On Independent Sets in Random Graphs*

Amin Coja-Oghlan, Charilaos Efthymiou
Goethe University, Mathematics Institute, Frankfurt 60054, Germany;
e-mail: acoghlan@math.uni-frankfurt.de; efthymiou@math.uni-frankfurt.de

Received 30 August 2012; accepted 17 October 2013
Published online 30 May 2014 in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/rsa.20550

ABSTRACT: The independence number of a sparse random graph $G(n,m)$ of average degree $d = 2m/n$ is well-known to be $(2-\varepsilon_d)n\ln(d)/d \le \alpha(G(n,m)) \le (2+\varepsilon_d)n\ln(d)/d$ with high probability, with $\varepsilon_d \to 0$ in the limit of large d. Moreover, a trivial greedy algorithm w.h.p. finds an independent set of size $n\ln(d)/d$, i.e., about half the maximum size. Yet in spite of 30 years of extensive research no efficient algorithm has emerged to produce an independent set with size $(1+\varepsilon)n\ln(d)/d$ for any *fixed* $\varepsilon > 0$ (independent of both d and n). In this paper we prove that the combinatorial structure of the independent set problem in random graphs undergoes a phase transition as the size k of the independent sets passes the point $k \sim n\ln(d)/d$. Roughly speaking, we prove that independent sets of size $k > (1+\varepsilon)n\ln(d)/d$ form an intricately rugged landscape, in which local search algorithms seem to get stuck. We illustrate this phenomenon by providing an exponential lower bound for the Metropolis process, a Markov chain for sampling independent sets. © 2014 Wiley Periodicals, Inc. Random Struct. Alg., 47, 436–486, 2015

Keywords: random graphs; independent set problem; Metropolis process; phase transitions

1. INTRODUCTION AND RESULTS

1.1. Probabilistic Analysis and the Independent Set Problem

In the early papers on the subject, the motivation behind the probabilistic analysis of algorithms was to alleviate the glum of worst-case analyses by painting a brighter 'average-case' picture [12, 30, 40]. Indeed, simple, greedy-type algorithms turned out to perform rather well on randomly generated input instances, at least for certain ranges of the parameters. Examples of such analyses include Grimmett and McDiarmid [23] (independent set problem), Wilf [41], Achlioptas and Molloy [2] (graph coloring) and Frieze and Suen [18] (k-SAT). By now, Wormald's "method of differential equations" has become a unifying tool for the

Correspondence to: Amin Coja-Oghlan
*Supported by EPSRC grant (EP/G039070/2); DIMAP.
A preliminary version of this work appeares in the proceedings of ACM-SIAM SODA 2011.

analysis of such greedy algorithms [42]. Yet, remarkably, in spite of 30 years of research, for many problems no efficient algorithms, howsoever sophisticated, have been found to outperform those early greedy algorithms markedly.

The independent set problem in random graphs $G(n,m)$ is a case in point. Recall that $G(n,m)$ is a graph on n vertices obtained by choosing m edges uniformly at random (without replacement). We say that $G(n,m)$ has a property *with high probability* if the probability that the property holds tends to 1 as $n \to \infty$. One of the earliest results in the theory of random graphs is a non-constructive argument showing that for $m = \frac{1}{2}\binom{n}{2}$ the independence number of $G(n,m)$ is $\alpha(G(n,m)) \sim 2\log_2(n)$ w.h.p. [7, 14, 31]. Grimmett and McDiarmid [23] analysed a simple algorithm that just constructs an inclusion-maximal independent set greedily on $G(n,m)$: it yields an independent set of size $(1+o(1))\log_2 n$ w.h.p., about half the maximum size. But no algorithm is known to produce an independent set of size $(1+\varepsilon)\log_2 n$ for any fixed $\varepsilon > 0$ in polynomial time with a non-vanishing probability, neither on the basis of a rigorous analysis, nor on the basis of experiments or other evidence. In fact, devising such an algorithm is probably the most prominent open problem in the algorithmic theory of random graphs [17,26]. (However, note that one can find a maximum independent set w.h.p. by trying all $n^{O(\ln n)}$ possible sets of size $2\log_2 n$.)

The situation is no better on sparse random graphs. If we let $d = 2m/n$ denote the average degree, then non-constructive arguments yield

$$\alpha(G(n,m)) \sim \frac{2\ln(d)}{d} \cdot n$$

for $1 \ll d = o(n)$. In the case $d \gg \sqrt{n}$, the proof of this is via a simple second moment argument [7,31]. By contrast, for $1 \ll d \ll \sqrt{n}$, the second moment argument breaks down and additional methods such as large deviations inequalities or a weighted second moment are needed [9, 16]. Yet in either case, no algorithm is known to find an independent set of size $(1+\varepsilon)\frac{\ln d}{d} \cdot n$ in polynomial time with a non-vanishing probability, while 'greedy' yields an independent set of size $(1+o(1))\frac{\ln d}{d} \cdot n$ w.h.p. In the sparse case, the time needed for exhaustive search scales as $\exp(\frac{2n}{d}\ln^2(d))$, i.e., the complexity grows as d decreases.

The aim of this paper is to explore the apparent difficulty of finding large independent sets in random graphs. The focus is on the sparse case, both conceptually and computationally the most challenging case. We exhibit a phase transition in the structure of the problem that occurs as the size of the independent sets passes the point $\frac{\ln d}{d} \cdot n$ up to which efficient algorithms are known to succeed. Roughly speaking, we show that independent sets of sizes bigger than $(1+\varepsilon)\frac{\ln d}{d} \cdot n$ form an intricately rugged landscape, which plausibly explains why local-search algorithms get stuck. Thus, ironically, instead of exhibiting a brighter 'average case' scenario, we end up suggesting that random graphs provide an excellent source of difficult examples. Taking into account the (substantially) different nature of the independent set problem, our work complements the results obtained in [1] for random constraint satisfaction problem such as k-SAT or graph coloring.

1.2. Results

Throughout the paper we will be dealing with sparse random graphs where the average degree $d = 2m/n$ is 'large' but remains bounded as $n \to \infty$. To formalise this sometimes

we work with functions ε_d that tend to zero as d gets large. [1] Unless otherwise specified, the asymptotics are w.r.t. n and we use the standard O-notation. Thus $\alpha(G(n,m)) = (2-\varepsilon_d)\frac{\ln d}{d}\cdot n$ and the greedy algorithm finds independent sets of size $(1+\varepsilon'_d)\frac{\ln d}{d}\cdot n$ w.h.p., where $\varepsilon_d, \varepsilon'_d \to 0$. However, no efficient algorithm is known to find independent sets of size $(1+\varepsilon'')\frac{\ln d}{d}\cdot n$ for any *fixed* $\varepsilon'' > 0$.

For a graph G and an integer k we let $\mathcal{S}_k(G)$ denote the set of all independent sets in G that have size exactly k. What we will show is that in $G(n,m)$ the set $\mathcal{S}_k(G(n,m))$ undergoes a phase transition as $k \sim \frac{\ln d}{d}n$. For two sets $S, T \subset V$ we let $S\triangle T$ denote the symmetric difference of S, T. Moreover, $\mathrm{dist}(S,T) = |S\triangle T|$ is the Hamming distance of S, T viewed as vectors in $\{0,1\}^V$.

To state the result for k smaller than $\frac{\ln d}{d}n$, we need the following concept. Let $\mathcal{S}$ be a set of subsets of V, and let $\gamma > 0$ be an integer. We say that $\mathcal{S}$ is γ-*connected* if for any two sets $\sigma, \tau \in \mathcal{S}$ there exist $\sigma_1, \ldots, \sigma_N \in \mathcal{S}$ such that $\sigma_1 = \sigma$, $\sigma_N = \tau$, and $\mathrm{dist}(\sigma_t, \sigma_{t+1}) \le \gamma$ for all $1 \le t < N$. If $\mathcal{S}_k(G(n,m))$ is γ-connected for some $\gamma = O(1)$, one can easily define various simple Markov chains on $\mathcal{S}_k(G)$ that are ergodic.

Theorem 1. *There exist $\varepsilon_d \to 0$ and $C_d > 0$ such that $\mathcal{S}_k(G(n,m))$ is C_d-connected w.h.p. for any*

$$k \le (1-\varepsilon_d)\frac{\ln d}{d}\cdot n.$$

The proof of Theorem 1 is 'constructive' in the following sense. Suppose given $G = G(n,m)$ we set up an auxiliary graph whose vertices are the independent sets $\mathcal{S}_k(G)$ with $k \le (1-\varepsilon_d)\frac{\ln d}{d}\cdot n$. In the auxiliary graph two independent sets $\sigma, \tau \in \mathcal{S}_k(G)$ are adjacent if $\mathrm{dist}(\sigma,\tau) \le C_d$. Then the proof of Theorem 1 yields an algorithm for finding paths of length $O(n)$ between any two elements of $\mathcal{S}_k(G)$ w.h.p. Thus, intuitively Theorem 1 shows that for $k \le (1-\varepsilon_d)\frac{\ln d}{d}\cdot n$ the set $\mathcal{S}_k(G(n,m))$ is easy to 'navigate' w.h.p.

By contrast, our next result shows that for $k > (1+\varepsilon_d)\frac{\ln d}{d}\cdot n$ the set $\mathcal{S}_k(G(n,m))$ is not just disconnected w.h.p., but that it shatters into exponentially many, exponentially tiny pieces.

Definition 2. *Let $k = k(n)$ be an integer sequence. We say that there occurs* **shattering** *for d, k if there exist constants $\gamma, \zeta > 0$ such that w.h.p. the set $\mathcal{S}_k(G(n,m))$ admits a partition into subsets such that*

1. *Each subset contains at most* $\exp(-\gamma n)\,|\mathcal{S}_k(G(n,m))|$ *independent sets.*
2. *For any σ, τ that belong to different subsets we have* $\mathrm{dist}(\sigma,\tau) \ge \zeta n$.

We prefer "shattering" over the term "clustering" that is common in statistical physics literature. This is because "clustering" does not necessarily provide that condition 1 holds. (For instance, one might say that there is "clustering" in the so-called condensation phase

[1] The reason why we need to speak about d 'large' is that the sparse random graph $G(n,m)$ is not connected. This implies, for instance, that algorithms can find independent sets of size $(1+\varepsilon_d)n\ln(d)/d$ for some $\varepsilon_d \to 0$ by optimizing carefully over the small tree components of $G(n,m)$. Our results/proofs actually carry over to the case that $d = d(n)$ tends to infinity as n grows, but to keep matters as simple as possible, we will confine ourselves to fixed d.

hypothesized in the physics literature, where shattering does *not* occur.) We emphasize that the definition of "shattering" does not require that the individual subsets into which $\mathcal{S}_k(G(n,m))$ decomposes are $O(1)$-connected.

Theorem 3. *There is $\varepsilon_d \to 0$ such that there occurs shattering for all d, k with*

$$(1+\varepsilon_d)\frac{\ln d}{d}\cdot n \le k \le (2-\varepsilon_d)\frac{\ln d}{d}\cdot n.$$

Theorems 1 and 3 deal with the geometry of a single 'layer' $\mathcal{S}_k(G(n,m))$ of independent sets of a specific size. The following two results explore if/how a 'typical' independent set in $\mathcal{S}_k(G(n,m))$ can be extended to a larger one. To formalize the notion of 'typical', we let $\Lambda_k(n,m)$ signify the set of all pairs (G,σ), where G is a graph on $V=\{1,\dots,n\}$ with m edges and $\sigma \in \mathcal{S}_k(G)$. Let $\mathcal{U}_k(n,m)$ be the probability distribution on $\Lambda_k(n,m)$ induced by the following experiment.

Choose a graph $G = G(n,m)$ at random.
If $\alpha(G) \ge k$, choose an independent set $\sigma \in \mathcal{S}_k(G)$ uniformly at random and output (G,σ).

We say a pair (G,σ) chosen from the distribution $\mathcal{U}_k(n,m)$ has a property $\mathcal{P}$ *with high probability* if the probability of the event $\{(G,\sigma)\in\mathcal{P}\}$ tends to one as $n\to\infty$.

Definition 4. *Let $\gamma,\delta \ge 0$, let G be a graph, and let σ be an independent set of G. We say that (G,σ) is (γ,δ)-**expandable** if G has an independent set τ such that $|\tau| \ge (1+\gamma)|\sigma|$ and $|\tau\cap\sigma| \ge (1-\delta)|\sigma|$.*

In the statement of the following theorem and throughout, we omit floor and ceiling signs to simplify the notation.

Theorem 5. *There are $\varepsilon_d, \delta_d \to 0$ such that for any $\varepsilon_d \le \varepsilon \le 1-\varepsilon_d$ the following is true. For $k = (1-\varepsilon)\frac{\ln d}{d}\cdot n$ a pair (G,σ) chosen from the distribution $\mathcal{U}_k(n,m)$ is $((2-\delta_d)\varepsilon/(1-\varepsilon),0)$-expandable w.h.p.*

Theorem 5 shows that w.h.p. in a random graph $G(n,m)$ almost all independent sets of size $k=(1-\varepsilon)\frac{\ln d}{d}\cdot n$ are contained in *some* bigger independent set of size $(1+\varepsilon)\frac{\ln d}{d}\cdot n$. That is, they can be expanded beyond the critical size $\frac{\ln d}{d}\cdot n$ where shattering occurs. However, as k approaches the critical size $\frac{\ln d}{d}\cdot n$, i.e., as $\varepsilon\to 0$, the typical potential for expansion diminishes.

Theorem 6. *There is $\varepsilon_d \to 0$ such that for any ε satisfying $\varepsilon_d \le \varepsilon \le 1-\varepsilon_d$ and $k=(1+\varepsilon)\frac{\ln d}{d}\cdot n$ w.h.p. a pair (G,σ) chosen from the distribution $\mathcal{U}_k(n,m)$ is not (γ,δ)-expandable for any $\gamma > \varepsilon_d$ and*

$$\delta < \gamma + \frac{2(\varepsilon-\varepsilon_d)}{1+\varepsilon}.$$

In other words, Theorem 6 shows that for $k=(1+\varepsilon)\frac{\ln d}{d}\cdot n$, a typical $\sigma\in\mathcal{S}_k(G(n,m))$ cannot be expanded to an independent set of size $(1+\gamma)k$, $\gamma>\varepsilon_d$ without first *reducing* its size below

$$(1-\delta)k = (1-\varepsilon-\gamma(1+\varepsilon)+2\varepsilon_d)\frac{\ln d}{d}\cdot n < \frac{\ln d}{d}\cdot n.$$

However, a random independent set of size $k \leq (2-\varepsilon_d)\ln(d)n/d$ is typically not inclusion-maximal because, for instance, it is unlikely to contain *all* isolated vertices of the random graph $G(n,m)$. For this reason, in Theorem 6, we have $\gamma > \epsilon_d$. (Yet in the situation of Theorem 6 typical independent sets are "almost" inclusion maximal in the sense that the number of vertices with no neighbor inside the independent set is tiny w.h.p.)

Metaphorically, the above results show that w.h.p. the independent sets of $G(n,m)$ form a rugged mountain range. Beyond the 'plateau level' $k \sim \frac{\ln d}{d} \cdot n$ there is an abundance of smaller 'peaks', i.e., independent sets of sizes $(1+\varepsilon)k$ for any $\varepsilon_d < \varepsilon < 1-\varepsilon_d$, almost all of which are not expandable (by much).

The algorithmic equivalent of a mountaineer aiming to ascend to the highest summit is a Markov chain called the *Metropolis process*, [28, 33]. For a given graph G its state space is the set of all independent sets of G. Let I_t be the state at time t. In step $t+1$, the chain chooses a vertex v of G uniformly at random. If $v \in I_t$, then with probability $1/\lambda$ the next state is $I_{t+1} = I_t \setminus \{v\}$, and with probability $1-1/\lambda$ we let $I_{t+1} = I_t$, where $\lambda \geq 1$ is called the *fugacity*. If $v \notin I_t \cup N(I_t)$ (with $N(I_t)$ the neighbourhood of I_t), then $I_{t+1} = I_t \cup \{v\}$. Finally, if $v \in N(I_t)$, then $I_{t+1} = I_t$.

The above process satisfies a set of technical conditions known as ergodicity[2]. In turn ergodicity implies that the process possesses a unique stationary distribution $\pi : \Omega \to [0,1]$, where $\Omega = \bigcup_k \mathcal{S}_k(G(n,m))$. By standard arguments, for the Metropolis process with fugacity λ it holds that $\pi(\sigma) = \lambda^{|\sigma|}/Z(G,\lambda)$, where

$$Z(G,\lambda) = \sum_{k=0}^{n} \lambda^k \cdot |\mathcal{S}_k(G)|$$

is the partition function. Hence, the larger λ, the higher the mass of large independent sets. Let

$$\mu(G,\lambda) = \frac{\partial \ln Z(G,\lambda)}{\partial \ln \lambda} = \sum_{k=0}^{n} k\lambda^k \cdot |\mathcal{S}_k(G)| \, / Z(G,\lambda)$$

denote the average size of an independent set of G under the stationary distribution.

Here, we are interested in finding the rate at which the Metropolis process converges to equilibrium. There are a number of ways of quantifying the closeness to stationarity. Let $P^t(\sigma,\cdot) : \Omega \to [0,1]$ denote the distribution of the state at time t given that σ was the initial state. The *total variation distance* at time t with respect to the initial state σ is

$$\Delta_\sigma(t) = \max_{S \subset \Omega} |P^t(\sigma,S) - \pi(S)| = \frac{1}{2}\sum_{\tau \in \Omega} |P^t(\sigma,\tau) - \pi(\tau)|.$$

Starting from σ, the rate of convergence to stationarity may then be measured by the function

$$\tau_\sigma = \min_t \{\Delta_\sigma(t') < e^{-1} \text{for all } t' > t\}.$$

The **mixing time** of the Metropolis process is defined as

$$T = \max_{\sigma \in \Omega} \tau_\sigma.$$

[2]For finite Markov chains, as the one we consider here, ergodicity is equivalent to the chain being irreducibile and aperiodic.

Our above results on the structure of the sets $\mathcal{S}_k(G(n,m))$ imply that w.h.p. the mixing time of the Metropolis process is exponential if the parameter λ is tuned so that the Metropolis process tries to ascend to independent sets bigger than $(1+\epsilon_d)\frac{\ln d}{d}\cdot n$.

Theorem 7. *There is $\varepsilon_d \to 0$ such that for $\lambda > 1$ with*

$$(1+\varepsilon_d)\frac{\ln d}{d}\cdot n \le \mathbb{E}\left[\mu(G(n,m),\lambda)\right] \le (2-\varepsilon_d)\frac{\ln d}{d}\cdot n. \tag{1}$$

the mixing time of the Metropolis process on $G(n,m)$ is $\exp(\Omega(n))$ w.h.p.

In fact, the proof of Theorem 7 implies that under the assumption (1) even with a "warm start" (i.e., with an initial state chosen from the stationary distribution) the mixing time of the Metropolis process is $\exp(\Omega(n))$ w.h.p.

1.3. Related Work

To our knowledge, the connection between transitions in the geometry of the 'solution space' (in our case, the set of all independent sets of a given size) and the apparent failure of *local algorithms* in finding a solution has been pointed first out in the statistical mechanics literature [19, 29, 32]. In that work, which mostly deals with CSPs such as k-SAT, the shattering phenomenon goes by the name of 'dynamic replica symmetry breaking.' Our present work is clearly inspired by the statistical mechanics ideas although we are unaware of explicit contributions from that line of work addressing the independent set problem in the case of random graphs with average degree $d \gg 1$ prior to this work. Generally, the statistical mechanics work is based on deep, insightful, but, alas, mathematically non-rigorous techniques.

In the case that the average degree d satisfies $d \gg \sqrt{n}$, the independent set problem in random graphs is conceptually somewhat simpler than in the case of $d = o(\sqrt{n})$. The reason for this is that for $d \gg \sqrt{n}$ the second moment method can be used to show that the *number* of independent sets is concentrated about its mean. As we will see in Corollary 19 below, this is actually untrue for sparse random graphs.

The results of the present paper extend the main results from Achlioptas and Coja-Oghlan [1], which dealt with constraint satisfaction problems such as k-SAT or graph coloring, to the independent set problem. This requires new ideas, because the natural questions are somewhat different (for instance, the concept of 'expandability' has no counterpart in CSPs). Furthermore, in [1] we conjectured but did not manage to prove the counterpart of Theorem 1 on the connectivity of $\mathcal{S}_k(G(n,m))$. On a technical level, we owe to [1] the idea of analysing the distribution $\mathcal{U}_k(n,m)$ via a different distribution $\mathcal{P}_k(n,m)$, the so-called 'planted model' (see Section 3 for details). However, the proof that this approximation is indeed valid (Theorem 18 below) requires a rather different approach. In [1] we derived the corresponding result from the second moment method in combination with sharp threshold results. By contrast, here we use an indirect approach that reduces the problem of estimating the number $|\mathcal{S}_k(G(n,m))|$ of independent sets of a given size to the problem of (very accurately) estimating the independence number $\alpha(G(n,m))$. Indeed, the argument used here carries over to other problems, particularly random k-SAT, for which it yields a conceptually simpler proof than given in [1] (details omitted).

Subsequently to [1], it was shown in [35] that in many random CSPs the threshold for the shattering of the solution space into exponentially small components coincides asymptotically with the *reconstruction threshold*. Roughly speaking, the reconstruction threshold marks the onset of long-range correlations in the Gibbs measure. More precisely, it is shown in [35] that for a class of 'symmetric' random CSPs the reconstruction threshold derives from the corresponding threshold on random trees, and that it happens to coincide with the shattering threshold. Our Theorem 3 determines the threshold for shattering in the independent set problem in random graphs. Furthermore, Bhatnagar, Sly, and Tetali [5] recently studied the reconstruction problem for the independent set problem on k-regular trees. It would be most interesting to obtain a result similar to [35], namely that the reconstruction threshold on the $G(n,m)$ random graph is given by the reconstruction threshold on trees and that it coincides with the shattering threshold from Theorem 3.

The work that is perhaps most closely related to ours is a remarkable paper of Jerrum [25], who studied the Metropolis process on random graphs $G(n,m)$ with average degree $d = 2m/n > n^{2/3}$. The main result is that w.h.p. *there exists* an initial state from which the expected time for the Metropolis process to find an independent set of size $(1+\varepsilon)\frac{\ln d}{d}\cdot n$ is superpolynomial. This is quite a non-trivial achievement, as it is a result about the *initial* steps of the process where the states might potentially follow a very different distribution than the stationary distribution. The proof of this fact is via a concept called 'gateways', which is somewhat reminiscent of the expandability property in the present work. However, Jerrum's proof hinges upon the fact that the number of independent sets of size $k \sim (1+\varepsilon)\frac{\ln d}{d}\cdot n$ is concentrated about its mean. The techniques from the present work (particularly Theorem 18 below) can be used to extend Jerrum's result to the sparse case quite easily, showing that the expected time until a large independent set is found is fully exponential in n w.h.p. Yet as also pointed out in [25], an unsatisfactory aspect of this type of result is that it only shows that *there exists* a 'bad' initial state, while it seems natural to conjecture that indeed most specific initial states (such as the empty set) are 'bad'. Since we are currently unable to establish such a stronger statement, we will confine ourselves to proving an exponential lower bound on the mixing time (Theorem 7).

For *extremely* sparse random graphs, namely $d < \mathrm{e} \approx 2.718$, finding a maximum independent set in $G(n,m)$ is easy. More specifically, the greedy matching algorithm of Karp and Sipser [27] can easily be adapted so that it yields a maximum independent set w.h.p. But this approach does not generalize to average degrees $d > \mathrm{e}$ (see, however, [20] for a particular type of weighted independent sets).

Recently Rossman [38] obtained a monotone circuit lower bound for the clique problem on random graphs that is exponential in the size of the clique. The setup of [38] is somewhat orthogonal to our contribution, as we are concerned with the case that the size of the desired object (i.e., the independent set) is linear in the number of vertices, while [38] deals with the case that the size of the clique is $O(1)$ in terms of the order of the graph. Nevertheless, the punchline of viewing random graphs as a potential source of hard problems is similar.

In the course of the analysis in this paper we need a lower bound on $\alpha(G(n,m))$ which is bigger than [16]. For this reason, in [8], a previous version of this work, we slightly improved the bounds on the likely value of $\alpha(G(n,m))$ provided in [16]. The proof is similar to [16] in that it combines a "vanilla" second moment with a large deviations inequality (Talagrand's inequality, to be specific). Independently Dani and Moore [9] obtained an even better bound by means of a weighted second moment argument. Roughly speaking, they show that a $G(n,m)$ of expected degree

$$d \leq 2(n/k)(\ln(n/k)+1) - O(\sqrt{n/k})$$

has an independent set of size k w.h.p. In comparison to [9], our bound on d in [8] is

$$d \leq 2(n/k)(\ln(n/k)+1) - O(\sqrt{\ln(n/k)\cdot(n/k)}).$$

To absolve our work from the tedious second moment calculations we make direct use of the result [9].

Subsequently to the present work there have been several of related results. Gamarnik and Sudan [21] use arguments similar to the ones developed here to establish shattering in order to disprove a conjecture by Hatami, Lovász, and Szegedy [24] as to the power of certain "local algorithms" for the maximum independent set problem in random regular graphs. In addition, a new Markov Chain for the clique problem on dense random graphs has been suggested [22]. It would be interesting to see if the present techniques for lower-bounding the mixing time extend to this chain. A further somewhat related problem is that of finding a large "planted" independent set (or clique) in a random graph [3, 6, 15], for which recently a new algorithm has been put forward [10].

Furthermore, the present paper has inspired a reconsideration of the (non-rigorous) statistical physics analysis of the independent set problem on random graphs [4]. In physics, the independent set problem on random graphs is viewed as a simple model of a so-called "lattice glass" [43]. According to [4], the prior physics work suggested that this model exhibits a phenomenon called "full replica symmetry breaking" in statistical physics. By contrast, [4] predicts that for sufficiently large average degrees there occurs a simpler type of phase transition called "one-step replica symmetry breaking". This last prediction is very much in line with the rigorous results presented in the present paper. For more details on the physics perspective on random CSP we refer to [34]. In addition, based on the "one-step replica symmetry breaking" scenario, in [4] a conjecture as to the independence number of random regular graphs is put forward; this conjecture has recently been proved rigorously [11].

1.4. Organisation of the Paper

The remaining material of this work is organised as follows: For completeness, in Section 2 we provide some very elementary results, which are either known or easy to derive. In Section 3 we analyse the so-called 'planted model' to approximate the distribution $\mathcal{U}_k(n,m)$. Then in Section 4 we show Theorem 1. In Section 5 we show Theorem 3. In Section 6 we show Theorem 5. In Section 7 we show Theorem 6. In Section 8 we show Theorem 7.

2. PRELIMINARIES AND NOTATION

In this section we collect a few basic concepts and results that are either known or follow from known arguments. We will need the following Chernoff bounds on the tails of a sum of independent Bernoulli variables.

Theorem 8. *Let $I_1, I_2 \ldots, I_n$ be independent Bernoulli variables. Let $X = \sum_{i=1}^{n} I_i$ and $\mu = \mathbb{E}[X]$. Then*

$$\mathbb{P}[X < (1-\delta)\mu] \leq \exp\left(-\mu\delta^2/2\right) \quad \text{for any } 0 < \delta \leq 1, \text{ and} \tag{2}$$

$$\mathbb{P}[X > (1+\delta)\mu] \leq \exp\left(-\mu\delta^2/4\right) \quad \text{for any } 0 < \delta < 2e-1. \tag{3}$$

Also, for any $x \geq 7\mathbb{E}[X]$ it holds that

$$\mathbb{P}[X \geq x] \leq \exp(-x). \tag{4}$$

The tail bounds in (2) and (3) are from [36] while (4) is from [37], Corollary 2.4.

Let $G^*(n,m)$ be a random graph on n vertices obtained as follows: choose m pairs of vertices independently out of all n^2 possible pairs; insert the $\leq m$ edges induced by these pairs, omitting self-loops and replacing multiple edges by single edges. For technical reasons it will sometimes be easier to first work with $G^*(n,m)$ and then transfer the results to $G(n,m)$. The two distributions are related as follows.

Lemma 9. *Let $\mathcal{A}$ be any (possibly infinite) set of graphs. For any fixed $c > 0$ and $m = cn$ we have*

$$\mathbb{P}[G(n,m) \in \mathcal{A}] \leq (1 + o(1)) \exp(c + c^2) \cdot \mathbb{P}[G^*(n,m) \in \mathcal{A}]$$

The proof of Lemma 9, which is based on standard arguments, is deferred to Section 9.1.

Corollary 10. *Suppose that $m = cn$ for a fixed $c > 0$. For a graph G let $Z_k(G) = |\mathcal{S}_k(G)|$. Then for any $1 \leq k \leq 0.99n$ we have*

$$\ln \mathbb{E}[Z_k(G^*(n,m))] = \ln \mathbb{E}[Z_k(G(n,m))] + O(1).$$

For a proof of Corollary 10 see Section 9.2.

Finally, we present an estimate that will be very useful in the course of this paper.

Lemma 11 (Expectation)**.** *Let $m = dn/2$ for a real $d > 0$. Let $0 < \beta < \ln d - \ln\ln d + 1 - \ln 2$ and set*

$$k = \frac{2n}{d}\left(\ln d - \ln\ln d + 1 - \ln 2 - \beta\right) > 0.$$

If $Z_k(G)$ is the number of independent sets of size k in G, then

$$\ln \mathbb{E}[Z_k(G^*(n,m))] = k\left[\beta - \ln\left(1 - \frac{\ln\ln d - 1 + \ln 2 + \beta}{\ln d}\right) - \frac{1-\epsilon_d}{2}\frac{k}{n}\right].$$

for $\epsilon_d \to 0$ as $d \to \infty$.

Proof. Since $G^*(n,m)$ is obtained by choosing m independent pairs of vertices, we have

$$\mathbb{E}[Z_k(G^*(n,m))] = \binom{n}{k}(1 - (k/n)^2)^m. \tag{5}$$

Let $s = \frac{k}{n}$. By Stirling's formula and the fact that for $x > 0$ it holds that $\ln(1-x) = -x - \frac{x^2}{2(1-\xi)^2}$ for some $0 < \xi < x$, we get that

$$\begin{aligned}\ln\binom{n}{k} &= -n(s\ln s + (1-s)\ln(1-s)) + o(n)\\ &= ns(-\ln s + 1 - s/2 - s^2/(2(1-\xi_1)^2) + o(n) \quad [\text{where } 0 < \xi_1 < s]\end{aligned}$$

$$= k\left[\ln d - \ln\ln d - \ln 2 + 1 - \ln(1-q_d) - k/(2n) + (k/n)^2/(2(1-\xi_1)^2)\right] + o(n), \tag{6}$$

where $q_d = \frac{\ln\ln d - 1 + \ln 2 + \beta}{\ln d}$. As $m = \frac{d}{2}n$, we obtain

$$\begin{aligned}\ln(1-s^2)^m &= -dn/2\left(s^2 + s^4/(2(1-\xi_2)^2)\right) \quad [\text{where } 0 < \xi_2 < s^2] \\ &= -ns[ds/2 + ds^3/(2(1-\xi_2))^2] \\ &= -k\left(\ln d - \ln\ln d - \ln 2 + 1 - \beta + d(k/n)^3/(2(1-\xi_2)^2)\right). \end{aligned} \tag{7}$$

Note that both ξ_1, ξ_2 tend to zero with d. Combining (6) and (7) yields the assertion. ∎

We also need the following theorem from Dani and Moore [9] on the independence number of $G^*(n,m)$.

Theorem 12. *There is a constant $\alpha_0 > 0$ such that for any $x > 4/e$ and any $k \le \alpha_0 n$ the following is true. Suppose that*

$$d \le 2(n/k)(\ln(n/k) + 1) - x\sqrt{n/k}$$

and let $m = dn/2$. Then $\alpha(G^*(n,m)) \ge k$ w.h.p.

Remark. In a previous version of this work [8] we derived a slightly weaker bound on d, i.e. $d \le 2(n/k)(\ln(n/k)+1) - O(\sqrt{\ln(n/k)\cdot(n/k)})$. As opposed to the weighted second moment in [9], our approach is based on "vanilla" second moment calculations and the use of a Talagrand type inequality, i.e. similar to that in [16].

From [9] we, also, have the following corollary.

Corollary 13. *Let $W(z)$ denote the largest positive root y of the equation $ye^y = z$. W.h.p. it holds that*

$$0 \le \frac{2}{d}W\left(\frac{ed}{2}\right) - \alpha(G^*(n,m)) \le y\sqrt{\frac{\ln d}{d^3}},$$

for any constant $y > 4\sqrt{2}/e$. Expanding $W(ed/2)$ asymptotically in d we have that

$$W\left(\frac{ed}{2}\right) = \ln d - \ln\ln d + 1 - \ln 2 + \frac{\ln\ln d}{\ln d} - \frac{1-\ln 2}{\ln d} + \frac{1}{2}\left(\frac{\ln\ln d}{\ln d}\right) - (2-\ln 2)\frac{\ln\ln d}{\ln^2 d} + \frac{3+\ln^2 2 - 4\ln 2}{2\ln^2 d} + O\left(\left(\frac{\ln\ln d}{\ln d}\right)^3\right).$$

It is well known that the independence number $\alpha(G^*(n,m))$ of the random graph is tightly concentrated. More precisely, the following lower tail bound follows from a standard application of Talagrand's large deviations inequality [39], similar to the one used in [37, Section 7.1] to establish concentration for $\alpha(G(n,p))$.

Theorem 14. *Suppose that d, k are as in Theorem 12. Then for $m = \frac{dn}{2}$ and for any positive integer $t < k$ it holds that*

$$\mathbb{P}[\alpha(G^*(n,m)) < t] \leq 12\exp\left(-\frac{(k-t+1)^2}{4k}\right).$$

Proof. Consider the graph $G(n,p)$ where $p = d/n$ and let $E(G(n,p))$ denote the number of its edges. It holds that

$$\begin{aligned}\mathbb{P}[\alpha(G(n,p)) \geq k] &= \sum_{M=0}^{\binom{n}{2}} \mathbb{P}[\alpha(G^*(n,M)) \geq k]\mathbb{P}[E(G(n,p)) = M] \\ &\geq \sum_{M \leq dn/2} \mathbb{P}[\alpha(G^*(n,m)) \geq k]\mathbb{P}[E(G(n,p)) = M] \quad [\text{where } m = dn/2] \\ &\geq \mathbb{P}[\alpha(G^*(n,m)) \geq k]\mathbb{P}\left[E(G(n,p)) \leq \frac{dn}{2}\right].\end{aligned}$$

From the above derivations and Theorem 12, it is direct that

$$\mathbb{P}[\alpha(G(n,p)) \geq k] \geq \frac{1}{3}\mathbb{P}[\alpha(G^*(n,m)) \geq k] \geq 1/4. \tag{8}$$

A standard vertex exposure argument allows us to apply Talagrand's large deviation inequality for the independence number of $G(n,p)$ (in the form that appears in [37], page 41 (2.39)). The following holds:

$$\mathbb{P}[\alpha(G(n,p)) < t]\mathbb{P}[\alpha(G(n,p)) \geq k] \leq \exp\left(-(k-t+1)^2/4k\right).$$

Using (8) we get

$$\mathbb{P}[\alpha(G(n,p)) < t] \leq 4\exp\left(-(k-t+1)^2/4k\right).$$

Working as in (8) we get that $\frac{1}{3}\mathbb{P}[\alpha(G^*(n,m)) < t] \leq \mathbb{P}[\alpha(G(n,p)) < t]$. The theorem follows. ∎

Corollary 15. *For an integer $k > 0$ let*

$$\delta_k = 2(n/k)\ln(n/k) + 2(n/k) - 8\sqrt{n/k}.$$

There is a constant $\alpha_0 > 0$ such that for $k < \alpha_0 n$ and $G^*(n,m)$ of expected degree $d \leq \delta_k$ it holds that

$$\mathbb{P}[\alpha(G^*(n,m)) < k] \leq 12\exp\left(-n/(d^2\ln^5 d)\right). \tag{9}$$

Also, for $d = \delta_k$ it holds that $\mathbb{E}|\mathcal{S}_k(G^*(n,m))| \leq \exp\left(14n\sqrt{\ln^5 d/d^3}\right)$.

Proof. Let $G^*(n,m)$ be of expected degree $d = 2(n/k)(\ln(n/k)+1) - 8\sqrt{n/k}$, where k is as in the statement. Also, let k' be such that $d = 2(n/k')(\ln(n/k')+1) - 2\sqrt{n/k'}$. By Theorem 14 we have that

$$\mathbb{P}[\alpha(G^*(n,m)) < k] \leq 12\exp\left(-\frac{(k'-k+1)^2}{4k'}\right) \leq 12\exp\left(-\frac{(k'-k+1)^2}{8k}\right), \tag{10}$$

where the last inequality follows from the fact that $k' < 2k$. The tail bound in (9) will follow by bounding appropriately $t = k' - k > 0$. We bound t by using the fact that

$$2(n/k)(\ln(n/k) + 1) - 8\sqrt{n/k} = 2(n/k')(\ln(n/k') + 1) - 2\sqrt{n/k'}.$$

Set $s = k/n$ and $q = t/k$. Let $h(s, q)$ be the difference of the l.h.s. minus r.h.s. in the above equality, written in terms of s, t. Clearly, it holds that that $h(s, q) = 0$. That is

$$h(s,q) = \frac{2\ln(1+q)}{s} + \frac{q}{1+q}(-\ln s - \ln(1+q) + 1) - \frac{2}{\sqrt{s}}\left(4 - \frac{1}{\sqrt{1+q}}\right) = 0.$$

For $1.5n \ln d/d < k, k' < 2n \ln d/d$, it is direct to verify that for $q = 10/\sqrt{d \ln^5 d}$ and sufficiently small s it holds that $h(s, q) < 0$. Furthermore, it is easy to see that

$$\frac{\partial}{\partial q}h(s,q) = \frac{2}{s(1+q)} + \frac{1}{(1+q)^2}(-\ln s - \ln(1+q) + 1 - q) - \frac{1}{\sqrt{s(1+q)^3}}.$$

For any $q \in [0, 1]$ and sufficiently small s we have that $\frac{\partial}{\partial q}h(s, q) > 0$. This entails that for any $q \le 10/\sqrt{d \ln^5 d}$ and sufficiently small s we have $h(s, q) < 0$. Thus, we get that $k' - k \ge 10k/\sqrt{d \ln^5 d}$. Plugging this into (10) we get that

$$\begin{aligned}\mathbb{P}[\alpha(G^*(n,m)) < k] &\le 12\exp\left(-\frac{100}{8}\frac{k}{d\ln^5 d}\right)\\ &\le 12\exp\left(-\frac{300}{16}\frac{n}{d^2\ln^4 d}\right), \quad [\text{as } k \ge 1.5n\ln d/d]\end{aligned}$$

which implies (9).

For the rest of the proof, consider $G^*(n, m)$ with expected degree $d = \delta_k$. Assume that we add to $G^*(n, m)$ edges at random so as to increase the expected degree to $d^+ = 2\frac{s\ln s + (1-s)\ln(1-s)}{\ln(1-s^2)}$ and get the graph $G^*(n, m')$. That is, we need to insert into $G^*(n, m)$ as many as $(d^+ - d)n/2$ random edges. Therefore, each independent set of size k in $G^*(n, m)$ is also an independent set of $G^*(n, m')$ with probability $\left(1 - (k/n)^2\right)^{(d^+-d)n/2}$. Let $s = (k/n)$. It is direct that

$$\mathbb{E}|\mathcal{S}_k(G(n,m'))| = \left(1 - s^2\right)^{(d^+-d)n/2}\mathbb{E}[|\mathcal{S}_k(G(n,m))|]. \tag{11}$$

Using Corollary 10 we get that

$$\begin{aligned}\frac{1}{n}\ln\mathbb{E}|\mathcal{S}_k(G(n,m'))| &= \frac{1}{n}\ln\left(\binom{n}{k}(1-(k/n)^2)^{d^+n/2}\right) + O\left(\frac{1}{n}\right)\\ &\sim -[s\ln s + (1-s)\ln(1-s)] + d^+\ln(1-s^2)/2 - \frac{\ln n}{2n}\\ &\sim -\frac{\ln n}{2n}.\end{aligned} \tag{12}$$

Furthermore, using the fact that $-\frac{x}{1-x} \le \ln(1-x) \le -x$, for $0 < x < 1$, it is direct that

$$2\frac{-\ln s + 1}{s} \le d^+ \le 2\frac{-\ln s + 1}{s} + 2. \tag{13}$$

Combining (11), (12) and (13), we get that

$$\frac{1}{n}\ln \mathbb{E}|\mathcal{S}_k(G(n,m))| \leq -\ln(1-s^2)(d^+-d)/2 - o(1) \quad \text{[by (11) and (12)]}$$
$$\leq 4\frac{s^{3/2}}{1-s^2} \quad \text{[by (13) and } 1-x > e^{-x/(1-x)} \text{ for } 0 < x < 1\text{]}.$$

The upper bound for $\mathbb{E}|\mathcal{S}_k(G(n,m))|$ follows by using the above inequality and noting that $k \leq 2n\ln d/d$, i.e. $s \leq 2\ln d/d$. ∎

Corollary 16. *For the graph $G(n,m)$ of expected degree d it holds that*

$$\mathbb{P}\left[\alpha(G(n,m)) \geq 2n(1-\epsilon_d)\ln d/d\right] \geq 1 - \exp\left[-8n/(d\ln^3 d)\right].$$

where $\epsilon_d \to 0$ as d increases.

Proof. Consider $G^*(n,m)$ of expected degree d and let k be such that

$$k/n = \frac{2}{d}\left(W(ed/2) - 10\sqrt{\ln d/d^3} - 2\frac{\ln\ln d}{\ln d}\right),$$

where $W(z)$ is defined in the statement of Corollary 13. Using Corollary 13 and Theorem 14, we get that

$$\mathbb{P}\left[\alpha(G^*(n,m)) \leq k\right] \leq \exp\left(-\frac{8(\ln\ln d)^2}{d\ln^3 d}n\right).$$

The corollary follows by using Lemma 9. ∎

The following is taken from [37, p. 156].

Lemma 17. *Let $d > 0$ be fixed and $m = dn/2$. Let Y be the number of isolated vertices in $G(n,m)$. Then $Y = (1+o(1))n\exp(-d)$ w.h.p.*

3. APPROACHING THE DISTRIBUTION $\mathcal{U}_k(n,m)$

3.1. The Planted Model

The main results of this paper deal with properties of 'typical' independent sets of a given size in a random graph, i.e., the probability distribution $\mathcal{U}_k(n,m)$. In the theory of random discrete structures often the conceptual difficulty of analysing a probability distribution is closely linked to the computational difficulty of sampling from that distribution (e.g., [37, Chapter 9]). This could suggest that analysing $\mathcal{U}_k(n,m)$ is a formidable task, because for $k > (1+\varepsilon)n\ln(d)/d$ there is no efficient procedure known for finding an independent set of size k in a random graph $G(n,m)$, let alone for sampling one at random. In effect, we do not know of an efficient method for sampling from $\mathcal{U}_k(n,m)$.

To get around this problem, we are going to 'approximate' the distribution $\mathcal{U}_k(n,m)$ by another distribution $\mathcal{P}_k(n,m)$ on the set $\Lambda_k(n,m)$ of graph/independent set pairs, the

so-called planted model, which is easy to sample from. This distribution is induced by the following experiment:

Choose a subset $\sigma \subset [n]$ of size k uniformly at random.
Choose a graph G with m edges *in which* σ *is an independent set* uniformly at random.
Output the pair (G, σ).

In other words, the probability assigned to a given pair $(G_0, \sigma_0) \in \Lambda_k(n, m)$ is

$$\mathbb{P}_{\mathcal{P}_k(n,m)}[(G_0, \sigma_0)] = \left[\binom{n}{k} \cdot \binom{\binom{n}{2} - \binom{k}{2}}{m}\right]^{-1}, \tag{14}$$

i.e., $\mathcal{P}_k(n, m)$ is nothing but the uniform distribution on $\Lambda_k(n, m)$. The key result that allows us to study the distribution $\mathcal{U}_k(n, m)$ is the following.

Theorem 18. *There is* $\varepsilon_d \to 0$ *such that for* $k < (2 - \varepsilon_d)n \ln(d)/d$ *the following is true. If* $\mathcal{B}$ *is an event such that*

$$\mathbb{P}_{\mathcal{P}_k(n,m)}[\mathcal{B}] = o\left(\exp\left(-14n\sqrt{\ln^5 d/d^3}\right)\right), \tag{15}$$

then $\mathbb{P}_{\mathcal{U}_k(n,m)}[\mathcal{B}] = o(1)$.

Hence, Theorem 18 allows us to bound the probability of some 'bad' event $\mathcal{B}$ in the distribution $\mathcal{U}_k(n, m)$ by bounding its probability in the distribution $\mathcal{P}_k(n, m)$.

To establish Theorem 18, we need to find a way to compare $\mathcal{P}_k(n, m)$ and $\mathcal{U}_k(n, m)$. Suppose that $k < (2-\varepsilon_d)n \ln(d)/d$ is such that $\alpha(G(n, m)) \geq k$ w.h.p. Then the probability of a pair $(G_0, \sigma_0) \in \Lambda_k(n, m)$ under the distribution $\mathcal{U}_k(n, m)$ is

$$\mathbb{P}_{\mathcal{U}_k(n,m)}[(G_0, \sigma_0)] \sim \left[\binom{\binom{n}{2}}{m} \cdot |\mathcal{S}_k(G_0)|\right]^{-1} \tag{16}$$

(because we first choose a graph uniformly, and then an independent set of that graph). Hence, the probabilities assigned to (G_0, σ_0) under (16) and (14) coincide (asymptotically) iff

$$|\mathcal{S}_k(G_0)| \sim \binom{n}{k}\binom{\binom{n}{2} - \binom{k}{2}}{m} \Big/ \binom{\binom{n}{2}}{m}. \tag{17}$$

A moment's reflection shows that the expression on the r.h.s. of (17) is precisely the *expected* number $\mathbb{E}|\mathcal{S}_k(G(n, m))|$ of independent sets of size k. Thus, $\mathcal{P}_k(n, m)$ and $\mathcal{U}_k(n, m)$ coincide asymptotically iff the number $|\mathcal{S}_k(G(n, m))|$ of independent sets of size k is concentrated about its expectation.

This is indeed the case in 'dense' random graphs with $m \gg n^{3/2}$. For this regime one can perform a 'second moment' computation to show that $|\mathcal{S}_k(G(n, m))| \sim \mathbb{E}|\mathcal{S}_k(G(n, m))|$ w.h.p., (e.g. see [37, Chapter 7]) whence the measures $\mathcal{P}_k(n, m)$ and $\mathcal{U}_k(n, m)$ are interchangeable. This fact forms (somewhat implicitly) the foundation of the proofs in [25].

By contrast, in the sparse case $m \ll n^{3/2}$ a straight second moment argument fails utterly. As it turns out, this is because the quantity $|\mathcal{S}_k(G(n, m))|$ simply it not concentrated about its expectation anymore. In fact, maybe somewhat surprisingly Theorem 18 can be used

to infer the following corollary, which shows that in sparse random graphs the expectation $\mathbb{E}|\mathcal{S}_k(G(n,m))|$ 'overestimates' the typical number of independent sets by an exponential factor w.h.p.

Corollary 19. *There exist functions $\varepsilon_d \to 0$ and $g(d) > 0$ such that for $10n/d < k < (2-\varepsilon_d)n\ln(d)/d$ we have*

$$|\mathcal{S}_k(G(n,m))| \leq \mathbb{E}|\mathcal{S}_k(G(n,m))| \cdot \exp(-g(d)n) \quad \text{w.h.p.}$$

The proof of Corollary 19 appears in Section 3.3.

Conversely, in order to prove Theorem 18 we need to bound the 'gap' between the typical value of $|\mathcal{S}_k(G(n,m))|$ and its expectation from above. This estimate can be summarized as follows.

Proposition 20. *There is $\varepsilon_d \to 0$ such that for $k < (2-\varepsilon_d)n\ln(d)/d$ we have*

$$|\mathcal{S}_k(G(n,m))| \geq \mathbb{E}|\mathcal{S}_k(G(n,m))| \cdot \exp\left(-14n\sqrt{\ln^5 d/d^3}\right)$$

with probability at least $1-\exp\left[-n/(2d^2\ln^4 d)\right]$.

Before we prove Proposition 20 in Section 3.2, let us indicate how it implies Theorem 18.

Corollary 21. *There is $\varepsilon_d \to 0$ such that for $k < (2-\varepsilon_d)n\ln(d)/d$ the following is true. Let*

$$\mathcal{Z} = \left\{(G,\sigma) \in \Lambda_k(n,m) : |\mathcal{S}_k(G)| \geq \mathbb{E}|\mathcal{S}_k(G(n,m))| \cdot \exp\left(-14n\sqrt{\ln^5 d/d^3}\right)\right\}. \quad (18)$$

Then $\mathbb{P}_{\mathcal{U}_k(n,m)}[\mathcal{Z}] = 1-o(1)$, and for any event $\mathcal{B} \subset \Lambda_k(n,m)$ we have

$$\mathbb{P}_{\mathcal{U}_k(n,m)}[\mathcal{B}|\mathcal{Z}] \leq (1+o(1))\exp\left(14n\sqrt{\ln^5 d/d^3}\right) \cdot \mathbb{P}_{\mathcal{P}_k(n,m)}[\mathcal{B}].$$

Proof. Proposition 20 directly implies that

$$\mathbb{P}_{\mathcal{U}_k(n,m)}[\mathcal{Z}] = 1-o(1). \quad (19)$$

Furthermore, by the definition (16) of the distribution $\mathcal{U}_k(n,m)$,

$$\begin{aligned}
\mathbb{P}_{\mathcal{U}_k(n,m)}[\mathcal{B}\cap\mathcal{Z}] &= \sum_{(G,\sigma)\in\mathcal{B}\cap\mathcal{Z}} \left[\binom{\binom{n}{2}}{m}|\mathcal{S}_k(G)|\right]^{-1} \\
&\leq \exp\left[14n\sqrt{\ln^5 d/d^3}\right] \sum_{(G,\sigma)\in\mathcal{B}\cap\mathcal{Z}} \left[\binom{\binom{n}{2}}{m}\mathbb{E}|\mathcal{S}_k(G(n,m))|\right]^{-1} \quad \text{[by (18)]} \\
&= \exp\left[14n\sqrt{\ln^5 d/d^3}\right] \mathbb{P}_{\mathcal{P}_k(n,m)}[\mathcal{B}\cap\mathcal{Z}] \quad \text{[by (14)]} \\
&\leq \exp\left[14n\sqrt{\ln^5 d/d^3}\right] \mathbb{P}_{\mathcal{P}_k(n,m)}[\mathcal{B}]. \quad (20)
\end{aligned}$$

The assertion is immediate from (19) and (20). ∎

Proof of Theorem 18. The theorem follows directly from Corollary 21. ∎

3.2. Proof of Proposition 20

Since the second moment method fails to yield a lower bound on the typical number of independent sets $|\mathcal{S}_k(G(n,m))|$, we need to invent a less direct approach to prove Proposition 20. Of course, the demise of the second moment argument also presented an obstacle to Frieze [16] in his proof that

$$\alpha(G(n,m)) \geq (2-\varepsilon_d)n\ln(d)/d \quad \text{w.h.p.} \tag{21}$$

However, unlike the *number* $|\mathcal{S}_k(G(n,m))|$ of independent sets $\alpha(G(n,m))$, the *size* of the largest one actually is concentrated about its expectation. In fact, an arsenal of large deviations inequalities applies (e.g., Azuma's and Talagrand's inequality), and [16] uses these to bridge the gap left by the second moment argument.Unfortunately, these large deviations inequalities draw a blank on $|\mathcal{S}_k(G(n,m))|$. Therefore, we are going to derive the desired lower bound on $|\mathcal{S}_k(G(n,m))|$ directly from (21).

To simplify our derivations we consider the model of random graphs $G^*(n,m)$ and we show the following proposition.

Proposition 22. *There is $\varepsilon_d \to 0$ such that for $k < (2-\varepsilon_d)n\ln(d)/d$ we have*

$$|\mathcal{S}_k(G^*(n,m))| \geq \mathbb{E}|\mathcal{S}_k(G^*(n,m))| \cdot \exp\left(-14n\sqrt{\ln^5 d/d^3}\right) \tag{22}$$

with probability at least $1-\exp\left[-n/(d\ln^2 d)^2\right]$.

Then, Proposition 20 follows by Lemmas 9 and 10.

Given some integer $k>0$ and $q\in[0,1]$, let $Z_k(G^*(n,m)) = |\mathcal{S}_k(G^*(n,m))|$ and let

$$M_k^q = \max\{m\in\mathbb{N} : \mathbb{P}[Z_k(G^*(n,m))>0] \geq 1-q\}.$$

In words, M_k^q is the largest number of edges that we can squeeze in while keeping the probability that $G^*(n,m)$ has an independent set of size k above $1-q$. The following lemma summarizes the key step of our proof of Proposition 22. The idea is that Lemma 23 gives a tradeoff between the *likely* number of independent sets of size k in the random graph with $m < M_k^q$ edges and the *expected* number of such independent sets in the random graph with M_k^q edges. More precisely, we show that it is very unlikely that the number of independent sets at 'time' m is smaller than its expectation by much more than a factor of $\mathbb{E}[Z_k(G^*(n,M_k^q))]$.

Lemma 23. *Suppose that $k,m>0, q\in[0,1]$ are such that $m<M_k^q$. Then*

$$\mathbb{P}\left[Z_k(G^*(n,m)) < \frac{\mathbb{E}[Z_k(G^*(n,m))}{2\mathbb{E}[Z_k(G^*(n,M_k^q))]}\right] \leq 2q.$$

Proof. Let $M = M_k^q$. The random graph $G^*(n,M)$ is obtained by choosing M pairs of vertices independently and inserting the corresponding edges (while omitting loops and reducing multiple edges to single edges). Let us think of the M pairs as being generated in two rounds. In the first round, we generate m pairs, which induce the random graph $G_1 = G^*(n,m)$. In the second round, we choose a further $M-m$ pairs independently and

add the corresponding edges to G_1 (again, omitting self-loops and reducing multiple edges to single edges) to obtain $G_2 = G^*(n, M)$.

By the linearity of the expectation and because the m (resp. M) pairs that the random graph G_1 (resp. G_2) consists of are chosen independently, we have (cf. (5))

$$\mathbb{E}[Z_k(G_1)] = \binom{n}{k}(1-(k/n)^2)^m, \quad \text{and}$$
$$\mathbb{E}[Z_k(G_2)] = \binom{n}{k}(1-(k/n)^2)^M = \mathbb{E}[Z_k(G_1)] \cdot (1-(k/n)^2)^{M-m}. \tag{23}$$

Furthermore, with respect to the number of independent sets of size k in G_2 *given* their number in the outcome G_1 of the 'first round', we have

$$\mathbb{E}[Z_k(G_2)|Z_k(G_1)] = Z_k(G_1)(1-(k/n)^2)^{M-m}. \tag{24}$$

Indeed, for each independent set Q of size k in G_1 each of the $M-m$ additional random pairs has its two vertices in Q with probability $(k/n)^2$. Hence, (24) follows because these $M-m$ pairs are independent and by the linearity of the expectation.

Now, let $\mathcal{E}_1$ be the event that

$$Z_k(G_1) < \frac{\mathbb{E}[Z_k(G_1)]}{2\mathbb{E}[Z_k(G_2)]}.$$

Then by Markov's inequality and (24),

$$\frac{1}{2} \le \mathbb{P}\left[Z_k(G_2) < 2\mathbb{E}[Z_k(G_2)|\mathcal{E}_1]\,\middle|\,\mathcal{E}_1\right] \le \mathbb{P}\left[Z_k(G_2) < \frac{\mathbb{E}[Z_k(G_1)]\cdot(1-(k/n)^2)^{M-m}}{\mathbb{E}[Z_k(G_2)]}\,\middle|\,\mathcal{E}_1\right],$$

whence

$$\mathbb{P}\left[Z_k(G_2) < \frac{\mathbb{E}[Z_k(G_1)]\cdot(1-(k/n)^2)^{M-m}}{\mathbb{E}[Z_k(G_2)]}\right] \ge \mathbb{P}[\mathcal{E}_1]/2. \tag{25}$$

Combining (25) and (23), we see that $\mathbb{P}[\mathcal{E}_1] \le 2\,\mathbb{P}[Z_k(G_2) < 1] \le 2q$, as claimed. ∎

Proof of Proposition 22. Consider $G^*(n,m)$ of expected degree d and let $k = \frac{2}{d}(\ln d - \ln\ln d + 1 - \ln 2)$. We are going to show that (22) holds for $G^*(n,m)$ and k with probability at least $1 - \exp\left[-n/(d\ln^2 d)^2\right]$.

Consider, now, the graph $G(n,M)$ of expected degree $d^+ = 2\frac{-\ln s+1}{s} + \frac{8}{\sqrt{s}}$, where $s = k/n$. According to Corollary 15 it holds that $\mathbb{P}[|S_k(G(n,M))| > 0] \ge 1 - 12\exp\left(-n/(d^2\ln^5 d)\right)$ and

$$\mathbb{E}|S_k(G(n,M))| \le \exp\left(14\sqrt{\frac{\ln^5 d}{d^3}}n\right).$$

The proposition will follow by just showing that $m < M$, i.e. $d^+ > d$, and using Lemma 23. Note, first, that

$$\begin{aligned} -\ln s + 1 &= \ln d - \ln\ln d + 1 - \ln 2 - \ln\left(1 - \frac{\ln\ln d - 1 + \ln 2}{\ln d}\right) \\ &\ge \ln d - \ln\ln d + 1 - \ln 2 + \frac{\ln\ln d - 1 + \ln 2}{\ln d}. \quad [\text{as } 1 - x \le e^{-x}] \end{aligned}$$

Using the above, it is elementary to derive that $2\frac{-\ln s+1}{s} \geq d$. Then, it follows that $d^+ > d$ as promised. ∎

3.3. Proof of Corollary 19

In this section we keep the assumptions of Corollary 19, i.e., we let k, d be such that $10n/d < k < (2-\varepsilon_d)n\ln(d)/d$, with $\varepsilon_d \to 0$ sufficiently slowly in the limit of large d.

First we are going to show that for in a pair (G,σ) chosen from the distribution $\mathcal{P}_k(n,m)$, the number of isolated vertices are somehow, exceedingly many, compared to those in $G(n,m)$

Lemma 24. *There exist numbers $\xi > 0$ and $\eta = \eta(d) > 0$ such that the following is true. Let (G,σ) be a pair chosen from the distribution $\mathcal{P}_k(n,m)$. Let X be the number of isolated vertices in G. Then*

$$\mathbb{P}\left[X \leq n(\eta + \exp(-d))\right] \leq \exp(-3\xi n). \tag{26}$$

Proof. Let $\alpha = k/n$. It is convenient to first consider the following variant of the planted distribution: given a set $\sigma \subset V$ of size k, let G' be the random graph obtained by including each of the $\binom{n}{2} - \binom{k}{2}$ possible edges that do not link two vertices in σ with probability

$$q = \frac{m}{\binom{n}{2} - \binom{k}{2}} \sim \frac{m}{\binom{n}{2}(1-\alpha^2)} \sim \frac{d}{n(1-\alpha^2)}$$

independently. Hence, the total number of edges in G' is binomially distributed with mean m. By Stirling's formula, the event $\mathcal{E}$ that G' has precisely m edges has probability $\Theta(m^{-1/2})$, and given that $\mathcal{E}$ occurs, the pair (G',σ) has the same distribution as the pair (G,σ) chosen from the distribution $\mathcal{P}_k(n,m)$. Therefore, for any event $\mathcal{A}$ we have

$$\mathbb{P}\left[(G,\sigma) \in \mathcal{A}\right] = \mathbb{P}\left[(G',\sigma) \in \mathcal{A} | \mathcal{E}\right] \leq O(\sqrt{m}) \cdot \mathbb{P}\left[(G',\sigma) \in \mathcal{A}\right]. \tag{27}$$

Now, consider the number X' of vertices in σ that are isolated in G'. Since each possible edge is present in G' with probability q independently, the degree of each vertex $v \in \sigma$ has a binomial distribution $\mathrm{Bin}(n-k,q)$ with mean

$$q(1-\alpha)n = d \cdot \frac{1-\alpha}{1-\alpha^2} = \frac{d}{1+\alpha}.$$

In particular, for each $v \in \sigma$ we have

$$\mathbb{P}\left[v \text{ is isolated in } G'\right] \sim \exp(-d/(1+\alpha)).$$

Hence,

$$\mathbb{E}\left[X'\right] \sim \alpha n \exp(-d/(1+\alpha)). \tag{28}$$

In addition, let X'' be the number of isolated vertices in $V \setminus \sigma$. Since for each $v \in X''$ the expected number of neighbors is $(n-1)q \sim d/(1-\alpha^2)$, we have

$$\mathbb{E}\left[X''\right] \sim (1-\alpha)n \exp(-d/(1-\alpha^2)). \tag{29}$$

Combining (28) and (29), we see that

$$n^{-1}\mathbb{E}[X] \sim \alpha \exp(-d/(1+\alpha)) + (1-\alpha)\exp(-d/(1-\alpha^2)). \tag{30}$$

If $\alpha \geq 10/d$ and d sufficiently large, then there is $\eta = \eta(d) > 0$ such that

$$\alpha \exp(-d/(1+\alpha)) + (1-\alpha)\exp(-d/(1-\alpha^2)) > \exp(-d) + 3\eta.$$

Hence, (31) yields

$$\mathbb{E}[X] > n(\exp(-d) + 2\eta). \tag{31}$$

Finally, the assertion follows from (31) and a standard application of Azuma's inequality. ∎

Proof of Corollary 19. Let $\mathcal{B} \subset \Lambda_k(n,m)$ be the set of all pairs (G,σ) such that G has fewer than $n(\eta + \exp(-d))$ isolated vertices. Lemmas 17 and 24 entail that

$$\mathbb{P}_{\mathcal{U}_k(n,m)}[\mathcal{B}] = 1 - o(1) \text{ while } \mathbb{P}_{\mathcal{P}_k(n,m)}[\mathcal{B}] \leq \exp(-\xi n). \tag{32}$$

Since $\mathcal{P}_k(n,m)$ is the uniform distribution over $\Lambda_k(n,m)$, (32) implies that

$$|\mathcal{B}| \leq |\Lambda_k(n,m)| \cdot \exp(-\xi n) = \binom{\binom{n}{2}}{m}\mathbb{E}|\mathcal{S}_k(G(n,m))| \cdot \exp(-\xi n). \tag{33}$$

Now, let $\mathcal{A} \subset \Lambda_k(n,m)$ be the set of all pairs (G,σ) such that $|\mathcal{S}_k(G)| \geq \exp(-\xi n/3)\mathbb{E}|\mathcal{S}_k(G(n,m))|$, and assume for contradiction that there is a fixed $\varepsilon > 0$ such that $\mathbb{P}_{\mathcal{U}_k(n,m)}[\mathcal{A}] \geq \varepsilon$. Then (32) implies that

$$\mathbb{P}_{\mathcal{U}_k(n,m)}[\mathcal{A} \cap \mathcal{B}] \geq \varepsilon - o(1)$$

Therefore,

$$\begin{aligned}|\mathcal{B}| \geq |\mathcal{A}\cap\mathcal{B}| &\geq \binom{\binom{n}{2}}{m}\mathbb{P}_{\mathcal{U}_k(n,m)}[\mathcal{A}\cap\mathcal{B}] \cdot \exp(-\xi n/3)\mathbb{E}|\mathcal{S}_k(G(n,m))| \\ &\geq (\varepsilon - o(1))\binom{\binom{n}{2}}{m}\exp(-\xi n/3)\mathbb{E}|\mathcal{S}_k(G(n,m))| \geq (\varepsilon - o(1))\exp(-\xi n/3) \cdot |\Lambda_k(n,m)|,\end{aligned}$$

which contradicts (33). Hence, $\mathbb{P}_{\mathcal{U}_k(n,m)}[\mathcal{A}] = o(1)$, as claimed. ∎

4. PROOF OF THEOREM 1

Instead of the random graph model $G(n,m)$ we consider the model $G(n,p)$, where $p = d/n$ for fixed real d and we prove the following theorem.

Theorem 25. *There is $\varepsilon_d \to 0$ such that $\mathcal{S}_k(G(n,d/n))$ is $20d$-connected for any $k \leq (1-\varepsilon_d)\frac{\ln d}{d} \cdot n$, with probability at least $1 - \exp\left(-\frac{\ln^{40} d}{d}n\right)$.*

Theorem 1 follows by using standard arguments, i.e. the following corollary.

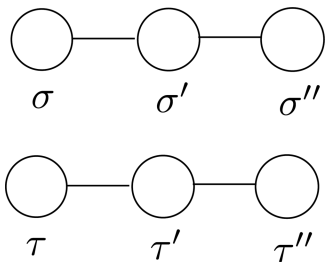

Fig. 1. Short chains of adjacent independent sets. E.g. σ' is adjacent to both σ and σ''.

Corollary 26. *For any fixed $d > 0$, $m = dn/2$ and any graph property A it holds that* $\mathbb{P}[G(n,m) \in A] \leq \Theta(\sqrt{n})\mathbb{P}[G(n,d/n) \in A]$.

Proof. Let E_d be the number of edges in $G(n,d/n)$. It holds that

$$\mathbb{P}[G(n,m) \in A] = \mathbb{P}[G(n,d/n) \in A | E_d = dn/2] \leq \frac{\mathbb{P}[G(n,d/n) \in A]}{\mathbb{P}[E_d = dn/2]}.$$

E_d is binomially distributed with parameters $\binom{n}{2}$ and d/n. Straightforward calculations yield that $\mathbb{P}[E_d = dn/2] = \Theta(1/\sqrt{n})$. The corollary follows. ∎

For every vertex u in $G(n,d/n)$ we let $N(u)$ denote the set of vertices which are adjacent to u. A sufficient condition for establishing the connectivity of $\mathcal{S}_k(G(n,d/n))$ is requiring this space to have what we call Property Γ:

Property Γ. For any two $\sigma, \tau \in \mathcal{S}_k(G(n,d/n))$ there exist "chains" $\sigma, \sigma', \sigma''$ and τ, τ', τ'' of independent sets in $\mathcal{S}_k(G(n,d/n)) \cup \mathcal{S}_{k+1}(G(n,d/n))$ such that

- the independent sets are connected as in Fig. 1; i.e.,

$$\operatorname{dist}(\sigma,\sigma'), \operatorname{dist}(\sigma',\sigma''), \operatorname{dist}(\tau,\tau'), \operatorname{dist}(\tau',\tau'') \leq 20d.$$

- $|\sigma''| = |\tau''| = k$,
- $\operatorname{dist}(\sigma'',\tau'') < \operatorname{dist}(\sigma,\tau)$, and thus $|\sigma'' \cap \tau''| = |\sigma \cap \tau| + 1$.

The following result is straightforward.

Corollary 27. *If $\mathcal{S}_k(G(n,d/n))$ has Property Γ, then it is connected.*

Using Corollary 27, Theorem 25 will follow by showing that with probability $1-o(1)$ the set $\mathcal{S}_k(G(n,d/n))$ has Property Γ, for $k < (1-\epsilon_d)n \ln d/d$. For this, we need to introduce the notion of "augmenting vertex".

Definition 28 (Augmenting vertex)**.** *For the pair $\sigma, \tau \in \mathcal{S}_k(G(n,d/n))$ the vertex $v \in V \backslash (\sigma \cup \tau)$ is* augmenting *if one of the following A, B holds.*

A. $N(v) \cap (\sigma \cup \tau) = \emptyset$
B. $N(v) \cap (\sigma \cap \tau) = \emptyset$ *and there are sets $I_v(\sigma)$ and $I_v(\tau)$ of size at most $7d$ such that*

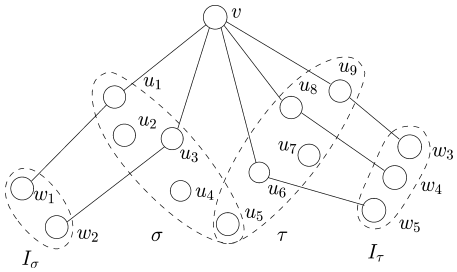

Fig. 2. σ, τ with Property Γ.

- $I_v(\sigma) \cup \{v\}$ *is an independent set of* $G(n, d/n)$
- $|I_v(\sigma)| = |N(v) \cap \sigma|$
- *each* $w \in I_v(\sigma)$ *has a unique neighbour in* σ *which is a neighbour of the augmenting vertex* v. *In symbols,*

$$\forall w \in I_v(\sigma) : |N(w) \cap \sigma| = 1 \wedge |N(w) \cap N(v) \cap \sigma| = 1.$$

The corresponding conditions hold for $I_v(\tau)$, *as well. We refer to the vertices in* $I_v(\sigma), I_v(\tau)$ *as* terminals.

Figure 2 shows an example of a pair of independent sets where the vertex v is an *augmenting vertex*. We emphasize that in **B.** we require that $N(v)$ does not share a vertex with the *intersection* $\sigma \cap \tau$, while **A.** requires that $N(v)$ does not contain a vertex from the *union* $\sigma \cup \tau$.

We will show that for a pair $\sigma, \tau \in \mathcal{S}_k(G(n, d/n))$ that has an *augmenting vertex* v we can find short chains $\sigma, \sigma', \sigma''$ and τ, τ', τ''. That is, if we can find an augmenting vertex for any two members of $\mathcal{S}_k(G(n, d/n))$, then $\mathcal{S}_k(G(n, d/n))$ has Property Γ.

First, let us show how we can create short chains as in Figure 1 for two independent sets σ, τ with augmenting vertex v. For this, we introduce a process called *Collider*. This process takes as an input σ, τ and the augmenting vertex v and returns the independent sets σ'' and τ'' of the chains.

Collider (σ, τ, v)**:**

Phase 1. /*Creation of σ' and τ'.*/

1. Derive σ' from σ by removing the all its vertices in $N(v) \cap \sigma$ and by inserting $\{v\} \cup I_v(\sigma)$.
2. Do the same for τ'.

Phase 2. /* Creation of σ'' and τ''*/.

1. σ'' is derived from σ' by deleting one (any) vertex from $\sigma' \setminus \tau'$.
2. τ'' is derived from τ' by deleting one (any) vertex from $\tau' \setminus \sigma'$.

Return σ'' and τ''.

End

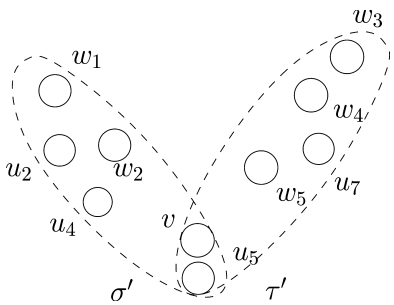

Fig. 3. The independent sets σ', τ'.

Figure 3 shows the changes that have taken place to the independent sets in Fig. 2 at the end of Phase 1. Note that after Phase 1 both σ', τ' contain the augmenting vertex v, i.e. the overlap has increased as $\sigma' \cap \tau' = (\sigma \cap \tau) \cup \{v\}$. After Phase 2, the independent sets in Fig. 3 are transformed to those in Fig. 4. There the vertices u_2 and u_7 are removed from σ' and τ', correspondingly.

In the following lemma we show that *Collider* has all the desired properties we promised above.

Fact 29. *Let $\sigma, \tau \in \mathcal{S}_k(G)$ with augmenting vertex v. Let σ'' and τ'' be the two sets of vertices that are returned from* Collider(σ, τ, v) *. The two sets have the following properties:*

1. $\sigma'', \tau'' \in \mathcal{S}_k(G)$,
2. $|\sigma'' \cap \tau''| = |\sigma \cap \tau| + 1$,
3. *There are $\sigma', \tau' \in \mathcal{S}_{k+1}(G)$ such that σ' (resp. τ') is adjacent to both σ and σ'' (resp. τ and τ').*

Fact 29 is immediate from Figs. 2–4.

Since for every pair $\sigma, \tau \in \mathcal{S}_k(G(n, d/n))$ with augmenting vertex we can construct short chains as in Fig. 1 by using Collider, we have the following corollary:

Corollary 30. *If for any two $\sigma, \tau \in \mathcal{S}_k(G)$ there is an augmenting vertex v, then $\mathcal{S}_k(G)$ has Property Γ.*

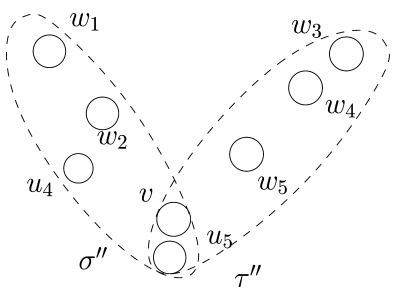

Fig. 4. Final sets.

We are going to use the first moment method to show that with probability $1 - o(1)$, the graph $G(n, d/n)$ has no pair of independent sets in $\mathcal{S}_k(G(n, d/n))$ without augmenting vertex. Corollary 30, then, implies that with probability $1 - o(1)$ the set $\mathcal{S}_k(G(n, d/n))$ has Property Γ. Then, Theorem 25 follows from Corollary 27.

We compute, first, the conditional probability that σ, τ have an augmenting vertex given that $\sigma, \tau \in \mathcal{S}_k(G(n, d/n))$.

Proposition 31. *For some integers i, k, consider σ, τ, two sets of vertices each of size k such that $|\sigma \cap \tau| = i$. Let $G_{\sigma,\tau}$ denote $G(n, d/n)$ conditional that each of σ, τ is an independent set. Also, let $p_{k,i}$ be the probability that the pair σ, τ has an augmenting vertex in $G_{\sigma,\tau}$. Then, there exists $\epsilon_d \to 0$ such that for any $\epsilon_d \le \epsilon \le 1 - \epsilon_d$ and $k = (1 - \epsilon)\frac{\ln d}{d} n$ the following is true*

$$p_{k,i} \ge 1 - \exp\left(-\frac{\ln^{90} d}{d} n\right) \quad \text{for any } i \le k.$$

Observe that the lower bound we get for $p_{k,i}$ is independent of i. The proof of Proposition 31 appears in Section 4.1.

Proof of Theorem 25. Let Ψ_k be the number of pairs of independent sets of size k in $G(n, d/n)$ that do not have an augmenting vertex. From Corollary 30 and Corollary 27, it suffices to show that $\mathbb{P}\left[\sum_{k \le K} \Psi_k > 0\right] = o(1)$, where $K = (1-\epsilon_d) n \ln d/d$ and $\epsilon_d \to 0$ with d. For this, we are going to use Markov's inequality, i.e. $\mathbb{P}\left[\sum_{k \le K} \Psi_k > 0\right] \le \mathbb{E}\left[\sum_{k \le K} \Psi_k\right]$ and we are going to show that $\mathbb{E}\left[\sum_{k \le K} \Psi_k\right] = o(1)$.

First consider the case where $\frac{1}{10}\frac{\ln d}{d} n \le k \le (1 - \epsilon_d)\frac{\ln d}{d} n$ and ϵ_d is as defined in the statement of Proposition 31. Using Proposition 31 we get that

$$\mathbb{E}[\Psi_k] \le \binom{n}{k}^2 \exp\left(-\frac{\ln^{90} d}{d} n\right). \tag{34}$$

It follows easily that $\binom{n}{k}^2 \le \binom{n}{\frac{\ln d}{d} n}^2 \le \left(\frac{de}{\log d}\right)^{2\frac{\ln d}{d} n} = \exp\left(3n \ln^2 d/d\right)$. Thus, from (34) we get that there is $\epsilon_d \to 0$ with d such that

$$\mathbb{E}[\Psi_k] \le \exp\left(-0.5 \frac{\ln^{90} d}{d} n\right),$$

for any $k = (1 - \epsilon)\frac{\ln d}{d} n$, where $\epsilon_d < \epsilon < 1 - \epsilon_d$.

Consider now the case where $k < n \ln d/(10d)$. For a pair of independent sets any vertex that is not adjacent to the vertices of the pair is an augmenting vertex. Let σ, τ be a pair of independent sets each of size $k \le (1 - \epsilon) n \ln d/d$, for $\epsilon \ge 0.9$. Let $R_{\sigma,\tau}$ be the vertices not in $\sigma \cup \tau$ but not adjacent to any vertex in $\sigma \cup \tau$, as well. Every $w \notin \sigma \cup \tau$ belongs to $R_{\sigma,\tau}$ independently of the other vertices with probability at least $(1-p)^{2k} = (d^\epsilon/d)^2$. Thus, $\mathbb{E}|R_{\sigma,\tau}| \ge (n - 2k)(d^\epsilon/d)^2$. Using Chernoff bounds we get

$$\mathbb{P}[|R_{\sigma,\tau}| = 0] \le \exp\left(-\frac{d^{2\epsilon}}{10d^2} n\right) \le \exp\left(-\frac{d^{0.8}}{10d} n\right) \quad [\text{since } \epsilon > 0.9].$$

Since $R_{\sigma,\tau}$ consists of augmenting vertices for the pair σ, τ, the probability for σ, τ not to have any augmenting vertex is upper bounded by $\mathbb{P}[|R_{\sigma,\tau}| = 0]$. For $k < n \ln d/(10d)$ it holds that

$$\mathbb{E}[\Psi_k] \leq \binom{n}{k}^2 \exp\left(-\frac{d^{0.8}}{10d}n\right) \leq \exp\left(3\frac{\ln^2 d}{d}n\right) \cdot \exp\left(-\frac{d^{0.8}}{10d}n\right) \leq \exp\left(-\frac{d^{0.8}}{15d}n\right).$$

The theorem follows. ∎

4.1. Proof of Proposition 31

Consider an arbitrary pair $\sigma, \tau \in \mathcal{S}_k(G(n, d/n))$ where $k = (1-\epsilon)n \ln d/d$ and $100\frac{\ln\ln d}{\ln d} \leq \epsilon \leq 1 - 100\frac{\ln\ln d}{\ln d}$. For the rest of the proof assume that $|\sigma \cap \tau| = ak$ where $a \in [0, 1]$. Also, let $\epsilon' \in \left[100\frac{\ln\ln d}{\ln d}, 1\right]$ be such that $1-\epsilon' = (1-a)(1-\epsilon)$. We consider two cases, in the first one we assume that $\epsilon' \leq 1 - 100\frac{\ln\ln d}{\ln d}$, in the second one we assume that $1 - 100\frac{\ln\ln d}{\ln d} < \epsilon'$.

Consider ϵ' as in the first case. We will show that with sufficiently large probability there exists a non-empty set Q_0 of augmenting vertices for the pair σ, τ. The set Q_0 contains a specific kind of augmenting vertices. To specify Q_0, we need the following definitions:

- **$\mathbf{Q_1}(\sigma)$:** $Q_1(\sigma) \subseteq V\setminus(\sigma \cup \tau)$ contains every vertex v that has exactly one neighbour in $\sigma\setminus\tau$ but it does not have exactly one neighbour in $\tau\setminus\sigma$.
- **$\mathbf{Q_2}(\sigma)$:** $Q_2(\sigma) \subseteq \sigma\setminus\tau$ is the set of vertices in $\sigma\setminus\tau$ that have at least one neighbour in $Q_1(\sigma)$.
- **$\mathbf{Q_3}(\sigma)$:** $Q_3(\sigma) \subseteq V\setminus(\sigma\cup\tau\cup Q_1(\sigma))$ contains every vertex w such that following holds:
  - **$\mathbf{S}_1$-** $N(w) \cap (\sigma\setminus\tau) \subseteq Q_2(\sigma)$ and $|N(w) \cap (\sigma\setminus\tau)| \leq 7d$.
  - **$\mathbf{S}_2$-** There exists $R \subseteq Q_1(\sigma)$ that contains exactly one neighbour of each $v \in N(w) \cap (\sigma\setminus\tau)$ in $Q_1(\sigma)$ and no other vertex. Furthermore, $R \cup \{w\}$ is an independent set.

In an analogous manner we define $Q_1(\tau)$, $Q_2(\tau)$ and $Q_3(\tau)$. Basically, the idea is that $Q_1(\sigma)$ contains the vertices that can possibly become terminals (the set I_σ in Fig. 2). Moreover, the reader should think of $Q_2(\sigma)$ as the "middle vertices" in $u_1, \ldots, u_5$ in Fig. 2, and of $Q_3(\sigma)$ as comprising the possible augmenting vertices.

Indeed, for the augmenting vertex $u \in Q_0$ the following holds: the terminal set $I_u(\sigma)$ is a subset of $Q_1(\sigma)$. Also, the neighbours of u inside $\sigma\setminus\tau$ are exclusively in $Q_2(\sigma)$. Finally, we require that $Q_0 \subseteq Q_3(\sigma)$. The corresponding holds w.r.t. independent set τ. To be more precise, for $u \in Q_0$ the following holds:

- $u \in Q_3(\sigma) \cap Q_3(\tau)$
- u has no neighbour in $\sigma \cap \tau$
- $N(u) \cap (\sigma\setminus\tau) \subseteq Q_2(\sigma)$ and $N(u) \cap (\tau\setminus\sigma) \subseteq Q_2(\tau)$,
- $I_u(\sigma) \subseteq Q_1(\sigma)$ and $I_u(\tau) \subseteq Q_1(\tau)$.

Consider a process where we reveal all the sets $Q_i(\sigma), Q_i(\tau)$, for $i = 1, 2, 3$ in steps. In each step we reveal a certain amount of information regarding these six sets. Since $Q_i(\sigma)$ is symmetric to $Q_i(\tau)$ we just present results related to $Q_i(\sigma)$, those regarding $Q_i(\tau)$ will be immediate. The results appear as a series of claims whose proofs appear after the proof of this proposition.

In `Step 1`, we reveal which vertex $u \notin \sigma \cup \tau$ has exactly one neighbour either in $\sigma \setminus \tau$ or in $\tau \setminus \sigma$ or in both. Clearly, this reveals the sets $Q_1(\sigma)$, $Q_1(\tau)$. There we have the following result.

Claim 32. *Let $X_1 = |Q_1(\sigma)|$. It holds that $\mathbb{E}[X_1] = \frac{(1-\epsilon')\ln d}{d^{1-\epsilon'}} n(1-\epsilon_d) - O(1)$, where $\epsilon_d \to 0$ as d grows. Furthermore, it holds that*

$$\mathbb{P}\left[|X_1 - \mathbb{E}[X_1]| \geq 0.5\mathbb{E}[X_1]\right] \leq 2\exp\left(-nd^{\epsilon'}/d\right).$$

In `Step 2`, we reveal the edges between $Q_1(\sigma)$ and σ (resp. $Q_1(\tau)$ and τ). By definition each vertex in σ which has a neighbour in $Q_1(\sigma)$, belongs to $Q_2(\sigma)$. Thus, in this step we also reveal $Q_2(\sigma)$ and $Q_2(\tau)$. Then we have the following result.

Claim 33. *Let $X_2 = |Q_2(\sigma)|$. For $\gamma = 1 - \ln^{-5} d$, it holds that*

$$\mathbb{P}[X_2 \leq \gamma \cdot |\sigma \setminus \tau| | \mathcal{F}_1] \leq \exp\left(-nd^{\epsilon'}/(4d\ln^5 d)\right),$$

where $\mathcal{F}_1 = \{|X_1 - \mathbb{E}[X_1]| < 0.5\mathbb{E}[X_1]\}$.

It remains to reveal the sets $Q_3(\sigma)$ and $Q_3(\tau)$. Revealing these sets is, technically, a more complex task. Let us make some observations regarding these sets. Assume that some vertex $u \in V \setminus (\sigma \cup \tau \cup Q_1(\sigma))$ satisfies $\mathbf{S_1}$, in the definition of $Q_3(\sigma)$. So as u to belong to $Q_3(\sigma)$ there should exist a set $R \subseteq Q_1(\sigma)$ as specified by $\mathbf{S_2}$. However, the possibility of edges between vertices in $Q_1(\sigma)$ leaves open whether we can have such a set for u. To this end consider the following.

Definition 34. *For every $i = 1 \ldots 7d$, let $\mathcal{A}_i$ be the family of subsets $B \subseteq Q_2(\sigma)$ with $|B| = i$ which have the following property: There exists an independent set $R \subseteq Q_1(\sigma)$ that contains exactly one neighbour of each $v \in B$ in $Q_1(\sigma)$ and no other vertex.*

That is, a vertex u which satisfies $\mathbf{S_1}$ satisfies also $\mathbf{S_2}$ only if either $N(u) \cap (\sigma \setminus \tau) \in \mathcal{A}_i$, for some appropriate $i > 0$, or $N(u) \cap (\sigma \setminus \tau) = \emptyset$.

In `Step 3` we reveal the edges with both ends in $Q_1(\sigma)$ (resp. in $Q_1(\tau)$). Observe that the families $\mathcal{A}_i$ are uniquely determined by the edges we reveal at this step. Thus, by revealing the aforementioned edges we get $\mathcal{A}_i$. Then, we get the following result.

Claim 35. *Let $\mathcal{F}_2 = \{\mathcal{F}_1$ and $X_2 > \gamma \cdot |\sigma \setminus \tau|\}$. For every $2 \leq i \leq 7d$ it holds that*

$$\mathbb{P}\left[|\mathcal{A}_i| \leq (1 - 2d^5/n)\binom{|Q_2(\sigma)|}{i} \middle| \mathcal{F}_2\right] \leq 2\exp\left(-nd^{2\epsilon'}/d\right).$$

Also $\mathcal{A}_1 = Q_2(\sigma)$.

Let the vertex set V' contain each v such that $u \notin \sigma \cup \tau$ and $|N(u) \cap \sigma \setminus \tau|, |N(u) \cap \tau \setminus \sigma| \neq 1$. The set V' contains all the vertices whose edges with σ and τ are not revealed during Step 1. Both $Q_3(\sigma), Q_3(\tau)$ will be subsets of V'.

In `Step 4`, we reveal the edges between each $v \in V'$ and the set $Q_1(\sigma) \cup Q_2(\sigma)$ as well as the edges between v and $Q_1(\tau) \cup Q_2(\tau)$. Once we have revealed these edges, it is direct to tell whether v belongs to $Q_3(\sigma)$ (resp. $Q_3(\tau)$ or not.

Claim 36. *Let the event* $\mathcal{F}_3 = \left\{\mathcal{F}_2 \text{ and } |\mathcal{A}_i| \geq (1-2d^5/n)\binom{|Q_2(\sigma)|}{i}\right\}$. *For every* $u \in V'$, *it holds that*

$$\mathbb{P}[u \in Q_3(\sigma)|\mathcal{F}_3] \geq 9/10.$$

For every $v \in V'$ let J_v be an indicator random variable such that $J_v = 1$ if $v \in Q_3(\sigma) \cap Q_3(\tau)$ and $J_v = 0$ otherwise. Observe that the edge events between v and $Q_1(\sigma) \cup Q_2(\sigma)$ are independent of the edge events between v and the vertices in $Q_1(\tau) \cup Q_2(\tau)$. The event $\mathcal{F}_3$ affects only the cardinality of V'. As long as V' is non empty J_vs for various $v \in V'$ are independent.

Let $X_3 = \sum_{v \in V'} J_v$. Using Claim 36 we get that

$$\mathbb{E}[X_3|\mathcal{F}_3] \geq \left(1 - 10d^{\epsilon'} \ln d/d\right) n \cdot (\mathbb{P}[u \in Q_3(\sigma)|\mathcal{F}_3])^2 \geq 8n/10,$$

It is not hard to see that $v \in Q_3(\sigma) \cap Q_3(\tau)$ independently of the other vertices in V'. That is, X_3 is a sum of independent identically distributed random variables. Applying Chernoff bounds for X_3 and we get that

$$\mathbb{P}[X_3 < 0.7n|\mathcal{F}_3] \leq \exp\left(-n/350\right). \tag{35}$$

In `Step 5`, the last step, we reveal the edges between every $v \in Q_3(\sigma) \cap Q_3(\tau)$ and $\sigma \cap \tau$. Clearly, every such vertex which does not have neighbours in $\sigma \cap \tau$ belongs to Q_0, i.e. it is augmenting. We have no information about the edges between the sets $Q_3(\sigma) \cap Q_3(\tau)$ and $\sigma \cap \tau$, as we never examined edges with an end in the later set.

Every $u \in Q_3(\sigma) \cap Q_3(\tau)$ is augmenting independently of all the rest vertices with probability $d^{-a(1-\epsilon)} + O(n^{-1})$, as $|\sigma \cap \tau| = \alpha k$. Let the event $\mathcal{F}_4 = \{\mathcal{F}_3 \text{ and } X_3 \geq 0.7n\}$. It is direct that $\mathbb{E}[|Q_0||\mathcal{F}_4] \geq 0.7nd^{-a(1-\epsilon)} - O(1)$. Since $a \in [0,1]$, there exists $\delta = \delta(\epsilon, a) > \epsilon$ such that $a(1-\epsilon) = 1-\delta$. Applying Chernoff bounds we get that

$$\mathbb{P}[|Q_0| = 0|\mathcal{F}_4] \leq \exp\left(-0.2d^{\delta}n/d\right) \leq \exp\left(-0.2d^{\epsilon}n/d\right) \quad [\text{as } \delta > \epsilon]. \tag{36}$$

Using (35) and Claims 32, 33 and 35 we get that $\mathbb{P}[\mathcal{F}_4] \geq 1 - 20\exp\left(-nd^{\epsilon'}/(4d\ln^5 d)\right)$. Combining this bound for $\mathbb{P}[\mathcal{F}_4]$ with (36) we get that

$$\mathbb{P}[|Q_0| = 0] \leq 30\exp\left(-nd^{\epsilon'}/(3d\ln^5 d)\right) \leq \exp\left(-n\ln^{90} d/d\right) \tag{37}$$

as $100\frac{\ln\ln d}{\ln d} \leq \epsilon' \leq 1 - 100\frac{\ln\ln d}{\ln d}$.

It remains to consider the case where $1 - 100\frac{\ln\ln d}{\ln d} < \epsilon' \leq 1$. There, it holds that $|\sigma \cup \tau| = k_0 \leq (1-\epsilon)\frac{\ln d}{d}n + 100\frac{\ln\ln d}{d}n$. Let $R_{\sigma,\tau}$ be the set of vertices, outside σ, τ, that are not adjacent to any vertex in $\sigma \cup \tau$. Every $w \notin \sigma \cup \tau$ belongs to $R_{\sigma,\tau}$ independently of the other vertices with probability $(1-p)^{k_0} \leq \left(d^{\epsilon/2}/d\right)$. Thus, $\mathbb{E}|R_{\sigma,\tau}| \geq (n-k_0)d^{\epsilon/2}/d$. Using Chernoff bounds we get

$$\mathbb{P}[|R_{\sigma,\tau}| = 0] \leq \exp\left(-\frac{d^{\epsilon/2}}{2d}n\right). \tag{38}$$

Since $R_{\sigma,\tau}$ consists of *augmenting vertices* for the pair σ, τ, the probability that there is no augmenting vertex is upper bounded by $\mathbb{P}[|R_{\sigma,\tau}| = 0]$. The proposition follows from (37) and (38).

Proof of Claim 32. Let r be the probability for a vertex v outside σ, τ, to have exactly one neighbour in $\sigma \setminus \tau$. It holds that

$$r = (1-a)kp(1-p)^{(1-a)k-1} = (1-\epsilon')\ln d/d^{1-\epsilon'} - O(n^{-1}).$$

Of course, with the same probability v has exactly one neighbour in $\tau \setminus \sigma$. Then, the probability for v to be in $Q_1(\sigma)$ is $p_1 = r(1-r)$. Observe that v belongs to $Q_1(\sigma)$ independently of the other vertices. It is direct that there exists $\epsilon_d \to 0$ such that

$$\mathbb{E}[X_1] = (n-2k)p_1 = \frac{(1-\epsilon')\ln d}{d^{1-\epsilon'}} n\,(1-\epsilon_d) - O(1).$$

The claim follows by applying Chernoff bounds. ∎

Proof of Claim 33. Due to symmetry each vertex $u \in Q_1(\sigma)$ is adjacent to exactly one random vertex in $\sigma \setminus \tau$, independently of the other vertices in $Q_1(\sigma)$. An equivalent way of looking at adjacencies between vertices in $Q_1(\sigma)$ and $\sigma \setminus \tau$ is by assuming that the vertices in $Q_1(\sigma)$ are balls and each vertex in $\sigma \setminus \tau$ is a bin and each ball is thrown into a uniformly *random* bin. The non-empty bins correspond to vertices in $Q_2(\sigma)$. The claim will follow by deriving an appropriate tail bound on the number of occupied bins.

Let N denote the number or balls and m denote the number of bins, it holds that $N \geq \frac{d^{\epsilon'}}{d} n$ and $m = (1-\epsilon')\frac{\ln d}{d} n$. For $c \in (0,1)$, let P_c be the probability that there is a subset of bins of size cm that contains all the balls. For B_c a fixed subset of bins of size cm and for a fixed ball r, it holds that

$$\begin{aligned} P_c &\leq \binom{m}{cm} (\mathbb{P}[r \text{ is placed into some bin in } B_c])^N \leq \left(\frac{me}{cm}\right)^{cm} c^N \\ &\leq \exp\left(cm(1-\ln c) + N \ln c\right). \end{aligned}$$

For $c_0 = (1 - \ln^{-5} d)$ we have that

$$\begin{aligned} P_{c_0} &\leq \exp\left(2\frac{\ln d}{d} n - \frac{d^{\epsilon'}}{2d\ln^5 d} n\right) \quad [\text{as } 1 - x \geq \exp(-x/(1-x) \text{ for } 0 < x < 0.1] \\ &\leq \exp\left(-nd^{\epsilon'}/(3d\ln^5 d)\right) \quad [\text{for large } d]. \end{aligned}$$

It is easy to check that for any $0 \leq c \leq c_0$ we have $P_c \leq P_{c_0}$. Hence, letting E_{c_0} be the event that there is a subset of at most $c_0 \cdot m$ bins that has all the balls, it holds that

$$\mathbb{P}[E_{c_0}] \leq \exp\left(-nd^{\epsilon'}/(4d\ln^5 d)\right).$$

The claim follows. ∎

Proof of Claim 35. The cardinality of each family $\mathcal{A}_i$, for $2 \leq i \leq 7d$, depends on the edges whose both ends are in $Q_1(\sigma)$. As a first step we estimate the number of these vertices conditional on the event $\mathcal{F}_2$.

Let R_1 be the set of edges whose both ends are in $Q_1(\sigma)$. The bound on X_1 and the cardinality of $Q_1(\sigma)$ that $\mathcal{F}_2$ specifies as well as the fact that each edge appears independently with probability d/n yields the following relation.

$$\mathbb{E}[|R_1||\mathcal{F}_2] = C\frac{d^{2\epsilon'}n}{d}(1-\epsilon')^2\ln^2 d,$$

where $1/8 < C < 9/8$. Chernoff bounds yield the following inequality.

$$\mathbb{P}\left[|R_1| \geq n/d^{1-3\epsilon'}|\mathcal{F}_2\right] \leq \exp\left(-nd^{2\epsilon'}/d\right). \tag{39}$$

Let the event $H = \{\mathcal{F}_2 \text{ and } |R_1| < n/d^{1-3\epsilon'}\}$.

Next, we compute $\mathbb{E}[|\mathcal{A}_i||H]$. Note that the event H specifies, only, an upper bound on $|R_1|$ and it does not specify where the edges are placed. That is, all subsets of $Q_2(\sigma)$ of cardinality i are symmetric thus they belong to $\mathcal{A}_i$ with the same probability. By the linearity of expectation we get that

$$\mathbb{E}[|\mathcal{A}_i\|H] = \binom{|Q_2(\sigma)|}{i}\mathbb{P}[L \notin \mathcal{A}_i|H] \quad [\text{for a fixed } L \subseteq Q_2(\sigma) \text{ and } |L| = i]$$

Let M_L be the family of subsets of $Q_1(\sigma)$, each of cardinality i, such that for each $\mathcal{W} \in M_L$ the following is true: The set $\mathcal{W}$ contains exactly one neighbour of each vertex $q \in L$ and no other vertex. By definition the family M_L must have at least one member. Moreover, if there exists one set in M_L which is independent, then $L \in \mathcal{A}_i$.

When we reveal the edges between the vertices in $Q_1(\sigma)$ it is easy to see that the probability that M_L contains no independent set is maximized when M_L is a singleton. Given $|R_1|$ and X_1, observe that each pair of vertices in $Q_1(\sigma)$ is adjacent with probability at most $|R_1|/\binom{X_1}{2}$. Each subset of $Q_1(\sigma)$ of cardinality i has expected number of adjacent vertices $\binom{i}{2}|R_1|/\binom{X_1}{2} \leq d^4/n$, for large d. That is, the probability that M_L does not contain an independent set is at most d^4/n. Thus,

$$\mathbb{E}[|\mathcal{A}_i\|H] \geq \left(1 - \frac{d^4}{n}\right)\binom{|Q_2(\sigma)|}{i}. \tag{40}$$

Having calculated a lower bound for $\mathbb{E}[|\mathcal{A}_i||H]$ we will show that given the event H, $|\mathcal{A}_i|$ is tightly concentrated about its expectation. Then, claim will be immediate. So as to show the concentration result, we use an edge exposure martingale argument for the edges in R_1 and then we apply Azuma's inequality (see e.g. [37] Theorem 2.25).

Observe that the revelation of each edge in R_1 cannot reduce the cardinality of $\mathcal{A}_i$ by more than $c = \binom{X_2-2}{i-2} \leq (X_2)^{i-2}/(i-2)!$ sets. Standard arguments with Azuma's inequality yield to that for any $\lambda > 0$ it holds that

$$\mathbb{P}\left[|\mathcal{A}_i| \leq \mathbb{E}[\mathcal{A}_i|H] - \lambda|H\right] \leq \exp\left(-\frac{\lambda^2}{2|R_1|c^2}\right).$$

Setting $\lambda = d^4X_2^{i-1}/i!$ we get that

$$\mathbb{P}\left[|\mathcal{A}_i| \leq \left(1 - 2\frac{d^5}{n}\right)\binom{Q_2(\sigma)}{i}\middle| H\right] \leq \exp\left(-\frac{d^8X_2^2}{2|R_1|i^2}\right) \leq \exp\left(-dn\right),$$

where the last derivation follows by using the fact that $1 \leq i \leq 7d$, $|R_1| \leq n/d^{1-3\epsilon'}$ and $100\ln\ln d/\ln d < 1 - \epsilon' < 1 - 100\ln\ln d/\ln d$. The claim follows by just using the law of total probability and get that

$$\begin{aligned}\mathbb{P}\left[|\mathcal{A}_i| \leq \left(1-2\frac{d^5}{n}\right)\binom{Q_2(\sigma)}{i}\middle|\mathcal{F}_2\right]&\\ \leq \mathbb{P}\left[|\mathcal{A}_i| \leq \left(1-2\frac{d^5}{n}\right)\binom{Q_2(\sigma)}{i}\middle|H\right] + \mathbb{P}\left[|R_1| \geq n/d^{1-3\epsilon'}\middle|\mathcal{F}_2\right]&\\ \leq 2\exp\left(-nd^{2\epsilon'}/d\right).&\end{aligned}$$

■

Proof of Claim 36. For some $u \in V$, let $d_{\sigma,\tau}(u)$ be the number of vertices in $\sigma\setminus\tau$ which are adjacent to u. Also, let the event $E_i = \{N(u) \cap (\sigma\setminus\tau) \in \mathcal{A}_i\}$ for $i > 0$ and $E_0 = \{N(u) \cap (\sigma\setminus\tau) = \emptyset\}$. By the law of total probability we get that

$$\begin{aligned}\mathbb{P}[u \in Q_3(\sigma)|\mathcal{F}_3] \geq \sum_{i=0}^{7d} \mathbb{P}[u \in Q_3(\sigma)|d_{\sigma,\tau} = i, E_i, \mathcal{F}_3]\cdot\mathbb{P}[E_i|d_{\sigma,\tau} = i, \mathcal{F}_3]&\\ \cdot\mathbb{P}[d_{\sigma,\tau} = i|\mathcal{F}_3].&\end{aligned} \tag{41}$$

We impose the bound $i \leq 7d$ since no vertex in $Q_3(\sigma)$ can have more than $7d$ neighbours in $Q_2(\sigma)$. Conditional on $d_{\sigma,\tau}(u) = i$, all the subsets of size i in $\sigma\setminus\tau$ are equally likely to be adjacent to u. Thus, we get that

$$\begin{aligned}\mathbb{P}[E_i|d_{\sigma,\tau} = i, \mathcal{F}_3] = \frac{|\mathcal{A}_i|}{\binom{|\sigma\setminus\tau|}{i}} \geq (1-2d^5/n)\frac{\binom{X_2}{i}}{\binom{|\sigma\setminus\tau|}{i}} \quad \text{[by Claim 35]}&\\ \geq \left(\frac{X_2}{|\sigma\setminus\tau|}\right)^i(1-o(1)) \geq \gamma^i(1-o(1)),&\end{aligned} \tag{42}$$

where $\gamma = 1-\ln^{-5} d$. Also, it is easy to see that

$$\mathbb{P}[u \in Q_3(\sigma)|d_{\sigma,\tau} = i, E_i, \mathcal{F}_3] \geq (1-d/n)^i \geq 1-7d^2/n. \quad [\text{as } 0 \leq i \leq 7d] \tag{43}$$

Let the event C be $d_{\sigma,\tau}(u) \neq 1$ and $d_{\sigma,\tau}(u) \leq 7d$. Observe that the variable $d_{\sigma,\tau}(u)$ is distributed as in $\mathcal{B}((1-a)k, d/n)$ conditional on the event C. Using this along with (43) and (42) we can rewrite (41) as follows:

$$\begin{aligned}&\mathbb{P}[u \in Q_3|\mathcal{F}_3]\\ &\geq \frac{1-o(1)}{\mathbb{P}[C|\mathcal{F}_3]}\left[\sum_{i=0}^{7d}\binom{(1-a)k}{i}p^i(1-p)^{(1-a)k-i}\gamma^i - \gamma\binom{(1-a)k}{1}p(1-p)^{(1-a)k-1}\right]\\ &\geq (1-o(1))\left[\sum_{i=0}^{7d}\binom{(1-a)k}{i}p^i(1-p)^{(1-a)k-i}\gamma^i - d^{-(1-\epsilon')}\ln d\right],\end{aligned} \tag{44}$$

where the last inequality follows from the fact that $\gamma, \mathbb{P}[C|\mathcal{F}_3] \leq 1$ and a simple derivation which implies that $\binom{(1-a)k}{1}p(1-p)^{(1-a)k-1} \leq d^{-(1-\epsilon')}\ln d$. Also, note that

$$\begin{aligned}\sum_{i=7d+1}^{(1-a)k}\binom{(1-a)k}{i}p^i(1-p)^{(1-a)k-i}\gamma^i \leq \sum_{i=7d+1}^{(1-a)k}\binom{(1-a)k}{i}p^i(1-p)^{(1-a)k-i} \quad [\text{as } 0 \leq \gamma < 1]&\\ \leq \exp(-7d).&\end{aligned} \tag{45}$$

The last inequality follows by noting that the summation on the l.h.s. of the first line is equal to the probability $\mathbb{P}[\mathcal{B}((1-a)k, d/n) > 7d]$ and bounding it by using Theorem 8, i.e. (4). Using (45), we get that

$$\begin{aligned}\sum_{i=0}^{7d}\binom{(1-a)k}{i}p^i(1-p)^{(1-a)k-i}\gamma^i &\geq (1-p\ln^{-5}d)^{(1-a)k} - \exp(-7d)\\ &\geq \exp\left[-(1-\epsilon')\ln^{-4}d - O(n^{-1})\right] - \exp(-7d)\\ &\qquad\qquad [\text{as } \ln(1-x) = -x - O(x^2)]\\ &\geq 1 - \frac{1-\epsilon'}{\ln^4 d} - \exp(-7d) - O(n^{-1}) \quad [\text{as } 1+x \leq e^x]\\ &\geq 95/100. \end{aligned} \tag{46}$$

The claim follows by plugging (46) into (44) and get that $\mathbb{P}[u \in Q_3 | \mathcal{F}_3] \geq 9/10$. ∎

5. PROOF OF THEOREM 3

The following proposition reduces the problem of establishing shattering to an exercise in calculus.

Proposition 37. *There exist a constant $d_0 > 0$ and $\varepsilon_d \to 0$ such that for all $d > d_0$ the following is true. Suppose that $s = (1+q)\ln d/d$ for $\epsilon_d \leq q \leq (1-\epsilon_d)$ and let*

$$\psi(x) = \psi_{d,s}(x) = xs(2 - 2\ln x - \ln s) + \frac{d}{2}\ln\left(1 - \frac{s^2(1-(1-x)^2)}{1-s^2}\right).$$

If there is a real $0 < b < 1$ such that

$$\psi(b) < -18qs \quad \text{and} \tag{47}$$

$$\sup_{x<b}\psi(x) < -s\ln(s) - (1-s)\ln(1-s) + \frac{d}{2}\ln(1-s^2) - 20s \tag{48}$$

then for $k = sn$ there occurs shattering.

To prove Theorem 3, consider a pair (G, σ) chosen from the planted model $\mathcal{P}_k(n,m)$. We are going to show that under the assumptions (47) and (48) the independent set σ belongs to a small 'cluster' of independent sets that is separated from the others by a linear Hamming distance with a probability very close to one. We will then use Theorem 18 to transfer this result to the distribution $\mathcal{U}_k(n,m)$. Let $Z_{k,\beta}$ be the number of independent sets $\tau \in \mathcal{S}_k(G)$ such that $|\sigma \cap \tau| = (1-\beta)k$.

Lemma 38. *We have $\frac{1}{n}\ln \mathbb{E}_{\mathcal{P}_k(n,m)}\left[Z_{k,\beta}\right] \leq \psi(\beta) + o(1)$.*

Proof. Let $\tau \subset V$ be such that $|\sigma \cap \tau| = (1-\beta)k$. The total number of graphs with m edges in which both σ, τ are independent sets equals

$$\binom{\binom{n}{2} - 2\binom{k}{2} + \binom{(1-\beta)k}{2}}{m}.$$

For we can choose any m edges out of those potential edges that do not join two vertices of either σ or τ. Since both σ, τ have size k and $|\sigma \cap \tau| = (1-\beta)k$, the number of such 'bad' potential edges is $2\binom{k}{2} - \binom{(1-\beta)k}{2}$ by inclusion/exclusion. Since G is chosen uniformly among all $\binom{\binom{n}{2}-\binom{k}{2}}{m}$ graphs in which σ is independent, we thus get

$$\begin{aligned}
\mathbb{P}\left[\tau \text{ is independent}\right] &= \binom{\binom{n}{2} - 2\binom{k}{2} + \binom{(1-\beta)k}{2}}{m} \Bigg/ \binom{\binom{n}{2} - \binom{k}{2}}{m} \\
&= \prod_{j=0}^{m-1} \frac{\binom{n}{2} - 2\binom{k}{2} + \binom{(1-\beta)k}{2} - j}{\binom{n}{2} - \binom{k}{2} - j} \leq \left(\frac{\binom{n}{2} - 2\binom{k}{2} + \binom{(1-\beta)k}{2}}{\binom{n}{2} - \binom{k}{2}}\right)^m \\
&= \left(1 - \frac{k^2 - ((1-\beta)k)^2}{n^2 - k^2} + O(1/n)\right)^m \\
&\leq O(1) \cdot \left(1 - \frac{s^2(1-(1-\beta)^2)}{1-s^2}\right)^m \quad [\text{as } k = sn]. \qquad (49)
\end{aligned}$$

Furthermore, the total number of ways to choose a set τ with $|\sigma \cap \tau| = (1-\beta)k$ equals $\binom{k}{(1-\beta)k} \cdot \binom{n-k}{\beta k}$ (choose the $(1-\beta)k$ vertices in the intersection $\sigma \cap \tau$ and then choose the remaining βk vertices). By the linearity of the expectation, we get from (49)

$$\begin{aligned}
\mathbb{E}\left[Z_{k,\beta}\right] &= O(1) \cdot \binom{k}{(1-\beta)k} \cdot \binom{n-k}{\beta k} \cdot \left(1 - \frac{s^2(1-(1-\beta)^2)}{1-s^2}\right)^m \\
&= O(1) \cdot \binom{k}{\beta k} \cdot \binom{n-k}{\beta k} \cdot \left(1 - \frac{s^2(1-(1-\beta)^2)}{1-s^2}\right)^m \\
&\leq O(1) \cdot \left(\frac{\mathrm{e}}{\beta}\right)^{\beta k} \left(\frac{\mathrm{e}(n-k)}{\beta k}\right)^{\beta k} \cdot \left(1 - \frac{s^2(1-(1-\beta)^2)}{1-s^2}\right)^m \\
&= O(1) \cdot \left(\frac{\mathrm{e}^2(1-s)}{s\beta^2}\right)^{\beta sn} \cdot \left(1 - \frac{s^2(1-(1-\beta)^2)}{1-s^2}\right)^{dn/2} \quad [\text{as } k = sn \text{ and } m = dn/2].
\end{aligned}$$

Taking logarithms and dividing by n completes the proof. ∎

Let us call an independent set σ of size k of a graph G (b_1, b_2, γ)-*good* if G has no independent set τ such that $(1-b_1)k \leq |\sigma \cap \tau| \leq (1-b_2)k$ and if $|\{\tau \in \mathcal{S}_k(G) : |\sigma \cap \tau| > (1-b_2)k\}| \leq \exp(-\gamma n)|\mathcal{S}_k(G)|$. Moreover, let

$$\mathcal{Z}_{d,k} = \left\{(G,\sigma) \in \Lambda_k(n,m) : |\mathcal{S}_k(G)| \geq \mathbb{E}|\mathcal{S}_k(G(n,m))| \cdot \exp\left(-14n\sqrt{\ln^5 d/d^3}\right)\right\}. \qquad (50)$$

Corollary 39. *Suppose that $b > 0$ is such that (47) and (48) hold. Then there exist $b_1, b_2, \gamma > 0$ such that*

$$\mathbb{P}_{\mathcal{U}_k(n,m)}\left[(G,\sigma) \text{ is } (b_1,b_2,\gamma)\text{-good} | \mathcal{Z}_{d,k}\right] \geq 1 - \exp(-\gamma n).$$

Proof. The function ψ is continuous. Therefore, if (47) and (48) are satisfied for some $b < 0$ then there exist $b_1 > b_2$ and $\zeta > 0$ such that

$$\sup_{b_2 \le \beta \le b_1} \psi(\beta) < -18qs - \zeta \quad \text{and} \tag{51}$$

$$\sup_{x < b_2} \psi(x) < -s\ln(s) - (1-s)\ln(1-s) + \frac{d}{2}\ln(1-s^2) - d^{-1.49} - \zeta. \tag{52}$$

Let $Z_{k,b_1,b_2}(G,\sigma)$ be the number of $\tau \in \mathcal{S}_k(G)$ such that $(1-b_1)k \le |\sigma \cap \tau| \le (1-b_2)k$. Then Lemma 38, (51), and Markov's inequality yield

$$\begin{aligned}\mathbb{P}_{\mathcal{P}_k(n,m)}\left[Z_{k,b_1,b_2} > 0\right] &\le \mathbb{E}_{\mathcal{P}_k(n,m)}\left[Z_{k,b_1,b_2}\right] \le \sum_{b_2 k \le j \le b_1 k} \mathbb{E}_{\mathcal{P}_k(n,m)}\left[Z_{k,j/k}\right] \\ &\le \exp\left[n\left(\sup_{b_2 \le \beta \le b_1} \psi(\beta) + o(1)\right)\right] \le \exp\left[-n\ln\ln d/d\right]. \end{aligned} \tag{53}$$

The last inequality follows by taking $q > 100\ln\ln d/\ln d$ and then $18qs \ge \ln\ln d/d$. Similarly, let $Z_{k,<b_2}(G,\sigma)$ be the number of $\tau \in |\mathcal{S}_k(G)|$ such that $|\sigma \cap \tau| > (1-b_2)k$. Moreover, let $s = k/n$ and let

$$\begin{aligned}\mu &= \mathbb{E}|\mathcal{S}_k(G(n,m))| \cdot \exp\left(-14n\sqrt{\ln^5 d/d^3}\right) \\ &= O(1)\binom{n}{k}(1-(k/n)^2)^m \cdot \exp\left(-14n\sqrt{\ln^5 d/d^3} + o(n)\right) \quad \text{[by Corollary 10]} \\ &= \exp\left[n\left(-s\ln(s) - (1-s)\ln(1-s) - \frac{d}{2}\ln(1-s^2) - 14\sqrt{\ln^5 d/d^3} + o(1)\right)\right],\end{aligned}$$

where in the last step we used Stirling's formula. Using (52) and Markov's inequality, we find that

$$\begin{aligned}\mathbb{P}_{\mathcal{P}_k(n,m)}\left[Z_{k,<b_2} > \mu\right] &\le \frac{\mathbb{E}_{\mathcal{P}_k(n,m)}\left[Z_{k,<b_2}\right]}{\mu} \le \sum_{j < b_2 k} \frac{\mathbb{E}_{\mathcal{P}_k(n,m)}\left[Z_{k,j/k}\right]}{\mu} \\ &\le \frac{1}{\mu}\exp\left[n\left(\sup_{\beta < b_2}\psi(\beta) + o(1)\right)\right] \le \exp\left[-n\ln d/d\right]. \end{aligned} \tag{54}$$

Combining (53) and (54) with Corollary 21, and letting, say, $\gamma = d^{-2}$, we see that

$$\begin{aligned}&\mathbb{P}_{\mathcal{U}_k(n,m)}\left[(G,\sigma) \text{ is not } (b_1,b_2,\gamma)\text{-good} | \mathcal{Z}_{d,k}\right] \\ &\quad \le \mathbb{P}_{\mathcal{U}_k(n,m)}\left[Z_{k,<b_2} > \mu \text{ or } Z_{k,b_1,b_2} > 0\right] \\ &\quad \le (1+o(1))\mathbb{P}_{\mathcal{P}_k(n,m)}\left[Z_{k,>b_2} > \mu \text{ or } Z_{k,b_1,b_2} > 0\right] \cdot \exp\left[14n\sqrt{\ln^5 d/d^3}\right] \\ &\quad \le \exp(-\gamma n),\end{aligned}$$

as claimed. ∎

Proof of Proposition 37. Let $\mathcal{Z}$ be the event that

$$|\mathcal{S}_k(G(n,m))| \ge \mathbb{E}|\mathcal{S}_k(G(n,m))| \cdot \exp\left(-14n\sqrt{\ln^5 d/d^3}\right).$$

Corollary 39 implies that there exists b_1, b_2, γ such that given $\mathcal{Z}$, w.h.p. $G = G(n,m)$ has the property that all but $\exp(-\gamma n)|\mathcal{S}_k(G(n,m))|$ independent sets $\sigma \in \mathcal{S}_k(G)$ are (b_1, b_2, γ)-good. Let $\mathcal{G}$ denote this event. As Lemma 20 ensures that $G(n,m) \in \mathcal{Z}$ w.h.p., we have

$$\mathbb{P}[\mathcal{G}] \geq \mathbb{P}[\mathcal{G} \cap \mathcal{Z}] = \mathbb{P}[\mathcal{G}|\mathcal{Z}] \cdot \mathbb{P}[\mathcal{Z}] = 1 - o(1).$$

As a consequence, we just need to show that the two conditions in Definition 2 are satisfied if $\mathcal{G}$ occurs.

Thus, let $G \in \mathcal{G}$. We construct a decomposition of $\mathcal{S}_k(G)$ into pairwise disjoint subsets $S_1, \ldots, S_N$ inductively as follows. Suppose $i \geq 1$. If the set $\mathcal{S}_k(G) \setminus \bigcup_{j=1}^{i-1} S_j$ does not contain a (b_1, b_2, γ)-good set anymore, let $N = i$, set

$$S_N = \mathcal{S}_k(G) \setminus \bigcup_{j=1}^{N-1} S_j$$

and stop. Otherwise, choose some $\sigma_i \in \mathcal{S}_k(G) \setminus \bigcup_{j=1}^{i-1} S_j$ that is (b_1, b_2, γ)-good, let

$$S_i = \{\tau \in \mathcal{S}_k(G) : |\sigma_i \cap \tau| > b_2 k\} \setminus \bigcup_{j=1}^{i-1} S_j$$

and proceed to $i + 1$.

Let $\zeta = k(b_1 - b_2)/n$. We claim that this construction satisfies the two conditions in Definition 2. Indeed, each σ_i is (b_1, b_2, γ)-good for all, we have $|S_i| \leq \exp(-\gamma n)\,|\mathcal{S}_k(G)|$ for all $i < N$. Furthermore, as $G \in \mathcal{G}$ we have $|S_N| \leq \exp(-\gamma n)\,|\mathcal{S}_k(G)|$. Thus, the partition $S_1, \ldots, S_N$ satisfies the first condition in Definition 2.

With respect to the second condition, let $\tau \in S_i$ and $\tau' \in S_j$ with $1 \leq i < j \leq N$. Assume for contradiction that $\mathrm{dist}(\tau, \tau') < \zeta n$. Then

$$\mathrm{dist}(\sigma_i, \tau') \leq \mathrm{dist}(\sigma_i, \tau) + \mathrm{dist}(\tau, \tau') = 2(k - |\sigma_i \cap \tau|) + \zeta n \leq 2b_2 k + \zeta n,$$

and thus $|\sigma_i \cap \tau'| = k - \mathrm{dist}(\sigma_i, \tau')/2 \leq (1 - b_2)k - \zeta n/2 \in [(1 - b_1)k, (1 - b_2)k]$. This contradicts the fact that σ_i is good (which implies that there is no independent set σ' such that $|\sigma_i \cap \sigma'| \in [(1 - b_1)k, (1 - b_2)k]$). Thus, we have established the second condition in Definition 2. ∎

Lemma 40. *There exist a constant $d_0 > 0$ and $\epsilon_d \to 0$ such that for all $d > d_0$ the following is true. If $s = (1 + q)\ln d/d$, where $\epsilon_d \leq q \leq 1 - \epsilon_d$, then for $b = 20/\ln d$ conditions (47) and (48) are satisfied.*

Proof. Let $\epsilon_d = 5 \ln\ln d/d$. Using the elementary inequality $\ln(1 - x) \leq -x$, we find

$$\begin{aligned}
\psi(x) &\leq sx(2 - 2\ln x - \ln s) - \frac{ds^2}{2}(1 - (1 - x)^2) \\
&= sx\,(2 - 2\ln x - \ln s - ds + dsx/2) \\
&\leq sx\,(2 - 2\ln x - \ln d - ds + dsx/2) \quad [\text{as } s \geq \ln d/d] \\
&\leq sx\,(2 - 2\ln x - \delta \ln d + dsx/2) \quad [\text{as } s \geq (1 + \epsilon_d)\ln d/d]. \qquad (55)
\end{aligned}$$

Hence, for $d \geq d_0$ sufficiently large our choice of s, b ensures that

$$\psi(b) \leq bs\,(22 + 2\ln\ln d - \ln 20 - q\ln d) \leq -\frac{9}{10}bsq\ln d \leq -18qs.$$

Thus, we have verified (47).

Starting from (55), we see that for any $\beta < b$ and $d > d_0$ large,

$$\begin{aligned}\psi(\beta) &\leq \beta s(22 - 2\ln\beta - 100\ln\ln d) \quad [\text{as } \beta ds < 40 \text{ and by the choice of } \epsilon_d]\\ &\leq -2\beta s\ln\beta < s,\end{aligned} \tag{56}$$

because $-x\ln x < 1/2$ for all $x > 0$. By comparison, for $s \leq (2 - \epsilon_d)\ln d/d$ we have

$$\begin{aligned}&-s\ln(s) - (1-s)\ln(1-s) + \frac{d}{2}\ln(1-s^2)\\ &\quad\geq -s\ln s + s - \frac{ds^2}{2} - \frac{ds^4}{2} \quad [\text{using } \ln(1-x) \geq -x - x^2]\\ &\quad\geq s\,(-\ln s - ds/2 + 1)\\ &\quad\geq s\left(\frac{1-q}{2}\ln d - \ln\ln d + 1\right) \geq 40s\ln\ln d.\end{aligned} \tag{57}$$

Combining (56) and (57), we obtain

$$\begin{aligned}\psi(\beta) &< -s\ln(s) - (1-s)\ln(1-s) + \frac{d}{2}\ln(1-s^2) - s\\ &< -s\ln(s) - (1-s)\ln(1-s) + \frac{d}{2}\ln(1-s^2) - 20s\end{aligned}$$

as $s \geq \ln d/d$. Thus, we have got (48). ∎

Finally, Theorem 3 is immediate from Proposition 37 and Lemma 40. ∎

6. PROOF OF THEOREM 5

In this section we assume that $d \geq d_0$ for some large enough constant $d_0 > 0$. Moreover, let $\varepsilon_d \to 0$ be a function of d that tends to 0 sufficiently slowly, and assume that $k = (1-\varepsilon)n\ln d/d$ for some $\varepsilon \in [\varepsilon_d, 1 - \varepsilon_d]$.

Our goal is to show that for a random pair (G, σ) chosen from $\mathcal{U}_k(n, m)$ w.h.p. there is a larger independent set τ in G that contains σ as a subset. More precisely, τ is supposed to have size $k(1 + \frac{2\varepsilon}{1-\varepsilon})$. In order to construct such a set τ we need the following concept.

Definition 41. *A vertex $v \in V\backslash\sigma$ is called* σ**-pure** *in G if it is not adjacent to any vertex in σ.*

Basically, in order to expand σ we are going to show that G has an independent set $I \subset V \setminus \sigma$ of size $|I| = 2\varepsilon k/(1-\varepsilon)$ consisting of σ-pure vertices. Then $\tau = \sigma \cup I$ is the desired larger independent set. We begin by estimating the number of σ-pure vertices and the density of the graph that they span.

Lemma 42. *Let (G,σ) be chosen from $\mathcal{P}_k(n,m)$, where $k = (1-\varepsilon)\frac{\ln d}{d}n$ with $\varepsilon \in [10\ln\ln d/\ln d, 1]$. Let Q be the set of σ-pure vertices. Then with probability at least $1-\exp\left(-\frac{n}{d}\right)$ the following two statements hold.*

1. *Let $N = |Q|$. Then $N \geq (1-o_d(1))d^{\varepsilon-1}n$.*
2. *Let M be the number of edges in the induced subgraph $G[Q]$. Then $M \leq (\frac{1}{2}+\delta)d^{2\varepsilon-1}n$, with $0 < \delta < 2d^{-\epsilon/3}$.*

Proof. Instead of working directly with the distribution $\mathcal{P}_k(n,m)$, let us consider the following variant $\mathcal{P}'_k(n,m)$. First, choose a set $\sigma' \subset V$ of size k uniformly at random. Then, construct a graph G' by inserting each of the $\binom{n}{2}-\binom{k}{2}$ possible edges that do not join two vertices in σ' with probability $p = m/(\binom{n}{2}-\binom{k}{2})$ independently.

Thus, the number of edges in G' is binomially distribution with mean m. Furthermore, given that G' has precisely m edges, it is a uniformly random graph with this property in which σ' is an independent set. Therefore, for any event $\mathcal{A}$ we have

$$\begin{aligned}\mathbb{P}_{\mathcal{P}_k(n,m)}[\mathcal{A}] &= \mathbb{P}_{\mathcal{P}'_k(n,m)}\left[\mathcal{A}\,|\,|E(G')| = m\right] \\ &\leq \frac{\mathbb{P}_{\mathcal{P}'_k(n,m)}[\mathcal{A}]}{\mathbb{P}\left[\mathrm{Bin}\left(\binom{n}{2}-\binom{k}{2},p\right)=m\right]} = \Theta(\sqrt{m})\cdot\mathbb{P}_{\mathcal{P}'_k(n,m)}[\mathcal{A}], \end{aligned} \tag{58}$$

where the last step follows from Stirling's formula.

Now, let N' be the number of σ'-pure vertices in G'. For each vertex $v \notin \sigma$ the number of neighbours in σ is binomially distributed with mean kp. In effect, v is pure with probability $(1-p)^k$. Since these events are mutually independent for all $v \notin \sigma$, N' has a binomial distribution $\mathrm{Bin}(n-k,(1-p)^k)$. Hence, letting $s = k/n = (1-\varepsilon)\ln d/d$, we have

$$\begin{aligned}\mathbb{E}\left[N'\right] &= (n-k)(1-p)^k \sim (1-s)n\exp(-kp) \sim (1-s)n\exp\left[-\frac{ds}{1-s^2}\right] \\ &\geq (1-s)n\exp\left[-ds\left(1+2s^2\right)\right] \geq 0.99nd^{\varepsilon-1},\end{aligned}$$

provided that d is sufficiently big. Letting $\gamma = d^{-\varepsilon/3} = o_d(1)$, we obtain from Theorem 8 (the Chernoff bound)

$$\mathbb{P}\left[N' < (1-\gamma)nd^{\varepsilon-1}\right] \leq \exp\left[-nd^{\varepsilon/3-1}/4\right] \leq \exp\left[-2n/d\right]$$

for d large enough. Together with (58) this implies the first assertion.

To prove the second assertion, we need an upper bound on N'. Once more by the Chernoff bound,

$$\mathbb{P}\left[N' > (1+\gamma)nd^{\varepsilon-1}\right] \leq \exp\left[-nd^{\varepsilon/3-1}/8\right] \leq \exp\left[-2n/d\right] \tag{59}$$

for d large enough. Let Q be the set of σ'-pure vertices in G'. Since each potential edge that does not link two vertices in σ' is present in G' with probability p independently, given the value of N' the number M' of edges spanned by Q is binomially distributed with mean $\binom{N'}{2}p$. Therefore,

$$\mathbb{E}\left[M'|N' \leq (1+\gamma)nd^{\varepsilon-1}\right] \leq \frac{(1+\gamma)^2n^2d^{2\varepsilon-2}}{2}\cdot\frac{dn/2}{\binom{n}{2}-\binom{k}{2}} \leq \frac{1+3\gamma}{2}nd^{2\varepsilon-1},$$

provided that d is large. Hence, by the Chernoff bound and (59),

$$\begin{aligned}\mathbb{P}\left[M' > \left(\frac{1}{2}+2\gamma\right)nd^{2\varepsilon-1}\right] &\le \mathbb{P}\left[M' > \left(\frac{1}{2}+2\gamma\right)nd^{2\varepsilon-1}\,\middle|\, N' \le (1+\gamma)nd^{\varepsilon-1}\right] \\ &\quad + \mathbb{P}\left[N' > (1+\gamma)nd^{\varepsilon-1}\right] \\ &\le \exp\left[-nd^{2\varepsilon-1}/8\right] + \exp\left[-2n/d\right] \le 2\exp\left[-2n/d\right]\end{aligned} \tag{60}$$

for d big. Finally, the second assertion follows from (58) and (60). ∎

Proof of Theorem 5. Suppose that $k = (1-\varepsilon)n\ln d/d$. Let (G,σ) be a pair chosen from the distribution $\mathcal{P}_k(n,m)$. Let Q be the set of σ-pure vertices and let N, M be as in Lemma 42. Crucially, given Q, N, M, the induced subgraph $G[Q]$ is just a uniformly random graph on N vertices with M edges, because the conditioning only imposes the absence of Q-σ-edges. In other words, $G[Q]$ is nothing but a random graph $G(N,M)$. We are going to use this observation to show that $G[Q]$ contains a large independent set w.h.p.

Let $\mathcal{A}$ be the event that $N \ge (1-o_d(1))d^{\varepsilon-1}n$ and $M \le (\frac{1}{2}+o_d(1))d^{2\varepsilon-1}n$. Then by Lemma 42

$$\mathbb{P}_{\mathcal{P}_k(n,m)}[\mathcal{A}] \ge 1-\exp(-n/d). \tag{61}$$

Given $\mathcal{A}$, the average degree of $G[Q]$ is

$$D = \frac{2M}{N} \le (1+o_d(1))\frac{d^{2\varepsilon-1}}{d^{\varepsilon-1}} = (1+o_d(1))d^{\varepsilon}.$$

Let $\mathcal{B}$ be the event that $\alpha(G[Q]) \ge (2-o_d(1))\frac{N\ln D}{D}$. Since $G[Q]$ is distributed as $G(N,M)$, Corollary 16 implies that

$$\mathbb{P}_{\mathcal{P}_k(n,m)}[\mathcal{B}|\mathcal{A}] \ge 1-\exp\left(-\frac{8n}{\epsilon^3 d\ln^3 d}\right). \tag{62}$$

Combining (61) and (62) with Theorem 18, we thus get

$$\mathbb{P}_{\mathcal{U}_k(n,m)}[\mathcal{A}\cap\mathcal{B}] = 1-o(1). \tag{63}$$

Now assume that $(G,\sigma) \in \mathcal{A}\cap\mathcal{B}$. Let I be the largest independent set of $G[Q]$. Then

$$|I| = (1-o_d(1))\frac{2d^{\varepsilon-1}n\cdot\ln(d^{\varepsilon})}{d^{\varepsilon}} = (1-o_d(1))\frac{2\varepsilon\ln d}{d} = (1-o_d(1))\frac{2\varepsilon k}{1-\varepsilon}. \tag{64}$$

Since $\sigma\cup I$ is independent, (64) shows that σ is $((2-o_d(1))\varepsilon/(1-\varepsilon),0)$-expandable. Thus, the assertion follows from (63). ∎

7. PROOF OF THEOREM 6

Let $\varepsilon_d = 3\ln\ln d/\ln d \to 0$. In this section we assume that $k = (1+\varepsilon)n\ln d/d$ with $\varepsilon_d \le \varepsilon \le 1-\varepsilon_d$, and that $d \ge d_0$ for some large enough constant $d_0 > 0$. Assuming that $\gamma, \delta > 0$ are reals such that

$$\gamma > \varepsilon_d \quad \text{and} \quad \delta < \gamma + \frac{2(\varepsilon-\varepsilon_d)}{1+\varepsilon}, \tag{65}$$

we are going to show that in a pair (G,σ) chosen from the distribution $\mathcal{U}_k(n,m)$, σ is not (γ,δ)-expandable.

To see why this is plausible, consider a pair (G,σ) chosen from the distribution $\mathcal{P}_k(n,m)$. (The following argument is not actually needed for our proof of Theorem 6; it is only included to facilitate understanding.) Then for each vertex $v \notin \sigma$ the *expected* number of neighbours of v inside of σ is greater than $kd/n = (1+\varepsilon)\ln d$. Indeed, one could easily show that for each vertex v the number of neighbours in σ dominates a Poisson variable $\mathrm{Po}((1+\varepsilon)\ln d)$. Hence, the probability that v is σ-pure is bounded by $\exp(-(1+\varepsilon)\ln d) = d^{-\varepsilon-1}$, and thus the expected number of σ-pure vertices is $\leq nd^{-\varepsilon-1} = o_d(1)\cdot k$. In effect, in order to expand σ significantly we would have to include some vertices that are *not* σ-pure. But each such vertex would 'displace' some other vertex from σ (by the very definition of σ-pure). In fact, most vertices that are not σ-pure have several neighbours in σ, and thus it seems impossible to expand σ substantially without first removing a significant share of its vertices.

To actually prove Theorem 6 we use a first moment argument. We begin by analysing the planted model.

Lemma 43. *With $d \geq d_0$ sufficiently large and k,γ,δ as above, we have*

$$P_{\mathcal{P}_k(n,m)}[\sigma \text{ is not } (\gamma,\delta)\text{-expandable }] \geq 1-\exp\left(-\frac{n}{d}\right).$$

Proof. Let $s = k/n$. For (G,σ) chosen from the distribution $\mathcal{P}_k(n,m)$, let X be the number of independent sets τ such that

$$|\tau| = (1+\gamma)k \quad \text{and} \quad |\tau\cap\sigma| \geq (1-\delta)k. \tag{66}$$

The total number of ways to choose a set $\tau \subset V$ satisfying (66) is

$$\mathcal{H} = \binom{k}{(1-\delta)k}\binom{n-k}{(\gamma-\delta)k} \tag{67}$$

(first choose $(1-\delta)k$ vertices from σ, then choose the remaining $(1+\gamma)k-(1-\delta)k = (\gamma-\delta)k$ vertices from $V\setminus\sigma$). Furthermore, for any $\tau\subset V$ satisfying (66) the probability of being independent is

$$\mathcal{P} = \binom{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\delta)k}{2}}{m}\Big/\binom{\binom{n}{2}-\binom{k}{2}}{m} \tag{68}$$

Indeed, in order for both σ and τ to be independent we have to forbid all edges that connect two vertices in either set, and the number of potential such edges is $\binom{|\sigma|}{2}+\binom{|\tau|}{2}-\binom{|\sigma\cap\tau|}{2}$ by inclusion/exclusion. This explains the numerator in (68), and the denominator simply reflects that G is chosen randomly from all graphs in which σ is independent.

Combining (67) and (68) and using the linearity of the expectation, we see that

$$\mathbb{E}[X] = \mathcal{H}\cdot\mathcal{P}. \tag{69}$$

We are going to show that $\mathbb{E}[X] < e^{-d/n}$ and then apply Markov's inequality to obtain the lemma.

We begin by estimating $\mathcal{H}$ and $\mathcal{P}$ separately. For $\mathcal{H}$ we get

$$\begin{aligned}\mathcal{H} &= \binom{k}{\delta k}\binom{(1-s)n}{(\gamma+\delta)sn} \le \left(\frac{\mathrm{e}}{\delta}\right)^{\delta k}\left(\frac{\mathrm{e}(1-s)}{(\gamma+\delta)s}\right)^{(\gamma+\delta)k} \\ &= \exp\left[s\left[\delta(1-\ln\delta)+(\gamma+\delta)\left(1+\ln\left(\frac{1-s}{(\gamma+\delta)s}\right)\right)\right]n\right] \\ &\le \exp\left[s\left[\delta(1-\ln\delta)+(\gamma+\delta)\left(1-\ln(\gamma+\delta)-\ln s\right)\right]n\right].\end{aligned}$$

As we assume that $s \ge \ln d/d$ and $\gamma \ge \varepsilon_d \ge 1/\ln d$ and $\delta \ge 0$, we have $-\ln s \le \ln d$ and $-\ln(\gamma+\delta) \le \ln\ln d$. Furthermore, the function $x \mapsto x(1-\ln x)$ is monotonically increasing for $x \le 1$. Hence, if $\gamma+\delta \le 1$, then $\delta(1-\ln\delta) \le (\gamma+\delta)(1-\ln(\gamma+\delta))$. If, on the other hand, $\gamma+\delta > 1$, then $\delta(1-\ln\delta) \le 1 < \gamma+\delta$. In either case we obtain

$$\frac{1}{n}\ln\mathcal{H} \le s(\gamma+\delta)(1+\ln\ln d-\ln d). \tag{70}$$

With respect to $\mathcal{P}$, we have

$$\begin{aligned}\mathcal{E} &= \binom{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\delta)k}{2}}{m}\Bigg/\binom{\binom{n}{2}-\binom{k}{2}}{m} \\ &= \prod_{j=0}^{m-1}\frac{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\delta)k}{2}-j}{\binom{n}{2}-\binom{k}{2}-j} \le \left(\frac{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\delta)k}{2}}{\binom{n}{2}-\binom{k}{2}}\right)^m \\ &= O(1)\cdot\left(\frac{1-s^2-(1+\gamma)^2s^2+(1-\delta)^2s^2}{1-s^2}\right)^m = O(1)\cdot\left(1-\frac{s^2(\gamma+\delta)(2+\gamma-\delta)}{1-s^2}\right)^m.\end{aligned}$$

Since $m = dn/2$ and $d = (1+\varepsilon)\ln d/d$, the elementary inequality $\ln(1-x) \le -x$ yields

$$\frac{1}{n}\ln\mathcal{E} \le \frac{d}{2}\ln\left(1-s^2(\gamma+\delta)(2+\gamma-\delta)\right) \le -s(\gamma+\delta)\left(1+\frac{\gamma-\delta}{2}\right)(1+\varepsilon)\ln d. \tag{71}$$

Finally, plugging (70) and (71) into (69), we get for $d \ge d_0$ large enough

$$\begin{aligned}\frac{1}{n}\ln\mathbb{E}[X] = \frac{1}{n}\ln\mathcal{H}+\frac{1}{n}\ln\mathcal{E} &\le s(\gamma+\delta)\left[1+\ln\ln d-\ln d-\left(1+\frac{\gamma-\delta}{2}\right)(1+\varepsilon)\ln d\right] \\ &\le s(\gamma+\delta)\left[1+\ln\ln d-\left(\varepsilon+\frac{\gamma-\delta}{2}\right)\ln d\right] \\ &\le s(\gamma+\delta)\left[1+\ln\ln d-\frac{\varepsilon_d}{2}\ln d\right] \quad \text{[by our assumption (65) and } \gamma,\delta] \\ &\le -s(\gamma+\delta) \quad [\text{as } \varepsilon_d = 3\ln\ln d/\ln d] \\ &\le -s\varepsilon_d \le -1/d \quad [\text{as } \gamma \ge \varepsilon \text{ and } s \ge \ln d/d].\end{aligned}$$

Thus, the assertion follows from Markov's inequality. ∎

Theorem 6 follows directly from Lemma 43 and Theorem 18.

8. PROOF OF THEOREM 7

Let $\varepsilon_d \to 0$ slowly. Throughout this section we assume that

$$(1+\varepsilon_d)\frac{\ln d}{d}\cdot n \le \mathbb{E}\left[\mu(G(n,m),\lambda)\right] \le (2-\varepsilon_d)\frac{\ln d}{d}\cdot n. \tag{72}$$

The proof of Theorem 7 is based on a conductance-type argument, similar in spirit to the ones used in [13,25]. We are going to show that the Metropolis process can be "trapped" in a tiny "cluster" of independent sets from which it is likely to escape only after an exponential number of steps. More specifically, we already know (from Theorem 3) that the large independent sets of our random graph shatter into an exponential number of tiny "clusters". Think of them as the peaks in a mountain range. We are going to show that, typically, to pass from one peak to another, the Metropolis process has to follow a narrow "ridge" consisting of relatively small independent sets. Moreover, under the stationary distribution the total mass of the "ridges" that from one cluster to the others typically is tiny by comparison to the mass of the cluster itself.

To be more specific, let

$$K = \left\{k : \left|\mathbb{E}\left[\mu(G_{n,m},\lambda)\right] - k\right| \le 4n/d\right\}. \tag{73}$$

We show that $\bigcup_{k\in K}\mathcal{S}_k$ can be partitioned into parts $\mathcal{C}_1,\ldots,\mathcal{C}_N$ disconnected with each other. That is, it is impossible for the process to move from one part to another without using independent sets of size much smaller than the minimum $k \in K$.

The "typical" independent sets in $\bigcup_{k\in K}\mathcal{S}_k$, belong only to some $\mathcal{C}_i$, for $i \in [N]$. We consider a process that starts from such a typical independent set, i.e. it will start from some $\mathcal{C}_i$. Then the time for the chain to reach equilibirium depends heavily on the number of transitions that are required to escape from $\mathcal{C}_i$. As we are going to show, this time is typically exponentially large. This will imply that the mixing time is exponentially large, too.

Before showing Theorem 7 we provide some auxiliary results. The following proposition shows that for a given parameter λ the stationary distribution of the Metropolis process concentrates on a small range of sizes of independent sets.

Proposition 44. *With probability at least* $1 - 2\exp\left[-n/(2d^2\ln^4 d)\right]$ *the random graph* $G = G(n,m)$ *has the following property.*

For an independent set $\mathcal{I}$ *chosen from the stationary distribution of the Metropolis process on* G *we have*

$$\mathbb{P}[|\mathcal{I}| \notin K] \le \exp(-n/d) \tag{74}$$

(where in (74) probability is taken over the choice of $\mathcal{I}$ *only).*

The proof of Proposition 44 appears in Section 8.1.

Lemma 45. *W.h.p. the random graph* $G = G(n,m)$ *has the following property. The set* $\bigcup_{k\in K}\mathcal{S}_k(G)$ *admits a partition into classes* $\mathcal{C}_1,\ldots,\mathcal{C}_N$ *such that the following three statements hold.*

C1. *The distance between any two independent sets in different classes is at least* 2.
C2. *For a random set $\mathcal{I}$ chosen from the stationary distribution of the Metropolis process we have*

$$\mathbb{P}[\mathcal{I} \in \mathcal{C}_i] \leq 5\exp\left(-n/(2d^2\ln^4 d)\right) \quad \textit{for each } i \leq i \leq N.$$

C3. *Furthermore,* $\mathbb{P}[\mathcal{I} \in \bigcup_{1\leq i\leq N} \mathcal{C}_i] \geq 1 - 5\exp\left(-n/(2d^2\ln^4 d)\right)$.

The proof of Lemma 45 appears in Section 8.2.

Proof of Theorem 7. Let K be as in (73) and assume that $G = G_{n,m}$ is such that $\bigcup_{k\in K} \mathcal{S}_k(G)$ has a partition $\mathcal{C}_1, \ldots, \mathcal{C}_N$ satisfying **C1–C3** in Lemma 45. We are going to show that the mixing time of the Metropolis process exceeds $\exp\left(n/d^3\right)$. The proof is by contradiction. Thus, assume that the mixing time of the Metropolis process is $T \leq \exp\left(n/d^3\right)$. Let $\mathcal{I}_t$ be the state of the Metropolis process at time ($t \geq 0$).

Let $t_1 = n^2T$ and $t_2 = 2n^2T$. Since T is the mixing time, for any $t_1 \leq t \leq t_2$ the distribution of $\mathcal{I}_t$ is extremely close to the stationary distribution. More precisely, if $\mathcal{I}_\infty$ chosen from the stationary distribution, then for any $t \in [t_1, t_2]$ we have

$$\|\mathcal{I}_t - \mathcal{I}_\infty\|_{tv} \leq \exp\left(-n^2\right).$$

Therefore, **C3** implies that for any $t \in [t_1, t_2]$,

$$\mathbb{P}\left[\mathcal{I}_t \notin \bigcup_{1\leq i\leq N} \mathcal{C}_i\right] \leq \mathbb{P}\left[\mathcal{I}_\infty \notin \bigcup_{1\leq i\leq N} \mathcal{C}_i\right] + \|\mathcal{I}_t - \mathcal{I}_\infty\|_{tv} \leq 6\exp\left[-n/(2d^2\ln^4 d)\right].$$

Applying the union bound, we get for $d \geq d_0$ large enough

$$\begin{aligned}\mathbb{P}\left[\exists t \in [t_1, t_2] : \mathcal{I}_t \notin \bigcup_{1\leq i\leq N} \mathcal{C}_i\right] &\leq 6\exp\left(-\frac{n}{2d^2\ln^4 d} + n/d^3\right)\\ &\leq \exp\left(-\frac{n}{3d^2\ln^4 d}\right). \qquad (75)\end{aligned}$$

In other words, we have shown that to get from $\mathcal{I}_{t_1}$ to $\mathcal{I}_{t_2}$, the Metropolis process very likely only passes through independent sets from $\bigcup_{1\leq i\leq N} \mathcal{C}_i$.

Most likely, the two independent sets $\mathcal{I}_{t_1}, \mathcal{I}_{t_2}$ belong to different classes of the partition $\mathcal{C}_1, \ldots, \mathcal{C}_N$, because the time difference $t_2 - t_1 = n^2T$ is much bigger than the mixing time T. Formally, if $\mathcal{I}_\infty$ is chosen from the stationary distribution and i_1 such that $\mathcal{I}_{t_1} \in \mathcal{C}_{i_1}$, then by **C2**

$$\mathbb{P}\left[\mathcal{I}_{t_2} \in \mathcal{C}_{i_1}\right] \leq \mathbb{P}\left[\mathcal{I}_\infty \in \mathcal{C}_{i_1}\right] + \left\|\mathcal{I}_{t_2-t_1} - \mathcal{I}_\infty\right\|_{tv} \leq 2\exp(-n/(3d^2\ln^4 d)). \qquad (76)$$

Combining (75) and (76), we thus get

$$\mathbb{P}[\exists i, j \in [N], i \neq j : \mathcal{I}_{t_1} \in \mathcal{C}_i \wedge \mathcal{I}_{t_2} \in \mathcal{C}_j] \geq 1 - \exp(-n/(3d^2\ln^4 d)). \qquad (77)$$

Thus, assume that there are two distinct $i, j \in [N]$ such that $\mathcal{I}_{t_1} \in \mathcal{C}_i$ and $\mathcal{I}_{t_2} \in \mathcal{C}_j$. Let $t > t_1$ be the first time that $\mathcal{I}_t \notin \mathcal{C}_i$. Then by definition of the Metropolis process, $dist(\mathcal{I}_t, \mathcal{I}_{t-1}) \leq 1$.

Consequently, $\mathcal{I}_t \notin \bigcup_{l\in N} \mathcal{C}_l$ because otherwise there would be two independent sets in different classes at distance one. Thus,

$$\mathbb{P}[\exists i,j \in [N], i \neq j : \mathcal{I}_{t_1} \in \mathcal{C}_i \wedge \mathcal{I}_{t_2} \in \mathcal{C}_j] \leq \mathbb{P}\left[\exists t_1 \leq t \leq t_2 : \mathcal{I}_t \notin \bigcup_{1\leq i\leq N} \mathcal{C}_i\right],$$

in contradiction to (75) and (77). ∎

8.1. Proof of Proposition 44

For a graph G, let

$$R_G(k,\lambda) = |\mathcal{S}_k(G)|\lambda^k.$$

It is easy to deduce from the definition of Metropolis process (see e.g. [25]) that for any set of integers $\mathcal{A}$ it holds that

$$\mathbb{P}[|\mathcal{I}| \in \mathcal{A}] \propto \sum_{k\in\mathcal{A}} R_G(k,\lambda).$$

Therefore, we have

$$\mathbb{P}[|\mathcal{I}| \notin \mathcal{A}] = \frac{\sum_{k\notin\mathcal{A}} R_G(k,\lambda)}{\sum_k R_G(k,\lambda)} \leq \frac{\sum_{k\notin\mathcal{A}} R_G(k,\lambda)}{\sum_{k\in\mathcal{A}} R_G(k,\lambda)}. \tag{78}$$

Consider some λ that satisfies (72). Then, Proposition 44 will follow by bounding appropriately the rightmost ratio above, for $\mathcal{A} = K$ (as defined in (73)) and G being a typical instance of $G(n,m)$.

Remark. Observe that when the graph G is distributed as in $G(n,m)$ the quantity R_G is a random variable which depends *only on the underlying graph.*

Before proving the proposition we need some preliminary results. With the parameter $\lambda > 0$ and the expected degree d in mind, for any $x \in (0,1)$ we define the following function:

$$f_\lambda(x) = -(x\ln x + (1-x)\ln(1-x)) + \frac{d}{2}\ln(1-x^2) + x\ln\lambda.$$

It is straightforward to verify that $\frac{1}{n}\ln\mathbb{E}[R_G(k,\lambda)] \sim f_\lambda(k/n)$. $f_\lambda(x)$ is twice differentiable, as a matter of fact it holds that

$$f'_\lambda(x) = \ln(1-x) - \ln x - d\frac{x}{1-x^2} + \ln\lambda \tag{79}$$

$$f''_\lambda(x) = -\frac{1}{x(1-x)} - d\frac{1+x^2}{(1-x^2)^2}. \tag{80}$$

For any λ and $x \in (0,1)$ it holds that $f''_\lambda(x) < 0$. That is, $f'_\lambda(x)$ is strictly decreasing. Furthermore, if for given λ, d there exists $x_0 \in (0,1)$ such that

$$\lambda = \frac{x_0}{1-x_0}\exp\left(d\frac{x_0}{1-x_0^2}\right), \tag{81}$$

then $f_\lambda(x_0)$ is a global maximum for f_λ. Since $f'_\lambda(x)$ is strictly decreasing, for any given $x' \in (0,1)$ and d, we can find unique $\lambda_0 > 0$ such that $f_{\lambda_0}(x)$ is maximized when $x = x'$.

Claim 46. *Take $x_0 \in (0,1)$ and let λ be such that $f_\lambda(x)$ is maximized for $x = x_0$. Then for any x such that $|x - x_0| = t$ it holds that*

$$f_\lambda(x) \le f_\lambda(x_0) - \frac{1}{2}t^2 d.$$

Proof. From (80) it is easy to show that for any $x \in (0,1)$, it holds that $f''_\lambda(x) < -d$. Also, for any $x \in (0,1)$ we can find appropriate $\xi \in [0,1)$ such that

$$\begin{aligned} f_\lambda(x) &= f_\lambda(x_0) + (x - x_0) f'_\lambda(x_0) + \frac{(x - x_0)^2}{2} f''_\lambda(\xi) \\ &\le f_\lambda(x_0) - \frac{(x - x_0)^2}{2} d, \quad [\text{as } f'_\lambda(x_0) = 0 \text{ and } f''_\lambda(x) < -d] \end{aligned}$$

as promised. ∎

Let λ_c be such that $f_{\lambda_c}(x)$ is maximized for $x = (1+c)\ln d/d$.

Lemma 47. *For $c \in [\epsilon_d, 1 - \epsilon_d]$ and $k = (1+c)\frac{\ln d}{d}n$, it holds that*

$$\mathbb{P}\left[R_{G(n,m)}(k,\lambda_c) \le \exp\left(-14n\sqrt{\ln^5 d/d^3}\right) \cdot \mathbb{E}[R_{G(n,m)}(k,\lambda_c)]\right] \le \exp\left[-n/(2d^2 \ln^4 d)\right].$$

Proof. The lemma follows directly from Proposition 20. ∎

Lemma 48. *For $c \in [\epsilon_d, 1 - \epsilon_d]$, let $k = (1+c)\frac{\ln d}{d}n$ and*

$$\mathcal{R}_c = \exp\left(-14n\sqrt{\ln^5 d/d^3}\right)\mathbb{E}[R_{G(n,m)}(k,\lambda_c)].$$

It holds that

$$\mathbb{P}\left[\sum_{k':|k-k'|>\frac{1.9n}{d}} R(k',\lambda_c) \ge \exp(-n/d)\,\mathcal{R}_c\right] \le \exp(-n/(2d)).$$

Proof. Observe that for any integer $0 \le k' \le 2n\ln d/d$ it holds that

$$\mathbb{E}[R_{G(n,m)}(k',\lambda_c)] = \exp\left[f(k'/n)n + o(n)\right].$$

Since the function $f_{\lambda_c}(x)$ is increasing for every $0 \le x < (1+c)\ln d/d$ and decreasing for $(1+c)\ln d/d < x < 1$, for $k_0 = k - 1.9n/d$ and sufficiently large n it holds that

$$\mathbb{E}[R_{G(n,m)}(k_0,\lambda_c)] \ge \max_{k':|k'-k|>1.9n/d}\left\{\mathbb{E}[R_{G(n,m)}(k',\lambda_c)]\right\}. \tag{82}$$

Furthermore, using Claim 46 we get that

$$\mathbb{E}[R_{G(n,m)}(k_0,\lambda_c)] \le \mathbb{E}[R_{G(n,m)}(k,\lambda_c)]\exp\left(-\frac{1.8n}{d} + o(n)\right). \tag{83}$$

Let $Q = \sum_{k':|k-k'|>\frac{1.9n}{d}} R(k', \lambda_c)$. It holds that

$$\begin{aligned}\mathbb{E}[Q] &= \sum_{k':|k-k'|>\frac{1.9n}{d}} \mathbb{E}[R(k', \lambda_c)] \\ &\leq n\mathbb{E}[R_{G(n,m)}(k_0, \lambda_c)] \quad \text{[from (82)]} \\ &\leq \mathbb{E}[R_{G(n,m)}(k, \lambda_c)] \exp\left(-\frac{1.8n}{d} + o(n)\right). \quad \text{[from (83)]} \end{aligned} \tag{84}$$

The lemma follows by applying Markov's inequality. That is, for sufficiently large d it holds that

$$\begin{aligned}\mathbb{P}\left[Q \geq \exp(-n/d)\,\mathcal{R}_c\right] &\leq \mathbb{P}\left[Q \geq \mathbb{E}[Q]\exp\left(\frac{n}{2d}\right)\right] \quad \text{[from (84)]} \\ &\leq \exp\left(-\frac{n}{2d}\right), \quad \text{[from Markov's inequality]}\end{aligned}$$

as promised. ∎

Proof of Proposition 44. Let $c \in (\epsilon_d, 1-\epsilon_d)$, for $\epsilon_d \to 0$.

Observe that quantity $\mu(G, \lambda)$ for fixed λ and G distributed as in $G(n, m)$ is a random variable which depends only on the graph G. We are going to show that for λ_c it holds that

$$\mathbb{P}\left[\left|\mu(G(n,m), \lambda_c) - (1+c)\frac{\ln d}{d}n\right| > \frac{1.95n}{d}\right] \leq \exp\left[-n/(2d)\right]. \tag{85}$$

Observe that once we have the above tail bound, the proposition follows easily from Lemma 48. In particular (85) implies that

$$\left|\mathbb{E}[\mu(G(n,m), \lambda_c)] - (1+c)\frac{\ln d}{d}n\right| \leq \frac{1.95n}{d} + n\exp\left[-n/(2d)\right]. \tag{86}$$

Also, from Lemma 48 and (78) we have the following: Consider the Metropolis process with underlying graph $G(n, m)$ and parameter λ_c. Then, with probability at least $1-\exp(-n/(2d))$ over the graph instances $G(n, m)$, if we choose $\mathcal{I}$ according to the stationary distribution of the Metropolis process, then

$$\mathbb{P}[\mathcal{I} \notin \hat{K}] \leq \exp(-n/d), \tag{87}$$

where $\hat{K} = \{k \in \mathbb{N} : |k - (1+c)\frac{\ln d}{d}n| \leq \frac{1.9n}{d}\}$. The proposition follows from (86) and (87).

It remains to show (85). By definition we have that for any fixed graph G it holds that $\mu(G, \lambda) = \frac{1}{Z(G,\lambda)}\sum_{k=1}^{n} kR_G(k, \lambda)$, where $Z(G, \lambda) = \sum_{k=1}^{n} R_G(k, \lambda)$. From Lemma 48 we have that with probability at least $1 - \exp\left[-n/(2d)\right]$ over the graph instances $G(n, m)$ it holds that

$$0 \leq Z(G(n,m), \lambda_c) - \sum_{k\in\hat{K}} R_{G(n,m)}(k, \lambda_c) \leq \exp(-n/d)\left(\sum_{k\in\hat{K}} R_{G(n,m)}(k, \lambda_c)\right) \tag{88}$$

and

$$0 \leq \sum_{k=0}^{n} kR_{G(n,m)}(k, \lambda_c) - \sum_{k\in\hat{K}} kR_{G(n,m)}(k, \lambda_c) \leq n\exp(-n/(2d))\left(\sum_{k\in\hat{K}} kR_{G(n,m)}(k, \lambda_c)\right). \tag{89}$$

Combining (88) and (89) we get that with probability at least $1-\exp[-n/(2d)]$ over $G(n,m)$ it holds that

$$\mu(G(n,m),\lambda_c) = (1+r)\sum_{k\in\hat{K}} k\frac{R_{G(n,m)}(k,\lambda_c)}{\sum_{k\in\hat{K}} R_{G(n,m)}(k,\lambda_c)},$$

for some $|r| \le 2n\exp(-n/(2d))$. Then, it is elementary to verify that the summation on the r.h.s. is a convex combination of values of k in K. That is, the summation is at most $\max\{k\in\hat{K}\}$ and at least $\min\{k\in\hat{K}\}$. Then (85) follows. ∎

8.2. Proof of Lemma 45

As in (50) let

$$\mathcal{Z}_{d,k} = \left\{(G,\sigma)\in\Lambda_k(n,m) : |\mathcal{S}_k(G)| \ge \mathbb{E}|\mathcal{S}_k(G(n,m))|\cdot\exp\left(-14n\sqrt{\ln^5 d/d^3}\right)\right\}.$$

Lemma 49. *Let $(G,\sigma)\in\Lambda_k(n,m)$ be distributed as in $\mathcal{U}_k(n,m)$, for $k\in K$, where K and $\mu(G,\lambda)$ are as in (73) and (1), respectively. The set $\bigcup_{k\in K}\mathcal{S}_k(G)$ admits a partition into classes $\mathcal{C}_1,\ldots,\mathcal{C}_N$ such that*

1. $\mathbb{P}[\sigma\in\mathcal{C}_i|\mathcal{Z}_{d,k}] \le \exp[-n/(2d^{1.2})]$, *for any* $i\in[N]$
2. $\mathbb{P}[\sigma\notin\bigcup_{i\in[N]}\mathcal{C}_i|\mathcal{Z}_{d,k}] \le \exp(-n/d)$
3. *The distance between two independent sets in different classes is at least* 2.

Proof of Lemma 45 (Given Lemma 49). Consider $G(n,m)$ and the Metropolis process with parameter λ, for λ as in (72). Let the independent set $\mathcal{I}$ be chosen according to the stationary distribution of the process.

Conditional that $|\mathcal{I}|=k$, $\mathcal{I}$ is distributed uniformly at random in $\mathcal{S}_k(G(n,m))$, for any k. For any $A\subset 2^{[n]}$ it holds that

$$\begin{aligned}\mathbb{P}[\mathcal{I}\in A|\mathcal{Z}_{d,k}] &\le \mathbb{P}[\mathcal{I}\in A|\mathcal{Z}_{d,k},|\mathcal{I}|\in K] + \mathbb{P}[|\mathcal{I}|\notin K|\mathcal{Z}_{d,k}]\\ &\le \max_{k\in K}\{\mathbb{P}[\mathcal{I}\in A|\mathcal{Z}_{d,k},|\mathcal{I}|=k]\} + \mathbb{P}[|\mathcal{I}|\notin K|\mathcal{Z}_{d,k}].\end{aligned}$$

the last inequality follows from the fact that $\mathbb{P}[\mathcal{I}\in A|\mathcal{Z}_{d,k},|\mathcal{I}|\in K]$ is a convex combination of $\mathbb{P}[\mathcal{I}\in A|\mathcal{Z}_{d,k},|\mathcal{I}|=j]$ for $j\in K$. Also, it holds that

$$\begin{aligned}\mathbb{P}[|\mathcal{I}|\notin K|\mathcal{Z}_{d,k}] &\le \frac{\mathbb{P}[|\mathcal{I}|\notin K]}{\mathbb{P}[\mathcal{Z}_{d,k}]} \le 2\mathbb{P}[|\mathcal{I}|\notin K] \quad \text{[from Proposition 20]}\\ &\le 4\exp\left(-n/(2d^2\ln^4 d)\right) \quad \text{[from Proposition 44]}.\end{aligned}$$

Hence,

$$\mathbb{P}[\mathcal{I}\in A|\mathcal{Z}_{d,k}] \le \max_{k\in K}\{\mathbb{P}[\mathcal{I}\in A|\mathcal{Z}_{d,k},|\mathcal{I}|=k]\} + 4\exp\left(-n/(2d^2\ln^4 d)\right). \qquad (90)$$

Also, from the law of total probability we get that

$$\begin{aligned}\mathbb{P}[\mathcal{I}\in A] &\le \mathbb{P}[\mathcal{I}\in A|\mathcal{Z}_{d,k}]+\mathbb{P}[\mathcal{Z}_{d,k}^c] \quad [\mathcal{Z}_{d,k}^c \text{ is the complement of } \mathcal{Z}_{d,k}]\\ &\le \mathbb{P}[\mathcal{I}\in A|\mathcal{Z}_{d,k}]+\exp\left(-n/(2d^2\ln^4 d)\right) \quad [\text{from Proposition 20}]\\ &\le \max_{k\in K}\{\mathbb{P}[\mathcal{I}\in A|\mathcal{Z}_{d,k},|\mathcal{I}|=k]\}+5\exp\left(-n/(2d^2\ln^4 d)\right) \quad [\text{from (90)}]. \quad (91)\end{aligned}$$

The statement $\mathbf{C}_1$ holds from the statement 3 in Lemma 49. Setting $A=\mathcal{C}_i$ in (91) and using Statement 1 from Lemma 49, we get the Statement $\mathbf{C}_2$. Similarly, Statement $\mathbf{C}_3$ follows by setting $A=\left(\bigcup_{k\in K}\mathcal{S}_k\right)\backslash\left(\bigcup_{i\in[N]}\mathcal{C}_i\right)$ in (91) and using Statement 2 from Lemma 49. ∎

8.3. Proof of Lemma 49

Consider a uniform pair $(G,\sigma)\in\Lambda_k(n,m)$, for some $k\in K$. For fixed $0<\beta<1$, and $|\gamma|<1$, let $Z_{k,\beta,\gamma}$ be the number of independent sets $\tau\in\mathcal{S}_{(1+\gamma)k}(G)$ such that $|\sigma\cap\tau|=(1-\beta)k$. Also, for $0<\beta_1<\beta_2<1$ consider $\vec{\beta}=(\beta_1,\beta_2)$ and let the independent set σ be called $(\vec{\beta},\gamma,\delta)$-*good* if G has no independent set τ such

- $\tau\in S_{k,\gamma}=\bigcup_{t=(1-\gamma)\cdot k}^{(1+\gamma)k}\mathcal{S}_t(G)$
- $(1-\beta_2)k<|\sigma\cap\tau|<(1-\beta_1)k$

while $|\{\tau'\in S_{k,\gamma}:(\sigma\cap\tau')>(1-\beta_1)k\}|<\exp(-\delta n)\,|\mathcal{S}_k(G)|$.

Lemma 50. *For $\psi(x)$ is as defined in statement of Proposition 37 and $s=k/n$, it holds that*

$$\frac{1}{n}\ln\mathbb{E}_{\mathcal{P}_k(n,m)}[Z_{k,\beta,\gamma}]\le\psi(\beta)+\xi(\beta,\gamma)+o(1),$$

where

$$\begin{aligned}\xi(x,y)=s[-x\ln(1+y/x)+y(1-\ln s-\ln(x+y))]\\ +\frac{d}{2}\ln\left(1-s^2\frac{2y+y^2}{1-(1+2x-x^2)s^2}\right).\end{aligned}$$

Proof. Let $\tau\subset V$ be such that $|\tau|=(1+\gamma)k$ and $|\sigma\cap\tau|=(1-\beta)k$. With application of inclusion/exclusion principle we get that the total number of graphs with m edges in which σ and τ are independent sets equals

$$\binom{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\beta)k}{2}}{m}.$$

Since G is chosen uniformly at random among all $\binom{\binom{n}{2}-\binom{k}{2}}{m}$ graphs on n vertices and m edges such that σ is an independent set, we get that

$$\begin{aligned}\mathbb{P}[\tau\text{ is independent}]&=\binom{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\beta)k}{2}}{m}\Bigg/\binom{\binom{n}{2}-\binom{k}{2}}{m}\\ &=\prod_{i=0}^{m-1}\frac{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\beta)k}{2}-i}{\binom{n}{2}-\binom{k}{2}-i}\end{aligned}$$

$$\leq \left(\frac{\binom{n}{2}-\binom{k}{2}-\binom{(1+\gamma)k}{2}+\binom{(1-\beta)k}{2}}{\binom{n}{2}-\binom{k}{2}}\right)^m$$
$$\leq \left(1-\frac{(1+\gamma)^2k^2-(1-\beta)^2k^2}{n^2-k^2}+O(1/n)\right)^m$$
$$\leq O(1)\cdot\left(1-s^2\frac{(1+\gamma)^2-(1-\beta)^2}{1-s^2}\right)^m \quad [\text{as } k=sn].$$

The total number of ways to choose a set of vertices τ of size $(1+\gamma)k$ such that $|\sigma\cap\tau| = (1-\beta)k$ is equal to $\binom{k}{(1-\beta)k}\binom{n-k}{(\gamma+\beta)k}$. By the linearity of expectation, we get that

$$\begin{aligned}\mathbb{E}[Z_{k,\beta,\gamma}] &= O(1)\cdot\binom{k}{(1-\beta)k}\cdot\binom{n-k}{(\gamma+\beta)k}\cdot\left(1-s^2\frac{(1+\gamma)^2-(1-\beta)^2}{1-s^2}\right)^m\\ &\leq O(1)\cdot\binom{k}{\beta k}\cdot\binom{n-k}{(\gamma+\beta)k}\cdot\left(1-s^2\frac{(1+\gamma)^2-(1-\beta)^2}{1-s^2}\right)^m\\ &\leq O(1)\cdot\left(\frac{e}{\beta}\right)^{\beta k}\cdot\left(\frac{(1-s)e}{(\gamma+\beta)s}\right)^{(\gamma+\beta)k}\cdot\left(1-s^2\frac{(1+\gamma)^2-(1-\beta)^2}{1-s^2}\right)^{dn/2}\\ &\leq O(1)\cdot\left(\frac{e}{\beta}\right)^{\beta k}\cdot\left(\frac{e}{(\gamma+\beta)s}\right)^{(\gamma+\beta)k}\cdot\left(1-s^2\frac{(1+\gamma)^2-(1-\beta)^2}{1-s^2}\right)^{dn/2}.\end{aligned} \tag{92}$$

By definition (see Proposition 37), it holds that

$$\exp(\psi(\beta)n) = \left(\frac{e}{\beta}\right)^{\beta k}\left(\frac{e}{\beta s}\right)^{\beta k}\left(1-s^2\frac{1-(1-\beta)^2}{1-s^2}\right)^{dn/2}. \tag{93}$$

Combining (92) and (93) we get that

$$\frac{\mathbb{E}[Z_{k,\beta,\gamma}]}{\exp(\psi(\beta)n)} \leq O(1)\left(\frac{\beta}{\beta+\gamma}\right)^{\beta k}\left(\frac{e}{(\gamma+\beta)s}\right)^{\gamma k}\left(1-s^2\frac{2\gamma+\gamma^2}{1-(2-(1-\beta)^2)s^2}\right)^{dn/2}, \tag{94}$$

since

$$\left(\frac{(1-s)e}{(\gamma+\beta)s}\right)^{(\gamma+\beta)k}\Bigg/\left(\frac{(1-s)e}{(\gamma+\beta)s}\right)^{\beta k} = \left(\frac{\beta}{\beta+\gamma}\right)^{\beta k}\left(\frac{(1-s)e}{(\gamma+\beta)s}\right)^{\gamma k} \quad \text{and}$$
$$\left(1-s^2\frac{(1+\gamma)^2-(1-\beta)^2}{1-s^2}\right)^{dn/2}\Bigg/\left(1-s^2\frac{1-(1-\beta)^2}{1-s^2}\right)^{dn/2}$$
$$= \left(1-s^2\frac{2\gamma+\gamma^2}{1-(2-(1-\beta)^2)s^2}\right)^{dn/2}.$$

Taking the logarithm and dividing by n the quantities in (94) we get the lemma. ∎

Lemma 51. *There is $\epsilon_d \to 0$ such that for $(1+\epsilon_d)n\ln d/d \leq k \leq (2-\epsilon_d)n\ln d/d$ the following is true: For $\gamma = 4/\ln d$, and $\delta = 1/d^{1.2}$ there is $\vec{\beta}\in[0,1]^2$ such that*

$$\mathbb{P}_{\mathcal{U}_k(n,m)}[(G,\sigma) \text{ is } (\vec{\beta},\gamma,\delta)\text{-good}|\mathcal{Z}_{k,d}] \geq 1-\exp(-n/d).$$

Proof. Let $\epsilon_d = 100 \ln\ln d/\ln d$. Assume that $k = (1+q)\ln d/d$ for some $q \in [\epsilon_d, 1-\epsilon_d]$. Consider the functions $\psi(x)$ and $\xi(x,y)$ as defined in the statement of Lemma 50. In what follows take $b = \frac{20}{\ln d}$. Let

$$\mathcal{H}_k(x) = \psi(x) + \max_{(\beta,\rho)\in\mathbb{A}} \xi(\beta,\rho),$$

where $\mathbb{A} = \{(\beta,\rho) \in [0,b]\times[-\gamma,\gamma] | \beta+\rho \geq 0\}$. Our choices for b and γ ensure that for any $(\beta,\rho)\in\mathbb{A}$ it holds that

$$\begin{aligned}
\xi(\beta,\rho) &= s[-\beta\ln(1+\rho/\beta) + \rho(1-\ln s - \ln(\beta+\rho))] \\
&\quad + \frac{d}{2}\ln\left(1 - s^2\frac{2\rho+\rho^2}{1-(1+2\beta-\beta^2)s^2}\right) \\
&\leq s[-(\beta+\rho)\ln(\beta+\rho) + \beta\ln(\beta) + \rho(1-\ln s)] - ds^2\rho - ds^2\rho^2/2. \\
&\leq s\left[25\frac{\ln\ln d}{\ln d} + \rho\,(1-\ln s - ds)\right] \\
&\qquad\qquad [-x\ln x \text{ is increasing for } 0<x<1/e \text{ and } \beta\ln\beta<0] \\
&\leq s\left[25\frac{\ln\ln d}{\ln d} + \gamma q\ln d\right] \quad [\text{as } s = (1+q)\ln d/d \text{ and } \rho \geq -\gamma] \\
&< 5qs \quad [\text{as } q \geq 100\ln\ln d/\ln d]. \qquad (95)
\end{aligned}$$

Using (95) and (47), from Lemma 40, we get that

$$\mathcal{H}_k(b) \leq -13qs \leq -1300\ln\ln d/d. \qquad (96)$$

The function $\mathcal{H}_k(x)$ is continuous, therefore there exist $b_2 > b_1 > 0$ and ζ such that

$$\begin{aligned}
\sup_{b_1<\beta<b_2} \mathcal{H}_k(\beta) &< -1300\ln\ln d/d - \zeta \\
\sup_{b>\beta} \mathcal{H}_k(\beta) &< -s\ln(s) - (1-s)\ln(1-s) + \frac{d}{2}\ln(1-s^2) - 15s - \zeta.
\end{aligned}$$

The last relation follows from (48), of Lemma 40 and (95).

Let $\Psi_{k,b_1,b_2}(G,\sigma)$, be the number of $\tau \in \bigcup_{t=(1-\gamma)k}^{(1+\gamma)k} \mathcal{S}_t(G)$ such that $(1-b_2)k \leq |\sigma\cap\tau| \leq (1-b_1)k$. Then, Markov's inequality yields

$$\mathbb{P}_{\mathcal{P}_k(n,m)}[\Psi_{k,b_1,b_2} > 0] \leq \mathbb{E}_{\mathcal{P}_k(n,m)}[\Psi_{k,b_1,b_2}] = \sum_{i\in A}\sum_{j\in B}\mathbb{E}_{\mathcal{P}_k(n,m)}[Z_{k,j/k,i/k}]$$

where $A = [-4k/\ln d, 4k/\ln d]$ and $B = [b_1k, b_2k]$. Using Lemma 50 we get

$$\mathbb{P}_{\mathcal{P}_k(n,m)}[\Psi_{k,b_1,b_2} > 0] \leq \exp\left[n\cdot\left(\sup_{b_2\leq\beta\leq b_1}\mathcal{H}(\beta) + o(1)\right)\right] \leq \exp(-10n/d). \qquad (97)$$

Let $\Psi_{k,b_1}(G,\sigma)$ be the number of $\tau \in \bigcup_{t=(1-\gamma)k}^{(1+\gamma)k} S_t(G)$ such that $|\sigma\cap\tau| > (1-b_1)k$. Moreover, let

$$\begin{aligned}
\mu &= \mathbb{E}[|\mathcal{S}_k(G)|]\exp\left(-n/d^{1.2}\right) \\
&= \exp\left[n\left(-s\ln s - (1-s)\ln(1-s) - \frac{d}{2}\ln(1-s^2) - n/d^{1.2} + o(1)\right)\right].
\end{aligned}$$

For the derivation in the second line, see in the proof of Corollary 39. For $A' = [-4k/\ln d, 4k/\ln d]$ and $B' = [0, b_1 k)$, it holds that

$$\mathbb{P}_{\mathcal{P}_k(n,m)}[\Psi_{k,b_1} > \mu] \le \frac{\mathbb{E}_{\mathcal{P}_k(n,m)}[\Psi_{k,b_1}]}{\mu} \le \sum_{i\in A'}\sum_{j\in B'} \frac{\mathbb{E}_{\mathcal{P}_k(n,m)}[Z_{k,j/k,i/k}]}{\mu}$$
$$\le \frac{1}{\mu}\exp\left[n\left(\sup_{\beta<b_1}\mathcal{H}(\beta)+o(1)\right)\right] \le \exp\left(-15n/d\right).$$

The lemma follows by noting the following for $\delta = 14\sqrt{\ln^5 d/d^3}$,

$$\begin{aligned}
&\mathbb{P}_{\mathcal{U}_k(n,m)}\left[(G,\sigma) \text{ is not } (\vec{\beta},\gamma,\delta)\text{-good}|\mathcal{Z}_{d,k}\right]\\
&\quad\le \mathbb{P}_{\mathcal{U}_k(n,m)}\left[\Psi_{k,b_1} > \mu \text{ or } \Psi_{k,b_1,b_2} > 0|\mathcal{Z}_{d,k}\right]\\
&\quad\le (1-o(1))\mathbb{P}_{\mathcal{P}_k(n,m)}\left[\Psi_{k,b_1} > \mu \text{ or } \Psi_{k,b_1,b_2} > 0|\mathcal{Z}_{d,k}\right]\cdot\exp\left[14n\sqrt{\ln^5 d/d^3}\right]\\
&\quad\le \exp(-n/d),
\end{aligned}$$

as claimed. ∎

Now, Lemma 49 follows from the above lemma and by using arguments very similar to those in the proof of Proposition 37.

9. REMAINING PROOF

9.1. Proof of Lemma 9

This is a standard counting argument. The random graph $G^*(n,m)$ is obtained by choosing one of the n^{2m} possible sequences of vertex pairs uniformly at random. Out of these n^{2m} sequences, precisely $2^m\binom{n}{2}_m$ sequences induce simple graphs with m edges (where $(\cdot)_m$ denotes the falling factorial). Indeed, each of the $\binom{\binom{n}{2}}{m}$ simple graph with m edges can be turned into a sequence of pairs by ordering the edges arbitrarily (a factor $m!$), and then choosing for each edge in which order its vertices appear in the sequence (a factor 2^m). Hence, letting Σ denote the event that $G^*(n,m)$ is a simple graph with m edges, we see that

$$\begin{aligned}
\mathbb{P}[G^*(n,m)\in\Sigma] = \frac{2^m\binom{n}{2}_m}{n^{2m}} &= \left(\frac{2}{n^2}\right)^m\cdot\prod_{j=0}^{m-1}\binom{n}{2}-j = \prod_{j=0}^{m-1}1-\frac{1}{n}-\frac{2j}{n^2}\\
&= \exp\left[\sum_{j=0}^{m}\ln\left(1-\frac{1}{n}-\frac{2j}{n^2}\right)\right]\\
&\sim \exp\left[-\sum_{j=0}^{m}\frac{1}{n}+\frac{2j}{n^2}\right] \quad [\text{using } \ln(1-x) = -x+O(x^2) \text{ as } x\to 0]\\
&\sim \exp\left[-c-c^2\right]. \qquad (98)
\end{aligned}$$

Furthermore, given that the event Σ occurs, $G^*(n,m)$ is just a uniformly distributed (simple) graph with m edges. Therefore, (98) yields

$$\begin{aligned}\mathbb{P}\left[G(n,m)\in\mathcal{A}\right] = \mathbb{P}\left[G^*(n,m)\in\mathcal{A}|\,\Sigma\right] &\leq \frac{\mathbb{P}\left[G^*(n,m)\in\mathcal{A}\right]}{\mathbb{P}\left[G^*(n,m)\in\Sigma\right]}\\ &\sim \exp\left[c+c^2\right]\mathbb{P}\left[G^*(n,m)\in\mathcal{A}\right],\end{aligned}$$

as claimed.

9.2. Proof of Corollary 10

Let $Q\subset V$ be a set of size k, and let $Z_Q(G)=1$ if Q is independent in G, and set $Z_Q(G)=0$ otherwise. The total number of sequences of m vertex pairs such that Q is an independent set in the corresponding graph $G^*(n,m)$ equals $(n^2-k^2)^m$ (just avoid the k^2 pairs of vertices in Q). Hence,

$$\mathbb{E}\left[Z_Q(G^*(n,m))\right] = \frac{(n^2-k^2)^m}{n^{2m}}, \quad \text{and similarly} \tag{99}$$

$$\mathbb{E}\left[Z_Q(G(n,m))\right] = \binom{\binom{n}{2}-\binom{k}{2}}{m}\Bigg/\binom{\binom{n}{2}}{m} = \frac{(\binom{n}{2}-\binom{k}{2})_m}{\binom{n}{2}_m}\,. \tag{100}$$

Combining (99) with (100) and using $\ln(1-x)=-x+O(x^2)$ as $x\to 0$, we obtain

$$\begin{aligned}\frac{\mathbb{E}\left[Z_Q(G^*(n,m))\right]}{\mathbb{E}\left[Z_Q(G(n,m))\right]} &= \frac{2^m\binom{n}{2}_m}{n^{2m}}\cdot\frac{(n^2-k^2)^m}{2^m(\binom{n}{2}-\binom{k}{2})_m} \overset{(98)}{\sim} \exp(-c-c^2)\frac{(n^2-k^2)^m}{2^m(\binom{n}{2}-\binom{k}{2})_m}\\ &= \exp\left[-c-c^2-\sum_{j=0}^{m-1}\ln\left(1-\frac{n-k}{n^2-k^2}-\frac{2j}{n^2-k^2}\right)\right]\\ &\sim \exp\left[-c-c^2+\frac{m(n-k)}{n^2-k^2}+\frac{m^2}{n^2-k^2}\right]\\ &= \exp\left[-c-c^2+\frac{c}{1+k/n}+\frac{c^2}{1-(k/n)^2}\right]\\ &= \exp\left[-\frac{ck}{n+k}+\frac{c^2k^2}{n^2-k^2}\right].\end{aligned}$$

Hence, by the linearity of expectation,

$$\begin{aligned}\mathbb{E}[Z_k(G^*(n,m))] &= \binom{n}{k}\cdot\mathbb{E}\left[Z_Q(G^*(n,m))\right]\\ &= \exp\left[-\frac{ck}{n+k}+\frac{c^2k^2}{n^2-k^2}\right]\cdot\binom{n}{k}\mathbb{E}\left[Z_Q(G(n,m))\right]\\ &= \exp\left[-\frac{ck}{n+k}+\frac{c^2k^2}{n^2-k^2}\right]\mathbb{E}[Z_k(G(n,m))].\end{aligned}$$

Taking logarithms and recalling that $k\leq 0.99n$ completes the proof.

REFERENCES

[1] D. Achlioptas and A. Coja-Oghlan, Algorithmic barriers from phase transitions, In proceedings of FOCS, Philadelphia, PA, USA, 2008, pp. 793–802.

[2] D. Achlioptas and M. Molloy, The analysis of a list-coloring algorithm on a random graph, In proceedings of FOCS, Miami Beach, Florida, USA, 1997, pp. 204–212.

[3] N. Alon, M. Krivelevich, and B. Sudakov, Finding a large hidden clique in a random graph, Random Struct Algor 13 (1998), 457–466.

[4] J. Barbier, F. Krzakala, Zdeborov'a L., and Pan Zhang, The hard-core model on random graphs revisited. Available at: http://arxiv.org/abs/1306.4121. Access date: 2013.

[5] N. Bhatnagar, A. Sly, and P. Tetali, Reconstruction threshold for the hardcore model, In proceedings of APPROX-RANDOM, Barcelona, Spain, 2010, pp. 434–447.

[6] A. Coja-Oghlan, Finding large independent sets in polynomial expected time, Comb Proba Comput 15 (2006), 731–751.

[7] B. Bollobás and P. Erdős, Cliques in random graphs, Math Proc Comb Phil Soc 80 (1976), 419–427.

[8] A. Coja-Oghlan and C. Efthmiou, On independent sets in random graphs, arXiv:1007.1378, 2010.

[9] V. Dani and C. Moore, Independent sets in random graphs from the weighted second moment method, In proceedings of APPROX-RANDOM, Princeton University, NJ, USA, 2011, pp. 472–482.

[10] Y. Deshpande and A. Montanari, Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time, arXiv:1304.7047, 2013.

[11] J. Ding, A. Sly, and N. Sun, Maximum independent sets on random regular graphs, arXiv:1310.4787, 2013.

[12] M. E. Dyer and A. M. Frieze, Fast algorithms for some random NP-hard problems, J Algor 10 (1986), 451–489.

[13] M. Dyer, A. Frieze, and M. Jerrum, On counting independent sets in sparse graphs, SIAM J Comput 31 (2002), 1527–1541.

[14] P. Erdős, Some remarks on the theory of graphs, Bull Am Math Soc 53 (1947), 292–294.

[15] U. Feige and R. Krauthgamer, Finding and certifying a large hidden clique in a semirandom graph, Random Struct Algor 16 (2000), 195–208.

[16] A. M. Frieze, On the independence number of random graphs, Discrete Math 81 (1990), 171–175.

[17] A. M. Frieze and C. McDiarmid, Algorithmic theory of random graphs, Random Struct Algor 10 (1997), 5–42.

[18] A. Frieze and S. Suen, Analysis of two simple heuristics on a random instance of k-SAT, J Algor 20 (1996), 312–355.

[19] Y. Fu and P. W. Anderson, Applications of statistical mechanics to NP-complete problems in combinatorial optimization, J Phys A 19 (1986), 1605.

[20] D. Gamarnik, T. Nowick, and G. Swirscsz, Maximum weight independent sets and matchings in sparse random graphs, Random Struct Algor 28 (2005), 76–106.

[21] D. Gamarnik and M. Sudan, Limits of Local Algorithm over sparse random graphs, arXiv preprint arXiv:1304.1831, 2013.

[22] A. Gaudilliére, B. Scoppola, E. Scoppola, and M. Viale, Phase transitions for the cavity approach to the clique problem on random graphs, J Stat Phys 145 (2011), 1127–1155.

[23] G. R. Grimmett and C. J. H. McDiarmid, On colouring random graphs, Math Proc Cambridge Philos Soc 77 (1975), 313–324.

[24] H. Hatami, L. Lovász, and B. Szegedy, Limits of local-global convergent graph sequences, arXiv preprint, Available at: http://arxiv.org/abs/1205.4356, 2012.

[25] M. R. Jerrum, Large cliques elude the Metropolis process, Random Struct Algor 3 (1992), 347–359.

[26] R. M. Karp, The probabilistic analysis of some combinatorial search algorithms, F. Traub (Editor), In Algorithms and Complexity: New Directions and Recent Results, Academic Press, USA, 1976, pp. 1–19.

[27] R. Karp and M. Sipser, Maximum matchings in sparse random graphs, In proceedings of FOCS, Nashville, Tennessee, USA, 1981, pp. 364–375.

[28] S. Kirkpatrick, C. Gelatt, and M. Vecchi, Optimisation by simulated annealing, science 220 (1983), 671–680.

[29] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjianc, and L. Zdeborova, Gibbs states and the set of solutions of random constraint satisfaction problems, Proc Natl Acad Sci USA 104 (2007), 10318–10323.

[30] L. Kučera, Graphs with small chromatic number are easy to color, Inf Process Lett 30 (1989), 233–236.

[31] D. Matula, The largest clique size in a random graph, Technical Report Department of Computer Science, Southern Methodist University, Dallas, Texas, 1976.

[32] M. Mezard, G. Parisi, and R. Zecchina, Analytic and Algorithmic Solution of Random Satisfiability Problems, Science 297 (2002), 812–815.

[33] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, J Chem Phys 21 (1953), 1087–1092.

[34] M. Mézard and A. Montanari, Information, physics and computation, Oxford University Press, UK, 2009.

[35] A. Montanari, R. Restrepo, and P. Tetali, Reconstruction and clustering thresholds in random CSPs, SIAMJ Discrete Math 25 (2011), 771–808.

[36] R. Motwani and P. Raghavan, Randomized algorithms, Cambridge University Press, 1995.

[37] S. Janson, T. Luczak, and A. Ruciński, Random graphs, Wiley and Sons, Inc., 2000.

[38] B. Rossman, The monotone complexity of k-clique on random graphs, In Proceedings of FOCS, Las Vegas, Nevada, USA, 2010, pp. 193–201.

[39] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, Instit Hautes Études Sci Publ Math 81 (1995), 73–205.

[40] J. S. Turner, Almost all k-colorable graphs are easy to color, J Algo 9 (1988), 63–82.

[41] H. S. Wilf, Backtrack: An expected O(1) time algorithm for the graph coloring problem, Inf Proc Lett 18 (1984), 119–122.

[42] N. Wormald, The differential equation method for random graph processes and greedy algorithms, M. Karoński and H. J. Proömel (Editors), In Lectures on Approximation and randomized algorithms, 1999, pp. 73–155.

[43] G. Biroli and M. Mézard, Lattice glass models, Physical Rev Lett 88 (2001), 025501.