

A pair-based approximation for simplicial contagion

Federico Malizia^{a,b}, Luca Gallo^c, Mattia Frasca^d, István Z. Kiss^{a,e}, Vito Latora^{b,f,g,h}, Giovanni Russoⁱ

^a Network Science Institute, Northeastern University London, London E1W 1LP, United Kingdom

^b Department of Physics and Astronomy, University of Catania, 95125 Catania, Italy

^c ANETI Lab, Corvinus Institute for Advanced Studies (CIAS), Corvinus University of Budapest, 1093 Budapest, Hungary

^d Department of Electrical, Electronics and Computer Science Engineering, University of Catania, 95125 Catania, Italy

^e Department of Mathematics, Northeastern University, Boston, MA 02115, USA

^f INFN Sezione di Catania, Via S. Sofia, 64, 95125 Catania, Italy

^g School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, UK

^h Complexity Science Hub Vienna, A-1080 Vienna, Austria

ⁱ Department of Mathematics and Computer Science, University of Catania, 95125 Catania, Italy

ARTICLE INFO

Keywords:

Contagion processes
Higher-order interactions
Simplicial complexes
Phase transitions

ABSTRACT

Higher-order interactions play an important role in complex contagion processes. Mean-field approximations have been used to characterize the onset of spreading in the presence of group interactions. However, individual-based mean-field models are unable to capture correlations between different subsets of nodes, which can significantly influence the dynamics of a contagion process. In this paper, we introduce a pair-based mean-field approximation that allows to study the dynamics of a SIS model on simplicial complexes by taking into account correlations at the level of pairs of nodes. Compared to individual-based mean-field approaches, the proposed approximation yields more accurate predictions of the dynamics of contagion processes on simplicial complexes. Specifically, the pair-based mean-field approximation provides higher accuracy in predicting the extent of the region of bistability, the type of transition from disease-free to endemic state, and the average time evolution of the fraction of infected individuals. Crucially, for the pair-based approximation we were able to obtain an analytical expression for the epidemic threshold, that elucidates the dependency on the parameters of the model. Through comparison with stochastic simulations, we show that our model correctly predicts that the onset of the epidemic outbreak in simplicial complexes depends on the strength of higher-order interactions. Overall, our findings highlight the importance of accounting for pair correlations when investigating contagion processes in the presence of higher-order interactions.

1. Introduction

Complex networks have been used to describe a large variety of dynamical processes involving interacting units, such as epidemic spreading [1,2], social contagion [3], random walks [4], and synchronization [5], among many others. However, networks have inherent limitations as they can only capture interactions between pairs of units [6,7]. Consequently, they cannot be employed to study complex systems characterized by interactions occurring in groups of three or more units [8]. In such cases, more sophisticated mathematical structures, such as hypergraphs [9] and simplicial complexes [10], are required. The inclusion of higher-order interactions has proved that these interactions give rise to novel collective phenomena in various dynamical processes [11], including diffusion [12,13], percolation [14], synchronization [15–17], and evolutionary games [18].

In particular, the framework of higher-order networks has proven to be essential in modeling social contagion processes, such as opinion formation, rumor spreading, or the adoption of novelties, where the exposure to multiple sources is needed to trigger the transmission [19]. Recent studies of social contagion in hypergraphs [20] and simplicial complexes [21] have revealed that higher-order interactions may radically change the characteristics of the spreading process. In the presence of higher-order interactions, the transition to an endemic state in SIS models can become discontinuous and a bistable regime where a disease-free and an endemic state co-exist can appear [20,21]. This behavior is closely connected to the microscopic organization of higher-order interactions as diverse structures may exhibit different types of transitions [22].

* Corresponding author at: Network Science Institute, Northeastern University London, London E1W 1LP, United Kingdom.
E-mail address: federico.malizia@nulondon.ac.uk (F. Malizia).

<https://doi.org/10.1016/j.chaos.2025.116776>

Received 4 April 2025; Received in revised form 16 June 2025; Accepted 17 June 2025

Available online 25 July 2025

0960-0779/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Most models for the analysis of social contagion processes with higher-order interactions rely on individual-based mean-field approaches, assuming that the individuals are homogeneously mixed and interact with each other at random. However, this assumption may lead to an oversimplified approximation, as it overlooks the dynamic correlations that arise within the underlying structure. For example, infected individuals are more likely to come into contact with other infected individuals [2]. In the context of complex networks, to obtain a more precise description of the spreading process, one can consider more sophisticated approximations. For instance, pair-based models can be employed [23–26]. In these models, the system dynamics is not characterized at the level of nodes but at the level of node pairs. That is, they do not follow the temporal evolution of the expected number of individuals in a given state, they analyze instead how the expected number of edges in a given state evolves in time.

Despite their increased complexity, pair-based models provide analytical insights into epidemic dynamics, yielding closed-form expressions for epidemic thresholds across diverse connectivity patterns and epidemiological frameworks [27,28]. Their versatility extends from homogeneous [29] to heterogeneous networks [30], while offering crucial guidance for intervention strategies including vaccination [31], contact tracing [32], and real-world epidemic control [33–35]. However, deriving pair-based mean-field models for spreading processes on structures with higher-order interactions remains challenging.

In this work, we focus on the complex contagion process introduced in Ref. [21], namely an SIS process in the presence of three-body simplicial interactions. We develop a mean-field model of the process based on a description at the level of node pairs, and we show that it reproduces the results of stochastic simulations of the process better than an individual-based mean-field model [21].

Recently, several novel approaches have been proposed to study and capture the emerging behavior in contagion processes in the presence of higher-order interactions. These methods include quenched group-based model [36], group-based approximate master equations [37], quenched pair approximation [38] and triadic approximation [39]. In particular, the authors of Ref. [38] have proposed a discrete-time model relying on a Markov-chain. The system dynamics is expressed in terms of joint probabilities of the microscopic states of the links and the nodes, and is thus described by a set of $N + L$ master equations, where N is the number of nodes and L the number of links. In our work, instead, we consider a framework that is continuous-time and with a number of governing equations that does not depend on the number of nodes and links. More recently, the authors of Ref. [39] have introduced a mean-field triadic approximation, which allows to describe higher-order contagion processes by associating state variables not only to nodes and pairs of nodes but also to groups of three nodes. This model correctly predicts key features of complex contagion, including the epidemic threshold and the stationary density of spreaders, and works for both simplicial complexes and hypergraphs. One of the main limitations of [36–39] is that they rely on the assumption that two three-body interactions (hyperedges or simplices) share at most one node. This can be a too strong assumption in structures with high clustering. The mean-field approximation we introduce in this paper stops at the level of pairs, but it allows one to properly account for correlations arising in motifs involving three and four nodes within a simplicial complex. Despite being limited to the case where higher-order interactions are represented by simplicial complexes our work provides a complementary approach to existing models, and can provide further insights into contagion processes in presence of higher-order interactions.

The paper is organized as follows. In Section 2 we provide the basic ideas and methods behind individual-based and pair-based mean-field approximations of the SIS model on networks. In Section 3 we focus instead on the case of the SIS model on simplicial complexes. We review the individual-based mean-field approximation proposed in

Ref. [21], and we introduce our mean-field pair-based approximation of the simplicial SIS model.

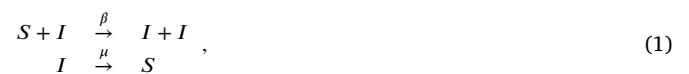
In Section 4, we show the advantages of a pair-based mean-field approximation compared to an individual-based one in reproducing the results of stochastic simulations of the simplicial SIS model. Furthermore, we derive an analytical expression of the epidemic threshold in the pair-based approximation, that elucidates how the onset of the epidemic outbreak also depends on the strength of higher-order interactions. Finally, in Section 5 we summarize and discuss the main results obtained.

2. Mean-field SIS models

In this section, we focus on the case of simple contagions, i.e. processes where transmission occurs exclusively through pairwise interactions. We discuss the two standard mean-field approximations for the SIS model on networks. Specifically, we describe the individual-based and the pair-based mean-field, highlighting the mathematical approach used to derive mean-field models at the individual- and pair-based levels. This will lay the groundwork for revisiting the individual-based approximation for the simplicial SIS model studied in Ref. [21] and for introducing our pair-based approximation in Section 3.2.

2.1. Individual-based mean-field SIS model

We begin by characterizing the processes governing the transition of an individual from one state to another in the standard SIS model. In this model, an individual can be either susceptible (S) or infected/infectious (I), and can transit from one compartment to another as infection or recovery takes place. A susceptible individual (S) becomes infected (S→I) after interacting with an infectious individual (I), who acts a mediator of the transition. The infection mechanism is, hence, a two-body nonlinear process. Infected individuals (I) recover after a given period of time, becoming once again susceptible (I→S). Contrarily to the infection mechanism, recovery is described by a one-body linear process. Formally, we can express these transitions in terms of two kinetic equations



where β and μ are the transition rates for the infection and the recovery processes, respectively.

The dynamics of the SIS model on network can be investigated using different approaches [1,2].

A commonly employed strategy is to perform stochastic simulations of the kinetic equations (1). While stochastic simulations can provide statistically exact results and can be implemented for very large systems [40,41], mean-field approaches offer complementary advantages. They allow for analytical insights into the relationship between model parameters and emerging collective behavior, facilitating the derivation of general theoretical results such as epidemic thresholds and phase transition conditions [1,2]. For this reason, it is common to use a deterministic representation of the process that focuses on population-level quantities, such as the density or the expected number of individuals in a given state. This deterministic approach assumes homogeneous mixing: interactions between individuals are treated as uniform and independent of the network structure, so that each individual can be considered statistically equivalent to any other. This hypothesis allows us to describe the dynamics of the system in terms of a so-called mean-field model.

The SIS process in homogeneous mixing is described by the following equations

$$\begin{aligned} \langle \dot{S} \rangle &= \mu \langle I \rangle - \beta k \langle SI \rangle \\ \langle \dot{I} \rangle &= -\mu \langle I \rangle + \beta k \langle SI \rangle \end{aligned} \quad (2)$$

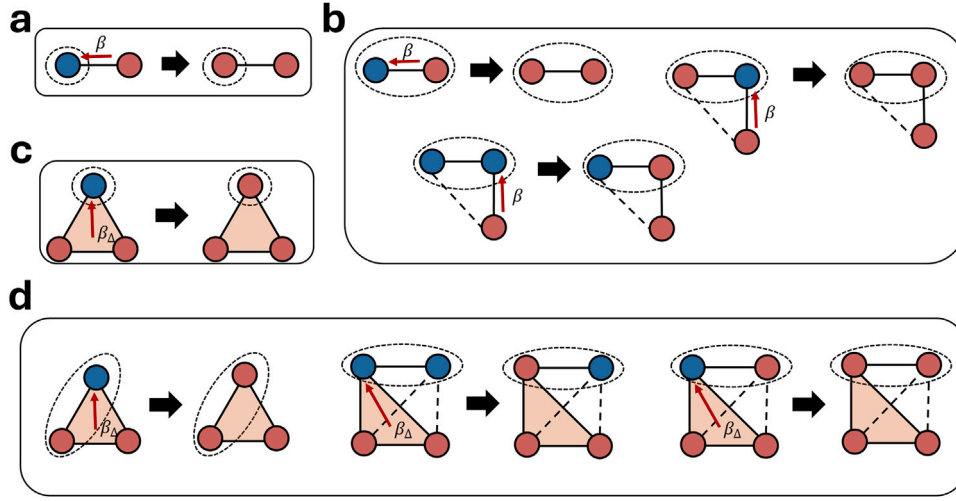


Fig. 1. Pictorial representation of infection processes among susceptible (in blue) and infected (in red) individuals in SIS models. (a) Infection of a node connected to an infected node through a link. This is the only infection process occurring in the individual-based SIS model. (b) Infection of a pair of nodes. In a pair-based description of contagion in a SIS model, the disease can be transmitted to one of the nodes of a pair by the other node of the pair or by other infected nodes not belonging to the pair. (c) Simplicial infection of a node by a group interaction with two infected nodes. (d) Infection of a pair of nodes in the presence of group interactions. The infection processes depicted in panel (d) arise in the pair-based description of the simplicial SIS model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $\langle S \rangle$ and $\langle I \rangle$ represent the densities of susceptible and infected individuals, respectively, $\langle SI \rangle$ is the density of edges in the state (S, I) , i.e., a susceptible individual in contact with an infectious one (note that pairs are unordered, meaning that we do not distinguish (S, I) from (I, S)), while k is the average number of contacts. The densities of individuals in state S and I are given by $\langle S \rangle = [S]/N$ and $\langle I \rangle = [I]/N$, where $[S]$ and $[I]$ are the expected number of susceptible and infected individuals, respectively, and $N = [S] + [I]$ is the total number of individuals in the population. Consequently, $\langle S \rangle$ and $\langle I \rangle$ are not independent and satisfy the conservation relation $\langle S \rangle + \langle I \rangle = 1$. As can be easily checked, the terms on the right hand sides of Eqs. (2) sum to zero, so that $\langle \dot{S} \rangle + \langle \dot{I} \rangle = 0$. The density $\langle SI \rangle$ is defined as $\langle SI \rangle = [SI]/kN$, where $[SI]$ represents the expected number of edges in state (S, I) . Note that $kN = 2L$, where L is the total number of edges within the network, assuming it to be undirected. The two terms in the equation for $\langle I \rangle$ represent the decrease in the density of infectious individuals due to recoveries, and the increase due to infections, respectively. The first depends on the transition rate μ and on the density of infected individuals, while the second is given by the transition rate β , the average degree k and the edge density $\langle SI \rangle$. The transition from susceptible to infected through the interaction with an infected individual is depicted in panel (a) of Fig. 1, from which one can derive the interaction terms appearing in (2). System (2) is exact but not closed, as the equations governing the dynamics of $\langle S \rangle$ and $\langle I \rangle$ depend on the quantity $\langle SI \rangle$. To obtain a closed system of equations, one may apply the law of mass action, assuming statistical independence at the level of individuals. This means to assume that infected individuals are randomly distributed in the network, so the probability that a neighbor is infected is simply given by $\langle I \rangle$ and does not depend on the state of the node itself. Under this assumption, we can approximate the fraction of edges in state (S, I) as

$$\langle SI \rangle \approx \langle S \rangle \langle I \rangle, \quad (3)$$

and substituting this expression in Eqs. (2), we obtain a closed form for the mean-field SIS model

$$\begin{aligned} \langle \dot{S} \rangle &= \mu \langle I \rangle - \beta k \langle S \rangle \langle I \rangle \\ \langle \dot{I} \rangle &= -\mu \langle I \rangle + \beta k \langle S \rangle \langle I \rangle \end{aligned} \quad (4)$$

2.2. Pair-based mean-field SIS model

Assuming that infected individuals are randomly distributed in the network may be a too rough approximation, as it does not account for

the dynamic correlations that exist within the contact network (e.g., infected nodes are more likely to come into contact with other infected nodes) [1]. To provide a more accurate description of the SIS dynamics on a network, we can consider a pair-based model, incorporating these dynamic correlations. In practice, one has to characterize the system dynamics at the level of pairs of nodes, describing how the expected number of edges in a given state evolves in time. With reference to Fig. 1, the equations describing the SIS process at the level of pairs of nodes are

$$\begin{aligned} \langle \dot{S} \rangle &= \mu \langle I \rangle - \beta k \langle SI \rangle \\ \langle \dot{I} \rangle &= -\mu \langle I \rangle + \beta k \langle SI \rangle \\ \langle \dot{SS} \rangle &= 2\mu \langle SI \rangle - 2\beta(k-1) \langle SSI \rangle \\ \langle \dot{SI} \rangle &= \mu \langle II \rangle - \mu \langle SI \rangle + \beta(k-1) \langle SSI \rangle \\ &\quad - \beta(k-1) \langle ISI \rangle - \beta \langle SI \rangle \\ \langle \dot{II} \rangle &= -2\mu \langle II \rangle + 2\beta(k-1) \langle ISI \rangle + 2\beta \langle SI \rangle \end{aligned} \quad (5)$$

where $\langle ISI \rangle$ represents the density of (both open and closed, and unordered) triplets in state (I, S, I) , and is defined as $\langle ISI \rangle = [ISI]/(k(k-1)N)$, with $[ISI]$ being the expected number of triplets in that state.¹ Note that the term $k(k-1)$ corresponds to the average number of triplets connected to each node. Indeed a node with k neighbors is at the center of $k(k-1)/2$ unordered triples. $\langle S \rangle$ and $\langle I \rangle$ can be obtained from $\langle SS \rangle$, $\langle SI \rangle$ and $\langle II \rangle$ through marginalization, namely $\langle S \rangle = \langle SS \rangle + \langle SI \rangle$ and $\langle I \rangle = \langle SI \rangle + \langle II \rangle$. Additionally, $\langle SS \rangle$, $\langle SI \rangle$ and $\langle II \rangle$ are not independent, as they are linked through the conservation relation $\langle SS \rangle + 2\langle SI \rangle + \langle II \rangle = 1$. It can be easily checked that, similarly to Eqs. (2), also Eqs. (5) satisfy $\langle \dot{S} \rangle + \langle \dot{I} \rangle = 0$. Moreover, Eqs. (5) satisfy $\langle \dot{SS} \rangle + \langle \dot{SI} \rangle = \langle \dot{S} \rangle$, $\langle \dot{SI} \rangle + \langle \dot{II} \rangle = \langle \dot{I} \rangle$, where we used the relation $\langle SSI \rangle + \langle ISI \rangle = \langle SI \rangle$.

Note that, with respect to the individual-based approximation, the pair-based SIS model contains a higher number of terms. This reflects the higher number of possible states and transitions that a pair can undergo compared to a single node.

To illustrate the terms in Eqs. (5), let us focus on the equation governing the dynamics of $\langle SI \rangle$. The first two terms are related to recoveries: the first represents the increase in the density $\langle SI \rangle$ due to the recovery of one of the nodes of the pairs in state (I, I) ; the second

¹ Note that factors 2 in the equations come from the fact that pairs of nodes are unordered [26].

models the decrease in $\langle SI \rangle$ due to the recovery of the infected nodes of the pairs in state (S, I) , i.e., the transition from (S, I) to (S, S) . The remaining three terms encode the transitions due to infections, as graphically represented in panel (b) of Fig. 1. In particular, the first term captures the increase in $\langle SI \rangle$ due to the infection of one of the nodes of the pairs in (S, S) , while the second represents the decrease in $\langle SI \rangle$ due to the infection of the susceptible nodes of the pairs in state (S, I) . Both transitions are due to the interactions with a third (infectious) node, among the $(k - 1)$ remaining neighbors of the susceptible nodes. Note also that the three nodes involved in the process can be arranged in two possible motifs, namely an open triangle, i.e., a wedge, or a closed one. This point will turn out to be crucial when considering the closure of the system. Finally, the last term corresponds to the decrease of $\langle SI \rangle$ due to the infection of the susceptible nodes in the pairs in state (S, I) . In this case, however, the transition is due to the infected nodes of the pair, and not to a third node.

System (5) is exact but not closed, as the equations governing the dynamics of $\langle SS \rangle$, $\langle SI \rangle$ and $\langle II \rangle$ depend on the quantities $\langle ISI \rangle$ and $\langle SSI \rangle$. Therefore, various approximations of the densities of the triplets have been proposed to close the system at the pair level. Here, we use one of the most common and studied closures [1], which consists of writing the probability that a triplet is in a given state as the product of the probabilities that its edges are in a certain state, normalized by the probabilities that the nodes in common to the pairs of edges are in a particular state. A key point of this closure is to distinguish between open and closed triangles. In the first case, the three nodes, namely x , y , and z form a wedge, that is, node y is linked to nodes x and z , but x and z are not connected. In the second case, there is a link also between x and z (we call this configuration, shortly, a triangle).

Let us first consider the case in which three nodes x , y , and z form a wedge are in states A , B , and C , respectively. We can write the density of wedges in state (A, B, C) as

$$\langle ABC^\wedge \rangle \approx \frac{\langle AB \rangle \langle BC \rangle}{\langle B \rangle}, \quad (6)$$

where $\langle AB \rangle = P(AB)$ and $\langle BC \rangle = P(BC)$ are the probabilities that links (x, y) and (y, z) are in states (A, B) and (B, C) , respectively, while $\langle B \rangle$ is the probability that node y is in state B . This formula can be obtained as follows. $\langle ABC^\wedge \rangle$ is given by

$$\langle ABC^\wedge \rangle = P(AB \cap BC) = P(BC)P(AB|BC)$$

where $P(AB|BC)$ denotes the conditional probability that x and y are in states A and B , given that y and z are in states B and C . Since x and z are not connected, we can approximate the conditional probability as

$$P(AB|BC) \approx P(AB|B) = P(AB)/P(B)$$

Combining these two relations we obtain Eq. (6). When the network has few short cycles, i.e., few closed triangles, one can use $\langle ABC \rangle \simeq \langle ABC^\wedge \rangle$ and Eq. (6) that, in this case, give an accurate approximation of the probability that a triplet of nodes is in state (A, B, C) . However, this closure fails to describe the epidemic dynamics on networks where the number of closed triplets is not negligible [1,23], as it overlooks the correlation between nodes x and z .

Let us now consider the case when the three nodes form a closed triangle. Here, to approximate the density of triangles in state (A, B, C) the so-called Kirkwood superposition [42], namely

$$\langle ABC^\Delta \rangle \approx \frac{\langle AB \rangle \langle BC \rangle \langle AC \rangle}{\langle A \rangle \langle B \rangle \langle C \rangle}. \quad (7)$$

can be used.² Note that, if we assume the dynamics of nodes x and z to be uncorrelated, i.e., $\langle AC \rangle = \langle A \rangle \langle C \rangle$, the approximation for $\langle ABC^\Delta \rangle$ recovers the one for $\langle ABC^\wedge \rangle$.

² The formula can be understood as follows: $P(AB \cap BC \cap CA) = P(AB \cap BC|CA)P(CA)$. Now $P(AB \cap BC|CA) = P(AB|BC \cap CA) \cdot P(BC|CA)$. We can write $P(AB|BC \cap CA) = \frac{P(AB)}{P(A)P(B)}$, and use the approximation

The final step is to calculate the probability that a generic triplet of nodes is in state (A, B, C) , namely $\langle ABC \rangle$, as a function of $\langle ABC^\wedge \rangle$ and $\langle ABC^\Delta \rangle$. To this aim, we consider the global clustering coefficient $\phi \in [0, 1]$, representing the fraction of closed triangles over all triplets in the structure, that we can calculate in a statistical meaning as

$$\phi = \frac{3\mathbb{E}(\mathcal{T})}{Nk(k-1)}, \quad (8)$$

where N is the number of nodes, k is the average number of links, and $\mathbb{E}(\mathcal{T})$ represents the expected number of triangles, which depends on the structure considered. At this point, we can calculate $\langle ABC \rangle$ as [1,23]:

$$\langle ABC \rangle = (1 - \phi)\langle ABC^\wedge \rangle + \phi\langle ABC^\Delta \rangle. \quad (9)$$

Note that the densities $\langle ABC^\wedge \rangle$ and $\langle ABC^\Delta \rangle$ are defined as $\langle ABC^\wedge \rangle = [ABC^\wedge]/(k(k-1)(1-\phi)N)$ and $\langle ABC^\Delta \rangle = [ABC^\Delta]/(k(k-1)\phi N)$, where $[ABC^\wedge]$ and $[ABC^\Delta]$ are the expected number of wedges and of triangles in state (A, B, C) , respectively.

Summing up, we consider Eq. (9) to close the compartmental models at the pair level. In particular, we will approximate the densities $\langle SSI \rangle$ and $\langle ISI \rangle$ as:

$$\begin{aligned} \langle SSI \rangle &\approx (1 - \phi) \frac{\langle SS \rangle \langle SI \rangle}{\langle S \rangle} + \phi \frac{\langle SS \rangle \langle SI \rangle^2}{\langle S \rangle^2 \langle I \rangle} \\ \langle ISI \rangle &\approx (1 - \phi) \frac{\langle SI \rangle^2}{\langle S \rangle} + \phi \frac{\langle SI \rangle^2 \langle II \rangle}{\langle S \rangle \langle I \rangle^2} \end{aligned} \quad (10)$$

Eqs. (5) along with the expressions (10) constitute a closed approximation for the pair-based mean-field SIS model.

3. Mean-field simplicial SIS models

In this section, we consider the case of complex contagion. We focus on the simplicial contagion model, and we begin by describing an individual-based approximation of such model [21]. Then, we introduce our novel pair-based mean-field approximation of simplicial contagion that extends the pair-based approximation on networks discussed in Section 2.2 to the case of simplicial complexes.

3.1. Individual-based mean-field simplicial SIS model

We now discuss the individual-based mean-field simplicial SIS model presented in [21]. We first describe the processes ruling the transition of an individual from one state to another. Compared to the SIS process on networks (see (1)), there is a further way in which an individual can transit from one compartment to another. Specifically, a susceptible individual (S) can become infected (S→I) through a three-body interaction, in which the other two individuals are infectious. In this higher-order interaction, two infected individuals at the same time act as mediators of the transition. We can represent the processes of the simplicial SIS model in terms of kinetic equations



where β and β_Δ are the transition rates for the two-body and the three-body infection process, respectively, while μ is the recovery rate.

We now derive a model that describes the system dynamics in terms of population-level quantities. In this framework, we consider each three-body interaction to also include all possible pairwise interactions among the three nodes. This represents the essential feature of

$P(AB|BC \cap CA) = \frac{P(AB)}{P(A)P(B)}$. Substituting in the previous expression one has $P(AB \cap BC|CA) = \frac{P(AB)}{P(A)P(B)} \cdot \frac{P(BC)}{P(C)}$ from which it follows $P(AB \cap BC \cap CA) = P(AB)P(BC)P(CA)/(P(A)P(B)P(C))$.

simplicial complexes, known as their inclusion property or downward closure [43], which makes them a very special case of hypergraphs. Under the homogeneous mixing hypothesis, the exact equations for the simplicial SIS model are given by

$$\begin{aligned} \langle \dot{S} \rangle &= \mu \langle I \rangle - \beta k \langle SI \rangle - \beta_{\Delta} \frac{k(k-1)\phi\delta}{2} \langle ISI^{\Delta} \rangle \\ \langle \dot{I} \rangle &= -\mu \langle I \rangle + \beta k \langle SI \rangle + \beta_{\Delta} \frac{k(k-1)\phi\delta}{2} \langle ISI^{\Delta} \rangle \end{aligned} \quad (12)$$

where $\delta \in [0, 1]$ is the fraction of triangles that are effectively 2-simplices, i.e., they represent a three-body interaction. Compared to Eqs. (2), these equations present an additional term, i.e., $\beta_{\Delta} k(k-1)\phi\delta \langle ISI^{\Delta} \rangle / 2$, which takes into account the infections of susceptible nodes due to the simultaneous interaction with two infected individuals, as shown in panel (c) of Fig. 1.

System (12) may be closed by applying the law of mass action, starting from the assumption that infected individuals are randomly distributed in the simplicial complex [21]. Accordingly, the density $\langle SI \rangle$ is approximated as in Eq. (3), while $\langle ISI^{\Delta} \rangle$ is approximated as

$$\langle ISI^{\Delta} \rangle \approx \langle I \rangle \langle S \rangle \langle I \rangle. \quad (13)$$

We can therefore write the individual-based mean-field model for the simplicial SIS model as

$$\begin{aligned} \langle \dot{S} \rangle &= \mu \langle I \rangle - \beta k \langle S \rangle \langle I \rangle - \beta_{\Delta} k_{\Delta} \langle S \rangle \langle I \rangle^2 \\ \langle \dot{I} \rangle &= -\mu \langle I \rangle + \beta k \langle S \rangle \langle I \rangle + \beta_{\Delta} k_{\Delta} \langle S \rangle \langle I \rangle^2 \end{aligned} \quad (14)$$

where $k_{\Delta} = k(k-1)\phi\delta/2$ represents the average number of 2-simplices connected to each node.

The system can be analytically investigated in terms of two parameters, $\lambda = k\beta/\mu$ and $\lambda_{\Delta} = k_{\Delta}\beta_{\Delta}/\mu$, representing the rescaled infectivity on 1-simplices, i.e., links, and on 2-simplices, respectively. In particular, the steady-state solutions of Eqs. (14) and their stability can be studied as a function of λ and λ_{Δ} . The analysis [21] shows that, when the higher-order interactions are weak, i.e., when $\lambda_{\Delta} < 1$, the system behaves similarly to the SIS model on networks, described by Eq. (2): for $\lambda < 1$ the disease-free equilibrium $\langle I \rangle^* = 0$ is the only solution, while for $\lambda > 1$ a stable endemic state $\langle I \rangle^* \neq 0$ exists.

In addition, the phase transition at $\lambda = 1$ is continuous. Instead, when the higher-order interactions are strong, i.e., $\lambda_{\Delta} > 1$, the system shows a different behavior. For λ smaller than $\lambda_c = 2\sqrt{\lambda_{\Delta}} - \lambda_{\Delta}$, only the disease-free equilibrium exists; for $\lambda_c < \lambda < 1$, the disease-free equilibrium and an endemic state coexist in a bistable regime, with the initial fraction of infected individuals determining whether the system reaches one equilibrium or the other: If the initial fraction of infected individuals is larger than a critical mass [21], the system evolves towards the endemic state; if not, the spreading cannot be sustained, and the system evolves towards the disease-free equilibrium. Finally, for $\lambda > 1$, the endemic state is the only stable equilibrium. Differently from the SIS model on networks, when $\lambda_{\Delta} > 1$ the phase transitions occurring at $\lambda = \lambda_c$ and $\lambda = 1$ are discontinuous.

3.2. Pair-based mean-field simplicial SIS model

In this section, we derive the pair-based mean-field simplicial SIS model. To this aim, we have to account for all ways a pair of nodes in a simplicial complex can transit from one state to another. Besides those related to recoveries and two-body infections (see Section 2), the transitions determined by three-body infections (illustrated in panel (d) of Fig. 1) should also be considered. Remarkably, two of these transitions depend on the state of 4-node motifs, and occur when the 2-simplex shares with the edge of interest a single node. The remaining one depends instead on the state of a 3-node motif, and corresponds to the case where the focal edge belongs to the 2-simplex. In the first two cases, infections occur when the shared node is susceptible and the other nodes of the 2-simplex are infected, while the other node in the pair can be either susceptible or infected. Both the densities of edges in states (S, S) and (S, I) are, hence, affected by this infection process,

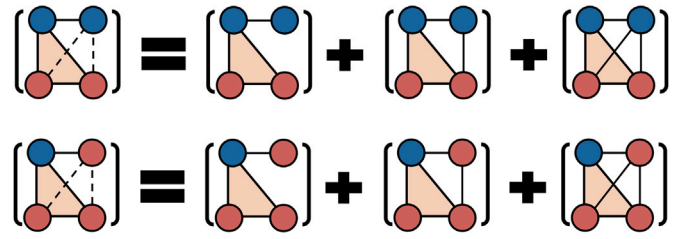


Fig. 2. Graphical representation of the three possible microscopical configurations of four-node motif states (I, I, S, S) (on top) and (I, I, S, I) (bottom) that contain a single 2-simplex at the state (I, S, I) , as defined in Eq. (15). Square brackets refer to the expected number of the configurations containing only one 2-simplex, which are obtained through the relation $[ABC^{\Delta}D] = Nk_{\Delta}(k-2)(1-\phi)^2(ABC^{\Delta}D)_0 + 2\phi(1-\phi)\langle ABC^{\Delta}D \rangle_1 + \phi^2\langle ABC^{\Delta}D \rangle_2$.

with a rate that is function of β_{Δ} and the density of 4-node motifs in the aforementioned states. To simplify, we only consider 4-node motifs that only contain a unique 2-simplex. As shown in Fig. 2, there are three possible motifs that contribute to this infection process. They differ for the number of links (zero, one or two) that connect the node external to the 2-simplex to the infected nodes of the 2-simplex. Hereby, we denote as $\langle IIS^{\Delta}S \rangle_0$, $\langle IIS^{\Delta}S \rangle_1$, and $\langle IIS^{\Delta}S \rangle_2$ the densities of these motifs in state (I, I, S, S) , while we denote as $\langle IIS^{\Delta}I \rangle_0$, $\langle IIS^{\Delta}I \rangle_1$, and $\langle IIS^{\Delta}I \rangle_2$ the density of motifs in state (I, I, S, I) . Using the clustering coefficient ϕ , we can write the density of quadruplets of nodes in states (I, I, S, S) and (I, I, S, I) as

$$\begin{aligned} \langle IIS^{\Delta}S \rangle &= (1-\phi)^2 \langle IIS^{\Delta}S \rangle_0 \\ &\quad + 2(1-\phi)\phi \langle IIS^{\Delta}S \rangle_1 \\ &\quad + \phi^2 \langle IIS^{\Delta}S \rangle_2 \\ \langle IIS^{\Delta}I \rangle &= (1-\phi)^2 \langle IIS^{\Delta}I \rangle_0 \\ &\quad + 2(1-\phi)\phi \langle IIS^{\Delta}I \rangle_1 \\ &\quad + \phi^2 \langle IIS^{\Delta}I \rangle_2 \end{aligned} \quad (15)$$

Next, to evaluate the terms in the model equations that quantify the transitions involving the state of 4-node motifs, we need to calculate the average number of motifs in which the pair of nodes is not part of the 2-simplex. As each node is connected to k links on average, the average number of motifs composed by a 2-simplex connected to a link is $k(k-1)\phi\delta(k-2)/2$. Indeed, the average number of 2-simplices connected to a node is $k(k-1)\phi\delta/2$, while the fourth node of the motif has to be chosen among the $(k-2)$ remaining neighbors of the node.

We now turn our attention to the three-body infection process involving 3-node motifs. In this case, the focal edge is part of the 2-simplex, and the only possible transition is the one from state (S, I) to state (I, I) . In fact, both the two other nodes of the 2-simplex have to be in the infectious state to yield a simplicial contagion, as expressed by the second kinetic equation (11). Similarly to the previous case, the overall contribution present in the model will depend on the infection rate β_{Δ} , on the density of 2-simplices in state (I, I, S) , which we denote as $\langle IIS^{\Delta} \rangle$, and on the average number of triangles an edge is part of. In particular, the latter is given by $(k-1)\phi\delta = 2k_{\Delta}/k$, where the factor 2 comes from the fact that each 2-simplex has two edges pointing to a node.

Finally, we can write the equations governing the simplicial SIS model at the pair-level

$$\begin{aligned} \langle \dot{S} \rangle &= \mu \langle I \rangle - \beta k \langle SI \rangle - \beta_{\Delta} k_{\Delta} \langle ISI^{\Delta} \rangle \\ \langle \dot{I} \rangle &= -\mu \langle I \rangle + \beta k \langle SI \rangle + \beta_{\Delta} k_{\Delta} \langle ISI^{\Delta} \rangle \\ \langle \dot{S}S \rangle &= 2\mu \langle SI \rangle - 2\beta(k-1) \langle SSI \rangle \\ &\quad - 2\beta_{\Delta} \frac{k_{\Delta}}{k} (k-2) \langle IIS^{\Delta}S \rangle \\ \langle \dot{S}I \rangle &= \mu \langle II \rangle - \mu \langle SI \rangle + \beta(k-1) \langle SSI \rangle \\ &\quad - \beta(k-1) \langle ISI \rangle - \beta \langle SI \rangle \\ &\quad + \beta_{\Delta} \frac{k_{\Delta}}{k} (k-2) \langle IIS^{\Delta}S \rangle - 2\beta_{\Delta} \frac{k_{\Delta}}{k} \langle ISI^{\Delta} \rangle \\ &\quad - \beta_{\Delta} \frac{k_{\Delta}}{k} (k-2) \langle IIS^{\Delta}I \rangle \\ \langle \dot{I}I \rangle &= -2\mu \langle II \rangle + 2\beta(k-1) \langle ISI \rangle + 2\beta \langle SI \rangle \\ &\quad + 4\beta_{\Delta} \frac{k_{\Delta}}{k} \langle ISI^{\Delta} \rangle + 2\beta_{\Delta} \frac{k_{\Delta}}{k} (k-2) \langle IIS^{\Delta}I \rangle \end{aligned} \quad (16)$$

We remark here a property of the model. As for the SIS process on networks, in the simplicial SIS contagion recovery is a 1-body mechanism and there are 2-body infections, so the dynamics of m -body variables in both the individual-based and the pair-based (exact) models depend on m -body and $(m+1)$ -body quantities. For instance, the dynamics of $\langle I \rangle$ (1-body variable) depends on $\langle I \rangle$ itself and on $\langle SI \rangle$ (2-body). In the simplicial SIS process, however, there is a further 3-body infection mechanism at play. Consequently, the equations governing the evolution of the m -body variables in the individual-based and the pair-based models are determined by $(m+2)$ -body quantities too. Hence, the dynamics of $\langle I \rangle$ (1-body variable) depends not only on $\langle I \rangle$ itself and on $\langle SI \rangle$ (2-body), but also on $\langle ISI^4 \rangle$ (3-body).

Similarly to what we have done for Eqs. (5), we now close Eqs. (16) at the level of pairs. We use the approximations for the triplet densities $\langle ABC \rangle$ at the pair level given by Eqs. (6), (7), and (9). For the densities of the 4-node motifs, we rely on the closures introduced in [44] (see Eqs. 15), re-adapting them to the simplicial case. Specifically, we consider

$$\begin{aligned} \langle IIS^4S \rangle_0 &= \frac{\langle SI \rangle^2 \langle II \rangle \langle SS \rangle}{\langle S \rangle^2 \langle I \rangle^2} \\ \langle IIS^4S \rangle_1 &= \frac{\langle SI \rangle^3 \langle SS \rangle \langle II \rangle}{\langle S \rangle^3 \langle I \rangle^3} \\ \langle IIS^4S \rangle_2 &= \frac{\langle SS \rangle \langle SI \rangle^4 \langle II \rangle}{\langle S \rangle^4 \langle I \rangle^4} \\ \langle IIS^4I \rangle_0 &= \frac{\langle SI \rangle^3 \langle II \rangle}{\langle S \rangle^2 \langle I \rangle^2} \\ \langle IIS^4I \rangle_1 &= \frac{\langle SI \rangle^3 \langle II \rangle^2}{\langle S \rangle^2 \langle I \rangle^4} \\ \langle IIS^4I \rangle_2 &= \frac{\langle SI \rangle^3 \langle II \rangle^3}{\langle S \rangle^2 \langle I \rangle^6} \end{aligned} \quad (17)$$

which represent all the closures at the level of pair state variables for the 4-node state depicted in Fig. 2. It is worth noting that this is an approximation, as the 4-node state should depend on the dynamical correlations of the 3-node state variables [44]. Moreover, by closing the system at the level of pairs and considering the relation in Eq. (7) to express the state of closed triangles, the conservation relations that are exact for the system before the closure, i.e., up to Eqs. (16) ($\langle S \rangle + \langle I \rangle = 1$, $\langle SS \rangle + \langle SI \rangle = \langle S \rangle$, $\langle SI \rangle + \langle II \rangle = \langle I \rangle$), are no longer valid [24].

Summing up, Eqs. (16) along with the expressions (6), (7), (9), and (17) constitute a closed approximation for the pair-based mean-field simplicial SIS model. Notice that, while the individual-based model in Section 3.1 can be easily extended to the more general scenario of hypergraphs [20] where correlations among three- and two-body interactions are not considered, the pair-based approximation above only applies to the case of simplicial complexes, which satisfy the inclusion property.

4. Results

In this section, we analyze the pair-based approximation of the simplicial SIS model. We first present an analytical derivation of the epidemic threshold. Then, we compare the predictions of the pair-based model with the results of stochastic simulations on synthetic simplicial complexes. We show that the pair-based simplicial SIS model is able to predict the behavior of the epidemic threshold, the nature of the phase transitions, i.e., continuous or discontinuous, and the temporal evolution of the spreading process.

4.1. Epidemic threshold for the pair-based mean-field simplicial SIS model

To derive the epidemic threshold, we consider the initial phase of the spreading. At this stage, the population is almost entirely made up of susceptible individuals, and we can determine the condition under which an outbreak can occur by looking at when the density of infected individuals grows. By replacing the density of triangles $\langle ISI^4 \rangle$ in the

second of Eqs. (16) by the approximation in Eq. (7), we can write the dynamics of $\langle I \rangle$ as:

$$\begin{aligned} \langle \dot{I} \rangle &= -\mu \langle I \rangle + \beta k \langle SI \rangle + \beta_\Delta k_\Delta \frac{\langle SI \rangle^2 \langle II \rangle}{\langle S \rangle \langle I \rangle^2} \\ &= \mu \langle I \rangle \left(\frac{\beta k \langle SI \rangle}{\mu \langle I \rangle} + \frac{\beta_\Delta k_\Delta}{\mu \langle S \rangle} \left(\frac{\langle SI \rangle}{\langle I \rangle} \right)^2 \frac{\langle II \rangle}{\langle I \rangle} - 1 \right). \end{aligned} \quad (18)$$

Moreover, by defining the quantity \mathcal{R} as

$$\mathcal{R} = \frac{\beta k \langle SI \rangle}{\mu \langle I \rangle} + \frac{\beta_\Delta k_\Delta}{\mu \langle S \rangle} \left(\frac{\langle SI \rangle}{\langle I \rangle} \right)^2 \frac{\langle II \rangle}{\langle I \rangle} - 1, \quad (19)$$

we can rewrite this equation as

$$\langle \dot{I} \rangle = \mu \langle I \rangle \mathcal{R}. \quad (20)$$

The quantity $\mu \langle I \rangle$ remains non-negative throughout the entire spreading process. Therefore, $\mathcal{R} > 0$ represents the condition under which contagion can occur at the early stage of the process. Conversely, if $\mathcal{R} < 0$, the contagion dies out. We assume that the values of all the parameters in \mathcal{R} , namely μ, β, β_Δ and k_Δ are known. Hence, in order to assess whether an outbreak can occur or not, we only require further information about the quantities $\frac{\langle SI \rangle}{\langle I \rangle}$ and $\frac{\langle II \rangle}{\langle I \rangle}$ at the early stage of the contagion process. Therefore, it is convenient to define the new variables [23,45]

$$\Pi = \frac{\langle SI \rangle}{\langle I \rangle}; \quad \Psi = \frac{\langle II \rangle}{\langle I \rangle}, \quad (21)$$

which take finite (possibly non-zero) values even in the limit $t \rightarrow 0$, when $(\langle S \rangle, \langle I \rangle, \langle SS \rangle, \langle SI \rangle, \langle II \rangle) \rightarrow (1, 0, 1, 0, 0)$, and are able to capture the early time correlation between susceptible and infected nodes. Compared to the dynamics of $\langle I \rangle$, $\langle SI \rangle$ and $\langle II \rangle$, the two new quantities Π and Ψ are *fast variables* [45], meaning that they quickly converge to a quasi-equilibrium state, which we denote as $(\bar{\Pi}, \bar{\Psi})$. The equations governing the time evolution of these fast variables, and all the details about their derivation can be found in Appendix A. The introduction of the quasi-stationary states $(\bar{\Pi}, \bar{\Psi})$ allows us to study the behavior of \mathcal{R} at the early stage of the contagion. Specifically, by defining the rescaled infection rates $\lambda = k\beta/\mu$ and $\lambda_\Delta = k_\Delta\beta_\Delta/\mu$, we can express Eq. (19) at the disease-free state equilibrium as

$$\mathcal{R} = \lambda \bar{\Pi} + \lambda_\Delta \bar{\Pi}^2 \bar{\Psi} - 1. \quad (22)$$

Given the condition for the epidemic outbreak, $\mathcal{R} > 0$, we can express the critical value λ^* above which the disease-free equilibrium is unstable as:

$$\lambda^* = \frac{1 - \lambda_\Delta \bar{\Pi}^2 \bar{\Psi}}{\bar{\Pi}}. \quad (23)$$

The complexity of the model makes formulating an exact closed-form solution for the quasi-stationary states $(\bar{\Pi}, \bar{\Psi})$ in terms of the model parameters not analytically tractable. However, the values of $(\bar{\Pi}, \bar{\Psi})$ can be numerically estimated (see Appendix A). More importantly, an approximate analytical expression for the stationary states of the fast variables which holds at the critical value λ^* can be derived. In turn, this makes it possible to derive an approximate analytical expression for λ^* . Assuming that λ_Δ is small and $\phi = 0$, the quasi-stationary states $(\bar{\Pi}, \bar{\Psi})$ can be approximated as

$$\begin{aligned} \bar{\Psi} &\approx \frac{1}{k} + \frac{\lambda_\Delta (k-1)^2}{k^4}, \\ \bar{\Pi} &\approx 1 - \bar{\Psi}. \end{aligned} \quad (24)$$

This approximation assumes that $\bar{\Pi} \approx 1 - \bar{\Psi}$. Note that the condition $\Pi(t) = 1 - \Psi(t)$ holds only approximately, since the system of Eqs. (16) is no longer exact under the closure in Eq. (7). By substituting the approximate solutions in Eq. (24) for the quasi-stationary states $(\bar{\Pi}, \bar{\Psi})$ into Eq. (23) and ignoring all the terms of order λ_Δ^2 and higher, we obtain an approximate expression for the critical λ^* as

$$\lambda^* \approx \frac{k}{k-1} - \lambda_\Delta \frac{k-2}{k^2}. \quad (25)$$

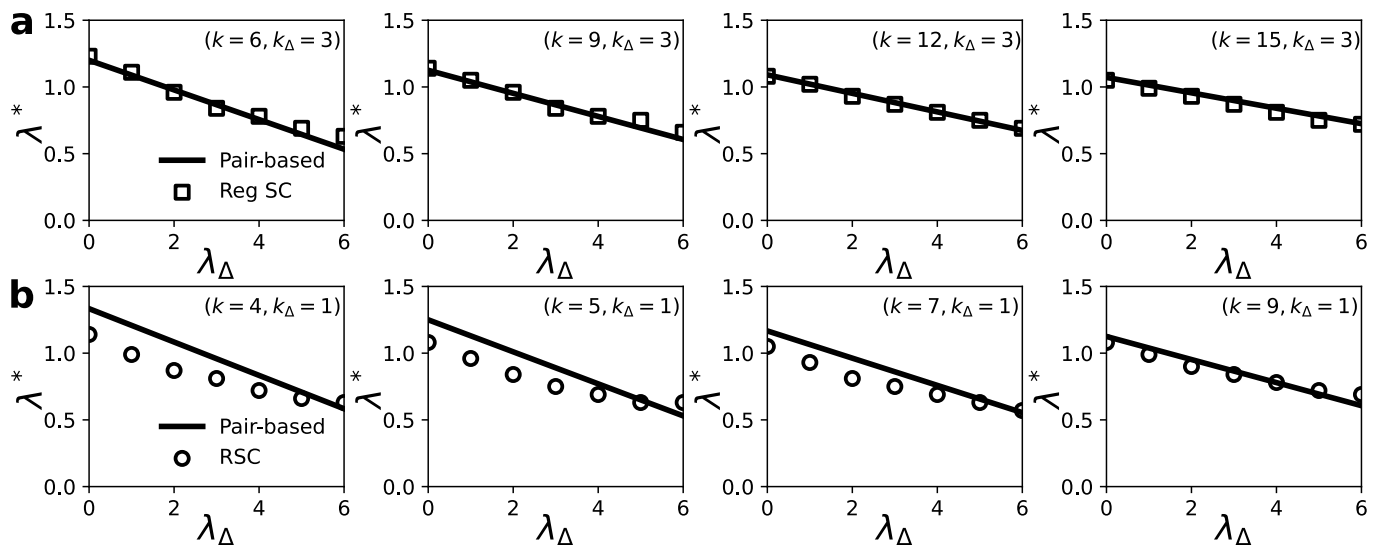


Fig. 3. Comparison of the epidemic threshold λ^* as predicted by Eq. (25) (black solid lines) with simulation results with different values of (k, k_Δ) in Regular Simplicial Complexes (black squares in panel a) and Random Simplicial Complexes (black circles in panel b).

A detailed analysis of this approximation is provided in Appendix A. Eq. (25) shows that the epidemic threshold, and so the stability of the disease-free equilibrium, depends on the presence of higher-order interactions. When $\lambda_\Delta = 0$, i.e., when there are no higher-order interactions, the epidemic threshold in Eq. (25) coincides with that of the SIS pair-based mean-field model on traditional networks [1]. As the strength of higher-order interactions increases, i.e. as λ_Δ increases, the value of λ^* decreases linearly. This indicates that, in presence of higher-order interactions, the system becomes more susceptible to an epidemic outbreak, highlighting the critical role of these interactions in the contagion. Additionally, Eq. (25) predicts that the impact of higher-order interactions on the epidemic threshold depends on the degree of pairwise connectivity. Specifically, for sparser simplicial complexes, higher levels of connectivity in the pairwise interactions reduce the susceptibility of the system to outbreaks. As we will show in the next subsection, this result is in agreement with the numerical simulations of the simplicial SIS model on synthetic simplicial complexes. In contrast, the individual mean-field approximation of Section 3.1 predicts an epidemic threshold equal to that of the standard SIS model with only pairwise interactions, i.e., $\lambda^* = 1$ [21], independently from the value of λ_Δ . Notice that the introduction of fast variables allows us to carry out the linear stability analysis of the disease-free steady state via the numerical evaluation of the leading eigenvalue of the Jacobian matrix of the system given in Eqs. (16). In Appendix B, we show that the leading eigenvalue becomes zero exactly at $\lambda = \lambda^*$, where λ^* is given by Eq. (23). However, due to the complexity of the system, the linear stability analysis alone would have prevented us from obtaining a closed-form expression for λ^* .

4.2. Comparison with stochastic simulations

We now compare the predictions of the pair-based mean-field approximation with the results of stochastic simulations of the simplicial SIS model on synthetic simplicial complexes. These complexes are generated using two distinct random models: (i) a Regular Simplicial Complex [46] (Reg SC), in which each node has the same number of 2- and 3-body interactions, and (ii) a higher-order generalization of the Erdős-Rényi random network model [21], namely the Random Simplicial Complex (RSC).

First, we examine the validity of our expression for the epidemic threshold, given by Eq. (25), then we analyze the time evolution of the fraction of infected individuals, $\rho(t) = \langle I(t) \rangle$, with particular focus on

the steady-state value ρ^* as a function of the model parameters λ and λ_Δ .

To assess the validity of Eq. (25), we perform stochastic simulations on different Regular Simplicial Complexes (Reg SCs), constructed using the methodology in [46]. In this way we have structures with an intra-order hyperedge overlap parameter set to zero, indicating that hyperedges of the same order at each node do not share any common nodes [22]. Specifically, we consider four different structures, all consisting of $N = 2000$ nodes with $k_\Delta = 3$ and different values of degree $k = \{6, 9, 12, 15\}$. We then run simulations on four different Random Simplicial Complexes (RSC), all consisting of $N = 2000$ nodes with $\langle k_\Delta \rangle \approx 1$, and with different values of $\langle k \rangle$, that is, $\langle k \rangle \approx \{4, 5, 7, 9\}$. For each structure, we conduct $M = 500$ simulation runs, initializing the system with an infection density of $\rho(0) = 0.001$. We calculate the thresholds λ^* from the simulations as the smallest value of λ for which a non-zero average value of the steady-state infection density, ρ^* , is obtained. We then compare these values of λ^* with those predicted by Eq. (25).

The results are shown in Fig. 3. Square symbols in Panel (a) indicate simulation results on Regular Simplicial Complexes, black circles in Panel (b) correspond to simulations on Random Simplicial Complexes, while solid black lines in both panels represent predictions of our model. Fig. 3 shows that the epidemic threshold λ^* decreases as the strength of higher-order interactions (controlled by λ_Δ) increases, for both classes of simplicial complexes we considered. Furthermore, we observe that higher-order interactions make sparser structures more susceptible to outbreaks, whereas they have a weaker impact on simplicial complexes with higher pairwise connectivity. This behavior is in excellent agreement with the prediction of Eq. (25), which indicates that the influence of higher-order interactions decreases with the degree k . This is consistent with previous studies [47,48] reporting that the impact of higher-order interactions is hindered by increased overall connectivity. Finally, although we derived Eq. (25) under the assumption of regular connectivity (which underpins the closure), it nevertheless captures the general behavior of the system even in the case of random structures, as shown in Fig. 3(b). While RSCs are homogeneous structures with narrow degree distributions, they differ from the perfectly regular case assumed in our theoretical derivation, demonstrating the robustness of our mean-field approximation.

We now study the capability of our pair-based model to predict the stationary density ρ^* , as a function of the rescaled infectivity rates. While for the individual-based approximation an analytical evaluation of both the stationary density ρ^* and the condition for the emergence of a saddle-node bifurcation (which marks the onset of the bistability

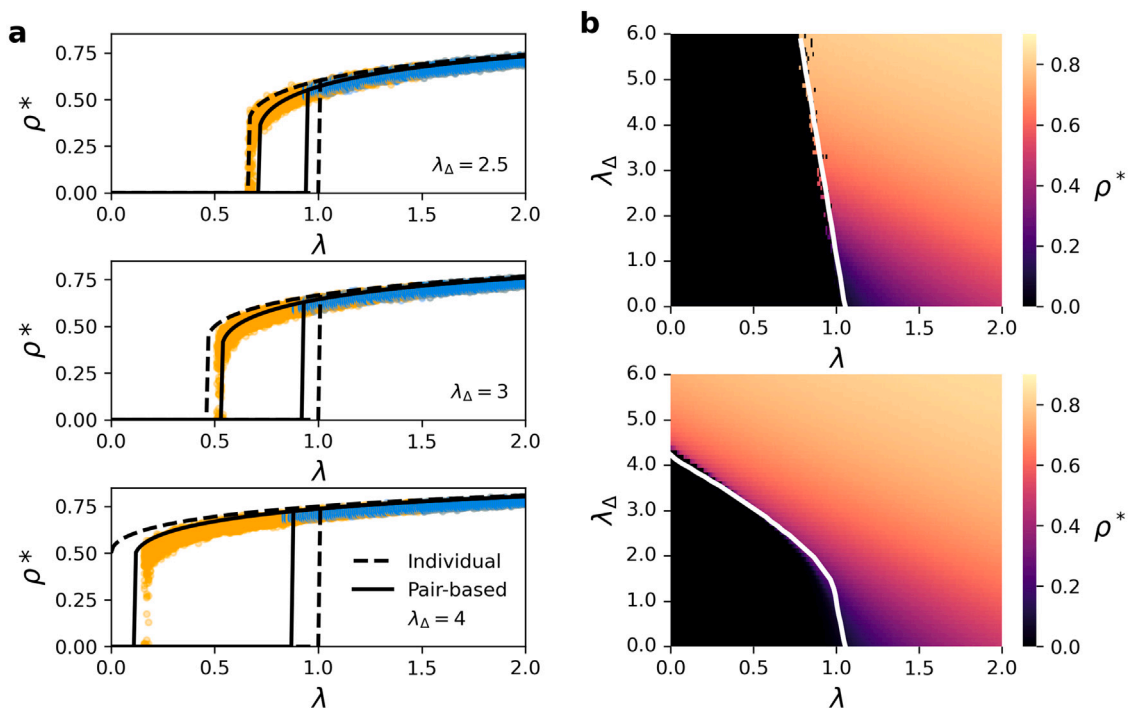


Fig. 4. Average fraction ρ^* of infected nodes in the stationary state for the SIS model on a random simplicial complex (RSC) with $N = 2000$ nodes, $k = 20$ and $k_\Delta = 6$. (a) Results of individual (black dashed lines) and pair-based (black solid lines) mean-field approximations are compared to the stochastic simulations for three different values of infectivity λ_Δ . The blue dots represent the values of the stationary densities obtained from $M = 100$ iterations of the stochastic simulations with $\rho(0) = 0.001$, while the orange dots represent the stationary densities obtained with $\rho(0) = 0.8$. (b) Phase diagrams reporting ρ^* as a function of the two parameters λ and λ_Δ . In the upper and lower panel we show the results of the stochastic simulations obtained by starting respectively with initial densities $\rho(0) = 0.001$ and $\rho(0) = 0.8$. The two solid white lines in the upper and lower panels indicate, respectively, the epidemic threshold λ^* as predicted by Eq. (25) and the critical value of λ , obtained via numerical integration with initial density $\rho(0) = 0.8$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

region) can be carried out, for the pair-based approximation this is not possible. Consequently, for the pair-based model, we obtain the stationary density ρ^* through numerical integration of Eqs. (16) over a sufficiently long time window, ensuring that the system has reached the steady-state. To investigate the possible onset of a bistable behavior in the pair-based approximation, for each parameter setting, we integrate the model using two distinct initial conditions: a very low fraction of initially infected individuals and a high (close to one) fraction. Having fixed the initial fraction of infected individuals $\langle I \rangle_0$, from which it follows $\langle S \rangle_0 = 1 - \langle I \rangle_0$, we set the initial conditions of the pair variables as

$$\begin{aligned} \langle SS \rangle_0 &= \langle S \rangle_0^2 \\ \langle SI \rangle_0 &= \langle S \rangle_0 \langle I \rangle_0 \\ \langle II \rangle_0 &= \langle I \rangle_0^2 \end{aligned} \quad (26)$$

Finally, to integrate the pair-based model, we need to evaluate ϕ , namely the fraction of closed triangles over all triplets in the simplicial complex, which is an input parameter of Eqs. (15). Henceforth, we will only compare the predictions of our model with the results of stochastic simulations obtained through a Random Simplicial Complex (RSC). Given that, for an RSC with N nodes, an average number of 1-simplices equal to k , and an average number of 2-simplices equal to k_Δ , the expected number of closed triplets is given by

$$\mathbb{E}(\mathcal{T}) \approx \binom{N}{3} \frac{k}{(N-1)} + \binom{N}{3} \frac{2k_\Delta}{(N-1)(N-2)}. \quad (27)$$

After some algebraic manipulations and using the definition in Eq. (8), we can approximate the global clustering coefficient ϕ for the RSC as

$$\phi \approx \frac{k^2(N-2)}{(N-1)^2(k-1)} + \frac{2k_\Delta}{k(k-1)}. \quad (28)$$

We carry out stochastic simulations on a RSC with $N = 2000$ nodes, $k \approx 20$ and $k_\Delta \approx 6$, setting $\rho(0)N$ randomly chosen nodes in the infectious state. For each setting of the model parameters, we perform $M = 100$ runs, each with a different instance of the RSC model, and

evaluate the stationary state ρ^* as the average over the last 100 values of $\rho(t)$. We run the simulations with two different values of $\rho(0)$, namely $\rho(0) = 0.001$ and $\rho(0) = 0.8$.

We now illustrate our results by comparing the stationary density ρ^* obtained with the stochastic simulations, the individual-based, and the pair-based model for different sets of parameters. Fig. 4(a) shows ρ^* as a function of the rescaled infectivity λ for three different values of λ_Δ . The outcome of the individual-based model is represented by a dashed black curve, while that of the pair-based model by a continuous black line. The results of the stochastic simulations with $\rho(0) = 0.001$ are depicted as blue dots, while those with $\rho(0) = 0.8$ as yellow dots.

We observe that both the individual-based and the pair-based approximations predict the existence of an interval of values of λ for which the system is bistable, and the transition from the disease-free state, $\rho^* = 0$, to an endemic equilibrium, $\rho^* > 0$, is discontinuous. However, the pair-based approximation, in general, better predicts the behavior of the stochastic simulations compared to the individual-based approximation. In particular, the individual-based model tends to overestimate the stationary fraction of infected individuals, while the pair-based model is in better agreement with the numerical simulations. Our analysis also shows that the pair-based approximation better identifies the two epidemic thresholds delimiting the region of bistability, while the individual-based model overestimates the width of this region. The individual-based model, in fact, underestimates the first threshold, representing critical point for the simulations starting from a high fraction of infected individuals, and, at the same time, overestimates the second threshold, representing the critical point for the simulations starting from a low fraction of infected individuals. In particular, the individual-based approximation predicts this second transition at $\lambda = 1$ for any value of λ_Δ , whereas the stochastic simulations display a discontinuity at values of $\lambda < 1$ that depend on λ_Δ . This behavior is instead correctly reproduced by the pair-based mean-field approximation. Moreover, the analytical expression of the epidemic

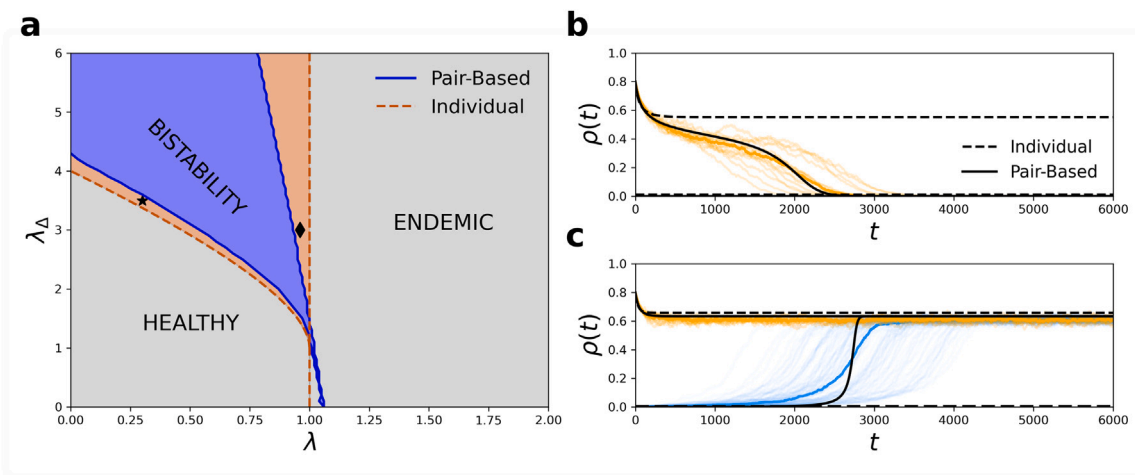


Fig. 5. Region of bistability of the SIS model and temporal evolution of the fraction of infected individuals. (a) Phase diagram, in the parameter space $(\lambda, \lambda_\Delta)$, highlighting the region of bistability obtained from the individual-based approximation (in orange) and the one derived from the pair-based approximation (in blue). (b) Temporal evolution of the fraction of infected individuals $\rho(t)$ for $(\lambda, \lambda_\Delta) = (0.3, 3.5)$, which corresponds to the point marked with a black star in (a), for the individual-based (dashed line) and pair-based (continuous line) approximations. (c) Temporal evolution of the fraction of infected individuals $\rho(t)$ for $(\lambda, \lambda_\Delta) = (0.95, 3)$, which corresponds to the point marked with a black diamond in (a), for the individual-based (dashed line) and pair-based (continuous line) approximations. In both panels (b) and (c), for each mean-field model we have illustrated the temporal evolution from two distinct initial conditions, namely $\rho(0) = 0.8$ and $\rho(0) = 0.001$. The colored curves depict $M = 100$ realizations of the spreading process on a RSC, with the colors representing the two different initial conditions of the simulations, namely $\rho(0) = 0.8$ (orange) and $\rho(0) = 0.001$ (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

threshold in Eq. (25) accurately predicts the value of λ at which the phase transition in the numerical simulations occurs.

We now investigate the behavior of ρ^* as a function of both λ and λ_Δ , comparing the results of the stochastic simulations on the RSC with the predictions of the pair-based mean field. Fig. 4(b) shows the results for $\rho(0) = 0.001$ (upper panel) and $\rho(0) = 0.8$ (lower panel). The color represents the values of ρ^* as obtained through the stochastic simulations, while the solid white lines denote the transition to a state with $\rho^* \neq 0$, as predicted by the pair-based model. Specifically, in the lower panel, representing the case where the initial density of infected nodes $\rho(0)$ is large, the white solid line corresponds to the leftmost threshold of the region of bistability, which we here indicate as λ_c . Instead, in the upper panel, obtained from simulations with a small $\rho(0)$, the white solid line corresponds to the epidemic threshold λ^* in Eq. (23). Also in this case, we observe that λ^* decreases as a function of λ_Δ , as predicted by Eq. (25). Moreover, λ_c decreases faster than λ^* as a function of λ_Δ suggesting that the width of the bistable region increases with the intensity of the three-body interactions. The dependence of both thresholds on λ_Δ is well reproduced by the pair-based approximation of the simplicial SIS model.

Next, we compare the region of bistability predicted by the individual-based and pair-based approximations. Fig. 5(a) shows the model behavior as a function of λ and λ_Δ , with the orange and blue areas indicating the region of bistability as predicted by the individual-based and the pair-based approximation, respectively. The darker lines represent the thresholds associated to the two approximations. Finally, the gray area represents the region that in both models corresponds to a single equilibrium, either disease-free $\rho^* = 0$ or endemic $\rho^* > 0$. Since the pair-based approximation well predicts the dependence of the critical points on λ_Δ , as shown in Fig. 4(b), then, Fig. 5(a) provides another indication that the individual-based model overestimates the width of the region of bistability, by simultaneously underestimating λ_c and overestimating λ^* . Overall, these results demonstrate that the pair-based approximation of the simplicial SIS process provides a better description of the spreading process on simplicial complexes compared to the individual-based mean-field approximation.

We conclude this section by illustrating in Fig. 5(b)–(c) the time evolution of $\rho(t)$ obtained from stochastic simulations on the RSC and from the individual-based and the pair-based approximations for two settings of $(\lambda, \lambda_\Delta)$, namely $(\lambda, \lambda_\Delta) = (0.3, 3.5)$ and $(\lambda, \lambda_\Delta) = (0.95, 3)$.

These values correspond to the points of the phase diagram in Fig. 5(a) marked with a star and a diamond, respectively. In both cases, the individual-based approximation predicts the existence of two stable equilibria, while the pair-based approximation a single one. Specifically, in the pair-based model, the steady-state equilibria correspond to a disease-free state for $(\lambda, \lambda_\Delta) = (0.3, 3.5)$ and to an endemic state for $(\lambda, \lambda_\Delta) = (0.95, 3)$, respectively. As shown in Fig. 5(b)–(c), the time evolution of $\rho(t)$ calculated from the stochastic simulations converges to a single equilibrium point, independently from the system initial conditions, as correctly predicted by the pair-based approximation.

5. Discussion and conclusions

In this work, we have derived a pair-based mean-field approximation of the simplicial contagion model and we have compared this approximation to the individual-based approximation. The key point of our derivation is the introduction of suitable approximations for densities of motifs of three and four nodes, to obtain closed equations at the level of pair of nodes. Consequently, our pair-based approximation can account for the effect of clustering and dynamical correlations that arise during a contagion process in the presence of both two-body and three-body interactions.

For the simplicial SIS model, an evident hierarchy emerges: the state of nodes depends on that of pairs and triples of nodes, the state of pairs depends on triples and quadruples, and so on. This dependency on larger motifs is curtailed using closures. In our model, variables at the level of triples or quadruples of nodes are approximated by node and pair-level variables. This leads to a “new” system that ideally preserves some of the desirable properties of the original system, particularly the conservation relationships at the level of nodes, namely, pairs, and so on. Indeed, the conservation relations of the state variables $\langle S \rangle + \langle I \rangle = 1$ and $\langle SS \rangle + 2\langle ST \rangle + \langle II \rangle = 1$ allow to reduce system in Eq. (16) to a restricted system of two equations. While it has been proved that the pair-based model is exact before closing the system [24], Eq. (7) does not preserve triple-level relationships, as it has been shown for pair-based models of SIR processes [49]. While an improved closure that preserves the conservation relations could be considered [32,45], using it for the pair-based approximation of the simplicial contagion model would have increased the system complexity without providing further understanding of the system behavior. Moreover, the revised

closure does not generally improve the agreement with the numerical simulation of the network process compared to classical closures [1]. Therefore, to avoid negative state variables and to limit the propagation of this error while integrating the model, we considered the full system of equations instead of the restricted one that uses conservation relations.

In this work we also derived an analytical expression for the epidemic threshold in the pair-based mean-field approximation. In absence of higher-order interactions, our formula correctly recovers the epidemic threshold for the classical SIS pair-based model on networked systems [1]. Otherwise, it predicts that the system is more susceptible to the epidemic outbreak in presence of higher-order interactions. Our analysis also allowed us to show that, while in a hypergraph with low correlations between two-body and three-body interactions the epidemic threshold is not affected by the higher-order interactions [39], in a simplicial complex — a specific type of hypergraph where an interaction among a group of d units necessarily includes all interactions among possible subgroups within that ensemble — it depends on the strength of higher-order interactions. Additionally, our analysis demonstrates that, when higher-order interactions are present, the sensitivity of the system to outbreaks in simplicial complexes decreases as pairwise connectivity increases.

Finally, we have compared the predictions of pair-based and individual-based approximations to the numerical simulations of the simplicial SIS model on synthetic random simplicial complexes. Our results show that a pair-based approach offers a more accurate description of the higher-order SIS contagion process than an individual-based one. This is evident in several different aspects, including the width of the bistability region, the nature of the transition from a disease-free to an endemic state, and the average time evolution of the fraction of infected individuals. In particular, the pair-based approximation correctly predicts that the epidemic thresholds depend on the higher-order infectivity. We notice that, although based on a different approach to derive the mean-field approximation, our conclusion on the dependence of the epidemic threshold from the higher-order infectivity is in agreement with [36,39] that show similar findings for epidemic processes occurring on higher-order structures that adhere to the inclusion property of simplicial complexes.

The pair-based approximation presented in our work is tailored for systems with higher-order interactions that can be represented as simplicial complexes. Although the model considered in this manuscript is based on the SIS process, the proposed pair-based framework can be extended to describe other epidemic dynamics, such as the SIR model, or more refined models that include an incubation period or additional infectious states, such as asymptomatic carriers. Hence, our approach is relevant in contexts such as human face-to-face contacts, which are well described in terms of simplicial complexes [50]. One important direction for future work, although not straightforward, is to extend our study to the more general case of hypergraphs. In general, hypergraphs do not necessarily satisfy the inclusion property, and incorporating dynamical correlations while maintaining analytical tractability in such structures may present significant challenges. Nonetheless, the framework developed here for simplicial complexes has proven valuable in identifying core features — such as the role of higher-order motifs and the importance of dynamical correlations — that must be taken into account when handling higher-order systems in general. Another possible direction to explore is the inclusion of both dynamical correlations and degree heterogeneity, so as to extend mean-field models to a larger class of real systems.

CRediT authorship contribution statement

Federico Malizia: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Luca Gallo:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal

analysis. **Mattia Frasca:** Writing – review & editing, Writing – original draft, Validation, Supervision. **István Z. Kiss:** Writing – review & editing, Methodology. **Vito Latora:** Writing – review & editing, Writing – original draft, Validation, Supervision. **Giovanni Russo:** Writing – review & editing, Writing – original draft, Validation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Derivation of fast variable dynamics

In this appendix, we provide the details for the calculation of the steady states of the fast variables Π and Ψ , which is the main ingredient in the derivation of the analytical expression for the epidemic threshold given in Eq. (25).

In the early stages of the spreading process, the fraction of infected individuals, $\langle I \rangle$, and the states of pairs involving infected individuals, namely $\langle SI \rangle$ and $\langle II \rangle$, tend to zero. Consequently, the quantity \mathcal{R} in Eq. (19), which determines whether the contagion occurs or dies out, is ill-defined. To resolve this issue, we consider the behavior of the system during the early stages of infection, when the population consists almost entirely of susceptible individuals. Under these conditions, the local correlations of infected states surrounding newly infected individuals develop at a much faster rate than the overall densities of susceptible and infected nodes [23]. Additionally, after an initial transient local correlations are expected to remain relatively constant as long as the clusters of infected nodes developing around the initially infected individuals (invading clusters) remain disconnected from each other. Therefore, we define the variables

$$\Pi = \frac{\langle SI \rangle}{\langle I \rangle}; \quad \Psi = \frac{\langle II \rangle}{\langle I \rangle}, \quad (29)$$

The quasi-stationary values of Π and Ψ can then be used to calculate the condition for the epidemic outbreak, namely the critical value in Eq. (22). This approach has been used to analytically determine the epidemic threshold in clustered networks [23,45,51,52]. Here, we extend this methodology to account for the presence of higher-order interactions.

The equation governing the dynamics of Π and Ψ are derived from the system in Eq. (16). To simplify the derivation and the tractability of the model, we consider for both 4-motif states $\langle IIS^4S \rangle$ and $\langle IIS^4I \rangle$, the microscopical configuration where no closed triangles are present in the motif, except the 2-simplex, namely

$$\begin{aligned} \langle IIS^4S \rangle &\approx (1 - \phi)^2 \langle IIS^4S \rangle_0 \\ \langle IIS^4I \rangle &\approx (1 - \phi)^2 \langle IIS^4I \rangle_0. \end{aligned} \quad (30)$$

Using the standard rule of differentiation and omitting the obvious time dependency, we have

$$\begin{aligned} \langle \dot{\Pi} \rangle &= \frac{\langle SI \rangle'}{\langle I \rangle} - \frac{\langle I \rangle' \langle SI \rangle}{\langle I \rangle^2} \\ \langle \dot{\Psi} \rangle &= \frac{\langle II \rangle'}{\langle I \rangle} - \frac{\langle I \rangle' \langle II \rangle}{\langle I \rangle^2}. \end{aligned} \quad (31)$$

Around the disease-free state equilibrium ($\langle S \rangle, \langle I \rangle, \langle SS \rangle, \langle SI \rangle, \langle II \rangle = (1, 0, 1, 0, 0)$), we obtain the following system of equations

$$\begin{aligned} \dot{\Pi} &= [(k-1)(1-\phi) - 1] \beta \Pi + \mu \Psi + [(k-1)\phi - k] \beta \Pi^2 \\ &\quad + \left[-\beta(k-1)\phi + \beta \Delta \frac{k_\Delta}{k} (k-2)(1-\phi)^2 - 2\beta \Delta \frac{k_\Delta}{k} \right] \Pi^2 \Psi - \beta \Delta k_\Delta \Pi^3 \Psi \\ \dot{\Psi} &= +2\beta \Pi - \mu \Psi - \beta k \Pi \Psi + \left[2\beta(k-1)\phi + 4\beta \Delta \frac{k_\Delta}{k} \right] \Pi^2 \Psi - \beta \Delta k_\Delta \Pi^2 \Psi^2. \end{aligned} \quad (32)$$

Note that $\Pi(t) + \Psi(t)$ does not sum in general to 1 as the system of Eqs. (16) is no longer exact under the closure introduced in Eq. (7). However, in Fig. 6(a) we show that the condition $\Pi(t) + \Psi(t) \approx 1$

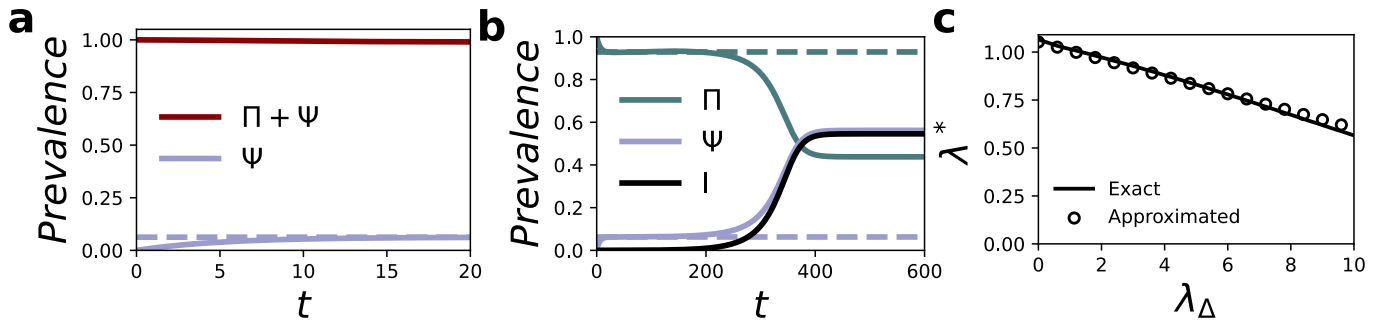


Fig. 6. (a) Early stage of the dynamical evolution of the fast variable Ψ (green solid line) and $\Psi + \Pi$ (red solid line), as obtained from numerical integration of the system in Eq. (16). (b) Dynamical evolution of the density of infected individuals (black solid line) and of the fast variables Π (green solid line) and Ψ (violet solid line), as obtained from numerical integration of the system in Eq. (16). The dashed lines represent the quasi-stationary values $\bar{\Pi}$ and $\bar{\Psi}$ as obtained from numerical integration of the system in Eq. (32). (c) Comparison between the critical value λ^* obtained from the exact epidemic threshold in Eq. (23) (solid line) and the approximation in Eq. (34) (dashed line), for a Random Simplicial Complex (RSC) with $N = 2000$, $k = 20$, $k_d = 6$, and $\phi \approx 0.05$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is consistent even before the variables have converged to their quasi-stationary states.

Since the system in Eq. (32) is derived around the steady-state equilibrium, the fixed points $(\bar{\Pi}, \bar{\Psi})$ of the fast variables coincide with the quasi-stationary states of the full system. However, due to the complexity of the system, an exact closed-form solution for the equilibria as functions of the model parameters is in general not analytically feasible. Therefore, we can evaluate $\bar{\Pi}$ and $\bar{\Psi}$ by numerically integrating the system in Eq. (32) across the entire set of model parameters and plug in these values in the expression of the critical threshold in Eq. (23). An example of the dynamical evolution of the fast variables Π and Ψ , obtained through numerical integration of Eq. (16), is illustrated in Fig. 6(b), along with the time evolution of the density of infected individuals. Fig. 6(a) zooms in on the initial phase of the spreading, showing how the fast variable Ψ (and Π , consequently) rapidly converges to a quasi-steady state.

We now show how to derive an approximate analytical expression for the stationary states of the fast variables at the critical value λ^* , as given in Eq. (23). We set $\dot{\Psi} = 0$ in Eqs. (32) and assume that near the stationary states $(\bar{\Pi}, \bar{\Psi})$, the condition $\bar{\Pi}(t) \approx 1 - \bar{\Psi}(t)$ holds. We consider the case where $\phi = 0$ and λ_d is small, and apply an asymptotic expansion of the quasi-stationary state $\bar{\Psi}$ around λ_d , namely $\bar{\Psi} = \bar{\Psi}_0 + \lambda_d \bar{\Psi}_1 + O(\lambda_d^2)$. Collecting all terms for $\bar{\Psi}_0$ and $\bar{\Psi}_1$, we obtain an approximate solution for the quasi-stationary states $(\bar{\Pi}, \bar{\Psi})$ as:

$$\begin{aligned} \bar{\Psi} &\approx \frac{1}{k} + \frac{\lambda_d(k-1)^2}{k^4}, \\ \bar{\Pi} &\approx 1 - \bar{\Psi}. \end{aligned} \quad (33)$$

By substituting this solution into Eq. (23), λ^* can be written as a rational function depending on λ_d . This can be further simplified by expanding it as an asymptotic series in λ_d , and neglecting all the terms of order λ_d^2 and higher, leading to $\lambda^* = \lambda_0 + \lambda_d \lambda_1 + O(\lambda_d^2)$. This leads to a simplified expression for the critical λ^* , namely

$$\lambda^* \approx \frac{k}{k-1} - \lambda_d \frac{k-2}{k^2}. \quad (34)$$

In Fig. 6(c), we show a comparison between the exact condition for the epidemic threshold in Eq. (23) and our approximation, in the case of a RSC with $N = 2000$, $k = 20$, $k_d = 6$ and $\phi \approx 0.05$. We observe that Eq. (34) represents a good approximation of the epidemic threshold for low values of λ_d , and gradually diverges from it for larger λ_d .

Appendix B. Numerical computation of the epidemic threshold

In this appendix, we show how the critical value λ^* , as determined by the condition in Eq. (23), is equivalent to the value obtained from the largest eigenvalue of the Jacobian matrix of the system around the disease-free state.

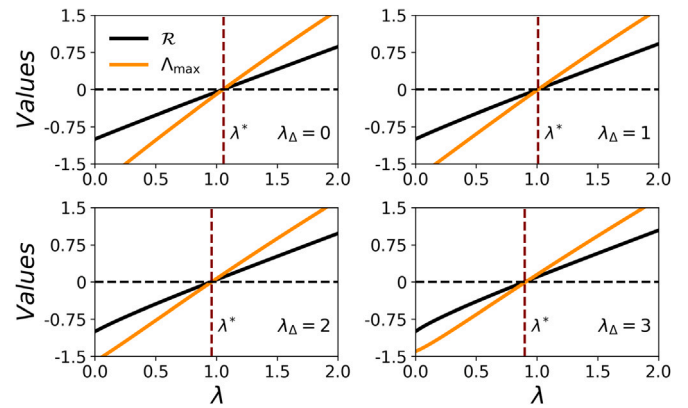


Fig. 7. Comparison between the numerical values of \mathcal{R} as given by Eq. (22) (black solid line) and the maximum eigenvalue of the Jacobian of the restricted system (orange line) as a function of λ for four different values of λ_d . The model parameters are $N = 2000$, $k = 20$, $k_d = 6$, $\mu = 0.05$, $\phi \approx 0.05$, and the values of $\bar{\Pi}$ and $\bar{\Psi}$ are obtained through Eq. (32). Both conditions are equal to zero at the same critical value, λ^* . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To streamline the calculations, we focus on the equations for $\langle S \rangle$, $\langle SI \rangle$, and $\langle SS \rangle$ from Eq. (16). We further simplify the derivation employing the approximation in Eq. (30). Finally, we use the closures for the triple states reported in Eq. (10) and for the four-node motif states in Eq. (17). We can thus express the restricted system of equations as given in Box I. The other variables of the system can be obtained considering the conservation of state variables, namely $\langle I \rangle = 1 - \langle S \rangle$ and $\langle II \rangle = 1 - 2\langle SI \rangle - \langle SS \rangle$.

To evaluate the stability of the disease-free state equilibrium, we calculate the Jacobian of the system, denoting

$$f_1(y) = \langle \dot{S} \rangle; \quad f_2(y) = \langle \dot{SI} \rangle; \quad f_3(y) = \langle \dot{SS} \rangle, \quad (36)$$

where $y = (S, SI, SS)$.

We can express the terms of the Jacobian matrix \mathbf{J} around the disease-free state equilibrium $y = (1, 0, 1)$ as given in Box II, where we have used the relation given in Eq. (29) to avoid the presence of ill-defined variables.

We focus on the condition at which the largest eigenvalue $\Lambda_{\max} = \Lambda_{\max}(\mathbf{J})$ of the Jacobian matrix crosses the imaginary axis, namely $\Lambda_{\max} > 0$.

In Fig. 7, we compare the numerical evaluation of both \mathcal{R} , defined in Eq. (22), and Λ_{\max} for $k = 20$, $k_d = 6$, $\mu = 0.05$, $\phi \approx 0.05$ and where the values of $\bar{\Pi}$ and $\bar{\Psi}$ are obtained through Eq. (32). Despite exhibiting different behaviors, both conditions converge to zero at the

$$\begin{aligned}
 \langle \dot{S} \rangle &= \mu \langle I \rangle - \beta k \langle SI \rangle - \beta_{\Delta} k_{\Delta} \frac{\langle SI \rangle^2 \langle II \rangle}{\langle S \rangle \langle I \rangle^2} \\
 \langle \dot{S} S \rangle &= 2\mu \langle SI \rangle - 2\beta(k-1)(1-\phi) \frac{\langle SS \rangle \langle SI \rangle}{\langle S \rangle} - 2\beta(k-1)\phi \frac{\langle S \rangle \langle SI \rangle^2}{\langle S \rangle^2 \langle I \rangle} - 2\beta_{\Delta} \frac{k_{\Delta}}{k} (k-2)(1-\phi)^2 \frac{\langle SI \rangle^2 \langle II \rangle \langle SS \rangle}{\langle S \rangle^2 \langle I \rangle^2} \\
 \langle \dot{S} I \rangle &= \mu \langle II \rangle - \mu \langle SI \rangle + \beta(k-1)(1-\phi) \frac{\langle SS \rangle \langle SI \rangle}{\langle S \rangle} + \beta(k-1)\phi \frac{\langle SS \rangle \langle SI \rangle^2}{\langle S \rangle^2 \langle I \rangle} - \beta(k-1)(1-\phi) \frac{\langle SI \rangle^2}{\langle S \rangle} - \beta(k-1)\phi \frac{\langle SI \rangle^2 \langle II \rangle}{\langle S \rangle \langle I \rangle^2} \\
 &\quad - \beta \langle SI \rangle + \beta_{\Delta} \frac{k_{\Delta}}{k} (k-2)(1-\phi)^2 \frac{\langle SI \rangle^2 \langle II \rangle \langle SS \rangle}{\langle S \rangle^2 \langle I \rangle^2} - 2\beta_{\Delta} \frac{k_{\Delta}}{k} \frac{\langle SI \rangle^2 \langle II \rangle}{\langle S \rangle \langle I \rangle^2} - \beta_{\Delta} \frac{k_{\Delta}}{k} (k-2)(1-\phi)^2 \frac{\langle SI \rangle^3 \langle II \rangle}{\langle S \rangle^2 \langle I \rangle^2}.
 \end{aligned} \tag{35}$$

Box I.

$$\begin{aligned}
 \left. \frac{\partial f_1(y)}{\partial \langle S \rangle} \right|_{y=(1,0,1)} &= -2\bar{\Pi}^2 \bar{\Psi} \beta_{\Delta} k_{\Delta} - \mu \\
 \left. \frac{\partial f_1(y)}{\partial \langle SI \rangle} \right|_{y=(1,0,1)} &= 2\bar{\Pi}^2 \beta_{\Delta} k_{\Delta} - 2\bar{\Pi} \bar{\Psi} \beta_{\Delta} k_{\Delta} - \beta k \\
 \left. \frac{\partial f_1(y)}{\partial \langle SS \rangle} \right|_{y=(1,0,1)} &= \bar{\Pi}^2 \beta_{\Delta} k_{\Delta} \\
 \left. \frac{\partial f_2(y)}{\partial \langle S \rangle} \right|_{y=(1,0,1)} &= -2\bar{\Pi}^2 \bar{\Psi} \beta \phi (k-1) + 2\bar{\Pi}^2 \bar{\Psi} \beta_{\Delta} \frac{k_{\Delta}}{k} (k-2)(1-\phi)^2 - 4\bar{\Pi}^2 \bar{\Psi} \beta_{\Delta} \frac{k_{\Delta}}{k} + \bar{\Pi}^2 \beta \phi (k-1) \\
 \left. \frac{\partial f_2(y)}{\partial \langle SI \rangle} \right|_{y=(1,0,1)} &= 2\bar{\Pi}^2 \beta \phi (k-1) - 2\bar{\Pi}^2 \beta_{\Delta} \frac{k_{\Delta}}{k} (k-2)(1-\phi)^2 + 4\bar{\Pi}^2 \beta_{\Delta} \frac{k_{\Delta}}{k} - 2\bar{\Pi} \bar{\Psi} \beta \phi (k-1) \\
 &\quad + 2\bar{\Pi} \bar{\Psi} \beta_{\Delta} (k-2)(1-\phi)^2 \frac{k_{\Delta}}{k} - 4\bar{\Pi} \bar{\Psi} \beta_{\Delta} \frac{k_{\Delta}}{k} + 2\bar{\Pi} \beta \phi (k-1) + \beta(1-\phi)(k-1) - \beta - 3\mu \\
 \left. \frac{\partial f_2(y)}{\partial \langle SS \rangle} \right|_{y=(1,0,1)} &= \bar{\Pi}^2 \beta \phi (k-1) - \bar{\Pi}^2 \beta_{\Delta} (1-\phi)^2 \frac{k_{\Delta}}{k} (k-2) + 2\bar{\Pi}^2 \beta_{\Delta} \frac{k_{\Delta}}{k} - \mu \\
 \left. \frac{\partial f_3(y)}{\partial \langle S \rangle} \right|_{y=(1,0,1)} &= -4\bar{\Pi}^2 \bar{\Psi} \beta_{\Delta} \frac{k_{\Delta}}{k} (k-2)(1-\phi)^2 - 2\bar{\Pi}^2 \beta \phi (k-1) \\
 \left. \frac{\partial f_3(y)}{\partial \langle SI \rangle} \right|_{y=(1,0,1)} &= 4\bar{\Pi}^2 \beta_{\Delta} \frac{k_{\Delta}}{k} (k-2)(1-\phi)^2 - 4\bar{\Pi} \bar{\Psi} \beta_{\Delta} \frac{k_{\Delta}}{k} (k-2)(1-\phi)^2 - 4\bar{\Pi} \beta \phi (k-1) - 2\beta(1-\phi)(k-1) + 2\mu \\
 \left. \frac{\partial f_3(y)}{\partial \langle SS \rangle} \right|_{y=(1,0,1)} &= 2\bar{\Pi}^2 \beta_{\Delta} \frac{k_{\Delta}}{k} (k-2)(1-\phi)^2
 \end{aligned} \tag{37}$$

Box II.

same value of λ . This convergence demonstrates the reliability of the relation obtained in Eq. (23) for evaluating the epidemic threshold in the pair-based model.

Data availability

Data will be made available on request.

References

- [1] Kiss István Z, Miller Joel C, Simon Péter L, et al. In: Mathematics of epidemics on networks, vol. 598, Cham: Springer; 2017.
- [2] Pastor-Satorras Romualdo, Castellano Claudio, Mieghem Piet Van, Vespignani Alessandro. Epidemic processes in complex networks. Rev Modern Phys 2015;87(3):925.
- [3] Gómez-Gardenes Jesus, Lotero L, Taraskin SN, Pérez-Reche FJ. Explosive contagion in networks. Sci Rep 2016;6(1):19767.
- [4] Noh Jae Dong, Rieger Heiko. Random walks on complex networks. Phys Rev Lett 2004;92(11):118701.
- [5] Gómez-Gardenes Jesús, Latora Vito. Entropy rate of diffusion processes on complex networks. Phys Rev E 2008;78(6):065102.
- [6] Boccaletti Stefano, Latora Vito, Moreno Yamir, Chavez Martin, Hwang D-U. Complex networks: Structure and dynamics. Phys Rep 2006;424(4-5):175-308.
- [7] Latora Vito, Nicosia Vincenzo, Russo Giovanni. Complex networks: principles, methods and applications. Cambridge University Press; 2017.
- [8] Battiston Federico, Cencetti Giulia, Iacopini Iacopo, Latora Vito, Lucas Maxime, Patania Alice, et al. Networks beyond pairwise interactions: structure and dynamics. Phys Rep 2020;874:1-92.
- [9] Berge Claude. In: Hypergraphs: combinatorics of finite sets, vol. 45, Elsevier; 1984.
- [10] Atkin RonaldH. From cohomology in physics to q-connectivity in social science. Int J Man-Mach Stud 1972;4(2):139-67.
- [11] Carletti Timoteo, Fanelli Duccio, Nicoletti Sara. Dynamical systems on hypergraphs. J Phys: Complex 2020;1(3):035006.
- [12] Carletti Timoteo, Battiston Federico, Cencetti Giulia, Fanelli Duccio. Random walks on hypergraphs. Phys Rev E 2020;101(2):022308.
- [13] Schaub Michael T, Benson Austin R, Horn Paul, Lippner Gabor, Jadbabaie Ali. Random walks on simplicial complexes and the normalized hodge 1-laplacian. SIAM Rev 2020;62(2):353-91.
- [14] Sun Hanlin, Bianconi Ginestra. Higher-order percolation processes on multiplex hypergraphs. Phys Rev E 2021;104(3):034306.
- [15] Skardal Per Sebastian, Arenas Alex. Abrupt desynchronization and extensive multistability in globally coupled oscillator simplexes. Phys Rev Lett 2019;122(24):248301.
- [16] Lucas Maxime, Cencetti Giulia, Battiston Federico. Multiorder laplacian for synchronization in higher-order networks. Phys Rev Res 2020;2(3):033410.
- [17] Gambuzza Lucia Valentina, Patti Francesca Di, Gallo Luca, Lepri Stefano, Romance Miguel, Criado Regino, et al. Stability of synchronization in simplicial complexes. Nat Commun 2021;12(1):1255.
- [18] Alvarez-Rodriguez Unai, Battiston Federico, Arruda Guilherme Ferraz de, Moreno Yamir, Perc Matjaž, Latora Vito. Evolutionary dynamics of higher-order interactions in social networks. Nat Hum Behav 2021;5(5):586-95.
- [19] Centola Damon, Macy Michael. Complex contagions and the weakness of long ties. Am J Sociol 2007;113(3):702-34.
- [20] Arruda Guilherme Ferraz de, Petri Giovanni, Moreno Yamir. Social contagion models on hypergraphs. Phys Rev Res 2020;2(2):023032.
- [21] Iacopini Iacopo, Petri Giovanni, Barrat Alain, Latora Vito. Simplicial models of social contagion. Nat Commun 2019;10(1):1-9.
- [22] Malizia Federico, Lamata-Otín Santiago, Frasca Mattia, Latora Vito, Gómez-Gardeñes Jesús. Hyperedge overlap drives explosive transitions in systems with higher-order interactions. Nat Commun 2025;16(1):555.
- [23] Keeling Matthew J. The effects of local spatial structure on epidemiological invasions. Proc R Soc Lond [Biol] 1999;266(1421):859-67.
- [24] Sharkey Kieran J, Kiss Istvan Z, Wilkinson Robert R, Simon Peter L. Exact equations for sir epidemics on tree graphs. Bull Math Biol 2015;77:614-45.
- [25] Frasca Mattia, Sharkey Kieran J. Discrete-time moment closure models for epidemic spreading in populations of interacting individuals. J Theoret Biol 2016;399:13-21.

- [26] Malizia Federico, Gallo Luca, Frasca Mattia, Latora Vito, Russo Giovanni. Individual-and pair-based models of epidemic spreading: Master equations and analysis of their forecasting capabilities. *Phys Rev Res* 2022;4(2):023145.
- [27] Trapman Pieter. On analytical approaches to epidemics on networks. *Theor Popul Biol* 2007;71(2):160–73.
- [28] House Thomas, Keeling Matt J. Insights from unifying modern approximations to infections on networks. *J R Soc Interface* 2011;8(54):67–73.
- [29] Bauch Chris T. The spread of infectious diseases in spatially structured populations: an invasy pair approximation. *Math Biosci* 2005;198(2):217–37.
- [30] Eames Ken TD, Keeling Matt J. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc Natl Acad Sci* 2002;99(20):13330–5.
- [31] House Thomas, Keeling Matt J. Epidemic prediction and control in clustered populations. *J Theoret Biol* 2011;272(1):1–7.
- [32] House Thomas, Keeling Matt J. The impact of contact tracing in clustered populations. *PLoS Comput Biol* 2010;6(3):e1000721.
- [33] Ferguson Neil M, Garnett Geoffrey P. More realistic models of sexually transmitted disease transmission dynamics: sexual partnership networks, pair models, and moment closure. *Sex Transm Dis* 2000;27(10):600–9.
- [34] Green DM, Kiss IZ, Kao RR. Modelling the initial spread of foot-and-mouth disease through animal movements. *Proc R Soc B: Biol Sci* 2006;273(1602):2729–35.
- [35] Keeling Matt J. Models of foot-and-mouth disease. *Proc R Soc B: Biol Sci* 2005;272(1569):1195–202.
- [36] Burgio Giulio, Arenas Alex, Gómez Sergio, Matamalas Joan T. Network clique cover approximation to analyze complex contagions through group interactions. *Commun Phys* 2021;4(1):111.
- [37] St-Onge Guillaume, Thibeault Vincent, Allard Antoine, Dubé Louis J, Hébert-Dufresne Laurent. Master equation analysis of mesoscopic localization in contagion dynamics on higher-order networks. *Phys Rev E* 2021;103(3):032301.
- [38] Matamalas Joan T, Gómez Sergio, Arenas Alex. Abrupt phase transition of epidemic spreading in simplicial complexes. *Phys Rev Res* 2020;2(1):012049.
- [39] Burgio Giulio, Gómez Sergio, Arenas Alex. Triadic approximation reveals the role of interaction overlap on the spread of complex contagions on higher-order networks. *Phys Rev Lett* 2024;132(7):077401.
- [40] Mata Angélica S, Ferreira Silvio C. Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks. *Europhys Lett* 2013;103(4):48003.
- [41] Jamieson-Lane Alastair, Blasius Bernd. Calculation of epidemic arrival time distributions using branching processes. *Phys Rev E* 2020;102(4):042301.
- [42] Kirkwood John G. Statistical mechanics of fluid mixtures. *J Chem Phys* 1935;3(5):300–13.
- [43] Salnikov Vsevolod, Cassese Daniele, Lambiotte Renaud. Simplicial complexes and complex systems. *Eur J Phys* 2018;40(1):014001.
- [44] House Thomas, Davies Geoffrey, Danon Leon, Keeling Matt J. A motif-based approach to network epidemics. *Bull Math Biol* 2009;71(7):1693–706.
- [45] Barnard Rosanna C, Berthouze Luc, Simon Péter L, Kiss István Z. Epidemic threshold in pairwise models for clustered networks: closures and fast correlations. *J Math Biol* 2019;79:823–60.
- [46] Lamata-Otín Santiago, Malizia Federico, Latora Vito, Frasca Mattia, Gómez-Gardeñes Jesús. Hyperedge overlap drives synchronizability of systems with higher-order interactions. *Physical Review E* 2025;111(3):034302.
- [47] Kiss István Z, Iacopini Iacopo, Simon Péter L, Georgiou Nicos. Insights from exact social contagion dynamics on networks with higher-order structures. *J Complex Netw* 2023;11(6):cnad044.
- [48] Kim Jihye, Lee Deok-Sun, Goh K-I. Contagion dynamics on hypergraphs with nested hyperedges. *Phys Rev E* 2023;108(3):034313.
- [49] Taylor Michael, Simon Péter L, Green Darren M, House Thomas, Kiss Istvan Z. From markovian to pairwise epidemic models and the performance of moment closure approximations. *J Math Biol* 2012;64:1021–42.
- [50] Landry Nicholas W, Young Jean-Gabriel, Eikmeier Nicole. The simpliciality of higher-order networks. *EPJ Data Sci* 2024;13(1):17.
- [51] Juher David, Ripoll Jordi, Saldaña Joan. Outbreak analysis of an sis epidemic model with rewiring. *J Math Biol* 2013;67:411–32.
- [52] Britton Tom, Juher David, Saldaña Joan. A network epidemic model with preventive rewiring: comparative analysis of the initial phase. *Bull Math Biol* 2016;78(12):2427–54.