Phase Transition of Random Non-uniform Hypergraphs

Élie de Panafieu*

Univ Paris Diderot, Sorbonne Paris Cité,
LIAFA, UMR 7089,
75013, Paris, France

Abstract. Non-uniform hypergraphs appear in several domains of computer science as in the satisfiability problems and in data analysis. We analyze their typical structure before and near the birth of the *complex* component, that is the first connected component with more than one cycle. The model of non-uniform hypergraph studied is a natural generalization of the *multigraph process* defined in the "giant paper" [1]. This paper follows the same general approach based on analytic combinatorics. We study the evolution of 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 hypergraphs with a given excess, including the expected number of edges.

Keywords: Hypergraph, phase transition, analytic combinatorics.

1 Introduction

In the seminal article [2], 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 appear for example for the modelisation of databases and xor-formulas.

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 can be found in [3] and [4].

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 [5], where the size of the largest connected component is obtained using probabilistic methods. In [6], Darling and Norris define the important Poisson random hypergraphs model and analyze its structure via fluid limits of pure jump-type Markov processes.

* This work was partially founded by the ANR Boole and the ANR Magnum.

T. Lecroq and L. Mouchard (Eds.): IWOCA 2013, LNCS 8288, pp. 127–139, 2013.

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 [7] 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* defined in [1].

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 expected number of edges. The critical excess at which the first complex component appears is obtained in section 4. 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 5. It is then used to derive the asymptotic of connected hypergraphs with n vertices and excess k up to a multiplicative factor independent of n. Finally, we present in section 6 a surprising result: although the critical excess is generally different for graphs and hypergraphs, both models share the same structure at their respective critical excess.

2 Definitions

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

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

The notion of *excess* was first used for graphs in [8], then named in [1], and finally extended to hypergraphs in [9]. The excess of a connected component C expresses how far from a tree it is: C is a tree if and only if its excess is -1 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 $\text{NumbSeq}(G)$ as the number of sequences of nonempty sequences of vertices that lead to G. For example, the sequences $(1,2),(2,3)$ and $(3,2),(1,2)$ represent the same hypergraph, but not $(2,1),(1,3)$. If G is simple, then $\text{NumbSeq}(G)$ 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

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

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

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

that is the extension to hypergraphs of the *compensation factor* defined in section 1 of [1]. 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 $[y^k] F(z/y, w/y, xy)$, where $[x^n] \sum_k a_k x^k$ denotes the coefficient a_n.

We define the exponential generating function of the edges as

$$\Omega(z) := \sum_{t \geq 1} \omega_t \frac{z^t}{t!}.$$

For 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 $\Omega(z) = z^d/d!$, and for hypergraphs with weight 1 for all size of edge $\Omega(z) = e^z$. 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 uniform hypergraphs.

The generating function of all hypergraphs is

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

This expression can be derived from (1) or using the symbolic method presented in [10]. 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,

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

Similar expressions have been derived in [7]. The authors use them to give a combinatorial interpretation of a functional equation of Bouwkamp and de Bruijn.

Comparing (1) with (3), simple hypergraphs may appear more natural than hypergraphs. But their generating function is more intricated and the asymptotics results on hypergraphs can often be extended to simple hypergraphs. This is another reason not to confine our study to simple hypergraphs.

So far, we have adopted an enumerative approach of the model, but there is a corresponding probabilistic description. We define $\mathrm{HG}_{n,k}$ (resp. $\mathrm{SHG}_{n,k}$) as the set of hypergraphs (resp. simple hypergraphs) with n vertices and excess k, and equippe it 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.

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

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 [10]. 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

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

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 $\frac{\Omega(y)}{y}n-\frac{\omega_2^2y^2}{4}-\frac{y\,\Omega''(y)}{2}+\mathcal{O}(1/n)$ 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,

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

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 $\text{HG}_{n,k}$ 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 [1].

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.

Corollary 1. *Let λ be a positive value and G a random hypergraph in* $\text{HG}_{n,k}$ *or in* $\text{SHG}_{n,k}$ *with $k=(\lambda-1)n$, then the asymptotic expectations and factorial moments of the number m of edges and size l of G are*

$$\mathbb{E}_{n,k}(m)\sim\tfrac{\Omega(\zeta)}{\zeta}n,$$
$$\forall t\geq 0,\ \mathbb{E}_{n,k}\big(m(m-1)\dots(m-t)\big)\sim\left(\tfrac{\Omega(\zeta)}{\zeta}n\right)^{t+1},$$
$$\mathbb{E}_{n,k}(l)\sim\Omega'(\zeta)n.$$

where $\Psi(z)$ denotes the function $\Omega'(z)-\frac{\Omega(z)}{z}$ and ζ is the solution of $\Psi(\zeta)=\lambda$.

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 discret random variable X takes the value t and $f(z)=\sum_n p_nz^n$, 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 $\text{SHG}_{n,k}$ are

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

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.

The variance of $m(G)$ for G in $\mathrm{HG}_{n,k}$ cannot be straightforward derived from this corollary, because the asymptotic approximations of the factorial moments cannot be summed. If more terms of the asymptotic expansion of the factorial moments are derived, this variance can be bounded. However, it varies greatly with the parameters (ω_t). For example, the variance for graphs is 0, since all the graphs with n vertices and excess k have exactly $k+n$ edges.

4 Birth of the Complex Component

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.

We follow the conventions established in [11]: a *walk* of a hypergraph G is a sequence $v_0, e_1, v_1, \dots, v_{t-1}, e_t, v_t$ 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 1. *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 from the tools presented in [10]. 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 [12] and studied in [13]. It can be checked that $z\partial_z U = T$. Unicycle components are cycles of rooted trees, which implies (8).

Theorem 3 counts the hypergraphs with no complex component. A phase transition occurs when $\frac{k}{n}$ reaches the critical value $\Lambda - 1$, defined in the next theorem, which corresponds to the coalescence of two saddle points. To extract the asymptotics, we need the following general theorem, borrowed from [14] and adapted for our purpose (in the original theorem, $\mu = 0$). It is also close to the lemma 3 of [1].

Theorem 2. *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 $k = \frac{\sigma}{h_1} n + x n^{1/\lambda}$ with x bounded, then for any real constant μ*

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

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 $z = \rho(1 - t n^{-1/\lambda})$

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

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 $-\infty$.

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

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

Theorem 3. *Let $\mathrm{thg}_{n,k}$ denote the sum of the weights of the hypergraphs with no complex component, n vertices and global excess k. Let $\Psi(z)$ denote the function $\Omega'(z) - \frac{\Omega(z)}{z}$, τ be implicitely defined by $\tau \Omega''(\tau) = 1$ and $\Lambda = \Psi(\tau)$.*

If $k = (\lambda - 1)n + \mathcal{O}(n^{1/3})$ with $0 < \lambda < \Lambda$, and $\Psi(\zeta) = \lambda$, then

$$\mathrm{thg}_{n,k} \sim \frac{n^{n+k}}{\sqrt{2\pi n}} \frac{e^{\frac{\Omega(\zeta)}{\zeta} n}}{\zeta^{n+k}} \frac{1}{\sqrt{\zeta \Omega''(\zeta) - \lambda}}. \tag{11}$$

If $k = (\Lambda - 1)n + x n^{2/3}$ where x is bounded, then $\mathrm{thg}_{n,k}$ is equivalent to

$$\frac{n^{n+k}}{\sqrt{2\pi n}} \frac{e^{\frac{\Omega(\tau)}{\tau} n}}{\tau^{n+k}} \frac{1}{\sqrt{1 - \Lambda}} \sqrt{\frac{3\pi}{2}} e^{-\frac{x^2}{2(1-\Lambda)} n^{1/3} - \frac{x^3}{3(1-\Lambda)^2}} G\left(\frac{3}{2}, \frac{1}{4}; -\frac{3^{2/3} \gamma^{1/3} x}{2(1 - \Lambda)}\right) \tag{12}$$

where $G(\lambda, \mu; x)$ *is defined in theorem 2 and* $\gamma = 1 + \tau^2\,\Omega'''(\tau)$. *For simple hypergraphs, there is an additional factor* $\exp\left(-\frac{\zeta\,\Omega''(\zeta)}{2} - \frac{\omega_2^2\zeta^2}{4}\right)$ *for the first asymptotic, and* $\exp\left(-\frac{1}{2} - \frac{\omega_2^2\tau^2}{4}\right)$ *for the second one.*

Proof. A hypergraph G with no complex component is a forest of 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, the excess of G is the opposite of the number of trees. The sum of the weights of the hypergraphs with no complex component, n vertices and excess k (which is negative) is

$$n![z^n]\frac{U^{-k}}{(-k)!}e^V = \frac{n!}{(-k)!}[z^n]\frac{U^{-k}}{\sqrt{1 - T\,\Omega''(T)}} = \frac{n!}{(-k)!}\frac{1}{2i\pi}\oint \frac{U^{-k}}{\sqrt{1 - T\,\Omega''(T)}}\frac{dz}{z^{n+1}}.$$

For $k = (\lambda - 1)n$, there are two saddle points: one implicitly defined by $\Psi(\zeta) = \lambda$ and the other at τ. Those two saddle points coalesce when $\lambda = \Psi(\tau)$. For smaller values of λ, the first saddle point dominates and an application of the large power theorem of [10] leads to (11). When k is around its critical value $(\Lambda - 1)n$, we apply theorem 2. The Newton-Puiseux expansions of T, e^V and U can be derived from lemma 1

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

where $\Psi(z) = \Omega'(z) - \frac{\Omega(z)}{z}$ and $\gamma = 1 + \tau^2\,\Omega'''(\tau)$. Using Theorem 2, 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 (12).

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

Combining theorems 1 and 3, 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 [15] for graphs. However, the evolution of this probability between the subcritical and the critical ranges of excess depends on the (ω_t).

Corollary 2. *For* $k = (\Lambda - 1)n + xn^{2/3}$ *with* x *bounded, the probability that a hypergraph in* $\mathrm{HG}_{n,k}$ *or in* $\mathrm{SHG}_{n,k}$ *has no complex component is*

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

Theorem 2 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 3 do not apply to 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 [5].

5 Kernels

In the seminal articles [8] and [16], 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 [1] and the notions of excess and kernels were extended to hypergraphs in [9].

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.

Following [9], 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 [9]. We give a new proof and an expression for the generating function of the *clean* kernels.

Theorem 4. *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*

$$c_k (1 + \omega_3 z^2)^{2k} \omega_2^{3k} z^{2k} \tag{13}$$

where $c_k = [z^{2k}] \log \sum_k \frac{(6k)!}{(3!)^{2k} 2^{3k} (3k)!} \frac{z^{2k}}{(2k)!}$ 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

$$k+n+m \geq 3m - m_2, \tag{14}$$
$$k+n+m \geq 3n - n_2, \tag{15}$$

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

$$k+n+m \geq 2m_2 + n_2. \tag{16}$$

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

This bound is reached if and only if (14), (15) and (16) 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 (16) $k+n+m \geq 2n_2+m_2$ 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 implies that the generating function of the connected kernels of excess k is a multivariate polynomial of degree $3k$ in ω_2. 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 functions: $z \leftarrow T(z)$, $w_2 \leftarrow \frac{\Omega''(T)}{1-T\,\Omega''(T)}$ and $w_t \leftarrow \Omega^{(t)}(T)$ for all $t>2$. Therefore, there exists a polynomial $P_k(X)$ in $\mathbb{Q}[X, \Omega(X), \Omega'(X), \dots]$ such that the generating function of connected hypergraphs of excess k is expressed as

$$\mathrm{chg}_k(z) = \frac{P_k(T)}{(1-T\,\Omega''(T))^{3k}}. \tag{17}$$

From there, a singularity analysis gives the asymptotics of connected hypergraphs in $\mathrm{HG}_{n,k}$

$$n![z^n]\,\mathrm{chg}_k(z) \sim \sqrt{2\pi} \frac{P_k(\tau)(2+2\tau^2\,\Omega'''(\tau))^{-\frac{3k}{2}}}{\Gamma\left(\frac{3k}{2}\right)} \frac{e^{(\Omega'(\tau)-1)n}}{\tau^n} n^{n+\frac{3k-1}{2}}$$

where τ, value of T at its dominant singularity, is characterized by $\tau\,\Omega''(\tau) = 1$.

This formula gives the asymptotic number of connected hypergraphs with respect to n, up to a constant factor $P_k(\tau)$, which is computable through the enumeration of the connected kernels of excess k. This is however unsatisfactory, because the complexity of this computation is too high. We believe that the approach developed in [17] for graphs may be the solution. It starts by considering hypergraphs as sets of trees, unicycle components and connected components of higher order

$$\mathrm{hg}(z,y) = \sum_n e^{\frac{\Omega(yn)}{y}} \frac{(z/y)^n}{n!} = \exp\left(y^{-1}U + V + \sum_k \frac{P_k(T)}{(1 - T\,\Omega''(T))^{3k}} \right),$$

from which it may be possible to extract informations on the values $P_k(\tau)$.

6 Hypergraphs with Complex Components of Fixed Excess

The next theorem describes the structure of critical hypergraphs. It generalizes theorem 5 of [1] about graphs. Interestingly, the result does not depend on the (ω_t).

Theorem 5. *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*

$$\left(\frac{4}{3}\right)^r \frac{r!}{(2r)!} \sqrt{\frac{2}{3}} \frac{c_1^{r_1}}{r_1!} \frac{c_2^{r_2}}{r_2!} \cdots \frac{c_q^{r_q}}{r_q!}. \tag{18}$$

where the (c_i) are defined as in Theorem 4. For $k = (\Lambda - 1)n + x n^{2/3}$ and x bounded, the limit of this probability is

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

Proof. Let $C_k(z)$ denote the generating function of connected hypergraphs of excess k. Those can be obtained by expansion of the connected kernels of excess k, so $C_k(z) = c_k (1 + T^2\,\Omega'''(T))^{2k} \frac{\Omega''(T)^{3k}}{(1 - T\,\Omega''(T))^{3k}} T^{2k} + \dots$ plus terms with a denominator $(1 - T\,\Omega''(T))$ of smaller order. Therefore, 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 $n![z^n] \frac{U^{K-k}}{(K-k)!} e^V \frac{C_1(z)^{r_1}}{r_1!} \frac{C_2(z)^{r_2}}{r_2!} \cdots \frac{C_q(z)^{r_q}}{r_q!}$ and an application of Theorem 2 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 3.

Remark. We have seen in the proof that around the critical value of the excess $k = (\Lambda - 1)n$, the kernel of a hypergraph is clean with high probability. In [1], the authors remark that the theorem holds true when q is unbounded, because the sum of the probabilities (18) over all finite sequences (r_t) is 1.

7 Future Directions

Much more information can be extracted from the generating functions (4) and (5), as the number of edges at which the first cycle appears [15], more statistics on the parameters n, m, k and l for random hypergraphs and more error terms on the asymptotics presented. In particular, connected non-uniform hypergraphs deserve a dedicated paper, with an expression for the constants $P_k(\tau)$ defined in (17).

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 [10] 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 [6] and [18], could lead to new developments and applications.

References

1. Janson, S., Knuth, D.E., Luczak, T., Pittel, B.: The birth of the giant component. Random Struct. Algorithms 4(3), 233–359 (1993)
2. Erdös, P., Rényi, A.: On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. 5, 17 (1960)
3. Karoǹski, M., Łuczak, T.: The phase transition in a random hypergraph. Journal of Computational and Applied Mathematics 42(1), 125–135 (2002); Probabilistic Methods in Combinatorics and Combinatorial Optimization
4. Ravelomanana, V.: Birth and growth of multicyclic components in random hypergraphs. Theor. Comput. Sci. 411(43), 3801–3813 (2010)
5. Schmidt-Pruzan, J., Shamir, E.: Component structure in the evolution of random hypergraphs. Combinatorica 5(1), 81–94 (1985)
6. Darling, R.W.R., Norris, J.R.: Structure of large random hypergraphs. Ann. Appl. Probab., 125–152 (2004)
7. Gessel, I.M., Kalikow, L.H.: Hypergraphs and a functional equation of Bouwkamp and de Bruijn. J. Comb. Theory, Ser. A 110(2), 275–289 (2005)
8. Wright, E.M.: The number of connected sparsely edged graphs. Journal of Graph Theory 1, 317–330 (1977)
9. Karoǹski, M., Łuczak, T.: The number of connected sparsely edged uniform hypergraphs. Discrete Mathematics 171(1-3), 153–167 (1997)
10. Flajolet, P., Sedgewick, R.: Analytic Combinatorics. Cambridge University Press (2009)

11. Berge, C.: Graphs and Hypergraphs. Elsevier Science Ltd. (1985)
12. Bergeron, F., Labelle, G., Leroux, P.: Combinatorial Species and Tree-like Structures. Cambridge University Press (1997)
13. Oger, B.: Decorated hypertrees. CoRR abs/1209.0941 (2012)
14. Banderier, C., Flajolet, P., Schaeffer, G., Soria, M.: Random maps, coalescing saddles, singularity analysis, and airy phenomena. Random Struct. Algorithms 19(3-4), 194–246 (2001)
15. Flajolet, P., Knuth, D.E., Pittel, B.: The first cycles in an evolving graph. Discrete Mathematics 75(1-3), 167–215 (1989)
16. Wright, E.M.: The number of connected sparsely edged graphs III: Asymptotic results. Journal of Graph Theory 4(4), 393–407 (1980)
17. Flajolet, P., Salvy, B., Schaeffer, G.: Airy phenomena and analytic combinatorics of connected graphs. The Electronic Journal of Combinatorics 11, 34 (2004)
18. Bollobás, B., Janson, S., Riordan, O.: Sparse random graphs with clustering, cite arxiv:0807 (2008)