Evolution of high-order connected components in random hypergraphs

Oliver Cooley [1,2], Mihyun Kang [1,2], and Christoph Koch [1,2]

Institute of Optimization and Discrete Mathematics
Graz University of Technology
8010 Graz, Austria

Abstract

We consider high-order connectivity in k-uniform hypergraphs defined as follows: Two j-sets are j-connected if there is a walk of edges between them such that two consecutive edges intersect in at least j vertices. We describe the evolution of j-connected components in the k-uniform binomial random hypergraph $\mathcal{H}^k(n,p)$. In particular, we determine the asymptotic size of the giant component shortly after its emergence and establish the threshold at which $\mathcal{H}^k(n,p)$ becomes j-connected with high probability. We also obtain a hitting time result for the related random hypergraph process $\{\mathcal{H}^k(n,M)\}_M$ – the hypergraph becomes j-connected exactly at the moment when the last isolated j-set disappears. This generalises well-known results for graphs and vertex-connectivity in hypergraphs.

Keywords: Random hypergraphs, high-order connectivity, hitting time, giant component, phase transition.

1 Evolution of random graphs

The theory of random graphs was founded in the late 1950s by Erdős and Rényi describing the evolution of the random graph process $\{G(n, M)\}_M$. The vertex set of this process is $[n] := \{1, \ldots, n\}$ and initially there are no edges present. In each step of the process, add an edge between a pair of vertices chosen uniformly at random amongst all pairs of vertices that do not already form an edge. In the early stages of this process, all connected components are small and then, within very short time, they merge into a single component of linear size – the *giant component*. This remarkable phenomenon, first proved in [8], is known as the *phase transition* of the random graph process $\{G(n, M)\}_M$.

It is often more convenient to analyse the binomial random graph $G(n, p)$: The vertex set is $[n]$ and every pair of vertices is connected by an edge with probability p independently. Incorporating various strengthenings the phase transition can be summarised as follows. (All asymptotic statements are with respect to $n \to \infty$ and by *whp* we abbreviate "with probability $\to 1$".)

Theorem 1.1 (Bollobás [2]; Łuczak [11]) *Let $\varepsilon = \varepsilon(n) > 0$ be a real function satisfying $\varepsilon \to 0$ and $\varepsilon^3 n \to \infty$.*

(i) *If $p = \frac{1-\varepsilon}{n}$, then whp all components in $G(n, p)$ have size $O(\varepsilon^{-2} \log(\varepsilon^3 n))$;*

(ii) *If $p = \frac{1+\varepsilon}{n}$, then whp the size of the largest component in $G(n, p)$ is $(1 \pm o(1))2\varepsilon n$, while all other components have size $O(\varepsilon^{-2} \log(\varepsilon^3 n))$.*

As we continue to add edges one by one, more and more components are consumed by the giant component and eventually the graph becomes connected. In fact, Bollobás and Thomason [4] showed that this happens precisely at the moment when the last isolated vertex disappears – thereby relating a *global graph property* to its minimal *local* obstruction. Denote the *hitting time* of connectivity by τ_c, i.e. τ_c is the minimal M such that $G(n, M)$ is connected, and the hitting time for the disappearance of the last isolated vertex by τ_i.

Theorem 1.2 (Bollobás & Thomason [4]) *Whp $\tau_c = \tau_i$ for $\{G(n, M)\}_M$.*

2 Evolution of random hypergraphs – Main results

Given an integer $k \geq 2$ a *k-uniform hypergraph* H consists of a set V of vertices and a set E of edges, where each edge contains precisely k vertices. In particular, 2-uniform hypergraphs are simply graphs. Given an integer $1 \leq j \leq k - 1$ we say that two *j-sets* (sets of j distinct vertices) J and J' are *j-connected* if there is a sequence of edges $e_1, \ldots, e_m$ such that $J \subset e_1$,

$J' \subset e_m$ and $|e_i \cap e_{i+1}| \geq j$ for all $1 \leq i \leq m-1$. A *j-component* is a maximal set of pairwise *j*-connected *j*-sets. The hypergraph H is *j-connected* if every two *j*-sets are *j*-connected. A *j*-set is called *isolated* if it is not contained in any edge. Note that connectivity in graphs corresponds to the case $k = 2$ and $j = 1$.

We consider the *k*-uniform random hypergraph process $\{\mathcal{H}^k(n, M)\}_M$: The vertex set is $[n]$ and initially there are no edges present. In each step of the process, we add an edge for a *k*-set chosen uniformly at random from all *k*-sets that do not already form an edge. Instead of analysing this process directly we usually consider the *k*-uniform binomial random hypergraph $\mathcal{H}^k(n, p)$ with vertex set $[n]$, where every *k*-set is an edge with probability p independently. In this model there are no dependencies between different edges. It is well-known that both models are very similar and results can be easily transferred from one to the other using standard techniques (e.g. [9]).

For any $k \geq 2$ the case of vertex-connectivity ($j = 1$) is well-studied and results analogous to Theorem 1.1 were obtained in [1,3,10,13]. Theorem 1.2 has also recently been extended for vertex-connectivity in *k*-uniform hypergraphs [12].

However, for high-order connectivity ($j > 1$) not much was known until recently. This is due to the fact that vertex-connectivity can usually be studied with tools which are very similar to those used for graphs. By contrast, analysing high-order connectivity is often significantly more sophisticated and thus these methods are usually not sufficient. Recently Cooley, Kang, and Person [7] showed that the sharp threshold for the emergence of the giant component in $\mathcal{H}^k(n, p)$ is

$$p_{\mathrm{g}} := \frac{1}{\left(\binom{k}{j} - 1\right)\binom{n}{k-j}}.$$

We strengthen this result and provide the asymptotic size of the largest *j*-component in the *weakly supercritical regime* of the phase transition.

Theorem 2.1 *Let $k \geq 2$ and $1 \leq j \leq k-1$ be integers. Let $\varepsilon = \varepsilon(n) > 0$ be a real function satisfying $\varepsilon \to 0$, $\varepsilon^3 n^j \to \infty$ and $\varepsilon^2 n^{1-2\delta} \to \infty$, for some constant $\delta > 0$. Then whp the size of the largest j-component $\mathcal{L}_1$ in $\mathcal{H}^k(n, (1+\varepsilon)p_{\mathrm{g}})$ satisfies*

$$|\mathcal{L}_1| = (1 \pm o(1))\frac{2\varepsilon}{\binom{k}{j} - 1}\binom{n}{j},$$

while all other j-components contain $o(\varepsilon n^j)$ j-sets.

The proof of Theorem 2.1 fundamentally uses a powerful tool, the *smooth boundary lemma*, which provides insight into the structure of the unique largest j-component in the supercritical regime. The precise statement of this lemma requires a large amount of additional notation and thus we omit it and refer to [5] for the details. Instead we explain the notion of 'smooth sets' on a more intuitive level in Section 3.1.

An elegant application of the notion of smoothness arises when studying the threshold for j-connectivity in $\mathcal{H}^k(n, p)$. The key idea is that the giant component contains a smooth set of large size.

Lemma 2.2 *Let $k \geq 2$ and $1 \leq j \leq k - 1$ be integers. Let $\gamma = \gamma(n) > 0$ be a real function satisfying $\gamma \to 0$ and $\gamma^3 n \to \infty$. Then whp the unique largest j-component $\mathcal{L}_1$ in $\mathcal{H}^k(n, (1 + \gamma)p_{\mathrm{g}})$ contains a subset $S \subset \mathcal{L}_1$ of at least $\gamma^3 n^j$ j-sets with the following property for all integers $0 \leq \ell < j$:*

Every ℓ-set $L \subset [n]$ is contained in $(1 \pm o(1)) \frac{|S|}{\binom{n}{j}} \binom{n}{j-\ell}$ j-sets of S.

Based on Lemma 2.2 we provide an elementary proof showing that the hypergraph process $\{\mathcal{H}^k(n, M)\}_M$ becomes j-connected exactly at the moment when the last isolated j-set disappears. Let $\tau_{\mathrm{c}} = \tau_{\mathrm{c}}(k, j)$ be the hitting time for j-connectivity in the random hypergraph process $\{\mathcal{H}^k(n, M)\}_M$ and let $\tau_{\mathrm{i}} = \tau_{\mathrm{i}}(k, j)$ denote the time-step of $\{\mathcal{H}^k(n, M)\}_M$ in which the last isolated j-set disappears.

Theorem 2.3 *Let $k \geq 3$ and $1 \leq j \leq k - 1$ be integers. Then whp $\tau_{\mathrm{c}} = \tau_{\mathrm{i}}$ for $\{\mathcal{H}^k(n, M)\}_M$.*

It follows that, in the binomial random hypergraph $\mathcal{H}^k(n, p)$, the properties of being j-connected and having no isolated j-sets share the common sharp threshold

$$p_{\mathrm{c}} := \frac{j \log n}{\binom{n}{k-j}}.$$

In fact we obtain a slightly stronger result.

Theorem 2.4 *Let $k \geq 3$ and $1 \leq j \leq k - 1$ be integers. Let $\omega = \omega(n) > 0$ be a real function satisfying $\omega \to \infty$ and $\omega = o(\log n)$.*

- *If $p = \frac{j \log n - \omega}{\binom{n}{k-j}}$, then whp $\mathcal{H}^k(n, p)$ contains isolated j-sets (and is therefore not j-connected);*
- *If $p = \frac{j \log n + \omega}{\binom{n}{k-j}}$, then whp $\mathcal{H}^k(n, p)$ is j-connected (and therefore contains no isolated j-sets).*

3 Proof outlines

3.1 Smooth sets in large components

The intuition behind the smooth boundary lemma is the following: For an arbitrary j-set J, we explore its component via a breadth-first search process, i.e. generation by generation. If the component of J happens to be large, the sizes of the generations have a tendency to grow and thus most generations should have a 'reasonable' size already early on in the process. However, once the generations are not too small, random fluctuations should start to even themselves out over time. Thus generations should begin to look 'smooth' in the sense that any set $L \subset [n]$ of at most $j - 1$ vertices is contained in approximately the 'right' number of j-sets of any smooth generation. Usually we will only run the exploration process for as long as necessary, resulting in a *partial component*. If so, the last generation that is discovered before the process was stopped is called the *boundary* and will be of special interest in the proof of Theorem 2.1, since it contains all the j-sets that are still *active*. This will be discussed further in Section 3.2.

In other scenarios, the boundary does not play such a crucial role or may not be large enough on its own, in which case we may consider the smooth set given by the union of all smooth generations in the partial component. It turns out that this union contains almost all j-sets of the partial component. As an immediate consequence we obtain Lemma 2.2. The details can be found in [5,6].

3.2 Emergence of the giant component

First we study the number of j-sets in 'large' components. For a given j-set we explore its component via a breadth-first search process and approximate this process by a supercritical branching process. In order to control the second moment we have to study two exploration processes and make sure that one of them being large does not increase the probability of the other becoming large too much. We first run one exploration process, stop it as soon as we know that whp it will grow large, and then consider a subprocess of the second exploration process in which no k-sets containing a j-set from the first (partial) component is present. The smooth boundary lemma ensures that the resulting process is still close to a (sufficiently) supercritical branching process. Therefore the number of j-sets in large components is concentrated around its expectation and Theorem 2.1 follows by a sprinkling argument. The details can be found in [5].

3.3 Hitting times and connectivity threshold

It is convenient to split the proof of Theorem 2.3 into two parts: First we anal-
yse the structure of the binomial random hypergraph $\mathcal{H}^k(n,p)$ in the regime
where the expected number of isolated j-sets begins to vanish. In particular,
we show that whp in this regime the random hypergraph consists only of one
non-trivial component and isolated j-sets. Then we transfer this structure to
the corresponding time-range in the random hypergraph process $\{\mathcal{H}^k(n,M)\}_M$
by classical contiguity results.

For integer-valued random variables $X_1, X_2, \dots$ and Y, we say X_n _con-
verges in distribution to Y_, denoted by $X_n \xrightarrow{\text{d}} Y$, if we have $\mathbb{P}(X_n = i) \to$
$\mathbb{P}(Y = i)$ for every integer i. We apply the Chen-Stein method for Poisson-
approximation to the number D_s of j-sets of degree $s \geq 0$ (i.e. j-sets that are
contained in precisely s edges).

Theorem 3.1 _Suppose $p = \frac{j \log n + s \log \log n + c_n}{\binom{n}{k-j}}$ for some real function c_n satis-
fying $c_n = o(\log n)$. Then we have, for any integer $s \geq 0$,_

$$
\begin{aligned}
\text{(i)} \quad & D_s = 0 \ whp & \text{if } c_n \to \infty; \\
\text{(ii)} \quad & D_s \xrightarrow{\text{d}} \mathrm{Po}\left(\frac{j^s \, \mathrm{e}^{-c}}{j! s!}\right) & \text{if } c_n \to c \in \mathbb{R}; \\
\text{(iii)} \quad & D_s \to \infty \ whp & \text{if } c_n \to -\infty.
\end{aligned}
$$

Theorem 2.4 is an immediate corollary of Theorem 2.3 and Theorem 3.1.
The details can be found in [6].

References

[1] Behrisch, M., A. Coga-Oghlan, and M. Kang, _The order of the giant component
of random hypergraphs_, Random Structures & Algorithms **36** (2010) 149–184.

[2] Bollobás, B., _The evolution of random graphs_, Trans. Amer. Math. Soc. **286**
(1984) 257–274.

[3] Bollobás, B. and O. Riordan, _Asymptotic normality of the size of the giant
component in a random hypergraph_, Random Structures & Algorithms **41**
(2012) 441–450.

[4] Bollobás, B. and A. Thomason, _Random graphs of small order_, in "Random
graphs '83 (Poznań, 1983)", North-Holland Math. Stud. **118** (1985) 47–97
(North-Holland, Amsterdam, 1985).

[5] Cooley, O., M. Kang, and C. Koch, *The size of the giant component in random hypergraphs*, submitted (2015). Preprint: arXiv:1501.07835.

[6] Cooley, O., M. Kang, and C. Koch, *Threshold and hitting time for high-order connectivity in random hypergraphs*, submitted (2015). Preprint: arXiv:1502.07289.

[7] Cooley, O., M. Kang, and Y. Person, *Largest components in random hypergraphs*, manuscript (2014). Preprint: arXiv:1412.6366.

[8] Erdős, P. and A. Rényi, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **5** (1960) 17–61.

[9] Janson, S., T. Łuczak, and A. Ruciński, "Random graphs", Wiley-Interscience Series in Discrete Mathematics and Optimization (Wiley-Interscience, New York, 2000).

[10] Karoński, M. and T. Łuczak, *The phase transition in a random hypergraph*, J. Comput. Appl. Math. **142** (2002), 125–135.

[11] Łuczak, T., *Component behavior near the critical point of the random graph process*, Random Structures & Algorithms **1** (1990), 287–310.

[12] Poole, D., *On the strength of connectedness of a random hypergraph*, manuscript (2014). Preprint: arXiv:1409.1489.

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