Generalised moment closure for discrete-state dynamics on networks

Bert Wuyts and Jan Sieber

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

(Dated: Received: date / Accepted: date)

Abstract

Moment closure is a method originating from statistical physics for finding low-dimensional mean-field models of dynamics on networks. It proceeds by deriving moment equations, expressions of the evolution of mean counts of subgraph states. Each moment equation for a given subgraph depends on mean counts of larger subgraphs, for which new equations need to be derived. To avoid an infinite system of equations, one relies on a closure scheme, a substitution of counts of the largest subgraphs by an expression involving counts of smaller ones, which is justified by an independence assumption. This approach leads to an exact description with few equations if the independence assumption is valid for sufficiently small subgraphs, such as in case of epidemic dynamics on trees without reinfection, but the mean-field models obtained in this way still approximate the dynamics well on networks with few short loops. A higher prevalence of loops requires increasing the order of the moment closure. Yet, no systematic method exists to obtain higher-order moment equations and the closure scheme required to truncate them. In this paper, we present a general method to obtain and truncate moment equations, applicable to any model of dynamics on networks with at most interactions between nearest neighbours and for arbitrary approximation order. We first obtain the moment equations in their general form, via a derivation from the master equation. Then, we show that closure schemes can be systematically obtained via a decomposition of the largest subgraphs into their smaller-diameter components, and, that this decomposition is exact when these components form a tree and there is independence at distances beyond their graph diameter, offering a theoretical justification for moment closure on non-tree networks. Applying our method to the SIS epidemic model on lattices and random networks, we find that the well-known long-range correlation near the epidemic threshold due to a continuous phase transition only leads to considerable bias in lower-order moment closures for low-dimensional lattices, because here, presence of loops of all sizes prevent decomposition of larger-distance correlations in terms of smaller-distance ones, unlike in random networks. Our method extends the practical applicability of moment closure to networks in which clustering due to a high density of short loops is particularly important. A Mathematica script that automates the moment closure is made available for download.

Keywords: moment closure, networks, graph theory, dynamics on networks, nonlinear dynamics, master equation, Markov networks, epidemic models, phase transitions, feedback control
## CONTENTS

I. Introduction 3

II. Main definitions and general overview 5

III. Derivation of the moment equations 9

IV. Closure scheme 15
   A. Definitions and background 15
   B. Motif decomposition 18
   C. Examples 20

V. Overview of steps to obtain MFk 23

VI. Application to the SIS model 23
   A. Definitions and implementation of the simulations 23
   B. Mean field equations 26
      1. MF1 27
      2. MF2 28
      3. MF3 30
      4. MF4 and MF5 32
   C. Comparison to simulations 32

VII. Conclusions 36

Acknowledgments 37

References 37

A. Supplementary figures and table 40

B. Closure examples 41

C. Differential invariants for the SIS model up to order 3 42

D. Higher order mean-field equations for the SIS model 43
I. INTRODUCTION

Mathematical analysis of complex systems usually starts from high-dimensional models. While such models may have more realism, they are harder to analyse than simpler low-dimensional models. Therefore, one often looks for a low-dimensional description via a coarse-graining method. The earliest coarse-graining methods were developed in statistical physics to study phase transitions in lattice models of materials [1–4] and are now known as mean-field approximations. The essence of mean-field approximations is that one can obtain a simpler description by neglecting some correlations in the system, via a replacement of an average (sum) of products with a product of averages (sums). While this substitution was initially only applied to terms in the energy function of equilibrium lattice models [1–4], it eventually came to be applied as well to occurrence probabilities in kinetic equations of non-equilibrium lattice models derived from the master equation [5–7]. Its domain of application was extended from materials to spatial population dynamics in ecology and evolution [8–10], and to epidemic dynamics on networks [11, 12]. In this generalised context, the method of deriving and truncating kinetic equations for dynamics of chemical or biological species came to be known as moment closure [13].

For discrete-state dynamics on networks [14–16], moment closure consists of first deriving a system of ordinary differential equations (ODEs), also known as moment equations, representing the evolution of the expectation of small subgraph states known as network motifs. The derivation proceeds from smaller to larger sized-motifs, with dynamics of mean motif frequencies of size \( m \) depending only on mean motif frequencies of size \( m \) and \( m + 1 \) if the dynamics has only nearest-neighbour interactions. Hence, a system of ODEs obtained in such a manner for motif frequencies up to a maximum considered size \( k \) (also referred to as the order of the moment closure) is always underdetermined, because it does not contain equations for motifs of size \( k + 1 \). Therefore, as the second step of moment closure, a closure approximation is applied – as in the mean-field approximation mentioned above – by formulating frequencies of \((k + 1)\)-size motifs as functions of frequencies of \(\{1, \ldots, k\}\)-size motifs, allowing the system of ODEs to be closed. The underlying assumption that may justify such a substitution is that mean frequencies or probabilities of larger-sized motifs factorise in terms of smaller-sized ones when the latter are (conditionally) independent. The validity of the independence assumption has been proven only for the second-order individual-level equations of the SIR epidemic model on trees when conditioning on susceptible nodes [17–19]. 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.
in the whole network) [20], which then permits a compact description in terms of population-level quantities. Despite these strong constraints, moment closure has been found to be robust to mild violations of these assumptions, showing little bias for networks that are only locally treelike, that have some heterogeneity and for dynamics that includes reinfection [15, 21]. Furthermore, when the effect of short loops is taken into account by including network motifs that contain them, the approximation can be steadily improved [18, 22, 23].

One drawback of moment closure is that it may quickly become impractical due to a combinatorial increase of the number of motif types, and hence equations, 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 for it to be also sufficiently representative of the original high-dimensional model. Another drawback of moment closure is that the independence assumptions underlying commonly used closure formulas are either justified only heuristically or for specific cases, and that there is hence no systematic method to derive them. Therefore, the underlying assumptions are often hard to state in general and considerable effort is often invested in testing the validity of closures [19]. As a consequence, focus has recently shifted to finding methods that are more efficient – at least for some important cases – and that rely on more clearly specified simplifying assumptions. For instance, approximate master equation methods better accommodate degree-heterogeneous networks by considering for every state and degree the number of connections to each state, assuming absence of higher-order structure [24–26], with later work relaxing this assumption [e.g. 27]. Our moment closure method presented here addresses the above-mentioned drawbacks by providing a general automated derivation that allows a larger number of equations to be considered than typically derived by hand and that includes a systematic procedure to derive closures starting from specified independence assumptions. The derivation consists of three steps: i. derivation of the ODEs for motifs of size $\{1, \ldots, k\}$ – later called ‘differential invariants’, ii. elimination of some equations from the ODEs via additional relations – later called ‘algebraic invariants’, iii. closure of the system of ODEs. We will refer to the final closed system of ODEs as the $k$-th order mean field, or MF$k$ in short.

A general overview with main concepts is given in Section II. The derivation (step i) uses a general formula (4) that we derive from the master equation in Section III. The expressions used for elimination of variables from the ODEs (step ii) follow from conservation of sums of variables and are shown in equations (19-20) of Section III. Finally, general relations to close the system of ODEs (step iii) are derived in Section IV – see (30) or (31) – by assuming conditional independence of states in the network beyond a distance $d$, which permits a factorisation in terms of its $d$-cliques.
if they form a tree. This follows from a generalisation of decomposition of Markov networks [28, 29]. An overview of the resulting procedure is given in Section V. In Section VI we set up MF1-5 models of the SIS epidemic model and compare their steady states to those of simulations stabilised by feedback control on a selection of networks. We focus in particular on the square lattice, for which low-order moment closures and approximate master equation methods fail, due to its large number of loops of any size, and we compare against random networks (and higher-dimensional lattices), for which the former methods work well.

We developed a Mathematica [30] package that derives MFk by performing steps (i-iii). It writes the resulting set of MATLAB [31] functions to an output file that can be used in continuation software such as COCO [32] and MatCont [33]. The required inputs for this algorithm are: i. the frequencies of all induced subgraphs in the considered network up to size k, ii. the matrices $R^0$, $R^1$ with conversion and interaction rates, and iii. the name for the output MATLAB file. The html output of an example is attached in Appendix D and the package will be made available for download.

II. MAIN DEFINITIONS AND GENERAL OVERVIEW

a. The underlying discrete-state continuous-time Markov chain  We consider a dynamical system on a fixed undirected graph $G$ with $N$ nodes and adjacency matrix $A \in \{0,1\}^{N \times N}$, 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 some $i \in \{1,\ldots,N\}$ is occupied by exactly one species in $\{1,\ldots,n\}$. A simple example is a square two-dimensional lattice of $\sqrt{N} \times \sqrt{N}$ nodes (say, periodic in both directions), where each node may have one of two states, susceptible or infected, in $\{\bullet, \circ\} = \{1,2\}$. The type of dynamics we consider is a continuous-time Markov chain with possible states $[X_i(t)]_{i=1}^N \in \{1,\ldots,n\}^N$ and two types of transitions, specified by a $n \times n$ matrix $R^0$ and a $n \times n \times n$ tensor $R^1$: i. spontaneous conversions: any node with state $a$ may change spontaneously into state $b$ with rate $R^0_{ab}$; ii. nearest-neighbour interactions: any node with state $a$ may change into state $b$ for each link to a node with state $c$, with rate $R^1_{abc}$. For the above example with states $X_i(t) \in \{\bullet, \circ\} = \{S,I\} = \{1,2\}$ for $t \in [0,\infty)$ and $i \in \{1,\ldots,N\}$, the matrix $R^0$ with a single non-zero entry $R^0_{2,1} = \gamma > 0$ (for spontaneous recovery) and the tensor $R^1$ with a single non-zero entry $R^1_{1,2,2} = \beta > 0$ (for infection along IS links, denoted by the symbol $\bullet$•$\circ$ below) describes a continuous-time SIS model on a square lattice with recoveries and infections occurring with rates $\gamma$ and $\beta$. 

5
b. Mean-field quantities and network motifs  Mean-field quantities for such Markov chains are total counts of induced subgraphs of \( \mathcal{G} \) with given state labels, where an induced subgraph of \( \mathcal{G} \) is a connected subset of nodes in \( \mathcal{G} \) and all the links between them. E.g., for our example, the count of \( I \) nodes or \( IS \) links are mean-field quantities. We use square brackets around a symbol to denote the count of occurrences of this symbol, i.e. the number of occurrences of \( I \) nodes is \([\circ]\). Hence, we can write e.g. respectively for \( I \) nodes, \( IS \) links and \( ISI \) chains

\[
[\circ] = \sum_{i=1}^{N} \delta_2(X_i), \quad [\circ \circ] = \sum_{i=1}^{N} \sum_{j \neq i} \delta_1(A_{ij})\delta_2(X_i)\delta_1(X_j), \\
[\circ \circ \circ] = \sum_{i=1}^{N} \sum_{j \neq i} \sum_{k \neq i, j} \delta_1(A_{ij})\delta_1(A_{jk})\delta_0(A_{ik})\delta_2(X_i)\delta_1(X_j)\delta_2(X_k), \\
\]

where we omitted the dependence on \( t \) to avoid clutter. The Kronecker delta, \( \delta_y(x) \), which is 1 if \( x \) equals \( y \) and 0 otherwise, helps count the instances where some quantity equals \( y \). By construction, the mean-field quantities on the left-hand side are random, since \( X(t) \) is random.

A network motif of size \( m \) generalizing the above examples, is a network with \( m \) nodes (only small \( m \) are typically considered), each of which are labelled with a state. It his hence fully characterised 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 \( a \in (\{1,\ldots,m\} \times \{1,\ldots,m\})^\mu \) of links between the \( m \) nodes (the indices of the non-zero entries in the adjacency matrix of the motif such that \( \mu \leq m(m-1)/2 \)), or a graphical representation of the connectivity. For instance, two linked nodes are displayed according to these representations as \((\circ,\circ)^{(1,2)}\), \((1,2)\), or \(\circ\cdots\circ\) 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 \( \mathbf{x} = (x_1,\ldots,x_m) \in \{1,\ldots,n\}^m \) with state labels \( x_p \) at positions \( p = 1,\ldots,m \). Hence, the pair of \( \mathbf{x} \) and \( a \) describes the motif, which we write as \( \mathbf{x}^a \). For example \((2,1)^{(1,2)} = (\circ,\circ)^{(1,2)} \) is the \( \circ \circ \) link, \((1,1,1)^{(1,2),(2,3),(1,3)} = (\circ,\circ,\circ)^{(1,2),(2,3),(1,3)} \) is the motif of nodes with states \( \circ \) connected in a triangle, later written as \( \circ - \circ - \circ \).

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

\[
[\mathbf{x}^a] = \sum_{i \in S(m,N)} \delta_a(A_i)\delta_\mathbf{x}(X_i), \quad (1)
\]

where \( S(m,N) \) is the set of all \( m \)-tuples from \( \{1,\ldots,N\} \) without repetition, which has size \(|S(m,N)| = N!/(N-m)!\). For instance, \( \sum_{i \in S(3,N)} = \sum_{i \in \{1,\ldots,N\}} \sum_{j \in \{1,\ldots,N\}\setminus i} \sum_{k \in \{1,\ldots,N\}\setminus \{i,j\}} \).
We use the convention that $A_i$ and $X_i$ are the restrictions of matrix $A$ and vector $X$ to the index set $i$. Counting via (1) leads to multiple counting of motifs with symmetries, with multiplicity equal to the number of symmetries. For instance, $\bullet \bullet$ is counted once; but $\bullet \bullet$ or $\bullet \bullet$ twice, and $\bullet \bullet \bullet$ six times (see Section III for more detail).

c. Differential invariants In our SIS network model example, the expected rate of change for the frequency of infected nodes is well known to satisfy

$$\frac{d}{dt} \langle [\bullet] \rangle = -\gamma \langle [\bullet] \rangle + \beta \langle [\bullet \bullet] \rangle,$$

(2)

where $\langle \cdot \rangle$ brackets denote expectations over many independent realisations of the underlying Markov chain. Relation (2) is exact for finite network sizes $N$ and can be derived from the Kolmogorov-forward (or master) equation for the Markov chain, as we will do in Section III. Note that the structure in (2) is such that the expected rate for the frequency of a motif of size $m = 1$ (here $\langle [\bullet] \rangle$) depends on the frequencies of motifs of size $m = 1$ (here $\langle [\bullet] \rangle$) from spontaneous conversions (here recovery) and size $m = 2$ (here $\langle [\bullet \bullet] \rangle$) from nearest-neighbour-induced conversions (here infection). This is true in general such that the mean-field quantity $[x^a](t)$ of size $m$ satisfies a differential invariance relation of form

$$\frac{d}{dt} \langle [x^a] \rangle = F_{x^a} \left( \langle [y_1^{b_1}] \rangle, \langle [y_2^{b_2}] \rangle, \ldots \right),$$

(3)

(using $|y|$ for the length of the vector $y$). On the right-hand side the $\langle [y_i^{b_i}] \rangle(t)$ stand for counts of motifs of size $m$ or $m+1$ on which the dynamics of $\langle [x^a] \rangle$ depend. In Section III we will show that the precise general expression for the differential equation (3) is

$$\frac{d}{dt} \langle [x^a] \rangle = \sum_{p=1}^{m} \sum_{k=1}^{n} \sum_{c=1}^{n} \left\{ \frac{R_{kx^a}}{n} + \kappa_{p,c}^a R_{kx^a}^1 \right\} \langle [k_{p,c}^a] \rangle - \left( \frac{R_{x^a}}{n} + \kappa_{p,c}^a R_{x^a}^1 \right) \langle [x^a] \rangle + \sum_{y^b \in \mathcal{N}_p^c(x^a)} R_{kx^a}^1 \langle [y^b] \rangle - \sum_{y^b \in \mathcal{N}_p^c(x^a)} R_{x^a}^1 \langle [y^b] \rangle \right\}. \tag{4}$$

Here, $k_{p,c}^a$ is the state label vector obtained by setting the state label of the $p$th element of $x$ to $k$. $\kappa_{p,c}^a$ is the $c$-degree at position $p$ in motif $x^a$, i.e. it is the number of connections node $p$ in motif $x^a$ has to nodes with state label $c$. The set $\mathcal{N}_p^c(x^a)$ is defined as containing all motifs $y^b$ of order $m+1$ such that $y$ extends the state label vector $x$ by one new node with state label $c$, and adjacency $b$ extends $a$ by the link between node $p$ and the new node (and, if present in $A$, additional links from the new node to other nodes of $a$). See (16) below for the formal definition of $\mathcal{N}_p^c$. The differential equation (4) shows how the expected count of $x^a$ is increased by transitions into $x^a$ – the positive terms – and decreased by transitions out of $x^a$ – the negative terms. This happens
through spontaneous conversions (terms with $R^0$), through nearest-neighbour interaction between nodes within the motif (first two terms with $R^1$), or through nearest-neighbour interaction with nodes outside the motif (last two terms with $R^1$). When we use (4) to express expected rates of change for the set of motifs up to a chosen maximum size $k$, we obtain a linear system of differential equations of the form

$$\frac{dx}{dt} = Ax + Cy,$$

where $x \in \mathbb{R}^K$ is a vector with all expected motif counts up to maximal size $k$, $y \in \mathbb{R}^L$ a vector with the expected counts of $(k + 1)$-order motifs on which the $k$-order motif counts in $x$ depend, $A$ is a matrix containing the coefficients in (4) for counts of motifs up to size $k$, and $C$ is a matrix containing the coefficients in (4) for counts of motifs of size $k + 1$. As $y \in \mathbb{R}^L$ with $L \geq 1$, (5) is an underdetermined system, with $L$ more unknowns than equations.

d. Algebraic invariants  Since our graph $G$ is fixed, for each adjacency matrix $a$ of possible motifs of size $m$ the algebraic invariance relation

$$\sum_{x:|x|=m} [x^a](t) = [a],$$

where $[a] = \sum_{i \in S(m,N)} \delta_a(A_i)$, (6)

holds. The overall count $[a]$ of induced subgraphs $a$ in graph $G$ is constant in time, and split up between all possible labellings. For our SIS network model example on a two-dimensional square lattice with periodic boundary conditions, the algebraic invariants for sizes $m = 1$ and $m = 2$ are the conservation of node counts and link counts:

$$[\bullet](t) + [\bullet](t) = N, \quad [\bullet\bullet](t) + [\bullet\bullet](t) + [\bullet\bullet](t) = \sum_{i} \sum_{j \neq i} \delta\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} A_{ij}\right) = \sum_{i} \sum_{j \neq i} A_{ij} = 4N.$$

The algebraic relations (6) can be used to reduce the number of variables in (5) via substitution. The right-hand side $[a]$ is also a natural normalisation factor for the motif counts $[x^a]$.

e. Moment closure  Section IV will discuss a general way (applicable at arbitrary order $k$) of expressing the undetermined parts $y \in \mathbb{R}^L$ in (5) through a function $f(x) \approx y$ [shown in (30, 31)], creating a closed system of ODEs for $x$ from (5). As (5) is often referred to as a system of moment equations, this method is commonly known as moment closure. We discuss in Section IV that, in addition to the assumption about independence of submotif occurrences, closure at the population level requires the further assumptions of spatial homogeneity and a large network size $N$. 

8
III. 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 (4) and (6). 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”) \( X := (X_1, X_2, ..., X_N) \in \{1, \ldots, n\}^N \), 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 [34]). The probability density \( P(X, t) \) for a particular state \( X \) at time \( t \) changes according to

\[
\frac{d}{dt} P(X, t) = \sum_{X' \neq X} \left[ w(X' \rightarrow X) P(X', t) - w(X \rightarrow X') P(X, t) \right],
\]

(7)

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

\[
\dot{P}(t) = WP(t),
\]

which describes the evolution of the density for all states as elements of a vector \( P \) [and not just of one particular state as in (7)]. 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, \( W \) 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 (7) as

\[
\frac{d}{dt} P(X, t) = \sum_{i=1}^{N} \sum_{k \neq X_i}^n \left[ w_{i}(k X \rightarrow X) P(k X, t) - w_{i}(X \rightarrow k X) P(X, t) \right],
\]

(8)

where

\[
X \rightarrow k_i X : (X_1, ..., X_i, ..., X_N) \mapsto (X_1, ..., k, ..., X_N)
\]

is the operator that replaces the species at the \( i \)th node by species \( k \), and \( w_i(\cdot) \) 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 $x^a$, as defined in Section II. Recall that these motifs, are defined by their state label vector $x$ of size $m$ and the adjacency structure between motif nodes $a$. We denote a single occurrence at a given $i \in S(m, N)$ as

$$[x^a_i](t) := \delta_a(A_i)\delta_x(X_i(t)),$$

which requires an exact match of the adjacency $a$ by $A_i$ and state label vector $x$ by $X_i$. For example, if $A$ contains a connected triangle between nodes 1, 2 and 3, then motifs with $a = \{(1, 2), (2, 3)\}$ would not occur on $i = (1, 2, 3)$. 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 $i \in S(m, N)$, as shown in (1). Since the large network structure is constant in time, we can use the counts of induced subgraphs $[a]$ given by (6) as a normalisation factor for the (variable) counts of motifs such that we may consider normalised motif frequencies

$$[x^a] := [x^a] / [a].$$

(10)

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

$$\langle [x^a_i] \rangle = \sum_X [x^a_i](X)P(X, t).$$

(11)

The master equation (8) for the density $P$ implies that the expectation satisfies the differential equation

$$\frac{d}{dt} \langle [x^a_i] \rangle = \sum_X [x^a_i](X)\frac{d}{dt}P(X, t)$$

$$= \sum_X \left\{ \delta_a(A_i)\delta_x(X_i) \sum_{i'} \sum_{k \neq X_{i'}}^N \left[ w_{i'}(k_iX \rightarrow X)P(k_iX, t) - w_{i'}(X \rightarrow k_iX)P(X, t) \right] \right\},$$

$$= \left\{ \delta_a(A_i) \sum_{i'} \sum_{k \neq X_{i'}}^N \left[ \delta_x((k_iX)_i) - \delta_x(X_i) \right] w_{i'}(X \rightarrow k_iX) \right\},$$

$$= \left\{ \delta_a(A_i) \sum_{p \neq X_{ip}}^m \left[ \delta_x((k_{ip}X)_i) - \delta_x(X_i) \right] w_{ip}(X \rightarrow k_{ip}X) \right\},$$

(12)
where in the second step, we substituted \( \delta_x(X_i)w_{ip}(\chi^k_j \to X)P(\chi^k_j X, t) \) for \( \delta_x((\chi^k_j X)_i)w_{ip}(X \to \chi^k_j X)P(X, t) \) 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 use the fact that only changes to the nodes belonging to \( i \) matter. In the last step, we factored out the common elements in the delta functions corresponding to all but the \( p \)th element of \( X_i \) and \( x \), and we used the subscript notation \( (-)_{p=0} \) to denote a vector with element \( p \) removed. The expression on the right-hand side in (12) can be understood independent of the prior algebraic manipulations: the expected change rate of \( \langle [x^n_i] \rangle \) equals the sum of expected rates for each of the nodes \( i_p \) in \( i \) changing its state to \( x_p \) minus the rate of node \( i_p \) changing its state from \( x_p \).

\[ d \frac{d}{dt} \langle [x^n_i] \rangle = \delta_x(A_i) \sum_p \left[ \sum_{a,c} R^{1}_{axcp} \delta_a(X_{ip}) \sum_j A_{ip,j} \delta_c(X_j) \right] \delta_{x_p=0}(X_{i\backslash ip}) \]

With the rates in (13), and noting that

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

the sum inside the averaging brackets in (12) has the form

\[ \sum_{a,c} R^{1}_{axcp} \delta_a(X_{ip}) \sum_j A_{ip,j} \delta_c(X_j) + \sum_{a} R^{0}_{axp} \delta_a(X_{ip}) \]

\[ - \sum_{k \neq X_{ip}} \left[ \sum_{c} R^{1}_{xpc} \sum_j A_{ip,j} \delta_c(X_j) + R^{0}_{xpc} \right] \delta_x(X_i), \]

where we used for the last two terms that \( \delta_{x_p}(X_{ip})\delta_{x_p=0}(X_{i\backslash ip}) \) can be combined to \( \delta_x(X_i) \).

d. **Conversion rates** Next, we will insert the two types of admissible transitions (as discussed in Section II) into (12), 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 neighbour 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 \( w_{ip} \) in (12) are

\[ w_{ip}(X \to \chi^k_j X) = \sum_{a,c} R^{1}_{akcp} \delta_a(X_{ip}) \sum_j A_{ip,j} \delta_c(X_j) + \sum_{a} R^{0}_{akp} \delta_a(X_{ip}). \]

\[ \text{(13)} \]

With the rates in (13), and noting that

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

the sum inside the averaging brackets in (12) has the form

\[ \sum_{a,c} R^{1}_{axcp} \delta_a(X_{ip}) \sum_j A_{ip,j} \delta_c(X_j) + \sum_{a} R^{0}_{axp} \delta_a(X_{ip}) \]

\[ - \sum_{k \neq X_{ip}} \left[ \sum_{c} R^{1}_{xpc} \sum_j A_{ip,j} \delta_c(X_j) + R^{0}_{xpc} \right] \delta_x(X_i), \]

where we used for the last two terms that \( \delta_{x_p}(X_{ip})\delta_{x_p=0}(X_{i\backslash ip}) \) can be combined to \( \delta_x(X_i) \).

e. **Differential invariants** When distributing the products and exploiting the linearity of the averaging brackets, the differential equation (12) for the expected rate of change at \( i \) becomes

\[ \frac{d}{dt} \langle [x^n_i] \rangle = \delta_x(A_i) \sum_p \left[ \sum_{a,c} R^{1}_{axcp} \delta_a(X_{ip}) \sum_j A_{ip,j} \delta_c(X_j) \right] \delta_{x_p=0}(X_{i\backslash ip}) + \sum_{a} R^{0}_{axp} \delta_x(X_i) \]

\[ - \sum_{k \neq X_{ip},c} R^{1}_{xpc} \sum_j A_{ip,j} \delta_c(X_j) - \sum_{k \neq X_{ip}} R^{0}_{xpc} \delta_x(X_i) \]

\[ \text{(13)} \]
After replacing index label $a$ by $k$, using the definition of $[x_i^a]$ in (9) and using $k^p x_i^a$ to indicate $x_i^a$ with its $p$th element replaced by species $k$, this becomes

$$\frac{d}{dt} \langle |x_i^a| \rangle = \sum_p \sum_k \left[ \sum_c R_{k_x,p,c} \left\langle \left[ k^p x_i^a \right] \sum_j A_{ip,j} \delta_c(x_j) \right\rangle + R_{k_x,p}^0 \left\langle \left[ k^p x_i^a \right] \right\rangle \right]$$

$$- \sum_c R_{x_y,p,k}^1 \left\langle [x_i^a] \sum_j A_{ip,j} \delta_c(x_j) \right\rangle - R_{x_y,p}^0 \left\langle [x_i^a] \right\rangle .$$

This shows that $\langle |x_i^a| \rangle$ can increase by a conversion from motifs that differ from $x_i^a$ in only one node (first two terms), or decrease by having any of the nodes in $x_i^a$ convert to another species (last two terms), with both increase and decrease possible via interaction with neighbours and via spontaneous conversion. The factors of form $\left\langle \left[ k^p x_i^a \right] \sum_j A_{ip,j} \delta_c(x_j) \right\rangle$ count the number of $c$-connections of motif $[\cdot]$ (located at $i$) at node $i_p$. The neighbouring 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 neighbourhood sums as follows,

$$\sum_j A_{ip,j} \delta_c(x_j) = \sum_{j \in i} A_{ip,j} \delta_c(x_j) + \sum_{j \notin i} A_{ip,j} \delta_c(x_j),$$

their products with $[x_i^a]$ and $[k^p x_i^a]$ count contributions of $c$-neighbours from within versus from outside the motif separately:

$$\left[ k^p x_i^a \right] \sum_j A_{ip,j} \delta_c(x_j) = \delta_c(x) a e_p \left[ k^p x_i^a \right] + \sum_{y^b \in N_p^c(x^a)} \sum_{j \notin i} y_{i,j}^b ,$$

$$[x_i^a] \sum_j A_{ip,j} \delta_c(x_j) = \delta_c(x) a e_p [x_i^a] + \sum_{y^b \in N_p^c(x^a)} \sum_{j \notin i} y_{i,j}^b .$$

On the right-hand side, $a e_p$ is an $m$-dimensional vector with a 1 at position $p$ and zeroes elsewhere, and $\delta_c(x)$ a vector Kronecker delta function that returns a vector of the size of $x$ with ones where the elements equal $c$ and zeroes elsewhere. Thus, the term $\delta_c(x) a e_p$ counts the number of $c$-connections at position $p$ in the motif $x^a$. We use the notation $i, j$ for the vector $i$ with an extra node index $j$ appended at position $m + 1$. We used $N_p^c(x^a)$ to denote the set of all $(m + 1)$th order connected motifs that can be obtained by linking a new $c$-node to the $p$th position in motif $x^a$, i.e.,

$$N_p^c(x^a) := \bigcup_{\ell=1}^{m+1} \left\{ y^a : |y| = m + 1, y_\ell = c, y_{\ell+1}^b = x^a, (\ell, p) \in b \right\} ,$$

where $y_{\ell+1}^b$ denotes the $m$th order connected motif obtained by deleting the $\ell$th node of $y^b$. The sum over elements of $N_p^c(\cdot)$ in (15) results from the fact that 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 Figure A1, we show this dependence structure (ignoring the labels).

With the substitutions from above, we obtain the \textit{individual-level differential invariants}:

\[
\frac{d}{dt} \langle [x_i^a] \rangle = \sum_{p} \sum_{k} \sum_{c} \left\{ \left( \sum_{c} \delta_c(x) \alpha_p R_{kx,p,c}^1 + R_{kx,p}^0 \right) \langle \left[ x_i^a \right] \rangle \right\} + \sum_{c} \sum_{y^b \in N_p^c(k,x^a)} \sum_{j \neq i} N \langle [ y_{i,j}^b ] \rangle \\
- \left( \sum_{c} \delta_c(x) \alpha_p R_{x_k,kc}^1 + R_{x_k,k}^0 \right) \langle [ x_i^a ] \rangle - \sum_{c} \sum_{y^b \in N_p^c(x^a)} \sum_{j \neq i} N \langle [ y_{i,j}^b ] \rangle ;
\]  

\hspace{2cm} (17)

The \textit{population-level differential invariants} are then obtained by taking the sum \( \sum_{i \in S(m,N)} \) [see definitions of \( [x^a] \) and \( [x_i^a] \) in (1) and (9)] of (17) over all index sets \( i \):

\[
\frac{d}{dt} \langle [x^a] \rangle = \sum_{p} \sum_{k} \sum_{c} \left\{ \left( \frac{R_{kx,p}^0}{n} + \delta_c(x) \alpha_p R_{kx,p,c}^1 \right) \langle \left[ x^a \right] \rangle \right\} - \left( \frac{R_{x_k,k}^0}{n} + \delta_c(x) \alpha_p R_{x_k,kc}^1 \right) \langle [x^a] \rangle \\
+ \sum_{y^b \in N_p^c(k,x^a)} \sum_{y^b \in N_p^c(x^a)} N \langle [ y^b ] \rangle ;
\]  

\hspace{2cm} (18)

where we collected the sums at the front. In closing this section, we make the following remarks.

\textit{i.} The evolution of expected motif counts of size \( m \) is a function of expected motif counts of size \( m \) and \( m+1 \). \textit{ii.} Equation (18) leads to the same equation for motifs in the network that are \textit{isomorphic}, where we call two motifs \( x^a \) and \( y^b \) isomorphic if there exists a permutation \( \pi \) of their node indices that maps them onto each other, i.e. \( \forall x^a, y^b : x^a \simeq y^b \iff \exists \pi : \pi x = y, \pi ax^T = b \). 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. \textit{iii.} The summation in (1) over index tuples \( i \in S(m,N) \) that we used to go from (17) to (18) leads to multiple counting of motifs that possess \textit{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. \( \text{Aut}(x^a) := \{ \pi : \pi x = x, \pi ax^T = a \} \). The multiplicity with which a particular motif \( x^a \) is counted [via (1)] is then equal to \( |\text{Aut}(x^a)|\).

\textit{f. Algebraic invariants} 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,

$$\sum_{x} [x^a] = [a], \quad \sum_{x} [x^a] = 1,$$

(19)

where the right is the normalised form of the left [via (10)]. For networks with homogeneous degree and motifs where $p$ is a stub (also called leaf), i.e. a node with degree 1, there is the additional conservation relation

$$\sum_{k} \binom{k}{p^x a} = \binom{a}{a \setminus p}, \quad \sum_{k} \binom{k}{p^x a} = \binom{x^a_p \emptyset}{a},$$

(20)

where $S$ is the set of stubs in the motif and where the right is again the normalised form of the left [via (10)]. Under the conditions mentioned above, $[a] / [a \setminus p]$ is the number of out-motif connections at the node that connects to the stub. As (19) and (20) 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 (20), 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 $\binom{n}{m}$ ways of labelling its $m$ nodes with $n$ species. The total number of equations from chains is then $n_{ch} = 2 + \sum_{m=2}^{k} \binom{n}{m}$. For networks in which the number of short loops 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 non-chain trees. When loops need to be taken into account however, there will be additional connected motif types at each order; see Table A1 column $n_g$ [35, Table 4.2.1]. To obtain the number of equations contributed by each of these, one has to consider all labelling orderings, knowing that some orderings lead to isomorphic motifs and hence do not add to the total. We have listed in Table A1 the total number of equations $n_{eq}$ resulting from our enumeration algorithm for dynamics with $n = 2$, such as the SIS model. 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 algebraic invariants. Column $n_4$ in Table A1 shows the number of equations for the square lattice and column $n_{4c}$ shows the remaining number after eliminating variables by using algebraic invariants.
IV. CLOSURE SCHEME

In this section, we show how and when motif counts can be decomposed using counts of their smaller components, which in turn can be used to close the moment equation hierarchy (5). Our final formula generalises closures hitherto most commonly used, as shown in e.g. House et al. [22]. We will show that the decomposition is valid when: i. component motif states are conditionally independent and the adjacency structure between them is a tree, ii. the network is spatially homogeneous, iii. the network is sufficiently large, such that the law of large numbers applies. We explain these requirements in detail below and end with a selection of examples from Appendix B.

A. Definitions and background

We will rely on the theory of decomposable Markov networks, following mostly the terminology of Pearl [28, Ch. 3]. We generalise the decomposition from a factorisation involving (1-)cliques to one involving $d$-cliques. The definitions here apply to a general graph $G$, but we will later only apply the decomposition to motifs.

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

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

\[ i \perp \perp_{M} j \mid k \implies X_i \perp \perp X_j \mid X_k, \tag{21} \]

where the $\perp \perp$ notation on the right refers to independence of the random variables: $P(X_i=x_i, X_j=x_j \mid X_k=x_k) = P(X_i=x_i \mid X_k=x_k)P(X_j=x_j \mid X_k=x_k)$. The pair $(X, M)$ defines what is known as a Markov network.

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

\[ G^d = (V, E^d), \text{ where } E^d = \{(i, j) \in V : \text{dist}_G(i, j) \leq d\}. \tag{22} \]
We then say that \((\mathcal{G}, X)\) has \textit{independence beyond distance} \(d\) if \(\mathcal{G}^d\) is the independence map of \(X\), or

\[
i \perp_{\mathcal{G}^d} j \mid k \implies X_i \perp_{\mathcal{G}^d} X_j \mid X_k.
\]  

This means that states of two non-neighbouring sets in \(\mathcal{G}^d\), which by definition (22) are further than \(d\) steps apart in \(\mathcal{G}\), are independent of each other given the state of their separator \(k\).

d. \textbf{Maximal} \(d\)-\textbf{cliques} and \(d\)-\textbf{clique graph} A maximal clique is a complete subgraph not contained in a larger complete subgraph [35]. As a generalisation, \textit{maximal} \(d\)-\textbf{cliques} are maximal subgraphs with distance between any two nodes not greater than \(d\) [30]. Correspondingly, maximal cliques in \(\mathcal{G}^d\) are maximal \(d\)-cliques in \(\mathcal{G}\). The graph \(\mathcal{C}^d\) is the \textit{\(d\)-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. \textbf{Junction graph of} \(d\)-\textbf{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 \(c_i, c_j \subseteq \mathcal{V}\), a link between \(i\) and \(j\) in \(\mathcal{C}^d\) is \textit{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 [29], applied to the \(d\)-clique graph.

f. \textbf{\(d\)-chordality} A graph \(\mathcal{G}\) is chordal when for every loop of length greater than 3, there exists a link in \(\mathcal{G}\) between two non-consecutive nodes of the loop (thus, giving a short-cut, also called \textit{chord} to the loop). As a generalisation, 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. Non-chordal graphs can always be converted to a chordal graph via \textit{triangulation}, i.e. adding chords to every chordless loop of length greater than 3. We will write below \(\text{tr}(\mathcal{G}^d)\) as a \textit{minimal triangulation} of a non-chordal \(\mathcal{G}^d\), obtained by adding the smallest number of links that leads to chordality, unless stated otherwise.

g. \textbf{Decomposability at distance} \(d\) If there is independence beyond distance \(d\) (23) and \(\mathcal{G}\) is \(d\)-chordal, then the joint probability of the network nodes \(\mathcal{G}\) being in a given state \(P(X = x)\) can
be factorised 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 $d$. In such cases, the factorisation is possible because the tree structure between $d$-cliques allows application of the chain rule of conditional probability:

$$P(X = x) = \prod_{i} P(X_{J_i} = x_{J_i} \mid X_{J_i \cap J_i} = x_{J_i \cap J_i}),$$

$$= \prod_{i} P(X_{J_i} = x_{J_i} \mid X_{pa(J_i)} = x_{pa(J_i)}),$$

$$= \prod_{i} P(X_{J_i} = x_{J_i} \mid X_{pa(J_i) \cap J_i} = x_{pa(J_i) \cap J_i}),$$

$$= \prod_{i} \frac{P(X_{J_i} = x_{J_i})}{P(X_{pa(J_i) \cap J_i} = x_{pa(J_i) \cap J_i})},$$

(24)

where we use $\mathcal{J}$ as shorthand for the set of $d$-cliques in $\mathcal{G}$, corresponding to the nodes of $\mathcal{J}^d(\mathcal{G})$. The $d$-cliques take any ordering consistent with $\mathcal{J}^d(\mathcal{G})$ and $pa(J_i)$ is the parent $d$-clique of $d$-clique $J_i$ in $\mathcal{G}$. The steps in (24) are explained as follows. As $d$-chordality makes $\mathcal{J}^d(\mathcal{G})$ a tree, one can:

i. choose any node of this tree as root (determining the ordering of $d$-cliques),

ii. given that any $d$-clique separates its neighbours, recursively use conditional independence of children given parents (line 1 to line 2),

iii. given that any two $d$-cliques in $\mathcal{G}$ are also separated by their intersection, condition instead on intersections (line 3).

h. Dealing with non-$d$-chordality

There are two alternative ways to treat non-chordal $\mathcal{G}^d$: i. it can be made chordal before applying (24), via triangulation, resulting in the altered independence map $\text{tr}(\mathcal{G}^d)$; ii. one can start from the non-tree $\mathcal{J}^d(\mathcal{G})$ and use the ad-hoc formula

$$P(X = x) \approx \zeta \frac{\prod_{i} P(X_{J_i} = x_{J_i})}{\prod_{i,j \neq i} P(X_{J_i \cap J_j} = x_{J_i \cap J_j})},$$

(25)

Because the fraction in (25) does not result from application of the chain rule as in (24), it is not a product of conditional probabilities and hence it does not guarantee the property that each of the node states in $x$ 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 $d$. $\zeta$ in (25) corrects for this inconsistency – see IVB for more detail. After applying (i) to non-chordal graphs, the subgraphs in the resulting $\mathcal{J}^d(\mathcal{G})$ are not all maximal $d$-cliques of $\mathcal{G}$ any more. More precisely, the subgraphs in the $d$-clique consist of the nodes of the maximal cliques of $\text{tr}(\mathcal{G}^d)$ and the links of $\mathcal{G}$. When applying (ii) to non-chordal 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.
B. Motif decomposition

a. Decomposition of motifs at the individual level

(24) may also be applied to motifs embedded in the network with connectivity \( \mathbf{a} \) and chordal independence map \( \mathbf{a}^d \). For now, we ignore that this may in some cases break the independence assumption, but see point IVBe for more detail on this issue. Restricting \( \mathbf{i} \) to nodes that have connectivity \( \mathbf{a} \) and taking \( P(X_i=x) \) as the probability that these are in states with labels \( x \), we can write the decomposition in our earlier notation, using (9, 11),

\[
P(X_i=x) = \sum_y P(X_i=x, X_{\mathbf{V}\setminus i}=y),
\]

\[
= \sum_{x',y} P(X_i=x', X_{\mathbf{V}\setminus i}=y) \delta_x(X_i),
\]

\[
= \langle [x_i^a] \rangle
\]

such that, from (24),

\[
\langle [x_i^a] \rangle = \prod_j \left\{ \frac{\langle [x_{i,J_j}] \rangle}{\langle [x_{i,p_a(J_j)}] \rangle} \right\},
\]

where now \( J \) is the set of \( d \)-cliques in \( \mathbf{a} \). Choosing \( d = \text{diam}(\mathbf{a}) - 1 \) ensures that the decomposition results in component motifs with diameter decreased by one compared to the decomposed motif. For motifs with non-chordal \( \mathbf{a}^d \) one can, as noted above, either triangulate \( \mathbf{a}^d \) first or use the ad-hoc approximation (25). We relied on the package Chordal Graph [36] for triangulation of non-chordal \( \mathbf{a}^d \). The ad-hoc formula (25) applied to motifs is

\[
\langle [x_i^a] \rangle \approx \frac{\prod_j \langle [x_{i,J_j}] \rangle}{\prod_{j,k \neq j} \langle [x_{i,J_j \cap J_k}] \rangle} \prod_{j \in i} \langle [x_j] \rangle^{\gamma_j}.
\]

The consistency correction [written as \( \zeta \) in (25)] here equals \( \prod_{j \in i} \langle [x_j] \rangle^{\gamma_j} \) and ensures that the ad-hoc extension of closures to motifs with non-chordal \( \mathbf{a}^d \) does not result in inconsistency with MF1 [condition 1 of 9, 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. \( \langle [x_i^a] \rangle \rightarrow \prod_{j \in i} \langle [x_j] \rangle \), the right hand side of (28) should reduce to MF1. Therefore, for each \( j, \gamma_j \) is chosen such that this is fulfilled. These ad-hoc steps usually result in violation of the algebraic invariants of Section III f [9]. In the approximations used in Section VI, the bias introduced due to this violation is small. For mitigation of this problem, see [9, 23]. Finally, (27, 28) may be used to close the individual-level equations (17).
b. Decomposition of motifs at the population level  If we take the following spatial homogeneity assumption for motifs of any type and size \( m \leq k \),

\[ \forall i, i' \in I(a) : \langle [x_i^a] \rangle = \langle [x_{i'}^a] \rangle = \frac{\langle [x^a] \rangle}{|a|} = \langle [x^a] \rangle, \tag{29} \]

where \( I(a) := \{ j \in S(m, N) : A_j = a \} \), then (27, 28) are independent of \( i \), such that we can write (27) as

\[ \langle [x^a] \rangle = \prod_j \frac{\langle [x_{J_j}] \rangle}{\langle [x_{pa(J_j) \cap J_j}] \rangle}, \tag{30} \]

and (28) as

\[ \langle [x^a] \rangle \approx \prod_j \frac{\langle [x_{J_j}] \rangle}{\langle [x_{pa(J_j) \cap J_j}] \rangle} \prod_p \langle [x_p] \rangle^{\gamma_p}, \tag{31} \]

which may be used to close the population-level equations (18). In (30, 31), node indexing of a motif is consistent with the node labels \( x \). Recall that we use a single consistent indexing for isomorphic motifs. Not shown in (30, 31) is that we take the average of various alternative ways of decomposing the motif if they exist.

c. Normalisation  We showed the closure formulas for normalised motifs, i.e. we first normalised the counts of motifs via (10), and then applied the closure. Hence, in this case, the counts of induced subgraphs in the network enter into the system of equations as normalisation factors in the unclosed system. One can also decide not to normalise (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 non-normalised motif counts contain them (to see this, substitute each motif count in (30, 31) as \( [y^b] \rightarrow [y^b]/|b| \)). Hence, structural information specific to the considered network enters the mean field equations either when normalising 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. In Mathematica [30], we used the IGraph [37] command IGLADSubisomorphismCount[g, G, Induced -> True], where \( G \) is the considered network and \( g \) the induced subgraph to be counted.

d. Law of large numbers  We will use the population-level closure to study the steady states in a single realisation of a given network. Mean-field quantities are then assumed to be equal to the total counts or frequencies in a single network, instead of their expectations over many realisations. 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 $G$ at distance $d$ is exact when $G$ is $d$-chordal and there is independence beyond distance $d$ (Section IV A). Applying the decomposition to motifs embedded in $G$ instead of to $G$ can be done without additional bias as long as $a$ is a distance-hereditary subgraph of $G$ (i.e. distances between nodes in $a$ are equal to those between corresponding nodes in $G$) and conditional independence relations implied by $a$ are also valid in $G$. As a counterexample, take for $G$ the six-node graph of example 11 in Appendix B and for $a$ its induced subgraph consisting of nodes $\{1, 4, 5, 3, 2\}$. Here, $a$ is not distance-hereditary because $\text{dist}_a(1, 2) = 4 \neq \text{dist}_G(1, 2) = 2$. As a second counterexample, take for $G$ the cyclically indexed square with indices $\{1, 2, 3, 4\}$ and for $a$ the 3-node chain $\{1, 2, 3\}$. In this case, $a$ is distance hereditary, but, while (when assuming independence beyond distance $d=1$) $a$ implies $\{1\} \perp \perp a \{3\} | \{2\}$, this is not true in $G$: $\{1\} \not\perp \perp_G \{3\} | \{2\}$. This occurs because the decomposed motifs are non-maximal $d$-cliques. Therefore, we expect a bias as a consequence of this in the closed mean-field equation hierarchy $[(17) \text{ with } (27) \text{ at individual level, or } (18) \text{ with } (30) \text{ at population level under spatial homogeneity}]$. This bias can be avoided when expressing the equations in terms of maximal $p$-cliques for $p \in \{0, \ldots, k+1\}$ and truncating at diameter $k$ instead of at a given size. Recall that, however, truncation of the equation hierarchy is usually done with regards to size and not with regards to diameter – and so do we. Our approach is to decompose each motif $\mathbf{x}^a$ of size $k+1$ and diameter $\text{diam}(\mathbf{x}^a)$, via a factorisation in terms of its $d$-cliques, with $d = \text{diam}(\mathbf{x}^a) - 1$. Hence, additionally to the problems mentioned here, decomposition of the lower-diameter size $k+1$ motifs will introduce more bias due to a more stringent independence assumption.

C. Examples

Appendix B shows 13 application examples of (30, 31) 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. Comments are added in the last column. Below, we derive the normalisation factors and the non-normalised closures of examples 1-3 of Appendix B for different network types. Note that, unlike in Appendix B, we use letter labels below, for consistency with the main text and the literature.

1. $\mathbf{x}^a = \bigodot \otimes \odot$: This diameter-2 motif has chordal independence map equal to $a$ and decomposes
with (30) as
\[ [\circ \circ \circ ] \approx \frac{[\circ \circ \circ ] [\circ \circ ]}{[\circ ]}, \] (32)
assuming conditional independence beyond distance \( d=1 \). Via (10) we obtain also the closure for the non-normalised counts:
\[ [\circ \circ \circ ] \approx [\circ \circ \circ ] \frac{[\circ \circ ] [\circ \circ ]}{[\circ ]^2}. \] (33)
The frequencies of the induced subgraphs of size 2 and 1, required for normalisation, are
\[ [\circ \circ ] = \kappa N, \quad [\circ ] = N, \] (34)
and total number of triples (3-node motifs) in the network is
\[ [\circ \circ \circ ] + [\circ \circ ] = \sum_i \kappa_i (\kappa_i - 1), \] (35)
with \( \kappa_i \) the number of neighbours of node \( i \) and \( \kappa \) the mean number of neighbours over the whole network. For particular network types (35) 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 \( [\circ \circ ] = 0 \), such that
\[ [\circ \circ \circ ] = \kappa (\kappa - 1) N. \] (36)
Using (33) and (36), we obtain
\[ [\circ \circ \circ ] \approx \frac{\kappa - 1}{\kappa} \frac{[\circ \circ ] [\circ \circ ]}{[\circ ]}. \] (37)
An early use of this closure for networks can be found in Keeling et al. [11].

(b) In a large Erdős-Rényi random network, we have \( \kappa_i \sim \text{Pois}(\kappa) \) and \( \frac{[\circ \circ ]}{([\circ \circ ] + [\circ \circ \circ ])} \approx 0 \) [38]. Hence
\[ [\circ \circ \circ ] = \sum_i N \kappa_i^2 - \sum_i N \kappa_i, \]
\[ = N \left[ \text{E}(\kappa_i^2) - \text{E}(\kappa_i) \right], \]
\[ = N \left[ \text{Var}(\kappa_i) + \text{E}(\kappa_i)^2 - \text{E}(\kappa_i) \right], \]
\[ = \kappa^2 N, \] (38)
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 \( \text{Var}(\cdot) := \text{E}((\cdot)^2) - (\text{E}(\cdot))^2 \), and on the fourth line
we used that, for $\kappa_i \sim \text{Pois}(\kappa)$, we have $E(\kappa_i) = \text{Var}(\kappa_i) = \kappa$. We could also have obtained this result directly from the large-$N$ limit of chains [38]. Using (33) and (38), we obtain

$$[\circ \circ \circ] \approx \frac{[\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ]}{[\circ \circ \circ]}.$$  \hspace{1cm} (39)

This closure was, to the best of our knowledge, first used for networks in Gross et al. [21].

2. $x^2 = \circ \circ \circ$: This diameter-1 motif can be decomposed into its three 0-cliques as

$$[\circ \circ \circ] \approx [\circ \circ \circ],$$  \hspace{1cm} (40)

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

$$[\circ \circ \circ] \approx [\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ] / [\circ \circ \circ].$$  \hspace{1cm} (41)

Alternatively, one can extend the usage of the ad-hoc formula (31) to include non-maximal cliques: using its three 1-cliques in (31), we obtain

$$[\circ \circ \circ] \approx \frac{[\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ]}{[\circ \circ \circ]},$$  \hspace{1cm} (42)

which is known as the Kirkwood closure for triangles [20]. While this often-used closure may perform better than (41), it cannot be rigorously justified based on the independence beyond distance 1 assumption because the chosen 1-cliques are non-maximal. Using (10), we obtain for the closure with non-normalised counts

$$[\circ \circ \circ] \approx \frac{[\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ]}{[\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ]}.$$  \hspace{1cm} (43)

The frequency $[\circ \circ \circ]$ depends on the network type. For instance, if we use the definition of the clustering coefficient $\phi := [\circ \circ \circ] / ([\circ \circ \circ] + [\circ \circ \circ])$ [39], we have [via (34) and (35)] for a network with fixed degree $[\circ \circ \circ] = \phi \kappa (\kappa - 1) N$, such that

$$[\circ \circ \circ] \approx \frac{\kappa - 1}{\kappa^2} \phi N \frac{[\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ]}{[\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ]},$$

which was first used for networks in Keeling [39].

3. $x^3 = \circ \circ \circ$: This diameter-2 motif has chordal independence map $\circ \circ \circ$ and decomposes with (30) as

$$[\circ \circ \circ] \approx \frac{[\circ \circ \circ] [\circ \circ \circ] [\circ \circ \circ]}{[\circ \circ \circ]^2},$$  \hspace{1cm} (44)
when assuming independence beyond distance $d=1$. Alternatively, extending the ad-hoc formula \((31)\) to the three non-maximal 3-cliques, we obtain
\[
\begin{bmatrix}
\mathcal{R} \\
\mathcal{C} \\
\mathcal{G}
\end{bmatrix} \approx
\begin{bmatrix}
\mathcal{G} \\
\mathcal{C} \\
\mathcal{G}
\end{bmatrix}
\begin{bmatrix}
\mathcal{C} \\
\mathcal{G} \\
\mathcal{C}
\end{bmatrix}
\begin{bmatrix}
\mathcal{G} \\
\mathcal{C} \\
\mathcal{G}
\end{bmatrix}
\begin{bmatrix}
\mathcal{C} \\
\mathcal{C} \\
\mathcal{G}
\end{bmatrix}
\begin{bmatrix}
\mathcal{C} \\
\mathcal{C} \\
\mathcal{C}
\end{bmatrix},
\]
where a consistency correction was required. The non-normalised form of this closure was first used in House et al. \cite{22}.

V. OVERVIEW OF STEPS TO OBTAIN MFk

The required steps to obtain the $k$th order moment closure of dynamics on networks are:
1. find the list of independent motifs of order $1, ..., k$ that occur in the considered network,
2. derive the algebraic invariants \((19)\) or \((20)\) and use them to decide which motifs can be eliminated from this list,
3. derive the differential invariants \((18)\) for the remaining motifs up to order $k$,
4. in \((18)\), substitute motifs not in the list of step 2 via algebraic invariants and set motifs not occurring in the considered network to zero,
5. substitute terms of order $(k+1)$ via a closure scheme as described in Section IV and apply the substitutions of step 4 to the resulting $k$th order motifs.

In Appendix D2, this procedure is applied to obtain the fourth-order mean field MF4 of the SIS model on a square lattice. When deriving the more compact MF1-3 in Section VIB for easier demonstration of our method (see \cite{22} for an earlier derivation), we did not follow steps (1-4) in strict order. I.e., all equations were derived first (steps 1 and 3), then variables were eliminated (step 2 and 4) and finally the system was closed (step 5). Doing step 2 before step 3 is especially recommended for higher-order approximations, because it skips derivation of unnecessary equations, saving computation time.

VI. APPLICATION TO THE SIS MODEL

A. Definitions and implementation of the simulations

We apply our method to the SIS model, which is a continuous-time discrete-state Markov chain description of epidemic spreading through a population of susceptibles \cite[e.g.][]{15}. The reaction scheme of the SIS model is
\[
\begin{align*}
\bullet & \xrightarrow{\gamma} \bullet, \\
\bullet & \xrightarrow{\beta} \bullet \bullet,
\end{align*}
\]
(46)
meaning that contagion of susceptibles $S$ occurs over $IS$ links at rate $\beta$, whereas recovery occurs spontaneously at rate $\gamma$. In the study of phase transitions and interacting particle systems, the SIS model is known as the contact process [40], which is typically studied on $d$-dimensional lattices and which has infection rate $\lambda = 2d\beta/\gamma$. In (non-equilibrium) phase transitions theory, it was found to belong to the directed percolation universality class, of which scaling properties have been widely studied [5–7, 41].

a. Implementation with 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

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

as was done in [42–44] for continuation of unstable vibrations in mechanical experiments. In (47), $r$ is one of the conversion rates in $R^0$ or $R^1$. The feedback control [45] makes this rate time dependent by coupling it to the motif frequency $[x^a](t)$ of a chosen motif $x^a$ through the relation (47). 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 (47) settles to an equilibrium $(r^*_c, [x^a]^*_c)$ (in the limit of large $N$) the point $(r^*_c, [x^a]^*_c)$ will be on the equilibrium branch of the simulation without feedback control [46–48]. For the SIS model, we choose

$$r = \beta, \quad x^a = \bullet, \quad g = \infty. \quad (48)$$

This limit for feedback control results in what is called the conserved contact process, as proposed by Tomé and De Oliveira [49]. While the SIS model does not have any unstable steady states, points that are marginally stable, such as its transcritical bifurcation (resulting from a continuous phase transition), are stabilised as well by the control. The steady states in the controlled and uncontrolled SIS model are the same, but the control gives them different stability properties, which makes them easier to study (see Figure 1): i. the set of states with $[\bullet](t) = 0$ is not absorbing, ii. there are no large fluctuations close to the bifurcation, which results in faster convergence of the mean.

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 $[\bullet]$ stays constant. In a simulation based on the Gillespie algorithm [50] this is done in the following steps:
FIG. 1. Feedback control of the SIS model 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.

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 neighbours,
4. advance time with \(\Delta t = -\log(\xi)/\langle [I] \rangle\), (where \(\xi\) is a uniform random variable on the interval \([0, 1]\)),
5. go to step 2.

This loop runs until we observe that \([I] (t)\) is stationary (call this time \(t_e\)), and then for some additional time \(T\) to observe the fluctuations of \([I] (t)\) around its mean. Tomé and De Oliveira [49] derived the effective infection rate \(\tilde{\beta}\) for each chosen count \([I]\) of infected nodes by noting that, because for every infection event there is a recovery event,

\[
\tilde{\beta}\langle [I]\rangle = \gamma [I],
\]

where \(\langle \cdot \rangle\) is an average over many independent realisations, such that

\[
\tilde{\beta} = \gamma \langle [I]\rangle,
\]

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

\[ \tilde{\beta} = \gamma \int_{t_\text{e}}^{t_\text{e} + T} \mathcal{E}(t) \, dt / T. \]

(49)

In this manner, the error bars of estimates of \( \tilde{\beta} \) can be made arbitrarily small.

**B. Mean field equations**

In the SIS model, we have \( n = 2 \) species. The \( 2 \times 2 \) matrix of spontaneous conversion rates and the \( 2 \times 2 \times 2 \) tensor of conversion rates due to nearest-neighbour interaction for the SIS model are

\[
R^0 = \begin{pmatrix} 0 & 0 \\ \gamma & 0 \end{pmatrix}, \quad R^1 = \begin{pmatrix} (0,0) & (0,\beta) \\ (0,0) & (0,0) \end{pmatrix}.
\]

We derived the mean-field models up to fifth order for the square lattice and up to second order for other networks, including cubic/hypercubic lattices, random regular networks and Erdős-Rényi random networks. The general differential invariants (18), are listed up to order 3, for this example in Appendix C. The coefficients of the general fourth-order differential invariants are shown in Appendix D1, and the closed fourth-order model applied to a square lattice is derived in Appendix D2 (both as output of the supplementary Mathematica script in html form). Recall that we write in the text \( \langle \cdot \rangle \) as \( [\cdot] \), assuming the LLN holds (in the htmls we use \( \langle \cdot \rangle \) instead).

We define the correlation between neighbouring node states \( a \) and \( b \) as in Keeling [39], i.e.

\[
C_{ab} = \frac{N}{\kappa} \frac{[\mathbf{G} \mathbf{B}]}{[\mathbf{G}][\mathbf{B}]}, \tag{50}
\]

(where \( \kappa \) is the mean degree) or when motif frequencies are normalised,

\[
C_{ab} = \frac{[\mathbf{G} \mathbf{B}]}{[\mathbf{G}][\mathbf{B}]}. \tag{51}
\]

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). In the discussion of the results for lattices, we will also show a correlation (only between infected nodes) at a given distance between the considered nodes, like the correlation functions used in the study of phase transitions [e.g. 6], i.e.

\[
C_{ab}^D = \frac{1}{\sum_{ij} \mathbf{1}_{\text{dist}(i,j)=D}} \sum_{ij} \frac{[\mathbf{G} \mathbf{B}]}{[\mathbf{G}][\mathbf{B}]} \mathbf{1}_{\text{dist}(i,j)=D}, \tag{52}
\]
where the distance \( \text{dist}(i, j) \) is the length of the shortest path between \( i \) and \( j \). Note that the correlations between neighbouring nodes (50) is a special case of this, i.e. \( C_{ab} = C_{ab}^1 \). An alternative way to obtain these correlations is via

\[
C_{ab}^D = \frac{\sum_{c_1, \ldots, c_{D-1}} [ac_1 \ldots c_{D-1}b]}{[\circ][\circ]},
\]

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

1. **MF1**

The first-order mean field originates from the molecular field approximation in statistical physics \([1, 2]\) and is now commonly known as the ‘mean field model’ \([14–16]\). MF1 only considers node states \((m = 1)\) and neglects correlations beyond distance 0. There are two motif types of size 1 and hence two equations (A1) and (A2), one for each species type, \( S \) and \( I \). Note that this is not a closed system of equations. To close the system at \( m = 1 \), \([\circ\circ] \) needs to be expressed in terms of \([\circ] \) and \([\circ] \). First we use the conservation relation \([\circ] + [\circ] = N \) to substitute \([\circ] = N - [\circ] \), such that

\[
\frac{d}{dt}[\circ] = \gamma(N - [\circ]) - \beta[\circ\circ].
\]

When assuming no correlation \((C_{SS} = C_{SI} = C_{II} = 1)\) between neighbouring species,

\[
[\circ\circ] \approx \kappa N \frac{[\circ]}{N} = \kappa \frac{N}{N} (N - [\circ]) [\circ],
\]

we obtain the first-order mean field:

\[
\frac{d}{dt}[\circ] = \left( \gamma - \frac{\kappa \beta}{N} [\circ] \right) (N - [\circ]),
\]

which, after normalisation [via (10)], yields

\[
\frac{d}{dt}[\circ] = (\gamma - \kappa \beta [\circ])(1 - [\circ]).
\]

The steady state solutions are then

\[
[\circ]_1^* = 1, [\circ]_2^* = \frac{\gamma}{\kappa \beta}.
\]

At \( \beta / \gamma = \kappa^{-1} \), the solution \([\circ]_1^* \) becomes unstable due to a transcritical bifurcation, also known as the epidemic threshold in epidemiology.
The first-order mean field 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 $[\bullet] + [\bullet] = N \to \infty$. A complete network implies $[\bullet \bullet] = [\bullet] [\bullet]$ and the large-$N$ limit implies $\langle [\bullet \bullet] [\bullet \bullet] \rangle = \langle [\bullet] \rangle \langle [\bullet] \rangle$. The latter can be seen via

\[
\langle [\bullet \bullet] [\bullet \bullet] \rangle = \langle [\bullet] (1 - [\bullet]) \rangle,
\]

\[
= \langle [\bullet] - [\bullet]^2 \rangle,
\]

\[
(N \to \infty) \Rightarrow \langle [\bullet] \rangle - \langle [\bullet] \rangle^2,
\]

\[
= \langle [\bullet] \rangle (1 - \langle [\bullet] \rangle),
\]

\[
= \langle [\bullet] \rangle \langle [\bullet] \rangle,
\]

where on the third line we apply the LLN, such that $\text{Var}([\bullet]) = \langle [\bullet] \rangle^2 - \langle [\bullet] \rangle^2 \to 0$ when $N \to \infty$.

2. **MF2**

The second-order mean field originates from the Bethe approximation in statistical physics [3] and is now commonly known as the ‘pair approximation’ [14–16]. MF2 neglects dependence beyond distance 1 and is obtained by considering all motifs up to size 2, i.e. equations (A1-A5). These depend in turn on two types of order 3 motifs: chains and triangles. The number of algebraic invariants used for elimination depends on whether the networks have a homogeneous degree.

a. **Networks with homogeneous degree** Before closing the system of equations we use the algebraic invariants

\[
[\bullet] + [\bullet] = N, \quad (55)
\]

\[
[\bullet \bullet] + [\bullet \bullet] = \kappa [\bullet], \quad (56)
\]

\[
[\bullet \bullet \bullet] + [\bullet \bullet \bullet] = \kappa [\bullet], \quad (57)
\]

to eliminate $[\bullet]$, $[\bullet \bullet]$ and $[\bullet \bullet \bullet]$, resulting in

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

\[
\frac{d}{dt} [\bullet \bullet] = 2 \gamma (\kappa [\bullet] - [\bullet \bullet]) - 2 \beta [\bullet \bullet \bullet] - 2 \beta \left[ \bullet \bullet \bullet \bullet \right].
\]
The networks considered in this paper do not contain triangular subgraphs, such that \( [\bullet \bullet \bullet] = 0 \). Applying the closure (37) for degree-homogeneous networks [resulting from (30)],

\[
[\bullet \bullet \bullet] \approx \kappa (\kappa - 1) N \frac{\kappa N}{\kappa N},
\]

\[
\approx \frac{\kappa - 1}{\kappa} [\bullet \bullet \bullet],
\]

\[
\approx \frac{\kappa - 1}{\kappa} (\kappa [\bullet] - [\bullet \bullet \bullet]),
\]

we obtain

\[
\frac{d}{dt} [\bullet] = \gamma (N - [\bullet]) - \beta [\bullet \bullet \bullet],
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet] = 2(\kappa [\bullet] - [\bullet \bullet \bullet])(\gamma - \beta \frac{\kappa - 1}{\kappa} [\bullet \bullet \bullet]).
\]

Normalising the equations via (10) results in

\[
\frac{d}{dt} [\bullet] = \gamma (1 - [\bullet]) - \kappa \beta ([\bullet] - [\bullet \bullet \bullet]),
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet] = 2([\bullet] - [\bullet \bullet \bullet]) [\gamma - \beta (\kappa - 1) [\bullet \bullet \bullet]].
\]

The steady state solutions for this second-order mean-field model for homogeneous networks are

\[
([\bullet]^{*}, [\bullet \bullet \bullet]^{*})_1 = (1, 1), \quad ([\bullet]^{*}, [\bullet \bullet \bullet]^{*})_2 = \left( \frac{1}{\kappa \beta (\kappa - 1)/\gamma - 1}, \frac{\kappa - 1}{\kappa \beta (\kappa - 1)/\gamma - 1} \right).
\]

Again, we have a transcritical bifurcation at the epidemic threshold, which is now located at \( \beta/\gamma = (\kappa - 1)^{-1} \). From (58), we can derive steady non-trivial state correlations via (51), yielding

\[
C^*_\text{II} = \frac{(\gamma - \beta \kappa)(\gamma - \beta (\kappa - 1)\kappa)}{\beta \kappa^2 (\beta (\kappa - 1) - \gamma)}, \quad C^*_{SI} = 1 - \frac{\gamma}{\beta \kappa (\kappa - 1)}, \quad C^*_{SS} = \frac{\beta (\kappa - 1) \kappa - \gamma}{\beta (\kappa - 1)^2}.
\]

b. Networks with heterogeneous degree

Here, (56) and (57) do not hold, but the total frequency of nodes, chains and triangles is still conserved. Hence, the algebraic invariants are (55) from order 1 and

\[
[\bullet \bullet] + [\bullet \bullet \bullet] + 2 [\bullet \bullet \bullet] = \kappa N.
\]

Using these algebraic invariants to eliminate [\bullet] and [\bullet \bullet \bullet] in (56) and (57), we obtain the three-dimensional model

\[
\frac{d}{dt} [\bullet] = \gamma (N - [\bullet]) - \beta [\bullet \bullet \bullet],
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet] = 2\gamma [\bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet],
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet \bullet] = \gamma \kappa N - \gamma [\bullet \bullet \bullet] - (\beta + 3\gamma) [\bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet] - \beta [\bullet \bullet \bullet] - \beta [\bullet \bullet \bullet].
\]
The only network type with heterogeneous degree distribution that we consider is the Erdős-Rényi random network. The probability of having short loops in ER random networks approaches zero as \( N \to \infty \) [38]. For large ER random networks, we can hence ignore any motif that has loops, and here we can accordingly set all \([\underbar{C}_C]=0\). Using the closure for ER networks (39) [resulting from (30)],

\[
[\bullet\bullet\bullet] \approx \frac{[\bullet\bullet]}{[\bullet]}, \quad [\bullet\bullet\bullet] \approx \frac{[\bullet]^2}{[\bullet]},
\]

we obtain

\[
\begin{align*}
\frac{d}{dt}[\bullet] &= \gamma(N - [\bullet]) - \beta[\bullet\bullet]\, , \\
\frac{d}{dt}[\bullet\bullet] &= 2[\bullet\bullet](\gamma - \beta\frac{[\bullet\bullet]}{[\bullet]}), \\
\frac{d}{dt}[\bullet\bullet\bullet] &= \gamma\kappa N - \gamma[\bullet\bullet] - (\beta + 3\gamma)[\bullet\bullet\bullet] + \beta\frac{[\bullet\bullet\bullet]}{[\bullet\bullet]}((\bullet\bullet\bullet) - [\bullet\bullet\bullet]),
\end{align*}
\]

which, after normalisation via (10) becomes

\[
\begin{align*}
\frac{d}{dt}[\bullet] &= \gamma(1 - [\bullet]) - \kappa\beta[\bullet\bullet]\, , \\
\frac{d}{dt}[\bullet\bullet] &= 2[\bullet\bullet](\gamma - \kappa\beta\frac{[\bullet\bullet]}{[\bullet]}), \\
\frac{d}{dt}[\bullet\bullet\bullet] &= \gamma(1 - [\bullet\bullet]) - (\beta + 3\gamma)[\bullet\bullet\bullet] + \kappa\beta\frac{[\bullet\bullet\bullet]}{[\bullet\bullet]}((\bullet\bullet\bullet) - [\bullet\bullet\bullet]).
\end{align*}
\]

There are three steady states, of which two are in the admissible range,

\[
\begin{align*}
([\bullet]^*, [\bullet\bullet]^*, [\bullet\bullet\bullet]^*)_1 &= (1, 1, 0), \\
([\bullet]^*, [\bullet\bullet]^*, [\bullet\bullet\bullet]^*)_2 &= \left(\frac{\sqrt{4\gamma + \beta(\kappa - 1)^2}}{2\sqrt{\beta}} - \frac{(\kappa - 1)/2}{2\kappa\beta\gamma}, \frac{\gamma\sqrt{4\gamma + \beta(\kappa - 1)^2} - \gamma\sqrt{3}(\kappa - 1)}{2\kappa\beta\gamma}, \frac{\gamma\sqrt{3}(\kappa + 1) - \gamma\sqrt{4\gamma + \beta(\kappa - 1)^2}}{2\kappa\beta\gamma}\right).
\end{align*}
\]

Also here we have a transcritical bifurcation at the epidemic threshold, which is now located at \( \beta/\gamma=\kappa^{-1} \), as in the first-order mean field model. The steady state correlations are

\[
\begin{align*}
C_{II}^* &= \frac{\sqrt{\beta\gamma} \left(\sqrt{\beta(\kappa - 1)^2 + 4\gamma} - \sqrt{\beta(\kappa + 1)}\right)^2}{\sqrt{\beta\kappa \left(\sqrt{\beta(\kappa - 1)^2 + 4\gamma} - \sqrt{3}(\kappa + 1)\right)^2}}, \\
C_{SI}^* &= C_{SS}^* = \frac{2\gamma}{\kappa\sqrt{\beta^2(\kappa - 1)^2 + 4\beta\gamma} - \beta(\kappa - 1)\kappa}.
\end{align*}
\]

3. MF3

Approximations of higher order than two correspond to cluster variation approximations in statistical physics [4]. MF3 results from neglecting dependence beyond distance 2 and is commonly
known as the ‘triple approximation’ [15, 22]. The differential invariants for motifs up to size 3 are (A1-A15). We will only apply MF3 to the square lattice, such that we can ignore all motifs that contain triangles, i.e. we can omit equations (A6-A9) and set \([\circ\circ\circ] = [\circ\circ\circ\circ] = [\circ\circ\circ\circ\circ] = 0\) in the 11 remaining equations (A1-A5, A10-A15). Here, we will normalise all motifs in advance for convenience. The 7 algebraic invariants with normalised motifs are

\[
\begin{align*}
[\bullet] + [\circ] &= 1, \\
[\bullet\bullet] + [\bullet\circ] &= [\bullet], \\
[\bullet\bullet\bullet] + [\bullet\bullet\circ] &= [\bullet], \\
[\bullet\bullet\circ\circ\circ] + [\bullet\bullet\circ\circ\circ\circ] &= [\bullet\circ], \\
[\bullet\circ\circ\circ\circ\circ] + [\bullet\circ\circ\circ\circ\circ\circ] &= [\bullet\circ], \\
[\bullet\circ\circ\circ\circ\circ\circ] + [\bullet\circ\circ\circ\circ\circ\circ\circ] &= [\bullet\circ], \\
[\bullet\circ\circ\circ\circ\circ\circ\circ] + [\bullet\circ\circ\circ\circ\circ\circ\circ\circ] &= [\bullet\circ].
\end{align*}
\]

Hence, after substitution we will have \(11 - 7 = 4\) remaining equations. We will eliminate \([\bullet\circ],[\bullet\circ\circ],[\bullet\circ\circ\circ],[\bullet\circ\circ\circ\circ],[\bullet\circ\circ\circ\circ\circ]\) using the algebraic invariants, which results in

\[
\begin{align*}
\frac{d}{dt}[\bullet\circ] &= \gamma(1 - [\bullet\circ]) - 4\beta([\bullet\circ] - [\bullet\circ\circ]), \\
\frac{d}{dt}[\bullet\circ\circ] &= 2\gamma([\bullet\circ] - [\bullet\circ\circ]) - 6\beta([\bullet\circ\circ] - [\bullet\circ\circ\circ]), \\
\frac{d}{dt}[\bullet\circ\circ\circ] &= \gamma[\bullet\circ\circ\circ] + 2\gamma([\bullet\circ\circ\circ] - [\bullet\circ\circ\circ\circ]) \\
&\quad - 14/3\beta[\bullet\circ\circ\circ\circ] - 2\beta[\bullet\circ\circ\circ\circ\circ] - 4/3\beta[\bullet\circ\circ\circ\circ\circ\circ], \\
\frac{d}{dt}[\bullet\circ\circ\circ\circ] &= 2\gamma([\bullet\circ\circ\circ] - [\bullet\circ\circ\circ\circ]) - (2\beta + \gamma)[\bullet\circ\circ\circ\circ\circ] \\
&\quad - 14/3\beta[\bullet\circ\circ\circ\circ\circ\circ] + 2\beta[\bullet\circ\circ\circ\circ\circ\circ\circ] - 4/3\beta[\bullet\circ\circ\circ\circ\circ\circ\circ\circ],
\end{align*}
\]

where we have used \(\kappa = 4\) for the square lattice. We apply (30) to the chains and star, and (31) to the loops (using the extension to non-maximal 2-cliques) to obtain the normalised closures (see also Appendix B, examples 3-5)

\[
\begin{align*}
[\bullet\circ\circ\circ\circ] &\approx \frac{[\bullet\circ\circ\circ\circ][\bullet\circ\circ\circ\circ]}{[\bullet\circ\circ\circ\circ]} = \frac{([\bullet\circ\circ\circ\circ] - [\bullet\circ\circ\circ\circ\circ\circ])}{[\bullet\circ\circ\circ\circ]}[\bullet\circ\circ\circ\circ\circ\circ], \\
[\bullet\circ\circ\circ\circ\circ] &\approx \frac{[\bullet\circ\circ\circ\circ\circ][\bullet\circ\circ\circ\circ\circ\circ]}{[\bullet\circ\circ\circ\circ\circ]} = \frac{([\bullet\circ\circ\circ\circ\circ] - 2[\bullet\circ\circ\circ\circ\circ\circ\circ])}{[\bullet\circ\circ\circ\circ\circ]}[\bullet\circ\circ\circ\circ\circ\circ\circ], \\
[\bullet\circ\circ\circ\circ\circ\circ\circ] &\approx \frac{[\bullet\circ\circ\circ\circ\circ\circ\circ][\bullet\circ\circ\circ\circ\circ\circ\circ\circ]}{[\bullet\circ\circ\circ\circ\circ\circ\circ]} = \frac{([\bullet\circ\circ\circ\circ\circ\circ\circ] - 2[\bullet\circ\circ\circ\circ\circ\circ\circ\circ\circ])}{[\bullet\circ\circ\circ\circ\circ\circ\circ]}[\bullet\circ\circ\circ\circ\circ\circ\circ\circ\circ],
\end{align*}
\]

31
and substitute them into (63, 64) such that

\[
\frac{d}{dt} J_K \approx \frac{[\bullet\bullet\bullet]^2 [\bullet\bullet\bullet]^2}{([\bullet\bullet])^4} = \frac{[\bullet\bullet\bullet]^2 ([\bullet] - 2[\bullet\bullet] + [\bullet\bullet\bullet])^2}{([\bullet] - [\bullet\bullet])^4},
\]

and substitute them into (63, 64) such that

\[
\frac{d}{dt} J_K = \gamma [\bullet\bullet\bullet] + 2\gamma ([\bullet\bullet\bullet] - [\bullet\bullet\bullet]) - \frac{14}{3}\beta ([\bullet\bullet\bullet] - [\bullet\bullet\bullet]) [\bullet\bullet\bullet]
\]

\[
-2\beta ([\bullet] - [\bullet\bullet\bullet]) [\bullet\bullet\bullet]^2 - \frac{4}{3}\beta [\bullet\bullet\bullet] ([\bullet\bullet\bullet] - [\bullet\bullet\bullet])^2 [\bullet\bullet\bullet]^2,
\]

\[
\frac{d}{dt} J_K = 2\gamma ([\bullet] - [\bullet\bullet\bullet]) - (2\beta + 3\gamma) [\bullet\bullet\bullet] - \frac{14}{3}\beta [\bullet\bullet\bullet] ([\bullet] - 2[\bullet\bullet\bullet] + [\bullet\bullet\bullet])
\]

\[
+2\beta ([\bullet] - [\bullet\bullet\bullet]) [\bullet\bullet\bullet]^2 - \frac{4}{3}\beta [\bullet\bullet\bullet]^2 ([\bullet] - 2[\bullet\bullet\bullet] + [\bullet\bullet\bullet])^2 [\bullet\bullet\bullet]^2.
\]

The steady state solutions of (61, 62, 66, 67) are roots of a ninth-order polynomial, of which two are admissible (see Figure 2 for numerical results).

4. **MF4 and MF5**

In Appendix D1, 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 algebraic invariants is respectively 14 and 37. The derivation of MF4 with closure is shown in Appendix D2. The steady states of MF4 and MF5 are shown in Figure 2 in Section VI C.

C. **Comparison to simulations**

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

a. **Square lattice** In Figure 2a, we show, for the square lattice, the steady state fraction of infecteds versus $\beta/\gamma$ 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 non-trivial (endemic) steady states when approaching the critical value of $\beta/\gamma$ from above. Figure 2b 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 $\beta/\gamma$ from above and has a
singularity at the bifurcation. E.g. for MF2, via (59), we have $C_{II}^*(\beta/\gamma) \to \infty$ for $\beta/\gamma \downarrow (\kappa-1)^{-1}$ (limit from above in the endemic equilibrium). We recall that the errors in the mean-field models visible in Figure 2a can be due to violation of the statistical dependence assumption beyond distance $d = \text{diam}(a) - 1$, non-chordality 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 [20]. 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 Figure 1b), spatial inhomogeneity can not be a source of bias here. This leaves statistical dependence and non-chordality as only sources of bias. In the square lattice there are always loops 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 need to consider motifs of diameter up to the network diameter to be exact, with any lower-order model being inexact.

As MF1 assumes no correlation between the states of neighbouring 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 non-chordality. 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 A2 in Appendix A shows the correlation as a function of distance from a central point [via (52)], for various values of $\beta/\gamma$. 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 [6].

Comparing the square lattice to the 4-neighbour random regular graph, we can see that there is also a phase transition, but there is substantially less bias than in the square lattice (Figure 3 blue $\times$ vs $+$). 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 [17, 18].

b. General cubic lattices and random regular networks We show in Figure 3 how the steady states and correlations in MF1, MF2 and simulations depend on the number of neighbours in d-
FIG. 2. Comparison of the mean-field approximations of order 1 to 5 (lines) with simulations (markers) of the SIS epidemic model on a square lattice as a function of $\beta/\gamma$: (a) Nontrivial steady states, (b) Correlations at distance 1 [shown for MF2 (solid), MF5 (dashed), and simulations (dots)].

dimensional cubic lattices and random regular networks. When $d$ is the lattice dimension, the lattice degree is $\kappa = 2d$. The observations of the square lattice generalise to cubic/hypercubic lattices and random regular networks (at least up to $\kappa = 10$): 

i. there is a transcritical bifurcation at a particular value of $\beta/\gamma$, 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,  

vi. the bias is larger closer to the bifurcation. According to MF1, the bifurcation occurs at $\kappa^{-1} = (2d)^{-1}$ and according to MF2 at $(2d - 1)^{-1} = (\kappa - 1)^{-1}$ (see Sections VIB1 and VIB2). Liggett [51] proved that for lattices, the critical value predicted by MF2 is a lower bound. Figure 3a shows that MF2 and this lower bound is approached increasingly closely when the lattice dimension increases. Due to higher clustering of neighbours, the epidemic threshold in lattices is higher than that in the corresponding random regular network [39], but this difference decreases with dimension/degree (Figure 3a). The steady states of a 5-dimensional hypercubic lattice and of a random regular network with degree 10 are indistinguishable from each other and from MF2. This is a consequence of the fact that two 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, having many intersections in $d \leq 4$ means that one cannot ignore alternative infection paths. In $d > 4$, 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 [18], MF2 should then be more accurate.

c. Erdős-Rényi random networks  Finally, we show in Figure 4 how the steady states and correlations in MF1, MF2 and simulations depend on the number of neighbours in an Erdős-Rényi
FIG. 3. Comparison of the mean-field approximations (dotted lines: MF1, solid lines: MF2) and simulations (markers) of the SIS epidemic model on lattices and random regular networks with number of neighbours 4,6,8,10 as a function of $\beta/\gamma$: (a) Nontrivial steady states, (b) Correlations: $C_{II}^*$ (above 1), $C_{SI}^*$ (below 1).

FIG. 4. Comparison of the mean-field approximations (dotted lines: MF1, solid lines: MF2) and simulations (markers) of the SIS epidemic model on an Erdős-Rényi random network with number of neighbours 4,6,8,10 as a function of $\beta/\gamma$: (a) Nontrivial steady states, (b) Correlations: $C_{II}^*$ (above 1), $C_{SI}^*$ (below 1).

random network. Recall that MF2 on an Erdős-Rényi random network is different from that of the networks above because it has one fewer algebraic invariant. 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 $\beta/\gamma$, 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 loops beyond closure distance is the dominant cause for mean field model biases.
VII. CONCLUSIONS

Previous work found that lower-order moment closures work well on networks that are locally treelike, i.e. that have few short loops [15, 17–20]. Our method developed in this paper shows how to systematically extend moment closure to networks with many short loops of any size, where higher presence of progressively longer loops requires a progressively higher order of approximation, particularly in parameter regimes with long-range correlations, such as near continuous phase transitions. It should hence lend itself well to study dynamics on networks with density and size of short loops larger than in random networks but smaller than in lattices, such as often observed in real-world networks [55], particularly when the network is structurally homogeneous. Also, despite our focus on the population-level equations and closure, the individual-level equations and closures derived in Sections III-IV may be used to study heterogeneous problems as well, at cost of an increase in the number of equations. Furthermore, while our treatment focused on static networks with at most nearest-neighbour interactions, it can be extended to adaptive networks – i.e. when $A$ is a random variable depending on the network state – such as those studied in [21, 56], and to dynamics with higher-order interactions [57] – requiring reaction rate tensors $R^p$ with $p \geq 2$. We also suspect that our generalisation of moment closure to higher orders can be used to extend the use of message-passing methods [58–60] for epidemic modelling to graphs with loops.

To end, we list some of the many open questions that remain. Can the differential invariants be generally and efficiently rewritten in terms of states on maximal $d$-cliques such that each order of approximation has equations for motifs of a given diameter instead of a given size, thereby avoiding bias due to non-inheritance of independence relations by certain motifs (see IV Be)? How do higher-order moment closures perform in systems with other types of phase transitions, or with other types of behaviours, such as bistability, oscillations, waves, or formation of patterns? Can general structural and dynamical properties be identified other than those mentioned here that make moment closures fail, quantitatively or qualitatively?
ACKNOWLEDGMENTS

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grants EP/N023544/1 and EP/V04687X/1.

[1] P. Weiss, J. Phys. Theor. Appl. 6, 661 (1907).
[2] W. L. Bragg and E. J. Williams, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 145, 699 (1934).
[3] H. A. Bethe, Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences 150, 552 (1935).
[4] R. Kikuchi, Phys. Rev. 81, 988 (1951).
[5] J. Marro and R. Dickman, Nonequilibrium Phase Transitions in Lattice Models (Cambridge University Press, 1999).
[6] M. Henkel, H. Hinrichsen, and S. Lübeck, Non-Equilibrium Phase Transitions, Theoretical and Mathematical Physics (Springer Netherlands, Dordrecht, 2008).
[7] T. Tomé and M. J. De Oliveira, Stochastic dynamics and irreversibility (Springer, 2015) p. 394.
[8] H. Matsuda, N. Ogita, A. Sasaki, and K. Sato, Progress of Theoretical Physics 88, 1035 (1992).
[9] U. Dieckmann, R. Law, and J. A. J. Metz, The geometry of ecological interactions: simplifying spatial complexity (Cambridge University Press, 2000) p. 564.
[10] S. Kéfi, M. Rietkerk, M. van Baalen, and M. Loreau, Theoretical Population Biology 71, 367 (2007).
[11] M. J. Keeling, D. A. Rand, and A. J. Morris, Proceedings of the Royal Society B: Biological Sciences 264, 1149 (1997).
[12] D. A. Rand, Advanced Ecological Theory, Tech. Rep. (1999).
[13] C. Kuehn, Control of self-organizing nonlinear systems, 253 (2016).
[14] M. A. Porter and J. P. Gleeson, Frontiers in Applied Dynamical Systems: Reviews and Tutorials 4 (2016).
[15] 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.
[16] M. Newman, Networks (Oxford university press, 2018).
[17] K. J. Sharkey, I. Z. Kiss, R. R. Wilkinson, and P. L. Simon, Bulletin of Mathematical Biology 77, 614 (2015).
[18] K. J. Sharkey and R. R. Wilkinson, Mathematical Biosciences 264, 74 (2015).
[19] L. Pellis, T. House, and M. J. Keeling, Journal of Theoretical Biology 382, 160 (2015).
[20] K. J. Sharkey, Journal of Mathematical Biology 57, 311 (2008).
[21] T. Gross, C. J. D’Lima, and B. Blasius, Physical Review Letters 96, 208701 (2006).
[22] T. House, G. Davies, L. Danon, and M. J. Keeling, Bulletin of Mathematical Biology 71, 1693 (2009).
[23] N. Peyrard, U. Dieckmann, and A. Franc, Theoretical Population Biology 73, 383 (2008).
[24] V. Marceau, P.-A. Noël, L. Hébert-Dufresne, A. Allard, and L. J. Dubé, Phys. Rev. E 82, 036116 (2010).
[25] J. P. Gleeson, Physical Review X 3, 021004 (2013).
[26] P. G. Fennell and J. P. Gleeson, SIAM Review 61, 92 (2019).
[27] G. St-Onge, V. Thibeault, A. Allard, L. J. Dubé, and L. Hébert-Dufresne, Phys. Rev. E 103, 032301 (2021).
[28] J. Pearl, Probabilistic reasoning in intelligent systems: networks of plausible inference (Elsevier, 1988).
[29] D. Barber, Bayesian Reasoning and Machine Learning (Cambridge University Press, 2012).
[30] Wolfram Research, Inc., “Mathematica, Version 12.3.1,” (2021), Champaign, IL.
[31] MathWorks, “MATLAB 2021a,” (2021).
[32] F. Schilder, H. Dankowicz, and M. Li, “Continuation core and toolboxes (coco),” (2021).
[33] A. Dhooge, W. Govaerts, and Y. A. Kuznetsov, ACM Transactions on Mathematical Software (TOMS) 29, 141 (2003).
[34] C. Gardiner, Stochastic Methods: A Handbook for the Natural and Social Sciences (Springer Series in Synergetics, 2009).
[35] F. Harary and E. M. Palmer, Graphical Enumeration (Elsevier, 1973).
[36] Y. Bulatov, “Chordal graph package,” (2011).
[37] S. Horvát, “IGraph/M—the igraph interface for Mathematica, version 0.5.1,” (2021).
[38] For an ER random network, the expected number of $m$-node loops and $m$-node chains (counting all ordered $m$-tuples without repetition) are $L(N) = \frac{N!}{(N-m)!}(\frac{k}{N-1})^m$ and $C(N) = \frac{N!}{(N-m)!}(1 - \frac{k}{N-1})^{m-1}$, where $\frac{k}{N-1}$ 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 $N$, loops 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.
[39] M. J. Keeling, Proceedings of the Royal Society B: Biological Sciences 266, 859 (1999).
[40] T. E. Harris, The Annals of Probability 2, 969 (1974).
[41] U. C. Täuber, Critical Dynamics: A Field Theory Approach to Equilibrium and Non-Equilibrium Scaling Behavior (Cambridge University Press, 2012) pp. 1–511.
[42] J. Sieber, A. Gonzalez-Buelga, S. A. Neild, D. J. Wagg, and B. Krauskopf, Phys. Rev. Lett. 100, 244101 (2008).
[43] F. Schilder, E. Bureau, I. F. Santos, J. J. Thomsen, and J. Starke, Journal of Sound and Vibration 358, 251 (2015).
[44] D. Barton, Mechanical Systems and Signal Processing 84, 54 (2017).
[45] (47) is the simplest possible form of feedback control as it changes only one input ($r$) depending on one output ([|x|]) and is static (no further processing of [|x|] 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).

[46] D. A. W. Barton and J. Sieber, Phys. Rev. E 87, 052916 (2013).
[47] J. Sieber, O. E. Omel’chenko, and M. Wolfrum, Phys. Rev. Lett. 112, 054102 (2014).
[48] L. Renson, D. Barton, and S. Neild, International Journal of Bifurcation and Chaos 27, 1730002 (2017).
[49] T. Tomé and M. J. De Oliveira, Physical Review Letters 86, 5643 (2001).
[50] D. T. Gillespie, Annual Review of Physical Chemistry 58, 35 (2007).
[51] T. M. Liggett, Interacting Particle Systems, Classics in Mathematics (Springer Berlin Heidelberg, Berlin, Heidelberg, 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] D. J. Watts and S. H. Strogatz, nature 393, 440 (1998).
[56] T. Gross and I. G. Kevrekidis, EPL (Europhysics Letters) 82, 38004 (2008).
[57] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, Physics Reports 874, 1 (2020).
[58] B. Karrer and M. E. Newman, Physical Review E 82, 016101 (2010).
[59] R. R. Wilkinson and K. J. Sharkey, Physical Review E 89, 022808 (2014).
[60] A. Koher, H. H. K. Lentz, J. P. Gleeson, and P. Hövel, Phys. Rev. X 9, 031017 (2019).
Appendix A: Supplementary figures and table

FIG. A1. Dependence of differential invariants of a given motif on motifs of one order higher due to nearest-neighbour interactions, up to order 4 (ignoring labels).

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

TABLE A1. 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 algebraic invariants (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 algebraic invariants (if $n = 2$).
FIG. A2. Steady state correlation function $C^{D}_{II}$ versus distance $D$ [see (52)] for simulations of the SIS model on the square lattice (lines). Colour scale indicates values of $\beta/\gamma$ consistent with the markers in Figure 2 and sort such that higher correlations occur for lower values of $\beta/\gamma$ (closer to the critical point). Markers show approximated correlations when assuming MF2 conditions hold (using (53) and factorising 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).

Appendix B: Closure examples

The table in this pdf file shows 13 examples of subgraph decompositions based on the method explained in Section IV. By column, it shows: 1. the example number, 2. the considered subgraph, 3. its diameter, 4. its independence map assuming independence beyond distance $\text{diam}(a - 1)$, 5. whether the independence map is chordal, 6. the derived junction graph of $d$-cliques where $d = \text{diam}(a - 1)$, 7. the resulting closure formula, 8. a triangulation of the independence map if the independence map is non-chordal, 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 non-maximal cliques for some subgraphs, 12. the resulting closure formula based on this extension, 13. comments. Our Mathematica script generates these closures automatically.
Appendix C: Differential invariants for the SIS model up to order 3

\[
\frac{d}{dt} [\bullet] = \gamma [\bullet] - \beta [\bullet \bullet], \quad (A1)
\]

\[
\frac{d}{dt} [\bullet] = -\gamma [\bullet] + \beta [\bullet \bullet], \quad (A2)
\]

\[
\frac{d}{dt} [\bullet \bullet] = 2\gamma [\bullet \bullet] - 2\beta [\bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet], \quad (A3)
\]

\[
\frac{d}{dt} [\bullet \bullet] = \gamma [\bullet \bullet] - (\beta + \gamma) [\bullet \bullet] + \beta [\bullet \bullet \bullet] + \beta [\bullet \bullet \bullet] - \beta [\bullet \bullet \bullet] - \beta [\bullet \bullet \bullet], \quad (A4)
\]

\[
\frac{d}{dt} [\bullet \bullet] = 2\beta [\bullet \bullet] - 2\gamma [\bullet \bullet] + 2\beta [\bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet], \quad (A5)
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet] = 3\gamma [\bullet \bullet \bullet] - 3\beta [\bullet \bullet \bullet] - 6\beta [\bullet \bullet \bullet] - 3\beta [\bullet \bullet \bullet], \quad (A6)
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet] = 2\gamma [\bullet \bullet \bullet] - (2\beta + \gamma) [\bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet] + \beta [\bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet]
\quad - 2\beta [\bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet] + \beta [\bullet \bullet \bullet], \quad (A7)
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet] = \gamma [\bullet \bullet \bullet] - 2(\beta + \gamma) [\bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet] - \beta [\bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet]
\quad + 2\beta [\bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet] - \beta [\bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet], \quad (A8)
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet \bullet] = -3\gamma [\bullet \bullet \bullet \bullet] + 6\beta [\bullet \bullet \bullet \bullet] + 3\beta [\bullet \bullet \bullet \bullet] + 6\beta [\bullet \bullet \bullet \bullet] + 3\beta [\bullet \bullet \bullet \bullet], \quad (A9)
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet \bullet] = \gamma [\bullet \bullet \bullet \bullet] + 2\gamma [\bullet \bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet \bullet] - \beta [\bullet \bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet \bullet] - 4\beta [\bullet \bullet \bullet \bullet]
\quad - 3\beta [\bullet \bullet \bullet \bullet], \quad (A10)
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet \bullet] = 2\gamma [\bullet \bullet \bullet \bullet] - (2\beta + \gamma) [\bullet \bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet \bullet]
\quad + 2\beta [\bullet \bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet], \quad (A11)
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet \bullet] = \gamma [\bullet \bullet \bullet \bullet] + \gamma [\bullet \bullet \bullet \bullet] - (\beta + \gamma) [\bullet \bullet \bullet \bullet] - \beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet] - \beta [\bullet \bullet \bullet \bullet]
\quad - 2\beta [\bullet \bullet \bullet \bullet] - \beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet] - \beta [\bullet \bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet], \quad (A12)
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet \bullet] = \gamma [\bullet \bullet \bullet \bullet] - (\beta + 2\gamma) [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet] - \beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet]
\quad + \beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet] - \beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet], \quad (A13)
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet \bullet] = \gamma [\bullet \bullet \bullet \bullet] - 2(\beta + \gamma) [\bullet \bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet \bullet] - \beta [\bullet \bullet \bullet \bullet] - 2\beta [\bullet \bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet \bullet]
\quad + 2\beta [\bullet \bullet \bullet \bullet] - \beta [\bullet \bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet \bullet], \quad (A14)
\]

\[
\frac{d}{dt} [\bullet \bullet \bullet \bullet] = -3\gamma [\bullet \bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet \bullet]
\quad + 2\beta [\bullet \bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet \bullet] + \beta [\bullet \bullet \bullet \bullet] + 2\beta [\bullet \bullet \bullet \bullet]. \quad (A15)
\]
Appendix D: Higher order mean-field equations for the SIS model

1. General unclosed fourth-order

2. Closed fourth-order for square lattice