Long cycles in random subgraphs of graphs with large minimum degree

Oliver Riordan*

August 16, 2013; revised May 24, 2014

Abstract

Let G be any graph of minimum degree at least k, and let G_p be the random subgraph of G obtained by keeping each edge independently with probability p. Recently, Krivelevich, Lee and Sudakov showed that if $pk \to \infty$ then with probability tending to 1 G_p contains a cycle of length at least $(1-o(1))k$. We give a much shorter proof of this result, also based on depth-first search.

Let $(G_k)_{k=1}^{\infty}$ be a sequence of graphs where G_k has minimum degree at least k (we make no assumption on the number of vertices of G_k), and let $0 \leqslant p = p(k) \leqslant 1$. Let $G_k[p]$ be the random spanning subgraph of G_k obtained by retaining each edge with probability p, independently of the others. (Thus, if G_k is complete, then $G_k[p]$ is the classical binomial random graph.) Recently, Krivelevich, Lee and Sudakov [1] initiated the study of this very general model, asking which results for classical random graphs have analogues in this setting. As they point out, many questions do not make much sense in this generality – for example, no condition on p can ever ensure that $G_k[p]$ is likely to contain a short cycle, since the graphs G_k may have large girth. However, some do, and indeed have positive answers.

One of the main results in [1] is the following; here and later we often suppress the dependence on k in the notation. An event holds *with high probability* or *whp* if its probability tends to 1 as $k \to \infty$.

Theorem 1. *Let $G_k[p]$ be defined as above, and suppose that $pk \to \infty$ as $k \to \infty$. Then whp $G_k[p]$ contains a cycle of length at least $(1-o(1))k$.*

The aim of this note is to give a short proof of this result, based on the same basic starting point as in [1], but (sadly!) using relatively simple deterministic arguments to avoid most of the complicated probabilistic reasoning used there. Of course, the arguments in [1] give significant structural information about $G_k[p]$ that cannot be deduced from the simple proof given here.

Following [1], the first step is to explore $G_p = G_k[p]$ by depth-first search, revealing a *rooted spanning forest* T of G_p. In other words, T will be the disjoint union of one or more rooted trees, one spanning each component of G_p. For completeness, we describe the algorithm and its key implication, although this is of course discussed in detail in [1].

The search algorithm maintains a *stack* of vertices. Initially the stack is empty, and all vertices are *unreached*. Each vertex v will be 'reached' at some point, and placed

*Mathematical Institute, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK. E-mail: riordan@maths.ox.ac.uk.

onto the stack. At some later point v will be removed from the stack (never to return) and will be declared *explored*. More precisely, the algorithm executes the following loop, until every vertex is explored:

If the stack is empty, pick any unreached vertex v and place it onto the stack. v will be the root of a new component of T.

Otherwise, let u be the top vertex on the stack. If there are any not-yet-tested edges of $G = G_k$ from u to unreached vertices w, pick one such edge, and test whether it is present in G_p. If so, add w to the top of the stack, and add uw to T. If there are no not-yet-tested edges of G from u to unreached vertices, remove u from the stack and label it as explored.

The key point is that if we find a neighbour w of u, then we postpone looking for further neighbours of u until we have checked for neighbours of w, and so on. A consequence is that the vertices on the stack always represent a path in T. Let us call a path in a rooted tree *vertical* if it does not contain the root as an interior vertex. Let U be the set of edges of G_k not tested during the exploration. The following two lemmas are the key starting points for the arguments of Krivelevich, Lee and Sudakov [1]. We include proofs for completeness.

Lemma 2. *Every edge e of U joins two vertices on some vertical path in T.*

Proof. Let $e = uv$ and suppose that u was explored (left the stack) before v. When u left the stack, v cannot have been unreached, or uv would have been tested. Also, v cannot have been explored, by choice of u. So v was on the stack, and the stack from v to u forms the required vertical path. □

Suppressing the dependence on k, from now on we write $n = n(k) = |G_k|$ for the number of vertices of G_k.

Lemma 3. *With high probability, at most $2n/p = o(kn)$ edges are tested during the exploration above.*

Proof. Each time an edge is tested, the test succeeds (the edge is found to be present) with probability p. Comparison with a binomial distribution implies that the probability that more than $2n/p$ tests are made but fewer than n succeed is $o(1)$ (in fact, exponentially small in $n \geqslant k$). But every successful test contributes an edge to the forest T, so at most $n - 1$ tests are successful. □

From now on let us fix an arbitrary (small) constant $0 < \varepsilon < 1/10$. We call a vertex v *full* if it is incident with at least $(1 - \varepsilon)k$ edges in U.

Corollary 4. *With high probability, all but $o(n)$ vertices of T are full.*

Proof. Since G has minimum degree at least k, each $v \in V(G) = V(T)$ that is not full is incident with at least εk tested edges. If for some constant $c > 0$ there are at least cn such vertices, then there are at least $c\varepsilon kn/2$ tested edges. But the probability of this is $o(1)$ by Lemma 3. □

To state the new part of the argument, we need a couple of definitions. Let v be a vertex of a rooted forest T. Then there is a unique path from v to the root of its component. We write $A(v)$ for the set of *ancestors* of v, i.e., vertices (excluding v) on this path. We write $D(v)$ for the set of *descendants* of v, again excluding v. Thus

$w \in D(v)$ if and only if $v \in A(w)$. The *distance* $d(u,v)$ between two vertices u and v on a common vertical path is just their graph distance along this path. We write $A_i(v)$ and $D_i(v)$ for the set of ancestors/descendants of v at distance exactly i, and $A_{\leqslant i}(v)$, $D_{\leqslant i}(v)$ for those at distance at most i. The *height* of a vertex v is $\max\{i : D_i(v) \neq \emptyset\}$.

Let us call a vertex v *rich* if $|D(v)| \geqslant \varepsilon k$, and *poor* otherwise. In the next two lemmas, $(T_k)_{k=1}^{\infty}$ is a sequence of rooted forests. We suppress the dependence on k in the notation, and write n for $|T_k|$.

Lemma 5. *Suppose that T contains $o(n)$ poor vertices. Then, for any constant C, all but $o(n)$ vertices of T are at height at least Ck.*

Proof. For each rich vertex v, let $P(v)$ be a set of $\lceil \varepsilon k \rceil$ descendants of v, obtained by choosing vertices of $D(v)$ one-by-one starting with those furthest from v. For every $w \in P(v)$ we have $D(w) \subsetneq P(v)$, so $|D(w)| < \varepsilon k$, i.e., w is poor. Consider the set $\mathcal{S}_1$ of ordered pairs (v,w) with v rich and $w \in P(v)$. Each of the $(1-o(1))n$ rich vertices appears in at least εk pairs, so $|\mathcal{S}_1| \geqslant (1-o(1))\varepsilon kn$.

For any vertex w we have $|A_{\leqslant i}(w)| \leqslant i$, since there is only one ancestor at each distance, until we hit the root. Since $(v,w) \in \mathcal{S}_1$ implies that w is poor and $v \in A(w)$, and there are only $o(n)$ poor vertices, at most $o(Ckn) = o(kn)$ pairs $(v,w) \in \mathcal{S}_1$ satisfy $d(v,w) \leqslant Ck$. Thus $\mathcal{S}_1' = \{(v,w) \in \mathcal{S}_1 : d(v,w) > Ck\}$ satisfies $|\mathcal{S}_1'| \geqslant (1-o(1))\varepsilon kn$. Since each vertex v is the first vertex of at most $\lceil \varepsilon k \rceil \sim \varepsilon k$ pairs in $\mathcal{S}_1 \supset \mathcal{S}_1'$, it follows that $n - o(n)$ vertices v appear in pairs $(v,w) \in \mathcal{S}_1'$. Since any such v has height at least Ck, the proof is complete. □

Let us call a vertex v *light* if $|D_{\leqslant (1-5\varepsilon)k}(v)| \leqslant (1-4\varepsilon)k$, and *heavy* otherwise. Let H denote the set of heavy vertices in T.

Lemma 6. *Suppose that $T = T_k$ contains $o(n)$ poor vertices, and let $X \subset V(T)$ with $|X| = o(n)$. Then, for k large enough, T contains a vertical path P of length at least $\varepsilon^{-2}k$ containing at most $\varepsilon^2 k$ vertices in $X \cup H$.*

Proof. Let $\mathcal{S}_2$ be the set of pairs (u,v) where u is an ancestor of v and $0 < d(u,v) \leqslant (1-5\varepsilon)k$. Since a vertex has at most one ancestor at any given distance, we have $|\mathcal{S}_2| \leqslant (1-5\varepsilon)kn$. On the other hand, by Lemma 5 all but $o(n)$ vertices u are at height at least k and so appear in at least $(1-5\varepsilon)k$ pairs $(u,v) \in \mathcal{S}_2$. It follows that only $o(n)$ vertices u are in more than $(1-4\varepsilon)k$ such pairs, i.e., $|H| = o(n)$.

Let $\mathcal{S}_3$ denote the set of pairs (u,v) where $v \in X \cup H$, u is an ancestor of v, and $d(u,v) \leqslant \varepsilon^{-2}k$. Since a given v can only appear in $\varepsilon^{-2}k$ pairs $(u,v) \in \mathcal{S}_3$, we see that $|\mathcal{S}_3| \leqslant \varepsilon^{-2}k|X \cup H| = o(kn)$. Hence only $o(n)$ vertices u appear in more than $\varepsilon^2 k$ pairs $(u,v) \in \mathcal{S}_3$.

By Lemma 5, all but $o(n)$ vertices are at height at least $\varepsilon^{-2}k$. Let u be such a vertex appearing in at most $\varepsilon^2 k$ pairs $(u,v) \in \mathcal{S}_3$, and let P be the vertical path from u to some $v \in D_{\lceil \varepsilon^{-2}k \rceil}(u)$. Then P has the required properties. □

Proof of Theorem 1. Fix $0 < \varepsilon < 1/10$. It suffices to show that whp $G_p = G_k[p]$ contains a cycle of length at least $(1-5\varepsilon)k$, say.

Explore $G_p = G_k[p]$ by depth-first search as described above (i.e., as in [1]), writing T for the spanning forest revealed, and U for the set of untested edges. Condition on the result of the exploration, noting that the edges of U are still present independently with probability p. By Lemma 2, $uv \in U$ implies that u is either an ancestor or a descendant

of v. As before, let $n = |G_k| = |T|$. By Corollary 4, we may assume that all but $o(n)$ vertices are full, i.e., are incident with at least $(1-\varepsilon)k$ edges of U.

Suppose that

$$\big|\{u : uv \in U,\ d(u,v) \geqslant (1-5\varepsilon)k\}\big| \geqslant \varepsilon k \tag{1}$$

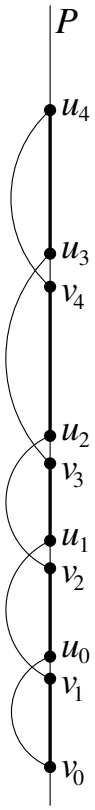
for some vertex v. Then, since $\varepsilon kp \to \infty$, testing the relevant edges uv one-by-one, whp we find one present in G_p, forming, together with T, the required long cycle. Suppose then that (1) fails for every v.

Suppose that some vertex v is full but poor. Since v has at most εk descendants, there are at least $(1-2\varepsilon)k$ pairs $uv \in U$ with $u \in A(v)$. Since v has only one ancestor at each distance, it follows that (1) holds for v, a contradiction.

We have shown that no poor vertex is full. Hence there are $o(n)$ poor vertices, and we may apply Lemma 6, with X the set of vertices that are not full. Let P be the path whose existence is guaranteed by the lemma, and let Z be the set of vertices on P that are full and light, so $|V(P) \setminus Z| \leqslant \varepsilon^2 k$.

For any $v \in Z$, since v is full, there are at least $(1-\varepsilon)k$ vertices $u \in A(v) \cup D(v)$ with $uv \in U$. Since (1) does not hold, at least $(1-2\varepsilon)k$ of these vertices satisfy $d(u,v) \leqslant (1-5\varepsilon)k$. Since v is light, in turn at least $2\varepsilon k$ of these u must be in $A(v)$. Recalling that a vertex has at most one ancestor at each distance, we find a set $U(v)$ of at least εk vertices $u \in A(v)$ with $uv \in U$ and $\varepsilon k \leqslant d(u,v) \leqslant (1-5\varepsilon)k \leqslant k$.

It is now easy to find a (very) long cycle whp. Recall that $Z \subset V(P)$ with $|V(P) \setminus Z| \leqslant \varepsilon^2 k$. Thinking of P as oriented upwards towards the root, let v_0 be the lowest vertex in Z. Since $|U(v_0)| \geqslant \varepsilon k$ and $kp \to \infty$, whp there is an edge v_0u_0 in G_p with $u_0 \in U(v_0)$. Let v_1 be the first vertex below u_0 along P with $v_1 \in Z$. Note that we go up at least εk steps from v_0 to u_0 and down at most $1 + |V(P) \setminus Z| \leqslant 2\varepsilon^2 k$ from u_0 to v_1, so v_1 is above v_0. Again whp there is an edge v_1u_1 in G_p with $u_1 \in U(v_1)$, and so at least εk steps above v_1. Continue downwards from u_1 to the first $v_2 \in Z$ below u_1, and so on. Since $\varepsilon^{-1} = O(1)$, whp we may continue in this way to find overlapping 'chords' v_iu_i for $0 \leqslant i \leqslant \lfloor 10\varepsilon^{-1} \rfloor$, say. (Note that we remain within P as each upwards step has length at most k.) These chords combine with P to form a cycle C, as shown in the figure. Since each chord v_iu_i corresponds to at least εk steps up P, and each of the segments v_iu_{i-1} of P has length at most $2\varepsilon^2 k$, we see that all of the $\lfloor 10\varepsilon^{-1} \rfloor + 1 \geqslant 10\varepsilon^{-1}$ segments of P included in C have length at least $(\varepsilon - 4\varepsilon^2)k \geqslant \varepsilon k/2$, so C has length at least $5k$.



References

[1] M. Krivelevich, C. Lee and B. Sudakov, Long paths and cycles in random subgraphs of graphs with large minimum degree, to appear in *Random Structures and Algorithms* (doi: 10.1002/rsa.20508). arXiv:1207.0312v3