Mean-field models of dynamics on networks via moment closure: An automated procedure

Bert Wuyts [*] and Jan Sieber

College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QF, United Kingdom

(Received 12 November 2021; revised 24 October 2022; accepted 25 October 2022; published 28 November 2022)

In the study of dynamics on networks, moment closure is a commonly used method to obtain low-dimensional evolution equations amenable to analysis. The variables in the evolution equations are mean counts of subgraph states and are referred to as moments. Due to interaction between neighbors, each moment equation is a function of higher-order moments, such that an infinite hierarchy of equations arises. Hence, the derivation requires truncation at a given order and an approximation of the highest-order moments in terms of lower-order ones, known as a closure formula. Recent systematic approximations have either restricted focus to closed moment equations for SIR epidemic spreading or to unclosed moment equations for arbitrary dynamics. In this paper, we develop a general procedure that automates both derivation and closure of arbitrary order moment equations for dynamics with nearest-neighbor interactions on undirected networks. Automation of the closure step was made possible by our generalized closure scheme, which systematically decomposes the largest subgraphs into their smaller components. We show that this decomposition is exact if these components form a tree, there is independence at distances beyond their graph diameter, and there is spatial homogeneity. Testing our method for SIS epidemic spreading on lattices and random networks confirms that biases are larger for networks with many short cycles in regimes with long-range dependence. A `Mathematica` package that automates the moment closure is available for download.

DOI: 10.1103/PhysRevE.106.054312

I. INTRODUCTION

The dynamics of complex systems are usually most accurately represented by high-dimensional stochastic simulation models. However, their large state space makes exact mathematical analysis prohibitive. Therefore, one often looks for low-dimensional approximations that permit analysis. In *moment closure*, one achieves this by studying the time evolution of a finite set of "moments" rather than that of the full probability distribution of the considered stochastic dynamical system [1]. The complete set of moment equations forms an infinite hierarchy of ordinary differential equations (ODEs), with lower-order moments depending on higher-order ones. The approximation consists then of truncating the hierarchy at a chosen order and replacing the highest order moments by functions of the lower-order moments. Such functions are known as closure formulas, and they can be obtained in various ways [1], such as via an assumption of statistical independence [e.g., 2–4], physical principles (e.g., maximum entropy [5]), timescale separation [6], and assumptions on the type of probability distribution [e.g., 7]. In this work, we only consider closures derived from an assumption of statistical independence, which is equivalent to a *mean-field approximation*, a method originating from the statistical physics of phase transitions in materials [8–11]. First-order moment closure assumes pairwise independence of species counts and corresponds to the "mean field" or "simple mean field" [e.g., 12–14], resulting in equations only for total counts of each species. Likewise, second order moment closure assumes independence of pair counts in larger units, and corresponds to the "pair approximation" [e.g., 15–20], which also includes equations for pair counts. We aim to exploit this connection between moment closure and mean-field approximations to generalize and automate the derivation of arbitrary-order mean-field models for arbitrary dynamics with nearest-neighbor interactions on undirected networks. In the rest of our introduction, we introduce our approach with more precision, discuss the relevant literature, state our aims, and provide an overview of the paper contents.

In general, moments in the moment closure for dynamics on networks represent the expected frequencies of small subgraph states known as *network motifs* [21]. Derivation of the moment equations proceeds from smaller- to larger-sized motifs, with dynamics of mean motif counts of size m depending only on mean motif counts of size m and $m+1$ if the dynamics has only nearest-neighbor interactions. Hence, a system of ODEs obtained in such a manner for motif counts up to a maximum considered size k (also referred to as the order of the moment closure) is always underdetermined, because it depends on motifs of size $k+1$ but does not contain equations for them. Therefore, as the second step of moment closure, a closure approximation is applied by expressing counts of $(k+1)$-size motifs as functions of counts of $\{1, \ldots, k\}$-size motifs, closing the system of ODEs. In this substitution, larger-sized motifs factorize in terms of smaller-sized ones, which we will justify, as mentioned above, by an assumption of statistical independence. For homogeneous networks, closures that are valid at the individual level (i.e., concerning states of given nodes) are also valid at the population level (i.e., concerning total counts or averages of states

[*]b.wuyts@ex.ac.uk

in the whole network) [22], which then permits a compact description in terms of population-level quantities. Yet, the number of motif types, and hence equations, increases combinatorially with k. It is then hoped that the derivation can be stopped at an order low enough for the resulting system of ODEs to be sufficiently amenable to analytical or numerical methods and high enough to satisfy the independence assumptions underlying the closure approximately. We note that various other types of approximations exist that focus on specific types of subgraphs, such as active motifs [23], star graphs [24,25], or hypergraphs of cliques [26,27] or of general motifs [28]. These have also been referred to as moment closure [e.g., in 29], but their truncation order and closure formulas are implicit in the method. At present, the equations obtained by the approach of references [24–28] are referred to as approximate master equations.

In the context of population dynamics on networks or lattices, moment closure methods have been used to study applications such as spatial ecology [e.g., 15,18,19], epidemics [e.g., 16,17,20,21,30,31], opinion formation [e.g., 29], evolution of cooperation [e.g., 32], among others. Despite the wide use of moment closure in applications, derivation of the moment equations is customarily done separately for each considered process and approximation order, while justifying the used closure formulas only heuristically. Comparatively few studies have shown how moment equations derive in general from the master equation or how closure formulas arise from precisely defined independence assumptions. Regarding the former, an automated algorithm to derive (unclosed) moment equations for arbitrary adaptive dynamics on directed networks was recently developed by Danos *et al.* [33]. Regarding the latter, attention has centered on the specific case of SIR epidemic spreading [2–4] because proving validity of low-order closures is least challenging here. In particular, Sharkey *et al.* [3] proved that an exact individual-level closure approximation exists for motifs that have an all-susceptible set of nodes which cuts all possible chains of infection between the remaining parts when removed. In tree networks, this is already possible with three nodes such that the largest required motif in the moment equations is of size 2. In case of nontree networks, larger motifs need to be taken into account, resulting in a larger number of equations. Hence, while Danos *et al.* [33] have shown that it is feasible to derive moment equations in a generic form, the work on SIR spreading [2–4] indicates which type of independence assumptions are required to obtain valid closures.

In this paper, we provide a first *fully automated* procedure for both derivation and closure of population-level moment equations up to any order and for arbitrary dynamics with at most nearest-neighbor interactions on undirected networks. Automated derivation is made possible by our generic moment equation (6), which we derived from the master equation. As mentioned above, more general derivations than ours exist [33]. Hence, our main contribution is in automating also the closure, which we show and justify in detail (Sec. VII). Our closure scheme [Eqs. (23) or (24)] generalizes previous insights [2–4] and relies on the theory of Markov networks [34] to make it applicable to motifs of any type and size, such that it can close any set of moment equations at arbitrary order. We show that our closure scheme is exact if the motifs that are decomposed by the closure form a tree and if there is independence beyond their graph diameter.

As shown in Fig. 1, the whole procedure consists of four main steps: (i) enumeration of all required motifs up to size k, (ii) derivation of the unclosed ODEs for these motifs, (iii) elimination via conservation relations, and (iv) closure of the system of ODEs. We will refer to the final closed system of ODEs as the k-th order mean field, or MFk in short. Elimination is not strictly necessary but it increases efficiency, particularly for higher-order approximations. We developed a `Mathematica` [35] package that derives MFk by performing steps (i)–(iv). The required inputs for this algorithm are: the counts of all induced subgraphs in the underlying network up to size k, and, the matrices $\mathbf{R}^0$, $\mathbf{R}^1$ with conversion and interaction rates. A code example with output is discussed in Appendix G and the package is available for download [36].

In Secs. II and III, the underlying Markov chain for the network and its motifs are introduced. Section IV explains the general formula (6) for step (ii), which follows from the master equation (for its derivation, see Appendix A). The conservation relations used for elimination of variables from the ODEs in step (iii) are shown in Eqs. (8) and (9) of Sec. V. The form of the system of moment equations up to a truncation order k (10) and the variable elimination are shown in Sec. VI, resulting in the unclosed system (13). Finally, general expressions to close the system of moment equations in step (iv) are derived in Sec. VII [Eqs. (23) or (24)], resulting in form (14). In Sec. VIII we set up MF1-5 models of SIS epidemic spreading and compare their steady states to those of simulations on a selection of networks. We focus in particular on the square lattice, for which low-order moment closures fail, due to its large number of cycles of any size, and we compare against random networks and higher-dimensional lattices, for which they work well.

II. THE UNDERLYING DISCRETE-STATE CONTINUOUS-TIME MARKOV CHAIN

We consider a dynamical system on a fixed undirected graph $\mathcal{G}$ with N nodes and adjacency matrix $\mathsf{A} \in \{0,1\}^{N\times N}$, where each node may have one out of n discrete states. We may see the nodes as locations and the states as species, such that the space at node $i \in \{1,\ldots,N\}$ is occupied by exactly one species in $\{1,\ldots,n\}$. We denote the state vector at time t by $X(t)$ such that $[X_i(t)]_{i=1}^N \in \{1,\ldots,n\}^N$. The type of dynamics we consider is a continuous-time Markov chain with two Poisson process transition types, with rates specified by a $n\times n$ matrix $\mathbf{R}^0$ and a $n\times n\times n$ tensor $\mathbf{R}^1$:

(i) $\mathbf{R}^0$ specifies *spontaneous conversion* rates. Any node with state a may change spontaneously into state b, with rate R^0_{ab};

(ii) $\mathbf{R}^1$ specifies *nearest-neighbor-induced conversion* rates. Any node with state a may change into state b for each link to a node with state c, with rate R^1_{abc}.

This corresponds to the reaction rules

$$ⓐ \overset{R^0_{ab}}{\to} ⓑ, \qquad ⓐ\!-\!ⓑ \overset{R^1_{abc}}{\to} ⓐ\!-\!ⓒ. \tag{1}$$

A simple example is *susceptible-infected-susceptible* (SIS) *epidemic spreading* on a square two-dimensional lattice of

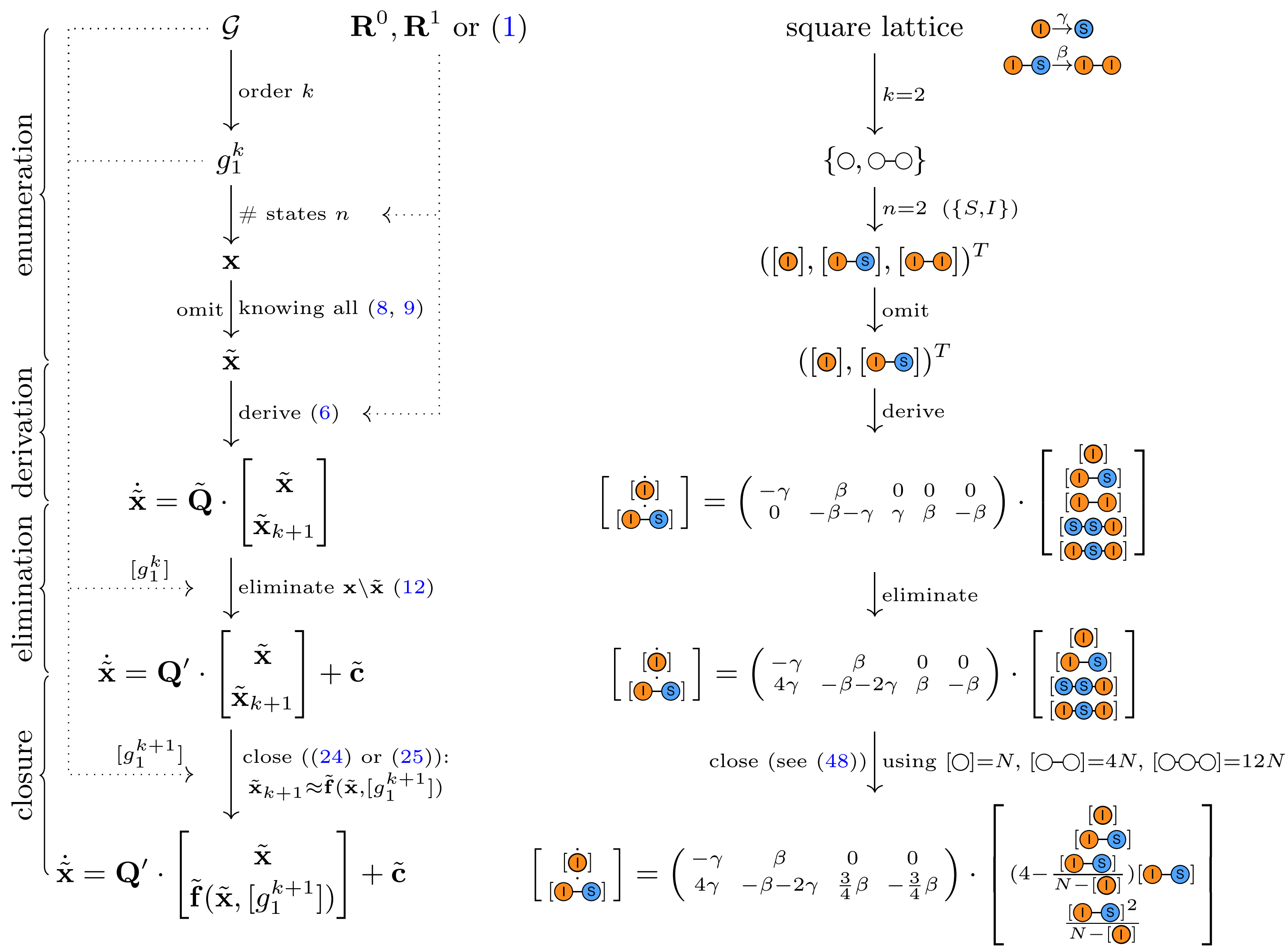


FIG. 1. Mean-field models for multistate dynamics on networks via moment closure. Left: general form. Right: example (MF2 for SIS spreading). Input (top): network $\mathcal{G}$, the considered process defined by the reaction rate matrices $\mathbf{R}^0, \mathbf{R}^1$. Output (bottom): closed system of differential equations representing mean frequencies of motif counts in the network up to a maximum size k. The four main steps are: (i) enumerate all required motifs, (ii) derive the moment equations, (iii) substitute to be eliminated terms via conservation relations, (iv) close the system via a closure scheme.

$\sqrt{N} \times \sqrt{N}$ nodes (say, periodic in both directions). For SIS spreading, each node may have one of two states, *susceptible* or infected, $\{\text{S}, \text{I}\} = \{1, 2\}$, for $t \in [0, \infty)$ and $i \in \{1, \ldots, N\}$. S nodes can become infected at rate β per infected neighbor and I nodes recover spontaneously at rate γ. Hence, for SIS spreading, the matrix $\mathbf{R}^0$ has a single nonzero entry $R^0_{2,1} = \gamma > 0$ (for spontaneous *recovery*) and the tensor $\mathbf{R}^1$ has a single nonzero entry $R^1_{1,2,2} = \beta > 0$ (for infection along IS links, denoted by the symbol I–S below), corresponding to the reaction rules

$$\text{I} \xrightarrow{\gamma} \text{S}, \qquad \text{I–S} \xrightarrow{\beta} \text{I–I}. \tag{2}$$

III. NETWORK MOTIFS AND THEIR COUNTS

We define network motifs as (typically small) graphs with given state labels. The order of a motif is the number of nodes it has. For example, in our example, I nodes and IS links are examples of first and second-order motifs. We use square brackets to denote the count of occurrences of motifs in $(\mathcal{G}, X)$, i.e., the number of occurrences of I nodes is [I]. Hence, we can write, e.g., respectively, for I nodes, IS links, and ISI chains

$$[\text{I}] = \sum_{i}^{N} \delta_2(X_i), [\text{I–S}] = \sum_{i,j\neq i}^{N,N} \delta_1(\mathsf{A}_{ij})\delta_2(X_i)\delta_1(X_j),$$

$$[\text{ISI}] = \sum_{i,j\neq i,k}^{N,N,N} \delta_1(\mathsf{A}_{ij})\delta_1(\mathsf{A}_{jk})\delta_0(\mathsf{A}_{ik})\delta_2(X_i)\delta_1(X_j)\delta_2(X_k),$$

where we omitted the dependence on t. The Kronecker δ, $\delta_y(x)$, is 1 if x equals y and 0 otherwise. By construction, the motif counts on the left-hand side are random, since $X(t)$ is random.

Generalizing the above examples, a *network motif* of *order* m, is a network with m nodes, each of which are labeled with a state. It his hence fully characterized by its connectivity pattern between nodes and its state labels on the nodes. The connectivity between motif nodes, i.e., the motif without labels, will be indicated by a, which, depending on the context, denotes the adjacency matrix of the motif, or a set $\mathsf{a} \in (\{1, \dots, m\} \times \{1, \dots, m\})^{\mu}$ of links between the m nodes [the indices of the nonzero entries in the adjacency matrix of the motif such that $\mu \leqslant m(m-1)/2$], or a graphical representation of the connectivity. For instance, two linked nodes are displayed according to these representations as $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\{(1, 2)\}$, or ○-○, respectively. As we focus only on undirected networks, each pair in the pair representation is bidirectional, i.e., we write $\{(1, 2)\}$ instead of $\{(1, 2), (2, 1)\}$. Motif labels will be indicated via a vector $\boldsymbol{x} = (x_1, \dots, x_m) \in \{1, \dots, n\}^m$ with state labels x_p at positions $p = 1, \dots, m$. Hence, the pair of $\boldsymbol{x}$ and a describes the motif, which we write as $\boldsymbol{x}^{\mathsf{a}}$. For example $(2, 1)^{(1,2)} =$ Ⓘ, Ⓢ$)^{(1,2)}$ is the Ⓘ-Ⓢ link.

For a general network motif $\boldsymbol{x}^{\mathsf{a}}$, the total count $[\boldsymbol{x}^{\mathsf{a}}] = [\boldsymbol{x}^{\mathsf{a}}](t)$ in the network $\mathcal{G}$ with fixed adjacency A and node labels $X = X(t)$ is

$$[\boldsymbol{x}^{\mathsf{a}}] = \sum_{\boldsymbol{i} \in S(m,N)} \delta_{\mathsf{a}}(\mathsf{A}_{\boldsymbol{i}}) \delta_{\boldsymbol{x}}(X_{\boldsymbol{i}}), \tag{3}$$

where $S(m, N)$ is the set of all m-tuples from $\{1, \dots, N\}$ without repetition, which has size $|S(m, N)| = N!/(N-m)!$. For instance, $\sum_{\boldsymbol{i} \in S(3,N)} = \sum_{i \in \{1,\dots,N\}} \sum_{j \in \{1,\dots,N\} \setminus i} \sum_{k \in \{1,\dots,N\} \setminus \{i,j\}}$. For an index set $\boldsymbol{i} \subset \{1, \dots, N\}^m$ of length m we use the convention that $\mathsf{A}_{\boldsymbol{i}} \in \{0, 1\}^{m \times m}$ and $X_{\boldsymbol{i}} \in \{1, \dots, n\}^m$ are the restrictions of matrix A and vector X to the index set $\boldsymbol{i}$. We count exact matches between the motif and the subgraph in $(\mathcal{G}, X)$. This means that the counted motif needs to have a as induced subgraph of $\mathcal{G}$, with matching state labels. Counting via Eq. (3) leads to multiple counting of motifs with symmetries (more precisely: automorphisms—see Appendix A), with multiplicity equal to the number of symmetries. For instance, Ⓘ-Ⓢ is counted once, but Ⓢ-Ⓢ or Ⓘ-Ⓘ is counted twice, and the triangle Ⓢ-Ⓢ-Ⓢ is counted six times.

IV. DIFFERENTIAL EQUATIONS FOR MOTIF COUNTS

In our SIS spreading example, the expected rate of change for the count of infected nodes is well known to satisfy

$$\frac{d}{dt}\langle[\text{Ⓘ}]\rangle = -\gamma\langle[\text{Ⓘ}]\rangle + \beta\langle[\text{Ⓘ-Ⓢ}]\rangle, \tag{4}$$

where $\langle\cdot\rangle$ brackets denote expectations over many independent realizations of the underlying Markov chain. Relation (4) is exact for finite network sizes N and can be derived from the Kolmogorov-forward (or master) equation for the Markov chain, see Appendix A. Note that the structure in Eq. (4) is such that the expected rate for the frequency of a motif of size $m = 1$ depends on the counts of motifs of size $m = 1$ (here $\langle[\text{Ⓘ}]\rangle$) from spontaneous conversions (here recovery) and size $m = 2$ (here Ⓘ-Ⓢ) from nearest-neighbor-induced conversions (here infection).

This is true in general such that the count of a general motif $[\boldsymbol{x}^{\mathsf{a}}](t)$ of size m satisfies an ordinary differential equation of form

$$\frac{d}{dt}\langle[\boldsymbol{x}^{\mathsf{a}}]\rangle = F_{\boldsymbol{x}^{\mathsf{a}}}\left(\left\langle\left[\boldsymbol{y}_1^{\mathsf{b}_1}\right]\right\rangle, \left\langle\left[\boldsymbol{y}_2^{\mathsf{b}_2}\right]\right\rangle, \dots\right), \tag{5}$$

where $|\boldsymbol{y}_i| \in \{m, m+1\}$. On the right-hand side the $[\boldsymbol{y}_i^{\mathsf{b}_i}](t)$ stand for counts of motifs of size m or $m+1$ on which the dynamics of $\langle[\boldsymbol{x}^{\mathsf{a}}]\rangle$ depend. Our package expresses the right-hand side of the differential equation (5) for arbitrary motifs $\boldsymbol{x}^{\mathsf{a}}$ of size m in the general form

$$\begin{aligned}\frac{d}{dt}\langle[\boldsymbol{x}^{\mathsf{a}}]\rangle = \sum_{p=1}^{m}\sum_{k=1}^{n}\sum_{c=1}^{n}\Bigg\{&\left(\frac{R^0_{kx_p}}{n} + \kappa^{\boldsymbol{x}^{\mathsf{a}}}_{p,c} R^1_{kx_pc}\right)\langle[\boldsymbol{x}^{\mathsf{a}}_{p\to k}]\rangle - \left(\frac{R^0_{x_pk}}{n} + \kappa^{\boldsymbol{x}^{\mathsf{a}}}_{p,c} R^1_{x_pkc}\right)\langle[\boldsymbol{x}^{\mathsf{a}}]\rangle \\ &+ \sum_{\boldsymbol{y}^{\mathsf{b}} \in \mathcal{N}^c_p(\boldsymbol{x}^{\mathsf{a}}_{p\to k})} R^1_{kx_pc}\langle[\boldsymbol{y}^{\mathsf{b}}]\rangle - \sum_{\boldsymbol{y}^{\mathsf{b}} \in \mathcal{N}^c_p(\boldsymbol{x}^{\mathsf{a}})} R^1_{x_pkc}\langle[\boldsymbol{y}^{\mathsf{b}}]\rangle\Bigg\}.\end{aligned} \tag{6}$$

Here, $\boldsymbol{x}_{p\to k}$ is the state label vector obtained by setting the state label of the pth element of $\boldsymbol{x}$ to k. $\kappa^{\boldsymbol{x}^{\mathsf{a}}}_{p,c}$ is the c degree at position p in motif $\boldsymbol{x}^{\mathsf{a}}$, i.e., it is the number of connections node p in motif $\boldsymbol{x}^{\mathsf{a}}$ has to nodes with state label c. The set $\mathcal{N}^c_p(\boldsymbol{x}^{\mathsf{a}})$ is defined as

$$\mathcal{N}^c_p(\boldsymbol{x}^{\mathsf{a}}) := \bigcup_{\ell=1}^{m+1}\left\{\boldsymbol{y}^{\mathsf{b}} : |\boldsymbol{y}| = m+1,\, y_\ell = c,\, \boldsymbol{y}^{\mathsf{b}}_{\ell\to\emptyset} = \boldsymbol{x}^{\mathsf{a}},\, (\ell, p) \in \mathsf{b}\right\},$$

and contains all motifs $\boldsymbol{y}^{\mathsf{b}}$ of order $m+1$ that extend the state label vector $\boldsymbol{x}$ by one new node with state label c and extend adjacency a by links to the new node, where $\boldsymbol{y}^{\mathsf{b}}_{\ell\to\emptyset}$ denotes the mth order connected motif obtained by deleting the ℓth node of $\boldsymbol{y}^{\mathsf{b}}$ and its links. In Fig. 5, we show the dependence of each subgraph type on subgraphs of size $m+1$ up to order 4. The differential equation (6) shows how the expected count of $\boldsymbol{x}^{\mathsf{a}}$ is increased by transitions into $\boldsymbol{x}^{\mathsf{a}}$—the positive terms—and decreased by transitions *out of* $\boldsymbol{x}^{\mathsf{a}}$—the negative terms. This happens through spontaneous conversions (terms with R^0), through nearest-neighbor interaction between nodes within the motif (first two terms with R^1), or through nearest-neighbor interaction with nodes outside the motif (last two terms with R^1). Note that equivalent motifs up to permutation (isomorphic motifs—see Appendix A) result in the same equation, such that we choose the same representative node indexing for each equivalence class.

V. CONSERVATION RELATIONS

Conservation relations are linear algebraic relations between motif counts. For example, for SIS spreading on a square lattice with periodic boundary conditions, we have for first and second-order motifs:

$$[\mathrm{S}](t) + [\mathrm{I}](t) = N, \quad [\mathrm{S}\text{-}\mathrm{S}](t) + [\mathrm{I}\text{-}\mathrm{S}](t) + [\mathrm{I}\text{-}\mathrm{I}](t) = 4N.$$

Such conservation of node and link counts occurs because our graph $\mathcal{G}$ is fixed. We can use the total counts on the right-hand side as normalizing factors such that we may write

$$[\![\mathrm{S}]\!](t) + [\![\mathrm{I}]\!](t) = 1, \quad [\![\mathrm{S}\text{-}\mathrm{S}]\!](t) + [\![\mathrm{I}\text{-}\mathrm{S}]\!](t) + [\![\mathrm{I}\text{-}\mathrm{I}]\!](t) = 1.$$

where $[\![\cdot]\!]$ is our notation for normalized motif counts. For networks with homogeneous degree, one can also write conservation equations of the type $[\![\mathrm{S}]\!] = [\![\mathrm{I}\text{-}\mathrm{S}]\!] + [\![\mathrm{S}\text{-}\mathrm{S}]\!]$.

In general, for each adjacency matrix a of possible motifs of size m the conservation relation

$$\sum_{\boldsymbol{x}:|\boldsymbol{x}|=m} [\boldsymbol{x}^{\mathsf{a}}](t) = [\mathsf{a}], \text{ where } [\mathsf{a}] = \sum_{\boldsymbol{i}\in S(m,N)} \delta_{\mathsf{a}}(\mathsf{A}_{\boldsymbol{i}}), \tag{7}$$

holds. The overall count $[\mathsf{a}]$ of induced subgraphs a in graph $\mathcal{G}$ is constant in time, and split up between all possible labelings. We can therefore use $[\mathsf{a}]$ as a normalization factor for the (variable) counts of motifs such that we may consider normalized motif counts

$$[\![\boldsymbol{x}^{\mathsf{a}}]\!] := [\boldsymbol{x}^{\mathsf{a}}]/[\mathsf{a}], \text{ such that } \sum_{\boldsymbol{x}:|\boldsymbol{x}|=m} [\![\boldsymbol{x}^{\mathsf{a}}]\!](t) = 1, \tag{8}$$

is the conservation relation for the normalized quantities. For networks with homogeneous degree there is the additional type of conservation relation

$$\sum_{k=1}^{n} [\![\boldsymbol{x}^{\mathsf{a}}_{p\to k}]\!] = [\![\boldsymbol{x}^{\mathsf{a}}_{p\to\emptyset}]\!], \tag{9}$$

for each stub p of the motif $\boldsymbol{x}^{\mathsf{a}}$ (a stub is a node with degree 1 in a). The conservation relations (8) and (9) can be used to reduce the number of variables in the moment equations via substitution. This can result in a substantial reduction in the number of moment equations (see Table II).

VI. TRUNCATION AND SUBSTITUTION

When we use Eq. (6) to express expected rates of change for the set of motifs up to a chosen maximum size k, we obtain a truncated hierarchy of moment equations. This linear system of differential equations has the form

$$\begin{bmatrix}\dot{\mathbf{x}}_1\\ \dot{\mathbf{x}}_2\\ \vdots\\ \dot{\mathbf{x}}_k\end{bmatrix} = \begin{bmatrix}\mathbf{Q}_1 & \mathbf{Q}_{12} & \mathbf{0} & \cdots & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_2 & \mathbf{Q}_{23} & \mathbf{0} & \cdots & \vdots\\ \vdots & \mathbf{0} & \ddots & \ddots & \mathbf{0} & \vdots\\ \vdots & \cdots & \mathbf{0} & \mathbf{Q}_{k-1} & \mathbf{Q}_{k-1,k} & \mathbf{0}\\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{Q}_k & \mathbf{Q}_{k,k+1}\end{bmatrix} \cdot \begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\\ \vdots\\ \mathbf{x}_k\\ \mathbf{x}_{k+1}\end{bmatrix} =: \begin{array}{c} \\ \dot{\mathbf{x}}_1\\ \dot{\mathbf{x}}_2\\ \vdots\\ \dot{\mathbf{x}}_{k-1}\\ \dot{\mathbf{x}}_k\end{array} \begin{array}{c} \begin{array}{cccccc}\mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{k-1} & \mathbf{x}_k & \mathbf{x}_{k+1}\end{array}\\ \begin{bmatrix}\mathbf{Q}_1 & \mathbf{Q}_{12} & \mathbf{0} & \cdots & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_2 & \mathbf{Q}_{23} & \mathbf{0} & \cdots & \vdots\\ \vdots & \mathbf{0} & \ddots & \ddots & \mathbf{0} & \vdots\\ \vdots & \cdots & \mathbf{0} & \mathbf{Q}_{k-1} & \mathbf{Q}_{k-1,k} & \mathbf{0}\\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{Q}_k & \mathbf{Q}_{k,k+1}\end{bmatrix}\end{array}, \tag{10}$$

where we defined a more compact notation on the right. In Eq. (10), $\mathbf{x}_m$ are vectors with all dynamically relevant motif counts of size m, $\mathbf{Q}_m$ the coefficients for $\dot{\mathbf{x}}_m := d\mathbf{x}_m/dt$ with motifs of the same size, and $\mathbf{Q}_{m,m+1}$ the coefficients for $\dot{\mathbf{x}}_m$ with motifs of the size $m+1$. The block diagonal form arises because the change of size-m motif counts depends only on motif counts of size m and of size $m+1$. In Appendix B, Eq. (10) is shown for SIS spreading up to a maximum motif size of $k = 3$.

The substitution via conservation relations can be written as

$$\mathbf{x} = \mathbf{E} \cdot \tilde{\mathbf{x}} + \mathbf{c}, \tag{11}$$

where $\mathbf{x} = \mathbf{x}_1, \ldots, \mathbf{x}_k$, $\tilde{\mathbf{x}}$ is $\mathbf{x}$ with the to be substituted elements omitted. $\mathbf{E}$, $\mathbf{c}$ contain the coefficients of linear dependence from Eqs. (7) and (9), with their ith row/element corresponding to the identity transformation for motifs that are not substituted. Substituting this into Eq. (10) results in the system of equations for the remaining motifs $\tilde{\mathbf{x}}$:

$$\dot{\tilde{\mathbf{x}}} = \tilde{\mathbf{Q}}_{1\cdots k,1\cdots k} \cdot (\mathbf{E} \cdot \tilde{\mathbf{x}} + \mathbf{c}) + \tilde{\mathbf{Q}}_{1\cdots k,k+1} \cdot \tilde{\mathbf{x}}_{k+1} \tag{12}$$

(where the tilde omits the to be substituted rows/elements), such that we can write the system in the same form as Eq. (10), but now with an added constant vector:

$$\dot{\tilde{\mathbf{x}}} = \mathbf{Q}' \cdot \begin{bmatrix}\tilde{\mathbf{x}}\\ \tilde{\mathbf{x}}_{k+1}\end{bmatrix} + \tilde{\mathbf{c}}. \tag{13}$$

For an example, see Sec. VIII A 2.

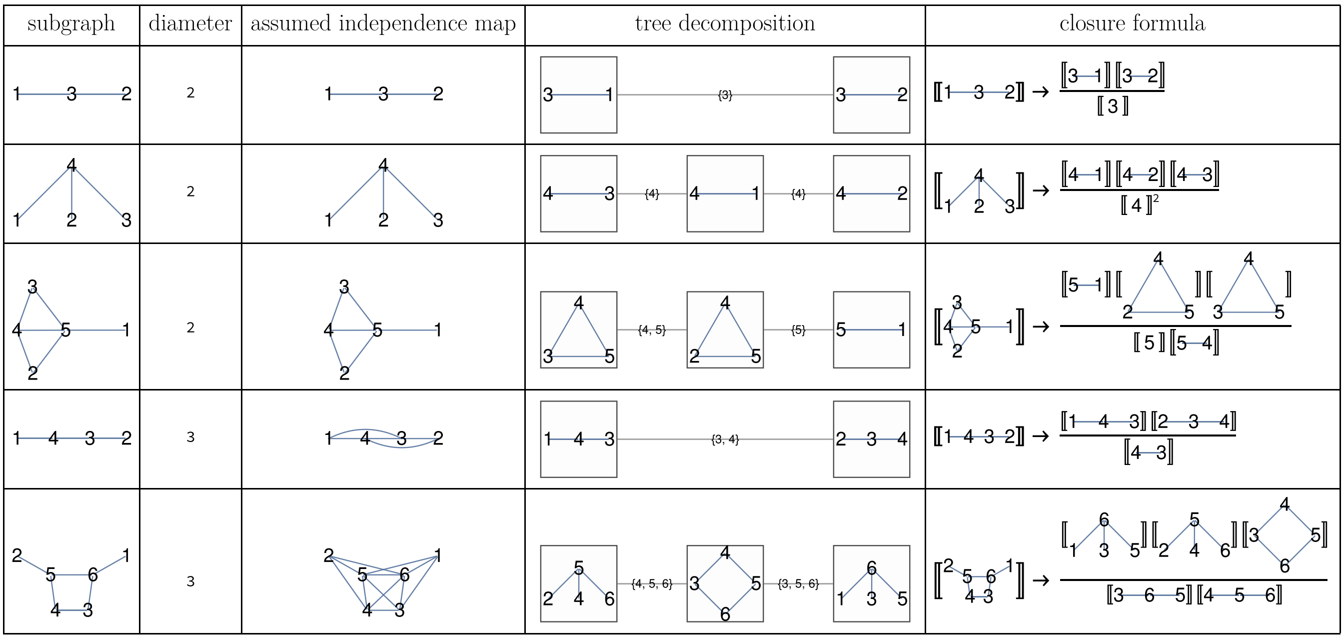
VII. CLOSURE SCHEME

Because in Eq. (13) counts for the largest motifs $\mathbf{x}_{k+1}$ appear on the right-hand side but not on the left-hand side, Eq. (13) is underdetermined. A closure scheme provides a way of expressing the undetermined parts $\tilde{\mathbf{x}}_{k+1}$ in Eq. (10) through a nonlinear function $\tilde{\mathbf{x}}_{k+1} \approx \tilde{\mathbf{f}}(\tilde{\mathbf{x}})$, creating a closed system of ODEs:

$$\dot{\tilde{\mathbf{x}}} = \mathbf{Q}' \cdot \begin{bmatrix}\tilde{\mathbf{x}}\\ \tilde{\mathbf{f}}(\tilde{\mathbf{x}})\end{bmatrix} + \tilde{\mathbf{c}}, \tag{14}$$

where $\tilde{\mathbf{f}}$ also depends on counts of induced subgraphs of order up to $k+1$ ($[g_1^{k+1}]$ in Fig. 1) if the motifs were not normalized in advance. In this section, we develop a closure scheme that decomposes $\tilde{\mathbf{x}}_{k+1}$ into its smaller-sized components. Our final formula generalizes closures hitherto most commonly used, as shown in, e.g., House *et al.* [21]. We will show that the decomposition is valid when: (i) counts of components are

TABLE I. Examples of our method to obtain closure formulas. Because the closures can be written independent of the state labels, the decomposition is shown for node-indexed graphs without reference to particular node states.



conditionally independent given the node states in their intersection and the adjacency structure between them is a tree, (ii) the network is spatially homogeneous, and (iii) the network is sufficiently large, such that the law of large numbers applies. We start with some introducing examples in Sec. VII A and defer detailed explanation to Secs. VII B and VII C. Examples are given in Secs. VII D and VII H.

A. Introduction

When truncating the moment hierarchy at a chosen order k, we approximate the order $k+1$ motifs appearing in the equations for order k motifs in terms of lower-order motifs. For example, looking at the moment equations for SIS spreading in Appendix B, when truncating at k=1, we would need an expression of [I-S] in terms of [I] and [S]. When assuming statistical independence of neighboring node states, the resulting expression is of the form [I-S] $\propto$ [I][S] (ignoring proportionality constants for now). Similarly, for truncation at $k=2$, we would need an expression for all the 3-chains and triangles on the right-hand side of Eqs. (B3) and (B4). For instance, the 3-chain [SSI] is typically decomposed as [SSI] $\propto$ [S-S][I-S]/[S]. Using the shorthand x_i for the event $X_i = x$, this is justified when there is a conditional independence relation of the form $P(S_i, S_j, I_k) = P(S_i, S_j)P(S_j, I_k|S_j)$ and if the component probabilities are the same everywhere in the network, such that node indices i, j, k do not matter. This can be generalized to larger chains, such as, e.g., [SSII] $\propto$ [SSI][SII]/[I-S], which similarly follows from the assumed conditional independence relation $P(S_i, S_j, I_k, I_l) = P(S_i, S_j, I_k)P(S_j, I_k, I_l|S_j, I_k)$ and homogeneity in the network.

It is possible to generalize the examples above to larger subgraphs by starting from the chain rule of probability,

$$P(\boldsymbol{x}) = P\big(\boldsymbol{x}_{\boldsymbol{i}_n} \mid \boldsymbol{x}_{\boldsymbol{i}_{n-1}}, \ldots, \boldsymbol{x}_{\boldsymbol{i}_2}, \boldsymbol{x}_{\boldsymbol{i}_1}\big)\ldots P\big(\boldsymbol{x}_{\boldsymbol{i}_2} \mid \boldsymbol{x}_{\boldsymbol{i}_1}\big)P\big(\boldsymbol{x}_{\boldsymbol{i}_1}\big),$$

where $\boldsymbol{x}$ is a vector of state labels on a given network motif, and subsequently simplify with assumed conditional independence relations. For instance, for our second example above, [SSI], we have $P(S_i, S_j, I_k) = P(I_k|S_i, S_j)P(S_j|S_i)P(S_i)$. If now node k is conditionally independent of node i, then we substitute $P(I_k|S_i, S_j) = P(I_k|S_j)$, such that we obtain the expression found above. In general, a simplification of the chain rule in terms of subgraphs is possible if we can order the chosen sets of subgraphs (with node indices $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_n$) without creating cycles and if the states of adjacent subgraphs are conditionally independent given their shared nodes. To make this precise, we need the concept of independence map. The independence map is a graph in which link absence between two nodes means that there is no direct dependence between them. For instance, if in the four-node graph shown in row 4 column 1 of Table I there is statistical independence between node states that are further than two steps removed, then its independence map is the graph shown in row 4 column 3 of Table I. The condition mentioned above for simplification of the chain rule in terms of subgraphs reduces to the requirement that the independence map be chordal, because this allows a tree composition in terms of maximal cliques, as shown in column 4 of Table I. A graph is chordal if every cycle greater than three is cut short by a link between two nonconsecutive nodes of the cycle.

In reality, we do not know the dependence structure between the graph nodes. However, the practical requirement of truncation of the moment hierarchy obliges us to assume an independence map for each of the largest motifs. To obtain a consistent decomposition method, we will choose as independence map the graph obtained by connecting all nodes to other nodes in their $d-1$ neighborhood, where d is the diameter of the motif. The chain rule then leads to a decomposition in terms of (maximal) cliques of the independence map in the numerator and the node sets that separate them in the denominator. Table I shows example graphs, with their diameter, their assumed independence maps, their tree decomposition and the resulting closure formula. The choice of dependence within a distance d may in some cases result in nonchordal independence maps, such that the decomposition cannot be made. In the next sections, we will explain our method in detail and further show how one can treat motifs with a nonchordal independence map.

B. Definitions and background

We will rely on the theory of decomposable Markov networks, following mostly the terminology of Pearl [34, Ch. 3]. We generalize the decomposition from a factorization involving (1-)cliques to one involving d-cliques. he definitions in this section apply to a general graph $\mathcal{G}$ with nodes states X, but we will apply the decomposition to motifs in Sec. VII C.

a. Separation. Given a graph $\mathcal{G}(\mathcal{V},\mathcal{E})$ and three disjoint subsets of nodes $\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}\subset\mathcal{V}$, $\boldsymbol{k}$ *separates* $\boldsymbol{i}$ and $\boldsymbol{j}$ in $\mathcal{G}$, written as $\boldsymbol{i}\perp\!\!\!\perp_{\mathcal{G}}\boldsymbol{j}\mid\boldsymbol{k}$, if every path between $\boldsymbol{i}$ and $\boldsymbol{j}$ has at least one vertex in $\boldsymbol{k}$. Here, $\boldsymbol{k}$ is called a separator, or also, a node cut set of $\boldsymbol{i}$ and $\boldsymbol{j}$ in $\mathcal{G}$.

b. Independence map. An independence map $\mathcal{M}$ is a graph that represents the independence between components of a set of random variables X such that separation in $\mathcal{M}$ guarantees conditional independence between corresponding subsets of X. More precisely, given three disjoint subsets of nodes $\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}\subset\mathcal{V}$, X possesses a spatial Markov property:

$$\boldsymbol{i}\perp\!\!\!\perp_{\mathcal{M}}\boldsymbol{j}\mid\boldsymbol{k}\Rightarrow X_{\boldsymbol{i}}\perp\!\!\!\perp X_{\boldsymbol{j}}\mid X_{\boldsymbol{k}},\tag{15}$$

where the $\perp\!\!\!\perp$ notation on the right refers to independence of the random variables: $P(X_{\boldsymbol{i}}{=}x_{\boldsymbol{i}},X_{\boldsymbol{j}}{=}x_{\boldsymbol{j}}|X_{\boldsymbol{k}}{=}x_{\boldsymbol{k}})=P(X_{\boldsymbol{i}}{=}x_{\boldsymbol{i}}|X_{\boldsymbol{k}}{=}x_{\boldsymbol{k}})P(X_{\boldsymbol{j}}{=}x_{\boldsymbol{j}}|X_{\boldsymbol{k}}{=}x_{\boldsymbol{k}})$. The pair $(X,\mathcal{M})$ defines what is known as a *Markov network*.

c. Independence beyond distance d. Let $\mathcal{G}=(\mathcal{V},\mathcal{E})$ be a graph where the nodes in $\mathcal{V}$ have (random) states X. We define $\mathcal{G}^d$ as the graph in which all nodes of $\mathcal{G}$ are neighbors if they are at most a shortest distance d away from each other, i.e.,

$$\mathcal{G}^d{=}(\mathcal{V},\mathcal{E}^d),\quad\text{with}\quad\mathcal{E}^d{=}\{(i,j){\in}\mathcal{V}:\mathrm{dist}_{\mathcal{G}}(i,j){\leqslant}d\}.\tag{16}$$

We then say that $(\mathcal{G},X)$ has *independence beyond distance d* if $\mathcal{G}^d$ is the independence map of X, or for all distinct subsets of nodes $\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}\subset\mathcal{V}$ holds

$$\boldsymbol{i}\perp\!\!\!\perp_{\mathcal{G}^d}\boldsymbol{j}\mid\boldsymbol{k}\Rightarrow X_{\boldsymbol{i}}\perp\!\!\!\perp X_{\boldsymbol{j}}\mid X_{\boldsymbol{k}}.\tag{17}$$

This means that states of two nonneighboring sets in $\mathcal{G}^d$, which by definition (16) are further than d steps apart in $\mathcal{G}$, are independent of each other given the state of their separator $\boldsymbol{k}$.

d. Maximal d-cliques and d-clique graph. A maximal clique is a complete subgraph not contained in a larger complete subgraph [37]. As a generalization, *maximal d-cliques* are maximal subgraphs with distance between any two nodes not greater than d [35]. Correspondingly, maximal cliques in $\mathcal{G}^d$ are maximal d-cliques in $\mathcal{G}$. The graph $\mathcal{C}^d$ is the *d-clique graph* of $\mathcal{G}$ if each node in $\mathcal{C}^d$ corresponds to a d-clique in $\mathcal{G}$ with links between nodes in $\mathcal{C}^d$ occurring when the corresponding d-cliques overlap. Hence, while nodes in $\mathcal{C}^d$ correspond to maximal d-cliques in $\mathcal{G}$, links in $\mathcal{C}^d$ correspond to intersections between overlapping maximal d-cliques in $\mathcal{G}$.

e. Junction graph of d-cliques. A junction graph of $\mathcal{G}$'s maximal d-cliques, denoted further as $\mathcal{J}^d(\mathcal{G})$, is a subgraph of the d-clique graph $\mathcal{C}^d$ obtained by removing redundant links from $\mathcal{C}^d$. Denoting d-cliques corresponding to nodes i, j in $\mathcal{C}^d$ as $c_i,c_j\subset\mathcal{V}$, a link between i and j in $\mathcal{C}^d$ is *redundant* when there is an alternative path between i and j in $\mathcal{C}^d$ passing by a series of other nodes in $\mathcal{C}^d$ of which the corresponding d-cliques all contain $c_i\cap c_j$. The junction graph $\mathcal{J}^d(\mathcal{G})$ is then obtained by iteratively removing redundant links from the $\mathcal{C}^d$ until there are no further redundant links. While the d-clique graph is unique, there may be several junction graphs $\mathcal{J}^d(\mathcal{G})$ of d-cliques for one graph $\mathcal{G}$. Note that for chordal graphs (defined below), the junction graph equals what is known as a junction tree, which can also be obtained via the junction tree algorithm [38], applied to the d-clique graph.

f. d-chordality. A graph $\mathcal{G}$ is chordal when for every cycle of length greater than 3, there exists a link in $\mathcal{G}$ between two nonconsecutive nodes of the cycle (thus, giving a short-cut, also called *chord* to the cycle). As a generalization, we will call a graph $\mathcal{G}$ d-chordal if $\mathcal{G}^d$ is chordal. If a graph $\mathcal{G}$ is d-chordal, then $\mathcal{J}^d(\mathcal{G})$ is a tree, or equivalently, if $\mathcal{G}^d$ is chordal, then $\mathcal{J}(\mathcal{G}^d)$ is a tree. Nonchordal graphs can always be converted to a chordal graph via *triangulation*, i.e., adding chords to every chordless cycle of length greater than 3. We will write below $\mathrm{tr}(\mathcal{G}^d)$ as a *minimal triangulation* of a nonchordal $\mathcal{G}^d$, obtained by adding the smallest number of links that leads to chordality, unless stated otherwise.

g. Decomposability at distance d. If there is independence beyond distance d (17) and $\mathcal{G}$ is d-chordal, then the joint probability of the network nodes of $\mathcal{G}$ being in a given state $P(X=\boldsymbol{x})$ can be factorized over the d-cliques of $\mathcal{G}$. We will call this property of the graph $\mathcal{G}$ and its node states X *decomposability at distance d*. We call $\mathcal{J}$ the set of d-cliques in $\mathcal{G}$, ordering its elements consistent with the resulting junction tree structure for $\mathcal{G}^d$ (such that $\mathcal{J}_1$ is the chosen root and parent nodes have lower index than their leaves), and call $\mathrm{pa}(\mathcal{J}_i)$ the parent node of $\mathcal{J}_i$. In such cases, the factorization is possible because the tree structure between d-cliques allows application of the chain rule of conditional probability:

$$\begin{aligned}P(X{=}\boldsymbol{x})&=\prod_{i=1}^{|\mathcal{J}|}P\big(X_{\mathcal{J}_i}{=}\boldsymbol{x}_{\mathcal{J}_i}\mid X_{\mathrm{pa}(\mathcal{J}_i)}{=}\boldsymbol{x}_{\mathrm{pa}(\mathcal{J}_i)}\big),\\&=\prod_{i=1}^{|\mathcal{J}|}P\big(X_{\mathcal{J}_i}{=}\boldsymbol{x}_{\mathcal{J}_i}\mid X_{\mathrm{pa}(\mathcal{J}_i)\cap\mathcal{J}_i}{=}\boldsymbol{x}_{\mathrm{pa}(\mathcal{J}_i)\cap\mathcal{J}_i}\big),\\&=\frac{\prod_{i=1}^{|\mathcal{J}|}P\big(X_{\mathcal{J}_i}{=}\boldsymbol{x}_{\mathcal{J}_i}\big)}{\prod_{i=2}^{|\mathcal{J}|}P\big(X_{\mathrm{pa}(\mathcal{J}_i)\cap\mathcal{J}_i}{=}\boldsymbol{x}_{\mathrm{pa}(\mathcal{J}_i)\cap\mathcal{J}_i}\big)}.\end{aligned}\tag{18}$$

The steps in Eq. (18) are explained as follows. As d-chordality makes $\mathcal{J}^d(\mathcal{G})$ a tree and any d-clique separates its neighbors, one can recursively use conditional independence of children given parents (line 1). We use the convention that $\mathrm{pa}(\mathcal{J}_1)=\emptyset$, such that the first factor is $P(X_{\mathcal{J}_1}{=}\boldsymbol{x}_{\mathcal{J}_1})$. In line 2 we exploit that any two d-cliques in $\mathcal{G}$ are also separated by their intersection to condition instead on intersections.

h. Non-d-chordal graphs. There are two alternative ways to decompose non-d-chordal $\mathcal{G}$: (i) perform the decomposition (18) on the (more conservative) independence map after triangulation, $\mathrm{tr}(\mathcal{G}^d)$. In this case, the factors in Eq. (18) may still contain subgraphs of diameter $d+1$, but of smaller size than $\mathcal{G}$, such that one may have to apply Eq. (18) recursively to achieve smaller diameter for all factors. Furthermore, the resulting decomposition will depend on the choice of triangulation. Alternatively, (ii) one can start from the nontree $\mathcal{J}^d(\mathcal{G})$ and use the ad hoc formula (without prior triangulation)

$$P(X{=}\boldsymbol{x}) \approx \zeta \frac{\prod_i^{|\mathcal{J}|} P(X_{\mathcal{J}_i}{=}\boldsymbol{x}_{\mathcal{J}_i})}{\prod_{i,j\neq i}^{|\mathcal{J}|} P(X_{\mathcal{J}_i\cap\mathcal{J}_j}{=}\boldsymbol{x}_{\mathcal{J}_i\cap\mathcal{J}_j})}. \tag{19}$$

Because the fraction in Eq. (19) does not result from application of the chain rule as in Eq. (18), it is not a product of conditional probabilities and hence it does not guarantee the property that each of the node states in $\boldsymbol{x}$ has to appear one more time in the numerator than in the denominator, which in turn leads to inconsistency between closure formulas that assume different d. The factor ζ in Eq. (19) corrects for this inconsistency—see Sec. VII C for more detail. After applying (i) to nonchordal graphs, the nodes in the resulting d-clique tree are not all maximal d-cliques of $\mathcal{G}$ any more [39]. When applying (ii) to nonchordal graphs, the subgraphs in $\mathcal{J}^d(\mathcal{G})$ are still maximal d-cliques of $\mathcal{G}$ but the clique graph is not a tree, thus, violating the assumptions behind the decomposition (18).

i. Maximum motif diameter. The decomposition explained in this section implies that if independence beyond distance d is valid in the whole network $(\mathcal{G}, X)$, then we know that we only need to consider motifs up to diameter d, justifying truncation of the moment hierarchy. Note, however, that truncation is usually done at a give size, not at a given diameter. We expand on this in Sec. VII C e.

C. Motif decomposition

a. Decomposition of motifs at the individual level. We apply the decomposition (18) to motifs with connectivity a and chordal independence map a^d embedded in the network. For now, we ignore that this may in some cases break the independence assumption, but see Sec. VII C e for more detail on this issue. Considering a set $\boldsymbol{i}$ of nodes that have connectivity a in our network $\mathcal{G}$ and taking $P(X_{\boldsymbol{i}}{=}\boldsymbol{x})$ as the probability that these are in states with labels $\boldsymbol{x}$, we can write the decomposition (18) for of $P(X_{\boldsymbol{i}}{=}\boldsymbol{x}) = \langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle$ to obtain

$$\langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle = \frac{\prod_{j=1}^{|\mathcal{J}|}\langle[\boldsymbol{x}_{\boldsymbol{i}_{\mathcal{J}_j}}]\rangle}{\prod_{j=2}^{|\mathcal{J}|}\langle[\boldsymbol{x}_{\boldsymbol{i}_{\mathrm{pa}(\mathcal{J}_j)\cap\mathcal{J}_j}}]\rangle}, \tag{20}$$

where now $\mathcal{J}$ is the set of d-cliques in a. Choosing $d = \mathrm{diam}(\mathsf{a}) - 1$ ensures that the decomposition results in component motifs with diameter decreased by one compared to the decomposed motif. For motifs with nonchordal a^d one can, as noted above, either triangulate a^d first or use the *ad-hoc* approximation (19). We relied on the package `Chordal Graph` [40] for triangulation of nonchordal a^d. The ad hoc formula (19) applied to the motif at $\boldsymbol{i}$ is

$$\langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle \approx \frac{\prod_{j}^{|\mathcal{J}|}\langle[\boldsymbol{x}_{\boldsymbol{i}_{\mathcal{J}_j}}]\rangle}{\prod_{j,k\neq j}^{|\mathcal{J}|}\langle[\boldsymbol{x}_{\boldsymbol{i}_{\mathcal{J}_j\cap\mathcal{J}_k}}]\rangle}\prod_{j\in\boldsymbol{i}}\langle[x_j]\rangle^{\gamma_j}. \tag{21}$$

The consistency correction [written as ζ in Eq. (19)] here equals $\prod_{j\in\boldsymbol{i}}\langle[x_j]\rangle^{\gamma_j}$ and ensures that the ad-hoc extension of closures to motifs with nonchordal a^d does not result in inconsistency with MF1 [condition 1 of 18, Ch. 21] under independence between node states: when all motifs of order greater than one are replaced by products of order one motifs, i.e., $\langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle \to \prod_{j\in\boldsymbol{i}}\langle[x_j]\rangle$, the right-hand side of Eq. (21) should reduce to MF1. Therefore, for each j, γ_j is chosen such that this is fulfilled. These ad hoc steps usually result in violation of the conservation relations of Sec. V [18]. In the approximations used in Sec. VIII, the bias introduced due to this violation is small. For mitigation of this problem, see Refs. [18,41].

b. Decomposition of motifs at the population level. If we take the following *spatial homogeneity* assumption for all motifs $\boldsymbol{x}^{\mathsf{a}}$ of sizes m up to our maximal considered size k,

$$\forall \boldsymbol{i}, \boldsymbol{i}' \in \mathcal{I}(\mathsf{a}) : \langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle = \langle[\boldsymbol{x}_{\boldsymbol{i}'}^{\mathsf{a}}]\rangle = \frac{\langle[\boldsymbol{x}^{\mathsf{a}}]\rangle}{[\mathsf{a}]} = \langle[\![\boldsymbol{x}^{\mathsf{a}}]\!]\rangle, \tag{22}$$

where $\mathcal{I}(\mathsf{a}) := \{\boldsymbol{j} \in S(m, N) : \mathsf{A}_{\boldsymbol{j}} = \mathsf{a}\}$, then Eqs. (20) and (21) are independent of $\boldsymbol{i}$, such that we can write Eq. (20) as

$$\langle[\![\boldsymbol{x}^{\mathsf{a}}]\!]\rangle \approx \frac{\prod_{j=1}^{|\mathcal{J}|}\langle[\![\boldsymbol{x}_{\mathcal{J}_j}]\!]\rangle}{\prod_{j=2}^{|\mathcal{J}|}\langle[\![\boldsymbol{x}_{\mathrm{pa}(\mathcal{J}_j)\cap\mathcal{J}_j}]\!]\rangle}, \tag{23}$$

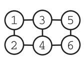
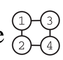
and Eq. (21) as

$$\langle[\![\boldsymbol{x}^{\mathsf{a}}]\!]\rangle \approx \frac{\prod_{j}^{|\mathcal{J}|}\langle[\![\boldsymbol{x}_{\mathcal{J}_j}]\!]\rangle}{\prod_{j,k}^{|\mathcal{J}|}\langle[\![\boldsymbol{x}_{\mathcal{J}_j\cap\mathcal{J}_k}]\!]\rangle}\prod_{p}^{m}\langle[\![x_p]\!]\rangle^{\gamma_p}, \tag{24}$$

which may be used to close the population-level equations (10). In Eqs. (23) and (24), node indexing of a motif is consistent with the node labels $\boldsymbol{x}$. Recall that we use a single consistent indexing for isomorphic motifs. The decomposition (23) and (24) is not unique. If the independence assumptions are satisfied, then each of the alternative ways to decompose motif $\boldsymbol{x}^{\mathsf{a}}$ should result in the same value. As we do not expect the independence to be perfectly valid, we take the average of the alternative ways of decomposing $\boldsymbol{x}^{\mathsf{a}}$ if they exist.

c. Normalization. We showed the closure formulas for normalized motifs, i.e., we first normalized the counts of motifs via Eq. (8), and then applied the closure. Hence, in this case, the counts of induced subgraphs in the network enter into the system of equations as normalization factors in the unclosed system. One can also decide not to normalize (or to do it after applying closure). In this latter case, the subgraph counts enter into the final system of equations when applying closure, as the closure formulas for the nonnormalized motif counts contain them (to see this, substitute each motif count in Eqs. (23) and (24) as $[\![\boldsymbol{y}^{\mathsf{b}}]\!] \to [\boldsymbol{y}^{\mathsf{b}}]/[\mathsf{b}]$). Hence, structural information specific to the considered network enters the mean-field equations either when normalizing the motif counts or when applying closure. In simple cases, such as lattices or random graphs, the subgraph counts can be found by hand without much effort. In other cases, one can resort to subgraph counting algorithms—we used `IGraph` [42] for `Mathematica` [35].

d. Law of large numbers. We will use the population-level closure to study the steady states in a single realization of a given network. Motif counts are then assumed to be the total counts in a single network, instead of their expectations over many realizations. As the closure formulas apply to expectations, we make the additional assumption that motif counts are close to their expectations.

e. Additional bias. Decomposing the whole network $\mathcal{G}$ at distance d is exact when $\mathcal{G}$ is d-chordal and there is independence beyond distance d (Sec. VII B). Applying the decomposition to motifs embedded in $\mathcal{G}$ instead of to $\mathcal{G}$ can be done without additional bias when a is a distance-hereditary subgraph of $\mathcal{G}$ (i.e., distances between nodes in a are equal to those between corresponding nodes in $\mathcal{G}$) and conditional independence relations implied by a are also valid in $\mathcal{G}$. As a counterexample for the former, take for $\mathcal{G}$ the six-node graph  and for a its induced subgraph consisting of nodes $\{1, 2, 4, 6, 5\}$. Here, a is not distance-hereditary because $\text{dist}_{\mathsf{a}}(1, 5)=4 \neq \text{dist}_{\mathcal{G}}(1, 5)=2$. As a counterexample for the latter, take for $\mathcal{G}$ the square  and for a the 3-node chain $\{1, 2, 4\}$. In this case, a is distance hereditary, but, while (when assuming independence beyond distance d=1) within a we have the independence relation $\{1\} \perp\!\!\!\perp_{\mathsf{a}} \{4\} \mid \{2\}$, this is not true in $\mathcal{G}$, where $\{1\} \not\perp\!\!\!\perp_{\mathcal{G}} \{4\} \mid \{2\}$, because the node cut set in $\mathcal{G}$ for $\{1\}$ and $\{4\}$ is $\{2, 3\}$. This occurs because the decomposed motifs are nonmaximal d-cliques. Therefore, we expect a bias as a consequence of this in the closed mean-field equation hierarchy [Eq. (10) with Eqs. (23) or (24) at population level under spatial homogeneity]. This bias can be avoided when expressing the equations in terms of maximal p-cliques for $p \in \{0, ..., k+1\}$ and truncating at given diameter instead of at given size.

D. Examples

Appendix H shows 13 application examples of Eqs. (23) and (24) in table form. As the closures can be written independent of the particular labels, they are shown for subgraphs only, with each node tagged with its index. We have also dropped the $\langle\cdot\rangle$, assuming that the law of large numbers applies, such that the counts approach their expectations almost surely for increasing network size N. The examples can be understood by reading the table from left to right. Below, we derive the normalization factors and the nonnormalized closures of examples 1–3 of Appendix H for different network types. Note that, unlike in Appendix H, we use letter labels below, for consistency with the main text and the literature.

(1) $x^a = \text{ⓐ-ⓑ-ⓒ}$: This diameter-2 motif has chordal independence map equal to a and decomposes with Eq. (23) as

$$[\![\text{ⓐ-ⓑ-ⓒ}]\!] \approx \frac{[\![\text{ⓐ-ⓑ}]\!][\![\text{ⓑ-ⓒ}]\!]}{[\![\text{ⓑ}]\!]}, \tag{25}$$

assuming conditional independence beyond distance $d = 1$. Via normalization (8) we obtain also the closure for the nonnormalized counts:

$$[\text{ⓐ-ⓑ-ⓒ}] \approx [\circ\!-\!\circ\!-\!\circ]\,\frac{[\circ]}{[\circ\!-\!\circ]^2}\,\frac{[\text{ⓐ-ⓑ}]\,[\text{ⓑ-ⓒ}]}{[\text{ⓑ}]}. \tag{26}$$

The counts of the induced subgraphs of size 2 and 1, required for normalization, are

$$[\circ\!-\!\circ] = \kappa N, \quad [\circ] = N, \tag{27}$$

and total number of triples (3-node motifs) in the network is

$$[\circ\!-\!\circ\!-\!\circ] + [\triangle] = \sum_i^N \kappa_i(\kappa_i - 1), \tag{28}$$

with κ_i the number of neighbors of node i and κ the mean number of neighbors over the whole network. For particular network types (28) can be simplified. Below are two examples.

(a) For a network with fixed degree without triangles (e.g., a square lattice), we have $\forall i : \kappa_i = \kappa$ and $[\triangle] = 0$, such that

$$[\circ\!-\!\circ\!-\!\circ] = \kappa(\kappa - 1)N. \tag{29}$$

Using Eqs. (26) and (29), we obtain

$$[\text{ⓐ-ⓑ-ⓒ}] \approx \frac{\kappa - 1}{\kappa}\,\frac{[\text{ⓐ-ⓑ}]\,[\text{ⓑ-ⓒ}]}{[\text{ⓑ}]}. \tag{30}$$

An early use of this closure for networks can be found in Keeling *et al.* [16].

(b) In a large Erdős-Rényi random network, we have $\kappa_i \sim \text{Pois}(\kappa)$ and $[\triangle]/([\triangle] + [\circ\!-\!\circ\!-\!\circ]) \approx 0$ [43]. Hence,

$$\begin{aligned}
[\circ\!-\!\circ\!-\!\circ] &= \sum_i^N \kappa_i^2 - \sum_i^N \kappa_i,\\
&= N\big[\mathrm{E}\big(\kappa_i^2\big) - \mathrm{E}(\kappa_i)\big],\\
&= N[\mathrm{Var}(\kappa_i) + \mathrm{E}(\kappa_i)^2 - \mathrm{E}(\kappa_i)],\\
&= \kappa^2 N,
\end{aligned} \tag{31}$$

where replacing the average by the expectation on the second line requires $N \to \infty$ (law of large numbers), on the third line we used $\mathrm{Var}(\cdot) := \mathrm{E}((\cdot)^2) - (\mathrm{E}(\cdot))^2$, and on the fourth line we used that, for $\kappa_i \sim \text{Pois}(\kappa)$, we have $\mathrm{E}(\kappa_i) = \mathrm{Var}(\kappa_i) = \kappa$. We could also have obtained this result directly from the large-N limit of chains [39]. Using Eqs. (26) and (31), we obtain

$$[\text{ⓐ-ⓑ-ⓒ}] \approx \frac{[\text{ⓐ-ⓑ}]\,[\text{ⓑ-ⓒ}]}{[\text{ⓑ}]}, \tag{32}$$

This closure was, to the best of our knowledge, first used for networks in Gross *et al.* [20].

(2) $\boldsymbol{x}^{\mathsf{a}} = \triangle_{\text{ⓐⓑⓒ}}$: This diameter-1 motif can be decomposed into its three 0-cliques as

$$[\![\triangle_{\text{ⓐⓑⓒ}}]\!] \approx [\![\text{ⓐ}]\!][\![\text{ⓑ}]\!][\![\text{ⓒ}]\!], \tag{33}$$

when assuming independence of nodes ($d = 0$). With nonnormalized counts, this becomes

$$[\![\text{ⓐ–ⓑ–ⓒ–ⓐ}]\!] \approx [\text{○–○–○–○}]\;[\text{ⓐ}]\;[\text{ⓑ}]\;[\text{ⓒ}]\;/\;[\text{○}]^3. \tag{34}$$

Alternatively, one can extend the usage of the ad hoc formula (24) to include nonmaximal cliques: using its three 1-cliques in Eq. (24), we obtain

$$[\![\text{ⓐ–ⓑ–ⓒ–ⓐ}]\!] \approx \frac{[\![\text{ⓐ–ⓑ}]\!][\![\text{ⓑ–ⓒ}]\!][\![\text{ⓒ–ⓐ}]\!]}{[\![\text{ⓐ}]\!][\![\text{ⓑ}]\!][\![\text{ⓒ}]\!]}, \tag{35}$$

which is known as the Kirkwood closure for triangles [22]. Using Eq. (8), we obtain for the closure with nonnormalized counts

$$[\text{ⓐ–ⓑ–ⓒ–ⓐ}] \approx [\text{○–○–○–○}]\,\frac{[\text{○}]^3}{[\text{○–○}]^3}\,\frac{[\text{ⓐ–ⓑ}]\;[\text{ⓑ–ⓒ}]\;[\text{ⓒ–ⓐ}]}{[\text{ⓐ}]\;[\text{ⓑ}]\;[\text{ⓒ}]}. \tag{36}$$

The frequency $[\text{○–○–○–○}]$ depends on the network type. For instance, if we use the definition of the clustering coefficient $\phi := [\text{○–○–○–○}]/([\text{○–○–○–○}] + [\text{○–○–○}])$ [30], then we have [via Eqs. (27) and (28)] for a network with fixed degree $[\text{○–○–○–○}] = \phi\kappa(\kappa - 1)N$, such that

$$[\text{ⓐ–ⓑ–ⓒ–ⓐ}] \approx \frac{\kappa - 1}{\kappa^2}\phi N\,\frac{[\text{ⓐ–ⓑ}]\;[\text{ⓑ–ⓒ}]\;[\text{ⓒ–ⓐ}]}{[\text{ⓐ}]\;[\text{ⓑ}]\;[\text{ⓒ}]},$$

which was first used for networks by Keeling [30].

(3) $x^a = \text{ⓐ–ⓑ–ⓓ},\ \text{ⓑ–ⓒ}$: This diameter-2 motif has chordal independence map a and decomposes with Eq. (23) as

$$[\![\text{ⓐ–ⓑ–ⓓ},\ \text{ⓑ–ⓒ}]\!] \approx \frac{[\![\text{ⓐ–ⓑ}]\!][\![\text{ⓑ–ⓒ}]\!][\![\text{ⓑ–ⓓ}]\!]}{[\![\text{ⓑ}]\!]^2}, \tag{37}$$

when assuming independence beyond distance $d = 1$. Alternatively, extending the ad-hoc formula (24) to the three nonmaximal 3-cliques, we obtain

$$[\![\text{ⓐ–ⓑ–ⓓ},\ \text{ⓑ–ⓒ}]\!] \approx \frac{[\![\text{ⓐ–ⓑ–ⓒ}]\!][\![\text{ⓐ–ⓑ–ⓓ}]\!][\![\text{ⓒ–ⓑ–ⓓ}]\!][\![\text{ⓑ}]\!]}{[\![\text{ⓐ–ⓑ}]\!][\![\text{ⓑ–ⓒ}]\!][\![\text{ⓑ–ⓓ}]\!]}, \tag{38}$$

where a consistency correction $[\![\text{ⓑ}]\!]$ was required. The nonnormalized form of this closure was first used in House *et al.* [21].

VIII. APPLICATION TO SIS EPIDEMIC SPREADING

We apply our method to SIS spreading, which is a continuous-time discrete-state Markov chain description of epidemic spreading through a population of susceptibles [e.g., 31]. As introduced in Sec. II, we have $n = 2$ species. The 2×2 matrix $\mathbf{R}^0$ of spontaneous conversion rates and the $2 \times 2 \times 2$ tensor $\mathbf{R}^1$ of conversion rates due to nearest-neighbor interaction for SIS spreading have only two positive entries, $R^0_{2,1} = \gamma$, $R^1_{1,2,2} = \beta$, corresponding to reaction scheme (2). Hence, contagion of susceptibles S occurs over IS links at rate β, whereas recovery occurs spontaneously at rate γ. In the study of phase transitions and interacting particle systems, SIS epidemic spreading is known as the contact process [44], which is typically studied on d-dimensional lattices. In this context, it was found to belong to the directed percolation universality class, of which scaling properties have been widely studied [12–14].

We run the simulations with a Gillespie algorithm [45] and stabilize them via feedback control [46–48], such that steady states can be obtained in a more efficient manner than when running regular simulations (see Appendix C).

A. Mean-field equations

We derived the mean-field models up to fifth order for the square lattice and up to second order for other networks, including cubic and hypercubic lattices, random regular networks and Erdős-Rényi random networks. In the main text, we only show a step-by-step derivation of the first and second-order mean-field models because they allow demonstration of our method in the simplest form. Recall that we write in the text $\langle[\cdot]\rangle$ as $[\cdot]$, assuming the LLN holds (in the `Mathematica` file for MF4 in Appendix G, we use the notation $\langle\cdot\rangle$ instead).

To gain insight in the strength of dependence between neighboring nodes in simulations and higher-order mean-field models, we will observe the correlation between neighboring node states a and b as in Keeling [30], defined by

$$C_{ab} = \frac{N}{\kappa}\frac{[\text{ⓐ–ⓑ}]}{[\text{ⓐ}]\;[\text{ⓑ}]}, \tag{39}$$

(where κ is the mean degree) or when motif counts are normalized,

$$C_{ab} = \frac{[\![\text{ⓐ–ⓑ}]\!]}{[\![\text{ⓐ}]\!][\![\text{ⓑ}]\!]}. \tag{40}$$

They are uncentered correlations between species types that are separated by one link. Values greater than 1 indicate clustering and values less than 1 avoidance (compared to a uniform random distribution). For a generalization of this correlation to arbitrary distances between end nodes, see Appendix D.

1. MF1

The first-order mean field originates from the molecular field approximation in statistical physics [8,9] and is now commonly known as the "mean-field model" [12–14,31,49,50]. MF1 only considers node states ($k = 1$) and neglects correlations beyond distance 0. It provides a picture of the dynamics when species are well mixed throughout a large domain. One way to achieve this is when the domain is a complete network on which susceptibles and infecteds have contact rate $\kappa\beta/N$, and when $N \to \infty$ [31].

There are two motif types of size 1: Ⓢ and Ⓘ. Of these, only Ⓘ is dynamically relevant. This means that we only need the equation

$$\frac{d}{dt}[\text{Ⓘ}] = \beta[\text{Ⓘ–Ⓢ}] - \gamma[\text{Ⓘ}].$$

To close the system at order 1, Ⓘ–Ⓢ needs to be expressed in terms of [Ⓘ] and [Ⓢ]. No correlation beyond distance 0 corresponds to [using Eq. (23)]

$$\frac{[\mathrm{I{-}S}]}{\kappa N} = \frac{[\mathrm{I}]}{N}\frac{[\mathrm{S}]}{N} \implies [\mathrm{I{-}S}] = \frac{\kappa}{N}[\mathrm{I}](N-[\mathrm{I}]), \tag{41}$$

where we have also used the conservation relation $[\mathrm{S}] + [\mathrm{I}] = N$ to substitute $[\mathrm{S}] = N - [\mathrm{I}]$. The final expression for the first-order mean field is

$$\frac{d}{dt}[\mathrm{I}] = \beta\frac{\kappa}{N}[\mathrm{I}](N-[\mathrm{I}]) - \gamma[\mathrm{I}],$$

which, after normalization [via Eq. (8)], yields

$$\frac{d}{dt}[\![\mathrm{I}]\!] = \beta\kappa[\![\mathrm{I}]\!](1-[\![\mathrm{I}]\!]) - \gamma[\![\mathrm{I}]\!]. \tag{42}$$

The steady-state solutions are then

$$[\![\mathrm{I}]\!]_1^* = 0, [\![\mathrm{I}]\!]_2^* = 1 - \frac{\gamma}{\kappa\beta}.$$

At $\beta/\gamma = \kappa^{-1}$, the solution $[\![\mathrm{I}]\!]_1^*$ becomes unstable due to a transcritical bifurcation, also known as the epidemic threshold in epidemiology.

2. *MF2*

The second-order mean field originates from the Bethe approximation in statistical physics [10] and is now commonly known as the "pair approximation" [15–20,31,49,50]. MF2 neglects dependence beyond distance 1 and is obtained by considering all dynamically relevant motifs up to size 2. Noting that motifs without infecteds are dynamically irrelevant and omitting zero blocks, we obtain

$$\begin{array}{c} \\ \dot{[\mathrm{I}]} \\ \dot{[\mathrm{I{-}S}]} \\ \dot{[\mathrm{I{-}I}]} \end{array}
\begin{array}{c}
\begin{array}{ccccccc} [\mathrm{I}] & [\mathrm{I{-}S}] & [\mathrm{I{-}I}] & [\mathrm{S{-}S{-}I}] & [\mathrm{I{-}S{-}I}] & [\triangle_{\mathrm{SSI}}] & [\triangle_{\mathrm{ISI}}] \end{array} \\
\left[\begin{array}{c|cc|cccc} -\gamma & \beta & 0 & & & & \\ \hline & -\beta-\gamma & \gamma & \beta & -\beta & \beta & -\beta \\ & 2\beta & -2\gamma & 0 & 2\beta & 0 & 2\beta \end{array}\right].
\end{array}$$

Hence, two types of order 3 motifs appear on the right-hand side: chains and triangles. For the networks we consider in this paper, triangular subgraphs are either not present (in case of square, cubic, hypercubic lattices) or negligible for large N (e.g., for Erdős-Rényi random networks [39]), so we only need to consider the system

$$\begin{array}{c} \\ \dot{[\mathrm{I}]} \\ \dot{[\mathrm{I{-}S}]} \\ \dot{[\mathrm{I{-}I}]} \end{array}
\begin{array}{c}
\begin{array}{ccccc} [\mathrm{I}] & [\mathrm{I{-}S}] & [\mathrm{I{-}I}] & [\mathrm{S{-}S{-}I}] & [\mathrm{I{-}S{-}I}] \end{array} \\
\left[\begin{array}{c|cc|cc} -\gamma & \beta & 0 & & \\ \hline & -\beta-\gamma & \gamma & \beta & -\beta \\ & 2\beta & -2\gamma & 0 & 2\beta \end{array}\right].
\end{array}$$

The number of conservation relations used for elimination depends on whether the networks have a homogeneous degree.

a. Networks with homogeneous degree. In this case, the conservation relations are

$$[\mathrm{I}] + [\mathrm{S}] = N, \tag{43}$$

$$[\mathrm{I{-}S}] + [\mathrm{S{-}S}] = \kappa\,[\mathrm{S}]\,, \tag{44}$$

$$[\mathrm{I{-}I}] + [\mathrm{I{-}S}] = \kappa\,[\mathrm{I}]\,,. \tag{45}$$

However, due to the dynamic irrelevance of $[\mathrm{S}]$ and $[\mathrm{S{-}S}]$, only Eq. (45) can be used to eliminate further variables. We use it to eliminate $[\mathrm{I{-}I}] = \kappa[\mathrm{I}] - [\mathrm{I{-}S}]$. Following Eqs. (11)–(13), this means

$$\mathbf{x} = \left[[\mathrm{I}]\ \ [\mathrm{I{-}S}]\ \ [\mathrm{I{-}I}]\right]^T,$$

$$\tilde{\mathbf{x}} = \begin{bmatrix} [\mathrm{I}] \\ [\mathrm{I{-}S}] \end{bmatrix},\ \tilde{\mathbf{x}}_3 = \begin{bmatrix} [\mathrm{S{-}S{-}I}] \\ [\mathrm{I{-}S{-}I}] \end{bmatrix},$$

and

$$\mathbf{E} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \kappa & -1 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

$$\tilde{\mathbf{Q}}_{1\cdots2,1\cdots2} = \begin{bmatrix} -\gamma & \beta & 0 \\ 0 & -\beta-\gamma & \gamma \end{bmatrix}, \quad \tilde{\mathbf{Q}}_{23} = [\beta \quad -\beta],$$

such that we can calculate $\mathbf{Q}'$, $\tilde{\mathbf{c}}$ to obtain

$$\begin{array}{c} \\ \dot{[\mathrm{I}]} \\ \dot{[\mathrm{I{-}S}]} \end{array}
\begin{array}{c}
\begin{array}{cccc} [\mathrm{I}] & [\mathrm{I{-}S}] & [\mathrm{S{-}S{-}I}] & [\mathrm{I{-}S{-}I}] \end{array} \\
\left[\begin{array}{c|c|cc} -\gamma & \beta & & \\ \hline \gamma\kappa & -\beta-2\gamma & \beta & -\beta \end{array}\right].
\end{array} \tag{46}$$

Applying the closure (30) for degree-homogeneous networks [resulting from Eq. (23)],

$$[\mathrm{S{-}S{-}I}] \approx \kappa(\kappa-1)N\frac{\frac{[\mathrm{S{-}S}]}{\kappa N}\frac{[\mathrm{I{-}S}]}{\kappa N}}{\frac{[\mathrm{S}]}{N}} = \frac{\kappa-1}{\kappa}\frac{[\mathrm{S{-}S}][\mathrm{I{-}S}]}{[\mathrm{S}]},$$

$$[\mathrm{I{-}S{-}I}] \approx \kappa(\kappa-1)N\frac{\frac{[\mathrm{I{-}S}]^2}{\kappa^2N^2}}{\frac{[\mathrm{S}]}{N}} = \frac{\kappa-1}{\kappa}\frac{[\mathrm{I{-}S}]^2}{[\mathrm{S}]}, \tag{47}$$

we obtain the final nonlinear system

$$\begin{array}{c} \\ \dot{[\mathrm{I}]} \\ \dot{[\mathrm{I{-}S}]} \end{array}
\begin{array}{c}
\begin{array}{cccc} [\mathrm{I}] & [\mathrm{I{-}S}] & \frac{[\mathrm{S{-}S}][\mathrm{I{-}S}]}{[\mathrm{S}]} & \frac{[\mathrm{I{-}S}]^2}{[\mathrm{S}]} \end{array} \\
\left[\begin{array}{c|c|cc} -\gamma & \beta & & \\ \hline \gamma\kappa & -\beta-2\gamma & \beta\frac{\kappa-1}{\kappa} & -\beta\frac{\kappa-1}{\kappa} \end{array}\right],
\end{array} \tag{48}$$

which, after elimination of $[\mathrm{S}]$ and $[\mathrm{S{-}S}]$ via Eqs. (43) and (44) and normalization via Eq. (8) becomes

$$\begin{array}{c} \\ \dot{[\![\mathrm{I}]\!]} \\ \dot{[\![\mathrm{I{-}S}]\!]} \end{array}
\begin{array}{c}
\begin{array}{cccc} [\![\mathrm{I}]\!] & [\![\mathrm{I{-}S}]\!] & \left(1-\frac{[\![\mathrm{I{-}S}]\!]}{1-[\![\mathrm{I}]\!]}\right)[\![\mathrm{I{-}S}]\!] & \frac{[\![\mathrm{I{-}S}]\!]^2}{1-[\![\mathrm{I}]\!]} \end{array} \\
\left[\begin{array}{c|c|cc} -\gamma & \kappa\beta & & \\ \hline \gamma & -\beta-2\gamma & \beta(\kappa-1) & -\beta(\kappa-1) \end{array}\right],
\end{array} \tag{49}$$

From the steady-state solutions (E1), we find that the epidemic threshold is now located at $\beta/\gamma = (\kappa-1)^{-1}$. We also derived nontrivial steady-state correlations via Eq. (40) in Eq. (E2).

b. Networks with heterogeneous degree. Here, Eqs. (44) and (45) do not hold, but the total frequency of any given subgraph is still conserved. Hence, the conservation relations are Eq. (43) from order 1 and

$$[\mathrm{S{-}S}] + [\mathrm{I{-}I}] + 2\,[\mathrm{I{-}S}] = \kappa N. \tag{50}$$

This means that we cannot eliminate $[\mathrm{I{-}I}]$ here (unlike in case of networks with homogeneous degree), such that we have three instead of two equations. Also applying the closure

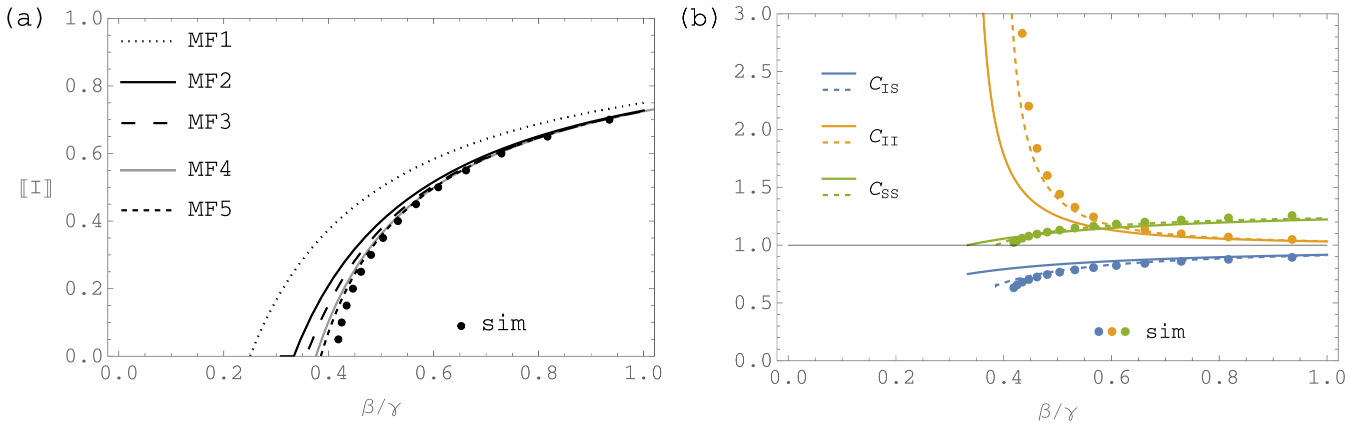


FIG. 2. Comparison of the mean-field approximations of order 1 to 5 (lines) with simulations (markers) of SIS epidemic spreading on a square lattice as a function of β/γ: (a) nontrivial steady states, (b) correlations at distance 1 [shown for MF2 (solid), MF5 (dashed), and simulations (dots)].

for ER networks (32) [resulting from Eq. (23)],

$$[S\text{-}S\text{-}I] \approx \frac{[S\text{-}S]\,[I\text{-}S]}{[S]}, \quad [I\text{-}S\text{-}I] \approx \frac{[I\text{-}S]^2}{[S]}, \tag{51}$$

we obtain

$$\begin{array}{cc} & \begin{array}{ccccc} [I] & [I\text{-}S] & [I\text{-}I] & \frac{[S\text{-}S][I\text{-}S]}{[S]} & \frac{[I\text{-}S]^2}{[S]} \end{array} \\ \begin{array}{c} \dot{[I]} \\ \dot{[I\text{-}S]} \\ \dot{[I\text{-}I]} \end{array} & \left[\begin{array}{c|cc|cc} -\gamma & \beta & 0 & & \\ \hline & -\beta-\gamma & \gamma & \beta & -\beta \\ & 2\beta & -2\gamma & 0 & 2\beta \end{array}\right]. \end{array} \tag{52}$$

After substitution of $[S]$ and $S\text{-}S$ via the conservation relations (43) and (50) and normalization via Eq. (8), this becomes

$$\begin{array}{cc} & \begin{array}{ccccc} [\![I]\!] & [\![I\text{-}S]\!] & [\![I\text{-}I]\!] & \frac{(1-2[\![I\text{-}S]\!]-[\![I\text{-}I]\!])[\![I\text{-}S]\!]}{1-[\![I]\!]} & \frac{[\![I\text{-}S]\!]^2}{1-[\![I]\!]} \end{array} \\ \begin{array}{c} \dot{[\![I]\!]} \\ \dot{[\![I\text{-}S]\!]} \\ \dot{[\![I\text{-}I]\!]} \end{array} & \left[\begin{array}{c|cc|cc} -\gamma & \kappa\beta & 0 & & \\ \hline & -\beta-\gamma & \gamma & \kappa\beta & -\kappa\beta \\ & 2\beta & -2\gamma & 0 & 2\kappa\beta \end{array}\right]. \end{array} \tag{53}$$

There are three steady states, of which two are in the admissible range (E3). Here, the epidemic threshold is located at $\beta/\gamma = \kappa^{-1}$, as in the first-order mean-field model. The steady-state correlations are given in Eq. (E4).

3. *MF3-MF5*

Approximations of higher order than two correspond to cluster variation approximations in statistical physics [11]. MF3 results from neglecting dependence beyond distance 2 and is commonly known as the "triple approximation" [21]. Its step-by-step derivation for the square lattice is shown in Appendix F and results in four equations (after substitution with conservation relations). In Appendix G, we show the derivation of the unclosed MF4 for a general network as `Mathematica` notebook output. We also derived the closed MF4 and MF5 for the square lattice. The number of equations after elimination with conservation relations is, respectively, 14 and 37. The derivation of MF4 with closure is shown in Appendix G point 2. The steady states of MF4 and MF5 are shown in Fig. 2 of Sec. VIII B.

B. Comparison to simulations

Here, we compare the steady states of MF1-5 of SIS epidemic spreading with those of the simulations on a selection of network types: lattices, regular random networks, and Erdős-Rényi random networks.

a. Square lattice. In Fig. 2(a), we show, for the square lattice, the steady-state fraction of infecteds versus β/γ for MF1-MF5 compared to simulations. The steady states of the mean-field models get closer to those of the simulation with increasing order. All mean-field models have an increasing bias in their nontrivial (endemic) steady states when approaching the critical value of β/γ from above. Figure 2(b) compares the steady-state distance-1 correlations between species types from MF2 and MF5 to those in simulations. Species of the same type cluster whereas different species tend to avoid each other (compared to a random distribution). The infected-infected correlation diverges when approaching the critical value of β/γ from above and has a singularity at the bifurcation. For example, for MF2, via Eq. (E2), we have $C^*_{II}(\beta/\gamma) \to \infty$ for $\beta/\gamma \searrow (\kappa-1)^{-1}$ (limit from above in the endemic equilibrium). We recall that the errors in the mean-field models visible in Fig. 2(a) can be due to violation of the statistical dependence assumption beyond distance k, nonchordality of the independence map, and violation of the spatial homogeneity assumption. Spatial homogeneity can be violated in two ways: via heterogeneity of structure and via heterogeneity of dynamics [22]. As the structure of a lattice is homogeneous and we only study steady states (i.e., there are no dynamics) that are spatially homogeneous [see Fig. 6(b)], spatial inhomogeneity can not be a source of bias here. This leaves statistical dependence and nonchordality as only sources of bias. In the square lattice there are always cycles with diameter larger than d for any chosen d along which unaccounted for information can spread, unless d is greater than or equal to the graph diameter of the entire lattice. This means that mean-field models that do not consider

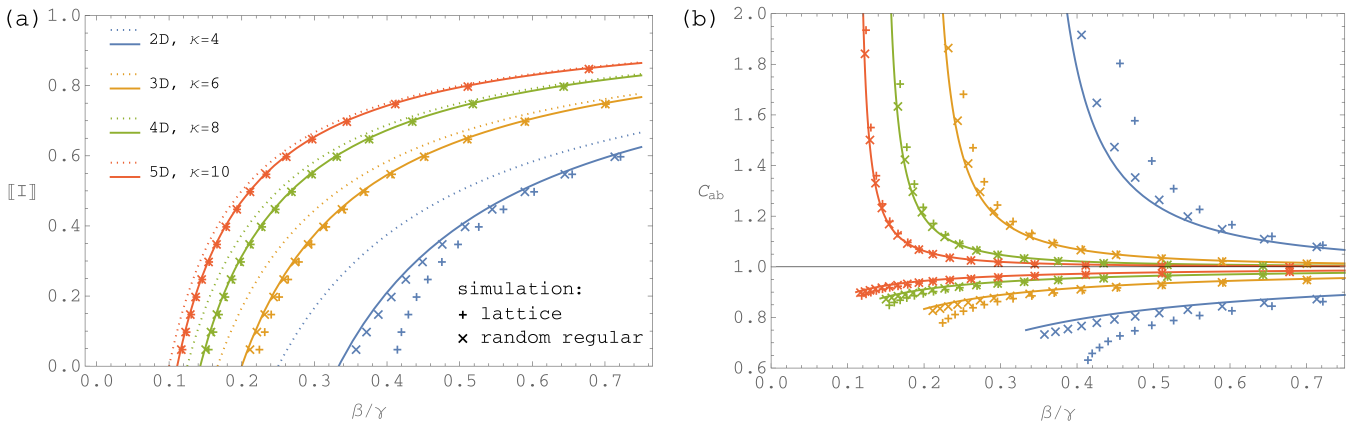


FIG. 3. Comparison of the mean-field approximations (dotted lines: MF1, solid lines: MF2) and simulations (markers) of SIS epidemic spreading on lattices and random regular networks with number of neighbors 4,6,8,10 as a function of β/γ: (a) nontrivial steady states, (b) correlations: C^*_{II} (above 1), C^*_{SI}(below 1).

motifs of size up to the network diameter minus one cannot be exact.

As MF1 assumes no correlation between the states of neighboring nodes, the distance between the horizontal line through 1 and the + markers of the correlations in the simulations is a measure of the bias of MF1 due to neglection of correlations in MF1 closures. Likewise, the distance between the steady-state MF2/MF5 correlations and the simulations is due to neglection of (higher-order and conditional) dependence in MF2/MF5 closures and higher-level nonchordality. All models have larger biases closer to the critical point, where higher-order correlations become more important. This is a well-known characteristic of continuous phase transitions, in which correlations occur on increasingly long ranges when approaching the phase transition (bifurcation). Figure 7 in Appendix D shows the correlation as a function of distance from a central point [via Eq. (D1)], for various values of β/γ. It shows, as expected from phase transitions theory, that correlations at any distance are larger closer to the critical point. At the critical point, theory shows that correlations occur at all distances [13].

Comparing the square lattice to the 4-neighbor random regular graph, we can see that there is also a phase transition, but there is substantially less bias than in the square lattice (Fig. 3, blue × versus +). This is because random regular graphs are locally treelike, and hence, unlike in the square lattice, correlations over longer distances can be captured well via a decomposition of larger motifs into links (MF2). The small remaining bias in the 4-regular random graph we suspect to be because the assumed conditional independence is not valid for all states [2,3].

b. General cubic lattices and random regular networks. We show in Fig. 3 how the steady states and correlations in MF1, MF2 and simulations depend on the number of neighbors in d-dimensional cubic lattices and random regular networks. When d is the lattice dimension, the lattice degree is $\kappa = 2\mathsf{d}$. The observations of the square lattice generalize to cubic and hypercubic lattices and random regular networks (at least up to $\kappa = 10$): (i) there is a transcritical bifurcation at a particular value of β/γ, where C^*_{II} becomes singular, (ii) MF1 and MF2 capture qualitatively the steady-state fraction of infecteds and correlations are captured qualitatively by MF2, (iii) MF2 is less biased than MF1, (iv) the bias is larger closer to the bifurcation. According to MF1, the bifurcation occurs at $\kappa^{-1} = (2\mathsf{d})^{-1}$ and according to MF2 at $(2\mathsf{d}-1)^{-1} = (\kappa-1)^{-1}$ (see Secs. VIII A 1 and VIII A 2). Liggett [51] proved that for lattices, the critical value predicted by MF2 is a lower bound. Figure 3(a) shows that MF2 and this lower bound is approached increasingly closely when the lattice dimension increases. Due to higher clustering of neighbors, the epidemic threshold in lattices is higher than that in the corresponding random regular network [30], but this difference decreases with dimension and degree [Fig. 3(a)]. The steady states of a five-dimensional hypercubic lattice and of a random regular network with degree 10 are indistinguishable from each other and from MF2. This is because random walks in space of dimension 5 or higher have a finite number of intersections almost surely [52–54]. If the path along which an infection travels is seen as a random walk, then having many intersections in $\mathsf{d} \leqslant 4$ means that one cannot ignore alternative infection paths. In $\mathsf{d} > 4$, it is harder for infections to travel via alternative paths to the same point, and hence those paths resemble trees more closely. Hence, as explained above and in Sharkey and Wilkinson [3], MF2 should then be more accurate.

c. Erdős-Rényi random networks. Finally, we show in Fig. 4 how the steady states and correlations in MF1, MF2, and simulations depend on the number of neighbors in an Erdős-Rényi random network. Recall that MF2 on an Erdős-Rényi random network is different from that of the networks above because it has one fewer conservation relation. As above, the behaviour of the steady-state solutions is captured qualitatively by MF1 and MF2 and of the correlations by MF2 alone. Also as above, networks with a larger degree have lower biases and MF2 is better than MF1, but now there seems to be a slight increase of bias with β/γ, at least in the range inspected. Despite spatial heterogeneity of the degree in Erdős-Rényi random networks, there is considerably less bias than in lattices, confirming that the presence of cycles beyond closure distance is the dominant

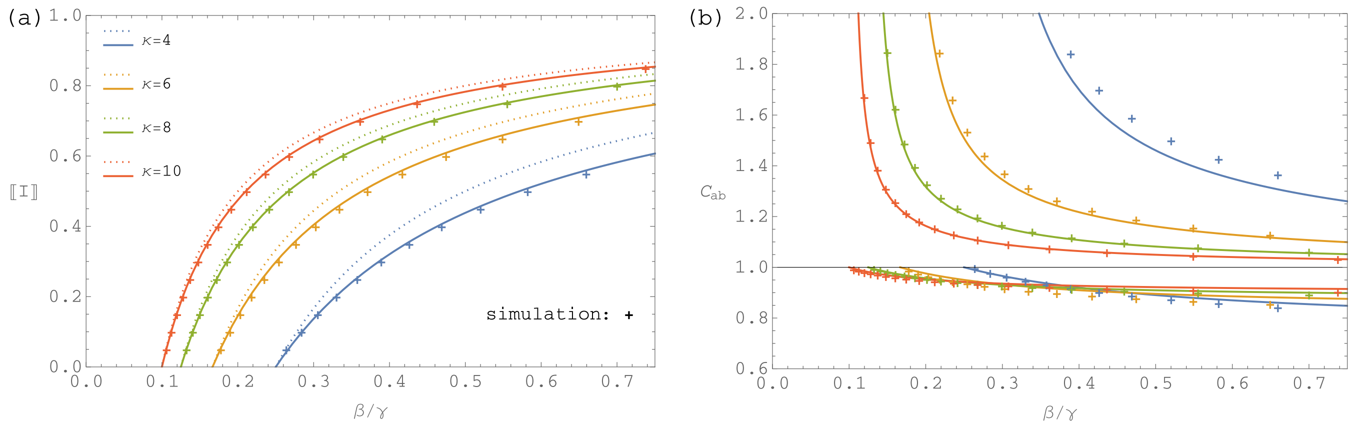


FIG. 4. Comparison of the mean-field approximations (dotted lines: MF1, solid lines: MF2) and simulations (markers) of SIS epidemic spreading on an Erdős-Rényi random network with number of neighbors 4, 6, 8, 10 as a function of β/γ: (a) nontrivial steady states, (b) correlations: C^*_{II} (above 1), C^*_{SI}(below 1).

cause for mean-field model biases in the steady states of SIS spreading.

IX. SUMMARY AND CONCLUSIONS

Previous work found that exact closed individual-level moment equations exist for SIR spreading on arbitrary networks, with the requirement to consider larger-sized motifs, and therefore more equations, for networks that are decreasingly treelike [3]. While for other dynamics than SIR spreading it may not be possible to prove exactness for a finite number of closed moment equations, it is generally found that accuracy increases with the order of approximation [e.g., 21,31] (which was confirmed here). Feasibility of automated derivation of exact closed moment equations for SIR epidemic spreading was shown by Ref. [3], while an automated procedure to derive unclosed moment equations for arbitrary dynamics was developed by Ref. [33]. We developed an automated procedure to both derive and close population-level moment equations for arbitrary dynamics on networks at any approximation order, allowing us to consider mean-field models of higher orders than typically derived by hand. For this purpose, we developed a method to derive closure schemes from predefined independence assumptions. Our closure formulas rely, besides the requirements of spatial homogeneity and large network size, on the assumption of conditional independence beyond distance k and k chordality of the considered network. Consistently, our simulations of SIS epidemic spreading showed that, at given approximation order, the largest biases occurred for networks with many short cycles of any size, such as lattices, in parameter regimes with long-range correlations, such as near continuous phase transitions. Note however that, for lattices, we found the bias of mean-field models to decrease with lattice dimension, which is consistent with results in percolation and phase transitions theory, where it was shown that the importance of cycles decreases with the lattice dimension [52–54]. We also showed that the conventional procedure of truncation at a maximum motif size instead of at a maximum motif diameter necessitates independence assumptions that are inconsistent for different approximated motifs or that may be incompatible with the network (see Sec. VII C e). This suggests that choosing a moment space that consists of motifs at increasing diameter instead of increasing size would lead to more accurate mean-field models.

Whereas our method still needs to be tested more widely, we expect it to lend itself well to study dynamics on networks with density and size of short cycles between that of random networks and (low-dimensional) lattices, particularly when the network is structurally homogeneous. In these cases, derivation by hand may be too tedious while the final set of moment equations is still more manageable than the Markov chain simulations. For networks with considerable degree heterogeneity and/or community structure, we expect approximate master equation methods [24–28] to be more efficient. Our approach focused on static networks with at most nearest-neighbor interactions, but it can be extended to adaptive networks such as those studied in [20,29,33], and to dynamics with higher-order interactions [55]—requiring reaction rate tensors $\mathbf{R}^p$ with $p \geqslant 2$. Our approximation scheme of Sec. VII can be applied more generally, to understand precisely which independence assumptions are taken in other existing types of mean-field approximations than moment closure, or to devise new mean-field approximations. It may also serve to extend the use of message-passing methods [56–58] for epidemic modeling to graphs with cycles.

While we were able to derive closure formulas by assuming statistical independence, leading to mean-field models, other assumptions can be used to obtain closures [1], such as maximum entropy [5], or timescale separation between moments at different orders [6]. Sometimes one can find an appropriate moment space and closure by taking account of the characteristic features of the process in consideration, leading to a description with greater efficiency compared to what is obtainable by using the size-based moment space and closing via independence assumptions [59]. We expect that for finding good moment spaces and closures, equation-free and machine-learning methods [60,61] will play an important role, in particular because the most appropriate low-dimensional descriptions or their closures may not necessarily be available in closed form [5,6,60,61]. It is subject of future work to explore if and how these different approaches relate.

ACKNOWLEDGMENT

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) Grants No. EP/N023544/1 and No. EP/V04687X/1.

APPENDIX A: DERIVATION OF THE MOMENT EQUATIONS

In this section, we derive expressions for the expected rate of change and conservation relations of motif counts, first shown in Eqs. (8) and (9) of Sec. V. Both can be seen as invariance relations, the former being of a differential type, and derived from the master equation for the Markov chain on the network, and the latter of an algebraic type, following from the property that the network is fixed.

a. Master equation for transitions in a Markov chain. We recall that the state of the system for our network with N nodes is given by (using $:=$ for "defined as") $X := (X_1, X_2, ..., X_N) \in \{1, \dots, n\}^N$, where X_i is the label of the species that occupies node i. Hence, the total number of states is n^N. The probabilistic transition from one state to another, following a discrete-state continuous-time Markov chain, defines an evolution equation for the probability of being in each of these states, the so-called master equation (also known as the Kolmogorov-forward equation for a Markov jump process [62]). The probability density $P(X, t)$ for a particular state X at time t changes according to

$$\frac{d}{dt}P(X,t) = \sum_{X'\neq X}[w(X' \to X)P(X',t) - w(X \to X')P(X,t)], \tag{A1}$$

where w denotes the transition rate between system states. If we define $\mathbf{W}$ as a $n^N \times n^N$ transition rate matrix with nondiagonal entries $w(X' \to X)$ and diagonal entries $-\sum_{X'\neq X} w(X \to X')$, then we can rewrite the master equation as

$$\dot{\mathbf{P}}(t) = \mathbf{W}\mathbf{P}(t),$$

which describes the evolution of the density for all states as elements of a vector $\mathbf{P}$ [and not just of one particular state as in Eq. (A1)]. Because almost surely at most one node can change state at any one time t, only states differing from each other in one node can be directly transitioned between. Therefore, $\mathbf{W}$ must be sparse, having only $N(n-1)$ entries in each row or column, as each of the N nodes can convert to any of the $n-1$ other species. This permits writing Eq. (A1) as

$$\frac{d}{dt}P(X,t) = \sum_{i=1}^{N}\sum_{k\neq X_i}^{n}[w_i(X_{i\to k} \to X)P(X_{i\to k},t) - w_i(X \to X_{i\to k})P(X,t)], \tag{A2}$$

where

$$X \to X_{i\to k} : (X_1, ..., X_i, ..., X_N) \mapsto (X_1, ..., k, ..., X_N)$$

is the operator that replaces the species at the ith node by species k, and $w_i(.)$ is the conversion rate at node i.

b. Motifs and their frequencies. In what follows, we will derive from the master equation the evolution of the frequency (or total count) of the *motifs* $\boldsymbol{x}^{\mathsf{a}}$, as defined in Sec. II. Recall that these motifs, are defined by their state label vector $\boldsymbol{x}$ of size m and the adjacency structure between motif nodes a. We denote a single *occurrence* at a given $\boldsymbol{i} \in S(m, N)$ as

$$[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}](t) := \delta_{\mathsf{a}}(\mathsf{A}_{\boldsymbol{i}})\delta_{\boldsymbol{x}}(X_{\boldsymbol{i}}(t)), \tag{A3}$$

which requires an exact match of the adjacency a by $\mathsf{A}_{\boldsymbol{i}}$ and state label vector $\boldsymbol{x}$ by $X_{\boldsymbol{i}}$. For example, if A contains a connected triangle between nodes 1, 2, and 3, then motifs with $\mathsf{a} = \{(1, 2), (2, 3)\}$ would not occur on $\boldsymbol{i} = (1, 2, 3)$. The total count of motifs in the large network is then the number of such exact matches, which is obtained by summing over all indices $\boldsymbol{i} \in S(m, N)$, as shown in Eq. (3). Since the large network structure is constant in time, we can use the counts of induced subgraphs $[\mathsf{a}]$ given by Eq. (8) as a normalization factor for the (variable) counts of motifs such that we may consider normalized motif frequencies

$$[\![\boldsymbol{x}^{\mathsf{a}}]\!] := [\boldsymbol{x}^{\mathsf{a}}]/[\mathsf{a}]. \tag{A4}$$

c. Evolution of expected counts. Total and normalized counts $[\boldsymbol{x}^{\mathsf{a}}]$ and $[\![\boldsymbol{x}^{\mathsf{a}}]\!]$ refer to realizations of states X on the large random network such that they are random variables. Similarly, $[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}](X)$ is a random variable in $\{0, 1\}$ for each index vector $\boldsymbol{i} \in S(m, N)$ once we take the randomness of states X into account. Its expectation is

$$\langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle = \sum_{X}[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}](X)P(X,t). \tag{A5}$$

The master equation (A2) for the density P implies that the expectation satisfies the differential equation

$$\begin{aligned}
\frac{d}{dt}\langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle &= \sum_{X}[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}](X)\frac{d}{dt}P(X,t) = \sum_{X}\left\{\delta_{\mathsf{a}}(\mathsf{A}_{\boldsymbol{i}})\delta_{\boldsymbol{x}}(X_{\boldsymbol{i}})\sum_{i'=1}^{N}\sum_{k\neq X_{i'}}^{n}[w_{i'}(X_{i'\to k} \to X)P(X_{i'\to k},t) - w_{i'}(X \to X_{i'\to k})P(X,t)]\right\}, \\
&= \left\langle \delta_{\mathsf{a}}(\mathsf{A}_{\boldsymbol{i}})\sum_{i'=1}^{N}\sum_{k\neq X_{i'}}^{n}[\delta_{\boldsymbol{x}}((X_{i'\to k})_{\boldsymbol{i}}) - \delta_{\boldsymbol{x}}(X_{\boldsymbol{i}})]w_{i'}(X \to X_{i'\to k})\right\rangle, \\
&= \left\langle \delta_{\mathsf{a}}(\mathsf{A}_{\boldsymbol{i}})\sum_{p=1}^{m}\sum_{k\neq X_{i_p}}^{n}\left[\delta_{\boldsymbol{x}}\left(\left(X_{i_p\to k}\right)_{\boldsymbol{i}}\right) - \delta_{\boldsymbol{x}}(X_{\boldsymbol{i}})\right]w_{i_p}\left(X \to X_{i_p\to k}\right)\right\rangle, \\
&= \left\langle \delta_{\mathsf{a}}(\mathsf{A}_{\boldsymbol{i}})\sum_{p=1}^{m}\sum_{k\neq X_{i_p}}^{n}\delta_{\boldsymbol{x}_{p\to\emptyset}}(X_{\boldsymbol{i}_{p\to\emptyset}})\left[\delta_{x_p}(k) - \delta_{x_p}\left(X_{i_p}\right)\right]w_{i_p}\left(X \to X_{i_p\to k}\right)\right\rangle,
\end{aligned} \tag{A6}$$

where in the second step, we substituted $\delta_{\boldsymbol{x}}(X_{\boldsymbol{i}})w_{i'}(X_{i'\to k} \to X)P(X_{i'\to k}, t)$ for $\delta_{\boldsymbol{x}}((X_{i'\to k})_{\boldsymbol{i}})w_{i'}(X \to X_{i'\to k})P(X, t)$ which corresponds to a reordering of the terms in $\sum_X$, and where we have used the notation $\langle\cdot\rangle$ for the expectation $\sum_X(\cdot)P(X, t)$. In the third step, we used that only changes to the nodes belonging to $\boldsymbol{i}$ matter. In the last step, we factored out the common elements in the δ functions corresponding to all but the pth element of $X_{\boldsymbol{i}}$ and $\boldsymbol{x}$, and we used the subscript notation $(\cdot)_{p\to\emptyset}$ to denote a vector with element p removed. The expression on the right-hand side in Eq. (A6) can be understood independent of the prior algebraic manipulations: the expected change rate of $\langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle$ equals the sum of expected rates for each of the nodes i_p in $\boldsymbol{i}$ changing its state *to* x_p minus the rate of node i_p changing its state *from* x_p.

d. Conversion rates. Next, we will insert the two types of admissible transitions (as discussed in Sec. II) into Eq. (A6), namely spontaneous conversions with rates given in matrix $\mathbf{R}^0 \in \mathbb{R}^{n\times n}$, and, conversions due to interactions with a single nearest neighbor with rates given in $\mathbf{R}^1 \in \mathbb{R}^{n\times n\times n}$. The diagonal entries of $\mathbf{R}^0, \mathbf{R}^1$ are zero without loss of generality. For these transition types, the rates w_{i_p} in Eq. (A6) are

$$w_{i_p}(X \to X_{i_p\to k}) = \sum_{a,c}^{n} R^1_{akc}\delta_a(X_{i_p}) \sum_j^N \mathsf{A}_{i_p j}\delta_c(X_j) + \sum_a^n R^0_{ak}\delta_a(X_{i_p}). \tag{A7}$$

With the rates in Eq. (A7), and noting that

$$\delta_{x_p}(k) - \delta_{x_p}(X_{i_p}) = \begin{cases} 1 & \text{if } x_p = k, \\ -1 & \text{if } x_p = X_{i_p}, \\ 0 & \text{if } k \neq x_p \text{ and } x_p \neq X_{i_p}, \end{cases}$$

the sum inside the averaging brackets in Eq. (A6) has the form

$$\left[\sum_{a,c}^{n} R^1_{ax_pc}\delta_a(X_{i_p}) \sum_j^N \mathsf{A}_{i_p j}\delta_c(X_j) + \sum_a^n R^0_{ax_p}\delta_a(X_{i_p})\right]\delta_{\boldsymbol{x}_{p\to\emptyset}}(X_{\boldsymbol{i}\setminus i_p}) - \sum_{k\neq X_{i_p}}^{n}\left[\sum_c^n R^1_{x_pkc}\sum_j^N \mathsf{A}_{i_p j}\delta_c(X_j) + R^0_{x_pk}\right]\delta_{\boldsymbol{x}}(X_{\boldsymbol{i}}),$$

where we used for the last two terms that $\delta_{x_p}(X_{i_p})\delta_{\boldsymbol{x}_{p\to\emptyset}}(X_{\boldsymbol{i}\setminus i_p})$ can be combined to $\delta_{\boldsymbol{x}}(X_{\boldsymbol{i}})$.

e. Differential equations for motif counts. When distributing the products and exploiting the linearity of the averaging brackets, the differential equation (A6) for the expected rate of change at $\boldsymbol{i}$ becomes

$$\frac{d}{dt}\langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle = \delta_{\mathsf{a}}(\mathsf{A}_{\boldsymbol{i}})\sum_p^m\left[\sum_{a,c}^n R^1_{ax_pc}\left\langle \delta_{\boldsymbol{x}_{p\to\emptyset}}(X_{\boldsymbol{i}_{p\to\emptyset}})\delta_a(X_{i_p})\sum_j^N \mathsf{A}_{i_p j}\delta_c(X_j)\right\rangle + \sum_a^n R^0_{ax_p}\langle\delta_{\boldsymbol{x}_{p\to\emptyset}}(X_{\boldsymbol{i}_{p\to\emptyset}})\delta_a(X_{i_p})\rangle - \sum_{k\neq X_{i_p},c}^n R^1_{x_pkc}\left\langle\delta_{\boldsymbol{x}}(X_{\boldsymbol{i}})\sum_j^N \mathsf{A}_{i_p j}\delta_c(X_j)\right\rangle - \sum_{k\neq X_{i_p}}^n R^0_{x_pk}\langle\delta_{\boldsymbol{x}}(X_{\boldsymbol{i}})\rangle\right].$$

After replacing index label a by k, using the definition of $[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]$ in Eq. (A3) and using $(\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}})_{p\to k}$ to indicate $\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}$ with its pth element replaced by species k, this becomes

$$\frac{d}{dt}\langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle = \sum_p^m\sum_k^n\left[\sum_c^n R^1_{kx_pc}\left\langle\left[(\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}})_{p\to k}\right]\sum_j^N \mathsf{A}_{i_p j}\delta_c(X_j)\right\rangle + R^0_{kx_p}\langle[(\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}})_{p\to k}]\rangle - \sum_c^n R^1_{x_pkc}\left\langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\sum_j^N \mathsf{A}_{i_p j}\delta_c(X_j)\right\rangle - R^0_{x_pk}\langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle\right].$$

This shows that $\langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle$ can increase by a conversion from motifs that differ from $\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}$ in only one node (first two terms), or decrease by having any of the nodes in $\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}$ convert to another species (last two terms), with both increase and decrease possible via interaction with neighbors and via spontaneous conversion. The factors of form $[\cdot]\sum_j^N \mathsf{A}_{i_p j}\delta_c(X_j)$ count the number of c-connections of motif $[\cdot]$ (located at $\boldsymbol{i}$) at node i_p. The neighboring node j with species c can be part of the motif $[\cdot]$, or it can be outside of $[\cdot]$, in which case it gives rise to higher-order motifs. Therefore, by splitting the neighborhood sums as follows:

$$\sum_j^N \mathsf{A}_{i_p j}\delta_c(X_j) = \sum_{j\in\boldsymbol{i}}^N \mathsf{A}_{i_p j}\delta_c(X_j) + \sum_{j\notin\boldsymbol{i}}^N \mathsf{A}_{i_p j}\delta_c(X_j),$$

their products with $[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]$ and $[(\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}})_{p\to k}]$ count contributions of c neighbors from within versus from outside the motif separately:

$$\left[(\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}})_{p\to k}\right]\sum_j^N \mathsf{A}_{i_p j}\delta_c(X_j) = \boldsymbol{\delta}_c(\boldsymbol{x})\mathsf{a}\boldsymbol{e}_p\left[(\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}})_{p\to k}\right] + \sum_{\boldsymbol{y}^{\mathsf{b}}\in\mathcal{N}_p^c((\boldsymbol{x}^{\mathsf{a}})_{p\to k})}\sum_{j\notin\boldsymbol{i}}^N[\boldsymbol{y}_{\boldsymbol{i},j}^{\mathsf{b}}], \tag{A8}$$

$$[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\sum_j^N \mathsf{A}_{i_p j}\delta_c(X_j) = \boldsymbol{\delta}_c(\boldsymbol{x})\mathsf{a}\boldsymbol{e}_p[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}] + \sum_{\boldsymbol{y}^{\mathsf{b}}\in\mathcal{N}_p^c(\boldsymbol{x}^{\mathsf{a}})}\sum_{j\notin\boldsymbol{i}}^N[\boldsymbol{y}_{\boldsymbol{i},j}^{\mathsf{b}}]. \tag{A9}$$

On the right-hand side, $\boldsymbol{e}_p$ is an m-dimensional vector with a 1 at position p and zeros elsewhere, and $\boldsymbol{\delta}_c(\boldsymbol{x})$ a vector Kronecker δ function that returns a vector of the size of $\boldsymbol{x}$ with ones where the elements equal c and zeros elsewhere.

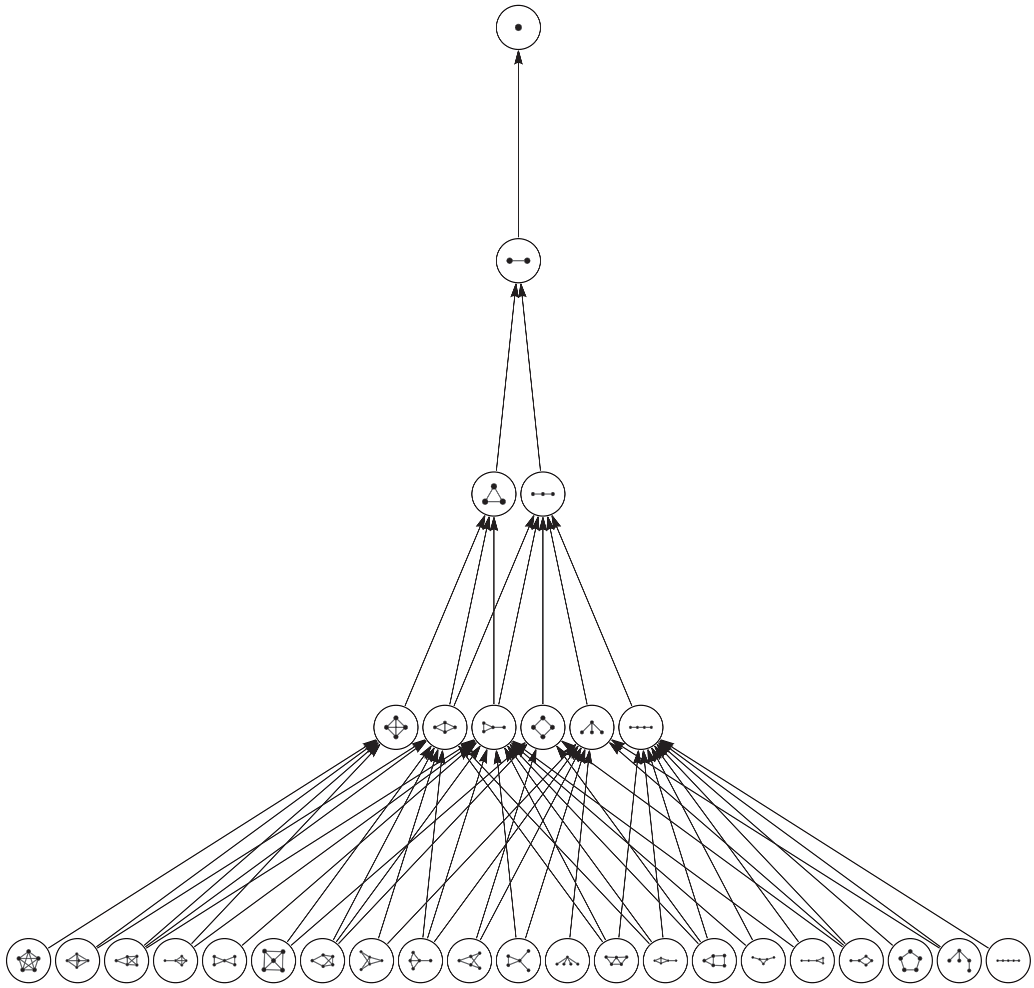


FIG. 5. Dependence of moment equations of a given motif on motifs of one order higher due to nearest-neighbor interactions, up to order 4 (ignoring labels).

Thus, the term $\boldsymbol{\delta}_c(\boldsymbol{x})\mathsf{a}\boldsymbol{e}_p$ counts the number of c-connections at position p in the motif $\boldsymbol{x}^{\mathsf{a}}$. We use the notation $\boldsymbol{i}, j$ for the vector $\boldsymbol{i}$ with an extra node index j appended at position $m+1$. We used $\mathcal{N}_p^c(\boldsymbol{x}^{\mathsf{a}})$ to denote the set of all $(m+1)$th order connected motifs that can be obtained by linking a new c-node to the pth position in motif $\boldsymbol{x}^{\mathsf{a}}$, i.e.,

$$\mathcal{N}_p^c(\boldsymbol{x}^{\mathsf{a}}) := \bigcup_{\ell=1}^{m+1} \big\{\boldsymbol{y}^{\mathsf{b}} : |\boldsymbol{y}| = m+1, y_\ell = c, \boldsymbol{y}^{\mathsf{b}}_{\ell\to\emptyset} = \boldsymbol{x}^{\mathsf{a}}, (\ell, p) \in \mathsf{b}\big\}, \tag{A10}$$

where $\boldsymbol{y}^{\mathsf{b}}_{\ell\to\emptyset}$ denotes the mth order connected motif obtained by deleting the ℓth node of $\boldsymbol{y}^{\mathsf{b}}$. The sum over elements of $\mathcal{N}_p^c(\cdot)$ in Eq. (A9) is taken because any of the other motif nodes can also link to the new node. The types of higher-order motifs appearing in the differential equation depend on the considered motif. In Fig. 5, we show this dependence structure (ignoring the labels).

With the substitutions from above, we obtain the *individual-level moment equations*:

$$\begin{aligned}\frac{d}{dt}\langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle = \sum_p^m \sum_k^n \Bigg\{ &\left(\sum_c^n \boldsymbol{\delta}_c(\boldsymbol{x})\mathsf{a}\boldsymbol{e}_p R^1_{kx_pc} + R^0_{kx_p}\right)\langle[(\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}})_{p\to k}]\rangle \\ &+ \sum_c^n \sum_{\boldsymbol{y}^{\mathsf{b}}\in\mathcal{N}_p^c(\boldsymbol{x}^{\mathsf{a}}_{p\to k})} \sum_{j\notin\boldsymbol{i}}^N R^1_{kx_pc}\langle[\boldsymbol{y}^{\mathsf{b}}_{\boldsymbol{i},j}]\rangle \\ &- \left(\sum_c^n \boldsymbol{\delta}_c(\boldsymbol{x})\mathsf{a}\boldsymbol{e}_p R^1_{x_pkc} + R^0_{x_pk}\right)\langle[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]\rangle \\ &- \sum_c^n \sum_{\boldsymbol{y}^{\mathsf{b}}\in\mathcal{N}_p^c(\boldsymbol{x}^{\mathsf{a}})} \sum_{j\notin\boldsymbol{i}}^N R^1_{x_pkc}\langle[\boldsymbol{y}^{\mathsf{b}}_{\boldsymbol{i},j}]\rangle\Bigg\}.\end{aligned} \tag{A11}$$

The *population-level moment equations* are then obtained by taking the sum $\sum_{\boldsymbol{i}\in S(m,N)}$ [see definitions of $[\boldsymbol{x}^{\mathsf{a}}]$ and $[\boldsymbol{x}_{\boldsymbol{i}}^{\mathsf{a}}]$ in Eqs. (3) and (A3)] of Eq. (A11) over all index sets $\boldsymbol{i}$:

$$\begin{aligned}\frac{d}{dt}\langle[\boldsymbol{x}^{\mathsf{a}}]\rangle = \sum_p^m \sum_k^n \sum_c^n \Bigg\{ &\left(\frac{R^0_{kx_p}}{n} + \boldsymbol{\delta}_c(\boldsymbol{x})\mathsf{a}\boldsymbol{e}_p R^1_{kx_pc}\right)\langle[(\boldsymbol{x}^{\mathsf{a}})_{p\to k}]\rangle \\ &- \left(\frac{R^0_{x_pk}}{n} + \boldsymbol{\delta}_c(\boldsymbol{x})\mathsf{a}\boldsymbol{e}_p R^1_{x_pkc}\right)\langle[\boldsymbol{x}^{\mathsf{a}}]\rangle \\ &+ \sum_{\boldsymbol{y}^{\mathsf{b}}\in\mathcal{N}_p^c(\boldsymbol{x}^{\mathsf{a}}_{p\to k})} R^1_{kx_pc}\langle[\boldsymbol{y}^{\mathsf{b}}]\rangle - \sum_{\boldsymbol{y}^{\mathsf{b}}\in\mathcal{N}_p^c(\boldsymbol{x}^{\mathsf{a}})} R^1_{x_pkc}\langle[\boldsymbol{y}^{\mathsf{b}}]\rangle\Bigg\},\end{aligned} \tag{A12}$$

where we collected the sums at the front. In closing this section, we make the following remarks. (i) The evolution of expected motif counts of size m is a function of expected motif counts of size m and $m+1$. (ii) Equation (A12) leads to the same equation for motifs in the network that are *isomorphic*, where we call two motifs $\boldsymbol{x}^{\mathsf{a}}$ and $\boldsymbol{y}^{\mathsf{b}}$ isomorphic if there exists a permutation π of their node indices that maps them onto each other, i.e., $\forall \boldsymbol{x}^{\mathsf{a}}, \boldsymbol{y}^{\mathsf{b}} : \boldsymbol{x}^{\mathsf{a}} \simeq \boldsymbol{y}^{\mathsf{b}} \iff \exists \pi : \pi \boldsymbol{x} = \boldsymbol{y},\ \pi \mathsf{a} \pi^T = \mathsf{b}$. Hence, different motifs that are isomorphic belong to the same equivalence class. Isomorphic motifs have equal total counts such that we may consider only one representative from each equivalence class. In our implementation, we therefore make sure that we count all isomorphic motifs under a single representative node indexing. (iii) The summation in Eq. (3) over index tuples $\boldsymbol{i} \in S(m, N)$ that we used to go from Eqs. (A11) to (A12) leads to multiple counting of motifs that possess *automorphisms* other than the identity transformation. A motif has an automorphism if it can be mapped onto itself by a permutation of its node indices. An automorphism is therefore an isomorphism with itself, i.e., $\mathrm{Aut}(\boldsymbol{x}^{\mathsf{a}}) := \{\pi : \pi \boldsymbol{x} = \boldsymbol{x},\ \pi \mathsf{a} \pi^T = \mathsf{a}\}$. The multiplicity with which a particular motif $\boldsymbol{x}^{\mathsf{a}}$ is counted [via Eq. (3)] is then equal to $|\mathrm{Aut}(\boldsymbol{x}^{\mathsf{a}})|$.

f. Conservation relations. In fixed networks, the frequencies of induced subgraphs of a given type (e.g., nodes, links, polygons, chains) remain fixed. The sum over all possible motif label orderings then yields these fixed frequencies, or using the notation from above,

$$\sum_{\boldsymbol{x}} [\boldsymbol{x}^{\mathsf{a}}] = [\mathsf{a}], \qquad \sum_{\boldsymbol{x}} [\![\boldsymbol{x}^{\mathsf{a}}]\!] = 1, \tag{A13}$$

where the right is the normalized form of the left [via Eq. (A4)]. For networks with homogeneous degree and motifs and every stub p of motif $\boldsymbol{x}^{\mathsf{a}}$ (also called leaf), i.e., a node with degree 1, there is the additional conservation relation

$$\sum_k [\boldsymbol{x}^{\mathsf{a}}_{p\to k}] = \frac{[\mathsf{a}]}{[\mathsf{a} \setminus p]} [\boldsymbol{x}^{\mathsf{a}}_{p\to\emptyset}], \qquad \sum_k [\![\boldsymbol{x}^{\mathsf{a}}_{p\to k}]\!] = [\![\boldsymbol{x}^{\mathsf{a}}_{p\to\emptyset}]\!], \tag{A14}$$

where the right is again the normalized form of the left [via Eq. (A4)]. Under the conditions mentioned above, $[\mathsf{a}]/[\mathsf{a} \setminus p]$ is the number of out-motif connections at the node that connects to the stub. As Eqs. (A13) and (A14) can be derived for each motif up to the chosen truncation order, they form an additional system of equations that can be used to reduce the dimensionality of the mean-field model via elimination. To avoid multiplicity of equations while deriving relations (A14), we set up one equation per set of stubs that lead to isomorphic variants of the considered subgraph when their indices are permuted.

g. Number of equations. A simple lower bound on the number of equations (before elimination) can be found by considering only chains. At each order there is only one chain graph. Order 1 contributes n equations, where n is the number of possible node states. Each chain of order $m > 1$ has $(\!(\begin{smallmatrix} n \\ m \end{smallmatrix})\!) = \binom{m+n-1}{n} = \frac{(m+n-1)!}{(m-1)!n!}$ ways of labeling its m nodes with n species. The total number of equations from chains is then $n_{\mathrm{ch}} = 2 + \sum_{m=2}^{k} (\!(\begin{smallmatrix} n \\ m \end{smallmatrix})\!)$. For networks in which the number of short cycles goes to zero with $N \to \infty$, as in Erdős-Rényi random networks, the total number of equations is equal to this lower bound plus the number of equations due to nonchain trees. When cycles need to be taken into account however, there will be additional connected motif types at each order; see Table II column n_g [37, Table 4.2.1]. To obtain the number of equations contributed by each of these, one has to consider all labeling orderings, knowing that some orderings lead to isomorphic motifs and hence do not add to the total. We have listed in Table II the total number of equations n_{eq} resulting from our enumeration algorithm for dynamics with $n = 2$, such as SIS epidemic spreading. For a particular network, these are still reduced by the number of motifs not occurring in the considered network and by the number of variables via the conservation relations. Column n_4 in Table II shows the number of equations for the square lattice and column n_{4c} shows the remaining number after eliminating variables by using conservation relations.

TABLE II. Cumulative number of equations (n_{ch}, n_{eq}, n_{4c}) or motif types (n_g) as a function of order k. n_{ch}: number of chain motifs (if the number of species $n = 2$), n_g: number of subgraph types, n_{eq}: total number of equations ignoring conservation relations (if $n = 2$), n_4: number of equations for the square lattice (if $n = 2$), n_{4c}: number of equations for the square lattice after elimination via conservation relations (if $n = 2$).

k	n_{ch}	n_g	n_{eq}	n_4	n_{4c}
1	2	1	2	2	1
2	5	2	5	5	2
3	11	4	15	11	4
4	21	10	65	35	14
5	36	31	419	113	38

APPENDIX B: MOMENT EQUATIONS FOR SIS SPREADING UP TO ORDER 3

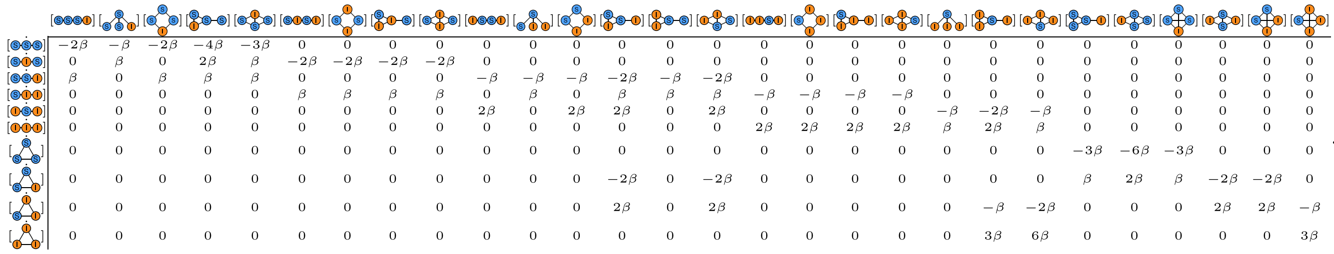
The moment equations up to third order can be written [via Eq. (10)] as

$$\begin{array}{c} \\ \dot{\mathbf{x}}_1 \\ \dot{\mathbf{x}}_2 \\ \dot{\mathbf{x}}_3 \end{array} \begin{array}{c} \begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \end{array} \\ \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_{12} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2 & \mathbf{Q}_{23} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_3 & \mathbf{Q}_{34} \end{bmatrix} \end{array}. \tag{B1}$$

Then, the coefficients for motifs up to order three are (omitting zero blocks)

$$\begin{array}{c|cc|cccccccc}
 & [\mathrm{I}] & [\mathrm{I{-}S}] & [\mathrm{I{-}I}] & [\mathrm{SIS}] & [\mathrm{SSI}] & [\mathrm{SII}] & [\mathrm{ISI}] & [\mathrm{III}] & [\triangle_{\mathrm{SSI}}] & [\triangle_{\mathrm{SII}}] & [\triangle_{\mathrm{III}}] \\
\dot{[\mathrm{I}]} & -\gamma & \beta & 0 & & & & & & & & \\
\hline
\dot{[\mathrm{I{-}S}]} & & -\beta-\gamma & \gamma & 0 & \beta & 0 & -\beta & 0 & \beta & -\beta & 0 \\
\dot{[\mathrm{I{-}I}]} & & 2\beta & -2\gamma & 0 & 0 & 0 & 2\beta & 0 & 0 & 2\beta & 0 \\
\hline
\dot{[\mathrm{SIS}]} & & & & -2\beta-\gamma & 0 & 2\gamma & 0 & 0 & 0 & 0 & 0 \\
\dot{[\mathrm{SSI}]} & & & & 0 & -\beta-\gamma & \gamma & \gamma & 0 & 0 & 0 & 0 \\
\dot{[\mathrm{SII}]} & & & & \beta & \beta & -\beta-2\gamma & 0 & \gamma & 0 & 0 & 0 \\
\dot{[\mathrm{ISI}]} & & & & 0 & 0 & 0 & -2\beta-2\gamma & \gamma & 0 & 0 & 0 \\
\dot{[\mathrm{III}]} & & & & 0 & 0 & 2\beta & 2\beta & -3\gamma & 0 & 0 & 0 \\
\dot{[\triangle_{\mathrm{SSI}}]} & & & & 0 & 0 & 0 & 0 & 0 & -2\beta-\gamma & 2\gamma & 0 \\
\dot{[\triangle_{\mathrm{SII}}]} & & & & 0 & 0 & 0 & 0 & 0 & 2\beta & -2\beta-2\gamma & \gamma \\
\dot{[\triangle_{\mathrm{III}}]} & & & & 0 & 0 & 0 & 0 & 0 & 0 & 6\beta & -3\gamma
\end{array},$$

while those for fourth-order motifs in $\mathbf{Q}_{34}$ are



When written out, this corresponds to the equations:

(B2) (B3) (B4) (B5) (B6) (B7) (B8) (B9) (B10) (B11) (B12)

APPENDIX C: OBTAINING STEADY STATES FROM SIMULATIONS

a. Feedback control. As we aim to compare the steady states of mean-field models to those of the simulation, we need a way to obtain the steady states of the simulations, even if they are unstable or marginally stable. Treating the simulation like an ideal physical experiment, the general approach to finding equilibria regardless of stability is to introduce a stabilizing feedback loop of the form

$$r(t) = r_0 + g([\boldsymbol{x}^{\mathrm{a}}](t) - [\boldsymbol{x}^{\mathrm{a}}]_{\mathrm{ref}}), \tag{C1}$$

as was done in Refs. [46–48] for continuation of unstable vibrations in mechanical experiments. In Eq. (C1), r is one of the conversion rates in $\mathbf{R}^0$ or $\mathbf{R}^1$. The feedback control [63] makes this rate time dependent by coupling it to the motif frequency $[\boldsymbol{x}^{\mathrm{a}}](t)$ of a chosen motif $\boldsymbol{x}^{\mathrm{a}}$ through the relation (C1). The factor g is called the feedback control gain and is problem specific. When performing bifurcation analysis, it is convenient if the rate r used as the control input is also the bifurcation parameter varied for the bifurcation diagram. In this case, whenever the simulation with feedback control (C1) settles to an equilibrium $(r_{\mathrm{c}}^*, [\boldsymbol{x}^{\mathrm{a}}]_{\mathrm{c}}^*)$ (in the limit of large N) the point $(r_{\mathrm{c}}^*, [\boldsymbol{x}^{\mathrm{a}}]_{\mathrm{c}}^*)$ will be on the equilibrium branch

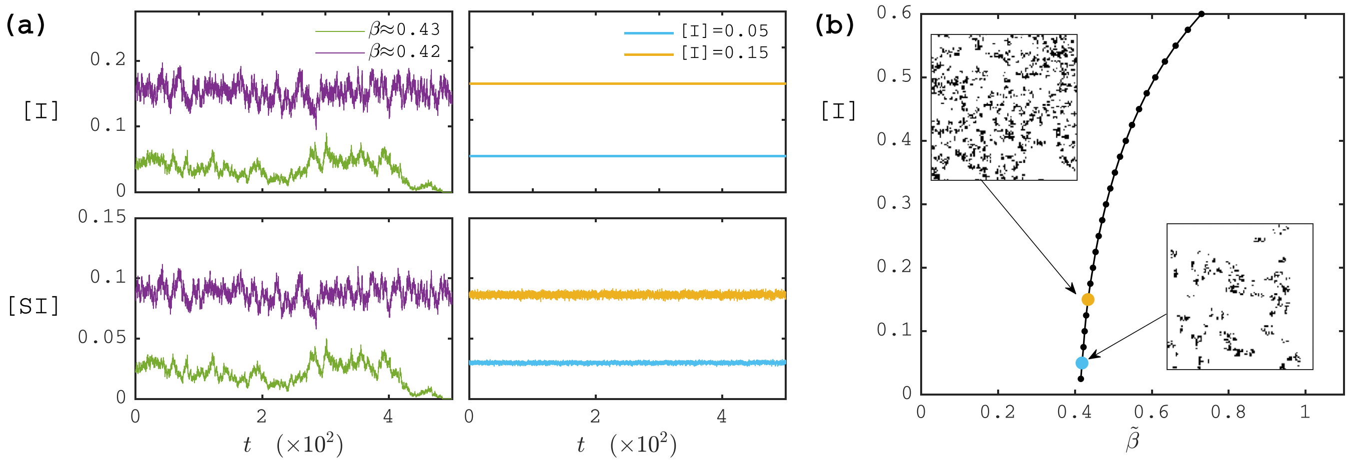


FIG. 6. Feedback control of SIS epidemic spreading on a square lattice: (a) example time profiles of infected fraction [I] and fraction of susceptible-infected links [SI] in conventional (left) and controlled (right) simulation, (b) bifurcation diagram with control; insets: example spatial patterns; grid size 100×100 (white=susceptible, black=infected), gain $g \gg 1$, recovery rate $\gamma = 1$. Near the critical point, the conventional simulation shows large fluctuations and short extinction times (absorption event occurs at the end of the time series for $\beta \approx 0.43$ in green), whereas the controlled simulation has small fluctuations and no extinction.

of the simulation without feedback control [64–66]. For SIS spreading, we choose

$$r = \beta, \quad \boldsymbol{x}^{\mathrm{a}} = \textcircled{I}, \quad g = \infty. \tag{C2}$$

This limit for feedback control results in what is called the conserved contact process, as proposed by Tomé and De Oliveira [67]. While SIS spreading does not have any unstable steady states, points that are marginally stable, as near the continuous phase transition (epidemic threshold) are stabilized as well by the control. This stabilization suppresses fluctuations, even close to the bifurcation, which results in faster convergence of the mean and absence of absorption for any positive $[\textcircled{I}]$ (see Fig. 6).

b. Simulation algorithm. In the limit of infinite gain g, each recovery event forces a simultaneous infection event, such that the number of infected nodes $[\textcircled{I}]$ stays constant. In a simulation based on the Gillespie algorithm [45] this is done in the following steps:

(1) start with a number of randomly distributed infected nodes,

(2) recover an infected node selected uniformly at random,

(3) infect a randomly selected susceptible node, with selection probability proportional to its number of infected neighbors,

(4) advance time with $\Delta t = -\log(\xi)/(\gamma[\textcircled{I}])$ (where ξ is a uniform random variable on the interval [0,1]),

(5) go to step 2.

This loop runs until we observe that $[\textcircled{I}\text{-}\textcircled{S}](t)$ is stationary (call this time t_{e}). Then for some additional time T we observe the fluctuations of $[\textcircled{I}\text{-}\textcircled{S}](t)$ around its mean. Tomé and De Oliveira [67] derived the effective infection rate $\tilde{\beta}$ for each chosen count $[\textcircled{I}]$ of infected nodes by noting that, because for every infection event there is a recovery event,

$$\tilde{\beta}\langle[\textcircled{I}\text{-}\textcircled{S}]\rangle = \gamma\,[\textcircled{I}]\,,$$

where $\langle\cdot\rangle$ is an average over many independent realizations, such that

$$\tilde{\beta} = \gamma\frac{[\textcircled{I}]}{\langle[\textcircled{I}\text{-}\textcircled{S}]\rangle},$$

which they found to lead to the same nontrivial steady states $[\textcircled{I}]^*(\tilde{\beta}/\gamma)$ as $[\textcircled{I}]^*(\beta/\gamma)$ in the model without control. As the model with control is ergodic (no absorbing states exist) we need to run only a single realization and compute the effective infection rate as

$$\tilde{\beta} = \gamma\frac{[\textcircled{I}]}{\int_{t_{\mathrm{e}}}^{t_{\mathrm{e}}+T}[\textcircled{I}\text{-}\textcircled{S}]\,(t)dt/T}. \tag{C3}$$

In this manner, the error bars of estimates of $\tilde{\beta}$ can be made arbitrarily small by increasing T.

APPENDIX D: CORRELATIONS AT GIVEN DISTANCE

In Fig. 7 we show for SIS epidemic spreading the correlation at given distance between two nodes (only between infected nodes shown). Its general definition is

$$C_{ab}^{D} = \frac{N^2}{\sum_{ij}\mathbb{1}_{\mathrm{dist}(i,j)=D}}\frac{\sum_{ij}[\textcircled{a}_i\textcircled{b}_j]_{\mathrm{dist}(i,j)=D}}{[\textcircled{a}][\textcircled{b}]}, \tag{D1}$$

or with normalized motifs,

$$C_{ab}^{D} = \frac{\sum_{ij}[\![\textcircled{a}_i\textcircled{b}_j]\!]_{\mathrm{dist}(i,j)=D}}{[\![\textcircled{a}]\!][\![\textcircled{b}]\!]}, \tag{D2}$$

where the distance dist(i, j) is the length of the shortest path between i and j. Note that the correlations between neighboring nodes (39) is a special case of this, i.e., $C_{ab} = C_{ab}^{1}$. An alternative way to write Eq. (D2) is

$$C_{ab}^{D} = \frac{\sum_{c_1,\ldots,c_{D-1}}[\![ac_1\ldots c_{D-1}b]\!]}{[\![\textcircled{a}]\!][\![\textcircled{b}]\!]}, \tag{D3}$$

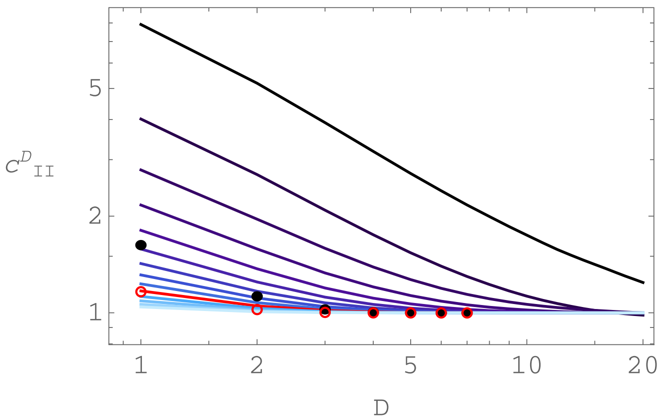


FIG. 7. Steady-state correlation function C_{II}^D versus distance D (D1) for simulations of SIS spreading on the square lattice (lines). Color scale indicates values of β/γ corresponding to the markers in Fig. 2 and sorted such that higher correlations occur for lower values of β/γ (closer to the critical point). Markers show approximated correlations when assuming MF2 conditions hold [using Eq. (D3)] and factorizing the numerator in pair fractions). Filled black markers are MF2 correlations for $\beta/\gamma = 0.4137$ (which for simulations is the top black line). Open red markers are MF2 correlations for $\beta/\gamma = 0.5602$ (which for simulations is the red line).

where $[\![ac_1...c_{D-1}b]\!]$ is a chain motif of size $D+1$ with indicated states. This definition was used to derive C_{ab}^D as approximated by mean-field models in Fig. 7, by applying the closure formula to the chain in the numerator.

APPENDIX E: STEADY STATES OF MF1-MF2

Here, we show the expressions for the steady states and steady-state correlations of MF1 and MF2.

1. MF1

The MF1 model equals SIS epidemic spreading under well-mixed conditions. Its steady states are the trivial and the endemic state,

$$[\![\text{I}]\!]_1^* = 0, \quad [\![\text{I}]\!]_2^* = 1 - \frac{\gamma}{\kappa\beta}.$$

2. MF2

For the MF2 model the equations and, hence, their steady states, depend on the type of network.

a. Degree-homogeneous networks

The steady states of the dynamic equations for MF2 on degree-homogeneous networks (49) are

$$([\![\text{I}]\!]^*, [\![\text{I–S}]\!]^*)_1 = (0,0), \quad ([\![\text{I}]\!]^*, [\![\text{I–S}]\!]^*)_2 = \left(1 - \frac{\kappa-1}{\kappa\beta(\kappa-1)/\gamma - 1}, \frac{\beta(\kappa-1)-1}{\beta(\beta\kappa(\kappa-1)-1)}\right). \tag{E1}$$

The nontrivial correlations, obtained by substitution of the above into Eq. (40), are

$$C_{II}^* = \frac{(\gamma-\beta\kappa)(\gamma-\beta(\kappa-1)\kappa)}{\beta\kappa^2(\beta(\kappa-1)-\gamma)}, \quad C_{SI}^* = 1 - \frac{\gamma}{\beta\kappa(\kappa-1)}, \quad C_{SS}^* = \frac{\beta(\kappa-1)\kappa-\gamma}{\beta(\kappa-1)^2}. \tag{E2}$$

b. Degree-heterogeneous networks

The steady states of the dynamic equations for MF2 on degree-heterogeneous networks (53) are

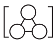
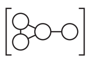
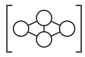
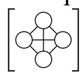
$$\begin{aligned}
&([\![\text{I}]\!]^*, [\![\text{I–I}]\!]^*, [\![\text{I–S}]\!]^*)_1 = (0,0,0),\\
&([\![\text{I}]\!]^*, [\![\text{I–I}]\!]^*, [\![\text{I–S}]\!]^*)_2 = \left(\frac{1}{2}\left(\kappa+1-\sqrt{(\kappa-1)^2+4\gamma/\beta}\right), \frac{\gamma}{2\beta}\left(\sqrt{(\kappa-1)^2+4\gamma/\beta}-\kappa-3\right)+\kappa,\right.\\
&\qquad\left.\frac{\gamma}{2\beta}\left(\kappa+1-\sqrt{(\kappa-1)^2+4\gamma/\beta}\right)\right).
\end{aligned} \tag{E3}$$

The nontrivial correlations, via substitution of the above into Eq. (40), are

$$C_{II}^* = \frac{4\beta^{3/2}\kappa - 2\sqrt{\beta}\gamma(\kappa+3) + 2\gamma\sqrt{\beta(\kappa-1)^2+4\gamma}}{\sqrt{\beta}\kappa[\sqrt{\beta(\kappa-1)^2+4\gamma} - \sqrt{\beta}(\kappa+1)]^2}, \tag{E4}$$

$$C_{SI}^* = C_{SS}^* = \frac{2\gamma}{\kappa\sqrt{\beta^2(\kappa-1)^2+4\beta\gamma} - \beta(\kappa-1)\kappa}. \tag{E5}$$

APPENDIX F: MF3 FOR SIS SPREADING

We only apply MF3 to the square lattice, such that we can ignore all motifs that contain triangles and set  =  =  =  $= 0$ in the remaining equations, such that we obtain

	$[\mathrm{I}]$	$[\mathrm{I{-}S}]$	$[\mathrm{I{-}I}]$	$[\mathrm{SIS}]$	$[\mathrm{SSI}]$	$[\mathrm{SII}]$	$[\mathrm{ISI}]$	$[\mathrm{III}]$	$[\mathrm{SSSI}]$	$[\mathrm{S}\langle\mathrm{SSI}\rangle]$	$[\diamond\,\mathrm{S;SS;I}]$	$[\mathrm{SISI}]$	$[\diamond\,\mathrm{I;SS;I}]$	$[\mathrm{ISSI}]$	$[\mathrm{S}\langle\mathrm{SII}\rangle]$	$[\diamond\,\mathrm{S;SI;I}]$	$[\mathrm{IISI}]$	$[\diamond\,\mathrm{I;SI;I}]$	$[\mathrm{S}\langle\mathrm{III}\rangle]$
$[\dot{\mathrm{I}}]$	$-\gamma$	β	0																
$[\dot{\mathrm{I}}{-}\mathrm{S}]$		$-\beta-\gamma$	γ	0	β	0	$-\beta$	0											
$[\dot{\mathrm{I}}{-}\mathrm{I}]$		2β	-2γ	0	0	0	2β	0											
$[\mathrm{S}\dot{\mathrm{I}}\mathrm{S}]$				$-2\beta-\gamma$	0	2γ	0	0	0	β	0	-2β	-2β	0	0	0	0	0	0
$[\mathrm{S}\dot{\mathrm{S}}\mathrm{I}]$				0	$-\beta-\gamma$	γ	γ	0	β	0	β	0	0	$-\beta$	$-\beta$	$-\beta$	0	0	0
$[\mathrm{S}\dot{\mathrm{I}}\mathrm{I}]$				β	β	$-\beta-2\gamma$	0	γ	0	0	0	β	β	0	β	0	$-\beta$	$-\beta$	0
$[\mathrm{I}\dot{\mathrm{S}}\mathrm{I}]$				0	0	0	$-2\beta-2\gamma$	γ	0	0	0	0	0	2β	0	2β	0	0	$-\beta$
$[\mathrm{I}\dot{\mathrm{I}}\mathrm{I}]$				0	0	2β	2β	-3γ	0	0	0	0	0	0	0	0	2β	2β	β

The seven conservation relations are

$$
\begin{aligned}
[\mathrm{I}] + [\mathrm{S}] &= N,\\
[\mathrm{I{-}S}] + [\mathrm{S{-}S}] &= \kappa[\mathrm{S}],\\
[\mathrm{I{-}I}] + [\mathrm{I{-}S}] &= \kappa[\mathrm{I}],\\
[\mathrm{SSI}] + [\mathrm{SSS}] &= (\kappa-1)[\mathrm{S{-}S}],\\
[\mathrm{SII}] + [\mathrm{SIS}] &= (\kappa-1)[\mathrm{I{-}S}],\\
[\mathrm{ISI}] + [\mathrm{SSI}] &= (\kappa-1)[\mathrm{I{-}S}],\\
[\mathrm{III}] + [\mathrm{SII}] &= (\kappa-1)[\mathrm{I{-}I}].
\end{aligned}
$$

We eliminate $[\mathrm{I{-}I}]$, $[\mathrm{SIS}]$, $[\mathrm{SSI}]$, $[\mathrm{SII}]$. Following Eqs. (11)–(13), this means

$$
\mathbf{x} = \begin{bmatrix} [\mathrm{I}] \\ [\mathrm{I{-}S}] \\ [\mathrm{I{-}I}] \\ [\mathrm{SIS}] \\ [\mathrm{SSI}] \\ [\mathrm{SII}] \\ [\mathrm{ISI}] \\ [\mathrm{III}] \end{bmatrix}, \quad
\tilde{\mathbf{x}} = \begin{bmatrix} [\mathrm{I}] \\ [\mathrm{I{-}S}] \\ [\mathrm{ISI}] \\ [\mathrm{III}] \end{bmatrix}, \quad
\tilde{\mathbf{x}}_4 = \begin{bmatrix} [\mathrm{ISSI}] \\ [\diamond\,\mathrm{S;SI;I}] \\ [\mathrm{IISI}] \\ [\diamond\,\mathrm{I;SI;I}] \\ [\mathrm{S}\langle\mathrm{III}\rangle] \end{bmatrix},
$$

and

$$
\tilde{\mathbf{Q}}_{1\cdots3,1\cdots3} = \begin{pmatrix}
-\gamma & \beta & 0 & 0 & 0 & 0 & 0 & 0\\
0 & -\beta-\gamma & \gamma & 0 & \beta & 0 & -\beta & 0\\
0 & 2\beta & -2\gamma & 0 & 0 & 0 & 2\beta & 0\\
0 & 0 & 0 & -2\beta-\gamma & 0 & 2\gamma & 0 & 0\\
0 & 0 & 0 & 0 & -\beta-\gamma & \gamma & \gamma & 0\\
0 & 0 & 0 & \beta & \beta & -\beta-2\gamma & 0 & \gamma\\
0 & 0 & 0 & 0 & 0 & 0 & -2\beta-2\gamma & \gamma\\
0 & 0 & 0 & 0 & 0 & 2\beta & 2\beta & -3\gamma
\end{pmatrix}, \quad
\mathbf{E} = \begin{pmatrix}
1 & 0 & 0 & 0\\
0 & 1 & 0 & 0\\
\kappa & -1 & 0 & 0\\
-\kappa(\kappa-1) & 2(\kappa-1) & 0 & 1\\
0 & \kappa-1 & -1 & 0\\
\kappa(\kappa-1) & 1-\kappa & 0 & -1\\
0 & 0 & 1 & 0\\
0 & 0 & 0 & 1
\end{pmatrix},
$$

$$
\tilde{\mathbf{Q}}_{34} = \begin{pmatrix} 2\beta & 2\beta & 0 & 0 & -\beta\\ 0 & 0 & 2\beta & 2\beta & \beta \end{pmatrix}, \qquad \mathbf{c} = (0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0)^T,
$$

which results in the four remaining equations ($\tilde{\mathbf{c}} = \mathbf{0}$ so not shown),

	$[\mathrm{I}]$	$[\mathrm{I{-}S}]$	$[\mathrm{ISI}]$	$[\mathrm{III}]$	$[\mathrm{ISSI}]$	$[\diamond\,\mathrm{S;SI;I}]$	$[\mathrm{IISI}]$	$[\diamond\,\mathrm{I;SI;I}]$	$[\mathrm{S}\langle\mathrm{III}\rangle]$
$[\dot{\mathrm{I}}]$	$-\gamma$	β							
$[\dot{\mathrm{I}}{-}\mathrm{S}]$	4γ	$2\beta-2\gamma$	-2β	0					
$[\mathrm{I}\dot{\mathrm{S}}\mathrm{I}]$			$-2\beta-2\gamma$	γ	2β	2β	0	0	$-\beta$
$[\mathrm{I}\dot{\mathrm{I}}\mathrm{I}]$	24β	-6β	2β	$-3\gamma-2\beta$	0	0	2β	2β	β

(F1)

where we have used $\kappa = 4$ for the square lattice. In normalized form this is

$$
\begin{array}{c|c|c|c|c|c|c|c|c|c|c}
 & [\![\mathrm{I}]\!] & [\![\mathrm{IS}]\!] & [\![\mathrm{ISI}]\!] & [\![\mathrm{III}]\!] & [\![\mathrm{ISSI}]\!] & [\![\square\,\mathrm{SIIS}]\!] & [\![\mathrm{IISI}]\!] & [\![\square\,\mathrm{IIIS}]\!] & [\![\mathrm{S}\{\mathrm{III}\}]\!] \\
\hline
\dot{[\![\mathrm{I}]\!]} & -\gamma & 4\beta & & & & & & & \\
\hline
\dot{[\![\mathrm{IS}]\!]} & \gamma & 2\beta-2\gamma & -6\beta & 0 & & & & & \\
\hline
\dot{[\![\mathrm{ISI}]\!]} & & & -2\beta-2\gamma & \gamma & \frac{14}{3}\beta & \frac{4}{3}\beta & 0 & 0 & -2\beta \\
\dot{[\![\mathrm{III}]\!]} & 2\beta & -2\beta & 2\beta & -3\gamma-2\beta & 0 & 0 & \frac{14}{3}\beta & \frac{4}{3}\beta & 2\beta
\end{array}. \tag{F2}
$$

To close the system of equations, we apply Eq. (23) to the chains and star, and Eq. (24) to the cycles (using the extension to nonmaximal 2-cliques) and obtain the normalized closures (see also Sec. VII D, examples 3–5)

$$
\begin{aligned}
[\![\mathrm{ISSI}]\!] &\approx \frac{[\![\mathrm{SSI}]\!]^2}{[\![\mathrm{SS}]\!]} = \frac{([\![\mathrm{IS}]\!]-[\![\mathrm{ISI}]\!])^2}{(1-[\![\mathrm{I}]\!]-[\![\mathrm{IS}]\!])},\\
[\![\mathrm{IISI}]\!] &\approx \frac{[\![\mathrm{SII}]\!][\![\mathrm{ISI}]\!]}{[\![\mathrm{IS}]\!]} = \frac{([\![\mathrm{I}]\!]-[\![\mathrm{III}]\!]-[\![\mathrm{IS}]\!])[\![\mathrm{ISI}]\!]}{[\![\mathrm{IS}]\!]},\\
[\![\mathrm{S}\{\mathrm{III}\}]\!] &\approx \frac{[\![\mathrm{IS}]\!]^3}{[\![\mathrm{S}]\!]^2} = \frac{[\![\mathrm{IS}]\!]^3}{(1-[\![\mathrm{I}]\!])^2},\\
[\![\square\,\mathrm{SIIS}]\!] &\approx \frac{[\![\mathrm{SII}]\!]^2[\![\mathrm{SSI}]\!]^2}{[\![\mathrm{IS}]\!]^2[\![\mathrm{SS}]\!][\![\mathrm{II}]\!]} = \frac{([\![\mathrm{I}]\!]-[\![\mathrm{III}]\!]-[\![\mathrm{IS}]\!])^2([\![\mathrm{IS}]\!]-[\![\mathrm{ISI}]\!])^2}{(1-[\![\mathrm{I}]\!]-[\![\mathrm{IS}]\!])([\![\mathrm{I}]\!]-[\![\mathrm{IS}]\!])[\![\mathrm{IS}]\!]^2},\\
[\![\square\,\mathrm{IIIS}]\!] &\approx \frac{[\![\mathrm{SII}]\!]^2[\![\mathrm{III}]\!][\![\mathrm{ISI}]\!]}{[\![\mathrm{IS}]\!]^2[\![\mathrm{II}]\!]^2} = \frac{[\![\mathrm{III}]\!]([\![\mathrm{I}]\!]-[\![\mathrm{III}]\!]-[\![\mathrm{IS}]\!])^2[\![\mathrm{ISI}]\!]}{([\![\mathrm{I}]\!]-[\![\mathrm{IS}]\!])^2[\![\mathrm{IS}]\!]^2},
\end{aligned} \tag{F3}
$$

where we also used the conservation relations to substitute any previously eliminated motif. The steady-state solutions of the final system are roots of a ninth-order polynomial, of which two are admissible (see Fig. 2 for numerical results).

APPENDIX G: MF4 FOR SIS SPREADING IN MATHEMATICA

We uploaded two application examples of our algorithm at fourth order to Mathematica's notebook archive (best viewed locally). In the first section of the file, the procedure as explained in the main text is followed for MF4 on the square lattice. The second section in the file shows the general coefficient matrix **Q** for the (unclosed) moment equations up to fourth order, with columns labeled by $\mathbf{x}_1, \dots, \mathbf{x}_5$ (stored in the variable `mots`). The algorithm can be instructed to write the resulting set of equations as `MATLAB` [68] functions in a format compatible with continuation software such as `COCO` [69] and `MatCont` [70].

APPENDIX H: CLOSURE EXAMPLES

Here we show 13 examples of subgraph decompositions based on the method explained in Sec. VII [Eqs. (23) and (24)]. They are shown in table form on the last page of the arxiv version of this article [71]. We also discussed three commonly used ones (1–3 in the table) in Sec. VII D.

As the closures can be written independent of the particular labels, they are shown in the table for subgraphs only, with each node tagged with its index. We have also dropped the $\langle\cdot\rangle$, assuming that the law of large numbers applies, such that the counts approach their expectations almost surely for increasing network size N. The examples can be understood by reading the table from left to right. Comments are added in the last column. By column, it shows: (1) the example number, (2) the considered subgraph, (3) its diameter, (4) its independence map assuming independence beyond distance diam($\mathsf{a}-1$), (5) whether the independence map is chordal, (6) the derived junction graph of d-cliques where $d = \mathrm{diam}(\mathsf{a}-1)$, (7) the resulting closure formula, (8) a triangulation of the independence map if the independence map is nonchordal, (9) the junction graph of d-cliques based on the triangulation, (10) the closure formula based on the triangulation, (11) an ad hoc extension of the method to nonmaximal cliques for some subgraphs, (12) the resulting closure formula based on this extension, (13) comments. Our `Mathematica` script [36] generates these closures automatically.

[1] C. Kuehn, *Control of Self-Organizing Nonlinear Systems* (Springer, New York, 2016), pp. 253–271.
[2] K. J. Sharkey, I. Z. Kiss, R. R. Wilkinson, and P. L. Simon, Bull. Math. Biol. **77**, 614 (2015).
[3] K. J. Sharkey and R. R. Wilkinson, Math. Biosci. **264**, 74 (2015).
[4] K. J. Sharkey, Theor. Popul. Biol. **79**, 115 (2011).
[5] T. Rogers, J. Stat. Mech.: Theory Exp. (2011) P05007.
[6] T. Gross and I. G. Kevrekidis, Europhys. Lett. **82**, 38004 (2008).
[7] V. Isham, Math. Biosci. **107**, 209 (1991).
[8] P. Weiss, J. Phys. Theor. Appl. **6**, 661 (1907).
[9] W. L. Bragg and E. J. Williams, Proc. R. Soc. London A **145**, 699 (1934).
[10] H. A. Bethe, Proc. R. Soc. London A **150**, 552 (1935).
[11] R. Kikuchi, Phys. Rev. **81**, 988 (1951).
[12] J. Marro and R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models* (Cambridge University Press, Cambridge, UK, 1999)
[13] M. Henkel, H. Hinrichsen, and S. Lübeck, *Non-equilibrium Phase Transitions*, Theoretical and Mathematical Physics (Springer Netherlands, Dordrecht, 2008).

[14] T. Tomé and M. J. De Oliveira, *Stochastic Dynamics and Irreversibility* (Springer, Berlin, 2015), p. 394.
[15] H. Matsuda, N. Ogita, A. Sasaki, and K. Sato, Prog. Theor. Phys. **88**, 1035 (1992).
[16] M. J. Keeling, D. A. Rand, and A. J. Morris, Proc. R. Soc. B: Biol. Sci. **264**, 1149 (1997).
[17] D. A. Rand, Advanced Ecological Theory, Tech. Rep. (1999).
[18] U. Dieckmann, R. Law, and J. A. J. Metz, *The Geometry of Ecological Interactions: Simplifying Spatial Complexity* (Cambridge University Press, Cambridge, UK, 2000), p. 564.
[19] S. Kéfi, M. Rietkerk, M. van Baalen, and M. Loreau, Theor. Popul. Biol. **71**, 367 (2007).
[20] T. Gross, Carlos J. D'Lima, and B. Blasius, Phys. Rev. Lett. **96**, 208701 (2006).
[21] T. House, G. Davies, L. Danon, and M. J. Keeling, Bull. Math. Biol. **71**, 1693 (2009).
[22] K. J. Sharkey, J. Math. Biol. **57**, 311 (2008).
[23] G. A. Böhme and T. Gross, Phys. Rev. E **83**, 035101(R) (2011).
[24] J. P. Gleeson, Phys. Rev. X **3**, 021004 (2013).
[25] P. G. Fennell and J. P. Gleeson, SIAM Rev. **61**, 92 (2019).
[26] V. Marceau, P.-A. Noël, L. Hébert-Dufresne, A. Allard, and L. J. Dubé, Phys. Rev. E **82**, 036116 (2010).
[27] G. St-Onge, V. Thibeault, A. Allard, L. J. Dubé, and L. Hébert-Dufresne, Phys. Rev. E **103**, 032301 (2021).
[28] K. Cui, W. R. Khudabukhsh, and H. Koeppl, Phys. Rev. E **105**, L042301 (2022).
[29] G. Demirel, F. Vazquez, G. A. Böhme, and T. Gross, Physica D **267**, 68 (2014).
[30] M. J. Keeling, Proc. R. Soc. B **266**, 859 (1999).
[31] I. Z. Kiss, J. C. Miller, and P. L. Simon, *Interdisciplinary Applied Mathematics*, Interdisciplinary Applied Mathematics, Vol. 46 (Springer International Publishing, Cham, 2017), pp. 1–413.
[32] G. Szabó and C. Hauert, Phys. Rev. Lett. **89**, 118101 (2002).
[33] V. Danos, T. Heindel, R. Honorato-Zimmer, and S. Stucki, in *Computational Methods in Systems Biology*, Lecture Notes in Computer Science, Vol. No. 12314 LNBI (Springer Science and Business Media Deutschland GmbH, 2020), pp. 3–26.
[34] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Elsevier, Amsterdam, 1988).
[35] Wolfram Research, Inc., Mathematica, Version 13.0.0 (2022), Champaign, IL, https://www.wolfram.com/mathematica.
[36] B. Wuyts, Mathematica package for mean-field models via moment closure (2022), https://github.com/b-wuyts/mfmcl.
[37] F. Harary and E. M. Palmer, *Graphical Enumeration* (Elsevier, Amsterdam, 1973).
[38] D. Barber, *Bayesian Reasoning and Machine Learning* (Cambridge University Press, Cambridge, UK, 2012).
[39] More precisely, they consist of the nodes of the maximal cliques of $\mathrm{tr}(\mathcal{G}^d)$ and the links of $\mathcal{G}$.
[40] Y. Bulatov, Chordal Graph Package (2011), https://mathematica-bits.blogspot.com/2011/02/chordal-graph-usage.html.
[41] N. Peyrard, U. Dieckmann, and A. Franc, Theor. Popul. Biol. **73**, 383 (2008).
[42] S. Horvát, IGraph/M—The igraph interface for Mathematica, version 0.5.1 (2021), http://szhorvat.net/pelican/igraphm-a-mathematica-interface-for-igraph.html.
[43] For an ER random network, the expected number of m-node cycles and m-node chains (counting all ordered m-tuples without repetition) are $L(N) = \frac{N!}{(N-m)!}(\frac{\kappa}{N-1})^m$ and $C(N) = \frac{N!}{(N-m)!}(1-\frac{\kappa}{N-1})(\frac{\kappa}{N-1})^{m-1}$, where $\frac{\kappa}{N-1}$ is the probability of having a link between two given nodes. Hence, $\lim_{N\to\infty}L(N) = \kappa^m = O(1)$ and $\lim_{N\to\infty}C(N) = \lim_{N\to\infty}\kappa^{m-1}N = O(N)$. Hence, as $L(N)$ remains finite and $C(N)$ grows with N, cycles can be ignored in the limit of large N. In the case $m = 3$ (and N large), we can hence safely assume that all triples are chains.
[44] T. E. Harris, Ann. Probabil. **2**, 969 (1974).
[45] D. T. Gillespie, Annu. Rev. Phys. Chem. **58**, 35 (2007).
[46] J. Sieber, A. Gonzalez-Buelga, S. A. Neild, D. J. Wagg, and B. Krauskopf, Phys. Rev. Lett. **100**, 244101 (2008).
[47] F. Schilder, E. Bureau, I. F. Santos, J. J. Thomsen, and J. Starke, J. Sound Vib. **358**, 251 (2015).
[48] D. Barton, Mech. Syst. Signal Process. **84**, 54 (2017).
[49] M. A. Porter and J. P. Gleeson, *Dynamical Systems on Networks*, Frontiers in Applied Dynamical Systems: Reviews and Tutorials (Springer, 2016), Vol. 4.
[50] M. Newman, *Networks* (Oxford University Press, Oxford, UK, 2018).
[51] T. M. Liggett, *Interacting Particle Systems*, Classics in Mathematics (Springer, Berlin, 2005).
[52] P. Erdős and S. J. Taylor, Acta Mathematica Academiae Scientiarum Hungaricae, Tech. Rep. 3-4 (1960).
[53] M. Heydenreich and R. van der Hofstad, *Progress in High-Dimensional Percolation and Random Graphs*, CRM Short Courses (Springer International Publishing, Cham, 2017), p. 285.
[54] G. F. Lawler and V. Limic, Random Walk: A Modern Introduction, Tech. Rep. (2010).
[55] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, Phys. Rep. **874**, 1 (2020).
[56] B. Karrer and M. E. J. Newman, Phys. Rev. E **82**, 016101 (2010).
[57] R. R. Wilkinson and K. J. Sharkey, Phys. Rev. E **89**, 022808 (2014).
[58] A. Koher, H. H. K. Lentz, J. P. Gleeson, and P. Hövel, Phys. Rev. X **9**, 031017 (2019).
[59] B. Wuyts and J. Sieber, Emergence of nonlinear dynamics from spatial structure in tropical forest-grassland landscapes, arXiv:2207.14118 (2022).
[60] Y. Kevrekidis and G. Samaey, Annu. Rev. Phys. Chem. **60**, 321 (2009).
[61] D. G. Patsatzis, L. Russo, I. G. Kevrekidis, and C. Siettos, Data-driven control of agent-based models: An equation/variable-free machine learning approach, arXiv:2207.05779 (2022).
[62] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, Springer Series in Synergetics (Springer, Berlin, 2009).
[63] Equation (C1) is the simplest possible form of feedback control as it changes only one input (r) depending on one output ($[\boldsymbol{x}^{\mathrm{a}}]$) and is static (no further processing of $[\boldsymbol{x}^{\mathrm{a}}]$ before feeding it back). Linear control theory ensures that unstable equilibria of ODEs can be stabilized with single-input-single-output dynamic feedback control using any single input and any single output satisfying some genericity conditions (linear controllability and observability).
[64] D. A. W. Barton and J. Sieber, Phys. Rev. E **87**, 052916 (2013).

[65] J. Sieber, O. E. Omel'chenko, and M. Wolfrum, Phys. Rev. Lett. **112**, 054102 (2014).

[66] L. Renson, D. Barton, and S. Neild, Int. J. Bifurcat. Chaos **27**, 1730002 (2017).

[67] T. Tomé and M. J. de Oliveira, Phys. Rev. Lett. **86**, 5643 (2001).

[68] MathWorks, MATLAB 2021a (2021), http://uk.mathworks.com.

[69] F. Schilder, H. Dankowicz, and M. Li, Continuation core and toolboxes (coco) (2021), https://sourceforge.net/projects/cocotools.

[70] A. Dhooge, W. Govaerts, and Y. A. Kuznetsov, ACM Trans. Math. Softw. **29**, 141 (2003).

[71] B. Wuyts and J. Sieber, Mean-field models of dynamics on networks via moment closure: an automated procedure, arXiv:2111.07643 (2022).