Journal of Combinatorial Theory, Series B 160 (2023) 144–162

Contents lists available at ScienceDirect

Journal of Combinatorial Theory, Series B

journal homepage: www.elsevier.com/locate/jctb

Isomorphisms between random graphs

Sourav Chatterjee [1], Persi Diaconis [2]

Departments of Mathematics and Statistics, Stanford University, United States of America

ARTICLE INFO

Article history:
Received 8 September 2021
Available online 17 January 2023

Keywords:
Random graph
Graph isomorphism
Concentration inequality

ABSTRACT

Consider two independent Erdős–Rényi $G(N, 1/2)$ graphs. We show that with probability tending to 1 as $N \to \infty$, the largest induced isomorphic subgraph has size either $\lfloor x_N - \varepsilon_N \rfloor$ or $\lfloor x_N + \varepsilon_N \rfloor$, where $x_N = 4\log_2 N - 2\log_2 \log_2 N - 2\log_2(4/e) + 1$ and $\varepsilon_N = (4\log_2 N)^{-1/2}$. Using similar techniques, we also show that if Γ_1 and Γ_2 are independent $G(n, 1/2)$ and $G(N, 1/2)$ random graphs, then Γ_2 contains an isomorphic copy of Γ_1 as an induced subgraph with high probability if $n \le \lfloor y_N - \varepsilon_N \rfloor$ and does not contain an isomorphic copy of Γ_1 as an induced subgraph with high probability if $n > \lfloor y_N + \varepsilon_N \rfloor$, where $y_N = 2\log_2 N + 1$ and ε_N is as above.

1. Introduction

This paper has two motivations. First, in giving expository talks about the Rado graph, we asked: "Let Γ_1 and Γ_2 be independent $G(N, 1/2)$. What's the chance they are isomorphic? Small? How small? Less than $N!2^{-\binom{N}{2}}$. So, when $N = 100$, less than 10^{-1300}. Now suppose $N = \infty$ (a $G(\infty, 1/2)$ graph is called a Rado graph). The chance

E-mail addresses: souravc@stanford.edu (S. Chatterjee), diaconis@math.stanford.edu (P. Diaconis).
[1] S. Chatterjee was partially supported by NSF grants DMS-1855484 and DMS-2113242.
[2] P. Diaconis was partially supported by NSF grant DMS-1954042.

https://doi.org/10.1016/j.jctb.2023.01.001
0095-8956/

Table 1
Results from simulating one instance of $G(N, 1/2)$.

N	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
$\lfloor x_N - \varepsilon_N \rfloor$	14	15	15	15	15	15	15	15	16	16	16	16	16	16	16
$\lfloor x_N + \varepsilon_N \rfloor$	15	15	15	15	15	16	16	16	16	16	16	16	16	17	17
L_N	14	14	14	14	15	15	15	15	15	15	15	15	16	16	16

that $\Gamma_1 \cong \Gamma_2$ is one!" To help think about this seeming discontinuity, we asked, for finite N, how large is the largest induced isomorphic subgraph of Γ_1 and Γ_2.

Theorem 1.1. *Let Γ_1 and Γ_2 be independent $G(N, 1/2)$ graphs. Let L_N be the size of the largest induced isomorphic subgraph. Then, with probability tending to one as $N \to \infty$, L_N is either $\lfloor x_N - \varepsilon_N \rfloor$ or $\lfloor x_N + \varepsilon_N \rfloor$, where $x_N = 4\log_2 N - 2\log_2\log_2 N - 2\log_2(4/e) + 1$ and $\varepsilon_N = (4\log_2 N)^{-1/2}$.*

To interpret the above result, note that $\varepsilon_N \leq 1/2$ when $N \geq 2$. This implies that $\lfloor x_N - \varepsilon_N \rfloor$ and $\lfloor x_N + \varepsilon_N \rfloor$ are either the same integer, or differ exactly by one. Thus, L_N is either concentrated at one point, or at two consecutive points, depending on N. The latter case can happen only when x_N is close to an integer (specifically, within less than ε_N of an integer).

Incidentally, our proof does not rule out the possibility that L_N is asymptotically concentrated on one point, and not two. Proving or disproving this would require a more delicate analysis of certain remainder terms than what we currently have (see the remark at the end of Section 2). Concentration on two points, although rare, is not unprecedented. For a classical example, see [5, Theorem 5.1].

Simulations run by Ciaran McCreesh and James Trimble using the McSplit algorithm [14] (personal communication) indicate that the asymptotics of Theorem 1.1 kick in for N as small as 31 (with a sample of size 1). For $N = 31$, we have $x_N \doteq 15.08$ and $\varepsilon_N \doteq 0.23$, and in the simulation, L_N turned out to be 14. (Theorem 1.1 says that with high probability, $L_N \in \{14, 15\}$.) And the results continue to match the prediction of Theorem 1.1 (sometimes off by one) all the way up to $N = 45$, when the task starts approaching computational limits (see Table 1).

Theorem 1.1 is proved in Section 2, which gives some background. The argument requires some work beyond the usual second moment method.

Our second motivation for the problems studied here came from asking experts (we were sure that Theorem 1.1 was in the literature). Svante Janson had just been asked the second question in the abstract (the one on subgraph isomorphism, described below) by Don Knuth. Subgraph isomorphism is a basic question in constraint satisfaction. Developers of programs had used $G(n, 1/2)$, $G(N, 1/2)$ as the center of extensive tests comparing algorithms. They had observed a sharp transition: When $n = 15$, $N = 150$, the chance of finding an induced copy of Γ_1 in Γ_2 is close to 1. When $n = 16$, $N = 150$, the chance is close to zero. Our theorem predicts this. To state a sharp result, let $P(n, N)$ be the probability that a random $G(N, 1/2)$ graph contains an induced copy of a random

$G(n, 1/2)$ graph. The following theorem shows that $P(n, N)$ drops from 1 to 0 in a window of size ≤ 2.

Theorem 1.2. *With above notation, as* $N \to \infty$, $P(\lfloor y_N - \varepsilon_N \rfloor, N) \to 1$ *and* $P(\lfloor y_N + \varepsilon_N \rfloor + 1, N) \to 0$, *where* $y_N = 2\log_2 N + 1$ *and* $\varepsilon_N = (4\log_2 N)^{-1/2}$.

For example, when $N = 150$, $y_N \doteq 15.46$ and $\varepsilon_N \doteq 0.19$. So the theorem says we expect $P(15, 150) \doteq 1$ and $P(16, 150) \doteq 0$, matching explicit computation [15].

Incidentally, Theorem 1.2 is related to the classic problem of figuring out the size of the smallest graph that contains all graphs on n vertices as induced subgraphs. After many years of partial progress, this problem was recently settled by Noga Alon [1], who showed that the minimum N is $(1+o(1))2^{(n-1)/2}$. The key step in Alon's proof is to show that if $\binom{N}{n}2^{-\binom{n}{2}} = \lambda$, then the probability that $G(N, 1/2)$ contains all graphs of size n as induced subgraphs is $(1 - e^{-\lambda})^2 + o(1)$ uniformly in n as $N \to \infty$. Without checking all the details, it seems to us that Alon's result shows that the quantity $\widetilde{P}(n, N) := \mathbb{P}(G(N, 1/2)$ contains all graphs of size $n)$ drops from 1 to 0 in a window of size 2. This window is slightly to the left of our window, at $2\log_2 N - 2\log_2\log_2 N + 2\log_2(e/2) + 1$. The reason is that the main contributor to $\widetilde{P}(n, N)$ is the probability that $G(N, 1/2)$ contains a copy of K_n (the complete graph of size n), which is the 'hardest' subgraph to contain. In particular, it is harder to contain K_n than it is to contain a copy of $G(n, 1/2)$, which is why Alon's window is shifted by $O(\log\log N)$ to the left.

The fact that the size of the largest clique is concentrated at the above point is known from old work of Matula [12,13] and Bollobás and Erdős [3]. Several other graph invariants have concentrated distributions but others can be proved to be 'non-concentrated' at a finite number of points. For a survey and fascinating work on the chromatic number, see [8].

Theorem 1.2 is proved in Section 3, which begins with a literature review on subgraph isomorphism. The final section has remarks and open problems.

Acknowledgments

We thank Maryanthe Malliaris and Peter Cameron for teaching us about the Rado graph, and Jacob Fox, Benny Sudakov, and Noga Alon for useful feedback and references. We thank Svante Janson and Don Knuth for pointing us to subgraph isomorphism and for careful technical reading of our paper, and several useful suggestions. We thank Ciaran McCreesh and James Trimble for running a set of very helpful simulations using their McSplit algorithm, and Don Knuth for facilitating our communication with them. Lastly, we thank the two anonymous referees for several helpful comments.

2. Isomorphic graphs

Because two $G(\infty, 1/2)$ graphs are isomorphic with probability 1, a random R from $G(\infty, 1/2)$ has come to be called *the* random graph. Non-random models for R abound: Let the vertex set be $\{0, 1, 2, \ldots\}$ and put an undirected edge from i to j if $i < j$ and the i^{th} binary digit of j is a 1 (labeling the rightmost digit as the zeroth digit). So $i = 0$ is connected to all odd j, $i = 1$ is connected to $j \equiv 2$ or 3 (mod 4) and also to 0, and so on. The amazing properties of R are beautifully exposited (and proved) in Peter Cameron's lovely article [4].

Logicians have developed facts about R (and higher cardinality versions). After all, a graph is just a symmetric relation. Some of this is used in Diaconis and Malliaris [6] to show that various algebraic problems are intractable because an associated commuting graph contains R as an induced subgraph.

The discontinuity between finite and infinite N is jarring. Aren't infinite limits supposed to shed light on large finite N? Exploring this led to Theorem 1.1.

A related question is pick an Erdős–Rényi graph and asks for the largest k such that it contains two disjoint isomorphic induced subgraphs with k vertices. A combinatorial application of this problem is studied in [10]. Similar questions can be asked for other combinatorial objects (largest disjoint isomorphic subtrees in a tree, largest disjoint order isomorphic pair of sub-permutations in a permutation — see [7] for applications of such problems).

In the remainder of this section, we prove Theorem 1.1. Although we use log base 2 in the statement of the theorem, we will work with log base e throughout the proof, which will be denoted by log, as usual. The proofs of both Theorems 1.1 and 1.2 make use of the following technical result. Take any $1 \le m \le n$. Let Γ be a $G(n, 1/2)$ random graph. Let $X(i, j)$ be the indicator that $\{i, j\}$ is an edge in Γ. Let

$$\phi(m, n) := \sum_{\pi \in S_n} \mathbb{P}(X(i, j) = X(\pi(i), \pi(j)) \text{ for all } 1 \le i < j \le m), \tag{2.1}$$

where the right side is interpreted as $n!$ if $m = 0$ or $m = 1$.

Proposition 2.1. *There are positive universal constants K_1 and K_2 such that for all $n \ge 1$ and $2n/3 \le m \le n$,*

$$\phi(m, n) \le K_1 e^{K_2 (n-m) \log(n-m)},$$

where $0 \log 0$ is interpreted as 0.

The proof of this proposition requires two lemmas.

Lemma 2.2. *Let $\pi \in S_n$. Let $A \subseteq \{1, \ldots, n\}$ be a set such that $\pi(i) \ne i$ for each $i \in A$. Then there exists $B \subseteq A$, such that $|B| \ge |A|/3$, and $\pi(i) \notin B$ for each $i \in B$.*

Proof. Let $C = (c_1, \ldots, c_k)$ be a cycle in π. In the following, we will sometimes treat C as the set $\{c_1, \ldots, c_k\}$. First, suppose that k is even. Let $A_1 := A \cap \{c_1, c_3, \ldots, c_{k-1}\}$ and $A_2 := A \cap \{c_2, c_4, \ldots, c_k\}$. Let A' be the larger of these two sets. Then $|A'| \geq \frac{1}{2}|A \cap C|$ and $\pi(i) \in C \setminus A'$ for each $i \in A'$. Next, suppose that k is odd. If $k = 1$, then A does not intersect C because A contains no fixed point of π. In this case, let $A' := \emptyset$. If $k \geq 3$ (and odd), let $A_1 := A \cap \{c_1, c_3, \ldots, c_{k-2}\}$ and $A_2 := A \cap \{c_2, c_4, \ldots, c_{k-1}\}$. Let A' be the larger of these two sets. Then again we have that $\pi(i) \in C \setminus A'$ for each $i \in A'$, and

$$|A'| \geq \frac{1}{2}|A \cap \{c_1, \ldots, c_{k-1}\}|. \tag{2.2}$$

Now, if $C \subseteq A$, then by the above inequality and the fact that $k \geq 3$, we get

$$|A'| \geq \frac{k-1}{2} \geq \frac{k}{3} = \frac{1}{3}|A \cap C|.$$

On the other hand, if there is some element of C that is not in A, then we may assume without loss of generality that $c_k \notin A$, because the cycle C can be alternatively represented as $(c_l, c_{l+1}, \ldots, c_k, c_1, c_2 \ldots, c_{l-1})$ for any l. Thus, in this case, (2.2) gives

$$|A'| \geq \frac{1}{2}|A \cap C|.$$

To summarize, given any cycle C, we have constructed a set $A' \subseteq A \cap C$, such that $|A'| \geq \frac{1}{3}|A \cap C|$, and $\pi(i) \in C \setminus A'$ for each $i \in A'$. Let B be the union of A' over all cycles C. It is easy to see that this B satisfies the two required properties. □

Lemma 2.3. *Take any $\pi \in S_n$. Let $k := |\{i \leq m : \pi(i) = i\}|$. Then*

$$\mathbb{P}(X(i,j) = X(\pi(i), \pi(j)) \text{ for all } 1 \leq i < j \leq m) \leq C_1 e^{-C_2(m-k)m},$$

where C_1 and C_2 are positive universal constants.

Proof. Let $F := \{i \leq m : \pi(i) = i\}$ and $A := \{i \leq m : \pi(i) \neq i\}$, so that $|F| = k$ and $|A| = m - k$. By Lemma 2.2, there exists $B \subseteq A$ such that $|B| \geq \frac{1}{3}|A| = \frac{1}{3}(m-k)$, and $\pi(i) \notin B$ for all $i \in B$. Moreover, since $\pi(\pi(i)) \neq \pi(i)$ for $i \in B$ (because otherwise, $\pi(i) = i \in B$), we conclude that $\pi(i) \notin F$. By independence of edges, this gives

$$\begin{aligned}
&\mathbb{P}(X(i,j) = X(\pi(i), \pi(j)) \text{ for all } 1 \leq i < j \leq m) \\
&\leq \mathbb{P}(X(i,j) = X(\pi(i), \pi(j)) \text{ for all } i \in B,\, j \in F) \\
&= \mathbb{P}(X(i,j) = X(\pi(i), j) \text{ for all } i \in B,\, j \in F) \\
&= 2^{-|B||F|} \leq 2^{-\frac{1}{3}(m-k)k}.
\end{aligned}$$

On the other hand, since $\pi(B) \cap B = \emptyset$, we have

$$\begin{aligned}&\mathbb{P}(X(i,j) = X(\pi(i),\pi(j)) \text{ for all } 1 \le i < j \le m)\\ &\le \mathbb{P}(X(i,j) = X(\pi(i),\pi(j)) \text{ for all } i,j \in B,\, i<j)\\ &= 2^{-\binom{|B|}{2}} \le 2^{-\frac{1}{36}(m-k)^2+\frac{1}{4}},\end{aligned}$$

where the last inequality holds because $|B| \ge \frac{1}{3}(m-k)$ and $\binom{a}{2} \ge \frac{1}{4}(a^2-1)$ for any nonnegative integer a. The proof is now completed by combining the two bounds (e.g., by taking a suitable weighted geometric mean). □

Proof of Proposition 2.1. Throughout this proof, $C_1, C_2, \ldots$ will denote positive universal constants. Two of these, C_1 and C_2, are already fixed from Lemma 2.3. For $k = 0, \ldots, m$, let T_k be the set of all $\pi \in S_n$ such that $|\{i \le m : \pi(i) = i\}| = k$. Then by Lemma 2.3,

$$\begin{aligned}\phi(m,n) &= \sum_{k=0}^{m} \sum_{\pi \in T_k} \mathbb{P}(X(i,j) = X(\pi(i),\pi(j)) \text{ for all } 1 \le i < j \le m)\\ &\le C_1 \sum_{k=0}^{m} |T_k| e^{-C_2(m-k)m}. \qquad (2.3)\end{aligned}$$

Now, to choose an element of T_k, we can first choose the locations of the k fixed points of π in $\{1, \ldots, m\}$, and then choose the remaining part of π. The first task can be done in $\binom{m}{k}$ ways, and having done the first task, the second task can be done in $\le (n-k)!$ ways. Thus,

$$|T_k| \le \binom{m}{k}(n-k)!.$$

We will now use the above to get an upper bound for k^{th} term in (2.3). First, suppose that $2m-n \le k \le m$ (noting that $2m-n \ge 0$, since $m \ge 2n/3$). Then $n-k \le 2(n-m)$, and hence

$$\begin{aligned}|T_k| e^{-C_2(m-k)m} &\le \binom{m}{k}(n-k)! e^{-C_2(m-k)m}\\ &\le \binom{m}{k}(2(n-m))^{2(n-m)} e^{-C_2(m-k)m},\end{aligned}$$

interpreting $0^0 = 1$ if $n = m$. This gives

$$\begin{aligned}&\sum_{2m-n \le k \le m} |T_k| e^{-C_2(m-k)m}\\ &\le \sum_{2n-m \le k \le m} \binom{m}{k}(2(n-m))^{2(n-m)} e^{-C_2(m-k)m}\end{aligned}$$

$$\leq (2(n-m))^{2(n-m)} \sum_{k=0}^{m} \binom{m}{m-k} e^{-C_2(m-k)m}$$
$$= (2(n-m))^{2(n-m)}(1+e^{-C_2 m})^m \leq C_3 e^{C_4(n-m)\log(n-m)}. \tag{2.4}$$

Next, suppose that $0 \leq k < 2m-n$. Then $n-k < 2(m-k)$, which gives

$$\begin{aligned} |T_k| e^{-C_2(m-k)m} &\leq \binom{m}{k}(n-k)! e^{-C_2(m-k)m} \\ &\leq \binom{m}{k}(2(m-k))^{2(m-k)} e^{-C_2(m-k)m} \\ &\leq \binom{m}{k} e^{C_5(m-k)\log m - C_2(m-k)m}. \end{aligned}$$

This shows that there is a sufficiently large number n_0, such that if $n \geq n_0$ (which implies that $m \geq 2n_0/3$), and $0 \leq k < 2m-n$, then we have

$$|T_k| e^{-C_2(m-k)m} \leq \binom{m}{k} e^{-C_6(m-k)m}.$$

Therefore,

$$\begin{aligned} \sum_{0\leq k<2m-n} |T_k| e^{-C_2(m-k)m} &\leq \sum_{0\leq k<2m-n} \binom{m}{m-k} e^{-C_6(m-k)m} \\ &\leq (1+e^{-C_6 m})^m \leq C_7. \end{aligned} \tag{2.5}$$

Combining (2.3), (2.4), and (2.5), we get the desired upper bound for sufficiently large n. We then get it for all n by just increasing the value of K_1. □

We now start towards our proof of Theorem 1.1. Take any N, and let $a = 4/\log 2$, $b = -2/\log 2$ and $c := \frac{1}{2}a(1-\log a)$. An easy verification shows that

$$\begin{aligned} &a\log N + b\log\log N + c \\ &= 4\log_2 N - 2\log_2\log_2 N - 2\log_2(4/e) = x_N - 1. \end{aligned} \tag{2.6}$$

Choose any integer n so that $|x_N - n| \leq 2$, and write n as

$$n = a\log N + b\log\log N + d,$$

so that

$$c + 1 - d = x_N - n. \tag{2.7}$$

Let $X(i,j)$ and $Y(i,j)$ be the indicators that $\{i,j\}$ is an edge in Γ_1 and Γ_2, respectively. Let $\mathcal{A}$ be the set of all ordered n-tuples of distinct numbers from $\{1,\ldots,N\}$. For $A,B \in \mathcal{A}$, we will write $A \simeq B$ if $X(a_i,a_j) = Y(b_i,b_j)$ for all $1 \le i < j \le n$. Let

$$W := |\{A,B \in \mathcal{A} : A \simeq B\}|.$$

Note that $L_N \ge n$ if and only if $W > 0$. We will prove an upper bound on $\mathbb{P}(W>0)$ using the first moment method, and a lower bound using the second moment method. In the following, we adopt the convention that for any function f, $O(f(N))$ denotes any quantity whose absolute value is bounded above by a constant times $f(N)$, where the constant has no dependence on N.

Lemma 2.4. *Let all notation be as above. Then*

$$\mathbb{P}(W>0) \le e^{2(c+1-d)\log N + O((\log\log N)^2)}.$$

Proof. First, note that

$$\mathbb{E}(W) = |\mathcal{A}|^2 2^{-\binom{n}{2}} \le N^{2n} 2^{-\binom{n}{2}}.$$

Next, note that if $A \simeq B$, then $A_\pi \simeq B_\pi$ for any $\pi \in S_n$, where A_π and B_π denote the lists $(a_{\pi(1)},\ldots a_{\pi(n)})$ and $(b_{\pi(1)},\ldots,b_{\pi(n)})$, respectively. Thus, $W>0$ if and only if $W \ge n!$. This gives

$$\mathbb{P}(W>0) = \mathbb{P}(W \ge n!) \le \frac{\mathbb{E}(W)}{n!} \le N^{2n} 2^{-\binom{n}{2}} n^{-(n+\frac{1}{2})} e^{n+O(1)}.$$

Plugging in the value of n, we get

$$\begin{aligned}\mathbb{P}(W>0) \le \exp\Bigg(& 2(a\log N + b\log\log N + d)\log N \\ & - \frac{1}{2}(a\log N + b\log\log N + d)(a\log N + b\log\log N + d - 1)\log 2 \\ & - \left(a\log N + b\log\log N + d + \frac{1}{2}\right)\log(a\log N + b\log\log N + d) \\ & + a\log N + b\log\log N + d + O(1)\Bigg).\end{aligned}$$

The third line in the above display is a little bit more complicated than the rest. To simplify, let us use $\log(1+x) = x + O(x^2)$, which gives

$$\left(a\log N + b\log\log N + d + \frac{1}{2}\right)\log(a\log N + b\log\log N + d)$$

$$= \left(a\log N + b\log\log N + d + \frac{1}{2}\right)\left(\log(a\log N) + \log\left(1 + \frac{b\log\log N + d}{a\log N}\right)\right)$$
$$= \left(a\log N + b\log\log N + d + \frac{1}{2}\right)(\log\log N + \log a) + b\log\log N + O(1).$$

Plugging this into the previous display, let us compute the coefficients of various terms. First, note that the coefficient of $(\log N)^2$ is

$$2a - \frac{a^2\log 2}{2},$$

which is zero since $a = 4/\log 2$. Next, the coefficient of $(\log N)\log\log N$ is

$$2b - ab\log 2 - a,$$

which, again, is zero since $a = 4/\log 2$ and $b = -2/\log 2$. The next highest term is $\log N$, whose coefficient is

$$2d - \frac{a}{2}(2d-1)\log 2 - a\log a + a = 2 - a\log a + a - 2d$$
$$= 2 + 2c - 2d,$$

since $a = 4/\log 2$ and $c = \frac{1}{2}a(1 - \log a)$. All other terms are of order $(\log\log N)^2$ or smaller. This completes the proof. □

Next, we get a lower bound for $\mathbb{P}(W > 0)$ using the second moment method. For that, we need an upper bound on $\mathbb{E}(W^2)$. Let $\mathcal{A}_0$ be the set of all pairs $(A, B) \in \mathcal{A}^2$ such that $A \cap B = \emptyset$ (considering A and B as sets rather than n-tuples). For each $1 \le m \le n$, each $1 \le i_1 < \cdots < i_m \le n$, and each m-tuple of distinct $j_1, \ldots, j_m \in \{1, \ldots, n\}$, let $\mathcal{A}_{i_1,\ldots,i_m;j_1,\ldots,j_m}$ be the set of all $(A, B) \in \mathcal{A}^2$ such that $a_{i_1} = b_{j_1}, \ldots, a_{i_m} = b_{j_m}$, and $a_i \ne b_j$ for all $i \notin \{i_1, \ldots, i_m\}$ and $j \notin \{j_1, \ldots, j_m\}$. Note that these sets are disjoint, and their union, together with $\mathcal{A}_0$, equals $\mathcal{A}$. For $A, B, C, D \in \mathcal{A}$, let

$$P(A, B, C, D) := \mathbb{P}(A \simeq B,\, C \simeq D).$$

Then note that

$$\mathbb{E}(W^2) = \sum_{A,B,C,D\in\mathcal{A}} P(A, B, C, D)$$
$$= \sum_{(B,D)\in\mathcal{A}_0} \sum_{A,C\in\mathcal{A}} P(A, B, C, D)$$
$$+ \sum_{m=1}^{n} \sum_{1\le i_1<\cdots<i_m\le n} \sum_{\substack{1\le j_1,\ldots,j_m\le n\\ \text{distinct}}} \sum_{(B,D)\in\mathcal{A}_{i_1,\ldots,i_m;j_1,\ldots,j_m}} \sum_{A,C\in\mathcal{A}} P(A, B, C, D).$$

Let $\mathbb{P}'$ denote the conditional probability given Γ_1. If $(B, D) \in \mathcal{A}_0$, then

$$\mathbb{P}'(A \simeq B,\, C \simeq D) = 2^{-2\binom{n}{2}},$$

and thus, the unconditional probability is also the same. Next, for $(B, D) \in \mathcal{A}_{i_1,\ldots,i_m;j_1,\ldots,j_m}$, independence of edges implies that

$$\begin{aligned}
&\mathbb{P}'(A \simeq B,\, C \simeq D)\\
&= \mathbb{P}'(X(a_i, a_j) = Y(b_i, b_j) \text{ and } X(c_i, c_j) = Y(d_i, d_j) \text{ for all } 1 \le i < j \le n)\\
&= 2^{-2(\binom{n}{2}-\binom{m}{2})}\mathbb{P}'(X(a_{i_p}, a_{i_q}) = Y(b_{i_p}, b_{i_q}) \text{ and } X(c_{j_p}, c_{j_q}) = Y(d_{j_p}, d_{j_q})\\
&\qquad\qquad \text{for all } 1 \le p < q \le m)\\
&= 2^{-2(\binom{n}{2}-\binom{m}{2})}\mathbb{P}'(Y(b_{i_p}, b_{i_q}) = X(a_{i_p}, a_{i_q}) = X(c_{j_p}, c_{j_q})\\
&\qquad\qquad \text{for all } 1 \le p < q \le m)\\
&= 2^{-2\binom{n}{2}+\binom{m}{2}}\mathbb{I}_{\{X(a_{i_p},a_{i_q})=X(c_{j_p},c_{j_q}) \text{ for all } 1\le p<q\le m\}},
\end{aligned}$$

where $\mathbb{I}_E$ denotes the indicator of an event E, and in going from the third to the fourth line we used the fact that $b_{i_p} = d_{j_p}$ for each p. Thus,

$$P(A, B, C, D) = 2^{-2\binom{n}{2}+\binom{m}{2}}\mathbb{P}(X(a_{i_p}, a_{i_q}) = X(c_{j_p}, c_{j_q}) \text{ for all } 1 \le p < q \le m).$$

Let $p(a_{i_1}, \ldots, a_{i_m}; c_{j_1}, \ldots, c_{j_m})$ denote the probability on the right. Combining the above observations, we get

$$\begin{aligned}
\mathbb{E}(W^2) = |\mathcal{A}_0||\mathcal{A}|^2 2^{-2\binom{n}{2}}&\\
+ \sum_{m=1}^{n} \sum_{1\le i_1<\cdots<i_m\le n} \sum_{\substack{1\le j_1,\ldots,j_m\le n\\ \text{distinct}}} \sum_{A,C\in\mathcal{A}} &|\mathcal{A}_{i_1,\ldots,i_m;j_1,\ldots,j_m}|2^{-2\binom{n}{2}+\binom{m}{2}}\\
&\cdot p(a_{i_1}, \ldots, a_{i_m}; c_{j_1}, \ldots, c_{j_m}).
\end{aligned}$$

Note that

$$|\mathcal{A}_0| \le N^{2n}, \quad |\mathcal{A}| \le N^n, \quad |\mathcal{A}_{i_1,\ldots,i_m;j_1,\ldots,j_m}| \le N^{2n-m}.$$

Now, given m, $i_1, \ldots, i_m$, $j_1, \ldots, j_m$ and A, note that by symmetry,

$$\begin{aligned}
&\sum_{C\in\mathcal{A}} p(a_{i_1}, \ldots, a_{i_m}; c_{j_1}, \ldots, c_{j_m})\\
&= \sum_{C\in\mathcal{A}} p(1, \ldots, m; c_{j_1}, \ldots, c_{j_m})
\end{aligned}$$

$$= (N-m)_{n-m} \sum_{\substack{1 \le e_1,\dots,e_m \le N \\ \text{distinct}}} p(1,\dots,m; e_1,\dots,e_m),$$

where we used the standard notation $(x)_y = x(x-1)\cdots(x-y+1)$. Let $\xi(m,N)$ denote the last sum. Then, using the above information in the expression for $\mathbb{E}(W^2)$ displayed above, we get

$$\mathbb{E}(W^2) \le N^{4n} 2^{-2\binom{n}{2}} \biggl(1 + \sum_{m=1}^{n} N^{-2m} 2^{\binom{m}{2}} \binom{n}{m} (n)_m \xi(m,N)\biggr). \tag{2.8}$$

Our goal now is to get an upper bound for $\xi(m,N)$. Take any $1 \le m \le n$, and any distinct $1 \le e_1,\dots,e_m \le N$. Let $l := |\{e_1,\dots,e_m\} \cap \{1,\dots,m\}|$ and let $1 \le p_1 < \cdots < p_l \le m$ be the indices such that $e_{p_i} \in \{1,\dots,m\}$ for each i. Let $\widetilde{\mathbb{P}}$ denote the conditional probability given $(X(i,j))_{1 \le i < j \le m}$. Then

$$\begin{aligned} &\widetilde{\mathbb{P}}(X(p,q) = X(e_p, e_q) \text{ for all } 1 \le p < q \le m) \\ &= 2^{-\binom{m}{2} + \binom{l}{2}} \mathbb{I}_{\{X(p_r,p_s) = X(e_{p_r}, e_{p_s}) \text{ for all } 1 \le r < s \le l\}}. \end{aligned}$$

In the sum defining $\xi(m,N)$, $e_1,\dots,e_m$ can be chosen as follows. First, we choose l between 0 and m. Then, given l, we choose $1 \le p_1 < \cdots < p_l \le m$. Given $p_1,\dots,p_l$, we choose distinct numbers $e_{p_1},\dots,e_{p_l} \in \{1,\dots,m\}$. Finally, we choose the rest of the e_i's from outside $\{1,\dots,m\}$ so that they are distinct. Breaking up the sum in this manner (and rewriting $f_1 = e_{p_1},\dots,f_l = e_{p_l}$, and using symmetry to replace p_i by i for $i = 1,\dots,l$), we get

$$\xi(m,N) \le \sum_{l=0}^{m} 2^{-\binom{m}{2} + \binom{l}{2}} N^{m-l} \binom{m}{l} \biggl(\sum_{\substack{1 \le f_1,\dots,f_l \le m \\ \text{distinct}}} p(1,\dots,l; f_1,\dots,f_l) \biggr).$$

By Proposition 2.1, the inner sum is bounded by $K_1 e^{K_2 (m-l)\log(m-l)}$ if $l \ge 2m/3$. If $l < 2m/3$, it is trivially bounded by m^l. Combining, we get that the sum is bounded by $C_1 e^{C_2 \min\{l, m-l\} \log m}$ for some universal constants C_1 and C_2, which is bounded by $C_1 e^{C_2 \min\{l, n-l\} \log n}$ since $n \ge m$. The same bound holds for $\binom{m}{l}$. Plugging these into the above display, and then using the resulting bound on $\xi(m,N)$ in (2.8), we get

$$\begin{aligned} \mathbb{E}(W^2) \le N^{4n} 2^{-2\binom{n}{2}} \biggl(1 + \sum_{l=0}^{n} \sum_{m=\max\{1,l\}}^{n} N^{-m-l} 2^{\binom{l}{2}} \binom{n}{m} (n)_m \\ \cdot C_1 e^{C_2 \min\{l, n-l\} \log n} \biggr). \end{aligned}$$

In the following, we will use $C_1, C_2, \dots$ to denote arbitrary universal constants, whose values may change from line to line. Throughout, we will implicitly assume that N is

large enough (depending only on our choice of d'), wherever required. For the innermost sum above, consider three cases. First, take $l = 0$. Then the sum over m is bounded by

$$\sum_{m=1}^{n} N^{-m} n^{2m} C_1 \le C_2 n^2 N^{-1}.$$

Next, take $1 \le l \le 2n/3$. Then the sum over m is bounded by

$$\sum_{m=l}^{n} N^{-m-l} n^{2m} 2^{\binom{l}{2}} C_1 e^{C_2 l \log n} \le C_3 n^{2l} N^{-2l} 2^{\binom{l}{2}} e^{C_2 l \log n} \le C_4 e^{-C_5 l \log N},$$

where the last inequality was obtained using

$$N^{-2l} 2^{\binom{l}{2}} \le e^{-2l \log N} 2^{ln/3} \le e^{-2l \log N} 2^{(5l/3) \log_2 N},$$

which holds because $l \le 2n/3$, and $n \le 5 \log_2 N$ when N is large enough. Next, using that $n \ge \frac{7}{2} \log_2 N$ for N large enough, we have that for any $l \ge 2n/3$,

$$\frac{N^{-2(l+1)} 2^{\binom{l+1}{2}}}{N^{-2l} 2^{\binom{l}{2}}} = N^{-2} 2^l \ge N^{-2} 2^{(7/3) \log_2 N} = N^{1/3},$$

which implies, by backward induction on l, that

$$N^{-2l} 2^{\binom{l}{2}} \le N^{-(n-l)/3} N^{-2n} 2^{\binom{n}{2}}.$$

Thus, for $l \ge 2n/3$, the sum over m is bounded by

$$\begin{aligned}\sum_{m=l}^{n} N^{-m-l} 2^{\binom{l}{2}} \binom{n}{l} n! C_1 e^{C_2 (n-l) \log n} &\le C_3 n! N^{-2l} 2^{\binom{l}{2}} e^{C_4 (n-l) \log n} \\ &\le C_5 n! N^{-2n} 2^{\binom{n}{2}} e^{-C_6 (n-l) \log N},\end{aligned}$$

where, in the second inequality, we used the previous display and the fact that $\log N \gg \log n$ when N is large enough. Now, from the proof of Lemma 2.4, we have that

$$n! N^{-2n} 2^{\binom{n}{2}} = e^{-2(c+1-d) \log N + O((\log \log N)^2)}.$$

Combining all of the above (and trivially bounding $e^{-C_6 (n-l) \log N} \le 1$), we get

$$\mathbb{E}(W^2) \le N^{4n} 2^{-2\binom{n}{2}} (1 + e^{-2(c+1-d) \log N + O((\log \log N)^2)}).$$

On the other hand,

$$\mathbb{E}(W) = ((N)_n)^2 2^{-\binom{n}{2}} \geq N^{2n} 2^{-\binom{n}{2}} \left(1 - \frac{n}{N}\right)^{2n}$$
$$\geq N^{2n} 2^{-\binom{n}{2}} \left(1 - \frac{2n^2}{N}\right).$$

Thus, by the second moment inequality,

$$\mathbb{P}(W > 0) \geq \frac{(\mathbb{E}(W))^2}{\mathbb{E}(W^2)}$$
$$\geq \frac{(1 - 2n^2/N)^2}{1 + e^{-2(c+1-d)\log N + O((\log\log N)^2)}}.$$

Let $\varepsilon_N := (4\log_2 N)^{-1/2}$, as in the statement of Theorem 1.1. By Lemma 2.4, $\mathbb{P}(W > 0) \to 0$ if $c+1-d < -\varepsilon_N$, and by the above lower bound, $\mathbb{P}(W > 0) \to 1$ if $c+1-d > \varepsilon_N$. But by (2.7), $c + 1 - d = x_N - n$. Since $\varepsilon_N \in (0, 1/2]$, this proves Theorem 1.1.

Remark 2.5. To prove or disprove that L_N concentrates on one point instead of two, one needs to carefully analyze and refine the $O((\log\log N)^2)$ error term in the above analysis and replace it by some explicit term plus $o(1)$ error, so that when x_N is within $O((\log N)^{-1})$ of some integer n, one can show that $\mathbb{P}(W > 0)$ is close to neither 1 or 0.

3. Subgraph isomorphism

Deciding if a graph Γ_1 appears as an induced subgraph of Γ_2 is a basic problem in the world of image analysis (does this person appear in this crowd scene?), chemistry, and database query. The problem is NP complete but modern constraint satisfaction algorithms can handle Γ_1 with hundreds of nodes and Γ_2 with thousands. A comprehensive review of programs and applications is in [15]. The forthcoming book by Knuth [9] features subgraph isomorphism as a basic problem of constraint satisfaction.

Comparing algorithms requires a suite of test problems. The authors of [15] noted that most tests were done on 'easy cases' where Γ_2 is fixed and Γ_1 is chosen by choosing a random set of n vertices of Γ_2 and taking that induced subgraph (so, Γ_1 appears in Γ_2). They noticed that taking Γ_1, Γ_2 from independent copies of $G(n,p)$, $G(N,q)$ led to different recommendations and conclusions.

As part of their extensive tests they fixed $N = 150$ and discovered the phase transition discussed in the introduction. Their results are much richer when p and q are varied — we only treat $p = q = 1/2$. We believe the techniques introduced in this paper will allow similar limit theorems (at least for p, q away from $\{0, 1\}$).

As mentioned earlier in the introduction, a closely related recent paper is that of Alon [1], which shows that if $\binom{N}{n}2^{-\binom{n}{2}} = \lambda$, then with probability $(1 - e^{-\lambda})^2 + o(1)$, $G(N, 1/2)$ contains every graph on n vertices as an induced subgraph. Theorem 1.2 is also related to the classical result about the concentration of the size of the largest clique in $G(N, 1/2)$, due to Matula [12,13] and Bollobás and Erdős [3].

Take any $1 \le n \le N$. Let Γ_1 and Γ_2 be independent $G(n, 1/2)$ and $G(N, 1/2)$ random graphs. In the remainder of this section, we prove Theorem 1.2. The proof is similar to that of Theorem 1.1 (in particular, we use Proposition 2.1), although a bit simpler because the expected number of copies of Γ_1 in Γ_2 has a simpler expression that the expected number of isomorphic pairs of induced subgraphs in two independent random graphs (which has an extra $n!$ in the denominator, leading to the $\log\log N$ correction).

Let $X(i,j)$ be the indicator that $\{i,j\}$ is an edge in Γ_1, and $Y(i,j)$ be the indicator that $\{i,j\}$ is an edge in Γ_2. Let $\mathcal{A}$ be the set of all ordered n-tuples of distinct numbers from $\{1,\ldots,N\}$, as in the previous section. We will write $A \simeq \Gamma_1$ if $Y(a_i, a_j) = X(i,j)$ for all $1 \le i < j \le n$. Let

$$W := |\{A \in \mathcal{A} : A \simeq \Gamma_1\}|.$$

Note that Γ_2 contains a copy of Γ_1 as an induced subgraph if and only if $W > 0$.

Lemma 3.1. *Suppose that $n = a\log N + b$ for $a = 2/\log 2$ and some $b \in \mathbb{R}$. Then*

$$\mathbb{P}(W > 0) \le N^{1-b}2^{-b(b-1)/2}.$$

Proof. Note that

$$\mathbb{E}(W) = |\mathcal{A}|2^{-\binom{n}{2}} \le N^n 2^{-\binom{n}{2}}.$$

Plugging in $n = a\log N + b$, this gives

$$\begin{aligned}\mathbb{E}(W) &\le \exp\left((a\log N + b)\log N - \frac{1}{2}(a\log N + b)(a\log N + b - 1)\log 2\right)\\ &= N^{1-b}2^{-b(b-1)/2}.\end{aligned}$$

By Markov's inequality, $\mathbb{P}(W > 0) = \mathbb{P}(W \ge 1) \le \mathbb{E}(W)$. This completes the proof. □

For each $(A, B) \in \mathcal{A}^2$, let

$$P(A,B) := \mathbb{P}(A \simeq \Gamma_1,\ B \simeq \Gamma_1).$$

Let $\mathcal{A}_0$ and $\mathcal{A}_{i_1,\ldots,i_m;j_1,\ldots,j_m}$ be as in the previous section. Then

$$\begin{aligned}\mathbb{E}(W^2) &= \sum_{A,B\in\mathcal{A}} P(A,B)\\ &= \sum_{(A,B)\in\mathcal{A}_0} P(A,B)\\ &+ \sum_{m=1}^{n} \sum_{1\le i_1<\cdots<i_m\le n} \sum_{\substack{1\le j_1,\ldots,j_m\le n\\ \text{distinct}}} \sum_{(A,B)\in\mathcal{A}_{i_1,\ldots,i_m;j_1,\ldots,j_m}} P(A,B). \qquad (3.1)\end{aligned}$$

We will now use the above expression to establish an upper bound for $\mathbb{E}(W^2)$.

Lemma 3.2. *We have*

$$\mathbb{E}(W^2) \le N^{2n} 2^{-2\binom{n}{2}} \biggl(1 + \sum_{m=1}^{n} \sum_{\substack{1 \le j_1, \ldots, j_m \le n \\ \text{distinct}}} \binom{n}{m} N^{-m} 2^{\binom{m}{2}} p(j_1, \ldots, j_m)\biggr),$$

where

$$p(j_1, \ldots, j_m) := \mathbb{P}(X(p,q) = X(j_p, j_q) \text{ for all } 1 \le p < q \le m).$$

Proof. Let $\mathbb{P}'$ denote the conditional probability given Γ_1. If $(A, B) \in \mathcal{A}_0$, then

$$\mathbb{P}'(A \simeq \Gamma_1,\, B \simeq \Gamma_1) = 2^{-2\binom{n}{2}},$$

and thus, the unconditional probability is also the same. Next, for $(A, B) \in \mathcal{A}_{i_1,\ldots,i_m;j_1,\ldots,j_m}$, we have

$$\begin{aligned}
&\mathbb{P}'(A \simeq \Gamma_1,\, B \simeq \Gamma_1) \\
&= \mathbb{P}'(Y(a_i, a_j) = Y(b_i, b_j) = X(i,j) \text{ for all } 1 \le i < j \le n) \\
&= 2^{-2(\binom{n}{2} - \binom{m}{2})} \mathbb{P}'(Y(a_{i_p}, a_{i_q}) = X(i_p, i_q) = X(j_p, j_q) \text{ for all } 1 \le p < q \le m) \\
&= 2^{-2\binom{n}{2} + \binom{m}{2}} \mathbb{I}_{\{X(i_p,i_q) = X(j_p,j_q) \text{ for all } 1 \le p < q \le m\}},
\end{aligned}$$

where $\mathbb{I}_E$ denotes the indicator of an event E. Thus,

$$P(A, B) = 2^{-2\binom{n}{2} + \binom{m}{2}} \mathbb{P}(X(i_p, i_q) = X(j_p, j_q) \text{ for all } 1 \le p < q \le m).$$

Combining the above observations, we get

$$\begin{aligned}
&\mathbb{E}(W^2) \\
&= |\mathcal{A}_0| 2^{-2\binom{n}{2}} + \sum_{m=1}^{n} \sum_{i_1 < \cdots < i_m} \sum_{\substack{j_1, \ldots, j_m \\ \text{distinct}}} |\mathcal{A}_{i_1,\ldots,i_m;j_1,\ldots,j_m}| 2^{-2\binom{n}{2} + \binom{m}{2}} \\
&\qquad \cdot \mathbb{P}(X(i_p, i_q) = X(j_p, j_q) \text{ for all } 1 \le p < q \le m).
\end{aligned}$$

Note that

$$|\mathcal{A}_0| \le N^{2n}, \quad |\mathcal{A}_{i_1,\ldots,i_m;j_1,\ldots,j_m}| \le N^{2n-m}.$$

Plugging these bounds into the previous display and using symmetry, we get the desired result. □

Henceforth, let us assume that

$$n = a \log N + b \quad \text{for} \quad a = \frac{2}{\log 2} \quad \text{and some} \quad b \in [-1, 1]. \tag{3.2}$$

Lemma 3.3. *Assume (3.2), and suppose that $m = \alpha n$ for some $\alpha \in [0, 1]$. Then*

$$-m \log N + \binom{m}{2} \log 2 \le -\alpha(1-\alpha) n \log N - \alpha(1 - \alpha b) \log N + \log 2.$$

Proof. Note that

$$\begin{aligned}
&- m \log N + \binom{m}{2} \log 2 \\
&= -\alpha a (\log N)^2 - \alpha b \log N + \frac{(\alpha a \log N + \alpha b)(\alpha a \log N + \alpha b - 1)}{2} \log 2 \\
&= -\alpha(1-\alpha) a (\log N)^2 + \alpha((2\alpha - 1)b - 1) \log N + \frac{\alpha b(\alpha b - 1)}{2} \log 2 \\
&\le -\alpha(1-\alpha)(a \log N + b) \log N - \alpha(1 - \alpha b) \log N + \log 2,
\end{aligned}$$

where in the last inequality we used $\alpha b(\alpha b - 1) \le 2$, which holds because $b \in [-1, 1]$ and $\alpha \in [0, 1]$. □

Lemma 3.4. *There are positive universal constants C_1, C_2 and N_0, such that if $N \ge N_0$, and (3.2) holds, then*

$$\mathbb{E}(W^2) \le N^{2n} 2^{-2\binom{n}{2}} (1 + C_1 N^{-C_2(1-b)}).$$

Proof. In this proof, $C_1, C_2, \ldots$ will denote arbitrary positive universal constants. Let $p(j_1, \ldots, j_m)$ be as in Lemma 3.2. First, note that

$$\sum_{\substack{j_1, \ldots, j_m \\ \text{distinct}}} p(j_1, \ldots, j_m) = \frac{\phi(m, n)}{(n-m)!},$$

where $\phi(m, n)$ is the quantity defined in (2.1). Therefore, by Proposition 2.1,

$$\begin{aligned}
&\sum_{2n/3 \le m \le n} \sum_{\substack{j_1, \ldots, j_m \\ \text{distinct}}} \binom{n}{m} N^{-m} 2^{\binom{m}{2}} p(j_1, \ldots, j_m) \\
&\le C_1 \sum_{2n/3 \le m \le n} N^{-m} 2^{\binom{m}{2}} e^{C_2(n-m)\log n}.
\end{aligned}$$

For $m \ge 2n/3$, Lemma 3.3 gives (using $(1-\alpha)n = n - m$ and $\alpha b \le (5b+1)/6$ for all $\alpha \in [2/3, 1]$ and $b \in [-1, 1]$)

$$-m\log N + \binom{m}{2}\log 2 \le -\alpha(n-m)\log N - \frac{5\alpha}{6}(1-b)\log N + \log 2$$
$$\le -\frac{2}{3}(n-m)\log N - \frac{5}{9}(1-b)\log N + \log 2.$$

Thus, for $N \ge N_0$, where N_0 is a sufficiently large universal constant, we have

$$\sum_{2n/3\le m\le n} N^{-m} 2^{\binom{m}{2}} e^{C_2(n-m)\log n}$$
$$\le \sum_{2n/3\le m\le n} 2e^{-C_3(n-m)\log N} N^{-5(1-b)/9} \le C_4 N^{-5(1-b)/9}.$$

If $1 \le m < 2n/3$, then Lemma 3.3 gives (using $\alpha n = m$ and $\alpha(1-\alpha b) \ge 0$)

$$-m\log N + \binom{m}{2}\log 2 \le -(1-\alpha)m\log N + \log 2$$
$$\le -\frac{1}{3}m\log N + \log 2.$$

Thus, using the trivial bound $p(j_1,\dots,j_m) \le 1$ and the assumption (3.2), we get

$$\sum_{1\le m<2n/3} \sum_{\substack{j_1,\dots,j_m \\ \text{distinct}}} \binom{n}{m} N^{-m} 2^{\binom{m}{2}} p(j_1,\dots,j_m)$$
$$\le \sum_{1\le m<2n/3} 2N^{-m/3} n^{2m} \le 2N^{-C_5},$$

provided that $N \ge N_0$ for some sufficiently large universal constant N_0. Combining all of the above, and applying Lemma 3.2, we get the required upper bound. □

Lemma 3.5. *There are positive universal constants N_0, C_1 and C_2 such that the following is true. If $N \ge N_0$ and (3.2) holds, then*

$$\mathbb{P}(W \ge 1) \ge 1 - C_1 N^{-C_2(1-b)}.$$

Proof. Note that

$$\mathbb{E}(W) = (N)_n 2^{-\binom{n}{2}} \ge N^n 2^{-\binom{n}{2}} \left(1 - \frac{n}{N}\right)^n$$
$$\ge N^n 2^{-\binom{n}{2}} \left(1 - \frac{n^2}{N}\right).$$

Combining this with the upper bound on $\mathbb{E}(W^2)$ from Lemma 3.4, and the second moment inequality, we get the desired result. □

Proof of Theorem 1.2. Note that $\lfloor y_N + \varepsilon_N \rfloor + 1 \geq 2\log_2 N + 1 + \varepsilon_N$. Therefore, by Lemma 3.1, $P(\lfloor y_N + \varepsilon_N \rfloor + 1, N) \to 0$ as $N \to \infty$. Similarly, note that $\lfloor y_N - \varepsilon_N \rfloor \leq 2\log_2 N + 1 - \varepsilon_N$. By Lemma 3.5, this shows that $P(\lfloor y_N - \varepsilon_N \rfloor, N) \to 1$ as $N \to \infty$. □

4. Remarks and problems

Theorem 1.1 does not capture the way the induced isomorphic subgraph varies from N to $N+1$. If Γ_1 and Γ_2 are grown by adding fresh vertices one at a time, it may well be that that largest isomorphic subgraph varies quite a bit (eventually becoming disjoint from earlier champions?). This makes the connection with the limiting R more tenuous and seems worth further study.

It seems natural to ask similar questions for other graph limit models [11]. In particular, all $G(\infty, p)$ graphs are isomorphic to the Rado graph, for any $p \in (0, 1)$. This makes understanding the largest isomorphic induced subgraph of two independent $G(N, p)$ graphs more interesting.

There are a variety of notions of Γ_1 being contained in Γ_2. Just isomorphic as a subgraph (without the 'induced' constraint)? As labeled graphs? The classic paper [2] relates such problems to the problem of finding maximal cliques.

We have focused on the yes/no question. There are further counting questions — how many copies of a pick from $G(n, p)$ appear in $G(N, q)$ and how is this number distributed? (See forthcoming work of Surya, Warnke and Zhu for a solution of this problem.)

Finally, Knuth's treatment [9] treats subgraph isomorphism as a special case of constraint satisfaction problems, and similar questions can be asked.

Data availability

No data was used for the research described in the article.

References

[1] N. Alon, Asymptotically optimal induced universal graphs, Geom. Funct. Anal. 27 (1) (2017) 1–32.
[2] H.G. Barrow, R.M. Burstall, Subgraph isomorphism, matching relational structures, and maximal cliques, Inf. Process. Lett. 4 (4) (1976) 83–84.
[3] B. Bollobás, P. Erdös, Cliques in random graphs, Math. Proc. Camb. Philos. Soc. 80 (3) (1976) 419–427.
[4] P.J. Cameron, The random graph, in: The Mathematics of Paul Erdős, II, Springer, Berlin, 1997, pp. 333–351.
[5] A. D'Aristotile, P. Diaconis, D. Freedman, On merging of probabilities, Sankhyā, Ser. B 50 (3) (1988) 363–380.
[6] P. Diaconis, M. Malliaris, Complexity and randomness in the Heisenberg groups (and beyond), arXiv preprint, arXiv:2107.02923, 2021.
[7] A. Dudek, J. Grytczuk, A. Ruciński, Variations on twins in permutations, arXiv preprint, arXiv: 2001.05589, 2020.
[8] A. Heckel, Non-concentration of the chromatic number of a random graph, J. Am. Math. Soc. 34 (1) (2021) 245–260.
[9] D.E. Knuth, The art of computer programming, Vol. 4C, 2021, forthcoming.

[10] C. Lee, P.S. Loh, B. Sudakov, Self-similarity of graphs, SIAM J. Discrete Math. 27 (2) (2013) 959–972.
[11] L. Lovász, Large Networks and Graph Limits, American Mathematical Society, Providence, RI, 2012.
[12] D.W. Matula, On the complete subgraphs of a random graph, in: Proc. Second Chapel Hill Conf. on Combinatorial Mathematics and Its Applications, Univ. North Carolina, Chapel Hill, N.C., 1970, Univ. North Carolina, Chapel Hill, N.C., 1970, pp. 356–369.
[13] D.W. Matula, The employee party problem, Not. Am. Math. Soc. 19 (2) (1972) A–382.
[14] C. McCreesh, P. Prosser, J. Trimble, A partitioning algorithm for maximum common subgraph problems, in: Proc. 26th Internat. Joint Conf. Artific. Intell. (IJCAI), 2017, pp. 712–719.
[15] C. McCreesh, P. Prosser, C. Solnon, J. Trimble, When subgraph isomorphism is really hard, and why this matters for graph databases, J. Artif. Intell. Res. 61 (2018) 723–759.