Random hypergraph processes with degree restrictions*

Catherine Greenhill†
Department of Mathematics and Statistics
University of Melbourne VIC 3010
Australia
csg@ms.unimelb.edu.au

Andrzej Ruciński‡
Department of Discrete Mathematics
Adam Mickiewicz University
Poznań, Poland
rucinski@amu.edu.pl

Nicholas C. Wormald§
Department of Mathematics and Statistics
University of Melbourne VIC 3010
Australia
nick@ms.unimelb.edu.au

25 November 2002

Abstract

A d-process for s-uniform hypergraphs starts with an empty hypergraph on n vertices, and adds one s-tuple at each time step, chosen uniformly at random from those s-tuples which are not already present as a hyperedge and which consist entirely of vertices with degree less than d. We prove that for $d \geq 2$ and $s \geq 3$, with probability which tends to 1 as n tends to infinity, the final hypergraph is saturated; that is, it has $dn - s\lfloor dn/s \rfloor$ vertices of degree d and the remaining vertices (if any) are of degree $d-1$. This generalises the result for $s = 2$ obtained by the second and third authors. In addition, when $s \geq 3$, we prove asymptotic equivalence of this process and the more relaxed process, in which the chosen s-tuple may already be a hyperedge (and which therefore may form multiple hyperedges).

*Some of this work was performed while the first and second authors were visiting the Isaac Newton Institute for Mathematical Sciences, Cambridge, United Kingdom.
†Research supported by an Australian Research Council Postdoctoral Fellowship.
‡Research supported by the KBN grant 2 P03A 15 23.
§Research supported by the Australian Research Council.

1 Introduction

Let s, d and n be positive integers. Consider the random process $\mathcal{G} = \mathcal{G}(s, d, n)$ which adds s-tuples, one by one, to an originally empty s-uniform hypergraph on n vertices, with the restriction that the next hyperedge is uniformly chosen out of all unused s-tuples containing just vertices which have current degree less than d. Let us call the process an *s-graph d-process*. After at most $\lfloor dn/s \rfloor$ steps this process gets stuck at a final hypergraph G_f, which, however, does not need to be d-regular, even when s divides dn. For instance, G_f may contain a clique component of k vertices, each of degree $\binom{k-1}{s-1} \le d-1$. Nevertheless, the process always lasts for at least $\lfloor dn/s \rfloor - c$ steps, where $c = c(s, d)$. (This is because the subgraph of G_f induced by the vertices with degree less than d must either have fewer than s vertices, or form a clique of size k where $\binom{k-1}{s-1} \le d-1$, as above.)

We say that an event holds *asymptotically almost surely* (a.a.s.) if the probability that it holds tends to 1 as $n \to \infty$. An s-graph d-process is said to *saturate* if the final hypergraph has $n - i$ vertices of degree d and the remaining i vertices all have degree $d - 1$, where $i = dn - s\lfloor dn/s \rfloor$. (If dn is divisible by s then this means that the final hypergraphs is d-regular.) Ruciński and Wormald [2] proved that for $s = 2$, that is for the case of graphs, the final graph of a d-process a.a.s. is d-regular, when dn is even, or has one vertex of degree $d-1$ and the rest of degree d, when dn is odd. In other words, the process a.a.s. saturates. We show in this paper that for $s \ge 3$, an s-graph d-process a.a.s. saturates.

We will achieve our goal by studying the following *relaxed s-graph d-process* $\mathcal{G}^M = \mathcal{G}^M(s, d, n)$, which allows multiple hyperedges. Call a vertex *unsaturated* if it has degree less than d. At each step a hyperedge is chosen uniformly at random from the set of all s-tuples of unsaturated vertices, regardless of whether it has been already chosen. The relaxed process may also get stuck before saturation, but the number of unsaturated vertices in the final hypergraph must be smaller than s. We will prove that the relaxed process a.a.s. saturates and that a.a.s. it does not contain multiple hyperedges. Finally, we establish asymptotic equivalence of the original (simple) process and the relaxed process.

Unless otherwise stated, all asymptotics are as $n \to \infty$, while both s and d are constants with $s \ge 3$ and $d \ge 2$.

2 Results and structure of the argument

In what follows, we use the term *s-graph* to mean an s-uniform hypergraph of maximum degree at most d, which may have multiple hyperedges. Let us now describe the processes $\mathcal{G}(s, d, n)$ and $\mathcal{G}^M(s, d, n)$ in more detail and establish some preliminary relation between them. (We will often abbreviate $\mathcal{G}(s, d, n)$ and $\mathcal{G}^M(s, d, n)$ to $\mathcal{G}$ and $\mathcal{G}^M$.) The underlying set of both processes is the set of sequences of s-graphs on n vertices, $\pi = (G_0, \dots, G_f)$ such that

(i) $|E(G_t)| = t$ for $0 \le t \le f$,

(ii) $E(G_t) \subset E(G_{t+1})$ for $0 \le t \le f-1$,

(iii) G_f is the final s-graph in the relevant process (simple or relaxed); that is, a maximal hypergraph (or multi-hypergraph) with maximum degree at most d.

Here $E(G)$ is the *multiset* of hyperedges of the s-graph G, and the value of f lies between $\lfloor dn/s \rfloor - c$ and $\lfloor dn/s \rfloor$, for $c = c(s,d)$ constant.

Note that different sequences are produced by the two processes. The simple process $\mathcal{G}$ may produce sequences where the final s-graph has s or more unsaturated vertices, while the relaxed process $\mathcal{G}^M$ only produces sequences where the final s-graph has at most $s-1$ unsaturated vertices. This difference causes some difficulties. For instance, a sequence produced by $\mathcal{G}$ can appear as a subsequence in $\mathcal{G}^M$, where it is continued for a few more steps. Let $\mathcal{M}$ be the set of all sequences satisfying (i)–(iii) above which contain some multiple hyperedges. Partition the remaining sequences as $\mathcal{S}_{<s} \cup \mathcal{S}_{\ge s}$, where $\mathcal{S}_{<s}$ is the set of all sequences with no multiple hyperedges and at most $s-1$ unsaturated vertices in G_f, while $\mathcal{S}_{\ge s}$ is the set of all sequences with no multiple hyperedges and at least s unsaturated vertices in G_f. Both $\mathcal{G}$ and $\mathcal{G}^M$ can be considered as (non-uniform) probability spaces with the underlying set

$$\Omega = \mathcal{M} \cup \mathcal{S}_{<s} \cup \mathcal{S}_{\ge s},$$

where the support of $\mathcal{G}$ is $\mathcal{S}_{<s} \cup \mathcal{S}_{\ge s}$ and the support of $\mathcal{G}^M$ is $\mathcal{M} \cup \mathcal{S}_{<s}$. Given $\pi = (G_0, \dots, G_f) \in \Omega$, let u_t be the number of unsaturated vertices in G_t, and let w_t be the number of hyperedges in G_t which contain only unsaturated vertices. We may now complete the definition of $\mathcal{G}$ and $\mathcal{G}^M$, by assigning for each $\pi \in \mathcal{S}_{<s} \cup \mathcal{S}_{\ge s}$

$$\mathbf{P}_{\mathcal{G}}(\pi) = \prod_{t=0}^{f-1} \left(\binom{u_t}{s} - w_t \right)^{-1},$$

while for $\pi \in \mathcal{M} \cup \mathcal{S}_{<s}$,

$$\mathbf{P}_{\mathcal{G}^M}(\pi) = \prod_{t=0}^{f-1} \binom{u_t}{s}^{-1}.$$

In particular, for all $\pi \in \mathcal{S}_{<s}$ we have $\mathbf{P}_{\mathcal{G}}(\pi) \ge \mathbf{P}_{\mathcal{G}^M}(\pi)$. Except for very few cases, this inequality is strict.

For every $\pi \in \mathcal{S}_{\ge s}$ there is a nonempty set $\mathcal{M}_\pi \subset \mathcal{M}$ of sequences which are extensions of π obtained by adding at most $c = c(s,d)$ multiple hyperedges. (This corresponds to taking the final s-graph G_f from the simple process $\pi \in \mathcal{S}_{\ge s}$, and running a few more steps of the *relaxed* process on this s-graph until it terminates.) Note that the sets $\mathcal{M}_\pi$ are mutually disjoint for all $\pi \in \mathcal{S}_{\ge s}$. Call a non-empty event $\mathcal{E} \subseteq \Omega$ *persistent* if $\mathcal{E} \cap \mathcal{M}_\pi \ne \emptyset$ for all $\pi \in \mathcal{E} \cap \mathcal{S}_{\ge s}$.

Lemma 1 *For every persistent event $\mathcal{E} \subseteq \Omega$,*

$$\mathbf{P}_{\mathcal{G}}(\mathcal{E}) = O\left(\mathbf{P}_{\mathcal{G}^M}(\mathcal{E})\right),$$

where the constant hidden in $O(\cdot)$ depends on s and d only.

Proof. First, let us compare the probabilities of the elementary events in both spaces. Take $\pi = (G_0, \dots, G_f) \in \mathcal{S}_{<s} \cup \mathcal{S}_{\geq s}$. Let $r_t = nd/s - t$ be the maximum *residual* time (at time t). It is easy to check that for $t = 0, \dots, f$, we have $w_t \leq (d-1)u_t/2$ and $r_t s/d = n - st/d \leq u_t \leq sr_t$. Hence for some constants C_0, C_1, C_2 depending on s and d only, we have

$$\begin{aligned}
\mathbf{P}_{\mathcal{G}}(\pi) &\leq \prod_{t=0}^{f-1} \binom{u_t}{s}^{-1} \cdot \prod_{t=0}^{f-1} \frac{\binom{u_t}{s}}{\binom{u_t}{s} - (d-1)u_t/s} \\
&\leq \prod_{t=0}^{f-1} \binom{u_t}{s}^{-1} \cdot \exp\left(\sum_{t=0}^{f-1} C_1 \, {u_t}^{-(s-1)} \right) \\
&\leq \prod_{t=0}^{f-1} \binom{u_t}{s}^{-1} \cdot \exp\left(\sum_{r_t = \lfloor dn/s \rfloor - f + 1}^{\lfloor dn/s \rfloor} C_2 \, r_t^{-(s-1)} \right) \\
&< C_0 \prod_{t=0}^{f-1} \binom{u_t}{s}^{-1}. \qquad (1)
\end{aligned}$$

When $\pi \in \mathcal{S}_{<s}$ this simply says

$$\mathbf{P}_{\mathcal{G}}(\pi) \leq C_0 \, \mathbf{P}_{\mathcal{G}^M}(\pi).$$

It remains to consider $\pi = (G_0, \dots, G_f) \in \mathcal{S}_{\geq s}$. For any $\pi' = (G_0, \dots, G_f, \dots, G_F) \in \mathcal{M}_\pi$, we have $F - f \leq c = c(s,d)$ and $u_t \leq u_f \leq sr_f \leq sc$ for $f \leq t \leq F-1$. It follows that

$$\mathbf{P}_{\mathcal{G}^M}(\pi') = \prod_{t=0}^{F-1} \binom{u_t}{s}^{-1} \geq \binom{sc}{s}^{-c} \cdot \prod_{t=0}^{f-1} \binom{u_t}{s}^{-1} = c_0 \prod_{t=0}^{f-1} \binom{u_t}{s}^{-1}$$

where $c_0 = c_0(s,d) = \binom{sc}{s}^{-c} < 1$. Combining this with (1) gives

$$\mathbf{P}_{\mathcal{G}}(\pi) \leq C_0 \prod_{t=0}^{f-1} \binom{u_t}{s}^{-1} \leq \frac{C_0}{c_0} \mathbf{P}_{\mathcal{G}^M}(\pi').$$

Now, let $\mathcal{E} \subseteq \Omega$ be a persistent event. For every $\pi \in \mathcal{E} \cap \mathcal{S}_{\geq s}$, fix one $\pi' \in \mathcal{E} \cap \mathcal{M}_\pi$. Then, by the above, and the obvious fact that $\mathbf{P}_{\mathcal{G}}(\mathcal{M}) = 0$,

$$\mathbf{P}_{\mathcal{G}}(\mathcal{E}) \leq C_0 \sum_{\pi \in \mathcal{E} \cap \mathcal{S}_{<s}} \mathbf{P}_{\mathcal{G}^M}(\pi) + \frac{C_0}{c_0} \sum_{\pi \in \mathcal{E} \cap \mathcal{S}_{\geq s}} \mathbf{P}_{\mathcal{G}^M}(\pi') \leq C \, \mathbf{P}_{\mathcal{G}^M}(\mathcal{E}),$$

with $C = C_0/c_0$. $\square$

If $\mathcal{E}$ is persistent then the above lemma allows one to prove that $\mathbf{P}_{\mathcal{G}}(\mathcal{E}) = o(1)$ by proving that $\mathbf{P}_{\mathcal{G}^M}(\mathcal{E}) = o(1)$, which is typically easier. Unfortunately, the event $\mathcal{E}$ that the process does not saturate is not persistent. To see this, consider $\pi \in \mathcal{S}_{\geq s}$ in which G_f has exactly s unsaturated vertices, which all have degree $d-1$ and span a hyperedge. Then G_f is the final graph of the simple process which has not saturated, so $\pi \in \mathcal{E}$. Now $\mathcal{M}_\pi$ has a unique element $\pi' = (G_0, G_1, \dots, G_f, G_F)$ where $F = f+1$ and G_F is the final s-graph in the relaxed process. Note that G_F has no unsaturated vertices, so the relaxed process has saturated and $\pi' \notin \mathcal{E}$.

Instead we work with a property $\mathcal{I}_\tau$ which is defined for a given sequence $\tau = \tau(n) \to \infty$. Namely, $\pi = (G_0, \dots, G_f) \in \Omega$ satisfies $\mathcal{I}_\tau$ if $\tau > c$ and the unsaturated vertices in the s-graph $G_{\lfloor dn/s \rfloor - \tau}$ are all of degree $d-1$ and form an independent set. This property trivially (and deterministically) implies saturation. Moreover, for n sufficiently large both the event $\mathcal{I}_\tau$ and its negation $\neg\mathcal{I}_\tau$ are persistent. (In fact something stronger is true: if $\pi \in \mathcal{I}_\tau \cap \mathcal{S}_{\geq s}$ then $\mathcal{M}_\pi \subseteq \mathcal{I}_\tau$, and similarly for $\neg\mathcal{I}_\tau$.) To show that $\mathcal{I}_\tau$ holds for $\mathcal{G}^M$ a.a.s., we will first focus on the event that the process achieves minimum degree $d-1$ at least some $\omega(n) \to \infty$ steps from the end, deferring the question of independence for later. For $j = 0, \dots, d-2$, let T_j (respectively, T_j^M) be the latest time t when the minimum degree of G_t in $\mathcal{G}$ (respectively, $\mathcal{G}^M$) is j. The following lemma lies at the heart of the whole argument.

Lemma 2 *For $s \geq 3$ and $d \geq 2$, there is a sequence $\omega(n) \to \infty$ such that a.a.s.*

$$T_{d-2}^M \leq \left\lfloor \frac{dn}{s} \right\rfloor - \omega(n).$$

Using an adaptation of a simple lemma from [2] we will be able to conclude the following. (The proofs of Lemma 2 and Corollary 1 are given later in this section.)

Corollary 1 *There is a sequence $\tau = \tau(n) \to \infty$ such that a.a.s. $\mathcal{G}^M \in \mathcal{I}_\tau$.*

It is now easy to prove the main result of this paper.

Theorem 1 *For $s \geq 3$ and $d \geq 2$, the s-graph d-process $\mathcal{G}(s,d,n)$ and the relaxed s-graph d-process $\mathcal{G}^M(s,d,n)$ both a.a.s. saturate.*

Proof. The fact that $\mathcal{G}^M(s,d,n)$ a.a.s. saturates follows immediately from Corollary 1. Next, recall that the event $\neg\mathcal{I}_\tau$ is persistent for n sufficiently large. Applying Lemma 1 to $\neg\mathcal{I}_\tau$ and using Corollary 1 we obtain the corresponding statement for $\mathcal{G}$; namely, there is a sequence $\tau = \tau(n) \to \infty$ such that a.a.s $\mathcal{G} \in \mathcal{I}_\tau$. This immediately implies that $\mathcal{G}(s,d,n)$ a.a.s. saturates. ∎

We now reveal our strategy of proof of Lemma 2. We will show that for some $\alpha = \alpha(n) \to \infty$, a.a.s. $T_0^M < \lfloor dn/s \rfloor - \alpha$. In that case, at time $T_0^M + 1$ all vertices have degree at least 1 and there is plenty of time to go. To look at the process in a slightly different way, define the *degree deficit* of a vertex i at time t to be $d - \deg_{G_t}(i)$. Then $T_0^M + 1$ is the first time at which the maximum degree deficit drops from d to $d-1$. This calls for induction on d, except that the initial situation now is different in that we have vertices of varying degrees (or degree deficits). In order to handle it, we introduce a *relaxed generalised s-graph d-process* on the vertex set $[n]$ as follows. Each vertex i is assigned an initial degree deficit m_i, where $1 \le m_i \le d$, with $d = \max_i m_i$. (Note that each initial deficit is *positive*.) The process produces a sequence of s-graphs $(G_0, G_1, \dots, G_f)$ on the vertex set $[n]$, starting with the empty s-graph G_0 (with no hyperedges.) A vertex i of G_t is *unsaturated* if $\deg_{G_t}(i) < m_i$. Let $\mathcal{U}_t$ be the set of unsaturated vertices in G_t. At time $t+1$, the s-graph G_{t+1} is formed from G_t by adding a new hyperedge, chosen uniformly at random from the set of all subsets of $\mathcal{U}_t$ of size s. Note that the sequences $(G_0, G_1, \dots, G_f)$ produced by a relaxed generalised s-graph d-process also satisfy conditions (i)–(iii) given at the start of this section, under a new definition of 'final' in (iii).

Let $m = \sum_{i=1}^{n} m_i$ be the total initial degree deficit. Then the generalised d-process runs for at most $\lfloor m/s \rfloor$ steps. Note that

$$n \le m \le dn$$

and that the total degree deficit at time t is $m - st$. A vertex i with $\deg_{G_t}(i) = 0$ is called an *isolate* (at time t). Define $U_t = |\mathcal{U}_t|$ to be a random variable counting the unsaturated vertices at time t, and let I_t be the number of isolates at time t with (maximum) deficit d. Clearly

$$\frac{m - st}{d} \le U_t \le m - st - (d-1)I_t. \tag{2}$$

We now state three further lemmas which will imply Lemma 2. The first is just [2, Lemma 3.1] adapted to hypergraphs. In fact, the situation here is easier, since we have no "forbidden hyperedges". We omit the proof.

Lemma 3 *For $s \ge 2$ and $d \ge 1$, consider a relaxed generalised s-graph d-process, with total initial deficit m. Fix j with $1 \le j \le s$. For $0 \le u < v \le \lfloor m/s \rfloor$, let $\mathbf{P}(j,u,v)$ be the conditional probability that the vertices $1, \dots, j$ remain isolated in G_v, given they were isolated in G_u. There exists an absolute constant $C > 0$ such that*

$$\mathbf{P}(j,u,v) \le C \left(\frac{\lfloor m/s \rfloor - v + 1}{\lfloor m/s \rfloor - u} \right)^j$$

for all $0 \le u < v \le \lfloor m/s \rfloor$.

The following lemma is proved in the next section using the differential equations method from [4].

Lemma 4 *For $s \ge 2$ and $d \ge 2$ consider a relaxed generalised s-graph d-process on n vertices, with total initial deficit m. Let I_t be the number of isolates of deficit d at time t. There exists a function $w(x)$ satisfying*

$$\lim_{x \to 1/s-} \frac{\log(1/s - x) w(x)}{(1 - sx)} = -\frac{1}{d-1}, \tag{3}$$

such that for all $\varepsilon > 0$, a.a.s.

$$I_t \leq m\, w(t/m) + o(m)$$

for $0 \leq t \leq m/s - \varepsilon m$. (Here 'a.a.s.' is with respect to m and uniform over all initial deficit assignments $m_1 + \cdots + m_n = m$.)

Let T_j^{RG} be the last time at which the maximum degree deficit (in the relaxed generalised s-graph d-process) is $d - j$, for $0 \leq j \leq d - 2$. (For the initial deficit assignment $m_i \equiv d$, this is the same as previously defined T_j^M.)

Lemma 5 *For $s \geq 2$ and $d \geq 2$ consider a generalised s-graph d-process on n vertices, with total initial deficit m. Then, for some $\alpha(n) \to \infty$ a.a.s. there are no isolates of deficit d at time $\lfloor m/s \rfloor - \alpha(n)$. Equivalently, $T_0^{RG} < \lfloor m/s \rfloor - \alpha(n)$ (or $I_{\lfloor m/s \rfloor - \alpha(n)} = 0$) a.a.s..*

Proof. The result will follow if we show that for each integer $\alpha \geq 2$,

$$\lim_{m \to \infty} \mathbf{P}(I_{\lfloor m/s \rfloor - \alpha} \geq 1) = 0. \tag{4}$$

Indeed, from this it is a simple exercise to show that there exists a sequence $\alpha(m) \to \infty$ for which (4) holds too. (Note that $m = m(n) \to \infty$ if and only if $n \to \infty$.)

Let $v_\alpha = m/s - \alpha$. By Lemma 3, for all $u < v_\alpha$

$$\begin{aligned} \mathbf{P}(I_{v_\alpha} \geq 1) &\leq \mathbf{E}I_{v_\alpha} = \mathbf{E}(\mathbf{E}(I_{v_\alpha} | I_u)) \\ &\leq \mathbf{E}I_u \cdot C\left(\frac{\alpha + 1}{m/s - u}\right) \end{aligned} \tag{5}$$

By Lemma 4, for all $\varepsilon > 0$, setting $u_\varepsilon = m/s - \varepsilon m$, a.a.s.

$$I_{u_\varepsilon} \leq mw(1/s - \varepsilon) + o(m).$$

Since always $I_u \leq m$, this yields that

$$\mathbf{E}I_{u_\varepsilon} \leq mw(1/s - \varepsilon) + o(m).$$

To prove (4), for each $\delta > 0$ choose ε so small that

$$w(1/s - \varepsilon) < \frac{\varepsilon\delta}{2C\alpha}.$$

By (3), such a choice is always possible, since $\lim_{\varepsilon \to 0} \log(1/\varepsilon) = \infty$. Now choose sufficiently large m_0 so that for all $m \geq m_0$ we have $u_\varepsilon < v_\alpha$, and hence

$$\mathbf{E}I_{u_\varepsilon} < \frac{\varepsilon\delta m}{C(\alpha + 1)}.$$

Consequently, by (5) with $u = u_\varepsilon$, $\mathbf{P}(I_{v_\alpha} \geq 1) < \delta$. ∎

With Lemma 5 at hand, we can quickly prove Lemma 2.

Proof of Lemma 2. We prove the statement in the lemma for relaxed generalized processes by induction on d. For $d = 2$ it coincides with the statement of Lemma 5. For $d \geq 3$, we first apply Lemma 5, and then pretend that we start a new (relaxed generalised) s-graph $(d-1)$-process at time $T_0 + 1$. More specifically, the new process has possibly a smaller number of vertices $\widehat{n}$, and smaller deficiencies $\widehat{m}_i$. Indeed, we ignore any saturated vertices at time $T_0 + 1$, namely those i with $\deg_{G_{T_0+1}}(i) = m_i$, and reset the value of m_i by subtracting $\deg_{G_{T_0+1}}(i)$, for all remaining vertices i. Hence the new total deficit $\widehat{m} = \Theta(\widehat{n})$. The new maximum length of the process $\lfloor \widehat{m}/s \rfloor \geq \alpha(n)$ tends to infinity as n tends to infinity. But $\widehat{n} \geq s\lfloor \widehat{m}/s \rfloor / d$, and so $\widehat{n}$ and $\widehat{m}$ also tend to infinity with n. We apply the inductive hypothesis to the new process, replacing d with $d - 1$. ∎

We conclude this section with a proof of Corollary 1.

Proof of Corollary 1. Pick up the process at time $T_{d-2} + 1$. By Lemma 2 we know that a.a.s there are at least $\omega(n)$ steps to go, and as in the proof of Lemma 2, we may view the remaining steps as a new, relaxed generalised s-graph 1-process. At time $\lfloor dn/s \rfloor - \omega$, in the underlying s-graph of the original process there are $O(\omega)$ hyperedges spanned by the unsaturated vertices. Apply Lemma 3 with $j = s$, $u = \lfloor dn/s \rfloor - \omega$ and $v = \lfloor dn/s \rfloor - \tau$ for $\tau = \omega^{1/2}$, to each of these hyperedges in turn, to conclude that the expected number of those of them which at time $\lfloor dn/s \rfloor - \tau$ still contain unsaturated vertices is

$$O\left(\omega \left(\frac{\tau}{\omega}\right)^s\right) = o(1).$$

Then a.a.s. the set of unsaturated vertices at remaining time τ is independent. ∎

3 Analysis with differential equations

In this section we prove Lemma 4 by applying the following result on approximation by differential equations [5, Theorem 3], which is a simplified version of [4, Theorem 6.1]. First some definitions. Given a variable m, let $(X_0, X_1, \ldots, X_{f_m})$ be a discrete-time Markov chain with the state space $\Lambda = \Lambda_m$. We are interested in asymptotics as $m \to \infty$.

Assume there are a random variables $Y_1, \ldots, Y_a$ of interest, where a is a constant, given by $Y_i(t) = y_i(X_t)$ where y_i is a deterministic function $y_i : \Lambda \to \mathbb{R}$ for $1 \leq i \leq a$. Note that $(Y_1(t), \ldots, Y_a(t))$, $t = 1, \ldots, f_n$, forms a random process whose scaled trajectory $(t/m, Y_1(t)/m, \ldots, Y_a(t)/m)$ is to be approximated.

For any domain $\mathcal{D} \subseteq \mathbb{R}^{a+1}$ define the *stopping time* $T_{\mathcal{D}}(Y_1, \ldots, Y_a)$ to be the minimum t such that $(t/m, Y_1(t)/m, \ldots, Y_a(t)/m) \notin \mathcal{D}$. This will be written as $T_{\mathcal{D}}$ for short. We say something is true *always* if it holds with probability 1, which, for discrete probability spaces, means that it holds for all elements of the underlying space which have non-zero probability. By $f : \mathbb{R}^s \to \mathbb{R}$ being *Lipschitz* on a set $\mathcal{D} \subseteq \mathbb{R}^s$, we mean that for some $C > 0$, for all $\varepsilon' > 0$, $|f(\mathbf{x}) - f(\mathbf{x}_0)| < C\varepsilon'$ whenever $||\mathbf{x} - \mathbf{x}_0|| < \varepsilon'$ with $\mathbf{x}$, $\mathbf{x}_0 \in \mathcal{D}$. (In particular we can take $|| \cdot ||$ to be the ℓ_∞-norm on $\mathbb{R}^s$.)

Theorem 2 *With notation as above, assume that $\mathcal{D} \subseteq \mathbb{R}^{a+1}$ is closed and bounded and contains the set*

$$\{(0, z_1, \ldots , z_a) : \mathbf{P}(Y_i(0) = z_i m, 1 \leq i \leq a) \neq 0 \textit{ for some } m\}$$

of all possible initial points of the scaled process. Furthermore, assume that for $1 \leq i \leq a$,

(i) for some constant β

$$\max_{1\leq i\leq a} |Y_i(t + 1) - Y_i(t)| \leq \beta$$

always for $t < T_{\mathcal{D}}$,

(ii) for some functions $f_i : \mathbb{R}^{a+1} \to \mathbb{R}$ which are Lipschitz on an open set containing $\mathcal{D}$,

$$\mathbf{E}(Y_i(t + 1) - Y_i(t) \mid X_t) = f_i(t/m, Y_1(t)/m, \ldots , Y_a(t)/m) + o(1)$$

always for $t < T_{\mathcal{D}}$.

Then the following are true.

(a) For all $(0, \hat{z}_1, \ldots , \hat{z}_a) \in \mathcal{D}$ the system of differential equations

$$\frac{dz_i}{dx} = f_i(x, z_1, \ldots , z_a), \qquad i = 1, \ldots , a$$

has a unique solution in $\mathcal{D}$ for $z_i : \mathbb{R} \to \mathbb{R}$ passing through $(0, \hat{z}_1, \ldots , \hat{z}_a)$, which extends for positive x past some point at the boundary of $\mathcal{D}$;

(b) Asymptotically almost surely,

$$Y_i(t) = mz_i(t/m) + o(m) \tag{6}$$

uniformly for $0 \leq t \leq \min\{\sigma m, T_{\mathcal{D}}\}$ and for each $i = 1, \ldots , a$, where $z_i(x)$ are as in (a) with $\hat{z}_i = Y_i(0)/m$, and σ denotes the least x-coordinate of the solution in (a) on the boundary of $\mathcal{D}$.

In part (b) of this theorem, "uniformly" refers to the convergence implicit in the $o(m)$ term of (6). If $\partial\mathcal{D}$ denotes the boundary of $\mathcal{D}$, then σ from part (b) is defined by

$$\sigma = \min \{x > 0 \mid (x, z_1(x), \ldots , z_d(x)) \in \partial\mathcal{D}\} .$$

Now we define the random variables to which we will apply this theorem. Take $X_t = G_t$ for $t \geq 0$, where $(G_0, G_1, ...)$ is the relaxed, generalised s-graph d-process on n

vertices, and let $Y_j^{(t)}$ be the number of vertices with degree deficit j in G_t for $1 \le j \le d$. Then $I_t = Y_d^{(t)}$, and $U_t = Y_1^{(t)} + \cdots + Y_d^{(t)}$. Since all initial deficits are positive, the total initial deficit $m = \sum_{j=1}^{d} jY_j^{(0)} = \Theta(n)$. We will do asymptotics with respect to m.

It is not difficult to show that

$$\mathbf{E}(Y_j^{(t+1)} - Y_j^{(t)} \mid G_t) = \begin{cases} s\left(Y_{j+1}^{(t)} - Y_j^{(t)}\right) / \left(Y_1^{(t)} + \cdots Y_d^{(t)}\right) & \text{for } 1 \le j < d, \\ -sY_d^{(t)} / \left(Y_1^{(t)} + \cdots Y_d^{(t)}\right) & \text{for } j = d. \end{cases}$$

Now we switch to continuous variables. The variable z_j will a.a.s. approximate $Y_j^{(t)}/m$. For $1 \le j \le d$, let

$$f_j(z_1, \ldots, z_d) = \begin{cases} s(z_{j+1} - z_j)/(z_1 + \cdots + z_d) & \text{for } 1 \le j < d, \\ -sz_d/(z_1 + \cdots + z_d) & \text{for } j = d. \end{cases}$$

Then

$$\mathbf{E}(Y_j^{(t+1)} - Y_j^{(t)} \mid G_t) = f_j(Y_1^{(t)}/m, \ldots, Y_d^{(t)}/m) \tag{7}$$

for $1 \le j \le d$.

Fix a constant $\varepsilon > 0$. We will define the domain $\mathcal{D}$ so that the solution exits at a convenient point arbitrarily close to the end of the process, while avoiding the singularities of the f_j. Specifically, let

$$\begin{aligned} \mathcal{D} \;=\; & \{(x, z_1, \ldots, z_d) \mid -\varepsilon \le x \le 1/s - \varepsilon,\ -\varepsilon \le z_j \le 1 + \varepsilon \ \text{ for } \ 1 \le j \le d, \\ & \quad z_1 + \cdots + z_d \ge s\varepsilon/2d\}\,. \end{aligned}$$

We check the conditions of Theorem 2. Clearly $\mathcal{D}$ is closed, bounded and independent of m. Note also that $\mathcal{D}$ contains the set of all (scaled) possible initial points of the process, namely the set

$$\left\{(0, \hat{z}_1, \ldots, \hat{z}_d) \mid \mathbf{P}(Y_j^{(0)} = \hat{z}_j\, m \ \text{ for } \ 1 \le j \le d) > 0 \ \text{ for some } m\right\}.$$

Furthermore, we have $|Y_j^{(t+1)} - Y_j^{(t)}| \le s$ always. So take $\beta = s$ for part (i). For $1 \le j \le d$, it is easy to see that f_j is Lipschitz on some open set $\mathcal{D}_0$ containing $\mathcal{D}$ (for instance, choose $\mathcal{D}_0$ open and bounded such that the closure of $\mathcal{D}_0$ does not contain any singularities of f_j). Hence (ii) follows from (7). Thus the conditions of the theorem are satisfied. We conclude that a.a.s.

$$Y_j^{(t)} = mz_j(t/m) + o(m) \tag{8}$$

uniformly for $0 \le t \le \min\{\sigma m, T_{\mathcal{D}}\}$, where for $0 \le j \le d$, the functions $z_j(x)$ form the unique solution of the system of differential equations

$$\frac{dz_j}{dx} = f_j(z_1, \ldots, z_d) \quad \text{for } \ 1 \le j \le d$$

on the domain $\mathcal{D}$, with $z_j(0) = Y_j^{(0)}/m$.

Now we determine the values of $T_{\mathcal{D}}$ and σ. By (2) we see that $U_t \geq (m - st)/d$. So $U_t \geq s\varepsilon m/d$ whenever $0 \leq t \leq m/s - \varepsilon m$. Clearly $0 \leq Y_j^{(t)} \leq m$ for all j, t. Therefore the scaled trajectory $(t/m, Y_1(t)/m, \ldots, Y_d(t)/m)$ cannot approach any boundary of $\mathcal{D}$ other than $x = 1/s - \varepsilon$. This proves that $T_{\mathcal{D}} = \lceil m(1/s - \varepsilon) \rceil$. So $\min\{\sigma m, T_{\mathcal{D}}\} = \sigma m$. Next we show that $\sigma = 1/s - \varepsilon$.

By (8), within $\mathcal{D}$, inequalities satisfied by the components of the scaled trajectory $(t/m, Y_1^{(t)}(t/m), \ldots, Y_d^{(t)}(t/m))$ are also satisfied (within additive error $o(1)$) by the corresponding components of the differential equation solution $F = (x, z_1, \ldots, z_d)$ for $t \leq \sigma m$. Thus, since the functions f_i are Lipschitz, no boundary of $\mathcal{D}$ other than $x = 1/s - \varepsilon$ can be approached, to within distance $o(1)$, by F (for m large enough). Hence the solution of the differential equation may not reach any boundary of $\mathcal{D}$ other than $x = 1/s - \varepsilon$, establishing that $\sigma = 1/s - \varepsilon$.

For the remainder of the proof we focus on $I_t = Y_d^{(t)}$, the number of isolates with deficit d at time t. By (8), I_t a.a.s. satisfies

$$I_t \leq m z_d(t/m) + o(m)$$

for $0 \leq t \leq m/s - \varepsilon m$. But

$$\frac{dz_d}{dx} = -\frac{s z_d}{z_1 + \cdots + z_d} \leq -\frac{s z_d}{1 - sx - (d-1) z_d}$$

for $0 \leq x \leq 1/s - \varepsilon$, using (8) and the right hand inequality of (2). For if dz_d/dx were greater than this right hand side, it would be greater by at least some constant amount, contradicting (8) for m large enough.

Let $w(x)$ be the function which agrees with $z_d(0)$ at $x = 0$ and which has derivative suggested by this upper bound. That is, $w(x)$ is the solution of

$$\frac{dw}{dx} = -\frac{sw}{1 - sx - (d-1)w}$$

with $w(0) = I_0/m$. Then $z_d(x) \leq w(x)$ for $0 \leq x \leq 1/s - \varepsilon$, and so a.a.s.

$$I_t \leq m w(t/n) + o(m) \tag{9}$$

for $0 \leq t \leq m/s - \varepsilon m$. Simple calculations (as in [2, 3]) show that $w(x)$ satisfies

$$\frac{\log(1/s - x) w(x)}{(1 - sx)} \to -\frac{1}{d-1}$$

as x tends to $1/s$ from below. This completes the proof of Lemma 4. $\square$

4 Asymptotic equivalence of the simple and relaxed processes

Informally, two (sequences of) probability spaces are said to be *asymptotically equivalent* if the probability of any event in one space differs by at most $o(1)$ from the probability of that event in the other space. (By 'event' we really mean a sequence of events indexed by n.) In this section we prove that the two processes $\mathcal{G}$ and $\mathcal{G}^M$ are asymptotically equivalent. (In particular, this implies that $\mathcal{G}$ and $\mathcal{G}^M$ are *contiguous*; see, e.g., [1].)

First we show that a.a.s. the relaxed process does not create multiple hyperedges. The calculations are similar to those in the proof of Lemma 1. Note that Lemma 6 does not hold for $s = 2$, that is, for graph d-processes.

Lemma 6 *For $s \geq 3$ and $d \geq 2$, the relaxed s-graph d-process a.a.s. creates no multiple hyperedges.*

Proof. By Corollary 1 we know that at time $\lfloor dn/s \rfloor - \tau(n)$, for some $\tau(n) \to \infty$, the unsaturated vertices all have degree $d-1$ and form an independent set. So, from that point onward there is no danger of creating multiple hyperedges. To complete the proof, it remains to show that a.a.s. no repeated hyperedges are formed in the first $\lfloor dn/s \rfloor - \tau(n)$ steps. As soon as a hyperedge is formed in the process, call that hyperedge *forbidden* for the remainder of the process. The maximum number of forbidden hyperedges at time t is $(d-1)U_t/s$. Note also that $U_t \geq r_t s/d$, where $r_t = nd/s - t$ is the remaining time. Therefore, the probability that a forbidden hyperedge is selected at time t is at most

$$\frac{(d-1)U_t}{s\binom{U_t}{s}} = O({U_t}^{-(s-1)}) = O({r_t}^{-(s-1)}).$$

Thus, setting $r = r_t$, the probability that a multiple hyperedge is created in the first $dn/s - \tau(n)$ steps is at most

$$O\left(\sum_{r=\tau(n)}^{\infty} r^{-(s-1)}\right) = O\left(\int_{r=\tau(n)}^{\infty} r^{-2} dr\right) = O(\tau(n)^{-1}) = o(1).$$

□

By Lemma 6, together with Lemma 1 and Theorem 1, the contiguity of $\mathcal{G}$ and $\mathcal{G}^M$ follows easily. We will now establish the asymptotic equivalence of the two processes, which is a much stronger statement.

Theorem 3 *For $s \geq 3$ and $d \geq 2$, the processes $\mathcal{G}$ and $\mathcal{G}^M$ are asymptotically equivalent in the sense that for every event $\mathcal{E} \subseteq \Omega$ we have*

$$\mathbf{P}_{\mathcal{G}}(\mathcal{E}) - \mathbf{P}_{\mathcal{G}^M}(\mathcal{E}) = o(1).$$

Proof. Define an auxilliary probability space $\mathcal{G}^*$ acting on the (infinite) set Ω^* of sequences of s-tuples which result from the following random experiment. Start with an empty sequence and an empty s-graph on n vertices. At each time step, choose an s-tuple of unsaturated vertices as in the relaxed process, i.e. with no concern about repetitions. Add this s-tuple to the sequence, but only add it to the s-graph if it is not already present. Continue in this manner so long as the evolving s-graph has an s-tuple of unsaturated vertices which is not a hyperedge (this may be true forever). If no such s-tuple is present, then stop provided that there are fewer than s unsaturated vertices, otherwise move to the second phase. In the second phase continue to randomly choose available s-tuples, but now add them to *both* the sequence and the s-graph, despite the fact that they are certainly already present as hyperedges in the s-graph. Stop the second phase when fewer than s unsaturated vertices are present in the s-graph.

Let us emphasize that the set Ω^* consists of sequences of all chosen s-tuples, but we associate each one with an s-graph obtained by crossing out all repeated s-tuples except those chosen in the final phase. The s-graph determines when the experiment stops. Note that the same s-graph can be obtained from (infinitely) many sequences in Ω^*.

The sequences in Ω^* split into two disjoint sets: $\Omega^* = \mathcal{M}^* \cup \mathcal{S}^*_{<s}$, where $\mathcal{M}^*$ consists of those with multiple s-tuples, and $\mathcal{S}^*_{<s}$ is a copy of $\mathcal{S}_{<s}$. For each $\pi \in \mathcal{S}_{<s}$ we have

$$\mathbf{P}_{\mathcal{G}^*}(\pi) = \mathbf{P}_{\mathcal{G}^M}(\pi). \tag{10}$$

Therefore, by Lemma 6,

$$\mathbf{P}_{\mathcal{G}^*}(\mathcal{M}^*) = 1 - \mathbf{P}_{\mathcal{G}^*}(\mathcal{S}^*_{<s}) = 1 - \mathbf{P}_{\mathcal{G}^M}(\mathcal{S}_{<s}) = \mathbf{P}_{\mathcal{G}^M}(\mathcal{M}) = o(1).$$

If $\pi' \in \Omega^*$, let $\rho(\pi')$ be the sequence obtained from π' by crossing out all but the first occurrence of any s-tuple in π'. By definition of $\mathcal{G}^*$, we have $\mathbf{P}_{\mathcal{G}}(\pi) = \mathbf{P}_{\mathcal{G}^*}(\rho^{-1}(\pi))$ for all $\pi \in \mathcal{S}_{<s}$. (Note that the sets $\rho^{-1}(\pi)$ are infinite.) Now, let $\mathcal{E} \subseteq \Omega$ be an arbitrary event. Without loss of generality we may assume that $\mathcal{E} \subseteq \mathcal{S}_{<s}$ (since the events $\mathcal{S}_{\geq s}$ and $\mathcal{M}$ a.a.s. do not hold in either $\mathcal{G}$ or $\mathcal{G}^M$). Define an auxilliary event $\mathcal{E}^* \subseteq \Omega^*$ by

$$\mathcal{E}^* = \{\pi' \in \Omega^* \mid \rho(\pi') \in \mathcal{E}\} = \bigcup_{\pi \in \mathcal{E}} \rho^{-1}(\pi),$$

where the union is disjoint. Notice that sequences in $\mathcal{E}^*$ never entered the final phase of the experiment. Now

$$\mathbf{P}_{\mathcal{G}}(\mathcal{E}) = \sum_{\pi \in \mathcal{E}} \mathbf{P}_{\mathcal{G}}(\pi) = \sum_{\pi \in \mathcal{E}} \mathbf{P}_{\mathcal{G}^*}(\rho^{-1}(\pi)) = \mathbf{P}_{\mathcal{G}^*}(\mathcal{E}^*)$$

by the definition of $\mathcal{E}^*$. Moreover, using (10), $\mathbf{P}_{\mathcal{G}^*}(\mathcal{E}^* \cap \mathcal{S}^*_{<s}) = \mathbf{P}_{\mathcal{G}^M}(\mathcal{E})$. Hence,

$$\mathbf{P}_{\mathcal{G}}(\mathcal{E}) = \mathbf{P}_{\mathcal{G}^*}(\mathcal{E}^*) = \mathbf{P}_{\mathcal{G}^*}(\mathcal{E}^* \cap \mathcal{S}^*_{<s}) + \mathbf{P}_{\mathcal{G}^*}(\mathcal{E}^* \cap \mathcal{M}^*) = \mathbf{P}_{\mathcal{G}^M}(\mathcal{E}) + o(1).$$

□

References

[1] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, John Wiley and Sons, New York (2000).

[2] A. Ruciński and N.C. Wormald, Random graph processes with degree restrictions, *Combinatorics, Probability and Computing* **1** (1992), 169–180.

[3] A. Ruciński and N.C. Wormald, Random graph processes with maximum degree two, *The Annals Applied Probab.* **7** (1997), 183–199.

[4] N. C. Wormald, The differential equations method for random graph processes and greedy algorithms, in *Lectures on Approximation and Randomized algorithms* (M. Karoński and H. J. Prömel, eds.), PWN, Warsaw, 1999, pp. 73–155.

[5] N.C. Wormald, Analysis of greedy algorithms on graphs with bounded degrees, *Discrete Mathematics* (to appear).