SHOTGUN ASSEMBLY OF RANDOM GEOMETRIC GRAPHS

KARTICK ADHIKARI AND SUKRIT CHAKRABORTY

Abstract. In a recent work, Huang and Tikhomirov considered the shotgun assembly for Erdős-Rényi graphs $G(n, p_n)$ with $p_n = n^{-\alpha}$, and showed that the graph is reconstructable if $0 < \alpha < \frac{1}{2}$ and not reconstructable if $\frac{1}{2} < \alpha < 1$ from its 1-neighbourhoods. In this article, we consider random geometric graphs $G(n, r)$, where $r^2 = n^{-\alpha}$ and $0 < \alpha < 1$, on flat torus. Interestingly, unlike the results for the Erdős-Rényi random graphs, we show that the random geometric graph is always reconstructable from its 1-neighbourhoods.

1. Introduction

One of the exciting fields of study in recent days is the shotgun assembly problem for random graphs, which was first introduced by Mossel and Ross [4]. The shotgun assembly of a graph means reconstructing the graph from a collection of vertex neighbourhoods. The problem of graph assembly essentially tells us whether the local structure contains all the information about its global formation. The motivation comes from DNA shotgun assembly (determining a DNA sequence from multiple short nucleobase chains), reconstruction of neural networks (reconstructing a big neural network from subnetworks), and the random jigsaw puzzle problem. See [4] and references therein.

The recent development of random jigsaw puzzles can be found in [1]. The graph shotgun assembly for various models was studied extensively in the last decade. The random regular graphs and the labelled graphs were considered in [5] and [4], respectively. The problem for Erdős–Rényi random graphs is well studied in [2–4]. In this article, we consider the shotgun assembly problem for random geometric graphs.

The problem of graph shotgun assembly is as follows. Let $G = (V, E)$ denote a graph with vertex set $V$ and edge set $E$. We say two vertices $x, y \in G$ are at a distance $\ell$ from each other if we need to visit at least $\ell$ edges to go from $x$ to $y$. For $v \in V$ and $t \in \mathbb{N}$, the $t$-neighbourhood of $v$, denoted by $N_t(v)$, is the subgraph induced by all the vertices at a distance (graph distance) at most $t$ from $v$ in the graph $G$.

Two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ are said to be graph isomorphic if there exists a bijective function $f : V_1 \rightarrow V_2$ such that whenever $(u, v) \in E_1$ we must have $(f(u), f(v)) \in E_2$. In that case, we write $G_1 \simeq G_2$. The aim is to reconstruct (up to isomorphism) $G$ from its $t$-neighbourhoods. In other words, the goal is to find a graph $\tilde{G}$, which is isomorphic to $G$, from a given collection of $t$-neighbourhoods of $G$.

Date: February 8, 2022.

Key words and phrases. Shotgun assembly, Random geometric graph.

2010 Mathematics Subject Classification: 05C80.
We write $G_1 \simeq_t G_2$ if two graphs $G_1(V,E_1)$ and $G_2(V,E_2)$ have the same $t$-neighbourhoods, that is, the $t$-neighbourhood around $v$ in $G_1$ is isomorphic to the $t$-neighbourhood around $v$ in $G_2$ for all $v \in V$. We say that the graph $G$ is reconstructable (up to isomorphism) from its $t$-neighbourhoods if $G \simeq_t \tilde{G}$ implies $\tilde{G} \simeq \tilde{G}$, for all graphs $\tilde{G} = (V,E_{\tilde{G}})$. A graph $G$ is said to be exactly reconstructable if it is possible to recover $G$ instead of recovering some graph $\tilde{G} \simeq G$.

Observe that if $t$ is more than the graph radius then graph is reconstructable. It is an appropriate question to ask what is the radius $t$, required for assembly with high probability. The reconstruction of Erdős-Rényi graphs from its 3-neighbourhoods and 2-neighbourhoods are considered in [4, Theorem 4.5] and [2, Theorem 4] respectively. The most interesting case is $t = 1$, that is, whether the graph is reconstructable from its 1-neighbourhoods. It is clear that if we can reconstruct the graph from its 1-neighbourhoods then it can be reconstructed from any neighbourhoods. It was shown in [3] that Erdős–Rényi graphs $G(n,p_n)$ with $p_n = n^{-\alpha}$ is reconstructable if $0 < \alpha < 1/2$ and not reconstructable if $1/2 < \alpha < 1$ from its 1-neighbourhoods.

We consider labelled random geometric graphs $G(n,r) = ([n],E)$ with vertex set $[n] := \{1,2,\ldots,n\}$, the vertices are placed uniformly on the unit square $[0,1]^2$. There is an edge between two distinct points $x, y \in [0,1]^2$ if and only if $d_{tor}(x,y) \leq r$, where $d_{tor}(x,y) := \inf\{\|x - y + a\|_2 : a \in \mathbb{Z}\}$. We assume that $r^2 = n^{-\alpha}$ with $0 < \alpha < 1$. We show that with high probability $G(n,r)$ is exactly reconstructable from its 1-neighbourhoods for all $\alpha \in (0,1)$. See Theorem 2. Consequently it follows that the random geometric graph is reconstructable from its $t$-neighbourhoods for any $t \in \mathbb{N}$. See Corollary 3.

The rest of the article is organized as follows: In Section 2, the main result of the article is stated and the same is proved in Section 3. Proofs of a couple of basic facts that are used in the proof of the main result are given in Section 4.

2. MAIN RESULT

In this section we state our main result, and give the key idea of the proof. Let us define the high probability events.

**Definition 1.** A sequence of events $A_n$ occurs with high probability if

$$P(A_n^c) = o\left(\frac{1}{n^s}\right),$$

for some $s > 0$. We write $a_n = o(b_n)$ for two sequence of numbers $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$ if $|a_n/b_n| \to 0$ as $n \to \infty$.

The main result of this article is given below.

**Theorem 2.** The random geometric graph $G(n,r)$ where $r^2 = n^{-\alpha}$ for $0 < \alpha < 1$ is exactly reconstructable from its 1-neighbourhoods with high probability.

Observe that the information of the $t$-neighbourhoods, where $t > 1$, of vertices of the graph gives the information of its 1-neighbourhoods. Thus we have the following corollary.

**Corollary 3.** The random geometric graph $G(n,r)$ where $r^2 = n^{-\alpha}$ for $0 < \alpha < 1$ is exactly reconstructable from its $t$-neighbourhoods with high probability for any $t \in \mathbb{N}$. 
We use a fingerprint idea from [6] to prove Theorem 2. This method was used to show that the Erdős-Rényi graph \( G(n, n - \alpha) \) is exactly reconstructable for \( \alpha \in (0, 1/3) \). See [2, Theorem 3]. For vertices \( u, v \) such that \((u, v)\) is an edge in the graph \( G \), let \( H_{u,v} \) denote the subgraph induced by the common neighbours of \( u \) and \( v \). The graph \( H_{u,v} \) is called a fingerprint for the edge \((u, v)\). It is obvious that \( H_{u,v} \cong H_{v,u} \). We recall the fingerprint lemma [2, Lemma 2].

Lemma 1 (Fingerprint Lemma). Suppose that \((x,y)\) and \((u,v)\) are the two same edges on a random graph \( G \) if and only if the neighbourhoods \( H_{x,y} \) and \( H_{u,v} \) are graph isomorphic. Then the graph \( G \) can be exactly reconstructed from the collection of its \( 1 \)-neighbourhoods.

Lemma 1 implies that, to prove Theorem 2, it is enough to show that the subgraphs \( H_{u,v} \neq H_{x,y} \) for two distinct edges \((u,v)\) and \((x,y)\) in \( G(n, r) \) with high probability. Observe that if the number vertices that are connected with both \( u,v \) is different from that of \( x,y \) then \( H_{u,v} \neq H_{x,y} \). Our main aim is to show that the number of common vertices are distinct for distinct edges with high probability. In fact prove a stronger result showing that the difference goes to infinity, see (6). It is worth to mention that the claim does not hold for Erdős-Rényi graphs.

3. Proof of Theorem 2

This section is dedicated to prove Theorem 2. The following lemmas will be used in the proof. Throughout we assume \( r^2 = n^{-\alpha} \) for \( 0 < \alpha < 1 \).

Lemma 2. Let \( W_{x,y} \) be the number of shared neighbours of vertices \( x \) and \( y \) in \( G(n, r) \). Let \( d = (1 - \alpha)/3 \). Then with high probability
\[
|W_{x,y} - E(W_{x,y})| \leq \frac{E(W_{x,y})}{n^d}.
\]

Lemma 3. Let \( R_{x,y} = R_x \cap R_y \), where \( R_x \) denotes the ball of radius \( r \) with center at the vertex \( x \). It’s area \( |R_{x,y}| \) is given by
\[
|R_{x,y}| = r^2 (\theta - \sin \theta) \text{ where } \theta = 2 \cos^{-1} \left( \frac{|x - y|}{2r} \right).
\]
In particular, we have \( r^2 \left( \frac{\pi}{4} - \frac{\sqrt{3}}{2} \right) \leq |R_{x,y}| \leq \pi r^2 \).

There is an abuse of notation, \(|\cdot|\) will be understood according to the situation. That is, whenever \( R \) is a region, then \(|R|\) denotes the area of the region \( R \), and if \( a, b \in \mathbb{R} \) or \([0,1]^2\) then \(|a-b|\) denotes the distance between \( a \) and \( b \).

Lemma 4. Let \( R_{x,y} \) be as defined in Lemma 3. Let \( \gamma = \frac{5\alpha+2}{12} \). Then, with high probability the following holds:
\[
| |R_{x,y}| - |R_{u,v}| | \geq \frac{C_1}{n^\gamma},
\]
for some constant \( C_1 > 0 \).

The last lemma shows that the difference between areas of \( R_{x,y} \) and \( R_{u,v} \) is bounded below by \( n^{-\gamma} \) for some \( 0 < \gamma < 1 \) with high probability. We now prove Theorem 2.
Proof of Theorem\textsuperscript{2} As mentioned earlier, as a consequence of Lemma 1, it is enough to show that \( W_{x,y} \neq W_{u,v} \) with high probability whenever \((x,y)\) and \((u,v)\) are two distinct edges to show that the graph \( G(n,r) \) is exactly reconstructable for any \( r \).

Observe that a point is connected with both \( x \) and \( y \) if and only if the point lies in the area \( R_{x,y} \), which occurs with probability \(|R_{x,y}|\). Thus \( W_{x,y} \) is distributed as a \( \text{Bin}(n-2,|R_{x,y}|) \) random variable, as all the points are placed independently. Then by Lemma 3 we get

\[
E(W_{x,y}) = (n-2)|R_{x,y}| \leq n^{1-\alpha}.
\]

Recall that \( d = (1-\alpha)/3 \). Therefore, by Lemma 2, Lemma 3 implies that there exists an \( s > 0 \) such that

\[
P\left( |W_{x,y} - E(W_{x,y})| \geq \pi n^{\frac{3}{2}(1-\alpha)} \right) = o(n^{-s}), \text{ as } n \to \infty.
\]

Note that (3) also holds for \( W_{u,v} \). Therefore, with high probability,

\[
|W_{x,y} - E(W_{x,y})| + |W_{u,v} - E(W_{u,v})| \leq 2\pi n^{\frac{3}{2}(1-\alpha)}.
\]

Again from Lemma 4 we have

\[
|E(W_{x,y}) - E(W_{u,v})| \geq \left( 1 - \frac{2}{n} \right) C_1 n^{1-\gamma} \geq \frac{1}{2} C_1 n^{\frac{10n-5n\alpha}{12}}
\]

with high probability. Note that \( 2\pi n^{\frac{3}{2}(1-\alpha)} < \frac{1}{2} C_1 n^{\frac{10n-5n\alpha}{12}} \) since \( n \frac{3n+2}{n^2} > 4C_1 \pi^{-1} \) for all large \( n \). Hence combining (1) and (5) we conclude that

\[
|W_{x,y} - W_{u,v}| \geq \frac{1}{2} C_1 n^{\frac{10n-5n\alpha}{12}} - 2\pi n^{\frac{3}{2}(1-\alpha)}
\]

\[
\geq \frac{1}{2} C_1 n^{\frac{10n-5n\alpha}{12}} (1 - o(1)),
\]

with high probability. In particular, \( W_{x,y} \neq W_{u,v} \) with high probability. This completes the proof. \( \blacksquare \)

The rest of the section is dedicated to prove Lemmas 2, 3 and 4. We note down the well known Chernoff’s bound from Lemma 3, which will be used in the proof of Lemma 2.

Lemma 5. [Chernoff’s bound] Let \( X_1, X_2, \ldots, X_n \) be independent indicator random variables and call \( X = \sum_{i=1}^{n} X_i \). Then for any \( \delta > 0 \),

\[
P(X \leq (1-\delta)E(X)) \leq \exp\left(-\frac{\delta^2}{2}E(X)\right) \text{ and}
\]

\[
P(X \geq (1+\delta)E(X)) \leq \exp\left(-\frac{\delta^2}{2+\delta}E(X)\right).
\]

Proof of Lemma 3 We need to show that for some \( s > 0 \),

\[
P\left( |W_{x,y} - E(W_{x,y})| \geq \frac{E(W_{x,y})}{n^d} \right) = o(n^{-s}), \text{ as } n \to \infty.
\]

Using Chernoff’s bound (Lemma 5), we have

\[
P\left( |W_{x,y} - E(W_{x,y})| \geq n^{-d}E(W_{x,y}) \right) \leq \exp\left(-\frac{n^{-2d}}{2+n^{-d}}E(W_{x,y})\right).
\]
Since $E(W_{x,y}) = (n - 2)|R_{x,y}|$, $d = (1 - \alpha)/3$ and $r^2 = n^{-\alpha}$, Lemma 3 implies that
\[
P\left((W_{x,y} - E(W_{x,y})) \geq n^{-d}E(W_{x,y})\right) \leq \exp\left(-\frac{n^{-2d}(n - 2)(\frac{2\pi}{3} - \sqrt{3})r^2}{4(2 + n^{-d})}\right) = o(n^{-s}), \quad \text{as } n \to \infty,
\]
for any $s > 0$. Similarly using the other inequality of Chernoff’s bound (Lemma 5), it can again be shown that for some $s > 0$,
\[
P\left((W_{x,y} - E(W_{x,y})) \leq -n^{-d}E(W_{x,y})\right) = o(n^{-s}) \quad \text{as } n \to \infty.
\]
Hence the lemma. ■

The following simple fact will be used in the proof of Lemma 3.

**Fact 1.** The area over a chord within a circle of radius $r$ where the chord makes an angle $\theta$ in the center (described in Figure 1) is equal to $\frac{1}{2}r^2(\theta - \sin \theta)$.

![Area under a cord](image)

**Figure 1.** Area under a cord

A proof of Fact 1 is provided in the Appendix.

**Proof of Lemma 3.** The region $R_{x,y}$ is described in Figure 2. Using the formula as given in Figure 1, we have $|R_{x,y}| = r^2(\theta - \sin \theta)$ where $\theta = 2\cos^{-1}(|x - y|/2r)$. 
Finally, the bounds $r^2 \left( \frac{2\pi}{3} - \frac{\sqrt{3}}{2} \right) \leq |R_{x,y}| \leq \pi r^2$ follows directly from the observation that $|R_{x,y}|$ is monotone and it approaches its supremum and infimum values as $|x - y|$ tends to 0 and $r$, respectively.

It remains to prove Lemma 4. The following fact will be used in the proof.

**Fact 2.** Let $X$ and $Y$ are independent and identically distributed (i.i.d.) uniform random variables on $[0, 1]^2$. Suppose $f_Z(z)$ denotes the probability density function of $Z = X - Y$. Then

$$f_Z(z) \leq 1 \text{ for } z \in \mathbb{R}^2.$$ 

For the sake of completeness we provide a proof of Fact 2 in the Appendix. Also the exact expression of $f_Z(z)$ is given in the proof. See (10).

**Proof of Lemma 4.** It is enough to show that for some $s > 0$,

$$P \left( ||R_{x,y}|| - ||R_{u,v}|| \leq C_1 n^{-s} \right) = o(n^{-s}) \text{ as } n \to \infty.$$ 

Observe that the difference between the area of $R_{x,y}$ and $R_{u,v}$ can be bounded below by the rectangle $R_0$, whose two adjacent sides are of lengths $||x - y|| - ||u - v||/2$ and $(4r^2 - ||x - y||^2)^{1/2}$. See Figure 3.
Hence area of $R_0$ can be bounded below in the following way:

$$|R_0| = \frac{1}{2} \left( |x - y| - |u - v| \right) \left( \sqrt{4r^2 - |x - y|^2} \right)$$

(7)

the last line follows from the fact that $|x - y| \leq r$. Note that, for $a > 0$, we have

$$P \left( \left| |x - y| - |u - v| \right| \leq a \right) = \int_{x',y' \in [0,1]^2} P[(u - v) \in B(x' - y', a)]dx'dy'.$$

(8)

Let $Z = u - v$. Fact 2 implies that

$$P[Z \in B(x' - y', a)] = \int_{z \in B(x' - y', a)} f_Z(z)dz \leq \pi a^2,$$

(9)

as $f_Z(z) \leq 1$. Using (9) from (8) we get

$$P \left( \left| |x - y| - |u - v| \right| \leq a \right) \leq \pi a^2.$$

In particular, if $a = n^{-\gamma_0}$ where $\gamma_0 = (2 - \alpha)/12$, we have

$$P \left( \left| |x - y| - |u - v| \right| > n^{-\gamma_0} \right) = 1 - O(n^{-2\gamma_0}).$$

Thus the above observation along with the equation (7) gives us that

$$P \left( |R_0| \geq \frac{\sqrt{3}}{2} n^{-(\gamma_0 + \frac{\alpha}{2})} \right) = 1 - O(n^{-2\gamma_0}).$$

The proof finishes by the fact that $||R_{x,y} - R_{u,v}|| \geq |R_0|$ and $\gamma = \gamma_0 + \frac{\alpha}{2}$. ■
4. Appendix

In this section, we provide proofs of Facts 1 and 2 for the sake of completeness.

Proof of Fact 1. Observe that the area of the region $OABC$ in Figure 1 is $\frac{1}{2}r^2\theta$. The length of the chord $AC$ and the height $OD$ of the triangle $\triangle OAC$ are $2r \sin(\theta/2)$ and $r \cos(\theta/2)$, respectively. Therefore, the area of $\triangle OAC$ is $\frac{1}{2}r^2\sin\theta$ and hence the required area over the chord is $\frac{1}{2}r^2(\theta - \sin\theta)$. ■

Proof of Fact 2. Let $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$, where $X_1, X_2, Y_1, Y_2$ are independent and uniformly distributed random variables in $[0, 1]$. Then

$$Z = X - Y = (X_1 - Y_1, X_2 - Y_2).$$

Let $f_X$ be the probability density function of $X$. Let $Z_1 := X_1 - Y_1$ and $Z_2 := X_2 - Y_2$. Then by the independence we have

$$f_Z(z) = f_{Z_1}(z_1)f_{Z_2}(z_2), \text{ for } z = (z_1, z_2) \in [-1, 1]^2. \tag{10}$$

Let $F_{Z_1}$ denote the cumulative distribution function of $Z_1$. Then, for $-1 \leq z_1 \leq 0$,

$$F_{Z_1}(z_1) = P(Z_1 \leq z_1) = P(X_1 \leq z_1 + Y_1)$$

$$= \int_{y_1 = -z_1}^{1} \int_{x_1 = 0}^{y_1 + z_1} dx_1 dy_1$$

$$= \frac{(1 + z_1)^2}{2}.$$

as $X_1$ and $Y_1$ are i.i.d. uniform random variables in $[0, 1]$. Again, for $0 \leq z_1 \leq 1$,

$$F_{Z_1}(z_1) = P(X_1 \leq z_1 + Y_1)$$

$$= \int_{y_1 = 0}^{1 - z_1} \int_{x_1 = 0}^{y_1 + z_1} dx_1 dy_1 + \int_{y_1 = 1 - z_1}^{1} \int_{x_1 = 0}^{1} dx_1 dy_1$$

$$= \frac{1}{2} - \frac{z_1^2}{2} + z_1.$$

Therefore the probability density function of $Z_1$ is given by

$$f_{Z_1}(z_1) = \begin{cases} 1 + z_1 & \text{if } -1 \leq z_1 \leq 0 \\ 1 - z_1 & \text{if } 0 \leq z_1 \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, $f_{Z_1}(z_1) \leq 1$ for $z_1 \in [-1, 1]$. Similarly, $f_{Z_2}(z_2) \leq 1$ for all $z_2 \in [-1, 1]$. Therefore from (10) we get

$$f_Z(z) \leq 1, \text{ for } z \in [-1, 1]^2.$$

Hence the result. ■

Acknowledgement

The research of KA was partially supported by the Inspire Faculty Fellowship: DST/INSPIRE/04/2020/000579. SC’s research was supported by the NBHM post-doctoral fellowship (order no. 0204/19/2021/R&D-II/11178).


References

[1] C. Bordenave, U. Feige, and E. Mossel. Shotgun assembly of random jigsaw puzzles. *Random Structures & Algorithms*, 56(4):998–1015, 2020.

[2] J. Gaudio and E. Mossel. Shotgun assembly of Erdős–Rényi random graphs. *arXiv preprint arXiv:2010.14661*, 2020.

[3] H. Huang and K. Tikhomirov. Shotgun assembly of unlabeled erdos-renyi graphs. *arXiv preprint arXiv:2108.09636*, 2021.

[4] E. Mossel and N. Ross. Shotgun assembly of labeled graphs. *IEEE Trans. Network Sci. Eng.*, 6(2):145–157, 2019.

[5] E. Mossel and N. Sun. Shotgun assembly of random regular graphs. *arXiv preprint arXiv:1512.08473*, 2015.

[6] L. Yartseva, J. E. Simoes, and M. Grossglauser. Assembling a network out of ambiguous patches. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 883–890. IEEE, 2016.

Department of Mathematics, Indian Institute of Science Education and Research, Bhopal 462066

Email address: kartickmath [at] gmail.com

Stat-Math Unit, Indian Statistical Institute, 203 B T Road, Kolkata 700108

Email address: sukrit049 [at] gmail.com