Reconstructing a point set from a random subset of its pairwise distances

António Girão[†] Freddie Illingworth[†] Lukas Michel[†]
Emil Powierski[†] Alex Scott[†]

Abstract

Let V be a set of n points on the real line. Suppose that each pairwise distance is known independently with probability p. How much of V can be reconstructed up to isometry?

We prove that $p = (\log n)/n$ is a sharp threshold for reconstructing all of V which improves a result of Benjamini and Tzalik. This follows from a hitting time result for the random process where the pairwise distances are revealed one-by-one uniformly at random. We also show that $1/n$ is a weak threshold for reconstructing a linear proportion of V.

1 Introduction

Let $V \subset \mathbb{R}^1$ be a finite set of points on the real line and suppose that all we know about the points are the distances between some pairs $\mathcal{P} \subset \binom{V}{2}$ of them. More precisely, labelling the points $v_1, \dots, v_n$, if the pair $v_i v_j \in \mathcal{P}$, then the distance between v_i and v_j is given[1]. How much of V can be reconstructed? Can all of V be reconstructed? *Reconstructing* a set $U \subseteq V$ means deducing the positions of the labelled points in U up to isometry or, equivalently, deducing all pairwise distances of points in U. In this paper we consider the case where each pairwise distance is known independently with probability p, that is, $(V, \mathcal{P})$ is distributed as the Erdős-Renyi binomial random graph $\mathcal{G}(n, p)$.

Benjamini and Tzalik [BT22] recently proved the following result.

Theorem 1.1 (Benjamini and Tzalik [BT22])**.** *Let V be a set of n points on the real line. There is a sufficiently large constant C such that if the graph G of known pairwise distances $(V, \mathcal{P})$ is distributed as $\mathcal{G}(n, p)$ where $p = C\log(n)/n$, then the whole of V can be reconstructed with high probability (whp).*

2020 MSC: 05C80 (Random graphs), 52C25 (Rigidity), 05C60 (Isomorphism problems in graph theory)

[†]Mathematical Institute, University of Oxford, United Kingdom ({girao,illingworth,michel,powierski, scott}@maths.ox.ac.uk). Research of AG, FI, and AS supported by EPSRC grant EP/V007327/1.

[1]The problem where the points are not identified and only the multiset of pairwise distances is given has also been studied, see [ACKR89, PRS03]. This is not in general enough to reconstruct a point set, as $A + B$ and $A - B$ are not distinguishable. However, sets in at most two dimensions are reconstructible from subsets of constant size.

We strengthen this result. Our first result identifies two important landmarks in the evolution of the largest reconstructible set as p increases. We improve on Theorem 1.1 by proving a sharp threshold result for reconstructing the whole of V, as well as identifying the threshold for reconstructing a positive fraction of V.

Theorem 1.2. *Let V be a set of n points on the real line. Suppose the graph G of known pairwise distances $(V, \mathcal{P})$ is distributed as $\mathcal{G}(n,p)$. Then the following hold whp.*

a. If $p \geqslant 42/n$, then there is a reconstructible set of size $\Omega(n)$.
b. If $pn \to \infty$, then there is a reconstructible set of size $(1-o(1))n$.
c. If $p \leqslant \frac{\log n + \log\log n - \omega(1)}{n}$, then it is not possible to reconstruct the whole of V.
d. If $p \geqslant \frac{\log n + \log\log n + \omega(1)}{n}$, then the whole of V can be reconstructed.

These results are best possible up to the constant factor in part a[2]. Indeed, when $p = c/n$ for constant $c < 1$, the largest connected component in G has size $\mathcal{O}(\log n)$. When $p = c/n$ for constant $c > 1$, whp the largest component of G is a giant component of size at most dn for some constant $d < 1$.

The final two parts of Theorem 1.2 follow from a stronger hitting time result. Suppose that the pairwise distances are revealed one-by-one in a random order: in other words, the graph of known pairwise distances follows the random graph process $(G_t : 0 \leqslant t \leqslant \binom{n}{2})$. We will prove that whp V is reconstructible when G_t has minimum degree at least two. However, V may also be reconstructible slightly before this. If some point u is incident to exactly one revealed distance, then u has two possible positions relative to its neighbour but it may be the case that one of these possible positions is already occupied and so in fact u can be reconstructed. To this end, we say an ordered pair of distinct points (u,v) is *secure* if $w = 2v - u$, the reflection of u over v, is also in V.

Theorem 1.3. *Let V be a set of n points on the real line. Suppose the distances between pairs of points in V are revealed one-by-one in a uniformly random order. Then whp V is reconstructible exactly at the first time that both the following hold.*

- *Every point is incident to at least one revealed distance.*
- *If a point u is incident to exactly one revealed distance, which is to point v, then (u,v) is secure.*

It follows that if every point is incident to at least two revealed distances, then whp the whole of V is reconstructible. In fact, this moment is whp the hitting time for full reconstructibility if and only if there are $o(n^2)$ secure pairs in V. This is because the unique distance incident to the final point with only one revealed distance is uniformly random amongst $\binom{V}{2}$. An example of a set with $\Omega(n^2)$ secure pairs is $V = \{1, 2, \dots, n\}$. In the other direction, we will show (Theorem 3.4) that if there are $\omega(1)$ points incident to only one revealed distance, then whp the whole of V is not reconstructible.

It is also possible to obtain an algorithmic hitting time result that does not need knowledge of the underlying point set V to determine the point at which the entirety of V is reconstructible.

Theorem 1.4. *There is an online algorithm with polynomial expected running time in n*

[2]We have not optimised the value 42 in part a: a short calculation shows the best our methods could give is 9, while we believe the correct threshold is $1+\varepsilon$ (see Section 4 for further discussion).

that takes the revealed distances one-by-one and whp recognises the first time at which V is reconstructible and reconstructs it.

Reconstructibility is strongly related to graph rigidity, which is concerned with generic embeddings of graphs in $\mathbb{R}^d$. An embedding of a set of vertices V in $\mathbb{R}^d$ is *generic* if the set of $d|V|$ coordinates of the vertices is algebraically independent over the rationals. A graph is *globally rigid in* $\mathbb{R}^d$ if it has some generic embedding in $\mathbb{R}^d$ which is reconstructible from its edge lengths. In fact, a graph is globally rigid if and only if all of its generic embeddings are reconstructible from their edge lengths [Con05, GHT10].

The (global) rigidity of the random graph in $\mathbb{R}^d$ has been extensively studied with work both on when the whole graph is (globally) rigid [JSS07, KT13, BDG+21, JT22] and when it has a linear sized rigid component [KMT11, BLM18]. Lew, Nevo, Peled, and Raz [LNPR22] recently gave a hitting time result: the random graph process in $\mathbb{R}^d$ becomes globally rigid at exactly the moment it has minimum degree $d+1$. This implies Theorem 1.3 for *generic* embeddings since there are no secure pairs in generic embeddings. However, the restriction to generic embeddings in the definition of global rigidity is a significant weakening. For example, it is folklore (see [JW17, Thm. 63.2.7]) that a graph is globally rigid in $\mathbb{R}$ if and only if it is 2-connected, while for reconstructing arbitrary point sets, the situation is rather different. Indeed, we show that there are graphs with arbitrarily high connectivity which can be embedded in $\mathbb{R}$ so that their vertex sets cannot be reconstructed from their edge lengths. This disproves a conjecture of Benjamini and Tzalik [BT22].

Theorem 1.5. *Let k be a positive integer. There are k-connected graphs on arbitrarily many vertices which can be embedded in the real line in such a way that the largest reconstructible subset of vertices has size k.*

Proof. Fix a positive integer k and let $n = \ell k$ be any multiple of k. Let $C_1, \dots, C_\ell$ be ℓ disjoint k-cliques. Write $C_i = \{a_1^{(i)}, \dots, a_k^{(i)}\}$. Let G be the graph obtained from the union of $C_1, \dots, C_\ell$ by adding matchings $\{a_1^{(i)}a_1^{(i+1)}, a_2^{(i)}a_2^{(i+1)}, \dots, a_k^{(i)}a_k^{(i+1)}\}$ between C_i and C_{i+1} for each $i = 1, \dots, \ell - 1$. Note that G is the Cartesian product of K_k and P_ℓ and is certainly k-connected.

Embed the C_i so that $a_1^{(i)}, \dots, a_k^{(i)}$ are k consecutive integers (in that order), that is, $|a_s^{(i)} - a_t^{(i)}| = |s - t|$. We specify that the length of each edge $a_s^{(i)}a_s^{(i+1)}$ is k^i. Certainly such an embedding is possible by placing each $a_s^{(i)}$ at $k^{i-1} + k^{i-2} + \cdots + k + s$. Further, having placed $C_1, \dots, C_i$ at these points, there are multiple options for where to place $C_{i+1}, \dots, C_\ell$: either to the right of C_i or to the left of C_1. In particular, any vertex subset of size at least $k+1$ contains vertices from at least two distinct C_i and is thus not reconstructible. □

The rest of the paper is organized as follows. We establish parts a and b of Theorem 1.2 in Section 2. The remainder of Theorem 1.2 as well as Theorems 1.3 and 1.4 are proved in Section 3. We conclude in Section 4 with some discussion of open problems.

Throughout we use standard asymptotic notation and assume that n is sufficiently large.

2 Reconstructing a linear proportion of the points

In this section, we prove parts a and b of Theorem 1.2. Fix a set V of n points on the real line. Our method for reconstructing a linear portion of V is via short cycles. We first show that, for most short cycles C whose vertices are in V, if the edges of C were all revealed distances, then the vertices of C can be reconstructed up to isometry. This will follow from the fact that multiple possible embeddings of a short cycle on the real line would yield a small "linear dependence" among the pairwise distances of adjacent points on the cycle. We show in Section 2.1 that this linear dependence can only occur for a small proportion of short cycles.

Then in Section 2.2 we assume that each pairwise distance is known independently with probability $p = 42/n$. We note that a random graph $G \sim \mathcal{G}(n, 42/n)$ will typically have $\Omega(n^2)$ pairs of vertices that are contained in a short cycle. Since most short cycles can be reconstructed, we can reconstruct the distances between $\Omega(n^2)$ pairs of points. This allows us to apply the following extremal result of Benjamini and Tzalik [BT22] to reconstruct a linear proportion of all points and so prove part a.

Theorem 2.1 (Benjamini and Tzalik [BT22])**.** *Let V be a set of n points on the real line. Suppose that the set of known pairwise distances $\mathcal{P}$ has size greater than $40n^{3/2}$. Then there is a set of $\Omega(|\mathcal{P}|/n)$ points that can be reconstructed.*

Finally, to obtain part b, we start with this linear proportion of all points that we have already reconstructed. Then, by increasing p and so sprinkling in some additional edges, we will find that for most points we will know at least two distances to the set that we have already reconstructed, and so we can reconstruct those points as well.

2.1 Reconstructible tuples in embeddings

In this section we show that most short cycles, if their edges are revealed distances, can be reconstructed. To this end we make the following definition.

Definition 2.2. A tuple $T = (v_1, v_2, \dots, v_k)$ of distinct points in V is *cycle-reconstructible* if, given just the pairwise distances $|v_2 - v_1|$, …, $|v_k - v_{k-1}|$, and $|v_1 - v_k|$, we can reconstruct $\{v_1, \dots, v_k\}$ up to isometry.

Consider some k-tuple $T = (v_1, v_2, \dots, v_k)$ and let $d_1 = |v_2 - v_1|, \dots, d_{k-1} = |v_k - v_{k-1}|$, and $d_k = |v_1 - v_k|$. Certainly there are some $\varepsilon_1, \dots, \varepsilon_{k-1} \in \{-1, 1\}$ such that

$$\sum_{i=1}^{k-1} \varepsilon_i d_i = d_k. \tag{1}$$

If T is not cycle-reconstructible, then there are at least two $(\varepsilon_1, \dots, \varepsilon_{k-1}) \in \{-1, 1\}^{k-1}$ satisfying (1). Subtracting two such expressions and halving the result shows that there is some non-zero vector $(\gamma_1, \dots, \gamma_{k-1}) \in \{-1, 0, 1\}^{k-1}$ with

$$\sum_{i=1}^{k-1} \gamma_i d_i = 0. \tag{2}$$

This allows us to bound the number of non-cycle-reconstructible k-tuples.

Lemma 2.3. *Let $k \leqslant 0.9 \log n$. At most an $n^{-0.01}$ fraction of all k-tuples are not cycle-reconstructible.*

Proof. Select a k-tuple $T = (v_1, v_2, \ldots, v_k)$ of distinct points in V uniformly at random. Let $d_1 = |v_2 - v_1|, \ldots, d_{k-1} = |v_k - v_{k-1}|$. We say that v_{s+1} *fails* (for $1 \leqslant s \leqslant k-1$) if there is some non-zero vector $(\gamma_1, \ldots, \gamma_{s-1}) \in \{-1, 0, 1\}^{s-1}$ with $\sum_{i<s} \gamma_i d_i = d_s$. By the preceding discussion, if T is not cycle-reconstructible, then some v_{s+1} fails.

Given $v_1, \ldots, v_s$ there are at most 3^{s-1} values of $d_s = |v_{s+1} - v_s|$ for which there is a vector $(\gamma_1, \ldots, \gamma_{s-1}) \in \{-1, 0, 1\}^{s-1}$ with $d_s = \sum_{i<s} \gamma_i d_i$. Each such d_s corresponds to two possible values of v_{s+1} (namely $v_s \pm d_s$). In particular, as there are $n - s$ possible choices for v_{s+1},

$$\mathbb{P}(v_{s+1} \text{ fails}) \leqslant \frac{2 \cdot 3^{s-1}}{n-s} \leqslant \frac{3^s}{n},$$

and so, taking a union bound,

$$\mathbb{P}(T \text{ not cycle-reconstructible}) \leqslant \sum_{s \leqslant 0.9 \log n} \frac{3^s}{n} \leqslant \frac{3^{0.9 \log n + 1}}{2n} = \frac{3}{2} n^{0.9 \log 3 - 1} \leqslant n^{-0.01}. \quad \square$$

We say that a pair of points u, v is *k-bad* if greater than an $n^{-0.005}$ fraction of the k-tuples containing both u and v are not cycle-reconstructible. Given the preceding lemma, it follows that there are only few k-bad pairs of points.

Lemma 2.4. *Let $k \leqslant 0.9 \log n$. At most an $n^{-0.005}$ fraction of all pairs are k-bad.*

Proof. Note that every pair of points is in the same number of k-tuples – call this common value A. Let u, v be a uniformly random pair of points in V and let X be the number of non-cycle-reconstructible k-tuples containing both u and v. We will double count the number of $(T, \{u, v\})$ where T is a non-cycle-reconstructible k-tuple and the pair u, v is contained in T. By Lemma 2.3, this number is at most

$$n^{-0.01} \cdot \text{number of } k\text{-tuples} \cdot \tbinom{k}{2} = n^{-0.01} \cdot A \tbinom{n}{2}.$$

On the other hand, the number is exactly equals $\mathbb{E}(X) \cdot \binom{n}{2}$. Hence $\mathbb{E}(X) \leqslant A n^{-0.01}$. Thus, by Markov's inequality,

$$\mathbb{P}(u, v \text{ is } k\text{-bad}) = \mathbb{P}(X > A n^{-0.005}) \leqslant \frac{\mathbb{E}(X)}{A n^{-0.005}} \leqslant n^{-0.005}. \quad \square$$

If a pair of points u, v is not k-bad for any $k \leqslant 0.9 \log n$, then a random short cycle in $\mathcal{G}(n, p)$ containing u and v is likely to allow us to reconstruct the distance between u and v. This motivates the following definition.

Definition 2.5. A pair of points u, v is *useless* if there is some $k \leqslant 0.9 \log n$ such that the pair u, v is k-bad. Otherwise u, v is *useful*.

Using the previous results, we can bound the number of useless pairs of points.

Corollary 2.6. *There are at most $n^{2-0.004}$ useless pairs of points.*

Proof. By Lemma 2.4, the number of useless pairs of points is at most

$$\sum_{k \leqslant 0.9 \log n} \text{number of } k\text{-bad pairs} \leqslant (0.9 \log n) n^{2-0.005} \leqslant n^{2-0.004}. \quad \square$$

2.2 Reconstructing from short cycles in $\mathcal{G}(n,p)$

Given the results from the previous section, it will be helpful to show that most pairs of points in $\mathcal{G}(n,p)$ are contained in a short cycle. We will use the following straightforward result about random graphs.

Lemma 2.7. *A random graph $G \sim \mathcal{G}(n, 42/n)$ contains whp $\Omega(n^2)$ pairs of vertices that are in a cycle of length at most $0.9 \log n$.*

Let $\mathrm{diam}(G)$ denote the largest distance between two vertices in the same component of G. We will use the following consequence of a (much more general) result by Riordan and Wormald [RW10, Thm. 1.1]. There are alternative elementary arguments that prove results similar to Lemma 2.7, but we use this result for brevity.

Lemma 2.8. *A random graph $G \sim \mathcal{G}(n, 21/n)$ satisfies whp $\mathrm{diam}(G) \leqslant 0.44 \log n$.*

Proof of Lemma 2.7. Let $V = V_1 \cup V_2$ with $|V_1| = |V_2| = n/2$. First reveal the edges inside V_1 and V_2, and let C_1 and C_2 be the largest components in $G[V_1]$ and $G[V_2]$ respectively. $G[V_1]$ and $G[V_2]$ are independently distributed as $\mathcal{G}(n/2, 21/(n/2))$.

By standard facts about random graphs (see, for example, [Bol01, Ch. 6]), C_1 and C_2 both have size $\Omega(n)$ whp. Also, Lemma 2.8 implies that whp C_1 and C_2 have diameter at most $0.44 \log n$.

Next, we reveal the edges between V_1 and V_2. For a fixed vertex $v \in C_1$ write X_v for the indicator random variable of the event that v has a neighbour in C_2. We have $\mathbb{P}(X_v = 0) = (1 - 42/n)^{|C_2|} \leqslant e^{-42|C_2|/n}$ which is at most some constant $a < 1$. Since the variables X_v for $v \in C_1$ are independent, by Chernoff, we have whp that $\sum_v X_v \geqslant (1-a)|C_1|/2$, and so the set A of vertices in C_1 that have a neighbour in C_2 has size at least $\Omega(n)$. Now, for any pair of vertices $u, v \in A$, concatenating the shortest path between them in C_1, their respective edges to vertices $w_u, w_v \in C_2$, and the shortest path between w_u and w_v in C_2 gives a cycle containing u and v of length at most $0.44 \log n + 0.44 \log n + 2 \leqslant 0.9 \log n$. □

Recall that $\mathcal{P}$ is the set of known distances. We say that a pair of points u, v is *close* if u and v are in a cycle in $(V, \mathcal{P})$ of length at most $0.9 \log n$. We say that the pair u, v is *deducible* if $|u - v|$ can be uniquely determined from $\mathcal{P}$. Note that if some cycle-reconstructible cycle in $\mathcal{P}$ contains both u and v, then the pair u, v is deducible. With this in mind, we bound the number of useful pairs of points that are in a short cycle, but where we cannot determine their distance.

Lemma 2.9. *If $(V, \mathcal{P})$ is distributed as $\mathcal{G}(n,p)$, then whp there are at most $n^{2-0.002}$ pairs of points that are close and useful but not deducible.*

Proof. Fix a pair of points $u, v \in V$ and let $k \leqslant 0.9 \log n$. Conditioned on the event that $(V, \mathcal{P})$ contains a k-cycle which contains u and v, the probability that any particular k-tuple containing u and v appears in $(V, \mathcal{P})$ as a cycle is uniform. If u, v is useful, then at most an $n^{-0.005}$ fraction of these k-tuples is non-reconstructible, and so

$$\mathbb{P}(u, v \text{ not deducible} \mid u, v \text{ in a } k\text{-cycle and useful}) \leqslant n^{-0.005}.$$

This holds for all $k \leqslant 0.9 \log n$ and so $\mathbb{P}(u, v \text{ not deducible} \mid u, v \text{ close and useful}) \leqslant n^{-0.005}$. Hence,

$$\mathbb{P}(u, v \text{ close and useful but not deducible}) \leqslant n^{-0.005}.$$

Let X denote the number of pairs which are close and useful but not deducible. Then $\mathbb{E}(X) \leqslant n^{-0.005} \cdot \binom{n}{2}$. By Markov's inequality,

$$\mathbb{P}(X > n^{2-0.002}) \leqslant \frac{\mathbb{E}(X)}{n^{2-0.002}} \leqslant n^{-0.003}. \qquad \square$$

We are now ready to prove the statements a and b of Theorem 1.2.

Proof of Theorem 1.2a. By Lemma 2.7, whp there are $\Omega(n^2)$ close pairs of points in $(V, \mathcal{P})$. Of the close pairs, at most $n^{2-0.004}$ pairs are useless by Corollary 2.6 and whp at most $n^{2-0.002}$ are useful but not deducible by Lemma 2.9. In particular, whp at least $\Omega(n^2)$ pairs of vertices in $(V, \mathcal{P})$ are deducible.

Finally, Theorem 2.1 implies that whp there is a set of $\Omega(n)$ vertices that can be reconstructed up to isometry. $\square$

Note that if we have reconstructed a set S of points and we know at least two distances from another point v to some points in S, this uniquely determines the position of v relative to S, and so we can reconstruct $S \cup \{v\}$. This allows us to prove part b.

Proof of Theorem 1.2b. By Theorem 1.2a, there is a constant $c > 0$ such that for $p = 42/n$ whp there is a set R of at least cn points that can be reconstructed. Let $d > c$ be a constant, $q = d/(cn)$ and consider $G \sim \mathcal{G}(n, q)$. Every vertex $v \notin R$ satisfies

$$\mathbb{P}(|N_G(v) \cap R| < 2) = (1-q)^{|R|} + |R|q(1-q)^{|R|-1} \leqslant 2nq(1-q)^{cn} \leqslant (2d/c)e^{-d}.$$

Since all of these events are independent, a Chernoff bound implies that whp at most $(4d/c)e^{-d}n$ points $v \notin R$ have at most one neighbour in R. Because every point with two neighbours in R can also be reconstructed relative to R, it follows that whp a set of at least $(1 - (4d/c)e^{-d})n$ points can be reconstructed in $\mathcal{G}(n, p) \cup \mathcal{G}(n, q)$ and therefore also in $\mathcal{G}(n, p+q) \sim \mathcal{G}(n, \omega/n)$ where $\omega = 42 + d/c$. Since $(4d/c)e^{-d} \to 0$ when $d \to \infty$, this proves the claim. $\square$

3 Hitting time for complete reconstruction

In this section, we prove Theorem 1.3. For this, we use the following standard coupling of random graphs. Let $(U_e)_{e \in E(K_n)}$ be a collection of independent random variables all uniformly distributed on $[0, 1]$. For $p \in [0, 1]$, let G_p be the subgraph of K_n whose edges are exactly those edges e with $U_e \leqslant p$. Note that G_p is distributed as $\mathcal{G}(n, p)$ and, as p is increased continuously from 0 to 1, G_p evolves as a random graph process where edges are added one-by-one in a uniformly random order (note that with probability 1 all U_e are distinct). We use this to model the process of revealing distances between pairs of points in V one-by-one uniformly at random. We will freely use the following facts about $\mathcal{G}(n, p)$ in this section (see, for example, [Bol01, Ch. 7]). For any positive

integer k: if $p \leqslant (\log n + (k-1)\log\log n - \omega(1))/n$, then whp $\mathcal{G}(n,p)$ has minimum degree at most $k-1$ and if $p \geqslant (\log n + (k-1)\log\log n + \omega(1))/n$, then whp $\mathcal{G}(n,p)$ has minimum degree at least k.

Three particular values of p will be important for our analysis. Theorem 1.2b implies that there is some large constant ω such that whp there is a set $R \subseteq V$ of size at least $0.9n$ that can be reconstructed in $G_{\omega/n}$. Let $p_1 = \omega/n$. Let $p_2 = p_1 + (0.9\log n)/n$ and $p_3 = (2\log n)/n$. Note that whp G_{p_2} still contains isolated vertices and so the whole of V is not yet reconstructible. We will show that the whole of V is reconstructible in G_{p_3} and so our analysis will focus on the range $[p_2, p_3]$.

Lemma 3.1. *Whp, for every edge uv in G_{p_3} with $u, v \notin R$, at least one of u, v has at least two neighbours among R in G_{p_2}.*

Proof. First, reveal all edges in G_{p_1}. This determines R and a Chernoff bound implies that whp G_{p_1} has at most $2p_1\binom{n}{2} \leqslant \omega n$ edges. Now, reveal the edges of G_{p_3} in the complement of R. This adds every edge independently with a probability of at most p_3 to the graph G_{p_1}. Hence, a Chernoff bound implies that whp this adds at most $2p_3\binom{n}{2} \leqslant 2n\log n$ edges to the complement of R, and so G_{p_3} has in total at most $\omega n + 2n\log n \leqslant 3n\log n =: m$ edges in the complement of R.

Finally, reveal the edges of G_{p_2} between R and the complement of R. Each such edge was already present in G_{p_1} or is added independently with probability at least $p := p_2 - p_1 = (0.9\log n)/n$ to the graph G_{p_1}. Therefore, every point $v \notin R$ satisfies

$$\begin{aligned}\mathbb{P}(|N_{G_{p_2}}(v) \cap R| < 2) &\leqslant (1-p)^{|R|} + |R|p(1-p)^{|R|-1} \leqslant 2np(1-p)^{|R|} \\ &\leqslant 2np(1-p)^{0.9n} \leqslant (2np)e^{-0.9pn} \leqslant (2\log n)n^{-0.81} \leqslant n^{-0.8}.\end{aligned}$$

Since these events are independent between distinct points $u, v \notin R$, this implies

$$\mathbb{P}(\max(|N_{G_{p_2}}(u) \cap R|, |N_{G_{p_2}}(v) \cap R|) < 2) \leqslant (n^{-0.8})^2 = n^{-1.6}.$$

Let X denote the number of edges uv in G_{p_3} such that $u, v \notin R$, but u and v both have at most one neighbour among R in G_{p_2}. Since G_{p_3} has at most m edges in the complement of R, this implies that $\mathbb{E}(X) \leqslant m \cdot n^{-1.6} \leqslant 3n^{-0.6}\log n \leqslant n^{-0.5}$, and so $\mathbb{P}(X \geqslant 1) \leqslant n^{-0.5}$. □

If a vertex v has two neighbours among a reconstructible set, then the position of v can be reconstructed with respect to that set. With this in mind, we define, for each $p \in [0,1]$, R'_p and R''_p as follows.

$$\begin{aligned}R'_p &= R \cup \{v \in V : |N_{G_p}(v) \cap R| \geqslant 2\}, \\ R''_p &= R'_p \cup \{v \in V : |N_{G_p}(v) \cap R'_p| \geqslant 2\}.\end{aligned}$$

We now collect some important facts about the R'_p and R''_p.

Lemma 3.2. *With high probability the following all hold.*

a. *For all $p \geqslant p_1$, R'_p and R''_p are reconstructible in G_p.*
b. *For all $p \in [p_2, p_3]$, every edge in G_p is incident to a vertex in R'_p.*
c. *For all $p \in [p_2, p_3]$, R''_p contains every vertex of degree at least two in G_p.*

Proof. Note that R is reconstructible in G_{p_1}. For any $p \geqslant p_1$, G_p contains G_{p_1} and so R is reconstructible in G_p. R'_p consists of R and all vertices with at least two neighbours in R and so is reconstructible in G_p. Similarly for R''_p. This proves a.

By Lemma 3.1, whp every edge in G_{p_3} has an end-point in R'_{p_2}. Now let $p \in [p_2, p_3]$ and let uv be an edge of G_p. Then uv is an edge of G_{p_3}. Hence, at least one of u, v is in $R'_{p_2} \subseteq R'_p$. This proves b.

Let u be a vertex of degree at least two in G_p. By b, either u is in R'_p or every neighbour of u is in R'_p. Either way, u is in R''_p which proves c. □

Now whp $\mathcal{G}(n, p_3)$ has minimum degree at least two and so, by the time that the random graph process from the coupling reaches minimum degree two, Lemma 3.2 tells us that whp the whole of V can be reconstructed. Theorem 1.2d follows immediately.

We say that an ordered pair of points (u, v) is *uncertain* in G_p if uv is the only edge incident to u and the point $w = 2v - u$ is a point of V with degree one in G_p. We call u uncertain in G_p, if (u, v) is uncertain in G_p for some $v \in V$.

Lemma 3.3. *Whp, for all $p \in [p_2, p_3]$, G_p contains no uncertain point.*

Proof. Let (u, v) be an ordered pair of points such that $w = 2v - u$ is a point of V. If (u, v) is uncertain in G_p for some $p \in [p_2, p_3]$, then there is a point x such that the edges uv and wx are present in $G_p \subseteq G_{p_3}$, but every other edge incident to u and w cannot be present in $G_p \supseteq G_{p_2}$. In particular, we get that

$$\begin{aligned}\mathbb{P}((u, v) \text{ is uncertain in some } G_p) &\leqslant np_3^2(1 - p_2)^{2n-5} \leqslant 2np_3^2(1 - p_2)^{2n} \\ &\leqslant (2np_3^2)e^{-2np_2} \leqslant 8(\log n)^2 n^{-1} n^{-1.8} \leqslant n^{-2.5}.\end{aligned}$$

Let X denote the number of ordered pairs of points (u, v) which are uncertain in some G_p. Since there are at most n^2 ordered pairs of points, this implies that $\mathbb{E}(X) \leqslant n^2 \cdot n^{-2.5} = n^{-0.5}$, and so $\mathbb{P}(X \geqslant 1) \leqslant n^{-0.5}$. □

We say that a point u is *undecidable* in G_p if it has degree one in G_p and (u, v) is not secure where v is the unique neighbour of u. Note that Theorem 1.3 says that whp G_p is reconstructible exactly when it has no isolated nor undecidable points.

Proof of Theorem 1.3. Using Lemma 3.2 and Lemma 3.3, we know that whp all of the following hold for all $p \in [p_2, p_3]$. Firstly, R''_p can be reconstructed in G_p. Secondly, every edge in G_p is incident to a vertex in R''_p. Thirdly, G_p contains no uncertain point. Fourthly, R''_p contains all vertices of degree at least two in G_p. Finally, G_{p_2} has isolated points while G_{p_3} has minimum degree at least two.

Now, consider the minimal p such that the graph G_p has no isolated and no undecidable points. By the final property, we know that $p_2 < p \leqslant p_3$. In particular, R''_p can be reconstructed in G_p by the first property. Let u be a point in the complement of R''_p. Since u cannot be an isolated point, u must have a unique neighbour v, and that neighbour must be in R''_p by the second property. So v has been reconstructed and we therefore know that u can only be at position u or $w = 2v - u$.

Given that u is not undecidable, (u, v) must be secure which means that w is a point of V. Since w cannot be isolated and u is not uncertain by the third property, it then follows that the degree of w must be at least two. Hence, by the fourth property, $w \in R''_p$, and so w has been reconstructed. So we know that the position w is already occupied by a point different from u, and so u can only be at a single possible position relative to the rest of the points. Hence, we can also reconstruct u. Because u was an arbitrary point of the complement of R''_p, it follows that all points of G_p can be reconstructed up to isometry.

Finally note that the graph preceding G_p in the random graph process has an isolated or undecidable point $v \in V$. Such a point has at least two possible positions relative to the rest of V and is therefore not determined up to isometry. Hence, G_p is the first graph in the random graph process that can be reconstructed up to isometry. □

Theorem 1.4 is now an easy consequence as we can simply follow the proofs from this section and the previous section to obtain the algorithm that we want.

Proof of Theorem 1.4. Our algorithm is composed of the following steps.

a. Reveal pairwise distances one-by-one until the graph of known pairwise distances has $42n$ edges. Call this graph G'.
b. Enumerate all walks (that do not repeat edges) of length at most $0.9 \log n$ in G' by starting at an arbitrary vertex and picking neighbours sequentially. From this, find a list of all cycles of length at most $0.9 \log n$ in G'.
c. Then check each cycle $v_1 \dots v_k$ of length at most $0.9 \log n$ in G'. If the distances between consecutive points on the cycle are $d_1, \dots, d_k$ and there are unique $\varepsilon_1, \dots, \varepsilon_{k-1} \in \{-1, 1\}$ such that $\sum_{i=1}^{k-1} \varepsilon_i d_i = d_k$, then the distance between v_i and v_j must be $|\sum_{l=i}^{j-1} \varepsilon_l d_l|$.
d. Using these new distances, apply Theorem 2.1 to reconstruct a set of points R (a polynomial time algorithm can be obtained from the proofs in [BT22]).
e. Reconstruct and add to R every vertex with two neighbours in R as well as every vertex with one neighbour in R whose other possible position is already occupied by a vertex of R.
f. Continue to reveal pairwise distances one-by-one. Each time a distance between points u, v is added where $v \in R$ and $u \notin R$, check if u has two neighbours in R or if the reflection of u over v is in R. If either of these two outcomes occur, then reconstruct u and add u to R.

Let p be such that G' corresponds to G_p in the coupling. By a Chernoff bound, with an exponential (in n) failure probability, $42/n < p < 168/n$. When $p < 168/n$, the expected number of walks in G_p (that do not repeat edges) of length at most $0.9 \log n$ is at most

$$\sum_{k \leqslant 0.9 \log n} p^k n^{k+1} < n \sum_{k \leqslant 0.9 \log n} 168^k < n^6,$$

and the number of such walks is always at most $n^{0.9 \log n}$. Hence, the expected number of such walks in G' is polynomial in n which implies that the expected runtime of step b is polynomial in n. It also follows that if X is the number of cycles of length at most $0.9 \log n$ in G', then $\mathbb{E}(X)$ is polynomial in n. For each cycle of length at most $0.9 \log n$,

there are at most $2^{0.9 \log n}$ possible $(\varepsilon_1, \dots, \varepsilon_{k-1})$ and so each check in step c takes time polynomial in n, implying that the runtime of part c is X times a polynomial in n. Hence, the expected runtime of the first three steps is polynomial in n. The final three steps can be performed in polynomial time.

The set R obtained by the algorithm is certainly reconstructible. We are left to check that whp $R = V$ exactly when V is reconstructible. As shown in the proofs of Theorem 1.2 and Theorem 1.3, whp the initial set R will have size $\Omega(n)$, and we will then add every point to R that has degree two or that has degree one and is not undecidable. Thus, by Theorem 1.3, whp this algorithm will reconstruct V exactly at the time when this is possible. □

We have already shown that if every point is incident to at least two known distances, then whp the whole of V is reconstructible. While in some cases the whole of V might be reconstructible even before that time, we now show that this does not happen much earlier.

Theorem 3.4. *Let $f : \mathbb{N} \to \mathbb{R}$ be any function going to infinity. Let V be a set of n points on the real line. Suppose the distances between pairs of points in V are revealed one-by-one in a uniformly random order. If there are at least $f(n)$ points that are incident to only one revealed distance, then whp it is not possible to reconstruct the whole of V.*

Proof. We may and will assume that $f(n) \leqslant n/100$ for all n. Let $G = (V, \mathcal{P})$ be the graph of known distances and suppose G has at least $f(n)$ points of degree one. It suffices to show that whp there is a point u whose unique edge uv is such that (u, v) is not secure.

We first give an upper bound on the number of secure pairs of points. Let the points in V be $v_1 < v_2 < \cdots < v_n$. Fix $k \leqslant n/2$. There are at most $k-1$ points $u \in \{v_{k+1}, \dots, v_n\}$ for which (u, v_k) is secure as the reflection of u over v_k must be one of $v_1, \dots, v_{k-1}$. In particular, there are at most $2(k-1)$ points $u \in V$ for which (u, v_k) is secure, and so the number of secure pairs whose second vertex is in the first half of V is

$$\sum_{k \leqslant n/2} 2(k-1) = 2\binom{\lfloor n/2 \rfloor}{2} \leqslant n^2/4.$$

Hence, the total number of secure pairs is at most $n^2/2$.

Now, perform the following random process. First pick a uniformly random degree-one vertex u_1 of G and let v_1 be the unique neighbour of u_1. For $s = 2, \dots, \lfloor f(n)/3 \rfloor$ let u_s be a uniformly random degree-one vertex in $V \setminus \{u_1, v_1, u_2, v_2, \dots, u_{s-1}, v_{s-1}\}$ and let v_s be the unique neighbour of u_s. Given $\{u_1, v_1, u_2, v_2, \dots, u_{s-1}, v_{s-1}\}$, note that (u_s, v_s) is a uniformly random ordered pair from

$$\{(x, y) : x \neq y,\ x \in V \setminus \{u_1, v_1, \dots, u_{s-1}, v_{s-1}\},\ y \in V \setminus \{u_1, u_2 \dots, u_{s-1}\}\}.$$

This set has size at least $3n^2/4$ and so, given $\{u_1, v_1, u_2, v_2, \dots, u_{s-1}, v_{s-1}\}$, the probability that (u_s, v_s) is secure is at most $(n^2/2)/(3n^2/4) = 2/3$. In particular, the probability that (u_s, v_s) is secure for all $s \leqslant \lfloor f(n)/3 \rfloor$ is at most $(2/3)^{\lfloor f(n)/3 \rfloor} = o(1)$. □

Now, if $p \leqslant (\log n + \log\log n - \omega(1))/n$, then whp $\mathcal{G}(n, p)$ has arbitrarily many degree-one vertices (see, for example, [Bol01, Ch. 7]) and so Theorem 1.2c follows.

4 Open problems

Theorem 1.2 shows that $1/n$ is a weak threshold for a linear sized reconstructible set of vertices. It would be interesting to determine whether there is a sharp threshold and it is natural to conjecture that this occurs at $1/n$ as this coincides with the appearance of the giant component in $\mathcal{G}(n,p)$.

Conjecture 4.1. *Let V be a set of n points on the real line. Suppose the graph of known pairwise distances $(V,\mathcal{P})$ is distributed as $\mathcal{G}(n,(1+\varepsilon)/n)$ where $\varepsilon > 0$ is a constant. Then whp there is a reconstructible set of size $\Omega_\varepsilon(n)$.*

We have not tried to optimize the constant in Theorem 1.2a, but our approach will not give a bound better than $9/n$, as for smaller p there are too few pairs in sufficiently short cycles. A bootstrapping argument might push the constant below 9, but would not get down to $p = (1+\varepsilon)/n$, as most pairs will only be in cycles of length at least $\Omega(\log(n)/\varepsilon)$.

It would also be interesting to give a characterisation of the giant reconstructible component. For global graph rigidity in one-dimension the largest reconstructible component is the largest 2-connected subgraph. For graph rigidity in two-dimensions the threshold for the emergence of a giant reconstructible component was determined in [KMT11] and a characterisation of this component was determined in [BLM18].

Reconstructibility in dimensions greater than one is particularly interesting. Reconstructing the whole of V is too much to ask: consider an embedding where $n-2$ points lie in a $(d-1)$-dimensional hyperplane and the other two points u,v do not. V can only be fully reconstructed if the distance between u and v is revealed (otherwise u,v could be on the same side of the hyperplane or on opposite sides). However, it is interesting to ask for the threshold at which a linear sized subset of V can be reconstructed. Is $1/n$ a weak threshold as it is for $d=1$?

Finally, Benjamini and Tzalik [BT22] also considered an even stronger notion of reconstructibility. Let $G=(V,E)$ be a graph. We say a subset $U \subseteq V$ is *adversarially reconstructible in* $\mathbb{R}$ if, for *every* embedding of V in $\mathbb{R}$, U is reconstructible from the distances $|u-v|$ for $uv \in E$. Note this is similar to the definition of global rigidity with the generic condition removed. It would be interesting to determine the thresholds for $\mathcal{G}(n,p)$ to be adversarially reconstructible in $\mathbb{R}$ and for some linear sized subset of $\mathcal{G}(n,p)$ to be adversarially reconstructible. In contrast to Theorem 1.2, minimum degree at least two is necessary as the embedding can be chosen so there are no secure pairs. Benjamini and Tzalik conjectured it is also sufficient: when distances are revealed one-by-one in a random order, the graph becomes adversarially reconstructible exactly at the first time that it has minimum degree two.

References

[ACKR89] Noga Alon, Yair Caro, Ilia Krasikov, and Yehuda Roditty (1989). Combinatorial reconstruction problems. *Journal of Combinatorial Theory, Series B* **47**(2), 153–161. ↑1

[BDG+21] Daniel Irving Bernstein, Sean Dewar, Steven J. Gortler, Anthony Nixon, Meera Sitharam, and Louis Theran (2021). Maximum likelihood thresholds via graph rigidity. arXiv:2108.02185. ↑3

[BLM18] J. Barré, M. Lelarge, and D. Mitsche (2018). On rigidity, orientability, and cores of random graphs with sliders. *Random Structures & Algorithms* **52**(3), 419–453. ↑3, 12

[Bol01] Béla Bollobás (2001). Random graphs. Cambridge Studies in Advanced Mathematics, 2nd edn. (Cambridge University Press). ↑6, 7, 11

[BT22] Itai Benjamini and Elad Tzalik (2022). Determining a points configuration on the line from a subset of the pairwise distances. arXiv:2208.13855. ↑1, 3, 4, 10, 12

[Con05] Robert Connelly (2005). Generic global rigidity. *Discrete & Computational Geometry. An International Journal of Mathematics and Computer Science* **33**(4), 549–563. ↑3

[GHT10] Steven J. Gortler, Alexander D. Healy, and Dylan P. Thurston (2010). Characterizing generic global rigidity. *American Journal of Mathematics* **132**(4), 897–939. ↑3

[JSS07] Bill Jackson, Brigitte Servatius, and Herman Servatius (2007). The 2-dimensional rigidity of certain families of graphs. *Journal of Graph Theory* **54**(2), 154–166. ↑3

[JT22] Tibor Jordán and Shin ichi Tanigawa (2022). Rigidity of random subgraphs and eigenvalues of stiffness matrices. *SIAM Journal on Discrete Mathematics* **36**(3), 2367–2392. ↑3

[JW17] Tibor Jordán and Walter Whiteley (2017). Global rigidity. *Handbook of Discrete and Computational Geometry* (edited by Jacob E. Goodman, Joseph O'Rourke, and Csaba D. Tóth), 3rd edn., chap. 63. ↑3

[KMT11] Shiva Prasad Kasiviswanathan, Cristopher Moore, and Louis Theran (2011). The rigidity transition in random graphs. *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, 1237–1252. arXiv:1010.3605. ↑3, 12

[KT13] Franz J. Király and Louis Theran (2013). Coherence and sufficient sampling densities for reconstruction in compressed sensing. arXiv:1302.2767. ↑3

[LNPR22] Alan Lew, Eran Nevo, Yuval Peled, and Orit E. Raz (2022). Sharp threshold for rigidity of random graphs. arXiv:2202.09917. ↑3

[PRS03] Luke Pebody, Jamie Radcliffe, and Alex Scott (2003). Finite subsets of the plane are 18-reconstructible. *SIAM Journal on Discrete Mathematics* **16**(2), 262–275. ↑1

[RW10] Oliver Riordan and Nicholas Wormald (2010). The diameter of sparse random graphs. *Combinatorics, Probability and Computing* **19**(5-6), 835–926. ↑6