Long paths and cycles
in random subgraphs of $\mathcal{H}$-free graphs

Michael Krivelevich[*]

School of Mathematical Sciences
Tel Aviv University
Tel Aviv 69978, Israel

krivelev@post.tau.ac.il

Wojciech Samotij[†]

School of Mathematical Sciences
Tel Aviv University
Tel Aviv 69978, Israel

and

Trinity College
Cambridge, CB2 1TQ, UK

samotij@post.tau.ac.il

Abstract

Let $\mathcal{H}$ be a given finite (possibly empty) family of connected graphs, each containing a cycle, and let G be an arbitrary finite $\mathcal{H}$-free graph with minimum degree at least k. For $p \in [0, 1]$, we form a p-random subgraph G_p of G by independently keeping each edge of G with probability p. Extending a classical result of Ajtai, Komlós, and Szemerédi, we prove that for every positive ε, there exists a positive δ (depending only on ε) such that the following holds: If $p \geqslant \frac{1+\varepsilon}{k}$, then with probability tending to 1 as $k \to \infty$, the random graph G_p contains a cycle of length at least $n_{\mathcal{H}}(\delta k)$, where $n_{\mathcal{H}}(k) > k$ is the minimum number of vertices in an $\mathcal{H}$-free graph of average degree at least k. Thus in particular G_p as above typically contains a cycle of length at least linear in k.

1 Introduction

Given a graph G and a real number $p \in [0, 1]$, we define the *p-random subgraph of G*, denoted by G_p, to be the random subgraph of G such that each edge of G belongs to G_p with probability p, independently of all other edges. The most studied case of the above model is when G is a complete graph. This particular model, usually denoted by $G(n, p)$,

---

where n is the number of vertices in the base (complete) graph, was first introduced in [10] and has since become one of the most popular objects of study in combinatorics.

In the groundbreaking paper of Erdős and Rényi [7], the following fundamental discovery was made: If we let $p(n)$ gradually increase from 0 to 1, then the connectivity structure of the random graph $G(n, p)$ undergoes a dramatic *phase transition* around $p(n) = \frac{1}{n}$. For any positive constant ε, if $p(n) \leqslant \frac{1-\varepsilon}{n}$, then asymptotically almost surely[1] (a.a.s.), the size of each connected component of $G(n, p)$ is at most logarithmic in n, whereas if $p(n) \geqslant \frac{1+\varepsilon}{n}$, then a.a.s. $G(n, p)$ has a unique component of linear size, traditionally called the *giant component*. The paper of Erdős and Rényi has had an enormous influence on the development of the theory of random graphs. Its main results have been given several different proofs and extended or improved in many different ways. For a detailed account of the theory of random graphs, we refer the reader to the two standard monographs [3, 11].

One of the better-known extensions of the main result of [7] is due to Ajtai, Komlós, and Szemerédi [1], who proved that if $p(n) \geqslant \frac{1+\varepsilon}{n}$, then not only the random graph $G(n, p)$ a.a.s. contains a giant component occupying a positive proportion of all the vertices, but also it typically has a path of length linear in n. An easy corollary of this fact is that if $p(n) \geqslant \frac{1+\varepsilon}{n}$, then a.a.s. $G(n, p)$ contains a cycle of length linear in n. A simple proof of this result has recently been given in [13].

The following natural generalizations of the classical results about the evolution of the random graph mentioned above were recently considered in [8, 12, 13, 16]. Suppose that $p_0 \colon \mathbb{N} \to [0, 1]$ is a *threshold function* for some (monotone) graph property $\mathcal{P}$ in $G(n, p)$. For example, one might let $p_0(n) = \frac{1}{n}$ and $\mathcal{P}$ be the property of containing a connected component (or a path / cycle) of size (length) linear in n. In particular, suppose that if $p \geqslant (1 + \varepsilon)p_0$ for some positive constant ε, then a.a.s. $G(n, p)$ possesses $\mathcal{P}$. Does this statement remain true if one replaces $G(n, p)$ with the p-random subgraph G_p of an *arbitrary* graph G with minimum degree $n - 1$? This has been answered in the affirmative in several cases, e.g., when $\mathcal{P}$ is the property of being non-planar [8], containing a path of length linear in n [13] or of length at least $n - 1$ [12], and having a cycle of length $n - o(n)$ [12]. As a by-product, more robust proofs of the corresponding statements in the case $G = K_n$ were obtained.

In this paper, we continue this study and introduce one additional twist. Namely, we fix a finite family $\mathcal{H}$ of graphs and further assume that the base graph G is $\mathcal{H}$-*free*, i.e., that G does not contain a copy of any $H \in \mathcal{H}$ as a subgraph. Since we allow the family $\mathcal{H}$ to be empty, which imposes no additional restrictions on G, our results will generalize some previous works. Our aim is to prove that if G is an $\mathcal{H}$-free graph with minimum degree at least k (from now on, we will use k to denote the lower bound on the minimum degree and n to denote the number of vertices of the graph G) and $p \geqslant \frac{1+\varepsilon}{k}$, then with probability approaching 1 as $k \to \infty$, the random graph G_p contains a long path and a long cycle. Since we are only interested in the asymptotic behavior of these probabilities, we will assume that there exist $\mathcal{H}$-free graphs with arbitrary large minimum degree. This implies, in particular, that the family $\mathcal{H}$ cannot contain any acyclic graphs. Moreover, we will assume for convenience that every graph in $\mathcal{H}$ is connected. We will

---

[1]That is, with probability tending to 1 as $n \to \infty$.

term such families *good*. That is, we will say that a family $\mathcal{H}$ of graphs is good if it is a finite collection of connected graphs, each containing a cycle. Finally, we will assume throughout the paper that the base graph G is finite.

Recall that the *Turán number for* $\mathcal{H}$, denoted by $\mathrm{ex}(n, \mathcal{H})$, is the maximum number of edges in an $\mathcal{H}$-free graph on n vertices. Observe that if G is an $\mathcal{H}$-free graph with average degree at least k, then the number n of vertices of G satisfies

$$nk \leqslant 2 \cdot \mathrm{ex}(n, \mathcal{H}). \tag{1}$$

Let $n_{\mathcal{H}}(k)$ be the smallest n for which (1) holds, that is, the smallest number of vertices in an $\mathcal{H}$-free graph of average degree at least k. Our main result is the following.

Theorem 1. *For every positive ε, there exists a positive constant δ such that the following is true. Let k be a sufficiently large integer, let $\mathcal{H}$ be a good family of graphs, and let G be an $\mathcal{H}$-free graph with minimum degree at least k. If $p \geqslant \frac{1+\varepsilon}{k}$, then*

$$\Pr\big(G_p \text{ contains a cycle of length at least } n_{\mathcal{H}}(\delta k)\big) \geqslant 1 - \exp(-\delta k).$$

It is plausible that $n_{\mathcal{H}}(\delta k)$ above could be replaced by $\delta n_{\mathcal{H}}(k)$, but unfortunately our methods do not yield this stronger statement, cf. Section 6.1. Note that $n_{\mathcal{H}}(k) \geqslant k + 1$ for every family $\mathcal{H}$ and that $n_{\mathcal{H}}(k) \leqslant 2k$ if $\mathcal{H}$ does not contain any bipartite graph. Consequently, the interesting cases will be either when $\mathcal{H}$ is empty or when it contains at least one bipartite graph. Indeed, when we let $\mathcal{H}$ be the empty family, we obtain the following corollary.

Corollary 2. *For every positive ε, there exists a positive constant δ such that the following is true. Let k be a sufficiently large integer and let G be an arbitrary graph with minimum degree at least k. If $p \geqslant \frac{1+\varepsilon}{k}$, then*

$$\Pr\big(G_p \text{ contains a cycle of length at least } \delta k\big) \geqslant 1 - \exp(-\delta k).$$

We remark that the statement obtained from Corollary 2 by replacing 'a cycle of length δk' with 'a path of length $\varepsilon^2 k/5$' was proved in [13]. Still, when G is an arbitrary graph, one cannot easily deduce Corollary 2 from its 'path version' using a standard double exposure (sprinkling) argument as, unlike the case when G is a complete graph, there is no guarantee that G contains any edges that close a given path of length $\Theta(k)$ into a cycle of comparable length. For example, the girth of G can be much larger than k.

Another case which seems especially interesting is when $\mathcal{H}$ is the family of all cycles of lengths ranging from 3 to some g. Note that in this case, requiring a graph to be $\mathcal{H}$-free is the same as requiring that its girth exceeds g. Since if g is even, then $\mathrm{ex}(n, C_g) = O(n^{1+2/g})$, as proved by Bondy and Simonovits [4], Theorem 1 has the following nice corollary.

Corollary 3. *For every $g \in \{2, 3, \ldots\}$ and every positive ε, there exists a positive constant δ such that the following is true. Let k be a sufficiently large integer and let G be an arbitrary graph with minimum degree at least k and girth larger than g. If $p \geqslant \frac{1+\varepsilon}{k}$, then*

$$\Pr\big(G_p \text{ contains a cycle of length at least } \delta k^{\lfloor g/2 \rfloor}\big) \geqslant 1 - \exp(-\delta k).$$

Observe that Corollary 3 remains true when one replaces the assumption that the girth of G is larger than g with the weaker assumption that G does not contain a cycle of length $2\lfloor \frac{g}{2} \rfloor$.

The study of circumference (the length of a longest cycle) of graphs with given girth and minimum degree was initiated by Ore [15] and has since attracted the attention of many researchers, see [5, 17] and references therein. Several years ago, this study culminated in a result of Sudakov and Verstraëte [17], who proved that every graph with minimum degree at least k whose girth exceeds g contains a cycle of length $\Omega(k^{\lfloor g/2 \rfloor})$. Actually, it is proved in [17] that if $\mathcal{H} = \{H\}$, where H is a bipartite graph containing a cycle, then every $\mathcal{H}$-free graph G with average degree k contains a family of cycles whose lengths are $n_{\mathcal{H}}(\delta k)$ consecutive even integers, for some positive constant δ. In particular, every such G contains a cycle of length at least $n_{\mathcal{H}}(\delta k)$. As every graph with average degree k contains a subgraph with minimum degree at least $k/2$, the last statement is also a straightforward corollary of our Theorem 1.

It is perhaps a good point to discuss yet another interpretation of our results, related to robustness of graph properties. The general approach of *robustness*, explicitly promoted in [12], suggests to investigate whether graph theoretic properties and statements typically remain valid under taking random subgraphs – which would then indicate that they are robust under (massive) random deletions. For example, it is elementary to prove that any graph G of minimum degree at least k, $k \geqslant 2$, contains a cycle of length at least $k + 1$. Corollary 2 shows that this property typically stays with the random subgraph G_p of G, even when the edge probability p is only a notch above the critical probability $p^* = \frac{1}{k}$. Moreover, if in addition G is assumed to be $\mathcal{H}$-free, then not only G contains deterministically a cycle of length at least $n_{\mathcal{H}}(\delta k)$, but the p-random subgraph G_p of G retains this property with probability exponentially (in k) close to 1, even for $p = \frac{1+\varepsilon}{k}$. Thus the property of containing long cycles is robust under taking random subgraphs. This complements in a substantial way the qualitative statement of [12, Theorem 1.3] (see also [16]), which says that if the minimum degree of G is at least k and $p = \frac{\omega(1)}{k}$, then with probability tending to 1 as $k \to \infty$, G_p contains a cycle of length $k - o(k)$.

Even though the existence of a cycle of length ℓ in a graph immediately implies the existence of a path of length $\ell - 1$, we give a separate, much shorter, argument to prove that if G is an $\mathcal{H}$-free graph with minimum degree at least k and $p \geqslant \frac{1+\varepsilon}{k}$, then with probability close to 1, the random graph G_p contains a path of length $n_{\mathcal{H}}(\delta k)$ for some positive contant δ, see Theorem 4 below. Our proof of Theorem 4 is a fairly straightforward adaptation of the argument given in [13].

Theorem 4. *Let $\varepsilon \in (0,1)$, let k be an integer, let $\mathcal{H}$ be a good family of graphs, and let ℓ be an integer satisfying*

$$\frac{\mathrm{ex}(6\ell/\varepsilon, \mathcal{H})}{\ell} \leqslant \frac{k}{2}. \tag{2}$$

If G is an $\mathcal{H}$-free graph with minimum degree at least k and $p \geqslant \frac{1+\varepsilon}{k}$, then

$$\Pr\left(G_p \text{ contains a path of length } \ell\right) \geqslant 1 - 3\exp\left(-\frac{\varepsilon^3 k}{300}\right).$$

Note the more explicit, as compared to Theorem 1, dependence of ℓ on $\mathcal{H}$, k, and ε. In order to see that Theorem 4 implies that G_p typically contains a path of length $n_{\mathcal{H}}(\delta k)$, note that (2) is satisfied when $\ell = n_{\mathcal{H}}(\delta k)$ and δ is sufficiently small (as a function of ε), cf. the definition of $n_{\mathcal{H}}(k)$ below (1) and Lemma 6. In particular, observe that when the family $\mathcal{H}$ is empty, then (2) is satisfied when $\ell = \varepsilon^2 k / 36$, and hence Theorem 4 implies that with probability very close to 1, G_p contains a path of length $\varepsilon^2 k / 36$. This is somewhat weaker than [13, Theorem 4], which asserts that under the same assumptions, i.e., $\delta(G) \geqslant k$ and $p \geqslant \frac{1+\varepsilon}{k}$, with probability tending to 1 as $k \to \infty$, the random graph G_p contains a path of length $\frac{\varepsilon^2 k}{5}$, but is still optimal up to a constant factor as when $p = \frac{1+\varepsilon}{k}$, then a.a.s. the longest path in $G(k+1, p)$ has length at most $2\varepsilon^2 k$, see, e.g., [11, Theorem 5.17].

At the heart of our proofs of Theorems 1 and 4 lies the analysis of the execution of the depth-first search algorithm on the random graph G_p. This approach to investigating the properties of random graphs near the threshold for the appearance of the giant component was considered in [13], and our work draws heavily from there. Having said that, we would like to stress the fact that our proof of Theorem 1 is not a mere adaptation of the arguments from [13] and employs several novel ideas.

The remainder of the paper is organized as follows. In Section 2, we introduce some notational conventions and list several auxiliary lemmas that we will refer to in the proofs of our main results. In Section 3, we describe the depth-first search algorithm and list some of its properties for later reference. Sections 4 and 5 contain proofs of Theorems 4 and 1, respectively. We close with some concluding remarks and open problems, in Section 6.

2 Preliminaries

2.1 Notation

We use standard graph theoretic notation. In particular, given a graph G, we denote its vertex set by $V(G)$ and the number of edges by $e(G)$. Given a set $A \subseteq V(G)$, we denote the subgraph of G induced by the set A by $G[A]$. For $v \in V(G)$ and $A \subseteq V(G)$, we denote the number of neighbors of v (the degree of v in G) and the number of neighbors of v in the set A by $\deg_G(v)$ and $\deg_G(v, A)$, respectively. For two disjoint sets $A, B \subseteq V(G)$, we write $e_G(A, B)$ to denote the number of edges of G with one endpoint in A and one endpoint in B.

We shall now introduce the notion of *excess edges*, which will play a crucial role in the proof of Theorem 1. Suppose that G is an n-vertex graph with r connected components of sizes $n_1, \ldots, n_r$, respectively. Since each connected component contains a spanning tree, then clearly $e(G) \geqslant n_1 - 1 + \ldots + n_r - 1 = n - r$. The number of excess edges of G, which we denote by $\mathrm{excess}(G)$, is the difference between $e(G)$ and this trivial lower bound. In other words, we let

$$\mathrm{excess}(G) = e(G) - |V(G)| + \#\text{connected components of } G. \qquad (3)$$

Since adding an edge to a graph is easily seen no to decrease its excess, it follows that if G is a graph and U is a (not necessarily induced) subgraph of G, then

$$\text{excess}(U) \leqslant \text{excess}(G). \tag{4}$$

We will use this simple observation in the proof of our main result.

Finally, let us remark that we will repeatedly omit rounding symbols whenever they are not crucial and treat large numbers as integers.

2.2 Tools

In our proofs, we will use the following standard estimate on tail probabilities of the binomial distribution, see, e.g., [2, Appendix A].

Lemma 5. *Let n be a positive integer, let $p \in [0, 1]$, and let $X \sim \text{Bin}(n, p)$.*

(i) *(Chernoff's inequality) For every positive a with $a \leqslant np/2$,*

$$P(|X - np| > a) < 2\exp\left(-\frac{a^2}{4np}\right).$$

(ii) *For every positive κ,*

$$P(X > \kappa np) \leqslant \left(\frac{e}{\kappa}\right)^{\kappa np}.$$

In the proof of Theorem 1, we will use the following simple estimates on the rate of growth of the Turán function, Lemmas 6 and 7 below.

Lemma 6. *Let $\mathcal{H}$ be an arbitrary family of graphs and let m and n be integers with $n \geqslant m \geqslant 2$. Then*

$$\text{ex}(n, \mathcal{H}) \leqslant \left(\frac{n-1}{m-1}\right)^2 \cdot \text{ex}(m, \mathcal{H}).$$

Proof. Let G be an $\mathcal{H}$-free graph with n vertices and $\text{ex}(n, \mathcal{H})$ edges. The subgraph of G induced by any set of m vertices has at most $\text{ex}(m, \mathcal{H})$ edges and therefore,

$$\binom{n}{m} \cdot \text{ex}(m, \mathcal{H}) \geqslant \binom{n-2}{m-2} \cdot \text{ex}(n, \mathcal{H}),$$

which easily implies the claimed inequality. □

Lemma 7. *Let $\mathcal{H}$ be a good family of graphs and let m and n be integers with $n \geqslant m \geqslant 2$. Then*

$$\frac{\text{ex}(m, \mathcal{H})}{m} \leqslant 2 \cdot \frac{\text{ex}(n, \mathcal{H})}{n}.$$

Proof. The claimed inequality is an immediate consequence of the simple observation that, since each graph in the family $\mathcal{H}$ is connected, an $\mathcal{H}$-free graph on n vertices can be obtained by taking $\lfloor \frac{n}{m} \rfloor$ vertex-disjoint copies of an $\mathcal{H}$-free graph with m vertices and $\text{ex}(m, \mathcal{H})$ edges, thus implying that $\text{ex}(n, \mathcal{H}) \geqslant \lfloor \frac{n}{m} \rfloor \cdot \text{ex}(m, \mathcal{H}) \geqslant \frac{n}{2m} \cdot \text{ex}(m, \mathcal{H})$. □

3 Depth-first search algorithm

At the heart of our approach lies the *depth-first search algorithm* (DFS algorithm for short), which is a well-known graph exploration method. We briefly describe it below.

The DFS algorithm takes as input a finite graph G on a vertex set V and, after visiting all vertices of G, outputs a rooted spanning forest F of G such that the connected components of F are the connected components of G. Since we are interested not only in the connectivity structure of G but also in its cycles, we make our algorithm eventually examine all the edges of G, not only those that belong to F. Consequently, our version of the DFS algorithm runs in two phases. In the first phase, the algorithm discovers the connected components of G and constructs the spanning forest F. In the second phase, it examines the remaining edges of G, whose both endpoints lie in the same tree of F (connected component of G).

At all times, the algorithm maintains a partition of V into three sets S, T, and U. At any given moment, the set S contains vertices whose exploration is complete (i.e., whose neighborhood in F has been fully determined), T is the set of vertices that have not yet been visited, and U consists of vertices that are being explored. The vertices in U are kept in a stack, that is, a last-in-first-out data structure. The algorithm starts with $T = V$ and $S = U = \emptyset$ and, as it examines the edges of G, the vertices of G are moved from T to U and from U to S. The algorithm switches from the first to the second phase when $S = V$ and $U = T = \emptyset$. Finally, the procedure terminates when all the edges of G have been examined.

We assume that the set V is equipped with some canonical linear order $\preccurlyeq$. The first phase of the execution of the DFS algorithm can be divided into $2|V|$ rounds. At the beginning of each round, the algorithm checks whether or not U is empty. If $U = \emptyset$, then the $\preccurlyeq$-smallest vertex of T is moved to U. Otherwise, the algorithm considers the top (most recently added) element v of U and queries whether $\{v, w\}$ is an edge of G for some $w \in T$, examining the vertices of T from the $\preccurlyeq$-smallest to the $\preccurlyeq$-largest. If v does have a neighbor w in T, then the $\preccurlyeq$-smallest such w is moved from T to U, becoming its top element; no more queries about whether or not $\{v, w'\}$ is an edge of G for some $w' \in T$ with $w \prec w'$ are asked in this round. Otherwise, if the top element v of U has no neighbors in T, then v is moved to S. Finally, unless $S = V$, the algorithm proceeds to the next round. Observe that in each round exactly one vertex moves, either from T to U or from U to S. Since at the end of the first phase all vertices have made their way from T to S, passing through U, the number of rounds is indeed $2|V|$.

At the end of the first phase, the DFS algorithm has constructed a rooted spanning forest F of the input graph G. The root of each tree in F is the first vertex of it that was moved from T to U. The edges of F are precisely those pairs $\{v, w\} \in G$ such that at some point during the execution of the algorithm, v was the top element of U and w was the $\preccurlyeq$-smallest neighbor of v in T. It is possible that at the end of the first phase, some pairs of vertices have not yet been queried. Note that if $\{v, w\}$ is such a pair, then necessarily v and w belong to the same tree component of F found by the algorithm and, moreover, v is a predecessor of w (or vice-versa) in this rooted tree, as otherwise

the algorithm would have queried $\{v, w\}$. In particular, each such v and w are connected by a unique path in this tree. We denote the length of this path by $\lambda(v, w)$. In order to complete the exploration of all edges of G, in the second phase of its execution, the algorithm queries all the previously not queried pairs $\{v, w\}$, ordered according to the value of $\lambda(v, w)$, from the smallest to the largest. We break ties arbitrarily, that is, the pairs with the same value of $\lambda(\cdot, \cdot)$ are queried in an arbitrary order.

Finally, we list several properties of the DFS algorithm for future reference:

1. The algorithm starts exploring a connected component C of G at the moment the $\preccurlyeq$-smallest vertex of C is moved into the (empty beforehand) set U and completes discovering C when U becomes empty again. At the moment when C is fully discovered, all of its vertices are in S.

2. If $T \neq \emptyset$, then every positively answered query increases the size of U by one.

3. At any stage, the algorithm has queried all pairs $\{v, w\}$ with $v \in S$ and $w \in T$ and found out that $\{v, w\} \notin G$.

4. The set U always spans a path in G.

In several of our proofs, we will analyze the execution of the DFS algorithm on some (random) subgraph of a given graph G. Our assumption will be that the base graph G is known to the algorithm and it thus asks queries only about the edges of G. We will often use the fact that our graph exploration algorithm implicitly defines a bijection φ between the set of all $\{0, 1\}$-sequences of length $e(G)$ and the family of all subgraphs of G, which pairs subgraphs with m edges with sequences containing exactly m ones. This bijection is defined as follows: Given a graph $G' \subseteq G$, we run the DFS algorithm with input G'. We start with $\varphi(G')$ being the empty sequence and each time the algorithm queries whether some $\{v, w\} \in G$ is an edge of G', we append to $\varphi(G')$ the answer to this query, i.e., 1 if $\{v, w\} \in G'$ and 0 otherwise. We will sometimes say that G' is represented by the sequence $\varphi(G')$.

Conversely, any $\{0, 1\}$-sequence (X_i) of length $e(G)$ represents a subgraph G' of G that is described as follows. We run the DFS algorithm and each time it queries whether some pair $\{v, w\} \in G$ is an edge of G', we let $\{v, w\} \in G'$ if $X_i = 1$ and $\{v, w\} \notin G'$ otherwise, where i is the number of the query (the algorithm examines every edge of G exactly once). In particular, if (X_i) is a sequence of i.i.d. Bernoulli random variables with success probability p, then G' is the random subgraph G_p of G.

4   Proof of Theorem 4

Let G be a finite $\mathcal{H}$-free graph with minimum degree at least k and assume that $p \geqslant \frac{1+\varepsilon}{k}$. Without loss of generality, we may assume that ℓ is the largest integer satisfying (2). This quantity is well defined as the assumptions that $\mathcal{H}$ is finite and that each graph in $\mathcal{H}$ contains a cycle imply that $\mathrm{ex}(n, \mathcal{H}) = \Omega(n^{1+\eta})$ for some positive constant η, since

for every integer $g \geqslant 3$, there are well-known constructions of n-vertex graphs with girth exceeding g and $\Omega(n^{1+1/(g-1)})$ edges, see, e.g., [6]. In particular,

$$\frac{36(\ell+1)}{2\varepsilon^2} \geqslant \frac{\mathrm{ex}\big(6(\ell+1)/\varepsilon, \mathcal{H}\big)}{\ell+1} \geqslant \frac{k}{2},$$

and hence

$$\ell > \frac{\varepsilon^2 k}{36} - 1. \tag{5}$$

Note also that the number n of vertices of G satisfies $nk \leqslant 2 \cdot \mathrm{ex}(n, \mathcal{H})$, see (1), which implies that $n > 6\ell/\varepsilon$, as otherwise Lemma 7 and (2) would yield

$$\frac{\mathrm{ex}(n, \mathcal{H})}{n} \leqslant 2 \cdot \frac{\mathrm{ex}(6\ell/\varepsilon, \mathcal{H})}{6\ell/\varepsilon} \leqslant \frac{\varepsilon k}{6} < \frac{k}{2},$$

a contradiction.

Consider a random $\{0, 1\}$-sequence X of length $e(G)$ whose all positions are i.i.d. Bernoulli random variables with success probability p. We will show that if we run the DFS algorithm on the random subgraph of G represented by X, then with probability at least $1 - 3\exp(-\varepsilon^3 k/300)$, at the moment when $|S \cup U|$ reaches $6\ell/\varepsilon$, the set U contains at least ℓ elements. (To see that such moment must occur, recall that in each round of the DFS algorithm, the size of $S \cup U$ does not change or increases by one and that eventually $|S| = n > 6\ell/\varepsilon$.) Since U spans a path in this random subgraph, see property (4) of the DFS algorithm, the assertion of the theorem will follow.

Consider the moment when $|S \cup U|$ reaches $6\ell/\varepsilon$ and suppose to the contrary that $|U| < \ell$. Let Q be the number of queries about edges of G that have been asked so far. By property (3) of the DFS algorithm,

$$Q \geqslant e(S, T) = \sum_{v \in S} \deg(v) - 2e(S) - e(S, U) \geqslant |S| \cdot k - 2e(S \cup U)$$
$$\geqslant \left(\frac{6\ell}{\varepsilon} - \ell\right) \cdot k - 2 \cdot \mathrm{ex}(6\ell/\varepsilon, \mathcal{H}) \geqslant \frac{6k\ell}{\varepsilon} - 2k\ell.$$

Let P be the number of positively answered queries among the first $(6/\varepsilon - 2)k\ell$ queries about edges of G. Since P is a binomial random variable, it follows from Chernoff's inequality (Lemma 5) that with probability at least $1 - 2\exp(-\varepsilon\ell/8)$, which, by (5), is at least $1 - 3\exp(-\varepsilon^3 k/300)$,

$$P > \left(1 + \frac{\varepsilon}{2}\right)\left(\frac{6}{\varepsilon} - 2\right)\ell > \frac{6\ell}{\varepsilon},$$

where in the last inequality we used the assumption that $\varepsilon < 1$. Finally, since T is still non-empty (as $n > 6\ell/\varepsilon = |S \cup U|$), it follows from property (2) of the DFS algorithm that $|S \cup U| \geqslant P$, a contradiction. $\qquad\square$

5 Proof of Theorem 1

5.1 Proof outline

Let us start by giving a brief outline of the proof. Suppose that G is an $\mathcal{H}$-free graph with n vertices and minimum degree at least k, for some sufficiently large integer k, and that $p \geqslant \frac{1+\varepsilon}{k}$. The key step in the proof is to show that with high probability, the random graph G_p contains $\Omega(n)$ excess edges. To this end, we first prove that with probability $\Omega(1)$, a positive proportion of the vertices of G_p lies in large connected components (Theorem 9). Second, we observe that it is extremely unlikely that there is a set B of $\Omega(n)$ vertices that belong only to large connected components in G_p but $G[B]$ has merely $o(|B| \cdot k)$ edges. It then follows from a standard double exposure argument that with positive probability, G_p has $\Omega(n)$ excess edges (Theorem 10). Finally, we note that the number of excess edges in a random graph is tightly concentrated around its expectation (Proposition 8) and therefore, $\mathrm{excess}(G_p) = \Omega(n)$ with very high probability (Corollary 11). This means that when we run the DFS algorithm on a typical G_p, then the number of queries about edges of G asked in the second phase of its execution is $\Omega(nk)$. Since the graph G is $\mathcal{H}$-free, which implies upper bounds on the densities of induced subgraphs of G, at least $\Omega(k^2)$ of these queries are about edges $\{u, v\} \in G$ such that $\lambda(u, v) = \Omega(n_{\mathcal{H}}(k))$. With high probability, one of these pairs is an edge of G_p; this edge closes a cycle of length $\Omega(n_{\mathcal{H}}(k))$ in G_p.

5.2 Bounding the number of excess edges

As mentioned above, the key ingredient in our proof of Theorem 1 is the fact that if G is a graph with minimum degree at least k and $p \geqslant \frac{1+\varepsilon}{k}$, then with high probability, the random graph G_p contains $\Omega(|V(G)|)$ excess edges, see Corollary 11 below. This statement will be an immediate consequence of the following two facts. First, in Proposition 8, using a martingale concentration result, we prove that for arbitrary graph G and probability p, the number of excess edges in the p-random subgraph of G is concentrated around its expectation. Second, in Theorem 10, we show that if $\delta(G) \geqslant k$ and $p \geqslant \frac{1+\varepsilon}{k}$, then this expectation is at least $c \cdot |V(G)|$ for some positive constant c.

Proposition 8. *Let G be an arbitrary graph and let $p \in [0, 1/2]$. Let μ be the expected number of excess edges in the random graph G_p. Then for every $\beta \in (0, 1]$,*

$$\Pr\left(|\mathrm{excess}(G_p) - \mu| \geqslant \beta p e(G)\right) \leqslant 2 \exp\left(-\frac{\beta^2 p e(G)}{3}\right).$$

Proof. Let $m = e(G)$ and fix an arbitrary ordering $e_1, \ldots, e_m$ of the edges of G. For each i, let X_i be the indicator random variable of the event $e_i \in G_p$. Fix an i, let A be an arbitrary subset of $\{e_1, \ldots, e_{i-1}\}$, and let $\mathcal{A}_{i,A}$ denote the event that $G_p \cap \{e_1, \ldots, e_{i-1}\} = A$. Following McDiarmid [14], we define $g_{i,A} \colon \{0, 1\} \to \mathbb{N}$ by

$$g_{i,A}(x) = \mathbb{E}\left[\mathrm{excess}(G_p) \mid \mathcal{A}_{i,A} \text{ and } X_i = x\right] - \mathbb{E}\left[\mathrm{excess}(G_p) \mid \mathcal{A}_{i,A}\right].$$

The function $g_{i,A}$ measures how much the expected number of excess edges in G_p changes when it is revealed whether e_i is or is not an edge of G_p. Observe crucially that the function excess($\cdot$) is edge Lipschitz, i.e., adding or deleting a single edge to/from a graph changes the number of excess edges by at most one. It follows that $|g_{i,A}(1) - g_{i,A}(0)| \leqslant 1$ for all i and A. Applying [14, Theorem 3.9] to the sequence (X_i) with $f = $ excess and $t = \beta pe(G)$, noting that $b = $ maxdev $\leqslant 1$ and $\hat{r}^2 \leqslant e(G)$, we get

$$\Pr\left(|\mathrm{excess}(G_p) - \mu| \geqslant \beta pe(G)\right) \leqslant 2\exp\left(-\frac{(\beta pe(G))^2}{2pe(G) + \beta pe(G)}\right) \leqslant 2\exp\left(-\frac{\beta^2 pe(G)}{3}\right).$$

$\square$

Our proof of Theorem 10 will use the following fact, which is implicit in many earlier works on the phase transition in $G(n, p)$, see, e.g., [3, 11]. Our proof here, which we include for the sake of completeness, follows the approach of [13].

Theorem 9. *Let $\varepsilon \in (0, 1/3)$, let k be an arbitrary integer, let G be a graph with minimum degree at least k, and let v be an arbitrary vertex of G. If $p \geqslant \frac{1+\varepsilon}{k}$, then*

$$\Pr\left(\text{the connected component of } v \text{ in } G_p \text{ has at least } \varepsilon k/2 \text{ vertices}\right) > \varepsilon/6.$$

Proof. Let G be a graph with minimum degree at least k, let v be an arbitrary vertex of G, and fix some arbitrary linear order $\preccurlyeq$ on V whose smallest element is v. Assume that $p \geqslant \frac{1+\varepsilon}{k}$ and consider the p-random $\{0, 1\}$-sequence X of length $e(G)$. We will show that if we run the DFS algorithm on the random subgraph of G represented by X, then with probability at least $\varepsilon/6$, the following event holds:

$\mathcal{A}$:   From the moment v is moved to U, the set U does not become empty until $|S| \geqslant \varepsilon k/2$.

This clearly implies the assertion of the theorem, see property (1) of the DFS algorithm.

With the aim of estimating the probability of $\mathcal{A}$, let us define for each $i \in \{1, \dots, e(G)\}$,

$$Y_i = \begin{cases} (1 - \varepsilon/2) \cdot k & \text{if } X_i = 1, \\ -1 & \text{if } X_i = 0, \end{cases} \qquad \text{and} \qquad Z_i = (1 - \varepsilon/2) \cdot k + \sum_{j=1}^{i} Y_j.$$

We claim that if $Z_i > 0$ for all $i \leqslant e(G)$, then $\mathcal{A}$ holds. Indeed, suppose that the sequence X is such that $Z_i > 0$ for all i and consider the execution of the DFS algorithm on the random subgraph of G defined by X. Suppose that $\mathcal{A}$ does not hold, that is, the set U becomes empty when $|S| < \varepsilon k/2$. Recall that initially U contains one element (the vertex v) and the size of U increases by one every time $X_i = 1$ for some i. Moreover, at the time a vertex w is moved from U to S, the DFS algorithm has already queried all the $\deg(w, T)$ edges of G that connect w and T and got a negative answer for each of these queries. Finally, note that clearly $\deg_G(w, T) \geqslant k - |S| - |U| \geqslant (1 - \varepsilon/2) \cdot k$. In particular, the algorithm gets at least $(1 - \varepsilon/2)k$ negative answers to queries about edges of G incident to w before moving w from U to S. These three facts readily imply that if $\mathcal{A}$ does not hold, then $Z_i \leqslant 0$ for some i.

Finally, we estimate the probability that $Z_i > 0$ for all i. To this end, let α be the smallest positive real that satisfies

$$f(\alpha) := p \cdot \alpha^{(1-\varepsilon/2)\cdot k} + (1-p) \cdot \alpha^{-1} = 1. \tag{6}$$

We now argue that $\alpha \leqslant e^{-\frac{\varepsilon}{4k}}$. To this end, observe that $\lim_{\alpha \to 0^+} f(\alpha) = \infty$ and, recalling that $\varepsilon < 1/3$ and that $e^x \leqslant 1 + x + x^2$ if $|x| \leqslant 1$,

$$
\begin{aligned}
f\left(e^{-\frac{\varepsilon}{4k}}\right) &= p \cdot e^{-\frac{\varepsilon}{4}\cdot\left(1-\frac{\varepsilon}{2}\right)} + (1-p) \cdot e^{\frac{\varepsilon}{4k}} \\
&\leqslant p \cdot \left(1 - \frac{\varepsilon}{4} + \frac{3\varepsilon^2}{16}\right) + (1-p) \cdot \left(1 + \frac{\varepsilon}{4k} + \frac{\varepsilon^2}{16k^2}\right) \\
&= 1 - \frac{\varepsilon}{4} \cdot \left(p - \frac{1}{k} - \frac{3\varepsilon p}{4}\right) - \frac{\varepsilon}{4k}\left(p - \frac{\varepsilon(1-p)}{4k}\right) < 1.
\end{aligned}
$$

Now, let $Z_0 = (1 - \varepsilon/2)k$ and for each i with $0 \leqslant i \leqslant e(G)$, let $M_i = \alpha^{Z_i}$. Since α satisfies (6), the sequence (M_i) is a martingale. Furthermore, let

$$T = \min\{i \colon Z_i \leqslant 0\} \qquad \text{and} \qquad M_i^* = M_{\min\{i,T\}}$$

and observe that the sequence (M_i^*) is also a martingale, as T is a stopping time with respect to (M_i). Therefore,

$$\mathbb{E}[M_{e(G)}^*] = \mathbb{E}[M_0^*] = \mathbb{E}[M_0] = \mathbb{E}[\alpha^{Z_0}] = \alpha^{\left(1-\frac{\varepsilon}{2}\right)k} \leqslant e^{-\frac{\varepsilon}{4}\left(1-\frac{\varepsilon}{2}\right)}.$$

On the other hand, since $\alpha < 1$ and therefore $M_{e(G)}^* \geqslant 1$ whenever $Z_i \leqslant 0$ for some i, then

$$\mathbb{E}[M_{e(G)}^*] \geqslant \Pr(Z_i \leqslant 0 \text{ for some } i).$$

It follows that, recalling again that $e^x \leqslant 1 + x + x^2$ if $|x| \leqslant 1$,

$$\Pr(Z_i > 0 \text{ for all } i) \geqslant 1 - e^{-\frac{\varepsilon}{4}\left(1-\frac{\varepsilon}{2}\right)} \geqslant \frac{\varepsilon}{4}\left(1 - \frac{\varepsilon}{2}\right) - \frac{\varepsilon^2}{16} > \frac{\varepsilon}{6}. \qquad \square$$

Theorem 10. *For every positive ε, there exists a positive constant c such that the following holds. Let k be a sufficiently large integer and let G be an n-vertex graph with minimum degree at least k. If $p \geqslant \frac{1+\varepsilon}{k}$, then the expected number of excess edges in G_p is at least cn.*

Proof. Let us first choose some parameters. Let

$$\gamma = \frac{\varepsilon}{24} \qquad \text{and} \qquad \delta = \frac{\gamma^2}{4e^4}$$

and assume that $k \geqslant \frac{32}{\delta\varepsilon^2}$. Without loss of generality, we may also assume that $\varepsilon \leqslant \frac{2}{3}$. Let G be a graph with minimum degree at least k and suppose that $p \geqslant \frac{1+\varepsilon}{k}$. We will expose the edges of G_p in two rounds. To this end, let $p_1 = \frac{1+\varepsilon/2}{k}$, let p_2 satisfy $(1 - p_1)(1 - p_2) = 1 - p$, and observe that $p_2 \geqslant p - p_1 \geqslant \frac{\varepsilon}{2k}$. Consider the following two events in G_{p_1}:

$\mathcal{A}$:   at least γn vertices of G_{p_1} lie in components of size at least $\varepsilon k/4$.

$\mathcal{B}$:   some $B \subseteq V(G)$ with $|B| \geqslant \gamma n$ and $e(G[B]) \leqslant \delta|B|k$ satisfies $e(G_{p_1}[B]) \geqslant |B|/2$.

Suppose now that $\mathcal{A}$ holds. In this case, there is a set $A \subseteq V(G)$ of at least γn vertices such that each connected component of $G_{p_1}[A]$ has at least $\varepsilon k/4$ vertices and therefore,

$$e(G_{p_1}[A]) \geqslant \left(1 - \frac{4}{\varepsilon k}\right)|A| > \frac{|A|}{2}.$$

Assume furthermore that $\mathcal{B}$ does not hold. Then necessarily $e(G[A]) > \delta|A|k$. Observe that each edge of $G \setminus G_{p_1}$ belongs to G_{p_2} with probability p_2. Hence, considering separately the cases $e(G_{p_1}[A]) \geqslant \frac{\delta|A|k}{2}$ and $e(G_{p_1}[A]) < \frac{\delta|A|k}{2}$, we obtain the following bound:

$$
\begin{aligned}
\mathbb{E}\big[e(G_p[A]) \mid \mathcal{A} \wedge \neg\mathcal{B}\big] &\geqslant \mathbb{E}\left[e(G_{p_1}[A]) + p_2 \cdot \big(e(G[A]) - e(G_{p_1}[A])\big) \mid \mathcal{A} \wedge \neg\mathcal{B}\right] \\
&\geqslant \min\left\{\frac{\delta|A|k}{2}, \left(1 - \frac{4}{\varepsilon k}\right)|A| + p_2 \cdot \frac{\delta|A|k}{2}\right\} \\
&\geqslant \left(1 - \frac{4}{\varepsilon k} + \frac{\varepsilon\delta}{4}\right)|A| \geqslant \left(1 + \frac{\varepsilon\delta}{8}\right)|A| \geqslant |A| + \frac{\varepsilon\delta\gamma}{8}n.
\end{aligned}
\tag{7}
$$

Consequently, since $\mathrm{excess}(G_p) \geqslant \mathrm{excess}(G_p[A])$ by (4), the expected number of excess edges in G_p, conditioned on $\mathcal{A}$ and $\neg\mathcal{B}$ holding simultaneously, satisfies

$$\mathbb{E}\big[\mathrm{excess}(G_p) \mid \mathcal{A} \wedge \neg\mathcal{B}\big] \geqslant \mathbb{E}\big[\mathrm{excess}(G_p[A]) \mid \mathcal{A} \wedge \neg\mathcal{B}\big] \geqslant \frac{\varepsilon\delta\gamma}{8}n.$$

It is therefore enough to prove the following claim.

Claim. $\Pr(\mathcal{A} \wedge \neg\mathcal{B}) \geqslant \gamma/2$.

Let L be the number of vertices of G_{p_1} that lie in components of size at least $\varepsilon k/4$. It follows from Theorem 9 that $\mathbb{E}[L] \geqslant 2\gamma n$ and hence, $\Pr(\mathcal{A}) \geqslant \gamma$, as clearly $L \leqslant n$. With the aim of estimating the probability of $\mathcal{B}$, fix an arbitrary set B of b vertices, where $b \geqslant \gamma n$, such that $e(G[B]) \leqslant \delta bk$. It follows from Lemma 5 that

$$\Pr\big(e(G_{p_1}[B]) \geqslant b/2\big) \leqslant \Pr\big(\mathrm{Bin}(\delta bk, 2/k) \geqslant b/2\big) \leqslant (4e\delta)^{b/2}.$$

Consequently,

$$\Pr(\mathcal{B}) \leqslant \sum_{b \geqslant \gamma n} \binom{n}{b}(4e\delta)^{b/2} \leqslant \sum_{b \geqslant \gamma n} \left(\frac{en}{b}\right)^b (4e\delta)^{b/2} \leqslant \sum_{b \geqslant \gamma n} \left(\frac{4e^3\delta}{\gamma^2}\right)^{b/2} = \sum_{b \geqslant \gamma n} e^{-b/2} < \gamma/2,$$

since $\gamma n \geqslant \gamma k \geqslant 4/\gamma$. This completes the proof of the claim and consequently, the proof of the theorem. $\qquad\square$

Corollary 11. *For every positive ε, there exists a positive constant c such that the following holds. Let k be a sufficiently large integer and let G be an n-vertex graph with minimum degree at least k. If $p \geqslant \frac{1+\varepsilon}{k}$, then with probability at least $1 - 2\exp(-c^2 n/12)$, the number of excess edges in G_p is at least cn.*

Proof. Let $c = \min\{1, c_{10}(\varepsilon)/2\}$ and suppose that k is sufficiently large, so that the assertion of Theorem 10 is true. Without loss of generality, we may assume that $pe(G) \leqslant 4n$ as otherwise by Chernoff's inequality (Lemma 5), $e(G_p) \geqslant 2n$ with probability at least $1 - \exp(-n/4)$, and consequently

$$\operatorname{excess}(G_p) \geqslant e(G_p) - n \geqslant n \geqslant cn.$$

It follows from Theorem 10 that $\mathbb{E}[\operatorname{excess}(G_p)] \geqslant 2cn$ and therefore by Proposition 8 with $\beta = \frac{cn}{pe(G)}$,

$$\Pr(\operatorname{excess}(G_p) < cn) \leqslant 2 \exp\left(-\frac{cn}{3pe(G)} \cdot cn\right) \leqslant 2 \exp\left(-\frac{c^2 n}{12}\right),$$

as $pe(G) \leqslant 4n$ by our assumption. $\qquad\square$

5.3 Finding a long cycle

Let G be an $\mathcal{H}$-free graph with minimum degree at least k and assume that $p \geqslant \frac{1+\varepsilon}{k}$. Let $c = c_{11}/3$, let $\delta = c^2/21$, and suppose that k is sufficiently large, so that, in particular, the assertion of Corollary 11 holds. Without loss of generality, we may assume that $\varepsilon \leqslant 1$. Let ℓ be the largest integer satisfying

$$\operatorname{ex}(\ell, \mathcal{H}) \leqslant \frac{ck\ell}{20}. \tag{8}$$

By the maximality of ℓ, we have $\operatorname{ex}(\ell + 1, \mathcal{H}) > ck(\ell + 1)/20$ and hence $\ell \geqslant n_{\mathcal{H}}(ck/10) \geqslant n_{\mathcal{H}}(\delta k)$. Consider the p-random $\{0, 1\}$-sequence X of length $e(G)$. Recall the definition of $\lambda(\cdot, \cdot)$ from Section 3. We will show that if we run the DFS algorithm on the random subgraph of G represented by X, then with probability at least $1 - 4\exp(-c^2 k/16)$, the following event holds:

$\mathcal{A}$:   At the moment when the DFS algorithm finishes discovering all the connected components, i.e., when $S = V$, the number of edges $\{v, w\} \in G$ such that $\{v, w\}$ has not been yet queried and $\lambda(v, w) \geqslant \ell$ is at least $ck\ell/2$.

This will clearly be sufficient, since it implies that each of the last $ck\ell/2$ edges of G that are queried by our graph exploration algorithm closes a cycle of length at least ℓ and with probability at least $1 - (1 - p)^{ck\ell/2}$, one of these queries will be answered positively. Since, as noted above, $\ell \geqslant n_{\mathcal{H}}(ck/10) \geqslant ck/10$, this probability is at least $1 - \exp(-c^2 k/20)$.

Let n be the number of vertices of G. We now argue that the following statement implies the claimed bound on the probability that G_p contains a cycle of length at least ℓ.

Claim. The event $\mathcal{A}$ contains the intersection of the following two events:

(i) The number of excess edges in G_p is at least $3cn$.

(ii) There are fewer than $3cn$ indices i with $i > e(G) - ckn$ such that $X_i = 1$.

Indeed, note that $n \geqslant k + 1$, that, by Corollary 11, (i) holds with probability at least $1 - 2\exp(-c^2 n/2)$, and that, by Chernoff's inequality (Lemma 5), (ii) holds with probability at least $1 - 2\exp(-cn/16)$.

It now suffices to prove the claim. Let $G' \subseteq G$ consist of pairs (edges of G) that have not been queried at the time when our graph exploration algorithm finishes discovering the connected components of G_p. Note that (i) and (ii) imply that $e(G') \geqslant ckn$ as the rooted spanning forest F of G_p constructed by the DFS algorithm contains no excess edges. Recall from Section 3 that each edge of G' connects a vertex w with its predecessor v in one of the rooted trees forming F and that $\lambda(v, w)$ denotes the length of the unique path joining v and w in this tree. Since the average degree of G' is at least $2ck$ and the endpoints of each edge of G' lie in the same connected component of F, there must be a rooted tree R in F such that the average degree of $G'[V(R)]$ is also at least $2ck$. We will now show that there are at least $ck\ell/2$ edges $\{v, w\} \in G'[V(R)]$ such that $\lambda(v, w) \geqslant \ell$.

To this end, for every subtree R^* of R, let

$$d(R^*) = \sum_{w \in V(R^*)} |\{v \in V(R): v \text{ is a predecessor of } w \text{ in } R \text{ and } \{v, w\} \in G'\}|$$

and note that $d(R) = e(G'[V(R)]) \geqslant ck|V(R)|$. We claim that R contains a subtree R^* with $\ell \leqslant |V(R^*)| < 2\ell$ with $d(R^*) \geqslant ck\ell$. Indeed, we can construct such R^* as follows. First, repeatedly delete from R every full subtree[2] R' such that $d(R') < ck|V(R')|$ until there are no such R' left. Since as a result of every such deletion, the ratio $d(R)/|V(R)|$ increases, we may assume that R contains no such R'. We claim that R still has at least ℓ vertices. Indeed, $G'[V(R)]$ is an $\mathcal{H}$-free graph with average degree at least $2ck$ and for every $m \leqslant \ell$, by (8) and Lemma 7,

$$\frac{ex(m, \mathcal{H})}{m} \leqslant 2 \cdot \frac{\mathrm{ex}(\ell, \mathcal{H})}{\ell} \leqslant \frac{ck}{10}.$$

We can obtain an R^* with the claimed property by running the following simple recursive procedure on R: If R has fewer than 2ℓ vertices, then we let $R^* = R$. Otherwise, if the full subtree R' rooted at one of the children of the root of R has at least ℓ vertices, then we let $R = R'$ and work with R', noting that $d(R') \geqslant ck|V(R')|$ by our assumption on the original tree. Else, R has at least 2ℓ vertices but each full subtree of R attached to its root has fewer than ℓ vertices. In this case, we can easily obtain a subtree R^* of R that has the required properties by deleting some of these full subtrees.

Finally, let R^+ be the subtree of R obtained from R^* by adding to it the first (at most) ℓ vertices on the unique path from the root of R^* to the root of R. By definition, if $\{v, w\}$ is an edge of G' such that $\lambda(v, w) < \ell$ and v is a predecessor of w, then $w \in V(R^*)$ implies that $v \in V(R^+)$. It follows that

$$|\{\{v, w\} \in G': \lambda(v, w) \geqslant \ell\}| \geqslant d(R^*) - e(G'[V(R^+)]) \geqslant ck\ell - \mathrm{ex}(3\ell, \mathcal{H}) \geqslant ck\ell/2.$$

---

[2]A subtree of a rooted tree that is induced by some vertex and all of its descendants.

To see the last inequality, note that by (8) and Lemma 6,

$$\mathrm{ex}(3\ell, \mathcal{H}) \leqslant 10 \cdot \mathrm{ex}(\ell, \mathcal{H}) \leqslant ck\ell/2.$$

This completes the proof of the claim and consequently, the proof of the theorem. $\qquad\square$

6 Concluding remarks

6.1 Dependence of ℓ on $\mathcal{H}$ and ε and the function $n_{\mathcal{H}}(\cdot)$

In our proof of Theorem 1, we made no attempts to optimize the dependence of δ on ε and $\mathcal{H}$. A careful analysis of the proof shows that δ has only polynomial dependence on ε, that is, Theorem 1 asserts the existence of a cycle of length $n_{\mathcal{H}}(c_{\mathcal{H}}\varepsilon^{\alpha_{\mathcal{H}}})$, where $c_{\mathcal{H}}$ and $\alpha_{\mathcal{H}}$ are positive constants depending only on $\mathcal{H}$. It would be interesting to know whether in the setting of Corollary 2, i.e., when $\mathcal{H}$ is the empty family, one can replace $\alpha_{\mathcal{H}}$ with 2, as in Theorem 4, where condition (2) is satisfied for $\ell = \varepsilon^2 k/36$.

A much more intriguing question is that about the relation between $n_{\mathcal{H}}(k)$ and $n_{\mathcal{H}}(\delta k)$. It is not difficult to prove that for every good family $\mathcal{H}$ and positive constant δ, we have $n_{\mathcal{H}}(\delta k) \leqslant (\delta + o(1))n_{\mathcal{H}}(k)$. Indeed, one may derive this estimate from Lemma 6 with $m = n_{\mathcal{H}}(\delta k) - 1$ and $n = n_{\mathcal{H}}(k)$. On the other hand, for most bipartite graphs H, since the known lower and upper bounds on $\mathrm{ex}(n, H)$ do not match, it is not even clear whether the ratio $n_{\{H\}}(k)/n_{\{H\}}(\delta k)$ is bounded for any constant $\delta < 1$.

6.2 Regular graphs

The proof of Theorem 1 can be somewhat simplified (at least conceptually) when one makes the assumption that the graph G is regular. Recall that the key step in our proof is establishing that with high probability, the graph G_p contains $\Omega(n)$ excess edges. If G is regular, then using the approach of [9], one can obtain a lower bound on the expected number of small tree components in G_p that is sufficiently strong to imply, via (3), that $\mathbb{E}[\mathrm{excess}(G_p)] \geqslant \varepsilon^3 n/12$. Proposition 8 will then imply the statement of Corollary 11 with c replaced by $\varepsilon^2/20$, which is sufficient for the proof of Theorem 1.

Moreover, following the approach of [9], one may also show that for some absolute constant C, with high probability only at most $C\varepsilon n$ vertices of G_p lie in connected components that are not small trees. Consequently, in the proof of Theorem 1, we may find a rooted tree R such that the average degree of $G'[V(R)]$ is at least $\frac{\varepsilon^2}{10C} \cdot k$. This is enough to show that with high probability, G_p contains a cycle of length at least $\frac{\varepsilon^2}{20C} \cdot k$.

6.3 Graphs with large average degree

It is natural to ask whether Theorem 1 is still true when one replaces the assumption that the minimum degree of G is at least k with the much weaker assumption that the *average* degree of G is at least k. Since each graph G with average degree k contains a

subgraph with minimum degree exceeding $k/2$, one can easily deduce such statement if one strengthens the assumption on p to $p \geqslant \frac{2+\varepsilon}{k}$. The following easy argument shows that this is not optimal.

Assume that G has n vertices and average degree k. For every $p \in [0,1]$, since the function $(0,\infty) \ni x \mapsto (1-p)^x$ is convex, the expected number of isolated vertices in G_p is at least $(1-p)^k \cdot n$. Let c_0 be the positive solution of the equation $\frac{c}{2} - 1 + e^{-c} = 0$ and observe that $c_0 \approx 1.6$. Using (3), we see that if $c > c_0$ and k is sufficiently large, then $\mathbb{E}[\mathrm{excess}(G_p)] = \Omega(n)$. It follows that the assumption on p can be weakened to $p \geqslant \frac{c_0+\varepsilon}{k}$. We believe that similarly as in the minimum degree case, the assumption $p \geqslant \frac{1+\varepsilon}{k}$ is sufficient, but at the moment we are unable to establish this claim.

Acknowledgments

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi. The longest path in a random graph. *Combinatorica*, 1:1–12, 1981.

[2] N. Alon and J. H. Spencer. *The probabilistic method.* Wiley-Interscience Series in Discrete Math. and Optimization. John Wiley & Sons Inc., Hoboken, NJ, third edition, 2008. With an appendix on the life and work of Paul Erdős.

[3] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, second edition, 2001.

[4] J. A. Bondy and M. Simonovits. Cycles of even length in graphs. *J. Combinatorial Theory Ser. B*, 16:97–105, 1974.

[5] M. N. Ellingham and D. K. Menser. Girth, minimum degree, and circumference. *J. Graph Theory*, 34:221–233, 2000.

[6] P. Erdős. Graph theory and probability. *Canad. J. Math.*, 11:34–38, 1959.

[7] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[8] A. Frieze and M. Krivelevich. On the non-planarity of a random subgraph. *Combin. Probab. Comput.*, 22:722–732, 2013.

[9] A. Frieze, M. Krivelevich, and R. Martin. The emergence of a giant component in random subgraphs of pseudo-random graphs. *Random Structures Algorithms*, 24:42–50, 2004.

[10] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30:1141–1144, 1959.

[11] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[12] M. Krivelevich, C. Lee, and B. Sudakov. Long paths and cycles in random subgraphs of graphs with large minimum degree. Random Structures Algorithms, to appear.

[13] M. Krivelevich and B. Sudakov. The phase transition in random graphs: a simple proof. *Random Structures Algorithms*, 43:131–138, 2013.

[14] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin, 1998.

[15] O. Ore. On a graph theorem by Dirac. *J. Combinatorial Theory*, 2:383–392, 1967.

[16] O. Riordan. Long cycles in random subgraphs of graphs with large minimum degree. arXiv:1308.3144.

[17] B. Sudakov and J. Verstraëte. Cycle lengths in sparse graphs. *Combinatorica*, 28:357–372, 2008.