Chapter 6
Community Networks with Equitable Partitions

Stefania Ottaviano, Francesco De Pellegrini, Stefano Bonaccorsi, Delio Mugnolo and Piet Van Mieghem

A community structure is an important non-trivial topological feature of a complex networks. Indeed community structures are a typical feature of social networks, tightly connected groups of nodes in the World Wide Web usually correspond to pages on common topics, communities in cellular and genetic networks are related to functional modules [46].

Thus, in order to investigate this topological feature, we consider, in this chapter, that the entire population is partitioned into communities (also called households, clusters, subgraphs, or patches). There is an extensive literature on the effect of network community structures on epidemics. Models utilizing this structure are commonly known as "metapopulation" models (see, e.g., [10, 125, 178]). Such models

Parts of this chapter have been written based on a thesis archived on Open Archives Initiative. http://eprints-phd.biblio.unitn.it/1684/.

S. Ottaviano · S. Bonaccorsi
University of Trento, Trento, Italy
e-mail: stefania.ottaviano@unitn.it

S. Bonaccorsi
e-mail: stefano.bonaccorsi@unitn.it

F. De Pellegrini (✉)
University of Avignon, Avignon, France
e-mail: francesco.de-pellegrini@univ-avignon.fr

D. Mugnolo
University of Hagen, Hagen, Germany
e-mail: delio.mugnolo@fernuni-hagen.de

P. Van Mieghem
Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands
e-mail: p.f.a.vanmieghem@tudelft.nl

E. Altman et al. (eds.), *Multilevel Strategic Interaction Game Models for Complex Networks*, https://doi.org/10.1007/978-3-030-24455-2_6

assume that each community shares a common environment or is defined by a specific relationship.

Some of the most common works on metapopulation regard a population divided into households with two level of mixing [31, 32, 220]. These models typically assume that contacts, and consequently infections, between nodes in the same group occur at a higher rate than those between nodes in different groups [35]. Thus, groups can be defined, e.g., in terms of spatial proximity, considering that between-group contact rates (and consequently the infection rates) depend in some way on spatial distance. In this type of heterogeneous contact networks each node can be theoretically infected by any other node. However, an underlying network contact structure, where infection can only be transmitted by node directly linked by an edge, may provide a more realistic approach for the study of the evolution of the epidemics, [35]. In turn, an important challenge is how to consider a realistic underlying structure and appropriately incorporate the influences of the network topology on the dynamics of epidemics [33, 34, 98, 212, 270, 283].

6.1 Equitable Partitions

We consider the diffusion of epidemics over an undirected graph $G = (V, E)$ with edge set E and node set V. The order of G, denoted N, is the cardinality of V, whereas the size of G is the cardinality of E, denoted L. Connectivity of the graph G is conveniently encoded in the $N \times N$ adjacency matrix A. We are interested in the case of networks that can be naturally partitioned into n communities: they are represented by a node set partition $\pi = \{V_1, ..., V_n\}$, i.e., a sequence of mutually disjoint nonempty subsets of V, called cells, whose union is V.

The dynamics of epidemic transmission is modeled by an SIS (susceptible-infected-susceptible) model, where we consider the continuous-time mean-field approximation of the exact Markovian SIS model, NIMFA [274]. We consider a curing rate δ equals for all nodes. Instead, compared to the homogeneous case, where the infection rate is the same for all pairs of nodes, in this framework we consider two infection rates: the *intra-community* infection rate β, for infecting individuals in the same community, and the *inter-community* infection rate $\varepsilon\beta$, i.e., the rate at which individuals among different communities get infected. We assume $0 < \varepsilon < 1$, the customary physical interpretation being that infection across communities occur at a much smaller rate. Indeed the presence of communities generates a strong mixing effect at local level (e.g., the rate of infection inside a community tends to be homogeneous) as opposed to the much lower speed of mixing (i.e., much larger inhomogeneity) within the whole population. Specifically, the contact newtork structure that we consider has an *equitable partition* of its node set. The original definition of equitable partition is due to Schwenk [234].

Definition 6.13 Let $G = (V, E)$ be a graph. The partition $\pi = \{V_1, ..., V_n\}$ of the node set V is called *equitable* if for all $i, j \in \{1, \ldots, n\}$, there is an integer d_{ij} such that

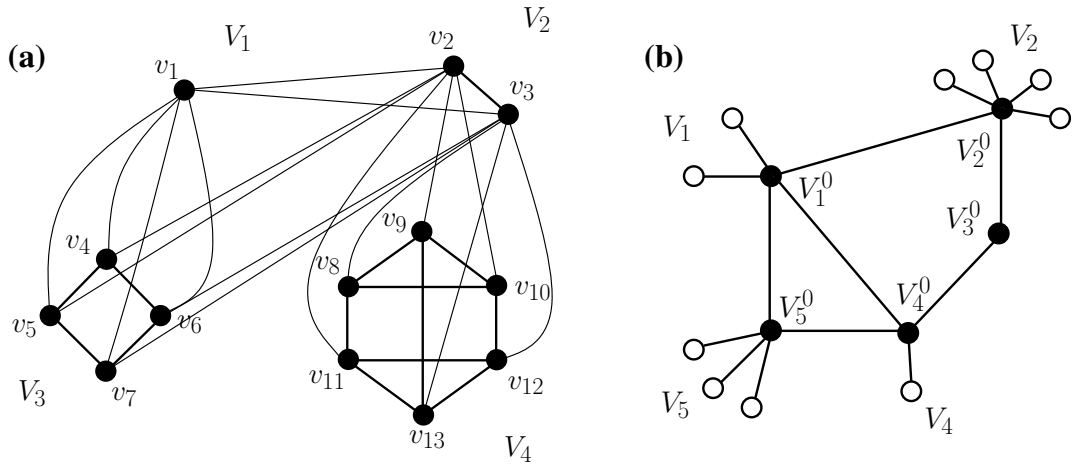

Fig. 6.1 A sample graphs with equitable partition. **a** $V = \{V_1, V_2, V_3, V_4\}$, **b** Interconnected star networks: $V = \{V_1^0, V_2^0, V_3^0, V_4^0, V_5^0, V_1, V_2, V_3, V_4, V_5\}$

$$d_{ij} = \deg(v, V_j) := \#\left\{e \in E : e = \{v, w\}, w \in V_j\right\}.$$

independently of $v \in V_i$.

We shall identify the set of all nodes in V_i with the i-th *community* of the whole population. Basically this means that all nodes belonging to the same community have the same internal degree: the subgraph G_i of $G(V, E)$ induced by V_i is regular for all i's (recall that $\pi = \{V_1, ..., V_n\}$ is a partition of the node set V, which is assumed to be given a priori). Furthermore, for any two subgraphs G_i, G_j, each node in G_i is connected with the same number of nodes in G_j (see as examples Fig. 6.1).

Thus, a network with an equitable partition of its node set posses a sort of interesting symmetry properties, where with the word symmetry we refer to a certain structural regularity of the graph connectivity. Such kind of network structure can be observed, e.g., in the architecture of some computer networks where clusters of clients connect to single routers, whereas the routers' network has a connectivity structure with nodes' degree constrained by the number of ports. An other real-word circumstance, that can be model by an equitable partition of the population, is the epidemics transmission between e.g. households, classes in a school or work offices in the same department, i.e. small communities whose members know each other. This framework can be modeled representing the internal structure of each community by a complete graph. Moreover given two connected communities, all of their nodes can be considered mutually linked, indeed each member of those two communities may potentially come into contact.

Equitable partitions appears also in the study of synchrony and pattern formation in coupled cell networks [108, 250] where they are named "balanced" partitions. Equitable partitions have been used also to analyze the controllability of multi-agent systems, for the case of a multi-leader setting [216], and for the leader-selection controllability problem, in characterizing the set of nodes from which a given net-

worked control system (NCS) is controllable/uncontrollable [7]. These works show interesting realistic scenarios for the use of equitable partitions. Since the size of some real networks might pose limitations in our ability to investigate their spectral properties, as we shall see, we can leverage on the structural regularity of network with equitable partition to reduce the dimensionality of our system.

The macroscopic structure of a network with an equitable partition of its node set can be described by a *quotient graph* G/π, which is a *multigraph* with cells $V_1, \ldots, V_n$ as vertices and $k_i d_{ij}$ edges between V_i and V_j. For the sake of explanation, in the following we will identify G/π with the (simple) graph having the same vertex set, and where an edge exists between V_i and V_j if at least one exists in the original multigraph. We shall denote by B the adjacency matrix of the graph G/π.

Remark 6.3 We use the notation lcm and gcd to denote the least common multiple and greatest common divisor, respectively. We can observe that the partition of a graph is equitable if and only if

$$d_{ij} = \alpha \frac{\text{lcm}(k_i, k_j)}{k_i}$$

where α is an integer satisfying $1 \leq \alpha \leq \gcd(k_i, k_j)$ and k_i the number of nodes in V_i, for all $i = 1, ..., n$.

Example 6.1 Let us assume that the adjacency matrix B of the quotient graph is given and that, for any $i, j \in \{1, \ldots, n\}$, $b_{ij} \neq 0$ implies $d_{ij} = k_j$, i.e., each node in V_i is connected with every node inside V_j. We can explicitly write the adjacency matrix A in a block form. Let $C_{V_i} = (c_{ij})_{k_i \times k_i}$ be the adjacency matrix of the subgraph induced by V_i and $J_{k_i \times k_j}$ is an all ones $k_i \times k_j$ matrix; then

$$A = \begin{bmatrix} C_{V_1} & \varepsilon J_{k_1 \times k_2} b_{12} & \ldots & \varepsilon J_{k_1 \times k_n} b_{1n} \\ \varepsilon J_{k_2 \times k_1} b_{21} & C_{V_2} & \ldots & \varepsilon J_{k_2 \times k_n} b_{2n} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & \ldots & C_{V_n} \end{bmatrix} \tag{6.1}$$

We observe that (6.1) represents a block-weighted version of the adjacency matrix A. The derivation of NIMFA for the case of two different infection rates, considered in this paper, results in the replacement of the unweighted adjacency matrix in the NIMFA system (6.6) with its weighted version.

A matrix smaller than the adjacency matrix A, that contains the relevant information for the evolution of the system, is associated with the quotient graph. Such a matrix is the *quotient matrix* Q of the equitable partition.

The quotient matrix Q can be defined for any equitable partition: in view of the internal structure of a graph with an equitable partition, it is natural to consider the cell-wise average value of a function on the node set, that is to say the projection of the node space into the subspace of cell-wise constant functions.

Definition 6.14 Let $G = (V, E)$ a graph. Let $\pi = \{V_i,\ i = 1, \ldots, n\}$ be any partition of the node set V, let us consider the $n \times N$ matrix $S = (s_{iv})$, where

$$s_{iv} = \begin{cases} \frac{1}{\sqrt{|V_i|}} & v \in V_i \\ 0 & \text{otherwise.} \end{cases}$$

The *quotient matrix* of G (with respect to the given partition) is

$$Q := SAS^T.$$

Observe that by definition $SS^T = I$.

In the case of equitable partitions, the expression for Q writes

$$Q = \text{diag}(d_{ii}) + (\sqrt{d_{ij}d_{ji}}\varepsilon b_{ij})_{i,j=1,\ldots n}.$$

A key feature of the model is that the spectral radius of this smaller quotient graph (which only captures the macroscopic structure of the community network) is all we need to know in order to decide whether the epidemics will go extinct in a reasonable time frame. Indeed, the spectral radius is related to the epidemic threshold of the system. NIMFA determines the epidemic threshold for the effective spreading rate β/δ as $\tau_c^{(1)} = \frac{1}{\lambda_1(A)}$, where $\lambda_1(A)$ is the spectral radius of A and the superscript (1) refers to the first-order meanfield approximation [50, 274]. Since Q and A have the same spectral radii [267, art. 62] we can compute the spectral radius of Q in order to estimate the epidemic threshold. This may lead to a significative computational advantage in the calculation of $\tau_c^{(1)}$, since the order of Q is smaller than that of A [50, Sect. 3.3].

6.1.1 Lower Bounds for the Epidemic Threshold

We can write $Q = D + \widehat{B}$, where $D = \text{diag}(d_{ii})$ and $\widehat{B} = (\sqrt{d_{ij}d_{ji}}\varepsilon b_{ij})_{i,j=1,\ldots n}$. By the Weyl's inequality [180] we have

$$\lambda_1(Q) \leq \lambda_1(D) + \lambda_1(\widehat{B}) = \max_{1\leq i\leq n} d_{ii} + \lambda_1(\widehat{B}). \tag{6.2}$$

Since

$$\tau_c^{(1)} = 1/\lambda_1(A) = 1/\lambda_1(Q),$$

a lower bound for the epidemic threshold can be derived from (6.2)

$$\tau_c^{(1)} \geq \tau^\star = \min_i \frac{1}{d_{ii} + \lambda_1(\widehat{B})}, \tag{6.3}$$

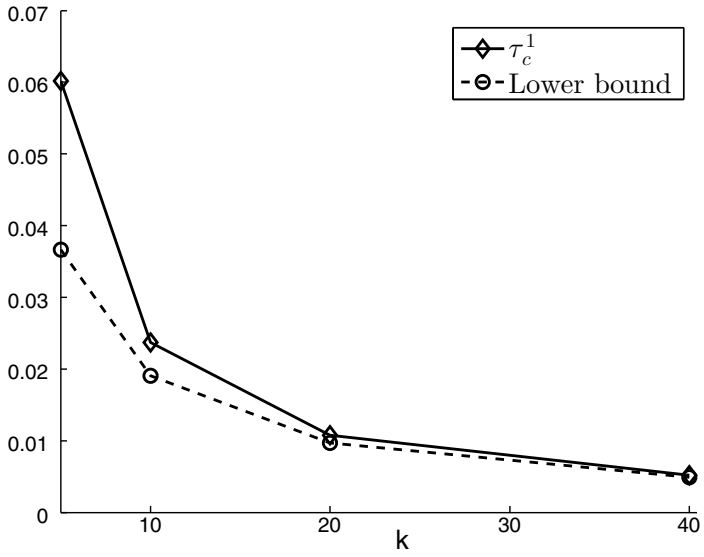

Fig. 6.2 Lower bound (6.3) versus epidemic threshold: comparison for different values of k in a 40-communities network. The internal structure of each community is a ring and $d_{ij} = 2$ for all $i, j = 1, \ldots, n$

Moreover let us note that $\lambda_1(\widehat{B}) \leq \max_i \sum_j \hat{b}_{ij}$ [181, pp. 24–26], hence

$$\tau_c^{(1)} \geq \frac{1}{\max_i (d_{ii} + \sum_j \hat{b}_{ij})}. \tag{6.4}$$

Figure 6.2 reports on the comparison of the lower bound (6.3) and the actual threshold value: it refers to the case of a sample equitable partition composed of interconnected rings for increasing values of the community order.

We observe that obtaining a lower bound for $\tau_c^{(1)}$ is meaningful because $\tau_c^{(1)}$ is itself a lower bound for the epidemic threshold τ_c of the exact stochastic model, i.e., $\tau_c = \alpha\tau_c^{(1)}$ with $\alpha \geq 1$ [61, 274]. In fact, smaller values of the effective spreading rate τ, namely $\delta > \beta/\tau_c^{(1)}$, correspond, in the exact stochastic model, to a region where the infectious dies out exponentially fast for sufficiently large times [106, 262, 269, 273]. Thus, in applications, when designing or controlling a network, τ^* (or the more conservative bound in (6.4)) can be adopted to determine a safety region $\{\tau \leq \tau^\star\}$ for the effective spreading rate that guarantees the extinction of epidemics in a reasonable time frame (above the threshold, the overall-healthy state is only reached after an unrealistically long time).

Equality can be attained in (6.3): consider for instance the graph described by the adjacency matrix A in (6.1). Furthermore, we may require that all V_i's have the same number of nodes $k_i = k$ and same internal degree $d_{ii} = d$, $i = 1, \ldots, n$. In this case $Q = d\,\mathrm{Id}_n + \widehat{B}$, where $\widehat{B} := (k\varepsilon b_{ij})_{i,j=1,\ldots n}$, and

$$\lambda_1(Q) = d + k\varepsilon\lambda_1(B),$$

which is the exact value of $\lambda_1(A)$ and consequently of $\tau_c^{(1)}$.

Remark 6.4 Let us underline that if we remove edges between the communities, or inside the communities, in a network whose set nodes has an equitable partition, the lower bound (6.3) still holds. This because the spectral radius of an adjacency matrix is monotonically non increasing under the deletion of edges.

6.1.2 Infection Dynamics for Equitable Partitions

The NIMFA model describes the process of diffusion of epidemics on a graph by expressing the time-change of the probability p_i that node i is infected.

Thus, node i obeys a following differential equation [274]

$$\frac{dp_i(t)}{dt} = \beta \sum_{j=1}^{N} a_{ij} p_j(t)(1 - p_i(t)) - \delta_i(t), \qquad i \in \{1, \ldots, N\} \tag{6.5}$$

The time-derivative of the infection probability of node i consists of two competing processes:

1. while healthy (with probability $1 - p_i(t)$), all infected neighbors, whose average number is $s_i(t)$, try to infect node i at rate β;
2. while node i is infected (with probability $p_i(t)$) it is cured at rate δ.

The following matrix representation of (6.5) holds

$$\frac{dP(t)}{dt} = (\beta A - \delta I)P(t) - \beta \,\mathrm{diag}(p_i(t))AP(t). \tag{6.6}$$

where $P(t) = (\, p_1(t)\; p_2(t) \ldots p_N(t)\,)^T$ and $\mathrm{diag}(p_i(t))$ is the diagonal matrix with elements $p_1(t), p_2(t), \ldots, p_N(t)$. Clearly we study the system for $(p_1, \ldots, p_N) \in I_N = [0, 1]^N$. It can be shown that the system (6.6) is positively invariant in I_N, i.e. if $P(0) \in I_N$ then $P(t) \in I_N$ for all $t > 0$ [163, Lemma 3.1].

The following theorem shows under which conditions the matrix Q can be used in order to express the epidemic dynamics introduced in (6.6). This allows us to describe the time-change of the infection probabilities by a system of n differential equations instead of N. For the proof we refer to [50].

Theorem 6.10 *Let $G = (V, E)$ a graph and $\pi = \{V_j,\ j = 1, \ldots, n\}$ an equitable partition of the node set V. Let G_j be the subgraph of $G = (V, E)$ induced by cell V_j. If $p_h(0) = p_w(0)$ for all $h, w \in G_j$ and for all $j = 1, \ldots, n$, then $p_h(t) = p_w(t)$ for all $t > 0$. In this case we can reduce the number of equations representing the time-change of infection probabilities using the quotient matrix Q.*

Basically the theorem shows that the following subset of I_N, defined by restricting nodes in the same community to have the same state

$$M = \left\{ P \in [0,1]^N | p_1 = \ldots = p_{k_1} = \overline{p}_1,\, p_{k_1+1} = \ldots = p_{k_1+k_2} = \overline{p_2}, \right.$$
$$\left. \ldots,\, p_{(k_1+..k_{n-1}+1)} = \ldots = p_N = \overline{p}_n \right\}$$

is a positively invariant set for the system (6.6).

Thus, let us consider $P(0) \in M$ and $\overline{P} = (\overline{p}_1, \ldots, \overline{p}_n)$, we can write

$$\frac{d\overline{p}_j(t)}{dt} = \beta(1 - \overline{p}_j(t)) \sum_{m=1}^{n} \varepsilon b_{jm} d_{jm} \overline{p}_m(t) \tag{6.7}$$
$$+ \beta d_j (1 - \overline{p}_j(t)) \overline{p}_j(t) - \delta \overline{p}_j(t), \qquad j = 1, \ldots, n$$

After some manipulations we arrive to the following matrix representation of (6.7)

$$\frac{d\overline{P}(t)}{dt} = \beta \left(\mathrm{I}_n - \mathrm{diag}(\overline{p}_j(t)) \right) \widetilde{Q} \overline{P}(t) - \delta \overline{P}(t), \tag{6.8}$$

where $\widetilde{Q} = \mathrm{diag}\left(\frac{1}{\sqrt{k_j}} \right) Q \, \mathrm{diag}(\sqrt{k_j})$. It is immediate to observe that $\sigma(Q) = \sigma(\tilde{Q})$.

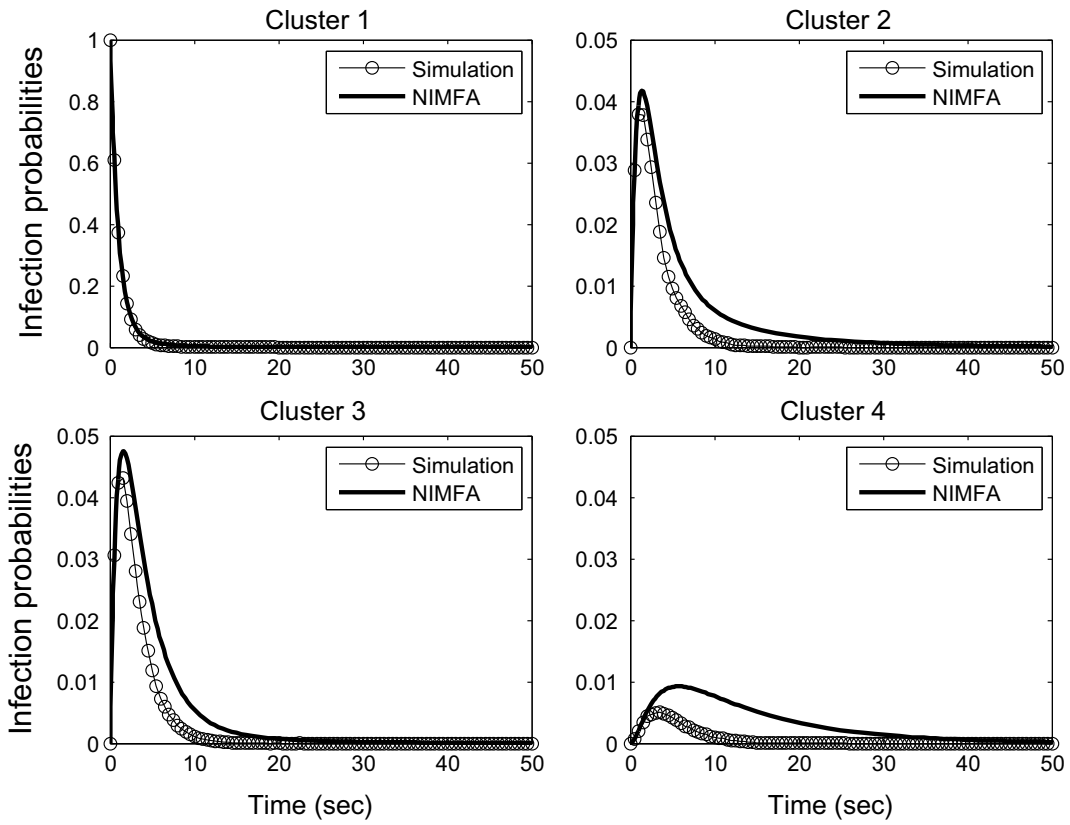

Fig. 6.3 Dynamics of infection probabilities for each community of the network in Fig. 6.1: simulation versus numerical solutions of (6.8); $\tau = \beta/\delta < \tau_c^{(1)} = 0.3178$, with $\beta = 0.29$ and $\delta = 1$, $\varepsilon = 0.3$. At time 0 the only infected node is node 1

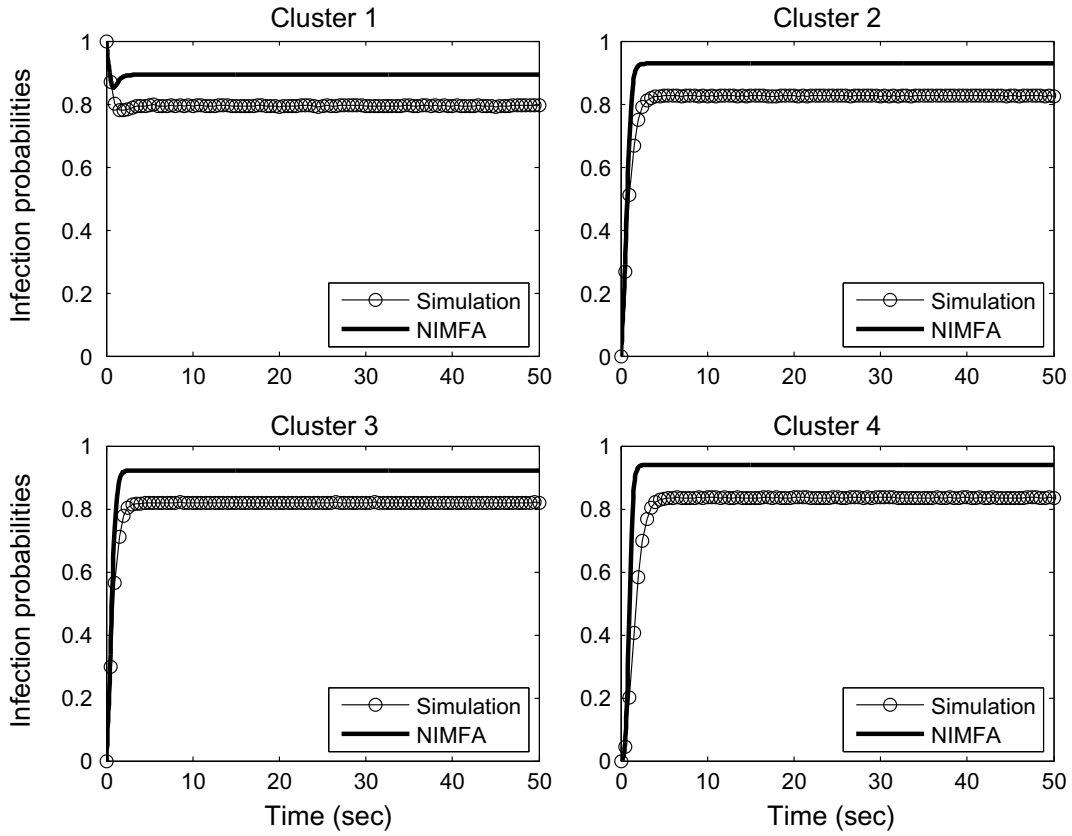

Fig. 6.4 Dynamics of infection probabilities for each community of the network in Fig. 6.1: simulation versus numerical solutions of (6.8); $\tau = \beta/\delta > \tau_c^{(1)} = 0.3178$, with $\beta = 1.5$ and $\delta = 0.3$, $\varepsilon = 0.3$; initial conditions as in Fig. 6.3

In Figs. 6.3 and 6.4 we provide a comparison between the solution of the reduced ODE system (6.8) for the graph in Fig. 6.1 and the averaged 50×10^4 sample paths resulting from a discrete event simulation of the exact SIS process. The discrete event simulation is based on the generation of independent Poisson processes for both the infection of healthy nodes and the recovery of infected ones. We observe that, as expected, NIMFA provides an upper bound to the dynamics of the infection probabilities.

Figure 6.5 depicts the same comparison in the case of a network with eighty nodes partitioned into four communities; each community is a complete graph and all nodes belonging to two linked communities are connected. The agreement between NIMFA and simulations improves compared to Fig. 6.4. This is expected, because the accuracy of NIMFA is known to increase with network order N, under the assumption that the nodes' degree also increases with the number of nodes. Conversely, it is less accurate, e.g., in lattice graphs or regular graphs with fixed degree not depending on N [264, 274].

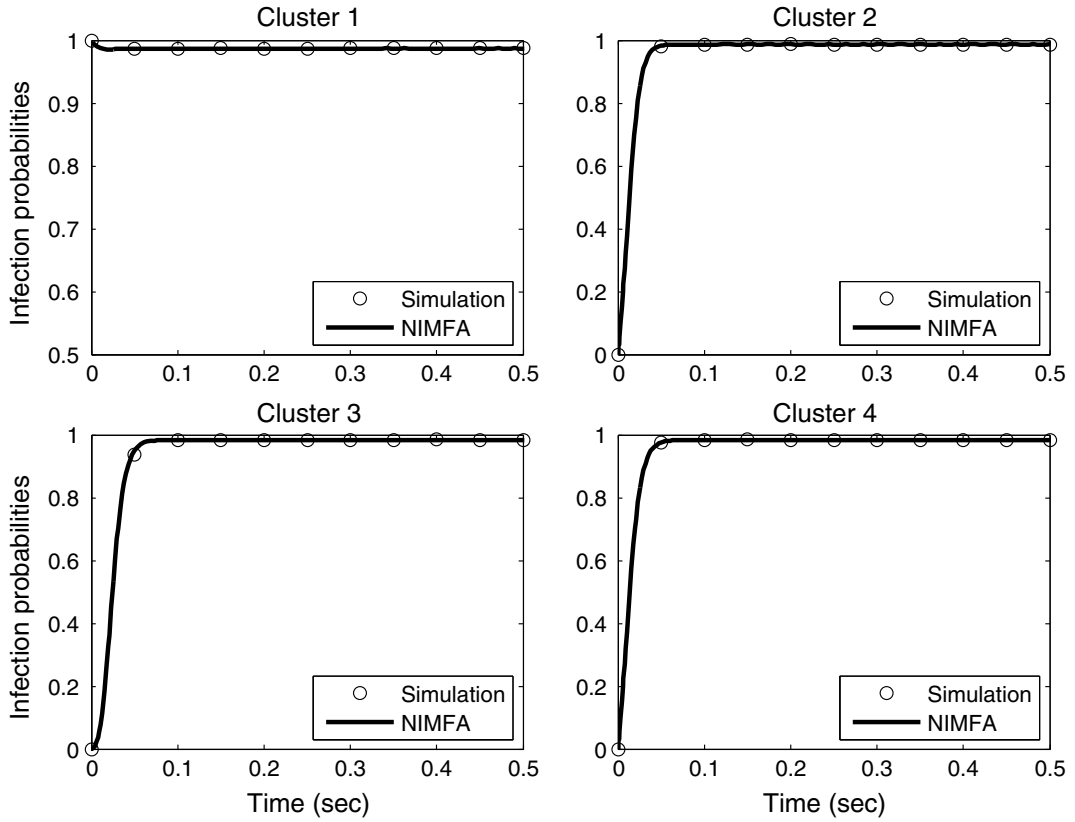

Fig. 6.5 Infection probabilities for each community in a network with $N = 80$, $d_{ii} = k_i - 1 = 19$ and $d_{ij} = 20$, for all $i, j = 1, .., 4$: simulation versus numerical solutions of (6.8); $\tau = \beta/\delta > \tau_c^{(1)} = 0.0348$, with $\beta = 5$ and $\delta = 2$, $\varepsilon = 0.3$; at time 0 all nodes of the 1-st community are infected

6.1.3 Steady State

Corollary 6.5 *When $\tau > \tau_c^{(1)}$ the metastable state P_∞ of the system (6.6) belongs to $M - \{0\}$.*

The result above is proved in [48]. Basically, Corollary 6.5 says that one can compute the $n \times 1$ vector, $\overline{P}_\infty$, of the reduced system (6.8) in order to obtain the $N \times 1$ vector, P_∞, of (6.6): indeed $p_{z\infty}, \ldots, p_{x\infty} = \overline{p}_{j\infty}$, for all $z, x \in G_j$ and $j = 1, \ldots, n$. This provides a computational advantage by solving a system of n equations instead of N. Moreover, since P_∞ is a globally asymptotically stable equilibrium in $I^N - \{0\}$, the trajectories starting outside M will approach those starting in $M - \{0\}$, as time elapses. The same holds clearly for trajectories starting in I^N and in M when $\tau \leq \tau_c^{(1)}$. Numerical experiments in Fig. 6.6 depict this fact.

We focus now on the computation of the steady-state $P_\infty = \left(p_{i\infty}\right)_{i=1,\ldots,N}$ of system (6.6). To this aim, by Corollary 6.5, we can compute the steady-state $\overline{P}_\infty = \left(\overline{p}_{j\infty}\right)_{j=1,\ldots,n}$ of the reduced system (6.8) and obtain

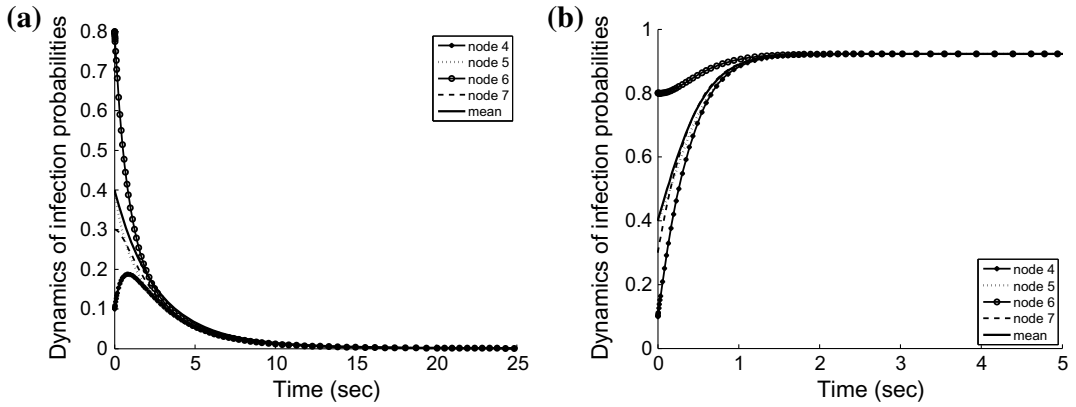

Fig. 6.6 Comparison between the dynamics of the original system (6.6) for each of the nodes belonging to V_3 in Fig. 6.1, for different initial conditions and the dynamics of the reduced system (6.8). In the latter case the initial conditions for each node are the mean value of the $p_i(0)$s. **a** case below the threshold: $\beta = 0.29$, $\delta = 1$, $\varepsilon = 0.3$ **b** case above the threshold: $\beta = 1.5$, $\delta = 0.3$, $\varepsilon = 0.3$

$$\beta(1-\overline{p}_{j\infty})\sum_{m=1}^{n}\left(\frac{k_j}{k_m}\right)^{-1/2} q_{jm}\overline{p}_{m\infty} - \delta\overline{p}_{j\infty} = 0, \qquad j = 1,\ldots,n$$

whence

$$\begin{aligned}\overline{p}_{j\infty} &= 1 - \frac{1}{1+\tau\sum_{m=1}^{n}\left(\frac{k_j}{k_m}\right)^{-1/2} q_{jm}\overline{p}_{m\infty}} \\ &= 1 - \frac{1}{1+\tau g_j\left(\overline{P}\right)}\end{aligned} \tag{6.9}$$

where

$$g_j\left(\overline{P}\right) := \left(d_{jj} + \varepsilon\sum_{m=1}^{n}\left(\frac{k_j}{k_m}\right)^{-1/2}\sqrt{d_{jm}d_{mj}}\right) - \sum_{m=1}^{n}\left(\frac{k_j}{k_m}\right)^{-1/2} q_{jm}(1-\overline{p}_{m\infty}).$$

By introducing $1-\overline{p}_{m\infty} = \frac{1}{1+\tau\sum_{z=1}^{n}\left(\frac{k_m}{k_z}\right)^{-1/2} q_{mz}\overline{p}_{z\infty}}$ in (6.9), we can express $\overline{p}_{j\infty}$ as a continued fraction iterating the formula

$$x_{j,s+1} = f(x_{1;s},..,x_{n;s}) = 1 - \frac{1}{1+\tau g_j(x_{1;s},..,x_{n;s})},$$

As showed in [274], after a few iterations of the formula above, one can obtain a good approximation of $\overline{p}_{j\infty}$, with a loss in the accuracy of the calculation around $\tau = \tau_c$.

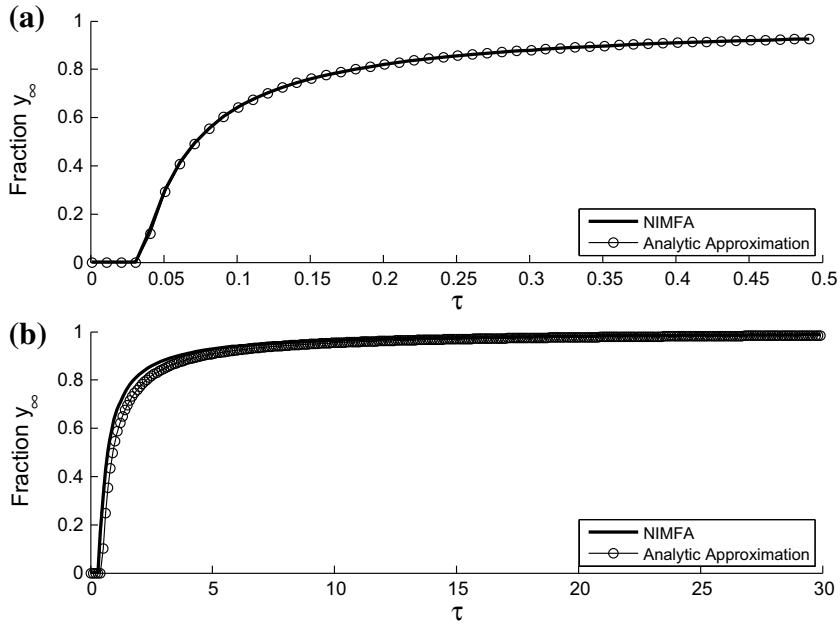

Fig. 6.7 Steady-state average fraction of infected nodes, for different values of τ: comparison between the approximation (6.10) and the exact computation through (6.6); **a** the graph is the one considered in Fig. 6.1a and **b** the one considered in Fig. 6.5

If we consider a regular graph where communities have the same number of nodes, then

$$\overline{p}_{j\infty} = 1 - \left(1/\tau \left(d_{jj} + \varepsilon \sum_{m=1}^{n} \left(\frac{k_j}{k_m}\right)^{-1/2} \sqrt{d_{jm} d_{mj}}\right)\right)$$

is the exact solution of (6.9).

Now let $r_j = d_{jj} + \varepsilon \sum_{m=1}^{n} \left(\frac{k_j}{k_m}\right)^{-1/2} \sqrt{d_{jm} d_{mj}}$ and $r(1) = \min_j r_j$; relying on the estimate $\overline{p}_{j\infty} \approx 1 - \left(1/\tau r_j\right)$ we can express the steady-state average fraction of infected nodes $y_\infty(\tau) = (1/N) \sum_{j=1}^{n} k_j p_{j\infty}(\tau)$ by

$$y_\infty(\tau) \approx 1 - \frac{1}{\tau N} \sum_{j=1}^{n} k_j \frac{1}{d_{jj} + \varepsilon \sum_{m=1}^{n} \left(\frac{k_j}{k_m}\right)^{-1/2} \sqrt{d_{jm} d_{mj}}}. \tag{6.10}$$

According to the analysis reported in [274], approximation (6.10) becomes the more precise the more the difference $r(2) - r(1)$ is small, where $r(2)$ is the second smallest of the r_j's (Fig. 6.7).

6.1.4 Clique Case

A *clique* of a graph is a set of vertices that induces a complete subgraph of that graph. Here we consider the specific case, analyzed in [49], where we have a *clique cover* of the graph, i.e., a set of cliques that partition its vertex set.

Thus, basically, all elements in a community are connected, i.e, $d_{ii} = k_i - 1$ for all $i = 1, ..., n$. Moreover we assume that all nodes belonging to two linked communities i and j are connected, i.e., $d_{ij} = k_j$ and $d_{ji} = k_i$. A sample graph is depicted in Fig. 6.8.

In [49] sufficient conditions for the extinction of epidemics have been found explicitly in terms of the dimension of the communities, their connectivity, and the parameters of the model. In the following we report the main results (for the derivation of the results see [49]).

Theorem 6.11 *Let $G = (V, E)$ be a graph with partition $\pi = \{V_i, i = 1, ..., n\}$, such that all V_i's induce a complete subgraph G_i of G, and all V_i's have the same order $k_i = k$. Moreover let us consider that whenever a node of G_i is connected with a node in G_j, then it is connected with all nodes in G_j. Therefore a sufficient condition for the uniqueness of the zero steady-state is the following:*

$$\frac{d_{\max}\varepsilon\beta + (1 - \frac{1}{k})\beta}{\delta} < \frac{1}{k},$$

where $d_{\max} = \max_i d_i$, and d_i is the number of communities with which the i-th is connected.

Theorem 6.12 *Let $G = (V, E)$ be a graph with partition $\pi = \{V_i, i = 1, ..., n\}$, such that all V_i's induce a complete subgraph G_i of G, each of arbitrary order k_i. Moreover let us consider that whenever a node of G_i is connected with a node in G_j, then it is connected with all nodes in G_j. Therefore a sufficient condition for the uniqueness of the zero steady state is the following:*

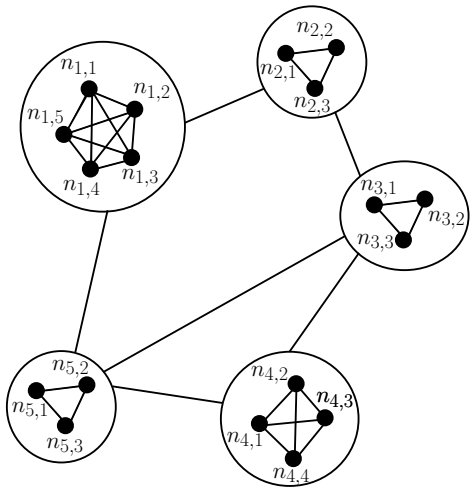

Fig. 6.8 Interconnected cliques. A link between two cliques means that each node in one clique is linked with all nodes in the other clique

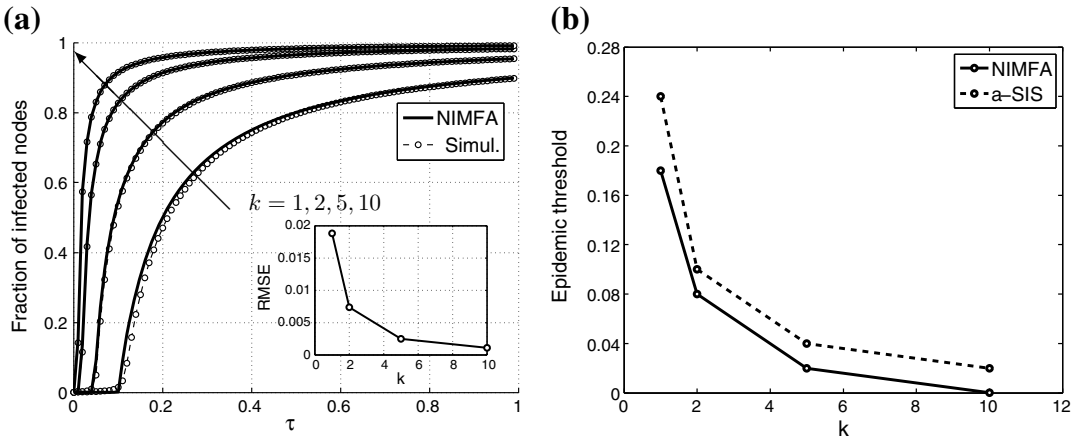

Fig. 6.9 **a** Fraction of infected nodes for different values of k as a function of $\tau = \beta/\delta$, with fixed the ratio $\varepsilon = 1/2$ and the value $\delta = 1$. The network of the communities is a regular graph with degree 10; the number of nodes is $N = 500$. The inserted plot represents the root mean square error between the simulated and the approximated fraction of infected nodes. **b** The corresponding value of the epidemic threshold for the NIMFA and the exact a-SIS model

$$\forall\, i = 1, \ldots, n: \qquad \frac{d_i\, \varepsilon\beta + (1 - \frac{1}{k_i})\beta}{\delta} < \frac{1}{k_i}.$$

Our NIMFA-like approximation is validated here by comparison with the exact SIS model. From the operative standpoint, we compare NIMFA with the a-SIS model [133, 271] where a nodal self-infection is allowed, at rate a. This model has no absorbing state and its stationary distribution, that can be computed for explicitly, can be made arbitrarily close to the quasi-stationary distribution of the original SIS model, by considering appropriate and small values of $a > 0$ [172, 271]. For a detailed explanation on the simulation process see [172].

Effect of community dimension. We depict first, in Fig. 6.9a, the impact of the community dimension k on the fraction of infected nodes in the steady-state, and compare the results of our model to the a-SIS model. The epidemic threshold of the a-SIS model is measured as the value of τ where the second derivative of the steady-state fraction of infected nodes equals zero. We consider a range for $\tau = \beta/\delta$, for constant ratio $\varepsilon = 1/2$ and fixed $\delta = 1$.

The sample network, representing the connections between the communities, has constant degree $d = 10$. The total number of nodes is $N = 500$. The number of elements k is the same for all communities: curves are drawn for increasing values of k ($k = 1, 2, 5, 10$), where $k = 1$ denotes the absence of local clusters. The threshold effect is well visible in the graphs depicted in Fig. 6.9a. As can be further observed, our model and the exact SIS model are in good agreement and the root mean square error between them decreases as k increases.

In Fig. 6.9b the corresponding value of the epidemic threshold for the NIMFA and the a-SIS model is reported. As expected from Theorem 6.11, the critical threshold above which a persistent infection exists decreases with the dimension of the com-

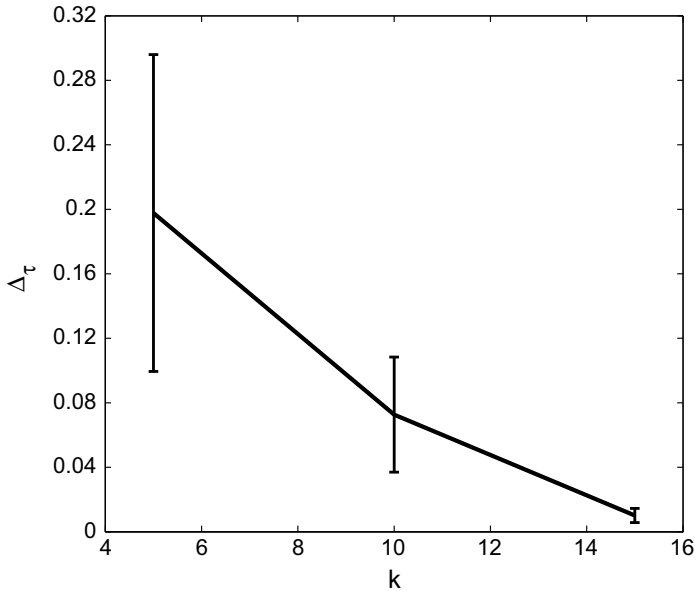

Fig. 6.10 Difference Δ_τ between the epidemic threshold in the case of homogeneous cluster distribution and inhomogeneous cluster distribution for different values of k (5, 10, 15), being fixed the ratio $\varepsilon = 1/8$. The difference was obtained averaging over 300 instances of tree graphs of 10 clusters, the level of confidence is set to 98%

munities. Thus, for large values of the community dimension, a very small value of τ is sufficient to cause epidemic outbreaks, irrespective of the actual network structure.

Effect of the heterogeneity of the community dimension.

One interesting question that concerns the two-scale epidemic model is the influence of the community dimension distribution onto the epidemic threshold. In general, it is not obvious whether, fixing all remaining system's parameters, a constant community dimension will lead to a lower or larger epidemic threshold for the same network.

In Fig. 6.10 we performed a test using a set of 300 sample tree graphs for depicting the connectivity of the communities. Each graph is the spanning tree of an Erdős-Rényi graph of order $n = 10$ and $p = 0.3$. The ratio ε is set to 1/8. The plot draws the difference Δ_τ, obtained averaging over the 300 sample graphs, between the epidemic threshold measured for homogeneous cluster distribution, and the epidemic threshold measured in the case of inhomogeneous cluster distribution.
In particular, for each sample tree, we considered different values of the average cluster dimension $k = 5, 10, 15$. In the case of heterogeneous cluster distribution half of the communities have dimension 2 and half of them have dimension $2k - 2$.

Figure 6.10 exemplifies that heterogeneity of communities' dimension lowers the epidemic threshold compared to the case of constant dimension. This observation agrees with the theory, indeed from the inequality [265, (3.34) on p. 47]:

$$\lambda_1 \geq \frac{2L}{N}\sqrt{1 + \frac{\mathrm{Var}[d]}{(\mathbb{E}[d])^2}},$$

where λ_1 is the spectral radius of a given graph with N nodes and L links, and d is the degree of a randomly chosen node in the graph, we have

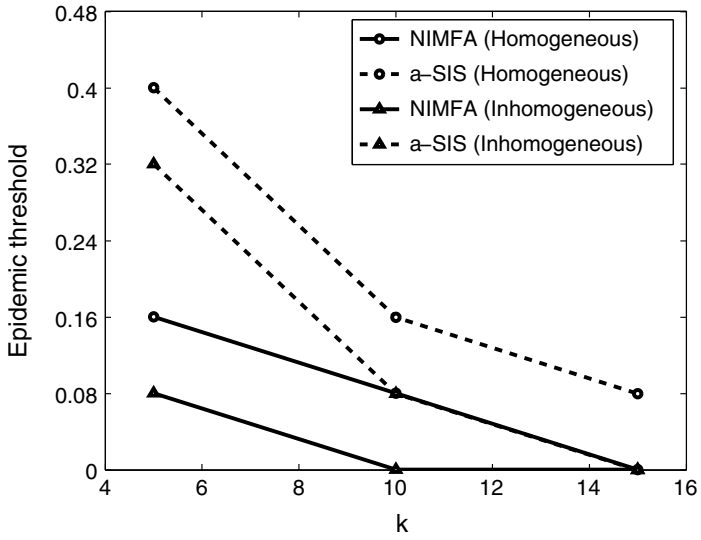

Fig. 6.11 The epidemic threshold in the case of homogeneous cluster distribution and inhomogeneous cluster distribution for different values of k, where the network of the communities is a spanning tree of an Erdős-Rényi graph of order $n = 10$ and $p = 0.3$. Both the NIMFA and the a-SIS thresholds are shown

$$\tau_c^{(1)} = \frac{1}{\lambda_1} \leq \frac{N}{2L} \frac{1}{\sqrt{1 + \frac{\text{Var}[d]}{(\mathbb{E}[d])^2}}}$$

implying that, the larger the variance in the degree d, the lower the NIMFA epidemic threshold $\tau_c^{(1)}$. Unfortunately, since $\tau_c^{(1)} \leq \tau_c$, we cannot conclude that an increase in Var$[d]$ also always lowers the exact epidemic threshold τ_c.

Figure 6.11 shows the epidemic threshold measured for homogeneous community dimension and the epidemic threshold measured for inhomogeneous community dimension, by considering one instance of the previous set of spanning trees of an Erdős-Rényi graph. We report both the results obtained for our model and the results obtained for the a-SIS model: the NIMFA epidemic threshold well estimates the a-SIS epidemic threshold in both community dimension distributions.

6.2 Almost Equitable Partions

In this section we consider graphs where the partition of the vertex set is *almost equitable*.

Definition 6.15 The partition $\pi = \{V_1, ..., V_n\}$ is called *almost equitable* if for all $i, j \in \{1, \ldots, n\}$ with $i \neq j$, there is an integer d_{ij} such that for all $v \in V_i$, it holds

$$d_{ij} = deg(v, V_j) := \#\left\{e \in E : e = \{v, w\}, w \in V_j\right\}$$

independently of $v \in V_i$.

The difference between equitable and almost equitable partitions is that, in the former case, subgraph G_i of G induced by V_i has regular structure, whereas the latter

definition does not impose any structural condition into G_i. Ideally we can think of a network $\tilde{G}$ whose node set has an almost equitable partition as a network G with equitable partition where links between nodes in one or more communities have been added or removed.

The objective is to obtain lower bounds on threshold $\tau_c^{(1)}$, useful in determining a safety region for the extinction of epidemics. We start assuming that links are added only.

To this aim, let us consider two graphs $G = (V, E)$ and $\tilde{G} = (V, \tilde{E})$ with the same partition $\{V_1, \ldots, V_n\}$, but different edge sets $E \subsetneq \tilde{E}$, and assume G to have an equitable partition but $\tilde{G}$ to have merely an almost equitable partition. Then if $\tilde{A}$ and A are the adjacency matrices of $\tilde{G}$ and G respectively it holds

$$\tilde{A} = A + R,$$

where $R = \mathrm{diag}(R_1, \ldots, R_n)$; the dimension of R_i is $k_i \times k_i$ for $i = 1, ..., n$, as before k_i is the order of G_i and n is the number of the communities.

The theorem of Weyl can be applied to $\tilde{A} = A + R$ and then it yields

$$\lambda_1(\tilde{A}) \leq \lambda_1(A) + \lambda_1(R). \tag{6.11}$$

Proposition 6.26 *Let $G = (V, E)$ and $\tilde{G} = (V, \tilde{E})$ be two graphs and consider a partition $\{V_1, \ldots, V_n\}$ of the set of vertices V; we shall denote by $G_i = (V_i, E_i)$ and $\tilde{G}_i = (V_i, \tilde{E}_i)$ the subgraph of G and $\tilde{G}$ induced by the cell V_i, respectively, for $i = 1, ...n$. Assume this partition to be equitable for G and almost equitable for $\tilde{G}$. Let $E \subset \tilde{E}$ with*

$$\tilde{E} \setminus E = \bigcup_{i=1}^{n} (\tilde{E}_i \setminus E_i)$$

(i.e., the edge sets can only differ within cells) and denote by R the adjacency matrix corresponding to a graph with $\tilde{E} \setminus E$ as edge set. Finally, let us denote by G_i^C the graph with edge set $\tilde{E}_i \setminus E_i$ and whose node set is simply the set of endpoints of its edges (i.e., no further isolated nodes).

1. *If $\Delta(G_i^C)$ denotes the maximal degree in G_i^C, $i = 1, \ldots, n$, then*

$$\lambda_1(R) \leq \max_{1 \leq i \leq n} \min \left\{ \sqrt{\frac{2e_i(k_i - 1)}{k_i}}, \Delta(G_i^C) \right\},$$

   where e_i is the number of edges added to G_i, i.e., $e_i = (|\tilde{E}_i| - |E_i|)$, and k_i is the number of nodes in V_i.
2. *If additionally G_i^C is connected for each $i = 1, \ldots, n$, then*

$$\lambda_1(R) \leq \max_{1 \leq i \leq n} \min \left\{ \sqrt{2e_i - k_i' + 1}, \Delta(G_i^C) \right\},$$

where k_i' is the number of nodes of G_i^C.

Thus, by using estimate (6.2) and Proposition 6.26, we can derive a lower bound for the epidemic threshold, actually

$$\tau_c^{(1)} = \frac{1}{\lambda_1(\tilde{A})} \geq \tau^\star = \frac{1}{\max\limits_{1\leq i\leq n} \lambda_1(C_{V_i}) + \lambda_1(\widehat{B}) + \max\limits_{1\leq i\leq n} \min\left\{\sqrt{\frac{2e_i(k_i-1)}{k_i}}, \Delta(G_i^C)\right\}}. \tag{6.12}$$

Now let us consider the case where we remove edges, inside the communities, in a network whose set nodes has an equitable partition, thus because the spectral radius of an adjacency matrix is monotonically non increasing under the deletion of edges, we have

$$\lambda_1(\tilde{A}) \leq \lambda_1(A)$$

whence

$$\frac{1}{\lambda_1(\tilde{A})} \geq \frac{1}{\lambda_1(A)} \geq \min_i \frac{1}{d_{ii} + \lambda_1(\widehat{B})}.$$

The bounds developed so far support the design of community networks with a safety region for the effective spreading rate, that guarantees the extinction of epidemics. E.g. if we consider some G_i, $i = 1, \ldots, n$, it is possible to connect them such in a way to form a graph $\tilde{G} = (V, \tilde{E})$ with an almost equitable partition. Now, any subgraph obtained from $\tilde{G}$, by removing edges inside the communities, will have smaller spectral radius than $\tilde{G}$ and, consequently, a larger epidemic threshold. Thus the lower bound in (6.12) still holds.

6.3 Heterogenous SIS on Networks

Several analytic studies in the literature have determined the conditions for the appearance of endemic infectious states over a population under the assumptions of homogeneous infection and recovery rates.

However, in many real situations, e.g., in social, biological and data communications networks, homogeneity is a demanding assumption and it appears more appropriate to consider instead an heterogeneous setting [263]. A concise overview on the literature considering heterogeneous populations can be found in [215, 289].

To this aim, we report some results in [50], where heterogeneous infection and curing rates have been included.

Thus, hereafter, we denote by β_{ij} the infection rate of node j towards node i, and we exclude self-infection phenomena, i.e., $\beta_{ii} = 0$. Thus, we include the possibility that the infection rates depend on the connection between two nodes, covering a much more general case, than e.g. in [263], where a node i can infect all neighbors with the same infection rate β_i. Basically we allow for the epidemics to spread over a

directed weighted graph. Moreover each node i recovers at rate δ_i, so that the curing rate is node specific.

As for the homogeneous case, the SIS model with heterogeneous infection and recovery rates is a Markovian process as well. The time for infected node j to infect any susceptible neighbor i is an exponential random variable with mean β_{ij}^{-1}. Also, the time for node j to recover is an exponential random variable with mean δ_j^{-1}. In the same way as in the homogeneous setting, we provide the NIMFA approximation. The NIMFA governing equation for node i in the heterogeneous setting writes as

$$\frac{dp_i(t)}{dt} = \sum_{j=1}^{N} \beta_{ij} p_j(t) - \sum_{j=1}^{N} \beta_{ij} p_i(t) p_j(t) - \delta_i p_i(t), \ i = 1, \ldots, N. \tag{6.13}$$

Let the vector $P = (p_1, \ldots, p_N)^T$ and let $\overline{A} = (\overline{a}_{ij})$ be the matrix defined by $\overline{a}_{ij} = \beta_{ij}$ when $i \neq j$, and $\overline{a}_{ii} = -\delta_i$; moreover let $F(P)$ be a column vector whose i-th component is $-\sum_{j=1}^{N} \beta_{ij} p_i(t) p_j(t)$. Then we can rewrite (6.14) in the following form:

$$\frac{dP(t)}{dt} = \overline{A} P(t) + F(P). \tag{6.14}$$

Let

$$r(\overline{A}) = \max_{1 \leq j \leq N} Re(\lambda_j(\overline{A}))$$

be the *stability modulus* [163] of $\overline{A}$, where $Re(\lambda_j(\overline{A}))$ denotes the real part of the eigenvalues of $\overline{A}$, $j = 1, \ldots, N$. We report a result from [163] that lead us to extend the stability analysis of NIMFA in [50] to the heterogeneous case (see [163, Theorem 3.1] for the proof).

Theorem 6.13 *If $r(\overline{A}) \leq 0$ then $P = 0$ is a globally asymptotically stable equilibrium point in $I_N = [0, 1]^N$ for the system (6.14), On the other hand if $r(\overline{A}) > 0$ then there exists a constant solution $P^\infty \in I_N - \{0\}$, such that P^∞ is globally asymptotically stable in $I_N - \{0\}$ for (6.14).*

Finally, in [205] we have defined the equitable partitions for the case of a directed weighted networks, and we have extended the analysis in [50] to this framework. For the purpose of modeling, nodes of the quotient graph can represent communities, e.g., villages, cities or countries. Link weights in the quotient graph in turn provide the strength of the contacts between such communities. In particular, the weight of a link may be (a non-negative) function of the number of people traveling per day between two countries; in fact, the frequency of contacts between them correlates with the propensity of a disease to spread between nodes.