On the intersecting family process

Patrick Bennett* Alan Frieze† Andrew Newman‡ Wesley Pegden§

August 12, 2024

Abstract

We study the intersecting family process initially studied in [6]. Here $k = k(n)$ and $E_1, E_2, \ldots, E_m$ is a random sequence of k-sets from $\binom{[n]}{k}$ where E_{r+1} is uniformly chosen from those k-sets that are not already chosen and that meet $E_i, i = 1, 2, \ldots, r$. We prove some new results for the case where $k = cn^{1/3}$ and for the case where $k \gg n^{1/2}$.

1 Introduction

We study the following process introduced by Bohman, Cooper, Frieze, Martin and Ruszinkó [6]: consider the random sequence $\mathcal{I}_k = (E_1, E_2, \ldots, E_m)$ where $E_i \in \binom{[n]}{k}$ for $i = 1, 2, \ldots, m$ and (i) E_1 is uniformly random and (ii) E_{i+1} is randomly chosen from the k-subsets of $[n]$ that are not already chosen and that intersect each of $E_1, E_2, \ldots, E_i$. The process continues until no further sets can be added i.e. until $\{E_1, E_2, \ldots, E_m\}$ is a maximal intersecting family. We will abuse terminology and sometimes consider $\mathcal{I}_k$ to be a set of edges (i.e. a k-uniform hypergraph) instead of a sequence of edges.

We denote the hypergraph comprising the first r accepted edges by $\mathcal{H}_r$. For a set $S \subseteq [n]$, we let $e_r(S)$ denote the number of edges $E_i, i \le r$ such that $E_i \cap S \neq \emptyset$. For our purposes, an intersecting family is *trivial* if it is of the form $\mathcal{A}_x = \left\{E \in \binom{[n]}{k} : x \in E\right\}$ where x is some fixed element of $[n]$. Now $|\mathcal{A}_x| = \binom{n-1}{k-1}$ and the famous result of Erdős, Ko and Rado [9] is

*Department of Mathematics, Western Michigan University, Kalamazoo MI49008, Research supported in part by Simons Foundation Grant #426894

†Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA 15213, Research supported in part by NSF grant DMS1952285

‡Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA 15213.

§Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA 15213, Research supported in part by NSF grant DMS1700365.

that if $k \leq n/2$ then any intersecting family of k-sets is bounded in size by $\binom{n-1}{k-1}$ and that the maximum is achieved only by trivial families. The aim of [6] was to see when the process $\mathcal{I}_k$ produces a trivial family w.h.p. Their main result is the following:

Theorem 1. *Let $\mathcal{E}_0$ be the event that $\mathcal{I}_k = \mathcal{A}_x$ for some $x \in [n]$. If $k = c_n n^{1/3} < n/2$ then*

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_0) = \begin{cases} 1 & c_n \to 0 \\ \frac{1}{1+c^3} & c_n \to c \\ 0 & c_n \to \infty \end{cases}.$$

So if $c_n \to c$ in this theorem, then in the limit, there is a positive probability of $\frac{c^3}{1+c^3}$ that $\mathcal{I}_k$ is not trivial. What can be said about this case?

Patkós [14] considered this question and showed that in the random intersecting process we study here that for $k = c_n n^{1/3}$ with $c_n \to c$ with probability

$$\left(\frac{c^3}{1+c^3}\right)\left(\frac{3}{3+c^3}\right)$$

$\mathcal{I}_k$ is a *Hilton–Milner*-type hypergraph. A Hilton–Milner-type hypergraph was first described in [11] and is a k-uniform hypergraph obtained by specifying a single vertex v and a single edge F that does not contain v and the edges are $F \cup \{E \in \binom{[n]}{k} \mid v \in E, F \cap E \neq \emptyset\}$. Our goal here is to extend beyond the two possibilities considered so far, the trivial system and the Hilton–Milner system, to further understand the full distribution of the asymptotic behavior in the critical regime.

Our first result (Theorem 2 below) applies to the case where $k = c_n n^{1/3}$ where $c_n \to c$. We describe two randomly generated families $\mathcal{I}^*, \mathcal{I}^{**}$, which are approximately the same size, such that w.h.p. $\mathcal{I}^* \subseteq \mathcal{I}_k \subseteq \mathcal{I}^{**}$. Furthermore $\mathcal{I}^*, \mathcal{I}^{**}$ can be determined from the very early evolution of the process.

We define two hitting times which will help determine $\mathcal{I}^*, \mathcal{I}^{**}$. We let r_0 be the first step r such that $\mathcal{H}_r$ has a vertex of degree three. We let J be the set of vertices of degree two in $\mathcal{H}_{r_0-1}$. A set $S \subseteq J$ is said to be *independent* if no edge of $\mathcal{H}_{r_0-1}$ contains more than one member of S. For $r \geq r_0$ we let $\mathcal{S}_r$ be the family of all $S \subseteq J$ that are independent and which meet every edge $E_{r_0}, \ldots, E_r$. Let r_1 be the first step r when $\mathcal{S}_r$ is an intersecting family, and set $\mathcal{S} := \mathcal{S}_{r_1}$ to be that intersecting family. (We show below that w.h.p. $\mathcal{S}_r$ becomes an intersecting family before the process terminates.) To summarise:

$$r_0 = \text{the first step } r \text{ such that } \mathcal{H}_r \text{ has a vertex of degree three.}$$
$$r_1 = \text{the first step } r \text{ when } \mathcal{S}_r \text{ is an intersecting family.}$$

In the limit as $n \to \infty$, $\mathcal{S}$ has a natural interpretation as an intersecting family of matchings of a complete graph. Both the size of this complete graph and the actual intersecting family are random, and in general $\mathcal{S}$ will not be a uniform hypergraph. In the appendix we explain this random process to generate $\mathcal{S}$ outside the broader context of its role in Theorem 2.

Theorem 2. *Let $k = \lfloor cn^{1/3} \rfloor$ for some positive constant c. Then we have the following.*

(a) Let $b = b(n) \to \infty$. Then w.h.p. $r_0 \le r_1 \le b$.

(b) W.h.p. $\mathcal{I}^* \subseteq \mathcal{I}_k \subseteq \mathcal{I}^{**}$ where

$$\mathcal{I}^* = \mathcal{H}_{r_0-1} \cup \{E : E \text{ contains some } S \in \mathcal{S} \text{ and intersects every edge of } \mathcal{H}_{r_0-1}\},$$
$$\mathcal{I}^{**} = \mathcal{H}_{r_0-1} \cup \{E : E \text{ intersects each } S \in \mathcal{S} \text{ and intersects every edge of } \mathcal{H}_{r_0-1}\}.$$

(c) W.h.p. $|\mathcal{I}^{**} \setminus \mathcal{I}^*| = o(|\mathcal{I}^*|)$ and

$$\frac{|\mathcal{I}_k|}{\binom{n}{k}} = (1 + o(1))\frac{\sum_{S\in\mathcal{S}} c^{3(r_0-1-|S|)}}{k^{r_0-1}} \text{ as } n \to \infty.$$

(d) The sequence

$$\left\{\frac{\lim_{n\to\infty} \mathbb{P}(r_0 = i+1)}{\lim_{n\to\infty} \mathbb{P}(r_0 > i+1)}\right\}_{i\ge 0}$$

has exponential generating function

$$e^x(e^{x^2/(2c^3)} - 1).$$

Theorem 2 recovers the eariler results of [6] that the probability of $\mathcal{E}_0$ is asymptotically $1/(1+c^3)$ and the result of [14] about the probabiliy of a Hilton–Milner-type hypergraph as special cases. After carefully describing the distribution of $\mathcal{S}$, we explain how these results are recovered in Example 11.

Note that the distribution of r_0 can be recovered from part (d). Regarding part (b), we note that the containment $\mathcal{I}^* \subseteq \mathcal{I}_k \subseteq \mathcal{I}^{**}$ is not in general strict at either end. For example $\mathcal{I}^*$ may not be maximally intersecting, and $\mathcal{I}^{**}$ may not be intersecting at all. We suspect that it may be possible to give a better estimate for where the final $\mathcal{I}_k$ falls in between $\mathcal{I}^*$ and $\mathcal{I}^{**}$, but that this information cannot be determined from the first few edges.

We say that a family $\mathcal{J} \subseteq \binom{[n]}{k}$ is a *j-junta* if there is some $J \subseteq [n]$ with $|J| = j > 0$ and a family $\mathcal{J}^*$ of subsets of J such that

$$\mathcal{J} = \{T \in \binom{[n]}{k} : T \cap J \in \mathcal{J}^*\}.$$

We say that $\mathcal{J}$ is *generated by* $(J, \mathcal{J}^*)$. Juntas are relevant here because, roughly speaking, they can provide a simple "certificate" that a family is intersecting. Specifically if $\mathcal{H} \subseteq \mathcal{J}$ for some junta $\mathcal{J}$ generated by $(J, \mathcal{J}^*)$ with $\mathcal{J}^*$ intersecting, then $\mathcal{H}$ must be intersecting. For example a trivial intersecting family is a 1-junta, sometimes called a *dictatorship*. For a second example a Hilton–Milner-type hypergraph is almost (with the exception of one edge) contained in a dictatorship. Using this terminology, Theorem 2 has the following corollary.

Corollary 3. *Let $k = \lfloor cn^{1/3} \rfloor$ where $c > 0$, and $b = b(n) \to \infty$. Then w.h.p. there is some b-junta $\mathcal{J}$ such that $|\mathcal{I}_k \setminus \mathcal{J}| = o(|\mathcal{I}_k|)$. Furthermore $\mathcal{J}$ is determined by the hypergraph $\mathcal{H}_b$.*

Bohman, Frieze, Martin, Ruszinkó and Smyth [7] considered the case where $n^{1/3} \ll k \ll n^{5/12}$. They prove that w.h.p. the structure of $\mathcal{I}_k$ satisfies the following: there exists a hypergraph $\mathcal{H}$ with τ edges and a vertex v such that the following holds: (i) $n\tau^3/6k^3$ converges to the exponential distribution with mean 1 and (ii) $\mathcal{I}_k$ consists of all $E \in \binom{[n]}{k}$ that (a) contain v and (b) meet every edge of $\mathcal{H}$ that does not contain v. So in this regime we also have that w.h.p. $\mathcal{I}_k$ is almost (with the exception of the $poly(n)$ edges of $\mathcal{H}$) contained in a dictatorship. It is interesting to note that, combining the results of [6], [7] and Theorem 2, for the whole regime $k \ll n^{5/12}$ we have that w.h.p. $\mathcal{I}_k$ is almost contained in a relatively small junta of order $k^2/n^{1/3}$, and furthermore this small junta is a dictatorship unless $k = \Theta(n^{1/3})$. This is not a shortcoming in our proof: indeed, for $k = \Theta(n^{1/3})$ and sufficently small fixed $\varepsilon > 0$, Theorem 2 implies that there is a probability bounded away from zero that no dictatorship contains more than $1 - \varepsilon$ proportion of $\mathcal{I}_k$.

We also make a little progress on the case where $k \gg n^{1/2}$. In particular we give non-trivial upper and lower bounds on $|\mathcal{I}_k|$. Let $N = \binom{n}{k}$ and $d = \binom{n-k}{k}$. There is a trivial lower bound of $N/(d+1)$ on the minimum size of a maximal intersecting family. We prove that w.h.p. $|\mathcal{I}_k|$ is significantly larger.

Theorem 4. *For all fixed and sufficiently small $\zeta > 0$ there exists some $\eta = \eta(\zeta) > 0$ such that we have the following. For all k such that $\zeta^{-1}n^{1/2}\log^{1/2} n \le k \le \left(\frac{1}{2} - \zeta\right) n$, w.h.p. $\Omega\left(\frac{N}{d}\log\frac{N}{d}\right) \le |\mathcal{I}_k| \le O\left(N\left(\frac{d}{N}\right)^{\eta}\right)$.*

Our proof of Theorem 4 will resemble the analysis of some similar processes that have been studied. Suppose we are given a graph G. In the *random greedy independent set process*, or just *independent process*, we choose a random sequence $(v_1, \dots, v_m)$ of vertices of G where v_{i+1} is randomly chosen from all vertices not already chosen and not adjacent to any chosen vertex. Then the process we are studying is equivalent to the independent process on the Kneser graph $K(n,k)$ which has vertex set $\left\{v_S : S \in \binom{[n]}{k}\right\}$, and where v_S is adjacent to $v_{S'}$ whenever $S \cap S' = \emptyset$. Wormald [15] was the first to study the independent process, which he analyzed on random regular graphs. Lauer and Wormald [13] extended the analysis from random regular graphs to all regular graphs of sufficiently high girth. Bennett and Bohman [4] analyzed a generalization of this process to d-regular hypergraphs with $d \to \infty$ sufficiently fast assuming a relatively mild upper bound on codegrees. So, one could hope that we could just use some existing analysis of the independent process and apply it to the Kneser graph. Unfortunately, all the results for general deterministic graphs (or at least the ones we could find) assume a similar upper bound on vertex codegrees which the Kneser graph does not satisfy. Indeed, for say $k = cn$ for constant c, the graph $K(n,k)$ has pairs of vertices whose codegree is the same order of magnitude as their degrees. Even worse, every vertex in the whole graph has high codegree with a few other vertices. Thus, our main contribution in this regime is to carry out an analysis of the process that resembles previous work but allows

for a few pairs of vertices with high codegree. The analysis uses the so-called differential equation method (see [5] for a gentle introduction).

To put the above results in some context we point out another related but distinct random model of intersecting families. Improving a result of Balogh, Das, Delcourt, Liu and Sharifzadeh [1], Balogh, Das, Liu, Sharifzadeh and Tran [2] and Frankl and Kupavskii [12] showed that if $n \geq 2k + \Omega((k \log k)^{1/2})$ then almost all intersecting families are trivial. Balogh, Garcia, Li and Wagner [3] reduced the lower bound on n to $2k + 100 \log k$. Dinur and Friedgut [8] proved that when $k = pn$ for constant $0 < p < 1/2$, every intersecting family is (up to a small number of members) contained in an $O(1)$-junta. In [8] they also proved that when $k = o(n)$ every intersecting family is almost contained in a trivial family (a dictatorship).

The organization of the paper is as follows. In Section 2 we prove Theorem 2. In Sections 3 and 4 we prove Theorem 4, where in each of the two sections we consider a certain range of k. The proofs for each regime are almost identical, but the error bounds are somewhat different. In Section 5 we give some concluding remarks.

2 Proof of Theorem 2

2.1 The first $o(\sqrt{\log n})$ steps: generating $\mathcal{S}$

We examine the first r steps in the process for $r = o(\sqrt{\log n})$. The most important part of the random intersecting process in these first few steps is the generation of a (not necessarily uniform) intersecting hypergraph $\mathcal{S}$ that will be the $\mathcal{S}$ in the definition of $\mathcal{I}^*$ and $\mathcal{I}^{**}$. The random intersecting hypergraph $\mathcal{S}$ in the limit as $n \to \infty$ is described on its own in the appendix. To describe $\mathcal{S}$ here we introduce some notation.

Denote by $V(\mathcal{H}_r)$ the set of all vertices of degree at least one in $\mathcal{H}_r$, and by $U(\mathcal{H}_r)$ the set of vertices of degree exactly one. For any set S of vertices of degree at least two in $\mathcal{H}_r$, let $E_r(S)$ denote the set of edges of $\mathcal{H}_r$ that contain a vertex in S and let $e_r(S) = |E_r(S)|$. Let $\chi_r(S) = e_r(S) - 2|S|$ and let $\chi_r^* = \max\{\chi_r(S) \mid S \subseteq V(\mathcal{H}_r) \setminus U(\mathcal{H}_r)\}$. At each step r, we let

$$\mathcal{S}_r = \{S \subseteq V(\mathcal{H}_r) \setminus U(\mathcal{H}_r) \mid \chi_r(S) = \chi_r^*\}$$

It will become clear in the proof that this $\mathcal{S}_r$ is indeed the same as the $\mathcal{S}_r$ mentioned in the introduction. Here we will describe three regimes for $\mathcal{S}_r$: a growing regime, a diminishing regime, and a stable regime. As the name suggests, $\mathcal{S}$ will denote the unchanging $\mathcal{S}_r$ within the stable regime.

The growing regime coincides with all steps before vertices of degree three show up. Letting r_0 denote the number of edges when the first vertex of degree three appears we have by the following lemma that w.h.p. for $r < r_0$, $\mathcal{S}_r$ is just the collection of all independent sets of

vertices of degree two in $\mathcal{H}_r$ and $\chi_r^* = 0$. A hypergraph H is *simple* if $E_1, E_2 \in E(H)$ implies that $|E_1 \cap E_2| \leq 1$.

Lemma 5. *If $r = o(\sqrt{\log n})$ and $\mathcal{H}_{r-1}$ is simple and has no vertices of degree at least 3 then the probability that $\mathcal{H}_r$ is not simple is $O(1/k^{1-o(1)})$.*

This lemma will follow as a corollary to another lemma we prove shortly, so we save the proof for later. By part (d) of Theorem 2, r_0 will take some value bounded by any slowly growing function of n, and we use without further remark that with high probability r_0 is smaller than any function that tends to infinity.

By simplicity, before vertices of degree three appear the size of $\mathcal{S}_r$ is given by the following lemma: a hypergraph will be called intersecting if its edge set defines an intersecting family.

Lemma 6. *If $\mathcal{H}$ is a k-uniform simple intersecting hypergraph on r edges with maximum degree two then for each $1 \leq m \leq r/2$ the number of independent sets of vertices of degree two of size m is*

$$\frac{r!}{(r-2m)!m!2^m}.$$

Proof. Let $I_{r,m}$ denote the number of independent sets of vertices of degree two of size m when $\mathcal{H}$ satisfies the assumptions. Then, $I_{r,m}$ satisfies the recurrence,

$$I_{r,m} = \frac{\binom{r}{2} I_{r-2,m-1}}{m}.$$

Indeed we can pick any of the $\binom{r}{2}$ vertices of degree two to start building an independent set. We then delete the selected vertex and the two edges that contain it and then take an independent set of vertices of degree two of size $m-1$ from the remaining hypergraph. However, this overcounts by a factor of m because of the choice of the first vertex.

Now clearly $I_{r,1} = \binom{r}{2} = \frac{r!}{(r-2)!1!2^1}$. And by induction using the recurrence above we have

$$I_{r,m} = \frac{r(r-1)(r-2)!}{2m(r-2-2m+2)!(m-1)!2^{m-1}} = \frac{r!}{(r-2m)!m!2^m} = \binom{r}{2m}\binom{2m}{m}2^{-m}.$$

□

We next show that once vertices of degree three appear χ_r^* starts to increase by one at each step, and that $\mathcal{S}_r$ decays, until it reaches an intersecting family at time r_1 with $\mathcal{S} := \mathcal{S}_{r_1}$.

Before we prove how the random intersecting process stabilizes to $\mathcal{S}$, we introduce some definitions:

Definition 1. *An edge E that extends the intersecting family $\mathcal{H}_r$ to an intersecting family $\mathcal{H}_{r+1}$ is said to be an* almost simple extension *if for every pair of distinct $x, y \in E$ that are already contained in a common edge in $\mathcal{H}_r$, both x and y have degree at least two in $\mathcal{H}_r$.*

Definition 2. *Given a k-uniform intersecting family $\mathcal{H}_r$ on r edges, an edge E that extends the intersecting family is* good *if it is an almost simple extension and the set of vertices S of $\mathcal{H}_r$ of degree at least two that belong to E satisfies $\chi_r(S) = \chi_r^*$. Otherwise we say the extension is* bad.

Lemma 7. *Suppose that $r = o(\sqrt{\log n})$ and that $\mathcal{H}_r$ has been generated by the random intersecting process via good extensions with $\mathcal{S}_r$ as defined above, then the probability that $\mathcal{H}_{r+1}$ is generated by selecting a bad extension of $\mathcal{H}_r$ is at most $1/k^{1-o(1)}$.*

Proof. We first show that the number of extensions that are not almost simple is at most $k^{-(r+1-\chi_r^*+o(1))}\binom{n}{k}$.

We have two cases to consider, either E contains two vertices of degree one in the same edge of $\mathcal{H}_r$ or else it contains one vertex of degree one and one vertex of degree two from an edge of $\mathcal{H}_r$. To build such an extension in the former case we first choose an edge (r choices) and then two vertices of degree one from it (at most k^2 choices). Next we choose a set S of vertices of degree at least two to belong to the new edge. Since $\mathcal{H}_r$ has been built out of good extensions, the number of vertices of degree at least two is at most $\binom{r}{2}$. So we have at most 2^{r^2} choices for S. Once the two vertices of degree one in a common edge and S have been selected for E we have covered $1 + e_r(S)$ edges, so we have $r - e_r(S) - 1$ edges still to cover. We cover these remaining edges with vertices of degree one since S already accounts for the vertices of degree larger than 1 that we will use. So we have to pick at least one vertex from the remaining $r - e_r(S) - 1$ edges. Lastly we choose another $k - (r - e_r(S) - 1) - |S| - 2$ vertices from the ground set. So, using the fact that $k \sim cn^{1/3}$, we see that the number of extensions that are not almost simple is at most

$$rk^2 2^{r^2} k^{r-e_r(S)-1}\binom{n}{k-(r-e_r(S)+|S|+1)} \leq k^{r-e_r(S)+1+o(1)}\left(\frac{k}{n}\right)^{r-e_r(S)+|S|+1}\binom{n}{k}$$
$$= \frac{1}{k^{r+1-(e_r(S)-2|S|)+o(1)}}\binom{n}{k}$$

This is therefore at most what we claimed since $e_r(S) - 2|S| \leq \chi_r^*$.

For the other type of not almost simple extensions we have the upper bound of

$$rkr^2 2^{r^2} k^{r-e_r(S)}\binom{n}{k-(r-e_r(S)+|S|+1)} \leq \frac{1}{k^{r+1-(e_r(S)-2|S|)+o(1)}}\binom{n}{k}$$

So again using $e_r(S) - 2|S| \leq \chi_r^*$ we arrive at the same conclusion as in the first case.

Next we turn our attention to almost simple extensions for which the chosen set S has $\chi_r(S) \leq \chi_r^* - 1$. Let $\nu(S)$ for S a collection of vertices of degree at least two in $\mathcal{H}_r$ be the number of almost simple extensions which contain S and otherwise meet each edge of $\mathcal{H}_r$ at a vertex of degree one. To build such an extension we choose a vertex of degree one from

the $r-e_r(S)$ edges not covered by S. For each such edge E there are between $k-r^2$ and k choices. This leaves $k-(r-e_r(S))-|S|$ vertices of E to be chosen from $[n]\setminus V(\mathcal{H}_r)$. Thus,

$$\big(k-r^2\big)^{r-e_r(S)}\binom{n-kr}{k-(r-e_r(S))-|S|}\le\nu(S)\le k^{r-e_r(S)}\binom{n}{k-(r-e_r(S))-|S|}. \quad (1)$$

Thus, since $r=o(k^{1/2})$,

$$\begin{aligned}\nu(S)=e^{O(r^2/k)}\frac{k^{2r-2e_r(S)+|S|}}{n^{r-e_r(S)+|S|}}\binom{n}{k}&=e^{O(r^2/k)}\frac{k^{2r-2e_r(S)+|S|}}{(c^{-1}k)^{3(r-e_r(S)+|S|)}}\binom{n}{k}\\&=e^{O(r^2/k)}\frac{c^{3(r-e_r(S)+|S|)}}{k^{r-e_r(S)+2|S|}}\binom{n}{k}\sim\frac{c^{3(r-e_r(S)+|S|)}}{k^{r-e_r(S)+2|S|}}\binom{n}{k}. \quad (2)\end{aligned}$$

Thus the number of almost simple extensions with $\chi_r(S)<\chi_r^*-1$ is at most

$$k^{-(r-e_r(S)+2|S|+o(1))}\binom{n}{k}\le k^{-(r+1+\chi_r^*+o(1))}\binom{n}{k}$$

The total number of extensions of $\mathcal{H}_r$ is at least the number of good extensions and by our estimate on $\nu(S)$ when $\chi_r(S)=\chi_r^*$ we have the number of good extensions is at least $k^{-(r+\chi_r^*+o(1)))}\binom{n}{k}$ and the claim follows. □

We see that Lemma 5 follows from Lemma 7.

Proof of Lemma 5. If $\mathcal{H}_r$ is simple with no vertices of degree at least 3 then $\mathcal{H}_r$ has been generated by good extensions with $\mathcal{S}_r$ as the collection of independent sets of vertices of degree two. If $\mathcal{H}_r$ has no vertices of degree at least 3 and is simple then the only good extensions of $\mathcal{H}_r$ are simple extensions, thus the probability of a non-simple extension is at most $1/k^{1-o(1)}$. □

Lemma 7 implies that w.h.p., step r_0 coincides exactly with the first time we take $S\in\mathcal{S}_r$ with $S\neq\emptyset$ and take a good extension with S as the set of vertices in $\mathcal{H}_r$ of degree at least two. It follows from (2) that $S\in\mathcal{S}_r$ is selected at each step for the next good extension with probability proportional $(c^{-3})^{|S|}$. So, if $\mathcal{H}_r$ has maximum degree two and $|\mathcal{S}_r|=\sigma=\binom{r}{2}$ then

$$\mathbb{P}(r_0=r+1\mid r_0>r)=\frac{e^{O(r^2/k)}\sum_{s\ge1}\binom{\sigma}{s}c^{-3s}}{\sum_{s\ge0}\binom{\sigma}{s}c^{-3s}}=e^{O(r^2/k)}\left(1-\frac{1}{(1+c^{-3})^\sigma}\right). \quad (3)$$

and so for $r=o(\sqrt{\log n})$,

$$\mathbb{P}(r_0>r)\le\prod_{\rho=1}^{r}\left(\frac{1}{(1+c^{-3})^{\binom{\rho}{2}}}+O\left(\frac{r^2}{k}\right)\right)<\frac{1+O(r^3/k)}{(1+c^{-3})^{r(r+1)/2}}.$$

Clearly a maximal intersecting family has size at least k. Thus w.h.p. $r_0 \le \omega = \omega(n)$ for any function $\omega \to \infty, \omega = o(k)$.

Note that in general $\chi_{r+1}(T) \le \chi_r(T)+1$ for $T \subseteq \mathcal{S}_r$. Now for any $T \in \mathcal{S}_r$ so that $T \cap S = \emptyset$, with S selected for the good extension to $\mathcal{H}_{r+1}$ we observe that $\chi_{r+1}(T) = \chi_r(T)$ while for any $T \in \mathcal{S}_r$ such that $T \cap S \neq \emptyset$, $\chi_{r+1}(T) = \chi_r(T) + 1$. Thus if $T \cap S = \emptyset$ then $T \notin \mathcal{S}_{r+1}$ while if $T \cap S \neq \emptyset$ then T remains in $\mathcal{S}_{r+1}$.

Suppose now that $\mathcal{S}_r$ is not intersecting and that $T \in \mathcal{S}_r$. Then if $S \cap T = \emptyset$ then $T \notin \mathcal{S}_{r+1}$. So,

$$\mathbb{P}(T \in \mathcal{S}_{r+1}) \le 1 - \frac{e^{O(r^2/k)}c^{-3|T|}}{\sum_{S \in \mathcal{S}_r} c^{-3|S|}}$$

and so, assuming $r_1 \ge r_0 + \rho$,

$$\mathbb{P}(T \in \mathcal{S}_{r_0+\rho}) \le \prod_{i=1}^{\rho} \left(1 - \frac{e^{O((r_0+i)^2/k)}c^{-3|T|}}{\sum_{S \in \mathcal{S}_{r_0}} c^{-3|S|}}\right)^i \le \exp\left\{-\rho \frac{e^{O((r_0+\rho)^2/k)}c^{-3\binom{r_0}{2}}}{\sum_{S \in \mathcal{S}_{r_0}} c^{-3|S|}}\right\}$$

This implies that if $\rho \gg \frac{c^{-3\binom{r_0}{2}}}{\sum_{S \in \mathcal{S}_{r_0}} c^{-3|S|}}$ then $r_1 \le r_0 + \rho$ w.h.p. So, without further remark we have that with high probability r_1 is smaller than any function of n tending to infinity. This verifies part (a) of Theorem 2.

2.2 Step $o(\sqrt{\log n})$ to step $n^{\omega(1)}$

We've shown that with high probability in the first $o(\sqrt{\log n})$ steps we have only made good extensions. Moreover we have also described the growing regime, diminishing regime, and stable regime for $\mathcal{S}_r$. We now want to show that conditioned on knowing the final, stable family $\mathcal{S}$, and the two hitting times r_0 and r_1 (each of which is a random positive integer that is sampled according to some asymptotic distribution) that we have $\mathcal{I}^* \subseteq \mathcal{I}_k \subseteq \mathcal{I}^{**}$, with $\mathcal{I}^*$ and $\mathcal{I}^{**}$ depending on r_0 and $\mathcal{S}$.

In order to prove $\mathcal{I}^* \subseteq \mathcal{I}_k \subseteq \mathcal{I}^{**}$ we first show that for any $r \ge r_1$, the probability that we add $E \in \binom{[n]}{k} \setminus \mathcal{I}^{**}$ is at most $O\left(\frac{1}{n^M}\right)$ where M can be set to be any constant. This means that $\mathcal{H}_r \subseteq \mathcal{I}^{**}$ for $n^{\omega(1)}$ steps; we handle the rest of the process in the next subsection.

We say that an edge is *open* after r steps if it is in $\mathcal{H}_r$ or if it meets $E_1, E_2, \ldots, E_r$, otherwise we say that it is closed, and let $\mathcal{O}_r$ be the set of open edges at step r. By what we showed in the first $o(\sqrt{\log n})$ steps we have $\mathcal{H}_{r_1} \subseteq \mathcal{I}^{**}$ w.h.p. For $r \ge r_1$, all of $\mathcal{I}^*$ remains open as long as we have $\mathcal{H}_r \subseteq \mathcal{I}^{**}$. If an edge E does not meet all of $\mathcal{H}_{r_1}$ then E is already closed at step r_1. Thus the only edges that concern us here are those edges E that meet all of $\mathcal{H}_{r_1}$ but are not in $\mathcal{I}^{**}$ because they are disjoint from some $S \in \mathcal{S}$.

We first lower bound $|\mathcal{I}^*|$. Let $S \in \mathcal{S}$. Then S meets $E_{r_0}, \ldots E_{r_1}$ and also $2|S|$ of the edges $E_1, \ldots E_{r_0-1}$. Thus to choose an edge $E \in \mathcal{I}^*$, $E \supseteq S$ we can choose one vertex from each of

the $r_0 - 1 - 2|S|$ edges of $\mathcal{H}_{r_1}$ that are disjoint from S, and then there is no restriction on the other $k - r_0 + 1 + |S|$ vertices of E. So, assuming that $\mathcal{H}_r \subseteq \mathcal{I}^{**}$, the number of edges in $\mathcal{I}^*$ at step r is at least

$$(k-r_0)^{r_0-1-2|S|}\binom{n-r_0k}{k-r_0+1+|S|} \sim k^{r_0-1-2|S|}\left(\frac{k}{n}\right)^{r_0-1-|S|}\binom{n}{k} = k^{-(r_0-1+o(1))}\binom{n}{k}. \quad (4)$$

Of course some (at most r) of the edges in $\mathcal{I}^*$ are only open because they are already in $\mathcal{H}_r$, but if $r \leq n^{\omega(1)}$ then this has a negligible effect on the above estimate.

Now we upper bound the number of open edges $E \notin \mathcal{I}^{**}$ at some step r with $r_1 \ll r \ll \sqrt{\log n}$. To determine the number of choices for E we first choose some subset T of the vertices of $\mathcal{H}_r$ of degree at least two with $T \cap S = \emptyset$ for some $S \in \mathcal{S}$. T will be the set of vertices of degree at least two in $\mathcal{H}_r$ that belong to E. The number of choices for T is at most $2^{r^2} = n^{o(1)}$. Once T is selected there is $S \in \mathcal{S}$ with $T \cap S = \emptyset$. For $\ell = \ell(r_1, r_2)$ we have that $\chi_r(T) \leq r - r_0 + 1 - \ell$. Now for each q with $r_1 \leq q = o(\sqrt{\log n})$ we have at least a $\frac{1}{(1+c^3)^{r_1^2}}$ probability to take S to be the set selected from $\mathcal{S}$ for the good extension. (The lower bound on probability is derived as in (3).) If $\frac{r}{(1+c^3)^{r_1^2}} \to \infty$ then w.h.p. we pick S at least ℓ times between steps r_1 and r, and at these steps $\chi_r(T)$ does not increase. Thus at step r we have the number of sets $E \notin I^{**}$ with T as the set of vertices of degree at least two in E is at most

$$k^{r-e_r(T)}\binom{n}{k-(|T|+r-e_r(T))} \leq k^{-(r_0-1+\ell+o(1))}\binom{n}{k}. \quad (5)$$

So at step r the number of open edges $E \notin \mathcal{I}^{**}$ is at most $n^{-M}\binom{n}{k}$ where M can be set as any large constant with an appropriate choice of ℓ. However as long as $\mathcal{H}_r$ is contained in $\mathcal{I}^{**}$ all edges of $\mathcal{I}^*$ remain open. For $r \leq o(\sqrt{\log n})$, by step r we have not yet selected an edge not in $\mathcal{I}^{**}$ and from then on the probability that we pick an edge outside of $\mathcal{I}^{**}$ is at most $n^{-\omega(1)}$. This follows from (4) and (5). Thus with high probability $\mathcal{H}_{n^{\omega(1)}} \subseteq I^{**}$.

2.3 From step $n^{\omega(1)}$ onward

In this section we handle the remainder of the process. We will show that for L sufficiently large (depending on r_0) $\mathcal{O}_{n^L} \subseteq \mathcal{I}^{**}$, which together with results in the previous subsection implies that $\mathcal{I}_k \subseteq \mathcal{I}^{**}$, and that in turn implies that $\mathcal{I}^* \subseteq \mathcal{I}_k$. Let $L = r_0 + 3$. For $E \in \binom{[n]}{k}$ and knowing r_0, r_1, and $\mathcal{S}$, let

$$\mathcal{D}(E) = \{F \in \mathcal{I}^* \mid F \cap E = \emptyset\}$$

Suppose $E' \in \mathcal{O}_{n^L} \setminus \mathcal{I}^{**}$, so that E' is disjoint with some set $S \in \mathcal{S}$. We lower bound $|\mathcal{D}(E')|$. Quite crudely, there is at least one choice for some set of vertices $S \in \mathcal{S}$ that is disjoint with E' and such that $|S| \leq r_0$, and from here we can pick from each edge $E_1, \cdots, E_{r_0-1}$ not

intersecting S a single vertex $v_j \in E_j \setminus E'$. Such a choice always exists since $E' \cap S = 0$ and $E' \neq E_1, \cdots, E_{r_0-1}$. So we count edges $E \in \mathcal{D}(E')$ to be at least

$$\binom{n-k-2r_0}{k-2r_0} \geq \frac{1}{k^{4r_0+o(1)}}\binom{n}{k}$$

If we pick a set from $\mathcal{D}(E')$ before we pick E' then E' can never be selected. For $r \leq n^{\omega(1)}$ the probability at each step that we pick a set in $\mathcal{D}(E')$ is at least

$$k^{-(3r_0+1+o(1))}.$$

This estimate again uses the estimate of $k^{-r_0+1+o(1)}\binom{n}{k}$ for the total number of extensions. Thus the probability that E' remains open at step n^L is at most

$$(1-k^{-(3r_0+1+o(1))})^{n^L} \leq \exp\left\{-k^{3L-3r_0+1+o(1)}\right\}.$$

Thus, the expected number of open edges $E' \notin \mathcal{I}^{**}$ at step n^L is at most

$$\binom{n}{k}\exp\left\{-k^{3L-3r_0+1+o(1)}\right\} \leq \exp\left\{3k\log k - k^{3L-(3r_0+1+o(1))}\right\} = o(1),$$

since $L = r_0 + 3$. This completes the proof of part (b) of Theorem 2.

2.4 The size of $\mathcal{I}^*$ and $\mathcal{I}^{**}$

Theorem 2 tells us that our final intersecting family $\mathcal{I}_k$ satisfies

$$\mathcal{I}^* \subseteq \mathcal{I}_k \subseteq \mathcal{I}^{**}.$$

While in general we do not have a complete description of $\mathcal{I}_k$, we do have enough to conclude that following

Lemma 8. *Conditional on the process having reached step r_1 with stable family $\mathcal{S}_{r_1} = \mathcal{S}$, we have w.h.p.*

$$\frac{|\mathcal{I}_k|}{\binom{n}{k}} = (1+o(1))\frac{\sum_{S\in\mathcal{S}} c^{3(r_0-1-|S|)}}{k^{r_0-1}} \qquad \text{as } n \to \infty.$$

Proof. It suffices to find a lower bound on $\mathcal{I}^*$ and an upper bound on $\mathcal{I}^{**}$. Both approximations are based largely on the fact that in the growing regime every $S \in \mathcal{S}_r$, $r \leq r_0 - 1$ has $\chi_r(S) = 0$ and in the diminishing regime and stable regime every $S \in \mathcal{S}_r$, $r_0 \leq r$ has $\chi_r(S) = r - r_0 + 1$. Thus $\chi_{r_1}(S) = r_1 - r_0 + 1$. We now bound $|\mathcal{I}^{**}|$. For any set T of vertices of degree at least two in $\mathcal{H}_{r_1}$ we have $\chi_{r_1}(T) \leq r_1 - r_0 + 1$ with $\chi_{r_1}(T) = r_1 - r_0 + 1$ if and only if $T \in \mathcal{S}$. For each choice of T the number of hyperedges of $\mathcal{I}^{**}$ so that T is the set of vertices of degree at least two in $\mathcal{H}_{r_1}$ belonging to the hyperedge is at most

$$k^{r_1-e(T)}\binom{n}{k-(|T|+r_1-e_{r_1}(T))}.$$

We sum this over the at most 2^{r^2} choices for T, but the largest order terms come from $T \in \mathcal{S}$, so we have

$$\begin{aligned} |\mathcal{I}^{**}| &\leq (1+o(1)) \sum_{S \in \mathcal{S}} \frac{c^{3(r_1-(e_{r_1}(S)-|S|))}}{k^{r_1-(e_{r_1}(S)-2|S|))}} \binom{n}{k} \\ &= (1+o(1)) \sum_{S \in \mathcal{S}} \frac{c^{3(r_0-1-|S|))}}{k^{r_0-1}} \binom{n}{k}, \end{aligned}$$

after using $\chi_{r_1}(S) = r_1 - r_0 + 1$.

On the other hand we have that for any fixed $S \in \mathcal{S}$ the number of sets that contain S and meet all other sets in $\mathcal{H}_{r_1}$ in a vertex of degree one is at least

$$(k-(r_1)^2)^{r_1-e(S)} \binom{n-kr_1}{k-(|S|+r_1-e_{r_1}(S))} = (1-o(1)) \frac{c^{r_0-1-|S|}}{k^{r_0-1}} \binom{n}{k}.$$

We can sum this over all choices of $S \in \mathcal{S}$ to arrive at the conclusion. □

This completes the proof of part (c) of Theorem 2.

We finally consider part (d). The distribution of r_0 can be recovered from the following generating function.

Theorem 9. *Consider the random intersecting process with $k = \lfloor cn^{1/3} \rfloor$ where $c > 0$ is a constant and let X be the random variable counting the number of steps until a vertex of degree three appears, then for $n \to \infty$ the following sequence*

$$\left\{ \frac{\lim_{n\to\infty} \mathbb{P}(X = r+1 \mid X > r)}{\lim_{n\to\infty} \mathbb{P}(X > r+1 \mid X > r)} \right\}_{r \geq 0}$$

has exponential generating function

$$e^x(e^{x^2/(2c^3)} - 1).$$

Proof. By Lemma 5 we have that if $r = o(\sqrt{\log n})$ then w.h.p. either $\mathcal{H}_r$ has a vertex of degree at least three or it is a simple hypergraph. Let $SIMPLE(r)$ denote the event that H_r is simple. We first verify the following lemma.

Lemma 10.

$$\frac{\mathbb{P}(X = r+1 \mid X > r)}{\mathbb{P}(X > r+1 \mid X > r)} \tag{6}$$

is asymptotically equal to

$$\frac{\mathbb{P}(X = r+1 \mid X > r, SIMPLE(r))}{\mathbb{P}(X > r+1 \mid X > r, SIMPLE(r))}. \tag{7}$$

Proof. The ratio of the expressions in (6), (7) is

$$\frac{\mathbb{P}(X = r+1, X > r, SIMPLE(r))}{\mathbb{P}(X = r+1, X > r)} \times \frac{\mathbb{P}(X > r+1)}{\mathbb{P}(X > r+1, SIMPLE(r))}. \tag{8}$$

Now we bound the probability p_r that $X > r$ holds but that $SIMPLE(r)$ fails to hold. Clearly

$$p_1 = \mathbb{P}(X > 1, \neg SIMPLE(1)) \leq \mathbb{P}(\neg SIMPLE(1)) = 0.$$

For $1 < r = o(\sqrt{\log n})$ we have

$$\begin{aligned} p_r &= \mathbb{P}(X > r, \neg SIMPLE(r)) \\ &= \mathbb{P}(X > r, \neg SIMPLE(r), SIMPLE(r-1)) + \mathbb{P}(X > r, \neg SIMPLE(r-1)) \\ &\leq \mathbb{P}(\neg SIMPLE(r) \mid X > r-1, SIMPLE(r-1)) + \mathbb{P}(X > r-1, \neg SIMPLE(r-1)) \\ &\leq \frac{1}{k^{1-o(1)}} + p_{r-1}, \end{aligned}$$

after using Lemma 5 to bound the first summand. Thus $p_r = O(\sqrt{\log n}/k^{1-o(1)}) = O(1/k^{1-o(1)})$.

Going back to (8),

$$1 - \frac{\mathbb{P}(X > r, \neg SIMPLE(r))}{\mathbb{P}(X = r+1, X > r)} \leq \frac{\mathbb{P}(X = r+1, X > r, SIMPLE(r))}{\mathbb{P}(X = r+1, X > r)} \leq 1.$$

Now clearly,

$$\mathbb{P}(X = r+1, X > r) = \mathbb{P}(X = r+1) \geq \mathbb{P}(X = r+1, SIMPLE(r)).$$

Moreover the rest of the proof of Theorem 9 shows that if $\mathcal{H}_r$ is simple then the number of simple extensions and the number of almost simple extensions that add vertices of degree three have the same order of magnitude, so for fixed r if $\mathcal{H}_r$ is simple there is a positive probability that $\mathcal{H}_{r+1}$ is simple as well, so inducitively we have that for r fixed,

$$\mathbb{P}(X = r+1, SIMPLE(r)) = \zeta_r$$

where ζ_r is some positive constant. So with our estimate on p_r we have that

$$\frac{\mathbb{P}(X = r+1, X > r, SIMPLE(r))}{\mathbb{P}(X = r+1, X > r)} = 1 - O(k^{-(1-o(1))}).$$

A similar argument deals with the second quotient in (8). □

If $\mathcal{H}_r$ is a simple k-uniform intersecting hypergraph with maximum degree two then each pair of edges of $\mathcal{H}_r$ intersect in a unique vertex of degree two. In this case $\mathcal{H}_{r+1}$ will have vertices of degree three if and only if we select a nonempty set S of the vertices of degree two of $\mathcal{H}_r$ to belong to the new edge. The event that S is not an independent set of vertices of degree two is negligible as is the event that the extension of $\mathcal{H}_r$ to $\mathcal{H}_{r+1}$ is not almost

simple by Lemma 7. Thus the number of almost simple extensions that do not add a vertex of degree three is

$$(1+o(1))k^r\binom{n}{k-r} \sim k^r\left(\frac{c^3}{k^2}\right)^r\binom{n}{k}. \tag{9}$$

On the other hand to count the number of almost simple extensions that do add a vertex of degree three we have a choice of how many vertices of degree three we add. To add m vertices of degree three we must (with high probability) select an independent set of m vertices of degree two in $\mathcal{H}_r$.

Applying Lemma 6 we see that the number of almost simple extensions of $\mathcal{H}_r$ that create vertices of degree three is

$$(1+o(1))\sum_{m=1}^{\lfloor r/2\rfloor}\frac{r!}{(r-2m)!m!2^m}k^{r-2m}\binom{n}{k-(r-2m)-m} \sim \sum_{m=1}^{\lfloor r/2\rfloor}\frac{r!}{c^{3m}(r-2m)!m!2^m}k^r\left(\frac{c^3}{k^2}\right)^r\binom{n}{k}. \tag{10}$$

Thus from (9) and (10), we have for each $r = o(\sqrt{\log n})$,

$$\frac{\mathbb{P}(X=r+1 \mid X>r)}{\mathbb{P}(X>r+1 \mid X>r)} \sim \sum_{m=1}^{\lfloor r/2\rfloor}\frac{r(r-1)\cdots(r-2m+1)}{m!(2c^3)^m} \sim \sum_{m=1}^{\infty}\frac{r(r-1)\cdots(r-2m+1)}{m!(2c^3)^m}$$

as $r\to\infty$.

Now we just verify the exponential generating function is what we claimed:

$$\begin{aligned}
\sum_{r=0}^{\infty}\sum_{m=1}^{\infty}\frac{r(r-1)\cdots(r-2m+1)}{r!m!(2c^3)^m}x^r &= \sum_{m=1}^{\infty}\sum_{r=2m}^{\infty}\frac{x^r}{(r-2m)!m!(2c^3)^m}\\
&= \sum_{m=1}^{\infty}\frac{(x^2/(2c^3))^m}{m!}\sum_{r=2m}^{\infty}\frac{x^{r-2m}}{(r-2m)!}\\
&= \sum_{m=1}^{\infty}\frac{(x^2/(2c^3))^m}{m!}\sum_{s=0}^{\infty}\frac{x^s}{s!}\\
&= \left(\exp\left\{x^2/(2c^3)\right\}-1\right)e^x.
\end{aligned}$$

□

Example 11. *Using this generating function and Theorem 2 we can recover two previous results. Using the generating function we have that* $\mathbb{P}(r_0=3)=\frac{1}{1+c^3}$. *In the case that* $r_0=3$ *we see that* $\mathcal{S}$ *is consists of just a singleton set and it stabilizes at the third step of the process, so* $r_1=3$. *In this case* $\mathcal{I}^*=\mathcal{I}^{**}$ *and is the star centered at that first vertex of degree three. So this recovers the* $c_n\to c$ *case of Theorem 1.*

Moreover we can also recover the Hilton–Milner-type statement of Patkós from our methods. If $r_0=4$ *then at the fourth step of the process we have three edges that all contain a single*

common vertex and there is a fourth edge not containing that vertex but meeting all three edges. In this case $\mathcal{I}^* = \mathcal{I}^{**}$ and is a Hilton–Milner type system. We have

$$\mathbb{P}(r_0 = 4) = \mathbb{P}(r_0 = 4 \mid r_0 > 3)\mathbb{P}(r_0 > 3) = \mathbb{P}(r_0 = 4 \mid r_0 > 3)\mathbb{P}(r_0 \neq 3).$$

Using the generating function

$$\mathbb{P}(r_0 = 4) = \left(\frac{3}{c^3+3}\right)\left(\frac{c^3}{1+c^3}\right),$$

and so we recover Corollary 1.6 of [14].

Naively we might expect that r_0 determines $\mathcal{S}$ which determines the final intersecting family. This however is only the case when r_0 happens to be small. When $r_0 \leq 6$ it isn't too difficult to see that $\mathcal{S}$ must stabilize to be a star, i.e. it will stabilize as all independent sets of vertices of degree two in $\mathcal{H}_{r_0-1}$ that contain some fixed vertex. If $r_0 = 7$ however, we see that we can get a $\mathcal{S}$ that is not a star. If $r_0 = 7$ it is possible that $\mathcal{S} = \{\{u,v\},\{v,w\},\{u,w\},\{u,w,v\}\}$ for $\{u,v,w\}$ an independent set of vertices of degree two in $\mathcal{H}_6$.

This completes the proof of Theorem 2.

3 Proof of Theorem 4 when $k = \Theta(n)$

In this section we assume $k = cn$ where $c = c(n)$ such that $\zeta \leq c \leq \frac{1}{2} - \zeta$ for some constant $\zeta > 0$. Let $G = K(n,k)$ be the Kneser graph. Recall that G has vertex set $\left(v_S : S \in \binom{[n]}{k}\right)$, and v_S is adjacent to v'_S when $S \cap S' = \emptyset$.

Let $g(x) := x\log(x)$ for $x > 0$ and let $g(0) = 0$. Stirling's formula gives us that for $a \geq b = \Omega(1)$ we have

$$\binom{an}{bn} = \exp\{[g(a) - g(b) - g(a-b)]n + O(\log n)\}. \tag{11}$$

Thus G has N vertices and is d-regular for

$$N = \binom{n}{cn} = e^{p_N(c)n + O(\log n)} \qquad \text{where } p_N(c) = -g(c) - g(1-c),$$
$$d = \binom{n-cn}{cn} = e^{p_d(c)n + O(\log n)} \qquad \text{where } p_d(c) = g(1-c) - g(c) - g(1-2c).$$

We claim that

$$N^{\varepsilon_1} < d < N^{1-\varepsilon_1} \tag{12}$$

for some $\varepsilon_1 = \varepsilon_1(\zeta) > 0$. Indeed, first note that both $p_d(c), p_N(c)$ are continuous and positive for $c \in (0, 1/2)$, and so $p_d(c), p_N(c) = \Theta(1)$ for $c \in (\zeta, 1/2 - \zeta)$. Now observe that

$0 < p_d(c) < p_N(c)$ for all $c \in (0, 1/2)$. Indeed, $p_d(0) = p_N(0) = 0$, and since $g'(x) = 1 + \log x$ we have

$$\frac{d}{dc}(p_N(c) - p_d(c)) = \frac{d}{dc}(-2g(1-c) + g(1-2c)) = 2\log(1-c) - 2\log(1-2c) > 0.$$

Thus $p_N(c) = p_d(c) + \Omega(1)$ for $c \in (\zeta, 1/2 - \zeta)$. Now since

$$\frac{\log d}{\log N} = \frac{p_d(c)n + O(\log n)}{p_N(c)n + O(\log n)} \sim \frac{p_d(c)}{p_N(c)} \in \Big(\Omega(1), 1 - \Omega(1)\Big)$$

we see that (12) holds.

Now we claim that

$$\text{each vertex has codegree at most } dN^{-\varepsilon_2} \text{ with all but at most } dN^{-\varepsilon_2} \text{ vertices, for some } \varepsilon_2 = \varepsilon_2(\zeta). \quad (13)$$

Indeed, suppose $|S \cap S'| \le (c-\delta)n$ for some $\delta > 0$. Then $|S \cup S'| \ge (c+\delta)n$, and so the number of sets S'' that are disjoint with both S and S' (i.e. the codegree of $v_S, v_{S'}$) is at most

$$\binom{n-(c+\delta)n}{cn} = e^{p_1(c,\delta)n + O(\log n)} \quad \text{where } p_1(c,\delta) = g(1-c-\delta) - g(c) - g(1-2c-\delta)]n. \quad (14)$$

We claim that $p_1(c,\delta) < p_d(c)$ for any $\delta > 0$ such that $p_1(c,\delta)$ is defined. Indeed, we have $p_1(c,0) = p_d(c)$ and

$$\frac{\partial p_1}{\partial \delta}(c,0) = -\log(1-c) + \log(1-2c) < 0.$$

Thus the codegree bound on line (14) is at most $dN^{-\varepsilon_2}$ for some $\varepsilon_2 = \varepsilon_2(\zeta)$.

Now for a fixed set S, the number of S' such that $|S \cap S'| > (c-\delta)n$ is at most

$$\binom{cn}{(c-\delta)n}\binom{n}{\delta n} = e^{p_2(c,\delta)n + O(\log n)} \qquad \text{where } p_2(c,\delta) = g(c) - g(c-\delta) - 2g(\delta) - g(1-\delta).$$

We claim that for some sufficiently small δ we have $p_2(c,\delta) < p_d(c)$. This follows from the fact that $p_d(c) > 0$, $p_2(c,0) = 0$ and p_2 is continuous in δ. This verifies (13).

We let $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ for the remainder of Section 3. Both of (12) and (13) remain true with ε_1 and ε_2 replaced by ε respectively.

3.1 The good event

Let G be any d-regular graph on N vertices where $N^\varepsilon < d < N^{1-\varepsilon}$. Assume that for each vertex v, there are at most $dN^{-\varepsilon}$ vertices u whose codegree with v is at least $dN^{-\varepsilon}$. We

run the random greedy independent set process on G: initially $I(0) = \emptyset$. After r steps, let $I(r) = \{v_1, \ldots, v_r\}$ be the set of vertices in the independent set constructed so far. Let $V(r)$ be the set of available vertices (i.e. the vertices not adjacent to any vertex in $I(r)$). We choose a random vertex $v \in V(r)$ and put $v_{r+1} = v$.

Let $D_v(r)$ be the set of available neighbors of v, for $v \in V(r)$. Let $B(r)$ be the set of vertices w (available or not) such that some vertex $u \in I(r)$ has codegree with w at least $dN^{-\varepsilon}$. Note that $B(r)$ is nondecreasing in r. Let $C_v(r) := D_v(r) \cap B(r)$. Let

$$t = t(r) = \frac{dr}{N}.$$

Define the error function

$$f(t) := N^{-\varepsilon/20} e^{10t}. \tag{15}$$

Let the *good event* $\mathcal{E}_r$ be the event that the following inequalities (16), (17) and (18) all hold for all $r' \le r$ and where $t' = t(r')$:

$$\left| |V(r')| - Ne^{-t'} \right| \le Nf(t') \tag{16}$$

$$\left| |D_v(r')| - de^{-t'} \right| \le df(t') \text{ for all } v \in V(r') \setminus B(r') \tag{17}$$

$$|C_v(r')| \le dN^{-\varepsilon/10} \text{ for all } v \in V(r'). \tag{18}$$

We will show that w.h.p. the good event $\mathcal{E}_{r_{end}}$ holds where

$$r_{end} := \frac{\varepsilon}{1000} \frac{N \log N}{d}.$$

In particular this will imply that w.h.p. the process lasts to at least step r_{end}, proving the lower bound on $|\mathcal{I}_k|$ in Theorem 4 for the case of $k = cn$. Indeed,

$$e^t \le N^{\varepsilon/1000} \text{ implying that } f(t) \le \min\left\{e^{-t}, N^{-\varepsilon/25}\right\} = o(1).$$

for all $r \le r_{end}$, inequalities (16) and (17) imply that for all $r \le r_{end}$ we have

$$|V(r)| = (1 + o(1))Ne^{-t}, \qquad |D_v(r)| = (1 + o(1))de^{-t} \text{ for all } v \in V(r) \setminus B(r).$$

In particular $|V(r_{end})|$ is positive and so the process lasts to step r_{end}. The upper bound in Theorem 4 comes from the upper bound on $|V(r_{end})|$ implied by (16). Indeed, (16) gives

$$|V(r_{end})| \le Ne^{-t(r_{end})} + Nf(t(r_{end})) \le N^{1-\varepsilon/1000} + N^{1-\varepsilon/25}.$$

We will sometimes use the bounds (which hold for all $r \le r_{end}$ in the good event $\mathcal{E}_r$)

$$|V(r)| \ge N^{1-\varepsilon/10}, \qquad |B(r)| \le N^{1-\varepsilon/2}. \tag{19}$$

The first bound follows from our estimate of $V(r_{end})$. The second bound follows since $|B(r_{end})| \le r_{end} dN^{-\varepsilon}$.

3.2 Dynamic concentration of $V(r)$

Here we bound the probability that $\mathcal{E}_{r_{end}}$ fails due to condition (16). We define variables V^+ and V^- as follows.

$$V^{\pm} = V^{\pm}(r) := \begin{cases} |V(r)| - N(e^{-t} \pm f(t)) & \text{if } \mathcal{E}_{r-1} \text{ holds,} \\ V^{\pm}(r-1) & \text{otherwise.} \end{cases} \tag{20}$$

For $r = 0$ above, we interpret $\mathcal{E}_{-1}$ as the trivial event that always holds. Note that if $\mathcal{E}_{r_{end}}$ fails due to condition (16) then we either have that $V^+(r_{end}) > 0$ or $V^-(r_{end}) < 0$. To show that those events are unlikely we will establish V^+ is a supermartingale, i.e. $\mathbb{E}[\Delta(|V^+(r)|)\ |I(r)] \leq 0$, where we use the notation $\Delta(|V^+(r)|) = |V^+(r+1)| - |V^+(r)|$. Similarly we will show that V^- is a submartingale.

First we show $\mathbb{E}[\Delta|V^+(r)|\ |I(r)] \leq 0$. If $\mathcal{E}_r$ fails then $\Delta V^+(r) = 0$ by definition, so we assume $\mathcal{E}_r$ holds. We have

$$\begin{aligned} \mathbb{E}[\Delta|V(r)|\ |I(r), \mathcal{E}_r] &= -\frac{1}{|V(r)|} \sum_{v \in V(r)} (1 + D_v(r)) \\ &\leq -\frac{1}{|V(r)|} \left(|V(r)| - |B(r)|\right) \left(de^{-t} - df(t)\right) \\ &\leq -\left(1 - \frac{N^{1-\varepsilon/2}}{N^{1-\varepsilon/10}}\right) \left(de^{-t} - df(t)\right) \\ &\leq -d\left(e^{-t} - f(t)\right) + O\left(dN^{-\varepsilon/3}\right). \end{aligned} \tag{21}$$

We turn to bounding the one-step change in $N(e^{-t} + f(t))$ (the deterministic part of V^+). We use Taylor's theorem:

Theorem 12. *Let $g : \mathbb{R} \to \mathbb{R}$ be a function twice differentiable on the closed interval $[a, b]$. Then, there exists a number τ between a and b such that*

$$g(b) - g(a) = g'(a)(b-a) + \frac{g''(\tau)}{2}(b-a)^2. \tag{22}$$

Since $\Delta t(r) = \frac{d}{N}$, we have for some $\tau \in (t(r), t(r+1))$ that

$$\begin{aligned} \Delta N(e^{-t} \pm f(t)) &= N(-e^{-t} \pm f'(t)) \cdot \frac{d}{N} + \frac{1}{2} N(e^{-\tau} \pm f''(\tau)) \cdot \frac{d^2}{N^2} \\ &= d(-e^{-t} \pm f'(t)) + O\left(\frac{d^2}{N}\right). \end{aligned}$$

Thus we have

$$\begin{aligned} \mathbb{E}[\Delta V^+(r)\ |I(r), \mathcal{E}_r] &\leq -d\left(e^{-t} - f(t)\right) - d(-e^{-t} + f'(t)) + O\left(dN^{-\varepsilon/3}\right) \\ &= d(f(t) - f'(t)) + O\left(dN^{-\varepsilon/3}\right) \\ &= -\Omega\left(dN^{-\varepsilon/20}\right). \end{aligned} \tag{23}$$

Thus V^+ is a supermartingale. Now we will show that V^- is a submartingale. From (21) we have

$$\begin{aligned}
\mathbb{E}[\Delta|V(r)|\ |I(r),\mathcal{E}_r] &= -\frac{1}{|V(r)|}\sum_{v\in V(r)}(1+D_v(r)) \\
&\geq -1-\frac{1}{|V(r)|}\left[(|V(r)|-|B(r)|)\left(de^{-t}+df(t)\right)+|B(r)|d\right] \\
&\geq -\left(de^{-t}+df(t)\right)-\frac{N^{1-\varepsilon/2}}{N^{1-\varepsilon/10}}d-1 \\
&\geq -\left(de^{-t}+df(t)\right)+O\left(dN^{-\varepsilon/3}\right).
\end{aligned}$$

Thus, similar to (23), we have

$$\begin{aligned}
\mathbb{E}[\Delta V^-(r)\ |I(r),\mathcal{E}_r] &\geq -d\left(e^{-t}+f(t)\right)-d(-e^{-t}-f'(t))+O\left(dN^{-\varepsilon/3}\right) \\
&= \Omega\left(dN^{-\varepsilon/20}\right).
\end{aligned}$$

We use the following concentration inequality of Freedman [10].

Theorem 13 (Freedman)**.** *Suppose $Y_0, Y_1, \ldots$ is a supermartingale such that $\Delta Y_j \leq C$ for all j, and let $W_m = \sum_{r\leq m} \mathbb{V}ar[\Delta Y_r|\mathcal{F}_r]$. Then, for all m and positive real λ,*

$$\mathbb{P}(W_m \leq b \text{ and } Y_m - Y_0 \geq \lambda) \leq \exp\left(-\frac{\lambda^2}{2(b+C\lambda)}\right). \tag{24}$$

We apply Freedman's theorem to the supermartingale V^+. In particular we want to bound the probability that it is positive at step r_{end}. We have at step 0 that $V^+(0) = -Nf(0) = -N^{1-\varepsilon/20}$. If $V^+(r_{end})$ is positive then $V^+(r_{end}) - V^+(0) > N^{1-\varepsilon/20}$. Thus we will use $\lambda = N^{1-\varepsilon/20}$. We bound the one-step change to determine C. We have $|\Delta|V(r)|| \leq d$ and $|\Delta N(e^{-t}+f(t))| \leq 2d$ so we can use $C = 3d$. Note that we have

$$\mathbb{V}ar[\Delta V^{\pm}(r)|\mathcal{F}_r,\mathcal{E}_r] = \mathbb{V}ar[\Delta|V(r)|\ |\mathcal{F}_r,\mathcal{E}_r] \leq \mathbb{E}[\Delta|V(r)|^2|\mathcal{F}_r,\mathcal{E}_r] \leq d^2.$$

Thus we have $W_{r_{end}} \leq r_{end}\cdot d^2 = O(Nd\log N)$ and we can take $b = O(Nd\log N)$. Freedman's theorem gives us that the probability that $V^+(r_{end})$ is positive is at most

$$\exp\left(-\frac{\lambda^2}{2(b+C\lambda)}\right) = \exp\left(-\Omega\left(\frac{N^{2-\varepsilon/10}}{Nd\log N + dN^{1-\varepsilon/20}}\right)\right) = o(1).$$

Similarly one can apply Freedman's theorem to the submartingale $-V^-$ using the same values of λ, C, b to show that w.h.p. $-V^-(r_{end})$ is not positive. Thus we have w.h.p. that $\mathcal{E}_{r_{end}}$ does not fail due to condition (16).

3.3 Dynamic concentration of $D_v(r)$

Here we bound the probability that $\mathcal{E}_{r_{end}}$ fails due to condition (17). We define variables D_v^+ and D_v^- as follows.

$$D_v^{\pm} = D_v^{\pm}(r) := \begin{cases} |D_v(r)| - d(e^{-t} \pm f(t)) & \text{if } \mathcal{E}_{r-1} \text{ holds and } v \notin B(r), \\ D_v^{\pm}(r-1) & \text{otherwise.} \end{cases} \tag{25}$$

If $\mathcal{E}_{r_{end}}$ fails due to condition (17) then we either have that $D_v^+(r_{end}) > 0$ or $D_v^-(r_{end}) < 0$ for some v. Similarly to the last subsection, we will show these events are unlikely using Freedman's theorem. First we verify that D_v^+ is a supermartingale. We can assume $\mathcal{E}_r$ holds, since otherwise $\Delta D_v^+(r) = 0$. Of course if $v \in B(r+1)$ we will also have $\Delta D_v^+(r) = 0$ by definition (25). Thus we just need the expected one-step change in $|D_v(r)|$ in the case when $v \notin B(r+1)$. We have

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_{v \notin B(r+1)} \cdot \Delta|D_v(r)| \mid I(r), \mathcal{E}_r] &= -\sum_{u \in D_v(r)} \mathbb{P}(u \notin V(r+1) \text{ and } v \notin B(r+1)) \\
&\leq -\sum_{u \in D_v(r)} \frac{D_u(r) - dN^{-\varepsilon}}{|V(r)|} \\
&\leq -\frac{(de^{-t} - df(t) - dN^{-\varepsilon/10})^2}{Ne^{-t} + Nf(t)} \\
&= -\frac{d^2}{N} e^{-t} \cdot \frac{(1 - e^t f(t) - e^t N^{-\varepsilon/10})^2}{1 + e^t f(t)} \\
&= -\frac{d^2}{N} e^{-t} \cdot \left(1 - 3e^t f(t)\right) + O\left(d^2 N^{-1-\varepsilon/15}\right)
\end{aligned}$$

Meanwhile the one-step change in $d(e^{-t} + f(t))$ is

$$\frac{d^2}{N}(-e^{-t} + f'(t)) + O\left(d^3 N^{-2}\right)$$

by Taylor's theorem. Thus we have

$$\begin{aligned}
\mathbb{E}[\Delta D_v^+(r) \mid I(r), \mathcal{E}_r] &\leq -\frac{d^2}{N} e^{-t} \cdot \left(1 - 3e^t f(t)\right) - \frac{d^2}{N}(-e^{-t} + f'(t)) + O\left(d^2 N^{-1-\varepsilon/15}\right) \\
&= \frac{d^2}{N}(3f(t) - f'(t)) + O\left(d^2 N^{-1-\varepsilon/15}\right) \\
&= -\Omega\left(d^2 N^{-1-\varepsilon/20}\right)
\end{aligned}$$

Thus D_v^+ is a supermartingale, and similarly D_v^- is a submartingale. Now we apply Freedman's theorem. As in the last section, bounding the probability that $-D_v^-(r_{end}) > 0$ is entirely similar to bounding the probability that $D_v^+(r_{end}) > 0$, so from here we will only

show the work for the latter. Recall that $|\Delta D_v^+(r)| = 0$ unless $v \notin B(r+1)$, in which case we have $|\Delta|D_v(r)|| \leq 1 + dN^{-\varepsilon}$. We have

$$\begin{aligned}|\Delta D_v^+(r)| &\leq |\Delta|D_v(r)|| + |\Delta de^{-t}| + |\Delta df(t)| \\ &\leq 1 + dN^{-\varepsilon} + O(d^2/N) \\ &\leq 2dN^{-\varepsilon}.\end{aligned}$$

Thus we can use $C = 2dN^{-\varepsilon}$. If $D_v^+(r_{end}) > 0$ then $D_v^+(r_{end}) - D_v^+(0) > df(0) = dN^{-\varepsilon/20}$ and so we use $\lambda = dN^{-\varepsilon/20}$. Note that

$$\begin{aligned}\mathbb{V}ar[\Delta D_v^+(r)|\mathcal{F}_r, \mathcal{E}_r] &\leq \mathbb{E}[(\Delta|D_v^+(r)|)^2|\mathcal{F}_r, \mathcal{E}_r] \\ &\leq C\mathbb{E}[\ |\Delta D_v^+(r)|\ |\mathcal{F}_r, \mathcal{E}_r] \\ &\leq C\mathbb{E}[\ |\Delta|D_v(r)|\ |\ |\mathcal{F}_r, \mathcal{E}_r] + C|\Delta de^{-t}| + C|\Delta df(t)| \\ &= O(d^3N^{-1-\varepsilon}),\end{aligned}$$

where the second line follows since we always have $|\Delta D_v^+(r)| \leq C$, and the third line follows from the triangle inequality. Thus we can take $b = O(r_{end} \cdot d^3N^{-1-\varepsilon}) = O(d^2N^{-\varepsilon/2})$. Freedman's theorem gives us a failure probability of at most

$$\exp\left\{-\frac{\lambda^2}{2(b+C\lambda)}\right\} = \exp\left\{-\Omega\left(\frac{d^2N^{-\varepsilon/10}}{d^2N^{-\varepsilon/2} + dN^{-\varepsilon}\cdot dN^{-\varepsilon/20}}\right)\right\} = o\left(\frac{1}{N}\right).$$

This probability is small enough to beat a union bound over N choices of v. Thus we have w.h.p. that $\mathcal{E}_{r_{end}}$ does not fail due to condition (17).

3.4 The upper bound on C_v

We now bound the probability that $\mathcal{E}_{r_{end}}$ fails due to condition (18). We define variables C_v^+ as follows.

$$C_v^+ = C_v^+(r) := \begin{cases}|C_v(r)| - d^2N^{-1-\varepsilon/2}r & \text{if } \mathcal{E}_{r-1} \text{ holds,} \\ C_v^+(r-1) & \text{otherwise.}\end{cases} \tag{26}$$

We show that C_v^+ is a supermartingale. We have

$$\mathbb{E}[\Delta|C_v(r)|\ |I(r), \mathcal{E}_r] \leq \frac{d\cdot dN^{-\varepsilon}}{|V(r)|} \leq \frac{d^2N^{-\varepsilon}}{N^{1-\varepsilon/10}} \leq d^2N^{-1-\varepsilon/2}$$

and so $\mathbb{E}[\Delta|C_v^+(r)|\ |I(r), \mathcal{E}_r] \leq d^2N^{-1-\varepsilon/2} - d^2N^{-1-\varepsilon/2} = 0$. We apply Freedman's theorem. We have $\Delta C_v^+(r) \leq \Delta C_v(r) \leq dN^{-\varepsilon}$ and so we use $C = dN^{-\varepsilon}$. Note that we have

$$\begin{aligned}\mathbb{V}ar[\Delta C_v^+(r)|\mathcal{F}_r, \mathcal{E}_r] = \mathbb{V}ar[\Delta C_v(r)|\mathcal{F}_r, \mathcal{E}_r] &\leq \mathbb{E}[\Delta C_v(r)^2|\mathcal{F}_r, \mathcal{E}_r] \\ &\leq dN^{-\varepsilon}\mathbb{E}[|\Delta C_v(r)||\mathcal{F}_r, \mathcal{E}_r] \\ &\leq dN^{-\varepsilon}\cdot d^2N^{-1-\varepsilon/2} \\ &\leq d^3N^{-1-3\varepsilon/2}\end{aligned}$$

Thus we have $W_{r_{end}} \le r_{end} \cdot d^3 N^{-1-3\varepsilon/2} = O(d^2 N^{-\varepsilon})$ and we can take $b = O(d^2 N^{-\varepsilon})$. Using $\lambda = dN^{-\varepsilon/4}$, Freedman's theorem gives us that the probability we ever have $C_v^+(r) > \lambda$ is at most

$$\exp\left(-\frac{\lambda^2}{2(b+C\lambda)}\right) = \exp\left(-\Omega\left(\frac{d^2N^{-\varepsilon/2}}{d^2N^{-\varepsilon} + dN^{-\varepsilon}\cdot dN^{-\varepsilon/4}}\right)\right) = o(1/N),$$

which is small enough to beat the union bound over the N choices for v. Thus w.h.p. we have for each v that $C_v^+(r) \le \lambda$ and so

$$C_v(r) \le d^2N^{-1-\varepsilon/2}r_{end} + \lambda < dN^{-\varepsilon/10}.$$

Thus w.h.p. $\mathcal{E}_{r_{end}}$ does not fail due to condition (18). This completes our proof of Theorem 4 for the case when $k = \Theta(n)$.

4 Proof of Theorem 4 when $k = o(n)$

In this section we assume $\zeta^{-1}n^{1/2}\log^{1/2} n \le k \le \zeta n$ for a sufficiently small constant $\zeta > 0$. We use the same notation as in Section 3, so $k = cn$.

For say $|x| \le 1/2$, we have $g(1-x) = -x + \frac{x^2}{2} + O(x^3)$ and so we can write

$$\begin{aligned} p_N(c) &= -g(c) + c - \frac{1}{2}c^2 + O(c^3), \\ p_d(c) &= -g(c) + c - \frac{3}{2}c^2 + O(c^3). \\ p_N(c) - p_d(c) &= c^2 + O(c^3). \end{aligned}$$

Thus, we have that

$$d = Ne^{-(c^2+O(c^3))n}.$$

With p_1, p_2 as in Section 3, we now have

$$\begin{aligned} p_1(c,\delta) - p_d(c) &= -c\delta + O(c^3). \\ p_2(c,\delta) &= \delta\log\left(\frac{c}{\delta^2}\right) + 2\delta - \frac{\delta^2}{c} + O(\delta^2). \end{aligned}$$

Let $\delta = 0.9c$, and

$$\begin{aligned} \gamma &:= \frac{d}{N} = \exp\{-c^2n + O(c^3n)\} \\ \alpha &:= d\exp\left\{-c\delta n + O(c^3n)\right\} = d\exp\{-0.9c^2n + O(c^3n)\} \\ \beta &:= \exp\left\{\left(\delta\log\left(\frac{c}{\delta^2}\right) + 2\delta - \frac{\delta^2}{c}\right)n + O(\delta^2 n)\right\} = \exp\left\{-0.9(c\log c)n + O(cn)\right\} \end{aligned}$$

Arguing as in Section 3, we see that for each vertex v, there are at most α vertices u whose codegree with v is at least β.

We assume that

$$\gamma \leq \log^{-100} N$$

which holds when c is at least some large constant times $n^{-1/2}\log^{1/2} n$ (i.e. for ζ small enough).

4.1 The good event

Let $t = t(r) = \gamma r$. Define the error function

$$f(t) := \gamma^{0.1} e^{10t}. \tag{27}$$

Note that this is different from our error function $f(t)$ from Section 3.1. Let the *good event* $\mathcal{E}_r$ be the event that the following inequalities (28), (29) and (30) all hold for all $r' \leq r$:

$$\left||V(r')| - Ne^{-t'}\right| \leq Nf(t') \tag{28}$$

$$\left||D_v(r')| - de^{-t'}\right| \leq df(t') \text{ for all } v \in V(r') \setminus B(r') \tag{29}$$

$$|C_v(r')| \leq \gamma^{-0.1}\alpha \text{ for all } v \in V(r'). \tag{30}$$

We will show that w.h.p. the good event $\mathcal{E}_{r_{end}}$ holds where

$$r_{end} := 0.001\gamma^{-1}\log\gamma^{-1} = \Omega\left(\frac{N}{d}\log\frac{N}{d}\right).$$

In particular this will imply that w.h.p. the process lasts to at least step r_{end}, proving the lower bound in Theorem 4 for the case $c_n = o(1)$. Indeed, we have $e^{10t} \leq \gamma^{-0.01}$ and so $e^{-t} \geq \gamma^{0.001}$ and $f(t) \leq \gamma^{0.09}$ for all $r \leq r_{end}$. Now lines (28) and (29) imply that for all $r \leq r_{end}$ we have

$$|V(r)| = (1+o(1))Ne^{-t}, \qquad |D_v(r)| = (1+o(1))de^{-t} \text{ for all } v \in V(r) \setminus B(r).$$

In particular $|V(r_{end})|$ is positive and so the process lasts to step r_{end}. The upper bound in Theorem 4 comes from the bound on $|V(r_{end})|$ implied by (28).

We will sometimes use the bounds (which hold for all $r \leq r_{end}$ in the good event $\mathcal{E}_r$)

$$|V(r)| \geq \frac{1}{2}Ne^{-\gamma r_{end}} = \frac{1}{2}\gamma^{0.001}N, \qquad |B(r)| \leq \alpha r_{end} = 0.001\alpha\gamma^{-1}\log\gamma^{-1}. \tag{31}$$

4.2 Dynamic concentration of $V(r)$

Here we bound the probability that $\mathcal{E}_{r_{end}}$ fails due to condition (28). We define variables V^+ and V^- exactly as in (20) (but now $f(t)$ is as in (27)). We now establish that V^+ is a supermartingale.

First we show $\mathbb{E}[\Delta|V^+(r)|\ |I(r)] \le 0$. If $\mathcal{E}_r$ fails then $\Delta V^+(r) = 0$ by definition, so we assume $\mathcal{E}_r$ holds. Also, as before we only need the expected one-step change in $|D_v(r)|$ in the case when $v \notin B(r+1)$. We have

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_{v\notin B(r+1)} \cdot \Delta|D_v(r)|\ |I(r), \mathcal{E}_r] &= -\frac{1}{|V(r)|}\sum_{v\in V(r)}(1 + D_v(r)) \\
&\le -\frac{1}{|V(r)|}\left(|V(r)| - |B(r)|\right)\left(de^{-t} - df(t)\right) \\
&\le -\left(1 - \frac{0.001\alpha\gamma^{-1}\log\gamma^{-1}}{\frac{1}{2}\gamma^{0.001}N}\right)\left(de^{-t} - df(t)\right) \\
&\le -d\left(e^{-t} - f(t)\right) + O\left(\alpha\gamma^{-0.001}\log\gamma^{-1}\right).
\end{aligned}$$

We turn to bounding the one-step change in $N(e^{-t} + f(t))$. Since $\Delta t(r) = \frac{d}{N}$, we have for some $\tau \in (t(r), t(r+1))$ that

$$\begin{aligned}
\Delta N(e^{-t} + f(t)) &= N(-e^{-t} + f'(t)) \cdot \frac{d}{N} + \frac{1}{2}N(e^{-\tau} + f''(\tau)) \cdot \frac{d^2}{N^2} \\
&= d(-e^{-t} + f'(t)) + O\left(d^2N^{-1}\right).
\end{aligned}$$

Thus we have

$$\begin{aligned}
\mathbb{E}[\Delta V^+(r)\ |I(r), \mathcal{E}_r] &\le -d\left(e^{-t} - f(t)\right) - d(-e^{-t} + f'(t)) + O\left(\alpha\gamma^{-0.001}\log\gamma^{-1}\right) \qquad (32) \\
&= d(f(t) - f'(t)) + O\left(\alpha\gamma^{-0.001}\log\gamma^{-1}\right) \\
&= -\Omega\left(d\gamma^{0.1}\right), \qquad \text{since } \alpha/d \le \gamma^{0.8}.
\end{aligned}$$

Thus V^+ is a supermartingale. To see that V^- is a submartingale, we replace (32) by

$$\mathbb{E}[\Delta V^-(r)\ |I(r), \mathcal{E}_r] \ge -d\left(e^{-t} - f(t)\right) - d(-e^{-t} - f'(t)) + O\left(\alpha\gamma^{-0.001}\log\gamma^{-1}\right) = \Omega\left(d\gamma^{0.1}\right).$$

We apply Freedman's theorem to the supermartingales V^+ and $-V^-$ to show w.h.p. neither of them becomes positive. The calculations for $-V^-$ is entirely similar to V^+ so we only show the latter. If $V^+(r_{end}) > 0$ then $V^+(r_{end}) - V^+(0) > Nf(0) = N\gamma^{0.1}$. Thus we use $\lambda = N\gamma^{0.1}$. We bound the one-step change to determine C. We have $|\Delta|V(r)|| \le d$ and $|\Delta N(e^{-t} + f(t))| \le 2d$ so we can use $C = 3d$. Note that we have

$$\mathbb{V}ar[\Delta V^+(r)|\mathcal{F}_r, \mathcal{E}_r] = \mathbb{V}ar[\Delta|V(r)|\ |\mathcal{F}_r, \mathcal{E}_r] \le \mathbb{E}[\Delta|V(r)|^2|\mathcal{F}_r, \mathcal{E}_r] \le d^2.$$

Thus we have $W_{r_{end}} \le d^2 r_{end}$ and we can take $b = d^2 r_{end} = O\left(d^2\gamma^{-1}\log\gamma^{-1}\right)$. Freedman's theorem gives us a failure probability of at most

$$\exp\left\{-\frac{\lambda^2}{2(b + C\lambda)}\right\} = \exp\left\{-\Omega\left\{\frac{N^2\gamma^{0.2}}{d^2\gamma^{-1}\log\gamma^{-1} + Nd\gamma^{0.1}}\right\}\right\} = o(1).$$

Thus we have w.h.p. that $\mathcal{E}_{r_{end}}$ does not fail due to condition (28).

4.3 Dynamic concentration of $D_v(r)$

Here we bound the probability that $\mathcal{E}_{r_{end}}$ fails due to condition (29). We define variables D_v^+ and D_v^- as in (25). We verify that D_v^+ is a supermartingale. As before we can assume $\mathcal{E}_r$ holds. We have

$$\begin{aligned}
\mathbb{E}[\Delta|D_v(r)|\ |I(r),\mathcal{E}_r] &\le -\sum_{u\in D_v(r)} \mathbb{P}(u\notin V(r+1) \text{ and } v\notin B(r+1)) \\
&\le -\sum_{u\in D_v(r)} \frac{D_u(r)-\alpha}{|V(r)|} \\
&\le -\frac{(de^{-t}-df(t)-\gamma^{-0.1}\alpha)^2}{Ne^{-t}+Nf(t)} \\
&= -\frac{d^2}{N}e^{-t}\cdot\frac{(1-e^tf(t)-e^t\gamma^{-0.1}\alpha d^{-1})^2}{1+e^tf(t)} \\
&= -\frac{d^2}{N}e^{-t}\cdot\left(1-3e^tf(t)+O\left(e^t\gamma^{-0.1}\alpha d^{-1}\right)\right) \\
&= -\frac{d^2}{N}e^{-t}\cdot\left(1-3e^tf(t)\right)+O\left(\gamma^{0.9}\alpha\right)
\end{aligned}$$

Meanwhile the one-step change in $d(e^{-t}+f(t))$ is

$$\frac{d^2}{N}(-e^{-t}+f'(t))+O\left(d^3N^{-2}\right)$$

by Taylor's theorem. Thus we have

$$\begin{aligned}
\mathbb{E}[\Delta D_v^+(r)\ |I(r),\mathcal{E}_r] &\le -\frac{d^2}{N}e^{-t}\cdot\left(1-3e^tf(t)\right)-\frac{d^2}{N}(-e^{-t}+f'(t))+O\left(\gamma^{0.9}\alpha\right) \\
&= \frac{d^2}{N}(3f(t)-f'(t))+O\left(\gamma^{0.9}\alpha\right) \\
&= -\Omega\left(\gamma^{1.1}d\right)
\end{aligned}$$

Thus D_v^+ is a supermartingale, and similarly D_v^- is a submartingale.

Now we apply Freedman's theorem. We have

$$\begin{aligned}
|\Delta D_v^{\pm}(r)| &\le |\Delta D_v(r)|+|\Delta de^{-t}|+|\Delta df(t)| \\
&\le \beta+O(d\gamma) \\
&\le O(d\gamma), \qquad (33)
\end{aligned}$$

where on the second line we used that $\Delta e^{-t}=O(\gamma e^{-t})=O(\gamma)$ and $\Delta f(t)=O(\gamma f'(t))=O(\gamma^{1.1}e^{10t})=O(\gamma^{1.09})$. Thus we can use $C=O(d\gamma)$. If $D_v^+(r_{end})>0$ then $D_v^+(r_{end})-D_v^+(0)>df(0)=d\gamma^{0.1}$ and so we use $\lambda=d\gamma^{0.1}$. Note that by (33) we have

$$\begin{aligned}
\mathbb{V}ar[\Delta D_v^+(r)|\mathcal{F}_r,\mathcal{E}_r]=\mathbb{V}ar[\Delta|D_v(r)|\ |\mathcal{F}_r,\mathcal{E}_r] &\le \mathbb{E}[(\Delta|D_v(r)|)^2|\mathcal{F}_r,\mathcal{E}_r] \\
&= O(d^2\gamma^2).
\end{aligned}$$

Thus we can take $b = O(d^2\gamma^2 r_{end}) = O(d^2\gamma \log \gamma^{-1})$. Freedman's theorem gives us a failure probability of at most

$$\exp\left(-\frac{\lambda^2}{2(b+C\lambda)}\right) = \exp\left(-\Omega\left(\frac{d^2\gamma^{0.2}}{d^2\gamma\log\gamma^{-1} + d^2\gamma^{1.1}}\right)\right) = o(1/N).$$

This probability is small enough to beat a union bound over N choices of v. Thus we have w.h.p. that $\mathcal{E}_{r_{end}}$ does not fail due to condition (29).

4.4 The upper bound on C_v

Here we bound the probability that $\mathcal{E}_{r_{end}}$ fails due to condition (30). We define variables C_v^+ analogously to (26):

$$C_v^+ = C_v^+(r) := \begin{cases} |C_v(r)| - 2\alpha\gamma^{0.999}r & \text{if } \mathcal{E}_{r-1} \text{ holds,} \\ C_v^+(r-1) & \text{otherwise.} \end{cases}$$

We show that C_v^+ is a supermartingale. We have

$$\mathbb{E}[\Delta|C_v(r)|\ |I(r), \mathcal{E}_r] \le \frac{d\alpha}{|V(r)|} \le 2d\alpha\gamma^{-0.001}N^{-1} = 2\alpha\gamma^{0.999}$$

and so $\mathbb{E}[\Delta|C_v^+(r)|\ |I(r), \mathcal{E}_r] \le 0$. We apply Freedman's theorem. We have $\Delta C_v^+(r) \le \Delta C_v(r) \le \alpha$ and so we use $C = \alpha$. Note that we have

$$\begin{aligned} \mathbb{V}ar[\Delta C_v^+(r)|\mathcal{F}_r, \mathcal{E}_r] = \mathbb{V}ar[\Delta C_v(r)|\mathcal{F}_r, \mathcal{E}_r] &\le \mathbb{E}[\Delta C_v(r)^2|\mathcal{F}_r, \mathcal{E}_r] \\ &\le \alpha\mathbb{E}[|\Delta C_v(r)||\mathcal{F}_r, \mathcal{E}_r] \\ &\le 2\alpha^2\gamma^{0.999} \end{aligned}$$

Thus we have $W_{r_{end}} = O(\alpha^2\gamma^{-.001}\log\gamma^{-1})$ and we can take $b = O(\alpha^2\gamma^{-.001}\log\gamma^{-1})$. Freedman's theorem gives us that the probability we ever have $C_v^+(r) > \lambda := \frac{1}{2}\gamma^{-0.1}\alpha$ is at most

$$\exp\left(-\frac{\lambda^2}{2(b+C\lambda)}\right) = \exp\left(-\Omega\left(\frac{\gamma^{-0.2}\alpha^2}{\alpha^2\gamma^{-.001}\log\gamma^{-1} + \gamma^{-0.1}\alpha^2}\right)\right) = o(1/N).$$

Otherwise we have $C_v^+(r) \le \frac{1}{2}\gamma^{-0.1}\alpha$ and so

$$\begin{aligned} C_v(r) &\le 2\alpha\gamma^{0.999}r_{end} + \frac{1}{2}\gamma^{-0.1}\alpha \\ &= O(\alpha\gamma^{-.001}\log\gamma^{-1}) + \frac{1}{2}\gamma^{-0.1}\alpha \\ &\le \gamma^{-0.1}\alpha. \end{aligned}$$

Thus w.h.p. $\mathcal{E}_{r_{end}}$ does not fail due to condition (30). This completes the proof of Theorem 4.

5 Summary

We have proved some new results about the random intersecting family process. There is still much to do. For example, can we say more about the structure of $\mathcal{I}_k$ in Theorem 2. Second, as far as Theorem 4, we say little about the structure of the final family and there is a large gap between the upper and lower bounds on the family size.

References

[1] J. Balogh, S. Das, Delcourt, Liu and M. Sharifzadeh, Intersecting families of discrete structures are typically trivial., *Journal of Combinatorial Theory, Series A* 132 (2015) 224-245.

[2] J. Balogh, S. Das, L. Liu, Sharifzadeh and T. Tran, Structure and supersaturation for intersecting families, *Electronic Journal of Combinatorics* 26(2), 2019.

[3] J. Balogh, R. Garcia, L. Li and A. Wagner, Intersecting families of sets are typically trivial.

[4] P. Bennett and T. Bohman, A note on the random greedy independent set algorithm, *Random Structures and Algorithms* 49 (2016) 479–502.

[5] P. Bennett and A. Dudek, A gentle introduction to the differential equation method and dynamic concentration, *Discrete Mathematics* 345 (2022).

[6] T. Bohman, C. Cooper, A.M. Frieze, R. Martin and M. Ruszinkó, On Randomly Generated Intersecting Hypergraphs, *Electronic Journal of Combinatorics* 10 (2003).

[7] T. Bohman, C. Cooper, A.M. Frieze, R. Martin, M. Ruszinkó and C. Smyth, On Randomly Generated Intersecting Hypergraphs II, *Random Structures and Algorithms* 30 (2007) 17-34.

[8] I. Dinur and E. Friedgut, Intersecting families are essentially contained in juntas, *Combintaorics, Probability and Computing* 18 (2009) 107–122.

[9] P. Erdős, C. Ko, R. Rado, Intersection theorems for systems of finite sets, *Quarterly Journal of Mathematics, Oxford Series (2)* 12 (1961) 313-320.

[10] D. Freedman, On tail probabilities for martingales, *Annals of Probability (3)* (1975) 100-118.

[11] A. Hilton and E. Milner, Some intersection theorems for systems of finite sets, *Quarterly Journal of Mathematics, Oxford Series (2)* 18 (1967) 369–384.

[12] P. Frankl and A. Kupavskii, Counting intersecting and pairs of cross-intersecting families, *Combinatorics, Probability and Computing* 27 (2018) 60–68.

[13] J. Lauer and N. Wormald, Large independent sets in regular graphs of large girth, *Journal of Combinatorial Theory Series B* 97 (2007) 999–1009.

[14] B. Patkos, On randomly generated non-trivially intersecting hypergraphs, *Electronic Journal of Combinatorics* 10 (2010).

[15] N. Wormald, Differential equations for random processes and random graphs, *Annals of Applied Probability* 5 (1995) 1217–1235.

A Appendix: Describing $\mathcal{S}$ as a random matching process

In the proof of Theorem 2 we described the process of generating $\mathcal{S}$ and r_0 which determine $\mathcal{I}^*$ and $\mathcal{I}^{**}$. Here we isolate just this process away from the broader context to set up an open problem that would have to be solved in order to fully describe the distribution of $\mathcal{I}_k$ when $k = cn^{1/3}$. Recall that at the beginning of the random intersecting process, before vertices of degree three appear, $\mathcal{H}_r$ is a simple hypergraph in which every pair of edges intersects in a unique vertex. If we represent the hyperedges as vertices then an independent set of vertices of degree two in $\mathcal{H}_r$ corresponds to a matching in the complete graph K_r. And moreover, once r_0 is determined $\mathcal{S}$ corresponds to an intersecting family $\mathcal{M}$ of matchings on K_{r_0-1}. For $k = c_n n^{1/3}$, and $c_n \to c$ we can describe how $\mathcal{S}$ is generated by Procedure 1 below. For a collection of finite sets $\mathcal{N}$, $\text{rand}_c(\mathcal{N})$ refers to sampling a set S from $\mathcal{N}$ with probability proportional to $c^{|S|}$.

By the proof of Theorem 2 we can see that $\mathcal{S}$ is generated by Procedure 1 with input equal to c^{-3} for $k = c_n n^{1/3}$, $c_n \to c$. The situation that Procedure 1 outputs the single edge on K_2 corresponds to the case that the system is trivial. The case that Procedure 1 outputs a single-edge matching on K_3 corresponds to the Hilton–Milner system described by [14].

For K_t, $t \leq 5$ there is only one combinatorial type for a maximal family of intersecting matching $\mathcal{M}$ on K_t, namely all matchings that contain some fixed edge (i.e. a star). For $t = 6$, however, there are two: All matchings that contain some fixed edge and all matchings that contain at least two out of three edges of some perfect matching. For $t = r_0 - 1 \leq 5$ we can therefore determine exactly what $\mathcal{S}$ will look like. When $t = 6$ and $c = 1$, a routine calculation shows that conditioned stopping at $t = 6$, there is a 123/128 chance of getting a star and a 5/128 chance of getting all matchings that contain at least two out of three edges of a fixed perfect matching. As t gets larger these families of matchings can become arbitrarily complicated, so a full description of the distribution on $\mathcal{S}$ would be quite difficult to work out even though we do have a generating function in part (d) of Theorem 2 that captures the distribution on t.

Procedure 1: Random matching procedure

Input : $c > 0$

Output: A random number n and a family $\mathcal{M}$ of pairwise intersecting matchings of the complete graph K_n

$n \leftarrow 1$;
$G \leftarrow K_n$;
$\mathcal{N} \leftarrow$ All matchings on G;
$\mathcal{M} \leftarrow \emptyset$;

while $\mathcal{M} = \emptyset$ **do**
    $M \leftarrow \mathrm{rand}_c(\mathcal{N})$;
    if $M = \emptyset$ **then**
        $n \leftarrow n + 1$;
        $G \leftarrow K_n$;
        $\mathcal{N} \leftarrow$ All matchings on G;
    else
        Add M to $\mathcal{M}$;
        $\mathcal{N} \leftarrow$ All matchings on G that intersect M;

while $\mathcal{M}$ *is not a maximal intersecting family of matchings on* K_n **do**
    $M \leftarrow \mathrm{rand}_c(\mathcal{N})$;
    if $M \notin \mathcal{M}$ **then**
        Add M to $\mathcal{M}$;
        $\mathcal{N} \leftarrow$ All matchings on G that intersect every matching in $\mathcal{M}$;

return $\mathcal{M}$ on K_n;