Journal of Combinatorics
Volume 0, Number 0, 1, 2019

The Chromatic Number of Squares Of Random Graphs*

Kalyan Garapaty[†]
Daniel Lokshtanov[†]
Hemanta K. Maji[‡]
Alex Pothen[§]

MSC subject codes: 68R10, 05C15

The Erdös-Rényi model is a simple and widely studied model for generating random graphs. Given a positive integer n and a real p between 0 and 1, $G(n,p)$ is the distribution over n-vertex graphs obtained by including, for every unordered pair $\{u,v\}$ of vertices, the edge uv in the edge set of G independently with probability p. The *square* of a graph G, denoted by G^2, is the graph obtained from G by also adding an edge between every pair of vertices that share at least one common neighbor. A proper k-coloring of a graph G is a function f that assigns to every vertex of G a color $f(v)$ from the set $\{1,\ldots,k\}$ such that no two neighbouring vertices get the same color, and the *chromatic number* of a graph G is the minimum k so that G has a k-coloring.

In a recent article, Cheng, Maji and Pothen [3] consider squares of sparse Erdős-Rényi graphs $G(n,p)$ with $p=\Theta(1/n)$ as interesting benchmark instances to evaluate parallel algorithms that color the input graph. These authors prove that if G is sampled from $G(n,p)$ with $p=\Theta(1/n)$ then, with high probability, the chromatic number of G^2 lies between $\Omega\left(\frac{\log n}{\log\log n}\right)$ and $\mathcal{O}(\log n)$. In this work we obtain a tight $\Theta\left(\frac{\log n}{\log\log n}\right)$ bound on the chromatic number of G^2. Along the way we also obtain asymptotically tight bounds for the maximum degree of the k-th power of graphs sampled from $G(n,p)$.

*Squares of Random Graphs

[†]The work of Lokshtanov and Garapaty is supported by BSF award 2018302 and NSF award CCF-2008838.

[‡]The work of Maji is supported in part by NSF awards CNS–1618822 and CNS–2055605.

[§]The work of Pothen is supported by NSF award CCF-1637534 and DOE award SC-0022260.

1. Introduction

We consider the problem of characterizing the chromatic number of squares and higher powers of random graphs. For a random bipartite graph on n vertices, choosing each edge with probability $\Theta(1/n)$, Cheng, Maji, and Pothen [3] had established a lower bound of $\Omega(\log n/\log\log n)$ for the chromatic number of the square of the graph induced by one vertex set. (Such a graph is called a random binomial intersection graph, and more details will be provided in the next section.) In this paper, we show a matching upper bound on the chromatic number. Indeed, we obtain for the k-th power of an Erdös-Rényi (non-bipartite) graph, an upper bound on the chromatic number of $\mathcal{O}\Big(\log n/\log_{(k)} n\Big)$, where the denominator involves the k times nested logarithmic function. We believe our techniques for obtaining the upper bound are of independent interest, since we show that the maximum degree of vertex in the k-th power graph is $\mathcal{O}\Big(\log n/\log_{(k+1)} n\Big)$.

This problem is motivated by earlier work on generating graphs of arbitrary size whose chromatic numbers are tunable and precisely known [3]. Achlioptas and Naor [1] and later Coja-Oghlan and Vilenchik [4] showed that the chromatic number of sparse Erdös-Rényi graphs could be precisely specified when the probability of an edge is $\Theta(1/n)$, and that it was independent of the number of vertices n. The coloring problems occurring in the efficient computation of Jacobian matrices for continuous optimization or solution of differential equations are related to the (partial) distance-2 coloring problem in bipartite graphs [6], and these correspond to the usual distance-1 coloring on the square of the bipartite graph induced by one vertex set. The several coloring problems that occur in the computation of Jacobians and Hessians are discussed by Gebremedhin, Manne and Pothen [6].

Note that unlike the distance-1 chromatic number, the distance-2 chromatic number of random Erdös-Rényi graphs increases with n when the probability of an edge is $\Theta(1/n)$. Knowing how it increases with n is practically helpful to assess the performance of coloring algorithms when test graphs with billions or more edges are generated and colored on massively parallel distributed-memory multiprocessors.

In this paper we will consider the powers of graphs sampled from the Erdös-Rényi model of random graphs for specific parameter regimes. We will begin by showing bounds on the number of vertices at a given distance from any vertex in a graph sampled according to the model. Using these bounds, we will give a tight bound on the chromatic number. We will then show that these results apply to the class of binomial random intersection graphs.

2. Preliminaries

Let X be a random variable over the sample space Ω. Consider any subset $\mathcal{E} \subseteq \Omega$, referred to as an *event*. The random variable $\mathbb{1}(\mathcal{E})$ over the sample space $\{0,1\}$ is the random variable "$X \in \mathcal{E}$." In particular, we have $\mathrm{E}[\mathbb{1}(\mathcal{E})] = \Pr[X \in \mathcal{E}]$.

The Erdös-Rényi model is denoted as $G(n,p)$, where n and p are parameters to the model. A graph of this model is sampled as follows - a set of n vertices is taken as the vertex set V; an edge is added between any pair of vertices in V independently with a probability p; and the resulting graph is output as the sample. In this paper, we will be looking at the graphs sampled with $p = \frac{\gamma}{n}$, where γ is a positive constant.

The k-th power of a graph G is another graph G^k, defined on the same set of vertices as G but it has an edge between any two vertices that are at a distance of at most k in G. The chromatic number of a graph is the minimum number of colors needed to be able to give each vertex of the graph a color such that no two vertices connected by a single edge have the same color.

The distribution of Binomial Random Intersection Graphs is represented by $G(n,m,p)$. A graph of this model is sampled as follows. Two sets of vertices V and W of sizes n and m respectively, are taken. The set V is interpreted as a set of vertices, and W as a set of features, in a bipartite graph H. Each of the $n \cdot m$ edges between the sets V and W is added to the graph independently with a probability p. The graph thus formed is called a *Bipartite Vertex-feature Inclusion Graph*, labelled as H. Now, the vertex set V of H is squared by adding an edge between each pair of vertices V if they have at least one common neighbor in W. The resulting graph over the set V, ignoring the vertices in W, is called a *binomial Random Intersection Graph*. In this paper we will be looking at the parameter regimes of $n = m$ and $p = \frac{\gamma}{n}$, where γ is a positive constant.

A proper k-coloring of a graph G is a function f that assigns to every vertex of G a color $f(v)$ from the set $\{1, \ldots, k\}$ such that no two neighboring vertices get the same color. The graph G is said to be d-degenerate if there exists an ordering of the vertices of G such that every vertex in the ordering has at most d forward edges, i.e., edges that join the vertex to higher numbered vertices. The minimum d such that G is d-degenerate is known as the *degeneracy* of G.

Let G be any graph with vertex set V and set of edges $E(G)$. A subset of vertices S of a graph G is called *sufficiently dense* if the induced subgraph of S in G has at least as many edges as there are vertices in S. For any vertex $v \in V$, we call the set of vertices at a distance exactly k from v the distance-k neighborhood of v. The distance between any two vertices is the number of edges in a shortest path that joins them. Since we are dealing with random graphs in this paper, note that for such a graph any of its properties, including these neighborhoods would be randomized. We will thus represent them by random variables defined as follows.

The set $\mathcal{N}_{v,k}$ will denote a random variable denoting the set of distance-k neighbors of v. The size of $\mathcal{N}_{v,k}$ will be represented by the random variable $N_{v,k}$. If $X_1, X_2, \cdots X_t$ are some fixed, mutually disjoint subsets of the vertex set V, we will represent them as a vector such as - $\vec{X_t} = (X_1, X_2, \cdots X_t)$ and define $\cup\vec{X_k} = \cup_{1\leqslant t\leqslant k} X_t$. Similarly, we will have a random variable $\vec{\mathcal{N}_{v,t}}$, which is a vector of size t with the entries denoting the values of the random variables $\mathcal{N}_{v,1}, \mathcal{N}_{v,2} \cdots \mathcal{N}_{v,t}$. Now, the event $\mathcal{N}_{v,1} = X_1$, $\mathcal{N}_{v,2} = X_2$, $\cdots \mathcal{N}_{v,t} = X_t$ will be represented as $\vec{\mathcal{N}_{v,t}} = \vec{X_t}$.

For any of the variables defined here, if the graph that they are describing is clear from the context, we write them as above. If we are talking about different graphs, such as G_1 and G_2, we add a superscript to the random variables defined above as - $\mathcal{N}_{v,t}^{G_1}$ or $\mathcal{N}_{v,t}^{G_2}$.

For any graph G, we have the following parameters:

- $\Delta(G)$ - maximum degree among its vertices
- $\chi(G)$ - its chromatic number
- $\mathcal{D}(G)$ - its degeneracy

If the graph being considered is clear from the context, we drop G from this notation.

Relation with Degeneracy ($\mathcal{D}$) Let $v_{o_1}, v_{o_2} \cdots v_{o_n}$ be an ordering of the vertices of G such that every vertex v_{o_i} is a neighbor to at most $\mathcal{D}$ of the vertices $v_{o_{i+1}}, v_{o_{i+2}} \cdots v_{o_n}$. Such an ordering does exist due to the definition of degeneracy. Now we take $\mathcal{D}+1$ colors to color the vertices $v_{o_n}, v_{o_{n-1}} \cdots v_{o_1}$ in that order. When v_{o_i} is to be colored, due to the order of coloring, only $v_{o_{i+1}}, v_{o_{i+2}} \cdots v_{o_n}$ are coloured. Thus, at most $\mathcal{D}$ neighbors of v_{o_i} are colored. As we have $\mathcal{D}+1$ colors, there will be at least one color that was not assigned to any of the neighbors of v_{o_i}. This color will be given to v_{o_i}. In this way all the vertices of G can be colored in $\mathcal{D}+1$ colors.

$$\chi \leqslant \mathcal{D} + 1.$$

For any real x and natural number k, we denote $\log_{(k)}(x) = \underbrace{\log\log\cdots\log}_{k\text{-times}}(x)$ and $x_k = \frac{\log x}{\log_{(k)}(x)}$. We use the well-known Big Oh, Big Omega, Big Theta, and small Oh asymptotic notations throughout; Θ_t. will denote that the constants involved in the relation depend on a constant parameter t.

As we are working with random graphs, the meaning of a bound on the value of any variable is different from the deterministic case. Here a bound means that the probability of the random variable being over (for upper bound) or below (for lower bound) the bound is negligible in terms of the input parameters.

Related Work. Now we summarize related work to put our contributions in context.

Chromatic number. The chromatic number of the random graph $G(n, d/n)$ can be specified for "almost all" values of d. Let $d_{k,1} = 2k \log k - \log k - 1 + o_k(1)$, where the last term goes to zero with increasing values of k, and let $d_{k,2} = 2k \log k$. Achlioptas and Naor [1] proved that in the interval $d \in \left(d_{(k-1),1}, d_{(k-1),2}\right)$, the chromatic number of $G(n, d/n)$ is k for large enough values of n with high probability. In the subsequent interval $d \in \left(d_{(k-1),2}, d_{k,1}\right)$, it is either k or $k+1$. This result has been strengthened by Coja-Oghlan and Vilenchik [4] in the following theorem.

Theorem 2.1. *There exists a constant k_0 such that the following statement is true. Let $S_k = (2(k-1)\log(k-1) - \log(k-1) - 0.99, 2k \log k - \log k - 1.38)$, $S = \bigcup_{k \geqslant k_0} S_k$, and $F(d) = k$ for all $d \in S_k$. Then S has asymptotic density 1 and for any $d \in S$, we have*

$$\lim_{n \to \infty} \Pr[\chi(G(n, d/n)) = F(d)] = 1.$$

Intersection graphs. A study of various properties of random intersection graphs is included in the book [5].

Lagerås and Lindholm [7] have studied the component structure of the binomial random intersection graph $G(n, m, p)$ with $m = \lfloor \beta/n \rfloor$ and $p = \gamma/n$, where β and γ are constants. The expected degree of a vertex is then $\mu = \beta\gamma^2$. If $\mu < 1$, then with high probability there is no connected component in G with more than $\mathcal{O}(\log n)$ vertices; if $\mu > 1$, then with high probability there exists a giant component with $(1 - \rho + o(1))n$ vertices, where $0 < \rho < 1$, and the size of the second largest component is $\mathcal{O}(\log n)$.

The chromatic number of binomial random intersection graphs has been studied by Behrisch, Taraz and Ueckerdt [2]. One of their theorems of interest in this context is the following:

Theorem 2.2. *Let $m = n^\alpha$, with $\alpha > 0$, and let $p = o(\sqrt{1/mn})$. Then with high probability the graph $G(n, m, p)$ can be colored optimally in linear time* *and $\chi(G(n,m,p)) = \omega(G(n,m,p))$, i.e., the chromatic number is equal to the size of the maximum clique.*

For these values of p, the graph G is a perfect graph with high probability, and the authors show that Greedy algorithm colors the graph optimally. We consider the graph for larger values of p, i.e., when $p = \Theta(1/n)$, due to the optimization problem that motivates this work.

3. Asymptotically Tight Degree Bounds for Powers Of Sparse Random Graphs

Let t be a natural number and $\gamma > 0$ be a positive real number. If we sample a random graph G on n vertices, including each edge independently with probability γ/n, what can we say about the maximum degree of G^t? We show that for sufficiently large n (depending on γ and t), with high probability it holds that

$$\Delta(G^t) = \Theta_t\left(\frac{\log n}{\log_{(t+1)} n}\right). \tag{1}$$

We will prove a stronger statement that also gives a bound of $\frac{1}{n^{0.7}}$ on the failure probability (i.e., the probability that a graph G drawn from $G(n,\gamma/n)$ fails to satisfy Equation 1) and gives bounds for the dependence of the constants hidden in the Θ_t-notation in Equation 1.

Theorem 3.1. *Define constants $\{\varepsilon_i := 0.05 \cdot 2^{-i}\}_{i \geqslant 0}$ and $\delta := 6$. Fix any natural number t and a positive constant γ. For sufficiently large $n \in \mathbb{N}$, the graph $G \sim G(n,\gamma/n)$ satisfies*

$$\Pr\left[\frac{\Delta\left(G^t\right)}{n_{t+1}} \in [\varepsilon_t, \delta]\right] \geqslant 1 - \frac{1}{n^{0.7}}.$$

We emphasize that the condition "n should be sufficiently large" depends on both the parameters t and γ. For example, the quantity $\log_{(t+1)} n$ requires n to be at least a height-t tower of exponentiations to be well-behaved. Furthermore, we clarify that here the subscript i is an index in the notation ε_i, unlike the definition of n_{t+1}.

The proof of Theorem 3.1 spans the remainder of this section. The statement (and proof) of Theorem 3.1 naturally splits in two parts: an *upper* bound and a *lower* bound for the maximum degree of the t'th power of a random graph. Lemma 3.2 shows the upper bound, while Lemma 3.3 and Lemma 3.4 imply the lower bound.

High-level proof overview.

1. For the upper bound, Lemma 3.2 proves that the number of vertices at distance $k \in \{0, 1, \ldots, t\}$ is $3n_{k+1}$ with $1 - 1/\mathsf{poly}(n)$ probability using the first moment technique. In particular, this result implies that the probability of any vertex having $\geqslant 6n_{t+1}$ degree in G^t is at most $1/\mathsf{poly}(n)$.
2. For the lower bound, we proceed using the second moment technique. Lemma 3.3 proves that the expected number of vertices with at least $\Theta(2^{-k}n_{k+1})$ vertices at distance k, where $k \in \{0, 1, \ldots, t\}$, is $n^{1-\Theta(1)}$. Next, Lemma 3.4 shows that family of events "a vertex v has at least $\Theta(2^{-k}n_{k+1})$ vertices at distance k" is (essentially) pairwise independent, for $v \in V(G)$; entailing a small covariance. Using the second moment technique, these two observations imply that there must be a vertex with $\Theta(2^{-t}n_{t+1})$ vertices at distance t, except with $1/\mathsf{poly}(n)$ probability.

Section 4 proves that the chromatic number of G^t is $\mathcal{O}(n_t)$ (refer to Theorem 4.1). This result relies only on Lemma 3.2. Readers interested in Theorem 4.1 can skip the details of Lemma 3.3 and Lemma 3.4, which pertain to the lower bound on $\Delta(G^t)$.

3.1. Degree Upper Bound

Suppose $G \sim G(n, p)$, where $p = \gamma/n$. Recall that $N_{v,k}$ represents the number of vertices in G at distance k from v.

Lemma 3.2. *Fix a constant $\gamma > 0$. There exists an infinite sequence of natural numbers $B(0, \gamma) \leqslant B(1, \gamma) \leqslant B(2, \gamma) \leqslant \cdots$ such that the following statement holds. For all integers $k \geqslant 0$ and $n \geqslant B(k, \gamma)$, the probability estimate below holds for any vertex v in the graph $G \sim G(n, \gamma/n)$.*

$$\Pr[N_{v,k} > 3 \cdot n_{k+1}] \leqslant \frac{k}{n^2}.$$

Proving the upper bound of Theorem 3.1 using Lemma 3.2. Fix $t \in \mathbb{N}$ and constant $\gamma > 0$. Consider a natural number $n \geqslant B(t, \gamma)$. For any vertex v in $G \sim G(n, \gamma/n)$ and $k \in \{0, 1, \ldots, t\}$, since $n \geqslant B(t, \gamma) \geqslant B(k, \gamma)$, Lemma 3.2 implies

$$\Pr[N_{v,k} > 3 \cdot n_{k+1}] \leqslant \frac{k}{n^2}.$$

By union bound over $k \in \{1, 2, \ldots, t\}$, we have

$$\Pr\left[\bigvee_{k=1}^{t} N_{v,k} > 3 \cdot n_{k+1}\right] \leqslant \frac{1 + 2 + \cdots + t}{n^2} \leqslant \frac{t^2}{n^2}.$$

Finally, using a union bound over $v \in V(G)$, we get

$$\Pr\left[\exists v \in V(G) \text{ s.t. } \bigvee_{k=1}^{t} N_{v,k} > 3 \cdot n_{k+1}\right] \leqslant \frac{t^2}{n}.$$

So, with probability at least $(1 - t^2/n)$, every vertex $v \in V(G)$ satisfies $N_{v,k} \leqslant 3 \cdot n_{k+1}$, for every $k \in \{1, 2, \ldots, t\}$. Therefore, any vertex $v \in V(G)$ in such a graph has degree at most $3 \cdot (n_2 + n_3 + \cdots + n_{t+1}) \leqslant 6 \cdot n_{t+1}$ in the graph G^t (by Lemma A.1). So,

$$\Pr[\Delta(G^t) \leqslant 6 \cdot n_{t+1}] \geqslant 1 - \frac{t^2}{n} = 1 - \mathrm{o}\left(\frac{1}{n^{0.7}}\right). \tag{2}$$

Proof of Lemma 3.2 . We proceed by induction on $k \in \{0, 1, \ldots\}$ and also show the existence of an infinite sequence of natural numbers $B(0, \gamma) \leqslant B(1, \gamma) \leqslant \cdots$.

For the base case of $k = 0$, define $B(0, \gamma) := 1$. For all $n \geqslant B(0, \gamma)$, $G \sim G(n, \gamma/n)$, and a vertex $v \in V(G)$, we have $N_{v,0} = 1 < 3 = 3 \cdot n_1$.

We now proceed with the inductive step for $k \in \{1, 2, \ldots\}$. Consider $n \geqslant B(k-1, \gamma)$. The value of $B(k, \gamma)$ shall be determined later. Consider $G \sim G(n, \gamma/n)$ and any vertex $v \in V(G)$. By law of total probability we obtain the following.

$$\begin{aligned}
\Pr[N_{v,k} > 3 \cdot n_{k+1}] &= \Pr[N_{v,k} > 3 \cdot n_{k+1}, N_{v,k-1} \leqslant 3 \cdot n_k] \\
&\quad + \Pr[N_{v,k} > 3 \cdot n_{k+1}, N_{v,k-1} > 3 \cdot n_k] \\
&\leqslant \Pr[N_{v,k} > 3 \cdot n_{k+1} | N_{v,k-1} \leqslant 3n_k] + \Pr[N_{v,k-1} > 3n_k] \\
&\leqslant \Pr[N_{v,k} > 3 \cdot n_{k+1} | N_{v,k-1} \leqslant 3n_k] + \frac{k-1}{n^2}. \qquad (3)
\end{aligned}$$

Here we applied the inductive hypothesis in the last transition. In light of Equation 3, to complete the proof, it suffices to show the upper bound

$$\Pr[N_{v,k} > 3 \cdot n_{k+1} | N_{v,k-1} \leqslant 3n_k] \leqslant \frac{1}{n^2}.$$

Intuitively, this result is equivalent to proving that a small vertex-set is unlikely to expand significantly. We proceed with estimating this probability expression below.

$$\begin{aligned}
&\Pr[N_{v,k} > 3 \cdot n_{k+1} | N_{v,k-1} \leqslant 3 \cdot n_k] \\
&\leqslant \max_{\substack{\vec{X}_{k-1} \text{ s.t.} \\ |X_{k-1}| \leqslant 3 \cdot n_k}} \Pr\Big[N_{v,k} > 3 \cdot n_{k+1} | \vec{\mathcal{N}}_{v,k-1} = \vec{X}_{k-1}\Big] \\
&\leqslant \max_{\substack{\vec{X}_{k-1} \text{ s.t.} \\ |X_{k-1}| \leqslant 3 \cdot n_k}} \Pr\Big[\vec{X}_{k-1} \text{ has } > 3 \cdot n_{k+1} \text{ neighbors in } V(G) \setminus \vec{X}_{k-1}\Big] \\
&\leqslant \left(\frac{\mathrm{e} n_k \gamma}{n_{k+1}}\right)^{3n_{k+1}}. \qquad (4)
\end{aligned}$$

The last bound follows from the following claim.

Claim 1. *Consider a subset $A \subseteq V(G)$. The probability of A having at least b neighbors in $V(G) \setminus A$ is at most*

$$\binom{n-a}{b} \cdot (1-q^a)^b \leqslant \binom{n}{b} \cdot \left(\frac{a\gamma}{n}\right)^b \leqslant \left(\frac{\mathrm{e} a\gamma}{b}\right)^b,$$

where $a = |A|$ and $q = (1 - \gamma/n)$.

The proof of this claim proceeds as follows. Consider a set $B \subseteq V(G) \setminus A$ of size b. A vertex in A is not adjacent to a vertex in B with probability q, and hence no vertex in A is adjacent to a vertex in B with probability q^a. Thus some vertex in A is adjacent to a vertex in B with probability $(1 - q^a)$, and it follows that the probability that the neighborhood of A contains B is at least

$$(1 - q^a)^b \leqslant (a\gamma/n)^b.$$

The simplification follows from

$$1 - q^a = 1 - (1 - \gamma/n)^a) \leqslant 1 - (1 - a\gamma/n) = a\gamma/n,$$

since $(1-x)^a \geqslant 1-ax$ for all $a \geqslant 1$ and $x \in [0, 1]$. The claim follows by taking a union bound over all possible subsets B and upper bounding the binomial coefficient $\binom{n}{b} \leqslant (\mathrm{e}n/b)^b$ (see Lemma A.2).

We continue the simplification of our probability expression as follows.

$$\Pr[N_{v,k} > 3 \cdot n_{k+1} | N_{v,k-1} \leqslant 3 \cdot n_k] \leqslant \left(\frac{\mathrm{e} n_k \gamma}{n_{k+1}}\right)^{3n_{k+1}}$$

$$= \left(\frac{\mathrm{e}\gamma \log_{(k+1)} n}{\log_{(k)} n}\right)^{3\log n/\log_{(k+1)} n}$$
$$= n^{-3+\mathrm{o}(1)} \leqslant \mathrm{o}\left(\frac{1}{n^2}\right).$$

The last step in the simplification uses the facts that

$$\left(\log_{(i)} n\right)^{\log n/\log_{(i+1)} n} = n, \text{ and } \left(\Theta(1)\cdot \log_{(i+1)} n\right)^{\log n/\log_{(i+1)} n} = n^{\mathrm{o}(1)}.$$

The natural number $B(k,\gamma)$ is chosen such that the $\mathrm{o}(1)$ function in the probability expression above becomes < 1. This completes the proof of Lemma 3.2. □

3.2. Degree Lower Bound

Lemma 3.3. *For a fixed constant $\gamma > 0$, there exists an infinite sequence of natural numbers $C(0,\gamma) \leqslant C(1,\gamma) \leqslant C(2,\gamma) \leqslant \cdots$ such that the following statement holds. For all integers $k \geqslant 0$, for $n \geqslant C(k,\gamma)$, and for any vertex v in the graph $G \sim G(n,\gamma/n)$, the number of vertices at distance k from v satisfies the following probability estimate:*

$$\Pr[N_{v,k} \geqslant \varepsilon_k \cdot n_{k+1}] \geqslant n^{-\sum_{i=0}^{k}\varepsilon_i} > \frac{1}{n^{0.1}},$$

where $\varepsilon_0 = 0.05$ and $\varepsilon_i = 2^{-i}\varepsilon_0$, for $i \geqslant 1$.

Proof. Observe that the sequence $\{\varepsilon_i\}_{i\geqslant 0} = \{0.05\cdot 2^{-i}\}_{i\geqslant 0}$ is a geometric progression. Therefore, $\sum_{i=0}^{k}\varepsilon_i < 2\varepsilon_0 = 0.1$, which implies the final estimate in the lemma.

Now, we proceed by induction on $k \in \{0,1,\dots\}$ to prove that

$$\Pr[N_{v,k} \geqslant \varepsilon_k \cdot n_{k+1}] \geqslant n^{-\sum_{i=0}^{k}\varepsilon_i},$$

and, simultaneously, show the existence of the appropriate (non-decreasing) integer sequence $\{C(i,\gamma)\}_{i\geqslant 0}$.

For the base case of $k=0$, define $C(0,\gamma)=1$. For $n \geqslant C(k,\gamma)$, it trivially holds that

$$\Pr[N_{v,0} = 1 \geqslant 0.05 = \varepsilon_0 \cdot n_1] = 1 \geqslant \frac{1}{n^{\varepsilon_0}}.$$

For the inductive step, consider $k \in \{1, 2, \dots\}$. Consider any $n \geqslant C(k-1, \gamma)$ and $G \sim G(n, \gamma/n)$. The exact constant $C(k, \gamma)$ shall be determined later.

Before we proceed, we present a high-level outline of our proof strategy. Our objective is to prove that $N_{v,k}$ is large with substantial probability. However, there are two events that can preclude this phenomenon. The first bad event is when $\sum_{i=1}^{k-1} N_{v,i}$ is "too large" and, thus, the pool of remaining vertices becomes too small. The second bad event is that the $N_{v,k-1}$ is "too small" and it is unlikely for the set of vertices at distance $(k-1)$ to sufficiently expand to have a large size-$N_{v,k}$ set of vertices at distance k. To avoid both these bad events, we define the following predicate.

$$\{\mathsf{true}, \mathsf{false}\} \ni \mathsf{Good}(v, G) := \left(\bigwedge_{i=0}^{k-1} N_{v,i} \leqslant 3n_{i+1} \right) \wedge \left(N_{v,k-1} \geqslant \varepsilon_{k-1} \cdot n_k \right).$$

We shall show that conditioned on $\mathsf{Good}(v, G)$ happening, the probability of $N_{v,k} \geqslant \varepsilon_k \cdot n_{k+1}$ is substantial. Furthermore, the probability of $\mathsf{Good}(v, G)$ is substantial as well (using the inductive hypothesis and Lemma 3.2).

We present two claims formalizing these intuitive statements.

Claim 2.

$$\Pr[\mathsf{Good}(v, G)] \geqslant \frac{1}{2} \cdot n^{-\sum_{i=0}^{k-1} \varepsilon_i}.$$

Claim 3.

$$\Pr[N_{n,k} \geqslant \varepsilon_k \cdot n_{k+1} | \mathsf{Good}(v, G)] \geqslant 2 \cdot n^{-\varepsilon_k}.$$

Using Claim 2 and Claim 3, the proof of the lemma is straightforward.

$$\begin{aligned}
\Pr[N_{n,k} \geqslant \varepsilon_k \cdot n_{k+1}] &\geqslant \Pr[N_{n,k} \geqslant \varepsilon_k \cdot n_{k+1}, \mathsf{Good}(v, G)] \\
&= \Pr[N_{n,k} \geqslant \varepsilon_k \cdot n_{k+1} | \mathsf{Good}(v, G)] \cdot \Pr[\mathsf{Good}(v, G)] \\
&\geqslant \left(2 \cdot n^{-\varepsilon_k}\right) \cdot \left(\frac{1}{2} \cdot n^{-\sum_{i=0}^{k-1} \varepsilon_i} \right) \\
&= n^{-\sum_{i=0}^{k} \varepsilon_i}.
\end{aligned}$$

This derivation completes the inductive proof of our lemma. □

Thus, all that remains is to prove Claim 2 and Claim 3.

Proof of Claim 2 . Observe that

$$\Pr[N_{v,k-1} \geqslant \varepsilon_{k-1} \cdot n_k]$$

$$
\begin{aligned}
&= \Pr\left[\left(\bigwedge_{i=0}^{k-1} N_{v,i} \leqslant 3n_{i+1}\right) \wedge (N_{v,k-1} \geqslant \varepsilon_{k-1} \cdot n_k)\right] \\
&\qquad + \Pr\left[\left(\bigvee_{i=0}^{k-1} N_{v,i} > 3n_{i+1}\right) \wedge (N_{v,k-1} \geqslant \varepsilon_{k-1} \cdot n_k)\right] \\
&= \Pr[\mathsf{Good}(v,G)] \\
&\qquad + \Pr\left[\left(\bigvee_{i=0}^{k-1} N_{v,i} > 3n_{i+1}\right) \wedge (N_{v,k-1} \geqslant \varepsilon_{k-1} \cdot n_k)\right] \\
&\leqslant \Pr[\mathsf{Good}(v,G)] + \Pr\left[\bigvee_{i=0}^{k-1} N_{v,i} > 3n_{i+1}\right]
\end{aligned}
$$

(By Lemma 3.2)

$$
\begin{aligned}
&\leqslant \Pr[\mathsf{Good}(v,G)] + \frac{0+1+\cdots+(k-1)}{n^2} \\
&< \Pr[\mathsf{Good}(v,G)] + \frac{k^2}{n^2}
\end{aligned}
$$

Rearranging this inequality, we get

$$
\begin{aligned}
\Pr[\mathsf{Good}(v,G)] &\geqslant \Pr[N_{v,k-1} \geqslant \varepsilon_{k-1} \cdot n_k] - \frac{k^2}{n^2} \\
\text{(By inductive hypothesis)} \qquad &\geqslant n^{-\sum_{i=0}^{k-1} \varepsilon_i} - \frac{k^2}{n^2} \\
&\geqslant \frac{1}{2} \cdot n^{-\sum_{i=0}^{k-1} \varepsilon_i}.
\end{aligned}
$$

We shall choose a sufficiently large $C(k,\gamma)$ such that for all $n \geqslant C(k,\gamma)$ the last inequality is satisfied. □

Proof of Claim 3 . The proof relies on a technical claim.

Claim 4. *Consider a subset $A \subseteq S \subseteq V(G)$. The probability of A having at least b neighbors in $V(G) \setminus S$ is at least*

$$
\binom{n-s}{b} \cdot (1-q^a)^b \cdot (q^a)^{n-s-b},
$$

where $s = |S|$, $a = |A|$, and $q = (1 - \gamma/n)$.

The probability above is lower-bounded by the probability of the event that A has exactly b neighbors in $V(G) \setminus S$. There are $\binom{n-s}{b}$ such sets. Fix a size-b subset $B \subseteq V(G) \setminus S$. The probability that A is neighbor to a vertex in B is $(1 - q^a)$, and the probability that A is a neighbor to every vertex in B is $(1 - q^a)^b$. The probability that A is *not* a neighbor to any vertex in $V(G) \setminus (S \cup B)$ is $q^{a\cdot(n-s-b)}$. Observe that there is no double-counting of configurations in the argument above because if A has exactly b neighbors in $V(G)\setminus S$ in a graph configuration then this configuration gets counted exactly once. Therefore, the expression in the claim is a lower bound to the desired probability.

In our case, A denotes the vertices at distance $(k - 1)$ from v in G, S denotes the vertices at distance $\leqslant (k-1)$ from v in G, and $b = \varepsilon_k \cdot n_{k+1}$ (the lower bound on the number of vertices at distance k from v in G). Since we condition on $\mathsf{Good}(v, G)$, we have $s \leqslant 3(n_1 + \cdots + n_k) \leqslant 6n_k$ (by Lemma A.1). Furthermore, we have $\varepsilon_{k-1} \cdot n_k \leqslant a \leqslant 3 \cdot n_k$.

We shall rely on the following individual estimates of the terms in the probability expression.

$$\binom{n-s}{b} \geqslant \left(\frac{n-s}{b}\right)^b = (1 - \mathrm{o}(1)) \cdot \left(\frac{n}{b}\right)^b$$

In this bound, we use the fact that $sb = \mathrm{o}(n)$ in our context.

$$\begin{aligned}(1-q^a)^b &= \left(1 - \left(1 - \frac{\gamma}{n}\right)^a\right)^b \geqslant (1 - \exp(-\gamma a/n))^b \\ &\geqslant \left(\frac{\gamma a}{2n}\right)^b.\end{aligned}$$

This bound relies on the fact that (i) $(1 - x) \leqslant \exp(-x)$, for all $x \in \mathbb{R}$, and (ii) $\exp(-x) \leqslant 1 - x/2$, for $x \in [0, 1/2]$. We also use the fact that $a = \mathrm{o}(n)$ in our context.

$$\begin{aligned}q^{a\cdot(n-s-b)} &\geqslant \left(1 - \frac{\gamma}{n}\right)^{a(n-s-b)} > \left(1 - \frac{\gamma}{n}\right)^{an} \\ &= (1 - \mathrm{o}(1)) \cdot \exp(-\gamma a).\end{aligned}$$

This inequality uses the fact that $a = \mathrm{o}(n)$ in our context. Consequently, the probability is lower-bounded as follows.

$$\binom{n-s}{b} \cdot (1 - q^a)^b \cdot q^{a(n-s-b)}$$

$$\geqslant (1 - \mathrm{o}(1)) \cdot \left(\frac{\gamma a}{2b}\right)^{b} \cdot \exp(-\gamma a)$$

$$\geqslant (1 - \mathrm{o}(1)) \cdot \left(\frac{\gamma \varepsilon_{k-1} n_k}{2\varepsilon_k n_{k+1}}\right)^{\varepsilon_k n_{k+1}} \cdot \exp\left(-3\gamma n_k\right)$$

(Because $\varepsilon_i = \varepsilon_{i-1}/2$)

$$\geqslant (1 - \mathrm{o}(1)) \cdot \left(\frac{\gamma n_k}{n_{k+1}}\right)^{\varepsilon_k n_{k+1}} \cdot \exp\left(-3\gamma n_k\right)$$

$$= (1 - \mathrm{o}(1)) \cdot \left(\frac{\gamma \exp\left(-3\gamma n_k/\varepsilon_k n_{k+1}\right) \log_{(k+1)} n}{\log_{(k)} n}\right)^{\varepsilon_k n_{k+1}}$$

(Because $n_k = \mathrm{o}(n_{k+1})$)

$$\geqslant (1 - \mathrm{o}(1)) \cdot \left(\frac{(\gamma/2) \log_{(k+1)} n}{\log_{(k)} n}\right)^{\varepsilon_k n_{k+1}}$$

$$= (1 - \mathrm{o}(1)) \cdot \left((\gamma/2) \log_{(k+1)} n\right)^{\varepsilon_k n_{k+1}} \cdot n^{-\varepsilon_k}$$

$$\geqslant 2 \cdot n^{-\varepsilon_k}.$$

Since $\left((\gamma/2) \log_{(k+1)} n\right)^{\varepsilon_k n_{k+1}} = \omega(1)$, one can set n sufficiently large to ensure that the final inequality is satisfied. The choice of $C(k, \gamma)$ should ensure this bound. $\square$

Lemma 3.3 proves that every given vertex has degree at least $\varepsilon_k \cdot n_{k+1}$ with probability at least $n^{-0.1}$. By linearity of expectation, the expected number of vertices with degree at least $\varepsilon_k \cdot n_{k+1}$ is at least $n^{0.9}$. This observation strongly suggests that, with high probability, there will be at least one such vertex in the graph. We now proceed to formally prove this intuition using Chebyshev's inequality. For this analysis, we prove an upper bound of $\mathrm{o}\left(n \log^2 n\right)$ on the variance of the number of vertices with degree at least $\varepsilon_k \cdot n_{k+1}$.

Lemma 3.4. *Fix a constant $\gamma > 0$. There exists an infinite sequence of natural numbers $D(0, \gamma) \leqslant D(1, \gamma) \leqslant D(2, \gamma) \leqslant \cdots$ such that the following statement holds. For all integers $k \geqslant 0$, $n \geqslant D(k, \gamma)$, and for any two distinct vertices v, v' in the graph $G \sim G(n, \gamma/n)$, we have the estimate*

$$\mathrm{CoVar}\left[\mathbb{1}\left(N_{v,k} \geqslant \varepsilon_k \cdot n_{k+1}\right), \mathbb{1}\left(N_{v',k} \geqslant \varepsilon_k \cdot n_{k+1}\right)\right] = o\left(\log^2 n/n\right),$$

where $\varepsilon_k = \varepsilon_0 \cdot 2^{-k}$ is defined as in Lemma 3.3.

Proof. It is instructive to present a high-level proof overview before proceeding with the formal details. For brevity, let E represent the event $N_{v,k} \geqslant \varepsilon_k \cdot n_{k+1}$

and E' represent the event that $N_{v',k} \geqslant \varepsilon_k \cdot n_{k+1}$. Our objective is to prove that the two events E and E' are "essentially independent" of each other. Towards this objective, we shall identify an event Good such that

1. $\Pr[\neg\mathsf{Good}] = \mathrm{o}(\log n/n)$, and
2. Conditioned on the Good event, the joint probability of the events E and E' is roughly the product of their marginal probabilities, i.e.,

$$\Pr[E, E'|\mathsf{Good}] \leqslant \Pr[E] \cdot \Pr[E'] + \mathrm{o}\Big(\log^2 n/n\Big).$$

The Good event occurs when the vertices v and v' are at a distance $> 2k$ from each other in the graph G (and few additional properties are satisfied by the vertices v and v'). This Good event, in particular, entails that the ball of vertices that are at distance $\leqslant k$ from v is disjoint from the ball of vertices at are at distance $\leqslant k$ from v'. This separation of the radius-k balls centered at v and v' suffices to make E and E' (essentially) independent.

Using the properties above, we can estimate the covariance as follows.

$$\begin{aligned}
\mathrm{CoVar}\left[\mathbb{1}(E), \mathbb{1}(E')\right] &= \Pr[E, E'] - \Pr[E] \cdot \Pr[E'] \\
&= \Pr[E, E', \mathsf{Good}] + \Pr[E, E', \neg\mathsf{Good}] - \Pr[E] \cdot \Pr[E'] \\
&\leqslant \Pr[E, E'|\mathsf{Good}] \cdot \Pr[\mathsf{Good}] + \Pr[\neg\mathsf{Good}] - \Pr[E] \cdot \Pr[E'] \\
&< \Big(\Pr[E] \cdot \Pr[E'] + \mathrm{o}\Big(\log^2 n/n\Big)\Big) \cdot 1 + \mathrm{o}(\log n/n) - \Pr[E] \cdot \Pr[E'] \\
&= \mathrm{o}\Big(\log^2 n/n\Big).
\end{aligned}$$

Formally, we define the $\mathsf{Good}(v, v', G)$ event as follows.

$$\begin{aligned}
\{\mathsf{true}, \mathsf{false}\} \ni \mathsf{Good}(v, v', G) = &\left(\bigwedge_{i=0}^{k} N_{v,i} \leqslant 3n_{i+1}\right) \wedge \left(\bigwedge_{i=0}^{k} N_{v',i} \leqslant 3n_{i+1}\right) \\
&\wedge \mathrm{dist}_G(v, v') > 2k.
\end{aligned}$$

The following claims formalizes the two intuitive statements made earlier.

Claim 5.

$$\Pr[\neg\mathsf{Good}(v, v', G)] = o\left(\frac{\log n}{n}\right).$$

Claim 6.

$$\Pr[E, E'|\mathsf{Good}(v, v', G)] \leqslant \Pr[E] \cdot \Pr[E'] + o\left(\frac{\log^2 n}{n}\right).$$

The natural numbers $D(k,\gamma)$ are chosen sufficiently large such that for all $n \geqslant D(k,\gamma)$ the claims above hold. As demonstrated above, the proof of our lemma follows from these two claims. □

Proof of Claim 5 . We already know by Lemma 3.2 that for any $i \in \{0,1,\ldots,k\}$, we have

$$\Pr[N_{v,i} > 3 \cdot n_{i+1}] \leqslant \frac{i}{n^2}.$$

Therefore, by union bound, we have

$$\Pr\left[\bigvee_{i=0}^{k} N_{v,i} > 3 \cdot n_{i+1}\right] \leqslant \frac{0+1+\cdots+k}{n^2} \leqslant \frac{k^2}{n^2} = \mathrm{o}\left(\frac{1}{n}\right). \tag{5}$$

Similarly, we also have

$$\Pr\left[\bigvee_{i=0}^{k} N_{v',i} > 3 \cdot n_{i+1}\right] \leqslant \frac{k^2}{n^2} = \mathrm{o}\left(\frac{1}{n}\right). \tag{6}$$

Finally, our objective is to estimate the probability that $\mathrm{dist}_G(v,v') \leqslant 2k$. By symmetry of the vertices, we have

$$\begin{aligned}
\Pr\left[\mathrm{dist}_G(v,v') \leqslant 2k \,\middle|\, \bigwedge_{i=0}^{2k} N_{v,i} \leqslant 3 \cdot n_{i+1}\right] &\leqslant \frac{3(n_1+n_2+\cdots+n_{2k+1})}{n} \\
\text{(By Lemma A.1)} \qquad &\leqslant \frac{6kn_{2k+1}}{n}. \\
\Pr\left[\mathrm{dist}_G(v,v') \leqslant 2k \wedge \left(\bigvee_{i=0}^{2k} N_{v,i} > 3 \cdot n_{i+1}\right)\right] &\leqslant \Pr\left[\bigvee_{i=0}^{2k} N_{v,i} > 3 \cdot n_{i+1}\right] \\
\text{(By Lemma 3.2)} \qquad &\leqslant \frac{1+2+\cdots+2k}{n^2} \leqslant \frac{4k^2}{n^2}.
\end{aligned}$$

We combine these two results to obtain the following conclusion.

$$\begin{aligned}
\Pr[\mathrm{dist}_G(v,v') \leqslant 2k] = {} & \Pr\left[\mathrm{dist}_G(v,v') \leqslant 2k \wedge \left(\bigwedge_{i=0}^{2k} N_{v,i} \leqslant 3 \cdot n_{i+1}\right)\right] \\
& + \Pr\left[\mathrm{dist}_G(v,v') \leqslant 2k \wedge \left(\bigvee_{i=0}^{2k} N_{v,i} > 3 \cdot n_{i+1}\right)\right] \\
\leqslant {} & \Pr\left[\mathrm{dist}_G(v,v') \leqslant 2k \,\middle|\, \bigwedge_{i=0}^{2k} N_{v,i} \leqslant 3 \cdot n_{i+1}\right]
\end{aligned}$$

$$+ \Pr\left[\operatorname{dist}_G(v, v') \leqslant 2k \wedge \left(\bigvee_{i=0}^{2k} N_{v,i} > 3 \cdot n_{i+1}\right)\right]$$

$$\leqslant \frac{6kn_{2k+1}}{n} + \frac{4k^2}{n^2} = \mathrm{o}\left(\frac{\log n}{n}\right) \tag{7}$$

By Equation 5, Equation 6, and Equation 7, we conclude that

$$\begin{aligned}\Pr[\neg\mathsf{Good}(v, v', G)] &\leqslant \Pr\left[\bigvee_{i=0}^{k} N_{v',i} > 8 \cdot n_{i+1}\right] + \Pr\left[\bigvee_{i=0}^{k} N_{v',i} > 8 \cdot n_{i+1}\right] \\ &\quad + \Pr[\operatorname{dist}_G(v, v') \leqslant 2k] \\ &= \mathrm{o}\left(\frac{\log n}{n}\right).\end{aligned}$$

This derivation concludes the proof of our claim. □

Proof of Claim 6 . A *pattern* in a graph specifies that some edges are included and some other edges are excluded. All remaining edges may or may not be included in the graph. Consequently, the probability of $G \sim G(n, \gamma/n)$ containing a (feasible) pattern ω is $\Pr[\omega \in G] = p^\alpha q^\beta$, where ω specifies α edge inclusions, β edge exclusions, $p = \gamma/n$ and $q = (1 - p)$. Therefore, the probability of $G \sim G(n, p)$ containing two patterns ω and ω' simultaneously is

$$\Pr[\omega \in G, \omega' \in G] = \Pr[\omega \in G] \cdot \Pr[\omega' \in G] \cdot p^{-\alpha^*} q^{-\beta^*},$$

where α^* edges are included both in ω and ω', and β^* edges are excluded in both ω and ω' (by the inclusion exclusion principle).

To analyze our problem, fix a vertex $v \in V(G)$ and its neighborhood $\mathcal{N}_{v,\leqslant k}$. Fix the graph H induced by $\mathcal{N}_{v,\leqslant k}$ arbitrarily. Since $\mathcal{N}_{v,\leqslant k}$ contains all vertices at distance $\leqslant k$, it is implicit that there are no edges between the vertex set $\mathcal{N}_{v,\leqslant k-1}$ and the vertex set $V(G) \setminus \mathcal{N}_{v,\leqslant k}$. Insisting that the neighborhood of v in G satisfies these constraints defines a pattern ω with $E(H)$ edge inclusions and $E(\overline{H}) + (n - N_{v,\leqslant k}) \cdot N_{v,\leqslant k-1}$ edge exclusions (see Figure 1 for intuition).

Similarly, fix a vertex $v' \in V(G)$ and its neighborhood $\mathcal{N}_{v',\leqslant k}$. Fix the graph H' induced by $\mathcal{N}_{v',\leqslant k}$ arbitrarily. This defines a template ω' with $E(H')$ edge inclusions and $E(\overline{H'}) + (n - N_{v',\leqslant k}) \cdot N_{v',\leqslant k-1}$ edge exclusions.

If $\operatorname{dist}_G(v, v') > 2k$, then the vertex sets $\mathcal{N}_{v,\leqslant k}$ and $\mathcal{N}_{v',\leqslant k}$ are disjoint. Therefore, the patterns ω and ω' have $\alpha^* = 0$ common edge inclusions and

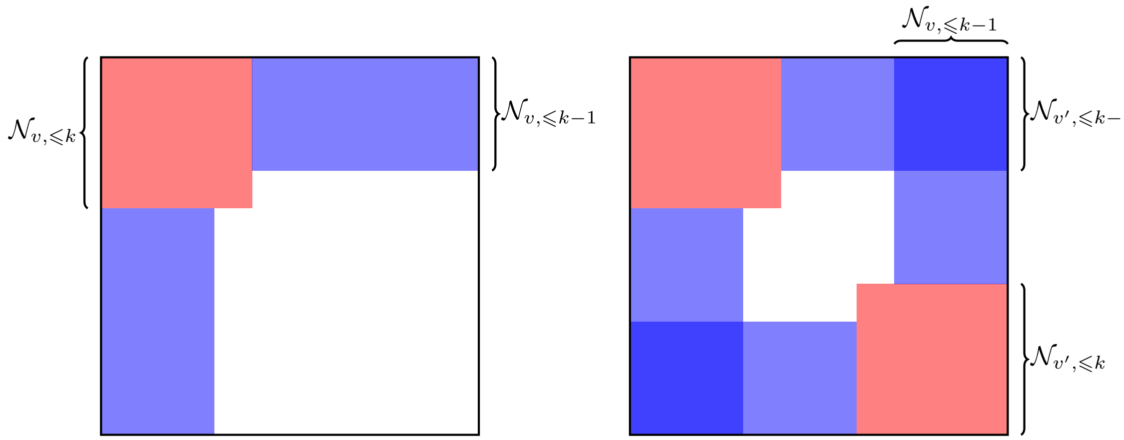

Figure 1: Visualization of the adjacency matrix of $G \sim G(n, \gamma/n)$ containing the pattern ω, and the compensation term in the probability expression when G contains two patterns ω and ω'.

$\beta^* = N_{v,\leqslant k-1} \cdot N_{v',\leqslant k-1}$ common edge exclusions (refer to Figure 1 for clarification of this argument). Therefore, we have

$$\Pr[\omega \in G, \omega' \in G] = \Pr[\omega \in G] \cdot \Pr[\omega' \in G] \cdot q^{-N_{v,\leqslant k-1} \cdot N_{v',\leqslant k-1}},$$

In particular, conditioned on $\mathsf{Good}(v, v', G)$, we have $\mathrm{dist}_G(v, v') > 2k$, $N_{v,\leqslant k-1} \leqslant 3(k-1)^2 n_k$, and $N_{v,\leqslant k-1} \leqslant 3(k-1)^2 n_k$. Hence

$$\frac{\Pr[\omega \in G, \omega' \in G]}{\Pr[\omega \in G] \Pr[\omega' \in G]} \leqslant \left(1 - \frac{\gamma}{n}\right)^{-9k^2 n_k^2} \leqslant \exp\left(\frac{18\gamma k^2 n_k^2}{n}\right) \leqslant 1 + \frac{36\gamma k^2 n_k^2}{n}.$$

This derivation relies on the fact that (i) $(1 - x) \geqslant \exp(-2x)$, when $x \in [0, 1/2]$, (ii) $\exp(x) \leqslant 1 + 2x$, for any $x \in [0, 1]$. The final inequality above relies on the fact that $n_k = \mathrm{o}(\log n)$ and $n_k^2/n = \mathrm{o}(1)$.

Let us summarize the discussion so far. We condition on the event $\mathsf{Good}(v, v', G)$. We arbitrarily choose $\mathcal{N}_{v,\leqslant k}$, $\mathcal{N}_{v',\leqslant k}$, H, and H'. We have proven then that the events "$\mathcal{N}_{v,\leqslant k}$ inducing H" and "$\mathcal{N}_{v',\leqslant k}$ inducing H'" are nearly independent. Furthermore, the joint probability is *multiplicatively-close* to the product of the marginal probabilities.

Therefore, due to this multiplicative closeness, overall we conclude that

$$\Pr[E, E' | \mathsf{Good}(v, v', G)] \leqslant \Pr[E] \cdot \Pr[E'] \cdot \left(1 + \frac{36\gamma k^2 n_k^2}{n}\right)$$

$$\leqslant \Pr[E] \cdot \Pr[E'] + \mathrm{o}\left(\frac{\log^2 n}{n}\right).$$

This derivation completes the proof. $\square$

Proving the lower bound of Theorem 3.1 using Lemma 3.3 and Lemma 3.4. Fix $t \in \mathbb{N}$ and constant $\gamma > 0$. Consider any $n \geqslant \max\{C(t,\gamma), D(t,\gamma)\}$ and $G \sim G(n, \gamma/n)$. Let X_v represent the indicator variable that $N_{v,t} \geqslant \varepsilon_t \cdot n_{t+1}$ in the graph G. Then, the random variable $X = \sum_{v \in V(G)} X_v$ counts the number of vertices with at least $\varepsilon_t \cdot n_{t+1}$ vertices at distance t in the graph G. Our objective is to prove that $X \geqslant 1$ with $1 - \mathrm{o}(1)$ probability; or, equivalently, prove that the probability of $X = 0$ is $\mathrm{o}(1)$. We proceed by the second moment technique (Chebyshev's inequality).

$$\begin{aligned}
\Pr[X = 0] \leqslant \Pr[\,|X - \mathrm{E}[X]| \geqslant \mathrm{E}[X]\,] &\leqslant \frac{\mathrm{Var}\,[X]}{\mathrm{E}[X]^2} \\
&\leqslant \frac{\mathrm{Var}\left[\sum_{v \in V(G)} X_v\right]}{\left(\sum_{v \in V(G)} \mathrm{E}[X_v]\right)^2} \\
\text{(By Lemma 3.3)} \qquad &\leqslant \frac{\sum_{v \in V(G)} \mathrm{Var}\,[X_v] + \sum_{v,v' \in V(G):\ v \neq v'} \mathrm{CoVar}\,[X_v, X_{v'}]}{(n \cdot n^{-0.1})^2} \\
&< \frac{n \cdot 1 + n^2\,\mathrm{CoVar}\,[X_v, X_{v'}]}{n^{1.8}} \\
\text{(By Lemma 3.4)} \qquad &= \frac{n + \mathrm{o}\left(n \log^2 n\right)}{n^{1.8}} \\
&= \mathrm{o}\left(n^{-0.7}\right).
\end{aligned}$$

A Generalization. This derivation indicates that a more general result holds: there are *several* high-degree vertices in a random graph. For example, the following large deviation bound is immediate.

$$\Pr[X \leqslant \mathrm{E}[X]/2] \leqslant \Pr[\,|X - \mathrm{E}[X]| \geqslant \mathrm{E}[X]/2\,] \leqslant 4 \cdot \frac{\mathrm{Var}\,[X]}{\mathrm{E}[X]^2} = \mathrm{o}\left(n^{-0.7}\right).$$

This bound shall be useful in Section 5, so we formalize it as the following lemma.

Lemma 3.5. *Fix a positive constant γ and $t \in \mathbb{N}$. For $G \sim G(n, \gamma/n)$ and sufficiently large $n \in \mathbb{N}$, there are (at least) $n^{0.9}/2$ vertices $v \in V(G)$ such that $N_{v,t} \geqslant \varepsilon_t \cdot n_{t+1}$ with probability $1 - o(n^{-0.7})$.*

4. Upper bound on the Chromatic Number of G^t

We shall prove the following theorem.

Theorem 4.1. *Fix any natural number t and a positive constant γ. For sufficiently large $n \in \mathbb{N}$, the graph $G \sim G(n, \gamma/n)$ satisfies*

$$\Pr\left[\chi\left(G^t\right) \leqslant 6 \cdot n_t + \frac{t \cdot \log_{(t)} n}{\log_{(2t+1)} n}\right] \geqslant 1 - o\left(\frac{1}{\sqrt{\log n}}\right).$$

We remind the reader that the maximum degree of G^t is $\Theta(n_{t+1})$ asymptotically almost surely. So, our proof proceeds by showing that the discrepancy of G^t is $\mathcal{O}(n_t)$, which is $\ll n_{t+1}$. This upper bound on the discrepancy of G^t translates into an upper bound on its chromatic number.

Furthermore, for $t = 2$, this bound is *tight*. Consider $G \sim G(n, \gamma/n)$. Arbitrarily partition $V(G) = L \cup R$ such that $|L| = |R| = n/2$. [3] proves that the induced subgraph $G[L, R]$ has a vertex with degree at least $c_\gamma \cdot n_2$ (a.a.s.), where $G \sim G(n, \gamma/n)$ and c_γ is a positive constant. Therefore, G^2 has a clique of size (at least) $c_\gamma \cdot n_2$.

Proof. Given a threshold τ, we say that the event $\mathsf{Good}_\tau(G)$ holds if, for all vertices $v \in V(G)$, the following conditions are true.

1. $N_{v,\leqslant t-1} \leqslant 6 \cdot n_t$,
2. $N_{v,\leqslant 2t} \leqslant 6 \cdot n_{2t+1}$, and
3. The induced subgraph $G\left[\mathcal{N}_{v,\leqslant 2t}\right]$ is a tree or $N_{v,\leqslant 2t} \leqslant \tau$.

Our proof follows as a consequence of Claim 7 and Claim 8. Claim 7 proves that the probability of the good event *not* occurring is o(1). Finally, Claim 8 proves that the chromatic number of G^t is at most τ, for any good graph G.

Claim 7. *Fix $\tau = 6 \cdot n_t + \frac{t \cdot \log_{(t)} n}{\log_{(2t+1)} n}$. For $G \sim G(n, \gamma/n)$, we have*

$$\Pr[\neg\mathsf{Good}_\tau(G)] \leqslant \frac{5t^2}{n} + \frac{(6t+1)\cdot\left(1 + (\mathrm{e}^2\gamma/2)^{6t}\right)}{\tau}.$$

Claim 8. *Fix $\tau = 6 \cdot n_t + \frac{t \cdot \log_{(t)} n}{\log_{(2t+1)} n}$. For any graph G, if $\mathsf{Good}_\tau(G)$ holds, then every vertex $v \in V(G^t)$ has at most τ neighbors with degree $\geqslant \tau$ in the graph G^t. Consequently, we have $\chi\left(G^t\right) \leqslant \tau$.*

All that remains is proving Claim 7 and Claim 8. □

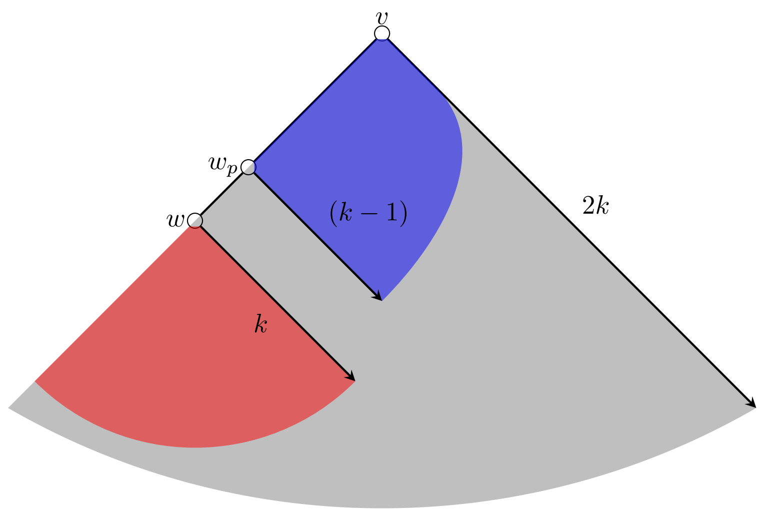

Figure 2: The gray region denotes vertices at distance $\leqslant 2t$ from v in G. The vertex w is at distance at most t from v. Vertices at distance $\leqslant t$ from w occur either in the blue or the red region. The blue region denotes vertices at distance $\leqslant (t-1)$ from w_p. The red region denotes vertices at distance $\leqslant t$ from w whose shortest path to w *does not* pass through its parent w_p. The blue and the red regions are entirely inside the gray region. We show that the blue region typically has at most $6 \cdot n_t$ vertices. Furthermore, typically, the number of vertices w having $\geqslant \frac{t \cdot \log_{(t)} n}{\log_{(2t+1)} n}$ vertices in the red region is at most $6 \cdot n_t$.

Proof of Claim 7 . We shall show that the probability of $\neg\mathsf{Good}_\tau(G)$ is very small. Similar to the proof of Theorem 3.1 using Lemma 3.2, we conclude from Equation 2 that

$$\Pr[\exists v \in V(G) \text{ s.t. } N_{v,\leqslant t-1} > 6 \cdot n_t] \leqslant \frac{(t-1)^2}{n}. \tag{8}$$

Furthermore,

$$\Pr[\exists v \in V(G) \text{ s.t. } N_{v,\leqslant 2t} > 6 \cdot n_{2t+1}] \leqslant \frac{(2t)^2}{n}. \tag{9}$$

These two bounds upper bound the first two ways of failing.

Finally, to bound the third failure event, we proceed at follows. We shall identity an alternative graph property that shall automatically imply that "for all $v \in V(G)$, we have $G[\mathcal{N}_{v,\leqslant t}]$ is a tree or $N_{v,\leqslant 2t} \leqslant \tau$." Fix distinct vertices $v_1, v_2, \ldots, v_s \in V(G)$. When $G \sim G(n, \gamma/n)$, the probability that the induced graph $G[\{v_1, \ldots, v_s\}]$ has at least s edges is

$$\leqslant \binom{\binom{s}{2}}{s} \cdot p^s \cdot 1^{\binom{s}{2}-s} \leqslant \left(\frac{\mathrm{e}sp}{2}\right)^s,$$

where $p = \gamma/n$.

Let X_ℓ represent the number of subsets $\{v_1, \ldots, v_s\} \subseteq V(G)$, where $s \in \{1, 2, \ldots, \ell\}$, such that the the induced subgraph $G[\{v_1, \ldots, v_s\}]$ has at least s edges. We refer to such subgraphs as "dense." Then, we have

$$\begin{aligned} \mathrm{E}[X_\ell] &\leqslant \sum_{s=1}^{\ell} \binom{n}{s} \left(\frac{\mathrm{e}sp}{2}\right)^s \leqslant \sum_{s=1}^{\ell} \left(\frac{\mathrm{e}^2\gamma}{2}\right)^s \\ &\leqslant \ell \cdot \max\left\{1, (\mathrm{e}\gamma/2)^\ell\right\} < \ell \cdot \left(1 + (\mathrm{e}\gamma/2)^\ell\right). \end{aligned}$$

Therefore, by Markov inequality, we have

$$\Pr[X_\ell \geqslant B] < \frac{\ell \cdot (1 + (\mathrm{e}\gamma/2)^\ell)}{B}. \tag{10}$$

In Equation 10, substitute $\ell = 6t+1$ and $B = \tau$ to get

$$\Pr[X_{6t+1} \geqslant \tau] < \frac{(6t+1) \cdot (1 + (\mathrm{e}\gamma/2)^{6t+1})}{\tau} = \Theta\left(\frac{1}{\tau}\right).$$

The above probability bound implies that it is unlikely for a random $G \sim G(n, \gamma/n)$ to have more than τ "dense subgraphs" of size $\leqslant 6t+1$.

Now, we can proceed to bound the third failure event. Consider a graph G that has at most τ "dense subgraphs" of size $\leqslant 6t+1$. Consider a vertex $v \in V(G)$ such that the induced subgraph $G[\mathcal{N}_{v,\leqslant 2t}]$ is *not* a tree. If $G[\mathcal{N}_{v,\leqslant 2t}]$ is not a tree, then there exists a cycle C in $G[\mathcal{N}_{v,\leqslant 2t}]$. Let $e = (u_1, u_2)$ be an edge in the cycle C such that $v \notin \{u_1, u_2\}$. Let V_1 be the set of vertices on the shortest path from v to u_1, and V_2 be the set of vertices on the shortest path from v to u_2. We know that $|V_1 \cup V_2| \leqslant (2t+1) + (2t+1) - 1 = 4t+1$. Observe that the induced subgraph $G[V_1 \cup V_2]$ contains a cycle.

Fix an arbitrary vertex $w \in \mathcal{N}_{v,\leqslant 2t}$. Let V_w be the set of vertices on the shortest path from v to w. We know that $|V_w| \leqslant 2t+1$. Therefore, $G_w := G[V_1 \cup V_2 \cup V_w]$ defines a dense subgraph of size $s \leqslant 6t+1$ (because it is a connected graph containing at least the cycle induced by the vertices $V_1 \cup V_2$).

Observe that, for each $w \in \mathcal{N}_{v,\leqslant 2t}$, the corresponding induced subgraph G_w is distinct. Therefore, the total number of such possible vertices w, i.e.,, $N_{v,\leqslant 2t}$, is at most τ (otherwise, the total number of small dense subgraphs will surpass τ). To conclude, for $G \sim G(n, \gamma/n)$, we have shown that

$$\begin{aligned} \Pr[\exists v \in V(G) \text{ s.t. } G[\mathcal{N}_{v,\leqslant 2t}] \text{ is not a tree } \wedge\ N_{v,2t} > \tau] \\ \leqslant \frac{(6t+1)\cdot\left(1+(\mathrm{e}\gamma/2)^{6t+1}\right)}{\tau}. \end{aligned} \tag{11}$$

By Equation 8, Equation 9, and Equation 11, our claim follows using the union bound. □

Proof of Claim 8 . Consider any G such that $\mathsf{Good}_\tau(G)$ holds. Then, for every $v \in V(G)$, we have the following guarantees.

1. $N_{v,t-1} \leqslant 6 \cdot n_t$,
2. $N_{v,2t} \leqslant 6 \cdot n_{2t+1}$, and
3. The subgraph H_v is a tree or $N_{v,\leqslant 2t} \leqslant \tau$, where $H_v := G[\mathcal{N}_{v,2t}]$.

Fix any $v \in V(G)$. Our first objective is to prove that the following set of "high degree vertices close to the vertex v" has size at most τ.

$$S_v := \{w \in \mathcal{N}_{v,\leqslant t} \colon N_{w,\leqslant t} \geqslant \tau\}.$$

We consider the following two exhaustive cases.

Case 1: H_v is not a tree. In this case, we know that the number of vertices in H_v is $\leqslant \tau$ (from the third property ensured by the Good_τ event).

Therefore, v has $\leqslant \tau$ neighbors in G^t. Consequently, the vertex v can have at most τ neighbors in G^t with degree $\geqslant \tau$ in G^t.

Case 2: H_v is a tree. Consider H_v rooted at v. Consider any $w \in \mathcal{N}_{v,\leqslant t}$ such that w has $\geqslant \tau$ neighbors in G^t. The neighbors of w in G^t are vertices that are connected to w via paths of length $\leqslant t$ in G. Let w_p be the parent of w in H_v. These paths of length $\leqslant k$ emanating from w are of two types (refer Figure 2) (a) the path goes through w_p (and their final vertex is in the blue region), or (b) the path does not go through w_p (and their final vertex is in the red region). Observe that $N_{w_p,\leqslant t-1} \leqslant 6n_t$ (guaranteed by the good event). Therefore, there are at least $(\tau - 6n_t)$ paths of length (at most) t from w that do not go through w_p.

Intuitively, every $w \in \mathcal{N}_{v,\leqslant t}$ with $\geqslant \tau$ degree in G^t has $\geqslant (\tau - 6n_t)$ paths going "downward" in the tree (path leading to vertices in the red region of Figure 2). The final vertices of these downward paths are neighbors of w in G^t.

Let $M := |S_v|$. Note that any vertex $x \in \mathcal{N}_{v,\leqslant 2t}$ can occur as the final vertex of at most t distinct "downward paths" (because the vertex x can be the end point of paths starting from its t ancestors). Therefore, we have the following upper bound

$$M \cdot (\tau - 6n_t) \leqslant t \cdot N_{v,\leqslant 2t} \leqslant 6t \cdot n_{2t+1}.$$

If $M > 6 \cdot n_t$, then $M \cdot (\tau - 6n_t) > 6t \cdot n_{2t+1}$, which violates the bound above. Therefore, $M \leqslant 6 \cdot n_t < \tau$ must hold, which proves that there are at most τ neighbors of v in G^t that have degree $\geqslant \tau$, proving the first part of our claim.

For the final part of our claim, we shall prove a general result for an arbitrary graph J. Assume that, for every vertex $v \in V(J)$, the number of neighbors of v with degree $\geqslant h$ is at most h (for some $h \in \mathbb{N}$). Then, we shall prove that $\chi(J) \leqslant h$. For the proof, order the vertices in a (weakly) increasing order of their degree. If possible let there exist a vertex v that has $(h+1)$ neighbors to its right. These neighbors, in turn, would have degree at least $(h+1)$ because the degree of the vertices are (weakly) increasing. Therefore, v has $(h+1)$ vertices with degree $\geqslant (h+1)$, a contradiction. Therefore, any vertex has at most h neighbors to its right, which upper bounds the degeneracy of J and, in turn, the chromatic number of J. This result proves that $\chi(G^t) \leqslant \tau$. □

5. Results for Bipartite Vertex-feature inclusion graph

Let Erdős-Rényi graph $G(n, n, p)$ represent the distribution over (undirected) bipartite graphs with size-n partite sets and each edge included in the graph independently with probability p. For a bipartite graph $G = (L, R, E)$, in this section, we interpret the left partite set L as the set of vertices and the right partite set R as the set of features. The square of this graph $H := G^2 = (L, E')$ is an undirected graph defined by

$$E' = \{(u, v) \colon u, v \in L, \exists w \in R \text{ s.t. } (u, w), (v, w) \in E\}\,.$$

We shall prove a tight estimate of the chromatic number of $H = G^2$ when $G \sim G(n, n, \gamma/n)$, for any positive constant γ.

Theorem 5.1. *Fix any positive constant γ. There exists a positive constant c_γ such that, for sufficiently large $n \in \mathbb{N}$, the bipartite graph $G \sim G(n, n, \gamma/n)$ satisfies*

$$\Pr\left[\frac{\chi(G^2)}{n_2} \in [c_\gamma, 7]\right] \geqslant 1 - o(1).$$

Proof. Cheng. Maji and Pothen [3] prove that

$$\Pr\left[\frac{\chi(G^2)}{n_2} \geqslant c_\gamma\right] \geqslant \Pr\left[\frac{\omega(G^2)}{n_2} \geqslant c_\gamma\right] \geqslant 1 - \mathrm{o}(1),$$

where $\omega(G^2)$ is the size of a maximum clique in the square of the graph.

For the upper bound, consider the undirected graph $G' \sim G(2n, \gamma/n)$, where the vertex set $V(G') = L \cup R$. Theorem 4.1 proves that, for sufficiently large $n \in \mathbb{N}$, we have

$$\Pr\left[\frac{\chi\left(\,(G')^2\,\right)}{n_2} \leqslant 7\right] \geqslant 1 - \mathrm{o}(1).$$

Observe that the induced subgraph $G := G'[L, R]$ is distributed identically to $G(n, n, \gamma/n)$, when $G' \sim G(2n, \gamma/n)$.

Fix the graphs G and G', such that $G = G'[L, R]$. Consider $H := G^2$ as defined for vertex-feature inclusion graphs. Note that $E(H) \subseteq E\left(\,(G')^2\,\right)$, because (unlike H) the length-($\leqslant 2$) paths in G' that start and end in L need not necessarily go through only the vertices in R. Therefore, $\chi(H) \leqslant \chi\left(\,(G')^2\,\right)$.

Putting things together, we have

$$\Pr\left[\frac{\chi(G^2)}{n_2} \leqslant 7 \colon G \sim G(n, n, \gamma/n)\right] \leqslant \Pr\left[\frac{\chi\left(\,(G')^2\,\right)}{n_2} \leqslant 7 \colon G' \sim G(2n, \gamma/n)\right]$$

$$\leqslant 1 - \mathrm{o}(1).$$

This derivation completes the proof of our upper bound. □

This result is particularly interesting because $\Delta(G^2) \geqslant \frac{1}{8} \cdot \varepsilon_2 \cdot n_3 \gg n_2$, when $G \sim G(n, n, \gamma/n)$ (similar to Theorem 3.1 for undirected graphs).

Lemma 5.2. *Fix a positive constant γ. For sufficiently large $n \in \mathbb{N}$ and bipartite graph $G \sim G(n, n, \gamma/n)$, the maximum degree of the undirected loopless graph $H := G^2$ is $\geqslant \frac{1}{8}\varepsilon_2 n_3$, with probability $1 - o(1)$.*

Proof. Consider $G' \sim G(2n, \gamma/n)$. Randomly partition the vertex set $V(G')$ into L and R subsets such that $|L| = |R| = n$. Consider the induced subgraph $G := G'[L, R]$. Observe that G is distributed identically to $G(n, n, \gamma/n)$, when $G' \sim G(2n, \gamma/n)$. Our objective is to prove that one of the vertices in $V(G')$ that has high degree in $(G')^2$ also has high-degree in the graph $H = G^2$. Recall that $x_k = \frac{\log x}{\log_{(k)} x}$.

We shall condition on a good event $\mathsf{Good}(G')$ that ensures

1. Every vertex $v \in V(G')$ has $N_{v,1}(G') \leqslant 6 \cdot (2n)_2$ and $N(v, 2)(G') \leqslant 6 \cdot (2n)_3$,
2. For every vertex $v \in V(G')$, the induced subgraph $G'[\mathcal{N}_{v,\leqslant 2}(G')]$ is a tree, or $N_{v,\leqslant 2}(G') \leqslant \varepsilon_2(2n)_3$, and
3. G' has (at least) $(2n)^{0.9}/2$ vertices $v \in V(G')$ such that $N_{v,2}(G') \geqslant \varepsilon_2 \cdot (2n)_3$.

The probability of a graph $G' \sim G(2n, \gamma/n)$ satisfying the first two conditions is $1 - \mathrm{o}(1)$ (by Claim 7), and the probability of a graph $G' \sim G(2n, \gamma/n)$ satisfying the last condition is $1 - \mathrm{o}(1)$ (by Lemma 3.5). By union bound, the probability of $G' \sim G(2n, \gamma/n)$ simultaneously satisfying all three conditions is $1 - \mathrm{o}(1)$.

Fix an arbitrary graph G' that is good. Define

$$V_{\textsf{high-deg}}(G') := \left\{ v \in V(G') \colon N_{v,2}(G') \geqslant \varepsilon_2 \cdot (2n)_3 \right\}.$$

We have $|V_{\textsf{high-deg}}(G')| \geqslant (2n)^{0.9}/2$. Over the random partition L and R of the vertex set $V(G')$ such that $|L| = |R| = n$, consider the random variable

$$S(G') := |L \cap V_{\textsf{high-deg}}(G')|.$$

Note that $S(G') = 0$ if and only if $V_{\mathsf{high\text{-}deg}}(G') \subseteq R$. The probability of $S(G') = 0$, therefore, is

$$\frac{\binom{|R|}{|V_{\mathsf{high\text{-}deg}}(G')|}}{\binom{|L|+|R|}{|V_{\mathsf{high\text{-}deg}}(G')|}} = \frac{\binom{n}{|V_{\mathsf{high\text{-}deg}}(G')|}}{\binom{2n}{|V_{\mathsf{high\text{-}deg}}(G')|}} = \frac{n(n-1)\cdots(n-|V_{\mathsf{high\text{-}deg}}(G')|+1)}{2n(2n-1)\cdots(2n-|V_{\mathsf{high\text{-}deg}}(G')|+1)} \leqslant \frac{1}{2^{|V_{\mathsf{high\text{-}deg}}(G')|}}.$$

So, $S(G') \geqslant 1$ with probability $1 - \mathrm{o}(1)$.

Fix a vertex $v \in L \cap V_{\mathsf{high\text{-}deg}}(G')$. The following analysis is over the random partition of $V(G')$ into equal sets L and R conditioned on the vertex v being in the partite set L. For brevity, define $\mathcal{N}_1 := \mathcal{N}_{v,1}(G')$ (i.e., the set of vertices in $V(G')$ at distance 1 from v in the graph G'), and $\mathcal{N}_2 := \mathcal{N}_{v,2}(G')$ (i.e., the set of vertices in $V(G')$ at distance 2 from v in the graph G').

Since $v \in V_{\mathsf{high\text{-}deg}}(G')$, we have $|\mathcal{N}_2| \geqslant \varepsilon_2 \cdot (2n)_3$. Therefore, $N_{v,\leqslant 2}(G') \geqslant \varepsilon_2 \cdot (2n)_3 + 2$. By the second condition of the good event, the induced subgraph $G'[\mathcal{N}_1 \cup \mathcal{N}_2]$ is a tree. Let T_v represent this tree rooted at v.

For $w \in \mathcal{N}_2$, let X_w be the indicator variable for the event that w is at distance 2 in the bipartite graph $G := G'[L, R]$. Let $w_p \in \mathcal{N}_1$ represent the parent of w in T_v. Note that $X_w = 1$ if and only if "$w \in L$ and $w_p \in R$" (recall that we have already conditioned on $v \in L$). Therefore, we have

$$\begin{aligned} \mathrm{E}[X_w] &= \Pr[w \in L | v \in L] \cdot \Pr[w_p \in R | v \in L, w \in L] \\ &= \frac{n-1}{2n-1} \cdot \frac{n}{2n-2} \\ &= \frac{1}{4} \cdot \frac{n}{n-1/2}. \end{aligned}$$

Let $X = \sum_{w \in \mathcal{N}_2} X_w$ represent the number of vertices in the bipartite graph G that are at distance 2 from the vertex v in this bipartite graph. That is, X represents the degree of the vertex v in the graph $H := G^2$. By the linearity of expectation, the expected value of X is

$$\mathrm{E}[X] = |\mathcal{N}_2| \cdot \frac{1}{4} \cdot \frac{n}{n-1/2} > \frac{1}{4} \cdot \varepsilon_2 (2n)_3. \tag{12}$$

Our objective is to show that X is typically close to this expected value using the second moment technique.

For two vertices $w, w' \in \mathcal{N}_2$, let w_p and w'_p represent their respective parents in the tree T_v.

Claim 9.

$$\mathrm{CoVar}\left[X_w, X_{w'}\right] \leqslant \begin{cases} 1, & \text{if } w_p = w'_p \\ \mathcal{O}(\frac{1}{n^2}), & \text{if } w_p \neq w'_p \end{cases}.$$

We emphasize that, in the claim above, the case $w = w'$ is covered in the case $w_p = w'_p$. We can use this claim to prove the following bound on the variance of X.

Claim 10.

$$\frac{\mathrm{Var}\,[X]}{\mathrm{E}[X]^2} \leqslant \mathcal{O}\left(\frac{\log_{(3)} n}{\log_{(2)} n}\right) = o(1).$$

Now, from Chebyshev's inequality, it follows that

$$\Pr[X \leqslant \mathrm{E}[X]/2] \leqslant 4 \cdot \frac{\mathrm{Var}\,[X]}{\mathrm{E}[X]^2} = \mathrm{o}(1).$$

Therefore, we conclude that the degree of v in the graph H is $\geqslant \frac{1}{8} \cdot \varepsilon_2 (2n)_3 \geqslant \frac{1}{8}\varepsilon_2 n_3$ with $1 - \mathrm{o}(1)$ probability. All that remains is proving Claim 9 and Claim 10, which are included below. $\square$

Proof of Claim 9 . The only non-trivial case is when $w_p \neq w'_p$ (the other one is immediate because X_w and $X_{w'}$ are indicator variables).

$$\begin{aligned}
\mathrm{E}[X_w \cdot X_{w'}] &= \Pr[w \in L | v \in L] \times \Pr[w' \in L | v \in L, w \in L] \\
&\quad \times \Pr[w_p \in R | v \in L, w \in L, w' \in L] \\
&\quad \times \Pr\Big[w'_p \in R \Big| v \in L, w \in L, w' \in L, w_p \in R\Big] \\
&= \frac{n-1}{2n-1} \cdot \frac{n-2}{2n-2} \cdot \frac{n}{2n-3} \cdot \frac{n-1}{2n-4} \\
&= \frac{1}{16} \cdot \frac{n(n-1)}{(n-1/2)(n-3/2)}
\end{aligned}$$

We already know that

$$\mathrm{E}[X_w] \cdot \mathrm{E}[X_{w'}] = \frac{1}{16} \cdot \frac{n^2}{(n-1/2)^2}.$$

Therefore, we have

$$\begin{aligned}
\mathrm{CoVar}\,[X_w, X_{w'}] &= \mathrm{E}[X_w \cdot X_{w'}] - \mathrm{E}[X_w] \cdot \mathrm{E}[X_{w'}] \\
&= \frac{1}{16} \cdot \frac{n(n-1)}{(n-1/2)(n-3/2)} - \frac{1}{16} \cdot \frac{n^2}{(n-1/2)^2} \\
&= \frac{1}{16} \cdot \frac{n}{n-1/2} \left(\frac{n-1}{n-3/2} - \frac{n}{n-1/2}\right)
\end{aligned}$$

$$= \frac{1}{32} \cdot \frac{n}{(n-1/2)^2(n-3/2)} = \mathcal{O}\left(\frac{1}{n^2}\right).$$

This derivation completes the proof of this claim. □

Proof of Claim 10 . Observe that

$$\begin{aligned}\mathrm{Var}\left[X\right] &= \sum_{w,w' \in \mathcal{N}_2} \mathrm{CoVar}\left[X_w, X_{w'}\right] \\ &= \sum_{w,w' \in \mathcal{N}_2 \colon w_p = w'_p} \mathrm{CoVar}\left[X_w, X_{w'}\right] + \sum_{w,w' \in \mathcal{N}_2 \colon w_p \neq w'_p} \mathrm{CoVar}\left[X_w, X_{w'}\right].\end{aligned}$$

Bounding the first term. For every vertex $\widetilde{w} \in \mathcal{N}_1$, let $a_{\widetilde{w}}$ represent the number of children of $\widetilde{w}$ in the tree T_v. By the first constraint of the good event, we have the following constraint

$$\sum_{\widetilde{w} \in \mathcal{N}_1} a_{\widetilde{w}} \leqslant |\mathcal{N}_2| \leqslant 6(2n)_3 =: A$$

Furthermore, the first constraint of the good event also implies that, for all $\widetilde{w} \in \mathcal{N}_1$, we have

$$a_{\widetilde{w}} \leqslant 6(2n)_2 =: M.$$

The second term of the variance is upper bounded by (due to Claim 10)

$$\left(\sum_{\widetilde{w} \in \mathcal{N}_1} a_{\widetilde{w}}^2\right) \cdot 1 \leqslant \sum_{\widetilde{w} \in \mathcal{N}_1} a_{\widetilde{w}} \cdot M = M \left(\sum_{\widetilde{w} \in \mathcal{N}_1} a_{\widetilde{w}}\right) \leqslant MA = \mathcal{O}(n_2 n_3).$$

Bounding the second term. The second term is upper-bounded by

$$|\mathcal{N}_2|^2 \cdot \mathcal{O}\left(\frac{1}{n^2}\right) = \mathrm{o}\left(\frac{\log^2 n}{n^2}\right),$$

because $|\mathcal{N}_2| \leqslant 6(2n)_3$ and Claim 9.
Putting things together. We conclude that

$$\mathrm{Var}\left[X\right] \leqslant \mathcal{O}(n_2 n_3) + \mathrm{o}\left(\log^2 n / n^2\right) = \mathcal{O}(n_2 n_3).$$

We know that $\mathrm{E}[X] \geqslant \frac{1}{16}\left(\varepsilon_2 (2n)_3\right)^2 = \Theta\left(n_3^2\right)$, which completes the proof of the claim. □

References

[1] D. Achlioptas and A. Naor, *The two possible values of the chromatic number of a random graph*, Annals of Mathematics, (2005), pp. 1335–1351. 2, 5

[2] M. Behrisch, A. Taraz, and M. Ueckerdt, *Coloring random intersection graphs and complex networks*, SIAM J. Discr. Math., 23 (2009), pp. 288–299. 5

[3] X. Cheng, H. K. Maji, and A. Pothen, *Graphs with tunable chromatic numbers for parallel coloring*, in Proceedings of the SIAM Workshop on Combinatorial Scientific Computing, 2020, pp. 54–64. 1, 2, 20, 25

[4] A. Coja-Oghlan and D. Vilenchik, *Chasing the k-colorability threshold*, in 54th Ann. Symp. Found, Comp.Sci., IEEE, 2013, pp. 380–389. 2, 5

[5] A. Frieze and M. Karoński, *Introduction to random graphs*, Cambridge University Press, 2016. 5, 31

[6] A. H. Gebremedhin, F. Manne, and A. Pothen, *What color is your Jacobian? Graph coloring for computing derivatives*, SIAM Review, 47 (2005), pp. 629–705. 2

[7] A. N. Lagerås and M. Lindholm, *A note on the component structure in random intersection graphs with tunable clustering*, The Electronic Journal of Combinatorics, 15 (2008), p. #N10. 5

Appendix A. Technical Results

Lemma A.1 (Sum of Inverse of Nested Logarithms)**.** *Fix* $t \in \{0, 1, \dots\}$*. For* $n \geqslant \mathrm{tower}(t, \mathrm{e})$*, the following bound holds.*

$$\sum_{i=0}^{t} \frac{1}{\log_{(i)} n} \leqslant \frac{2}{\log_{(t)} n}.$$

Proof. For $n \geqslant \mathrm{tower}(t, \mathrm{e})$, observe that $\log_{(i)} n \geqslant \log_{(t)} n > 0$. Next, we shall use the fact that $\log x \leqslant x/\mathrm{e}$, for all $x \in \mathbb{R}$, in the following derivation.

$$\sum_{i=0}^{t} \frac{1}{\log_{(i)} n} \leqslant \sum_{i=0}^{t} \frac{1}{\exp(t-i)\log_{(t)} n} < \frac{1}{1-1/\mathrm{e}} \cdot \frac{1}{\log_{(t)} n} < \frac{2}{\log_{(t)} n}.$$

□

The following estimates of the binomial coefficients are well-established (refer, for example, [5]).

Lemma A.2 (Estimation of Binomial Coefficients)**.** *For $a, b \in \mathbb{N}$, the following bounds hold.*

$$\left(\frac{a}{b}\right)^b \leqslant \binom{a}{b} \leqslant \left(\frac{\mathrm{e}a}{b}\right)^b$$

Lemma A.3 (Estimation of Logarithm)**.** *The following inequalities hold.*

1. *For $x \in [0,1]$, $\log(1-x) \leqslant -x - x^2/2$*
2. *For $x \in [0,1/2]$, $\log(1-x) \geqslant -x - x^2$.*

Lemma A.4 (Application of Jensen's Inequality)**.** *For any $a \geqslant 1$ and $x > 0$, we have $(1-x)^a \geqslant (1-ax)$.*

Kalyan Garapaty
E-mail address: kalyankumar.garapaty@gmail.com

Daniel Lokshtanov
E-mail address: daniello@ucsb.edu

Hemanta K. Maji
E-mail address: hmaji@purdue.edu

Alex Pothen
E-mail address: apothen@purdue.edu