COMBINATORICA 5 (1) (1985) 81—94

COMPONENT STRUCTURE IN THE EVOLUTION OF RANDOM HYPERGRAPHS

Jeanette SCHMIDT-PRUZAN and Eli SHAMIR

Received 28 March 1984

The component structure of the most general random hypergraphs, with edges of differen sizes, is analyzed. We show that, as this is the case for random graphs, there is a "double jump" in the probable and almost sure size of the greatest component of hypergraphs, when the average vertex degree passes the value 1.

1. Introduction

A hypergraph $\mathbf{H}=(V, E)$ is a set V of vertices and a family E of subsets of V called edges. The edges of $\mathbf{H}$ may have various sizes; in many cases the size of its largest edge, called the *range* of $\mathbf{H}$, has special importance. The probable structure of random *graphs*, (hypergraphs of range 2), with n labelled vertices and given density, has been studied in a series of articles by Erdős and Rényi, [6], where many striking results have been established.

For a number of fundamental structural properties P, there exists a "*threshold function*" $P(n)$, such that the probability that the graph $\mathbf{G}$ possesses P increases from 0 to 1, when the density parameter $D_G(n)$ of $\mathbf{G}$, $\left(0\leqq D_G(n)\leqq\binom{n}{2}\right)$, ranges from $P(n)/\omega(n)$ to $P(n)\omega(n)$, (for any $\omega(n)$ s.t. $\lim_{n\to\infty}\omega(n)=\infty$). This is the case for connectedness, and for the presence of certain subgraphs of given type and order (such as perfect matchings, trees, cycles or connected components). Naturally the question arises whether these results generalize to hypergraphs of arbitrary range with edges of various sizes. All monotone properties have "*threshold intervals*" $[\underline{P}(n), \overline{P}(n)]$, meaning that Prob ($\mathbf{H}$ possesses P) increases from 0 to 1, when $D_H(n)$ ranges from $\underline{P}(n)$ to $\overline{P}(n)$, however it is not clear for which monotone properties these intervals become narrow enough as to define a threshold function. In a previous article [9] we made the attempt to find a threshold function, or rather a narrow "*threshold interval*", for the presence of perfect matchings in random hypergraphs, and it is an open question whether a threshold function for this hypergraph-property exists. In the present article we investigate the growth of the greatest (connected)

AMS subject classification (1980): 05 C 65, 60 C 05 component in random hypergraphs. For random graphs, the situation can be summarized as follows, [Erdős and Rényi, 5]:

If $C(d)$ is the size of the greatest connected component in a random graph of order n with average vertex degree d, then with probability tending to 1 when n tends to ∞:

$$\begin{aligned} C(d) &= O(\log n) && \text{for} \quad d < 1 \\ C(d) &= O(n^{(2/3)}) && \text{for} \quad d \approx 1 \qquad (1.1) \\ C(d) &= O(n) && \text{for} \quad d > 1. \end{aligned}$$

This "double jump" of the size of the greatest component when d passes the value 1 is one of the most impressive facts concerning random graphs. It generalizes to uniform as well as mixed hypergraphs and the corresponding result reads:

If $C(d^*(n))$ is the size of the greatest connected component in a random hypergraph of order n, range t, with $2 \leq t \leq O(\ln n)$, and with average vertex degree $d^*(n)$, then with probability tending to 1 when n tends to infinity:

$$\begin{aligned} C(d^*(n)) &= O(t \log n) && \text{for} \quad d < 1 \\ C(d^*(n)) &= O(n^{(2/3)}) && \text{for} \quad d \approx 1 \qquad (1.2) \\ C(d^*(n)) &= O(n/t) && \text{for} \quad d > 1. \end{aligned}$$

The theorems in Section 3 prove the relations stated in (1.2), and include also information on the expected structure of the hypergraph-components.

One can consider components of random hypergraphs as an outcome of a branching process. This explains the jump from a negligeable size if $d^*(n) < 1$ to a very large size if $d^*(n) > 1$, but this approach is not expected to yield the refined and precise results we state. These are achieved by elaborating the counting and estimation method used in [5].

However proving the existence of a large path for $d^*(n) > 1$ may necessitate an intricate branching process analysis like the one given in [1] for random graphs.

We shall introduce a few definitions in order to simplify the formulation of our Theorems.

2. Definitions and Notations

Probability Spaces

Probability spaces of graphs have been studied in a series of articles by Erdős and Rényi [6]. These spaces generalize to hypergraphs in a natural way. This has been done extensively in [9]. Random hypergraph-spaces have been studied as well in [8, 10, 11, 12] and mentioned in [4].

$\mathscr{H}(n, \vec{N})$: $\vec{N} = N_2, \ldots, N_t$. $\mathscr{H}(n, \vec{N})$ is the space of all hypergraphs with n vertices, and N_i edges of size i. All hypergraphs in this space have equal probability.

$\mathscr{H}(n, \vec{p})$: $\vec{p} = (p_2, p_3, \ldots, p_t)$. $\mathscr{H}(n, \vec{p})$ is the space of all hypergraphs with n vertices. The probability of a hypergraph in this space is calculated by the following condition:

edges of size i occur independently from each other and from edges of other size, with probability p_i. Hence all hypergraphs with the same number of edges, for all sizes i, have the same probability in $\mathcal{H}(n, \vec{p})$. Spaces of ordinary graphs are included here with $p_i=0$ for all $i\neq 2$. When $\mathbf{H}$ is t-uniform, ($p_i=0$ for all $i\neq t$), we denote the corresponding space by $\mathcal{H}(n, p_t)$.

Many definitions mentioned here were introduced by Berge [2, 3], some others extend these definitions or other common notions in graph theory.

Notations

In the following notations all hypergraphs $\mathbf{H}(V, E)$ are of range t and belong either to the space $\mathcal{H}(n, \vec{N})$ or to the space $\mathcal{H}(n, \vec{p})$.

A k-edge is an edge with k vertices.

$E=\bigcup_{k=2}^{t} E_k$, E_k is the set of k-edges of $\mathbf{H}$.

The *edge-degree* of $v\in V$ is the number of edges containing v.

The *vertex-degree* of $v\in V$ are the number of pairs (v_i, e_i) such that $\{v, v_i\}\in e_i$. (Note that vertices which have multiple links with v are counted accordingly.)

$\mathbf{E}(Y)$ is the expectation of the random variable Y.

$$(2.1)\quad d_k(n)=\begin{cases}\binom{n-1}{k-1}p_k=\mathbf{E}\left(\frac{k|E_k|}{n}\right) & \text{for}\quad \mathbf{H}\in\mathcal{H}(n,\vec{p}),\\ \frac{kN_k}{n}=\frac{k|E_k|}{n} & \text{for}\quad \mathbf{H}\in\mathcal{H}(n,\vec{N}).\end{cases}\qquad d(n)=\sum_{k=2}^{t} d_k(n)$$

$d_k(n)$ is the average k-edge degree of $\mathbf{H}\in\mathcal{H}(n, \vec{N})$, (or the *expected* average if $\mathbf{H}\in\mathcal{H}(n, \vec{p})$), $d(n)$ is the (expected) average edge degree of $\mathbf{H}$.

$$(2.2)\qquad d_k^*(n)=(k-1)d_k(n)\qquad d^*(n)=\sum_{k=2}^{t}(k-1)d_k(n)$$

$d_k^*(n)$ and $d^*(n)$ are respectively the (expected) average k-vertex degree and the (expected) average vertex degree of $\mathbf{H}$.

Definitions

A *sparse* hypergraph is a hypergraph with average vertex degree less than one.

A *probability space of sparse hypergraphs* is a space $\mathcal{H}$, in which the expected average vertex degree of $\mathbf{H}\in\mathcal{H}$ is less then one.

A *hyperpath* between s_1 and s_f is a tuple $(s_1, e_1, s_2, \ldots, e_t, s_{t+1}=s_f)$, such that for $1\leq i\leq t$, $e_i\in E$, and $\{s_i, s_{i+1}\}\in e_i$.

A *cycle* in a hypergraph is a tuple $(s_1, e_1, s_2, \ldots, e_t, s_1)$, such that for $1\leq i, j\leq t$, $e_i\neq e_j$, $e_i\in E$, and $\{s_i, s_{i+1}\}\in e_i$. Note that all edges of a cycle are distinct.

A subset S of V in $\mathbf{H}$ is *connected* if there is a hyperpath between any two vertices of S.

A subset S of V in $\mathbf{H}$ is isolated if there is no hyperpath between the vertices of S and $V-S$, (S and $V-S$ are disconnected).

A subset S of V in $\mathbf{H}$ is *self-connected* if S is connected and all the connecting edges are subsets of S.

A subset S of V in $\mathbf{H}$ is a *connected component* if S is (self)-connected and isolated.

A spanning *hypertree (h-tree)* on $S \subseteq V$ in $\mathbf{H}(V, E)$ is a hypergraph $\mathbf{H}'(S, E_S)$, $E_S \subseteq E$, where S is connected in $\mathbf{H}'$, (and hence self-connected in $\mathbf{H}$), and the removal of any edge from E_S will disconnect S in $\mathbf{H}'$.

Note that spanning h-trees may contain many cycles; an edge $\{v_1, v_2, v_3\}$ in the h-tree might be the only one covering v_3 but nevertheless close a cycle between v_1 and v_2.

A simple consequence of these definitions is that if S is self-connected then there is an h-tree in $\mathbf{H}$ spanning precisely the vertices of S. Note that for simple graphs (edges of size 2), the notions of connected and self-connected are the same.

A *hyper-breath-first-search (hBFS)* with root v_0 of a connected component $\mathbf{H}'$ is defined as follows:

— v_0 is at level 0 of the search.

— Level $i+1$ of the search consists of all new vertices which are connected to some vertex of level i, (the edges contributing the vertices intersect level i, but do not intersect any lower level).

Note that if $\mathbf{H}'$ is an h-tree then every edge of $\mathbf{H}'$ contributes at least one vertex to an hBFS rooted at any of its vertices.

A *cycling* edge in an hBFS is an edge of size i, contributing less than $(i-1)$ vertices, (and hence adding at least one cycle to the structure).

3. Main theorems and proofs

In order to investigate the probability of the appearance of connected components with a given number of vertices in a random hypergraph $\mathbf{H} \in \mathscr{H}(n, \vec{p})$, we shall point out a few facts, which are obvious enough as not to need an explicit proof.

Observation 3.1. *If $\mathbf{S}(V_S, E_S)$ is a connected component of $\mathbf{H}(V, E)$, then $\mathbf{S}$ contains an isolated (however not necessarily cycle-free) h-tree $\mathbf{S}'$ spanning V_S with a subset of E_S.* ▮

Observation 3.2. *In a t-uniform hypergraph $\mathbf{H}(V, E)$ a spanning h-tree on s vertices has between $\lceil (s-1)/(t-1) \rceil$ and $(s-t+1)$ edges, and a cycle free h-tree has exactly $\lceil (s-1)/(t-1) \rceil$ vertices.* ▮

Observation 3.3. *In a hypergraph $\mathbf{H}(V, E)$ (with edges of different sizes) a spanning h-tree on $S \subseteq V, |S|=s$, has t edges, $t=\sum_i t_i$, (t_i edges of size i), such that $\sum_i (i-1) t_i \geqq s-1 \geqq \sum_i t_i$.* ▮

Observation 3.4. *The cycling index of a connected component with s vertices and t_i edges of size i is defined as $\left(\left(\sum (i-1) t_i\right) - (s-1)\right)$. A connected component with cycling index c contains at least c cycles.* ▮

Observation 3.5. *$\mathscr{H}(n, \vec{p})$ is a space of sparse hypergraphs if $\vec{p}=(p_2, \ldots, p_t)$, $p_i = \alpha_i \left[(i-1) \binom{n-1}{i-1} \right]^{-1}$ and $\sum \alpha_i = \alpha < 1$. $\mathscr{H}(n, \vec{N})$ is a space of sparse hypergraphs if $\vec{N}=(N_2, \ldots, N_t)$, and $\sum i(i-1)(N_i/n) = \alpha < 1$. (Note that $\alpha_i = d_i^*(n)$ and $\alpha = d^*(n)$, as defined in (2.2)).* ▮

Theorem 3.6. *Let C and ε be two positive constants such that $C \geqq 1$ and $0 < \varepsilon < 1$ and let* **H** *be a sparse random hypergraph, of order n, range t and average vertex degree α. Under the above conditions there are constants $\overline{K}_{\alpha,C}$ and $\underline{K}_{\alpha,\varepsilon}$, such that, with probability $1-o(n^{-C})$, the largest connected component in* **H** *contains less than $\overline{K}_{\alpha,C}(t \ln n)$ vertices and with probability $1-o(n^{-\varepsilon})$ more than $\underline{K}_{\alpha,\varepsilon} \ln n$ vertices. Furthermore the cycling index of the connected components of* **H** *is, with probability $1-o(n^{-k+o(1)})$, at most k.*

Theorem 3.7. *Given a random hypergraph* **H**, *of order n, range t and average vertex degree $\alpha=1$, with probability $1-o(1)$ the largest connected component in* **H** *contains less than $\omega_1(n)n^{2/3}$ and more than $n^{2/3}/\omega(n)$ vertices, where $\omega(n)$ is any function satisfying $\lim_{n\to\infty} \omega(n)=\infty$, and $\omega_1(n) \approx O(\ln n)$.*

Theorem 3.8. *With probability $1-o(1)$ the largest connected component in a random hypergraph* **H**, *of order n, range t and average vertex degree $\alpha > 1$, contains more than $\beta_\alpha n/t$ vertices, where $0 < \beta_\alpha < 1$, is a constant, (independent of n and t).*

Before proceeding to the proofs of the three theorems, we shall outline their main ideas.

Outline of proofs (Thm. 3.6 and 3.7). Let $C(S)$ be the number of connected components of size S in **H**, and $\mathbf{E}(C(S))$ the expectation of $C(S)$. We shall show that both for $\alpha=1$ and for $\alpha<1$ the following holds:

(1) *if* $\mathbf{E}(C(S)) = o(1)$ *then* $\text{Prob}\,(C(S)=0) = 1-o(1)$

for $\alpha \leqq 1$

(2) $$\text{Prob}\,\left(\sum C(S) > 0\right) \geqq 1 - 1/\mathbf{E}\left(\sum C(S)\right).$$

Relation (1) is evident, whereas relation (2) will be proven by the second moment method. Note that relation (2) is trivial if $\mathbf{E}(C(S)) < 1$, we shall actually use it only for the case $\mathbf{E}(C(S)) \to \infty$, to show that the corresponding probabilities are $1-o(1)$.

Outline of proofs (Thm. 3.8). We shall show that when $d^*(n)=\alpha$ passes the value one, there is a constant $\beta<1$ such that, for $\Omega(\log(n)) \leqq S < \beta n$, $\mathbf{E}(C(S))=o(1)$. It follows that, with probability $1-o(1)$, **H** does not contain connected components in the above range. On the other hand, with probability $1-o(1)$, not all connected components in those graphs have less than $O(\log(n))$ vertices, (this is obvious even from the case $d^*(n)=1$). We then conclude that with probability $(1-o(1))$ **H** contains one connected component with more than βn vertices.

From the outline of the proofs we see that they are based on a precise evaluation of $\mathbf{E}(C(S))$. We shall show that the easier task, a precise evaluation of $\mathbf{E}(T(S))$, the expected number of isolated *cycle-free* h-trees in **H**, shall suffice. $\mathbf{E}(T(S))$ is indeed easier to evaluate than $\mathbf{E}(C(S))$, since

1. The T_i edges of size i of any cycle-free h-tree with S vertices satisfy the relation $\sum (i-1)T_i = S-1$, while the edges of a connected component with S vertices merely satisfy the relation $\sum (i-1)\binom{S}{i-1} \geqq \sum (i-1)T_i \geqq S-1$.

2. As we shall show in Lemmas (3.11, 3.12) we can give exact formulas for the number of *different* cycle-free h-trees on S labelled vertices, (while the same task for connected components with S vertices is very hard).

In Lemma (3.13) we evaluate $\mathbf{E}(T(S))$, using the count established in (3.11) and (3.12). In Lemma (3.14), (3.15) we derive $\mathbf{E}(C(S))$ from $\mathbf{E}(T(S))$. Finally in Conclusion (3.16) we draw the final conclusions necessary for the proofs of the three theorems. First however (Lemma 3.10) we shall prove Relation (2), without explicitly evaluating $\mathbf{E}(C(S))$, or even $\mathbf{E}(T(S))$.

Observation 3.9. *If S_1 and S_2 are two connected components of* $\mathbf{H}$, *spanning respectively V_1 and V_2 then either $V_1 = V_2$ or $V_1 \cap V_2 = \emptyset$.* ∎

Lemma 3.10. *In a hypergraph $\mathbf{H}(V, E) \in \mathcal{H}(n, \bar{p})$ with average vertex degree $d^*(n) = \alpha \leqq 1$ the following holds for all $1 \leqq S_1 < S_2 \leqq n$*

$$\mathrm{Prob}\,\{C([S_1, S_2]) > 0\} \geqq 1 - \frac{1}{\mathbf{E}(C([S_1, S_2]))}, \quad \textit{where} \quad C([S_1, S_2]) = \sum_{S=S_1}^{S_2} C(S).$$

Proof. As already mentioned we shall prove the above by the second moment method, [7]. We shall actually prove the inequality for a singleton S, namely $\mathrm{Prob}\,(C(S) > 0) \geqq 1 - 1/\mathbf{E}(C(S))$. This proves of course a slightly weaker claim, but the stronger version (needed only for the case $\alpha = 1$ in Lemma 3.13) is easy to verify, using exactly the same proof, but requiring more complicated notations.

$$\mathrm{Prob}\,\{C(S) \leqq 0\} \leqq \mathrm{Prob}\,\{|C(S) - \mathbf{E}(C(S))| > \mathbf{E}(C(S))\}$$

$$\leqq \frac{\mathrm{Var}\,(C(S))}{\mathbf{E}^2(C(S))} = \frac{\mathbf{E}(C(S)^2)}{\mathbf{E}^2(C(S))} - 1.$$

We shall show that

$$\frac{\mathbf{E}(C(S)^2)}{\mathbf{E}^2(C(S))} \leqq 1 + 1/\mathbf{E}(C(S)).$$

For this purpose we shall order the $\binom{n}{S}$ subsets with S vertices of V, such that for $1 \leqq i \leqq S$ we can refer to the i^{th} subset, and then define

$$C(S) = \sum_{i=1}^{\binom{n}{S}} X_i \quad \text{where} \quad X_i = \begin{cases} 1 & \text{the } i^{\text{th}} \text{ subset is a con. comp. of } \mathbf{H} \\ 0 & \text{otherwise.} \end{cases}$$

From Observation 3.9 it follows that there are exactly $\binom{n}{S}\binom{n-S}{S}$ ordered pairs of (X_i, X_j), for which $\mathbf{E}(X_i \wedge X_j) > 0$. Furthermore, for all these pairs, the event $(X_j = 1)$ is almost independent from the event $(X_i = 1)$, except what concerns edges between the i^{th} and j^{th} subset. We conclude that

$$\mathbf{E}(X_i \wedge X_j) = \mathbf{E}(X_i)\mathbf{E}(X_j) \prod_{i=2}^{t} (1 - p_i)^{-\sum_{j=2}^{i} \left[\binom{2S}{j} - 2\binom{S}{j}\right]\binom{n-2S}{i-j}}$$

$$\leqq \mathbf{E}(X_i)\mathbf{E}(X_j) \prod_{i=2}^{t} (1 - p_i)^{-S^2\binom{n}{i-2}}.$$

We can now proceed to evaluate

$$\frac{\mathbf{E}(C(S)^2)}{\mathbf{E}^2(C(S))} \leqq \frac{\binom{n}{S}\binom{n-S}{S}\mathbf{E}(X_i)\mathbf{E}(X_j)}{\binom{n}{S}\binom{n}{S}\mathbf{E}^2(X_i)} \prod_{i=2}^{t}(1-p_i)^{-S^2\binom{n}{i-2}} + \frac{\binom{n}{S}\mathbf{E}(X_i)}{\binom{n}{S}\binom{n}{S}\mathbf{E}^2(X_i)}$$

$$= \frac{\binom{n-S}{S}}{\binom{n}{S}} \prod_{i=2}^{t}(1-p_i)^{-S^2\binom{n}{i-2}} + \frac{1}{\mathbf{E}(C(S))}.$$

The first factor of the first term is bounded above by $(n-S)^S/n^S$, whereas the second factor is bounded by $(1-S\alpha/n)^{-S}$. We conclude that, for $\alpha \leqq 1$, the whole sum is bounded by $1+1/\mathbf{E}(C(S))$. This finishes the proof of the Lemma. Note that exactly the same proof holds for $T(S)$. ▌

Lemma 3.11. *The number of cycle-free t-uniform h-trees spanning* $[S+1]=(t-1)T_t$ *labelled vertices, equals the following expression:*

(3.11) $$\frac{(S+1)!\,(S+1)^{(T_t-1)}}{(S+1)(T_t)!\,(t-1)!^{T_t}}.$$

Lemma 3.12. *The number of cycle-free h-trees of range t spanning* $[S+1]$ *labelled vertices, equals the following expression:*

(3.12) $$\sum_{\{T_i \mid \sum(i-1)T_i=S\}} \frac{(S+1)!\,(S+1)^{(\sum T_i-1)}}{(S+1)\Pi_i(T_i)!\,(i-1)!^{\,T_i}}.$$

T_i stands for the number of edges of size i, in the h-tree.

Proof (of both lemmas). A cycle-free h-tree, with $[S+1]$ vertices is parameterized by exactly one hBFS for each of its potential $[S+1]$ roots. The edges of size i, occuring in the hBFS, can be decomposed in contributed (new), and one connecting (old) vertices. Once the partition of the n vertices into $[\sum T_i]$ contributions and the root v_r is determined the number of hBFS with precisely these contributions is $(S+1)^{\sum T_i-1}$. This is a generalized Cayley formula, which we shall demonstrate right below by showing that the number of hBFS $(v_r, \{T_i\})$, in which vertex v has edge-degree d_v+1 is exactly $\binom{\sum T_i-1}{d_r, d_{v_1}, \dots, d_{v_S}}$. All the h-trees $(v_r, \{T_i\})$ can be generated by the following procedure:

1. All $\sum T_i$ sets are free (unlocked) and not frozen. All vertices are free. Lock the root and all vertices v with $d_v=0$.

2. **While** V contains free vertices **do begin** *(**Construct subtree rooted at free vertex**)*
   (a) v is any free vertex in a free set. v chooses d_v neighboring sets among the free sets. All chosen sets, as well as v become locked, the set of v becomes frozen.

(b) **While** V contains free verticesi n locked sets **do begin**
v_f is any free vertex in a locked set. v_f chooses d_{v_f} neighbors among the free and not frozen sets. All chosen sets as well as v_f become locked.
end
Unfreeze the set of v_f.
end

3. All the free sets become the neighbors of the root. (There are exactly $\sum T_i - \sum d_{v_j} = d_r + 1$ free sets left at this point).

All hBFS with degree sequences d_v are generated by the above procedure. Furthermore it is easy to verify that the number of different hBFS is exactly $\binom{\sum T_i - 1}{d_r, d_{v_1}, \ldots, d_{v_S}}$, and hence their total number is

$$\sum_{d_v \geq 0} \binom{\sum T_i - 1}{d_r, d_{v_1}, \ldots, d_{v_S}} = (S+1)^{\Sigma T_i - 1}.$$

(3.12) follows. ∎

Lemma 3.13. *Let $T(S)$ be the number of isolated cycle-free h-trees with S vertices, in a hypergraph $\mathbf{H} \in \mathscr{H}(n, \bar{p})$, of range t with $\left(p_i = \alpha_i \left((i-1)\binom{n}{i-1}\right)^{-1}\right)$ and $\sum \alpha_i = \alpha$, then*

$$\text{if} \quad \alpha < 1 \begin{cases} \mathbf{E}(T(S)) = o(n^{-C}) & \text{for} \quad S \geq \overline{K}_{\alpha,C} t \ln n = C\overline{K}_\alpha t \ln n \\ \mathbf{E}(T(S)) \geq n^\delta & \text{for} \quad S < \underline{K}_\delta \ln n \end{cases}$$

$$\text{if} \quad \alpha = 1 \begin{cases} \mathbf{E}(T(S)) = o\left(\frac{n^{-2/3}}{e^{\omega^3(n)}}\right) & \text{for} \quad S \geq \omega(n) n^{2/3} \\ \mathbf{E}(\Sigma_{\mathscr{I}} T(S)) \geq \omega(n) & \text{for} \quad \mathscr{I} = \left\{S \,\middle|\, \frac{1}{2} n^{2/3} < \omega(n) S < n^{2/3}\right\} \end{cases}$$

$$\text{if} \quad \alpha > 1 \quad \mathbf{E}(T(S)) = o(e^{-\beta(S/t)}) \quad \text{for} \quad \underline{K}'_\alpha t \ln n < S < \overline{K}'_\alpha n/t.$$

Proof. Let $S' = S+1$. Denote: $[n]_k = n!/(n-k)!$.

A specific isolated cycle-free h-tree with T_i edges of size i in $\mathbf{H}$ with probability p_i for edges of size i has probability:

$$\prod_i p_i^{T_i} \prod_i (1-p_i)^{\sum_{j=1}^{i} \binom{S'}{j}\binom{n-S'}{i-j}}.$$

The first product accounts for the T_i edges of the h-tree, while the second factor accounts for the fact that the h-tree is isolated. It follows now from (3.12), and the formula for p_i that $\mathbf{E}(T(S'))$ equals:

$$\sum_{\{T_i | \Sigma(i-1)T_i = S'-1\}} \frac{[n]_{S'}}{\prod(([n]_{i-1}))^{T_i}} \frac{1}{S'^2} \prod_i \frac{(S'\alpha_i/(i-1))^{T_i}}{(T_i)!} \prod_i (1-p_i)^{\sum_{j=1}^{i} \binom{S'}{j}\binom{n-S'}{i-j}}.$$

Since $\sum (i-1)T_i = S'-1 = S$, it is convenient to express the formula in terms of S, instead of S', which essentially will make no difference.

$$\text{(3.13)} \quad = \sum_{\{T_i|\dots\}} \frac{O(1)[n]_S}{\Pi([n]_{i-1})^{T_i}} \frac{n}{S^2} \prod_i \frac{(S\alpha_i/(i-1))^{T_i}}{(T_i)!} \prod_i (1-p_i)^{S\binom{n}{i-1}-\binom{S}{2}\binom{n}{i-2}}$$

$$= \sum_{\{T_i|\dots\}} \varrho(S,\alpha_i) \prod_i \frac{(S\alpha_i/(i-1))^{T_i}}{(T_i)!}.$$

Where $\varrho(S,\alpha_i)$ is given by the following expression:

$$\text{(3.13a)} \qquad \varrho(S,\alpha_i) = O(1)\frac{n}{S^2} e^{-S(\Sigma_i \alpha_i/i-1)} e^{-(1-\alpha)S^2/2n} e^{-O(S^3/n^2)} e^{(t-1)S/n}.$$

The following lines contain the evaluation of (3.13), for any α, and finally the distinction into the three cases ($\alpha<1$, $\alpha=1$, $\alpha>1$), leading the different threshold values for S.

We shall show that the function

$$f(\vec{T}) = \frac{\prod_i \left(\alpha_i \frac{S}{i-1}\right)^{T_i}}{T_i!}$$

defined on all $\{T_i\}$ for which $\sum (i-1)T_i = S$, has an absolute maximum at $T_i^0 = (\alpha_i+\varepsilon_i)S/(i-1)$ for some $\{\varepsilon_i\}$ for which $\sum(\alpha_i+\varepsilon_i)=1$. All the ε_i are positive if $\sum \alpha_i \leqq 1$, and all the ε_i are negative if $\sum \alpha_i > 1$; their precise value shall be given in the course of the proof.)

Let $\vec{T} = T_2, \dots, T_t$ be any vector $\vec{T}$ for which $\sum (i-1)T_i = S$. We shall denote by i all the indices for which $T_i < T_i^0$, and by j all those for which $T_j \geqq T_j^0$. Then

$$\begin{aligned} &\forall i \quad T_i + c_i = T_i^0, \quad c_i \geqq 0 \\ &\forall j \quad T_j - c_j = T_j^0, \quad c_j \geqq 0 \end{aligned} \qquad \text{s.t.} \quad \sum (i-1)c_i = \sum (j-1)c_j.$$

Now

$$\text{(a)} \qquad f(\vec{T}) = f(\vec{T}^0) \frac{\prod_i \left(\alpha_j \frac{S}{j-1}\right)^{c_j} \prod_i [T_i^0]_{c_i}}{\prod_i \left(\alpha_i \frac{S}{i-1}\right)^{c_i} \prod_j [T_j^0 + c_j]_{c_j}}$$

$$\text{(b)} \qquad = f(\vec{T}^0) \frac{\prod_j \left(\frac{\alpha_j S/(j-1)}{T_j^0}\right)^{c_j}}{\prod_i \left(\frac{\alpha_i S/(i-1)}{T_i^0}\right)^{c_i}} g(\vec{c}) \quad \text{where} \quad g(\vec{c}) = e^{-O(\Sigma c_i^2/T_i^0)} \leqq 1.$$

Upon substituting $T_k^0 = (\alpha_k+\varepsilon_k)S/(k-1)$ for all $k=2 \dots t$

$$\text{(c)} \qquad = f(\vec{T}^0) \frac{\prod_j \left(\frac{\alpha_j}{\alpha_j+\varepsilon_j}\right)^{c_j}}{\prod_i \left(\frac{\alpha_i}{\alpha_i+\varepsilon_i}\right)^{c_i}} g(\vec{c}).$$

Set $\varepsilon_k=\alpha_k(e^{(k-1)\beta}-1)$ where β is the unique real solution of

$$\sum \alpha_k e^{(k-1)\beta} = 1, \quad (\beta \geqq 0 \quad \text{if} \quad \alpha \leqq 0 \quad \text{and} \quad \beta \leqq 0 \quad \text{if} \quad \alpha \geqq 0).$$

Upon substituting ε_k one obtains:

(d) $$f(\vec{T}) = f(\vec{T}^0)e^{\beta(\Sigma(i-1)c_i-\Sigma(j-1)c_j)}g(\vec{c}) = f(\vec{T}^0)g(\vec{c}).$$

This proves that $f(\vec{T})\leqq f(\vec{T}^0)$ for any $\vec{T}$. To evaluate (3.13) it remains to evaluate $\varrho(S,\alpha_i)f(\vec{T}^0)$

(e) $$\varrho(S,\alpha_i)f(\vec{T}^0) = \varrho(S,\alpha_i)\prod_{i=2}^{t}\frac{(\alpha_i S/(i-1))^{T_i^0}}{T_i^0!}$$

(f) $$\leqq \varrho(S,\alpha_i)\prod_{i=2}^{t}\left(\frac{\alpha_i e}{\alpha_i+\varepsilon_i}\right)^{(\alpha_i+\varepsilon_i)S/(i-1)}$$

(g) $$\leqq \varrho(S,\alpha_i)e^{\Sigma S/(i-1)[\alpha_i-\varepsilon_i^2/2(\alpha_i+\varepsilon_i)]},$$

where $L=2$ if $\alpha<1$, $L=8$ if $\alpha>1$. Note that $\sum \varepsilon_i^2/(\alpha_i+\varepsilon_i)\geqq(1-\sum\alpha_i)^2$, and substitute $\varrho(S,\alpha_i)$ as given in (3.13a), to obtain the bound of:

(h) $$\leqq \frac{n}{S^2}e^{-(1-\alpha)^2S/L(t-1)}e^{-(1-\alpha)S^2/2n}.$$

1. Sparse hypergraphs (i.e. $\alpha<1$)

Since there are at most $\binom{S}{t-2}$ terms in the sum, we conclude that (3.13) is bounded by

$$\binom{S}{t-2}\frac{n}{S^2}e^{-(1-\alpha)^2S/2(t-1)}.$$

For $S=\overline{K}_{\alpha,C}t\ln n$, with $\overline{K}_{\alpha,C}=C\overline{K}_\alpha\approx 4C/(1-\alpha)^2$, we obtain that

$$\mathbf{E}(T(S)) = o(n^{-C}).$$

For $S<\underline{K}_\delta \ln n$, with $\underline{K}_\delta\approx(1-\delta)$, it follows from (e) that the single term corresponding to $\vec{T}^0$ is greater than n^δ, and hence $\mathbf{E}(T(S))>n^\delta$.

2. $\alpha=1$

In this case we could have stopped the estimates at (e), and conclude that (3.13) equals

(e1) $$\sum_{\{c_i,c_j|\Sigma(i-1)c_i=\Sigma(j-1)c_j\}}\varrho(S\alpha_i)f(\vec{T}^0)g(\vec{c}).$$

We proceed to estimate (e1):

$$= \sum_{\{c_i,c_j|\dots\}}O(1)\frac{n}{S^2}\prod_{i=1}^{t}\frac{1}{\sqrt{T_i^0}}e^{-O(\Sigma c_i^2/T_i^0)}e^{-O(S^3/n^2)}$$

$$= O(1)\frac{n}{S^2}e^{-O(S^3/n^2)}\sum_{\{c_i,c_j|\dots\}}\prod_{i=1}^{t}\frac{1}{\sqrt{T_i^0}}e^{-O(\Sigma c_i^2/T_i^0)}$$

$$= O(1)\frac{n}{S^2}\frac{1}{\sqrt{S}}e^{-O(S^3/n^2)}.$$

For $S > \omega(n) n^{2/3}$, we obtain that

$$\mathbf{E}(T(S)) = o(n^{-2/3} e^{-\omega^3(n)}), \quad \text{and} \quad \mathbf{E}(\sum T(S)) = o(e^{-\omega^3(n)}).$$

Whereas for $\mathscr{J} = [n^{2/3}/2\omega(n),\ n^{2/3}/\omega(n)]$

$$\sum_{S \in \mathscr{J}} \mathbf{E}(T(S)) = O(\omega(n)^{1.5}).$$

3. $\alpha > 1$

We continue the development from (h) one step further and obtain the expression in the following form:

$$\text{(i)} \qquad \frac{n}{S^2} e^{-\left\{\frac{(\alpha-1)^2}{8(t-1)} - \frac{(\alpha-1)S}{2n}\right\} S}.$$

For $S = 4\beta'(\alpha-1)n/(t-1) = 4$, $\beta' < 1$, the expression in the brackets is greater than

$$\frac{(1-\beta')(\alpha-1)^2}{8(t-1)} = \frac{\beta''}{t}$$

and hence (i) is $n/S^2\, o(e^{-\beta'' S/t})$.

For $S > (2/\beta'')t \ln n = \underline{K}'_\alpha t \ln n$ and $\beta = \beta''/2$ (i) is bounded by

$$o(e^{-\beta S/t}).$$

This concludes the proof of the lemma. ∎

Clearly $\mathbf{E}(T(S)) \leqq \mathbf{E}(C(S))$, and hence for all S for which $\mathbf{E}(T(S)) = \infty$ $\mathbf{E}(C(S)) = \infty$ holds. It remains to show that, for most values of S, if $\mathbf{E}(T(S)) = o(1)$, then $\mathbf{E}(C(S)) = o(1)$. The next two lemmas show that $\mathbf{E}(C(S))$ is, in all relevant cases not considerably larger than $\mathbf{E}(T(S))$.

Lemma 3.14. *If $C_k(S)$ is the number of connected components with cycling index k in $\mathbf{H}(V, E) \in \mathscr{H}(n, \vec{p})$ with average vertex degree $\sum \alpha_i = \alpha \leqq 1$, then:*

$$\mathbf{E}(C_k(S)) \leqq \frac{(O(S^2/n))^k}{k!} \mathbf{E}(T(S+k)).$$

Proof. A connected component with S vertices and cycling index k, can be represented as several cycle-free h-tree with $S+k$ vertices, composed of the original S vertices and k additional copies of vertices of S. (This can be done for example during a hBFS in which all edges of size i will be forced to contribute $i-1$ vertices by dynamically adding additional copies of vertices which already appeared in the hBFS.) It follows that

$$C_k(S) \binom{n}{S}^{-1} \leqq \binom{S+k}{k} T(S+k) \binom{n}{S+k}^{-1}.$$

For $\mathbf{E}(C_k(S))$ we need the probability that a specific connected component with cycling index k and T_i edges of size i, is present in $\mathbf{H} \in \mathscr{H}(n, \vec{p})$; for $\mathbf{E}(T(S))$ we used the probability that a specific cycle free h-tree with T_i edges of size i is present in $\mathbf{H} \in \mathscr{H}(n, \vec{p})$. These probabilities are given respectively by:

$$\prod_i p_i^{T_i} (1-p_i)^{\sum_{j \geq 1} \binom{S}{j}\binom{n-S}{i-j}} \quad \text{and} \quad \prod_i p_i^{T_i} (1-p_i)^{\sum_{j \geq 1} \binom{S+k}{j}\binom{n-S-k}{i-j}}.$$

Note that for both structures $\sum (i-1)T_i = S-1+k$. It follows that:

$$\mathbf{E}(C_k(S)) \leqq \frac{\binom{n}{S}}{\binom{n}{S+k}} \binom{S+k}{k} \mathbf{E}(T(S+k)) \prod_i (1-p_i)^{\sum \binom{S}{j}\binom{n-S}{i-j} - \binom{S+k}{j}\binom{n-S-k}{i-j}}$$

$$\leqq \frac{([S+k]_k)^2}{k!} \frac{1}{[n-S]_k} e^{[\sum \alpha_i/(i-1)]k}$$

$$\leqq \frac{\{O(S^2/n)\}^k}{k!} \mathbf{E}(T(S+k)). \quad \blacksquare$$

Lemma 3.15. *(The structure of connected components) The expected number of cycles,* $\mathbf{E}(\text{cycl})$, *in* $\mathbf{H}(V, E) \in \mathscr{H}(n, \vec{p})$, *with average vertex degree* $\sum \alpha_i = \alpha \leqq 1$, *is* $O(\ln n)$ *for* $\alpha = 1$, *and* $O(1)$ *for* $\alpha < 1$. *Moreover, for each constant* C, *there is a constant* K_C *such that the probability that* $\mathbf{H}(V, E) \in \mathscr{H}(n, \vec{p})$ *contains more than* $K_C \ln n\, \mathbf{E}(\text{cycl})$ *cycles is less than* n^{-C}.

Proof. The expected number of cycles of order c in $\mathbf{H}(V, E)$ is

$$\binom{n}{c}(c-1)!\,(1-\prod_i (1-p_i)^{\binom{n-2}{i-2}})^c \leqq \frac{1}{c}\,(\textstyle\sum_i \alpha_i)^c = \frac{\alpha^c}{c}.$$

We conclude that: $\mathbf{E}(\text{cycl}) \leqq \sum_{c=2}^{n} \frac{\alpha^c}{c}$.

This is indeed $O(\ln n)$ for $\alpha = 1$, and $O(1)$ for $\alpha < 1$.

Moreover, we can divide the cycles into $2\mathbf{E}(\text{cycl})$ order groups such that the expected number of cycles in each group is at most 1/2. It follows that for edge disjoint cycles the probability of having more than $C(\ln n)$ cycles in any group is $o(n^{-C})$, and hence the probability of a total of more than $K_C(\ln n)$ disjoint cycles is $o(\mathbf{E}(\text{cycl})n^{-C})$. It is easy to verify that the probability of intersecting cycles is even smaller. $\blacksquare$

3.16. *(Conclusion of the proof of Theorems* (3.6), (3.7), *and* (3.8)*)*

1. $\alpha < 1$ *(Theorem* 3.6*)*

(a) *(upper bound)* By Lemma (3.15) with probability $1-o(n^{-C})$, $\mathbf{H}(V, E)$ contains at most $O(\ln n)$ cycles and hence by Lemma (3.14)

$$\mathbf{E}(C(S)) \leqq \sum_{k \leqq O(\ln n)} \frac{\{O(S^2/n)\}^k}{k!} \mathbf{E}(T(S+k)).$$

For $S = C\overline{K}_\alpha t \ln n$ (by lemma (3.13)) $\mathbf{E}(T(S)) = o(n^{-C})$, and hence

$$\mathbf{E}(C(S)) \leqq \sum_k \frac{\{O(S^2/n)\}^k}{k!} o(n^{-S/(\overline{K}_\alpha t \ln n)}) = o(n^{-C(1-o(1))}).$$

(b) *(lower bound)* For $S < \underline{K}_\delta \ln n$ $\mathbf{E}(C(S)) \geqq \mathbf{E}(T(S)) \geqq n^\delta$ and hence by Lemma (3.10) Prob $(C(S) > 0) = 1 - n^{-\delta}$. Moreover again by Lemma (3.14) for all S

for which $\mathbf{E}(C(S)) \neq o(1)$, (i.e. $S < \bar{K}_\alpha t \ln n$)

$$\mathbf{E}(C_k(S)) \leqq \frac{\{O(S^2/n)\}^k}{k!} \mathbf{E}(T(S+k)) \leqq \frac{\{O(S^2/n)\}^{k-1}}{k!}$$

and we conclude that the cycling index of the connected components of $\mathbf{H}$ is with probability $1-o(n^{-k+o(1)})$, at most k. (This holds for any positive integer k.)

2. $\alpha=1$ *(Theorem 3.7)*

(a) *(upper bound)* Again by Lemma (3.15) with probability $1-o(1)$ $\mathbf{H}(V, E)$ contains at most $O(\ln n)$ cycles and hence by lemma (3.14)

$$\sum_{S > \omega(n) n^{2/3}} \mathbf{E}(C(S)) \leqq \sum_{k < O(\ln n)} \frac{\{O(S^2/n)\}^k}{k!} \mathbf{E}(T(S+k))$$

$$\leqq \sum_{\{S | S = \omega(n) n^{2/3}\}} \sum_k \frac{\{O(S^2/n)\}^k}{k!} o(n^{-2/3} e^{-\omega^3(n)})$$

which is $o(1)$ if $\omega(n) > \ln n$.

(b) *(lower bound)* For $\mathcal{S} = \{S | n^{2/3}/2 < S\omega(n) < n^{2/3}\}$

$$\mathbf{E}(\textstyle\sum_{\mathcal{S}} C(S)) > \mathbf{E}(\sum_{\mathcal{S}} T(S)) = \infty$$

and hence by Lemma (3.10) Prob $(\sum_{\mathcal{S}} C(S) > 0) = 1 - o(1)$. This concludes the proof of Theorem (3.7).

3. $\alpha > 1$ *(Theorem 3.8)* By Lemma (3.14),

$$\sum_{S=\underline{K}'_\alpha t \ln n}^{K'_\alpha n/t} \mathbf{E}(C(S)) \leqq \sum_k \frac{\{O(S^2/n)\}^k}{k!} \mathbf{E}(T(S+k))$$

$$\leqq \sum_{S=\underline{K}'_\alpha t \ln n}^{K'_\alpha n/t} e^{S^2/n} o(e^{-\beta S/t}).$$

This is $o(1)$ for $S \leqq \min(\bar{K}'_\alpha, \beta) n/t = O(n/t)$. It follows that in this case $\mathbf{H}$ has with probability $1-o(1)$ at least one connected component with $O(n/t)$ vertices. This terminates the proof of Theorem (3.8).

This also terminates the proof of all our Theorems. ∎

References

[1] M. Ajtai, J. Komlós and E. Szemerédi, The longest path in a random graph, *Combinatorica*, **1** (1981), 1—12.

[2] C. Berge, *Graphes et Hypergraphes*, Dunod, Paris, (1970).

[3] C. Berge, *Introduction à la Theorie des Hypergraphes*, Le presse de l'Université de Montréal, Seminaire de Math. Superieur, été 1971.

[4] B. Bollobás and P. Erdős, Cliques in random graphs, *Math. Proc. Camb. Phil. Soc.*, **80** (1976), 419—427.

[5] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. of the Math. Inst. of the Hung. Acad. Sci.*, **5** (1960), 17—61.

[6] P. Erdős, *The Art of Counting*, Selected Writings, MIT Press, Cambridge/Massachusetts and London/England, 1973.

[7] P. ERDŐS and J. SPENCER, *Probabilistic Methods in Combinatorics*, Academic Press, New York, 1974.
[8] W. FERNANDEZ DE LA VEGA, Sur la cardinalité maximum des couplages d'hypergraphes aléatoire uniformes, *Discrete Math.* **40**, (1982), 315—318.
[9] J. SCHMIDT and E. SHAMIR, A threshold for perfect matchings in *d*-pure random hypergraphs, *Discrete Math.*, **45** (1983), 287—295.
[10] J. SCHMIDT-PRUZAN, E. SHAMIR, and E. UPFAL, Random hypergraph coloring algorithms and the weak chromatic number, *Journal of Graph Theory (to appear)*, (Preliminary version: Technical Report (CS83-09) Dept. Appl. Math. Weizmann Inst. of Sc. Rehovot Israel).
[11] J. SCHMIDT-PRUZAN, Probabilistic analysis of strong hypergraph coloring algorithms and the strong chromatic number. *Submitted to Discrete Math.*, (Preliminary version: Technical Report (CS83-10) Dept. Appl. Math. Weizmann Inst. of Sc. Rehovot Israel).
[12] I. TOMESCU, Asymptotical estimations for the number of cliques of uniform hypergraphs, *Annals of Discrete Math.*, **11**, *Studies on graphs and Discrete Programming* (1981), (Edited by P. Hansen), North-Holland Publishing Comp., 345—358.

Jeanette Schmidt-Pruzan, Eli Shamir

Department of Applied Mathematics
The Weizmann Institute of Science
Rehovot, 76100, Israel

and

The Institute of Mathematics and Comp. Sci.
The Hebrew University, Jerusalem, Israel