Weizmann Institute of Science

Thesis for the degree
Master of Science

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

Generalized Girth Problems in Graphs and Hypergraphs

Author:
Tal Wagner

Supervisor:
Prof. Uriel Feige

November 2013

Abstract

We study the asymptotic value of several extremal problems on graphs and hypergraphs, that arise as generalized notions of girth. Apart from being combinatorially natural questions, they are motivated by computational-theoretic applications.

1. An *ℓ-subgraph* is a subgraph with ℓ edges per vertex, or equivalently, average degree 2ℓ. What is the optimal upper bound $S_\ell(n,d)$, such that any graph of size n and density d must contain an ℓ-subgraph of size at most $S_\ell(n,d)$?

   The $\ell = 1$ case coincides with the girth problem, and the answer is well known to be $\Theta(\log_{d-1} n)$. For $\ell \geq 2$ we give nearly tight upper and lower bounds:

   $$\forall \epsilon > 0, \quad \Omega(n/d^{\frac{\ell}{\ell-1}}) \;\leq\; S_\ell(n,d) \;\leq\; O(n/d^{\frac{\ell}{\ell-1}-\frac{1}{\ell^3-2\ell+1}-\epsilon})$$

   For example for $\ell = 2$, every graph of size n and density d contains a subgraph of size $O(n/d^{1.8-\epsilon})$ and average degree 4. We further improve the upper bound to $O(n^{2-\epsilon})$, nearly meeting the lower bound, for graphs with bounded density or large girth, and conjecture it should hold for general graphs as well.

2. The *ℓ-girth* of a graph is the size of its smallest subgraph with minimum degree ℓ. What is the optimal upper bound $g_\ell(n,d)$, such that any graph of size n and density d has ℓ-girth at most $g_\ell(n,d)$?

   Erdős et al. [EFRS90] proved that $g_\ell(n,d) = O(n/d)$. We prove,

   $$\forall \epsilon > 0, \quad \Omega(n/d^{1+\frac{2}{\ell-2}}) \;\leq\; g_\ell(n,d) \;\leq\; O(n/d^{1+\frac{1}{\ell-2}-\frac{1}{\ell^3-3\ell^2+\ell+2}-\epsilon})$$

   For example for 3-girth, we get $\Omega(n/d^3) \leq g_3(n,d) \leq O(n/d^{1.8-\epsilon})$.

3. The $S_\ell(n,d)$ question is naturally posed on hypergraphs as well. We give upper bounds on the size of (2/3)-subgraphs in 3-uniform hypergraphs, progressing towards a conjecture of Feige [Fei08] that was raised in the context of efficient random 3CNF refutation.

Acknowledgements

I thank my advisor Prof. Uriel Feige for his insightful guidance through this research, and for allowing me a glimpse into the depths of his thought. It has been inspiring to work with him.

In parallel to this work I was fortunate to collaborate with Prof. Robert Krauthgamer. I thank him for his time, patience, and the many things I learned from him. His added guidance has been extremely valuable to me.

Contents

1 Introduction

This work deals with several extremal problems in Graph and Hypergraph Theory, that can be described in terms of generalized notions of girth. The *girth* of a graph is the size of the smallest cycle it contains (or infinity if there are none), and it poses a fundamental graph-theoretic notion that arises in many contexts. The problems we study arise as natural generalizations, and are further motivated by applications to analysis of algorithms as will be detailed below.

All graphs under discussion are simple and undirected unless stated otherwise. The main problem we address is as follows. An *ℓ-subgraph* of a graph is a subgraph with at least ℓ edges per vertex, or equivalently, with average degree $\geq 2\ell$.

Question 1.1. *What is the optimal upper bound $S_\ell(n,d)$, such that any graph of size n and density[1] d must contain an ℓ-subgraph of size at most $S_\ell(n,d)$?*

Observe that a smallest 1-subgraph is a smallest cycle and vice-versa, and therefore, the size of the smallest ℓ-subgraph may be viewed as a type of generalized girth. The $\ell = 1$ case of Question 1.1 then asks how large can the girth of a graph be in terms of its size and density, and it has been studied thoroughly. It is well known that $S_1(n,d) = \Theta(\log_{d-1} n)$, that is, any graph of size n and density d contains a cycle of at most such size, and there are graphs that exclude any smaller cycles. To the best of our knowledge (and somewhat surprisingly), the question for larger values of ℓ has not yet been addressed.

A closely related notion of generalized girth, that has not eluded previous attention, is the *ℓ-girth*, defined as the size of the smallest subgraph with minimal degree ℓ (or infinity if there are none). Similarly to Question 1.1, the following question arises.

Question 1.2. *What is the optimal upper bound $g_\ell(n,d)$, such that any graph of size n and density d has ℓ-girth at most $g_\ell(n,d)$?*

Here the $\ell = 2$ case coincides with the usual girth, hence $g_2(n,d) = S_1(n,d) = \Theta(\log_{d-1} n)$. Question 1.2 for $\ell > 2$ has been addressed in the past, and existing bounds are reviewed later in this section.

The third question we consider concerns hypergraphs. A *hypergraph* is a pair of a vertex set V and a *hyperedge* set $E \subset 2^V$, and is called *r-uniform* (or just a *r-hypergraph*) if each hyperedge has size r. The $r = 2$ case clearly coincides with graphs in the usual sense. The following is therefore a natural extension of Question 1.1.

Question 1.3. *Answer Question 1.1 for r-uniform hypergraphs, with any $r \geq 2$.*

Remark on density measures. Our upper bounds on $S_\ell(n,d)$ and $g_\ell(n,d)$ will be proven with d being the average degree, and our lower bounds with d being the minimal degree. In both cases this is the stronger form of the result. Informally we refer to d as the "density", and think of it in terms of asymptotic dependence on n.

[1]By *density* we typically mean the average degree; see a later remark.

1.1 Motivation and Related Work

Girth. The girth problem, which is to determine $S_1(n,d)$ (or equivalently $g_2(n,d)$), has been intensively studied. The aforementioned result $S_1(n,d) = \Theta(\log_{d-1} n)$ is in fact easy to achieve: The upper bound is simply by a breadth-first search on the graph, and the lower bound follows from a standard application of the probabilistic deletion method. Both can be found in [Bol04, page 104, Theorems 1.1 and 1.2], and for the convenience of the reader, they are re-proven here (in a formulation more suitable with this work) in Appendix D.

A major problem has been to identify the optimal leading constant c of the $\log_{d-1} n$ term in the bound, which to date remains open, despite concentrated efforts. The best upper bound is $c \leq 2$ due to Alon, Hoory and Linial [AHL02], who have extended a trivial upper bound on d-regular graphs (known as the Moore bound and obtained by the aforementioned BFS) to arbitrary graphs with average degree d, thus resolving a long standing open question. The best lower bound is $c \geq 4/3$, originally achieved by Margulis [Mar82, Mar88] and (independently) by Lubotzky, Phillips and Sarnak [LPS88], and subsequently reproved and extended in a long sequence of works [Imr84, BB90, Mor94, LU95, Dah13]. Notably, their proofs are by explicit constructions of extremal graphs (known as "Ramanujan graphs"), and have attracted considerable attention in many other regards.

ℓ-Girth. The ℓ-girth has been studied since the late 80's, and in particular, Question 1.2 was addressed by Erdős et al. [EFRS90] and by Kézdy and Markert [KM87, Kez91]. In [EFRS90, Theorem 2] it is shown that $g_\ell(n,d) \leq 2\ell n/d$, and to our knowledge this is the only known bound for general densities.

Apart from this result, research has primarily focused on borderline low densities. The precise d that guarantees $g_\ell(n,d) < \infty$ is known (see [Kez91, Lemma 23]), and graphs with that density and ℓ-girth exactly n are studied in [Kez91, Chapter 5] and [KM87]. Upper bounds in the presence of one extra edge are discussed in [EFRS90], and some properties of graphs one edge short are studied by Erdős et al. [EFGS88] and by Bollobás and Brightwell [BB89]. With respect to restricted types of graphs, the ℓ-girth problem for cycle powers was tackled by Bermond and Peyrat [BP89] and Brandt et al. [BMRR10].

From a complexity-theoretic point of view, the computational problem of determining the ℓ-girth of an input graph was considered by Amini et al. [ASS12, APP$^+$12], who have shown various hardness of approximation results for $\ell \geq 3$. (The $\ell = 2$ case is easily seen to be polynomial-time solvable, by performing a BFS starting at each vertex.) We remark that in Appendix A we discuss some computational aspects related to $S_\ell(n,d)$.

Hypergraphs. To motivate the discussion of hypergraphs, we present yet another notion of generalized girth: An *even cover* is a subset of hyperedges by which each vertex is covered an even (possibly zero) number of times. In graphs ($r = 2$), a smallest even cover coincides with (the edge set of) a smallest cycle, so again we are dealing with a girth-type problem.

Question 1.4. *What is the optimal upper bound $EC_r(n,d)$, such that any r-uniform hypergraph of size n and density γd, with γ a sufficiently large constant, must contain an even cover of size at most $EC_r(n,d)$?*

This question has been directly addressed in several works. Feige posed the following conjecture [Fei08, Conjecture 1.2],

Conjecture 1.5. *$EC_r(n,d) = \tilde{O}(n/d^{2/(r-2)})$, where the $\tilde{O}$ notation may hide a multiplicative polylogarithmic term.*

This meets a known lower bound (exhibited by random hypergraphs), and hence offers a full resolution of Question 1.4. The conjecture was priorly shown by Feige, Kim and Ofek [FKO06] to hold for random hypergraphs, and for arbitrary hypergraphs, some progress has been made on high densities. Naor and Verstraëte [NV08] proved the highest density case for all even r, i.e. $EC_r(n, n^{(r-2)/2}) = O(\log n)$ (see [Fei08, Proposition 2.2]). The odd r case appears more difficult; focusing on $r = 3$, [Fei08] and [NV08] obtain logarithmic bounds on $EC_3(n,d)$ for various densities in the regime $\tilde{O}(\sqrt{n})$, slightly above the borderline high case of Conjecture 1.5 which is $d = \sqrt{n}$. For yet higher densities, Dellamonica et al. [DHL+12] prove[2] that $EC_3(n, n^{1/2+\Theta(1/m)}) = m$ for any m.

The connection between Questions 1.3 and 1.4 is twofold. On one hand, an even cover in a r-hypergraph spans a $(2/r)$-subgraph, as each hyperedge covers r vertices and each spanned vertex is covered at least twice. In graphs ($r = 2$) this simply means that a cycle has at least as many edges as vertices. Question 1.3 is therefore suggested in [Fei08] as an intermediate step towards proving Conjecture 1.5. On the other hand, an ℓ-subgraph with $\ell > 1$ must contain an even cover: Viewing each hyperedge as an indicator vector for its vertices, a sub-hypergraph with more hyperedges than vertices contains a linear dependency over $\mathbb{F}_2$, i.e. a subset of vectors that sum to zero modulo 2, and that subset forms an even cover.[3] Therefore, upper bounds on Question 1.3 in the $\ell > 1$ regime apply directly to Question 1.4.[4] For example, Alon and Feige [AF09, Lemma 3.3] prove that any 3-hypergraph of size n and density d contains a subgraph of size $O(n \log n/d)$ with strictly more hyperedges than vertices, and hence an even cover of that size. For general densities this is currently the best bound on the $r = 3$ case of both Question 1.3 with $\ell = 1$ and Question 1.4.

The study of even covers, and in particular of Question 1.4, is well motivated by applications to Theory of Computation. In [NV08] it arises in the context of *sparse parity-check matrices*, a key notion in Coding Theory. The motivation in [Fei08] is the design of *refutation algorithms* for random 3CNF formulas, which is an average-case variant of the fundamental 3-satisfiability problem (3SAT). This task poses a main challenge in Computational Complexity and further has implication to hardness of approximation [Fei02]. The approach centered at even covers has so far been the most successful: Apart from achieving the best bounds for random refutation [FO07], it has also been successfully applied to non-deterministic refutation in parameters regimes for which deterministic algorithms are unknown [FKO06], and to refutation of semi-random formulas, in which the assignment of variables to clauses is adversarial and the randomness is confined to variable polarities [Fei07]. This latter work is

[2] Their result assumes that any two hyperedges may intersect on at most one vertex, but by [Fei08, Lemma 2.4] removing this assumption has only a negligible effect on the density.

[3] In the even r case this holds already for $\ell = 1$, i.e. at least as many hyperedges as vertices. This is because each hyperedge e satisfies $\mathbf{1}^T e = 0$ over $\mathbb{F}_2$ (where $\mathbf{1}$ is the all-1's vector), so the hyperedges reside in a subspace with co-dimension 1.

[4] Our work does not handle this regime, and only applies to Question 1.3 with ℓ approaching 1 from below.

especially notable in our context, as it relies directly on upper bounds for Question 1.4 (in the $r = 3$ case).

Turán-type problems. Questions 1.1 to 1.4 fit into a broader theme in Extremal Combinatorics known as *Turán-type* problems, where the goal is to determine the maximum number of edges that a graph (or hypergraph) may have while avoiding certain subgraphs. Typically, however, these problems are concerned with a single forbidden subgraph (the original Turán Theorem avoids a clique) or a rather restricted forbidden family, whereas in our case the "forbidden family" is very general (say in Question 1.1, all ℓ-subgraphs up to some size). Yet, a certain well studied line of Turán-type problems on hypergraphs relates to Question 1.3 for very small sized ℓ-subgraphs. Brown, Erdős and Sós [BES73b, BES73a] initiated the study of the asymptotic growth of the maximum number of hyperedges in a r-hypergraph that excludes all subgraphs with e hyperedges and v vertices, for small constants v, e. A celebrated result of Ruzsa and Szemerédi [RS76] resolved the $r = 3, e = 6, v = 3$ case, that became known as *the* $(6, 3)$*-problem*, settling the answer at $o(n^2)$. Phrased in terms of Question 1.3, it states that for any $\epsilon > 0$ and sufficiently large n, density ϵn in 3-hypergraphs forces a $1/2$-subgraph of size 6. This was extended by Erdős, Frankl and Rödl [EFR86] to $1/(r-1)$-subgraphs of size $3r - 3$ for any uniformity r. These have been cornerstone results in their field: They are among the earliest applications of the Szemerédi Regularity Lemma, and their proofs contain the first versions of the Triangle and Graph Removal Lemmas; all three are by now recognized as highly consequential in various areas of Mathematics and Computer Science. For some more recent bounds of this flavour, see Alon and Shapira [AS06].

1.2 Our Results

ℓ-Subgraphs. We give nearly tight bounds on the asymptotic value of $S_\ell(n, d)$, placing it slightly above $n/d^{\ell/(\ell-1)}$. Generally we show for every $\epsilon > 0$,

$$\Omega(n/d^{\frac{\ell}{\ell-1}}) \;\le\; S_\ell(n, d) \;\le\; O(n/d^{\frac{\ell}{\ell-1} - \frac{1}{\ell^3 - 2\ell + 1} - \epsilon})$$

(i) Lower bounds (Section 5). For all densities d we show $S_\ell(n, d) = \Omega(n/d^{\ell/(\ell-1)})$. For high densities we show this bound holds even for regular graphs. In the highest density case we show $S_\ell(n, d) = \omega(n/d^{\ell/(\ell-1)})$.

(ii) Upper bounds (Sections 3 and 4). We focus on the $\ell = 2$ case for concreteness. The above lower bound is then n/d^2, and we attempt to establish its tightness in two natural senses: The first is by making a small compromise on the edge-to-vertex ratio of the target subgraph, from 2 to $2 - \epsilon$. We prove that for every $\epsilon > 0$, there are $(2 - \epsilon)$-subgraphs of size $O(n/d^2)$. That is, $S_{2-\epsilon}(n, d) = O(n/d^2)$.

The other sense of tightness is a small compromise on the size of the target subgraph, from n/d^2 to $n/d^{2-\epsilon}$. This turns out far more challenging to analyse, and constitutes the main part of our work. Our main result is the following:

Theorem (main; informal). *Let $\epsilon > 0$. Every graph of size n and density d contains a 2-subgraph of size $O(n/d^{2-\epsilon})$ if,*

- $\epsilon > 1/5$*, or*
- $d \leq n^{1/\Theta(\log(1/\epsilon))}$*, or*
- *G has girth $\geq \Theta(\log(1/\epsilon))$.*

(The constants suppressed in the Θ notation are very small.)

We conjecture that $S_2(n,d) = O(n/d^{2-\epsilon})$ for every $\epsilon > 0$, i.e. that none of the bullets is necessary. The first bullet may be rephrased as $S_2(n,d) = O(n/d^{1.8-\epsilon'})$ for every $\epsilon' > 0$. The second and third bullets close the gap towards our conjecture gradually, by either density or girth; in particular, the conjecture is proven for low densities ($d \leq \operatorname{polylog}(n)$) and for graphs with super-constant girth.

ℓ-Girth (Section 6). As immediate consequences of the above, we get for every $\epsilon > 0$:

$$\Omega(n/d^{1+\frac{2}{\ell-2}}) \quad \leq \quad g_\ell(n,d) \quad \leq \quad O(n/d^{1+\frac{1}{\ell-2}-\frac{1}{\ell^3-3\ell^2+\ell+2}-\epsilon})$$

The upper bound is a significant improvement over $O(n/d)$ by [EFRS90]. The lower bound is the first we are aware of. For concreteness, let us write the bounds for 3-girth:

$$\Omega(n/d^3) \quad \leq \quad g_3(n,d) \quad \leq \quad O(n/d^{1.8-\epsilon})$$

As in the case of $S_2(n,d)$, the upper bound improves gradually by either density or girth, and reaches $O(n/d^{2-\epsilon})$ for graphs with polylogarithmic density or with super-constant girth.

Hypergraphs (Section 7). We show how upper bounds on $S_\ell(n,d)$ extend to hypergraphs, and in particular we derive results for the $\ell = \frac{2}{3}$ case in 3-hypergraphs. (Recall this is the minimum edge-to-vertex ratio of any even cover.)

For every $\epsilon > 0$, we get an upper bound of $O(n/d^2)$ on the size of $(\frac{2}{3} - \epsilon)$-subgraphs, and $O(n/d^{1.8-\epsilon})$ on the size of $\frac{2}{3}$-subgraphs. The latter bound gradually improves as the density lowers and reaches $O(n/d^{2-\epsilon})$ for polylogarithmic densities, which nearly meets a conjectured bound of $\tilde{O}(n/d^2)$ by Feige [Fei08, Conjecture 1.7].

2 Preliminaries

In this section we record some simple facts that we will be of repeated use in our proofs.

2.1 Concentration Lemmas

We will often need to argue that with sufficiently high probability, a random variable does not stray too far away from its expected value. We begin with a lemma for arbitrarily distributed random variables, with bounds on the support.

Lemma 2.1. *Let X be distributed over $0, \dots, n$ with $\mathbb{E}[X] = \mu$. Then $\Pr[X \geq \frac{1}{2}\mu] \geq \frac{\mu}{2n}$.*

Proof. For each $i = 0, \dots, n$ denote $p_i = \Pr[X = i]$, and let $p = \sum_{i \geq \frac{1}{2}\mu} p_i$. We have,

$$\mu = \sum_{i=0}^{n} p_i i = \sum_{i<\frac{1}{2}\mu} p_i i + \sum_{i\geq\frac{1}{2}\mu} p_i i \leq \sum_{i<\frac{1}{2}\mu} p_i \frac{1}{2}\mu + \sum_{i\geq\frac{1}{2}\mu} p_i n = \frac{1}{2}\mu(1-p) + np \leq \frac{1}{2}\mu + np$$

Rearranging gives $p \geq \frac{\mu}{2n}$. □

The next lemma offers a trade-off between the multiplicative deviation from the expectation, and the probability for that deviation.

Lemma 2.2. *Let X be a non-negative random variable. There is an integer $t \geq 1$ such that $\Pr[X \geq \frac{2}{9} \cdot 2^t \cdot \mathbb{E}X] \geq (2^t t^2)^{-1}$. Furthermore if M is an upper bound on X, then $t \leq \log(\frac{9}{2}M/\mathbb{E}[X])$.*

Proof. Suppose no t satisfies the statement, then

$$\mathbb{E}[X] \leq \frac{2\mathbb{E}[X]}{9} + \sum_{t\geq 1} \frac{2\mathbb{E}[X]}{9} 2^{t+1} \frac{2^{-t}}{t^2} = \frac{2\mathbb{E}[X]}{9}(1 + 2\sum_{t\geq 1}\frac{1}{t^2}) = \frac{2\mathbb{E}[X]}{9}(1 + 2\frac{\pi^2}{6}) < \mathbb{E}[X],$$

a contradiction. This proves the first assertion of the lemma. The second assertion follows, because a value of $\frac{2}{9} \cdot 2^t \cdot \mathbb{E}[X]$ or larger is now known to be in the support of X, and hence it cannot exceed the bound M. Rearranging $\frac{2}{9} \cdot 2^t \cdot \mathbb{E}[X] \leq M$ gives $t \leq \log(\frac{9}{2}M/\mathbb{E}[X])$. □

We will also need the following case of the Chernoff bound.

Lemma 2.3 (a Chernoff bound)**.** *If X is binomially distributed, then,*

$$\Pr[X < 2\mathbb{E}[X]] \geq 1 - (0.25e)^{\mathbb{E}[X]}.$$

The proof is well known and can be found, for example, in [WS11, Theorem 5.24].

2.2 Minimum Degree Guarantees

Following is a sequence of lemmas that show how in various situations, we can ensure a lower bound on the degrees of a subset of vertices in a graph.

Lemma 2.4. *A graph with average degree d has a subgraph with minimum degree $\geq \lfloor \frac{1}{2}d+1 \rfloor$.*

Proof. Let G be a graph on n vertices and average degree d, so $\frac{1}{2}dn$ edges. Iteratively, as long as there are vertices with degree $\leq \frac{1}{2}d$, pick an arbitrary one and remove it from the graph. The resulting subgraph H has minimum degree greater than $\frac{1}{2}d$ as long as it is non-empty.

Supposing by contradiction that H is empty, we have performed n iterations; in each one we have removed at most $\frac{1}{2}d$ edges, and in the last one no edges were removed (as the graph then contained only a single vertex). Altogether we have removed at most $\frac{1}{2}d(n-1)$ edges, less than the total number of edges in G, and hence there are surviving edges H, which contradicts its being empty. $\square$

Lemma 2.5 (half-matching)**.** *Let $G(V, U; E)$ be a bipartite graph such that each $v \in V$ has degree d, and each $u \in U$ has degree at most 2. There is a subset of edges E' such that in $G'(V, U; E')$, each $v \in V$ has degree at least $\lfloor \frac{1}{2}d \rfloor$, and each $u \in U$ has degree at most 1.*

Proof. Construct an auxiliary bipartite graph $H(V_H, U; E_H)$ from G, by replacing each $v \in V$ with $\lfloor \frac{1}{2}d \rfloor$ copies, each connected to the neighbours of v in U. For $W \subset V_H$, each $w \in W$ has d outgoing edges, and each $u \in U$ has (in H) at most $2\lfloor \frac{1}{2}d \rfloor$ incoming edges, so W has neighbourhood of size at least $N(W) \geq d|W|/2\lfloor \frac{1}{2}d \rfloor \geq |W|$. Hence H satisfies the condition of Hall's Theorem, and thus has a perfect matching $E'_H \subset E_H$. Re-unify all copies of each vertex $v \in V$ into a single vertex. The resulting subgraph of G is $G'(V, U; E')$. $\square$

Corollary 2.6 (edge orientation)**.** *Let G be a graph with minimum degree d. Its edges can be oriented such that each vertex has at least $\lfloor \frac{1}{2}d \rfloor$ edges oriented towards it.*

Proof. Consider the bipartite *incidence graph* B_G of G which has sides V and E, and $v \in V$, $e \in E$ are adjacent in B_G iff e is incident to v in G. Apply Lemma 2.5 on B_G to get an assignment of each edge in G to at most one of its end vertices. Orient the edges according to that assignment. $\square$

Lemma 2.7. *Let $G(V, U; E)$ be a bipartite graph such that side V has average degree d and maximum degree D. There is a subset $V' \subset V$ with size $|V'| \geq \frac{d}{2D}|V|$ such that each $v \in V'$ has degree $\geq \frac{1}{2}d$.*

Proof. By Lemma 2.1 with $\mu = d$, $n = D$, and X the degree of a random vertex in V. $\square$

3 Degree Compromise

As a first step towards tackling our main problem, we consider a relaxation that allows a small compromise on the target average degree: Instead of seeking ℓ-subgraphs, we settle for $(\ell - \epsilon)$-subgraphs for an arbitrarily small $\epsilon > 0$. We give the following result, stating that $S_{\ell-\epsilon}(n, d) = O(n/d^{\ell/(\ell-1)})$ (To see this, plug $\Delta = 2\ell$ in the statement). Note that the constant hidden in the big O notation depends on ϵ.

Theorem 3.1 (degree compromise)**.** *Let $\Delta > 1$ be an integer and $\epsilon > 0$. There is a constant $C = C(\Delta, \epsilon)$, such that every graph on n vertices with average degree d (satisfying $\Delta - \epsilon \le d \le O(n^{(\Delta-2)/\Delta})$) contains a subgraph of size at most $C \cdot n/d^{\Delta/(\Delta-2)}$ with average degree $\ge \Delta - \epsilon$. (So at least $\frac{1}{2}(\Delta - \epsilon)$ edges per vertex.)*

The proof is fairly straight-forward, and we use it to lay the foundations towards the more involved proofs of Section 4. To keep the presentation simple, we focus on the following special case that captures the main ideas. It is restricted to $\ell = 2$ and to only a partial range of the possible densities. The full proof of Theorem 3.1 is conceptually similar but rather cumbersome, and is deferred to Appendix B.

Theorem 3.2. *Let $\epsilon > 0$ be arbitrary. There is a constant $C = C(\epsilon)$ such that every graph with n vertices and average degree d satisfying $4 - \epsilon \le d = o(\sqrt{n/\log n})$, contains a subgraph of size at most $C \cdot n/d^2$ with average degree $\ge 4 - \epsilon$.*

We begin with an overview of the proof. The idea is to sample a random subset A of vertices and then count certain induced sub-structures. We "mark" each vertex with independent probability $p = \alpha/d^2$ (where α is a large constant that depends on ϵ), thus including it A. We denote by B the subset of vertices that have two marked neighbours. Each vertex is included in B with probability $\binom{d}{2}p^2$ (for picking two of its neighbours and marking them), which roughly equals αp, so we get $\mathbb{E}|A| = np$ and $\mathbb{E}|B| = \alpha np$. Note that both sizes are $O(n/d^2)$. By probabilistic existence arguments, there is a marking of vertices that realizes these expected sizes.

In the subgraph H induced by A and B, each vertex of B comes with two edges connecting it to A, thus "paying for itself" towards the end of attaining average degree 4. The vertices in A are not paid for, but since $|B|/|A| = \alpha$, their number can be reduced to an arbitrarily small constant fraction of the vertices in H by choosing α sufficiently large (namely, $\alpha \sim 1/\epsilon$). This brings the average degree of H arbitrarily close to 4.

The remainder of this section is dedicated to the formal proof.

Proof of Theorem 3.2

Let $G(V, E)$ be a graph as in the statement of the theorem. By Lemma 2.4 we may assume, up to a slight variation of constants, that G has minimum degree d (by applying the proof on the subgraph given in the lemma). Moreover it is enough to prove the theorem for all sufficiently large values of d, as the lower values can then be handled by a proper choice of constant C. We will use $o(1)$ to denote a term that tends to 0 as d grows.

Let α be a large constant that will be determined later. Sample a random subset $A \subset V$ by including each vertex in A with independent probability $p = \alpha/d^2$. We refer to vertices in A as *marked*. Note that $|A|$ is binomially distributed with parameters n, p, and $\mathbb{E}|A| = \alpha n/d^2$.

For each vertex v we fix an arbitrary subset $N(v)$ of exactly d of its neighbours. Define B to be the random subset of vertices v that are not marked, and have (exactly) two marked neighbours in $N(v)$. We then have,

$$\Pr[v \in B] = (1-p) \cdot \binom{d}{2} p^2 (1-p)^{d-2} = \frac{1}{2}(1-o(1))p^2 d^2 = \frac{1-o(1)}{2}\alpha^2/d^2$$

where the middle equality holds since $(1-p)^{d-1} = (1-\frac{1}{d^2})^{d-1} = 1-o(1)$. Hence by linearity of expectation, $\mathbb{E}|B| = \frac{1-o(1)}{2}\alpha^2 n/d^2$.

By the Chernoff bound (Lemma 2.3) applied to $|A|$, we get:

$$\Pr\left[|A| < 2\mathbb{E}|A|\right] \geq 1 - (0.25e)^{\alpha n/d^2} > 1 - o(d^{-2})$$

where the final inequality is by the assumption $d = o(\sqrt{n/\log n})$ and by choosing α sufficiently large. On the other hand, by Lemma 2.1, $|B|$ attains half its expected value with probability at least $\frac{\mathbb{E}|B|}{2n} = \frac{1-o(1)}{4}\alpha^2/d^2 = \Omega(d^{-2})$. Summing the bounds yields:

$$\Pr\left[|A| \leq 2\mathbb{E}|A|\right] + \Pr\left[|B| \geq \frac{1}{2}\mathbb{E}|B|\right] \geq 1 - o(d^{-2}) + \Omega(d^{-2})$$

The bound on the right-hand side is strictly more than 1 for sufficiently large d, and hence there is a positive probability that both of the events $|A| \leq 2\mathbb{E}[|A|]$ and $|B| \geq \frac{1}{2}\mathbb{E}[|B|]$ occur. We fix this event from now on, and arbitrarily remove vertices from B until $|B| = \frac{1}{2}\mathbb{E}[|B|]$. The following bounds now hold:

$$|A| + |B| \leq 2\mathbb{E}|A| + \frac{1}{2}\mathbb{E}|B| = \left(2\alpha + \frac{1-o(1)}{4}\alpha^2\right) \cdot \frac{n}{d^2} \tag{3.1}$$

$$\frac{|B|}{|A|} \geq \frac{\frac{1}{2}\mathbb{E}|B|}{2\mathbb{E}|A|} = \frac{1-o(1)}{8} \cdot \alpha \tag{3.2}$$

We take our target subgraph H to be that induced by $A \cup B$. By eq. (3.1), its size is bounded by $C \cdot n/d^2$ for, say, $C = \frac{1}{2}\alpha^2$. To bound its average degree, note that each vertex in B is incident to two edges connecting it to A, and since A and B are disjoint (recall that vertices in B are not marked), each such edge has a unique end in B. Hence we count at least $2|B|$ different edges in H, and find that its average degree is:

$$\operatorname{avgdeg}(H) \geq 2 \cdot \frac{2|B|}{|A|+|B|} = 4 - \frac{4}{1+\frac{|B|}{|A|}} \geq 4 - \frac{4}{1+\frac{1-o(1)}{8}\cdot\alpha}$$

using eq. (3.2) for the final inequality. The bound on the right-hand side is guaranteed to be at least $4-\epsilon$ as long as we pick $\alpha > \frac{33}{\epsilon}$, and the proof is complete. $\square$

4 Size Compromise

4.1 Overview

We now turn to the primary goal of this work, which is to establish upper bounds on $S_\ell(n,d)$. We focus on the $\ell = 2$ case, the first non-trivial one, which is further motivated by the applications described in Section 1. An extension to general values of ℓ is also discussed.

The $\ell = 2$ case. In Section 5 we will show that $S_2(n,d) = \omega(n/d^2)$, and in Section 3 we have shown that $S_{2-\epsilon}(n,d) = O(n/d^2)$ for any $\epsilon > 0$. It is therefore natural to postulate that $S_2(n,d) = O(n/d^{2-\epsilon})$ for any $\epsilon > 0$. Let us formally record it as a conjecture.

Conjecture 4.1.1. *Let $\epsilon > 0$. There is a constant $C = C(\epsilon)$ such that every graph G with n vertices and average degree d (satisfying $4 \le d \le O(\sqrt{n})$) contains a subgraph of size at most $C \cdot n/d^{2-\epsilon}$ and average degree 4.*

We will devise a framework towards proving the conjecture, and use it to prove the following special cases:

Theorem 4.1.2 (summary of positive results[5])**.** *Conjecture 4.1.1 holds under each of the following additional conditions (separately and independently of each other):*

1. *$\epsilon > \frac{1}{5}$. (See Corollary 4.2.9)*
2. *$d = O(n^{1/t})$ where $t = \lfloor \log(\frac{8}{3}(\frac{1}{\epsilon} - 2)) \rfloor$. (See Theorem 4.4.1)*
3. *G has girth $\ge 2t - 1$ where $t = \lfloor \log(\frac{8}{3}(\frac{1}{\epsilon} - 2)) \rfloor$. (See Corollary 4.2.10)*
4. *$\epsilon > \frac{1}{11}$ and G is square-free, i.e. contains no cycles of length 4. (See Theorem 4.5.1)*

Item 1 of the theorem arises as a special case of both Item 2 and Item 3, since for $\epsilon > \frac{1}{5}$ we get $t \le 2$ (and then $d = O(n^{1/t})$ holds by hypothesis, and girth ≥ 3 holds for any simple graph)[6]. It is stated separately for emphasis, as it is a non-trivial result that applies to all graphs.

Item 2 is the interesting one of the four. It provides a gradual improvement to the ϵ-range that we can handle according to the density d. On the highest end $d \sim \sqrt{n}$ we get Item 1, and on the low end $d = \text{polylog}\,(n)$ we get the full range of $\epsilon > 0$ (as then $d = O(n^{1/t})$ holds for arbitrarily large t). This in fact proves Conjecture 4.1.1 for polylogarithmic densities. Similarly, Item 3 proves Conjecture 4.1.1 for graphs with super-constant girth.

For the formal statements and proofs of the four items of Theorem 4.1.2, see the pointers in the theorem itself.

[5] Our proofs work for $d = O(n^{1/(2-\delta)})$ with any $\delta < \epsilon$, but for clarity we state our results with $d = O(\sqrt{n})$.

[6] Note the use of the floor operator in the value of t. Plugging $\epsilon = 0.2$ yields $t = 3$; plugging $\epsilon > 0.2$ yields $\log(\frac{8}{3}(\frac{1}{\epsilon} - 2)) < 3$, and hence $t \le 2$.

General ℓ values. The proofs in this section can be applied to any $\ell \geq 2$ with straightforward adjustments. This is a merely technical treatment and is mostly omitted from this work; we explicitly extend only the result for general graphs, i.e. Item 1 of Theorem 4.1.2. In accordance with Conjecture 4.1.1, we postulate that a bound of $S_\ell(n,d) = O(n/d^{\ell/(\ell-1)-\epsilon})$ for any $\epsilon > 0$ should hold, but only manage to prove it for $\epsilon > 1/(\ell^3 - 2\ell + 1)$. Note that the gap between this upper bound (formalized in the next theorem) and the lower bounds given in Section 5 becomes narrower as ℓ grows.

Theorem 4.1.3. *Let $\ell > 1$ be an integer and $\epsilon > 0$. There is a constant $C = C(\ell, \epsilon)$ such that every graph on n vertices with average degree d (satisfying $2\ell \leq d \leq O(n^{(\ell-1)/\ell})$) contains a subgraph of size at most $C \cdot n/d^{\frac{\ell}{\ell-1} - \frac{1}{\ell^3-2\ell+1} - \epsilon}$ with average degree 2ℓ.*

The proof is given in Appendix C. Apart from demonstrating how the extension to general ℓ can be done, it can be beneficial as a simplified, self-contained proof of Item 1 of Theorem 4.1.2 (outside the framework that allows the proofs of the other items).

Organization of this section. In Section 4.2 we present the framework, which is a reduction to a setting similar to that used in Section 3. Items 1 and 3 of Theorem 4.1.2 will follow as corollaries. The reduction itself is proven in Section 4.3. Items 2 and 4 of Theorem 4.1.2 are then proven in Sections 4.4 and 4.5, respectively.

4.2 Approach

Much like the proof of Theorem 3.2, our approach is to mark each vertex in the graph with independent probability p, and then use structures induced by the marking as building-blocks in constructing the desired subgraph. The key property a structure should satisfy is that each non-marked vertex contributes two edges, thus "paying for itself" when used to construct a 2-subgraph. The building-block used in Section 3 was a non-marked vertex attached to two marked neighbours; here we will be using larger structures with the same property. For example, consider a non-marked vertex with a marked neighbour and a non-marked neighbour, the latter having two marked neighbours of its own. Such a structure has 2 non-marked vertices and 4 edges and hence is usable for our purposes.

In general, we can visualize structures in a natural way as full[7] binary trees in which the leaves are marked and the internal vertices are non-marked. We therefore refer to non-marked vertices in a structure as *roots*. A full binary tree with k leaves has exactly $2k-2$ edges (this is easy to verify), so in a graph with minimum degree d, we intuitively expect each vertex to be the root of a structure with k marked vertices with probability roughly $d^{2k-2}p^k$, for choosing each edge out of d possible edges and for marking the leaves. Notice how this coincides with Section 3, where each structure would occur with probability $\sim d^2p^2$.

With the notion of roots in mind, we shift our attention to a related one that we call *excited vertices*. An excited vertex is a neighbour of a root, so if we mark one of its other neighbours it would become a root by itself. In a sense, it is "half-way" into being a root. In accordance with the intuition described above, we expect each vertex to be excited with

[7] A binary tree is *full* if each vertex has either 2 or 0 children.

probability roughly $d^{2k-1}p^k$ (for any k), as we have d choices for a neighbour and probability roughly $d^{2k-2}p^k$ for that neighbour to be a root.

We formally reduce the problem of finding 2-subgraphs to showing that each vertex is indeed excited with at least that probability. This would be the case had the graph been a tree, but in general the topology of the graph may create dependencies that affect the probability. The task we face within our framework is to show that regardless of the specific graph we are given, the probability for each vertex to be excited is not significantly reduced. As an example, it is straight-forward to observe that in a graph with minimum degree d and $p = o(d^{-1})$, each vertex has probability $\geq (1 - o(1))pd$ to have a marked neighbour, i.e. to be excited with $k = 1$. In our framework this immediately implies Item 1 of Theorem 4.1.2. The larger k for which we can prove this, the closer to Conjecture 4.1.1 we would get.

Lastly, we emphasize the crucial benefit of our reduction: In Section 3, the non-marked vertices paid for themselves, but the marked vertices were not paid for. Therefore we came short of attaining average degree 4 and settled for $4 - \epsilon$. Here the entire goal is to avoid this loss. Our reduction exploits the fact that excited vertices are "half-way" into being roots, and combines them in a way that covers also for the marked vertices, thus achieving average degree 4 at the expense of a slight increase in the size of the subgraph (ϵ in Conjecture 4.1.1). This is a delicate issue that poses a main challenge in proving the reduction. It is done in a black-box fashion, so when using the framework to prove Theorem 4.1.2, we only need to count excited vertices and need not worry about paying for the marked vertices at all.

4.2.1 Definitions

Definition 4.2.1 (graph with random subset model). *Let $G(V, E)$ be a graph. For $p \in [0, 1]$, we define $G(V, E, p)$ to be a random model in which each vertex in V is marked with independent probability p.*

Definition 4.2.2 (root). *In $G(V, E, p)$, a vertex is a* 1-root *if it is marked. For an integer $k > 1$, a vertex is a* k-root *if, inductively, it has two distinct neighbours which are a k_1-root and a k_2 root, and $k_1 + k_2 = k$.*

Definition 4.2.3 (excited vertex). *In $G(V, E, p)$, for an integer $k \geq 1$, a vertex is* k-excited *if it has a k-root neighbour.*

(Remark: A vertex may be a k-root concurrently for several values of k, or for the same k due to several combinations of neighbours. The same goes for being k-excited. Moreover, a vertex may be a k-root and k-excited at the same time.)

Definition 4.2.4 (tree-like graph). *For an integer $k \geq 1$, $p \in [0, 1]$ and $\gamma > 0$, a graph $G(V, E)$ with minimum degree d is* (k, p, γ)-tree-like *if in $G(V, E, p)$, each vertex is k-excited with probability at least* $\min\{\gamma \cdot p^k d^{2k-1}, 0.99\}$.

The meaningful case that should be in mind is when d is arbitrarily large, $p \ll d^{-1}$, and k and γ are constants. In this case, the following proposition is important to note. It motivates the terminology of Definition 4.2.4.

Claim 4.2.5. *Let $t \geq 2$ be a constant integer and $k = 2^{t-2}$. Let $T(V, E)$ be a d-ary tree with d arbitrarily large, rooted by $v \in V$ and with all leaves in level t. (The root is considered to be in level 1.) Then in $T(V, E, p)$ with $p = o(d^{-1})$, the probability that v is k-excited is* $\min\{\Omega(p^k d^{2k-1}), 0.99\}$.

Proof. We only sketch the proof as it is very clear. Let $u_1, \ldots, u_d$ be the children of v, and let A_i denote the event that u_i is a k-root. We show by induction on t that $\Pr[A_i] = \Omega(p^k d^{2k-2})$: In the base case $t = 2$ we get $k = 1$, so we need to show that u_i is a 1-root w.p. p, which holds by definition. For $t > 2$, by induction, each child of u_i is a $(\frac{1}{2}k)$-root w.p. $q = \Omega(p^{k/2} d^{k-2})$, and these events are independent since T is a tree. Hence the number of $(\frac{1}{2}k)$-roots among the children of u_i is binomially distributed with parameters (d, q), and the probability that two of them are $(\frac{1}{2}k)$-roots is $\Omega(d^2 q^2) = \Omega(p^k d^{2k-2})$. This renders u_i a k-root, so the proof by induction is complete.

The events $A_1, \ldots, A_d$ are independent since T is a tree, so the number of k-roots among $u_1, \ldots, u_d$ is binomially distributed with parameters (d, p') for $p' = \Omega(p^k d^{2k-2})$. If this number is at least one then v is k-excited, and this occurs w.p. $\Omega(d, p) = \Omega(p^k d^{2k-1})$. □

We see that in the setting described above, a graph is tree-like if for the purpose of counting excited vertices in $G(V, E, p)$, it behaves roughly as if the k-neighbourhood of each vertex was a tree. We now get to our main definition.

Definition 4.2.6 (good graph)**.** *For integer $k \geq 1$, $p \in [0, 1]$ and $\gamma > 0$, a graph G is* (k, p, γ)-good *if for each vertex v in G, the graph $G - \{v\}$ is (k, p, γ)-tree-like.*

Put simply, a good graph is a "robust" tree-like graph. We will reduce Conjecture 4.1.1 to the problem of proving that graphs are good. As we will be interested in very general classes of graphs, our "goodness" proofs will only use global properties (such as minimum degree, girth, etc.) and not rely on any specific topologies. Hence we will actually be proving that the graphs under discussion are tree-like, and the robustness will follow without additional effort. See for example the proofs of Corollaries 4.2.9 and 4.2.10 below.

4.2.2 The Reduction

The reduction simply states that Conjecture 4.1.1 holds for good graphs. This is formalized in the following theorem. Its proof is deferred to Section 4.3.

Theorem 4.2.7 (reduction theorem)**.** *Let $k \geq 1$ be an integer and $\epsilon > 1/(3k + 2)$. There is a constant $C = C(k, \epsilon)$ such that for sufficiently large d the following holds: If G is a graph with $n \geq \Omega(d^2)$ vertices and minimum degree d, and G is (k, p, γ)-good with $p = 1/d^{2-\epsilon}$ and $\gamma = \Omega(1/\mathrm{polylog}\,(d))$, then G contains a subgraph of average degree 4 and size at most $C \cdot n/d^{2-\epsilon}$. (The power of the* polylog (d) *term in the bound on γ may depend on k.)*

We remark that in order to prove Theorem 4.1.2 it suffices to let γ be constant. Yet in the statement of Theorem 4.2.7 we let it be as low as $\Omega(1/\mathrm{polylog}\,(d))$, since it does not complicate the proof and may be useful in future contexts.

The following conjecture is posed just to clarify our mindset towards proving Theorem 4.1.2. In practice we will prove some very particular cases of it.

Conjecture 4.2.8. *Let $k \geq 1$ be an integer. For sufficiently large d, every graph with $n > \Omega(d^2)$ vertices and minimum degree d, either contains a subgraph of average degree 4 and constant size, or is $(k, p, \Omega(1))$-good for every $d^{-2} \ll p \ll d^{-1}$.*

Conjecture 4.1.1 is implied by Conjecture 4.2.8, by choosing k large enough so that $\epsilon > 1/(3k+2)$ and applying Theorem 4.2.7. Evidently, various weaker versions of Conjecture 4.2.8 would also suffice (for example, it is not necessary to handle the full range of p for each k).

With Theorem 4.2.7 at hand, we can prove special cases of Conjecture 4.1.1 by showing that some restricted families of graphs are good for restricted values of k. As warm-up examples, we can immediately derive the following two corollaries.

Corollary 4.2.9 (Item 1 of Theorem 4.1.2)**.** *Let $\epsilon > 0.2$. There is a constant $C = C(\epsilon)$ such that every graph G with n vertices and average degree d (satisfying $4 \leq d \leq O(\sqrt{n})$) contains a subgraph of size at most $C \cdot n/d^{2-\epsilon}$ and average degree 4.*

Proof. We prove for sufficiently large d, and the lower values can then be handled by a proper choice of constant C. Moreover by Lemma 2.4 we may assume that d is the minimum degree (as in the proof of Theorem 3.2). Set $p = d^{-(2-\epsilon)}$. In $G(V, E, p)$, the probability for each vertex to be 1-excited (that is, to have a marked neighbour) is $\geq (1 - o(1))dp$, hence G is $(1, p, 1 - o(1))$-tree-like. This holds even if we remove any single vertex from G, hence it is $(1, p, 1 - o(1))$-good. The corollary now follows from Theorem 4.2.7, as the condition $\epsilon > 1/(3k + 2)$ is met for $k = 1$ by hypothesis. □

Put equivalently, the corollary follows by simply observing that a vertex with all its neighbours form a tree with two levels, and applying the same reasoning used in Claim 4.2.5.

Corollary 4.2.10 (Item 3 of Theorem 4.1.2)**.** *Let $\epsilon > 0$. There is a constant $C(\epsilon)$ such that every graph with n vertices, average degree d (satisfying $4 \leq d \leq O(\sqrt{n})$) and girth $\geq 2t - 1$ for $t = \lfloor \log(\frac{8}{3}(\frac{1}{\epsilon} - 2)) \rfloor$, contains a subgraph of average degree 4 and size at most $C(\epsilon) \cdot n/d^{2-\epsilon}$.*

Proof. Let $G(V, E)$ be a graph as in the statement. Again we prove for sufficiently large d and assume d is the minimum degree. Let $k = 2^{t-2}$; by rearranging, one may verify that the value set for t in the statement is the smallest integer for which $\epsilon > 1/(3k + 2)$.

Since G has no cycles of length $2t - 1$ or less, the radius-$(t - 1)$ neighbourhood of each vertex is a tree. In other words, each vertex v is the root of a d-ary tree with all leaves in level t. Hence by Claim 4.2.5, v is k-excited with probability $\Omega(p^k d^{2k-1})$ for $p = d^{-(2-\epsilon)}$. This means G is $(k, p, \Omega(1))$-tree-like, and the argument remains intact (up to a small change of constants) even if we remove any single vertex from G, so it is $(k, p, \Omega(1))$-good. The corollary now follows from Theorem 4.2.7. □

4.3 Proof of the Reduction

In this section we prove Theorem 4.2.7. Let us first present an overall description of the proof. As explained in Section 4.2, we use roots in $G(V, E, p)$ to identify small structures with two edges per non-marked vertex, and our plan is to use them as building-blocks to construct a 2-subgraph. These structures are referred to as *arrangements* and are discussed in Section 4.3.1.

The key to the proof is to combine arrangements in a way that earns an additional edge, that would eventually be used to pay for the marked vertices. This is done in the main lemma, proven in Section 4.3.2. It states that with sufficiently high probability, we can identify structures in which one vertex contributes *three* edges, and not just two. That extra edge makes all the difference in achieving average degree 4.

In Section 4.3.3 we prove Theorem 4.2.7, in a way quite similar to the proof of Theorem 3.2. However, one additional point requires attention: The structures we use to build our target subgraph are rather large and may overlap in edges, causing us to pay with the same edge for two different vertices (its two endpoints). In Section 3 this issue did not arise, as each edge in a structure had only one non-marked endpoint and was counted to pay for it. To solve this now, we equip each arrangement with an orientation of its edges and decide that each edge pays for its destination vertex.

4.3.1 Arrangements in $G(V, E, p)$

We now formalize the notion explained above, of structures that have two edges per each non-marked vertex.

Definition 4.3.1 (arrangement; arrangeable subgraph). *Given a fixed marking of vertices sampled from $G(V, E, p)$, an* arrangement *is a pair (G', O) of a subgraph $G'(V', E')$ of G, and an orientation O of the edges in E' such that each non-marked vertex in V' has at least two edges oriented towards it.*

G' is an arrangeable subgraph *if there is an orientation O of its edges such that (G', O) is an arrangement.*

The following lemma states that under a fixed marking of vertices, arrangeable subgraphs are closed under union.

Lemma 4.3.2 (union of arrangements). *Given a fixed marking of vertices sampled from $G(V, E, p)$, let H_1 and H_2 be arrangeable subgraphs in G. The union subgraph $H = H_1 \cup H_2$ is arrangeable.*

Proof. Let O_1, O_2 be orientations of the edges in H_1, H_2 respectively, such that (H_1, O_1) and (H_2, O_2) are arrangements. We orient the edges in H as follows: For an edge e, if it is present in H_1 then we orient it by O_1. Otherwise e is present in H_2, and then we orient it by O_2. Call the resulting orientation O.

To see that (H, O) is an arrangement, let $v \in H$ be any vertex. If $v \in H_1$ then some two edges $e, e' \in H_1$ are oriented towards v in O_1, and by the above they are oriented towards v also in O. Otherwise we have $v \in H_2$, and then some two edges $e, e' \in H_2$ are oriented towards v in O_2. Neither of them can be in H_1 (as $v \notin H_1$), so they are oriented in O as they are in O_2, towards v. Thus, every vertex in H has two edges oriented towards it by O. □

The next lemma states that k-roots are part of arrangements with small (constant) size.

Proposition 4.3.3. *Let $k \geq 1$ be an integer. Given a fixed marking of vertices sampled from $G(V, E, p)$, if v is a k-root, then v is part of an arrangeable subgraph with at most $k-1$ non-marked vertices.*

Proof. By induction on k: The base case $k = 1$ is trivial, since a 1-root is just a marked vertex, which by itself constitutes an arrangement with 0 non-marked vertices.

For $k > 1$, by definition v has two distinct neighbours u_1 and u_2 which are a k_1-root and a k_2-roots respectively, such that $k_1 + k_2 = k$. By induction, u_1 is part of an arrangeable subgraph H_1 with at most $k_1 - 1$ non-marked vertices, and u_2 is part of an arrangeable subgraph H_2 with at most $k_2 - 1$ non-marked vertices.

Let H be the union of H_1 and H_2; it is an arrangeable subgraph by Lemma 4.3.2, and it has at most $k - 2$ non-marked vertices. Let O be an orientation such that (H, O) is an arrangement. If $v \in H$, then we are done. Otherwise, we add v to H together with the two edges connecting v to u_1 and u_2, orienting both towards v. (Note that $v \notin H$ implies that the edges vu_1 and vu_2 are not already orientated by O, so we can orient them as we wish.) We have added a single non-marked vertex to H, so it now has at most $k - 1$ non-marked vertices, as required. □

4.3.2 Main Lemma: That Extra Edge

Lemma 4.3.4 (main). *Let $k \geq 1$ be an integer, $M > 0$ an arbitrary constant, and $\epsilon > 1/(3k+2)$. For sufficiently large d, the following holds: If $G(V, E)$ is a graph with minimum degree d that is (k, p, γ)-good for $p = d^{-(2-\epsilon)}$ and $\gamma = \Omega(1/\operatorname{polylog}(d))$, then in $G(V, E, 2p)$, each vertex $v \in V$ has probability $\geq M \cdot p$ to be part of an arrangement H_v with the following properties:*

- *v has three edges oriented towards it in H_v.*
- *H_v has at most $3k + 1$ non-marked vertices.*

(The power of the $\operatorname{polylog}(d)$ *term in the bound on γ may depend on k.)*

Proof. We assume d is sufficiently large wherever needed, possibly without explicitly stating so. Moreover in order to simplify the presentation, we do not attempt to optimize the constants involved in the proof.

Fix a vertex $v \in V$ for the remainder of the proof. We begin by removing v from G, along with all of its incident edges. Since G is (k, p, γ)-good, the resulting graph $G - \{v\}$ is (k, p, γ)-tree-like. To ease notation, we will refer to $G - \{v\}$ as $G(V, E)$, keeping in mind that it no longer contains v.

To sample from $G(V, E, 2p)$, we mark the vertices of G in two independent phases, each with probability p. That is, a vertex that was not marked in the first phase has probability p to be marked in the second phase, hence total probability of $2p - p^2$ to be marked at all. Since $p \ll 1$, this is equivalent to $G(V, E, 2p)$ up to a small variation of constants. Importantly, note that the first phase is a sample from $G(V, E, p)$.

Let $N(v)$ denote the set of neighbours of v. We restrict our attention to exactly d neighbours of v (arbitrarily chosen), so $|N(v)| = d$. Our current (and main) goal is to show that with probability $M \cdot p$, three vertices in $N(v)$ are $(k + 1)$-roots after the second phase.

Preliminaries. We begin by orienting the edges incident to vertices in $N(v)$. In principle we would like all edges incident to $u \in N(v)$ to point towards u, but there might be edges with two ends in $N(v)$. Bypassing this issue is a mere technicality: Since each vertex in $N(v)$ has at least d incident edges,[8] we can use Corollary 2.6 to find an orientation by which each vertex has $\frac{1}{2}d$ edges oriented towards it. Fix this orientation henceforth.

Consider an arbitrary vertex $u \in N(v)$, and let $e_i = w_i u$ be the edges oriented towards it, for $i = 1, \ldots, \frac{1}{2}d$. Suppose u is k-excited after the first phase, which means it has a k-root neighbour w. In this case, we call each edge $e_i = uw_i$ with $w_i \neq w$ an *excited* edge. So if u is k-excited after the first phase, it is incident to $\frac{1}{2}d$ excited edges[9] that are oriented towards it. (There may be additional excited edges incident to u but oriented away from it, towards other vertices in $N(v)$.)

The source vertex w_i of an excited edge e_i will be called a *touched* vertex. Observe that if w_i is marked in the second phase, then u becomes a $(k+1)$-root. This is illustrated in fig. 1. Lastly, we say that a touched vertex is *light* if it is the source of one or two excited edges, and *heavy* if it is the source of at least three excited edges.

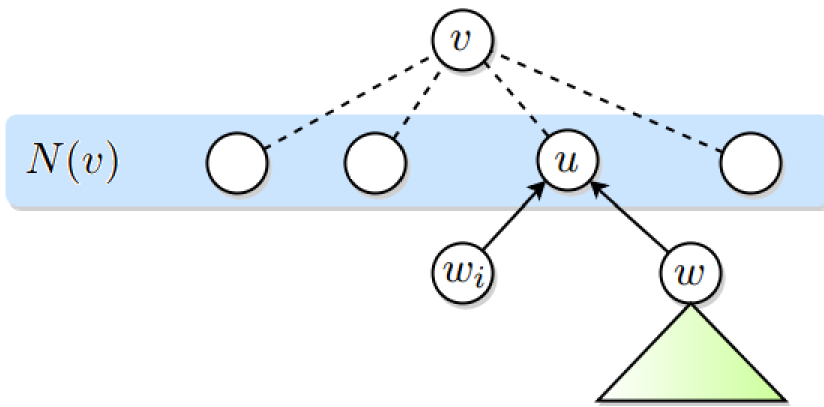

Figure 1: After the first phase, w is a k-root, rendering u a k-excited vertex in $N(v)$. The edge $w_i u$ is thus *excited*, and w_i is a *touched* vertex (which is either heavy or light, depending on how many additional excited edge are oriented away from it). If w_i is marked in the second phase, u would become a $(k+1)$-root.

Counting excited edges. Let X denote the subset of k-excited vertices in $N(v)$ after the first phase. Since G is (k, p, γ)-tree-like, each vertex has probability $q = \gamma p^k d^{2k-1}$ to be k-excited, and since $|N(v)| = d$, we get $\mathbb{E}|X| = dq$.

Let Y be the set of excited edges. As each vertex in X renders $\frac{1}{2}d$ edges excited, we have

$$|Y| = \frac{1}{2}d|X| \tag{4.1}$$

and therefore,

$$\mathbb{E}|Y| = \frac{1}{2}d\mathbb{E}|X| = \frac{1}{2}qd^2 \tag{4.2}$$

[8]In fact $d-1$ edges, having removed v from G, but we suppress the -1 for simplicity.

[9]In fact u is incident to either $\frac{1}{2}d$ (if w is one of the w_i's) or $\frac{1}{2}d - 1$ (otherwise) excited edges, but again we suppress the -1.

Applying Lemma 2.2 to $|Y|$, we get:

$$\Pr[|Y| \geq \frac{1}{9} \cdot 2^t q d^2] \geq \frac{1}{2^t t^2} \tag{4.3}$$

for some integer $t \geq 1$.

Proposition 4.3.5. *$t < 2\log d$.*

Proof. We take the bound M in the statement Lemma 2.2 of to be $\frac{1}{2}d^2$, as this is the total number of edges that can be excited ($\frac{1}{2}d$ edges per vertex in $N(v)$). Hence:

$$t < \log(\frac{9}{2}M/\mathbb{E}|Y|) \leq \log(9/2q)$$

(plugging eq. (4.2) for $\mathbb{E}|Y|$.) And by recalling that $p = d^{-(2-\epsilon)}$ and $\gamma = \Omega(1/\text{polylog}\,(d))$:

$$q = \gamma p^k d^{2k-1} = \Omega(1/\text{polylog}\,(d))\frac{d^{k\epsilon}}{d} > \frac{1}{d}$$

(for sufficiently large d.) Combining the latter two inequalities, we obtain: $t < \log(\frac{9}{2}) + \log d < 2\log d$. □

Recall that an excited edge has a touched vertex as its source. Each touched vertex is either light or heavy, so for each fixed marking of the vertices in phase 1, either half the excited edges are sourced in light vertices, or half are sourced in heavy vertices. By applying an averaging argument on eq. (4.3), we get that with probability $\frac{1}{2}(2^t t^2)^{-1}$, one of these two cases holds concurrently with the event $|Y| \geq \frac{1}{9} \cdot 2^t q d^2$. We now handle each case separately.

Case 1 - Light vertices: With probability $\frac{1}{2}(2^t t^2)^{-1}$, we have at least $\frac{1}{9} \cdot 2^t q d^2$ excited edges, half of which are sourced in light vertices.

In this case, we want the second phase to turn three k-excited vertices in $N(v)$ into $(k+1)$-roots by marking a light vertex for each. See fig. 2 for illustration. Let L be the set of light vertices. Since each light vertex is incident to at most two excited edges, we have $|L| \geq \frac{1}{2}|Y| = \frac{1}{4}d|X|$, where the last equality is by eq. (4.1).

Our intention is now to uniquely assign light vertices to k-excited vertices in $N(v)$ (i.e. those in X). Towards this end, we consider the bipartite graph with sides X and L and with the excited edges as the edge set. In fact, X and L may intersect; in such case we make two copies of each vertex in the intersection, putting one copy on the X-side and the other on the L-side. Note that all the edges are oriented from L to X.

Side X has average degree $|X|/|L| \geq \frac{1}{4}d$ and maximum degree $\frac{1}{2}d$ (as each vertex in X is the end of only $\frac{1}{2}d$ excited edges), so by Lemma 2.7, there is a subset $X' \subset X$ with size $|X'| \geq \frac{1}{4}|X|$, such that each vertex in X' is adjacent to at least $\frac{1}{8}d$ vertices in L. Now consider the bipartite graph with sides X' and L: Side X' has minimum degree $\frac{1}{8}d$, and side L has degree at most 2 (since a light vertex is adjacent to at most two vertices in X, and hence in X'). So by applying Lemma 2.5, we can assign the vertices in L to the vertices in X' in a way that each L-vertex is assigned to at most one X'-vertex, and each X'-vertex has at least $\frac{1}{16}d$ L-vertices assigned to it. This concludes the assignment of light vertices to (a large subset of the) k-excited vertices.

The event we are interested in, denoted henceforth as $\mathcal{A}$, is that the second phase turns three vertices in X into $(k+1)$-roots. Given the assignment we have just worked out, it is sufficient to pick three vertices in X' and for each of them to mark in the second phase an assigned vertex in L. Since each vertex in X' has $\frac{1}{16}d$ uniquely assigned vertices in L, the probability to mark one of them is $\frac{1}{16}dp(1-p)^{d/16}$, which is lower-bounded by $\frac{1}{32}dp$ for sufficiently large d. Moreover these events are independent for distinct vertices in X' (by the unique assignment), hence in total we get:

$$\Pr[\mathcal{A}] \geq \frac{1}{2 \cdot 2^t t^2} \cdot \binom{|X'|}{3} \cdot \left(\frac{1}{32} d \cdot p\right)^3 \geq \frac{1}{2 \cdot 2^t t^2} \cdot \frac{|X'|^3}{3^3} \cdot \left(\frac{1}{32} d \cdot p\right)^3 \tag{4.4}$$

(Recall that $\binom{a}{b} \geq (\frac{a}{b})^b$ for any integers a, b.) To lower-bound eq. (4.4), we recall that $|X'| \geq \frac{1}{4}|X|$, that $|X| = \frac{2}{d}|Y|$ (by eq. (4.1)) and that $|Y| \geq \frac{1}{9} \cdot 2^t q d^2$ (by the assumption of the current case). Putting these together, we get $|X'| \geq \frac{1}{18} \cdot 2^t q d$, and plugging into eq. (4.4):

$$\Pr[\mathcal{A}] \geq \frac{1}{2 \cdot (3 \cdot 32 \cdot 18)^3} \cdot \frac{2^{2t}}{t^2} \cdot (qd^2 \cdot p)^3$$

Next we recall that $1 \leq t < 2 \log d$ and $q = \gamma p^k d^{2k-1} = \Omega(1/\text{polylog}\,(d)) \cdot p^k d^{2k-1}$. Plugging these into the above, we get:

$$\Pr[\mathcal{A}] \geq \frac{(p^{k+1} d^{2k+1})^3}{\Omega(\text{polylog}\,(d))}$$

We need this bound to be at least $M \cdot p$. Suppressing the constant M into the Ω notation, we need the following to hold:

$$p^{3k+2} d^{6k+3} \geq \Omega(\text{polylog}\,(d))$$

Recalling that $p = d^{-(2-\epsilon)}$, the latter is rewritten as:

$$d^{(3k+2)\epsilon - 1} \geq \Omega(\text{polylog}\,(d))$$

For this to hold for sufficiently large d, it is enough to require:

$$(3k+2)\epsilon - 1 > 0$$

and this holds by the hypothesis of the lemma. Hence we have obtained: $\Pr[\mathcal{A}] \geq M \cdot p$.

Case 2 - Heavy vertices: With probability $\frac{1}{2}(2^t t^2)^{-1}$, we have at least $\frac{1}{9} \cdot 2^t q d^2$ excited edges, half of which are incident to heavy vertices.

Let L be the subset of heavy vertices. Each such vertex is incident to at most d excited edges (as there are only d vertices in $N(v)$), so we have $|L| \geq \frac{1}{18} \cdot 2^t q d$.

Again we let $\mathcal{A}$ denote the event that the second phase turns three vertices in X into $(k+1)$-roots. A heavy vertex is the source of at least three excited edges, so it is adjacent to three k-excited vertices in $N(v)$. Hence, marking any heavy vertex in the second phase is

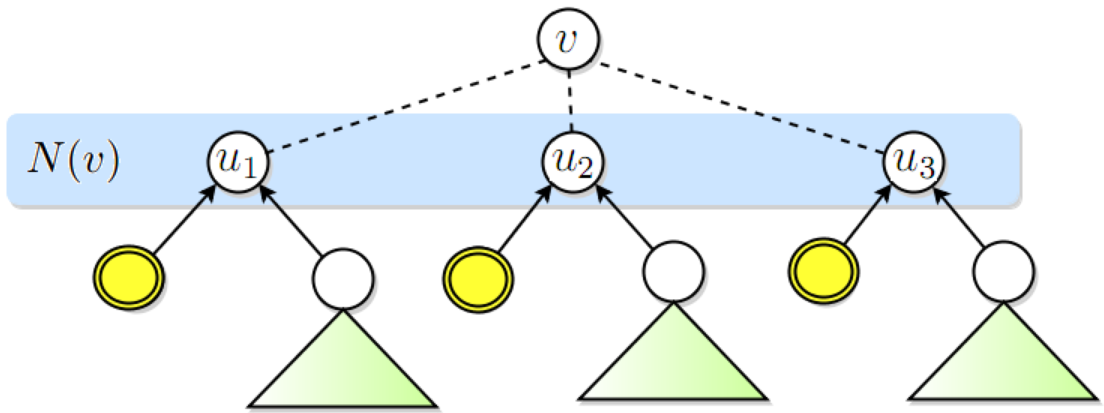

Figure 2: u_1, u_2, u_3 were k-excited after the first phase. In the second phase, marking a light vertex (in double circle) for each turns them into a $(k+1)$-roots.

enough for $\mathcal{A}$ to occur. See fig. 3 for illustration. The probability to mark a heavy vertex in the second phase is $(1-o(1))|L|p$, and hence:

$$\Pr[\mathcal{A}] \geq \frac{1}{2 \cdot 2^t t^2} \cdot (1-o(1))|L|p \geq \frac{(1-o(1))qdp}{36 \cdot t^2}$$

Plugging $t < 2\log d$ and $q = \gamma p^k d^{2k-1} = \Omega(1/\text{polylog}\,(d)) \cdot p^k d^{2k-1}$, we get

$$\Pr[\mathcal{A}] \geq \frac{p^{k+1} d^{2k}}{\Omega(\text{polylog}\,(d))}$$

We need this bound to be at least $M \cdot p$. Suppressing the constant M into the Ω notation, we need the following to hold:

$$p^k d^{2k} \geq \Omega(\text{polylog}\,(d))$$

Recalling that $p = d^{-(2-\epsilon)}$, the latter is equivalent to:

$$d^{\epsilon k} \geq \Omega(\text{polylog}\,(d))$$

and this clearly holds for sufficiently large d. Hence we have obtained: $\Pr[\mathcal{A}] \geq M \cdot p$.

Conclusion. In both of the above cases, we have shown that with probability at least $M \cdot p$ over the two phases, there are three $(k+1)$-roots in $N(v)$. By Proposition 4.3.3, each such root is part of an arrangement with at most k non-marked vertices, and by Lemma 4.3.2 we can unite the three arrangements. In summation, we've shown that with probability $M \cdot p$, the graph $G - \{v\}$ has an arrangement H'_v that contains three neighbours of v, say u_1, u_2, u_3, and has at most $3k$ non-marked vertices.

The claim now follows easily: All the above was shown to hold for $G - \{v\}$, and hence holds also for G with the guarantee that $v \notin H'_v$. We therefore can construct an arrangement H_v as required in the claim, by adding v to H'_v, and orienting towards it the three edges connecting it to u_1, u_2, u_3. The proof of the claim is complete. $\square$

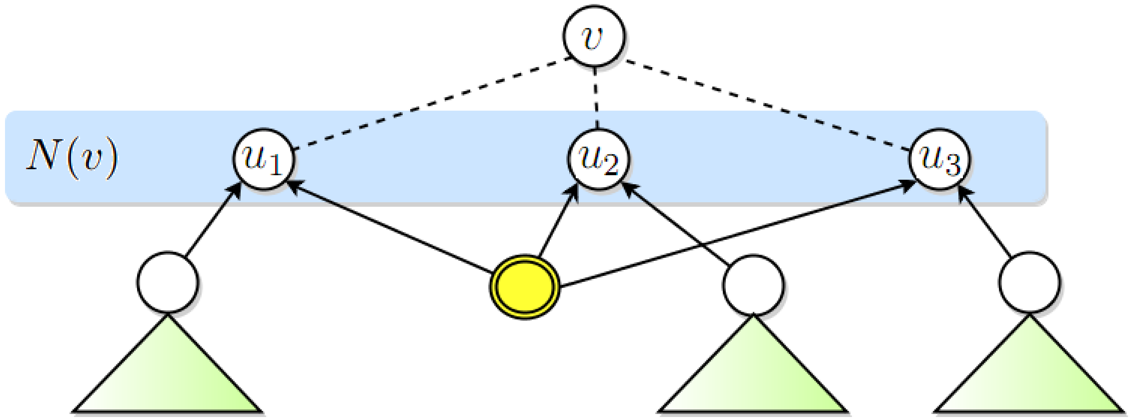

Figure 3: u_1, u_2, u_3 were k-excited after the first phase. In the second phase, marking a single heavy vertex (in double circle) turns all of them into $(k+1)$-roots.

4.3.3 Proof of Theorem 4.2.7

We restate the theorem to ease reading.

Theorem 4.2.7 (restated)**.** *Let $k \geq 1$ be an integer and $\epsilon > 1/(3k+2)$. There is a constant $C = C(k, \epsilon)$ such that for sufficiently large d the following holds: If G is a graph with $n \geq \Omega(d^2)$ vertices and minimum degree d, and G is (k, p, γ)-good with $p = 1/d^{2-\epsilon}$ and $\gamma = \Omega(1/\mathrm{polylog}\,(d))$, then G contains a subgraph of average degree 4 and size at most $C \cdot n/d^{2-\epsilon}$. (The power of the* polylog (d) *term in the bound on γ may depend on k.)*

Proof. Consider a sample from $G(V, E, 2p)$. Let A denote the subset of marked vertices, so $\mathbb{E}[|A|] = 2np$. Let $\mathcal{O}$ be a random orientation of the edges in G, sampled as follows: Each edge is oriented with independent probability $\frac{1}{2}$ towards each of its two ends. We define B to be the random subset that contains each vertex v if,

- v satisfies the conclusion of Lemma 4.3.4. That is, v is part of an arrangement H_v with at most $3k+1$ non-marked vertices, and has three edges oriented towards it in H_v.
- The orientation of H_v coincides with the orientation $\mathcal{O}$.

By Lemma 4.3.4, for sufficiently large d, each vertex v has probability $\geq M \cdot p$ to satisfy the first item, for a constant M that we will set right away. As for the second item, H_v has at most $3k+1$ non-marked vertices that each of whom has two edges oriented towards it (by definition of arrangement), plus one additional edge oriented towards v (by the assertion of Lemma 4.3.4). Hence H_v has at most $6k+3$ edges, and therefore, $\mathcal{O}$ has probability $\geq (\frac{1}{2})^{6k+3}$ to coincide with the orientation of H_v. In conclusion, v has probability $\geq (\frac{1}{2})^{6k+3} Mp$ to be in B, over the choices of both $\mathcal{O}$ and the sample of $G(V, E, 2p)$. We set $M = 16 \cdot 2^{6k+3}$ and obtain that $\mathbb{E}[|B|] \geq 16np$.

By applying the Chernoff bound Lemma 2.3 to $|A|$ and plugging $n \geq \Omega(d^2)$, we get:

$$\Pr\left[|A| < 2\mathbb{E}|A|\right] \geq 1 - (0.25e)^{2np} = 1 - (0.25e)^{2nd^{\epsilon-2}} \geq 1 - (0.25e)^{\Omega(d^\epsilon)}$$

On the other hand, by Lemma 2.1, $|B|$ attains half its expected value with probability at least $\frac{\mathbb{E}|B|}{2n} = 8p = 8d^{-(2-\epsilon)}$. Summing the bounds yields:

$$\Pr\left[|A| \leq 2\mathbb{E}|A|\right] + \Pr\left[|B| \geq \frac{1}{2}\mathbb{E}|B|\right] \geq 1 - (0.25e)^{\Omega(d^{\epsilon})} + 8d^{-(2-\epsilon)}$$

Since $(0.25e)^{\Omega(d^{\epsilon})} \ll 8d^{-(2-\epsilon)}$ for sufficiently large d, the above RHS is strictly larger than 1. Therefore, there is a positive probability that both the events $|A| \leq 2\mathbb{E}|A|$ and $|B| \geq \frac{1}{2}\mathbb{E}|B|$ occur. These imply $|A| \leq 4np$ and $|B| \geq 8np$, respectively. We fix this event from now on (note that this also fixes an orientation $\mathcal{O}$), and arbitrarily remove vertices from B until $|B| = 8np$. We take our target subgraph H to be the one induced by the union of A and the arrangements H_v for all $v \in B$.

Size of H. For each $v \in B$, H_v has at most $3k+1$ non-marked vertices. Hence:

$$|H| \leq |A| + (3k+1)|B| \leq (2 + 4(3k+1))np = (12k+6)n/d^{2-\epsilon}$$

which is as required, if we set $C(k) = 12k + 6$.

Average degree of H. To show that H has average degree at least 4, we need to count two edges per vertex. We use the orientation $\mathcal{O}$ to assign edges to vertices, to ensure that each edge is counted in favour of only one vertex.

Consider a non-marked vertex u in H. It is part of an arrangement H_v for some $v \in B$, and being an arrangement, u has two edges oriented towards it in H_v. Since the orientation of H_v coincides with $\mathcal{O}$, we see that u has two edges oriented towards it in $\mathcal{O}$.

Additionally, each $v \in B$ has a third edge oriented towards it in H_v, and hence in $\mathcal{O}$. Together, we have $|B|$ edges that we have not yet used, and we now count them to cover for the marked vertices. Since the number of marked vertices is $|A| \leq 4np$, and the number of remaining edges is $|B| = 8np$, we can indeed count two edges per marked vertex. In conclusion, we see that H has average degree ≥ 4, and the proof is complete. $\square$

4.4 Tighter Bounds for Lower Densities

We now prove Item 2 of Theorem 4.1.2, that shows that as the density of the input graph lowers, we can get a gradual improvement of the value of ϵ in the size bound of the target subgraph. The next theorem restates the result formally.

Theorem 4.4.1. *Let $\epsilon > 0$. There is a constant $C = C(\epsilon)$ such that every graph with n vertices and average degree d satisfying $4 \leq d \leq O(n^{1/t})$ for $t = \lfloor \log(\frac{8}{3}(\frac{1}{\epsilon} - 2)) \rfloor$, contains a subgraph of average degree 4 and size at most $C \cdot n/d^{2-\epsilon}$.*

The idea underlying the proof is that either some vertex neighbourhood is dense enough to constitute our target subgraph, or all neighbourhoods are sparse enough to be considered as trees, which be handled as in Claim 4.2.5.

Lemma 4.4.2. *Let $t \geq 1$ be a fixed integer, $0 < \alpha < 1$, and $r > 0$ sufficiently large. Let T be an r-ary tree with all leaves in level t.*[10] *Suppose we remove all but at least an α-fraction of the leaves from T. The remaining tree contains an r'-ary tree with $r' = (\frac{1}{2})^{t-1}\alpha r$, with the same root as T and with all leaves in level t.*

Proof. By induction on t. In the base case $t = 1$ there is nothing to show, as the root itself is the only leaf and $\alpha > 0$, so we cannot remove any leaves. Suppose now $t > 1$. Let L be the subset of remaining leaves after the removal, and let M be the subset of vertices in the $(t-1)^{th}$ level of T. In the bipartite graph with sides L and M, side M has maximum degree r and average degree $\frac{|L|}{|M|} = \frac{\alpha \cdot r^{t-1}}{r^{t-2}} = \alpha r$. Hence by Lemma 2.7, there is a subset $M' \subset M$ with size $|M'| \geq \frac{\alpha}{2}|M|$ such that each vertex in M' has $\frac{1}{2}\alpha r$ neighbours in L.

Consider the r-ary tree T' given by the top $t-1$ levels of T. The subset of its leaves is M, and suppose we remove all leaves but those in M'. This removes all but at least a $\frac{1}{2}\alpha$-fraction of the leaves, so by the inductive hypothesis, T' contains an r'-ary tree T" with $r' = (\frac{1}{2})^{t-2} \cdot \frac{1}{2}\alpha \cdot r = (\frac{1}{2})^{t-1}\alpha r$, which has the same root as T and all leaves in M'. We extend T" by one more level, by picking for each leaf in M' an arbitrary subset of r' neighbours in L, which it is guaranteed to have by the choice of M' (as explained above). T" constitutes the required subtree of T. $\square$

Proposition 4.4.3 (main towards proving Theorem 4.4.1)**.** *Let $t \geq 1$ be an integer. There is a constant $\gamma_t > 0$ such that for every graph G with minimum degree d sufficiently large and an arbitrary vertex v in G, at least one of the following holds:*

- *G contains a subgraph of size at most $\gamma_t d^{t-2}$ and average degree ≥ 4.*
- *G contains a $(\gamma_t d)$-ary tree rooted by v, with all leaves in level t.*

Proof. For each vertex u in G we restrict our attention to an arbitrary subset of exactly d of its neighbours, and refer only to them as its neighbours. This approach has been taken in our proofs before; as usual, it may happen that for an adjacent pair or vertices u, u' we consider u' to be a neighbour of u but not vice-versa, and this would not interfere with our reasoning. Moreover, keep in mind that we will assume d is sufficiently large wherever necessary.

We go by induction on t. The base case $t = 1$ is trivial, since v alone is a tree with one level (of any arity). Now fix $t > 1$. By induction, G either contains a subgraph of size d^{t-3} and average degree ≥ 4, or a $(\gamma_{t-1}d)$-ary tree T rooted by v with all leaves in level $t-1$. The former immediately implies the proposition, so we now focus on the latter. Let L denote the set of leaves in T. Note that $|L| = (\gamma_{t-1}d)^{t-2}$ and,

$$|T| = \frac{(\gamma_{t-1}d)^{t-1} - 1}{\gamma_{t-1}d - 1} \leq 2(\gamma_{t-1}d)^{t-2}.$$

Let E_L be the subset of edges incident to vertices in L. Since G has minimum degree d, we have $|E_L| \geq d|L|$. We write E_L as a disjoint union $E_L = E_L^{in} \cup E_L^{out}$, where E_L^{in} are the edges going back into T (i.e. have both endpoints in T), and E_L^{out} are all the other edges. If

[10] Recall that as in Claim 4.2.5, we consider the root to be in level 1.

$|E_L^{in}| \geq 2|T|$ then T is a subgraph with average degree 4 and size $\leq 2(\gamma_{t-1}d)^{t-2}$, which meets the requirement of the proposition if we set $\gamma_t = 2\gamma_{t-1}^{t-2}$. Otherwise,

$$|E_L^{out}| > |E_L| - 2|T| \geq \gamma_{t-1}^{t-2}d^{t-1} - 4(\gamma_{t-1}d)^{t-2} \geq \tfrac{1}{2}\gamma_{t-1}^{t-2} \cdot d^{t-1} \geq \frac{1}{2}d|L| \tag{4.5}$$

for sufficiently large d.

Let N_L be the set of all endpoints of edges in E_L^{out} which are not in T. (Note that each edge in E_L^{out} has exactly one endpoint not in T.) We say a vertex in N_L is *light* if it is a neighbour of at most two vertices in L, and *heavy* otherwise. Now, either half the edges in E_L^{out} are incident to light vertices, or half are incident to heavy vertices. We handle the two cases separately.

- *Case I - Light vertices:* Half the edges in E_L^{out} are incident to light vertices in N_L.

  Let $N_l \subset N_L$ denote the set of light vertices. Since each light vertex is incident to at most 2 edges in E_L^{out}, we have $|N_l| \geq \frac{1}{4}|E_L^{out}|$. Hence in the bipartite graph with sides L and N_l, side L has maximum degree d and average degree (using eq. (4.5)):

  $$\frac{|N_l|}{|L|} \geq \frac{\frac{1}{4}|E_L^{out}|}{|L|} \geq \frac{\frac{1}{8}d|L|}{|L|} = \tfrac{1}{8}d$$

  Now by Lemma 2.7, there is a subset $L' \subset L$ of size at least $\frac{1}{16}|L|$ such that each vertex in L' has $\frac{1}{16}d$ neighbours in N_l. Remove from T all leaves except those in L'. By Lemma 4.4.2 (with $\alpha = \frac{1}{16}$ and $r = \gamma_{t-1}d$), T contains a $((\frac{1}{2})^{t+2}\gamma_{t-1}d)$-ary subtree T' with all leaves in level $t-1$, that is in L'. We extend T by one more level using Lemma 2.5: Recall that in the bipartite graph with sides L' and N_l, side L has minimum degree $\frac{1}{16}d$ and side N_l has maximum degree 2. Hence, Lemma 2.5 picks a subset of edges such that each vertex in N_l is adjacent to at most one vertex in L', and each vertex in L' has at least $\frac{1}{32}d$ neighbours in N_l, thus adding a t^{th} level to T' while keeping it r-ary with $r = \min\{(\frac{1}{2})^{t+2}\gamma_{t-1}d, \frac{1}{32}d\}$. This meets the requirement of the proposition.

- *Case II - Heavy vertices:* Half the edges in E_L^{out} are incident to heavy vertices in N_L.

  Let $N_h \subset N_L$ be the subset of heavy vertices.

  - If $|N_h| < 2|L|$, consider the subgraph H induced by $L \cup N_h$. It has at most $3|L|$ vertices, and it contains all the edges in E_L^{out} which are incident to heavy vertices. By the assumption of the current case there are at least $\frac{1}{2}|E_L^{out}|$ such edges, and by eq. (4.5), this is at least $\frac{1}{2}d|L|$. Hence for $d \geq 12$, H has average degree ≥ 4.
  - If $|N_h| \geq 2|L|$, pick an arbitrary subset $N_h' \subset N_h$ of size exactly $2|L|$, and consider the subgraph H induced by $L \cup N_h'$. It has $3|L|$ vertices and at least $3|N_h'| = 6|L|$ edges, as each vertex in N_h' is incident to at least 3 edges in E_L^{out}. Hence H has average degree ≥ 4.

  In both cases the size of H is bounded by $3|L| = 3(\gamma_{t-1}d)^{t-2}$, and hence satisfies the requirement of the proposition if we set $\gamma_t = 3\gamma_{t-1}^{t-2}$.

Conclusion. Considering all the cases that have arisen, the claim is proven with $\gamma_t = \min\{\frac{1}{16}, (\frac{1}{2})^{t+1}\gamma_{t-1}, 3\gamma_{t-1}^{t-2}\}$. □

Proof of Theorem 4.4.1

As usual (for example in the proofs of Corollaries 4.2.9 and 4.2.10), we suppose that d is the minimum degree and that it is sufficiently large. Set $k = 2^{t-2}$ (and note this adheres to Claim 4.2.5). The motivation for the choice t in the statement is that in order to apply the Reduction Theorem 4.2.7, we need to pick t such that $\epsilon > 1/(3k+2)$. As already remarked in the proof of Corollary 4.2.10, one may verify by rearranging that the value set for t is the smallest integer that meets this requirement.

We apply Proposition 4.4.3 to each vertex in G, and handle two cases:

- Case I: For some vertex v, the first option of Proposition 4.4.3 is met. This means G contains a subgraph of size at most $\gamma_t d^{t-2}$ and average degree 4.

- Case II: For each vertex v, the second option of Proposition 4.4.3 is met. This means v is the root of a $(\gamma_t d)$-ary tree with all leaves in level t, so by Claim 4.2.5, v has probability $\Omega(1) \cdot d^{2k-1} p^k$ to be k-excited with $p = 1/d^{2-\epsilon}$. This holds for each vertex, hence G is $(k, p, \Omega(1))$-tree-like, and this holds even if we remove any single vertex from the graph, hence G is $(k, p, \Omega(1))$-good. By the Reduction Theorem 4.2.7, there is a constant $C = C(k, \epsilon)$ such that G contains a subgraph of size $C \cdot n/d^{2-\epsilon}$ and average degree 4.

Combining the two cases, we obtain that G contains a subgraph of average degree 4 and size at most $\max\{\gamma_t d^{t-2}, Cn/d^{2-\epsilon}\}$. We will finish the proof by showing,

$$\gamma_t d^{t-2} \leq Cn/d^{2-\epsilon} \tag{4.6}$$

for which it is enough to show:

$$\gamma_t d^{t-2} \leq \beta n/d^2 \tag{4.7}$$

for an arbitrary constant $\beta > 0$, as then eq. (4.6) follows for sufficiently large d. And Indeed, by hypothesis we have $d \leq O(n^{1/t})$, hence $\gamma_t d^t \leq O(n)$, so eq. (4.7) is satisfied. □

4.5 Improved Bound for Square-Free Graphs

We now prove Item 4 of Theorem 4.1.2, formally stated next. The analysis allowing this proof is patterned by that of Coja-Oghlan et al. [COFKR13, Lemma 24]. Note that we exclude squares (length-4 cycles) but allow triangles (length-3 cycles). However even if we didn't, i.e. in the girth ≥ 5 case, Item 4 would still improve over Item 3 as the former handles $\epsilon > 1/11$ whereas the latter only handles $\epsilon > 1/8$.

Theorem 4.5.1. *Let $\epsilon > 1/11$. There is a constant $C = C(\epsilon)$ such that every graph with n vertices, average degree d (satisfying $4 \leq d \leq O(\sqrt{n})$) and no cycles of length 4, contains a subgraph of average degree 4 and size at most $C \cdot n/d^{2-\epsilon}$.*

Proof. As usual we suppose that d is the minimum degree and that it is sufficiently large (see the proofs of Corollaries 4.2.9 and 4.2.10 for details). Moreover for each vertex in G we restrict attention to exactly d of its neighbours. We use $o(1)$ for a term that tends to 0 as d grows, and $o(f)$ for $o(1) \cdot f$.

Set $p = d^{-21/11}$. We will show that G is $(3, p, \Omega(1))$-good, and the theorem would then follow from the Reduction Theorem 4.2.7.

Proposition 4.5.2. *Each vertex in G is 2-excited w.p.* $\geq (1-o(1))d\binom{d-2}{2}p^2$.

Proof. Let v be any vertex and let $N(v)$ be its subset of neighbours. A vertex $u \in N(v)$ may have at most one neighbour in $N(v)$, as otherwise it would form a square with its two neighbours in $N(v)$ and with v. Hence u has at least $d-2$ neighbours not in $\{v\} \cup N(v)$.

Let L denote the subset of all vertices that are neighbours of vertices in $N(v)$, and that are not in $\{v\} \cup N(v)$. By the above, each vertex in $N(v)$ has at least $d-2$ neighbours in L. On the other hand, a vertex $w \in L$ has a unique neighbour in $N(v)$, since if it had two neighbours u', u" it would form a square $v - u' - w - u$". Furthermore, w is not a neighbour of v since L was defined to exclude all neighbours of v.

We conclude that v is the root of a tree with all leaves in level 3, such that v (which is level 1) has d children and each vertex in level 2 has $d-2$ children. To make v 2-excited, we need to pick one of its children, and for that child, pick two children and mark them. The total probability is $(1-o(1))d\binom{d-2}{2}p^2$. □

(We remark that $(1-o(1))d\binom{d-2}{2}p^2 = \Omega(d^3p^2)$, and hence the above proposition essentially proves Theorem 4.5.1 for $\epsilon > 1/8$.)

Fix a vertex v for the rest of the proof, with neighbours $v_1, \ldots, v_d$. Let X_i be an indicator random variable for the event that v_i is 2-excited, and let $X = \sum_i X_i$.

Proposition 4.5.3. $\Pr[X \geq \frac{1}{2}\gamma p^2 d^4)] > \frac{1}{2}$ *for some constant* $\gamma > 0$.

Proof. By Proposition 4.5.2 we have $\Pr[X_i] = \Omega(p^2d^3)$, and hence $\mathbb{E}[X] \geq \gamma p^2 d^4$ for some $\gamma > 0$. We now wish to show that X behaves roughly as its expectation using the second moment method, so we turn to bounding $\mathbb{E}[X^2]$. To this end we take $i \neq j$ and upper-bound $\Pr[X_i \wedge X_j]$. Since G is square-free, v_i and v_j have v as a unique mutual neighbour. Hence one possibility to make both v_i and v_j 2-excited is that v is a 2-root, which happens with probability $\frac{1}{2}(1-o(1))d^2p^2$ (as is easily verified).

The other possibility is that v_i and v_j are neighbours of distinct 2-roots, and this is the event we analyse now. Let N_i and N_j denote the subset of neighbours of v_i, v_j respectively, excluding v. Then N_i, N_j are disjoint and of size $d-1$ each, and every $u \in N_i$, $w \in N_j$ have at most one mutual neighbour. Dependency between X_i, X_j is maximized when each such pair u, w has *exactly* one mutual neighbour, so this is the case we consider. We visualize this as a $(d-1) \times (d-1)$ matrix with rows corresponding to N_i, columns corresponding to N_j, and entries corresponding to (some of the) vertices in G, such that $u \in N_i$ is adjacent to the entries in its row, and $w \in N_i$ is adjacent to the entries in its column.

We are concerned with the event that some $u \in N_i$, $w \in N_j$ are each a 2-root, which corresponds to having a row and a column, each with two marked entries. This can occur in either of two constellations, as illustrated in fig. 4:

- Four marked entries. We need to choose a row r and a column c ($(d-1)^2$ options), then two columns c', c" of the remaining columns excluding c (so $\binom{d-2}{2}$ options) and two rows r', r" of the remaining rows exlucing r (so again $\binom{d-2}{2}$ options), and finally, to mark the four entries $(r, c'), (r, c"), (r', c), (r", c)$. See fig. 4 to make this clearer (the row and column marked with arrows are r and c). The total probability is hence: $(1-o(1))((d-1)\binom{d-2}{2}p^2)^2$.

- Three marked entries. The total probability is $(1-o(1))((d-1)(d-2))^2p^3$, for considerations similar to the above; details are omitted.

Adding everything together, we have:

$$\Pr[X_i \wedge X_j] = \frac{1}{2}(1-o(1))p^2d^2 + (1-o(1))((d-1)(d-2))^2p^3 + (1-o(1))((d-1)\binom{d-2}{2}p^2)^2$$

The first term behaves roughly like d^2p^2, the second like d^2p^3, and the third like d^6p^4. Since $pd^2 \gg 1$, the third term is asymptotically dominant and hence:

$$\Pr[X_i \wedge X_j] = (1+o(1))\left((d-1)\binom{d-2}{2}p^2\right)^2$$

By Proposition 4.5.2 we have $\Pr[X_i] \geq (1-o(1))d\binom{d-2}{2}p^2$, and a similar bound holds for $\Pr[X_j]$. Hence,

$$\Pr[X_i \wedge X_j] \leq (1+o(1))\left(d\binom{d-2}{2}p^2\right)^2 \leq \frac{1+o(1)}{1-o(1)}\Pr[X_i]\Pr[X_j] \leq (1+o(1))\Pr[X_i]\Pr[X_j]$$

Consequently:

$$\mathbb{E}[X^2] = \sum_{i,j}\mathbb{E}[X_iX_j] = \sum_{i,j}\Pr[X_i \wedge X_j] = \sum_i \Pr[X_i] + \sum_{i \neq j}\Pr[X_i \wedge X_j] \leq$$

$$\mathbb{E}[X]+(1+o(1))\sum_{i\neq j}\Pr[X_i]\Pr[X_j] = \mathbb{E}[X]+(1+o(1))\sum_{i\neq j}\mathbb{E}[X_i]\mathbb{E}[X_j] \leq \mathbb{E}[X]+(1+o(1))\left(\mathbb{E}[X]\right)^2$$

and therefore, $\mathbf{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \leq \mathbb{E}X + o(1)\left(\mathbb{E}X\right)^2$. Since $\mathbb{E}X \gg 1$ we have $\mathbb{E}X = o((\mathbb{E}X)^2)$ so we can write $\mathbf{Var}(X) = o((\mathbb{E}X)^2)$. Now by Chebyshev's inequality:

$$\Pr[|X - \mathbb{E}X| > 0.5\mathbb{E}X] < \frac{4\mathbf{Var}(X)}{(\mathbb{E}X)^2} \xrightarrow[d\to\infty]{} 0$$

In particular, for sufficiently large d, the probability that X is less than half its expected value is less than $\frac{1}{2}$. Since $\mathbb{E}X \geq \gamma p^2d^4$, the proposition is proven. □

Now we can prove the theorem. In $G(V,E,p)$, we suppose the vertices are marked in two independent phases, each with probability $\frac{1}{2}p$. This approach has already been taken in the proof of Lemma 4.3.4, and as explained there, this indeed simulates $G(V,E,p)$ up to a small variation to the constants involved.

By Proposition 4.5.3, after the first phase there is probability $\frac{1}{2}$ for v to have $\frac{1}{2}p^2d^4$ neighbours that are 2-excited. Let Y denote their subset. For $w \in Y$, let $A(w)$ be the event that the second phase marks a neighbour of w which is neither v nor the neighbour that made 2-excited in the first phase. Since w has $d-2$ such neighbours, $A(w)$ occurs with probability (roughly) $(d-2)p$. Furthermore, the events $\{A(w) : w \in Y\}$ are independent because all vertices in Y have v as a mutual neighbour, and cannot share additional mutual neighbours without forming a square. Therefore, the probability for at least one of the events $A(w)$ to

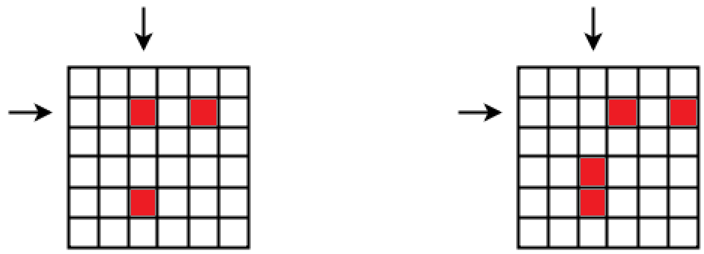

Figure 4: Two possible constellations for having a row and column with two marked entries each: On the left with a total of three marked entries, and on the right with a total of four marked entries.

occur is (again, roughly) $|Y| \cdot (d-2)p = \Omega(p^3 d^5)$. This occurrence renders w a 3-root, and hence makes v 3-excited.

We have proven this for an arbitrary v, so G is $(3, p, \Omega(1))$-tree-like. All properties used remain intact even if we remove any single vertex from G (up to a small variation of constants), hence G is $(3, p, \Omega(1))$-good. Theorem 4.5.1 now follows from Theorem 4.2.7. $\square$

5 Negative Results

In this section we provide lower bounds on $S_\ell(n,d)$, by establishing the existence of arbitrarily large graphs that exclude all ℓ-subgraphs up to a certain size.

We note that a graph may have an ℓ-subgraph of a certain size but not of any larger size, so in a sense the property is non-monotone. (To illustrate this, consider a $(2\ell+1)$-clique joined with arbitrarily many isolated vertices.) Therefore in order to prove $S_\ell(n,d) \geq s$, we need explicitly to rule out ℓ-subgraph of all sizes up to s, and not just s.

5.1 A Random Graph Model

Definition 5.1. *The distribution $\mathcal{G}_{min}(n,d)$ over simple graphs on the vertex set $[n] = \{1,\dots,n\}$ is defined by the following sampling process:*

- *In the first stage, each vertex chooses uniformly at random a subset of size d of the remaining $n-1$ vertices, and connects to them with an (undirected) edge. Parallel edges are allowed.*
- *In the second stage, parallel edges are unified into a single edge. The resulting (simple) graph is the output sample.*

Proposition 5.2. *Let G be sampled from $\mathcal{G}_{min}(n,d)$. Then,*

1. *G has minimum degree at least d.*
2. *G has average degree between d and $2d$.*
3. *Any subset F of possible edges on the vertex set $[n]$ occurs in G w.p. $\leq (\frac{4d}{n})^{|F|}$.*

Proof. (*1*) In the first stage of the sampling process, each vertex chooses d neighbours, and remains connected to all of them after the second stage.

(*2*) The first stage places exactly dn edges in G. Then, as each pair of vertices is connected with at most two parallel edges, the second stage removes at most half the edges.

(*3*) Each edge occurs in G with probability:

$$p = 1 - \left(\frac{\binom{n-2}{d}}{\binom{n-1}{d}}\right)^2 = 1 - \left(\frac{n-1-d}{n-1}\right)^2 = \frac{2d}{n-1} - \frac{d^2}{(n-1)^2} \leq \frac{4d}{n}$$

and concurrent appearance of edges is either independent (if they are vertex-disjoint) or negatively correlated (otherwise). □

Our reason for preferring $\mathcal{G}_{min}(n,d)$ over the more standard $G(n,p)$ model is the firm bound on the minimum degree, that automatically strengthens our results. We could carry out the upcoming proofs of Theorems 5.3 and 5.4 with $G(n, p = \frac{d}{n})$, and get similar but slightly weaker results, with d being the average degree rather than the minimum degree.[11]

[11] In some of the cases, we could then transfer to minimum degree with additional effort.

5.2 Negative Result for All Densities

The next theorem states that $S_\ell(n,d) = \Omega(n/d^{\ell/(\ell-1)})$, for all densities d.

Theorem 5.3. *Let $\ell > 1$. There is a constant $c_\ell > 0$ such that for all sufficiently large n and $d = O(n^{(\ell-1)/\ell})$, there is a graph on n vertices with minimum degree d, without any ℓ-subgraphs of size $\leq c_\ell \cdot n/d^{\ell/(\ell-1)}$.*

Proof. Let G be a sample of $\mathcal{G}_{min}(n,d)$. For each subset U of $[n]$, let A_U denote the event that G contains an ℓ-subgraph on the vertex set U. We recall this means that there is a subset F of edges with size $\lceil \ell|U| \rceil$ and with all endpoints in U. We assume for simplicity that $\ell|U|$ is an integer (even though this is not necessary), so $|F| = \ell|U|$. By Proposition 5.2, each such F occurs in G w.p. $\leq \left(\frac{4d}{n}\right)^{\ell|U|}$, so a union bound over the possible choices of F out of the edges that may be induced by U, implies:

$$\Pr[A_U] \leq \binom{|U|^2}{\ell|U|} \cdot \left(\frac{4d}{n}\right)^{\ell|U|} \leq \left(\frac{4ed|U|}{\ell n}\right)^{\ell|U|} \tag{5.1}$$

where we have applied the known bound $\binom{m}{k} \leq \left(\frac{em}{k}\right)^k$. For $s > 0$, we apply another union bound to bound the probability p_s that G contains an ℓ-subgraph of size at most s:

$$\begin{aligned} p_s &\leq \sum_{U\subset[n],|U|\leq s} \Pr[A_U] = \sum_{t=1}^{s} \sum_{U\subset[n],|U|=t} \Pr[A_U] \leq \sum_{t=1}^{s} \binom{n}{t} \left(\frac{4edt}{\ell n}\right)^{\ell t} \leq \sum_{t=1}^{s} \left(\frac{en}{t}\right)^t \left(\frac{4edt}{\ell n}\right)^{\ell t} \\ &= \sum_{t=1}^{s} \left(e\left(\frac{4e}{\ell}\right)^\ell \cdot \frac{d^\ell}{n^{\ell-1}} \cdot t^{\ell-1}\right)^t \leq \sum_{t=1}^{s} \left(e\left(\frac{4e}{\ell}\right)^\ell \cdot \frac{d^\ell}{n^{\ell-1}} \cdot s^{\ell-1}\right)^t \end{aligned} \tag{5.2}$$

Plugging $s = c_\ell \cdot n/d^{\ell/(\ell-1)}$ with $c_\ell = \left(\frac{1}{3e}\left(\frac{\ell}{4e}\right)^\ell\right)^{1/(\ell-1)}$, we find that $p_s \leq \sum_{t=1}^{s}\left(\frac{1}{3}\right)^t < \sum_{t=1}^{\infty}\left(\frac{1}{3}\right)^t = \frac{1}{2} < 1$. Hence there is a sample G without ℓ-subgraphs of size at most s. ☐

5.3 Strong Negative Result for High Density

The next theorem states that $S_\ell(n,d) = \omega(n/d^{\ell/(\ell-1)})$ for the highest density case, $d = \Theta(n^{(\ell-1)/\ell})$. This amounts to finding arbitrarily large graphs of such density without any constant-sized ℓ-subgraphs.

Theorem 5.4. *Let $\ell > 1$ and $c, s > 0$ be arbitrary constants. For all sufficiently large n, there is a graph on n vertices with minimum degree $d = c \cdot n^{(\ell-1)/\ell}$, without any ℓ-subgraphs of size $\leq s$.*

Proof. Our proof parallels that of Hoory [Hoo02, Theorem A.4], which addresses the closely related problem of showing there are graphs with large girth. Let G be a sample of $\mathcal{G}_{min}(n,d)$. We use the notation of events $\{A_U : U \subset [n]\}$ as in the above proof of Theorem 5.3. For U with size $|U| = t \leq s$, we have:

$$\Pr[A_U] \leq \left(\frac{4edt}{\ell n}\right)^{\ell t} \leq \frac{\alpha}{n^t} \tag{5.3}$$

for a sufficiently large constant α. (The first inequality was established as eq. (5.1) in the proof of Theorem 5.3.) We need to show that with positive probability, none of the events $\{A_U : |U| \leq s\}$ occurs. To this end we invoke the Local Lemma, stated next. For a proof see [AS11, Lemma 5.1.1].

Lemma 5.5 (Local Lemma)**.** *Let $\mathcal{A}$ be a finite set of events in an arbitrary probability space. For $A \in \mathcal{A}$, let $\Gamma(A) \subset \mathcal{A}$ be such that A is independent of the collection of events $\mathcal{A} \setminus (A \cup \Gamma(A))$. If there is an assignment of reals $x : \mathcal{A} \to (0, 1)$ such that for all $A \in \mathcal{A}$,*

$$\Pr[A] \leq x(A) \cdot \prod_{B \in \Gamma(A)} (1 - x(B)) \tag{5.4}$$

then with positive probability, none of the events in $\mathcal{A}$ occurs.

Observe that in $\mathcal{G}_{min}(n, d)$, the event A_U is determined solely by the choices of the vertices in U, and hence is independent of $A_{U'}$ for all $U' \subset [n]$ such that $U \cap U' = \emptyset$. In other words, A_U may be dependent only of events $A_{U'}$ for which U, U' have mutual vertices. Therefore, applying Lemma 5.5 with $x(A_U) = 2 \Pr[A_U]$, the condition eq. (5.4) becomes:

$$\Pr[A_U] \leq 2 \Pr[A_U] \cdot \prod_{r=1}^{s} \prod_{\substack{|U'|=r \\ U \cap U' \neq \emptyset}} (1 - 2 \Pr[A_{U'}]) \tag{5.5}$$

The number of subsets U' of size r that share any vertices with U is:

$$\binom{n}{r} - \binom{n - |U|}{r} \leq \frac{n^r}{r!} - \frac{(n - r - |U| + 1)^r}{r!} \leq \beta n^{r-1}$$

for a sufficiently large constant β. Using this and eq. (5.3) we get,

$$\prod_{r=1}^{s} \prod_{\substack{|U'|=r \\ U \cap U' \neq \emptyset}} (1 - 2 \Pr[A_{U'}]) \geq \prod_{r=1}^{s} \left(1 - \frac{2\alpha}{n^r}\right)^{\beta n^{r-1}} \geq \exp\left(-\sum_{r=1}^{s} \frac{4\alpha\beta}{n}\right) \geq \frac{1}{2}$$

for sufficiently large n. (The middle inequality is because $1 - z \geq \exp(-2z)$ holds for all, say, $0 < z < \frac{1}{2}$.) Consequently eq. (5.5) is satisfied, so by Lemma 5.5 we get positive probability that none of the events occurs, and the theorem follows. □

5.4 Negative Result for Regular Graphs

Next we give a result specialized to regular graphs. It essentially shows that for high densities, the bound $S_\ell(n, d) = \Omega(n/d^{\ell/(\ell-1)})$ from Theorem 5.3 holds even under this restriction. More precisely, let $S_\ell^{\text{reg}}(n, d)$ be the optimal upper bound S such that every d-regular graph on n vertices contains an ℓ-subgraph of size at most S. Clearly, $S_\ell(n, d) \geq S_\ell^{\text{reg}}(n, d)$. Theorem 5.6 states that $S_\ell^{\text{reg}}(n, d) = \omega(f(n, d))$ for all functions $f(n, d)$ that can be made constant w.r.t. n by plugging some $d = o(n^{(\ell-1)/\ell})$.

Theorem 5.6. *Let $\ell > 1$ and $s > 0$ be arbitrary constants. For all sufficiently large n and $d = d(n)$ such that $d = o(n^{(\ell-1)/\ell})$ and $d = \omega(1)$, there is a d-regular graph without any ℓ-subgraph of size $\leq s$.*

The proof relies on the interesting main lemma of Kim, Sudakov and Vu [KSV07], that essentially states that in a uniformly random d-regular graph, any fixed constant-sized subgraph occurs with roughly the same probability as in $G(n,p)$ with $p = d/n$ (and also as in $\mathcal{G}_{min}(n,d)$, by Proposition 5.2). In particular, the probability is determined (up to low order terms) solely by the number of edges in the subgraph.

Lemma 5.7 (Lemma 2.1 in [KSV07][12])**.** *For integer n, suppose $d = d(n)$ satisfies $d = \omega(1)$ and $d = o(n)$. Let G be a uniformly random d-regular graph on the vertex set $[n]$. Let F be a fixed collection of edges on this vertex set, of constant size t. Then,*

$$\Pr[F \subset G] = (1 + o(1))(d/n)^t$$

Proof of Theorem 5.6. Let G be a uniformly random d-regular graph on the vertex set $[n]$. We repeat the proof of Theorem 5.3, taking two union bounds to derive eq. (5.2), but with the use of Proposition 5.2 replaced by Lemma 5.7. We obtain $p_s \leq \sum_{t=1}^{s} q^t$, where p_s is the probability that G contains an ℓ-subgraph of size at most s, and $q = C \cdot d^\ell / n^{\ell-1}$ for a constant C that depends on ℓ and s. Since $d = o(n^{(\ell-1)/\ell})$, we have $q \leq \frac{1}{3}$ for sufficiently large n, in which case $p_s < \sum_{t=1}^{\infty} (\frac{1}{3})^t < 1$. This implies that there is a sample G without any ℓ-subgraphs of size up to s. ∎

[12]The statement in [KSV07] is slightly more general, to also apply to nearly-regular random graphs.

6 Bounds on ℓ-Girth

Our bounds on $g_\ell(n,d)$ follow as immediate consequences of the results of Sections 4 and 5.

Upper Bounds.

Corollary 6.1 (upper bound on $g_\ell(n,d)$)**.** *Let $\ell > 1$ be an integer and $\epsilon > 0$. There is a constant $C = C(\ell,\epsilon)$ such that every graph G on n vertices with average degree*[13] *d has ℓ-girth at most $C \cdot n/d^{1+\frac{1}{\ell-2}-\frac{1}{\ell^3-3\ell^2+\ell+2}-\epsilon}$.*

Proof. Applying Theorem 4.1.3 with $\ell-1$, we get subgraph with average degree $2\ell-2$ and size as in the corollary. By Lemma 2.4 it contains a subgraph with minimum degree ℓ. □

Corollary 6.2 (improved upper bounds on 3-girth)**.** *For every $\epsilon > 0$ there is a constant $C = C(\epsilon)$ such that the following holds: Let G be a graph on n vertices with average degree d. If one of the following is satisfied:*

- *$\epsilon > \frac{1}{5}$, or*
- *$d = O(n^{1/t})$ where $t = \lfloor \log(\frac{8}{3}(\frac{1}{\epsilon}-2)) \rfloor$, or*
- *G has girth $\geq 2t-1$ where $t = \lfloor \log(\frac{8}{3}(\frac{1}{\epsilon}-2)) \rfloor$, or*
- *$\epsilon > \frac{1}{11}$ and G is square-free,*

then G has 3-girth at most $C \cdot n/d^{2-\epsilon}$.

Proof. Similar to Corollary 6.1, but with Theorem 4.1.2 instead of Theorem 4.1.3. □

Lower Bounds.

Corollary 6.3. *Let $\ell > 1$. There is a constant $c_\ell > 0$ such that for all sufficiently large n and $d = O(n^{(\ell-2)/(\ell-1)})$, there is a graph on n vertices with minimum degree d, with ℓ-girth $\geq c_\ell \cdot n/d^{1+\frac{2}{\ell-2}}$.*

Corollary 6.4. *Let $\ell > 1$ and $c, s > 0$ be arbitrary constants. For all sufficiently large n, there is a graph on n vertices with minimum degree $d = c \cdot n^{(\ell-2)/(\ell-1)}$, with ℓ-girth $\geq s$.*

Corollary 6.5. *Let $\ell > 1$ and $s > 0$ be arbitrary constants. For all sufficiently large n and $d = d(n)$ such that $d = o(n^{(\ell-2)/(\ell-1)})$ and $d = \omega(1)$, there is a d-regular graph with ℓ-girth $\geq s$.*

Proof of Corollaries 6.3, 6.4 and 6.5. Apply Theorems 5.3, 5.4 and 5.6, respectively, with $\ell/2$. This gives a graph that has no subgraphs with average degree $\geq \ell$ and size up to the bound in the corollaries. In particular, it has no such subgraphs with minimum degree ℓ. □

Similarly to Section 5, we remark that for high densities, Corollary 6.4 states that $g_\ell(n,d) = \omega(n/d^{1+\frac{2}{\ell-2}})$ and Corollary 6.5 shows that the bound $g_\ell(n,d) = \Omega(n/d^{1+\frac{2}{\ell-2}})$ holds even for regular graphs.

[13] To ease the statement, trivially required upper and lower bounds on d are omitted.

7 Hypergraphs and Multigraphs

In this section we give upper bounds on the size of ℓ-subgraphs in hypergraphs, using the results of Sections 3 and 4. The starting point of this work was in fact the following conjecture of Feige [Fei08].

Conjecture 7.1 (Conjecture 1.7 from [Fei08])**.** *Let c be sufficiently large. Every 3-uniform hypergraph on n vertices and $m = c \cdot dn$ hyperedges (with $1 < d \leq O(\sqrt{n})$) has a set of $n' \leq \tilde{O}(n/d^2)$ vertices that induce at least $2n'/3$ hyperedges. (The $\tilde{O}$ notation may suppress a* polylog (n) *multiplicative factor.)*

Put otherwise, Conjecture 7.1 states that a 3-hypergraph must contain a $\frac{2}{3}$-subgraph of size $\tilde{O}(n/d^2)$. Note that it is a close analogue of Conjecture 4.1.1, up to a tighter bound on the size of the subgraph: the multiplicative deviation from n/d^2 is allowed to be polylog (n) in the former and n^ϵ in the latter.

We will show how our positive results extend to hypergraphs. This brings us short of proving Conjecture 7.1, but we obtain the following variants, which are analogues of Theorems 3.1 and 4.1.2 respectively.

Theorem 7.2 (degree compromise)**.** *Let $\epsilon > 0$. There is a constant $C = C(\epsilon)$ such that every 3-uniform hypergraph with n vertices and dn edges (where $\frac{2}{3} \leq d \leq O(\sqrt{n})$) contains a sub-hypergraph of size at most $C \cdot n/d^2$ with at least $\frac{2}{3} - \epsilon$ edges per vertex.*

Theorem 7.3 (size compromise)**.** *Let $\epsilon > 0$. There is a constant $C = C(\epsilon)$ such that the following holds: Let H be a 3-uniform hypergraph with n vertices and dn edges (where $\frac{2}{3} \leq d \leq O(\sqrt{n})$). If one of the following holds:*

1. *$\epsilon > \frac{1}{5}$, or*
2. *$d = O(n^{1/t})$ for $t = \lfloor \log(\frac{8}{3}(\frac{1}{\epsilon} - 2)) \rfloor$, or*
3. *$\epsilon > \frac{1}{11}$ and H contains no loose cycles[14] of length 2, 3 and 4,*

then H contains a sub-hypergraph of size at most $C \cdot n/d^{2-\epsilon}$ with at least $\frac{2}{3}$ edges per vertex.

Our approach is simply to replace each hyperedge with a clique of pair-wise edges, and apply the results from Sections 3 and 4 on the resulting multigraph. We indeed have to consider the latter as a multigraph (rather than a simple graph) in order to avoid a fatal loss in the number of edges. Hence an intermediate step towards handling hypergraphs is to handle multigraphs.

Lower bound. The bound in Theorem 7.3 cannot be improved further than $O(n/d^2)$. That is, for some constant $c > 0$, there are 3-hypergraphs of size n with dn hyperedges excluding all $\frac{2}{3}$-subgraphs of size up to cn/d^2. The proof is very similar to that of Theorem 5.3, by a union bound in a random hypergraph model in which each hyperedge is present with independent probability $p = d/n^2$. Details are omitted.

[14] A *loose cycle* of length k in a hypergraph $H(V, E)$ is a sequence of vertices $v_0, v_1 \ldots, v_{2k}$ with $v_0 = v_{2k}$, such that each hyperedge $\{v_{2i}, v_{2i+1}, v_{2i+2}\}$ is present in E. (This definition is standard.)

7.1 From Graphs to Multigraphs

Recall that a *multigraph*, as opposed to a *simple* graph, may have self-loops and parallel edges. Our upper bounds in Sections 3 and 4 were proven for simple graphs, but the following simple observation extends them to multigraphs as well.

Proposition 7.4. *Each of Theorems 3.1, 4.1.2 and 4.1.3 holds for multigraphs, with only a change to the constant C in the statement.*

Remark. Even Item 3 of Theorem 4.1.2 extends if we define the girth of a multigraph to be the length of the shortest cycle *that does not contain any parallel edges*. This clearly coincides with the usual definition for simple graphs.

Proof. We prove for $\ell = 2$ for simplicity; the generalization is straight-forward. Let G be a multigraph with n vertices and average degree d, so $\frac{1}{2}dn$ edges. If any two vertices in G are connected with 4 parallel edges, they induce a 2-subgraph of size 2. If any vertex in G has 2 self-loops, it induces a 2-subgraph of size 1. In both cases the proposition holds.

Otherwise, we move to the simple subgraph G' of G obtained by eliminating all self-loops and unifying parallel edges into a single edge. We claim that their average degrees are roughly the same. To see this, first remove the self-loops, of which there are at most n (one per vertex). We are left with $\geq (\frac{1}{2}d - 1)n$ edges. Then unify the parallel edges: Since each adjacent pair in G is connected with at most 3 edges, this removes at most two-thirds of them, and we are left with at least $\frac{1}{3}(\frac{1}{2}d - 1)n$ edges. This means G' has average degree $\geq \frac{1}{3}(d-2)$, which is similar to d up to a multiplicative constant. We can now apply either of Theorems 3.1, 4.1.2 and 4.1.3 on G' to get the proposition, suppressing the loss in the average degree into the constant C in their statements. $\square$

7.2 From Multigraphs to Hypergraphs

We now present the reduction from 3-uniform hypergraphs to graphs.

Definition 7.5 (skeleton multigraph of a hypergraph)**.** *Let $H(V, E)$ be a hypergraph. The* skeleton multigraph *G_H of H has vertex set V, and for each $u, v \in V$, the number of parallel edges connecting u, v in G_H is the number of hyperedges in E covering both u and v.*

(Equivalently, each hyperedge in H induces an edge in G_H between each pair it covers.)

Lemma 7.6 (main for Section 7)**.** *Let $H(V, E)$ be a 3-uniform hypergraph with skeleton multigraph G_H. Suppose G_H has a $(2-\epsilon)$-subgraph G' of size k, for some $0 \leq \epsilon < 2$. Then H has a sub-hypergraph of size at most $3k$ with at least $\frac{2}{3} - \frac{1}{3}\epsilon$ edges per vertex.*

Proof. We may assume w.l.o.g. that G' is an induced subgraph. Let E' the subset of hyperedges in H that induce any edges in G'. We can partition E' to disjoint subsets A, B as follows:

- A - hyperedges that induce exactly *one* edge in G'. Let $a = |A|$.
- B - hyperedges that induce exactly *three* edge in G'. Let $b = |B|$.

(Note that since we have assumed G' is an induced subgraph, there are no hyperedges that induce two edges.) With this notation G' has exactly $a + 3b$ edges, and hence:

$$a + 3b \geq (2 - \epsilon)k \tag{7.1}$$

We now arbitrarily remove edges from A (and thus from E') either until A is empty, or until $a + 3b \leq 2k$ (whichever happens first). Note that in both cases, eq. (7.1) remains satisfied.

Let V' be the subset of vertices in H that are covered by any edges in E', and consider the subgraph $H'(V', E')$ of H. Let $U \subset V'$ the subset of vertices in H' that are not in G'. Observe that each hyperedge in A contributes one vertex to U, and a hyperedge in B contributes no vertices to U. Hence $|U| \leq a$ (note that a vertex in U may contributed by more than one hyperedge in A), and we conclude:

$$|V'| = k + |U| \leq k + a \tag{7.2}$$

We bound the size of H'. Recall that by the above, one of two cases must hold: Either $A = \emptyset$, in which case $U = \emptyset$ and hence $|V'| = k$, or $a + 3b \leq 2k$, in which case $a \leq 2k$ and hence $|V'| = k + |U| \leq k + a \leq 3k$. In both cases we have $|V'| \leq 3k$.

We turn to bounding the hyperedge-to-vertex ratio of H'. Using eq. (7.2) and then eq. (7.1), we get:

$$|V'| \leq k + a \leq \frac{a + 3b}{2 - \epsilon} + a = \frac{3a + 3b - \epsilon a}{2 - \epsilon} = \frac{3|E'| - \epsilon a}{2 - \epsilon}$$

(recalling that $|E'| = a + b$), and hence:

$$\frac{|E'|}{|V'|} \geq \frac{(2 - \epsilon)|E'|}{3|E'| - \epsilon a} = \frac{2 - \epsilon}{3 - \frac{\epsilon a}{|E'|}} \geq \frac{2 - \epsilon}{3}$$

as required. □

7.3 Proof of Theorems 7.2 and 7.3

Proof of Theorem 7.2. Let $\epsilon > 0$. Let H be a 3-uniform hypergraph with n vertices and dn hyperedges, and let G_H be its skeleton multigraph. By Definition 7.5 we see that G_H has $3dn$ edges, hence average degree $6d$. By Theorem 3.1 (combined with Proposition 7.4) applied with 3ϵ, G_H contains a $(2 - 3\epsilon)$-subgraph of size $k \leq O(n/d^2)$. Now by Lemma 7.6 applied with 3ϵ, H contains a $(\frac{2}{3} - \epsilon)$-subgraph of size at most $3k$. □

Proof of Theorem 7.3. Similar to the previous proof, only we use Theorem 4.1.2 instead of Theorem 3.1, and apply Lemma 7.6 with $\epsilon = 0$. The only point that requires attention is Item 3 of Theorem 7.3, that relies on the absence of small loose cycles from the hypergraph. It can be easily seen that if a H excludes any loose cycles of lengths 2, 3 and 4, then G_H is square-free (even though it clearly contains many triangles). Hence, Item 3 of Theorem 7.3 follows from Item 4 of Theorem 4.1.2. □

References

[AF09] Noga Alon and Uriel Feige, *On the power of two, three and four probes*, proceedings of the 20th annual ACM-SIAM symposium on discrete algorithms (SODA), 2009, pp. 346–354.

[AHI02] Yuichi Asahiro, Refael Hassin, and Kazuo Iwama, *Complexity of finding dense subgraphs*, Discrete Applied Mathematics **121** (2002), no. 1-3, 15–26.

[AHL02] Noga Alon, Shlomo Hoory, and Nathan Linial, *The moore bound for irregular graphs*, Graphs and Combinatorics **18** (2002), no. 1, 53–57.

[APP+12] Omid Amini, David Peleg, Stéphane Pérennes, Ignasi Sau, and Saket Saurabh, *On the approximability of some degree-constrained subgraph problems*, Discrete Applied Mathematics **160** (2012), no. 12, 1661–1679.

[AS06] Noga Alon and Asaf Shapira, *On an extremal hypergraph problem of Brown, Erdős and Sós*, Combinatorica **26** (2006), no. 6, 627–645.

[AS11] Noga Alon and Joel H. Spencer, *The probabilistic method*, Wiley Series in Discrete Mathematics and Optimization, Wiley, 2011.

[ASS12] Omid Amini, Ignasi Sau, and Saket Saurabh, *Parameterized complexity of finding small degree-constrained subgraphs*, J. Discrete Algorithms **10** (2012), 70–83.

[BB89] Béla Bollobás and Graham Brightwell, *Long cycles in graphs with no subgraphs of minimal degree 3*, Discrete Mathematics **75** (1989), no. 1-3, 47–53.

[BB90] Norman L. Biggs and Alan G. Boshier, *Note on the girth of ramanujan graphs*, J. Comb. Theory, Ser. B **49** (1990), no. 2, 190–194.

[BES73a] William G. Brown, Paul Erdős, and Vera T. Sós, *On the existence of triangulated spheres in 3-graphs, and related problems*, Periodica Mathematica Hungarica **3** (1973), no. 3-4.

[BES73b] William G. Brown, Paul Erdős, and Vera T. Sós, *Some extremal problems on r-graphs*, New Directions in the Theory of Graphs, Proceedings of the 3rd Ann Arbor Conference on Graph Theory (1973).

[BMRR10] Stephan Brandt, Janina Müttel, Dieter Rautenbach, and Friedrich Regen, *Minimum degree and density of binary sequences*, Eur. J. Comb. **31** (2010), no. 7, 1936–1945.

[Bol04] Béla Bollobás, *Extremal graph theory*, Dover Books on Mathematics Series, Dover Publications, 2004.

[BP89] Jean-Claude Bermond and Claudine Peyrat, *Induced subgraphs of the power of a cycle*, SIAM J. Discrete Math. **2** (1989), no. 4, 452–455.

[COFKR13] Amin Coja-Oghlan, Uriel Feige, Michael Krivelevich, and Daniel Reichman, *Contagious sets in expanders*, in arXiv.org, CoRR abs/1306.2465 (2013).

[Dah13] Xavier Dahan, *Regular graphs of large girth and arbitrary degree*, in arXiv.org, CoRR abs/1110.5259v4. To appear in Combinatorica (2013).

[DHL$^+$12] Domingos Jr. Dellamonica, Penny E. Haxell, Tomasz Luczak, Dhruv Mubayi, Brendan Nagle, Yury Person, Vojtech Rödl, Mathias Schacht, and Jacques Verstraëte, *On even-degree subgraphs of linear hypergraphs*, Combinatorics, Probability & Computing **21** (2012), no. 1-2, 113–127.

[EFGS88] Paul Erdős, Ralph J. Faudree, András Gyárfás, and Richard H. Schelp, *Cycles in graphs without proper subgraphs of minimum degree 3*, Ars Combinatorica **25(B)** (1988), 195201.

[EFR86] Paul Erdős, Peter Frankl, and Vojtěch Rödl, *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs and Combinatorics **2** (1986), no. 1, 113–121 (English).

[EFRS90] Paul Erdős, Ralph J. Faudree, Cecil C. Rousseau, and Richard H. Schelp, *Subgraphs of minimal degree k*, Discrete Mathematics **85** (1990), no. 1, 53–58.

[Fei02] Uriel Feige, *Relations between average case complexity and approximation complexity*, proceedings of the 34th annual ACM symposium on theory of computing (STOC), 2002, pp. 534–543.

[Fei07] Uriel Feige, *Refuting smoothed 3CNF formulas*, proceedings of the 48th annual symposium on foundations of computer science (FOCS), 2007, pp. 407–417.

[Fei08] Uriel Feige, *Small linear dependencies for binary vectors of low weight*, Building Bridges Between Mathematics and Computer Science, Bolyai Society Mathematical Studies **19** (2008), 283–307.

[FKO06] Uriel Feige, Jeong Han Kim, and Eran Ofek, *Witnesses for non-satisfiability of dense random 3CNF formulas*, proceedings of the 47th annual symposium on foundations of computer science (FOCS), 2006, pp. 497–508.

[FO07] Uriel Feige and Eran Ofek, *Easily refutable subformulas of large random 3CNF formulas*, Theory of Computing **3** (2007), no. 1, 25–43.

[FS97] Uriel Feige and Michael Seltser, *On the densest k-subgraph problem*, Technical Report CS97-16, Faculty of Mathematics and Computer Science, Weizmann Institute of Science (1997).

[Hoo02] Shlomo Hoory, *On graphs of high girth*, Hebrew University of Jerusalem, 2002.

[Imr84] Wilfried Imrich, *Explicit construction of regular graphs without small cycles*, Combinatorica **4** (1984), no. 1, 53–59 (English).

[Kez91] André E. Kezdy, *Studies in connectivity*, University of Illinois at Urbana-Champaign, 1991.

[KM87] André E. Kézdy and Matthew Markert, *Fragile graphs*, manuscript (1987).

[KSV07] Jeong Han Kim, Benny Sudakov, and Van Vu, *Small subgraphs of random regular graphs*, Discrete Mathematics **307** (2007), no. 15, 1961 – 1967.

[LPS88] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak, *Ramanujan graphs*, Combinatorica **8** (1988), no. 3, 261–277.

[LU95] Felix Lazebnik and Vasiliy A Ustimenko, *Explicit construction of graphs with an arbitrary large girth and of large size*, Discrete Applied Mathematics **60** (1995), no. 1, 275–284.

[Mar82] Gregori A. Margulis, *Explicit constructions of graphs without short cycles and low density codes*, Combinatorica **2** (1982), no. 1, 71–78.

[Mar88] Gregori A. Margulis, *Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators*, Probl. Peredachi Inf. **24** (1988), no. 1, 39–46.

[Mor94] Moshe Morgenstern, *Existence and explicit constructions of q+1 regular Ramanujan graphs for every prime power q*, Journal of Combinatorial Theory, Series B **62** (1994), no. 1, 44–62.

[NV08] Assaf Naor and Jacques Verstraëte, *Parity check matrices and product representations of squares*, Combinatorica **28** (2008), no. 2, 163–185.

[RS76] Imre Z. Ruzsa and Endre Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics, Proc. Colloq. Math. Soc. János Bolyai **2** (1976), no. 18, 939–945.

[WS11] David P. Williamson and David B. Shmoys, *The design of approximation algorithms*, 1st ed., Cambridge University Press, New York, NY, USA, 2011.

A Appendix: Computational Hardness

From a Complexity Theory standpoint, Question 1.1 naturally raises the following computational problem.

Definition A.1. *Let $\Delta > 0$. The computational problem* Minimum Subgraph of Average Degree $\geq \Delta$, *denoted* MSAD_Δ, *is defined as:*

- *Input: A simple graph G.*
- *Output: A smallest subgraph of G with average degree at least Δ, or a report that no such subgraph exists.*

We are not aware of a previous treatment of this problem, but some close variants have been studied, as detailed below. It seems natural to postulate that MSAD_Δ is **NP**-hard for $\Delta > 2$, but despite some effort we were not able to prove this. (The $\Delta \leq 2$ case is polynomial-time solvable, see below.) It is left here as an open problem.

Problem A.2. *Prove or disprove that* MSAD_Δ *is* **NP**-*hard for every* $\Delta > 2$.

In this appendix we review some related work to this problem, then sketch an (obvious) polynomial-time algorithm for $\Delta \leq 2$, and finally we show that when considering the hardness of MSAD_Δ, it is sufficient to look at an arbitrarily small range of Δ values above 2.

Related work. First observe that if we let Δ be part of the input to MSAD_Δ, then the problem is **NP**-hard since we can use it to solve k-Clique by setting $\Delta = k - 1$. Problem A.2 is different in that Δ is a fixed constant, say 4.

As mentioned in Section 1.1, Amini et al. [ASS12, APP+12] obtained some hardness results for MSMD_Δ, which is a problem similar to MSAD_Δ except that the output subgraph is required to have *minimal* degree $\geq \Delta$, rather than average degree.

Some close variants of MSAD_Δ have been shown to be **NP**-hard, particularly by Asahiro, Hassin and Iwama [AHI02] and Feige and Seltser [FS97]: Let $(k, f(k))$-DSP be the problem of deciding whether an input graph has a subgraph of size k with at least $f(k)$ edges. The case $f(k) = \frac{1}{2}\Delta k$ is the decision version of MSAD_Δ. The aforementioned works show that for $0 < \epsilon < 1$, $(k, f(k))$-DSP is **NP**-complete for $f(k) = \Theta(k^{1+\epsilon})$ [AHI02, Theorem 1], and for $f(k) = k + k^\epsilon$ [FS97, Corollary 3.1]. Note that setting either $\epsilon = 0$ in the former or $\epsilon = 1$ in the latter would show that MSAD_Δ is **NP**-hard for a single value of Δ,[15] however these results do not handle the respective ϵ value.

The $\Delta \leq 2$ case. MSAD_Δ is easily seen to be polynomial-time solvable for $\Delta \leq 2$: If $\Delta = 2$, the problem reduces to finding a smallest cycle, which can be done by performing a BFS from each vertex. If $\Delta < 2$, the output subgraph could be either a cycle or a tree. A tree of size t has $t - 1$ edges and hence average degree $2 - \frac{1}{t}$, so we need one with size $t = \lceil \frac{1}{2-\Delta} \rceil$, which can again be found by a BFS from each vertex. We then return either the tree (if it exists) or a smallest cycle (if it exists), depending on which is smaller.

[15] This Δ value would be 4 in [FS97, Corollary 3.1], or twice the constant hidden in the Θ notation of [AHI02, Theorem 1].

Low densities above 2 are as hard as any. We prove the following:

Proposition A.3. *Let $\beta > 0$. For every $\Delta > 2$, there is $\Delta' \in (2, 2+\beta)$, such that there is a polynomial-time reduction of MSAD_Δ to $\mathsf{MSAD}_{\Delta'}$.*

Corollary A.4. *Let $\beta > 0$. If all the problems $\{\mathsf{MSAD}_\Delta : 2 < \Delta < 2+\beta\}$ are polynomial-time solvable, then all the problems $\{\mathsf{MSAD}_\Delta : \Delta > 2\}$ are polynomial-time solvable.*

The proof is by sparsifying the input graph as follows: We split each edge to two edges by adding a new "dummy" vertex, which is in fact equivalent to constructing a bipartite graph that describes the edge-vertex incidences in the original graph. We then repeat the process sufficiently many times.

Definition A.5. *Let $G(V, E)$ be a simple graph. The* incidence graph *B_G of G is the bipartite graph with sides V and E, such that every $v \in V$ and $e \in E$ are connected with an edge in B_G iff v is an endpoint of e in G.*

For a graph G, let us denote by $f_\Delta(G)$ the size of its smallest subgraph with average degree *exactly* Δ. (If no such subgraph exists, set $f_\Delta(G) = \infty$.)

Lemma A.6. *Let $G(V, E)$ be a simple graph with incidence graph B_G, let $\Delta > 2$, and $k > 0$. Denote $\Delta' = 4 - \frac{8}{\Delta+2}$. Then, $f_\Delta(G) \le k$ if and only if $f_{\Delta'}(G) \le k + \frac{1}{2}\Delta k$.*

Proof. Suppose $f_\Delta(G) \le k$. Let $H(U, F)$ be a subgraph of G with size $|U| \le k$ and average degree Δ. Let H' be the subgraph of B_G induced by $U \cup F$. Then H' has $|U| + |F|$ vertices and $2|F|$ edges, as each edge in F is incident in G to two vertices in U, and hence is adjacent in B_G to two vertices in U. Since H has average degree Δ we have $|F| = \frac{1}{2}\Delta|U|$, and since $|U| \le k$, we see that H' has size:

$$|H'| = |U| + |F| = |U| + \tfrac{1}{2}\Delta|U| \le k + \tfrac{1}{2}\Delta k$$

Calculating the average degree of H', we get:

$$\text{avgdeg}(H') = 2 \cdot \frac{2|F|}{|F| + |U|} = 4 - \frac{4|U|}{|F| + |U|} = 4 - \frac{4}{\frac{|F|}{|U|} + 1} = 4 - \frac{8}{\Delta + 2} = \Delta'$$

and this proves one direction.

Conversely, suppose $f_{\Delta'}(G) \le k + \frac{1}{2}\Delta k$. Let H' be a minimum-sized subgraph of B_G with average degree exactly Δ'. By hypothesis we have $\Delta > 2$, and this is easily seen to imply $\Delta' > 2$. Hence by the minimality of the size of H' and Lemma 2.4, we may assume that H' has minimum degree 2.

Recall that B_G is bipartite with sides V and E. Hence we can partition the vertices of H' into two disjoins subsets, $U \subset V$ and $F \subset E$. Since H' has minimum degree 2, and each $e \in F$ has degree exactly 2 in B_G, we infer that e has degree 2 in H'. This means e is adjacent in H' to two vertices in U, so in G, the edge e has both its endpoints in U. In other words, each edge in F is spanned by the vertices in U, so (U, F) is a subgraph H of G.

To calculate the average degree of H, observe that H' has $|U| + |F|$ vertices and $2|F|$ edges, and by hypothesis its average degree is $\Delta' = 4 - \frac{8}{\Delta+2}$. Hence:

$$2 \cdot \frac{2|F|}{|U| + |F|} = 4 - \frac{8}{\Delta + 2}$$

Rearranging gives $\Delta = \frac{2|F|}{|U|}$, and this is the average degree of H.

To calculate the size of H, note that the above equality $\Delta = \frac{2|F|}{|U|}$ is equivalent to $|F| = \frac{1}{2}\Delta|U|$, and recall that by assumption we have $|U| + |F| \leq k + \frac{1}{2}\Delta k$ (since $|U| + |F|$ is the size of H'). Putting these together and rearranging, we get:

$$|U| + \frac{1}{2}\Delta|U| \leq k + \frac{1}{2}\Delta k$$

and hence $|U| \leq k$, so H has size at most k. The proof is complete. □

Proof of Proposition A.3. Let $\Delta > 2$. By Lemma A.6, we can reduce MSAD_Δ to MSAD_{Δ_1} with $\Delta_1 = 4 - \frac{8}{\Delta+2}$. Then we use Lemma A.6 again to reduce MSAD_{Δ_1} to MSAD_{Δ_2}, with $\Delta_2 = 4 - \frac{8}{\Delta_1+2}$. We repeat this sufficiently many times. The sequence $\Delta_{m+1} = 4 - \frac{8}{\Delta_m+2}$ converges to 2, so eventually we get to $\Delta_m < 2+\beta$. The number of repetitions needed depends only on the initial Δ, which is constant, and hence this is a polynomial-time reduction of MSAD_Δ to MSAD_{Δ_m}, for some $\Delta_m < 2 + \beta$. □

B Appendix: Proof of Theorem 3.1

Theorem B.1 (restatement of Theorem 3.1)**.** *Let $\Delta > 1$ be an integer and $c, \epsilon > 0$. There is a constant $C = C(\Delta, \epsilon, c)$, such that every graph on n vertices with average degree d satisfying $\Delta - \epsilon \leq d \leq c \cdot n^{(\Delta-2)/\Delta}$, contains a subgraph of size at most $C \cdot n/d^{\Delta/(\Delta-2)}$ with average degree $\geq \Delta - \epsilon$.*

Proof. Let $G(V, E)$ be a graph as in the statement of the theorem. By Lemma 2.4 we may assume, up to a slight variation of constants, that G has minimum degree d. Moreover it is enough to prove the theorem for all sufficiently large values of d, as the lower values can then be handled by a proper choice of constant C. We use $o(1)$ to denote a term that tends to 0 as d grows.

For each vertex, we fix an arbitrary subset of exactly d of its neighbours, and refer only to them as its neighbours.

We first assume that Δ is even, $\Delta = 2\ell$ for an integer $\ell \geq 2$. We handle two separate cases, according to the range of the density d.

Case I - Low density: Suppose $d \leq n^{(\ell-1)/1.99\ell}$. The proof in this case is very similar to that of Theorem 3.2. Let α be a large constant that will be determined later. Sample a random subset $A \subset V$ by including each vertex in A with independent probability $p = \alpha/d^{\ell/(\ell-1)}$. We refer to vertices in A as *marked*. Note that $|A|$ is binomially distributed with parameters n, p, and that $\mathbb{E}|A| = \alpha n/d^{\ell/(\ell-1)}$.

For each vertex v we fix an arbitrary subset $N(v)$ of exactly d of its neighbours. Define B to be the random subset of vertices v that are not marked, and have exactly ℓ marked neighbours in $N(v)$. We then have,

$$\Pr[v \in B] = (1-p) \cdot \binom{d}{\ell} p^\ell (1-p)^{d-\ell} \geq \tfrac{(1-o(1))}{\ell^\ell}(dp)^\ell = \tfrac{(1-o(1))}{\ell^\ell}\alpha^\ell d^{-\ell/(\ell-1)}$$

where the inequality is by the known bound $\binom{a}{b} \geq (\frac{a}{b})^b$ and by observing that $1-p = 1-o(1)$. By linearity of expectation we get $\mathbb{E}|B| = \frac{(1-o(1))}{\ell^\ell}\alpha^\ell n/d^{\ell/(\ell-1)}$.

By the Chernoff bound Lemma 2.3 applied to $|A|$, we get:

$$\Pr\left[|A| < 2\mathbb{E}|A|\right] \geq 1 - (0.25e)^{\mathbb{E}|A|} = 1 - (0.25e)^{\alpha n/d^{\ell/(\ell-1)}} \geq 1 - (0.25e)^{\alpha d^{0.99\ell/(\ell-1)}}$$

where the final inequality is by the assumption of the current case, $d \leq n^{(\ell-1)/1.99\ell}$. On the other hand, by Lemma 2.1, $|B|$ attains half its expected value with probability at least $\frac{\mathbb{E}|B|}{2n} = \frac{(1-o(1))}{\ell^\ell}\alpha^\ell d^{-\ell/(\ell-1)} = \Omega(d^{-\ell/(\ell-1)})$. Summing the bounds yields:

$$\Pr\left[|A| \leq 2\mathbb{E}|A|\right] + \Pr\left[|B| \geq \tfrac{1}{2}\mathbb{E}|B|\right] \geq 1 - (0.25e)^{\alpha d^{0.99\ell/(\ell-1)}} + \Omega(d^{-\ell/(\ell-1)})$$

The second term in the above right-hand side shrinks exponentially in d, whereas the third term shrinks polynomially. Hence for sufficiently large d the above right-hand side is strictly more than 1, and hence there is a positive probability that both of the events $|A| \leq 2\mathbb{E}[|A|]$

and $|B| \geq \frac{1}{2}\mathbb{E}[|B|]$ occur. We fix this event from now on, and arbitrarily remove vertices from B until $|B| = \frac{1}{2}\mathbb{E}[|B|]$. The following bounds now hold:

$$|A| + |B| \leq 2\mathbb{E}|A| + \frac{1}{2}\mathbb{E}|B| = \left(2\alpha + \tfrac{(1-o(1))}{\ell^\ell}\alpha^\ell\right) \cdot \frac{n}{d^{\ell/(\ell-1)}} \tag{B.1}$$

$$\frac{|B|}{|A|} \geq \frac{\frac{1}{2}\mathbb{E}|B|}{2\mathbb{E}|A|} = \frac{1-o(1)}{4\ell^\ell} \cdot \alpha^{\ell-1} \tag{B.2}$$

We take our target subgraph H to be that induced by $A \cup B$. By eq. (B.1), its size is bounded by $C \cdot n/d^{\ell/(\ell-1)}$, which equals $C \cdot n/d^{\Delta/(\Delta-2)}$, for $C = 2\alpha + \frac{(1-o(1))}{\ell^\ell}\alpha^\ell$. To bound its average degree, note that each vertex in B is incident to ℓ edges connecting it to A, and since A and B are disjoint (recall that vertices in B are not marked), each such edge has a unique end in B. Hence we count at least $\ell|B|$ different edges in H, and find that its average degree is:

$$\text{avgdeg}(H) \geq 2 \cdot \frac{\ell|B|}{|A| + |B|} = 2\ell - \frac{2\ell}{1 + \frac{|B|}{|A|}} \geq 2\ell - \frac{2\ell}{1 + \frac{1-o(1)}{4\ell^\ell} \cdot \alpha^{\ell-1}}$$

using eq. (B.2) for the final inequality. The bound on the right-hand side is guaranteed to be at least $2\ell - \epsilon$ (which equals $\Delta - \epsilon$) as long as we pick α such that $\alpha^{\ell-1} > 16\ell^{\ell+1}/\epsilon$, and the proof for this case is complete.

Case II - High density: Suppose $d > n^{(\ell-1)/1.99\ell}$. In this case, we choose the subset A uniformly at random over all subsets of V with size exactly $a = \alpha n/d^{\ell/(\ell-1)}$. (Again, $\alpha > 0$ is a large constant that will be set later.) Again we refer to vertices in A as *marked*, and define B similarly to the previous case, as the subset of non-marked vertices with exactly ℓ marked neighbours. We now lower-bound the probability of a vertex $v \in V$ to be in B:

$$\Pr[v \in B] = \frac{\binom{d}{\ell}\binom{n-d-1}{a-\ell}}{\binom{n}{a}} \geq \frac{\left(\frac{d}{\ell}\right)^\ell \cdot \frac{(n-d-1)!}{(a-\ell)!(n-d-1-a+\ell)!}}{\frac{n!}{a!(n-a)!}} = \frac{\left(\frac{d}{\ell}\right)^\ell \cdot \frac{a!}{(a-\ell)!} \cdot \frac{(n-d-1)!}{(n-d-1-a+\ell)!}}{\frac{n!}{(n-a)!}}$$

Next we apply the inqealities $(m-k+1)^k \leq \frac{m!}{(m-k)!} \leq m^k$ that holds for all positive integers m, k, and obtain:

$$\Pr[v \in B] \geq \frac{(\frac{d}{\ell})^\ell \cdot (a - \ell + 1)^\ell \cdot (n - d - a + \ell)^{a-\ell}}{n^a} \geq \frac{(\frac{d}{\ell})^\ell \cdot (\frac{1}{2}a)^\ell \cdot (n - d - a + \ell)^{a-\ell}}{n^a}$$

the second inequality is because a tends to infinity with d, so for sufficiently large d we have $a - \ell + 1 \geq \frac{1}{2}a$. (We remark that this is where the constant c from the statement of the theorem comes into play: the growth rate of a depends on c.) We can now rearrange and write:

$$\Pr[v \in B] \geq \left(\frac{ad}{2\ell n}\right)^\ell \cdot \left(\frac{n - d - a + \ell}{n}\right)^{a-\ell} = \left(\frac{ad}{2\ell n}\right)^\ell \cdot \left(1 - \frac{d + a - \ell}{n}\right)^{a-\ell}$$

We claim that the term $\left(1-\frac{d+a-\ell}{n}\right)^{a-\ell}$ is $1-o(1)$. For this it is sufficient to show that $\frac{d+a-\ell}{n}\cdot(a-\ell)=o(1)$, and since ℓ is constant, this is equivalent to showing that $\frac{d+a}{n}\cdot a=o(1)$. Indeed, recalling that $a=\alpha n/d^{\ell/(\ell-1)}$,

$$\frac{d+a}{n}\cdot a=\frac{da}{n}+\frac{a^2}{n}=\frac{\alpha}{d^{1/(\ell-1)}}+\frac{\alpha^2 n}{d^{2\ell/(\ell-1)}}=o(1)+o(1)=o(1)$$

where we have used the assumption of the current case, that $n<d^{1.99\ell/(\ell-1)}$. We conclude,

$$\Pr[v\in B]\geq(1-o(1))\left(\frac{ad}{2\ell n}\right)^{\ell}=\frac{1-o(1)}{(2\ell)^{\ell}}\cdot\alpha^{\ell}\cdot\frac{1}{d^{\ell/(\ell-1)}}$$

and from this point the proof proceeds as in the low density case. This concludes the proof for the case Δ is even.

Odd values of Δ. The proof in this case is a close variant of the above proof for even values of Δ, so we only sketch the differences. Suppose $\Delta=2\ell+1$. Again we mark each vertex with independent probability $p=\alpha/d^{\Delta/(\Delta-2)}$, so the subset A of marked vertices has the "correct" expected size, $\mathbb{E}|A|=\alpha n/d^{\Delta/(\Delta-2)}$.

The difference is that we define B to be the subset of *edges* that each of their two endpoints is non-marked, and has exactly ℓ marked neighbours. As in the even-Δ case, the probability for a vertex to be non-marked and to have ℓ marked neighbours is roughly $q=(dp)^{\ell}$. We claim that in the current setting, each edge has probability roughly q^2 to be in B.

This is a subtle but technical point. Fix an edge $e=uv$, let X_u denote the event that u is non-marked and has ℓ marked neighbours, and similarly define X_v for v. As stated above, each of X_u and X_v occurs with probability q, and we wish to show that both occur (which means e is included in B) with probability $\Omega(q^2)$. In the low density case random model, where the vertices are marked independently, the events X_u, X_v are either positively correlated (if u,v have any mutual neighbours) or independent (otherwise), so the probability for both to occur is indeed at least q^2. In the high density case random model, where we pick a random subset of marked vertices with fixed size, the events X_u, X_v may in fact be negatively correlated, but it is a technicality to calculate the probability for both X_u, X_v to occur and to see that it remains approximately q^2.

Having established that each edge in included in B with probability about q^2, we recall that there are $\frac{1}{2}nd$ edges in G, and hence the expected size of B is roughly (recall that $p=\alpha/d^{\Delta/(\Delta-2)}=\alpha/d^{(2\ell+1)/(2\ell-1)}$):

$$\mathbb{E}|B|=nd\cdot q^2=nd(dp)^{2\ell}=\alpha^{2\ell}n/d^{(2\ell+1)/(2\ell-1)}=\alpha^{2\ell}n/d^{\Delta/(\Delta-2)}$$

which is also the "correct" expected size. The target subgraph H is taken to be the one induced by A and all the endpoints of edges in B. The proof then proceeds as in the even-Δ case. Full details are omitted. $\square$

C Appendix: Proof of Theorem 4.1.3

Theorem 4.1.3 (restated)**.** *Let $\ell > 1$ be an integer and $\epsilon > 0$. There is a constant $C = C(\ell, \epsilon)$ such that every graph on n vertices with average degree d (satisfying $2\ell \leq d \leq O(n^{(\ell-1)/\ell})$) contains a subgraph of size at most $C \cdot n/d^{\frac{\ell}{\ell-1} - \frac{1}{\ell^3 - 2\ell+1} - \epsilon}$ with average degree 2ℓ.*

Proof. Let $G(V, E)$ be a graph as in the statement. We prove the theorem for sufficiently large d, and the lower values can then be handled by a proper choice of constant C. We will use $o(1)$ to denote a term that tends to 0 as d grows.

By Lemma 2.4 we may assume that G has minimal degree $\geq \frac{1}{2}d$. Now by Corollary 2.6, we can orient the edges in G such that each vertex has at least $\frac{1}{4}d$ edges oriented towards it. For each vertex v, we then fix an arbitrary subset of exactly $\frac{1}{4}d$ edges oriented towards v, and call their other endpoints the *in-neighbours* of v.

Set $p = 1/d^{\ell/(\ell-1)-\epsilon}$. We mark the vertices of G in two independent phases: In the first phase, each vertex is marked with independent probability p, and in the second phase, each non-marked vertex is again marked with independent probability p. Let A be the subset of marked vertices; we have $\mathbb{E}[|A|] = (2p - p^2)n$.

Definition C.1. *Let $k \geq 1$ be an integer, and consider a fixed marking of the vertices in G. A vertex $v \in V$ is a* k-root *if it is non-marked and has exactly k marked in-neighbours.*

Proposition C.2. *After the first phase of marking vertices in G, each $u \in V$ has probability $\geq \gamma_\ell \cdot (dp)^{\ell-1}$ to be an $(\ell-1)$-root, where γ_ℓ is a constant that depends only on ℓ.*

Proof. The number of marked in-neighbours of u is binomially distributed with parameters $\frac{1}{4}d$ and p, so its probability to be an $(\ell-1)$-root is at least:

$$(1-p) \cdot \binom{\frac{1}{4}d}{\ell-1} p^{\ell-1}(1-p)^{\frac{1}{4}d-(\ell-1)} \geq \gamma_\ell \cdot (dp)^{\ell-1}$$

for $\gamma_\ell \geq \frac{1}{2}(4(\ell-1))^{-(\ell-1)}$. Note that the leading $(1-p)$ in the above is the probability for u to be non-marked. For the lower bound, we have used the known bound $\binom{m}{k} \geq (\frac{m}{k})^k$, and $(1-p)^{\frac{1}{4}d-(\ell-1)+1} \geq \frac{1}{2}$, which holds for sufficiently large d as we recall $p = 1/d^{\ell/(\ell-1)-\epsilon}$. □

Lemma C.3 (main)**.** *Let $c > 0$ be any constant. If $\epsilon > 1/(\ell^3 - 2\ell + 1)$, then for d sufficiently large, each $v \in V$ has probability $\geq cp$ to be non-marked and to have at least $\ell + 1$ in-neighbours which are ℓ-roots.*

Proof. Fix $v \in V$ and let $N(v)$ be the set of in-neighbours of v. (Recall that $|N(v)| = \frac{1}{4}d$.) Consider G after the first phase of marking vertices. Suppose $u \in N(v)$ is an $(\ell-1)$-root. We call an edge e oriented towards u an *excited* edge, if its source vertex is a non-marked in-neighbour of u. Recall that if u is an $(\ell-1)$-root then it has exactly $\ell - 1$ marked in-neighbours, so at least $\frac{1}{4}d - (\ell-1) \geq \frac{1}{8}d$ non-marked in-neighbours. This means u renders at least $\frac{1}{8}d$ edges excited, and for simplicity, we assume henceforth that it renders *exactly* $\frac{1}{8}d$ edges excited (arbitrarily chosen).

Let X be the set of $(\ell-1)$-roots in $N(v)$ (after only the first phase of marking vertices), and let Y be the set of excited edges. By Proposition C.2, each $u \in N(v)$ has probability $\geq \gamma_\ell (pd)^{\ell-1}$ to be an $(\ell-1)$-root and hence,

$$\mathbb{E}|X| \geq \tfrac{1}{4} d \cdot \gamma_\ell (pd)^{\ell-1} = \tfrac{1}{4} \gamma_\ell d^\ell p^{\ell-1}$$

and since each vertex in X renders $\frac{1}{8}d$ edges excited,

$$\mathbb{E}|Y| \geq \tfrac{1}{8} d \cdot \mathbb{E}|X| \geq \tfrac{1}{32} \gamma_\ell d^{\ell+1} p^{\ell-1} \tag{C.1}$$

Applying Lemma 2.2 we get that for some $t \geq 1$,

$$\Pr[|Y| \geq \tfrac{2^t}{9 \cdot 16} \gamma_\ell d^{\ell+1} p^{\ell-1}] \geq (2^t t^2)^{-1} \tag{C.2}$$

Proposition C.4. *$t < 2 \log d$.*

Proof. Recall that $p = d^{\ell/(\ell-1)-\epsilon}$ and hence,

$$(dp)^{\ell-1} = d^{1-(\ell-1)\epsilon} > d^{-1} \tag{C.3}$$

Now observe that there can be at most $\frac{1}{4}d$ excited edges per vertex in $N(v)$, so a total of at most $\frac{1}{16}d^2$ excited edges. Hence, when applying Lemma 2.2 to derive eq. (C.2), we can set $M = \frac{1}{16}d^2$ in the statement of Lemma 2.2 and obtain:

$$t \leq \log\left(\frac{9M}{2\mathbb{E}|Y|}\right) \leq \log\left(\frac{9}{\gamma_\ell (dp)^{\ell-1}}\right) < \log\left(\frac{9}{\gamma_\ell} \cdot d\right) < 2 \log d$$

where the first inequality is by the guarantee of Lemma 2.2; the second one is by plugging eq. (C.1) for $\mathbb{E}|Y|$ and $M = \frac{1}{16}d^2$; the next one is by eq. (C.3); and the final inequality holds for sufficiently large d, since γ_ℓ is constant. □

We proceed with the proof of the Main Lemma C.3. By eq. (C.2), after the first phase, there is probability $(2^t t^2)^{-1}$ to have $\frac{2^t}{9 \cdot 16} d^{\ell+1} p^{\ell-1}$ excited edges. Let W be the set of source vertices of all the excited edges. We recall that by definition (of an excited edge), each vertex in W is a non-marked in-neighbour of a vertex in X.

For each $w \in W$, we say w is *light* if it is the source of at most ℓ excited edges, and *heavy* if it is the source of at least $\ell + 1$ excited edges. For a fixed marking of the vertices in G after the first phase, either half the excited edges are sourced at light vertices, or half are sourced at heavy vertices. By an averaging argument applied to eq. (C.2), we see that one of these cases must hold with probability $\frac{1}{2}(2^t t^2)^{-1}$. We handle the two cases separately.

Case I - Light vertices: With probability $\frac{1}{2}(2^t t^2)^{-1}$ we have $\frac{2^t \gamma_\ell}{9 \cdot 16} d^{\ell+1} p^{\ell-1}$ excited edges, and half of them are sourced at light vertices of W.

Let L be the subset of light vertices. We have $\frac{2^t \gamma_\ell}{9 \cdot 32} d^{\ell+1} p^{\ell-1}$ excited edges incident to light vertices, and each light vertex is the source of at most ℓ excited edges. Hence, $|L| \geq \frac{2^t \gamma_\ell}{9 \cdot 32 \cdot \ell} d^{\ell+1} p^{\ell-1}$. Arbitrarily remove vertices from L until equality holds. Moreover, the number of excited edges is $\frac{1}{8}d \cdot |X|$, and hence we get $|X| \geq \frac{2^t \gamma_\ell}{36} d^\ell p^{\ell-1}$.

Our intention is now to uniquely assign L-vertices to X-vertices. To this end we consider the bipartite graph with sides X and L and with the excited edges as the edge set. In fact, X and L may intersect; in such case we make two copies of each vertex in the intersection, putting one copy on the X-side and the other on the L-side. Note that all the edges are oriented from L to X.

Side X has maximum degree $\frac{1}{8}d$ and average degree:

$$\frac{|L|}{|X|} \geq \frac{\frac{2^t \gamma_\ell}{9 \cdot 32 \cdot \ell} d^{\ell+1} p^{\ell-1}}{\frac{2^t \gamma_\ell}{36} d^\ell p^{\ell-1}} = \frac{2^t}{36 \cdot \ell} d \geq \frac{1}{36 \cdot \ell} d$$

Hence by Lemma 2.7, there is a subset $X' \subset X$ with size $|X'| \geq \frac{1}{4.5\ell}|X|$ such that each vertex in X' is adjacent to at least $\frac{1}{72\ell}d$ vertices in L. Now consider the bipartite graph with sides X' and L: Side X' has degree (at least) $\frac{1}{72\ell}d$, and side L has degree at most ℓ (since a light vertex is adjacent to at most ℓ vertices in X, and hence in X'). We use following lemma:

Lemma C.5. *Let $G(V, U; E)$ be a bipartite graph such that each $v \in V$ has degree d, and each $u \in U$ has degree at most ℓ. There is a subset of edges E' such that in $G'(V, U; E')$, each $v \in V$ has degree at least $\lfloor \frac{1}{\ell} d \rfloor$, and each $u \in U$ has degree at most 1.*

Proof. Similar to the proof of Lemma 2.5, with every "2" replaced by "ℓ". □

By Lemma C.5, there is an assignment of L-vertices to X'-vertices such that each L-vertex is uniquely assigned, and each X'-vertex has at least $\frac{1}{72\ell^2}d$ vertices assigned to it. Fixed this assignment henceforth.

Let $u \in X'$. Recall that u is in X, which means it is an $(\ell-1)$-roots after the first phase. Moreover by the above it has $\frac{1}{16\ell^2}d$ neighbours in L assigned to it. If one of them is marked in the second phase, then u turns into an ℓ-root. Denote this event by A_u. Its probability is $(1 - o(1)) \cdot \frac{1}{72\ell^2} d \cdot p$, for picking an assigned neighbour and marking it.

Let $\mathcal{A}$ denote the event that $\ell+1$ of the events $\{A_u : u \in X'\}$ occur concurrently after the second phase. If $\mathcal{A}$ occurs, then v has $\ell+1$ neighbours which are ℓ-roots, and the conclusion of Lemma C.3 (which we are now proving) is satisfied. To lower bound its probability, we observe that the events $\{A_u : u \in X'\}$ are pairwise independent, by the uniqueness of the assignment of L-vertices to X'-vertices. Hence (recalling that the probability for the current case is $\frac{1}{2}(2^t t^2)^{-1}$):

$$\Pr[\mathcal{A}] \geq \frac{(1-o(1))}{2 \cdot 2^t t^2} \binom{|X'|}{\ell+1} \left(\frac{1}{72\ell^2} dp\right)^{\ell+1} \geq \frac{(1-o(1))}{2 \cdot 2^t t^2} \cdot \left(\frac{|X'|}{\ell+1} \cdot \frac{1}{72\ell^2} dp\right)^{\ell+1} \tag{C.4}$$

where we have used the known bound $\binom{a}{b} \geq (\frac{a}{b})^b$. We recall:

$$|X'| \geq \frac{1}{4.5\ell}|X| \geq \frac{2^t \gamma_\ell}{162\ell} d^\ell p^{\ell-1}$$

Plugging this back into eq. (C.4), we get:

$$\Pr[\mathcal{A}] \geq \frac{(1-o(1))}{2 \cdot 2^t t^2} \cdot \left(\frac{\frac{2^t \gamma_\ell}{162\ell}}{\ell+1} \cdot \frac{1}{72\ell^2} \cdot d^{\ell+1} p^\ell\right)^{\ell+1}$$

Now applying Proposition C.4 and suppressing all the constants into a constant C_ℓ, we get:

$$\Pr[\mathcal{A}] \geq \frac{C_\ell}{\log^2 d} \cdot d^{\ell^2+2\ell+1} p^{\ell^2+\ell}$$

We need this bound to be at least cp. Rearranging and suppressing c into C_ℓ, we need:

$$\frac{C_\ell}{\log^2 d} \cdot d^{\ell^2+2\ell+1} p^{\ell^2+\ell-1} \geq 1$$

We recall that $p = d^{-\ell/(\ell-1)+\epsilon}$ and plug this into the above, which then becomes:

$$\frac{C_\ell}{\log^2 d} \cdot d^{\epsilon(\ell^3-2\ell+1)-1} \geq 1$$

and this is satisfied as long as $\epsilon > 1/(\ell^3 - 2\ell + 1)$, which holds by hypothesis (of the current Lemma C.3). The proof for this case is complete.

Case II - Heavy vertices: With probability $\frac{1}{2}(2^t t^2)^{-1}$ we have $\frac{2^t \gamma_\ell}{9 \cdot 16} d^{\ell+1} p^{\ell-1}$ excited edges, and half of them are sourced at heavy vertices of W.

A heavy vertex is the source of at least $\ell + 1$ excited edges. Each such edge is oriented towards a vertex in X, which we recall is an $(\ell-1)$-root. Hence marking a single heavy vertex in the second phase turns $\ell + 1$ neighbours of v into ℓ-roots, as the lemma requires.

We have $\frac{2^t \gamma_\ell}{9 \cdot 32} d^{\ell+1} p^{\ell-1}$ excited edges incident to heavy vertices. Each heavy vertex is the source of at most $\frac{1}{4}d$ excited edges (since an excited edge is oriented towards a vertex in $X \subset N(v)$, and $|N(v)| = \frac{1}{4}d$). Hence, the number of heavy vertices is at least:

$$\frac{\frac{2^t \gamma_\ell}{9 \cdot 32} d^{\ell+1} p^{\ell-1}}{\frac{1}{4}d} = \frac{2^t \gamma_\ell}{72} d^\ell p^{\ell-1}$$

so the probability to mark one in the second phase is at least:

$$\frac{1}{2}(2^t t^2)^{-1} \cdot (1 - o(1))p \cdot \frac{2^t \gamma_\ell}{72} d^\ell p^{\ell-1} = \frac{1 - o(1)}{2t^2} \cdot \frac{\gamma_\ell}{72}(dp)^\ell \geq \frac{1 - o(1)}{16 \log^2 d} \cdot \frac{\gamma_\ell}{72}(dp)^\ell$$

using $t < 2 \log d$ by Proposition C.4 for the last inequality. To prove the lemma, we need this lower bound on the probability to be at least cp. Rearranging, we need:

$$p \geq \frac{16 \cdot 72 \cdot c \cdot \log^2 d}{(1 - o(1))\gamma_\ell} \cdot d^{-\ell/(\ell-1)}$$

By recalling that $p = d^{-\ell/(\ell-1)+\epsilon}$, we see that the latter inequality indeed holds for sufficiently large d, which proves the current case.

Concluding the proof of Lemma C.3. Having handled both the light and heavy cases, we have proven that v has probability cp to have $\ell + 1$ in-neighbours which are ℓ-roots. We further need it to be non-marked, which happens with probability $(1-p)^2$ (over the two phases), independently of the marking of any other vertices. Hence the probability that v satisfies the conclusion of the lemma is $(1-p)^2 cp = (1 - o(1))cp$. The lemma is proven by slightly rescaling c. $\square$

Back to the proof of Theorem 4.1.3. Suppose $\epsilon > 1/(\ell^3 - 2\ell + 1)$. Consider the graph after both phases of marking vertices. Recall that A is the subset of marked vertices. Denote by B the subset of vertices that are non-marked, and have $\ell+1$ in-neighbours which are ℓ-roots. Note that by Lemma C.3, each vertex is included in B with probability $\geq cp$, for a constant $c > 0$ of our choice (that will be set later).

We recall that $\mathbb{E}|A| = (2p - p^2)n \geq pn$, that $p = d^{-\ell/(\ell-1)+\epsilon}$, and that $n \geq \Omega(d^{\ell/(\ell-1)})$ (by hypothesis of the theorem). Together we get $\mathbb{E}|A| \geq \Omega(d^{\epsilon})$. Hence by a Chernoff bound (Lemma 2.3) applied to $|A|$,

$$\Pr\left[|A| < 2\mathbb{E}|A|\right] \geq 1 - (0.25e)^{\mathbb{E}|A|} \geq 1 - (0.25e)^{\Omega(d^{\epsilon})}$$

On the other hand we have $\mathbb{E}|B| = cpn$, and by Lemma 2.1:

$$\Pr\left[|B| \geq \tfrac{1}{2}\mathbb{E}|B|\right] \geq \tfrac{\mathbb{E}|B|}{2n} = \tfrac{1}{2}cp = \tfrac{1}{2}cd^{-\ell/(\ell-1)+\epsilon}$$

Summing the bounds yields:

$$\Pr\left[|A| \leq 2\mathbb{E}|A|\right] + \Pr\left[|B| \geq \tfrac{1}{2}\mathbb{E}|B|\right] \geq 1 - (0.25e)^{\Omega(d^{\epsilon})} + \tfrac{1}{2}cd^{-\ell/(\ell-1)+\epsilon}$$

The second term shrinks exponentially in d whereas the third term shrinks polynomially, and hence for sufficiently large d, the above RHS is strictly larger than 1. Therefore, there is a positive probability that both the events $|A| \leq 2\mathbb{E}|A|$ and $|B| \geq \frac{1}{2}\mathbb{E}|B|$ occur. These imply $|A| \leq 4np$ (as we recall $\mathbb{E}|A| \leq 2np$) and $|B| \geq \frac{1}{2}cnp$, respectively. We fix this event from now on, and arbitrarily remove vertices from B until $|B| = \frac{1}{2}cnp$.

Recall that each vertex in B has $\ell+1$ in-neighbours which are ℓ-roots. Let Z be the subset of all the ℓ-roots in-neighbours of vertices in B. Note that B and Z may intersect, and that $|Z| \leq (\ell+1)|B|$. We take our target subgraph H to be one induced by $A \cup B \cup Z$. Its size is bounded by,

$$|H| \leq |A| + |B| + |Z| \leq |A| + (\ell+2)|B| \leq (4 + (\ell+2)\tfrac{1}{2}c)np = (4 + (\ell+2)\tfrac{1}{2}c) \cdot n/d^{\ell/(\ell-1)-\epsilon}$$

which is as required if we set $C = 4 + (\ell+2)\frac{1}{2}c$. We move on to establish that H has ℓ edges per vertex. Each $v \in Z$ is an ℓ-root, i.e. has ℓ marked in-neighbours. Those in-neighbours are in A and hence in H, so the edges oriented from them to v are present in H, and we count them in its favour. Each $v \in B$ has $\ell+1$ in-neighbours in Z, so the edges going from those neighbours into v are present in H. We count ℓ of them in favour of v. Note that so far, each edge was counted in favour of its destination endpoint, and hence was counted only once. We are left to handle vertices in A. Note that for each vertex in B, there is one edge oriented towards it (from a vertex in Z) that we have not yet counted. Together we have $|B|$ edges yet unused, and we now count them in favour of the vertices in A. We now only need $|B| \geq \ell|A|$ to hold; by recalling that $|A| \leq 4np$ and $|B| = \frac{1}{2}cnp$, we achieve this by choosing $c = 8\ell$.

The proof is complete with graph size at most $C \cdot n/d^{\ell/(\ell-1)-\epsilon}$, under the assumption $\epsilon > 1/(\ell^3 - 2\ell + 1)$. This is equivalent to the statement of the theorem, by rescaling $\epsilon' = \epsilon - 1/(\ell^3 - 2\ell + 1)$. ∎

D Appendix: Logarithmic Girth Bounds

We provide known proofs that show $S_1(n, d) = \Theta(\log_{d-1} n)$.

Remark: In Section 1.1 we discussed the open problem of determining the optimal leading constant c to the $\log_{d-1} n$ term in the bound, and mentioned that the best currently known bounds are $\frac{4}{3} \le c \le 2$. The bounds proven here are weaker, $1 \le c \le 4$.

Theorem D.1. *Every graph of size n with average degree d, contains a cycle of length at most $4 \log_{d-1}(n + 1)$.*

Proof. Denote $t = \log_{d-1}(n + 1)$. First suppose that d is the minimum degree. Fix an arbitrary vertex v and consider all paths of length t starting at v. If any two paths intersect at a vertex other than v, then their union contains a cycle of length at most $2t$. Otherwise, all the paths form a tree with $t + 1$ levels such that every vertex has at least $d - 1$ children, so the tree has $\frac{(d-1)^{t+1}-1}{(d-1)-1} > n$ vertices, a contradiction.

Now suppose d is the average degree. By Lemma 2.4, the graph contains a subgraph with minimum degree $\frac{1}{2}d$ or more. By the above, the subgraph contains a cycle of size at most $2 \log_{d/2-1}(n + 1) \le 4t$. $\square$

Theorem D.2. *For every n and d, there is a graph with n vertices and average degree $\ge d$, with girth $> \log_{d+2} n$.*

Proof. Consider the following process for sampling a random graph G with girth $> \log_{d+2} n$:

1. Sample a $G(n, p)$ graph G' with $p = \frac{d+2}{n-1}$.
2. Remove from G' one (arbitrary) edge from each cycle of length $\le \log_{d+2} n$, to get G.

For $k = 3, \ldots, \log_{d+2} n$, let C_k denote the number of length-k cycles in G'. There are $\binom{n}{k}(k-1)! = \frac{n!}{(n-k)!k}$ possible such cycles, and each occurs with probability p^k, hence:

$$\mathbb{E}[C_k] = \frac{n!}{(n-k)!k} p^k = \frac{n!}{(n-k)!k} \left(\frac{d+2}{n-1}\right)^k \le \frac{(d+2)^k}{k} \le \frac{(d+2)^{\log_{d+2} n}}{\log_{d+2} n} = \frac{n}{\log_{d+2} n}$$

Denoting by $C = \sum_k C_k$ the total number of short cycles, we get $\mathbb{E}[C] \le n$. Let E denote the number of edges in G'. Step 2 of the above sampling process removes C edges from G' to obtain G, and hence the average degree of G is $\Delta = \frac{2(E-C)}{n}$. Moving to expectation,

$$\mathbb{E}[\Delta] = \frac{2}{n} (\mathbb{E}[E] - \mathbb{E}[C]) \ge \frac{2}{n} \left(\frac{d+2}{n-1} \binom{n}{2} - n \right) = d$$

by recalling that $\mathbb{E}[E] = p\binom{n}{2}$ and our choice $p = \frac{d+2}{n-1}$. We therefore conclude there is a sample of G with average degree $\Delta \ge d$. $\square$