

Replicating hypergraph disease dynamics with lower-order interactions

Eugene Tan^{*,†}*The Kids Research Institute Australia, Nedlands 6009, Australia*Michael Small[‡] and Shannon D. Algar[§]*The Complex Systems Group, The University of Western Australia, Crawley 6009, Australia*

(Received 27 August 2025; accepted 8 January 2026; published 23 January 2026)

Disease spreading models such as the ubiquitous SIS compartmental model and its numerous variants are widely used to understand and predict the behavior of a given epidemic or information diffusion process. A common approach to imbue more realism to the spreading process is to constrain simulations to a network structure, where connected nodes update their disease state based on pairwise interactions along the edges of their local neighborhood. Simplicial contagion models (SCM) extend this to hypergraphs such that groups of three nodes are able to interact and propagate the disease along higher-order hyperedges (triangles). Though more flexible, it is not clear the extent to which the inclusion of these higher-order interactions results in dynamics that are characteristically different from those attained from simpler pairwise interactions. Here, we propose an agent-based model that unifies the classical SIS/SIR compartmental model and SCM, and extends it to allow for interactions along hyperedges of arbitrary order. Using this model, we demonstrate how the steady-state dynamics of pairwise interactions can be made to replicate those of simulations that include higher-order topologies by linearly scaling disease parameters based on a proposed measure of network activity. By allowing disease parameters to dynamically vary over time, lower-order pairwise interactions can be made to closely replicate both the transient and steady-state dynamics of higher-order simulations. We demonstrate that this relationship is robust to misspecification in the assumed higher-order interaction model, and applies to non-clique complex hypergraphs with nontrivial heterogeneous topology. For the latter case, it is found that heterogeneities in hypergraph topology result in weakened approximations of higher-order dynamics by pairwise interactions.

DOI: [10.1103/6whh-kvk7](https://doi.org/10.1103/6whh-kvk7)

I. INTRODUCTION

Spreading processes over a given topology are ubiquitous across many social and physical systems. From epidemics to information diffusion, many studies aim to understand the governing rules of spreading processes in social contexts [1–4]. Nevertheless, the influence of the underlying topology on which spreading occurs can sometimes be as important as the mechanistic rules governing the spreading process [5,6]. This consideration is crucial in the study of contagion and epidemics on social networks, where local properties like degree and clustering can mediate nontrivial dynamics such as super-spreader events and endemic states [6,7].

Recent approaches propose the inclusion of interactions along higher-order topologies such as simplices (e.g., triangles and tetrahedrons) to account for social interactions that include more individuals than pairwise interactions [5,8,9]. The inclusion of higher-order interactions has been shown to promote critical transitions and bistability in epidemics [3,10–12]. However, the importance that they play in altering

the overall epidemic trajectory remains an active area of research. Current work has also explored the effect of simplicial contagion processes on infection paths on hypergraphs [13]. Several analytical methods have also been developed to better understand the governing factors of unique simplicial contagion behaviours such as hysteresis, bistability, and reduced epidemic thresholds [14,15].

In this work, we present a discrete numerical model that generalizes the classical SIS/SIR compartmental model with homogeneous mixing to account for hyperedge interactions of arbitrary order. Comparing dynamics with and without higher-order interactions, we find that topology primarily affects epidemic transients. If the network structure is known, one can calculate normalization values related to the potential network activity. These values may be used to define disease parameters for first-order pairwise interactions that replicate the steady state dynamics of those that include higher-order interactions, but are insufficient for reproducing epidemic transients. However, we find a duality where the effects of network topology can be overcome by allowing for temporally varying disease parameters, allowing first-order dynamics to closely approximate both transient and steady-state trajectories of higher-order systems.

In addition to requiring knowledge of the order scaling of interaction dynamics, the normalization approach makes two assumptions regarding the hypergraph topology. Namely, that (1) all hypergraphs are defined as clique complexes, and

^{*}Also at The Complex Systems Group, The University of Western Australia.

[†]Contact author: eugene.tan@thekids.org.au

[‡]Contact author: michael.small@uwa.edu.au

[§]Contact author: shannon.algar@uwa.edu.au

(2) macroscale topological features are uniformly distributed across the network. To investigate the robustness of the results, tests are conducted on cases where the scaling function is incorrect. We show robustness to model misspecification provided the size scales of the maximal infection amplification between the true and misspecified scaling functions are of similar order. These results are broadly aligned with the recent analytical results by [16] where mean-field approximations of pairwise interactions were able to capture features of simplicial contagion. We also extend analyses to artificial and empirical networks that do not adhere to Assumptions (1) and (2) and show robustness of results when disease parameters are allowed to temporally vary. Further inspection of the network structure and component hyperedge activities suggests that nontrivial macroscale heterogeneities in hypergraph topologies limit the approximation of higher-order dynamics by pairwise interactions.

II. DISEASE MODEL

We begin by presenting a variation of the classical compartmental ODE model [17], focusing on the three-compartment case with susceptible-infected-removed (SIR) states,

$$\frac{dS}{dt} = -\beta SI + \mu I, \quad (1a)$$

$$\frac{dI}{dt} = \beta SI - \mu I - \alpha I, \quad (1b)$$

$$\frac{dR}{dt} = \alpha I, \quad (1c)$$

where $N = S + I + R$ is the total population, and S, I, R are the susceptible, infected, and removed populations, respectively. The constants β , μ , and α are disease parameters corresponding to the rates of infection, recovery, and removal, respectively. For $\alpha = 0$, the SIR system reduces to the simpler SIS system.

The simple SIR model assumes that populations homogeneously mix. However, this is not true for human populations where disease spread occurs on a social contact network. Here, individuals—represented as nodes—interact within a small neighborhood with occasional branching, clustering, or long-range connections [5, 18, 19]. One can simulate more realistic spread dynamics by employing an agent-based approach where randomly selected nodes interact with all or a collection of their neighbors every time step [20]. Whilst this node-based approach is more realistic for modeling the diffusion of abstract quantities such as information, it is not necessarily adequate for disease contagion as it treats individuals rather than interactions as facilitators of infection. It is unlikely for an individual to interact with all of their neighbors in a given time step. Instead, we propose simulating diseases on networks using an edge-based approach.

Let $G = (E, V)$ be a hypergraph consisting of a set of hyperedges E and nodes V . A k -th order hyperedge is a connection between $k + 1$ different nodes. Each simulation time step consists of m iterations. A random hyperedge $e \in E$ is selected in each iteration. Each susceptible node s_i in e has

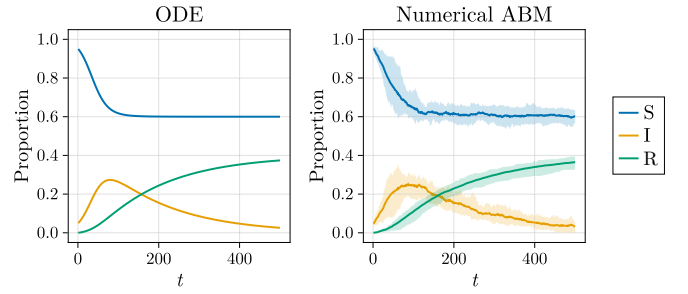


FIG. 1. Epidemic trajectories from the ODE model assuming homogeneous mixing, and 20 simulations of the numerical approximation using an $N = 500$ node fully connected network. Mean and 90% quantiles shown. Parameters are $\beta_{\text{ODE}} = 0.0003$, $\mu_{\text{ODE}} = 0.5N\beta_{\text{ODE}}$, $\alpha_{\text{ODE}} = 0.04N\beta_{\text{ODE}}$.

a probability β_e of getting infected and transitioning to state I ,

$$\beta_e = \min \left[f \left(\frac{n_I(e)}{n_S(e) - 1} \right) n_I(e) \beta_{\text{ABM}}, 1 \right], \quad (2)$$

where $n_I(e)$ and $n_S(e)$ are the number of nodes in e that are infected and susceptible respectively, β_{ABM} is a base infection probability for a pairwise interaction, and f is a scaling function that controls the extent to which multiple infected nodes in a given hyperedge amplify the probability of infection. In simplicial social contagion models (SCM) where interactions occur up to order $k = 2$, susceptible nodes whose neighbors are all infected experience an amplified probability of infection [5]. Whilst SCM has only been applied to social dynamics (e.g., knowledge spreading), we argue that a similar principle is also relevant when studying diseases where the infection probability may be mediated by the concentration of disease particles within an individual's immediate vicinity during an interaction. The allowance for variable infectivity has been explored previously by Fu *et al.* [21]. Inspired by this, we use the following scaling function to describe the disease spread dynamics on hyperedges:

$$f(p) = \begin{cases} 2p, & 0 \leq p < 0.5 \\ 1 + 2k_y(2p - 1), & p \geq 0.5 \end{cases}, \quad (3)$$

where k_y is the maximum amount that a hyperedge interaction can amplify the infection probability. After interaction, all infected nodes recover or die with probabilities $P(I \rightarrow S) = \mu_{\text{ABM}}$ and $P(I \rightarrow R) = \alpha_{\text{ABM}}$.

The proposed numerical model achieves two goals. First, it unifies the classical homogeneous ODE disease model with the existing SCM up to order $k_{\text{max}} = 2$. For $k_{\text{max}} = 1$ and a fully connected graph G , this model is a numerical approximation for the SIR/SIS model using the following disease parameters:

$$\beta_{\text{ABM}} = \frac{N(N-1)}{2} \beta_{\text{ODE}} \delta t, \quad (4a)$$

$$\mu_{\text{ABM}} = \mu_{\text{ODE}} \delta t, \quad (4b)$$

$$\alpha_{\text{ABM}} = \alpha_{\text{ODE}} \delta t, \quad (4c)$$

where $\delta t = 1/m$ is the integration time step. We validate simulations against the analytical ODE model and find good agreement in epidemic trajectories (see Fig. 1). For $k_{\text{max}} = 2$, the model is equivalent to an edge-based formulation of the

simplicial contagion model (SCM) by Iacopini *et al.* [5], with the conversion

$$k_\gamma = \frac{\beta + \beta_\Delta}{2\beta}, \quad (5)$$

where β_Δ is the simplicial infection rate. Second, this generalizes simplicial contagion to account for simplices of arbitrary order and varying proportions of infected individuals within each hyperedge.

III. NETWORK ACTIVITIES

Identifying dynamical differences between networks with and without interactions on higher-order topologies requires an appropriate normalization that allows for comparison. Thus, we consider two candidate quantities related to the rate of new infection $\frac{dI}{dt}$.

The first quantity, called the combinatorial network activity $\hat{\lambda}^{(K)}$ calculated up to order K for a hypergraph G with respect to a baseline infection rate β , is defined as

$$\hat{\lambda}^{(K)}(G, \beta) = \frac{\sum_{k=1}^K N_k \hat{\beta}^{(k)}}{\sum_{k=1}^K N_k}, \quad (6)$$

where

$$\hat{\beta}^{(k)} = \frac{\sum_{i=0}^{k+1} \binom{k+1}{i} (k+1-i) \min[f(\frac{i}{k}) i \beta, 1]}{\sum_{i=0}^{k+1} \binom{k+1}{i}}, \quad (7)$$

and approximates $\frac{dI}{dt}$ under the assumption that individuals are equally likely to be infected or susceptible. Therefore, $\hat{\lambda}^{(K)}$ depends only on the network topology and the scaling function f .

Similarly, an exact network activity $\lambda^{(K)}$ up to order K can be calculated as

$$\lambda^{(K)}(G, \beta, X) = \frac{\sum_{e \in E_K} n_S(e) \beta_e}{|E_K|}, \quad (8)$$

where $E^{(K)}$ is the set of all hyperedges of order up to $k_{\max} = K$, and X are the nodes' states. Unlike $\hat{\lambda}^{(K)}$, $\lambda^{(K)}$ does not assume a distribution of node states and is instead a function of the full network state and approximates the instantaneous value of $\frac{dI}{dt}$.

IV. RESULTS

A. Initial network activity normalization

To identify the role of higher-order interactions in disease spreading dynamics, we pose the following question: Can pairwise interactions replicate higher-order dynamics using an appropriate normalization?

We address this question using the case of SIS dynamics with baseline disease parameters $(\beta, \mu, \alpha) = (0.05, 0.0001, 0)$. The SIS case is analyzed first as it possesses steady states on which measures can be defined. We later extend our analyses to the more general SIR case.

Given an undirected N node network with a pairwise adjacency matrix A , we define two different hypergraphs $G_1 = (E_1, V)$ and $G_K = (E_K, V)$ with hyperedges defined up to the order of 1 and a K clique complex of A , respectively. Parallel simulations with the same initial conditions are conducted

using disease parameters (β_1, μ, α) and (β_K, μ, α) with infection values normalized using the ratio of both networks' combinatorial network activity,

$$\beta_1 = \frac{\hat{\lambda}^{(K)}(G_K, \beta)}{\hat{\lambda}^{(1)}(G_1, \beta)} \beta_K, \quad (9a)$$

$$\beta_K = \beta. \quad (9b)$$

This normalization ensures that simulations on both G_1 and G_K begin with approximately similar network activity. Thus, deviations in the epidemic trajectories of both simulations are the result of topology rather than disease parameters.

This test is run on two $N = 500$ node networks with different topologies: a 2D triangular lattice and an Erdős-Rényi (ER) random network with connection probability $p = 0.1$. The triangular lattice represents the case where the probability of higher-order interactions is greatest (i.e., $\frac{|E_K|}{|E_1|}$ is maximised) for $K = 2$. The random network is the case that allows for multiple orders of interaction, with the maximum order chosen to be $K = 4$. We run 50 and 10 simulations of the triangular lattice and random network, respectively, with a resolution of $m = 50$ (see Fig. 2).

A base case with unnormalized infection rates is run as a control. Enabling higher-order interactions after the pairwise system settles pushes the steady state align with the higher-order scenario. For the test case, applying a normalization based on the ratio of each network's combinatorial network activity results in epidemic trajectories whose steady state in the $K = 1$ case closely mimics those of $K > 1$. However, the initial adjustment of disease parameters is insufficient for replicating epidemic transients in earlier portions of the trajectory. These findings suggest that network topology and higher-order interactions are critical during epidemic transients, but wash out in the endemic steady state. In the early stages of the epidemic, infection spread is primarily mediated by the local network structures in the neighborhood of infected agents, which in turn govern the slopes of the trajectory. Once steady state is achieved, pairwise first-order $K = 1$ interactions, given an appropriate normalization, are sufficient to reproduce spread dynamics.

B. Adaptive network activity normalization

The capacity for networks with the same initial activity to reproduce epidemic steady states naturally leads the discussion to a second related question: Under what minimal conditions can pairwise interactions be used to replicate the *full dynamics* as those of higher-order? We again perform similar tests consisting of two parallel simulations with separate disease parameters, with the exception that the infection rate of the $K = 1$ case is allowed to vary over time, with parameters $(\beta_1(t), \mu, \alpha)$ and (β_K, μ, α) . This is plausible in cases where disease strains evolve, or intervention policies are enforced (i.e., dynamic network structures, altered infection parameters).

For simulating adaptive disease parameters, we first calculate the moving average of the exact network activity ratios in both hypergraphs,

$$\xi(t) = \frac{\lambda^{(K)}(G_K, \beta, X_K(t))}{\lambda^{(1)}(G_1, \beta_1(t), X_1(t))}, \quad (10a)$$

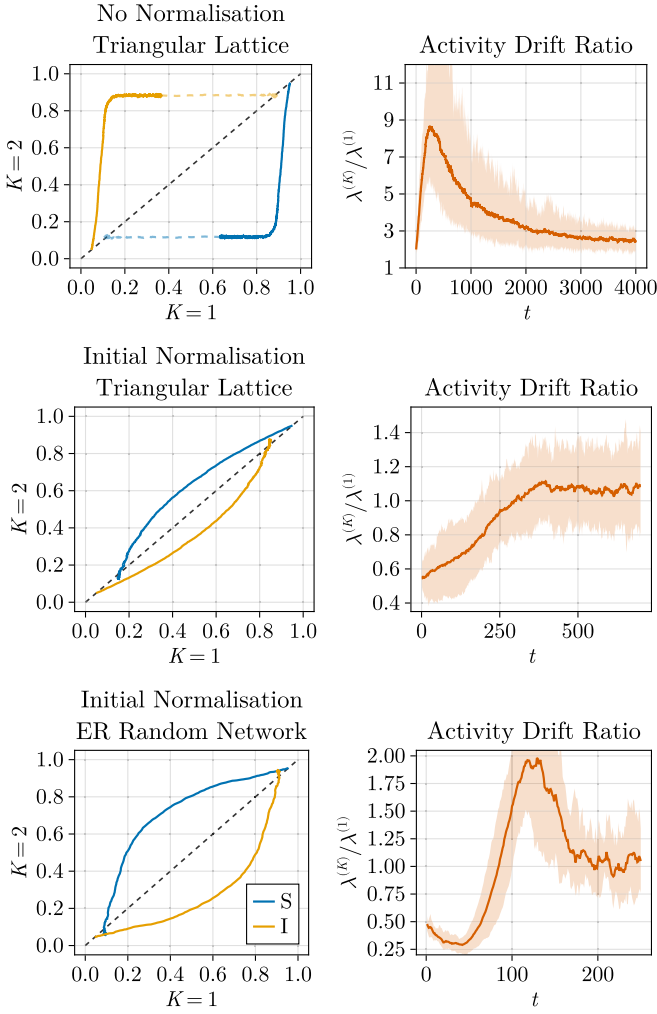


FIG. 2. SI epidemic trajectories between the pairwise and higher-order simulations. Accompanying activity ratios are shown on the right, revealing a saturation to 1 as the epidemic reaches the endemic steady state. Dotted trajectory in the unnormalized case corresponds to enabling higher-order interactions in pairwise simulation.

$$\bar{\xi}(t) = \frac{1}{\tau} \sum_{n=t}^{t+\tau} \xi(n), \quad (10b)$$

where $\tau = 5$ is a set window length. The pairwise infection rate parameter $\beta_1(t)$ is dynamically adjusted with respect to a conservative set threshold $\rho = 0.5$,

$$\beta_1(t) = \begin{cases} \beta_1(t-1), & |\bar{\xi}(t) - 1| < \rho \\ \bar{\xi}(t)\beta_1(t-1), & |\bar{\xi}(t) - 1| \geq \rho \end{cases}. \quad (11)$$

The adjustment of $\beta_1(t)$ based on a threshold of the moving average $\bar{\xi}(t)$ limits the extent to which the infection parameter is allowed to vary, and thus aims to mimic a more realistic evolution of disease dynamics where base infection rates are less volatile.

Results for the normalization tests with dynamically varying disease parameters are shown in Fig. 3. We find that the dynamical adjustment of disease parameters allows for first-order (pairwise) simulations to replicate those of higher order in both transient and steady-state regimes. Furthermore, we

note that the adjustments of disease parameters are infrequent, with two to four adjustment events occurring across 700 simulated time steps in each simulation. A majority of adjustments occur during the first half of the transient regime. These results suggest a duality between network topology and disease parameters. Namely, the effects of higher-order interactions may be partially accounted for in first-order approximations through the presence of dynamic disease parameters.

C. SIR dynamics

All of the previous tests focus on the SIS disease scenario, which is characterized by an epidemic transient followed by a steady-state regime where the disease is either eradicated or remains in endemic. A more interesting, and arguably challenging case, is that of SIR, where infection trajectories exhibit a single peak during the epidemic's transient phase.

We extend our analyses to this case and test the dynamic adjustment of infection rates applied across a range of base disease parameter values $\beta_K \in [0.004, 0.08]$ and $\mu_K \in [0.000025, 0.0005]$ with maximum order $K = 4$ and $k_\gamma = 3$. Phase plots of the peak time, and peak and final proportions are given in Fig. 4. Overall, we find that normalizations using activity ratios enable higher-order interaction dynamics to be replicated by those of lower order.

D. Robustness and model misspecification

While the thresholds imposed for the adaptive adjustment of disease parameters are relatively mild, the network activity normalization method requires extensive knowledge about the scaling of spreading dynamics up to the k -th order hyperedge. This is not easily defined in practice, and direct application will almost certainly contain errors in the specification of the scaling function $f(p)$. To test the robustness of our findings, the same numerical experiment is repeated using a misspecified scaling function $\hat{f}(p)$ when calculating activity ratios ξ . We test two types of misspecification: scale $\hat{f}_1(p)$, and functional form $\hat{f}_2(p)$ given by

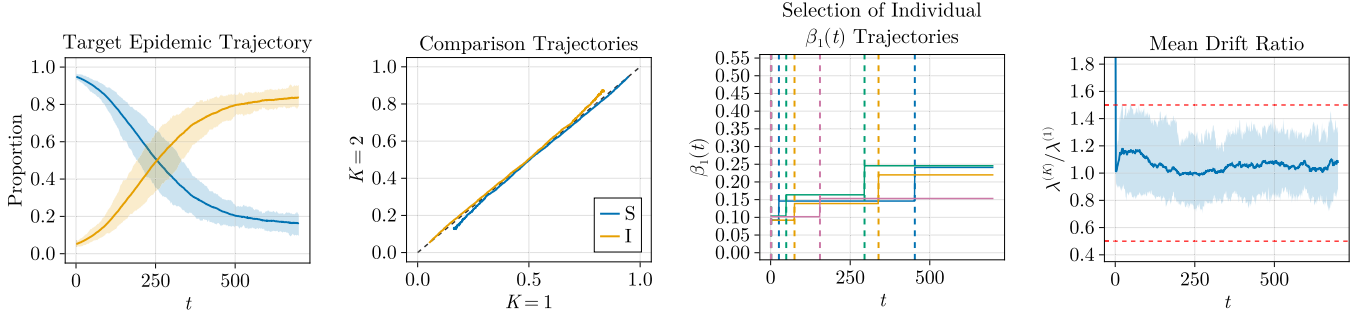
$$\hat{f}_1(p) = \begin{cases} 2p, & 0 \leq p < 0.5 \\ 1 + 2k_\gamma\eta(2p - 1), & p \geq 0.5 \end{cases}, \quad (12a)$$

$$\hat{f}_2(p) = \begin{cases} (2p)^{1/n}, & 0 \leq p < 0.5 \\ 1 + 2k_\gamma(2p - 1)^n, & p \geq 0.5 \end{cases}, \quad (12b)$$

The parameter η controls the misspecification in the amplification infection rates due to higher-order interactions, and n affects the degree of nonlinearity from the true scaling function $f(p)$ (see Fig. 5).

We simulate disease dynamics for trajectory pairs of $K = 1$ and $K = 4$ on a 500-node ER random network for $\eta \in \{0.5, 0.8, 1.2, 1.5, 2\}$ and $n \in \{2, 3, 4, 5\}$. In both cases, dynamic normalization of infection rates based on network activity ratios calculated with \hat{f}_1, \hat{f}_2 remained sufficient for pairwise interactions to replicate trajectories from simulations that include higher-order interactions (see Fig. 6). We observe that scale errors are more detrimental to the accuracy of the normalization errors.

Triangular Lattice



ER Random Network

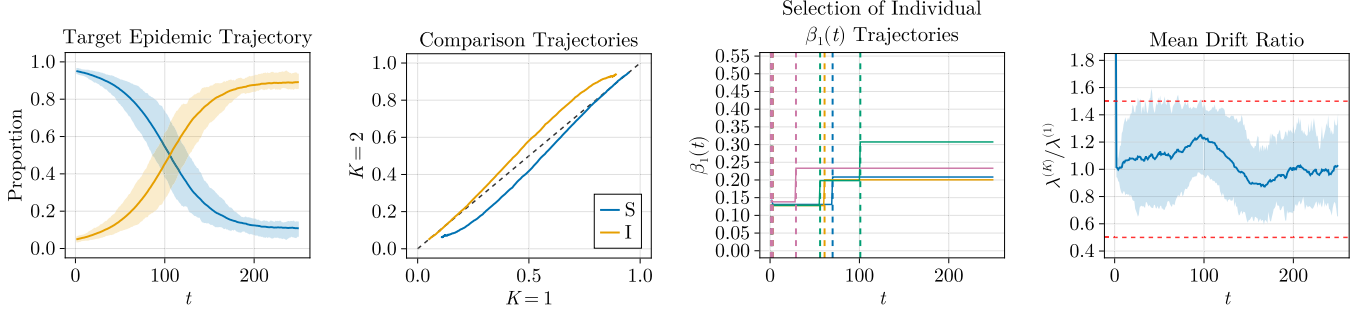


FIG. 3. Simulations with dynamic disease parameters with $K = 1$ for triangular lattice and random networks. Shaded regions correspond to 90% CI. Left to right: target trajectories from the higher order simulation, comparison trajectories between pairwise and higher-order cases, four randomly chosen trajectories of $\beta_1(t)$ with adjustment events indicated by vertical lines, and the mean drift ratio $\bar{\xi}(t)$, with the ρ threshold envelope in red.

E. Topological effects

1. Network topology assumptions

All of the cases analyzed so far make two key assumptions pertaining to the network topology on which disease dynamics occur: (1) All hypergraphs are defined as clique complexes and (2) macroscale topological features are uniformly distributed across the network. We first elaborate further on the importance and meaning of these two characteristics, and present additional tests on their impact on the pairwise to higher-order relationship.

The assumption that all hypergraphs belong to a family of clique complexes of some maximal order K places constraints on the existence of hyperedges in the network. Namely, the existence of a given hyperedge e of order K between a set $v(e) = \{v_i\}_{i=1}^{K+1}$ of $K + 1$ nodes in the network implies that all hyperedges of order $k < K$ consisting of a subset of nodes from $v(e)$ also exist. If this condition is met, interactions on hyperedges present the same connectivity as the pairwise case, with the only difference being an acceleration of the disease dynamics. While computationally convenient, such an assumption is not necessarily true in real social networks. As

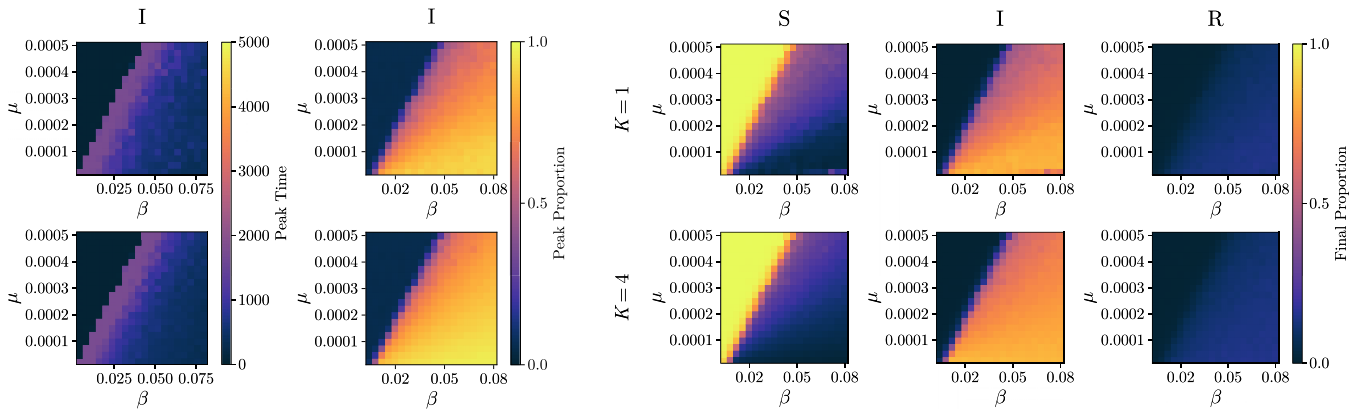


FIG. 4. Phase plots of the peak time, peak, and final proportions for the simulated SIR model with dynamic disease parameters. The top and bottom rows correspond to the pairwise $K = 1$ and higher-order $K = 4$ simulations for $N = 500$ node ER random network. Dynamic adjustment shows good replication of phase plots and the epidemic outbreak threshold.

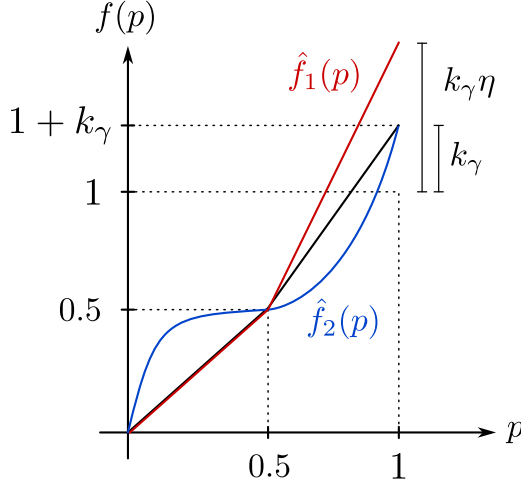


FIG. 5. Schematic of scaling function $f(p)$ and tested misspecified forms $\hat{f}_1(p)$ and $\hat{f}_2(p)$.

an illustrative example, one can consider a sports team consisting of $K + 1$ players who regularly engage in close contact as a group by virtue of their vocation. However, subgroups of players may never interact outside this group setting.

For the second assumption, we clarify “macroscale topological features” to include the various centrality measures of a given node and “uniformly distributed” to mean that the occurrence of a locally significant topological feature (e.g., hubs and high degree) can occur on any node in the network. Therefore, this assumption describes that a given node’s connectivity is not conditioned on its neighbors’ local topology.

This assumption can be described more generally by considering the deviation of local properties from their neighborhood. For networks, let $\mu(x_i)$ be some measure (in the statistical sense) of a node’s property for the i -th node. The deviation in measure μ between a node and its neighbors is given by

$$\Delta\mu(x_i) = \mu(x_i) - \frac{1}{|\mathcal{N}_m(i)|} \sum_{j \in \mathcal{N}_m(i)} \mu(x_j),$$

where \mathcal{N}_m is the set of neighbors for node i with a distance less than or equal to m . Large values of $\Delta\mu(x_i)$ imply that

the property of a node differs substantially from that of its neighborhood.

For example, in the case of small-world networks—such as those generated using the Watts-Strogatz algorithm [22]—shortcut nodes may arise anywhere in the network. For a random Erdős-Rényi network, local heterogeneity—nodes with abnormally high degrees—have equal probability to occur anywhere within the network. In terms of node deviations, this implies

$$P(\Delta\mu(x_i) = \Delta\mu^*),$$

where $\Delta\mu^*$ is some constant quantity that is independent of i .

One can also consider large deviations in terms of distributions. For networks with small deviations, given a local network measure of interest $\mu(x_i)$, and an embedding of the network into a space $\{x_i\}_{i=1}^N \rightarrow \mathcal{M} \subseteq \mathbb{R}^n$, the probability of observing $P(\mu(x_i) = \mu^*)$ in a randomly generated network is uniformly distributed in \mathcal{M} .

Examples that do not possess this characteristic are those that arise from preferential-attachment mechanisms such as the Barabási-Albert (BA) scale-free network model [23]. In this case, high-degree nodes (i.e., hubs) will always remain close together in the network due to the preferential attachment mechanism, and thus result in a highly localized distribution of $P(\mu)$ in \mathcal{M} .

2. Pathological network cases

One point of interest is to test the robustness of the network activity normalization methods on network topologies that do not adhere well to the abovementioned topological assumptions. For this, we consider both artificial randomly generated networks and two different empirical networks taken from contact data that exhibit nontrivial local topologies. For the artificial case, we choose to study a randomly generated BA scale-free $N = 300$ node network with additional edges added using the standard ER random network algorithm with varying probability $p \in \{0, 0.005, 0.02\}$, where $p = 0$ corresponds to the standard BA scale-free model of preferential attachment. Increasing values of p dilute the scale-free property of the network, resulting in a topology that better adheres to Assumption (2).

To demonstrate application to real data, two empirical networks describing human contact were taken from existing publicly available resources. The first is a network of Australian politicians with edges representing mutually liked Facebook pages from November 2017 [24]. The second is a fictional contact network derived from conflicts present in the entire collection of movies from the Marvel Cinematic Universe (MCU) from the Aleph Zero Heroes dataset compiled by Roughan *et al.* [25]. Visualizations of both networks are given in Fig. 7. Both cases showcase features suggestive of preferential attachment with a high-degree center and tree-like or low-degree exteriors.

In addition to examining the impact of base network structures, we also test the importance of the clique complex Assumption (1) by randomly removing hyperedges such that higher-order hyperedges no longer contain all their subcliques and thus breaking any implied downward closure

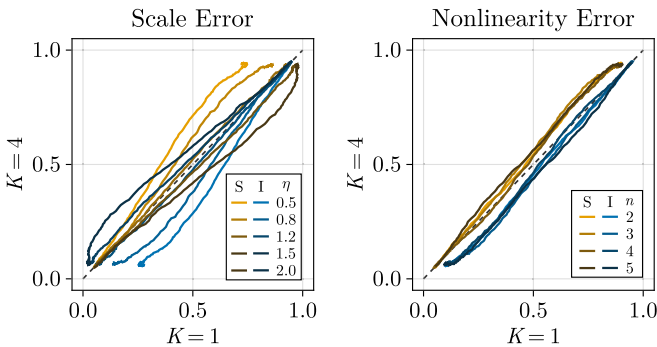


FIG. 6. Comparison trajectories showing robust performance across misspecification due to scale and functional form, with the former having more detrimental effects on performance.

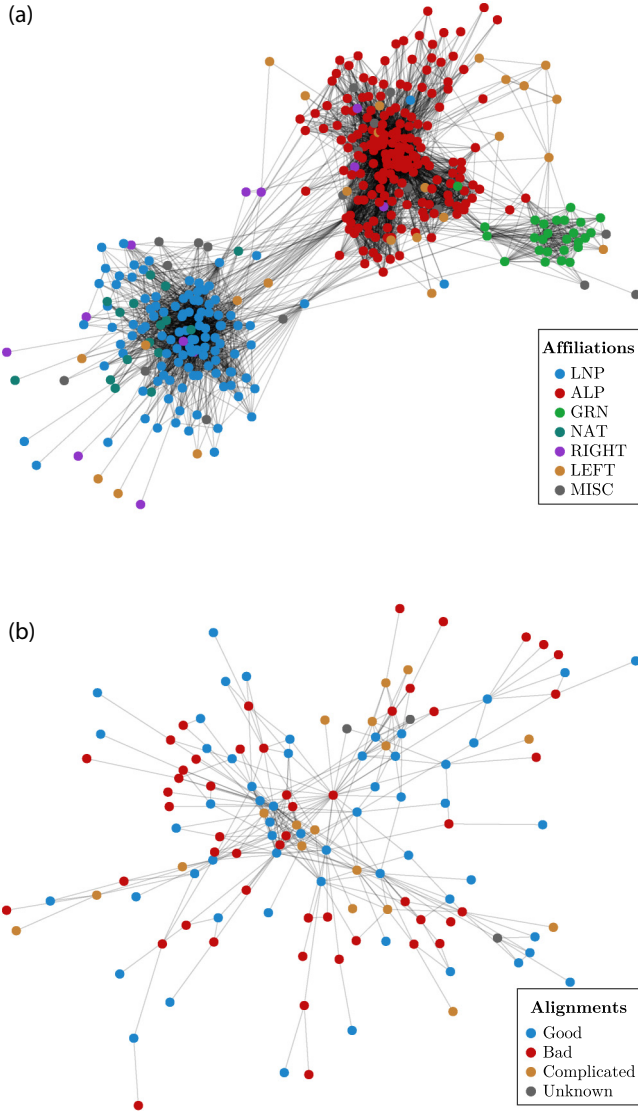


FIG. 7. Visualization of tested empirical networks with associated node classifications shown. Politician affiliations correspond to either Australian political parties or general political alignment. Hero alignment corresponds to the classification of an MCU hero as either a protagonist or antagonist. (a) Australian politicians Facebook interactions and (b) marvel cinematic universe (MCU) conflicts.

of higher-order interactions [26]. This is done using the following pruning algorithm:

(1) Given an undirected pairwise adjacency matrix A_0 , construct its clique complex hypergraph G_K up to order K .

(2) Let $k = K$ be the starting maximal order. For each k -th order hyperedge e_k , delete e_k with probability p_d .

(a) If no deletion occurred, continue to the next hyperedge.

(b) If deletion occurred, identify all child clique complexes of order $k - 1 > q > 1$ and recursively repeat for all subsequent children.

(c) If a child is a hyperedge of order $q = 1$, then only allow the deletion of the edge if the pairwise network remains connected.

(3) Repeat step 2 with decreasing starting order $k \leftarrow k - 1$ until $k = 1$.

An important feature of the above edge deletion algorithm is that lower-order hyperedges experience more deletion events due to the recursive structure of the algorithm. For low to moderate deletion probabilities p_d , this results in a skewing of the proportion of hyperedges toward those of higher order.

3. Topological robustness

We repeat similar tests as those outlined in Secs. IV A and IV B across the five previously mentioned networks, and $p_d \in (0, 0.6)$. Five different normalization settings are tested, consisting of the standard initial normalization using the combinatorial network activity $\hat{\lambda}$, adaptive normalization of the exact network activity with decreasing thresholds $\rho \in \{0.5, 0.2, 0.1\}$, and a related variant of the combinatorial network activity weighted according to the hyperdegree of network nodes. Results are shown in Fig. 8.

In all tested networks, we find that a single initial normalization of disease parameters using the combinatorial network activity $\hat{\lambda}$ is negatively affected by the highly localised “macroscale” structures for all networks, such as the clustered groups of hubs. Furthermore, as the influence of higher-order interactions increases (i.e., higher p_d), the validity of this normalization also fails.

To better understand the cause for this failure, a variant of the combinatorial network activity is used as an alternative normalization. In this case, a surrogate network state is constructed where half of all nodes are assigned an infected state I with priority ranking based on their hyperdegree. The network activity is then calculated for the surrogate network state to yield an approximate network activity $\lambda^*(p_I)$, which is subsequently used for normalization. In all cases, this method was found to correspond to better agreement in the steady-state values of the epidemic across all networks.

The superior performance of the hyperdegree weighted network activity is due to a closer resemblance between the surrogate network state and the likely true network state, where higher-degree nodes will tend to have a higher probability of being infected. This can be seen by looking at the distribution of activities associated with each hyperedge in the combinatorial, degree weighted, and true exact network activity cases at a given infection proportion p_I (see Fig. 9). Notably, the combinatorial approach, which inherently assumes a binomial distribution ($p = 0.5$) for infections, and results in infected nodes that are uniformly spread throughout the entire network. However, node states from contagion dynamics have a tendency to bias infections toward nodes with high degree, which is well replicated by the degree weighting approach.

Interestingly, we note that the degree-weighted approach still fails to properly replicate epidemic transients. One may consider a naïve extension of the degree-weighted network activity that is normalized based on the true current proportions of infected p_I in both the pairwise and higher-order cases. This method is attractive as it potentially allows for an approximation of the true network activity without requiring knowledge of the full state. However, we find that such an approach fails due to the presence of highly local macroscale structures that exacerbate the deviation $|\lambda^*(p_I) - \lambda|$ as $p_I \rightarrow 1$. This deviation can be partially attributed to the random

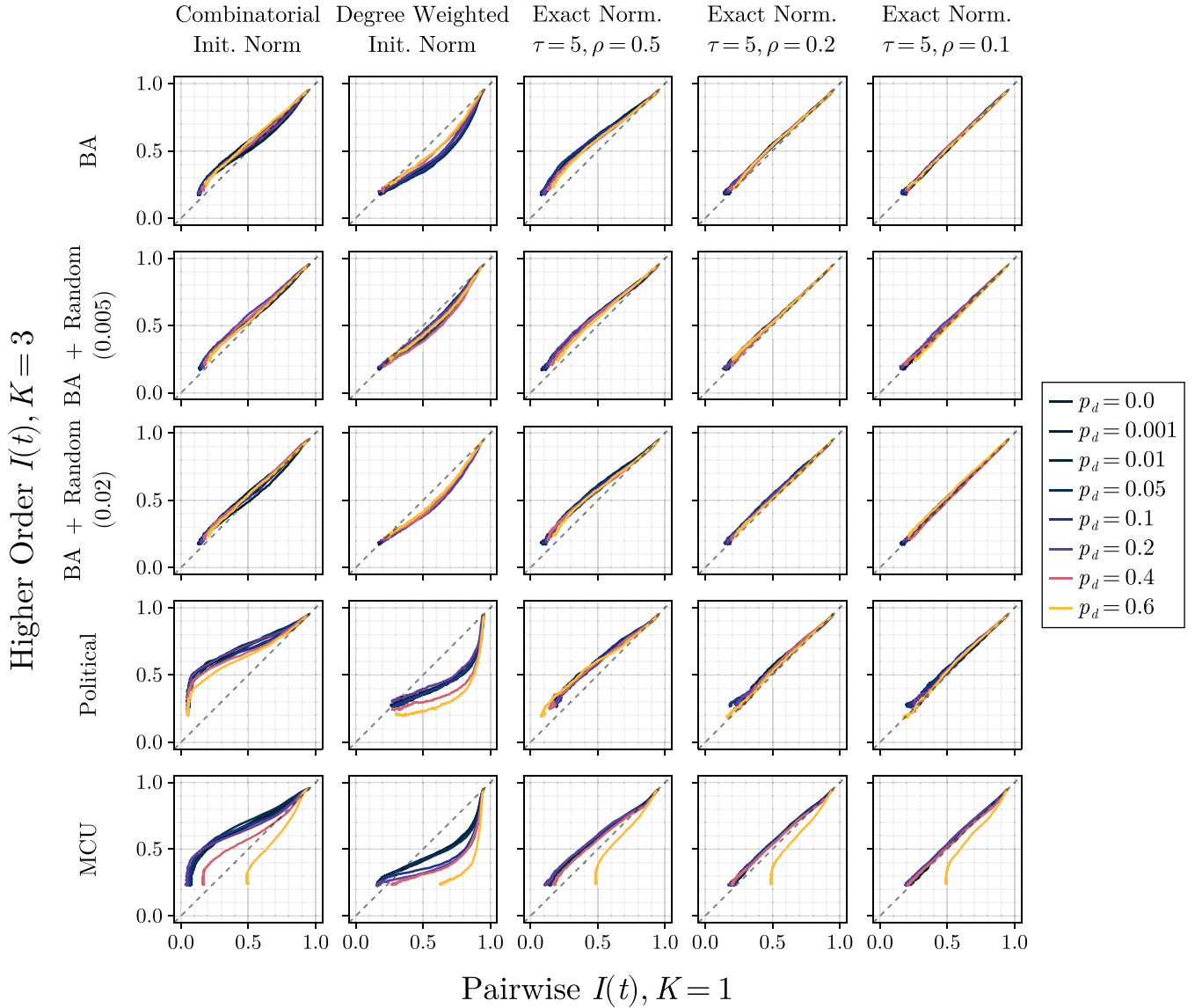


FIG. 8. Comparison of pairwise vs higher-order simulation trajectories for five tested networks with varying edge deletion probabilities p_d .

recovery of individual nodes within the high-degree clusters of the network, whose hyperedges boost the underlying network activity.

As seen in Fig. 8, the exact activity normalization methods performed well across all tested networks and for moderately high levels of edge deletion, with the exception of $p_d = 0.6$ in the case of the MCU network. This performance is improved for increasingly stringent thresholds ρ . However, we note that the performance of the exact method is lower than when applied to the triangular and random networks tested in Sec. IV B. This suggests that, whilst topologically robust, network activity has a nontrivial relationship with network topology.

V. CONCLUSION

In this work, we quantify the effect of higher-order topologies on the spreading dynamics on networks. Focusing on the single disease case, we have presented an agent-based model

that unifies the classical compartmental SIR model of homogeneous mixing with pairwise simplicial contagion models on networks, and extends it to account for interactions on hyperedges of arbitrary order. We use this model in conjunction with a normalization method based on the notion of network activity to compare spread dynamics across hypergraphs of different maximal orders.

We first consider the restricted case of clique complexes with uniformly distributed macroscale topological features for both SIS and SIR dynamics. The inclusion of higher-order topology is found to primarily affect the transient dynamics in epidemic trajectories. Using carefully chosen disease parameters such that initial combinatorial network activities are approximately equal, simulations restricted to pairwise (order 1) interactions are able to replicate the steady-state behaviours resulting from higher-order interactions. Furthermore, a sufficiently accurate normalization can be calculated purely based on the network topology, network state, and the scaling function describing the higher-order interaction. We

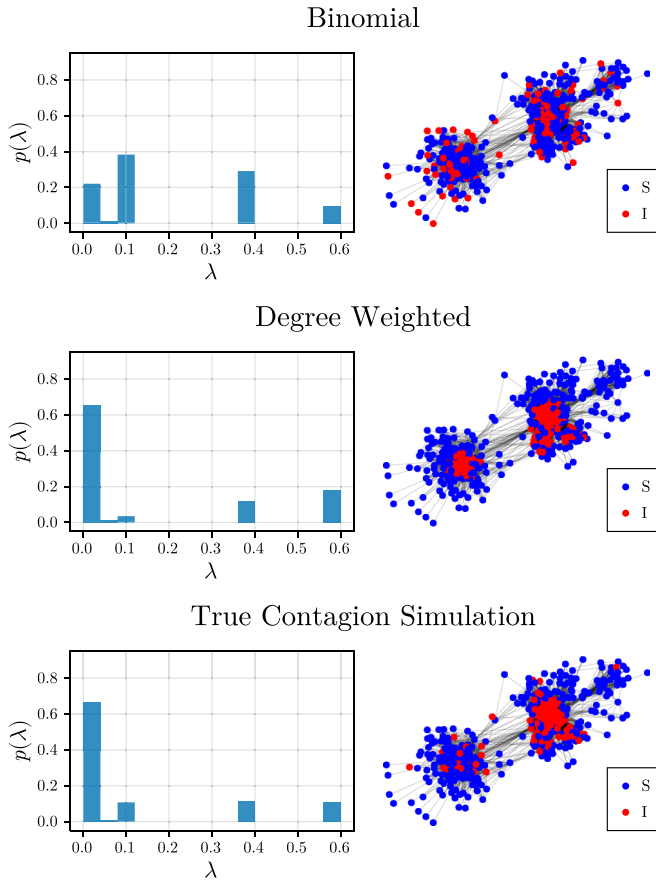


FIG. 9. Frequency distributions of network edge activities assuming an infected proportion of $p_I = 0.3$ for the combinatorial (binomial), degree weighted, and true contagion simulation. Surrogate and exact network states are shown on the right. A degree-weighted network state surrogate results in a distribution of network activity that more closely resembles the true network activity.

find that allowing disease parameters to dynamically vary over time is sufficient for pairwise simulations to reproduce full trajectories of higher-order interactions. Thus, we conclude that there is a duality between topological effects and the stationarity of disease parameters.

The proposed normalization method based on network activities makes several assumptions, such as having knowledge of the scaling function, and having a clique complex structure with relatively mild macroscale heterogeneities. While the findings are theoretically interesting, these assumptions impose heavy constraints on applicability as real-world systems rarely fulfill these conditions. To this end, we test the robustness of the network activity approach toward model misspecification and non-clique hypergraphs with nontrivial topology taken from empirical data. Overall, all results are relatively robust to model misspecification due to scale and functional form errors. In contrast, we find that network topology plays a larger role where networks with macroscale heterogeneities—such as social networks or those arising from preferential attachment mechanisms—where higher-order interactions are not as easily or robustly approximated by pairwise interactions. This result is particularly true for cases where disease parameters are kept static.

The ability for pairwise interaction models to reproduce epidemic trajectories from models containing higher-order complexity features raises interesting questions on the need for higher-order features to be included in disease models. Furthermore, it encourages a reconsideration of whether dynamics, observed or modeled, should be attributed as the result of complex topological structure or temporally varying spreading mechanisms. The above tests focus on epidemic trajectories, which represent summaries of the disease averaged across the entire network. These results invite further questions on the similarity of spreading dynamics on the local scale of a given network and whether the order of infection events between individuals differs between interactions of different orders. One can also consider the effect of incubation and immunity refractory periods, which may impart a form of time delay in the infection process, resulting in more dynamic disease parameters. On application, network activities offer a potential way to construct adaptive rewiring strategies that can be used to maintain a sustainable level of infection.

DATA AVAILABILITY

The data that support the findings of this article are openly available [27].

- [1] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Epidemic processes in complex networks, *Rev. Mod. Phys.* **87**, 925 (2015).
- [2] M. Jalili and M. Perc, Information cascades in complex networks, *J. Complex. Netw.* **5**, 665 (2017).
- [3] G. Ferraz de Arruda, A. Aleta, and Y. Moreno, Contagion dynamics on higher-order networks, *Nat. Rev. Phys.* **6**, 468 (2024).
- [4] D. J. Daley and D. G. Kendall, Epidemics and rumours, *Nature (London)* **204**, 1118 (1964).
- [5] I. Iacopini, G. Petri, A. Barrat, and V. Latora, Simplicial models of social contagion, *Nat. Commun.* **10**, 2485 (2019).
- [6] M. D. Shirley and S. P. Rushton, The impacts of network topology on disease spread, *Ecol. Complex.* **2**, 287 (2005).
- [7] M. Small, C. K. Tse, and D. M. Walker, Super-spreaders and the rate of transmission of the SARS virus, *Physica D* **215**, 146 (2006).
- [8] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, Networks beyond pairwise interactions: Structure and dynamics, *Phys. Rep.* **874**, 1 (2020).
- [9] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno, *et al.*, The physics of higher-order interactions in complex systems, *Nat. Phys.* **17**, 1093 (2021).
- [10] L. Zhao, H. Wang, H. Yang, C. Gu, and J. M. Moore, Susceptible-infected-recovered-susceptible processes competing on simplicial complexes, *Phys. Rev. E* **110**, 064311 (2024).

- [11] M. Lucas, I. Iacopini, T. Robiglio, A. Barrat, and G. Petri, Simplicially driven simple contagion, *Phys. Rev. Res.* **5**, 013201 (2023).
- [12] Á. Bodó, G. Y. Katona, and P. L. Simon, SIS epidemic propagation on hypergraphs, *Bull. Math. Biol.* **78**, 713 (2016).
- [13] D. A. Contreras, G. Cencetti, and A. Barrat, Infection patterns in simple and complex contagion processes on networks, *PLoS Comput. Biol.* **20**, e1012206 (2024).
- [14] I. Z. Kiss, C. Bick, and P. L. Simon, Decoding how higher-order network interactions shape contagion dynamics, *J. Math. Biol.* **91**, 29 (2025).
- [15] Y. Zhao, C. Li, D. Shi, G. Chen, and X. Li, Higher order epidemic spreading in simplicial networks, *IEEE Trans. Syst. Man Cybern.: Syst.* **55**, 7714 (2025).
- [16] F. Malizia, L. Gallo, M. Frasca, I. Z. Kiss, V. Latora, and G. Russo, A pair-based approximation for simplicial contagion, *Chaos Solitons Fractals* **199**, 116776 (2025).
- [17] W. O. Kermack and A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. R. Soc. London, Ser. A* **115**, 700 (1927).
- [18] S. Boccaletti, P. De Lellis, C. Del Genio, K. Alfaro-Bittner, R. Criado, S. Jalan, and M. Romance, The structure and dynamics of networks with higher order interactions, *Phys. Rep.* **1018**, 1 (2023).
- [19] I. Z. Kiss, J. C. Miller, and P. L. Simon, *Mathematics of Epidemics on Networks: From Exact to Approximate Models*, (Springer, New York, 2017), Vol. 46.
- [20] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Dynamical and correlation properties of the internet, *Phys. Rev. Lett.* **87**, 258701 (2001).
- [21] X. Fu, M. Small, D. M. Walker, and H. Zhang, Epidemic dynamics on scale-free networks with piecewise linear infectivity and immunization, *Phys. Rev. E* **77**, 036113 (2008).
- [22] D. J. Watts and S. H. Strogatz, Collective dynamics of ‘small-world’ networks, *Nature (London)* **393**, 440 (1998).
- [23] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *Science* **286**, 509 (1999).
- [24] R. A. Rossi and N. K. Ahmed, The network data repository with interactive graph analytics and visualization, in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (AAAI Press, Austin, Texas, 2015), pp. 4292–4293.
- [25] M. Roughan, L. Mitchell, and T. South, How the Avengers assemble: Ecological modeling of effective cast sizes for movies, *PLoS ONE* **15**, e0223833 (2020).
- [26] L. Torres, A. S. Blevins, D. Bassett, and T. Eliassi-Rad, The why, how, and when of representations for complex systems, *SIAM Rev.* **63**, 435 (2021).
- [27] E. Tan, M. Small, and S. D. Algar, DiseaseHypergraphs (Accompanying data) (2025), <https://github.com/eugenetkj98/DiseaseHypergraphs>.