Phase Transition of Random Non-Uniform Hypergraphs☆

Élie de Panafieu

Research Institute for Symbolic Computation (RISC), J. Kepler University Linz, Austria

Abstract

Non-uniform hypergraphs appear in various domains of computer science as in the satisfiability problems and in data analysis. We analyse a general model where the probability for an edge of size t to belong to the hypergraph depends of a parameter ω_t of the model. It is a natural generalization of the models of graphs presented in [1] and in [2]. The present paper follows the same general approach based on analytic combinatorics. We show that many analytic tools developed for the analysis of graphs can be extended surprisingly well to non-uniform hypergraphs. More specifically, we analyze their typical structure before and near the birth of the *complex* components, that are the connected components with more than one cycle, and derive the asymptotic number of sparse connected hypergraphs as their complexity, defined as the *excess*, increases. Although less natural than the number of edges, this parameter allows a precise description of the structure of hypergraphs. Finally, we compute some statistics of the model to link number of edges and excess.

Keywords: Hypergraph, phase transition, analytic combinatorics

1. Introduction

In the seminal article [3], Erdös and Rényi discovered an abrupt change of the structure of a random graph when the number of edges reaches half the number of vertices. It corresponds to the emergence of the first connected component with more than one cycle, immediately followed by components with even more cycles. The combinatorial analysis of those components improves the understanding of the objects modeled by graphs and has application in the analysis and the conception of graph algorithm. The same motivation holds for hypergraphs which are used, among others, to represent databases and xor-formulas.

☆This work was partially founded by the ANR Boole, the ANR Magnum and the University Paris Diderot, Sorbonne Paris Cité.

This is the long version of the homonym paper accepted in the proceedings of IWOCA 2013.

Much of the literature on hypergraphs is restricted to the uniform case, where all the edges contain the same number of vertices. In particular, the analysis of the birth of the complex component in terms of the size of the component and the order of the phase transition can be found in [4], [5], [6], [7] and [8].

There is no canonical choice for the size of a random edge in a hypergraph; thus several models have been proposed. One is developed in [9], where the size of the largest connected component is obtained using probabilistic methods. It is our opinion that to be general, a non-uniform hypergraph model needs one parameter for each possible size of edges, in order to quantify how often those edges appear. In [10], Darling and Norris define such a model, the Poisson random hypergraphs model, and analyze its structure via fluid limits of pure jump-type Markov processes.

We have not found in the literature much use of the generating function of non-uniform hypergraphs to investigate their structure, and we intend to fill this gap. However, similar generating functions have been derived in [11] for a different purpose: Gessel and Kalikow use it to give a combinatorial interpretation for a functional equation of Bouwkamp and de Bruijn. The underlying hypergraph model is a natural generalization of the *multigraph process*.

In Section 2 we introduce the hypergraph models, the probability distribution and the corresponding generating functions. The important notion of *excess* is also defined. Section 3 is dedicated to the asymptotic number of hypergraphs with n vertices and excess k. Some statistics on the random hypergraphs are derived, including the limit distribution of the number of edges under some technical condition. Section 4 focuses on hypergraphs with small excess, which are composed only of trees and unicycle components with high probability. The critical excess at which the first complex component appears is obtained in Section 5. For a range of excess near and before this critical value, we compute the probability that a random hypergraph contains no complex component. The classical notion of *kernel* is introduced for hypergraphs in Section 6. It is then used to derive the asymptotic of connected hypergraphs with n vertices and fixed excess k.

We derive in Section 7 the structure of random hypergraphs in the critical window, and obtain a surprising result: although the critical excess is generally different for graphs and hypergraphs, both models share the same structure distribution exactly at their respective critical excess. Finally, we give an intuitive explanation of the birth of the *giant component* in Section 8.

2. Presentation of the Model

In this paper, a hypergraph G is a multiset $E(G)$ of $m(G)$ edges. Each edge e is a multiset of $|e|$ vertices in $V(G)$, where $|e| \geq 2$. The vertices of the hypergraph are labelled from 1 to $n(G)$. We also set $l(G)$ for the *size* of G, defined by

$$l(G) = \sum_{e \in E(G)} |e| = \sum_{v \in V(G)} \deg(v).$$

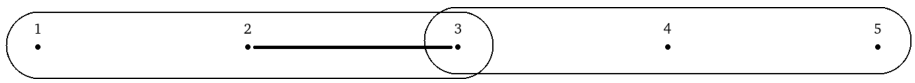

Figure 1: Hypergraph with $n = 5$ vertices, $m = 3$ edges, excess $k = 0$, size $l = 8$ and $\mathrm{NumbSeq} = 432$. There is one cycle, which links the vertices 2 and 3.

Those notions are illustrated in figure 1.

The notion of *excess* was first used for graphs in [12], then named in [2], and finally extended to hypergraphs in [13]. The excess of a connected component C is always greater or equal to -1. It expresses how far from a tree it is: C is a tree if and only if its excess is -1, contains exactly one cycle if its excess is 0, and is said to be *complex* if its excess is strictly positive. Intuitively, a connected component with high excess is "hard" to treat for a backtracking algorithm. The excess $k(G)$ of a hypergraph G is defined by

$$k(G) = l(G) - n(G) - m(G).$$

A hypergraph may contain several copies of the same edge and a vertex may appear more than once in an edge; thus we are considering multihypergraphs. A hypergraph with no loop nor multiple edge is said to be *simple*. Let us recall that a sequence is by definition an ordered multiset. We define $\mathrm{NumbSeq}(G)$ as the number of sequences of nonempty sequences of vertices that lead to G. For example, one of the sequences that lead to the hypergraph of figure 1 is $(5, 3, 4), (3, 2), (2, 1, 3)$, but $(5, 1, 4), (1, 2), (2, 3, 1)$ would describe a different hypergraph. If G is simple, then $\mathrm{NumbSeq}(G)$ is equal to $m(G)! \prod_{e \in E(G)} |e|!$, otherwise it is smaller. We associate to any family $\mathcal{F}$ of hypergraphs the generating function

$$F(z, w, x) = \sum_{G \in \mathcal{F}} \frac{\mathrm{NumbSeq}(G)}{m(G)!} \left(\prod_{e \in E(G)} \frac{\omega_{|e|}}{|e|!} \right) w^{m(G)} x^{l(G)} \frac{z^{n(G)}}{n(G)!} \tag{1}$$

where ω_t marks the edges of size t, w the edges, x the size of the graph and z the vertices. Therefore, we count hypergraphs with a *weight* κ

$$\kappa(G) = \frac{\mathrm{NumbSeq}(G)}{m(G)!} \prod_{e \in E(G)} \frac{\omega_{|e|}}{|e|!} \tag{2}$$

that is the extension to hypergraphs of the *compensation factor* defined in Section 1 of [2]. If $\mathcal{F}$ is a family of simple hypergraphs, then we obtain the simpler and natural expression

$$F(z, w, x) = \sum_{G \in \mathcal{F}} \Big(\prod_{e \in E(G)} \omega_{|e|} \Big) w^{m(G)} x^{l(G)} \frac{z^{n(G)}}{n(G)!}. \tag{3}$$

Remark that the generating function of the subfamily of hypergraphs of excess k is $[y^k] F\left(z/y, w/y, xy\right)$, where $[x^n] \sum_k a_k x^k$ denotes the coefficient a_n.

We define the exponential generating function of the edges as

$$\Omega(z) := \sum_{t \geq 2} \omega_t \frac{z^t}{t!}.$$

From now on, the (ω_t) are considered as a bounded sequence of nonnegative real numbers with $\omega_0 = \omega_1 = 0$. The value ω_t represents how likely an edge of size t is to appear. Thus, for graphs we get $\Omega(z) = z^2/2$, for d-uniform hypergraphs (*i.e.* with all edges of size d) we have $\Omega(z) = z^d/d!$, for hypergraphs with sizes of edges restricted to a set S we have $\Omega(z) = \sum_{s \in S} z^s$ and for hypergraphs with weight 1 for all size of edge $\Omega(z) = e^z - 1 - z$. To simplify the saddle point proofs, we also suppose that $\Omega(z)/z$ cannot be written as $f(z^d)$ for an integer $d > 1$ and a power serie f with a non-zero radius of convergence. This implies that $e^{\Omega(z)/z}$ is aperiodic. Therefore, we do not treat the important, but already studied, case of d-uniform hypergraphs for $d > 2$.

The generating function of all hypergraphs is

$$\mathrm{hg}(z, w, x) = \sum_n e^{w\,\Omega(nx)} \frac{z^n}{n!}. \tag{4}$$

This expression can be derived from (1) or using the symbolic method presented in [14]. Indeed, $\Omega(nx)$ represents an edge of size marked by x and n possible types of vertices, and $e^{w\,\Omega(nx)}$ a set of edges. For the family of simple hypergraphs,

$$\mathrm{shg}(z, w, x) = \sum_n \left(\prod_t (1 + \omega_t x^t w)^{\binom{n}{t}} \right) \frac{z^n}{n!}. \tag{5}$$

Similar expressions have been derived in [11]. The authors use them to give a combinatorial interpretation of a functional equation of Bouwkamp and de Bruijn.

A hypergraph with n vertices and m labelled edges can be represented by a (n, m)-matrix M with nonnegative integer coefficients, the coefficient $M_{v,e}$ being the number of occurences of the vertex v in the edge e. In this representation, multigraphs correspond to matrices where the sum of the coefficients on each column is equal to 2. Simple hypergraphs correspond to matrices with $\{0, 1\}$ coefficients that do not contain two identical columns. Let us consider a hypergraph G and a matrix representation M of it. A hypergraph H is said to be the *dual* of G if the transpose M^T of M represents it. In other words, H is obtained from G by reversing the roles of vertices and edges, of degrees and sizes of the edges. Therefore, the choice of weighting the edges depending of their size can be transposed into weights on the vertices with respect to their degrees. Figure 2 displays a dual of the hypergraph of figure 1. This notion will be usefull in the proof of Theorem 8.

Comparing (1) with (3), simple hypergraphs may appear more natural than hypergraphs. But their generating function is more intricate, their matrix representations satisfy more complex constraints and the asymptotic results on hypergraphs can often be extended to simple hypergraphs. Furthermore, experience

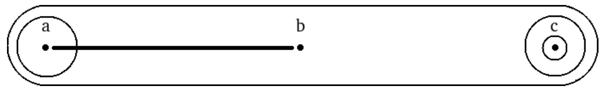

Figure 2: One of the duals of the hypergraph of figure 1. The vertex c, of degree 3, corresponds to the edge $(3, 4, 5)$, of size 3, in figure 1.

has shown that multigraphs appear as often as simple graphs in applications. This is why we do not confine our study to simple hypergraphs.

So far, we have adopted an enumerative approach of the model, but there is a corresponding probabilistic description. Let us define $\mathrm{HG}_{n,k}$ (resp. $\mathrm{SHG}_{n,k}$) as the set of hypergraphs (resp. simple hypergraphs) with n vertices and excess k, equipped with the probability distribution induced by the weights (2). Therefore, the hypergraph G occurs with probability $\kappa(G)/\sum_{H\in\mathrm{HG}_{n,k}}\kappa(H)$.

3. Hypergraphs with n Vertices and Excess k

In this section, we derive the asymptotic of hypergraphs and simple hypergraphs with n vertices and global excess k. This result is interesting by itself and is a first step to find the excess k at which the first component with strictly positive excess is likely to appear. Statistics on the number of edges are also derived.

Theorem 1. *Let λ be a strictly positive real value and $k = (\lambda - 1)n$, then the sum of the weights of the hypergraphs in $\mathrm{HG}_{n,k}$ is*

$$\mathrm{hg}_{n,k} \sim \frac{n^{n+k}}{\sqrt{2\pi n}}\frac{e^{\frac{\Omega(\zeta)}{\zeta}n}}{\zeta^{n+k}}\frac{1}{\sqrt{\zeta\,\Omega''(\zeta)-\lambda}}$$

where $\Psi(z)$ denotes the function $\Omega'(z) - \frac{\Omega(z)}{z}$ and ζ is defined by $\Psi(\zeta) = \lambda$. A similar result holds for simple hypergraphs:

$$\mathrm{shg}_{n,k} \sim \frac{n^{n+k}}{\sqrt{2\pi n}}\frac{e^{\frac{\Omega(\zeta)}{\zeta}n}}{\zeta^{n+k}}\frac{\exp\left(-\frac{\omega_2^2\zeta^2}{4}-\frac{\zeta\,\Omega''(\zeta)}{2}\right)}{\sqrt{\zeta\,\Omega''(\zeta)-\lambda}}.$$

More precisely, if $k = (\lambda-1)n+xn^{2/3}$ where x is bounded, then the two previous asymptotics are multiplied by a factor $\exp\left(\frac{-x^2}{2(\zeta\,\Omega''(\zeta)-\lambda)}n^{1/3}+\frac{x^3}{6}\frac{\zeta^2\,\Omega'''(\zeta)+\lambda}{(\zeta\,\Omega''(\zeta)-\lambda)^3}\right)$.

Proof. With the convention (1), the sum of the weights of the hypergraphs with n vertices and excess k is

$$n![z^n y^k]\,\mathrm{hg}(z/y, 1/y, y) = n![z^n y^k]\sum_n e^{\frac{\Omega(ny)}{y}}\frac{(z/y)^n}{n!} = n^{n+k}[y^{n+k}]e^{\frac{\Omega(y)}{y}n}.$$

The asymptotic is then extracted using the *large power* scheme presented in [14]. Remark that $\Psi(z) = \sum_t \omega_t (t-1) \frac{z^{t-1}}{t!}$ has nonnegative coefficients, so there is a unique solution of $\Psi(\zeta) = \lambda$, and that $\Psi(\zeta) = \lambda$ implies $\zeta\,\Omega''(\zeta) - \lambda > 0$.

For simple hypergraphs, the coefficient we want to extract from (5) is now

$$[y^{n+k}] \prod_t (1 + \omega_t y^{t-1})^{\binom{n}{t}} = \frac{n^{n+k}}{2i\pi} \oint \exp\left(\textstyle\sum_t \binom{n}{t} \log\left(1 + \omega_t \left(\frac{y}{n}\right)^{t-1}\right)\right) \frac{dy}{y^{n+k+1}}.$$

The sum in the exponential can be rewritten

$$\frac{\Omega(y)}{y} n + \sum_t \binom{n}{t} \left(\log(1 + \omega_t \left(\frac{y}{n}\right)^{t-1}) - \omega_t \left(\frac{y}{n}\right)^{t-1}\right) - \left(\frac{n^t}{t!} - \binom{n}{t}\right) \omega_t \left(\frac{y}{n}\right)^{t-1}$$

which is $\frac{\Omega(y)}{y} n - \frac{\omega_2^2 y^2}{4} - \frac{y\,\Omega''(y)}{2} + \mathcal{O}(1/n)$ when y is bounded (we use here the hypothesis that $\omega_0 = \omega_1 = 0$). In the saddle point method, y is close to ζ, which in our case is fixed with respect to n. Therefore,

$$n![z^n y^k] \operatorname{shg}\left(\frac{z}{y}, \frac{1}{y}, y\right) \sim \exp\left(-\frac{\omega_2^2 \zeta^2}{4} - \frac{\zeta\,\Omega''(\zeta)}{2}\right) \mathrm{hg}_{n,k}\,.$$

The constraint $k = (\lambda - 1)n + x n^{2/3}$ is equivalent to $k = (\bar\lambda - 1)n$ with $\bar\lambda = \lambda + x n^{-1/3}$. Since x is bounded, so is $\bar\lambda$ and the first part of the theorem can be applied. Let us consider the solution $\bar\zeta$ of $\Psi(\bar\zeta) = \bar\lambda$. With the help of maple, we find

$$\frac{e^{n \frac{\Omega(\bar\zeta)}{\bar\zeta}}}{\bar\zeta^{n+k}} = \frac{e^{n \frac{\Omega(\zeta)}{\zeta}}}{\zeta^{n+k}} \exp\left(-\frac{x^2 n^{1/3}}{2(\zeta\,\Omega''(\zeta) - \lambda)} + \frac{x^3}{6} \frac{\zeta\,\Omega'''(\zeta) + \lambda}{(\zeta\,\Omega''(\zeta) - \lambda)^3} + \mathcal{O}(n^{-1/3})\right).$$

The factor $\exp\left(-\frac{\omega_2^2\zeta^2}{4} - \frac{\zeta\,\Omega''(\zeta)}{2}\right)$ is the asymptotic probability for a hypergraph in $\mathrm{HG}_{n,k}$ to be simple. For graphs, with $\Omega(z) = z^2/2$ and $\lambda = 1/2$, we obtain the same factor $e^{-3/4}$ as in [2].

We study the evolution of hypergraphs as their excess increases. This choice of parameter is less natural than the number of edges, but it significantly simplifies the equations. On the other hand, we can compute statistics on the number of edges of hypergraphs with n edges and excess k.

Theorem 2. *Let $\Psi(z)$ and ζ be defined as in Theorem 1, and G a random hypergraph in $\mathrm{HG}_{n,k}$ or in $\mathrm{SHG}_{n,k}$ with $k = (\lambda - 1)n$, then the number m of edges of G admits a limit law that is gaussian with parameters*

$$\mathbb{E} = \frac{\Omega(\zeta)}{\zeta} n,$$
$$\mathbb{V} = \left(\frac{\Omega(\zeta)}{\zeta} - \frac{\lambda^2}{\zeta\,\Omega''(\zeta) - \lambda}\right) n$$

if $\mathbb{V}$ is non-zero. Whether this condition is satisfied or not, the asymptotic expectations and factorial moments of the number m of edges are

$$\mathbb{E}_{n,k}(m) \sim \tfrac{\Omega(\zeta)}{\zeta} n,$$
$$\forall t \geq 0,\ \mathbb{E}_{n,k}\big(m(m-1)\dots(m-t)\big) \sim \left(\tfrac{\Omega(\zeta)}{\zeta} n\right)^{t+1}.$$

Reversely, the expectation and variance of the excess k of a random hypergraph with n vertices and m edges are

$$\mathbb{E}_{n,m}(k) = nm\frac{\Omega'(n)}{\Omega(n)} - n - m,$$
$$\mathbb{V}_{n,m}(k) = \frac{nm}{\Omega(n)} \left(n\,\Omega''(n) - n\frac{\Omega'(n)^2}{\Omega(n)} + \Omega'(n)\right).$$

Proof. Let us recall that if p_t denotes the probability that a discrete random variable X takes the value t and $f(z) = \sum_n p_n z^n$, then the expectation of X is $f'(1)$ and its kth factorial moment is $\mathbb{E}(X(X-1)\dots(X-k)) = \partial^{t+1} f(1)$. By extraction from (4), the generating functions of the hypergraphs with n vertices and excess k (resp. m edges) and of the simple hypergraphs in $\mathrm{SHG}_{n,k}$ are

$$\mathrm{hg}_{n,k}(w) = n^{n+k}[y^{n+k}]e^{w\frac{\Omega(y)}{y}n},$$
$$\mathrm{hg}_{n,m}(y) = \frac{\Omega(ny)^m}{y^{n+m}m!},$$
$$\mathrm{shg}_{n,k}(w) = n^{n+k}[y^{n+k}]e^{w\frac{\Omega(y)}{y}n}e^{-\frac{y\,\Omega''(y)}{2}w-\frac{\omega_2^2 y^2}{4}w^2+\mathcal{O}(1/n)}$$

where w and y mark respectively the number of edges and the excess. Therefore, the probability generating function corresponding to the distribution of m is $\mathrm{hg}_{n,k}(w)/\mathrm{hg}_{n,k}(1)$, and similarly for k. The asymptotics are then derived as in the proof of Theorem 1.

To prove the normal limit distribution of the number of edges in $\mathrm{HG}_{n,k}$, we write

$$\frac{\mathrm{hg}_{n,k}(e^s)}{\mathrm{hg}_{n,k}} = e^{nA(s)+B(s)}\left(1+\mathcal{O}\left(n^{-1/2}\right)\right)$$

where

$$A(s) = \frac{\Omega(\zeta)}{\zeta}s + \left(\frac{\Omega(\zeta)}{\zeta} - \frac{\lambda^2}{\zeta\,\Omega''(\zeta)-\lambda}\right)\frac{s^2}{2} + \mathcal{O}(s^3)$$

and apply a lemma of Hwang [15] that can also be found in [14] as Lemma IX.1.

The condition $\mathbb{V} \neq 0$ for the limit law of the number of edges in a hypergraph from $\mathrm{HG}_{n,k}$ to be gaussian is not always satisfied. For example, the variance for graphs is 0, since all the graphs with n vertices and excess k have exactly $k+n$ edges.

4. Subcritical Hypergraphs

We follow the conventions established in [16]: a *walk* of a hypergraph G is a sequence $v_0, e_1, v_1, \ldots, v_{t-1}, e_t, v_t$ where for all i, $v_i \in V(G)$, $e_i \in E(G)$ and $\{v_{i-1}, v_i\} \subset e_i$. A *path* is a walk in which all v_i and e_i are distinct. A walk is a cycle if all v_i and e_i are distinct, except $v_0 = v_t$. Connectivity, trees and rooted trees are then defined in the usual way.

A *unicycle component* is a connected hypergraph that contains exactly one cycle. We also define a *path of trees* as a path that contains no cycle, plus a rooted tree hooked to each vertex, except to the two ends of the path. It can equivalently be defined as an unrooted tree with two distinct marked leaves.

Lemma 3. *Let T, U, V and P denote the generating functions of rooted trees, unrooted trees, unicycle components and paths of trees, using the variable z to mark the number of vertices, then*

$$T(z) = z e^{\Omega'(T(z))}, \tag{6}$$

$$U(z) = T(z) + \Omega(T(z)) - T(z)\,\Omega'(T(z)), \tag{7}$$

$$V(z) = \frac{1}{2} \log \frac{1}{1 - T(z)\,\Omega''(T(z))}, \tag{8}$$

$$P(z) = \frac{\Omega''(T(z))}{1 - T(z)\,\Omega''(T(z))}. \tag{9}$$

PROOF. Those expressions can be derived using the *symbolic method* presented in [14]. The generating function of edges is $\Omega(z)$. If one vertex is marked, it becomes $z\,\Omega'(z)$ and $z\,\Omega''(z)$ if another vertex is deleted. Equation (6) means that a rooted tree is a vertex (the root) and a set of edges from which a vertex has been removed and the other vertices replaced by rooted trees. Equation (7) is a classical consequence of the dissymmetry theorem described in [17] and studied in [18]. It can be checked that $z\partial_z U = T$, which, in a symbolic method, means that a tree with a vertex marked is a rooted tree. Unicycle components are cycles of rooted trees, which implies (8).

Combining the enumeration of hypergraphs with the enumeration of forests, we can investigate the birth of the first cycle and the limit distribution of the number of cycles in a hypergraph with small excess.

Theorem 4. *Let $\Psi(z)$ denote the function $\Omega'(z) - \frac{\Omega(z)}{z}$, τ be implicitly defined by $\tau\,\Omega''(\tau) = 1$ and $\Lambda = \Psi(\tau)$. Let us consider an excess $k = (\lambda - 1)n$ where $0 < \lambda < \Lambda$ and the value ζ such that $\Psi(\zeta) = \lambda$.*

With high probability, a hypergraph in $\mathrm{HG}_{n,k}$ or $\mathrm{SHG}_{n,k}$ contains no component with two cycles. The limit distribution of the number of cycles of such a hypergraph follows a Poisson law of parameter

$$\frac{1}{2} \log \left(\frac{1}{1 - \zeta\,\Omega''(\zeta)} \right)$$

if the hypergraph is in $\mathrm{HG}_{n,k}$*, and*

$$\frac{1}{2}\log\left(\frac{1}{1-\zeta\,\Omega''(\zeta)}\right)-\frac{\omega_2^2\zeta^2}{4}-\frac{\zeta\,\Omega''(\zeta)}{2}$$

if it is in $\mathrm{SHG}_{n,k}$*.*

Proof. Let $\mathrm{THG}_{n,k}$ denote the set of hypergraphs in $\mathrm{HG}_{n,k}$ that contains only trees and unicycle components. The excess of a tree is -1, the excess of a unicycle component is 0. Since the excess of a hypergraph is the sum of the excesses of its components, each hypergraph in $\mathrm{THG}_{n,k}$ contains exactly $-k$ trees. The generating function of the number of cycles in hypergraphs of $\mathrm{THG}_{n,k}$ is

$$n![z^n]\frac{U(z)^{-k}}{(-k)!}e^{uV(z)}=\frac{n!}{(-k)!}\frac{1}{2i\pi}\oint(T+\Omega(T)-T\,\Omega'(T))^{-k}e^{\frac{u}{2}\log\left(\frac{1}{1-T\,\Omega''(T)}\right)}\frac{dz}{z^{n+1}}$$

where u marks the cycles. We use the *large power* Theorem VIII.8 of [14] to extract the asymptotic. For $k=(\lambda-1)n$, after the change of variable $z\to T$, the dominant saddle point is characterized by $\Psi(\zeta)=\lambda$. The computations lead to

$$n![z^n]\frac{U(z)^{-k}}{(-k)!}e^{uV(z)}\sim\frac{n^{n+k}}{\sqrt{2\pi n}}\frac{e^{\frac{\Omega(\zeta)}{\zeta}n}}{\zeta^{n+k}}\frac{e^{\frac{u-1}{2}\log\left(\frac{1}{1-T\,\Omega''(T)}\right)}}{\sqrt{\zeta\,\Omega''(\zeta)-\lambda}}.$$

Dividing by the cardinality of $\mathrm{HG}_{n,k}$ derived in Theorem 1, we obtain the generating function of the limit probabilities of the number of cycles in $\mathrm{THG}_{n,k}$:

$$\sum_t\mathbb{P}(G\in\mathrm{THG}_{n,k}\text{ and has }t\text{ cycles}\mid G\in\mathrm{HG}_{n,k})u^t=e^{\frac{u-1}{2}\log\left(\frac{1}{1-T\,\Omega''(T)}\right)}.$$

For $u=1$, it is equal to 1, so with probability tending to 1, a hypergraph in $\mathrm{HG}_{n,k}$ has no component with more than one cycle. For $u=e^{it}$, we recognize the characteristic function of a Poisson law with parameter $\frac{1}{2}\log\left(\frac{1}{1-T\,\Omega''(T)}\right)$.

The same computations hold for the analysis of simple hypergraphs, except the generating function $V(z)$ has to be replaced by $V(z)-\frac{T\,\Omega''(T)}{2}-\frac{\omega_2^2T^2}{4}$ to avoid loops and multiple edges (in unicycle components, those can only be two edges of size 2).

More informations on the length of the first cycle and the size of the component that contains it could be extracted, following the approach of [1].

5. Birth of the complex components

Let us recall that a connected hypergraph is *complex* if its excess is strictly positive. In order to locate the global excess k at which the first complex component appears, we compare the asymptotic numbers of hypergraphs and hypergraphs with no complex component.

Theorem 6 describes the limit probability for a hypergraph not to contain any complex component. A phase transition occurs when $\frac{k}{n}$ reaches the critical value $\Lambda-1$, defined in Theorem 4. On an analytic point of view, this corresponds to the coalescence of two saddle points. In this context, the large power scheme ceases to apply, so we replace it by the following general theorem, borrowed from [19] (see also Theorem IX.16 of [14] for discussions and links with the *stable laws* of probability theory) and adapted for our purpose (in the original theorem, $\mu = 0$). It is also close to Lemma 3 of [2].

Theorem 5. *We consider a generating function $H(z)$ with nonnegative coefficients and a unique isolated singularity at its radius of convergence ρ. We also assume that it is continuable in $\Delta := \{z \mid |z| < R, z \notin [\rho, R]\}$ and there is a $\lambda \in]1;2[$ such that $H(z) = \sigma - h_1(1-z/\rho) + h_\lambda(1-z/\rho)^\lambda + \mathcal{O}((1-z/\rho)^2)$ as $z \to \rho$ in Δ. Let $k = \frac{\sigma}{h_1}n + xn^{1/\lambda}$ with x bounded, then for any real constant μ*

$$[z^n]\frac{H^k(z)}{(1-z/\rho)^\mu} \sim \sigma^k \rho^{-n} \frac{1}{n^{(1-\mu)/\lambda}} (h_1/h_\lambda)^{(1-\mu)/\lambda} G\left(\lambda, \mu; \frac{h_1^{1+1/\lambda}}{\sigma h_\lambda^{1/\lambda}} x\right) \qquad (10)$$

where $G(\lambda, \mu; x) = \frac{1}{\lambda\pi} \sum_{k \geq 0} \frac{(-x)^k}{k!} \sin\left(\pi \frac{1-\mu+k}{\lambda}\right) \Gamma\left(\frac{1-\mu+k}{\lambda}\right)$.

Proof. In the Cauchy integral that represents $[z^n]\frac{H^k(z)}{(1-z/\rho)^\mu}$ we choose for the contour of integration a positively oriented loop, made of two rays of angle $\pm\pi/(2\lambda)$ that intersect on the real axis at $\rho - n^{-1/\lambda}$, we set $z = \rho(1 - tn^{-1/\lambda})$

$$[z^n]\frac{H^k(z)}{(1-z/\rho)^\mu} \sim \frac{-\sigma^k \rho^{-n}}{2i\pi n^{(1-\mu)/\lambda}} \int t^{-\mu} e^{\frac{h_\lambda}{h_1} t^\lambda} e^{-x \frac{h_1}{\sigma} t} dt$$

The contour of integration comprises now two rays of angle $\pm\pi/\lambda$ intersecting at -1. Setting $u = t^\lambda h_\lambda / h_1$, the contour transforms into a classical Hankel contour, starting from $-\infty$ over the real axis, winding about the origin and returning to $-\infty$.

$$\frac{-\sigma^k \rho^{-n}}{2i\pi n^{(1-\mu)/\lambda}} \frac{1}{\lambda} (h_1/h_\lambda)^{(1-\mu)/\lambda} \int_{-\infty}^{(0)} e^u e^{-xu^{1/\lambda} h_1^{1+1/\lambda}/(\sigma h_\lambda^{1/\lambda})} u^{\frac{1-\mu}{\lambda}-1} du$$

Expanding the exponential, integrating termwise, and appealing to the complement formula for the Gamma function finally reduces this last form to (10).

Theorem 6. *Let $\Psi(z)$, τ and Λ be defined as in Theorem 4, $G(\lambda, \mu; x)$ as in Theorem 5 and $\gamma = 1 + \tau^2\,\Omega'''(\tau)$. We consider an excess $k = (\Lambda-1)n + xn^{2/3}$ where x is bounded. Then the sum of the weights of the hypergraphs in $\mathrm{HG}_{n,k}$ with no complex component is equivalent to*

$$\frac{n^{n+k}}{\sqrt{2\pi n}} \frac{e^{\frac{\Omega(\tau)}{\tau} n}}{\tau^{n+k}} \frac{1}{\sqrt{1-\Lambda}} \sqrt{\frac{3\pi}{2}} e^{-\frac{x^2}{2(1-\Lambda)} n^{1/3} - \frac{x^3}{6(1-\Lambda)^2}} G\left(\frac{3}{2}, \frac{1}{4}; -\frac{3^{2/3}\gamma^{1/3} x}{2(1-\Lambda)}\right). \qquad (11)$$

For simple hypergraphs, this sum is

$$\frac{n^{n+k}}{\sqrt{2\pi n}} \frac{e^{\frac{\Omega(\tau)}{\tau}n}}{\tau^{n+k}} \frac{\exp\left(-\frac{1}{2} - \frac{\omega_2^2\tau^2}{4}\right)}{\sqrt{1-\Lambda}} \sqrt{\frac{3\pi}{2}} e^{-\frac{x^2}{2(1-\Lambda)}n^{1/3} - \frac{x^3}{6(1-\Lambda)^2}} G\left(\frac{3}{2}, \frac{1}{4}; -\frac{3^{2/3}\gamma^{1/3}x}{2(1-\Lambda)}\right).$$

Proof. For $k = (\lambda - 1)n$, there are two saddle points in T: one implicitly defined by $\Psi(\zeta) = \lambda$ and the other at $\zeta = \tau$. Those two saddle points coalesce when $\lambda = \Psi(\tau)$. When k is around its critical value $(\Lambda - 1)n$, we apply Theorem 5. The Newton-Puiseux expansions of T, e^V and U can be derived from Lemma 3

$$T(z) \sim \tau - \tau\sqrt{\frac{2}{\gamma}}\sqrt{1 - z/\rho},$$
$$e^{V(z)} \sim (2\gamma)^{-1/4}(1 - z/\rho)^{-1/4},$$
$$U(z) = \tau(1 - \Psi(\tau)) - \tau(1 - z/\rho) + \tau\frac{2}{3}\sqrt{\frac{2}{\gamma}}(1 - z/\rho)^{3/2} + \mathcal{O}(1 - z/\rho)^2,$$

where $\rho = \tau e^{-\Omega'(\tau)}$. Using Theorem 5, we obtain

$$\mathrm{thg}_{n,k} \sim \frac{n!}{(-k)!} \frac{\sqrt{3}}{2} \frac{(\tau(1-\Lambda))^{-k}}{\rho^n \sqrt{n}} G\left(\frac{3}{2}, \frac{1}{4}; -\frac{3^{2/3}\gamma^{1/3}x}{2(1-\Lambda)}\right)$$

which reduces to (11).

As in the proof of Theorem 4, in the analysis of simple hypergraphs, the generating function $V(z)$ is replaced by $V(z) - \frac{T\,\Omega''(T)}{2} - \frac{\omega_2^2 T^2}{4}$.

Combining Theorems 1 and 6, we deduce that when $k = (\lambda - 1)n + \mathcal{O}(n^{1/3})$ with $\lambda < \Lambda$, the probability that a random hypergraph in $\mathrm{HG}_{n,k}$ has no complex component approaches 1 as n tends towards infinity. When $k = (\Lambda - 1)n + \mathcal{O}(n^{1/3})$, this limit becomes $\sqrt{2/3}$ because $G(2/3, 1/4; 0)$ is equal to $2/(3\sqrt{\pi})$. It is remarkable that this value does not depend on Ω, therefore it is the same as in [1] for graphs. However, the evolution of this probability between the subcritical and the critical ranges of excess depends on the (ω_t).

Corollary 7. *Let τ, Λ and γ be defined as in Theorem 6. For $k = (\lambda - 1)n$ and $\lambda < \Lambda$, a hypergraph in $\mathrm{HG}_{n,k}$ or in $\mathrm{SHG}_{n,k}$ has no complex component with high probability. For $k = (\Lambda - 1)n + xn^{2/3}$ with x bounded, the limit probability that such a hypergraph has no complex component is*

$$\sqrt{\frac{3\pi}{2}} \exp\left(\frac{-x^3\gamma}{6(1-\Lambda)^3}\right) G\left(\frac{3}{2}, \frac{1}{4}; -\frac{3^{2/3}\gamma^{1/3}x}{2(1-\Lambda)}\right)$$

where G is the function defined in Theorem 5.

Proof. From the second assertion of Theorem 1 we deduce the asymptotic number of hypergraphs in $\mathrm{HG}_{n,k}$ when $k = (1 - \Lambda)n + xn^{2/3}$

$$\mathrm{hg}_{n,k} \sim \frac{n^{n+k}}{\sqrt{2\pi n}} \frac{e^{\frac{\Omega(\tau)}{\tau}n}}{\tau^{n+k}} \frac{e^{\frac{-x^2}{2(1-\Lambda)}n^{1/3} + \frac{x^3}{6}\frac{\gamma-1+\Lambda}{(1-\Lambda)^3}}}{\sqrt{1-\Lambda}}.$$

Equation 11 divided by this estimation of $\mathrm{hg}_{n,k}$ leads to the result announced. The computations are the same to simple hypergraphs.

Theorem 5 does not apply when $H(z)$ is periodic. This is why we restricted $\Omega(y)/y$ not to be of the form $f(z^d)$ where $d > 1$ and $f(z)$ is a power serie with a strictly positive radius of convergence. An unfortunate consequence is that Theorems 1 and 6 do not apply to the important but already analyzed case of d-uniform hypergraphs. However, the expression of the critical excess is still valid. For the d-uniform hypergraphs, $\Omega(z) = \frac{z^d}{d!}$, $\Psi(z) = \frac{(d-1)}{d!} z^{d-1}$ and $\tau^{d-1} = (d-2)!$, so we obtain $k = \frac{1-d}{d} n$ for the critical excess, which corresponds to a number of edges $m = \frac{n}{d(d-1)}$, a result already derived in [9].

6. Kernels and Connected Hypergraphs

In the seminal articles [12] and [20], Wright establishes the connection between the asymptotic of connected graphs with n vertices and excess k and the enumeration of the connected *kernels*, which are multigraphs with no vertex of degree less than 3. This relation was then extensively studied in [2] and the notions of excess and kernels were extended to hypergraphs in [13].

A kernel is a hypergraph with additional constraints that ensure that:

- each hypergraph can be reduced to a unique kernel,
- the excesses of a hypergraph and its kernel are equal,
- for any integer k, there is a finite number of kernels of excess k,
- the generating function of hypergraphs of excess k can be derived from the generating function of kernels of excess k.

Remark that the two last requirements oppose each other: the third one impose the kernels to be *simple*, but the fourth one means they should keep trace of the structure of the hypergraph. Following [13], we define the *kernel* of a hypergraph G as the result of the repeated execution of the following operations:

1. delete all the vertices of degree ≤ 1,
2. delete all the edges of size ≤ 1,
3. if two edges (a, v) and (v, b) of size 2 have one common vertex v of degree 2, delete v and replace those edges by (a, b),
4. delete the connected components that consist of one vertex v of degree 2 and one edge (v, v) of size 2.

The following theorem has already been derived for uniform hypergraphs in [13]. We give a new and more general proof. We also define *clean* kernels and derive an expression for their generating function. As we will see in the proof of Theorem 10, with high probability the kernel of a random hypergraph in the critical window is clean.

Theorem 8. *The number of kernels of excess k is finite and each of them contains at most $3k$ edges of size 2. We say that a kernel is* clean *if this bound is reached. The generating functions of connected clean kernels of excess k is*

$$c_k(1+\omega_3 z^2)^{2k}\omega_2^{3k}z^{2k} \tag{12}$$

where $c_k = [z^{2k}]\log\sum_n \frac{(6n)!}{(3!)^{2n}2^{3n}(3n)!}\frac{z^{2n}}{(2n)!}$ and the variables w and x have been omitted.

Proof. By definition, $k+n+m = \sum_{e\in E}|e| = \sum_{v\in V}\deg(v)$. By construction, the vertices (resp. edges) of a kernel have degree (resp. size) at least 2, so

$$k+n+m \geq 3m - m_2, \tag{13}$$

$$k+n+m \geq 3n - n_2, \tag{14}$$

where n_2 (resp. m_2) is the number of vertices of degree 2 (resp. edges of size 2). Furthermore, each vertex of degree 2 belongs to an edge of size at least 3, so

$$k+n+m \geq 2m_2 + n_2. \tag{15}$$

Summing those three inequalities, we obtain $3k \geq m_2$.

This bound is reached if and only if (13), (14) and (15) are in fact equalities. Therefore, the vertices (resp. edges) of a clean kernel have degree (resp. size) 2 or 3, each vertex of degree 2 belongs to exactly one edge of size 3 and all the vertices of degree 3 belongs to edges of size 2. Consequently, any connected clean kernel can be obtained from a connected cubic multigraph with $2k$ vertices through substitutions of vertices of degree 3 by groups of three vertices of degree 2 that belong to a common edge of size 3. This means that if $f(z)$ represent the cubic multigraphs where z marks the vertices, then the generating function of clean kernels is $f(z+\omega_3 z^3)$. The generating function of cubic multigraphs of excess k is $\frac{(6k)!}{(3!)^{2k}2^{3k}(3k)!}\frac{z^{2k}}{(2k!)}$, and a cubic multigraph is a set of connected cubic multigraphs, so the value $(2k)!c_k$ defined in the theorem is the sum of the weights of the connected cubic multigraphs.

To prove that the total number of kernels of excess k is bounded, we introduce the *dualized* kernels, which are kernels where each edge of size 2 contains a vertex of degree at least 3. This implies the dual inequality of (15) $k+n+m \geq 2n_2+m_2$ that leads to $7k \geq n+m$. Finally, each dualized kernel matches a finite number of normal kernels by substitution of an arbitrary set of vertices of degree 2 by edges of size 2.

The previous theorem gives a way to construct all connected hypergraphs of fixed excess k from a finite set of kernels. This allows us to derive the asymptotic number of connected hypergraphs with fixed excess. The corresponding result for uniform hypergraphs can be found in [13].

Theorem 9. *Let $C_k(z)$ denote the generating function of connected hypergraphs with excess k, then near its dominant singularity ρ, we have*

$$C_k(z) \sim c_k \left(\frac{\sqrt{\gamma}}{2^{3/2}\tau} \right)^k (1 - z/\rho)^{-3k/2}$$

where c_k is defined in Theorem 8, τ is the solution of $\tau\,\Omega''(\tau) = 1$, $\rho = \tau e^{-\,\Omega'(\tau)}$ and $\gamma = 1 + \tau^2\,\Omega'''(\tau)$. The number of connected hypergraphs with excess k and n vertices is

$$n![z^n]C_k(z) \sim \frac{c_k\sqrt{2\pi}}{\Gamma\left(\frac{3k}{2}\right)} \left(\frac{\sqrt{\gamma}}{2^{3/2}\tau} \right)^k \left(\frac{e}{\rho} \right)^n n^{n+(3k-1)/2}.$$

The same results apply to connected simple hypergraphs.

Proof. Theorem 8 implies that the generating function of the connected kernels of excess k is a multivariate polynomial with variables $z, \omega_2, \omega_3, \ldots$. Let us write it as the sum of two polynomials, P_k and Q_k, one corresponding to clean kernels and the other to the rest of the kernels of excess k. According to Theorem 8, P_k is equal to $c_k(1 + \omega_3 z^2)^{2k} \omega_2^{3k} z^{2k}$. By definition of the clean kernels, the degree of Q_k with respect to ω_2 is strictly less than $3k$.

One can develop a kernel into a hypergraph by adding rooted trees to its vertices, replacing its edges of size 2 by paths of trees and adding rooted trees into the edges of size greater than 2. This matches the following substitutions in the generating function of kernels: $z \leftarrow T(z)$, $w_2 \leftarrow \frac{\Omega''(T)}{1 - T\,\Omega''(T)}$ and $w_t \leftarrow \Omega^{(t)}(T)$ for all $t > 2$. Applying this substitution to $P_k + Q_k$, we obtain for the generating functions $C_k(z)$ of connected hypergraphs of excess k

$$C_k(z) = c_k(1 + z^2\,\Omega'''(T(z)))^{2k} \left(\frac{\Omega''(T(z))}{1 - T(z)\,\Omega''(T(z))} \right)^{3k} z^{2k} + \ldots$$

where the "…" hides terms with a denominator $1 - T(z)\,\Omega''(T(z))$ at a power at most $3k - 1$. The Puiseux developments already derived for the proof of Theorem 6, lead to the expression stated in the theorem, from which the asymptotic enumeration result follows.

We now prove that the result holds for connected simple hypergraphs. As shown in the first part of the proof, we can restrict our investigation to hypergraphs with clean kernels. Among them, let us consider the set of connected hypergraphs with excess k that are not simple. Each one contains a loop or two edges of size 2 linking the same vertices. Therefore, the kernel of each of them has at least one edge of size 2 that is not replaced by a (non-empty) path of threes in the hypergraph. It follows that the generating function of those hypergraphs has a denominator $1 - T(z)\,\Omega''(T(z))$ at a power at most $3k - 1$, so the cardinality of this set is negligible compared to the number of connected hypergraphs with excess k.

An other and more intuitive way to understand it is that at fixed excess, adding more and more vertices into a kernel, the chances that an edge of size 2 does not break into a non-empty path of threes are negligible.

To derive a complete asymptotic expansion of connected hypergraphs, one needs to take into account non-clean kernels. For any fixed k, one can enumerate all the kernels of excess k (since it is a finite set), then apply the substitution described in the previous proof to obtain the generating function of all connected hypergraphs of excess k, from which a complete asymptotic expansion follows. Although computable, this construction is heavy. The purely analytic approach of [21], that adresses this problem for graphs, may allow a simpler expression.

The asymptotic enumeration of connected hypergraphs in $\mathrm{HG}_{n,k}$ when k tends toward infinity is more challenging. Since the original result for graphs of [22], other proofs have been proposed, as [23], which may be generalized to hypergraphs.

7. Structure of Hypergraphs in the Critical Window

The next theorem describes the structure of hypergraphs with an excess at or close to the critical value $k = (\Lambda - 1)n$ introduced in Theorem 6. It generalizes Theorem 5 of [2] about graphs. Interestingly, the result at the critical excess does not depend on the (ω_t).

Theorem 10. *Let $\Psi(z)$, τ, Λ and γ be defined as in Theorem 6. Let $r_1, \dots, r_q$ denote a finite sequence of integers and $r = \sum_{t=1}^{q} t\, r_t$, then the limit of the probability for a hypergraph or simple hypergraph with n vertices and global excess $k = (\Lambda - 1)n + \mathcal{O}(n^{1/3})$ to have exactly r_t components of excess t for t from 1 to q is*

$$\left(\frac{4}{3}\right)^r \frac{r!}{(2r)!} \sqrt{\frac{2}{3}} \frac{c_1^{r_1}}{r_1!} \frac{c_2^{r_2}}{r_2!} \cdots \frac{c_q^{r_q}}{r_q!}. \tag{16}$$

where the (c_i) are defined as in Theorem 8. For $k = (\Lambda - 1)n + xn^{2/3}$ and x bounded, the limit of this probability is

$$3^{-r} \frac{c_1^{r_1}}{r_1!} \frac{c_2^{r_2}}{r_2!} \cdots \frac{c_q^{r_q}}{r_q!} \sqrt{\frac{3\pi}{2}} \exp\left(\frac{-x^3\gamma}{6(1-\Lambda)^3}\right) G\left(\frac{3}{2}, \frac{1}{4} + \frac{3r}{2}; -\frac{3^{2/3}\gamma^{1/3}x}{2(1-\Lambda)}\right).$$

Proof. Let $C_k(z)$ denote the generating function of connected hypergraphs of excess k. From Theorem 9, when z tends towards the dominant singularity ρ of $T(z)$,

$$C_k(z) \sim c_k \left(\frac{\sqrt{\gamma}}{2^{3/2}\tau}\right)^k (1 - z/\rho)^{-3k/2}.$$

The sum of the weights of hypergraphs with global excess k and r_t components of excess t is

$$n![z^n] \frac{U^{r-k}}{(r-k)!} e^V \frac{C_1(z)^{r_1}}{r_1!} \frac{C_2(z)^{r_2}}{r_2!} \cdots \frac{C_q(z)^{r_q}}{r_q!}$$

and an application of Theorem 5 ends the proof, with $G(3/2, 1/4 + 3r/2; 0) = \frac{2}{3\sqrt{\pi}} \frac{4^r r!}{(2r)!}$. Those computations are the same as in Theorem 6.

In [2] page 52, the authors remark that the theorem holds true when q is unbounded, because the sum of the probabilities (16) over all finite sequences (r_t) is 1.

8. Birth of the Giant Component

Erdős and Rényi [3] analyzed random graphs with a large number n of vertices and m edges such that m/n tends toward a constant c. They proved that when c is strictly greater than $1/2$, with high probability the graph has a *giant component*, which contains a constant fraction of the vertices and has an excess going to infinity with n. Similar results have been derived for various models of hypergraphs [9], [4], [5], [6], [7].

We consider random hypergraphs with n vertices and excess $k = (\lambda - 1)n$. We have seen in Theorem 4 that when $\lambda < \Lambda$, with high probability the hypergraph contains only trees and unicyclic components. We also have derived the limit distribution of the excesses of the components in the critical window, *i.e.* for hypergraphs with excess $k = (\Lambda - 1)n + xn^{2/3}$ with x bounded. In this section, we investigate the case $\lambda > \Lambda$.

Molloy and Reed [24] and Newman, Strogatz and Watts [25] gave an intuitive explanation of the birth of the giant component in graphs with known degree distribution[1]. Starting with a vertex, we can determine the component in which it lies by exploring its neighbors, then the neighbors of its neighbors and so on. This branching process is likely to stop rapidly if the expected number of new neighbors is smaller than 1. On the other hand, the component is likely to be large if this means is greater than 1. We will not derive results on the giant component as precise as Erdős and Rényi [3] did. Instead, we explain why the expected number of new neighbors is smaller than 1 only for subcritical hypergraphs.

Let us define the *excess degree* of a vertex v in an edge e as the sum over all the other edges that contain v of their sizes minus 1

$$\text{excess degree}(v, e) = \sum_{\substack{v \in \tilde{e}, \\ \tilde{e} \neq e}} (|\tilde{e}| - 1) .$$

This is the number of neighbors we discover when we arrive at the vertex v from the edge e, assuming they are distinct. We now prove that the expected excess degree is smaller than 1 only for subcritical hypergraphs. This provides an intuitive explanation for the birth of the giant component.

Theorem 11. *Let us consider a random hypergraph with n vertices and excess $k = (\lambda - 1)n$, and a uniformly chosen pair (v, e), where the vertex v belongs to the edge e. With the notations of Theorem 4, the expected excess degree of v in e is smaller than (resp. equal to or greater than) 1, if λ is smaller than (resp. equal to or greater than) Λ.*

Proof. Let $F(u)$ denote the generating function of the degree excess of a marked vertex in a marked edge of a hypergraph with n vertices and excess

[1] We thank an anonymous referee for those references.

$k = (\lambda - 1)n$. The marked vertex and edge represent v and e. Such a hypergraph can be decomposed into a hypergraph on $n-1$ vertices, an edge with one vertex marked – this is the edge e that contains the vertex v – and a set of edges with one vertex marked and removed. Those last edges contain the neighbors of v that are counted by the excess degree. Introducing the variable u to mark the excess degree of v and the variable y for the excess of the hypergraph, we obtain

$$F(u) = [y^{n+k}] e^{\frac{\Omega((n-1)y)}{y}} \Omega'(ny) e^{\Omega'(nyu)}.$$

After the change of variable $ny \to y$, this expression becomes

$$F(u) = n^{n+k} [y^{n+k}] e^{n \frac{\Omega\left(y - \frac{y}{n}\right)}{y}} \Omega'(y) e^{\Omega'(yu)},$$

which can be approximated by

$$F(u) = n^{n+k} [y^{n+k}] e^{n \frac{\Omega(y)}{y} - \Omega'(y) + \mathcal{O}(1/n)} \Omega'(y) e^{\Omega'(yu)}.$$

The expected excess degree of v in e is then $F'(1)/F(1)$ (see [14, Part C])

$$\mathbb{E}(\text{excess degree}) = \frac{F'(1)}{F(1)} = \frac{[y^{n+k}] e^{n \frac{\Omega(y)}{y} - \Omega'(y) + \mathcal{O}(1/n)} \Omega'(y) y \Omega''(y) e^{\Omega'(y)}}{[y^{n+k}] e^{n \frac{\Omega(y)}{y} - \Omega'(y) + \mathcal{O}(1/n)} \Omega'(y) e^{\Omega'(y)}}.$$

The asymptotics of the coefficient extractions in the computation of $F(1)$ and $F'(1)$ are obtained using the *Large Powers Theorem* [14, Theorem VIII.8]. The saddle-point ζ is characterized by

$$\Psi(\zeta) = \lambda,$$

and the limit value of the expectation is

$$\lim_{n \to \infty} \mathbb{E}(\text{excess degree}) = \zeta \Omega''(\zeta).$$

Let us recall the equalities $\Lambda = \Psi(\tau)$ and $\tau \Omega''(\tau) = 1$. Since $\Psi(z)$ and $z\Omega''(z)$ are increasing functions, it follows that when λ is smaller than (resp. equal to or greater than) Λ, then $\zeta \Omega''(\zeta)$ is smaller than (resp. equal to or greater than) 1.

9. Future Directions

In the present paper, for the sake of the simplicity of the proofs, we restrained our work to the case where $e^{\Omega(z)/z}$ is aperiodic. This technical condition can be waived in the same way Theorem VIII.8 of [14] can be extended to periodic functions.

In the model we presented, the weight ω_t of an edge only depends on its size t. For some applications, one may need weights that also vary with the number of vertices n. It would be interesting to measure the impact of this modification on the phase transition properties described in this paper.

More generally, the study of the relation to other models, as the one presented in [10] and [26], could lead to new developments and applications.

References

[1] P. Flajolet, D. E. Knuth, B. Pittel, The first cycles in an evolving graph, Discrete Mathematics 75 (1-3) (1989) 167–215.

[2] S. Janson, D. E. Knuth, T. Łuczak, B. Pittel, The birth of the giant component, Random Structures and Algorithms 4 (3) (1993) 233–358.

[3] P. Erdős, A. Rényi, On the evolution of random graphs, Publication of the Mathematical Institute of the Hungarian Academy of Sciences 5 (1960) 17.

[4] M. Karoǹski, T. Łuczak, The phase transition in a random hypergraph, Journal of Computational and Applied Mathematics 142 (1) (2002) 125–135.

[5] C. Cooper, The cores of random hypergraphs with a given degree sequence, Random Struct. Algorithms 25 (4) (2004) 353–375.

[6] A. Dembo, A. Montanari, Finite size scaling for the core of large random hypergraphs, Ann. Appl. Probab. 18 (5) (2008) 1993–2040.

[7] G. Ghoshal, V. Zlatić, G. Caldarelli, M. Newman, Random hypergraphs and their applications, Physical Review E 79 (6) (2009) 066118.

[8] V. Ravelomanana, Birth and growth of multicyclic components in random hypergraphs, Theoretical Computer Science 411 (43) (2010) 3801–3813.

[9] J. Schmidt-Pruzan, E. Shamir, Component structure in the evolution of random hypergraphs, Combinatorica 5 (1) (1985) 81–94.

[10] R. W. R. Darling, J. R. Norris, Structure of large random hypergraphs, Annals of Applied Probability 23 (6) (2004) 125–152.

[11] I. M. Gessel, L. H. Kalikow, Hypergraphs and a functional equation of Bouwkamp and de Bruijn, Journal of Combinatorial Theory, Series A 110 (2) (2005) 275–289.

[12] E. M. Wright, The number of connected sparsely edged graphs, Journal of Graph Theory 1 (1977) 317–330.

[13] M. Karoǹski, T. Łuczak, The number of connected sparsely edged uniform hypergraphs, Discrete Mathematics 171 (1-3) (1997) 153–167.

[14] P. Flajolet, R. Sedgewick, Analytic Combinatorics, Cambridge University Press, 2009.

[15] H.-K. Hwang, On convergence rates in the central limit theorems for combinatorial structures, European Journal of Combinatorics 19 (3) (1998) 329–343.

[16] C. Berge, Graphs and Hypergraphs, Elsevier Science Ltd, 1985.

[17] F. Bergeron, G. Labelle, P. Leroux, Combinatorial Species and Tree-like Structures, Cambridge University Press, 1997.

[18] B. Oger, Decorated hypertrees, Journal of Combinatorial Theory, Series A 120 (7) (2013) 1871–1905.

[19] C. Banderier, P. Flajolet, G. Schaeffer, M. Soria, Random maps, coalescing saddles, singularity analysis, and Airy phenomena, Random Structures Algorithms 19 (3-4) (2001) 194–246.

[20] E. M. Wright, The number of connected sparsely edged graphs III: Asymptotic results, Journal of Graph Theory 4 (4) (1980) 393–407.

[21] P. Flajolet, B. Salvy, G. Schaeffer, Airy phenomena and analytic combinatorics of connected graphs, Electronic Journal of Combinatorics 11 (1).

[22] E. A. Bender, E. R. Canfield, B. D. McKay, The asymptotic number of labeled connected graphs with a given number of vertices and edges, Random Structures and Algorithm 1 (1990) 129–169.

[23] B. Pittel, N. C. Wormald, Counting connected graphs inside-out, Journal of Combinatorial Theory, Series B 93 (2) (2005) 127–172.

[24] M. Molloy, B. Reed, A critical point for random graphs with a given degree sequence, Random Structures and Algorithms 6 (1995) 161–180.

[25] M. E. J. Newman, S. H. Strogatz, D. J. Watts, Random graphs with arbitrary degree distributions and their applications, Phys. Rev. E 64 (2) (2001) 026118.

[26] B. Bollobás, S. Janson, O. Riordan, Sparse random graphs with clustering, Random Structures and Algorithms 38 (3) (2011) 269–323.