

# Shotgun assembly of random graphs

Tom Johnston\*      Gal Kronenberg<sup>†‡</sup>      Alexander Roberts  
   Alex Scott<sup>†§</sup>

June 4, 2023

## Abstract

In the graph shotgun assembly problem, we are given the balls of radius  $r$  around each vertex of a graph and asked to reconstruct the graph. We study the shotgun assembly of the Erdős-Rényi random graph  $\mathcal{G}(n, p)$  from a wide range of values of  $r$ . We determine the threshold for reconstructibility for each  $r \geq 3$ , extending and improving substantially on results of Mossel and Ross for  $r = 3$ . For  $r = 2$ , we give upper and lower bounds that improve on results of Gaudio and Mossel by polynomial factors. We also give a sharpening of a result of Huang and Tikhomirov for  $r = 1$ .

## 1 Introduction

When can we reconstruct a graph from local information? In the *shotgun assembly problem*, we are given the balls  $N_r(v)$  of radius  $r$  around each vertex of a graph  $G$  and want to reconstruct the graph from this information. Problems of this type arise naturally in DNA shotgun assembly, where the goal is to reconstruct a DNA sequence from a collection of shorter stretches of the sequence (see [16, 4, 34] among many references), and have also been considered in the neuronal network context [43]. The shotgun assembly problem for random graphs was introduced in an influential paper of Mossel and Ross [32], which also raised a number of interesting variants such as the reconstruction of random jigsaws (see [40, 28, 6, 29, 11]) and random colourings of lattices (see [41, 14]). There has also recently been work on the closely related problem of reconstructing random pictures [39].

In this paper we will be concerned with the shotgun assembly of an Erdős-Rényi random graph  $G \in \mathcal{G}(n, p)$ , where each edge is open independently with probability  $p = p(n)$ . This model has been extensively studied [32, 17, 21, 15] (there is also interesting work on other random graphs including random regular graphs [33], random geometric

---

\*School of Mathematics, University of Bristol, Bristol, BS8 1UG, UK and Heilbronn Institute for Mathematical Research, Bristol, UK. Email: [tom.johnston@bristol.ac.uk](mailto:tom.johnston@bristol.ac.uk).

†Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, United Kingdom.

‡Email: [kronenberg@maths.ox.ac.uk](mailto:kronenberg@maths.ox.ac.uk). Supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska Curie grant agreement No. 101030925.

§Email: [scott@maths.ox.ac.uk](mailto:scott@maths.ox.ac.uk). Research supported by EPSRC grant EP/X013642/1.

graphs [2] and random simplicial complexes [1]). Let us start by defining the problem more carefully. For a graph  $G$ , let  $N_r^{(G)}(v)$  be the graph induced by the vertices at distance at most  $r$  from  $v$ , where the vertices are unlabelled except for the vertex  $v$ . For an integer  $r \geq 1$  and graphs  $G$  and  $H$ , we say  $G$  and  $H$  *have isomorphic  $r$ -neighbourhoods* if there is a bijection  $\phi : V(G) \rightarrow V(H)$  such that for each vertex  $v$  of  $G$  there is an isomorphism from the  $r$ -neighbourhood  $N_r(v)$  around  $v$  in  $G$  to the  $r$ -neighbourhood  $N_r(\phi(v))$  around  $\phi(v)$  in  $H$  which maps  $v$  to  $\phi(v)$ . We say that  $G$  is *reconstructible from its  $r$ -neighbourhoods* (or  *$r$ -reconstructible*) if every graph with  $r$ -neighbourhoods isomorphic to those of  $G$ , is in fact isomorphic to  $G$ . The general problem is to determine for what range of  $p$  a random graph  $G \in \mathcal{G}(n, p)$  is reconstructible (or non-reconstructible) from its  $r$ -neighbourhoods with high probability (i.e. with probability tending to 1 as  $n$  tends to infinity). We improve on previous bounds for all values of  $r$ , and give a fairly complete picture for  $r \geq 3$ .

For very small  $p$ , the general picture is similar for all  $r$ . Indeed, we show that at every radius  $r$  there is a phase transition when  $p$  is around  $n^{-\frac{2r+1}{2r}}$ . If  $p = o(n^{-\frac{2r+1}{2r}})$ , every component of the graph contains at most  $2r$  vertices and is contained entirely in an  $r$ -ball, and we reconstruct the graph by iteratively identifying and removing the largest components. On the other hand, if  $p$  grows slightly faster than  $n^{-\frac{2r+1}{2r}}$ , then with high probability we obtain a graph that is not  $r$ -reconstructible.

The more difficult question is what happens for larger  $p$ . It seems likely that for every radius  $r$  there should be a second phase transition around some threshold  $t = t(n)$ . By which we mean that, if  $p = \omega(t(n))$ , then  $G$  is with high probability reconstructible from its  $r$ -neighbourhoods, while if  $p = o(t(n))$  and  $p = \omega(n^{-\frac{2r+1}{2r}})$ , then with high probability  $G$  is not reconstructible from its  $r$ -neighbourhoods. This was not previously known at any radius. Our results here prove the existence of this second phase transition for all  $r \geq 3$ , and narrow the gap for  $r = 1, 2$ . We start by giving our main results regarding  $r \geq 3$ , and then discuss reconstruction from the 1- and 2-neighbourhoods and give some small improvements.

*Radius 3:* We begin by looking at reconstruction from balls of radius 3, and give the correct threshold for when  $\mathcal{G}(n, p)$  is 3-reconstructible with high probability. Mossel and Ross [32] considered reconstruction from balls of radius 3 and showed that  $G \in \mathcal{G}(n, p)$  is with high probability 3-reconstructible when  $p = \omega(\log^2 n/n)$ . We improve on this result, and show that there are two phase transitions: the first is around  $n^{-7/6}$ , and the second is around  $\frac{\log^2 n}{n(\log \log n)^3}$ .

**Theorem 1.1.** *Let  $p = p(n)$  and  $G \in \mathcal{G}(n, p)$ . There exist  $\beta > \alpha > 0$  such that the following hold.*

- (i) *If  $p = o(n^{-7/6})$ , then  $G$  is reconstructible from its 3-neighbourhoods with high probability.*
- (ii) *If  $p = \omega(n^{-7/6})$  and  $p \leq \alpha \frac{\log^2 n}{n(\log \log n)^3}$ , then with high probability  $G$  is not reconstructible from its 3-neighbourhoods.*
- (iii) *If  $p \geq \beta \frac{\log^2 n}{n(\log \log n)^3}$ , then  $G$  is reconstructible from its 3-neighbourhoods with high probability.*

*Radius 4 or more:* A similar picture holds for any fixed radius  $r \geq 4$  (and in fact even when  $r$  grows slowly with  $n$ ), except that the second phase transition comes earlier by a factor of roughly  $\frac{\log n}{(\log \log n)^3}$ .

**Theorem 1.2.** *Let  $p = p(n)$  and  $G \in \mathcal{G}(n, p)$ . There exist  $\beta > \alpha > 0$  such that the following hold for all  $4 \leq r = o(\log n)$ .*

- (i) *If  $p = o(n^{-\frac{2r+1}{2r}})$ , then  $G$  is reconstructible from its  $r$ -neighbourhoods with high probability.*
- (ii) *If  $p = \omega(n^{-\frac{2r+1}{2r}})$  and  $p \leq \alpha \frac{\log n}{rn}$ , then with high probability  $G$  is not reconstructible from its  $r$ -neighbourhoods.*
- (iii) *If  $p \geq \beta \frac{\log n}{rn}$ , then  $G$  is reconstructible from its  $r$ -neighbourhoods with high probability.*

Recently, Gaudio, Rácz and Sridhar [18] independently studied the special case of  $r = 4$  as part of their work on local canonical labellings of Erdős-Rényi graphs and showed that  $G \in \mathcal{G}(n, p)$  is 4-reconstructible with high probability when  $np \geq (1 + \delta) \log n$ .

*Radius 2:* We next move to the case where  $r = 2$ . It is not hard to see that if  $p = \omega(\sqrt{\log n/n})$ , then  $G \in \mathcal{G}(n, p)$  is 2-reconstructible with high probability as the diameter of  $G$  is at most 2 with high probability (and so the 2-balls are the entire graph). Better bounds were given by Gaudio and Mossel [17] who showed that, for any  $\varepsilon > 0$ ,  $G$  is 2-reconstructible with high probability when  $n^{-3/5+\varepsilon} \leq p \leq n^{-1/2-\varepsilon}$  or  $p \geq n^{-1/2+\varepsilon}$ . We extend the range at the lower end, and remove the gap in the middle.

**Theorem 1.3.** *Let  $p = p(n)$  and  $G \in \mathcal{G}(n, p)$ . There exists a constant  $\delta > 0$  such that the following holds. If  $p \geq n^{-2/3-\delta}$ , then  $G$  is reconstructible from its 2-neighbourhoods with high probability.*

For slightly sparser graphs, Gaudio and Mossel [17] showed that  $G$  fails to be 2-reconstructible with high probability when  $n^{-1+\varepsilon} \leq p \leq n^{-3/4-\varepsilon}$ . We extend this range in both directions as follows.

**Theorem 1.4.** *Let  $p = p(n)$  and  $G \in \mathcal{G}(n, p)$ . If  $p \leq \frac{1}{3}n^{-3/4} \log^{1/4} n$  and  $p = \omega(n^{-5/4})$ , then with high probability  $G$  cannot be reconstructed from its 2-neighbourhoods.*

Once again, the lower bound on  $p$  in Theorem 1.4 is best possible.

**Theorem 1.5.** *Let  $p = p(n)$  and  $G \in \mathcal{G}(n, p)$ . If  $p = o(n^{-5/4})$ , then with high probability  $G$  is reconstructible from its 2-neighbourhoods.*

We note that there is still a gap where we do not know whether  $\mathcal{G}$  can be reconstructed with high probability, and it would be interesting to remove this.

**Question.** Determine when  $\mathcal{G}(n, p)$  is 2-reconstructible. Is there a threshold around  $n^{-3/4}$  (up to a polylogarithmic factor)?

*Radius 1:* We finish this section by looking at reconstruction from balls of radius 1. Gaudio and Mossel [17] showed that, for any  $\varepsilon > 0$ , a random graph  $G \in \mathcal{G}(n, p)$  is 1-reconstructible with high probability when  $n^{-1/3+\varepsilon} \leq p \leq n^{-\varepsilon}$ ; and fails to be 1-reconstructible with high probability when  $n^{-1+\varepsilon} \leq p \leq n^{-1/2-\varepsilon}$ . This was recently improved in an impressive paper of Huang and Tikhomirov [21] which showed that there are constants  $c, C > 0$  such that  $G$  is 1-reconstructible with high probability when  $n^{-1/2} \log^C n \leq p \leq c$ , while  $G$  fails to be 1-reconstructible if  $p = o(1/\sqrt{n})$  and  $p = \omega(\log n/n)$ . This shows that there is a change of behaviour around  $n^{-1/2}$ , up to a polylogarithmic gap. We give a small improvement on the region where  $G$  fails to be 1-reconstructible: we improve the lower bound, and give a slight sharpening of the upper bound. Note that in particular this shows that some polylogarithmic factor is indeed necessary.

**Theorem 1.6.** *Let  $p = p(n)$  and  $G \in \mathcal{G}(n, p)$ . If  $p = \omega(n^{-3/2})$  and  $p \leq \sqrt{\frac{\log n}{25n}}$ , then with high probability  $G$  cannot be reconstructed from its 1-neighbourhoods.*

We further show that the lower bound is sharp.

**Theorem 1.7.** *Let  $p = p(n)$  and  $G \in \mathcal{G}(n, p)$ . If  $p = o(n^{-3/2})$ , then with high probability  $G$  is reconstructible from its 1-neighbourhoods.*

We note that, for very sparse graphs, there are results for even larger radii. Mossel and Ross [32] showed that if  $p = \lambda/n$  with  $\lambda \neq 1$ , then there are constants  $c_\lambda, C_\lambda$  such that  $G$  is with high probability  $r$ -reconstructible if  $r \geq C_\lambda \log n$  and with high probability not  $r$ -reconstructible if  $r \leq c_\lambda \log n$ . Very recently sharp asymptotics were obtained by Ding, Jiang and Ma [15] (including for the case  $\lambda = 1$ ).

As with most graph reconstruction problems, the shotgun assembly problem is closely related to the famous *reconstruction conjecture* of Kelly and Ulam [23, 24, 46]. The conjecture asserts that every graph  $G$  with at least 3 vertices can be determined up to isomorphism from its vertex-deleted subgraphs (i.e. from the multiset  $\{G - v : v \in V(G)\}$  of unlabelled subgraphs). There has been substantial work by many different authors over many years on this conjecture (see e.g. [10, 9, 5, 26] for surveys and background), and on variants with less information such as using fewer subgraphs (see e.g. [38, 36, 37, 7, 31, 12]) and smaller subgraphs (see e.g. [19, 35, 25, 44, 20]). Müller 1976 [35] and Bollobás 1990 [7] showed that the conjecture holds for almost all graphs. In fact, they showed that for reconstructing a random graph one needs significantly less information, for example, only a few of the vertex-deleted subgraphs are needed. The shotgun assembly problem can thus be viewed as a variant of the reconstruction problem using just *local* information.

The paper is organised as follows. In the next section, we give a brief discussion of our proof techniques, and state some probabilistic lemmas that we will use throughout the rest of the paper. In Section 3 we give skeleton proofs for Theorems 1.1 and 1.2, breaking the full proof into a series of (technical) claims that will be proved in Section 6. In Section 4 we prove Theorem 1.3, and in Section 5 we prove Theorem 1.4 and Theorem 1.6.

## 2 Discussion and definitions

In this section we give short descriptions of some of the main ideas in our proofs. We will use a very simple but powerful tool for reconstructing graphs, known as the ‘overlap method’, which was introduced in the paper of Mossel and Ross [32]. Intuitively, it seems reasonable that if the neighbourhoods of different vertices are very different from each other, then one might be able to identify vertices in the neighbourhoods of other vertices and reconstruct the graph. In  $N_r(v)$  we can see the entire  $(r - 1)$ -neighbourhood of the neighbours of  $v$ , so if all the  $(r - 1)$ -neighbourhoods are unique, then we can identify the neighbours of  $v$  from its  $r$ -neighbourhood. This leads to the following lemma.

**Lemma 2.1** ([32, Lemma 2.4]). *Suppose that a graph  $G$  has unique  $(r - 1)$ -neighbourhoods. Then it is reconstructible from its  $r$ -neighbourhoods.*

We will use this lemma when we prove reconstructibility in the proofs of Theorem 1.1(iii) and Theorem 1.2(iii). However, proving the uniqueness of neighbourhoods is not always a simple task, especially for such a large range of  $p$ . Moreover, for large values of  $r$ , we will not have uniqueness of  $(r - 1)$ -neighbourhoods for the entire range of  $p$  we consider and we cannot apply the method as is. We will instead use the idea of the overlap method to handle high-degree vertices and then apply a different argument for low degree vertices.

Reconstructibility below the first phase transition, that is reconstructibility when  $p = o(n^{-\frac{2r+1}{2r}})$ , will follow easily from the fact that all components are with high probability small enough to be guaranteed to be fully contained in an  $r$ -neighbourhood.

For showing non-reconstructibility, we need to prove that with high probability there is a second graph  $H$  which is not isomorphic to  $G$  but has isomorphic  $r$ -neighbourhoods. When considering smaller values of  $p$ , that is, closer to the first phase transition, our reasoning for non-reconstructibility will lie in the small components. Indeed, for such values of  $p$  there will be components that are paths with  $2r + 1$  vertices with high probability. The non-reconstructibility will follow from the fact that the collection of  $r$ -neighbourhoods of two disjoint copies of  $P_{2r+1}$  (a path with  $2r + 1$  vertices) is isomorphic to the collection of  $r$ -neighbourhoods of disjoint copies of  $P_{2r}$  and  $P_{2r+2}$ , and therefore graphs containing these cannot be uniquely identified. Interestingly, for  $r \geq 4$  being non-reconstructible coincides with the existence of these small components, and the second threshold for reconstructibility is around the point where we stop seeing two disjoint copies of  $P_{2r+1}$  as components. For  $r \leq 3$  however, a different phenomena occurs and with high probability it is not possible to reconstruct  $G$  even after the disappearance of these small paths. Roughly speaking, it turns out that (with high probability) we can find two edges  $uv$  and  $xy$ , where the  $(r - 1)$ -neighbourhoods of the end vertices are isomorphic, but the  $r$ -neighbourhoods are not. We can replace the edges by  $uy$  and  $xv$  to get a graph with the same collection of  $r$ -neighbourhoods, but which is in a different isomorphism class. This property will continue beyond the existence of two isolated copies of  $P_{2r+1}$  for  $r \leq 3$ , and for  $r = 3$  it is instead the disappearance of this property which coincides the second phase transition.

We use the following notation to distinguish between different types of neighbourhoods. For a vertex  $v$ , we let  $\Gamma_r(v)$  be the set of vertices that are at distance *exactly*  $r$  from  $v$ . We write  $|\Gamma_r(v)|$  for the number of such vertices. In the special case that  $r = 1$

we simply write  $\Gamma(v)$  and we use  $d(v) = |\Gamma(v)|$  to denote the degree of the vertex  $v$ . As mentioned above, we let  $N_r^{(G)}(v)$  be the graph induced by the vertices at distance at most  $r$  from  $v$ , where the vertices are unlabelled except for the vertex  $v$ . We also use  $\Gamma_{\leq r}(v)$  to denote the set of vertices of the graph  $N_r^{(G)}(v)$  (i.e. the vertices at distance at most  $r$  from  $v$ ). In some proofs we will consider subgraphs consisting of neighbourhoods of several vertices and we will give the relevant notation as and when it is needed.

**Remark 2.2.** One can also consider exact reconstructibility. A graph  $G$  is said to be *exactly reconstructible* from its  $r$ -neighbourhoods if  $G$  is the unique labelled graph with its collection of  $r$ -neighbourhoods, i.e. for any  $H$  such that  $N_r^{(G)}(v) \simeq N_r^{(H)}(v)$  for every  $v \in V(G)$ , we have  $H = G$ . Lemma 2.1 holds for exact reconstructibility, but not all reconstructible graphs are exactly reconstructible. For example, any graph with two disjoint edges as components cannot be reconstructed exactly from its neighbourhoods. In particular, this means there is some  $\alpha > 0$  such that  $\mathcal{G}(n, p)$  is not exactly reconstructible with high probability when  $p$  is both  $\omega(1/n^2)$  and at most  $\alpha \log n/n$ . This contrasts with Theorems 1.1(i), 1.2(i), 1.7 and 1.5 which show that  $\mathcal{G}(n, p)$  is reconstructible for some this range. When  $p \leq 1/2$  and  $p = \omega(\log^4 n / (n \log \log n))$ , the degree neighbourhoods of vertices are unique with high probability [13]. When this is true, exact reconstructibility from  $r$ -neighbourhoods is the same as non-exact reconstructibility for all  $r \geq 2$ . It follows that, when  $p \leq 1/2$ , we have exact reconstructibility in Theorem 1.3. A minor adaptation of the proof of Theorem 1.1(iii) would give exact reconstructibility.

## 2.1 Useful facts

In this section we state some well known probabilistic bounds which will be useful later in the paper. We start by stating a simple fact about the median(s) of the binomial distribution.

**Fact 2.3.** Let  $X \sim \text{Bin}(n, p)$ . Then  $\mathbb{P}(X > \lceil np \rceil) \leq 1/2$ .

We will make frequent use of the following well-known bounds on the tails of the binomial distribution, known as Chernoff bounds (see e.g. [3], [22], [30]).

**Lemma 2.4** (Follows from Theorem 4.4 in [30]). *Let  $X \sim \text{Bin}(n, p)$  and  $\varepsilon > 0$ . Then*

$$\begin{aligned} \mathbb{P}(X \geq (1 + \varepsilon)np) &\leq \exp\left(-\frac{\varepsilon^2 \mu}{2 + \varepsilon}\right), \\ \mathbb{P}(X \leq (1 - \varepsilon)np) &\leq \exp\left(-\frac{\varepsilon^2 \mu}{2}\right). \end{aligned}$$

We will also be interested in tail bounds for binomial distributions where  $\mu \rightarrow 0$  as  $n \rightarrow \infty$ , for which we use the following simple bound.

**Lemma 2.5.** *Let  $X \sim \text{Bin}(n, p)$  and  $k \in \mathbb{N}$ . Then*

$$\mathbb{P}(X \geq k) \leq e(np)^k.$$

*Proof.* We have

$$\mathbb{P}(X \geq k) = \sum_{j=k}^n \binom{n}{j} p^j (1-p)^{n-j} \leq \sum_{j=k}^n \frac{n^j}{j!} p^j \leq (np)^k \sum_{j=0}^{\infty} \frac{1}{j!},$$

and the result is immediate.  $\square$

We will also want to bound the probability that a binomial (or Poisson binomial) random variable takes a specific value, and we now give several useful lemmas bounding these probabilities. The first, due to Rogozin [42], bounds the probability of a mode of independent discrete random variables.

**Theorem 2.6** (Theorem 2 in [42]). *Let  $X_1, \dots, X_n$  be a sequence of independent discrete random variables, and let  $S = X_1 + \dots + X_n$ . Let  $p_i = \sup_x \mathbb{P}(X_i = x)$ . Then*

$$\sup_x \mathbb{P}(S = x) \leq \frac{C}{\sqrt{\sum_{i=1}^n (1-p_i)}}$$

where  $C$  is an absolute constant.

The following estimation can be derived from the proofs of Theorem 1.2 and Theorem 1.5 in [8].

**Theorem 2.7.** *Suppose  $X \sim \text{Bin}(n, p)$  where  $p = p(n)$  may depend on  $n$ . Let  $q = 1 - p$  and define  $\sigma(n)$  by  $\sigma = \sqrt{pqn}$ . If  $\sigma \rightarrow \infty$  as  $n \rightarrow \infty$ , then uniformly over all  $0 \leq h \leq \sigma^{5/4}$  such that  $pn + h \in \mathbb{Z}$ , we have*

$$\mathbb{P}(X = pn + h) = (1 + o_{\sigma}(1)) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{h^2}{2\sigma^2}\right).$$

In the proof of Theorem 1.3, we will approximate the sum of Bernoulli random variables with a Poisson random variable for which we use the following result. The first version of this result was given by Le Cam [27] in 1960, but there are now several variations and different proofs, and we refer the reader to [45] for more discussion. We will use the following the version.

**Theorem 2.8** (Le Cam Theorem). *Let  $X_1, \dots, X_n$  be independent Bernoulli random variables with success probabilities  $p_1, \dots, p_n$ . Let  $S = X_1 + \dots + X_n$  and let  $\mu$  denote the expectation of  $S$  (i.e.  $\mu = \mathbb{E}[S] = \sum_{i=1}^n p_i$ ). Then*

$$\sum_{k=0}^{\infty} \left| \mathbb{P}(S = k) - \frac{\mu^k e^{-\mu}}{k!} \right| < 2 \min\left\{1, \frac{1}{\mu}\right\} \sum_{i=1}^n p_i^2.$$

### 3 Reconstruction from $r$ -neighbourhoods, $r \geq 3$

In this section we use a series of lemmas to prove Theorem 1.1 and Theorem 1.2, but we delay proving the lemmas until the later sections. Both of these proofs employ different arguments for different ranges of  $p$ , although the proofs of parts (i) and (ii) are very similar in both cases.

We start by recording some simple facts about the structure of random graphs.

**Lemma 3.1.** *Let  $r \geq 1$ . If  $p = o(n^{-\frac{2r+1}{2r}})$ , then with high probability the largest component of a random graph  $G \in \mathcal{G}(n, p)$  has size at most  $2r$ .*

*Proof.* It is enough to show that, with high probability,  $G$  does not contain a tree on  $2r + 1$  vertices as a subgraph. This follows immediately from Markov's inequality, as the expected number of such subtrees is  $O(n^{2r+1}p^{2r}) = o(1)$ .  $\square$

**Lemma 3.2.** *There exists an  $\alpha > 0$  such that the following holds for all  $1 \leq r = o(\log n)$ . If  $p$  is such that  $pn^{\frac{2r+1}{2r}} = \omega(1)$  and  $p \leq \alpha \frac{\log n}{rn}$ , then  $G \in \mathcal{G}(n, p)$  contains two paths of  $2r + 1$  vertices as components with high probability.*

*Proof.* Fix  $\alpha < 1/6$ . Let  $X$  be the number of path components with  $2r + 1$  vertices. The expectation of  $X$  is

$$f(r, n, p) := \frac{1}{2} \binom{n}{2r+1} (2r+1)! p^{2r} (1-p)^{(2r+1)(n-2r-1) + \binom{2r+1}{2} - 2r}.$$

We may assume that  $r \leq \beta \log n$  and  $p \geq \lambda n^{-\frac{2r+1}{2r}}$  where  $\beta = \beta(n)$  and  $\lambda = \lambda(n)$  are functions that slowly tend to 0 and infinity respectively. For fixed  $n$  and  $r$ , the minimum of  $f(r, n, p)$  over  $p \in [\lambda n^{-\frac{2r+1}{2r}}, \alpha \frac{\log n}{rn}]$  is attained at one of the end points. When  $p = \lambda n^{-\frac{2r+1}{2r}}$ , a calculation shows that  $f(r, n, p) = \Omega(\exp(2r(\log \lambda - 2\lambda \exp(-1/\beta))))$ , and this is  $\omega(1)$  if  $\lambda$  grows sufficiently slowly compared to  $\beta$ . At the other end point, we have  $f(r, n, p) = \Omega(\exp((1 - 6\alpha) \log n + 2r \log(\alpha \log n/r)))$ , which is  $\omega(1)$  provided  $\alpha < 1/6$ . In particular,  $\mathbb{E}[X] \rightarrow \infty$  as  $n \rightarrow \infty$ .

We now bound  $\mathbb{E}[X^2]$ . Let  $\gamma$  be the probability that a specific set of  $2r + 1$  vertices induces a path component. Note that distinct components cannot share vertices, so  $\mathbb{E}[X^2]$  decomposes as  $\mathbb{E}[X]$  plus a sum over disjoint pairs of  $(2r + 1)$ -sets. The probability that two specific disjoint sets of  $2r + 1$  vertices both induce path components is  $\gamma^2(1-p)^{-(2r+1)^2}$ , as there are  $(2r + 1)^2$  potential edges between the sets. Thus  $\mathbb{E}[X^2] = \mathbb{E}[X] + \mathbb{E}[X(X - 1)] \leq (1 + o(1)) \mathbb{E}[X]^2$ . By Chebyshev's inequality, we obtain that with high probability  $X \geq 2$ .  $\square$

Combining the two lemmas above gives the following lemma, which handles the first phase transition.

**Lemma 3.3.** *Let  $G \in \mathcal{G}(n, p)$ . There is a constant  $\alpha > 0$  such that, for all  $r \geq 1$ ,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(G \text{ is } r\text{-reconstructible}) = \begin{cases} 1, & \text{if } p = o\left(n^{-\frac{2r+1}{2r}}\right), \\ 0, & \text{if } p = \omega\left(n^{-\frac{2r+1}{2r}}\right) \text{ and } p \leq \alpha \frac{\log n}{rn}. \end{cases}$$

*Proof.* The dense regime follows immediately from Lemma 3.2 and the fact that the graph consisting of two paths of  $2r + 1$  vertices is not reconstructible (see Section 2).

For the sparse regime, we note first that if a component has at most  $2r$  vertices, then it is contained in the  $r$ -ball around some vertex, and if there is a component with at least  $2r + 1$  vertices, there must be an  $r$ -ball with at least  $2r + 1$  vertices.

Suppose there is no  $r$ -ball with  $2r + 1$  vertices. Then we start by choosing an  $r$ -ball with as many vertices as possible: this gives us an entire component  $C$ , and from this we



can determine the  $r$ -balls of all vertices in  $C$ . We now delete all these  $r$ -balls from our collection, and repeat on the remaining  $r$ -balls (which are exactly the  $r$ -balls of  $G$  with  $C$  deleted). This will reconstruct the graph  $G$ , and the claim follows since Lemma 3.1 implies that every component has at most  $2r$  vertices with high probability.  $\square$

The following lemma will be useful when proving Theorem 1.2(iii).

**Lemma 3.4.** *There exists  $\beta > 0$  such that the following holds for all  $4 \leq r = o(\log n)$ , and  $p \geq \beta \frac{\log n}{rn}$ . Let  $G \in \mathcal{G}(n, p)$ , and let  $H$  be the subgraph of  $G$  induced by the vertices with degree at most  $np/2$ . Then with high probability the maximum component size of  $H$  is at most  $r - 3$ .*

*Proof.* Fix  $\beta > 30$ . It is enough to bound the probability of the event  $E$  that there is a set  $A$  of  $r - 2$  vertices such that  $G[A]$  is connected and each vertex in  $A$  has at most  $np/2$  neighbours outside  $A$ . For fixed  $A$ , these properties are independent. Note that there are  $O(n^{r-2})$  choices for  $A$ , and the probability that  $G[A]$  is connected is  $O(p^{r-3})$ . Let  $X \sim \text{Bin}(n - r + 2, p)$ . Then the probability that  $v \in A$  has at most  $np/2$  neighbours outside  $A$  equals  $\mathbb{P}(X \leq np/2)$ , which by a Chernoff bound (Lemma 2.4) is at most  $e^{-\frac{1}{5}np}$ . Thus

$$\mathbb{P}(E) = O(n^{r-2} p^{r-3} e^{-\frac{1}{5}np(r-2)}) = o(1),$$

for  $\beta > 30$ ,  $r = o(\log n)$  and  $p > \frac{\beta \log n}{rn}$ .  $\square$

We will also need several facts about small balls in random graphs. The proofs of these are more complicated so we postpone them to Section 6.

**Lemma 3.5.** *For any  $\varepsilon > 0$ , there exists  $\beta > 0$  such that, for  $\beta \frac{\log^2 n}{n(\log \log n)^3} \leq p \leq n^{-2/3-\varepsilon}$ , the 2-neighbourhoods of  $G \in \mathcal{G}(n, p)$  are unique with high probability.*

**Lemma 3.6.** *Suppose  $\frac{\log^{2/3} n}{n} \leq p \leq \frac{\log^2 n}{n}$ . Then, with high probability, there are no two vertices  $x, y$  of  $G \in \mathcal{G}(n, p)$  with degree at least  $np/2$  such that the 3-neighbourhoods around  $x$  and  $y$  are isomorphic (i.e. the 3-neighbourhoods around vertices with degree at least  $np/2$  are unique).*

**Lemma 3.7.** *Let  $\alpha > 0$  be a sufficiently small constant and suppose  $\frac{\log^{2/3} n}{n} \leq p \leq \alpha \frac{\log^2 n}{n(\log \log n)^3}$ . Then, for  $G \in \mathcal{G}(n, p)$ , with high probability there are distinct vertices  $x, y, u, v$  such that  $xy, uv \in E(G)$  and  $xv, yu \notin E(G)$  and the graph  $G'$  obtained from  $G$  by deleting  $xy, uv$  and adding  $xv, yu$  satisfies the following:*

1.  $G$  and  $G'$  are not isomorphic.
2.  $G$  and  $G'$  have the same collection of 3-balls.

We now piece together the lemmas above to give proofs of Theorem 1.1 and Theorem 1.2.

*Proof of Theorem 1.1(i) and Theorem 1.2(i).* Follows immediately from Lemma 3.3.  $\square$

*Proof of Theorem 1.1(ii) and Theorem 1.2(ii).* Theorem 1.2(ii) follows immediately from Lemma 3.3, but the lemma does not give the entire range of  $p$  needed in Theorem 1.1(ii), and we will use a different argument for larger  $p$ . To cover the remaining region, it is enough to show that there exists  $\alpha > 0$  such that  $G \in \mathcal{G}(n, p)$  is not reconstructible from its 3-neighbourhoods with high probability when  $\frac{\log^{2/3} n}{n} \leq p \leq \alpha \frac{\log^2 n}{n(\log \log n)^3}$ , and this is exactly the content of Lemma 3.7.  $\square$

*Proof of Theorem 1.1(iii).* Theorem 1.3 shows there is a constant  $\delta > 0$  such that the graph can be reconstructed from its 2-neighbourhoods with high-probability when  $p \geq n^{-2/3-\delta}$ . Hence, we can assume that  $\beta \frac{\log^2 n}{n(\log \log n)^3} \leq p \leq n^{-2/3-\delta/2}$ , and it follows from Lemma 3.5 that the 2-neighbourhoods are unique with high probability. The result now follows immediately by applying Lemma 2.1.  $\square$

*Proof of Theorem 1.2(iii).* By Theorem 1.1(iii),  $G \in \mathcal{G}(n, p)$  is reconstructible with high probability from its 3-neighbourhoods when  $p = \Omega(\log^2 n / (n \log \log n))$ , so we may assume that  $p = O(\log^2 n / n)$ . We use the overlap method to reconstruct the portion of the graph induced by vertices of moderately large degree; a further argument is needed to reconstruct the rest of the graph.

Let  $V_1$  be the vertices of  $G$  with degree at least  $np/2$  and let  $V_2 = V(G) \setminus V_1$ . For  $i = 1, 2$ , let  $H_i$  be the subgraph induced by  $V_i$ . For each vertex  $v$ , we can determine from its 4-ball whether  $v \in V_1$  or  $v \in V_2$ . When the 3-balls (in  $G$ ) around the vertices in  $V_1$  are unique, we can easily reconstruct  $H_1$  using the overlap method, and this event happens with high probability by Lemma 3.6.

Now consider  $H_2$ . By Lemma 3.4 we may assume that all components of  $H_2$  have at most  $r - 3$  vertices, and note that we can easily check that this holds from the  $r$ -balls. Consider a component  $C$  of  $H_2$ . For each vertex  $v$  of  $C$ , the  $(r - 4)$ -ball around  $v$  contains all vertices of  $C$ , so the  $(r - 3)$ -ball contains all vertices of  $V_1$  that are adjacent to a vertex of  $C$ . The  $r$ -ball around  $v$  contains the 3-balls around these vertices, and we assume these are all unique. It follows that by looking at the  $r$ -ball around  $v$ , we can identify  $C$  (up to isomorphism), and for each vertex of  $C$ , we can determine which vertices of  $V_1$  it is adjacent to. We obtain this information  $|C|$  times for each component  $C$  of  $H_2$  (once for each vertex of  $C$ ), and so allowing for multiplicities we can reconstruct all components of  $H_2$  and the way they are attached to  $H_1$ .  $\square$

## 4 Reconstruction from 2-neighbourhoods

In this section we prove Theorem 1.3. Since Gaudio and Mossel [17] proved that, for all  $\varepsilon > 0$ , a random graph  $G \in \mathcal{G}(n, p)$  can be reconstructed from its collection of 2-balls if  $n^{-1/2+\varepsilon} \leq p$  with high probability, we may assume that  $p \leq n^{-16/35}$ .

We use an approach similar to that of Gaudio and Mossel [17]. We will colour each edge  $uv$  by a colour which can be determined from the 2-neighbourhoods of both  $u$  and  $v$  and we attempt to reconstruct the graph from the edge-coloured stars around the vertices. Gaudio and Mossel [17] showed that this information is sufficient to reconstruct an edge-coloured graph when no two edges have the same colour. In order to prove our result, we will use colourings which satisfy a slightly weaker condition which is easier to show.

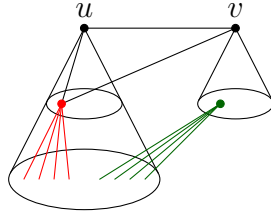


Figure 1: We will show the  $C_{uv}$  are unique by consider the number of edges each vertex in  $\Gamma_1(v)$  has to  $\Gamma_2(u) \setminus \Gamma_1(v)$ . The vertex adjacent to  $u$  and  $v$  shown in red will be problematic and we will view its degree as an “error”.

**Lemma 4.1.** *Let  $G$  be an edge-coloured graph such that every pair of edges of the same colour share a vertex. Then by looking only at the number of edges of each colour adjacent to each vertex,  $G$  can be reconstructed exactly.*

*Proof.* Let our edge-coloured stars be  $S_1, \dots, S_n$ , and label the corresponding centres  $v_1, \dots, v_n$ . Fix a colour  $c$  and consider the subgraph  $H$  consisting of all edges with this colour. From the degree sequence of  $H$  we can check if  $H$  (up to isolated vertices) is a triangle or a star, and note that these are the only graphs with no disjoint edges so  $H$  must be one of these graphs. In either case, we can reconstruct  $H$  by joining  $v_i$  and  $v_j$  with an edge in colour  $c$  whenever one of  $v_i$  and  $v_j$  is a vertex of largest degree in colour  $c$ . The graph  $G$  is the union (over all colours) of these subgraphs.  $\square$

We now give the edge colouring we will use and show that with high probability no two disjoint edges have the same colour. For an edge  $uv$ , let  $C_{uv}$  be the subgraph of  $G$  induced by the vertices at distance at most 2 from both  $u$  and  $v$ , where we distinguish the edge  $uv$ . We write  $C_{uv} \simeq C_{xy}$  if there is a bijection  $f : V(C_{uv}) \rightarrow V(C_{xy})$  such that  $ab \in E(C_{uv})$  if and only if  $f(a)f(b) \in E(C_{xy})$ , and  $\{f(u), f(v)\} = \{x, y\}$ . We will refer to each such isomorphism class as a *colour*. Theorem 1.3 follows immediately from Lemma 4.1 and the following.

**Lemma 4.2.** *There exists a constant  $\delta > 0$  such that the following holds. Suppose  $n^{-2/3-\delta} \leq p \leq n^{-16/35}$ , and let  $u, v, x, y$  be distinct vertices. The probability that  $uv$  and  $xy$  are edges, and  $C_{uv} \simeq C_{xy}$  is  $o(n^{-4})$ .*

Before proving Lemma 4.2, we explain how it implies Theorem 1.3.

*Proof of Theorem 1.3.* For each edge  $uv$  in  $G \in \mathcal{G}(n, p)$ , we colour the edge  $uv$  with the isomorphism class of  $C_{uv}$ , and note that for each vertex  $u$  it is possible to determine the colour of all edges incident with  $u$  from the 2-ball around  $u$ . Indeed, if  $x$  is a vertex at distance at most 2 from  $u$  and  $vw x$  is a path from  $v$  to  $x$ , then  $v, w$  and  $x$  are all contained in the 2-ball around  $u$ . This means we can determine which vertices in the 2-ball around  $u$  are also in the 2-ball around  $v$ , and we can determine the isomorphism class of  $C_{uv}$ . It follows from Lemma 4.2 that with high probability no two disjoint edges have the same colour, and by Lemma 4.1, we can then reconstruct  $G$ .  $\square$

Before giving the full details of the proof of Lemma 4.2, let us sketch our strategy. Suppose that  $C_{uv}$  and  $C_{xy}$  are isomorphic with  $u$  mapping to  $x$  and  $v$  mapping to  $y$ . Then it must be the case that the unordered degree sequence of  $\Gamma_1(v)$  into  $\Gamma_2(u) \setminus \Gamma_1(v)$  and

of  $\Gamma_1(y)$  into  $\Gamma_2(x) \setminus \Gamma_1(y)$  are equal, and we will show that the probability of this event is  $o(n^{-4})$ . We note that although we cannot see the whole of  $\Gamma_2(u)$  in  $C_{uv}$ , we do see all the edges from  $\Gamma_1(v)$  to  $\Gamma_2(u)$  and we can therefore read off the degree sequence of  $\Gamma_1(v)$  into  $\Gamma_2(u) \setminus \Gamma_1(v)$ . By symmetry, the probability of an isomorphism which maps  $u$  to  $y$  and  $v$  to  $x$  will also be  $o(n^{-4})$ .

Fix  $u$  and  $v$  and suppose that  $uv$  is an edge. We reveal the edges from  $u$  and  $v$ , and then from  $\Gamma_1(u)$ . Given a vertex  $i$  in  $\Gamma_1(v) \setminus \Gamma_1(u)$  which is not  $u$ , we have not revealed any of its edges to  $\Gamma_2(u) \setminus \Gamma_1(v)$  so the number of such edges  $b(i)$  is a binomial random variable with  $|\Gamma_2(u)|$  trials and success probability  $p$ . When  $p$  is only a little bit bigger than  $n^{-2/3}$ , we have  $|\Gamma_2(u)| = \Theta(n^2 p^2)$  and  $b(i)$  takes each of the  $np^{3/2}$  most likely values with probabilities which are  $\Theta(n^{-1} p^{-3/2})$ . If we ignore problematic vertices and assume that every vertex is an independent binomial, the number of vertices  $N_k$  in  $\Gamma_1(v)$  with a fixed likely degree  $k$  is a binomial random variable with  $\Theta(np)$  trials and success probability  $\Theta(n^{-1} p^{-3/2})$ . We also do the same thing for the edge  $xy$  to find that the probability that there are  $N_k$  vertices in  $\Gamma_1(y)$  with  $k$  edges to  $\Gamma_2(x) \setminus \Gamma_1(y)$  is  $O(p^{-1/2})$ . By considering multiple values of  $k$ , we can show that the probability that  $C_{uv}$  is isomorphic to  $C_{xy}$  is  $o(n^{-4})$ .

Unfortunately, this sketch has glossed over many details, most notably the dependencies between the different values we consider, and we will have to work considerably harder to make the argument rigorous. At various points we will see different behaviour for different values of  $p$  in the range (e.g. the number of vertices in  $\Gamma_2(u)$  is not  $\Theta(n^2 p^2)$  when  $p = \omega(n^{-1/2})$ ), and we will have to employ different arguments for different ranges of  $p$ .

We remark that our proof actually gives an efficient algorithm for reconstructing a random graph  $G \in \mathcal{G}(n, p)$  from its 2-neighbourhoods which succeeds with high probability. Instead of colouring the edge  $uv$  by the isomorphism class of  $C_{uv}$ , we can colour it by a combination of the unordered degree sequence of  $\Gamma_1(v)$  into  $\Gamma_2(u) \setminus \Gamma_1(v)$  and the unordered degree sequence of  $\Gamma_1(u)$  into  $\Gamma_2(v) \setminus \Gamma_1(u)$ . The proof of Lemma 4.2 shows that any two disjoint edges get the same colour with probability  $o(n^{-4})$ , and Lemma 4.1 applies with high probability. These degree sequences can clearly be calculated efficiently.

*Proof of Lemma 4.2.* Fix four vertices  $u, v, x$  and  $y$ , and condition on the event that  $uv$  and  $xy$  are edges. Let  $M$  be the set of vertices which are adjacent to at least 2 of the vertices in  $\{x, y, u, v\}$ . These vertices introduce dependence between the degree sequences we care about, and we will view these vertices as introducing an ‘‘error’’ of size at most  $|M|$ . We are therefore interested in an upper bound for  $|M|$ . There are 6 pairs of vertices from  $\{x, y, u, v\}$  and the probability that a vertex is adjacent to a given pair is  $p^2$ , so  $|M|$  is dominated by a  $\text{Bin}(n, 6p^2)$  random variable.

**Claim 4.3.** *Let*

$$m = \begin{cases} 12n^{1/9} & p > n^{-11/20}, \\ 40 & p \leq n^{-11/20}. \end{cases}$$

*Then*

$$\mathbb{P}(|M| > m) = o(n^{-4}).$$

*Proof.* The first case follows almost immediately from the Chernoff bound in Lemma 2.4. Indeed, since  $p \leq n^{-16/35} \leq n^{-4/9}$ ,  $|M|$  is clearly dominated by a  $\text{Bin}(n, 6n^{-8/9})$  random

variable, and the probability that this exceeds  $12n^{1/9}$  is at most  $\exp(-2n^{1/9}) = o(n^{-4})$ .

The second case follows from Lemma 2.5. In this case,  $m$  is stochastically dominated by a  $\text{Bin}(n, 6n^{-11/10})$  random variable and

$$\mathbb{P}(m \geq 41) \leq e(6n^{-1/10})^{41} = o(n^{-4}).$$

□

We now look to bound the size of the neighbourhood of a vertex.

**Claim 4.4.** *Fix a vertex  $i$ , and let*

$$\lambda(i) = (n - 1 - d(i))(1 - (1 - p)^{d(i)}).$$

*Then with probability  $1 - o(n^{-4})$  we have*

$$\frac{np}{2} \leq d(i) \leq 2np,$$

*and*

$$||\Gamma_2(i)| - \lambda(i)| \leq (np)^{5/4}.$$

*Proof.* The degree of  $i$  follows a  $\text{Bin}(n - 1, p)$  distribution so using a Chernoff bound (see Lemma 2.4), the probability that  $d(i)$  is less than  $np/2$  is at most

$$4 \exp\left(-\frac{(n - 2)^2 p}{8(n - 1)}\right) = \exp(-\Theta(np)) = o(n^{-4}).$$

In the other direction, the other bound in Lemma 2.4 shows that the probability  $d(i) \geq 2np$  is also at most  $4 \exp(-np/3) = o(n^{-4})$ .

Given  $d(i)$ , the size of the second neighbourhood of  $i$  is distributed like

$$X \sim \text{Bin}(n - 1 - d(i), 1 - (1 - p)^{d(i)}),$$

so  $\mathbb{E}[X] = \lambda(i)$ . If  $\lambda(i) = \omega(\log^8 n)$ , then

$$\mathbb{P}(|X - \lambda(i)| \geq \lambda(i)^{9/16}) \leq \exp(-\Theta(\lambda(i)^{1/8})) = o(n^{-4}).$$

Hence, it suffices to prove that with probability  $o(n^{-4})$  we have  $\lambda(i) = \omega(\log^8 n)$  and (for large enough  $n$ )  $\lambda(i)^{9/16} \leq (np)^{5/4}$ .

For the first statement, we may assume that  $np/2 \leq d(i) \leq 2np$ . Using that  $1 - t \leq e^{-t} \leq 1 - t/2$  for all  $t \in [0, 1]$ , we have

$$\begin{aligned} \lambda(i) &= (n - 1 - d(i))(1 - (1 - p)^{d(i)}) \\ &\geq \frac{n}{2}(1 - (1 - p)^{np/2}) \\ &\geq \frac{n}{2}(1 - e^{-np^2/2}) \\ &\geq \frac{n}{2} \min\{1 - e^{-1}, np^2/4\} \end{aligned}$$

for large enough  $n$ . This is  $\omega(\log n)$  in our range of  $p$ .

For the second statement, note that  $\lambda(i) \leq n(1 - (1 - p)^{2np}) \leq 2n^2 p^2$ , by Bernoulli's inequality. □

We will shortly reveal the edges from  $\Gamma_1(u)$  and from  $\Gamma_1(x)$  to discover their second neighbourhoods. Unfortunately, this may reveal some edges from  $\Gamma_1(v)$  to  $\Gamma_2(u) \setminus \Gamma_1(v)$ . For example, if  $i \in \Gamma_1(v)$ , then we will be revealing all edges from  $i$  to  $\Gamma_1(x)$ . Some of the vertices in  $\Gamma_1(x)$  may also be in  $\Gamma_2(u) \setminus \Gamma_1(v)$ , so we have revealed some of the edges from  $i$  to  $\Gamma_2(u) \setminus \Gamma_1(v)$ . We will use the following lemma to control how many edges have been revealed.

**Claim 4.5.** *Let  $i \in \{u, v, x, y\}$ . If  $n^{-11/20} \leq p \leq n^{-4/9}$ , then the probability there exists a vertex  $j \notin \{i\} \cup \Gamma_1(i)$  which is adjacent to at least  $(n^2 p^3)^{1/4}$  vertices in  $\Gamma_1(i)$  is  $o(n^{-4})$ .*

*If  $p \leq n^{-11/20}$ , then the probability there exists a vertex  $j \notin \{i\} \cup \Gamma_1(i)$  which is adjacent to at least 51 vertices in  $\Gamma_1(i)$  is  $o(n^{-4})$ .*

*Proof.* Suppose first that  $n^{-11/20} \leq p \leq n^{-4/9}$ . For a given vertex  $j$ , the number of neighbours in  $\Gamma_1(i)$  is a binomial random variable with  $d(i) = |\Gamma_1(i)|$  trials and success probability  $p$ . We may assume that  $d(i) \leq 2np$ , and applying a Chernoff bound (Lemma 2.4), we find that the probability that  $j$  is adjacent to at least  $(n^2 p^3)^{1/4}$  vertices in  $\Gamma_1(i)$  is at most

$$\exp(-\Theta(n^2 p^3)^{1/4}),$$

provided  $np^{5/2} \rightarrow 0$ . There are at most  $n$  choices for  $j$  and applying a union bound completes the proof.

To prove the second part of the claim where  $p \leq n^{-11/20}$ , we use Lemma 2.5. For a given vertex  $j$ , the number of neighbours is dominated by binomial random variable with mean  $2np^2 \leq 2n^{-1/10}$ . Hence, by Lemma 2.5, the probability that a vertex has at least 51 neighbours in  $\Gamma_1(i)$  is  $O(n^{-51/10})$ . Taking a union bound over all choices for the vertex  $j$ , the probability that any suitable  $j$  is adjacent to at least 51 vertices is  $o(n^{-4})$  as required.  $\square$

We now reveal the edges from  $u, v, x$  and  $y$ , the edges from  $\Gamma_1(u)$  and  $\Gamma_1(x)$  and the edges between the neighbours of  $u, v, x$  and  $y$ . None of the other edges need to be revealed and they are still each present independently with probability  $p$ . We also check that the following have all occurred and note that each of them occurs with probability  $1 - o(n^{-4})$ .

- $|M|$  is bounded above by  $m$ ,
- $d(u), d(v), d(x)$  and  $d(y)$  are all in  $[np/2, 2np]$ ,
- $||\Gamma_2(u)| - \lambda(u)| \leq (np^{5/4})$  and  $||\Gamma_2(x)| - \lambda(x)| \leq (np^{5/4})$ ,
- for every vertex  $a \in \Gamma_1(v)$ , the sets  $\Gamma_1(a) \cap \Gamma_1(u)$ ,  $\Gamma_1(a) \cap \Gamma_1(x)$  and  $\Gamma_1(a) \cap \Gamma_1(y)$  have size at most  $(n^2 p^3)^{1/4}$  if  $n^{-11/20} \leq p \leq n^{-4/9}$ , or 51 if  $p \leq n^{-11/20}$
- for every vertex in  $b \in \Gamma_1(y)$ , and the sets  $\Gamma_1(b) \cap \Gamma_1(u)$ ,  $\Gamma_1(b) \cap \Gamma_1(v)$  and  $\Gamma_1(b) \cap \Gamma_1(x)$  have size at most  $(n^2 p^3)^{1/4}$  if  $n^{-11/20} \leq p \leq n^{-4/9}$ , or 51 if  $p \leq n^{-11/20}$ .

If there is an isomorphism from  $C_{uv}$  to  $C_{xy}$  which maps  $u$  to  $x$ , then we must have  $d(u) = d(x)$ , and we also assume that this event occurs. This means that  $\lambda(u) = \lambda(x)$  and we denote the single quantity by  $\lambda$ .

Having assumed the above properties, we are ready to begin looking at the the number of edges from each vertex in  $\Gamma_1(v)$  to  $\Gamma_2(u) \setminus \Gamma_1(v)$  and bound the probability that this unordered degree sequence equals the one from  $\Gamma_1(y)$  to  $\Gamma_2(x) \setminus \Gamma_1(y)$ . For any  $i, j \in V(G)$ , let  $X_{i,j}$  be the indicator that the edge  $\{i, j\}$  is present in  $G$ , and let

$$A = \{x, y, u, v\} \cup \Gamma_1(u) \cup \Gamma_1(v) \cup \Gamma_1(x) \cup \Gamma_1(y).$$

For a vertex  $i \in \Gamma_1(v)$ , let  $Y_i$  be the number of edges from  $i$  to  $\Gamma_2(u) \setminus \Gamma_1(v)$ , that is

$$Y_i = \sum_{w \in \Gamma_2(u) \setminus A} X_{i,w} + \sum_{w \in (\Gamma_2(u) \setminus \Gamma_1(v)) \cap A} X_{i,w}.$$

The second term consists of (indicators for the) edges adjacent to  $u, v, x$  or  $y$  and edges between the neighbourhoods of those vertices. In particular, the second term is already known (as these edges have been revealed) and we denote it by  $\varepsilon_i$ . The assumptions we have made imply that  $\varepsilon_i \leq \varepsilon$  where we have  $\varepsilon = 3(n^2 p^3)^{1/4} + 4$  if  $n^{-11/20} \leq p \leq n^{-16/35}$  and  $\varepsilon = 157$  if  $p \leq n^{-11/20}$ . Provided that  $i \notin \{u, v, x, y\} \cup M$ , we have not revealed any of the indicator variables in the first sum, and  $Y_i - \varepsilon_i$  is a binomial random variable with  $\lambda + O((np)^{5/4})$  trials and success probability  $p$ .

Similarly, for  $j \in \Gamma_1(y)$ , let  $Y'_j$  be the number of edges from  $j$  to  $\Gamma_2(x) \setminus \Gamma_1(y)$ , that is

$$Y'_j = \sum_{w \in \Gamma_2(x) \setminus A} X_{j,w} + \sum_{w \in (\Gamma_2(x) \setminus \Gamma_1(y)) \cap A} X_{j,w},$$

and let  $\varepsilon'_j = \sum_{w \in (\Gamma_2(x) \setminus \Gamma_1(y)) \cap A} X_{j,w}$ . Define  $B_1$  and  $B_2$  by  $B_1 = \Gamma_1(v) \setminus (M \cup \{u, v, x, y\})$  and  $B_2 = \Gamma_1(y) \setminus (M \cup \{u, v, x, y\})$ , so that the random variables

$$\{Y_i - \varepsilon_i : i \in B_1\} \cup \{Y'_j - \varepsilon'_j : j \in B_2\}$$

are independent binomial random variables, each with success probability  $p$ . Indeed, if  $Y_{i_1} - \varepsilon_{i_1}$  and  $Y_{i_2} - \varepsilon_{i_2}$  ( $i_1 \neq i_2$ ) are not independent, then there must be  $w_1, w_2 \in \Gamma_2(u) \setminus A$  such that  $\{i_1, w_1\} = \{i_2, w_2\}$ . Since  $i_1 \neq i_2$ , we would have  $i_1 = w_2 \in \Gamma_2(u) \setminus A$ , but  $i_1 \in A$ . If there are  $i \in B_1$  and  $j \in B_2$  such that  $Y_i - \varepsilon_i$  and  $Y'_j - \varepsilon_j$  are not independent, there must be  $w_1 \in \Gamma_2(u) \setminus A$  and  $w_2 \in \Gamma_2(x) \setminus A$  such that  $\{i, w_1\} = \{j, w_2\}$ . Since  $i \notin M$  and  $i \in \Gamma_1(v)$ , we cannot have  $i \in \Gamma_1(y)$  and so  $i \neq j$ . This means  $i = w_2$ , but  $w_2 \in \Gamma_1(v) \subseteq A$ , a contradiction.

If  $C_{uv}$  is isomorphic to  $C_{xy}$  with  $u$  mapping to  $x$ , then the multisets  $\{Y_i : i \in \Gamma_1(v)\}$  and  $\{Y'_j : j \in \Gamma_1(y)\}$  must be equal. Equivalently, the number of  $Y_i$  and  $Y'_j$  equal to  $k$  must be equal for every choice of  $k$ . The  $Y_i$  with  $i \notin B_1$  are potentially problematic, but there are at most  $m + 4$  of them and so we ignore them and consider the multiset  $\{Y_i : i \in B_1\}$  which is “close” to the multiset  $\{Y_i : i \in \Gamma_1(v)\}$ . Likewise we can consider the multiset  $\{Y'_j : j \in B_2\}$  which is “close” to the multiset  $\{Y'_j : j \in \Gamma_1(y)\}$ . Since we have deleted at most  $m + 4$  elements from each multiset, the number of  $Y_i$  and  $Y'_j$  equal to  $k$  in the resulting multisets may differ by at most  $m + 4$ .

Let  $Z_k$  be the number of the  $Y_i$ , where  $i \in B_1$ , which are equal to  $k$  and note that  $Z_k$  is the sum of  $|B_1|$  independent Bernoulli random variables (with potentially different probabilities due to different  $\varepsilon_i$ ). Similarly, let  $Z'_k$  be the number of the  $Y'_j$ , with  $j \in B_2$  which are equal to  $k$ .

Let  $\mu = |\Gamma_2(u) \setminus A|p$  and  $\mu' = |\Gamma_2(x) \setminus A|p$ , so that  $\mathbb{E}[Y_i - \varepsilon_i] = \mu$  and  $\mathbb{E}[Y'_j - \varepsilon'_j] = \mu'$ . Since  $|A| = O(np)$  and  $\Gamma_2(u)$  and  $\Gamma_2(x)$  are both  $\lambda + O((np)^{5/4})$ , both  $\mu$  and  $\mu'$  are  $p\lambda + O(n^{5/4}p^{9/4})$ . Let  $k_i = \max\{\lceil \mu \rceil, \lceil \mu' \rceil\} + \varepsilon + i$ , and let  $\ell$  be a quantity to be determined. We will reveal the values of  $Z_{k_i}$  for  $i \in [\ell]$  and call these our *target values*. If there is an isomorphism mapping  $C_{uv}$  to  $C_{xy}$  which sends  $u$  to  $x$ , it must be the case that  $|Z_{k_i} - Z'_{k_i}| \leq m + 4$  for all  $i \in [\ell]$ , and we will iteratively bound the probability that  $|Z_{k_i} - Z'_{k_i}| \leq m + 4$ , conditional on the event that such a bound held for the values  $k_1, \dots, k_{i-1}$ . If this event does not occur, then  $C_{uv}$  and  $C_{xy}$  are not isomorphic and we are done. If the event does occur, we reveal the vertices in  $B_2$  which have  $k_i$  edges to  $\Gamma_2(x) \setminus \Gamma_1(y)$  and carry on.

We now prove a series of claims which we will use to ensure that the probability that  $|Z_{k_i} - Z'_{k_i}| \leq m + 4$  is small for every  $i$ . We start by showing that knowing that  $Y'_j$  has not already been revealed only changes the probability that it is revealed in the next step by a constant factor. We will then show that the probability that  $Y'_j$  takes a particular value  $k_i$  is small, for which we use two different approximations depending on the value of  $p$ .

**Claim 4.6.** *For any  $\ell > 0$ ,*

$$\mathbb{P}(Z_{k_1} + \dots + Z_{k_\ell} \leq 3|B_2|/4) = 1 - o(n^{-4}).$$

*Proof.* We first bound the probability that a given  $Y_i$  is in  $\{k_1, \dots, k_\ell\}$ , or equivalently, that  $Y_i - \varepsilon_i \in \{k_1 - \varepsilon_i, \dots, k_\ell - \varepsilon_i\}$ . Since  $k_1 - \varepsilon_i > \lceil \mu \rceil$ , this is clearly bounded above by the probability that  $Y_i - \varepsilon_i > \lceil \mu \rceil$ . The random variable  $Y_i - \varepsilon_i$  follows a binomial distribution and hence the median is  $\lfloor \mu \rfloor$  or  $\lceil \mu \rceil$ . This means

$$\mathbb{P}(Y_i \in \{k_1, \dots, k_\ell\}) \leq \frac{1}{2}.$$

In particular, the random variable  $Z_{k_1} + \dots + Z_{k_\ell}$  is dominated by a binomial random variable with  $|B_1| = \Theta(np)$  trials and success probability  $1/2$ . Using Lemma 2.4, the probability that such a random variable exceeds  $2|B_1|/3$  is at most  $\exp(-|B_1|/6^3) = o(n^{-4})$ . The result is now immediate since  $|B_2| = (1 + o(1))|B_1|$ .  $\square$

**Claim 4.7.** *For all  $i \in [\ell]$ ,*

$$\mathbb{P}(Y'_j = k_i) \leq \mathbb{P}(Y'_j = k_i | Y'_j \notin \{k_1, \dots, k_{i-1}\}) \leq 2\mathbb{P}(Y'_j = k_i).$$

*Proof.* The claim follows immediately from  $\mathbb{P}(Y'_j \in \{k_1, \dots, k_\ell\}) \leq \frac{1}{2}$  and

$$\mathbb{P}(Y'_j = k | Y'_j \notin \{k_1, \dots, k_{i-1}\}) = \frac{\mathbb{P}(Y'_j = k)}{1 - \mathbb{P}(Y'_j \in \{k_1, \dots, k_{i-1}\})}.$$

$\square$

We now assume that  $Z_{k_1} + \dots + Z_{k_\ell} \leq 3|B_2|/4$ . Our goal is to apply Theorem 2.6 for which we need to bound the probability that  $Y'_j = k_i$ , given that  $Y'_j \notin \{k_1, \dots, k_{i-1}\}$ . We use different approaches for different values of  $p$ , and we now split the proof into two parts.



**Claim 4.8.** *Suppose  $p = \omega(n^{-2/3})$  and  $p \leq n^{-16/35}$ . There exist constants  $\alpha, \beta > 0$  such that, for all  $j \in B_2$  and  $i \in [\sqrt{p\lambda}]$ , we have*

$$\frac{\alpha}{\sqrt{\mu'}} \leq \mathbb{P}(Y'_j = k_i) \leq \frac{\beta}{\sqrt{\mu'}}.$$

*Proof.* Note that  $\mathbb{P}(Y'_j = k_i) = \mathbb{P}(Y'_j - \varepsilon'_j = k_i - \varepsilon'_j)$  and that  $Y'_j - \varepsilon'_j$  is a binomial random variable whose variance tends to infinity. By Theorem 2.7 it is enough to show that there is a constant  $M$  such that  $|k_i - \varepsilon'_j - \mu'| \leq M\sqrt{\mu'}$  for all  $j \in B_2$  and  $k_i$ . Note that

$$\begin{aligned} |k_i - \varepsilon'_j - \mu'| &\leq |\max\{\lceil \mu \rceil, \lceil \mu' \rceil\} - \mu'| + |\varepsilon'_j| + i \\ &\leq O(n^{5/4}p^{9/4}) + \varepsilon + \sqrt{p\lambda}. \end{aligned}$$

As seen in the proof of Claim 4.4, we have  $\lambda \geq \frac{n}{2} \min\{1 - e^{-1}, np^2/4\}$  for large enough  $n$ . In particular, there are constants  $a$  and  $b$  such that  $\sqrt{\mu'} \geq \min\{a\sqrt{np}, b\sqrt{n^2p^3}\}$ . Since  $p \leq n^{-16/35} \leq n^{-3/7}$ , we have  $n^{5/4}p^{9/4} \leq \sqrt{np}$ , and  $n^{5/4}p^{9/4} \leq \sqrt{n^2p^3}$  provided  $p \leq n^{-1/3}$ . This shows  $n^{5/4}p^{9/4} = O(\sqrt{\mu'})$ . We also need to check that  $\varepsilon = O(\sqrt{\mu'})$ . Note that  $\sqrt{\mu'}$  is  $\omega(1)$ , so  $157 = O(\sqrt{\mu'})$ , and it is easy to see that  $(n^2p^3)^{1/4} = O(\sqrt{\mu'})$  as well.  $\square$

Suppose we are at stage  $i$ , and so we have already revealed the vertices with degrees  $k_1, \dots, k_{i-1}$  and are interested in the event that  $|Z_{k_i} - Z'_{k_i}| \leq m + 4$ . Since we have already revealed  $Z_{k_i}$ , it suffices to bound the probability that  $Z'_{k_i}$  takes one of the  $2m + 9$  most likely values. The random variable  $Z'_{k_i}$  is the sum of independent Bernoulli random variables, and we may apply Theorem 2.6. By Claim 4.6 there are at least  $|B_2|/4$  trials and by Claim 4.7 the success probability of each trial is at least  $\alpha/\sqrt{\mu'}$  and at most  $2\beta/\sqrt{\mu'}$ . Since  $\mu' \rightarrow \infty$  as  $n \rightarrow \infty$ , we may assume  $2\beta/\sqrt{\mu'} < 1/2$ . In particular, each unrevealed  $j \in B_2$  is equal to  $k_i$  with probability less than  $1/2$ . Applying Theorem 2.6 we have

$$\sup_x \mathbb{P}(Z'_{k_i} = x) \leq C \left( \frac{\alpha|B_2|}{4\sqrt{\mu'}} \right)^{-1/2} = O(p^{1/4}),$$

and

$$\mathbb{P}(|Z_{k_i} - Z'_{k_i}| \leq m + 4) = O(mp^{1/4}).$$

Since  $p \leq n^{-16/35}$  and  $m \leq 12n^{1/9}$ , we have  $mp^{1/4} = O(n^{-1/315})$ . The only condition on  $\ell$  in this argument comes from the application of Claim 4.8 where we required that  $\ell \leq \sqrt{p\lambda}$ . Since  $\sqrt{p\lambda} = \omega(1)$ , we may take  $\ell > 1260$  to be a constant, in which case the probability that all  $\ell$  steps succeed is  $O(n^{-\ell/315}) = o(n^{-4})$  as required.

We now consider the case where  $n^{-2/3-\delta} \leq p \leq n^{-2/3} \log \log n$ . Instead of applying a local limit theorem as in Claim 4.8, we approximate  $Y'_j - \varepsilon'_j$  by a Poisson random variable and use this to bound the probability that  $Y'_j - \varepsilon'_j$  equals  $k_i$ .

**Claim 4.9.** *Suppose  $n^{-2/3-\delta} \leq p \leq n^{-2/3} \log \log n$ . Then, for all  $i > 0$ , we have*

$$\frac{(\mu')^{k_i - \varepsilon} \exp(-\mu')}{(k_i - \varepsilon)!} + O(n^2p^4) \leq \mathbb{P}(Y'_j = k_i) \leq 1/5 + O(n^2p^4).$$

*Proof.* By the Le Cam Theorem (see Theorem 2.8), the total variation distance between  $Y'_j - \varepsilon'_j$  and a Poisson random variable with mean  $\mu'$  is at most  $2p\mu' = O(n^2p^4)$ . Hence,

$$\mathbb{P}(Y'_j = k_i) = \mathbb{P}(Y'_j - \varepsilon'_j = k_i - \varepsilon'_j) = \frac{(\mu')^{k_i - \varepsilon'_j} \exp(-\mu')}{(k_i - \varepsilon'_j)!} + O(n^2p^4).$$

The probability mass function of a Poisson distribution is decreasing above its mean, and so the right hand side is a decreasing function of  $k_i - \varepsilon'_j$ . The lower bound now follows since  $\varepsilon'_j \leq \varepsilon$ . For the upper bound, note that  $k_i - \varepsilon'_j \geq \lceil \mu' \rceil + 1$ , and it suffices to bound

$$\frac{t^{\lceil t+1 \rceil} \exp(-t)}{\lceil t+1 \rceil!}$$

over all values of  $t > 0$ . This is bounded above by  $1/5$ .  $\square$

The random variable  $Z_{k_i}$  is the sum of at least  $|B_2|/4$  independent Bernoulli random variables, each with probability at least  $(\mu')^{k_i - \varepsilon} \exp(-\mu') / (k_i - \varepsilon)! + O(n^2p^4)$  and at most  $2/5 + O(n^2p^4)$ . Hence, by Theorem 2.6,

$$\sup_t \mathbb{P}(Z'_{k_i} = t) \leq C \left( \frac{|B_2|}{4} \cdot \frac{(\mu')^{k_i - \varepsilon} \exp(-\mu')}{(k_i - \varepsilon)!} + O(n^3p^5) \right)^{-1/2}. \quad (4.1)$$

Note that  $t^t \exp(-t)$  is bounded below by  $1/e$  and that  $\mu' = p\lambda + O(n^{5/4}p^{9/4})$ . Since  $\lambda \leq 2n^2p^2$ , we may assume  $\mu' \leq 3(\log \log n)^3$  for large enough  $n$  (provided  $\delta < 1/4$ ). Hence, for large enough  $n$ ,

$$\begin{aligned} \frac{|B_2|}{4} \cdot \frac{(\mu')^{k_i - \varepsilon} \exp(-\mu')}{(k_i - \varepsilon)!} &= \frac{|B_2|}{4} \cdot (\mu')^{\mu'} \exp(-\mu') \cdot \frac{(\mu')^{i + \lceil \mu' \rceil - \mu'}}{(\lceil \mu' \rceil + i)!} \\ &\geq \frac{|B_2|}{4e} \cdot \frac{n^{-3\delta(i+1)}}{(3(\log \log n)^3 + \ell + 1)!}. \end{aligned}$$

For any fixed  $\ell$  and  $\delta$ , the quantity  $(3(\log \log n)^3 + \ell + 1)!$  is less than  $n^{-3\delta}$  for large  $n$ . We also have that  $|B_2| \geq np/2 - (m+4) \geq n^{1/3-\delta}/3$  for large  $n$ . Hence,

$$\frac{|B_2|}{4} \cdot \frac{(\mu')^{k_i - \varepsilon} \exp(-\mu')}{(k_i - \varepsilon)!} \geq \frac{n^{1/3-\delta-3\delta(i+2)}}{12e}.$$

Substituting this bound into (4.1) gives

$$\sup_t \mathbb{P}(Z'_{k_i} = t) \leq C \left( \frac{n^{1/3-\delta-3\delta(i+2)}}{12e} + O(n^3p^5) \right)^{-1/2}.$$

Hence, the probability that all  $\ell$  steps complete is at most

$$\prod_{i=1}^{\ell} (2m+4) C \left( \frac{n^{1/3-\delta-3\delta(i+2)}}{12e} + O(n^3p^5) \right)^{-1/2} = O(n^{-(\ell/6 - 7\ell\delta/2 - 3\delta\ell(\ell+1)/4)}).$$

For any  $\ell > 24$ , one can choose  $\delta$  sufficiently small such that

$$\ell/6 - 7\ell\delta/2 - 3\delta\ell(\ell+1)/4 > 4$$

which completes the proof.  $\square$

## 5 Non-reconstructibility from 1-neighbourhoods and 2-neighbourhoods

In this section we prove Theorem 1.4 and Theorem 1.6. The proofs are quite similar, but differ in the technical details. We start in Section 5.1 with the proof of Theorem 1.6 since it is slightly simpler, and then we move on to the proof of Theorem 1.4 in Section 5.2.

### 5.1 1-neighbourhoods

In this subsection we prove Theorem 1.6. When  $p = O\left(\frac{\log n}{n}\right)$  and  $p = \omega(n^{-3/2})$ , we can appeal directly to Lemma 3.3. It is therefore sufficient to show that if  $p \leq \sqrt{\frac{\log n}{25n}}$  and  $p = \omega(n^{-1})$ , a random graph  $G \in \mathcal{G}(n, p)$  is not 1-reconstructible with high probability.

*Proof.* Suppose that  $p = \omega(n^{-1})$  and  $p \leq c\sqrt{\frac{\log n}{n}}$  for some small constant  $c > 0$  (which we will later take to be  $1/5$ ). We will show that with high probability, there exist four vertices  $u, v, x, y \in V(G)$  such that

1. the pairs  $xy, uv \in E(G)$ , and  $xu, xv, yu, yv \notin E(G)$ ,
2. all the degrees  $d(u), d(v), d(x), d(y)$  are different,
3. the degrees  $d(u), d(v), d(x), d(y)$  are at most  $(np)^{2/3}$  from  $np$ , and
4. the neighbourhoods  $\Gamma(u), \Gamma(v), \Gamma(x)$  and  $\Gamma(y)$  are all pairwise disjoint.

It is straightforward to see that this implies that the graph  $G$  is not reconstructible from its 1-neighbourhoods. Indeed, the graphs  $G$  and  $G' = (G \setminus \{xy, uv\}) \cup \{xu, yv\}$  have the same collection of 1-neighbourhoods, but they are not isomorphic as there is one fewer edge between vertices of degree  $d(x)$  and  $d(y)$  in  $G'$  than in  $G$ .

It thus remains to prove that there exist four such vertices with high probability. Let  $A = (a_1, a_2, a_3, a_4) \subseteq V(G)$  be an ordered tuple of four vertices, and let  $X_A$  be the indicator of the event that the vertices of  $A$  satisfy the conditions above with  $a_1 = u, a_2 = v, a_3 = x$  and  $a_4 = y$ . Let  $X = \sum_{A \subseteq V} X_A$  be the total number of such ‘good’ tuples. Then  $\mathbb{E}[X] = \sum_{A \subseteq V} \mathbb{E}[X_A] = 4! \binom{n}{4} \mathbb{P}(X_{(1,2,3,4)} = 1)$ . Let  $R_1, R_2, R_3$  and  $R_4$  be the events that  $(1, 2, 3, 4)$  satisfies the conditions 1, 2, 3 and 4 respectively. The probability of the event  $R_1$  is simply  $p^2(1-p)^4$ . Given that  $R_1$  occurs, the degree of a vertex in  $A$  is distributed like a  $\text{Bin}(n-4, p)$  random variable plus one. The degrees are independent so the probability that two of the vertices have the same degree is at most 6 times the probability that two  $\text{Bin}(n-4, p)$  random variables are equal, and this is  $o(1)$  by Theorem 2.6. Further, an application of Lemma 2.4 shows that  $\mathbb{P}(R_3^c \mid R_1) = o(1)$ , and hence,  $\mathbb{P}(R_2 \cap R_3 \mid R_1) = 1 - o(1)$ .

We now consider  $R_4$ . Given  $n'$  and  $a$  with  $|n' - n| \leq 8$  and  $|a - np| \leq (np)^{2/3}$ , the probability that four uniformly chosen sets from  $[n']$  of size  $a$  are pairwise disjoint is

$$\frac{\binom{n'}{a} \binom{n'-a}{a} \binom{n'-2a}{a} \binom{n'-3a}{a}}{\binom{n'}{a}^4} = (1 - o(1))e^{-6a^2/n} = (1 - o(1))e^{-6np^2}. \quad (5.1)$$

Given  $R_1, R_2$  and  $R_3$  the probability that  $R_4$  occurs can be bounded above by the probability that four uniformly chosen sets from  $[n]$  of size  $np - (np)^{2/3}$  are pairwise disjoint, and bounded below by the probability that four uniformly chosen sets from  $[n-4]$  of size  $np - (np)^{2/3}$  are pairwise disjoint. By (5.1) both probabilities are  $(1 - o(1))e^{-6np^2}$ .

Combining the above we have  $\mathbb{P}(X_A) = (1 - o(1))p^2 \exp(-6np^2)$ , and so

$$\mathbb{E}[X] = (1 + o(1))n^4 p^2 \exp(-6np^2) = \Omega(n^{2-6c^2}). \quad (5.2)$$

We next show that  $\mathbb{E}[X^2] \leq (1 + o(1))\mathbb{E}[X]^2$ , so that  $\text{Var}(X) = o(\mathbb{E}[X]^2)$  and Chebyshev's inequality completes the proof. Write

$$\mathbb{E}[X^2] = \sum_{A_1, A_2} \mathbb{E}[X_{A_1} X_{A_2}] = \sum_{k=0}^4 \sum_{\substack{A_1, A_2, \\ |A_1 \cap A_2|=k}} \mathbb{P}((X_{A_1} = 1) \wedge (X_{A_2} = 1)).$$

We first consider when  $A_1$  and  $A_2$  intersect (with  $|A_1 \cup A_2| = 8 - k$ ). If both  $A_1$  and  $A_2$  satisfy condition 1, then there are at least  $4 - k/2$  edges which must each be present. This happens with probability at most  $p^{4-k/2}$ . Hence, summing over the at most  $n^{8-k}$  choices for  $A_1$  and  $A_2$  for each  $k$  and noting that  $n^2 p = \omega(1)$ , we have

$$\sum_{k=1}^4 \sum_{\substack{A_1, A_2, \\ |A_1 \cap A_2|=k}} \mathbb{P}((X_{A_1} = 1) \wedge (X_{A_2} = 1)) \leq \sum_{k=1}^4 n^{8-k} p^{4-k/2} \leq 4n^7 p^{7/2}.$$

Considering (5.2), we see that for small enough  $c$  this sum is  $o(\mathbb{E}[X]^2)$ . Indeed,  $n^7 p^{7/2} = O(n^{15/2} p^4)$  while  $\mathbb{E}[X]^2 = \Omega(n^{8-12c^2} p^4)$ , and it suffices to take  $c = 1/5$ . It therefore suffices to show that the sum over the choices of  $A_1$  and  $A_2$  with no intersection contributes at most  $(1 + o(1))\mathbb{E}[X]^2$ .

Now suppose that there is no intersection between  $A_1$  and  $A_2$ . We loosen the requirements given by 1, 2, 3 and 4, by ignoring the edges between  $A_1$  and  $A_2$ , and ignoring condition 2. Condition 1 is unchanged, and condition 4 is weaker as we allow the neighbourhoods to intersect in  $A_1$  and  $A_2$ . We modify condition 3 so that the degree of each vertex is at most  $(np)^{2/3} + 4$  away from  $np$  ignoring any edges between  $A_1$  and  $A_2$ , and note that this has a negligible difference on the probability. Let  $X'_{A_1, A_2}$  be the indicator of the event that both  $A_1$  and  $A_2$  pass these conditions which, since we have weakened the conditions, dominates the event that  $X_{A_1} = 1$  and  $X_{A_2} = 1$ . Repeating the calculation from before shows that  $\mathbb{P}(X'_{A_1, A_2} = 1) = (1 + o(1))\mathbb{P}(X_{(1,2,3,4)} = 1)^2$ . It then follows that  $\sum_{A_1 \subseteq V} \sum_{A_2 \subseteq V \setminus A_1} \mathbb{P}((X_{A_1} = 1) \wedge (X_{A_2} = 1)) \leq (\sum_{A \subseteq V} (1 + o(1))\mathbb{P}(X_A = 1))^2 = (1 + o(1))\mathbb{E}[X]^2$ , as required.  $\square$

## 5.2 2-neighbourhoods

In this subsection we prove Theorem 1.4. When  $p = O(\frac{\log n}{n})$  and  $p = \omega(n^{-5/4})$ , we can appeal directly to Lemma 3.3, so it suffices to consider  $p$  where  $p \leq \frac{1}{3} \left( \frac{\log^{1/3} n}{n} \right)^{3/4}$  and  $p = \omega(n^{-1} \log \log n)$ . We will show that for such  $p$  a random graph  $G \in \mathcal{G}(n, p)$  is not 2-reconstructible with high probability.

*Proof of Theorem 1.4.* Suppose that  $p = \omega(n^{-1} \log \log n)$  and  $p \leq c \left( \frac{\log^{1/3} n}{n} \right)^{3/4}$  for some small constant  $c > 0$  (which we will later take to be  $1/3$ ). For 2 vertices  $i \sim j$ , define the ‘one-sided 2-neighbourhood’ of  $i$  with respect to  $ij$  to be  $N_2^{ij}(i) = (\Gamma_1(i) \setminus \{j\}) \cup (\Gamma_2(i) \setminus \Gamma_1(j))$ . We will show that with high probability, there exist four vertices  $u, v, x, y \in V(G)$  such that

1. the pairs  $xy, uv \in E(G)$ , and  $xu, xv, yu, yx \notin E(G)$ ,
2.  $d(x) = d(v)$  and  $d(y) = d(u)$ ,
3. the degrees  $d(u), d(v), d(x), d(y)$  are at most  $(np)^{2/3}$  from  $np$ ,
4. the sizes of the second neighbourhoods:  $|\Gamma_2(x)|, |\Gamma_2(y)|, |\Gamma_2(u)|, |\Gamma_2(v)|$ , are all different,
5. the sizes of the second neighbourhoods  $|\Gamma_2(x)|, |\Gamma_2(y)|, |\Gamma_2(u)|, |\Gamma_2(v)|$  are at most  $(n^2 p^2)^{2/3}$  from  $n^2 p^2$ ,
6. the graphs induced by the first neighbourhoods are all empty:  
 $G[\Gamma(x)], G[\Gamma(y)], G[\Gamma(u)], G[\Gamma(v)]$  contain no edges,
7. The one-sided 2-neighbourhoods  $N_2^{xy}(x), N_2^{xy}(y), N_2^{uv}(v)$  and  $N_2^{uv}(u)$  are disjoint.

It is straightforward to see that this implies that the graph  $G$  is not reconstructible from its 2-neighbourhoods. Indeed, conditions 1, 2, 6 and 7 ensure the graphs  $G$  and  $G' = (G \setminus \{xy, uv\}) \cup \{xu, yv\}$  have the same collection of 2-neighbourhoods, but the number of edges  $ij$  where  $|\Gamma_2(i)| = |\Gamma_2(x)|$  and  $|\Gamma_2(j)| = |\Gamma_2(y)|$  (or the other way round) is one less in  $G'$ .

It thus remains to prove that there exist four such vertices with high probability. Let  $A = (a_1, a_2, a_3, a_4) \subseteq V(G)$ , and let  $X_A$  be the event that the vertices of  $A$  satisfy the conditions above with  $a_1 = u, a_2 = v, a_3 = x, a_4 = y$ . Let  $X = \sum_{A \subseteq V} X_A$  be the total number of such ‘good’ pairs. Then  $\mathbb{E}[X] = \sum_{A \subseteq V} \mathbb{E}[X_A] = 4! \binom{n}{4} \mathbb{P}(X_{(1,2,3,4)} = 1)$ . For  $i \in [7]$ , let  $R_i$  be the event that  $(1, 2, 3, 4)$  satisfies the condition  $i$  above. The probability of the event  $R_1$  is simply  $p^2(1-p)^4$ . Further, an application of Lemma 2.4 gives  $\mathbb{P}(R_3^c | R_1) = o(1)$ . Given that  $R_1$  occurs, the degree of a vertex in  $A$  is distributed like a  $\text{Bin}(n-4, p)$  random variable plus one. Given  $R_1$ , the degrees  $d(u), d(v), d(x)$  and  $d(y)$  are all independent so, since  $p = \omega(n^{-1})$ , an application of Theorem 2.7 shows that  $\mathbb{P}(R_2 | R_1) = \Theta(\frac{1}{np})$ .

Now reveal the edges between  $u, v, x$  and  $y$  and the degrees  $d(u), d(v), d(x)$  and  $d(y)$ , and assume that  $R_1, R_2$  and  $R_3$  hold. Then  $|\Gamma_2(x)|$  is distributed like a  $\text{Bin}(n-2-d(x)-d(y), 1-(1-p)^{d(x)-1})$  random variable plus  $d(y)-1$ . Hence, the probability that  $|\Gamma_2(x)| = |\Gamma_2(y)|$  (given  $R_1, R_2$  and  $R_3$ ) is  $O(1/\sqrt{np})$ , and it follows that the probability of  $R_4$  is  $1 - o(1)$ . Applying Lemma 2.4 also shows that the probability that  $R_5$  holds is  $1 - o(1)$ .

We are left with  $R_6$  and  $R_7$ . For them to hold, we first consider the probability that  $G[\Gamma(x)], G[\Gamma(y)], G[\Gamma(u)], G[\Gamma(v)]$  are all disjoint and empty, and then the probability that the second neighbourhoods are disjoint, and also disjoint from the first neighbourhoods.

Given  $n'$  and  $a'$  with  $|n' - n| \leq 8$  and  $|a' - np| \leq (np)^{2/3}$ , the probability that four uniformly chosen sets from  $[n']$  of size  $a'$  are pairwise disjoint is

$$\frac{\binom{n'}{a'} \binom{n'-a'}{a'} \binom{n'-2a'}{a'} \binom{n'-3a'}{a'}}{\binom{n'}{a'}^4} = (1 - o(1))e^{-6a'^2/n} = (1 - o(1))e^{-6np^2} = 1 - o(1). \quad (5.3)$$

Given conditions 1 through 5 hold, the probability that  $G[\Gamma(x)], G[\Gamma(y)], G[\Gamma(u)], G[\Gamma(v)]$  are disjoint can be bounded above by the probability that four uniformly chosen sets from  $[n]$  of size  $np - (np)^{2/3}$  are pairwise disjoint, and bounded below by the probability that four uniformly chosen sets from  $[n - 4]$  of size  $np + (np)^{2/3}$  are pairwise disjoint. By (5.3) this is  $(1 - o(1))$ . The probability that they are all empty (given that they are disjoint, and given  $R_1, R_2, R_3, R_4, R_5$ ) is bounded from below by  $1 - 4\mathbb{P}(\text{Bin}(\binom{\hat{d}}{2}, p) > 0)$ , where  $\hat{d} = np + (np)^{2/3}$ . Since  $\mathbb{E}[\text{Bin}(\binom{\hat{d}}{2}, p)] = o(1)$  for our range of  $p$ , we obtain that the conditioned probability is  $(1 - o(1))$  by applying Markov's inequality. Finally, to complete  $R_7$ , note again that the probability that four uniformly chosen sets of size  $a = n^2p^2 + O((n^2p^2)^{2/3})$  chosen from  $[n']$  where  $|n' - n| = O(np)$  are pairwise disjoint is

$$\frac{\binom{n'}{a} \binom{n'-a}{a} \binom{n'-2a}{a} \binom{n'-3a}{a}}{\binom{n'}{a}^4} = (1 - o(1))e^{-6a^2/n} = (1 - o(1))e^{-6n^3p^4}. \quad (5.4)$$

Given  $R_1, R_2, R_3, R_4, R_5$  and that the first neighbourhoods are disjoint and empty, the probability that the second neighbourhoods are disjoint can be bounded above by the probability that four uniformly chosen sets from  $[n]$  of size  $n^2p^2 - (n^2p^2)^{2/3}$  are pairwise disjoint, and bounded below by the probability that four uniformly chosen sets from  $[n']$  of size  $n^2p^2 + (n^2p^2)^{2/3}$  are pairwise disjoint, where  $n' = n - 4 - 4np - 4(np)^{2/3}$ . By (5.4) this is  $(1 - o(1))e^{-6n^3p^4}$ .

Combining the above gives that  $\mathbb{E}[X] = \Theta(n^3p \exp(-6n^3p^4))$ .

We next show that  $\mathbb{E}[X^2] \leq (1 + o(1))\mathbb{E}[X]^2$ , so that  $\text{Var}(X) = o(\mathbb{E}[X]^2)$  and Chebyshev's inequality completes the proof. As before,

$$\mathbb{E}[X^2] = \sum_{k=0}^4 \sum_{\substack{A_1, A_2, \\ |A_1 \cap A_2|=k}} \mathbb{P}((X_{A_1} = 1) \wedge (X_{A_2} = 1)),$$

and we first consider when  $A_1$  and  $A_2$  intersect (with  $|A_1 \cup A_2| = 8 - k$ ). For condition 1 to be satisfied for both  $A_1$  and  $A_2$ , there are at least  $4 - k/2$  edges which must each be present and this happens with probability at most  $p^{4-k/2}$ . Summing over the at most  $n^{8-k}$  choices for  $A_1$  and  $A_2$  for each  $k$  we have

$$\sum_{k=1}^4 \sum_{\substack{A_1, A_2, \\ |A_1 \cap A_2|=k}} \mathbb{P}((X_{A_1} = 1) \wedge (X_{A_2} = 1)) \leq 4n^7p^{7/2} \leq 4c^{3/2}n^{47/8} \log^{3/8} n \cdot p^2.$$

We have  $\mathbb{E}[X]^2 = \Omega(n^{6-12c^4}p^2)$ , so for  $c = 1/3$  the sum over the  $A_1$  and  $A_2$  that intersect is  $o(\mathbb{E}[X]^2)$ . It therefore suffices to show that the sum over the instances of  $A_1$  and  $A_2$  with no intersection contributes at most  $(1 + o(1))\mathbb{E}[X]^2$ .

As in the proof of Theorem 1.6, we count the disjoint pairs of tuples  $(a_1, a_2, a_3, a_4)$  and  $(a'_1, a'_2, a'_3, a'_4)$  which satisfy slightly weaker conditions. Again, these make a negligible difference to the calculations above, and we find that the expected number of pairs of tuples is  $(1 + O(1))\mathbb{E}[X^2]$ , but we omit the details.  $\square$

## 6 Properties of random graphs

The aim of this section is to prove the claims from Section 3, and in doing so complete the proofs of Theorem 1.1 and Theorem 1.2.

We prove several lemmas concerning the uniqueness of  $r$ -balls. In Section 6.1 we show that for appropriate values of  $p$ , the 2-balls of a random graph  $G \in \mathcal{G}(n, p)$  are typically unique, proving Lemma 3.5. Then, in Section 6.2 we show that the 3-balls of vertices of large degree are all unique (again, for appropriate values of  $p$ ), proving Lemma 3.6. In Section 6.3 we consider when we can swap two edges, keeping the set of 3-balls in the graph unchanged, proving Lemma 3.7, and thus completing the proof for non-reconstructibility from 3-neighbourhoods.

### 6.1 Uniqueness of 2-balls

In this section, we prove Lemma 3.5 which gives a region for  $p$  for which the 2-balls of a random graph  $G \in \mathcal{G}(n, p)$  are all distinct with high probability. We build on the argument of Mossel and Ross in [32] and extend their result to smaller values of  $p$ . We take a similar approach and we will also show that in  $\mathcal{G}(n, p)$ , with high probability, the degree sequences  $(\{d(w)\}_{w \in \Gamma(v)})_{v \in [n]}$  are distinct.

For a vertex  $v$ , let us denote the multiset of the degrees of the neighbours of  $v$  by  $D(v) = \{d(w)\}_{w \in \Gamma(v)}$ .

*Proof of Lemma 3.5.* Suppose

$$\zeta^2 \frac{\log^2 n}{n(\log \log n)^3} \leq p \leq n^{-2/3-\varepsilon}$$

for some large  $\zeta$  that we will fix later. Note that we can assume  $\varepsilon < 1/3$ , as a larger  $\varepsilon$  only makes the problem easier. We show that for each pair of vertices  $x, y$ , the event  $D(x) = D(y)$  occurs with probability  $o(n^{-2})$ . Taking a union bound over the  $x, y$ , shows that  $\mathcal{G}(n, p)$  has unique 2-neighbourhoods with high probability.

Fix vertices  $x, y$ . We first reveal the set  $A$  of vertices adjacent to at least one of  $x$  and  $y$  excluding  $x$  and  $y$  themselves, i.e.  $A = (\Gamma(x) \cup \Gamma(y)) \setminus \{x, y\}$ . So each vertex  $u \in V \setminus \{x, y\}$  is in  $A$  independently with probability  $1 - (1 - p)^2$ . Note that we do not yet reveal the set of edges between  $\{x, y\}$  and  $A$ , just that each vertex in  $A$  has at least one neighbour in  $\{x, y\}$ .

Next we reveal the vertices in  $A$  adjacent to both  $x$  and  $y$ , and the edges inside  $A$ . That is, for each vertex in  $A$  we connect it to both  $x$  and  $y$  with probability  $p^2/(1 - (1 - p)^2)$ , while each edge inside  $A$  is present independently with probability  $p$ .

We discount some low-probability events via the following claims.

**Claim 6.1.** *Let  $R_1$  be the event  $\{np/2 \leq |A| \leq 3np\}$ . Then  $\mathbb{P}(R_1) = 1 - o(n^{-2})$ .*

**Claim 6.2.** *The following hold.*

(i) Let  $R_3$  be the event  $\{|\Gamma(x) \cap \Gamma(y)| \leq 6\}$ . Then  $\mathbb{P}(R_3) = 1 - o(n^{-2})$ .

(ii) Let  $R_4$  be the event that there are at most  $1/\varepsilon$  edges inside  $A$ . Then  $\mathbb{P}(R_4 \mid R_1) = 1 - o(n^{-2})$ .

Note that independently each vertex in  $A$  which is not adjacent to both  $x$  and  $y$ , is connected to  $x$  with probability  $1/2$  and otherwise it is connected to  $y$  (though we do not yet reveal the adjacencies). Next we reveal every edge which is not incident with  $x$  or  $y$ . For all  $k \in \mathbb{N}$  such that  $|k - np| \leq \frac{1}{4}\sqrt{np \log(np)}$  define  $A_k$  by

$$A_k = \{z \in A : |\Gamma(z) \setminus (A \cup \{x, y\})| = k\}.$$

That is  $A_k$  is the set of vertices which have  $k$  neighbours in the rest of the graph. We would like to think of vertices in  $A_k$  as the vertices in  $A$  with degree exactly  $k+1$ , although this is not quite correct since there are vertices which are connected to both  $x$  and  $y$  and to other vertices in  $A$ . We will therefore only consider a subset of the possible values for  $k$ , and we will make sure to choose only  $k$  for which  $A_k$  is definitely the vertices in  $A$  of degree exactly  $k+1$ . When  $D(x) = D(y)$ , the vertices  $x$  and  $y$  must have the same number of the neighbours of degree  $k+1$ . If we are sure that  $A_k$  is exactly the vertices in  $A$  of degree  $k+1$ , the vertices in  $A_k$  must be evenly split between being neighbours of  $x$  and neighbours  $y$ , and this is unlikely to occur if  $A_k$  is “large”.

For each  $k$ , we say that  $A_k$  is *large* if  $|A_k| \geq (np)^{1/4}$ , and we say that  $A_k$  is *small* otherwise. We claim that most  $A_k$  are large, and we will ignore the small  $A_k$ .

**Claim 6.3.** *Let  $R_2$  be the event  $\{\#\{\text{small } A_k\} \leq (np)^{1/4}\}$ . Then  $\mathbb{P}(R_2 \mid R_1) = 1 - o(n^{-2})$ .*

Suppose  $v \in A_k$ . Then  $v$  has degree at least  $k+1$ , but it may be higher:  $v$  might be a neighbour of both  $x$  and  $y$  which would increase the degree by 1 (over the minimum); there are also at most  $1/\varepsilon$  edges between vertices of  $A$  with high probability, and they could all be incident to  $v$ , further increasing the degree by  $1/\varepsilon$ . In particular, the degree of  $v$  is  $k+1$  if none of these “bad” events occur, but could be as high as  $k+2+1/\varepsilon$ . This motivates the following definition of a good  $A_k$ .

We say that a large  $A_k$  is *good* if for all  $s$  such that  $|s - k| \leq 2/\varepsilon$  the following hold.

1. Each  $z \in A_s$  is connected to exactly one of  $x$  and  $y$ .
2. Each  $z \in A_s$  has no neighbours in  $A$ , i.e.  $\Gamma(z) \cap A = \emptyset$ .

We otherwise say that  $A_k$  is *bad*. We wish to show that there are many good  $A_k$ .

Suppose that  $R_i$  holds for  $i = 1, \dots, 4$ . We claim we have few bad  $A_k$ . Indeed, we have at most  $(np)^{1/4}$  small  $A_k$ . Each vertex in  $\Gamma(x) \cap \Gamma(y)$  causes at most  $3/\varepsilon$  sets  $A_k$  to fail condition (1), so altogether the (at most 6) vertices in  $\Gamma(x) \cap \Gamma(y)$  cause at most  $18/\varepsilon$  bad  $A_k$ . Similarly each edge inside  $A$  causes at most  $6/\varepsilon$  (doubled for each end of the edge)  $A_k$  to fail condition (2), and these edges cause at most  $6/\varepsilon^2$  bad  $A_k$ . Altogether we have  $O((np)^{1/4})$  bad  $A_k$ , and so we have at least  $\frac{1}{3}\sqrt{np \log(np)}$  good  $A_k$  for sufficiently large  $n$ .

Recall that when  $D(x) = D(y)$ , for each good  $A_k$  we must have  $|A_k \cap \Gamma(y)| = |A_k \cap \Gamma(x)|$ . Each vertex in a good  $A_k$  is adjacent to  $x$  with probability  $1/2$  and otherwise adjacent to  $y$ , and so, independently for each good  $A_k$ , the quantity  $|A_k \cap \Gamma(x)|$  is



distributed like a  $\text{Bin}(|A_k|, 1/2)$  random variable. For every  $m \geq 1$ , the probability that a  $\text{Bin}(m, 1/2)$  random variable takes the value  $m/2$ , is at most  $1/\sqrt{m}$ . Hence, for every good  $A_k$ , the probability exactly half of the vertices in  $A_k$  are connected to  $x$  (and half to  $y$ ) is at most  $(np)^{-1/8}$ . Assuming that the events  $R_1, R_2, R_3$  and  $R_4$  all happen, the number of good  $A_k$  is at least  $\frac{1}{3}\sqrt{np \log(np)}$ . This means that

$$\mathbb{P}(D(x) = D(y) | R_1, \dots, R_4) \leq \left( \frac{1}{(np)^{1/8}} \right)^{\frac{1}{3}\sqrt{np \log(np)}} = \exp\left(-\frac{1}{24}\sqrt{np} \log^{3/2} n\right).$$

Since  $p \geq \zeta^2 \frac{\log^2 n}{n(\log \log n)^3}$ , this is at most  $\exp(-\frac{1}{6}\zeta \log n)$  for large enough  $n$ , and this is  $o(n^{-2})$  for large enough  $\zeta$ . By Claims 6.1, 6.2 and 6.3 the probability that any of  $R_1, R_2, R_3$  and  $R_4$  do not hold is also  $o(n^{-2})$ , and this proves the result for  $\beta = \zeta^2$ .  $\square$

It remains to prove the claims.

*Proof of Claim 6.1.* First, note that  $d(x) - 1 \leq |A| \leq d(x) + d(y)$ , so it suffices to bound  $d(x)$  and  $d(y)$ . Using Lemma 2.4 we have

$$\mathbb{P}(d(x) - 1 \leq np/2) \leq \exp(-(1 + o(1))np/8) = o(n^{-2}),$$

which proves the first inequality. For the second inequality, note that at least one of  $d(x)$  and  $d(y)$  must be at least  $3np/2$ , and we can again use Lemma 2.4 to bound this as follows.

$$\mathbb{P}(d(x) + d(y) \geq 3np) \leq 2\mathbb{P}(d(x) \geq 3np/2) \leq \exp(-np/10) = o(n^{-2}). \quad \square$$

*Proof of Claim 6.2.* (i) Note that independently each  $z \neq x, y$  is connected to both  $x$  and  $y$  with probability  $p^2$ . Thus,  $|\Gamma(x) \cap \Gamma(y)|$  is distributed like a  $\text{Bin}(n - 2, p^2)$  random variable and

$$\mathbb{P}(|\Gamma(x) \cap \Gamma(y)| \geq 6) \leq en^6 p^{12} = O(n^{-2-12\varepsilon}) = o(n^{-2}).$$

(ii) Conditional on  $R_1$ , the number of edges inside  $A$  is stochastically dominated by a  $\text{Bin}(6(np)^2, p)$  random variable, and

$$\mathbb{P}(\text{Bin}(6(np)^2, p) \geq 2/\varepsilon) \leq e(6n^2 p^3)^{1/\varepsilon} \leq e(6n^{-3\varepsilon})^{1/\varepsilon} = o(n^{-2}). \quad \square$$

*Proof of Claim 6.3.* For each  $z \in A$ , define  $d'(z) = |\Gamma(z) \setminus (A \cup \{x, y\})|$ . Conditionally given  $|A|$ , the  $d'(z)$  are distributed like independent  $\text{Bin}(n - (|A| + 2), p)$  random variables. Hence, for  $r \in \mathbb{N}$  such that  $|r - np| \leq \frac{1}{4}\sqrt{np \log(np)}$  and  $m \in [np/2, 3np]$ , Theorem 2.7

gives

$$\begin{aligned}
\mathbb{P}(d'(z) = r \mid |A| = m) &= \mathbb{P}(\text{Bin}(n - m - 2, p) = r) \\
&\geq (1 + o(1)) \frac{1}{\sqrt{2\pi np}} \exp\left(-\frac{\left(\frac{1}{4}\sqrt{np \log(np)} + 4np^2\right)^2}{2(1-p)(np - 4np^2)}\right) \\
&= (1 + o(1)) \frac{1}{\sqrt{2\pi np}} \exp\left(-(1 + o(1)) \frac{\log(np)}{32}\right) \\
&= \frac{1}{\sqrt{2\pi}} (np)^{-\frac{1+o(1)}{32} - \frac{1}{2}}.
\end{aligned}$$

For large enough  $n$ , this is certainly at least  $(np)^{-5/8}$ .

Given  $|A| = m \in [np/2, 3np]$ , each  $|A_r|$  stochastically dominates a  $\text{Bin}(m, (np)^{-5/8})$  random variable and, for any values  $r_1, \dots, r_k$  such that  $|r_i - np| \leq \frac{1}{4}\sqrt{np \log(np)}$  for all  $i \in [k]$ , the number of vertices in  $A_{r_1} \cup \dots \cup A_{r_k}$  dominates a  $\text{Bin}(m, k(np)^{-5/8})$  random variable. If  $A_{r_1}, \dots, A_{r_k}$  are all small, then this set contains at most  $k(np)^{1/4}$  vertices and, by Lemma 2.4, we have

$$\mathbb{P}(\text{Bin}(m, k(np)^{-5/8}) \leq k(np)^{1/4}) \leq \exp\left(-\frac{(1 - 2(np)^{-1/8})^2 k(np)^{3/8}}{4}\right).$$

Hence, for large enough  $n$ , the probability that there are at least  $(np)^{1/4}$  small  $A_r$  is at most

$$\left(\sqrt{np \log(np)}\right)^{(np)^{1/4} + 1} \exp(-(np)^{5/8}/8) = o(n^{-2}).$$

□

## 6.2 Uniqueness of 3-balls

We next turn to the proof of Claim 3.6. Recall that  $\frac{\log^{2/3} n}{n} \leq p \leq \frac{\log^2 n}{n}$ , and we aim to show that with high probability the 3-balls around vertices with degree at least  $np/2$  are unique. This is done by considering the degree sequences of the neighbours of a vertex. That is, for a vertex  $x$  we consider the collection of multisets of the form  $\{d(w) : w \in \Gamma(u) \setminus \{x\}\}$ , for each neighbour  $u$  of  $x$ . Given two vertices  $x$  and  $y$ , it would be nice to appeal to a level of independence and assume the degrees of vertices at distance 2 from  $x$  or  $y$  are i.i.d. binomial random variables. Therefore, our first step in the proof is to restrict ourselves to parts of the 2-balls around  $x$  and  $y$  which do not interact or overlap, so that we may assume this independence. We then bound the probability of two multisets of i.i.d. binomial random variables being equal, and finally pull everything together and appeal to a union bound over pairs of vertices  $x$  and  $y$ .

*Proof of Claim 3.6.* Let  $G \in \mathcal{G}(n, p)$ , and fix two vertices  $x, y \in V(G)$ . Suppose that  $d = d(x) = d(y)$ , and denote the neighbourhoods of  $x$  and  $y$  by  $\{u_1, \dots, u_d\}$  and  $\{v_1, \dots, v_d\}$  respectively. For a vertex  $w \in V(G)$ , let  $D(w)$  be the multiset of the degrees of the neighbours of  $w$ , that is,  $D(w) = \{d(z) : z \in \Gamma(w)\}$ . Let  $\mathcal{D}_x = \{D(u_i) : i \in [d]\}$ , and

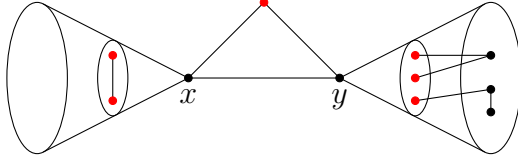


Figure 2: An edge  $uv$  with examples of vertices failing the conditions 2, 3 and 4 shown in red.

$\mathcal{D}_y = \{D(v_i) : i \in [d]\}$ . Clearly, if the 3-balls around  $x$  and  $y$  are isomorphic, then  $\mathcal{D}_x = \mathcal{D}_y$  as multisets, and we will show that the probability that this happens is  $o(n^{-2})$ .

We say that a vertex  $v \in \Gamma(x) \cup \Gamma(y)$  is *bad* if any of the following hold, and otherwise we say that it is *good*.

1.  $v \in \{x, y\}$ ,
2.  $v$  is adjacent to both  $x$  and  $y$ ,
3.  $v$  is adjacent to a vertex in  $(\Gamma(x) \cup \Gamma(y)) \setminus \{x, y\}$
4. there is a neighbour of  $v$  adjacent to a vertex at distance at most 2 from  $x$  or  $y$  and which is not  $v$ ,
5. the degree of  $v$  is less than  $np/2$ .

We first claim that, with probability  $1 - o(n^{-2})$ , there are at most  $2 \log^{1/2} n$  bad vertices. Note that we will only be interested in applying this when  $d \geq np/2 \geq \log^{2/3} n/2$ , and so the proportion of bad vertices will tend to 0.

**Claim 6.4.** *For any two vertices  $x$  and  $y$ , the number of bad vertices in  $\Gamma(x) \cup \Gamma(y)$  is at most  $2 \log^{1/2} n$  with probability  $1 - o(n^{-2})$ .*

We now reveal the 2-balls around  $x$  and  $y$ . If  $d(x) \neq d(y)$ , then the 3-balls are not isomorphic and we are done, and if  $d = d(x) = d(y)$  is less than  $np/2$ , there is nothing to prove. From the 2-balls, we can also check which of the vertices in  $\Gamma(x) \cup \Gamma(y)$  are bad, and we assume that there are at most  $2 \log^{1/2} n$  of them. The degree of a vertex is dominated by a  $\text{Bin}(n, \log^2 n/n)$  random variable so, by Lemma 2.4, we may also assume that  $d \leq 2 \log^2 n$  and that the union of the 2-balls around  $x$  and  $y$  contains at most  $9 \log^4 n$  vertices. If  $w$  is a neighbour of a good vertex (and not  $x$  or  $y$ ), then  $d(w) - 1$  is a binomial random variable, and moreover, the degrees for such vertices are i.i.d. random variables. Hence, if  $u_i$  is a good vertex, the set  $D(u_i)$  consists of  $d(u_i)$  i.i.d. binomial random variables with at least  $n - 9 \log^4 n$  trials and success probability  $p$ . The following claim shows that the probability that  $D(u_i) = D(v_j)$  is small (for  $i$  and  $j$  such that  $u_i$  and  $v_j$  are both good).

**Claim 6.5.** *Let  $A_1, \dots, A_d$  and  $B_1, \dots, B_d$  be i.i.d. binomial random variables with  $n - \sqrt{n} \leq N \leq n$  trials and success probability  $p \leq 1/2$ , and suppose that  $d \geq np/2$ . If  $np \rightarrow \infty$ , then the probability that  $A = \{A_1, \dots, A_d\}$  and  $B = \{B_1, \dots, B_d\}$  are equal as multisets is at most  $\exp(-\Omega(\sqrt{np} \log(np)))$ .*

If  $\mathcal{D}_x$  and  $\mathcal{D}_y$  are equal as multisets, then there is a permutation  $\sigma$  such that  $D(u_i) = D(v_{\sigma(i)})$  for all  $i \in [d]$ . We show that, given that there are not too many bad vertices, the probability this holds for any particular choice of  $\sigma$  is  $o(1/(n^2 d!))$ , and a union bound over the possible permutations and then the choices for  $x$  and  $y$  completes the proof. Let  $\pi$  be a permutation of  $[d]$ , and consider each  $i = 1, \dots, d$  in turn. If at least one of  $u_i$  or  $v_{\pi(i)}$  is bad, we continue onto the next  $i$ . If neither  $u_i$  nor  $v_{\pi(i)}$  is bad, then Claim 6.5 shows that the probability that  $D(u_i) = D(v_{\pi(i)})$  is at most  $\exp(-\Omega(\sqrt{np} \log(np)))$ . Since we have assumed that there are at most  $2 \log^{1/2} n$  vertices which are bad, we skip at most  $4 \log^{1/2} n$  choices for  $i$ . Hence, the probability that  $D(u_i) = D(v_{\pi(i)})$  for all  $i \in [d]$  is at most  $\exp(-\Omega(d\sqrt{np} \log(np)))$ . By the union bound, the probability that  $\mathcal{D}_x$  and  $\mathcal{D}_y$  are equal is at most

$$\mathbb{P}(\mathcal{D}_x = \mathcal{D}_y) = o(n^{-2}) + \exp(-\Omega(d\sqrt{np} \log(np))) + d \log d.$$

Since  $np \rightarrow \infty$  and  $d \leq 2(np)^3$ , this is  $o(n^{-2}) + \exp(-\Omega(d\sqrt{np} \log(np))) = o(n^{-2})$ .

Finally, taking a union bound over the vertices  $x$  and  $y$  completes the proof.  $\square$

We now prove the two claims made in the proof above.

*Proof of Claim 6.4.* We will bound the number of vertices that fail each of the conditions in the definition of being good. Clearly at most two vertices fail the first condition. The number of vertices which fail the second condition is given by a  $\text{Bin}(n-2, p^2)$  random variable, which is dominated by a  $\text{Bin}(n, \log^4 n/n^2)$  random variable. Hence, using Lemma 2.5, the probability there are at least three vertices which fail the second condition is at most  $e \log^{12} n/n^3 = o(n^{-2})$ .

Consider the vertices in  $\Gamma(x) \cup \Gamma(y)$  which are not one of  $x$  or  $y$ . Using Lemma 2.4, we may assume that there are at most  $4 \log^2 n$  of them. At this point, we have only revealed the edges incident to  $x$  and  $y$ , and so each edge  $uv$  between two of these vertices is present independently with probability  $p$ . Hence, the number of such edges is at most 3 with probability  $o(n^{-2})$ , and at most six vertices fail the third condition.

We split the fourth condition into two parts. First, we consider the number of  $v$  that fail due to one of their neighbours being adjacent to another vertex in  $\Gamma(x) \cup \Gamma(y)$ . A vertex  $z \notin \{x, y\} \cup \Gamma(x) \cup \Gamma(y)$  has a binomial number of neighbours in  $\Gamma(x) \cup \Gamma(y)$  with at most  $4 \log^2 n$  trials and success probability at most  $\log^2 n/n$ . Hence, the probability that  $z$  has at least 4 such neighbours is  $o(n^{-3})$ , and with probability  $1 - o(n^{-2})$ , there is no choice for  $z$  with at least 4 neighbours. The probability that a vertex  $z \notin \{x, y\} \cup \Gamma(x) \cup \Gamma(y)$  has at least two neighbours in  $\Gamma(x) \cup \Gamma(y)$  is at most  $e(4p \log^2 n)^2$ , and so the number of such  $z$  is at dominated by a  $\text{Bin}(n, 16e \log^8 n/n^2)$  random variable. In particular, with probability  $1 - o(n^{-2})$ , there are at most 2 vertices adjacent to least 2 vertices in  $\Gamma(x) \cup \Gamma(y)$ , and they are adjacent to at most 3 vertices. Hence, at most six vertices fail during the first part of the fourth condition.

Let  $W$  be the set of  $v \in \Gamma(x) \cup \Gamma(y)$  which have not already failed. We can reveal the set  $W$  by checking the edges from  $x$  and  $y$  and from  $\Gamma(x)$  and  $\Gamma(y)$ , and note that we may assume that  $|\Gamma(W) \setminus \{x, y\}| \leq 8 \log^4 n$  as this happens with probability  $1 - o(n^{-2})$ . Hence, the number of edges between vertices in  $\Gamma(W) \setminus \{x, y\}$  is dominated by a  $\text{Bin}(64 \log^8 n, \log^2 n/n)$  random variable. In particular, there are at most two edges with probability  $1 - o(n^{-2})$ . Each of these can rule out at most two  $v \in W$ . Hence, at most a further four  $v$  fail here.

Let  $W' = (\Gamma(x) \cup \Gamma(y)) \setminus \{x, y\}$ . We now consider the number of vertices in  $W'$  which have degree less than  $np/2$ . Such a vertex must have less than  $np/2$  neighbours in  $V \setminus (\{x, y\} \cup \Gamma(x) \cup \Gamma(y))$ . We assume that we have revealed the edges from  $x$  and  $y$  and the edges between vertices in  $\Gamma(x) \cup \Gamma(y)$ , but no other edges. We may assume that there are at most  $4 \log^2 n$  vertices in  $\Gamma(x) \cup \Gamma(y)$ . For a given vertex in  $v \in W'$ , the number of neighbours in  $V \setminus (\{x, y\} \cup \Gamma(x) \cup \Gamma(y))$  dominates a binomial random variable with  $n - 4 \log^2 n - 2$  trials and success probability  $p$ . Hence, the probability that it is less than  $np/2$  is at most

$$\exp\left(\frac{-n^2 p}{8(n - 4 \log^2 n - 2)}\right) \leq \exp(-np/16)$$

for large enough  $n$ . Since each vertex  $v \in W'$  satisfies this independently, the number of vertices in  $W'$  which have degree less than  $np/2$  is dominated by a binomial random variable with  $4 \log^2 n$  trials and success probability  $\exp(-np/16)$ . Hence, the probability there are more than  $\log^{1/2} n$  such vertices is at most

$$e\left(4 \log^2 n \exp\left(-\frac{\log^{2/3} n}{16}\right)\right)^{\log^{1/2} n} = \exp\left(\log^{1/2} n \left(\Theta(\log \log n) - \Theta(\log^{2/3} n)\right)\right),$$

which is  $o(n^{-2})$ . Hence, with probability  $o(n^{-2})$ , the number of vertices which are bad is at most  $2 + 2 + 6 + 4 + \log^{1/2} n$ , as required.  $\square$

We now prove Claim 6.5. The general strategy here is similar to the approach used in Lemma 4.2 when we also wanted to show that the probability that two multisets were equal was small: we count the number of  $A_i$  and  $B_i$  which are equal to  $k$  for  $\sqrt{np}$  values of  $k$  close to the mean. The probability that these two quantities are equal is  $O(1/\sqrt{dnp})$ , and this holds even after we have revealed this for  $\sqrt{np}$  choices of  $k$ . However, while the general strategy is similar, this time it is much simpler as the  $A_i$  and  $B_i$  are i.i.d. binomial random variables.

*Proof of Claim 6.5.* Let  $Z_k$  be the number of  $A_1, \dots, A_d$  which are equal to  $k$  and similarly define  $Z'_k$  to be the number of  $B_1, \dots, B_d$  equal to  $k$ . Let  $\ell = \lceil \sqrt{np} \rceil - 1$ , and define  $k_i = \lceil np \rceil + i$  for  $i \in [\ell]$ . By Fact 2.3, we have

$$\mathbb{P}(B_1 \in \{k_1, \dots, k_\ell\}) \leq \mathbb{P}(B_1 > \lceil np \rceil) \leq 1/2.$$

Hence,

$$\mathbb{P}(Z'_{k_1} + \dots + Z'_{k_\ell} \leq d/4) \leq \mathbb{P}(\text{Bin}(d, 1/2) \geq 3d/4) \leq \exp(-d/18).$$

Suppose this is the case and reveal the values  $Z'_{k_i}$ , which we call our *target values*. We will iteratively reveal the  $A_j$  which are equal to  $k_i$ , and check if there are  $Z'_{k_i}$  of them. Suppose we are about to reveal the  $A_j$  equal to  $k_i$ , so we have already revealed the values  $Z_{k_1}, \dots, Z_{k_{i-1}}$  and they are equal to  $Z'_{k_1}, \dots, Z'_{k_{i-1}}$ . We will show that the probability that  $Z_{k_i}$  is equal to  $Z'_{k_i}$  is  $O(1/np)$ . Suppose that  $A_j$  has not been revealed, so we know that  $A_j$  is not equal to  $k_1, \dots, k_{i-1}$ . We have

$$|k_i - np| \leq |k_i - np| + |np - np| \leq i + 1 + \sqrt{np} \leq 2\sqrt{np}$$

for large  $n$ , and so by Lemma 2.7, there are constants  $\alpha$  and  $\beta$  such that

$$\frac{\alpha}{\sqrt{Np(1-p)}} < \mathbb{P}(A_j = k_i | A_j \notin \{k_1, \dots, k_{i-1}\}) < \frac{2\beta}{\sqrt{Np(1-p)}}.$$

Hence, the number of unrevealed  $A_j$  which are equal to  $k_i$  is a binomial random variable with at least  $d/4$  trials and success probability  $\Theta(1/\sqrt{Np})$ . In particular,

$$\mathbb{P}(Z_{k_i} = Z'_{k_i}) \leq \sup_x \mathbb{P}(Z_{k_i} = x) = O\left(\frac{1}{\sqrt{dNp}}\right) = O\left(\frac{1}{\sqrt{dnp}}\right).$$

If  $A$  and  $B$  are equal as multisets, then either  $Z'_{k_1} + \dots + Z'_{k_\ell} > d/4$  or all of the steps succeed, and both of these happen with probability  $\exp(-\Omega(\sqrt{np} \log(np)))$ .  $\square$

### 6.3 The set of 3-balls after swapping edges

In this section we prove Lemma 3.7, that is, we show that there is a constant  $\alpha > 0$  such that if  $\frac{\log^{2/3} n}{n} \leq p \leq \alpha \frac{\log^2 n}{n(\log \log n)^3}$ , a random graph  $G \in \mathcal{G}(n, p)$  is not 3-reconstructible with high probability. The main idea of the proof will be to show that, with high probability, there exist two edges  $xy, uv$  in  $G$  such that by deleting these edges and adding  $xv, yu$  we obtain a graph  $G'$  which is not isomorphic to  $G$ , but has the same collection of 3-balls. Lemma 3.6 shows that we may assume the 3-balls around vertices of “large” degree are all distinct, in which case, if  $u, v, x$  and  $y$  all have large degree, the graphs  $G$  and  $G'$  are not isomorphic. To find the edges to swap we consider the structures  $H_{uv}$  defined as follows. For an edge  $uv$ , let  $H_{uv}$  be the subgraph  $G[\Gamma_{\leq 2}(u) \cup \Gamma_{\leq 2}(v)]$  induced by the vertices at distance at most 2 from  $u$  or  $v$ , and distinguish the edge  $uv$ . We will only consider the  $H_{uv}$  for “good” edges whose 5-balls are trees and where all the vertices in  $H_{uv}$  have “typical” degrees. There are many good edges but not that many isomorphism classes for the  $H_{uv}$ , and so, by the pigeonhole principle, there must be two edges  $uv$  and  $xy$  with  $H_{uv} \simeq H_{xy}$ . This is not quite enough to guarantee that the switch does not change the 3-balls by introducing extra edges and we will also require the edges to be far apart.

*Proof of Lemma 3.7.* Let  $G \in \mathcal{G}(n, p)$  where  $\frac{\log^{2/3} n}{n} \leq p \leq \alpha \frac{\log^2 n}{n(\log \log n)^3}$ . We will show there exist vertices  $u, v, x, y$  as claimed using a pigeonhole argument over the  $H_{uv}$  of good edges. We say that an edge  $uv$  is a *good* edge if  $G[\Gamma_{\leq 5}(u) \cup \Gamma_{\leq 5}(v)]$  is a tree and  $|d(z) - (n-1)p| < 10\sqrt{np \log(np)}$  for every  $z \in \Gamma_{\leq 2}(u) \cup \Gamma_{\leq 2}(v)$ . We will need the following claim which bounds the number of “pigeonholes”.

**Claim 6.6.** *The number of isomorphism classes for the  $H_{uv}$  of the good edges is at most*

$$400np \log(np) \exp\left(42\left((np)^{1/2} \log^{3/2}(np)\right)\right).$$

Having bounded the number of pigeonholes we have, we now consider the number of pigeons, or the number of good edges  $uv$  in  $G$ . The following claim shows that there are at least  $n^2 p/8$  good edges with high probability.

**Claim 6.7.** *With probability  $1 - o(1)$ , the graph  $G$  satisfies the following:*

- (i) The number of edges of  $G$  contained in a cycle of length at most 12 is at most  $\log^{24} n$ .
- (ii) The maximum degree of  $G$  is at most  $\log^2 n$ .
- (iii) The number of vertices  $z$  with degree  $d(z)$  such that  $|d(z) - (n-1)p| > 10\sqrt{np \log(np)}$  is at most  $n^{-31}p^{-32}$ .
- (iv)  $G$  contains at least  $n^2p/4$  edges.
- (v) The 3-balls around vertices of degree at least  $np/2$  are all distinct.

Let us denote the subgraph of  $G$  induced by the vertices at distance at most 5 from  $u$  or  $v$  by  $N_5(u, v)$ , i.e.  $N_5(u, v) = G[\Gamma_{\leq 5}(u) \cup \Gamma_{\leq 5}(v)]$ . We note that if  $N_5(u, v)$  is not a tree, then it contains a cycle of length at most 12, so it will be enough to count the number of edges  $uv$  such that  $N_5(u, v)$  does not contain a cycle of length at most 12 and every  $z \in V(H_{uv})$  satisfies the degree condition that  $|d(z) - (n-1)p| \leq 10\sqrt{np \log(np)}$ . For this we will first bound the number of edges  $uv$  for which there is a vertex  $z \in H_{uv}$  with  $|d(z) - (n-1)p| > 10\sqrt{np \log(np)}$ , and then we will bound the number of edges  $uv$  for which there is an edge  $e \in N_5(u, v)$  that is contained in a cycle of length 12 in  $G$ . The sum of these two bounds will be an upper bound on the number of bad edges.

Assume that the graph  $G$  satisfies the conditions Claim 6.7. Then the second condition implies that there are at most  $\log^{2k} n$  vertices in the  $k$ th neighbourhood of a vertex, and hence every vertex  $x$  is in at most  $\log^2 n(\log^4 n + \log^2 n + 1) \leq 2\log^6 n$  of the  $H_{uv}$ . Indeed, the number of vertices  $u$  such that  $x \in \Gamma_{\leq 2}(u)$  is at most  $1 + \log^2 n + \log^4 n$ , and there are at most  $\log^2 n$  possible different subgraphs  $H_{uv}$  for each vertex  $u$ . In particular, a vertex  $z$  with  $|d(z) - (n-1)p| > 10\sqrt{np \log(np)}$  can be contained in at most  $2\log^6 n$  subgraphs  $H_{uv}$ . Thus, given the third condition above, the number of edges  $uv$  such that  $H_{uv}$  contains such a vertex  $z$  is at most  $n^{-31}p^{-32}2\log^6 n$ . Similarly, each vertex is in at most  $2\log^{12} n$  of the  $N_5(u, v)$  so clearly each edge is in at most  $2\log^{12} n$  of the  $N_5(u, v)$ . Thus, given the first condition above, the number of edges  $uv$  such that  $N_5(u, v)$  contains an edge which is in a cycle of length at most 12 is  $2\log^{12} n \log^{24} n$ .

Using (i) and (iii), the number of bad edges for our range of  $p$  is at most

$$2\log^{12} n \cdot \log^{24} n + 4\log^6 n \cdot n^{-31}p^{-32} \leq 2\log^{36} n + 4n \log^{-14} n \leq n$$

for large enough  $n$ . From the fourth condition  $G$  has at least  $n^2p/4 \geq n \log^{2/3} n/4$  edges and therefore (crudely) there are at least  $n^2p/8$  good  $H_{uv}$  for large enough  $n$ .

We now use Claim 6.6 to finish the proof. There must be some isomorphism class of  $H_{uv}$  that occurs at least

$$\frac{n^2p}{8 \cdot 400np \log(np) \exp(42((np)^{1/2} \log^{3/2}(np)))} \geq \exp(\log n - \log \log(np) - 43((np)^{1/2} \log^{3/2}(np)))$$

times. That is, there is some good structure  $J$  which appears as  $H_{uv}$  for at least this many edges  $uv$ . Noting that  $p \leq \beta \frac{\log^2 n}{n(\log \log n)^3}$ , this is at least

$$\exp\left(\left(1 - 43\sqrt{8\beta}\right) \log n - O(\log n / (\log \log n)^{3/2})\right) \geq 4\log^{14} n + 1,$$

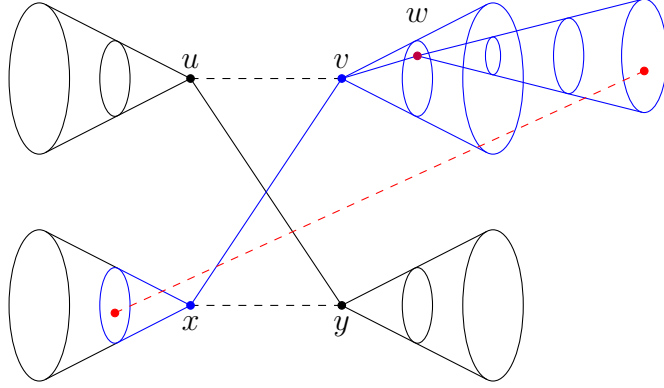


Figure 3: The 3-ball around a vertex  $w$  in the neighbourhood of  $v$  in  $G'$  is shown in blue. The assumption that  $H_{uv} \simeq H_{xy}$  does not rule out the existence of the red edge, but this edge would create a path from  $v$  to  $x$  of length 6 in  $G$ .

if  $\beta$  is sufficiently small. Suppose that  $H_{uv} \simeq J$ . There are at most  $2(\log^2 n)^6$  vertices at distance at most 6 from any vertex  $w$ , and there are at most  $4 \log^{12} n$  vertices at distance at most 6 from  $u$  or  $v$ . Hence, there are at most  $4 \log^{14} n$  edges where at least one vertex is at distance at most 6 from  $u$  or  $v$ . Thus, there is a good edge  $xy$  such that  $H_{xy} \simeq J$  and both  $x$  and  $y$  are at distance at least 7 from both  $u$  and  $v$ .

Fix an isomorphism from  $H_{uv}$  to  $H_{xy}$  and suppose without loss of generality that  $u$  is mapped to  $x$ . Let  $G' = (G \setminus \{uv, xy\}) \cup \{uy, vx\}$ . We claim that  $G'$  has the same collection of 3-balls as  $G$  and that  $G'$  is not isomorphic to  $G$ .

Note that the 3-ball of a vertex  $w$  is clearly unchanged if  $w$  is not in the 2-ball of one of  $u, v, x$  or  $y$ , so suppose it is in  $\Gamma_{\leq 2}(v)$ . Since  $N_5(u, v)$  is a tree, the 3-ball of  $w$  in  $G$  is a tree  $T$ . As  $H_{uv} \simeq H_{xy}$  (with  $u$  mapping to  $x$ ), the 3-ball of  $u$  in  $G'$  certainly contains a copy  $T'$  of  $T$ , but this condition alone does not rule out the possibility that  $w$  contains extra edges between  $T' \cap T$  and  $T' \setminus T$  (see Figure 3 for an example). However, any extra edge would create a cycle of length at most 7 and it must use the edge  $vx$ . This means that  $v$  and  $x$  are at distance at most 6 in  $G$ , which contradicts the choice of  $xy$ .

The graphs  $G$  and  $G'$  cannot be isomorphic as the 3-balls around vertices of degree at least  $np/2$  are unique and  $G$  contains an edge between a vertex with 3-ball  $N_3(u)$  and a vertex with 3-ball  $N_3(v)$  while  $G'$  does not.  $\square$

It remains to prove our technical claims.

*Proof of Claim 6.6.* When  $uv$  is a good edge, the structure  $H_{uv}$  is a tree with a distinguished edge where each vertex  $z \in V(H_{uv})$  satisfies  $|d(z) - (n-1)p| < 10\sqrt{np \log(np)}$ . It suffices to bound the number of different options for  $d(u)$ ,  $d(v)$  and the multisets  $\{d(z) : z \in \Gamma(u) \setminus v\}$  and  $\{d(z) : z \in \Gamma(v) \setminus u\}$ . The condition  $|d(z) - (n-1)p| < 10\sqrt{np \log(np)}$  means that all the degrees are one of  $N = \lfloor 20\sqrt{np \log(np)} \rfloor$  options. Hence, the multiset  $\{d(z) : z \in \Gamma(u) \setminus v\}$  is a multiset of  $d(u) - 1$  entries spread across at most  $N$  options,



and so there are at most

$$\begin{aligned} \binom{d(u) + N - 2}{N - 1} &\leq (d(u) + N)^N \\ &\leq \left(np + 30\sqrt{np \log(np)}\right)^{20\sqrt{np \log(np)}} \\ &\leq \exp\left(21((np)^{1/2} \log^{3/2}(np))\right) \end{aligned}$$

possible multisets. The same is true for the multiset  $\{d(z) : z \in \Gamma(v) \setminus u\}$ . This means there are at most

$$\left(20\sqrt{np \log(np)} \exp\left(21(\sqrt{np} \log^{3/2}(np))\right)\right)^2 = 400np \log(np) \exp\left(42\left(\sqrt{np} \log^{3/2}(np)\right)\right)$$

possible isomorphism classes for the  $H_{uv}$  of a good edge, as required.  $\square$

*Proof of Claim 6.7.* Let  $G \in \mathcal{G}(n, p)$ . We show that each of the conditions holds with probability  $1 - o(1)$ , and the union bound over the five events completes the proof.

- (i) For each  $k \in \{3, \dots, 12\}$ , let  $C_k$  be the number of cycles of length  $k$  in  $G$ . Then  $\mathbb{E}[C_k] \leq n^k p^k$ . For the range of  $p$  that we consider, we have  $np = o(\log^2 n)$  and so the expected number of edges in cycles of length at most 12 is bounded by

$$\sum_{k=3}^6 k \mathbb{E}[C_k] \leq \sum_{k=3}^{12} kn^k p^k = o(\log^{24} n).$$

The claim now follows from Markov's Inequality.

- (ii) Note that the degree  $d(z)$  of a vertex  $z$  is distributed like a  $\text{Bin}(n - 1, p)$  random variable. For large enough  $n$ , we have  $p \leq \log^2 n / (2n - 2)$  and so Lemma 2.4 gives

$$\mathbb{P}(d(z) \geq \log^2 n) \leq \mathbb{P}\left(\text{Bin}\left(n - 1, \frac{\log^2 n}{2n - 2}\right) \geq \log^2 n\right) \leq \exp\left(-\frac{2}{3} \log^2 n\right) = o(n^{-1}).$$

The claim now follows from a union bound.

- (iii) Again applying Lemma 2.4 we get

$$\mathbb{P}\left(|d(x) - (n - 1)p| > 10\sqrt{np \log(np)}\right) \leq 2 \exp\left(-\frac{100}{3} \log(np)\right) \leq 2(np)^{-33}.$$

Thus, the expected number of vertices  $z$  with  $|d(z) - (n - 1)p| > 10\sqrt{np \log(np)}$  is bounded by  $2n^{-32}p^{-33}$ . We are then done by Markov's Inequality since  $np \rightarrow \infty$ .

- (iv) The number of edges in  $G$  is distributed like a  $\text{Bin}\left(\binom{n}{2}, p\right)$  random variable so the result follows from Lemma 2.4.

- (v) This follows from Lemma 3.6.  $\square$

## References

- [1] K. Adhikari and S. Chakraborty. Shotgun assembly of Linial-Meshulam model. *arXiv preprint arXiv:2209.10942*, 2022.
- [2] K. Adhikari and S. Chakraborty. Shotgun assembly of random geometric graphs. *arXiv preprint arXiv:2202.02968*, 2022.
- [3] N. Alon and J. H. Spencer. *The probabilistic method*. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, fourth edition, 2016.
- [4] R. Arratia, D. Martin, G. Reinert, and M. S. Waterman. Poisson process approximation for sequence repeats, and sequencing by hybridization. *Journal of Computational Biology*, 3(3):425–463, 1996.
- [5] K. J. Asciak, M. A. Francalanza, J. Lauri, and W. Myrvold. A survey of some open questions in reconstruction numbers. *Ars Combinatoria*, 97:443–456, 2010.
- [6] P. Balister, B. Bollobás, and B. Narayanan. Reconstructing random jigsaws. In *Multiplex and multilevel networks*, pages 31–50. Oxford Univ. Press, Oxford, 2019.
- [7] B. Bollobás. Almost every graph has reconstruction number three. *Journal of Graph Theory*, 14(1):1–4, 1990.
- [8] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.
- [9] J. A. Bondy. A graph reconstructor’s manual. In *Surveys in combinatorics, 1991 (Guildford, 1991)*, volume 166 of *London Math. Soc. Lecture Note Ser.*, pages 221–252. Cambridge Univ. Press, Cambridge, 1991.
- [10] J. A. Bondy and R. L. Hemminger. Graph reconstruction—a survey. *Journal of Graph Theory*, 1(3):227–268, 1977.
- [11] C. Bordenave, U. Feige, and E. Mossel. Shotgun assembly of random jigsaw puzzles. *Random Structures & Algorithms*, 56(4):998–1015, 2020.
- [12] A. Bowler, P. Brown, and T. Fenner. Families of pairs of graphs with a large number of common cards. *Journal of Graph Theory*, 63(2):146–163, 2010.
- [13] T. Czajka and G. Pandurangan. Improved random graph isomorphism. *Journal of Discrete Algorithms*, 6(1):85–92, 2008.
- [14] J. Ding and H. Liu. Shotgun assembly threshold for lattice labeling model. *Probability Theory and Related Fields*, 2023.
- [15] J. Ding, J. Yang, and H. Ma. Shotgun threshold for sparse Erdős-Rényi graphs. *arXiv preprint arXiv:2208.09876*, 2022.
- [16] M. Dyer, A. Frieze, and S. Suen. The probability of unique solutions of sequencing by hybridization. *Journal of Computational Biology*, 1(2):105–110, 1994.

- [17] J. Gaudio and E. Mossel. Shotgun assembly of Erdős-Rényi random graphs. *Electronic Communications in Probability*, 27:Paper No. 5, 14, 2022.
- [18] J. Gaudio, M. Z. Rácz, and A. Sridhar. Local canonical labeling of Erdős-Rényi random graphs. *arXiv preprint arXiv:2211.16454*, 2022.
- [19] W. B. Giles. Reconstructing trees from two point deleted subtrees. *Discrete Mathematics*, 15(4):325–332, 1976.
- [20] C. Groenland, T. Johnston, A. Scott, and J. Tan. Reconstructing trees from small cards. *arXiv preprint arXiv:2103.13359*, 2021.
- [21] H. Huang and K. Tikhomirov. Shotgun assembly of unlabeled Erdős-Rényi graphs. *arXiv preprint arXiv:2108.09636*, 2021.
- [22] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.
- [23] P. J. Kelly. *On isometric transformations*. PhD thesis, University of Wisconsin, 1942.
- [24] P. J. Kelly. A congruence theorem for trees. *Pacific Journal of Mathematics*, 7:961–968, 1957.
- [25] A. V. Kostochka, M. Nahvi, D. B. West, and D. Zirlin. 3-regular graphs are 2-reconstructible. *European Journal of Combinatorics*, 91:Paper No. 103216, 10, 2021.
- [26] J. Lauri and R. Scapellato. *Topics in graph automorphisms and reconstruction*, volume 432 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, second edition, 2016.
- [27] L. Le Cam. An approximation theorem for the Poisson binomial distribution. *Pacific Journal of Mathematics*, 10:1181–1197, 1960.
- [28] A. Martinsson. Shotgun edge assembly of random jigsaw puzzles. *arXiv preprint arXiv:1605.07151*, 2016.
- [29] A. Martinsson. A linear threshold for uniqueness of solutions to random jigsaw puzzles. *Combinatorics, Probability and Computing*, 28(2):287–302, 2019.
- [30] M. Mitzenmacher and E. Upfal. *Probability and computing*. Cambridge University Press, Cambridge, second edition, 2017. Randomization and probabilistic techniques in algorithms and data analysis.
- [31] R. Molina. Correction of a proof on the ally-reconstruction number of a disconnected graph. Correction to: “The ally-reconstruction number of a disconnected graph” [Ars Combin. **28** (1989), 123–127; MR1039138 (90m:05094)] by W. J. Myrvold. *Ars Combinatoria*, 40:59–64, 1995.
- [32] E. Mossel and N. Ross. Shotgun assembly of labeled graphs. *IEEE Transactions on Network Science and Engineering*, 6(2):145–157, 2019.

- [33] E. Mossel and N. Sun. Shotgun assembly of random regular graphs. *arXiv preprint arXiv:1512.08473*, 2015.
- [34] A. S. Motahari, G. Bresler, and D. N. C. Tse. Information theory of DNA shotgun sequencing. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, 59(10):6273–6289, 2013.
- [35] V. Müller. Probabilistic reconstruction from subgraphs. *Commentationes Mathematicae Universitatis Carolinae*, 17(4):709–719, 1976.
- [36] W. Myrvold. The ally-reconstruction number of a disconnected graph. *Ars Combinatoria*, 28:123–127, 1989.
- [37] W. Myrvold. The ally-reconstruction number of a tree with five or more vertices is three. *Journal of Graph Theory*, 14(2):149–166, 1990.
- [38] W. J. Myrvold. *Ally and adversary reconstruction problems*. PhD thesis, University of Waterloo, 1988.
- [39] B. Narayanan and C. Yap. Reconstructing random pictures. *arXiv preprint arXiv:2210.09410*, 2022.
- [40] R. Nenadov, P. Pfister, and A. Steger. Unique reconstruction threshold for random jigsaw puzzles. *Chicago Journal of Theoretical Computer Science*, pages Art. 2, 16, 2017.
- [41] M. Przykucki, A. Roberts, and A. Scott. Shotgun reconstruction in the hypercube. *Random Structures & Algorithms*, 60(1):117–150, 2022.
- [42] B. A. Rogozin. An estimate for concentration functions. *Theory of Probability & Its Applications*, 6(1):94–97, 1961.
- [43] D. Soudry, S. Keshri, P. Stinson, M.-h. Oh, G. Iyengar, and L. Paninski. Efficient “shotgun” inference of neural connectivity from highly sub-sampled activity data. *PLoS computational biology*, 11(10):e1004464, 2015.
- [44] H. Spinoza and D. B. West. Reconstruction from the deck of  $k$ -vertex induced subgraphs. *Journal of Graph Theory*, 90(4):497–522, 2019.
- [45] J. M. Steele. Le Cam’s inequality and Poisson approximations. *American Mathematical Monthly*, 101(1):48–54, 1994.
- [46] S. M. Ulam. *A collection of mathematical problems*. Interscience Tracts in Pure and Applied Mathematics, no. 8. Interscience Publishers, New York-London, 1960.