

EPIDEMIC OUTBREAKS IN NETWORKS WITH EQUITABLE OR ALMOST-EQUITABLE PARTITIONS *

S. BONACCORSI*, S. OTTAVIANO*†, D. MUGNOLO‡, AND F. DE PELLEGRINI †

Abstract. We study the diffusion of epidemics on networks that are partitioned into local communities. The gross structure of hierarchical networks of this kind can be described by a *quotient graph*. The rationale of this approach is that individuals infect those belonging to the same community with higher probability than individuals in other communities. In community models the nodal infection probability is thus expected to depend mainly on the interaction of a few, large interconnected clusters. In this work, we describe the epidemic process as a continuous-time individual-based susceptible–infected–susceptible (SIS) model using a first-order mean-field approximation.

A key feature of our 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 overall healthy-state defines a globally asymptotically stable or an unstable equilibrium. Indeed, the spectral radius is related to the epidemic threshold of the system.

Moreover we prove that, above the threshold, another steady-state exists that can be computed using a lower-dimensional dynamical system associated with the evolution of the process on the quotient graph. Our investigations are based on the graph-theoretical notion of *equitable partition* and of its recent and rather flexible generalization, that of *almost equitable partition*.

Key words. susceptible-infected-susceptible model, hierarchical networks, graph spectra, equitable and almost equitable partitions

AMS subject classifications.

1. Introduction. Metapopulation models of epidemics consider the entire population partitioned into communities (also called households, clusters or subgraphs). Such models assume that each community shares a common environment or is defined by a specific relationship (see, e.g., [1, 2, 3]).

Several authors also account for the effect of migration between communities [4, 5]. Conversely, the model we are interested in suits better the diffusion of computer viruses or stable social communities, which do not change during the infection period; hence we do not consider migration.

In this work, we study 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 $\pi = \{V_1, \dots, V_n\}$, i.e., a sequence of mutually disjoint nonempty subsets of V , called cells, whose union is V .

The epidemic model adopted in the rest of the paper is a continuous-time Markovian individual-based susceptible–infected–susceptible (SIS) model. In the SIS model a node can be repeatedly infected, recover and yet be infected again. The viral state of a node i , at time t , is thus described by a Bernoulli random variable $X_i(t)$, where we set $X_i(t) = 0$ if i is healthy and $X_i(t) = 1$ if i is infected. Every node at time t is either infected with probability $p_i(t) = \mathbb{P}(X_i(t) = 1)$ or healthy (but susceptible) with probability $1 - p_i(t) = \mathbb{P}(X_i(t) = 0)$. Each node becomes infected following a

*This work has been partially supported by the European Commission within the framework of the CONGAS project FP7-ICT-2011-8-317672 (see <http://www.congas-project.eu>)

*Mathematics Department, University of Trento, via Sommarive 14, 38123 Povo (Trento), Italy

†CREATE-NET, Via alla Cascata 56/d, 38123 Povo (Trento), Italy

‡Institute of Analysis, University of Hagen, 58084 Hagen, Germany

Poisson process with rate β . Also, i recovers following a Poisson process with rate δ . We further assume that infection and curing processes are independent [6]. The ratio $\tau = \beta/\delta$ is called the *effective spreading rate*.

Recently, also non Markovian types of epidemic spread were introduced in the literature, by choosing other than the exponential interaction time for infection and/or curing (see for instance [7, 8]). However, this seems to complicate the analysis considerably and it is beyond the scope of this work.

Compared to the homogeneous case where the infection rate is the same for all pairs of nodes, in our model 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. Clearly the model can be extended to the case $\varepsilon \geq 1$.

Further models where the epidemic process within communities is faster compared to the rate at which it spreads across communities, have been studied in literature [9, 10, 11, 12].

As described in [6, 13], the SIS process developing on a graph with N nodes is modeled as a continuous-time Markov process with 2^N states. The dynamics of the nodal infection probability is obtained by the corresponding Kolmogorov differential equations, but the resulting dynamical system consists of 2^N linear differential equations, not a viable approach for large networks. Hence, often, an approximation of the SIS process is needed. In this work we consider the first-order mean-field approximation NIMFA, proposed by Van Mieghem et al. in [6, 14, 15].

NIMFA replaces the original 2^N linear differential equations by N non-linear differential equations; they represent the time-change of the infection probability of a node. As typical in first-order approximations of SIS dynamics, the only approximation required by NIMFA is that the infectious state of two nodes in the network are uncorrelated, i.e., $E[X_i(t)X_j(t)] = E[X_i(t)]E[X_j(t)]$.

1.1. Long-term prediction and epidemic threshold. For a network with finite order N , the exact SIS Markov process will always converge towards its unique absorbing state, that is the zero-state where all nodes are healthy. The other states form a transient class, from which one can reach the zero-state with positive probability. Because transitions from the zero-state have zero probability* the stochastic model predicts that the virus will disappear from the network [16].

However the waiting time to absorption is a random variable whose distribution depends on the initial state of the system, and on the parameters of the model [17, 18]. In fact there is a critical value τ_c of the effective spreading rate $\tau = \beta/\delta$, whereby if $\tau > \tau_c$ the time to absorption grows exponentially in N , while for $\tau < \tau_c$ the infection vanishes exponentially fast in time. The critical value τ_c is often called the *epidemic threshold* [6, 19, 20, 21].

Thus above the threshold, a typical realization of the epidemic process may experience a very long waiting time before absorption to the zero-state. During such waiting time, the so-called *quasi-stationary distribution* can be used in order to approximate the probability distribution of occupancy of the system's states. The quasi-stationary distribution is obtained by conditioning on the fact that there is no extinction [17, 18].

*Some models, as, e.g., the ε -SIS model [13], include the possibility of a nodal self-infection, thus making the whole process irreducible.

The quasi-stationary distribution can be regarded as the limiting conditional distribution, useful in representing the long-term behavior of the process “*that in some sense terminates, but appears to be stationary over any reasonable time scale*”[23].

In fact, numerical simulations of SIS processes also reveal that, already for reasonably small networks ($N \geq 100$) and when $\tau > \tau_c$, the overall-healthy state is only reached after an unrealistically long time. Hence, the indication of the model is that, in the case of real networks, one should expect that the extinction of epidemics is hardly ever attained [22, 24]. For this reason the literature is mainly concerned with establishing the value of the epidemic threshold, being a key parameter behind immunization strategies related to the network protection against viral infection.

For an SIS process on graphs, τ_c depends on the spectral radius $\lambda_1(A)$ of the adjacency matrix A [25, 6]. NIMFA determines the epidemic threshold for the effective spreading rate as $\tau_c^{(1)} = \frac{1}{\lambda_1(A)}$, where the superscript (1) refers to the first-order mean-field approximation [6, 31]. Farther, in Thm. 2.1, we shall study the asymptotic behavior of the solutions of the NIMFA system, both above and below the threshold.

We observe that, with respect to the exact Markovian SIS model, the state of nodes was recently proved to be not negatively correlated [27]. It is hence possible to prove that, due to the assumption of independence, NIMFA yields an upper bound for the probability of infection of each node, as well as a lower bound for the epidemic threshold, i.e., $\tau_c = \alpha \tau_c^{(1)}$ with $\alpha \geq 1$.

From the application standpoint, a key issue is to determine for which networks of given order NIMFA performs worst, meaning that $\alpha = \frac{\tau_c}{\tau_c^{(1)}}$ is largest. To this respect, further efforts have been made to satisfactorily quantify the accuracy of the model [28].

Finally, when $\tau > \tau_c^{(1)}$, a limiting occupancy probability appears as the second constant solution[†] of the NIMFA non-linear system which exists, apart from the zero-vector solution. Such non-zero steady-state reflects well the observed viral behavior [14]: it can be seen as the analogous of the quasi-stationary distribution of the exact stochastic SIS model.

1.2. Outline and main results. As already observed in [9], 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. In [9] a complete graph represents the internal structure of each community. Such assumption appears natural for small community orders, for example, because the members of a small community usually know each other, as they may be friends, relatives, members of a common club, employees of the same department, etc. Moreover, given two connected communities, all of their nodes are mutually linked.

In this work, instead, we allow for the case of sparser community structures. More precisely we consider an *equitable partition* of the graph. First of all this means that all nodes belonging to the same community have the same internal degree: formally the subgraph G_i of $G(V, E)$ induced by V_i is regular for all i 's (recall that $\pi = \{V_1, \dots, V_n\}$ 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 .

[†]We remember that all bounded trajectories of an autonomous first-order differential equation tend to an equilibrium, i.e., to a constant solution of the equation.

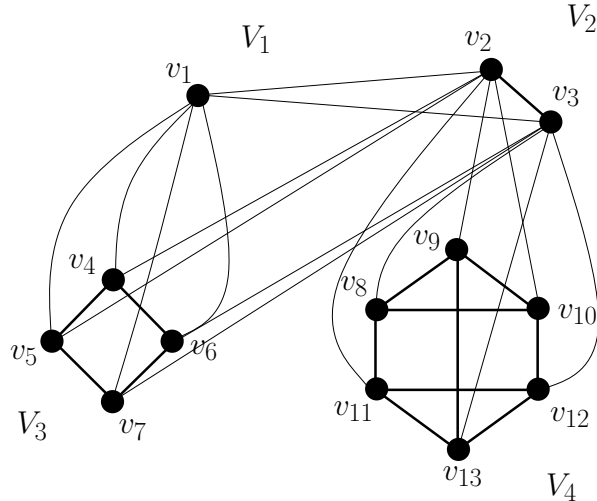


Fig. 1: A sample graph with equitable partition $V = \{\{v_1\}, \{v_2, v_3\}, \{v_3, v_4, v_5, v_6\}, \{v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}\}\}$.

The macroscopic structure of such a network can be described by the *quotient graph* G/π , an oriented graph (possibly) featuring loops and multiple edges. The nodes of the quotient graph are the cells V_1, \dots, V_n in G . In the last part of the work we extend our study to the case of *almost equitable partitions* that does not require any specific structural condition inside each G_i .

Such 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. Also, graphs representing multi-layer networks may be characterized using equitable and almost equitable partitions [29].

In Sec. 2 we describe the NIMFA differential equations and provide an analysis of the global dynamics that allows us to identify the epidemic threshold $\tau_c^{(1)}$. In Sec. 3, after defining equitable partitions, we introduce the so-called quotient matrix Q that is related to G/π . Since matrix Q has the same spectral radius of adjacency matrix A , a novel expression is found for the bound on the epidemic threshold $\tau_c^{(1)}$ as a function of network metrics. Thus, a relation between the epidemic threshold and the spectral properties of the corresponding quotient matrix is obtained.

In Sec. 4 we show under which conditions the matrix Q can be used in order to express the whole epidemic dynamics by a system of n equations instead of N , where $n < N$. We prove the existence of a positively invariant set for the original system of N differential equations that contains the equilibrium points. Moreover we show that, above the threshold, when a second non-zero equilibrium point appears, we can use the reduced system for its computation.

In Sec. 5 we finally extend our investigations to the case of almost equitable partitions. We consider the special case of almost equitable partitions obtained by perturbing an equitable one, i.e., by adding/deleting a certain set of edges from an equitable partition. Thus, we relax the assumption that the internal structure of each community is regular. Even in this case we obtain a lower bound for the epidemic

threshold.

2. The epidemic model. 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 [6]

$$(2.1) \quad \frac{dp_i(t)}{dt} = (1 - p_i(t))\beta \left(\sum_{j=1}^N a_{ij}p_j(t) \right) - \delta p_i(t), \quad i = 1, \dots, N.$$

In (2.1) 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 $\sum_{j=1}^N a_{ij}p_j(t)$, 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 (2.1) holds

$$(2.2) \quad \frac{dP(t)}{dt} = \beta AP(t) - \text{diag}(p_i(t))(\beta AP(t) + \delta u),$$

where $P(t) = (p_1(t) \ p_2(t) \ \dots \ p_N(t))^T$, $\text{diag}(p_i(t))$ is the diagonal matrix with elements $p_1(t), p_2(t), \dots, p_N(t)$ and u is the all-one vector. From (2.2), considering $P(t) = \text{diag}(p_i(t))u$, we can write

$$(2.3) \quad \begin{aligned} \frac{dP(t)}{dt} &= \beta AP(t) - \delta \text{diag}(p_i(t))u - \text{diag}(p_i(t))\beta AP(t) \\ &= (\beta A - \delta I)P(t) - \beta \text{diag}(p_i(t))AP(t). \end{aligned}$$

Clearly we study the system for $(p_1, \dots, p_N) \in I_N = [0, 1]^N$. It can be shown that the system (2.3) is positively invariant in I_N , i.e. if $P(0) \in I_N$ then $P(t) \in I_N$ for all $t > 0$ [33, Lemma 3.1].

The analysis of the global dynamics of (2.3) leads to identify the epidemic threshold $\tau_c^{(1)}$ in terms of the effective spreading rate $\tau = \beta/\delta$ where, as mentioned in Sec. 1.1,

$$(2.4) \quad \tau_c^{(1)} = \frac{1}{\lambda_1(A)},$$

with $\lambda_1(A)$ spectral radius of A . This critical value separates the absorbing phase from the endemic phase. We shall prove this, in Thm 2.1, by studying the stability of the equilibrium points of (2.3), that are solutions of the equation

$$(2.5) \quad P = \frac{\beta}{\delta}(I - \text{diag}(p_i))AP.$$

To this aim we shall adapt the results in [33] to our individual-based SIS model. Let us denote by f the right hand side of (2.3), i.e., (2.3) can be re-written as a vector-valued differential equation

$$(2.6) \quad \frac{dP}{dt} = f(P),$$

where $f : [0, 1]^N \rightarrow \mathbb{R}^N$ is a C^∞ function. Let $P_0 = 0$ be the vector of all zero components, one can easily check that P_0 is an equilibrium point of the system (2.6), i.e. $f(P_0) = 0$. Also, the following holds

THEOREM 2.1. *If $\tau \leq 1/\lambda_1(A)$ then P_0 is a globally asymptotically stable equilibrium of (2.3).*

If $\tau > 1/\lambda_1(A)$, P_0 is unstable and there exists another equilibrium point P_∞ that is globally asymptotically stable in $I_N - \{0\}$.

Proof. We can rewrite the system (2.6) in the following form (see [32, p. 108])

$$(2.7) \quad \dot{P} = D_f P + F(P),$$

where D_f is the Jacobian matrix of f at P_0 and $F(P)$ is a column vector whose i -th component is $-\beta \sum_{j=1}^N a_{ij} p_i p_j$.

From (2.3) we have

$$(Df(P_0))_{ij} = \begin{cases} \beta a_{ij} & i \neq j \\ -\delta & i = j \end{cases}$$

that is $D_f = \beta A - \delta I$. Since adjacency matrix A is real and symmetric its eigenvalues are real. Hence, the eigenvalues of D_f are real as well and of the form

$$\lambda_i(D_f) = \beta \lambda_i(A) - \delta.$$

In particular, let $\lambda_1(D_f) = \max_i \lambda_i(D_f)$, since the spectral radius of A is positive we have

$$\lambda_1(D_f) = \beta \lambda_1(A) - \delta.$$

Now we can apply [33, Thm. 3.1] to the system (2.7) and assert that when $\lambda_1(D_f) \leq 0$, i.e., $\tau \leq 1/\lambda_1(A)$, P_0 is a globally asymptotically stable equilibrium of (2.3).

Conversely, if $\lambda_1(D_f) > 0$, i.e. $\tau > 1/\lambda_1(A)$, there exists another equilibrium point P_∞ . P_0 and P_∞ are the only equilibrium points in I_N and P_∞ is globally asymptotically stable in $I_N - \{0\}$.

Finally, since $\tau > 1/\lambda_1(A)$, we have $\lambda_1(D_f) > 0$. From Lyapunov's Linearization (or First) Method, it follows that P_0 is an unstable equilibrium point in I^N . \square

3. Equitable Partitions. In this section we describe the SIS individual-based model for graphs with equitable partitions. The original definition of equitable partition is due to Schwenk [30].

DEFINITION 3.1. *Let $G = (V, E)$ be a graph. The partition $\pi = \{V_1, \dots, V_n\}$ of the node set V is called equitable if for all $i, j \in \{1, \dots, n\}$, there is an integer d_{ij} such that*

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

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. In particular, each V_i induces a subgraph of G that is necessarily regular.

REMARK 1. *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 $i = 1, \dots, n$.

An equitable partition generates the *quotient graph* G/π , which is a *multigraph* with cells as vertices and 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 cells 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 2. In [9] it has been considered the special case when each community has a clique structure, i.e. $d_{ii} = k_i - 1$ for all $i = 1, \dots, n$. Moreover all nodes belonging to two linked communities i and j are connected, $d_{ij} = k_j$. By considering the theory of equitable partition, we generalize the cited work and consider any kind of regular graph to represent the internal structure of each community. Moreover, unlike before, if two communities i and j are connected, each node in community i is connected with $d_{ij} \leq k_j$ nodes in community j .

3.1. Example. Let us assume that the adjacency matrix B of the quotient graph is given and that, for any $i, j \in \{1, \dots, n\}$, $b_{ij} \neq 0$ 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

$$(3.1) \quad A = \begin{bmatrix} C_{V_1} & \varepsilon J_{k_1 \times k_2} b_{12} & \cdot & \cdot & \varepsilon J_{k_1 \times k_n} b_{1n} \\ \varepsilon J_{k_2 \times k_1} b_{21} & C_{V_2} & \cdot & \cdot & \varepsilon J_{k_2 \times k_n} b_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & C_{V_n} \end{bmatrix}$$

We observe that (3.1) 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 (2.3) with its weighted version (see [26] for a deeper explanation).

3.2. The quotient matrix. We search for a smaller matrix Q that contains the relevant information for the evolution of the system. Such a matrix is the *quotient matrix* of the equitable partition. In Prop. 3.3 we will see that Q and A have the same spectral radii. As a consequence, we can compute its spectral radius in order to estimate the epidemic threshold, instead of computing the spectral radius of matrix A .

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 3.2. Let $G = (V, E)$ a graph. Let $\pi = \{V_i, i = 1, \dots, n\}$ 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 the example in Sec. 3.1 the form of Q is rather simple:

$$q_{ii} = \sum_{h=1}^{k_i} \left(\frac{1}{\sqrt{k_i}} \right)^2 \sum_{k=1}^{k_i} (C_{V_i})_{kh} = \frac{1}{k_i} \sum_{h,k=1}^{k_i} (C_{V_i})_{kh}$$

and

$$q_{ij} = \frac{1}{\sqrt{k_i k_j}} \sum_{z \in V_i, l \in V_j} a_{zl} = \sqrt{k_i k_j} \varepsilon b_{ij}.$$

Hence we obtain that

$$Q = \text{diag}(d_{ii}) + (\sqrt{k_i k_j} \varepsilon b_{ij})_{i,j=1,\dots,n},$$

where $d_{ii} = \frac{1}{k_i} \sum_{h,k=1}^{k_i} (C_{V_i})_{kh}$ is the internal degree of the subgraph induced by V_i .

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

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

There exists a close relationship between the spectral properties of Q and that of A . Being the order of Q smaller of that of A , a result in [35] basically shows that $\sigma(Q) \subseteq \sigma(A)$, where with $\sigma(A)$ we refer, hereafter, to the spectrum of a square matrix A . Furthermore it holds the following

PROPOSITION 3.3. *Let $G = (V, E)$ a graph. Let $\pi = \{V_i, i = 1, \dots, n\}$ be an equitable partition of the node set V . The adjacency matrix A and the quotient matrix Q have the same spectral radius, i.e.*

$$\lambda_1(Q) = \lambda_1(A).$$

Proof. See [36, art. 62]. \square

3.3. Complexity reduction. Prop. 3.3 further details that, once the network structure is encoded in the connectivity of a quotient graph Q , then the epidemic threshold $\tau_c^{(1)}$ is expressed by the spectral radius of Q .

Now, since the order of Q is smaller than the order of A , this can provide a computational advantage. The complexity reduction can be evaluated easily, e.g. in the case of the power iteration method [34]. The power iteration method is a numerical technique for approximating a dominant eigenpair of a diagonalizable matrix L , using the following iteration

$$y^h = L y^{h-1}, \quad h = 1, 2, \dots$$

for a given initial vector y^0 . As the iteration step h increases, y^h approaches a vector which is proportional to a dominant eigenvector of L . If we order the eigenvalues of L such as $|\lambda_1(L)| \geq |\lambda_2(L)| \geq \dots \geq |\lambda_n(L)|$, the rate of convergence of the method is ruled by $|\lambda_2(L)|/|\lambda_1(L)|$.

In our case, for the Perron-Frobenius Theorem the dominant eigenvalue $\lambda_1(A)$ is positive and by Prop. 3.3, $\lambda_1(A) = \lambda_1(Q)$. Furthermore $\sigma(Q) \subseteq \sigma(A)$, hence $\max_{i \geq 2} |\lambda_i(A)| \geq \max_{i \geq 2} |\lambda_i(Q)|$: this means that the convergence of power iteration for matrix Q is never slower than for matrix A . Finally, it is immediate that at each step the computational complexity is $O(n^2)$ for Q whereas for A it is $O(N^2)$.

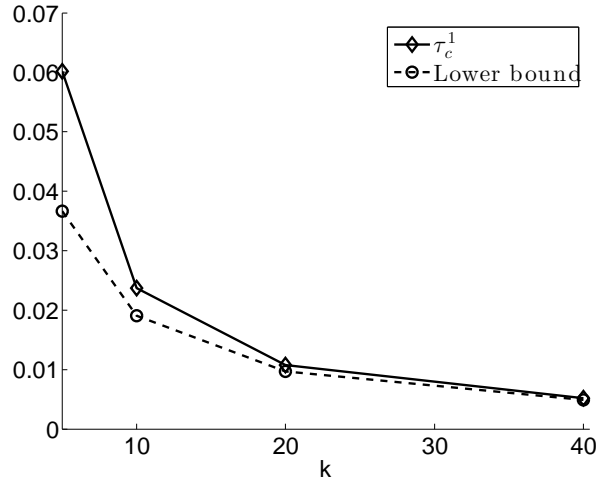


Fig. 2: Lower bound (3.3) versus epidemic threshold: comparison for different values of k in a 40-communities network. The internal structure of each community is a ring and $d_{ij} = 2$ for all $i, j = 1, \dots, n$.

3.4. A lower bound for $\tau_c^{(1)}$. We can write $Q = D + \widehat{B}$, where $D = \text{diag}(d_{ii})$ and $\widehat{B} = (\sqrt{d_{ij}d_{ji}}\varepsilon b_{ij})_{i,j=1,\dots,n}$. By the Weyl's theorem [34] we have

$$(3.2) \quad \lambda_1(Q) \leq \lambda_1(D) + \lambda_1(\widehat{B}) = \max_{1 \leq i \leq n} d_{ii} + \lambda_1(\widehat{B}).$$

From (2.4) and by Proposition 3.3

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

thus a lower bound for the epidemic threshold can be derived from (3.2)

$$(3.3) \quad \tau_c^{(1)} \geq \tau^* = \min_i \frac{1}{d_{ii} + \lambda_1(\widehat{B})},$$

In applications, when designing or controlling a network, this value can be adopted to determine a safety region $\{\tau \leq \tau^*\}$ for the effective spreading rate that guarantees the extinction of epidemics.

Fig. 2 reports on the comparison of the lower bound 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 $\alpha \geq 1$, as anticipated in Sec. 1.1. 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 decay towards the healthy state decreases exponentially fast in time. By forcing the effective spreading rate below τ^* , one ensures that the epidemic will go extinct in a reasonable time frame (we recall that, above the threshold, the overall-healthy state is only reached after an unrealistically long time.).

Equality can be attained in (3.3): consider for instance the graph described by the adjacency matrix A in (3.1). Furthermore, we may require that all V_i 's have the same number of nodes $k_i = k$ and same internal degree $d_{ii} = d$, $i = 1, \dots, n$. In this case $Q = d\text{Id}_n + \widehat{B}$, where $\widehat{B} := (k\varepsilon b_{ij})_{i,j=1,\dots,n}$, and

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

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

4. Infection Dynamics for Equitable Partitions. In this section we show under which conditions matrix Q can be used in order to express the epidemic dynamics introduced in (2.3). This allows us to describe the time-change of the infection probabilities by a system of n differential equations instead of N .

THEOREM 4.1. *Let $G = (V, E)$ a graph and $\pi = \{V_j, j = 1, \dots, n\}$ 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 $j = 1, \dots, n$, 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 .*

Proof. Let $\bar{p}_j(t) = \frac{1}{k_j} \sum_{h \in G_j} p_h(t)$ be the average value of the infection probabilities at time t of nodes in G_j . Then starting from (2.3), we can write a new system of differential equations

$$(4.1) \quad \begin{aligned} \frac{d(p_h(t) - \bar{p}_j(t))}{dt} &= -\delta(p_h(t) - \bar{p}_j(t)) + \beta(1 - p_h(t)) \sum_{z=1}^N a_{hz} p_z(t) \\ &\quad - \frac{1}{k_j} \beta \sum_{l \in G_j} (1 - p_l(t)) \sum_{z=1}^N a_{lz} p_z(t), \quad \forall h \in G_j, \quad j = 1, \dots, n. \end{aligned}$$

From (4.1) we have

$$\begin{aligned} \frac{d(p_h(t) - \bar{p}_j(t))}{dt} &= -\delta(p_h(t) - \bar{p}_j(t)) + \beta \left(\sum_{m=1}^n \sum_{z \in G_m} a_{hz} p_z(t) - \frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} a_{lz} p_z(t) \right) \\ &\quad - \beta \left(p_h(t) \sum_{m=1}^n \sum_{z \in G_m} a_{hz} p_z(t) - \frac{1}{k_j} \sum_{l \in G_j} p_l(t) \sum_{m=1}^n \sum_{z \in G_m} a_{lz} p_z(t) \right), \end{aligned}$$

that can be written as

$$\begin{aligned} \frac{d(p_h(t) - \bar{p}_j(t))}{dt} &= -\delta(p_h(t) - \bar{p}_j(t)) + \beta \left(\frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} (a_{hz} - a_{lz}) p_z(t) \right) \\ &\quad - \beta \frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} (a_{hz} p_h(t) - a_{lz} p_l(t)) (p_z(t) - \bar{p}_m(t)) \\ &\quad - \beta \frac{1}{k_j} \left(\sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} (a_{hz} p_h(t) - a_{lz} p_l(t)) \right) \bar{p}_m(t). \end{aligned}$$

Whence, since $\sum_{z \in G_m} a_{hz} = d_{jm}$, for $h \in G_j$ and for all $m = 1, \dots, n$, we have

$$(4.2) \quad \begin{aligned} \frac{d(p_h(t) - \bar{p}_j(t))}{dt} &= - \left[\sum_{m=1}^n \beta d_{jm} \bar{p}_m(t) + \delta \right] (p_h(t) - \bar{p}_j(t)) \\ &\quad + \beta \frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} (a_{hz} - a_{lz}) p_z(t) \\ &\quad - \beta \frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} (a_{hz} p_h(t) - a_{lz} p_l(t)) (p_z(t) - \bar{p}_m(t)). \end{aligned}$$

Now, we note that

$$-\frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} (a_{hz} p_h(t) - a_{lz} p_l(t)) (p_z(t) - \bar{p}_m(t))$$

can be written as

$$\begin{aligned} &-\frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} ((p_h(t) - \bar{p}_j(t)) a_{hz} - (p_l(t) - \bar{p}_j(t)) a_{lz}) (p_z(t) - \bar{p}_m(t)) \\ &-\frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} \bar{p}_j(t) (a_{hz} - a_{lz}) (p_z(t) - \bar{p}_m(t)), \end{aligned}$$

whence we can rewrite (4.2) as

$$\begin{aligned} \frac{d(p_h(t) - \bar{p}_j(t))}{dt} &= - \left[\sum_{m=1}^n \beta d_{jm} \bar{p}_m(t) + \delta \right] (p_h(t) - \bar{p}_j(t)) \\ &\quad + \beta \frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} (a_{hz} - a_{lz}) (p_z(t) - \bar{p}_m(t) + \bar{p}_m(t)) \\ &\quad - \beta \frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} ((p_h(t) - \bar{p}_j(t)) a_{hz} - (p_l(t) - \bar{p}_j(t)) a_{lz}) (p_z(t) - \bar{p}_m(t)) \\ &\quad - \beta \frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} \bar{p}_j(t) (a_{hz} - a_{lz}) (p_z(t) - \bar{p}_m(t)). \end{aligned}$$

Finally, since $\frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} (a_{hz} - a_{lz}) \bar{p}_m(t) = 0$, we can consider the following system

$$\begin{aligned} \frac{d(p_h(t) - \bar{p}_j(t))}{dt} &= - \left[\sum_{m=1}^n \beta d_{jm} \bar{p}_m(t) - \delta \right] (p_h(t) - \bar{p}_j(t)) \\ &\quad + \beta \frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} (a_{hz} - a_{lz}) (p_z(t) - \bar{p}_m(t)) (1 - \bar{p}_j(t)) \\ &\quad - \beta \frac{1}{k_j} \sum_{l \in G_j} \sum_{m=1}^n \sum_{z \in G_m} ((p_h(t) - \bar{p}_j(t)) a_{hz} - (p_l(t) - \bar{p}_j(t)) a_{lz}) (p_z(t) - \bar{p}_m(t)), \end{aligned}$$

$\forall h \in G_j, \quad j = 1, \dots, n$

Now let us denote by $g(t)$ the solution of (4.1), where $g : \mathbb{R} \rightarrow \mathbb{R}^N$ and consider the case where

$$(4.3) \quad p_h(0) - \bar{p}_j(0) = 0, \quad \forall h \in G_j, \quad j = 1, \dots, n,$$

i.e., $p_h(0) = p_w(0)$ for all $h, w \in G_j$. Then, from (4.3), we can easily see that the identically zero function $g \equiv 0$ is the unique solution of (4.1) with initial conditions (4.3). Indeed $g \equiv 0$ means that for all $t \geq 0$, $p_h(t) = p_w(t)$ for all $h, w \in G_j$, $j = 1, \dots, n$. Moreover the vector $P(t)$ such that $p_h(t) = p_w(t)$ for all $h, w \in G_j$, $j = 1, \dots, n$, is a solution of (2.3) and it is unique in $[0, 1]^N$ with respect to the initial conditions (4.3), [32, Cap. 2, Sec. 2.2]. Thus we can conclude that also $g = 0$ is a unique solution of (4.1) in $[-1, 1]^N$.

Basically we have shown that the following subset of I_N

$$M = \left\{ P \in [0, 1]^N \mid p_1 = \dots = p_{k_1} = \bar{p}_1, p_{k_1+1} = \dots = p_{k_1+k_2} = \bar{p}_2, \dots, p_{(k_1+\dots+k_{n-1}+1)} = \dots = p_N = \bar{p}_n \right\}$$

is a positively invariant set for the system (2.3). This allows us to reduce the system (2.3) of N differential equations and describe the time-change of the infection probabilities by a system of n equations involving the matrix Q .

Indeed, let us consider $P(0) \in M$ and $\bar{P} = (\bar{p}_1, \dots, \bar{p}_n)$, we can write

$$(4.4) \quad \begin{aligned} \frac{d\bar{p}_j(t)}{dt} &= \beta(1 - \bar{p}_j(t)) \sum_{m=1}^n \varepsilon b_{jm} d_{jm} \bar{p}_m(t) \\ &\quad + \beta d_j(1 - \bar{p}_j(t)) \bar{p}_j(t) - \delta \bar{p}_j(t), \quad j = 1, \dots, n \end{aligned}$$

Hence, based on Thm. 2.1 in [35], we observe that

$$q_{ij} = (k_j/k_i)^{1/2} d_{ji},$$

This relation in our case brings

$$d_{jm} = \left(\frac{k_j}{k_m} \right)^{-1/2} \frac{q_{mj}}{\varepsilon} = \left(\frac{k_j}{k_m} \right)^{-1/2} \frac{q_{jm}}{\varepsilon},$$

where the last equality holds because Q is symmetric. We can rewrite (4.4) as

$$(4.5) \quad \begin{aligned} \frac{d\bar{p}_j(t)}{dt} &= \beta(1 - \bar{p}_j(t)) \sum_{m=1, m \neq j}^n \left(\frac{k_j}{k_m} \right)^{-1/2} q_{jm} p_m(t) \\ &\quad + \beta q_{jj}(1 - \bar{p}_j(t)) \bar{p}_j(t) - \delta \bar{p}_j(t); \quad j = 1, \dots, n \end{aligned}$$

where $q_{jj} = d_{jj} = \lambda_1(C_{V_j})$. The matrix representation of (4.5) is the following

$$(4.6) \quad \frac{d\bar{P}(t)}{dt} = \beta (\mathbf{I}_n - \text{diag}(\bar{p}_j(t))) \tilde{Q} \bar{P}(t) - \delta \bar{P}(t),$$

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

COROLLARY 4.2. *When $\tau > \tau_c^{(1)}$ the non-zero steady-state P_∞ of the system (2.3) belongs to $M - \{0\}$.*

Proof. In Theorem 2.1 we have shown that when $\tau > \tau_c^{(1)}$, the system (2.3) has a globally asymptotically stable equilibrium P_∞ in $I_N - \{0\}$; hence for any initial state $P(0) \in I_N - \{0\}$

$$\lim_{t \rightarrow \infty} \|P(t) - P_\infty\| = 0.$$

We have proved in Thm. 4.1 that if $P(0) \in M$ then $P(t) \in M$ for all $t > 0$, thus we can conclude that P_∞ must be in $M - \{0\}$ when $\tau > \tau_c^{(1)}$. \square

Basically, Corollary 4.2 says that one can compute the $n \times 1$ vector, \bar{P}_∞ , of the reduced system (4.6) in order to obtain the $N \times 1$ vector, P_∞ , of (2.3): indeed $p_{z\infty}, \dots, p_{x\infty} = \bar{p}_{j\infty}$, for all $z, x \in G_j$ and $j = 1, \dots, n$. 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 $M - \{0\}$. The same holds clearly for trajectories starting in I^N and in M when $\tau \leq \tau_c^{(1)}$. Numerical experiments in Fig. 6 depict this fact.

The statements proved above can be easily verified, with a direct computation, in the simple case of graphs considered in [9] (see Remark 2). Indeed for all $h, w \in G_j$, $j = 1, \dots, n$, we have

$$\begin{aligned} \frac{d(p_h(t) - p_w(t))}{dt} = & -\delta(p_h(t) - p_w(t)) + \beta \sum_{z \notin G_j} [(1 - p_h(t))a_{hz} - (1 - p_w(t))a_{wz}] p_z(t) \\ & + \beta \sum_{z \in G_j, z \neq h, w} [(1 - p_h(t))a_{hz} - (1 - p_w(t))a_{wz}] p_z(t) \\ (4.7) \quad & + \beta \sum_{z=h, w} [(1 - p_h(t))a_{hz} - (1 - p_w(t))a_{wz}] p_z(t) \end{aligned}$$

Since in this special case $a_{hz} = a_{wz}$, for all $z \in V$ s.t. $z \neq h, j$, we can rewrite (4.7) as

$$\frac{d(p_h(t) - p_w(t))}{dt} = - \left[\delta + \beta \left(\sum_{z=1, z \neq h, w}^N a_{hz} p_z(t) + 1 \right) \right] (p_h(t) - p_w(t)).$$

whence

$$p_h(t) - p_w(t) = (p_h(0) - p_w(0)) e^{-\int_0^t \delta + \beta (\sum_{z=1, z \neq h, w}^N a_{hz} p_z(s) + 1) ds}.$$

Thus if $p_h(0) = p_w(0)$ for the uniqueness of solution it will occur $p_h(t) = p_w(t)$ for all $t > 0$, as we have proved in Thm. 4.1, but if the initial conditions are different, the distance between $p_w(t)$ and $p_z(t)$ decreases exponentially.

REMARK 3. *The framework of quotient graphs extends the NIMFA model to graphs with prescribed community network structure. It reduces to the original NIMFA model when $k_j = 1$ for all $j = 1, \dots, n$.*

4.1. Steady-state. We focus now on the computation of the steady-state $P_\infty = (p_{i\infty})_{i=1, \dots, N}$ of system (2.3). To this aim, by Corollary 4.2, we can compute the steady-state $\bar{P}_\infty = (\bar{p}_{j\infty})_{j=1, \dots, n}$ of the reduced system (4.6) and obtain

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

whence

$$(4.8) \quad \begin{aligned} \bar{p}_{j\infty} &= \frac{\beta \sum_{m=1}^n \left(\frac{k_j}{k_m}\right)^{-1/2} q_{jm} \bar{p}_{m\infty}}{\beta \sum_{m=1}^n \left(\frac{k_j}{k_m}\right)^{-1/2} q_{jm} \bar{p}_{m\infty} + \delta} = 1 - \frac{1}{1 + \tau \sum_{m=1}^n \left(\frac{k_j}{k_m}\right)^{-1/2} q_{jm} \bar{p}_{m\infty}} \\ &= 1 - \frac{1}{1 + \tau g_j(\bar{P})} \end{aligned}$$

where

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

From (4.8) follows that the steady-state infection probability of any node j is bounded by

$$(4.9) \quad 0 \leq \bar{p}_{j\infty} \leq 1 - \frac{1}{1 + \tau (d_{jj} + \varepsilon \sum_{m=1}^n \left(\frac{k_j}{k_m}\right)^{-1/2} \sqrt{d_{jm} d_{mj}})},$$

where the inequality holds true because $\bar{p}_{j\infty} \in [0, 1]$ for all $j = 1, \dots, n$.

By introducing $1 - \bar{p}_{m\infty} = \frac{1}{1 + \tau \sum_{z=1}^n \left(\frac{k_m}{k_z}\right)^{-1/2} q_{mz} \bar{p}_{z\infty}}$ in (4.8), we can express $\bar{p}_{j\infty}$ as a continued fraction iterating the formula

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

As showed in [6], after a few iterations of the formula above, one can obtain a good approximation of $\bar{p}_{j\infty}$, with a loss in the accuracy of the calculation around $\tau = \tau_c$. Ultimately, such numerical estimation can be used to improve the bound in (4.9).

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

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

is the exact solution of (4.8).

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 $\bar{p}_{j\infty} \approx 1 - (1/\tau r_j)$ 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

$$(4.10) \quad y_\infty(\tau) \approx 1 - \frac{1}{\tau N} \sum_{j=1}^n k_j \frac{1}{d_{jj} + \varepsilon \sum_{m=1}^n \left(\frac{k_j}{k_m}\right)^{-1/2} \sqrt{d_{jm} d_{mj}}}.$$

According to the analysis reported in [6], approximation (4.10) 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. In Sec. 4.3 we report on some related numerical experiments.

4.2. Examples. In Fig. 1 we provide an example of a graph which has an equitable partition with respect to $V_1 = \{v_1\}$, $V_2 = \{v_2, v_3\}$, $V_3 = \{v_3, v_4, v_5, v_6\}$, $V_4 = \{v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}\}$.

The corresponding quotient matrix reads

$$Q = \begin{bmatrix} 0 & \varepsilon\sqrt{2} & \varepsilon 2 & 0 \\ \varepsilon\sqrt{2} & 1 & \varepsilon\sqrt{2} & \varepsilon\sqrt{3} \\ \varepsilon 2 & \varepsilon\sqrt{2} & 2 & 0 \\ 0 & \varepsilon\sqrt{3} & 0 & 3 \end{bmatrix}$$

From (4.6) we have that the steady-state can be computed by

$$\bar{P}_\infty = \frac{\beta}{\delta} (\mathbf{I}_n - \text{diag}(\bar{p}_\infty)) \text{diag}(s_j) Q \text{diag}(1/s_j) \bar{P}_\infty,$$

where s_j is the j -th entry of vector $s = (1, \sqrt{2}, 2, \sqrt{6})$.

4.3. Numerical experiments. In Figures 3 and 4 we provide a comparison between the solution of the reduced ODE system (4.6) for the graph in Fig. 1 and the averaged $50 \cdot 10^4$ 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. Also, in Fig. 3 we observe that the dynamics for the communities that are initially healthy is characterized by a unique maximum for the infection probability, which decreases afterwards. The communities initially infected, conversely, show a monotonic decrease of the infection probability.

Fig. 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 (see Remark 2). The agreement between NIMFA and simulations improves compared to Fig. 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 [6, 28].

Fig. 6 depicts the solutions of system (2.3) for each node belonging to V_3 in the graph of Fig. 1; here nodes in V_3 have different initial infection probabilities $p_i(0)$'s. These solutions are compared with the one computed using the reduced system (4.6), in the case when the initial conditions for those nodes are the same, precisely equal to the mean value of the $p_i(0)$'s. As expected, trajectories starting outside invariant set M described in Thm. 4.1 tend to approach the one starting in M as time elapses. Finally, we report on numerical experiments about the steady-state average fraction of infected nodes. More precisely, Fig. 7 compares the value obtained by solving the original system (2.5) and the value obtained from approximation (4.10), as a function of τ .

In Fig. 8, instead, we have reported on the comparison between the steady-state average fraction of infected nodes, as function of τ , computed via NIMFA and via simulations. We consider a graph of regular degree $d = 10$ and $N = 500$, whose communities are clique, each with the same number of elements k . We repeat the same calculation for different values of k in the communities. As it can be observed, our model and the exact SIS model are in good agreement and the root mean square error between them decreases as k increases.

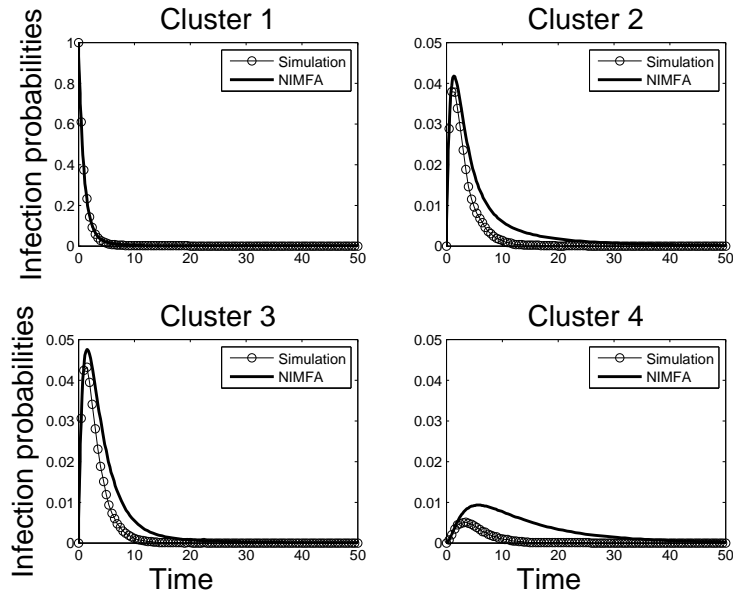


Fig. 3: Dynamics of infection probabilities for each community of the network in Fig.1: simulation versus numerical solutions of (4.6); $\tau = \beta/\delta < \tau_c^{(1)} = 0.3178$, with $\beta = 0.29$ and $\delta = 1$, $\varepsilon = 0.3$. At time 0 the only infected node is node 1.

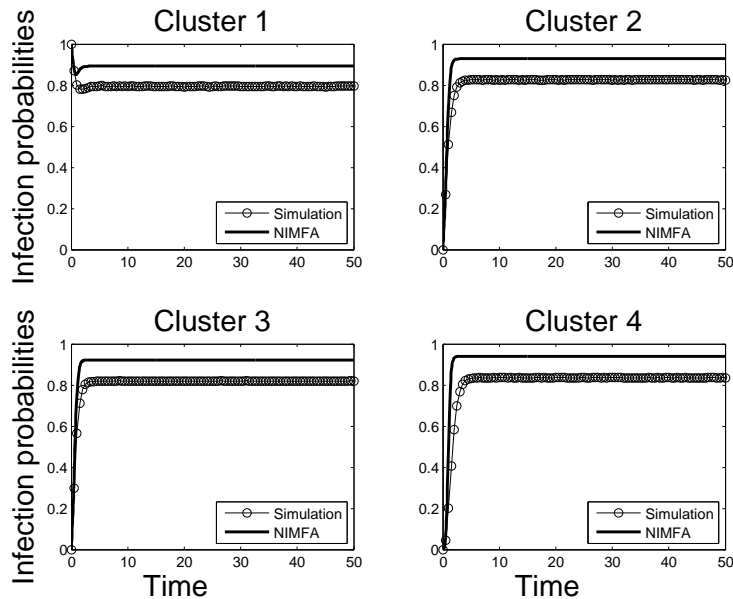


Fig. 4: Dynamics of infection probabilities for each community of the network in Fig.1: simulation versus numerical solutions of (4.6); $\tau = \beta/\delta > \tau_c^{(1)} = 0.3178$, with $\beta = 1.5$ and $\delta = 0.3$, $\varepsilon = 0.3$; initial conditions as in Fig. 3.

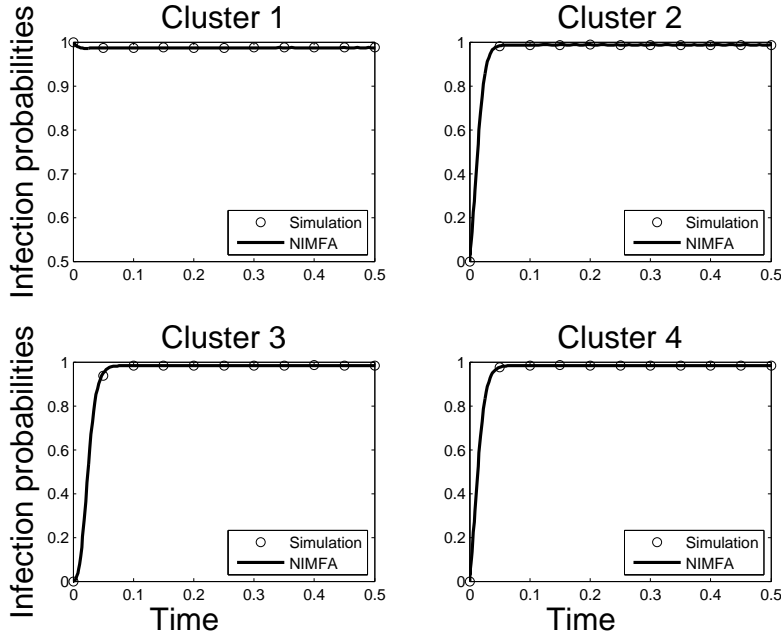


Fig. 5: Infection probabilities for each community in a network with $N = 80$, $d_{ii} = k_i - 1 = 19$ and $d_{ij} = 20$, for all $i, j = 1, \dots, 4$: simulation versus numerical solutions of (4.6); $\tau = \beta/\delta > \tau_c^{(1)} = 0.0348$, with $\beta = 5$ and $\delta = 2$, $\varepsilon = 0.3$; at time 0 all nodes of the 1-st community are infected.

5. Almost equitable partitions. In this section we consider graphs where the partition of the vertex set is *almost equitable*. Thus, we can relax the initial assumption on the regularity of the internal community structure implied by the definition of equitable partition.

DEFINITION 5.1. *The partition $\pi = \{V_1, \dots, V_n\}$ is called almost equitable if for all $i, j \in \{1, \dots, n\}$ with $i \neq j$, there is an integer d_{ij} such that for all $v \in V_i$, it holds*

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

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 $\{V_1, \dots, V_n\}$, 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}

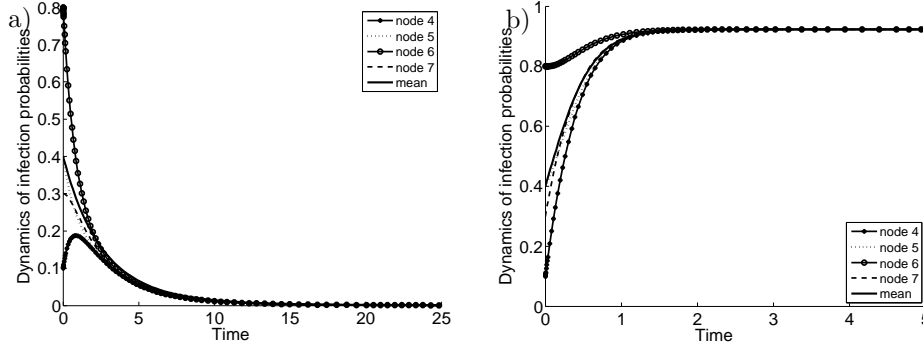


Fig. 6: Comparison between the dynamics of the original system (2.3) for each of the nodes belonging to V_3 in Fig. 1, for different initial conditions and the dynamics of the reduced system (4.6). In the latter case the initial conditions for each node are the mean value of the $p_i(0)$ s. a) case below the threshold: $\beta = 0.29$, $\delta = 1$, $\varepsilon = 0.3$ b) case above the threshold: $\beta = 1.5$, $\delta = 0.3$, $\varepsilon = 0.3$.

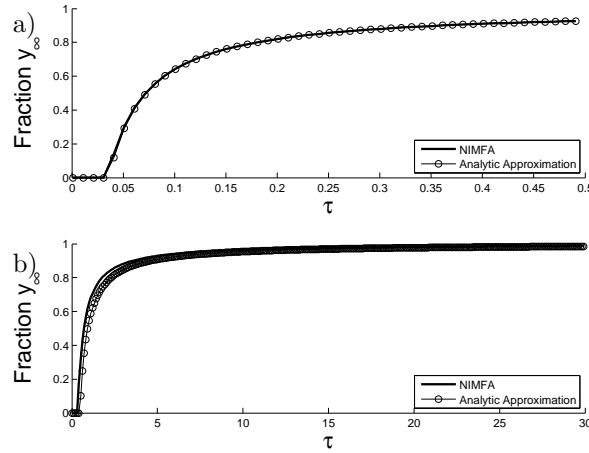


Fig. 7: Steady-state average fraction of infected nodes, for different values of τ : comparison between the approximation (4.10) and the exact computation (2.5); a) the graph is the one considered in Fig. 1 and b) the one considered in Fig. 5.

and A are the adjacency matrices of \tilde{G} and G respectively it holds

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

where $R = \text{diag}(R_1, \dots, R_n)$; the dimension of R_i is $k_i \times k_i$ for $i = 1, \dots, n$, 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

$$(5.1) \quad \lambda_1(\tilde{A}) \leq \lambda_1(A) + \lambda_1(R).$$

In the following we shall provide a more explicit formulation of the right hand side of (5.1) involving the number of added edges.

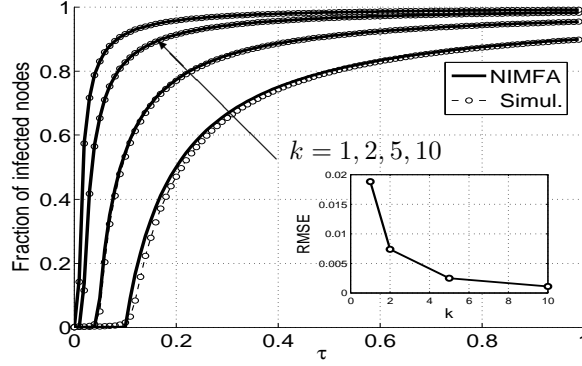


Fig. 8: Steady-state average fraction of infected nodes for different values of k and τ , for a graph of regular degree $d = 10$ and $N = 500$. The internal structure of each community is a clique. Both NIMFA and simulations are shown. The inserted plot represents the root mean square error between the simulated and the approximated fraction of infected nodes.

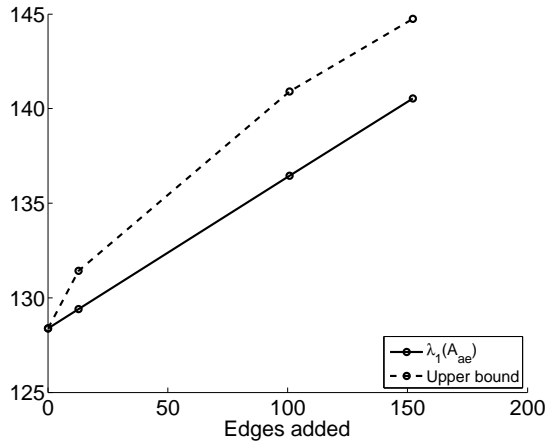


Fig. 9: Comparison of the bound and the spectral radius for a 40-communities network. Each community has $k = 25$ nodes, whose internal structure is initially a ring; the perturbation graph is obtained by adding in each of them the same increasing number of links. The spectral radius of the adjacency matrix \tilde{A} (A_{ae} in the legend, where the subscript “ae” stays for “almost equitable”) is compared to the upper bound as a function of the links added in each community.

PROPOSITION 5.2. *Let $G = (V, E)$ and $\tilde{G} = (V, \tilde{E})$ be two graphs and consider a partition $\{V_1, \dots, V_n\}$ 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 $i = 1, \dots, n$. 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 , $i = 1, \dots, n$, 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 $i = 1, \dots, n$, 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 .

Proof. (1) By assumption, R is a diagonal block matrix whose blocks R_i are the adjacency matrices of the induced subgraphs G_i^C . Thus, $\lambda_1(R)$ is the maximum of all spectral radii $\lambda_1(R_i)$. On the other hand, one has by [36, (3.45)] that

$$\lambda_1(R_i) \leq \min \left\{ \sqrt{\frac{2e_i(k_i - 1)}{k_i}}, \Delta(G_i^C) \right\}.$$

and the claim follows.

(2) By Gershgorin's theorem, the spectral radius of an adjacency matrix of a graph without loops is never larger than the graph's maximal degree, i.e., $\lambda_1(R_i) \leq \Delta(G_i^C)$. By assumption, there exists a permutation of the vertices in V_i such that the matrix R_i has the form

$$R_i = \begin{bmatrix} R'_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

where R'_i is the adjacency matrix of a connected graph with k'_i nodes (i.e., the block R'_i has dimension $k'_i \times k'_i$). Now, we deduce from [36, art. 50] that

$$\lambda_1(R'_i) \leq \sqrt{2e_i - k'_i - 1},$$

and since $\lambda(R_i) = \lambda(R'_i)$, the statement follows. \square

By using estimate (3.2) and Proposition 5.2 in the first and the second term on the right hand side of (5.1), respectively, we deduce

$$(5.2) \quad \lambda_1(\tilde{A}) \leq \max_{1 \leq i \leq n} \lambda_1(C_{V_i}) + \lambda_1(\hat{B}) + \max_{1 \leq i \leq n} \min \left\{ \sqrt{\frac{2e_i(k_i - 1)}{k_i}}, \Delta(G_i^C) \right\}.$$

The inequality in (5.2) gives us a lower bound for the epidemic threshold in the case of a graph whose partition of nodes set is almost equitable. Actually

$$(5.3) \quad \tau_c^{(1)} = \frac{1}{\lambda_1(\tilde{A})} \geq \tau^* = \frac{1}{\max_{1 \leq i \leq n} \lambda_1(C_{V_i}) + \lambda_1(\hat{B}) + \max_{1 \leq i \leq n} \min \left\{ \sqrt{\frac{2e_i(k_i - 1)}{k_i}}, \Delta(G_i^C) \right\}}.$$

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

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

The bounds developed so far support the design of community networks with safety region for the effective spreading rate, that guarantees the extinction of epidemics. E.g. if we consider some G_i , $i = 1, \dots, n$, 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 (5.3) still holds.

6. Conclusion. In this work we have discussed the relation between the epidemic threshold of a given graph with equitable partitioning of its node set, and the spectral properties of the corresponding quotient matrix. Because the quotient matrix Q has the same spectral radius of A , this may lead to a significative computational advantage in the calculation of $\lambda_1(A)$ and, consequently, of $\tau_c^{(1)}$, since the order of Q is smaller than that of A .

A novel expression has been derived for the lower bound on $\tau_c^{(1)}$ as function of network metrics, e.g., the maximum among the internal degrees of the nodes over all communities. In practice this value can be adopted to determine a safety region for the extinction of epidemics, i.e., by forcing the effective spreading rate below the lower bound; it can be also useful in order to design new network architectures robust to long-term, massive infections.

In the analysis, we have showed that it is possible to reduce the number of equations representing the time-change of infection probabilities using the quotient matrix Q , when all nodes belonging to the same community have the same initial conditions. After proving the existence of a positively invariant set for the original system of N differential equations, we have shown that the non-zero steady-state infection probabilities belongs to this invariant set, and that it can be computed by the reduced system of n equations.

Finally, we have also considered the case when the partition is almost equitable. An input graph whose partition is equitable can be perturbed, by adding or removing edges inside communities, in order to obtain a graph with an almost equitable partition. A lower bound for the epidemic threshold has been derived, and the effect of perturbations of the communities' structure has been explored.

Acknowledgments. The authors would like to thank Piet Van Mieghem for providing interesting and useful comments to an early draft of this work. The helpful comments of two anonymous reviewers are also gratefully acknowledged.

- [1] I. HANSKI AND O. OVASKAINEN *Metapopulation theory for fragmented landscapes*, Theoretical population biology, 64(1), 119-127, 2003.
- [2] N. MASUDA *Effects of diffusion rates on epidemic spreads in metapopulation networks*, PLoS ONE, 5(3), 2010.
- [3] L. J. S. ALLEN, B. M. BOLKER, Y. LOU AND A. L. NEVAI, *Asymptotic profiles of the steady states for an SIS epidemic patch model*, SIAM Journal on Applied Mathematics, 67(5), 1283-1309, 2007.
- [4] V. COLIZZA AND A. VESPIGNANI, *Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations*. J. Theoret. Biol., 251(3):450-467, 2008.
- [5] C. POLETTI, S. MELONI, V. COLIZZA, Y. MORENO, AND A. VESPIGNANI, *Host mobility drives pathogen competition in spatially structured populations*. PLoS Computational Biology, 9(8):e1003169, 2013.
- [6] P. VAN MIEGHEM, J. OMIC AND R. E. KOOLJ, *Virus Spread in Networks*, IEEE/ACM Transaction on Networking, 17:1-14, 2009.
- [7] P. VAN MIEGHEM AND R. VAN DE BOVENKAMP, *Non-Markovian infection spread dramatically alters the SIS epidemic threshold in networks*, Physical Review Letters, 110(10):108701, 2013.
- [8] E. CATOR, R. VAN DE BOVENKAMP, AND P. VAN MIEGHEM, *Susceptible-Infected-Susceptible epidemics on networks with general infection and curing times*, Physical Review E, 87(6):062816, 2013.
- [9] S. BONACCORSI, S. OTTAVIANO, F. DE PELLEGRINI, A. SOCIEVOLE AND P. VAN MIEGHEM, *Epidemic outbreaks in two-scale community networks*, Physical Review E, 90(1): 012810, 2014.
- [10] F. BALL, D. MOLLISON, AND G. SCALIA-TOMBA, *Epidemics with two levels of mixing*, The Annals of Applied Probability, 7(1):46-89, 1997.
- [11] F. G. BALL AND P. J. NEAL, *Network epidemic models with two levels of mixing*, Math. Biosci, 212:69-87, 2008.
- [12] J. V. ROSS, T. HOUSE, AND M. J. KEELING, *Calculation of disease dynamics in a population of households*. PLoS ONE, 5(3), 2010.
- [13] P. VAN MIEGHEM AND E. CATOR, *Epidemics in networks with nodal self-infection and the epidemic threshold*, Physical Review E, 86(1):016116, 2012.
- [14] P. VAN MIEGHEM, *The viral conductance of a network*, Computer Communications, 35(12):1494-1506, 2012.
- [15] P. VAN MIEGHEM, *The N-Intertwined SIS epidemic network model*, Springer Computing, 93(2):147-169, 2011.
- [16] P. K. POLLETT AND A. J. ROBERTS, *A description of the long-term behaviour of absorbing continuous-time Markov chains using a centre manifold*, Advances in applied probability, 111-128, 1990.
- [17] I. NASELL, *The quasi-stationary distribution of the closed endemic SIS model*, Advances in Applied Probability, 895-932, 1996.
- [18] I. NASELL, *Stochastic models of some endemic infections*, Mathematical biosciences, 179.1 : 1-19, 2002.
- [19] N.T.J. BAILEY, *The mathematical theory of infectious diseases and its applications*, Charles Griffin, London, 1975.
- [20] D.J. DALEY, AND J. GANI, *Epidemic Modelling: An Introduction*, Cambridge Univ. Press, Cambridge, 1999.
- [21] R. PASTOR-SATORRAS, AND A. VESPIGNANI, *Epidemic spreading in scale free networks*, Physical Review Letters, 86(14): 3200, 2001.
- [22] P. VAN MIEGHEM, *Decay towards the overall-healthy state in SIS epidemics on networks*, arXiv:1310.3980, 2014
- [23] P. K. POLLETT, *Quasi-stationary distributions: a bibliography*, <http://www.maths.uq.edu.au/~pkp/papers/qsds/qsds.pdf>, 2008.
- [24] M. DRAIEF AND L. MASSOULIÉ, *Epidemics and Rumours in Complex Networks*, London Math. Society Lecture Note Ser. 369, Cambridge University Press, Cambridge, 2010.
- [25] Y. WANG, D. CHAKRABARTI, C. WANG, AND C. FALOUTSOS, *Epidemic spreading in real networks: an eigenvalue viewpoint*, Proc. 22nd Int. Symposium on Reliable Distributed Systems, IEEE, 25-33, 2003
- [26] F. D. SAHNEH, C. SCOGGIO, AND P. VAN MIEGHEM, *Generalized epidemic mean-field model for spreading processes over multilayer complex networks*, Networking, IEEE/ACM Transactions on 21.5 :1609-1620, 2013
- [27] E. CATOR AND P. VAN MIEGHEM, *Nodal infection in Markovian SIS and SIR epidemics on networks are non-negatively correlated*, Physical Review E, 89(5): 052802, 2014.
- [28] P. VAN MIEGHEM, AND R. VAN DE BOVENKAMP, *Accuracy criterion for the mean-field approximation in SIS epidemics on networks*, Physical Review E, 91(3): 032812, 2015.

- [29] R. J. SANCHEZ-GARCÍA, E. C., AND Y. MORENO. *Dimensionality reduction and spectral properties of multilayer networks* Physical Review E, 89, 052815 (2014).
- [30] A.J. SCHWENK, *Computing the characteristic polynomial of a graph*, Graphs and Combinatorics (Proc. Washington D.C. 1973), Springer, New York, 153-172, 1974
- [31] P. VAN MIEGHEM, *Exact Markovian SIR and SIS epidemics on networks and an upper bound for the epidemic threshold*, arXiv:1402.1731, 2014
- [32] L. PERKO, *Differential Equations and Dynamical Systems*, Springer-Verlag, New York, 2001.
- [33] A. LAJMANOVICH AND J. A. YORKE , *A deterministic model for gonorrhea in a nonhomogeneous population*, Mathematical Biosciences 28.3, 221-236, 1976
- [34] R.A. HORN AND C.R. JOHNSON, *Topics in Matrix Analysis* SIAM, 1990.
- [35] C.D. GODSIL AND B.D. MCKAY. *Feasibility conditions for the existence of walk-regular graphs*, Linear Algebra and its Applications, 30: 15-61, 1980.
- [36] P. VAN MIEGHEM. *Graph Spectra for Complex Networks*. Cambridge University Press, Cambridge, 2011.