

---

---

# The Threshold Probability for Long Cycles

---

ROMAN GLEBOV<sup>1†</sup> HUMBERTO NAVES<sup>2‡</sup> and BENNY SUDAKOV<sup>3§</sup>

<sup>1</sup>School of Computer Science and Engineering, Hebrew University, Jerusalem 9190401, Israel  
(e-mail: roman.1.glebov@gmail.com)

<sup>2</sup>Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, MN 55455, USA  
(e-mail: hnaves@ima.umn.edu)

<sup>3</sup>Department of Mathematics, ETH, 8092 Zurich, Switzerland  
(e-mail: benjamin.sudakov@math.ethz.ch)

*Received 15 August 2014; revised 31 May 2016; first published online 8 September 2016*

For a given graph  $G$  of minimum degree at least  $k$ , let  $G_p$  denote the random spanning subgraph of  $G$  obtained by retaining each edge independently with probability  $p = p(k)$ . We prove that if  $p \geq (\log k + \log \log k + \omega_k(1))/k$ , where  $\omega_k(1)$  is any function tending to infinity with  $k$ , then  $G_p$  asymptotically almost surely contains a cycle of length at least  $k + 1$ . When we take  $G$  to be the complete graph on  $k + 1$  vertices, our theorem coincides with the classic result on the threshold probability for the existence of a Hamilton cycle in the binomial random graph.

2010 *Mathematics subject classification*: Primary 05C80  
Secondary 05C38, 05C85, 05D40, 05C05

## 1. Introduction

Given a graph  $G$  and a real  $p \in [0, 1]$ , let  $G_p$  be the probability space of subgraphs of  $G$  obtained by taking each edge of  $G$  independently with probability  $p$ . We sometimes use the notation  $(G)_p$  to avoid ambiguity. For a given graph property  $\mathcal{P}$  and sequences of graphs  $\{G_i\}_{i=1}^\infty$  and probabilities  $\{p_i\}_{i=1}^\infty$ , we say that  $(G_i)_{p_i} \in \mathcal{P}$  *asymptotically almost surely*, or a.a.s. for brevity, if the probability that  $(G_i)_{p_i} \in \mathcal{P}$  tends to 1 as  $i$  goes to infinity. In

<sup>†</sup> Previous affiliation: Department of Mathematics, ETH, 8092 Zurich, Switzerland. Research supported by the ERC grant ‘High-dimensional Combinatorics’ at the Hebrew University.

<sup>‡</sup> Previous affiliation: Department of Mathematics, ETH, 8092 Zurich, Switzerland, and Department of Mathematics, UCLA, Los Angeles, CA 90095, USA. Research supported in part by the Institute for Mathematics and its Applications with funds provided by the National Science Foundation.

<sup>§</sup> Research supported in part by SNSF grant 200021-149111 and by a USA–Israel BSF grant.

this paper, when  $G$  and  $p$  depend upon some parameter, we abuse notation and consider  $G$  and  $p$  as sequences obtained by taking the parameter to tend to infinity, and we say that  $G_p$  has  $\mathcal{P}$  a.s. if the sequence does.

When the host graph  $G$  is the complete graph on  $n$  vertices, the random graph model  $G_p$  coincides with the classic binomial random graph model  $\mathbb{G}(n, p)$ , introduced independently by Gilbert in [7] and by Erdős and Rényi in [6]. This important model has been studied extensively for the past few decades. A result of Pósa [15] states that for some large constant  $C > 0$ , if  $p \geq (C \log n)/n$  then  $\mathbb{G}(n, p)$  a.s. contains a Hamilton cycle. This result was later strengthened by Korshunov [11], Komlós and Szemerédi [10], and independently by Bollobás [3]. They proved that the same statement holds for  $p \geq (\log n + \log \log n + \omega_n(1))/n$ , provided  $n$  is large.

In this paper we extend the aforementioned result to a more general class of graphs. More precisely, we would like to replace the host graph  $G$ , taken to be the complete graph in the classic setting, by a graph with minimum degree at least  $k$ , and to find a.s. a cycle of length at least  $k + 1$  in the random subgraph  $G_p$ . Our main result is as follows.

**Theorem 1.1.** *Let  $G$  be a graph with minimum degree at least  $k$ . If*

$$p = p(k) \geq \frac{\log k + \log \log k + \omega_k(1)}{k},$$

*then  $G_p$  a.s. contains a cycle of length at least  $k + 1$ .*

Our results are complementary to those of Krivelevich, Lee and Sudakov [12] and Riordan [16]. They proved that for  $p = \omega_k(1)/k$ , the graph  $G_p$  a.s. contains a cycle of length at least  $(1 + o(1))k$ , which might be slightly less than  $k + 1$ . Since the property stated in the main theorem is monotone increasing, we assume throughout the paper that  $p \leq (\log k + 2 \log \log k)/k$ .

The rest of this paper is organized as follows. Section 2 contains a variety of tools, which are used to prove Theorem 1.1. All propositions, statements and lemmas in that section are stated without proofs. In Section 3, we prove our main theorem. The final section contains some concluding remarks.

**1.1. Notation**

A graph  $G = (V, E)$  is given by a pair of its (finite) vertex set  $V(G)$  and edge set  $E(G)$ . We use  $|G|$  or  $|V(G)|$  to denote the order of the graph. For a subset  $X$  of vertices, we use  $e(X)$  to denote the number of edges spanned by  $X$ , and for two disjoint sets  $X, Y$ , we use  $e(X, Y)$  to denote the number of edges with one endpoint in  $X$  and the other in  $Y$ . Let  $G[X]$  denote the subgraph of  $G$  induced by a subset of vertices  $X$ . We write  $N(X)$  to denote the collection of vertices outside  $X$  that have at least one neighbour in  $X$ . When  $X$  consists of a single vertex, we abbreviate  $N(v)$  for  $N(\{v\})$ , and let  $\deg(v)$  denote the cardinality of  $N(v)$ , that is, the degree of  $v$ . For two graphs  $G_1$  and  $G_2$ , not necessarily over the same vertex set, we define their intersection as

$$G_1 \cap G_2 = (V(G_1) \cap V(G_2), E(G_1) \cap E(G_2))$$

and their union as

$$G_1 \cup G_2 = (V(G_1) \cup V(G_2), E(G_1) \cup E(G_2)).$$

Moreover, if  $X$  is a set of vertices, we let  $G \setminus X$  be the induced subgraph  $G[V(G) \setminus X]$ . Finally, if  $G$  is a graph and  $E$  is a collection of unordered pairs of vertices from  $V(G)$ , let  $G + E$  denote the graph obtained from  $G$  by adding the edges in  $E$  which are not already in  $G$ . When there are several graphs under consideration, we use subscripts such as  $N_G(X)$  indicating the relevant graph of interest.

The probability space  $G_p$  is a simple product space. When sampling from this model, one could unveil the graph  $G_p$  by successively answering queries of the form ‘does  $e$  belong to  $G_p$ ?’ for each edge  $e \in E(G)$ . Since the answers to these queries are independent, this process can be carried out regardless of the order of the queries, as long as each edge of  $G$  is queried exactly once. Throughout the paper we expose  $G_p$  in this manner. The edges of  $G$  not yet queried in  $G_p$  shall be named *untested*, while the others are called *tested*. When an edge  $e$  from  $G$  is queried and the outcome turns out to be positive, we say that  $e$  was *successfully tested*, or equivalently,  $e$  was *successfully exposed*. We write *partially exposed*  $G_p$  as a reminder that not all edges of  $G$  were tested in  $G_p$ . All probabilistic statements involving a partially exposed  $G_p$  must be conditioned on the outcome of the tested edges at that particular moment of the exposure process. More precisely, if  $Q$  is the set of tested edges, and  $E \subseteq Q$  is the set of successfully tested edges of the partially exposed  $G_p$ , then for each subgraph  $\Gamma \subseteq G$ , the probability that we obtain the graph  $\Gamma$  after we expose all the remaining untested edges is  $\mathbb{P}[G_p = \Gamma \mid E(G_p) \cap Q = E]$ .

To simplify the presentation, we often omit floor and ceiling signs whenever these are not crucial, and make no attempt to optimize the absolute constants involved. We also assume that the parameter  $k$  (which always denotes the minimum degree of the host graph) tends to infinity and therefore is sufficiently large whenever necessary. All our asymptotic notation symbols ( $O$ ,  $o$ ,  $\Omega$ ,  $\omega$ ,  $\Theta$ ) are relative to this variable  $k$ , unless otherwise specified with a subscript. Finally, all logarithms are to base  $e \approx 2.718$ .

## 2. Preliminaries

### 2.1. Probabilistic tools

We use extensively the following well-known bounds on the lower and upper tails of the binomial distribution due to Chernoff (see e.g. [1, Theorems A.1.11, A.1.13, and A.1.12]).

**Lemma 2.1.** *If  $X \sim \text{Bin}(n, p)$ , then*

- $\mathbb{P}[X < (1 - a)np] < \exp\left(-\frac{a^2 np}{2}\right)$  for every  $a > 0$ ,
- $\mathbb{P}[X > (1 + a)np] < \exp\left(-\frac{a^2 np}{3}\right)$  for every  $0 < a < 1$ .

**Lemma 2.2.** *Let  $X \sim \text{Bin}(n, p)$  and  $a \in \mathbb{N}$ . Then  $\mathbb{P}[X \geq a] \leq (enp/a)^a$ .*

## 2.2. Depth-first search algorithm

Depth-first search (DFS) is a well-known graph exploration algorithm, usually applied to discover the connected components of an input graph. The algorithm visits all vertices of a graph  $H$  (the input of the DFS) and produces a rooted spanning forest  $T$  of  $H$  (the output). It also maintains a stack  $S$  (last-in/first-out data structure) of vertices. Initially, the stack is empty, and all vertices of  $H$  are *active*. Each active vertex  $v$  eventually gets reached, henceforth becoming inactive, and is then pushed into  $S$ . At some point later, the same vertex  $v$  is popped from  $S$  and is declared *explored*. Once a vertex becomes explored, it never changes its state back to active again. Indeed, the algorithm ends when all the vertices of  $H$  become explored. The main loop of the DFS is as follows.

- (i) If  $S$  is empty, choose an active vertex  $v$ , deactivate it, and push it onto the stack. The vertex  $v$  is the root of a new tree in  $T$ .
- (ii) Otherwise, let  $u$  be the unique vertex on top of the stack  $S$ . The algorithm then queries for active neighbours of  $u$  in  $H$ , that is, active vertices  $w$  such that  $uw$  forms an edge in  $H$ . If there is such an edge, we remove  $w$  from the set of active vertices and place it on top of  $S$ . Otherwise, we just pop  $u$  from the top of  $S$  and mark it as explored.

Notice that we specified neither how to choose the new vertex  $v$  in (i) nor the order in which the neighbours of  $u$  should be queried in (ii). It was implicitly assumed that these choices were made according to some predetermined order – the *priority* of the DFS.

The rooted spanning forest  $T$  produced by the DFS induces a partial order on the vertices of  $H$ . Namely, we say that  $u \leq_T v$  if  $u$  belongs to the (unique) path connecting  $v$  to a root of  $T$ . In this case, we say that  $u$  is an *ancestor* of  $v$ , or equivalently,  $v$  is a *descendant* of  $u$  with respect to  $T$ . Whenever  $uv \in E(T)$ , we say that  $v$  is an *immediate descendant* of  $u$ , or, equivalently,  $u$  is an *immediate ancestor* of  $v$ . A key observation is the following.

**Proposition 2.3.** *For every edge  $uv$  of  $H$ ,  $u$  and  $v$  are comparable with respect to  $\leq_T$ .*

In our setting, we utilize the DFS algorithm on the random graph  $G_p$ , and expose an edge only at the moment when its existence is queried by the algorithm. Note that the input graph  $G_p$  might already be partially exposed at the moment we start the DFS. In this case it is perfectly possible that the algorithm re-uses some of the successfully exposed edges (the algorithm never queries the same edge twice). We discuss this topic in more detail in Section 3.2.

Regardless of the portion of  $G_p$  that has already been exposed, the following is always true.

**Proposition 2.4.** *The rooted forest  $T$  produced by the DFS algorithm running on a partially exposed  $G_p$  contains all successfully tested edges revealed by the algorithm.*

For instance, if we apply the DFS to  $G_p$  with all the edges of  $G$  initially untested, since the resulting forest  $T$  has at most  $n - 1$  edges, the algorithm must necessarily stop after the first  $n - 1$  successfully exposed edges. Moreover, the connected components of  $G_p$ ,

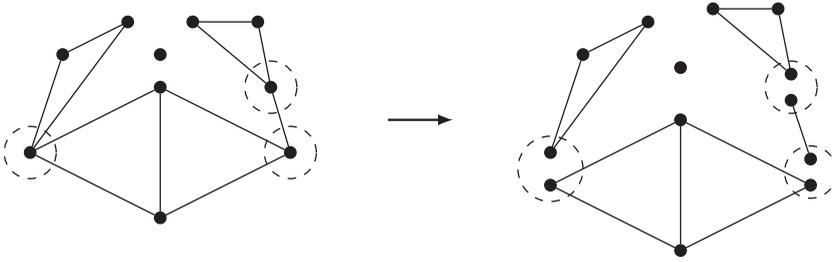


Figure 1. The block decomposition of a graph.

when viewed as vertex subsets of  $V(G)$ , are the same as the components of  $T$ , regardless of the outcome of the remaining untested edges from  $G_p$ . One noteworthy advantage of the DFS algorithm is that it produces this ‘certificate’ for the connected components of a random graph by testing very few of its edges. For more details on the application of the depth-first search algorithm to random graphs, we refer the reader to [13].

### 2.3. Block algorithm

Let us briefly recall some standard definitions and notions in graph theory. Let  $H$  be a graph. A vertex in  $H$  is a *cut-vertex* if by removing it we increase the number of connected components of  $H$ . A maximal connected subgraph of  $H$  without a cut-vertex is called a *block*. A *2-connected graph* is a graph of order at least 3 having no cut-vertex. In general, a  *$t$ -connected graph* is a graph  $H$  of order at least  $t + 1$  such that  $H \setminus X$  is connected for all subsets  $X \subseteq V(H)$  of size smaller than  $t$ .

In our quest to find long cycles, we will need to merge some already revealed cycles into longer ones. To merge two disjoint cycles, we need to find a collection of vertex-disjoint paths connecting them. A classic result of Menger [14] enables us to find these paths.

**Theorem 2.5 (Menger).** *Let  $H$  be a  $t$ -connected graph. For every pair of subsets  $A$  and  $B$  of  $V(H)$ , there are at least  $\min\{t, |A|, |B|\}$  vertex-disjoint paths in  $H$  that connect  $A$  and  $B$ .*

We extensively apply Menger’s result inside the blocks of  $G_p$ . This can be done because a block having at least three vertices is necessarily 2-connected. To discover the blocks of  $G_p$  we use another algorithm. Our proposed algorithm produces a similar ‘certificate’ for the blocks of  $G_p$ , just like the DFS does for the connected components of  $G_p$ .

Unlike the connected components of a graph  $H$ , the blocks of  $H$  must not necessarily be disjoint, as Figure 1 shows. In fact two blocks can intersect, but in at most one vertex. Moreover, it is well known that blocks form a forest-like structure. More formally, let  $H_{\text{block}}$  be the bipartite graph on the vertex set  $\mathcal{A} \cup \mathcal{B}$ , where  $\mathcal{A}$  is the set of all cut-vertices of  $H$ ,  $\mathcal{B}$  is the set of all blocks of  $H$ , and the edges are formed by pairs  $\{v, B\}$  satisfying  $v \in \mathcal{A}$ ,  $B \in \mathcal{B}$  and  $v \in B$ . The resulting graph  $H_{\text{block}}$ , referred to as the *block decomposition* of  $H$ , is always cycle-free. This graph is also commonly known as the *block-cutpoint graph* of  $H$ .

We summarize some of the properties of the block decomposition in the next proposition. For more details, we refer the interested reader to [5, Chapter 3] and [17, Chapter 4].

**Proposition 2.6.** *Let  $H$  be a graph and let  $H_{\text{block}}$  be its block decomposition with vertex set  $\mathcal{A} \cup \mathcal{B}$ .*

- (i) *The equality  $\bigcup_{B \in \mathcal{B}} V(B) = V(H)$  holds, and for every two distinct blocks  $B, B' \in \mathcal{B}$ , their intersection  $B \cap B'$  is either empty or contains exactly one cut-vertex from  $\mathcal{A}$ . Furthermore, we have  $|\mathcal{B}| \leq |V(H)|$ .*
- (ii) *The sets  $E(B)$  for  $B \in \mathcal{B}$  form a partition of  $E(H)$ .*
- (iii) *The graph  $H_{\text{block}}$  is always cycle-free. Moreover,  $H_{\text{block}}$  is a tree if  $H$  is connected.*

*Furthermore, if  $H$  and  $H^*$  are two graphs having the same number of connected components, where  $H$  is a spanning subgraph of  $H^*$ , then the following statements hold.*

- (iv) *Every cut-vertex from  $H^*$  is also a cut-vertex in  $H$ .*
- (v) *If  $H_{\text{block}} \not\cong H^*_{\text{block}}$  then there exists an edge  $e \in E(H^*) \setminus E(H)$  such that no block of  $H$  contains both of its endpoints.*

Algorithms that efficiently find the block decomposition of a graph are already known: see for instance [9] and [17, Chapter 4]. Let us briefly describe one possible approach to finding such a decomposition, which we shall call the *block algorithm*. The description of the algorithm is first given in the deterministic setting, and is later extended to the random setting.

Motivated by Proposition 2.6(v), we say that an unordered pair of vertices  $uv$ , where  $u, v \in V(H)$ , is *crossing* for  $H$  if  $u$  and  $v$  lie in the same connected component of  $H$ , and there is no block  $B$  in  $H_{\text{block}}$  containing both  $u$  and  $v$ . Note that by Proposition 2.6(ii), a crossing pair is necessarily a non-edge of  $H$ . Another important property of crossing pairs is the following.

**Proposition 2.7.** *Let  $e$  be a crossing pair for  $H$ . Then the number of blocks of  $H + \{e\}$  is strictly smaller than the number of blocks of  $H$ .*

The input of the block algorithm consists of a pair  $(H, H^*)$  of graphs, where  $H$  is a spanning subgraph of  $H^*$  having the same number of connected components as  $H^*$ . This requirement might seem rather artificial at first, but it greatly simplifies the description of the algorithm. The output of the block algorithm is a graph  $M$  such that  $H \subseteq M \subseteq H^*$  and  $M_{\text{block}} \simeq H^*_{\text{block}}$ .

Let  $M$  be the running graph. Initially we have  $M := H$ . The main loop of the algorithm proceeds as follows.

If there exists a crossing pair  $e \in E(H^*) \setminus E(M)$  for the graph  $M$ , we add  $e$  to  $M$  and iterate the loop again. Otherwise we stop and output  $M$ .

Clearly, at the end of the algorithm we obtain a graph  $M$  satisfying the required properties. Moreover, by Proposition 2.7, the number of iterations performed by the

algorithm is less than the number of blocks of  $H$ , as every new edge added to the running graph reduces the number of blocks of the graph  $M$ .

In the random setting, the input parameter  $H^*$  is a partially exposed random graph, and  $H$  is the graph containing the successfully exposed edges from  $H^*$ . As we did in the DFS algorithm, we only expose the edges of  $H^*$  when their existence is queried by the algorithm. One subtlety that should be remarked is that  $H_{\text{block}}^*$  is not known *a priori*, since the graph  $H^*$  is random. The algorithm works regardless. Moreover, Proposition 2.7 implies the following.

**Proposition 2.8.** *The number of edges successfully tested by the block algorithm with input  $(H, H^*)$  is less than the number of blocks of  $H$ .*

Recall that we need to ensure that  $H$  and  $H^*$  have the same number of connected components. To guarantee this assumption, before we start the block algorithm, we run the DFS on  $H^*$  and we always choose an input parameter  $H$  that contains the rooted spanning forest produced by the DFS.

#### 2.4. Pósa's rotation–extension technique

In this section we present yet another technique for showing the existence of long paths and cycles in graphs. This technique was introduced by Pósa [15] in his research on Hamiltonicity of random graphs.

In quite informal terms, Pósa's lemma guarantees that expanding graphs not only have long paths, but also provide a very convenient structure for augmenting a graph to a Hamiltonian one by adding new (random) edges. To formalize this assertion, we need some definitions. A graph  $H$  is an  $(m, 2)$ -expander if  $|N_H(X)| \geq 2|X|$  holds for every subset  $X \subseteq V(G)$  of size  $|X| \leq m$ . Given a non-Hamiltonian graph  $H$ , a non-edge  $e$  of  $H$  is called a *booster* if  $H + \{e\}$  is either Hamiltonian, or contains a path which is longer than any path in  $H$ . The following consequence of Pósa's technique (see e.g. [4, Lemma 8.5]) shows that every connected and non-Hamiltonian graph  $H$  with good expansion properties has many boosters.

**Lemma 2.9.** *If  $H$  is a connected non-Hamiltonian  $(m, 2)$ -expander, then the number of boosters for  $H$  is at least  $(m + 1)^2/2$ .*

### 3. Proof of the main result

For the rest of the paper, let  $\varepsilon = \varepsilon(k) := \log^{-1/10} k$  and let  $n$  be the number of vertices of  $G$ . We begin with the analysis of the structure of  $G$ . For that purpose, we make use of the following definition.

**Definition.** A subset  $C \subseteq V(G)$  of the vertices of  $G$  is a *pseudo-clique* if its size is bounded by  $(1 - 4\varepsilon)k < |C| \leq (1 + \varepsilon)k$ , and the minimum degree of  $G[C]$  is at least  $(1 - 4\varepsilon)k$ .

This important notion plays a fundamental role in our analysis of  $G$ . We later prove that if  $G$  is not covered by many pseudo-cliques with very few remaining vertices uncovered, then a.s.  $G_p$  contains a cycle of length at least  $k + 1$ . To make this statement more precise, let  $\mathcal{C}$  be a collection of vertex-disjoint pseudo-cliques in  $G$  such that the union of their vertices  $\bigcup_{C \in \mathcal{C}} C$  has maximum size. Vertices of  $G$  not in  $\bigcup_{C \in \mathcal{C}} C$  are called *outcast* vertices, and let  $\ell$  denote the number of such vertices. We prove the following.

**Lemma 3.1.** *If  $\ell > 10^7 \cdot n/(\varepsilon k)$  then a.s.  $G_p$  has a cycle of length at least  $k + 1$ .*

For the case when  $\ell$  is small, we have the following.

**Lemma 3.2.** *If  $\ell \leq 10^7 \cdot n/(\varepsilon k)$ , then either a.s.  $G_p$  contains a cycle of length at least  $k + 1$ , or there exist a pseudo-clique  $C \in \mathcal{C}$  and a set  $N$  of size  $|N| \leq 10$  such that there are at most  $\varepsilon k$  edges in  $G$  connecting  $C \setminus N$  to vertices not in  $C \cup N$ .*

But if there exists such a pair  $(C, N)$  as stated in Lemma 3.2,  $G_p$  must also have a cycle of length at least  $k + 1$  a.s., as the next lemma shows.

**Lemma 3.3.** *If there exist a pseudo-clique  $C \in \mathcal{C}$  and a set  $N \subseteq V(G)$  of size at most 10 such that  $e_G(C \setminus N, V(G) \setminus (C \cup N)) \leq \varepsilon k$ , then a.s.  $G_p[C \cup N]$  has a cycle of length at least  $k + 1$ .*

One can verify that lemmas 3.1, 3.2, and 3.3 together imply Theorem 1.1. In the next subsections, we devote ourselves to the proofs of these lemmas. Our argument is divided into six steps. In each step, we may reveal a portion of  $G_p$  by testing some of the edges from  $G$ . The six steps are as follows.

Step 1: Pseudo-cliques were named for one clear reason: with respect to  $G_p$  they behave similarly as if they were cliques. We formalize this claim by exposing the edges inside pseudo-cliques and showing that a typical pseudo-clique contains a relatively long cycle in  $G_p$ . We further delete from  $G$  a few vertices such that in the remainder, every pseudo-clique induces a (large) Hamiltonian graph in  $G_p$ . Finally, we prove that this deletion does not affect the host graph much.

Step 2: We run a modified DFS algorithm on the resulting graph from Step 1, handling pseudo-cliques as if they were single vertices. This way, the number of edges revealed in this step is small and bounded by a function that depends only on  $\ell$  and the number of pseudo-cliques.

Step 3: We proceed with the block algorithm. The number of edges that are revealed in this step is bounded similarly as in Step 2. Hence, after this step, we know the vertex sets of the blocks of  $G_p$ , and a.s. most outcast vertices still have almost  $k$  untested edges incident to them.

Step 4: The study of the internal structure of the blocks provides some insight into how pseudo-cliques can interact with each other and with other cycles. For instance, we prove

that if a block contains at least two pseudo-cliques then we have already exposed all the edges of a cycle of length at least  $k + 1$ .

Step 5: We use the results from the previous step, combined with some double-counting arguments, to prove Lemmas 3.1 and 3.2. The only remaining case for the next step is the existence of a block in our graph with one pseudo-clique, just a constant number of outcast vertices, and only a few edges between the pseudo-clique and the vertex set outside the block.

Step 6: Finally, we analyse the case that remained after the previous step. In some sense, this case is very close to the usual  $\mathbb{G}(n, p)$  model: almost all vertices have degree close to  $k$  inside the block, and almost no edges leave the pseudo-clique to the outside the block. Using expansion properties of the random subgraph of the block, we show that in this case too, we find a cycle of length at least  $k + 1$  asymptotically almost surely.

### 3.1. Step 1: preparing the pseudo-cliques

Pseudo-cliques behave similarly as if they were cliques in  $G$ . When exposed in  $G_p$ , pseudo-cliques typically contain large cycles of length close to  $k$ . However, there might be a certain small proportion of them not behaving in this typical way. The aim of this subsection is to show that this seldom happens, and therefore does not affect the remainder of the graph much.

Formally, let us consider a two-round exposure process. Recall that we fixed a collection  $\mathcal{C}$  of disjoint pseudo-cliques. In the first round, we test edges inside pseudo-cliques with probability  $p_1$ , where  $p_1$  is such that  $1 - p = (1 - p_1)^2$ . Observe that  $p_1$  is roughly  $\log k/2k$  and testing an edge with probability  $p$  (unsuccessfully) is the same as testing it twice (unsuccessfully) with probability  $p_1$ . Denote by  $G^-$  the resulting random subgraph. Let  $W_1$  be the set of vertices that have degree at most  $\log k/100$  inside their pseudo-cliques in  $G^-$ .

In the second round we again expose with probability  $p_1$  the edges inside pseudo-cliques in  $\mathcal{C}$  that were not successfully exposed during the first round; the resulting supergraph of  $G^-$  is denoted by  $G^+$ . For technical reasons, we would like the remainders  $C \setminus W_1$  of pseudo-cliques  $C \in \mathcal{C}$  to satisfy the properties:

- (P1)  $C \cap W_1$  has fewer than  $\varepsilon k/2$  vertices,
- (P2)  $e_{G^-}(X, Y) > 0$  for any two disjoint sets  $X, Y \subseteq C \setminus W_1$  of size at least  $6\varepsilon k$ ,
- (P3) the induced graph  $G^+[C \setminus W_1]$  is Hamiltonian.

We now define the set  $W_2$  to be the union of those pseudo-cliques  $C \in \mathcal{C}$ , for which the above properties do not simultaneously hold for  $C \setminus W_1$ . We refer to the set  $W := W_1 \cup W_2$  as the *waste*. The set  $W$  contains the vertices we aim to delete from the graph  $G$  to obtain the new graph  $G' := G \setminus W$ . Finally, let  $Z_1$  be the set of all outcast vertices  $u$  such that at least an  $\varepsilon/3$ -proportion of its neighbours from  $G$  belong to  $W$ . The probability that  $u \in Z_1$  is bounded by the following statement.

**Lemma 3.4.** *Let  $u$  be an outcast vertex. Then  $\mathbb{P}[u \in Z_1] \leq 1/k^3$ .*

We split the proof of Lemma 3.4 into several propositions, from which the statement of the lemma is a trivial consequence. The first proposition of the series ensures that a.s. most outcast vertices do not have many neighbours in  $W_1$ .

**Proposition 3.5.** *Let  $u$  be an outcast vertex and denote by  $d \geq k$  its degree in  $G$ . The probability that at least  $\varepsilon d/6$  neighbours of  $u$  belong to  $W_1$  is at most  $1/k^4$ .*

**Proof.** The probability that a vertex  $v$  from a pseudo-clique  $C$  has degree at most  $\log k/100$  in  $G^-[C]$  is already sufficiently small. However, these events are not independent: the event that  $v$  has small degree in  $G^-[C]$  is positively correlated with another vertex from the same pseudo-clique getting small degree in  $G^-[C]$ . Since the statement of the proposition is far from being tight, one possibility to overcome this technicality is the following. Let  $\vec{G}$  be the digraph obtained from  $G$  by replacing each edge  $vw \in E(G)$  with two oriented edges  $v\vec{w}, \vec{w}v \in E(\vec{G})$ . We test each of the  $2|E(G[C])|$  oriented edges corresponding to the edges of  $G[C]$  independently with probability  $p_2$ , where  $p_2$  is such that  $1 - p_1 = (1 - p_2)^2$ , and roughly

$$p_2 \approx \frac{p_1}{2} \approx \frac{\log k}{4k}.$$

Next, we say that we successfully exposed the (non-oriented) edge  $vw \in E(G)$  if we successfully exposed at least one of the oriented edges  $v\vec{w}$  or  $\vec{w}v$ . In this model, all non-oriented edges are exposed independently at random with probability  $p_1$ . Thus, we can assume that each edge  $vw$  of  $G[C]$  that became a non-edge also had two corresponding oriented non-edges,  $v\vec{w}$  and  $\vec{w}v$ , in the random digraph. Hence, in order for  $v$  to get at most  $\log k/100$  non-oriented edges, all but at most  $\log k/100$  of the oriented edges going out from  $v$  to other vertices of  $C$  must become non-edges. Now, these events ('all but at most  $\log k/100$  oriented edges going out from a fixed vertex from  $C$  to other vertices in  $C$  were tested as non-edges') are indeed independent for any two vertices from  $C$ .

For one vertex  $v \in C$ , since the minimum degree in  $G[C]$  is at least  $(1 - 4\varepsilon)k$ , the probability of this event is at most

$$\mathbb{P}[\text{Bin}((1 - 4\varepsilon)k, p_2) \leq \log k/100] < k^{-1/5} \tag{3.1}$$

due to Lemma 2.1. Thus, the probability that at least  $\varepsilon d/6$  neighbours of  $u$  belong to  $W_1$  is bounded by  $\mathbb{P}[\text{Bin}(d, k^{-1/5}) > \varepsilon d/6]$ , and another application of Lemma 2.1 finishes the proof of the proposition. □

For a pseudo-clique  $C$ , let us denote by  $C^-$  the remainder  $C \setminus W_1$ . Similarly to Proposition 3.5, we need to ensure that also for a vertex from a pseudo-clique  $C$ , after the first round of exposure, a.s. only a few neighbours of this vertex are in  $C \cap W_1$ . The proof of this proposition follows the lines of the proof of Proposition 3.5 and is therefore omitted.

**Proposition 3.6.** *For fixed  $C \in \mathcal{C}$  and  $u \in C$ , the probability that, in  $G^-$ , at least  $\log k/200$  of the neighbours of  $u$  are in  $C \cap W_1$  is at most  $1/k^6$ .*

We remark that one could have replaced  $\log k/200$  by a large constant in the statement of Proposition 3.6. Indeed, the number of neighbours of  $u$  in  $C \cap W_1$  can be roughly bounded by a binomial random variable of  $O(\log k)$  trials with success probability  $k^{-1/5}$ . However, we do not require such tight estimates.

The very same calculation also shows that a.a.s.  $C^-$  is large enough, as required to satisfy (P1).

**Proposition 3.7.** *For fixed  $C \in \mathcal{C}$ , with probability at least  $1 - 1/k^5$ , we have  $|C \cap W_1| < \epsilon k/2$ .*

Notice that the inequalities in Proposition 3.7 are again far from being sharp, but they already suffice for our purposes.

Recall that for the second property (P2), we need  $G^-[C^-]$  to have edges between any two reasonably large disjoint sets. The next proposition ensures that a.a.s. this is indeed the case.

**Proposition 3.8.** *For every  $C \in \mathcal{C}$ , with probability at least  $1 - 1/k^5$ , we have  $e_{G^-}(X, Y) > 0$  for any two disjoint sets  $X, Y \subseteq C^-$ , each of size at least  $6\epsilon k$ .*

**Proof.** In  $G$ , for every choice of the sets  $X, Y \subseteq C$ , we have  $e_G(X, Y) \geq 6\epsilon^2 k^2$ , as every vertex from  $X$  has at least  $|Y| - 5\epsilon k \geq \epsilon k$  neighbours in  $Y$ . This is because every vertex in a pseudo-clique  $C$  has at most  $5\epsilon k$  non-neighbours in  $G[C]$ . Thus the probability that  $e_{G^-}(X, Y) = 0$  is at most

$$(1 - p_1)^{6\epsilon^2 k^2} \leq \exp(-k\sqrt{\log k}).$$

Since there are at most  $4^k$  possible choices for the pair  $X, Y$ , a simple application of the union bound finishes the proof. □

For the last property (P3), required to ensure that a pseudo-clique  $C$  is not put into  $W_2$ , we need  $G^+[C^-]$  to be Hamiltonian. To prove the Hamiltonicity of  $G^+[C^-]$  we first show in the next proposition that  $G^-[C^-]$  is a good expander.

**Proposition 3.9.** *For  $C \in \mathcal{C}$ , with probability at least  $1 - 3/k^5$ , the induced graph  $G^-[C^-]$  is a  $(k/6000, 2)$ -expander.*

**Proof.** Suppose that there exists a set  $A \subset C^-$  of size  $|A| \leq k/6000$  such that  $|N_{G^-[C^-]}(A)| < 2|A|$ . Also assume that the conclusion of Proposition 3.6 does not hold for any vertex in  $C^-$ , that is, no vertex in  $C^-$  has more than  $\log k/200$  neighbours in  $C \cap W_1$ . This happens with probability at least  $1 - 2/k^5$  by the union bound. Thus, if  $u \in C^-$ , we have

$$\deg_{G^-[C^-]}(u) \geq \deg_G(u) - \log k/200 \geq \log k/200.$$

Now let  $B = A \cup N_{G^-[C^-]}(A)$ . Then  $|B| < 3|A| \leq k/2000$ , and

$$|E(G^-[B])| \geq |A| \log k/400 \geq |B| \log k/1200.$$

On the other hand, by Lemma 2.2 and the union bound, we have

$$\begin{aligned} & \mathbb{P}[\exists B \subset C, |B| \leq k/2000 : |E(G^-[B])| \geq |B| \log k/1200] \\ & \leq \sum_{b \leq k/2000} \binom{|C|}{b} \mathbb{P}\left[\text{Bin}\left(\binom{b}{2}, p_1\right) > b \log k/1200\right] \\ & \leq \sum_{b \leq k/2000} \binom{|C|}{b} \left(\frac{ep_1 \binom{b}{2}}{b \log k/1200}\right)^{b \log k/1200} < 1/k^5, \end{aligned}$$

where in the last inequality we used that  $|C| \leq (1 + \varepsilon)k$  and that  $p_1 \approx \log k/2k$ . This concludes the proof of the proposition.  $\square$

Finally, we show that with sufficiently high probability,  $G^+[C^-]$  is Hamiltonian. Notice that we could strengthen the statement and ask for  $G^+[C^-]$  to be Hamilton connected. However, Hamiltonicity suffices for our proof, and it is technically slightly easier to show.

**Proposition 3.10.** *For every  $C \in \mathcal{C}$ , with probability at least  $1 - 5/k^5$  all properties (P1), (P2), and (P3) hold for  $C$ .*

**Proof.** By Propositions 3.7, 3.8 and 3.9, we can assume  $G^-[C^-]$  is a connected  $(k/6000, 2)$ -expander on at least  $(1 - 5\varepsilon)k$  vertices, and satisfies properties (P1) and (P2). The connectivity of  $G^-[C^-]$  is a consequence of Proposition 3.9, which implies that every connected component of  $G^-[C^-]$  has at least  $k/2000$  vertices, together with Proposition 3.8. Conditioned on these assumptions, we would like to show that then  $G^+[C^-]$  is Hamiltonian with probability at least  $1 - 1/k^5$ . Indeed, if a supergraph  $H$  of  $G^-[C^-]$  is not Hamiltonian, Lemma 2.9 guarantees a quadratic number of boosters. Now, let us look at the second round of exposure as a random process, with non-edges of  $G^-[C^-]$  turning into edges one by one, analogous to the standard random process coupling  $\mathbb{G}(n, p)$  and  $\mathbb{G}(n, M)$ . The new edges are exposed in a random order, their number  $|E(G^+[C^-]) \setminus E(G^-[C^-])|$  is binomially distributed, and thus by Lemma 2.1 with probability at least  $1 - e^{-k}$ , there are  $\Omega(k \log k)$  new successfully exposed edges. After every exposed edge, we update the set of boosters – bearing in mind that there are still quadratically many of them. Hence, every successfully exposed edge is a booster with probability at least a constant bounded away from zero. Thus we expect that the number of additional exposed edges needed for the graph induced by  $C^-$  to become Hamiltonian is at most linear. Furthermore, we can use Lemma 2.1 to say that the probability that we expose  $\omega(k)$  edges and we do not make the graph on  $C^-$  Hamiltonian, is at most  $e^{-k}$ , and the statement of the proposition follows.  $\square$

The following statement can be derived in the same way as Proposition 3.10, hence we omit its proof.

**Proposition 3.11.** *We may assume that there is no set  $X \subseteq V(G)$  of size  $(1 + \varepsilon/2)k \leq |X| \leq (1 + 2000\varepsilon)k$  such that the minimum degree of  $G[X]$  is at least  $(1 - 10\varepsilon)k$ , as otherwise  $G_p[X]$  a.a.s. would contain a cycle of length at least  $k + 1$ .*

The last proposition allows us to further assume from this point on that all pseudo-cliques in  $\mathcal{C}$  have size less than  $(1 + \varepsilon/2)k$ . We are ready to prove Lemma 3.4.

**Proof of Lemma 3.4.** Let  $u \in Z_1$ , and let  $d$  denote the degree of  $u$  in  $G$ . Our aim is to bound the number of neighbours of  $u$  that are in  $W$ . We remark that the following estimations for the number of neighbours of  $u$  which belong to  $W$  are true even if we drop the assumption that  $u$  is outcast.

Either  $u$  has  $\varepsilon d/6$  neighbours in  $W_1$ , or it has the same amount of neighbours in  $W_2$ . Proposition 3.5 bounds the probability of the first case to happen by at most  $1/k^4$ . For the second case, notice that Proposition 3.10 implies that  $\mathbb{P}[w \in W_2] \leq 5/k^5$  for all  $w \in \bigcup_{C \in \mathcal{C}} C$ . By Markov’s inequality, we have that

$$\mathbb{P}[|N(u) \cap W_2| > \varepsilon d/6] < \frac{30}{\varepsilon k^5} < \frac{1}{k^4}.$$

Therefore, by the union bound,  $\mathbb{P}[u \in Z_1] < 1/k^3$ , concluding the proof of the lemma.  $\square$

Lemma 3.4 bounds the number of outcast vertices that lost a significant proportion of their neighbours after the deletion of the waste from  $G$  to obtain  $G'$ . By Markov’s inequality, asymptotically almost surely, the size of  $Z_1$  is bounded by

$$|Z_1| \leq \frac{n}{k^2}. \tag{3.2}$$

This inequality tells us that the influence of the waste is not too large, so for most of our subsequent arguments, we can completely ignore the vertices from  $W$ . Also, from the definition of  $Z_1$ , if  $v \in G' \setminus Z_1$  is an outcast vertex then  $\deg_{G'}(v) \geq (1 - \varepsilon/3)k$ , and hence  $v$  still retains most of its degree after the deletion of  $W$ . However, in the final part of the proof of our main theorem, we have to use the full structure of  $G$  and incorporate the waste vertices back. Therefore, we need a lemma to state what typically happens to a pseudo-clique after we delete the vertices from the waste.

**Lemma 3.12.** Consider an arbitrary pseudo-clique  $C \in \mathcal{C}$ , and denote by  $D_1$  the set of vertices from  $C$  having more than  $\varepsilon k$  neighbours in  $G$  outside  $C$ . Let  $D'_1$  be the union of  $D_1 \cap W$  together with the vertices in  $D_1 \setminus W$  that lost more than a  $\frac{1}{100}$ -proportion of its neighbours outside  $C$  after the removal of the waste vertices from  $W$ . Furthermore, let  $D_2$  be the set of vertices not in  $C$  that have at least  $\varepsilon k$  neighbours in  $C$  in the graph  $G$ . Finally, let  $\mathcal{E}$  denote the set of edges from  $G$  connecting  $C \setminus D_1$  to a vertex not in  $C \cup D_2$ . Then a.a.s. we have

$$|\mathcal{E} \setminus E(G')| \leq |\mathcal{E}|/100, \quad |D_2 \cap W| \leq |D_2|/100 \quad \text{and} \quad |D'_1| \leq |D_1|/100. \tag{3.3}$$

Therefore, a.a.s. at least  $(1 + o(1))|\mathcal{C}|$  pseudo-cliques in  $\mathcal{C}$  satisfy (3.3).

**Sketch of proof.** For each fixed vertex  $u$ , the probability that  $W$  contains  $u$  is either zero (if  $u$  is outcast) or tiny, as the inequality (3.1) together with Proposition 3.10 imply that both  $\mathbb{P}[u \in W_1]$  and  $\mathbb{P}[u \in W_2]$  are small. Similarly, for each fixed edge  $e$ , the probability that one of its endpoints belongs to  $W$  is also very small. In expectation, we have

$\mathbb{E}[|\mathcal{E} \setminus E(G')|] = o(|\mathcal{E}|)$ , hence by Markov's inequality we know that a.s.  $|\mathcal{E} \setminus E(G')| \leq |\mathcal{E}|/100$ . Similarly, we have  $\mathbb{E}[|D_1 \cap W|] = o(|D_1|)$  and  $\mathbb{E}[|D_2 \cap W|] = o(|D_2|)$ . Moreover, for each vertex  $u \in D_1$ , if we denote by  $d$  the number of edges connecting  $u$  to a vertex outside  $C$ , and by  $d'$  the number of edges connecting  $u$  to a vertex in  $W \setminus C$ , then  $\mathbb{E}[d'] = o(d)$ . Thus, by Markov's inequality, we know that  $\mathbb{P}[v \in D'_1] = o(1)$ , and the lemma follows by another application of Markov's inequality.  $\square$

**3.2. Step 2: exploring the connected components**

Recall that at this point, some of the edges of  $G$  have already been tested in  $G_p$ , namely all the edges inside pseudo-cliques from  $\mathcal{C}$ . Let  $Q_1$  be the set of tested edges from the partially exposed  $G_p$  that live inside  $G'$ , and let  $E_1 \subseteq Q_1$  be the subset of the successfully tested edges. To find the connected components of the partially exposed  $G'_p = G_p \setminus W$  using the DFS algorithm, we adopt the following DFS priority:

Whenever the DFS reaches a vertex  $v$  from a pseudo-clique  $C \in \mathcal{C}$ , the algorithm, instead of testing new edges, walks through an already exposed Hamilton cycle in  $G_p[C \setminus W]$  using the edges from  $E_1$ , until it visits all vertices from  $C \setminus W$ .

In the rooted spanning output forest  $T$ , this Hamilton cycle forms a path, and the algorithm saved many edge tests in this way. This observation is stated more formally as follows.

**Observation 3.13.** *For every pseudo-clique  $C \in \mathcal{C}$  such that  $C \not\subseteq W$ , there exists a path in  $T$  whose vertices are precisely the vertices in  $C \setminus W$ .*

Let  $Q'_2$  be the set of tested edges, and let  $E_2 \subseteq Q'_2$  be the set of successfully tested edges in this exploration of  $G'_p$  by the DFS. Clearly  $E(T) \subseteq E_1 \cup E_2$ , and  $|E_2| < \ell + |\mathcal{C}|$ , since once we reach a pseudo-clique  $C$ , we do not need to test edges until all the vertices of  $C \setminus W$  are reached.

Next, we query all the untested edges connecting vertices from  $G'$  which have distance at least  $k + 1$  with respect to the forest  $T$ . Let  $Q''_2$  be the set of all such edges. We test the edges in  $Q''_2$  one by one, in an arbitrary order. If by chance we successfully expose one edge from  $Q''_2$ , we automatically obtain a cycle of length at least  $k + 1$  in  $G'_p$ , as desired in Theorem 1.1, and we stop the whole procedure. In particular, the total number of edges in  $Q''_2$  must be very small, say  $|Q''_2| < \varepsilon^2 k/2$ , as otherwise we would have a long cycle a.s.. Let  $Q_2$  be the union of  $Q'_2$  with the tested edges from  $Q''_2$ . We can estimate the total number of edges  $Q_2$  using Proposition 2.4 and Lemma 2.1, obtaining the next statement.

**Corollary 3.14.** *Asymptotically almost surely, we have*

$$|Q_2| \leq \frac{1.1}{p} \cdot |E_2| + \varepsilon^2 k/2 + \varepsilon^2 k/2 \leq \frac{1.2}{p} \left( \ell + \frac{n}{k} \right) + \varepsilon^2 k.$$

*Moreover, if not all the edges in  $Q''_2$  were tested at this point, then we have already exposed a cycle of length at least  $k + 1$  in  $G'_p$ .*

We would like to remark that the expression  $\ell + n/k$  is not guaranteed to tend to infinity with  $k$ , so the inequality  $|Q'_2| \leq 1.1/p \cdot |E_2|$  is not guaranteed to hold asymptotically almost surely. However, Corollary 3.14 is true because of the extra  $\varepsilon^2 k/2$  term, as

$$\mathbb{P} \left[ \text{Bin} \left( \frac{1.2}{p} \cdot \left( \ell + \frac{n}{k} \right) + \varepsilon^2 k/2, p \right) \geq \ell + \frac{n}{k} \right] \rightarrow 1 \quad \text{as } k \rightarrow \infty.$$

**3.3. Step 3: the block decomposition**

In this subsection, we apply the block algorithm to the input  $(H, H^*)$ , where  $H = T + E_1$  and  $H^*$  is the current partially exposed  $G'_p$ . Recall that  $T + E_1$  might have some large cycles already, coming from the exposed pseudo-cliques in Step 1. Thus we can bound the number of blocks of  $T + E_1$  from above by  $\ell + |\mathcal{C}|$ . This is because for every pseudo-clique  $C \in \mathcal{C}$  which is not completely inside the waste  $W$ , the vertex set  $C \setminus W$  necessarily induces a Hamiltonian graph  $(T + E_1)[C \setminus W]$ .

Let  $Q_3$  be the set of edges from  $G'_p$  tested during the execution of the block algorithm, and let  $E_3 \subseteq Q_3$  be the subset of the successfully tested edges. From Proposition 2.8, we know that the number of successfully tested edges revealed by the block algorithm is at most the number of blocks of  $T + E_1$ . Moreover, by the observation discussed in the last paragraph, we also know that the total number of blocks of  $T + E_1$  is at most  $\ell + |\mathcal{C}|$ , hence  $|E_3| \leq \ell + |\mathcal{C}|$ , and by Lemma 2.1 we have the following corollary.

**Corollary 3.15.** *Asymptotically almost surely*

$$|Q_3| \leq \frac{1.2}{p} \left( \ell + \frac{n}{k} \right) + \varepsilon^2 k.$$

We would like to draw the reader’s attention to the fact that we added the term  $\varepsilon^2 k$  to the right-hand side of the inequality in Corollary 3.15. This is because we want to make sure that the right side, when multiplied by  $p$ , tends to infinity with  $k$ . We recall that a similar ‘trick’ was used in Corollary 3.14.

**3.4. Step 4: the structure inside the blocks**

Let  $\mathcal{B}$  be the family of all blocks of the partially exposed  $G'_p$  obtained in Step 3. Here, the edges of every block  $B \in \mathcal{B}$  consist of those successfully exposed in  $G'_p$  so far, that is,  $E(B) \subseteq E_1 \cup E_2 \cup E_3$ . Moreover, the cut-vertices of  $G'_p$  are precisely the same as the cut-vertices of  $T + (E_1 \cup E_3)$ . One should also observe the following.

**Observation 3.16.** *For each  $B \in \mathcal{B}$ , the graph  $T \cap B$  is a tree. Moreover, if  $B_1$  and  $B_2$  are two distinct blocks from  $\mathcal{B}$  having a vertex  $v$  in common, then  $v$  is the smallest vertex (with respect to  $\leq_T$ ) from at least one of the two blocks  $B_1$  or  $B_2$ .*

The content of the previous observation is illustrated in Figure 2. In the picture, each connected component represents a subtree of the form  $T \cap B$  for some  $B \in \mathcal{B}$ . The dashed ovals represent the cut-vertices from  $G'_p$  (all small solid circles inside the dashed ovals actually represent the same cut-vertex).

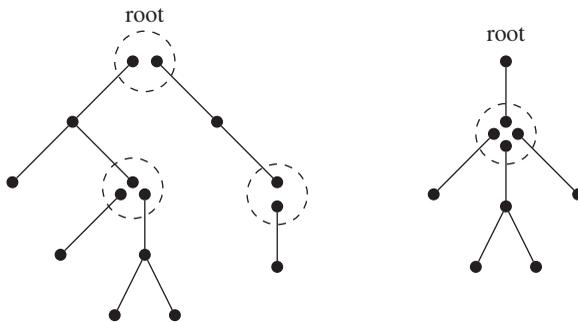


Figure 2. Blocks and cut-vertices of  $G'_p$  together with the rooted forest  $T$ .

Note that for all pseudo-cliques  $C \in \mathcal{C}$  with  $C \not\subseteq W$ , there exists a unique block  $B \in \mathcal{B}$  such that  $C \setminus W \subseteq V(B)$ . This is because cycles are 2-connected. In this case, with slight abuse of notation, we say that  $B$  contains the pseudo-clique  $C$ .

The next proposition shows that a block containing more than one pseudo-clique already has a long cycle.

**Proposition 3.17.** *If  $B \in \mathcal{B}$  is a block that contains two distinct pseudo-cliques  $C_1, C_2 \in \mathcal{C}$ , then  $B$  contains a cycle of length at least  $k + 1$ .*

Before we prove Proposition 3.17, let us prove an auxiliary statement.

**Proposition 3.18.** *Let  $C \in \mathcal{C}$  be such that  $C \not\subseteq W$ , and let  $B \in \mathcal{B}$  be the block containing  $C$ . Then for every two distinct vertices  $u, v \in C \setminus W$ , the induced graph  $B[C \setminus W]$  contains a path of length at least  $(1 - 20\epsilon)k$  connecting  $u$  to  $v$ .*

**Proof.** We want to show that there exists a path  $P$  in  $B[C \setminus W]$  connecting  $u$  and  $v$  of length at least  $(1 - 20\epsilon)k$ . Because of property (P3) stated in Section 3.1, we know that  $B[C \setminus W_1] = B[C \setminus W]$  is Hamiltonian. Let  $J$  be a Hamilton cycle in  $C \setminus W$ . Next, consider the two paths  $P_1$  and  $P_2$  obtained from the cycle  $J$  connecting the vertices  $u$  and  $v$ . Assume that  $P_2$  is no longer than  $P_1$ . By property (P1),  $J$  is of length at least  $(1 - 5\epsilon)k$ , so  $P_1$  has at least  $(1 - 5\epsilon)k/2$  vertices. If the length of  $P_1$  is greater than  $(1 - 20\epsilon)k$ , our proposition immediately follows by taking  $P := P_1$ . Otherwise the path  $P_2$  has at least  $15\epsilon k$  vertices. Let  $X$  be the set of the  $6\epsilon k$  vertices from the path  $P_1$  which are closest to the endpoint  $v$ . Similarly, let  $Y$  be the set of the  $6\epsilon k$  vertices from the path  $P_2$  which are closest to the other endpoint  $u$ . Using property (P2), we know that  $B[C \setminus W]$  has an edge  $e = xy$  connecting a vertex  $x$  from  $X$  to a vertex  $y$  from  $Y$ , as shown in Figure 3.

We can build the longer path  $P$  by patching two segments from  $P_1$  and  $P_2$  together with the edge  $e$  as follows. The initial segment of  $P$  consists of the path in  $P_1$  connecting  $u$  to  $x$ , while the final segment of  $P$  consists of the path in  $P_2$  connecting  $y$  to  $v$ , and these two segments are interconnected by  $e$ . The total length of  $P$  is at least the length of  $J$  minus  $12\epsilon k$ , hence the length of  $P$  is at least  $|J| - 12\epsilon k > (1 - 20\epsilon)k$ , finishing the proof of the proposition. □

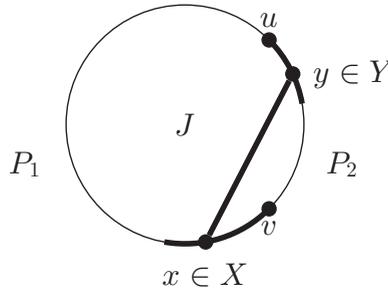


Figure 3. Rotation inside pseudo-cliques.

We are ready to prove Proposition 3.17.

**Proof of Proposition 3.17.** In the proof of this proposition, we use Theorem 2.5 to merge long cycles. Since  $B$  is 2-connected, Theorem 2.5 asserts the existence of two vertex-disjoint paths  $P_1, P_2$  in  $B$  connecting  $C_1 \setminus W$  to  $C_2 \setminus W$ . Let  $u_1, v_1$  be the endpoints of  $P_1, P_2$  (respectively) in  $C_1 \setminus W$ . Similarly, let  $u_2, v_2$  be the endpoints of  $P_1, P_2$  in  $C_2 \setminus W$ .

By Proposition 3.18, we can obtain two paths  $P_3, P_4$  both having length at least  $(1 - 20\epsilon)k$ , where  $P_3$  is a path in  $C_1 \setminus W$  connecting  $u_1$  to  $v_1$ , and  $P_4$  is a path in  $C_2 \setminus W$  connecting  $u_2$  to  $v_2$ . By patching together  $P_1, P_3, P_2$ , and  $P_4$  in that order, we obtain a cycle of total length larger than  $(2 - 40\epsilon)k > k + 1$ , thereby proving the proposition.  $\square$

As we have seen in the proof Proposition 3.17, we can use pseudo-cliques to obtain long cycles, which can then be merged into even longer cycles. We do not need to use the full strength of pseudo-cliques in order to merge cycles. In the proof of Proposition 3.18, the edge  $e$  played an important role, as it allowed us to ‘rotate’ inside the relatively long cycle. In what follows, we describe a weaker structure that also allows this ‘rotation’ operation. We say that a cycle  $J$ , formed by some of the successfully exposed edges from a partially exposed  $G'_p$ , is a *rotating cycle* if all properties below hold simultaneously:

- (P1 $\star$ )  $J$  has at least  $(1 - 4\epsilon)k$ , but at most  $k$  vertices,
- (P2 $\star$ ) all but one edge of  $J$  belong to the forest  $T$  revealed in Step 2,
- (P3 $\star$ ) if  $u \in V(J)$  is the largest vertex with respect to the order  $\leq_T$  (we call  $u$  the *pivot* of  $J$ ), then there exist at least  $(1 - 4\epsilon)k$  untested edges in the partially exposed  $G'_p$  connecting  $u$  to another vertex of  $J$ .

The properties listed above bear some resemblance to those enumerated in Section 3.1. For instance (P1) and (P1 $\star$ ) both state some bounds about the size of the structure under consideration. Property (P2 $\star$ ) might look somewhat artificial at first, but we observe that for every pseudo-clique  $C \in \mathcal{C}$  such that  $C \not\subseteq W$ , the graph  $G'_p[C \setminus W]$  contains a Hamilton path that is entirely contained in  $T$ , a consequence of the priority of the DFS remarked in Section 3.2. Finally, (P3 $\star$ ) is the property that will allow us to perform the rotation *per se*, and note that (P3 $\star$ ) clearly implies the lower bound of the length of  $J$  in (P1 $\star$ ).

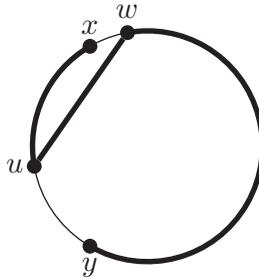


Figure 4. Rotation inside the rotating cycle.

The next proposition describes the operation of rotation, which is similar to the rotation described in Proposition 3.18.

**Proposition 3.19.** *Let  $J$  be a rotating cycle with pivot  $u \in V(J)$ , and fix any two distinct vertices  $x, y \in V(J) \setminus \{u\}$ . After exposing the untested edges connecting  $u$  to the other vertices in  $J$ , a.a.s. we can find a path in  $G'_p[V(J)]$  between  $x$  and  $y$  of length at least  $(2 - 10\epsilon)k/3$ .*

**Proof.** Let  $P_1$  and  $P_2$  be the two paths between  $x$  and  $y$  obtained from the cycle  $J$ , with lengths  $l_1$  and  $l_2$ , respectively. Assume, without loss of generality, that  $u \in V(P_1)$ , and that the distance from  $u$  to  $y$  is no larger than the distance from  $u$  to  $x$  in the path  $P_1$ . Let  $N$  be the set of all vertices  $w$  of  $J$  such that  $uw$  is an untested edge of  $G'_p$ , and let  $N_1 = N \cap P_1$  and  $N_2 = N \cap P_2$ . If either  $N_1$  or  $N_2$  has size at least  $(2 - 10\epsilon)k/3$  then we are done, since  $l_1 \geq |N_1|$  and  $l_2 \geq |N_2|$ . Otherwise, both  $N_1$  and  $N_2$  have size at least  $(1 - 2\epsilon)k/3$ , because (P3\*) implies that  $|N| \geq (1 - 4\epsilon)k$ . Next, we test all the edges connecting  $u$  to the  $\epsilon k$  vertices from  $N_2$  which are closest to  $x$  with respect to the path  $P_2$ . This is possible because  $|N_2| > \epsilon k$ . Asymptotically almost surely, we can find a successfully exposed edge  $e = uw$  where  $w$  belongs to this subset of  $N_2$  of size  $\epsilon k$ .

We can then obtain a path  $P$  as follows. We use the segment from  $P_1$  connecting  $x$  to  $u$ , and then we traverse the edge  $e$ , to reach the vertex  $w$ , and then use the segment from  $P_2$  connecting  $w$  to  $y$  as illustrated in Figure 4. The length of  $P$  is at least

$$l_1/2 + |N_2| - \epsilon k \geq |N_1|/2 + |N_2| - \epsilon k = (|N| + |N_2|)/2 - \epsilon k \geq (2 - 10\epsilon)k/3,$$

concluding the proof. □

Analogous to pseudo-cliques, rotating cycles are also somewhat tied to the block structure of  $G'_p$ . For every rotating cycle  $J$ , there exists a unique block  $B \in \mathcal{B}$  such that  $V(J) \subseteq V(B)$ . The equivalent of Proposition 3.17 for rotating cycles is the next statement.

**Proposition 3.20.** *Suppose the partially exposed  $G'_p$  contains two vertex-disjoint rotating cycles  $J_1$  and  $J_2$  whose vertices are contained in the same block  $B \in \mathcal{B}$ . Then after we expose the remaining untested edges of  $G'_p$ , a.a.s. we can find a cycle of length at least  $k + 1$  in  $G'_p[V(B)]$ .*

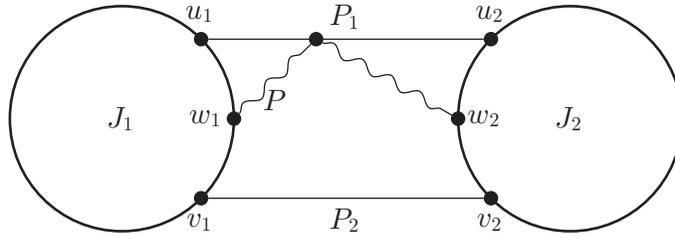


Figure 5. Merging disjoint cycles.

**Proof.** Here, we again use Theorem 2.5 to merge cycles. Let  $P$  be the path in  $T$  connecting  $J_1$  to  $J_2$ . This path exists and is unique because  $T \cap B$  is a tree. Let  $w_1$  be the endpoint of  $P$  in  $J_1$  and let  $w_2$  be the other endpoint of  $P$  in  $J_2$ . We may assume, without loss of generality, that  $w_1$  is the smallest vertex with respect to  $\leq_T$  in  $J_1$ . To see why this assumption can be made, observe that (P2\*) implies that both  $V(J_1)$  and  $V(J_2)$  induce paths in  $T \cap B$ .

Since  $B$  is a 2-connected graph, we can use the edges of  $B$  to obtain two vertex-disjoint paths  $P_1, P_2$  connecting  $J_1$  to  $J_2$ . Let  $u_1, v_1$  be the endpoints of  $P_1, P_2$  in  $J_1$ , respectively. Similarly, let  $u_2, v_2$  be the endpoints of  $P_1, P_2$  in  $J_2$ . If neither  $u_1$  nor  $v_1$  is the pivot of  $J_1$ , then we can obtain the long cycle in the following way. By using Proposition 3.19, we obtain a.s.a. a path  $P_3$  of length at least  $(2 - 10\epsilon)k/3$  between  $u_1$  and  $v_1$  in  $G'_p[V(J_1)]$ , and clearly there exists a path  $P_4$  of length at least  $(1 - 4\epsilon)k/2$  between  $u_2$  and  $v_2$  in  $J_2$  (just take the longest of the two paths connecting  $u_2$  to  $v_2$  in the cycle  $J_2$ ). Putting together  $P_1, P_4, P_2$ , and  $P_3$  in that order, we obtain a cycle of length at least  $(7 - 32\epsilon)k/6 > k + 1$ , thereby proving the proposition.

Otherwise, assume without loss of generality that  $u_1$  is the pivot of  $J_1$ . One of the edges in the cycle  $J_1$  connects  $u_1$  to  $w_1$  (recall that  $u_1$  is the largest vertex with respect to  $\leq_T$ , while  $w_1$  is the smallest). The idea now is to modify one of the paths  $P_1$  or  $P_2$  so that either the endpoint of  $P_1$  in  $J_1$  is no longer  $u_1$ , or the endpoint of  $P_2$  is no longer  $v_1$ , but  $w_1$  instead. To do this, we follow the path  $P$  from  $w_1$  to  $w_2$ , until it hits  $P_1, P_2$ , or  $J_2$ . If  $P$  hits  $P_1$  first, we replace the initial segment of  $P_1$  with the initial segment of  $P$  as illustrated in Figure 5. If  $P$  hits  $P_2$  first, we modify  $P_2$  similarly. Otherwise,  $P$  never hits  $P_1$  or  $P_2$ , so we can just replace the whole path  $P_2$  by  $P$ .

If  $P_1$  was modified, we can use the ideas described in the second paragraph of this proof to obtain the long cycle asymptotically almost surely. Otherwise, if  $P_2$  was modified, we can use a path connecting  $u_1$  to  $w_1$  in  $J_1$  that uses all of its vertices. In this way we obtain a cycle of length at least  $3(1 - 4\epsilon)k/2 > k + 1$ , and we did not need to test any edge for this case. □

Proposition 3.20 dealt with the case of vertex-disjoint rotating cycles. But what if the cycles intersect? The next proposition shows that even if the intersection is not empty, it is still possible to merge the rotating cycles, provided that their intersection is not too large.

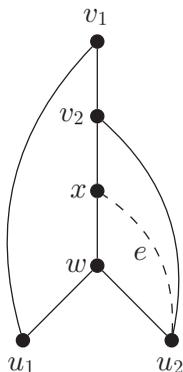


Figure 6. Merging intersecting cycles.

**Proposition 3.21.** *Assume the partially exposed  $G'_p$  contains two intersecting rotating cycles  $J_1$  and  $J_2$  contained in the same block  $B \in \mathcal{B}$  whose intersection  $J_1 \cap J_2$  has at most  $(1 - 15\epsilon)k$  vertices. Then after we expose the remaining untested edges of  $G'_p$ , a.s. we can find a cycle of length at least  $k + 1$  in  $G'_p[V(B)]$ .*

**Proof.** Let  $u_1$  and  $u_2$  be the pivots of  $J_1$  and  $J_2$ , respectively. Since  $|J_1 \cap J_2| \leq (1 - 15\epsilon)k$ , we must necessarily have  $u_1 \neq u_2$ , as shown in Figure 6. In fact, if  $u_1 = u_2$  then either  $V(J_1) \subseteq V(J_2)$  or  $V(J_2) \subseteq V(J_1)$ , and hence we would have

$$|J_1 \cap J_2| = \min\{|J_1|, |J_2|\} > (1 - 15\epsilon)k,$$

which is a contradiction. Moreover, let  $v_1$  and  $v_2$  be the smallest vertices in  $J_1$  and  $J_2$  respectively, with respect to the order  $\leq_T$ . Furthermore, let  $w$  be the largest vertex in  $J_1 \cap J_2$  with respect to the same order. We must have either  $v_1 \leq_T v_2$  or  $v_2 \leq_T v_1$ , because otherwise (P2\*) would imply that  $J_1$  and  $J_2$  are disjoint. Assume  $v_1 \leq_T v_2$ . The intersection  $J_1 \cap J_2$  comprises the path in  $T$  joining  $v_2$  to  $w$ . We divide the remainder of the proof into two cases.

In the first case we have  $|J_1 \cap J_2| < 100\epsilon k$ . We can obtain a long cycle  $P$  as follows: we start at  $v_1$ , traverse the edge to  $u_1$ , walk the path in  $J_1$  from  $u_1$  to  $w$  (we choose the path that does not contain  $v_1$ ), then walk the path in  $J_2$  from  $w$  to  $u_2$  (again choosing the path that does not contain  $v_2$ ), move to  $v_2$  using an edge from  $J_2$ , and finish the cycle with the path from  $v_2$  to  $v_1$  in  $J_1$ . The length of  $P$  is at least

$$|J_1| + |J_2| - 2|J_1 \cap J_2| > k + 1,$$

and we are done.

In the second case, we have  $|J_1 \cap J_2| \geq 100\epsilon k$ . Let  $X$  be the set of the  $\epsilon k$  vertices  $x$  in the path  $J_1 \cap J_2$  which are closest to  $w$  such that  $xu_2$  is an untested edge. Such a set  $X$  exists because of (P3\*) and  $|J_1 \cap J_2| \geq 100\epsilon k$ . Observe that no vertex in  $X$  is more than  $4\epsilon k + \epsilon k$  vertices away from  $w$ , as (P3\*) implies. Next, we expose the untested edges joining  $u_2$  to a vertex in  $X$ . Asymptotically almost surely we can find a successfully exposed edge  $e = xu_2$ . We can now obtain a long cycle  $P$  in a way very similar to what we did before: we start at

$v_1$ , traverse the edge to  $u_1$ , walk the path in  $J_1$  from  $u_1$  to  $w$ , then walk the path in  $J_2$  from  $w$  to  $u_2$ , move to  $x$  using the edge  $e$  that we recently exposed, and finish the cycle with the path from  $x$  to  $v_1$  in  $J_1$ . The length of  $P$  is at least  $|J_1| + |J_2| - |J_1 \cap J_2| - 5\epsilon k > k + 1$ , and we are done.  $\square$

So far, we only have analysed the cases where the block has either two pseudo-cliques or two rotating cycles. To conclude this subsection, we state a proposition that handles the case when these two different structures are mixed together in the same block.

**Proposition 3.22.** *Let  $J$  be a rotating cycle of the partially exposed  $G'_p$ , and let  $B \in \mathcal{B}$  be the unique block containing  $J$ . Assume that  $B$  contains a pseudo-clique  $C \in \mathcal{C}$ , and that the intersection  $V(J) \cap C$  has at most  $(1 - 30\epsilon)k$  vertices. Then after we expose the remaining untested edges of  $G'_p$ , a.a.s. we can find a cycle of length at least  $k + 1$  in  $G'_p[V(B)]$ .*

**Proof.** The proof is identical to the previous proof and is therefore omitted.  $\square$

### 3.5. Step 5: double-counting the poor and the full

In this subsection we study the rotating cycles of  $G'_p$ . For that purpose, we assume that the statements of Proposition 3.11, Lemma 3.12, Corollary 3.14 and Corollary 3.15 hold. We further assume that (3.2) holds, and all the edges from  $Q'_2$  were tested (as otherwise we would already have a long cycle). Thus, the reader should bear in mind that any probabilistic statement in this subsection should be conditioned on the event that all these assumptions hold.

Let  $U$  be the set of all untested edges from  $G'_p$  so far. More precisely, let  $U = E(G') \setminus (Q_1 \cup Q_2 \cup Q_3)$ . In the next few paragraphs, we adopt some definitions motivated by the work of Riordan [16]. We say that a vertex  $v$  in a block  $B \in \mathcal{B}$  is *poor* in  $B$  if the number of descendants (with respect to  $T$ ) of  $v$  in  $B$  is at most  $\epsilon k$ . Otherwise we say that  $v$  is *rich* in  $B$ . Observe that every rich vertex in a block  $B$  has at least  $\epsilon k$  poor descendants in  $B$ . Finally, we say that a vertex  $v \in V(B)$  is *full* in  $B$  if the number of vertices  $u \in V(B)$  such that  $uv \in U$  is at least  $(1 - \epsilon)k$ .

Observation 3.13 stated that the vertices of a pseudo-clique induce a path in the rooted forest  $T$ . A consequence of this fact is the following observation.

**Observation 3.23.** *Let  $C \in \mathcal{C}$  be a pseudo-clique satisfying  $C \not\subseteq W$ , and let  $B \in \mathcal{B}$  be the unique block containing  $C$ . The total number of vertices in  $C$  which are poor in  $B$  is at most  $\epsilon k$ .*

The next proposition shows how to obtain a rotating cycle from full vertices.

**Proposition 3.24.** *Let  $v$  be a full vertex in  $B \in \mathcal{B}$  such that the number of edges  $vu \in U$  for which  $u \in V(B)$  is a descendant of  $v$  with respect to  $T \cap B$  is at most  $\epsilon k$ . Then by testing some of the untested edges incident to  $v$ , a.a.s. we can obtain a rotating cycle with pivot  $v$ .*

**Proof.** We would like to remind the reader that we are using the fact that all the edges of  $Q_2''$  were tested (the edges of  $Q_2''$  connect vertices at distance greater than  $k$  with respect to  $T$ ), as assumed in the beginning of this subsection. Let  $X$  be the set of all vertices  $u \in V(B)$ , such that  $vu \in U$ . Because of the property of DFS forests stated in Proposition 2.3, we know that for each  $u \in X$ ,  $u$  is either a descendant or an ancestor of  $v$ . By the hypothesis of the proposition, the set  $X$  has at least  $(1 - 2\varepsilon)k$  ancestors of  $v$ . We also know that none of the vertices in  $X$  have distance more than  $k$  to  $v$  with respect to the tree  $T \cap B$ . Let  $Y \subseteq X$  consist of the  $\varepsilon k$  vertices in  $X$  which are ancestors of  $v$  and are as far from  $v$  as possible, with respect to the same distance on the tree  $T \cap B$ . Asymptotically almost surely, if we test the edges of  $U$  connecting  $v$  to vertices in  $Y$ , we obtain a successfully tested edge  $e = vu$ . We claim that the path from  $v$  to  $u$  in the tree  $T \cap B$  together with the edge  $e$  forms a rotating cycle  $J$  with pivot  $v$ . This assertion is clear, as one can immediately verify that properties (P1\*), (P2\*), and (P3\*) hold.  $\square$

The careful reader will notice that the conditions of Proposition 3.24 are trivially satisfied when  $v$  is full and poor, hence we have the following corollary.

**Corollary 3.25.** *Let  $v$  be a full poor vertex in  $B \in \mathcal{B}$ . Then by testing some of the untested edges incident to  $v$ , a.a.s. we can obtain a rotating cycle with pivot  $v$ .*

We turn to identify the set of full vertices in the blocks of  $\mathcal{B}$ . Let  $Z_2$  be the set of all vertices  $v$  from  $G' \setminus Z_1$  such that  $v$  is incident to at least  $\varepsilon k/3$  tested edges from  $Q_2 \cup Q_3$ . From our assumptions at the beginning of this subsection, more specifically from Corollaries 3.14 and 3.15, we have

$$|Z_2| \leq \frac{15}{\varepsilon pk} \cdot \left( \ell + \frac{n}{k} \right) + 12\varepsilon. \tag{3.4}$$

To avoid future issues with double-counting arguments, we would like to identify the set of vertices  $v \in V(G') \setminus (Z_1 \cup Z_2)$ , such that there exists a unique block  $B \in \mathcal{B}$  for which all but at most  $\varepsilon k$  neighbours of  $v$  in  $G'$  belong to  $V(B)$ . If  $v$  is not a cut-vertex of  $G'_p$ , this is trivial (recall that the cut-vertices of  $G'_p$  are precisely the cut-vertices of  $T + (E_1 \cup E_3)$ ). Otherwise, let  $Z_3$  be the set of cut-vertices  $v$  from  $G'_p$  not in  $Z_1 \cup Z_2$  such that  $v$  is the smallest (with respect to the order  $\leq_T$ ) of a block containing a pseudo-clique from  $\mathcal{C}$ . Moreover, let  $Z_4$  be the set of all cut-vertices  $v$  from  $G'_p$  not in  $Z_1 \cup Z_2 \cup Z_3$ , such that there are at least  $\varepsilon k/3$  edges  $vw$  in  $U$  for which  $v$  is the smallest vertex in the unique block that contains both  $v$  and  $w$ . We claim the following.

**Proposition 3.26.**  $|Z_3| \leq |\mathcal{C}|$  and

$$|Z_4| \leq \frac{3(\ell + |\mathcal{C}|)}{\varepsilon k}.$$

**Proof.** To prove  $|Z_3| \leq |\mathcal{C}|$ , note that for each pseudo-clique  $C \in \mathcal{C}$  there exists a unique block  $B \in \mathcal{B}$  such that  $B$  contains  $C$ . Moreover, there is a unique vertex  $v$  which is the

smallest vertex of  $B$  with respect to the order  $\leq_T$ . The map given by  $C \mapsto v$  covers every vertex from  $Z_3$ , hence  $|Z_3| \leq |C|$ .

To prove the other inequality, observe that if  $vw \in U$ , where  $v \in Z_4$  and  $w$  is a vertex that belongs to a block  $B$  where  $v$  is the smallest vertex, then either  $w$  is an outcast vertex, or there exists  $C \in \mathcal{C}$  such that  $w \in C$ . But if  $w$  belongs to the pseudo-clique  $C$ , we claim that  $w$  must be the smallest vertex in the unique block that contains  $C$ . To see this, first observe that since  $v \notin Z_3$ ,  $B$  does not contain  $C$ . Let  $B' \neq B$  be the unique block containing  $C$ . By Observation 3.16,  $w \in V(B) \cap V(B')$  must be the smallest vertex of either  $B$  or  $B'$ . But because  $v$  is the smallest vertex from  $B$ , we infer that  $w$  is the smallest vertex from  $B'$ , proving our claim. Since  $w$  is either outcast or the smallest vertex of a block that contains a pseudo-clique, we must conclude that there are at most  $\ell + |C|$  different choices for  $w$ .

We claim that for every vertex  $w \in V(G')$ , there is at most one edge in  $U$  connecting  $w$  to an ancestor of  $w$  which is the smallest vertex of some block. To prove this claim, suppose towards contradiction that there exist two such edges  $wx_1$  and  $wx_2$ . Let  $B_1$  and  $B_2$  be the corresponding blocks containing  $wx_1$  and  $wx_2$ , respectively. The intersection of the blocks  $B_1$  and  $B_2$  contains  $w$ , but  $w$  is not the smallest in either of them, contradicting Observation 3.16, and proving our second claim.

Therefore, there are at most  $\ell + |C|$  edges  $vw \in U$  such that  $v \in Z_4$  and  $v$  is the smallest vertex in the unique block containing both  $v$  and  $w$ . This immediately implies that

$$|Z_4| \leq \frac{3(\ell + |C|)}{\varepsilon k},$$

concluding the proof of the proposition. □

Let  $Z = Z_1 \cup Z_2 \cup Z_3 \cup Z_4$ . Combining (3.2), (3.4) and Proposition 3.26, we obtain that a.a.s.

$$|Z| \leq \frac{1.05n}{k} + \frac{16}{\varepsilon pk} \cdot \left( \ell + \frac{n}{k} \right) + 12\varepsilon. \tag{3.5}$$

**Proposition 3.27.** *Let  $v \in V(G')$  be an outcast vertex such that  $v \notin Z$ . Then there exists a unique block  $B \in \mathcal{B}$  such that  $v$  is full in  $B$ . Moreover, for any other block  $B' \in \mathcal{B}$  such that  $v \in V(B')$  and  $B' \neq B$ ,  $v$  is necessarily the smallest vertex in  $B'$  with respect to the order  $\leq_T$ .*

**Proof.** Since  $v$  is an outcast vertex and  $v \notin Z_1$ , we know that

$$\deg_{G'}(v) \geq \left(1 - \frac{\varepsilon}{3}\right) \deg_G(v) \geq \left(1 - \frac{\varepsilon}{3}\right)k.$$

Moreover, because  $v \notin Z_2$ , at most  $\varepsilon k/3$  edges from  $Q_2 \cup Q_3$  are incident to  $v$ . Hence at least  $(1 - 2\varepsilon/3)k$  edges from  $U$  are incident to  $v$ . Now we split the analysis into two cases.

In the first case,  $v$  is not a cut-vertex from  $G'_p$ . Then there exists a unique block  $B \in \mathcal{B}$  such that  $v \in V(B)$ . Clearly all the edges in  $U$  incident to  $v$  are of the form  $vu$ , for some

$u \in V(B)$ . Thus  $v$  is full in  $B$ , and  $v$  does not belong to any other block, concluding the analysis in this case.

In the second case,  $v$  is a cut-vertex from  $G'_p$ . We claim that there exists a unique block  $B \in \mathcal{B}$  such that  $v$  is not the smallest element in  $B$  with respect to the order  $\leq_T$ . To see this, observe that if  $v$  is the smallest vertex in every block in which it belongs, then  $v$  must be a root of the rooted forest  $T$ , and hence  $v \in Z_3 \cup Z_4$ , as  $v$  is incident to more than  $ek/3$  edges from  $U$ . But this is not the case, so there exists at least one block  $B$  such that  $v \in V(B)$  and  $v$  is not the smallest vertex in  $B$ . By Observation 3.16 we know that such a  $B$  must be unique. Thus, for all edges in  $U$  of the form  $vu$ , where  $u \notin V(B)$ , the vertex  $v$  must necessarily be the smallest in the unique block that contains both  $u$  and  $v$ . But because  $v \notin Z_3 \cup Z_4$ , there can be at most  $ek/3$  of such edges, so  $v$  is full in  $B$ , concluding the proof of the proposition.  $\square$

One immediate consequence of Proposition 3.27 is the following corollary.

**Corollary 3.28.** *Let  $v \in V(B) \setminus Z$  be a vertex which is not the smallest in  $B \in \mathcal{B}$  with respect to  $\leq_T$ . Then either  $v$  is full in  $B$  or there exists a pseudo-clique  $C$  such that  $v \in C$  and  $B$  contains  $C$ .*

**Proof.** If  $v$  is an outcast vertex, then since  $v$  is not the smallest of  $B$  and  $v \notin Z$ , Proposition 3.27 implies that  $v$  must be full in  $B$ . Otherwise there exists a pseudo-clique  $C \in \mathcal{C}$  such that  $v \in C$ . We claim that  $B$  contains  $C$ . Suppose for a contradiction that  $B$  does not contain  $C$ . Let  $B'$  be the unique block containing  $C$ . By Observation 3.16,  $v$  must be the smallest vertex of  $B'$ , since  $v$  is not the smallest vertex in  $B$ . Therefore  $v$  is the smallest vertex of the block  $B'$  which contains the pseudo-clique  $C$ , hence  $v \in Z_3 \subseteq Z$ , a contradiction, concluding the proof of the corollary.  $\square$

Finally, we turn to the analysis of the poor vertices in the blocks of  $\mathcal{B}$ . We say that a block  $B \in \mathcal{B}$  is *good* if the proportion of vertices from  $Z$  in  $B$  is at most  $\varepsilon/10^6$ . We have the following.

**Lemma 3.29.** *If the proportion of poor vertices inside a good block  $B$  is at most  $\varepsilon/1000$ , then after we expose the remaining untested edges of the partially exposed  $G'_p$ , a.a.s.  $G'_p[V(B)]$  contains a cycle of length at least  $k + 1$ .*

Before we prove Lemma 3.29 we need to prove some auxiliary results. For that purpose, let us introduce additional notation.

For each vertex  $v \in V(B)$ , let  $D(v)$  denote the set of all vertices  $u \in V(B)$  which are descendants of  $v$  (recall that every  $v$  is a descendant of itself) and let  $A(v)$  denote the set of all vertices  $u \in V(B)$  which are ancestors of  $v$  with respect to the tree  $T \cap B$ . The block  $B$  should be clear from the context whenever we use the notation for ancestors and descendants. We shall add the subscript ' $\leq d$ ' to either  $D(v)$  or  $A(v)$ , as in the expression  $D_{\leq d}(v)$ , to refer to the subset obtained by keeping the vertices at distance at most  $d$  from  $v$  with respect to the same tree  $T \cap B$ . Similarly, we add the superscript ' $(p)$ '/' $(r)$ ' to select

only the poor/rich vertices of the indicated set in the notation, such as in the expression  $D^{(p)}(v)$ .

We say that a vertex  $v \in V(B)$  is *branching* if there exist at least two distinct rich vertices  $u_1, u_2 \in V(B)$  which are immediate descendants of  $v$  with respect to  $T$ . Similarly to what we did previously, we reserve the superscript ‘(b)’ to denote the branching vertices of the set under consideration. We claim the following.

**Proposition 3.30.** *For each  $v$  such that  $D^{(b)}(v) \neq \emptyset$ , we have  $|D^{(p)}(v)| \geq \varepsilon k(|D^{(b)}(v)| + 1)$ .*

**Proof.** Assume  $D(v)$  contains at least one branching vertex. In this case  $v$  must be rich. Let  $T'$  be the subtree of  $T \cap B$  containing all the rich descendants of  $v$  (including itself). Since every branching vertex in  $D^{(b)}(v)$  has degree at least 3 in  $T'$  (except possibly the root  $v$ ), the number of leaves in  $T'$  is at least  $|D^{(b)}(v)| + 1$ . But every leaf of  $T'$  contains at least  $\varepsilon k$  poor descendants in  $T \cap B$ , thereby proving the proposition.  $\square$

Proposition 3.30 yields an upper bound on the total number of branching vertices in a block, namely it is at most  $1/\varepsilon k$  times the number of poor vertices in the same block.

The next proposition allow us to find a structure that resembles a path with a small number of ‘pendant’ vertices in a block with very few poor vertices.

**Proposition 3.31.** *If the proportion of poor vertices inside a good block  $B$  is at most  $\varepsilon/1000$  then there exist two vertices  $u, v \in V(B)$  such that  $u$  is a descendant of  $v$  at distance  $30k$  with respect to the tree  $T \cap B$ , and*

- (1) *the number of vertices in  $D(v) \setminus D(u)$  is at most  $(30 + \varepsilon/10)k$ , and*
- (2) *the number of vertices in  $(D(v) \setminus D(u)) \cap Z$  is at most  $\varepsilon k/10$ .*

**Proof.** Let

$$q(x) = \frac{1}{\varepsilon} \cdot |D_{\leq d}^{(p)}(x)| + k \cdot |D_{\leq d}^{(b)}(x)| + \frac{1}{\varepsilon} \cdot |D_{\leq d}(x) \cap Z|,$$

where  $d := 40k$ . We would like to estimate  $q := \sum_{x \text{ rich}} q(x)$ . We have

$$\begin{aligned} q &= \sum_{x \text{ rich}} \left( \sum_{y \in D_{\leq d}^{(p)}(x)} 1/\varepsilon + \sum_{y \in D_{\leq d}^{(b)}(x)} k + \sum_{y \in Z \cap D_{\leq d}(x)} 1/\varepsilon \right) \\ &= \sum_{y \text{ poor}} \frac{1}{\varepsilon} \cdot |A_{\leq d}^{(r)}(y)| + \sum_{y \text{ branching}} k \cdot |A_{\leq d}^{(r)}(y)| + \sum_{y \in Z} \frac{1}{\varepsilon} \cdot |A_{\leq d}^{(r)}(y)| \\ &\leq \sum_{y \text{ poor}} \frac{d}{\varepsilon} + \sum_{y \text{ branching}} k \cdot d + \sum_{y \in Z} \frac{d}{\varepsilon} \\ &\leq \frac{d}{\varepsilon} \cdot \frac{\varepsilon}{1000} \cdot |B| + k \cdot d \cdot \frac{1}{\varepsilon k} \cdot \frac{\varepsilon}{1000} \cdot |B| + \frac{d}{\varepsilon} \cdot \frac{\varepsilon}{10^6} \cdot |B| < \frac{d}{450} \cdot |B|, \end{aligned}$$

where for the penultimate inequality we used Proposition 3.30 to estimate the number of branching vertices. Note that all sums are taken over vertices in  $B$ . By averaging, there

exists a rich vertex  $v \in V(B)$  such that

$$q(v) \leq \frac{d}{450} |B| / ((1 - \varepsilon/1000)|B|) < d/400 = k/10.$$

In particular, we must have  $D_{\leq d}^{(b)}(v) = \emptyset$ ,  $|D_{\leq d}^{(p)}(v)| < \varepsilon k/10$  and  $|D_{\leq d}(v) \cap Z| < \varepsilon k/10$ . Let  $d' = 30k$ . We claim that for each rich vertex  $x \in D_{\leq d'}(v)$ , there exists exactly one rich vertex  $x'$  which is an immediate descendant of  $x$  with respect to  $T \cap B$ . Clearly there are no two such vertices  $x'$ , since otherwise  $x$  would be branching, and this cannot happen because  $D_{\leq d}^{(b)}(v) = \emptyset$ . To finish the proof of the claim, notice that if all the immediate descendants of  $x$  were poor, then

$$D(x) = \{x\} \cup D^{(p)}(x) = \{x\} \cup D_{\leq \varepsilon k}^{(p)}(x),$$

which together with  $d > d' + \varepsilon k$  implies that  $|D_{\leq d}^{(p)}(v)| \geq |D_{\leq \varepsilon k}^{(p)}(x)| \geq \varepsilon k$ , a contradiction.

By the claim we proved in the previous paragraph, we know that the set  $D_{\leq d'}^{(r)}(v)$  induces a path in  $T$ . Let  $u$  be the unique rich vertex in  $D(v)$  at distance exactly  $d'$ . We claim that the pair  $u, v$  satisfies the conditions stated in the proposition. For the first condition, observe that

$$D(v) \setminus D(u) \subseteq D_{\leq d'}^{(r)}(v) \cup D_{\leq d}^{(p)}(v),$$

hence clearly

$$|D(v) \setminus D(u)| < d' + \varepsilon k/10.$$

For the second condition, we have that

$$Z \cap (D(v) \setminus D(u)) \subseteq D_{\leq d}(v) \cap Z,$$

hence

$$|Z \cap (D(v) \setminus D(u))| < \varepsilon k/10,$$

finishing the proof of the proposition. □

We have the necessary tools to prove Lemma 3.29.

**Proof of Lemma 3.29.** We assume, without loss of generality, that  $B$  contains at most one pseudo-clique from  $\mathcal{C}$ , as otherwise Proposition 3.17 would already imply the conclusion of this lemma.

We start the proof by applying Proposition 3.31 to  $B$ , thus obtaining the pair  $u, v$ . Let  $P$  be the path between  $u$  and  $v$  in  $T$ . We have  $|P| = 30k$ , the number of vertices in  $V(P) \cap Z$  is at most  $\varepsilon k/10$ , and the number of ‘pendant’ vertices from  $P$  is at most  $\varepsilon k/10$ . In particular, for each vertex  $w \in V(P)$  at distance at least  $k$  from  $u$  with respect to  $P$ , there are at most  $\varepsilon k/10$  edges in  $U$  from  $w$  to one of its descendants not in  $P$ .

We redefine  $P$  to be the subpath of length  $28k$  obtained by removing the two segments of length  $k$  closest to the two endpoints from the original path. For each vertex  $x \in V(P) \setminus Z$ , we know (by Corollary 3.28) that either  $x$  is full in  $B$ , or  $B$  contains a pseudo-clique  $C \in \mathcal{C}$  such that  $x \in C$ . But we know, by our initial assumption, that there is at most one



Figure 7. The long cycle  $J$  formed by the solid lines.

such  $C$ , and if it exists, then  $V(P) \cap C$  should be a segment of  $P$  (because pseudo-cliques induce paths in  $T$ , see Observation 3.13). Hence, we can always find in  $P$  two disjoint segments  $L_1, L_2$ , each of length  $8k$ , such that for every  $x \in (V(L_1) \cup V(L_2)) \setminus Z$ ,  $x$  is full in  $B$  and the distance between these segments along the path  $P$  is at least  $k$ . In other words, almost all vertices from  $L_1 \cup L_2$  are full in  $B$ .

Let us divide the rest of the proof into two cases. In the first case, we assume that there exist two vertices  $x_1 \in L_1$  and  $x_2 \in L_2$ , both full in  $B$ , such that for each  $i \in \{1, 2\}$ , there are at most  $\varepsilon k$  descendants  $w$  of  $x_i$  in  $B$  for which  $w x_i$  is an edge in  $U$ . By Proposition 3.24, then a.a.s. we can obtain two rotating cycles, with pivots  $x_1$  and  $x_2$  and by Proposition 3.20 we can merge these two disjoint rotating cycles and obtain the desired long cycle, proving the lemma in this first case.

In the second case, we assume that there is no such pair of vertices  $x_1, x_2$ . Hence we might also assume that, without loss of generality, for each full vertex  $x$  in  $L_1$ , there exist at least  $\varepsilon k$  descendants  $y$  of  $x$  for which  $x y \in U$ . Out of these descendants, at most  $\varepsilon k/10$  do not belong to  $P$  (recall that the number of ‘pendant’ vertices from  $P$  is at most  $\varepsilon k/10$ ). Thus  $x$  sends at least  $9\varepsilon k/10$  untested edges to its descendants in  $P$ . Furthermore, at most  $4\varepsilon k/5$  of these descendants are of distance at most  $4\varepsilon k/5$ , thus there are at least  $\varepsilon k/10$  edges in  $U$  connecting  $x$  to one of its descendants in  $P$  at distance at least  $4\varepsilon k/5$  from  $x$ . Observe that we can a.a.s. obtain a cycle of length at least  $4\varepsilon k/5$  by testing the  $\varepsilon k/10$  edges in  $U$  going from a full vertex to its descendants on the path  $P$ . The key idea in what comes next is to merge  $O(1/\varepsilon)$  of these small cycles.

In  $P$ , and hence in  $L_1$ , there can be at most  $\varepsilon k/10$  non-full vertices. This implies that for each subsegment  $L$  in  $L_1$  of length  $\varepsilon k/5$ , at least  $\varepsilon k/10$  of its vertices are full, thus there exists a set  $E_L$  of at least  $\varepsilon^2 k^2/100$  edges in  $U$  of the form  $x y$ , where  $x \in V(L)$ , and  $y \in V(P)$  is a descendant of  $x$  at distance at least  $4\varepsilon k/5$ . Clearly the distance between  $x$  and  $y$  in  $P$  is at most  $k$ , as all edges of  $Q'_2$  were tested according to our assumption in the beginning of this subsection. By the union bound and by Lemma 2.1, if we test the edges in  $E_L$  for every segment  $L$  in  $L_1$ , a.a.s. we can find one successfully exposed edge in each  $E_L$ .

To obtain the long cycle is straightforward. We start with a segment  $L^{(1)}$  of length  $\varepsilon k/5$  containing the endpoint of  $L_1$  which is smallest with respect to the order  $\leq_T$ . In this segment, we can find an edge  $u_1 v_1 \in E_{L^{(1)}}$  which was successfully exposed, where  $u_1 \in V(L^{(1)})$ . We then proceed recursively for each  $j$  as follows. Let  $L^{(j+1)}$  be the segment of  $L_1$  of length  $\varepsilon k/5$  whose smallest vertex with respect to  $\leq_T$  is the ancestor of  $v_j$  at distance  $\varepsilon k/5 + 1$  (hence  $L^{(j+1)}$  does not contain  $v_j$ ). Then choose an edge  $u_{j+1} v_{j+1} \in E_{L^{(j+1)}}$  which was successfully exposed, where  $u_{j+1} \in V(L^{(j+1)})$ . Repeat this process while  $v_j$  has distance at most  $5k$  from the smallest vertex from  $L_1$ . Assume the last segment chosen was  $L^{(t)}$ . The cycle  $J$  we seek can be easily seen from Figure 7.

The reader can check that the distance between  $u_j$  and  $v_{j-1}$  is always smaller than the distance between  $u_j$  and  $v_{j-2}$  with respect to the path  $P$ . In fact, we always jump ‘downwards’ by at least  $4\epsilon k/5$  and move ‘upwards’ by  $\epsilon k/5 + 1$ . So the total length of  $J$  is at least half of the distance between  $u_1$  and  $v_t$ , therefore  $|J| \geq \frac{5}{2}k$ , which proves the lemma.  $\square$

Now that we have studied the case when the number of poor vertices is small, it is natural to consider the case where this number is sufficiently large. In the next lemma, we show the existence of the desired cycle in this situation.

**Lemma 3.32.** *If the number of poor vertices not in  $Z$  inside a block  $B$  is at least  $200\epsilon k$ , then after we expose the remaining untested edges of  $G'_p$ , a.a.s.  $G'_p[V(B)]$  contains a cycle of length at least  $k + 1$ .*

Before we prove the previous lemma, we list and prove some technical results. The first one is Theorem 3.1(ii) proved by Krivelevich, Lee and Sudakov [12].

**Theorem 3.33.** *Let  $\mathcal{G}$  be a bipartite graph of minimal degree at least  $k$ . Then asymptotically almost surely  $\mathcal{G}_p$  contains a path of length  $(2 + o(1))k$  whenever  $p = \omega_k(1)/k$ , for any function  $\omega_k(1) < k$  that tends to infinity with  $k$ .*

The second technical result is inspired by the same paper [12].

**Proposition 3.34.** *Let  $B \in \mathcal{B}$  be a block. Suppose there exists a pair  $(P, X)$ , where  $P$  is a path in  $T \cap B$  of size  $(1 - 5\epsilon)k \leq |P| \leq (1 + \epsilon)k$ , and  $X \subseteq V(B) \setminus V(P)$  is a set of at least  $180\epsilon k$  vertices, such that for each  $v \in X$  there are at least  $(1 - 80\epsilon)k$  vertices  $u \in V(P)$  satisfying  $w \in U$ . After exposing all the edges in  $U$  connecting  $P$  to  $X$ , a.a.s. we can find a cycle in  $G'_p[V(B)]$  of length at least  $k + 1$ .*

**Proof.** The main idea here goes along the lines of the proof of Theorem 1.2 (Case 1) in the above-mentioned paper [12]. Let  $w$  be an arbitrary vertex in  $X$ . We test all the edges in  $U$  connecting  $w$  to  $P$ , and a.a.s. we obtain two successfully tested edges  $wu_1$  and  $wu_2$ , such that  $u_1$  and  $u_2$  are at distance at least  $(1 - 82\epsilon)k$  with respect to  $P$ . We redefine  $P$  as the segment of itself connecting  $u_1$  to  $u_2$ . We now know that  $(1 - 82\epsilon)k \leq |P| \leq (1 + \epsilon)k$ , and that for each vertex  $v \in X \setminus \{w\}$ , there exists a set of at least  $(1 - 163\epsilon)k$  vertices  $u \in V(P)$  such that  $wv \in U$ .

Let  $Y$  be an arbitrary subset of  $X \setminus \{w\}$  of size  $175\epsilon|P| < 180\epsilon k$ . We partition  $P$  into  $1/(175\epsilon)$  segments  $P_1, \dots, P_{1/175\epsilon}$ , each of length  $175\epsilon|P|$ . By an averaging argument, there exists an interval  $P_i$  for which  $e(P_i, Y) \geq (1 - 200\epsilon)|P_i||Y|$ . Consider a bipartite graph  $\Gamma$  formed by the edges of  $U$  with the vertex set being the union of the two parts  $P_i$  and  $Y$ . Note that the number of non-adjacent pairs is at most  $200\epsilon|P_i||Y|$  (also note that  $|Y| = |P_i|$ ). We repeatedly remove vertices of degree at most  $(1 - 40\epsilon^{1/2})|Y|$  from  $\Gamma$ . As long as the total number of deleted vertices is at most  $20\epsilon^{1/2}|Y|$ , each deletion accounts for at least  $20\epsilon^{1/2}|Y|$  non-adjacent pairs of  $\Gamma$ . So, if we continue this removal process

for at least  $20\epsilon^{1/2}|Y|$  vertices, the total number of non-adjacent pairs we removed from  $\Gamma$  is at least  $400\epsilon|Y|^2$ , which is a contradiction. Thus, this process must stop before we remove  $20\epsilon^{1/2}|Y|$  vertices, and we obtain a subgraph  $\Gamma_1$  of minimum degree at least  $(1 - 40\epsilon^{1/2})|Y|$ .

Let  $P_{i,0}$  and  $P_{i,1}$  be the two segments of  $P_i$  of length  $45\epsilon^{1/2}|Y|$  closest to the two endpoints of  $P_i$ . Even after removing the vertices in  $P_{i,0} \cup P_{i,1}$  from  $\Gamma_1$ , we are left with a graph  $\Gamma_2$  of minimum degree at least

$$(1 - 40\epsilon^{1/2})|Y| - 90\epsilon^{1/2}|Y| > \frac{9}{10}|Y|.$$

By Theorem 3.33,  $(\Gamma_2)_p$  a.a.s. contains a path of length at least

$$2\left(\frac{9}{10} + o(1)\right)|Y| > \frac{5}{3}|Y|.$$

By removing at most two vertices, we may assume that the endpoints  $x$  and  $y$  of this path are both in  $Y$ . Since  $\Gamma_1$  has minimum degree at least  $(1 - 40\epsilon^{1/2})|P_i|$ , both of these endpoints have at least  $5\epsilon^{1/2}|P_i| > \epsilon^{3/2}k$  neighbours in both  $P_{i,0}$  and  $P_{i,1}$ . By Lemma 2.1,  $(\Gamma_1)_p$  a.a.s. contains two edges  $xv_0$  and  $yv_1$ , where  $v_0 \in P_{i,0}$  and  $v_1 \in P_{i,1}$ . We found a path in  $(\Gamma_1)_p$  of length at least  $\frac{5}{3}|Y|$ , which starts at  $v_0$  and ends at  $v_1$ , and uses only vertices from  $Y \cup (P_i \setminus (P_{i,0} \cup P_{i,1}))$  as internal vertices. Together with the path  $P$  and the two edges  $wu_1$  and  $wu_2$ , we obtain a cycle of length at least

$$|P| - |P_i| + \frac{5}{3}|Y| > k + 1,$$

concluding the proof of the proposition. □

Our last auxiliary statement studies the set of poor vertices inside a block.

**Proposition 3.35.** *Let  $B$  be a block. Suppose  $x_1$  and  $x_2$  are two full poor vertices in  $B$ . If  $x_1$  and  $x_2$  have distance at least  $30\epsilon k$  with respect to  $T$ , then after we expose the remaining untested edges of  $G'_p$ , a.a.s.  $G'_p[V(B)]$  contains a cycle of length at least  $k + 1$ .*

**Proof.** Proposition 3.24 guarantees that, by testing some edges in  $U$  incident to  $x_1$  and  $x_2$ , we will obtain a.a.s. two rotating cycles  $J_1$  and  $J_2$  with pivots  $x_1$  and  $x_2$ , respectively. If  $J_1$  and  $J_2$  are disjoint, by Proposition 3.20 we can merge them and obtain the desired long cycle. Otherwise, if  $J_1$  and  $J_2$  intersect, then the unique path in  $T$  from  $x_1$  to  $x_2$  is contained in  $J_1 \cup J_2$  and is edge-disjoint from  $J_1 \cap J_2$ . Thus

$$|J_1| + |J_2| - |J_1 \cap J_2| = |J_1 \cup J_2| \geq |J_1 \cap J_2| + 30\epsilon k, \tag{3.6}$$

which implies  $|J_1 \cap J_2| < (1 - 15\epsilon)k$ . By Proposition 3.21, we can merge these two intersecting rotating cycles, thereby proving the proposition. □

We turn to prove Lemma 3.32.

**Proof of Lemma 3.32.** Observation 3.23 says that each pseudo-clique can contribute at most  $\epsilon k$  poor vertices to the block in which it is contained. Moreover, if  $B$  contains

more than one pseudo-clique, Proposition 3.17 would already imply the conclusion of this lemma. Thus, we may assume that among the poor vertices not in  $Z$ , at least  $199\epsilon k$  of them do not belong to pseudo-cliques that are contained in  $B$ . Let  $X$  denote the set of such vertices. We have  $|X| \geq 199\epsilon k$ , and clearly  $X$  does not contain the smallest vertex of  $B$  with respect to  $\leq_T$  (otherwise the smallest element would be poor, and hence  $|B| \leq \epsilon k$ ).

By Corollary 3.28, each vertex in  $X$  must necessarily be full in  $B$ . If there exist two full poor vertices  $v_1$  and  $v_2$  at distance at least  $30\epsilon k$  with respect to  $T$ , we can obtain the long cycle by using Proposition 3.35. Thus we may assume that all the poor full vertices of  $B$  are close to each other, that is, they have distance at most  $30\epsilon k$ . Let  $v$  be any full poor vertex, and let  $P$  be the path containing all the ancestors of  $v$  in  $B$  at distance less than  $k$ . Clearly  $(1 - 2\epsilon)k \leq |P| \leq k$ . For every vertex  $v' \in X$ , since  $v$  and  $v'$  are at distance at most  $30\epsilon k$ , there exists at least  $(1 - 80\epsilon)k$  pairs  $uv' \in U$ , where  $u \in P$  (recall that no edge in  $U$  connects pairs of vertices at distance larger than  $k$ ). Thus we can apply Proposition 3.34, and a.s. obtain a cycle of length at least  $k + 1$  in  $G'_p[V(B)]$ , finishing the proof of the lemma.  $\square$

The last case to be solved is when the number of poor vertices is not too large and not too small. The next lemma investigates this case.

**Lemma 3.36.** *Suppose the proportion of poor vertices inside a good block  $B \in \mathcal{B}$  is at least  $\epsilon/1000$ , but the total number of poor vertices in  $B \setminus Z$  is at most  $200\epsilon k$ . We have either*

- *$B$  contains a pseudo-clique  $C$  and all the other vertices in  $V(B) \setminus C$  belong to  $Z$ , or*
- *after testing all the edges of  $U$  joining two vertices from  $B$ , a.s.  $G'_p[V(B)]$  contains a cycle of length at least  $k + 1$ .*

As before, we need some technical statements in preparation for the proof of Lemma 3.36. The first statement strengthens Proposition 3.35.

**Proposition 3.37.** *Let  $B$  be a block. Suppose  $x_1, x_2 \notin Z$  are two vertices in  $B$ , such that for each  $i \in \{1, 2\}$ , there are at most  $\epsilon k$  edges in  $U$  connecting  $x_i$  to one of its descendants in  $B$ . If  $x_1$  and  $x_2$  have distance at least  $60\epsilon k$  with respect to  $T$  and are not comparable with respect to  $\leq_T$ , then after we expose the remaining untested edges of  $G'_p$ , a.s.  $G'_p[V(B)]$  contains a cycle of length at least  $k + 1$ .*

**Proof.** First observe that the smallest vertex in  $B$  does not belong to the set  $\{x_1, x_2\}$ , because otherwise  $x_1$  and  $x_2$  would be comparable with respect to  $\leq_T$ . Let  $L$  be the union of all pseudo-cliques  $C \in \mathcal{C}$  which are contained in  $B$ .

In the first case, both  $x_1$  and  $x_2$  belong to  $L$ . Let  $C_1, C_2 \in \mathcal{C}$  be the pseudo-cliques contained in  $B$  such that  $x_1 \in C_1$  and  $x_2 \in C_2$ . We claim that  $C_1 \neq C_2$ . To prove our claim, assume that  $C_1 = C_2$ . But by Observation 3.13, we know that the vertices of  $C_1$  induce a path in  $T$ , and hence in  $T \cap B$ . But this would imply that  $x_1$  and  $x_2$  are comparable with respect to  $\leq_T$ , contradicting the hypothesis of the proposition. Hence  $C_1 \neq C_2$ . But by Proposition 3.17, we can merge the two cycles in the pseudo-cliques and obtain the desired long cycle. This finishes the analysis of the first case.

In the second case, neither  $x_1$  or  $x_2$  belong to  $L$ . By Corollary 3.28, both  $x_1$  and  $x_2$  are full in  $B$ , since they do not belong to  $L \cup Z$ . After testing some of the edges of  $U$  incident to  $x_1$  and  $x_2$  we can obtain two rotating cycles  $J_1$  and  $J_2$  respectively, as Proposition 3.24 assures. Either  $J_1$  is disjoint from  $J_2$  or their intersection is of size at most  $(1 - 30\epsilon)k$  (we use the same strategy as in (3.6) to estimate the size of  $J_1 \cap J_2$ ). In any case, by either Proposition 3.20 or Proposition 3.21, a.a.s. we can obtain the long cycle after we test the remaining edges of  $U$  in  $B$ , concluding the analysis of this case.

In the last remaining case, we assume that  $x_1 \in L$  but  $x_2 \notin L$ . As before, we know that  $x_2$  is full in  $B$ . Using Proposition 3.24, we obtain a.a.s. a rotating cycle  $J$  in  $B$  for which  $x_2$  is its pivot. Moreover, since  $x_1 \in L$ , we also know that  $x_1$  belongs to a pseudo-clique  $C \in \mathcal{C}$  which is contained in  $B$ . We claim that  $|V(J) \cap C| < (1 - 30\epsilon)k$ . If  $V(J)$  and  $C$  are disjoint, then our claim is trivially true. Otherwise, if they intersect, then there is a cycle  $J'$  in  $B$  containing the vertices of  $C \setminus W$ , such that  $J' \cap T$  is a path (Observation 3.13). By a calculation analogous to (3.6), we obtain that  $|J \cap J'| = |V(J) \cap C| < (1 - 30\epsilon)k$ , proving our claim. Finally, we finish the proof of this proposition with a final application of Proposition 3.22 to obtain the long cycle.  $\square$

Our second auxiliary result allows us to estimate the number of poor vertices in a block.

**Proposition 3.38.** *Let  $\delta > 1$ ,  $B$  be a block,  $v$  be a poor vertex in  $B$ , and let  $P$  be the unique path from  $v$  to the smallest vertex from  $B$  with respect to  $\leq_T$ . If  $|V(B) \cap Z| < \epsilon k$ , and every poor vertex of  $B$  not in  $Z \cup \{v\}$  is at distance at most  $\delta \epsilon k$  from  $v$ , then there is no rich vertex in  $B$  outside  $P$  at distance at least  $\delta \epsilon k$  from  $v$ . Furthermore,  $B$  contains at least  $(|B| - |P|)/(\delta + 1)$  poor vertices.*

**Proof.** Let  $X$  be the set of poor vertices in  $B$ . The proof of the first part of the proposition goes by contradiction. Assume that there exists a rich vertex  $u$  outside  $P$  at distance at least  $\delta \epsilon k$  from  $v$ . Recall that  $D(u)$  denotes the set of descendants of  $u$  in  $B$ . Observe that all vertices from  $D(u) \cap X$  must have distance greater than  $\delta \epsilon k$  from  $v$ . On the other hand,  $|D(u) \cap X| \geq \epsilon k$  (since every rich vertex has at least  $\epsilon k$  poor descendants), and every vertex in  $D(u) \cap X$  must belong to  $Z$  by the assumptions of the proposition. We then have a contradiction, because  $|D(u) \cap X| \geq \epsilon k > |V(B) \cap Z|$ . This contradiction proves the first statement of the proposition.

For each  $u \notin V(P)$  such that  $u$  has an immediate ancestor in  $P$ , we shall prove that

$$|D(u) \cap X| \geq \frac{|D(u)|}{\delta + 1}. \tag{3.7}$$

We identify the set  $F$  of the rich vertices in  $D(u)$  that have no rich descendant. For every  $w \in F$ , since  $w$  is rich, we have  $|D(w)| > \epsilon k$ , and all vertices in  $D(w)$  are poor, except for  $w$  itself. Hence  $|D(u) \cap X| \geq \epsilon k|F|$ . Furthermore, no rich vertex in  $D(u)$  is at distance larger than  $\delta \epsilon k$  from  $u$ , since it would have distance at least  $\delta \epsilon k$  from  $v$  as well, which is impossible. In other words, every rich vertex in  $D(u)$  belongs to some path in  $T$  from a vertex in  $F$  to  $u$ . Since all these paths have length at most  $\delta \epsilon k$ , the total number of rich vertices in  $D(u)$  is at most  $\delta \epsilon k|F|$ , thereby proving (3.7). Therefore the total

number of poor vertices in  $B$  is at least  $(|B| - |P|)/(\delta + 1)$ , concluding the proof of the proposition.  $\square$

The last auxiliary result is to handle the case of a block containing a long path and few vertices in  $Z$ . The statement is as follows.

**Proposition 3.39.** *Let  $B$  be a block, let  $v$  be a poor vertex in  $V(B) \setminus Z$ , and let  $P$  be the unique path from  $v$  to the smallest vertex from  $B$  with respect to  $\leq_T$ . If*

$$|V(B) \cap Z| < 3\epsilon k/10, \quad |P| \geq (1 + 1000\epsilon)k,$$

*and every poor vertex of  $B$  not in  $Z \cup \{v\}$  is at distance at most  $100\epsilon k$  from  $v$ , then after testing the remaining untested edges from  $G'_p$ , a.a.s. we can find a cycle of length at least  $k + 1$  in  $G'_p[V(B)]$ .*

**Proof.** Let  $L$  be the union of all pseudo-cliques that are contained in  $B$ . Observe that  $L$  is the union of at most one pseudo-clique, as otherwise we would have a cycle of length at least  $k + 1$  by Proposition 3.17.

We claim that either  $v \in L$ , or after possibly testing a few edges from  $U$  incident to  $v$ , we obtain a.a.s. a rotating cycle  $J$  with pivot  $v$ . To see this, observe that if  $v \notin L$ , then Corollary 3.28 implies that  $v$  must be full in  $B$ . Using Corollary 3.25, we obtain a.a.s. such a rotating cycle  $J$ , proving our claim. In any case, we can assume that  $v$  belongs to a cycle (either because  $v$  belongs to a pseudo-clique  $C$  contained in  $B$  or because it is the pivot of a rotating cycle  $J$ ) of length at least  $(1 - 5\epsilon)k$ .

Suppose there is vertex  $w$  in  $P$  at distance at least  $100\epsilon k$  from the endpoint  $v$  of  $P$  that satisfies:

- (i)  $w \notin Z \cup L$ , and  $w$  is not the smallest vertex in  $B$ , and
- (ii) there are at most  $\epsilon k$  edges of  $U$  connecting  $w$  to one of its descendants.

By (i) and by Corollary 3.28, we know that  $w$  must be full in  $B$ , and hence by Proposition 3.24, after testing the far-reaching edges in  $U$  incident to  $w$ , we obtain a.a.s. a rotating cycle  $J'$  with pivot  $w$ . We could then merge  $J'$  with the large cycle containing  $v$  (which could be either from a pseudo-clique  $C \in \mathcal{C}$  if  $v \in L$ , or from the rotating cycle  $J$  with pivot  $v$ , if  $v \notin L$ ) by using Proposition 3.22, hence obtaining the desired long cycle, and we would be done.

From the discussion in the last paragraph, we may assume without loss of generality that there is no vertex  $w$  satisfying both (i) and (ii). Thus every vertex  $w \in V(P)$  at distance at least  $100\epsilon k$  from  $v$  satisfying (i) must have at least  $\epsilon k$  descendants  $u$  in  $B$  such that  $uw \in U$ .

Let us denote by  $z$  the smallest vertex with respect to  $\leq_T$  on  $P$  such that there are at least  $\epsilon k/40$  vertices  $u$  of distance at most  $100\epsilon k$  from  $v$  with  $zu \in U$ . Notice that such a vertex  $z$  exists, since, for example,  $v$  itself, as a poor full vertex, has at least  $\epsilon k/40$  neighbours in  $U$  of distance at most  $100\epsilon k$  from  $v$ . Then, for every  $w \in V(P)$  with  $w <_T z$  of distance at least  $100\epsilon k$  from  $v$  and satisfying (i), of the at least  $\epsilon k$  descendants  $u$  in  $B$  such that  $uw \in U$ , at most  $|Z \cap V(B)| + \epsilon k/40 < 13\epsilon k/40$  do not belong to  $P$ . To see this,

observe that  $w$  does not have a rich descendant  $u$  outside  $P$  of distance at least  $100\epsilon k$  from  $v$ , as otherwise it would contradict the conclusion of Proposition 3.38 (applied with  $\delta = 100$ ). Furthermore, every poor descendant of  $w$  of distance at most  $100\epsilon k$  from  $v$  must belong to  $Z \cap B$ . So the total number of descendants of  $w$  outside  $P$  is at most  $13\epsilon k/40$ . In particular,  $w$  has at least  $27\epsilon k/40$  descendants  $u$  such that  $u \in V(P)$  and  $uw \in U$ .

At most  $13\epsilon k/20$  of these edges connect  $w$  to a vertex in  $P$  at distance at most  $13\epsilon k/20$  from  $w$ . Hence there are at least  $\epsilon k/40$  edges in  $U$  connecting  $w$  to one of its descendants at distance at least  $13\epsilon k/20$  on  $P$ . If we test these edges, a.s. we can find a successfully tested edge connecting  $w$  to one of its deep descendants in  $P$ , thus forming a small cycle of length at least  $13\epsilon k/20$ . We can now use the same technique as in the proof of Lemma 3.29 to finish the proof of the proposition (see Figure 7). In the next few paragraphs, we briefly sketch this technique. We also remark that in our case we only need to merge constantly many small cycles, which simplifies the union bound argument.

The idea is to start at a full vertex  $w_0$  that satisfies (i) and is at distance between  $(1 + 998\epsilon)k$  and  $(1 + 999\epsilon)k$  from  $v$ , and repeat the following loop. For each  $i = 0, 1, 2, \dots$ , by testing some edges of  $U$  incident to  $w_i$ , a.s. we can find a neighbour  $w'_i$  of  $w_i$  at distance at least  $13\epsilon k/20$  from  $w_i$  which is a descendant of  $w_i$  on  $P$ . Then we go ‘upwards’ (in direction to the smallest vertex from  $B$ ) the path  $P$  starting from  $w'_i$  until we find another full vertex  $w_{i+1}$ . Recall that we need to go ‘upwards’ at most  $3\epsilon k/10$  vertices to reach this full vertex, as long as we move entirely outside  $L$ . Also observe that the small cycles do not ‘double-overlap’, as  $13\epsilon k/20 > 2 \cdot (3\epsilon k/10)$ . We repeat the loop until we either hit the interior of  $J$  (if it exists), or a vertex from  $L$  which is not the smallest vertex in  $L$ , or a vertex below  $z$ .

Recall that  $v$  belongs to a cycle, which could be formed by vertices from either  $J$  or  $L$ . Since this cycle has size between  $(1 - 5\epsilon)k$  and  $(1 + \epsilon)k$ , we must necessarily stop this procedure after constantly many iterations of the loop. More precisely, if  $T$  denotes the time we stopped, then

$$T < \frac{999\epsilon k}{7\epsilon k/20} < 300.$$

Assume that at the very last step, the vertex  $w'_T$  either belongs to the interior of  $J$  (if it exists) or is in  $L$  (but is not the smallest vertex in the pseudo-clique). In the first case, we can close the cycle we are forming with  $J$ , because  $v$  has a.s. a neighbour which is an ancestor of  $w'_T$  at distance at most  $5\epsilon k$  from  $w'_T$ . In the latter case, when  $w'_T \in L$ , from Proposition 3.22 one can deduce that the smallest vertex from  $L$  in the block is at distance at most  $(1 + 40\epsilon)k$  from  $v$ . In addition, by Proposition 3.18, we can close the cycle using the majority of vertices from  $L$ . More specifically, there is a path in  $G'_p[L]$  of length at least  $(1 - 20\epsilon)k$  which connects  $w'_T$  and the smallest vertex from  $L$ . Therefore, regardless of what happens in the last iteration of our procedure, the merged cycle always has size at least

$$(1 + 998\epsilon)k - 45\epsilon k - 20\epsilon k - 300 \cdot \frac{3\epsilon k}{10} > k.$$

The case when, at the very last step, the vertex  $w'_T$  is below  $z$  (but  $w'_T$  neither belongs to the interior of  $J$  (if it exists) nor is a non-smallest vertex of  $L$ ), proceeds analogously to

the above argument: we expose the (at least  $\varepsilon k/40$ ) edges from  $U$  connecting  $z$  to vertices at distance at most  $100\varepsilon k$  from  $v$ , obtaining at least one neighbour  $v'$  of  $z$ . By definition of  $z$ , it has distance at most  $(1 + 100\varepsilon)k$  from  $v$ , since  $z$  has neighbours at distance at most  $100\varepsilon k$  from  $v$ , and all far-reaching edges were already tested unsuccessfully in  $Q'_2$ . Therefore, since  $w'_T$  must have distance at least  $(1 - 5\varepsilon)k$  from  $v$ , the segment of  $P$  connecting  $z$  and  $w'_T$  has length at most  $105\varepsilon k$ . The merged cycle using the path from  $w'_T$  to  $v'$  has size at least

$$(1 + 998\varepsilon)k - 105\varepsilon k - 100\varepsilon k - 300 \cdot \frac{3\varepsilon k}{10} > k,$$

thereby proving the proposition. □

We are ready to prove Lemma 3.36.

**Proof of Lemma 3.36.** Let  $X$  be the set of poor vertices of  $B$ , and let  $Y = X \setminus Z$ . Since  $|Z \cap V(B)| \leq 10^{-6}\varepsilon|B|$  (because  $B$  is good) and  $|X| \geq 10^{-3}\varepsilon|B|$ , we must conclude that  $|Y| \geq \varepsilon|B|/1100$ . We also know that  $|Y| \leq 200\varepsilon k$ , and this implies that  $|B| < 3 \cdot 10^5 k$  (we will improve this bound later), thus  $|Z \cap V(B)| < 3\varepsilon k/10$ . Finally, we have

$$|X| \leq |Y| + |Z \cap V(B)| \leq 201\varepsilon k.$$

Fix a vertex  $v \in Y$  arbitrarily. Using Proposition 3.37 one can see that  $v$  has distance (with respect to  $T$ ) of at most  $60\varepsilon k$  from any other vertex from  $Y$ , as otherwise we would obtain the long cycle and the second conclusion of the lemma would hold. Indeed, since we are considering poor vertices, there can be no situation when one of them is a descendant of the other of distance at least  $60\varepsilon k$ , and therefore Proposition 3.37 can be applied.

Let  $P$  be the path in  $T$  joining  $v$  to the smallest vertex  $v_0$  of  $B$  with respect to  $\leq_T$ . By applying Proposition 3.38 for  $\delta = 60$ , we obtain that the total number of poor vertices in  $B$  is at least  $(|B| - |P|)/61$ , which implies that

$$|B| \leq |P| + 61|X| \leq |P| + 15000\varepsilon k.$$

We might assume then that  $|P| < (1 + 1000\varepsilon)k$ . This is because the conclusion of the lemma would be true otherwise, as Proposition 3.39 shows. In particular, we must have  $|B| < (1 + 16000\varepsilon)k$ . If  $|V(B) \setminus Z| > (1 + \varepsilon)k$ , then we claim that  $\Gamma := G'[V(B) \setminus (Z \cup \{v_0\})]$  is a graph with minimum degree at least  $(1 - 5\varepsilon)k$ . Indeed, every vertex in  $\Gamma$  which does not belong to a pseudo-clique contained in  $B$  is full in  $B$ , and every vertex in  $\Gamma$  which does belong to a pseudo-clique contained in  $B$  has degree at least  $(1 - 5\varepsilon)k$  in  $G'[V(B) \setminus (Z \cup \{v_0\})]$  because  $|Z \cap V(B)| < 3\varepsilon k/10$  and every pseudo-clique not completely inside the waste lost at most  $\varepsilon k/2$  vertices to  $W$ , and the claim follows.

Thus  $\Gamma$  is a graph with minimum degree at least  $(1 - 6\varepsilon)k$  satisfying  $(1 + \varepsilon)k \leq |\Gamma| \leq (1 + 16000\varepsilon)k$ . However, as assumed at the beginning of this subsection, such a graph cannot exist, because it would violate the assumption that Proposition 3.11 holds. This implies that  $|V(B) \setminus Z| \leq (1 + \varepsilon)k$ , hence  $|B| \leq (1 + 2\varepsilon)k$ .

Next, we claim that there exists a pseudo-clique  $C \in \mathcal{C}$  such that  $B$  contains  $C$ . To prove our claim, suppose, for a contradiction, that  $B$  does not contain any pseudo-clique from  $\mathcal{C}$ .

Then every vertex  $u$  in  $V(B) \setminus (Z \cup \{v_0\})$  must be outcast. Indeed, if  $u$  belongs to a pseudo-clique  $C' \in \mathcal{C}$ , then  $u$  is the smallest vertex in the unique block that contains  $C'$ , hence  $u \in Z_3 \subseteq Z$ , which is a contradiction. Moreover, since every vertex in  $V(B) \setminus (Z \cup \{v_0\})$  is full in  $B$ , we have that  $\Gamma$  is a graph with minimum degree at least  $(1 - 2\varepsilon)k$  whose vertex set consists of only outcast vertices. This fact, together with  $|\Gamma| < (1 + \varepsilon)k$ , implies that the set  $V(B) \setminus (Z \cup \{v_0\})$  forms a pseudo-clique in  $G$  which is disjoint from all the other pseudo-cliques from  $\mathcal{C}$ . But this contradicts the maximality of the union  $\bigcup_{C \in \mathcal{C}} C$ , since we chose the collection of disjoint pseudo-cliques that covers the maximum number of vertices possible. This contradiction proves that  $B$  contains exactly one pseudo-clique  $C$  from  $\mathcal{C}$ .

It remains to show that  $V(B) \subseteq Z \cup C$ , or equivalently  $V(B) \setminus (C \cup Z) = \emptyset$ . Suppose not. Clearly  $v_0 \in Z_3 \subseteq Z$  (because  $v_0$  is the smallest vertex in  $B$ , and  $B$  contains a pseudo-clique). Every vertex in  $V(B) \setminus (Z \cup C)$  must be outcast and full in  $B$ , hence the graph  $\Gamma' := G'[C \cup (V(B) \setminus Z)]$  is a graph with minimum degree  $(1 - 4\varepsilon)k$ . By the assumption that Proposition 3.11 holds, we have that

$$|\Gamma'| < \left(1 + \frac{\varepsilon}{2}\right)k.$$

Hence the vertices of  $\Gamma'$  form a pseudo-clique in  $G$ , and if we replace  $C$  by  $C' := C \cup (V(B) \setminus Z)$  (recall that  $C' \setminus C$  consists only of outcast vertices), we obtain a family of pseudo-cliques whose union is larger than before, contradicting the maximality of  $\bigcup_{C \in \mathcal{C}} C$ . This final contradiction establishes the lemma. □

We turn to prove Lemma 3.1. One important fact that will be used in subsequent double-counting arguments is the following consequence of Observation 3.16: by removing the smallest vertex with respect to  $\leq_T$  from each block in  $\mathcal{B}$ , we obtain a family of pairwise vertex-disjoint graphs.

**Proof of Lemma 3.1.** Assume, for a contradiction, that  $\ell > 10^7 \cdot n/\varepsilon k$ , but  $G_p$  does not a.a.s. contain a cycle of length at least  $k + 1$ . Lemmas 3.29, 3.32 and 3.36 combined imply that either  $G_p$  a.a.s. contains a cycle of length at least  $k + 1$ , or all the good blocks contain a pseudo-clique inside, and the remaining vertices not in the pseudo-clique are in  $Z$ . We can bound the number  $t$  of vertices not in good blocks as follows.

By definition, every non-good block  $B$  contains at least  $(\varepsilon/10^6)|B|$  vertices from  $Z$ . Furthermore, as it was previously remarked, if we remove the smallest vertex from each block in  $\mathcal{B}$ , we obtain a family of disjoint graphs. Hence, if  $t_0$  denotes the number of blocks in  $\mathcal{B}$  of size 1, then

$$|Z| \geq t_0 + \sum_{B \in \mathcal{B}} (|Z \cap B| - 1) \geq t_0 + \frac{\varepsilon}{2 \cdot 10^6} \sum_{\substack{B \text{ not good} \\ |B| > 1}} |B| \geq \frac{\varepsilon t}{2 \cdot 10^6}. \tag{3.8}$$

Thus

$$t \leq \frac{2 \cdot 10^6}{\varepsilon} \cdot |Z|,$$

where  $t$  denotes the number of vertices in non-good blocks. By (3.5) we obtain

$$t \leq 10^6 \cdot \left( \frac{3n}{\varepsilon k} + \frac{32}{\varepsilon^2 p k} \cdot \left( \ell + \frac{n}{k} \right) + 24 \right).$$

Using Lemma 3.36, we can estimate the number of outcast vertices by adding the estimation of  $|Z|$  in (3.5) with our previous bound for  $t$  in (3.8) for  $t$ . This is true because if an outcast vertex is in a good block, then it must belong to  $Z$ . Hence we have  $\ell \leq t + |Z|$ , which implies that  $\ell \leq 10^7 \cdot n/\varepsilon k$ , a contradiction that establishes Lemma 3.1. □

Next is Lemma 3.2.

**Proof of Lemma 3.2.** Suppose that  $\ell \leq 10^7 \cdot n/\varepsilon k$ , but  $G_p$  does not a.a.s. have a cycle of length at least  $k + 1$ . Clearly  $|\mathcal{C}|$  is roughly  $n/k$ , as the number of outcast vertices is  $\ell = o(n)$ . Let  $\mathcal{B}'$  be the sub-family of  $\mathcal{B}$  consisting of the good blocks. If we plug the inequality  $\ell \leq 10^7 \cdot n/\varepsilon k$  into the bound (3.5), we obtain  $|Z| \leq 1.1n/k$ , since  $\varepsilon^2 p k \rightarrow \infty$  as  $k \rightarrow \infty$ . Moreover, using inequality (3.8), we obtain that the number of vertices of  $G'$  not in good blocks is at most  $10^7 \cdot n/\varepsilon k = o(n)$ . Because of Lemma 3.36, every member  $B$  of  $\mathcal{B}'$  contains a pseudo-clique  $C \in \mathcal{C}$  and  $V(B) \setminus C \subseteq Z$ , hence  $|\mathcal{B}'| \approx n/k$ , or more specifically,  $0.99n/k \leq |\mathcal{B}'| \leq 1.01n/k$ . Furthermore, by Corollaries 3.14 and 3.15, the total number of edges in  $Q_2 \cup Q_3$  (recall that  $Q_2$  is the set of edges tested by DFS, and  $Q_3$  is the set of edges tested by the block algorithm) is at most

$$10^8 \cdot \frac{n}{\varepsilon p k} + 2\varepsilon^2 k = o(\varepsilon n).$$

We claim that every block  $B \in \mathcal{B} \setminus \mathcal{B}'$  of size  $1 \leq |B| < (1 - 4\varepsilon)k$  must contain at least  $\max\{1, |B| - 1\}$  vertices from  $Z$  (and hence  $B$  is necessarily not good). This is because  $B$  can not contain a pseudo-clique (its size is too small) and every vertex of  $V(B) \setminus Z$  which is not the smallest with respect to  $\leq_T$  must be full in  $B$  (see Corollary 3.28). However  $B$  contains no full vertex, as  $|B| < (1 - 4\varepsilon)k$ , therefore  $B$  contains at least  $|B| - 1$  vertices from  $Z$ . When  $|B| = 1$ , the unique vertex in  $B$  is isolated in  $G'_p$ , hence it belongs to  $Z$  (and belongs to no other block in  $\mathcal{B}$ ) thus proving our claim.

The number of blocks  $B \in \mathcal{B} \setminus \mathcal{B}'$  having size  $|B| \geq (1 - 4\varepsilon)k$  is  $o(n/k)$ . This is because the total number of vertices not in good blocks in  $G'$  is at most  $o(n)$ . Thus the number of blocks in  $B \in \mathcal{B} \setminus \mathcal{B}'$  having size  $|B| < (1 - 4\varepsilon)k$  is at most  $|Z| \leq 1.1n/k$ . Combining these observations, we obtain  $|\mathcal{B}| \leq 3|\mathcal{B}'|$ .

The above implies the following four statements.

- (i) Fewer than  $\frac{1}{4} \cdot |\mathcal{B}'|$  blocks in  $\mathcal{B}'$  have more than five vertices from  $Z$ . This is a consequence of the inequality  $|Z| \geq \sum_{B \in \mathcal{B}'} (|Z \cap V(B)| - 1)$ .
- (ii) The number of blocks in  $\mathcal{B}'$  having more than five cut-vertices is less than

$$\frac{1}{5} \cdot |\mathcal{B}| \leq \frac{3}{5} \cdot |\mathcal{B}'|.$$

- This is because the total number of cut-vertices is at most  $|\mathcal{B}| - 1$ , and after the removal of the smallest vertex from each block in  $\mathcal{B}$ , those blocks containing more than five cut-vertices still contribute at least five to this total.
- (iii) At least  $0.99|\mathcal{B}'|$  blocks  $B \in \mathcal{B}'$  are such that there are fewer than  $\epsilon k/3$  edges from  $G'$  having one endpoint in  $V(G') \setminus V(B)$  and the other being a non-cut-vertex of  $B$ . This is true since every such edge is necessarily tested and belongs to  $Q_2 \cup Q_3$ , and the total number of edges in  $Q_2 \cup Q_3$  is  $o(\epsilon n)$ .
  - (iv) At least  $0.99|\mathcal{B}'|$  blocks in  $\mathcal{B}'$  contain pseudo-cliques that satisfy the condition (3.3) in Lemma 3.12.

Hence, there exists a block  $B \in \mathcal{B}'$  satisfying the conditions stated in (i)–(iv). Let  $C \in \mathcal{C}$  be the unique pseudo-clique contained in  $B$ .

It is time to incorporate the waste vertices back. Let  $N$  denote the union of  $V(B) \cap Z$  with all cut-vertices from  $B$ . By (i) and (ii), the set  $N$  has size at most 10 and clearly  $V(B) \subseteq C \cup N$ . Let  $F$  be the set of edges in  $G$  connecting  $C \setminus N$  to a vertex outside  $C \cup N$ . We prove that  $|F| \leq \epsilon k$ . In order to prove this inequality, we will apply Lemma 3.12. Let us recall the definitions of the sets  $D_1, D'_1, D_2$  and  $\mathcal{E}$  in (3.3). The set  $D_1$  consists of all the vertices in  $C$  that have more than  $\epsilon k$  neighbours outside  $C$  in  $G$ . The subset  $D'_1 \subseteq D_1$  is the union of  $D_1 \cap W$  with all vertices in  $D_1 \setminus W$  which lost more than a  $\frac{1}{100}$  proportion of its neighbours outside  $C$  after the deletion of  $W$ . The set  $D_2$  is the set of all vertices from  $G$  not in  $C$  that have at least  $\epsilon k$  neighbours in  $C$ . At last,  $\mathcal{E}$  is the set of all edges from  $G$  connecting  $C \setminus D_1$  to a vertex outside not in  $C \cup D_2$ .

We claim that all vertices in  $D_1 \setminus D'_1$  are cut-vertices, and thus  $D_1 \setminus D'_1 \subseteq N$ . Every vertex in  $D_1 \setminus D'_1$  sends at least  $0.99\epsilon k$  edges outside  $C$  in  $G'$ ; in particular, it also sends at least  $0.99\epsilon k - |N| > \epsilon k/3$  edges outside  $B$ . But by (iii), every vertex in  $B$  that sends at least  $\epsilon k/3$  edges to the outside  $B$  must be a cut-vertex, hence  $D_1 \setminus D'_1 \subseteq N$ .

Next, we claim that  $D'_1 = \emptyset$ . By the discussion in the previous paragraph, we have  $D_1 \setminus D'_1 \subseteq N$ , hence  $|D_1 \setminus D'_1| \leq 10$ . The inequality (3.3) states that  $|D'_1| \leq |D_1|/100$ , thus  $|D_1 \setminus D'_1| \geq 99|D'_1|$ . Hence we have  $|D'_1| < 10/99 < 1$ , which implies  $D'_1 = \emptyset$ .

Our next claim states that  $D_2 \setminus W \subseteq N$ . To prove this, let us first show that every vertex in  $D_2 \setminus W$  belongs to  $B$ . A vertex in  $D_2 \setminus W$  sends at least  $\epsilon k$  edges to  $C$  in  $G$ , and since  $|C \cap W| < \epsilon k/2$  (as otherwise  $C$  would be completely thrown away to the waste), this implies that every vertex in  $D_2 \setminus W$  has more than  $\epsilon k/2$  neighbours in  $C \setminus W$  in the graph  $G'$ , hence more than  $\epsilon k/2 - |N|$  neighbours in  $C \setminus (W \cup N)$ . By (iii), any such vertex must belong to  $B$ , hence  $D_2 \setminus W \subseteq V(B)$ . On the other hand, since  $D_2$  and  $C$  are disjoint, clearly we must have  $D_2 \setminus W \subseteq V(B) \cap Z \subseteq N$ , proving our claim.

Similarly to the proof of  $D'_1 = \emptyset$ , let us now show that  $D_2 \cap W = \emptyset$ . Because  $D_2 \setminus W \subseteq N$ , we have  $|D_2 \setminus W| \leq 10$ , and by (3.3), we must have  $|D_2 \setminus W| \geq 99|D_2 \cap W|$ , therefore  $|D_2 \cap W| < 10/99 < 1$ , which implies  $D_2 \cap W = \emptyset$ .

We turn to prove  $|F| < \epsilon k$ . Assume not. Because  $D_1, D_2 \subseteq N$ , we have  $F \subseteq \mathcal{E}$ . Moreover, every vertex in  $C \setminus D_1$  can send at most  $\epsilon k$  edges to the outside  $C$  in  $G$ , and every vertex not in  $D_2$  can send at most  $\epsilon k$  edges to  $C$  in  $G$ . Thus

$$|\mathcal{E} \setminus F| \leq |N \setminus D_1|\epsilon k + |N \setminus D_2|\epsilon k \leq 20\epsilon k,$$

and hence

$$|F| \leq |\mathcal{E}| \leq |F| + 20\epsilon k \leq 21|F|.$$

By the inequality (3.3), we know that  $|\mathcal{E} \setminus E(G')| \leq |\mathcal{E}|/100$ , hence  $|F \setminus E(G')| \leq 21|F|/100$ . This implies that there are at least  $0.79\epsilon k > \epsilon k/3$  edges in  $G'$  connecting a vertex from  $C \setminus N$  to a vertex outside  $C \cup N \supseteq V(B)$ , which contradicts (iii), hence  $|F| < \epsilon k$ . Therefore the pair  $(C, N)$  satisfies the statement of Lemma 3.2. This concludes the proof of the lemma. □

### 3.6. Step 6: finishing the proof

It remains to prove Lemma 3.3.

**Proof of Lemma 3.3.** If we remove the vertices from  $N$  that have fewer than  $\epsilon k$  neighbours in  $C \cup N$ , we might increase the number of edges between  $C \setminus N$  and  $V(G) \setminus (C \cup N)$  to at most  $\epsilon k + |N|\epsilon k \leq 11\epsilon k$ . So we can assume that every vertex from  $N$  has at least  $\epsilon k$  neighbours from  $G$  in  $C \cup N$ , and that  $e_G(C \setminus N, V(G) \setminus (C \cup N)) \leq 11\epsilon k$ . Let  $X = C \cup N$ . From now on, we only deal with the graph  $G[X]$ .

Observe for the beginning that  $|X| \geq k + 1$ . Indeed, the minimum degree in  $G$  is at least  $k$ , and the number of edges between  $C \setminus N$  and  $V(G) \setminus X$  is less than the size of  $C \setminus N$ . In the following, we show that a.s.  $G_p[X]$  is Hamiltonian.

The general framework of the proof is to show some expansion properties of  $G_p[X]$  and then to deduce the Hamiltonicity from them. There are several recent papers dedicated to or just using Hamiltonicity of expanders, and the notion of expanders is slightly different every time, depending on the setting it should be applied in. Here we go with the notion used by Glebov and Krivelevich [8]: a graph  $H$  with the vertex set  $[m]$  is called a  $p'$ -expander, if there exists a set  $D \subset [m]$  such that  $H$  and  $D$  satisfy the following properties.

- $|D| \leq m^{0.09}$ .
- The graph  $H$  does not contain a non-empty path of length at most  $(2 \log m)/(3 \log \log m)$  such that both of its (possibly identical) endpoints lie in  $D$ .
- For every set  $S \subset [m] \setminus D$  of size  $|S| \leq 1/p'$ , its neighbourhood satisfies

$$|N(S)| \geq \frac{mp'}{1000}|S|.$$

Let us denote  $F = G[X]$ , and let  $m = |X|$  be its order. Furthermore, let us define for convenience  $p' = (\log m + \log \log m)/m$ . We first show that for every  $f_1, f_2 = \omega_m(1)$ ,  $f_1 < f_2 < \log \log m$ , every graph  $H$  satisfying  $F_{p_1} \subseteq H \subseteq F_{p_2}$  with  $p_i = (\log m + \log \log m + f_i)/m$  is a.s. a  $p'$ -expander. (Notice that we are coupling  $F_{p_1}$  and  $F_{p_2}$ , so that  $F_{p_1} \subseteq F_{p_2}$ .) Indeed, let us fix

$$D = \{v \in X : d_{F_{p_1}}(v) < mp'/100\}$$

to be the set of all vertices from  $X$  with degree less than  $mp'/100$  in  $F_{p_1}$ . The proof of the first property is similar to the proof of Claim 4.3 in [2], and the second property is shown to hold similarly to Claim 4.4 in [2]. Finally, the proof of the third bullet follows the lines

of the corresponding proof in Lemma 10 in [8]. Furthermore, observe that Lemma 2.1 guarantees us that a.a.s. every vertex from  $C$  has degree at least two in  $H$ , and for the vertices in  $N$  this also holds a.a.s. by Lemma 2.1. Hence, the random graph  $H$  is a.a.s. a  $p'$ -expander with minimum degree at least 2. Applying Lemma 11 from [8], we see that  $H$  is either Hamiltonian or has quadratically many boosters.

With this statement in our toolbox, the proof is similar to the proof of Proposition 3.10. We fix  $p_1$  such that  $p - p_1 = \omega_m(1)$ , and let  $p_2 = p$ . We start with  $H = F_{p_1}$  and successively add random edges to its edge set until we obtain  $F_p$ . We update the set of boosters after each new edge. Every such edge has at least constant probability to be a booster for the current  $H$  as long as  $H$  is not Hamiltonian. Every added edge that is a booster increases the length of the longest cycle in the current graph by at least one, or makes it Hamiltonian. Therefore, after at most  $k$  added boosters, the process would end with a Hamiltonian graph. On the other hand, the total number of added edges is a binomial random variable  $|E(F_p)| - |E(F_{p_1})|$  with  $\binom{m}{2}$  trials and probability  $p - p_1 = \omega_m(1)$ . By Lemma 2.1, with probability at least  $1 - \exp(-m)$ , the number of new edges that are added to obtain  $F_p$  from  $F_{p_1}$  is  $\omega_m(m)$ . Hence, Lemma 2.1 guarantees us that a.a.s. we get sufficiently many boosters to make the graph  $F_p$  Hamiltonian, proving the lemma.  $\square$

#### 4. Concluding remarks and open questions

In this paper, we studied random subgraphs of graphs with large minimum degree. Our goal was to extend classical results on random graphs to a more general model, where we replace the host graph by a graph with large minimum degree. We determined the threshold probability for having cycle of length at least  $k + 1$  in the random subgraph of graph with minimum degree at least  $k$ , showing that the assertion about Hamiltonicity of  $\mathbb{G}(k + 1, p)$  can be extended to this setting.

We believe that there are further interesting statements that one can deduce from our proof. One of them is the bipartite version of Theorem 1.1. Namely, that in a bipartite graph with minimum degree at least  $k$ , the random subgraph (with the same probability as in this paper) a.a.s. contains a cycle of length at least  $2k$ . However, since the paper is already quite long, we do not check all the technical details needed for the proof of this statement.

Another fact that can be shown similarly to Theorem 1.1 is as follows. Let  $G$  be a graph with minimum degree at least  $k$ , and fix a constant  $c$ . If

$$p = p(k) \geq \frac{\log k + \log \log k + c}{k},$$

then  $G_p$  contains a cycle of length at least  $k + 1$  with probability at least  $e^{-e^{-c}} - o(1)$ . This particular statement is an analogue of the well-known result on the probability of  $\mathbb{G}(k + 1, p)$  being Hamiltonian in the range of  $p$  where the probability of having one vertex of degree at most one is a constant (see e.g. [4]). The only difference in the proof compared to Theorem 1.1 would be the proof of the corresponding version of Lemma 3.3, since this is the only place where we use the additional summand  $\omega(1)$  in the definition of  $p$ .

One natural question is to determine whether the results of this paper, as well as several previous ones on this topic, hold if one weakens the condition of minimum degree of the host graph. One possibility here would be to only require the host graph  $G$  to have *average degree* at least  $k$ . Does this still guarantee cycles of length  $(1 - o(1))k$  and  $k + 1$  in  $G_p$ , for the same value of  $p$  as in [12] and in this paper?

Finally, it would be interesting to find more monotone properties  $\mathcal{P}$  for which the threshold probability in the binomial random graph model is the smallest among all host graphs of given minimum degree. Formally, these are the properties  $\mathcal{P}$  such that if  $\mathbb{G}(n, p)$  a.a.s. satisfies  $\mathcal{P}$ , then this holds a.a.s. also for a random subgraph  $G_p$  of a graph  $G$  with minimum degree at least  $n - 1$ .

### Acknowledgement

We would like to thank an anonymous referee for the thorough review of the paper and helpful comments, especially regarding Proposition 3.39.

### References

- [1] Alon, N. and Spencer, J. (2008) *The Probabilistic Method*, Wiley.
- [2] Ben-Shimon, S., Krivelevich, M. and Sudakov, B. (2011) On the resilience of Hamiltonicity and optimal packing of Hamilton cycles in random graphs. *SIAM J. Discrete Math.* **25** 1176–1193.
- [3] Bollobás, B. (1984) The evolution of sparse graphs. In *Graph Theory and Combinatorics: Proc. Cambridge Combinatorial Conference in honor of Paul Erdős*, Academic Press, pp. 35–57.
- [4] Bollobás, B. (2001) *Random Graphs*, second edition, Cambridge University Press.
- [5] Diestel, R. (2010) *Graph Theory*, fourth edition, Vol. 173 of *Graduate Texts in Mathematics*, Springer.
- [6] Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5** 17–61.
- [7] Gilbert, E. (1959) Random graphs. *Ann. Math. Statist.* **30** 1141–1144.
- [8] Glebov, R. and Krivelevich, M. (2013) On the number of Hamilton cycles in sparse random graphs. *SIAM J. Discrete Math.* **27** 27–42.
- [9] Hopcroft, J. and Tarjan, R. (1973) Algorithm 447: Efficient algorithms for graph manipulation. *Comm. Assoc. Comput. Mach.* **16** 372–378.
- [10] Komlós, J. and Szemerédi, E. (1983) Limit distributions for the existence of Hamilton circuits in a random graph. *Discrete Math.* **43** 55–63.
- [11] Korshunov, A. (1976) Solution of a problem of Erdős and Rényi on Hamiltonian cycles in non-oriented graphs. *Soviet Math. Dokl.* **17** 760–764.
- [12] Krivelevich, M., Lee, C. and Sudakov, B. (2015) Long paths and cycles in random subgraphs of graphs with large minimum degree. *Random Struct. Alg.* **46** 320–345.
- [13] Krivelevich, M. and Sudakov, B. (2013) The phase transition in random graphs: A simple proof. *Random Struct. Alg.* **43** 131–138.
- [14] Menger, K. (1927) Zur allgemeinen Kurventheorie. *Fund. Math.* **10** 96–115.
- [15] Pósa, L. (1976) Hamiltonian circuits in random graphs. *Discrete Math.* **14** 359–364.
- [16] Riordan, O. (2014) Long cycles in random subgraphs of graphs with large minimum degree. *Random Struct. Alg.* **45** 764–767.
- [17] West, D. (2007) *Introduction to Graph Theory*, Prentice Hall.