

Network clique cover approximation to analyze complex contagions through group interactions

Giulio Burgio¹, Alex Arenas ¹, Sergio Gómez ¹✉ & Joan T. Matamalas ²

Contagion processes have been proven to fundamentally depend on the structural properties of the interaction networks conveying them. Many real networked systems are characterized by clustered substructures representing either collections of all-to-all pair-wise interactions (cliques) and/or group interactions, involving many of their members at once. In this work, focusing on interaction structures represented as simplicial complexes, we present a discrete-time microscopic model of complex contagion for a susceptible-infected-susceptible dynamics. Introducing a particular edge clique cover and a heuristic to find it, the model accounts for the higher-order dynamical correlations among the members of the substructures (cliques/simplices). The analytical computation of the critical point reveals that higher-order correlations are responsible for its dependence on the higher-order couplings. While such dependence eludes any mean-field model, the possibility of a bi-stable region is extended to structured populations.

¹Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona, Catalonia, Spain. ²Center for Interdisciplinary Cardiovascular Sciences, Cardiovascular Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ✉email: sergio.gomez@urv.cat

Epidemics¹, rumor spreading², adoption³, and opinion dynamics⁴ are well-known manifestations of real-world contagion processes. One of the most remarkable achievements of network science has been the characterization of the dependence of contagion processes on the structural properties of the interaction network through which they spread^{5–8}. In particular, many real-world networks of interest, especially the social ones, boast a clustered and cycles-rich structure. Triadic closures are indeed renowned to be a distinguishing feature of social systems^{9,10}, together with the presence of larger communities in which every element is connected to (nearly) any other element in them¹¹. Households and workplaces are common contact-based examples of that, while online social communities and groups in messaging apps are information-based ones. In the language of graph theory, such all-to-all substructures are called cliques. Specifically, a n -clique (or clique of order n) consists of n nodes all pair-wisely connected to each other, i.e., a complete subgraph of n nodes.

Apart from a collection of dyadic interactions, cliques can be also regarded as the pair-wise projection of richer substructures representing group interactions (also known as ‘higher-order’ interactions), involving more than two agents (nodes) at once. In fact, from few years on, a growing branch of literature dedicated to the study of various dynamical processes involving group interactions^{12–18}, has been showing that such interactions can heavily affect the dynamics, and neglecting them can therefore lead to wrong predictions.

The interaction patterns can be properly formalized by means of hypergraphs^{19,20}, a generalization of graphs in which the nodes can be grouped in hyperedges of any order—not only in pairs. Specifically, here we make use of simplicial complexes (SCs)²¹, defined as sets of faces, whereby a face is a set of nodes with the hereditary property, stating that each of its subsets is also a face of the SC. Associating a group interaction to each face, a SC then represents the entire set of interactions among the nodes in it.

Going to the dynamics, we are interested in complex contagion processes^{22–26}, in which the outcome of a potentially spreading interaction depends on how many different contagious agents take part to it, and not—as in simple contagions—only on the strength of the interaction. The presence of complex contagion mechanisms has been assessed in various contexts²⁷, but largely in online social networks and forums^{28,29}, thanks to their unique data traceability.

Several studies have recently appeared showing the dynamical effects of group interactions on complex contagion. They provide either a qualitative understanding by means of mean-field approximations^{30–32}, finding such interactions as responsible for critical mass effects; or a more quantitative one, reinforcing the previous qualitative findings, but for very particular hypergraphs^{33–35}. All the studies considered continuous-time dynamics and, noteworthy, uncorrelated nodes’ states. For the case of two-dimensional SCs (i.e., consisting of edges and triangles), Matamalas et al.³⁶ provided a microscopic discrete-time model accounting for first-order (two-nodes) and, only partially, second-order (three-nodes) dynamical correlations. However, as shown afterwards, the same subtle inconsistencies that make the model applicable to any two-dimensional SC, also impede the computation of the critical point.

In this work, looking for a discrete-time model holding for SCs of any dimension, we reveal that fixing those inconsistencies puts a topological condition on the interaction structure the model can describe, namely that two group interactions can only share one node. In order to satisfy this constraint, we introduce the notion of edge-disjoint edge clique cover (EECC) of a SC and a heuristic to find it. By means of our Microscopic Epidemic Clique Equations (MECLE), we provide a two-fold extension of the existent

discrete-time complex contagion models, accounting for higher-order correlations and group interactions (if any) at the same time. We prove those dynamical correlations to be essential to describe how the critical point depends on the higher-order couplings. Lastly, different approaches to treat group interactions sharing multiple nodes are also discussed, while hinting at easy adaptations of ours.

Results and discussion

The over-counting problem. Let us first introduce all the basic notions needed to state the problem. We start from the interaction structures we used, SCs. A SC \mathcal{K} is a subset of the power-set 2^V of a vertex set V , endowed with the hereditary property: given $f \in \mathcal{K}$ and $f' \subseteq f$, then $f' \in \mathcal{K}$. Note that we can neglect the empty set from 2^V , for it does not have any practical interest here. The elements of \mathcal{K} are called faces, and a n -dimensional face (or n -face) is a subset of V made of $n + 1$ nodes. Given a face f , its power-set 2^f is called a simplex. If f is a n -face, 2^f is a n -simplex. Giving a geometrical interpretation to \mathcal{K} , a n -simplex is also the n -dimensional polytope being the convex closure of its $n + 1$ vertices. For example, given $V = \{i, j, k\}$ and $\mathcal{K} = 2^V$, $\{i\}$ is a 0-simplex (a point), $\{\{i, j\}, \{i\}, \{j\}\}$ is a 1-simplex (a segment), $\{\{i, j, k\}, \{i, j\}, \{i, k\}, \{j, k\}, \{i\}, \{j\}, \{k\}\}$ is a 2-simplex (a triangle).

If a simplex in \mathcal{K} is not included in any other simplex in \mathcal{K} , then it is said to be maximal. If d is the maximum dimension of the faces in \mathcal{K} , then \mathcal{K} is d -dimensional and is called simplicial d -complex. The underlying graph $\mathcal{K}^{(1)}$ of \mathcal{K} is the graph induced by the 1-simplices in \mathcal{K} , i.e., the graph whose node set is V and whose edge set is the set of all the 1-faces in \mathcal{K} . Consequently, a n -simplex induces a $(n + 1)$ -clique, made of $(0.0ptn + 12)$ edges (i.e., 1-faces), in $\mathcal{K}^{(1)}$. If a clique is not part of a larger one, it is said to be maximal. Finally, \mathcal{K} is said to be q -connected if, given any two simplices $s_1, s_2 \subset \mathcal{K}$, there exists a sequence of simplices connecting s_1 and s_2 such that any two adjacent simplices of the sequence have (at least) a q -face in common; if \mathcal{K} is connected, then $\mathcal{K}^{(1)}$ is 0-connected—as any other connected graph.

To identify whether a n -clique c in $\mathcal{K}^{(1)}$ corresponds to a $(n - 1)$ -simplex in \mathcal{K} or it is just a n -clique also in \mathcal{K} , we introduce a binary variable g , the group classifier, defined to be 1 or 0, respectively; by which c is regarded as a (g, n) -clique. Since a 1-simplex is equivalent to a 2-clique, we choose to assign $g = 0$ to any 2-clique, so that $n \geq 3$ when $g = 1$. From now on, unless explicitly specified, we refer to the g -classified cliques in $\mathcal{K}^{(1)}$ simply as ‘cliques’. These are the building blocks of our description.

Going to the dynamics of interest, we adopt, without loss of generality, a standard epidemiological terminology. We consider discrete-time susceptible-infected-susceptible (SIS) dynamics on a SC. Let $\beta^{(n)}$ be the probability with which a susceptible node in a group of $n + 1$ nodes (a n -face of the SC) gets infected when all the other n nodes are infected; and μ be the probability with which an infected node recovers. Due to the hereditary property, if n nodes are infected, and thus can pass the infection as a group of n nodes, then also any subset of that group can: for any $k \leq n$, there are $(0.0ptnk)$ subsets passing the infection with probability $\beta^{(k)}$ as a group of k nodes. In other words, to each state configuration of the n nodes (i.e., which of them are infected) corresponds a set of concurrent channels of infections, and these channels are correlated. Compared to existing models³⁶, we take a first step forward by accounting for such correlation, allowing us to compute the critical point. Then, when computing the probability for a node to get infected within a group, the contribution coming from each state configuration of the group consists of a product over the concurrent channels that

configuration admits. In addition, each contribution is weighted by the probability for the group to be in that configuration.

Now, given a clique, no matter whether it conveys ($g = 1$) or not ($g = 0$) group interactions, we want to account also for the dynamical correlations among the states of the nodes in it. However, if s of such cliques have $m \geq 2$ nodes in common, the contribution to the infection of one of the m nodes, coming from the others $m - 1$, is counted s times instead of once. This is because the $m - 1$ nodes would appear in the state configuration probabilities associated with each of the s cliques. To note that if the cliques are just edges, then necessarily $m = 1$, meaning that the over-counting is excluded in pair-wise models.

Edge-disjoint edge clique cover. To avoid the over-counting, we aim at covering all the edges of $\mathcal{K}^{(1)}$ by means of a set of cliques such that any two of them share at most one node, giving what we call here an EECC. Cliques sharing more than one node are consequently decomposed in lower-order cliques. This comes with no essential repercussions when the decomposed cliques have all $g = 0$. Otherwise, some group interactions would be ignored, implying the model to strictly apply for group interactions sharing at most one node. Interestingly, as shown later on, the model remains reliable when the over-counted interactions are relatively scarce.

Covering the structure by cliques only, we are able to give a unique equation holding for any order n ; an impossible task via generic, less symmetric substructures. Furthermore, since we want to capture as many correlations and group interactions as possible, we want the cover to consist of the least possible number of cliques. Finding such minimal set of edge-disjoint cliques is closely related to the edge clique cover problem, known to be NP-complete³⁷. Heuristics are thus necessary to estimate the solution in large graphs. For convenience, from now on, we reserve the acronym EECC to sets which are solutions to the problem.

If all the maximal cliques in $\mathcal{K}^{(1)}$ are edge-disjoint, then $\mathcal{K}^{(1)}$ admits a unique EECC, simply given by the set of the maximal cliques in it. Otherwise, $\mathcal{K}^{(1)}$ generally admits multiple EECCs. See Fig. 1 for illustration.

To estimate an EECC, we propose the following greedy heuristic (the best among several options we conceived for this task). Given a graph G , the heuristic proceeds as follows:

1. Find the set C of all the maximal cliques in G
2. Include in the EECC and remove from C all the elements in C that do not share edges with other elements in C
3. While C is not empty
 - (a) For every maximal clique $c \in C$, compute the score r_c , defined as the fraction of edges that c shares with the other elements in C
 - (b) Consider the elements of C with the lowest score; include in the EECC and remove from C (one, randomly chosen, of) the element(s) of highest order among them

Noteworthy, when dealing with highly modular structures—as those representing many real social systems—in which communities of nodes are loosely connected among them, the search for an (edge-disjoint) edge clique cover can be speeded up. Indeed, each time two regions of the structure are joined by bridging cliques (which are evidently maximal), the problem of finding the optimal cover of the whole structure reduces to that of finding it in each of the two, smaller regions.

Now, whenever the maximal cliques in $\mathcal{K}^{(1)}$ are not all edge-disjoint, some cliques are forced to be decomposed in sub-cliques during step 4. Since decomposing a $(1, \cdot)$ -clique in $(0, \cdot)$ -connected sub-cliques means also neglecting some group interactions, whenever a $(0, \cdot)$ -clique and a $(1, \cdot)$ -clique are not edge-disjoint, we prefer to include the latter in the EECC. This additional difficulty disappears whenever the SC is 0-connected or when it has dimension 1 (i.e., when it is a graph). Thus, being \mathcal{K} a SC and $\mathcal{K}^{(1)}$ its g -classified underlying graph, a EECC of \mathcal{K} is constructed as follows:

- I. Consider the subgraph $G_1 \subseteq \mathcal{K}^{(1)}$ induced by the nodes in the maximal $(1, \cdot)$ -cliques in \mathcal{K} . Find a EECC of G_1 ; let us call $\mathcal{S}(\mathcal{K})$ the resulting EECC
- II. Consider the subgraph $G_0 = \mathcal{K}^{(1)} \setminus G_1$. Find a EECC of G_0 ; let us call it $\mathcal{C}(\mathcal{K})$
- III. $\mathcal{D}(\mathcal{K}) \equiv (\mathcal{S}(\mathcal{K}), \mathcal{C}(\mathcal{K}))$ is the estimated EECC of \mathcal{K}

A basic question is about the dependence of the prediction made by the model when different EECCs are estimated for a given structure. Indeed, while the effectiveness of the heuristic ensures better performance of the model, its robustness is an indispensable quality, as we look for a reliable model giving

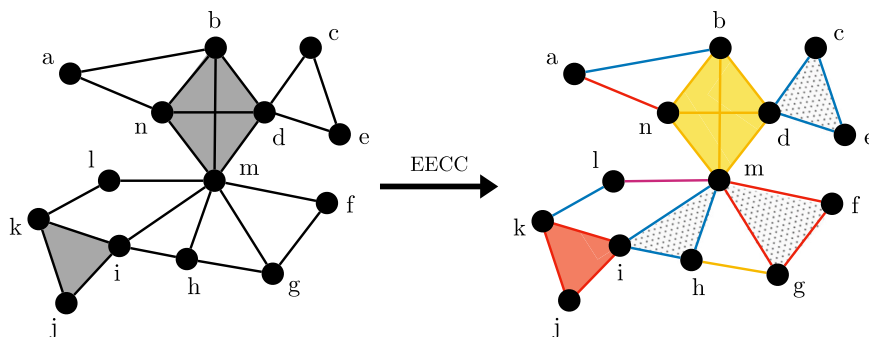


Fig. 1 Edge-disjoint edge clique cover (EECC) of a small simplicial complex (SC). The SC, shown on the left, consists of fourteen nodes (0-simplices), identified via letters, connected by one 3-simplex, one 2-simplex, and fourteen 1-simplices. Gray areas indicate r -simplices with $r \geq 2$, including $r + 1$ nodes each. The EECC of the SC is shown on the right, where colored and dotted areas are used to visualize, respectively, the $(1, r)$ -cliques and $(0, r)$ -cliques in it, with colors carrying no specific meanings. The SC is decomposed in: one $(1, 4)$ -clique, $\{b, d, m, n\}$; one $(1, 3)$ -clique, $\{i, j, k\}$; three $(0, 3)$ -cliques, $\{c, d, e\}$, $\{f, g, m\}$, $\{h, i, m\}$; and five $(0, 2)$ -cliques, $\{a, b\}$, $\{a, n\}$, $\{g, h\}$, $\{k, l\}$, $\{l, m\}$. The underlying subgraph induced by the subset $\{a, b, d, m, n\}$ originally consists of a $(1, 4)$ -clique and a $(0, 3)$ -clique. To preserve the group interaction mediated by the $(1, 4)$ -clique, is preferable to include this in the EECC and then break the $(0, 3)$ -clique into two (non-maximal) $(0, 2)$ -cliques. Besides, the underlying subgraph induced by the subset $\{f, g, h, i, m\}$ is made of three overlapping $(0, 3)$ -cliques, and the EECC is in this case obtained by including $\{f, g, m\}$ and $\{h, i, m\}$ (and then the remaining edge $\{g, h\}$), instead of $\{g, h, m\}$ first.

certain results when fed up with a certain structure, of which a minimal EECC is estimated. In Supplementary Note 1 and Supplementary Figs. 1 and 2, our model is shown to be robust under EECC variability. Therefore, given a SC, it is safe to make use of the first EECC computed for it.

Calling m_1 the maximum order of the cliques we want to include in $\mathcal{S}(\mathcal{K})$ (i.e., considering simplices of dimension up to $m_1 - 1$ in \mathcal{K}), m_1 could be smaller than ω_1 , the maximum order of the cliques in G_1 . In such case, when looking for an EECC of G_1 , those maximal cliques in G_1 of order greater than m_1 must be decomposed in edge-disjoint sub-cliques (corresponding to sub-simplices in \mathcal{K}) of variable order $m' \in \{1, \dots, m_1\}$. Clearly, the higher is m_1 , the higher is the order of the group interactions (and of the correlations within them) included in the description. Overall, as long as the proportion of obviated group interactions is small enough, the deviations from the complete dynamics are comparatively negligible; or alternatively, the error made by including non-0-connected simplices is negligible.

In the building of an EECC, different values $m_0 \geq 2$ for the maximum order to be considered for $(0, \cdot)$ -cliques can also be chosen. The higher is m_0 , the higher is the order of the captured dynamical correlations within the cliques in \mathcal{K} . In any case, $m_0 \leq \omega_0$, being ω_0 the maximum order of the cliques in G_0 .

Summarizing, the couple (m_0, m_1) identifies the considered implementation of the MECLE.

Microscopic epidemic clique equations. Given a (g, n) -clique $\{i_1, \dots, i_n\}$ in $\mathcal{D}(\mathcal{K})$, with $n \leq m \equiv \max\{m_0, m_1\}$, we indicate with $P_{i_1 \dots i_n, g}^{\sigma_{i_1} \dots \sigma_{i_n}}$ the joint probability that node i_1 is in the state σ_{i_1} , node i_2 is in the state σ_{i_2} , etc., where $\{\sigma_{i_1}, \dots, \sigma_{i_n}\} \in \{S, I\}^n$. Besides, with $P_{i_1 \dots i_n, g}^{\sigma_{i_1} \dots \sigma_{i_{k-1}} \sigma_{i_{k+1}} \dots \sigma_{i_n} | \sigma_{i_k}}$ we indicate the conditional probability that nodes $i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_n$ are in their respective states $\sigma_{i_1}, \dots, \sigma_{i_{k-1}}, \sigma_{i_{k+1}}, \dots, \sigma_{i_n}$, given that node i_k is in the state σ_{i_k} . Clearly, the normalization condition must hold:

$$\sum_{\{\sigma_{i_k}\}_{k=1, \dots, n}} P_{i_1 \dots i_n, g}^{\sigma_{i_1} \dots \sigma_{i_n}} = 1 \tag{1}$$

Indicating with $\{\sigma_{i_k}\}_{k=1, \dots, n}$ the states at time t and with $\{\sigma'_{i_k}\}_{k=1, \dots, n}$ those at time $t + 1$, the MECLE model dynamic equation governing the evolution of the state of a (g, n) -clique $\{i_1, \dots, i_n\}$ reads

$$P_{i_1 \dots i_n, g}^{\sigma'_{i_1} \dots \sigma'_{i_n}}(t + 1) = \sum_{\{\sigma_{i_1}, \dots, \sigma_{i_n}\}} P_{i_1 \dots i_n, g}^{\sigma_{i_1} \dots \sigma_{i_n}}(t) \Phi_g(\{\sigma_{i_k}\}, \{\sigma'_{i_k}\}; \{\beta^{(s)}\}, \mu) \tag{2}$$

where

$$\Phi_g(\{\sigma_{i_k}\}, \{\sigma'_{i_k}\}; \{\beta^{(s)}\}, \mu) = \prod_{k=1}^n \phi_{i_k, g}(\{\sigma_{i_k}\}, \sigma'_{i_k}; \{\beta^{(s)}\}, \mu) \tag{3}$$

is the transition probability from the starting state $\{\sigma_{i_k}\}_{k=1, \dots, n}$ to the arrival state $\{\sigma'_{i_k}\}_{k=1, \dots, n}$. It is understood that Φ_g is computed at time t . This is expressed as a product over the single-node

transition probabilities $\{\phi_{i_k, g}\}$, given by

$$\begin{aligned} \phi_{i_k, g}(\{\sigma_{i_k}\}, \sigma'_{i_k}; \{\beta^{(s)}\}, \mu) &= \mathbb{1}[\sigma_{i_k} = I, \sigma'_{i_k} = I](1 - \mu) \\ &+ \mathbb{1}[\sigma_{i_k} = I, \sigma'_{i_k} = S] \mu \\ &+ \mathbb{1}[\sigma_{i_k} = S, \sigma'_{i_k} = I] \left[1 - \frac{w_{N_I, g}^{(n-1)}}{q_{i_k(-i_k), g}^{(n-1)} \prod q_{i_k, g'}^{(r)}} \right] \\ &+ \mathbb{1}[\sigma_{i_k} = S, \sigma'_{i_k} = S] \left[\frac{w_{N_I, g}^{(n-1)}}{q_{i_k(-i_k), g}^{(n-1)} \prod q_{i_k, g'}^{(r)}} \right] \end{aligned} \tag{4}$$

where $\mathbb{1}[p]$ gives 1 if condition p is fulfilled and 0 otherwise; $N_I = |\{i_{k=1, \dots, n} | \sigma_{i_k} = I\}|$ is the number of infected nodes in the starting state; $w_{N_I, g}^{(n-1)} \equiv w_{N_I, g}^{(n-1)}(\{\beta^{(s)}\})$ is the probability that a susceptible node (i_k) does not get infected within a (g, n) -clique $(\{i_1, \dots, i_k, \dots, i_n\})$ whose state configuration ($N_I \leq n - 1$) is known, reading

$$w_{N_I, g}^{(n-1)} = \mathbb{1}[g = 0](1 - \beta^{(1)})^{N_I} + \mathbb{1}[g = 1] \prod_{s=1}^{N_I} (1 - \beta^{(s)})^{(N_I)} \tag{5}$$

and $q_{i_k, g'}^{(r)} \equiv q_{i_k, g'}^{(r)}(\{\beta^{(s)}\})$ is the probability that node i_k does not get infected via any of the $(g', r + 1)$ -cliques incident on it, that is

$$\begin{aligned} q_{i_k, g'}^{(r)} &= \prod_{\substack{\{i_1, \dots, i_r\} \\ \in \Gamma_{i_k, g'}^{(r)}}} \left[1 - \sum_{l=1}^r \frac{1 - w_{l, g'}^{(r)}}{l!(r - l)!} \right] \\ &\times \sum_{k_1 \neq \dots \neq k_r = 1}^r P_{j_{k_1} \dots j_{k_r} j_{k_1+1} \dots j_{k_r} | i_k, g'}^{I \dots IS \dots S | S} \end{aligned} \tag{6}$$

where $\Gamma_{i_k, g'}^{(r)}$ indicates the set of r -tuples of indexes corresponding to subsets of r nodes forming a $(g', r + 1)$ -clique with i_k , and $q_{i_k(-i_k), g}^{(n-1)}$ coincides with $q_{i_k, g}^{(n-1)}$ expect for excluding the considered (g, n) -clique $\{i_1, \dots, i_k, \dots, i_n\}$ from the product. Finally, the products in Eq. (4) are performed over the couples $\{(g', r)\}$ such that $2 \leq r \leq m_0$ for $g' = 0$, and $3 \leq r \leq m_1$ for $g' = 1$.

In Eq. (3) the single-node transition probabilities, ϕ , are treated as independent from each other within the time step. This merely derives from the implicit assumption that all the events within a given time step are simultaneous, and therefore not causally related. Simply put, the state of a node at time step $t + 1$ only depends on its state and on the states of its neighbors at the previous time step t , as in any markovian model.

Importantly, to get the expression for $q_{i_k, g'}^{(r)}$ we have adopted the following closure

$$P_{j_{k_1} \dots j_{k_r} j_{k_1+1} \dots j_{k_r} | i_k, g'}^{I \dots IS \dots S | S} = \frac{P_{j_{k_1} \dots j_{k_r} j_{k_1+1} \dots j_{k_r} | i_k, g'}^{I \dots IS \dots S} P_{i_1 \dots i_k \dots i_n, g}^{\sigma_{i_1} \dots \sigma_{i_n}}}{P_{i_k}^S} \tag{7}$$

(the classifier g is assigned only to the cliques in $\mathcal{D}(\mathcal{K})$). Therefore, by definition of conditional probability, the probability $P_{j_{k_1} \dots j_{k_r} j_{k_1+1} \dots j_{k_r} | i_k \dots i_n}^{I \dots IS \dots S | S}$ appears in Eq. (6) as $P_{j_{k_1} \dots j_{k_r} j_{k_1+1} \dots j_{k_r} | i_k, g'}^{I \dots IS \dots S | S}$. Intuitively, each clique in $\mathcal{D}(\mathcal{K})$ is treated as an independent dynamical unit of the system, accounting for the correlations among the states of the nodes it includes. In the form of a generalization of the classical pair approximation³⁸, the dynamical correlations between two adjacent cliques are conveyed by the state probability of the node they share (in the denominator to avoid double counting). In this regard, the larger the adjacent

cliques, the smaller the expected influence of the state of the shared node on all the other ones. For this reason, the presented closure—and consequently the MECLE—is expected to gain further accuracy with the increasing of the order of the cliques.

The probability P_i^I for the single node i of being infected is computed as a marginal probability from any (g, n) -clique including i , as

$$P_i^I(t + 1) = \sum_{\{\sigma_{-i}\}} P_{i\{-i\},g}^{I\{\sigma_{-i}\}}(t + 1) = P_i^I(t)(1 - \mu) + P_i^S(t) \left(1 - \prod_{(g,r)} q_{i,g}^{(r)} \right) \quad (8)$$

in which $\{-i\}$ and $\{\sigma_{-i}\}$ indicate, respectively, the set of the other $n - 1$ nodes in the clique and their states. Equation (8) is, consistently, also found taking $n = 1$ in Eq. (2) ($w_{N,g}^{(0)} = 1$, $q_{i,g}^{(0)} = 1$). The average value ρ of the single node probabilities,

$$\rho(t) = \frac{1}{N} \sum_{i=1}^N P_i^I(t) \quad (9)$$

is the epidemic prevalence, which is also the order parameter of the system.

It is important to note that the used closure, Eq. (7), apart from preserving the n -node state correlations of a considered (g, n) -clique, is the only one making feasible the marginalization of Eq. (2) to get Eqs. (8) and (9) and, consequently, get the expression for the epidemic threshold, through Eqs. (15)–(19) in “Methods”.

At any time step t , Eq. (8) yields one constraint for each of the n nodes in a (\cdot, n) -clique, leaving $2^n - n - 1$ independent state probabilities to be determined, one being fixed by the normalization. Therefore, if \mathcal{K} has N nodes and $\mathcal{D}(\mathcal{K})$ consists of $C^{(n)}$ (\cdot, n) -cliques, $n \leq m$, the MECLE is defined by a system of $N + \sum_{n=2}^m (2^n - n - 1)C^{(n)}$ independent equations.

To help in the understanding of the model, we show in Supplementary Note 2 all the equations for the particular case of the simplicial 2-complexes, i.e., the (3, 3) implementation of the MECLE.

Finally, we can frame the existent models in the (m_0, m_1) notation. As $m_1 = 2$ implies $S(\mathcal{K}) = \emptyset$, $\{(m_0, 2)\}_{2 \leq m_0 \leq \omega_0}$ is the class of models accounting for correlations within cliques of order up to m_0 on graphs (simple contagion). In particular, (2, 2) gives the Epidemic Link Equations (ELE) model³⁹. The Microscopic Markov Chain Approach (MMCA)⁴⁰ is then recovered from (2, 2) by assuming that $P_{ji}^{IS} = P_j^I P_i^S$ (i.e., $P_{ji}^{IS} = P_j^I$ in Eq. (6)). Considering group interactions, the simplicial ELE and MMCA³⁶ fall outside the MECLE class of models, for they do not account for the correlations among the concurrent channels of infection within simplices. This, alongside the consideration of higher-order dynamical correlations among nodes’ states, is precisely the refinement made here.

Results for simplicial 2-complexes. We apply here the developed formalism to the case of simplicial 2-complexes. As we are only using synthetic structures, the SCs are constructed from some graph (the future underlying graph) converting into (1, 3)-cliques (i.e., 2-faces) a fraction p_Δ of the 3-cliques allowed by the EECC of the graph, while considering as (0, 3)-cliques the remaining fraction $1 - p_\Delta$. Specifically, if all the 3-cliques are converted ($p_\Delta = 1.0$), the resulting SCs are the clique complexes of the respective original graphs. To better appreciate the improvement made here, we mainly show results for clique complexes, although notable improvements can be generally found for any value of p_Δ depending on the used structure. Conveniently, if p_Δ is not specified then the structure is understood to be a clique complex.

We identify a such generated SC adding ‘SC’ to the name of the graph model used for it: a ‘Dorogovtsev-Mendes SC’, for instance, is built upon a graph given by the Dorogovtsev-Mendes generative model⁴¹. Specifically, the ‘random SC’ is obtained as follows: first we generate a simplicial 2-complex through the random SC model³⁰; then we consider its underlying graph and compute its EECC; finally, the random SC is got, as before, by converting 3-cliques in 2-faces, in this way ensuring the SC to be 0-connected.

In Fig. 2 we compare the prevalence ρ obtained using Monte Carlo (MC) simulations (see Methods for details), the MECLE model, and the other discrete-time markovian models, i.e., the simplicial ELE and MMCA models³⁶, for different 0-connected clique complexes. For all structures, the improvement brought by the MECLE with respect to the simplicial ELE is substantial, in both predicting the epidemic prevalence and, even more, locating the critical point, for which the relative errors $\varepsilon_{\beta^{(1)}}$ are reported. Note that the predictions is expected to improve for increasing order of the cliques, since the precision of the used closure, Eq. (7), grows with it as well, especially in the case of populations arranged in lowly inter-connected dense communities.

We then leverage the prominent jump of the discontinuous transition found for random SCs, to illustrate, in Fig. 3, the existence of a hysteresis cycle, enclosing a bi-stable region. This is in line with recent findings^{30,31,33–36}. Again, the simplicial ELE is outperformed by the MECLE, especially in predicting the ‘backward’ curves, as the overestimation made by the former (see next subsection) is emphasized in that case.

More interestingly, our analysis clearly shows that models with uncorrelated nodes’ states generally fail to pinpoint the (‘forward’) epidemic threshold at even a qualitative level. Indeed, by neglecting those dynamical correlations, Eq. (22) (see “Methods”) predicts $\beta_{cr}^{(1)}$, the value of $\beta^{(1)}$ at which the epidemic-free state becomes unstable, to be independent from the values of the higher-order infection probabilities, $\{\beta^{(s)}\}_{s>1}$. In particular, in the case of two-dimensional SCs, increasing enough $\beta^{(2)}$ results in the appearance of a bi-stable region, but $\beta_{cr}^{(1)}$ does not change in those models^{30,31}. On the contrary, when the actual structure of the interactions is retained, increasing $\beta^{(2)}$ does lead $\beta_{cr}^{(1)}$ to decrease, as arises from MC simulations and correctly predicted by the MECLE. We show this dependence in Fig. 4, while it can also be grasped by comparing panels (c) and (d) in Fig. 2.

The calculation of the critical point in the MECLE (see “Methods”), together with the evidences coming from the MC simulations, reveal the necessity of accounting for higher-order dynamical correlations. An interaction (infection) of order s , with coupling (infection probability) $\beta^{(s)}$, requires s nodes in a simplex to be active (infected). Therefore, in order to preserve its contribution near the critical point, the state probability with s active nodes must not be neglected—as done instead in models with uncorrelated nodes’ states. In other words, near the epidemic threshold, higher-order interactions rely exclusively on comparatively higher-order correlations, so that their effect when varying the higher-order couplings, $\{\beta^{(s)}\}_{s>1}$, is observed only if such correlations are preserved. Accordingly, such dependence eludes the MF and MMCA approximations of the system. Besides, as second-order correlations within triangles are partially accounted for in the simplicial ELE, the latter predicts this dependence, but not as accurately as MECLE does. Finally, in the endemic state and away from the threshold, low-order correlations suffice to sustain higher-order interactions. Nevertheless, the higher the order of the accounted correlations, the more accurate is the quantification of the prevalence.

While a closed equation for $\beta_{cr}^{(1)}$ is generally inaccessible for complex interaction structures, in Supplementary Note 3 and

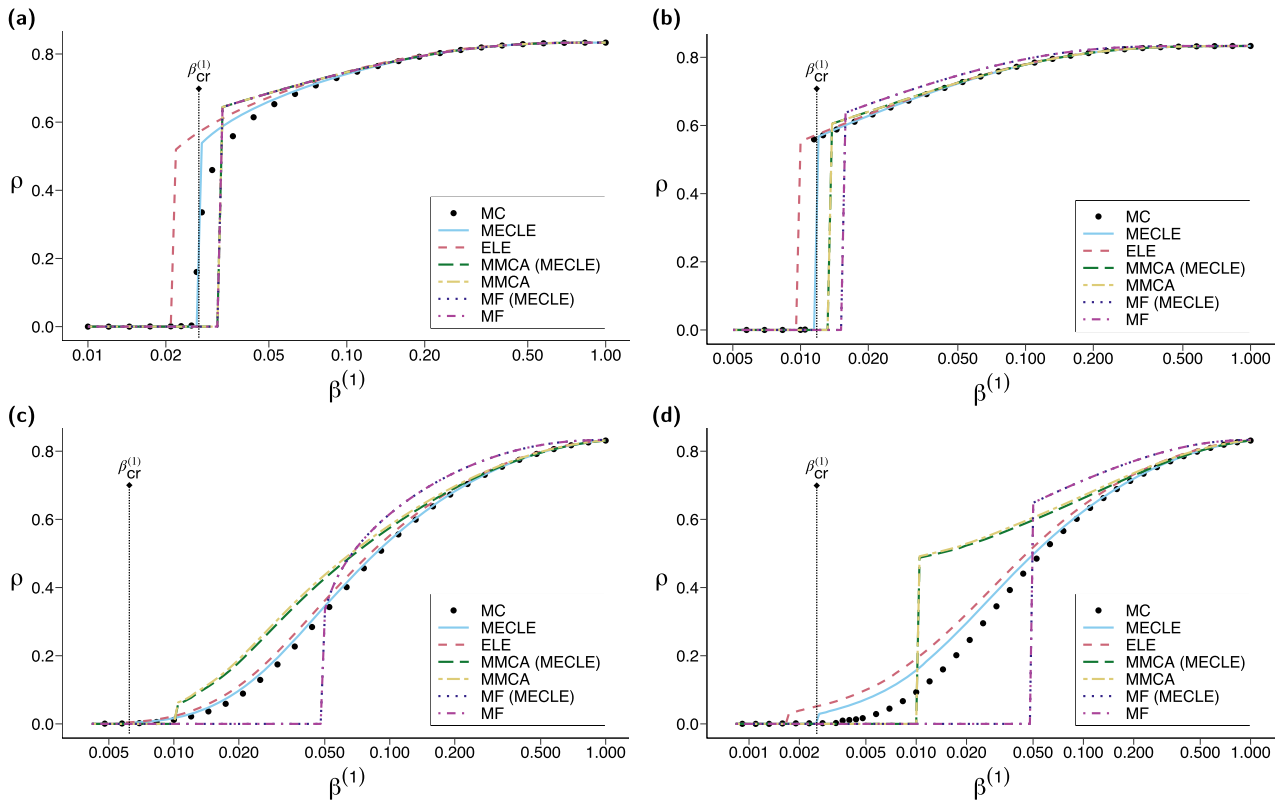


Fig. 2 Epidemic prevalence ρ as a function of the edge infection probability $\beta^{(1)}$ on different 2-dimensional simplicial complexes (SCs). Results obtained from Monte Carlo (MC) simulations are depicted by dots, while lines represent the analytically computed prevalence using the indicated models. MMCA and MMCA(MECL) refer to the Microscopic Markov Chain approximation of, respectively, the simplicial Epidemic Link Equations (ELE) model and the Microscopic Epidemic Clique Equations (MECLE) model, as obtained by considering the state probabilities of the nodes as uncorrelated; while MF and MF (MECLE) refer to their homogeneous mean-field approximations (see Methods). Note that MF and MF(MECL) are indistinguishable at the used scale. The value of the epidemic threshold, as computed in the MECLE through Eqs. (15)–(19) in Methods, is marked with a vertical dotted line. The recovery probability is fixed to $\mu = 0.2$. **a** Periodic triangular SC with $\bar{k}^{(0,1)} = 0.00$, $\bar{k}^{(0,2)} = 0.00$ and $\bar{k}^{(1,2)} = 3.00$, being $\bar{k}^{(g,r)}$ the mean number of $(g, n + 1)$ -cliques incident on a node, and triangle infection probability $\beta^{(2)} = 0.25$; the relative error in locating the epidemic threshold is $\epsilon_{\beta^{(1)}} \approx 0.08$ for MECLE and $\epsilon_{\beta^{(1)}} \approx 0.12$ for ELE. **b** Random SC with $\bar{k}^{(0,1)} = 4.10$, $\bar{k}^{(0,2)} = 0.00$ and $\bar{k}^{(1,2)} = 3.95$, and $\beta^{(2)} = 0.15$; $\epsilon_{\beta^{(1)}} \approx 0.06$ for MECLE and $\epsilon_{\beta^{(1)}} \approx 0.09$ for ELE. **c** Dorogovtsev-Mendes SC with $\bar{k}^{(0,1)} = 1.10$, $\bar{k}^{(0,2)} = 0.00$ and $\bar{k}^{(1,2)} = 1.45$, and $\beta^{(2)} = 0.25$; $\epsilon_{\beta^{(1)}} \approx 0.07$ for MECLE and $\epsilon_{\beta^{(1)}} \approx 0.16$ for ELE. **d** Same as **c** but with $\beta^{(2)} = 0.50$; $\epsilon_{\beta^{(1)}} \approx 0.23$ for MECLE and $\epsilon_{\beta^{(1)}} \approx 0.50$ for ELE.

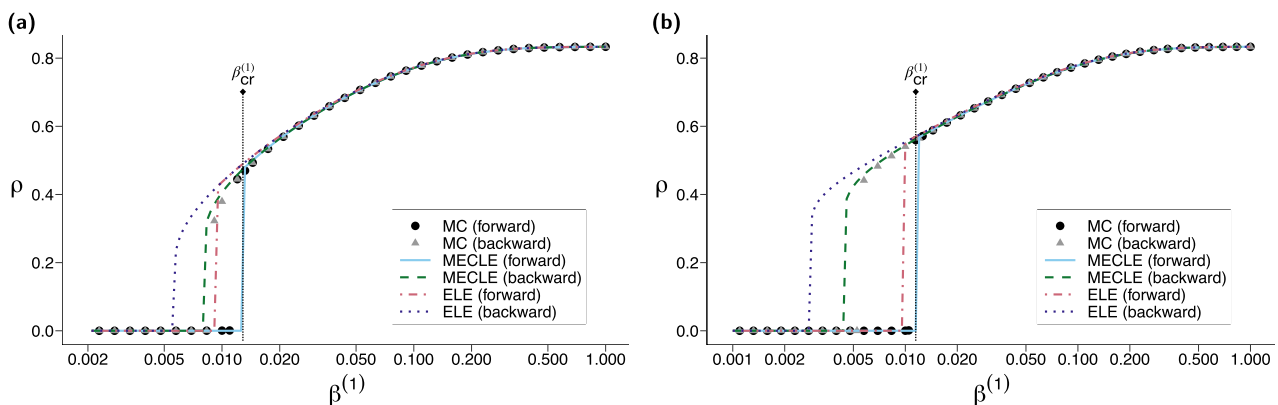


Fig. 3 Hysteresis cycle of the epidemic prevalence ρ with respect to the edge infection probability $\beta^{(1)}$ on different two-dimensional simplicial complexes (SCs). Results obtained from Monte Carlo (MC) simulations are depicted by dots, while lines represent the analytically computed prevalence using the simplicial Epidemic Link Equations (ELE) model and the Microscopic Epidemic Clique Equations (MECLE) model. The ‘forward’ and ‘backward’ curves are obtained through small equilibrium transformations taking as initial value ρ_0 the equilibrium value of ρ got at the next smaller and next greater value of $\beta^{(1)}$, respectively. The hysteresis cycle reveals the bi-stable region. The computed value of the (forward) epidemic threshold is marked with a vertical dotted line. The recovery probability is fixed to $\mu = 0.2$. **a** Random SC with $\bar{k}^{(0,1)} = 4.10$, $\bar{k}^{(0,2)} = 1.58$ and $\bar{k}^{(1,2)} = 2.37$ ($p_{\Delta} = 0.6$), being $\bar{k}^{(g,r)}$ the mean number of $(g, n + 1)$ -cliques incident on a node, and triangle infection probability $\beta^{(2)} = 0.25$; the relative errors in locating the epidemic threshold are $\epsilon_{\beta^{(1)}}^{\text{for.}} \approx 0.14$ and $\epsilon_{\beta^{(1)}}^{\text{back.}} \approx 0.10$ for MECLE, and $\epsilon_{\beta^{(1)}}^{\text{for.}} \approx 0.18$ and $\epsilon_{\beta^{(1)}}^{\text{back.}} \approx 0.38$ for ELE. **b** Random SC with $\bar{k}^{(0,1)} = 4.10$, $\bar{k}^{(0,2)} = 0.00$ and $\bar{k}^{(1,2)} = 3.95$ ($p_{\Delta} = 1.0$), and $\beta^{(2)} = 0.15$; $\epsilon_{\beta^{(1)}}^{\text{for.}} \approx 0.06$ and $\epsilon_{\beta^{(1)}}^{\text{back.}} \approx 0.15$ for MECLE, and $\epsilon_{\beta^{(1)}}^{\text{for.}} \approx 0.09$ and $\epsilon_{\beta^{(1)}}^{\text{back.}} \approx 0.47$ for ELE.

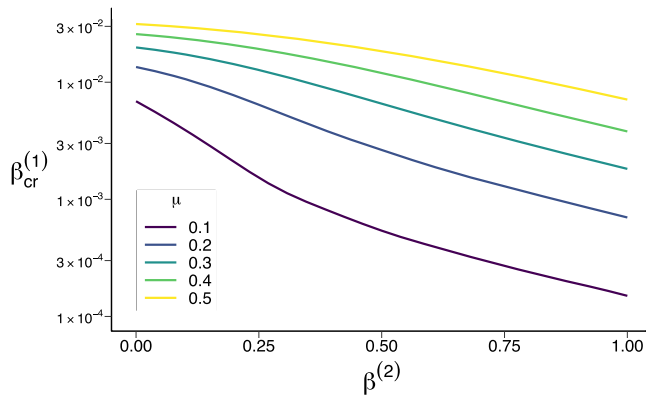


Fig. 4 Dependence of the epidemic threshold on the triangle infection probability $\beta^{(2)}$. $\beta_{cr}^{(1)}$, computed via Eqs. (15)–(19) in Methods, is shown against $\beta^{(2)}$ for a Dorogovtsev-Mendes simplicial complex with $\bar{k}^{(0,1)} = 1.10$, $\bar{k}^{(0,2)} = 0.00$ and $\bar{k}^{(1,2)} = 1.45$, being $\bar{k}^{(g,r)}$ the mean number of $(g, n + 1)$ -cliques incident on a node. Note that Eq. (22), disregarding dynamical correlations, wrongly predicts $\beta_{cr}^{(1)} = \mu/\bar{k} = \mu/4, \forall \beta^{(2)}$.

Supplementary Figs. 3 and 4, we explicitly derive the monotonous decrease of $\beta_{cr}^{(1)}$ with respect to the higher-order infection probabilities for selected symmetrical structures. In particular, we study regular SCs, like the herein studied periodic triangular SC, and SCs built upon Friendship graphs, taken as proxies for homogeneous and heterogeneous structures, respectively. We prove that $\beta_{cr}^{(1)}$ decays with $\beta^{(2)}$ whatever the system size N , and that the dependence considerably increases with the structural heterogeneity.

It should be noted that the herein observed and proved dependence of the critical point on the higher-order couplings, already showed up in previously reported numerical simulations, especially in SCs built from real data³⁰. However, it went seemingly overlooked, eventually leading to claim⁴² that only pair-wise interactions govern the value of $\beta_{cr}^{(1)}$. Even though we have considered a discrete-time dynamic, instead of the continuous-time used in those works upon which that claim is based on, we predict the qualitative shift brought by our analysis to hold in continuous-time as well. The continuous-time limit is here recovered by neglecting all those terms in the equations appearing as second or greater powers of any combination of the infection probabilities $\{\beta^{(s)}\}$ and the recovery probability μ , i.e., allowing only single-node state changes. Still, the linear terms proportional to any of the infection probabilities show up in Eq. (19), thus contributing to the value of the critical point. In Supplementary Note 4, we derive the continuous-time limit of the MECLE equations for simplicial 2-complexes. Taking the example of a Dorogovtsev-Mendes SC, the predicted dependence of the critical point on $\beta^{(2)}$ is shown in Supplementary Fig. 5.

Lastly, considering clique complexes with some fraction p_s of edges shared by two or more 2-faces, we have studied how the MECLE behaves out of the bounds of the 0-connectedness. As shown in Fig. 5, when p_s is low enough, it performs comparably to or still better than the simplicial ELE. As expected, the value of p_s above which the MECLE performs worse turns out to specifically depend on the structure, preventing us to find a simple relation. Precisely, that value resulted to be around 0.06 for Dorogovtsev-Mendes SCs and around 0.04 for RSCs.

Possible approaches for group interactions sharing multiple nodes. In discrete-time models, allowing groups to share two or more nodes comes with the drawback of impeding the

computation of the epidemic threshold, as shown in “Methods”. Forgoing the latter, let us discuss some approaches which may still come in handy.

The first option is the simplicial ELE model³⁶. It describes a SIS dynamics in simplicial 2-complexes by means of a pair approximation that, via a specific triangle closure, is able to partially account for second-order correlations. However, considering the multiple channels of infection within a simplex as mutually uncorrelated, it is not possible to marginalize the 1-faces equations to get the single nodes one, hence neither the epidemic threshold. To further elucidate the effects of disregarding the correlations among the concurrent infections within a simplex, let us consider the toy example of a triangle upon the set of nodes $\{i, j, k\}$. In the MECLE model, the probability for i to be infected by j and k , reads

$$\beta^{(1)} \left(P_{jki}^{IS|S} + P_{jki}^{SI|S} \right) + \left[1 - (1 - \beta^{(1)})^2 (1 - \beta^{(2)}) \right] P_{jki}^{II|S}$$

In particular, taking $\beta^{(1)} = 1$, the dependence on $\beta^{(2)}$ correctly disappears: no matter the infectiousness $\beta^{(2)}$ of the couple $\{j, k\}$, the probability for node i to be infected equals the probability that at least one node between j and k is infected. On the contrary, the simplicial ELE neglects that correlation, and the above infection probability becomes

$$1 - \left(1 - \beta^{(1)} P_{jii}^{II|S} \right) \left(1 - \beta^{(1)} P_{kii}^{II|S} \right) \left(1 - \beta^{(2)} P_{jki}^{II|S} \right)$$

where $P_{jki}^{II|S}$ is expressed in terms of edge probabilities. Here, taking $\beta^{(1)} = 1$, there is still a dependence on $\beta^{(2)}$, with the effect of both overestimating the probability of infection, hence the prevalence, and wrongly anticipating the position of the epidemic threshold. Evidently, the error grows with both $\beta^{(2)}$ and $P_{jki}^{II|S}$. To notice that this analysis does not make any reference to the correlations among nodes’ states a model considers. Consequently, a similar comparison holds also between the MMCA and MF approximations of the MECLE and those of the simplicial ELE, with the latter overestimating more the prevalence.

Curiously, when the 0-connectedness is heavily broken (making the MECLE unreliable), the simplicial ELE appears to gain accuracy, especially in locating the critical point³⁶. However, while we could explain the reasons why the MECLE outperforms the simplicial ELE in (nearly) 0-connected SCs, the unexpected improvement of the latter for higher connectedness remains unclear. Changes in topological factors, like the spectral dimension (note that, independently from their geometrical dimension, 0-connected SCs have spectral dimension very close to that of a random tree⁴³) or the triangles’ percolation, are probably able to soften the approximations of the model. Future work addressing the role of those factors may help to better understand the limits of the simplicial ELE, while giving new, general insights about the relation between dynamical correlations and topology.

An alternative approach, assuming the considered contagion would make it usable, is to opt for more general hypergraphs by dropping the hereditary property out. One could then generalize the approach of Matamalas et al.³⁶ to any hypergraph, but at the aforementioned cost of accepting some inconsistencies in the marginalization of the probabilities. Besides, referring to the class of simple hypergraphs¹⁹, i.e., those in which a hyperedge—what in a SC is a face—cannot be a subset of any other hyperedge, one can still resort to the fully consistent approach of the MECLE. Indeed, in such structures, to each configuration of the group corresponds a unique channel of infection, so the necessity of constraining groups to share not more than a single node can be relaxed. A model can then be easily constructed adapting Eq. (2) by solely modifying the form of $w_{i,1}^{(g)}$, being $g = 1$ for hyperedges

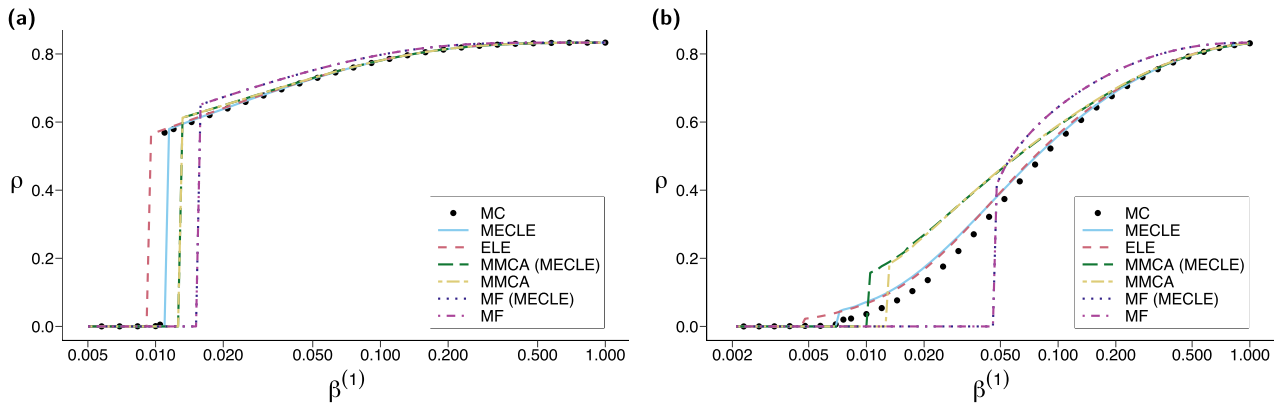


Fig. 5 Epidemic prevalence ρ as function of the link infection probability $\beta^{(1)}$ on 2-dimensional simplicial complexes (SCs) with low percentage p_s of shared non-maximal edges. p_s is computed as the fraction of edges within 2-faces which are included in more than one 2-face. Besides, with \bar{s} we indicate the average number of 2-faces in which the edges corresponding to the fraction p_s are included ($\bar{s} \geq 2$). Results obtained from Monte Carlo (MC) simulations are depicted by dots, while lines represent the analytically computed prevalence using the indicated models. MMCA and MMCA(MECL) refer to the Microscopic Markov Chain approximation of, respectively, the simplicial Epidemic Link Equations (ELE) model and the Microscopic Epidemic Clique Equations (MECLE) model, as obtained by considering the state probabilities of the nodes as uncorrelated; while MF and MF(MECL) refer to their homogeneous mean-field approximations (see “Methods”). Note that MF and MF(MECL) are indistinguishable at the used scale. The recovery probability is fixed to $\mu = 0.2$. **a** Random SC with $\bar{k}^{(0,1)} = 2.25$, $\bar{k}^{(0,2)} = 0.00$ and $\bar{k}^{(1,2)} = 5.20$, being $\bar{k}^{(g,r)}$ the mean number of $(g, n + 1)$ -cliques incident on a node; $p_s = 0.03$, $\bar{s} = 2.01$, and triangle infection probability $\beta^{(2)} = 0.15$. **b** Dorogovtsev-Mendes SC with $\bar{k}^{(0,1)} = 1.10$, $\bar{k}^{(0,2)} = 0.00$ and $\bar{k}^{(1,2)} = 1.60$, $p_s = 0.05$, $\bar{s} = 2.06$, and $\beta^{(2)} = 0.25$.

including more than two nodes. For example, supposing the contagion to be maximally conservative, we would write $w_{l,l}^{(r)} = \mathbb{1}[l = r](1 - \beta^{(r)})$. Moreover, for linear hypergraphs¹⁹, in which two hyperedges can only share one node, the critical point is again calculable.

Lastly, in a regime of both high and rapid infectiousness and recovery, each set of nodes shared by a sufficient number of groups (e.g., two partners carrying out many of their activities together) could be effectively treated as a single super-node, whose state always represents that of each of the nodes it contains. This effective approach can be combined with any of the ones that have been discussed.

Methods

Epidemic threshold. Here we derive the critical point $\beta_{cr}^{(1)}$, defined as the value of $\beta^{(1)}$ at which the inactive (epidemic-free) state becomes unstable, thus marking the onset of the active (endemic) state. In the presence of a bi-stable region, it identifies the rightmost transition.

We linearize Eq. (2) by regarding of the same order $\epsilon \ll 1$, all the state probabilities containing at least one infected node, i.e., $P_{i_1 \dots i_n, g}^{S \dots S} \sim \mathcal{O}(\epsilon)$ iff $\exists k : \sigma_k = I$; and consequently, $P_{i_1 \dots i_n, g}^{S \dots S} \sim 1 - \mathcal{O}(\epsilon)$. Without this assumption, higher-order dynamical correlations would be lost, and the critical point would not be correctly located. This can be interpreted as an extension of the results in Matamalas et al.³⁹, where it is shown that $P^I \sim \mathcal{O}(\epsilon)$ for the state probability of an edge having both nodes infected.

Being interested in stationary states, the value of the time step is omitted from now on. All the states appearing in $q_{i_k, g'}^{(r)}$, see Eq. (6), include at least one infected node, therefore it takes the form,

$$q_{i_k, g'}^{(r)} = 1 - \sum_{\substack{\{i_1, \dots, i_r\} \\ \in \mathcal{I}_{i_k, g'}^{(r)}}} \left[\sum_{l=1}^r \frac{1 - w_{l, g'}^{(r)}}{\ell!(r - \ell)!} \times \sum_{k_1 \dots k_r \neq k_r=1}^r P_{j_{k_1} \dots j_{k_r}, i_{k_1+1} \dots i_{k_r}, i_k, g'}^{I \dots IS \dots SS} \right] + \mathcal{O}(\epsilon^2) \tag{10}$$

where the squared brackets contain $\mathcal{O}(\epsilon)$ terms only. It is important to remark that, if in the considered clique there is at least another node $i_{\bar{k}}$ forming with i_k an edge included in some other clique, let us say a (g', r) -clique, $q_{i_k, g'}^{(r)}$ cannot be linearized, since the product corresponding to that (g', r) -clique would be made of state probabilities conditioned to the state of both i_k and $i_{\bar{k}}$ (not of i_k only, as in Eq. (10)). Indeed, those states in which $i_{\bar{k}}$ is in state I , would give $\mathcal{O}(1)$ terms, for both

numerator and denominator would be $\mathcal{O}(\epsilon)$. This is the reason why in markovian models, when aiming to account for the dynamical correlations within some subsets of nodes (e.g., cliques), a consistent expression for the critical point can be given only when those subsets are edge-disjoint.

Returning to the derivation, since $P_{i_1 \dots i_n, g}^{S \dots S}$ is fixed by the normalization condition,

$$P_{i_1 \dots i_n, g}^{S \dots S} = 1 - \sum_{\substack{\{\sigma_1, \dots, \sigma_n\} \\ \neq \{S, \dots, S\}}} P_{i_1 \dots i_n, g}^{\sigma_1 \dots \sigma_n} \tag{11}$$

we only need the linearized equations for arrival states with at least one infected node, i.e., $\{\sigma'_1, \dots, \sigma'_n\} \neq \{S, \dots, S\}$. Retaining the $\mathcal{O}(\epsilon)$ terms in $\phi_{i_k, g}$, after some algebra, we find

$$P_{i_1 \dots i_n, g}^{\sigma'_1 \dots \sigma'_n} = \sum_{\substack{\{\sigma_1, \dots, \sigma_n\} \\ \neq \{S, \dots, S\}}} P_{i_1 \dots i_n, g}^{\sigma_1 \dots \sigma_n} \left[(1 - \mu)^{N_{I \rightarrow I}} \mu^{N_{I \rightarrow S}} (1 - w_{N_{I, g}}^{(n-1)})^{N_{S \rightarrow I}} (w_{N_{I, g}}^{(n-1)})^{N_{S \rightarrow S}} \right] + \mathbb{1}[\exists k : \sigma'_k = I] \left\{ \sum_{(g', r) \neq (g, n-1)} \sum_{\substack{\{i_1, \dots, i_r\} \\ \in \mathcal{I}_{i_k, g'}^{(r)}}} \left[\sum_{l=1}^r \frac{1 - w_{l, g'}^{(r)}}{\ell!(r - \ell)!} \sum_{k_1 \dots k_r \neq k_r=1}^r P_{j_{k_1} \dots j_{k_r}, i_{k_1+1} \dots i_{k_r}, i_k, g'}^{I \dots IS \dots SS} \right] \right\} + \sum_{\substack{\{i_1, \dots, i_{n-1}\} \\ \in \mathcal{I}_{i_k, g}^{(n-1)} \setminus \{i_k\}}} \left[\sum_{l=1}^{n-1} \frac{1 - w_{l, g}^{(n-1)}}{\ell!(n-1 - \ell)!} \sum_{k_1 \dots k_{l-1} \neq k_{l-1}=1}^{l-1} P_{j_{k_1} \dots j_{k_{l-1}}, i_{k_1+1} \dots i_{k_{l-1}}, i_k, g}^{I \dots IS \dots SS} \right] \right\} + \mathcal{O}(\epsilon^2) \tag{12}$$

where $N_{\sigma \rightarrow \sigma'} = |\{i_{k=1, \dots, n} | \sigma_k = \sigma, \sigma'_k = \sigma'\}|$ is the number of nodes going from state σ to state σ' . The terms in curly brackets derive from those transitions starting in state $\{S, \dots, S\}$ and arriving to a state with exactly one infected node.

In particular, Eq. (8), the dynamic equation for a single node, becomes

$$P_i^I = P_i^I(1 - \mu) + \sum_{(g, r)} \sum_{\substack{\{i_1, \dots, i_r\} \\ \in \mathcal{I}_{i, g}^{(r)}}} \left[\sum_{l=1}^r \frac{1 - w_{l, g}^{(r)}}{\ell!(r - \ell)!} \times \sum_{k_1 \dots k_r \neq k_r=1}^r P_{j_{k_1} \dots j_{k_r}, i_{k_1+1} \dots i_{k_r}, i, g}^{I \dots IS \dots SS} \right] + \mathcal{O}(\epsilon^2) \tag{13}$$

At this point, to get an expression for the critical threshold, we put Eq. (13) in the form of an eigenvalue equation for the vector of single-node probabilities $\mathbf{P}^I = (\{P_i^I\}_{i \in V})$. To this purpose, given a (\cdot, n) -clique $\{i_1, \dots, i_n\}$, we need to express every joint probability over its nodes states as a linear combination of the marginal probabilities, $P_{i_1}^I, \dots, P_{i_n}^I$. Eq. (12) provides a linearized equation for each of the $2^n - 1$ unknown state probabilities, hence a system admitting a unique solution. At this point, instead of the n equations for the transition to a state with a single node in state I (second term in Eq. (12)), we use the n consistency relations for the marginal probabilities, i.e., $P_{i_k}^I = \sum_{\{\sigma_{i_k}\}} P_{i_k, \{i_{\bar{k}}\}, g}^{I \dots \sigma_{i_k}}$, $\forall k \in \{1, \dots, n\}$. In this way, the

system is still made of $2^n - 1$ equations, hence is determined, but now it includes the marginal probabilities. Eventually, with some algebra, one gets the decomposition with its linear coefficients. Alternatively, asserted the uniqueness of the solution of the linear system, the problem can be approached in the other way around. That is, firstly expressing each of the joint probabilities as the most general linear combination of the marginal probabilities and then inserting them into the $2^n - 1$ equations. Doing so, we get a new, determined linear system whose unknowns are the linear coefficients of the original system. In the end, given a (g, n) -clique $\{i_1, \dots, i_n\}$, the most general and proper linear decomposition of $P_{i_1, \dots, i_n}^{\sigma_{i_1}, \dots, \sigma_{i_n}}$ in terms of $P_{i_1}^l, \dots, P_{i_n}^l$, reads

$$P_{i_1, \dots, i_n}^{\sigma_{i_1}, \dots, \sigma_{i_n}} = X_{N_I, g}^{(n-1)} \sum_{k: \sigma_{i_k}=l} P_{i_k}^l + Y_{N_I, g}^{(n-1)} \sum_{k: \sigma_{i_k}=S} P_{i_k}^l + \mathcal{O}(\epsilon^2) \quad (14)$$

where we can take $Y_{n, g}^{(n-1)} = 0$, being null the term it multiplies, as there are no nodes in state S for $N_I = n$. We get one coefficient for $N_I = n$ and two coefficients for each $N_I \in \{1, \dots, n-1\}$, leading to $2n - 1$ of them in total. Summing for every order n from 2 to m_0 for $g = 0$, and from 3 to m_1 for $g = 1$, we get a maximum of $m_0^2 + m_1^2 - 5$ linear coefficients to fix. These coefficients, as functions of all the m_1 microscopic parameters of the model, $\{\beta^{(s)}\}_{s=1, \dots, m_1-1}$ and μ , weigh the probability of finding a clique in a given state, when the system approaches the critical point.

Once all the coefficients have been found by insertion of Eq. (14) in the original linear system of $2^n - 1$ equations, we substitute them in Eq. (13) to finally get an eigenvalue equation. To this end, we define the set of $\mathcal{D}(\mathcal{K})$ -dependent adjacency matrices $\left\{ \{A^{(0,r)}\}_{r \in \{1, \dots, m_0-1\}}, \{A^{(1,r)}\}_{r \in \{2, \dots, m_1-1\}} \right\}$, such that $A_{ij}^{(g,r)}$ equals 1 if nodes i and j share a common incident $(g, r+1)$ -clique in $\mathcal{D}(\mathcal{K})$, and 0 otherwise. Besides, we define the (g, r) degree of node i , $k_i^{(g,r)}$, as the number of $(g, r+1)$ -cliques incident on node i in $\mathcal{D}(\mathcal{K})$, computed as $k_i^{(g,r)} = \sum_{j=1}^N A_{ij}^{(g,r)}$. Being the decomposition edge-disjoint, only one of those matrices can have a non-zero element in the position corresponding to a given pair of nodes. Consequently, the $\mathcal{D}(\mathcal{K})$ -independent adjacency matrix of $\mathcal{K}^{(1)}$, the underlying graph of \mathcal{K} , is simply obtained by the sum of all those $\mathcal{D}(\mathcal{K})$ -dependent adjacency matrices. It also follows that the degree k_i of node i in $\mathcal{K}^{(1)}$ is computed as $k_i = k_i^{(0)} + k_i^{(1)}$, where $k_i^{(0)} = \sum_{r=1}^{m_0} r k_i^{(0,r)}$ and $k_i^{(1)} = \sum_{r=2}^{m_1} r k_i^{(1,r)}$ are the total number of neighbors of i within, respectively, $(0, \cdot)$ -cliques and $(1, \cdot)$ -cliques.

Substituting Eq. (14) in Eq. (13), and doing some algebra and combinatorics, we get

$$MP^l = DP^l \quad (15)$$

where we have defined the matrices M and D , of elements

$$M_{ij} = \sum_{(g,r)} A_{ij}^{(g,r)} \sum_{l=1}^r \left(1 - w_{l,g}^{(r)}\right) \left[\binom{r-1}{l-1} X_{l,g}^{(r)} + \binom{r-1}{r-l-1} Y_{l,g}^{(r)} \right] \quad (16)$$

$$D_{ij} = \delta_{ij} \left[\mu - \sum_{(g,r)} k_i^{(g,r)} \sum_{l=1}^r \binom{r}{l} \left(1 - w_{l,g}^{(r)}\right) Y_{l,g}^{(r)} \right] \quad (17)$$

being δ_{ij} the elements of the $N \times N$ identity matrix. Equations (15)–(17) define a generalized eigenvalue problem. The form taken by M and D is easily understood in this way. Each sum multiplying $A_{ij}^{(g,r)}$ in M_{ij} represents the marginal contribution to the infection of node i coming from its neighbor j through the $(g, r+1)$ -clique they share. Given j in state I , $(0, 0)ptr - 1a$ is the number of ways in which a out of $r-1$ nodes can be chosen to be in state I ($a = l-1$) or S ($a = r-l-1$). Similarly, each sum multiplying $k_i^{(g,r)}$ in D_{ij} represents the contribution coming from all the configurations of any of the $(g, r+1)$ -cliques incident on node i , in which i is in state S . $(0, 0)ptrl$ is the number of ways in which l out of r nodes can be chosen to be in state I .

Now, M is non-negative. Indeed, for any fixed (g, r) , the sum multiplying $A_{ij}^{(g,r)}$ must be positive whenever $A_{ij}^{(g,r)} > 0$, since it represents the by-definition positive contribution to the infection of a node (i) coming from one of its neighbors (j). Moreover, being \mathcal{K} undirected and so M symmetric, it follows that M is also irreducible⁴⁴. Looking now at D , which is diagonal, the non-negativity of M implies the diagonal elements of D to be positive. Indeed, Eq. (15) holds for any value of the microscopic parameters; so let us suppose $\mu = 0$. Since the non-zero elements of M are positive $\forall \mu \in [0, 1]$, the sum multiplying $k_i^{(g,r)}$ in D_{ij} must be negative whenever $A_{ij}^{(g,r)} > 0$, proving any diagonal element of D to be positive. Therefore, D is invertible and its inverse as well, with elements $[D^{-1}]_{ii} = (D_{ii})^{-1}, \forall i = 1, \dots, N$. Applying D^{-1} to both sides of Eq. (15), we finally get the sought eigenvalue equation,

$$M'P^l = P^l \quad (18)$$

where we have defined the matrix $M' \equiv D^{-1}M$, of elements $M'_{ij} = M_{ij}/D_{ii}$. Thus, M' is a non-negative irreducible matrix as well and, by the Perron-Frobenius theorem⁴⁴, it admits a unique leading eigenvector P^l . This is the only one associated with the largest eigenvalue $\Lambda_{\max}(M')$ and the only one with all its entries

positive, hence representing the unique physically acceptable expected state of the system at the onset of the epidemic. Fixed the values of the recovery probability, μ , and the higher-order infection probabilities, $\{\beta^{(s)}\}_{s>1}$, the critical threshold $\beta_{cr}^{(1)}$ is implicitly found as the smallest non-negative value of $\beta^{(1)}$ such that

$$\Lambda_{\max}(M') = 1 \quad (19)$$

Mean-field approximation. The homogeneous mean-field (MF) approximation of Eq. (8) is found by neglecting both the state correlations and the local structural heterogeneity among the nodes, i.e., regarding every node as the "average node" in the structure⁴⁵. Given any (g, r) -clique $\{j_1, \dots, j_r\}$, it follows $P_{j_1, \dots, j_r}^{I, \dots, I, S, \dots, S} = \rho^l (1 - \rho)^{r-l}$, and, for any node i , $k_i^{(g,r)} = \bar{k}^{(g,r)}$, where $\bar{k}^{(g,r)}$ is the average value of the (g, r) -degree of the nodes in the structure. Thus, Eq. (8) becomes

$$\rho(t+1) = \rho(t)(1 - \mu) + (1 - \rho(t)) \left(1 - \prod_{(g,r)} \bar{q}_g^{(r)} \right) \quad (20)$$

where

$$\bar{q}_g^{(r)} = \left[1 - \sum_{l=1}^r \binom{r}{l} \left(1 - w_{l,g}^{(r)}\right) \rho^l (1 - \rho)^{r-l} \right]^{\bar{k}^{(g,r)}} \quad (21)$$

The stationary solution is then got imposing $\rho(t+1) = \rho(t)$ in Eq. (20). The linearization around the epidemic-free state is then implemented by taking $\rho = \epsilon \ll 1$. Looking at Eq. (21), the only $\mathcal{O}(\epsilon)$ terms are given by $l=1$, whatever the couple (g, r) . That is, only the pair-wise probability $\beta^{(1)}$ contributes in the MF approximation. With few algebra, one gets the renowned formula

$$\beta_{cr}^{(1)} = \frac{\mu}{\bar{k}} \quad (22)$$

where $\bar{k} = \frac{1}{N} \sum_{i=1}^N k_i$, is the average degree of a node in $\mathcal{K}^{(1)}$.

More generally, Eq. (22) holds for any model treating the nodes states as independent, thus including the MMCA approximation of the MECLC, the MMCA and MF approximations of the simplicial ELE, and also, except for substituting \bar{k} with \bar{k}^2/\bar{k} , the heterogeneous MF approximation⁴⁵ of both. The same result is found in continuous time³¹.

Numerical simulations. The equilibrium value of the prevalence ρ is computed using synchronous Monte Carlo simulations and the quasistationary state (QS) method⁴⁶. In the specific case of the simplicial 2-complexes, for each node $i \in V$, a simulated time step proceeds as follows: (1) if i is currently infected, it recovers with probability μ ; (2) if i is currently susceptible, (2.1) it gets infected with

probability $1 - (1 - \beta^{(1)})^{n_i^{(1)}}$, being $n_i^{(1)}$ the number of currently infected neighbors of i through edges; (2.2) if i is still susceptible after sub-step (2.1), it gets infected with probability $1 - (1 - \beta^{(2)})^{n_i^{(2)}}$, being $n_i^{(2)}$ the number of currently infected couples of neighbors of i through triangles. For higher dimensions, $n > 2$, step (2) consists of $n-2$ additional sub-steps, analogously defined and ordered as shown here.

In accordance with the QS method, every time the absorbing state $\rho = 0$ is reached, it is replaced by one of the previously stored active states of the system, i.e., one of those states with at least an active individual. Since for finite systems, when approaching the critical point, a large number of realizations end up in the absorbing state, the QS method properly reduces to a single run the wasteful method of performing many simulations. We have made use of 50 stored active states and an update probability of 0.25. We have given the systems a transient time of 10^5 time steps, and then calculated ρ as an average over 2×10^4 additional time steps.

Connectivity structures. We have referred to various interaction structures. For the numerical evaluation, we have made use of synthetic simplicial 2-complexes with around $N = 10^4$ nodes, presenting dissimilar structural properties: regular structures, as the clique complexes built from a triangular lattice; homogeneous structures, generated from the random SC model³⁰, in which both the $(\cdot, 1)$ - and the $(\cdot, 2)$ -degree follow a Poisson distribution; heterogeneous structures, derived from the Dorogovtsev-Mendes model⁴¹, having $(\cdot, 1)$ - and $(\cdot, 2)$ -degree distributions nearly following power-laws with exponent around 3.

Data availability

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Code availability

The code for estimating a minimal edge-disjoint edge clique cover (EECC) of a graph has been implemented as the DisjointCliqueCover.jl⁴⁷ package for the Julia language, available at [github](https://github.com) and archived at zenodo.org.

Received: 11 January 2021; Accepted: 4 May 2021;
Published online: 31 May 2021

References

- Anderson, R. M. & May, R. M. *Infectious Diseases Of Humans: Dynamics and Control* (Oxford university press, 1992).
- Daley, D. J. & Kendall, D. G. Stochastic rumours. *IMA J. Appl. Math.* **1**, 42–55 (1965).
- Rogers, E. M. *Diffusion of Innovations* (Simon and Schuster, 2010).
- Katz, E. & Lazarsfeld, P. F. *Personal Influence, The Part Played by People in the Flow of Mass Communications* (Transaction Publishers, 1966).
- Valente, T. W. *Network Models of the Diffusion of Innovations* (Hampton Press, 1995).
- Cowan, R. & Jonard, N. Network structure and the diffusion of knowledge. *J. Econ. Dyn. Control* **28**, 1557–1575 (2004).
- Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200 (2001).
- Watts, D. J. & Dodds, P. S. Influentials, networks, and public opinion formation. *J. Consum. Res.* **34**, 441–458 (2007).
- Watts, D. J. & Strogatz, S. H. Collective dynamics of ‘small-world’ networks. *Nature* **393**, 440–442 (1998).
- Bianconi, G., Darst, R. K., Iacovacci, J. & Fortunato, S. Triadic closure as a basic generating mechanism of communities in complex networks. *Phys. Rev. E* **90**, 042806 (2014).
- Palla, G., Derényi, I., Farkas, I. & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814–818 (2005).
- Battiston, F. et al. Networks beyond pairwise interactions: structure and dynamics. *Phys. Rep.* **874**, 1(2020).
- Burgio, G., Matamalas, J. T., Gómez, S. & Arenas, A. Evolution of cooperation in the presence of higher-order interactions: from networks to hypergraphs. *Entropy* **22**, 744 (2020).
- Dai, X. et al. D-dimensional oscillators in simplicial structures: odd and even dimensions display different synchronization scenarios. Preprint at <https://arxiv.org/abs/2010.14976> (2020).
- Ghorbanchian, R., Restrepo, J. G., Torres, J. J. & Bianconi, G. Higher-order simplicial synchronization of coupled topological signals. Preprint at <https://arxiv.org/abs/2011.00897> (2020).
- Tadić, B. & Gupte, N. Hidden geometry and dynamics of complex networks: spin reversal in nanoassemblies with pairwise and triangle-based interactions (a). *Europhys. Lett.* **132**, 60008 (2021).
- Sun, H., Ziff, R. M. & Bianconi, G. Renormalization group theory of percolation on pseudofractal simplicial and cell complexes. *Phys. Rev. E* **102**, 012308 (2020).
- Andjelković, M., Tadić, B. & Melnik, R. The topology of higher-order complexes associated with brain hubs in human connectomes. *Sci. Rep.* **10**, 1–10 (2020).
- Bretto, A. *Hypergraph Theory. An Introduction*. (Springer, 2013).
- Lambiotte, R., Rosvall, M. & Scholtes, I. From networks to optimal higher-order models of complex systems. *Nat. Phys.* **15**, 313–320 (2019).
- Jonsson, J. *Simplicial Complexes of Graphs* (Springer Science & Business Media, 2007).
- Granovetter, M. Threshold models of collective behavior. *Am. J. Sociol.* **83**, 1420–1443 (1978).
- Centola, D. & Macy, M. Complex contagions and the weakness of long ties. *Am. J. Sociol.* **113**, 702–734 (2007).
- Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010).
- Melnik, S., Ward, J. A., Gleeson, J. P. & Porter, M. A. Multi-stage complex contagions. *Chaos* **23**, 013124 (2013).
- Watts, D. J. A simple model of global cascades on random networks. *Proc. Natl Acad. Sci. USA* **99**, 5766–5771 (2002).
- Guilbeault, D., Becker, J. & Centola, D. Complex contagions: A decade in review. In *Complex spreading phenomena in social systems*, 3–25 (Springer, 2018).
- Ugander, J., Backstrom, L., Marlow, C. & Kleinberg, J. Structural diversity in social contagion. *Proc. Natl Acad. Sci. USA* **109**, 5962–5966 (2012).
- Weng, L., Flammini, A., Vespignani, A. & Menczer, F. Competition among memes in a world with limited attention. *Sci. Rep.* **2**, 335 (2012).
- Iacopini, I., Petri, G., Barrat, A. & Latora, V. Simplicial models of social contagion. *Nat. Commun.* **10**, 1–9 (2019).
- Landry, N. W. & Restrepo, J. G. The effect of heterogeneity on hypergraph contagion models. *Chaos* **30**, 103117 (2020).
- St-Onge, G., Sun, H., Allard, A., Hébert-Dufresne, L. & Bianconi, G. Bursty exposure on higher-order networks leads to nonlinear infection kernels. Preprint at <https://arxiv.org/abs/2101.07229> (2021).
- Bodó, Á., Katona, G. Y. & Simon, P. L. Sis epidemic propagation on hypergraphs. *Bull. Math. Biol.* **78**, 713–735 (2016).
- Jhun, B., Jo, M. & Kahng, B. Simplicial sis model in scale-free uniform hypergraph. *J. Stat. Mech. Theory Exp.* **2019**, 123207 (2019).
- de Arruda, G. F., Petri, G. & Moreno, Y. Social contagion models on hypergraphs. *Phys. Rev. Res.* **2**, 023032 (2020).
- Matamalas, J. T., Gómez, S. & Arenas, A. Abrupt phase transition of epidemic spreading in simplicial complexes. *Phys. Rev. Res.* **2**, 012049 (2020).
- Kou, L. T., Stockmeyer, L. J. & Wong, C.-K. Covering edges by cliques with regard to keyword conflicts and intersection graphs. *Commun. ACM* **21**, 135–139 (1978).
- Henkel, M., Hinrichsen, H. & Lübeck, S. *Non-equilibrium Phase Transitions: Volume 1: Absorbing Phase Transitions* (Springer Science & Business Media, 2008).
- Matamalas, J. T., Arenas, A. & Gómez, S. Effective approach to epidemic containment using link equations in complex networks. *Sci. Adv.* **4**, eaau4212 (2018).
- Gómez, S., Arenas, A., Borge-Holthoefer, J., Meloni, S. & Moreno, Y. Discrete-time markov chain approach to contact-based disease spreading in complex networks. *Europhys. Lett.* **89**, 38009 (2010).
- Dorogovtsev, S. N., Mendes, J. F. & Samukhin, A. N. Size-dependent degree distribution of a scale-free growing network. *Phys. Rev. E* **63**, 062101 (2001).
- Barrat, A., de Arruda, G. F., Iacopini, I. & Moreno, Y. Social contagion on higher-order structures. <https://arxiv.org/abs/2103.03709> (2021).
- Dankulov, M. M., Tadić, B. & Melnik, R. Spectral properties of hyperbolic nanonetworks with tunable aggregation of simplexes. *Phys. Rev. E* **100**, 012309 (2019).
- Meyer, C. D. *Matrix Analysis and Applied Linear Algebra*, vol. 71 (SIAM, 2000).
- Gómez, S., Gómez-Gardenes, J., Moreno, Y. & Arenas, A. Nonperturbative heterogeneous mean-field approach to epidemic spreading in complex networks. *Phys. Rev. E* **84**, 036105 (2011).
- Ferreira, S. C., Castellano, C. & Pastor-Satorras, R. Epidemic thresholds of the susceptible-infected-susceptible model on networks: a comparison of numerical and theoretical results. *Phys. Rev. E* **86**, 041125 (2012).
- Burgio, G., Arenas, A., Gómez, S. & Matamalas, J. T. DisjointCliqueCover.jl: v0.1. <https://doi.org/10.5281/zenodo.4723748> (2021).

Acknowledgements

G.B. acknowledges financial support from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 945413. We acknowledge support by Ministerio de Economía y Competitividad (PGC2018-094754-BC21, FIS2017-90782-REDT and RED2018-102518-T), Generalitat de Catalunya (2017SGR-896 and 2020PANE00098), and Universitat Rovira i Virgili (2019PFR-URV-B2-41). A.A. acknowledges also ICREA Academia and the James S. McDonnell Foundation (220020325).

Author contributions

G.B., A.A., S.G. and J.T.M. designed the research and wrote the manuscript. G.B. and J.T.M. made the analytical calculations. G.B. performed the numerical analysis and simulations.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42005-021-00618-z>.

Correspondence and requests for materials should be addressed to S.G.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021