

On the maximum running time in graph bootstrap percolation

Béla Bollobás*

Department of Pure Mathematics and Mathematical Statistics
University of Cambridge, U.K.

Department of Mathematical Sciences
University of Memphis, U.S.A.

b.bollobas@dpmms.cam.ac.uk

Michał Przykucki

Oliver Riordan

Julian Sahasrabudhe

Mathematical Institute
University of Oxford, U.K.

Department of Mathematical Sciences
University of Memphis, U.S.A.

{przykucki,riordan}@maths.ox.ac.uk

julian.sahasra@gmail.com

Submitted: Dec 8, 2015; Accepted: Apr 10, 2017; Published: May 5, 2017

Mathematics Subject Classifications: 05D99, 05D40, 05C35, 05C80

Abstract

Graph bootstrap percolation is a simple cellular automaton introduced by Bollobás in 1968. Given a graph H and a set $G \subseteq E(K_n)$ we initially ‘infect’ all edges in G and then, in consecutive steps, we infect every $e \in K_n$ that completes a new infected copy of H in K_n . We say that G *percolates* if eventually every edge in K_n is infected. The extremal question about the size of the smallest percolating sets when $H = K_r$ was answered independently by Alon, Kalai and Frankl. Here we consider a different question raised more recently by Bollobás: what is the maximum time the process can run before it stabilizes? It is an easy observation that for $r = 3$ this maximum is $\lceil \log_2(n - 1) \rceil$. However, a new phenomenon occurs for $r = 4$ when, as we show, the maximum time of the process is $n - 3$. For $r \geq 5$ the behaviour of the dynamics is even more complex, which we demonstrate by showing that the K_r -bootstrap process can run for at least $n^{2-\varepsilon_r}$ time steps for some ε_r that tends to 0 as $r \rightarrow \infty$.

Keywords: Bootstrap percolation; weak saturation; cellular automata; monotone cellular automata; extremal combinatorics

*Research supported in part by NSF grant DMS-1301614 and EU MULTIPLEX grant 317532.

1 Introduction

Graph bootstrap percolation was introduced by Bollobás in 1968 [10] under the name weak saturation. For a fixed graph H , a graph G on n vertices is called *weakly H -saturated* if there is no copy of H in G , but there is an ordering of the missing edges of G so that if they are added one-at-a-time, every added edge creates a new copy of H . Bollobás conjectured that the minimum size of a weakly K_r -saturated graph is $\binom{n}{2} - \binom{n-r+2}{2}$. This conjecture was confirmed independently by Alon [1], Frankl [15] and Kalai [19].

Recently, Balogh, Bollobás and Morris [3] observed that weak saturation is strongly related to bootstrap percolation, a dynamical process introduced by Chalupa, Leath and Reich [12] in 1979 to model the behaviour of ferromagnets. In bootstrap percolation on a graph $G = (V, E)$ with infection threshold $r \geq 2$ we choose a set $A \subseteq V$ of initially ‘infected’ vertices and we declare the remaining vertices ‘healthy’. Then, in consecutive rounds, we infect all healthy vertices with at least r infected neighbours. We say that A *percolates* if, starting from A as the set of the initially infected vertices, we eventually infect every vertex in V . More precisely, we set $A_0 = A$ and for $t = 1, 2, 3, \dots$ we define

$$A_t = A_{t-1} \cup \{v \in V : |N_G(v) \cap A_{t-1}| \geq r\},$$

where $N_G(v)$ is the neighbourhood of the vertex v in G . Hence A percolates if we have $\bigcup_{t=0}^{\infty} A_t = V(G)$.

Let us then redefine weak saturation in the language of bootstrap percolation. We fix a graph H and we choose a set $G \subseteq E(K_n)$ of edges that we initially infect. (We shall switch back and forth between thinking of G as a graph and as a set of edges.) Then, in consecutive steps, we infect every $e \in K_n$ that completes a new infected copy of H in K_n . Formally, we take $G_0 = G$ and for $t = 1, 2, 3, \dots$ let

$$G_t = G_{t-1} \cup \{e \in E(K_n) : G_{t-1} + e \text{ contains more copies of } H \text{ than } G_{t-1}\}. \quad (1)$$

We call this process *H -bootstrap percolation* and we say that G *percolates* (or *H -percolates*) if eventually every edge in K_n is infected. Hence, G is weakly H -saturated if and only if it percolates in H -bootstrap percolation and contains no copy of H .

In bootstrap percolation we are usually interested in the following setup: for $0 < p < 1$, we initially infect every vertex in V independently at random with probability p . We then look for the critical probability $p = p_c(G, r)$ beyond which percolation is more likely to occur than not. A large number of results, often very sharp, have been obtained in the search for critical probabilities for various graphs G and values of r . To name a few, the values of the critical probabilities $p_c(\mathbb{Z}^d, r)$ for infinite grids were found by van Enter [13] ($r = d = 2$) and Schonmann [25] (for general r and d). Holroyd [17] and Balogh, Bollobás, Duminil-Copin and Morris [2] obtained sharp bounds on the critical probabilities for finite grids. Various families of random graphs were studied in this context by Balogh and Pittel [6], Janson, Łuczak, Turova and Vallier [18] and Bollobás, Gunderson, Holmgren, Janson and Przykucki [11].

An analogous question in graph bootstrap percolation was considered in [3]. There, the authors look at the random graph $G(n, p)$ obtained by choosing every edge of K_n

independently at random with probability $p = p(n)$. For fixed r , they obtain bounds on the values of p for which $G(n, p)$ is likely to percolate in the K_r -bootstrap process. In a similar vein, Gunderson, Koch and Przykucki [16] studied the density threshold for which $G(n, p)$ percolates in exactly T steps, in the graph bootstrap process.

Another family of problems considered in bootstrap percolation concerns the extremal properties of the process. For example, the size of the smallest percolating sets in various graphs was studied by Pete [5], Balogh, Bollobás, Morris and Riordan [4] and by Morrison and Noel [22]. Morris [21] and Riedl [24] analysed the size of the largest minimal percolating sets for the $n \times n$ square grid and the hypercube respectively, both for $r = 2$. Przykucki [23] and Benevides and Przykucki [7, 8] studied the maximum time the infection process on a square grid or a hypercube can take before it stabilizes for $r = 2$.

In this paper we study a question due to Bollobás [9], who initiated the study of the maximum running time of the graph bootstrap process. Let us define the *maximum running time* of the K_r -bootstrap process to be

$$M_r(n) = \max\{t : \text{there exists } G \subseteq E(K_n) \text{ s.t. } G_t \neq G_{t-1} \text{ in the } K_r\text{-bootstrap process}\},$$

where G_t is defined as in (1). We say that the process *stabilizes at time* t if t is the smallest integer such that $G_t = G_{t+1}$. Hence $M_r(n)$ is the maximum time the K_r -bootstrap process takes before it stabilizes starting from any initially infected graph with n vertices.

It is an easy observation that $M_3(n) = \lceil \log_2(n - 1) \rceil$. Indeed, at every step of the K_3 -bootstrap process, any two vertices at distance two are joined by an edge. Hence, if G_t is connected then we have $\text{diam}(G_{t+1}) = \lceil \text{diam}(G_t)/2 \rceil$. If G_t is not connected then the components of G_t remain separate in G_{t+1} . Therefore to maximize the time until the process stabilizes we should take G_0 to be a connected graph with as large diameter as possible, i.e., a path on n vertices. It then takes $\lceil \log_2(n - 1) \rceil$ steps until we obtain a clique.

We shall prove two main results.

Theorem 1. *We have $M_4(n) = n - 3$ for all $n \geq 3$.*

Theorem 2. *For each fixed $r \geq 5$, we have $M_r(n) \geq n^{2-\alpha_r-o(1)}$ as $n \rightarrow \infty$, where $\alpha_r = \frac{r-2}{\binom{r}{2}-2}$.*

The rest of this paper is organised as follows. In Sections 2 and 3 we prove Theorems 1 and 2 respectively. We prove Theorem 2 by giving a probabilistic argument showing the existence of initially infected sets of edges satisfying our bound. In Section 4 we give a deterministic construction of an initially infected set that in the K_5 -bootstrap process runs for at least $\frac{1}{200}n^{3/2}$ time steps before it stabilizes. The construction can be easily adapted for all $r \geq 5$, always guaranteeing the running time to be at least $c_r n^{3/2}$ (this is weaker than Theorem 2, but the proof is much simpler). Finally, in Section 5 we discuss some open problems.

For each $r \geq 3$, the parameter $M_r(n)$ has a natural sister parameter $M'_r(n)$, defined as the maximum running time of a *percolating* graph on n vertices. These quantities are closely related: it is trivial that $M_r(n) \geq M'_r(n)$, while, given a graph on n vertices for

which the K_r -process ($r \geq 5$) runs for time T , it is not hard to construct a percolating graph on $O(n)$ vertices for which the process runs for time $> T$.

Independently of the present work, Kilian Matzke [20] has proved Theorem 1 and a weaker form of Theorem 2, giving $M_r(n) = \Omega(n^{3/2})$ for each $r \geq 5$.

2 Maximum running time in the K_4 -bootstrap process

In this section we prove Theorem 1, that is, we show that $M_4(n) = n - 3$ for all $n \geq 3$. We prove this in two steps: we first give a simple construction to show $M_4(n) \geq n - 3$; then we prove the corresponding upper bound. Given a graph G , in this section we define G_t as in (1) with $H = K_4$. Thus G_t is the graph obtained after running the K_4 -process for t steps, starting with the graph G .

Lemma 3. *For all integers $n \geq 3$, we have $M_4(n) \geq n - 3$.*

Proof. The case $n = 3$ is trivial. We shall show by induction that for each $n \geq 4$ there is an n -vertex graph G that stabilizes in time exactly $n - 3$, for which $G_{n-3} = K_n$. The base case is trivial taking $G = K_4^-$, a complete graph on 4 vertices with one edge deleted.

Suppose then that $n \geq 4$ and G is such a graph; it remains to construct a corresponding graph G^+ on $n + 1$ vertices. Let xy be any edge of $G_{n-3} \setminus G_{n-4}$, and let $G^+ = G + xz + yz$ be formed from G by adding one new vertex, z , and two new edges, xz and yz . Let $t \leq n - 3$ be the first time that there is a copy of K_4^- present in G_t^+ but not in G_t . Until this time, exactly the same edges are added in the percolation processes starting from G and from G^+ , so $G_t^+ = G_t + xz + yz$. Any K_4^- present in this graph but not in G_t must contain z , and hence x, y and one other vertex w . Since the edge zw is missing, xy must be present, so $t \geq n - 3$. Conversely, since G_{n-3} is complete, G_{n-3}^+ consists of a complete graph on n vertices with an extra vertex z joined to x and to y , and it follows that all edges zw are added in the next step, so G_{n-2}^+ is complete, as required. \square

We now turn to the upper bound in Theorem 1.

Observation 4. *Suppose that $K_1, K_2 \subseteq V(G)$ are such that K_1, K_2 induce cliques in G_t with $|K_1 \cap K_2| \geq 2$. Then $K_1 \cup K_2$ induces a clique in G_{t+1} .*

To prove that $M_4(n) \leq n - 3$, we prove the following stronger statement by induction. Once we have established it, the upper bound will easily follow. For a set A and $k \in \mathbb{N}$, we define $A^{(k)}$ to be the collection of subsets of A of size k . So $A^{(2)}$ is the edge-set of the complete graph with vertex set A .

Lemma 5. *Let $t \geq 1$ and suppose that $e = ij \in G_t \setminus G_{t-1}$ in the K_4 -bootstrap process. Then for some $d = d(e) \in \{0, 1\}$ the edge e is contained in a clique of size $t + 3 + d$ in G_{t+d} .*

Proof. We prove the lemma by induction on t . The case $t = 1$ is immediate. So assume that $t \geq 2$ and that the statement is true for $t - 1$. Now suppose that $e = ij \in G_t \setminus G_{t-1}$.

We want to find $d(e) \in \{0, 1\}$ such that e is contained in a clique of size $t + 3 + d(e)$ in $G_{t+d(e)}$.

The edge ij appears in G_t by virtue of some 4-element set K such that $i, j \in K$ and $(K^{(2)} - ij) \subseteq G_{t-1}$. Note that we cannot have $(K^{(2)} - ij) \subseteq G_{t-2}$, otherwise we would have $ij \in G_{t-1}$. So there is another edge f of $K^{(2)}$ in $G_{t-1} \setminus G_{t-2}$. By induction there exists some $d(f)$ and a set $C \subseteq V(G)$ such that $f \in C^{(2)}$, $|C| \geq (t-1) + 3 + d(f)$ and $C^{(2)} \subseteq G_{t-1+d(f)}$. We consider the following cases.

1. Suppose first that $i, j \in C$. Then we cannot have $d(f) = 0$ as this would imply that C induces a clique in G_{t-1} and hence the edge e is in G_{t-1} , which contradicts the choice of e . Hence we have $d(f) = 1$ and, taking $d(e) = d(f) - 1 = 0$, we observe that C induces a clique in $G_{t+d(e)} = G_t$ of size $(t-1) + d(f) + 3 = t + d(e) + 3 = t + 3$ that contains i, j . Hence, we are done in this case.
2. Suppose instead that at least one of the vertices i, j is not in C . We further divide this case into two subcases.

(a) If $d(f) = 1$ then K and C both induce cliques in $G_{t-1+d(f)} = G_t$. Since $f \in K^{(2)} \cap C^{(2)}$, we have $|K \cap C| \geq 2$ and, by Observation 4, at time $t + 1$, K and C will merge to form a single clique of size at least $|C| + 1 \geq t + 3 + 1$. Hence, we are done choosing $d(e) = 1$.

(b) If $d(f) = 0$ then both K and C induce cliques in G_t . If $|K \cap C| = 2$ then by Observation 4 at time $t + 1$ they merge to form a complete graph of size $|C| + 2 \geq t + 1 + 3$ in G_{t+1} . So again we are done taking $d(e) = 1$.

If $|K \cap C| = 3$ then C contains exactly one endpoint of e . Assume that $i \in C, j \notin C$. Since $K^{(2)} - e \subseteq G_{t-1}$, j has two neighbours $j_1, j_2 \in K \cap C$ in G_{t-1} . This implies that both C and $\{j, j_1, j_2\}$ induce cliques in G_{t-1} and that they intersect in two points, i.e., in j_1 and j_2 . Hence by Observation 4 we have that $C \cup \{j\}$ induces a clique of size $|C| + 1 \geq t + 3$ in G_t . Hence, taking $d(e) = 0$ we are done. This completes the proof of Lemma 5. \square

Theorem 1 follows easily from Lemma 5.

Proof of Theorem 1. Let T be the time at which the K_4 -bootstrap process stabilizes starting from an arbitrary graph $G = G_0$. For any edge e , let $t(e)$ be the time at which the edge e is added to the infected set. For all $e \in G_T$ we have that

$$n \geq t(e) + d(e) + 3,$$

since by Lemma 5 we know that $G_{t(e)+d(e)}$, a graph on n vertices, contains a clique of size $t(e) + d(e) + 3$. Now since $d(e) \in \{0, 1\}$, we have

$$n \geq \max_{e \in G_T} \{t(e) + d(e) + 3\} \geq \max_{e \in G_T} \{t(e) + 3\} = T + 3.$$

This completes the proof of Theorem 1. \square

3 Lower bounds on $M_r(n)$ for $r \geq 5$

In this section we use a probabilistic argument to prove Theorem 2. In other words, taking $\alpha_r = \frac{r-2}{\binom{r}{2}-2}$, we show that for every $\varepsilon > 0$ and for n large enough (depending on r and ε) there exists an n -vertex graph $G = G(r, n)$ such that the K_r -bootstrap process started from G stabilizes after at least $n^{2-\alpha_r-\varepsilon}$ time steps. Throughout we fix $r \geq 5$ and $\varepsilon > 0$; all constants in what follows may depend on r and ε .

The simplest possible way to construct a graph that takes a fairly long time to stabilize in the K_r -bootstrap process is as follows. Consider a sequence $e_0, H_1, e_1, H_2, \dots, e_{t-1}, H_t, e_t$ where each H_i is a copy of K_r , each e_i is an edge, and all of e_1, \dots, e_t and H_1, \dots, H_t are *vertex disjoint* except that both e_{i-1} and e_i are (vertex disjoint) edges of H_i . Take G_0 to be the union of all the edges contained in H_1, \dots, H_t in the chain, with all edges e_i , *except* e_0 , deleted. Then, clearly, e_i is the unique edge added at step i , and the process stabilizes at time t . Of course, this construction only gives the rather weak lower bound $M_r(n) \geq \lfloor (n-2)/(r-2) \rfloor = \Theta(n)$.

To obtain a stronger bound, the idea is to relax the disjointness conditions, but in a way that will not affect the percolation process. More precisely, by a K_r -*chain* (within K_n) we mean a sequence H_1, H_2, \dots, H_t where each H_s is a complete graph on r vertices (contained within K_n), and H_i and H_j are *edge disjoint* unless $|i-j|=1$, in which case they share exactly one edge. Given such a chain we let e_i be the edge shared by H_i and H_{i+1} , let e_0 be some edge of H_1 other than e_1 , and let e_t be some edge of H_t other than e_{t-1} . Sometimes we describe the chain by the list $e_0, H_1, e_1, H_2, \dots, e_{t-1}, H_t, e_t$. In this list, the edges e_i are distinct, and e_i is an edge of H_j if and only if $i \in \{j-1, j\}$.

Let K_r^- denote the graph formed by deleting an edge from K_r . Given a K_r -chain as above, an *external* K_r^- is a (not necessarily induced) subgraph of the graph union $\bigcup_{i=1}^t H_i$ which is isomorphic to K_r^- and is not contained in any individual H_i . A K_r -chain is *good* if it contains no external K_r^- .

Observation 6. *If $e_0, H_1, e_1, H_2, \dots, e_{t-1}, H_t, e_t$ is a good K_r -chain, then the graph $G_0 = \bigcup_{i=1}^t H_i - \{e_1, e_2, \dots, e_t\}$ stabilizes at time t in the K_r -bootstrap process.*

Proof. Although this is immediate, let us spell out the details. Define G_i as in (1), with $H = K_r$. We claim that for $i \leq t$ we have $G_i = G_0 \cup \{e_1, \dots, e_i\}$. Clearly this holds for $i = 0$. Suppose it holds for some $0 \leq i \leq t-1$; then G_i is a subgraph of $\bigcup_{j \leq t} H_j$, so any copy of K_r^- in this graph has vertex set $V(H_j)$ for some j . Within $V(H_j)$, exactly two edges are missing in $G_0 - e_0$, namely e_{j-1} and e_j . Since (by induction) G_i contains e_k if and only if $k \leq i$, we have an induced copy of K_r^- if and only if $j = i+1$, so e_{i+1} is the unique edge added at step $i+1$. It follows that G_0 stabilizes at time t . \square

Our aim is to show that we can fit a very long good K_r -chain into a set of n vertices, i.e., to prove the following result. Here and in what follows we ignore rounding to integers when it makes no essential difference.

Theorem 7. Let $r \geq 5$ and $\varepsilon > 0$ be given, and set $\alpha_r = \frac{r-2}{\binom{r}{2}-2}$. If n is large enough, then there is a good K_r -chain $e_0, H_1, e_1, \dots, e_{T-1}, H_T, e_T$ with $T = n^{2-\alpha_r-\varepsilon}$ using at most n vertices in total.

By Observation 6 above, Theorem 7 implies Theorem 2. The rest of this section is devoted to the proof of Theorem 7, for which we use a random construction. We outline this informally before giving the formal proof. Throughout, $r \geq 5$ and $\varepsilon > 0$ are fixed, n is sufficiently large for various conditions that arise below to hold, and α_r and T are defined as in Theorem 7.

In what follows, we construct a good K_r -chain randomly. There is a snag, however: if we follow the obvious method (just pick $r - 2$ new random vertices each time) then, because there are many steps, we are likely to get stuck (create an external K_r^-) relatively early, at a point where each individual step is unlikely to cause a problem. Intuitively, we should be able to keep going as long as there is a decent chance that the next step succeeds. To make this precise, we consider a number $m = \lceil \log n \rceil$ of attempts at choosing the next $r - 2$ vertices, and continue if one of these succeeds. Of course this leads to a lot of dependence (which choice succeeds depends on what happened at previous steps). But we can get around this in the analysis by considering *all possible* ways that our choices could lead to an external K_r^- .

Outline construction. Let e_0 be a uniformly random edge in K_n . For $1 \leq t \leq T$ and $1 \leq a \leq m$, let $X_{t,a}$ be a uniformly random set of $r - 2$ vertices of K_n , and $e_{t,a}$ a uniformly random edge within $X_{t,a}$, with these choices independent over all t and a . Once these have been selected, for t running from 0 to $T - 1$ do the following: pick the smallest index $a = a_{t+1}$ so that certain conditions specified later hold (if this is possible). Then set $X_{t+1} = X_{t+1,a}$, $e_{t+1} = e_{t+1,a}$, and let H_{t+1} be the complete graph on $e_t \cup X_{t+1}$.

We think of $X_{t+1,a}$ as the a th *candidate* set of new vertices in step $t + 1$, and X_{t+1} as the selected set; similarly, $e_{t+1,a}$ is a candidate next edge in the chain, and e_{t+1} the actual next edge. A candidate set $X_{t+1,a}$ is *successful* if it satisfies certain conditions (C0)–(C4) below. Later we shall write \mathcal{F}_{t+1} for the event that our construction *fails* at step $t + 1$, in the sense that none of the candidates $X_{t+1,a}$ for X_{t+1} is successful. We write

$$\mathcal{G}_t = \mathcal{F}_1^c \cap \dots \cap \mathcal{F}_t^c$$

for the ‘good’ event that the construction succeeds up to (at least) step t .

The first of the conditions referred to above is

$$X_{t+1,a} \text{ is disjoint from } e_t. \tag{C0}$$

This condition ensures that $e_t \cup X_{t+1}$ is indeed a set of r vertices.

We shall always write $G_t = \bigcup_{1 \leq s \leq t} H_s$ for the graph constructed so far. Our second condition is the following:

$$\text{No pair } A \neq V(e_t) \text{ of vertices contained in } e_t \cup X_{t+1,a} \text{ is contained in any } H_s, \tag{C1} \\ s \leq t.$$

Note that if we select $X_{t+1,a}$ to be X_t , and so take H_{t+1} to be the complete graph on $e_t \cup X_{t,a+1}$, then condition (C1) applies to every edge A of H_{t+1} other than e_t . Thus, if conditions (C0) and (C1) hold at each step $t' \leq t$, then $e_0, H_1, e_1, \dots, H_t, e_t$ is a K_r -chain: each H_s is a complete graph on r vertices, and for $i < j$ (using (C1) at step j) we see that H_i and H_j are edge disjoint unless $j = i + 1$, in which case they share only one edge, e_i .

For $2 \leq i \leq r - 1$, we say that a set A of i vertices is *dangerous at time t* (or sometimes just *dangerous*) if there is a set B of r vertices with $A \subset B$ and $B \neq V(H_s)$ for all $s \leq t$ such that the graph $G_t \cup A^{(2)}$ includes all but at most one of the edges within B . In other words, A is dangerous if adding all edges within A would create an external K_r^- which contains A . The next condition involves avoiding such sets:

No set $A \neq V(e_t)$ with $2 \leq i \leq r - 1$ vertices and with $A \subseteq e_t \cup X_{t+1,a}$ was dangerous at time t . (C2)

Note that the conditions above depend on the particular choice of X_1, \dots, X_t , which seems difficult to analyse. However, we will not deal directly with these conditions; rather, we control the probability of \mathcal{F}_{t+1} in terms of the (more manageable) event that any *a priori* possible choice of the X_1, \dots, X_t (from the potential candidates) violates one of the conditions above.

Observation 8. *If, for each step t , $1 \leq t \leq T$, we manage to choose a candidate $X_t = X_{t,a}$ so that conditions (C0)–(C2) hold, then $(e_0, H_1, \dots, H_T, e_T)$ is a good K_r -chain.*

Proof. As noted above, (C0) and (C1) ensure that we get a K_r -chain. Suppose it is not good. Then at some step t an external K_r^- must have been generated. Let its vertex set be B . Clearly, $G_t \setminus G_{t-1}$ contains at least one edge within B . Hence $A = B \cap V(H_t)$ is at set of at least 2 and (by definition of an external K_r^-) at most $r - 1$ vertices with $A \neq V(e_t) \subset V(H_{t-1})$. But then A was dangerous at step $t - 1$, so condition (C2) did not hold. □

If steps $1, 2, \dots, t$ succeed, i.e., \mathcal{G}_t holds, then by definition conditions (C0)–(C2) were satisfied at steps $1, \dots, t$. Hence

\mathcal{G}_t implies that H_1, \dots, H_t is a good K_r -chain.

Our aim is to show that with positive probability, we can satisfy conditions (C0)–(C2) above. Since our construction is inductive, when considering the candidates for X_{t+1} we may assume that \mathcal{G}_t holds.

A key step in our proof is to show that there are not too many dangerous i -sets. More generally, we shall show that for any graph F with r vertices, if we exclude copies arising inside a single H_s , it is unlikely that G_t contains many more copies of F than a random graph with the same overall edge density as G_t . To show this, we fix a set A of r vertices, and think about all possible ways that the chain H_1, H_2, \dots, H_t can meet A , considering only cliques H_s that contribute an edge within A , i.e., satisfy $|V(H_s) \cap A| \geq 2$. Since successive copies of K_r in the chain are related in a different way from those further apart,

we shall split these copies of K_r in the chain that meet A into subchains. This (hopefully) motivates the following definition.

Let A be a set of r vertices. By a *chain within A* we mean a sequence (S_1, \dots, S_k) of subsets of A such that

- (i) $2 \leq |S_i| \leq r - 1$ for $1 \leq i \leq k$ and
- (ii) $|S_i \cap S_{i+1}| \leq 2$ for $1 \leq i \leq k - 1$.

By a *chain cover \mathcal{C}* of A we mean a sequence $(S_1^1, \dots, S_{k_1}^1), (S_1^2, \dots, S_{k_2}^2), \dots, (S_1^\ell, \dots, S_{k_\ell}^\ell)$ of chains within A . Note that a chain cover need not cover all of the edges in A . Rather, these chain covers correspond to all possible ways that our long ‘chain’ of copies of K_r could possibly intersect A (excluding the case where one copy has vertex set A). We say that a chain cover is *minimal* if every S_i^j spans some edge not spanned by any other set $S_{i'}^{j'}$ in the chain cover. Clearly, a minimal chain cover contains at most $\binom{r}{2}$ sets S_i^j in total, so there are a finite number (depending on r) of minimal chain covers of any r -set A . Also, if \mathcal{C} is minimal, then each of its constituent chains is minimal in the natural sense (i.e., as a chain cover with one chain).

The *cost* of a chain $C = (S_1, \dots, S_k)$ is

$$c(C) = |S_1| + \sum_{i=1}^{k-1} |S_{i+1} \setminus S_i|,$$

informally corresponding to the total number of ‘new’ vertices that must hit A in order to realise the intersection pattern described by C . The *edge set* $E(C)$ of C is simply $\bigcup_{i=1}^k S_i^{(2)}$. The *benefit* $b(C)$ of C is $|E(C)|$. For a chain cover $\mathcal{C} = (C_1, \dots, C_j)$, we define its cost and benefit by $c(\mathcal{C}) = \sum c(C_i)$ and $b(\mathcal{C}) = \sum b(C_i)$. (For $b(\mathcal{C})$ it might be more natural to take the total number of edges without repetition; this makes no difference in the argument below.)

Given a set A of r vertices and a chain cover

$$\mathcal{C} = ((S_1^1, \dots, S_{k_1}^1), (S_1^2, \dots, S_{k_2}^2), \dots, (S_1^\ell, \dots, S_{k_\ell}^\ell))$$

of A , we say that our random K_r -chain H_1, \dots, H_t *meets A in \mathcal{C}* if there exist $0 \leq t_1 < \dots < t_\ell$ such that

- $t_j + k_j \leq t_{j+1} - 1$ for each $1 \leq j \leq \ell - 1$ and $t_\ell + k_\ell \leq t$, and
- $V(H_{t_j+i}) \cap A = S_i^j$ for each $j \leq \ell$ and $1 \leq i \leq k_j$.

In other words, our K_r -chain H_1, \dots, H_t *meets A in \mathcal{C}* if there exist ℓ ‘time’ intervals (with gaps between them) so that the copies of K_r in the j th interval intersect A in the manner described by the j th chain in \mathcal{C} . We impose no condition on how other cliques H_s may meet A .

Observation 9. *Suppose that \mathcal{G}_t holds. Let A be a set of r vertices spanning at least q edges in $G_t = \bigcup_{s \leq t} H_s$, with $A \neq V(H_s)$ for each $s \leq t$. Then there is some minimal chain cover \mathcal{C} with $b(\mathcal{C}) \geq q$ such that $(H_s)_{s=1}^t$ meets A in \mathcal{C} .*

Proof. Simply consider a minimal subset of the cliques H_s that between them contain all $\geq q$ edges of G_t inside A , and, if necessary, split this set of cliques into intervals with gaps of at least one between them. Each of these H_s shares at least two and at most $r - 1$ vertices with A , and consecutive H_s meet in at most two vertices, since \mathcal{G}_t holds so (H_1, \dots, H_t) is a K_r -chain. \square

Lemma 10 and its consequence Corollary 11 below are key to our analysis; the latter says, roughly speaking, that the most likely way for a set A (or some other graph F) to be covered by a random K_r -chain of length T is when the chain meets it $\binom{r}{2} - 1$ times ($|E(F)|$ times), in just one edge, in non-consecutive cliques H_i . This condition is loosely analogous to the ‘2-balanced’ condition that appears in many small subgraph problems. We start by showing that for a single chain, a certain ‘cost-benefit ratio’ is maximized in the single-edge case.

Lemma 10. *Let A be a set of r vertices, and $C = (S_1, \dots, S_k)$ a minimal chain within A . Then $\alpha_r(b(C) - 1) \leq c(C) - 2$.*

Proof. Recalling the definition of α_r , and writing $b = b(C)$ and $c = c(C)$, the inequality claimed can be rewritten as

$$(r - 2)(b - 1) \leq \left(\binom{r}{2} - 2\right)(c - 2). \quad (2)$$

Now $|\bigcup S_i| \leq c$, and $b = |\bigcup S_i^{(2)}| \leq |(\bigcup S_i)^{(2)}|$, so $b \leq \binom{c}{2}$. The inequality

$$(r - 2) \left(\binom{c}{2} - 1\right) \leq \left(\binom{r}{2} - 2\right)(c - 2)$$

is an equality for $c = 2$, and is easily seen (by multiplying out) to hold (strictly) for $c = r - 1$. Hence (since the left-hand side is convex and the right-hand side linear) it holds for $2 \leq c \leq r - 1$; since $b \leq \binom{c}{2}$ this proves (2) in these cases.

For $c \geq r + 1$ we use the trivial bound $b \leq \binom{r}{2}$; it is easy to check that this implies (2) in this case. This leaves only the case $c = r$. In this case (2) reduces to $b \leq \binom{r}{2} - 1$, i.e., we must show that if $c(C) = r$ then there is at least one edge in $A^{(2)}$ missing from $E(C)$. To see this, recall first that $|S_i| \leq r - 1$ by definition, so $k \geq 2$. Also, if $|\bigcup S_i| < c = r$, then clearly $b(C) \leq \binom{r-1}{2} < \binom{r}{2} - 1$. Hence, we may assume that $\bigcup S_i = A$ and that only consecutive sets S_i meet. By minimality there is some $u \in S_1 \setminus S_2$, and some $v \in S_k \setminus S_{k-1}$. But then $uv \notin E(C)$ and we are done. \square

Remark. We do not need it here, but in fact we have strict inequality in the result above unless $k = 1$, $c = 2$ and so $b = 1$. To see this we must show in the final case above that at least two edges are missing, which is not hard.

For what follows, let us recall that $T = n^{2-\alpha_r-\varepsilon}$ is the number of steps in our construction and m is the number of candidates at each step.

Corollary 11. *Let \mathcal{C} be a minimal chain cover of some r -set A consisting of ℓ chains and having total cost c and benefit b . Then $T^\ell n^{-c} \leq n^{-\alpha_r b - \varepsilon}$.*

Proof. If $\mathcal{C} = (C_1, \dots, C_\ell)$ then $c = \sum_{i=1}^\ell c(C_i)$ and $b = \sum_{i=1}^\ell b(C_i)$. Applying Lemma 10 to each chain C_i and summing, we see that $\alpha_r(b - \ell) \leq c - 2\ell$. Rearranging gives

$$(2 - \alpha_r)\ell - c \leq -\alpha_r b.$$

The result follows recalling that $T = n^{2-\alpha_r-\varepsilon}$ and noting that $\ell \geq 1$. □

Fix a set A of r vertices. Given a minimal chain $C = (S_1, \dots, S_k)$ within A and an integer $0 \leq t \leq T - k$, let

$$[A, t, C]$$

denote the event that there *exist* indices $a_t, a_{t+1}, \dots, a_{t+k} \in [m]$ such that if we set $V_s = e_{s-1, a_{s-1}} \cup X_{s, a_s}$ for $s = t+1, \dots, t+k$, then we have $V_{t+j} \cap A = S_j$ for $j = 1, \dots, k$. In other words, $[A, t, C]$ is the event that it is conceivable (considering all *a priori* possible choices for which candidate set/edge is chosen at steps $t, \dots, t+k$) that the consecutive copies of K_r , H_{t+1}, \dots, H_{t+k} , meet A in the chain C . The key here is that the event $[A, t, C]$ depends only on the random variables $X_{s,a}$ for $t \leq s \leq t+k$, $a \in [m]$.

Recalling the definition of $c(C)$, it is not hard to see that for the event $[A, t, C]$ to hold, we need c of the relevant candidate vertices to fall in A which (as we shall now show) is an event with probability $O^*(n^{-c})$; here, as usual, the O^* notation hides bounded powers of $\log n$. To spell this out, suppose that $[A, t, C]$ holds and that a_t, \dots, a_{t+k} are indices as above. Then, recalling that by definition $e_{t, a_t} \subset X_{t, a_t}$, for $j = 2, \dots, k$ we have

$$X_{t+j, a_{t+j}} \supseteq V_{t+j} \setminus V_{t+j-1} \supseteq S_j \setminus S_{j-1}.$$

Also,

$$X_{t, a_t} \cup X_{t+1, a_{t+1}} \supseteq e_{t, a_t} \cup X_{t+1, a_{t+1}} = V_{t+1} \supseteq S_1.$$

Since the sets $X_{s,a}$ are independent random subsets of $r-2 = O(1)$ vertices chosen uniformly from n vertices, for each j the event $X_{t+j, a_{t+j}} \supseteq S_j \setminus S_{j-1}$ has probability $O(n^{-|S_j \setminus S_{j-1}|})$. Furthermore, these events are independent for different j . Similarly, the event $X_{t, a_t} \cup X_{t+1, a_{t+1}} \supseteq S_1$ has probability $O(n^{-|S_1|})$, and is independent of the events involving $X_{t+j, a_{t+j}}$, $j \geq 2$. Combining these bounds, and taking a union bound over the m^{k+1} possible choices for a_t, \dots, a_{t+k} , we see that

$$\mathbb{P}([A, t, C]) = O(m^{k+1} n^{-|S_1| - |S_2 \setminus S_1| - \dots - |S_k \setminus S_{k-1}|}) = O(m^{k+1} n^{-c(C)}) = O^*(n^{-c(C)}), \quad (3)$$

where we used the bound $m \leq \log n$, and also that k is bounded, since $k \leq \binom{r}{2}$.

For $\mathcal{C} = (C_1, \dots, C_\ell)$ a minimal chain cover of A with each chain C_j having length k_j , let $\mathcal{B}_{A, \mathcal{C}}$ be the event that there exist $0 \leq t_1 < t_2 < \dots < t_\ell$ such that $t_j + k_j \leq t_{j+1} - 1$ for each $1 \leq j \leq \ell - 1$ and $t_\ell + k_\ell \leq T$, and

$$[A, t_j, C_j] \text{ holds for } j = 1, 2, \dots, \ell. \quad (4)$$

Note that if the K_r -chain $(H_s)_{s=1}^t$ meets A in \mathcal{C} , then $\mathcal{B}_{A, \mathcal{C}}$ holds. Indeed, informally speaking, $\mathcal{B}_{A, \mathcal{C}}$ is the event that it is conceivable that our chain might meet A in \mathcal{C} . The

advantage of working with $\mathcal{B}_{A,\mathcal{C}}$ is that it avoids the complicated dependence introduced by the rule for selecting the next K_r in our chain.

Since $[A, t_j, C_j]$ depends only on the random variables $X_{t,a}$, $t_j \leq t \leq t_j + k_j$, $1 \leq a \leq m$, with \mathcal{C} and the t_j fixed the ℓ events appearing in (4) are independent. Hence, using (3) and summing over the at most T^ℓ choices for t_1, \dots, t_ℓ , we have

$$\mathbb{P}(\mathcal{B}_{A,\mathcal{C}}) = O^*(T^\ell n^{-c(C_1)-c(C_2)\dots-c(C_\ell)}) = O^*(T^\ell n^{-c(\mathcal{C})}) = O^*(n^{-\alpha_r b(\mathcal{C})-\varepsilon}),$$

where the final step follows from Corollary 11.

For $1 \leq q \leq \binom{r}{2}$, let $\mathcal{B}_{A,q}$ be the event that there exists some minimal chain cover \mathcal{C} of A with $b(\mathcal{C}) \geq q$ such that $\mathcal{B}_{A,\mathcal{C}}$ holds. Since there are $O(1)$ minimal chain covers of A , we see that

$$\mathbb{P}(\mathcal{B}_{A,q}) = O^*(n^{-\alpha_r q - \varepsilon}). \tag{5}$$

By Observation 9, for any $t \leq T$, if \mathcal{G}_t holds and A spans at least q edges in $G_t = \bigcup_{s \leq t} H_s$, with $A \neq V(H_s)$ for each $s \leq t$, then $\mathcal{B}_{A,q}$ holds. This allows us to prove our key lemma on the number of dangerous i -sets.

Recall that a set A of i vertices is dangerous at time t if adding all edges inside A to G_t would create an external copy of K_r^- (or of K_r) with vertex set $B \supset A$. We will consider the ‘bad’ event \mathcal{B}_t^1 which we define as

$$\mathcal{G}_t \cap \{ \text{for some } 2 \leq i \leq r-1, \text{ there are more than } n^{i-\varepsilon/2} \text{ dangerous } i\text{-sets at time } t \}.$$

Lemma 12. *For $r \geq 5$ and $\varepsilon > 0$ fixed, setting $T = n^{2-\alpha_r-\varepsilon}$, we have $\mathbb{P}(\bigcup_{t \leq T} \mathcal{B}_t^1) = o(1)$.*

Proof. For $2 \leq i \leq r-1$ let $Y_{t,i}$ denote the number of dangerous i -sets at time t . Let Z_i denote the number of r -sets A such that \mathcal{B}_{A,q_i} holds, where $q_i = \binom{r}{2} - 1 - \binom{i}{2}$. If \mathcal{G}_t holds and an i -set A' is dangerous at time t , then it is contained in an r -set A , not the vertex set of any H_s , $s \leq t$, such that A spans at least q_i edges in G_t . But then \mathcal{B}_{A,q_i} holds, as noted above. Since a particular r -set can be responsible for at most $\binom{r}{i} \leq 2^r$ i -sets being dangerous, it follows that, for any $t \leq T$, if \mathcal{G}_t holds then $Y_{t,i} \leq 2^r Z_i$. Thus it suffices to show that

$$\mathbb{P}(Z_i \geq 2^{-r} n^{i-\varepsilon/2}) = o(1), \tag{6}$$

for each i .

We claim that, for $2 \leq i \leq r-1$, we have

$$r-i \leq \alpha_r \left(\binom{r}{2} - 1 - \binom{i}{2} \right). \tag{7}$$

This holds for $i=2$ by definition of α_r . For $i=r-1$ it simplifies to $1 \leq \alpha_r(r-2)$ and hence to $(r-2)^2 \geq \binom{r}{2} - 2$, which is easily seen to hold for $r \geq 5$. Hence, by convexity, (7) holds for all $2 \leq i \leq r-1$. By (5) we have

$$\mathbb{E}[Z_i] = \sum_A \mathbb{P}(\mathcal{B}_{A,q_i}) = O^*(n^{r-\alpha_r q_i - \varepsilon}) = o(n^{r-\alpha_r q_i - \varepsilon/2}).$$

But then, by (7), $\mathbb{E}[Z_i] = o(n^{i-\varepsilon/2})$, and (6) follows by Markov’s inequality. □

Lemma 12 shows that one ‘global bad event’ is unlikely. The next lemma considers another, much simpler one.

Lemma 13. *Let $r \geq 5$, $\varepsilon > 0$ and set $T = n^{2-\alpha_r-\varepsilon}$. Let \mathcal{B}_t^2 be the event that the maximum degree of G_t is at least $n^{1-\varepsilon}$. Then $\mathbb{P}(\bigcup_{t \leq T} \mathcal{B}_t^2) = o(1)$.*

Proof. Suppose that a vertex v has degree at least $n^{1-\varepsilon}$ in some G_t . Then v must be in (crudely) at least $n^{1-\varepsilon}/r$ of the cliques H_s , $s \leq t$, and hence in at least $n^{1-\varepsilon}/(2r)$ of the selected sets X_s . Thus, v is in at least $n^{1-\varepsilon}/(2r)$ of the candidate sets $X_{s,a}$, $1 \leq s \leq T$, $1 \leq a \leq m$. For each v , let N_v be the number of these candidate sets that contain v . Then N_v has a binomial distribution with mean $m(T+1)(r-2)/n = O^*(n^{1-\alpha_r-\varepsilon}) = o(n^{1-\varepsilon}/(2r))$. Thus the probability that $N_v \geq n^{1-\varepsilon}/(2r)$ is $\exp(-\Omega(n^{1-\varepsilon})) = o(n^{-100})$, say, and the probability that there exists such a v is $o(1)$. \square

After this preparation we are now ready to prove Theorem 7.

Proof of Theorem 7. As above, we fix $r \geq 5$ and $\varepsilon > 0$, and define $m = \lceil \log n \rceil$ and $T = n^{2-\alpha_r-\varepsilon}$, where $\alpha_r = (r-2)/\binom{r}{2}$. We consider the outline construction described above, saying that a candidate choice $X_{t+1,a}$ for X_{t+1} is *successful* if it satisfies conditions (C0)–(C2) above, and (C3) and (C4) below. At each step $t+1 \leq T$ we let X_{t+1} be the successful candidate $X_{t+1,a}$ with the smallest index a if there is one, otherwise, our construction *fails at step $t+1$* .

Let \mathfrak{F}_t be the (finite, of course) σ -algebra generated by $(X_{s,a})_{s \leq t, a \leq m}$, i.e., by all information ‘revealed’ by time t . Notice that \mathfrak{F}_t encodes everything about our construction ‘up to time t ’. In particular, we know how $\bigcup_{s=1}^t H_s$ is defined at this point. Our key claim is that for $0 \leq t \leq T-1$ we have

$$\mathbb{P}(\text{candidate } X_{t+1,1} \text{ succeeds} \mid \mathfrak{F}_t) \geq 99/100 \tag{8}$$

whenever all previous steps have succeeded, and neither \mathcal{B}_t^1 nor \mathcal{B}_t^2 holds. Here, as usual, the probability of an event E , conditional on a σ -algebra \mathfrak{F} , is the random variable defined by the conditional expectation of the indicator of E with respect to \mathfrak{F} .

Suppose for the moment that the inequality (8) holds; we shall see that Theorem 7 easily follows. Indeed, because the $X_{t+1,a}$ are independent of each other and of \mathfrak{F}_t , and identically distributed, if (8) holds, then recalling that \mathcal{F}_{t+1} denotes the event that our construction fails at step $t+1$, we have

$$\mathbb{P}(\mathcal{F}_{t+1} \mid \mathfrak{F}_t) = \prod_{a=1}^m \mathbb{P}(\text{candidate } X_{t+1,a} \text{ fails} \mid \mathfrak{F}_t) \leq 100^{-m} \leq n^{-2}$$

whenever $(\mathcal{B}_t^1 \cup \mathcal{B}_t^2)^c \cap \mathcal{G}_t = (\mathcal{B}_t^1 \cup \mathcal{B}_t^2 \cup \mathcal{F}_1 \cup \dots \cup \mathcal{F}_t)^c$ holds. Since this latter event is \mathfrak{F}_t -measurable, it follows that

$$\mathbb{P}(\mathcal{F}_{t+1} \cap (\mathcal{B}_t^1 \cup \mathcal{B}_t^2 \cup \mathcal{F}_1 \cup \dots \cup \mathcal{F}_t)^c) \leq \mathbb{P}(\mathcal{F}_{t+1} \mid (\mathcal{B}_t^1 \cup \mathcal{B}_t^2 \cup \mathcal{F}_1 \cup \dots \cup \mathcal{F}_t)^c) \leq n^{-2}.$$

Also we have $\mathbb{P}(\mathcal{F}_1) \leq n^{-2}$. But then, considering the least t for which $\mathcal{B}_t^1 \cup \mathcal{B}_t^2 \cup \mathcal{F}_t$ holds, we see that

$$\mathbb{P}\left(\bigcup_{t \leq T} \mathcal{F}_t\right) \leq Tn^{-2} + \mathbb{P}\left(\bigcup_{t \leq T} \mathcal{B}_t^1\right) + \mathbb{P}\left(\bigcup_{t \leq T} \mathcal{B}_t^2\right) = o(1),$$

using Lemmas 12 and 13. Hence, with probability $1 - o(1) > 0$ (for n large enough), the construction succeeds for T steps. Then, as noted above, the fact that conditions (C0)–(C2) are satisfied implies that we construct a good K_r -chain of the required length. It remains only to establish (8).

From now on, we condition on \mathfrak{F}_t (i.e., on all $X_{s,a}$, $s \leq t$), and assume that our construction succeeded at steps $1, \dots, t$ and that neither \mathcal{B}_t^1 nor \mathcal{B}_t^2 holds. The only relevant randomness remaining is the uniform choice of $X_{t+1,1}$ from all sets of $r - 2$ vertices. Let \mathcal{E}_i be the event that $X_{t+1,1}$ fails to satisfy condition (Ci). To establish (8) it suffices to show that $\mathbb{P}(\mathcal{E}_i) = o(1)$, $i = 0, \dots, 4$. Clearly, $\mathbb{P}(\mathcal{E}_0) = 1 - \binom{n-2}{r-2} / \binom{n}{r-2} = o(1)$.

The case $i = 1$ is also easy. If \mathcal{E}_1 holds then either (i) $X_{t+1,1}$ includes two vertices forming an edge uv of G_t or (ii) for some vertex u of e_t , $X_{t+1,1}$ includes a vertex v such that $uv \in E(G_t)$. The first event has probability at most $\binom{r-2}{2}e(G_t)/\binom{n}{2} = O(e(G_t)n^{-2}) = O(Tn^{-2}) = o(1)$. Since \mathcal{B}_t^2 does not hold and there are only two choices for u , the second has probability $O(\Delta(G_t)/n) = o(1)$.

Before turning to the details, let us outline the argument for \mathcal{E}_2 . If \mathcal{E}_2 holds, then for some $0 \leq j \leq 2$ and $1 \leq i \leq r - 2$ with $i + j < r$, some subset of $V(e_t)$ of size j and some subset of $X_{t+1,1}$ of size i combine to form a dangerous set of size $i + j$. The case $j = 0$ will cause no problems, since we assume \mathcal{B}_t^1 does not hold, so there are $o(n^i)$ dangerous sets of size i . (This is analogous to (i) above.) For $j > 0$, we will have a problem only if some subset J of $V(e_t)$ of size j is in $\Theta(n^i)$ dangerous $(i + j)$ -sets. To avoid this, we consider a further condition

$$\text{For each } j = 1, 2 \text{ and } i \geq 1 \text{ with } i + j < r, \text{ no set of } j \text{ vertices of } X_{t+1,a} \text{ is contained in more than } n^{i-\varepsilon/4} \text{ dangerous } (i + j)\text{-sets at time } t. \quad (\text{C3})$$

This is not quite what we need, however – we would like to use this condition one step earlier to say that $V(e_t) \subset X_t$ has the properties we want. The problem is that condition (C3) having been satisfied by X_t only tells us that subsets of $V(e_t)$ are not in many sets that *were dangerous at time $t - 1$* , rather than at time t . On the other hand, the bound $n^{i-\varepsilon/4}$ in condition (C3) is much stronger than the bound $o(n^i)$ that we need. So we will be able to deal with this ‘off-by-one’ problem by showing that not too many sets become dangerous in one step. To make this argument work we will need one final condition. We say that a set A of vertices with $2 \leq |A| \leq r - 1$ is *deadly* at time t if there is a set B of r vertices with $A \subset B$ such that $G_t \cup A^{(2)}$ contains *all* edges within B , with B not equal to $V(H_s)$ for any $s \leq t$. Note that a deadly set is dangerous.

$$\text{For } j \in \{2, 3\}, \text{ no set of } j \text{ vertices of } e_t \cup X_{t+1,a} \text{ containing at most one vertex of } e_t \text{ and at least one vertex of } X_{t+1,a} \text{ is contained in more than } n^{1-\varepsilon/8} \text{ deadly } (j + 1)\text{-sets at time } t. \quad (\text{C4})$$

As noted above, it remains only to show that $\mathbb{P}(\mathcal{E}_i) = o(1)$ for $i = 2, 3, 4$. Before doing this, we establish some deterministic consequences of our conditions (C0)–(C4), i.e., of

succeeding with our choice at step $t+1$. In the following we assume that n is large enough, depending on the constant c .

Claim 1. *Let $c > 0$ be any positive constant. If the candidate $X_{t+1,1}$ satisfies conditions (C0)–(C4), then for $j = 1, 2$ and each $i \geq 1$ with $i + j < r$, every set J of j vertices from $X_{t+1,1}$ is in at most cn^i $(i + j)$ -sets that are dangerous at time $t + 1$.*

Proof. Suppose the claim fails for some i, j and J . Because condition (C3) is satisfied, we see that there are $\Theta(n^i)$ i -sets I disjoint from J such that $I \cup J$ is dangerous at time $t + 1$ but not at time t . Since $H_{t+1} = (e_t \cup X_{t+1,1})^{(2)}$ contains $r = O(1)$ vertices and $\Delta(G_{t+1}) \leq \Delta(G_t) + r - 1 = o(n)$ by our assumption that \mathcal{B}_t^2 does not hold, there are $\Theta(n^i)$ of these sets I with the additional property that no vertex in I is in H_{t+1} , or is adjacent in G_{t+1} to any vertex in H_{t+1} . For each such set I , by the definition of dangerous there is an r -set $A \supset I \cup J$ such that $G_{t+1} \cup (I \cup J)^{(2)}$ contains all but at most one of the edges within A . Furthermore, since $I \cup J$ is not dangerous at time t , there is such a set A such that $G_{t+1} \setminus G_t$ includes at least one edge within A but not within $I \cup J$. In particular, $A \setminus (I \cup J)$ contains at least one vertex of H_{t+1} . Let $L = V(H_{t+1}) \cap (A \setminus (I \cup J))$, so $L \neq \emptyset$.

By choice of I , there are no edges from I to H_{t+1} in G_{t+1} . Hence there are no edges from I to L in G_{t+1} . Since I, J and L are disjoint, in $G_{t+1} \cup (I \cup J)^{(2)}$ there are still no edges from I to L . Hence, in this final graph, there are at least $|I||L|$ missing edges inside A ; as at most one edge is missing, we conclude that $|I| = |L| = 1$. Furthermore, the only missing edge in A in $G_{t+1} \cup (I \cup J)^{(2)}$ is the I – L edge. Hence in $G_{t+1} \cup (I \cup J \cup L)^{(2)}$ the set A is complete. Since $V(H_{t+1}) \cap A = J \cup L$, we see that A is also complete in $G_t \cup (I \cup J \cup L)^{(2)}$, so $I \cup J \cup L$ is deadly at time t . Note that $|I \cup J \cup L| = j + 2 < r$.

Since we find one such deadly set $I \cup J \cup L$ for each of $\Theta(n)$ choices for I (recall that now $i = 1$), and there are at most r choices for L , we see that some set $J \cup L$ of $j + 1 = 2$ or 3 vertices of H_{t+1} (with at most one in e_t , namely that in L) is contained in $\Theta(n)$ deadly $(j + 2)$ -sets $I \cup J \cup L$. But this violates our assumption that condition (C4) holds, completing the proof of the claim. \square

Claim 2. *If the candidate $X_{t+1,1}$ satisfies conditions (C0)–(C4), then for $i \in \{2, 3\}$ each vertex of $X_{t+1,1}$ is contained in at most $2n^{i-\varepsilon/4}$ deadly $(i + 1)$ -sets at time $t + 1$.*

Proof. The proof is very similar to that of Claim 1. Adopting similar notation, if the claim fails there is some $i \in \{2, 3\}$ and some singleton set $J = \{u\} \subset X_{t+1,1}$ such that there are $2n^{i-\varepsilon/4}$ i -sets I disjoint from J with $I \cup J$ deadly at time $t + 1$. Since condition (C3) holds and a deadly set is dangerous, at least $n^{i-\varepsilon/4}$ of these sets I are such that $I \cup J$ is deadly at time $t + 1$ but not at time t . As before, since $\Delta(G_{t+1}) = O(n^{1-\varepsilon}) = o(n^{1-\varepsilon/4})$, we may find $\Theta(n^{i-\varepsilon/4}) > 0$ such sets I which are disjoint from H_{t+1} and send no edges to H_{t+1} in the graph G_{t+1} . Fix any such I . Then there is an r -set $A \supset I \cup J$ such that $G_{t+1} \cup (I \cup J)^{(2)}$ contains all edges in A with (as before) $L = V(H_{t+1}) \cap (A \setminus (I \cup J))$ non-empty. (Otherwise $I \cup J$ was deadly at time t also.) Since $L \subset V(H_{t+1})$, by choice of I there are no I – L edges in G_{t+1} , so the I – L edges (of which there is at least one) are missing in $G_{t+1} \cup (I \cup J)^{(2)}$ also, a contradiction. \square

We now show that for $k = 2, 3, 4$ the probability of the event \mathcal{E}_k that our first candidate $X_{t+1,1}$ fails to satisfy condition (Ck) is $o(1)$. Recall that we are assuming that our algorithm succeeded at earlier steps and that neither \mathcal{B}_t^1 nor \mathcal{B}_t^2 holds. If condition (C2) fails then for some $0 \leq j \leq 2$ and $1 \leq i \leq r - 2$ with $i + j < r$ there is an i -element subset I of $X_{t+1,1}$ and a j -element subset J of e_t such that $I \cup J$ is dangerous (at time t). For $j = 0$, since \mathcal{B}_t^1 does not hold, there are $o(n^i)$ dangerous i -sets in G_t . The probability that $X_{t+1,1}$ includes any given one is $O(n^{-i})$, so the probability that such I, J exist with $j = 0$ is $o(1)$. For $j \in \{1, 2\}$, applying Claim 1 with t replaced by $t - 1$ (for $t > 0$; for $t = 0$ there is no problem), we see that for each j -element subset J of $e_t \subset X_t$ there are $o(n^i)$ i -sets I such that $I \cup J$ is dangerous. It follows that $\mathbb{P}(\mathcal{E}_2) = o(1)$.

Turning to \mathcal{E}_3 , fix $j \in \{1, 2\}$ and $i \geq 1$ with $i + j < r$. Call a j -set *bad* if it is in at least $n^{i-\varepsilon/4}$ $(i + j)$ -sets that are dangerous at time t . Since at most $n^{i+j-\varepsilon/2}$ $(i + j)$ -sets are dangerous at time t , there are at most $\binom{i+j}{j} n^{j-\varepsilon/4} = O(n^{j-\varepsilon/4})$ bad j -sets. For condition (C3) to fail, $X_{t+1,1}$ must contain such a bad j -set, an event of probability $O(n^{j-\varepsilon/4} n^{-j}) = o(1)$.

Finally, we turn to \mathcal{E}_4 ; the argument is somewhat similar to that for \mathcal{E}_2 . Firstly, as \mathcal{B}_t^1 does not hold and a deadly set is dangerous, there are at most $(j + 1)n^{j-\varepsilon/4}$ j -sets J with the property that J is contained in more than $n^{1-\varepsilon/4}$ deadly $(j + 1)$ -sets at time t . Hence the probability that $X_{t+1,1}$ contains such a j -set is $O(n^{j-\varepsilon/4}/n^j) = o(1)$. For $j = 2, 3$ it remains to bound the probability that for some vertex v of e_t , the set $X_{t+1,1}$ contains a $(j - 1)$ -set J such that $J \cup \{v\}$ is in more than $n^{1-\varepsilon/8}$ deadly $(j + 1)$ -sets. This probability is $o(1)$ as required unless v is in $\Theta(n^{j-1})$ j -sets each contained in more than $n^{1-\varepsilon/8}$ deadly $(j + 1)$ -sets. But if this holds, v is in $\Theta(n^{j-\varepsilon/8})$ deadly $(j + 1)$ -sets. Since $v \in e_t \subset X_t$, this contradicts Claim 2 applied at time $t - 1$. \square

4 Deterministic constructions

In this section we present a deterministic algorithm that constructs a graph which stabilizes after at least $\frac{1}{200}n^{3/2}$ time steps in the K_5 -bootstrap process. This construction generalizes immediately to any $r \geq 6$ to obtain graphs that stabilize after at least $c_r n^{3/2}$ time steps in the K_r -bootstrap process. The result is not as strong as Theorem 2, but the proof, in addition to being deterministic, is much simpler, and perhaps is of independent interest.

Lemma 14. *For all $t \leq T = \frac{1}{200}n^{3/2}$ we can find a collection $\mathcal{A} = \{A_1, \dots, A_t\} \subseteq [n]^{(5)}$ and distinct $e_0 = \{1, 2\}, e_1, \dots, e_t \in [n]^{(2)}$ with the following properties:*

1. $e_0 \in A_1$, and $e_i = A_i \cap A_{i+1}$ for $i = 1, \dots, t - 1$;
2. $|A_i \cap A_j| \leq 1$ if $|i - j| \geq 2$;
3. there are no distinct $u, v, w \in [n]$ and $1 \leq i < j < k \leq t$, such that $\{u, v\} \subset A_i$, $\{u, w\} \subset A_j$ and $\{v, w\} \subset A_k$ (i.e., there are no ‘external triangles’ in the set system);

4. for all $u \in [n]$ we have $\deg(u) = |\{A_i \in \mathcal{A} : u \in A_i\}| \leq \frac{1}{20}\sqrt{n}$.

Proof. Let us define the neighbourhood $N(u)$ of $u \in [n]$ to be

$$N(u) = \{v \in [n] : \text{there exists some } A_i \in \mathcal{A} \text{ such that } u, v \in A_i\}.$$

Also define the second neighbourhood of u to be

$$N^2(u) = \bigcup_{v \in N(u)} N(v).$$

Notice that $u \in N(u) \subseteq N^2(u)$. Also, if for all $u \in [n]$ we have $\deg(u) \leq \Delta$, then $|N^2(u)| < 25\Delta^2$.

Let us proceed by induction on t . Assume that we have sets A_1, \dots, A_{t-1} , and pairs $e_0 = \{1, 2\}, e_1, \dots, e_{t-1}$, satisfying the lemma. We want to find some A_t and e_t while $t \leq n^{3/2}/200$. Let $e_{t-1} = \{u, v\}$. We shall take $A_t = e_{t-1} \cup \{w_1, w_2, w_3\}$ for some distinct $w_1, w_2, w_3 \notin e_{t-1}$, and then set $e_t = \{w_1, w_2\}$. We do this by choosing w_i , one at a time, so that, for $i = 1, 2, 3$, the vertex w_i is not contained in

$$N^2(u) \cup N^2(v) \cup N^2(w_1) \cup \dots \cup N^2(w_{i-1}),$$

where in defining $N^2(\cdot)$ we include the set $A_t = \{u, v, w_1, \dots, w_{i-1}\}$. By the condition on the degrees of all vertices,

$$|N^2(u)| + |N^2(v)| + |N^2(w_1)| + \dots + |N^2(w_{i-1})| < 100\left(\frac{1}{20}\sqrt{n} + 1\right)^2 < n/2. \quad (9)$$

Choosing w_i that does not belong to any of the above second neighbourhoods ensures that there are no external triangles formed by the addition of A_t and that $|A_t \cap A_j| \leq 1$ for all $j \leq t - 2$.

By (9), if the maximum degree of a vertex is at most $\frac{1}{20}\sqrt{n}$, then there are at least $\frac{1}{2}n$ possible choices for each of w_1, w_2, w_3 . Now, for each $i = 1, 2, 3$, choose w_i to be the vertex with the smallest degree among these choices. If this choice makes the degree of w_i larger than $\frac{1}{20}\sqrt{n}$, then all of these $\frac{1}{2}n$ vertices have degree at least $\frac{1}{20}\sqrt{n}$. Hence there are at least $T = \frac{1}{200}n^{3/2}$ sets in \mathcal{A} . This completes the proof. \square

Let us show how to turn the set system \mathcal{A} into a graph that has a long running time in the K_5 -bootstrap process. For $T = \frac{1}{200}n^{3/2}$, given \mathcal{A} and e_1, \dots, e_t as defined above, we define $G = ([n], E)$, where

$$E = \left(\bigcup_{i=1}^T A_i^{(2)} \right) \setminus \{e_1, \dots, e_T\}. \quad (10)$$

We shall show that the K_5 -bootstrap process starting from $G_0 = G$ takes T steps to stabilize. To do this, we need one further definition. Let \mathcal{B} be a collection of subsets of $[n]$ and let H be a graph with vertex set $[n]$. We say that H is *covered* by \mathcal{B} if every edge $e \in H$ is contained in some set $B \in \mathcal{B}$. We say that H is *simply covered* by \mathcal{B} if there exists some $B \in \mathcal{B}$ such that $e \subseteq B$ for every $e \in H$. The following observation follows immediately from the third condition in Lemma 14.

Observation 15. Let \mathcal{A} be as in Lemma 14. For any distinct $u, v, w \in [n]$, if \mathcal{A} covers $\{u, v, w\}^{(2)}$ then it simply covers it.

Lemma 16. Let $G_0 = G$ be defined as in Equation (10) and let $(G_t)_{t \leq T}$ be the steps of the K_5 -process, starting from G . Then for all $1 \leq t \leq T$ we have $G_t \setminus G_{t-1} = \{e_t\}$.

Proof. Suppose that the lemma is false and let $t' \geq 1$ be the first time when $G_{t'} \setminus G_{t'-1} \neq \{e_{t'}\}$. Hence, for all $1 \leq t < t'$ we have $G_t \setminus G_{t-1} = \{e_t\}$. Clearly $A_{t'}$ induces a K_5^- in $G_{t'-1}$ because $e_{t'-1} \in G_{t'-1}$ and, by the definition of \mathcal{A} , all edges $\{e_1, \dots, e_{t'}\}$ are distinct which implies that $e_{t'} \notin G_{t'-1}$. Hence we have $e_{t'} \in G_{t'} \setminus G_{t'-1}$ and, by the definition of t' , we must have some $e \neq e_{t'}$ such that $e \in G_{t'} \setminus G_{t'-1}$.

If e is included at time t' then it appears by virtue of some K_5^- that is contained in $G_{t'-1}$. If there was no A_i that induced all the edges of this K_5^- in $G_{t'-1}$ then, since $G_{t'-1}$ is a subset of $\bigcup_{i=1}^{t'} A_i^{(2)}$, the set system \mathcal{A} would cover a triangle that it would not simply cover. That would contradict Observation 15.

Hence $e = e_k$ for some $k > t'$. However, only two sets in \mathcal{A} contain e_k namely, A_k and A_{k+1} . By the definition of t' we have $e_{k-1}, e_k, e_{k+1} \notin G_{t'-1}$ therefore both A_k and A_{k+1} induce $\binom{5}{2} - 2$ edges in $G_{t'-1}$. Hence it is impossible that $e_k \in G_{t'} \setminus G_{t'-1}$. \square

Lemma 16 immediately implies that the K_5 -bootstrap process starting from G stabilizes after at least $\frac{1}{200}n^{3/2}$ time steps.

5 Open problems

In this paper we considered the problem of finding the maximum running time of the K_r -bootstrap process. For $r \geq 5$ the exact answer remains to be found and it is the obvious open problem in this direction. In the proof of Theorem 2 we were a little careless with the $o(1)$ term in the exponent; a more careful version of the argument will give a slightly stronger bound (presumably $n^{2-\alpha r}$ divided by some power of $\log n$), but still $o(n^{2-\alpha r})$. It seems unlikely that any random construction of this type can go (significantly, if at all) beyond $n^{2-\alpha r}$, since at this point, in a random graph with the relevant number of edges, many missing edges e have the property that adding e would create a K_r^- . It is tempting to think that therefore $M_r(n) = n^{2-\alpha r + o(1)}$ for $r \geq 5$, but we have no real reason to believe this. Instead, since proving any non-trivial upper bound on $M_r(n)$ is open for $r \geq 5$, we make the following much weaker conjecture.

Conjecture 1. For all $r \geq 5$ we have $M_r(n) = o(n^2)$.

To support our conjecture, we use an argument of Erdős, Frankl, Ródl [14] to note that any graph ‘resembling’ those constructed in Lemma 16 and Theorem 2 cannot run for time $\Omega(n^2)$. In particular, suppose G is a graph that is constructed by chaining together 5-sets that do not admit any external triangles (as in condition 3 of Lemma 14) and such that each 5-set in the chain intersects at most two other 5-sets (again, as in Lemma 14). Now let T be the running time of the K_5 -process on G and observe that G_T will have at most $O(n^2)$ triangles, as no external triangles were created. Hence, by the triangle removal

lemma (see, for example Theorem 1.5 in [14]), we may remove a collection of $o(n^2)$ edges so that the remaining graph is triangle free. However, every edge is contained in at most 2 of the 5-sets, from which the graph was constructed, and therefore the number of such 5-sets must be $o(n^2)$, as each contains a triangle. Since at least one of our collection of 5-sets becomes filled in at each step of the K_5 -process, the process can run for at most $o(n^2)$ steps.

References

- [1] N. Alon. An extremal problem for sets with applications to graph theory. *Journal of Combinatorial Theory, Series A*, 40:82–89, 1985.
- [2] J. Balogh, B. Bollobás, H. Duminil-Copin, and R. Morris. The sharp threshold for bootstrap percolation in all dimensions. *Transactions of the American Mathematical Society*, 364:2667–2701, 2012.
- [3] J. Balogh, B. Bollobás, and R. Morris. Graph bootstrap percolation. *Random Structures and Algorithms*, 41:413–440, 2012.
- [4] J. Balogh, B. Bollobás, R. Morris, and O. Riordan. Linear algebra and bootstrap percolation. *Journal of Combinatorial Theory, Series A*, 119:1328–1335, 2012.
- [5] J. Balogh and G. Pete. Random disease on the square grid. *Random Structures and Algorithms*, 13:409–422, 1998.
- [6] J. Balogh and B. G. Pittel. Bootstrap percolation on the random regular graph. *Random Structures and Algorithms*, 30:257–286, 2007.
- [7] F. S. Benevides and M. Przykucki. On slowly percolating sets of minimal size in bootstrap percolation. *Electronic Journal of Combinatorics*, 20(2):#P46, 2013.
- [8] F. S. Benevides and M. Przykucki. Maximum percolation time in two-dimensional bootstrap percolation. *SIAM Journal on Discrete Mathematics*, 29:224–251, 2015.
- [9] B. Bollobás. Graph bootstrap percolation. To appear.
- [10] B. Bollobás. Weakly k -saturated graphs. In *In Beiträge zur Graphentheorie (Kolloquium, Manebach, 1967)*, H. Sachs, H. J. Voss, and H. Walther (Editors), Teubner, Leipzig, Barth, pages 25–31, 1968.
- [11] B. Bollobás, K. Gunderson, C. Holmgren, S. Janson, and M. Przykucki. Bootstrap percolation on Galton–Watson trees. *Electronic Journal of Probability*, 19:1–27, 2014.
- [12] J. Chalupa, P.L. Leath, and G.R. Reich. Bootstrap percolation on a Bethe lattice. *Journal of Physics C*, 12:L31–L35, 1979.
- [13] A. van Enter. Proof of Straley’s argument for bootstrap percolation. *Journal of Statistical Physics*, 48:943–945, 1987.
- [14] P. Erdős, P. Frankl, and V. Rödl. The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs and Combinatorics*, 2:113–121, 1968.

- [15] P. Frankl. An extremal problem for two families of sets. *European Journal of Mathematics*, 3:125–127, 1982.
- [16] K. Gunderson, S. Koch, and M. Przykucki. Time of graph bootstrap percolation. To appear, *Random Structures and Algorithms*.
- [17] A. E. Holroyd. Sharp metastability threshold for two-dimensional bootstrap percolation. *Probability Theory and Related Fields*, 125:195–224, 2003.
- [18] S. Janson, T. Łuczak, T. Turova, and T. Vallier. Bootstrap percolation on the random graph $G_{n,p}$. *Annals of Applied Probability*, 22:1989–2047, 2012.
- [19] G. Kalai. Weakly saturated graphs are rigid. In *Convexity and graph theory*, M. Rosenfeld and J. Zaks (Editors), *Annals of Discrete Mathematics*, pages 189–190, 1984.
- [20] K. Matzke. The saturation time of graph bootstrap percolation. Preprint, <http://arxiv.org/abs/1510.06156>.
- [21] R. Morris. Minimal percolating sets in bootstrap percolation. *Electronic Journal of Combinatorics*, 16(1):#R2, 2009.
- [22] N. Morrison and J. Noel. Extremal bounds for bootstrap percolation in the hypercube. Preprint, <http://arxiv.org/abs/1506.04686v1>.
- [23] M. Przykucki. Maximal percolation time in hypercubes under 2-bootstrap percolation. *Electronic Journal of Combinatorics*, 19(2):#P41, 2012.
- [24] E. Riedl. Largest minimal percolating sets in hypercubes under 2-bootstrap percolation. *Electronic Journal of Combinatorics*, 17:#R80, 2010.
- [25] R.H. Schonmann. On the behaviour of some cellular automata related to bootstrap percolation. *Annals of Probability*, 20:174–193, 1992.