A Simple Algorithm for Graph Reconstruction

Claire Mathieu, Hang Zhou

A Simple Algorithm for Graph Reconstruction*

Claire Mathieu† Hang Zhou‡

Abstract

How efficiently can we find an unknown graph using distance queries between its vertices? We assume that the unknown graph is connected, unweighted, and has bounded degree. The goal is to find every edge in the graph. This problem admits a reconstruction algorithm based on multi-phase Voronoi-cell decomposition and using $\tilde{O}(n^{3/2})$ distance queries [27].

In our work, we analyze a simple reconstruction algorithm. We show that, on random Δ-regular graphs, our algorithm uses $\tilde{O}(n)$ distance queries. As by-products, with high probability, we can reconstruct those graphs using $\log^2 n$ queries to an all-distances oracle or $\tilde{O}(n)$ queries to a betweenness oracle, and we bound the metric dimension of those graphs by $\log^2 n$.

Our reconstruction algorithm has a very simple structure, and is highly parallelizable. On general graphs of bounded degree, our reconstruction algorithm has subquadratic query complexity.

Keywords. reconstruction, network topology, random regular graphs, metric dimension

*This is the full version of the extended abstract that appears in the proceedings of the European Symposium on Algorithms (ESA) 2021. This work was partially funded by the grant ANR-19-CE48-0016 from the French National Research Agency (ANR).

†CNRS Paris, France; e-mail: claire.mathieu@irif.fr.

‡École Polytechnique, France; e-mail: hzhou@lix.polytechnique.fr.

1 Introduction

Discovering the topology of the Internet is a crucial step for building accurate network models and designing efficient algorithms for Internet applications. The topology of Internet networks is typically investigated at the router level, using `traceroute`. It is a common and reasonably accurate assumption that `traceroute` generates paths that are shortest in the network. Unfortunately, sometimes routers block `traceroute` requests due to privacy and security concerns. As a consequence, the inference of the network topology is rather based on the end-to-end delay information on those requests, which is roughly proportional to the shortest-path distances in the network.

In the *graph reconstruction* problem, we are given the vertex set V of a hidden connected, undirected, and unweighted graph and have access to information about the topology of the graph via an oracle, and the goal is to find every edge in E. Henceforth, unless explicitly mentioned, all graphs studied are assumed to be connected. This assumption is standard and shared by almost all references on the subject, e.g., [7, 14, 27, 39, 41]. The efficiency of an algorithm is measured by the *query complexity*, i.e., the number of queries to the oracle. Motivated by `traceroute`, the literature has explored several types of query oracles.

- One type consists of *all-shortest-paths* and *all-distances* queries, when querying a vertex yields either shortest paths from that vertex to all other vertices [7, 41] or distances from that vertex to all other vertices [14]. The latter, of course, is less informative.
- A more refined type of query oracles, suggested in [7, 14], consists of *shortest-path* and *distance* queries, when querying a pair of vertices yields either a shortest path or the distance between them [27, 38, 39]. Again, the latter is less informative.

In this work, we focus on the weakest of those four query oracles, that takes as input a pair of vertices a and b and returns the distance $\delta(a, b)$ between them. Reyzin and Srivastava [38] showed that graph reconstruction requires $\Omega(n^2)$ distance queries on general graphs, so we focus on the bounded degree case. For graphs of bounded degree, Kannan, Mathieu, and Zhou [27] gave a reconstruction algorithm based on multi-phase Voronoi-cell decomposition and using $\tilde{O}(n^{3/2})$ distance queries, and raised an open question of whether $\tilde{O}(n)$ is achievable.[1]

We provide a partial answer to that open question by analyzing a simple reconstruction algorithm (Algorithm 1). We show that, on (uniformly) random Δ-regular graphs, where every vertex has the same degree Δ, our reconstruction algorithm uses $\tilde{O}(n)$ distance queries (Theorem 1). As by-products, with high probability, we can reconstruct those graphs using $\log^2 n$ queries to an all-distances oracle (Corollary 2) or using $\tilde{O}(n)$ queries to a betweenness oracle (Corollary 3), and we bound the metric dimension of those graphs by at most $\log^2 n$ (Corollary 5).

Our analysis exploits the *locally tree-like* property of random Δ-regular graphs, meaning that these graphs contain a small number of short cycles. Our method might be applicable to other locally tree-like graphs, such as Erdös-Rényi random graphs and *scale-free* graphs. In particular, many real world networks, such as Internet networks, social networks, and peer-to-peer networks, are believed to have scale-free properties [6, 25, 34]. We defer the reconstruction of those networks for future work.

Our reconstruction algorithm has a very simple structure, and is highly parallelizable (Lemma 8). On general graphs of bounded degree, the same reconstruction algorithm has subquadratic query complexity (Theorem 6).

1.1 Related Work

The problem of reconstructing a graph using queries that reveal partial information has been extensively studied in different contexts and has many applications.

Reconstruction of Random Graphs The gist of our paper deals with random graphs. The graph reconstruction problem has already attracted much interest in the setting of random graphs. On Erdös-Rényi random graphs, Erlebach, Hall, and Mihal'ák [15] studied the approximate network reconstruction using all-shortest-paths queries; Anandkumar, Hassidim, and Kelner [4] used end-to-end measurements between a subset of vertices to approximate the network structure. Experimental results to reconstruct random graphs using shortest-path queries were given in [8, 20].

On random Δ-regular graphs, Achlioptas et al. [2] studied the bias of `traceroute` sampling in the context of the network reconstruction. They showed that the structure revealed by `traceroute` sampling on random Δ-regular graphs admits a power-law degree distribution [2], a common phenomenon as in Erdös-Rényi random graphs [31] and Internet networks [16].

[1] The notation $\tilde{O}(f(n))$ stands for $O(f(n) \cdot \operatorname{polylog} f(n))$.

Metric Dimension and Related Problems Our work yields an upper bound on the *metric dimension* of random Δ-regular graphs. The metric dimension problem was first introduced by Slater [42] and Harary and Melter [21], see also [5, 12, 13, 23, 29, 36, 37, 40]. The metric dimension of a graph is the cardinality of a smallest subset S of vertices such that every vertex in the graph has a unique vector of distances to the vertices in S. On regular graphs, the metric dimension problem was studied in special cases [13, 24]. In Erdös-Rényi random graphs, the metric dimension problem was studied by Bollobás, Mitsche, and Prałat [11]. Mitsche and Rué [32] also considered the random forest model.

A related problem is the *identifying code* of a graph [28], which is a smallest subset of vertices such that every vertex of the graph is uniquely determined by its neighbourhood within this subset. The identifying code problem was studied on random Δ-regular graphs [17] and on Erdös-Rényi random graphs [19]. Other related problems received attentions on random graphs as well, such as the *sequential metric dimension* [35] and the *seeded graph matching* [33].

Betweenness Oracle There exists an oracle that is even weaker than the distance oracle: the *betweenness* oracle [1], which receives three vertices u, v, and w and returns whether w lies on a shortest path between u and v. Our work yields a reconstruction algorithm using $\tilde{O}(n)$ betweenness queries on random Δ-regular graphs. For graphs of bounded degree, Abrahamsen et al. [1] generalized the $\tilde{O}(n^{3/2})$ result in the distance oracle model from [27] to the betweenness oracle model.

Tree Reconstruction and Parallel Setting Our paper focuses on the distance oracle and bounded degree, and considers the parallel setting. All of those aspects were previously raised in the special case of the *tree reconstruction*. Indeed, motivated by the reconstruction of a phylogenetic tree in evolutionary biology, the tree reconstruction problem using a distance oracle is well-studied [22, 30, 43], in particular assuming bounded degree [22]. Afshar et al. [3] studied the tree reconstruction in the parallel setting, analyzing both the *round complexity* and the *query complexity* in the relative distance query model [26].

1.2 Our Results

Our reconstruction algorithm, called Simple, is given in Algorithm 1. It takes as input the vertex set V of size n and an integer parameter $s \in [1, n]$.

Algorithm 1 Simple (V, s)

1: $S \leftarrow$ sample of s vertices selected uniformly and independently at random from V
2: **for** $u \in S$ and $v \in V$ **do** Query(u, v)
3: $\hat{E} \leftarrow$ set of vertex pairs $\{a, b\} \subseteq V$ such that, for all $u \in S$, $|\delta(u, a) - \delta(u, b)| \leq 1$
4: **for** $\{a, b\} \in \hat{E}$ **do** Query(a, b)
5: **return** set of vertex pairs $\{a, b\} \in \hat{E}$ such that $\delta(a, b) = 1$

Intuitively, the set $\hat{E}$ constructed in Simple consists of all vertex pairs $\{a, b\} \subseteq V$ that *might* be an edge in E. In order to obtain the edge set E, it suffices to query uniquely the vertex pairs in $\hat{E}$. We remark that Simple correctly reconstructs the graph for any parameter $s \in [1, n]$, and that choosing an appropriate s only affects the query complexity, see Lemma 8.

Simple correctly reconstructs any connected graph. We analyze the expected query complexity of Simple on random Δ-regular graphs in Theorem 1 and on bounded degree graphs in Theorem 6.

1.2.1 Random Regular Graphs

Our first main result shows that Simple (Algorithm 1) uses $\tilde{O}(n)$ distance queries on random Δ-regular graphs for an appropriately chosen s and uses 2 parallel rounds (Theorem 1). The analysis exploits the *locally tree-like* property of random Δ-regular graphs. The proof of Theorem 1 consists of several technical novelties, based on a new concept of *interesting vertices* (Definition 2). See Section 3.

Theorem 1. *Consider a uniformly random Δ-regular graph with $\Delta = O(1)$. Let $s = \log^2 n$. In the distance query model,* Simple *(Algorithm 1) is a reconstruction algorithm using $\tilde{O}(n)$ queries in expectation. In addition,* Simple *can be parallelized using 2 rounds.*

We extend Simple and its analysis to reconstruct random Δ-regular graphs in the all-distances query model and in the betweenness query model with high probability (Corollaries 2 and 3). These extensions are based on the observation that the set $\hat{E}$ constructed in Simple equals the edge set E with high probability (Lemma 17),[2] see Section 4.

[2]This property (i.e., $\hat{E} = E$ with high probability) does not hold on general graphs of bounded degree.

Corollary 2. *Consider a uniformly random Δ-regular graph with $\Delta = O(1)$. In the all-distances query model, there is an algorithm that uses $\log^2 n$ queries and reconstructs the graph with probability $1 - o(1)$.*

Corollary 3. *Consider a uniformly random Δ-regular graph with $\Delta = O(1)$. In the betweenness query model, there is an algorithm that uses $\tilde{O}(n)$ queries and reconstructs the graph with probability $1 - o(1)$.*

Remark 4. *The algorithms in Corollaries 2 and 3 require the promise that the graph is uniformly random Δ-regular, and may output incorrect results on other graphs.*

We further extend the analysis of SIMPLE to study the metric dimension of random Δ-regular graphs (Corollary 5), by showing (in Lemma 20) that a random subset of $\log^2 n$ vertices is almost surely a *resolving set* (Definition 3) for those graphs, see Section 5.

Corollary 5. *Consider a uniformly random Δ-regular graph with $\Delta = O(1)$. With probability $1 - o(1)$, the metric dimension of the graph is at most $\log^2 n$.*

With extra work, the parameter $s = \log^2 n$ in Theorem 1 can be reduced to $\log n \cdot (\log\log n)^{2+\epsilon}$, for any $\epsilon > 0$, see Remark 16. As a consequence, the query complexity in the all-distances query model (Corollary 2) and the upper bound on the metric dimension (Corollary 5) can both be improved to $O(\log n \cdot (\log\log n)^{2+\epsilon})$.

1.2.2 Bounded-Degree Graphs

On general graphs of bounded degree, SIMPLE (Algorithm 1) has subquadratic query complexity (Theorem 6), see Section 6.

Theorem 6. *Consider a general graph of bounded degree $\Delta = O(\text{polylog}\, n)$. Let $s = n^{2/3}$. In the distance query model,* SIMPLE *(Algorithm 1) is a reconstruction algorithm using $\tilde{O}(n^{5/3})$ queries in expectation. In addition,* SIMPLE *can be parallelized using 2 rounds.*

We note that the MULTI-PHASE algorithm[3] from [27] also reconstructs graphs of bounded degree in the distance query model. How does SIMPLE compare to MULTI-PHASE? In terms of query complexity, on general graphs of bounded degree, SIMPLE uses $\tilde{O}(n^{5/3})$ queries, so is not as good as MULTI-PHASE using $\tilde{O}(n^{3/2})$ queries; on random Δ-regular graphs, SIMPLE is more efficient than MULTI-PHASE: $\tilde{O}(n)$ versus $\tilde{O}(n^{3/2})$. In terms of round complexity, SIMPLE can be parallelized using 2 rounds on general graphs of bounded degree, and even $1 + o(1)$ rounds on random Δ-regular graphs; while MULTI-PHASE requires up to $3 \log n$ rounds due to a multi-phase selection process for centers.[4] In terms of structure, SIMPLE is much simpler than MULTI-PHASE, which is based on multi-phase Voronoi-cell decomposition.

In worst case instances of graphs of bounded degree, the query complexity of SIMPLE is higher than linear. For example, when the graph is a complete binary tree, SIMPLE would require $\Omega(n\sqrt{n})$ queries (the complexity of SIMPLE is minimized when s is roughly $\sqrt{n}$). Thus the open question from [27] of whether general graphs of bounded degree can be reconstructed using $\tilde{O}(n)$ distance queries remains open and answering it positively would require further algorithmic ideas.

2 Notations and Preliminary Analysis

Let $G = (V, E)$ be a connected, undirected, and unweighted graph, where V is the set of vertices such that $|V| = n$ and E is the set of edges. We say that $\{a, b\} \subseteq V$ is a *vertex pair* if both a and b belong to V such that $a \neq b$. The *distance* between a vertex pair $\{a, b\} \subseteq V$, denoted by $\delta(a, b)$, is the number of edges on a shortest a-to-b path. Throughout the paper, we use $\log(\cdot)$ to indicate $\log_2(\cdot)$.

Definition 1 (Distinguishing)**.** For a vertex pair $\{a, b\} \subseteq V$, we say that a vertex $u \in V$ *distinguishes* a and b, or equivalently that u is a *distinguisher* of $\{a, b\}$, if $|\delta(u, a) - \delta(u, b)| > 1$. Let $D(a, b) \subseteq V$ denote the set of vertices $u \in V$ distinguishing a and b.

Let $s \in [1, n]$ be an integer parameter. The set S constructed in SIMPLE consists of s vertices selected uniformly and independently at random from V.

The set $\hat{E}$ constructed in SIMPLE consists of the vertex pairs $\{a, b\} \subseteq V$ such that a and b are not distinguished by any vertex in S, i.e., $D(a, b) \cap S = \emptyset$, or equivalently, $|\delta(u, a) - \delta(u, b)| \leq 1$ for all $u \in S$. For any edge $(a, b) \in E$, it is easy to see that $|\delta(u, a) - \delta(u, b)| \leq 1$ for all $u \in V$, which implies that $\{a, b\} \in \hat{E}$. Hence the following inclusion property.

Fact 7. $E \subseteq \hat{E}$.

[3] Algorithm 3 in [27].

[4] The number of rounds in MULTI-PHASE is implicit in the proof of Lemma 2.3 from [27].

We show that SIMPLE is correct and we give a preliminary analysis on its query complexity as well as on its round complexity, in Lemma 8.

Lemma 8. *The output of* SIMPLE *(Algorithm 1) equals the edge set E. The number of distance queries in* SIMPLE *is $n \cdot s + |\hat{E}|$. In addition,* SIMPLE *can be parallelized using 2 rounds.*

Proof. The output of SIMPLE consists of the vertex pairs $\{a, b\} \in \hat{E}$ such that $\{a, b\}$ is an edge in E. Since $E \subseteq \hat{E}$ (Fact 7), the output of SIMPLE equals the edge set E.

Observe that the distance queries in SIMPLE are performed in two stages. The number of distance queries in the first stage is $|V| \cdot |S| = n \cdot s$. The number of distance queries in the second stage is $|\hat{E}|$. Thus the query complexity of SIMPLE is $n \cdot s + |\hat{E}|$. The distance queries in each of the two stages can be performed in parallel, so SIMPLE can be parallelized using 2 rounds. □

From Lemma 8, in order to further study the query complexity of SIMPLE, it suffices to analyze $|\hat{E}|$, which equals $|E| + |\hat{E} \setminus E|$ according to Fact 7. Since $|E| \leq \Delta n$ in a graph of bounded degree Δ, our focus in the subsequent analysis is $|\hat{E} \setminus E|$.

Lemma 9. *Let $s = \omega(\log n)$ be an integer parameter. Let B be the set of vertex pairs $\{a, b\} \subseteq V$ such that $\delta(a, b) \geq 2$ and $|D(a, b)| \leq 3n \cdot (\log n)/s$. We have $\mathbb{E}_S\big[|\hat{E} \setminus E|\big] \leq |B| + o(1)$.*

Proof. Let Z denote the set $\hat{E} \setminus E$. Observe that $|Z| \leq |B| + |Z \setminus B|$. Since B is independent of the random set S, we have $\mathbb{E}_S\big[|Z|\big] \leq |B| + \mathbb{E}_S\big[|Z \setminus B|\big]$. It suffices to show that $\mathbb{E}_S\big[|Z \setminus B|\big] = o(1)$.

We claim that for any vertex pair $\{a, b\} \subseteq V$ such that $\{a, b\} \notin B$, the probability that $\{a, b\} \in Z$ is $o(n^{-2})$. To see this, fix a vertex pair $\{a, b\} \notin B$. By definition of B, either $\delta(a, b) = 1$, or $|D(a, b)| > 3n \cdot (\log n)/s$. In the first case, $\{a, b\} \notin Z$ since Z does not contain any edge of E. In the second case, the event $\{a, b\} \in Z$ would imply that $\{a, b\} \in \hat{E}$, hence $D(a, b) \cap S = \emptyset$. Therefore,

$$\begin{aligned}
&\mathbb{P}_S\big[\{a, b\} \in Z \mid \{a, b\} \notin B\big] \\
\leq &\mathbb{P}_S\big[D(a, b) \cap S = \emptyset \mid \{a, b\} \notin B\big] \\
< &\left(1 - \frac{3n \cdot (\log n)/s}{n}\right)^s \\
= &o(n^{-2}),
\end{aligned}$$

where the second inequality follows since $|D(a, b)| > 3n \cdot (\log n)/s$ and the set S consists of s vertices selected uniformly and independently at random, and the last step follows since $s = \omega(\log n)$.

There are at most $n(n-1)/2$ vertex pairs $\{a, b\} \notin B$. By the linearity of expectation, the expected number of vertex pairs $\{a, b\} \notin B$ such that $\{a, b\} \in Z$ is at most $o(n^{-2}) \cdot n(n-1)/2 = o(1)$, so $\mathbb{E}_S\big[|Z \setminus B|\big] = o(1)$. Therefore, $\mathbb{E}_S\big[|Z|\big] \leq |B| + \mathbb{E}_S\big[|Z \setminus B|\big] = |B| + o(1)$. □

3 Reconstruction of Random Regular Graphs (Proof of Theorem 1)

In this section, we analyze SIMPLE (Algorithm 1) on random Δ-regular graphs in the distance query model. We assume that $\Delta \geq 2$ and that Δn is even since otherwise those graphs do not exist.

We bound the expectation of $|\hat{E} \setminus E|$ on random Δ-regular graphs, in Lemma 10.

Lemma 10. *Let G be a uniformly random Δ-regular graph with $\Delta = O(1)$. Let $s = \log^2 n$. Let $S \subseteq V$ be a set of s vertices selected uniformly and independently at random from V. We have $\mathbb{E}_{G,S}\big[|\hat{E} \setminus E|\big] = o(1)$.*

Proof of Theorem 1 using Lemma 10. By Lemma 8, SIMPLE is a reconstruction algorithm using $n \cdot s + |\hat{E}| = n \cdot \log^2 n + |\hat{E}|$ distance queries, and in addition, SIMPLE can be parallelized using 2 rounds. From Fact 7, $|\hat{E}| = |E| + |\hat{E} \setminus E|$. Since G is Δ-regular, $|E| = \Delta n/2$. By Lemma 10, $\mathbb{E}_{G,S}\big[|\hat{E} \setminus E|\big] = o(1)$. Therefore, the expected number of distance queries in SIMPLE is $n \cdot \log^2 n + \Delta n/2 + o(1)$, which is $\tilde{O}(n)$ since $\Delta = O(1)$. □

It remains to prove Lemma 10 in the rest of this section.

3.1 Configuration Model and the Structural Lemma

We consider a random Δ-regular graph generated according to the *configuration model* [9, 44]. Given a partition of a set of Δn points into n cells $v_1, v_2, \ldots, v_n$ of Δ points, a *configuration* is a perfect matching of the points into $\Delta n/2$ pairs. It corresponds to a (not necessarily connected) *multigraph* G' in which the cells are regarded as vertices and the pairs as edges: a pair of points $\{x, y\}$ in the configuration corresponds to an edge (v_i, v_j) of G' where $x \in v_i$ and $y \in v_j$. Since each Δ-regular graph has exactly $(\Delta!)^n$ corresponding configurations, a

Δ-regular graph can be generated uniformly at random by rejection sampling: choose a configuration uniformly at random,[5] and reject the result if the corresponding multigraph G' is not simple or not connected. The configuration model enables us to show properties of a random Δ-regular graph by analyzing a multigraph G' corresponding to a random configuration.

Based on the configuration model, we are ready to state the following *Structural Lemma*, which is central in our analysis.

Lemma 11 (Structural Lemma)**.** *Let $\Delta = O(1)$ be such that $\Delta \geq 3$. Let G' be a multigraph corresponding to a uniformly random configuration. Let $\{v, w\}$ be a vertex pair in G' such that $\delta(v, w) \geq 2$. With probability $1 - o(n^{-2})$, we have $|D(v, w)| > 3n/\log n$.*

In Section 3.2, we prove the Structural Lemma, and in Section 3.3, we show Lemma 10 using the Structural Lemma.

3.2 Proof of the Structural Lemma (Lemma 11)

Let G' be a multigraph corresponding to a uniformly random configuration, and let V be the vertex set of G'. Let $\{v, w\} \subseteq V$ be a vertex pair such that $\delta(v, w) \geq 2$. For a vertex $x \in V$, let $\ell(x) \in \mathbb{Z}$ denote the distance in G' between x and the vertex pair $\{v, w\}$, i.e., $\ell(x) = \min(\delta(x, v), \delta(x, w))$. For any integer $k \geq 0$, let $U_k \subseteq V$ denote the set of vertices $x \in V$ such that $\ell(x) = k$. Let $U_{\leq k}$ denote $\bigcup_{j \leq k} U_j$.

To construct the multigraph G' from a random configuration, we borrow the approach from [10], which proceeds in n phases to construct the edges in G', exploring vertices $x \in V$ in non-decreasing order of $\ell(x)$. We start at the vertices of $U_0 = \{v, w\}$. Initially, i.e., in the 0-th phase, we construct all the edges incident to v or incident to w. Suppose at the beginning of the k-th phase, for each $k \in [1, n-1]$, we have constructed all the edges with at least one endpoint belonging to $U_{\leq k-1}$. During the k-th phase, we construct the edges incident to the vertices in U_k one by one, till the degree of all the vertices in U_k reaches Δ. The ordering of the edge construction within the same phase is arbitrary. Let G' be the resulting multigraph in the end of the construction.[6] The ordering of the edges in G' is defined according to the above edge construction.

An edge (a, b) in G' is *indispensable* if it explores either the vertex a or the vertex b for the first time in the edge construction. In the first case, b is the *predecessor* of a; and in the second case, a is the *predecessor* of b. An edge is *dispensable* if it is not indispensable, in other words, if each of its endpoints either belongs to $\{v, w\}$ or is an endpoint of an edge constructed previously.

Fact 12. *Neither v or w has a predecessor. For any vertex in V, its predecessor, if exists, is unique. If vertex a is the predecessor of vertex b, then $\ell(b) = \ell(a) + 1$.*

We introduce the concept of *interesting vertices*, which is a key idea in the analysis.

Definition 2 (Interesting Vertices)**.** A vertex $x \in V$ is *v-interesting* if, for all vertices $z \in V \setminus \{v\}$ with $\delta(v, z) + \delta(z, x) = \delta(v, x)$, the edges incident to z are indispensable. Similarly, a vertex $x \in V$ is *w-interesting* if, for all vertices $z \in V \setminus \{w\}$ with $\delta(w, z) + \delta(z, x) = \delta(w, x)$, the edges incident to z are indispensable.

For any finite integer $k \geq 1$, let $I_k(v) \subseteq V$ denote the set of v-interesting vertices $x \in V$ such that $\delta(v, x) = k$, and let $I_k(w) \subseteq V$ denote the set of w-interesting vertices $x \in V$ such that $\delta(w, x) = k$.

We show in Lemma 13 that interesting vertices distinguish the vertex pair $\{v, w\}$, and we provide a lower bound on the number of interesting vertices in Lemma 14. These two lemmas are main technical novelties in our work. Their proofs are in Sections 3.2.1 and 3.2.2, respectively.

Lemma 13. *For any finite integer $k \geq 1$, we have $I_k(v) \cup I_k(w) \subseteq D(v, w)$.*

Lemma 14. *Let $\Delta = O(1)$ be such that $\Delta \geq 3$. Let k be any positive integer such that $k \leq \lceil \log_{\Delta-1}(3n/\log n) \rceil + 2$. With probability $1 - o(n^{-2})$, we have $|I_k(v) \cup I_k(w)| > (\Delta - 2 - o(1))(\Delta - 1)^{k-1}$.*

The Structural Lemma (Lemma 11) follows easily from Lemmas 13 and 14, see Section 3.2.3.

3.2.1 Proof of Lemma 13

Fix a finite integer $k \geq 1$. From the symmetry of v and w, it suffices to prove $I_k(v) \subseteq D(v, w)$.

Let x be any vertex in $I_k(v)$. By definition, x is v-interesting and $\delta(v, x) = k$. Let $a_0 = v, a_1, \ldots, a_k = x$ be any shortest v-to-x path. For any vertex a_i with $i \in [1, k]$, the edges incident to a_i are indispensable according to Definition 2.

[5] To generate a random configuration, the points in a pair can be chosen sequentially: the first point can be selected using any rule, as long as the second point in that pair is chosen uniformly from the remaining points.

[6] When a multigraph corresponding to a random configuration is not connected, the resulting G' consists of the union of the components of v and of w, respectively, in that multigraph. Note that any vertex $x \in V$ outside those two components cannot distinguish v and w (i.e., $x \notin D(v, w)$), thus x is irrelevant to $|D(v, w)|$ in the statement of Lemma 11.

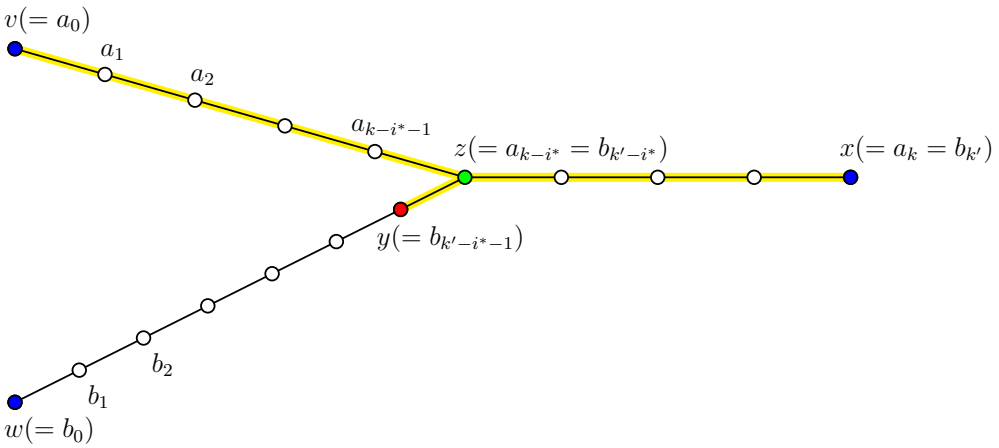

Figure 1: $a_0, a_1, \ldots, a_k$ is a shortest v-to-x path, and $b_0, b_1, \ldots, b_{k'}$ is a shortest w-to-x path. The vertex z represents the branching point of these two paths. Since the vertex x is v-interesting, the highlighted edges are indispensable.

We claim that, for any $i \in [1, k]$, a_{i-1} is the predecessor of a_i, and in addition, $\ell(a_i) = i$. The proof is by induction. First, consider the case when $i = 1$. The edge (a_0, a_1) is incident to the vertex a_1, so is indispensable. Thus either a_0 is the predecessor of a_1, or a_1 is the predecessor of a_0. Since a_0 $(= v)$ has no predecessor (Fact 12), a_1 cannot be the predecessor of a_0, so a_0 is the predecessor of a_1. Again using Fact 12, we have $\ell(a_1) = \ell(a_0) + 1$. Since $\ell(a_0) = \ell(v) = 0$, we have $\ell(a_1) = 1$. Next, consider the case when $i \geq 2$, and assume that the claim holds already for $1, \ldots, i-1$. The edge (a_{i-1}, a_i) is incident to the vertex a_i, so is indispensable. Thus either a_{i-1} is the predecessor of a_i, or a_i is the predecessor of a_{i-1}. By induction, a_{i-2} is the predecessor of a_{i-1}. Since the predecessor of a_{i-1} is unique (Fact 12), a_i cannot be the predecessor of a_{i-1}, so a_{i-1} is the predecessor of a_i. Again using Fact 12, we have $\ell(a_i) = \ell(a_{i-1}) + 1$. Since $\ell(a_{i-1}) = i - 1$ by induction, we have $\ell(a_i) = i$.

In order to show that $x \in D(v, w)$, we prove in the following that $\delta(w, x) \geq k + 2$. Indeed, since $\delta(v, x) = k$, the event $\delta(w, x) \geq k + 2$ implies that $x \in D(v, w)$ by Definition 1.[7]

Let $b_0 = w, b_1, \ldots, b_{k'} = x$ be any shortest w-to-x path, for some integer k'. See Fig. 1. Let $i^* \in [0, k]$ be the largest integer such that $a_{k-j} = b_{k'-j}$ for all $j \in [0, i^*]$. Let z denote the vertex a_{k-i^*}, which equals $b_{k'-i^*}$. If $i^* = k$, the v-to-x path $a_0, a_1, \ldots, a_k$ is a subpath of the w-to-x path $b_0, b_1, \ldots, b_{k'}$. Since $\delta(w, v) \geq 2$, we have $\delta(w, x) = \delta(w, v) + \delta(v, x) \geq 2 + k$, which implies that $x \in D(v, w)$. From now on, it suffices to consider the case when $i^* < k$.

Let y denote the vertex $b_{k'-i^*-1}$. Since y is on a shortest w-to-x path, we have

$$\delta(w, x) = \delta(w, y) + \delta(y, x) = \delta(w, y) + (i^* + 1) \geq \ell(y) + (i^* + 1), \tag{1}$$

where the inequality follows from the definition of $\ell(y)$. It remains to analyze the value of $\ell(y)$.

The edge (z, y) is incident to the vertex z $(= a_{k-i^*})$, so is indispensable. Thus either y is the predecessor of z, or z is the predecessor of y. From the previous claim, a_{k-i^*-1} is the predecessor of z. Since the predecessor of z is unique (Fact 12) and $y \neq a_{k-i^*-1}$ (by definition of i^*), y cannot be the predecessor of z, so z is the predecessor of y. Again by Fact 12, $\ell(y) = \ell(z) + 1$. Since $\ell(z) = \ell(a_{k-i^*}) = k - i^*$ by the previous claim, we have $\ell(y) = k - i^* + 1$. We conclude from Eq. (1) that

$$\delta(w, x) \geq (k - i^* + 1) + (i^* + 1) = k + 2,$$

which implies that $x \in D(v, w)$.

We proved that $I_k(v) \subseteq D(v, w)$. Similarly, $I_k(w) \subseteq D(v, w)$. Therefore, $I_k(v) \cup I_k(w) \subseteq D(v, w)$.

We complete the proof of Lemma 13.

3.2.2 Proof of Lemma 14

To begin with, we show that there are relatively few dispensable edges within a neighborhood of $\{v, w\}$. This property, also called the *locally tree-like* property, was previously exploited by Bollobás [10] for three levels of neighborhoods on random Δ-regular graphs in the context of automorphisms of those graphs. In Lemma 15, we extend the analysis from [10] to show the locally tree-like property for $M = \lceil \log \log n \rceil$ levels of neighborhoods.

[7] When $\delta(w, x)$ is infinite (i.e., w and x are not connected in G'), it is trivial that $x \in D(v, w)$, since $\delta(v, x)$ is finite. Therefore, it suffices to consider the case when $\delta(w, x)$ is finite in the rest of the proof.

Lemma 15. *Let $M = \lceil \log \log n \rceil$. We can construct two non-decreasing sequences $\{g_i\}_{1 \le i \le M}$ and $\{L_i\}_{1 \le i \le M}$, such that all of the following properties hold when n is large enough:*

1. *$g_1 = 3$; and for any $i \in [2, M]$, $g_i = o\left((\Delta-1)^{L_{i-1}}/M\right)$;*
2. *$L_M \ge \lceil \log_{\Delta-1}(3n/\log n) \rceil + 2$;*
3. *With probability $1 - o(n^{-2})$, for all $i \in [1, M]$, strictly less than g_i edges are dispensable among the edges incident to vertices in $U_{\le L_i}$.*

Proof of Lemma 15. First, we define two sequences $\{g_i\}_{1 \le i \le M}$ and $\{f_i\}_{1 \le i \le M}$ as follows: $g_1 = 3$, $f_1 = \lceil n^{1/8} \rceil$, and for each $i \in [2, M]$, let

$$g_i = \left\lceil n^{1-7/2^{i+1}}/(\log n)^{1/2} \right\rceil,$$

$$f_i = \left\lceil n^{1-7/2^{i+2}}/(\log n)^{1/3} \right\rceil.$$

Next, we define the sequence $\{L_i\}_{1 \le i \le M}$ as follows: for each $i \in [1, M]$, let

$$L_i = \lceil \log_{\Delta-1} f_i \rceil - 6.$$

It is easy to see that all of the three sequences $\{g_i\}_{1 \le i \le M}$, $\{f_i\}_{1 \le i \le M}$, and $\{L_i\}_{1 \le i \le M}$ are non-decreasing.

To show Property 1 of the statement, observe that for any $i \in [2, M]$,

$$g_i \cdot M = \left\lceil n^{1-7/2^{i+1}}/(\log n)^{1/2} \right\rceil \cdot \lceil \log \log n \rceil = o\left(n^{1-7/2^{i+1}}/(\log n)^{1/3}\right).$$

Thus $g_i \cdot M = o(f_{i-1})$ by definition of f_{i-1}. From the definition of L_{i-1} and the fact that $\Delta = O(1)$, we have $f_{i-1} = \Theta\left((\Delta-1)^{L_{i-1}}\right)$. Therefore, $g_i \cdot M = o\left((\Delta-1)^{L_{i-1}}\right)$, hence $g_i = o\left((\Delta-1)^{L_{i-1}}/M\right)$.

To show Property 2 of the statement, observe that

$$f_M \ge n^{1-7/2^{(\log\log n)+2}}/(\log n)^{1/3} = 2^{-7/4} \cdot n/(\log n)^{1/3} > (\Delta-1)^8 \cdot 3n/\log n,$$

where the last inequality follows since $\Delta = O(1)$ and n is large enough. Therefore, $L_M = \lceil \log_{\Delta-1} f_M \rceil - 6 \ge \lceil \log_{\Delta-1}(3n/\log n) \rceil + 2$.

It remains to show Property 3 of the statement. Consider any integer $i \in [1, M]$. Since the graph is Δ-regular, the number of vertices in $U_{\le L_i}$ is at most

$$2 + 2\Delta \cdot \sum_{j=0}^{L_i - 1} (\Delta-1)^j = 2 + \frac{2\Delta(\Delta-1)^{L_i} - 2\Delta}{\Delta-2} < \frac{2\Delta(\Delta-1)^{L_i}}{\Delta-2}.$$

Let n_i be the number of edges incident to vertices in $U_{\le L_i}$. Since each vertex is incident to Δ edges, we have $n_i < \frac{2\Delta^2(\Delta-1)^{L_i}}{\Delta-2}$. Since $L_i = \lceil \log_{\Delta-1} f_i \rceil - 6 < \left(\log_{\Delta-1} f_i\right) - 5$, we have $n_i < \frac{2\Delta^2}{(\Delta-2)(\Delta-1)^5} \cdot f_i$, which is less than f_i since $\Delta \ge 3$.

In order to bound the number of dispensable edges incident to vertices in $U_{\le L_i}$, it suffices to bound the number of dispensable edges among the first f_i edges in the ordering of edge construction.

For any integer $t \in [1, \Delta n/2]$, let $p(t)$ denote the probability that the t-th edge in the construction is dispensable. We use the argument of Bollobás [10] to bound $p(t)$ as follows. Before constructing the t-th edge, the previously constructed $t-1$ edges are incident to at most $t+1$ vertices. For each of these $t+1$ vertices, at most $\Delta - 1$ incident edges are not yet constructed. Thus $p(t) \le \frac{(\Delta-1)(t+1)}{\Delta n - 2(t-1)}$, which is less than $\frac{2t}{n}$ as soon as $t = o(n)$.

From the definition of f_i, we have $f_i \le n/(\log n)^{1/3} = o(n)$, thus $p(f_i) < \frac{2f_i}{n}$. The probability that there exist g_i dispensable edges among the first f_i edges is at most

$$\binom{f_i}{g_i} \cdot \left(\frac{2f_i}{n}\right)^{g_i} < \left(\frac{e \cdot f_i}{g_i}\right)^{g_i} \cdot \left(\frac{2f_i}{n}\right)^{g_i},$$

where the inequality follows from Stirling's formula. When $i = 1$, we have

$$\left(\frac{e \cdot f_1}{g_1}\right)^{g_1} \cdot \left(\frac{2f_1}{n}\right)^{g_1} = \left(\frac{2e \cdot \left(\lceil n^{1/8} \rceil\right)^2}{3n}\right)^3 = o(n^{-17/8}),$$

and when $i \ge 2$, we have

$$\left(\frac{e \cdot f_i}{g_i}\right)^{g_i} \cdot \left(\frac{2f_i}{n}\right)^{g_i} = O\left(\left(\frac{2e}{(\log n)^{1/6}}\right)^{g_i}\right) = o(n^{-17/8}),$$

by definition of g_i and f_i and by observing that $g_i \geq n^{1/8}/(\log n)^{1/2}$ for any $i \geq 2$.

Thus for any $i \in [1, M]$, with probability $1 - o(n^{-17/8})$, strictly less than g_i edges are dispensable among the first f_i edges, hence strictly less than g_i edges are dispensable among the edges incident to vertices in $U_{\leq L_i}$.

Therefore, with probability $1 - o(M \cdot n^{-17/8}) = 1 - o(n^{-2})$, for all $i \in [1, M]$, strictly less than g_i edges are dispensable among the edges incident to vertices in $U_{\leq L_i}$. This completes the proof for Property 3 of the statement. □

We condition on the occurrence of the high probability event in Property 3 of Lemma 15. Let $\mathcal{E}$ denote this event.

We say that a dispensable edge is *trivial* if it is incident to v or incident to w, and *non-trivial* otherwise. Let E_0 be the set of trivial dispensable edges. Let E_1 be the set of non-trivial dispensable edges that are incident to vertices in $U_{\leq L_1}$. The event $\mathcal{E}$ implies that strictly less than $g_1(= 3)$ edges are dispensable among the edges incident to vertices in $U_{\leq L_1}$. Hence $|E_0| + |E_1| \leq 2$.

Let $F_0 \subseteq U_1$ be the set of vertices $u \in U_1$ such that u is not incident to any trivial dispensable edge. We claim that $|F_0| \geq 2\Delta - 2|E_0|$. If $E_0 = \emptyset$, it is clear that $|F_0| = 2\Delta$. If $E_0 \neq \emptyset$, there are three cases for each trivial dispensable edge in E_0: (1) a self-loop at v or at w, (2) a parallel edge incident to v or incident to w, and (3) an edge (v, u) when u is a neighbor of w, or an edge (w, u) when u is a neighbor of v. In all the three cases, the existence of each trivial dispensable edge in E_0 decreases the size of F_0 by at most 2. Hence $|F_0| \geq 2\Delta - 2|E_0|$.

For each $u \in F_0$, define

$$T(u) = \{x \in U_{\leq L_1} \mid \ell(x) = \delta(x, u) + 1\}.$$

Let $F \subseteq F_0$ be the set of vertices $u \in F_0$ such that $T(u)$ contains no vertex incident to a dispensable edge in E_1. Since each dispensable edge in E_1 is incident to two vertices, we have

$$|F| \geq |F_0| - 2|E_1| \geq 2\Delta - 2|E_0| - 2|E_1| \geq 2(\Delta - 2).$$

Since $F \subseteq F_0 \subseteq U_1$, one of v and w has at least $|F|/2 \geq \Delta - 2$ neighbors in F.

Without loss of generality, we assume that v has at least $\Delta - 2$ neighbors in F. We show that, under this assumption, $|I_k(v)| \geq (\Delta - 2 - o(1))(\Delta - 1)^{k-1}$.

Our proof proceeds in increasing order on $k \geq 1$.

First, consider any integer $k \in [1, L_1]$. Let u be any neighbor of v in F. Since $T(u)$ contains no vertex incident to a dispensable edge, $T(u)$ corresponds to a complete $(\Delta - 1)$-ary tree. Consider any vertex $x \in T(u)$ such that $\delta(v, x) = k$. Any vertex $z \in V \setminus \{v\}$ such that $\delta(v, z) + \delta(z, x) = \delta(v, x)$ belongs to the (unique) shortest x-to-u path. Since the shortest x-to-u path is completely within $T(u)$, we have $z \in T(u)$, thus the edges incident to z are indispensable. Hence x is v-interesting according to Definition 2. Since $\delta(v, x) = k$, we have $x \in I_k(v)$. There are at least $\Delta - 2$ choices of u, and for a fixed u, there are $(\Delta - 1)^{k-1}$ choices of x. Therefore, the size of $I_k(v)$ is at least $(\Delta - 2)(\Delta - 1)^{k-1}$.

Next, consider any integer $k \in [L_1 + 1, L_2]$. For any vertex $x \in I_{L_1}(v)$, define

$$T'(x) = \{y \in U_{\leq L_2} \mid \ell(y) = \delta(y, x) + L_1\}.$$

Let $F' \subseteq I_{L_1}(v)$ be the set of vertices $x \in I_{L_1}(v)$ such that $T'(x)$ contains no vertex incident to a dispensable edge. The event $\mathcal{E}$ implies that strictly less than g_2 dispensable edges are incident to vertices in $U_{\leq L_2}$. Since each dispensable edge is incident to two vertices, we have $|F'| > |I_{L_1}(v)| - 2g_2$. Let x be any vertex in F'. Since $T'(x)$ contains no vertex incident to a dispensable edge, $T'(x)$ corresponds to a complete $(\Delta - 1)$-ary tree. Consider any vertex $y \in T'(x)$ such that $\delta(v, y) = k$. Any vertex $z \in V \setminus \{v\}$ such that $\delta(v, z) + \delta(z, y) = \delta(v, y)$ belongs either to the (unique) shortest x-to-v path or to the (unique) shortest y-to-x path. In the first case, since x is v-interesting, the edges incident to z are indispensable by Definition 2. In the second case, since the shortest y-to-x path is completely within $T'(x)$, we have $z \in T'(x)$, thus the edges incident to z are indispensable. Hence y is v-interesting according to Definition 2. Since $\delta(v, y) = k$, we have $y \in I_k(v)$. There are $|F'| > |I_{L_1}(v)| - 2g_2$ choices of x, and for a fixed x, there are $(\Delta - 1)^{k - L_1}$ choices of y. Therefore,

$$\begin{aligned} |I_k(v)| &> (|I_{L_1}(v)| - 2g_2) \cdot (\Delta - 1)^{k - L_1} \\ &\geq ((\Delta - 2)(\Delta - 1)^{L_1 - 1} - 2g_2) \cdot (\Delta - 1)^{k - L_1} \\ &= (\Delta - 2 - o(1/M))(\Delta - 1)^{k-1}, \end{aligned}$$

where the equality follows because $g_2 = o((\Delta - 1)^{L_1}/M)$ from Lemma 15 and since $\Delta = O(1)$.

We move on to larger values of k. Let i be any integer in $[3, M]$. From Lemma 15, strictly less than g_i edges are dispensable among the edges incident to vertices in $U_{\leq L_i}$ and $g_i = o\left((\Delta - 1)^{L_{i-1}}/M\right)$. For any integer $k \in [L_{i-1} + 1, L_i]$, by extending the previous argument, we have

$$|I_k(v)| > (\Delta - 2 - i \cdot o(1/M))(\Delta - 1)^{k-1} = (\Delta - 2 - o(1))(\Delta - 1)^{k-1}.$$

We conclude that for any $k \in [1, L_M]$, we have $|I_k(v)| \geq (\Delta - 2 - o(1))(\Delta - 1)^{k-1}$. In the other case that w has at least $\Delta - 2$ neighbors in F, similarly, we have $|I_k(w)| \geq (\Delta - 2 - o(1))(\Delta - 1)^{k-1}$. Hence $|I_k(v) \cup I_k(w)| \geq (\Delta - 2 - o(1))(\Delta - 1)^{k-1}$. The event $\mathcal{E}$, on which the above analysis is conditioned, occurs with probability $1 - o(n^{-2})$ according to Lemma 15. Therefore, with probability $1 - o(n^{-2})$, we have

$$|I_k(v) \cup I_k(w)| \geq (\Delta - 2 - o(1))(\Delta - 1)^{k-1}, \text{ for any } k \in [1, L_M].$$

Again by Lemma 15, we have $L_M \geq \lceil \log_{\Delta-1}(3n/\log n) \rceil + 2$. Thus the above inequality holds for any positive integer $k \leq \lceil \log_{\Delta-1}(3n/\log n) \rceil + 2$.

We complete the proof of Lemma 14.

3.2.3 Proof of the Structural Lemma (Lemma 11) using Lemmas 13 and 14

We set $k = \lceil \log_{\Delta-1}(3n/\log n) \rceil + 2$. By Lemma 13, $|D(v, w)| \geq |I_k(v) \cup I_k(w)|$. By Lemma 14, with probability $1 - o(n^{-2})$, we have

$$|I_k(v) \cup I_k(w)| > (\Delta - 2 - o(1))(\Delta - 1)^{k-1} \geq (\Delta - 2 - o(1))(\Delta - 1) \cdot (3n/\log n),$$

where the last inequality follows from the definition of k. Since $\Delta \geq 3$, we have $(\Delta - 2 - o(1))(\Delta - 1) > 1$. Thus with probability $1 - o(n^{-2})$, we have $|I_k(v) \cup I_k(w)| > 3n/\log n$, which implies that $|D(v, w)| > 3n/\log n$.

We complete the proof of Lemma 11.

3.3 Proof of Lemma 10 using the Structural Lemma

Let G be a random graph and let S be a random subset of vertices, both defined in the statement of Lemma 10. According to Lemma 9, $\mathbb{E}_{G,S}\big[|\hat{E} \setminus E|\big] \leq \mathbb{E}_G\big[|B|\big] + o(1)$. It suffices to prove that $\mathbb{E}_G\big[|B|\big] = o(1)$.

First, we consider the case when $\Delta = O(1)$ is such that $\Delta \geq 3$. Our analysis is based on the configuration model. Let G' be a multigraph corresponding to a uniformly random configuration. Let $\mathbb{E}_{G'}\big[|B|\big]$ denote the expected size of the set B defined on G'. Since each Δ-regular graph corresponds to the same number of configurations and because the probability spaces of configurations and of Δ-regular graphs, respectively, are uniform, we have $\mathbb{E}_G\big[|B|\big] \leq \mathbb{E}_{G'}\big[|B|\big]/p$, where p is the probability that G' is both simple and connected. According to [44], when $\Delta \geq 3$, $p \sim e^{(1-\Delta^2)/4}$, which is constant since $\Delta = O(1)$. Thus $\mathbb{E}_G\big[|B|\big] = O(\mathbb{E}_{G'}\big[|B|\big])$.

In order to bound $\mathbb{E}_{G'}\big[|B|\big]$, consider any vertex pair $\{v, w\}$ in G' such that $\delta(v, w) \geq 2$. From the Structural Lemma (Lemma 11), the event $|D(v, w)| \leq 3n/\log n$ occurs with probability $o(n^{-2})$. Equivalently, the event $|D(v, w)| \leq 3n \cdot (\log n)/s$ occurs with probability $o(n^{-2})$, since $s = \log^2 n$. Thus the event $\{v, w\} \in B$ occurs with probability $o(n^{-2})$ according to the definition of B in Lemma 9. There are $n(n-1)/2$ vertex pairs $\{v, w\}$ in G'. By linearity of expectation, $\mathbb{E}_{G'}\big[|B|\big]$ is at most $o(n^{-2}) \cdot n(n-1)/2 = o(1)$. Hence $\mathbb{E}_G\big[|B|\big] = O(\mathbb{E}_{G'}\big[|B|\big]) = o(1)$.

In the special case when $\Delta = 2$, a 2-regular graph G is a ring. Consider any vertex pair $\{v, w\}$ in G such that $\delta(v, w) \geq 2$. It is easy to see that at least $n - 4$ vertices u in the ring G are such that $|\delta(u, v) - \delta(u, w)| > 1$, so $|D(v, w)| \geq n - 4$ by Definition 1. When n is large enough, $n - 4 > 3n/\log n$, so $|D(v, w)| > 3n/\log n$. Equivalently, we have $|D(v, w)| > 3n \cdot (\log n)/s$, since $s = \log^2 n$. Thus $\{v, w\} \notin B$ according to the definition of B in Lemma 9. Therefore, $B = \emptyset$ and $\mathbb{E}_G\big[|B|\big] = 0$.

We conclude that $\mathbb{E}_G\big[|B|\big] = o(1)$ for any $\Delta = O(1)$. Thus $\mathbb{E}_{G,S}\big[|\hat{E} \setminus E|\big] \leq \mathbb{E}_G\big[|B|\big] + o(1) = o(1)$.

We complete the proof of Lemma 10.

Remark 16. *With more care in the construction of the sequences in Lemma 15, we can improve the bound in Property 2 of Lemma 15 by* $L_M \geq \lceil \log_{\Delta-1}(3n/(\log\log n)^{2+\epsilon}) \rceil + 2$*, for any* $\epsilon > 0$*. As a result, the range of* k *in Lemma 14 can be extended to* $k \leq \lceil \log_{\Delta-1}(3n/(\log\log n)^{2+\epsilon}) \rceil + 2$*, and consequently, the event in Lemma 11 can be replaced by* $|D(v, w)| > 3n/(\log\log n)^{2+\epsilon}$*. Therefore, Lemma 10 holds for* $s = \log n \cdot (\log\log n)^{2+\epsilon}$*. This implies that the parameter* s *in Theorem 1 can be reduced to* $\log n \cdot (\log\log n)^{2+\epsilon}$*.*

4 Other Reconstruction Models (Proofs of Corollaries 2 and 3)

In this section, we study the reconstruction of random Δ-regular graphs in the all-distances query model and in the betweenness query model.

We start by presenting another reconstruction algorithm for uniformly random Δ-regular graphs in the distance query model, called Simple-Modified (see Algorithm 2). On a uniformly random Δ-regular graph, with high probability, Simple-Modified succeeds the reconstruction and returns the set of edges of the graph; otherwise it returns "failure". On other graphs, Simple-Modified may output incorrect results. The first three lines in Simple-Modified are the same as in Simple, by setting the parameter s to $\log^2 n$. The differences with Simple are in the last two lines of the algorithm.

By extending the analysis from Section 3, we observe that the set $\hat{E}$ constructed in Simple-Modified equals the edge set E with high probability, in Lemma 17.

Algorithm 2 SIMPLE-MODIFIED (V)

$S \leftarrow$ sample of $s = \log^2 n$ vertices selected uniformly and independently at random from V
for $u \in S$ and $v \in V$ **do** QUERY(u, v)
$\hat{E} \leftarrow$ set of vertex pairs $\{a, b\} \subseteq V$ such that, for all $u \in S$, $|\delta(u, a) - \delta(u, b)| \leq 1$
if $|\hat{E}| = \Delta n/2$ **then return** $\hat{E}$
else return "failure"

Lemma 17. *Let G be a uniformly random Δ-regular graph with $\Delta = O(1)$. Let $s = \log^2 n$. Let $S \subseteq V$ be a set of s vertices selected uniformly and independently at random from V. With probability $1 - o(1)$, $|\hat{E}| = \Delta n/2$. In addition, the event $|\hat{E}| = \Delta n/2$ implies $\hat{E} = E$.*

Proof. From Lemma 10, $\mathbb{E}_{G,S}\left[|\hat{E} \setminus E|\right] = o(1)$. By Markov's inequality, the event that $|\hat{E} \setminus E| \geq 1$ occurs with probability $o(1)$. Thus with probability $1 - o(1)$, we have $\hat{E} \subseteq E$. On the other hand, $E \subseteq \hat{E}$ by Fact 7. Therefore, the event that $\hat{E} = E$ occurs with probability $1 - o(1)$, and this event occurs if and only if $|\hat{E}| = |E|$. The statement follows since $|E| = \Delta n/2$ in a Δ-regular graph. □

From Lemma 17, if SIMPLE-MODIFIED returns a set of edges $\hat{E}$, then the output is correct, i.e., equals E.

Corollary 18. *Let G be a uniformly random Δ-regular graph with $\Delta = O(1)$. In the distance query model,* SIMPLE-MODIFIED *(Algorithm 2) uses $\tilde{O}(n)$ queries and reconstructs the graph with probability $1 - o(1)$.*

4.1 All-Distances Query Model (Proof of Corollary 2)

We extend SIMPLE-MODIFIED from the distance query model to the all-distances query model. Observe that in SIMPLE-MODIFIED, the distance queries are performed between each sampled vertex $u \in S$ and all vertices in the graph. This is equivalent to a single query at each sampled vertex $u \in S$ in the all-distances query model. Hence the distance queries in SIMPLE-MODIFIED correspond to $|S| = \log^2 n$ all-distances queries. Therefore, in the all-distances query model, an algorithm equivalent to SIMPLE-MODIFIED uses $\log^2 n$ all-distances queries and reconstructs the graph with probability $1 - o(1)$.

4.2 Betweenness Query Model (Proof of Corollary 3)

In the betweenness query model, Abrahamsen et al. [1] showed that $\tilde{O}(\Delta^2 \cdot n)$ betweenness queries suffice to compute the distances from a given vertex to all vertices in the graph (it is implicit in Lemma 16 from [1]), so an all-distances query can be simulated by $\tilde{O}(\Delta^2 \cdot n)$ betweenness queries, where $\Delta = O(1)$. As a consequence of Corollary 2, there is an algorithm that uses $\tilde{O}(\Delta^2 \cdot n \cdot \log^2 n) = \tilde{O}(n)$ betweenness queries and reconstructs the graph with probability $1 - o(1)$.

5 Metric Dimension (Proof of Corollary 5)

In this section, we study the metric dimension of random Δ-regular graphs. To begin with, we show an elementary structural property of random Δ-regular graphs, in Lemma 19, based on a classical result on those graphs.

Lemma 19. *Let $G = (V, E)$ be a uniformly random Δ-regular graph with $\Delta = O(1)$. With probability $1 - o(1)$, for any edge (a, b) of the graph G, there exists a vertex $c \in V \setminus \{a, b\}$ that is adjacent to b but is not adjacent to a.*

Proof. First, consider the case when $\Delta = 2$. A 2-regular graph is a ring. Let (a, b) be any edge of the graph. The vertex b has two neighbors, the vertex a and another vertex, let it be c. We have $c \in V \setminus \{a, b\}$ and c is not adjacent to a (as soon as $n > 3$). The statement of the lemma follows.

Next, consider the case when $\Delta = O(1)$ is such that $\Delta \geq 3$. Let $\mathcal{E}$ denote the event that, for any edge (a, b) of G, there do not exist two vertices c_1 and c_2 in G, such that all of the 4 edges $(a, c_1), (a, c_2), (b, c_1), (b, c_2)$ belong to G. We show that $\mathcal{E}$ occurs with probability $1 - o(1)$. Indeed, if for some edge (a, b) of G, there exist two vertices c_1 and c_2 such that $(a, c_1), (a, c_2), (b, c_1), (b, c_2)$ are edges of G, then the induced subgraph on $\{a, b, c_1, c_2\}$ consists of at least 5 edges. A classical result on random Δ-regular graphs shows that, for any constant integer k, the probability that there exists an induced subgraph of k vertices with at least $k + 1$ edges is $o(1)$, see, e.g., Lemma 11.12 in [18]. Therefore, $\mathcal{E}$ occurs with probability $1 - o(1)$.

We condition on the occurrence of $\mathcal{E}$. For any edge (a, b) of G, let $N(a)$ be the set of $\Delta - 1$ neighbors of a that are different from b, and let $N(b)$ be the set of $\Delta - 1$ neighbors of b that are different from a. Since $\Delta \geq 3$,

we have $|N(a)| = |N(b)| \geq 2$. The event $\mathcal{E}$ implies that $N(a) \neq N(b)$, so there exists a vertex $c \in N(b) \setminus N(a)$. By definition, c is adjacent to b but is not adjacent to a, and $c \in V \setminus \{a, b\}$. Since $\mathcal{E}$ occurs with probability $1 - o(1)$, we conclude that, with probability $1 - o(1)$, for any edge (a, b) of the graph G, there exists a vertex $c \in V \setminus \{a, b\}$ that is adjacent to b but is not adjacent to a. □

Definition 3 (e.g., [5, 13])**.** A subset of vertices $S \subseteq V$ is a *resolving set* for a graph $G = (V, E)$ if, for any pair of vertices $\{a, b\} \subseteq V$, there is a vertex $u \in S$ such that $\delta(u, a) \neq \delta(u, b)$. The *metric dimension* of G is the smallest size of a resolving set for G.

Based on the analysis of Simple from Lemma 17 and the structural property from Lemma 19, we show that, with high probability, a random subset of $\log^2 n$ vertices is a resolving set for a random Δ-regular graph, in Lemma 20.

Lemma 20. *Let $G = (V, E)$ be a uniformly random Δ-regular graph with $\Delta = O(1)$. Let $S \subseteq V$ be a sample of $s = \log^2 n$ vertices selected uniformly and independently at random from V. With probability $1 - o(1)$, the set S is a resolving set for the graph G.*

Proof. Let $\mathcal{E}_1$ denote the event that, for any edge (a, b) of the graph G, there exists a vertex $c \in V \setminus \{a, b\}$ that is adjacent to b but is not adjacent to a. By Lemma 19, the event $\mathcal{E}_1$ occurs with probability $1 - o(1)$. Let $\mathcal{E}_2$ denote the event $\hat{E} = E$. By Lemma 17, the event $\mathcal{E}_2$ occurs with probability $1 - o(1)$. Thus with probability $1 - o(1)$, both events $\mathcal{E}_1$ and $\mathcal{E}_2$ occur simultaneously. We condition on the occurrences of both events $\mathcal{E}_1$ and $\mathcal{E}_2$ in the subsequent analysis.

First, consider any vertex pair $\{a, b\} \subseteq V$ such that $\delta(a, b) \geq 2$. The event $\mathcal{E}_2$ implies that $\{a, b\} \notin \hat{E}$. By definition, there exists some vertex $u \in S$ such that $|\delta(u, a) - \delta(u, b)| \geq 2$, which implies that $\delta(u, a) \neq \delta(u, b)$.

Next, consider any vertex pair $\{a, b\} \subseteq V$ such that $\delta(a, b) = 1$. The event $\mathcal{E}_1$ implies that there exists a vertex $c \in V \setminus \{a, b\}$ that is adjacent to b but is not adjacent to a. Since $\delta(a, c) \geq 2$, the event $\mathcal{E}_2$ implies that $\{a, c\} \notin \hat{E}$. By definition, there exists some vertex $u \in S$ such that $|\delta(u, a) - \delta(u, c)| \geq 2$. Using an elementary inequality of $|x - y| + |y - z| \geq |x - z|$ for any three real numbers x, y, and z, we have

$$\begin{aligned}
|\delta(u, a) - \delta(u, b)| &\geq |\delta(u, a) - \delta(u, c)| - |\delta(u, b) - \delta(u, c)| \\
&\geq |\delta(u, a) - \delta(u, c)| - \delta(b, c) && \text{(by the triangle inequality)} \\
&\geq 2 - \delta(b, c) && \text{(by the definition of } u\text{)} \\
&\geq 1 && \text{(since } (b, c) \text{ is an edge in } G\text{)}.
\end{aligned}$$

Thus $\delta(u, a) \neq \delta(u, b)$.

Therefore, conditioned on the occurrences of both events $\mathcal{E}_1$ and $\mathcal{E}_2$, for any vertex pair $\{a, b\} \subseteq V$, there exists a vertex $u \in S$ such that $\delta(u, a) \neq \delta(u, b)$.

We conclude that, with probability $1 - o(1)$, the set S is a resolving set for G. □

From Lemma 20, with probability $1 - o(1)$, the metric dimension of a random Δ-regular graph is at most $\log^2 n$. This completes the proof of Corollary 5.

6 Reconstruction of Bounded-Degree Graphs (Proof of Theorem 6)

In this section, we analyze Simple (Algorithm 1) on general graphs of bounded degree in the distance query model. Recall that a set B of vertex pairs $\{a, b\} \subseteq V$ is defined in Lemma 9. For every vertex $a \in V$, we define the set of vertices $B(a) \subseteq V$ as

$$B(a) = \big\{b \in V \mid \{a, b\} \in B\big\}.$$

Intuitively, $B(a)$ consists of the vertices $b \in V$ that has few distinguishers with a. We bound the size of the set $B(a)$ for any vertex a, in Lemma 21.

Lemma 21. *Let G be a general graph of bounded degree Δ. For any vertex $a \in V$, $|B(a)| \leq 9\Delta^3 \cdot n^2 \cdot (\log^2 n)/s^2$.*

We defer the proof of Lemma 21 for the moment and first show how it implies Theorem 6.

Proof of Theorem 6 using Lemma 21. By Lemma 8, Simple is a reconstruction algorithm using $n \cdot s + |\hat{E}|$ distance queries, and in addition, Simple can be parallelized using 2 rounds. It remains to further analyze the query complexity.

From Fact 7, $|\hat{E}| = |E| + |\hat{E} \setminus E|$. Since the graph has bounded degree Δ, $|E| \leq \Delta n$. From Lemma 9, $\mathbb{E}_S\big[|\hat{E} \setminus E|\big] \leq |B| + o(1)$. Therefore, the expected number of distance queries in Simple is at most $n \cdot s + \Delta n + |B| + o(1)$. It suffices to analyze $|B|$.

Observe that $|B| \leq \sum_{a \in V} |B(a)|$ by definition of $\{B(a)\}_{a \in V}$. From Lemma 21, $|B(a)| \leq 9\Delta^3 \cdot n^2 \cdot (\log^2 n)/s^2$, for any vertex $a \in V$. Hence $|B| \leq (9\Delta^3 \cdot n^2 \cdot (\log^2 n)/s^2) \cdot n$. Thus the expected number of distance queries in SIMPLE is at most $n \cdot s + \Delta n + (9\Delta^3 \cdot n^2 \cdot (\log^2 n)/s^2) \cdot n + o(1)$, which is $\tilde{O}(n^{5/3})$ since $s = n^{2/3}$ and $\Delta = O(\text{polylog}\, n)$. □

The rest of the section is dedicated to prove Lemma 21.

Let a be any vertex in V. Let T be an (arbitrary) shortest-path tree rooted at a and spanning all vertices in V. For any vertex $b \in V$, let *the shortest a-to-b path* denote the path between a and b in the tree T. To simplify the presentation, we assume that, for any $b \in B(a)$, $\delta(a, b)$ is even, so that the *midpoint vertex* of the shortest a-to-b path is uniquely defined. We extend our analysis to the general setting in the end of the section.

For any vertex $m \in V$, define the set $B(a, m) \subseteq B(a)$ as

$$B(a, m) = \big\{b \in B(a) \mid \text{the midpoint vertex of the shortest } a\text{-to-}b \text{ path is } m\big\}.$$

Define the set $M(a) \subseteq V$ as

$$M(a) = \big\{m \in V \mid B(a, m) \neq \emptyset\big\}.$$

In other words, $M(a)$ consists of the vertices $m \in V$ such that m is the midpoint vertex of the shortest a-to-b path for some $b \in B(a)$. From the construction, we have

$$B(a) = \bigcup_{m \in M(a)} B(a, m). \tag{2}$$

In order to bound the size of $B(a)$, first we bound the size of $B(a, m)$ for any midpoint $m \in M(a)$, in Lemma 22, and then we bound the number of distinct midpoints, in Lemma 23.

Lemma 22. *For any $m \in M(a)$, $|B(a, m)| \leq 3\Delta \cdot n \cdot (\log n)/s$.*

Proof. For any $b \in B(a, m)$, the vertex m is the midpoint vertex of the shortest a-to-b path by definition. From the assumption, $\delta(a, b)$ is even for any $b \in B(a, m)$, so there exists some positive integer ℓ, such that $\delta(m, a) = \ell$ and $\delta(m, b) = \ell$ for any $b \in B(a, m)$.

For every neighbor m' of m such that $\delta(a, m') = \delta(a, m) + 1$, define a set $Y(m') \subseteq B(a, m)$ that consists of the vertices $b \in B(a, m)$ such that m' is on the shortest a-to-b path. Let $\hat{m}$ be a neighbor of m such that $\delta(a, \hat{m}) = \delta(a, m) + 1$ and that $|Y(\hat{m})|$ is maximized, see Fig. 2. Since the graph has bounded degree Δ, we have $|B(a, m)| \leq \Delta \cdot |Y(\hat{m})|$. It suffices to bound $|Y(\hat{m})|$.

The main observation is that any vertex of $Y(\hat{m})$ distinguishes a and any other vertex of $Y(\hat{m})$. To see this, let b_0 be any vertex in $Y(\hat{m})$. By definition, $\delta(a, \hat{m}) = \delta(a, m) + 1 = \ell + 1$. Since $\hat{m}$ is on the shortest a-to-b_0 path, we have $\delta(\hat{m}, b_0) = \delta(a, b_0) - \delta(a, \hat{m}) = \ell - 1$, thus $\delta(\hat{m}, b_0) = \delta(\hat{m}, a) - 2$. For any vertex $b_1 \in Y(\hat{m})$, from the triangle inequalities on δ, we have

$$\delta(b_1, b_0) \leq \delta(b_1, \hat{m}) + \delta(\hat{m}, b_0) = \delta(b_1, \hat{m}) + \delta(\hat{m}, a) - 2 = \delta(b_1, a) - 2.$$

According to Definition 1, the vertex b_1 distinguishes a and b_0, and equivalently, $b_1 \in D(a, b_0)$. Thus we have $Y(\hat{m}) \subseteq D(a, b_0)$, hence $|Y(\hat{m})| \leq |D(a, b_0)| \leq 3n \cdot (\log n)/s$ using the fact that $b_0 \in Y(\hat{m}) \subseteq B(a)$ and the definition of B in Lemma 9.

We conclude that $|B(a, m)| \leq \Delta \cdot |Y(\hat{m})| \leq 3\Delta \cdot n \cdot (\log n)/s$. □

Lemma 23. $|M(a)| \leq 3\Delta \cdot n \cdot (\log n)/s$.

Proof. For each vertex $m \in M(a)$, let x_m denote the second-to-last vertex on the shortest a-to-m path. Let $X(a) \subseteq V$ denote the set of vertices x_m for all $m \in M(a)$. See Fig. 3. Since G has bounded degree Δ, we have $|M(a)| \leq \Delta \cdot |X(a)|$. It suffices to bound $|X(a)|$.

Let b^* be a vertex in $B(a)$ such that $\delta(a, b^*)$ is maximized. From the assumption, $\delta(a, b^*)$ is even. Let ℓ be a positive integer such that $\delta(a, b^*) = 2\ell$.

The main observation is that any vertex of $X(a)$ distinguishes a and b^*. To see this, let x be any vertex in $X(a)$. Let m be any vertex in $M(a)$ such that x is the second-to-last vertex on the shortest a-to-m path.[8] We have $\delta(a, m) \leq \ell$ and $\delta(a, x) = \delta(a, m) - 1 \leq \ell - 1$. By the triangle inequality on the distances, $\delta(b^*, x) \geq \delta(a, b^*) - \delta(a, x) \geq 2\ell - (\ell - 1) = \ell + 1$. Thus $\delta(b^*, x) - \delta(a, x) \geq 2$. According to Definition 1, the vertex x distinguishes a and b^*, and equivalently, $x \in D(a, b^*)$. Thus $X(a) \subseteq D(a, b^*)$, hence $|X(a)| \leq |D(a, b^*)| \leq 3n \cdot (\log n)/s$ using the fact that $b^* \in B(a)$ and the definition of B in Lemma 9.

We conclude that $|M(a)| \leq \Delta \cdot |X(a)| \leq 3\Delta \cdot n \cdot (\log n)/s$. □

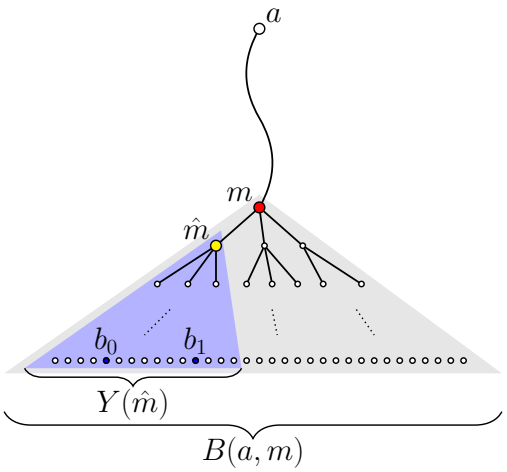

Figure 2: The vertex m is the midpoint of the shortest path between a and any vertex in $B(a,m)$. The vertex $\hat{m}$ is a well-chosen neighbor of m. Consider any vertex $b_0 \in Y(\hat{m})$. We can show that any vertex $b_1 \in Y(\hat{m})$ distinguishes a and b_0.

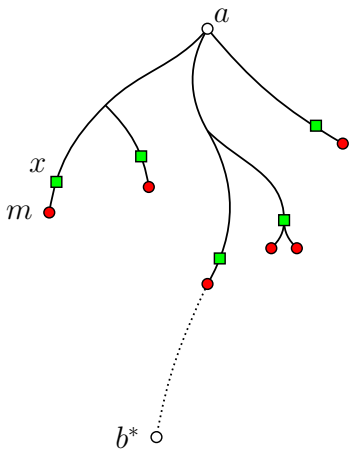

Figure 3: Solid circular nodes represent the vertices $m \in M(a)$. Solid curves represent the shortest a-to-m paths. Solid square nodes represent the vertices in $X(a)$. Let b^* denote a vertex in $B(a)$ that is farthest from a. We can show that any vertex $x \in X(a)$ distinguishes a and b^*.

From Eq. (2), $|B(a)| \leq \sum_{m \in M(a)} |B(a,m)|$. From Lemma 22, $|B(a,m)| \leq 3\Delta \cdot n \cdot (\log n)/s$ for every $m \in M(a)$. From Lemma 23, $|M(a)| \leq 3\Delta \cdot n \cdot (\log n)/s$. Therefore, $|B(a)| \leq 9\Delta^2 \cdot n^2 \cdot (\log^2 n)/s^2$.

Finally, consider the general setting in which $\delta(a,b)$ is not necessarily even for any $b \in B(a)$. For a vertex m on the shortest a-to-b path, we say that m is the *midpoint vertex* of that path if $\delta(a,m) = \lfloor \delta(a,b)/2 \rfloor$. The definitions of $B(a,m)$ and $M(a)$ remain the same. Lemma 23 holds in the same way. In Lemma 22, the upper bound of $|B(a,m)|$ is replaced by $3\Delta^2 \cdot n \cdot (\log n)/s$. Indeed, to extend the proof of Lemma 22, instead of considering vertex m' (resp., vertex $\hat{m}$) that is a neighbor of m, we consider m' (resp., $\hat{m}$) that is at distance 2 from m. We have $|B(a,m)| \leq \Delta^2 \cdot |Y(\hat{m})|$. The bound $|Y(\hat{m})| \leq 3n \cdot (\log n)/s$ remains the same, so we have $|B(a,m)| \leq 3\Delta^2 \cdot n \cdot (\log n)/s$. Hence $|B(a)| \leq 9\Delta^3 \cdot n^2 \cdot (\log^2 n)/s^2$.

We complete the proof of Lemma 21. Therefore, we obtain Theorem 6.

References

[1] Mikkel Abrahamsen, Greg Bodwin, Eva Rotenberg, and Morten Stöckel. Graph Reconstruction with a Betweenness Oracle. In *Symposium on Theoretical Aspects of Computer Science*, pages 5:1–5:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016.

[2] Dimitris Achlioptas, Aaron Clauset, David Kempe, and Cristopher Moore. On the bias of traceroute sampling: Or, power-law degree distributions in regular graphs. *Journal of the ACM*, 56(4):21:1–21:28, 2009.

[8] Such a vertex m exists according to the construction of $X(a)$.

[3] Ramtin Afshar, Michael T. Goodrich, Pedro Matias, and Martha C. Osegueda. Reconstructing biological and digital phylogenetic trees in parallel. In *European Symposium on Algorithms*, volume 173, pages 3:1–3:24. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020.

[4] Animashree Anandkumar, Avinatan Hassidim, and Jonathan Kelner. Topology discovery of sparse random graphs with few participants. *Random Structures & Algorithms*, 43(1):16–48, 2013.

[5] Robert F. Bailey and Peter J. Cameron. Base size, metric dimension and other invariants of groups and graphs. *Bulletin of the London Mathematical Society*, 43(2):209–242, 2011.

[6] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[7] Zuzana Beerliova, Felix Eberhard, Thomas Erlebach, Alexander Hall, Michael Hoffmann, Matús Mihal'ak, and L. Shankar Ram. Network discovery and verification. *IEEE Journal on Selected Areas in Communications*, 24(12):2168–2181, 2006.

[8] Vincent D. Blondel, Jean-Loup Guillaume, Julien M. Hendrickx, and Raphaël M. Jungers. Distance distribution in random graphs and application to network exploration. *Physical Review E*, 76(6):066101, 2007.

[9] Béla Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316, 1980.

[10] Béla Bollobás. Distinguishing vertices of random graphs. *North-Holland Mathematics Studies*, 62:33–49, 1982.

[11] Béla Bollobás, Dieter Mitsche, and Paweł Prałat. Metric dimension for random graphs. *The Electronic Journal of Combinatorics*, 20(4):P1, 2013.

[12] José Cáceres, Carmen Hernando, Mercè Mora, Ignacio M. Pelayo, María L. Puertas, Carlos Seara, and David R. Wood. On the metric dimension of cartesian products of graphs. *SIAM Journal on Discrete Mathematics*, 21(2):423–441, 2007.

[13] 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(1-3):99–113, 2000.

[14] Thomas Erlebach, Alexander Hall, Michael Hoffmann, and Matúš Mihal'ák. Network discovery and verification with distance queries. *Algorithms and Complexity*, pages 69–80, 2006.

[15] Thomas Erlebach, Alexander Hall, and Matúš Mihal'ák. Approximate discovery of random graphs. In *International Symposium on Stochastic Algorithms*, pages 82–92. Springer, 2007.

[16] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. *ACM SIGCOMM*, 29(4):251–262, 1999.

[17] Florent Foucaud and Guillem Perarnau. Bounds for identifying codes in terms of degree parameters. *Electronic Journal of Combinatorics*, 19(P32), 2012.

[18] Alan Frieze and Michał Karoński. Introduction to random graphs. `https://www.math.cmu.edu/~af1p/BOOK.pdf`.

[19] Alan Frieze, Ryan Martin, Julien Moncel, Miklós Ruszinkó, and Cliff Smyth. Codes identifying sets of vertices in random networks. *Discrete Mathematics*, 307(9):1094–1107, 2007.

[20] Jean-Loup Guillaume and Matthieu Latapy. Complex network metrology. *Complex systems*, 16(1):83, 2005.

[21] Frank Harary and Robert A. Melter. On the metric dimension of a graph. *Ars Combinatoria*, 2(191-195), 1976.

[22] Jotun J. Hein. An optimal algorithm to reconstruct trees from additive distance data. *Bulletin of Mathematical Biology*, 51(5):597–603, 1989.

[23] Carmen Hernando, Merce Mora, Ignacio M. Pelayo, Carlos Seara, and David R. Wood. Extremal graph theory for metric dimension and diameter. *Electronic Notes in Discrete Mathematics*, 29:339–343, 2007. European Conference on Combinatorics, Graph Theory and Applications.

[24] Imran Javaid, M. Tariq Rahim, and Kashif Ali. Families of regular graphs with constant metric dimension. *Utilitas mathematica*, 75:21–34, 2008.

[25] Mihajlo Jovanović, Fred Annexstein, and Kenneth Berman. Modeling peer-to-peer network topologies through small-world models and power laws. In *IX Telecommunications Forum, TELFOR*, pages 1–4. Citeseer, 2001.

[26] Sampath Kannan, Eugene L. Lawler, and Tandy Warnow. Determining the evolutionary tree using experiments. *Journal of Algorithms*, 21(1):26 – 50, 1996.

[27] Sampath Kannan, Claire Mathieu, and Hang Zhou. Graph reconstruction and verification. *ACM Transactions on Algorithms*, 14(4):1–30, 2018.

[28] Mark G. Karpovsky, Krishnendu Chakrabarty, and Lev B. Levitin. On a new class of codes for identifying vertices in graphs. *IEEE Transactions on Information Theory*, 44(2):599–611, 1998.

[29] Samir Khuller, Balaji Raghavachari, and Azriel Rosenfeld. Landmarks in graphs. *Discrete applied mathematics*, 70(3):217–229, 1996.

[30] Valerie King, Li Zhang, and Yunhong Zhou. On the complexity of distance-based evolutionary tree reconstruction. In *Symposium on Discrete Algorithms*, pages 444–453. SIAM, 2003.

[31] Anukool Lakhina, John W. Byers, Mark Crovella, and Peng Xie. Sampling biases in IP topology measurements. In *Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 1, pages 332–341. IEEE, 2003.

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

[33] Elchanan Mossel and Jiaming Xu. Seeded graph matching via large neighborhood statistics. *Random Structures & Algorithms*, 57(3):570–611, 2020.

[34] Mark E. J. Newman, Duncan J. Watts, and Steven H. Strogatz. Random graph models of social networks. *Proceedings of the national academy of sciences*, 99(suppl 1):2566–2572, 2002.

[35] Gergely Odor and Patrick Thiran. Sequential metric dimension for random graphs, 2020. `arXiv:1910.10116`.

[36] Ortrud R. Oellermann and Joel Peters-Fransen. The strong metric dimension of graphs and digraphs. *Discrete Applied Mathematics*, 155(3):356–364, 2007.

[37] Yunior Ramírez-Cruz, Ortrud R. Oellermann, and Juan A. Rodríguez-Velázquez. The simultaneous metric dimension of graph families. *Discrete Applied Mathematics*, 198:241–250, 2016.

[38] Lev Reyzin and Nikhil Srivastava. Learning and verifying graphs using queries with a focus on edge counting. In *Algorithmic Learning Theory*, pages 285–297. Springer, 2007.

[39] Guozhen Rong, Wenjun Li, Yongjie Yang, and Jianxin Wang. Reconstruction and verification of chordal graphs with a distance oracle. *Theoretical Computer Science*, 859:48–56, 2021.

[40] András Sebő and Eric Tannier. On metric generators of graphs. *Mathematics of Operations Research*, 29(2):383–393, 2004.

[41] Sandeep Sen and V. N. Muralidhara. The covert set-cover problem with application to network discovery. In *WALCOM*, pages 228–239. Springer, 2010.

[42] Peter J. Slater. Leaves of trees. In *Southeastern Conference on Combinatorics, Graph Theory, and Computing*, pages 549–559, 1975.

[43] Michael S. Waterman, Temple F. Smith, M. Singh, and W. A. Beyer. Additive evolutionary trees. *Journal of Theoretical Biology*, 64(2):199–213, 1977.

[44] Nicholas C. Wormald. Models of random regular graphs. *London Mathematical Society Lecture Note Series*, pages 239–298, 1999.