Sub-optimality of local algorithms on sparse random graphs

by

Mustazee Rahman

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Mathematics
University of Toronto

© Copyright 2015 by Mustazee Rahman
Abstract

Sub-optimality of local algorithms on sparse random graphs

Mustazee Rahman
Doctor of Philosophy
Graduate Department of Mathematics
University of Toronto
2015

This thesis studies local algorithms for solving combinatorial optimization problems on large, sparse random graphs. Local algorithms are randomized algorithms that run in parallel at each vertex of a graph by using only local information around each vertex. In practice, they generate important structures on large graphs such as independent sets, matchings, colourings and eigenfunctions of the graph Laplacian with efficient run-time and little memory usage. They have also been used to solve instances of constraint satisfaction problems such as \(k\)-SAT.

Hatami, Lovász and Szegedy conjectured that all reasonable optimization problems on large random \(d\)-regular graphs can be approximately solved by local algorithms. This is true for matchings: local algorithms can produce near perfect matchings in random \(d\)-regular graphs. However, this conjecture was recently shown to be false for independent sets by Gamarnik and Sudan. They showed that local algorithms cannot generate maximal independent sets on large random \(d\)-regular graphs if the degree \(d\) is sufficiently large.

We prove an optimal quantitative measure of failure of this conjecture for the problem of percolation on graphs. The basic question is this. Consider a large integer \(\tau\), which is a threshold parameter. Given some large graph \(G\), find the maximum sized induced subgraph of \(G\) whose connected components have size no bigger than \(\tau\). When \(\tau\) equals 1 this is the problem of finding the maximum independent sets of \(G\).

We show that for large values of the degree \(d\), local algorithms cannot solve the percolation problem on random \(d\)-regular graphs and Erdős-Rényi graphs of expected average degree \(d\). Roughly speaking, we prove that for any given threshold the largest induced subgraphs that local algorithms can find have size no more than half of the maximal ones. This is optimal because there exist local algorithms that find such induced subgraphs of half the maximum size.

Part of this thesis represents joint work with Bálint Virág.
Dedication

To my parents and my teachers.
Acknowledgements

This thesis would not have been possible without the guidance of my wonderful advisor, Bálint Virág. His constant tutelage has made me a better thinker and his many problems have helped me learn a lot of interesting mathematics. He taught me the importance of working on meaningful problems and thinking in concrete terms. His care and advice on matters beside mathematics have also been invaluable. Thank you, Bálint, for everything; I could not have hoped for a better advisor and friend.

I am sincerely grateful to Steve Tanny for peeking my interest in mathematical research and for providing me an opportunity to work on open problems as an undergraduate. I probably would not have continued with mathematics if it weren’t for my fantastic experience while working with him. His encouragement and support has meant a lot to me over the years.

I would also like to thank Miklós Abért for hosting me at the Rényi Institute for several months at a time. The engaging atmosphere at the institute along with Miklós’s kindness and hospitality provided a great environment for carrying out parts of my research. I shared fruitful conversations with several members of the institute and I thank them all for their interest and time. Special thanks go to Ági Backhausz, Viktor Harangi, Gábor Pete and Máté Vizer for valuable discussions on research.

Of course, my professors at University of Toronto have provided me with much guidance over the years. I am especially grateful to Almut Burchard, Jim Colliander, Larry Guth, Pierre Milman, Mike Molloy and Jeremy Quastel.

The departmental staff has also been very helpful as they guided me through many administrative matters. I thank Ida Bulat and Jemima Merisca in particular for being so kind to me and for always being there to help.

I was supported by NSERC graduate scholarships and departmental research assistantships during several stages of this work. These scholarships have benefitted my research greatly.

Finally, I want to thank my parents, my sister and all my friends who have given me so much love and support during these years. To my family – thank you for encouraging me to follow my own path. And to my friends – thank you for sharing good times and comforting me through the bad; I have learned so much that I will always remain indebted to you.
## List of notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{E}[X]$</td>
<td>The expectation of the random variable $X$.</td>
</tr>
<tr>
<td>$\mathbb{P}(E)$</td>
<td>The probability of the event $E$.</td>
</tr>
<tr>
<td>$1_{{E}}$</td>
<td>The indicator function of the event $E$.</td>
</tr>
<tr>
<td>$T_d$</td>
<td>The rooted $d$-regular tree.</td>
</tr>
<tr>
<td>$\text{PGW}_\lambda$</td>
<td>Poisson-Galton-Watson tree of expected degree $\lambda$.</td>
</tr>
<tr>
<td>$\mathcal{G}_{n,d}$</td>
<td>Random $d$-regular graph on $n$ vertices.</td>
</tr>
<tr>
<td>$\text{ER}(n,p)$</td>
<td>Erdős–Rényi graph on $n$ vertices and edge inclusion probability $p$.</td>
</tr>
<tr>
<td>$\text{Bin}(n,p)$</td>
<td>Binomial random variable with parameter values $n$ (positive integer) and $p \in [0,1]$.</td>
</tr>
<tr>
<td>Poisson($\lambda$)</td>
<td>Poisson random variable with mean $\lambda$.</td>
</tr>
<tr>
<td>$X$</td>
<td>IID random variables over a graph with Uniform$[0,1]$ marginals.</td>
</tr>
<tr>
<td>$\text{den}(Y)$</td>
<td>The density of a factor of iid percolation process $Y$.</td>
</tr>
<tr>
<td>$\text{corr}(Y)$</td>
<td>The correlation of a factor of iid percolation process $Y$.</td>
</tr>
<tr>
<td>$\text{avdeg}(Y)$</td>
<td>The average degree of a factor of iid percolation process $Y$.</td>
</tr>
<tr>
<td>$h(x)$</td>
<td>The function $h(x) = -x \log(x)$ with the convention that $h(0) = 0$.</td>
</tr>
<tr>
<td>$H(\mu)$</td>
<td>The entropy of a finitely supported probability measure $\mu$.</td>
</tr>
<tr>
<td>$\Psi(c)$</td>
<td>The function $\Psi(c) = c \log(c) - c + 1$ with $\Psi(0) = 1$.</td>
</tr>
<tr>
<td>$(P,\pi)$</td>
<td>The edge–profile induced by a colouring of a graph.</td>
</tr>
<tr>
<td>$(u,v)$</td>
<td>A directed edge in a graph from vertex $u$ to vertex $v$.</td>
</tr>
</tbody>
</table>
# Contents

1 Introduction
   1.1 Background and motivation ........................................ 1
   1.2 Summary of results .................................................. 4
   1.3 Organization of the thesis ......................................... 10

2 Percolation with small clusters in random graphs .................. 11
   2.1 A brief history of the problem ...................................... 11
   2.2 Preliminaries and terminology ...................................... 12
   2.3 Random regular graphs ............................................. 13
   2.4 Erdős–Rényi graphs ................................................ 24

3 Factor of IID processes on regular trees ................................. 28
   3.1 Connection to processes on random regular graphs ............... 28
   3.2 The fundamental entropy inequality ................................ 31

4 Factor of IID percolation on regular trees ............................... 35
   4.1 Bounds on the density in terms of correlation ..................... 36
   4.2 Percolation with finite clusters – an asymptotically optimal bound .... 40
   4.3 Density bounds for small values of $d$ ................................ 50
   4.4 FIID processes on the edges of $T_d$ ................................ 52

5 Percolation on Poisson–Galton–Watson trees .......................... 55
   5.1 An asymptotically optimal upper bound ........................... 55
   5.2 A lower bound from regular trees .................................. 64

6 Concluding remarks and open problems ............................... 68

Bibliography ............................................................... 71
Chapter 1

Introduction

We study a class of randomized distributed algorithms, called local algorithms, that are used to solve combinatorial problems on large, sparse graphs. Due to the local nature of these algorithms they can be efficiently implemented in practice, have reasonable memory requirements and run fast. They provide a practical framework to solve computational problems on large graphs such as neural networks, social networks, the world wide web, etc. Local algorithms have been used to produce graph structures such as independent sets, matchings, cuts sets and approximate eigenvectors of the graph Laplacian in constant time (see \[16, 17, 22, 28, 30, 33, 34, 35\] and the references therein). Iterative local algorithms, such as Belief propagation, have also been used in physics and computer science to find satisfying assignments for constraint satisfaction problems such as \(k\)-SAT and NAE-SAT \[26, 32, 39\]. My contribution, however, is of a negative variety and proves a ‘hardness of approximation’ result for using local algorithms to solve certain combinatorial problems on sparse random graphs.

1.1 Background and motivation

A local algorithm is a randomized algorithm that runs in parallel at each vertex of a graph to produce global structures by using only local information around each vertex. An intuitive description of local algorithms is as follows. The input to the algorithm is a graph \(G\). The algorithm decorates \(G\) with independent, identically distributed (IID) random variables \((X(v), v \in V(G))\). The output is \((f(i(v)); v \in G)\) where the ‘rule’ \(f\) depends on the isomorphism class \(i(v)\) of the labelled, rooted \(r\)-neighbourhood of \(v\) for some fixed \(r\) (not dependent on the size of \(G\)). The process \((f(i(v)); v \in G)\) generated by the local algorithm is called a factor of IID (FIID) process. The prototypical example of an FIID process is Bernoulli percolation: for a parameter \(0 \leq p \leq 1\) a random subset of \(G\) is constructed by including each vertex independently with probability \(p\).

FIID processes arise naturally in the theory of sparse graph limits as developed by Hatami, Lovász and Szegedy \[28\]. The authors pose the question of what processes over large graphs can be realized as FIID processes on \(T_d\), and make the conjecture that all reasonable optimization
problems over random d-regular graphs can be approximated by FIID processes over $T_d$ (see [28], Conjecture 7.13). Although this conjecture is true for perfect matchings [12, 17, 35], it was refuted for maximal independent sets by Gamarnik and Sudan [25]. Our work provides an optimal quantitative bound on the failure on this conjecture for independent sets, along with a similar quantitative bound on its failure for another natural problem: finding maximal induced subgraphs having connected components of bounded size. Roughly speaking, we prove that for asymptotically large d local algorithms on random d-regular graphs can only produce independent sets, or more generally induced subgraphs with bounded components, having size no more than half the maximum possible value. An analogous statement is proved for sparse Erdős-Rényi graphs having large average degree.

1.1.1 Local algorithms in terms of factors of IID

To describe the motivation behind our work in some detail we need to introduce a precise definition of FIID processes. We specialize to the case where the processes are defined over a regular tree. Later on we explain analogous results for FIID processes that are defined for Erdős-Rényi graphs.

Let $T_d$ denote the rooted d-regular tree with a specified vertex $\circ$ being the root. Let $\chi$ be a finite set. An FIID process on $T_d$ is an invariant random process $\Phi = (\phi(v), v \in V(T_d)) \in \chi^{V(T_d)}$ defined as follows. The automorphism group $\text{Aut}(T_d)$ acts on $[0, 1]^{V(T_d)}$ via $\gamma \cdot \omega(v) = \omega(\gamma^{-1}v)$ where $\gamma \in \text{Aut}(T_d)$ and $\omega \in [0, 1]^{V(T_d)}$. Let $f : [0, 1]^{V(T_d)} \to \chi$ be a measurable function with respect to the Borel $\sigma$-algebra that is also invariant under root-preserving automorphisms (i.e. $f$ is spherically symmetric about $\circ$). A random labelling of $T_d$ is the IID process $X = (X(v), v \in V(T_d)) \in [0, 1]^{V(T_d)}$, where each $X(v) \sim \text{Uniform}[0, 1]$. Since $\text{Aut}(T_d)$ acts transitively on $V(T_d)$, for each $v \in V(T_d)$ let $\gamma_{v \to \circ}$ be an automorphism such that $\gamma_{v \to \circ}v = \circ$. The process $\Phi$ is then defined by

$$\phi(v) = f(\gamma_{v \to \circ} \cdot X).$$

The value of $f$ does not depend on the choice of $\gamma_{v \to \circ}$ since $f$ is invariant under root-preserving automorphisms. The function $f$ is called the factor associated to $\Phi$.

The following is an example on an FIID process that produces independent sets over $T_d$ (an independent set of a graph is a set of vertices that have no edges between them). The factor is $f_{\text{IND}}(\omega) = 1_{\{\omega(\circ) \leq \min\{\omega(v); v \in N_{T_d}(\circ)\}\}}$. In other words, the process selects the set of all local minimums of the random labelling. Observe that this process can also be naturally defined over any finite graph. This is a general feature of FIID processes and is the reason for these being used as a model of local algorithms.

We are interested in FIID percolation processes, namely, $\phi \in \{0, 1\}^{V(T_d)}$. In this case we think of $\phi$ as a random subset of $V(T_d)$ given by the vertices that take the value 1. The density of such a process is $\text{den}(\phi) = \mathbb{P}[\phi(\circ) = 1] = \mathbb{E}[f(X)]$, where $f$ is the factor associated to $\phi$. The density does not depend on the choice of the root due to invariance. A percolation process $\phi$ has
finite clusters if the connected components of the induced subgraph $T_d[\{v \in V(T_d) : \phi(v) = 1\}]$ are finite with probability one. Observe that independent sets have clusters of size one. Define

$$\alpha^\text{IND}(T_d) = \sup\{\text{den}(I) : I \text{ is an FIID independent set}\},$$

$$\alpha^\text{perc}(T_d) = \sup\{\text{den}(\phi) : \phi \text{ is an FIID percolation with finite clusters}\}.$$  

1.1.2 Conjecture of Hatami, Lovász and Szegedy on local algorithms

We now relate the above to random regular graphs and explain the conjecture of Hatami, Lovász and Szegedy. Let $G_{n,d}$ denote a random $d$-regular graph on $n$ vertices. To be precise, we sample $G_{n,d}$ from the well-known configuration model: each of $n$ labelled vertices get $d$ half-edges and these $nd$ half-edges are paired uniformly at random (see [31]). The graph $G_{n,d}$ locally looks like the tree $T_d$ in the following sense. If $\circ_n$ is a uniform random vertex of $G_{n,d}$ then for every $r \geq 0$ the $r$-neighbourhood of $\circ_n$, $N_r(\circ_n, G_{n,d})$, is a tree with probability $1 - o(1)$ as $n \to \infty$. This follows from the fact that for large $n$, the number of cycles in $G_{n,d}$ whose size is of constant order in $n$ follows a Poisson distribution with finite mean (see [31] chapter 9.2).

The local tree-like structure allows us to project FIID percolation processes on $T_d$ down to $G_{n,d}$ with high degree of accuracy. If a percolation process $\phi$ has a factor $f$ that depends only on a finite size neighbourhood of the root, say of radius $r$, then we can copy it to $G_{n,d}$. Simply decorate $G_{n,d}$ with a random labelling $X$ and define $\phi_n(v)$ to be the evaluation of $f$ on the decorated $r$-neighbourhood of $v$ if $N_r(v, G_{n,d})$ is a tree. Otherwise, set $\phi_n(v) = 0$. Under this scheme, the local statistics of $\phi_n$ are close to those of $\phi$, and in particular, $|\{v : \phi_n(v) = 1\}|/n \to \text{den}(\phi)$ in probability as $n \to \infty$. In case that $\phi$ does not have a factor which depends on a finite size neighbourhood of the root, we can approximate $\phi$ in distribution with percolation processes having such factors [34]. Therefore, there is no loss of generality in assuming that all factors depend on a finite size neighbourhood of the root. (We explain the details of this projection scheme in Section 3.1 where it serves as a technical tool in the proofs of our main results.)

The fact that FIID processes can be copied to $G_{n,d}$ means that combinatorial structures sampled by such processes, such as independent sets and percolation with finite clusters, can also be observed in $G_{n,d}$ with high probability. So the natural question is whether $G_{n,d}$ contains isomorphism invariant combinatorial structures that do not come from FIID processes on $T_d$. Hatami, Lovász and Szegedy conjectured that such structures do not exist under natural restrictions. This is be stated for independent sets and percolation in the following manner. Define

$$\alpha^\text{IND}(G) = \max\{|I|/|V(G)| : I \text{ is an independent set of } G\},$$

$$\alpha^\tau(G) = \max\{|S|/|V(G)| : G[S] \text{ has components of size at most } \tau\}.$$  

A beautiful argument by Bayati, Gamarnik and Tetali [6] proved that $\alpha^\text{IND}(G_{n,d})$ converges
almost surely to a non-random limit, say $\alpha^{\text{IND}}(d)$. (By almost surely we mean that for each $n$ a
graph is sampled according to $G_{n,d}$ independently of all other $n$). The same argument also shows
that $\alpha^{\tau}(G_{n,d})$ converges, almost surely, to a non-random quantity $\alpha^{\tau}(d)$ as $n \to \infty$ ($\tau$ is a fixed
integer). Recently, Ding, Sly and Sun [20] have proved an exact, but implicit, formula for $\alpha^{\text{IND}}(d)$
for all sufficiently large $d$. Their proof rigorously verifies some of the predictions from statistical
physics about the nature of random constraint satisfaction problems. The conjecture of Hatami,
Lovász and Szegedy implies that $\alpha^{\text{IND}}(T_d) = \alpha^{\text{IND}}(d)$ and also that $\alpha^{\text{perc}}(T_d) = \sup_{\tau} \alpha^{\tau}(d)$ for
every $d$. The refutation by Gamarnik and Sudan proved that $\alpha^{\text{IND}}(T_d) \leq (\frac{1}{2} + \frac{1}{\sqrt{8}})\alpha^{\text{IND}}(d)$ for
all sufficiently large $d$.

### 1.2 Summary of results

We begin by describing an asymptotically optimal approximation gap for generating independent
sets and percolation with finite clusters via local algorithms on $G_{n,d}$. Part of this represents
joint work with Bálint Virág. This is followed by similarly optimal approximation gaps for
independent sets and percolation with finite clusters using local algorithms on Erdős-Rényi
graphs.

#### 1.2.1 Sub–optimality on random regular graphs

**Theorem 1.2.1.** For independent sets, $\lim_{d \to \infty} \frac{\alpha^{\text{IND}}(T_d)}{\alpha^{\text{IND}}(d)} = 1/2$.

Let us explain how this is optimal. It is known due to an upper bound by Bollobás [10] and
a lower bound by Frieze and Luczak [38] that $\lim_{d \to \infty} \frac{\alpha^{\text{IND}}(d)}{(\log d)/d} = 2$. Also, several construction
exists in the literature (see [23, 33, 40, 41]) to establish the lower bound $\lim_{d \to \infty} \frac{\alpha^{\text{IND}}(T_d)}{(\log d)/d} \geq 1$.
In Section [4] we describe one of these construction that is due to Lauer and Wormald [33]. So to
establish Theorem 1.2.1 we only had to prove that $\limsup_{d \to \infty} \frac{\alpha^{\text{IND}}(T_d)}{(\log d)/d} \leq 1$.

In a related vein to our work Coja-Oghlan and Efthymiou [14] has shown that the density
$(\log d)/d$ is the ‘clustering threshold’ for independent sets in $G_{n,d}$, provided that $d$ is large. In
practice, algorithms that are of a local nature have not been able to find independent sets in
$G_{n,d}$ of density exceeding the clustering threshold of $(\log d)/d$. Theorem 1.2.1 provides some
evidence as to why this is the case.

### A remark on the domination number

A related problem to finding maximum independent
sets is constructing minimum dominating sets. A dominating set $D$ of a graph is a set of vertices
such that any vertex outside $D$ has a neighbour in $D$. Note that a maximal independent set
is a dominating set. Let $\beta(G)$, the domination ratio, denote the size density of the smallest
dominating set in $G$. It is shown in [1] that asymptotically in $d$ the domination ratio $\beta(G_{n,d})$ is
at least $\frac{\log d}{d}$ with high probability as $n \to \infty$. Therefore, any FIID dominating set of $T_d$ must
have asymptotic density at least $\frac{\log d}{d}$. On the other hand, any FIID maximal independent set of
Theorem 1.2.2. Let \( G_{n,d} \) be a sequence of FIID percolations where \( Y_d \in \{0, 1\}^{V(T_d)} \) is a percolation process on \( T_d \). If \( \text{corr}(Y_d) \to 0 \) as \( d \to \infty \) then

\[
\limsup_{d \to \infty} \frac{\text{den}(Y_d)}{(\log d)/d} \leq 1.
\]

Alternatively, if \( \text{avdeg}(Y_d) = o(\log d) \) as \( d \to \infty \) then \( \limsup_{d \to \infty} \frac{\text{den}(Y_d)}{(\log d)/d} \leq 1 \).

The result stated in terms of the average degree follows from the one in terms of the correlation because \( \text{corr}(Y_d) = \text{avdeg}(Y_d)/(d \text{den}(Y_d)) \). Thus, assuming WLOG that \( \text{den}(Y_d) \geq (\log d)/d \), we deduce that if \( \text{avdeg}(Y_d) = o(\log d) \) then \( \text{corr}(Y_d) \to 0 \).

Corollary 1.2.3. Let \( \{Y_d\} \) be a sequence of FIID percolations where \( Y_d \) is a percolation process on \( T_d \). Suppose that the connected components of the subgraph of \( T_d \) induced by the vertices \( \{v \in V(T_d) : Y_d(v) = 1\} \) are finite with probability one. Then

\[
\limsup_{d \to \infty} \frac{\text{den}(Y_d)}{(\log d)/d} \leq 1.
\]

The natural question that arises following Theorem 1.2.2 is whether there is a bound on the density of FIID percolation on \( T_d \) in terms of the correlation. That is, suppose \( \{Y_d\} \) is a sequence of FIID percolation such that \( \text{corr}(Y_d) \leq c \). What can be concluded about \( \text{den}(Y_d) \) in terms of \( c \)?

There is no density bound if \( c = 1 \). Indeed, Bernoulli percolation on \( T_d \) has correlation 1 and density \( p \) for any desired \( p \in [0, 1] \). Also, Lyons [34] constructs, as a weak limit of FIID percolations on \( T_d \), a density 1/2 percolation on \( T_d \) with correlation \( 1 - O(1/\sqrt{d}) \) for every \( d \).
His construction can be utilized to get any desired density. On the other hand if \( c < 1 \) then there is a bound on the density of order \( O((\log d)/d) \).

**Theorem 1.2.4.** Let \( \{Y_d\} \) be a sequence of FIID percolation processes where \( Y_d \) is a percolation process on \( \mathbb{T}_d \). Let \( \Psi(c) = 1 - c + c \log c \) for \( c \geq 0 \) \((\Psi(0) = 1)\). If \( \text{corr}(Y_d) \leq c < 1 \) for every \( d \) or \( \text{corr}(Y_d) \geq c > 1 \) for every \( d \) then

\[
\text{den}(Y_d) \leq \frac{2}{\Psi(c)d} \cdot (\log d - \log \log d + 1 + o(1)) \quad \text{as } d \to \infty.
\]

Also, given \( 0 \leq c < 1 \) there exists a sequence of percolations \( \{Z_d\} \) with \( \text{corr}(Z_d) = c \) and

\[
\text{den}(Z_d) = \frac{1 - o(1)}{\sqrt{1 - c}} \cdot \frac{\log d}{d} \quad \text{as } d \to \infty.
\]

We are not aware of any sequence of FIID percolation processes whose correlations satisfy the inequality \( \text{corr}(Y_d) \geq c > 1 \) for every \( d \). We also expect the bounds in Theorem 1.2.4 to be suboptimal but do not know how to improve upon them in a general setting.

To compare the upper bound on the density of FIID percolation processes on \( \mathbb{T}_d \) having finite clusters with corresponding percolation densities in \( G_{n,d} \), we prove the following.

**Theorem 1.2.5.** Let \( G_{n,d} \) be a random \( d \)-regular graph on \( n \) vertices. Let \( \epsilon_d \) be a sequence of real numbers such that \( 0 < \epsilon_d \leq 1 \) and \( \epsilon_d \to 0 \) as \( d \to \infty \). Let \( \tau = \epsilon_d \log d / d \). Given \( \epsilon > 0 \) there exists a \( d_0 = d_0(\epsilon, \{\epsilon_d\}) \) such that if \( d \geq d_0 \), then with high probability any induced subgraph of \( G_{n,d} \) with components of size at most \( \tau \) has size at most

\[
(2 + \epsilon) \frac{\log d}{d} n.
\]

This means that if \( E_n \) is the event that all induced subgraphs of \( G_{n,d} \) with components of size at most \( \tau \) have size at most \((2 + \epsilon) \log d / d \) \( n \), then for \( d \geq d_0 \), we have that \( P[E_n] \to 1 \) as \( n \to \infty \).

An important corollary of Theorem 1.2.5 is the following. Recall, for a fixed integer \( \tau \geq 1 \) the quantity \( \alpha^\tau(d) \), which is the (almost sure) limiting value of the maximum size density of subsets in \( G_{n,d} \) whose induced subgraphs have components of size at most \( \tau \).

**Corollary 1.2.6.** We have that

\[
\limsup_{d \to \infty} \sup_{\tau \geq 1} \frac{\alpha^\tau(d)}{(\log d)/d} \leq 2.
\]

Although it is not immediate from Theorem 1.2.5, a careful adaptation of the proof of Theorem 1.2.5 towards Corollary 1.2.6 shows that

\[
\sup_{\tau \geq 1} \alpha^\tau(d) \leq \frac{2(\log d + 2 - \log 2)}{d} \quad \text{for } d \geq 5.
\]
As we already have that $\alpha^{\text{IND}}(d) = (2 + o(1)) \frac{\log d}{d}$, it follows from the above that $\sup_{\tau \geq 0} \alpha^{\tau}(d) = (2 + o(1)) \frac{\log d}{d}$ as $d \to \infty$. Hence,

$$\lim_{d \to \infty} \frac{\alpha^{\text{perc}}(T_d)}{\sup_{\tau \geq 0} \alpha^{\tau}(d)} = \frac{1}{2}.$$

### 1.2.2 FIID process on the edges of $T_d$

The natural analogue to FIID process defined over the vertices of $T_d$, that is, taking values in $\chi^{V(T_d)}$, are FIID processes over the edges of $T_d$. These are defined analogous to vertex-indexed processes. Since $\text{Aut}(T_d)$ acts transitively on the edges of $T_d$ the formal definition for vertex-indexed processes carries over naturally. Such a process $\phi \in \chi^{E(T_d)}$ is determined by a measurable function $F : [0, 1]^E(T_d) \to \chi^{E(T_d)}$ satisfying $F(\gamma \cdot \omega) = \gamma \cdot F(\omega)$ for all $\gamma \in \text{Aut}(T_d)$. Then $\phi = F(X)$ where $X$ is a random labelling of the edges of $T_d$.

Consider an FIID edge-indexed percolation process $\phi \in \{0, 1\}^E(T_d)$. Its density is $\text{den}(\phi) = \mathbb{P}[\phi(e) = 1]$, where $e = \{0, 0'\}$ is an arbitrary edge. The clusters of $\phi$ are the connected components of the subgraph induced by the open edges $\{e \in E(T_d) : \phi(e) = 1\}$. Such processes are called FIID bond percolation processes.

The prototypical example is Bernoulli bond percolation: every edge is open independently with probability $p$, where $0 \leq p \leq 1$ is the density. In this case the clusters are finite if and only if $p \leq \frac{1}{d} - 1$ (see [36] Theorem 5.15). It is natural to ask about the largest possible density of a FIID bond percolation process on $T_d$ with finite clusters.

**Theorem 1.2.7.** The maximum density of an invariant bond percolation on $T_d$ with finite clusters is $\frac{2}{d}$ for every $d \geq 2$.

This bound is not achieved by any invariant bond percolation process with finite clusters. However, there is an FIID bond percolation process of density $\frac{2}{d}$ which is a weak limit of FIID bond percolation processes with finite clusters. This process is a 2-factor: every vertex of $T_d$ is incident to two open edges.

2-factors are not the only possible such limit points in the class of invariant bond percolation processes. The wired spanning forest of $T_d$ has density $\frac{2}{d}$ and every cluster has one topological end.

We also prove the following result about edge orientations of $T_d$. It is a curious result that may be of independent interest, and at first sight may seem unexpected.

Given any edge orientation of a graph $G$, a vertex $v$ is called a source if all edges that are incident to $v$ are oriented away from it. If all incident edges are oriented towards $v$, then $v$ is called a sink. For $d \geq 3$, we construct an FIID edge orientation of $T_d$ with no sources or sinks. Such an orientation cannot exist for the bi-infinite path, that is, for $T_2$. Indeed, there are only two edge orientations of the path with no sources or sinks: either all edges point to the right or all point to the left. However, the uniform distribution on these two orientations cannot be an FIID process because it is not mixing.
Theorem 1.2.8. For every $d \geq 3$ there exists an FIID orientation of the edges of $\mathbb{T}_d$ that contains no sources or sinks.

1.2.3 Sub-optimality on Erdős–Rényi graphs

The Erdős–Rényi graph ER$(n, p)$ is a random graph on the vertex set $[n]$ where every pair of vertices $\{i, j\}$ is independently included with probability $p$. Our interest lies with the random graphs ER$(n, \lambda/n)$ where $\lambda > 0$ is fixed.

Local algorithms on sparse Erdős–Rényi graphs are projections of FIID processes on Poisson-Galton-Watson (PGW) trees. We define the appropriate notion of FIID processes on PGW trees. A PGW tree with average degree $\lambda$ contains no sources or sinks.

Before we can define the notion of FIID processes on PGW trees we will need some notation. Let $\Lambda_r$ 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 $\text{dist}(u, v) \leq r$ where $\text{dist}$ denotes the graph distance, and (3) $x \in [0, 1]^{V(H)}$ is a labelling of $H$. $\Lambda_r$ has a natural $\sigma$-algebra, $\Sigma_r$, 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 $\gamma : (H, v) \rightarrow (H', v')$, any labelling $x$ of $(H, v)$ induces a labelling $\gamma \cdot x$ of $(H', v')$ by defining $\gamma \cdot x(i) = x(\gamma^{-1}(i))$, and vice-versa. Given a finite set $\chi$, a function $f : \Lambda_r \rightarrow \chi$ is a factor if it is $\Sigma_r$ measurable and $f(H, v, x) = f(\gamma(H), \gamma(v), \gamma \cdot x)$ for all isomorphisms $\gamma$ of $H$, and all $H$.

For $0 \leq r < \infty$, let $f : \Lambda_r \rightarrow \chi$ be a factor. Consider a PGW$_\lambda$ tree with a random labelling $X$. Let $N_r(\text{PGW}_\lambda, v)$ denote the $r$-neighbourhood of a vertex $v$ in PGW$_\lambda$ and let $X(\text{PGW}_\lambda, v, r)$ be the restriction of $X$ to $N_r(\text{PGW}_\lambda, v)$. Define a $\chi$-valued process $\phi$ on the vertices of PGW$_\lambda$ by setting

$$\phi(v) = f(N_r(\text{PGW}_\lambda, v), v, X(\text{PGW}_\lambda, v, r)).$$

We say $\phi$ is a percolation with finite clusters if $\chi = \{0, 1\}$ and the connected components of the subgraph of PGW$_\lambda$ induced by $\{v \in V(\text{PGW}_\lambda) : \phi(v) = 1\}$ are finite with probability 1 (w.r.t. the random labelled tree $(\text{PGW}_\lambda, X)$).

The distribution of the random variable $\phi(v)$ does not depend on the choice of the vertex $v$. This is because in a PGW tree the distribution of the neighbourhoods $N_r(\text{PGW}_\lambda, v)$ does not
depend on the choice of $v$. So let $PGW(\lambda, r)$ 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 an FIID percolation process $Y$ on $PGW_{\lambda}$ is

$$\text{den}(Y) = E[f(PGW(\lambda, r, \circ, X)] .$$

Define the quantity $\alpha_{\text{perc}}(PGW_{\lambda})$ by

$$\alpha_{\text{perc}}(PGW_{\lambda}) = \sup_{0 \leq r < \infty} \left\{ \text{den}(Y) : Y \text{ is a percolation process on } PGW_{\lambda} \text{ with finite clusters and factor } f : \Lambda_r \rightarrow \{0,1\} \right\} .$$

**Theorem 1.2.9.** The limit $\lim_{\lambda \to \infty} \frac{\alpha_{\text{perc}}(PGW_{\lambda})}{(\log \lambda) / \lambda} = 1 .

This result is interesting for two reasons. First, the lower bound $\liminf_{\lambda \to \infty} \frac{\alpha_{\text{perc}}(PGW_{\lambda})}{(\log \lambda) / \lambda} \geq 1$ is proved via an indirect argument, whereby we show the existence of local independent sets in $ER(n, \lambda/n)$ of density $(\log \lambda) / \lambda$ from the existence of such independent sets in $G_{n,d}$. Second, the proof of the upper bound $\limsup_{\lambda \to \infty} \frac{\alpha_{\text{perc}}(PGW_{\lambda})}{(\log \lambda) / \lambda} \leq 1$ is more complex. The difficulty arises because the underlying local structure is random. We show that having randomness in the local structure, i.e. the PGW$_{\lambda}$ tree, does not provide local algorithms with extra power.

Finally, we compare the largest density of percolation sets with finite clusters in $ER(n, \lambda/n)$ given by local algorithms, that is, $\alpha_{\text{perc}}(PGW_{\lambda})$, to arbitrary such percolation sets in $ER(n, \lambda/n)$. Recall the quantity $\alpha^\tau(G)$ for a finite graph $G$. The proof technique of Bayati, Gamarnik and Tetali [6] can be utilized in a straightforward manner to show that $\alpha^\tau(ER(n, \lambda/n))$ converges almost surely to a non-random limit $\alpha^\tau(ER(\lambda))$ for every fixed integer $\tau \geq 1$. The following theorem implies that

$$\sup_{\tau \geq 1} \alpha^\tau(ER(\lambda)) \leq (2 + o(1)) \frac{\log \lambda}{\lambda} \text{ as } \lambda \to \infty .$$

**Theorem 1.2.10.** Suppose $\lambda \geq 5$ and set $\tau = \log \lambda(n) - \log \log \log (n) - \log (\omega_n)$ where $\omega_n \to \infty$ with $n$. With high probability any induced subgraph of $ER$ with components of size at most $\tau$ has size at most

$$\frac{2}{\lambda} \left( \log \lambda + 2 - \log 2 \right) n .$$

The lower bound on independent sets provided by Frieze [37] implies that $\alpha^1(ER(\lambda)) \geq (2 + o(1)) \frac{\log \lambda}{\lambda}$. So $\sup_{\tau \geq 1} \alpha^\tau(ER(\lambda)) = (2 + o(1)) \frac{\log \lambda}{\lambda} \text{ as } \lambda \to \infty$. Together with Theorem 1.2.9 this provides the following sub–optimality result for local algorithms on Erdős-Rényi graphs.

$$\lim_{\lambda \to \infty} \frac{\alpha_{\text{perc}}(PGW_{\lambda})}{\sup_{\tau \geq 1} \alpha^\tau(ER(\lambda))} = \frac{1}{2} .$$
1.3 Organization of the thesis

We begin with the proofs of Theorems 1.2.5 and 1.2.10 in Chapter 2. This is a good place to start as the proofs are elementary, and at the same time, they illustrate some of the basic combinatorial techniques used throughout the thesis.

In Chapter 3 we introduce the important relations between FIID processes on $T_d$ and their projection on $G_{n,d}$. We use these relations and combinatorial arguments involving $G_{n,d}$ to deduce a fundamentally important ‘entropy inequality’ that is satisfied by FIID processes on $T_d$. The application of this entropy inequality is vital in the proofs of our main results.

In Chapter 4 we prove Theorems 1.2.2 and 1.2.4, and explain how Corollary 1.2.3 follows. As Theorem 1.2.1 is a special case of Theorem 1.2.2 we need not supply a separate proof. We then discuss FIID percolation processes with finite clusters on $T_d$ for small values of $d$, and heuristically explain a result of Csóka [15] which shows that $\alpha_{\text{perc}}(3) = 3/4$. We conclude Chapter 4 with the proofs of Theorem 1.2.7 and Theorem 1.2.8 in Section 4.4.

Finally, in Chapter 5 we consider FIID processes on Poisson-Galton-Watson trees, which yield local algorithms on Erdős-Rényi graphs. There we prove Theorem 1.2.9.
Chapter 2

Percolation with small clusters in random graphs

A subset $S$ of a graph $G$ is a percolation set with clusters of size at most $\tau$ if all the components of the induced subgraph $G[S]$ have size at most $\tau$. We say $S$ is a percolation set with small clusters when we do not want to mention the parameter $\tau$ explicitly.

Recall that for an integer $\tau \geq 1$ and finite graph $G$ we defined

$$\alpha^{\tau}(G) = \max\left\{ \frac{|S|}{|V(G)|} : S \subset V(G) \text{ is a percolation set with clusters of size at most } \tau \right\}.$$

In this section we prove Theorem 1.2.5 and Theorem 1.2.10.

2.1 A brief history of the problem

Edwards and Farr [21] studied the problem of determining the size of the largest percolation sets in some general deterministic classes of graphs. They termed this notion ‘graph fragmentability’. They consider a natural $\tau \to \infty$ version of the problem. Namely, for a sequence of graphs $\{G_n\}$ with $|G_n| \to \infty$ they consider the quantity

$$\sup_{\tau \geq 1} \inf_n \alpha^{\tau}(G_n).$$

They provide upper and lower bounds on the above quantity for bounded degree graphs. Their bound is sharp for the family of graphs with maximum degree 3, and optimal, in a sense, for several families of graphs such as trees, planar graphs or graphs with a fixed excluded minor. However, their bounds are not of the correct order of magnitude for random $d$-regular graphs.

For random graphs, a lower bound on the density of percolation sets with small clusters can be deduced from just considering the largest independent sets, that is, the $\tau = 1$ case. As we mentioned in the introduction, Bollobás [10] proved that with high probability the density of the largest independent sets in a random $d$-regular graph is at most $2(\log d)/d$ as $d \to \infty$. The
same bound was proved for Erdős-Rényi graphs of average degree \( d \) by several authors (see Theorem 11.25). Frieze and Luczak then provided matching lower bounds by using a non-constructive argument.

Problems closely linked to percolation sets with small clusters, such as maximum induced forests and maximum \( k \)-independent sets, have been well studied in the literature. Bau et al. provide bounds on the density of maximum induced forests in random regular graphs. Hoppen and Wormald discuss algorithms and lower bounds for maximum induced forests and maximum \( k \)-independent in random regular graphs. (See the references within these papers for further discussion on these and related problems.) The techniques of this paper can be used to prove upper bounds on the maximum density of induced forests and \( k \)-independent sets in random \( d \)-regular graphs for large \( d \), as we explain in Section 2.3.3.

### 2.2 Preliminaries and terminology

Let \( V(G) \) and \( E(G) \) denote the set of vertices and edges of a graph \( G \), respectively.

We say that a sequence of events \( E_n \), generally associated to \( \mathcal{G}_{n,d} \), occurs with high probability if \( \Pr[E_n] \to 1 \) as \( n \to \infty \).

We use the configuration model (see chapter 2.4) as the probabilistic method to sample a random \( d \)-regular graph \( \mathcal{G}_{n,d} \) on \( n \) labelled vertices. The graph \( \mathcal{G}_{n,d} \) is sampled in the following manner. Each of the \( n \) distinct vertices emit \( d \) distinct half-edges, and we pair up these \( nd \) half-edges uniformly at random. (We assume that \( nd \) is even.) These \( nd/2 \) pairs of half-edges can be glued into full edges to yield a random \( d \)-regular graph. There are \((nd)!! = (nd-1)(nd-3)\cdots3\cdot1\) such graphs.

The resulting random graph may have loops and multiple edges, that is, it is a multigraph. However, the probability that \( \mathcal{G}_{n,d} \) is a simple graph is uniformly bounded away from zero as \( n \to \infty \). In fact, Bender and Canfield and Bollobás showed that as \( n \to \infty \)

\[
\Pr[\mathcal{G}_{n,d} \text{ is simple}] \to e^{1-d^2/4}.
\]

Also, conditioned on \( \mathcal{G}_{n,d} \) being simple its distribution is a uniformly chosen \( d \)-regular simple graph on \( n \) labelled vertices. It follows from these observations that any sequence of events that occur with high probability for \( \mathcal{G}_{n,d} \) also occur with high probability for a uniformly chosen simple \( d \)-regular graph.

We denote by \( \text{ER}(n,p) \) the Erdős-Rényi random graph on \( n \) vertices where each pair of distinct vertices \( \{i,j\} \) is included as an edge independently with probability \( p \). We are interested in the sparse regime: \( p = \lambda/n \) where \( \lambda \) is constant.

Throughout the rest of the thesis we set the function \( h(x) = -x \log(x) \) for \( 0 \leq x \leq 1 \) with the convention that \( h(0) = 0 \). We are going to use the following properties of \( h(x) \) throughout.
The proof of Theorem 1.2.5 is based on two main lemmas. We state these lemmas and then prove them. The lemmas are then proved in Section 2.3.1 and 2.3.2 respectively.

A finite (multi)-graph $H$ is $k$-sparse if $|E(H)|/|V(H)| \leq k$, that is, the average degree of $H$ is at most $2k$. For example, finite trees are 1-sparse. Any subgraph of a $d$-regular graph is $d/2$-sparse. The first lemma shows that linear sized subgraphs of a random $d$-regular graph are likely to be $k$-sparse so long as their size density is sufficiently small.

**Lemma 2.3.1.** Let $G_{n,d}$ be a random $d$-regular graph on $n$ vertices. Suppose $d \geq 12$ and $3.5 < k \leq (1 - \frac{1}{\sqrt{2}})d$. Set $C_{k,d} = e^{-4(2k/d)\frac{1}{1+k}}$. With high probability, any subgraph in $G_{n,d}$ of size at most $C_{k,d} \cdot n$ is $k$-sparse. The probability that this property fails in $G_{n,d}$ is $O_d(n^{3.5-k})$.

The next lemma states that induced $k$-sparse subgraphs of $G_{n,d}$ are actually not very large if $k = o(\log d)$ and $d$ is sufficiently large.

**Lemma 2.3.2.** Let $G_{n,d}$ be a random $d$-regular graph on $n$ vertices. Let $k = \epsilon_d \log d$ where $0 < \epsilon_d \leq 1$ and $\epsilon_d \to 0$ as $d \to \infty$. Given any $\epsilon > 0$ there is a $d_0 = d_0(\epsilon, \{\epsilon_d\})$ such that if $d \geq d_0$, then with high probability any $k$-sparse induced subgraph of $G_{n,d}$ has size at most

$$\left(2 + \epsilon\right)\frac{\log d}{d} n.$$

We do not attempt to provide bounds on $d_0$.

**Proof of Theorem 1.2.5** Let $\epsilon_d$ be as in the statement of the theorem. First we show that it is possible to choose $k \geq 4$ satisfying both the constraints that $k = o(\log d)$ and $e^{-4(2k/d)\frac{1}{1+k-1}} \geq \epsilon_d \frac{\log d}{d}$ for all large $d$. Let $\epsilon'_d = \max\{\frac{4}{\log(\epsilon_d)}, \frac{4}{\log d}\}$. Note that $\epsilon'_d \to 0$ as $d \to \infty$. We assume that $d$ is large enough that $\epsilon_d \leq e^{-6}$. Set $k = \epsilon'_d \log d$. 

\[\begin{align*}
(1) & \quad h(xy) = xh(y) + yh(x). \\
(2) & \quad h(1-x) \geq x - x^2/2 - x^3/2 \text{ for } 0 \leq x \leq 1. \\
(3) & \quad h(1-x) \leq x - x^2/2 \text{ for } 0 \leq x \leq 1.
\end{align*}\]
We begin by showing that $e^{-4}(2k/d)^{1+1/(k-1)} \geq \epsilon_d \frac{\log d}{d}$ for all large $d$. As $(2k/d) \leq 1$ we have

$$
\left(\frac{2k}{d}\right)^{1+1/(k-1)} \geq \left(\frac{2k}{d}\right)^{1+2/k} \geq \left(\frac{k}{d}\right)^{1+2/k} \geq \left(\frac{\epsilon_d \log d}{d}\right)^{1+\frac{2}{\epsilon_d \log d}}.
$$

We now show that $\left(\frac{\epsilon_d \log d}{d}\right)^{\frac{2}{\epsilon_d \log d}} \geq \epsilon_d^{1/2}$, which would imply that the very last term above is greater than $\left(\epsilon_d^{1/2}\right) \log d$. Note that

$$
\log \left(\frac{\epsilon_d^{1/2} \log d}{d}\right)^{\frac{2}{\epsilon_d \log d}} = \frac{2}{\epsilon_d} \left(\frac{\log(\epsilon_d^{1/2}) + \log \log d}{\log d} - 1\right)
$$

$$
\geq \frac{2}{\epsilon_d} \left(\frac{\log 4}{\log d} - 1\right) \quad \text{(as } \epsilon_d \geq \frac{4}{\log d})
$$

$$
\geq \frac{-2}{\epsilon_d}.
$$

We conclude that $\left(\frac{\epsilon_d^{1/2} \log d}{d}\right)^{\frac{2}{\epsilon_d \log d}} \geq e^{-2/\epsilon_d}$, and as $\epsilon_d \geq \frac{4}{\log(\epsilon_d)}$, we deduce that $e^{-2/\epsilon_d} \geq e^{-\frac{1}{2} \log(\epsilon_d)} = \epsilon_d^{1/2}$.

So far we have seen that $(2k/d)^{1+1/(k-1)} \geq (\epsilon_d^{1/2}) \frac{\log d}{d}$ for all large $d$ (large $d$ is required to ensure that $\epsilon_d \leq 1$). Now we show that $e^{-4}\epsilon_d^{1/2} \geq \epsilon_d$ for larger $d$. As $\epsilon_d \geq \frac{4}{\log(\epsilon_d)}$, the inequality $e^{-4}\epsilon_d^{1/2} \geq \epsilon_d$ holds if $|\log(\epsilon_d)| \leq 4e^{-4}\epsilon_d^{-1/2}$. Since $\epsilon_d \leq 1$, this inequality is the same as $\log(\epsilon_d^{-1}) \leq 4e^{-4}\epsilon_d^{-1/2}$. A simple calculation shows that $\log(x) \leq 4e^{-4}x^{1/2}$ if $x \geq 8/4$. Therefore, $e^{-4}\epsilon_d^{1/2} \geq \epsilon_d$ whenever $\epsilon_d \leq 4e^{-8}$. This certainly holds for all large $d$.

We have thus concluded that it is possible to choose $k \geq 4$ satisfying both the constraints that $k = o(\log d)$ and $e^{-4}(2k/d)^{1+1/(k-1)} \geq \epsilon_d \frac{\log d}{d}$ for all large $d$. We are now able to finish the proof. Set $k = \epsilon_d^{-1/2} \log d$ in the following.

Let $A = A(\mathcal{G}_{n,d})$ be the event that all subgraphs of $\mathcal{G}_{n,d}$ containing at most $\epsilon_d \frac{\log d}{d} n$ vertices are $k$-sparse. From the conclusion derived above we see that there exists $d_1$ such that if $d \geq d_1$ then $k/d < 1 - 1/\sqrt{2}$ and $e^{-4}(2k/d)^{1+1/(k-1)} \geq \epsilon_d \frac{\log d}{d}$. Lemma 2.3.1 implies that $\mathbb{P}[A] \to 1$ as $n \to \infty$.

Let $B = B(\mathcal{G}_{n,d})$ be the event that any induced subgraph of $\mathcal{G}_{n,d}$ that is $k$-sparse contains at most $(2 + \epsilon) \frac{\log d}{d} n$ vertices. From Lemma 2.3.2 we conclude that there exists a $d_2$ such that if $d \geq d_2$ then $\mathbb{P}[B] \to 1$ as $n \to \infty$.

If $d \geq \max\{d_1, d_2\}$ then $\mathbb{P}[A \cap B] \to 1$ as $n \to \infty$. Indeed, $\mathbb{P}[A \cap B] \geq \mathbb{P}[A] + \mathbb{P}[B] - 1$. Let $D = D(\mathcal{G}_{n,d})$ be the event that all induced subgraphs of $\mathcal{G}_{n,d}$ with components of size at most $\tau = \epsilon_d \frac{\log d}{d} n$ have size at most $(2 + \epsilon) \frac{\log d}{d} n$. We show that $A \cap B \subset D$ for all $d \geq \max\{d_1, d_2\}$.

Suppose a $d$-regular graph $G$ on $n$ vertices satisfies properties $A$ and $B$. If $S \subset V(G)$ induces a subgraph with components of size at most $\tau = \epsilon_d \frac{\log d}{d} n$ then all components of $S$ are $k$-sparse because $G$ satisfies property $A$. Hence, $S$ itself induces a $k$-sparse subgraph. As $G$ also satisfies property $B$ we deduce that $S$ contains at most $(2 + \epsilon) \frac{\log d}{d} n$ vertices. This means that $G$ satisfies property $D$, as required.
The proof of Theorem 1.2.5 is now complete because if \( d \geq \{d_1, d_2\} \) then \( \mathbb{P}[D] \geq \mathbb{P}[A \cap B] \to 1 \) as \( n \to \infty \).

In Section 2.3.1 we prove Lemma 2.3.1 and in Section 2.3.2 we prove Lemma 2.3.2.

### 2.3.1 Proof of Lemma 2.3.1

We prove Lemma 2.3.1 by showing that the expected number subgraphs of \( G_{n,d} \) that are of size at most \( C_{k,d} \cdot n \) and that are not \( k \)-sparse is vanishingly small as \( n \to \infty \). The first moment bound implies that the probability is vanishingly small as well.

Let \( Z_{i,j} = Z_{i,j}(G_{n,d}) \) be the number of subsets \( S \subset V(G_{n,d}) \) such that \( |S| = i \) and \( e(S) = j \).

Notice that \( Z_{i,j} = 0 \) unless \( j \leq (d/2)i \).

Let \( N \) be the number of subgraphs of \( G_{n,d} \) that have size at most \( C_{k,d} \cdot n \) and that are not \( k \)-sparse. We have

\[
N = \sum_{i=1}^{C_{k,d}n} \sum_{j=ki}^{(d/2)i} Z_{i,j}.
\]

(2.2)

In the following sequence of lemmas we compute \( \mathbb{E}[Z_{i,j}] \) in order to bound \( \mathbb{E}[N] \).

**Lemma 2.3.3.** For \( 1 \leq i \leq n \) and \( 0 \leq j \leq (d/2)i \), the expectation of \( Z_{i,j} \) is

\[
\mathbb{E}[Z_{i,j}] = \binom{n}{i} \times \frac{(id)! ((n-i)d)! (nd/2)! 2^{id-2j}}{(id-2j)! j! (\frac{nd}{2} - id + j)! (nd)!}.
\]

(2.3)

Proof. For a fixed subset \( S \subset V(G_{n,d}) \) the number of pairings in the configuration model satisfying \( |S| = i \) and \( e(S) = j \) is

\[
\binom{id}{id-2j} \binom{(n-i)d}{id-2j} (id-2j)! (2j-1)!! \left( (n-2i)d + 2j - 1 \right)!!.
\]

The product of binomial coefficients counts the number of ways to choose \( id - 2j \) half edges from \( S \) and \( S^c \) that are to be paired by against each other. The factorial \( (id-2j)! \) is the number of ways to pair up these half-edges. The double factorial \( (2j-1)!! \) counts the ways to pair up the remaining half-edges of \( S \) with themselves, and analogously \( ((n-2i)d + 2j - 1)!! \) counts it for \( S^c \).

As each pairing occurs with probability \( 1/(nd-1)!! \), the probability that a fixed \( S \subset V(G_{n,d}) \) of size \( |S| = i \) satisfies \( e(S) = j \) is

\[
\left( \frac{id}{id-2j} \right) \left( \frac{(n-i)d}{id-2j} \right) (id-2j)! (2j-1)!! \left( (n-2i)d + 2j - 1 \right)!! \times \frac{1}{(nd-1)!!}.
\]

(2.3)

As \( (m-1)!! = \frac{m!}{2^m m!(m/2)!} \) for even integers \( m \geq 2 \), and \( 0!! = 1 \), we may simplify (2.3). Also, there are \( \binom{n}{i} \) subsets \( S \) of size \( i \) and \( \mathbb{E}[Z_{i,j}] \) is the sum over each such \( S \) of the probability that \( e(S) = j \). Therefore,
\[ \mathbb{E}[Z_{i,j}] = \binom{n}{i} \times \frac{(id)! \cdot ((n-i)d)! \cdot (nd/2)! \cdot 2^{id-2j}}{(id-2j)! \cdot j! \cdot \left(\frac{(n-2i)d}{2} + j\right)! \cdot (nd)!}. \]  

(2.4)

Lemma 2.3.4. Suppose \(1 \leq k \leq d/2\) and \(1 \leq i \leq (2k/d)n\). For \(ki \leq j \leq (d/2)i\), \(\mathbb{E}[Z_{i,j}]\) is maximized at \(j = ki\).

Proof. From the equation for \(\mathbb{E}[Z_{i,j}]\) in (2.4) we deduce that the ratio

\[ \frac{\mathbb{E}[Z_{i,j+1}]}{\mathbb{E}[Z_{i,j}]} = \frac{(id-2j-1)(id-2j)}{4(j+1)\left(\frac{(n-2i)d}{2} + j + 1\right)}. \]

If \(i \leq (2k/d)n\) then this ratio is at most 1 provided that \(ki \leq j \leq (id)/2\). Indeed, subtracting the denominator from the numerator gives

\[ \frac{(id-2j-1)(id-2j)}{(n-2i)d + j + 1}. \]

If \(i \leq (2k/d)n\) then this ratio is at most 1 provided that \(ki \leq j \leq (id)/2\) if and only if

\[ ki \geq \frac{1}{2}(id)(id-1) - (n-2i)d - 2 \]

\[ \frac{1}{nd + 3}. \]  

(2.5)

In order to show that (2.5) holds for \(1 \leq i \leq (2k/d)n\) it suffices to show that \(ki \geq \frac{(id)^2}{2nd}\) because the latter term is larger than the right hand side of inequality (2.5). Since \(i \geq 1\), \(ki \geq \frac{(id)^2}{2nd}\) if and only if \(k \geq \frac{id}{2n}\), which is indeed assumed.

It follows from Lemma 2.3.4 and (2.2) that

\[ \mathbb{E}[N] \leq \sum_{i=1}^{C_kdn} \frac{(id/2)\mathbb{E}[Z_{i,ki}]}{\mathbb{E}[Z_{i,j}]} \leq dn^2 \max_{1 \leq i \leq C_kdn} \mathbb{E}[Z_{i,ki}]. \]  

(2.6)

To get a bound on \(\mathbb{E}[Z_{i,ki}]\) which is suitable for asymptotic analysis we first introduce some notation. For a graph \(G\) and subsets \(S, T \subset V(G)\) let

\[ m(S, T) = \frac{|\{(u, v) : u \in S, v \in T, \{u, v\} \in E(G)\}|}{2|E(G)|}. \]

The **edge profile** of \(S\) associated to \(G\) is the \(2 \times 2\) matrix

\[ M(S) = \begin{bmatrix} m(S, S) & m(S, S^c) \\ m(S^c, S) & m(S^c, S^c) \end{bmatrix} \]

where \(S^c = S \setminus V(G)\). If \(|S| = i\) and \(e(S) = j\) then

\[ M(S) = \begin{bmatrix} \frac{2j}{nd} & i - \frac{2j}{nd} \\ \frac{i}{n} - \frac{2j}{nd} & 1 - \frac{i}{n} + \frac{2j}{nd} \end{bmatrix} \]
We denote the matrix above by \( M(i/n, j/(nd)) \) – it is the edge-profile of any subset \( S \) with \( |S| = i \) and \( e(S) = j \). In particular, \( Z_{i,j} \) is the number of \( S \subset V(\mathcal{G}_{n,d}) \) such that \( M(S) = M(i/n, j/(nd)) \). The entropy of a finitely supported probability distribution \( \pi \) is

\[
H(\pi) = \sum_{x \in \text{support}(\pi)} -\pi(x) \log \pi(x).
\]

**Lemma 2.3.5.** For \( 1 \leq i \leq n - 1 \) and \( 0 \leq j \leq id/2 \), we have that

\[
\mathbb{E}[Z_{i,j}] \leq A \times \exp \left\{ n \left[ \frac{d}{2} H(M(i/n, j/(nd))) - (d - 1)H(i/n, 1 - (i/n)) \right] \right\}
\]

where \( A = O(d\sqrt{n}) \) for a universal big \( O \) constant.

**Proof.** We use Stirling’s approximation of \( m! \) to simplify (2.4):

\[
1 \leq \frac{m!}{\sqrt{2\pi m(m/e)^m}} \leq e^{1/12m}.
\]

First we consider \( \binom{n}{m} \). For an integer \( 1 \leq i \leq n - 1 \), Stirling’s approximation shows that \( \binom{n}{i} \leq O(1)\sqrt{n/i(n-i)}e^{nH(i/n, 1-i/n)} \). But \( n/i(n-i) \leq n/(n-1) \leq 2 \) for \( n \geq 2 \). As \( H(0, 1) = H(1, 0) = 0 \), we conclude that \( \binom{n}{m} \leq O(1) \cdot e^{nH(1-\alpha)} \).

Now we consider the term in (2.3). Stirling’s approximation implies that the polynomial order term (in \( n \)) of (2.3) is bounded from above, up to an universal multiplicative constant, by

\[
\left[ \frac{d(nd/2)}{(id-2j)j((nd/2)-id+j)} \right]^{1/2}.
\]

We may assume that each of the terms \( id - 2j, j, \text{ and } (nd/2) - id + j \) are positive integers, for if one of these were zero then the corresponding factorial in \( d \) would be 1, and we could ignore that term from the calculation. Thus, \( (id - 2j)j((nd/2) - id + j) \geq 1 \), which implies that (2.7) is bounded above by \( d\sqrt{n} \).

To deal with terms of exponential order in (2.3) we may substitute \( (m/e)m \) for every \( m! \), and after algebraic simplifications we have

\[
\frac{(id)^i (n-i)^d (nd)^{nd/2}}{(id-2j)^{id-2j} (2j)^i} \left[ (n-2i)d + 2j \right]^{((nd/2)-id+j)} (nd)^{nd}.
\]

Simplifying this further to cancel out powers of \( nd \) results in

\[
\left[ \frac{(i/n)(i/n)(1 - (i/n))(1 - (i/n))}{((i/n) - 2j/nd)^{(i/n) - (2j/nd)^{j/nd} (2j/nd)^{j/nd} (1 - 2(i/n) + 2j/nd)^{1/2 - (i/n) + j/nd} \right]}^{nd}
\]

\[
= \exp \left\{ n \left[ \frac{d}{2} H(M(i/n, j/(nd))) - dH(i/n, 1 - (i/n)) \right] \right\}.
\]
Therefore, (2.4) is bounded from above by
\[ O(d\sqrt{n}) \exp \left\{ n \left[ \frac{d}{2} H\left( M(i/n, j/(nd)) \right) - (d - 1)H(i/n, 1 - (i/n)) \right] \right\}. \]

Since we want to bound \( \mathbb{E} [Z_{i,ki}] \), we proceed with an analysis of the maximum of \( (d/2)H(M(i/n, ki/nd)) - (d - 1)H(i/n, 1 - (i/n)) \) over the range \( 1 \leq i \leq C_{k,d} \cdot n \). Lemma [2.3.5] implies that \( \mathbb{E} [Z_{i,ki}] \) is bounded from above by
\[ O(d\sqrt{n}) \times \exp \{ n[(d/2)H(M(i/n, ki/nd)) - (d - 1)H(i/n, 1 - (i/n))] \}. \]

It is convenient to work with the analytic continuation of the terms involving the entropy. Recall that \( h(x) = -x \log x \). If we set \( \alpha = i/n \) then \( (d/2)H(M(i/n, ki/nd)) - (d - 1)H(i/n, 1 - (i/n)) \) equals
\[ (d/2)[h(\alpha(2k/d)) + 2h(\alpha - \alpha(2k/d)) + h(1 - 2\alpha + \alpha(2k/d))] - (d - 1)H(\alpha, 1 - \alpha). \quad (2.8) \]

Here \( \alpha \) lies in the range \( 1/n \leq \alpha \leq C_{k,d} \). We will show that (2.8) is decreasing in \( \alpha \) if \( 0 \leq \alpha \leq C_{k,d} \). We will then evaluate its value at \( \alpha = 1/n \) to show that the leading term (in \( n \)) is \( (1 - k)(\log n)/n \). This will allow us to conclude Lemma [2.3.1].

**Lemma 2.3.6.** Suppose that \( 2 \leq k \leq (1 - 1/\sqrt{2})d \). Then the entropy term in (2.8) is decreasing as a function of \( \alpha \) for \( 0 \leq \alpha \leq C_{k,d} \).

**Proof.** We will differentiate (2.8) and show that it is negative for \( 0 < \alpha < C_{k,d} \). Notice that the derivative \( h'(\alpha) = -1 - \log(\alpha) \). Differentiating (2.8) in \( \alpha \) and simplifying gives
\[ \frac{d}{2} \left( h\left( \frac{2k}{d} \right) + 2h\left( 1 - \frac{2k}{d} \right) \right) + (k - 1) \log(\alpha) + (d - 1)(-\log(1 - \alpha)) - (d - k)(-\log(1 - 2\alpha + \frac{2k}{d}\alpha)). \]

First, we deal with the term \((d - 1)(-\log(1 - \alpha)) - (d - k)(-\log(1 - 2\alpha + \frac{2k}{d}\alpha))\) and show that it is negative for \( 0 < \alpha < 1/2 \). We will use the following inequalities for \(-\log(1 - x)\), which can be deduced from Taylor expansion. If \( 0 \leq x \leq 1/2 \) then \(-\log(1 - x) \leq x + (1/2)x^2 + (2/3)x^3\). If \( 0 \leq x \leq 1 \) then \(-\log(1 - x) \geq x + (1/2)x^2 + (1/3)x^3\). From these inequalities we conclude that \((d - 1)(-\log(1 - \alpha)) - (d - k)(-\log(1 - 2\alpha + \frac{2k}{d}\alpha))\) is bounded from above by
\[ (d - 1)(\alpha + \frac{\alpha^2}{2} + \frac{\alpha^3}{3}) - (d - k)(2(1 - \frac{k}{d})\alpha + 2(1 - \frac{k}{d})^2 \alpha^2 + \frac{8}{3}(1 - \frac{k}{d})^3 \alpha^3). \]

The term \((1 - \frac{k}{d})\) is positive and decreasing in \( k \) if \( 2 \leq k \leq (1 - 1/\sqrt{2})d \). Its minimum value
is $1/\sqrt{2}$. Thus, $(1 - \frac{k}{d})^2 \leq 1/2$ and $(1 - \frac{k}{d})^3 \leq 1/\sqrt{8}$. We deduce from this that

$$(d - 1)(\alpha + \frac{\alpha^2}{2} + \frac{\alpha^3}{3}) - (d - k)[2(1 - \frac{k}{d})\alpha + 2(1 - \frac{k}{d})^2\alpha^2 + \frac{8}{3}(1 - \frac{k}{d})^3\alpha^3] \leq$$

$$= (d - 1)(\alpha + \frac{\alpha^2}{2} + \frac{\alpha^3}{3}) - \frac{d}{\sqrt{2}}[\sqrt{2}\alpha + \alpha^2 + \frac{\sqrt{8}}{3}\alpha^3] =$$

$$= -\alpha - \frac{(\sqrt{2} - 1)d + 1}{2}\alpha^2 - \frac{d + 3}{3}\alpha^3.$$ 

The last term is clearly negative for positive $\alpha$. This shows what we had claimed.

Now we consider the term $\frac{d}{2}(h(\frac{2k}{d}) + 2h(1 - \frac{2k}{d})) + (k - 1)\log(\alpha)$ and show that it is negative for $0 < \alpha < C_{k,d}$. By property (2) of $h(x)$ from (2.1) we have $h(1 - x) \leq x$. Therefore, $h(1 - \frac{2k}{d}) \leq 2k/d$, and $(d/2)[h(\frac{2k}{d}) + 2h(1 - \frac{2k}{d})] \leq k\log(d/2k) + 2k$. Therefore,

$$\frac{d}{2}(h(\frac{2k}{d}) + 2h(1 - \frac{2k}{d})) + (k - 1)\log(\alpha) \leq k\log(d/2k) + 2k + (k - 1)\log(\alpha).$$

The latter term in increasing in $\alpha$ because $k \geq 2$, and it tends to $-\infty$ as $\alpha \to 0$. It is thus negative until its first zero, which is the value $\alpha^*$ satisfying

$$-\log(\alpha^*) = \frac{k\log(d/2k) + 2k}{k - 1}.$$ 

However, $\frac{k\log(d/2k) + 2k}{k - 1} \leq (1 + \frac{1}{k - 1})\log(d/2k) + 4$ since $k \geq 2$.

Consequently, $\alpha^* \geq e^{-4}(2k/d)^{1+1/(k-1)}$, and we conclude that $(d/2)[h(\frac{2k}{d}) + 2h(1 - \frac{2k}{d})] + (k - 1)\log(\alpha)$ is negative for $0 < \alpha < C_{k,d}$. The proof is now complete since we have shown that the derivative of (2.8) is negative for $0 < \alpha < C_{k,d}$ if $2 \leq k \leq (1 - 1/\sqrt{2})d$. 

**Lemma 2.3.7.** Suppose that $2 \leq k \leq (1 - 1/\sqrt{2})d$ and $0 \leq \alpha \leq 1$. Then the entropy term (2.8) is bounded from above by

$$\alpha(k\log(d) + 1) + h(\alpha)(1 - k) + (d/2)\alpha^3.$$ 

**Proof.** We use the properties of $h(x)$ from (2.1). We have that $h(\frac{2k}{d}\alpha) = \alpha h(\frac{2k}{d}) + \frac{2k}{d}h(\alpha)$, $h(\alpha - \frac{2k}{d}\alpha) = \alpha h(1 - \frac{2k}{d}) + (1 - \frac{2k}{d})h(\alpha)$, and $h(1 - 2\alpha + \frac{2k}{d}\alpha) \leq (2\alpha - \frac{2k}{d}\alpha) - \frac{1}{2}(2\alpha - \frac{2k}{d}\alpha)^2$. Therefore,

$$h(\alpha(2k/d)) + 2h(\alpha - \alpha(2k/d)) + h(1 - 2\alpha + \alpha(2k/d)) \leq$$

$$\alpha\left(h(\frac{2k}{d}) + 2h(1 - \frac{2k}{d}) + 2 - \frac{2k}{d}\right) + 2h(\alpha)(1 - \frac{k}{d}) - 2\alpha^2(1 - \frac{k}{d})^2.$$

Now, $H(\alpha, 1 - \alpha) = h(\alpha) + h(1 - \alpha) \geq h(\alpha) + \alpha - (1/2)\alpha^2 - (1/2)\alpha^3$ by property (3) of (2.1). As a result, (2.8) is bounded from above by

$$\alpha(k\log(d) + 1) + h(\alpha)(1 - k) + (d/2)\alpha^3.$$
\[ \alpha \left( \frac{d}{2} h \left( \frac{2k}{d} \right) + dh \left( 1 - \frac{2k}{d} \right) + 1 - k \right) - (k - 1)h(\alpha) + \alpha^2 \left( \frac{d}{2} - d(1 - \frac{k}{d})^2 \right) + \frac{d}{2} \alpha^3. \] (2.9)

The term \( \frac{d-1}{2} - d(1 - \frac{k}{d})^2 \) is increasing in \( k \) and maximized when \( k = (1 - 1/\sqrt{2})d \), where it equals \(-1/2\). Thus \( \alpha^2 \left( \frac{d-1}{2} - d(1 - \frac{k}{d})^2 \right) \) is negative. The term \( \frac{d}{2} h \left( \frac{2k}{d} \right) + dh \left( 1 - \frac{2k}{d} \right) + 1 - k \) simplifies to \( k \log(d) - k \log(2k) + k + 1 \leq k \log(d) + 1 \) because \( k - k \log(2k) < 0 \) if \( k \geq 2 \). Consequently, (2.9) is bounded from above by \( \alpha(k \log(d) + 1) + h(\alpha)(1 - k) + (d/2)\alpha^3 \) as required.

**Completion of the proof of Lemma 2.3.1** Recall that \( N \) was defined to be the number of subsets \( S \subset V(G_{n,d}) \) of size at most \( C_{k,d} \cdot n \) such that \( S \) is not \( k \)-sparse. From (2.6) we have

\[ \mathbb{E} [N] \leq dn^2 \max_{i \leq \alpha \leq C_{k,d} n} \mathbb{E} [Z_{i,ki}]. \]

By Lemma 2.3.5 \( \mathbb{E} [Z_{i,ki}] \) is bounded from above by \( O(d\sqrt{n}) \times \exp \{ n[(d/2)H(M(i/n, ki/nd)) - (d - 1)H(i/n, 1 - (i/n))] \} \). Now,

\[ \max_{1 \leq i \leq C_{k,d} n} \frac{(d/2)H(M(i/n, ki/nd)) - (d - 1)H(i/n, 1 - (i/n))}{(d/2)H(M(\alpha, (k/d)\alpha)) - (d - 1)H(\alpha, 1 - \alpha)} \]

(2.10)

where \( \alpha \) is a continuous parameter. Lemma 2.3.6 shows that the supremum of (2.10) is achieved at \( \alpha = 1/n \) provided that \( 2 \leq k \leq (1-1/\sqrt{2})d \). Lemma 2.3.7 implies that when \( 2 \leq k \leq (1-1/\sqrt{2})d \), the term in (2.10) is bounded from above at \( \alpha = 1/n \) by \( \frac{1}{n}(k \log(d) + 1) + \frac{\log(n)}{n}(1 - k) + \frac{d}{2n^3} \).

Therefore, we deduce that for \( 2 \leq k \leq (1-1/\sqrt{2})d \),

\[ \mathbb{E} [N] \leq O(d^2 n^{2.5}) \exp \left\{ n \left[ \frac{1}{n}(k \log(d) + 1) + \frac{\log(n)}{n}(1 - k) + \frac{d}{2n^3} \right] \right\}. \]

If \( n \geq \sqrt{d} \) then we see that \( \mathbb{E} [N] \leq O(d^{k+2})n^{3.5-k} \). In particular, if \( k > 3.5 \) then \( \mathbb{E} [N] \to 0 \) as \( n \to \infty \). Hence, \( \mathbb{P} [N \geq 1] \leq \mathbb{E} [N] \to 0 \), and this is precisely the statement of Lemma 2.3.1.

### 2.3.2 Density of \( k \)-sparse graphs: proof of Lemma 2.3.2

We begin with the following elementary lemma about the density of \( k \)-sparse sets.

**Lemma 2.3.8.** Let \( S \) be a \( k \)-sparse set in a finite \( d \)-regular graph \( G \). Then \( |S|/|G| \leq \frac{d}{2d-2k} \).

**Proof.** Set \( |G| = n \), and so \( |E(G)| = nd/2 \). Consider the edge-profile \( M(S) \) of \( S \). We have that \( |S|/n = m(S,S) + m(S, S^c) \). Since \( S \) is \( k \)-sparse, \( m(S,S) \leq 2k|S|/(nd) \). The number of edges from \( S \) to \( S^c \) is at most \( d|S^c| \) because \( G_{n,d} \) is \( d \)-regular. Therefore, \( m(S,S^c) \leq |S^c|/n \).

Consequently, \( |S|/n \leq \frac{2k}{d} - 1)|S|/n + 1 \), which implies that \( |S|/n \leq \frac{d}{2d-2k} \). \( \square \)
Chapter 2. Percolation with small clusters in random graphs

Let $E$ denote the event that $G_{n,d}$ contains an induced $k$-sparse subgraph of size $\alpha n$. We bound the probability of $E$ by using the first moment method as well. We will call a subset $S \subset V(G_{n,d})$ $k$-sparse if it induces a $k$-sparse subgraph. By definition, any $k$-sparse set $S$ has the property that $e(S) \leq k|S|$. 

Let $Z = Z(\alpha, G_{n,d})$ be the number of $k$-sparse sets in $G_{n,d}$ of size $\alpha n$. Thus, $\mathbb{P}[E] \leq \mathbb{E}[Z]$. Recall the notation $Z_{i,j}$ from Section 2.3.1. Let $Z_j = Z_{\alpha n,j}(G_{n,d})$ be the number of subsets $S \subset G_{n,d}$ such that $|S| = \alpha n$ and the number of edges in $G_{n,d}[S]$ is $j$. Then,

$$\mathbb{E}[Z] = \sum_{j=0}^{k\alpha n} \mathbb{E}[Z_j].$$

(2.11)

Lemma 2.3.3 implies that $\mathbb{E}[Z_j]$ is of exponential order in $n$. So the sum in (2.11) is dominated by the largest term. From Lemma 2.3.3 applied to $i = \alpha n$ and $j$ we conclude that

$$\mathbb{E}[Z_j] = \binom{n}{\alpha n} \frac{(\alpha n)!((1-\alpha)nd)!((nd/2)!2^{\alpha nd - 2j}}{(\alpha nd - 2j)!j!(1-2\alpha)^{nd + j}(nd)!}.$$

(2.12)

Lemma 2.3.9. If $\alpha > \frac{2k}{d}$ then the expectation of $Z_i$ is maximized when $i = k\alpha n$ for all sufficiently large $n$. Note that $k\alpha n$ is the maximum number of edges contained in a $k$-sparse set.

Proof. We argue as in the proof of Lemma 2.3.4. From the equation for $\mathbb{E}[Z_j]$ in (2.12) we deduce that

$$\frac{\mathbb{E}[Z_{j+1}]}{\mathbb{E}[Z_j]} = \frac{(\alpha nd - 2j - 1)(\alpha nd - 2j)}{4(j + 1)(\frac{1-2\alpha}{2}nd + j + 1)}.$$

This ratio is at least 1 for all $0 \leq j \leq k\alpha n$ if $n$ is sufficiently large, provided that $\alpha > 2k/d$. Indeed, subtracting the denominator from the numerator of the ratio gives $\alpha nd(\alpha nd - 1) - 2(1 - 2\alpha)nd - 4 - 2j(nd + 3)$. This is non negative for all $0 \leq j \leq k\alpha n$ if and only if

$$k\alpha n \leq \frac{1}{2}(\alpha nd(\alpha nd - 1) - (1 - 2\alpha)nd - 2}{nd + 3}.$$

(2.13)

If the inequality in (2.13) fails to hold for all sufficiently large $n$ then after dividing through by $n$ and letting $n \to \infty$ we conclude that $k\alpha \geq (1/2)\alpha^2 d$. This implies that $\alpha \leq 2k/d$, which contradicts our assumption. \hfill $\square$

From Lemma 2.3.5 applied to $\mathbb{E}[Z_{i,j}]$ for $i = \alpha n$ and $j = k\alpha n$ we conclude that

$$\mathbb{E}[Z_j] \leq O(\sqrt{n}) \exp\left\{n \left[\frac{d}{2}H(M(\alpha, j/nd)) - (d - 1)H(\alpha, 1 - \alpha)\right]\right\}.$$

(2.14)

For the rest of this section we assume that $\alpha \geq (\log d)/d$ and $d$ is large enough such that $(\log d)/d > 2k/d$. This holds due to $k = o(\log d)$. If $\alpha < (\log d)/d$ then there is nothing to prove.
We conclude from Lemma 2.3.9, (2.14) and (2.11) that

\[
E[Z] \leq (kn)E[Z_{kon}]
\]

\[
\leq O(kdn^{3/2}) \exp \left\{ n \left[ \frac{d}{2} H \left( M(\alpha, \frac{k}{d} \alpha) \right) - (d - 1)H(\alpha, 1 - \alpha) \right] \right\}. \tag{2.15}
\]

Note that \(M(\alpha, \frac{k}{d} \alpha)\) equals

\[
M(\alpha, \frac{k}{d} \alpha) = \left[ \begin{array}{cc} \frac{2k\alpha}{d} & \alpha - \frac{2k\alpha}{d} \\ \alpha - \frac{2k\alpha}{d} & 1 - 2\alpha + \frac{2k\alpha}{d} \end{array} \right].
\]

This matrix may depend on \(n\) through \(\alpha\). If it does then we replace \(\alpha\) by its limit supremum as \(n \to \infty\). By an abuse of notation we denote the limit supremum by \(\alpha\) as well.

For \(d \geq 3\) define \(\alpha_d = \alpha_{d,k}\) by

\[
\alpha_d = \sup \{ \alpha : 0 \leq \alpha \leq 1 \text{ and } \frac{d}{2} H(M(\alpha, \frac{k}{d} \alpha)) - (d - 1)H(\alpha, 1 - \alpha) > 0 \}.
\]

Thus, if \(\alpha > \alpha_d\) then from the continuity of the entropy function \(H\) we conclude that for all sufficiently large \(n\) the function \(\frac{d}{2} H \left( M(\alpha, \frac{k}{d} \alpha) \right) - (d - 1)H(\alpha, 1 - \alpha) < 0\). Consequently, from (2.21) and (2.15) we conclude that \(\lim_{n \to \infty} \mathbb{P}[E] = 0\). We devote the rest of this section to bounding the entropy functional in order to show that \(\alpha_d \leq (2 + \epsilon)\frac{\log d}{d}\) as \(d \to \infty\).

First, we show that \(\alpha_d \to 0\) as \(d \to \infty\). Suppose otherwise, that \(\limsup_{d \to \infty} \alpha_d = \alpha_\infty > 0\). Lemma 2.3.8 implies that \(\alpha_\infty \leq 1/2\) because \(\alpha_d \leq d/(2d - 2k)\) and \(k = o(\log d)\). Then after passing to an appropriate subsequence in \(d\), noting that \(2k/d \to 0\) as \(d \to \infty\) due to \(k = o(\log d)\), and using the continuity of \(H\) we see that

\[
\lim_{d \to \infty} \frac{1}{2} H(M(\alpha_d, \frac{k}{d} \alpha_d)) - H(\alpha_d, 1 - \alpha_d) = \frac{1}{2} H(M(\alpha_\infty, 0)) - H(\alpha_\infty, 1 - \alpha_\infty).
\]

However, \((1/2)H(M(x, 0)) - H(x, 1 - x) = (1 - x) \log(1 - x) - (1/2)(1 - 2x) \log(1 - 2x)\), and this is negative for \(0 < x \leq 1/2\). This can be seen by noting that the derivative of the expression is negative for \(x > 0\) and the expression vanishes at \(x = 0\). Therefore, for all large \(d\) along the chosen subsequence we have \(\frac{d}{2} H \left( M(\alpha_d, \frac{k}{d} \alpha_d) \right) - (d - 1)H(\alpha_d, 1 - \alpha_d) < 0\); a contradiction.

We now analyze the supremum of the entropy functional for large \(d\) in order to bound \(\alpha_d\). From the properties of \(h(x)\) in (2.1) we deduce that

\[
H(M(\alpha)) = h \left( \frac{2k\alpha}{d} \right) + 2h \left( \alpha - \frac{2k\alpha}{d} \right) + h \left( 1 - 2\alpha + \frac{2k\alpha}{d} \right)
\]

\[
\leq 2[h(\alpha) + \alpha - \alpha^2] + \frac{2k}{d}[\alpha - h(\alpha) + \alpha \log \frac{d}{2k} + 2\alpha^2] + O(\alpha^3), \tag{2.16}
\]

\[
H(\alpha, 1 - \alpha) = h(\alpha) + \alpha - \frac{1}{2} \alpha^2 + O(\alpha^3). \tag{2.17}
\]
From (2.16) and (2.17) we get that the entropy functional \( \frac{d}{2}H(M(\alpha)) - (d - 1)H(\alpha, 1 - \alpha) \) is at most
\[
- \frac{d}{2} \alpha^2 + k[\alpha - h(\alpha) + \alpha \log \left( \frac{d}{2k} \right) + 2\alpha^2] + \alpha + h(\alpha) + O(\alpha^3). \tag{2.18}
\]

Now, \( k(\alpha + 2\alpha^2) + \alpha \leq 4k\alpha \). Also, \( \log(d/2k) \leq \log(d/k) \). Hence, (2.18) is bounded by
\[
- \frac{d}{2} \alpha^2 + k[\alpha \log(d/k) - h(\alpha)] + \alpha + 4k\alpha + O(\alpha^3). \tag{2.19}
\]

Let us write \( \alpha = \beta \frac{\log d}{d} \) where \( \beta \geq 1 \). In terms of \( \beta \) we have that \( h(\alpha) = \beta \frac{\log^2 d - \log d \log \log d}{d} + h(\beta) \frac{\log d}{d} \). Since \( \beta \geq 1 \) the term \( h(\beta) \leq 0 \) and we get that \( -\frac{d}{2} \alpha^2 + h(\alpha) \leq \left( -\frac{\beta^2}{2} + \beta \right) \frac{\log^2 d}{d} \).

Now we consider \( \alpha \log(d/k) - h(\alpha) \). It equals \( \beta \frac{\log d \log d \log d - \log k}{d} + \beta \log \beta \frac{\log d}{d} \). Substituting \( k = \epsilon_d d \log d \) and combining these inequalities we see that (2.19) is bounded by
\[
[1 - (1/2)\beta + \epsilon_d \log \beta - \epsilon_d \log(\epsilon_d) + 4\epsilon_d] \beta \frac{\log^2 d}{d} + O(\frac{\beta^3 \log^3 d}{d^2}). \tag{2.20}
\]

Some elementary calculus shows that in order for \( 1 - (1/2)\beta + \delta \log \beta - \delta \log(\delta) + C\delta \) to be non-negative, \( \beta \) must satisfy \( \beta \leq 2 - 2\delta \log(\delta) + 2C\delta \), provided that \( 0 \leq \delta \leq 1 \). Furthermore, as \( \beta = o(d/(\log d)) \) we see that \( \beta^3 (\log^3 d)/d^2 \) is of order \( o(\beta^2 (\log^2 d)/d) \).

We can thus conclude that there is a function \( \delta(d) = \delta(\epsilon_d) \) such that \( \delta(d) \to 0 \) as \( d \to \infty \) and (2.20) is negative unless \( \beta \leq 2 + \delta(d) \). As a result, for all large \( d \) we have that \( \alpha_d \leq (2 + \delta(d)) \frac{\log d}{d} \), and the latter is bounded by \( (2 + \epsilon) \frac{\log d}{d} \) for all large \( d \). This completes the proof of Lemma 2.3.2.

### 2.3.3 Remarks on percolation with small clusters for fixed \( d \)

Recall from the introductory section that the quantity \( \alpha^\tau(d) = \lim_{n \to \infty} \alpha^\tau(G_{n,d}) \) exits and is non-random. We briefly discuss what is known about \( \alpha^\tau(d) \) for small values of \( d \). For independent sets, McKay [18] proved that \( \alpha^1(3) \leq 0.4554 \) and this bound was improved recently by Barbier et al. [4] to \( \alpha^1(3) \leq 0.4509 \). Regarding lower bounds, Csóka et al. [16] showed by way of local algorithms that \( \alpha^1(3) \geq 0.4361 \) and this was improved to \( \alpha^1(3) \geq 0.4375 \) by Hoppen and Wormald [30] via local algorithms as well.

A problem closely related to percolation with small clusters is that of the size density of the largest induced forest. Hoppen and Wormald [29] provide a lower bound for the largest size density of an induced forest in \( G_{n,d} \) and their construction can be used to get the same lower bound for \( \sup \alpha^\tau(d) \). Upper bounds on the density of the largest induced forest of \( G_{n,d} \) were given by Bau et al. [5] with numerical values for \( d \leq 10 \). Their upper bounds also serve as upper bounds for \( \sup \alpha^\tau(d) \). On the other hand it is known that \( \sup \alpha^\tau(3) = 3/4 \) through general results on the fragmentability of graphs by Edwards and Farr [21].

The techniques used to prove Theorem 1.2.5 can be applied with little modification to show that the size density of the largest induced forest in \( G_{n,d} \) is, with high probability, at most \( (2 + o(1)) \frac{\log d}{d} \) as \( d \to \infty \). The same conclusion holds for the size density of the largest
Chapter 2. Percolation with small clusters in random graphs

$k$-independent sets in $G_{n,d}$ for every fixed $k$. (A $k$-independent set is a subset of vertices such that the induced subgraph has maximum degree $k$.)

2.4 Erdős–Rényi graphs

In this section we prove Theorem 1.2.10. But first a remark on the statement of the theorem. The truth is that it is not possible to beat the $2(\log \lambda) / \lambda$ bound on the size density of induced subgraphs even if the components are allowed to be of linear size. In other words, a result similar to Theorem 1.2.5 also holds for Erdős–Rényi graphs. However, we prove a weaker statement to keep things simple. Theorem 1.2.10 is strong enough to show sub–optimality of local algorithms.

Throughout the rest of this section let $G \sim \text{ER}(n, \lambda/n)$.

**Lemma 2.4.1.** The expected number of cycles of length no more than $\tau$ in $G$ is at most $(\lambda^\tau \log \tau) / 2$.

**Proof.** Let $C_\ell$ denote the number of cycles of length $\ell \geq 3$ in $G$. The number of cycles of length at most $\tau$ is $C_{\leq \tau} = C_3 + \cdots + C_\tau$. The number of possible configurations of $\ell$ vertices that can form a cycle in ER is $\binom{n}{\ell} \ell! / (2\ell)$. As each of the $\ell$ edges in such a configuration is included independently with probability $d/n$ we conclude that

$$
\mathbb{E}[C_\ell] = \binom{n}{\ell} \ell! (d/n)\ell \leq \frac{\lambda^\ell}{2\ell}.
$$

Note that $\sum_{\ell=3}^{\tau} 1 / (2\ell) \leq \int_2^{\tau} \frac{1}{t} dt = \log(\tau/2) \leq \log \tau$. Thus,

$$
\mathbb{E}[C_{\leq \tau}] = \sum_{\ell=3}^{\tau} \mathbb{E}[C_\ell] \leq \sum_{\ell=3}^{\tau} \frac{\lambda^\ell}{2\ell} \leq \frac{\lambda^\tau \log \tau}{2}.
$$

Let $X_{n,\tau}$ be the number of cycles of length at most $\tau$ in $G$. It follows from Lemma 2.4.1 that if $\tau = \log_3(n) - \log \log \log(n) - \log(\omega_n)$ then $\mathbb{E}[X_{n,\tau}] = O(n/\omega_n)$ as $n \to \infty$.

Let $E$ denote the event that $G$ contains a percolation set of size $\alpha n$ with clusters of size at most $\tau$. We can assume that $\alpha > (2e) / \lambda$ for otherwise there is nothing to prove due to $\lambda \geq 5$. Also, $\alpha n \leq n - 1$ as noted earlier. We bound the probability of $E$ by using the first moment method. From this we will show that if $\alpha n$ is bigger than the bound in the statement of Theorem 1.2.10 then $\mathbb{P}[E] \to 0$ as $n \to \infty$.

Set $\mu_n = \mathbb{E}[X_{n,\tau}] = O(n/\omega_n)$ for $\tau$ in the statement of Theorem 1.2.10. Fix $\delta > 0$ and note that $\mathbb{P}[X_{n,\tau} \geq \mu_n / \delta] \leq \delta$ from Markov’s inequality.

Let $Z = Z(\alpha, G)$ be the number of percolation sets in $G$ of size $\alpha n$ with clusters of size at most $\tau$. From the observation above we have that

$$
\mathbb{P}[E] \leq \mathbb{P}[E \cap \{X_{n,\tau} \leq \mu_n / \delta\}] + \delta \leq \mathbb{E}[Z; X_{n,\tau} \leq \mu_n / \delta] + \delta
$$

(2.21)
1.2.10 For then we have that \( \limsup_{n \to \infty} P[E] \leq \delta \) for any \( \delta > 0 \), and thus, \( P[E] \to 0 \).

We now make a simple observation about percolation sets with small clusters that is crucial to our analysis. Let \( S \) be a percolation set with clusters of size at most \( \tau \). If we remove an edge from every cycle of the induced graph \( G[S] \) of length at most \( \tau \) then the components of \( G[S] \) become trees. In that case the number of remaining edges in \( G[S] \) is at most \( |S| \). Therefore, the number of edges in \( G[S] \) is at most \( |S| + (\mu_n/\delta) \). This bound is important because it shows that the subgraph included by percolation sets with small clusters is much more sparse relative to the original graph.

Let \( M = M(\alpha, \tau, \delta, G) \) be the number of subsets \( S \subset G \) such that \( |S| = \alpha n \) and the number of edges in \( G[S] \) is at most \( |S| + (\mu_n/\delta) \). Notice that the number of edges in \( G[S] \) is distributed as the binomial random variable \( \Bin((\binom{n}{2}), d/n) \). The observation above implies that

\[
\mathbb{E}[Z; X_{n,\tau} \leq \mu_n/\delta] \leq \mathbb{E}[M] = \binom{n}{\alpha n} P[\Bin(\frac{\alpha n}{2}, d/n) \leq \alpha n + (\mu_n/\delta)].
\]

(2.22)

We need to following lemma about binomial probabilities.

**Lemma 2.4.2.** Let \( \Bin(m, p) \) denote a binomial random variable with parameters \( m \geq 1 \) and \( 0 \leq p \leq 1 \). If \( 0 < p \leq 1/2 \) and \( 0 < \mu \leq 1 \) then the following bound holds:

\[
P[\Bin(m, p) \leq \mu mp] = O(\sqrt{mp}) \times e^{-m(\mu p \log(1-\mu p) - \mu p)}.
\]

*Proof.* Since \( P[\Bin(m, p) = k] = \binom{m}{k} p^{k}(1-p)^{m-k} \), it is easy to check that \( P[\Bin(m, p) = k - 1] \leq P[\Bin(m, p) = k] \) if \( k \leq \mu mp \). Therefore,

\[
P[\Bin(m, p) \leq \mu mp] = P[\Bin(m, p) \leq \mu mp] \leq \mu mp P[\Bin(m, p) = \mu mp].
\]

We can estimate \( P[\Bin(m, p) = \mu mp] \) by \( \binom{m}{\mu mp} p^{\mu mp}(1-p)^{m-\mu mp} \) with a multiplicative error term of constant order. Now, from Stirling’s approximation we get

\[
\binom{m}{\mu mp} p^{\mu mp}(1-p)^{m-\mu mp} \leq \frac{O(1)}{\sqrt{\mu mp(1-\mu p)}} e^{mH(\mu p)} [p^{\mu p}(1-p)^{1-\mu p}]^m \\
\leq \frac{O(1)}{\sqrt{\mu mp}} e^{m[-\mu p \log(1-\mu p) + (1-\mu p) \log(1-\mu p) - \log(1-p)]} \quad (\text{as } 1 - \mu p \geq 1/2).
\]

From the Taylor expansion of \(-\log(1-x)\) we can deduce that \( x \leq -\log(1-x) \leq x + x^2 \) if
0 ≤ x ≤ 1/2. Thus, \( \log(1 - \mu p) - \log(1 - p) \geq p - \mu p - \mu^2 p^2 \), and hence

\[
-\mu p \log \mu - (1 - \mu p)(\log(1 - \mu p) - \log(1 - p)) \leq -\mu p \log \mu - (1 - \mu)p + \mu p^2.
\]

After combining this inequality with the inequality for \( \mathbb{P}[\text{Bin}(m, p) = \lfloor \mu mp \rfloor] \) above we deduce the conclusion of the lemma.

We now use Lemma 2.4.2 to provide an upper bound for \( \mathbb{P}[\text{Bin}(\frac{\alpha n}{2}, \lambda/n) \leq \alpha n + (\mu n/\delta)] \). We require that \( n \geq 2\lambda \) and see that the value of the parameter \( \mu \) is \( \mu = \frac{\alpha n + (\mu n/\delta)}{(\frac{\alpha n}{2})(\lambda/n)} \).

Recall that \( \alpha > 2e/\lambda \). With this assumption and for \( n \geq 2\lambda \) it is easy to show that \( \mu \leq \frac{2}{\lambda^2} + O(1/\omega_n) \) where the big O constant depend on \( \lambda \). For all large \( n \) we thus have that \( \mu \leq e^{-1} < 1 \). From Lemma 2.4.2 we deduce that

\[
\mathbb{P}\left[\text{Bin}\left(\frac{\alpha n}{2}, \lambda/n\right) \leq \alpha n + (\mu n/\delta)\right] \leq O(\sqrt{n\lambda})e^{-\frac{\alpha^2}{2}(\lambda/n)\delta} \left[\mu \log(\mu) + 1 - \frac{\lambda}{n} + O(1/\omega_n)\right]. \tag{2.23}
\]

We now simplify the exponent in (2.23). The function \( x \rightarrow x \log x \) is decreasing for \( 0 \leq x \leq e^{-1} \). Hence, as \( \mu \leq \frac{2}{\alpha^2} + O(1/\omega_n) \), we see that \( \mu \log(\mu) \geq (\frac{2}{\alpha^2} + O(1/\omega_n)) \log \left( \frac{2}{\alpha^2} + O(1/\omega_n) \right) \).

Using some basic estimates about \( \log x \), which we omit, it follows that

\[
\mu \log(\mu) + 1 - \mu \geq \frac{2}{\alpha^2} \log \left( \frac{2}{\alpha^2} \right) + 1 - \frac{2}{\alpha^2} - O(1/\omega_n).
\]

Also, \( \frac{\alpha^2}{2} \leq -O(1) \). Combining these estimates we gather that the exponent in (2.23) is at most

\[
-n \left( \alpha \log \left( \frac{2}{\alpha^2} \right) + \frac{\alpha^2}{2} - \alpha \right) + O\left( \max\{n/\omega_n, 1\} \right).
\]

Now we can provide an upper bound to \( \mathbb{E}[M] \) from (2.22). As we have seen from Stirling’s approximation, \( \binom{n}{\alpha n} \leq 2e^{nH(\alpha, 1-\alpha)} \). Combining this with the bound on the binomial probability we get that

\[
\mathbb{E}[M] \leq O_d(\sqrt{n})e^n \left[ H(\alpha, 1-\alpha) + \alpha \log \left( \frac{\alpha^2}{\alpha^2} \right) + \frac{\alpha^2}{2} - \alpha \right] + O(1/\omega_n, 1) \right].
\]

However, \( H(\alpha, 1-\alpha) + \alpha \log \left( \frac{\alpha^2}{\alpha^2} \right) + \frac{\alpha^2}{2} - \alpha = h(1-\alpha) + \alpha(1 + \log(\lambda/2)) - (\lambda/2)\alpha^2 \).

Since \( h(1-\alpha) \leq \alpha \) from (3) of (2.1) we deduce that \( h(1-\alpha) + \alpha(1 + \log(\lambda/2)) - (\lambda/2)\alpha^2 \leq \alpha(2 + \log(\lambda/2) - (\lambda/2)\alpha) \). This implies that

\[
\mathbb{E}[M] \leq O(\lambda(\sqrt{n})e^n \left[ 2 + \log(\lambda/2) - (\lambda/2)\alpha \right] + O(1/\omega_n, 1)).
\]

From (2.21) and (2.22) we have \( \mathbb{P}[E] \leq \mathbb{E}[Z; X_{n, \tau} \leq \mu n/\delta] + \delta \leq \mathbb{E}[M] + \delta \), and thus,

\[
\mathbb{P}[E] \leq O(\lambda(\sqrt{n})e^n \left[ 2 + \log(\lambda/2) - (\lambda/2)\alpha \right] + O(1/\omega_n, 1)) + \delta.
\]

If \( 2 + \log(\lambda/2) - (\lambda/2)\alpha < 0 \) then we get that \( \limsup_{n \to \infty} \mathbb{P}[E] \leq \delta \) for all \( \delta > 0 \). Thus
implies that $\mathbb{P}[E] \to 0$ as $n \to \infty$. Therefore, with high probability ER does not contain induced subgraphs of size larger than $\alpha n$ such that its components have size at most $\tau = \log_\lambda(n) - \log \log \log(n) - \log(\omega_n)$. The condition $2 + \log(\lambda/2) - (\lambda/2)\alpha < 0$ is equivalent to $\alpha > \frac{2}{\lambda}(\log \lambda + 2 - \log 2)$, which is precisely the bound in the statement of Theorem 1.2.10.
Chapter 3

Factor of IID processes on regular trees

The proof of Theorem 1.2.2 is based on the close connection between FIID processes on \( T_d \) and their projections onto random \( d \)-regular graphs. We will explain how to project an FIID percolation on \( T_d \) to a large, random \( d \)-regular graph. This will result in a percolation process on random \( d \)-regular graphs whose local statistics, such as the density and correlation, are close to that of the original with high probability. Then we use combinatorial properties of random \( d \)-regular graphs to bound the probability of observing a percolation process whose statistics are close to that of the original. This procedure results in a fundamental non negativity condition for an entropy-type functional that can be associated to the FIID percolation process. From there we proceed to show that the entropy functional is non-negative only if the density is small with respect to the correlation. The entropy functional that we come upon has been studied by Bowen [13], who developed an isomorphism invariant for free group actions on measure spaces in order to resolve questions in ergodic theory. Backhausz and Szegedy [2] have also used similar entropy arguments to show that certain processes in random \( d \)-regular graphs cannot be represented on trees.

3.1 Connection to processes on random regular graphs

Let \( \mathcal{G}_{n,d} \) denote a random \( d \)-regular graph sampled according to the configuration model (see Section 2.2 for the definition). Let \( G_{n,d} \) denote the set of all pairings in the configuration model (so that \( \mathcal{G}_{n,d} \) is a uniform random element of \( G_{n,d} \)).

In order to project FIID processes on \( T_d \) onto \( \mathcal{G}_{n,d} \) it is necessary to understand the local structure of \( \mathcal{G}_{n,d} \). It is well known (see [31] chapter 9.2) that the number of cycles of length \( \ell \) in \( \mathcal{G}_{n,d} \) converges in moments to a Poisson random variable with mean \( (d - 1)^\ell / 2\ell \). Consequently, for every constant \( L \), the expected number of cycles in \( \mathcal{G}_{n,d} \) of length at most \( L \) remains bounded as \( n \to \infty \).
It follows from the fact above that locally $G_{n,d}$ is a tree around most vertices. Indeed, if the $r$-neighbourhood, $N_G(v, r)$, of a vertex $v$ in a graph $G$ is not a tree then it contains a cycle of length at most $2r$. Thus, $N_G(v, r)$ is a tree if $v$ is not within graph distance $r$ of any cycle in $G$ with length at most $2r$. Since $N_{G_{n,d}}(v, r)$ contains at most $d^r$ vertices, and the expected number of cycles in $G_{n,d}$ of length at most $2r$ is bounded in $n$, the expected number of vertices in $G_{n,d}$ such that $N_{G_{n,d}}(v, r)$ is not a tree remains bounded in $n$ for every fixed $r$. An equivalent formulation is that if $o_n$ is a uniform random vertex of $G_{n,d}$ then for every $r \geq 0$

$$
\mathbb{P} \left[ N_{G_{n,d}}(o_n, r) \cong N_{T_d}(o, r) \right] = 1 - O_d(r(1/n)) \quad \text{as } n \to \infty. \quad (3.1)
$$

We now explain the procedure that projects FIID processes on $T_d$ onto $G_{n,d}$. For simplicity let $T_{d,r}$ denote the rooted subtree $N_{T_d}(o, r)$. Following Lyons [34], we say a factor $f$ for an FIID process $\Phi \in \chi^{V(T_d)}$ is a block factor if there exists a finite $r$ such that $f : [0, 1]^{V(T_{d,r})} \to \chi$. The smallest such $r$ is the radius of the factor. It is well known [34] that given $\Phi$ with factor $f$ it is possible to find block factors $f_n$ such that the corresponding processes $\Phi^n$ converge to $\Phi$ in the weak topology. As such, for any FIID percolation process $Y$ there exists approximating percolation processes $Y^n$ having block factors such that $\text{den}(Y^n) \to \text{den}(Y)$ and $\text{corr}(Y^n) \to \text{corr}(Y)$. Henceforth, we assume that all factors associated to FIID percolation processes on $T_d$ are block factors.

Suppose that an FIID process $\Phi$ has a block factor $f$ of radius $r$. We project it to a fixed $d$-regular graph $G \in G_{n,d}$. As $f$ maps to a finite set $\chi$, we simply take $\chi = \{0, 1, \ldots, q\}$. We begin with a random labelling $X = (X(v), v \in V(G))$ of $G$. Given any vertex $v \in G$ if its $r$-neighbourhood is a tree then set $\Phi_G(v) = f(X(u), u \in N_G(v, r))$. This is allowed since $N_G(v, r) = T_{d,r}$ by assumption. Otherwise, set $\Phi_G(v) = 0$. The process $\Phi_G$ is the projection of $\Phi$ onto $G$.

If we project $\Phi$ to $G_{n,d}$ then the projection $\Phi_{G_{n,d}}$ has the property that its local statistics converge to the local statistics of $\Phi$ as $n \to \infty$. To be precise, let $(o_n, o'_n)$ be a uniform random directed edge on $G_{n,d}$. Note this implies that $o_n$ is a uniform random vertex of $G_{n,d}$. Also, let $o$ be the root of $T_d$ with a neighbouring vertex $o'$. Then for any $i, j \in \chi$, the probabilities

$$
\mathbb{P} \left[ \Phi_{G_{n,d}}(o_n) = i, \Phi_{G_{n,d}}(o'_n) = j \right] \to \mathbb{P} \left[ \Phi(o) = i, \Phi(o') = j \right] \quad \text{as } n \to \infty. \quad (3.2)
$$

Consequently, $\mathbb{P} \left[ \Phi_{G_{n,d}}(o_n) = i \right] \to \mathbb{P} \left[ \Phi(o) = i \right]$ as well.

To see this, observe that the value of the pair $(\Phi_{G_{n,d}}(o_n), \Phi_{G_{n,d}}(o'_n))$ depends only on the labels on the $r$-neighbourhood of the edge $(o_n, o'_n)$. If this is a tree then $N_{G_{n,d}}((o_n, o'_n), r)$ is isomorphic to $N_{T_d}((o, o'), r)$, and $(\Phi_{G_{n,d}}(o_n), \Phi_{G_{n,d}}(o'_n))$ agrees with $(\Phi(o), \Phi(o'))$ if the labels on $N_{G_{n,d}}((o_n, o'_n), r)$ are lifted to $N_{T_d}((o, o'), r)$ in the natural way. On the other hand, if $N_{G_{n,d}}((o_n, o'_n), r)$ is not a tree then $N_{G_{n,d}}(o_n, r + 1)$ is also not a tree since $N_{G_{n,d}}((o_n, o'_n), r) \subset N_{G_{n,d}}(o_n, r + 1)$. From the local structure of $G_{n,d}$ discussed earlier in this section (equation (3.1) we know that $\mathbb{P} \left[ N_{G_{n,d}}(o_n, r + 1) \not\cong T_{d,r+1} \right] = O_d(r(1/n))$. As a result, it is straightforward to
conclude that \( |P[\Phi_{G_n,d}(\sigma_n) = i, \Phi_{G_n,d}(\sigma'_n) = j] - P[\Phi(\sigma) = i, \Phi(\sigma') = j]| = O_{d,r}(1/n). \)

In our proof we need more than the convergence of local statistics. We require that the random quantities \( \#\{(u,v) \in E(\mathcal{G}_{n,d}) : \Phi_{G_n,d}(u) = i, \Phi_{G_n,d}(v) = j\} \) are close to their expectations. The tuple \((u,v)\) denotes a directed edge so that \((u,v) \neq (v,u)\), and we count the number of directed edges \((u,v)\) such that \( \Phi_{G_n,d}(u) = i \) and \( \Phi_{G_n,d}(v) = j \). Note that

\[
P[\Phi_{G_n,d}(\sigma_n) = i, \Phi_{G_n,d}(\sigma'_n) = j] = \mathbb{E}\left[ \frac{\#\{(u,v) \in E(\mathcal{G}_{n,d}) : \Phi_{G_n,d}(u) = i, \Phi_{G_n,d}(v) = j\}}{nd} \right].
\]

In other words, we require that the empirical values of the local statistics of \( \Phi_{G_n,d} \) are close to their expected value with high probability. Given the convergence of the local statistics of \( \Phi_{G_n,d} \) shown above, this would imply that the empirical local statistics of \( \Phi_{G_n,d} \) are close to the local statistics of \( \Phi \). To achieve this we use the Hoeffding–Azuma concentration inequality (see [11] Theorem 1.20). Consider a product probability space \( (\Omega^n, \mu^n) \) and a bounded function \( h : \Omega^n \to \mathbb{R} \). If \( h \) is \( L \)-Lipschitz with respect to the Hamming distance in \( \Omega^n \) then the Hoeffding–Azuma inequality states that

\[
\mu^n (|h - \mathbb{E}\mu^n[h]| > \lambda) \leq 2e^{-\frac{\lambda^2}{2L^2n}}.
\]

In our case we fix a graph \( G \in G_{n,d} \) and consider the probability space generated by a random labelling \( X \) of \( G \). The function \( h \) is taken to be \( \#\{(u,v) \in E(G) : \Phi_G(u) = i, \Phi_G(v) = j\} \). We verify that \( h \) is Lipschitz as a function of the labels with Lipschitz constant at most \( 2d + 1 \) (recall that \( r \) is the radius of the factor associated with \( \Phi \)). If the value \( X(w) \) of the label at \( w \) is changed to \( X'(w) \) then the value of the pair \( (\Phi_G(u), \Phi_G(v)) \) can only change for edges \((u,v)\) that are within graph distance \( r \) of \( w \). Otherwise, the labels used to evaluate \( (\Phi_G(u), \Phi_G(v)) \) does not contain \( X'(w) \). Therefore, the number of directed edges \((u,v)\) where the value of \( \Phi_G \) can change is at most \( 2d + 1 \), which is a trivial upper bound on the number of edges that meet \( N_G(w, r) \).

From the Hoeffding–Azuma inequality we conclude that

\[
P \left( \left| \frac{\#\{(u,v) \in E(\mathcal{G}_{n,d}) : \Phi_{G_n,d}(u) = i, \Phi_{G_n,d}(v) = j\}}{nd} - \mathbb{E}[\Phi_{\mathcal{G}_{n,d}}(\sigma_n) = i, \Phi_{\mathcal{G}_{n,d}}(\sigma'_n) = j] \right| > \epsilon \right) \leq 2e^{-\frac{\epsilon^2 n}{8d^2r}}.
\]

Due to the convergence of local statistics we can replace \( P[\Phi_{G_n,d}(\sigma_n) = i, \Phi_{G_n,d}(\sigma'_n) = j] \) with \( P[\Phi(\sigma) = i, \Phi(\sigma') = j] \) in the inequality above at the expense of adding an error term \( \text{err}(n) \) such that \( \text{err}(n) \to 0 \) as \( n \to \infty \). Moreover, by taking a union bound over all \( i,j \in \chi \) we deduce from the inequality above that for some constant \( C_{d,r,q} = O(d^2r \log q) \) and error term
err(n,q) → 0 as n → ∞, we have that

\[ P \left[ \max_{i,j \in \chi} \frac{\# \{(u,v) \in E(G_{n,d}) : \Phi_{G_{n,d}}(u) = i, \Phi_{G_{n,d}}(v) = j\}}{nd} > \epsilon \right] \leq e^{-\frac{\epsilon^2 n}{C_{d,r,q}}} + \text{err}(n,q). \] (3.3)

### 3.2 The fundamental entropy inequality

From (3.3) we see that we can prove bounds on the local statistics of \( \Phi \) if we can use the structural properties of \( G_{n,d} \) to deduce bounds on the empirical value of the local statistics of \( \Phi_{G_{n,d}} \). We achieve the latter by combinatorial arguments.

On any graph \( G \in G_{n,d} \), the process \( \Phi_G \) creates an ordered partition \( \Pi \) of \( V(G) \) with \( q + 1 = \# \chi \) cells, namely, the sets \( \Pi(i) = \Phi_G^{-1}(\{i\}) \) for \( i \in \chi \). (It is ordered in the sense that the cells are distinguishable.) Associated to \( \Pi \) is its edge profile, defined by the quantities

\[ P(i,j) = \frac{\# \{(u,v) \in E(G) : \Phi_G(u) = i, \Phi_G(v) = j\}}{nd} \]
\[ \pi(i) = \frac{\# \{v \in V(G) : \Phi_G(v) = i\}}{n}. \]

As we have noted earlier, if we pick a uniform random directed edge \( (\circ_n, \circ'_n) \) of \( G \) then \( P(i,j) = P[\Phi_G(\circ_n) = i, \Phi_G(\circ'_n) = j] \) and \( \pi(i) = P[\Phi_G(\circ_n) = i] \). In particular, the matrix \( P = [P(i,j)]_{i,j \in \chi} \) and the vector \( \pi = (\pi(i))_{i \in \chi} \) are probability distributions over \( \chi^2 \) and \( \chi \), respectively. Also, the marginal distribution of \( P \) along its rows is \( \pi \), and \( P \) is symmetric because \( (\circ_n, \circ'_n) \) has the same distribution as \( (\circ'_n, \circ_n) \).

Our chief combinatorial ingredient, in particular where we use the structural property of \( G_{n,d} \), is computing the expected number of ordered partitions of \( G_{n,d} \) that admit a given edge profile. From (3.3) we see that any FIID process \( \Phi \) admits, with high probability, an ordered partition of \( V(G) \) with edge profile close to \( P_\Phi = [P[\Phi(\circ) = i, \Phi(\circ') = j]]_{i,j} \) and \( \pi_\Phi = (P[\Phi(\circ) = i])_i \). Thus, if the probability of observing a partition of \( G_{n,d} \) with this edge profile is vanishingly small then the process \( \Phi \) cannot exist on \( T_d \). Bounding the probability of observing partitions with a given edge profile is difficult, so instead we bound the expected number of such partitions, which serves as an upper bound to the probability.

Let \( Z(P, \pi) = Z(P, \pi, G) \) be the number of ordered partitions of \( V(G) \) with edge profile \( (P, \pi) \). Recall that the entropy of a discrete probability distribution \( \mu \) is

\[ H(\mu) = \sum_{x \in \text{support}(\mu)} -\mu(x) \log \mu(x). \]
Lemma 3.2.1. For the random graph $G_{n,d}$,

$$\mathbb{E}[Z(P, \pi, G_{n,d})] \leq \text{poly}(n, d) \times e^{nH(P)-(d-1)H(\pi)}$$

where $\text{poly}(n, d)$ is a polynomial factor.

Proof. Recall that in Lemma 2.3.3 we proved this result for 2-colouring of random graphs. Here we generalize the argument.

The number of candidates for ordered partitions $\Pi$ on $V(G_{n,d})$ that induce the edge profile $(P, \pi)$ on $G_{n,d}$ is equal to the number of partitions of $V(G_{n,d})$ into $q+1$ distinguishable cells such that the $i$-th cell has size $\pi(i)n$. (Of course, we may assume that $(P, \pi)$ is a valid edge profile so that the numbers $\pi(i)n$ and $P(i,j)nd$ are all integers.) The number of such partitions is given by the multinomial term $\frac{n^{\pi(i)n}}{\pi(i)n; i \in \chi} \times \prod_{i} \frac{\pi(i)^{ndP(i,j)} \prod_{\{i,j\}: i \neq j} (ndP(i,j))! \prod_{i} (ndP(i,i) - 1)!}{(nd)!!!}.$

The first product counts, for every $0 \leq i \leq q$, the number of ways to partition the $nd\pi(i)$ half edges from $\Pi(i)$ into distinguishable sub-cells $\Pi(i,j)$ such that $\#\Pi(i,j) = ndP(i,j)$. The half edges in $\Pi(i,j)$ are to be paired with half edges from $\Pi(j,i)$. The second and third products count the number of ways to achieve these pairings. As each configuration of pairings appear with probability $1/(nd)!!!$, the formula above follows.

From the linearity of the expectation it follows that $\mathbb{E}[Z(P, \pi, G_{n,d})]$ equals

$$\left(\frac{n}{\pi(i)n; i \in \chi}\right) \times \frac{\prod_{i} \pi(i)^{ndP(i,j)} \prod_{\{i,j\}: i \neq j} (ndP(i,j))! \prod_{i} (ndP(i,i) - 1)!}{(nd)!!!} \times \prod_{i} \frac{\pi(i)^{ndP(i,j)} \prod_{\{i,j\}: i \neq j} (ndP(i,j))! \prod_{i} (ndP(i,i) - 1)!}{(nd)!!!}.$$ (3.4)

To analyze the asymptotic behaviour of (3.4) we use Stirling’s approximation: $m! = \sqrt{2\pi m}(m/e)^m(1 + O(1/m))$. Note that $m!! = \frac{m!}{2^{(m/2)!m/2}}$ for even $m$. From Stirling’s approximation we can easily conclude that

$$\left(\frac{m}{\rho_i m; 0 \leq i \leq q}\right) = O(m^{-q/2})e^{mH(\rho_0, \ldots, \rho_q)} \text{ and } \frac{m!!}{\sqrt{m!}} = O(m^{-1/4}).$$

Here, the $\rho_i m$ are non-negative integers such that $\rho_0 + \ldots + \rho_q = 1$.

From these two estimates we can simplify (3.4). We ignore writing polynomial factors in $n$ and $d$ explicitly. We get

- $\left(\frac{n}{\pi(i)n; i \in \chi}\right) \leq \text{poly}(n, d) e^{nH(\pi)}$, 

- $e^{nH(P)-(d-1)H(\pi)}$. 


Then

\[ \Pi_i \left( n_d \phi_{(i)} \right) \leq \text{poly}(n, d) e^{nd} \left[ \sum_i \pi(i) H((P(i,j)/\pi(i); 0 \leq j \leq q)) \right], \]

- Finally, \( \frac{\Pi_{(i,j):i \neq j}(n_d P(i,j))!}{(n_d - 1)!} \) is at most

\[ \text{poly}(n, d) \times \left[ \frac{\Pi_{(i,j):i \neq j}(n_d P(i,j))!}{(n_d)!} \right]^{1/2}. \]

From Stirling's formula and after some algebraic simplifications the latter term becomes

\[ \text{poly}(n, d) e^{-\frac{nd}{2} H(P)}. \]

The term \( \sum_i \pi(i) H((P(i,j)/\pi(i); 0 \leq j \leq q)) \) is the conditional entropy of \( P \) given \( \pi \). Algebraic simplification shows that it equals \( H(P) - H(\pi) \). Consequently,

\[ \mathbb{E} [Z(P, \pi, \mathcal{G}_{n,d})] \leq \text{poly}(n, d) e^{n[H(\pi) + dH(P) - dH(\pi) - (d/2)H(P)]}, \]

and the above simplifies to the formula in the statement of the lemma.

We now have the ingredients to prove a key inequality about FIID processes on trees.

**Theorem 3.2.2.** Let \( \Phi \) be an FIID process on \( \mathbb{T}_d \) taking values in \( \chi^{\mathbb{T}_d} \). Let \( (\circ, \circ') \) be a directed edge of \( \mathbb{T}_d \). Let \( P_\Phi \) and \( \pi_\Phi \) be the distributions of \( (\Phi(\circ), \Phi(\circ')) \) and \( \Phi(\circ) \), respectively. Then

\[ \frac{d}{2} H(P_\Phi) - (d - 1)H(\pi_\Phi) \geq 0. \] (3.5)

The term \( \frac{d}{2} H(P_\Phi) - (d - 1)H(\pi_\Phi) \) will be denoted the **entropy functional** for \( \Phi \).

**Proof.** For any two matrices \( M \) and \( M' \), set \( ||M - M'|| = \max_{i,j} |M(i,j) - M'(i,j)| \). For \( 0 < \epsilon < 1 \) consider the event

\[ A(\epsilon) = \{ \mathcal{G}_{n,d} \text{ admits an ordered partition with edge profile } (P, \pi) \text{ satisfying } ||P - P_\Phi|| \leq \epsilon \}. \]

Observe that there are at most \( (2nd)^{(q+1)^2} \) edge profiles \( (P, \pi) \) that satisfy \( ||P - P_\Phi|| \leq \epsilon \). Indeed, any such \( P \) has entries \( P(i,j) \) that are rational numbers of the form \( a/(nd) \) with \( 0 \leq a \leq (P_\Phi(i,j) + \epsilon)nd \). Since \( P_\Phi(i,j) + \epsilon \leq 2 \) for every entry \( P_\Phi(i,j) \), there are no more than \( 2nd \) choices for \( a \). As there are \( (q + 1)^2 \) entries, the bound follows.

From Lemma [3.2.1] and a union bound we observe that

\[ \mathbb{P} [A(\epsilon)] \leq \sum_{(P,\pi): ||P - P_\Phi|| \leq \epsilon} \mathbb{E} [Z(P, \pi, \mathcal{G}_{n,d})] \]

\[ \leq \text{poly}(n, d, q) \times \sup_{(P,\pi): ||P - P_\Phi|| \leq \epsilon} e^{n[H(P) - (d - 1)H(\pi)]. \]


The map \((P, \pi) \to (d/2)H(P) - (d - 1)H(\pi)\) is continuous with respect to the norm \(\| \cdot \|\). Continuity implies that there exists \(\delta(\epsilon) \to 0\) as \(\epsilon \to 0\) such that \(\frac{d}{2}H(P) - (d - 1)H(\pi) \leq \frac{d}{2}H(P_\Phi) - (d - 1)H(\pi_\Phi) + \delta(\epsilon)\) if \(\|P - P_\Phi\| \leq \epsilon\). As a result,

\[
\mathbb{P}[A(\epsilon)] \leq \text{poly}(n, d) \times e^{n[\frac{d}{2}H(P_\Phi) - (d - 1)H(\pi_\Phi) + \delta(\epsilon)]}.
\] (3.6)

On the other hand, from the concentration inequality \((3.3)\) we see that \(\mathbb{P}[A(\epsilon)] \to 1\) as \(n \to \infty\) for every \(\epsilon\) since

\[
\mathbb{P}[A(\epsilon)] \geq \mathbb{P}[\|P_{\Phi_{n,d}} - P_\Phi\| \leq \epsilon] \geq 1 - e^{-\frac{\epsilon^2 n}{Cd,r,q} - \text{err}(n, q)}.
\]

If \(\frac{d}{2}H(P_\Phi) - (d - 1)H(\pi_\Phi) < 0\) then by choosing \(\epsilon\) small enough we can ensure that \(\frac{d}{2}H(P_\Phi) - (d - 1)H(\pi_\Phi) + \delta(\epsilon) < 0\). This implies from \((3.6)\) that \(\mathbb{P}[A(\epsilon)] \to 0\) and provides a contradiction. \(\square\)
Chapter 4

Factor of IID percolation on regular trees

A construction of FIID independent sets due to Lauer and Wormald

Lauer and Wormald study a greedy algorithm that produces FIID independent sets on $T_d$ for every $d \geq 3$ [33]. We describe this algorithm as it will be referred to during the course of the thesis. Our result implies that the lower bound given by this algorithm for $\alpha_{\text{IND}}(d)$ and $\alpha_{\text{per}}(d)$ is asymptotically optimal. We should mention that lower bounds for $\alpha_{\text{IND}}(d)$ of the same asymptotic order are also known in the literature from different constructions (see [23, 33, 40, 41] and the references therein).

The algorithm Fix a percolation density $p \in (0, 1)$ and integer $k \geq 1$. Let $U_0 = V(T_d)$, 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 $T_d$. Consider the subset $I' = \cup_{i=1}^k S_i$. $I'$ may not be an independent set only because some $S_i$ can 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 a subset $I \subset I'$ that is an independent set of $T_d$.

The random set $I$ is an FIID independent set because 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 $T_d$. Furthermore, 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 are arbitrarily close to

$$\beta(d) := \frac{1 - (d - 1)^{-2/(d-2)}}{2}.$$

Consequently, $\alpha_{\text{IND}}(d) \geq \beta(d)$. A simple analysis shows that

$$\frac{\log(d-1)}{d-2} - 2\left(\frac{\log(d-1)}{d-2}\right)^2 \leq \beta(d) \leq$$

35
4.1 Bounds on the density in terms of correlation

The proof of the upper bound in Theorem 1.2.4 is similar to that of Theorem 1.2.5. Except, we can now use the entropy inequality and move directly to the analysis of the entropy functional. During the analysis we need to keep track of the correlation parameter.

Let \( \{Y_d\} \) be a sequence of FIID percolation processes on \( T_d \) such that \( \text{corr}(Y_d) \leq c < 1 \). Set \( P_{Y_d}(i, j) = \mathbb{P}[Y_d(o) = i, Y_d(o') = j] \) where \( (o, o') \) is a directed edge of \( T_d \). In terms of the correlation and density of \( Y_d \) we have

\[
P_{Y_d} = \begin{bmatrix}
corr(Y_d)\text{den}(Y_d)^2 & \text{den}(Y_d) - corr(Y_d)\text{den}(Y_d)^2 \\
\text{den}(Y_d) - corr(Y_d)\text{den}(Y_d)^2 & 1 - 2\text{den}(Y_d) + corr(Y_d)\text{den}(Y_d)^2
\end{bmatrix}.
\]

Recall the function \( h(x) = -x \log x \) for \( 0 \leq x \leq 1 \). For convenience we write \( \rho_d = \text{corr}(Y_d) \) and \( \alpha_d = \text{den}(Y_d) \). By applying the entropy inequality (3.5) from Theorem 3.2.2 to \( Y_d \) we conclude that

\[
\frac{d}{2}[h(\rho_d\alpha_d^2) + 2h(\alpha_d - \rho_d\alpha_d^2) + h(1 - 2\alpha_d + \rho_d\alpha_d^2)] - (d - 1)[h(\alpha_d) + h(1 - \alpha_d)] \geq 0. \tag{4.1}
\]

Before we can analyze this inequality we need to conclude from the fact that \( \rho_d \leq c \) that \( \alpha_d \to 0 \) as \( d \to \infty \). Suppose otherwise, for the sake of a contradiction. By moving to a subsequence in \( d \) we may assume that \( \alpha_d \to \alpha_\infty > 0 \) and \( \rho_d \to \rho_\infty \leq c < 1 \). Then dividing (4.1) through by \( d \) and taking the limit in \( d \) we see that

\[
\frac{1}{2}[h(\rho_\infty\alpha_\infty^2) + 2h(\alpha_\infty - \rho_\infty\alpha_\infty^2) + h(1 - 2\alpha_\infty + \rho_\infty\alpha_\infty^2)] - h(\alpha_\infty) - h(1 - \alpha_\infty) \geq 0.
\]

It is straightforward to show from differentiating with respect to \( \rho_\infty \) that the term above it uniquely maximized at \( \rho_\infty = 1 \), for any \( \alpha_\infty > 0 \). However, when \( \rho_\infty = 1 \), the three terms involving \( \rho_\infty \) simplify to give \( 2[h(\alpha_\infty) + h(1 - \alpha_\infty)] \). (This, of course, follows from subadditivity of entropy and the fact that the entropy of a pair of distributions is maximal when they are independent.) Thus, when \( \rho_\infty = 1 \) the term above equals zero, but is otherwise negative for \( \rho_\infty < 1 \) and \( \alpha_\infty > 0 \). This contradicts the inequality above and allows us to conclude that \( \alpha_d \to 0 \).

Now, we analyze the asymptotic behaviour of \( \alpha_d \) from (4.1). We will use the three properties of \( h(x) \) from 2.1.
We get that
\[
    h(\rho_d \alpha_d^2) + 2h(\alpha_d - \rho_d \alpha_d^2) = h(\rho_d)\alpha_d^2 + 2\alpha h(1 - \rho_d \alpha_d) + 2h(\alpha_d) \\
    \leq [h(\rho_d) + 2\rho_d]\alpha_d^2 + 2h(\alpha_d) - \rho_d^2 \alpha_d^3
\]
because \(h(1 - \rho_d \alpha_d) \leq \rho_d \alpha_d\) via property (3) of (2.1). Similarly,
\[
    h(1 - 2\alpha_d + \rho_d \alpha_d^2) \leq 2\alpha_d - \rho_d \alpha_d^2 - (1/2)(2\alpha_d - \rho_d \alpha_d^2)^2.
\]
After simplifying, we get that \(H(\mathcal{P}_Y) \leq [h(\rho_d) + \rho_d - 2\alpha_d^2 + 2(\alpha_d + h(\alpha_d) + \rho_d \alpha_d^3]).\) Note that \(h(\rho_d) + \rho_d - 1 = -\Psi(\rho_d),\) if we recall that \(\Psi(c) = c \log c - c + 1.\)

Since \(h(1 - \alpha) \geq \alpha_d - (1/2)\alpha_d^2 - (1/2)\alpha_d^3\) by property (2) in (2.1), we also have that \((d - 1)H(\pi_Y) \geq (d - 1)[h(\alpha_d) + \alpha_d] - (d/2)[\alpha_d^2 + \alpha_d^3].\) Therefore,
\[
    \frac{d}{2} H(\mathcal{P}_Y) - (d - 1)H(\pi_Y) \leq -\frac{d}{2} \Psi(\rho_d)\alpha_d^2 + \alpha_d + h(\alpha_d) + \frac{(2\rho_d + 1)d}{2} \alpha_d^3.
\]

The top order terms are \(h(\alpha_d)\) and \(-(d/2)\Psi(\rho_d)\alpha_d^2,\) which are of the same order only if \(\alpha_d\)
is of order \((\log d)/d.\) This explains the emergence of \((\log d)/d\) in the density bounds. If we write \(\alpha_d = \beta_d \log d\) then the inequality translates to
\[
    \frac{d}{2} H(\mathcal{P}_Y) - (d - 1)H(\pi_Y) \leq \left( -\Psi(\rho_d) \beta_d + 1 + \frac{1 - \log \log d}{\log d} + \frac{(2\rho_d + 1)\log d}{2d} \beta_d \right) \frac{d \log^2 d}{d}.
\]

From Theorem 3.2.2 we conclude that \(g(\beta_d) \geq 0\) where
\[
    g(x) = -\frac{\Psi(\rho_d)}{2} x + 1 + \frac{1 - \log \log d}{\log d} + \frac{(2\rho_d + 1)\log d}{2d} x^2.
\]
The two roots of \(g(x)\) are
\[
    \frac{d}{(2\rho_d + 1)\log d} \left( \frac{\Psi(\rho_d)}{2} \pm \sqrt{\frac{\Psi^2(\rho_d)}{4} - \frac{(4\rho_d + 2)}{d} (\log d - \log \log d + 1)} \right).
\]

The quadratic \(g(x)\) is negative in between its two roots and non-negative outside. We show that \(\beta_d\) can not be larger than the smaller root of \(g\) for all large \(d.\) Then we will provide the required asymptotic value of the smaller root as stated in Theorem 1.2.4

First, we show that \(\Psi(\rho_d)/(2\rho_d + 1)\) can not tend to zero in \(d.\) Suppose that \(\rho_d \leq c < 1\). Then \(\Psi(\rho_d) \geq \Psi(c)\) and \(2\rho_d + 1 \leq 3,\) so \(\Psi(\rho_d)/(2\rho_d + 1) \geq \Psi(c)/3.\) Now suppose that \(\rho_d \geq c > 1.\) Then \(\Psi(\rho_d)/(2\rho_d + 1) \geq \Psi(\rho_d)/(3\rho_d),\) and \(\Psi(\rho_d)/\rho_d = \log(\rho_d) - 1 + 1/(\rho_d).\) The function \(\log(\rho_d) - 1 + 1/(\rho_d)\) is increasing if \(\rho_d \geq 1\) and equals zero only if \(\rho_d = 1.\) Therefore, that there are strictly positive quantities \(\ell_+(c)\) and \(\ell_-(c)\) such that \(\Psi(\rho_d)/(2\rho_d + 1) \geq \ell_+(c)\) according to whether \(\rho_d \geq c > 1\) or \(\rho_d \leq c < 1,\) respectively.
Now we claim that \( g(x) \) has two real roots for sufficiently large \( d \). Observe that \( \Psi(\rho_d) \geq \Psi(c) \) in both the cases \( \rho_d > c > 1 \) and \( \rho_d \leq c < 1 \). Hence, \( \Psi(\rho_d)^2/(2\rho_d + 1) \geq \Psi(c)\ell\pm(c) \), where we choose \( \ell_+ \) or \( \ell_- \) according to the assumed bound on \( c \). Consequently, for all \( d \geq d_1 \) we have that \( (\Psi^2(\rho_d)/4) - \frac{(4\rho_d+2)}{d}(\log d - \log \log d + 1) > 0 \) because \( (\log d - \log \log d + 1)/d \to 0 \). The quadratic \( g(x) \) has two real roots when this inequality holds.

Assuming that \( g \) has two roots we now show that \( \beta_d \) is not larger than the smaller root for all large \( d \). Recall that \( \beta_d = o\left(\frac{d}{\log d}\right) \) because \( \alpha \to 0 \). Therefore, given \( c \) there exists a \( d_2 \) such that for \( d \geq d_2 \) we have \( \beta_d \leq \ell\pm(d)/(3\log d) \). But note that the larger root of \( g(x) \) is at least \( \frac{d}{(2\rho_d+1)\log d} \cdot \frac{\Psi(\rho_d)}{d} \geq \ell\pm(d)/(2\log d) \) since \( \Psi(\rho_d)/(2\rho_d + 1) \geq \ell\pm(c) \). Therefore, \( \beta_d \) can not be bigger than the larger root. Then \( \beta_d \) is bounded from above by the smaller root as \( g \) is negative between its two roots.

We have deduced that for \( d \geq \max\{d_1, d_2\} \), \( \beta_d \) is bounded from above by the smaller root of \( g \). An elementary simplification shows that the smaller root equals

\[
\frac{2(1 - \frac{\log d}{\log d} + \frac{1}{\log d})}{\Psi(\rho_d)^2 + \sqrt{\Psi^2(\rho_d) - \frac{4(\rho_d+2)}{d}(\log d - \log \log d + 1)}}.
\]

As \( \sqrt{1-x} \geq 1-x \) for \( 0 \leq x \leq 1 \), we see that \( \sqrt{(\Psi^2(\rho_d)/4) - \frac{4(\rho_d+2)}{d}(\log d - \log \log d + 1)} \) is bounded from below by \( (\Psi(\rho_d)/2) - \frac{8\rho_d+4}{\Psi(\rho_d)} \cdot \frac{\log d - \log \log d + 1}{d} \).

As a result the smaller root is bounded from above for all large \( d \) by

\[
\frac{2}{\Psi(\rho_d)} \times \frac{1 - \frac{\log d}{\log d} + \frac{1}{\log d}}{1 - \frac{8\rho_d+4}{(\log d - \log \log d + 1)} \cdot \frac{\log d - \log \log d + 1}{d\Psi^2(\rho_d)}}
= \frac{2}{\Psi(\rho_d)} \left(1 - \frac{\log d}{\log d} + O\left(\frac{(2\rho_d + 1)\log d}{\Psi^2(\rho_d)d}\right)\right).
\]

This shows that

\[
\beta_d \leq \frac{2}{\Psi(\rho_d)} \left(1 + \frac{1 - \log d}{\log d} + O\left(\frac{(2\rho_d + 1)\log d}{\Psi^2(\rho_d)d}\right)\right).
\]

The upper bound of Theorem 1.2.4 follows immediately because \( \Psi(\rho_d) \geq \Psi(c) > 0 \) and \( (2\rho_d + 1)/\Psi^2(\rho_d) \leq 1/(\Psi(c)\ell\pm(c)) \).

The lower bound

In order to establish the lower bound in Theorem 1.2.4 we interpolate between Bernoulli percolation on \( T_d \) and FIID independent sets of large density. In Section 3 we saw that there exists FIID independent sets \( I_d \) on \( T_d \) with density \( (1 - o(1))\frac{\log d}{d} \). To construct \( Z_d \) from Theorem 1.2.4 we first fix parameters \( 0 < p < 1 \) and \( x > 0 \). Using two independent sources of random labelings of \( T_d \), we generate \( I_d \) and a Bernoulli percolation, \( \text{Ber}_d \), having density \( x\frac{\log d}{d} \). Then for each vertex \( v \) we define \( Z_d(v) = I_d(v) \) with probability \( p \), or \( Z_d(v) = \text{Ber}_d(v) \) with probability
1 − p. The decisions to chose between \( I_d \) or \( \text{Ber}_d \) are made independently between the vertices. Clearly, \( Z_d \) is an FIID percolation on \( T_d \). Since \( \text{corr}(I_d) = 0 \) and \( \text{corr}(\text{Ber}_d) = 1 \), it is easy to calculate that

\[
\text{den}(Z_d) = p \text{den}(I_d) + (1 - p) \text{den}(\text{Ber}_d), \quad \text{and} \\
\text{corr}(Z_d) = \frac{(1 - p)^2 \text{den}(\text{Ber}_d)^2 + 2p(1 - p) \text{den}(I_d) \text{den}(\text{Ber}_d)}{\text{den}(Z_d)^2}.
\]

In the following calculation we ignore the \( 1 - o(1) \) factor from \( \text{den}(I_d) \) for tidiness. It does not affect the conclusion and introduces the \( o(1) \) term in the statement of the lower bound from Theorem 1.2.4. Continuing from the above we see that

\[
\text{den}(Z_d) = \frac{p}{1 - p} \left( \frac{1}{\sqrt{1 - c}} - 1 \right). 
\]

Consequently, the density of \( Z_d \) is \( \frac{p}{\sqrt{1 - c}} \cdot \frac{\log d}{d} \). By letting \( p \to 1 \) we deduce the lower bound.

### The case for correlation near 1

If \( \text{corr}(Y_d) \to 1 \) then we cannot conclude from our entropy inequality that \( \text{den}(Y_d) \to 0 \). This in turn means that we cannot use Taylor expansions to analyze the entropy functional since that requires \( \text{den}(Y_d) \to 0 \) in order to control the error terms. However, from a result of Backhausz, Szegedy and Virág (see [3] Theorem 3.1) it is possible to conclude that \( \text{den}(Y_d) \to 0 \) if the correlation satisfies \( \text{corr}(Y_d) = 1 - \delta_d \) where \( \delta_d \to 0 \) at a rate such that \( \sqrt{d} \delta_d \to \infty \). Indeed, their correlation bound implies that for any FIID percolation process \( Y_d \) on \( T_d \) one has

\[
\left| \frac{\mathbb{P}[Y_d(\circ) = 1, Y_d(\circ') = 1] - \text{den}(Y_d)^2}{\text{den}(Y_d)(1 - \text{den}(Y_d))} \right| \leq \frac{10}{\sqrt{d}}.
\]

It follows directly from the bound above that

\[
\text{den}(Y_d) \leq \frac{10}{\sqrt{d} \left| \text{corr}(Y_d) - 1 \right|}.
\]  \( (4.2) \)

Now if \( \text{den}(Y_d) \to 0 \) then we can analyze the terms of entropy functional about \( \alpha = \text{den}(Y_d) \) and proceed as before. We will be able to conclude that for the entropy functional to remain non-negative the density of \( Y_d \) is bounded by \( O((\log d)/d \Psi(\text{corr}(Y_d))) \). If \( \text{corr}(Y_d) = 1 - \delta_d \) then \( \Psi(1 - \delta_d) = \delta_d^2 + O(\delta_d^3) \). Hence, if \( \text{corr}(Y_d) = 1 - \delta_d \) with \( \delta_d \to 0 \) and \( \sqrt{d} \delta_d \to \infty \) then

\[
\text{den}(Y_d) = O \left( \frac{\log d}{\delta_d^2 d} \right) \text{ as } d \to \infty.
\]

This bound beats the bound in (4.2) only if \( \delta_d \sqrt{d}/(\log d) \to \infty \).
4.2 Percolation with finite clusters – an asymptotically optimal bound

From Theorem 1.2.4 we conclude that \( \limsup_{d \to \infty} \frac{\text{den}(Y_d)}{(\log d)/d} \leq 2 \). In order to prove Theorem 1.2.2 we also use the entropy inequality, but apply it to many copies of a given percolation process \( Y_d \), where the copies are coupled in a particular way. It turns out that the entropy inequality provides better bounds when it is applied to many copies of \( Y_d \). This important idea is borrowed from statistical physics where it was initially observed through the overlap structure of the Sherrington-Kirkpatrick model [42]. Since then it has been successfully used to study other combinatorial and statistical physics models as well (see [39]).

To derive bounds on \( \text{den}(Y_d) \) from the entropy inequality applied to several copies of \( Y_d \), we need to find an upper bound to the general entropy functional in terms of the density of each colour class. This bound follows from an important convexity argument. We begin with the derivation of this upper bound.

Upper bound for the entropy functional

We provide an upper bound to the entropy functional in terms of the probability distribution \( \pi \) associated with an edge profile \((P, \pi)\). The subadditivity of entropy implies that for any edge profile \((P, \pi)\), the entropy \( H(P) \leq 2H(\pi) \), which in turn implies that \((d/2)H(P) - (d-1)H(\pi) \leq H(\pi)\). However, this is useless towards analyzing the entropy inequality since \( H(\pi) \geq 0 \). In fact, this upper bound is sharp if and only if \( P(i, j) = \pi(i)\pi(j) \) for all pairs \((i, j)\). However, any high density percolation process \( Y_d \) whose correlation satisfies the hypothesis of Theorem 1.2.2 cannot have an edge profile with this property. In order to derive a suitable upper bound we have to bound the entropy functional subject to constraints on the edge profile induced by small correlations.

We derive such a bound from convexity arguments that take into account the constraints put forth by small correlations. Although our bound is not sharp, it appears to become sharp as the degree \( d \to \infty \). This is the reason as to why we get an asymptotically optimal bound on the density as \( d \to \infty \). When the correlations do not converge to zero the bound becomes far from sharp and we cannot get bounds on the density that improve upon Theorem 1.2.4 in a general setting.

Before proceeding with the calculations we introduce some notation and provide the setup. Let \((P, \pi)\) be an edge profile such that the entries of \( P \) are indexed by \( \chi^2 \), where \( \#\chi = q + 1 \geq 2 \). Suppose that there is a subset \( \Lambda \subset \chi^2 \) (which represents the pairs \((i, j)\) where \( P(i, j) \) is very small) with the following properties.

- \( \Lambda \) is symmetric, that is, if \((i, j) \in \Lambda \) then \((j, i) \in \Lambda \).
- There exists an element \( 0 \in \chi \) such that \( \Lambda \) does not contain any of the pairs \((0, j)\) for every \( j \in \chi \).
There are positive constants $K \leq 1/(eq)$ and $J$ such that $P(i, j) \leq K$ for all $(i, j) \in \Lambda$ and $\pi(i) \leq J$ for all $i \neq 0$.

Set $\Lambda_i = \{j \in \chi : (i, j) \in \Lambda\}$, and define the quantities

$$
\pi(\Lambda_i) = \sum_{j \in \Lambda_i} \pi(j) \\
P(\Lambda_i) = \sum_{j \in \Lambda_i} P(i, j) \\
\pi^2(\Lambda) = \sum_i \pi(i)\pi(\Lambda_i) = \sum_{(i, j) \in \Lambda} \pi(i)\pi(j).
$$

In our application both $K$ and $J$ with converge to zero with $d$ while $\chi$ stays fixed, so the bound stipulated on $K$ will be satisfied for all large $d$.

**Lemma 4.2.1.** With the setup as above the following inequality holds:

$$
\frac{d}{2}H(P) - (d - 1)H(\pi) \leq H(\pi) - \frac{d}{2}\pi^2(\Lambda) + O_q\left(dK + dK \log \left(\frac{J^2}{K}\right)\right).
$$

The constant in the $O_q(\cdot)$ term is at most $2q^2$.

**Proof.** We use the concavity of the function $h(x) = -x \log x$, which can be verified from taking second derivatives. Using the identity $h(xy) = xh(y) + yh(x)$ we get

$$
H(P) = \sum_{(i,j)} h\left(\pi(i) \cdot \frac{P(i,j)}{\pi(i)}\right) = \sum_{(i,j)} \pi(i)h\left(\frac{P(i,j)}{\pi(i)}\right) + H(\pi).
$$

To bound $\sum_i \pi(i)h\left(\frac{P(i,j)}{\pi(i)}\right)$ we consider the summands for $i \in \Lambda_j$ and $i \notin \Lambda_j$ separately. Jensen’s inequality applied to $h(x)$ implies that

$$
\sum_{i \in \Lambda_j} \pi(i)h\left(\frac{P(i,j)}{\pi(i)}\right) \leq \pi(\Lambda_j)h\left(\frac{P(\Lambda_j)}{\pi(\Lambda_j)}\right)
$$

and

$$
\sum_{i \notin \Lambda_j} \pi(i)h\left(\frac{P(i,j)}{\pi(i)}\right) \leq (1 - \pi(\Lambda_j))h\left(\frac{\pi(j) - P(\Lambda_j)}{1 - \pi(\Lambda_j)}\right).
$$

Since $\pi(\Lambda_0) = P(\Lambda_0) = 0$, the first sum from (4.3) is empty for $j = 0$ and the second equals $h(\pi(0))$. We analyze the bound derived for the second sum from (4.3) for $j \neq 0$.

$$
(1 - \pi(\Lambda_j))h\left(\frac{\pi(j) - P(\Lambda_j)}{1 - \pi(\Lambda_j)}\right) = - (\pi(j) - P(\Lambda_j))\log \left(\frac{\pi(j) - P(\Lambda_j)}{1 - \pi(\Lambda_j)}\right)
$$

$$
= h(\pi(j) - P(\Lambda_j)) + (\pi(j) - P(\Lambda_j)) \log(1 - \pi(\Lambda_j)).
$$
Now, \( h(\pi(j) - P(\Lambda_j)) = h(\pi(j) \cdot (1 - \frac{P(\Lambda_j)}{\pi(j)})) = (1 - \frac{P(\Lambda_j)}{\pi(j)})h(\pi(j)) + \pi(j)h(1 - \frac{P(\Lambda_j)}{\pi(j)}) \). The first of these two terms, \((1 - \frac{P(\Lambda_j)}{\pi(j)})h(\pi(j))\), simplifies to \(h(\pi(j)) + P(\Lambda_j)\log(\pi(j))\). The second term satisfies \(\pi(j)h(1 - \frac{P(\Lambda_j)}{\pi(j)}) \leq P(\Lambda_j)\) because \(h(1 - x) \leq x\).

Also, \(\log(1 - \pi(\Lambda_j)) \leq -\pi(\Lambda_j)\) because \(\log(1 - x) \leq -x\) for \(0 \leq x \leq 1\), and as a result, \((\pi(j) - P(\Lambda_j))\log(1 - \pi(\Lambda_j)) \leq -\pi(j)\pi(\Lambda_j) + P(\Lambda_j)\). So we conclude that

\[
(1 - \pi(\Lambda_j)h\left(\frac{\pi(j) - P(\Lambda_j)}{1 - \pi(\Lambda_j)}\right) \leq h(\pi(j)) - \pi(j)\pi(\Lambda_j) + P(\Lambda_j)\log(\pi(j)) + 2P(\Lambda_j).
\]

Since \(\pi(\Lambda_j)h\left(\frac{P(\Lambda_j)}{\pi(\Lambda_j)}\right) = P(\Lambda_j)\log\left(\frac{\pi(\Lambda_j)}{P(\Lambda_j)}\right)\), we deduce from (4.3) that if \(j \neq 0\) then

\[
\sum_{i} \pi(i)h\left(\frac{P(i, j)}{\pi(i)}\right) \leq h(\pi(j)) - \pi(j)\pi(\Lambda_j) + 2P(\Lambda_j) + P(\Lambda_j)\log\left(\frac{\pi(j)\pi(\Lambda_j)}{P(\Lambda_j)}\right).
\] (4.4)

The last two terms of (4.4) contribute to the error term involving \(J\) and \(K\). From the hypotheses we have \(\pi(j) \leq J\) and \(\pi(\Lambda_j) \leq \#\Lambda \leq qJ\), so that \(\pi(j)\pi(\Lambda_j) \leq qJ^2\). The function \(h(x)\) is also increasing for \(x \leq 1/e\), and since \(P(\Lambda_j) \leq qK \leq 1/e\), we see that

\[
P(\Lambda_j)\log\left(\frac{\pi(j)\pi(\Lambda_j)}{P(\Lambda_j)}\right) = h(\pi(\Lambda_j)) + P(\Lambda_j)\log\left(\pi(j)\pi(\Lambda_j)\right) \leq h(qK) + qK\log(qJ^2) = qK\log\left(\frac{J^2}{K}\right).
\]

This implies from (4.4) that for \(j \neq 0\), \(\sum_{i} \pi(i)h\left(\frac{P(i, j)}{\pi(i)}\right) \leq h(\pi(0)) + \sum_{j \neq 0} [h(\pi(j)) - \pi(j)\pi(\Lambda_j)] + O_q\left(K + K\log\left(\frac{J^2}{K}\right)\right)\)

\[
= H(\pi) - \pi^2(\Lambda) + O_q\left(K + K\log\left(\frac{J^2}{K}\right)\right).
\]

Consequently, \(H(P) \leq 2H(\pi) - \pi^2(\Lambda) + O_q\left(K + K\log(J^2/K)\right)\), and the inequality from the statement of the lemma follows immediately.

\[\blacksquare\]

4.2.1 Coupling of percolation processes and the stability variable

In this section we set up some of the tools and terminology that we will use to prove Theorem 1.2.2 in the following section.
Many overlapping processes from a given percolation process

To prove Theorem 1.2.2, we construct a coupled sequence of exchangeable percolation processes $Y^0_d, Y^1_d, \ldots$ from a given percolation process $Y_d$. We then apply the entropy inequality to the FIID process $(Y^1_d, \ldots, Y^k_d)$ to derive bounds on $\text{den}(Y_d)$ that improve upon Theorem 1.2.4 by taking $k$ to be arbitrarily large. The process $Y^i_d$ is defined by varying the random labels that serve as input to the factor of $Y_d$. The labels are varied by first generating a random subset of vertices of $T_d$ via a Bernoulli percolation and then repeatedly re-randomizing the labels over this subset to produce the random labelling for the $Y^i_d$. The processes are coupled since their inputs agree on the complement of this random subset. The details follow.

Let $X^0_d, X^1_d, X^2_d, \ldots$ be independent random labellings of $T_d$. Fix a parameter $0 \leq p \leq 1$, and let $\text{Ber}_d$ be a Bernoulli percolation on $T_d$ having density $p$. If $f_{Y_d}$ is the factor associated to $Y_d$ then the FIID percolation process $Y^i_d$ has factor $f_{Y^i_d}$ but the random labelling used for it, say $W = (W^i(v), v \in V(T_d))$, is defined by $W^i(v) = X^i_d(v)$ if $\text{Ber}_d(v) = 1$ and $W^i(v) = X^0(v)$ if $\text{Ber}_d(v) = 0$.

For $k \geq 1$ consider the FIID process $(Y^1_d, \ldots, Y^k_d)$, which takes values in $\chi^{V(T_d)}$ where $\chi = \{0, 1\}^k$. By identifying $\{0, 1\}^k$ with subsets of $\{1, \ldots, k\}$ the edge profile of this process can be described as follows. For subsets $S, T \subset \{1, \ldots, k\},$

$$
\pi(S) = \mathbb{P} \left[ Y^i_d(\emptyset) = 1 \text{ for } i \in S \text{ and } Y^i_d(\emptyset) = 0 \text{ for } i \notin S \right]
$$

$$
P(S, T) = \mathbb{P} \left[ Y^i_d(\emptyset) = 1, Y^j_d(\emptyset') = 1 \text{ for } i \in S, j \in T \\
\text{ and } Y^i_d(\emptyset) = 0, Y^j_d(\emptyset') = 0 \text{ for } i \notin S, j \notin T \right]
$$

Observe that $\pi(\{i\}) = \text{den}(Y_d)$ for every $i$, and if $S \cap T \neq \emptyset$ then

$$
P(S, T) \leq \mathbb{P} \left[ Y^i_d(\emptyset) = 1, Y^j_d(\emptyset') = 1 \text{ for } i \in S \cap T \right] \leq \text{corr}(Y_d) \text{den}(Y_d)^2. \tag{4.5}
$$

We describe some important properties of the coupled processes that will be used to complete the proof of Theorem 1.2.2. Since the statement of the theorem is about the scaled density $\pi(\{1\}) \frac{d}{\log d}$, we find a probabilistic interpretation for these scaled quantities. We use the coupling to find such an interpretation for the ratio $\mathbb{P} \left[ Y^i_d(\emptyset) = 1 \text{ for } i \in S \right] / \text{den}(Y_d)$. Define the normalized density $\alpha(S)$ for any non empty and finite set $S \subset \{1, 2, 3, \ldots\}$ by

$$
\alpha(S) \frac{\log d}{d} = \mathbb{P} \left[ Y^i_d(\emptyset) = 1 \text{ for all } i \in S \right]. \tag{4.6}
$$

Since the coupled percolation processes are exchangeable, $\alpha(S) = \alpha(\{1, \ldots, \#S\})$ for all $S$. For convenience, we write $\alpha_{i,d} = \alpha(\{1, \ldots, i\})$ and call these the intersection densities of the processes $Y^1_d, Y^2_d, \ldots$ (although, these densities are normalized by the factor of $(\log d)/d$). Note that $\alpha_{i,d}$ also depends on the parameter $p$ but we ignore writing this explicitly. Later in
the proof we will set \( p \) to a value according to \( d \). When we apply the entropy inequality to \((Y_d^1, \ldots, Y_d^k)\) we will get an expression in terms of the intersection densities of the \( k \) percolation processes. In order to analyze this expression we have to realize the ratios \( \alpha_{i,d}/\alpha_{1,d} \) as the moments of a random variable \( Q_{d,p} \). We define \( Q_{d,p} \) in the following.

### The stability variable

The random variable \( Q_{d,p} \) is defined on a new probability space, which is obtained from the original probability space generated by the random labels \( X_d \) by essentially restricting to the support of the factor \( f_{Y_d} \). Formally, the new sample space is the set \( \{Y_d^0(\circ) \equiv 1\} \) considered as a subset of the joint sample space of \( X_d^0, X_d^1, \ldots, \) and \( \text{Ber}_p^d \). The new \( \sigma \)-algebra is the restriction of the \( \sigma \)-algebra generated by \( X_d^0, X_d^1, \ldots, \) and \( \text{Ber}_p^d \) to \( \{Y_d^0(\circ) \equiv 1\} \). The new expectation operator \( E^* \) is defined by

\[
E^*[U] = \frac{E[Y_d^0(\circ)U]}{E[Y_d^0(\circ)]}
\]

for any measurable random variable \( U \) defined on \( \{Y_d^0(\circ) \equiv 1\} \). In the following, we write \( Y_i \) to stand for \( Y_d^i \).

If \( F \) is a \( \sigma \)-algebra containing the \( \sigma \)-algebra generated by \( Y_d^0(\circ) \), then for any random variable \( U \) defined on the original probability space we have

\[
E^*[U | F] = E[U | F].
\]

This is to be interpreted by restricting \( F \) to \( \{Y_d^0(\circ) \equiv 1\} \) on the left and the random variable \( E[U | F] \) to \( \{Y_d^0(\circ) \equiv 1\} \) on the right. To prove this suppose that \( Z \) is a \( F \)-measurable random variable. Then,

\[
E^*[Z E[U | F]] = \frac{E[Y_d^0(\circ)Z E[U | F]]}{E[Y_d^0(\circ)]}
= \frac{E[E[Y_d^0(\circ)ZU | F]]}{E[Y_d^0(\circ)]} \quad \text{(since } Y_d^0(\circ) \text{ and } Z \text{ are } F \text{-measurable)}
= \frac{E[Y_d^0(\circ)ZU]}{E[Y_d^0(\circ)]}
= E^* [ZU].
\]

Define a \([0, 1]\)-valued random variable \( Q_{d,p} = Q_d(\text{Ber}_d^p, X_d^0) \), which we denote the *stability*, on the restricted probability space as follows. Set

\[
Q_{d,p} = E^* [Y_1(\circ) | X_d^0, \text{Ber}_d^p] = E[Y_1(\circ) | X_d^0, \text{Ber}_d^p].
\]

In an intuitive sense the stability is the conditional probability, given the root is included in the percolation process, that it remains to be included after re-randomizing the labels.
The ratio of the intersection densities are moments of the stability: \( \mathbb{E}^* \left[ Q_{d,p}^{i-1} \right] = \frac{\alpha_{i,d}}{\alpha_{1,d}} \) for \( i \geq 1 \). Indeed, note that \( Q_{d,p} = \mathbb{E} \left[ Y_j(o) \mid X_d^0, Ber_d^p \right] \) for every \( j \). Hence,

\[
\mathbb{E}^* \left[ Q_{d,p}^{i-1} \right] = \frac{\mathbb{E} \left[ Y_0(o) \left( \mathbb{E} \left[ Y_1(o) \mid X_d^0, Ber_d^p \right] \right)^{i-1} \right]}{\mathbb{E} \left[ Y_0(o) \right]}
= \frac{\mathbb{E} \left[ Y_0(o) \left( \prod_{j=1}^{i-1} \mathbb{E} \left[ Y_j(o) \mid X_d^0, Ber_d^p \right] \right) \right]}{\mathbb{E} \left[ Y_0(o) \right]}.
\]

Observe that the \( Y_j \) are independent of each other conditioned on \( (X_d^0, Ber_d^p) \). Hence,

\[
\prod_{j=1}^{i-1} \mathbb{E} \left[ Y_j(o) \mid X_d^0, Ber_d^p \right] = \mathbb{E} \left[ \prod_{j=1}^{i-1} Y_j(o) \mid X_d^0, Ber_d^p \right].
\]

Since \( Y_0 \) is measurable with respect to \( X_d^0 \) we see that

\[
\mathbb{E} \left[ Y_0(o) \left( \prod_{j=1}^{i-1} \mathbb{E} \left[ Y_j(o) \mid X_d^0, Ber_d^p \right] \right) \right] = \mathbb{E} \left[ Y_0(o) \prod_{j=1}^{i-1} Y_j(o) \right]
= \mathbb{P} \left[ Y_j(o) = 1, 0 \leq j \leq i - 1 \right].
\]

Thus, \( \mathbb{E}^* \left[ Q_{d,p}^{i-1} \right] = \frac{\mathbb{P}[Y_j(o) = 1, 0 \leq j \leq i - 1]}{\text{den}(Y_d)} = \frac{\alpha_{i,d}}{\alpha_{1,d}}. \)

We will require the following continuity lemma about the stability.

**Lemma 4.2.2.** For each \( u \geq 0 \), the moment \( \mathbb{E}^* \left[ Q_{d,p}^u \right] \) is a continuous function of \( p \). When \( p = 0 \), \( \mathbb{E}^* \left[ Q_{d,p}^u \right] = 1 \), and when \( p = 1 \), \( \mathbb{E}^* \left[ Q_{d,p}^u \right] = \text{den}(Y_d)^u \).

**Proof.** Recall we had assumed that the factor \( f_{Y_d} \) is a block factor with some radius \( r = r_d < \infty \). The parameter \( p \) enters into \( \mathbb{E}^* \left[ Q_{d,p}^u \right] \) through the random finite set \( \{ v \in V(\mathbb{T}_{d,r}) : Ber_d^p(v) = 1 \} \). If \( S \subset V(\mathbb{T}_{d,r}) \) then

\[
\mathbb{P} \left[ Ber_d^p(v) = 1 \text{ for } v \in S \text{ and } Ber_d^p(v) = 0 \text{ for } v \in V(\mathbb{T}_{d,r}) \setminus S \right] = p^{\#S} (1 - p)^{\#(V(\mathbb{T}_{d,r}) \setminus S)}.
\]

This is a polynomial in \( p \), and by conditioning on the output of \( Ber_d^p \) restricted to \( V(\mathbb{T}_{d,r}) \), it follows that \( \mathbb{E}^* \left[ Q_{d,p}^u \right] \) can be expressed as a convex combination of terms that are free of \( p \) with coefficients given by these probabilities. Thus, \( \mathbb{E}^* \left[ Q_{d,p}^u \right] \) is a polynomial in \( p \) as well.

When \( p = 0 \) the process \( Ber_d^0 \equiv 0 \), and \( Y^0 = Y^1 \). Therefore, conditioning on \( X_d^0 \) and restricting to \( \{ Y^0_d(o) \equiv 1 \} \) makes \( Q_{d,0} \equiv 1 \). When \( p = 1 \) the process \( Ber_d^1 \equiv 1 \). So \( Y^1 \) becomes independent of \( X_d^0 \). Then the conditioning has no effect and \( Q_{d,1} = \mathbb{E} \left[ Y^1(o) \right] = \text{den}(Y_d). \)

**4.2.2 Completing the proof**

Let \( \{ Y_d \} \) be a sequence of FIID percolation processes with \( Y_d \in \{ 0, 1 \}^{V(\mathbb{T}_d)} \). For the sake of a contradiction we assume that \( \text{corr}(Y_d) \to 0 \) while \( \text{den}(Y_d) \geq \alpha \frac{\log d}{d} \) for some \( \alpha > 1 \). (Technically...
speaking, we need to move to a subsequence in $d$, but we can assume WLOG that this is the case.) Note that it follows from Theorem 1.2.4 that $\alpha \leq 2$.

For fixed values of $k \geq 1$ and $0 \leq p \leq 1$ consider the coupling $(\mathbf{Y}_d^1, \ldots, \mathbf{Y}_d^k)$ described in Section 4.2.1, and denote the edge profile of this coupled process as $(P, \pi)$ where $P = \{P(S,T) | S,T \subset \{1, \ldots, k\}$ Due to Theorem 1.2.4 we can assume by taking $d$ large enough that $	ext{den}(\mathbf{Y}_d) \leq 10^{\log_d d}$, say. Then for any subset $S \neq \emptyset$, the quantity $\pi(S) \leq \pi(\{1\}) \leq 10^{\log_d d}$. Writing $\text{corr}(\mathbf{Y}_d) = \epsilon_d \to 0$, we have from (4.5) that $P(S,T) \leq \epsilon_d \log_d d = o(2^d \frac{\log^d d}{d})$.

We apply the upper bound from Lemma 4.2.1 to the entropy functional associated to the edge profile $(P, \pi)$. We take $\chi = \{0,1\}^k$, the role of the element 0 is taken by the empty set, and $\Lambda = \{(S,T) : S \cap T \neq \emptyset\}$. For all large $d$ we may take $K = 100\epsilon_d \frac{\log^d d}{d^2}$, which follows from (4.5) and the assumption that $\text{den}(\mathbf{Y}_d) \leq 10^{\log_d d}$. Also, we may take $J = 10^{\log_d d}$.

Applying Lemma 4.2.1 we conclude that

$$
\frac{d}{2} H(P) - (d - 1)H(\pi) \leq H(\pi) - \frac{d}{2} \pi^2(\Lambda) + O_k\left(\epsilon_d + h(\epsilon_d) \log_d d\right).
$$

Note that $\epsilon_d + h(\epsilon_d) \to 0$ if $\epsilon_d \to 0$. Therefore, the big O term is of order $o\left(\frac{\log^d d}{d}\right)$ as $d \to \infty$.

It remains to find the contribution from the term $H(\pi) - (d/2)\pi^2(\Lambda)$, which will be of order $\frac{\log^2 d}{d}$. For $S \neq \emptyset$ we define $\beta(S)$ by writing $\pi(S) = \beta(S) \log_d d$. We ignore explicitly writing the dependence of $\beta(S)$ on $d$. Due to the assumption on $\text{den}(\mathbf{Y}_d)$ we know that $\beta(S) \leq 10$ for all large $d$. With this notation we have

$$
\pi^2(\Lambda) = \sum_{(S,T): S \cap T \neq \emptyset} \beta(S) \beta(T) \frac{\log^2 d}{d^2}.
$$

Similarly, for $S \neq \emptyset$, $h(\pi(S)) \leq \beta(S) \log_d d + h(\beta(S)) \frac{\log d}{d}$ and $h(\beta(S)) \leq 1$ since $h(x) \leq 1$ for all $x \geq 0$. On the other hand, $\pi(\emptyset) = 1 - O_k\left(\frac{\log d}{d}\right)$ because $\pi(S) \leq 10 \frac{\log d}{d}$ for every non empty $S$. Hence, $h(\pi(\emptyset)) = O_k\left(\frac{\log d}{d}\right)$ as $h(1 - x) \leq x$. Therefore,

$$
H(\pi) - \frac{d}{2} \pi^2(\Lambda) \leq \left(\sum_{S \neq \emptyset} \beta(S) - \frac{1}{2} \sum_{(S,T): S \cap T \neq \emptyset} \beta(S) \beta(T)\right) \frac{\log^2 d}{d} + O_k\left(\frac{\log d}{d}\right).
$$

This provides the following critical upper bound on the entropy functional.

$$
\frac{d}{2} H(P) - (d - 1)H(\pi) \leq \left(\sum_{S \neq \emptyset} \beta(S) - \frac{1}{2} \sum_{(S,T): S \cap T \neq \emptyset} \beta(S) \beta(T)\right) \frac{\log^2 d}{d} + o_k\left(\frac{\log^2 d}{d}\right). \quad (4.7)
$$

From the non-negativity of the entropy functional in Theorem 3.2.2 we deduce that the right hand side of (4.7) must be non-negative for all large $d$. In particular, the coefficient of the leading order term is non-negative for all large $d$. Note also that the quantities $\beta(S)$ depend on
the coupling parameter $p$. However, our estimates so far have been independent of the value of $p$. As such, we have that

$$\liminf_{d \to \infty} \inf_{p \in [0,1]} \sum_{S \neq \emptyset} \beta(S) - \frac{1}{2} \sum_{(S,T): S \cap T \neq \emptyset} \beta(S) \beta(T) \geq 0. \quad (4.8)$$

For the analysis of (4.8) it is convenient to parametrize $\pi$ through the intersection densities of the coupled process, as is defined in (4.6), and then express (4.8) in terms of those densities. The principle of inclusion and exclusion provides the following relation between the $\beta(S)$ and the intersection densities $\alpha(S)$ defined in (4.6).

$$\alpha(S) = \sum_{T: S \subset T} \beta(T) \quad (4.9)$$

$$\beta(S) = \sum_{T: S \subset T} (-1)^{\#(T \setminus S)} \alpha(T). \quad (4.10)$$

We now show that the quantity in (4.8) can be rewritten as

$$\sum_{S \neq \emptyset} \beta(S) - \frac{1}{2} \sum_{(S,T): S \cap T \neq \emptyset} \beta(S) \beta(T) = \sum_{i} (-1)^{i-1} \left( \binom{k}{i} \alpha_{i,d} - \frac{1}{2} \alpha_{i,d}^2 \right). \quad (4.11)$$

The term $\sum_{S \neq \emptyset} \beta(S)$ equals $\sum_{i} (-1)^{i-1} \left( \binom{k}{i} \alpha_{i,d} \right)$ because both terms are equal to the normalized density $P[Y_d(\circ) = 1$ for some $i \leq k] \cdot \frac{d}{\log d}$. The relations (4.9) and (4.10) imply that the quantity $\alpha(S)^2 = \sum_{(T_1,T_2): S \subset T_1 \cap T_2} \beta(T_1) \beta(T_2)$. Thus,

$$\sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \alpha_{i,d}^2 = - \sum_{S \neq \emptyset} (-1)^{\#S} \alpha(S)^2$$

$$= - \sum_{S \neq \emptyset} (-1)^{\#S} \sum_{(T_1,T_2): S \subset T_1 \cap T_2} \beta(T_1) \beta(T_2)$$

$$= - \sum_{(T_1,T_2)} \beta(T_1) \beta(T_2) \sum_{S: S \subset T_1 \cap T_2, S \neq \emptyset} (-1)^{\#S}$$

$$= - \sum_{(T_1,T_2): T_1 \cap T_2 \neq \emptyset} \beta(T_1) \beta(T_2) \sum_{i=1}^{\#T_1 \cap T_2} (-1)^{i} \left( \binom{\#T_1 \cap T_2}{i} \right).$$

Recall the binomial identity $\sum_{i=1}^{t} (-1)^{i} \binom{t}{i} = (1 - 1)^{t} - 1$ for any integer $t \geq 1$. Using this to simplify the last equation above implies

$$\sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \alpha_{i,d}^2 = \sum_{(S,T): S \cap T \neq \emptyset} \beta(S) \beta(T), \text{ as required.}$$
We now express (4.8) in terms of the stability variable introduced in Section 4.2.1. Let $Q_{d,p}$ denote the stability of $Y_d$. Recall that $\alpha_{i,d} = \alpha_{1,d}E^* \left[ Q_{d,p}^{i-1} \right]$ for all $i \geq 1$. Henceforth, we denote the operator $E^* \left[ \cdot \right]$ by $E \left[ \cdot \right]$. Let $R_{d,p}$ denote an independent copy of $Q_{d,p}$; then $\alpha_{i,d}^2 = \alpha_{1,d}^2E \left[ (Q_{d,p}R_{d,p})^{i-1} \right]$. Using the identity $\sum_{i=1}^k (-1)^{i-1}(k)x^{i-1} = \frac{1-(1-x)^k}{x}$ for $0 \leq x \leq 1$, we translate the inequality from (4.8) via the identity (4.11) into

$$\lim inf_{d \to \infty} \inf_{p \in [0,1]} \alpha_{1,d}E \left[ \frac{1 - (1 - Q_{d,p})^k}{Q_{d,p}} \right] - \frac{\alpha_{1,d}^2}{2}E \left[ \frac{1 - (1 - Q_{d,p}R_{d,p})^k}{Q_{d,p}R_{d,p}} \right] \geq 0.$$  

(4.12)

Now we make a choice for the value of the coupling parameter $p$ for each value of $d$. Fix a parameter $u > 0$ and for all large $d$, pick $p = p(d, u)$ in the construction of the coupling so that $E \left[ Q_{d,p} \right] = 1/\alpha$. This can be done due to the continuity of expectations given by Lemma 4.2.2 and the assumption that $\alpha > 1$. We denote $Q_{d,p(d,u)}$ by $Q_d$, ignoring the dependence on $u$ for convenience. As of now it is mysterious as to why we set $p$ in this manner. However, in the following argument we will see that setting $p$ this way is a judicious choice.

Let $s_k(x) = \frac{1-(1-x)^k}{x}$ for $0 \leq x \leq 1$. Since $s_k(x) = 1 + (1-x) + \cdots + (1-x)^{k-1}$, it follows that $s_k(x)$ is continuous, decreasing on $[0,1]$ with maximum value $s_k(0) = k$ and minimum value $s_k(1) = 1$.

Since probability distributions on $[0,1]$ are compact in the weak topology, we choose a subsequence $(Q_{d_i}, R_{d_i})$ such that it converges in the weak limit to $(Q, R)$. The random variables $Q$ and $R$ are independent and identically distributed with values in $[0,1]$. Since $s_k(x)$ is continuous, we have that $E \left[ s_k(Q_{d_i}) \right] \to E \left[ s_k(Q) \right]$ and $E \left[ s_k(Q_{d_i}, R_{d_i}) \right] \to E \left[ s_k(QR) \right]$. Now pass to the subsequence $d_i$ and take limits in (4.12). The quantity $\alpha_{1,d}$ may not have a limit but since we assume, for sake of a contradiction, that $\alpha_{1,d} \geq \alpha > 1$, we deduce from taking limits in the $d_i$ that for all $k \geq 1$

$$E \left[ s_k(Q) \right] \geq \frac{\alpha}{2}E \left[ s_k(QR) \right].$$  

(4.13)

We want to take the limit in $k$ as well but we must be careful. The function $s_k(x)$ monotonically converges to $1/x$ for $x \in [0,1]$. So by taking limits in $k$ in (4.13) and using the independence of $Q$ and $R$ we conclude that $2E \left[ 1/Q \right] \geq \alpha E \left[ 1/Q \right]^2$. Of course, we do not know a priori that $E \left[ 1/Q \right] < \infty$. Even if it were, we can only conclude that $\alpha \leq 2/E \left[ 1/Q \right]$, which leaves us with the seemingly contradictory task of showing that $E \left[ 1/Q \right]$ is large but finite.

To deal with these issues we have to use the fact that we have chosen $p = p(d, u)$ such that $E \left[ Q_{d,p}^u \right] = 1/\alpha$. In particular, as $x \to x^u$ is continuous for $0 \leq x \leq 1$, we see that $E \left[ Q^u \right] = \lim_{u \to \infty} E \left[ Q_{d,p}^u \right] = 1/\alpha$. We have to control the expectation $E \left[ 1/Q \right]$ through our control of $E \left[ Q^u \right]$.

Three cases can arise: $P \left[ Q = 0 \right] > 0$, or $P \left[ Q = 0 \right] = 0$ but $E \left[ 1/Q \right] = \infty$, or $E \left[ 1/Q \right] < \infty$.

**Case 1:** $P \left[ Q = 0 \right] = q > 0$. The majority of the contribution to $E \left[ s_k(Q) \right]$ results from $\{Q = 0\}$. More precisely, $\frac{s_k(x)}{k} \to 1_{x=0}$ as $k \to \infty$, and $\frac{s_k(x)}{k} \in [0,1]$ for all $k$ and $x \in [0,1]$. From the
bounded convergence theorem we deduce that $E[s_k(Q)/k] \to P[Q = 0]$ and $E[s_k(QR)/k] \to P[QR = 0]$ as $k \to \infty$. The latter probability is $2q - q^2$ due to $Q$ and $R$ being independent and identically distributed. By dividing the inequality in (4.13) by $k$ and taking a limit we conclude that

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

Now, since for $x \in [0, 1]$, we have that $1_{x=0} \leq 1 - x^\alpha$. It follows from here that $q \leq 1 - E[Q^u] = 1 - 1/\alpha$. Hence,  

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

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

**Case 2:** $P[Q = 0] = 0$ but $E\left[\frac{1}{Q}\right] = \infty$. Now, most of the contribution to $E[s_k(Q)]$ occurs when $Q$ is small. Recall that $s_k(x) \not\to 1/x$ as $k \to \infty$. By the monotone convergence theorem, $E[s_k(Q)] \to \infty$ as $k \to \infty$.

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

$$E[s_k(Q)] = E[s_{k \leq \epsilon}(Q)] + E[s_{k > \epsilon}(Q)] \leq E[s_{k \leq \epsilon}(Q)] + \epsilon^{-1}. \quad (4.14)$$

Therefore, $E[s_{k \leq \epsilon}(Q)] \to \infty$ with $k$ since $E[s_k(Q)] \to \infty$.

Observe from the positivity of $s_k$ that

$$E[s_k(QR)] \geq E[s_k(QR); Q \leq \epsilon, R > \epsilon] + E[s_k(QR); Q > \epsilon, R \leq \epsilon].$$

The latter two terms are equal by symmetry, so $E[s_k(QR)] \geq 2E[s_k(QR); Q \leq \epsilon, R > \epsilon]$. The fact that $s_k(x)$ is decreasing in $x$ and $R \leq 1$ imply that $s_k(QR) \geq s_k(Q)$. Together with the independence of $Q$ and $R$ we deduce that

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

Consequently,

$$E[s_k(QR)] \geq 2E[s_{k \leq \epsilon}(Q)]P[Q > \epsilon]. \quad (4.15)$$

The inequality in (4.13) is $\frac{\alpha}{2} \leq \frac{E[s_k(Q)]}{E[s_k(QR)]}$. The bounds from (4.14) and (4.15) imply that

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

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

$$\alpha \leq \frac{1}{P[Q > \epsilon]}.$$
Chapter 4. Factor of IID percolation on regular trees

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

**Case 3:** $\mathbb{E}[1/Q]$ is finite. Since $s_k(x)$ monotonically converges to $1/x$ as $k \to \infty$, we deduce from (4.13) that

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

Since $\mathbb{E}[1/(QR)] = \mathbb{E}[1/Q]^2$, the inequality above becomes $\alpha \leq 2\mathbb{E}[1/Q]^{-1}$. Jensen’s inequality implies that $\mathbb{E}[1/Q]^{-1} \leq \mathbb{E}[Q^u]^{1/u}$. However, $\mathbb{E}[Q^u] = 1/\alpha$ and this means that $\alpha \leq 2^{\frac{1}{\alpha + 1}}$. As the previous two cases led to contradictions we conclude that $\alpha \leq 2^{\frac{1}{\alpha + 1}}$ for all $u > 0$. However, as $u \to 0$ we see that $\alpha \leq 1$. This final contradiction completes the proof of Theorem 1.2.2.

**Proof of Corollary 1.2.3** If $Y$ is an FIID percolation process on $T_d$ that has finite components with probability one, then the components of $Y$ are all finite trees. Thus, the component of the root induces a measure on finite, rooted trees with the root being picked uniformly at random (due to invariance and transitivity). If a tree has $n$ vertices then the expected degree of a uniform random root is $2(n - 1)/n \leq 2$. This implies that the average degree of the root of $T_d$, given that it is included in $Y$, cannot be larger than 2. As such, the hypothesis of Theorem 1.2.2 applies in this situation.

### 4.3 Density bounds for small values of $d$

Although the entropy inequality may be used in principal to get a bound on the maximal density of FIID percolation on $T_d$ with finite clusters for every $d$, these bounds are not the best available for small values of $d$. The best known explicit upper bounds for small values of $d$ ($d \leq 10$) are due to Bau et. al. [5]. For $d = 3$, an unpublished combinatorial construction of Csóka [15] shows that the largest density of FIID percolation on $T_3$ with finite components is $3/4$. We briefly explain Csóka’s construction in this section.

An important tool used in Csóka’s argument is the Mass Transport Principle (MTP) for invariant processes on trees. Suppose $F : V(T_d) \times V(T_d) \times \chi^V(T_d) \to [0, \infty)$ is a measurable function which is diagonally invariant in the sense that $F(u, v, \omega) = F(\gamma u, \gamma v, \gamma \cdot \omega)$ for any $\gamma \in \text{Aut}(T_d)$. If $\Phi \in \chi^V(T_d)$ is an Aut($T_d$) invariant process and $\circ$ is the root of $T_d$ then the MTP states that

$$\sum_{v \in V(T_d)} \mathbb{E}[F(\circ, v, \Phi)] = \sum_{v \in V(T_d)} \mathbb{E}[F(v, \circ, \Phi)].$$  

(4.16)

To see why this holds note that for any vertex $v$ there exists an automorphism $\gamma$ of $T_d$ that swaps $v$ and $\circ$. Consequently, $F(\circ, v, \Phi) = F(v, \circ, \gamma \cdot \Phi)$ and after taking expectations and using the invariance of $\Phi$ we conclude that $\mathbb{E}[F(\circ, v, \Phi)] = \mathbb{E}[F(v, \circ, \Phi)]$. The identity follows from summing over all vertices.

We can deduce from the MTP that any invariant percolation process on $T_d$ with finite components has density at most $d/(2d - 2)$. Indeed, for any such percolation process $Y$
we consider \( F(u, v, Y) = 1_{(u,v)\in E(T_d), Y(u)=1, Y(v)=0} \). Then, \( \sum_{v\in V(T_d)} \mathbb{E} [F(o, v, Y)] = (d - \text{avdeg}(Y))\text{den}(Y) \), and \( \sum_{v\in V(T_d)} \mathbb{E} [F(v, o, \Phi)] \leq d/(1 - \text{den}(Y)) \). From the MTP we deduce that \( \text{den}(Y) \leq d/(2d - \text{avdeg}(Y)) \). Recall from the proof of Corollary \([1.2.3]\) that if \( Y \) has finite components then \( \text{avdeg}(Y) \leq 2 \). As a result, \( d/(2d - \text{avdeg}(Y)) \leq d/(2d - 2) \).

This upper bound is sharp for \( T_2 \), which is a bi-infinite path, because Bernoulli percolation has finite components at any density \( p < 1 \). However, if \( d > 4 \) then the aforementioned result of Bau et al. \([5]\) shows that FIID percolation processes on \( T_d \) can not achieve density arbitrarily close to \( d/(2d - 2) \). These bounds are derived by using more sophisticated counting arguments than MTP or our entropy inequality. For \( d = 4 \) the upper bound of 2/3 for \( T_4 \) can not be ruled out by these counting arguments, but the best known lower bound is 0.6045 due to a construction of Hoppen and Wormald \([29]\). We suspect that the upper bound of 2/3 is optimal.

We briefly explain Csóka’s construction which shows that the bound of 3/4 is optimal for \( d = 3 \). The idea is to construct, as a weak limit of FIID processes, an invariant orientation of the edges of \( T_3 \) such that the out-degree of every vertex is either 1 or 3. The percolation process is taken to be the set of vertices with out-degree 1. Let us call these vertices blue. To see that they have density 3/4 let \( \{o \rightarrow o'\} \) denote the event that the edge \( \{o, o'\} \) is oriented from \( o \) to \( o' \). By invariance, \( P[o \rightarrow o'] = 1/2 \). Also by invariance, \( P[o \rightarrow o' | o \text{ is blue}] = 1/3 \). Therefore, computing \( P[o \rightarrow o'] \) by conditioning on the out-degree of \( o \) we deduce that

\[
\frac{1}{2} = \frac{P[o \text{ is blue]}}{3} + (1 - P[o \text{ is blue}]), \quad \text{which implies that } P[o \text{ is blue}] = \frac{3}{4}.
\]

The components of the blue vertices are infinite because the outgoing edge from any blue vertex points to another blue vertex, and following these outgoing edges gives an infinite blue ray emanating from any blue vertex. However, the average degree of the process is 2. This can be shown via the MTP. Indeed, given any blue vertex send mass 1 along its unique outgoing edge to the other endpoint. If the root \( o \) is blue then an unit mass leaves \( o \) towards the unique blue neighbour at which one of its edges point. Moreover, an unit mass enters \( o \) from each of its remaining blue neighbours. The sum of all these masses, entering and leaving \( o \), is then its number of blue neighbours. By the MTP the expectation of the sum of masses is 2 because the amount of mass leaving is always 1. So the expected number of blue neighbours of \( o \), conditional on \( o \) being blue, is 2 as claimed.

It is known (see \([36] \) Proposition 8.17) that if an invariant percolation process on \( T_d \) has infinite components and average degree 2 then its components have at most 2 topological ends (meaning that each component contains at most two edge disjoint infinite rays). Because of this the components can be made finite through a Bernoulli percolation at any density \( p < 1 \). This allows us to achieve a process of blue vertices with finite clusters and density arbitrarily close to 3/4.

Although Csóka’s construction realizes the optimal percolation process as a weak limit of FIID processes, it is not necessarily true that the limiting process itself is an FIID process. In
fact, there exists invariant, but not FIID, processes on $T_d$ that are weak limits of FIID process (see [27] Theorem 6).

### 4.4 FIID processes on the edges of $T_d$

In this section we prove Theorem 1.2.7 and Theorem 1.2.8. Recall that FIID process over the edges of $T_d$ can be defined analogously to vertex-indexed processes because Aut($T_d$) acts transitively on the edges of $T_d$ as well. See Section 1.2.2 from the introduction.

Recall that an FIID bond percolation process is a FIID process $\phi \in \{0, 1\}^{E(T_d)}$. Its clusters are the connected components of the subgraph induced by the open edges $\{e \in E(T_d): \phi(e) = 1\}$. Its density is $\text{den}(\phi) = P[\phi(e) = 1]$, where $e = \{v, u\}$ is an arbitrary edge.

An important, and perhaps simplest, example of bond percolation processes are perfect matchings: every vertex $v \in V(T_d)$ is incident to exactly one edge $e = \{v, u\}$ that is open in the percolation process. In other words, every vertex $v \in V(T_d)$ has degree 1 in the bond percolation process. Lyons and Nazarov construct FIID perfect matchings on $T_d$ for every $d \geq 3$ [35]. The density of an FIID perfect matching on $T_d$ is $1/d$ and its clusters are isolated edges.

The largest density of an FIID bond percolation process on $T_d$ with finite clusters is straightforward to find in light of the Lyons and Nazarov construction. Using their construction we provide the lower bound that is required for Theorem 1.2.7.

**Proof of Theorem 1.2.7**

**Proof.** To construct a 2-factor on $T_d$ we simply remove $(d - 2)$ disjoint FIID perfect matchings from $T_d$ (if $d = 2$ then there is noting to do). This can be done iteratively. Removing one FIID perfect matching from $T_d$ results in a forest whose clusters are isomorphic to $T_d - 1$. Proceeding iteratively with each cluster until every vertex has degree 2 gives the desired construction. Clearly, the density of an FIID 2-factor is $2/d$.

The clusters of any 2-factor must be bi-infinite paths, that is, isomorphic to $T_2$. These infinite paths are made finite through a Bernoulli bond percolation at any density $p < 1$; the resulting density of the bond percolation process being $(2/d)p$. These processes are an example of a sequence of FIID bond percolation processes with finite clusters that converge weakly to an FIID 2-factor as $p \to 1$.

In order to show that $2/d$ is the optimal density, we simply use the invariance of the process. If $\phi$ is an FIID bond percolation on $T_d$ with finite clusters then the expected number of open edges incident to a vertex $v$ is $d \cdot \text{den}(\phi)$. On the other hand, the cluster of $v$, say $C(v)$, is a random finite tree rooted at $v$. The root is distributed uniformly at random conditional on observing $C(v)$ as an unrooted tree. This follows due to the invariance of $\phi$ and the fact that Aut($T_d$) acts transitively on $T_d$. Thus, if $C(v) = T$ then $v$ is a uniform random vertex of $T$. This implies that $E[\text{deg}_\phi(v) | C(v) = T] = 2(|T| - 1)/|T|$. As $2(|T| - 1)/|T| < 2$, it follows from conditioning of $C(v)$ that $E[\text{deg}_\phi(v)] < 2$. Therefore, $d \cdot \text{den}(\phi) < 2$. 

Finally, the wired spanning forest is an invariant bond percolation process $T_d$ such that the expected degree of the root is 2. The clusters of the wired spanning forest have exactly one topological end almost surely. Thus, they are not isomorphic to $T_2$. Once again, the clusters can be made finite via a Bernoulli bond percolation at any density less than one. For a definition of the wired spanning forest and proofs of these claims see [8] or chapter 10 in [36].

**Proof of Theorem 1.2.8** The proof of Theorem 1.2.8 also uses FIID perfect matchings.

![Figure 4.1: Orientation of a bi-infinite path $L$. Edges of the perfect matching that meet $L$ are in blue. The edges of $L_n$ are red while edges of $L_{n+1}$ are green. The black edge between $v_n$ and $v_{n+1}$ connects $L_n$ to $L_{n+1}$.](image)

**Proof.** As $d \geq 3$, we select an FIID perfect matching of $T_d$ and orient its edges independently and uniformly at random. Then we remove the oriented edges of the matching and are left with a forest whose components are isomorphic to $T_{d-1}$. We continue with this procedure of selecting an FIID perfect matching from the components, orienting those edges at random and then removing them from the graph until we are left with a forest whose components are isomorphic to $T_3$. We will be successful if we can orient just the edges of $T_3$ as an FIID process such that there are no sources or sinks.

Henceforth, we explain how to orient the edges of $T_3$ in such a manner. Begin by finding an FIID perfect matching on $T_3$ and then orient the edges of the matching independently and uniformly at random. The un-oriented edges span a graph whose components are bi-infinite paths. Consider any such path $L$. The oriented perfect matching partitions $L$ into contiguous finite paths $L_n, n \in \mathbb{Z}$, which are characterized by the following properties.

First, $L_n$ is incident to $L_{n+1}$ via an edge $\{v_n, v_{n+1}\}$ in $L$. Second, for every $n$, the edges of the perfect matching that meet $L_n$ are all oriented in ‘parallel’, either all pointing towards $L_n$ or away from it. Finally, the edges of the matching that meet $L_{n+1}$ are oriented in the opposite direction compared to the (common) orientation of those edges of the matching that meet $L_n$. This setup is shown by the blue edges in Figure 4.1. The reason the paths $L_n$ are finite is because the edges of the matching that are incident to $L$ are oriented independently. We can think of the $L_n$s that are pointing towards the matching as the clusters of a Bernoulli percolation on $L$ at density $1/2$.
To complete the orientation we first orient all the edges on the path $L_n$ in the same direction. As there are two possible directions, we choose one at random. This is done independently for each $L_n$ on every un-oriented bi-infinite path $L$. The finiteness of the $L_n$ is crucial to ensure that these orientations can be done as an FIID. Following this, any vertex in the interior of the path $L_n$ has one edge directed towards it and another directed away. This ensures that such a vertex cannot form a source or a sink.

The vertices that can form sources or sinks are those at the endpoints of any path $L_n$. Also, the un-oriented edges that remain are the $\{v_n, v_{n+1}\}$ which connect two contiguous paths $L_n$ and $L_{n+1}$ on a common bi-infinite path $L$. Observe that the endpoints of these edges are precisely the vertices that can form sources and sinks (see the black edge in Figure 4.1). Given such an edge $\{v_n, v_{n+1}\}$ let $u_n$ (resp. $u_{n+1}$) be the neighbour of $v_n$ (resp. $v_{n+1}$) such that the edge $\{v_n, u_n\}$ (resp. $\{v_{n+1}, u_{n+1}\}$) lies in the matching. The edges $\{v_n, u_n\}$ and $\{v_{n+1}, u_{n+1}\}$ have opposite orientations by design (see Figure 4.1). Suppose that $u_n \to v_n$ and $v_{n+1} \to u_{n+1}$. We orient $\{v_n, v_{n+1}\}$ as $v_n \to v_{n+1}$, thus ensuring that both $v_n$ and $v_{n+1}$ have one incoming edge and another outgoing one. This completes the FIID orientation of $\mathbb{T}_3$ and produces no sources or sinks.
Chapter 5

Percolation on Poisson–Galton–Watson trees

Recall from Section 1.2.3 the definition of a FIID process on a PGW tree. Recall also that local algorithms on sparse Erdős-Rényi graphs are projections of FIID processes on PGW trees. In this chapter we will prove Theorem 1.2.9. In Section 5.1 we will prove that \( \lim \sup_{\lambda \to \infty} \frac{\alpha_{\text{perc}}(\text{PGW}_\lambda)}{(\log \lambda)/\lambda} \leq 1 \), and follow up in Section 5.2 by proving that \( \lim \inf_{\lambda \to \infty} \frac{\alpha_{\text{perc}}(\text{PGW}_\lambda)}{(\log \lambda)/\lambda} \geq 1 \). The proof of the upper bound will employ the strategy used for \( d \)-regular trees in Chapter 4. We will highlight the key differences but be brief with parts of the argument that are analogous to the case for \( d \)-regular trees. The lower bound will be proved by using a coupling of PGW trees with \( d \)-regular trees that will enable us to construct FIID independent sets on PGW trees from those defined on \( d \)-regular trees. We will then use the existence of FIID independent sets with asymptotic density \( \frac{\log d}{d} \) on \( d \)-regular trees to deduce the existence of such independent sets on PGW trees.

5.1 An asymptotically optimal upper bound

Suppose that \( Y \) is a FIID percolation process on PGW with finite clusters and factor \( f : \Lambda_r \to \{0, 1\} \). Using the same technique that was used for \( d \)-regular trees, \( Y \) yields a FIID percolation set \( Y_n \) of \( G_n \sim \text{ER}(n, \lambda/n) \) with the following properties. First, there exists a factor \( f_n : \Lambda_r \to \{0, 1\} \) such that if \( X \) is a random labelling of \( G_n \) then \( Y_n(v) = f_n(N_r(G_n, v), v, X(G_n, v, r)) \). Second, the components of the induced subgraph \( G_n[Y_n] = G_n[\{v : Y_n(v) = 1\}] \) are trees, that is, \( G_n[Y_n] \) is a forest. This is because the components of \( Y \) are finite trees and this property is local (so the projection can be made to retain it, perhaps by increasing \( r \)). Finally, we have that \( \mathbb{E} \left[ \frac{|Y_n|}{n} \right] \to \text{den}(Y) \) as \( n \to \infty \).

To prove that \( \lim \sup_{\lambda \to \infty} \frac{\alpha_{\text{perc}}(\text{PGW}_\lambda)}{(\log \lambda)/\lambda} \leq 1 \) we assume to the contrary. Then, as is the case for \( d \)-regular graphs, we can find \( \alpha > 1 \) and a subsequence of \( \lambda \to \infty \), such that for each \( \lambda \) there exist a FIID percolation set \( Y_{n, \lambda} \) of \( G_n \sim \text{ER}(n, \lambda/n) \) with factor \( f_{n, \lambda} : \Lambda_{r, \lambda} \to \{0, 1\} \) and the property that the subgraph induced by \( Y_n \) is a forest. Also, for each such \( \lambda \), we have that

55
$E|Y_{n,\lambda}|/n \geq \alpha \frac{\log \lambda}{\lambda}$ for all sufficiently large $n$. We can assume WLOG that these statements hold for all $\lambda$ and $n$. By setting $E|Y_{n,\lambda}|/n = \alpha_{1,n,\lambda} \frac{\log \lambda}{\lambda}$, we have that $\alpha_{1,n,\lambda} \geq \alpha > 1$.

At this point if we follow the previous approach we would introduce correlated copies of $Y_{n,\lambda}$ and show that their intersection densities must satisfy an inequality of the form (4.12), which we would derive from a structure theorem about Erdős-Rényi graphs à la Theorem 3.2.2. It turns out that the same structure theorem holds for Erdős-Rényi graphs and it yields an upper bound on $\alpha_{1,n,\lambda}$ different for Erdős-Rényi graphs. In the previous approach a contradiction was derived from inequality (4.12) based on modifying the random labelling. However, for Erdős-Rényi graphs it is possible that the factor does not use the labels at all, but only the random neighbourhoods to produce its output. In that case changing the labels would accomplish little towards establishing a bound based on the previous approach.

The point is to show that allowing for the local structure to be random, that is, have the distribution of $PGW_{\lambda}$, does not provide factors with any extra power. Intuitively this seems clear because the local structure of $PGW_{\lambda}$ and $T_d$ ‘appear’ to be the same for large $\lambda$ and $d$. This intuition could be formalized by construction a coupling between $PGW_{\lambda}$ and $T_d$ that, for large $\lambda$ and $d$, would convert factors on $PGW_{\lambda}$ to similar factors on $T_d$ and vice-versa. We have been partially successful at this, allowing us to derive the lower bound for $PGW$ trees from $d$-regular trees. We will derive the upper bound for $PGW$ trees by making a technical modification to the approach that was used to analyze inequality (4.12) for $d$-regular trees.

### 5.1.1 A coupling of FIID processes on Erdős-Rényi graphs

For $0 \leq p \leq 1$ 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 $\lambda/n$. 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 $ER(n, \lambda/n)$; 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 $W^k$, correlated with $X$, by $W^k(v) = W^k(v)$ if $v \in S$, and $W^k(v) = X(v)$ otherwise. Let $G^1, G^2, \ldots$ be copies of $G'_n$ obtained from $G_n$ and $S$ such that the induced subgraphs $G^1[S], G^2[S], \ldots$ are independent. In other words, $G^i$ is generated from $G_n$ and $S$ using the recipe for $G'_n$ but the rewiring of the edges for each $G^i$ is done independently for each $i$. We define percolation sets $Y^1, Y^2, \ldots$ by letting $Y^k$ be generated by the factor $f_{n,\lambda}$ with input graph $G^k$ and labelling $W^k$. Thus, $Y^k$ is a FIID percolation set of $G^k$. Since all these graphs have a common vertex set, namely $[n]$, we can consider intersections of the $Y^k$. Note that for any finite subset $T$ the expected intersection density $E \left[ |\cap_{i \in T} Y^i|/n \right]$
Theorem 5.1.1. The following inequality holds for each vertex. We restrict our probability space to the support of $W$ according to the $\tau$ and $S\{\ldots\}$ for given $(a$ bounded measurable function on $[0, Q\sigma a d$ graph as we has done previously for $d$-regular graphs: $\mathcal{G}$, the graphs $X, X^1, \ldots$, the random subset $S$ and the independent trials that determine the graphs $G_n, G^1, G^2, \ldots$. The new $\sigma$-algebra is the restriction on the original $\sigma$-algebra to the new sample space. Let $E^*$ be the expectation operator $E$ restricted to the new space as we had done for $d$-regular graphs:

$$E^* [Y] = E \left[ f_{n,\lambda}(N_{r_1}(G_n, \circ), \circ, X(G_n, \circ, r_\lambda)) \right].$$

Notice that we define the new probability space on finite graphs instead of on the infinite limiting graph as we has done previously for $d$-regular graphs.

One can check as before that if $f_{n,\lambda}(N_{r_1}(G_n, \circ), \circ, X(G_n, \circ, r_\lambda))$ is measurable with respect to a $\sigma$-algebra $\mathcal{F}$ then $E^* [Y | \mathcal{F}] = E [Y | \mathcal{F}]$ for all random variables $Y$ on the restricted probability space. Using the notation introduced earlier define $Q_{n,\lambda,p} = Q_{n,\lambda}(G_n, \circ, S, X)$ on the new probability space as follows:

$$Q_{n,\lambda,p} = E^* \left[ f_{n,\lambda}(N_{r_1}(G^1, \circ), \circ, W^1(G^1, \circ, r_\lambda) | G_n, \circ, S, X) \right].$$

One can check, as before, that for every $k \geq 1$ the moment $E^* \left[ Q_{n,\lambda,p}^{k-1} \right] = \frac{\alpha_{k,\lambda,p}}{\alpha_{1,\lambda}}$.

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 $E^* [g(Q_{n,\lambda,p})]$ is Lipschitz continuous in $p$.

Indeed, let $p_1 \leq p_2$. We couple the labelled graphs $(G^1(S_{p_1}), \circ, W^1_{p_1})$ and $(G^1(S_{p_2}), \circ, W^1_{p_2})$ 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 $\lambda/n$. 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 $W^1_{p_1}$ (resp. $W^1_{p_2}$) 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)

\[ \mathbb{E}^* [g(Q_{n,\lambda,p1})] - \mathbb{E}^* [g(Q_{n,\lambda,p2})] = \]

\[ \mathbb{E}^* \left[ g \left( \mathbb{E}^* \left[ f(G^1(S_{p1}), \circ, W_{p1}^1) \mid G_n, \circ, X, S_{p1} \right] \right) - g \left( \mathbb{E}^* \left[ f(G^1(S_{p2}), \circ, W_{p2}^1) \mid G_n, \circ, X, S_{p2} \right] \right) \right]. \]

The point to note is that if for every \( v \in [n] \) the label \( Z(v) \not\in (p_1, p_2) \) then \( S_{p1} = S_{p2} \), and hence, \( (G^1(S_{p1}), W_{p1}^1) = (G^1(S_{p2}), W_{p2}^1) \). This forces \( \mathbb{E}^* \left[ f(G^1(S_{p1}), \circ, W_{p1}^1) \mid G_n, \circ, X, S_{p1} \right] = \mathbb{E}^* \left[ f(G^1(S_{p2}), \circ, W_{p2}^1) \mid G_n, \circ, X, S_{p2} \right] \) and the difference of the two expectations above is zero on this event. By an union bound, the complementary probability that \( Z(v) \in (p_1, p_2) \) for some vertex \( v \) is at most \( n|p_1 - p_2| \). Therefore, from the triangle inequality is follows that

\[ |\mathbb{E}^* [g(Q_{n,\lambda,p1})] - \mathbb{E}^* [g(Q_{n,\lambda,p2})]| \leq (2||g||_{\infty,n})|p_1 - p_2|. \]

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 \( W^1 = X \) 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 \( (G^1, W^1) \) 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 \( N_r(\operatorname{ER}(n, \lambda/n), v) \) does not depend on \( v \). Hence, \( Q_{n,\lambda,1} \equiv \alpha_{1,n,\lambda} \frac{\log \lambda}{\lambda} \).

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.

### 5.1.2 Proof of Theorem 5.1.1

We will show that the existence of the FIID percolation set \( Y^i \) on the graph \( G^i \) implies that with high probability each graph \( G^i \) contains a subset \( Y^i \) such that \( Y^i \) is a forest in \( G^i \), and the empirical intersection densities of the \( Y^1, \ldots, Y^k \) are close to the quantities \( \alpha_{k,n,\lambda} \frac{\log \lambda}{\lambda} \). Then we will upper bound this probability by counting the expected number of \( k \)-tuples \( (Y^1, \ldots, Y^k) \) such that \( Y^i \) is a forest in \( G^i \), and the empirical intersection densities of the \( Y^1, \ldots, Y^k \) are close to \( \alpha_{i,n,\lambda} \frac{\log \lambda}{\lambda} \) for each \( 1 \leq i \leq k \). We will show that this probability is vanishingly small unless Theorem 5.1.1 holds.

Consider a subset \( S^i \subset V(G^i) \). As all the graphs \( G^i \) have a common vertex set, namely \([n]\), we define the density profile of the \( k \)-tuple \( (S^1, \ldots, S^k) \) as the vector \( \rho = (\rho(T); T \subset [k]) \) where

\[ \rho(T) = \frac{|\cap_{i\in T} S^i|}{n}. \]

Suppose there exists an FIID percolation set \( Y^i \) on the graph \( G^i \) that is generated from a common factor \( f_{n,\lambda} \) but random labelling \( W^i \) (as introduced in Section 5.1.1). The sets \( Y^1, \ldots, Y^k \) have the property that \( Y^i \) is a forest in \( G^i \), and for every \( 1 \leq i \leq k \), \( \mathbb{E} \left[ |\cap_{j=1}^i Y^j|/n \right] = \alpha_{i,n,\lambda,p} \frac{\log \lambda}{\lambda} \).
Fix $0 < \epsilon < 1$. Let $A(\epsilon, p)$ be the following event. For each $1 \leq i \leq k$, the graph $G^i$ contains a subset $Y^i$ such that the subgraph induced by $Y^i$ in $G^i$ is a forest. Also, the density profile $\rho$ that is induced by $(Y^1, \ldots, Y^k)$ satisfies the following for all $T \subset [k]$:

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

We show that $\mathbb{P}[A(\epsilon, p)] \rightarrow 1$ as $n \rightarrow \infty$. This follows immediately if we can show the following

$$
\mathbb{P} \left[ \max_{T \subset [k]} \left\{ \frac{|\cap_{i \in T} Y^i|}{n} - \mathbb{E} \left[ \frac{|\cap_{i \in T} Y^i|}{n} \right] \right\} > \epsilon \right] \rightarrow 0.
$$

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

**Lemma 5.1.2.** With $G^1, \ldots, G^k$ as defined and corresponding forests $Y^1, \ldots, Y^k$ as defined via the factor $f_{n, \lambda}$, one has that for all $\epsilon > 0$, as $n \rightarrow \infty$,

$$
\mathbb{P} \left[ \max_{T \subset [k]} \left\{ \frac{|\cap_{i \in T} Y^i|}{n} - \mathbb{E} \left[ \frac{|\cap_{i \in T} Y^i|}{n} \right] \right\} > \epsilon \right] \rightarrow 0.
$$

**Proof.** We show that $\mathbb{E} \left[ \left| \cap_{i \in T} Y^i \right| - \mathbb{E} \left[ \left| \cap_{i \in T} Y^i \right| \right] \right]^2 = o(n^2)$ where the little $o$ term may depend on $\lambda, r_\lambda$, and $k$. The statement of the lemma then follows from Chebyshev’s inequality and an union bound over $T \subset [k]$. We write

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

Now, $|1\{v \in \cap_{i \in T} Y^i\} - \mathbb{P}[v \in \cap_{i \in T} Y^i]| \leq 2$. Hence,

$$
\mathbb{E} \left[ (1\{v \in \cap_{i \in T} Y^i\} - \mathbb{P}[v \in \cap_{i \in T} Y^i])^2 \right] \leq 4.
$$

Also, for two vertices $u$ and $v$ if the graph distance $\text{dist}_{G^i}(u, v) > 2r_\lambda$ then the events $\{u \in Y^i\}$ and $\{v \in Y^i\}$ are independent with respect to the random labelling of $G^i$. This is because the factor $f_{n, \lambda}$ makes decisions based on the labels along the $r_\lambda$-neighbourhood of a vertex and the $r_\lambda$-neighbourhoods about $u$ and $v$ are disjoint. Consequently, the correlation term

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

is at most $4\mathbb{P}\{\text{dist}_{G^i}(u, v) \leq 2r_\lambda \text{ for some } i\}$. These two observations imply that

$$
\mathbb{E} \left[ \left| \cap_{i \in T} Y^i \right| - \mathbb{E} \left[ \left| \cap_{i \in T} Y^i \right| \right] \right]^2 \leq 4n + 4\mathbb{E} \left[ \# \{(u, v) : \text{dist}_{G^i}(u, v) \leq 2r_\lambda \text{ for some } i\} \right].
$$
Using the fact that the random graphs $ER(n, \lambda/n)$ converge locally to $PGW_\lambda$, it is a standard exercise to show that the expected number of pairs $(u, v)$ in $ER(n, \lambda/n)$ that satisfy $\text{dist}(u, v) > R$ is $o(n^2)$ (the little $o$ term depends on $\lambda$ and $R$). From this observation and a union bound over $i$ we deduce that $\mathbb{E} [\# \{ (u, v) : \text{dist}_{G^i}(u, v) \leq 2r_\lambda \text{ for some } i \}] = o(n^2)$. This proves the estimate for the squared expectation and completes the proof.

Now we must bound $\mathbb{P} [A(\epsilon, p)]$ from above. In doing so we will come upon the binomial sum that is found in the statement of Theorem 5.1.1. In order to bound $\mathbb{P} [A(\epsilon, p)]$ we need a procedure to sample the graphs $G^1, \ldots, G^k$.

**Sampling the graphs** $(G^1, \ldots, G^k)$ 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 $\lambda/n$ 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 $\lambda/n$ or they are all zero with the complementary probability.

The sampling procedure above will allow us to compute expectations involving percolation sets in the $G^i$. Let $Y^i \subset [n]$ be a subset set of $G^i$ such that the induced graph $G^i[Y^i]$ is a forest. Consider the partition $\Pi$ of $[n]$ given by cells

$$\Pi(S) = \big( \cap_{i \in S} Y^i \big) \cap \big( \cap_{i \notin S} [n] \setminus Y^i \big), \quad \text{where } S \subset [k].$$

Notice that for subsets $S, T \subset [k]$ if $i \in S \cap T$ then both $\Pi(S)$ and $\Pi(T)$ are contained in $Y^i$. Consequently, any edge connecting $\Pi(S)$ to $\Pi(T)$ must be present in $G^i[Y^i].$

For disjoint sets $A, B \subset [n]$ let $e_i(A, B)$ be the number of edges in $G^i$ that connect a vertex in $A$ to a vertex in $B$. As $G^i \sim ER(n, \lambda/n)$, we have that $e_i(A, B)$ is a binomial random variable. Namely, $e_i(A, B) \sim \text{Bin}(|A| \cdot |B|, \lambda/n)$. In particular, for any non-empty subset $S \subset [k]$ we have that $\mathbb{P} [e_i(A, B) \leq x \text{ for all } i \in S] \leq \mathbb{P} [\text{Bin}(|A| \cdot |B|, \lambda/n) \leq x].$

Now we make an observation, rather trivial, that will be much useful in what follows. The number of edges in any finite forest $F$ is $|F| - \kappa(F)$, where $\kappa(F)$ is the number of connected components of $F$. In particular, for a set of vertices $A \subset V(F)$ we have that $|E(F[A])| = e(A, A) \leq |A|$ because $F[A]$ is also a forest.

If we apply this observation to our percolation sets, we see that if $S, T \subset [k]$ have a non-empty intersection then $e_i(\Pi(S), \Pi(T)) \leq |\Pi(S)| + |\Pi(T)|$ for all $i \in S \cap T$. Also, if $S = T$ then $e_i(\Pi(S), \Pi(S)) \leq |\Pi(S)|$ for all $i \in S$. Thus, if $(Y^1, \ldots, Y^k)$ is a $k$-tuple of sets with $Y^i$ being a forest in $G^i$, then the associated partition $\Pi$ has the following property.

Let $\xi(S, T)$ be the event that $e_i(\Pi(S), \Pi(T)) \leq |\Pi(S)| + |\Pi(T)|$ for every $i \in S \cap T$ if $S \neq T$, and $e_i(\Pi(S), \Pi(S)) \leq |\Pi(S)|$ for $S \neq \emptyset$. Then the events $\xi(S, T)$ must occur for every pair $\{S, T\}$ with $S \cap T \neq \emptyset$. 

**Chapter 5. Percolation on Poisson–Galton–Watson trees** 60
Chapter 5. Percolation on Poisson–Galton–Watson trees

Let $Z$ be the number of $k$-tuples of percolation sets with a given density profile. Recall that the density profile $\rho = (\rho(T), T \subset [k])$ of a $k$-tuple of percolation sets $(Y^1, \ldots, Y^k)$ is given by $\rho(T) = |\cap_{i \in T} Y^i|/n$. The density profile $\rho$ determines a probability distribution $\pi = (\pi(T); T \subset [k])$ by equation (5.4) below.

\[
\pi(T) = \sum_{T' : T \subset T'} (-1)^{|T' \setminus T|} \rho(T'), \quad \text{(5.4)}
\]

\[
\rho(T) = \sum_{T' : T \subset T'} \pi(T'). \quad \text{(5.5)}
\]

This $\pi$ has the property that if $\Pi$ is the partition of $[n]$ generated by $(Y^1, \ldots, Y^k)$ then $\pi(S)n = |\Pi(S)|$. The $\pi$ and $\rho$ are in bijection with each other via the two equations (5.4) and (5.5). Let $Z(\rho)$ be the number of $k$-tuples of subsets $(Y^1, \ldots, Y^k)$ of $[n]$ such that it has density profile $\rho$ and $G^i[Y^i]$ is a forest of $G^i$. Then $\mathbb{E}[Z(\rho)]$ equals

\[
\sum_{\text{Partitions } \Pi \text{ with profile } \rho} \mathbb{P}\left[\{(Y^1, \ldots, Y^k) \subset [n]^k \text{ satisfy } G^i[Y^i] \text{ is a forest in } G^i \text{ and generates partition } \Pi\}\right].
\]

We can bound each of the summands by using (5.3) and Lemma 2.4.2. The probability on the right hand side of the inequality (5.3) depends on $\pi$, which can be derived from $\rho$ and otherwise does not depend on the partition $\Pi$ generated by a specific $(Y^1, \ldots, Y^k)$. Moreover, the number
of partitions \( \Pi \) that has density profile \( \rho \) is \( (\pi(S)_{n, S \subseteq [k]}) \). Therefore, there are \( (\pi(S)_{n, S \subseteq [k]}) \) terms in the sum above, and each term is bounded from above by \( (5.3) \). As a result,

\[
\mathbb{E}[Z(\rho)] \leq \left( \pi(S)_{n, S \subseteq [k]} \right)^n \prod_{S \neq \emptyset} \mathbb{P}\left[ \text{Bin}\left( \frac{\pi(S)n}{2}, \frac{\lambda}{n} \right) \leq \pi(S)n \right] \prod_{\{S,T\}: S \neq T, S \cap T \neq \emptyset} \mathbb{P}\left[ \text{Bin}(\pi(S)\pi(T)n^2, \frac{\lambda}{n}) \leq (\pi(S) + \pi(T))n \right].
\] (5.6)

Now we use Lemma 2.4.2. We set \( p = \lambda/n, m = \left( \frac{\pi(S)n}{2} \right) = (1/2)\pi(S)^2n^2 - O(1/n) \) when \( S = T \), or \( m = \pi(S)\pi(T)n^2 \) when \( S \neq T \). In both cases we have that \( \mu = \frac{\pi(S) + \pi(T)}{\lambda\pi(S)\pi(T)} \) with an additive error term of order \( O(1/n) \) when \( S = T \). But this error term only influences the binomial probabilities on the right hand side of \( (5.3) \) by at most a constant multiplicative term. So from Lemma 2.4.2 we get that for all pairs \( S \) and \( T \),

\[
\mathbb{P}\left[ \text{Bin}(\pi(S)\pi(T)n^2, \frac{\lambda}{n}) \leq (\pi(S) + \pi(T))n \right] \leq O(\sqrt{n\lambda}) \times \\
\exp \left\{-n \left[ \lambda\pi(S)\pi(T) + (\pi(S) + \pi(T)) \left( \log\left( \frac{\pi(S) + \pi(T)}{\lambda\pi(S)\pi(T)} \right) - 1 \right) \right] + \lambda(\pi(S) + \pi(T)) \right\}.
\]

We write \( \pi(S) = \beta(S)\frac{\log \lambda}{\lambda} \) for \( S \neq \emptyset \). Due to Theorem 1.2.10 we can assume that \( |Y|/n \leq 3\lambda + o(1) \) for all large values of \( \lambda \). If \( S \neq \emptyset \) then \( \pi(S) \subseteq Y \) for any \( i \in S \), and hence, \( \beta(S) \leq 3 \). Also, we may assume that all \( \beta(S) > 0 \) because if some \( \beta(S) = 0 \) then the corresponding \( \Pi(S) = \emptyset \). Then we may ignore \( \Pi(S) \) from the calculations, that is, ignore probabilities of all events involving \( \Pi(S) \).

Analyzing the principal component of the exponential rate above we get

1. \( \lambda\pi(S)\pi(T) = \beta(S)\beta(T)\frac{\log^2 \lambda}{\lambda} \).

2. \( \frac{\pi(S) + \pi(T)}{\lambda\pi(S)\pi(T)} = \frac{\beta(S) + \beta(T)}{\beta(S)\beta(T)} \). As \( \beta(S) + \beta(T) \geq 2\sqrt{\beta(S)\beta(T)} \), we see that \( \frac{\beta(S) + \beta(T)}{\beta(S)\beta(T)} \geq 2(\beta(S)\beta(T))^{-1/2} \geq 2/3 \). Consequently, \( \frac{\pi(S) + \pi(T)}{\lambda\pi(S)\pi(T)} \geq \frac{2}{3\log \lambda} \) and taking logarithms we get

\[
\log\left( \frac{\pi(S) + \pi(T)}{\lambda\pi(S)\pi(T)} \right) - 1 \geq \log(2/3) - 1 - \log \log \lambda. 
\]

Note that \( \log(2/3) - 1 - \log \log \lambda \leq -2 \log \log \lambda \) if \( \lambda \geq e^{3/2} \). Therefore, for all large \( \lambda \) we get

\[
(\pi(S) + \pi(T))\left( \log\left( \frac{\pi(S) + \pi(T)}{\lambda\pi(S)\pi(T)} \right) - 1 \right) \geq -2(\beta(S) + \beta(T)) \frac{\log \lambda \log \log \lambda}{\lambda} \\
\geq -12 \frac{\log \lambda \log \log \lambda}{\lambda}.
\]

The term \( \lambda(\pi(S) + \pi(T)) \leq 6 \log \lambda \), and so \( e^{\lambda(\pi(S) + \pi(T))} \leq \lambda^6 \). We can now conclude that
\begin{equation}
\mathbb{P}\left[ \text{Bin}(\pi(S)\pi(T)n^2, \lambda/n) \leq (\pi(S) + \pi(T))n \right] \leq O(\sqrt{n\lambda^2}) \times \exp\left\{ n\left[-\beta(S)\beta(T)\frac{\log^2 \lambda}{\lambda} + 12\frac{\log \lambda \log \log \lambda}{\lambda}\right]\right\}. 
\end{equation}

Plugging this estimate into the right hand of (5.6) we deduce that for all large \( \lambda \), up to a polynomial factor \( \text{poly}(n, \lambda) \) which has degree almost \( 4k/2 \) in \( n \) and \( 6 \cdot 4^k \) in \( \lambda \), we have

\begin{equation}
\mathbb{E}[Z(\rho)] \leq \text{poly}(n, \lambda) \times \left( \pi(S), S \subset [k] \right) \times \prod_{\{S,T\}: S \cap T \neq \emptyset} \exp\left\{ n\left[-\beta(S)\beta(T)\frac{\log^2 \lambda}{\lambda} + 12\frac{\log \lambda \log \log \lambda}{\lambda}\right]\right\} 
\end{equation}

\begin{equation}
\leq \text{poly}(n, \lambda) \times \left( \pi(S), S \subset [k] \right) \times \exp\left\{ n\left[-\frac{\log^2 \lambda}{\lambda} \left( \sum_{\{S,T\}: S \cap T \neq \emptyset} \beta(S)\beta(T) \right) + O_k\left(\frac{\log \lambda \log \log \lambda}{\lambda}\right)\right]\right\}.
\end{equation}

Recall from Stirling’s approximation that \( \left( \pi(S), S \subset [k] \right) \leq \text{poly}(n) e^{nH(\pi)} \). We also observed in Section 4.2 (see the paragraph before (4.7)) that

\[ H(\pi) = \left( \sum_{S \neq \emptyset} \beta(S) \right) \frac{\log^2 \lambda}{\lambda} + O_k\left(\frac{\log \lambda \log \log \lambda}{\lambda}\right). \]

As a result we get an analogous inequality to Lemma 3.2.1 for all large \( \lambda \)

\begin{equation}
\mathbb{E}[Z(\rho)] \leq \text{poly}(n, \lambda) \times \exp\left\{ n\left[-\frac{\log^2 \lambda}{\lambda} \left( \sum_{S \neq \emptyset} \beta(S) - \sum_{\{S,T\}: S \cap T \neq \emptyset} \beta(S)\beta(T) \right) + O_k\left(\frac{\log \lambda \log \log \lambda}{\lambda}\right)\right]\right\}.
\end{equation}

Finally, recall that the term \( \sum_{S \neq \emptyset} \beta(S) - \sum_{\{S,T\}: S \cap T \neq \emptyset} \beta(S)\beta(T) \) can be written in terms of the density profile \( \rho \) via the principal of inclusion–exclusion [5.4]. In particular, suppose that \( \rho(T) = \alpha_{|T|} \frac{\log \lambda}{\lambda} \) for every non-empty \( T \subset [k] \). As explained in Section 4.2 (recall 4.11), we have that

\begin{equation}
\sum_{S \neq \emptyset} \beta(S) - \frac{1}{2} \sum_{\{S,T\}: S \cap T \neq \emptyset} \beta(S)\beta(T) = \sum_{i=1}^{k} (-1)^{i-1} \binom{k}{i} \left( \alpha_i - \frac{1}{2} \alpha_i^2 \right). 
\end{equation}

From this point onward the proof of Theorem 5.1.1 proceeds in the same manner as for the case with \( d \)-regular graphs, which is the argument from Theorem 3.2.2. Recall the event \( A(\epsilon, p) \): the graph \( G^i \) contains a set \( Y^i \) such that the subgraph induced by \( Y^i \) is a forest in \( G^i \) and the
density profile of \((Y^1, \ldots, Y^k)\) satisfies
\[ \rho(T) \in [(1 - \epsilon)\alpha_T, (1 + \epsilon)\alpha_T] \] for all \(T \subset \{1, \ldots, k\}\).

Arguing as in the proof of Theorem 3.2.2 we use a first moment bound along with 5.8 and 5.9 to conclude that \(\mathbb{P}[A(\epsilon, p)]\) satisfies an upper bound equivalent to (3.6):
\[ \mathbb{P}[A(\epsilon, p)] \leq \exp \left\{ n \left( \frac{\log^2 \lambda}{\lambda} \sum_i (-1)^{i-1} \binom{k}{i} \left( \alpha_{i, n, \lambda, p} - \frac{1}{2} \alpha_{i, n, \lambda, p}^2 \right) + \text{err}(\epsilon) \right) + O_k \left( \frac{\log \lambda \log \log \lambda}{\lambda} \right) \right\} \]
where \(\text{err}(\epsilon) \to 0\) as \(\epsilon \to 0\) uniformly in \(n, \lambda\) and \(p\).

Now suppose that for some \(\delta > 0\) we have
\[ \liminf_{\lambda \to \infty} \liminf_{n \to \infty} \inf_{p \in [0, 1]} \sum_i (-1)^{i-1} \binom{k}{i} \left( \alpha_{i, n, \lambda, p} - \frac{1}{2} \alpha_{i, n, \lambda, p}^2 \right) = -\delta < 0. \]

For every \(\lambda\) and \(n\) we can find a \(p' = p'(\lambda, n)\) such that
\[ \inf_{p \in [0, 1]} \sum_i (-1)^{i-1} \binom{k}{i} \left( \alpha_{i, n, \lambda, p'} - \frac{1}{2} \alpha_{i, n, \lambda, p'}^2 \right) \geq \sum_i (-1)^{i-1} \binom{k}{i} \left( \alpha_{i, n, \lambda, p'} - \frac{1}{2} \alpha_{i, n, \lambda, p'}^2 \right) - \frac{\delta}{2}. \]

Therefore, \(\liminf_{\lambda \to \infty} \liminf_{n \to \infty} \sum_i (-1)^{i-1} \binom{k}{i} \left( \alpha_{i, n, \lambda, p'} - \frac{1}{2} \alpha_{i, n, \lambda, p'}^2 \right) \leq -\delta/2\). As \(\text{err}(\epsilon) \to 0\) uniformly in \(n, \lambda\) and \(p\), we can find a \(\epsilon' > 0\) such that \(\text{err}(\epsilon') < \delta/4\) for all \(n, \lambda\) and \(p\). Consequently, there exists a sequence of \(\lambda_i \to \infty\) with the following property. For every \(\lambda = \lambda_i\), there is a \(n_{\lambda}\) such that if \(n \geq n_{\lambda}\) then
\[ \mathbb{P}[A(\epsilon', p')] \leq \exp \left\{ n \left[ -\frac{\delta \log^2 \lambda}{4 \lambda} + O_k \left( \frac{\log \lambda \log \log \lambda}{\lambda} \right) \right] \right\}. \]

This implies that \(\mathbb{P}[A(\epsilon', p')] \to 0\) as \(n \to \infty\) for all sufficiently large \(\lambda_i\). However, we have shown that \(\mathbb{P}[A(\epsilon', p')] \to 1\) as \(n \to \infty\) for every \(\lambda\). Thus,
\[ \liminf_{\lambda \to \infty} \liminf_{n \to \infty} \inf_{p \in [0, 1]} \sum_i (-1)^{i-1} \binom{k}{i} \left( \alpha_{i, n, \lambda, p} - \frac{1}{2} \alpha_{i, n, \lambda, p}^2 \right) \geq 0, \]

as required.

### 5.2 A lower bound from regular trees

We will show that FIID independent sets of regular trees can be used to construct such independent sets on PGW trees as well. Let \(I\) be a FIID independent set on the regular tree \(T_d\) 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 \(\text{PGW}_\lambda\) and all of its neighbours have degree at most \(d\).
Theorem 5.2.1. Given $I$ as above there exists a FIID independent set $J$ of $\text{PGW}_\lambda$ whose density satisfies the bound
\[
density(I) \mathbb{P}[E(\lambda, d)] \leq \density(J) \leq \density(I).
\]

Proof. We construct $J$ in three stages.

The edge removal stage We remove edges from $\text{PGW}_\lambda$ via a FIID process such that all vertices will have degree at most $d$ after the removal procedure. Begin with a random labelling $X$ of $\text{PGW}_\lambda$. For each vertex $v$ consider all the neighbours $u$ of $v$ such that the variables $X_u$ are the degree$(v) - d$ highest in value (provided, of course, that degree$(v) > d$). Mark the 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 $\text{degree}_G(v) < d$, then attach to it $d - \text{degree}_G(v)$ 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 of $G$ it may not be an independent set of the original tree $\text{PGW}_\lambda$ 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 FIID process on $\text{PGW}_\lambda$. $J$ is also an independent set because if $(u, v)$ is an edge of $\text{PGW}_\lambda$ 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 of $G$). 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 $\text{PGW}_\lambda$ 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
\[
density(I) \mathbb{P}[E(\lambda, d)] \leq \density(J) \leq \density(I).
\]

Lemma 5.2.2. If $\lambda = d - d^u$ for any $1/2 < u < 1$ then the probability $\mathbb{P}[E(\lambda, d)] \to 1$ as $d \to \infty$. 

\[\Box\]
Proof. This is a calculation involving Poisson tail probabilities. Recall that the moment generating function of a Poisson(μ) random variable is e^{μ(e^x−1)}. Let X denote the degree of the root in a PGW tree of expected degree λ. Let Z_1, ..., Z_X denote the number of offspring of the neighbors of the root. Recall that X has distribution Poisson(λ), and that conditioned on X the random variables Z_1, ..., Z_X are i.i.d. with distribution Poisson(λ).

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

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

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

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

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

\[
\geq e^{-\lambda p(\lambda, d-1)} - p(\lambda, d-1)
\]

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

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

Setting \( \lambda = d - d^u \) for \( 1/2 < u < 1 \), we see from the bound above that \( p(d - d^u, d) \leq e^{d^u} (1 - d^u)^d = e^{d^u + d \log(1 - d^u)} \). Since

\[
\log(1 - x) = - \sum_{k \geq 1} \frac{x^k}{k} \leq -x - x^2/2 \quad \text{for } 0 \leq x < 1,
\]

by setting \( x = d^{u-1} < 1 \) we conclude that

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

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.

Given \( \lambda \), set \( d = \lfloor \lambda + \lambda^{3/4} \rfloor \). From the definition of \( \alpha_{\text{perc}}(\text{PGW}_\lambda) \), \( \alpha_{\text{IND}}(\mathbb{T}_d) \) and the conclusion of Theorem 5.2.1, we have that \( \alpha_{\text{perc}}(\text{PGW}_\lambda) \geq \alpha_{\text{IND}}(\mathbb{T}_d) \cdot P[E(\lambda, d)] \). As a result,

\[
\frac{\alpha_{\text{perc}}(\text{PGW}_\lambda)}{(\log \lambda)/\lambda} \geq \frac{\alpha_{\text{IND}}(\mathbb{T}_d)}{(\log d)/d} \cdot \frac{(\log d)/d}{(\log \lambda)/\lambda} \cdot P[E(\lambda, d)].
\]

Due to Theorem 1.2.1, we have that \( \lim_{\lambda \to \infty} \alpha_{\text{IND}}(\mathbb{T}_d) \cdot (\log d)/d = 1 \). By our choice of \( d \) as a function of \( \lambda \) we have \( \lim_{\lambda \to \infty} \frac{(\log d)/d}{(\log \lambda)/\lambda} = 1 \). By Lemma 5.2.2, we have that the probability \( P[E(\lambda, d)] \to 1 \)
as $\lambda \to \infty$. As a result, we conclude from the inequality above that

$$\liminf_{\lambda \to \infty} \frac{\alpha_{\text{perc}}(\text{PGW}_\lambda)}{(\log \lambda)/\lambda} \geq 1.$$ 

This lower bound completes the proof of Theorem 1.2.9.
Chapter 6

Concluding remarks and open problems

We conclude the thesis with a brief discussion of open problems.

Locality of the independence ratio of random 3-regular graphs  Our thesis studies the maximal density of FIID percolation on sparse graphs where the sparsity parameter (average degree) tends to infinity. However, it is still a very interesting problem to compute the maximum density for various classes of FIID processes on $d$-regular graphs for fixed values of $d$. There have been recent progress in this regard for independent sets on 3-regular graphs. Csoka et al [16] use Gaussian processes to construct FIID independent sets on $T_3$ of density at least 0.436, while Hoppen and Wormald [30] improve the bound to 0.437 via another local algorithm. McKay [18] has shown that the maximum density of independent sets on random 3-regular graphs is at most 0.4554, and this was recently improved to 0.4509 by Barbier et al [4]. Is it true that $\alpha_{\text{IND}}(3) = \alpha_{\text{IND}}(T_3)$?

Maximal density of FIID percolation with a given correlation  We have proved in Theorem 1.2.4 that if an FIID percolation process $\phi \in \{0,1\}^{V(T_d)}$ satisfies $\text{corr}(\phi) \leq c$ then $\text{den}(\phi) \leq \frac{2+o_d(1)}{1-c+c\log c} \frac{\log d}{d}$ for large $d$. It should be possible to improve upon the factor of $2/(1 - c + c\log c)$ with a more technical analysis of the entropy inequality applied to multiple correlated copies of $\phi$. However, it is not clear that this method will lead to an optimal bound. What is the optimal bound on the density on $\frac{\text{den}(\phi)}{\log d}/d$ for asymptotically large $d$? What FIID percolation processes improve upon the lower bound of $\frac{1}{\sqrt{1-c}} \frac{\log d}{d}$ from Theorem 1.2.4?

Are MAXCUT and MINCUT local?  A (balanced) cut set in a finite graph $G$ is a subset of vertices $S$ such that $|S| - |S^c| \leq 1$ (if $|G|$ is even then $|S| = |G|/2$). The density of the cut is the ratio of number of edges from $S$ to $S^c$ to twice the total number of edges: $|E(S, S^c)|/2|E|$. If one picks a uniform random directed edge of $G$ then the density of a cut is the probability
that the head of the edge lies in $S$ and the tail lies in $S^c$. The max and min cut ratio of $G$ are defined as

$$\text{MAXCUT}(G) = \max \left\{ \frac{|E(S, S^c)|}{2|E(G)|} : S \text{ is a cut set in } G \right\},$$

$$\text{MINCUT}(G) = \min \left\{ \frac{|E(S, S^c)|}{2|E(G)|} : S \text{ is a cut set in } G \right\}.$$

It has been recently shown that $\mathbb{E}[\text{MAXCUT}(G_n, d)] \rightarrow \text{MAXCUT}(d)$ where $\text{MAXCUT}(d) = 1/4 + c/\sqrt{d} + o(d^{-1/2})$ for some explicit constant $c \approx 0.3816$ [19]. An analogous result holds for $\mathbb{E}[\text{MINCUT}(G_n, d)]$ with the limiting value equal to $\text{MINCUT}(d) = 1/4 - c/\sqrt{d} + o(d^{-1/2})$.

A local cut is an FIID process $\sigma \in \{-1, 1\}^T_d$ such that $\mathbb{E}[\sigma(\circ)] = 0$. The cut density of $\sigma$ is $P[\sigma(\circ) = 1, \sigma(\circ') = -1]$ where $(\circ, \circ')$ is a directed edge of $T_d$. Define

$$\text{LOCMAXCUT}(d) = \sup \left\{ P[\sigma(\circ) = 1, \sigma(\circ') = -1] : \sigma \text{ is a local cut in } T_d \right\},$$

$$\text{LOCMINCUT}(d) = \inf \left\{ P[\sigma(\circ) = 1, \sigma(\circ') = -1] : \sigma \text{ is a local cut in } T_d \right\}.$$

Is it true that $\sqrt{d}[\text{LOCMAXCUT}(d) - (1/4)] \rightarrow c$ as $d \rightarrow \infty$ and $\sqrt{d}[\text{LOCMINCUT}(d) - (1/4)] \rightarrow -c$ as $d \rightarrow \infty$? So far as we know the best bounds for local cuts in $T_d$, for general $d$, are due to Lyons [34] who proves that $\text{LOCMAXCUT}(d) \geq 1/4 + 1/(\pi \sqrt{d})$ and $\text{LOCMINCUT}(d) \leq 1/4 - 1/(\pi \sqrt{d})$. Recently, Gamarnik and Li [24] prove that $\text{LOCMAXCUT}(3) \geq (1.36)/3$.

**Maximal density FIID bond percolation** We have seen in Theorem 1.2.7 that the maximal density of an FIID bond percolation in $T_d$ with finite clusters in exactly 2/d. In fact, our proof shows that the maximal density remains to be 2/d so long as the component of the root has at most two topological ends. It is known (see [36] Example 8.5) that with probability one the number of ends of the component of the root in any FIID bond percolation in $T_d$ is 0, 1, 2 or $\infty$. Therefore, 2/d is the maximal density of an FIID bond percolation in $T_d$ if the component of the root has a finite number of ends with probability one.

Let $K$ be the set of all probability measures on $\{0, 1\}^E(T_d)$ that arise from FIID bond percolation processes $\phi$ on $T_d$ with the property that $\text{den}(\phi) = 2/d$ and the component of the root in $\phi$ has at most 2 ends. The set $K$ is convex. What are the extreme points of $K$?

Let $\phi$ be an FIID bond percolation process in $T_d$ with density 2/d. Let $C(\circ)$ be the component of the root in $\phi$ and suppose that it has at most two ends. The probability that $C(\circ)$ is finite is zero, for otherwise, the argument in the proof of Theorem 1.2.7 would show that $\text{den}(\phi) < 2/d$. Let $p = \mathbb{P}[C(\circ) \text{ has 1 end}]$ and suppose that $0 < p < 1$. Then for any event $E$ we have

$$\mathbb{P}[E] = p \mathbb{P}[E \mid C(\circ) \text{ has 1 end}] + (1 - p) \mathbb{P}[E \mid C(\circ) \text{ has 2 ends}].$$
This shows that the measure associated to $\phi$ is not an extreme point of $K$. Therefore, if $\phi$ is an extreme point of $K$ then with probability one $\mathcal{C}(\circ)$ has either one end or two ends.

**Transference of local algorithms between regular trees and Poisson Galton Watson trees** We have shown in Theorem 5.2.1 that an FIID independent set $I$ on $\mathbb{T}_d$ can be used to construct an FIID independent set $J$ on $\text{PGW}_\lambda$ satisfying $c_{d,\lambda} \leq \frac{\text{den}(J)}{\text{den}(I)} \leq 1$ with $c_{d,\lambda} \to 1$ if $|\lambda/d - 1| = O(d^{-1/3})$. This allowed us to derive lower bounds on the density of FIID independent sets in $\text{PGW}_\lambda$ from FIID independent sets in $\mathbb{T}_d$. Is there a transference theorem of this kind from $\text{PGW}_\lambda$ trees to $\mathbb{T}_d$? In other words, given an FIID independent set $J$ in $\text{PGW}_\lambda$ does there exist a coupling $(A, B)$ such that $A$ is an FIID independent set in $\text{PGW}_\lambda$, $B$ is an FIID independent set in $\mathbb{T}_{d,\lambda}$ for some $d, \lambda$, and $A \sim J$? Moreover, we require that $d, \lambda / \lambda \to 1$ as $\lambda \to \infty$ and $\text{den}(A)/\text{den}(B) \to 1$ as $\lambda \to \infty$.

More generally, and also vaguely, is there a coupling of FIID processes on regular trees with FIID processes on Poisson Galton Watson trees such that the local statistics of the coupled processes on the two trees are asymptotically equal as the degree tends to infinity?
Bibliography


