Triadic Approximation Reveals the Role of Interaction Overlap on the Spread of Complex Contagions on Higher-Order Networks

Giulio Burgio, Sergio Gómez, and Alex Arenas

Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain

Contagion processes relying on the exposure to multiple sources are prevalent in social systems, and are effectively represented by hypergraphs. In this Letter, we derive a mean-field model that goes beyond node- and pair-based approximations. We reveal how the stability of the contagion-free state is decided by either two- or three-body interactions, and how this is strictly related to the degree of overlap between these interactions. Our findings demonstrate the dual effect of increased overlap: it lowers the invasion threshold, yet produces smaller outbreaks. Corroborated by numerical simulations, our results emphasize the significance of the chosen representation in describing a higher-order process.

Behaviors, strategies, and conventions often require some form of reinforcement for their adoption [1–6]. Spreading among individuals through learning and imitation, their diffusion can be studied as a contagion process [7–10]. Differently from epidemics, however, repeated exposures to the same contagious source (e.g., an individual with a given behavior) are not always sufficient for transmission and multiple sources are required, defining a complex contagion [1,4,5,11–13]. When transmission depends on the exposures being simultaneous, as when interacting in a group, a spreading event becomes a many-body interaction [14]. Similarly, in biochemical systems, oftentimes a species needs simultaneous exposure to one or more other species in order for the reaction to occur [15,16].

Here we focus on social contagion [1,3,10,17–19]. This can be mapped to a susceptible-infectious-susceptible (SIS) process on a hypergraph [14,20,21], a generalization of a graph where an n-edge (i.e., an edge incident on n nodes) is used to represent an n-body interaction [22]. We assume that a susceptible node (individual) becomes infected at rate $\beta^{(1)}$ in a two-body interaction (2-edge) with an infected individual $(S + I \xrightarrow{\beta^{(1)}} 2I)$, at rate $\beta^{(2)}$ in a three-body interaction (3-edge) with two infected individuals $(S + 2I \xrightarrow{\beta^{(2)}} 3I)$, and so on for larger groups. Infected individuals recover at rate μ $(I \xrightarrow{\mu} S)$. Within this higher-order setting, Iacopini *et al.* [14] found a phenomenology (the appearance of a saddle-node bifurcation implying critical-mass behavior) that is effectively equivalent to the one already uncovered by Dodds *et al.* [3]. The latter used a threshold model (where transmission generally requires multiple exposures) over two-body interactions only. One might thus believe that accounting for higher-order, group interaction is only a marginal refinement, if not an unnecessary complication.

In this Letter, we challenge this belief by demonstrating that the outcome of the contagion process is fundamentally linked to how interactions of different orders are arranged in the system. Having developed a clique-based mean-field model that accounts for local dynamical correlations, we reveal that, contrary to the predictions made by node-based approximations [14,20], the invasion threshold, $\beta_{\mathrm{cr}}^{(1)}$, at which the inactive (contagion-free) state becomes unstable, does depend on $\beta^{(2)}$ [21]. This dependence is proven to be strictly related to the degree of overlap between three- and two-body interactions. Having derived an explicit expression for the critical surface, we demonstrate that the overlap has a double-edged effect: it lowers the invasion threshold, but also makes the outbreaks generally smaller.

Let us start from the closure approximation we apply to the exact microscopic equations on hypergraphs. We track the state evolution of subsets of nodes which form maximal cliques (i.e., cliques not subsets of larger ones) in the projection graph constructed by associating cliques to edges of the hypergraph. Accordingly, considering up to three-body interactions, we account for the evolution of the state probability $P_i^{\sigma_i}$ for node i to be in state σ_i, $P_{ij}^{\sigma_i \sigma_j}$ for the maximal link ij to be in state $\sigma_i \sigma_j$, $P_{ijl}^{\sigma_i \sigma_j \sigma_l}$ for the (maximal) 3-clique ijl to be in state $\sigma_i \sigma_j \sigma_l$. Notice that a 3-clique, when projected back to the hypergraph, comes in one of three flavors: a length-3 cycle (or 3-cycle), conveying three two-body interactions, a 3-edge, conveying a three-body interaction, or a 2-simplex (or triangle), conveying all of them.

The state probability of other local structures is approximated in terms of the maximal cliques composing it. We consider random hypergraphs that are sparse to the extent that the probability for two maximal cliques to share more than one node vanishes in the infinite-size limit [23]. We thus need a closure only for the following local structures:

two connected maximal links, a maximal link connected to a 3-clique, and two connected 3-cliques. We approximate their state as follows [21]:

$$P_{i\underline{j}l}^{\sigma_i\sigma_j\sigma_l} \approx P_{ij}^{\sigma_i\sigma_j} P_{jl}^{\sigma_j\sigma_l}/P_j^{\sigma_j}, \tag{1a}$$

$$P_{ij\underline{l}h}^{\sigma_i\sigma_j\sigma_l\sigma_h} \approx P_{ij}^{\sigma_i\sigma_j} P_{jlh}^{\sigma_j\sigma_l\sigma_h}/P_j^{\sigma_j}, \tag{1b}$$

$$P_{ij\underline{l}hk}^{\sigma_i\sigma_j\sigma_l\sigma_h\sigma_k} \approx P_{ijl}^{\sigma_i\sigma_j\sigma_l} P_{lhk}^{\sigma_l\sigma_h\sigma_k}/P_l^{\sigma_l}, \tag{1c}$$

where the underline indicates the shared node. We refer to Eqs. (1) as the *triadic approximation* [24].

The higher-order interaction structure is encoded in the following binary tensors: $A^{(1)}$, such that $A_{ij}^{(1)} = 1$ if the maximal link ij exists; $A^{(1,0)}$ and $A^{(0,1)}$, such that $A_{ijl}^{(1,0)} = 1$

($A_{ijl}^{(1,0)} = 0$) and $A_{ijl}^{(0,1)} = 0$ ($A_{ijl}^{(0,1)} = 1$) if ijl is a 3-cycle (3-edge); and such that $A_{ijl}^{(1,0)}A_{ijl}^{(0,1)} = 1$ if ijl is a triangle (for later convenience, we introduce $A^{(1,1)} = A^{(1,0)} \odot A^{(0,1)}$). Specifically, if for any 3-clique ijl, $A_{ijl}^{(0,1)} = 1 \Rightarrow A_{ijl}^{(1,0)} = 1$, the hypergraph is a simplicial 2-complex, for the existence of a 3-edge implies the existence of the 2-edges it includes [26]. If, instead, $A_{ijl}^{(0,1)} = 1 \Rightarrow A_{ijl}^{(1,0)} = 0$, it is a linear hypergraph, for any two edges will share at most one node [26]. Any hypergraph is located in between these two limits, depending on the degree of overlap between three- and two-body interactions.

Having rescaled time by μ, the process is described by the following system of microscopic equations,

$$\dot{P}_i^I = -P_i^I + \beta^{(1)}\sum_j A_{ij}^{(1)} P_{ij}^{SI} + \frac{1}{2}\sum_{j,l}\left[A_{ijl}^{(1,0)}\beta^{(1)}(P_{ijl}^{SSI} + P_{ijl}^{SIS} + 2P_{ijl}^{SII}) + A_{ijl}^{(0,1)}\beta^{(2)} P_{ijl}^{SII}\right], \tag{2a}$$

$$\dot{P}_{ij}^{SI} = -(1 + \beta^{(1)})P_{ij}^{SI} + P_{ij}^{II} - \beta^{(1)}\sum_{l\neq j}A_{il}^{(1)} P_{ji\underline{l}}^{ISI} + \beta^{(1)}\sum_{l\neq i}A_{jl}^{(1)} P_{ij\underline{l}}^{SSI}$$
$$- \frac{1}{2}\sum_{l,h}\left[A_{ilh}^{(1,0)}\beta^{(1)}(P_{ji\underline{l}h}^{ISIS} + P_{ji\underline{l}h}^{ISSI} + 2P_{ji\underline{l}h}^{ISII}) + A_{ilh}^{(0,1)}\beta^{(2)} P_{ji\underline{l}h}^{ISII}\right] + \{i \leftrightarrow j\}, \tag{2b}$$

$$\dot{P}_{ijl}^{SSI} = -(1 + 2A_{ijl}^{(1,0)}\beta^{(1)})P_{ijl}^{SSI} + P_{ijl}^{ISI} + P_{ijl}^{SII}$$
$$- \beta^{(1)}\sum_{h\neq j,l}A_{ih}^{(1)} P_{jl\underline{i}h}^{SISI} - \beta^{(1)}\sum_{h\neq i,l}A_{jh}^{(1)} P_{il\underline{j}h}^{SISI} + \beta^{(1)}\sum_{h\neq i,j}A_{lh}^{(1)} P_{ij\underline{l}h}^{SSSI}$$
$$- \frac{1}{2}\sum_{h,k\neq j,l}\left[A_{ihk}^{(1,0)}\beta^{(1)}(P_{jl\underline{i}hk}^{SISIS} + P_{jl\underline{i}hk}^{SISSI} + 2P_{jl\underline{i}hk}^{SISII}) + A_{ihk}^{(0,1)}\beta^{(2)} P_{jl\underline{i}hk}^{SISII}\right] - \{i \leftrightarrow j\} + \{i \leftrightarrow l\}, \tag{2c}$$

$$\dot{P}_{ijl}^{SII} = -(2 + 2A_{ijl}^{(1,0)}\beta^{(1)} + A_{ijl}^{(0,1)}\beta^{(2)})P_{ijl}^{SII} + A_{ijl}^{(1,0)}\beta^{(1)}(P_{ijl}^{SSI} + P_{ijl}^{SIS}) + P_{ijl}^{III}$$
$$- \beta^{(1)}\sum_{h\neq j,l}A_{ih}^{(1)} P_{jl\underline{i}h}^{IISI} + \beta^{(1)}\sum_{h\neq i,l}A_{jh}^{(1)} P_{il\underline{j}h}^{SISI} + \beta^{(1)}\sum_{h\neq i,j}A_{lh}^{(1)} P_{ij\underline{l}h}^{SISI}$$
$$- \frac{1}{2}\sum_{h,k\neq j,l}\left[A_{ihk}^{(1,0)}\beta^{(1)}(P_{jl\underline{i}hk}^{IISIS} + P_{jl\underline{i}hk}^{IISSI} + 2P_{jl\underline{i}hk}^{IISII}) + A_{ihk}^{(0,1)}\beta^{(2)} P_{jl\underline{i}hk}^{IISII}\right] + \{i \leftrightarrow j\} + \{i \leftrightarrow l\}, \tag{2d}$$

where $\{i \leftrightarrow j\}$ denotes that obtained by swapping i and j in the explicit term (excluding the sign in front) on the same line and taking i and j in state S. The other state probabilities are found as $P_i^S = 1 - P_i^I$, $P_{ij}^{SS} = 1 - P_i^I - P_{ij}^{SI}$, $P_{ij}^{II} = P_i^I - P_{ij}^{IS}$, $P_{ijl}^{SSS} = 1 - P_i^I - P_{ijl}^{SII} - P_{ijl}^{SSI} - P_{ijl}^{SIS}$, and $P_{ijl}^{III} = P_i^I - P_{ijl}^{ISI} - P_{ijl}^{IIS} - P_{ijl}^{ISS}$. Equations (2) are closed through Eqs. (1). The system consists then of $N + L + 2(T^{(1,0)} + T^{(0,1)} + T^{(1,1)})$ equations, being $N, L, T^{(1,0)}, T^{(0,1)}$, and $T^{(1,1)}$, the number of nodes, maximal links, 3-cycles, 3-edges, and triangles, respectively.

To make this model analytically tractable, we perform a mean-field approximation by regarding all the nodes and cliques as equivalent to their average counterparts. Accordingly, every node is assumed to be part of the same number of maximal links $k^{(1)}$—3-cliques, $k^{(1,0)}$; 3-edges, $k^{(0,1)}$; and triangles, $k^{(1,1)}$—and thus participates in $\kappa^{(1)} = k^{(1)} + 2(k^{(1,0)} + k^{(1,1)})$ two-body interactions and $\kappa^{(2)} = k^{(0,1)} + k^{(1,1)}$ three-body interactions (see Fig. 1 for illustration). The state probabilities P_i^σ, $P_{ij}^{\sigma\sigma'}$, and $P_{ijl}^{\sigma\sigma'\sigma''}$, with $\sigma, \sigma', \sigma'' \in \{S, I\}$, are taken equal to their respective

averages, $P^\sigma = \sum_i P_i^\sigma / N$, $P^{\sigma\sigma'} = \sum_{i,j} A_{ij}^{(1)} P_{ij}^{\sigma\sigma'} / Nk^{(1)}$, and $P_x^{\sigma\sigma'\sigma''} = \sum_{i,j,l} A_{ijl}^x P_{ijl}^{\sigma\sigma'\sigma''} / 2Nk^x$, the index $x \in \{(1,0),$ $(0,1),(1,1)\}$ indicating the type of the considered 3-clique [27].

Using the indicator function $\mathbb{1}_p$, giving 1 if condition p is fulfilled and 0 otherwise, the reduced system reads

$$\dot{P}^I = -P^I + \beta^{(1)}k^{(1)}P^{SI} + 2\beta^{(1)}\left[k^{(1,0)}(P_{(1,0)}^{SSI} + P_{(1,0)}^{SII}) + k^{(1,1)}(P_{(1,1)}^{SSI} + P_{(1,1)}^{SII})\right] + \beta^{(2)}\left[k^{(0,1)}P_{(0,1)}^{SII} + k^{(1,1)}P_{(1,1)}^{SII}\right], \quad (3a)$$

$$\dot{P}^{SI} = -(1+\beta^{(1)})P^{SI} + P^{II} - \beta^{(1)}(k^{(1)}-1)P^{SI}\frac{P^{SI}-P^{SS}}{P^S}$$

$$- \left\{2\beta^{(1)}\left[k^{(1,0)}(P_{(1,0)}^{SSI} + P_{(1,0)}^{SII}) + k^{(1,1)}(P_{(1,1)}^{SSI} + P_{(1,1)}^{SII})\right] + \beta^{(2)}\left[k^{(0,1)}P_{(0,1)}^{SII} + k^{(1,1)}P_{(1,1)}^{SII}\right]\right\}\frac{P^{SI}-P^{SS}}{P^S}, \quad (3b)$$

$$\dot{P}_x^{SSI} = -2(1+\beta^{(1)}\mathbb{1}_{x\neq(0,1)})P_x^{SSI} + 2P_x^{SII} - \beta^{(1)}k^{(1)}P^{SI}\frac{2P_x^{SSI}-P_x^{SSS}}{P^S}$$

$$- 2\beta^{(1)}\left[(k^{(1,0)}-\mathbb{1}_{x=(1,0)})(P_{(1,0)}^{SSI}+P_{(1,0)}^{SII}) + (k^{(1,1)}-\mathbb{1}_{x=(1,1)})(P_{(1,1)}^{SSI}+P_{(1,1)}^{SII})\right]\frac{2P_x^{SSI}-P_x^{SSS}}{P^S}$$

$$- \beta^{(2)}\left[(k^{(0,1)}-\mathbb{1}_{x=(0,1)})P_{(0,1)}^{SII} + (k^{(1,1)}-\mathbb{1}_{x=(1,1)})P_{(1,1)}^{SII}\right]\frac{2P_x^{SSI}-P_x^{SSS}}{P^S}, \quad (3c)$$

$$\dot{P}_x^{SII} = -(2+2\beta^{(1)}\mathbb{1}_{x\neq(0,1)} + \beta^{(2)}\mathbb{1}_{x\neq(1,0)})P_x^{SII} + 2\beta^{(1)}\mathbb{1}_{x\neq(0,1)}P_x^{SSI} + P_x^{III} - \beta^{(1)}k^{(1)}P^{SI}\frac{P_x^{SII}-2P_x^{SSI}}{P^S}$$

$$- 2\beta^{(1)}\left[(k^{(1,0)}-\mathbb{1}_{x=(1,0)})(P_{(1,0)}^{SSI}+P_{(1,0)}^{SII}) + (k^{(1,1)}-\mathbb{1}_{x=(1,1)})(P_{(1,1)}^{SSI}+P_{(1,1)}^{SII})\right]\frac{P_x^{SII}-2P_x^{SSI}}{P^S}$$

$$- \beta^{(2)}\left[(k^{(0,1)}-\mathbb{1}_{x=(0,1)})P_{(0,1)}^{SII} + (k^{(1,1)}-\mathbb{1}_{x=(1,1)})P_{(1,1)}^{SII}\right]\frac{P_x^{SII}-2P_x^{SSI}}{P^S}, \quad (3d)$$

where $P^S = 1 - P^I$, $P^{SS} = 1 - P^I - P^{SI}$, $P^{II} = P^I - P^{SI}$, $P^{SSS} = 1 - P^I - P^{SII} - 2P^{SSI}$, $P^{III} = P^I - P^{SSI} - 2P^{SII}$.

To correctly locate the phase transition, we linearize Eqs. (3) around the inactive state by regarding of the same order $\epsilon \ll 1$ the probabilities of infected states, i.e., $P^I, P^{SI}, P^{II}, P_x^{SSI}, P_x^{SII}, P^{III} \in \mathcal{O}(\epsilon)$ [21]. The rightmost eigenvalue of the Jacobian matrix associated with the resulting linear system crosses the imaginary axis when the following is satisfied [28]:

$$k^{(1)}\frac{\beta^{(1)}}{1+\beta^{(1)}} + k^{(1,0)}\frac{2\beta^{(1)}(1+\beta^{(1)})}{1+2\beta^{(1)}(1+\beta^{(1)})}$$

$$+ k^{(1,1)}\frac{\beta^{(1)}(2+2\beta^{(1)}+\beta^{(2)})}{1+\beta^{(1)}(2+2\beta^{(1)}+\beta^{(2)})} = 1. \quad (4)$$

two 3-cycles, two 3-edges   one 3-cycle, one 3-edge, one triangle   two triangles

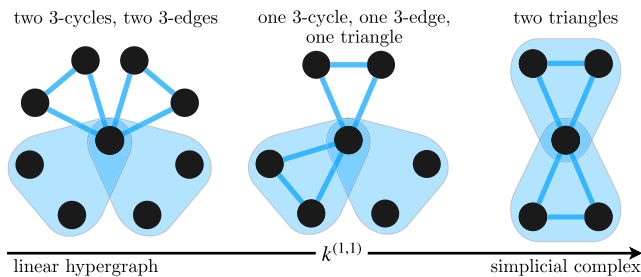linear hypergraph $k^{(1,1)}$ simplicial complex

FIG. 1. Example showing how the neighborhood of a focal node changes with $k^{(1,1)}$ for fixed $\kappa^{(1)}$, $k^{(1)}$ (zero here), and $\kappa^{(2)}$. The node takes part to $\kappa^{(1)} = 4$ two-body and $\kappa^{(2)} = 2$ three-body interactions, but their degree of overlap changes.

Equation (4) defines the critical surface in the parameter space [see Fig. 2(a)]. There is no solution when either $\beta^{(1)} = 0$ or there are only three-body interactions ($\kappa^{(2)} = k^{(0,1)}$, $\kappa^{(1)} = 0$). This reveals that a three-body interaction cannot affect the stability of the inactive state unless "activated" by the presence of two-body interactions within the same subset of nodes. Two-body interactions are thus needed to destabilize the inactive state. That activation occurs in triangles. Since each term in the l.h.s. of Eq. (4) is a strictly increasing function of the infection rates, given $k^{(1,1)} > 0$, the larger is $\beta^{(2)}$ ($\beta^{(1)}$), the smaller is $\beta^{(1)} = \beta_{\mathrm{cr}}^{(1)}$ ($\beta^{(2)} = \beta_{\mathrm{cr}}^{(2)}$) solving Eq. (4). In particular, when triangles can percolate the structure (i.e., $k^{(1,1)} > 1$), $\beta_{\mathrm{cr}}^{(1)}$ can be made arbitrarily small by increasing $\beta^{(2)}$. Moreover, by imposing $\beta^{(2)} = 0$ in Eq. (4), we find that a simple contagion suffices to cause extensive spreads at $\beta_{\mathrm{cr}}^{(1)} \in [\beta_{(1)}^{(1)}, \beta_{(1,0)}^{(1)}]$, being $\beta_{(1)}^{(1)} = 1/(\kappa^{(1)}-1)$ and $\beta_{(1,0)}^{(1)} = [\sqrt{1+4/(\kappa^{(1)}-2)}-1]/2$ the critical point for, respectively, a locally treelike network ($\kappa^{(2)} = 0$, $\kappa^{(1)} = k^{(1)}$) and a 3-cycle-based network ($\kappa^{(2)} = 0$, $\kappa^{(1)} = 2k^{(1,0)}$). In agreement with previous studies proving clustering to raise the critical point of simple [29,30] and slightly nonlinear [31] contagions, $\beta_{(1,0)}^{(1)} > \beta_{(1)}^{(1)}$.

Considering then a triangle-based network ($\kappa^{(2)} = k^{(1,1)}$, $\kappa^{(1)} = 2k^{(1,1)}$), i.e., a homogeneous simplicial 2-complex, we find that the critical point, $\beta^{(1)}_{(1,1)}$, reads

$$\beta^{(1)}_{(1,1)} = \frac{\beta^{(2)} + 2}{4} \left[\sqrt{1 + \frac{16}{(\kappa^{(1)} - 2)(\beta^{(2)} + 2)^2}} - 1 \right]. \quad (5)$$

$\beta^{(1)}_{(1,1)}$ thus vanishes as $1/\beta^{(2)}$ for large $\beta^{(2)}$ [see Fig. 2(b)]. Observe that Eq. (5) reflects the fact that extensive contagions are possible only for $\kappa^{(1)} > 2$ ($k^{(1,1)} > 1$), when a giant connected component can exist.

To isolate the effect of the overlap between two- and three-body interactions, we fix $\kappa^{(1)}$, $k^{(1)}$, and $\kappa^{(2)}$, and increase $k^{(1,1)}$ from 0 to $\kappa^{(2)}$ (correspondingly, $k^{(1,0)}$ and $k^{(0,1)}$ both decrease). Importantly, a larger $k^{(1,1)}$ implies a smaller and more redundant neighborhood (see Fig. 1). One may thus expect that the critical point increases with $k^{(1,1)}$.

As shown in detail in Fig. 2(a), instead, either $\beta^{(1)}_{\text{cr}}$ and $\beta^{(2)}_{\text{cr}}$ decrease with $k^{(1,1)}$, taking the lowest values in a simplicial complex and the highest in a linear hypergraph (for which, being $k^{(1,1)} = 0$, $\beta^{(1)}_{\text{cr}}$ is unaffected by $\beta^{(2)}$). Since 3-edges yield a negligible contribution around the inactive state, exchanging them for triangles helps the spread to thrive. As Figs. 2(c) and 2(d) show, this holds also for the equilibrium fraction of infected nodes, $I^\star$, when $\beta^{(1)}$ is close enough to the threshold for the simplicial complex. For larger infection rates, however, that redundancy becomes detrimental, for potentially infectious edges lead to nodes which are already infected. The largest spreads are thus found for linear hypergraphs, ensuring the least-redundant, widest neighborhoods. Finally, notice in Fig. 2(c) how solely varying the overlap can change the nature of the phase transition.

We test the model on random regular hypergraphs, generated through a standard configuration model in which every node is assigned the same degrees $k^{(1)}$, $k^{(0,1)}$, $k^{(1,0)}$,
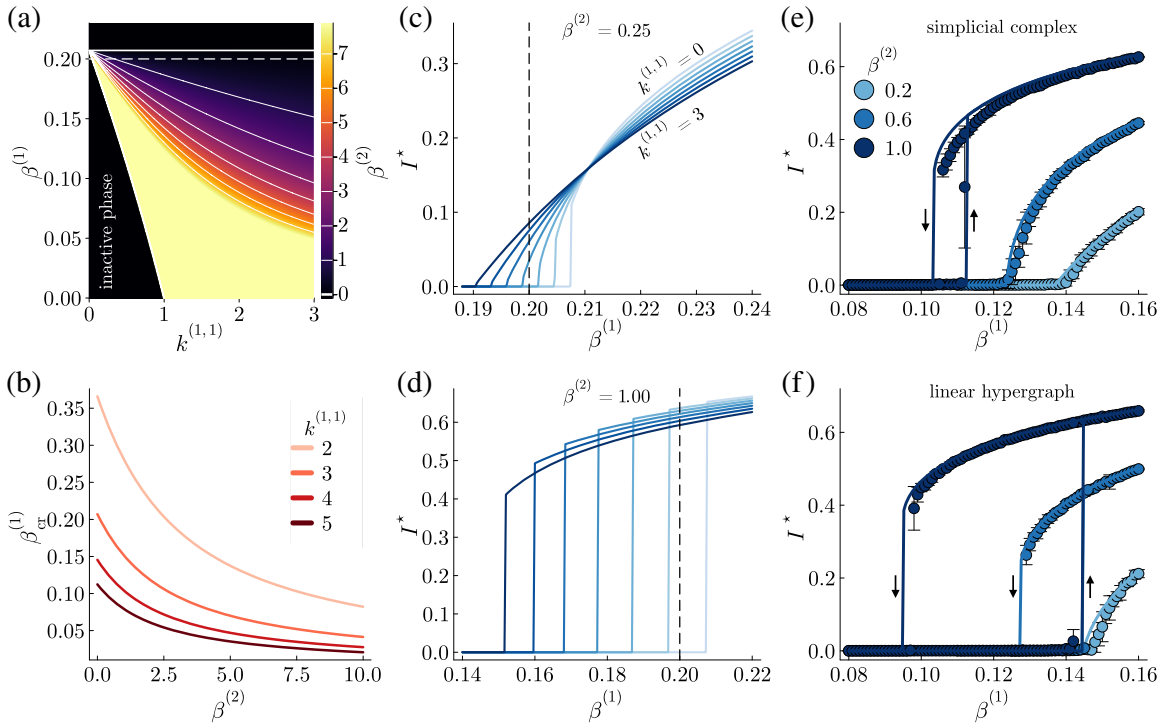

FIG. 2. Predictions from the mean-field model. (a) Critical surface $(k^{(1,1)}, \beta^{(1)}, \beta^{(2)})_{\text{cr}}$ defined by Eq. (4) for hypergraphs with $\kappa^{(1)} = 6$ two-body ($k^{(1)} = 0$) and $\kappa^{(2)} = 3$ three-body interactions per node. The white solid curves correspond to the values of $\beta^{(2)}$ indicated in the color bar (limited at 7 for better readability), the curve $\beta^{(2)} = 0$ being thicker. The dashed line denotes $\beta^{(1)}_{(1)} = 1/5$, associated to the locally treelike network with the same $\kappa^{(1)}$. (b) Critical threshold $\beta^{(1)}_{\text{cr}} \equiv \beta^{(1)}_{(1,1)}$, Eq. (5), for homogeneous simplicial 2-complexes. (c)–(d) Equilibrium prevalence, $I^\star$, for $\beta^{(2)} \in \{0.25, 1.00\}$, and $k^{(1,1)}$ from 0 (lightest shade) to 3 (darkest shade) in steps of 0.5. The dashed line indicates $\beta^{(1)}_{(1)} = 1/5$. (e)–(f) Comparison of the model (solid lines) with numerical simulations performed on random regular hypergraphs with $N = 5000$, $\kappa^{(1)} = 8$, $k^{(1)} = 2$, $\kappa^{(2)} = 3$, and $k^{(1,1)} = 3$ (simplicial complex) and $k^{(1,1)} = 0$ (linear hypergraph). Points and error bars denote averages and standard errors computed over 20 random initializations. The arrows help to distinguish the forward and backward curves in hysteresis cycles.

and $k^{(1,1)}$. As reported in Figs. 2(e) and 2(f), numerical simulations confirm the (quantitative) predictions made by the mean-field model.

We further test those predictions on hypergraphs constructed from real-world datasets, containing record of face-to-face interactions during a conference [32], and proximity data within a university campus [33]. We refer to the Supplemental Material [34] for the procedure used to convert each dataset into a binary network. The hypergraphs are then constructed by adding three-body interactions to either 3-cycles with probability h, converting them in triangles, or to randomly selected triplets of unconnected nodes otherwise, forming 3-edges. Even though the basic assumptions of homogeneity and sparseness that we made are heavily violated in more realistic structures (see [34]), the numerical results reported in Fig. 3 show that the qualitative phenomenology uncovered by the mean-field theory remains valid. This confirms its structural origin: the overlap between three- and two-body interactions. We thus conjecture that a similar picture holds for other contagion models as well, as we already verified for a SIR process (see Supplemental Material [34]).

Through a more refined mean-field model, this study reveals a fundamental relation between the behavior of complex contagion processes and the way interactions are arranged in the higher-order structure. Extending beyond node- and pair-based approximations, our analysis establishes how three-body interactions contribute to destabilizing the inactive state, proving their contribution is contingent on overlapping with two-body interactions. Examining the boundary structures, we demonstrated that simplicial complexes and linear hypergraphs—having maximal and no overlap, respectively—exert diametrically opposed dynamical effects. The former lower the critical point, while often resulting in smaller spreads; the latter heighten the critical point, yet typically leading to larger spreads. Complementing recent findings in synchronization [39], our investigation underscores the necessity of identifying the most suitable representation for specific higher-order processes.
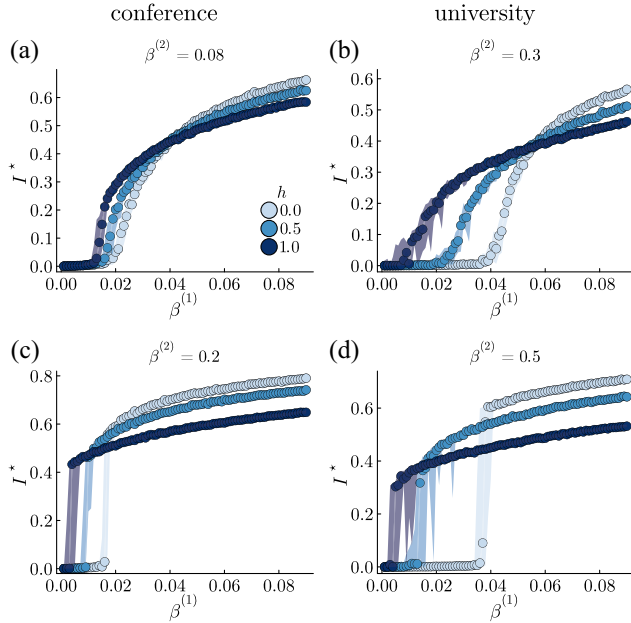
FIG. 3. Numerical simulations performed on the hypergraphs constructed from the conference's [32] [(a) and (c)] and the university campus's [33] [(b) and (d)] datasets. Points denote medians computed over 20 random initializations; ribbons cover from the 5th to the 95th percentile. The added three-body interactions form triangles with probability $h = 0$ (linear hypergraph), $h = 0.5$, and $h = 1$ (simplicial complex).

[1] M. Granovetter, Am. J. Sociology **83**, 1420 (1978).

[2] D. J. Watts, Proc. Natl. Acad. Sci. U.S.A. **99**, 5766 (2002).

[3] P. S. Dodds and D. J. Watts, Phys. Rev. Lett. **92**, 218701 (2004).

[4] D. Centola and M. Macy, Am. J. Sociology **113**, 702 (2007).

[5] D. Guilbeault, J. Becker, and D. Centola, Complex contagions: A decade in review, in *Complex Spreading Phenomena in Social Systems: Influence and Contagion in Real-World Social Networks* (Springer International Publishing, Cham, 2018), pp. 3–25.

[6] P. Romanczuk and B. C. Daniels, Phase transitions and criticality in the collective behavior of animals–self-organization and biological function, in *Order, Disorder and Criticality: Advanced Problems of Phase Transition Theory* (World Scientific, Singapore, 2023), pp. 179–208.

[7] W. Goffman and V. Newill, Nature (London) **204**, 225 (1964).

[8] D. J. Daley and D. G. Kendall, Nature (London) **204**, 1118 (1964).

[9] D. A. Levy and P. R. Nail, Genet. Soc. Gen. Psychol. Monogr. **119**, 233 (1993), http://europepmc.org/abstract/MED/8405969.

[10] A. L. Hill, D. G. Rand, M. A. Nowak, and N. A. Christakis, PLoS Comput. Biol. **6**, e1000968 (2010).

[11] D. Centola, Science **329**, 1194 (2010).

[12] D. J. O'Sullivan, G. J. O'Keeffe, P. G. Fennell, and J. P. Gleeson, Front. Phys. **3**, 71 (2015).

[13] D. Guilbeault and D. Centola, Nat. Commun. **12**, 4430 (2021).

[14] I. Iacopini, G. Petri, A. Barrat, and V. Latora, Nat. Commun. **10,** 2485 (2019).

[15] S. Klamt, U.-U. Haus, and F. Theis, PLoS Comput. Biol. **5,** e1000385 (2009).

[16] J. Jost and R. Mulas, Adv. Math. **351,** 870 (2019).

[17] G. Tarde, *The Laws of Imitation* (Henry Holt and Company, New York, 1890).

[18] H. Blumer, Collective behavior, in *Principles of Sociology*, edited by R. E. Park (Barnes & Noble, New York, 1939), pp. 219–288.

[19] G. Le Bon, *The Crowd: A Study of the Popular Mind* (Courier Corporation, 2002).

[20] N. W. Landry and J. G. Restrepo, Chaos **30,** 103117 (2020).

[21] G. Burgio, A. Arenas, S. Gómez, and J. T. Matamalas, Commun. Phys. **4,** 111 (2021).

[22] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, Phys. Rep. **874,** 1 (2020).

[23] The number of maximal edges of any cardinality must scale slower than quadratically with the system size.

[24] Notice that this closure becomes exact for non-recursive dynamics such as susceptible-infected-recovered (SIR) processes when the structure has no cycles of more than three nodes [25].

[25] I. Z. Kiss, J. C. Miller, P. L. Simon *et al.*, *Mathematics of Epidemics on Networks: From Exact to Approximate Models* (Springer, Cham, 2017).

[26] A. Bretto, *An Introduction. Mathematical Engineering* (Springer, Cham, 2013).

[27] Notice that permuting the superscripts σ, σ', and σ'', has no effect.

[28] This can be more readily found by imposing stationarity in Eq. (3) and solving the resulting homogeneous system.

[29] J. C. Miller, Phys. Rev. E **80,** 020901(R) (2009).

[30] L. Hébert-Dufresne, P.-A. Noël, V. Marceau, A. Allard, and L. J. Dubé, Phys. Rev. E **82,** 036115 (2010).

[31] L. A. Keating, J. P. Gleeson, and D. J. O'Sullivan, Phys. Rev. E **105,** 034306 (2022).

[32] M. Génois and A. Barrat, EPJ Data Sci. **7,** 1 (2018).

[33] P. Sapiezynski, A. Stopczynski, D. D. Lassen, and S. Lehmann, Sci. Data **6,** 315 (2019).

[34] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.132.077401 for the procedure used to generate the hypergraphs from real-world proximity data, for a detailed discussion about the quantitative limitations of the mean-field theory, and for the results found for a generalized SIR process. Refs. [35–38] are included.

[35] G. D'Agostino, A. Scala, V. Zlatić, and G. Caldarelli, Europhys. Lett. **97,** 68006 (2012).

[36] J. P. Gleeson, S. Melnik, J. A. Ward, M. A. Porter, and P. J. Mucha, Phys. Rev. E **85,** 026106 (2012).

[37] X. Lv, D. Fan, Q. Li, J. Wang, and L. Zhou, Physica (Amsterdam) **627A,** 129131 (2023).

[38] M. E. J. Newman, Phys. Rev. Lett. **89,** 208701 (2002).

[39] Y. Zhang, M. Lucas, and F. Battiston, Nat. Commun. **14,** 1605 (2023).

A Triadic Approximation Reveals the Role of Interaction Overlap on the Spread of Complex Contagions on Higher-Order Networks

Supplemental Material

Giulio Burgio, Sergio Gómez, and Alex Arenas

Departament d'Enginyeria Informàtica i Matemàtiques,
Universitat Rovira i Virgili, 43007 Tarragona, Spain

S1. Hypergraphs from real-world data

S1.1. Generation and properties

We first describe the procedure used to generate the hypergraphs used in Fig. 3 of the main text. We used two real-world dataset. One dataset contains face-to-face interactions recorded during a conference [3], the other one proximity data recorded within a university campus [8]. The procedure is the same for both datasets.

These consist of time-resolved interactions (each representing a face-to-face interaction or proximity between two people) which, once aggregated, yields very dense networks [3]. Therefore, we first build a static pairwise network where each edge is assigned a weight equal to its number of appearances (i.e., how many times the interaction between the two agents has been detected throughout the entire observation time) and then threshold it.

Starting from an empty network with only the nodes in the dataset ($N = 219$ for the conference's and $N = 672$ for the university campus'), edges are listed in decreasing order of weight and added to the network starting from the first one. Since some nodes only participate to edges with very low weight, waiting until all nodes are included would yield a network identical to the original one, except for just few missing edges. To avoid this, we stop including edges when the 95% of the nodes has been connected to some other node (notice that disconnected components may still exist at this point), indeed thresholding the original network[1]. The remaining degree-0 nodes are then connected to the other nodes at random. If the network is disconnected, the connected components are connected to the largest connected component by adding an edge at random between each of them and the latter. In practice, at the moment in which the 95% of the nodes is reached, there exists a component containing almost all nodes and very few other components of very few nodes. Consequently, the few edges added to connect the network do not affect the properties of the thresholded network.

The binary network obtained in this way represents the backbone to which we add three-body interactions in order to get rank-3 hypergraphs. To do this, we first list all the 3-cliques in the network[2]. Then, to each 3-clique, we add a three-body interaction (i.e., a 3-edge containing the three nodes) with probability h, such that, if the addition occurs, a 2-simplex (triangle)

---

[1]The original networks include many large cliques. Since our model assumes cliques of up to 3 nodes, when an edge is included as above we check whether a 4-clique formed, in which case the edge is ignored. Including 4-cliques or larger ones does not change qualitatively the results. On the other hand, avoiding them makes the network less dense and the phenomenology easier to appreciate.

[2]Notice that some 3-cliques share two nodes with other 3-cliques, meaning that those 2-cliques which are part of more than one 3-clique appear multiple times in the list. Each appearance is considered as a different interaction. An alternative method that would avoid repeated 2-cliques consists in finding an edge-disjoint edge clique cover of the network [1], where, for instance, the motif made of two 3-cliques sharing a 2-clique would be decomposed in (and considered as) one 3-clique (that includes the shared 2-clique) and two 2-cliques. Using this method just changes the number of 2-cliques and 3-cliques in the network, but the results remain qualitatively unaffected.

is formed. Otherwise (occurring with probability $1 - h$), the 3-edge is added to three uncon-nected nodes chosen at random, so that a three-body interaction not overlapped with two-body interactions is formed. Notice that the total number of three-body interactions added is inde-pendent from h; only their distribution over the system changes with it. Setting $h = 0$ yields a linear hypergraph. Increasing h, more and more frequently three-body interactions overlap with two-body interactions. At $h = 1$, the structure becomes a simplicial 2-complex.

Conference's dataset. The hypergraphs generated from this dataset consists of $N = 219$ nodes. The 2-degree $\kappa^{(1)}$ (number of 2-edges incident on a node) is distributed heterogeneously. The first and the second raw moments of the 2-degree distribution are $\langle \kappa^{(1)} \rangle \approx 33.05$ and $\langle \kappa^{(1)2} \rangle \approx 2034.16$, giving a high variance of $\mathrm{var}(\kappa^{(1)}) \approx 941.84$. The structure also shows 2-degree assortativity (coefficient $r = 0.1$ [7]). At last, the first and the second raw moments of the 3-degree (number of 3-edges incident on a node) distribution are $\langle \kappa^{(2)} \rangle \approx 16.22$ and $\langle \kappa^{(2)2} \rangle \approx 280.03$, giving a low variance of $\mathrm{var}(\kappa^{(2)}) \approx 16.97$.

University campus's dataset. The hypergraphs generated from this dataset consists of $N = 672$ nodes. The 2-degree $\kappa^{(1)}$ is distributed heterogeneously. The first and the second raw moments of the 2-degree distribution are $\langle \kappa^{(1)} \rangle \approx 15.3$ and $\langle \kappa^{(1)2} \rangle \approx 479.67$, giving a high variance of $\mathrm{var}(\kappa^{(1)}) \approx 245.56$. The structure also shows 2-degree assortativity (coefficient $r = 0.19$ [7]). At last, the first and the second raw moments of the 3-degree distribution are $\langle \kappa^{(2)} \rangle \approx 7.06$ and $\langle \kappa^{(2)2} \rangle \approx 57.68$, giving a low variance of $\mathrm{var}(\kappa^{(2)}) \approx 7.84$.

S1.2. Comparison with the mean-field model

In Fig. S1 we report the results shown in Fig. 3 of the main text, with the addition of the predictions made by the mean-field model (Eq. (3)). We also show the 2-degree and 3-degree distributions. These are well reproduced by exponential and gaussian distributions, respectively. Given the 2-degree heterogeneity and assortativity of the generated networks, is no surprise the poor performance of the mean-field approximation, especially in predicting the invasion threshold. The latter is always heavily overestimated. The reasons for this are multiple.

A first reason, valid for any h, is the 2-degree heterogeneity (see panels (b) and (f) in Fig. S1). Given that adding three-body interactions can only lower the invasion threshold, an upper bound for the threshold is found considering only two-body interactions, or equivalently, $h = 0$ (since, as we proved, 3-edges not overlapped with 2-edges do not affect the threshold). A rough estimate of the upper bound is provided by $\langle \kappa^{(1)} \rangle / \langle \kappa^{(1)2} \rangle$, as predicted by a heterogeneous (node-based) mean-field approximation, which yields a much smaller threshold than a homogeneous mean-field approximation when $\mathrm{var}(\kappa^{(1)})$ is large[3].

A second reason is that, on one hand, it has been shown that degree assortativity lowers the threshold [2]; on the other hand, mean-field approximations have been observed to perform exceptionally poorly against assortative networks [4]. These two facts together strongly suggest an additional contribution to the error made by the mean-field model.

Finally, increasing h, this model further overestimates the threshold, for it does not account for triangles sharing two nodes (i.e., a 2-edge), indeed present in the generated hypergraphs. In fact, once two nodes belonging to n triangles are both infected, n different three-body interac-tions become simultaneously active. This results in a smaller threshold than the one predicted by the mean-field model, given the sparsity it assumes (i.e., $n = 1$ only). To notice that the presence of triangles with shared 2-edges also explains the much lower prevalence observed for high h compared to that predicted by the mean-field model. Indeed, a set of n triangles sharing a given 2-edge, involves only $n + 2$ nodes. If they shared only one node, the involved nodes would be $2n + 1$; if they were all disconnected, the involved nodes would become $3n$. That is,

---

[3]Actually, the prediction made by the heterogeneous mean-field approximation underestimates the threshold. We are indeed dealing with highly clustered networks, for which the threshold is higher than for random networks with the same degree distribution.
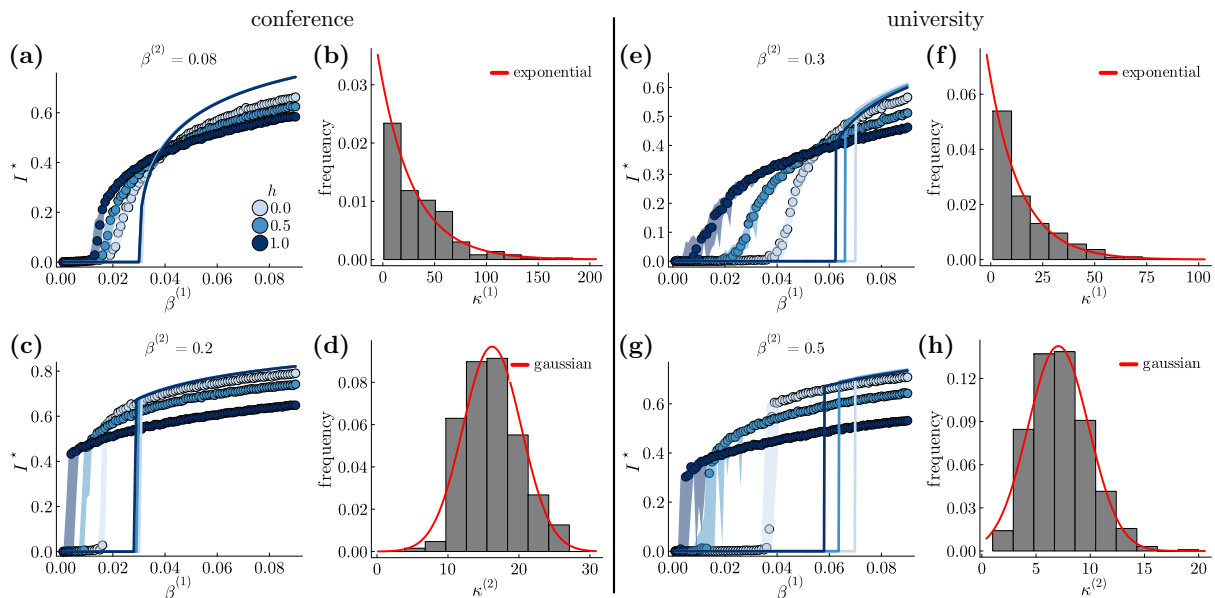
Figure S1: Comparison between the predictions made by the mean-field model (solid lines) and the numerical simulations (points) performed on the hypergraphs constructed from the conference's ((a) and (c)) and the university campus' ((e) and (f)) datasets. Points denote medians computed over 20 random initializations; ribbons cover from the 5-th to the 95-th percentile. The added three-body interactions form triangles with probability $h = 0$ (linear hypergraph), $h = 0.5$, and $h = 1$ (simplicial complex). ((b) and (d)) 2-degree and 3-degree distributions of the hypergraphs generated from the conference's dataset, shown to be well reproduced by an exponential distribution (mean: $1/\langle\kappa^{(1)}\rangle$) and a gaussian distribution (mean: $\langle\kappa^{(2)}\rangle$; variance: $\text{var}(\kappa^{(2)})$), respectively (red curves). ((f) and (h)) As panels (b) and (d), but for the hypergraphs generated from the university campus' dataset.

the more frequently triangles share nodes, the more redundant is the structure and, in turn, the smaller are the outbreaks.

It should be noted that a mean-field model is very much expected to fail in providing quantitatively accurate predictions for quenched structures [5]. In fact, the accuracy our mean-field theory shows for the configuration-model hypergraphs in Fig. 2 is not so obvious. For instance, a less refined node-based mean-field model would be very imprecise even in that case, especially in predicting the invasion threshold. In light of this, we can rather appreciate the significant aspect of the results reported in Fig. S1: that the phenomenology our theory revealed is, not only still valid for structures violating the assumptions of homogeneity and sparseness, but is actually greatly emphasized!

S2. Results for the SIR model

We show here the results of numerical simulations we performed using a susceptible-infectious-recovered (SIR) contagion model, where upon recovery individuals move to the recovered (R) compartment instead of entering back the susceptible one (S). This higher-order generalization of the SIR model has been very recently analyzed in detail by Lv et al. [6]. The authors showed that, as in the SIS model, sufficiently high values of the three-body infection rate, $\beta^{(2)}$, make the phase transition discontinuous in both homogeneous and heterogeneous simplicial complexes. Using however a node-based approximation, their approach is insensible to the way in which two- and three-body interactions are arranged in the structure, hence to their degree of overlap. Consequently, it is not possible to discern a simplicial complex from a linear hypergraph (or any other intermediate structure).

In Fig. S2, we show the results for hypergraphs generated from the university campus' dataset. The order parameter for a SIR-like model is the final attack rate, R_∞, which is the total
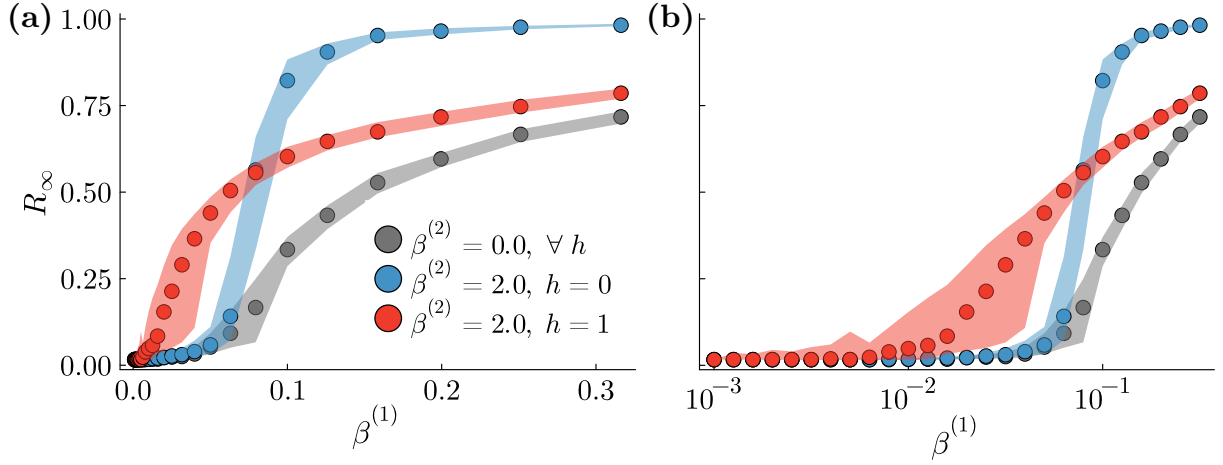
3

Figure S2: Numerical simulations performed on the hypergraphs constructed from the university campus' [8] datasets when the contagion dynamics is a SIR process. (a) Final attack rate, R_∞, versus the two-body infection rate, $\beta^{(1)}$. Points denote medians computed over 100 random initializations; ribbons cover from the 25-th to the 75-th percentile. The added three-body interactions form triangles with probability $h = 0$ (linear hypergraph; blue curve) and $h = 1$ (simplicial complex; red curve). Clearly, turning off the three-body interactions ($\beta^{(2)} = 0$), varying h has no effect on the dynamics (grey curve). (b) Same as panel (a), but with logarithmic abscissa to stretch the low-$\beta^{(1)}$ interval and better appreciate the shift of the invasion threshold.

fraction of nodes that got the infection (and eventually recovered) during the entire outbreak (the infected compartment is empty at equilibrium). The simulations confirm the generality of the phenomenology our mean-field model revealed: (i) three-body interactions affect the invasion threshold only if they overlap with two-body interactions ($h > 0$); (ii) a larger overlap (higher h) implies lower invasion thresholds but also smaller outbreaks; and (iii) varying exclusively the degree of overlap can change the nature of the phase transition. About the last point, it should be noted that the discontinuity of the transition for $\beta^{(2)} = 2$ and $h = 0$ is blurred by strong finite-size effects.

References

[1] G. Burgio, A. Arenas, S. Gómez, and J. T. Matamalas. Network clique cover approximation to analyze complex contagions through group interactions. *Communications Physics*, 4(1):111, 2021.

[2] G. D'Agostino, A. Scala, V. Zlatić, and G. Caldarelli. Robustness and assortativity for diffusion-like processes in scale-free networks. *Europhysics Letters*, 97(6):68006, 2012.

[3] M. Génois and A. Barrat. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Science*, 7(1):1–18, 2018.

[4] J. P. Gleeson, S. Melnik, J. A. Ward, M. A. Porter, and P. J. Mucha. Accuracy of mean-field theory for dynamics on real-world networks. *Physical Review E*, 85(2):026106, 2012.

[5] I. Z. Kiss, J. C. Miller, P. L. Simon, et al. Mathematics of epidemics on networks. *Cham: Springer*, 598:31, 2017.

[6] X. Lv, D. Fan, Q. Li, J. Wang, and L. Zhou. Simplicial sir rumor propagation models with delay in both homogeneous and heterogeneous networks. *Physica A: Statistical Mechanics and its Applications*, 627:129131, 2023.

[7] M. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20):208701, 2002.

[8] P. Sapiezynski, A. Stopczynski, D. D. Lassen, and S. Lehmann. Interaction data from the copenhagen networks study. *Scientific Data*, 6(1):315, 2019.