Shotgun assembly of Erdős-Rényi random graphs[*]

Julia Gaudio[†]   Elchanan Mossel[‡]

Abstract

Graph shotgun assembly refers to the problem of reconstructing a graph from a collection of local neighborhoods. In this paper, we consider shotgun assembly of Erdős–Rényi random graphs $G(n, p_n)$, where $p_n = n^{-\alpha}$ for $0 < \alpha < 1$. We consider both reconstruction up to isomorphism as well as exact reconstruction (recovering the vertex labels as well as the structure). We show that given the collection of distance-1 neighborhoods, G is exactly reconstructable for $0 < \alpha < \frac{1}{3}$, but not reconstructable for $\frac{1}{2} < \alpha < 1$. Given the collection of distance-2 neighborhoods, G is exactly reconstructable for $\alpha \in \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, \frac{3}{5}\right)$, but not reconstructable for $\frac{3}{4} < \alpha < 1$.

Keywords: random graphs; shotgun assembly.
MSC2020 subject classifications: 05C80.
Submitted to ECP on January 13, 2021, final version accepted on January 5, 2022.

1 Introduction

Graph shotgun assembly refers to the reconstruction of a graph from a collection of local neighborhoods. The terminology is inspired by DNA shotgun assembly, which is the problem of determining a DNA sequence from multiple short sequences. Mossel and Ross [9] introduced the problem of graph shotgun assembly for generative models.

We now describe the graph shotgun assembly problem. Let $G = (V, E)$ be a graph with vertices labeled from $[n] = \{1, 2, \ldots, n\}$. For a vertex $v \in V$, let $\mathcal{N}_r(v)$ be the subgraph induced by all vertices at distance at most r from v in the graph G. For each vertex $v \in V$, we observe the neighborhood $\mathcal{N}_r(v)$. The central vertex v in this neighborhood is labeled, and other vertices are unlabeled.

We say two graphs $G = (V_G, E_G)$ and $H = (V_H, E_H)$ are *isomorphic* if there exists a bijection $f : V_G \to V_H$ such that $(u, v) \in E_G \iff (f(u), f(v)) \in E_H$. In that case, we write $G \sim H$. The goal is to reconstruct G from the set of r-neighborhoods, up to isomorphism. In other words, we wish to find a graph H isomorphic to G given the unlabeled r-neighborhoods of G.

We say that two graphs $H_1(V, E_1)$ and $H_2(V, E_2)$ have the same r-neighborhoods if for all $v \in V$, the r-neighborhood around v in H_1 is isomorphic to the r-neighborhood around v in H_2. In that case we write $H_1 \sim_r H_2$. We say that the graph G is reconstructable from its r-neighborhoods if for all graphs $H = (V, E_H)$, $G \sim_r H \implies G \sim H$. In other

[†]Northwestern University. E-mail: julia.gaudio@northwestern.edu. Part of this work was completed while J.G. was at the Massachusetts Institute of Technology.
[‡]Massachusetts Institute of Technology. E-mail: elmos@mit.edu.

words, G is reconstructable from its r-neighborhoods if the only other graphs with the same r-neighborhoods are those that are isomorphic to G.

In addition to the above reconstructability question, we consider exact reconstructability. We say that a graph G is *exactly reconstructable* if it is possible to recover G (rather than just $H \sim G$). Note that exact reconstructability implies reconstructability up to isomorphism. For an example where reconstruction up to isomorphism is possible, but exact reconstruction is not, consider the graph on vertices $\{1, 2, 3, 4\}$ with edges $(1, 2)$ and $(3, 4)$. This graph has the same neighborhoods as the graph on vertices $\{1, 2, 3, 4\}$ with edges $(1, 3)$ and $(2, 4)$

Along with introducing the problem of graph shotgun assembly for generative models, [9] considered several random structure models. First, they considered the d-dimensional n-lattice, where the vertices are additionally labeled i.i.d. according to some distribution on $[q]$. Next, they considered the reconstruction of Erdős–Rényi random graphs, both in the dense and sparse case. In the sparse case where $p_n = \frac{\lambda}{n}$, it was shown that the asymptotic threshold for reconstructability is a radius of $r = \log(n)$, as long as $\lambda \neq 1$. In the dense case, it was shown that $r = 3$ suffices for reconstruction when $\lim_{n \to \infty} \frac{n p_n}{\log^2(n)} = \infty$. Finally, [9] introduced the random jigsaw puzzle, for which there has been much follow-up work ([1], [2], [6], [7], [11]). A jigsaw puzzle is an $n \times n$ grid where the border of two adjacent pieces is uniformly assigned to one of q possible shapes called "jigs." Martinsson ([6], [7]) showed that reconstruction is possible if $q \geq (2 + \epsilon)n$, and impossible if $q \leq \frac{2}{\sqrt{e}}n$. Other follow-up work includes a study of reconstruction in random regular graphs [10] and the hypercube [12].

In this paper, we continue the study of graph shotgun assembly for Erdős–Rényi random graphs. Let $G = (V, E)$ be a labeled Erdős-Rényi graph drawn from the model $G(n, p_n)$. The following general observation was used in [9] to establish reconstruction from 3-neighborhoods.

Lemma 1.1 (Lemma 2.4 in [9])**.** *If $\mathcal{N}_{r-1}(v) \not\sim \mathcal{N}_{r-1}(w)$ for all vertices $v \neq w$, then there is an algorithm for recovering the graph from r-neighborhoods.*

The algorithm is as follows: Given a neighborhood $\mathcal{N}_r(v)$, we can uniquely label any neighbor v_0 of v by examining $\mathcal{N}_{r-1}(v_0)$, which is contained within $\mathcal{N}_r(v)$. We call this algorithm the "overlap method." Using Lemma 1.1, [9] showed the following.

Theorem 1.2 (Theorem 4.5 in [9])**.** *$G(n, p_n)$ is reconstructable from its 3-neighborhoods with high probability, for p_n satisfying $\lim_{n \to \infty} \frac{n p_n}{\log^2(n)} = \infty$.*

The proof of Theorem 1.2 shows that no two vertices $u, v \in V$ have the same degree neighborhoods with high probability. (We say that two vertices u, v have the same degree neighborhoods if they have the same degree and the degrees of their neighbors are equal as multi-sets.)

In this paper, we extend the results of [9], by establishing regimes where reconstruction is possible and when it is not, given the collection of 1- or 2-neighborhoods.

1.1 Main Results

We first investigate reconstructablity from 1-neighborhoods (Section 2). First, we establish a range where reconstruction is possible using a "fingerprinting" idea from [13], which studied reconstruction in another random graph model.

Theorem 1.3. *Let $p_n = n^{-\alpha}$ for $0 < \alpha < \frac{1}{3}$. Then $G(n, p_n)$ is exactly reconstructable from its 1-neighborhoods with high probability.*

We outline the proof. For vertices u, v such that $(u, v) \in E$, let $H_{u,v}$ denote the graph induced by the common neighbors of u and v. The graph $H_{u,v}$ is a "fingerprint" for the edge (u, v) (note that $H_{u,v} \sim H_{v,u}$). The following lemma establishes that under a

uniqueness of fingerprints condition, the graph can be reconstructed.

Lemma 1.4 (Fingerprint Lemma). *Suppose $H_{u,v} \sim H_{x,y}$ if and only if (u,v) and (x,y) are the same edge. Then we can exactly reconstruct G from the collection of 1-neighborhoods.*

Given the Fingerprint Lemma, it remains to establish the uniqueness of the $H_{u,v}$ graphs.

We then show a negative result about the reconstructability of $G(n, p_n)$ from 1-neighborhoods using a counting argument.

Theorem 1.5. *Let $p_n = n^{-\alpha}$ for $\frac{1}{2} < \alpha < 1$. Then with high probability, $G(n, p_n)$ cannot be reconstructed from its 1-neighborhoods.*

Next, we consider reconstruction from 2-neighborhoods (Section 3). Theorem 1.6 establishes that $G(n, p_n)$ is exactly reconstructable from 2-neighborhoods for $p_n = n^{-\alpha}$ where $0 < \alpha < \frac{3}{5}$. When $0 < \alpha < \frac{1}{2}$, the graph has diameter 2 with high probability (Lemma 3.1). We then construct a canonical labeling, which is guaranteed by Lemma 3.2. For values $\frac{1}{2} < \alpha < \frac{3}{5}$, we demonstrate reconstructability by a fingerprint argument similar to the proof of Theorem 1.3.

Theorem 1.6. *Let $p_n = n^{-\alpha}$ for $\alpha \in \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, \frac{3}{5}\right)$. Then $G(n, p_n)$ is exactly reconstructable from its 2-neighborhoods with high probability. Moreover, there is an efficient algorithm for reconstruction when $\alpha < \frac{1}{2}$.*

On the other hand, when $\frac{3}{4} < \alpha < 1$, we show by a counting argument that $G(n, p_n)$ cannot be reconstructed from its 2-neighborhoods.

Theorem 1.7. *Let $p_n = n^{-\alpha}$ for $\frac{3}{4} < \alpha < 1$. Then with high probability, $G(n, p_n)$ cannot be reconstructed from its 2-neighborhoods.*

These above results do not cover $\alpha \in \left\{\frac{1}{2}\right\} \cup \left[\frac{3}{5}, \frac{3}{4}\right]$. Proposition 1.8 gives a partial answer for the range $\alpha \in \left(\frac{2}{3}, \frac{3}{4}\right]$.

Proposition 1.8. *Let $p_n = n^{-\alpha}$ for $\frac{2}{3} < \alpha < 1$. Then with high probability, $G(n, p_n)$ cannot be exactly reconstructed from its 2-neighborhoods using the overlap method.*

Finally, recall that the center of each neighborhood was assumed to be labeled. In Section 4, we show how to find the center if it is unlabeled.

1.2 Follow-up work

After posting of this paper to arXiv, Huang and Tikhomirov [4] contributed further results on the reconstruction of Erdős–Rényi random graphs from 1-neighborhoods. Their first result refined Theorem 1.5, showing that $G(n, p_n)$ is a.a.s. non-reconstructable from its 1-neighborhoods if p_n satisfies $p_n = \omega(n^{-1} \log n)$ and $p_n = o(n^{-\frac{1}{2}})$. Their second result showed that there exist universal constants $C, c > 0$ such that if $n^{-\frac{1}{2}} \log^C(n) \le p_n \le \frac{c}{n}$ for large n, then $G(n, p_n)$ is a.a.s. reconstructable from its 1-neighborhoods. Therefore, $p = \tilde{\Theta}(n^{-\frac{1}{2}})$ marks the transition between reconstructability and non-reconstructability, addressing an open problem that we had raised.

1.3 Preliminaries

We collect some results that will be used to establish the theorems.

Lemma 1.9 (Chernoff Bound [8]). *Let $\{X_i\}_{i=1}^m$ be independent indicator random variables, and let $X = \sum_{i=1}^m X_i$. Then for any $\epsilon > 0$,*

$$\mathbb{P}\left(X \le (1-\epsilon)\mathbb{E}[X]\right) \le \exp\left(-\frac{\epsilon^2}{2}\mathbb{E}[X]\right) \quad \text{and} \quad \mathbb{P}\left(X \ge (1+\epsilon)\mathbb{E}[X]\right) \le \exp\left(-\frac{\epsilon^2}{2+\epsilon}\mathbb{E}[X]\right).$$

Corollary 1.10. *Let X be a binomial random variable with parameters $(n, n^{-\beta})$ for constant $\beta > 0$. For any constant $c > 0$,*

$$\mathbb{P}\left(X \geq n^{1-\beta+c}\right) \leq \exp\left(-\frac{(n^c - 1)^2}{n^c + 1} n^{1-\beta}\right) = \exp\left(-\Theta(n^{1-\beta+c})\right).$$

Lemma 1.11. *Let X and Y be random variables such that conditioned on Y, X is distributed as a binomial random variable with parameters (Y, p). Let $Z(m)$ be a binomial random variable with parameters (m, p). Then*

$$\mathbb{P}\left(X \leq t_1 \mid Y \geq t_2\right) \leq \mathbb{P}\left(Z(t_2) \leq t_1\right) \quad \text{and} \quad \mathbb{P}\left(X \geq t_1 \mid Y \leq t_2\right) \leq \mathbb{P}\left(Z(t_2) \geq t_1\right).$$

Proof. The inequalities follow from the observation that $\mathbb{P}\left(Z(m_1) \geq t\right) \leq \mathbb{P}\left(Z(m_2) \geq t\right)$ for $m_1 < m_2$. $\qquad\square$

2 Reconstruction from 1-neighborhoods

Proof of Lemma 1.4. To determine whether a pair of vertices u, v is connected by an edge, we examine the neighborhoods of u and v, observing graphs H_{u,u_0} and H_{v,v_0} for neighbors $u_0 \sim u$ and $v_0 \sim v$. We declare that u and v are connected if we observe vertices $u_0 \sim u$ and $v_0 \sim v$ in the neighborhoods of u and v respectively such that $H_{u,u_0} \sim H_{v,v_0}$. Clearly this occurs if $(u, v) \in E$. For the other direction, suppose there exist vertices $u_0 \sim u$ and $v_0 \sim v$ such that $H_{u,u_0} \sim H_{v,v_0}$. By the assumption of the lemma, (u, u_0) and (v, v_0) are the same edge, and we conclude that u and v are connected. Continuing this process, we recover the whole graph. $\qquad\square$

Proof of Theorem 1.3. Applying Lemma 1.4, it suffices to show that with high probability, whenever $(u, v), (x, y) \in E$ are distinct edges, then $H_{u,v} \nsim H_{x,y}$.

Let W_{ab} denote the number of shared neighbors of vertices a and b. If $H_{u,v} \sim H_{x,y}$, then $W_{uv} = W_{xy}$. Let $Z = \mathbb{1}\{u \sim x, v \sim x\} + \mathbb{1}\{u \sim y, v \sim y\}$, and suppose $W_{uv} = W_{xy} = \lambda + Z$. In other words, λ is the number of common neighbors of u and v excluding x and y. Let Y be the number of shared neighbors of u, v, x, and y. Suppose also that $Y = \mu$. Let G_1 be the subgraph induced by the $\lambda - \mu$ shared neighbors of u and v (excluding x and y) that are not neighbors of both x and y. Let G_2 be the subgraph induced by all $\lambda + Z$ shared neighbors of x and y. The graphs G_1 and G_2 are disjoint. Observe that if $H_{u,v} \sim H_{x,y}$, then G_1 must be a subgraph of G_2, which we write as $G_1 \subset G_2$.

We upper-bound the probability of the event $\{G_1 \subset G_2\}$ by the First Moment Method. Observe that G_1 is an Erdős–Rényi graph, while we may have already revealed some edges of G_2. We note that up to $2\mu + 1$ edges have already been revealed in G_2; if u, v, x, and y form a clique, then the edges from u and v to the shared neighbors of u, v, x, and y have been revealed, contributing 2μ edges, with the last edge revealed being (u, v) itself. Let $E(H)$ denote the number of edges of a graph H. We have

$$
\begin{aligned}
&\mathbb{P}\left(H_{u,v} \sim H_{x,y} \mid W_{uv} = W_{xy} = \lambda + Z, Y = \mu, E(G_1) = k\right) \\
&\leq \mathbb{P}\left(G_1 \subset G_2 \mid W_{uv} = W_{xy} = \lambda + Z, Y = \mu, E(G_1) = k\right) \\
&\leq \binom{\lambda + 2}{\lambda - \mu}(\lambda - \mu)! p_n^{k - 2\mu - 1} \\
&\leq (\lambda + 2)^{\lambda - \mu} p_n^{k - 2\mu - 1} \\
&\leq \exp\left(\lambda \log(n) + (k - 2\mu - 1)\log(p_n)\right) \\
&= \exp\left(\log(n)\left(\lambda - \alpha(k - 2\mu - 1)\right)\right) \\
&\leq \exp\left(\log(n)\left((2\alpha + 1)\lambda + \alpha - \alpha k\right)\right)
\end{aligned}
\tag{2.1}
$$

We now provide high probability bounds on $W_{uv} - Z \leq W_{uv}$ and $E(G_1)$. Observe that W_{uv} is distributed as a binomial random variable with parameters $(n-2, p_n^2)$. Let $c > 0$, to be determined. By Corollary 1.10,

$$\mathbb{P}\left(W_{uv} \geq n^c(n-2)p_n^2\right) \leq \exp\left(-\Theta\left(n^{1+c-2\alpha}\right)\right).$$

For this to be a high probability bound, we need $1 + c - 2\alpha > 0 \iff c > 2\alpha - 1$. In that case, we obtain

$$\mathbb{P}\left(W_{uv} \geq n^{1+c-2\alpha}\right) = o\left(\frac{1}{n^4}\right). \tag{2.2}$$

Next, observe that $E(G_1)$ is a binomial random variable with parameters $\left(\binom{\lambda-\mu}{2}, p_n\right)$, conditioned on $W_{uv} - Y - Z = \lambda - \mu$. In turn, $W_{uv} - Y - Z$ is a binomial random variable. If the vertices u, v, x, and y are distinct, then $W_{uv} - Y - Z$ is binomial with parameters $\left(n-4, p_n^2(1-p_n^2)\right)$. Otherwise, $W_{uv} - Y - Z$ is binomial random variable with parameters $\left(n-3, p_n^2(1-p_n)\right)$. Fix $0 < \epsilon < \frac{1}{2}$. In the case of distinct vertices, we have

$$\mathbb{P}\left(W_{uv} - Y - Z \leq (1-\epsilon)(n-4)p_n^2(1-p_n^2)\right) \leq \exp\left(-\frac{\epsilon^2}{2}(n-4)p_n^2(1-p_n^2)\right)$$
$$= \exp\left(-\Theta\left(n^{1-2\alpha}\right)\right) = o\left(\frac{1}{n^4}\right).$$

A similar conclusion holds if the vertices are not distinct, so that

$$\mathbb{P}\left(W_{uv} - Y - Z \leq \frac{1}{2}np_n^2\right) = o\left(\frac{1}{n^4}\right), \tag{2.3}$$

for n sufficiently large. Applying (2.3) and Lemma 1.11, we obtain

$$\mathbb{P}\left(E(G_1) \leq (1-\epsilon)p_n\binom{\frac{1}{2}np_n^2}{2}\right) \leq \exp\left(-\frac{\epsilon^2}{2}p_n\binom{\frac{1}{2}np_n^2}{2}\right) + o\left(\frac{1}{n^4}\right)$$
$$= \exp\left(-\Theta\left(n^{2-5\alpha}\right)\right) + o\left(\frac{1}{n^4}\right). \tag{2.4}$$

Using (2.2) and (2.4) in (2.1), we arrive at

$$\mathbb{P}\left(H_{u,v} \sim H_{x,y}\right) \leq \exp\left(\Theta\left(n^{1+c-2\alpha}\log(n)\right) - \Theta\left(n^{2-5\alpha}\log(n)\right)\right) + o\left(\frac{1}{n^4}\right).$$

For this to be a high probability bound, we need $1 + c - 2\alpha < 2 - 5\alpha \iff c < 1 - 3\alpha$.

Summarizing, the conditions are $\max\{0, 2\alpha - 1\} < c < 1 - 3\alpha$. This is a consistent condition, since $\alpha < \frac{1}{3}$. We choose $c = \frac{1}{2}\left(\max\{0, 2\alpha - 1\} + 1 - 3\alpha\right)$. Applying a union bound, we conclude that $H_{u,v} \not\sim H_{x,y}$ whenever (u,v) and (x,y) are distinct edges, with high probability. $\qquad\square$

Remark 2.1. Intuitively, $H_{u,v}$ looks like $G\left(np_n^2, p_n\right)$, and therefore we expect that if it is supercritical, i.e. $np_n^3 \gg 1$, then it will not be isomorphic to an independent copy. Of course, $H_{u,v}$ and $H_{x,y}$ are not independent, which needs to be accounted for in the proof. Similarly, for radius r, the expected number of vertices in the analogue of $H_{u,v}$ is of order $n\left((np_n)^r/n\right)^2 = n^{2r-1}p_n^{2r}$, so reconstruction should be possible when $G\left(n^{2r-1}p_n^{2r}, p_n\right)$ is supercritical, i.e. $p_n \gg n^{\frac{1-2r}{1+2r}}$. On the other hand, [9] showed that for $r = 3$ it suffices that $\lim_{n\to\infty}\frac{np_n}{\log^2(n)} = \infty$, a weaker condition than $\lim_{n\to\infty} n^{\frac{5}{7}}p_n = \infty$. This suggests that the fingerprint argument may not be optimal.

Proof of Theorem 1.5. Let $\mathcal{N}_r(G)$ denote the collection of r-neighborhoods of a graph G. Consider a particular reconstruction algorithm. Given a collection of neighborhoods, the algorithm outputs a graph. We will find a set S of neighborhood collections such that with high probability, $\mathcal{N}_1(G) \in S$. Next, suppose we know that $|E| = m$. Given this information, there are $\binom{\binom{n}{2}}{m}$ possible labeled graphs, which are sampled with equal probability. The algorithm maps each element of S to an isomorphism class, which corresponds to at most $n!$ graphs. Therefore, conditioned on $|E| = m$, the algorithm fails with probability at least

$$\mathbb{P}\left(\mathcal{N}_1(G) \in S \mid |E| = m\right) - \frac{n!|S|}{\binom{\binom{n}{2}}{m}} = \frac{\binom{\binom{n}{2}}{m} - n!|S|}{\binom{\binom{n}{2}}{m}} - \mathbb{P}(\mathcal{N}_1(G) \notin S \mid |E| = m).$$

To see this, observe that the algorithm fails whenever G is such that $\mathcal{N}_1(G) \in S$ but G is not one of the at most $n!|S|$ graphs corresponding to the algorithm's output on $|S|$. Let $S_E \subset \left[\binom{n}{2}\right]$. The overall failure probability of the algorithm is then at least

$$\sum_{m \in S_E} \frac{\binom{\binom{n}{2}}{m} - n!|S|}{\binom{\binom{n}{2}}{m}} \mathbb{P}\left(|E| = m\right) - \mathbb{P}\left(\mathcal{N}_1(G) \notin S \cap |E| \in S_E\right)$$

$$\geq \mathbb{P}\left(|E| \in S_E\right) \min_{m \in S_E} \frac{\binom{\binom{n}{2}}{m} - n!|S|}{\binom{\binom{n}{2}}{m}} - \mathbb{P}\left(\mathcal{N}_1(G) \notin S\right).$$

Therefore, it suffices to show $\mathcal{N}_1(G) \in S$ with high probability, $|E| \in S_E$ with high probability, and $n!|S| = o\left(\binom{\binom{n}{2}}{m}\right)$ for all $m \in S_E$.

We now construct the set S. Fix $0 < \epsilon < 1$, and let $q_n = (1+\epsilon)p_n$. Let $c > 0$, to be determined, and let $t_n = (1 + n^c)p_n$. Let S be the set of possible 1-neighborhood collections where the degree of each central vertex is less than $q_n n$ and each neighborhood has fewer than $\frac{1}{2}q_n^2 t_n n^2$ neighbor edges. We need to show that the collection of 1-neighborhoods of G appears in S with high probability. By a Chernoff bound,

$$\mathbb{P}\left(\deg(v) < q_n n\right) \geq \mathbb{P}\left(\deg(v) < q_n(n-1)\right)$$
$$= 1 - \mathbb{P}\left(\deg(v) \geq (1+\epsilon)p_n(n-1)\right)$$
$$\geq 1 - \exp\left(-\frac{\epsilon^2 p_n(n-1)}{3}\right). \tag{2.5}$$

Let $Z(v)$ be the number of edges among the neighbors of v. Conditioned on $\deg(v)$, $Z(v)$ is distributed as a binomial random variable with parameters $\left(\binom{\deg(v)}{2}, p_n\right)$. We have

$$\mathbb{P}\left(Z(v) \geq \frac{1}{2}q_n^2 t_n n^2\right) \leq \mathbb{P}\left(Z(v) \geq \frac{1}{2}q_n^2 t_n n^2 \mid \deg(v) < q_n n\right) + \mathbb{P}\left(\deg(v) \geq q_n n\right). \tag{2.6}$$

By Lemma 1.11 and a Chernoff bound,

$$\mathbb{P}\left(Z(v) \geq \frac{1}{2}q_n^2 t_n n^2 \mid \deg(v) < q_n n\right) \leq \mathbb{P}\left(Z(v) \geq \frac{1}{2}q_n^2 t_n n^2 \mid \deg(v) = \lfloor q_n n \rfloor\right)$$
$$\leq \mathbb{P}\left(Z(v) \geq \binom{\lfloor q_n n \rfloor}{2} t_n \mid \deg(v) = \lfloor q_n n \rfloor\right)$$
$$\leq \exp\left(-\frac{n^{2c}}{2+n^c}\binom{\lfloor q_n n \rfloor}{2} p_n\right). \tag{2.7}$$

Substituting (2.5) and (2.7) into (2.6), we arrive at

$$\mathbb{P}\left(Z(v) \geq \frac{1}{2}q_n^2 t_n n^2\right) \leq \exp\left(-\frac{n^{2c}}{2+n^c}\binom{\lfloor q_n n \rfloor}{2} p_n\right) + \exp\left(-\frac{\epsilon^2 p_n(n-1)}{3}\right).$$

By a union bound,

$$\mathbb{P}\left(\bigcap_{v \in V}\left\{\deg(v) < q_n n \cap Z(v) < \frac{1}{2}q_n^2 t_n n^2\right\}\right)$$

$$\geq 1 - n\exp\left(-\frac{n^{2c}}{2+n^c}\binom{\lfloor q_n n\rfloor}{2}p_n\right) - n\exp\left(-\frac{\epsilon^2 p_n(n-1)}{3}\right)$$

$$= 1 - n\exp\left(-\Theta(n^{2+c-3\alpha})\right) - n\exp\left(-\Theta(n^{1-\alpha})\right).$$

We therefore need $\alpha < 1$ and $2+c-3\alpha > 0 \iff c > 3\alpha-2$ for this to be a high probability bound. If these constraints are satisfied, then $\mathcal{N}_1(G) \in S$ with high probability.

We now bound $|S|$. For clarity of presentation, we omit floor functions; the resulting off-by-one errors will not affect asymptotics. In each neighborhood, there are $q_n n$ choices for the degree of the central vertex, and $\frac{1}{2}q_n^2 t_n n^2$ choices for the number of neighbor edges. If $t_n = o(1) \iff c < \alpha$, then the number of choices for the set of edges is upper-bounded by $\binom{\binom{q_n n}{2}}{\frac{1}{2}q_n^2 t_n n^2}$. Therefore,

$$|S| \leq \left(q_n n \cdot \frac{1}{2}q_n^2 t_n n^2 \binom{\binom{q_n n}{2}}{\frac{1}{2}q_n^2 t_n n^2}\right)^n \leq \left(\frac{1}{2}q_n^3 t_n n^3 \left(\frac{e(q_n n)^2}{q_n^2 t_n n^2}\right)^{\frac{1}{2}q_n^2 t_n n^2}\right)^n$$

$$= \left(\frac{1}{2}q_n^3 t_n n^3 \left(\frac{e}{t_n}\right)^{\frac{1}{2}q_n^2 t_n n^2}\right)^n. \tag{2.8}$$

We now choose S_E to be $\{m \in \mathbb{N} : \left|m - \binom{n}{2}p_n\right| < \epsilon\binom{n}{2}\}$. By a Chernoff bound,

$$\mathbb{P}\left(|E| \notin S_E\right)) = \mathbb{P}\left(\left||E| - \binom{n}{2}p_n\right| \geq \epsilon\binom{n}{2}\right) \leq 2\exp\left(-\frac{\epsilon^2}{3}\binom{n}{2}p_n\right) = o(1).$$

Therefore, $|E| \in S_E$ with high probability.

We now compute

$$\min_{m \in S_E}\frac{\binom{\binom{n}{2}}{m} - n!|S|}{\binom{\binom{n}{2}}{m}} = 1 - \max_{m \in S_E}\frac{n!|S|}{\binom{\binom{n}{2}}{m}}.$$

We have

$$\min_{m \in S_E}\binom{\binom{n}{2}}{m} = \binom{\binom{n}{2}}{(1-\epsilon)\binom{n}{2}p_n} \geq \left(\frac{1}{(1-\epsilon)p_n}\right)^{(1-\epsilon)\binom{n}{2}p_n}. \tag{2.9}$$

Using (2.8), (2.9), and the bound $n! \leq \exp(n\log(n))$, we have

$$\max_{m \in S_E}\frac{n!|S|}{\binom{\binom{n}{2}}{m}} \leq \frac{\exp(n\log(n))\left(\frac{1}{2}q_n^3 t_n n^3 \left(\frac{e}{t_n}\right)^{\frac{1}{2}q_n^2 t_n n^2}\right)^n}{\left(\frac{1}{(1-\epsilon)p_n}\right)^{(1-\epsilon)\binom{n}{2}p_n}}$$

$$= \exp\left\{n\log(n) + n\left[\log\left(\frac{1}{2}q_n^3 t_n n^3\right) + \frac{1}{2}q_n^2 t_n n^2 \log\left(\frac{e}{t_n}\right)\right]\right.$$

$$\left. - (1-\epsilon)\binom{n}{2}p_n\log\left(\frac{1}{(1-\epsilon)p_n}\right)\right\}$$

$$= \exp\left\{n\log(n) + \Theta(n\log\left(n^{3+c-4\alpha}\right)) + \Theta(n^{3+c-3\alpha}\log(n)) - \Theta(n^{2-\alpha}\log(n))\right\}$$

$$= \exp\left\{\Theta(n\log\left(n\right)) + \Theta(n^{3+c-3\alpha}\log(n)) - \Theta(n^{2-\alpha}\log(n))\right\}.$$

We need $3 + c - 3\alpha < 2 - \alpha \iff c < 2\alpha - 1$ for this bound to go to zero.

Summarizing, we require

$$\max\{0, 3\alpha - 2\} < c < \min\{\alpha, 2\alpha - 1\}.$$

This is consistent for $\frac{1}{2} < \alpha < 1$. Then $\min\{\alpha, 2\alpha - 1\} = 2\alpha - 1$. The proof is completed by choosing

$$c = \frac{1}{2}\left(\max\{0, 3\alpha - 2\} + 2\alpha - 1\right). \qquad \square$$

3 Reconstruction from 2-neighborhoods

The following lemmas will be used to prove Theorem 1.6 for the case $0 < \alpha < \frac{1}{2}$.

Lemma 3.1 ([5])**.** *Let $p_n = c\frac{\sqrt{\log n}}{\sqrt{n}}$. If $c > \sqrt{2}$, then the diameter is 2 with high probability. If $c < \sqrt{2}$, then the diameter is at least 3 with high probability.*

A canonical labeling of a graph involves assigning a unique label to each vertex, in a way that is invariant under isomorphism.

Lemma 3.2 ([3])**.** *Suppose $p_n \le \frac{1}{2}$ and $p_n = \omega\left(\frac{\log^4(n)}{n\log(\log(n))}\right)$, and let $G(V, E) \sim G(n, p_n)$. Then with high probability, any $u, v \in V$ have different degree neighborhoods. Furthermore, one can produce a canonical labeling of G by sorting the vertices in lexicographic order by their degree neighborhoods.*

Remark 3.3. The proof of Theorem 1.2 (Theorem 4.5 in [9]) shows that the statement of Lemma 3.2 is true for $p_n = \omega\left(\frac{\log^2(n)}{n}\right)$, which strengthens the result of [3].

Proof of Theorem 1.6 for $0 < \alpha < \frac{1}{2}$. By Lemma 3.1, the diameter is equal to 2 with high probability, so we see the entire graph from any 2-neighborhood. To label the graph, we produce the canonical labeling of any 2-neighborhood, which is possible (and efficient) with high probability by Lemma 3.2. We then sort the vertices lexicographically by their degree neighborhoods in order to determine the canonical label of each vertex. This procedure exactly reconstructs G. $\qquad \square$

Proof of Theorem 1.6 for $\frac{1}{2} < \alpha < \frac{3}{5}$. The proof is similar to the proof of Theorem 1.3. This time, for $u, v \in V$ where $(u, v) \in E$, we define $L_{u,v}$ as the subgraph induced by those vertices which are at distance exactly 2 from both u and v. Lemma 1.4 applies to the $L_{u,v}$ graphs as well. It therefore suffices to show that with high probability, whenever (u, v) and (x, y) are distinct edges, then $L_{u,v} \not\sim L_{x,y}$.

Consider two distinct edges (u, v) and (x, y). Our goal is to show that $\mathbb{P}\left(L_{u,v} \sim L_{x,y}\right) = o\left(n^{-4}\right)$. Similarly to the proof of Theorem 1.3, we will identify two graphs G_1 and G_2 such that $\mathbb{P}\left(L_{u,v} \sim L_{x,y}\right) \le \mathbb{P}\left(G_1 \subset G_2\right)$. Note that by revealing which vertices are at distance exactly 2 from u and v, we have not revealed any of the edges among those distance-2 neighbors.

Let W_{ab} be the number of shared distance-2 neighbors of a and b. If $L_{u,v} \sim L_{x,y}$, then $W_{uv} = W_{xy}$. Let $d(a, b)$ be the graph distance between vertices a and b, and let

$$Z = \mathbb{1}\{d(u, x) = d(v, x) = 2\} + \mathbb{1}\{d(u, y) = d(v, y) = 2\}.$$

Suppose $W_{uv} = W_{xy} = \lambda + Z$, so that λ counts all shared distance-2 neighbors of u and v except for x and y. Let $Y = |\mathcal{N}_2(u) \cap \mathcal{N}_2(v) \cap \mathcal{N}_2(x) \cap \mathcal{N}_2(y)|$ be the number of shared distance-2 neighbors of u, v, x, and y. Suppose also that $Y = \mu$. Let G_1 be the subgraph induced by the $\lambda - \mu$ shared distance-2 neighbors of u and v that are not x, y, or a shared distance-2 neighbor of x and y. Let G_2 be the subgraph induced by all distance-2

neighbors of x and y. Finally, suppose $|\mathcal{N}_1(u)| + |\mathcal{N}_1(v)| = d$. Similarly to the proof of Theorem 1.3, we claim

$$\mathbb{P}\left(L_{u,v} \sim L_{x,y} \mid W_{uv} = W_{xy} = \lambda + Z, Y = \mu, E(G_1) = k, |\mathcal{N}_1(u)| + |\mathcal{N}_1(v)| = d\right)$$

$$\leq \binom{\lambda+2}{\lambda-\mu}(\lambda-\mu)! p_n^{k-d} \leq \exp\left(\log(n)\left(\lambda - \alpha(k-d)\right)\right). \tag{3.1}$$

To see the claim, consider the following way of determining which vertices belong in G_1 and G_2. First, determine which vertices are at distance-1 and distance-2 from x, and do the same for y. From this information, we determine which vertices are in G_2. We have not revealed any of the edges in G_2, unless u and v are both shared neighbors of x and y. In that case, we have revealed the edge (u, v). Next, we reveal the distance-1 neighbors of u and v. This step reveals at most $|\mathcal{N}_1(u)| + |\mathcal{N}_1(v)|$ edges in G_2. Finally, to determine which vertices are in G_1, we consider all vertices that are not in G_2. From among those vertices, we place those that are shared distance-2 neighbors of u and v into G_1. This last step does not reveal any edges in G_2. Therefore, G_2 has at most $|\mathcal{N}_1(u)| + |\mathcal{N}_1(v)|$ edges revealed in this process, so the probability that any particular collection of k edges exists is at most $p_n^{k-(|\mathcal{N}_1(u)|+|\mathcal{N}_1(v)|)}$.

We will now find high probability bounds on λ, $E(G_1)$, and $|\mathcal{N}_1(u)| + |\mathcal{N}_1(v)|$. Since $W_{uv} - 2 \leq \lambda \leq W_{uv}$, we will bound W_{uv}. Note that conditioned on $|\mathcal{N}_1(u)|$, $|\mathcal{N}_1(v)|$, and $|\mathcal{N}_1(u) \cap \mathcal{N}_1(v)|$, the number of shared distance-2 neighbors is distributed as a binomial random variable with parameters $(n - |\mathcal{N}_1(u) \cup \mathcal{N}_1(v)|, q_n)$, where

$$q_n = 1 - (1-p_n)^{|\mathcal{N}_1(u) \cap \mathcal{N}_1(v)|}$$
$$+ (1-p_n)^{|\mathcal{N}_1(u) \cap \mathcal{N}_1(v)|} \left(1 - (1-p_n)^{|\mathcal{N}_1(u) \setminus \mathcal{N}_1(v)|-1}\right) \left(1 - (1-p_n)^{|\mathcal{N}_1(v) \setminus \mathcal{N}_1(u)|-1}\right).$$

To see this, observe that a candidate shared distance-2 neighbor must either be connected to some vertex in $\mathcal{N}_1(u) \cap \mathcal{N}_1(v)$, or failing this, must be connected to some vertex in $\mathcal{N}_1(u) \setminus \{\mathcal{N}_1(v) \cup \{v\}\}$ and some vertex in $\mathcal{N}_1(v) \setminus \{\mathcal{N}_1(u) \cup \{u\}\}$. To derive bounds on W_{uv}, we use bounds on q_n:

$$q_n \leq 1 - (1-p_n)^{|\mathcal{N}_1(u) \cap \mathcal{N}_1(v)|} + \left(1 - (1-p_n)^{|\mathcal{N}_1(u)|}\right) \left(1 - (1-p_n)^{|\mathcal{N}_1(v)|}\right)$$

$$q_n \geq \left(1 - (1-p_n)^{|\mathcal{N}_1(u) \setminus \mathcal{N}_1(v)|-1}\right) \left(1 - (1-p_n)^{|\mathcal{N}_1(v) \setminus \mathcal{N}_1(u)|-1}\right).$$

Fix $0 < \epsilon < 1$. We have that $|\mathcal{N}_1(u)|, |\mathcal{N}_1(v)| \leq (1+\epsilon)n^{1-\alpha}$ with probability $1 - o(n^{-4})$. Noting that $\mathbb{E}\left[|\mathcal{N}_1(u) \cap \mathcal{N}_1(v)|\right] = (n-2)p_n^2 = \Theta\left(n^{1-2\alpha}\right)$ and applying Corollary 1.10, we have $|\mathcal{N}_1(u) \cap \mathcal{N}_1(v)| \leq n^{1-2\alpha+c}$ with probability $1 - o(n^{-4})$ for c satisfying $1 - 2\alpha + c > 0 \iff c > 2\alpha - 1$. Therefore, using the upper bound on q_n, we have that with probability $1 - o(n^{-4})$,

$$W_{uv} \leq (1+\epsilon)n \left[1 - (1-p_n)^{n^{1-2\alpha+c}} + \left(1 - (1-p_n)^{(1+\epsilon)n^{1-\alpha}}\right)^2\right]$$

$$\leq (1+\epsilon)n \left[1 - \exp\left(-2n^{1-3\alpha+c}\right) + \left(1 - \exp\left(-2(1+\epsilon)n^{1-2\alpha}\right)\right)^2\right]$$

$$\leq (1+\epsilon)n \left[2n^{1-3\alpha+c} + \left(2(1+\epsilon)n^{1-2\alpha}\right)^2\right]$$

$$= \Theta\left(n^{2-3\alpha+c}\right) + \Theta\left(n^{3-4\alpha}\right),$$

where we have used the inequalities $1 - x \geq e^{-2x}$ for $x \leq \frac{1}{2}$, and $1 - e^{-x} \leq x$. Taking $c > 2\alpha - 1$ sufficiently small, the second term dominates and we arrive at $W_{uv} = O\left(n^{3-4\alpha}\right)$ with probability $1 - o(n^{-4})$.

Next, observe that $E(G_1)$ is a binomial random variable with parameters $\left(\binom{\lambda-\mu}{2}, p_n\right)$, conditioned on $W_{uv} - Y - Z = \lambda - \mu$. We similarly lower bound W_{uv}, using the lower bound on q_n. Observe that $|\mathcal{N}_1(u) \setminus N_1(v)|, |\mathcal{N}_1(v) \setminus N_1(u)| \geq (1 - \epsilon)n^{1-\alpha}$ with probability $1 - o(n^{-4})$. Therefore, with probability $1 - o(n^{-4})$,

$$W_{uv} \geq (1 - \epsilon)n \left(1 - (1 - p_n)^{(1-\epsilon)n^{1-\alpha}-1}\right)^2 = \Omega\left(n^{3-4\alpha}\right).$$

By similar reasoning, we have $\mu = O\left(n^{5-8\alpha}\right)$ with probability $1 - o(n^{-4})$ if the four vertices are distinct, and otherwise $\mu = O\left(n^{4-6\alpha}\right)$ with probability $1 - o(n^{-4})$. Since $5 - 8\alpha, 4 - 6\alpha < 3 - 4\alpha$, we have $\binom{\lambda-\mu}{2} = \Theta\left(n^{6-8\alpha}\right)$, so that $E(G_1) = \Theta\left(n^{6-9\alpha}\right)$, with probability $1 - o(n^{-4})$.

Finally, examining the bound (3.1), we first compare $E(G_1)$ to $|\mathcal{N}_1(u)| + |\mathcal{N}_1(v)|$. We see that the order of $E(G_1)$ dominates when

$$6 - 9\alpha > 1 - \alpha \iff \alpha < \frac{5}{8}.$$

Since $\frac{5}{8} > \frac{3}{5}$, the edge count indeed dominates. Next, we compare λ to $E(G_1)$. We see that the order of $E(G_1)$ dominates when

$$6 - 9\alpha > 3 - 4\alpha \iff \alpha < \frac{3}{5}.$$

We have shown that for a given pair of distinct edges (u, v) and (x, y), we have $\mathbb{P}(L_{u,v} \sim L_{x,y}) = o(n^{-4})$. Taking a union bound completes the proof. $\qquad\square$

Proof of Theorem 1.7. The proof is similar to the proof of Theorem 1.5. Again we set $S_E = \{m \in \mathbb{N} : |m - \binom{n}{2}p_n| < \epsilon\binom{n}{2}\}$. It remains to find a set S of collections of 2-neighborhoods such that $\mathcal{N}_2(G) \in S$ with high probability, and show that

$$\max_{m \in S} \frac{n!|S|}{\binom{\binom{n}{2}}{m}} = o(1).$$

We now construct the set S. Fix $0 < \epsilon < 1$ to be determined, and let $q_n = (1 + \epsilon)p_n$. For some $c > 0$ to be determined, let $t_n = (1 + n^c)p_n$. Let $X(v)$ and $Y(v)$ respectively be the number of vertices and edges in the graph $\mathcal{N}_2(v)$. Let S be the set of possible 2-neighborhood collections where for each $v \in V$, the following three conditions hold: (1) $\deg(v) < q_n n$; (2) $\deg(w) < q_n n$ for each $w \sim v$; (3) $Y(v) < \frac{1}{2}q_n^4 t_n n^4 + X(v) - 1$. We need to show that the collection of 2-neighborhoods of G appears in S with high probability. As was shown in the proof of Theorem 1.5

$$\mathbb{P}\left(\cap_{v \in V} \{\deg(v) < q_n n\}\right) = 1 - o(1).$$

Therefore, the first two properties hold for every neighborhood with high probability. Let E_v be the event that properties (1) and (2) hold in the 2-neighborhood around v.

Next, we bound the number of edges. Our goal is to lower-bound

$$\mathbb{P}\left(Y(v) \geq \frac{1}{2}q_n^4 t_n n^4 + X(v) - 1 \mid E_v\right).$$

We reveal the neighborhood $\mathcal{N}_2(v)$ in a particular way. We start from v, and reveal its neighbors (and therefore the edges $\{(v, w) : w \sim v\}$). Then we choose an arbitrary order for the neighbors. Starting from the first neighbor w_1, we reveal all of its additional neighbors apart from v. Next, we come to the second neighbor w_2, and reveal all of its neighbors apart from v and w_1. Continuing through the neighbors of v, we reveal all the

vertices of the neighborhood using $X(v) - 1$ edges (since each vertex except v is revealed by exactly one other vertex). The number of additional edges in $\mathcal{N}_2(v)$ is dominated by a binomial random variable with parameters $\left(\binom{\lfloor (q_n n)^2 \rfloor}{2}, p_n\right)$. Let $Z \sim \text{Bin}\left(\binom{\lfloor (q_n n)^2 \rfloor}{2}, p_n\right)$. We therefore have

$$
\begin{aligned}
\mathbb{P}\left(Y(v) \geq \frac{1}{2} q_n^4 t_n n^4 + X(v) - 1 \mid E_v\right) &\leq \mathbb{P}\left(Z \geq \frac{1}{2} q_n^4 t_n n^4\right) \\
&\leq \mathbb{P}\left(Z \geq \binom{\lfloor (q_n n)^2 \rfloor}{2} t_n\right) \\
&\leq \exp\left(-\frac{n^{2c}}{2 + n^c}\binom{\lfloor (q_n n)^2 \rfloor}{2} p_n\right) \\
&= \exp\left(-\Theta\left(n^{4+c-5\alpha}\right)\right).
\end{aligned}
$$

For this to be a high probability bound, we need $4 + c - 5\alpha > 0 \iff c > 5\alpha - 4$. We then have

$$
\mathbb{P}\left(\bigcap_{v \in V}\left\{\deg(v) < q_n n, Y(v) < \frac{1}{2} q_n^4 t_n n^4 + X(v) - 1\right\}\right) = 1 - o(1),
$$

We conclude that $\mathcal{N}_2(G) \in S$ with high probability.

We now bound $|S|$. Consider one collection of 2-neighborhoods in the set S, and a particular vertex $v \in V$. There are $q_n n$ choices for the number of neighbors of v. Next, there are fewer than $q_n n(q_n n - 1) < (q_n n)^2$ vertices which are at distance 2 from v. We reveal the distance-2 neighbors sequentially, as described above. Each distance-2 neighbor is assigned to one of the distance-1 neighbors, according to which distance-1 neighbor revealed it. Using a stars-and-bars argument, the number of ways this assignment can happen is upper-bounded by

$$
\binom{q_n n(q_n n - 1) + q_n n - 1}{q_n n - 1} = \binom{(q_n n)^2 - 1}{q_n n - 1} < \binom{(q_n n)^2}{q_n n}.
$$

So far, $X(v) - 1$ edges have been revealed. There are fewer than $q_n^4 t_n n^4$ remaining edges, and at most $\binom{\binom{(q_n n)^2 + 1}{2}}{q_n^4 t_n n^4}$ choices for their locations, as long as $q_n^4 t_n n^4 < \frac{1}{2}\binom{(q_n n)^2 + 1}{2}$, which is satisfied for $c < \alpha$. We therefore have

$$
\begin{aligned}
|S| &\leq \left[q_n n \cdot (q_n n)^2 \cdot \binom{(q_n n)^2}{q_n n} \cdot q_n^4 t_n n^4 \cdot \binom{\binom{(q_n n)^2 + 1}{2}}{q_n^4 t_n n^4}\right]^n \\
&\leq \left[q_n^7 t_n n^7 \left(e q_n n\right)^{q_n n}\left(\frac{e\binom{(q_n n)^2 + 1}{2}}{q_n^4 t_n n^4}\right)^{q_n^4 t_n n^4}\right]^n \\
&= \exp\left\{n\left(\log\left(q_n^7 t_n n^7\right) + q_n n \log\left(e q_n n\right) + q_n^4 t_n n^4 \log\left(\frac{e\binom{(q_n n)^2 + 1}{2}}{q_n^4 t_n n^4}\right)\right)\right\} \\
&= \exp\Big\{n \log\left(q_n^7 t_n n^7\right) + (1 + \epsilon)n^{2-\alpha}\log(e(1 + \epsilon)) + (1 - \alpha)(1 + \epsilon)n^{2-\alpha}\log(n) \\
&\qquad + q_n^4 t_n n^5 \log\left(\frac{e\binom{(q_n n)^2 + 1}{2}}{q_n^4 t_n n^4}\right)\Big\} \\
&= \exp\{(7 + c - 8\alpha)\Theta\left(n\log(n)\right) + \Theta\left(n^{2-\alpha}\right) + (1 - \alpha)(1 + \epsilon)n^{2-\alpha}\log(n) \\
&\qquad + \Theta\left(n^{5+c-5\alpha}\log(n)\right)\} \\
&= \exp\left\{\Theta\left(n^{2-\alpha}\right) + (1 - \alpha)(1 + \epsilon)n^{2-\alpha}\log(n) + \Theta\left(n^{5+c-5\alpha}\log(n)\right)\right\}.
\end{aligned}
$$

As noted in the proof of Theorem 1.5,

$$\min_{m \in S_E} \binom{\binom{n}{2}}{m} \geq \left(\frac{1}{(1-\epsilon)p_n} \right)^{(1-\epsilon)\binom{n}{2}p_n} \geq \exp\left(\frac{1}{2}(1-\epsilon)(n-1)^2 n^{-\alpha} \log\left(\frac{n^\alpha}{(1-\epsilon)} \right) \right).$$

Using this bound and the bound on $|S|$, we have

$$\max_{m \in S} \frac{n!|S|}{\binom{\binom{n}{2}}{m}} \leq \exp\Big\{ \Theta\left(n^{2-\alpha}\right) + (1-\alpha)(1+\epsilon)n^{2-\alpha}\log(n) + \Theta\left(n^{5+c-5\alpha}\log(n)\right)$$

$$- \frac{1}{2}(1-\epsilon)(n-1)^2 n^{-\alpha} \log\left(\frac{n^\alpha}{(1-\epsilon)} \right) \Big\}$$

$$= \exp\Big\{ \Theta\left(n^{2-\alpha}\right) + (1-\alpha)(1+\epsilon)n^{2-\alpha}\log(n) + \Theta\left(n^{5+c-5\alpha}\log(n)\right)$$

$$- \frac{\alpha}{2}(1-\epsilon)(n-1)^2 n^{-\alpha} \log(n) \Big\}.$$

Therefore, it suffices to have $5 + c - 5\alpha < 2 - \alpha \iff c < 4\alpha - 3$ as well as

$$(1-\alpha)(1+\epsilon) < \frac{\alpha}{2}(1-\epsilon) \iff \alpha > \frac{2+2\epsilon}{3+\epsilon}.$$

Taking $\epsilon = \frac{1}{6}$ satisfies the above. We then obtain

$$\max_{m \in S} \frac{n!|S|}{\binom{\binom{n}{2}}{m}} \leq \exp\left(\Theta\left(n^{2-\alpha}\right) - \Theta\left(n^{2-\alpha}\log(n)\right) \right) = o(1).$$

Summarizing, we require $\max\{0, 5\alpha - 4\} < c < \min\{\alpha, 4\alpha - 3\}$. This is a consistent condition for $\frac{3}{4} < \alpha < 1$. We then choose $c = \frac{1}{2}\left(\max\{0, 5\alpha - 4\} + \min\{\alpha, 4\alpha - 3\} \right)$. □

Proof of Proposition 1.8. Let $\beta = \frac{1}{2}\left(\alpha + \frac{2}{3}\right)$, so that $\frac{2}{3} < \beta < \alpha$. We will show that with probability $1 - o(1)$, there exists some $k_n \leq n^{1-\beta}$ such that there are at least $\frac{1}{2}n^\beta$ stars of degree k_n in G_n, implying that the 1-neigborhoods are not unique.

Consider an arbitrary vertex v. By the Markov inequality,

$$\mathbb{P}\left(\deg(v) > n^{1-\beta}\right) \leq \frac{n^{1-\alpha}}{n^{1-\beta}} = n^{\beta-\alpha} = o(1). \tag{3.2}$$

Let S_v be the indicator that $\mathcal{N}_1(v)$ is a star of degree at most $n^{1-\beta}$. Let $Z \sim \text{Bin}\left(\binom{n^{1-\beta}}{2}, p_n\right)$. Observe that $\mathbb{E}[Z] < \frac{1}{2}n^{2-2\beta-\alpha} = o(1)$. Applying (3.2), we have

$$\mathbb{P}\left(S_v = 1\right) = \mathbb{P}\left(S_v = 1 \mid \deg(v) \leq n^{1-\beta}\right)(1 - o(1)) + o(1)$$

$$\geq \mathbb{P}\left(S_v = 1 \mid \deg(v) = n^{1-\beta}\right)(1 - o(1))$$

$$= 1 - \mathbb{P}\left(Z \geq 1\right) - o(1)$$

$$\geq 1 - \mathbb{E}[Z] - o(1) = 1 - o(1).$$

By the Markov inequality, $\mathbb{P}\left(\sum_{v \in V} S_v < \frac{3n}{4}\right) = \mathbb{P}\left(\sum_{v \in V}(1 - S_v) > \frac{n}{4}\right) = o(1)$. We conclude that with probability $1 - o(1)$, there are at least $\frac{3n}{4}$ vertices v such that $\mathcal{N}_1(v)$ is a star of degree at most $n^{1-\beta}$. Then by the Pigeonhole Principle, there is at least one $k_n \leq n^{1-\beta}$ such that the number of stars of degree k_n is at least $\frac{3n}{4(n^{1-\beta}+1)} \geq \frac{n}{2n^{1-\beta}}$. □

4 Finding the neighborhood centers

Previously we assumed that the central vertex v of each neighborhood $\mathcal{N}_r(v)$ was labeled. In this section, we show how to determine the center if it is not labeled.

4.1 Finding centers of 1-neighborhoods

The following lemma shows that in order to identify the center of each neighborhood, we should simply take the vertex that is connected to all others.

Lemma 4.1. *Suppose $p_n = n^{-\alpha}$ for $0 < \alpha < 1$. Then with high probability, for all $v \in V$, the vertex v is the only other vertex connected to every other vertex in $\mathcal{N}_1(v)$.*

Proof. Consider $v \in V$ and the neighborhood $\mathcal{N}_1(v)$. The degree of v is distributed according to a binomial distribution with parameters $(n-1, p_n)$. Furthermore, for $0 < \epsilon < 1$, it holds that $\deg(v) \geq (1-\epsilon)(n-1)p_n$ with high probability. Conditioned on $\deg(v) = m \geq (1-\epsilon)p_n(n-1)$, the probability that a given neighbor $w \sim v$ is connected to all other neighbors of v is then p_n^{m-1}. The claim follows by a union bound. $\square$

4.2 Finding centers of 2-neighborhoods

To identify the center of a neighborhood, we first prune the neighborhood by removing all vertices with degree less than $\frac{1}{2}n^{1-\alpha}$. We then return the vertex with the highest degree in the subgraph induced by the remaining vertices (which is also the only vertex connected to all the others). Recall that if $\alpha < \frac{1}{2}$, then the graph G has diameter 2 with high probability.

Lemma 4.2. *Suppose $p_n = n^{-\alpha}$ for $\frac{1}{2} < \alpha < 1$. Then with high probability, for all $v \in V$, the above approach recovers the center v in $\mathcal{N}_2(v)$.*

Proof. Consider the neighborhood $\mathcal{N}_2(v)$. Observe that with high probability, the total number of vertices in the 2-neighborhood is of order at most $(np_n)^2 = n^{2-2\alpha}$. Each neighbor u of v has $(n-1)p_n = \Theta\left(n^{1-\alpha}\right)$ expected neighbors, with exponential concentration. On the other hand, each vertex w which is at a distance of 2 from v has $1 + (|V(\mathcal{N}_2(v))| - 3)p_n = O\left(n^{2-3\alpha}\right)$ expected neighbors in $\mathcal{N}_2(v)$. Comparing exponents, we have $2 - 3\alpha < 1 - \alpha$ for $\alpha > \frac{1}{2}$. Therefore, by removing all vertices with fewer than $\frac{1}{2}n^{1-\alpha}$ neighbors, we remove all vertices which are at distance 2 from v, while not removing any neighbors of v, with high probability. By a union bound, this process leaves us with $\mathcal{N}_1(v)$ for each $v \in V$ with high probability. By Lemma 4.1, we can identify the center of each pruned neighborhood. $\square$

5 Conclusion and open problems

In this paper, we have studied the problem of shotgun assembly of Erdős–Rényi graphs from 1- and 2-neighborhoods. We have established regimes for reconstructability and non-reconstructability in terms of the edge probability $p_n = n^{-\alpha}$. Our work leaves some open problems:

1. Is reconstruction from 2-neighborhoods possible for $\alpha = \frac{1}{2}$ or $\frac{3}{5} \leq \alpha \leq \frac{3}{4}$?
2. Our results show that there is an efficient algorithm for reconstruction from 2-neighborhoods in the regime $0 < \alpha < \frac{1}{2}$. Can we find efficient algorithms for reconstruction in other cases?

References

[1] Paul Balister, Béla Bollobás, and Bhargav Narayanan, *Reconstructing random jigsaws*, arXiv preprint arXiv:1707.04730 (2017). MR3931403

[2] Charles Bordenave, Uriel Feige, and Elchanan Mossel, *Shotgun assembly of random jigsaw puzzles*, Random Structures & Algorithms **56** (2020), no. 4, 998–1015. MR4101351

[3] Tomek Czajka and Gopal Pandurangan, *Improved random graph isomorphism*, Journal of Discrete Algorithms **6** (2008), no. 1, 85–92. MR2398207

[4] Han Huang and Konstantin Tikhomirov, *Shotgun assembly of unlabeled erdos-renyi graphs*, arXiv preprint arXiv:2108.09636 (2021).

[5] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński, *Random graphs*, John Wiley & Sons, 2000. MR1782847

[6] Anders Martinsson, *Shotgun edge assembly of random jigsaw puzzles*, arXiv preprint arXiv:1605.07151 (2016).

[7] Anders Martinsson, *A linear threshold for uniqueness of solutions to random jigsaw puzzles*, Combinatorics, Probability and Computing **28** (2019), no. 2, 287–302. MR3922781

[8] Michael Mitzenmacher and Eli Upfal, *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*, Cambridge University Press, 2017. MR3674428

[9] Elchanan Mossel and Nathan Ross, *Shotgun assembly of labeled graphs*, IEEE Transactions on Network Science and Engineering **6** (2019), no. 2, 145–157. MR3969756

[10] Elchanan Mossel and Nike Sun, *Shotgun assembly of random regular graphs*, arXiv preprint arXiv:1512.08473 (2015).

[11] Rajko Nenadov, Pascal Pfister, and Angelika Steger, *Unique reconstruction threshold for random jigsaw puzzles*, Chicago Journal of Theoretical Computer Science **16** (2017), no. 2. MR3672682

[12] Michał Przykucki, Alexander Roberts, and Alex Scott, *Shotgun reconstruction in the hypercube*, arXiv preprint arXiv:1907.07250 (2019). MR4340476

[13] Lyudmila Yartseva, Jefferson Elbert Simoes, and Matthias Grossglauser, *Assembling a network out of ambiguous patches*, Fifty-fourth Annual Allerton Conference (2016).