A natural barrier in random greedy hypergraph matching

Patrick Bennett* Tom Bohman†

Abstract

Let $r \geq 2$ be a fixed constant and let $\mathcal{H}$ be an r-uniform, D-regular hypergraph on N vertices. Assume further that $D \to \infty$ as $N \to \infty$ and that degrees of pairs of vertices in $\mathcal{H}$ are at most L where $L = D/(\log N)^{\omega(1)}$. We consider the random greedy algorithm for forming a matching in $\mathcal{H}$. We choose a matching at random by iteratively choosing edges uniformly at random to be in the matching and deleting all edges that share at least one vertex with a chosen edge before moving on to the next choice. This process terminates when there are no edges remaining in the graph. We show that with high probability the proportion of vertices of $\mathcal{H}$ that are not saturated by the final matching is at most $(L/D)^{\frac{1}{2(r-1)}+o(1)}$. This point is a natural barrier in the analysis of the random greedy hypergraph matching process.

1 Introduction

Let $r \geq 2$ be a fixed constant and let $\mathcal{H}$ be an r-uniform, D-regular hypergraph on vertex set V where $|V| = N$ and $D \to \infty$ as $N \to \infty$. We study the evolution of the random greedy matching algorithm on $\mathcal{H}$. This process forms a matching (i.e. a collection of pairwise disjoint edges) in $\mathcal{H}$ by making a series of random choices. We begin with $\mathcal{M}(0) = \emptyset$, $\mathcal{H}(0) = \mathcal{H}$ and $V(0) = V$. In iteration i an edge E_i is chosen uniformly at random from $\mathcal{H}(i-1)$ and added to $\mathcal{M}(i-1)$ to form the matching $\mathcal{M}(i)$. We then form $\mathcal{H}(i)$ by setting $V(i) = V(i-1) \setminus E_i$ and deleting from $\mathcal{H}(i-1)$ all edges that intersect E_i. The process proceeds until the step M where $\mathcal{H}(M)$ is empty. We are interested in the likely value of M; that is, we are interested in the number of edges in the matching produced by the random greedy process.

The random greedy packing algorithm for producing a partial Steiner system is an important special case of this process. Let $1 < \ell < k$ be fixed integers. Define $\mathcal{H}_{\ell,k}$ to be the hypergraph on vertex set $\binom{[n]}{\ell}$ with edge set consisting of all sets of the form $\binom{A}{\ell}$ where $A \in \binom{[n]}{k}$. Note that a matching in $\mathcal{H}_{\ell,k}$ corresponds to a

*Mathematics Department, Western Michigan University, Kalamazoo, MI 49008, USA. Email: patrick.bennett@wmich.edu. Research supported in part by NSF grant DMS-1001638 and Simons Foundation grant #426894.

†Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Email: tbohman@math.cmu.edu. Research supported in part by NSF grants DMS-1001638 and DMS-1100215.

collection of k-element subsets of $[n]$ with the property that the intersect of any pair of sets in the collection has cardinality less than ℓ; that is, a matching in $\mathcal{H}_{\ell,k}$ gives a partial (n,k,ℓ)-Steiner system. The random greedy matching algorithm applied to $\mathcal{H}_{\ell,k}$ is also known as random greedy packing. This process is related to the celebrated Rödl nibble [10], which is a semi-random variation on random greedy packing. The Rödl nibble was introduced in the solution of the Erdős and Hanani conjecture [6], which states that for every fixed ℓ, k there is a matching in $\mathcal{H}_{\ell,k}$ that saturates $(1-o(1))\binom{n}{\ell}$ vertices.

In this paper we study the general random greedy matching algorithm by establishing dynamic concentration of the number of edges and the vertex degrees in the remaining hypergraph $\mathcal{H}(i)$. Let $Q(i)$ be the number of edges in $\mathcal{H}(i)$ and let $d_v(i)$ be the degree of vertex v in $\mathcal{H}(i)$. We aim to show that $Q(i)$ and $d_v(i)$, appropriately scaled, are tightly concentrated around expected trajectories that we express as smooth functions on the reals. In order to describe the trajectories we introduce a continuous time t which we relate to the steps of the process by setting

$$t = t(i) = \frac{i}{N}.$$

Our study is guided by the following probabilistic intuition: we suspect that $\mathcal{H}(i)$ resembles a subhypergraph of $\mathcal{H}$ chosen uniformly at random from the collection of all subhypergraphs induced by $N - ir$ vertices. So we anticipate that $\mathcal{H}(i)$ resembles a subhypergraph of $\mathcal{H}$ induced by a random subset of the vertices where each vertex is included independently with probability

$$p = 1 - ir/N = 1 - rt.$$

(Note that this probability can be viewed as either a function of either i or t; we pass between these interpretations without comment.) It follows from this assumption that the probability an edge $E \in \mathcal{H}$ is in $\mathcal{H}(i)$ should be about p^r, and therefore we ought to have

$$Q(i) \approx |\mathcal{H}|p^r = NDp^r/r. \tag{1}$$

Furthermore, if a vertex v is not saturated by $\mathcal{M}(i)$ then we should have

$$d_v(i) \approx Dp^{r-1}. \tag{2}$$

Our main result (see Theorem 2.1 below) is that estimates (1) and (2) hold for most of the evolution of the process. This is a generalization of a result of Bohman, Frieze and Lubetzky [3], who proved an analogous result for the special case of $\mathcal{H}_{2,3}$.

In order to discuss our main result in more detail, we define the random variable

$$X = X(\mathcal{H}) := 1 - Mr/N$$

where M is the number of steps before the random greedy matching algorithm on $\mathcal{H}$ terminates. In other words, X is the proportion of vertices left unsaturated by the matching produced by the random greedy algorithm. The following bound is a Corollary of Theorem 2.1.

Theorem 1.1. *Let $r \geq 2$ and $\mathcal{H}$ be an r-uniform, D-regular hypergraph on N vertices. If the maximum degree L of a pair of vertices in $\mathcal{H}$ satisfies $L = D/(\log N)^{\omega(1)}$ and $X(\mathcal{H})$ is the proportion of vertices that are not saturated by the matching produced by the random greedy algorithm then with high probability we have*

$$X(\mathcal{H}) \leq \left(\frac{L}{D}\right)^{\frac{1}{2(r-1)}+o(1)}.$$

Previous analyses of the random greedy matching algorithm due to Spencer [12] and, independently, Rödl and Thoma [10] showed that if $L = o(D)$ then we have $X(\mathcal{H}) = o(1)$ with high probability. Note that this result applied to the hypergraph $\mathcal{H}_{\ell,k}$ gives an alternate proof of the Erdős–Hanani conjecture. Wormald [15] applied the differential equations method for random graph processes to show that if $\mathcal{H}$ is an r-uniform, D-regular hypergraph on N vertices such that $D = o(N)$ but $D \to \infty$ sufficiently quickly as $N \to \infty$ then $X(\mathcal{H}) < D^{-\frac{1}{9r(r-1)+3}+o(1)}$ with high probability.

Theorem 1.1 takes the analysis of random greedy matching up to a natural barrier. To describe this barrier we assume estimates (1) and (2) hold. For a fixed vertex v let L_v be the set of vertices u such that the degree of $\{u, v\}$ in $\mathcal{H}$ is L. Note that $|L_v|$ can be as large as $(r-1)D/L$. Now early in the process (when $p = 1/2$, say) the expected number of vertices in L_v that are not saturated by $\mathcal{M}$ can be as large pD/L and thus can have variation as large as $\sqrt{D/L}$, roughly speaking. This yields variations in vertex degrees that are as large as $\sqrt{D/L} \cdot L = \sqrt{DL}$. If these early variations in vertex degree persist then at the point when $Dp^{r-1} = \sqrt{DL}$ these variations will be as large as the expected degree itself. So, if these variations indeed persist then when we reach this point vertex degrees could be zero even though the expected vertex degree is large. Note that this is point where Theorem 2.1 no longer holds. One would expect that in order to prove better bounds one would have to show that the variations in vertex degree *decrease* as the process evolves.

But where do we expect the random greedy matching algorithm to finally terminate? If we assume that estimates (1) and (2) hold all the way to termination then when $NDp^r = Np$ the number of unsaturated vertices should be roughly the same as the number of remaining edges. At this stage a positive proportion of the unsaturated vertices should be in no remaining edges; these vertices would remain unsaturated to termination. Thus, it is natural to guess that random greedy matching terminates when the proportion of unsaturated vertices is roughly $D^{-1/(r-1)}$. (We note in passing that this line of reasoning is suspect if $L > D^{1-\frac{1}{r-1}}$. In this case, one suspects that we will reach a point where degrees of pairs of vertices in $\mathcal{H}(i)$ are larger than degrees of individual vertices before the supposed termination point.) In the context of random greedy packing, this line of reasoning leads to the following conjecture.

Conjecture 1.2 (folklore)**.** *Let $1 < \ell < k$ be fixed. With high probability*

$$X(\mathcal{H}_{\ell,k}) = n^{-\frac{k-\ell}{\binom{k}{\ell}-1}+o(1)}.$$

The $\ell = 2, k = 3$ case of this conjecture was recently proved by Bohman, Frieze and Lubetzky [4] who establish estimates for vertex degrees in $\mathcal{H}_{2,3}(i)$ with error bounds that decrease as the process evolves. These self-correcting estimates are proved using the critical interval method that is featured in this paper and was introduced in [3]. It should be noted that the sharp result given in [4] requires a large, carefully selected ensemble of random variables.

The related problem of proving the existence of a large matching in an r-uniform, D-regular hypergraph $\mathcal{H}$ has been widely studied (see [9] [1] [8]). The best known results are due to Vu [14] who used a semi-random (i.e. Rödl nibble type) method to show that there exists a matching in $\mathcal{H}$ that saturates all but at most

$$\left(\frac{L}{D}\right)^{\frac{1}{r-1}+o(1)}$$

vertices where L is the maximum degree of pairs of vertices in $\mathcal{H}$. Vu obtained stronger results when one adds degree assumptions for larger sets of vertices.

The remainder of this paper is organized as follows. In the next Section we give a precise statement of our dynamic concentration result. The proof follows in Section 3. This proof uses the critical interval method introduced by Bohman, Frieze and Lubetzky in [3], where they prove Theorem 1.1 for the special case $\mathcal{H}_{2,3}$. In this note we show that the techniques introduced in [3] are robust enough to handle the general case (with the introduction of some delicate calculations necessitated by the large pairwise degrees).

2 Dynamic Concentration

Throughout this section we assume that $\mathcal{H}$ is an r-uniform, D-regular hypergraph on N vertices where r is a fixed constant and $D \to \infty$ as $N \to \infty$. We also assume that the maximum degree L of a pair of vertices in $\mathcal{H}$ satisfies $L = o(D/\log^5 N)$.

In order to make the estimates (1) and (2) precise we introduce error bounds for Q and d_v. Define

$$\begin{aligned} e_q &= 90r^2 N L p^{2-r} \log N \,(1 - r\log p)^2 \\ e_d &= \sqrt{6rLD\log N}\,(1 - r\log p) \end{aligned}$$

Further define the stopping time T to be the first step i such that

$$\begin{aligned} &\left| Q(i) - \frac{ND}{r}p^r \right| > e_q, \text{ or} \\ &|d_v(i) - Dp^{r-1}| > e_d \text{ for some } v \in V(i) \end{aligned}$$

Theorem 2.1. *With high probability we have*

$$N - Tr = O\left(N \cdot \left(\frac{L}{D}\right)^{\frac{1}{2(r-1)}} \log^{\frac{5}{2(r-1)}} N \right).$$

3 Proof

We begin with a brief overview of the critical interval method, which is a refinement of the differential equations method for proving dynamic concentration. In a standard application of the differential equations method, we have a sequence of random variables $Z(0), Z(1), \ldots$ that is determined by some combinatorial random process on n points, and our dynamic concentration statement is

$$Z(i) = z(i/n) \pm e_z(i/n) \qquad \text{for} \qquad i = 0, 1, \ldots, M(n) \tag{3}$$

with high probability. Note that we use the symbol "$\pm$" in two distinct ways: sometimes we write $a = b \pm c$ meaning that a is in the interval $[b-c, b+c]$ whereas other times we simply use "$\pm$" as a symbol that could either be "$+$" or "$-$." The meaning should be clear from context. The deterministic trajectory function z is usually determined by the one-step expected changes in $Z(i)$ and the initial condition $z(0) = Z(0)$. The error function e_z is a carefully chosen, slowly growing, function. It is often convenient to introduce a continuous time variable t that we relate to the steps of the process by setting $t = t(i) = i/n$. This allows us to view the function $z(t)$ as a scaling limit for the sequence $Z(i)$.

In a standard application of the differential equations method we prove the dynamic concentration statement (3) by two applications of a martingale deviation inequality. We introduce a stopping time T, which is defined to be the minimum of $M = M(n)$ and the first step i at which (3) fails. We then define the two sequences of random variables $\mathcal{D}Z^+(i), \mathcal{D}Z^-(i)$ as follows:

$$\mathcal{D}Z^{\pm}(i) = Z(i \wedge T) - z(t \wedge (T/n)) \pm e_z(t \wedge (T/n)).$$

Note that violation of the upper bound in (3) is equivalent to $\mathcal{D}Z^-(T) = \mathcal{D}Z^-(M) > 0$ and violation of the lower bound in (3) is equivalent to $\mathcal{D}Z^+(T) = \mathcal{D}Z^+(M) < 0$. Note further that $\mathcal{D}Z^-(0) = -e_z(0)$ and $\mathcal{D}Z^+(0) = e_z(0)$. If $\mathcal{D}Z^-$ is a supermartingale and $\mathcal{D}Z^+$ is a submartingale, then violation of (3) is contained in the event that one of these martingales has a large deviation. We choose the error functon $e_z(t)$ so that $\mathcal{D}Z^-$ is a supermartingale and $\mathcal{D}Z^+$ is a submartingale and $e_z(0)$ is sufficiently large to make the probabilities of these martingale deviations small. We emphasize that the introduction of this stopping time T is an important detail in the proof as it allows us to assume the bounds in (3) when we establish the martingale condition and apply the martingale inequality.

Our proof of Theorem 2.1 requires even greater control over the random variable Z when we are establishing the martingale condition. This is what the critical interval method provides. For each variable Z treated by Theorem 2.1 and each bound (i.e. upper and lower) we introduce a *critical interval* $I_Z(t) = [a_Z, b_Z]$ which has one end at the bound we are trying to establish and the other end slightly closer to the trajectory $z(t)$. The upper critical interval is

$$I_Z(t) = [z(t) + e_z(t) - f_z(t), z(t) + e_z(t)]$$

where the width $f_z(t)$ will be chosen below. Simillarly, the lower critical interval is

$$I_Z(t) = [z(t) - e_z(t), z(t) - e_z(t) + f_z(t)]$$

We can view violation of the dynamic concentration statement given by Theorem 2.1 as the event that some variable manages to cross one of its critical intervals. In order to bound the probability of this event we consider a large collection of martingales. We have one such martingale for each variable, each bound (upper and lower), and each step of the process that the random variable in question might enter the critical interval for the last time before crossing the interval.

Consider a random variable Z in the collection of random variables treated by Theorem 2.1, some step j of the process, and the upper bound on Z. We introduce a stopping time that is specialized to the event that variable Z enters its upper critical interval at step j and proceeds to cross the interval without leaving it. Define $T_{Z,j}$ to be the minimum of the global stopping time T (which is defined in Section 2 above) and the first step $i \geq j$ when $Z(i)$ is not in its upper critical interval. We simply have $T_{Z,j} = j$ if $Z(j)$ is not in the upper critical interval. We consider the sequence of random variables

$$\mathcal{D}Z_j^-(i) = Z(i \wedge T_{Z,j}) - z(t \wedge (T_{Z,j}/N)) - e_z(t \wedge (T_{Z,j}/N)) \quad \text{for} \quad i = j, \dots$$

Now, assuming that we have a suitable bound on the one step changes in each variable Z, the event $T = i$ and $Z(i) > z(t) + e_z(t)$ is contained in the event that there exists a $j < i$ such that $\mathcal{D}Z_j^-(j) \approx -f_z(j/N)$ and $\mathcal{D}Z_j^-(i) > 0$. If $\mathcal{D}Z_j^-$ is a supermartingale then each such event is the event that this martingale has a large deviation. We establish bounds on these events that are small enough that a simple application of the union bound – taking the union over all variables, bounds and starting points j – shows that the probability that of any event in the collection occuring is small. Theorem 2.1 follows.

We stress that the introduction of the stopping time $T_{Z,j}$ allows us to assume that Z is in the critical interval when we are establishing the martingale condition for Z. (Of course the other random variables are not so constrained.) The reason that we focus our attention on these critical intervals is the fact that the expected one-step changes in the variables we consider have self-correcting terms. These terms introduce a drift back toward the expected trajectory when Z is far from the expected trajectory. By restricting our attention to the critical intervals we make full use of these terms. See [13] and [5] for early applications of this self-correcting phenomenon in applications of the differential equations method for proving dynamic concentration. As we noted above, the critical interval method we use here was introduced in [3].

We close this preamble with some notation conventions and a lemma that we use below. For an arbitrary random variable Z we define

$$\Delta Z(i) = Z(i+1) - Z(i).$$

We let $\mathcal{F}_i$ be the filtration of the probability space given by the first i edges chosen by the random greedy matching process.

Lemma 3.1. *Suppose $(x_i)_{i\in I}$ and $(y_i)_{i\in I}$ are real numbers such that $|x_i - x| \le \delta$ and $|y_i - y| < \epsilon$ for all $i \in I$. Then we have*

$$\left|\sum_{i\in I} x_i y_i - \frac{1}{|I|}\left(\sum_{i\in I} x_i\right)\left(\sum_{i\in I} y_i\right)\right| \le 2|I|\delta\epsilon$$

Proof. The triangle inequality gives

$$\left|\sum_{i\in I}(x_i - x)(y_i - y)\right| \le |I|\delta\epsilon.$$

Rearranging this inequality gives

$$\begin{aligned}\sum_{i\in I} x_i y_i &= x\sum_{i\in I} y_i + y\sum_{i\in I} x_i - |I|xy \pm |I|\delta\epsilon \\ &= \frac{1}{|I|}\left(\sum_{i\in I} x_i\right)\left(\sum_{i\in I} y_i\right) - |I|\left(\frac{1}{|I|}\sum_{i\in I} x_i - x\right)\left(\frac{1}{|I|}\sum_{i\in I} y_i - y\right) \pm |I|\delta\epsilon.\end{aligned}$$

□

3.1 Vertex degrees

Let v be a fixed vertex. As usual in applications of the differential equations method for establishing dynamic concentration, we begin with the expected one-step change in d_v (i.e. we begin with the trend hypothesis). We have

$$E\left[\Delta d_v(i)|\mathcal{F}_i\right] = -\frac{1}{Q}\sum_{E\in\mathcal{H}(i):v\in E}\ \sum_{u\in E\setminus\{v\}} d_u(i) \pm d_v(i)\binom{r}{2}\frac{L}{Q}, \tag{4}$$

where $\mathcal{F}_i$ is the filtration defined by the random greedy matching process. We note that (4) does not take into account the contribution to the expected change in d_v that comes from the selection of an edge that contains v itself. Of course, this event causes a rather dramatic change in d_v, which could complicate our analysis. Furthermore, we are no longer interested in d_v after v leaves $V(i)$. This is handled formally by setting $d_v(i+1) = d_v(i)$ if $v \notin V(i+1)$, and (4) takes this convention into account.

We begin with the upper bound on d_v. Our critical interval is

$$[Dp^{r-1} + e_d - f_d, Dp^{r-1} + e_d],$$

where

$$f_d = \sqrt{6rLD\log N} \qquad \text{and} \qquad e_d = f_d\left(1 - r\log p\right).$$

Note that f_d does not change in time and that e_d is increasing.

For each step j of the process we define the sequence of random variables

$$\mathcal{D}d^+_{v,j}(i) := d_v(i) - Dp^{r-1} - e_d \qquad \text{for } i \geq j$$

with the stopping time $T_{v,j}$ defined to be the minimum of T and the smallest index $i \geq j$ such that $d_v(i)$ is not in the critical interval or $v \notin V(i)$. Note that if $d_v(j)$ is not in the critical interval then we simply have $T_{v,j} = j$. We prove dynamic concentration by considering the sequence of random variables $\mathcal{D}d^+_{v,j}(j), \ldots, \mathcal{D}d^+_{v,j}(T_{v,j})$. We chose f_d and e_d (with foresight) so that this sequence is a supermartingale with respect to the natural filtration $\mathcal{F}_i$. For $j \leq i < T_{v,j}$ we have

$$\begin{aligned} E\left[\Delta \mathcal{D}d^+_{v,j} | \mathcal{F}_i\right] &\leq -\frac{1}{Q} \sum_{E \in \mathcal{H}(i): v \in E} \sum_{u \in E \setminus \{v\}} d_u(i) + \frac{Dr(r-1)}{N} p^{r-2} - \frac{1}{N} e'_d \\ &\qquad\qquad + O\left(\frac{Ld_v}{Q} + \frac{D}{N^2} p^{r-3} + \frac{1}{N^2} e''_d\right) \\ &\leq -\frac{\left(Dp^{r-1} + e_d - f_d\right)(r-1)\left(Dp^{r-1} - e_d\right)}{NDp^r/r + e_q} + \frac{Dr(r-1)}{N} p^{r-2} \\ &\qquad\qquad - \frac{1}{N} e'_d + O\left(\frac{Ld_v}{Q} + \frac{D}{N^2} p^{r-3} + \frac{1}{N^2} e''_d\right) \\ &\leq \frac{r(r-1)}{Np} f_d - \frac{1}{N} e'_d \\ &\qquad\qquad + O\left(\frac{(e_d - f_d) e_d}{NDp^r} + \frac{e_q}{N^2 p^2} + \frac{Ld_v}{Q} + \frac{D}{N^2} p^{r-3} + \frac{1}{N^2} e''_d\right) \end{aligned}$$

Note that we use the assumption that $d_v(i)$ lies in the critical interval. Also note that in order to get the desired supermartingale condition it is necessary to choose e_d and f_d so that

$$e'_d > \frac{r(r-1)}{p} f_d. \tag{5}$$

(Of course, this equation plays a central in our choice of the functions f_d and e_d.)

For the given error functions e_d, e_q, we have

$$\begin{aligned} &\frac{(e_d - f_d) e_d}{NDp^r} + \frac{e_q}{N^2 p^2} + \frac{Ld_v}{Q} + \frac{D}{N^2} p^{r-3} + \frac{1}{N^2} e''_d \\ &\qquad \leq \frac{e_d}{Np} \cdot O\left(\frac{e_d p^{1-r}}{D} + \frac{e_q}{e_d N p} + \frac{L}{e_d} + \frac{D}{N e_d} p^{r-2} + \frac{1}{Np}\right) \qquad (6) \\ &\qquad \leq \frac{e_d}{Np} \cdot O\left(\frac{\sqrt{L} (\log N)^{3/2} p^{1-r}}{\sqrt{D}}\right) + \frac{e_d}{Np} \cdot o\left(\frac{\sqrt{L}}{\sqrt{D}} + \frac{\sqrt{D}}{N\sqrt{L}} + \frac{1}{\sqrt{N}}\right). \end{aligned}$$

(We note that these estimates make repeated use of the simple inequality $D < NL$.) By assuming that p is a sufficiently large constant times

$$\left(\frac{L}{D}\right)^{\frac{1}{2(r-1)}} \log^{\frac{5}{2(r-1)}} N$$

we see that the expression in (6) can be made smaller than any constant times $e_d/(Np \log N)$. As the error functions f_d and e_d satisfy (5), the supermartingale condition is satisfied.

We use a supermartingale inequality to bound the probability that the random variable $\mathcal{D}d^+_{v,j}(T_{v,j})$ is positive. We use the following Lemma (see [2] for a proof).

Lemma 3.2. *Let $X(i)$ be a supermartingale, such that $-\Theta \leq \Delta X(i) \leq \theta$ for all i, where $\theta < \frac{\Theta}{10}$. Then for any $a < \theta m$ we have*

$$Pr(X(m) - X(0) > a) \leq \exp\left(-\frac{a^2}{3\theta\Theta m}\right).$$

Since d_v is non-increasing, Dp^{r-1} is decreasing and e_d is increasing, the one step change in $\mathcal{D}d^+_{v,j}$ is bounded above by the one step change in Dp^{r-1}, which is at most

$$\theta = \frac{D(r-1)}{N}(1 + o(1)).$$

For a lower bound on $\Delta d^+_{v,j}$, note that the one step change in e_d is negligible compared to the maximum possible one step change in d_v, which occurs when we pick an edge containing a vertex that has pairwise-degree L with v. So we can set $\Theta = rL(1 + o(1))$.

Now, if d_v crosses the upper boundary of its critical interval at the stopping time T, then there is some step j (with $T = T_{v,j}$) such that

$$\mathcal{D}d^+_{v,j}(j) \leq -f_d(t(j)) + \frac{D(r-1)}{N}(1 + o(1))$$

and $d^+_{v,j}(T_{v,j}) > 0$. Applying the lemma (and noting $D/N = o(f_d)$) we see that the probability of the supermartingale $d^+_{v,j}$ having such a large upward deviation has probability at most

$$\exp\left\{-\frac{f_d^2}{3\frac{D(r-1)}{N}(rL)(\frac{Np}{r})}(1 + o(1))\right\}.$$

As there are $O(N^2)$ such supermartingales, we would like the above expression to be $o(N^{-2})$. Thus, it suffices to take

$$f_d = \sqrt{6rLD \log N}.$$

Furthermore this choice also satisfies (5). (Note that, in fact, this condition together with (5) essentially determines the error functions e_d.)

Thus, the probability that T is less than bound stated in Theorem 2.1 due to a violation of the upper bound on d_v goes to zero as N tends to infinity.

The lower bound for d_v is similar.

3.2 Number of edges

We again begin with the trend hypothesis. We have

$$E\left[\Delta Q(i)|\mathcal{F}_i\right] = -\frac{1}{Q}\sum_{A\in\mathcal{H}(i)}\sum_{v\in A} d_v(A) + O(L) = -\frac{1}{Q}\sum_{v\in V(i)} d_v^2(i) + O(L)$$

For $i < T$ we have

$$\sum_{v\in V(i)} d_v^2 = \frac{(rQ)^2}{Np} \pm 2Npe_d^2,$$

by an application of Lemma 3.1, and therefore

$$E\left[\Delta Q(i)|\mathcal{F}_i\right] = -\frac{r^2 Q}{Np} \pm \frac{2Npe_d^2}{Q} + O(L).$$

We work with the upper bound on $Q(i)$. Our critical interval is

$$\left[\frac{ND}{r}p^r + e_q - f_q, \frac{ND}{r}p^r + e_q\right],$$

where

$$f_q = 6r^2 NL\log Np^{2-r} \qquad \text{and} \qquad e_q = 15 f_q\left(1 - r\log p\right)^2.$$

Note that both f_q and e_q are non-decreasing in time. For each step j of the process we define the sequence of random variables

$$\mathcal{D}Q_j^+(i) := Q(i) - \frac{ND}{r}p^r - e_q$$

with the stopping time T_j defined to be the minimum of T and the smallest index $i \geq j$ such that $Q(i)$ is not in the critical interval.

We begin by showing that $\mathcal{D}Q_j^+(j), \dots, \mathcal{D}Q_j^+(T_j)$ is a supermartingale. For $j \leq i < T_j$ we have

$$\begin{aligned}
E\left[\Delta\mathcal{D}Q_j^+(i)|\mathcal{F}_i\right] &\leq -\frac{r^2 Q}{Np} + rDp^{r-1} - \frac{1}{N}e_q' + \frac{2Npe_d^2}{Q} + O\left(L + \frac{D}{N}p^{r-2} + \frac{1}{N^2}e_q''\right) \\
&\leq -\frac{r^2(e_q - f_q)}{Np} - \frac{1}{N}e_q' + \frac{(2r + o(1))p^{1-r}e_d^2}{D} \\
&\qquad + O\left(L + \frac{D}{N}p^{r-2} + \frac{1}{N^2}e_q''\right)
\end{aligned}$$

In order to get the supermartingale condition it suffices, up to constant factors, to take

$$e_q > e_d^2 Np^{2-r}/D.$$

Note that this determines the main terms in the choice of e_q above. As $f_q = 6r^2 NL\log Np^{2-r}$, we have

$$-\frac{r^2(e_q - f_q)}{Np} + \frac{(2r + o(1))p^{1-r}e_d^2}{D} \leq -Lp^{1-r}(\log N)(1 - r\log p)^2.$$

This clearly dominates the remaining error terms (note that $e_q' > 0$) and therefore the sequence $\mathcal{D}Q^+(j), \ldots, \mathcal{D}Q^+(T_j)$ is a supermartingale.

Now we apply the Hoeffding-Azuma inequality to bound the probability that the random variable $\mathcal{D}Q^+(T_j)$ is positive. The lemma we use is as follows:

Lemma 3.3. *Let X_j be a supermartingale, with $|\Delta X_i| \le c_i$ for all i. Then*

$$P(X_m - X_0 \ge a) \le \exp\left(-\frac{a^2}{2\sum\limits_{i\le m} c_i^2}\right).$$

Since $i < T$ implies bounds on degrees, we have

$$|\Delta \mathcal{D}Q^+| \le (1+o(1))re_d \le \sqrt{7r^3 LD\log N}(1 - r\log p).$$

Thus, if Q crosses its upper boundary at the stopping time T, then there is some step j (with $T = T_j$) such that

$$\mathcal{D}Q^+(j) \le -f_q(t(j)) + O(\sqrt{LD}\log^{3/2} N)$$

and $\mathcal{D}Q^+(T_j) > 0$. Applying the Hoeffding-Azuma we see that the probability of the supermartingale $\mathcal{D}Q^+$ having such a large upward deviation has probability at most

$$\begin{aligned}&\exp\left\{-\frac{[(1+o(1))6r^2NL\log Np^{2-r}]^2}{2(Np)[7r^3LD\log N(1-r\log p)^2]}\right\}\\ &\qquad\qquad \le \exp\left\{-(1+o(1))\frac{18r}{7}\cdot\frac{NL}{D}\cdot\frac{p^{3-2r}}{(1-r\log p)^2}\cdot\log N\right\} = o(N^{-1})\end{aligned}$$

where $p = p(j)$. Note that we have used $D < NL$ again and that the constants have been chosen to deal with p constant. As there are at most $O(N)$ such supermartingales, the probability that T is less than the bound stated in Theorem 2.1 due to $Q(i)$ breaching the upper bound tends to zero as N tends to infinity.

The lower bound for Q is similar.

References

[1] N. Alon, J. Kim, J. Spencer, Nearly perfect matchings in regular simple hypergraphs. *Israel Journal of Mathematics* **100** (1997) 171–187.

[2] T. Bohman, The triangle-free process, *Advances in Mathematics* **221** (2009) 1653-1677.

[3] T. Bohman, A. Frieze, E. Lubetzky, A note on the random greedy triangle packing algorithm. *Journal of Combinatorics* **1** (2010), 477–488.

[4] T. Bohman, A. Frieze, E. Lubetzky, Random triangle removal. *Advances in Mathematics* **280** (2015), 379–438.

[5] T. Bohman, M. Picollelli, Evolution of SIR epidemics on random graphs with a fixed degree sequence. *Random Structures and Algorithms* **41** (2012) 179–214.

[6] P. Erdős, H. Hanani, On a limit theorem in combinatorial analysis. *Publicationes Mathematicae Debrecen* **10** (1963), 10–13.

[7] D. Grable, On random greedy triangle packing. *Electronic Journal of Combinatorics* **4** (1997), R11, 19 pp.

[8] A. Kostochka, V. Rödl, Partial Steiner systems and matchings in hypergraphs. *Random Structures and Algorithms* **13** (1998), 335–347.

[9] N. Pippenger, J. Spencer, Asymptotic behavior of the chromatic index for hypergraphs. *Journal of Combinatorial Theory, Series A* **51** (1989) 24–42.

[10] V. Rödl, On a packing and covering problem. *European Journal of Combinatorics* **6** (1985) 69–78.

[11] V. Rödl, L. Thoma, Asymptotic packing and the random greedy algorithm. *Random Structures and Algorithms* **8** (1996) 161–177.

[12] J. Spencer, Asymptotic packing via a branching process. *Random Structures and Algorithms* **7** (1995) 167–172.

[13] A. Telcs, N. Wormald, S. Zhou, Hamiltonicity of random graphs produced by 2-processes. *Random Structures and Algorithms* **31** (2007) 450–481.

[14] V. Vu, New bounds on nearly perfect matchings in hypergraphs: Higher codegrees do help. *Random Structures and Algorithms* **17** (2000) 29–63.

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