AM205

Raphaël Pellegrin, Gaël Ancel, Tale Lokvenec

December 2020

1 Abstract

Graphs provide a formalism to represent dyadic interactions occurring in real-life networks - for instance, online social networks, such as Facebook, can be represented as mathematical networks with links corresponding to friendship on Facebook. Hypergraphs generalise graphs by allowing edges of different cardinalities. Hypergraphs can represent complex interactions, such as co-authorship hypernetworks. JT Matamalas and I Iacopini studied contagion models on simplicial complexes [MGA20, IPBL19], which present an abrupt phase transition. We generalize some of their analyses to the more general situation of hypergraphs and to higher orders.

2 Introduction

Hypergraphs are mathematical structures that were introduced by French mathematician Claude Berge in the 1970s [Ber73]. They generalise graphs in the sense that an edge can have any number of vertices. Formally, a hypergraph H is a pair $H = (V, E)$, where V is a set of vertices and E is a set of non-empty subsets of V called hyperedges.

To model real-life situations, dyadic relations are not always enough - there are some cases in which hypergraph representations of the real network are indispensable. The first application of hypergraphs for representing social network dates back to 1981 [ERV06], but scenarios that can be modelled as hypergraphs are ubiquitous. When studying the co-authorship of articles, it might be relevant to consider hyperedges for authors who wrote papers together [XOL16, PPV17, TCR10]. Another application is protein hypergraphs, which, unlike protein-protein interaction networks (networks in which nodes represent proteins and edges links two interacting proteins), take into account the multi-protein complexes [ERV06]. Directed hypergraphs, where each hyperedge is a pair of tail-nodes and head-nodes, are used to study brain networks, where more than one neuron might act [GLPN93, PET+14, LKC+12]. One might also consider a hypergraph with hyperedges corresponding to actors who played in

a movie together [RDPS04].

As for networks, hypergrpahs in the real world often present a power-law degree distribution [JJK19].

Applications of hypergraphs include higher-order link prediction, where one considers data where higher-order structures evolve over time. For instance, this is used to predict which sets of authors will write a paper together [BAS+18]. Moreover, higher-order link prediction could be used to find missing data in static hypernetworks [BAS+18].

Many interesting questions arise for hypergraphs. They include, among others, hypergraph visualization and hypergraph colouring. Models for networks such as the Erdös-Rényi model or the configuration model can be generalized to hypergraphs [CB16, YPVP17]. We use such generalizations in our simulations.

Dynamics on hypergraphs offer a more accurate description of contagion processes on complex systems [LRS19]. Hypergraphs models are fundamentally different from graph models, as they don't consider only dyadic relations. The use of hypergraphs can be used for the study of social contagion phenomena: peer pressure and the size of the social group (the hyperedge) might be important. An interesting example is the study of peer pressure on Facebook networks [KSS+16]. Other examples are the study of the adoption of innovation, the acceptance of norms, the diffusion of knowledge and opinion formation, which follow complex nonlinear patterns. When including complex dynamics of peer influence and reinforcement mechanisms, one might assume that the simplicity of pairwise interaction fails to capture the complexity of the situation. It is important to note that we want to model more than just the superposition pairwise interactions (which we would model as a multigraph). For example, in the setting of adoption of innovation, a person could adopt an innovative product because a single friend persuaded them to adopt it or because a group of friends simultaneously persuaded them to adopt the innovative product.

In public health, the use of hypergaph is not as obvious. In epidemic spreading, transmission usually occurs from one infectious person to a healthy one, through a dyadic interaction. It is possible to think of situations where hypergraphs might be useful, though. To capture the spread of a disease in a cafe (with a fixed square footage), we could model a population by adding a hyperedge between guests that sat at the same table. Because the size of the table is fixed, the higher the order of a hyperedge (ie the more people there were at the table), the closer the guests were to each other. In this case, modelling the interactions between the healthy individuals and the infectious ones using pairwise interactions fails to capture the non-linearity induced by the different distance between people. In other words, in this case, the interaction that a person can spend in at a table accompanied by a group of infectious individuals cannot be modelled as a superposition of interaction the person could spend in

the same setting with just one infectious individual at a time.

Simplicial complexes are a particular type of hypergraphs where if a hyperedge σ is present then all the subsimplices $v \subset \sigma$ are also contained in E. By convention, an ω-simplex contains $\omega + 1$ nodes. We could model group interactions between four individuals without taking into account the lower-order interactions with a hypergraph, but not with a simplicial complex.

Depending on the situation, we can assume that the presence of higher-order interactions implies the presence of the lower-order interactions [IPBL19]. Simplicial complexes have proven useful to study high dimensional data sets with the advances of topological data analysis (TDA) [SCL18], for instance using persistent homologies [OPT$^+$17, PET$^+$14, PPV17, LKC$^+$12]. But they might not be appropriate for all situations. For instance, it might well be that three authors have written a paper together [XOL16, PPV17, TCR10], while no paper was produced by a subset of two of the authors.

The remainder of this article is outlined as follows. The main goal is to study the abrupt phase transition of epidemic spreading in simplicial complexes [BKS16, MAG18, MGA20, IPBL19, dAPM19, JJK19], as studied mainly by JT Matamalas. We first discuss the generalisations of the discrete-time MMCA equations. They represent a Reactive Process (RP), meaning that all neighbours are contacted at each time step [GABH$^+$10]. We then discuss generalisations of the epidemic link equations for 2-simplicial complexes, recently proposed [MAG18, MGA20]. These concepts seem to only have been studied for 2-simplicial complexes in the literature. We thus generalise most of the discrete-time analysis of [MGA20], including the discrete-time MMCA equations and the ELE, to higher-order simplicial complexes - and (trivially) to hypergraphs. We illustrate the phase transition on synthetic networks - we perform both a replication of experiments by I Iacopini, and we extend the simulations to higher-order simplicial complexes [IPBL19]. We also discuss the epidemic importance of a link [MAG18], and possible generalisations. Finally, we discuss future research directions.

3 Results

Polyadic interactions can reveal higher-order dynamical effects that are not captured by traditional dyadic network models [NML19]. While multi-body dynamical effects can only appear if the interaction function is non-linear (linear dynamics on hypergraphs can always be rewritten as a dynamics on a standard, two-body network [NML19]), generalisations of SIS models to hypergraphs are interesting because both the theoretical analysis and numerical simulations show the emergence of a discontinuous transition induced by higher-order interactions [MGA20, IPBL19]. This is verified even for low-order simplicial complexes: the presence of 2-simplices can change the nature of the phase transition of

the epidemic spreading of the endemic state from continuous to discontinuous. Replicating exeperiments by I Iacopini, We perform simulations on synthetic hypernetworks, but the discontinuous transition is also observed during simulations on empirical simplicial complexes [IPBL19]. Higher-order interactions also lead to the appearance of a bi-stable region where both susceptible and infectious asymptotic states co-exist. The asymptotic state depends on the the initial proportion of infectious nodes - a critical mass is needed to reach the endemic state [IPBL19]. The theoretical analysis showing the discontinuous transition is realised (for 2-simplicial complexes only) via the microscopic Markov chain approach (MMCA) in [MGA20]. We show that the analysis carries to higher-orders, which doesn't seem to appear anywhere in the literature. The discontinuous transition is also analysed through a mean field approach [IPBL19].

These transitions contrast with classical contagion models on complex network, which instead display continuous transitions. Typical examples are the SIS or SIR models [dAPM20]. Understanding the origin of phase transition is an important issue for social contagion or epidemic spreading [JJK19], suggesting that the use of hypergraphs models is appropriate.

We also present the epidemic link equations (ELE) [MAG18], introduced for simplicial complexes consisting of 1 and 2-simplices. We try to generalise the ELE to situations incorporating higher-order simplicial complexes and hypergraphs. Furthermore, we discuss ways to extend the epidemic importance of a link [MAG18] to give a measure of the importance of hyperlinks. Both of these questions do not seem to appear in the literature.

Higher-order dependencies have been shown to either speed up or slow down dynamical processes, change node rankings [SWG16, Ben19, REL$^+$14, ERV06], and alter community structures [BGL16].

3.1 Microscopic Markov Chain Approach (MMCA)

The generalisation of the MMCA approach [GABH$^+$10, GGGMA11, GGA13] to 2-simplicial complexes [MGA20] presents a probabilistic formalisation of the higher-order interactions of an epidemic process represented by the SIS model. However, the authors restrict their attention to simplicial complexes with only one and two simplices present (edges and full triangles). We shall try to extend the model to general simplicial complexes and to hypergraphs.

In the classical SIS model on a connected undirected graph, the vertices represent individuals who are in one of two states: infectious (I) or susceptible (S) [Luc13]. To each vertices, we assign a binary value from the set $\{0, 1\}$ - 0 model susceptible individuals, and 1 models infectious individuals. We let $s_i(t) \in \{0, 1\}$ represent the state of individual i at time t. The infection propagates, through pairwise interactions, from infectious individual to susceptible

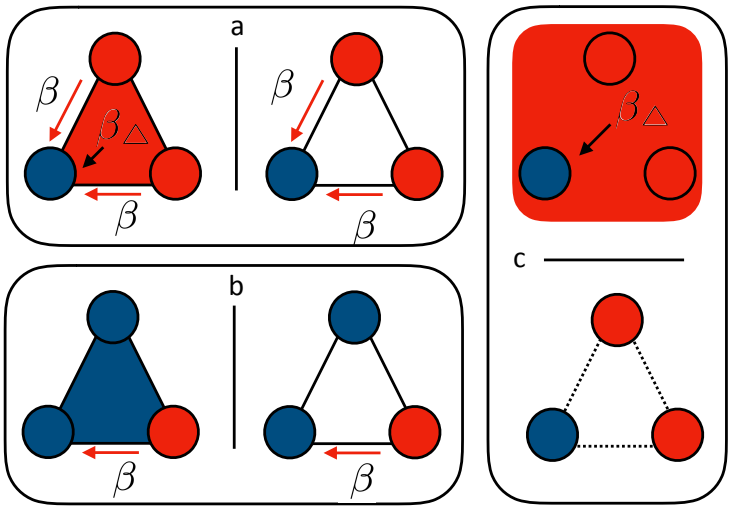

Figure 1: Possible infection channels via 2-simplices. For simplicial complexes (a,b), the presence of higher-order simplices (left) implies the presence of lower-order ones (right). This is not the case for hypergraphs (c).

individuals with probability β_{S_1} (we use S_1 to indicate that we consider 1-simplicial complexes), and infectious nodes recover with probability μ.

For 2-simplicial complexes, we can consider as well triangular interactions as in [MGA20]. We extend the model to higher-order (D) simplicial complexes (and to hypergraphs): nodes also interacts within the ω-simplices ($\omega \leq D$) with the ω neighbours at unison, with an infection probability β_{S_ω}.

To extend the model to hypergraphs, we drop the requirement that subsets of hyperedges are hyperedges as well. This means that we could have an infection through a "full triangle" but not through edges of this triangle.

Generalising the analysis of [MGA20], where the authors consider 2-simplicial complexes, we can then define a system of N discrete-time MMCA equations to describe the evolution over time of the probability p_i of node i being in an infected state at time t as:

$$p_i(t+1) = (1 - p_i(t))(1 - \Pi_{\omega=1}^{D} q_i^{S_\omega}(t)) + p_i(t)(1-\mu) \tag{1}$$

$q_i^{S_\omega}(t)$ denotes the probability that node i is not infected by any ω simplex it participates in. If node i is infectious it will recover with probability μ. Otherwise the node is susceptible and could get infected through interaction in ω-simplices it takes part in ($\omega \leq D$). The probability that the node will become infectious is $(1 - q_i^{S_1}(t) \dots q_i^{S_D}(t))$. Indeed, the node is not infected if it's not

infected through any ω-simplex. The model also directly extends to general hypergraphs.

The probability that node i is not infected by pairwise relations is:

$$q_i^{S_1}(t) = \Pi_{j \in S_1}(1 - \beta_{S_1} p_j(t)) \tag{2}$$

where S_1 is the set of neighbours (via links) of node i. Indeed, the node is not infected at the 2-simplex level if and only if it is not infected by any of its neighbours at this level.

Similarly, the probability that node i is not infected by any ω-simplex interaction is:

$$q_i^{S_\omega}(t) = \Pi_{n_1,\ldots,n_\omega \in S_\omega}(1 - \beta_{S_\omega} p_{n_1}(t) \ldots p_{n_\omega}(t)) \tag{3}$$

where S_ω is the set of ω-simplices (from which we exclude i) containing node i - we slightly abuse notation and we mean that the product ranges over such simplices. We note that contagion only happens when all ω neighbours are infectious. We can drop the requirement that subsets of hyperedges are hyperedges as well and consider general hypergraphs. We thus have generalised the MMCA approach for 2-simplicial complexes [MGA20] to general D-simplicial complexes and to hypergraphs.

When considering 2-simplicial complexes, the authors of [MGA20] show that the system of equations obtained is a contraction map:

$$T_{\beta_{S_1},\beta_{S_2},\mu} : \vec{p}(t) \to \vec{p}(t+1)$$

and the existence of fixed points is guaranteed by the Banach fixed point theorem [Luc13].

A naive approximation in [MGA20] assumes that all nodes have the same degree (k_{S_1}) and belong to the same number of 2-simplices (k_{S_2}), with the same probability of getting infected.

In [MGA20], expanding up to second order in p, and developing the equation at the stationary state, the authors arrive at:

$$0 \approx (1-p)[k_{s_1}\beta_{S_1} p + (k_{S_2}\beta_{S_2} - \beta_{S_1}^2 k_{S_1}(k_{S_1}-1)/2)p^2] - \mu p \tag{4}$$

To arrive at this equation, we assume that all nodes have the same probability of being infected $p_i(t) = p(t)$, and we expand up to second order in p at the

stationary state $p(t) = p$, as explained in [MGA20].

$$
\begin{aligned}
p =&(1-p)(1-q^{S_1}q^{S_2}) + p(1-\mu)\\
0 =&(1-p)(1-q^{S_1}q^{S_2}(t)) - p\mu\\
0 =&(1-p)(1-(1-\beta_{S_1}p)^{k_{S_1}}(1-\beta_{S_2}p^2)^{k_{S_2}}) - p\mu\\
0 \approx&(1-p)(1-(1-k_{S_1}\beta_{S_1}p + \binom{k_{S_1}}{2}(\beta_{S_1}p)^2)(1-k_{S_2}\beta_{S_2}p^2)) - p\mu\\
0 \approx&(1-p)[k_{S_1}\beta_{S_1}p + (k_{S_2}\beta_{S_2} - \beta_{S_1}^2\binom{k_{S_1}}{2})p^2] - p\mu
\end{aligned}
\tag{5}
$$

This third-order algebraic equation has a trivial solution, $p = 0$, and two more roots, that, depending on the parameters, provide up to two additional physical solutions [MGA20]. For the parameters with three physical solutions, the stability analysis shows that the middle one is unstable, thus being responsible for the abrupt transition [MGA20].

The good news is that this is easily verified in our D-simplicial complexes model (or hypergraph model) as we can recover the same approximation for the steady state equation. For hypergraphs, we also assume that all nodes belong to the same number of ω-simplices k_{S_ω}, and generally have the same probability of getting infected $p_i(t) = p(t)$. We expand up to second order in p at the stationary state $p(t) = p$ and we get exactly the same approximation (see Supplementary information). Thus we see analytically the same behaviour as in the 2-simplicial complexes case.

$$
\begin{aligned}
p =& (1-p)(1-\Pi_{\omega=1}^{D} q^{S_\omega}) + p(1-\mu)\\
0 =& (1-p)(1-\Pi_{\omega=1}^{D} q^{S_\omega}) - \mu p\\
0 =& (1-p)(1-\Pi_{\omega=1}^{D} (1-\beta_{S_\omega}p^{\omega})^{k_{S_\omega}}) - \mu p\\
0 \approx& (1-p)(1-(1-k_{S_1}\beta_{S_1}p + \binom{k_{S_1}}{2}(\beta_{S_1}p)^2)(1-k_{S_2}\beta_{S_2}p^2)) - \mu p\\
0 \approx& (1-p)(k_{S_1}\beta_{S_1}p + k_{S_2}\beta_{S_2}p^2 - \binom{k_{S_1}}{2}(\beta_{S_1}p)^2) - \mu p\\
0 =& (1-p)[k_{S_1}\beta_{S_1}p + (k_{S_2}\beta_{S_2} - \beta_{S_1}^2\binom{k_{S_1}}{2})p^2] - \mu p
\end{aligned}
\tag{6}
$$

We could have expanded to third order and we would have gotten more solutions. We would thus get a trivial solution $p = 0$ and possibly more physical solutions. This might create an hysteresis loop that is more interesting than the ones described in [MGA20, IPBL19]. We will explore this with numerical simulations, as well as in the Mean-field section.

3.2 Epidemic Link Equations (ELE)

Throughout this section, for clarity and tractability, we will first explain the notions introduced in [MGA20] using only the case of 2-simplicial complexes, as is done in the article, before proposing generalisations.

As observed in [MGA20], 1 and 4 use the implicit assumption that the probability of being infected by one neighbour is independent on the probability of being infected by any other neighbour. This approximation is a mean field approximation, whose validity is undermined by the fact that we consider higher-order simplices, and hence neighbours are unlikely to be independent [MGA20]. To circumvent this problem, the authors of [MGA20] define the simplicial model at the level of links and introduce the Epidemic Link Equations (ELE) (a system of $3L$ equations, where L is the number of links, as we need to take into account, for each edge between i and j, four possible states; the states $P_{ij}^{SS}(t)$, $P_{ij}^{SI}(t)$, P_{ij}^{IS}, and we can use $P_{ij}^{II}(t) = 1 - P_{ij}^{SS}(t) - P_{ij}^{SI}(t) - P_{ij}^{IS}$ [MAG18], where P_{ij}^{SS} is the joint probability that node i is susceptible and j is susceptible, P_{ij}^{SI} is the joint probability that node i is susceptible and j is infectious, etc...). For a node i, $P_i^S = P_{ij}^{SS} + P_{ij}^{SI}$ and $P_i^I = P_{ij}^{II} + P_{ij}^{IS}$. The size of this system is larger than the MMCA system of N equations at the level of nodes but the authors exploit the fact that the node marginal probabilities have to be the same for every link we consider to reduce the model to $L + N$ equations [MGA20]. Working in the 2-simplicial case, the epidemic link equations, ELE [MGA20], for every node i, are then:

$$P_i^I(t+1) = (1 - P_i^I(t))(1 - q_i^{S_1}(t)q_i^{S_2}(t)) + P_i^I(t)(1-\mu) \tag{7}$$

where $q_i^{S_1}(t)$ defines the probability that node i is not infected by any pairwise interaction with its neighbours,

$$q_i^{S_1}(t) = \Pi_{j \in S_1} \left(1 - \beta_{S_1} \frac{P_{ij}^{SI}(t)}{P_i^S(t)}\right) \tag{8}$$

and $q_i^{S_2}(t)$ is the probability that node i is not infected by any of the interactions at the 2-simplicial level,

$$q_i^{S_2} = \Pi_{j,r \in S_2} \left(1 - \beta_{S_2} \frac{P_{ijr}^{SII}(t)}{P_i^S(t)}\right) \tag{9}$$

Bayes theorem gives us that the probability that node j is infectious given that node i is susceptible is $P_{ji}^{I|S}(t) = P_{ij}^{SI}(t)/P_i^S(t)$ and $P_{jri}^{II|S} = P_{ijr}^{SII}/P_i^S$ [MGA20]. The system still requires L equations, one for every link, that account for the probability of having a link connecting two nodes in the infectious states II, transitioning from the four possible states SS, SI, IS, II [MGA20]. The probability of node i not being infected by any neighbour different from j through a

link; the probability of node i not being infected by any 2-simplex not containing node j; and the probability of node i not being infected by any 2-simplex containing node j are respectively [MGA20]:

$$q_{ij}^{S_1}(t) = \Pi_{r\in S_1, r\neq j}\left(1-\beta_{S_1}\frac{P_{ir}^{SI}(t)}{P_i^S(t)}\right) \tag{10}$$

$$q_{ij}^{S_2}(t) = \Pi_{r,l\in S_2, r\neq j, l\neq j}\left(1-\beta_{S_2}\frac{P_{irl}^{SII}(t)}{P_i^S(t)}\right) \tag{11}$$

$$u_{ij}^{S_2}(t) = \Pi_{r\in S_2'}\left(1-\beta_{S_2}\frac{P_{ijr}^{SII}(t)}{P_{ij}^{SI}(t)}\right) \tag{12}$$

where S_2' are the 2-simplices (from which we exclude i and j) containing both i and j. If both node i and j are already infectious at time t, they will both stay infectious at time $t+1$ with probability $(1-\mu)^2$. If they are both susceptible at time t, they will be infectious at time $t+1$ if i gets infected (which happens with probability $1-q_{ij}^{S_1}(t)q_{ij}^{S_2}(t)$ as the infection cannot be via the link ij or a 2-simplex containing node j) and j gets infected (which happens with probability $1-q_{ji}^{S_1}(t)q_{ji}^{S_2}(t)$ as the infection cannot come through link ji or a 2-simplex containing node i). If at time t node i is susceptible and node j is infectious, then both nodes will be infectious at time $t+1$ if node j remains infectious (that occurs with probability $(1-\mu)$) and node i gets infected. We will show that the probability of this happening is $(1-(1-\beta_{S_1})q_{ij}^{S_1}(t)u_{ij}^{S_2}(t)q_{ij}^{S_2}(t))$ by showing that the probability that node i stays susceptible is $(1-\beta_{S_1})q_{ij}^{S_1}(t)u_{ij}^{S_2}(t)q_{ij}^{S_2}(t)$. The only way node i doesn't become infectious is that it is not infected by node j through the link ij (this has probability $1-\beta_{S_1}$); it is not infected by any neighbour different from j through a link (this has probability $q_{ij}^{S_1}$); it is not infected by any 2-simplex not containing node j (this has probability $q_{ij}^{S_2}(t)$); it is not infected by any 2-simplex containing node j (this has probability $u_{ij}^{S_2}(t)$). The case when i is infectious and j is susceptible is similar. Thus we can express the probability that at time $t+1$ both node i and j are infectious as the sum of four terms:

$$P_{ij}^{SS}(t)(1-q_{ij}^{S_1}(t)q_{ij}^{S_2}(t))(1-q_{ji}^{S_1}(t)q_{ji}^{S_2}(t)) \tag{13}$$

$$P_{ij}^{SI}(t)(1-(1-\beta_{S_1})q_{ij}^{S_1}(t)u_{ij}^{S_2}(t)q_{ij}^{S_2}(t))(1-\mu) \tag{14}$$

$$P_{ij}^{IS}(t)(1-(1-\beta_{S_1})q_{ji}^{S_1}(t)u_{ji}^{S_2}(t)q_{ji}^{S_2}(t))(1-\mu) \tag{15}$$

$$P_{ij}^{II}(t)(1-\mu)^2 \tag{16}$$

In this model, we still need a closure for the triadic joint probabilities like P_{ijr}^{SII} found in 9, 11 and 12 . To break the hierarchy of clusters produced

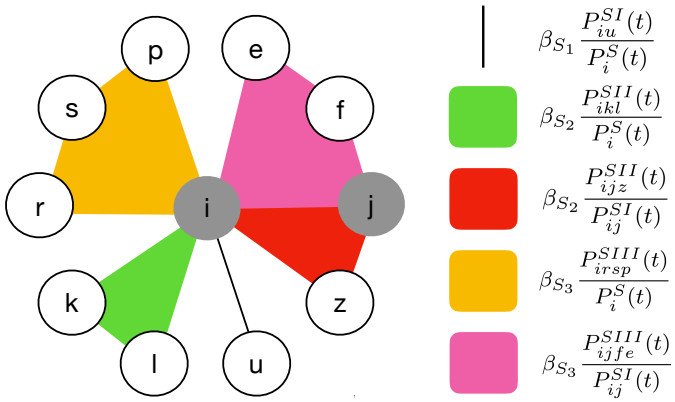

Figure 2: Possible contributions to the joint probability between the states of node i and j, via different simplices. We assume that i is susceptible and j is infectious.

by these terms, we must rely on approximations [MGA20]. As explained by the authors of [MGA20], the classical pair approximation in statistical physics modified for the context of epidemics [MF13] would consist in approximating $P_{ijr}^{SII} \approx P_{ij}^{SI} P_{jr}^{II} / P_j^I$. The problem with these proposals is that the symmetry of the elements in the 2-simplices is broken - "this concurs in a degeneration of solutions" [MGA20]. We could choose between three different closures for P_{ijr}^{SII}: $P_{ijr}^{SII} \approx P_{ij}^{SI} P_{jr}^{II} / P_j^I$, $P_{ijr}^{SII} \approx P_{ir}^{SI} P_{rj}^{II} / P_r^I$, $P_{ijr}^{SII} \approx P_{ji}^{IS} P_{ir}^{SI} / P_i^S$. Each alternative makes use of one node and two links probabilities, but it is not clear which one should be used. The authors of [MGA20] thus propose that the closure should be symmetric and hence uses the approximation [MGA20]:

$$P_{ijr}^{SII} \approx \frac{P_{ij}^{SI} P_{ir}^{SI} P_{jr}^{II}}{P_i^S P_j^I P_r^I} \tag{17}$$

In this way, all the node and link probabilities of the 2-simplex structures are used, avoiding asymmetries of the classical pair approximation [MGA20].

We generalise these results to the D-simplicial case (and its trivial extension to hypergraphs) to what could be called the epidemic hyperlink equations. This does not seem to have been done anywhere.

$$P_i^I(t+1) = (1 - P_i^I(t))(1 - \Pi_{\omega=1}^{D} q_i^{S_\omega}(t)) + P_i^I(t)(1-\mu) \tag{18}$$

The probability that i is not infected through any ω-simplices, $q_i^{S_\omega}(t)$, is:

$$q_i^{S_\omega} = \Pi_{n_1,\ldots,n_\omega \in S_\omega} \left(1 - \beta_{S_\omega} \frac{P_{in_1\ldots n_\omega}^{SI\ldots I}(t)}{P_i^S(t)}\right) \tag{19}$$

Indeed, $P^{I\dots I|S}_{n_1\dots n_\omega i} = P^{SI\dots I}_{in_1\dots n_\omega}/P^S_i$ by Bayes. We express $P^{II}_{ij}(t+1)$ as the sum of the following four terms:

$$P^{SS}_{ij}(t)(1-\Pi^D_{\omega=1}q^{S_\omega}_{ij}(t))(1-\Pi^D_{\omega=1}q^{S_\omega}_{ji}(t)) \tag{20}$$

$$P^{SI}_{ij}(t)(1-(1-\beta_{S_1})q^{S_1}_{ij}(t)\Pi^D_{\omega=2}[u^{S_\omega}_{ij}(t)q^{S_\omega}_{ij}(t)])(1-\mu) \tag{21}$$

$$P^{IS}_{ij}(t)(1-(1-\beta_{S_1})q^{S_1}_{ji}(t)\Pi^D_{\omega=2}[u^{S_\omega}_{ji}(t)q^{\omega}_{ji}(t)])(1-\mu) \tag{22}$$

$$P^{II}_{ij}(t)(1-\mu)^2 \tag{23}$$

where $q^{S_\omega}_{ij}$ is the probability of node i not being infected through a ω-simplex not containing node j, and $u^{S_\omega}_{ij}$ is the probability of node i not being infected by any ω simplex containing node j (as in 2). Thus:

$$q^{S_\omega}_{ij}(t) = \Pi_{\substack{n_1,\dots,n_\omega\in S_\omega \\ n_1,\dots,n_\omega\neq j}}\left(1-\beta_{S_\omega}\frac{P^{SI\dots I}_{in_1\dots n_\omega}(t)}{P^S_i(t)}\right) \tag{24}$$

$$u^{S\omega}_{ij}(t) = \Pi_{n_1,\dots,n_{\omega-1}\in S'_\omega}\left(1-\beta_{S_\omega}\frac{P^{SII\dots I}_{ijn_1\dots n_{\omega-1}}(t)}{P^{SI}_{ij}(t)}\right) \tag{25}$$

where S'_ω are the ω-simplices (from which we exclude i and j) containing both i and j.

The system then requires either approximations of $P^{SI\dots I}_{in_1\dots n_\omega}$ for $2\leq\omega\leq D$, or to totally generalise the ELE by adding additional equations for every ω-simplices, for $\omega\leq D-1$. These additional equations give the probability of an ω-simplex transitioning from $2^{\omega+1}$ possible states (eg: for $\omega=2$ the states are SSS, SSI, SIS, SII, ISS, ISI, IIS, III). The maximal number of equations needed would be $\sum_{\omega=1}^{D-1}L_\omega(2^{\omega+1}-1)$, where L_ω is the total number of ω-simplices, but there might be further reductions as in the case considered in [MGA20]. The maximal number of equations needed would be $\sum_{\omega=1}^{D-1}L_\omega(2^{\omega+1}-1)$ because for each ω-simplex (of which there are L_ω), we would need to take into account the $2^{\omega+1}$ states, spare one using the fact that the probability that we are in one of these states is 1. Even with further reductions, we still have to estimate terms of the form $P^{SI\dots I}_{in_1\dots n_D}$ where, we should follow a logic similar to [MGA20] and avoid asymmetries.

If we wanted to do the full generalization, as mentioned, additional equations give the probability of an ω-simplex transitioning from $2^{\omega+1}$ possible states. For 2-simplices, there are 8 infected states SSS, SSI, SIS, SII, ISS, ISI, IIS, III.

$$
\begin{aligned}
P_{ijr}^{III}(t+1) =& P_{ijr}^{SSS}(t)(1-\Pi_{\omega=1}^{D} q_i^{S_\omega})(1-\Pi_{\omega=1}^{D} q_j^{S_\omega})(1-\Pi_{\omega=1}^{D} q_r^{S_\omega}) \\
&+ P_{ijr}^{SSI}(t)(1-\mu)(1-\Pi_{\omega=1}^{D} q_i^{S_\omega})(1-\Pi_{\omega=1}^{D} q_j^{S_\omega}) \\
&+ P_{ijr}^{SIS}(t)(1-\mu)(1-\Pi_{\omega=1}^{D} q_i^{S_\omega})(1-\Pi_{\omega=1}^{D} q_r^{S_\omega}) \\
&+ P_{ijr}^{SII}(t)(1-\mu)^2(1-\Pi_{\omega=1}^{D} q_i^{S_\omega}) \\
&+ P_{ijr}^{ISS}(t)(1-\mu)(1-\Pi_{\omega=1}^{D} q_j^{S_\omega})(1-\Pi_{\omega=1}^{D} q_r^{S_\omega}) \\
&+ P_{ijr}^{ISI}(t)(1-\mu)^2(1-\Pi_{\omega=1}^{D} q_j^{S_\omega}) \\
&+ P_{ijr}^{IIS}(t)(1-\mu)^2(1-\Pi_{\omega=1}^{D} q_r^{S_\omega}) \\
&+ P_{ijr}^{III}(t)(1-\mu)^3
\end{aligned} \tag{26}
$$

where $q_r^{S_\omega}$ is the probability that node r is not infected through a ω-interactions, which we would have to decompose as in [MAG18] between cases where the hyperedge contains i and j, just one of i or j, or none of them.

3.3 Mean-field approach

Under the name "Simplicial Contagion Model (SCM)" [IPBL19], the authors study the same model as the one we have been discussing via the discrete-time MMCA using a mean field approach. It also correctly predicts the steady-state dynamics, the position and nature of the transition and the location of the bistable region.

We summarize the mean field approach of [JJK19, IPBL19]. We can write a MF expression for the temporal evolution of the density of infectious nodes $\rho(t)$ as [IPBL19]:

$$
d_t\rho(t) = -\mu\rho(t) + \sum_{\omega=1}^{D} \beta_{S_\omega} \langle k_{S_\omega} \rangle \rho^{\omega}(t)[1-\rho(t)] \tag{27}
$$

where, for each $\omega = 1, \ldots, D$, $\langle k_{S_\omega} \rangle$ is the average number of ω-simplices incident to a node. For $D = 1$, we recover the MF equation for the SIS model [IPBL19]. Note the similarities with 4! To derive 27, we note that infectious nodes become susceptible with rate per unit time μ, and the susceptible nodes (their density is $(1-\rho(t))$) can be infected by any simplices, ranging from links to D-simplices. Contagion happens through a ω-simplex, by definition with rate β_{S_ω}, if and only if all other nodes are infectious, hence the term $\rho^{\omega}(t)$.

We can expand the case $D = 2$ [IPBL19] and solve $d_t\rho(t) = 0$ - the steady state equation. This has three acceptable solutions in the range $[0, 1]$ [IPBL19]. The solution $\rho_1^* = 0$ corresponds to the usual absorbing epidemic-free state, in

which all the individuals recover and the spreading dies out 3. The two other solutions are:

$$\rho^*_{2\pm} = \frac{\lambda_{S_2} - \lambda_{S_1} \pm \sqrt{(\lambda_{S_2} - \lambda_{S_1})^2 - 4\lambda_{S_2}(1 - \lambda_{S_1})}}{2\lambda_{S_2}} \tag{28}$$

where $\lambda_{S_1} := \beta_{S_1}\langle k_{S_1}\rangle/\mu$ and $\lambda_{S_2} := \beta_{S_2}\langle k_{S_2}\rangle/\mu$ [IPBL19]. We show how to obtain these solutions now:

$$\begin{aligned}
d_t\rho(t) =& -\mu\rho(t) + \sum_{\omega=1}^{2} \beta_{S_\omega}\langle k_{S_\omega}\rangle\rho^\omega(t)[1-\rho(t)] \\
=& -\mu\rho(t) + \beta_{S_1}\langle k_{S_1}\rangle\rho(t)[1-\rho(t)] + \beta_{S_2}\langle k_{S_2}\rangle\rho^2(t)[1-\rho(t)] \\
=& -\mu\rho(t) + \beta_{S_1}\langle k_{S_1}\rangle\rho(t) - \beta_{S_1}\langle k_{S_1}\rangle\rho^2(t) + \beta_{S_2}\langle k_{S_2}\rangle\rho^2(t) \\
& - \beta_{S_2}\langle k_{S_2}\rangle\rho^3(t) \\
=& -\mu\rho(t)[1 - \lambda_{S_1} + \lambda_{S_1}\rho(t) - \lambda_{S_2}\rho(t) + \lambda_{S_2}\rho^2(t)]
\end{aligned}$$

Thus $p^*_{2\pm}$ are the solutions of $1 - \lambda_{S_1}(1-\rho) - \lambda_{S_2}(\rho - \rho^2)$ which we obtain by the quadratic formula.

When $\lambda_{S_2} \leq 1$, the transition is continuous, reflecting the transition of the standard SIS model [IPBL19].

When $\lambda_{S_2} > 1$, the transition is discontinuous at $\lambda^c = 2\sqrt{\lambda_{S_2}} - \lambda_{S_2}$ [IPBL19]. A careful analysis shows that for $\lambda^c < \lambda_{S_1} < 1$ the final state depends on the initial density of infectious nodes [IPBL19], as illustrated in 3. For $\lambda^c < \lambda_{S_1}$, we have two possibilities to consider. If $\lambda_{S_1} > 1$, we can show that $\rho^*_{2-} < 0 < \rho^*_{2-}$. 27 shows then that, for small $\rho(t)$, $d_t\rho(t) > 0$: then, the absorbing state $\rho^*_1 = 0$ is unstable and the density of infectious nodes tends to ρ^*_{2+}. If $\lambda_{S_1} < 1$, then $0 < p^*_{2-} < p^*_{2+}$. Then p^*_1 and p^*_{2+} are absorbing states. The density in the large time limit is determined by the initial density of incetious nodes. If it is below the unstable steady-state value ρ^*_{2-}, the absorbing state ρ^*_1 is reached:

$$\rho(t) \longrightarrow_{t\to\infty} \rho^*_1 \tag{29}$$

If the initial density of infectious nodes is above ρ^*_{2-}, the system tends to a finite density of infectious nodes equal to ρ^*_{2+} [IPBL19]:

$$\rho(t) \longrightarrow_{t\to\infty} \rho^*_{2+} \tag{30}$$

This shows that a critical mass is needed to reach the endemic state [IPBL19].

We now try to generalize this analysis for 3-simplicial complexes. In this case, we have:

$$d_t\rho(t) = -\mu\rho(t) + \sum_{\omega=1}^{3} \beta_{S_\omega}\langle k_{S_\omega}\rangle\rho^\omega(t)[1-\rho(t)] \tag{31}$$

The steady state equation is:

$$0 = -\mu\rho[1 - \lambda_{S_1} + \lambda_{S_1}\rho - \lambda_{S_2}\rho + \lambda_{S_2}\rho^2 - \lambda_{S_3}\rho^2 + \lambda_{S_3}\rho^3]$$

where $\lambda_{S_i} := \beta_{S_i}\langle k_{S_i}\rangle/\mu$. The solution $\rho_1^* = 0$ corresponds to the usual absorbing epidemic-free state, in which all the individuals recover and the spreading dies. The three other solutions are given by the cubic formula. Let:

$$\Delta_0 = (\lambda_{S_2} - \lambda_{S_3})^2 - 3\lambda_{S_3}(\lambda_{S_1} - \lambda_{S_2})$$
$$\Delta_1 = 2(\lambda_{S_2} - \lambda_{S_3})^3 - 9\lambda_{S_3}(\lambda_{S_2} - \lambda_{S_3})(\lambda_{S_1} - \lambda_{S_2}) + 27\lambda_{S_3}^2(1 - \lambda_{S_1})$$
$$C = \sqrt[3]{\frac{\Delta_1 \pm \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}}$$

where we take C to be one of the three possible cubic roots. The solutions are then given by:

$$\rho_{2,3,4}^* = -\frac{1}{3\lambda_{S_3}}(\lambda_{S_2} - \lambda_{S_3} + C + \frac{\Delta_0}{C}) \tag{32}$$

It remains to do the bifurcation analysis for this situation. From our numerical simulations, it seemed that there are only two absorbing states at most even in this case. Note that for simplicial-complexes of order $D \geq 2$, the steady states equations are:

$$0 = -\mu\rho[1 - \lambda_{S_1} + \lambda_{S_1}\rho + \sum_{i=2}^{D}(-\lambda_{S_i}\rho^{i-1} + \lambda_{S_i}\rho^i] \tag{33}$$

For 4-simplicial complexes, we could solve the equation using the quartic equation. For higher orders, we would have to use numerical methods.

3.4 Epidemic importance of a link

The ELE are originally proposed in [MAG18], for the case $D = 1$, to study the contribution of links to spreading processes. This is useful to study the containment of epidemics based on deactivating the most important links in transmitting the disease, rather that deactivating nodes. To determine which link should be removed first, one needs a measure of the importance of each link in the spreading of the epidemics. For networks, we could use the relative change of the spectral radius when the edge is removed [MAG18]. However, the spectral radius only depends on the structure of the network, not on the epidemic parameters. Moreover, it doesn't take into account the participation of the link in the spreading of the disease. In [MAG18], the authors introduce the epidemic importance of a link in networks, $I_{ij} := \bar{n}_{ij} + \bar{n}_{ji}$, where:

$$\bar{n}_{ij} := \beta_{S_1} P_{ji}^{SI} \sum_{r=1}^{N} A_{jr}\beta_{S_1} P_{rj}^{S|I} \tag{34}$$

A link in state SS or II is not going to be used by the epidemic in the next time step - there are either no infectious node to propagate the disease, or both nodes are already infectious [MAG18]. To propagate the epidemics, a link must either be in the SI or IS state. Suppose that a link (i, j) is in state IS. With probability β_{S_1}, node i can infect node j through this link. If this happens, the newly infected node j may infect its neighbours. Had we removed link (i, j), we would have annihilated this channel of infections starting at node i. The degree of j and the probability of its neighbours being susceptible when j becomes infectious are important: if the neighbours of j are infectious nodes, then removing link (i, j) is not going to have a big impact on the overall incidence of the epidemics [MAG18]. The expected number of infectious nodes produced in this way can be expressed by $\bar{n}_{ij}$ [MAG18]. This measure is asymmetric, we need to consider the paths of infections starting at i and using (i, j) as well as the paths of infections starting at j and using (j, i). The epidemic importance of a link is thus defined to take both sets of paths into account, as the sum of $\bar{n}_{ij}$ and $\bar{n}_{ji}$ [MAG18].

The epidemic link equations do not seem to have been generalised to the cases when $D > 1$ anywhere. Furthermore, there does not seem to be definitions of the epidemic importance of a hyperlink that would quantify the importance of hyperedges. We propose the following. When working with D-simplicial complexes, in the definition of the epidemic importance of a link, we still need to take into account both states IJ and JI, but directly generalising [MAG18], $\bar{n}_{ij}$ now equals:

$$\bar{n}_{ij} = \beta_{S_1} P_{ji}^{SI} \sum_{\omega=1}^{D} \sum_{l_1,\dots,l_\omega \in S_\omega} \sum_{k=1}^{\omega} \beta_{S_\omega} P^{I\dots IS|I}_{l_1\dots\hat{l}_k\dots l_\omega l_k j} \tag{35}$$

As above, at the link level, i can infect j, with probability β_{S_1} if i is infectious and j is susceptible. Then, an infectious j can go on to infect other nodes through ω-simplices (with probability β_{S_ω}), if all but one node in the ω-simplex in question is infectious. As always, S_ω represents the set of ω-simplices j takes part in (from which we exclude j).

3.5 Epidemic importance of a hyperlink

For hyperlinks -ie ω-simplices - between nodes $i_1, \dots, i_{\omega+1}$, the epidemic dynamic is going to use the simplex only if all but one node are infectious (there are therefore $\omega + 1$ configurations that the epidemic dynamics use). Thus we let $\bar{n}_{i_1,\dots,\hat{i}_j,\dots,i_{\omega+1};i_j} = \bar{n}_{i_j}$ represents the expected number of infected node when all nodes but i_j are infectious and i_j is susceptible. Then we propose the following extension, where $S_{\omega'}$ is the set of ω'-simplices containing i_j (i_j excluded):

$$\bar{m}_{i_j} := \sum_{\omega'=1}^{D} \sum_{l_1,\dots,l_{\omega'} \in S_{\omega'}} \sum_{k=1}^{\omega'} \beta_{S_{\omega'}} P^{I\dots IS|I}_{l_1\dots\hat{l}_k\dots l_{\omega'} l_k i_j} \tag{36}$$

$$\bar{n}_{i_j} = \beta_{S_\omega} P^{SI\dots I}_{i_j i_1 \dots \hat{i}_j \dots i_{\omega+1}} \bar{m}_{i_j} \tag{37}$$

$$I_{i_1,\dots,i_{\omega+1}} := \sum_{j=1}^{\omega+1} \bar{n}_{i_j} \tag{38}$$

4 Discussion

We have presented the mathematical formulation of the SIS in simplicial complexes, expanding the discrete-time probabilistic description of the process in the node approximation (MMCA) to higher-order simplicial complexes and hypergraphs. We proposed a generalisation of the discrete-time probabilistic description of the process in the link approximation (ELE). For 2-simplicial complexes, both descriptions predict an abrupt transition in the case of 2-simplicial complexes [MAG18]. We have shown that the same is true for the node approximation (MMCA) in the case of higher-order simplicial complexes and hypergraphs (by doing a second order expansion). Whether we can make such a prediction when considering the proposed extensions for the link approximation (ELE) remains to be investigated. This is a particularly important question, as, as the authors of [MGA20] note, the predictions for ELE are more accurate.

We now illustrate the existence of the abrupt transition through numerical simulations in synthetic 2-simplicial complexes. We first replicate experiments made by I Iacopini and discuss models to construct synthetic hypergraphs.

4.1 Synthetic hypergraphs

Different models of random simplicial complexes and hypergraphs have been proposed, such as the configuration model for hypergraphs [CB16, Cho19].

In [IPBL19], the authors construct random simplicial models of dimension D via their Random Simplicial Complexes (RSC) model. It has $D+1$ parameters, the number of vertices N and D probabilities $\{p_1, \dots, p_D\}$, with $p_\omega \in [0,1]$ to control the creation of ω-simplices up to dimension D [IPBL19]. It is a direct generalisation of the Erdös-Rényi model for random graphs. Let us explain in more detail the cases for 2-simplicial complexes and 3-simplicial complexes. In [IPBL19], the authors construct a 2-simplicial complex with a target for the average degree $\langle k_{S_1} \rangle$ and the average number of 2-simplices $\langle k_{S_2} \rangle$ any node participates in. Thus, the probabilities needed are:

$$p_1 = \frac{\langle k_{S_1} \rangle - 2\langle k_{S_2} \rangle}{(N-1) - 2\langle k_{S_2} \rangle} \tag{39}$$

$$p_2 = \frac{\langle k_{S_2} \rangle}{\binom{N-1}{2}} \tag{40}$$

This is because given any node v, there are $(N-1)$ remaining node, so $\binom{N-1}{2}$ possible 2-simplices including node v. As we want an average of $\langle k_{S_2} \rangle$ participations in 2-simplices, we use p_2. Now, as we work with simplicial complexes, subset of hyperedges are themselves hyperedges. Thus, the creation of a 2-simplex that includes v induces the creation of 2 1-simplices (edges) that include v (and one more 1-simplex, that does not include v). Thus, there are on average $(N-1) - 2\langle k_{S_2} \rangle$ nodes that v is not connected to, and we want to create $\langle k_{S_1} \rangle - 2\langle k_{S_2} \rangle$ edges. Thus, as we want an average of $\langle k_{S_1} \rangle$ participations in 1-simplices, we use p_1.

The probabilities are found similarly for 3-simplicial complexes. Indeed, given any node v, there are $(N-1)$ remaining nodes, meaning $\binom{N-1}{3}$ possible 3-simplices including node v. As we want an average of $\langle k_{S_3} \rangle$ participations in 3-simplices, we use p_3. The creation of a 3-simplex that includes v induces the creation of 3 2-simplices that include v. Thus, there are on average $\binom{N-1}{2} - 3\langle k_{S_2} \rangle$ possible pairs that v is not connected to via 2-simplices. We want to create $\langle k_{S_2} \rangle - 3\langle k_{S_3} \rangle$ 2-simplices that include v. Thus, as we want an average of $\langle k_{S_2} \rangle$ participations in 2-simplices, we use p_2. The same reasoning applied to 2-simplices is done to derive p_1.

$$p_1 = \frac{\langle k_{S_1} \rangle - 2\langle k_{S_2} \rangle}{(N-1) - 2\langle k_{S_2} \rangle} \tag{41}$$

$$p_2 = \frac{\langle k_{S_2} \rangle - 3\langle k_{S_3} \rangle}{\binom{N-1}{2} - 3\langle k_{S_3} \rangle} \tag{42}$$

$$p_3 = \frac{\langle k_{S_3} \rangle}{\binom{N-1}{3}} \tag{43}$$

The same reasoning is used for 4-simplicial complexes. Given any node v, there are $(N-1)$ remaining nodes, meaning $\binom{N-1}{4}$ possible 4-simplices including node v. We consequently derive p_4 knowing we want an average of $\langle k_{S_4} \rangle$ participations in 4-simplices. To obtain the expression of p_3, we observe that the creation of a 4-simplex that includes v induces the creation of 4 3-simplices that include node v. There are on average $\binom{N-1}{3} - 4\langle k_{S_4} \rangle$ possible triple that v is not connected to via 3-simplices; $\binom{N-1}{3}$ represents the total number of possible 3-simplices including node v. We want to create $\langle k_{S_3} \rangle - 4\langle k_{S_4} \rangle$ edges. Thus, we use p_3. The same reasoning applied to 3-simplices and 2-simplices is used to derive p_2 and p_1.

$$p_1 = \frac{\langle k_{S_1} \rangle - 2\langle k_{S_2} \rangle}{(N-1) - 2\langle k_{S_2} \rangle} \tag{44}$$

$$p_2 = \frac{\langle k_{S_2} \rangle - 3\langle k_{S_3} \rangle}{\binom{N-1}{2} - 3\langle k_{S_3} \rangle} \tag{45}$$

$$p_3 = \frac{\langle k_{S_3} \rangle - 4\langle k_{S_4} \rangle}{\binom{N-1}{3} - 4\langle k_{S_4} \rangle} \tag{46}$$

$$p_4 = \frac{\langle k_{S_4} \rangle}{\binom{N-1}{4}} \tag{47}$$

When we want to create D-simplicial complexes, we would use:

$$p_i = \frac{\langle k_{S_i} \rangle - (i+1)\langle k_{S_2} \rangle}{\binom{N-1}{i} - (i+1)\langle k_{S_2} \rangle} \text{ for } 1 \leq i < D \tag{48}$$

$$p_D = \frac{\langle k_{S_D} \rangle}{\binom{N-1}{D}} \tag{49}$$

Note that we assume that the hypernetwork is big enough (ie that the number of nodes in the hypernetwork is large enough) so that the hyperdeges created intersect in v only.

4.2 Numerical simulations on 2-simplicial complexes

We follow the analysis of [IPBL19] closely to illustrate the abrupt phase transition and we restrict to 2-simplicial complexes. We start with an initial density ρ_0 of infectious nodes in a synthetic 2-simplicial complex of 211 nodes, with $\langle k_{S_1} \rangle \approx 23$ and $\langle k_{S_2} \rangle \approx 7$, and consider all possible channels of infection. We stop a simulation if an absorbing state is reached - all the nodes are S, or all the nodes are I, otherwise we compute the average stationary density of infectious node ρ^* by averaging the values measured in the last 100 times-step. The results are averaged over 100 runs obtained by randomly generating a simplicial complex and randomly placing initial infectious nodes with the same density ρ_0 [IPBL19]. We use the re-scaled parameters $\lambda_{S_1} = \beta_{S_1}\langle k_{s_1} \rangle/\mu$ and $\lambda_{S_2} = \beta_{S_2}\langle k_{S_2} \rangle/\mu$ as already defined. We plot the average fraction of infectious nodes ρ^* in the stationary state as a function of λ_{S_1} for different values of λ_{S_2}. The case where $\lambda_{S_2} = 0$ corresponds to the standard SIS model. We use two values for ρ_0: 1% and 40%. For $\lambda_{S_2} \leq 1$, no matter the initial density of infectious nodes, ρ_0, we reach the same absorbing state.

For $\lambda_{S_2} = 2$, we observe the appearance of an endemic state with $\rho^* > 0$ at a value of λ^c well below the epidemic threshold of the standard SIS model. Furthermore, this transition is discontinuous, and a hysteresis loop appears in a bi-stable region, where both healthy $\rho^* = 0$ and endemic $\rho^* > 0$ states can co-exist 3 [IPBL19]. In this parameter region, the final state depends on the initial density of infectious nodes ρ_0. This agrees with our previous analysis.

The dynamics of the SCM are very similar on real-world simplicial complexes which also shows an abrupt phase transition [IPBL19], even though real-world complexes usually present a fat-tailed distribution of degrees. It remains to see if we could collect real-data to model a population as a 3-simplicial complex and conduct simulations on the resulting hypergraph.

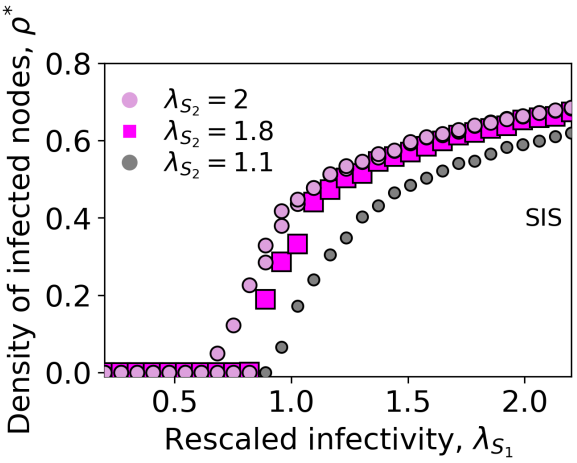

Figure 3: SCM on a RSC with $\langle k_{S_1} \rangle \approx 23$, $\langle k_{S_2} \rangle \approx 7$ and $\mu = 0.06$. We plot ρ^* against λ_{S_1} for different values of λ_{S_2}. We observe a discontinuous transition at a value approximately equal to $\lambda^c = 2\sqrt{\lambda_{S_2}} - \lambda_{S_2}$ when $\lambda_{S_2} = 2$.

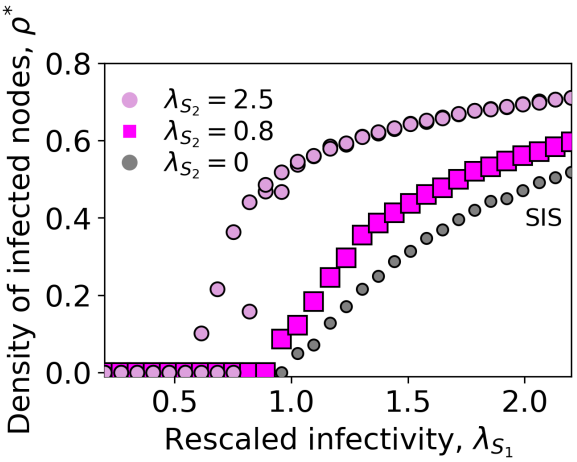

Figure 4: SCM on a RSC with $\langle k_{S_1} \rangle \approx 23$, $\langle k_{S_2} \rangle \approx 7$ and $\mu = 0.06$. We plot ρ^* against λ_{S_1} for different values of λ_{S_2}. We observe a discontinuous transition at a value approximately equal to $\lambda^c = 2\sqrt{\lambda_{S_2}} - \lambda_{S_2}$ when $\lambda_{S_2} = 2.5$.

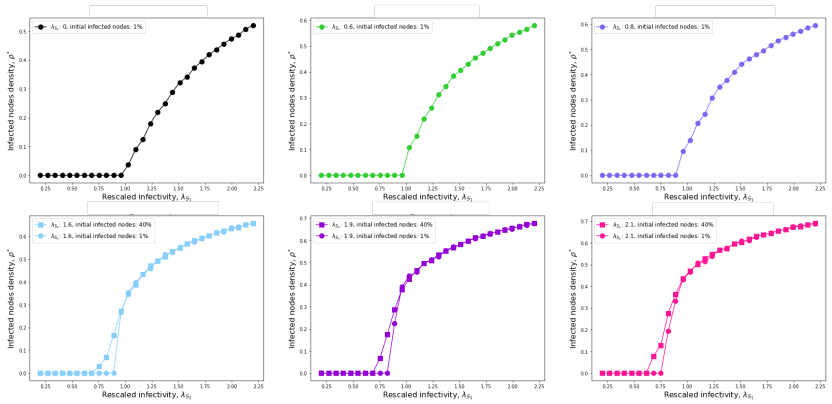

Figure 5: Illustration of the hysteresis loop: SCM on a RSC with $\langle k_{S_1} \rangle \approx 23$, $\langle k_{S_2} \rangle \approx 7$ and $\mu = 0.06$.

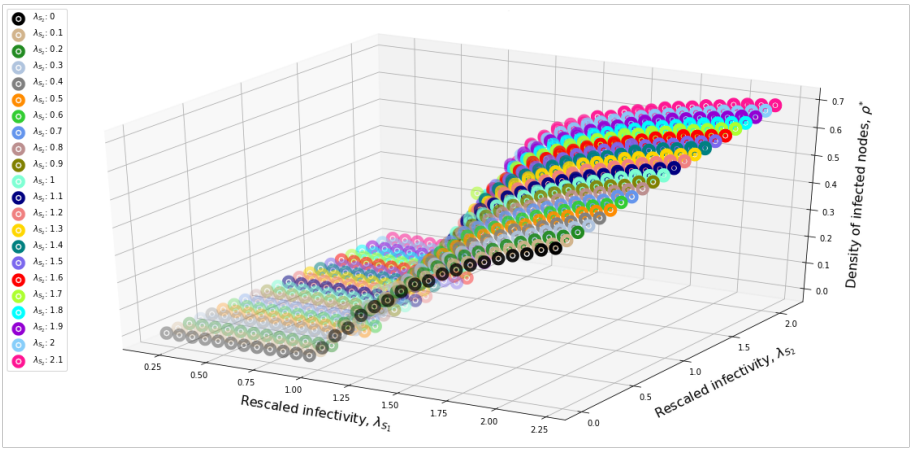

Figure 6: SCM on a RSC with $\langle k_{S_1} \rangle \approx 23$, $\langle k_{S_2} \rangle \approx 7$ and $\mu = 0.06$. We plot ρ^* against λ_{S_1} for different values of λ_{S_2}.

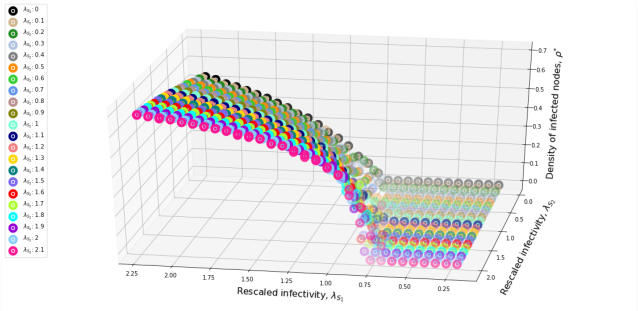

Figure 7: The same plot, from a different angle. Note that we only included simulations for which we chose an initial density of 40% of infected nodes. The hysteresis loop is thus not visible on this plot.

4.3 Numerical simulations on 2-simplicial complexes with seasonality

In this section we go beyond the analysis in [IPBL19] to illustrate the effect of seasonality on the abrupt phase transition and we restrict our attention to 2-simplicial complexes. We introduce seasonality by varying the infectivity parameters β_{S_1} and β_{S_2} through time using $\beta_{S_1} = \beta_{S_2} = cos(t/100)$. This means that no infection occurs when the cosinus is negative.. This approach was motivated directly by the impact seasonality has on the spread of the COVID-19. We were interested in the role of seasonality on the abrupt phase transition. The results obtained show that the effect of seasonality is not as impressive as one would imagine.

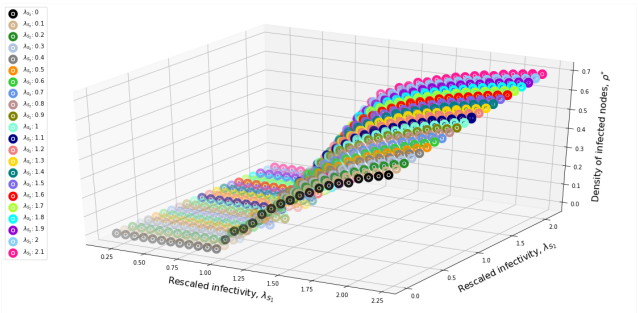

Figure 8: SCM on a RSC with $\langle k_{S_1} \rangle \approx 23$, $\langle k_{S_2} \rangle \approx 7$ and $\mu = 0.06$. We plot ρ^* against λ_{S_1} for different values of λ_{S_2}, where β_{S_1} and β_{S_2} are now the maximal possible infection parameters.

4.4 Numerical simulations on 3-simplicial complexes

We start with an initial density ρ_0 of infectious nodes in a synthetic 3-simplicial complex of 401 nodes, with $\langle k_{S_1} \rangle \approx 31$ and $\langle k_{S_2} \rangle \approx 7$, $\langle k_{S_3} \rangle \approx 2$, and consider all possible channels of infection. We stop a simulation if an absorbing state is reached - all the nodes are S, or all the nodes are I, otherwise we compute the average density of infectious node ρ^* by averaging the values measured in the last 100 times-step of the total of 1300 time-steps (note that this is not enough to get to the stationary density but would appropriately model the spread of an epidemic after 1300 days, for example). The results are averaged over 150 runs obtained by randomly generating a simplicial complex and randomly placing initial infectious nodes with the same density ρ_0. We use the re-scaled parameters $\lambda_{S_1} = \beta_{S_1}\langle k_{s_1} \rangle/\mu$, $\lambda_{S_2} = \beta_{S_2}\langle k_{S_2} \rangle/\mu$ and $\lambda_{S_3} = \beta_{S_3}\langle k_{S_3} \rangle/\mu$ as already defined. We plot the average fraction of infectious nodes ρ^* at the end of the process as a function of λ_{S_1} for the values $\lambda_{S_2} = 1.8$ and $\lambda_{S_3} = 1$. The case where $\lambda_{S_2} = \lambda_{S_3} = 0$ would correspond to the standard SIS model. We use different values for ρ_0: 1%, 5%, 25%, 45% and 99%.

We still observe the appearance of an endemic state with $\rho^* > 0$ at a value of λ^c well below the epidemic threshold of the standard SIS model. This transition seems to be again discontinuous, with a hysteresis loop appearing in a region where both healthy $\rho^* = 0$ and endemic $\rho^* > 0$ states can co-exist. We would need to conduct the experiment for more time steps as the curves show that $\rho(t)$ did not reach the theoretical limit $\lim_{t\to\infty} \rho(t)$ given by the mean-field approach.

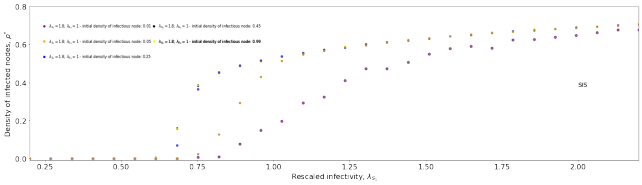

Figure 9: SCM on a RSC with $\langle k_{S_1} \rangle \approx 31$, $\langle k_{S_2} \rangle \approx 7$, $\langle k_{S_3} \rangle \approx 2$ and $\mu = 0.06$. We see a multi-stable region.

We would need to thoroughly conduct the bifurcation analysis (for all values of $\lambda_{S_1}, \lambda_{S_2}, \lambda_{S_3}$). In particular, in the plot above, the real solutions of the steady states equation:

$$0 = -\mu\rho[1 - \lambda_{S_1} + \lambda_{S_1}\rho - \lambda_{S_2}\rho + \lambda_{S_2}\rho^2 - \lambda_{S_3}\rho^2 + \lambda_{S_3}\rho^3]$$

are $\rho_1^* = 0$ and $\rho_{2,3,4}^* = a, b, c$, where:

$$[1 - \lambda_{S_1} + \lambda_{S_1}\rho - 1.8\rho + 1.8\rho^2 - \rho^2 + \rho^3] = (\rho - a)(\rho - b)(\rho - c)$$

where $-abc = 1 - \lambda_{S_1}$ and $ab + ac + bc = \lambda_{S_1} - \lambda_{S_2} = \lambda_{S_1} - 1.8$ and where $-a - b - c = \lambda_{S_2} - \lambda_{S_3} = 1.8 - 1 = 0.8$. Thus we cannot get three real solutions

in $(0,1)$. It is possible to get two real roots in $(0,1)$ - for example for $\lambda_{S_1} = 0.8$ we get roots approximately equal to $\rho_2^* \approx -1.53$, $\rho_3^* \approx 0.3$ and $\rho_4^* \approx 0.44$, as well as the root $\rho_1^* = 0$. In this case, $d_t\rho(t)$ is negative between ρ_1^* and ρ_3^* and positive between ρ_3^* and ρ_4^*. It is negative between ρ_4^* and 1. This indicates that we should expect to reach a final state of $\rho_1^* = 0$ density of infected notes in the large time limit if the initial density of infected nodes is below ρ_3^* and we should expect to reach a final state of ρ_4^* density of infected notes in the large time limit if the initial density of infected nodes is above ρ_3^*.

4.5 Numerical simulations on 4-simplicial complexes

We start with an initial density ρ_0 of infectious nodes in a synthetic 4-simplicial complex of 501 nodes, with $\langle k_{S_1}\rangle \approx 43$ and $\langle k_{S_2}\rangle \approx 17$, $\langle k_{S_3}\rangle \approx 5$, $\langle k_{S_4}\rangle \approx 1$, and consider all possible channels of infection.

We stop a simulation if an absorbing state is reached - all the nodes are S, or all the nodes are I, otherwise we compute the average density of infectious node ρ^* by averaging the values measured in the last 100 times-step of the total of 3333 time-steps. The results are averaged over 150 runs obtained by randomly generating a simplicial complex and randomly placing initial infectious nodes with the same density ρ_0. We use the re-scaled parameters $\lambda_{S_1} = \beta_{S_1}\langle k_{s_1}\rangle/\mu$, $\lambda_{S_2} = \beta_{S_2}\langle k_{S_2}\rangle/\mu$, $\lambda_{S_3} = \beta_{S_3}\langle k_{S_3}\rangle/\mu$ and $\lambda_{S_4} = \beta_{S_4}\langle k_{S_4}\rangle/\mu$ as already defined. We plot the average fraction of infectious nodes ρ^* at the end of the process as a function of λ_{S_1} for the values $\lambda_{S_2} = 1.065$ and $\lambda_{S_3} = 0$, $\lambda_{S_4} = 5$. The case where $\lambda_{S_2} = \lambda_{S_3} = \lambda_{S_4} = 0$ would correspond to the standard SIS model. We use different values for ρ_0: 1%, 5%, 25%, 45% and 99%.

We study the real solutions of the steady states equation:

$$0 = -\mu\rho[1 - \lambda_{S_1} + \lambda_{S_1}\rho - \lambda_{S_2}\rho + \lambda_{S_2}\rho^2 - \lambda_{S_3}\rho^2 + \lambda_{S_3}\rho^3 - \lambda_{S_4}\rho^3 + \lambda_{S_4}\rho^4]$$

For example for $\lambda_{S_1} = 0.999$ we get roots approximately equal to $\rho_2^* \approx 0.02244$, $\rho_3^* \approx 0.06978$, $\rho_4^* \approx 0.17407$ and $\rho_5^* \approx 0.73371$, as well as the root $\rho_1^* = 0$. In this case, $d_t\rho(t)$ is negative between $\rho_1^* = 0$ and ρ_2^*. It is positive between ρ_2^* and ρ_3^*. It is negative between ρ_3^* and ρ_4^* and it is positive between ρ_4^* and ρ_5^*. It is negative between ρ_5^* and 1.

This means that $\rho_1^* = 0$, $\rho_3^* = 0.06978$ as well as $\rho_5^* = 0.73371$ are stable roots. This indicates that we should expect to reach a final state of $\rho_1^* = 0$ density of infected notes in the large time limit if the initial density of infected nodes is below ρ_2^*; we should expect to reach a final state of ρ_3^* density of infected notes in the large time limit if the initial density of infected nodes is above ρ_2^* and below ρ_4^* and we should expect to reach a final state of ρ_5^* density of infected notes in the large time limit if the initial density of infected nodes is

above ρ_4^*.

5 Conclusion

We have studied hypergraphs models of contagion where the contagion of a node i occurs if all other nodes in the hyperedge are infectious. We could consider modified models where we require that more than half of other nodes are infectious. This has just very recently been proposed and is discussed in [dAPM20], but many simulations and mathematical analysis remain to be done. This could for instance model a social contagion phenomenon, where a person is influenced by its social spheres with a certain probability if more than half of the sphere believes in a rumour. We could also investigate temporal simplicial complexes and hypergraphs - this would model a population were the interactions are not fixed in time. Finally, we should explore generalizations to weighted graphs and weighted hypergraphs.

We also introduced the epidemic importance of a hyperlink. It remains to conduct simulations and epidemic containment strategies were we deactivate links and hyperlinks using the most important hyperlinks according to this quantity.

Code availability. The code used to produce Fig. 3 and 4 is slightly modified from code developed by I Iacopini (SimplagionModel Class) available under the GNU General Public License v3.0 [IPBL19]. We adapted the code to deal with higher-order simplicial complexes.

References

[BAS+18] Austin R Benson, Rediet Abebe, Michael T Schaub, Ali Jadbabaie, and Jon Kleinberg. Simplicial closure and higher-order link prediction. *Proceedings of the National Academy of Sciences*, 115(48):E11221–E11230, 2018.

[Ben19] Austin R Benson. Three hypergraph eigenvector centralities. *SIAM Journal on Mathematics of Data Science*, 1(2):293–312, 2019.

[Ber73] Claude Berge. Graphs and hypergraphs. 1973.

[BGL16] Austin R Benson, David F Gleich, and Jure Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.

[BKS16] Ágnes Bodó, Gyula Y Katona, and Péter L Simon. Sis epidemic propagation on hypergraphs. *Bulletin of mathematical biology*, 78(4):713–735, 2016.

[CB16] Owen T Courtney and Ginestra Bianconi. Generalized network structures: The configuration model and the canonical ensemble of simplicial complexes. *Physical Review E*, 93(6):062311, 2016.

[Cho19] Philip S Chodrow. Configuration models of random hypergraphs and their applications. *arXiv preprint arXiv:1902.09302*, 2019.

[dAPM19] Guilherme Ferraz de Arruda, Giovanni Petri, and Yamir Moreno. Social contagion models on hypergraphs. *arXiv preprint arXiv:1909.11154*, 2019.

[dAPM20] Guilherme Ferraz de Arruda, Giovanni Petri, and Yamir Moreno. Social contagion models on hypergraphs. *Physical Review Research*, 2(2):023032, 2020.

[ERV06] Ernesto Estrada and Juan A Rodríguez-Velázquez. Subgraph centrality and clustering in complex hyper-networks. *Physica A: Statistical Mechanics and its Applications*, 364:581–594, 2006.

[GABH+10] Sergio Gómez, Alexandre Arenas, J Borge-Holthoefer, Sandro Meloni, and Yamir Moreno. Discrete-time markov chain approach to contact-based disease spreading in complex networks. *EPL (Europhysics Letters)*, 89(3):38009, 2010.

[GGA13] Clara Granell, Sergio Gómez, and Alex Arenas. Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Physical review letters*, 111(12):128701, 2013.

[GGGMA11] Sergio Gómez, Jesús Gómez-Gardenes, Yamir Moreno, and Alex Arenas. Nonperturbative heterogeneous mean-field approach to epidemic spreading in complex networks. *Physical Review E*, 84(3):036105, 2011.

[GLPN93] Giorgio Gallo, Giustino Longo, Stefano Pallottino, and Sang Nguyen. Directed hypergraphs and applications. *Discrete applied mathematics*, 42(2-3):177–201, 1993.

[IPBL19] Iacopo Iacopini, Giovanni Petri, Alain Barrat, and Vito Latora. Simplicial models of social contagion. *Nature communications*, 10(1):1–9, 2019.

[JJK19] Bukyoung Jhun, Minjae Jo, and B Kahng. Simplicial sis model in scale-free uniform hypergraph. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):123207, 2019.

[KSS+16] Kerk F Kee, Lisa Sparks, Daniele C Struppa, Mirco A Mannucci, and Alberto Damiano. Information diffusion, facebook clusters, and the simplicial model of social aggregation: A computational simulation of simplicial diffusers for community health interventions. *Health communication*, 31(4):385–399, 2016.

[LKC+12] Hyekyoung Lee, Hyejin Kang, Moo K Chung, Bung-Nyun Kim, and Dong Soo Lee. Persistent brain network homology from the perspective of dendrogram. *IEEE transactions on medical imaging*, 31(12):2267–2277, 2012.

[LRS19] Renaud Lambiotte, Martin Rosvall, and Ingo Scholtes. From networks to optimal higher-order models of complex systems. *Nature physics*, 15(4):313–320, 2019.

[Luc13] Adam R Lucas. A fixed point theorem for a general epidemic model. *Journal of Mathematical Analysis and Applications*, 404(1):135–149, 2013.

[MAG18] Joan T Matamalas, Alex Arenas, and Sergio Gómez. Effective approach to epidemic containment using link equations in complex networks. *Science advances*, 4(12):eaau4212, 2018.

[MF13] Angélica S Mata and Silvio C Ferreira. Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks. *EPL (Europhysics Letters)*, 103(4):48003, 2013.

[MGA20] Joan T Matamalas, Sergio Gómez, and Alex Arenas. Abrupt phase transition of epidemic spreading in simplicial complexes. *Physical Review Research*, 2(1):012049, 2020.

[NML19] Leonie Neuhäuser, Andrew Mellor, and Renaud Lambiotte. Multibody interactions and non-linear consensus dynamics on networked systems. *arXiv preprint arXiv:1910.09226*, 2019.

[OPT+17] Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6(1):17, 2017.

[PET+14] Giovanni Petri, Paul Expert, Federico Turkheimer, Robin Carhart-Harris, David Nutt, Peter J Hellyer, and Francesco Vaccarino. Homological scaffolds of brain functional networks. *Journal of The Royal Society Interface*, 11(101):20140873, 2014.

[PPV17] Alice Patania, Giovanni Petri, and Francesco Vaccarino. The shape of collaborations. *EPJ Data Science*, 6(1):18, 2017.

[RDPS04] José J Ramasco, Sergey N Dorogovtsev, and Romualdo Pastor-Satorras. Self-organization of collaboration networks. *Physical review E*, 70(3):036106, 2004.

[REL+14] Martin Rosvall, Alcides V Esquivel, Andrea Lancichinetti, Jevin D West, and Renaud Lambiotte. Memory in network flows and its effects on spreading dynamics and community detection. *Nature communications*, 5(1):1–13, 2014.

[SCL18] Vsevolod Salnikov, Daniele Cassese, and Renaud Lambiotte. Simplicial complexes and complex systems. *European Journal of Physics*, 40(1):014001, 2018.

[SWG16] Ingo Scholtes, Nicolas Wider, and Antonios Garas. Higher-order aggregate networks in the analysis of temporal networks: path structures and centralities. *The European Physical Journal B*, 89(3):61, 2016.

[TCR10] Carla Taramasco, Jean-Philippe Cointet, and Camille Roth. Academic team formation as evolving hypergraphs. *Scientometrics*, 85(3):721–740, 2010.

[XOL16] Zheng Xie, Zhenzheng Ouyang, and Jianping Li. A geometric graph model for coauthorship networks. *Journal of Informetrics*, 10(1):299–311, 2016.

[YPVP17] Jean-Gabriel Young, Giovanni Petri, Francesco Vaccarino, and Alice Patania. Construction of and efficient sampling from the simplicial configuration model. *Physical Review E*, 96(3):032312, 2017.