LOCAL ALGORITHMS FOR INDEPENDENT SETS ARE HALF-OPTIMAL

BY MUSTAZEE RAHMAN[1] AND BÁLINT VIRÁG[2]

University of Toronto

We show that the largest density of factor of i.i.d. independent sets in the d-regular tree is asymptotically at most $(\log d)/d$ as $d \to \infty$. This matches the lower bound given by previous constructions. It follows that the largest independent sets given by local algorithms on random d-regular graphs have the same asymptotic density. In contrast, the density of the largest independent sets in these graphs is asymptotically $2(\log d)/d$. We prove analogous results for Poisson–Galton–Watson trees, which yield bounds for local algorithms on sparse Erdős–Rényi graphs.

1. Introduction. Local algorithms are randomized algorithms that run in parallel at each vertex of a graph by using only local information around each vertex. They produce important structures in large graphs, such as independent sets, matchings and colourings, with only constant running time (see [7–9, 11, 14, 16, 19, 20] and the references therein). In this paper, we investigate local algorithms for high density independent sets in random d-regular graphs. We find an optimal bound for the density of such independent sets as the degree becomes large. It turns out that in this limit local algorithms can only yield independent sets with half the maximum possible density. The motivation for our work comes from questions that arose in the theory of graph limits (see [11, 16] and the references therein). In particular, Hatami, Lovász, and Szegedy conjecture ([16] Conjecture 7.13) that most optimization problems over typical, sparse graphs can be solved by local algorithms.

We use the following notion of local algorithm introduced in [16]. The input to the algorithm is a graph G. The algorithm decorates G by putting i.i.d. labels on the vertices. The output is $(f(i(v)); v \in G)$ where f depends on the isomorphism class $i(v)$ of the labelled, rooted r-neighbourhood of v for some fixed r. The process $(f(i(v)); v \in G)$ generated by the local algorithm will be called a *factor of i.i.d.* process. See Section 2 for a more formal definition.

While the conjecture of Hatami, Lovász and Szegedy was verified for maximal matchings [5, 9, 20] and covariance structures [1], Gamarnik and Sudan [15]

showed that it fails for maximal independent sets. An independent set in a graph is a set of vertices that have no edges between them.

It is known from [2] that for each d the size density of the largest independent sets in a random d-regular graph on n vertices converges almost surely as $n \to \infty$. Furthermore, Bollobás [3] and McKay [21] proved that with high probability the size density of the largest independent sets in random d-regular graphs is at most $2(\log d)/d$ for every $d \geq 3$. Frieze and Łuczak [12, 13] provided lower bounds of matching asymptotic order for large d. Recently, precise formulae were given for large d by Ding, Sly and Sun [10]. On the other hand, several authors have produced local algorithms on d-regular graphs of large girth that yield independent sets of density $(\log d)/d$ for large d (see [14, 19, 23]). These algorithms use greedy strategies to construct independent sets and can be easily adapted to random d-regular graphs.

Thus, for large d, the density of the largest independent sets in random d-regular graphs is of order $2(\log d)/d$ while local algorithms have only produced independent sets with density of order $(\log d)/d$. The conjecture of Hatami, Lovász and Szegedy would imply that local algorithms can in fact produce independent sets in random d-regular graphs of density $2(\log d)/d$.

Gamarnik and Sudan [15] disprove this conjecture by showing that for large d local algorithms cannot find independent sets in random d-regular graphs of density larger than $(1 + \frac{1}{\sqrt{2}})(\log d)/d$. Their crucial step is to prove that with high probability any two high density independent sets in random d-regular graphs have a substantially large or substantially small intersection. This observation was guided by predictions from statistical physics regarding the solution-space geometry of constraint satisfaction problems [22]. In particular, the so called clustering phenomenon is expected to hold for independent sets in sparse random graphs. Rigorous results have been established in this regard by Coja-Oghlan and Efthymiou [6] and in the aforementioned work of Ding, Sly and Sun [10]. It is shown that for large enough d, some of the properties that determine clustering emerge for independent sets in random d-regular graphs at size density $(\log d)/d$.

In this paper, we analyze the intersection densities of many independent sets in random regular graphs. We show that with high probability (i.e., with probability tending to one as the size of the graphs tends to infinity) the intersection densities must satisfy various inequalities. These structural results on the admissible intersection densities imply quantitative bounds on the density of independent sets that can be generated from local algorithms. With the help of these inequalities we prove that for any $\varepsilon > 0$, local algorithms cannot find independent sets in random d-regular graphs of density larger than $(1 + \varepsilon)(\log d)/d$ if d is sufficiently large. In practice, iterative search algorithms that use local moves at each step fail to find independent sets with density exceeding the critical threshold of $(\log d)/d$ in random d-regular graphs. Our result provides some evidence as to why this is the case.

We also consider local algorithms for independent sets in Poisson–Galton–Watson trees. These yield local algorithms for independent sets in sparse Erdős–Rényi graphs. We prove that the maximal density of local independent sets in a Poisson–Galton–Watson tree of expected degree λ is of asymptotic order $(\log \lambda)/\lambda$ as $\lambda \to \infty$. The aforementioned results of Bollobás [3], Frieze and Łuczak [12, 13] show that the largest independent sets in Erdős–Rényi graphs of average degree λ have density of asymptotic order $2(\log \lambda)/\lambda$ as $\lambda \to \infty$.

The challenge in proving upper bounds to the density of local independent sets in Poisson–Galton–Watson trees is showing that the randomness of the tree does not provide local algorithms with extra power. Also, in order to show the existence of local independent sets having density close to $(\log \lambda)/\lambda$ we employ a coupling argument that produces independent sets in Poisson–Galton–Watson trees from independent sets in regular trees.

1.1. *Organization of the paper.* In Section 2, we define the notion of a local algorithm for independent sets in the d-regular tree and relate it to local algorithms on finite d-regular graphs. Our main result about the density of local independent sets in regular trees is stated in Theorem 2.1. In Section 2.1, we introduce the key inequality, stated in Theorem 2.2, that is satisfied by the intersection densities of any finite collection of local independent sets in the d-regular tree. Using this inequality, we prove Theorem 2.1 in Section 2.2. In Section 3, we prove Theorem 2.2 by employing combinatorial arguments involving random regular graphs. In Section 4, we state and prove our main result, Theorem 4.1, on local independent sets in Poisson–Galton–Watson trees.

2. Local algorithms for independent sets in regular graphs. We define the notion of local algorithms for independent sets in regular trees. Let $\mathbb{T}_d$ denote the rooted d-regular tree, and for $r \geq 0$ let $\mathbb{T}_{d,r}$ denote the rooted r-neighbourhood of $\mathbb{T}_d$. A *labelling* of $\mathbb{T}_d$ is a vector $x \in [0, 1]^{\mathbb{T}_d}$, and a *random labelling* is a labelling X where the co-ordinates $X(v)$, $v \in \mathbb{T}_d$, are independent, uniformly distributed random variables on $[0, 1]$. A *factor* on $\mathbb{T}_d$ is a measurable function $f : [0, 1]^{\mathbb{T}_d} \to \{0, 1\}$ (w.r.t. the Borel σ-algebra) such that f is invariant under all root preserving automorphisms of $\mathbb{T}_d$. In other words, f is spherically symmetric about the root. We say that f depends on the r-neighbourhood of the root if f is defined on $[0, 1]^{\mathbb{T}_{d,r}}$.

Any factor f on $\mathbb{T}_d$ defines a set-valued stochastic process I on $\mathbb{T}_d$ as follows. Any graph automorphism ϕ of $\mathbb{T}_d$ acts on labels $x \in [0, 1]^{\mathbb{T}_d}$ by $\phi \cdot x(v) = x(\phi^{-1}(v))$. Since the automorphism group of $\mathbb{T}_d$ acts transitively on the vertices, given any vertex v let ϕ_v be an automorphism that maps v to the root. For a random labelling X of $\mathbb{T}_d$, define $I(v) = f(\phi \cdot X)$. Due to f being invariant under root preserving automorphisms I is well defined. We call I a factor of i.i.d. process on $\mathbb{T}_d$. Note that the distribution of I is invariant under the action of the automorphism

group of $\mathbb{T}_d$ (however, factor of i.i.d. processes are more restrictive than invariant process).

A factor of i.i.d. independent set in $\mathbb{T}_d$ is a factor of i.i.d. process I such that I is an independent in $\mathbb{T}_d$ with probability 1. Since the distribution of $I(v)$ does not depend on the vertex v, we define the *density* of I as

$$\text{density}(I) = \mathbb{P}\big[I(\text{root}) = 1\big] = \mathbb{E}\big[f(X)\big].$$

It is easy to see that a factor that generates independent sets can be approximated by similar factors that depend on finite size neighbourhoods of the root (see [16], Section 12). In this manner, a factor of i.i.d. independent set of density ρ can be approximated by finite neighbourhood factor of i.i.d. independent sets whose densities converge to ρ. Hence, there is no harm in assuming that all our factors for independent sets depend on finite size neighbourhoods of the root.

Example: *A construction of Lauer and Wormald.*   In [19], the authors analyze the following algorithm that generates factor of i.i.d. independent sets in $\mathbb{T}_d$. Fix $p \in (0, 1)$ and an integer $k \geq 1$. Let $U_0 = V(\mathbb{T}_d)$ and for $1 \leq i \leq k$ do the following. Let $S_i \subset U_{i-1}$ be a random subset resulting from the output of a Bernoulli percolation on U_{i-1} at density p. Set $U_i = U_{i-1} \setminus (S_i \cup N(S_i))$, where $N(S_i)$ is the one-neighbourhood of the set S_i in $\mathbb{T}_d$. Consider the subset $I' = \bigcup_{i=1}^k S_i$. I' may not be an independent set only because some S_i may contain both vertices along an edge. If a vertex $v \in I'$ has one of its neighbours also included in I' then exclude v from I'. This results in an independent set $I \subset I'$.

The random set I is a factor of i.i.d. independent set since the decision rule to include a vertex is (deterministically) invariant of the vertex, and the rule depends on the outcome of k independent Bernoulli percolations on $\mathbb{T}_d$. A little thought shows that the factor for I depends only on the $(k + 1)$-neighbourhood of a vertex. Lauer and Wormald show that taking $k = \frac{c}{p}$ and then letting $p \to 0$, followed by $c \to \infty$, results in independent sets whose densities converge to $\beta(d) := \frac{1-(d-1)^{-2/(d-2)}}{2}$. A simple analysis shows that $\frac{\log(d-1)}{d-2} - 2(\frac{\log(d-1)}{d-2})^2 \leq \beta(d) \leq \frac{\log(d-1)}{d-2}$.

From trees to finite graphs.   Given a factor of i.i.d. independent set I in $\mathbb{T}_d$, we can construct a (random) independent set in any d-regular graph G on n vertices via the following procedure. Recall that I uses a factor f that computes $I(v)$ by looking only at the isomorphism class of the labelled r-neighbourhood of v in $\mathbb{T}_d$. We begin with a random labelling X of the vertices of G. Given any vertex $v \in G$ if its $(r + 1)$-neighbourhood, $N_{r+1}(G, v)$, is a tree then set $I_G(v) = f(X(u); u \in N_r(G, v))$. This is allowed since $N_r(G, v) = \mathbb{T}_{d,r}$ by assumption. Otherwise, set $I_G(v) = 0$.

We verify that I_G is an independent set. For any edge (u, v) such that both $N_{r+1}(G, u)$ and $N_{r+1}(G, v)$ are trees, the pair of values $(I_G(u), I_G(v))$ is the same as the values $(I(a), I(b))$ for any edge (a, b) of $\mathbb{T}_d$ with labels $(X_i; i \in$

$N_r(G, u) \cup N_r(G, v))$ lifted to $N_r(\mathbb{T}_d, a) \cup N_r(\mathbb{T}_d, b)$ in the natural way. Thus, $(I_G(u), I_G(v)) \neq (1, 1)$ as required. On the other hand, if one of $N_{r+1}(G, u)$ or $N_{r+1}(G, v)$ is not a tree then at least one of $I_G(u)$ or $I_G(v)$ is 0. Consequently, I_G is an independent set in G. Notice also that if $B(G)$ is the number of vertices of G whose $(r + 1)$-neighbourhood is not a tree then the expected size density of I_G is $\mathbb{E}[|I_G|/n] = \text{density}(I)(1 - B(G)/n)$.

We are going to use this technique to project factor of i.i.d. independent sets from $\mathbb{T}_d$ to finite, d-regular graphs. The resulting processes on the finite graphs will be referred to as independent sets from local algorithms. We are now prepared to state our main result for independent sets in d-regular graphs. Define the quantity α_d as follows:

$$(2.1) \quad \alpha_d \frac{\log d}{d} = \sup\{\text{density}(I) : I \text{ is a factor of i.i.d. independent set in } \mathbb{T}_d\}.$$

THEOREM 2.1. *The following inequality holds for α_d:*

$$\limsup_{d \to \infty} \alpha_d \leq 1.$$

In other words, for any $\varepsilon > 0$ there exists a D such that if $d > D$ then there are no local algorithms that generate independent sets in $\mathbb{T}_d$ having density larger than $(1 + \varepsilon)\frac{\log d}{d}$.

2.1. *Key inequality for intersection densities of local independent sets.* We prove Theorem 2.1 by way of contradiction. Assuming otherwise, we pass to a subsequence in d and assume that for some $\alpha > 1$ we have $\alpha_d > \alpha$ for every d along the subsequence. It follows that for each such d there exists a factor of i.i.d. independent set in $\mathbb{T}_d$, say I_d, such that the density of I_d is $\alpha\frac{\log d}{d}$. Let $f_d : [0, 1]^{\mathbb{T}_d} \to \{0, 1\}$ denote the factor associated to I_d. Recall we may assume that f_d depends on a finite size neighbourhood of the root. So we assume that f_d depends on the r_d-neighbourhood of $\mathbb{T}_d$.

Now we construct many copies of I_d that are correlated with each other via a parameter that we will control. Fix $p \in [0, 1]$ and let $S_d = S_d(p)$ denote a random subset of the vertices of $\mathbb{T}_d$ generated by a Bernoulli percolation of density p. Also, let X_i for $i \geq 0$ denote independent random labellings of $\mathbb{T}_d$. We construct independent sets $I_{d,i}$ for $i \geq 0$ by letting $I_{d,i}$ be generated from the factor f_d with labels $X_0(v)$ for $v \in V(\mathbb{T}_d) \setminus S_d$ and $X_i(v)$ for $v \in S_d$.

As f_d is defined on $[0, 1]^{\mathbb{T}_{d,r_d}}$ it follows that $I_{d,i}(\text{root})$ depends only on the labels $X_0(v)$ and $X_i(v)$ for $v \in V(\mathbb{T}_{d,r_d})$ and the subset $S_{d,r_d} = S_d \cap V(\mathbb{T}_{d,r_d})$. Also, the joint distribution of the sets $I_{d,i}$ is exchangeable over i and each $I_{d,i}$ follows the distribution of I_d. This implies that the intersection of any k of these local independent sets have a common density, which we denote $\alpha_{k,d,p}\frac{\log d}{d}$. Note that $\alpha_{1,d} = \alpha$. To reduce notational clutter, we will denote r_d by r and S_{d,r_d} by S until the end of Section 2.

We will achieve a contradiction by first showing that these intersection densities are constrained to satisfy an inequality for each k. Second, we will violate these inequalities by tuning the coupling parameter p (under the assumption that $\alpha > 1$). The next theorem introduces these key inequalities. Their proof, discussed in Section 3, is based on a structure theorem about independent sets in random d-regular graphs.

THEOREM 2.2.   *For each $k \geq 1$, the quantities $\alpha_{i,d,p}$ for $1 \leq i \leq k$ satisfy the following*:

$$(2.2) \qquad \liminf_{d \to \infty} \inf_{p \in [0,1]} \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \alpha_{i,d,p}(2 - \alpha_{i,d,p}) \geq 0.$$

Theorem 2.2 is proved by counting the expected number of k-tuples of independent sets $(I_1, \ldots, I_k)$ in random d-regular graphs such that their intersection densities are close to the quantities $\alpha_{i,d,p} \frac{\log d}{d}$ for $1 \leq i \leq k$. We show that if (2.2) fails then the probability of observing such k-tuples of independent sets in random d-regular graphs is vanishingly small as the size of the graphs tend to infinity. On the other hand, Lemma 3.4 implies that the existence of the local independent sets $(I_{d,1}, \ldots, I_{d,k})$ allows us to observe such k-tuples of independent sets in random d-regular graphs with high probability and so (2.2) must hold.

Relation to the approach of Gamarnik and Sudan.   In their paper [15], Gamarnik and Sudan derive inequality (2.2) for $k = 2$. The $k = 2$ case gives

$$\inf_{p \in [0,1]} 2\alpha(2 - \alpha) - \alpha_{2,d,p}(2 - \alpha_{2,d,p}) \geq 0 \qquad \text{for all large } d.$$

To minimize this in p, we certainly want to set $\alpha_{2,d,p} = 1$ for every d. It turns out that $\alpha_{2,d,p}$ is continuous in p (see Lemma 2.3) with $\alpha_{2,d,0} = \alpha$ and $\alpha_{2,d,1} = \alpha^2(\frac{\log d}{d})$. So if $\alpha > 1$ then for all large d we can find a value of p such that $\alpha_{2,d,p} = 1$. This implies that the density α satisfies $\alpha(2-\alpha) \geq 1/2$, or equivalently, that $\alpha \leq 1 + \frac{1}{\sqrt{2}}$. This is the conclusion of Gamarnik and Sudan.

We may also analyze (2.2) for $k = 3$ to conclude that $\alpha \leq 1 + \frac{1}{\sqrt{3}}$. Indeed, we have that $3\alpha(2 - \alpha) - 2\alpha_{2,d,p}(2 - \alpha_{2,d,p}) + \alpha_{3,d,p}(2 - \alpha_{3,d,p}) \geq 0$ for large d. If $\alpha > 1$, then for all large d we may choose a value of p such that $\alpha_{2,d,p} = 1$. Also, observe that $\alpha_{3,d,p}(2 - \alpha_{3,d,p}) \leq 1$. Thus, we conclude from (2.2) that $3\alpha(2 - \alpha) - 2 + 1 = 3\alpha(2 - \alpha) - 1 \geq 0$. This implies that $\alpha \leq 1 + \frac{1}{\sqrt{3}}$.

We do not know how to solve the minimization problem in p exactly for $k \geq 4$. In order to analyze (2.2) for large values of k, we are going to make a choice of p for each d (and fixed k) that allows us to bound the sum in (2.2) from above as $d \to \infty$. This upper bound is going to be a quantity that we can analyze in the large k limit. From there, we will derive a contradiction to the assumption that $\alpha > 1$.

2.2. *Proof of Theorem* 2.1 *from Theorem* 2.2. Given the setup thus far we begin by interpreting the $\alpha_{k,d,p}$ in a probabilistic manner. We show that the values $\frac{\alpha_{k,d,p}}{\alpha_{1,d}}$ can be realized as the moments of a random variable. This random variable is defined on a new probability space, which is obtained from the original probability space by essentially restricting to the support of the factor f_d. Formally, the new sample space is the set $\{f_d(X_0) \equiv 1\}$ considered as a subset of the joint sample space of $X_0, X_1, \ldots,$ and S. The new σ-algebra is the restriction of the σ-algebra generated by $S, X_0, X_1, \ldots$ to $\{f_d(X_0) \equiv 1\}$. The new expectation operator $\mathbb{E}^*$ is defined by

$$\mathbb{E}^*[U] = \frac{\mathbb{E}[f_d(X_0)U]}{\mathbb{E}[f_d(X_0)]}$$

for any random variable U defined on $\{f_d(X_0) \equiv 1\}$.

If $\mathcal{F}$ is a σ-algebra such that $f_d(X_0)$ is $\mathcal{F}$-measurable, then for any random variable U defined on the original probability space, we have

$$\mathbb{E}^*[U|\mathcal{F}] = \mathbb{E}[U|\mathcal{F}].$$

This is to be interpreted by restricting $\mathcal{F}$ to $\{f_d(X_0) \equiv 1\}$ on the left-hand side and the random variable $\mathbb{E}[U|\mathcal{F}]$ to $\{f_d(X_0) \equiv 1\}$ on the right-hand side. To prove this, suppose that Z is a $\mathcal{F}$-measurable random variable. Then

$$\begin{aligned}
\mathbb{E}^*[Z\mathbb{E}[U|\mathcal{F}]] &= \frac{\mathbb{E}[f_d(X_0)Z\mathbb{E}[U|\mathcal{F}]]}{\mathbb{E}[f_d(X_0)]} \\
&= \frac{\mathbb{E}[\mathbb{E}[f_d(X_0)ZU|\mathcal{F}]]}{\mathbb{E}[f_d(X_0)]} \quad \left(f_d(X_0) \text{ and } Z \text{ are } \mathcal{F}\text{-measurable}\right) \\
&= \frac{\mathbb{E}[f_d(X_0)ZU]}{\mathbb{E}[f_d(X_0)]} \\
&= \mathbb{E}^*[ZU].
\end{aligned}$$

Define a sequence of $[0,1]$-valued random variables $Q_{d,p} = Q_d(S, X_0)$, which we denote the *stability*, on the restricted probability space as follows. Let

$$f_{d,i} = f_d(X_0(v); v \notin S, X_i(v); v \in S) = I_{d,i}(\text{root}).$$

Set

$$Q_{d,p} = \mathbb{E}^*[f_{d,1}|X_0, S] = \mathbb{E}[f_{d,1}|X_0, S].$$

Roughly speaking, the stability is the conditional probability, given the root is included in the independent set, that it remains to be included after re-randomizing the labels on S.

The key observation is that $\mathbb{E}^*[Q_{d,p}^{k-1}] = \frac{\alpha_{k,d,p}}{\alpha_{1,d}}$ for $k \geq 1$. Indeed, as $Q_{d,p}$ has the same distribution as $\mathbb{E}[f_{d,i}|X_0, S]$ for every i we have that

$$\mathbb{E}^*[Q_{d,p}^{k-1}] = \frac{\mathbb{E}[f_{d,0}(\mathbb{E}[f_{d,1}|X_0, S])^{k-1}]}{\mathbb{E}[f_{d,0}]} = \frac{\mathbb{E}[f_{d,0}\prod_{i=1}^{k-1}\mathbb{E}[f_{d,i}|X_0, S]]}{\mathbb{E}[f_{d,0}]}.$$

The random variables $f_{d,i}$ are independent of each other conditioned on (X_0, S). Hence,

$$\prod_{i=1}^{k-1} \mathbb{E}[f_{d,i}|X_0, S] = \mathbb{E}\left[\prod_{i=1}^{k-1} f_{d,i}|X_0, S\right].$$

Furthermore, $f_{d,0}$ is measurable w.r.t. (X_0, S) and so we conclude that

$$\mathbb{E}\left[f_{d,0}\left(\prod_{i=1}^{k-1} \mathbb{E}[f_{d,i}|X_0, S]\right)\right] = \mathbb{E}\left[f_{d,0}\prod_{i=1}^{k-1} f_{d,i}\right] = \text{density}\left(\bigcap_{i=1}^{k} I_{d,i}\right).$$

Consequently, $\mathbb{E}^*[Q_{d,p}^{k-1}] = \frac{\text{density}(\bigcap_{i=1}^{k} I_{d,i})}{\text{density}(I_{d,1})} = \frac{\alpha_{k,d,p}}{\alpha_{1,d}}$.

Henceforth, all expectations involving $Q_{d,p}$ will simply be denoted by $\mathbb{E}$ instead of $\mathbb{E}^*$. We will need the following lemma regarding the continuity of the stability in terms of the coupling parameter p.

LEMMA 2.3. *Let* $g : [0, 1] \to \mathbb{R}$ *be a continuous function. The moment* $\mathbb{E}[g(Q_{d,p})]$ *is a continuous function of* p. *When* $p = 0$, $\mathbb{E}[g(Q_{d,0})] = g(1)$, *and when* $p = 1$, $\mathbb{E}[g(Q_{d,1})] = g(\alpha\frac{\log d}{d})$.

PROOF. The parameter p enters into $\mathbb{E}[g(Q_{d,p})]$ only through the random finite subset $S \subset T_{d,r}$. For each $W \subset \mathbb{T}_{d,r}$, the probability $\mathbb{P}[S = W] = p^{|W|}(1 - p)^{|T_{d,r} \setminus W|}$. This probability is a polynomial in p. By conditioning on the output of S we note that $\mathbb{E}[g(Q_{d,p})]$ can be expressed as a convex combination of terms that are free of p, namely $\mathbb{E}[g(Q_{d,p})|S = W]$, with corresponding coefficient $\mathbb{P}[S = W]$. Thus, $\mathbb{E}[g(Q_{d,p})]$ is also a polynomial in p.

When $p = 0$ the set S is empty and $f_{1,d} = f_{d,0}$. Therefore, conditioning on X_0 and restricting to $\{f_{d,0} \equiv 1\}$ forces $Q_{d,p} \equiv 1$. When $p = 1$ the set S equals $\mathbb{T}_{d,r}$, and hence $f_{1,d}$ becomes independent of the random labelling X_0, and hence of $f_{d,0}$ as well. Consequently, the conditioning has no effect and $Q_{d,p} = \mathbb{E}[f_d] = \alpha\frac{\log d}{d}$. This implies that $\mathbb{E}[g(Q_{d,0})] = g(1)$ and $\mathbb{E}[g(Q_{d,1})] = g(\alpha\frac{\log d}{d})$. □

We now translate the inequality from (2.2) in terms of the stability. Our goal is to rewrite (2.2) as an expectation of a function of the stability, which we can then analyze for large values of d and k.

Note that $\alpha_{i,d,p} = \alpha_{1,d}\mathbb{E}[Q_{d,p}^{i-1}] = \alpha\mathbb{E}[Q_{d,p}^{i-1}]$. To deal with the terms $\alpha_{i,d,p}^2$ we introduce an independent copy of $Q_{d,p}$, which we denote $R_{d,p}$. Thus, $\alpha_{i,d,p}^2 = \alpha^2\mathbb{E}[(Q_{d,p}R_{d,p})^{i-1}]$. This implies that

$$\alpha_{i,d,p}(2 - \alpha_{i,d,p}) = 2\alpha\mathbb{E}[Q_{d,p}^{i-1}] - \alpha^2\mathbb{E}[(Q_{d,p}R_{d,p})^{i-1}].$$

Observe the following identity that results from the binomial theorem:

$$(2.3) \qquad \sum_{i=1}^{k}(-1)^{i-1}\binom{k}{i}x^{i-1} = \frac{1 - (1-x)^k}{x} \qquad \text{for } 0 \le x \le 1.$$

Let $s_k(x) = \frac{1-(1-x)^k}{x}$ for $x \in [0, 1]$ and $k \geq 1$. Note that $s_k(0) = \lim_{x \to 0} s_k(x) = k$. We may now translate the inequality from (2.2) into

$$(2.4) \qquad \liminf_{d \to \infty} \inf_{p \in [0,1]} 2\alpha \mathbb{E}[s_k(Q_{d,p})] - \alpha^2 \mathbb{E}[s_k(Q_{d,p}R_{d,p})] \geq 0.$$

We make a particular choice of p for every d in order to analyze (2.4) in the large d limit. Fix a parameter $u > 0$ that we will tune later. In the statement of Lemma 2.3, take $g(x) = x^u$ for $0 \leq x \leq 1$. From the assumption that $\alpha > 1$, we employ Lemma 2.3 and deduce that for all sufficiently large d we can select a $p = p(d, u)$ such that

$$\mathbb{E}[Q_{d,p(d,u)}^u] = 1/\alpha.$$

We denote $Q_{d,p(d,u)}$ by Q_d. At this point our reasoning behind this choice is mysterious. The idea, of course, is that by choosing p this way we try to minimize the left-hand side of (2.4) in a manner that we can analyze as $k \to \infty$. The argument that follows will show that our choice is judicious.

Recall that probability distributions on $[0, 1]$ are compact with respect to convergence in distribution. Therefore, from the sequence (Q_d, R_d) we can choose a subsequence (Q_{d_i}, R_{d_i}) that converges in distribution to limiting random variables (Q, R). The random variables Q and R are independent and identically distributed with values in $[0, 1]$.

Observe that $s_k(x) = 1 + (1 - x) + \cdots + (1 - x)^{k-1}$. Thus, $s_k(x)$ is a continuous, decreasing function on $[0, 1]$ with maximum value $s_k(0) = k$ and minimum value $s_k(1) = 1$. Distributional convergence of (Q_{d_i}, R_{d_i}) to (Q, R) implies that $\mathbb{E}[s_k(Q_{d_i})] \to \mathbb{E}[s_k(Q)]$ and $\mathbb{E}[s_k(Q_{d_i}R_{d_i})] \to \mathbb{E}[s_k(QR)]$.

By passing to the subsequence d_i and taking limits in i the inequality (2.4) becomes

$$(2.5) \qquad\qquad 2\mathbb{E}[s_k(Q)] \geq \alpha \mathbb{E}[s_k(QR)].$$

This holds for every $k \geq 1$. Taking the limit as $k \to \infty$ of (2.5) results in the inequality $2\mathbb{E}[1/Q] \geq \alpha \mathbb{E}[1/Q]^2$. If $\mathbb{E}[1/Q]$ is finite then we have $\alpha \leq 2\mathbb{E}[1/Q]^{-1}$. We are thus left with the seemingly contradictory task of showing that $\mathbb{E}[1/Q]$ is finite but large. Unfortunately, distributional convergence of Q_{d_i} to Q is not sufficient to get a lower bound on $\mathbb{E}[1/Q]$. Furthermore, it is not a priori clear that this expectation is finite, or even that $\mathbb{P}[Q = 0] = 0$. To work around these difficulties, we have to control the distribution of the Q_d well enough to be able to conclude that $\mathbb{P}[Q = 0]$ is small while $\mathbb{E}[1/Q]$ is large. We derive a contradiction to the hypothesis $\alpha > 1$ from analyzing (2.5) based upon 3 cases: $\mathbb{P}[Q = 0] > 0$, or $\mathbb{P}[Q = 0] = 0$ but $\mathbb{E}[1/Q] = \infty$, or $\mathbb{E}[1/Q] < \infty$.

In order to bound $\mathbb{E}[1/Q]$ and $\mathbb{P}[Q = 0]$ we recall that we had set p such that $\mathbb{E}[Q_d^u] = 1/\alpha$ for all large d. Since $x \to x^u$ is continuous and bounded on $[0, 1]$ we conclude that $\mathbb{E}[Q^u] = \lim_i \mathbb{E}[Q_{d_i}^u] = 1/\alpha$. Now, an upper bound on $\mathbb{E}[Q^u]$ implies a lower bound on $\mathbb{E}[1/Q]$ due to $\mathbb{E}[1/Q] \geq \mathbb{E}[Q^u]^{-1/u}$, which follows

from Jensen's inequality. Also, a lower bound on $\mathbb{E}[Q^u]$ gives a upper bound on $\mathbb{P}[Q = 0]$ because $\mathbf{1}_{Q=0} \leq 1 - Q^u$. We now analyze (2.5) over all k based upon the 3 cases mentioned in the previous paragraph.

Case 1: $\mathbb{P}[Q = 0] = q > 0$. In this case, most of the contribution to $\mathbb{E}[s_k(Q)]$ results from $\{Q = 0\}$. More precisely, $\frac{s_k(x)}{k} = \mathbf{1}_{x=0} + \frac{s_k(x)}{k}\mathbf{1}_{x>0}$, and $\frac{s_k(x)}{k}\mathbf{1}_{x>0} \to 0$ as $k \to \infty$. Also, $\frac{s_k(x)}{k} \in [0, 1]$ for all k and $x \in [0, 1]$. Therefore, from the bounded converge theorem we deduce that $\mathbb{E}[s_k(Q)/k] \to \mathbb{P}[Q = 0]$ as $k \to \infty$, and similarly, $\mathbb{E}[s_k(QR)/k] \to \mathbb{P}[QR = 0]$. The latter probability is $2q - q^2$ due to Q and R being independent and identically distributed. Upon dividing the inequality in (2.5) through by k and taking a limit, we conclude that

$$2q - \alpha(2q - q^2) \geq 0, \quad \text{or equivalently that} \quad \alpha \leq \frac{2}{2 - q}.$$

For $x \in [0, 1]$, we have that $\mathbf{1}_{x=0} \leq 1 - x^u$. It follows from this that $q \leq 1 - \mathbb{E}[Q^u] = 1 - 1/\alpha$. Thus,

$$\alpha \leq \frac{2}{2 - q} \leq \frac{2}{1 + \alpha^{-1}}.$$

Simplifying the latter inequality gives $\alpha \leq 1$; a contradiction.

Case 2: $\mathbb{P}[Q = 0] = 0$ but $\mathbb{E}[\frac{1}{Q}] = \infty$. In this case, most of the contribution to $\mathbb{E}[s_k(Q)]$ occurs when Q is small. Note that $s_k(x) \nearrow 1/x$ as $k \to \infty$. Hence, the monotone convergence theorem implies that $\mathbb{E}[s_k(Q)] \to \infty$ as $k \to \infty$.

Fix $0 < \varepsilon < 1$, and write $s_k(x) = s_{k,\leq\varepsilon}(x) + s_{k,>\varepsilon}(x)$ where $s_{k,\leq\varepsilon}(x) = s_k(x)\mathbf{1}_{x\leq\varepsilon}$. Note that $s_{k,>\varepsilon}(x) \leq \varepsilon^{-1}$ for all k. We have that

$$(2.6) \qquad \mathbb{E}[s_k(Q)] = \mathbb{E}[s_{k,\leq\varepsilon}(Q)] + \mathbb{E}[s_{k,>\varepsilon}(Q)] \leq \mathbb{E}[s_{k,\leq\varepsilon}(Q)] + \varepsilon^{-1}.$$

Thus, $\mathbb{E}[s_{k,\leq\varepsilon}(Q)] \to \infty$ with k because $\mathbb{E}[s_k(Q)] \to \infty$.

We also observe from the positivity of s_k that

$$\mathbb{E}[s_k(QR)] \geq \mathbb{E}[s_k(QR); Q \leq \varepsilon, R > \varepsilon] + \mathbb{E}[s_k(QR); Q > \varepsilon, R \leq \varepsilon].$$

The latter two terms are equal by symmetry. Therefore, $\mathbb{E}[s_k(QR)]$ is at least $2\mathbb{E}[s_k(QR); Q \leq \varepsilon, R > \varepsilon]$. Now the fact that $s_k(x)$ is decreasing in x and $R \leq 1$ implies that $s_k(QR) \geq s_k(Q)$. Together with the independence of Q and R, we deduce that

$$\mathbb{E}[s_k(QR); Q \leq \varepsilon, R > \varepsilon] \geq \mathbb{E}[s_k(Q); Q \leq \varepsilon, R > \varepsilon]$$
$$= \mathbb{E}[s_{k,\leq\varepsilon}(Q)]\mathbb{P}[R > \varepsilon].$$

Consequently,

$$(2.7) \qquad\qquad \mathbb{E}[s_k(QR)] \geq 2\mathbb{E}[s_{k,\leq\varepsilon}(Q)]\mathbb{P}[Q > \varepsilon].$$

The inequality in (2.5) is $\frac{\alpha}{2} \leq \frac{\mathbb{E}[s_k(Q)]}{\mathbb{E}[s_k(QR)]}$. The bounds from (2.6) and (2.7) imply that

$$\frac{\alpha}{2} \leq \frac{\mathbb{E}[s_{k,\leq\varepsilon}(Q)] + \varepsilon^{-1}}{2\mathbb{E}[s_{k,\leq\varepsilon}(Q)]\mathbb{P}[Q > \varepsilon]}.$$

Since $\mathbb{E}[s_{k,\leq\varepsilon}(Q)] \to \infty$ with k, we can take a limit in k to conclude that

$$\alpha \leq \frac{1}{\mathbb{P}[Q > \varepsilon]}.$$

As $\varepsilon \to 0$ the probability $\mathbb{P}[Q > \varepsilon] \to \mathbb{P}[Q > 0] = 1$, by assumption. Thus, $\alpha \leq 1$; a contradiction.

Case 3: $\mathbb{E}[1/Q]$ is finite. In the final case, we again use the fact that $s_k(x)$ increases to $1/x$ for $0 \leq x \leq 1$. Hence, $s_k(Q) \nearrow 1/Q$ almost surely and $s_k(QR) \nearrow 1/(QR)$ almost surely. Taking a limit of the inequality in (2.5) and using the monotone convergence theorem, it follows that

$$2\mathbb{E}\left[\frac{1}{Q}\right] - \alpha\mathbb{E}\left[\frac{1}{QR}\right] \geq 0.$$

Since $\mathbb{E}[\frac{1}{QR}] = \mathbb{E}[\frac{1}{Q}]^2$ the above inequality reduces to

$$\alpha \leq 2\mathbb{E}\left[\frac{1}{Q}\right]^{-1} \leq 2\mathbb{E}[Q^u]^{1/u}.$$

In the last step, we have used the power-mean/Jensen's inequality. Since $\mathbb{E}[Q^u] = 1/\alpha$, we see that

$$\alpha \leq 2^{\frac{u}{u+1}}.$$

Due to the contradiction resulting from the previous two cases, we deduce that for all $u > 0$ we have $\alpha \leq 2^{\frac{u}{u+1}}$. By letting $u \to 0$, we conclude that $\alpha \leq 1$. This is the final contradiction.

3. Inequalities for intersection densities: Proof of Theorem 2.2. We will prove Theorem 2.2 by reducing it to a problem about densities of independent sets in large, finite, d-regular graphs. First, we begin with some terminology. Let $\mathcal{G}_{n,d}$ denote a random d-regular graph on n vertices sampled according to the configuration model (see [4] Chapter 2.4): each of the n distinct vertices emit d distinct half-edges, and we pair up these nd half-edges uniformly at random. These $nd/2$ pairs of half-edges can be glued into full edges to yield a labelled, random, d-regular graph. Note that the resulting graph can have loops and multiple edges. There are $(nd - 1)!! = (nd - 1)(nd - 3) \cdots 3 \cdot 1$ possible pairings, or outcomes, of the model. Let $G_{n,d}$ denote the set of all these outcomes. So $\mathcal{G}_{n,d}$ is picked uniformly at random from $G_{n,d}$.

3.1. *Projecting independent sets from* $\mathbb{T}_d$ *to* $\mathcal{G}_{n,d}$. Recall that given the factor of i.i.d. independent set I_d on $\mathbb{T}_d$ we can project it to a (random) independent set I_G on any given $G \in G_{n,d}$. If $B(G)$ is the number of vertices of G whose $(r+1)$-neighbourhood is not a tree then $\mathbb{E}[|I_G|/n] = \text{density}(I_d)(1 - \frac{B(G)}{n})$. We can model the independent sets $I_{d,i}$ from Section 2.1 in the random graph $\mathcal{G} = \mathcal{G}_{n,d}$. To do so, we first choose a random subset $S_{\mathcal{G}} \subset V(\mathcal{G}) = [n]$ via a Bernoulli percolation with density p. Then we fix independent random labellings X_i of $\mathcal{G}$ for $i \geq 0$. We define $I_{\mathcal{G},i}$ to be the projection of I_d with input $X_0(v)$ for $v \notin S_{\mathcal{G}}$ and $X_i(v)$ for $v \in S_{\mathcal{G}}$. As the $I_{\mathcal{G},i}$ are exchangeable, for any finite subset $T \subset \{1, 2, \ldots\}$ we have $\mathbb{E}[|\bigcap_{i \in T} I_{\mathcal{G},i}|/n] = \alpha_{|T|,d,p} \frac{\log d}{d}(1 - \frac{\mathbb{E}[B(\mathcal{G})]}{n})$. It is well known that $\mathbb{E}[B(\mathcal{G})]$ is bounded in n for every d [18], Chapter 9.2. So $\mathbb{E}[|\bigcap_{i \in T} I_{\mathcal{G},i}|/n] \to \alpha_{|T|,d,p} \frac{\log d}{d}$ as $n \to \infty$ for every $T \subset [k]$.

3.2. *The expected number of independent sets satisfying a given density profile.* From the construction above, we see that the factor of i.i.d. independent set I_d on $\mathbb{T}_d$ can be used to produce k-tuples of independent sets $(I_{\mathcal{G},1}, \ldots, I_{\mathcal{G},k})$ in $\mathcal{G} = \mathcal{G}_{n,d}$ such that the intersection densities of these k independent sets are close to those of $I_{d,1}, \ldots, I_{d,k}$, defined in Section 2.1. We will compute the expected number of k-tuples of independent sets in $\mathcal{G}$ with some given intersection densities. This will allow us to bound, from above, the probability of observing a k-tuple of independent sets in $\mathcal{G}$ whose intersection densities are close to that of $I_{d,1}, \ldots, I_{d,k}$. We will show that the expected number of k-tuples of independent sets in $\mathcal{G}$ with some prescribed intersection densities is an exponential term of the form e^{nR} (Theorem 3.1). The dominating contribution to the rate R is from the binomial sum of the left-hand side of inequality (2.2) for the prescribed intersection densities (Lemma 3.2 and Lemma 3.3). This will allow us to conclude that the only k-tuples of independent sets in $\mathcal{G}$ that exists (with nonvanishing probability as $n \to \infty$) are those for which the corresponding rate R is nonnegative. Then in the final step we will show via concentration inequalities that there exists k-tuples of independent sets in $\mathcal{G}$ whose intersection densities are close to those given by $I_{d,1}, \ldots, I_{d,k}$ (Lemma 3.4). This is the strategy behind the proof of Theorem 2.2. Before proceeding, we introduce some terminology.

For a k-tuple of independent sets $(I_{G,1}, \ldots, I_{G,k})$ in $G \in G_{n,d}$, the *density profile* associated to this k-tuple is the vector $\rho = (\rho(T); T \subset [k])$ defined by $\rho(T) = |\bigcap_{i \in T} I_{G,i}|/n$ [set $\rho(\varnothing) = 1$]. Associated to this k-tuple is also an ordered partition Π of $V(G)$ into 2^k cells defined as follows:

$$\Pi = \{\Pi(T) : T \subset [k]\} \qquad \text{with } \Pi(T) = \left(\bigcap_{i \in T} I_{G,i}\right) \cap \left(\bigcap_{i \notin T}(V(G) \setminus I_{G,i})\right).$$

In other words, $\Pi(T)$ consists of vertices that belong to all the sets $I_{G,i}$ for $i \in T$ and none of the other sets. The partition Π defines a probability measure $\pi =$

$(\pi(T); T \subset [k])$ on $2^{[k]}$ by $\pi(T) = |\Pi(T)|/n$. This correspondence between k-tuples $(I_{G,1}, \ldots, I_{G,k})$ and ordered partitions Π is bijective, and by the inclusion–exclusion principle we have that

$$(3.1) \qquad \pi(T) = \sum_{T':T\subset T'} (-1)^{|T'\setminus T|} \rho(T'),$$

$$(3.2) \qquad \rho(T) = \sum_{T':T\subset T'} \pi(T').$$

Finally, corresponding to G and Π is a $2^k \times 2^k$ matrix M that we denote the *edge profile* of Π. For $T, T' \subset [k]$, define

$$M(T, T') = \frac{|\{(u, v) \in E(G) : u \in \Pi(T), v \in \Pi(T')\}|}{nd}.$$

The tuple (u, v) refers to a directed edge; so $(u, v) \neq (v, u)$ unless $u = v$. The number of directed edges of G is $2|E(G)| = nd$. Notice that $M(T, T')$ is the probability that a uniformly chosen directed edge of G starts in $\Pi(T)$ and ends in $\Pi(T')$. Clearly, M is a symmetric matrix with nonnegative entries that sum to 1. Also, the marginal of M along either the rows or columns is π. A crucial observation is that if $T \cap T' \neq \varnothing$ then $M(T, T') = 0$. Indeed, in this case both $\Pi(T)$ and $\Pi(T')$ lie in the common independent set $I_{G,i}$ for any $i \in T \cap T'$, and thus, there cannot be any edges joining $\Pi(T)$ to $\Pi(T')$.

Conversely, suppose we begin with an ordered partition Π as above that induces an edge profile M on G. If the edge profile satisfies the constraints $M(T, T') = 0$ whenever $T \cap T' \neq \varnothing$ then the k-tuple of subsets $(I_{G,1}, \ldots, I_{G,k})$ of $V(G)$ corresponding to Π will be independents sets in G. Indeed, for any i, the number of edges of G that have both endpoints in $I_{G,i}$ is $(nd)/2$ times $\sum_{(T,T'):i\in T\cap T'} M(T, T')$, and all terms in the sum are zero. In this case, the density profile ρ of $(I_{G,1}, \ldots, I_{G,k})$ is given by (3.2) with π being the marginal of M along its rows.

With this terminology and bijection in mind, let $Z(\rho) = Z(\mathcal{G}, \rho)$ denote the number of k-tuples of independent sets in $\mathcal{G}$ with density profile ρ. Let $Z(\rho, M)$ denote the number of ordered partitions of $\mathcal{G}$ into 2^k cells such that the partitions induce the edge profile M, and M is compatible with ρ in the following sense. The marginal, π, of M along its rows is given by ρ via (3.1), and $M(T, T') = 0$ whenever $T \cap T' \neq \varnothing$. It is clear from the discussion above that $Z(\rho) = \sum_M Z(\rho, M)$, where the sum is over all M that is compatible with ρ.

THEOREM 3.1. *Given the setup as above, define the entropies*

$$H(M) = \sum_{(T,T')} -M(T, T') \log(M(T, T'))$$

and

$$H(\pi) = \sum_T -\pi(T) \log(\pi(T))$$

with the convention $0 \log 0 = 0$. *The expectation of* $Z(\rho, M)$ *satisfies*

$$(3.3) \qquad \mathbb{E}[Z(\rho, M)] \le O_k(\sqrt{n}d^{2^k}) \exp\left\{n\left[\frac{d}{2}H(M) - (d-1)H(\pi)\right]\right\}.$$

PROOF.   To compute the expectation, we sum the probabilities of outcomes where each outcome uniquely specifies a pairing of half-edges in the configuration model that gives rise to a partition Π with edge profile M. To specify such an outcome, do the following:

1. Partition the vertex set $[n]$ into 2^k distinguishable cells $\Pi(T)$, $T \subset [k]$ with $|\Pi(T)| = n\pi(T)$.
2. Given the partition Π from (1), and each subset $T \subset [k]$, partition the $nd\pi(T)$ half-edges attached to the vertices of $\Pi(T)$ into 2^k distinguishable cells $\Pi(T, T')$, $T' \subset [k]$, such that $|\Pi(T, T')| = ndM(T, T')$.
3. For each pair $\{T, T'\}$ with $T \neq T'$ pair up the half-edges from $\Pi(T, T')$ with those from $\Pi(T', T)$ in a specific way. Then for each T pair the half-edges from $\Pi(T, T)$ with themselves in a specific way.

Each outcome has probability $1/(nd-1)!!$ from definition of the configuration model. We compute the number of outcomes in the following. But first, we should mention some conventions that we use in the following calculations. For an even integer $m \ge 2$ we denote $(m-1)!! = (m-1)(m-3) \cdots 1$, and if $m = 0$ then $(m-1)!! = 1$. Also, note that in any valid edge profile M the quantities $ndM(T, T')$ have to be nonnegative integers. Furthermore, $ndM(T, T)$ has to be even for every T because for any $G \in G_{n,d}$ the number of half edges from $\Pi(T)$ to itself is twice the number of edges present in the subgraph of G induced by $\Pi(T)$. We may assume that M has all these properties. We now compute the number of outcomes.

- The number of partitions of $[n]$ that satisfies the properties in (1) above is the multinomial coefficient

$$\binom{n}{n\pi(T); T \subset [k]}.$$

- Given a partition Π satisfying (1) from above, the number of partitions of the half-edges that satisfy the properties in (2) is

$$\prod_{T \subset [k]} \binom{nd\pi(T)}{ndM(T, T'); T' \subset [k]}.$$

- Given the two partitions arising from (1) and (2), the number of pairings that satisfy (3) is

$$\left[\prod_{(T,T'):T \neq T'} (ndM(T, T'))!\right]^{1/2} \prod_{T \subset [k]} (ndM(T, T) - 1)!!.$$

The total number of outcomes is the product of the three terms above. From the linearity of expectation, we conclude that $\mathbb{E}[Z(\rho, M)]$ equals

$$\binom{n}{n\pi(T);\, T \subset [k]} \times \prod_{T \subset [k]} \binom{nd\pi(T)}{ndM(T,T');\, T' \subset [k]}$$

$$\times \left[\prod_{(T,T'):T \neq T'} (ndM(T,T'))! \right]^{1/2}$$

$$\times \prod_{T \subset [k]} (ndM(T,T) - 1)!! \times \frac{1}{(nd-1)!!}.$$

Now we do the asymptotics in n by using Stirling's approximation of $m!$. We have that $\sqrt{2\pi m}(m/e)^m \leq m! \leq (1 + \frac{1}{12m})\sqrt{2\pi m}(m/e)^m$. Also, for an even integer m, $(m-1)!! = \frac{m!}{2^{m/2}(m/2)!}$. In the following, we need to consider only those values of $\pi(T)$ and $M(T,T')$ that are strictly positive. We begin by simplifying the term

$$\prod_{T \subset [k]} \binom{nd\pi(T)}{ndM(T,T');\, T' \subset [k]} \times \left[\prod_{(T,T'):T \neq T'} (ndM(T,T'))! \right]^{1/2}$$

$$\times \prod_{T \subset [k]} (ndM(T,T) - 1)!!$$

$$= \prod_T (nd\pi(T))! \times \left[\prod_{(T,T'):T \neq T'} (ndM(T,T'))! \right]^{-1/2}$$

$$\times 2^{-nd/2 \sum_T M(T,T)} \left[\prod_T \left(\frac{nd}{2} M(T,T) \right)! \right]^{-1}.$$

Incorporating the remaining two terms, we see that the expectation is

$$\binom{n}{n\pi(T);\, T \subset [k]} \times \binom{nd}{nd\pi(T);\, T \subset [k]}^{-1} \times (nd/2)!$$

$$\times 2^{\left(\frac{nd}{2}(1 - \sum_T M(T,T)) \right)}$$

$$\times \left[\prod_{(T,T'):T \neq T'} (ndM(T,T'))! \right]^{-1/2} \times \left[\prod_T \left(\frac{nd}{2} M(T,T) \right)! \right]^{-1}.$$

Using Stirling's approximation, we can verify that (with universal constants)

$$\binom{n}{n\pi(T);\, T \subset [k]} \binom{nd}{nd\pi(T);\, T \subset [k]}^{-1} = O(d^{\frac{(2^k-1)}{2}}) \exp\{-n(d-1)H(\pi)\}.$$

Similarly,

$$\prod_T \left(\frac{nd}{2}M(T,T)\right)! = O\left((nd\pi)^{2^{k-1}}\prod_T M(T,T)^{\frac{1}{2}}\right)\prod_T \left(\frac{ndM(T,T)}{2e}\right)^{\frac{nd}{2}M(T,T)};$$

$$\prod_{\substack{(T,T'):\\T\neq T'}}(ndM(T,T'))! = O\left((2\pi nd)^{\binom{2^k}{2}}\prod_{T\neq T'}M(T,T')^{\frac{1}{2}}\right)$$

$$\times \left(\frac{nd}{e}\right)^{[nd\sum_{(T,T'):T\neq T'}M(T,T')]}$$

$$\times \prod_{\substack{(T,T'):\\T\neq T'}}M(T,T')^{ndM(T,T')}.$$

Now, $(nd/2)! = O(nd)(\frac{nd}{2e})^{nd/2}$. From this and the previous two equations, we can check that all terms involving powers of $(nd)/e$ and powers of 2 algebraically cancel out from the expression for $\mathbb{E}[Z(\rho,M)]$. Therefore, after algebraic simplifications we conclude that

$$\mathbb{E}[Z(\rho,M)] = O_k\left(\left(\frac{n}{d}\right)^{\frac{1-2^k}{2}}(nd)^{-\frac{1}{2}\binom{2^k}{2}}\prod_T M(T,T)^{-1/2}\prod_{\substack{(T,T'):\\T\neq T'}}M(T,T')^{-1/4}\right)$$

$$\times \exp\{-n(d-1)H(\pi)\}\times\prod_T M(T,T)^{-\frac{nd}{2}M(T,T)}$$

$$\times \prod_{\substack{(T,T'):\\T\neq T'}}M(T,T')^{-\frac{nd}{2}M(T,T')}$$

$$= O_k\left(n^{\frac{1}{2}}d^{2^k}\right)\exp\left\{n\left[\frac{d}{2}H(M)-(d-1)H(\pi)\right]\right\}.$$

The O_k term results because $ndM(T,T') \geq 1$ for every $M(T,T') > 0$, from which it is easy to deduce that n and d appear with exponents at most $1/2$ and 2^k, respectively. $\quad\square$

The number of summands in $\sum_M Z(\rho,M)$ over edge profiles M compatible with ρ is bounded by a polynomial in n. Indeed, M has 4^k non-negative entries of the form $m(T,T')/nd$ with the integers $m(T,T')$ satisfying $\sum_{(T,T')}m(T,T') = nd$. There are at most $(nd)^{4^k}$ such solutions. This allows us to conclude that $\mathbb{E}[Z(\rho)]$ is dominated by the largest exponential term, or in other words, the term with the largest value of $(d/2)H(M)-(d-1)H(\pi)$ optimized over M that are compatible with ρ. We bound this optimum in the following.

Let $M = [M(T, T')]_{\{T, T' \subset [k]\}}$ be an edge profile matrix with the property that M is symmetric, the support of M is contained in the set $\{(T, T') : T \cap T' = \varnothing\}$ and that the marginal of M along its row is a fixed probability distribution $\pi = (\pi(T); T \subset [k])$. Define the weights

$$w(T) = \sum_{T':T' \cap T = \varnothing} \pi(T').$$

Note that $w(\varnothing) = 1$, and

$$(3.4) \qquad \sum_{T:T \cap T' = \varnothing} \frac{\pi(T)}{w(T')} = 1.$$

LEMMA 3.2. *With a matrix M and vectors π, w as above we have*

$$H(M) \le 2H(\pi) + \sum_{S \subset [k]} \pi(T) \log(w(T)).$$

PROOF. Set $h(x) = -x \log(x)$ for $0 \le x \le 1$ ($0 \log 0 = 0$). Note that $h(x)$ is a smooth and strictly concave function on its domain. We have that

$$H(M) = \sum_{T'} \sum_{T:T \cap T' = \varnothing} \pi(T) \frac{h(M(T, T'))}{\pi(T)}$$

$$= \sum_{T'} \sum_{T:T \cap T' = \varnothing} \pi(T) h\left(\frac{M(T, T')}{\pi(T)}\right) + H(\pi).$$

For the second equality, we used that $h(xy) = xh(y) + yh(x)$.

By Jensen's inequality applied to $h(x)$ and the identity (3.4), we deduce that

$$\sum_{T:T \cap T' = \varnothing} \frac{\pi(T)}{w(T')} h\left(\frac{M(T, T')}{\pi(T)}\right) \le h\left(\sum_{T:T \cap T' = \varnothing} \frac{M(T, T')}{w(T')}\right) = h\left(\frac{\pi(T')}{w(T')}\right).$$

From this, we conclude that

$$H(M) \le \sum_{T'} w(T') h\left(\frac{\pi(T')}{w(T')}\right) + H(\pi) = 2H(\pi) + \sum_{T'} \pi(T') \log(w(T')). \qquad \square$$

Using Lemma 3.2 and Theorem 3.1, we conclude that for any density profile ρ

$$(3.5) \qquad \mathbb{E}[Z(\rho)] \le \text{poly}(n, d) \times \exp\left\{n\left[H(\pi) - \frac{d}{2}\hat{H}(\pi)\right]\right\},$$

where $\hat{H}(\pi) = \sum_T \pi(T) \log(w(T))$, and $\text{poly}(n, d)$ is a polynomial in n and d of degree at most 4^k.

For the purposes of our analysis, we will be interested in density profiles ρ such that $\rho(T) \in [\rho_{|T|} - \varepsilon, \rho_{|T|}]$ with $\rho_i = \alpha_{i,d,p} \frac{\log d}{d}$. To this end, let us fix $1 = \rho_0 \ge$

$\rho_1 \geq \cdots \geq \rho_k$ with $\rho_i = \alpha_{i,d,p}\frac{\log d}{d}$. Define the density profile ρ by $\rho(T) = \rho_{|T|}$ for $T \subset [k]$. Let π denote the probability distribution associated to ρ as given by (3.1). For $T \neq \varnothing$, define the quantities $\beta(T)$ by $\pi(T) = \beta(T)\frac{\log d}{d}$. Note that $\pi(\varnothing) = 1 - [\sum_{T \neq \varnothing} \beta(T)]\frac{\log d}{d}$. By setting $\alpha(T) = \alpha_{|T|,d,p}$ and using the relation between ρ and π from (3.1) and (3.2), we conclude the following relation between α and β:

$$(3.6) \qquad \alpha(T) = \sum_{T':T \subset T'} \beta(T),$$

$$(3.7) \qquad \beta(T) = \sum_{T':T \subset T'} (-1)^{|T' \setminus T|} \alpha(T').$$

Note that $\alpha_{1,d,p} = \alpha \leq 2$. Indeed, recall the result of Bollobás [3] mentioned in the Introduction: if $A_{n,d}$ is the event that all independent sets in $\mathcal{G}_{n,d}$ have size at most $2\frac{\log d}{d}n$ then $\mathbb{P}[A_{n,d}] \to 1$ as $n \to \infty$ for every $d \geq 3$. Recall from Section 3.1 that $\alpha\frac{\log d}{d} = \lim_{n\to\infty} \mathbb{E}[|I_{\mathcal{G}_{n,d},1}|/n]$. However, $\mathbb{E}[|I_{\mathcal{G}_{n,d},1}|] = \mathbb{E}[|I_{\mathcal{G}_{n,d},1}|; A_{n,d}] + \mathbb{E}[|I_{\mathcal{G}_{n,d},1}|; A_{n,c}^c] \leq 2\frac{\log d}{d}n + n\mathbb{P}[A_{n,d}^c]$. Dividing through by n and then taking limits in n, we deduce that $\alpha \leq 2$.

From the fact that $\alpha \leq 2$, we see that $0 \leq \alpha_{k,d,p} \leq \cdots \leq \alpha_{1,d,p} \leq 2$. From (3.7), it follows that $\beta(T) \leq 2^{k+1}$ for all $T \subset [k]$. In particular, this estimate is uniform in d and p.

LEMMA 3.3.   *With π, α and β as above, we have that*

$$H(\pi) - \frac{d}{2}\hat{H}(\pi) \leq \left[\sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \alpha_{i,d,p}(2 - \alpha_{i,d,p}) \right] \frac{\log^2 d}{2d} + O_k\left(\frac{\log d}{d}\right),$$

where the big O term depends only on k.

PROOF.   We need the asymptotic behaviour of $H(\pi) - \frac{d}{2}\hat{H}(\pi)$ where the entries of π are on the scale of $(\log d)/d$. By definition,

$$w(T) = 1 - \frac{\log d}{d} \sum_{T':T' \cap T \neq \varnothing} \beta(T').$$

From Taylor expansion, we observe that $-\log(1 - x) \geq x$. Hence, for $T \neq \varnothing$ we have

$$-\pi(T)\log(w(T)) \geq \beta(T)\left(\frac{\log d}{d}\right)^2 \sum_{T':T' \cap T \neq \varnothing} \beta(T').$$

Since $w(\varnothing) = 1$, we have

$$\hat{H}(\pi) = \sum_{T \neq \varnothing} -\pi(T)\log(w(T)) \geq \left(\frac{\log d}{d}\right)^2 \sum_{(T,T'):T \cap T' \neq \varnothing} \beta(T)\beta(T').$$

To analyze $H(\pi)$, we consider the terms $h(\pi(\varnothing))$ and $h(\pi(T))$ with $T \neq \varnothing$ separately. We note from Taylor expansion that $h(1 - x) \leq x$ for $0 \leq x \leq 1$. Thus,

$$h(\pi(\varnothing)) = h\left(1 - \frac{\log d}{d} \sum_{T \neq \varnothing} \beta(T)\right) \leq \frac{\log d}{d} \sum_{T \neq \varnothing} \beta(T).$$

Since $\beta(T) \leq 2^{k+1}$ for $T \neq \varnothing$, we see that $h(\pi(\varnothing)) = O_k(\frac{\log d}{d})$. On the other hand, for $T \neq \varnothing$ the quantity $h(\pi(T))$ equals

$$h\left(\beta(T)\frac{\log d}{d}\right) = \beta(T)h\left(\frac{\log d}{d}\right) + h(\beta(T))\frac{\log d}{d}$$

$$\leq \beta(T)\frac{\log^2 d}{d} + \frac{\log d}{ed}.$$

The inequality follows because $h(\frac{\log d}{d}) \leq \frac{\log^2 d}{d}$ and $h(x) \leq e^{-1}$ for all $x \geq 0$.

Therefore, $H(\pi) \leq \frac{\log^2 d}{d} \sum_{T \neq \varnothing} \beta(T) + O_k(\frac{\log d}{d})$.

From the above, we conclude that

$$H(\pi) - \frac{d}{2}\hat{H}(\pi) \leq \frac{\log^2 d}{d}\left[\sum_{T \neq \varnothing} \beta(T) - \frac{1}{2} \sum_{\substack{(T, T'): \\ T \cap T' \neq \varnothing}} \beta(T)\beta(T')\right] + O_k\left(\frac{\log d}{d}\right).$$

Finally, it follows by inclusion–exclusion that

$$\sum_{T \neq \varnothing} \beta(T) - \frac{1}{2} \sum_{\substack{(T, T'): \\ T \cap T' \neq \varnothing}} \beta(T)\beta(T') = \frac{1}{2}\sum_{i=1}^{k}(-1)^{i-1}\binom{k}{i}\alpha_{i,d,p}(2 - \alpha_{i,d,p}).$$

The details are as follows. From the relations between α and β in (3.7) and (3.6), it follows immediately that

$$\sum_{i=1}^{k}(-1)^{i-1}\binom{k}{i}\alpha_{i,d,p} = \sum_{T \neq \varnothing} \beta(T)$$

because both terms equal $(|\bigcup_{i=1}^{k} I_{G,i}|/n) \cdot \frac{d}{\log d}$.

These relations also imply $\alpha(T)^2 = \sum_{(T_1, T_2): T \subset T_1 \cap T_2} \beta(T_1)\beta(T_2)$. Hence,

$$\sum_{i=1}^{k}(-1)^{i-1}\binom{k}{i}\alpha_{i,d,p}^2 = \sum_{T \neq \varnothing}(-1)^{|T|-1}\alpha(T)^2$$

$$= \sum_{T \neq \varnothing}(-1)^{|T|-1} \sum_{\substack{(T_1, T_2): \\ T \subset T_1 \cap T_2}} \beta(T_1)\beta(T_2)$$

$$= \sum_{\substack{(T_1,T_2): \\ T_1 \cap T_2 \neq \varnothing}} \beta(T_1)\beta(T_2) \sum_{\substack{T: \\ T \subset T_1 \cap T_2, T \neq \varnothing}} (-1)^{|T|-1}$$

$$= \sum_{\substack{(T_1,T_2): \\ T_1 \cap T_2 \neq \varnothing}} \beta(T_1)\beta(T_2) \sum_{i=1}^{|T_1 \cap T_2|} (-1)^{i-1} \binom{|T_1 \cap T_2|}{i}.$$

Now recall the binomial identity $\sum_{i=1}^{t}(-1)^{i-1}\binom{t}{i} = 1 - (1-1)^t = 1$ for any integer $t \geq 1$. This identity implies that

$$\sum_{i=1}^{k}(-1)^{i-1}\binom{k}{i}\alpha_{i,d,p}^2 = \sum_{(T,T'):T \cap T' \neq \varnothing} \beta(T)\beta(T').$$

With this the proof of the final claim is complete. $\quad\square$

Let $E_{d,p}(\varepsilon) = E(\alpha, \varepsilon, n, d, p)$ be the event that $\mathcal{G}_{n,d}$ contains some k-tuple of independent sets $(I_1, \ldots, I_k)$ whose density profile ρ satisfies the property that for every $T \subset [k]$,

$$(3.8) \qquad \rho(T) \in \left[\alpha_{|T|,d,p}\frac{\log d}{d} - \varepsilon, \alpha_{|T|,d,p}\frac{\log d}{d} + \varepsilon\right].$$

We can bound $\mathbb{P}[E_{d,p}(\varepsilon)]$ from above via (3.5) and Lemma 3.3. Define the density profile ρ_α by $\rho_\alpha(T) = \alpha_{|T|,d,p}\frac{\log d}{d}$. Let π_{ρ_α} be the corresponding probability vector obtained from (3.1). For any admissible density profile ρ for the occurrence of the event $E_{d,p}(\varepsilon)$, the corresponding π_ρ satisfies $|\pi_\rho(T) - \pi_{\rho_\alpha}| = O_k(\varepsilon)$. We employ Lemma 3.3 for π_ρ. We get that

$$H(\pi_\rho) - \frac{d}{2}\hat{H}(\pi_\rho) = H(\pi_{\rho_\alpha}) - \frac{d}{2}\hat{H}(\pi_{\rho_\alpha}) + \text{err}_{d,k}(\varepsilon).$$

The error term $\text{err}_{d,k}$ is such that $\text{err}_{d,k}(\varepsilon) \to 0$ as $\varepsilon \to 0$, and this holds uniformly in $p \in [0,1]$. This follows from the fact that π is obtained from ρ by a smooth transformation [see (3.1)], and that H and $\hat{H}$ are smooth functions. The reason $\text{err}_{d,k}(\varepsilon)$ tends to 0 uniformly in p is because it depends on the $\alpha_{i,d,p}$ smoothly and only through their absolute values. However, the $\alpha_{i,d,p}$ are all bounded as $0 \leq \alpha_{k,d,p} \leq \cdots \leq \alpha_{1,d,p} = \alpha \leq 2$. A careful analysis will actually show that $\text{err}_{d,k}(\varepsilon) = O_k(\frac{\log^2 d}{d}\varepsilon)$.

From Lemma 3.3 applied to ρ_α it follows that for any admissible ρ for the occurrence of the event $E_{d,p}(\varepsilon)$,

$$(3.9) \qquad \begin{aligned} H(\pi_\rho) - \frac{d}{2}\hat{H}(\pi_\rho) &\leq \frac{\log^2 d}{d}\sum_{i=1}^{k}(-1)^{i-1}\binom{k}{i}\alpha_{i,d,p}(2 - \alpha_{i,d,p}) \\ &\quad + O_k\left(\frac{\log d}{d}\right) + \text{err}_{d,k}(\varepsilon). \end{aligned}$$

Now note that the number of density profiles ρ that is admissible for the event $E_{d,p}(\varepsilon)$ is at most $O(n^{2^k})$ where the big O constant is uniformly bounded in n because quantities of the form $\alpha_{i,d,p} \frac{\log d}{d}$ are all of constant order in n. Taking an union bound over all such admissible ρ, using the first moment method and employing the bounds in (3.5) and (3.9), we conclude that $\mathbb{P}[E_{d,p}(\varepsilon)]$ is bounded above by a polynomial term $\text{poly}(n,d)$ times the exponential term

(3.10)
$$\exp\left\{ n\left[\frac{\log^2 d}{d} \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \alpha_{i,d,p}(2-\alpha_{i,d,p}) \right.\right.$$
$$\left.\left. + O_k\left(\frac{\log d}{d}\right) + \text{err}_{d,k}(\varepsilon) \right]\right\}.$$

3.3. *Concentration of the density profile of $(I_{\mathcal{G},1}, \ldots, I_{\mathcal{G},k})$ about its mean.* Having derived an upper bound to $\mathbb{P}[E_{d,p}(\varepsilon)]$, we need a lower bound on this probability that violates the upper bound and provides a contradiction. We now show that $\mathbb{P}[E_{d,p}(\varepsilon)] \to 1$ as $n \to \infty$ via concentration inequalities.

Recall the terminology of Section 3. The factor of i.i.d. independent set I_d is used to construct independent sets $I_{G,1}, \ldots, I_{G,k}$ of $G \in G_{n,d}$ using random labellings $X_0, \ldots, X_k$ of G and a random subset $S \subset V(G)$ resulting from Bernoulli percolation on G. Set $Y(v) = (X_0(v), \ldots, X_k(v), \mathbf{1}_{v\in S})$ for $v \in V(G)$. Then $Y = (Y(v); v \in V(G))$ is an i.i.d. process on G and $I_{G,1}, \ldots, I_{G,k}$ is a function of Y. Also, for any $T \subset [k]$ the expected density $\mathbb{E}[|\bigcap_{i\in T} I_{G,i}|/n] \in [\alpha_{i,d,p} \frac{\log d}{d}(1 - \frac{B(G)}{n}), \alpha_{i,d,p} \frac{\log d}{d}]$, where $B(G)$ is the number of vertices in G whose $(r+1)$-neighbourhood is not a tree.

LEMMA 3.4. *For any $G \in G_{n,d}$, the independent sets $I_{G,1}, \ldots, I_{G,k}$ satisfy the following with $C_{r,d} = O(r^2 d^{2r})$:*

(3.11)
$$\mathbb{P}\left[\max_{T\subset[k]} \left| \frac{|\bigcap_{i\in T} I_{G,i}| - \mathbb{E}[|\bigcap_{i\in T} I_{G,i}|]}{n} \right| > \varepsilon \right] \leq 2^{k+1} e^{-\frac{\varepsilon^2 n}{C_{r,d}}}.$$

PROOF. For each $T \subset [k]$, the set $\bigcap_{i\in T} I_{G,i}$ is a function of $y = (y(v); v \in V(G))$, where each $y(v) \in [0,1]^{k+1} \times \{0,1\}$ [the set of values of the random variable $Y(v)$]. Modifying some entry $y(v)$ to $y'(v)$ can switch the state of inclusion of a vertex u within $\bigcap_{i\in T} I_{G,i}$ only if u is in $N_G(r,v)$, where r is the radius of the factor associated to I_d. Therefore, such a modification to y can cause the size of $\bigcap_{i\in T} I_{G,i}$ to change by at most $|N_G(r,v)| = O(rd^r)$ since G is d-regular. Since the random input Y is an i.i.d. process it follows from the Hoeffding–Azuma inequality [4], Theorem 1.20, that

$$\mathbb{P}\left[\left| \left|\bigcap_{i\in T} I_{G,i}\right| - \mathbb{E}\left[\left|\bigcap_{i\in T} I_{G,i}\right| \right] \right| > x \right] \leq 2e^{\frac{x^2}{2nC_{r,d}}}.$$

The lemma follows by taking an union bound over $T \subset [k]$ and replacing x by $n\varepsilon$.

$\square$

Recall that for the random graph $\mathcal{G}_{n,d}$ we have

$$\alpha_{|T|,d,p} \frac{\log d}{d} \left(1 - \frac{\mathbb{E}[B(\mathcal{G}_{n,d})]}{n} \right) \leq \mathbb{E}\left[\frac{|\bigcap_{i \in T} I_{\mathcal{G}_{n,d},i}|}{n} \right] \leq \alpha_{|T|,d,p} \frac{\log d}{d}.$$

As we mentioned in Section 3.1, $\mathbb{E}[B(\mathcal{G}_{n,d})] = O_{d,r}(1)$ in n.

We may find an n_d such that $\alpha_{k,d,p} \frac{\log d}{d} \mathbb{E}[B(\mathcal{G}_{n,d})] \leq (\varepsilon/2)n$ for $n \geq n_d$. This ensures that

$$\alpha_{|T|,d,p} \frac{\log d}{d} \left(1 - \frac{\mathbb{E}[B(\mathcal{G}_{n,d})]}{n} \right) \geq \alpha_{|T|,d,p} \frac{\log d}{d} - \frac{\varepsilon}{2} \qquad \text{for all } T \subset [k].$$

If $||\bigcap_{i \in T} I_{\mathcal{G}_{n,d},i}| - \mathbb{E}[|\bigcap_{i \in T} I_{\mathcal{G}_{n,d},i}|]| \leq (\varepsilon/2)n$, then the independent sets $I_{\mathcal{G}_{n,d},i}$ satisfy, for any $T \subset [k]$,

$$\alpha_{|T|,d,p} \frac{\log d}{d} - \varepsilon \leq \frac{|\bigcap_{i \in T} I_{\mathcal{G}_{n,d},i}|}{n} \leq \alpha_{|T|,d,p} \frac{\log d}{d} + \varepsilon.$$

Therefore, the event $E_{d,p}$ occurs for $n \geq n_d$ [see the definition of $E_{d,p}$ in (3.8)]. From Lemma 3.4, we conclude that for $n \geq n_d$,

$$(3.12) \qquad \mathbb{P}[E_{d,p}(\varepsilon)] \geq 1 - 2^{k+1} e^{-\frac{\varepsilon^2}{C_{r,d}} n} \longrightarrow 1 \qquad \text{as } n \to \infty.$$

3.3.1. *Conclusion of the proof of Theorem* 2.2. Suppose for some $\delta > 0$ we have that

$$\liminf_{d \to \infty} \inf_{p \in [0,1]} \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \alpha_{i,d,p} (2 - \alpha_{i,d,p}) = -\delta.$$

For each d, we pick a $p' = p'(d)$ such that

$$\inf_{p \in [0,1]} \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \alpha_{i,d,p} (2 - \alpha_{i,d,p})$$

$$\geq \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \alpha_{i,d,p'} (2 - \alpha_{i,d,p'}) - \frac{\delta}{2}.$$

We deduce that,

$$(3.13) \qquad \liminf_{d \to \infty} \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \alpha_{i,d,p'} (2 - \alpha_{i,d,p'}) \leq -\frac{\delta}{2}.$$

Recall the upper bound for $\mathbb{P}[E_{d,p}(\varepsilon)]$ in (3.10), which we now consider for $p = p'(d)$. The error term $\mathrm{err}_{d,k}(\varepsilon) \to 0$ as $\varepsilon \to 0$, uniformly in p. So for each

d pick an ε_d such that $\mathrm{err}_{d,k}(\varepsilon_d) \leq (\delta/4)\frac{\log^2 d}{d}$. The inequalities (3.10) and (3.13) imply that there exists a subsequence $d_i \to \infty$ such that

$$\mathbb{P}\big[E_{d_i, p'(d_i)}(\varepsilon_d)\big] \leq \mathrm{poly}(n, d_i) e^{n\left[-\frac{\delta \log^2 d_i}{4d_i} + O_k\left(\frac{\log d_i}{d_i}\right)\right]} \longrightarrow 0 \qquad \text{as } n \to \infty$$

for all sufficiently large values of d_i. However, we have already concluded from (3.12) that $\mathbb{P}[E_{d_i, p'(d_i)}(\varepsilon_d)] \to 1$ as $n \to \infty$ for all such d_i. This provides a contradiction and completes the proof.

4. Local algorithms for independent sets in Erdős–Rényi graphs. Local algorithms on sparse Erdős–Rényi graphs are projections of factor of i.i.d. processes on Poisson–Galton–Watson (PGW) trees. We will define the appropriate notion of factor of i.i.d. independent sets in PGW trees and prove the same asymptotic upper and lower bounds as for regular trees. Recall that a PGW tree with average degree λ, which we denote PGW_λ, is a random tree resulting from a Galton–Watson branching process with a Poisson(λ) offspring distribution. Before we can define the notion of factor of i.i.d. independent sets in PGW trees, we will need some notation.

Let Λ_r denote the collection of all triples (H, v, x) where (1) (H, v) is a finite, connected, rooted graph with root v, (2) for all vertices $u \in V(H)$ we have $\mathrm{dist}(u, v) \leq r$ where dist denotes the graph distance, and (3) $x \in [0, 1]^{V(H)}$ is a labelling of H. Λ_r has a natural σ-algebra, Σ_r, generated by sets of the form $(H, v) \times B$, where (H, v) satisfies properties (1) and (2) above and B is a Borel measurable subset of $[0, 1]^{V(H)}$. We consider two rooted graphs to be isomorphic if there exists a graph isomorphism between them that maps one root to the other. Given an isomorphism $\phi : (H, v) \to (H', v')$, any labelling x of (H, v) induces a labelling $\phi \cdot x$ of (H', v') by defining $\phi \cdot x(i) = x(\phi^{-1}(i))$, and vice versa. A function $f : \Lambda_r \to \{0, 1\}$ is a factor if it is Σ_r measurable and $f(H, v, x) = f(\phi(H), \phi(v), \phi \cdot x)$ for all isomorphisms ϕ of H, and all H.

For $0 \leq r < \infty$, let $f : \Lambda_r \to \{0, 1\}$ be a factor. Consider a PGW_λ tree with a random labelling X. Let $N_r(\mathrm{PGW}_\lambda, v)$ denote the r-neighbourhood of a vertex v in PGW_λ and let $X(\mathrm{PGW}_\lambda, v, r)$ be the restriction of X to $N_r(\mathrm{PGW}_\lambda, v)$. Define a subset I of the vertices of PGW_λ by setting

$$I(v) = f\big(N_r(\mathrm{PGW}_\lambda, v), v, X(\mathrm{PGW}_\lambda, v, r)\big).$$

We say that I is a factor of i.i.d. independent set in PGW_λ if I is an independent set in this tree with probability 1 [w.r.t. the random labelled tree $(\mathrm{PGW}_\lambda, X)$].

The distribution of the random variable $I(v)$ does not depend on the choice of the vertex v. This is because in a PGW tree the distribution of the neighbourhoods $N_r(\mathrm{PGW}_\lambda, v)$ does not depend on the choice of v. So let $\mathrm{PGW}(\lambda, r)$ denote the tree following the common distribution of these r-neighbourhoods, rooted at a

vertex $\circ$, and let X be a random labelling. The density of the subset I is defined to be the expectation

$$\text{density}(I) = \mathbb{E}\big[f\big(\text{PGW}(\lambda, r), \circ, X\big)\big].$$

Define $\alpha(\lambda)$ so that

$$\alpha(\lambda)\frac{\log \lambda}{\lambda} = \sup_{0 \le r < \infty} \big\{\text{density}(I) : I \text{ is an independent set in PGW}_\lambda$$

$$\text{with factor } f : \Lambda_r \to \{0, 1\}\big\}.$$

THEOREM 4.1.   *The limit* $\lim_{\lambda \to \infty} \alpha(\lambda) = 1$.

In Section 4.1, we prove that $\limsup_{\lambda \to \infty} \alpha(\lambda) \le 1$, and in Section 4.2 that $\liminf_{\lambda \to \infty} \alpha(\lambda) \ge 1$. The proof of the upper bound will employ the strategy used for regular trees in Section 2. We will highlight the key differences but be brief with parts of the argument that are analogous to the case for regular trees.

4.1. *Upper bound on density of factor of i.i.d. independent sets in PGW trees.* Recall the Erdős–Rényi graph $\text{ER}(n, p)$ is a random graph on the vertex set $[n]$ where every pair of vertices $\{u, v\}$ is independently included with probability p. Our interest lies with the random graphs $\text{ER}(n, \lambda/n)$ where $\lambda > 0$ is fixed. Throughout this section, let G_n denote a random graph that is distributed according to the $\text{ER}(n, \lambda/n)$ model. It is well known (see [4] Chapter 4) that the sequence of random graphs G_n converges in the *local weak limit* to the tree PGW_λ. This means that for every fixed $r \ge 0$, if $\circ_n \in [n]$ is chosen uniformly at random then for any finite rooted graph $(H, \circ)$ the probability $\mathbb{P}[N_r(G_n, \circ_n) \cong (H, \circ)] \to \mathbb{P}[\text{PGW}(\lambda, r) \cong (H, \circ)]$ as $n \to \infty$.

Consequently, using the same technique that was used for regular trees, a factor of i.i.d. independent set I of PGW_λ with factor $f : \Lambda_r \to \{0, 1\}$ yields a factor of i.i.d. independent set I_n of G_n in the following sense. These exists a factor $f_n : \Lambda_{r+1} \to \{0, 1\}$ such that if X is a random labelling of G_n then $I_n(v) = f_n(N_{r+1}(G_n, v), v, X(G_n, v, r + 1))$. Furthermore, $\mathbb{E}[|I_n|/n] \to \text{density}(I)$ as $n \to \infty$.

To prove that $\limsup \alpha(\lambda) \le 1$, we assume to the contrary. Then we can find $\alpha > 1$ and a subsequence of $\lambda \to \infty$ such that for each λ there exists a factor of i.i.d. independent set $I_{n,\lambda}$ of G_n with factor $f_{n,\lambda} : \Lambda_{r_\lambda} \to \{0, 1\}$, and $E|I_{n,\lambda}|/n \ge \alpha\frac{\log \lambda}{\lambda}$ for all sufficiently large n. We can assume w.l.o.g. that these statements hold for all λ and n. By setting $E|I_{n,\lambda}|/n = \alpha_{1,n,\lambda}\frac{\log \lambda}{\lambda}$ we have that $\alpha_{1,n,\lambda} \ge \alpha > 1$.

4.1.1. *A coupling of local algorithms on Erdős-Rényi graphs.* For $0 \le p \le 1$, let $S = S_{n,p}$ be a random subset of $V(G_n)$ chosen by doing a Bernoulli percolation with density p. Let $G'_n = G'_n(G_n, S)$ be the random graph that is obtained from G_n

by independently resampling the edge connections between each pair of vertices $\{u, v\} \subset S$ with inclusion probability λ/n. In other words, G'_n retains all edges of G_n that do not connect S to itself, and all possible edge connections between vertices within S are resampled according to the Erdős–Rényi model. Note that G'_n is also distributed according to $\mathrm{ER}(n, \lambda/n)$; if $p = 0$ then $G'_n = G_n$, and if $p = 1$ then G'_n is independent of G_n.

Now fix G_n and S as above and let X be a random labelling of G_n. Let $X^1, X^2, \ldots$ be new, independent random labellings and define labellings Y^k, correlated with X, by $Y^k(v) = Y^k(v)$ if $v \in S$, and $Y^k(v) = X(v)$ otherwise. Generate $G^1, G^2, \ldots$ from G_n and S by using the recipe for G'_n, but rewire the edges for each G^i independently. In other words, the induced subgraphs $G^1[S], G^2[S], \ldots$ are independent. Now consider independent sets $I^1, I^2, \ldots$ by letting I^k be generated by the factor $f_{n,\lambda}$ with input graph G^k and labelling Y^k. Thus, I^k is a factor of i.i.d. independent set of G^k. Since all these graphs have a common vertex set, namely $[n]$, we can consider intersections of the I^k. Note that for any finite subset T the expected intersection density $\mathbb{E}[|\bigcap_{t \in T} I^t|/n]$ depends only of $|T|$ due to exchangeability of the I^k. Define the parameters $\alpha_{k,n,\lambda}$ by

$$(4.1) \qquad \alpha_{k,n,\lambda} \frac{\log \lambda}{\lambda} = \mathbb{E}\left[\frac{|\bigcap_{t=1}^{k} I^t|}{n}\right].$$

THEOREM 4.2. *The following inequality holds for each $k \geq 1$:*

$$(4.2) \qquad \liminf_{\lambda \to \infty} \liminf_{n \to \infty} \inf_{p \in [0,1]} \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \alpha_{i,n,\lambda,p}(2 - \alpha_{i,n,\lambda,p}) \geq 0.$$

Now we establish the upper bound by using Theorem 4.2. We define stability variables $Q_{n,\lambda,p}$ in a manner analogous to what we did for regular graphs. First, let $\circ \in [n]$ be a uniform random vertex. We restrict our probability space to the support of $f_{n,\lambda}(N_{r_\lambda}(G_n, \circ), \circ, X(G_n, \circ, r_\lambda))$, considered as a subset of the original probability space determined by $\circ$, the random labellings $X, X^1, \ldots$, the random subset S and the independent trials that determine the graphs $G_n, G^1, G^2, \ldots$. Let $\mathbb{E}^*$ be the expectation operator $\mathbb{E}$ restricted to the new space:

$$\mathbb{E}^*[U] = \frac{\mathbb{E}[U f_{n,\lambda}(N_{r_\lambda}(G_n, \circ), \circ, X(G_n, \circ, r_\lambda))]}{\mathbb{E}[f_{n,\lambda}(N_{r_\lambda}(G_n, \circ), \circ, X(G_n, \circ, r_\lambda))]}.$$

Notice that we define the new probability space on finite graphs instead of on the infinite limiting graph as we had done for regular graphs. This coupling takes into account the randomness in the local structure of the underlying Erdős–Rényi graphs, which is not an issue for regular graphs.

Define the stability $Q_{n,\lambda,p} = Q_{n,\lambda}(G_n, \circ, S, X)$ on the new probability space by

$$Q_{n,\lambda,p} = \mathbb{E}^*\left[f_{n,\lambda}(N_{r_\lambda}(G^1, \circ), \circ, Y^1(G^1, \circ, r_\lambda)|G_n, \circ, S, X)\right].$$

Calculating as before, for every $k \geq 1$ the moment $\mathbb{E}^*[Q_{n,\lambda}^{k-1}] = \frac{\alpha_{k,n,\lambda,p}}{\alpha_{1,n,\lambda}}$.

We now show that expectations involving $Q_{n,\lambda,p}$ are continuous with respect to p, and that $Q_{n,\lambda,p}$ has the right values at the endpoints $p = 0$ and $p = 1$. Observe that $Q_{n,\lambda,p} \in [0, 1]$. If g is a bounded measurable function on $[0, 1]$ then $\mathbb{E}^*[g(Q_{n,\lambda,p})]$ is Lipschitz in p.

Indeed, let $p_1 \leq p_2$. We couple the labelled graphs $(G^1(S_{p_1}), \circ, Y_{p_1}^1)$ and $(G^1(S_{p_2}), \circ, Y_{p_2}^1)$ given $(G_n, \circ, X)$ through the percolation subsets. Let Z be a random labelling of $[n]$, and let $\tau_{\{u,v\}}$ for $\{u, v\} \subset [n]$ be independent Bernoulli trials of expectation λ/n. Set $S_{p_1} = \{v : Z(v) \leq p_1\}$ and $S_{p_2} = \{v : Z(v) \leq p_2\}$. The resampled edges of $G^1(S_{p_1})$ [resp. $G^1(S_{p_2})$] are determined according to the $\tau_{\{u,v\}}$ for $u, v \in S_{p_1}$ (resp., for $u, v \in S_{p_2}$). Similarly, the labelling $Y_{p_1}^1$ (resp., $Y_{p_2}^1$) agrees with X^1 on S_{p_1} (resp., S_{p_2}) and agrees with X otherwise. With this coupling, we have that (ignoring some formalities with the notation)

$$
\begin{aligned}
&\mathbb{E}^*\big[g(Q_{n,\lambda,p_1})\big] - \mathbb{E}^*\big[g(Q_{n,\lambda,p_2})\big] \\
&\quad = \mathbb{E}^*\big[g\big(\mathbb{E}^*[f(G^1(S_{p_1}), \circ, Y_{p_1}^1)|G_n, \circ, X, S_{p_1}]\big) \\
&\qquad\quad - g\big(\mathbb{E}^*[f(G^1(S_{p_2}), \circ, Y_{p_2}^1)|G_n, \circ, X, S_{p_2}]\big)\big].
\end{aligned}
$$

Observe that if $Z(v) \notin (p_1, p_2)$ for every $v \in [n]$ then $S_{p_1} = S_{p_2}$, and hence, $(G^1(S_{p_1}), Y_{p_1}^1) = (G^1(S_{p_2}), Y_{p_2}^1)$. On this event, the conditional expectations $\mathbb{E}^*[f(G^1(S_{p_1}), \circ, Y_{p_1}^1)|G_n, \circ, X, S_{p_1}] = \mathbb{E}^*[f(G^1(S_{p_2}), \circ, Y_{p_2}^1)|G_n, \circ, X, S_{p_2}]$. So the difference of the two expectations above is zero on this event. By an union bound, the probability that $Z(v) \in (p_1, p_2)$ for some vertex v is at most $n|p_1 - p_2|$. Therefore, the triangle inequality implies that

$$
\big|\mathbb{E}^*\big[g(Q_{n,\lambda,p_1})\big] - \mathbb{E}^*\big[g(Q_{n,\lambda}, p_2)\big]\big| \leq (2\|g\|_\infty n)|p_1 - p_2|.
$$

The endpoint values of $Q_{n,\lambda,p}$ are the same as before. When $p = 0$, the resampled graph G^1 equals G_n, and the labelling $Y^1 = X$ due to S being empty. Consequently, $Q_{n,\lambda,0} \equiv 1$ on the restricted probability space. On the other hand, if $p = 1$ then (G^1, Y^1) is independent of (G_n, X) and the conditioning has no effect due to S being the entire vertex set. Note that the common root $\circ$ does not affect the calculation because the distribution of $N_r(\mathrm{ER}(n, \lambda/n), v)$ does not depend on v. We thus have $Q_{n,\lambda,1} \equiv \alpha_{1,n,\lambda} \frac{\log \lambda}{\lambda}$.

With these observations, we can now proceed with the proof exactly the same way as before. We skip the remainder of the argument for brevity and prove Theorem 4.2 in the following.

4.1.2. *Proof of Theorem* 4.2. We will show that the existence of the factor of i.i.d. independent sets I^i on the graph G^i implies that with high probability each graph G^i contains a subset S^i such that S^i is an independent set in G^i, and the empirical intersection densities of the $S^1, \ldots, S^k$ are close to the quantities $\alpha_{k,n,\lambda} \frac{\log \lambda}{\lambda}$.

Then we will bound the probability of observing such a k-tuple of independent sets, and prove that this probability is vanishingly small unless Theorem 4.2 holds.

Consider subsets $S^i \subset V(G^i)$. The density profile of $(S^1, \ldots, S^k)$ is the vector $\rho = (\rho(T); T \subset [k])$ where $\rho = (\rho(T); T \subset [k])$ defined as $\rho(T) = \frac{|\bigcap_{i \in T} S^i|}{n}$. Consider the independent sets I^i of G^i defined in Section 4.1.1. They satisfy $\mathbb{E}[|\bigcap_{j=1}^i I^j|/n] = \alpha_{i,n,\lambda,p} \frac{\log \lambda}{\lambda}$ for every $1 \leq i \leq k$.

Fix $0 < \varepsilon < 1$. Let $A(\varepsilon, p)$ be the following event. For each $1 \leq i \leq k$, G^i contains an independent set S^i such that the density profile of $(S^1, \ldots, S^k)$ satisfies the following for all $T \subset [k]$:

$$\rho(T) \in \left[(1 - \varepsilon)\alpha_{|T|,n,\lambda,p} \frac{\log \lambda}{\lambda}, (1 + \varepsilon)\alpha_{|T|,n,\lambda,p} \frac{\log \lambda}{\lambda} \right].$$

We show that $\mathbb{P}[A(\varepsilon, p)] \to 1$ as $n \to \infty$. This follows if we show that

$$\mathbb{P}\left[\max_{T \subset [k]} \left| \frac{|\bigcap_{i \in T} I^i|}{n} - \mathbb{E}\left[\frac{|\bigcap_{i \in T} I^i|}{n} \right] \right| > \varepsilon \right] \to 0.$$

Indeed, given a realization of the graphs $G^1, \ldots, G^k$ and random labellings $Y^1, \ldots, Y^k$, we take $S^i = I^i$ on G^i. If $|\frac{\bigcap_{i \in T} I^i}{n} - \mathbb{E}[\frac{\bigcap_{i \in T} I^i}{n}]| \leq \varepsilon$ for every $T \subset [k]$ then the conditions for $A(\varepsilon, p)$ to occur are satisfied.

LEMMA 4.3. *With $G^1, \ldots, G^k$ as defined and corresponding independent sets $I^1, \ldots, I^k$ as defined via the factor $f_{n,\lambda}$, one has that for all $\varepsilon > 0$, as $n \to \infty$,*

$$\mathbb{P}\left[\max_{T \subset [k]} \left| \frac{|\bigcap_{i \in T} I^i|}{n} - \mathbb{E}\left[\frac{|\bigcap_{i \in T} I^i|}{n} \right] \right| > \varepsilon \right] \to 0.$$

PROOF. We show that $\mathbb{E}[||\bigcap_{i \in T} I^i| - \mathbb{E}[|\bigcap_{i \in T} I^i|]|^2] = o(n^2)$ where the little o term may depend on λ, r_λ, and k. The lemma then follows from Chebyshev's inequality and an union bound over $T \subset [k]$. We write

$$\left| \bigcap_{i \in T} I^i \right| - \mathbb{E}\left[\left| \bigcap_{i \in T} I^i \right| \right] = \sum_{v=1}^n \mathbf{1}\left\{ v \in \bigcap_{i \in T} I^i \right\} - \mathbb{P}\left[v \in \bigcap_{i \in T} I^i \right].$$

Observe that $|\mathbf{1}\{v \in \bigcap_{i \in T} I^i\} - \mathbb{P}[v \in \bigcap_{i \in T} I^i]| \leq 1$. For two vertices u and v, if the graph distance $\text{dist}_{G^i}(u, v) > 2r_\lambda$ then the events $\{u \in I^i\}$ and $\{v \in I^i\}$ are independent with respect to the random labelling of G^i because the factor $f_{n,\lambda}$ makes decisions based on the labels along the r_λ-neighbourhood of a vertex. Therefore, $\{u \in \bigcap_{i \in T} I^i\}$ is independent of $\{v \in \bigcap_{i \in T} I^i\}$ if $\text{dist}_{G^i}(u, v) \geq 2r_\lambda$ for every i. Consequently,

$$\mathbb{E}\left[\left(\mathbf{1}\left\{ u \in \bigcap_{i \in T} I^i \right\} - \mathbb{P}\left[u \in \bigcap_{i \in T} I^i \right] \right) \cdot \left(\mathbf{1}\left\{ v \in \bigcap_{i \in T} I^i \right\} - \mathbb{P}\left[v \in \bigcap_{i \in T} I^i \right] \right) \right]$$

$$\leq \mathbb{P}[\text{dist}_{G^i}(u, v) \leq 2r_\lambda \text{ for some } i].$$

These observations imply that

$$\mathbb{E}\left[\left|\left\|\bigcap_{i \in T} I^i\right\| - \mathbb{E}\left[\left\|\bigcap_{i \in T} I^i\right\|\right]\right|^2\right]$$

$$\leq n + \mathbb{E}\left[\#\{(u, v) : \mathrm{dist}_{G^i}(u, v) \leq 2r_\lambda \text{ for some } i\}\right].$$

Using the fact that the random graphs $\mathrm{ER}(n, \lambda/n)$ converge locally to PGW_λ, it is a standard exercise to show that the expected number of pairs (u, v) in $\mathrm{ER}(n, \lambda/n)$ that satisfy $\mathrm{dist}(u, v) > R$ is $o(n^2)$ (the little o term depends on λ and R). From this observation and a union bound over i, we deduce that $\mathbb{E}[\#\{(u, v) : \mathrm{dist}_{G^i}(u, v) \leq 2r_\lambda \text{ for some } i\}] = o(n^2)$. This proves the estimate for the squared expectation and completes the proof. $\quad\square$

In order to bound $\mathbb{P}[A(\varepsilon, p)]$ from above, we need a procedure to sample the graphs $G^1, \dots, G^k$.

Sampling the graphs $(G^1, \dots, G^k)$. Let $\tau_{i,u,v}$ for $1 \leq i \leq k$ and $\{u, v\} \subset [n]$ be the indicator of the event that the edge $\{u, v\}$ belongs to G^i. Then the random vectors $(\tau_{i,u,v}; 1 \leq i \leq k)$ are independent of each other as $\{u, v\}$ varies. Let $S \subset [n]$ be a random subset chosen by a Bernoulli percolation with density p. If both $u, v \in S$, then $(\tau_{i,u,v}; 1 \leq i \leq k)$ are independent Bernoulli trials of expectation λ/n for each i. Otherwise, $(\tau_{i,u,v}; 1 \leq i \leq k)$ satisfies $\tau_{1,u,v} = \cdots = \tau_{k,u,v}$. In the latter case, all k of these indicators take the value 1 with probability λ/n or they are all zero with the complementary probability.

The sampling procedure above will allow us to compute expectations involving independent sets in the G^i. Let $I^i \subset [n]$ be an independent set of G^i. Defining $\rho(T) = |\bigcap_{t \in T} I^t|/n$ for $T \subset [k]$, the density profile associated to these k independent sets is $\rho = (\rho(T); T \subset [k])$. The density profile ρ determines a probability distribution $\pi = (\pi(T); T \subset [k])$ by equation (3.1). Let $Z(\rho)$ be the number of k-tuple of subsets $(I^1, \dots, I^k)$ of $[n]$ such that they have density profile ρ and I^i is an independent set of G^i.

LEMMA 4.4. *The expectation of $Z(\rho)$ satisfies*

$$\mathbb{E}[Z(\rho)] \leq \binom{n}{\pi(T)n; T \subset [k]} \left(1 - \frac{\lambda}{n}\right)^\kappa,$$

where $\kappa = \sum_{T \neq \varnothing} \binom{\pi(T)n}{2} + \frac{1}{2} \sum_{(T,T'): T \neq T', T \cap T' \neq \varnothing} \pi(T)\pi(T')n^2$.

PROOF. Consider a k-tuple $(I^1, \dots, I^k)$, each $I^i \subset [n]$, with the property that its density profile is ρ. Given $v \in [n]$, let $T_v = \{i \in [k] : v \in I^i\}$. Let $E_{\{u,v\}}$ be the event that the edge $\{u, v\}$ is absent is all G^i for which $i \in T_u \cap T_v$, that is, $E_{\{u,v\}} = \{\tau_{i,u,v} = 0 \text{ for all } i \in T_u \cap T_v\}$. The subsets $I^1, \dots, I^k$ have the property

that I^i is an independent set of G^i if and only if the events $E_{\{u,v\}}$ occur for all pairs $\{u, v\}$.

From the sampling procedure for the graphs $G^1, \ldots, G^k$, we note that the events $E_{\{u,v\}}$ are independent. Conditioning on the random subset S and using the sampling procedure, we conclude that

$$\mathbb{E}[Z(\rho)|S] = \sum_{\substack{(I^1, \ldots, I^k) \\ \text{with density profile } \rho}} \prod_{\{u,v\} \subset S} \left(1 - \frac{\lambda}{n}\right)^{|T_u \cap T_v|} \prod_{\{u,v\} \not\subset S} \left(1 - \frac{\lambda}{n}\right)^{\mathbf{1}_{\{T_u \cap T_v \neq \varnothing\}}}.$$

Observe that $(1 - \frac{\lambda}{n})^{|T_u \cap T_v|} \leq (1 - \frac{\lambda}{n})^{\mathbf{1}_{\{T_u \cap T_v \neq \varnothing\}}}$. Therefore, no matter the outcome of S we have that

$$\mathbb{E}[Z(\rho)|S] \leq \sum_{\substack{(I^1, \ldots, I^k) \\ \text{with density profile } \rho}} \prod_{\{u,v\}} \left(1 - \frac{\lambda}{n}\right)^{\mathbf{1}_{\{T_u \cap T_v \neq \varnothing\}}}.$$

This implies the same inequality for the expectation $\mathbb{E}[Z(\rho)]$. We show that $\sum_{\{u,v\}} \mathbf{1}_{\{T_u \cap T_v \neq \varnothing\}} = \kappa$. Recall the probability distribution $(\pi(T); T \subset [k])$ derived from ρ from equation (3.1). Consider the ordered partition Π associated to any $(I^1, \ldots, I^k)$ with profile ρ. The partition Π has 2^k ordered cells $(\Pi(T); T \subset [k])$ defined by

$$\Pi(T) = \left(\bigcap_{t \in T} I^t\right) \cap \left(\bigcap_{t \notin T} [n] \setminus I^t\right).$$

The inclusion–exclusion principle implies that $|\Pi(T)| = \pi(T)n$. For a fixed k-tuple $(I^1, \ldots, I^k)$, the sum $\sum_{\{u,v\}} \mathbf{1}_{\{T_u \cap T_v \neq \varnothing\}}$ can be represented by accounting for the contribution of each pair of subsets $\{T, T'\}$ to it.

$$\sum_{\{u,v\}} \mathbf{1}_{\{T_u \cap T_v \neq \varnothing\}} = \sum_{\{T, T'\} : T \cap T' \neq \varnothing} \#\{\{u, v\} : T_u = T, T_v = T'\}.$$

Observe that $\Pi(T_u)$ is the cell of Π that contains u, that is, $T_u = T$ if and only if $u \in \Pi(T)$. Therefore, $\#\{\{u, v\} : T_u = T, T_v = T'\}$ equals $\#\{\{u, v\} : u \in \Pi(T), v = \in \Pi(T')\}$, and the latter is

$$|\Pi(T)||\Pi(T')| - \binom{|\Pi(T) \cap \Pi(T')|}{2} - |\Pi(T) \cap \Pi(T')|.$$

Since $|\Pi(T)| = \pi(T)n$ and $\Pi(T) \cap \Pi(T') = \varnothing$ for $T \neq T'$, it readily follows that $\sum_{\{u,v\}} \mathbf{1}_{\{T_u \cap T_v \neq \varnothing\}}$ equals κ. [The factor of $1/2$ appears in κ because we sum over all ordered pairs (T, T').] Thus,

$$\mathbb{E}[Z(\rho)] \leq \#\{(I^1, \ldots, I^k) \text{ with density profile } \rho\} \times \left(1 - \frac{\lambda}{n}\right)^{\kappa}.$$

The bijection between k-tuples $(I^1, \ldots, I^k)$ and ordered partitions implies that the number of k-tuples with density profile ρ is equal to the number of ordered partitions $(\Pi(T); T \subset [k])$ of $[n]$ such that $|\Pi(T)| = \pi(T)n$. The latter number is the multinomial coefficient $\binom{n}{\pi(T)n; T \subset [k]}$. $\quad\square$

To analyze the asymptotic behaviour of $\mathbb{E}[Z(\rho)]$, write $\pi(T) = \beta(T)\frac{\log \lambda}{\lambda}$ for $T \neq \varnothing$. Then,

$$\sum_{T \neq \varnothing} \binom{\pi(T)n}{2} + \frac{1}{2} \sum_{\substack{(T, T') \\ T \neq T', T \cap T' \neq \varnothing}} \pi(T)\pi(T')n^2$$

$$= \frac{n^2}{2} \sum_{\substack{(T, T') \\ T \cap T' \neq \varnothing}} \pi(T)\pi(T') - \frac{n}{2} \sum_{T \neq \varnothing} \pi(T).$$

In terms of β, the latter is $\frac{n^2 \log^2 \lambda}{2\lambda^2} \sum_{(T, T') T \cap T' \neq \varnothing} \beta(T)\beta(T') - \frac{n}{2}(1 - \pi(\varnothing))$. Also, considering only the nonzero $\pi(T)$ and using Sterling's approximation, we have $\binom{n}{\pi(T)n; T \neq \varnothing} = O_k(\sqrt{n}) \exp\{nH(\pi)\}$, where H is the previously introduced entropy function. Using the fact that $1 - \frac{\lambda}{n} \leq e^{-\frac{\lambda}{n}}$, and $1 - \pi(\varnothing) \leq 1$, we conclude that

$$\mathbb{E}[Z(\rho)] = O_k(\sqrt{n}e^{\lambda/2}) \exp\left\{ n\left[H(\pi) - \frac{\log^2 \lambda}{2\lambda} \sum_{\substack{(T, T') \\ T \cap T' \neq \varnothing}} \beta(T)\beta(T') \right] \right\}.$$

Lemma 3.3 shows that $H(\pi) = \frac{\log^2 \lambda}{\lambda} \sum_{T \neq \varnothing} \beta(T) + O_k(\frac{\log \lambda}{\lambda})$, where the big O constant depends on k. Consequently,

$$\begin{aligned}
& H(\pi) - \frac{\log^2 \lambda}{2\lambda} \sum_{\substack{(T, T'): \\ T \cap T' \neq \varnothing}} \beta(T)\beta(T') \\
(4.3) \\
& = \frac{\log^2 \lambda}{\lambda} \left(\sum_{T \neq \varnothing} \beta(T) - \frac{1}{2} \sum_{\substack{(T, T'): \\ T \cap T' \neq \varnothing}} \beta(T)\beta(T') \right) + O_k\left(\frac{\log \lambda}{\lambda} \right).
\end{aligned}$$

Lemma 3.3 also shows that if $\rho(S) = \alpha_{|S|}\frac{\log \lambda}{\lambda}$ for $S \neq \varnothing$ $[\rho(\varnothing) = 1]$, then

$$(4.4) \qquad \sum_{T \neq \varnothing} \beta(T) - \frac{1}{2} \sum_{\substack{(T, T') \\ T \cap T' \neq \varnothing}} \beta(T)\beta(T') = \frac{1}{2} \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} (2 - \alpha_i)\alpha_i.$$

Recall the event $A(\varepsilon, p)$: the graph G^i contains an independent set S^i such that the density profile of $(S^1, \ldots, S^k)$ satisfies

$$\rho(T) \in \left[(1 - \varepsilon)\alpha_{|T|, n, \lambda, p} \frac{\log \lambda}{\lambda}, (1 + \varepsilon)\alpha_{|T|, n, \lambda, p} \frac{\log \lambda}{\lambda} \right] \qquad \text{for all } T \subset [k].$$

We employ a first moment bound along with Lemma 4.4 to bound $\mathbb{P}[A(\varepsilon, p)]$; simplifying via (4.3) and (4.4) we get

$$\mathbb{P}[A(\varepsilon, p)] \leq \exp\left\{ n \left[\frac{\log^2 \lambda}{\lambda} \left(\sum_i (-1)^{i-1} \binom{k}{i} \left(\alpha_{i, n, \lambda, p} - \frac{1}{2} \alpha_{i, n, \lambda, p}^2 \right) \right. \right. \right.$$
$$\left. \left. \left. + \, \mathrm{err}(\varepsilon) \right) + O_k\left(\frac{\log \lambda}{\lambda} \right) \right] \right\},$$

where $\mathrm{err}(\varepsilon) \to 0$ as $\varepsilon \to 0$ uniformly in n, λ and p.

The proof of Theorem 4.2 is now completed in the same manner as for regular graphs, which is the argument from Section 3.3.1.

4.2. *A lower bound from regular trees.* We will show that factor of i.i.d. independent sets in regular trees can be used to construct such independent sets in PGW trees as well. Let I be a factor of i.i.d. independent set in the regular tree $\mathbb{T}_d$ such that the factor is a function of the labels in a finite size neighbourhood of the root. Let $E(\lambda, d)$ denote the event that the root of PGW_λ and all of its neighbours have degree at most d.

THEOREM 4.5. *Given I as above there exists a factor of i.i.d. independent set J of PGW_λ whose density satisfies the bound*:

$$\mathrm{density}(I)\mathbb{P}[E(\lambda, d)] \leq \mathrm{density}(J) \leq \mathrm{density}(I).$$

PROOF. We construct J in three stages.

The edge removal stage. We remove edges from PGW_λ via a factor of i.i.d. process such that all vertices will have degree at most d after the removal procedure. Begin with a random labelling X of PGW_λ. For each vertex v, consider all the neighbours u of v such that the variables X_u are the $\mathrm{degree}(v) - d$ highest in value [provided, of course, that $\mathrm{degree}(v) > d$]. Mark the $\mathrm{degree}(v) - d$ edges connecting v to these neighbours.

Following the marking procedure remove all the edges that have been marked. After the removal of edges, all vertices have degree at most d. The remaining graph is a disjoint collection of trees with a countable number of components. Denote it G.

The filling out stage. If a vertex v in G has degree $\mathrm{degree}_G(v) < d$, then attach to it $d - \mathrm{degree}_G(v)$ copies of a $(d-1)$-ary tree via $d - 1$ separate edges connecting v to these trees. Following this procedure, the graph G becomes a disjoint collection of d-regular trees. Randomly label G by a new set of labels Y that are independent of X.

The inclusion stage. Since G is a disjoint collection of d-regular trees, we can use the factor associated to I with input Y to construct an independent set I' of G with the same density as I. Although I' is an independent set in G, it may not be an independent set in the original tree PGW_λ due to the removal of edges. To construct the independent set J, we include in J all vertices $v \in I'$ such that no edges incident to v were removed during the edge removal stage.

By design, the random subset J is a factor of i.i.d. process on PGW_λ. J is also an independent set because if (u, v) is an edge of PGW_λ with both $u, v \in I'$, then the edge connecting u and v must have been removed during the edge removal stage (due to I' being an independent set in G). Thus, neither u nor v belong to J.

To bound the density of J, we note that $J \subset I'$. Also, for any $v \in I'$, if v and all of its neighbours in PGW_λ has degree at most d then none of the edges incident to v are removed during the edge removal stage. Consequently, v will be included in J. These two observations readily imply that

$$\mathrm{density}(I)\mathbb{P}\big[E(\lambda, d)\big] \leq \mathrm{density}(J) \leq \mathrm{density}(I). \qquad \square$$

LEMMA 4.6. *If $\lambda = d - d^u$ for any $1/2 < u < 1$, then the probability $\mathbb{P}[E(\lambda, d)] \to 1$ as $d \to \infty$.*

PROOF.   This is a calculation involving Poisson tail probabilities. Recall that the moment generating function of a Poisson(μ) random variable is $e^{\mu(e^t - 1)}$. Let X denote the degree of the root in a PGW tree of expected degree λ. Let $Z_1, \ldots, Z_X$ denote the number of offsprings of the neighbours of the root. Recall that X has distribution Poisson(λ), and that conditioned on X the random variables $Z_1, \ldots, Z_X$ are i.i.d. with distribution Poisson(λ).

Let $p(\lambda, d) = \mathbb{P}[\mathrm{Poisson}(\lambda) > d]$. Then

$$\mathbb{P}\big[E(\lambda, d)\big] = \mathbb{E}\left[\mathbf{1}_{X \leq d} \prod_{i=1}^{X} \mathbf{1}_{Z_i \leq d-1}\right]$$

$$= \mathbb{E}\left[\mathbf{1}_{X \leq d}\mathbb{E}\left[\prod_{i=1}^{X} \mathbf{1}_{Z_i \leq d-1} | X\right]\right]$$

$$= \mathbb{E}\big[\mathbf{1}_{X \leq d}(1 - p(\lambda, d-1))^X\big]$$

$$= \mathbb{E}\big[(1 - p(\lambda, d-1))^X\big] - \mathbb{E}\big[\mathbf{1}_{X > d}(1 - p(\lambda, d-1))^X\big]$$

$$\geq e^{-\lambda p(\lambda, d-1)} - p(\lambda, d-1).$$

We can bound the tail probability $p(\lambda, d-1)$ by using the exponential moment method. For simplicity, we replace $d - 1$ by d, which makes no difference to the analysis for large d. A simple and well-known computation gives

$$p(\lambda, d) \leq e^{d - \lambda}\left(\frac{\lambda}{d}\right)^d \qquad \text{if } \lambda < d.$$

Setting $\lambda = d - d^u$ for $1/2 < u < 1$, we see from the bound above that $p(d - d^u, d) \leq e^{d^u}(1 - d^{u-1})^d = e^{d^u + d \log(1 - d^{u-1})}$. Since $\log(1 - x) = -\sum_{k \geq 1} \frac{x^k}{k} \leq -x - x^2/2$ for $0 \leq x < 1$, by setting $x = d^{u-1} < 1$ we conclude that

$$p(d - d^u, d) \leq e^{d^u - d(d^{u-1} + \frac{d^{2u-2}}{2})} = e^{-\frac{d^{2u-1}}{2}}.$$

Due to $u > 1/2$, the latter quantity tends to 0 exponentially fast as $d \to \infty$. As a result, both $(d - d^u)p(d - d^u, d)$ and $p(d - d^u, d)$ tend to 0 with d. This implies the lemma. $\square$

Given λ, set $d = \lceil \lambda + \lambda^{3/4} \rceil$. From the definition of $\alpha(\lambda), \alpha_d$, and the conclusion of Theorem 4.5 we have that

$$\alpha(\lambda) \frac{\log \lambda}{\lambda} \geq \alpha_d \cdot \frac{\log d}{d} \cdot \mathbb{P}[E(\lambda, d)].$$

Recall the construction of Lauer and Wormald mentioned in Section 2 which shows that $\liminf_{d \to \infty} \alpha_d \geq 1$. By our choice to d as a function of λ we have $(\frac{\log d}{d})/(\frac{\log \lambda}{\lambda}) \to 1$ as $\lambda \to \infty$. By Lemma 4.6, we have that the probability $\mathbb{P}[E(\lambda, d)] \to 1$ as $\lambda \to \infty$. As a result, we conclude from the inequality above that

$$\liminf_{\lambda \to \infty} \alpha(\lambda) \geq 1.$$

This lower bound completes the proof of Theorem 4.1.

5. Concluding remarks. Our results are concerned with density of factor of i.i.d. independent sets in sparse graphs where the sparsity parameter (degree) tends to infinity. However, it is still a very interesting problem to compute the maximum density for various classes of factor of i.i.d. processes on d-regular graphs for fixed values of d. There have been some recent progress is this regard for independent sets in 3-regular graphs. In [8], the authors use Gaussian processes to construct factor of i.i.d. independent sets in $\mathbb{T}_3$ of density at least 0.436, and in [17] the authors improve the bound to at least 0.437 via another local algorithm. It is known due to McKay [21] that the maximum density of independent sets in random 3-regular graphs is at most 0.456, which provides an upper bound for factor of i.i.d. independent sets in $\mathbb{T}_3$.

Another question is whether there is a gap between the density of the maximum cut in random d-regular graphs and the maximal density of cuts derived from local algorithms. The density of the maximum cut of $\mathcal{G}_{n,d}$, denoted $\gamma(\mathcal{G}_{n,d})$, is the largest values of $|\partial_E S|/nd$, where $\partial_E S$ is the set of all edges $(u, v) \in E(\mathcal{G}_{n,d})$ with $u \in S$ and $v \notin S$. The expectation $\mathbb{E}[\gamma(\mathcal{G}_{n,d})] \to \gamma(d)$ as $n \to \infty$ for every d [2]. A local cut of $\mathbb{T}_d$ is a factor of i.i.d. process $\sigma \in \{-1, +1\}^{\mathbb{T}_d}$; its density is $\mathbb{P}[\sigma(\circ) \neq \sigma(\circ')]$, where $(\circ, \circ')$ is a fixed edge of $\mathbb{T}_d$. Is it true that $\gamma(d)$ equals the supremum over the density of local cuts of $\mathbb{T}_d$?

REFERENCES

[1] BACKHAUSZ, Á., SZEGEDY, B. and VIRÁG, B. (2015). Ramanujan graphings and correlation decay in local algorithms. *Random Structures Algorithms* **47** 424–435. MR3385741

[2] BAYATI, M., GAMARNIK, D. and TETALI, P. (2013). Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. *Ann. Probab.* **41** 4080–4115. MR3161470

[3] BOLLOBÁS, B. (1981). The independence ratio of regular graphs. *Proc. Amer. Math. Soc.* **83** 433–436. MR0624948

[4] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. Cambridge Univ. Press, Cambridge. MR1864966

[5] BORDENAVE, C., LELARGE, M. and SALEZ, J. (2013). Matchings on infinite graphs. *Probab. Theory Related Fields* **157** 183–208. MR3101844

[6] COJA-OGHLAN, A. and EFTHYMIOU, C. (2015). On independent sets in random graphs. *Random Structures Algorithms* **47** 436–486. MR3385742

[7] CSÓKA, E. (2012). Random local algorithms. Preprint. Available at arXiv:1202.1565.

[8] CSÓKA, E., GERENCSÉR, B., HARANGI, V. and VIRÁG, B. (2015). Invariant Gaussian processes and independent sets on regular graphs of large girth. *Random Structures Algorithms* **47** 284–303. MR3382674

[9] CSÓKA, E. and LIPPNER, G. (2012). Invariant random matchings in Cayley graphs. Preprint. Available at arXiv:1211.2374.

[10] DING, J., SLY, A. and SUN, N. (2013). Maximum independent sets in random regular graphs. Preprint. Available at arXiv:1310.4787.

[11] ELEK, G. and LIPPNER, G. (2010). Borel oracles. An analytical approach to constant-time algorithms. *Proc. Amer. Math. Soc.* **138** 2939–2947. MR2644905

[12] FRIEZE, A. M. (1990). On the independence number of random graphs. *Discrete Math.* **81** 171–175. MR1054975

[13] FRIEZE, A. M. and ŁUCZAK, T. (1992). On the independence and chromatic numbers of random regular graphs. *J. Combin. Theory Ser. B* **54** 123–132. MR1142268

[14] GAMARNIK, D. and GOLDBERG, D. A. (2010). Randomized greedy algorithms for independent sets and matchings in regular graphs: Exact results and finite girth corrections. *Combin. Probab. Comput.* **19** 61–85. MR2575098

[15] GAMARNIK, D. and SUDAN, M. (2014). Limits of local algorithms over sparse random graphs [extended abstract]. In *ITCS'14—Proceedings of the* 2014 *Conference on Innovations in Theoretical Computer Science* 369–375. ACM, New York. MR3359490

[16] HATAMI, H., LOVÁSZ, L. and SZEGEDY, B. (2014). Limits of locally-globally convergent graph sequences. *Geom. Funct. Anal.* **24** 269–296. MR3177383

[17] HOPPEN, C. and WORLMALD, N. (2013). Local algorithms, regular graphs of large girth, and random regular graphs. Preprint. Available at arXiv:1308.0266v2.

[18] JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs*. Wiley, New York. MR1782847

[19] LAUER, J. and WORMALD, N. (2007). Large independent sets in regular graphs of large girth. *J. Combin. Theory Ser. B* **97** 999–1009. MR2354714

[20] LYONS, R. and NAZAROV, F. (2011). Perfect matchings as IID factors on non-amenable groups. *European J. Combin.* **32** 1115–1125. MR2825538

[21] McKAY, B. D. (1987). Independent sets in regular graphs of high girth. In *Proceedings of the Singapore Conference on Combinatorial Mathematics and Computing* (*Singapore*, 1986) **23** 179–185. Ars Combin., A. MR0890138

[22] MÉZARD, M. and MONTANARI, A. (2009). *Information, Physics, and Computation*. Oxford Univ. Press, Oxford. MR2518205

[23] SHEARER, J. B. (1991). A note on the independence number of triangle-free graphs. II. *J. Combin. Theory Ser. B* **53** 300–307. MR1129557

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TORONTO
40 ST. GEORGE STREET
TORONTO, ON M5S 2E4
CANADA
E-MAIL: mustazee@math.toronto.edu
        balint@math.toronto.edu