

Isomorphisms between dense random graphs

Erlang Surya*

Lutz Warnke†

Emily Zhu‡

June 13, 2023; revised January 28, 2025

Abstract

We consider two variants of the induced subgraph isomorphism problem for two independent binomial random graphs with constant edge-probabilities p_1, p_2 . In particular, (i) we prove a sharp threshold result for the appearance of G_{n,p_1} as an induced subgraph of G_{N,p_2} , (ii) we show two-point concentration of the size of the maximum common induced subgraph of G_{N,p_1} and G_{N,p_2} , and (iii) we show that the number of induced copies of G_{n,p_1} in G_{N,p_2} has an unusual limiting distribution.

These results confirm simulation-based predictions of McCreesh, Prosser, Solnon and Trimble, and resolve several open problems of Chatterjee and Diaconis. The proofs are based on careful refinements of the first and second moment method, using extra twists to (a) take some non-standard behaviors into account, and (b) work around the large variance issues that prevent standard applications of these methods.

1 Introduction

Applied benchmark tests for the famous ‘subgraph isomorphism problem’ empirically discovered interesting phase transitions in random graphs. More concretely, these phase transitions were observed in two induced variants of the ‘subgraph containment problem’ widely-studied in random graph theory. In this paper we prove that the behavior of these two new random graph problems is surprisingly rich, with unexpected phenomena such as (a) that the form of the answer changes for constant edge-probabilities, (b) that the classical second moment method fails due to large variance, and (c) that an unusual limiting distribution arises.

To add more context, in many applications such as pattern recognition, computer vision, biochemistry and molecular science, it is a fundamental problem to determine whether an induced copy of a given graph F (or a large part of a given graph F) is contained in another graph G ; see [8, 19, 7, 9, 10, 13, 5, 17]. In this paper we consider two probabilistic variants of this problem, where the two graphs F and G are both independent binomial random graphs with constant edge-probabilities $p_1, p_2 \in (0, 1)$. Our main results are threefold:

- We prove a sharp threshold result for the appearance of G_{n,p_1} as an induced subgraph of G_{N,p_2} , and discover that the sharpness differs between the cases $p_2 = 1/2$ and $p_2 \neq 1/2$; see Theorem 1 and Corollary 3.
- We show that the number of induced copies of G_{n,p_1} in G_{N,p_2} has a Poisson limiting distribution for $p_2 = 1/2$, and a ‘squashed’ log-normal limiting distribution for $p_2 \neq 1/2$; see Theorem 2.
- We show two-point concentration of the maximum common induced subgraph of G_{N,p_1} and G_{N,p_2} , and discover that the form of the maximum size changes as we vary $p_1, p_2 \in (0, 1)$; see Theorem 4.

These results confirm simulation-based phase transition predictions of McCreesh, Prosser, Solnon and Trimble [18, 17], and resolve several open problems of Chatterjee and Diaconis [6]. Our proofs are based on careful refinements of the first and second moment method, using several extra twists to (a) take the non-standard phenomena into account, and (b) work around the large variance issues that prevent standard applications of these moment based methods, using in particular pseudorandom properties and multi-round exposure arguments to tame the variance.

*Department of Mathematics, University of California at San Diego, La Jolla, CA 92093, USA. Email: esurya@ucsd.edu. Supported by NSF CAREER grant DMS-1945481.

†Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, USA. E-mail: lwarnke@ucsd.edu. Supported by NSF CAREER grant DMS-1945481, and a Sloan Research Fellowship.

‡Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, USA. E-mail: e9zhu@ucsd.edu. Supported by NSF Graduate Research Fellowship Program grant DGE-2038238, and a Sloan Research Fellowship.

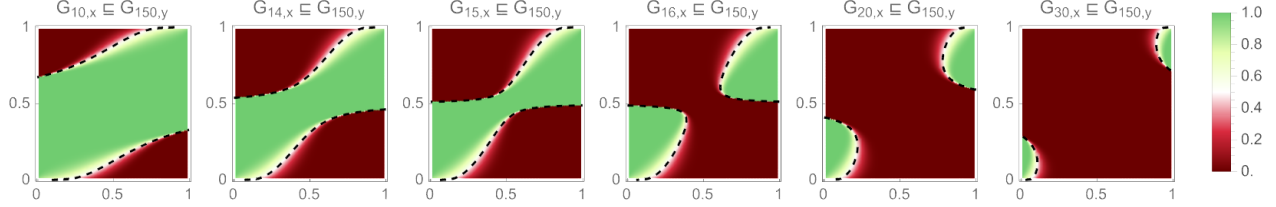


Figure 1: Theorem 1 establishes a sharp threshold around $n^* = n^*(p_1, p_2, N) := 2 \log_a N + 1$ for the appearance of the binomial random graph G_{n,p_1} as an induced subgraph of the independent random graph G_{N,p_2} . It also yields the induced containment probability estimate $\mathbb{P}(G_{n^*+c,p_1} \subseteq G_{N,p_2}) \approx f_{p_1,p_2}(c)$, which allows us to reproduce an idealized version of Figure 5 in [17], where $\mathbb{P}(G_{n,x} \subseteq G_{N,y})$ with $N = 150$ is empirically plotted for all $x, y \in [0, 1]$ and $n \in \{10, 14, 15, 16, 20, 30\}$; the dashed line corresponds to the threshold n^* . Previous work [6] applied to the special case $p_1 = p_2 = 1/2$, i.e., only reproduced the central point in each plot.

1.1 Induced subgraph isomorphism problem for random graphs

In the *induced subgraph isomorphism problem* the objective is to determine whether F is isomorphic to an induced subgraph of G , i.e., whether G contains an induced copy of F . In this paper we focus on this problem for two independent binomial random graphs $F = G_{n,p_1}$ and $G = G_{N,p_2}$, with constant edge-probabilities $p_1, p_2 \in (0, 1)$ and $n \ll N$ many vertices. The motivation here is that, in applied work on benchmark tests for this NP-hard problem, it was empirically discovered [18, 17] that such random graphs can be used to generate algorithmically hard problem-instances, leading to intriguing phase transitions. Knuth [14, 6] asked for mathematical explanations of these phase transitions, which are illustrated in Figure 1 via containment probability phase-diagram plots. The central points of these plots for $p_1 = p_2 = 1/2$ were resolved by Chatterjee and Diaconis [6] (and independently by Alon [1, 2]), who emphasized in talks that the more interesting general case seems substantially more complicated. In this paper we resolve the general case with edge-probabilities $(p_1, p_2) \in (0, 1)^2$: besides explaining the phase-diagram plots in Figure 1, we discover that the non-uniform case $p_2 \neq 1/2$ gives rise to new phenomena not anticipated by earlier work, including a different sharpness of the phase transition and an unusual ‘squashed’ log-normal limiting distribution.

Our first result establishes a sharp threshold for the appearance of the binomial random graph G_{n,p_1} as an induced subgraph of the independent random graph G_{N,p_2} , resolving an open problem of Chatterjee and Diaconis [6]. Below the abbreviation $F \subseteq G$ means that G contains an induced copy of F .

Theorem 1 (Sharp threshold). *Let $p_1, p_2 \in (0, 1)$ be constants. Define $a := 1/(p_2^{p_1}(1-p_2)^{1-p_1})$ and $\varepsilon_N := (\log \log N)^2 / \log N$. Then the following holds, for independent binomial random graphs G_{n,p_1} and G_{N,p_2} :*

(i) *If $p_2 = 1/2$, then $a = 2$ and*

$$\lim_{N \rightarrow \infty} \mathbb{P}(G_{n,p_1} \subseteq G_{N,p_2}) = \begin{cases} 1 & \text{if } n \leq 2 \log_a N + 1 - \varepsilon_N, \\ 0 & \text{if } n \geq 2 \log_a N + 1 + \varepsilon_N. \end{cases} \quad (1)$$

(ii) *If $p_2 \neq 1/2$, then*

$$\lim_{N \rightarrow \infty} \mathbb{P}(G_{n,p_1} \subseteq G_{N,p_2}) = \begin{cases} 1 & \text{if } n - (2 \log_a N + 1) \rightarrow -\infty, \\ f(c) & \text{if } n - (2 \log_a N + 1) \rightarrow c \in (-\infty, +\infty), \\ 0 & \text{if } n - (2 \log_a N + 1) \rightarrow \infty, \end{cases} \quad (2)$$

where $f(c) = f_{p_1,p_2}(c) := \mathbb{P}(N(0, \sigma^2) \geq c)$ for a normal random variable $N(0, \sigma^2)$ with mean 0 and variance $\sigma^2 := 2p_1(1-p_1)\log_a^2(1/p_2 - 1)$.

In concrete words, Theorem 1 shows that, around the threshold $n \approx 2 \log_a N + 1$, the induced containment probability $\mathbb{P}(G_{n,p_1} \subseteq G_{N,p_2})$ drops from $1 - o(1)$ to $o(1)$ in a window of size at most two when $p_2 = 1/2$, whereas this window has unbounded size when $p_2 \neq 1/2$. Our proof explains why this new phenomenon happens in the non-uniform case $p_2 \neq 1/2$: here the asymmetry of edges and non-edges

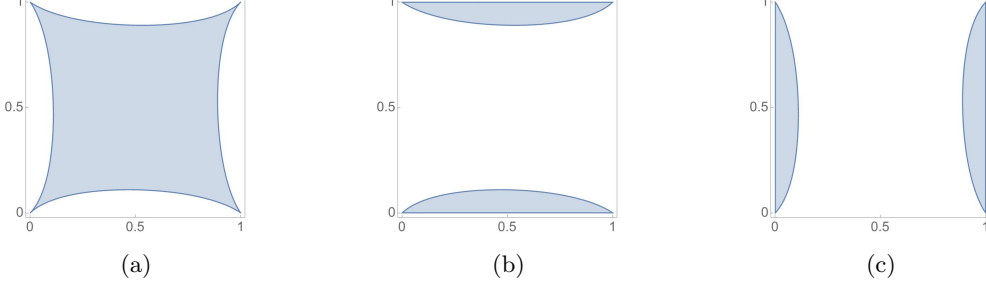


Figure 2: Theorem 4 establishes two-point concentration around n_N of the maximum common induced subgraph of two independent binomial random graphs G_{N,p_1} and G_{N,p_2} . The plots illustrate how the form of n_N subdivides the unit square of edge-probabilities $(p_1, p_2) \in (0, 1)^2$: we have $n_N = x_N^{(0)}(\hat{p}) \sim 4 \log_{b_0(\hat{p})} N$ in region (a), $n_N \sim x_N^{(1)}(p_1^*) \sim 2 \log_{b_1(p_1^*)} N$ in region (b), and $n_N \sim x_N^{(2)}(p_2^*) \sim 2 \log_{b_2(p_2^*)} N$ in region (c), where \hat{p} is defined as in (9) and p_i^* is the unique solution of $\log b_0(p) = 2 \log b_i(p)$; see Lemma 9 in Appendix A. Previous work [6] applied to the special case $p_1 = p_2 = 1/2$, i.e., only to the central point in region (a).

makes $\mathbb{P}(G_{n,p_1} \subseteq G_{N,p_2} \mid G_{n,p_1})$ strongly dependent on the number of edges in G_{n,p_1} , which does not occur in the uniform case $p_2 = 1/2$ considered in previous work [6]. This edge-deviation effect turns out to be the driving force behind the limiting probability $f(c) \in (0, 1)$ in (2); see Section 1.3.1 for more proof heuristics.

Theorem 1 confirms the simulation based predictions of McCreesh, Prosser, Solnon and Trimble [18, 17], who empirically plotted the induced containment probability $\mathbb{P}(G_{n,x} \subseteq G_{N,y})$ for $N = 150$ and predicted a phase transition near $n \approx 2 \log_a N$; see Figure 5 and Section 3.1 in [17]. Figure 1 illustrates that the fuzziness they found near their predicted threshold can be explained by the limiting probability $f(c)$ in (2), whose existence was not predicted in earlier work (of course, the ‘small number of vertices’ effect also leads to some discrepancies in the plots, in particular for very small and large edge-probabilities in Figure 1).

The classical problems of determining the size of the largest independent set and clique of G_{N,p_2} are both related to Theorem 1, as they would correspond to the excluded edge-probabilities $p_1 \in \{0, 1\}$. These two classical parameters $\alpha(G_{N,p_2})$ and $\omega(G_{N,p_2})$ are well-known [4, 15] to typically have size $2 \log_a N - \Theta(\log \log N)$ for $a \in \{1/(1-p_2), 1/p_2\}$, and this additive $\Theta(\log \log N)$ left-shift compared to the threshold from Theorem 1 stems from an important conceptual difference: n -vertex cliques and independent sets have an automorphism group of size $n!$, whereas G_{n,p_1} typically has a trivial automorphism group. This is one reason why our proof needs to take pseudorandom properties of random graphs into account; see also Sections 1.3.1 and 1.3.3.

As a consequence of our proof of Theorem 1, we are able to determine the asymptotic distribution of the number of induced copies of G_{n,p_1} in G_{N,p_2} , resolving another open problem of Chatterjee and Diaconis [6]. In concrete words, Theorem 2 (i) shows that the number of induced copies has a Poisson distribution for $p_2 = 1/2$ and n close to the sharp threshold location $2 \log_a N + 1$. Furthermore, Theorem 2 (ii) shows that the number of induced copies has a ‘squashed’ log-normal distribution for $p_2 \neq 1/2$ and $n = 2 \log_a N + \Theta(1)$, which is a rather unusual limiting distribution for random discrete structures (that intuitively arises since the number of such induced copies is so strongly dependent on the number of edges in G_{n,p_1} ; see Section 2.2). Below $(N)_n := N(N-1) \dots (N-n+1)$ denotes the falling factorial.

Theorem 2 (Asymptotic distribution). *Let $p_1, p_2 \in (0, 1)$ be constants. Define $a := 1/(p_2^{p_1}(1-p_2)^{1-p_1})$ and $\varepsilon_N := (\log \log N)^2 / \log N$ as in Theorem 1. For independent binomial random graphs G_{n,p_1} and G_{N,p_2} , let $X = X_{n,N}$ denote the number of induced copies of G_{n,p_1} in G_{N,p_2} . Then the following holds, as $N \rightarrow \infty$:*

- (i) *If $p_2 = 1/2$ and $2 \log_a N - 1 + \varepsilon_N \leq n \leq N$, then X has asymptotically Poisson distribution with mean $\mu = \mu_{n,N} := (N)_n 2^{-\binom{n}{2}} = (N \cdot 2^{-(n-1)/2 + O(n/N)})^n$, i.e.,*

$$d_{TV}(X, \text{Po}(\mu)) \rightarrow 0. \quad (3)$$

- (ii) *If $p_2 \neq 1/2$ and $n - (2 \log_a N + 1) \rightarrow c \in (-\infty, \infty)$ as $N \rightarrow \infty$, then*

$$\frac{\log(1+X)}{\log N} \xrightarrow{d} \text{SN}\left(-c, 2p_1(1-p_1)\log_a^2(1/p_2-1)\right), \quad (4)$$

where $\text{SN}(\mu, \sigma^2)$ has cumulative distribution function $F(x) := \mathbb{1}_{\{x \geq 0\}} \mathbb{P}(\text{N}(\mu, \sigma^2) \leq x)$.

For the uniform case $p_2 = 1/2$, Svante Janson pointed out that one can sharpen the containment probability estimate from Theorem 1 (i) a little further, by simply invoking the limiting Poisson result (3) for those n to which (1) does not apply. This readily leads to the following refinement of Theorem 1 (i), which (a) shows that in (1) we only need to assume $\varepsilon_N \cdot \log N \rightarrow \infty$, and (b) also determines the limiting probability inside the ‘transition window’ that was missing in (1) and previous work of Chatterjee and Diaconis [6].

Corollary 3 (Refined sharp threshold in special case $p_2 = 1/2$). *Let $p_1 \in (0, 1)$ be a constant. Then the following refinement of (1) holds, for independent binomial random graphs G_{n,p_1} and $G_{N,1/2}$:*

$$\lim_{N \rightarrow \infty} \mathbb{P}(G_{n,p_1} \sqsubseteq G_{N,1/2}) = \begin{cases} 1 & \text{if } (n - (2 \log_2 N + 1)) \log_2 N \rightarrow -\infty, \\ 1 - e^{-2^{-c}} & \text{if } (n - (2 \log_2 N + 1)) \log_2 N \rightarrow c \in (-\infty, \infty), \\ 0 & \text{if } (n - (2 \log_2 N + 1)) \log_2 N \rightarrow \infty. \end{cases} \quad (5)$$

1.2 Maximum common induced subgraph problem for random graphs

In the *maximum common induced subgraph problem* the objective is to determine the maximum size of an induced subgraph of F that is isomorphic to an induced subgraph of G , where the maximum size is with respect to the number of vertices (this generalizes the induced subgraph isomorphism problem, since the maximum common induced subgraph is F if and only if F is isomorphic to an induced subgraph of G). In this paper we focus on this problem for two independent binomial random graphs $F = G_{N,p_1}$ and $G = G_{N,p_2}$, with constant edge-probabilities $p_1, p_2 \in (0, 1)$. One motivation here comes from combinatorial probability theory [6], where the following paradox was recently pointed out: two independent infinite random graphs with edge-probabilities $p_1, p_2 \in (0, 1)$ are isomorphic with probability one, but two independent finite random graphs G_{N,p_1} and G_{N,p_2} are isomorphic with probability tending to zero as $N \rightarrow \infty$. This discontinuity of limits raises the question of finding the size of the maximum common induced subgraph of G_{N,p_1} and G_{N,p_2} , which also is a natural random graph problem in its own right. Chatterjee and Diaconis [6] answered this question in the special case $p_1 = p_2 = 1/2$. In this paper we resolve the general case with edge-probabilities $(p_1, p_2) \in (0, 1)^2$: we discover that the general form of the maximum size is significantly more complicated than for the uniform case, which in fact is closely linked to large variance difficulties.

Our next result establishes two-point concentration of the size I_N of the maximum common induced subgraph of two independent binomial random graphs G_{N,p_1} and G_{N,p_2} , resolving an open problem of Chatterjee and Diaconis [6]. In concrete words, Theorem 4 shows that I_N equals, with probability tending to one as $N \rightarrow \infty$, one of (at most) two consecutive integers; see (6)–(7) below.

Theorem 4 (Two-point concentration). *Let $p_1, p_2 \in (0, 1)$ be constants. For independent binomial random graphs G_{N,p_1} and G_{N,p_2} , define I_N as the size of the maximum common induced subgraph. Then*

$$\lim_{N \rightarrow \infty} \mathbb{P}(I_N \in \{\lfloor n_N - \varepsilon_N \rfloor, \lfloor n_N + \varepsilon_N \rfloor\}) = 1, \quad (6)$$

where $\varepsilon_N := (\log \log N)^2 / \log N$ and the parameter $n_N = n_N(p_1, p_2)$ defined in Remark 1 satisfies

$$n_N = \frac{4 \log N + O(\log \log N)}{\min_{p \in [0,1]} \max\{\log b_0(p), 2 \log b_1(p), 2 \log b_2(p)\}}, \quad (7)$$

where, using the convention $0^0 = 1$, we have

$$b_0 = b_0(p) := \left(\frac{p}{p_1 p_2}\right)^p \left(\frac{1-p}{(1-p_1)(1-p_2)}\right)^{1-p} \quad \text{and} \quad b_i = b_i(p) := \left(\frac{p}{p_i}\right)^p \left(\frac{1-p}{1-p_i}\right)^{1-p}. \quad (8)$$

Interestingly, Figure 2 shows that the form of the two-point concentration location $n_N = n_N(p_1, p_2)$ changes as we vary the edge-probabilities $(p_1, p_2) \in (0, 1)^2$, which is a rather surprising phenomenon for random graphs with constant edge-probabilities. In Section 1.3.2 we discuss how the three different forms of n_N heuristically arise (due to containment in G_{N,p_1} , containment in G_{N,p_2} , and containment in both). The value of p which attains the minimum in (7) is an approximation of the edge-density of a maximum common

induced subgraph in G_{N,p_1} and G_{N,p_2} . There is a natural guess for the ‘correct’ edge-density: if we condition on G_{n,p_1} and G_{n,p_2} being equal, then using linearity of expectation the expected edge-density equals

$$\mathbb{E}\left(\frac{e(G_{n,p_1})}{\binom{n}{2}} \middle| G_{n,p_1} = G_{n,p_2}\right) = \frac{p_1 p_2}{p_1 p_2 + (1-p_1)(1-p_2)} =: \hat{p}(p_1, p_2) = \hat{p}. \quad (9)$$

It turns out that \hat{p} is indeed the correct edge-density for a range of edge-probabilities $(p_1, p_2) \in (0, 1)^2$; see region (a) in Figure 2. In those cases Corollary 5 gives the two-point concentration location explicitly (the derivation is deferred to Appendix A), which in the special case $p_1 = p_2 = 1/2$ recovers [6, Theorem 1.1].

Corollary 5 (Special cases). *Let $p_1, p_2 \in (0, 1)$ be constants satisfying $\log b_0(\hat{p}) > \max\{2 \log b_1(\hat{p}), 2 \log b_2(\hat{p})\}$ for $b_j(p)$ is as defined in (8), which in particular holds when $p_1 = p_2$. For independent binomial random graphs G_{N,p_1} and G_{N,p_2} , define I_N as the size of the maximum common induced subgraph. Then*

$$\lim_{N \rightarrow \infty} \mathbb{P}(I_N \in \{ \lfloor x_N^{(0)}(\hat{p}) - \varepsilon_N \rfloor, \lfloor x_N^{(0)}(\hat{p}) + \varepsilon_N \rfloor \}) = 1, \quad (10)$$

where $\varepsilon_N := (\log \log N)^2 / \log N$, $\hat{p} = \hat{p}(p_1, p_2)$ is defined as in (9), $b_0 = b_0(p)$ is defined as in (8), and

$$x_N^{(0)}(p) := 4 \log_{b_0} N - 2 \log_{b_0} \log_{b_0} N - 2 \log_{b_0} (4/e) + 1. \quad (11)$$

In general, the two-point concentration location $n_N = n_N(p_1, p_2)$ is defined in (12) as the solution of an optimization problem over all edge-densities $p \in [0, 1]$ of a potential maximum common subgraph (as indicated above, in Corollary 5 the maximum is attained by \hat{p} , i.e., we have $n_N = x_N^{(0)}(\hat{p})$ in (12) below; see Lemma 9 in Appendix A). In Section 1.3.2 we discuss how this more complicated form of n_N stems from large variance difficulties that can arise in the second moment method arguments (where the typical number of copies can be zero even when the expected number of copies tends to infinity). While the definition (12) of n_N involves two implicitly defined (13) parameters $x_N^{(i)}(p)$, we remark that when $n_N = x_N^{(i)}(p)$ holds for $i \in \{1, 2\}$, then $x_N^{(i)}(p)$ has the explicit form¹ given in (14) below; see Lemma 9 in Appendix A for further estimates of n_N .

Remark 1 (Explicit expression for n_N). *The proof of Theorem 4 uses*

$$n_N = n_N(p_1, p_2) := \max_{p \in [0, 1]} \min \{ x_N^{(0)}(p), x_N^{(1)}(p), x_N^{(2)}(p) \}, \quad (12)$$

where $x_N^{(0)}(p)$ is as defined in (11), and $x_N^{(i)}(p) = x_N^{(i)}(p)$ with $i \in \{1, 2\}$ is the unique solution to

$$e \cdot N = x_N^{(i)}(p) \cdot b_i^{(x_N^{(i)}(p)-1)/2}, \quad (13)$$

where $b_i = b_i(p)$ is as defined in (8). Furthermore, if $n_N = x_N^{(i)}(p)$ holds in (12) for $i \in \{1, 2\}$, then

$$x_N^{(i)}(p) = 2 \log_{b_i} N - 2 \log_{b_i} \log_{b_i} N - 2 \log_{b_i} (2/e) + 1 + O\left(\frac{\log \log N}{\log N}\right). \quad (14)$$

1.3 Intuition and proof heuristics

The proofs of our main results are based on involved applications of the first and second moment method, which each require an extra twist due to large variance (that prevents standard applications of these methods). In the following we highlight some of the key intuition and heuristics going into our arguments.

¹One might be tempted to think that the estimate (14) could be used to explicitly define $x_N^{(i)}(p)$ for $i \in \{1, 2\}$, by simply ignoring the additive error term. Unfortunately, this does not work for a subtle technical reason: for fixed N this would lead to $\lim_{p \rightarrow p_i} x_N^{(i)}(p) = -\infty$ (since $\lim_{p \rightarrow p_i} b_i(p) = b_i(p_i) = 1$), making it inadequate for the definition (12) of n_N .

1.3.1 Induced subgraph isomorphism problem

We now heuristically discuss the sharp threshold for induced containment $G_{n,p_1} \sqsubseteq G_{N,p_2}$. Besides outlining the reasoning behind our proof approach for Theorem 1, we also motivate why the threshold is located around $n \approx 2 \log_a N + 1$, and clarify why the case $p = 1/2$ behaves so differently than the case $p \neq 1/2$.

The natural proof approach would be to apply the first and second moment method to the random variable X that counts the number of induced copies of G_{n,p_1} in G_{N,p_2} . While this approach can indeed be used to establish (1) when $p_2 = 1/2$ (as done in [6] for $p_1 = p_2 = 1/2$), it fails when $p_2 \neq 1/2$: the reason is $\text{Var } X \gg (\mathbb{E}X)^2$, i.e., that the variance of X is too large to apply the second moment method.

We overcome this second moment challenge for $p_2 \neq 1/2$ by identifying the key reason for the large variance, which turns out to be random fluctuations of the number of edges in G_{n,p_1} . To work around the effect of these edge-fluctuations we use a multi-round exposure approach, where we first reveal G_{n,p_1} and then G_{N,p_2} (which conveniently allows us to deal with one source of randomness at a time). When we reveal G_{n,p_1} , we exploit that $H := G_{n,p_1}$ will typically be asymmetric and satisfy other pseudorandom properties. Writing X_H for the number of induced copies of H in G_{N,p_2} , we then focus on the conditional probability

$$\mathbb{P}(G_{n,p_1} \sqsubseteq G_{N,p_2} \mid G_{n,p_1} = H) = \mathbb{P}(H \sqsubseteq G_{N,p_2}) = \mathbb{P}(X_H \geq 1). \quad (15)$$

To see how the number of edges in H affects this containment probability, suppose for concreteness that H has $m = \binom{n}{2} p_1 + \delta n/2$ edges. Recalling $a = 1/(p_2^{p_1}(1-p_2)^{1-p_1})$, it turns out (see (32)–(33) in Section 2.1 for the routine details) that the expected number of induced copies of H in G_{N,p_2} satisfies

$$\mathbb{E}X_H = (N)_n \cdot p_2^m (1-p_2)^{\binom{n}{2}-m} = \left[\left(\frac{p_2}{1-p_2} \right)^{\delta/2} N e^{-\log a \cdot (n-1)/2 + O(n/N)} \right]^n. \quad (16)$$

Intuitively, the expected number $\mathbb{E}X_H$ depends on H only through its number of vertices $n = v(H)$ and edges $m = e(H)$, since (a) for asymmetric graphs H the number $(N)_n$ of potential copies depends only on n , and (b) the containment probability $p_2^m (1-p_2)^{\binom{n}{2}-m}$ of any such copy of H depends only on n and m .

For $n = 2 \log_a N + 1 + c$ and $p_2 \neq 1/2$ it follows from (16) that the value of the edge-deviation parameter δ determines whether $\mathbb{E}X_H$ goes to ∞ or 0, i.e., depending on whether $\delta \log(p_2/(1-p_2))$ is smaller or larger than $c \log a$ (here we also see why the case $p_2 = 1/2$ behaves so differently: in (16) the term involving δ equals one and thus disappears). With some technical effort, we can make the first moment heuristic that $\mathbb{E}X_H \rightarrow \infty$ implies $\mathbb{P}(X_H \geq 1) \rightarrow 1$ rigorous, i.e., use the first and second moment method to show that

$$\mathbb{P}(X_H \geq 1) = \begin{cases} o(1) & \text{if } \delta \log(p_2/(1-p_2)) < c \log a, \\ 1 - o(1) & \text{if } \delta \log(p_2/(1-p_2)) > c \log a, \end{cases} \quad (17)$$

where a good control of the ‘overlaps’ of different H in the variance calculations requires us to identify and exploit suitable pseudorandom properties of H (which we can ‘with foresight’ insert into our argument when we first reveal $H = G_{n,p_1}$). The crux is that the containment conditions in (17) only depend on deviations in the number $e(G_{n,p_1}) = \binom{n}{2} p_1 + \delta n/2$ of edges in G_{n,p_1} , which in view of (15) intuitively translates into

$$\mathbb{P}(G_{n,p_1} \sqsubseteq G_{N,p_2}) = \mathbb{P}\left(\frac{e(G_{n,p_1}) - \binom{n}{2} p_1}{n/2} \cdot \log(p_2/(1-p_2)) > c \log a \right) + o(1) \quad (18)$$

for $n = 2 \log_a N + 1 + c$. This in turn makes the threshold result (2) plausible via the Central Limit Theorem (in fact, (18) is also consistent with the form of (1) for $p_2 = 1/2$, since then $\log(p_2/(1-p_2)) = 0$); see Section 2.1 for the full technical details of the proof of Theorem 1.

With this knowledge about X_H in hand, the distributional result Theorem 2 follows without much extra conceptual work. Indeed, for $p \neq 1/2$ we obtain the unusual limiting distribution (4) by exploiting that the event $X_H \geq 1$ in the second moment method based $1 - o(1)$ statement in (17) can be strengthened to $X_H \approx \mathbb{E}X_H$; see Section 2.2. Furthermore, for $p = 1/2$ we obtain the Poisson limiting distribution (3) by refining our variance estimates in order to apply the Stein-Chen method to X_H ; see Section 2.3.

1.3.2 Maximum common induced subgraph problem

We now heuristically discuss the two-point concentration of the size I_N of the maximum common induced subgraph in G_{N,p_1} and G_{N,p_2} . Our main focus is on motivating the ‘correct’ typical value $I_N \approx n_N$ in Theorem 4, which requires a refinement of the vanilla first and second moment approach. For ease of exposition, here we shall restrict our attention to the first-order asymptotics of $n_N = n_N(p_1, p_2)$ from (7).

Armed with the first and second moment insights for X_H from the heuristic discussion in Section 1.3.1, the natural approach towards determining the typical value of I_N would be to focus on the random variables $X_{n,m} := \sum_{H \in \mathcal{G}_{n,m}} X_H^{(1)} X_H^{(2)}$, where $\mathcal{G}_{n,m}$ denotes the set of all unlabeled graphs with n vertices and m edges, and $X_H^{(i)}$ counts the number of induced copies of H in G_{N,p_i} . Here the crux is that

$$I_N = \max\{n : X_{n,m} \geq 1 \text{ for some } 0 \leq m \leq \binom{n}{2}\}. \quad (19)$$

Using $|\mathcal{G}_{n,m}| = \binom{\binom{n}{2}}{m} e^{O(n \log n)}$ and the independence of G_{N,p_1} and G_{N,p_2} , similar to (16) it turns out (see (61) and (69) in Section 3.1 for the details) that for edge-density $p = p(m, n) := m/\binom{n}{2}$ we have

$$\mathbb{E}X_{n,m} = e^{O(n \log n)} \binom{\binom{n}{2}}{m} \cdot \prod_{i \in \{1,2\}} (N)_n p_i^m (1 - p_i)^{\binom{n}{2} - m} = \left[N^2 e^{-\frac{n}{2} \log b_0(p) + O(\log n)} \right]^n. \quad (20)$$

Taking into account when these expectations go to 0 and ∞ , in view of (19) and standard first moment heuristics it then is natural to guess² that the typical value of I_N should be approximately

$$n_N \approx \max_{p \in [0,1]} 4 \log_{b_0(p)} N = \frac{4 \log N}{\min_{p \in [0,1]} \log b_0(p)} = 4 \log_{b_0(\hat{p})} N, \quad (21)$$

where $\hat{p} = \hat{p}(p_1, p_2)$ from (9) turns out to be the unique minimizer of $\log b_0(p)$ from (8). For this choice of n_N it turns out that we indeed typically have $I_N \approx n_N$ for a range of edge-probabilities $(p_1, p_2) \in (0, 1)^2$, including the special case $p_1 = p_2$; see Corollary 5 and region (a) in Figure 2.

For general edge-probabilities $(p_1, p_2) \in (0, 1)^2$ the ‘correct’ form of n_N is more complicated than our first guess (21), and the key reason turns out to be that due to large variance we can have $\mathbb{P}(X_{n,m} \geq 1) \rightarrow 0$ despite $\mathbb{E}X_{n,m} \rightarrow \infty$. We overcome this difficulty by realizing that containment in G_{N,p_i} can, from a ‘first moment perspective’, sometimes be harder than containment in both G_{N,p_1} and G_{N,p_2} , i.e., that we can have $\mathbb{E}X_{n,m}^{(i)} \rightarrow 0$ for $X_{n,m}^{(i)} := \sum_{H \in \mathcal{G}_{n,m}} X_H^{(i)}$ despite $\mathbb{E}X_{n,m} \rightarrow \infty$. Similarly to (20) and (16) it turns out (see (59) in Section 3.1 for the details) that for edge-density $p = p(m, n) = m/\binom{n}{2}$ we have

$$\mathbb{E}X_{n,m}^{(i)} = e^{O(n \log n)} \binom{\binom{n}{2}}{m} \cdot (N)_n p_i^m (1 - p_i)^{\binom{n}{2} - m} = \left[N e^{-\frac{n}{2} \log b_i(p) + O(\log n)} \right]^n. \quad (22)$$

Note that $X_{n,m} \geq 1$ implies both $X_{n,m}^{(1)} \geq 1$ and $X_{n,m}^{(2)} \geq 1$. Taking into account when the two expectations in (22) go to 0, we thus refine our first guess (21) for the typical value of I_N to approximately

$$\begin{aligned} n_N &\approx \max_{p \in [0,1]} \min\{4 \log_{b_0(p)} N, 2 \log_{b_1(p)} N, 2 \log_{b_2(p)} N\} \\ &= \frac{4 \log N}{\min_{p \in [0,1]} \max\{\log b_0(p), 2 \log b_1(p), 2 \log b_2(p)\}}. \end{aligned} \quad (23)$$

This turns out to be the ‘correct’ value of n_N up to second order correction terms, see (7) and Figure 2. Indeed, with substantial technical effort (more involved than for the induced subgraph isomorphism problem) we can use refinements of the first and second moment method to prove two-point concentration of form $I_N \approx n_N$ for all edge-probabilities $(p_1, p_2) \in (0, 1)^2$; see Section 3 for the full technical details of the proof of Theorem 4.

²The first moment heuristic for n_N is as follows. For any $n \geq (1+\xi)n_N$, the definition of n_N and (20) imply that $\sum_m \mathbb{E}X_{n,m} \leq n^2 \cdot N^{-\Theta(\xi n)} \rightarrow 0$. Similarly, for any $n \leq (1-\xi)n_N$ we have $\mathbb{E}X_{n,m} \geq N^{\Theta(\xi n)} \rightarrow \infty$ for $m = \binom{n}{2}\hat{p}$.

1.3.3 Pseudorandom properties of random graphs

To enable the technical variance estimates in the proofs of our main results, we need to get good control over the ‘overlaps’ of different induced copies. To this end it will be important to take certain pseudorandom properties of random graphs into account, by restricting our attention to (labelled) graphs H from the following three classes:

- \mathcal{A}_n : The set of all n -vertex graphs H where, for all vertex-subsets $L \subseteq [n]$ of size $|L| \geq n - n^{2/3}$, the induced subgraph $H[L]$ is asymmetric, i.e., its automorphism group satisfies $|\text{Aut}(H[L])| = 1$.
- $\mathcal{E}_{n,m}$: The set of all n -vertex graphs H where, for all non-empty vertex-subsets $L \subseteq [n]$, the induced subgraph $H[L]$ contains $e(H[L]) = \binom{|L|}{2}m/\binom{n}{2} \pm n^{2/3}(n - |L|)$ edges.
- $\mathcal{F}_{n,p}$: The set of all n -vertex graphs H with $e(H) = \binom{n}{2}p \pm n^{4/3}$ edges.

In concrete words, the first property \mathcal{A}_n formalizes the folklore fact that all very large subsets of random graphs are asymmetric; see [4, 11, 16]. The second and third properties $\mathcal{E}_{n,m}$ and $\mathcal{F}_{n,p}$ formalize the well-known fact that the edges of a random graph are well-distributed; see [3, 4, 12]. The following auxiliary lemma states that these are indeed typical properties of dense random graphs (we defer the proof to Appendix B, since it is rather tangential to our main arguments). Below $G_{n,m}$ denotes the uniform random graph, which is chosen uniformly at random from all n -vertex graphs with exactly m edges. Furthermore, the abbreviation whp (with high probability) means with probability tending to one as the number of vertices goes to infinity.

Lemma 6 (Random graphs are pseudorandom). *Let $p, \gamma \in (0, 1)$ be constants. Then the following holds:*

- (i) *For all $\gamma\binom{n}{2} \leq m \leq (1 - \gamma)\binom{n}{2}$, the uniform random graph whp satisfies $G_{n,m} \in \mathcal{A}_n \cap \mathcal{E}_{n,m}$.*
- (ii) *The binomial random graph whp satisfies $G_{n,p} \in \mathcal{A}_n \cap \mathcal{F}_{n,p}$.*

The reason for why we only consider the edge-property $\mathcal{E}_{n,m}$ for the uniform random graph $G_{n,m}$ is conceptual: the edges of $G_{n,m}$ are simply more well-distributed than the edges of the binomial random $G_{n,p}$ with $p := m/\binom{n}{2} \in [\gamma, 1 - \gamma]$, where for large vertex-subsets $L \subseteq [n]$ of size $|L| = n - o(n^{1/3})$ we expect that typically $|e(G_{n,p}[L]) - \binom{|L|}{2}p| = \Omega(|L|\sqrt{p(1-p)}) \gg n^{2/3}(n - |L|)$ holds.

2 Induced subgraph isomorphism problem

This section is devoted to the induced subgraph isomorphism problem for two independent random graphs G_{n,p_1} and G_{N,p_2} with constant edge-probabilities $p_1, p_2 \in (0, 1)$. It naturally splits into several parts: in the core Section 2.1 we prove the sharp threshold result Theorem 1, and in the subsequent Sections 2.2–2.3 we then separately prove the two parts of the distributional result Theorem 2.

2.1 Sharp Threshold: Proof of Theorem 1

In this section we prove Theorem 1, i.e., establish a sharp threshold for the appearance of an induced copy of G_{n,p_1} in G_{N,p_2} , where $p_1, p_2 \in (0, 1)$. Our proof-strategy is somewhat roundabout, since we need to deal with the difficulty that for $p_2 \neq 1/2$ the induced containment event $G_{n,p_1} \sqsubseteq G_{N,p_2}$ is sensitive towards small variations in the number of edges of G_{n,p_1} (as discussed in Sections 1.1 and 1.3.1). For this reason we first prove a sharp threshold result for the event that G_{N,p_2} contains an induced copy of a particular n -vertex pseudorandom graph H with m edges (see Lemma 7 below). Since G_{n,p_1} will typically be pseudorandom by Lemma 6, this then effectively reduces the problem of estimating $\mathbb{P}(G_{n,p_1} \sqsubseteq G_{N,p_2})$ to understanding deviations in the number of edges of G_{n,p_1} , which is a well-understood and much simpler problem.

Turning to the details, to restrict our attention to graphs with pseudorandom properties we introduce

$$\mathcal{T}_{n,m} := \mathcal{A}_n \cap \mathcal{E}_{n,m}, \tag{24}$$

where the asymmetry property \mathcal{A}_n and the edge-distribution property $\mathcal{E}_{n,m}$ are defined as in Section 1.3.3. For edge-counts m of interest, Lemma 6 shows that the uniform random graph whp satisfies $G_{n,m} \in \mathcal{T}_{n,m}$, so Lemma 7 effectively gives a sharp threshold result for the induced containment $G_{n,m} \sqsubseteq G_{N,p_2}$.

Lemma 7 (Sharp threshold for pseudorandom graphs). *Let $p_1, p_2 \in (0, 1)$ be constants. Define $a := 1/(p_2^{p_1}(1-p_2)^{1-p_1})$, $\varepsilon_N := (\log \log N)^2 / \log N$, and $\psi := \log_a(p_2/(1-p_2))$. If $\delta_m := \lfloor m - \binom{n}{2}p_1 \rfloor / (n/2)$ and $c_N := n - (2 \log_a N + 1)$ satisfy $|\psi \delta_m| = o(n)$ and $|c_N| = o(\log_a N)$, then the following holds: as $N \rightarrow \infty$, for any graph $H \in \mathcal{T}_{n,m}$:*

$$\mathbb{P}(H \subseteq G_{N,p_2}) = \begin{cases} o(1) & \text{if } \psi \delta_m - c_N \leq -\varepsilon_N, \\ 1 - o(1) & \text{if } \psi \delta_m - c_N \geq \varepsilon_N. \end{cases} \quad (25)$$

Remark 2. *The $o(1)$ statement in (25) remains valid for any n -vertex graph H with $e(H) = m$ edges.*

Before giving the (first and second moment based) proof of Lemma 7, we first show how it implies Theorem 1 using a multi-round exposure argument, where we first reveal the number of edges of G_{n,p_1} , then reveal the random graph G_{n,p_1} , and afterwards try to embed G_{n,p_1} into the random graph G_{N,p_2} , each time discarding atypical outcomes along the way (so that we can focus on pseudorandom G_{n,p_1}).

Proof of Theorem 1. To estimate the probability that $G_{n,p_1} \subseteq G_{N,p_2}$ holds, by monotonicity³ in n it suffices to consider the case where $c_N = o(\log \log N)$ holds (in fact $c_N = O(1)$ suffices, since $\lim_{c \rightarrow \infty} f(c) = 0$ and $\lim_{c \rightarrow -\infty} f(c) = 1$ when $p_2 \neq 1/2$). Since G_{n,p_1} conditioned on having m edges has the same distribution as $G_{n,m}$, by using Lemma 6 to handle outcomes of G_{n,p_1} with an atypical number of edges $e(G_{n,p_1})$ it follows that

$$\left| \mathbb{P}(G_{n,p_1} \subseteq G_{N,p_2}) - \sum_{m: |m - \binom{n}{2}p_1| \leq n^{4/3}} \mathbb{P}(G_{n,m} \subseteq G_{N,p_2}) \mathbb{P}(e(G_{n,p_1}) = m) \right| \leq \mathbb{P}(G_{n,p_1} \notin \mathcal{F}_{n,p_1}) = o(1). \quad (26)$$

If $|m - \binom{n}{2}p_1| \leq n^{4/3}$, then by using Lemma 6 to handle the atypical event $G_{n,m} \notin \mathcal{T}_{n,m}$ it follows that

$$\left| \mathbb{P}(G_{n,m} \subseteq G_{N,p_2}) - \sum_{H \in \mathcal{T}_{n,m}} \mathbb{P}(H \subseteq G_{N,p_2}) \mathbb{P}(G_{n,m} = H) \right| \leq \mathbb{P}(G_{n,m} \notin \mathcal{T}_{n,m}) = o(1). \quad (27)$$

In particular, using the sharp threshold result Lemma 7 for the event $H \subseteq G_{N,p_2}$ (which applies since $|\psi \delta_m| = o(n)$ holds) and Lemma 6 for the typical event $G_{n,m} \in \mathcal{T}_{n,m}$, it follows that (27) implies

$$\mathbb{P}(G_{n,m} \subseteq G_{N,p_2}) = \begin{cases} o(1) & \text{if } \psi \delta_m - c_N \leq -\varepsilon_N, \\ 1 - o(1) & \text{if } \psi \delta_m - c_N \geq \varepsilon_N. \end{cases} \quad (28)$$

If $p_2 = 1/2$, then in the setting of (1) we have $\psi = 0$ and $|c_N| \geq \varepsilon_N$, so the sharp threshold result (1) follows by first inserting estimate (28) into (26) and finally using Lemma 6 to infer $\mathbb{P}(G_{n,p_1} \in \mathcal{F}_{n,p_1}) = 1 - o(1)$.

We focus on the case $p_2 \neq 1/2$ in the remainder of the proof. Note that estimate (28) does not apply when $|\psi \delta_m - c_N| < \varepsilon_N$, but we shall now argue that we can effectively ignore this small range of edge-counts m in (26) above. Recall that $n \rightarrow \infty$ as $N \rightarrow \infty$. By the Central Limit Theorem it thus follows that, as $N \rightarrow \infty$,

$$Z := \psi \delta_{e(G_{n,p_1})} = \psi \cdot \frac{e(G_{n,p_1}) - \binom{n}{2}p_1}{n/2} \xrightarrow{d} N(0, 2p_1(1-p_1)\psi^2). \quad (29)$$

Recalling that $\varepsilon_N = o(1)$ and $0 < \text{Var } Z = 2p_1(1-p_1)\psi^2 = \Theta(1)$, it follows that

$$\mathbb{P}(|Z - c_N| < \varepsilon_N) = o(1). \quad (30)$$

(This alternatively follows from $\max_k \mathbb{P}(e(G_{n,p_1}) = k) \leq O(1)/\sqrt{n^2 p_1(1-p_1)} \leq O(1/n)$, say.) Inserting estimates (28) and (30) into (26), using Lemma 6 to infer $\mathbb{P}(G_{n,p_1} \in \mathcal{F}_{n,p_1}) = 1 - o(1)$ it follows that

$$|\mathbb{P}(G_{n,p_1} \subseteq G_{N,p_2}) - \mathbb{P}(Z - c_N \geq 0)| = o(1), \quad (31)$$

which together with the convergence result (29) and $\varepsilon_N = o(1)$ establishes the threshold result (2). \square

³A standard coupling argument shows that $\mathbb{P}(G_{n,p_1} \subseteq G_{N,p_2})$ is monotone decreasing in n . The crux is that $G_{n+1,p_1}[\{1, \dots, n\}]$, i.e., the induced subgraph of G_{n+1,p_1} on vertex set $\{1, \dots, n\}$, has the same distribution as G_{n,p_1} . It thus follows that $\mathbb{P}(G_{n+1,p_1} \subseteq G_{N,p_2}) \leq \mathbb{P}(G_{n+1,p_1}[\{1, \dots, n\}] \subseteq G_{N,p_2}) = \mathbb{P}(G_{n,p_1} \subseteq G_{N,p_2})$, as claimed.

In the following proof of Lemma 7 we shall apply the first and second moment method to the random variable X_H , which we define as the number of induced copies of H in G_{N,p_2} . Here the restriction to pseudorandom graphs $H \in \mathcal{T}_{n,m}$ will be key for controlling the expectation $\mathbb{E}X_H$ and variance $\text{Var } X_H$.

Proof of Lemma 7 and Remark 2. Let H be any n -vertex graph with $e(H) = m$ edges (to clarify: with an eye on Remark 2 we initially do not assume $H \in \mathcal{T}_{n,m}$). Note that $X_H = \sum_{S \in \binom{[N]}{n}} I_S$, where I_S is the indicator random variable for the event that the induced subgraph $G_{N,p_2}[S]$ is isomorphic to H . Note that there are exactly $n! / |\text{Aut}(H)|$ distinct embeddings of H into S . Using linearity of expectation and $(N)_n = N! / (N-n)!$, in view of $m = \binom{n}{2} p_1 + \delta_m n / 2$, $a = 1 / (p_2^{p_1} (1-p_2)^{1-p_1})$, $p_2 / (1-p_2) = a^\psi$ and $(n-1)/2 = \log_a N + c_N / 2$ it follows that the expected number of induced copies of H in G_{N,p_2} satisfies

$$\begin{aligned} \mathbb{E}X_H &= \sum_{S \in \binom{[N]}{n}} \mathbb{E}I_S = \binom{N}{n} \cdot \frac{n!}{|\text{Aut}(H)|} \cdot p_2^{e(H)} (1-p_2)^{\binom{n}{2}-e(H)} \\ &= \frac{1}{|\text{Aut}(H)|} \cdot (N)_n \cdot p_2^m (1-p_2)^{\binom{n}{2}-m} \\ &= \frac{1}{|\text{Aut}(H)|} \cdot \left[N e^{O(n/N)} \cdot \left(\frac{p_2}{1-p_2} \right)^{\delta_m/2} a^{-(n-1)/2} \right]^n \\ &= \frac{1}{|\text{Aut}(H)|} \cdot \left[a^{\psi \delta_m - c_N + o(\varepsilon_N)} \right]^{n/2}, \end{aligned} \tag{32}$$

where we used $n/N = \Theta((\log N)/N) = o(\varepsilon_N)$ and $1 < a = O(1)$ for the last step. If $\psi \delta_m - c_N \leq -\varepsilon_N$, then using $|\text{Aut}(H)| \geq 1$ and Markov's inequality together with $\varepsilon_N n \rightarrow \infty$ as $N \rightarrow \infty$, we see that

$$\mathbb{P}(X_H > 0) \leq \mathbb{E}X_H \leq a^{-\Omega(\varepsilon_N n)} = o(1),$$

establishing the $o(1)$ statement in (25) and also Remark 2 (since we so far did not assume $H \in \mathcal{T}_{n,m}$).

In the remainder we focus on the $1 - o(1)$ statement in (25), i.e., we henceforth fix $H \in \mathcal{T}_{n,m}$ and assume that $\psi \delta_m - c_N \geq \varepsilon_N$. Since $H \in \mathcal{T}_{n,m} \subseteq \mathcal{A}_N$ implies $|\text{Aut}(H)| = 1$, using estimate (32) we infer that

$$\mu := \mathbb{E}X_H = (N)_n \cdot p_2^m (1-p_2)^{\binom{n}{2}-m} = \left[a^{\psi \delta_m - c_N + o(\varepsilon_N)} \right]^{n/2} \geq a^{\Omega(\varepsilon_N n)} \rightarrow \infty. \tag{33}$$

To complete the proof of (25), using Chebyshev's inequality it thus suffices to show that $\text{Var } X_H = o((\mathbb{E}X_H)^2)$. Recall that $X_H = \sum_{S \in \binom{[N]}{n}} I_S$, where I_S is the indicator random variable for the event that the induced subgraph $G_{N,p_2}[S]$ is isomorphic to H . Since I_R and I_S are independent when $|R \cap S| \leq 1$, we have

$$\text{Var } X_H \leq \underbrace{\sum_{2 \leq \ell \leq n} \sum_{R, S \in \binom{[N]}{n}: |R \cap S| = \ell} \mathbb{E}(I_R I_S)}_{=: w_\ell} = \sum_{2 \leq \ell < n} w_\ell + \underbrace{\mathbb{E}X_H}_{=\mu}. \tag{34}$$

To bound the parameter w_ℓ defined in (34) for $2 \leq \ell < n$, we first note that there are $\binom{N}{n} \binom{n}{\ell} \binom{N-n}{n-\ell}$ ways of choosing $R, S \in \binom{[N]}{n}$ with $|R \cap S| = \ell$. To then get a handle on $\mathbb{E}(I_R I_S)$, we count the number of ways we can embed two induced copies of H into $R \cup S$ so that $G_{N,p_2}[R]$ and $G_{N,p_2}[S]$ are both isomorphic to H : there are $n!$ ways of embedding an induced copy of H into R , at most $\binom{n}{n-\ell} \cdot (n-\ell)! = \binom{n}{\ell} \cdot (n-\ell)! = (n)_{n-\ell}$ ways of choosing and embedding $n-\ell$ vertices of a second induced copy of H into $S \setminus R$, and at most $\max_{H \in \mathcal{T}_{n,m}, |L|=\ell} |\text{Aut}(H[L])|$ ways of embedding the remaining ℓ vertices of the second induced copy of H into $R \cap S$, where the maximum is taken over all vertex-subsets $L \subseteq V(H)$ of size $|L| = \ell$. As these embeddings determine all edges and non-edges in $G_{N,p_2}[R \cup S]$, it follows that

$$\begin{aligned} w_\ell &\leq \binom{N}{n} \binom{n}{\ell} \binom{N-n}{n-\ell} \cdot n! \cdot (n)_{n-\ell} \cdot \max_{H \in \mathcal{T}_{n,m}, |L|=\ell} |\text{Aut}(H[L])| \\ &\quad \cdot \max_{H \in \mathcal{T}_{n,m}, |L|=\ell} p_2^{2m-e(H[L])} (1-p_2)^{2\binom{n}{2}-\binom{\ell}{2}-2m+e(H[L])}, \end{aligned} \tag{35}$$

where the two maxima in (35) are each taken over all vertex-subsets $L \subseteq V(H)$ of size $|L| = \ell$, as above. Note that $n = (2 + o(1)) \log_a N$, which implies $N - n > n - \ell$ for all sufficiently large N . Recalling that $\mu = (N)_n p_2^m (1 - p_2)^{\binom{n}{2} - m}$ by (33), using $n^2/N = o(1)$ it follows that

$$\frac{w_\ell}{\mu^2} \leq \underbrace{\frac{(N - n)_{n-\ell}}{(N)_n}}_{=(1+o(1))N^{-\ell}} \cdot \binom{n}{\ell}^2 \cdot \max_{H \in \mathcal{T}_{n,m}, |L|=\ell} |\text{Aut}(H[L])| \cdot \underbrace{\max_{H \in \mathcal{T}_{n,m}, |L|=\ell} \left(\frac{1-p_2}{p_2} \right)^{e(H[L])}}_{=: P_\ell} (1-p_2)^{-\binom{\ell}{2}}. \quad (36)$$

Setting $\zeta := 1/\max\{-2\log_a(1-p_2), -2\log_a(p_2), 2\} > 0$, we now bound w_ℓ/μ^2 further using a case distinction.

Case $2 \leq \ell \leq \zeta n$: Here we use $0 \leq e(H[L]) \leq \binom{\ell}{2}$ to deduce that the parameter P_ℓ from (36) satisfies

$$P_\ell \leq \max_{0 \leq k \leq \binom{\ell}{2}} p_2^{-k} (1-p_2)^{-\binom{\ell}{2}+k} \leq \left[\max\left\{ \frac{1}{p_2}, \frac{1}{1-p_2} \right\} \right]^{\binom{\ell}{2}}. \quad (37)$$

Inserting this estimate and the trivial bound $|\text{Aut}(H[L])| \leq |L|!$ into (36), using $\binom{n}{\ell} \leq n^\ell/\ell!$ together with $n = (2 + o(1)) \log_a N$ and $\ell \leq \zeta n \leq (1 + o(1)) \log_{\max\{1/p_2, 1/(1-p_2)\}} N$, it follows for all large enough N that

$$\frac{w_\ell}{\mu^2} \leq 2N^{-\ell} \cdot \binom{n}{\ell}^2 \ell! \cdot P_\ell \leq \left[N^{-1} n^2 \cdot \left[\max\left\{ \frac{1}{p_2}, \frac{1}{1-p_2} \right\} \right]^{(\ell-1)/2} \right]^\ell \leq N^{-\ell/3} \leq a^{-\Omega(n)}. \quad (38)$$

Case $\zeta n \leq \ell < n$: Here we improve the estimate (37) for P_ℓ using that $H \in \mathcal{T}_{n,m} \subseteq \mathcal{E}_{n,m}$: this pseudorandom edge-property implies that, for any vertex-subset $L \subseteq [n]$ of size $|L| = \ell$, the induced number of edges satisfies $e(H[L]) = \binom{\ell}{2} m / \binom{n}{2} + O(n^{2/3}(n-\ell))$. Recalling $a = 1/(p_2^{p_1}(1-p_2)^{1-p_1})$, using $m/\binom{n}{2} = p_1 + \delta_m/n + O(1/n^2)$ and $(1-p_2)/p_2 = a^{-\psi}$ it follows that the parameter P_ℓ from (36) satisfies

$$P_\ell = a^{\binom{\ell}{2}} \cdot \left(\frac{1-p_2}{p_2} \right)^{\binom{\ell}{2} \delta_m/n + O(n^{2/3}(n-\ell))} = a^{(1-\psi\delta_m/n)\binom{\ell}{2} + O(\psi n^{2/3}(n-\ell))}. \quad (39)$$

For $\zeta n \leq \ell < n - n^{2/3}$ we then estimate $|\text{Aut}(H[L])| \leq \ell! \leq n^\ell \leq e^{O(n^{2/3}(n-\ell))}$, and for $n - n^{2/3} \leq \ell < n$ we use $H \in \mathcal{T}_{n,m} \subseteq \mathcal{A}_n$ to obtain $\max_{|L|=\ell} |\text{Aut}(H[L])| = 1$. Note that $1 < a = O(1)$ and $\psi = O(1)$. We now insert these estimates and $\binom{n}{\ell} = \binom{n}{n-\ell} \leq n^{n-\ell}$ into (36). Then, using $\ell - 1 = (n - 1) - (n - \ell)$ and $|\psi\delta_m/n| = o(1)$ as well as $(n - 1)/2 = \log_a N + c_N/2$ and $n - \ell \geq 1$, it follows for all large enough N that

$$\begin{aligned} \frac{w_\ell}{\mu^2} &\leq 2N^{-\ell} n^{2(n-\ell)} \cdot a^{(1-\psi\delta_m/n)\binom{\ell}{2}} \cdot e^{O(n^{2/3}(n-\ell))} \\ &\leq \left[N^{-1} a^{(1-\psi\delta_m/n)(\ell-1)/2} \cdot e^{O(n^{2/3}(n-\ell)/\ell)} \right]^\ell \\ &\leq \left[N^{-1} a^{(n-1)/2 - \psi\delta_m/2 + o(1) - (1-o(1))(n-\ell)/2} \cdot e^{o(n-\ell)} \right]^\ell \\ &\leq \left[a^{c_N - \psi\delta_m} \right]^{\ell/2} \leq a^{-\Omega(\varepsilon_N n)} \leq n^{-\Omega(\log n)}, \end{aligned} \quad (40)$$

where for the second last inequality we used that $c_N - \psi\delta_m = -(\psi\delta_m - c_N) \leq -\varepsilon_N$.

To sum up: inserting the estimates (38) and (40) into the variance bound (34), using (33) it follows that

$$\frac{\text{Var } X_H}{(\mathbb{E} X_H)^2} \leq \frac{\sum_{2 \leq \ell < n} w_\ell + \mu}{\mu^2} \leq o(1) + \frac{1}{\mu} = o(1),$$

which completes the proof of Lemma 7, as discussed. \square

Remark 3. If $H \in \mathcal{T}_{n,m}$ and $\psi\delta_m - c_N \geq \varepsilon_N$, then the above proof of Lemma 7 shows that

$$\mathbb{P}(|X_H - \mathbb{E} X_H| < \xi \mathbb{E} X_H) = 1 - o(1)$$

for any $\xi > 0$, where the expected value $\mathbb{E} X_H$ satisfies the asymptotic estimate (33).

2.2 Asymptotic Distribution: Proof of Theorem 2 (ii)

In this section we prove the distributional result Theorem 2 (ii) as a corollary of the results from Section 2.1, i.e., establish that for $p_2 \neq 1/2$ the number of induced copies of G_{n,p_1} in G_{N,p_2} has a ‘squashed’ log-normal limiting distribution. Here the crux is X_H is strongly dependent on the number of edges: indeed, X_H quickly changes from 0 to $(1 + o(1))\mathbb{E}X_H \rightarrow \infty$ as the number of edges $m = e(H)$ passes through the threshold $\psi\delta_m \sim c_N$, see Remarks 2 and 3. This makes it plausible that $\log(1 + X_H)/\log N$ changes abruptly around that threshold, which together with $c_N \rightarrow c$ and the normal convergence result (29) for $\psi\delta_{e(G_{n,p_1})}$ intuitively explains the form of the limiting distribution of $\log(1 + X_{G_{n,p_1}})/\log N$ in the convergence result (4).

Proof of Theorem 2 (ii). Note that $c_N \rightarrow c$ and thus $c_N = O(1)$ by assumption. In view of Remark 3, define \mathcal{T}_n as the union of all $\mathcal{T}_{n,m}$ from Section 2.1 with $|m - \binom{n}{2}p_1| \leq n^{4/3}$. With analogous reasoning as for (26)–(27), using Lemma 6 it follows that

$$\mathbb{P}(G_{n,p_1} \notin \mathcal{T}_n) \leq \mathbb{P}(G_{n,p_1} \notin \mathcal{F}_{n,p_1}) + \max_{m: |m - \binom{n}{2}p_1| \leq n^{4/3}} \mathbb{P}(G_{n,m} \notin \mathcal{T}_{n,m}) = o(1). \quad (41)$$

Furthermore, with an eye on the form (33) of $\mathbb{E}X_H$ in Lemma 7, as in (29) we have

$$Z := \psi \cdot \delta_{e(G_{n,p_1})} = \psi \cdot \frac{e(G_{n,p_1}) - \binom{n}{2}p_1}{n/2} \xrightarrow{d} N(0, 2p_1(1 - p_1)\psi^2). \quad (42)$$

We now condition on $G_{n,p_1} = H \in \mathcal{T}_n$, so that $X|_{G_{n,p_1}=H} = X_H$ and $|Z| \leq 2|\psi|n^{1/3} = o(\sqrt{\log N})$. In particular, if $Z - c_N \leq -\varepsilon_N$ holds, then by applying Lemma 7 it follows that whp

$$\frac{\log(1 + X_H)}{\log N} = \frac{\log 1}{\log N} = 0. \quad (43)$$

Furthermore, if $Z - c_N \geq \varepsilon_N$ holds, then by combining Remark 3, estimate (33) and $\varepsilon_N = o(1/\sqrt{\log N})$ with $c_N = O(1)$ and $|Z| = o(\sqrt{\log N})$ as well as $n = 2\log_a N + O(1)$ and $a = O(1)$ it follows that whp

$$\frac{\log(1 + X_H)}{\log N} = \frac{\log(\mathbb{E}X_H)}{\log N} + \frac{O(1)}{\log N} = \frac{n(Z - c_N)}{2\log_a N} + \frac{o(1)}{\sqrt{\log N}} = Z - c_N + \frac{o(1)}{\sqrt{\log N}}. \quad (44)$$

Finally, by combining estimates (41) and (30) with the conclusions of (43)–(44), it follows that

$$\mathbb{P}\left(\left|\frac{\log(1 + X)}{\log N} - (Z - c_N)\mathbb{1}_{\{Z - c_N \geq \varepsilon_N\}}\right| \geq \frac{1}{\sqrt{\log N}}\right) = o(1),$$

which together with (42) as well as $c_N \rightarrow c$ and $\varepsilon_N \rightarrow 0$ establishes the convergence result (4). \square

Remark 4. Note that the above proof only uses the first and second moment method, i.e., does not require the asymptotics of $\text{Var } X_H$. Given the somewhat complicated limiting distribution of X_H , we leave it as an interesting open problem to complement (4) with near-optimal estimates on the rate of convergence. Furthermore, as suggested out by a referee, it would also be worthwhile to investigate the limiting distribution of X_H in the case when $n - (2\log_a N + 1) \rightarrow \infty$.

2.3 Asymptotic Poisson Distribution: Proof of Theorem 2 (i)

In this section we complete the proof of Theorem 2 by proving Theorem 2 (i), i.e., establishing that for $p_2 = 1/2$ the number of induced copies of G_{n,p_1} in G_{N,p_2} has asymptotically Poisson distribution. To this end we shall use a version of the Stein-Chen method for Poisson approximation together with a two-round exposure argument and a refinement of the variance estimates from Section 2.1 for $p_2 = 1/2$.

Proof of Theorem 2 (i). Note that $X = X_{n,N}$ conditional on $G_{n,p_1} = H$ has the same distribution as X_H . This enables us to again use a two-round exposure argument, where we first reveal G_{n,p_1} and then afterwards

count the number of induced copies of G_{n,p_1} in $G_{N,1/2}$. To this end, let \mathcal{G}_n be the set of all n -vertex graphs. Together with $\mathcal{A}_n \subseteq \mathcal{G}_n$ as defined in Section 1.3.3, by applying Lemma 6 it follows that

$$\begin{aligned} d_{\text{TV}}(X, \text{Po}(\mu)) &= \sup_{S \subseteq \mathbb{N}} |\mathbb{P}(X \in S) - \mathbb{P}(\text{Po}(\mu) \in S)| \\ &= \sup_{S \subseteq \mathbb{N}} \left| \sum_{H \in \mathcal{G}_n} \mathbb{P}(G_{n,p_1} = H) \cdot [\mathbb{P}(X_H \in S) - \mathbb{P}(\text{Po}(\mu) \in S)] \right| \\ &\leq \sum_{H \in \mathcal{A}_n} \mathbb{P}(G_{n,p_1} = H) \cdot \sup_{S \subseteq \mathbb{N}} |\mathbb{P}(X_H \in S) - \mathbb{P}(\text{Po}(\mu) \in S)| + \mathbb{P}(G_{n,p_1} \notin \mathcal{A}_n) \\ &\leq \max_{H \in \mathcal{A}_n} d_{\text{TV}}(X_H, \text{Po}(\mu)) + o(1). \end{aligned}$$

It thus remains to show that $d_{\text{TV}}(X_H, \text{Po}(\mu)) = o(1)$ for any $H \in \mathcal{A}_n$. Fix a graph $H \in \mathcal{A}_n$. As in Section 2.1 we write $X_H = \sum_{S \in \binom{[N]}{n}} I_S$, where I_S is the indicator random variable for the event that $G_{N,p_2}[S]$ is isomorphic to H . Since $H \in \mathcal{A}_n$ implies $|\text{Aut}(H)| = 1$, using (32) with $p_2 = 1/2$ it follows that

$$\mathbb{E}X_H = (N)_n \left(\frac{1}{2}\right)^{\binom{n}{2}} = \mu \quad \text{for all } H \in \mathcal{A}_n. \quad (45)$$

Note that $d_{\text{TV}}(X_H, \text{Po}(\mu)) = o(1)$ immediately follows when $\mu \rightarrow 0$ (since then X_H and $\text{Po}(\mu)$ are both whp zero). We may thus henceforth assume that $\mu = \Omega(1)$, which in view of $\mu = \binom{N}{n} 2^{-(n-1)/2 + O(n/N)}^n$ and $n \geq 2 \log_2 N - 1 + \varepsilon_N$ implies $n = 2 \log_2 N + O(1)$. Since I_R and I_S are independent when $|R \cap S| \leq 1$, by applying the well-known version of the Stein-Chen method for Poisson approximation (based on so-called dependency graphs) stated in [15, Theorem 6.23] it routinely follows that

$$d_{\text{TV}}(X_H, \text{Po}(\mu)) \leq \underbrace{\min\{\mu^{-1}, 1\}}_{=O(\mu^{-1})} \cdot \left[\underbrace{\sum_{\substack{R, S \in \binom{[N]}{n}: \\ 2 \leq |R \cap S| \leq n}} \mathbb{E}(I_R I_S)}_{=: \Lambda_1} + \underbrace{\sum_{\substack{R, S \in \binom{[N]}{n}: \\ 2 \leq |R \cap S| \leq n}} \mathbb{E}I_R \mathbb{E}I_S}_{=: \Lambda_2} \right]. \quad (46)$$

To establish the convergence result (3), it thus remains to show that Λ_1 and Λ_2 are both $o(\mu)$.

We first estimate Λ_2 using basic counting arguments. In particular, note that

$$\Lambda_2 = \sum_{R \in \binom{[N]}{n}} \mathbb{E}I_R \sum_{\substack{S \in \binom{[N]}{n}: \\ 2 \leq |R \cap S| \leq n}} \mathbb{E}I_S \leq \mu \cdot \sum_{2 \leq k \leq n} \binom{n}{k} \binom{N-n}{n-k} n! 2^{-\binom{n}{2}}.$$

Recalling that $n = 2 \log_2 N + O(1)$, for all $2 \leq k \leq n-1$ we see (with room to spare) that

$$\frac{\binom{n}{k} \binom{N-n}{n-k}}{\binom{n}{k+1} \binom{N-n}{n-k-1}} = \frac{(k+1)(N-2n+k+1)}{(n-k)^2} > 1$$

for all sufficiently large N . Using $n = 2 \log_2 N + O(1)$ and $n \geq 2 \log_2 N - 1 + \varepsilon_N$, it follows that

$$\begin{aligned} \frac{\Lambda_2}{\mu} &\leq n \cdot \binom{n}{2} \binom{N-n}{n-2} \cdot n! 2^{-\binom{n}{2}} \\ &\leq n^5 N^{n-2} \cdot 2^{-\binom{n}{2}} \\ &\leq N^{-2} \cdot \left(n^{5/n} N \cdot 2^{-(n-1)/2} \right)^n \\ &\leq O(2^{-n}) \cdot \left(2^{1-\varepsilon_N/2+o(\varepsilon_N)} \right)^n \leq 2^{-\Omega(\varepsilon_N n)} \leq n^{-\Omega(\log n)}. \end{aligned}$$

Finally, we estimate Λ_1 from (46) by refining the variance estimates from the proof of Lemma 7. Namely, bounding the following parameter w_ℓ from (34) as in (35), by setting $p_2 = 1/2$ we infer that

$$w_\ell := \sum_{\substack{R, S \in \binom{[N]}{n}: \\ |R \cap S| = \ell}} \mathbb{E}(I_R I_S) \leq \binom{N}{n} \binom{n}{\ell} \binom{N-n}{n-\ell} \cdot n! \cdot (n)_{n-\ell} \cdot \max_{H \in \mathcal{T}_n, |L|=\ell} |\text{Aut}(H[L])| \cdot 2^{\binom{\ell}{2} - 2\binom{n}{2}}.$$

Using similar (but simpler) arguments as for inequality (36), in view of (45) it then follows that

$$\frac{w_\ell}{\mu} \leq N^{n-\ell} \cdot \binom{n}{\ell}^2 \cdot \max_{H \in \mathcal{T}_n, |L|=\ell} |\text{Aut}(H[L])| \cdot 2^{\binom{\ell}{2} - \binom{n}{2}}. \quad (47)$$

We now bound w_ℓ/μ further using a case distinction.

Case $2 \leq \ell \leq n - n^{2/3}$: Here we insert the trivial bound $|\text{Aut}(H[L])| \leq \ell!$ into inequality (47). Writing $\binom{\ell}{2} - \binom{n}{2} = -(n-\ell)(n+\ell-1)/2$, using $\binom{n}{\ell} \leq n^\ell/\ell!$ and $n \geq 2 \log_2 N - 1 + \varepsilon_N$ it follows that

$$\begin{aligned} \frac{w_\ell}{\mu} &\leq N^{n-\ell} \cdot \binom{n}{\ell}^2 \cdot \ell! \cdot 2^{\binom{\ell}{2} - \binom{n}{2}} \\ &\leq \left(N \cdot n^{2\ell/(n-\ell)} \cdot 2^{-(n-1+\ell)/2} \right)^{n-\ell} \\ &\leq \left(2^{2-\varepsilon_N - \ell(1-4(\log n)/(n-\ell))} \right)^{(n-\ell)/2} \leq 2^{-\Omega(\varepsilon_N n)} \leq n^{-\Omega(\log n)}, \end{aligned} \quad (48)$$

where for the second last inequality we optimized over all $2 \leq \ell \leq n - n^{2/3}$, using $(\log n)/n = o(\varepsilon_N)$.

Case $n - n^{2/3} \leq \ell < n$: Here we exploit $H \in \mathcal{A}_n$ to insert $|\text{Aut}(H[L])| = 1$ into inequality (47). Writing again $\binom{\ell}{2} - \binom{n}{2} = -(n-\ell)(n+\ell-1)/2$, using $\binom{n}{\ell} = \binom{n}{n-\ell} \leq n^{n-\ell}$ and $\ell - 2 + \varepsilon_N = \Omega(n)$ it follows that

$$\begin{aligned} \frac{w_\ell}{\mu} &\leq N^{n-\ell} \cdot \binom{n}{\ell}^2 \cdot 2^{\binom{\ell}{2} - \binom{n}{2}} \\ &\leq \left(N \cdot n^2 \cdot 2^{-(n-1+\ell)/2} \right)^{n-\ell} \\ &\leq \left(2^{-(\ell-2+\varepsilon_N)+4 \log n} \right)^{(n-\ell)/2} \leq 2^{-\Omega(n)}. \end{aligned} \quad (49)$$

To sum up: combining (48)–(49) implies $\Lambda_1 = \sum_{2 \leq \ell < n} w_\ell = o(\mu)$, which completes the proof of (3). \square

Remark 5. The above calculations yield $\text{Var } X_H = (1 + o(1))\mathbb{E}X_H$ by (34), and can thus also be used to give an alternative (and compared to Section 2.1 more direct) proof of the sharp threshold result Theorem 1 (i) for $p_2 = 1/2$. We leave the optimal rate of convergence in (3) as an intriguing open problem.

3 Maximum common induced subgraph problem

In this section we prove Theorem 4, i.e., establish two-point concentration of the size I_N of the maximum common induced subgraph of two independent random graphs G_{N,p_1} and G_{N,p_2} with constant edge-probabilities $p_1, p_2 \in (0, 1)$. It naturally splits into two parts: we prove the whp upper bound $I_N \leq \lfloor n_N + \varepsilon_N \rfloor$ in Section 3.1, and prove the whp lower bound $I_N \geq \lfloor n_N - \varepsilon_N \rfloor$ in Section 3.2.

To analyze the parameter n_N defined in (12), it will be convenient to study the auxiliary function

$$g(p) := \max\{\log b_0(p), 2 \log b_1(p), 2 \log b_2(p)\} \quad \text{for } p \in [0, 1], \quad (50)$$

where the functions $b_0(p)$ and $b_i(p)$ are defined as in (8). Note that $g(p)$ depends on $p_1, p_2 \in (0, 1)$, but not on N . The following key analytic lemma establishes lower bounds on b_0, b_1 and b_2 , as well as properties of g (the proof is based on standard calculus techniques, and thus deferred to Appendix A).

Lemma 8. The functions $g(p)$ and $b_j(p)$ with $j \in \{0, 1, 2\}$ have the following properties:

- (i) The function $g : [0, 1] \rightarrow (0, \infty)$ has a unique minimizer $p_0 \in (0, 1)$. Furthermore, there is $\xi = \xi(p_1, p_2) \in (0, 1/2)$ such that $g(p) \geq g(p_0) + \xi$ for all $p \in [0, \xi] \cup [1 - \xi, 1]$.
- (ii) We have $\min_{p \in [0, 1]} b_i(p) \geq 1$ for $i \in \{1, 2\}$. Furthermore, there is $\lambda = \lambda(p_1, p_2) > 1$ such that $\min_{p \in [0, 1]} b_0(p) \geq \lambda$, and for sufficiently large $N \geq N_0(p_1, p_2)$ the following holds for any $p = p(N) \in [0, 1]$ and $i \in \{1, 2\}$: if $x_N^{(i)}(p) \leq x_N^{(0)}(p)$, then $b_i(p) \geq \lambda$. In particular, for any $p = p(N) \in [0, 1]$ we have

$$\min\{x_N^{(0)}(p), x_N^{(1)}(p), x_N^{(2)}(p)\} = \frac{4 \log N}{g(p)} + O(\log \log N), \quad (51)$$

where the implicit constant in $O(\log \log N)$ depends only on p_1, p_2 (and not on p).

Note that, using the uniformity of estimate (51), now (7) and $n_N = \Theta(\log N)$ follow readily from

$$n_N = \max_{p \in [0,1]} \min \left\{ x_N^{(0)}(p), x_N^{(1)}(p), x_N^{(2)}(p) \right\} = \frac{4 \log N}{g(p_0)} + O(\log \log N). \quad (52)$$

3.1 Upper bound: No common induced subgraph of size $\lceil n_N + \varepsilon_N \rceil$

In this section we prove the upper bound in (6) of Theorem 4. More precisely, we show that whp there is no common induced subgraph of size $\lceil n_N + \varepsilon_N \rceil$ in G_{N,p_1} and G_{N,p_2} , which implies the desired whp upper bound $I_N \leq \lceil n_N + \varepsilon_N \rceil - 1 \leq \lfloor n_N + \varepsilon_N \rfloor$. Our proof-strategy employs a refinement of the standard first moment method: the idea is to apply different first moment bounds for different densities of the potential common induced subgraphs, which in turn deal with the three containment bottlenecks discussed in Section 1.3.2 (i.e., containment in G_{N,p_1} , containment in G_{N,p_2} , and containment in both). As we shall see, these estimates are enabled by the corresponding three terms appearing in the definition (12) of n_N .

Turning to the details, to avoid clutter we define

$$n := \lceil n_N + \varepsilon_N \rceil. \quad (53)$$

Writing \mathcal{B}_m for the ‘bad’ event that G_{N,p_1} and G_{N,p_2} have a common induced subgraph with n vertices and m edges, using a standard union bound argument it follows that

$$\mathbb{P}(I_N \geq \lceil n_N + \varepsilon_N \rceil) \leq \sum_{0 \leq m \leq \binom{n}{2}} \mathbb{P}(\mathcal{B}_m). \quad (54)$$

Since we are only interested in equality of common subgraphs up to isomorphisms, we define $\mathcal{G}_{n,m}$ as the set of all unlabeled graphs with n vertices and m edges. Let $X_H^{(1)}$ and $X_H^{(2)}$ denote the number of induced copies of H in G_{N,p_1} and G_{N,p_2} , respectively. The crux is that if \mathcal{B}_m holds, then the three inequalities $\sum_{H \in \mathcal{G}_{n,m}} X_H^{(1)} \geq 1$ and $\sum_{H \in \mathcal{G}_{n,m}} X_H^{(2)} \geq 1$ as well as $\sum_{H \in \mathcal{G}_{n,m}} X_H^{(1)} X_H^{(2)} \geq 1$ all hold. Invoking Markov’s inequality three times, it thus follows that the bad event \mathcal{B}_m holds with probability at most

$$\mathbb{P}(\mathcal{B}_m) \leq \min \left\{ \mathbb{E} \sum_{H \in \mathcal{G}_{n,m}} X_H^{(1)}, \mathbb{E} \sum_{H \in \mathcal{G}_{n,m}} X_H^{(2)}, \mathbb{E} \sum_{H \in \mathcal{G}_{n,m}} X_H^{(1)} X_H^{(2)} \right\}, \quad (55)$$

where the three expectations correspond to the three containment bottlenecks discussed in Section 1.3.2. Let

$$p = p(m, n) := \frac{m}{\binom{n}{2}}. \quad (56)$$

The form of (55) suggests that we might need a good estimate of $|\mathcal{G}_{n,m}|$, but it turns out that we can avoid this: by double-counting labeled graphs on n vertices with m edges, the crux is that we obtain the identity

$$\sum_{H \in \mathcal{G}_{n,m}} \frac{n!}{|\text{Aut}(H)|} = \binom{\binom{n}{2}}{m}, \quad (57)$$

which in view of (32) for G_{N,p_2} interacts favorably with the form of the expectations $\mathbb{E} X_H^{(1)}$ and $\mathbb{E} X_H^{(2)}$. We shall further approximate (57) using the following consequence of Stirling’s approximation formula:

$$\binom{\binom{n}{2}}{m} = e^{O(\log n)} \left(\frac{1}{p^p (1-p)^{1-p}} \right)^{\binom{n}{2}} \quad \text{when } m = p \binom{n}{2}. \quad (58)$$

The heuristic idea for bounding $\mathbb{P}(\mathcal{B}_m)$ is to focus on the smallest expectation in (55) for $m = p \binom{n}{2}$, but due to the definition (12) of the parameter n_N in $n = \lceil n_N + \varepsilon_N \rceil$ it will be easier to use a case distinction depending on which term attains the minimum among $x_N^{(0)}(p)$, $x_N^{(1)}(p)$ and $x_N^{(2)}(p)$.

Case $x_N^{(i)}(p) = \min\{x_N^{(0)}(p), x_N^{(1)}(p), x_N^{(2)}(p)\}$ for $i \in \{1, 2\}$: Here we focus on $\mathbb{E} \sum_{H \in \mathcal{G}_{n,m}} X_H^{(i)}$ in (55). Invoking the estimates (57)–(58) to bound $\sum_{H \in \mathcal{G}_{n,m}} 1/|\text{Aut}(H)|$, by using the definition (8) of $b_i = b_i(p) = (p/p_i)^p[(1-p)/(1-p_i)]^{1-p}$ and $n^2/N = O((\log N)^2/N) = o(1)$ it follows similarly to (32) that

$$\begin{aligned} \mathbb{P}(\mathcal{B}_m) &\leq \mathbb{E} \sum_{H \in \mathcal{G}_{n,m}} X_H^{(i)} = \sum_{H \in \mathcal{G}_{n,m}} \mathbb{E} X_H^{(i)} \\ &= \sum_{H \in \mathcal{G}_{n,m}} \frac{1}{|\text{Aut}(H)|} \cdot (N)_n \cdot p_i^{p \binom{n}{2}} (1-p_i)^{(1-p) \binom{n}{2}} \\ &= e^{O(\log n)} \frac{N^n}{n!} b_i^{-\binom{n}{2}}. \end{aligned} \quad (59)$$

Since $x_N^{(i)}(p) \leq x_N^{(0)}(p)$, by Lemma 8 (ii) we have $b_i = b_i(p) \geq \lambda$ for some constant $\lambda = \lambda(p_1, p_2) > 1$. Inserting Stirling's approximation formula $n! = e^{O(\log n)}(n/e)^n$ and the identity $eN = x_N^{(i)} b_i^{(x_N^{(i)}-1)/2}$ from (13) into estimate (59), using $n \geq n_N + \varepsilon_N = x_N^{(i)} + \varepsilon_N$ (which implies $x_N^{(i)} \leq n$ and $n - x_N^{(i)} \geq \varepsilon_N$) it follows that

$$\mathbb{P}(\mathcal{B}_m) \leq \left[e^{O(\log n/n)} \frac{eN}{n} b_i^{-(n-1)/2} \right]^n = \left[e^{O(\log n/n)} \frac{x_N^{(i)}}{n} b_i^{-(n-x_N^{(i)})/2} \right]^n \leq \lambda^{-\Omega(\varepsilon_N n)} \leq n^{-\Omega(\log n)}. \quad (60)$$

Case $x_N^{(0)}(p) = \min\{x_N^{(0)}(p), x_N^{(1)}(p), x_N^{(2)}(p)\}$: Here we focus on $\mathbb{E} \sum_{H \in \mathcal{G}_{n,m}} X_H^{(1)} X_H^{(2)}$ in (55). Exploiting independence of the two random graphs G_{N,p_1} and G_{N,p_2} , using $|\text{Aut}(H)|^2 \geq |\text{Aut}(H)|$, and applying the definition (8) of $b_0 = b_0(p)$ it follows similarly to (59) that

$$\begin{aligned} \mathbb{P}(\mathcal{B}_m) &\leq \mathbb{E} \sum_{H \in \mathcal{G}_{n,m}} X_H^{(1)} X_H^{(2)} = \sum_{H \in \mathcal{G}_{n,m}} \mathbb{E} X_H^{(1)} \mathbb{E} X_H^{(2)} \\ &= \sum_{H \in \mathcal{G}_{n,m}} \frac{1}{|\text{Aut}(H)|^2} \cdot (N)_n^2 \cdot (p_1 p_2)^{p \binom{n}{2}} ((1-p_1)(1-p_2))^{(1-p) \binom{n}{2}} \\ &\leq e^{O(\log n)} \frac{N^{2n}}{n!} b_0^{-\binom{n}{2}}. \end{aligned} \quad (61)$$

By Lemma 8 (ii) we have $b_0 = b_0(p) \geq \lambda$ for some constant $\lambda = \lambda(p_1, p_2) > 1$. Recalling that $\log N = \Theta(n)$ by (52), observe that the definition (11) of $x_N^{(0)} = x_N^{(0)}(p)$ ensures that

$$x_N^{(0)} b_0^{(x_N^{(0)}-1)/2} = \frac{x_N^{(0)}}{4 \log_{b_0}(N)} \cdot eN^2 = \left(1 + O\left(\frac{\log \log N}{\log N}\right)\right) \cdot eN^2 = e^{O((\log n)/n)} eN^2. \quad (62)$$

Inserting Stirling's approximation formula $n! = e^{O(\log n)}(n/e)^n$ and (62) into (61), using $n \geq n_N + \varepsilon_N = x_N^{(0)} + \varepsilon_N$ (which implies $x_N^{(0)} \leq n$ and $n - x_N^{(0)} \geq \varepsilon_N$) it follows that

$$\mathbb{P}(\mathcal{B}_m) \leq \left[e^{O(\log n/n)} \frac{eN^2}{n} b_0^{-(n-1)/2} \right]^n = \left[e^{O(\log n/n)} \frac{x_N^{(0)}}{n} b_0^{-(n-x_N^{(0)})/2} \right]^n \leq \lambda^{-\Omega(\varepsilon_N n)} \leq n^{-\Omega(\log n)}. \quad (63)$$

To sum up: inserting (60) and (63) into (54) readily gives $\mathbb{P}(I_N \geq \lceil n_N + \varepsilon_N \rceil) = o(1)$, which as discussed completes the proof of the upper bound in (6), i.e., that whp $I_N \leq \lfloor n_N + \varepsilon_N \rfloor$.

3.2 Lower bound: Common induced subgraph of size $\lfloor n_N - \varepsilon_N \rfloor$

In this section we prove the lower bound in (6) of Theorem 4. More precisely, we establish the desired whp lower bound $I_N \geq \lfloor n_N - \varepsilon_N \rfloor$ by showing that whp there is a common induced subgraph of size $\lfloor n_N - \varepsilon_N \rfloor$ in G_{N,p_1} and G_{N,p_2} . Our proof-strategy is inspired by that of Lemma 7 from Section 2.1, though the technical details are significantly more involved: the idea is to pick an ‘optimal’ edge-density p , and then apply the second moment method to the total number of pairs of induced copies of H in G_{N,p_1} and G_{N,p_2} , where we consider only pseudorandom graphs H with $\lfloor n_N - \varepsilon_N \rfloor$ vertices and $\lfloor p \binom{n}{2} \rfloor$ edges. Here the restriction to pseudorandom H will again be key for controlling the expectation and variance, with the extra wrinkle

that the resulting involved variance calculations share some similarities with fourth moment arguments (that require some new ideas to control the ‘overlaps’ of different H , including more careful enumeration arguments).

Turning to the details, we pick p as a maximizer of $\min\{x_N^{(0)}(p), x_N^{(1)}(p), x_N^{(2)}(p)\}$. By comparing the asymptotic estimate (51) with the asymptotics (52) of n_N , it follows from Lemma 8 that there is a constant $\xi = \xi(p_1, p_2) \in (0, 1/2)$ such that $p \in [\xi, 1 - \xi]$ for all sufficiently large N (otherwise the first order asymptotics of (51) and (52) would differ, contradicting our choice of p). To avoid clutter, we define

$$n := \lfloor n_N - \varepsilon_N \rfloor \quad \text{and} \quad m := \lfloor p \binom{n}{2} \rfloor. \quad (64)$$

Recall that in Section 2.1 we introduced the set $\mathcal{T}_{n,m} = \mathcal{A}_n \cap \mathcal{E}_{n,m}$ of pseudorandom graphs with n vertices and m edges, where \mathcal{A}_n and $\mathcal{E}_{n,m}$ are defined as in Section 1.3.3. Since we are only interested in the existence of isomorphisms between induced subgraphs, we now define $\mathcal{U}_{n,m}$ as the unlabeled variant of $\mathcal{T}_{n,m}$, which can formally be constructed by ignoring labels of the graphs in $\mathcal{T}_{n,m}$. Since any graph in $\mathcal{T}_{n,m} \subseteq \mathcal{A}_n$ is asymmetric, we have $|\mathcal{U}_{n,m}| = |\mathcal{T}_{n,m}|/n!$, and so it follows from Lemma 6 that

$$|\mathcal{U}_{n,m}| = (1 + o(1)) \binom{\binom{n}{2}}{m} \frac{1}{n!}. \quad (65)$$

As in Section 3.1, let $X_H^{(1)}$ and $X_H^{(2)}$ denote the number of induced copies of H in G_{N,p_1} and G_{N,p_2} , respectively. We then define the random variable

$$X := \sum_{H \in \mathcal{U}_{n,m}} X_H^{(1)} X_H^{(2)}, \quad (66)$$

where $X > 0$ implies that G_{N,p_1} and G_{N,p_2} have a common induced subgraph on n vertices, i.e., that $I_N \geq n$. To complete the proof of $\mathbb{P}(I_N \geq \lfloor n_N - \varepsilon_N \rfloor) = 1 - o(1)$, using Chebyshev’s inequality it thus suffices to show that $\mathbb{E}X \rightarrow \infty$ and $\text{Var } X = o((\mathbb{E}X)^2)$ as $N \rightarrow \infty$.

We start by showing that $\mathbb{E}X \rightarrow \infty$ as $N \rightarrow \infty$. Analogous to Section 2.1 we have $X_H^{(i)} = \sum_{R_i \in \binom{[N]}{n}} I_{H,R_i}^{(i)}$, where $I_{H,R_i}^{(i)}$ is the indicator random variable for the event that the induced subgraph $G_{N,p_i}[R_i]$ is isomorphic to H . Note that every unlabeled graph $H \in \mathcal{U}_{n,m}$ satisfies the asymmetry property $\mathcal{A}_{n,m}$ from Section 1.3.3, so that $|\text{Aut}(H)| = 1$. It thus follows similarly to (61) in Section 3.1 that

$$\mu := \mathbb{E}X = \sum_{H \in \mathcal{U}_{n,m}} \mathbb{E}X_H^{(1)} \mathbb{E}X_H^{(2)} = |\mathcal{U}_{n,m}| \cdot \binom{N}{n}^2 \cdot \mu_1 \mu_2, \quad (67)$$

where

$$\mu_i := \mathbb{E}I_{H,R_i}^{(i)} = n! p_i^m (1 - p_i)^{\binom{n}{2} - m}. \quad (68)$$

Inspecting the form of (67) and (57), note that the asymptotic estimate (65) of $|\mathcal{U}_{n,m}|$ allows us to estimate $\mathbb{E}X$ analogously to $\mathbb{E} \sum_{H \in \mathcal{G}_{n,m}} X_H^{(1)} X_H^{(2)}$ in (61)–(63). Indeed, using $b_0 = b_0(p) \geq \lambda = \lambda(p_1, p_2) > 1$ and $n \leq n_N - \varepsilon_N \leq x_N^{(0)} - \varepsilon_N$ (which implies $x_N^{(0)} \geq n$ and $x_N^{(0)} - n \geq \varepsilon_N$) it here follows that

$$\begin{aligned} \mathbb{E}X &= (1 + o(1)) \binom{\binom{n}{2}}{m} \frac{1}{n!} \cdot (N)_n^2 (p_1 p_2)^{p \binom{n}{2}} ((1 - p_1)(1 - p_2))^{(1-p) \binom{n}{2}} \\ &= e^{O(\log n)} \frac{N^{2n}}{n!} b_0^{-\binom{n}{2}} = \left[e^{O(\log n/n)} \frac{x_N^{(0)}}{n} b_0^{(x_N^{(0)} - n)/2} \right]^n \geq \lambda^{\Omega(\varepsilon_N n)} \geq n^{\Omega(\log n)} \rightarrow \infty. \end{aligned} \quad (69)$$

The remainder of this section is devoted to showing that $\text{Var } X = o((\mathbb{E}X)^2)$. Note that $I_{H,R_1}^{(1)}$ and $I_{H',R_2}^{(2)}$ are always independent. Since $I_{H,R_i}^{(i)}$ and $I_{H',S_i}^{(i)}$ are independent when $|R_i \cap S_i| \leq 1$, it follows that

$$\text{Var } X \leq \underbrace{\sum_{\substack{0 \leq \ell_1, \ell_2 \leq n: \\ \max\{\ell_1, \ell_2\} \geq 2}} \sum_{H, H' \in \mathcal{U}_{n,m}} \left(\sum_{\substack{R_1, S_1 \in \binom{[N]}{n}: \\ |R_1 \cap S_1| = \ell_1}} \mathbb{E}I_{H,R_1}^{(1)} I_{H',S_1}^{(1)} \right) \left(\sum_{\substack{R_2, S_2 \in \binom{[N]}{n}: \\ |R_2 \cap S_2| = \ell_2}} \mathbb{E}I_{H,R_2}^{(2)} I_{H',S_2}^{(2)} \right)}_{=: w_{\ell_1, \ell_2}}. \quad (70)$$

Our upcoming estimates of this somewhat elaborate variance expression use a case distinction depending on whether $\ell_1 \geq \max\{\ell_2, 2\}$ or $\ell_2 \geq \max\{\ell_1, 2\}$ holds. Both cases can be handled by the same argument (with the roles of G_{N,p_1} and G_{N,p_2} in the below definition (71) of u_{ℓ_1} and v_{ℓ_2} interchanged), so we shall henceforth focus on the case where $\ell_1 \geq \max\{\ell_2, 2\}$ holds. In this case we find it convenient to estimate

$$\frac{w_{\ell_1, \ell_2}}{\mu^2} \leq \underbrace{\frac{\sum_{H, H' \in \mathcal{U}_{n,m}} \sum_{R_1, S_1: |R_1 \cap S_1| = \ell_1} \mathbb{E} I_{H, R_1}^{(1)} I_{H', S_1}^{(1)}}{|\mathcal{U}_{n,m}|^2 \binom{N}{n}^2 \mu_1^2}}_{=: u_{\ell_1}} \cdot \underbrace{\max_{H, H' \in \mathcal{U}_{n,m}} \frac{\sum_{R_2, S_2: |R_2 \cap S_2| = \ell_2} \mathbb{E} I_{H, R_2}^{(2)} I_{H', S_2}^{(2)}}{\binom{N}{n}^2 \mu_2^2}}_{=: v_{\ell_2}}, \quad (71)$$

where the sums are taken over all vertex sets $R_1, S_1 \in \binom{[N]}{n}$ and $R_2, S_2 \in \binom{[N]}{n}$, as in (70) above. This splitting allows us to deal with one random graph G_{N, p_i} at a time, which we shall exploit when bounding u_{ℓ_1} and v_{ℓ_2} in the upcoming Sections 3.2.1–3.2.2. For later reference, we now define the (sufficiently small) constant

$$\zeta := \min \left\{ \frac{\log b_0}{4 \log \max \left\{ \frac{1}{p_1}, \frac{1}{1-p_1}, \frac{1}{p_2}, \frac{1}{1-p_2} \right\}}, \frac{1}{2} \right\}. \quad (72)$$

3.2.1 Contribution of G_{N, p_2} to variance: Bounding v_{ℓ_2}

We first bound the parameter v_{ℓ_2} defined in (71) for $0 \leq \ell_2 \leq n$, using similar arguments as for the variance calculations from Section 2.1. To avoid clutter, in the following we will drop the index from ℓ_2 and simply write $\ell = \ell_2$ for brevity. We start with the pathological case $0 \leq \ell \leq 1$, where in view of (68) we trivially (due to independence of $I_{H, R_2}^{(2)}$ and $I_{H', S_2}^{(2)}$ when $|R_2 \cap S_2| \leq 1$) have

$$v_\ell = \max_{H, H' \in \mathcal{U}_{n,m}} \frac{\sum_{R_2, S_2: |R_2 \cap S_2| = \ell} \mathbb{E} I_{H, R_2}^{(2)} \mathbb{E} I_{H', S_2}^{(2)}}{\binom{N}{n}^2 \mu_2^2} \leq 1 \quad \text{when } 0 \leq \ell \leq 1. \quad (73)$$

It remains to bound v_ℓ for $2 \leq \ell \leq n$. Here we shall reuse some ideas from Section 2.1: bounding the numerator in the definition (71) of v_ℓ as in (35)–(36), in view of $\mu_2 = n! p_2^m (1-p_2)^{\binom{n}{2}-m}$ it follows that

$$\begin{aligned} v_\ell &\leq \frac{\binom{N}{n} \binom{n}{\ell} \binom{N-n}{n-\ell} \cdot n! \cdot (n)_{n-\ell}}{\binom{N}{n}^2 (n!)^2} \cdot \max_{H \in \mathcal{U}_{n,m}, |L|=\ell} |\text{Aut}(H[L])| \\ &\quad \cdot \max_{H \in \mathcal{U}_{n,m}, |L|=\ell} \frac{p_2^{2m-e(H[L])} (1-p_2)^{2\binom{n}{2}-2m-\binom{\ell}{2}+e(H[L])}}{p_2^{2m} (1-p_2)^{2\binom{n}{2}-2m}} \\ &\leq \underbrace{\frac{(N-n)_{n-\ell}}{\binom{N}{n}}}_{=(1+o(1))N^{-\ell}} \cdot \binom{n}{\ell}^2 \cdot \max_{H \in \mathcal{U}_{n,m}, |L|=\ell} |\text{Aut}(H[L])| \cdot \underbrace{\max_{H \in \mathcal{U}_{n,m}, |L|=\ell} p_2^{-e(H[L])} (1-p_2)^{-\binom{\ell}{2}+e(H[L])}}_{=: P_{\ell,2}}, \end{aligned} \quad (74)$$

where the four maxima in (74) are each taken over all vertex-subsets $L \subseteq V(H)$ of size $|L| = \ell$, as before. We define the parameter $P_{\ell,1}$ analogously to $P_{\ell,2}$ (by replacing p_2 with p_1). Since every unlabeled graph $H \in \mathcal{U}_{n,m}$ satisfies the pseudorandom edge-properties of $\mathcal{E}_{n,m}$ from Section 1.3.3, it follows that

$$P_{\ell,i} \leq \max_{k: |k - \binom{\ell}{2} m / \binom{n}{2}| \leq n^{2/3}(n-\ell)} p_i^{-k} (1-p_i)^{-\binom{\ell}{2}+k} \quad (75)$$

for any $0 \leq \ell \leq n$, which in view of $m = \lfloor p \binom{n}{2} \rfloor$ yields, similarly to (37) and (39) from Section 2.1, that

$$P_{\ell,i} \leq \begin{cases} \left[\max \left\{ \frac{1}{p_i}, \frac{1}{1-p_i} \right\} \right]^{\binom{\ell}{2}} & \text{if } 0 \leq \ell \leq \zeta n, \\ e^{O(n^{2/3}(n-\ell)+1)} \cdot (p_i^p (1-p_i)^{1-p})^{-\binom{\ell}{2}} & \text{if } \zeta n \leq \ell \leq n. \end{cases} \quad (76)$$

After these preparations, we are now ready to bound v_ℓ further using a case distinction, where ζ is defined as in (72).

Case $2 \leq \ell \leq \zeta n$: Here we proceed similarly to (37)–(38), and exploit that for all large enough N we have

$$\ell \leq \zeta n \leq \zeta \cdot n_N \leq \zeta \cdot x_N^{(0)}(p) \leq \zeta \cdot \frac{4 \log N}{\log b_0} \leq \frac{\log N}{\log \max\left\{\frac{1}{p_1}, \frac{1}{1-p_1}, \frac{1}{p_2}, \frac{1}{1-p_2}\right\}} \quad (77)$$

by our choice of ζ . Inserting (76) and the trivial bound $|\text{Aut}(H[L])| \leq |L|! = \ell!$ into (74), using $\binom{n}{\ell} \leq n^\ell/\ell!$ and (77) it follows for all large enough N that

$$v_\ell \leq 2N^{-\ell} \cdot \binom{n}{\ell}^2 \ell! \cdot P_{\ell,2} \leq \left(N^{-1} n^2 \cdot \left[\max\left\{ \frac{1}{p_2}, \frac{1}{1-p_2} \right\} \right]^{(\ell-1)/2} \right)^\ell \leq N^{-\ell/3} \leq e^{-\Omega(n)}. \quad (78)$$

Case $\zeta n \leq \ell \leq n$: Here we refine the previous argument, proceeding similarly to (40). For $\zeta n \leq \ell < n - n^{2/3}$ we estimate $|\text{Aut}(H[L])| \leq \ell! \leq n^\ell \leq e^{O(n^{2/3}(n-\ell))}$, and for $n - n^{2/3} \leq \ell \leq n$ we have $|\text{Aut}(H[L])| = 1$ since every unlabeled graph $H \in \mathcal{U}_{n,m}$ satisfies the asymmetry property $\mathcal{A}_{n,m}$ from Section 1.3.3. Inserting these estimates and $\binom{n}{\ell} = \binom{n}{n-\ell} \leq n^{n-\ell} = e^{O(n^{2/3}(n-\ell))}$ into (74), using (76) it follows that

$$v_\ell \leq 2N^{-\ell} \cdot \binom{n}{\ell}^2 \cdot e^{O(n^{2/3}(n-\ell)+1)} \cdot P_{\ell,2} \leq O(1) \cdot \left[N^{-1} (p_2^p (1-p_2)^{1-p})^{-(\ell-1)/2} e^{O(n^{-1/3}(n-\ell))} \right]^\ell, \quad (79)$$

which in Section 3.2.3 will later turn out to be a useful upper bound.

3.2.2 Contribution of G_{N,p_1} to variance: Bounding u_{ℓ_1}

We now bound the parameter u_{ℓ_1} defined in (71) for $2 \leq \ell_1 \leq n$. To avoid clutter, in the following we will drop the index from ℓ_1 and simply write $\ell = \ell_1$ for brevity. For the simpler case $2 \leq \ell \leq \zeta n$ we shall reuse the argument from Section 3.2.1, while the more elaborate case $\zeta n \leq \ell \leq n$ requires further new ideas. Recall that ζ is defined as in (72).

Case $2 \leq \ell \leq \zeta n$: Using the pair $H, H' \in \mathcal{U}_{n,m}$ which maximizes the summand as an upper bound reduces this case to the analogous bounds for v_ℓ from Section 3.2.1. Indeed, by proceeding this way we obtain that

$$u_\ell \leq \max_{H, H' \in \mathcal{U}_{n,m}} \frac{\sum_{R_1, S_1: |R_1 \cap S_1| = \ell} \mathbb{E} I_{H, R_1}^{(1)} I_{H', S_1}^{(1)}}{\binom{N}{n}^2 \mu_1^2} \leq e^{-\Omega(n)}, \quad (80)$$

where the last inequality follows word-by-word (with p_2 replaced by p_1 , and $P_{\ell,2}$ replaced by $P_{\ell,1}$) from the arguments leading to (77)–(78).

Case $\zeta n \leq \ell \leq n$: Here one key idea is to interchange the order of summation in the numerator to obtain

$$u_\ell = \frac{\sum_{R_1, S_1: |R_1 \cap S_1| = \ell} \sum_{H, H' \in \mathcal{U}_{n,m}} \mathbb{E} I_{H, R_1}^{(1)} I_{H', S_1}^{(1)}}{\binom{N}{n}^2 |\mathcal{U}_{n,m}|^2 \mu_1^2}, \quad (81)$$

which then allows us to exploit that not too many choices of the pseudorandom graphs $H, H' \in \mathcal{U}_{n,m}$ can intersect in a ‘compatible’ way. Indeed, if we proceeded similarly to the argument leading to (74), (78) and (80), then after choosing $R_1, S_1 \in \binom{[N]}{n}$ with $|R_1 \cap S_1| = \ell$, in the numerator of (81) we would use

$$|\mathcal{U}_{n,m}|^2 \cdot n! \cdot (n)_{n-\ell} \cdot \max_{H \in \mathcal{U}_{n,m}, |L|=\ell} |\text{Aut}(H[L])| \leq |\mathcal{U}_{n,m}|^2 (n!)^2$$

as a simple upper bound on the number $\Lambda_\ell(R_1, S_1)$ of labeled graphs F on $R_1 \cup S_1$ such that $F[R_1]$ and $F[S_1]$ are isomorphic to some H and H' , respectively, where $H, H' \in \mathcal{U}_{n,m}$. Here our key improvement idea is to more carefully enumerate all possible such graphs F , by first choosing the edges in the intersection $R_1 \cap S_1$ of size $|R_1 \cap S_1| = \ell$, and only then the remaining edges of F . The crux is that since all possible $H, H' \in \mathcal{U}_{n,m}$ satisfy the pseudorandom edge-properties of $\mathcal{E}_{n,m}$ from Section 1.3.3, we know in advance that all possible

numbers k of edges inside $R_1 \cap S_1$ must satisfy $|k - \binom{\ell}{2} m / \binom{n}{2}| \leq n^{2/3}(n - \ell)$. Hence, by first choosing k edges in the intersection $R_1 \cap S_1$, and then the remaining $2(m - k)$ edges, using $m = \lfloor p \binom{n}{2} \rfloor$ it follows that

$$\begin{aligned} \Lambda_\ell(R_1, S_1) &\leq \sum_{k: |k - \binom{\ell}{2} m / \binom{n}{2}| \leq n^{2/3}(n - \ell)} \binom{\binom{\ell}{2}}{k} \binom{2[\binom{n}{2} - \binom{\ell}{2}]}{2(m - k)} \\ &\leq n^2 \cdot \binom{\binom{\ell}{2}}{\lfloor p \binom{\ell}{2} \rfloor} \binom{2[\binom{n}{2} - \binom{\ell}{2}]}{\lfloor 2p[\binom{n}{2} - \binom{\ell}{2}] \rfloor} \cdot n^{O(n^{2/3}(n - \ell) + 1)} \\ &\leq (p^p(1 - p)^{1-p})^{-2\binom{n}{2} + \binom{\ell}{2}} \cdot e^{O(\log n + n^{3/4}(n - \ell))}, \end{aligned} \quad (82)$$

where for the last inequality we used Stirling's approximation formula similarly to (58). Using again Stirling's approximation formula similarly to (58), from the asymptotic estimate (65) for $|\mathcal{U}_{n,m}|$ it follows that

$$|\mathcal{U}_{n,m}|^2 (n!)^2 = (1 + o(1)) \binom{\binom{n}{2}}{m}^2 = e^{O(\log n)} (p^p(1 - p)^{1-p})^{-2\binom{n}{2}}, \quad (83)$$

which in view of $\binom{\ell}{2} = \Theta(n^2)$ makes it transparent that the refined upper bound (82) on $\Lambda_\ell(R_1, S_1)$ is significantly smaller than the simple upper bound $|\mathcal{U}_{n,m}|^2 (n!)^2$ mentioned above. After these preparations, we are now ready to estimate u_ℓ as written in (81): namely, (i) we bound the numerator as in (74), the key difference being that we use $\Lambda_\ell(R_1, S_1)$ to account for the choices of $H, H' \in \mathcal{U}_{n,m}$ and their embeddings into R_1 and S_1 , and (ii) we also use (83) to estimate the denominator in (81). Taking these differences to (74) into account, using $\mu_1 = n! p_1^m (1 - p_1)^{\binom{n}{2} - m}$ it then follows that

$$\begin{aligned} u_\ell &\leq \frac{\binom{N}{n} \binom{n}{\ell} \binom{N-n}{n-\ell}}{\binom{N}{n}^2 |\mathcal{U}_{n,m}|^2 (n!)^2} \cdot \max_{R_1, S_1: |R_1 \cap S_1| = \ell} \Lambda_\ell(R_1, S_1) \cdot \max_{H \in \mathcal{U}_{n,m}, |L| = \ell} \frac{p_1^{2m - e(H[L])} (1 - p_1)^{2\binom{n}{2} - 2m - \binom{\ell}{2} + e(H[L])}}{p_1^{2m} (1 - p_1)^{2\binom{n}{2} - 2m}} \\ &\leq (1 + o(1)) N^{-\ell} \cdot \binom{n}{\ell}^2 \ell! \cdot (p^p(1 - p)^{1-p})^{\binom{\ell}{2}} e^{O(\log n + n^{3/4}(n - \ell))} \cdot P_{\ell,1}, \end{aligned}$$

where $P_{\ell,1}$ is defined as below (74). Combining the upper bound (75)–(76) for $P_{\ell,1}$ with the definition (8) of $b_1 = b_1(p) = (p/p_1)^p [(1 - p)/(1 - p_1)]^{1-p}$, using Stirling's approximation formula $\ell! = e^{O(\log n)} (\ell/e)^\ell$ and $\binom{n}{\ell}^2 = \binom{n}{n-\ell}^2 \leq n^{2(n-\ell)} = e^{O(n^{3/4}(n-\ell))}$ it follows that

$$u_\ell \leq \left(\frac{\ell}{eN} \right)^\ell \cdot b_1^{\binom{\ell}{2}} \cdot e^{O(\log n + n^{3/4}(n - \ell))} \leq \left[e^{O(\log n/n + n^{-1/4}(n - \ell))} \frac{\ell}{eN} b_1^{(\ell-1)/2} \right]^\ell. \quad (84)$$

We now find it convenient to treat the case where b_1 is close to 1 separately from the case where b_1 is bounded away from 1. From (52) we know that $\ell \leq n = O(\log N)$, so there exists a constant $\rho = \rho(p_1, p_2) > 0$ such that $b_1 \leq 1 + \rho$ implies $b_1^{(\ell-1)/2} \leq e^{\rho(\ell-1)/2} \leq \sqrt{N}$. In case of $b_1 \leq 1 + \rho$ we thus infer that

$$u_\ell \leq \left[e^{O(n^{3/4})} \frac{\ell}{eN} b_1^{(\ell-1)/2} \right]^\ell \leq \left[e^{O(n^{3/4})} \frac{\ell}{\sqrt{N}} \right]^\ell \leq n^{-\Omega(\log n)}. \quad (85)$$

In case of $b_1 \geq 1 + \rho$ we insert the identity $eN = x_N^{(1)} b_1^{(x_N^{(1)} - 1)/2}$ from (13) into estimate (84), and then use $\zeta n \leq \ell \leq n \leq n_N - \varepsilon_N \leq x_N^{(1)} - \varepsilon_N$ (which implies $|n - \ell| \leq x_N^{(1)} - \ell$ and $\ell \leq x_N^{(1)}$ as well as $x_N^{(1)} - \ell \geq \varepsilon_N$ and $\ell \geq \zeta n$) to infer that

$$u_\ell \leq \left[e^{O(\log n/n) + o(n - \ell)} \frac{\ell}{x_N^{(1)}} b_1^{-(x_N^{(1)} - \ell)/2} \right]^\ell \leq \left[e^{O(\log n/n)} (1 + \rho)^{-(x_N^{(1)} - \ell)/3} \right]^\ell \leq e^{-\Omega(\varepsilon_N n)} \leq n^{-\Omega(\log n)}. \quad (86)$$

To sum up: from the two estimates (85)–(86) it follows that we always have

$$u_\ell \leq n^{-\Omega(\log n)} \quad \text{when } \zeta n \leq \ell \leq n. \quad (87)$$

3.2.3 Putting things together: Bounding $w_{\ell_1, \ell_2}/\mu^2$

Using the estimates from Sections 3.2.1–3.2.2, in the case $\ell_1 \geq \max\{\ell_2, 2\}$ we are now ready to bound the parameter $w_{\ell_1, \ell_2}/\mu^2 \leq u_{\ell_1} v_{\ell_2}$ from (71) using a case distinction based on whether or not $\ell_2 \leq \zeta n$.

Case $0 \leq \ell_2 \leq \zeta n$ and $2 \leq \ell_1 \leq n$: From (73) and (78), we infer that $v_{\ell_2} \leq 1$ for all $0 \leq \ell_2 \leq \zeta n$. From (80) and (87), we infer that $u_{\ell_1} \leq n^{-\Omega(\log n)}$ for all $2 \leq \ell_1 \leq n$. Hence

$$u_{\ell_1} v_{\ell_2} \leq n^{-\Omega(\log n)}. \quad (88)$$

Case $\zeta n \leq \ell_2 \leq \ell_1 \leq n$: Here we split our analysis into two cases, based on whether or not the upper bound (79) on v_{ℓ_2} is effective on its own. We start with the case where $N(p_2^p(1-p_2)^{1-p})^{(\ell_2-1)/2} \geq e^{n^{3/4}}$ holds: here estimate (79) implies $v_{\ell_2} \leq e^{-\Omega(\ell_2 n^{3/4})} = o(1)$, which together with (87) yields that

$$u_{\ell_1} v_{\ell_2} \leq n^{-\Omega(\log n)}. \quad (89)$$

We henceforth consider the remaining case, where in view of $\ell_1 = \Omega(n)$ we have

$$N(p_2^p(1-p_2)^{1-p})^{(\ell_1+\ell_2-1)/2} \leq e^{n^{3/4}} \cdot (p_2^p(1-p_2)^{1-p})^{\ell_1/2} \leq e^{-\Omega(n)}.$$

By combining the estimates (84) and (79) for u_{ℓ_1} and v_{ℓ_2} with $\binom{\ell_2}{2} = \binom{\ell_1}{2} - (\ell_1 - \ell_2)(\ell_1 + \ell_2 - 1)/2$ as well as $b_0 = b_1 \cdot p_2^{-p}(1-p_2)^{-(1-p)}$ and the identity (62), in view of $\ell_2 = O(n)$ and $\ell_1 = \Omega(n)$ it follows that

$$\begin{aligned} u_{\ell_1} v_{\ell_2} &\leq \left(\left(\frac{\ell_1}{\varepsilon_N} \right)^{\ell_1} b_1^{\binom{\ell_1}{2}} \right) \cdot \left(N^{-\ell_2} (p_2^p(1-p_2)^{1-p})^{-\binom{\ell_2}{2}} \right) \cdot e^{O(\log n + n^{3/4}(n-\ell_2))} \\ &\leq \left[\frac{\ell_1}{\varepsilon_N^2} b_0^{(\ell_1-1)/2} \right]^{\ell_1} \cdot \left[N(p_2^p(1-p_2)^{1-p})^{(\ell_1+\ell_2-1)/2} \right]^{\ell_1-\ell_2} \cdot e^{O(\log n + n^{3/4}(n-\ell_1) + n^{3/4}(\ell_1-\ell_2))} \\ &\leq \left[e^{O(\log n/n) + o(n-\ell_1)} \frac{\ell_1}{x_N^{(0)}} b_0^{-(x_N^{(0)}-\ell_1)/2} \right]^{\ell_1} \cdot e^{-\Omega(n(\ell_1-\ell_2))}. \end{aligned}$$

Recall that by Lemma 8 (ii) we have $b_0 = b_0(p) \geq \lambda$ for some constant $\lambda = \lambda(p_1, p_2) > 1$. Using $\zeta n \leq \ell_1 \leq n \leq n_N - \varepsilon_N \leq x_N^{(0)} - \varepsilon_N$ (which implies $|n - \ell_1| \leq x_N^{(0)} - \ell_1$ and $\ell_1 \leq x_N^{(0)}$ as well as $x_N^{(0)} - \ell_1 \geq \varepsilon_N$ and $\ell_1 \geq \zeta n$) it follows that

$$u_{\ell_1} v_{\ell_2} \leq \left[e^{O(\log n/n)} \lambda^{-(x_N^{(0)}-\ell_1)/3} \right]^{\ell_1} \leq e^{-\Omega(\varepsilon_N n)} \leq n^{-\Omega(\log n)}. \quad (90)$$

To sum up: in the case $\ell_1 \geq \max\{\ell_2, 2\}$ we each time obtained $w_{\ell_1, \ell_2}/\mu^2 \leq n^{-\Omega(\log n)}$. The same argument (with the roles of G_{N, p_1} and G_{N, p_2} interchanged in the definition (71) of u_{ℓ_1} and v_{ℓ_2}) also gives $w_{\ell_1, \ell_2}/\mu^2 \leq n^{-\Omega(\log n)}$ when $\ell_2 \geq \max\{\ell_1, 2\}$. Inserting these estimates into (70) readily implies $\text{Var } X = o((\mathbb{E}X)^2)$. As discussed, this completes the proof of the lower bound in (6), i.e., that whp $I_N \geq \lfloor n_N - \varepsilon_N \rfloor$, which together with the upper bound from Section 3.1 completes the proof of Theorem 4. \square

Remark 6. From Theorem 4 it readily follows that there is a set $M \subseteq \mathbb{N}$ with density 1 such that, for any $N \in M$, we have $\lfloor n_N - \varepsilon_N \rfloor = \lfloor n_N + \varepsilon_N \rfloor$ and thus whp $I_N = \lfloor n_N - \varepsilon_N \rfloor$. We leave it as an interesting open problem to show, that for infinitely many N , the size I_N can be equal to the each of the two different numbers $\lfloor n_N - \varepsilon_N \rfloor$ and $\lfloor n_N + \varepsilon_N \rfloor$ with probabilities bounded away from 0 (similar as for cliques in $G_{N, p}$; see [4, Theorem 11.7]).

4 Concluding remarks

In this paper we resolved the induced subgraph isomorphism problem and the maximum common induced subgraph problem for dense random graphs, i.e., with constant edge-probabilities. Besides the convergence questions already mentioned in Sections 2.2, 2.3 and 3 (see Remarks 4, 5 and 6), there are two main directions for further research: extensions of our main results to sparse random graphs, where edge-probabilities $p_i \rightarrow 0$ are allowed (Problem 1), and generalizations of the graph-sizes (Problem 2).

Problem 1 (Edge-Sparsity). *Prove versions of the induced subgraph isomorphism results Theorems 1–2 and maximum common induced subgraph result Theorem 4 for sparse random graphs.*

As a first step towards such sparse extensions of our main results, one can initially aim at slightly weaker results in the sparse case. For example, in Theorem 4 one could instead try to show that typically $I_N = (1+o(1))\Lambda_N$ for some explicit $\Lambda_N = \Lambda_N(p_1, p_2)$. Furthermore, in Theorem 1 one could instead try to determine some explicit $n^* = n^*(N, p_1, p_2)$ such that $\mathbb{P}(G_{n, p_1} \subseteq G_{N, p_2})$ changes from $1 - o(1)$ to $o(1)$ for $n \leq (1 - \varepsilon)n^*$ and $n \geq (1 + \varepsilon)n^*$, respectively. As another example, in Theorem 4 with $p_1 = p_2 = p$ and $p = p(N) \rightarrow 0$ we wonder if two-point concentration of I_N remains valid for $p \gg N^{-2/3+\varepsilon}$.

Problem 2 (Generalization). *Fix constants $p_1, p_2 \in (0, 1)$. Determine, as $N_1, N_2 \rightarrow \infty$, the typical size I_{N_1, N_2} of the maximum common induced subgraph of the independent random graphs G_{N_1, p_1} and G_{N_2, p_2} .*

This problem aims at a common generalization of our main results, since for $N_1 \leq N_2$ the two extreme cases $N_1 \leq 2 \log_a N_2 - \omega(1)$ and $N_1 = N_2$ are already covered by Theorems 1 and 4. So the real question is what happens in-between: will, similar as in this paper, new concentration phenomena occur?

Acknowledgements. We are grateful to the referees for useful suggestions concerning the presentation. We also would like to thank Svante Janson for suggesting the inclusion of Corollary 3.

References

- [1] N. Alon. Asymptotically optimal induced universal graphs. *Geometric and Functional Analysis* **27** (2017), 1–32.
- [2] N. Alon. Personal communication (2023).
- [3] P. Balister, B. Bollobás, J. Sahasrabudhe, and A. Veremyev. Dense subgraphs in random graphs. *Discrete Applied Mathematics* **260** (2019), 66–74.
- [4] B. Bollobás, *Random Graphs*. 2nd ed., Cambridge University Press, Cambridge (2001).
- [5] V. Bonnici, R. Giugno, A. Pulvirenti, D. Shasha, and A. Ferro. A subgraph isomorphism algorithm and its application to biochemical data. *BMC Bioinformatics* **14** (2013), 996–1010.
- [6] S. Chatterjee and P. Diaconis. Isomorphisms between random graphs. *Journal of Combinatorial Theory, Series B* **160** (2023), 144–162.
- [7] D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence* **18** (2004), 265–298.
- [8] D.J. Cook and L.B. Holder. Substructure discovery using minimum description length and background knowledge. *Journal of Artificial Intelligence Research* **1** (1994), 231–255.
- [9] G. Damiani, C. Solnon, C. de la Higuera, J.-C. Janodet, and É. Samuel. Polynomial algorithms for subisomorphism of nd open combinatorial maps. *Computer Vision and Image Understanding* **115** (2011), 996–1010.
- [10] H.-C. Ehrlich and M. Rarey. Maximum common subgraph isomorphism algorithms and their applications in molecular science: a review. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1** (2011), 680–79.
- [11] P. Erdős and A. Rényi. Asymmetric graphs. *Acta Mathematica Hungarica* **14** (1963), 295–315.
- [12] N. Fountoulakis, R.J. Kang, and C. McDiarmid. Largest sparse subgraphs of random graphs. *European Journal of Combinatorics* **35** (2014), 232–244.
- [13] R. Giugno, V. Bonnici, N. Bombieri, A. Pulvirenti, A. Ferro, and D. Shasha. Grapes: A software for parallel searching on biological graphs targeting multi-core architectures. *PLOS ONE* **8** (2013), 1–11.
- [14] S. Janson. Personal communication (2022).
- [15] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. Wiley-Interscience, New York (2000).
- [16] J. H. Kim, B. Sudakov, and V.H. Vu. On the asymmetry of random regular graphs and random graphs. *Random Structures & Algorithms* **21** (2002), 216–224.
- [17] C. McCreesh, P. Prosser, C. Solnon, and J. Trimble. When subgraph isomorphism is really hard, and why this matters for graph databases. *Journal of Artificial Intelligence Research* **61** (2018), 723–759.
- [18] C. McCreesh, P. Prosser, and J. Trimble. Heuristics and really hard instances for subgraph isomorphism problems. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI’16)*, pp.631–638, AAAI Press, New York (2016).
- [19] J. W. Raymond and P. Willett. Maximum common subgraph isomorphism algorithms for the matching of chemical structures. *Journal of Computer-Aided Molecular Design* **16** (2002), 1573–4951.

A Locating the parameter n_N from Theorem 4

In this appendix we approximately determine the value of the parameter n_N from Theorem 4, which in (12) of Section 1.2 is defined as the solution to an optimization problem over all edge-densities $p \in [0, 1]$. In particular, Lemma 9 below locates n_N in terms of the unique minimizer p_0 of the auxiliary function

$$g(p) = g_{p_1, p_2}(p) := \max\{\log b_0(p), 2 \log b_1(p), 2 \log b_2(p)\} \quad \text{for } p \in [0, 1], \quad (91)$$

where the functions $b_0(p)$ and $b_i(p)$ that depend on p_1, p_2 are defined as in (8). Using the standard convention that $0 \log 0 = \lim_{x \searrow 0} x \log x = 0$ (which is consistent with $0^0 = 1$), for $i \in \{1, 2\}$ we here continuously extend $\log b_i(p)$ to $p \in \{0, 1\}$, as usual. As in Section 1.2. we also introduce the parameter

$$\hat{p} = \hat{p}(p_1, p_2) := \frac{p_1 p_2}{p_1 p_2 + (1 - p_1)(1 - p_2)} \quad (92)$$

that occurs in the different cases of Lemma 9, which each assume information about the form of $g(\hat{p})$. In (93)–(94) below we use $a_N \sim b_N$ as a shorthand for $a_N = (1 + o(1))b_N$ as $N \rightarrow \infty$, as usual.

Lemma 9 (Locating the parameter n_N). *Fix $p_1, p_2 \in (0, 1)$. Then the function $g = g_{p_1, p_2} : [0, 1] \rightarrow (0, \infty)$ from (91) is strictly convex, and has a unique minimizer $p_0 \in (0, 1)$. Furthermore, writing $\hat{p} = \hat{p}(p_1, p_2)$ as in (92), the following holds for p_0 and the parameter n_N from Theorem 4 defined in (12):*

- (i) *If $\log b_0(\hat{p}) > \max\{2 \log b_1(\hat{p}), 2 \log b_2(\hat{p})\}$, then $p_0 = \hat{p}$. Moreover, for all large enough N , we have*

$$n_N = x_N^{(0)}(\hat{p}) \sim 4 \log_{b_0(\hat{p})} N. \quad (93)$$

- (ii) *If $\log b_0(\hat{p}) \leq 2 \log b_i(\hat{p})$ for $i \in \{1, 2\}$, then p_0 is the unique solution of $\log b_0(p) = 2 \log b_i(p)$, and p_0 lies between \hat{p} and p_i , with $p_0 \neq p_i$. Moreover, for all large enough N , we have*

$$n_N = x_N^{(i)}(p_0) + O(\log \log N) \sim 2 \log_{b_i(p_0)} N. \quad (94)$$

The proof of Lemma 9 is based on standard calculus techniques (mainly using convexity), and is spread across the following subsections. As a byproduct of these techniques, in Appendix A.1 and A.2 we also give the deferred proofs of the closely related results Corollary 5, Remark 1 and Lemma 8.

A.1 Proofs of Remark 1 and Lemma 8

The functions $g(p)$ and $\log b_j(p)$ appearing in the following auxiliary lemma of course depend on $p_1, p_2 \in (0, 1)$, but to avoid clutter we henceforth suppress the dependence on p_1, p_2 in the notation, as usual.

Lemma 10. *Fix $p_1, p_2 \in (0, 1)$. The functions $\log b_0(p)$, $\log b_1(p)$ and $\log b_2(p)$ are strictly convex functions for $p \in [0, 1]$, and achieve their unique minima in $[0, 1]$ at \hat{p} , p_1 and p_2 , respectively. Furthermore, the function $g : [0, 1] \rightarrow (0, \infty)$ is strictly convex, and has a unique minimizer in $[0, 1]$ at $p_0 \in (0, 1)$.*

Proof. We start with the functions $\log b_j(p)$. By routine calculus, the first and second derivatives are

$$\begin{aligned} \frac{\partial}{\partial p} \log b_0(p) &= \log \left(\frac{p(1-p_1)(1-p_2)}{(1-p)p_1 p_2} \right), \\ \frac{\partial}{\partial p} \log b_i(p) &= \log \left(\frac{p(1-p_i)}{(1-p)p_i} \right) \quad \text{for } i \in \{1, 2\}, \\ \frac{\partial^2}{\partial^2 p} \log b_j(p) &= \frac{1}{p(1-p)} \quad \text{for } j \in \{0, 1, 2\}. \end{aligned} \quad (95)$$

Note that the $\log b_j(p)$ with $j \in \{0, 1, 2\}$ are all strictly convex functions for $p \in [0, 1]$, since each function is continuous on $[0, 1]$ with a strictly positive second derivative for $p \in (0, 1)$. Note that the first derivatives (95) of the $\log b_j(p)$ are all negative near 0 and positive near 1. So by solving for the zeroes of the first derivative, it follows that $\log b_0(p)$, $\log b_1(p)$ and $\log b_2(p)$ achieve their unique minima at \hat{p} , p_1 and p_2 , respectively.

We now turn to the function $g(p)$. Letting $f_0(x) = \log b_0(x)$, $f_1(x) = 2 \log b_1(x)$ and $f_2(x) = 2 \log b_2(x)$, note that for any $a, b \in [0, 1]$ and $t \in (0, 1)$ with $a \neq b$, by strict convexity of the functions $f_j(x)$ it follows that

$$g(ta + (1-t)b) = \max_{j \in \{0, 1, 2\}} \{f_j(ta + (1-t)b)\} < \max_{j \in \{0, 1, 2\}} \{tf_j(a) + (1-t)f_j(b)\} \leq tg(a) + (1-t)g(b),$$

which establishes that $g(p)$ is a strictly convex function for $p \in [0, 1]$. As uniqueness of the minimizer p_0 of $g(p)$ over $p \in [0, 1]$ follows from strict convexity, it suffices to check that the minimum of $g(p)$ is not attained at the

endpoints $p \in \{0, 1\}$. This follows from the behavior of the first derivatives (95) of the $\log b_j(p)$ established above, which imply that $g(p)$ is decreasing near 0 and increasing near 1. Finally, it remains to determine the range of $g(p)$ for all $p \in [0, 1]$: the upper bound $g(p) \leq \max\{g(0), g(1)\} < \infty$ follows from strict convexity of $g(p)$ and the convention $0 \log 0 = 0$ (mentioned at the beginning of Appendix A), and the lower bound $g(p) \geq \log b_0(p) \geq \log b_0(\hat{p}) > 0$ follows from the properties of $\log b_0(p)$ established above. \square

Proof of Lemma 8. (i): From Lemma 10 we know that $g : [0, 1] \rightarrow (0, \infty)$ has a unique minimizer $p_0 \in (0, 1)$, and that $g(p)$ is strictly convex, which together readily establishes Lemma 8 (i).

(ii): From Lemma 10, for any $p \in [0, 1]$ we know that $b_0(p) \geq b_0(\hat{p}) > 1$ and $b_i(p) \geq b_i(p_i) \geq 1$ for $i \in \{1, 2\}$. This enables us to pick $\eta = \eta(p_1, p_2) > 0$ sufficiently small such that, for all sufficiently large N , we have

$$(4 \log_{b_0(\hat{p})} N + 1) \cdot (1 + \eta)^{4 \log_{b_0(\hat{p})} N} \leq N \quad \text{and} \quad \log b_0(\hat{p}) \geq 3 \log(1 + \eta). \quad (96)$$

Setting $\lambda := 1 + \eta$, we first analyze properties of $x_N^{(0)}(p)$. Since $\log b_0(p) \geq \log b_0(\hat{p}) > \log \lambda > 0$, by inspecting the definition (11) of $x_N^{(0)}(p)$ we infer that, for all sufficiently large N (depending only on λ), we have

$$x_N^{(0)}(p) \leq 4 \log_{b_0(\hat{p})} N + 1, \quad (97)$$

$$x_N^{(0)}(p) = 4 \log_{b_0(p)} N + O(\log \log N), \quad (98)$$

where the asymptotic estimate in (98) is uniform in p , i.e., the implicit error term $O(\log \log N)$ depends only on p_1, p_2 (and not on p). We next turn to $x_N^{(i)}(p)$ with $i \in \{1, 2\}$, where $b_i = b_i(p) \geq 1$ holds. Using the definition (13) it is easy to see that $x_N^{(i)}(p) > 1$ holds for all $N \geq 1$ (since $x_N^{(i)}(p) \leq 1$ implies that $e \leq eN \leq 1 \cdot b_i^{-(1-x_N^{(i)}(p))/2} \leq 1$). For technical reasons we now distinguish whether $b_i = b_i(p)$ is smaller or larger than λ . We start with the case $1 \leq b_i < \lambda$, where using (13) and (96) we infer that

$$x_N^{(i)}(p) \cdot \lambda^{(x_N^{(i)}(p)-1)/2} \geq eN > (4 \log_{b_0(\hat{p})} N + 1) \cdot \lambda^{[(4 \log_{b_0(\hat{p})} N + 1) - 1]/2},$$

which in turn, by noting that $x \lambda^{(x-1)/2}$ is increasing for $x \geq 1$, implies together with (97) that

$$x_N^{(i)}(p) > 4 \log_{b_0(\hat{p})} N + 1 \geq x_N^{(0)}(p). \quad (99)$$

In this case, gearing up towards (51), using (96) we also infer that

$$\frac{2 \log N}{\log b_i(p)} \geq \frac{2 \log N}{\log \lambda} > \frac{4 \log N}{\log b_0(p)}. \quad (100)$$

We next consider the case $b_i \geq \lambda > 1$. Applying a bootstrapping argument to the implicit definition (13) of $x_N^{(i)}(p)$, i.e., $eN = x_N^{(i)}(p) b_i^{(x_N^{(i)}(p)-1)/2}$, for all sufficiently large N (depending only on λ) it follows that

$$\begin{aligned} x_N^{(i)}(p) &= 2 \log_{b_i} N - 2 \log_{b_i} x_N^{(i)}(p) + 2 \log_{b_i} e + 1 \\ &= 2 \log_{b_i} N - 2 \log_{b_i} (2 \log_{b_i} N - 2 \log_{b_i} x_N^{(i)}(p) + 2 \log_{b_i} e + 1) + 2 \log_{b_i} (e) + 1 \\ &= 2 \log_{b_i} N - 2 \log_{b_i} \log_{b_i} N - 2 \log_{b_i} (2/e) + 1 + O\left(\frac{\log \log N}{\log N}\right) \\ &= 2 \log_{b_i} N + O(\log \log N), \end{aligned} \quad (101)$$

where all the asymptotic estimates in (101) are uniform, i.e., do not depend on p (here we exploit that $b_i \geq \lambda > 1$ in this case). In both cases, by combining (98) with (99)–(100) and (101), we obtain the uniform estimate

$$\min\{x_N^{(0)}(p), x_N^{(i)}(p)\} = \min\left\{\frac{4 \log N}{\log b_0(p)}, \frac{2 \log N}{\log b_i(p)}\right\} + O(\log \log N). \quad (102)$$

Applying (102) for both $i \in \{1, 2\}$ then establishes the desired estimate (51) by definition of $g(p)$, completing the proof of Lemma 8 (ii). \square

Proof of Remark 1. The definition of n_N is exactly as in the proof of Theorem 4, so we only need to establish the asymptotic estimate (14). Here the crux is that from (99) we know that $x_N^{(i)}(p) \leq x_N^{(0)}(p)$ implies $b_i \geq \lambda$, so that estimate (101) applies to $x_N^{(i)}(p)$, which immediately establishes (14), as desired. \square

A.2 Proofs of Corollary 5 and Lemma 9

Proof of Corollary 5. By assumption, Lemma 9 (i) implies that $n_N = x_N^{(0)}(\hat{p})$ for all N large enough, Invoking Theorem 4, it thus only remains to verify that $\log b_0(\hat{p}) > \max\{2 \log b_1(\hat{p}), 2 \log b_2(\hat{p})\}$ holds when $p_1 = p_2 = p$. To this end, we start with the trivial inequality

$$\frac{p^4 + (1-p)^4}{p^2 + (1-p)^2} < p^2 + (1-p)^2.$$

Since $\log(x)$ is a concave function, using $\hat{p} = \hat{p}(p, p) = p^2/(p^2 + (1-p)^2)$ and Jensen's inequality it follows that

$$\hat{p} \log p^2 + (1-\hat{p}) \log(1-p)^2 \leq \log \left(\frac{p^4 + (1-p)^4}{p^2 + (1-p)^2} \right) < \log(p^2 + (1-p)^2).$$

Exponentiating both sides, we infer that

$$p^{2\hat{p}}(1-p)^{2(1-\hat{p})} < p^2 + (1-p)^2.$$

Dividing both sides by $(p^2 + (1-p)^2)^2 > 0$, using $p_1 = p_2 = p$ we conclude that

$$b_1(\hat{p})^2 = b_2(\hat{p})^2 = \left(\frac{\hat{p}}{p} \right)^{2\hat{p}} \left(\frac{1-\hat{p}}{1-p} \right)^{2(1-\hat{p})} = \frac{p^{2\hat{p}}(1-p)^{2(1-\hat{p})}}{(p^2 + (1-p)^2)^2} < \frac{1}{p^2 + (1-p)^2} = b_0(\hat{p}),$$

which establishes that $\log b_0(\hat{p}) > \max\{2 \log b_1(\hat{p}), 2 \log b_2(\hat{p})\}$, as desired. \square

Proof of Lemma 9. (i): From Lemma 10, we know that $\log b_0(p)$ achieves its unique minimum at $p = \hat{p}$. Our assumption implies that $g(\hat{p}) = \log b_0(\hat{p})$. For any $p \in [0, 1]$ we thus have

$$g(p) \geq \log b_0(p) \geq \log b_0(\hat{p}) = g(\hat{p}),$$

which establishes that \hat{p} is the unique (see Lemma 10) minimizer p_0 of the function $g(p)$.

It remains to prove that $n_N = x_N^{(0)}(\hat{p})$ for all large enough N . Our assumption implies $\frac{1}{4} \log b_0(\hat{p}) > \max\{\frac{1}{2} \log b_1(\hat{p}), \frac{1}{2} \log b_2(\hat{p})\}$, which by comparing the first order terms in (102), (99) and (101) implies

$$x_N^{(0)}(\hat{p}) < \min \left\{ x_N^{(1)}(\hat{p}), x_N^{(2)}(\hat{p}) \right\} \quad (103)$$

for all N large enough (depending only on the functions $\log b_j(\hat{p})$ and thus p_1, p_2). Fix $p \in [0, 1]$ with $p \neq \hat{p}$. Setting $r := \log b_0(\hat{p}) > 0$ and $\gamma := \log b_0(p) - r > 0$ (where $\gamma > 0$ follows from $p \neq \hat{p}$, as \hat{p} is the unique minimizer of $\log b_0(p)$ by Lemma 10), using $\log(1 + \gamma/r) \leq \gamma/r$ it follows for all N large enough (depending only on $\log b_0(\hat{p})$ and thus p_1, p_2) that

$$\begin{aligned} & x_N^{(0)}(\hat{p}) - x_N^{(0)}(p) \\ &= \frac{4 \log N - 2 \log \log N - 2 \log(4/e) + 2 \log r}{r} - \frac{4 \log N - 2 \log \log N - 2 \log(4/e) + 2 \log(r + \gamma)}{r + \gamma} \\ &= \frac{\gamma}{r(r + \gamma)} \left[4 \log N - 2 \log \log N - 2 \log(4/e) - \frac{2r}{\gamma} \log \left(1 + \frac{\gamma}{r} \right) + 2 \log r \right] \\ &\geq \frac{\gamma}{r(r + \gamma)} \left[2 \log N - 2 + 2 \log \log b_0(\hat{p}) \right] \geq \frac{\gamma \log N}{r(r + \gamma)} > 0. \end{aligned} \quad (104)$$

Combining inequalities (103) and (104) with the definition (12) of n_N , it then follows that

$$x_N^{(0)}(\hat{p}) = \min \left\{ x_N^{(0)}(\hat{p}), x_N^{(1)}(\hat{p}), x_N^{(2)}(\hat{p}) \right\} \leq n_N \leq \max_{p \in [0, 1]} x_N^{(0)}(p) = x_N^{(0)}(\hat{p}).$$

Hence $n_N = x_N^{(0)}(\hat{p})$ for all large enough N (depending only on p_1, p_2), which similar to the asymptotic reasoning for (97) implies (93) by the definition of $x_N^{(0)}(\hat{p})$, and thus completes the proof of Lemma 9 (i).

(ii): First, we show that the solution p^* of $\log b_0(p) = 2 \log b_i(p)$ is unique, and that p^* lies between \hat{p} and p_i , with $p^* \neq p_i$. To this end we introduce the auxiliary function

$$h(p) := 2 \log b_i(p) - \log b_0(p) = p \log \left(\frac{pp_1 p_2}{p_i^2} \right) + (1-p) \log \left(\frac{(1-p)(1-p_1)(1-p_2)}{(1-p_i)^2} \right).$$

Using (95) we see that the second derivative is $h''(p) = 1/[p(1-p)]$. Hence $h(p)$ is a strictly convex function for $p \in [0, 1]$, since it is a continuous function on $[0, 1]$ with a strictly positive second derivative for $p \in (0, 1)$. The above proof of Corollary 5 shows that $p_1 = p_2$ falls in the previous case (i), so that we here have $p_1 \neq p_2$. Consequently, $h(0)$ and $h(1)$ are both not equal to 0, and in particular have different signs (using $0 \log 0 = 0$, as mentioned at the beginning of Appendix A). Using strict convexity it follows that $h(p)$ has a unique zero in $[0, 1]$, which by construction is the unique solution p^* of $\log b_0(p) = 2 \log b_i(p)$. Furthermore, since our assumptions imply $2 \log b_i(\hat{p}) \geq \log b_0(\hat{p})$ and $p_1, p_2 \in (0, 1)$, for $j \in \{1, 2\} \setminus \{i\}$ it follows that

$$h(\hat{p}) \geq 0 \quad \text{and} \quad h(p_i) = p_i \log p_j + (1-p_i) \log(1-p_j) < 0.$$

which establishes that p^* lies between \hat{p} and p_i , with $p^* \neq p_i$.

Second, we claim that $f(p) := \max\{\log b_0(p), 2 \log b_i(p)\}$ is decreasing in $[0, p^*]$ and increasing in $[p^*, 1]$, which together with $\log b_0(p^*) = 2 \log b_i(p^*)$ establishes the useful fact

$$\min_{p \in [0, 1]} \max\{\log b_0(p), 2 \log b_i(p)\} = 2 \log b_i(p^*). \quad (105)$$

By Lemma 10, $\log b_0(p)$ is decreasing in $[0, \hat{p}]$ and increasing in $[\hat{p}, 1]$. Furthermore, $2 \log b_i(p)$ is decreasing in $[0, p_1]$ and increasing in $[p_1, 1]$. Hence $f(p)$ is increasing in $[\max\{p_1, \hat{p}\}, 1]$. Furthermore, recalling that p^* lies between \hat{p} and p_1 , it follows that in $[p^*, \max\{p_1, \hat{p}\}]$ one of $\log b_0(p)$ or $2 \log b_i(p)$ is increasing, whereas the other one is decreasing. Since $\log b_0(p^*) = 2 \log b_i(p^*)$, it follows that $f(p)$ is increasing in $[p^*, \max\{p_1, \hat{p}\}]$, establishing that $f(p)$ is increasing in $[p^*, 1]$. A similar argument shows that $f(p)$ is decreasing in $[0, p^*]$.

Third, we show that p^* is the minimizer of $g(p)$. To see this, note that $\log b_0(p^*) = 2 \log b_i(p^*)$ implies

$$b_0(p^*) = b_0(p^*) \cdot \frac{b_0(p^*)}{b_i(p^*)^2} = \left[\left(\frac{1}{p_j} \right)^p \left(\frac{1}{1-p_j} \right)^{1-p} \right]^2 \geq b_j(p^*)^2$$

for $j \in \{1, 2\} \setminus \{i\}$, so that $g(p^*) = f(p^*) = 2 \log b_i(p^*)$. Using (105), for any $p \in [0, 1]$ it then follows that

$$g(p) \geq \max\{\log b_0(p), 2 \log b_i(p)\} \geq 2 \log b_i(p^*) = g(p^*),$$

which establishes that p^* is the unique (see Lemma 10) minimizer p_0 of the function $g(p)$.

Finally, from Lemma 8 (ii) and (52), using $p_0 = p^* \in (0, 1)$ and $g(p_0) = 2 \log b_i(p_0) > 0$ it follows that

$$n_N = \frac{4 \log N}{g(p_0)} + O(\log \log N) = x_N^{(i)}(p_0) + O(\log \log N) = (1 + o(1)) 2 \log_{b_i(p_0)} N,$$

which establishes (94) and thus completes the proof of Lemma 9 (ii). \square

B Proof of Lemma 6: Pseudorandom properties

Proof of Lemma 6. We start with the binomial random $G_{n,p}$, whose number of edges has a binomial distribution with expected value $\binom{n}{2} \cdot p$. Using Chebyshev's inequality (or Chernoff bounds), it easily follows that

$$\mathbb{P}(G_{n,p} \notin \mathcal{F}_{n,p}) = o(1). \quad (106)$$

Furthermore, the estimate

$$\mathbb{P}(G_{n,p} \text{ is not asymmetric}) \leq e^{-\Theta(np(1-p))} \quad (107)$$

follows, for instance, by the proof of Theorem 3.1 in [16] when replacing the ε in their proof by some small constant (their proof in fact works for any $p = p(n)$ satisfying $np(1-p) \gg \log n$). Note that the

induced subgraph $G_{n,p}[L]$ has the same distribution as $G_{|L|,p}$. Taking a union bound over all possible vertex-subsets $L \subseteq [n]$ of size $\ell := |L| \geq n - n^{2/3}$, using (107) and $\binom{n}{\ell} = \binom{n}{n-\ell} \leq n^{n-\ell} \leq n^{n^{2/3}}$ we obtain

$$\mathbb{P}(G_{n,p} \notin \mathcal{A}_n) \leq \sum_{n-n^{2/3} \leq \ell \leq n} \binom{n}{\ell} \cdot e^{-\Theta(\ell p(1-p))} \leq e^{-\Theta(np(1-p))} = o(1), \quad (108)$$

which in fact holds for any edge-probability $p = p(n)$ satisfying $n^{1/3}p(1-p) \gg \log n$. Combining (106) and (108) establishes that $G_{n,p} \in \mathcal{A}_n \cap \mathcal{F}_{n,p}$ whp.

We next consider the uniform random graph $G_{n,m}$. Using Pittel's inequality (see [4, Theorem 2.2]) with $p' = p'(n) := m/\binom{n}{2} \in [\gamma, 1-\gamma]$, it routinely follows from inequality (108) that

$$\mathbb{P}(G_{n,m} \notin \mathcal{A}_n) \leq 3\sqrt{m} \cdot \mathbb{P}(G_{n,p'} \notin \mathcal{A}_n) \leq O(n) \cdot e^{-\Theta(np'(1-p'))} = o(1). \quad (109)$$

Turning to $\mathcal{E}_{n,m}$, fix any $L \subseteq [n]$ and write $\ell := |L|$ for the size of L , as before. To estimate the number of edges contained in the induced subgraph $G_{n,m}[L]$, we use that by definition of $G_{n,m}$ we have

$$\underbrace{|E(G_{n,m})|}_{=m} = |E(G_{n,m}[L])| + \underbrace{|E(G_{n,m}) \setminus E(G_{n,m}[L])|}_{=:X_L}. \quad (110)$$

For $L = [n]$ this already gives $|E(G_{n,m}[L])| = \binom{|L|}{2}m/\binom{n}{2}$, so it remains to consider the case $L \subsetneq [n]$. The random variable X_L has a hypergeometric distribution with parameters $K := \binom{n}{2}$, $m = \Theta(n^2)$ and

$$k_\ell := \binom{n}{2} - \binom{\ell}{2} = \frac{(n-\ell)(n+\ell-1)}{2},$$

with expected value $\mathbb{E}X_L = k_\ell \cdot m/K = \Theta(n(n-\ell))$, since we have k_ℓ random draws (without replacement) out of a total of $K = \binom{n}{2}$ potential edges, where $m \in [\gamma\binom{n}{2}, (1-\gamma)\binom{n}{2}]$ out of the potential edges are present. Since standard Chernoff bounds for binomial random variables with expected value $k_\ell \cdot m/K$ also apply to X_L , see [15, Theorem 2.1 and 2.10], it routinely follows that

$$\mathbb{P}(|X_L - k_\ell m/K| \geq n^{2/3}(n-\ell)) \leq 2 \exp\left(\frac{-n^{4/3}(n-\ell)^2}{3\mathbb{E}X_L}\right) \leq e^{-\Theta(n^{1/3}(n-\ell))}.$$

Writing $\mathcal{E}'_{n,m}$ for the event that $|X_L - k_\ell m/K| < n^{2/3}(n-\ell)$ for all $L \subsetneq [n]$, using a standard union bound argument and $\binom{n}{\ell} = \binom{n}{n-\ell} \leq n^{n-\ell}$ it readily follows that

$$\mathbb{P}(G_{n,m} \notin \mathcal{E}'_{n,m}) \leq \sum_{0 \leq \ell \leq n-1} \binom{n}{\ell} \cdot e^{-\Theta(n^{1/3}(n-\ell))} = o(1). \quad (111)$$

Combining (109) and (111) establishes that $G_{n,m} \in \mathcal{A}_n \cap \mathcal{E}'_{n,m}$ whp. It remains to show that $G_{n,m} \in \mathcal{E}'_{n,m}$ implies $G_{n,m} \in \mathcal{E}_{n,m}$. To see this note that, for any $L \subsetneq [n]$, equation (110) and $G_{n,m} \in \mathcal{E}'_{n,m}$ imply

$$|E(G_{n,m}[L])| = m - X_L = m - \frac{[\binom{n}{2} - \binom{|L|}{2}]m}{\binom{n}{2}} \pm n^{2/3}(n-|L|) = \frac{\binom{|L|}{2}m}{\binom{n}{2}} \pm n^{2/3}(n-|L|),$$

which completes the proof of Lemma 6 (since this edge-estimate is trivially true for $L = [n]$, as discussed). \square