Creating small subgraphs in Achlioptas processes with growing parameter

Michael Krivelevich*

School of Mathematical Sciences
Tel Aviv University
69978 Tel Aviv, Israel
krivelev@post.tau.ac.il

Reto Spöhel†

Algorithms and Complexity Group
Max-Planck-Institut für Informatik
66123 Saarbrücken, Germany
rspoehel@mpi-inf.mpg.de

Abstract. We study the problem of creating a copy of some fixed graph H in the Achlioptas process on n vertices with parameter r, where $r = r(n)$ is a growing function of n. We prove general upper and lower bounds on the threshold of this problem, and derive exact threshold functions for the case where H is a tree, a cycle, or the complete graph on four vertices.

1. Introduction

Consider the following random graph process: starting with the empty vertex set $V = [n]$, in each round we are offered r distinct vertex pairs sampled uniformly at random from all non-edges in the current graph. We are required to choose exactly one of the offered vertex pairs for inclusion in the evolving graph, immediately and irrevocably. Thus after N rounds, we have seen rN random vertex pairs (not necessarily all of them distinct), and selected exactly N distinct edges for inclusion in the graph. Our goal when selecting edges is to either *delay* or *accelerate* the occurence of some given monotone graph property by as much as possible, compared to the well-known Erdős-Rényi random graph process in which edges appear one by one uniformly at random.

This process was first considered in [5], and is known in the literature as the *Achlioptas process with parameter* r. It is the most prominent example of a random graph process that is not completely random but involves some limited amount of choice (by some 'player' or 'online algorithm'). Other graph processes of that type are the *Ramsey process* [3, 4, 6, 11, 16, 17], in which random edges appear one by one and have to be colored with one of r available colors, and the *Balanced Ramsey process* [12, 14, 15, 20], in which at each step r random edges appear and have be colored using each of the r available colors exactly once (this can be seen as a combination of the previous two processes). The idea of introducing some limited amount of control into an otherwise random setting has also proved very fruitful in various areas of computer science [18], most notably in the famous load-balancing result of [2].

The Achlioptas process has been studied by many researchers, both for fixed values of r and under the assumption that $r = r(n)$ is a growing function of n. The property that received by far the most attention in this context is the property of containing a linear-sized (so-called 'giant') component [1, 5, 7, 8, 10, 21, 22]. Only recently, other properties have been studied: the problem of accelerating Hamiltonicity in Achlioptas processes was investigated in [14], and the problem of delaying the occurrence of a given fixed graph as a subgraph was studied in [13, 19]. In this work we are concerned with the opposite problem of the latter; i.e., throughout this paper our goal is to *create* a copy of some given fixed graph H as quickly as possible.

*Research supported in part by USA-Israel BSF Grants 2006322 and 2010115, and by grant 1063/08 from the Israel Science Foundation. Part of this research was performed when the author was visiting ETH Zurich.

†Research supported by a fellowship of the Swiss National Science Foundation. Part of this research was performed when the author was still at ETH Zurich, and when the author visited Tel Aviv University.

A standard multi-round exposure argument (see e.g. [16, Lemma 7]) shows that for any graph H and any given $r = r(n)$, there exists a threshold function $N_0 = N_0(n)$ in the following sense: For any $N \gg N_0$,[1] there is an edge-selection strategy that succeeds a.a.s.[2] in creating a copy of H within N rounds of the Achlioptas process with parameter r, and for any $N \ll N_0$, any possible strategy a.a.s. fails to create a copy of H within N rounds of the process. Thus informally speaking, $N_0 = N_0(n)$ is the typical number of rounds an optimal edge-selection strategy needs to create a copy of H in the Achlioptas parameter with parameter r.

Note that the above defines $N_0(n)$ only up to constant factors. Nevertheless it is convenient to talk about 'the' threshold $N_0(H, r, n)$ of specific instances of the problem under study. When we say that this threshold satisfies $N_0(H, r, n) \le f(n)$, this is to be understood as 'For any $N \gg f(n)$, there is an edge-selection strategy that succeeds a.a.s. in creating a copy of H within N rounds of the process with parameter r', and similarly for the other direction.

Note that for $r = 1$ the Achlioptas process reduces to the already mentioned Erdős-Rényi process, in which edges appear one by one uniformly at random (and no edge-selection strategies are involved). The following classical result due to Bollobás states the threshold for the appearance of a copy of some fixed graph H in the Erdős-Rényi process. Throughout this paper, we say that a graph is nonempty if it has at least one edge. By $e(G)$ and $v(G)$ we denote the number of edges and vertices, respectively, of a graph G.

Theorem 1 (Bollobás [9]). *For any nonempty graph H, the threshold for the appearance of a copy of H in the Erdős-Rényi process is*

$$N_0(H, n) = n^{2-1/m(H)} \quad ,$$

where

$$m(H) := \max \left\{ \frac{e(J)}{v(J)} \;\middle|\; J \subseteq H \;\; \text{and} \;\; v(J) \ge 1 \right\} \quad .$$

From Theorem 1 we immediately obtain two general bounds on the threshold for creating copies of H in the Achliopas process with parameter r. On the one hand, by simply picking one of the r offered vertex pairs uniformly at random in each round, we can emulate the Erdős-Rényi process, and will therefore a.a.s. create a copy of H after any $N \gg n^{2-1/m(H)}$ rounds. On the other hand, if we imagine for a moment that in each round we are allowed to pick all r vertex pairs offered, and that moreover these are sampled from all vertex pairs never offered before, we obtain an Erdős-Rényi process with rN edges, which a.a.s. does not contain a copy of H if $rN \ll n^{2-1/m(H)}$. Hence the threshold for creating a copy of H in the Achlioptas process with parameter r satisfies

$$\frac{n^{2-1/m(H)}}{r} \le N_0(H, r, n) \le n^{2-1/m(H)} \quad . \tag{1}$$

Note that if r is a fixed integer, the two bounds coincide in order of magnitude and yield a threshold of $N_0(H, r, n) = n^{2-1/m(H)}$. In the following we therefore assume $r = r(n)$ to be a growing function of n. For convenience, we assume moreover that $r = r(n)$ is at most subpolynomial in n, i.e., $r = o(n^\varepsilon)$ for any fixed $\varepsilon > 0$. All of our results hold in fact under the weaker assumption that $r = o(n^\alpha)$ for some appropriate $\alpha > 0$, but the specific value of α differs from case to case, and trying to make this value explicit or to optimize it would introduce a number of unpleasant technicalities.

Our first theorem states two bounds that improve on the trivial bounds stated in (1) for every graph H with maximum degree at least two.

[1] We write $f \gg g$ for $f = \omega(g)$ and $f \ll g$ for $f = o(g)$.
[2] asymptotically almost surely, i.e. with probability tending to 1 as n tends to infinity

Theorem 2 (General bounds)**.** *For any nonempty graph H and any function $r = r(n)$ that grows at most subpolynomially, the threshold for creating a copy of H in the Achlioptas process with parameter r satisfies*

$$\frac{n^{2-1/m(H)}}{r^{1-1/k^*(H)}} \leq N_0(H,r,n) \leq \left(\frac{n}{\sqrt{r}}\right)^{2-1/m(H)} ,$$

where

$$k^*(H) := \min\left\{ e(J) \;\middle|\; J \subseteq H \;\; \text{and} \;\; v(J) \geq 1 \;\; \text{and} \;\; m(H) = \frac{e(J)}{v(J)} \right\}$$

(in words, $k^*(H)$ is the number of edges of a smallest densest subgraph of H).

Our next two results show that the lower bound in Theorem 2 is tight if H is a tree or the complete graph on four vertices.

Theorem 3 (Trees)**.** *For any nonempty tree T and any function $r = r(n)$ that grows at most subpolynomially, the threshold for creating a copy of T in the Achlioptas process with parameter r is*

$$N_0(T,r,n) = \left(\frac{n}{r}\right)^{1-1/e(T)} .$$

Theorem 4 (K_4)**.** *For any function $r = r(n)$ that grows at most subpolynomially, the threshold for creating a copy of the complete graph K_4 in the Achlioptas process with parameter r is*

$$N_0(K_4,r,n) = \frac{n^{4/3}}{r^{5/6}} .$$

Unfortunately, the proof of of Theorem 4 does not generalize to larger complete graphs; see Remark 9 below for an explanation of the underlying issue. Nevertheless we obtain an improved upper bound for the case where $H = K_5$ is the complete graph on five vertices. Note that Theorem 2 only yields an upper bound of

$$N_0(K_5,r,n) \leq \frac{n^{3/2}}{r^{3/4}} .$$

Theorem 5 (K_5)**.** *For any function $r = r(n)$ that grows at most subpolynomially, the threshold for creating a copy of the complete graph K_5 in the Achlioptas process with parameter r satisfies*

$$\frac{n^{3/2}}{r^{9/10}} \leq N_0(K_5,r,n) \leq \frac{n^{3/2}}{r^{7/8}}$$

(where the lower bound follows from Theorem 2).

The results we have presented so far do not rule out the possibility that the lower bound stated in Theorem 2 in fact determines the general threshold of the problem. Our last result shows that this is not the case, and that in general neither of the two bounds stated in Theorem 2 is tight.

Note that for the case where $H = C_\ell$ is a cycle of length ℓ, Theorem 2 yields the bounds

$$\frac{n}{r^{1-1/\ell}} \leq N_0(C_\ell,r,n) \leq \frac{n}{r^{1/2}} .$$

Theorem 6 (Cycles)**.** *For any integer $\ell \geq 3$ and any function $r = r(n)$ that grows at most subpolynomially, the threshold for creating a cycle C_ℓ of length ℓ in the Achlioptas process with parameter r is*

$$N_0(C_\ell,r,n) = \frac{n}{r^{\lceil \ell/2 \rceil/(\lceil \ell/2 \rceil+1)}} .$$

Note that the last result also shows that in general the threshold of the problem depends on more structural properties of H than just its maximum density $m(H)$. It would be very interesting to determine exact threshold functions for other fixed graphs H, in particular for cliques of arbitrary size.

Organization of this paper. We prove the general bounds stated in Theorem 2 in Section 2. We then prove our results for the case where H is a tree, a clique, and a cycle (in that order) in Sections 3–5. We give some concluding remarks in Section 6.

Preliminaries. We omit ceiling and floor signs whenever they are not essential.

Whenever we prove upper bounds on the threshold (i.e., analyze the performance of 'good' edge-selection strategies), we may and will assume w.l.o.g. that all rN vertex pairs are sampled (u.a.r. and independently) from *all* $\binom{n}{2}$ vertex pairs *with replacement*. Thus the player might be offered less than r distinct vertex pairs in a given round, and he might be offered vertex pairs that are already present as an edge in the current graph.

In our upper bound proofs we will often divide the process into a constant number of phases, with a specific goal to be achieved in each phase. The following lemma conveniently allows us to analyze what happens in a given phase without looking at the individual rounds making up that phase. On an intuitive level, it states that if the 'good' edges occur not too frequently, we can take almost all of the good edges we ever see. In other words, we lose very little by only being allowed to take an $1/r$-fraction of the vertex pairs we see.

Lemma 7. *Consider the Achlioptas process with parameter $r = r(n) \to \infty$, and let $\mathcal{G} \in \binom{[n]}{2}$ denote a set of 'good' vertex pairs. If $|\mathcal{G}| = \mathcal{O}(n^2/r)$ and $N = \mathcal{O}(n^2/r)$ are such that $|\mathcal{G}| \cdot Nr/n^2 \gg 1$, then there is an edge-selection strategy that ensures that after N rounds of the process, the resulting graph a.a.s. contains a uniformly random subset of $\mathcal{G}$ of size $\Theta(|\mathcal{G}| \cdot Nr/n^2)$.*

Proof. Let $p := |\mathcal{G}|/\binom{n}{2}$, and note that our assumption on $\mathcal{G}$ implies that $p = \mathcal{O}(1/r)$. Our strategy is the following: Whenever exactly one of the r presented vertex pairs is in $\mathcal{G}$, we take it; otherwise, we take an arbitrary edge (and ignore it if happens to be in $\mathcal{G}$). The probability that we take a random edge from $\mathcal{G}$ in a given round is $rp(1-p)^{r-1} = \Theta(rp)$, where we used that $pr = \mathcal{O}(1)$. Consequently, the probability p' that a fixed vertex pair in $|\mathcal{G}|$ is picked in a fixed round of the process is $p' = \Theta(rp/|\mathcal{G}|) = \Theta(r/n^2)$. For simplicity, we only count the vertex pairs in $\mathcal{G}$ that are picked exactly once. The probability that a given vertex pair in $\mathcal{G}$ is picked exactly once is $N \cdot p'(1-p')^{N-1} = \Theta(Np') = \Theta(Nr/n^2)$, where we used that $p'N = \Theta(Nr/n^2) = \mathcal{O}(1)$. Hence by standard Chernoff bounds, a.a.s. the number of distinct vertex pairs from $\mathcal{G}$ picked is $\Theta(|\mathcal{G}| \cdot Nr/n^2)$, where we used that $|\mathcal{G}| \cdot Nr/n^2 \gg 1$. By symmetry, the subset of $\mathcal{G}$ picked in this way is distributed uniformly. $\square$

2. General bounds

In this section we prove Theorem 2.

Proof of Theorem 2 (upper bound). Let $N \gg (n/\sqrt{r})^{2-1/m(H)}$ be given, and assume w.l.o.g. that moreover $N \ll n^2/r$. We fix a set $S \subseteq V$ of size $s := n/\sqrt{r}$, and select an edge inside S whenever possible (if we have a choice between several edges inside S, we take one of them uniformly at random). The probability that we can do so in a given round is

$$1 - \left(1 - \frac{\binom{s}{2}}{\binom{n}{2}}\right)^r \geq 1 - \mathrm{e}^{-s^2 r/n^2} \geq 1/2 \ ,$$

so by Chernoff bounds a.a.s. we can pick an edge inside S in at least $N/3$ rounds. Moreover, by a similar calculation as in the proof of Lemma 7 the number of edges in S picked several times is negligible. Hence we obtain an Erdős-Rényi process with $\Theta(N)$ edges on S, which by Theorem 1 a.a.s. forms a copy of H for N as given. $\square$

For the proof of the lower bound in Theorem 2, we first prove the following auxiliary lemma.

Lemma 8. *Let J be a nonempty graph, and let $r = r(n)$ be a function that grows at most subpolynomially. For any $N \ll \frac{n^{2-v(J)/e(J)}}{r^{1-1/e(J)}}$, any possible strategy a.a.s. fails to create a copy of J within N rounds of the Achlioptas process with parameter r.*

Let us give some intuition for the upcoming proof of Lemma 8. By a standard calculation, the expected number of copies of J in the graph formed by all rN edges seen (not necessarily picked) is of order

$$n^{v(J)} \left(\frac{N}{n^2}\right)^{e(J)} r^{e(J)} \quad .$$

We will show by an inductive argument that no strategy can expect to create more than an r^{-1}-fraction of this number of copies. To this end, we think of the copies of J as being created in $e(J)$ steps, one for each edge that arrives. The key observation is that any strategy will 'lose' a factor of r in the very first step, as it sees rN random edges but can take only N of them. (This corresponds to the base case $e(J) = 1$ of our induction.) As our argument shows, this missing factor of r cannot be recovered in later steps, even if everything goes as well as one can hope for.

Proof. Consider an arbitrary fixed strategy. We prove by induction on $e(J)$ that the expected number of copies of J created in N rounds is at most of order

$$n^{v(J)} \left(\frac{N}{n^2}\right)^{e(J)} r^{e(J)-1} \quad .$$

Note that this implies Lemma 8 by Markov's inequality.

Clearly, the inductive statement is true if $e(J) = 1$. For the induction step, let $\mathcal{J}_-$ denote the family of graphs obtained by removing exactly one edge (and no vertices) from J. Let X denote the random variable that counts the total number of copies of graphs from $\mathcal{J}_-$ created by our fixed strategy in N rounds of the process. Clearly, the number of copies of J created is bounded by the number of edges seen (not necessarily picked!) that complete a copy of a graph $J_- \in \mathcal{J}_-$ to a copy of J. It follows that *conditional on* X the expected number of copies of J created is at most

$$rN \cdot cX / \binom{n}{2} = X \cdot 3c \cdot \frac{rN}{n^2}$$

for some appropriate constant $c = c(J)$. As by induction we have

$$\mathbb{E}[X] = \mathcal{O}\left(n^{v(J)} \left(\frac{N}{n^2}\right)^{e(J)-1} r^{e(J)-2}\right) \quad ,$$

the claimed bound on the expected number of copies of J follows.

This proves the inductive statement, and as mentioned the lemma follows by Markov's inequality. □

Proof of Theorem 2 (lower bound). Applying Lemma 8 for a graph

$$J \in \arg\min \left\{ e(J) \;\middle|\; J \subseteq H \text{ and } v(J) \geq 1 \text{ and } m(H) = \frac{e(J)}{v(J)} \right\}$$

shows that for $N \ll n^{2-1/m(H)}/r^{1-1/k^*(H)}$, any possible strategy a.a.s. fails to create a copy of J, and hence also fails to create a copy of H. □

3. Trees

Observing that the lower bound in Theorem 3 follows from Theorem 2, it remains to prove the upper bound part.

Proof of Theorem 3 *(upper bound).* Let $N \gg \left(\frac{n}{r}\right)^{1-1/e(T)}$ be given, and assume w.l.o.g. that moreover $N \ll n/r$.

We prove by induction on $e(T)$ that in N rounds of the process we can a.a.s. create a family $\mathcal{T}$ of pairwise vertex-disjoint copies of T, of size

$$|\mathcal{T}| = \Theta\left(N^{e(T)}\left(\frac{r}{n}\right)^{e(T)-1}\right) \quad .$$

As this is a growing function of n for N as given, the claimed upper bound follows.

Using that $N \ll n$, it is not hard to see that the inductive claim is true if T is a single edge (it suffices to select edges randomly and consider all isolated edges). For the induction step, we divide the process in two phases of length $N/2$ each. Let T_- denote an arbitrary tree obtained by removing exactly one leaf from T. Applying the induction hypothesis, in the first phase a.a.s. we can create a family $\mathcal{T}_-$ of pairwise vertex-disjoint copies of T_-, of size $|\mathcal{T}_-| = \Theta\left(N^{e(T)-1}\left(\frac{r}{n}\right)^{e(T)-2}\right)$. Note that due to our assumption that $N = \mathcal{O}(n/r)$ we have $|\mathcal{T}_-| \ll n/r$.

Let $V' \subseteq V$ be the set of vertices that are not part of one of these copies, and note that $|V'| = n - \Theta(|\mathcal{T}_-|) = \Theta(n)$. In the second phase, we consider a vertex pair good if it connects a copy of T_- in $\mathcal{T}_-$ and a vertex in V' to a copy of T. Note that the number of such good edges is $\Theta(|\mathcal{T}_-| \cdot n) = \mathcal{O}(n^2/r)$. Hence by Lemma 7, a.a.s. we can take a random subset of size $t = \Theta(|\mathcal{T}_-| \cdot n \cdot Nr/n^2) = \Theta\left(N^{e(T)}\left(\frac{r}{n}\right)^{e(T)-1}\right)$ of these good edges. Moreover, the endpoints of these edges are distributed independently and u.a.r. in V' on one side, and independently and u.a.r. among the appropriate vertices of the copies of T_- on the other side. We can view this as throwing t balls into $|V'| = \Theta(n)$ bins on one side, and independently into $\Theta(|\mathcal{T}_-|)$ bins on the other side. Using that $t \ll n$ and $t \ll |\mathcal{T}_-|$, standard arguments yield that a.a.s. on both sides there are at least $2t/3$ 'balls' that end up in pairwise different 'bins'. Hence we obtain at least $t/3$ pairwise vertex-disjoint copies of T a.a.s., which proves the inductive claim. □

4. Cliques

Observing that the lower bound in Theorem 4 follows from Theorem 2, it remains to prove the upper bound part.

Proof of Theorem 4 *(upper bound).* Let $r = r(n)$ growing subpolynomially and $N \gg \frac{n^{4/3}}{r^{5/6}}$ be given, and assume w.l.o.g. that moreover $N = \mathcal{O}(n^{3/2}/r)$.

Our strategy proceeds in two phases as follows. In the first phase we fix a set $S \subseteq V$ of size $s := n/r$, and consider a vertex pair good if exactly one of its vertices is in S. The number of such good vertex pairs is $s(n-s) = \Theta(n^2/r)$. Hence by Lemma 7 we can take $\Theta(n^2/r \cdot Nr/n^2) = \Theta(N)$ distinct such good edges a.a.s., yielding an average degree of $\Theta(N/s) = \Theta(Nr/n)$ for the vertices of S into $V \setminus S$. By standard Chernoff bound arguments, it follows that a.a.s. each vertex in S has degree at least cNr/n into $V \setminus S$ for some appropriate constant $c > 0$. We will condition on this degree property throughout the remainder of this proof.

For each vertex $v \in S$, we let $D_v \subseteq V \setminus S$ denote the endpoints of the first cNr/n edges incident to v that are picked. Thus we have exactly s sets D_v, $v \in S$, of size exactly $d := cNr/n$ each, sampled independently and uniformly at random from $V \setminus S$. In particular, a given vertex in $V \setminus S$ has a probability of

$$p_1 := \frac{d}{|V \setminus S|} = \Theta(Nr/n^2)$$

to be contained in D_v for some specific vertex $v \in S$.

In the second phase, we consider a vertex pair good if both its vertices are in the same set D_v, $v \in S$; i.e., we define

$$\mathcal{G} := \bigcup_{v \in S} \binom{D_v}{2} .$$

Observing that $|\mathcal{G}| \leq |S| \cdot \binom{d}{2} = \Theta(N^2 r/n) = \mathcal{O}(n^2/r)$ by our assumption that $N = \mathcal{O}(n^{3/2}/r)$, we obtain with Lemma 7 that a.a.s. we can take $m := c' \cdot |\mathcal{G}| \cdot Nr/n^2$ distinct vertex pairs from $\mathcal{G}$, for some appropriate constant $c' > 0$. In particular, a fixed good vertex pair is picked with probability

$$p_2 := \frac{m}{|\mathcal{G}|} = \Theta(Nr/n^2) .$$

Note that whenever three vertex pairs forming a triangle in some set D_v are picked, they form a copy of K_4 with the corresponding vertex $v \in S$. Let the random variable X denote the number of K_4's that are completed in this way during the second phase. We will show by the second moment method that (conditional on a 'good' outcome of the first phase) a.a.s. we have $X \geq 1$. Defining

$$\mathcal{T} := \bigcup_{v \in S} \left(\{v\} \times \binom{D_v}{3} \right) ,$$

we have

$$\mathbb{E}[X] = |\mathcal{T}| \cdot \frac{\binom{|\mathcal{G}|-3}{m-3}}{\binom{|\mathcal{G}|}{m}} = \Theta\left(s \cdot d^3 \cdot (p_2)^3\right) = \Theta\left(\frac{N^6 r^5}{n^8}\right) \to \infty \tag{2}$$

(where the randomness is that of the second phase, and we conditioned on the first phase satisfying the mentioned degree property).

To calculate the variance of X, we distinguish five different types of pairs of elements from $\mathcal{T}$:

(0) pairs of form $\big((v_1, T_1), (v_2, T_2)\big)$, where $v_1 \neq v_2 \in S$ and $T_1 \in \binom{D_{v_1}}{3}$, $T_2 \in \binom{D_{v_2}}{3}$ with $|T_1 \cap T_2| \leq 1$.

Note that for such pairs the two corresponding indicator random variables are negatively correlated, and that therefore the contribution of these pairs to the variance is negative.

(1) trivial pairs $\big((v, T), (v, T)\big)$, where $v \in S$ and $T \in \binom{D_v}{3}$,
(2) pairs of form $\big((v, T_1), (v, T_2)\big)$, where $v \in S$ and T_1, $T_2 \in \binom{D_v}{3}$ with $|T_1 \cap T_2| = 2$,
(3) pairs of form $\big((v_1, T), (v_2, T)\big)$, where $v_1 \neq v_2 \in S$ and $T \in \binom{D_{v_1}}{3} \cap \binom{D_{v_2}}{3}$,
(4) pairs of form $\big((v_1, T_1), (v_2, T_2)\big)$, where $v_1 \neq v_2 \in S$ and $T_1 \in \binom{D_{v_1}}{3}$, $T_2 \in \binom{D_{v_2}}{3}$ with $|T_1 \cap T_2| = 2$.

Letting A_i denote the number of pairs of type i, $i = 1, \ldots, 4$, by similar calculations as in (2) the variance of X satisfies

$$\mathrm{Var}[X] = \mathcal{O}\Big((A_1 + A_3) \cdot (p_2)^3 + (A_2 + A_4) \cdot (p_2)^5\Big) . \tag{3}$$

Clearly, we have

$$A_1 = s \cdot \binom{d}{3} \qquad\quad = \Theta(sd^3) = \Theta(sn^3(p_1)^3) \ ,$$
$$A_2 = s \cdot \binom{d}{4} \cdot \Theta(1) = \Theta(sd^4) = \Theta(sn^4(p_1)^4) \ .$$

(Recall that we condition on the first phase satisfying the degree property.)

The numbers A_3 and A_4 of pairs of type 3 and 4 are random variables that depend on the outcome of the first phase. Their expected values are

$$\mathbb{E}[A_3] = \binom{|S|}{2} \cdot \binom{|V \setminus S|}{3} \cdot \Theta(1) \cdot \left(\frac{\binom{|V\setminus S|-3}{d-3}}{\binom{|V\setminus S|}{d}} \right)^2 = \Theta\left(s^2n^3(p_1)^6\right) \ ,$$
$$\mathbb{E}[A_4] = \binom{|S|}{2} \cdot \binom{|V \setminus S|}{4} \cdot \Theta(1) \cdot \left(\frac{\binom{|V\setminus S|-3}{d-3}}{\binom{|V\setminus S|}{d}} \right)^2 = \Theta\left(s^2n^4(p_1)^6\right) \ .$$

Letting ω_n denote a very slowly growing function of n (say $\omega_n = \log r$ for concreteness), we have by Markov's inequality that a.a.s.

$$A_3 \leq \omega_n s^2 n^3 (p_1)^6 \ , \tag{4a}$$
$$A_4 \leq \omega_n s^2 n^4 (p_1)^6 \tag{4b}$$

(where the randomness is that of the first phase).

For the remaining calculations, note that both p_1 and p_2 are of order $p := rN/n^2$. Plugging our bounds for A_i, $i = 1, \ldots, 4$, into (3), we obtain that a.a.s. the variance of X in the second phase satisfies

$$\mathrm{Var}[X] = \mathcal{O}\left(sn^3p^6 + [\omega_n s^2n^3p^9] + sn^4p^9 + \omega_n s^2n^4p^{11}\right) \ ,$$

where the term in brackets is dominated due to $\omega_n s/n = o(1)$. Observing that according to (2) we have

$$\mathbb{E}[X] = \Theta(sd^3p^3) = \Theta(sn^3p^6) \to \infty$$

we finally obtain

$$\frac{\mathrm{Var}[X]}{(\mathbb{E}[X])^2} = \mathcal{O}\left(\frac{1}{sn^3p^6} + \frac{1}{sn^2p^3} + \frac{\omega_n}{n^2p}\right) = o(1) \ .$$

It follows by Chebyshev's inequality that a.a.s. $X \geq 1$, where the randomness is that of the second phase and we conditioned on the first phase satisfying the degree property and (4). □

Remark 9. In the above proof we started with a large set S of vertices of degree $\Theta(Nr/n)$, and then constructed copies of K_4 by considering the triangles formed in the neighborhoods of the vertices of S. Note that $|S|$ must be $\mathcal{O}(n/r)$, as we only have N edges in total.

For our argument to work it was crucial that there were at most $\mathcal{O}(n^2/r)$ good vertex pairs in the second phase so we could apply Lemma 7. This was ensured by our assumption that $N = \mathcal{O}(n^{3/2}/r)$. Unfortunately, we cannot make this assumption when trying to create copies of larger complete graphs, as in these cases the interesting range for N is higher than $n^{3/2}/r$. However, for complete graphs of size five we can still obtain a non-trivial upper bound due to the fact that the range of N we need to consider only exceeds $n^{3/2}/r$ by a factor polynomial in r. Specifically, we ensure that there are only $\mathcal{O}(n^2/r)$ good edges in the second phase by choosing S slightly smaller than $\Theta(n/r)$.

Proof of Theorem 5 *(upper bound).* Let $r = r(n)$ growing subpolynomially and $N \gg \frac{n^{3/2}}{r^{7/8}}$ be given, and assume w.l.o.g. that moreover $N = \mathcal{O}(n^2/r)$.

Our strategy proceeds in two phases similarly to before. In the first phase we fix a set $S \subseteq V$ of size

$$s := \frac{n^4}{N^2 r^3} \quad \left[\ll \frac{n}{r^{5/4}}\right] \ ,$$

and consider a vertex pair good if exactly one of its vertices is in S. By the same arguments as in the previous proof, we obtain sets D_v, $v \in S$, with the same properties as before. As before, we use the notations $d = |D_v| = \Theta(Nr/n)$ and $p_1 = d/|V \setminus S| = \Theta(Nr/n^2)$.

In the second phase, we consider a vertex pair good if both its vertices are in the same set D_v, $v \in S$; i.e., we define

$$\mathcal{G} := \bigcup_{v \in S} \binom{D_v}{2} \ .$$

Observing that $|\mathcal{G}| \leq |S| \cdot \binom{d}{2} = \Theta(n^2/r)$ by our choice of s (!), we obtain with Lemma 7 that a.a.s. we can take $m := c' \cdot |\mathcal{G}| \cdot Nr/n^2$ distinct vertex pairs from $\mathcal{G}$, for some appropriate constant c'. In particular, a fixed good vertex pair is picked with probability $p_2 = m/|\mathcal{G}| = \Theta(Nr/n^2)$ as in the previous proof.

Note that whenever six vertex pairs forming a K_4 in some set D_v are picked, they form a copy of K_5 with the corresponding vertex $v \in S$. Let the random variable X denote the number of K_5's that are completed in this way during the second phase. We will show by the second moment method that (conditional on a 'good' outcome of the first phase) a.a.s. we have $X \geq 1$. Defining

$$\mathcal{T} := \bigcup_{v \in S} \left(\{v\} \times \binom{D_v}{4}\right) \ ,$$

we obtain similarly to the previous proof

$$\mathbb{E}[X] = |\mathcal{T}| \cdot \frac{\binom{|\mathcal{G}|-4}{m-4}}{\binom{|\mathcal{G}|}{m}} = \Theta\left(s \cdot d^4 \cdot (p_2)^6\right) = \Theta\left(\frac{N^8 r^7}{n^{12}}\right) \to \infty \ . \tag{5}$$

To calculate the variance of X, we distinguish seven different types of pairs of elements from $\mathcal{T}$:

(0) pairs of form $\big((v_1, T_1), (v_2, T_2)\big)$, where $v_1 \neq v_2 \in S$ and $T_1 \in \binom{D_{v_1}}{4}$, $T_2 \in \binom{D_{v_2}}{4}$ with $|T_1 \cap T_2| \leq 1$.

As before, the contribution of these pairs to the variance is negative.

(1) trivial pairs $\big((v, T), (v, T)\big)$, where $v \in S$ and $T \in \binom{D_v}{4}$,
(2) pairs of form $\big((v, T_1), (v, T_2)\big)$, where $v \in S$ and T_1, $T_2 \in \binom{D_v}{4}$ with $|T_1 \cap T_2| = 3$,
(3) pairs of form $\big((v, T_1), (v, T_2)\big)$, where $v \in S$ and T_1, $T_2 \in \binom{D_v}{4}$ with $|T_1 \cap T_2| = 2$,
(4) pairs of form $\big((v_1, T), (v_2, T)\big)$, where $v_1 \neq v_2 \in S$ and $T \in \binom{D_{v_1}}{4} \cap \binom{D_{v_2}}{4}$,
(5) pairs of form $\big((v_1, T_1), (v_2, T_2)\big)$, where $v_1 \neq v_2 \in S$ and $T_1 \in \binom{D_{v_1}}{4}$, $T_2 \in \binom{D_{v_2}}{4}$ with $|T_1 \cap T_2| = 3$,
(6) pairs of form $\big((v_1, T_1), (v_2, T_2)\big)$, where $v_1 \neq v_2 \in S$ and $T_1 \in \binom{D_{v_1}}{4}$, $T_2 \in \binom{D_{v_2}}{4}$ with $|T_1 \cap T_2| = 2$.

Letting A_i denote the number of pairs of type i, $i = 1, \ldots, 6$, the variance of X satisfies

$$\mathrm{Var}[X] = \mathcal{O}\Big((A_1 + A_4) \cdot (p_2)^6 + (A_2 + A_5) \cdot (p_2)^9 + (A_3 + A_6) \cdot (p_2)^{11}\Big) \ , \tag{6}$$

where similarly to the previous proof

$$\begin{aligned} A_1 &= \Theta(sd^4) = \Theta(sn^4(p_1)^4) \ , \\ A_2 &= \Theta(sd^5) = \Theta(sn^5(p_1)^5) \ , \\ A_3 &= \Theta(sd^6) = \Theta(sn^6(p_1)^6) \ , \end{aligned}$$

and a.a.s.

$$A_4 \leq \omega_n s^2 n^4 (p_1)^{12} \ , \tag{7a}$$

$$A_5 \leq \omega_n s^2 n^5 (p_1)^{12} \ , \tag{7b}$$

$$A_6 \leq \omega_n s^2 n^6 (p_1)^{12} \ . \tag{7c}$$

For the remaining calculations, note that both p_1 and p_2 are of order $p := rN/n^2$. Plugging our bounds for A_i, $i = 1, \ldots, 6$, into (6), we obtain that a.a.s. the variance of X in the second phase satisfies

$$\mathrm{Var}[X] = \mathcal{O}\left(sn^4p^{10} + [\omega_n s^2 n^4 p^{18}] + sn^5p^{14} + [\omega_n s^2 n^5 p^{21}] + sn^6 p^{17} + \omega_n s^2 n^6 p^{23}\right) \ ,$$

where the two terms in brackets are dominated due to $\omega_n s/n = o(1)$ and $p \leq 1$. Observing that according to (5) we have

$$\mathbb{E}[X] = \Theta(sd^4p^6) = \Theta(sn^4p^{10}) \to \infty \ ,$$

we obtain

$$\frac{\mathrm{Var}[X]}{(\mathbb{E}[X])^2} = \mathcal{O}\left(\frac{1}{sn^4p^{10}} + \frac{1}{sn^3p^6} + \frac{1}{sn^2p^3} + \frac{\omega_n p^3}{n^2}\right) = o(1) \ .$$

It follows by Chebyshev's inequality that a.a.s. $X \geq 1$, where the randomness is that of the second phase and we conditioned on the first phase satisfying the degree property and (7). $\square$

5. Cycles

We first prove the upper bound in Theorem 6, proceeding along similar lines as in the previous proofs. Recall that in our upper bound proofs we assume w.l.o.g. that all rN vertex pairs are sampled (u.a.r. and independently) from all $\binom{n}{2}$ vertex pairs with replacement.

Proof of Theorem 6 *(upper bound).* Set $k := \lceil \ell/2 \rceil$, let $N \gg \frac{n}{r^{k/(k+1)}}$ be given, and assume w.l.o.g. that moreover $N \ll n$. Define $x_i := N \cdot \left(\frac{n}{Nr}\right)^{k-1-i}$, $i = 0, \ldots, k-1$, and note that

$$1 \ll \frac{n}{r^{k-1}} \ll x_0 \ll x_1 \ll \ldots \ll x_{k-2} = \frac{n}{r} \ . \tag{8}$$

We distinguish two cases depending on whether ℓ is odd or even.

If $\ell = 2k-1$ is odd, we divide the process in k phases of equal length N/k.

In the first phase, we fix a set $X_0 \subseteq V$ of size $|X_0| = x_0$, and consider a vertex pair good if it has exactly one endpoint in X_0. The number of such vertex pairs is $\Theta(x_0 n)$, which is $\mathcal{O}(n^2/r)$ by (8). Lemma 7 thus guarantees that a.a.s. we can take $\Theta(x_0 n \cdot Nr/n^2) = \Theta(x_1)$ distinct good edges uniformly at random, yielding an average degree of $\Theta(Nr/n)$ [$\gg r^{1/(k+1)}$] for the vertices of X_0 into $V \setminus X_0$. By standard Chernoff bound arguments, a.a.s. at least half of the vertices in X_0 have degree at least $d := cNr/n$ into $V \setminus X_0$ for some appropriate constant $c > 0$, and the corresponding endpoints in $V \setminus X_0$ are distributed uniformly at random. We fix a set X_0' of such vertices, $|X_0'| = |X_0|/2$, and consider for each vertex in X_0' only the first d (distinct) incident edges. Let X_1 denote the set of vertices in $V \setminus X_0$ that are incident to exactly one of those $|X_0'| \cdot d$ edges. Letting $p := d/|V \setminus X_0|$, the probability that a fixed vertex in $V \setminus X_0$ is in X_1 is

$|X_0'|p(1-p)^{|X_0'|-1} = \Theta(|X_0'| \cdot p)$, where we used that $|X_0'| \cdot p = \mathcal{O}(1)$. Hence by Chernoff bounds we have $|X_1| = \Theta(|X_0'|pn) = \Theta(x_1)$ a.a.s.

In the second phase we consider a vertex pair good if it has one endpoint in X_1, and the other one in $V \setminus (X_0 \cup X_1)$. By the same arguments as before, we a.a.s. get a set $X_1' \subseteq X_1$, $|X_1'| = |X_1|/2$, of vertices which have degree $\Theta(Nr/n)$ into $V \setminus (X_0 \cup X_1)$, and a set $X_2 \subseteq V \setminus (X_0 \cup X_1)$ of vertices which are incident to exactly one vertex of X_1', of size $|X_2| = \Theta(|X_1| \cdot Nr/n) = \Theta(x_2)$.

Continuing like this for $k-1$ phases in total, we eventually obtain a set $X_{k-1} \subseteq V \setminus (X_0 \cup \cdots \cup X_{k-2})$ of size $|X_{k-1}| = \Theta(x_{k-1}) = \Theta(N)$. By construction, from each vertex in this set there is a unique path of length $k-1$ to a vertex in X_0', and from there there are $\Theta((Nr/n)^{k-1})$ paths to other vertices in X_{k-1}. In total, we have thus created $\Theta(N \cdot (Nr/n)^{k-1})$ paths of length $2k-2 = \ell - 1$, and no two such paths have the same vertex pair as their endpoints.

In the last phase of the process we consider a vertex pair good if it completes such a path to a cycle of length ℓ. Note that we are done as soon as we are presented one such good vertex pair. The probability p' that a random vertex pair is good is $p' = \Theta(N \cdot (Nr/n)^{k-1})/\binom{n}{2}$, and consequently the expected number of good vertex pairs seen is

$$N/k \cdot r \cdot p' = \Theta\left(\frac{N^{k+1}r^k}{n^{k+1}}\right) \quad .$$

Due to $N \gg \frac{n}{r^{k/(k+1)}}$ this expectation tends to infinity, and by Chernoff bounds we can complete a cycle of length ℓ a.a.s. in this last phase.

If $\ell = 2k$ is even, we divide the process into $k+1$ phases of equal length. In the first $k-1$ phases, we proceed exactly as in the proof for the case where ℓ is odd. In the k-th phase however, we continue our path-building strategy for one more phase and consider an edge good if it has one endpoint in X_{k-1} and the other one in $V \setminus (X_0 \cup \cdots \cup X_{k-1})$. As $|X_{k-1}| = \Theta(N) \gg n/r$ we can no longer apply Lemma 7. Instead we argue as follows: The probability that we cannot take a good edge in a given round is $(1 - \frac{\Theta(nN)}{n^2})^r \leq \mathrm{e}^{-\Theta(Nr/n)} = o(1)$. Hence by Chernoff bounds a.a.s. we can take a good edge in $\Theta(N)$ rounds, yielding constant average degree for the vertices of X_{k-1} into $V \setminus (X_0 \cup \cdots \cup X_{k-1})$. Let X_k denote the set of all vertices in $V \setminus (X_0 \cup \cdots \cup X_{k-1})$ that are incident to exactly one vertex of X_{k-1}. A similar argument as before shows that $|X_k| = \Theta(N)$ a.a.s. Thus we now have created $\Theta(N \cdot (Nr/n)^{k-1})$ paths of length $2k-1 = \ell - 1$, and no two such paths have the same vertex pair as their endpoints. Hence in the last phase the proof can be completed exactly as before. $\square$

For the proof of a matching lower bound, we first bound the number of paths of length $\ell - 1$ that any strategy can create, and then apply Markov's inequality to infer that a.a.s. none of the rN random vertex pairs seen creates a cycle of length ℓ.

For the first part of the argument, we relax the setting as follows: the player is shown exactly $m := rN$ distinct edges at once, and is allowed to select an arbitrary set of exactly N edges from those. Thus we work in an offline setting, and ignore the restriction that the vertex pairs come in sets of size r and the player has to select exactly one edge from each such set. As usual, we denote by $G_{n,m}$ a graph drawn uniformly at random from all graphs on n vertices with m edges.

The next two lemmas establish two properties that hold a.a.s. for $G_{n,m}$ with $m = rN$ and N slightly above the threshold we wish to prove. The desired bound on the number of paths of length $\ell - 1$ will follow deterministically from these properties.

Lemma 10. *Consider $G_{n,m}$ with $m \gg n$, and let X_k denote the number of vertices of degree k. Then a.a.s. we have*

$$X_k \leq n(2/3)^k$$

for all $k \geq 12m/n$ simultaneously.

Proof. A fixed vertex has expected degree $2m/n$. By a Chernoff-type bound, a fixed vertex has degree $k \geq 12m/n$ with probability at most 2^{-k}. It follows with Markov's inequality that

$$\Pr[X_k \geq n(2/3)^k] \leq \frac{\mathbb{E}[X_k]}{n(2/3)^k} \leq (3/4)^k \ .$$

Taking a union bound over all $k \geq 12m/n$ and using that $m \gg n$ concludes the proof. $\square$

We say that a graph G has *maximum density at most d* if $e(H)/v(H) \leq d$ for all subgraphs $H \subseteq G$ (i.e., if $m(G)$ as defined in Theorem 1 is at most d).

Lemma 11. *For any $\varepsilon > 0$ there exists $d = d(\varepsilon)$ such that the following holds. For any $r \gg 1$ and $N \leq \frac{n}{r^{1/2+\varepsilon}}$ the following is true: A.a.s. the random graph $G_{n,rN}$ is such that every subgraph with at most N edges has maximum density at most d.*

Proof. Fix d such that $(2d-1)\varepsilon > 1/2$. We will show that a.a.s. $G_{n,rN}$ satisfies the following: for $v = 1, \ldots, 2N$, no set of v vertices induces more than dv edges in $G_{n,rN}$. Clearly, this implies the claim for any subgraph with at most N edges.

The expected number of vertex sets violating the above statement is bounded by

$$\sum_{v=1}^{2N} \binom{n}{v} \binom{\binom{v}{2}}{dv} \left(\frac{3rN}{n^2}\right)^{dv} \leq \sum_{v=1}^{2N} \left(\frac{en}{v}\right)^v \left(\frac{3evrN}{2dn^2}\right)^{dv} \leq \sum_{v=1}^{2N} \left(K \cdot \frac{v^{d-1} r^d N^d}{n^{2d-1}}\right)^v \ , \tag{9}$$

where K is a constant depending on d only. The expression in parentheses is largest for $v = 2N$; then it evaluates to

$$K \cdot 2^{d-1} \cdot r^d (N/n)^{2d-1} \leq K \cdot 2^{d-1} \cdot r^{1/2-(2d-1)\varepsilon} = o(1) \ .$$

Thus the entire sum (9) is $o(1)$, and the claim follows with Markov's inequality. $\square$

The last missing ingredient for our lower bound proof is the following purely deterministic result, which might be of independent interest.

Theorem 12 (Counting paths)**.** *For any $d > 0$ and $\ell \in \mathbb{N}_0$ there exists $C_0 = C_0(d, \ell)$ such that the following is true: If G is a graph with N edges, maximum degree Δ, and maximum density at most d, then G contains at most $C_0 N \Delta^{\lfloor \ell/2 \rfloor}$ paths of length ℓ.*

We postpone the proof of Theorem 12 and proceed with the proof of the desired lower bound.

Proof of Theorem 6 (lower bound). Let $N \ll \frac{n}{r^{\lceil \ell/2 \rceil/(\lceil \ell/2 \rceil+1)}}$ be given, and assume w.l.o.g. that moreover $N \geq n/r^{1-\delta}$ for some small $\delta > 0$. We show that a.a.s. the number of $P_{\ell-1}$'s created by any strategy is of order at most $N \cdot (Nr/n)^{\lfloor (\ell-1)/2 \rfloor} = N \cdot (Nr/n)^{\lceil \ell/2 \rceil - 1}$. This immediately implies that the expected number of edges closing a C_ℓ seen (not necessarily picked) is of order at most

$$N \cdot (Nr/n)^{\lceil \ell/2 \rceil - 1} \cdot Nr/n^2 = \frac{N^{\lceil \ell/2 \rceil+1} r^{\lceil \ell/2 \rceil}}{n^{\lceil \ell/2 \rceil+1}} \ll 1 \ ,$$

which implies the claim by Markov's inequality.

Recall that we work in a relaxed offline setting, where the player may select an arbitrary set of N edges from $m = rN$ distinct random vertex pairs offered. By Lemma 11 (with $\varepsilon := 1/10$, say), a.a.s. the maximum density of the random graph $G_{n,m}$ formed by these vertex pairs is bounded by some constant d, and regardless of the player's strategy the same holds for the graph G_N formed by the selected edges. Set $\Delta := 12m/n = 12Nr/n$, and let G_Δ denote the subgraph of G_N obtained by removing all edges incident to vertices of degree larger than Δ. Clearly also the maximum density of G_Δ is bounded by d, and we obtain with Theorem 12 that a.a.s. the graph G_Δ contains at most the claimed number of paths of length $\ell - 1$.

It remains to bound the number of paths of length $\ell - 1$ in G_N that involve one of the vertices of degree larger than Δ. Denoting by X_k the number of vertices of degree k (in $G_{n,rN} = G_{n,m}$) as before, the number of such paths is at most

$$\sum_{k>\Delta} X_k \cdot \ell \cdot k^{\ell-1} \ ,$$

where each term accounts for the paths whose highest-degree vertex has degree k. By Lemma 10, this is a.a.s. bounded by

$$\mathcal{O}\left(n \sum_{k>\Delta} (2/3)^k k^{\ell-1}\right) = n\,\mathrm{e}^{-\Omega(r^\delta)} = o(n/r) \ ,$$

where we used that due to our assumption on N we have $\Delta = \Omega(r^\delta)$. This is in fact a stronger bound than we needed to show, and thus concludes the proof. $\square$

It remains to prove Theorem 12, which we will do by induction. For any $u_1, u_2, \Delta > 0$, we define

$$f_\ell(u_1, u_2, \Delta) = \max_{\substack{x,y \in \mathbb{N}_0: \\ x+y=\ell}} \min\{\Delta^x \cdot u_1, \Delta^y \cdot u_2\} \ , \qquad \ell \geq 0 \ . \tag{10}$$

We will prove the following.

Claim 13 (Counting paths inductively)**.** *For any $d > 0$ and $\ell \in \mathbb{N}_0$ there exists $C = C(d, \ell)$ such that the following is true: If $G = (V, E)$ is a graph with maximum degree Δ and maximum density at most d, then any two (not necessarily disjoint) sets $U_1, U_2 \subseteq V$ are connected by at most $C \cdot f_\ell(|U_1|, |U_2|, \Delta)$ paths of length ℓ.*

Proof of Claim 13. Let G as in the claim be given. For ease of notation, we drop the third argument of f_ℓ as defined in (10) and abbreviate $f_\ell(u_1, u_2, \Delta)$ by $f_\ell(u_1, u_2)$ throughout.

We prove the claim by induction on ℓ. For $\ell = 0$, the quantity we need to bound is simply the size of $U_1 \cap U_2$, which clearly does not exceed $f_0(|U_1|, |U_2|) = \min\{|U_1|, |U_2|\}$. Thus we may set $C(d, 0) := 1$ for any $d > 0$. For the rest of the proof we assume w.l.o.g. that $d \geq 1$.

For $\ell \geq 1$, observe that the number of P_ℓ's connecting U_1 and U_2 is bounded by

$$\sum_{v_1 \in \Gamma(U_1)} (\text{number of } P_{\ell-1}\text{'s connecting } v_1 \text{ and } U_2) \cdot \deg_{U_1}(v_1) \ , \tag{11}$$

where $\Gamma(U_1)$ denotes the set of vertices with at least one neighbour in U_1, and $\deg_{U_1}(v_1)$ denotes the number of neighbours of v_1 in U_1. In the following we split the set $\Gamma(U_1)$ into

$$W_1 := \{v \in \Gamma(U_1) \,|\, \deg_{U_1}(v) \geq 2d\} \qquad \text{and} \qquad \overline{W}_1 := \Gamma(U_1) \setminus W_1 \ ,$$

and bound the contributions of these sets to the sum (11) separately.

Observing that

$$|\overline{W}_1| \leq |\Gamma(U_1)| \leq \Delta |U_1| \ , \tag{12}$$

the contribution of the low-degree vertices to the sum (11) can be bounded as

$$\begin{aligned} &\sum_{v_1 \in \overline{W}_1} (\text{number of } P_{\ell-1}\text{'s connecting } v_1 \text{ and } U_2) \cdot \deg_{U_1}(v_1) \\ &\leq (\text{number of } P_{\ell-1}\text{'s connecting } \overline{W}_1 \text{ and } U_2) \cdot 2d \\ &\overset{\text{I.H.}}{\leq} C(\ell-1, d) \cdot f_{\ell-1}(|\overline{W}_1|, |U_2|) \cdot 2d \\ &\overset{(10),(12)}{\leq} C(\ell-1, d) \cdot 2d \cdot f_\ell(|U_1|, |U_2|) \ , \end{aligned} \tag{13}$$

where in the second-to-last step we used the induction hypothesis.

Note that by definition of W_1 the set $W_1 \cup U_1$ spans at least $2d|W_1| - |E(W_1 \cap U_1)| \geq 2d|W_1| - d|W_1 \cap U_1|$ edges, where the inequality is due to the assumption that G has maximum density at most d. If $|W_1| > |U_1|$, this is strictly larger than $d(|W_1| + |U_1| - |W_1 \cap U_1|) = d|W_1 \cup U_1|$, contradicting the assumption that G has maximum density at most d. Consequently we have $|W_1| \leq |U_1|$, which in turn implies that the set $W_1 \cup U_1$ spans at most $d(|W_1| + |U_1|) \leq 2d|U_1|$ edges. In particular, we have

$$\sum_{v_1 \in W_1} \deg_{U_1}(v_1) \leq 4d|U_1| \ . \tag{14}$$

Consider now the sum

$$\sum_{v_1 \in W_1} (\text{number of } P_{\ell-1}\text{'s connecting } v_1 \text{ and } U_2) \cdot \deg_{U_1}(v_1) \ .$$

To derive an upper bound for this quantity, we may replace the $\deg_{U_1}(v_1)$ terms by a weight function $w_1 : W_1 \to [0, \Delta]$ satisfying $\sum_{v_1 \in W_1} w_1(v_1) \leq 4d|U_1|$ (recall (14)) that can be optimized independently from the edges forming the $P_{\ell-1}$'s.

Clearly, if $|U_1| \leq \Delta/(4d)$, for any fixed choice of these other edges, the best choice for w_1 is to assign weight $4d|U_1|$ to the vertex $v \in W_1$ that contributes the most to the sum. Otherwise, i.e., if $|U_1| > \Delta/(4d)$, the best choice for w_1 is to assign weight Δ to the $\min\{\lfloor 4d|U_1|/\Delta \rfloor, |W_1|\}$ vertices $v_1 \in W_1$ that contribute the most to the sum, and possibly some remaining weight to one other vertex. In either case, it follows that we may assume w.l.o.g. that there is a set $W_1^* \subseteq W_1$ with $|W_1^*| \leq \lceil 4d|U_1|/\Delta \rceil$ such that $\deg_{U_1}(v_1) = 0$ for all $v_1 \in W_1 \setminus W_1^*$.

To compute the contribution of the vertices in W_1 to the sum (11), we first consider the case where $|U_1| > \Delta/(4d)$. In that case we have

$$|W_1^*| \leq \lceil 4d|U_1|/\Delta \rceil \leq 8d|U_1|/\Delta \tag{15}$$

and we obtain

$$\begin{aligned}
&\sum_{v_1 \in W_1} (\text{number of } P_{\ell-1}\text{'s connecting } v_1 \text{ and } U_2) \cdot \deg_{U_1}(v_1) \\
&\leq (\text{number of } P_{\ell-1}\text{'s connecting } W_1^* \text{ and } U_2) \cdot \Delta \\
&\overset{\text{I.H.}}{\leq} C(\ell-1, d) \cdot f_{\ell-1}(|W_1^*|, |U_2|) \cdot \Delta \\
&\overset{(10),(15)}{\leq} C(\ell-1, d) \cdot 8d \cdot f_\ell(|U_1|, |U_2|) \ ,
\end{aligned} \tag{16}$$

where in the second-to-last step we used the induction hypothesis.

If $|U_1| \leq \Delta/(4d)$, the set W_1^* consists of a single vertex v_1^*. Thus we obtain

$$\begin{aligned}
&\sum_{v_1 \in W_1} (\text{number of } P_{\ell-1}\text{'s connecting } v_1 \text{ and } U_2) \cdot \deg_{U_1}(v_1) \\
&\leq (\text{number of } P_{\ell-1}\text{'s connecting } v_1^* \text{ and } U_2) \cdot |U_1| \\
&\overset{\text{I.H.}}{\leq} C(\ell-1, d) \cdot f_{\ell-1}(1, |U_2|) \cdot |U_1| \\
&\leq C(\ell-1, d) \cdot f_\ell(|U_1|, |U_2|) \ ,
\end{aligned} \tag{17}$$

where the last inequality follows from the fact that for $u_1 \leq \Delta$ we obtain from the definition in (10) that

$$\begin{aligned} f_{\ell-1}(1, u_2) \cdot u_1 &= \max_{\substack{x,y \in \mathbb{N}_0: \\ x+y=\ell-1}} \min\{\Delta^x \cdot u_1, \Delta^y \cdot u_2 \cdot u_1\} \\ &\leq \max_{\substack{x,y \in \mathbb{N}_0: \\ x+y=\ell-1}} \min\{\Delta^x \cdot u_1, \Delta^{y+1} \cdot u_2\} \\ &\leq f_\ell(u_1, u_2) \ . \end{aligned}$$

Combining (13), (16), and (17) we obtain the claim for $C(\ell, d) = 10d \cdot C(\ell-1, d)$, i.e., for $C(\ell, d) = (10d)^\ell$. □

Proof of Theorem 12. We bound the number of paths in G by applying Claim 13 with $U_1 = U_2$ being the set of non-isolated vertices. Clearly we have $|U_1| = |U_2| \leq 2N$, which yields with (10) that

$$f_\ell(|U_1|, |U_2|, \Delta) \leq \max_{\substack{x,y \in \mathbb{N}_0: \\ x+y=\ell}} \min\{\Delta^x \cdot 2N, \Delta^y \cdot 2N\} = 2N \cdot \Delta^{\lfloor \ell/2 \rfloor} \ ,$$

where in the last step we used that $\Delta \geq 1$. Theorem 12 follows for $C_0 := 2C$. □

6. Concluding remarks

We conclude this paper by explicitly stating some open questions that we would like to see answered.

The findings presented in this paper lead us to conjecture the following.

Conjecture 14. *For any graph H that has maximum degree at least two, there exists a rational number $q = q(H)$, $0 < q < 1$, such that the following holds: For any function $r = r(n)$ that grows at most subpolynomially, the threshold for creating a copy of H in the Achlioptas process with parameter r is*

$$N_0(H, r, n) = \frac{n^{2-1/m(H)}}{r^q} \ .$$

In view of our result for cycles (Theorem 6), we do not believe that there is an 'easy' general formula for $q = q(H)$. But, assuming Conjecture 14 is true, is there a finite procedure that computes q for any given graph H?

Another line of research would be to determine explicit threshold functions for special classes of graphs H, in particular for complete graphs of arbitrary size. The main difficulty that prevents us from tackling these and other questions is the lack of generally applicable techniques for proving *lower* bounds on the threshold. The approach used for the lower bound proof in Theorem 6 is unlikely to generalize, as it relies crucially on the fact that $rN \gg n \gg N$ in the regime of interest.

As a last open question we mention the problem of *avoiding* copies of some given graph H in Achlioptas processes with growing parameter $r = r(n)$. For fixed values of r, the general threshold function of this problem was determined in [19], but for the case where r grows the question is wide open.

References

[1] D. Achlioptas, R. M. D'Souza, and J. Spencer. Explosive percolation in random networks. *Science*, 323(5920):1453–1455, 2009.

[2] Y. Azar, A. Broder, A. Karlin, and E. Upfal. Balanced allocations. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing*, pages 593–602. ACM, 1994.

[3] J. Balogh and J. Butterfield. Online Ramsey games for triangles in random graphs. *Discrete Mathematics*, 310(24):3653–3657, 2010.

[4] M. Belfrage, T. Mütze, and R. Spöhel. Probabilistic one-player Ramsey games via deterministic two-player games. ArXiv e-print 0911.3810, November 2009.

[5] T. Bohman and A. Frieze. Avoiding a giant component. *Random Structures Algorithms*, 19(1):75–85, 2001.

[6] T. Bohman, A. Frieze, M. Krivelevich, P.-S. Loh, and B. Sudakov. Ramsey games with giants. *Random Structures Algorithms*, 38(1-2):1–32, 2011.

[7] T. Bohman, A. Frieze, and N. Wormald. Avoidance of a giant component in half the edge set of a random graph. *Random Structures Algorithms*, 25(4):432–449, 2004.

[8] T. Bohman and D. Kravitz. Creating a giant component. *Combin. Probab. Comput.*, 15(4):489–511, 2006.

[9] B. Bollobás. Threshold functions for small subgraphs. *Math. Proc. Cambridge Philos. Soc.*, 90(2):197–206, 1981.

[10] A. Flaxman, D. Gamarnik, and G. Sorkin. Embracing the giant component. *Random Structures Algorithms*, 27(3):277–289, 2005.

[11] E. Friedgut, Y. Kohayakawa, V. Rödl, A. Ruciński, and P. Tetali. Ramsey games against a one-armed bandit. *Combin. Probab. Comput.*, 12(5-6):515–545, 2003. Special issue on Ramsey theory.

[12] L. Gugelmann and R. Spöhel. On balanced coloring games in random graphs. *Electronic Notes in Discrete Mathematics*, 38(0):425 – 430, 2011. Proceedings of the Sixth European Conference on Combinatorics, Graph Theory and Applications (EuroComb) 2011.

[13] M. Krivelevich, P.-S. Loh, and B. Sudakov. Avoiding small subgraphs in Achlioptas processes. *Random Structures Algorithms*, 34(1):165–195, 2009.

[14] M. Krivelevich, E. Lubetzky, and B. Sudakov. Hamiltonicity thresholds in Achlioptas processes. *Random Structures Algorithms*, 37(1):1–24, 2010.

[15] M. Marciniszyn, D. Mitsche, and M. Stojaković. Online balanced graph avoidance games. *European J. Combin.*, 28(8):2248–2263, 2007.

[16] M. Marciniszyn, R. Spöhel, and A. Steger. Online Ramsey games in random graphs. *Combin. Probab. Comput.*, 18(1-2):271–300, 2009.

[17] M. Marciniszyn, R. Spöhel, and A. Steger. Upper bounds for online Ramsey games in random graphs. *Combin. Probab. Comput.*, 18(1-2):259–270, 2009.

[18] M. Mitzenmacher, A. Richa, and R. Sitaraman. The power of two random choices: a survey of techniques and results. In *Handbook of randomized computing, Vol. I, II*, volume 9 of *Comb. Optim.*, pages 255–312. Kluwer Acad. Publ., Dordrecht, 2001.

[19] T. Mütze, R. Spöhel, and H. Thomas. Small subgraphs in random graphs and the power of multiple choices. *Journal of Combinatorial Theory, Series B*, 101(4):237–268, 2011.

[20] A. Prakash, R. Spöhel, and H. Thomas. Balanced online Ramsey games in random graphs. *Electron. J. Combin.*, 16(1):Research Paper 11, 22, 2009.

[21] O. Riordan and L. Warnke. Achlioptas process phase transitions are continuous. ArXiv e-print 1102.5306, February 2011. To appear in *The Annals of Applied Probability*.

[22] J. Spencer and N. Wormald. Birth control for giants. *Combinatorica*, 27(5):587–628, 2007.