Constrained graph processes

Béla Bollobás and Oliver Riordan

Department of Mathematical Sciences
University of Memphis, Memphis TN 38152

Trinity College, Cambridge CB2 1TQ, England
bollobas@msci.memphis.edu, O.M.Riordan@dpmms.cam.ac.uk

Submitted: 25th June 1999; Accepted: 23rd February 2000

Keywords: random graphs
MR Subject Code: 05C80

Abstract

Let $\mathcal{Q}$ be a monotone decreasing property of graphs G on n vertices. Erdős, Suen and Winkler [5] introduced the following natural way of choosing a random maximal graph in $\mathcal{Q}$: start with G the empty graph on n vertices. Add edges to G one at a time, each time choosing uniformly from all $e \in G^c$ such that $G + e \in \mathcal{Q}$. Stop when there are no such edges, so the graph G_∞ reached is maximal in $\mathcal{Q}$. Erdős, Suen and Winkler asked how many edges the resulting graph typically has, giving good bounds for $\mathcal{Q} = \{\text{bipartite graphs}\}$ and $\mathcal{Q} = \{\text{triangle free graphs}\}$. We answer this question for C_4-free graphs and for K_4-free graphs, by considering a related question about standard random graphs $G_p \in \mathcal{G}(n,p)$.

The main technique we use is the 'step by step' approach of [3]. We wish to show that G_p has a certain property with high probability. For example, for K_4 free graphs the property is that every 'large' set V of vertices contains a triangle not sharing an edge with any K_4 in G_p. We would like to apply a standard Martingale inequality, but the complicated dependence involved is not of the right form. Instead we examine G_p one step at a time in such a way that the dependence on what has gone before can be split into 'positive' and 'negative' parts, using the notions of up-sets and down-sets. The relatively simple positive part is then estimated directly. The much more complicated negative part can simply be ignored, as shown in [3].

1 Introduction

A property $\mathcal{R}$ of graphs on n vertices is called *monotone increasing* (*monotone decreasing*) if it is preserved by the addition (deletion) of edges. Let V be a fixed set of n vertices, and let $N = \binom{n}{2}$. A *standard random graph process* on V is a random sequence $\tilde{G} = (G_t)_0^N$ of graphs on V, where $G_{t-1} \subset G_t$, $e(G_t) = t$, and all $N!$ such sequences are taken equally likely. A basic question in the theory of random graphs is when does a monotone increasing property $\mathcal{R}$ arise in such a process. More precisely, one would like to know as much as possible about the distribution of the *hitting time* $\tau_{\mathcal{R}}(\tilde{G})$, the minimum t such that $G_t \in \mathcal{R}$ (see, e.g., [1]).

Here we shall consider monotone decreasing properties $\mathcal{Q}$, and one could consider similarly the *leaving time* $\sigma_{\mathcal{Q}}(\tilde{G}) = \tau_{\mathcal{Q}^c}(\tilde{G}) - 1$, and the random graph $G = G_{\sigma_{\mathcal{Q}}(\tilde{G})} \in \mathcal{Q}$.

We wish to consider random *maximal* graphs in a monotone decreasing property $\mathcal{Q}$. The maximal graphs in $\mathcal{Q}$ are of interest both from the point of view of extremal combinatorics, and because they may provide a relatively compact description of the entire property $\mathcal{Q}$. Note that the random $G \in \mathcal{Q}$ described above does satisfy $G + e \notin \mathcal{Q}$ for some edge e, but need not be a maximal element of $\mathcal{Q}$.

At first sight the most natural measure on maximal $G \in \mathcal{Q}$ is the uniform one. Another natural possibility would be taking the probability of G proportional to $\binom{N}{e(G)}^{-1}$. However, both these measures are rather intractable in general—generating a random sample from either seems difficult, as we do not know how many $G \in \mathcal{Q}$ are maximal, or the distribution of the number of edges of such graphs.

In [5] Erdős, Suen and Winkler introduced a rather different measure on the set of maximal $G \in \mathcal{Q}$. This is also very natural, and is defined in terms of the 'greedy algorithm' for generating maximal $G \in \mathcal{Q}$, and so is easy to sample in practice. The procedure for constructing a random maximal $G_\infty \in \mathcal{Q}$ with this measure is as follows. Start with G the empty graph on n vertices. Add edges to G one by one, at each stage choosing uniformly from among all edges $e \in G^c$ such that $G \cup \{e\} \in \mathcal{Q}$. Stop when there are no such edges, i.e., when the graph G_∞ reached is a maximal graph in $\mathcal{Q}$. From now on when we refer to a *random maximal graph* from $\mathcal{Q}$ we are using this definition. Note that it is very different from any of the other models for random graphs from $\mathcal{Q}$ described above.

In [5] Erdős, Suen and Winkler asked the general question of how many edges G_∞ has on average. For the case $\mathcal{Q} = \{\text{bipartite graphs}\}$ they gave a very precise answer, and for $\mathcal{Q} = \{\text{triangle free graphs}\}$ the answer to within a $\log n$ factor. Here we give answers within powers of $\log n$ for the cases of C_4-free graphs and K_4-free graphs, using the 'step by step' methods of [3].

For convenience we shall not work with the process above, but with an equivalent one, $\tilde{G}_{\mathcal{Q}} = (G_{\mathcal{Q}}(t))_0^N$, also used in [5]. Fix a set V of n vertices. Let $N = \binom{n}{2}$, and let $e_1, \dots, e_N$ be all elements of $V^{(2)}$, listed in a uniformly chosen random order. Let $G_{\mathcal{Q}}(0) = \emptyset$. For $1 \le t \le N$ let

$$G_{\mathcal{Q}}(t) = \begin{cases} G_{\mathcal{Q}}(t-1) \cup \{e_t\} & \text{if } G_{\mathcal{Q}}(t-1) \cup \{e_t\} \in \mathcal{Q} \\ G_{\mathcal{Q}}(t-1) & \text{otherwise,} \end{cases}$$

and let $G_\infty = G_{\mathcal{Q}}(N)$. This definition is equivalent to the description above, where the edge to be added was chosen from all $e \notin G$ such that $G + e \in \mathcal{Q}$. The reason is that if we do not add the edge e_t at stage t, we have $G_{\mathcal{Q}}(s) \cup \{e_t\} \notin \mathcal{Q}$ for all $s \ge t$, so we never need to consider adding the edge e_t at a later stage.

We shall couple $G_{\mathcal{Q}}(t)$ with two processes that are easier to analyze, and which approximate $G_{\mathcal{Q}}(t)$. For $0 \le t \le N$ let $G_0(t) = (V, \{e_i, i \le t\})$, so $e(G_0(t)) = t$, and $(G_0(t))_{t=0}^N$ is a standard random graph process with $G_{\mathcal{Q}}(t) \subset G_0(t)$. Let $M(\mathcal{Q}^c)$ consist of all the minimal elements of $\mathcal{Q}^c$, so $G \notin \mathcal{Q}$ if and only if G contains some graph in $M(\mathcal{Q}^c)$. In the cases we consider, $\mathcal{Q}$ is all graphs not containing a copy of some fixed graph H, so $M(\mathcal{Q}^c)$ just consists of all copies of H on V. Let $G'_{\mathcal{Q}}(t)$ consist of those edges e of $G_0(t)$ which are not contained in some $F \subset G_0(t)$ with $F \in M(\mathcal{Q}^c)$. Then we have $G'_{\mathcal{Q}}(t) \subset G_{\mathcal{Q}}(t)$—indeed if $e = e_s \in G_0(t) \setminus G_{\mathcal{Q}}(t)$ then because e was not added at stage s, there is a graph $F \subset G_{\mathcal{Q}}(s-1) \cup \{e_s\}$ with $F \in M(\mathcal{Q}^c)$. But then $F \subset G_0(s) \subset G_0(t)$, so $e \notin G'_{\mathcal{Q}}(t)$.

In fact we shall not work with graph processes at all, but rather with a random graph $G_p \in \mathcal{G}(n,p)$ chosen by joining each pair of vertices independently with probability p. We obtain a graph G'_p from G_p by deleting any edge contained in some $F \subset G_p$ with $F \in M(\mathcal{Q}^c)$. We can couple the random variables G_p, G'_p with the processes above: let $T \sim \mathrm{Bi}(N,p)$. Then the graph $G_0(T)$ is a random graph G_p from $\mathcal{G}(n,p)$ with the correct distribution. Also, the graph $G'_{\mathcal{Q}}(T)$ has the correct distribution for G'_p. Since every $G'_{\mathcal{Q}}(t)$ is contained in $G_{\mathcal{Q}}(t)$ and thus G_∞, we have $G'_p \subset G_\infty$. This is *all* we shall use from now on, not only for lower bounds but, somewhat surprisingly, even to get upper bounds on $e(G_\infty)$.

In vague terms, as p increases from 0 to 1 the graphs G'_p first get larger, and then smaller again. We shall show that, in the cases we consider, G_∞ is not much larger than the largest G'_p. We suspect that this holds in many other cases, though it is not at all true for $\mathcal{Q} = \{\text{bipartite graphs}\}$, for example.

The rest of the paper is organized as follows. In §2 we state our main results, giving probabilistic upper and lower bounds on $e(G_\infty)$ for the properties $\{G \text{ is } C_4\text{-free}\}$ and $\{G \text{ is } K_4\text{-free}\}$. In §3 we give the simple proof of the lower bound. In §4 we quote two basic lemmas needed in the rest of the paper. In §5 we prove a lemma concerning the number of copies of a fixed graph H containing some edge $xy \in G_p$. This lemma, which is used in both the subsequent sections, is likely to be of interest in its own right. In §6 we give the upper bound for C_4-free graphs, and in §7 that for K_4-free graphs. In the final section we briefly discuss possible generalizations.

Throughout the paper we shall assume that the number n of vertices is larger than some very large fixed n_0, even when this is not explicitly stated. We shall use the notation $f = O(g)$ to mean that f/g is bounded for $n \geq n_0$, $f = \Theta(g)$ to mean $f = O(g)$ and $g = O(f)$, and $f = O^*(g)$ to mean that $f = O((\log n)^k g)$ for some fixed k.

2 Results

Throughout we take $\mathcal{Q}$ to be $\mathcal{Q}_H$, the set of H-free graphs with vertex set $V = [n] = \{1, 2, \ldots, n\}$, i.e., the set of graphs on V not containing a copy (induced or otherwise) of a fixed graph H. We shall consider the cases $H = C_4$ and $H = K_4$. Parts of the argument are the same for both cases, and work for a much larger class of graphs, which we now describe.

Let H be a fixed graph. For $0 \leq v < |H|$ let $e_H(v)$ be the maximum number of edges spanned by v vertices of H. Let

$$\alpha_H(v) = \frac{e(H) - e_H(v)}{|H| - v}.$$

We say that H is *edge-balanced* if H is connected, $|H| \geq 3$, and $\alpha_H(v) > \alpha_H(2)$ for $2 < v < |H|$. Writing $\mathrm{aut}(H)$ for the number of automorphisms of H, we shall prove the following lower bound on $e(G'_p)$ when G'_p is defined with respect to $\mathcal{Q}_H$.

Theorem 1. *Let H be a fixed edge-balanced graph, λ and ϵ positive constants, and*

$$p = \lambda n^{-\frac{|H|-2}{e(H)-1}}.$$

Then with G'_p defined as above with respect to $\mathcal{Q} = \mathcal{Q}_H$,

$$\mathbb{P}\left(e(G'_p) < (1-\epsilon)\left(\frac{\lambda}{2} - \frac{\lambda^{e(H)} e(H)}{\mathrm{aut}(H)}\right) n^{2-\frac{|H|-2}{e(H)-1}}\right) = o(1),$$

as $n \to \infty$.

This has the following immediate corollary.

Corollary 2. *Let H be a fixed edge-balanced graph, and let G_∞ be a random maximal H-free graph. Then there is a constant $c = c(H) > 0$ such that*

$$\mathbb{P}\left(e(G_\infty) < c n^{2-\frac{|H|-2}{e(H)-1}}\right) = o(1), \tag{1}$$

and $\mathbb{E}\, e(G_\infty) \geq (c/2) n^{2-\frac{|H|-2}{e(H)-1}}$.

Proof. The second statement follows from the first as $e(G_\infty) \geq 0$. For the first, we have $G_\infty \supset G'_p$ for all p. Taking $\epsilon = \frac{1}{2}$, say, and $\lambda = (\mathrm{aut}(H)/4e(H))^{1/(e(H)-1)}$, Theorem 1 implies (1) with $c = \lambda/8$. $\square$

In the other direction we shall prove the following results for $H = C_4$ and $H = K_4$, writing $\Delta(G)$ for the maximum degree of G.

Theorem 3. *For G_∞ a random maximal C_4-free graph we have*

$$\mathbb{P}\left(\Delta(G_\infty) > 13(\log n)^3 n^{1/3}\right) = o(n^{-2}).$$

In particular,

$$\mathbb{P}\left(e(G_\infty) > 7(\log n)^3 n^{4/3}\right) = o(n^{-2}),$$

and $\mathbb{E}\, e(G_\infty) \leq 8(\log n)^3 n^{4/3}$.

Theorem 4. *There is a constant C such that for G_∞ a random maximal K_4-free graph we have*

$$\mathbb{P}\left(\Delta(G_\infty) > 2C(\log n) n^{3/5}\right) = o(n^{-2}).$$

In particular,

$$\mathbb{P}\left(e(G_\infty) > C(\log n) n^{8/5}\right) = o(n^{-2}),$$

and $\mathbb{E}\, e(G_\infty) \leq 2C(\log n) n^{8/5}$.

Note that $2 - \frac{|H|-2}{e(H)-1}$ is equal to $\frac{4}{3}$ for $H = C_4$, and to $\frac{8}{5}$ for $H = K_4$, so by Corollary 2 in these cases we have found $e(G_\infty)$ to within a log factor for almost every G_∞. In fact our proofs of Theorems 3 and 4 give error bounds smaller than n^{-k} for any fixed k, and possibly even $n^{-\delta \log n}$ for $\delta > 0$ small enough.

In the next section we give the straightforward proof of Theorem 1. The heart of the paper is the proofs of the upper bounds.

3 Proof of the lower bound

We shall use Janson's inequality [6] in the following form. Let H be a fixed graph, and V a set of n vertices. Let $H_1, \ldots, H_h$ be all copies of H with vertices in V, so $h = (n)_{|H|}/\mathrm{aut}(H)$. Let $X = X_H(G_p)$ be the number of copies of H present in G_p, so $\mu = \mathbb{E}\, X = hp^{e(H)}$, and let

$$\Delta = \sum_{e(H_i \cap H_j) > 0} \mathbb{P}(H_i \cup H_j \subset G_p).$$

Then for $\gamma > 0$,

$$\mathbb{P}(X \leq (1-\gamma)\mu) < e^{-\frac{\gamma^2 \mu}{2+2\Delta/\mu}}, \tag{2}$$

and for $\epsilon > 0$,

$$\mathbb{P}(X \geq (1+\epsilon)\mu) \leq \frac{\gamma + e^{-\gamma^2\mu/(2+2\Delta/\mu)}}{\epsilon}. \tag{3}$$

Note that (2) implies (3) as

$$\mu \geq (1-\gamma)\mu\, \mathbb{P}(X \geq (1-\gamma)\mu) + \epsilon\mu\, \mathbb{P}(X \geq (1+\epsilon)\mu)).$$

Proof of Theorem 1. The graph G'_p is formed from G_p by deleting the edges of each copy of H in G_p, so $e(G'_p) \geq e(G_p) - e(H)X$, where $X = X_H(G_p)$. Writing N for $\binom{n}{2}$,

$$\mathbb{E}\, e(G_p) = pN \sim \frac{\lambda}{2} n^{2-\frac{|H|-2}{e(H)-1}},$$

while

$$\mu = \mathbb{E}\, X \sim \frac{\lambda^{e(H)}}{\operatorname{aut} H} n^{2-\frac{|H|-2}{e(H)-1}},$$

so it suffices to show that

$$e(G_p) \geq (1 - o(1))\, \mathbb{E}\, e(G_p) \tag{4}$$

and

$$X \leq (1 + o(1))\mu \tag{5}$$

hold with probability $1 - o(1)$.

As $pN \to \infty$, (4) is immediate from standard binomial bounds. For (5) we use Janson's inequality. Consider one particular copy H_1 of H on V. Then by symmetry

$$\Delta \leq hp^{e(H)} \sum_{i:e(H_i \cap H_1)>0} \mathbb{P}(H_i \subset G_p \mid H_1 \subset G_p).$$

Writing K for the complete graph on the vertex set of H_1 we thus have

$$\Delta \leq \mu \sum_{i:V(H_i \cap K)\geq 2} \mathbb{P}(H_i \subset G_p \mid K \subset G_p).$$

We can choose H_i by deciding v, the number of vertices to take from K, which v vertices to take from K, which $|H| - v$ vertices outside K to take, and how to arrange H_i on these $|H|$ vertices. As H_i has at least $e(H) - e_H(v)$ edges outside K, we have

$$\begin{aligned} \Delta &\leq \mu \sum_{v=2}^{|H|} \binom{|H|}{v} \binom{n}{|H|-v} |H|! p^{e(H)-e_H(v)} \\ &= O\left(\mu \sum_{v=2}^{|H|} n^{|H|-v} p^{e(H)-e_H(v)} \right). \end{aligned}$$

For $v = 2$ or $v = |H|$ the summand above is $O(1)$. Also, as H is edge-balanced, for $2 < v < |H|$ we have

$$(e(H) - e_H(v)) \frac{|H| - 2}{e(H) - 1} > |H| - v,$$

so the remaining terms of the sum are all $o(1)$. Thus $\Delta = O(\mu)$. Now fix $\epsilon > 0$ and set $\gamma = \epsilon^2$. Since $\Delta = O(\mu)$ and $\mu \to \infty$, inequality (3) implies that

$$\mathbb{P}(X \geq (1+\epsilon)\mu) \leq \frac{\epsilon^2 + o(1)}{\epsilon},$$

which is less than 2ϵ for n large. As ϵ was arbitrary, (5) holds almost surely, completing the proof. $\square$

Note that Theorem 1 can be strengthened in two ways. We can remove the factor $e(H)$ from the term $\lambda^{e(H)}e(H)/\operatorname{aut}(H)$ if we define G'_p by deleting only one edge from each copy of H in G_p. Choosing this edge to be the last edge in a random order on $V^{(2)}$, we can still couple this larger G'_p with G_∞ so that $G'_p \subset G_\infty$. Independently, we can obtain much smaller error probabilities (for example n^{-k} for any fixed k) by using the Azuma-Hoeffding inequality together with Lemma 8 from §4.

4 Basic lemmas

In the rest of the paper we shall need the following results: Janson's inequality (2), some standard bounds concerning the binomial distribution, and a lemma from [3] concerning up-sets and down-sets. To bound the tail of the binomial distribution we use the following lemma from [3], itself an immediate consequence of the Chernoff bounds [4] (see also [1], p.11).

Lemma 5. *Let X be a $\mathrm{Bi}(n,p)$ random variable, with $0 < p \leq \frac{1}{2}$. Then*

$$(a) \qquad \mathbb{P}(X < \frac{1}{2}pn) < \left(\frac{2}{e}\right)^{\frac{pn}{2}} < e^{-\frac{pn}{8}},$$

and if $k \geq 1$ and $\frac{pn}{k} < e^{-2}$ then

$$(b) \qquad \mathbb{P}(X > k) < \left(\frac{epn}{k}\right)^k < e^{-k}.$$

The main tool in the proofs of Theorems 3 and 4 will be the 'step by step' approach of [3], making use of up-sets and down-sets. An *up-set* $\mathcal{U}$ on a set W is a collection of subsets of W such that $A \in \mathcal{U}$ and $A \subset B \subset W$ imply $B \in \mathcal{U}$. A *down-set* $\mathcal{D}$ is one where $A \in \mathcal{D}$ and $B \subset A$ imply $B \in \mathcal{D}$. In the graph context, W is just the set $V^{(2)}$ of possible edges.

We wish to check that G_p satisfies a certain rather complicated condition with very high probability. We do this by considering a (completely impractical) algorithm which checks whether G_p satisfies this condition 'a bit at a time'. At each stage the algorithm tests whether the edges in a certain set E are all present in G_p, basing its subsequent behaviour on the yes/no answer. We design the algorithm so that the event $\mathcal{A}$ that the algorithm reaches any particular state has the form $\mathcal{A} = \mathcal{U} \cap \mathcal{D}$, where $\mathcal{U}$ is a very simple up-set, and $\mathcal{D}$ is some down-set. We can then bound the probability that $E \subset G_p$ given $\mathcal{A}$ using the following lemma from [3], itself a simple consequence of Kleitman's Lemma [7], which states that up-sets are positively correlated (see also [2], §19).

Lemma 6. *Let $\mathbf{p} = (p_1, \ldots, p_N)$, where each p_i lies between 0 and 1. Let $Q_\mathbf{p}$ be the weighted cube, i.e., the probability space with underlying set $\mathcal{P}([N])$ where a random subset $X \subset W = [N]$ is obtained by selecting elements of W independently, with $\mathbb{P}(i \in X) = p_i$, $i = 1, \ldots, N$. Let $\mathcal{U}_1$ and $\mathcal{U}_2$ be up-sets with $\mathbb{P}(\mathcal{U}_1 \cap \mathcal{U}_2) = \mathbb{P}(\mathcal{U}_1)\,\mathbb{P}(\mathcal{U}_2)$ and let $\mathcal{D}$ be a down-set. Then*

$$\mathbb{P}(\mathcal{U}_1 \cap \mathcal{U}_2 \cap \mathcal{D}) \leq \mathbb{P}(\mathcal{U}_1)\,\mathbb{P}(\mathcal{U}_2 \cap \mathcal{D}).$$

For the rest of the paper we work with the probability space $\mathcal{G}(n,p)$ of graphs on a fixed vertex set V. In this context an up-set (down-set) is just a monotone increasing (decreasing) property of graphs on V. Note that we shall not distinguish sets $\mathcal{A}$ of graphs on V from the corresponding events $\{G_p \in \mathcal{A}\}$. With this notation the most convenient form of Lemma 6 is the following.

Lemma 7. *Let G_p be a random graph from $\mathcal{G}(n,p)$, let A, B be fixed graphs on V and let $\mathcal{D}$ be a down-set. Then*

$$\mathbb{P}(G_p \supset B \mid \{G_p \supset A\} \cap \mathcal{D}) \le p^{e(B \setminus A)}.$$

Proof. We identify $\mathcal{G}(n,p)$ with the weighted cube $Q_{\mathbf{p}}$, where $W = [N]$, $N = \binom{n}{2}$, and all p_i are equal to p. Let $\mathcal{U}_1 = \{G_p \supset B \setminus A\}$, $\mathcal{U}_2 = \{G_p \supset A\}$, so $\mathcal{U}_1, \mathcal{U}_2$ are independent up-sets. From Lemma 6 we have

$$\begin{aligned} \mathbb{P}(G_p \supset B \mid \{G_p \supset A\} \cap \mathcal{D}) &= \mathbb{P}(G_p \supset B \setminus A \mid \{G_p \supset A\} \cap \mathcal{D}) \\ &\le \mathbb{P}(G_p \supset B \setminus A) = p^{e(B \setminus A)}, \end{aligned}$$

as required. □

In the next section we present an application of this lemma common to the cases $H = C_4$ and $H = K_4$, and in fact much more general.

5 Subgraphs containing a given edge

In this section we shall show that if H is edge-balanced, then copies of H containing a particular edge of G_p arise 'more or less independently'.

For $x, y \in V(G_p)$, let $\mathcal{H}(x,y)$ be the set of graphs S on V such that $xy \notin S$ and $S \cup \{xy\}$ is isomorphic to H. Let $U_H(G_p,x,y)$ be the union of all subgraphs S of G_p with $S \in \mathcal{H}(x,y)$, and let $X_H(G_p,x,y)$ be the number of such subgraphs S. Thus for $H = C_4$, the graph $U_H(G_p,x,y)$ is the union of all x-y paths of length three in G_p, and $X_H(G_p,x,y)$ is the number of such paths; if the edge xy is present in G_p, then $U_H(G_p,x,y)$ is the union of all C_4s in G_p containing xy, and $X_H(G_p,x,y)$ the number of such C_4s. As before we write $X_H(G_p)$ for the total number of copies of H in G_p, and N for $\binom{n}{2}$.

Lemma 8. *Let H be a fixed edge-balanced graph. Suppose that $p = p(n)$ is chosen such that*

$$\mathbb{E}(X_H(G_p)) = \lambda p N,$$

with $\lambda = \lambda(n)$ bounded as n tends to infinity. Then with probability $1 - o(n^{-2})$ we have
(i) $e(U_H(G_p,x,y)) \le \log n$ for all $x, y \in V(G_p)$, and
(ii) $X_H(G_p,x,y) \le \log n$ for all $x, y \in V(G_p)$.

Proof. Fix distinct vertices x, $y \in V$, and consider $\mathcal{H} = \mathcal{H}(x,y)$. Note that we shall never consider graphs with isolated vertices, so we may identify a graph S with the set E of its edges.

The idea of the proof is as follows. It is easy to bound the maximum number X_0 of disjoint $E \in \mathcal{H}$ present in G_p. We consider an algorithm for finding $U_H(G_p,x,y)$ that proceeds as follows. First find $\mathbf{H}_0 \subset G_p$, where $\mathbf{H}_0$ is a union of X_0 disjoint $E \in \mathcal{H}$, $E \subset G_p$. We will define a random variable $\mathbf{M}_t \subset G_p$, the set of 'marked edges', starting with $\mathbf{M}_0 = \mathbf{H}_0$. The variable $\mathbf{M}_t$ will represent the set of edges known to be present in G_p after t steps of the algorithm. At each step the algorithm considers an $E \in \mathcal{H}$ not yet considered, and tests whether $E \subset G_p$. If so, the edges of E are also marked. Thus $U_H(G_p,x,y)$ is the set of edges marked when we have considered all $E \in \mathcal{H}$. The key point is that the event that the algorithm reaches a particular state will be such that we can apply Lemma 7. This will give an upper bound on the conditional probability that $E \subset G_p$ at each stage.

Note that we expect $\mathbf{H}_0$ to be almost all of $U_H(G_p,x,y)$. The reason is that H is edge-balanced. This means that the increase in the conditional probability that $E \subset G_p$ due to E containing marked edges is outweighed by the reduction in the number of choices for such $E \in \mathcal{H}$—such E must share at least three vertices (including x and y) with the marked edges.

In what follows we often consider both a random subgraph of G_p, and possible values of this subgraph. We shall use bold type for the former, and italics for the latter. Thus $\mathbf{H}_0 \subset G_p$ will be a random variable, and H_0 will represent any possible value of this random variable. We now turn to the proof itself.

As described above we first consider disjoint sets $E \in \mathcal{H}$. For each $E \in \mathcal{H}$ the probability that $E \subset G_p$ is $p^{e(H)-1}$. Thus, counting the expectation of $e(H)X_H(G_p)$ in two different ways, we have $e(H)\lambda pN = e(H)\,\mathbb{E}\,X_H(G_p) = pN|\mathcal{H}|p^{e(H)-1}$. Writing $X_0 = X_0(G_p, x, y)$ for the maximum number of disjoint $E \in \mathcal{H}$ contained in G_p, we have

$$\begin{aligned}\mathbb{P}(X_0 \geq C) &\leq \binom{|\mathcal{H}|}{C} p^{C(e(H)-1)} \\ &\leq \left(\frac{e|\mathcal{H}|p^{e(H)-1}}{C}\right)^C = \left(\frac{e\lambda e(H)}{C}\right)^C = o(n^{-4}),\end{aligned}$$

if $C \geq \log n/2e(H)$, since then $e\lambda e(H)/C \leq e^{-9e(H)}$, for n large. We thus have

$$\mathbb{P}(X_0 \geq \log n/2e(H)) = o(n^{-4}). \tag{6}$$

In order to start the algorithm described above we need an event to condition on which is in a suitable form for Lemma 7. Let $A_1, A_2, \ldots, A_k = \emptyset$ be all edge sets that are disjoint unions of sets $E \in \mathcal{H}$. We order the A_i so that their sizes decrease, but otherwise arbitrarily. Let $\mathbf{H}_0 = \mathbf{H}_0(G_p)$ be the subgraph of G_p defined by $E(\mathbf{H}_0) = A_i$ for $i = \min\{j : A_j \subset G_p\}$. Then $E(\mathbf{H}_0)$ is the union of a largest collection of disjoint $E \in \mathcal{H}$, $E \subset G_p$, so $e(\mathbf{H}_0) = X_0(e(H)-1)$. Thus, from (6),

$$\mathbb{P}(|\mathbf{H}_0| > \log n) = o(n^{-4}). \tag{7}$$

Note that the event $\{\mathbf{H}_0 = A_i\}$ is of the form $\{A_i \subset G_p\} \cap \mathcal{D}$, where $\mathcal{D} = \bigcap_{j<i}\{A_j \not\subset G_p\}$ is a down-set. This is needed in the analysis of the algorithm outlined at the start of the proof, which we now describe precisely.

Enumerate the sets $E \in \mathcal{H}$ in an arbitrary way, so $\mathcal{H} = \{E_1, E_2, \ldots, E_h\}$. Set $\mathbf{M}_0 = \mathbf{H}_0$, $n_0 = 0$, and for $1 \leq t \leq h$ define $\mathbf{M}_t$, n_t by

$$\begin{aligned}\mathbf{M}_t &= \begin{cases} \mathbf{M}_{t-1} & \text{if } E_t \not\subset G_p \\ \mathbf{M}_{t-1} \cup E_t & \text{if } E_t \subset G_p \end{cases} \\ n_t &= \begin{cases} n_{t-1} & \text{if } \mathbf{M}_t = \mathbf{M}_{t-1} \\ n_{t-1}+1 & \text{otherwise.} \end{cases}\end{aligned}$$

Thus $n_t = n_{t-1}$ unless $E_t \subset G_p$ and $E_t \not\subset \mathbf{M}_{t-1}$. Now the event that $\mathbf{H}_0 = A_i$ and $\mathbf{M}_t = M \supset A_i$ is the event

$$\{M \subset G_p\} \cap \bigcap_{j<i}\{A_j \not\subset G_p\} \cap \bigcap_{s<t:E_s \not\subset M} \{E_s \not\subset G_p\},$$

which is of the form $\{M \subset G_p\} \cap \mathcal{D}$, where $\mathcal{D}$ is a down-set. Lemma 7 thus tells us that for any possible H_0 and M, and any $E \subset V^{(2)}$, we have

$$\mathbb{P}(E \subset G_p \mid \mathbf{H}_0 = H_0, \mathbf{M}_t = M) \leq p^{e(E \setminus M)}.$$

Considering the first t for which $n_t \geq s$ shows that the event that $\mathbf{H}_0 = H_0$ and $n_h \geq s$ is a disjoint union of events of the form

$$\mathcal{A} = \{\mathbf{H}_0 = H_0, \mathbf{M}_t = M\},$$

where $0 \leq t \leq h$, and M is a union of H_0 and s sets $E \in \mathcal{H}$, so $e(M) \leq e(H_0) + se(H)$. Given such an $\mathcal{A}$, we have $n_h \geq s+1$ if and only if there is some $E \subset G_p$ with $E \in \mathcal{H}$ and $E \not\subset M$. Any such E must meet $H_0 \subset M$, from the definition of $\mathbf{H}_0$. We thus have

$$p_{s,\mathcal{A}} = \mathbb{P}(n_h \geq s+1 \mid \mathcal{A}) \leq \sum \mathbb{P}(E \subset G_p \mid \mathcal{A}) \leq \sum p^{e(E \setminus M)},$$

where the sums are over $E \in \mathcal{H}$ with $E \not\subset M$ and $E \cap M \neq \emptyset$. We split this sum according to the number v of vertices that E shares with M, noting that $e(E \setminus M) \geq e(H) - e_H(v)$ if $v < |H|$, while in any case $e(E \setminus M) \geq 1$. This gives, being very generous,

$$p_{s,\mathcal{A}} \leq |M|^{|H|} p + \sum_{v=3}^{|H|-1} |M|^v n^{|H|-v} p^{e(H)-e_H(v)}.$$

Suppose that $|M| = n^{o(1)}$. Since $n^{|H|-2}p^{e(H)-1} = \Theta(\lambda) \leq n^{o(1)}$, and $\alpha_H(v) > \alpha_H(2)$ for $2 < v < |H|$, there is a positive ϵ such that every term in the above sum is bounded by $n^{-2\epsilon}$, say, taking n sufficiently large. Thus $p_{s,\mathcal{A}} < n^{-\epsilon}$. Since this holds for every $\mathcal{A}$ we are almost done: for every H_0 with $|H_0| = n^{o(1)}$ we have for $s = n^{o(1)}$ that

$$\mathbb{P}(n_h \geq s+1 \mid n_h \geq s, \mathbf{H}_0 = H_0) \leq n^{-\epsilon},$$

and hence that

$$\mathbb{P}(n_h \geq 5/\epsilon \mid \mathbf{H}_0 = H_0) = o(n^{-4}).$$

Now this holds for every H_0 with $|H_0| = n^{o(1)}$, so using (7) we obtain

$$\mathbb{P}(n_h \geq 5/\epsilon) = o(n^{-4}). \tag{8}$$

Recalling that $U_H(G_p, x, y)$ is the union of $\mathbf{H}_0$ and n_h sets $E \in \mathcal{H}$ we have

$$e(U_H(G_p, x, y)) \leq (e(H) - 1)(X_0 + n_h),$$

and from (6) and (8),

$$\mathbb{P}(e(U_H(G_p, x, y)) \geq \log n) = o(n^{-4}).$$

As this holds for all x and $y \in V(G_p)$, we have proved part (i) of the lemma.

For the second part we decompose $\mathbf{H}_0$ as $\mathbf{H}_1 \cup \mathbf{H}_2$, where $\mathbf{H}_1$ is the union of those $E \in \mathcal{H}$, $E \subset \mathbf{H}_0$ that share no edge with any other $E \in \mathcal{H}$, $E \subset G_p$, and $\mathbf{H}_2 = \mathbf{H}_0 \setminus \mathbf{H}_1$. Thus the sets $E \in \mathcal{H}$, $E \subset \mathbf{H}_0$ are all disjoint from each other, but those contained in $\mathbf{H}_2$ each meet some $E' \in \mathcal{H}$ with $E' \subset G_p$. Since any $E' \in \mathcal{H}$, $E' \subset G_p$ is by definition contained in $U_H(G_p, x, y)$, we have that each of the $E \in \mathcal{H}$, $E \subset \mathbf{H}_2$ shares an edge with $U_H(G_p, x, y) \setminus \mathbf{H}_0$, which consists of at most $n_h(e(H) - 1)$ edges. Since the sets E are edge disjoint, we have $e(\mathbf{H}_2) \leq n_h(e(H) - 1)^2$. Now any $E \in \mathcal{H}$, $E \subset G_p$ is either one of at most X_0 disjoint such sets in $\mathbf{H}_1 \subset \mathbf{H}_0$, or is formed from edges of $U_H(G_p, x, y) \setminus \mathbf{H}_1 = \mathbf{H}_2 \cup (U_H(G_p, x, y) \setminus \mathbf{H}_0)$. Thus,

$$X_H(G_p, x, y) \leq X_0 + \binom{n_h(e(H) - 1)^2 + n_h(e(H) - 1)}{e(H)},$$

which, with probability $1 - o(n^{-4})$, is at most X_0 plus a large constant depending on H. Together with (6) this completes the proof of the lemma. $\square$

Remarks. (i) In the particular cases $H = C_4$ and $H = K_4$, Lemma 8 can be proved much more simply. We give the proof above for two reasons: it is much more general, and it gives a simple illustration of the techniques used in the rest of the paper.

(ii) The same proof works with $\mathbb{E}\, X_H(G_p) = \lambda pN$ where $\lambda \to \infty$, as long as $\lambda < n^\epsilon$ for some $\epsilon > 0$ depending on H. Also, the probability that $e(U_H(G_p, x, y))$ exceeds its expectation by a factor C can be bounded by $2\left(\frac{e}{C}\right)^C$ for C up to n^ϵ. Thus copies of H containing xy do arise 'almost independently' in a rather strong sense.

(iii) Essentially the same proof can be applied to copies of $H \subset G_p$ containing a particular set of k vertices, with $0 \le k < |H|$. The edge-balanced condition must be replaced by $\alpha_H(v) > \alpha_H(k)$ for $|H| > v > k$. A weak form of the special case $H = K_r$ was Lemma 13 of [3]. Note that the condition on α_H is necessary, otherwise once we find a suitable K_{k+1} in G_p we find many more copies of H than expected.

6 The upper bound for C_4-free graphs

In this section we prove Theorem 3. Throughout the section we take $p = \frac{1}{2}n^{-2/3}$, $m = \lfloor n^{1/3}(\log n)^3 \rfloor$, and G_p a random graph from $\mathcal{G}(n,p)$. As before, the graph G_p' is formed from G_p by deleting any edge contained in a C_4 in G_p. Recall that we shall always assume that n is larger than some very large fixed n_0, even when this is not explicitly stated. The result we shall actually prove is the following.

Lemma 9. *With probability* $1 - o(n^{-2})$ *the graph* G_p *is such that every* C_4*-free graph* $G'' \supset G_p'$ *has maximum degree at most* $13m$.

This implies Theorem 3 since, using the coupling described in the introduction, G_∞ is a C_4-free graph containing G_p'.

The condition described in Lemma 9 is rather complicated when we express it in terms of G_p, which we need to do in order to calculate. We start by establishing some global properties of G_p that hold almost surely. Then we shall finish with the 'step by step' approach described in §4. Most of the time we shall work with G_p itself, rather than with G_p'. Thus, any graph theoretic notation we use, such as $\Gamma(x)$ for the set of neighbours of x, or $d(x)$ for the degree of x, will refer to the graph G_p unless explicitly stated otherwise.

As before, we write V for $V(G_p)$, a fixed set of n vertices. Let $\mathcal{B}_1$ be the event that some set $X \subset V$ with $100 \le k = |X| \le n^{2/5}$ spans at least $3k$ edges of G_p. Then we have

$$\begin{aligned}
\mathbb{P}(\mathcal{B}_1) &\le \sum_{k=100}^{n^{2/5}} \binom{n}{k}\binom{\binom{k}{2}}{3k} p^{3k} \\
&\le \sum_{k=100}^{n^{2/5}} \left(\frac{ne}{k}\right)^k \left(\frac{ke}{6}\right)^{3k} p^{3k} \\
&\le \sum_{k=100}^{n^{2/5}} (e^4 nk^2p^3)^k.
\end{aligned}$$

For $k \le n^{2/5}$ we have $nk^2p^3 = O(n^{1+4/5-6/3}) = O(n^{-1/5})$, so $\mathbb{P}(\mathcal{B}_1) = o(n^{-2})$.

For a set $X \subset V$ let $\Gamma_2(X)$ be the set of vertices $y \notin X$ with $|\Gamma(y) \cap X| \ge 2$, recalling that $\Gamma(y)$ is the set of neighbours of y in the graph G_p. For $X \subset V$ with $|X| = 2m$, each $y \notin X$ has probability

$p_0 \le \binom{2m}{2} p^2 \le 2m^2p^2$ of sending at least two edges to X. These events are independent, so

$$|\Gamma_2(X)| \sim \mathrm{Bi}(n-2m, p_0)$$

with mean at most $np_0 \le 2m^2p^2n = O^*(m)$. Thus by Lemma 5(b) we have

$$\mathbb{P}(|\Gamma_2(X)| > mn^{1/100}) \le e^{-mn^{1/100}}.$$

Let $\mathcal{B}_2$ be the event that some set $X \subset V$ with $|X| = 2m$ has $|\Gamma_2(X)| > mn^{1/100}$. Then

$$\mathbb{P}(\mathcal{B}_2) \le \binom{n}{2m} e^{-mn^{1/100}} \le e^{2m\log n - mn^{1/100}} = o(n^{-2}).$$

If neither $\mathcal{B}_1$ nor B_2 holds, then every set X with $|X| = 13m$ contains a set of m vertices that is rather well behaved. To formulate this precisely call a set $X \subset V$ *good* if it is an independent set (in G_p), and every $x \in X$ sends at most $n^{1/50}$ edges (of G_p) to $\Gamma_2(X)$.

Lemma 10. *Suppose that neither $\mathcal{B}_1$ nor $\mathcal{B}_2$ holds, $X_0 \subset V$ and $|X_0| = 13m$. Then X_0 contains a good set X with $|X| = m$.*

Proof. As $\mathcal{B}_1$ does not hold, every $Y \subset X_0$ with $100 \le |Y| \le n^{2/5}$ induces a graph $G_p[Y]$ with minimum degree less than 6. We may thus number the vertices of X_0 as $x_1, x_2, \ldots$ so that each of the first $13m - 100 \ge 12m$ vertices x_i sends at most 5 edges to later x_j. We can properly colour $G_p[X_0]$ as follows: colour the last m vertices arbitrarily, and proceed backwards, colouring x_i with one of the colours 1 to 6 which does not appear among its later neighbours. One of the colour classes 1 to 6 in this colouring has at least $12m/6 = 2m$ vertices, so we have found $X_1 \subset X_0$ spanning no edges of G_p, with $|X_1| = 2m$.

We claim that there is a set $X \subset X_1$ with $|X| = m$, such that each $x \in X$ sends at most $n^{1/50}$ edges to $\Gamma_2(X)$. Such an X would be a suitable good set. Suppose that the claim is false. For any $X \subset X_1$ we have $\Gamma_2(X) \subset Z = X_1 \cup \Gamma_2(X_1)$. More than m vertices of X_1 must have degree at least $n^{1/50}$ in $G_p[Z]$, as we could otherwise take m other vertices of X_1 to form X. Thus Z spans at least $\frac{1}{2}mn^{1/50}$ edges of G_p. As $\mathcal{B}_2$ does not hold, we have $|Z| \le 2m + mn^{1/100} \le 2mn^{1/100}$. But then Z spans more than $3|Z|$ edges, while $100 \le 2m \le |Z| \le n^{2/5}$, contradicting the assumption that $\mathcal{B}_1$ does not hold, and proving the claim, and hence the lemma. □

We shall define two further 'bad' events, $\mathcal{B}_3$ and $\mathcal{B}_4$. Let $\mathcal{B}_3$ be the event that for some edge $e = xy \in E(G_p)$ the graph $U_{C_4}(G_p, x, y)$ defined in §4 has more than $\log n$ edges. Then by Lemma 8 we have $\mathbb{P}(\mathcal{B}_3) = o(n^{-2})$.

Given $X \subset V$ we shall consider paths $xyz \subset G_p$ with $x, z \in X$, $y \notin X$. We shall need these paths not to share more vertices than necessary; we say that a set A of such paths is *independent* (with respect to X) if any two paths in A have distinct midpoints, and share at most one endpoint. We write $a(X)$ for the maximum number of paths in such a set A. Let X have size m. We can find a set A as above in the following way: start with $A = \emptyset$ and consider each $y \notin X$ in turn. If y is joined to exactly two vertices x, z of X, and x, z are not the endpoints of a path in A, add xyz to A and continue to the next $y \notin X$. Otherwise just continue. Having found a paths, for each vertex y considered the probability of adding a new path is exactly

$$\left(\binom{m}{2} - a\right) p^2(1-p)^{m-2}.$$

If $a < c_0 = nm^2p^2/8$ then $a \ll \binom{m}{2}$, and as $mp = o(1)$ this probability is at least $m^2p^2/3$. As we have $n-m$ vertices to consider, the probability that $a(X) < c_0$ is at most the probability that a $\mathrm{Bi}(n-m, m^2p^2/3)$ random variable is less than c_0. By Lemma 5(a) this is at most $e^{-nm^2p^2/32}$.

Let $\mathcal{B}_4$ be the event that some set $X \subset V$ with $|X| = m$ has $a(X) < c_0$. Then

$$\mathbb{P}(\mathcal{B}_4) \le \binom{n}{m} e^{-nm^2p^2/32} \le e^{m\log n - nm^2p^2/32} = o(n^{-2}),$$

as $nmp^2 \gg \log n$.

The significance of the paths counted by $a(X)$ is shown by the next definition. We say that a set $X \subset V$ is *unusable* if G'_p contains two paths $x_1y_1z_1$, $x_2y_2z_2$ with $x_i, z_i \in X$ and $y_1 \ne y_2$. We say that X is *usable* if it is not unusable. If $G'' \supset G'_p$ is C_4-free and $x \in V$, then $X = \Gamma_{G''}(x)$ must be usable—otherwise there are paths as above in $G'_p \subset G''$ with at least one y_i distinct from x. But then $xx_iy_iz_i$ is a C_4 in G''. To prove Lemma 9 it thus suffices to show that, with high probability, every set of $13m$ vertices is unusable.

Recall that a set $X \subset V$ is *good* if X spans no edges of G_p, and every $x \in X$ sends at most $n^{1/50}$ edges to $\Gamma_2(X)$. We shall show that for a fixed $X \subset V$, $|X| = m$, we have

$$\mathbb{P}(X \text{ is good and usable}, \bigcap_{i=1}^{4} \mathcal{B}_i^c) = o(n^{-m}). \tag{9}$$

The events $\mathcal{B}_i^c$, $1 \le i \le 3$, and the event that X is good are all down-sets, so we are happy to condition on these events. The event $\mathcal{B}_4^c$ is not a down-set, however, so we must treat this event differently. In particular, instead of (9) we prove the stronger statement

$$\mathbb{P}(X \text{ is usable}, \mathcal{B}_4^c \mid X \text{ is good} \cap \bigcap_{i=1}^{3} \mathcal{B}_i^c) = o(n^{-m}). \tag{10}$$

Before turning to the proof of (10) we show that (9), and hence (10), does indeed imply Lemma 9. Suppose that (9) holds, and let $\mathcal{B}_5$ be the event that there is some $X \subset V$ with $|X| = m$ which is good and usable. Then (9) implies that $\mathbb{P}(\mathcal{B}_5 \setminus (\bigcup_{i=1}^4 \mathcal{B}_i)) = o(\binom{n}{m}n^{-m}) = o(n^{-2})$. We have already shown that $\mathbb{P}(\mathcal{B}_i) = o(n^{-2})$ for $i = 1, \ldots, 4$, so we have $\mathbb{P}(\bigcup_{i=1}^5 \mathcal{B}_i) = o(n^{-2})$. Suppose now that no $\mathcal{B}_i$ holds. If X_0 is a set of $13m$ vertices, then X_0 contains a good set X with $|X| = m$ by Lemma 10. If X_0 is usable, then X is good and usable, contradicting the assumption that $\mathcal{B}_5$ does not hold. Hence, if no $\mathcal{B}_i$ holds, no set of $13m$ vertices is usable. As described above, if $G'' \supset G'_p$ is C_4-free, then the neighbourhood in G'' of any vertex x is usable. In summary, if (10) holds, so does (9). Then $\mathbb{P}(\bigcup_{i=1}^5 \mathcal{B}_i) = o(n^{-2})$, and with probability $1 - o(n^{-2})$ any C_4-free graph $G'' \supset G'_p$ has maximum degree less than $13m$. To prove Lemma 9, and hence Theorem 3, it thus suffices to prove (10).

Proof of Lemma 9. Fix $X \subset V$ with $|X| = m$. We consider all possible sets $A \subset X \times (V \setminus X)$ that are unions of independent paths as described in the definition of $a(X)$. Let $A_1, A_2, \ldots$ be these sets listed in decreasing order of size, and define a random variable $\mathbf{A}$ by $\mathbf{A} = A_i$ where $i = \min\{j : A_j \subset G_p\}$. Then $\mathbf{A}$ is a largest such set present in G_p, and by definition $\mathbf{A}$ consists of $a(X)$ independent paths. Our goal is to prove (10); since the event that X is good and $\bigcap_{i=1}^3 \mathcal{B}_i^c$ holds is a disjoint union of events $\mathcal{E}$ of the form

$$\mathcal{E} = \{\mathbf{A} = A, X \text{ is good}, \bigcap_{i=1}^{3} \mathcal{B}_i^c\},$$

it suffices to show that

$$\mathbb{P}(X \text{ is usable}, \mathcal{B}_4^c \mid \mathcal{E}) = o(n^{-m}) \tag{11}$$

separately for each $\mathcal{E}$. Note that as

$$\{\mathbf{A} = A\} = \{A \subset G_p\} \cap \bigcap_{j<i} \{A_j \not\subset G_p\}$$

when $A = A_i$, and the events $\{X \text{ is good}\}$ and $\mathcal{B}_i^c$, $i \leq 3$ are down-sets, each event $\mathcal{E}$ is of the form

$$\mathcal{E} = \{A \subset G_p\} \cap \mathcal{D}, \tag{12}$$

with $\mathcal{D}$ a down-set.

We must prove (11) for each possible event $\mathcal{E}$. Suppose first that the set A that defines $\mathcal{E}$ consists of fewer than $c_0 = nm^2p^2/8$ paths. Then $\mathcal{E}$ implies that $a(X) < c_0$ and hence that $\mathcal{B}_4$ holds, so (11) holds in this case—the conditional probability is actually zero.

From now on we suppose that A consists of at least c_0 paths, and that $\mathcal{E}$ holds. Let $P_1, \ldots, P_{c_0}$ be the first c_0 paths in A in some fixed order. We shall examine each P_t in turn, looking for a C_4 in G_p sharing an edge with P_t; whenever there is no such C_4, the path P_t is present in G_p'. As before we inductively define a set of 'marked' edges, and examine the graph in such a way that at each stage the information we have is that the marked edges are present, and some other 'negative' information. This will allow us to use Lemma 7. We now make this precise.

Suppose that the event $\mathcal{E}$ holds. Let $\mathbf{M}_0 = A$. At stage t, $1 \leq t \leq c_0$, we define $\mathbf{M}_t$ as follows.

1. If $\mathbf{M}_{t-1}$ contains a C_4 sharing an edge with P_t, set $\mathbf{M}_t = \mathbf{M}_{t-1}$; otherwise continue to step 2.
2. If the midvertex of P_t lies in $n^{1/50}$ edges of $\mathbf{M}_{t-1}$, set $\mathbf{M}_t = \mathbf{M}_{t-1}$; otherwise continue to step 3.
3. Form a list $E_{t,1}, \ldots, E_{t,\ell_t}$ of all edge sets $E \subset V^{(2)}$ that satisfy $E \cong C_4$, $E \not\subset \mathbf{M}_{t-1}$ and $E \cap P_t \neq \emptyset$. If one of these sets is present in G_p, set $\mathbf{M}_t = \mathbf{M}_{t-1} \cup E_{t,i_t}$, with $i_t = \min\{i : E_{t,i} \subset G_p\}$; otherwise continue to step 4.
4. Set $\mathbf{M}_t = \mathbf{M}_{t-1}$ and continue to stage $t+1$, noting that $P_t \subset G_p'$.

We shall show that when running this procedure we reach step 3 at least $c_0/(5\log n)$ times, and that each time we do, we have at least a 50% chance of proceeding to step 4. This will show that, almost certainly, many P_t are present in G_p'.

Let n_i, $i = 1, \ldots, 4$, be the number of times we reach step i above, so $n_1 = c_0$ and there are at least n_4 paths P_t in G_p'. Suppose that for some particular t we stop at step 1. Then there must be a four-cycle $C \subset \mathbf{M}_{t-1}$ sharing an edge with P_t. As there are no C_4s in $\mathbf{M}_0 = \mathbf{A}$ (using the independence of the paths), C shares an edge with E_{s,i_s} for some s for which we reached step 3. In other words, there is an edge in

$$P_t \cap \bigcup_{xy \in E_{s,i_s}} U_{C_4}(G_p, x, y).$$

But as we assume $\mathcal{B}_3$ does not hold, each $U_{C_4}(G_p, x, y)$ has at most $\log n$ edges. As the P_t are edge disjoint, we stop at step 1 at most $4\log n$ times for each of the n_3 possible s, so

$$n_2 \geq c_0 - 4(\log n)n_3.$$

Suppose now that for some particular t we stop at step 2. Then the midpoint y_t of P_t lies on $n^{1/50}$ edges of $\mathbf{M}_{t-1} \subset \mathbf{M}_{c_0}$. As the y_t are distinct this happens at most $2e(\mathbf{M}_{c_0})/n^{1/50}$ times. Now as $\mathcal{B}_2$ does not hold

$$e(\mathbf{M}_0) = 2a(X) \leq 2|\Gamma_2(X)| \leq 2mn^{1/100}.$$

Hence, as $e(\mathbf{M}_s \setminus \mathbf{M}_{s-1}) \leq 3$ for all s, we have $e(\mathbf{M}_{c_0}) \leq 2mn^{1/100} + 3c_0 = O^*(c_0 n^{1/100})$. Thus $n_2 - n_3 = O^*(c_0 n^{-1/100}) = o(c_0)$, and we have

$$n_3 \geq c_0 - 4(\log n)n_3 - o(c_0),$$

so

$$n_3 \geq \frac{c_0}{5\log n},$$

completing the first part of the argument.

To show that each time we reach step 3 we have a reasonable chance of continuing to step 4 we need to know a little more about the set $\mathbf{M}_t$.

For $x \in X$ let

$$\begin{aligned} S_1 &= \{xy \in E(\mathbf{M}_{c_0}) : y \in \Gamma_2(X)\}, \\ S_2 &= \{xy \in E(\mathbf{M}_{c_0}) : y \notin \Gamma_2(X)\}. \end{aligned}$$

We are assuming that X is good, so $|S_1| \leq n^{1/50}$. For $xy \in S_2$ the vertex y necessarily has only the one neighbour x in X. This means that $xy \notin \mathbf{M}_0$, so there is a unique t, $1 \leq t \leq c_0$, with $xy \in \mathbf{M}_t \setminus \mathbf{M}_{t-1}$. For this t we have $xy \in E = E_{t,i_t} \subset G_p$. Using the assumption that X spans no edges of G_p we see that E must be of the form $xyzw$ with $y, z, w \notin X$. Now E shares an edge with P_t, which must be the edge xw. Thus $|S_2|$ is at most the number of t for which P_t contains x. As X is good, this is at most $n^{1/50}$. We thus have that for every x and t

$$d_{\mathbf{M}_t}(x) \leq d_{\mathbf{M}_{c_0}}(x) \leq |S_1| + |S_2| \leq 2n^{1/50}.$$

Suppose that when considering P_t we reach step 3. We claim that the probability that some $E_{t,i}$ is in G_p is at most $1/2$, given everything we know at that stage. More precisely, let

$$p_t = p_t(A, M_1, \dots, M_{t-1}) = \mathbb{P}(\exists i : E_{t,i} \subset G_p \mid \mathcal{E}, \mathbf{M}_s = M_s \text{ for all } s < t).$$

Then for any possible sequence $M_1, \dots, M_{t-1}$,

$$p_t \leq \sum_{i=1}^{\ell_t} \mathbb{P}(E_{t,i} \subset G_p \mid \mathcal{E}, \mathbf{M}_s = M_s \text{ for all } s < t).$$

If we know $\mathbf{M}_s$ for $s < t$ then we know which sets $E_{s,i}$ were considered by the algorithm, and which ones are present in G_p. In particular we know that $\mathbf{M}_{t-1} \subset G_p$, and that certain other sets of edges are not contained in G_p. Thus from (12) the event $\mathcal{E} \cap \{\mathbf{M}_s = M_s, s < t\}$ is of the form $\{M_{t-1} \subset G_p\} \cap \mathcal{D}'$, where $\mathcal{D}'$ is a down-set. Hence, from Lemma 7,

$$p_t \leq \sum_{i=1}^{\ell_t} p^{e(E_{t,i} \setminus M_{t-1})}. \tag{13}$$

We split the sum above according to the relationship between $E_{t,i}$ and M_{t-1}, noting that by definition we have $E_{t,i} \not\subset M_{t-1}$. We write $E = E_{t,i}$ as $xyzw$ with $x \in X$ and xy an edge of P_t.

For E with $E \cap M_{t-1} = \{xy\}$ we have two choices for the edge xy of P_t, then at most n choices for each vertex z, w. As $e(E \setminus M_{t-1}) = 3$, such terms contribute at most $2n^2p^3 \leq 1/4$ to the sum in (13). As we shall see, this is actually the main term. In fact almost all the work in the proof is to show that the other terms are insignificant.

For E with $zw \in M_{t-1}$ we have two choices for the edge xy, at most $e(M_{t-1}) = O^*(n^{1/3+1/100})$ choices for zw, two ways of completing the C_4, and $e(E \setminus M_{t-1}) \geq 1$. Such terms thus contribute at most $O^*(n^{1/3+1/100}p) = o(1)$ to the sum.

For E with $E \cap M_{t-1} = \{xy, xw\}$ we have two choices for xy, at most $d_{M_{t-1}}(x) \leq 2n^{1/50}$ choices for w, and n choices for z. We have $e(E \setminus M_{t-1}) = 2$, so such terms contribute at most $4n^{1+1/50}p^2 = o(1)$.

For E with $E \cap M_{t-1} = \{xy, yz\}$, since we reached step 3 we have $d_{M_{t-1}}(y) \leq n^{1/50}$. Thus such terms contribute at most $2n^{1+1/50}p^2 = o(1)$.

Finally for E with $E \cap M_{t-1} = \{xy, yz, xw\}$ we have at most $4n^{1/25}$ choices, so such terms contribute at most $4n^{1/25}p = o(1)$ to the sum in (13).

Putting all this together, we have that if n is large enough, then $p_t \leq 1/2$. Thus for each of the $n_3 \geq c_0/(5\log n)$ times we reach step 3 we have a chance of at least $1/2$ of proceeding to step 4. Thus, given $\mathcal{E}$, the distribution of n_4 stochastically dominates a $\mathrm{Bi}(\lceil c_0/(5\log n)\rceil, 1/2)$ distribution. From Lemma 5(a) we have

$$\mathbb{P}(n_4 < 2 \mid \mathcal{E}) \leq e^{c_0/(80\log n)} = o(n^{-m}),$$

as $c_0 \gg m(\log n)^2$. As X usable implies $n_4 < 2$, we have thus proved (11). As this holds for all $\mathcal{E}$ we have proved (10), and thus (9), completing the proof of the lemma, and hence of Theorem 3. □

We now turn to the proof of Theorem 4, which is slightly simpler, but uses many of the same ideas.

7 The upper bound for K_4-free graphs

Throughout this section we consider the probability space $\mathcal{G}(n,p)$ with $p = \frac{1}{2}n^{-2/5}$, and write m for $\lfloor C(\log n)n^{3/5}\rfloor$, for some large constant C. As in the previous section we take n larger than some very large fixed n_0, even when this is not explicitly stated. We shall redefine the concepts of 'usable' and 'good', as well $\mathcal{B}_1$ to $\mathcal{B}_4$, keeping a close correspondence with the previous section whenever possible. We first prepare the ground, showing that certain 'bad' events have small probability. Then we shall prove Theorem 4 by the 'step by step' method.

Let $\mathcal{B}_0$ be the event that some set $X \subset V$ with $|X| = m$ contains $2m$ edge disjoint K_4s. For a fixed X there are $\binom{m}{4} \leq m^4$ possible K_4s, and so at most $\binom{m^4}{2m} \leq (em^3)^{2m}$ choices for the union of $2m$ edge disjoint ones. Each such union has $12m$ edges, so the probability that one is present is at most

$$(em^3)^{2m}p^{12m} = (em^3p^6)^{2m} = o(n^{-m}),$$

as $m^3p^6 = O^*(n^{-3/5}) = o(n^{-1/2})$. Thus

$$\mathbb{P}(\mathcal{B}_0) \leq \binom{n}{m}n^{-m} = o(n^{-2}).$$

By a *k-book* we mean k triangles sharing a common edge e, the *spine* of the book. Such a graph has $k+2$ vertices and $2k+1$ edges. Let $\mathcal{B}_1$ be the event that some set $X \subset V$ with $|X| = 4m$ contains m 25-books with vertex disjoint spines. Since books with vertex disjoint spines are edge disjoint we have

$$\begin{aligned}\mathbb{P}(\mathcal{B}_1) &\le \binom{n}{4m}(4m)^{27m}p^{51m}\\ &\le \left(O(n^4m^{27}p^{51})\right)^m\\ &= \Big(O(n^{-1/5})\Big)^m = o(n^{-2}).\end{aligned}$$

For a set $X \subset V$ let $\mathrm{dt}(X)$ be the maximum number of edge disjoint triangles in $G_p[X]$. For X with $|X| = 2m$ we have

$$\begin{aligned}\mathbb{P}(\mathrm{dt}(X) \ge k) &\le \binom{\binom{2m}{3}}{k}p^{3k}\\ &\le \binom{8m^3}{k}p^{3k} \le \left(\frac{8em^3p^3}{k}\right)^k.\end{aligned}$$

As $m^3p^3 = O(m(\log n)^2)$, taking $k = m(\log n)^3$ we have

$$\mathbb{P}(\mathrm{dt}(X) \ge m(\log n)^3) \le e^{-m(\log n)^3} \le n^{-2m}.$$

Let $\mathcal{B}_2$ be the event that some set $X \subset V$ with $|X| = 2m$ has $\mathrm{dt}(X) \ge m(\log n)^3$. Then

$$\mathbb{P}(\mathcal{B}_2) \le \binom{n}{2m}n^{-2m} = o(n^{-2}).$$

This time we say that a set $X \subset V$ is *good* if X contains no 25-book, and no vertex of X lies in $(\log n)^4$ triangles in X.

Lemma 11. *Suppose that $\mathcal{B}_1$, $\mathcal{B}_2$ do not hold, $X_0 \subset V$ and $|X_0| = 4m$. Then X_0 contains a good set X with $|X| = m$.*

Proof. We claim that some set $X_1 \subset X_0$ with $|X_1| = 2m$ contains no 25-book. If this is not the case, then X_0 contains a 25-book with spine $\{x_1, y_1\}$, say. As $|X_0 \setminus \{x_1, y_1\}| > 2m$, the set $X_0 \setminus \{x_1, y_1\}$ contains a 25-book with spine $\{x_2, y_2\}$ disjoint from $\{x_1, y_1\}$. Continuing in this way we find m 25-books in X_0 with disjoint spines, contradicting the assumption that $\mathcal{B}_1$ does not hold.

Fix a set $X_1 \subset X_0$ with $|X_1| = 2m$ that contains no 25-book. We are done unless m vertices of X_1 each lie in $(\log n)^4$ triangles in X_1. But then X_1 contains $m(\log n)^4/3$ triangles. As X_1 contains no 25-book, each of these triangles shares an edge with at most $3 \times 23 = 69$ others, and using the greedy algorithm we can find at least $m(\log n)^4/210 > m(\log n)^3$ edge disjoint triangles in X_1, contradicting the assumption that $\mathcal{B}_2$ does not hold, and completing the proof. $\square$

Continuing our preparation for the proof of Theorem 4, we shall define two further 'bad' events. Let $\mathcal{B}_3$ be the event that for some edge $e = xy \in E(G_p)$ the graph $U_{K_4}(G_p, x, y)$ defined in §4 has more than $\log n$ edges. Then by Lemma 8 we have $\mathbb{P}(\mathcal{B}_3) = o(n^{-2})$.

Let $\mathcal{B}_4$ be the event that some set $X \subset V$ with $|X| = m$ does not contain $\lfloor \mu/2 \rfloor$ triangles, where

$$\mu = \binom{m}{3}p^3 = \Theta(n^{3/5}(\log n)^3).$$

Fixing an m-set X and considering all possible triangles $T_i \in X^{(3)}$ we have

$$\begin{aligned}\Delta &= \sum_{e(T_i \cap T_j)>0} \mathbb{P}(T_i \cup T_j \subset G_p) \\ &= \binom{m}{2}\binom{m-2}{2}p^5 \\ &= O^*(m^4p^5) = O^*(n^{2/5}) = o(\mu).\end{aligned}$$

Thus applying (2) with $\gamma = 1/2$ we have

$$\begin{aligned}\mathbb{P}(\mathcal{B}_4) &\leq \binom{n}{m}e^{-\mu/10} \\ &= e^{O(n^{3/5}(\log n)^2)-\Theta(n^{3/5}(\log n)^3)} \\ &= o(n^{-2}).\end{aligned}$$

This time we say that a set $X \subset V$ is *unusable* if X contains a triangle in G'_p, and that X is *usable* if it is not unusable. If $G'' \supset G'_p$ is K_4-free and $x \in V$, then $X = \Gamma_{G''}(x)$ must be usable—otherwise there is a triangle in $G'_p[X] \subset G''[X]$, and hence a K_4 in G''.

Proof of Theorem 4. Fix a set $X \subset V$ with $|X| = m$. We claim that

$$\mathbb{P}(X \text{ is good and usable}, \bigcap_{i=0}^{4} \mathcal{B}_i^c) = o(n^{-m}) \tag{14}$$

implies Theorem 4. The argument is exactly analogous to that following equation (10) in the previous section. Indeed, let us write $\mathcal{B}_5$ for the event that there is some $X \subset V$ with $|X| = m$ which is good and usable. Then (14) implies that $\mathbb{P}(\mathcal{B}_5 \setminus (\bigcup_{i=0}^4 \mathcal{B}_i)) = o(\binom{n}{m}n^{-m}) = o(n^{-2})$, so $\mathbb{P}(\bigcup_{i=0}^5 \mathcal{B}_i) = o(n^{-2})$. If no $\mathcal{B}_i$ holds, then every set X_0 of $4m$ vertices is unusable, since a usable set of this size would contain a good, usable m-set by Lemma 11, contradicting the assumption that $\mathcal{B}_5$ does not hold. Thus G_∞, which is a K_4 free graph containing G'_p, has maximum degree at most $4m \leq 4C(\log n)n^{3/5}$. This proves Theorem 4 with C replaced by $2C$.

As before we wish to condition on a simple up-set $\mathcal{U}$ intersected with some down-set $\mathcal{D}$. The events $\mathcal{B}_i^c$, $0 \leq i \leq 3$ are all down-sets, but $\mathcal{B}_4^c$ is not, so we treat $\mathcal{B}_4$ differently. We shall show that

$$\mathbb{P}(X \text{ is usable}, \mathcal{B}_4^c \mid X \text{ is good}, \bigcap_{i=0}^{3} \mathcal{B}_i^c) = o(n^{-m}),$$

which implies (14).

Let $c_0 = \lfloor \mu/4 \rfloor$, where $\mu = \binom{m}{3}p^3$ as above. We consider all possible sets $A \subset X^{(2)}$ that are unions of *at most* c_0 triangles in $X^{(3)}$ and contain no K_4. Let $A_1, A_2, \ldots$ be these sets listed in decreasing order of size, and define a random variable $\mathbf{A}$ by $\mathbf{A} = A_i$ where $i = \min\{j : A_j \subset G_p\}$. The event that X is good and $\bigcap_{i=0}^3 \mathcal{B}_i^c$ holds is a disjoint union of events $\mathcal{E}$ of the form

$$\mathcal{E} = \{\mathbf{A} = A, X \text{ is good}, \bigcap_{i=0}^{3} \mathcal{B}_i^c\},$$

so it suffices to show that

$$\mathbb{P}(X \text{ is usable}, \mathcal{B}_4^c \mid \mathcal{E}) = o(n^{-m}) \tag{15}$$

separately for each $\mathcal{E}$.

As before, each event $\mathcal{E}$ is of the form

$$\mathcal{E} = \{A \subset G_p\} \cap \mathcal{D},$$

with $\mathcal{D}$ a down-set.

All that remains is to prove (15) for each possible event $\mathcal{E}$. We assume from now on that $\mathcal{E}$ holds.

Suppose first that the set A that defines $\mathcal{E}$ contains fewer than c_0 triangles. We claim that in this case $\mathcal{B}_4$ holds, proving (15). Indeed, suppose $\mathcal{B}_4$ does not hold, so X contains $2c_0$ triangles. As we are assuming that $\mathcal{B}_0$ does not hold, X does not contain $2m$ edge disjoint K_4s, and there is a set of at most $12m$ edges in X meeting every K_4 in X. As X is good, these edges together meet at most $24 \times 12m = o(c_0)$ triangles, so there is some set $A_i \subset G_p$ containing at least c_0 triangles and no K_4. Thus A must contain at least c_0 triangles, contradicting our assumption.

From now on we suppose that A contains at least c_0 triangles, and is hence the union of exactly c_0 distinct triangles $T_1, \ldots, T_{c_0}$. We shall examine each T_t in turn, looking for a K_4 in G_p sharing an edge with T_t; whenever there is no such K_4, the triangle T_t is present in G'_p.

Let $\mathbf{M}_0 = A$. At stage t, $1 \le t \le c_0$, we define $\mathbf{M}_t$ as follows.

1. If $\mathbf{M}_{t-1}$ contains a K_4 sharing an edge with T_t, set $\mathbf{M}_t = \mathbf{M}_{t-1}$; otherwise continue to step 2.
2. Form a list $E_{t,1}, \ldots, E_{t,\ell_t}$ of all edge sets $E \subset V^{(2)}$ that satisfy $E \cong K_4$, $E \not\subset \mathbf{M}_{t-1}$ and $E \cap T_t \ne \emptyset$. If one of these sets is present in G_p, set $\mathbf{M}_t = \mathbf{M}_{t-1} \cup E_{t,i_t}$, with $i_t = \min\{i : E_{t,i} \subset G_p\}$; otherwise continue to step 3.
3. Set $\mathbf{M}_t = \mathbf{M}_{t-1}$ and continue to stage $t+1$, noting that $T_t \subset G'_p$, so X is unusable.

As before let n_i, $i = 1, \ldots, 3$, be the number of times we reach step i above, so $n_1 = c_0$ and there are n_3 triangles T_t in G'_p. Suppose that for some particular t we stop at step 1. Then there must be a K_4 in $\mathbf{M}_{t-1}$ sharing an edge with T_t. As there are no K_4s in $\mathbf{M}_0 = A$ (by definition of $\mathbf{A}$), this K_4 is in the union of $U_{K_4}(G_p, x, y)$ over $xy \in E_{s,i_s}$ for all s such that we reach step 2. Since we are assuming $\mathcal{B}_3$ does not hold, this union has at most $6(\log n)n_2$ edges. As X is good, each edge is in at most 24 triangles T_t, so $n_1 - n_2 \le 150(\log n)n_2$, and

$$n_2 \ge \frac{c_0}{151 \log n}.$$

We have shown that we reach step 2 a reasonable number of times. It only remains to show that each time we have probability at least $1/2$ of continuing to step 3. Again we need to bound both the total size of $\mathbf{M}_t$ and the degrees in $\mathbf{M}_t$ of vertices in X. Now

$$e(\mathbf{M}_{c_0}) \le e(A) + 5c_0 \le 8c_0 = O^*(n^{3/5}), \tag{16}$$

using the fact that A is the union of c_0 triangles. Fix $x \in X$. As X is good, there are at most $2(\log n)^4$ edges xy contained in triangles in X. We count the number of edges $xy \in \mathbf{M}_{c_0}$ not contained in a triangle in X. Each such edge must be in one of the E_{t,i_t}. If the edge zw that E_{t,i_t} shares with T_t does not contain x, then xzw is a triangle in X—this can happen at for at most $(\log n)^4$ values of t, as X is good. On the other hand, if one of z, w is equal to x, then x is a vertex of T_t. Again this can happen for at most $(\log n)^4$ values of t. As each E_{t,i_t} contributes at most 3 to the degree of x in $\mathbf{M}_{c_0}$ we have

$$d_{\mathbf{M}_s}(x) \le d_{\mathbf{M}_{c_0}}(x) \le 2(\log n)^4 + 3(\log n)^4 + 3(\log n)^4,$$

so the degree of x in $\mathbf{M}_s$ is at most $8(\log n)^4 = n^{o(1)}$ for all s.

As in the previous section, if we reach step 2 for T_t with $\mathbf{M}_{t-1} = M_{t-1}$, the probability p_t that we find some $E_{t,i}$ in G_p satisfies

$$p_t \leq \sum_{i=1}^{\ell_t} p^{e(E_{t,i} \setminus M_{t-1})}. \tag{17}$$

We split the sum above according to the relationship between $E_{t,i}$ and M_{t-1}, noting that by definition we have $E_{t,i} \not\subset M_{t-1}$. We label the vertices of $E = E_{t,i}$ as x, y, z, w, with x, $y \in X$ and xy an edge of T_t. We say that an edge is *marked* if it is in M_{t-1}.

For E with $E \cap M_{t-1} = \{xy\}$ we have three choices for the edge xy of T_t, then at most n choices for each vertex z, w. As $e(E \setminus M_{t-1}) = 5$, such terms contribute at most $3n^2p^5 \leq 1/4$ to the sum in (17). As before this will be the main term.

For E with $zw \in M_{t-1}$ we have three choices for xy, and, from (16), at most $e(M_{t-1}) = O^*(n^{3/5})$ choices for zw. Such terms with at least two unmarked edges thus contribute $o(1)$ to the sum. With only one unmarked edge, say xz, then as yw and yz are marked and the number of marked edges from y is at most $n^{o(1)}$, there are only $n^{o(1)}$ choices for E. The single factor of p ensures that such terms contribute $o(1)$ to (17).

For E with zw not marked, suppose first that one or more of xz, yz is marked, and one or more of xw, yw. Then we again have $n^{o(1)}$ choices, so such terms contribute $o(1)$. Otherwise, one or more of xz, yz is marked, say, and none of xw, yw. We then have $n^{o(1)}$ choices for z, n choices for w, and at least 3 factors of p, from the edges xw, yw, zw. Such terms thus contribute $o(1)$ to (17).

The above case checking shows that $p_t \leq 1/2$ for n large. Thus

$$\mathbb{P}(n_3 = 0, \mathcal{B}_4^c \mid \mathcal{E}) \leq \left(\frac{1}{2}\right)^{c_0/(151 \log n)} \leq e^{-c_0/(300 \log n)}. \tag{18}$$

Now $c_0 = \lfloor \mu/4 \rfloor \sim C^3 (\log n)^3 n^{3/5}/192$. Since $n^{-m} = e^{-C(\log n)^2 n^{3/5}}$, and X usable implies $n_3 = 0$, we have that (18) implies (15) for $C = 1000$, completing the proof of Theorem 4. □

8 Conclusions

Theorems 3 and 4 have a very natural interpretation in vague terms. These results show that for $H = C_4$ and $H = K_4$ a random maximal H-free graph G_∞ has roughly the number of edges at which a random graph would be expected to contain an average of one copy of H per edge. Erdős, Suen and Winkler [5] showed that this is also the case for $H = C_3$. Thus, in these cases, the random process constructing G_∞ proceeds by adding edges more or less uniformly at random until a significant number of possible conflicts (copies of H) arise, and beyond this point not very many more edges are added. In other words, the constrained random process does not 'organize' itself significantly as the set of possible edges to add becomes smaller. This contrasts strongly with the case of bipartite graphs, where the graph G_∞ is highly 'organized'—it is a complete bipartite graph with nearly equal class sizes, as shown in [5]. It is thus natural to ask for which properties these two kinds of behaviour arise, and what happens in between.

On the one hand, it should be fairly straightforward to show that a random maximal k-colourable graph will be almost as large as it can be. On the other hand, we would expect results like Theorems 3 and 4 for H-free graphs, for many fixed graphs H. When H is edge-balanced, a case which includes complete graphs, cycles and complete bipartite graphs, the basic method used here may well be applicable. However, this application may not be easy: in the proof of Lemma 8, which applies to all edge-balanced H, the configurations $\mathbf{M}_t$ found by the algorithm had $n^{o(1)}$ edges. This meant that 'new' copies of H would almost always share at most one edge with $\mathbf{M}_t$, as H is edge-balanced. In the proofs of Theorems 3 and 4, however, configurations with n^α edges were considered, for some fixed $\alpha > 0$. This meant that separate arguments were needed in each case to deal with 'new' copies of H sharing more than one edge with $\mathbf{M}_t$, so it may not be so easy to extend these proofs to all edge-balanced H.

References

[1] B. Bollobás, *Random Graphs,* Academic Press, London 1985, xvi+447pp.

[2] B. Bollobás, *Combinatorics,* Cambridge University Press, 1986, xii+177pp.

[3] B. Bollobás and O. Riordan, Colorings generated by monotone properties, *Random Structures and Algorithms* **12** (1998), 1–25.

[4] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Statistics* **23** (1952) 493–507.

[5] P. Erdős, S. Suen and P. Winkler, On the size of a random maximal graph, *Random Structures and Algorithms* **6** (1995), 309–318.

[6] S. Janson, Poisson approximation for large deviations, *Random Structures and Algorithms* **1** (1990), 221–229.

[7] D. J. Kleitman, Families of non-disjoint subsets, *J. Combin. Theory* **1** (1966), 153–155.