Chapter 4
Mean-field approximations for homogeneous networks

As seen in Chapters 2 and 3, because of the high-dimensionality of exact mathematical models describing spreading processes on networks, the models are often neither tractable nor numerically solvable for networks of realistic size. We can avoid this fundamental difficulty by refocusing our attention on expected population-scale quantities, such as the expected prevalence or the expected number of edges where one node is susceptible while the other is infectious. This opens up a range of possibilities to formulate so-called mean-field models (i.e. typically low-dimensional ODEs or PDEs) that are widespread in the physics and mathematical biology literature. These are used to approximate stochastic processes, with the potential to be exact in the large system or "thermodynamic" limit, see (for example, [151, 166, 167, 243] and the literature overview in Section 4.8).

Figure 4.1 shows the outcomes of many stochastic realisations of SIS and SIR epidemics on an Erdős–Rényi network with average degree $\langle K \rangle = 30$. The realisations form the cloud plot, which highlights the stochasticity of the spreading processes. The same figure also shows the average prevalence based on all realisations together with predictions of the simplest mean-field approximations, system (4.8) for the SIS case and (4.9) for the SIR case. The derivation and analysis of mean-field models capable of describing such average behaviours is the focus of a significant part of this book. Many such models have been and can be derived. The simplest models typically require stronger assumptions. We set out to present these models in a unified framework, highlighting the motivation and applicability of these models. Some of these models may be network- or process-specific and only perform well for certain networks or processes, e.g. giving good results for SIR dynamics but not for SIS.

The typical recipe for deriving a mean-field model is to first identify some quantities of practical interest and derive equations for how the average value of these quantities change in time. Often, these equations rely on new variables. We iterate the process, deriving equations for the new variables. Occasionally, this process terminates quickly, yielding a small, self-consistent system of equations. Other times,

I.Z. Kiss et al., *Mathematics of Epidemics on Networks*, Interdisciplinary Applied Mathematics 46, DOI 10.1007/978-3-319-50806-1_4

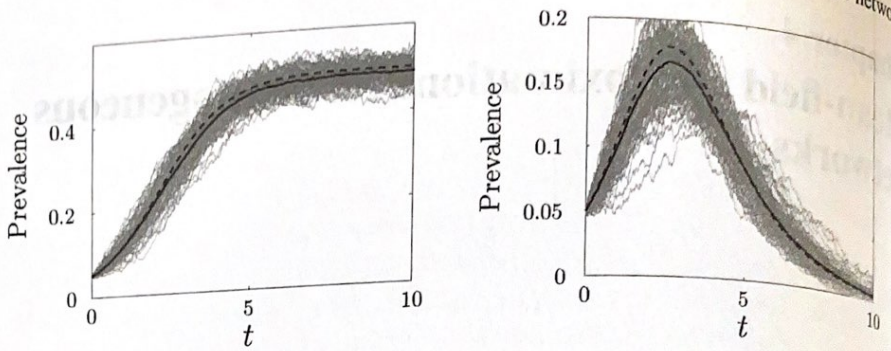


Fig. 4.1: The time dependence of the expected number of infected nodes for an Erdős–Rényi random graph from 100 agent-based stochastic simulations (grey curves) and from the mean-field system closed at the level of pairs (black dashed curves) for (a) an SIS epidemic and (b) an SIR epidemic. The averages of the simulations are shown with black solid curves. The parameter values are $N = 1000$, $\langle K \rangle = 30$, recovery rate $\gamma = 1$ and infection rate $\tau = 1/15$. We note that as the number of nodes increases, the spread of simulation around the mean decreases.

the iteration would continue until the full system size is reached; then, we typically "close" the model by approximating the newly introduced variables in terms of already existing ones. There is some "art" in the process: sometimes, there are options for what variables or approximations to use, and the choice determines the accuracy and complexity of the model.

This chapter and the following chapter systematically develop mean-field models, starting with the most basic ones that usually work for simpler networks, e.g. regular and Erdős–Rényi, and moving towards the more sophisticated models that can handle networks with high degree heterogeneity (which are the focus of Chapter 5), preferential mixing or clustering. We present results concerning steady states, stability and model consistency for many mean-field models to aid the readers' understanding of this important area. We note that these models have been extended to account for more complex structural properties of networks, such as households and networks with arbitrary subgraph distribution as well as directed or weighted networks. These will not be dealt with in the book, but relevant references are given in the concluding section of this chapter.

4.1 Exact, unclosed models

For models of infectious disease, the variables of interest are typically the average number of nodes having a given status. We denote these by $[S](t)$, $[I](t)$ and $[R](t)$ for susceptible, infected and recovered, respectively. The expected number of edges connecting nodes of status A to nodes of status B is denoted by $[AB](t)$, with $A, B \in \{S, I, R\}$. The expected number of triples, where a middle B node is connected to

a node of status A and to a node of status C, is denoted by $[ABC](t)$. A formal definition of these quantities is given in the next subsection. Note that there is an implicit direction, so each edge is counted twice. An edge from a susceptible node to an infected node is counted towards $[SI]$, but it will also be counted towards $[IS]$.

4.1.1 The variables of mean-field models: population-level counts

To define the population-level counts, we start from the random variable $X_i(t)$ that determines the type of node i at time t, e.g. $X_i(t) = I$ if node i is infected at time t. The above expected values can be defined formally as

$$[A](t) = \sum_{i=1}^{N} P(X_i(t) = A),$$

where $A \in \{S,I,R\}$. The expected number of edges in a given status can be defined similarly as follows:

$$[AB](t) = \sum_{i=1}^{N}\sum_{j=1}^{N} g_{ij} P(X_i(t) = A, X_j(t) = B),$$

with $A,B \in \{S,I,R\}$. It is important to note here that edges connecting two susceptible nodes contribute twice to the $[SS]$ count, since the pairs (i,j) and (j,i) are both counted in the $[SS]$ class when nodes i and j are susceptible. As an example, consider the graph with two infected and two susceptible nodes, shown in Fig. 4.2. The nodes are labelled clockwise from the top left, with nodes 1 and 2 being infected. Let us for a moment assume that this is a snapshot of a given realisation of the epidemic. We determine the different counts denoting them also by $[\cdot]$, although these are not expectations. In the situation given in Fig. 4.2, we have $[I] = 2$ and $[S] = 2$. Counting the pairs, we find that $(3,2)$, $(4,2)$ and $(4,1)$ are of SI type and, hence $[SI] = 3$. The II pairs are $(1,2)$ and $(2,1)$; therefore, $[II] = 2$. In a similar way, we get $[IS] = 3$ and $[SS] = 2$; the total number of pairs is 10. Let us finally turn to the counting of triples. The number of ABC triples is defined in general as

$$[ABC](t) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} g_{ij} g_{jk} P(X_i(t) = A, X_j(t) = B, X_k(t) = C),$$

with $A,B,C \in \{S,I,R\}$. Using this definition, the number of SSI triples in the graph shown in Fig. 4.2 is $[SSI] = 3$, namely the SSI triples are $(3,4,1)$, $(3,4,2)$ and $(4,3,2)$. The number of other triples can be similarly obtained as $[ISS] = 3$, $[IIS] = 3$, $[SII] = 3$, $[ISI] = 2$ and $[SIS] = 2$; there are 16 triples altogether.

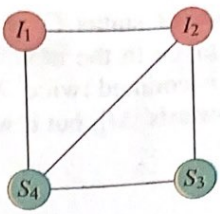


Fig. 4.2: Counting the number of pairs and triples.

The expected values introduced above obey some conservation relations. The simplest one, for the SIS dynamics, is based on the fact that $P(X_i(t) = S) + P(X_i(t) = I) = 1$ for any $i = 1,2,\ldots,N$, since any node is either susceptible or infected. This immediately implies that

$$[S](t) + [I](t) = N$$

for any time instant t. For the SIR epidemic, the corresponding equation is $[S](t) + [I](t) + [R](t) = N$. The conservation relation for the pairs is based on the simple fact that for the SIS epidemic, a pair can be in one of the following four statuses: SS, SI, IS or II. Hence, $P(X_i(t) = S, X_j(t) = S) + P(X_i(t) = S, X_j(t) = I) + P(X_i(t) = I, X_j(t) = S) + P(X_i(t) = I, X_j(t) = I) = 1$, implying

$$[SS](t) + [SI](t) + [IS](t) + [II](t) = \sum_{i=1}^{N} \sum_{j=1}^{N} g_{ij} = Nn,$$

where $n = \langle k \rangle$ is the average degree defined in Section 1.2.2. We note that n is used to denote the average degree in this chapter for the sake of brevity and to follow the widely used notation of pairwise models. The definition of pairs immediately implies that $[SI] = [IS]$; hence, the above relation can be formulated as

$$[SS](t) + 2[SI](t) + [II](t) = Nn.$$

Similar arguments lead to further pair conservation relations in the SIS case, namely

$$[SS](t) + [SI](t) = n_S(t)[S](t), \quad [SI](t) + [II](t) = n_I(t)[I](t),$$

where $n_S(t)$ and $n_I(t)$ denote the average degree of susceptible and infected nodes, respectively. The pair conservation relations in the SIR case can be formulated similarly. Returning to the SIS case, triple conservation relations can also be formulated. The relation

$$[SSI](t) + [ISI](t) = (n_S(t) - 1)[SI](t)$$

will play an important role. Besides deriving it formally from the triple definitions, one can argue as follows to prove the relation. Taking an arbitrary SI edge, the S node has $(n_S - 1)$ further neighbours; hence, the total number of triples containing an SI pair is $(n_S - 1)[SI]$. On the other hand, the same quantity can be obtained as the sum of those triples that contain an SI pair, namely $[SSI] + [ISI]$. In a similar way, one can derive further triple conservation relations, for example, $[SSS](t) + [ISS](t) = (n_S(t) - 1)[SS](t)$ and $[SIS](t) + [IIS](t) = (n_I(t) - 1)[IS](t)$.

4.1.2 Exact differential equations for the singles and pairs

We use heuristic arguments here to derive exact differential equations for the expected number of nodes and edges in given statuses. Later in this chapter, we derive these from the master equations, system (2.6), consisting of 2^N equations governing the full SIS dynamics. We sacrifice having information about the precise statuses of each node to get a much smaller system of equations for the expected numbers of nodes in each status. The main parameters of the epidemic processes are the infection and recovery rates denoted by τ and γ, respectively.

Theorem 4.1 *For the SIS epidemic on an arbitrary network (undirected and not weighted), the expected values $[S]$ and $[I]$ satisfy the following system*

$$[\dot{S}] = \gamma[I] - \tau[SI], \tag{4.1a}$$

$$[\dot{I}] = \tau[SI] - \gamma[I]. \tag{4.1b}$$

The proof of the theorem is presented in Section 4.6. The differential equations can be obtained heuristically following the top two flow diagrams in Fig. 4.3. The rate of transmission to S nodes is τ times the number of SI edges. The rate of recovery of infectious nodes back into a susceptible status is γ times the number of infectious nodes. Thus, the rate of change of $[S]$ is $\gamma[I] - \tau[SI]$. We similarly find the $[I]$ equation. For an SIR epidemic, the I nodes do not become susceptible again; instead their status changes to R. The equations take a different form.

Theorem 4.2 *For the SIR epidemic on an arbitrary network (undirected and not weighted), the expected values $[S]$, $[I]$ and $[R]$ satisfy the following system*

$$[\dot{S}] = -\tau[SI], \tag{4.2a}$$

$$[\dot{I}] = \tau[SI] - \gamma[I], \tag{4.2b}$$

$$[\dot{R}] = \gamma[I]. \tag{4.2c}$$

We note that the variables in the SIS and SIR systems thus far are not independent because of conservation relations, taking the form $[S] + [I] = N$ in the SIS case and $[S] + [I] + [R] = N$ in the SIR case. Hence, for the exact systems one of the equations could be omitted in both cases. However, for systems with approximate closures, the conservation laws do not hold automatically; their validity may depend on the choice of the closure. Therefore, differential equations for all singles and pairs that are needed to get a self-contained system will be considered. This will always be followed by investigating whether conservation laws hold as they enable us to reduce the system.

As illustrated in Fig. 4.3, the dynamics of the expected number of S nodes ($[S]$) and I nodes ($[I]$), i.e. singles, depends on the number of SI pairs ($[SI]$); hence, the system depends on pairs, for which we need additional equations. Similarly, the number of pairs depends on the number of triples. For example, the number of SS pairs decreases due to infection from outside the pair, i.e. it changes proportionally

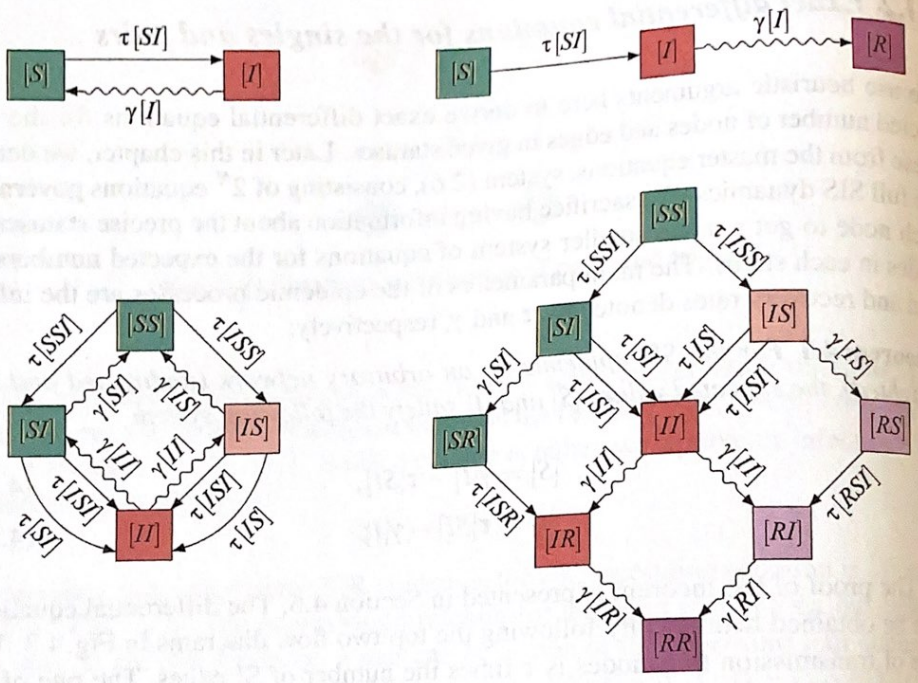


Fig. 4.3: Flow diagrams showing the flux between compartments of singles (top) and compartments of pairs (bottom). The SIS case is on the left and SIR on the right. In the compartments of pairs, solid lines denote infections coming from within the pair (with a rate depending on a pair) or from outside the pair (with a rate depending on a triple), and the wiggle lines denote a recovery. The colour indicates the status of the "first" node in the edge. Symmetry allows us to conclude that some of the variables (lighter shade, on the right of each diagram) must equal the symmetric version (e.g. $[RS] = [SR]$), so we do not need to directly calculate both.

to the number of SSI triples ($[SSI]$) with rate $2\tau[SSI]$ (recall the SS pair is counted twice). In an SI or IS pair, the infected node can recover with rate γ; hence, in the case of an SIS epidemic, the number of SS pairs increases at rate $\gamma([SI]+[IS])$. Since the numbers of SI and IS pairs are equal, we can use $2\gamma[SI]$. Extending this simple heuristic reasoning to SI and II pairs and by accounting for all within- and outside-pair transitions, we arrive at the following theorems.

Theorem 4.3 *For the SIS epidemic on an arbitrary network (undirected and not weighted) the expected values of* $[S]$, $[I]$, $[SI]$, $[II]$, *and* $[SS]$ *satisfy the following system of differential equations*

$$[\dot{S}] = \gamma[I] - \tau[SI], \tag{4.3a}$$

$$[\dot{I}] = \tau[SI] - \gamma[I], \tag{4.3b}$$

$$[\dot{SI}] = \gamma([II] - [SI]) + \tau([SSI] - [ISI] - [SI]), \tag{4.3c}$$

$$\dot{[SS]} = 2\gamma[SI] - 2\tau[SSI], \tag{4.3d}$$

$$\dot{[II]} = -2\gamma[II] + 2\tau([ISI] + [SI]). \tag{4.3e}$$

This result can also be derived directly from the master equations, system (2.6) (see [299]).

As before, for SIR epidemics the I nodes do not become susceptible again; hence, the terms due to recovery show up in different equations. We find the following.

Theorem 4.4 *For the SIR epidemic on an arbitrary network (undirected and not weighted), the expected values of* $[S]$, $[I]$, $[SI]$ *and* $[SS]$ *satisfy the following system of differential equations:*

$$\dot{[S]} = -\tau[SI], \tag{4.4a}$$

$$\dot{[I]} = \tau[SI] - \gamma[I], \tag{4.4b}$$

$$\dot{[R]} = \gamma[I], \tag{4.4c}$$

$$\dot{[SI]} = -\gamma[SI] + \tau([SSI] - [ISI] - [SI]), \tag{4.4d}$$

$$\dot{[SS]} = -2\tau[SSI], \tag{4.4e}$$

The differential equations for $[II]$, $[SR]$, $[IR]$ and $[RR]$ can be formulated similarly, but we leave them out because the other variables do not depend on them and they are not generally of epidemiological interest.

4.2 Closures at the pair and triple level and the resulting models

We now present the ideas leading to closures of the systems above.

4.2.1 Closures

We now assume that the network is homogeneous: each node has the same degree n. There are $[I]$ infected nodes, making up a proportion $[I]/N$ of the population. Assuming that infected nodes are distributed randomly, an average susceptible node has $n[I]/N$ infected neighbours. This assumption makes the closed system inexact since infected nodes are more likely to be in contact with other infected nodes because of how infection propagates. Using this assumption, however, the total number of SI edges is approximated by

$$[SI] \approx \frac{n}{N}[S][I]. \tag{4.5}$$

This relation is referred to as a pair closure since it replaces a term for a pair to yield a "closed" system of equations. Using this, the dependence on higher order moments in system (4.1) is avoided. This closure is used also for the SIR epidemic

in system (4.2). We note that sometimes $\frac{n}{N-1}$ is used instead of $\frac{n}{N}$, because after choosing a susceptible node, the remaining population contains only $N-1$ nodes.

To improve our accuracy, we need a better accounting of the fact that infected nodes are not uniformly distributed. We take system (4.3) and close it by assuming an algebraic expression for $[SSI]$ and for $[ISI]$ in terms of singles and pairs. To derive this, we start again with a susceptible node and determine what proportion of the edges starting from this node lead to infected or susceptible nodes. The total number of edges starting from susceptible nodes is $n[S]$. The total number of SI edges is $[SI]$; hence, a proportion $[SI]/n[S]$ of the edges starting from susceptible nodes lead to infected nodes. Similarly, the ratio of edges leading to susceptible nodes is $[SS]/n[S]$. Thus, if we choose a susceptible node u and two neighbours v and w (arbitrarily calling v "first"), the probability that v is susceptible and w infected is $[SS][SI]/n^2[S]^2$. There are $n(n-1)$ ways to choose v and w. Thus, the expected number of SSI triples is $[S]n(n-1)[SS][SI]/n^2[S]^2 = (n-1)[SS][SI]/n[S]$. Assuming that these are uniformly distributed, we conclude that

$$[SSI] \approx \frac{n-1}{n}\frac{[SS][SI]}{[S]}. \tag{4.6}$$

Similar analysis leads to the closure

$$[ISI] \approx \frac{n-1}{n}\frac{[SI]^2}{[S]}. \tag{4.7}$$

As before, this is only an approximation: for an SIS epidemic, a previously infected, but recently recovered node will tend to have more infected neighbours than other susceptible nodes. Thus, the distribution of infected neighbours to susceptible nodes is not truly uniform. Further, for an SIR epidemic, if there are many short cycles, then one neighbour of a susceptible node u being infected is correlated with other neighbours of u being infected, so again the distribution is not uniform. If there are few short cycles, we will see that this closure is accurate for SIR epidemics.

4.2.2 *Closed systems*

We now write out the closed systems. We first apply closure (4.5) to system (4.1). Since the closure is only approximate, the variables in the closed system are, strictly speaking, different from the original ones, so we use a different notation: $[S]_f$ and $[I]_f$ are the approximations given by the closed system.

SIS homogeneous mean-field at single level

$$[\dot{S}]_f = \gamma[I]_f - \tau\frac{n}{N}[S]_f[I]_f, \tag{4.8a}$$

$$[\dot{I}]_f = \tau\frac{n}{N}[S]_f[I]_f - \gamma[I]_f. \tag{4.8b}$$

The subscript "f" refers to "first" since this can be considered as the first approximation when the expected number of pairs is expressed in terms of the expected number of singles. The system and variables resulting by applying the triple closure will be referred to as the second approximation. In the case of the SIR epidemic, the simplest closed system takes the form below.

SIR homogeneous mean-field at single level

$$[\dot{S}]_f = -\tau \frac{n}{N}[S]_f[I]_f, \tag{4.9a}$$

$$[\dot{I}]_f = \tau \frac{n}{N}[S]_f[I]_f - \gamma[I]_f, \tag{4.9b}$$

$$[\dot{R}]_f = \gamma[I]_f. \tag{4.9c}$$

It is worth noting that these closures lead to the usual SIS and SIR disease models. These equations are an approximation for epidemics on static networks. However, if we consider an alternate problem, in which nodes select a new set of random neighbours at each moment, our assumption that $[SI] \approx \frac{n}{N}[S][I]$ is correct. So these equations are correct if we consider a system in which nodes are quickly changing their neighbours. Thus, the difference between these closed equations and the full equations is due to the duration of partnerships creating correlations between the status of neighbouring nodes.

We now apply the triple closures (4.6) and (4.7) to system (4.3). This is the second approximation, so we use the notation $[S]_s$, $[I]_s$, $[SI]_s$, $[SS]_s$ and $[II]_s$, yielding the system below.

SIS homogeneous pairwise

$$[\dot{S}]_s = \gamma[I]_s - \tau[SI]_s, \tag{4.10a}$$

$$[\dot{I}]_s = \tau[SI]_s - \gamma[I]_s, \tag{4.10b}$$

$$[\dot{SI}]_s = \gamma([II]_s - [SI]_s) + \tau \frac{n-1}{n} \frac{[SI]_s([SS]_s - [SI]_s)}{[S]_s} - \tau[SI]_s, \tag{4.10c}$$

$$[\dot{SS}]_s = 2\gamma[SI]_s - 2\tau \frac{n-1}{n} \frac{[SI]_s[SS]_s}{[S]_s}, \tag{4.10d}$$

$$[\dot{II}]_s = -2\gamma[II]_s + 2\tau \frac{n-1}{n} \frac{[SI]_s^2}{[S]_s} + 2\tau[SI]_s. \tag{4.10e}$$

We will show later that this system conserves certain quantities; hence, we do not need all of the differential equations to determine the number of susceptible and in-

fected nodes. In the next section, it will be shown that two differential equations are enough to determine all the unknown functions, because there is one conservation relation for singles and two for pairs.

For SIR epidemics, the same closures are applied to system (4.4), leading to the following.

SIR homogeneous pairwise

$$[\dot{S}]_s = -\tau[SI]_s, \tag{4.11a}$$

$$[\dot{I}]_s = \tau[SI]_s - \gamma[I]_s, \tag{4.11b}$$

$$[\dot{R}]_s = \gamma[I]_s, \tag{4.11c}$$

$$[\dot{SI}]_s = -\gamma[SI]_s + \tau\frac{n-1}{n}\frac{[SI]_s([SS]_s - [SI]_s)}{[S]_s} - \tau[SI]_s, \tag{4.11d}$$

$$[\dot{SS}]_s = -2\tau\frac{n-1}{n}\frac{[SI]_s[SS]_s}{[S]_s}, \tag{4.11e}$$

The system could be augmented by further equations if one had a reason to be interested in the values of pairs such as $[II]$, $[SR]$, $[IR]$ or $[RR]$.

When the closed systems are solved, initial conditions are needed for all model variables. Typically, the initial number of susceptible, infected and recovered nodes are given; these can be used as initial conditions in the mean-field models at single level. In the case of pairwise models, further initial conditions are needed for the initial number of pairs. Assuming that the different types of nodes are distributed randomly initially, the initial condition for AB pairs can be given as

$$[AB]_0 = \frac{n}{N}[A]_0[B]_0,$$

where $[A]_0$ and $[B]_0$ denote the initial number of nodes of status A and B, respectively. Obviously, this is not the only choice for the initial conditions of the pairs. For example, correlations in the initial position of infected nodes can be built in, but these will dampen in time.

4.2.3 *Clustered pairwise model*

The closures above fail to describe clustered graphs in a satisfactory way. The clustering coefficient of a network is simply the ratio of triangles to triples (both open and closed). More formally, it can be defined by using the adjacency matrix G of the network. Let the network be undirected without weights; that is, G is symmetric and each entry is 0 or 1. For a matrix A, we will use the notations $\|A\| = \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}$ and $\mathrm{Tr}(A) = \sum_{i=1}^{N} a_{ii}$ for the trace of the matrix. The number of pairs, i.e. edges in

our terminology, is $\|G\|$. It can be shown that the number of triples is $\|G^2\| - \mathrm{Tr}(G^2)$ and the number of triangles is $\mathrm{Tr}(G^3)$. The clustering coefficient of the network is defined as the ratio of the number of triangles and triples:

$$\phi = \frac{\mathrm{Tr}(G^3)}{\|G^2\| - \mathrm{Tr}(G^2)}.$$

In Chapter 3, we have shown that the bottom-up approach can handle certain classes of clustered networks, but the number of resulting equations can be forbidding. Hence, it seems more natural to consider mean-field and percolation models to capture this extra degree of complexity in the structure of the network. Below, we give a succinct overview of some extensions. In the spirit of this chapter, we start with extensions of the pairwise model. Here, we present two closures that attempt to capture clustering within the framework of pairwise models. The classical closure for clustered graphs was first introduced in [166, 256], the idea of which will be presented now. Take an SI edge and consider the $(n-1)$ other neighbours of the susceptible node. The average number of those neighbours that are not connected to the infected node is $(1-\phi)(n-1)$, while the average number of the neighbours connected to it is $\phi(n-1)$. For those that are not connected to the I node, we can apply the original idea of closure, i.e. we can say that the proportion of S neighbours is $[SS]/n[S]$, and hence the number of SSI triples given by these neighbours is

$$[SI](1-\phi)(n-1)\frac{[SS]}{n[S]} = (1-\phi)\frac{n-1}{n}\frac{[SS][SI]}{[S]}.$$

Consider now those neighbours of the S node that are connected to the I node as well. The proportion of susceptible nodes among these neighbours is scaled with the correlation $C_{SI} = \frac{N}{n}\frac{[SI]}{[S][I]}$, leading to

$$[SI]\phi(n-1)\frac{[SS]}{n[S]}C_{SI} = \phi\frac{n-1}{n}\frac{[SS][SI]}{[S]}\frac{N}{n}\frac{[SI]}{[S][I]}.$$

Thus, adding the above expressions we approximate the total number of SSI triples in a network with clustering coefficient ϕ as

$$[SSI] \approx \frac{n-1}{n}\frac{[SS][SI]}{[S]}\left((1-\phi)+\phi\frac{N}{n}\frac{[SI]}{[S][I]}\right).$$

By introducing C_{II} in the same manner, the number of ISI triples can be similarly approximated as

$$[ISI] \approx \frac{n-1}{n}\frac{[SI]^2}{[S]}\left((1-\phi)+\phi\frac{N}{n}\frac{[II]}{[I]^2}\right).$$

Intuitively, the scaling factors or correlations measure the propensity of nodes with certain statuses to be more or less likely to be neighbours than if they were randomly distributed in the network. The performance of these closure relations can

be investigated by comparison to simulations. However, there is a theoretical issue concerning them, namely, the closed systems given by these relations do not preserve the conservation relations at the level of pairs. That is, even if $[SS]+[SI]=n[S]$ and $[SI]+[II]=n[I]$ initially, these equations fail later if the closures are used. These discrepancies can be fixed by using a modified scaling factor instead of the correlation C_{SI}. In [149], the scaling factor

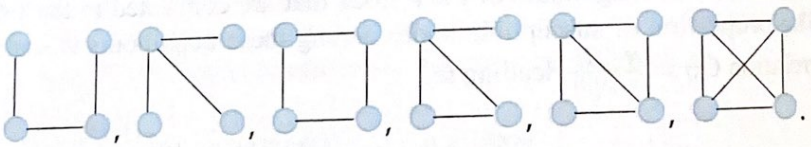
$$\frac{C_{AI}}{p_{S|S}C_{SI}+p_{I|S}C_{II}}$$

is introduced, where $p_{A|S}=\frac{[AS]}{n[S]}$ for $A\in\{S,I\}$. This yields the closure approximation

$$[ASI]\approx\frac{n-1}{n}\frac{[AS][SI]}{[S]}\left(1-\phi+\phi\frac{n[S][I][AI]}{([SS][I]+[S][II])[A][SI]}\right).$$

Applying these closure relations in system (4.3), we obtain a closed clustered pairwise model for SIS epidemic dynamics. Similarly, the closures can be used in system (4.4) to get a clustered pairwise model for SIR epidemic dynamics. Note that system (4.4) must be augmented by an equation for $[II]$ in order to apply this closure.

Other developments in this area include the continuation of the pairwise equations beyond pairs by writing down equations for triples, which now will include quadruples [152]. The complexity grows since triples themselves can be open or closed, and quadruples can be of the following types:



The network structure is accounted for by writing down closures separately for each quadruple type, hoping to more accurately capture the local structure. The performance of this model seems to be mildly better than that of the simple pairwise; however the exploration of this model is far from trivial as it relies on a good parametrisation of the network model, where quadruples can be controlled well. The dimensionality of such a model increases and therefore the insight gained may be less compared to that gained from simpler models.

Clustering can have a significant impact on the epidemic threshold, final epidemic size or endemic state. For example, it is widely accepted that the value of the transmission rate needed to generate an epidemic is larger for networks which are clustered when compared to an equivalent network with the same degree distribution but no clustering. Where neighbours of a given node are likely to be connected leads to local depletion of susceptibles where transmission events are effectively wasted [122, 166, 213, 214].

Perhaps the most significant progress in this direction has been made by using a percolation theory approach [117, 213, 237]. However, the analytical tractability of these models requires the consideration of specific classes of networks. In this

case, the networks usually are built based on allocating a number of triangle corners or hyperstubs to each node according to some degree distribution, which later on, with some probability, will either decompose into two classical edges or will go on to form triangles. This construction allowed the authors to use the probability generating function machinery coupled with percolation theory (see Chapter 6). This then led to finding analytical expressions for the size of the giant component and the location of the percolation threshold.

Several further developments of such models have been proposed. First of all, Karrer and Newman [164] have extended the network models to take into account arbitrary distributions of subgraphs and investigate the size of the giant component, the location of the phase transition at which the giant component appears, and percolation properties for both site and bond percolation on networks generated by the model. On the other hand, two other significant developments can be noted. In [318], the authors have successfully extended the edge-based compartmental model (EBCM) framework of Chapter 6 to provide a compact mean-field model that gives excellent agreement with simulations and accurately describes the temporal evolution of the disease. Finally, combining the results in [164] and [318], Ritchie et al. [265] extended the EBCM framework further to model SIR epidemics on graphs with arbitrary subgraph distributions. This is also a mean-field model that describes the temporal evolution of the epidemic and provides a strong framework for modelling SIR disease spread on clustered networks.

4.3 Analysis of the closed systems

4.3.1 SIS homogeneous mean-field equations at the single level

Consider first the SIS system closed at the level of pairs, i.e. system (4.8). Note that adding the two equations, we get that $[S]_f(t)+[I]_f(t)$ is constant in time. If the initial condition satisfies $[S]_f(0)+[I]_f(0)=N$, then by using $[S]_f(t)=N-[I]_f(t)$, the system can be reduced to the single equation

$$[\dot{I}]_f = \tau\frac{n}{N}(N-[I]_f)[I]_f - \gamma[I]_f.$$

We can analyse the behaviour of this dynamical system. It has a disease-free steady state $[I]_f^{df}=0$ and another fixed point at $[I]_f^e = N(1-\frac{\gamma}{n\tau})$. This second point is only biologically meaningful if it is positive, i.e. $\gamma<n\tau$, in which case it gives an endemic equilibrium. Differentiating the right-hand side of the differential equation with respect to $[I]_f$, i.e. linearising the right-hand side, one obtains that for $\gamma>n\tau$ the disease-free state is stable, while for $\gamma<n\tau$ the endemic state is stable. Thus at $\gamma=n\tau$ a transcritical bifurcation occurs, the disease-free steady state loses its stability, and the stable endemic steady state appears. Moreover, these stabilities are global for $N>[I]_f>0$. For $\gamma>n\tau$, the right-hand side is negative, i.e. $[I]_f$ is decreasing in time and converges to zero, starting from any meaningful initial condition. In