Article:

Diskin, S. orcid.org/0000-0002-5586-8152, Hoshen, I. and Zhukovskii, M. orcid.org/0000-0001-8763-9533 (2025) A jump of the saturation number in random graphs? Random Structures & Algorithms, 66 (4). e70009. ISSN 1042-9832

https://doi.org/10.1002/rsa.70009

Random Structures & Algorithms

WILEY

RESEARCH ARTICLE OPEN ACCESS

A Jump of the Saturation Number in Random Graphs?

Sahar Diskin[1] | Ilay Hoshen[1] | Maksim Zhukovskii[2]

[1]School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel | [2]School of Computer Science, The University of Sheffield, Sheffield, UK

Correspondence: Sahar Diskin (sahardiskin@mail.tau.ac.il)

Received: 6 April 2023 | **Revised:** 6 February 2025 | **Accepted:** 3 April 2025

Funding: The authors received no specific funding for this work.

Keywords: probablistic combinatorics | random graphs | saturation number

ABSTRACT

For graphs G and F, the saturation number $sat(G, F)$ is the minimum number of edges in an inclusion-maximal F-free subgraph of G. In 2017, Korándi and Sudakov initiated the study of saturation in random graphs. They showed that for constant $p \in (0, 1)$, **whp** $sat\left(G(n, p), K_s\right) = (1 + o(1))n \log_{\frac{1}{1-p}} n$. We show that for every graph F and every constant $p \in (0, 1)$, **whp** $sat(G(n, p), F) = O(n \ln n)$. Furthermore, if every edge of F belongs to a triangle, then the above is the right asymptotic order of magnitude, that is, **whp** $sat(G(n, p), F) = \Theta(n \ln n)$. We further show that for a large family of graphs $\mathcal{F}$ with an edge that does not belong to a triangle, which includes all bipartite graphs, for every $F \in \mathcal{F}$ and constant $p \in (0, 1)$, **whp** $sat(G(n, p), F) = O(n)$. We conjecture that this sharp transition from $O(n)$ to $\Theta(n \ln n)$ depends only on this property, that is, that for any graph F with at least one edge that does not belong to a triangle, **whp** $sat(G(n, p), F) = O(n)$. We further generalize the result of Korándi and Sudakov, and show that for a more general family of graphs $\mathcal{F}'$, including all complete graphs K_s and all complete multipartite graphs of the form $K_{1,1,s_3,\dots,s_\ell}$, for every $F \in \mathcal{F}'$ and every constant $p \in (0, 1)$, **whp** $sat(G(n, p), F) = (1 + o(1))n \log_{\frac{1}{1-p}} n$. Finally, we show that for every complete multipartite graph $K_{s_1,s_2,\dots,s_\ell}$ and every $p \in \left[\frac{1}{2}, 1\right)$, $sat\left(G(n, p), K_{s_1,s_2,\dots,s_\ell}\right) = (1 + o(1))n \log_{\frac{1}{1-p}} n$.

1 | Introduction

1.1 | Background and Main Results

For two graphs G and F, a subgraph $H \subseteq G$ is said to be F-saturated in G if it is a maximal F-free subgraph of G, that is, H does not contain any copy of F as a subgraph, but adding any missing edge $e \in E(G) \setminus E(H)$ creates a copy of F (throughout the paper, we assume for convenience that F does not contain isolated vertices and note that it does not cause any loss of generality). The minimum number of edges in an F-saturated subgraph in G is called the *saturation number*, which we denote by $sat(G, F)$. Zykov [1], and independently Erdős, Hajnal, and Moon [2] initiated the study of the saturation number of graphs, specifically of $sat(K_n, K_s)$. Since then, there has been an extensive study of $sat(K_n, F)$ for different graphs F. Note that the maximum number of edges in an F-saturated graph is $ex(n, F)$, and hence the saturation problem of finding $sat(K_n, F)$ is, in some sense, the opposite of Turán problem. Of particular relevance is the following result due to Kászonyi and Tuza [3], which shows that for every graph F, there exists some constant $c = c(F)$ such that $sat(K_n, F) \le cn$. We refer the reader to [4] for a comprehensive survey on results on saturation numbers of graphs.

In 2017, Korándi and Sudakov initiated the study of the saturation problem for random graphs, that is, when the host graph G is the binomial random graph $G(n, p)$ for constant p. Considering the saturation number for cliques in the binomial random graph, $sat(G(n, p), K_s)$, they showed the following:

Theorem 1.1 (*Theorem 1.1 in* [5]). *Let $0 < p < 1$ be a constant and let $s \ge 3$ be an integer. Then* **whp**[1]

$$sat\left(G(n, p), K_s\right) = (1 + o(1))n \log_{\frac{1}{1-p}} n$$

If F is a graph such that every $e \in E(F)$ belongs to a triangle, it is not hard to see that $sat(G(n, p), F) = \Omega(n \ln n)$. Indeed, the neighborhood $N_H(v)$ of every vertex v in the saturated subgraph H should dominate its neighborhood in $G(n, p)$, and thus **whp** $|N_H(v)|$ should be at least of logarithmic order in n. More precisely, the lower bound in Reference [5] comes from the following:

Theorem 1.2 (*Theorem 2.2 in* [5]). *Let $0 < p < 1$ be a constant. Let F be a graph such that every $e \in E(F)$ belongs to a triangle in F. Then* ***whp***

$$sat(G(n, p), F) \ge (1 + o(1))n \log_{\frac{1}{1-p}} n$$

Following [5], there has been subsequent work on the saturation number $sat(G(n, p), F)$ for other concrete graphs F. Mohammadian and Tayfeh-Rezaie [6] and Demyanov and Zhukovskii [7] proved tight asymptotics for $F = K_{1,s}$. Demidovich, Skorkin and Zhukovskii [8] proved that **whp** $sat\left(G(n, p), C_m\right) = n + \Theta\left(\frac{n}{\ln n}\right)$ when $m \ge 5$, and showed that **whp** $sat(G(n, p), C_4) = \Theta(n)$.

In this paper, we revisit the problem of saturation number in $G(n, p)$. Our first main result gives a general upper bound, holding for all graphs F.

Theorem 1. *Let $0 < p < 1$ be a constant. Let F be an arbitrary graph. Then* ***whp***

$$sat(G(n, p), F) = O(n \ln n)$$

Comparing with the result of [3], Theorem 1 shows that the saturation number in random graphs can be larger than the saturation number in K_n by at most a factor of $O(\ln n)$, whereas Theorem 1.2 shows that this is asymptotically tight.

We note that the hidden constant in $O(n \ln n)$, which we obtain in the proof, may depend on the probability p and the graph F. Note that if every edge of F belongs to a triangle in F, then Theorems 1 and 1.2 imply that **whp** $sat(G(n, p), F) = \Theta(n \ln n)$. In fact, we conjecture that the asymptotics of the saturation number are dictated by the assumption of Theorem 1.2. That is:

Conjecture 1.3. Let $0 < p < 1$ be a constant. If F is a graph such that *every* $e \in E(F)$ belongs to a triangle in F, then **whp**

$$sat(G(n, p), F) = \Theta(n \ln n)$$

On the other hand, if F is a graph such that there *exists* $e \in E(F)$ which does *not* belong to a triangle in F, then **whp**

$$sat(G(n, p), F) = O(n)$$

Our second main result further advances us towards settling this conjecture. We define a family of graphs for which the saturation number in $G(n, p)$ is typically linear in n. We say that a graph F has property ($\not\triangleright$) if there exists a connected

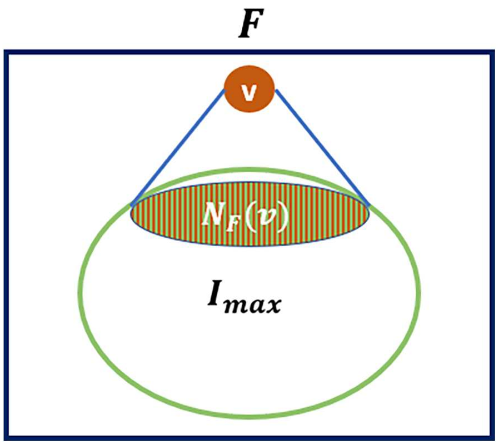

FIGURE 1 | An illustration of a connected graph satisfying property ($\ntriangleright$).

component F' of F which satisfies the following. There exists a largest color class $I_{\max} \subseteq V(F')$, among all proper colorings of F' with $\chi(F')$ colors (where $\chi(F')$ is the chromatic number of F'), and a vertex $v \in V(F') \setminus I_{\max}$ such that $N_{F'}(v) \subseteq I_{\max}$ (see Figure 1). Note that every bipartite graph satisfies property ($\ntriangleright$).

Theorem 2. *Let $0 < p < 1$ be a constant. Let F be a graph with property ($\ntriangleright$). Then* ***whp***

$$sat(G(n, p), F) = O(n)$$

Once again, the hidden constant in $O(n)$ which we obtain in the proof may depend on the probability p and the graph F. Furthermore, observe that the vertex v in Theorem 2 does not belong to a triangle in G. In particular, since we assume F has no isolated vertices and thus v is not an isolated vertex, there exists at least one edge which does not lie in a triangle in F, supporting the second part of Conjecture 1.3, although not resolving it completely. Furthermore, Theorem 2 extends the *asymptotic* results of [8], as every $F = C_m$ also satisfies property ($\ntriangleright$).

The second part of the paper aims for tight asymptotic bounds. Indeed, more ambitiously, one could try and aim for tight asymptotics in the case where every edge of F belongs to a triangle (as in Theorems 1.1 and 1.2). Our next two results aim at extending the tight asymptotics of Theorem 1.1 to a wider family of graphs. The first one extends Theorem 1.1 to complete multipartite graphs for $p \geq \frac{1}{2}$ (note that the case of bipartite graphs, that is $\ell = 2$, is covered by Theorem 2).

Theorem 3. *Let $\ell \geq 3$, $s_1 \leq s_2 \leq \cdots \leq s_\ell$, and let $\frac{1}{2} \leq p < 1$ be a constant. Then* ***whp***

$$sat\left(G(n, p), K_{s_1, \ldots, s_\ell}\right) = (1 + o(1))n \log_{\frac{1}{1-p}} n$$

Remark 1.4. Though it may be possible that the same result holds for constant $p < 1/2$, our proof does not work in this case since one of its main ingredients, Lemma 2.1, fails to be true when $p < 1/2$.

The second one defines a family of graphs for which the above is an asymptotic upper bound for all values of p. For two graphs A and B, we say that a graph B is *A-degenerate* if every two-vertex-connected subgraph of it is a subgraph of A. Furthermore, we say that a graph $F = (V, E)$ has the property ($\star$) if there is an edge $\{u, v\} = e \in E$ such that for every independent set $I \subseteq V$, we have that $F[V \setminus I]$ is non-$F[V \setminus \{u, v\}]$-degenerate, that is, there exists a two-vertex-connected subgraph of $F[V \setminus I]$ which is not a subgraph of $F[V \setminus \{u, v\}]$.

Theorem 4. *Let $0 < p < 1$ be a constant and let F be a graph with property ($\star$). Then* ***whp***

$$sat(G(n, p), F) \leq (1 + o(1))n \log_{\frac{1}{1-p}} n$$

Let us mention a family of graphs satisfying property ($\star$), which could be of particular interest. Let F be a graph with an edge $\{u, v\}$ such that there exists a proper coloring of F with $\chi(F)$ colors, where $\{u\}$ and $\{v\}$ are distinct color classes (see Figure 2), and let us further suppose that for every independent set $I \subseteq V(F)$, $F[V(F) \setminus I]$ is two-vertex-connected.

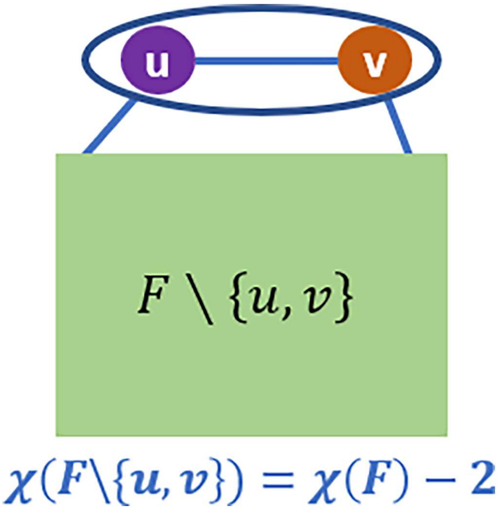

FIGURE 2 | An illustration of the edge $\{u, v\}$ and the remaining graph $F \setminus \{u, v\}$.

Then $F \setminus \{u, v\}$ is $\chi(F) - 2$ colorable, whereas for any independent set I, $F[V \setminus I]$ requires at least $\chi(F) - 1$ colors, and therefore F satisfies property $(\star)$.

Note that if, in addition, every edge of F lies in a triangle, then by Theorem 1.2, we obtain sharp asymptotics. That is, **whp** $sat(G(n, p), F) = (1 + o(1))n \log_{\frac{1}{1-p}} n$. In particular, we obtain the following corollary.

Corollary 1.5. *Let $0 < p < 1$ be a constant. Let $\ell \geq 3$ and $s_3, \ldots, s_\ell \geq 1$ be integers. Let $F = K_{1,1,s_3,\ldots,s_\ell}$. Then **whp***

$$sat(G(n, p), F) = (1 + o(1))n \log_{\frac{1}{1-p}} n$$

Observe that Theorem 3 provides the same asymptotic as Corollary 1.5 for any multipartite graph F, that is, without the requirement that $s_1 = s_2 = 1$, however only for $p \in \left[\frac{1}{2}, 1\right)$.

Theorem 1.2 together with Theorems 3 and 4 suggest that one may make a more ambitious claim in Conjecture 1.3, stating that if every edge of F lies in a triangle, then **whp** $sat(G(n, p), F) = (1 + o(1))n \log_{\frac{1}{1-p}} n$ (see more on that in Section 5).

1.2 | Main Methods and Organization

Let us begin with some conventions that will be useful for us throughout the paper. Given two graphs $H \subseteq G$, and a graph F, we say that an edge $e \in E(G)$ is *completed by* H (or that H *completes* e), if there is a subgraph of $H' \subseteq H$ such that $e \cup H' \cong F$. In this case, we also say that e *completes* (or *closes*) a copy of F. Furthermore, we use the notion of an $\mathcal{F}$*-saturated* graph for a family of graphs $\mathcal{F}$. We say that a graph H is $\mathcal{F}$-saturated in G if H does not contain a copy of *any* $F \in \mathcal{F}$, yet adding any edge $e \in E(G) \setminus E(H)$ closes a copy of*at least one* $F \in \mathcal{F}$.

The structure of the paper is as follows. In Section 2, we present and establish several lemmas that we will use throughout the paper.

In Section 3, we prove Theorems 1 and 2. Both proofs rely on a construction of an $\mathcal{F}$-saturated subgraph of $G(n, p)$ given in the proof of Lemma 3.1. The proof of Theorem 2 follows rather immediately from the above-mentioned construction, whereas the proof of Theorem 1 utilizes an inductive argument based on Lemma 3.1.

In Section 4, we prove Theorems 3 and 4. The proof of Theorem 4 utilizes the construction of [5], refined with the following lemma. For every two graphs G and A, and $\varepsilon > 0$, we say that a graph G is ε*-dense* with respect to A if every induced subgraph of G on at least $\varepsilon|V(G)|$ vertices contains a copy of A. Finally, given a family of graphs $\mathcal{F}$, we say that G is $\mathcal{F}$-free if it does not contain a subgraph isomorphic to F for every $F \in \mathcal{F}$. We will make use of the following result:

Lemma 1.6 ([9, *Theorem 2.1 and Remark 2*]). *Let $p \in (0,1)$. Let A be a graph and let $\mathcal{F}$ be a family of graphs such that every $F \in \mathcal{F}$ is non-A-degenerate. Then for every sufficiently small $\delta > 0$,* ***whp*** *there is a spanning subgraph in $G(n,p)$ which is $\mathcal{F}$-free and $n^{-\delta}$-dense with respect to A.*

We note that in Reference [9], it was shown that the condition on $\mathcal{F}$ in the above lemma is not only sufficient, but also necessary. In particular, this implies some limitations in the results that can be obtained using the construction of [5] (see Section 5 for more details on the matter).

The proof of Theorem 3 is more delicate and is the most involved in this paper. Let us briefly outline the key ideas in the construction, while comparing them with the construction of [5] for cliques.

Very roughly, in Reference [5], one takes a set A on $\Theta(\log n)$ vertices, and let $B := [n] \setminus A$. One can then find in A a spanning subgraph of $G(n,p)$ that is K_{s-1}-free, but $(1/\ln^3|A|)$-dense with respect to K_{s-2}, denote this subgraph by A'. Let H be the subgraph of $G(n,p)$ with all the edges between A and B, all the edges of the subgraph A', and no edges inside B. Then, typically almost any pair of vertices u, v in B is likely to have many common neighbors in A and one can find a copy of K_{s-2} in A' induced by this common neighborhood, such that together with the edge $\{u, v\}$ they close a copy of K_s, while on the other hand the graph is K_s-free since A' is K_{s-1}-free and B is empty.

Thus, in the case of a complete graph, one is mainly concerned with the property that the endpoints of every edge in $G(n,p)[B]$ have a large common neighborhood in A. However, in the case of a complete r-partite graph F with all parts being non-trivial, one cannot assume that in every large induced subgraph of A' there is a copy of F minus an edge, as otherwise it is easy to see that our graph is not F-free. In particular, one cannot consider an empty graph in B. Moreover, a suitable subgraph that we take in B should satisfy the property that every vertex and its B-neighbors have a large common neighborhood in A. To show the likely existence of such a subgraph in B, we use a coupling with an auxiliary random graph in a Hamming space and prove a tight bound for its independence number using a covering-balls argument (see Lemma 4.8 and the paragraph after its statement).

Finally, in Section 5, we discuss the obtained results and some of their limitations, and mention some questions as well as open problems.

2 | Preliminary Lemmas

Given a graph H and a vertex $v \in V(H)$, we denote by $N_H(v)$ the neighborhood of v in H and by $d_H(v) = |N_H(v)|$. Given a subset $S \subseteq V(H)$, we denote by $N_H(S)$ the common neighborhood of all $v \in S$ in H and by $d_H(S) = |N_H(S)|$. That is, $N_H(S) := \bigcap_{v\in S} N_H(v)$. Finally, given subsets $S_1, S_2 \subseteq V(H)$, we denote by $N_H(S_1|S_2)$ the common neighborhood of S_1 in S_2 in the graph H and by $d_H(S_1|S_2) = |N_H(S_1|S_2)|$. That is, $N_H(S_1|S_2) := N_H(S_1) \cap S_2$. When the graph H is clear from context, we may omit the subscript. We denote by K_s^ℓ the complete ℓ-partite graph where each part is of size s. We omit rounding signs for the sake of clarity of presentation.

We will make use of the following bounds on the tail of binomial distribution (see, e.g., [10, Theorem 2.1], [11, Theorem A.1.12], and [5, Claim 2.1]).

Lemma 2.1. *Let $N \in \mathbb{N}$, $p \in [0,1]$, and $X \sim Bin(N,p)$. Then, for $0 \le a \le Np$ and for $b \ge 0$,*

$$\mathbb{P}(|X - Np| \ge a) \le 2\exp\left(-\frac{a^2}{3Np}\right) \tag{1}$$

$$\mathbb{P}(X > bNp) \le \left(\frac{e}{b}\right)^{bNp} \tag{2}$$

$$\mathbb{P}\left(X \le \frac{N}{\ln^2 N}\right) \le (1-p)^{N - \frac{N}{\ln N}} \tag{3}$$

We will require the following probabilistic lemma, which shows that large enough sets are very likely to have a vertex whose number of neighbors in this set deviates largely from the expectation.

For every $p \in (0, 1)$, we set

$$\rho(p) := \frac{1}{1-p} \tag{4}$$

When the choice of p is clear, we may abbreviate $\rho := \rho(p)$.

Lemma 2.2. *Let $p \in \left[\frac{1}{2}, 1\right)$, and let $G \sim G(n, p)$. For every $\gamma \geq 0$, there exists a sufficiently small $\epsilon > 0$ such that the following holds. Let X and Y be disjoint sets of vertices in $V(G)$ of sizes $(1+\epsilon)\log_\rho n$ and at least $\frac{n}{\ln^3 n}$, respectively. Then,*

$$\mathbb{P}\left(\forall y \in Y \ d(y|X) < (1 + (1-\gamma)\epsilon)\log_\rho n\right) \leq \exp(-n^\epsilon)$$

Proof. Note that for every vertex $y \in Y$, the number of neighbors of y in X in the graph G is distributed according to $Bin(|X|, p)$. Hence, for a fixed $y \in Y$,

$$\begin{aligned}
\mathbb{P}\left(d(y|X) \geq (1+(1-\gamma)\epsilon)\log_\rho n\right) &\geq \mathbb{P}\left(d(y|X) = (1+(1-\gamma)\epsilon)\log_\rho n\right) \\
&= \binom{|X|}{(1+(1-\gamma)\epsilon)\log_\rho n} p^{(1+(1-\gamma)\epsilon)\log_\rho n}(1-p)^{\gamma\epsilon\log_\rho n} \\
&\geq \left(\frac{(1+\epsilon)(1-p)}{\gamma\epsilon}\right)^{\gamma\epsilon\log_\rho n} p^{(1+(1-\gamma)\epsilon)\log_\rho n} \\
&\geq (1-p)^{\gamma\epsilon\log_{1-p}\left(\frac{1+\epsilon}{\gamma\epsilon}\right)\log_\rho n + \gamma\epsilon\log_\rho n}(1-p)^{(\log_{1-p} p)(1+(1-\gamma)\epsilon)\log_\rho n} \\
&= n^{-\gamma\epsilon\log_{1-p}\left(\frac{1+\epsilon}{\gamma\epsilon}\right) - (\log_{1-p} p)(1+(1-\gamma)\epsilon) - \gamma\epsilon}
\end{aligned}$$

Since $p \geq \frac{1}{2}$, we have that $\log_{1-p} p \leq 1$ and thus

$$\begin{aligned}
\mathbb{P}\left(d(y|X) \geq (1+(1-\gamma)\epsilon)\log_\rho n\right) &\geq n^{-\gamma\epsilon\log_{1-p}\left(\frac{1+\epsilon}{\gamma\epsilon}\right) - 1 - \epsilon} \\
&\geq n^{-1+2\epsilon}
\end{aligned}$$

where the last inequality is true since for sufficiently small ϵ we have $-\gamma\epsilon\log_{1-p}\left(\frac{1+\epsilon}{\gamma\epsilon}\right) \geq 3\epsilon$.

Therefore, since $d(y|X)$ are independent for $y \in Y$, we obtain that

$$\begin{aligned}
\mathbb{P}\left(\forall y \in Y \ d(y|X) \leq (1+(1-\gamma)\epsilon)\log_\rho n\right) &\leq \left(1 - n^{-1+2\epsilon}\right)^{|Y|} \\
&\leq \exp(-n^\epsilon)
\end{aligned}$$

where in the last inequality we used our assumption that $|Y| \geq \frac{n}{\ln^3 n}$. □

We will also utilize the fact that random graphs typically have relatively small chromatic number (see, e.g., Chapter 7 in Reference [12]):

Lemma 2.3. *Let $0 < p < 1$ be a constant. Then* ***whp*** $\chi(G(n, p)) = O\left(\frac{n}{\ln n}\right)$.

3 | Global Bounds

We begin by giving a construction showing that, for any family of graphs which contains at least one bipartite graph, **whp** the saturation number in $G(n, p)$ is linear in n. This construction will be key for both the proofs of Theorems 1 and 2.

Lemma 3.1. *Let $p \in (0, 1)$. Let $\mathcal{F}$ be a family of graphs that contains at least one bipartite graph. Then* ***whp***

$$sat(G(n, p), \mathcal{F}) = O(n)$$

Proof. Note that if there exist two graphs $F_1, F_2 \in \mathcal{F}$ such that $F_1 \subseteq F_2$ then any graph G is $\mathcal{F}$-saturated if and only if G is $\mathcal{F} \setminus \{F_2\}$-saturated, and thus $sat(G, \mathcal{F}) = sat(G, \mathcal{F} \setminus \{F_2\})$. Hence, we may assume that there are no such two graphs $F_1, F_2 \in \mathcal{F}$.

Set

$$\ell := \min\big\{|V_1| : F \in \mathcal{F},\ F' \subseteq F \text{ is a connected component of } F,\ \chi(F') = 2,\\ V_1 \text{ and } V_2 \text{ are color classes of } F' \text{ and } V(F') = V_1 \sqcup V_2\big\}$$

Let F and its connected component F' be graphs that achieve the minimum above.

We construct a subgraph $H \subseteq G(n,p)$ which is $\mathcal{F}$-saturated such that **whp** $|E(H)| = O(n)$ in a couple of stages. Let $k = |F \setminus F'|$. Fix $v_1 \in V(G(n,p))$, and consider $N_1 := N_{G(n,p)}(v_1)$. Then, for every $2 \le i \le k$, choose v_i from N_{i-1} and set $N_i = N_{i-1} \cap N_{G(n,p)}(v_i)$. Note that for any constant k, **whp** we are able to find such $\{v_1, \dots, v_k\}$, and that they form a clique in $G(n,p)$. Noting that we have not revealed any of the edges induced by $[n] \setminus \{v_1, \dots, v_k\}$, we have that the graph on $[n] \setminus \{v_1, \dots, v_k\}$ is distributed as $G(n-k,p)$, which for constant k and p is essentially the same distribution as $G(n,p)$. Thus, **whp** we are able to find a copy of $F \setminus F'$ in $G(n,p)$, with the rest of the graph distributed as $G \sim G(n-k,p)$. We then set H to be the graph whose vertex set is $V(G(n,p))$, and its set of edges contains only the edges of this copy of $F \setminus F'$. Note that H is $\mathcal{F}$-free by the assumption that there are no two graphs $F_1, F_2 \in \mathcal{F}$ such that $F_1 \subseteq F_2$. Further, we set $V := V(G)$.

If $\ell = 1$, then, as long as there exists an edge in G which does not close a copy of some $\tilde{F} \in \mathcal{F}$ in H, we may add it to H. Note that in this way we increase the degree of every vertex in H by at most $|V(F')| - 2$. Further, as long as there exists an edge not induced by V in $G(n,p)$ (i.e., from the edges touching the copy of $F \setminus F'$ we set aside at the beginning), which does not close a copy of some $\tilde{F} \in \mathcal{F}$ in H, we may add it to H. Since k is a constant, we have added $O(n)$ edges to H, and H is $\mathcal{F}$-saturated.

We may thus assume that $\ell \ge 2$. Let τ be the smallest integer satisfying

$$(1 - p^{\ell-1})^{\tau} n \le n^{2/5} \tag{5}$$

Let us fix τ vertex disjoint sets of size $\ell - 1$ from V, denote them by $A_1, \dots, A_\tau$ and set $A = \bigcup_i A_i$. We then proceed iteratively. In the first step, we set B_1 to be the set of all common neighbors of A_1 in G among the vertices outside A. At the i-th step, where $1 < i \le \tau$, we set B_i to be the set of all common neighbors of A_i in G among the vertices outside $A \cup \bigcup_{j<i} B_j$. Let us add to H all the edges between A_i and B_i in G for every $1 \le i \le \tau$.

Observe that H remains $\mathcal{F}$-free since at this stage we have added to H only bipartite graphs admitting a 2-coloring with one color class of size at most $\ell - 1$. Moreover, every edge in H is incident to B_i, for some $1 \le i \le \tau$. Hence, the number of edges in H thus far is at most

$$e(F \setminus F') + \sum_i |B_i|(\ell - 1) \le e(F \setminus F') + (\ell - 1)n$$

We now turn to add edges to H such that it becomes $\mathcal{F}$-saturated. First, we consider edges whose both endpoints are in B_i, for some $1 \le i \le \tau$. For every $1 \le i \le \tau$, as long as there is an edge in $G[B_i]$ which does not close a copy of some $\tilde{F} \in \mathcal{F}$, we add it to H. Note that the degree of every vertex increased by at most $|F'| - \ell - 1$. Indeed, if for some $1 \le i \le \tau$ there is a vertex $v \in B_i$ with degree $|F'| - \ell$ in $H[B_i]$, then we can form a copy of $K_{\ell,|F'|-\ell}$ with v, $N_{H[B_i]}(v)$, and A_i. However, $F' \subseteq K_{\ell,|F'|-\ell}$, a contradiction since this copy will close a copy of F together with $F \setminus F'$ we set aside at the start of the construction. Hence, the number of edges that are added to H in this step is at most

$$\sum_i |B_i|(|F'| - \ell - 1) \le (|F'| - \ell - 1)n$$

Now, for every $1 \le i \le \tau$, as long as there is an edge between $V \setminus \cup_{j \le i} B_j$ and B_i in G which does not close a copy of some $\tilde{F} \in \mathcal{F}$, we add it to H. Note that, by the same argument as before, at every step i the degree of every vertex $v \in [n] \setminus \cup_{j \le i} B_j$ increased by at most $|F'| - \ell - 1$. For every vertex $v \notin A$, the probability that $v \notin \cup_{j=1}^{i} B_j$ is $(1 - p^{\ell-1})^i$. By Lemma 2.1 (1), the probability that there are at least $2(1-p^{\ell-1})^i n$ such vertices is at most $\exp(-n^{1/4})$. Thus, by the union bound over all less than n choices of $i \le \tau$, **whp**

$$n - \sum_{j=1}^{i} |B_j| \le 2(1-p^{\ell-1})^i n, \quad \forall 1 \le i \le \tau \tag{6}$$

Hence, by (6), the number of edges we add to H in this step is **whp** at most

$$\sum_{i \le \tau} \left(n - \sum_{j=1}^{i} |B_j| \right)(|F'| - \ell - 1) \le \sum_i 2(1-p^{\ell-1})^i n(|F'| - \ell - 1) = O(n)$$

Let us now consider the edges induced by $V \setminus \bigcup_i B_i$. As before, as long as there exists an edge in $G[V \setminus \bigcup_{j=1}^{\tau} B_i]$ which does not close a copy of some $\tilde{F} \in \mathcal{F}$, we add it to H. By (6) (for $i = \tau$) and our choice of τ and (5), we have that **whp** $\left|V \setminus \bigcup_{j=1}^{\tau} B_i\right| \le 2(1-p^{\ell-1})^{\tau} n = o(\sqrt{n})$. Therefore, we only add $o\left(\binom{\left|V \setminus \bigcup_{j=1}^{\tau} B_i\right|}{2}\right) = o(n)$ many edges in this step.

Lastly, there are $O(n)$ edges in $G(n,p)$ which are not induced by V (edges touching the copy of $F \setminus F'$ we set aside at the beginning). Hence, we can add to H these edges one by one, until the resulting graph becomes $\mathcal{F}$-saturated. We thus obtain the required $\mathcal{F}$-saturated subgraph H, where **whp** $|E(H)| = O(n)$. □

The proof of Theorem 2 follows a similar construction to the one in Lemma 3.1.

Proof of Theorem 2. Let F be a graph satisfying property ($\not\triangleright$). Then, there exists a connected component $F' \subseteq F$ with $I_{\max} \subseteq V(F')$ being a color class of maximum size among all proper colorings of F' with $\chi(F')$ colors, and a vertex $v \in V(F') \setminus I_{\max}$ such that $N_{F'}(v) \subseteq I_{\max}$. Let $k := |V(F) \setminus V(F')|$.

As in the proof of Lemma 3.1, we may assume that there are no two connected components F_1, F_2 of F such that $F_1 \subseteq F_2$. Similarly to the proof of Lemma 3.1, let us find a copy of $F \setminus F'$ in $G(n,p)$, and consider the remaining graph G which is distributed as $G(n-k,p)$. Let H be the graph whose vertex set is $V(G(n,p))$ and whose edge set contains only the edges of this copy of $F \setminus F'$. In particular, H is F-free. We now continue to construct an F'-saturated graph in G. Then, the F'-saturated graph in G together with the copy of $F \setminus F'$ (that we set aside) forms an F-free graph in $G(n,p)$, and any edge we add in G closes a copy of F.

The case where F' is a star may be treated in the same manner as in the proof of Lemma 3.1; thus, we may assume that F' is not a star. Set $V := V(G)$, and let τ be as in Equation (5). Let $U \subseteq V$ be a subset of n^{ϵ} vertices, for some small $\epsilon > 0$. **Whp**, we can find a $K_{|V(F')|}$-factor in $G[U]$ (see, i.e., [13]). Let us take τ vertex-disjoint copies of $F' \setminus (I_{\max} \cup \{v\})$ from $G[U]$, denote them by $A_1, \dots, A_{\tau}$, and add them to H. Then, we take B_1 to be the set of common neighbors of A_1 in $V \setminus U$, and for every $1 < i \le \tau$, we set B_i to be the set of all common neighbors of A_i in $V \setminus (U \cup \bigcup_{j=1}^{i-1} B_i)$. We then add to H all the edges between A_i and B_i for every $1 \le i \le \tau$. By the same arguments as in Lemma 3.1, **whp** we added only $O(n)$ edges to H.

Note that, for any $i \in [\tau]$, the graph $H[A_i \sqcup B_i]$ does not contain a copy of F'. Indeed, if we had a copy of F', denote it by $\tilde{F}'$, then

$$\chi(\tilde{F}'[A_i \cap V(\tilde{F}')]) \le \chi(\tilde{F}'[A_i]) \le \chi(F' \setminus (I_{\max} \cup \{v\})) \le \chi(F') - 1$$

and thus we could color $V(\tilde{F}'[B_i \cap V(\tilde{F}')])$ with one color and $V(\tilde{F}'[A_i \cap V(\tilde{F}')])$ with $\chi(F') - 1$ colors, obtaining a color class of size $|V(\tilde{F}'[B_i \cap V(\tilde{F}')])| \ge |I_{\max}| + 1$—a contradiction to the assumption that $I_{\max}$ is a color class of F' of maximum size among all proper colorings with $\chi(F')$ colors.

As in the construction in the proof of Lemma 3.1, we first add to H edges induced by B_i, for all $1 \le i \le \tau$, as long as it remains F-free. Next, for every $1 \le i \le \tau$, as long as there is an edge between $V \setminus \cup_{j \le i} B_j$ and B_i in G which does not close a copy of some $\tilde{F} \in \mathcal{F}$, we add it to H. Observe that connecting any $v \notin A_i$ to a set $S \subseteq B_i$ of size $|I_{\max}|$ creates a copy of F' with v, S, and A_i in H and thus creates a copy of F in H together with the copy of $F \setminus F'$ we set aside at the beginning, and we thus added only $O(n)$ edges at this step. The obtained graph $H[V(G)]$ is F'-saturated in G, and has **whp** $O(n)$ edges. The graph H is, in fact, F-free, and any edge we add in G will create a copy of F in H. Then, again as in the proof of Lemma 3.1, we may add to H all the edges not induced by V (i.e., edges touching the copy of $F \setminus F'$ we set aside at the

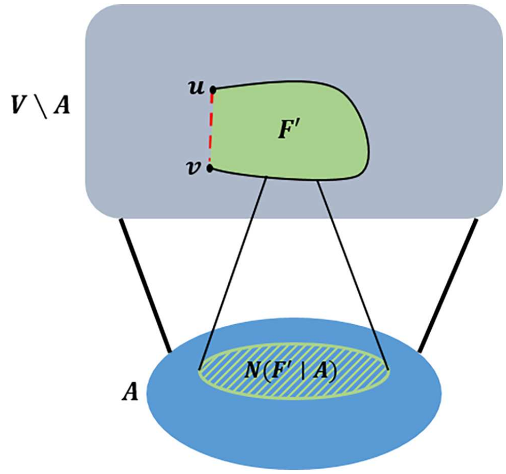

FIGURE 3 | In dashed red line, there is a missing edge u, v which closes a copy of $F' \in \hat{\mathcal{F}}$. Together with its common neighborhood in A (colored light green and blue), this closes a copy of $F \in \mathcal{F}$.

beginning of the construction) one by one, until H is F-saturated in $G(n,p)$. Since k is a constant, we have added only $O(n)$ edges to H.

Utilizing Lemma 3.1, we can now prove Theorem 1.

Proof of Theorem 1. In fact, we will prove a slightly stronger statement: That for any finite family of graphs $\mathcal{F}$, we have that **whp** $sat(G(n,p),\mathcal{F}) = O(n\log n)$ for any fixed $p \in (0,1)$—as we prove by induction, this will help us with the inductive step.

Set $\chi_0 := \chi(\mathcal{F}) = \min_{F\in\mathcal{F}} \chi(F)$. We prove by induction on χ_0. If $\chi_0 = 2$, we are done by Lemma 3.1. We may assume now that $\chi_0 \geq 3$, and that the statement holds for any family $\mathcal{F}'$ with $\chi(\mathcal{F}') < \chi_0$.

We now construct $H \subseteq G(n,p)$ such that **whp** H is $\mathcal{F}$-saturated, and $e(H) = O(n\ln n)$. We begin by letting H be the empty graph. Let A be a set of $C\ln n$ vertices for some large enough constant $C := C(\mathcal{F},p) > 0$. Set

$$\hat{\mathcal{F}} := \{F \setminus I \,:\, F \in \mathcal{F} \text{ and } I \text{ is an independent set of } F\}$$

We stress that we go over all the possible pairs (F, I) where $F \in \mathcal{F}$ and I is an independent set of F. We thus have that for some $F' \in \hat{\mathcal{F}}$, $\chi(F') = \chi_0 - 1$ (note that this holds also when F is disconnected). By induction, **whp** there exists $H' \subseteq [n] \setminus A$ which is $\hat{\mathcal{F}}$-saturated in $G(n,p)[[n] \setminus A]$ with $O(n\ln n)$ edges. Furthermore, for a fixed set X of order $\max_{F\in\mathcal{F}} |V(F)|$ in $[n] \setminus A$, $d_{G(n,p)}(X|A)$ is distributed according to $Bin\left(|A|, p^{|X|}\right)$. Thus, by Lemma 2.1 (1),

$$\mathbb{P}\left(d_{G(n,p)}(X|A) \leq \max_{F\in\mathcal{F}} |V(F)|\right) \leq \exp\left(-\frac{C(\ln n)p^{|X|}}{4}\right) \leq n^{-2\max_{F\in\mathcal{F}}|V(F)|}$$

for C large enough. Thus, by the union bound, **whp** for every set of order $\max_{F\in\mathcal{F}} |V(F)|$ in $[n] \setminus A$, we can find a set of $\max_{F\in\mathcal{F}} |V(F)|$ common neighbors in A. Therefore, when we add a missing edge from $G(n,p)[V \setminus A]$, we close a copy of some $F' \in \hat{\mathcal{F}}$. This copy of F' together with its common neighbors in A form a copy of some $F \in \mathcal{F}$ (see Figure 3).

Let E' be the set of edges of $G(n,p)$ between $[n] \setminus A$ and A. Let H be H' together with E'.

First, note that H is $\mathcal{F}$-free. Indeed, since $H[A]$ is an empty graph, if there is a copy of F in H, $F[A]$ must be an independent set. However, by definition of H', $F[[n] \setminus A]$ is free of any $F \setminus I$ for any independent set I of F—a contradiction.

Furthermore, by construction, every edge of $G(n,p) \setminus E(H)$ in $[n] \setminus A$ closes a copy of $F \in \mathcal{F}$. Since $|E'| = O(n\ln n)$ and **whp** $|E(H')| = O(n\ln n)$, we have that **whp** $|E(H)| = O(n\ln n)$. To ensure that H is $\mathcal{F}$-saturated, the only edges to consider are those where both endpoints are in A, and there are at most $O(\ln^2 n)$ such edges. Hence, **whp** there exists H which is $\mathcal{F}$-saturated with $O(n\ln n)$ edges.

4 | Sharp Bounds

In this section, we prove Theorems 3 and 4. We begin with the proof of Theorem 3.

Let us begin with an outline of the proof. We show that **whp** there exists a subgraph $H \subseteq G(n,p)$ which is $K_{s_1,\dots,s_\ell}$-saturated and $e(H) \le (1+o(1))n\log_{\frac{1}{1-p}} n$. Note that if we find a subgraph $H \subseteq G(n,p)$, with $e(H) \le (1+o(1))n\log_{\frac{1}{1-p}} n$, which completes all but at most $o(n\ln n)$ edges, then we can add edges one by one if necessary, and obtain a subgraph which is $K_{s_1,\dots,s_\ell}$-saturated and has at most $(1+o(1))n\log_{\frac{1}{1-p}} n$ many edges.

Very roughly, we take a subset A of order $\Theta(\ln n)$ and B from $[n]\setminus A$. We then find a subgraph in $G(n,p)[A]$ which is $\left\{K_\ell, K_{s_1}^{(\ell-1)}\right\}$-free and such that there is a copy of $K_{s_1-1,s_3,\dots,s_\ell}$ in every large enough subset of A (where if $s_1=1$, $K_{s_1-1,s_3,\dots,s_\ell} = K_{s_3,\dots,s_\ell}$). In the case of $s_\ell = 1$, as in Reference [5], it suffices to set B as the empty graph and draw all the edges between A and B, and then to show that almost all the edges are completed. However, as in Reference [5], there will still be a small (yet non-negligible) amount of edges that will not be completed, as the co-degree of their endpoints in A is too small. For these type of edges, some additional technical work is required, which will force us to maintain two additional small sets outside of B—A_2 and A_3—which, as in Reference [5], will allow us to deal with these problematic edges.

The case of $s_\ell \ge 2$ is naturally more delicate, as B cannot be taken to be an empty graph, but instead requires some special properties. Using a novel construction, we find a subgraph in $G(n,p)[B]$ which is almost-(s_2-1)-regular (i.e., almost all its vertices are of degree s_2-1, and the others might have smaller degree) and K_{s_1,s_2-s_1+1}-free in $G(n,p)[B]$. This subgraph will have another crucial property—the vertices of any copy of K_{1,s_2-1} have a large common neighborhood in A, which we show through coupling and covering-balls arguments (the construction of this graph is the most involved part of the proof and includes key new ideas, this appears in Lemma 4.1). In this way, almost all edges in B close a copy of K_{1,s_2} such that this copy has a large common neighborhood in A, in which we can find a copy of $K_{s_1-1,s_3,\dots,s_\ell}$ (as in the clique case, some additional technical work is required to deal with the other edges). These two copies form a copy of $K_{s_1,\dots,s_\ell}$ as needed.

Note that the requirement that the subgraph in $G(n,p)[B]$ is K_{s_1,s_2-s_1+1}-free is necessary, as otherwise a copy of $K_{s_1,\dots,s_\ell}$ could be formed when drawing the edges between B and A.

4.1 | Proof of Theorem 3

We may assume that $s_\ell \ge 2$, as the case of cliques has been dealt with in Reference [5].

Let $\gamma, \epsilon > 0$ be sufficiently small constants. Let $G \sim G(n,p)$. Let $L := L(s_1, \dots, s_\ell)$ be a constant large enough with respect to $s_1, \dots, s_\ell$. Let $\rho = \rho(p)$ and set

$$a_1 = \frac{1}{p}(1+(1+\gamma)\epsilon)\log_\rho n, \quad a_2 = L\log_\rho n, \quad a_3 = \frac{a_2}{\sqrt{\ln a_2}}$$

Let A_1, A_2, and A_3, be disjoint sets of vertices of sizes a_1, a_2, and a_3, respectively. Set $B := V \setminus (A_1 \cup A_2 \cup A_3)$. Set

$$I := \left[(1+\epsilon)\log_\rho n,\ (1+(1+2\gamma)\epsilon)\log_\rho n\right]$$

We say that a vertex $v \in B$ is A_1-*good* if $d_G(v|A_1) \in I$. Otherwise, we say that v is A_1-*bad*. Let $B_1 \subseteq B$ be the set of A_1-good vertices, and set $B_2 = B \setminus B_1$. Note that **whp** $|B_2| = O\left(\frac{n}{\ln n}\right)$. Indeed, for every vertex $v \in B$, we have $d_G(v|A_1) \sim \mathrm{Bin}(a_1, p)$. By Lemma 2.1 (1), for every vertex $v \in B$,

$$\mathbb{P}(v \text{ is} A_1 - \text{bad}) \le \exp(-c\ln n)$$

for some constant $c > 0$. Thus, $\mathbb{E}[|B_2|] = O\left(n^{1-c}\right)$. By Markov's inequality, **whp** $|B_2| = O\left(\frac{n}{\log_\rho n}\right)$.

As mentioned prior to the proof, a key element in the proof is finding in B_1 a subgraph H_{B_1} of G which is K_{s_1,s_2-s_1+1}-free and almost-(s_2-1)-regular in the edges in $G[B_1]$. Moreover, we want H_{B_1} to satisfy the following property. Every vertex

$v \in B_1$ and its neighbors in H_{B_1} have a large common neighborhood in A_1 in G. The proof of this key lemma is deferred to the end of this proof.

Lemma 4.1. ***Whp*** *there exists $H_{B_1} \subseteq G$ with $V(H_{B_1}) = B_1$ such that the following holds.*

- *H_{B_1} is $K_{s_1, s_2 - s_1 + 1}$-free.*
- *The maximum degree of H_{B_1} is $s_2 - 1$, and all but $O\left(\frac{n}{\log n}\right)$ of its vertices are of degree $s_2 - 1$ in H_{B_1}.*
- *For every $u, v \in V(H_{B_1})$ such that $\{u, v\} \in E(H_{B_1})$, $d_G(u, v|A_1) \geq (1 + (1 - 6\gamma)\epsilon) \log_\rho n$.*

We move the vertices of degree less than $s_2 - 1$ in H_{B_1} from B_1 to B_2. Note that by Lemma 4.1, **whp** we moved $O\left(\frac{n}{\log n}\right)$ vertices from B_1 to B_2, and this has not affected the above properties of H_{B_1}.

Since K_ℓ and $K_{s_1}^{(\ell-1)}$ are two-vertex-connected graphs (as $\ell \geq 3$ by assumption) and are non-$K_{s_1-1, s_3, \ldots, s_\ell}$-degenerate, by Lemma 1.6, we can take a spanning subgraph $H_{A_1} \subseteq G[A_1]$ which is $(1/\ln^3|A_1|)$-dense with respect to $K_{s_1-1, s_3, \ldots, s_\ell}$ and $\{K_\ell, K_{s_1}^{(\ell-1)}\}$-free.

Let H_1 be the graph on $A_1 \cup B$ with the edges H_{A_1}, H_{B_1}, and all the edges between A_1 and B in G. We now continue with a series of self-contained claims.

Claim 4.2. **Whp**, H_1 completes all but $o(n \ln n)$ of the vertex pairs in B not induced by B_2.

Proof. Fix $S \subset A_1$, $u \in B$. We say that u *avoids* S if $|N_G(u, S)| < \frac{|S|}{\ln^2|S|}$. The number of neighbors of u in S in G is distributed according to $\mathrm{Bin}(|S|, p)$. Thus, by Lemma 2.1 (3), the probability that u avoids S is at most $(1-p)^{|S| - |S|/\ln|S|}$. In particular, if $|S| = \left(1 + \frac{1}{2}\epsilon\right) \log_\rho n$, then the probability that u avoids S is at most

$$(1-p)^{|S|-|S|/\ln|S|} \leq (1-p)^{\left(1+\frac{1}{3}\epsilon\right)\log_\rho n} = n^{-1-\frac{1}{3}\epsilon}$$

Fix a vertex $v \in B$ and expose the edges in G from v to A_1. Assume that v is A_1-good. Fix $S \subset N_G(v, A_1)$ of size $\left(1 + \frac{1}{2}\epsilon\right) \log_\rho n$. Let X_S be the random variable counting the number of vertices $u \in B \setminus \{v\}$ that avoid S. Then X_S is stochastically dominated by $\mathrm{Bin}\left(|B| - 1, n^{-1-\frac{1}{3}\epsilon}\right)$. Since $|B| - 1 < n$, by Lemma 2.1 (2),

$$\mathbb{P}\left(X_S \geq \sqrt{\ln n}\right) \leq \left(\frac{e}{\sqrt{\ln n} \cdot n^{\frac{1}{3}\epsilon}}\right)^{\sqrt{\ln n}} \leq \exp\left(-\frac{1}{3}\epsilon \ln^{\frac{3}{2}} n\right)$$

Note that since v is A_1-good, then $|N_G(v, A_1)| \leq (1 + (1 + 2\gamma)\epsilon) \log_\rho n$. Hence, by the union bound, the probability that there exists $S \subset N_G(v, A_1)$ of size $\left(1 + \frac{1}{2}\epsilon\right) \log_\rho n$ such that $X_S > \sqrt{\ln n}$ is at most

$$\begin{aligned} &\binom{(1 + (1 + 2\gamma)\epsilon) \log_\rho n}{\left(1 + \frac{1}{2}\epsilon\right) \log_\rho n} \exp\left(-\frac{1}{3}\epsilon \ln^{\frac{3}{2}} n\right) \\ &\leq \exp\left(\frac{3}{2}\epsilon \ln\left(\frac{e(1 + 2\epsilon)}{\frac{3}{2}\epsilon}\right) \log_\rho n\right) \exp\left(-\frac{1}{3}\epsilon \ln^{\frac{3}{2}} n\right) = o\left(\frac{1}{n}\right) \end{aligned}$$

Thus, by the union bound **whp** there are no vertices $v \in B_1$ such that $X_S > \sqrt{\ln n}$ for some $S \subset N_G(v, A_1)$ of size $\left(1 + \frac{1}{2}\epsilon\right) \log_\rho n$.

Now, fix an edge $\{u, v\} \subset B$ in G but not in H_1, such that $v \in B_1$. Since $v \in B_1$, we have that $v \in V(H_{B_1})$ and has degree $s_2 - 1$ in H_{B_1}. Therefore, this edge closes a copy of K_{1, s_2} with the neighbors of v in H_{B_1}, denote them by $v_1, \ldots, v_{s_2 - 1}$. Let

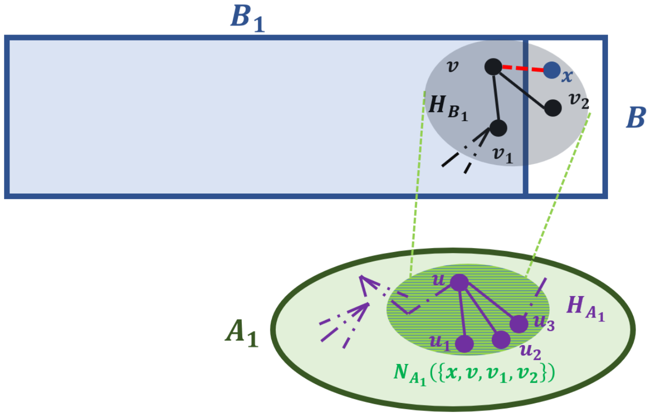

FIGURE 4 | An illustration of how the edge $\{x, v\}$, which is not induced by B_2, is completed by H_1. In this case, we consider $K_{2,3,3}$. In H_{B_1}, we have the vertices $v, v_1 \in B_1$ and $v_2 \in B_2$, with the edges $\{v, v_1\}$ and $\{v, v_2\}$. In the common neighborhood in A_1 of $\{x, v, v_1, v_2\}$, we can find inside H_{A_1} a copy of $K_{1,3}$. Note that with the edges of H_1 and the edge $\{x, v\}$, we now have a copy of $K_{2,3,3}$, with its parts being $\{v, u\}, \{v_1, v_2, x\}, \{u_1, u_2, u_3\}$.

$S := S(v, v_1, \dots, v_{s_2-1})$ be the set of common neighbors of $v, v_1, \dots, v_{s_2-1}$ in A_1 in the graph G. We note that the size of S must be at least $\left(1 + \frac{1}{2}\epsilon\right) \log_\rho n$. Indeed, we have

$$
\begin{aligned}
|S| &\geq |N_G(v, A)| - \sum_{i=1}^{s_2-1} |N_G(v, A) \setminus N_G(v_i, A)| \\
&\geq |N_G(v, A)| - (s_2 - 1)(8\gamma\epsilon) \log_\rho n \\
&\geq (1 + \epsilon - 8 s_2 \gamma \epsilon) \log_\rho n \\
&\geq \left(1 + \frac{1}{2}\epsilon\right) \log_\rho n
\end{aligned}
$$

where the second inequality follows from the fact that, by Lemma 4.1, given $\{x, y\} \in E(H_{B_1})$ we have that $d_G(x, y|A_1) \geq (1 + (1 - 6\gamma)\epsilon) \log_\rho n$, and the last inequality is true if $\gamma \leq \frac{1}{16 s_2}$. Note that if u has more than $\frac{|S|}{\ln^2 |S|}$ neighbors in S, then since H_{A_1} is $(1/\ln^3 |A_1|)$-dense with respect to $K_{s_1-1, s_3, \dots, s_\ell}$, we can find a copy of $K_{s_1-1, s_3, \dots, s_\ell}$ in S such that the edge $\{v, u\}$, joined with the neighbors of v in B_1, completes a copy of $K_{s_1, \dots, s_\ell}$ (see Figure 4). Thus, by this and the above, **whp** there are at most $n\sqrt{\ln n} = o(n \ln n)$ non-completed edges from B not induced by B_2. □

Let us continue with the construction of H. The set A_2 will be crucial when dealing with edges induced by B_2. By Lemma 1.6, we can take a spanning subgraph $H_{A_2} \subseteq G[A_2]$ which is $(1/\ln |A_2|)$-dense with respect to $K_{s_1-1, s_3, \dots, s_\ell}$ and $\{K_\ell, K_{s_1}^{(\ell-1)}\}$-free. Let H_2 be the graph on $A_2 \cup B_2$ with the edges of H_{A_2} and all the edges between A_2 and B_2 in G. We will prove how we close edges induced by B_2 at the last step of the construction—then, we will able to add edges induced by B_2 to H one by one, retaining the property that H is $K_{s_1, \dots, s_\ell}$-free, and show that at that step we have added only $O(n/\ln n) = o(n \ln n)$ edges.

Observe that **whp** we have $\Theta(n \log n)$ edges between A_2 and B_1 in G. We now utilize the set A_3 to complete them. Indeed, by Lemma 2.3 **whp** $k := \chi(G[A_2]) = O\left(\frac{a_2}{\ln a_2}\right)$. We can then split A_2 to k color classes $A_2^1, \dots, A_2^k$. Thus, there are no edges of G (and thus of H_{A_2}) inside A_2^i, for every $i \in [k]$. We further partition the vertices of A_3 to $2k$ (almost) equal parts $A_3^1, \dots, A_3^{2k}$ of size $a_4 := \frac{a_3}{2k} = \Theta(\sqrt{\ln a_2})$. For every $i \in [2k]$, by Lemma 1.6, the probability that there exists a subgraph $H_{A_3^i} \subset G[A_3^i]$ which is $(1/\ln \ln \ln^4 n)$-dense with respect to $K_{s_1-1, s_3, \dots, s_\ell}$ and $\{K_\ell, K_{s_1}^{(\ell-1)}\}$-free is $1 - o(1)$. Hence, by standard binomial tail bounds, **whp** there exist distinct $i_1, \dots, i_k$ such that there exists such a subgraph in each $A_3^{i_j}$, for every $j \in [k]$. Therefore, for every $j \in [k]$, there is $H_{A_3^{i_j}} \subseteq G[A_3^{i_j}]$ which is $(1/\ln \ln \ln^4 n)$-dense with respect to $K_{s_1-1, s_3, \dots, s_\ell}$ and $\{K_\ell, K_{s_1}^{(\ell-1)}\}$-free. We let $A_3^j = A_3^{i_j}$ and $H_{A_3^j} = H_{A_3^{i_j}}$ for simplifying notations and without loss of generality.

Let H_3 be the graph with edges of $H_{A_2}, H_{A_3^1}, \dots, H_{A_3^k}$, together with the edges between A_3 and B_1 in G, and together with the edges between A_3^i and A_2^i, for every $i \in [k]$, in G. Set $H := H_1 \cup H_2 \cup H_3$.

Claim 4.3. **Whp**, H completes all but at most $o(n \log n)$ edges from B_1 to A_2.

Proof. Fix two vertices $v \in B_1$ and $u \in A_2^i$, for some $i \in [k]$.

Recall that v has $s_2 - 1$ neighbors in B. Thus, together with u, we close a copy of K_{1,s_2}. If this copy has more than $\frac{a_4}{\log^3 a_4}$ common neighbors in A_3^i, then we can find among these neighbors a copy of $K_{s_1-1,s_3,\dots,s_\ell}$ which closes a copy of $K_{s_1,\dots,s_\ell}$. By Lemma 2.1 (3), the probability that this copy of K_{1,s_2} has less than $\frac{a_4}{\log^3 a_4}$ common neighbors is at most

$$(1-p^{s_2+1})^{a_4 - a_4/\ln a_4} \leq (1-p^{s_2+1})^{\Omega(\sqrt{\ln a_2})}$$

Hence, the expected number of uncompleted edges $e = \{v, u\}$ with $v \in B_1$ and $u \in A_2$ is at most

$$(1-p^{s_1+1})^{\Omega(\sqrt{\ln a_2})} \cdot |B_1| \cdot |A_2| = (1-p^{s_1+1})^{\Omega(\sqrt{\ln a_2})} \cdot O(n \log_\rho n)$$

Therefore, by Markov's inequality, **whp** the number of uncompleted such pairs is $o(n \log_\rho n)$. □

Claim 4.4. **Whp**, $e(H) = (1 + \Theta(\epsilon))n \log_\rho n$.

Proof. Indeed, **whp**

$$\begin{aligned} e(H) &\leq e(H_1) + e(H_2) + e(H_3) \\ &\leq (1+o(1))a_1 np + e(H_{B_1}) + e(H_{A_1}) + a_2|B_2|p + e(H_{A_2}) + e(H_{A_3}) + a_3 np \\ &\leq (1+o(1))(1+(s_1+1)\epsilon)n \log_\rho n = (1+\Theta(\epsilon))n \log_\rho n \end{aligned}$$

□

Claim 4.5. H is $K_{s_1,\dots,s_\ell}$-free.

Proof. Suppose towards contradiction that there exists a copy of $K_{s_1,\dots,s_\ell}$ in H, denote this copy by F.

Assume first that $V(F) \cap A_3 \neq \emptyset$. Set $\tilde{F} := V(F) \cap A_3$. Suppose further that $V(F) \cap (B_2 \cup A_1) \neq \emptyset$. Since there are no edges between $B_2 \cup A_1$ and A_3, all the vertices in $V(F) \cap (B_2 \cup A_1 \cup A_3)$ must belong to the same independent set of F. Thus, $V(F) \cap (B_1 \cup A_2)$ must contain a copy of $K_{s_1,\dots,s_{\ell-1}}$. Since there are no edges between B_1 and A_2, this copy is contained entirely in B_1 or entirely in A_2. Since $H[B_1]$ is K_{s_1,s_2-s_1+1}-free and $K_{s_1,s_2-s_1+1} \subseteq K_{s_1,\dots,s_{\ell-1}}$, and $H[A_2]$ is $K_{s_1}^{(\ell-1)}$-free and $K_{s_1}^{(\ell-1)} \subseteq K_{s_1,\dots,s_{\ell-1}}$, this is a contradiction, and hence $V(F) \cap (B_2 \cup A_1) = \emptyset$. Since $H[A_3]$ is K_ℓ-free, then at least one full part I' of $K_{s_1,\dots,s_\ell}$ must come from $B_1 \cup A_2$, as all its vertices must be adjacent to every vertex of $\tilde{F}$. If $F \setminus I'$ lies entirely in A_3, then $H[A_3]$ must contain a copy of $K_{s_1}^{(\ell-1)}$—a contradiction, as by construction $H[A_3]$ is composed of vertex disjoint subsets, each of which is $K_{s_1}^{(\ell-1)}$-free. Hence, there must be another vertex $v \in F \setminus I'$ which is in $B_1 \cup A_2$. Thus, $B_1 \cup A_2$ contains at least one full part as well as an additional vertex from another part of $K_{s_1,\dots,s_\ell}$. As this additional vertex is adjacent to I' in $B_1 \cup A_2$ and there are no edges between B_1 and A_2, all these vertices must belong exclusively to B_1 or exclusively to A_2. If they belong to B_1, then since the maximum degree of H_{B_1} is $s_2 - 1$ we have that the full part must be of size s_1. Furthermore, we have that H_{B_1} is K_{s_1,s_2-s_1+1}-free. Thus, if they belong to B_1, $H[A_3]$ must contain $K_{s_1}^{(\ell-1)}$ in order to complete $K_{s_1,\dots,s_\ell}$, as there are $\ell - 1$ parts missing at least s_1 vertices, once again leading to contradiction. Finally, if they belong to A_2, then they should be split between different independent sets A_2^i and A_2^j, but there are no vertices in A_3 which are adjacent to both A_2^i and A_2^j in H—a contradiction.

Let us now assume that $V(F) \cap A_3 = \emptyset$ and $V(F) \cap A_2 \neq \emptyset$. Set $\tilde{F} := V(F) \cap A_2$. Similarly to before, we may assume that $V(F) \cap (B_1 \cup A_1) = \emptyset$, as otherwise $V(F) \cap B_2$ must contain $K_{s_1,\dots,s_{\ell-1}}$ and in particular a vertex of degree s_2, contradicting the fact that every $v \in B$ has degree at most $s_2 - 1$ in H. Since $H[A_2]$ is K_ℓ-free, then a full part of $K_{s_1,\dots,s_\ell}$ must come from B_2 and, as before, at least one additional vertex. This vertex is adjacent to all the vertices in the full part, and thus, since every $v \in B$ has degree at most $s_2 - 1$ in H, the full part must be of size s_1. Moreover, there are at most $s_2 - s_1$ vertices of F in B_2 that do not belong to this part of size s_1 (otherwise, $H[B_2]$, and in particular $H[B]$ would contain a copy of K_{s_1,s_2-s_1+1}, contradicting the fact that the set of edges of $H[B]$ is the set of edges of H_{B_1}). Therefore, each part of $\tilde{F}$ must contain at least s_1 vertices, creating a copy of $K_{s_1}^{(\ell-1)}$—a contradiction to the fact that $H[A_2]$ is $K_{s_1}^{(\ell-1)}$-free.

Suppose now that $V(F) \cap A_3 = V(F) \cap A_2 = \emptyset$ and $V(F) \cap A_1 \neq \emptyset$. Once again, by our assumptions on $H[A_1]$, we obtain that there must be a full side and an additional vertex of F in B. Since the degree of every $v \in B$ in H is at most $s_2 - 1$, we have that this side is of size s_1. Since $H[B]$ (whose set of edges is as the set of edges of H_{B_1}) is K_{s_1,s_2-s_1+1}-free, there are at most $s_2 - s_1$ vertices of F in B that do not belong to this part of size s_1. Thus, each part of $V(F) \cap A_3$ must contain at least s_1 vertices, creating a copy of $K_{s_1}^{(\ell-1)}$—a contradiction.

We may thus assume that $V(F) \subseteq B$. However, $H[B]$ is K_{s_1,s_2-s_1+1}-free and $K_{s_1,s_2-s_1+1} \subseteq K_{s_1,\dots,s_\ell}$—a contradiction, thus completing the proof. □

In conclusion, **whp** by Claim 4.5 H is $K_{s_1,\dots,s_\ell}$-free and by Claims 4.2 and 4.3 H completes all but at most $o(n \ln n)$ of the edges not induced by B_2. Let us add each of these edges (not induced by B_2) which do not close a copy of F to H until none remain. By Claim 4.4, we have now that **whp** $e(H) = (1 + \Theta(\epsilon))n \log_\rho n + o(n \ln n) = (1 + \Theta(\epsilon))n \log_\rho n$. We are thus left with the edges of G induced by B_2. Let us show that we can add each of these edges that do not close a copy of F to H until none remain, and in doing so, add at most $o(n \log n)$ edges. As we may choose ϵ arbitrarily small, we will then obtain a graph H which is $K_{s_1,\dots,s_\ell}$-saturated with $(1 + o(1))n \log_\rho n$ edges.

To the task at hand, recall that $L := L(s_1, \dots, s_\ell)$ is a constant large enough with respect to $s_1, \dots, s_\ell$, and that $|A_2| = L \log_\rho n$. We may thus choose L large enough such that **whp** every $s_2 + 1$ vertices from B_2 have at least $\frac{a_2 p^{s_2+1}}{10}$ common neighbors in A_2 in G. Indeed, for a fixed set of $s_2 + 1$ vertices, the number of their common neighbors in A_2 is distributed according to $Bin(|A_2|, p^{s_2+1})$. Lemma 2.1 together with a union bound on the less than n^{s_2+1} choices of such sets in B_2, completes this claim. Since $H[A_2]$ is $(1/\ln|A_2|)$-dense with respect to $K_{s_1-1,s_3,\dots,s_\ell}$, these common neighbors in A_2 induce a copy of $K_{s_1-1,s_3,\dots,s_\ell}$ in $H[A_2]$. Hence, there cannot be a copy of K_{1,s_2} in $H[B_2]$ (i.e., a vertex of degree s_2 in $H[B_2]$), as it closes a copy of $K_{s_1,\dots,s_\ell}$ together with its common neighbors in A_2. Now, as long as there are edges of G induced by B_2 which do not close a copy of $K_{s_1,\dots,s_\ell}$, we add them to H. We thus only added at most $s_2 \cdot |B_2| = O(n/\log n)$ edges to H in this final step.

4.2 | Proof of Lemma 4.1

We build such a subgraph iteratively. In each iteration, we find a large matching in $G[B_1]$, which we add to the subgraph, such that the matching satisfies the following:

1. The union of the previous matching together with this matching does not induce a copy of K_{s_1,s_2-s_1+1}; and,
2. the endpoints of every edge in the matching have a large common neighborhood in A_1 in G.

Denote by Γ the auxiliary graph with vertex set B_1 and the set of edges defined as follows. For every two vertices $v \neq u$ in B_1,

$$\{u, v\} \in E(\Gamma) \Leftrightarrow d_G(u, v|A_1) \geq (1 + (1 - 6\gamma)\epsilon) \log_\rho n$$

Set $p' = 1 - (1 - p)^{1/(s_2-1)}$. For every $i \in [s_2 - 1]$, denote by Γ_i the subgraph of Γ obtained by retaining every edge independently with probability p'. We have $\Gamma_p := \Gamma \cap G(n, p)$. Note that Γ_p has the same distribution as $\bigcup_{i=1}^{s_2-1} \Gamma_i$.

We will, in fact, prove the following equivalent lemma:

Lemma 4.6. ***Whp*** *there exists $H_{B_1} \subseteq \Gamma_p$ which is K_{s_1,s_2-s_1+1}-free, has maximum degree $s_2 - 1$, and all but $O\left(\frac{n}{\log n}\right)$ of its vertices have degree $s_2 - 1$.*

We consider two cases separately. In the first case, we assume that $p = \frac{1}{2}$. While some details will be different when $p > \frac{1}{2}$, we believe the key ideas—in particular the ball-covering technique (see Claim 4.9)—are clearer in this case. Afterwards, we mention how to complete the proof for the range of $p > \frac{1}{2}$, where, in particular, Claim 4.7 no longer necessarily holds.

4.2.1 | Proof of Lemma 4.1, P=0.5

Recall that we are seeking a subgraph of Γ_p which is K_{s_1,s_2-s_1+1}-free, has maximum degree $s_2 - 1$ and all but $O\left(\frac{n}{\log n}\right)$ of its vertices are of degree $s_2 - 1$.

As for the first requirement, note that a graph whose maximum degree is $s_2 - 1$ and is C_4-free, is $K_{s_1, s_2 - s_1 + 1}$-free graph. Indeed, if $s_1 = s_2$ or $s_1 = 1$, we have that $K_{s_1, s_2 - s_1 + 1} = K_{1, s_2}$, and thus asking for the maximum degree to be $s_2 - 1$ suffices. Otherwise, $1 < s_1 < s_2$ and we have that any copy of $K_{s_1, s_2 - s_1 + 1}$ contains $K_{2,2}$, that is, C_4.

As for the second requirement, our strategy will then be to find a sufficiently large matching M_i in Γ_i, for every $i \in [s_2 - 1]$, such that there are no copies of C_4 in $M := \bigcup_{i=1}^{s_2-1} M_i$. We will show that for every $i \in [s_2 - 1]$, there are at most $O\left(\frac{n}{\log n}\right)$ vertices that are unmatched. Thus, the subgraph whose edges are the edges of M would then be the desired subgraph of Γ_p.

Fix $i \in \{2, \dots, s_2 - 1\}$. Assume that there exist edge-disjoint matchings $M_1 \subseteq \Gamma_1, \dots, M_{i-1} \subseteq \Gamma_{i-1}$ such that $\bigcup_{j=1}^{i-1} M_j$ is C_4-free. We will find a matching $M_i \subseteq \Gamma_i$ such that $\bigcup_{j=1}^{i} M_j$ is C_4-free.

Recall that $I = \left[(1+\epsilon)\log_\rho n,\ (1 + (1+2\gamma)\epsilon)\log_\rho n\right]$. Set

$$W = \{x \subseteq A_1 : |x| \in I\}$$

Denote by G_W the graph with vertex set W and the set of edges defined as follows. For every $x \neq y \in W$,

$$\{x, y\} \in E(G_W) \Leftrightarrow |x \cap y| \geq (1 + (1 - 6\gamma)\epsilon)\log_\rho n \tag{7}$$

Define $\phi : V(\Gamma) \to V(G_W)$ such that $\phi(v) = N_G(v|A_1)$ for every $v \in V(\Gamma)$. Note that this definition is valid since if $v \in V(\Gamma)$, then v is A_1-good and thus $N_G(v|A_1) \in W$.

Claim 4.7. **Whp** ϕ is injective.

Proof. Fix $u \in B$. If $u \in V(\Gamma)$, then for every vertex $v \in V(\Gamma)$, we have $\phi(u) = \phi(v)$ if and only if $N_G(v|A_1) = N_G(u|A_1)$. For every vertex $v \in B$,

$$\mathbb{P}\big(N_G(v|A_1) = N_G(u|A_1)\big) = p^{d_G(u|A_1)}(1-p)^{a_1 - d_G(u|A_1)} = \left(\frac{1}{2}\right)^{a_1}$$
$$= \left(\frac{1}{2}\right)^{\frac{1}{\rho}(1+(1+\gamma)\epsilon)\log_2 n} = o(1/n^2)$$

Hence, by the union bound, **whp** there are no two vertices $u \neq v \in V(\Gamma)$ such that $\phi(u) = \phi(v)$. □

Set $\widetilde{G}_W = G_W[\phi(V(\Gamma))]$. Denote by $\widetilde{G}_W(p')$ the random subgraph of $\widetilde{G}_W$ obtained by retaining every edge of $\widetilde{G}_W$ independently with probability p'. By Claim 4.7, **whp** ϕ is injective. Therefore, $\widetilde{G}_W \cong \Gamma$. Recall that we want to find a matching in Γ_i. We will show that **whp** $\alpha(\Gamma_i) \leq \frac{n}{\log_\rho n}$ and later we will construct the desired matching. Since $\widetilde{G}_W \cong \Gamma$, it suffices to prove the following lemma.

Lemma 4.8. ***Whp***

$$\alpha(\widetilde{G}_W(p')) \leq \frac{n}{\log_\rho n}$$

Before proving this lemma, let us outline how we shall use the notion of ball-covering in the proof. Let us recall that x, y are not connected in $\widetilde{G}_W$ if $|x \cap y| < (1 + (1 - 6\gamma)\epsilon)\log_\rho n$. Consider a Hamming ball around $x \in W$, containing all $y \in W$ such that $|x \cap y|$ is sufficiently large. If we can find m vertices, such that the Hamming balls around them cover all the vertices of W (and the respective edges are retained in $\widetilde{G}_W(p')$), then the independence number of the graph is at most m—indeed, any set of more than m vertices must have two vertices in the same Hamming ball, and thus there must be an edge between them (here we will use the fact that for every $v \in B_1$, the number of neighbors of v in A_1 lies in I). Let us proceed with the detailed proof.

Proof. For every $x \in W$, denote by $B(x)$ the following Hamming ball around x:

$$B(x) := \{y \in W : |x \cap y| \geq (1 + (1 - \gamma)\epsilon)\log_\rho n\}$$

Set $m := \frac{n}{\log(n)}$ and $m' := \frac{n}{\log^3 n}$. Let $B' = \{y_1, \dots, y_{m'}\} \subseteq V(\Gamma)$ be an arbitrary subset of size m'.

Claim 4.9. **Whp**, for every $x \in W$, there exists a vertex $v \in B'$ such that

$$d_G(v|x) \geq (1 + (1 - \gamma)\epsilon) \log_\rho n$$

□

Proof. For every $x \in W$, set

$$B'(x) = \{v \in B' : d_G(v|x) \geq (1 + (1 - \gamma)\epsilon) \log_\rho n\}$$

Fix $x \in W$. Since $|B'| \geq \frac{n}{\log^3 n}$, by Lemma 2.2 the probability that $B'(x) = \emptyset$ is at most $\exp\left(-n^{0.5\epsilon}\right)$. Note that $|W| = n^{O(1)}$. Thus, by the union bound over W, the probability that there exists a vertex $x \in W$ such that $B'(x) = \emptyset$ is at most

$$n^{O(1)} \cdot \exp\left(-n^{0.5\epsilon}\right) = o(1)$$

□

Let $\mathcal{A}$ be the event that for every $x \in W$, there exists a vertex $v \in B'$ such that $d_G(v|x) \geq (1 + (1 - \gamma)\epsilon) \log_\rho n$. By Claim 4.9, we have that $\mathbb{P}(\mathcal{A}) = 1 - o(1)$. In the following claim, we assume that $\mathcal{A}$ holds deterministically.

Claim 4.10. For every $J = \{x_1, \dots, x_m\} \subset V(\widetilde{G}_W)$, we have $|E(\widetilde{G}_W[J])| \geq 0.25n \log_\rho n$. □

Proof. For every $i \in [m']$, set

$$Y_i := \{x \in J : N_G(y_i|A_1) \in B(x)\}$$

Note that by $\mathcal{A}$, we have that $\bigcup_{i=1}^{m'} Y_i = J$. Indeed, for every $x \in J$, there exists a vertex $y \in B'$ such that $N_G(y|A_1) \in B(x)$ and thus there exists $i \in [m']$ such that $x \in Y_i$.

We now show that $\widetilde{G}_W[Y_i]$ is a clique for every $i \in [m']$. Fix $i \in [m']$ and two different vertices $x, x' \in Y_i$. By the definition of Y_i,

$$|x \cap N_G(y_i|A_1)| = |N_G(y_i|x)| \geq (1 + (1 - \gamma)\epsilon) \log_\rho n$$

and

$$|x' \cap N_G(y_i|A_1)| = |N_G(y_i|x')| \geq (1 + (1 - \gamma)\epsilon) \log_\rho n$$

Recall that $y_i \in B_1$, so y_i is A_1-good and thus $d_G(y_i|A_1) \leq (1 + (1 + 2\gamma)\epsilon) \log_\rho n$. Then,

$$|N_G(y_i|A_1) \setminus x| \leq (3\gamma\epsilon) \log_\rho n \quad \text{and} \quad |N_G(y_i|A_1) \setminus x'| \leq (3\gamma\epsilon) \log_\rho n$$

Hence,

$$\begin{aligned} |x \cap x'| &\geq |x \cap x' \cap N_G(y_i|A_1)| \\ &\geq |N_G(y_i|A_1)| - |N_G(y_i|A_1) \setminus x| - |N_G(y_i|A_1) \setminus x'| \\ &\geq (1 + \epsilon - 6\gamma\epsilon) \log_\rho n \end{aligned}$$

where the last inequality is true since $d_G(y_i|A_1) \geq (1 + \epsilon) \log_\rho n$ because y_i is A_1-good. Therefore, by (7), $\{x, x'\} \in E(\widetilde{G}_W)$ implying that $\widetilde{G}_W[Y_i]$ is indeed a clique.

For every $i \in [m']$, set

$$Y_i' := Y_i \setminus \left(\bigcup_{j=1}^{i-1} Y_j\right)$$

Since for every $i \in [m']$ we have that $\widetilde{G}_W[Y_i]$ is a clique, $\widetilde{G}_W[Y_i'] \subseteq \widetilde{G}_W[Y_i]$ is also a clique. Thus,

$$|E(\widetilde{G}_W[J])| \geq \sum_{i=1}^{m'} \binom{|Y'_i|}{2} \geq m' \cdot \binom{\frac{m}{m'}}{2} \geq m' \cdot \left(\frac{m}{2m'}\right)^2$$
$$= \frac{m^2}{4m'} = \frac{1}{4} n \log_\rho n$$

where the second inequality is true by Jensen's inequality and the fact that $m = |J| = \sum_{i=1}^{m'} |Y'_i|$. □

By Claim 4.10,

$$\mathbb{P}_{p'}(|E(\widetilde{G}_W(p')[J])| = 0 | \mathcal{A}) = (1-p')^{|E(\widetilde{G}_W[J])|} \leq (1-p')^{0.25 n \log_\rho n}$$

We have,

$$\binom{|W|}{m} \leq |W|^m = \exp(\Theta(m \log n)) = \exp(\Theta(n))$$

Therefore, by the union bound, the probability that there is an independent set in $\widetilde{G}_W(p')$ of size m is at most

$$\mathbb{P}(\neg\mathcal{A}) + \binom{|W|}{m}(1-p')^{0.25 n \log_\rho n} = o(1)$$

□

Recall that our goal is to find a matching $M_i \subseteq \Gamma_i \setminus (\cup_{j=1}^{i-1} M_j)$ such that $\cup_{j=1}^{i} M_j$ is C_4-free. By Claim 4.8 and the fact that $\widetilde{G}_W \cong \Gamma$, **whp**

$$\alpha(\Gamma_i) \leq \frac{n}{\log_\rho n}$$

Let $M_i \subseteq \Gamma_i \setminus (\cup_{j=1}^{i-1} M_j)$ be a matching of the maximum size such that $\bigcup_{j=1}^{i} M_j$ is C_4-free.

Let $U \subseteq V(\Gamma_i)$ be the set of unmatched vertices. The next claim bounds the maximum degree $\Delta(\Gamma_i[U])$.

Claim 4.11. $\Delta(\Gamma_i[U]) \leq s_2^3$.

Proof. Suppose towards contradiction that $\Delta(\Gamma_i[U]) > s_2^3$ and take a vertex $v \in U$ such that $d_{\Gamma_i[U]}(v) > s_2^3$.

Note that there are at most s_2^3 many paths with three edges in $\bigcup_{j=1}^{i} M_j$ which start with the vertex v. Thus, there exists a vertex $u \in N_{\Gamma_i[U]}(v)$ such that u is not adjacent to another endpoint of the above paths and thus $\{u, v\}$ does not close a copy of C_4 in $\bigcup_{j=1}^{i} M_j$. Hence, we can add $\{u, v\}$ to the matching M_i, a contradiction to the maximality of M_i. □

We finish with the following claim.

Claim 4.12. **Whp** $|U| \leq (s_2^3 + 1)\frac{n}{\log_\rho n}$.

Proof. Suppose towards contradiction that $|U| > (s_2^3 + 1)\frac{n}{\log_\rho n}$. By Claim 4.11, $\Delta(\Gamma_i[U]) \leq s_2^3$. Hence,

$$\alpha\left(\Gamma_i[U]\right) \geq \frac{|U|}{\Delta(\Gamma_i[U]) + 1} > \frac{n}{\log_\rho n}$$

a contradiction to Lemma 4.8. □

The desired subgraph in Lemma 4.1 is the subgraph H_{B_1} with its edges from $\cup_{i=1}^{s_2-1} M_i$—indeed, note that by Claim 4.12, there are at most $O\left(\frac{n}{\log n}\right)$ unmatched vertices at each round, and thus at most $O\left(\frac{n}{\log n}\right)$ vertices of degree at most $s_2 - 1$.

4.2.2 | Proof of Lemma 4.1, $p > 0.5$

Let us now explain how to complete the proof for $p > 0.5$. Indeed, note that Claim 4.7 is not necessarily true, and hence we require a more delicate treatment to overcome the lack of isomorphism. Recall that our main goal is to show that $\alpha(\Gamma_i) \leq \frac{n}{\log_p n}$. We prove this with the following series of relatively short claims and lemmas, where the key idea is that one can consider equivalence classes under ϕ, and show that **whp** either all such classes are of bounded order, or polynomial order (Claim 4.13).

For every $v, u \in V(\Gamma)$, we say that $v \sim u$ if and only if $\phi(v) = \phi(u)$. Let $C_1, \dots, C_\ell$ be the equivalence classes under this relation. With a slight abuse of notation, given $C_j = \{v_1, \dots, v_k\}$ we write $\phi(C_j) := \phi(v_1)$.

Claim 4.13. **Whp** one of the following holds.

1. There exists a constant C such that
$$|C_j| \leq C, \quad \forall j \in [\ell]$$
2. There exists $\beta > 0$ such that
$$|C_j| \geq n^\beta, \quad \forall j \in [\ell]$$

Proof. For every $v \in [n] \setminus A_1$, set $X_v = 0$ if v is A_1-bad and X_v to be the number of vertices $u \sim v$ otherwise. Fix $v \in [n] \setminus A_1$ and consider the random variable $\epsilon' = \epsilon'(v)$ such that $|N_{G(n,p)}(v) \cap A_1| = (1 + \epsilon') \log_p n$. Note that if v is A_1-good, then $\epsilon' \in [\epsilon, (1 + 2\gamma)\epsilon]$. We have

$$\begin{aligned}
\mathbb{E}[X_v | N_{G(n,p)}(v) \cap A_1 \text{ and } v \text{ is } A_1 - \text{good}] &= (1 + o(1))np^{|N_{G(n,p)}(v)\cap A_1|}(1 - p)^{|A_1|-|N_{G(n,p)}(v)\cap A_1|} \\
&= (1 + o(1))np^{(1+\epsilon')\log_p n}(1 - p)^{\frac{1+(1+\gamma)\epsilon}{p}\log_p n-(1+\epsilon')\log_p n} \\
&= (1 + o(1))n(1 - p)^{(\log_{1-p} p)(1+\epsilon')\log_p n}(1 - p)^{\frac{1+(1+\gamma)\epsilon}{p}\log_p n-(1+\epsilon')\log_p n} \\
&= (1 + o(1))n^{1-(1+\epsilon')(\log_{1-p} p)-\frac{1+(1+\gamma)\epsilon}{p}+1+\epsilon'} \\
&= (1 + o(1))n^{2-\log_{1-p} p-\frac{1}{p}+\epsilon'-\epsilon'\log_{1-p} p-\frac{(1+\gamma)\epsilon}{p}}
\end{aligned}$$

Set $f(x) := 2 - \log_{1-x}(x) - \frac{1}{x}$. Let x' be in $(0, 1)$ such that $f(x') = 0$, noting that $f(x)$ is increasing in $x \in (0, 1)$ and $f(x) = 0$ around $x \approx 0.64$. We then have that for ϵ small enough and for some constant $c > 0$, $\epsilon' - \epsilon' \log_{1-x'} x' - \frac{(1+\gamma)\epsilon}{x'} < -c\epsilon$. Thus, if $p \leq x'$, we have that $\mathbb{E}[X_v \mid v \text{is} A_1 - \text{good}] \leq n^{-c\epsilon}$ for every $v \in [n] \setminus A_1$, where we stress that ϵ can depend on p. By Lemma 2.1 (2),

$$\mathbb{P}(X_v > C \mid v \text{ is } A_1 - \text{good}) \leq \left(\frac{e}{\frac{C}{\mathbb{E}[X_v]}}\right)^C \leq n^{-C\cdot c\epsilon} = o(1/n)$$

Hence, by the union bound over all $v \in [n] \setminus A_1$, **whp** $X_v \leq C$ for every $v \in V(\Gamma)$, and thus the first item of the claim holds.

If $p > x'$, then we may choose ϵ small enough such that $\mathbb{E}[X_v | v \text{ is } A_1 - \text{good}] \geq n^\beta$, for some $\beta > 0$, for every $v \in [n] \setminus A_1$. By Lemma 2.1 (1),

$$\mathbb{P}(X_v \leq 0.5\mathbb{E}\left[X_v | v \text{ is } A_1 - \text{good}\right] | v \text{ is } A_1 - \text{good}) \leq \exp\left(-\Theta(n^\beta)\right)$$

Hence, by the union bound over all $v \in V(\Gamma)$, **whp** $X_v \geq 0.5n^\beta$ for every $v \in [n] \setminus A_1$ which is A_1-good, and thus the second item of the claim holds as well. □

We complete the proof with the following two lemmas.

Lemma 4.14. *If there exists a constant C such that*

$$|C_j| \leq C, \quad \forall j \in [\ell]$$

then **whp** $\alpha(\Gamma_i) \leq \frac{n}{\log_p n}$.

Proof. Set $\widetilde{G}_W$. Consider the following coupling $(\widetilde{G}_W(p'), \Gamma_{p'})$. $\Gamma_{p'}$ is obtained by retaining each edge independently in Γ with probability p'. For every $v, u \in V(\widetilde{G}_W)$, there is an edge $\{u, v\}$ in $\widetilde{G}_W(p')$ if and only if in $\Gamma_{p'}$ there are all the edges between $\phi^{-1}(v)$ and $\phi^{-1}(u)$. Note that

$$\frac{\alpha(\Gamma_{p'})}{C} \leq \alpha(\widetilde{G}_W(p'))$$

Indeed, let I be a maximum independent set in $\Gamma_{p'}$. Assume that I has vertices from m different equivalence classes. Observe that $m \geq \frac{|I|}{C}$ as otherwise, by the pigeonhole principle, we have an equivalence class larger than C. Notice that $\phi(I)$ is also an independent set in $\widetilde{G}_W(p')$ since, for every $v, u \in I$, there is at least one edge missing in $\Gamma_{p'}$ between the equivalence classes of v and u, and thus there is no edge between $\phi(v)$ and $\phi(u)$ in $\widetilde{G}_W(p')$. Hence, $\phi(I) \leq \alpha(\widetilde{G}_W(p'))$ and thus,

$$\frac{\alpha(\Gamma_{p'})}{C} \leq m = |\phi(I)| \leq \alpha(\widetilde{G}_W(p'))$$

By the coupling above, every edge in $\widetilde{G}_W(p')$ appears with probability at least $(p')^{C^2}$. Thus, it suffices to show that **whp** the independence number of the binomial random subgraph of $\widetilde{G}_W$ obtained by retaining every edge with probability $(p')^{C^2}$ is at most $\frac{n}{\log_\rho n}$. The rest of the proof is identical to the proof of Claim 4.8. □

Lemma 4.15. *If there exists $\beta > 0$ such that*

$$|C_j| \geq n^\beta, \quad \forall j \in [\ell]$$

then **whp** $\alpha(\Gamma_i) \leq \frac{n}{\log_\rho n}$.

Proof. Let I be a maximum independent set in Γ_i. For every $j \in [\ell]$, set $I_j := I \cap C_j$.

Notice that, for every $j \in [\ell]$, C_j is a clique in Γ. The probability that there is an independent set in $G(n, p)$ of size $n^\beta / \log_\rho n$ is at most

$$\binom{n}{\frac{n^\beta}{\log_\rho n}}(1-p)^{\binom{n^\beta/\log_\rho n}{2}} \leq \exp\left(2n^\beta - \frac{pn^{2\beta}}{2\ln^2 n}\right)$$

There are at most $n^{1-\beta}$ many C_js, and thus, by the union bound, **whp** $I_j \leq \frac{n^\beta}{\log_\rho n}$, for every $j \in [\ell]$. Hence,

$$|I| \leq n^{1-\beta} \cdot \frac{n^\beta}{\log_\rho n} = \frac{n}{\log_\rho n}$$

□

4.3 | Theorem 4

Let us first recall that since F satisfies property $(\star)$, we may find an edge $\{u, v\}$ such that for every independent set $I \subseteq V(F)$, $F[V \setminus I]$ is non-$F[V \setminus \{u, v\}]$-degenerate. Let $\mathcal{I} := \{I \subseteq V : I \text{ is an independent set of } F\}$, and let $\mathcal{F} := \{F[V \setminus I] : I \in \mathcal{I}\}$. By Lemma 1.6, **whp** there exists a graph which is $\mathcal{F}$-free, and is $n^{-\delta}$-dense with respect to $F[V \setminus \{u, v\}]$.

The proof of Theorem 4 follows then from the construction given in the proof of Theorem 3, where B is taken to be the empty graph (similarly to [5]), and in A_1 we take H_{A_1} to be the graph guaranteed by Lemma 1.6, as detailed in the above paragraph. Again, there can be a small, yet non-negligible amount of edges in $G[B]$ such that their endpoints have small common degree in A_1. For these edges, we define B_2 to be the set of vertices which are not A_1-good in B, and we use A_2 and A_3 in the same manner. We note that Lemma 4.1 is not relevant here.

5 | Discussion and Open Problems

In this paper, we present Conjecture 1.3 and make progress towards resolving it, in particular obtaining a universal bound of $O(n \ln n)$ for the saturation number $sat(G(n, p), F)$ for all F (Theorem 1), and characterizing a family of graphs for which the bound is linear in n (Theorem 2). On the same matter, let us also iterate the question raised in the introduction:

Question 5.1. Is it true that, for all constant $0 < p < 1$ and all graphs F where every edge of F is in a triangle, **whp** $sat(G(n,p), F) = (1 + o(1))n \log_{\frac{1}{1-p}} n$?

For specific graph families, we answer positively. In particular, we extended the sharp asymptotics results of [5], both to a wide family of graphs (Theorem 4), and to complete multipartite graphs when $p \geq \frac{1}{2}$ (Theorem 3). However, our proof of Theorem 3 requires our graph to be dense enough, so that we may find a subgraph A of an appropriate size and such that the neighborhoods of the vertices outside this subgraph form a dense enough subset in a Hamming space with the domain A. Unfortunately, this no longer holds when p is a small constant (with the technicalities explicit in Lemma 2.2). It would be interesting to know whether a similar construction, with different probabilistic or combinatorial tools, could extend the result for $p < \frac{1}{2}$.

Finally, let us reiterate that it was shown in Reference [9] that Lemma 1.6 is, in fact, tight, in the sense that the conditions for it are both sufficient and necessary. This implies that the results that can be obtained by the construction given in Reference [5], where B is taken as an empty graph, are limited to those of Theorem 4. Indeed, already trying to extend this result to complete multipartite graphs, which may not adhere to the conditions of Lemma 1.6, required a delicate (and at times technical) treatment, utilizing a coupling with an auxiliary random graph in a Hamming space and proving a tight bound for its independence number using a covering-balls argument.

Acknowledgments

The authors wish to thank the anonymous referees for their helpful comments, and in particular for their corrections in Section 4.2.2. The authors further wish to thank Michael Krivelevich for helpful comments and suggestions, and Dor Elboim for fruitful discussions.

Data Availability Statement

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Endnotes

[1] With high probability, that is, with probability tending to 1 as n tends to infinity.

References

1. A. A. Zykov, "On some properties of linear complexes," *Math. Sbornik Novaia Seriia* 24, no. 66 (1949): 163–188.

2. P. Erdős, A. Hajnal, and J. W. Moon, "A problem in graph theory," *Am. Math. Mon* 71 (1964): 1107–1110.

3. L. Kászonyi and Z. Tuza, "Saturated graphs with minimal number of edges," *J. Graph Theory* 10, no. 2 (1986): 203–210.

4. J. R. Faudree, R. J. Faudree, and J. R. Schmitt, "A survey of minimum saturated graphs," *Electron. J. Comb* DS19(Dynamic Surveys) 36 (2011).

5. D. Korándi and B. Sudakov, "Saturation in random graphs," *Random Struct. Algoritm* 51, no. 1 (2017): 169–181.

6. A. Mohammadian and B. Tayfeh-Rezaie, "Star saturation number of random graphs," *Discret. Math* 341, no. 4 (2018): 1166–1170.

7. S. Demyanov and M. Zhukovskii, "Tight concentration of star saturation number in random graphs," *Discret. Math* 346, no. 10 (2023): 113572.

8. Y. Demidovich, A. Skorkin, and M. Zhukovskii, "Cycle saturation in random graphs," *SIAM J. Discret. Math* 37, no. 3 (2023): 1359–1385.

9. S. Diskin, I. Hoshen, M. Krivelevich, and M. Zhukovskii, "On vertex Ramsey graphs with forbidden subgraphs," *Discret. Math* 347, no. 3 (2024): 113806.

10. S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization (Wiley-Interscience, 2000).

11. N. Alon and J. H. Spencer, *The probabilistic method*, 4th ed. (John Wiley & Sons, 2016).

12. A. Frieze and M. Karoński, *Introduction to random graphs* (Cambridge University Press, 2016).

13. A. Johansson, J. Kahn, and V. Vu, "Factors in random graphs," *Random Struct. Algoritm* 33, no. 1 (2008): 1–28.