

Paths and cycles in random subgraphs of graphs with large minimum degree

Stefan Ehard

Institut für Optimierung und Operations Research
Universität Ulm
Ulm, Germany

stefan.ehard@uni-ulm.de

Felix Joos

School of Mathematics
University of Birmingham
Birmingham, United Kingdom

f.joos@bham.ac.uk

Submitted: Jan 20, 2017; Accepted: May 9, 2018; Published: May 25, 2018

© The authors. Released under the CC BY-ND license (International 4.0).

Abstract

For a graph G and $p \in [0, 1]$, let G_p arise from G by deleting every edge mutually independently with probability $1 - p$. The random graph model $(K_n)_p$ is certainly the most investigated random graph model and also known as the $G(n, p)$ -model. We show that several results concerning the length of the longest path/cycle naturally translate to G_p if G is an arbitrary graph of minimum degree at least $n - 1$.

For a constant $c > 0$ and $p = \frac{c}{n}$, we show that asymptotically almost surely the length of the longest path in G_p is at least $(1 - (1 + \epsilon(c))ce^{-c})n$ for some function $\epsilon(c) \rightarrow 0$ as $c \rightarrow \infty$, and the length of the longest cycle is at least $(1 - O(c^{-\frac{1}{5}}))n$. The first result is asymptotically best-possible. This extends several known results on the length of the longest path/cycle of a random graph in the $G(n, p)$ -model to the random graph model G_p where G is a graph of minimum degree at least $n - 1$.

Mathematics Subject Classifications: 05C38, 05C80

1 Introduction

Around 1960 Erdős and Renyi proved the first results about random graphs – especially about graphs on n vertices where every possible edge is present independently with probability p , which is nowadays known as the $G(n, p)$ -model. It is not an overstatement saying that this field has grown enormously since then and for numerous graph parameters the typical value is (precisely) known for large n . In particular, the lengths of paths and cycles are investigated. As for any $\epsilon > 0$ and $p \geq \frac{(1+\epsilon)\log n}{n}$ a.a.s.¹ a graph in $G(n, p)$ is

¹For a sequence of graphs $(G_i)_{i \in \mathbb{N}}$ and numbers $(p_i)_{i \in \mathbb{N}}$ in $[0, 1]$, we say that the graph sequence has property \mathcal{P} asymptotically almost surely (a.a.s.) if $\mathbb{P}[(G_i)_{p_i} \in \mathcal{P}] \rightarrow 1$ as $i \rightarrow \infty$.

hamiltonian, we consider the length of a longest path/cycle if $p = \frac{c}{n}$ for some constant $c > 1$. A series of papers [1, 3, 4, 5, 6] finally led to the following theorem, where

$$\alpha(c) = \sup_{\alpha \geq 0} \left\{ G \in G(n, cn^{-1}) \text{ contains a path of length at least } \alpha n \text{ a.a.s.} \right\}$$

and $\beta(c)$ analogously for the length of cycles.

Theorem 1. *There exists a function $\epsilon(c) \rightarrow 0$ as $c \rightarrow \infty$ such that*

$$\alpha(c), \beta(c) = 1 - (1 + \epsilon(c))ce^{-c}.$$

Let us consider a more general random graph model. For a graph G , we denote by G_p the random subgraph obtained by deleting every edge independently with probability $1 - p$ from the edge set of G . Thus $(K_n)_p$ has the same distribution as $G(n, p)$. In this paper we consider the typical asymptotic behavior of $(G_k)_p$ instead of $(K_n)_p$ where G_k is a simple graph of minimum degree at least k . In our setting p depends on k instead of the order of G_k . To be precise, we consider $p = \frac{c}{k}$ where $c > 0$ and k is large enough in terms of c . We denote by \mathcal{G} the set of all graph sequences G_1, G_2, \dots such that G_k has minimum degree at least k . We define

$$\alpha'(c) = \inf_{(G_k)_{k \geq 1} \in \mathcal{G}} \sup_{\alpha \geq 0} \left\{ (G_k)_{\frac{c}{k}} \text{ contains a path of length at least } \alpha k \text{ a.a.s.} \right\}$$

and $\beta'(c)$ analogously for cycles. It is clear that $\alpha'(c) \leq \alpha(c)$ and $\beta'(c) \leq \beta(c)$. We prove that there is essentially no difference between $\alpha'(c)$ and $\alpha(c)$ and our second contribution is a lower bound on $\beta'(c)$.

Theorem 2. *There exists a function $\epsilon(c) \rightarrow 0$ as $c \rightarrow \infty$ such that*

$$\alpha'(c) = 1 - (1 + \epsilon(c))ce^{-c}.$$

Theorem 3. *We have $\beta'(c) = 1 - O(c^{-\frac{1}{5}})$ as $c \rightarrow \infty$.*

Thus Theorem 2 describes precisely the asymptotic behavior of $\alpha'(c)$ as $c \rightarrow \infty$ improving a result due to Krivelevich, Lee and Sudakov [8] who showed that $\alpha'(c) = 1 - O(c^{-\frac{1}{2}})$. In addition, it generalizes the statement about paths in Theorem 1.

Theorem 3 improves a result of Krivelevich, Lee and Sudakov [8] and Riordan [10] implying $\beta'(c) = 1 - o(1)$. It also generalizes early results of the length of the longest cycle in the $G(n, p)$ -model [1, 5].

Note that the questions of hamiltonicity in the $G(n, p)$ setting naturally translates to the question whether G_k has a cycle of length at least $k + 1$. These extensions are successfully settled by Krivelevich, Lee and Sudakov [8], and by Glebov, Naves and Sudakov [7].

We do not believe that the bound in Theorem 3 is tight and conjecture that $\alpha'(c) = \beta'(c)$ holds. However, due to the different setting as in the $G(n, p)$ -model, the construction of cycles is much more complicated as a graph of minimum degree at least k can clearly have arbitrarily large girth. Thus any cycle in G_p might be much larger than k and is

therefore difficult to detect. This may be one reason why the approach of Glebov, Naves and Sudakov in [7] needs so much more effort and additional tools in comparison to a proof that shows the same threshold behavior in the $G(n, p)$ -model. We think that a proof that describes the behavior of $\beta'(c)$ may be similarly complicated. In comparison to the approach in [7], the proof of Theorem 3 is rather short.

2 Preliminaries

We use standard notation. In particular, we denote the vertex set of a graph G by $V(G)$. For a path P and two vertices $u, v \in V(P)$, we denote by uPv the subpath of P with endvertices u, v .

We will frequently need to show that a binomial random variable is very close to its expected value and use for these purposes Chernoff's inequality.

Theorem 4 (Chernoff's inequality [2]). *If X is a binomial distributed random variable with $X \sim \text{Bin}(n, p)$ and $0 < \lambda \leq np = \mathbb{E}X$, then*

$$\mathbb{P}[|X - np| \geq \lambda] \leq 2e^{-\frac{\lambda^2}{3np}}.$$

Several results in this paper are based on the depth-first-search algorithm (DFS-algorithm) which is a frequently used exploration method of graphs. We briefly describe this algorithm and introduce some notation along the way. To the best of our knowledge, Krivelevich and Sudakov were the first who used this method in connection with random graphs [9]. This paper and [10] present very nice and short proofs of previously known and new results.

The DFS-algorithm is an algorithm traversing a graph such that all vertices of a given graph G are finally visited and outputs a rooted spanning forest T of G . It proceeds in the following way.

At any step, there is a partition of the vertex set $V(G)$ into three sets R , S and U . The set U contains the vertices that have not yet been visited during the exploration, R denotes the set of vertices whose exploration is complete, and all the remaining vertices that are currently under exploration are contained in S . The vertices of S are kept in a stack, which is a last-in-first-out data structure.

The algorithm starts with $U = V(G)$ and $R = S = \emptyset$ and executes the following rounds (one round is the execution of (a) or (b)) until every vertex is explored, i.e. $R = V(G)$ and $S = U = \emptyset$.

- (a) If $S = \emptyset$, then some unreached vertex v in U is moved to S . This vertex v will be the root of a new component of our rooted spanning forest T .
- (b) Otherwise, let v be the top element of the stack S (the *last-in* vertex). The algorithm queries whether v has some neighbor w in U . If so, w is placed on top of the stack S . If v has no neighbor in U , it is completely explored and is moved to R .

In each round of the algorithm there is exactly one vertex moved either from U to S or from S to R . So indeed, after $2|V(G)|$ rounds every vertex has been moved from U to R through S and the algorithm terminates with a rooted spanning forest T .

The following properties of the DFS-algorithm are important to us:

- (I) Every positively answered query about a neighbor in U increases the size of $R \cup S$ by exactly one.
- (II) The set S always spans a path in T .
- (III) At any round of the algorithm, all possible edges between the set R and U have been queried and answered negatively.
- (IV) Every edge $e = uv$ of the graph G which is not tested during the exploration of G joins two vertices on some vertical path in the rooted spanning forest T (because otherwise the algorithm would have queried the edge uv during the exploration).

We will use the DFS-algorithm to explore the random graph G_p . Therefore, we assume that the algorithm already knows the underlying graph G and all the edges of G . The DFS-algorithm only queries about these edges of G during the exploration of G_p . That is, if the DFS-algorithm looks for neighbors of some vertex v , it only considers the neighbors w of v in G , and queries whether this vertex is also a neighbor of v in G_p . We receive a positive answer of each such query independently with probability p . In this way, following this algorithm, we explore a rooted spanning forest of our random graph G_p . Note that by definition the answer of a query does not depend on the answers of the previous queries. We say an edge of G is *tested* if the DFS-algorithm queried whether this edge is in G_p and otherwise we say it is *untested*.

Throughout the paper we consider graphs G_k of minimum degree at least k . Several inequalities in our computations are only correct if k is large enough. For the purpose of readability we often drop the index k and simply write G . Additionally, we leave out necessary roundings where it does not affect the argument.

3 Long Cycles

In this section we prove Theorem 3. Let G be a graph of minimum degree at least k on n vertices and let $p = \frac{c}{k}$ for c sufficiently large and k sufficiently large in terms of c .

We consider a rooted forest T which is an output of the DFS-algorithm described in Section 2. We emphasize that every untested edge of G is present in G_p independently of T .

This proof is based on ideas of Riordan [10] and follows its strategy. In particular, the first two short lemmas naturally transfer to our setting. Note that he considers $p = \omega(k^{-1})$; that is, in our case p is significantly smaller. As a consequence, in our case many smaller error quantities do not vanish as $k \rightarrow \infty$. Therefore, we carefully introduced a suitable hierarchy of these small quantities. Beside that a few new ideas are needed to deal with a smaller edge probability p .

The strategy of the proof is as follows. At first we show that for most of the vertices in G almost all incident edges are not queried by the DFS-algorithm (Lemma 5 and 6). By (IV), all untested edges connect two vertices in the same component of T . Thus we may assume that there are not too many untested edges joining two vertices in distance at least $(1 - 5c^{-1/5})k$ in T . Using this we show that T contains a long root-to-leaf path P with certain properties (Lemma 7 and 8). Finally, we use P and all the untested edges that join two vertices on P to find a long cycle in G_p .

Lemma 5. *During the DFS-algorithm on G_p a.a.s. at most $\frac{2n}{p} = \frac{2nk}{c}$ many edges are tested.*

Proof. We run the DFS-algorithm on G_p . Note that the rooted spanning forest T of G_p has at most $n - 1$ edges and that every positively answered query contributes an edge to our exploration of this forest. Let X be the number of tested edges. For a positive integer ℓ , let A_ℓ be the event that out of ℓ queries at most $n - 1$ are answered positively. By comparison with a binomial distribution and using Chernoff's inequality, we obtain

$$\mathbb{P}[X > 2n/p] \leq \mathbb{P}[A_{2n/p}] \leq 2e^{-\frac{n}{3}} = o(1),$$

which completes the proof. \square

From now on, let $\epsilon = c^{-1/5}$. In view of the statement, we assume from now always that $\epsilon < 10^{-4}$. Let E_u be the set of untested edges of G during the DFS-algorithm. We call a vertex *free* if it is incident with at least $(1 - \epsilon)k$ untested edges in E_u .

Lemma 6. *A.a.s., at most $4\epsilon^4 n$ vertices of the rooted forest T produced by the DFS-algorithm are not free.*

Proof. Let $v \in V(T)$ be a vertex that is not free. Since the minimum degree of G is at least k , the vertex v is incident with at least ϵk tested edges. Assume that there are more than $4\epsilon^4 n$ vertices that are not free. Hence, we have more than $\frac{1}{2}4\epsilon^4 n \cdot \epsilon k = \frac{2nk}{c}$ many tested edges in total. By Lemma 5, the probability of this is $o(1)$, which implies the statement. \square

Following the proof of Riordan [10], for a rooted forest T and a vertex $v \in V(T)$, we introduce the following notation.

- Let $A(v)$ be the set of ancestors of v in T excluding v and let $D(v)$ be the set of descendants of v in T excluding v .
- Let $A_i(v)$ and $D_i(v)$ be the sets of ancestors and descendants of v at distance exactly i , respectively, and let $A_{\leq i}(v)$ and $D_{\leq i}(v)$ be the sets of ancestors and descendants of v at distance at most i .
- The height of the vertex v is defined as $\max\{i : D_i(v) \neq \emptyset\}$.

- For two vertices u, v , let $d(u, v)$ be the number of edges on a shortest u, v -path in T if such a path exists, otherwise set $d(u, v) = \infty$.
- We say the vertex v is *up* if $|D(v)| \geq \epsilon k$. If this is not the case, then v is *down*.
- We call the vertex v *skinny* if $|D_{\leq (1-5\epsilon)k}(v)| \leq (1-4\epsilon)k$. Let Y denote the set of vertices in T which are not skinny.

Lemma 7. *If a rooted forest T contains at most $5\epsilon^4 n$ down vertices, then, for any constant $h \geq 1$, at most $6h\epsilon^3 n$ vertices of T are at height less than hk .*

Proof. We define the set

$$\mathcal{S}_1 = \{(v, w) : v \text{ is up, } v \text{ is at height at most } hk, \text{ and } w \in D(v) \text{ is down}\}.$$

As $A(w)$ forms a path, every vertex w which is down is contained in at most hk pairs $(v, w) \in \mathcal{S}_1$. Since there are at most $5\epsilon^4 n$ down vertices, we conclude $|\mathcal{S}_1| \leq 5hk\epsilon^4 n$. As a vertex v which is up and at height at most hk is contained in at least ϵk pairs $(v, w) \in \mathcal{S}_1$, there are at most $5h\epsilon^3 n$ such vertices. That is, at most $5h\epsilon^3 n + 5\epsilon^4 n \leq 6h\epsilon^3 n$ vertices in T are at height less than hk . \square

Lemma 8. *If a rooted forest T contains at most $5\epsilon^4 n$ down vertices and $X \subseteq V(T)$ such that $|X| \leq 5\epsilon^4 n$, then, T contains a vertical path P of length at least $4k$ containing at most $\frac{1}{4}\epsilon k$ vertices in $X \cup Y$.*

Proof. Let X be a subset of $V(T)$ of size at most $5\epsilon^4 n$. First we show that the set $Y \subseteq V(T)$ is small enough for our purposes. (Recall that Y denotes the set of vertices which are not skinny.) We define the set

$$\mathcal{S}_2 = \{(v, w) : v \in A(w), d(v, w) \leq (1-5\epsilon)k\}.$$

Since a vertex has at most one ancestor at any given distance, we conclude

$$|\mathcal{S}_2| \leq (1-5\epsilon)kn.$$

By Lemma 7, all but at most $6\epsilon^3 n$ vertices v are at height at least k and thus, each such v appears in at least $(1-5\epsilon)k$ pairs $(v, w) \in \mathcal{S}_2$. This contributes at least

$$(1-5\epsilon)(1-6\epsilon^3)kn \geq (1-5\epsilon)kn - 6\epsilon^3 kn \tag{1}$$

pairs to the set \mathcal{S}_2 . Any vertex which is not skinny contributes another ϵk pairs to the lower bound (1). As $|\mathcal{S}_2| \leq (1-5\epsilon)kn$, we conclude $|Y| \leq \frac{6\epsilon^3 kn}{\epsilon k} = 6\epsilon^2 n$.

Next we want to find the desired path P . We define the set

$$\mathcal{S}_3 = \{(v, w) : w \in X \cup Y, v \in A(w), d(v, w) \leq 4k\}.$$

Since a vertex has at most one ancestor at each distance, for a pair $(v, w) \in \mathcal{S}_3$, the vertex w can appear in at most $4k$ different pairs in \mathcal{S}_3 . We obtain

$$\begin{aligned} |\mathcal{S}_3| &\leq 4k \cdot |X \cup Y| \\ &\leq 4k \cdot (5\epsilon^4 n + 6\epsilon^2 n) \\ &\leq 25\epsilon^2 kn. \end{aligned}$$

This implies that the number of vertices v that can appear in more than $\frac{1}{4}\epsilon k$ pairs $(v, w) \in \mathcal{S}_3$, is bounded from above by

$$\frac{25\epsilon^2 kn}{\frac{1}{4}\epsilon k} = 100\epsilon n.$$

By Lemma 7, all but at most $24\epsilon^3 n$ vertices of T are at height at least $4k$ and from above follows that all but at most $100\epsilon n$ vertices v appear in at most $\frac{1}{4}\epsilon k$ pairs $(v, w) \in \mathcal{S}_3$. Hence, for c sufficiently large such that ϵ is small enough, there exists a vertex v at height at least $4k$ that appears in at most $\frac{1}{4}\epsilon k$ pairs $(v, w) \in \mathcal{S}_3$. Let P be the vertical path from v to some vertex in $D_{4k}(v)$. Then P has length $4k$ and by the choice of v , the path P contains at most $\frac{1}{4}\epsilon k$ vertices in $X \cup Y$. \square

Proof of Theorem 3. Recall, G is a graph of minimum degree at least k and $p = \frac{c}{k}$ for c sufficiently large.

We run the DFS-algorithm on G_p . Let T be the spanning forest and let E_u be the set of untested edges of G that we obtain from this algorithm. By Lemma 6, we may assume that all but at most $4\epsilon^4 n$ vertices of T are free, that is, incident with at least $(1 - \epsilon)k$ untested edges. Due to property (IV) of the DFS-algorithm, for every untested edge $uv \in E_u$, either $u \in A(v)$ or $u \in D(v)$.

Suppose we have

$$\left| \{u : uv \in E_u, d(u, v) \geq (1 - 5\epsilon)k\} \right| \geq \epsilon k \quad (2)$$

for more than $\log k$ vertices v . This means, we can find at least $\epsilon k \log k$ untested edges $uv \in E_u$ in G with $d(u, v) \geq (1 - 5\epsilon)k$. With probability

$$1 - (1 - p)^{\epsilon k \log k} \leq 1 - e^{-c^{4/5} \log k} = 1 - o(1)$$

at least one of these edges present in G_p . Such an edge uv forms together with the u, v -path in T the desired cycle of length at least $(1 - 5\epsilon)k$ in G_p .

Hence we may assume that for all vertices v except for at most $\log k$, we have

$$\left| \{u : uv \in E_u, d(u, v) \geq (1 - 5\epsilon)k\} \right| < \epsilon k. \quad (3)$$

Let V_0 be the set of vertices v that do not satisfy (3); that is, $|V_0| \leq \log k$.

Claim. *A.a.s., there are at most $5\epsilon^4 n$ down vertices.*

Proof of the claim: Assume that some vertex $v \in V(T) \setminus V_0$ is free and down. Since $|D(v)| < \epsilon k$ and v is free, there are at least $(1 - \epsilon)k - \epsilon k = (1 - 2\epsilon)k$ untested edges $uv \in E_u$ with $u \in A(v)$. Since each vertex has at most one ancestor at each distance, v has at least $(1 - 2\epsilon)k - (1 - 5\epsilon)k = 3\epsilon k$ ancestors u with $uv \in E_u$ and $d(u, v) \geq (1 - 5\epsilon)k$, which is a contradiction as $v \notin V_0$. Therefore, no down vertex in $V(T) \setminus V_0$ is free. By Lemma 6, a.a.s. all but $4\epsilon^4 n$ vertices are free. Hence, at most

$$4\epsilon^4 n + |V_0| \leq 4\epsilon^4 n + \log k \leq 5\epsilon^4 n$$

vertices are down, where the last inequality holds for sufficiently large k since $n \geq k$. ■

Thus we may apply Lemma 8, where X is the union of V_0 and the set of vertices that are not free, that is, $|X| \leq 4\epsilon^4 n + \log k \leq 5\epsilon^4 n$ (Lemma 6). Recall that Y is the set of vertices that are not skinny. Let P be the path of length $4k$ that is given by the Lemma 8 and let Z be the set of vertices of $V(P) \setminus V_0$ that are free and skinny. By Lemma 8, we obtain

$$|V(P) \setminus Z| = |(X \cup Y) \cap V(P)| \leq \frac{1}{4}\epsilon k.$$

For any vertex $v \in Z$, there are at least $(1 - \epsilon)k$ untested edges $uv \in E_u$ with $u \in A(v) \cup D(v)$. We want to show that there are sufficiently many of these vertices u in $A(v)$.

Since $v \in Z$ implies $v \notin V_0$, at least $(1 - 2\epsilon)k$ of these vertices u with $uv \in E_u$ satisfy $d(u, v) \leq (1 - 5\epsilon)k$. Moreover, as v is skinny, at least $(1 - 2\epsilon)k - (1 - 4\epsilon)k = 2\epsilon k$ vertices u must be ancestors of v with $d(u, v) \leq (1 - 5\epsilon)k$. We define a set of ancestors of v within a certain distance, namely

$$B(v) = \{u \in A(v) : uv \in E_u, \epsilon k \leq d(u, v) \leq (1 - 5\epsilon)k\}.$$

Again, since v has only one ancestor at each distance, we obtain $|B(v)| \geq \epsilon k$.

Let $u_1 \in V(P)$ be the vertex on the path P , which is at height k in P . Let V_1 be the set of the first descendants of u_1 on P , such that $|V_1 \cap Z| \geq \log k$. Since $|V(P) \setminus Z| \leq \frac{1}{4}\epsilon k$, we have

$$V_1 \subseteq D_{\leq \frac{1}{4}\epsilon k + \log k}(u_1) \cap V(P).$$

For each of these vertices $v \in V_1 \cap Z$, we have $|B(v)| \geq \epsilon k$. Hence, there are at least $\epsilon k \log k$ untested edges $uv \in E_u$ such that $v \in V_1 \cap Z$ and $u \in B(v)$. With probability at least $1 - (1 - p)^{\epsilon k \log k} = 1 - o(k^{-1})$ one of these edges is present. Let $v_1 u_2$ be one of it such that $v_1 \in V_1$, $u_2 \in B(v_1)$.

Assume for some $i \geq 2$ we have defined vertices $v_1, u_1, v_2, u_2, \dots, u_{i-1}, u_i$ that appear in this bottom-to-top order on P such that

- $v_j u_{j+1} \in E(G_p)$,
- $u_{j+1} \in B(v_j)$,

- $v_j \in D_{\leq \frac{1}{4}\epsilon k + \log k}(u_j)$ and $|V(u_j P v_j) \cap Z| \leq 1 + \log k$ for all $1 \leq j \leq i-1$, and
- $d(u_1, u_i) \leq 2k$ where $d(u, v)$ denotes the distance between two vertices u, v in T .

Next we show how to define v_i and u_{i+1} . Let V_i be the set of the first descendants of u_i on P such that $|V_i \cap Z| \geq \log k$. Thus for every vertex $w \in V_i$, we have $d(w, u_{i-1}) \geq \epsilon k - \frac{1}{4}\epsilon k - 2\log k > \frac{\epsilon}{2}k$. Again, as $|V_i \cap Z| \geq \log k$, there is an edge $v_i u_{i+1}$ present in G_p with $v_i \in V_i$ and $u_{i+1} \in B(v_i) \subseteq V(P)$ with probability $1 - o(k^{-1})$.

We may continue this iterative procedure to find such edges $v_j u_{j+1}$ until we reach a vertex u_{i+1} with $d(u_1, u_{i+1}) > 2k$. Since $d(u_j, u_{j+1}) > \frac{1}{2}\epsilon k$, this implies $i \leq 4\epsilon^{-1}$. Thus the procedure does not fail with probability $1 - o(\epsilon^{-1}k^{-1}) = 1 - o(1)$.

Suppose i is even. Consider the following cycle C obtained by the concatenation of the following paths:

$$\begin{aligned} &v_1 u_2, u_2 P v_3, v_3 u_4, u_4 P v_5, v_5 u_6, u_6 P v_7 \dots \\ &v_{i-1} u_i, u_i P u_{i+1}, u_{i+1} v_i, v_i P u_{i-1}, u_{i-1} v_{i-2}, v_{i-2} P u_{i-3}, \dots \\ &u_3 v_2, v_2 P v_1. \end{aligned}$$

Note that the number of vertices of $V_1 \cup \dots \cup V_i$ that are not contained in C is at most $|V(P) \setminus Z| + i \log k \leq \frac{1}{4}\epsilon k + 4\epsilon^{-1}k$. Since $d(u_1, u_{i+1}) > 2k$, the length of C is at least

$$2k - \frac{1}{4}\epsilon k - 4\epsilon^{-1} \log k > k.$$

A similar argument applies if i is odd. □

4 Long Paths and the DFS-algorithm

Before we prove Theorem 2, in we cite and prove some results for later use. The first one uses a nice and direct analysis of the DFS-algorithm.

Lemma 9 (Krivelevich, Lee, Sudakov [8]). *Let $p = \frac{c}{k}$ for c sufficiently large, and let G be a graph of minimum degree at least k . If G is bipartite, then G_p a.a.s. contains a path of length $(2 - 6c^{-1/2})k$.*

The next lemma is of a similar flavor as the last one. We suitably modify a result of [8] for our purposes.

Lemma 10. *Let $p = \frac{c}{k}$ for c sufficiently large, and let G be a graph of minimum degree at least k . If $V_0 \subseteq V(G)$ with $|V_0| \geq \log k$, then G_p a.a.s. contains a path of length $(1 - 2c^{-1/2})k$ which starts at a vertex in V_0 .*

Proof. Let $\epsilon = c^{-1/2}$. Let $V_0 \subseteq V(G)$, and we may assume that $|V_0| = \lceil \log k \rceil$. We modify the DFS-algorithm as follows.

Recall that the stack S denotes the vertices that are currently under exploration. If $S = \emptyset$ in some step of the algorithm, then as long as possible we select a vertex of $V_0 \cap U$

as the new root of a component and put it onto the stack S . Hence, by this modified DFS-algorithm, at least up to the point when we explored $\log k$ vertices, the root of the current component is in V_0 .

We stop this modified DFS-algorithm when we reach $|R \cup S| = (1 - \epsilon)k$. Let \mathcal{A} be the event that $S = \emptyset$ at some moment after $\frac{1}{2} \log k$ steps of the algorithm and let \mathcal{B} be the event that there are less than $(1 - \epsilon)k$ positive answers among the first $\frac{k}{p} = \epsilon^2 k^2$ tested edges.

Claim. $\mathbb{P}[\mathcal{A} \cup \mathcal{B}] = o(1)$.

Assuming this claim we can a.a.s. find a path of length $(1 - \epsilon)k$ starting in a vertex of V_0 as follows.

Suppose neither \mathcal{A} nor \mathcal{B} holds. Consider the step of the DFS-algorithm at which we reach $|R \cup S| = (1 - \epsilon)k$. Thus the root of the current component is contained in V_0 , as \mathcal{A} does not hold. Due to property (I) such a step exists. Recall that the vertices in S form a path (property (II)). If $|S| \geq (1 - 2\epsilon)k$, then the statement of the lemma follows directly. Thus, we may assume that

$$|S| < (1 - 2\epsilon)k \quad (4)$$

which implies $|R| > \epsilon k$. Moreover, each vertex in R has at least $k - |R \cup S| \geq \epsilon k$ neighbors in G in the set of unreached vertices U . Due to property (III), all these edges between R and U have been queried and answered negatively. Hence at least $|R| \cdot \epsilon k > \epsilon^2 k^2$ queries are answered negatively and less than $(1 - \epsilon)k$ are answered positively. Thus \mathcal{B} holds, which is a contradiction.

We complete the proof of the lemma by the proof of claim. For a positive integer i , let \mathcal{A}_i be the event that we complete exploring a component when $|R| = i$. Since every vertex has degree at least k , in this moment of the algorithm every vertex in R has at least $k - i \geq \epsilon k$ neighbors in U (for $i \leq (1 - \epsilon)k$) and all these edges are queried negatively. Thus we queried at least $i\epsilon k$ edges in total, and had at most i positive answers. The probability that this occurs is at most the probability that a binomial distributed random variable X_i with $X_i \sim \text{Bin}(i\epsilon k, p)$ is at most i . Hence $\mathbb{E}X_i = i\epsilon c = ic^{1/2}$. By Chernoff's inequality, we obtain

$$\mathbb{P}[\mathcal{A}_i] \leq \mathbb{P}[X_i \leq i] \leq \mathbb{P}\left[\left|X_i - ic^{1/2}\right| \geq \frac{ic^{1/2}}{2}\right] \leq 2e^{-\frac{ic^{1/2}}{12}} \leq \frac{1}{2^i}.$$

Using the union bound leads to the desired result

$$\mathbb{P}[\mathcal{A}] \leq \mathbb{P}\left[\bigcup_{i=\frac{1}{2} \ln k}^{(1-\epsilon)k} \mathcal{A}_i\right] \leq \sum_{i=\frac{1}{2} \ln k}^{(1-\epsilon)k} \mathbb{P}[\mathcal{A}_i] \leq \sum_{i=\frac{1}{2} \ln k}^{(1-\epsilon)k} \frac{1}{2^i} = o(1).$$

An upper bound for the event \mathcal{B} follows by a direct applications of Chernoff's inequality. Let Y be a binomial distributed random variable with $Y \sim \text{Bin}\left(\frac{k}{p}, p\right)$. Then,

$$\mathbb{P}[\mathcal{B}] \leq \mathbb{P}[Y \leq (1 - \epsilon)k] \leq 2 \exp\left(-\frac{\epsilon^2 k}{3}\right) = o(1).$$

This implies $\mathbb{P}[\mathcal{A} \cup \mathcal{B}] = o(1)$, which completes the proof of the claim and thus the proof of the lemma. \square

5 Long Cycles in Pseudo-Cliques

Consider the well-known $G(n, p)$ -model and with our notation $(K_n)_p$. It is very natural and intuitive that $(K_n)_p$ and H_p typically have the same properties if H is a graph on n vertices which is almost a clique. In this section we indicate that a result of Frieze [6] can be suitably modified.

Let $\gamma > 0$ be a constant sufficiently small, say $\gamma < 10^{-5}$. We call a graph G on n vertices a k -pseudo-clique (or simply pseudo-clique) if its minimum degree is at least k and $n \leq (1 + \gamma)k$.

We say a vertex v has *small* degree if $d_G(v) \leq \frac{c}{10}$ and otherwise its degree is *large*. Let S and L be the set of all vertices of small and large degree in G_p , respectively. For $1 \leq i \leq 4$, let W_i be the set of all vertices v of small degree such that there is a vertex w of small degree and a v, w -path of length i or v is contained in a cycle of length i . We set $W = W_1 \cup \dots \cup W_4$.

The following lemmas are extensions of the results of Frieze [6], who prove the analogous results for $G = K_{k+1}$. Lemma 11 summarizes a several properties of G_p . The proof is very similar as the original proof of Frieze. The extra factor $(1 + \gamma)$ for the larger order disappears quickly in all the calculations. The same applies for the proofs of Lemma 12 and 13. Theorem 14 needs even fewer changes as it uses the properties provided by Lemma 11–13. Thus we decided to omit the proofs.

We denote by $e(G)$ the number of edges of G and for two disjoint sets $X, Y \subseteq V(G)$, we denote by $G[X, Y]$ the bipartite subgraph of graph G spanned by the vertices X and Y .

Lemma 11. *Let G be a k -pseudo-clique on n vertices, $p = \frac{c}{k}$ and let $\ell \geq 7$ be a fixed integer. Then a.a.s. G_p has the following properties,*

- (a) $|\{v \in V(G) : d_{G_p}(v) \leq \frac{c}{10} + 1\}| \leq (1 + \gamma)ke^{-\frac{2}{3}c},$
- (b) *for all sets $Z \subseteq V(G)$ with $|Z| \geq ke^{-c}$, we have $|\{e \in E(G_p) : e \cap Z \neq \emptyset\}| \leq 4c|Z|,$*
- (c) $\Delta(G_p) \leq 4 \log k,$
- (d) $|W| \leq c^4 e^{-\frac{4c}{3}} k,$
- (e) $\emptyset \neq Z \subseteq L$ and $|Z| \leq \frac{k}{2\ell}$ implies $|N_{G_p}(Z)| \geq \ell|Z|$, and
- (f) $Z \subseteq V(G)$ and $\frac{k}{2\ell} \leq |Z| \leq \frac{1}{2}k$ implies $e(G_p[Z, V(G) \setminus Z]) \geq \frac{c|Z|}{3\ell}.$

Let us comment on the changes that need to be done to the calculations due to the original Lemma 2.1 due to Frieze. The statements (a)-(f) refer to (2.1)-(2.6) in Lemma 2.1. As (a) follows directly by a standard application of Chernoff's inequality and the 'linearity

of expectations' no further changes are needed. For (b) a factor of $(1 + \gamma)^{4+1/c} < 11/10$ has to be added. Since $\sum_{i \geq ke^{-c}} \left(\frac{11e^{5+1/c}}{2560}\right)^{ci}$ still vanishes as $k \rightarrow \infty$, this proves (b). The calculation for (c) is easy to modify and would even work for $\gamma = 1$. For (d), observe that $e^{-2/3}$ is once estimated by $e^{-0.669}$, so we have here some room to spare to add another factor of $(1 + \gamma)$. For (e), observe that in the proof of (2.5) the factor $e^{-\frac{3 \cdot 20}{7}}$ is once estimated by e^{-8} , again giving enough room to add the corresponding factors of $1 + \gamma$. For (f), the same estimation as in the proof of (2.6) holds for c and k large enough as $(1 + \gamma)/c \rightarrow 0$ for $c \rightarrow \infty$ and hence, it still holds

$$2 \sum_{i=\lfloor k/(2\ell) \rfloor}^{\lfloor k/2 \rfloor} \left(\frac{(1 + \gamma)ke}{i} \right)^i \left(\frac{3\ell i((1 + \gamma)k - i)e}{ci} \right)^{ci/(3\ell)} \left(\frac{c}{k} \right)^{ci/(3\ell)} e^{-ci/3} = o(1).$$

Lemma 12. *Let G be a k -pseudo-clique on n vertices, $p = \frac{c}{k}$, and let X_1, X_2, \dots be a sequence obtained by the following rule*

$$X_i = \left\{ v \in V(G) : \left| N_{G_p}(v) \cap \left(S \cup \bigcup_{j=1}^{i-1} X_j \right) \right| \geq 2 \right\}.$$

If $X = \bigcup_{j \geq 1} X_j$, then $|X| \leq 500c^4 e^{-\frac{4c}{3}} k$ a.a.s.

Again, let us comment on the original Lemma 2.2 due to Frieze. It originally states that $|X| \leq 2e^4 c^4 e^{-4c/3} n$ a.a.s. By observing that $2e^4(1 + \gamma) \leq 500$, it can be easily seen that the proof of Lemma 12 works analogously as the original proof of Frieze. This also explains why the additional factor of $1 + \gamma$ vanishes in our statement.

Let V_2 be vertex set of the largest subgraph of G_p with minimum degree 2 ($G_p[V_2]$ is also known as the 2-core). Moreover, let Y be the set of all vertices v in G_p which have degree 2 and have a neighbor in X . Let $A = V_2 \setminus (W \cup X \cup Y)$.

Lemma 13. *Let G be a k -pseudo-clique on n vertices and $p = \frac{c}{k}$. Then, a.a.s.*

$$|A| \geq (1 - (1 + \epsilon(c))ce^{-c}) k,$$

where $\epsilon(c) \rightarrow 0$ as $c \rightarrow \infty$.

The above lemma refers to Lemma 2.4 of Frieze. Note that there is a little typo in Frieze's version and the corrected statement (2.27) should be

$$|A_{k'}| \geq \left(1 - (1 + \epsilon(c)) \frac{c^{k'-1}}{(k' - 1)!} e^{-c} \right) n.$$

In our case, we have $k' = 2$. As $n \geq k$ for the minimum degree k , no further changes are needed.

Having proved these three lemmas for pseudo-cliques, one can go once again along the lines of the result of Frieze to obtain the following.

Theorem 14. Let G be a k -pseudo-clique on n vertices and $p = \frac{c}{k}$. Then, a.a.s. G_p contains a cycle of length at least

$$(1 - (1 + \epsilon(c))ce^{-c})k,$$

where $\epsilon(c) \rightarrow 0$ as $c \rightarrow \infty$.

As the proof of Theorem 14 only exploits the statements of the lemmas before, no alterations are necessary.

6 Long Paths

This section is devoted to the proof of Theorem 2. This proof is inspired by a result in [8] which states that a.a.s. the random subgraph G_p of a graph G of minimum degree at least k contains a path of length k if $p = \frac{(1+\epsilon)\log k}{k}$ for any fixed $\epsilon > 0$.

Our strategy for the proof is as follows. If G contains a pseudo-clique, then Theorem 14 gives rise to the desired path. Otherwise, we use Theorem 3 that guarantees a cycle C of length $(1 - O(c^{-1/5}))k$. Next we either try to extend C by a path that leads away from it or we use a set of vertices, which is small, but large enough for our purposes, that forms together with C a larger cycle.

Proof of Theorem 2. Let c be sufficiently large, k be sufficiently large in terms of c , and let $\epsilon = 5\left(\frac{c}{3}\right)^{-1/5}$. If G contains a set $V' \subseteq V(G)$ such that

$$\left(1 - \frac{1}{\log k}\right)k \leq |V'| \leq (1 + 10\epsilon)k \tag{5}$$

and the minimum degree of the graph $G[V']$ is at least $(1 - \frac{2}{\log k})k$, then by Theorem 14, G_p a.a.s. contains a cycle of length at least

$$(1 - (1 + \delta(c))ce^{-c})k,$$

for some function $\delta(c) \rightarrow 0$ as $c \rightarrow \infty$, which implies the statement. Hence, we may assume that G does not contain such a set V' .

In the following, we use a technique which is known as *sprinkling*. In our case, we expose the edges of G_p in three rounds and in each round we suppose an edge to be present independently with probability $\frac{c}{3k}$. Thus we consider the union of three graphs $G_{p_1} \cup G_{p_2} \cup G_{p_3}$, where $p_i = \frac{c}{3k}$. As

$$1 - (1 - p_1)(1 - p_2)(1 - p_3) = \left(1 - \frac{c}{3k} + \frac{c^2}{27k^2}\right) \frac{c}{k} \leq p,$$

the union of these three graphs underestimates the model G_p . Therefore, if we can show that $G_{p_1} \cup G_{p_2} \cup G_{p_3}$ a.a.s. contains a path of the desired length, then also G_p a.a.s. contains such a path.

By Theorem 3, we know that G_{p_1} a.a.s. contains a cycle C of length at least $(1 - \epsilon)k$ (in the proof we use explicitly this constant). Moreover, we may assume that $|C| < (1 - (1 + \delta(c))ce^{-c})k$. Let $A \subseteq V(G) \setminus V(C)$ be the set of vertices having at least $(1 - 20\epsilon)k$ neighbors in $V(C)$ and let $B = V(G) \setminus (V(C) \cup A)$.

We divide the proof into two parts. First, we suppose that $|A| \leq 10\epsilon k$. Hence, if $B \neq \emptyset$, then $G[B]$ has minimum degree at least $10\epsilon k$.

Suppose first that at least $4k \log k$ edges join B and C in G and denote this set by E . Consider an ordering b_1, b_2, \dots of the vertices in B and consider an ordering e_1, e_2, \dots of the edges in E which respects the ordering on B , that is, if $i < j$, then the indices of the edges incident to b_i are smaller than the indices of the edges incident to b_j . For $1 \leq i \leq \lceil 2 \log k \rceil$, let $E_i = \{e_j : (2i - 2)k < j \leq (2i - 1)k\}$. This implies that there is no vertex $b \in B$ incident to an edge in E_i and E_j for $i \neq j$, since a vertex in B has at most $|V(C)| \leq k$ neighbors in C . Moreover, with probability at least $1 - e^{-\frac{\epsilon}{3}}$, every set E_i contains at least one edge in G_{p_2} independently for every i . Thus by Chernoff's inequality, at least $\log k$ sets E_i contain an edge in G_{p_2} with probability $1 - o(1)$. Let S be a set of $\log k$ vertices in B incident to an edge in G_{p_2} . By Lemma 10, with probability $1 - o(1)$, there is a path in $G_{p_3}[B]$ starting in S of length, say, ϵk . The union of C , a suitable edge in some E_i , and this path leads to a path in $G_{p_1} \cup G_{p_2} \cup G_{p_3}$ of length at least k with probability $1 - o(1)$.

Therefore, we may assume that at most $4k \log k$ edges join B and C in G . Observe that

$$\text{there are at most } \sqrt{k} \text{ vertices } v \text{ in } C \text{ with } d_B(v) \geq k^{2/3}, \quad (6)$$

where $d_B(v)$ denotes the number of neighbors that v has in B . Moreover,

$$|A \cup C| \geq k - 5 \log k, \quad (7)$$

otherwise every vertex in C has at least $5 \log k$ neighbors in B contradicting our assumption.

Next, we suppose that there exists a set $A' \subseteq A$ with at least \sqrt{k} many vertices having at least $k^{\frac{2}{3}}$ many neighbors in B . Observe that for any vertex $v \in A$ (and so any in A') the probability that v has an edge in G_{p_2} that connects v with C is at least $2/3$. Thus with probability $1 - o(1)$, there exists a set $A'' \subseteq A'$ of size at least $\frac{|A'|}{2}$ such that every vertex in A'' is adjacent to C in G_{p_2} . By a similar argument as before, with probability $1 - o(1)$, there are $\log k$ vertices in B such that each of them has a neighbor in A'' in G_{p_2} . Again, with probability $1 - o(1)$, there is a path in $G_{p_3}[B]$ of length at least ϵk starting in one of these vertices in B and this leads to a path of length at least k in $G_{p_1} \cup G_{p_2} \cup G_{p_3}$ with probability $1 - o(1)$.

Therefore, we may assume that there are at most \sqrt{k} vertices $v \in A$ with $d_B(v) \geq k^{\frac{2}{3}}$. Together with (6), we conclude that there are at most $2\sqrt{k}$ vertices v in $A \cup C$ with $d_B(v) \geq k^{\frac{2}{3}}$. Let Z be obtained from $A \cup C$ by deleting all these vertices. Using (7), this implies $|Z| \geq k - 3\sqrt{k}$. As $|Z| \leq (1 + 10\epsilon)k$, the set Z is a set as in (5), which is a contradiction.

Thus from now on, we may assume that $|A| \geq 10\epsilon k$. Let $A_1 \subseteq A$ with $|A_1| = 10\epsilon k$. We partition C into $\frac{1}{10\epsilon}$ cycle segments S_1, S_2, \dots each of length almost $10\epsilon k$. As every

vertex in A_1 has at least $(1 - 20\epsilon)k$ neighbors in C , by a simple average argument, there is a segment, say S_1 , such that the number of edges between S_1 and A_1 is at least $(1 - 20\epsilon)|A_1||S_1|$. Let H be the bipartite subgraph of G which is induced by A_1 and S_1 . This implies that the bipartite complement of H has at most $20\epsilon \cdot (10\epsilon k)^2 = 2000\epsilon^3 k^2$ edges. Of course, this graph contains at most $100\epsilon^{\frac{3}{2}}k$ vertices of degree at least $100\epsilon^{\frac{3}{2}}k$. Let H' be the graph obtained by deleting these vertices from H . Thus H' has minimum degree at least $(1 - 20\sqrt{\epsilon}) \cdot 10\epsilon k$.

For some orientation of C , let L and R be the first and last ϵk vertices on C in S_1 . Moreover, remove an arbitrary subset of A_1 to obtain from the graph $H' \setminus (R \cup L)$ a balanced bipartite graph H'' . Thus H'' has minimum degree at least $(1 - 25\sqrt{\epsilon}) \cdot 8\epsilon k$.

By Lemma 9, H'' contains a path P of length $15\epsilon k$ in G_{p_2} with probability $1 - o(1)$. Let P_1 and P_2 be the subpaths at the beginning and at the end of P of length ϵk , respectively. With probability at least $1 - (1 - p_3)^{\epsilon^2 k^2 / 2} \geq 1 - o(1)$, there exists an edge e_1 in G_{p_3} that joins a vertex in L and $V(P_1) \cap A_1$ and an edge e_2 joining a vertex in R and $V(P_2) \cap A_1$.

Combining the subpath of C between the endpoints of e_1 and e_2 that contains the segment S_2 , the subpath of P between the endpoints of e_1 and e_2 , and the edges e_1 and e_2 results in a cycle in $G_{p_1} \cup G_{p_2} \cup G_{p_3}$ of length at least $(1 - 11\epsilon)k + 13\epsilon k \geq k$ and this completes the proof. \square

References

- [1] M. Ajtai, J. Komlós, and E. Szemerédi, *The longest path in a random graph*, *Combinatorica* **1** (1981), 1–12.
- [2] N. Alon and J. H. Spencer, *The probabilistic method*, John Wiley & Sons, 2004.
- [3] B. Bollobás, *Long paths in sparse random graphs*, *Combinatorica* **2** (1982), 223–228.
- [4] B. Bollobás, T. I. Fenner, and A. M. Frieze, *Long cycles in sparse random graphs*, *Graph theory and combinatorics* (Cambridge, 1983), Academic Press, London, 1984, pp. 59–64.
- [5] W. Fernandez de la Vega, *Long paths in random graphs*, *Studia Sci. Math. Hungar.* **14** (1979), 335–340.
- [6] A. M. Frieze, *On large matchings and cycles in sparse random graphs*, *Disc. Math.* **59** (1986), 243–256.
- [7] R. Glebov, H. Naves, and B. Sudakov, *The threshold probability for long cycles*, *Comb., Probab. Comput.* **26** (2017), 208–247.
- [8] M. Krivelevich, C. Lee, and B. Sudakov, *Long paths and cycles in random subgraphs of graphs with large minimum degree*, *Random Structures Algorithms* **46** (2015), 320–345.
- [9] M. Krivelevich and B. Sudakov, *The phase transition in random graphs: a simple proof*, *Random Structures Algorithms* **43** (2013), 131–138.
- [10] O. Riordan, *Long cycles in random subgraphs of graphs with large minimum degree*, *Random Structures Algorithms* **45** (2014), 762–765.