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.
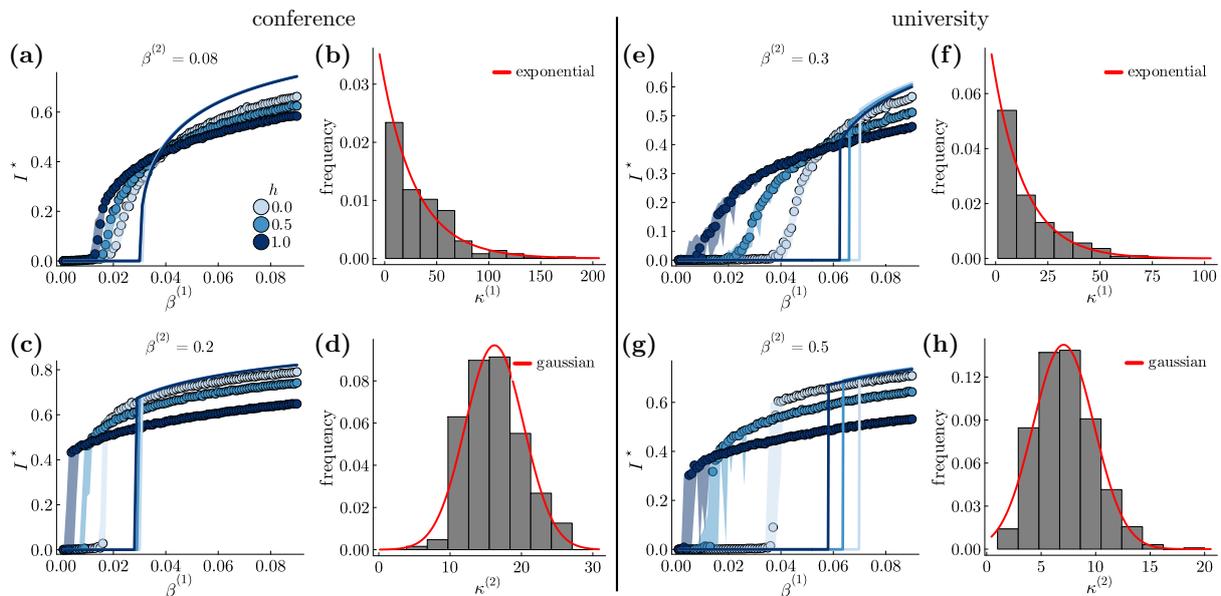
Figure S1: Comparison between the predictions made by the mean-field model (solid lines) and the numerical simulations (points) performed on the hypergraphs constructed from the conference's ((a) and (c)) and the university campus' ((e) and (f)) datasets. Points denote medians computed over 20 random 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
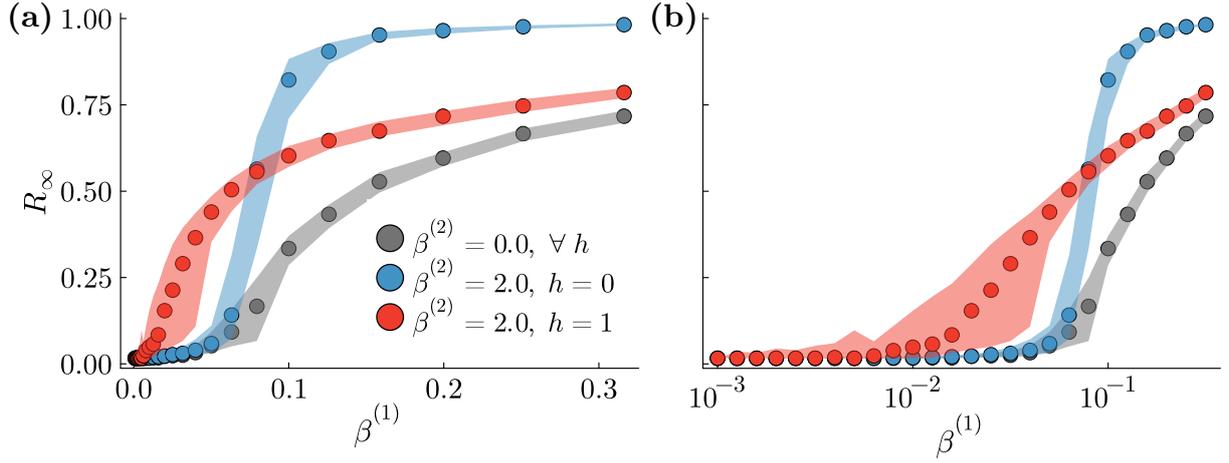
Figure S2: Numerical simulations performed on the hypergraphs constructed from the university campus' [8] datasets when the contagion dynamics is a SIR process. (a) Final attack rate, R_∞, versus the two-body infection rate, $\beta^{(1)}$. Points denote medians computed over 100 random initializations; 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.