# Metric Dimension

Richard C. Tillquist
richard.tillquist@colorado.edu

Rafael M. Frongillo
raf@colorado.edu

Manuel E. Lladser
manuel.lladser@colorado.edu

The **metric dimension** of a graph is the smallest number of vertices from which the vector of distances to every vertex in the graph is unique. It may be regarded as a generalization of the concept of trilateration in the two-dimensional real plane, the idea underpinning the Global Positioning System (GPS).

## 1 Definition

Let $G$ be a graph with vertex set $V$ and edge set $E$, and let $d(u, v)$ denote the shortest path or geodesic distance between two vertices $u, v \in V$. $G$ is not forced to be simple (though all examples in this article are) and may contain weighted edges, multi-edges, or self loops. A set $R \subseteq V$ is called **resolving** if for all $u, v \in V$ with $u \neq v$ there is at least one $r \in R$ such that $d(u, r) \neq d(v, r)$. In this case $r$ is said to *resolve* or *distinguish* $u$ and $v$. By definition, if an ordering on the vertices of $R = \{r_1, \ldots, r_n\}$ is given, any $u \in V$ may be uniquely represented by the vector $\Phi_R(u) := (d(u, r_1), \ldots, d(u, r_n))$ (see Figure 1). The **metric dimension** of $G$, denoted $\beta(G)$, is the smallest size of resolving sets on $G$; formally, $\beta(G) = \min\{|R| : R \subseteq V,\ R \text{ is resolving}\}$. If $R$ is a resolving set on $G$ and $|R| = \beta(G)$, $R$ is called a **minimal resolving set** of $G$, also called a *basis set*, or *reference set* [12, 23].

Intuitively, this concept is closely related to that employed by the Global Positioning System (GPS), called **trilateration**, where the location of any object on Earth can be determined by its distances to three satellites in orbit. More generally, given a point $x \in \mathbb{R}^2$, we may partition the space into equivalence classes of points with equal Euclidean distance to $x$, where $y, z \in \mathbb{R}^2$ belong to the same class if and only if $d(y, x) = d(z, x)$ (these classes form circles centered at $x$). A set of points $R \subset \mathbb{R}^2$ may be used to partition the space in a similar way. Now $y$ and $z$ belong to the same class if and only if $d(y, r) = d(z, r)$ for all $r \in R$. When $R$ contains a subset of three affinely independent points, every point in $\mathbb{R}^2$ belongs to its own equivalence class and $R$ may be said to resolve the plane.

## 2 Brute Force Calculation

Given an arbitrary graph $G = (V, E)$, the brute force method for determining $\beta(G)$ requires that every subset of $(\beta(G) - 1)$ vertices be established as non-resolving and that at least one resolving set of size $\beta(G)$ be found. Since $\beta(G) \leq |V| - 1$ [6], starting with sets of size one, $\sum_{k=1}^{|V|-2} \binom{|V|}{k} = O(2^{|V|})$ subsets must be examined in the worst case. In order to determine whether or not $R \subseteq V$ is resolving, every pair of vertices $u, v \in V$ must be compared across $|R|$ distances. This requires $O(|R||V|^2)$ time, bringing the total time necessary to find $\beta(G)$ to $|V|^2 \sum_{k=1}^{|V|-2} \binom{|V|}{k} k = O(|V|^3 2^{|V|})$.

The above assumes that all pairwise distances between nodes in $G$ have been precomputed. There are a host of algorithms for finding shortest path distances in graphs. When $G$ is directed and may have positive or negative edge weights, the Floyd-Warshall algorithm and Johnson's algorithm are among the most popular

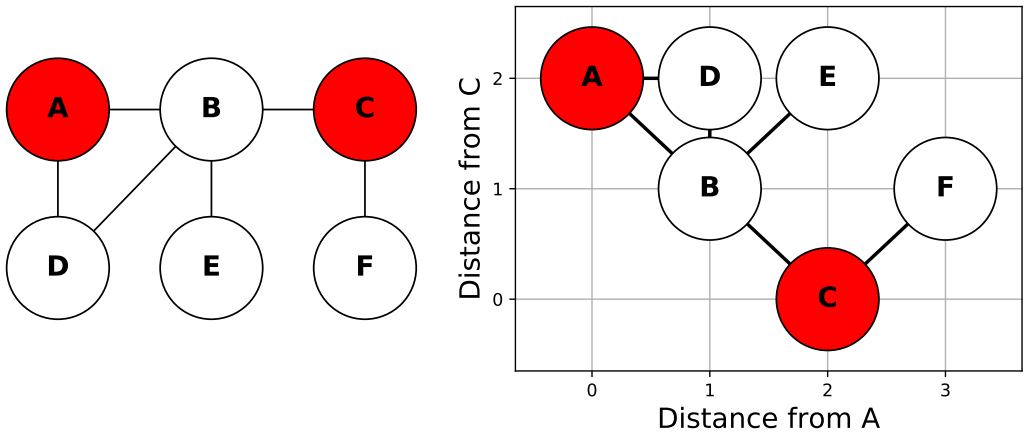


Figure 1: A graph with metric dimension 2 and minimal resolving set $R = \{A, C\}$. Based on this set, $\Phi_R(A) = (0, 2), \Phi_R(B) = (1, 1), \Phi_R(C) = (2, 0), \Phi_R(D) = (1, 2), \Phi_R(E) = (2, 2)$, and $\Phi_R(F) = (3, 1)$. This corresponds to the embedding of the graph in $\mathbb{R}^2$ on the right.

techniques. These have asymptotic run times $O(|V|^3)$ [9] and $O(|V||E| + |V|^2 \log |V|)$ [15], respectively. An algorithm based on a component hierarchy [25] can solve this problem in $O(|V||E| + |V|^2 \log \log |V|)$ time [22]. When $G$ is undirected and edge weights are guaranteed to take integer values, a similar approach can be used to determine all shortest path lengths in $O(|V||E|)$ time [25].

## 3 Complexity and Approximation Algorithms

The brute force approach to computing $\beta(G)$ is intractable even for small graphs. In fact, this problem is NP-hard and the associated decision problem, determining whether the metric dimension of a graph is less than a specified integer, has been shown to be NP-complete via reduction from 3-SAT [16] and 3-dimensional matching [10]. As a result, a number of algorithms for estimating metric dimension exist. Methods employing genetic algorithms [17] and a variable neighborhood search [21] can find small resolving sets but do not provide approximation guarantees which bound how far from $\beta(G)$ the result may be. The **Information Content Heuristic (ICH)**, on the other hand, ensures an approximation ratio of $1 + (1 + o(1)) \cdot \ln(|V|)$, the best possible ratio for metric dimension [13].

A brief description of the ICH algorithm follows. Let $u_R = \Phi_R(u)$ be the vector of distances from $u \in V$ to the elements of $R \subseteq V$. Let $S_R = \{u_R | u \in V\}$ be the set of all such vectors for a given graph and $B_R = [u_R | u \in V]$ be the bag or multiset associated with $S_R$. The ICH algorithm takes an information theoretic perspective, using $H(B_R)$, the discrete entropy over the multiset of vertex representations on $V$ imposed by $R$, to measure how far $R$ is from being resolving. Notice $H(B_R)$ is maximized precisely when $R$ is a resolving set, i.e. $|S_R| = |V|$ so that every vertex has a unique representation. At its core, the ICH algorithm is a greedy search for an $R$ achieving this maximum value, $H(B_R) = \log |V|$. Starting with $R_0 = \emptyset$, $R_i$ is built recursively by finding $v^* = \text{argmax}_{v \in V \setminus R_{i-1}} H(R_{i-1} \cup \{v\})$ and setting $R_i = R_{i-1} \cup \{v^*\}$.

With a run time complexity of $O(|V|^3)$, ICH is only practical for small and medium-sized graphs. Nevertheless, using parallel computing, it is possible to reduce the run time of the ICH algorithm further.

## 4 Metric Dimension of Specific Graph Families

While determining the exact metric dimension of an arbitrary graph is a computationally difficult problem, efficient algorithms, exact formulae, and useful bounds have been established for a variety of graphs. This section presents descriptions of the metric dimension of several common families of graphs. For a list of results related to the join and cartesian product of graphs, see [4].

**Fully Characterized Graphs:** Graphs on $n$ vertices with a metric dimension of 1, $(n-1)$, and $(n-2)$ have been fully characterized [6]. The first two cases are simple to describe:

- The metric dimension of a graph is 1 if and only if the graph is a path (see Figure 2).
- The metric dimension of a graph with $n$ nodes is $(n-1)$ if and only if the graph is the complete graph on $n$ nodes (see Figure 3).

For the third case, let us introduce notation, following [6]. Let $G \cup H$ be the **disjoint union** of two graphs $G$ and $H$, i.e. if $G = (V_1, E_1)$ and $H = (V_2, E_2)$, $G \cup H = (V_1 \sqcup V_2, E_1 \sqcup E_2)$, where $\sqcup$ denotes disjoint set union [1]. Further, let $G + H$ be the graph $G \cup H$ with additional edges joining every node in $G$ with every node in $H$. Finally, define $K_n$ to be the complete graph on $n$ nodes, $\overline{K_n}$ to be the graph with $n$ nodes and no edges, and $K_{n,m}$ to be the complete bipartite graph with partitions of size $n$ and $m$. Then the metric dimension of a graph with $n$ nodes is $(n-2)$ if and only if the graph is one of the following:

- $K_{s,t}$ with $s, t \geq 1$, and $n = s + t$.
- $K_s + \overline{K_t}$ with $s \geq 1$, $t \geq 2$, and $n = s + t$.
- $K_s + (K_1 \cup K_t)$ with $s, t \geq 1$, and $n = s + t + 1$.

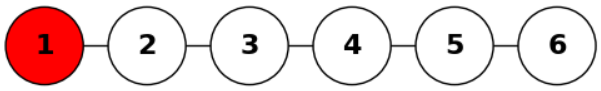


Figure 2: The path graph of size 6, $P_6$. $R = \{1\}$ is a minimal resolving set of this graph. In general, any set of the form $\{v\}$, with $v$ a node of degree 1 in $P_n$, is a minimal resolving set on $P_n$.

**Trees:** The introduction of metric dimension in the mid 1970s also brought a characterization of the metric dimension of trees, via a simple formula [12,23]. Let $T$ be a tree that is not a path and define $\ell(T)$ to be the

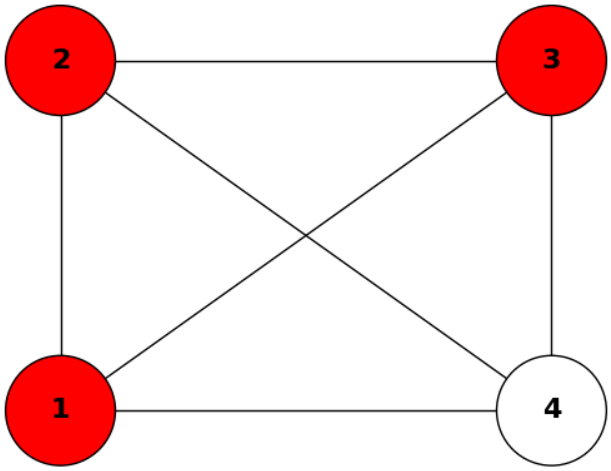


Figure 3: The complete graph of size 4, $K_4$. $R = \{1, 2, 3\}$ is a minimal resolving set of this graph. In general, any set of nodes of cardinality $(n-1)$ is a minimal resolving set of $K_n$.

number of **leaves** (nodes of degree 1) in $T$. Further, define $\sigma(T)$ as the number of **exterior major vertices** in $T$, that is vertices with degree at least 3 which are also connected to at least one leaf by a path of vertices of degree 2. Then the metric dimension of $T$ is $\beta(T) = \ell(T) - \sigma(T)$. A resolving set of this size may be constructed by taking the set of all leaves and removing exactly one element associated with each exterior major vertex [6] (see Figure 4). This construction may be carried out using a modified depth first search in $O(|V| + |E|)$ time.

**Hamming Graphs:** For positive integers $k$ and $a$, the Hamming graph $H_{k,a}$ consists of $a^k$ vertices each labeled with a unique string of length $k$ using an alphabet of size $a$. Two vertices in $H_{k,a}$ are adjacent when their labels differ in exactly one position; thus, the shortest path distance $d(u, v)$ is the total number of mismatches between the labels of $u$ and $v$ (i.e. the **Hamming distance** between $u$ and $v$). While determining $\beta(H_{k,a})$ exactly is difficult, it has been shown that, in general, $\beta(H_{a,k}) \leq \beta(H_{a,k+1}) \leq \beta(H_{a,k}) + \lfloor \frac{a}{2} \rfloor$. Furthermore, given a resolving set on $H_{k,a}$ of size $s$ it is possible to efficiently construct a resolving set on $H_{k+1,a}$ of size $s + \lfloor \frac{a}{2} \rfloor$ [26]. This implies that $\beta(H_{k,a})$ grows at most linearly with $k$ and allows small resolving sets to be generated despite how quickly Hamming graphs grow in size with increasing $k$.

Connections between coin weighing problems and $Q_k = H_{k,2}$, or hypercubes, lead to the asymptotic formula $\lim_{k\to\infty} \beta(Q_k)\frac{\log(k)}{k} = 2$ [8, 19]. Even with a binary alphabet, $\beta(Q_k)$ is known exactly only up to $k = 10$ (see Table 1).

| k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | *11* | *12* | *13* | *14* | *15* | *16* | *17* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta(Q_k)$ | 1 | 2 | 3 | 4 | 4 | 5 | 6 | 6 | 7 | 7 | *8* | *8* | *8* | *9* | *9* | *10* | *10* |

Table 1: Exact values of $\beta(Q_k)$ for $1 \leq k \leq 10$, and upper bounds for $11 \leq k \leq 17$ [21].

The Hamming graph $H_{k,a}$ may also be thought of as the **Cartesian product** of $k$ complete graphs of size $a$. That is, $H_{k,a} = K_a^{\square k} = K_a \square K_a \square \ldots \square K_a$, with $k$ copies of $K_a$. In general, $G \square H$, the Cartesian product of $G = (V_1, E_1)$ and $H = (V_2, E_2)$, has vertex set $V = \{(u, v) | u \in V_1, v \in V_2\}$ and edge set $E$

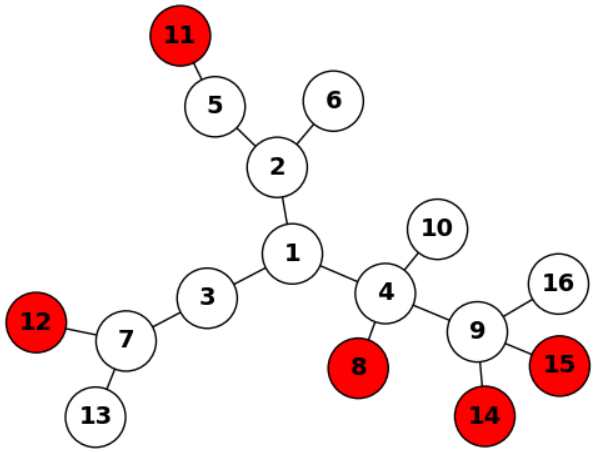


Figure 4: A tree of size 16. The vertices 2, 4, 7, and 9 are exterior major vertices and 6, 8, 10, 11, 12, 13, 14, 15, and 16 are leaves. Note that node 1 is not an exterior major vertex as every path from this vertex to a leaf includes at least one other vertex of degree greater than two. $R = \{8, 11, 12, 14, 15\}$ is a resolving set of minimum size.

defined as follows: $\{(u, v), (u', v')\} \in E$ if and only if $u = u'$ and $\{v, v'\} \in E_2$, or $v = v'$ and $\{u, u'\} \in E_1$. Working from this perspective, it has been shown that $\beta(H_{2,a}) = \lfloor \frac{2}{3}(2a-1) \rfloor$ [5]. This approach also allows a generalization of the asymptotic behavior of $\beta(Q_k)$ to show that $\lim_{k\to\infty} \beta(H_{k,a}) \frac{\log_a(k)}{k} = 2$. A similar asymptotic result holds for $G^{\square n}$ for other graphs $G$ including paths and cycles [14].

**Random Graphs:** In a study related to the graph isomorphism problem, it was shown that the set of $\lceil \frac{3\ln(n)}{\ln(2)} \rceil$ high degree vertices in a graph of size $n$ can be used to differentiate two random graphs with high probability [2]. Indeed, this set of nodes is highly likely to resolve the Erdös-Rényi random graph $G_{n,1/2}$. This bound has been generalized to encompass arbitrary values of $p$ so that, with high probability, $\beta(G_{n,p}) \leq \frac{-3\ln(n)}{\ln(p^2+(1-p)^2)}$ as $n$ goes to infinity and any set of nodes of this size resolves the graph with high probability [27]. Focusing closely on different regimes of $p$ as a function of the graph size, much more precise bounds on $\beta(G_{n,p})$ have been established [3].

Closely related to Erdös-Rényi random graphs are graphs generated via the Stochastic Block Model (SBM). This model groups a set of $n$ vertices into communities defined by a partition $C$ of $\{1, \ldots, n\}$. Adjacency probabilities for vertices in different communities are defined by a matrix $P$. By focusing on this adjacency information, general bounds on $G \sim SBM(n; C, P)$ have been established as have several efficient algorithms for finding small resolving sets on $G$ when $n$ is large enough to render the ICH algorithm impractical [27].

Random trees and forests have also been investigated with respect to metric dimension [20]. The exact formula and polynomial time algorithm for finding minimal resolving sets on trees allow the limiting distribution of $\beta(T_n)$, the metric dimension of a tree or forest chosen uniformly at random from all trees or forests of size $n$, to be determined precisely. In particular,

$$\frac{\beta(T_n) - \mu n(1 + o(1))}{\sqrt{\sigma^2 n(1 + o(1))}} \to N(0, 1),$$

where the convergence is in distribution as $n \to \infty$, and $\mu \simeq 0.14076941$ and $\sigma^2 \simeq 0.063748151$.

# 5 Applications

Despite the fact that finding minimal resolving sets of general graphs is computationally difficult, the ability to uniquely identify all vertices in a graph based on distances has proven to be quite useful. Applications regarding chemical structure [6] and robot navigation [16] have served as inspiration for the theoretical study of metric dimension. Deep connections between the metric dimension of Hamming graphs and a complete understanding and analysis of the game Mastermind [7] and various coin weighing problems [8, 19] have also been established. Resolving sets have proven valuable in a number of other applications as well.

**Source Localization:** Resolving sets are a natural tool to identify the source of a diffusion across a network. For instance, the ability to determine where a disease began as it spreads across a community has the potential to be valuable in a variety of contexts. If the time at which the spread began is known, and inter-node distances are deterministic and known, resolving sets give a direct solution. In more realistic settings, however, the notion of resolvability must be augmented to take into account an unknown start time and random transmission delays between nodes. The former may be addressed using doubly resolving sets. Whereas for every pair of different nodes $u, v \in V$ a resolving set $R \subseteq V$ need only contain a single element $r \in R$ such that $d(u, r) \neq d(v, r)$, a doubly resolving set $D \subseteq V$ must have nodes $r_1, r_2 \in D$ such that $d(u, r_1) - d(u, r_2) \neq d(v, r_1) - d(v, r_2)$. Successfully identifying the source of a spread is highly dependent on the variance associated with random inter-node distances [24].

**Representing Genetic Sequences:** Many machine learning algorithms assume numeric vectors as input. In contrast, sequences of nucleotides or amino acids from biological applications are symbolic in nature; as such, they must be transformed before they can be analyzed using machine learning techniques. One such transformation is an embedding based on resolving sets, which can be used to efficiently generate concise feature vectors for large sequences. In this approach, all possible sequences of length $k$ are encoded as nodes in a Hamming graph $H_{k,a}$, where $a$ is a reference alphabet size; given a resolving set $R$ of $H_{k,a}$, each vertex $v$ maps to the point $\Phi_R(v) \in \mathbb{R}^{|R|}$ (see Figure 1). For example, consider $H_{8,20}$, the Hamming graph used to represent amino acid sequences of length $k = 8$. This graph has approximately $25.6$ billion vertices and $1.9$ trillion edges, making many state-of-the-art graph embedding methods like multidimensional scaling [18] and Node2Vec [11] impractical. On the other hand, a resolving set of size $82$ is known for this graph, which was constructed by augmenting a resolving set for $H_{3,20}$ using bounds described in Section 4 [26]. This resolving set gives rise to an embedding into $\mathbb{R}^{82}$, whereas traditional techniques used to embed biological sequences, like binary vectors, require almost twice as many dimensions.

# 6 Acknowledgements

This article was partially funded by NSF IIS grant 1836914.

# References

[1] *Disjoint union*. `https://en.wikipedia.org/wiki/Disjoint_union`, Accessed: 2019-08-04.

[2] László Babai, Paul Erdös, and Stanley M Selkow, *Random graph isomorphism*, SIAM Journal on Computing **9** (1980), no. 3, 628–635.

[3] B. Bollobás, D. Mitsche, and P. Pralat, *Metric dimension for random graphs*, The Electronic Journal of Combinatorics **20** (2013), no. 4.

[4] José Cáceres, Carmen Hernando, Merce Mora, Ignacio M Pelayo, Maria L Puertas, Carlos Seara, and David R Wood, *On the metric dimension of some families of graphs*, Electronic Notes in Discrete Mathematics **22** (2005), no. 2, 129–133.

[5] _____, *On the metric dimension of cartesian products of graphs*, SIAM Journal on Discrete Mathematics **21** (2007), no. 2, 423–441.

[6] Gary Chartrand, Linda Eroh, Mark A Johnson, and Ortrud R Oellermann, *Resolvability in graphs and the metric dimension of a graph*, Discrete Applied Mathematics **105** (2000), no. 1, 99–113.

[7] Vasek Chvátal, *Mastermind*, Combinatorica **3** (1983), no. 3-4, 325–329.

[8] Paul Erdös and Alfréd Rényi, *On two problems of information theory*, Magyar Tud. Akad. Mat. Kutató Int. Közl **8** (1963), 229–243.

[9] Robert W Floyd, *Algorithm 97: shortest path*, Communications of the ACM **5** (1962), no. 6, 345.

[10] Michael R Garey and David S Johnson, *Computers and intractability: A guide to the theory of NP-completeness*, WH Freeman and Company, New York, 1979.

[11] Aditya Grover and Jure Leskovec, *Node2vec: Scalable feature learning for networks*, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 855–864.

[12] Frank Harary and Robert A Melter, *On the metric dimension of a graph*, Ars Combinatoria **2** (1976), no. 191-195, 1.

[13] Mathias Hauptmann, Richard Schmied, and Claus Viehmann, *Approximation complexity of metric dimension problem*, Journal of Discrete Algorithms **14** (2012), 214–222.

[14] Zilin Jiang and Nikita Polyanskii, *On the metric dimension of cartesian powers of a graph*, Journal of Combinatorial Theory, Series A **165** (2019), 1–14.

[15] Donald B Johnson, *Efficient algorithms for shortest paths in sparse networks*, Journal of the ACM (JACM) **24** (1977), no. 1, 1–13.

[16] Samir Khuller, Balaji Raghavachari, and Azriel Rosenfeld, *Landmarks in graphs*, Discrete Applied Mathematics **70** (1996), no. 3, 217–229.

[17] Jozef Kratica, Vera Kovačević-Vujčić, and Mirjana Čangalović, *Computing the metric dimension of graphs by genetic algorithms*, Computational Optimization and Applications **44** (2009), no. 2, 343–361.

[18] Wojtek J Krzanowski, *Principles of multivariate analysis: A user's perspective*, OUP Oxford, 2000.

[19] Bernt Lindström, *On a combinatory detection problem I*, I. Magyar Tud. Akad. Mat. Kutató Int. Közl **9** (1964), 195–207.

[20] Dieter Mitsche and Juanjo Rué, *On the limiting distribution of the metric dimension for random forests*, European Journal of Combinatorics **49** (2015), 68–89.

[21] Nenad Mladenović, Jozef Kratica, Vera Kovačević-Vujčić, and Mirjana Čangalović, *Variable neighborhood search for metric dimension and minimal doubly resolving set problems*, European Journal of Operational Research **220** (2012), no. 2, 328–337.

[22] Seth Pettie, *A new approach to all-pairs shortest paths on real-weighted graphs*, Theoretical Computer Science **312** (2004), no. 1, 47–74.

[23] Peter J Slater, *Leaves of trees*, Congressus Numerantium **14** (1975), no. 549-559, 37.

[24] Brunella Marta Spinelli, Elisa Celis, and Patrick Thiran, *Observer placement for source localization: the effect of budgets and transmission variance*, 54th Annual Allerton Conference on Communication, Control, and Computing, 2016.

[25] Mikkel Thorup, *Undirected single-source shortest paths with positive integer weights in linear time*, Journal of the ACM (JACM) **46** (1999), no. 3, 362–394.

[26] Richard C. Tillquist and Manuel E. Lladser, *Low-dimensional representation of genomic sequences*, Journal of Mathematical Biology **79** (2019), no. 1, 1–29.

[27] _____, *Multilateration of random networks with community structure* (2019). In progress.