Class of models for random hypergraphs

Marc Barthelemy *

Université Paris-Saclay, CEA, CNRS, Institut de Physique Théorique, 91191 Gif-sur-Yvette, France
and Centre d'Analyse et de Mathématique Sociales (CNRS/EHESS) 54 Avenue de Raspail, 75006 Paris, France

(Received 23 October 2022; accepted 1 December 2022; published 16 December 2022)

Despite the recently exhibited importance of higher-order interactions for various processes, few flexible (null) models are available. In particular, most studies on hypergraphs focus on a small set of theoretical models. Here, we introduce a class of models for random hypergraphs which displays a similar level of flexibility of complex network models and where the main ingredient is the probability that a node belongs to a hyperedge. When this probability is a constant, we obtain a random hypergraph in the same spirit as the Erdos-Renyi graph. This framework also allows us to introduce different ingredients such as the preferential attachment for hypergraphs, or spatial random hypergraphs. In particular, we show that for the Erdos-Renyi case there is a transition threshold scaling as $1/\sqrt{EN}$ where N is the number of nodes and E the number of hyperedges. We also discuss a random geometric hypergraph which displays a percolation transition for a threshold distance scaling as $r_c^* \sim 1/\sqrt{E}$. For these various models, we provide results for the most interesting measures, and also introduce new ones in the spatial case for characterizing the geometrical properties of hyperedges. These different models might serve as benchmarks useful for analyzing empirical data.

DOI: 10.1103/PhysRevE.106.064310

I. INTRODUCTION

Complex networks became increasingly important for describing a large number of processes and were the subject of many studies for more than 20 years now [1–3]. Recent analysis of complex systems [4–7], however, showed that networks provide a limited view. Indeed, networks (or graphs) describe a set of pairwise interactions and exclude any higher-order interactions involving groups of more than two units. With the increasing amount of data, many higher-order interactions were observed in a large variety of contexts, including systems biology [8], face-to-face systems [9], collaboration teams and networks [10,11], ecosystems [12], the human brain [13,14], document clusters in information networks, multicast groups in communication networks, etc. [15,16]. Modeling these higher-order interactions with graphs might lead to erroneous interpretations, calling for the need of a more flexible framework. In addition, these higher-order interactions are highly relevant for all possible processes that take place in these systems [6]. These processes include contagion where a disease can spread in a nondyadic way [17–19], diffusion [20,21], cooperative processes [22], and synchronization [23–25]. Models for hypergraphs—null or spatial, for example—are then much needed for analyzing processes that involve the interaction between more than two nodes.

In order to go beyond usual graphs, a natural extension consists in allowing edges that can connect an arbitrary number of nodes. These "hyperedges" constructed over a set of vertices define what is called a hypergraph [26,27]. More formally, a hypergraph is defined as $H = (V, E)$ where V is a set of elements (the vertices or nodes) and E is the set of hyperedges where each hyperedge is a nonempty subset of V (a simple example is shown in Fig. 1). For a directed hypergraph, the hyperedges are not sets, but an ordered pair of subsets of V, constituting the tail and head of the hyperedge [28–30].

The number $N = |V|$ of vertices is called the order of the hypergraph, and the number of hyperedges $M = |E|$ is usually called the size of the hypergraph. The size of a hyperedge $|e_i|$ is the number of its vertices. The degree of a vertex is then simply given by the number of hyperedges to which it is connected. A simpler hypergraph considered in many studies is obtained when all hyperedges have the same cardinality d and is then called a d-uniform hypergraph (the rank of a hypergraph is $r = \max_E |e|$ and the antirank $\overline{r} = \min_E |e|$, and when both quantities are equal the hypergraph is uniform). A two-uniform hypergraph is then a standard graph.

Many measures are available for hypergraphs. Walks, paths, and centrality measures can be defined and other measures such as the clustering coefficient can be extended to hypergraphs [5,16,31–33]. A recent study investigated the occurence of higher-order motifs [34] and community detection was also considered [35,36]. New measures can, however, be defined and we will discuss in particular the statistics of the intersection between two hyperedges as being the number of nodes they have in common [36]. For spatial hypergraphs, the geometrical structure of hyperedges is naturally of interest and we will discuss some quantities that characterize it.

It has already been noted in [16] that few flexible null models were proposed in the context of interactions occurring within groups of vertices of arbitrary size. In [16,37], a null model for hypergraphs at fixed node degree and edge dimensions is proposed. This generalizes to hypergraphs the usual configuration model [38]. Other models for higher-order

*marc.barthelemy@ipht.fr

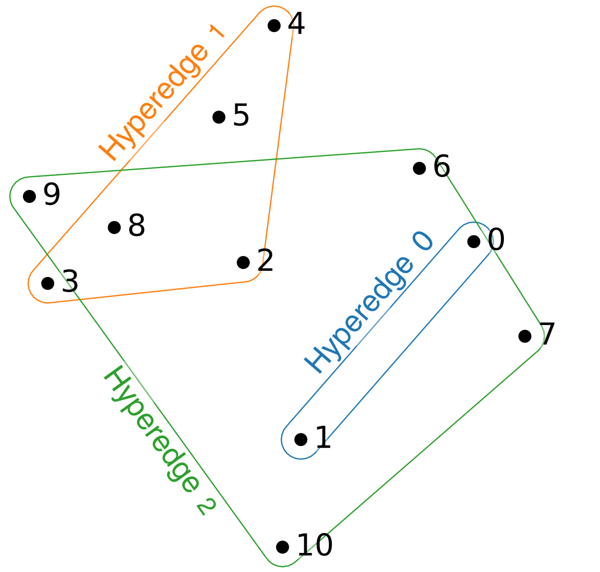

FIG. 1. Classical representation of a hypergraph. We have here $|V| = N = 10$ vertices and $E = 3$ hyperedges. The hyperedges are $e_0 = \{0, 1\}$, $e_1 = \{2, 3, 4, 5, 8\}$, and $e_2 = \{0, 6, 7, 9, 10\}$, with sizes 2, 5, and 5, respectively. All the nodes have degree $k = 1$, except for nodes 0, 1, 2, and 8 which have a degree $k = 2$.

interactions were described in the review [5] such as bipartite models, exponential graph models, or motifs models, but in general hypergraph models are less developed. Most of these models are introduced in the mathematical literature and are usually thought of as immediate generalizations of classical graph models. In many of these models, it is usually assumed that all hyperedges have the same size, which is a strong constraint. We note that an interesting hypernetwork growth model was proposed in [39] where both the idea of hyperedge growth and hyperedge preferential attachment were introduced.

Here, the goal is to present a mechanism that can generate (in a simple way and with low complexity) a family of different hypergraphs and which displays the same level of flexibility of complex network models. We propose here such a framework and introduce a class of models that relies on a single ingredient: the probability that a node belongs to a hyperedge. We will consider various illustrations of this class of models. We will start with the simplest one that could be seen as some sort of "Erdos-Renyi hypergraph", and also discuss preferential attachment. We then introduce space in different ways and in particular discuss in more detail a random geometric hypergraph where the probability that a vertex belongs to a hyperedge is one if its distance is less than a threshold, zero otherwise. For all these illustrations, we analyze simple measures (such as the degree of vertices or sizes of the hyperedges), the structural transition for the giant component, and also introduce new measures. In particular, for spatial hypergraphs, we characterize the spatial properties of hyperedges.

II. A CLASS OF HYPERGRAPH MODELS

Hypergraphs can be represented as bipartite graphs between the nodes and hyperedges (see Fig. 2 for an example).

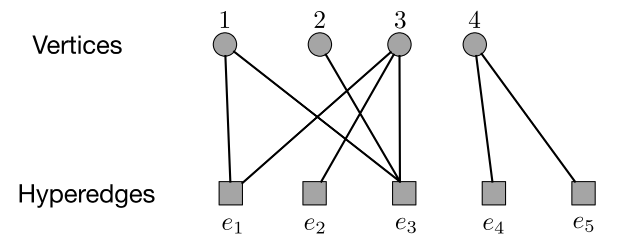

FIG. 2. Example of a small hypergraph with $N = 4$ vertices and $E = 5$ hyperedges represented as a bipartite graph. The degree of the vertices are $k_1 = 2$, $k_2 = 1$, $k_3 = 3$, and $k_4 = 2$, and the sizes of the hyperedges are $|e_1| = 2$, $|e_2| = 1$, $|e_3| = 3$, and $|e_4| = |e_5| = 1$. There are two connected components in this hypergraph $\{1, 2, 3\}$ and $\{4\}$. The intersection between some of the hyperedges is: $e_1 \cap e_2 = \{3\}$, $e_1 \cap e_3 = \{1, 3\}$, etc. The sum $\sum_i k_i$ is equal to the sum of hyperedge sizes $\sum_j |e_j|$ and equal to the number of links.

In this representation the links between nodes and hyperedges indicate a membership relation: there is a link between node i and hyperedge e if $i \in e$. The main idea of the class of models discussed here is to introduce the connection probability $P(v \in e)$ that a node v belongs to a hyperedge e. This is directly related to the incidence matrix I of the hypergraph which is a $N \times E$ matrix with elements I_{ie} equal to one if $i \in e$ and zero otherwise. This connection probability can be written under the form

$$P(v \in e) = F(e, d(v, e), ...), \tag{1}$$

where F is, in general, a function of the hyperedge e, its vertices, or some distance between v and e (we will see various examples below). We note here that in contrast with a remark in [35] that it is "unclear how to impose properties on a hypergraph when a bipartite representation is used", we actually believe that this representation provides a simple framework for introducing various mechanisms. We could even think of a generalization of Eq. (1) in the spirit of the hidden-variable model for networks. In these models, some characteristics are assigned to nodes such as fitnesses [40,41] or coordinates in a latent space [42–44]. The connection function F in Eq. (1) could in principle depend on these hidden variables. For example, if each vertex v has a fitness η_v, a hyperedge composed of vertices $v_{i_1}, v_{i_2}, \dots, v_{i_m}$ has then a fitness which will be a function of the fitnesses of all its vertices $\eta(e) = G(\eta_{i_1}, \eta_{i_2}, \dots, \eta_{i_m})$, and the connection function could then be chosen as

$$P(v \in e) = F(\eta(e)). \tag{2}$$

In this article, we will essentially consider the case where the function depends on some properties of the hyperedege e, including its position in space.

This definition [Eq. (1)] is obviously very general and we will focus here on different functions. First, we will consider the constant case $F = p \in [0, 1]$ which reminds us of the Erdos-Renyi model [45]. We will then consider the preferential attachment case where the function F depends on the size of the hyperedge. We then end this paper by considering cases where the nodes are in a 2D space (we will mainly consider the $D = 2$ case but the generalization to larger dimensions is trivial) and where the function F depends on a distance (to be defined) between the vertex v and the edge e. For all these models, we will present the result for usual measures but also develop new measures tailored to spatial hypergraphs.

Throughout this work, we denote by N the order of the hypergraph (which is the number of nodes) and E the number of hyperedges. For all the models considered here, we will assume that the number E of hyperedges is given. Once an initial set of hyperedges is given (in cases studied here, the initial hyperedges comprise single nodes chosen at random), we can iterate over all nodes and apply Eq. (1). Once all nodes are tested, we get the final hypergraph that we measure. The simplest measure is the degree k_i of a node i (which is the number of hyperedges to which it belongs). The degree can also be computed as the sum of row elements of the incidence matrix I_{ie}: $k_i = \sum_e I_{ie}$ (see, for example, the review [5]). The size $m_l = |e_l|$ of the hyperedge l is the number of nodes it contains. Naturally, the average degree and size (and the distribution of k and m) are important quantities, and it should be noted that they are not independent. Indeed, if we use the links in the bipartite representation of the hypergraph (i.e., there is a link between node i and hyperedge e if $i \in e$, see Fig. 2), their number L can be expressed in two different ways as $L = \sum_i k_i = \sum_j |e_j|$. This relation can be rewritten as

$$N\langle k\rangle = E\langle m\rangle, \tag{3}$$

where $\langle k\rangle$ denotes the average degree and $\langle m\rangle$ the average size of hyperedges. The distributions $P(k)$ and $P(m)$ are then natural objects to study. We will also focus on other quantities: the intersection between two hyperedges (a quantity that is trivial for graphs and equal to one), and for spatial hypergraphs the spatial extension $s(e)$ of a hyperedge e. We will also discuss connectivity properties of these hypergraphs and in particular we will focus on abrupt structural changes characterized by the emergence of a giant component (that will need to be defined) scaling with N.

III. THE SIMPLEST RANDOM HYPERGRAPH

A. Definition

There is no unique definition of random hypergraphs and various generalization of the classical Erdos-Renyi graph were proposed (see for example [11]). In particular, mathematicians like to think of a set of all possible hypergraphs—given some parameters (number of nodes, etc)—and to consider a uniform distribution over this set [46]. More specifically, many papers focus on k-uniform hypergraphs for which the size of hyperedges is constant and equal to k. Given a set of V vertices and a set of subsets of these vertices we can construct the natural analog of Erdos-Renyi graphs [45]: each k-tuple of vertices is a hyperedge with probability p [47] (more formally, the set $H^k(n, p)$ denotes the random k-uniform hypergraph with vertex set $[n] = \{1, 2, ..., n\}$ in which each of the $\binom{n}{k}$ possible edges is present independently with probability p). In other studies, every hypergraph of E hyperedges on N nodes has the same probability [48] (the classical Erdos-Renyi random graph is then recovered in the case $k = 2$). This type of model was discussed by mathematicians in particular about the phase transition for the giant component [4,46,47,49–51]. For $k = 2$, we recover usual graphs and from Erdos and Renyi [45], we know that there is a transition for $E = cN$ with $c = 1/2$ (which corresponds to an average degree $\langle k\rangle = 2E/M = 1$). This transition is characterized by an abrupt structural change with the emergence of a giant component. The general result for k-uniform hypergraphs obtained in [49] is similar and states that a giant component appears for $c = 1/k(k-1)$. We refer the reader interested in some mathematical aspects of this problem to the book [52].

Here, we will construct an "Erdos-Renyi hypergraph" in the context of Eq. (1). For the usual Erdos-Renyi graph, the connection probability between two nodes is constant, and similarly for the hypergraph we will then assume that for each vertex $v \in V$ and for each hyperedge e, there is a constant probability p that v belongs to e. If this probability is constant and equal to p, we can then write

$$P(v \in e) = p, \tag{4}$$

where $p \in [0, 1]$. Starting from a set of N nodes, we choose a random set of E hyperedges (which are E nodes chosen at random), and recursively add nodes to these hyperedges following the rule of Eq. (4).

B. Degree and size

This random hypergraph is a very simple model and most properties are trivial. In particular, the degree k and size m distributions are easy to compute and are binomials

$$P(k_i = k) = \binom{E}{k} p^k (1-p)^{N-k}, \tag{5}$$

$$P(|e_i| = m) = \binom{N}{m} p^m (1-p)^{E-m} \tag{6}$$

and the average degree is then $\langle k\rangle = pE$ and the average hyperedge size $\langle m\rangle = pN$.

If nodes are in space—in the disk of radius r_0 for example (we will see below in more details examples of spatial hypergraphs constructed over such a set)—we can compute the average spatial extent of a hyperedge e_i given by

$$s(e_i) = \frac{1}{m_i(m_i-1)} \sum_{v_j, v_l \in e_i} d_E(v_j, v_l), \tag{7}$$

where $m_i = |e_i|$ is the size of the hyperedge and where $d_E(v_j, v_l)$ is the euclidean distance between nodes v_j and v_l. In the Erdos-Renyi hypergraph case, the nodes of hyperedges are distributed uniformly and the extent is given by the average distance between randomly chosen nodes in the disk [53]

$$s = s_0 = \frac{128 r_0}{45\pi}, \tag{8}$$

which is verified in our numerical simulations (figure not shown).

C. Hyperedge intersection

The intersection I between two hyperedges e_i and e_j is the number of nodes they have in common [54] and is denoted by

$$I = |e_i \cap e_j|, \tag{9}$$

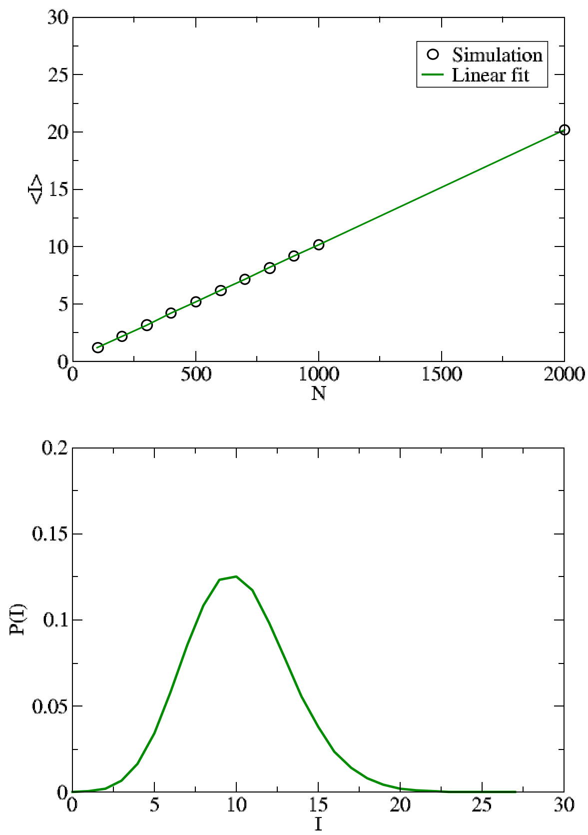

FIG. 3. (a) Average intersection $\langle I\rangle$ versus N. The line is a linear fit of slope p^2 (here $p = 0.1$ and results were obtained by averaging over 100 configurations). (b) Probability distribution $P(I)$ of the intersection between hyperedges defined in Eq. (9). Simulations were obtained for $N = 1000$, $E = 100$, $p = 0.1$ and averaged over 100 configurations.

where $|\cdot|$ denotes the cardinal of a set. For the random hypergraph discussed here, the probability that a given node belongs to two different hyperedges is p^2. The average intersection is then given by $\langle I\rangle = p^2N$ which we verified by simulations [see Fig. 3(a)]. The probability distribution of the intersection I between two hyperedges is then a binomial of parameters N and p^2

$$P(I = n) = \binom{N}{n} p^{2n}(1-p^2)^{N-n}, \tag{10}$$

which can be verified numerically [see Fig. 3(b)].

We can go further and define the intersection I_{jl} of two hyperedges of size j and l, respectively [54]. The intersection I_{jl} is then a random variable and can be expressed for the random hypergraph defined here. Indeed, the probability $P(I_{jl} = k)$ is given by the number $N_{jl}(k)$ of events where the intersection of a subset of size j and a subset of size l is of size k (see Fig. 4). The number $N_{jl}(k)$ is then given by the following multinomial coefficient

$$N_{jl}(k) = \frac{N!}{k!(j-k)!(l-k)!(N-j-l+k)!} \tag{11}$$

and the corresponding probability is $P_{jl}(k) = N_{jl}(k)/\sum_k N_{jl}(k)$ (if needed a large N analysis could then be performed).

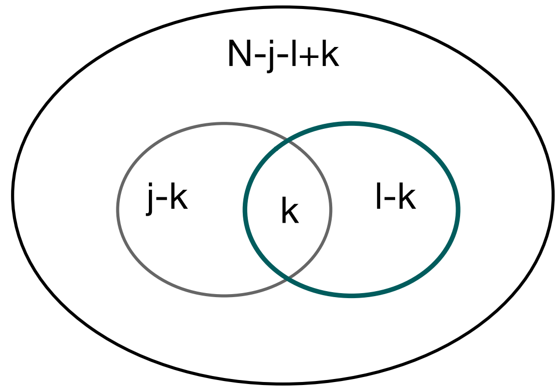

FIG. 4. Schematic illustration of the intersection between two hyperedges of respective sizes j and l.

D. Giant component

We now consider the behavior of the giant component when p varies. This problem actually motivated many studies of random graphs, starting with the original paper by Erdos and Renyi [45]. In order to define the giant component we assume that all nodes in the same hyperedge are connected to each other (equivalently that each hyperedge is a complete graph) and that two hyperedges are connected if their intersection is at least equal to one. Other definitions are possible using the concept of a high-order hypergraph walk [33]. A hypergraph walk is called a s walk where the order s controls the minimum edge intersection over which the walk takes place. A one-walk is then the usual walk and this is the connectivity that we are using here for defining the giant component: two nodes are connected if there is at least one one-walk between them (see [33] for a discussion about larger values of s).

In the case of random hypergraphs considered here, the nodes in each hyperedge form a connected clique and the giant component problem lies essentially in the connection between the hyperedges. The probability that there is at least one intersection between two hyperedges is given by

$$P(I \geqslant 1) = 1-(1-p^2)^N$$
$$\approx p^2N, \tag{12}$$

for $p^2N \ll 1$. If the number of hyperedges E is large, there is a giant component in the hypergraph if the E hyperedges constitute a giant component. The classical result for Erdos-Renyi random graph states that there is a giant component appearing at average degree equal to one which leads here to the condition $p_c^2NE = 1$. The threshold then behaves as

$$p_c \sim \frac{1}{\sqrt{NE}}. \tag{13}$$

We note that this result obtained by a simple argument has been already found by more rigorous methods for random bipartite graphs in [55].

IV. PREFERENTIAL ATTACHMENT FOR HYPERGRAPHS

The preferential model for networks states that the probability that a new node n connects to an existing one i is proportional to the degree k_i of i [56]. It might be interesting to extend this reinforcement mechanism to hypergraphs. Few studies discussed this apart from the notable exception of [39] and [37,57,58]. In the model proposed by [39], two important ingredients were introduced for the growth of a random hypergraph. First, at every time step, a new hyperedge e is constructed (either with m new nodes and a randomly chosen node in the existing hypergraph or from a random number of nodes selected at random in the existing hypergraph and a new node such as in [59]). Second, they introduced a hyperedge preferential attachment where the hyperedge e is connected to an existing node i with probability proportional to the degree of i (also called hyperdegree in this study) [39]. This was generalized in [60]. In [39], it was shown that the hypergraph constructed in this way shares many similar features with complex networks such as scale-free property of the degree distribution, etc. [39,59].

Clearly, there are several ways to introduce preferential attachment in the hypergraph formation, but in the framework defined by Eq. (1), the natural choice is to write the connection probability as a function of the hyperedge size $m = |e|$:

$$P(v \in e) = F[|e|], \tag{14}$$

and the simplest function is the linear one

$$P(v \in e_i) = \frac{|e_i|}{\sum_j |e_j|}, \tag{15}$$

which introduces a rich-get-richer process through the size of hyperedges. Other choices could include, for example, the average degree of nodes contained in the hyperedge e, etc.

In this simple model, the degree of each vertex is $k = 1$ (a simple generalization would consist in taking for each vertex n possible connections to different hyperedges). The probability that a given vertex connects to a given hyperedge is given by Eq. (15) and this problem is exactly a Polya urn with E colors. The limiting distribution for large time (and therefore at large number of nodes N) can be shown to be the Dirichlet multinomial distribution [61]

$$\begin{aligned} &P_t(m_1, m_2, \dots, m_E|\alpha_1, \alpha_2, \dots, \alpha_E) \\ &= \frac{\Gamma(\alpha_0)\Gamma(N+1)}{\Gamma(N+\alpha_0)} \prod_{j=1}^{E} \frac{\Gamma(m_j+\alpha_j)}{\Gamma(\alpha_j)\Gamma(m_j+1)}, \end{aligned} \tag{16}$$

where $m_j = |e_j|$, α_j denotes the initial size of hyperedge j, $\alpha_0 = \sum \alpha_m$ and $\Gamma(x)$ is the gamma function. In particular, if we start with all the hyperedges having the same size, the limiting distribution is uniform. More generally, it shows that in this case the limiting distribution depends crucially on the initial structure of hyperedges, which certainly renders the empirical identification of a preferential attachment mechanism difficult.

This multinomial Dirichlet distribution is a bit difficult to test but we can easily compute its first two cumulants. Indeed, for an initial condition given by $\alpha_1, \alpha_2, \dots, \alpha_E$, the average size of edge m_i is given by

$$\langle m_i \rangle = N\frac{\alpha_i}{\alpha_0}. \tag{17}$$

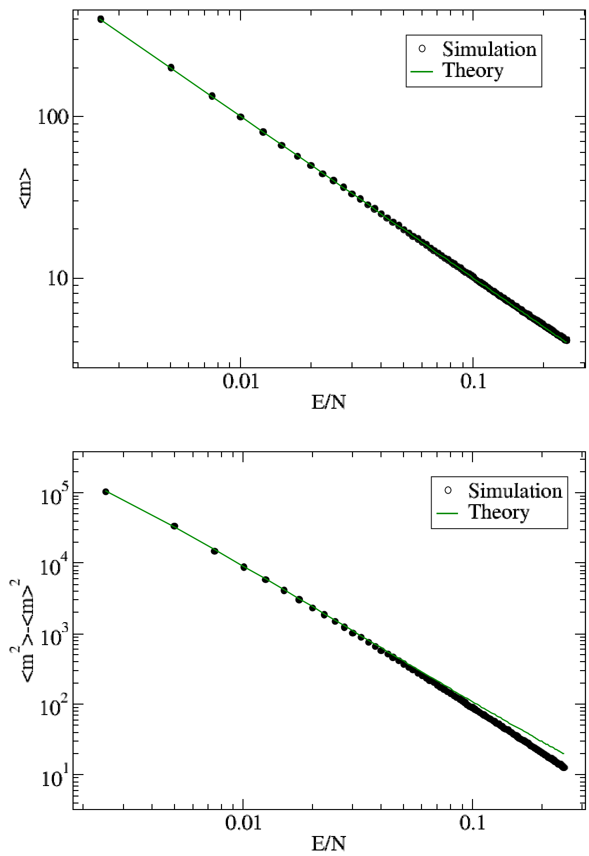

FIG. 5. Comparison of simulations and theoretical results obtained from the multinomial Dirichlet distribution Eq. (16) for the uniform case. The symbols are obtained for numerical simulations for $N = 1000$ (averages are computed over 1000 configurations) and the line represent the theoretical result Eqs. (17), (18). (a) Average hyperedge size versus the number E of hyperedges. (b) Variance of the hyperedge size versus E.

For the variance, we note that the marginal distribution is a Beta-Binomial which then leads to the result

$$\mathrm{Var}(m_i) = N\frac{\alpha_i}{\alpha_0}\left(1 - \frac{\alpha_i}{\alpha_0}\right)\frac{N+\alpha_0}{1+\alpha_0} \tag{18}$$

In the uniform case (all the α_is are equal to some value α), we then obtain $\langle m_i \rangle = N/E$, and $\mathrm{Var}(m) = N/E(1 - 1/E)(N + E\alpha)(1 + E\alpha)$. We can test this result by varying E for example, and we obtain the results shown in Fig. 5. The agreement between the numerical simulation and these results is excellent in the case of the average [Fig. 5(a)]. In the case of the variance, we observe a small deviation from the theoretical result for E becoming closer to N. In this case, the average size of the hyperedge can be small (of order N/E) and fluctuations can be large (the deviation then decreases with the number of configurations).

V. RANDOM SPATIAL HYPERGRAPHS

It is reasonable to think that in some instances, introducing space is necessary. For example, contagion among a group of individuals naturally involves space through the proximity needed to transmit an infectious disease. Other systems where

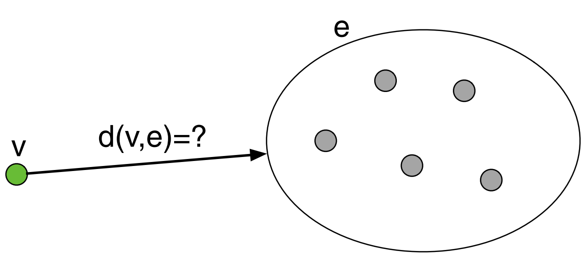

FIG. 6. Schematic illustration of the choice when constructing a spatial hypergraph. A new node v enters the hypergraph and will connect to the hyperedge e. The problem is how to compute the distance $d(v,e)$ between the node and the hyperedge.

space is relevant (communications networks, neural networks, etc) and where higher-order interactions take place can be described by what we could call "spatial hypergraphs" and we will discuss here some simple examples of such objects,

We assume that the N nodes are distributed uniformly on a disk of radius r_0. Each node i has a position x_i and it is then natural to consider a model defined by the following connection probability

$$P(v\in e)=F[d(v,e)], \tag{19}$$

where F is a given function, and $d(v,e)$ measures the distance between the vertex v and the hyperedge e. There are many choices for defining $d(v,e)$ and many models are possible (Fig. 6).

It is obviously hopeless (and probably useless, too) to try to explore all possible cases, and we will focus on two main models. First, we will assume that the function F decreases with the distance as an exponential $F(d)\sim\exp(-d/r_c)$. We will then consider a model close—in spirit at least—to the random geometric graph [62].

A. Exponential case

We consider the case where F is an exponential function decreasing with distance (we expect similar results for other decreasing functions). This corresponds to the intuitive idea that it is more difficult for a node to belong to a distant hyperedge. The range of the exponential is denoted by r_c and the connection probability then reads as

$$P(v\in e)=p\mathrm{e}^{-d(v,e)/r_c}, \tag{20}$$

where $p\in[0,1]$ and $d(v,e)$ is a measure of the distance between the node v and the hyperedge e. When $r_c\gg r_0$, the exponential term is essentially one, space is then irrelevant and we recover the random hypergraph model discussed above. For the distance $d(v,e)$, many choices are possible, such as the average over all nodes, the minimum or maximum distance among the nodes, and we will essentially present the results for the average distance

$$d(v,e)=\frac{1}{m}\sum_{w\in e}d_E(v,w), \tag{21}$$

where $m=|e|$ is the size of the hyperedge e and $d_E(v,w)$ is the euclidean distance between v and w.

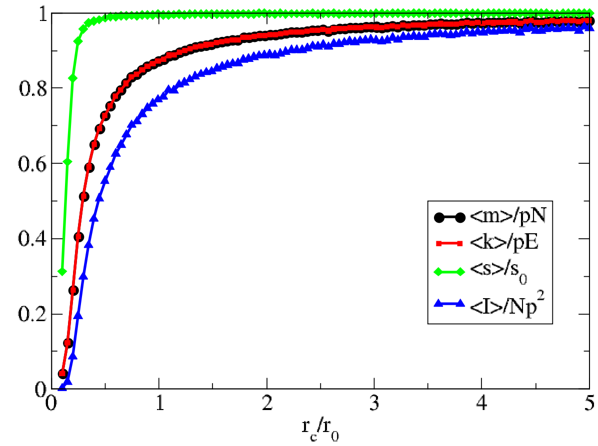

FIG. 7. Average degree $\langle k\rangle$, hyperedge size m, intersection $\langle I\rangle$, and hyperedge extent $\langle s\rangle$ versus r_c/r_0 and normalized by their value for the random hypergraph (for $p=0.1$, $N\in[100,1000]$, $E\in[100,N]$, 100 configurations).

For $r_c\gg r_0$ we recover the random hypergraph for which we know most quantities. We can then rescale the average degree by pE, the average hyperedge size by pN, the average intersection by Np^2, and the average spatial extent of hyperedges by $128r_0/45\pi$. These quantities versus r_c are shown in Fig. 7. The total number of links is given by $\sum_j k_j=\sum_i|e_i|$, which implies that $\langle m\rangle/pN=\langle k\rangle/pE$ as observed in Fig. 7. Also, and as expected, all these quantities grow with r_c, but at different speeds.

We can consider other choices for the distance $d(v,e)$ instead of Eq. (21). For example, the minimum distance is also a reasonable choice that can make sense for some systems

$$d(v,e)=\min_{w\in e}d_E(v,w). \tag{22}$$

Also, in the case where we define the centroid $c(e)$ of the hyperedge with coordinates $(x_c(e),y_c(e))=1/|e|\sum_{i\in e}(x(i),y(i))$, the distance can be computed from this centroid

$$d(v,e)=d(v,c(e)). \tag{23}$$

The corresponding hypergraph model bears some similarity with the k-means clustering method (see for example [63]). This method partitions N observations into k clusters, with the rule that each node belongs to the cluster with the nearest centroid location. In the hypergraph case, each time a new node is attached to a hyperedge, the centroid position is recalculated, as in the Lloyd algorithm [64]. We averaged here over all possible initial positions of the hyperedges, but as in the k-means clustering case, some further work is probably needed in order to understand the effect of initialization on the resulting hypergraph structure.

We compare the convergence to the random case of the average spatial extent of hyperedges $\langle s\rangle$ for these three different choices Eqs. (21), (22), and (23) and the results for $\langle s\rangle$ versus r_c are shown in Fig. 8. We observe different speeds to convergence to the random case. In particular, the cases Eqs. (21) and (23) behave similarly, while the case Eq. (22) seems to converge in a much faster way. The average spatial

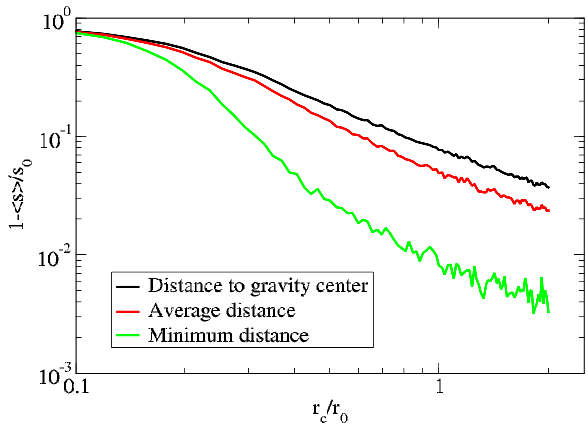

FIG. 8. One minus the average normalized spatial extent of hyperedges $1-\langle s\rangle/s_0$ versus r_c shown in loglog (results are obtained for $E=100$, $N=1000$ and 100 configurations). A power law fit over the last decade gives an exponent going from 1.2 to 1.5 for the different cases.

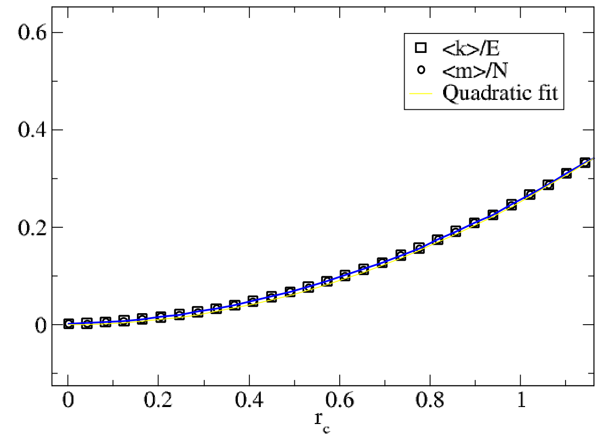

FIG. 9. Average hyperedge size $\langle m\rangle/N$ and degree $\langle k\rangle/E$ normalized by the number of nodes and the number of hyperedges, respectively. The line is a quadratic fit of the form ar_c^2 where $a=0.25$ here ($N=1000$, $E=100$, 100 configurations).

extent of hyperedges thus seems to be very sensitive to the choice of the connection probability.

B. A random geometric hypergraph

1. Definition

The random geometric graph is a classical spatial graph introduced by Gilbert [62] where nodes are located in the 2D plane and are connected if their distance is less than a threshold r_c (for mathematical properties of this object, see [65]). If we denote by $\rho=N/A$ the density of nodes in the area A, the average degree is given by

$$\langle k\rangle=\rho\pi r_c^2. \tag{24}$$

There is a critical value k_c for this quantity, above which there is a giant cluster of size of order N. The value of k_c is not exactly known but is approximately given by $k_c\approx 4.5$ (see for example [66] and references therein).

There are various possibilities to extend to hypergraphs this idea of the random geometric graph. For example, in [35], the authors added hyperedges connecting all nodes that are at a distance less than a threshold r_c. Varying this threshold r_c then gives a sequence of hyperedges included in each other. In the framework discussed here, we choose for the connection function the following form

$$P(v\in e)=\theta(r_c-d(v,e)), \tag{25}$$

where $d(v,e)$ is a distance between the vertex v and the hyperedge e ($\theta(x)$ is the Heaviside function). As in the previous section, there are several possible choices for defining $d(v,e)$ and here we choose

$$d(v,e)=\max_{w\in e} d(v,w). \tag{26}$$

This choice corresponds to the intuitive idea that a vertex belongs to a hyperedge if all of its vertices are close enough (and at a distance less than r_c).

2. Average degree and size

Some properties of this "random geometric hypergraph" can be discussed with simple scaling arguments. First, the probability that a vertex belongs to a hyperedge is (for a uniform distribution of nodes) given by $\pi r_c^2/\pi r_0^2$ (here the size of the disk is $r_0=1$). The average degree is then

$$\langle k\rangle=Er_c^2. \tag{27}$$

Similarly, the average size of the hyperedges is given by the number of nodes in their vicinity at a distance less than r_c. Their size is then simply given by

$$\langle m\rangle\approx\rho r_c^2\sim Nr_c^2. \tag{28}$$

These results are consistent with the general relation $\langle m\rangle/N=\langle k\rangle/E$, and imply a behavior r_c^2. Both these results are perfectly verified in numerical simulations (in Fig. 9 we show the quadratic fit).

3. Hyperedge intersection

The average extent of a hyperedge here is trivially given by $\langle s\rangle\approx r_c$ and is not a very interesting measure here. More interesting is the intersection I between two hyperedges. In order to estimate this quantity, we first consider the area $A(r,\ell)$ defined by the intersection of two disks with the same radius r and separated by a distance ℓ. Its expression can be found by elementary geometry

$$A(r,\ell)=2\cos^{-1}(\ell/2r)r^2-\frac{\ell}{2}\sqrt{4r^2-\ell^2}, \tag{29}$$

for $\ell\leqslant 2r$ and $A=0$ for $\ell>2r$. The probability p_I that a node belongs to the intersection of two hyperedges separated by a typical distance $\ell\sim 1/\sqrt{E}$ is then given by

$$p_I\approx\frac{A(r_c,1/\sqrt{E})}{\pi r_0^2}. \tag{30}$$

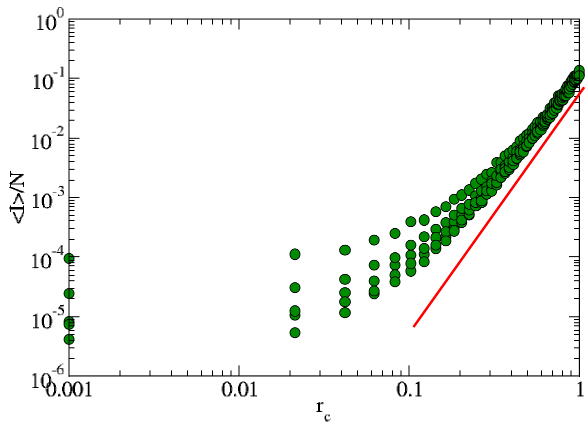

FIG. 10. Average intersection $\langle I \rangle$ rescaled by N versus r_c for various values of N from 200 to 1000 (here $E = 100$). We observe for large $r_c \gg 1/\sqrt{E}$ a good collapse behaving as r_c^2. The straight line meant as a guide to the eye is a power law with exponent two.

and the average intersection is $\langle I \rangle = p_I N$. It is zero for $r_c \lesssim 1/\sqrt{E}$ and for large r_c behaves as $\langle I \rangle \sim N r_c^2$. We thus plot $\langle I \rangle / N$ versus r_c and we indeed observe a quadratic behavior for large r_c (see Fig. 10).

4. Giant component: Transition

As discussed for the random hypergraph defined above, we need a definition for the connectivity in order to compute the giant component. As above, we will consider one-walks and that all nodes in the same hyperedge are connected to each other (equivalently that each hyperedge is a clique or a complete subgraph) and that two hyperedges are connected if their intersection is at least equal to one. With this definition, we can compute the largest component and see how it varies with r_c. We obtain the result shown in Fig. 11(a) displaying an abrupt transition for a value $r_c = r_c^*$. In order to estimate this threshold r_c^*, we propose the following argument. The hyperedges can be seen as different clusters of size r_c, and the existence of a giant component can then be mapped to the problem of continuum percolation of E disks of radius r_c. It is well known (see for example [67]) that percolation in this case is reached for

$$\rho_D a = \eta_c, \tag{31}$$

where $\rho_D = E/A$ is the density of disk (here A is the total area given by $A = \pi r_0^2$) and $a = \pi r_c^2$ is the area of the disks. The threshold quantity η_c has been estimated numerically and is approximately $\eta_c \approx 1.12$ for 2D continuum percolation (see for example [68]). The critical value for r_c^* is then behaving for large E as

$$r_c^* \sim \frac{r_0}{\sqrt{E}} \tag{32}$$

This result is consistent with simulations shown in Fig. 11(b).

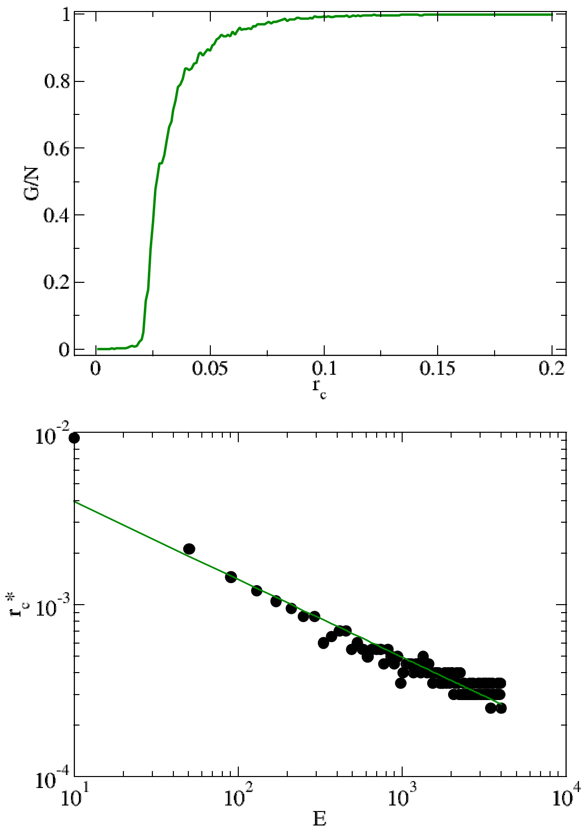

FIG. 11. (a) Size G of the giant component normalized by the number of nodes N versus the interaction range r_c for the random geometric hypergraph defined in Eq. (25) (here $N = 2000$, $E = 100$, and averaged over 100 configurations). (b) Critical radius value r_c^* versus E. The line is a power law fit of the form $r_c^* \sim 1/E^\tau$ with $\tau \approx 0.45$ ($r^2 = 0.95$). Simulations are done for 100 configurations and $N = 5000$.

VI. DISCUSSION

The observed relevance of higher-order interactions in empirical data pushed the scientists interested in complex systems to go beyond usual graphs and to consider random hypergraphs (or other models). It could even be possible that in the future, we have to extend hypergraphs to multilayered structures as discussed in [58,69]. The literature about hypergraph modeling is very heterogeneous, and sometimes difficult to grasp, and didn't reach the state-of-the-art observed for complex networks.

Here, we contributed to the modeling of these higher-order interactions and explored a particular class of random hypergraphs where the number of hyperedges is given and where their size is determined by some sort of hidden-variable modeling. Many alternatives are certainly possible, but the main advantage of this framework is its flexibility (with the drawback of fixing the number of hyperedges, a constraint that could probably be lift off in future models). An important purpose of this article is to highlight the vast space of possible hypergraph models that are left to be explored.

Many directions for future studies can be envisioned. In particular, for spatial hypergraphs, it would be interesting to generalize the standard models of graphs such as the Gabriel or Delaunay graphs, beta-skeletons, etc. Such models could in particular be helpful for understanding the impact of space on some processes over hypergraphs such as contagion or diffusion, for example. It would also be interesting to consider hypergraph models based on optimal considerations. For example, can we construct the equivalent of the minimum spanning tree for spatial hypergraphs, or more generally, can we define optimal hypergraphs? The field of hypergraphs is certainly not as mature as complex networks but the recently revealed interest in these higher-order interaction structures will certainly trigger many interesting studies and we can hope to see beautiful results in the future.

ACKNOWLEDGMENTS

I warmly thank Ginestra Bianconi, Jean-Marc Luck, Kirone Mallick, and Erwan Taillanter for stimulating discussions.

[1] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, UK, 2008).
[2] V. Latora, V. Nicosia, and G. Russo, *Complex Networks: Principles, Methods and Applications* (Cambridge University Press, Cambridge, UK, 2017).
[3] F. Menczer, S. Fortunato, and C. A. Davis, *A First Course in Network Science* (Cambridge University Press, Cambridge, UK, 2020).
[4] G. Ghoshal *et al.*, Random hypergraphs and their applications, Phys. Rev. E **79**, 066118 (2009).
[5] F. Battiston *et al.*, Networks beyond pairwise interactions: Structure and dynamics, Phys. Rep. **874**, 1 (2020).
[6] F. Battiston *et al.*, The physics of higher-order interactions in complex systems, Nature Phys. **17**, 1093 (2021).
[7] G. Bianconi, *Higher Order Networks: An Introduction to Simplicial Complexes* (Cambridge University Press, Cambridge, UK, 2021).
[8] S. Krieger and J. Kececioglu, *Fast Approximate Shortest Hyperpaths for Inferring Pathways in Cell Signaling Hypergraphs, Schloss Dagstuhl-Leibniz-Zentrum fur Informatik GmbH* (Dagstuhl Publishing, Wadern, Germany, 2021).
[9] G. Cencetti, F. Battiston, B. Lepri, and M. Karsai, Temporal properties of higherorder interactions in social networks, Sci. Rep. **11**, 7028 (2021).
[10] A. Patania, G. Petri, and F. Vaccarino, The shape of collaborations, EPJ Data Sci. **6**, 18 (2017).
[11] J. L. Juul, A. R. Benson, and J. Kleinberg, Hypergraph patterns and collaboration structure, arXiv:2210.02163.
[12] J. Grilli, G. Barabas, M. J. Michalska-Smith, and S. Allesina, Higher-order interactions stabilize dynamics in competitive network models, Nature (London) **548**, 210 (2017).
[13] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer, and F. Vaccarino, Homological scaffolds of brain functional networks, J. R. Soc. Interface **11**, 20140873 (2014).
[14] C. Giusti, R. Ghrist, and D. S. Bassett, Two's company, three (or more) is a simplex, J. Comput. Neurosci. **41**, 1 (2016).
[15] J. Gao, Q. Zhao, W. Ren, A. Swami, R. Ramanathan, and A. Bar-Noy, Dynamic shortest path algorithms for hypergraphs, IEEE/ACM Trans. Networking **23**, 1805 (2014).
[16] P. S. Chodrow, Configuration models of random hypergraphs, Journal of Complex Networks **8**, cnaa018 (2020).
[17] I. Iacopini, G. Petri, A. Barrat, and V. Latora, Simplicial models of social contagion, Nat. Commun. **10**, 2485 (2019).
[18] G. F. de Arruda, G. Petri, and Y. Moreno, Social contagion models on hypergraphs, Phys. Rev. Res. **2**, 023032 (2020).
[19] A. Barrat, G. Ferraz de Arruda, I. Iacopini, and Y. Moreno, *Social Contagion on Higher-Order Structures, Inhigher-Order Systems 2022* (Springer, Cham), pp. 329-346.
[20] M. T. Schaub, A. R. Benson, P. Horn, G. Lippner, and A. Jadbabaie, Random walks on simplicial complexes and the normalized Hodge Laplacian, SIAM Rev. **62**, 353 (2020).
[21] T. Carletti, F. Battiston, G. Cencetti, and D. Fanelli, Random walks on hypergraphs, Phys. Rev. E **101**, 022308 (2020).
[22] U. Alvarez-Rodriguez, F. Battiston, G. Ferraz de Arruda, Y. Moreno, M. Perc, and V. Latora, Evolutionary dynamics of higher-order interactions in social networks, Nat. Hum. Behav. **5**, 586 (2021).
[23] P. S. Skardal and A. Arenas, Higherorder interactions in complex networks of phase oscillators promote abrupt synchronization switching, Commun Phys **3**, 218 (2020).
[24] A. P. Millan, J. J. Torres, and G. Bianconi, Explosive Higher-Order Kuramoto Dynamics on Simplicial Complexes, Phys. Rev. Lett. **124**, 218301 (2020).
[25] M. Lucas, G. Cencetti, F. Battiston, A multi-order Laplacian for synchronization in higherorder networks, Phys. Rev. Research **2**, 033410 (2020).
[26] C. Berge, *Hypergraphs: Combinatorics of Finite Sets* (Elsevier, Amsterdam, Netherlands, 1984), Vol. 45.
[27] A. Bretto, *Hypergraph Theory. An Introduction. Mathematical Engineering* (Springer, Cham, 2013).
[28] G. Gallo, G. Longo, S. Pallottino, and S. Nguyen, Directed hypergraphs and applications, Discrete Applied Mathematics **42**, 177 (1993).
[29] H. Galeana-Sánchez and M. Manrique, Directed hypergraphs: A tool for researching digraphs and hypergraphs, Discuss. Math. Graph Theory **29**, 313 (2009).
[30] G. Ausiello and L. Laura, Directed hypergraphs: Introduction and fundamental algorithms-a survey, Theor. Comput. Sci. **658**, 293 (2017).
[31] V. Zlatic, G. Ghoshal, and G. Caldarelli, Hypergraph topological quantities for tagged social networks, Phys. Rev. E **80**, 036118 (2009).
[32] C. A. Joslyn *et al.*, *Hypernetwork Science: From Multidimensional Networks to Computational Topology, International Conference on Complex Systems* (Springer, Cham, 2020).
[33] S. G. Aksoy, C. Joslyn, C. O. Marrero, B. Praggastis, and E. Purvine, Hypernetwork science via high-order hypergraph walks, EPJ Data Science **9**, 16 (2020).
[34] Q. F. Lotito, F. Musciotto, F. Battiston, and A. Montresor, Exact and sampling methods for mining higher-order motifs in large hypergraphs, arXiv:2209.10241.
[35] K. Turnbull, S. Lunagómez, C. Nemeth, and E. Airoldi, Latent Space Modelling of Hypergraph Data, arXiv:1909.00472.
[36] P. S. Chodrow, N. Veldt, and A. R. Benson, Generative hypergraph clustering: From blockmodels to modularity, Sci. Adv. **7**, eabh1303 (2021).

[37] O. T. Courtney and G. Bianconi, Weighted growing simplicial complexes, Phys. Rev. E **95**, 062301 (2017).

[38] M. Molloy and B. Reed, The size of the giant component of a random graph with a given degree sequence, Combinator. Probab. Comp. **7**, 295 (1998).

[39] J. W. Wang, L. L. Rong, Q. H. Deng, and J. Y. Zhang, Evolving hypernetwork model, Eur. Phys. J. B **77**, 493 (2010).

[40] G. Bianconi and A.-L. Barabasi, Competition and multiscaling in evolving networks, Europhys. Lett. **54**, 436 (2001).

[41] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Munoz, Scale-Free Networks from Varying Vertex Intrinsic Fitness, Phys. Rev. Lett. **89**, 258702 (2002).

[42] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguna, Hyperbolic geometry of complex networks, Phys. Rev. E **82**, 036106 (2010).

[43] M. A. Serrano, M. Boguna, and F. Sagues, Uncovering the hidden geometry behind metabolic networks, Mol. BioSyst. **8**, 843 (2012).

[44] M. Kitsak, F. Papadopoulos, and D. Krioukov, Latent geometry of bipartite networks, Phys. Rev. E **95**, 032309 (2017).

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

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

[47] O. Cooley, M. Kang, and C. Koch, The size of the giant high-order component in random hypergraphs, Random structures & algorithms **53**, 238 (2018).

[48] W. Fernandez de la Vega, Sur la cardinalité maximum des couplages d'hypergraphes aléatoires uniformes, Discrete Mathematics **40**, 315 (1982).

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

[50] O. Cooley, M. Kang, and Y. Person, Largest components in random hypergraphs, Combinator. Probab. Comp. **27**, 741 (2018).

[51] J. H. Kim and K. I. Goh, Higher-order components in hypergraphs, arXiv:2208.05718.

[52] A. Frieze, and M. Karoński, *Introduction to Random Graphs* (Cambridge University Press, Cambridge, UK, 2016).

[53] R. García-Pelayo, Distribution of distance in the spheroid, J. Phys. A: Math. Gen. **38**, 3475 (2005).

[54] P. S. Chodrow and A. Mellor, Annotated hypergraphs: Models and applications, Appl. Network Sci. **5**, 9 (2020).

[55] T. Johansson, *The Giant Component of the Random Bipartite Graph* (Chalmers University of Technology, Gothenburg, Sweden, 2012).

[56] A.-L. Barabasi and R. Albert, Emergence of scaling in random networks, Science **286**, 509 (1999).

[57] G. Bianconi and C. Rahmede, Emergent hyperbolic network geometry, Sci. Rep. **7**, 41974 (2017).

[58] H. Sun and G. Bianconi, Higher-order percolation processes on multiplex hypergraphs, Phys. Rev. E **104**, 034306 (2021).

[59] G. Y. Yang and J. G. Liu, A local-world evolving hypernetwork model, Chinese Phys. B **23**, 018901 (2014).

[60] J. L. Guo, X. Y. Zhu, Q. Suo, and J. Forrest, Non-uniform evolving hypergraphs and weighted evolving hypergraphs, Sci. Rep. **6**, 36648 (2016).

[61] H. Mahmoud, *Pólya Urn Models* (Chapman and Hall/CRC, London, UK, 2008).

[62] E. N. Gilbert, Random graphs, The Annals of Mathematical Statistics **30**, 1141 (1959).

[63] A. K. Jain, M. N. Murty, and P. J. Flynn, Data clustering: A review, ACM Comput. Surv. **31**, 264 (1999).

[64] S. Lloyd, Least squares quantization in PCM, IEEE Trans. Inf. Theory **28**, 129 (1982).

[65] M. Penrose, *Random Geometric Graphs* (Oxford University Press, Oxford, 2003), Vol. 5.

[66] M. Barthelemy, *Spatial Networks: A Complete Introduction: From Graph Theory and Statistical Physics to Real-World Applications* (Springer Nature, Berlin, Germany, 2022).

[67] R. Meester and R. Roy, *Continuum percolation* (Cambridge University Press, Cambridge, UK, 1996), Vol. 119.

[68] S. Mertens and C. Moore, Continuum percolation thresholds in two dimensions, Phys. Rev. E **86**, 061109 (2012).

[69] A. Vazquez, Complex hypergraphs, arXiv:2209.07331.