Nodal infection in Markovian susceptible-infected-susceptible and susceptible-infected-removed epidemics on networks are non-negatively correlated

E. Cator[1,*] and P. Van Mieghem[2,†]

[1]*Faculty of Science, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands*
[2]*Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Delft, The Netherlands*

By invoking the famous Fortuin, Kasteleyn, and Ginibre (FKG) inequality, we prove the conjecture that the correlation of infection at the same time between any pair of nodes in a network cannot be negative for (exact) Markovian susceptible-infected-susceptible (SIS) and susceptible-infected-removed (SIR) epidemics on networks. The truth of the conjecture establishes that the N-intertwined mean-field approximation (NIMFA) upper bounds the infection probability in any graph so that network design based on NIMFA always leads to safe protections against malware spread. However, when the infection or/and curing are *not* Poisson processes, the infection correlation between two nodes can be negative.

I. INTRODUCTION

Epidemic processes on graphs are worth studying for a couple of reasons. First of all, the possible applications are numerous: for example, from biological disease spreading in contact networks [1,2] to information propagation in communications networks [3,4] and brain networks [5]; economic transactions in sector networks [6], security issues and spreading of opinions or sentiments in social networks [7]. Second, epidemics are described by local rules that give rise to nontrivial global behavior, of which a phase transition or threshold behavior is perhaps the most fascinating [8]. The simple epidemic models, susceptible-infected-susceptible (SIS) and susceptible-infected-removed (SIR), on any finite network can be described exactly [9], when we assume that the infection and curing processes are independent Poisson processes so that Markov theory applies. Apart from random walks on a graph, these stochastic Markovian SIS and SIR models are about the simplest processes that allow us to understand the influence of the topology on the dynamic epidemic process. Understanding the interplay between processes *on* and the topology *of* a network is a major goal of network science, which can lead to improved and/or new design rules for networks.

Here, we will mainly consider SIS epidemics on a general finite graph $G = (V,E)$. For each node $i \in V$ and time t, we describe by $X_i(t)$ the infectious state at node i, where $X_i(t) = 0$ means not infected, and $X_i(t) = 1$ means infected. Earlier [4], it was conjectured that $X_i(t)$ and $X_j(t)$ are non-negatively correlated, when starting from a deterministic initial state:

$$E[X_i(t)X_j(t)] \geqslant E[X_i(t)]E[X_j(t)], \qquad (1)$$

which is, by the Bernoulli nature of X, equivalent to

$$\Pr[X_i(t) = 1, X_j(t) = 1] \geqslant \Pr[X_i(t) = 1]\Pr[X_j(t) = 1]. \qquad (2)$$

In terms of conditional probabilities, (2) becomes

$$\Pr[X_i(t) = 1 | X_j(t) = 1] \geqslant \Pr[X_i(t) = 1],$$

which means that an infection at some node j in G will not decrease the probability of infection of node i.

Although the simple Markovian SIS and SIR model on networks can be described exactly, their numerical solution is prohibitive, which has naturally led to mean-field approaches. For continuous-time Markovian SIS epidemics on networks, the heterogeneous mean-field (HMF) approximation [10] and the N-intertwined mean-field approximation (NIMFA) [4,11] are the most important. In particular, NIMFA only makes one approximation: absence of correlation between the infectious states of two nodes in the network, i.e., assuming equality in (2). A similar neglect of correlations is implicitly made in HMF [12]. An important consequence of this independence assumption and the inequalities (1, 2) is that NIMFA always upper bounds the probability of infection of each node in the network and hence lower bounds the epidemic threshold. From a practical point of view, when designing [13] or controlling [14,15] a network against the epidemics by using NIMFA, the upper bound property ensures that the network is safeguarded from long-term, massive infection.

Apart from second-order mean-field studies [16–18], in which the joint probabilities $\Pr[X_i(t) = 1, X_j(t) = 1]$ appear, we found that few results on the correlation $E[X_iX_j]$ have been published (see, e.g., the recent review [8]). Barrat *et al.* [19] found that the average number of passengers w_{ij} traveling between any two airports i and j could be predicted by $w_{ij} \sim \sqrt{d_id_j}$, where d_i is the degree of node i. Since transport can be associated with a diffusion process, the strength of epidemic "traffic" $E[X_iX_j]$ between node i and j could be similarly interpreted as the weights w_{ij} in the air transportation network.

Although the conjecture (1, 2) seems natural and intuitive, since more infected nodes lead to more infections, recently concerns have been raised that this positive correlation might not be true for certain graphs. In this paper, we will rigorously prove that $X_i(t)$ and $X_j(t)$ are non-negatively correlated for all finite graphs and that the inequalities (1) and (2) are correct at any time t and for any node pair (i,j) in the graph. The

*e.cator@science.ru.nl
†p.f.a.vanmieghem@tudelft.nl

proof will rely on the so called FKG inequality (Fortuin, Kasteleyn, and Ginibre [20]). In Sec. III, we introduce the necessary notation for the FKG inequalities and prove (1), but we will start by introducing a different way to view the SIS process in Sec. II A. Then we approximate the process through a time discretization in Sec. II B. Section IV A exemplifies a simple graph with non-Markovian epidemics [21,22], in the sense that the infection and curing processes are not Poisson processes anymore, but general renewal processes, for which there can be a negative correlation. Thus, non-Markovian epidemic processes on graphs can violate the inequalities (1) and (2). However, the example in Sec. IV A illustrates that a negative correlation requires a very special construction and we are tempted to believe that, in general, infection correlations between node pairs in a graph will be non-negative.

II. DIFFERENT APPROACH TO THE SIS AND SIR MODEL

To apply the FKG inequalities, stated in Sec. III, it is essential that we do *not* consider the SIS epidemics as a process in time.

A. Another description of the epidemic process

Our description is inspired by the graphical construction introduced in [23] by Harris, which is used many times in the interacting particle processes literature. The idea is that we consider one probability measure on a large space that keeps track of all the events in our process, at all times. These events will be the infecting attempts made by diseased vertices and the healing events of diseased vertices. To code all these events, we also need to realize that an infecting event takes place along an edge of our graph, but it also has a direction: from the diseased vertex to the healthy one. We therefore define the set of directed edges

$$F = \{(i,j) \in V \times V \mid \{i,j\} \in E\},$$

and we define

$$W = (V,F),$$

which signifies the collection of all vertices and all directed edges in our graph.

At each time s there can be infecting or healing events on different elements of (V,F): healing events on vertices in V, and infecting events on directed edges in F. The trick is to determine beforehand whether these events at all possible times will take place or not, and then check afterwards whether a healing event on the vertex i actually heals this vertex (this happens if i is infected at the considered time s), or whether an infecting event on a directed edge $e = (i,j)$ actually infects a new site (this happens if at the considered time s the vertex i is infected and the vertex j is healthy). Let us introduce the "time line" $\{i\} \times [0,t]$ for each vertex i and the time line $\{e\} \times [0,t]$ for each directed edge $e = (i,j)$. On these time lines, Poisson processes are active that generate events: healing events on each $\{i\} \times [0,1]$ with intensity δ_i, possibly depending on the vertex i, and infection events from vertex i to vertex j with intensity β_e, possibly depending on the directed edge e. We denote the space of all finite subsets of $[0,t]$ by

$$\mathcal{M} = \{A \mid A \subset [0,t], A \text{ is finite}\}.$$

Each A represents a finite set of events on $[0,t]$. Now, we define a function

$$Z : W \to \mathcal{M},$$

which has the following meaning: if $s \in Z(e)$ for an edge $e = (i,j) \in F$, then at time s the vertex i will attempt to infect the vertex j. Of course, this will only happen if i is infected at time s. If $s \notin Z(e)$, then the vertex i will certainly not try to infect the vertex j at time s. Similarly, if $s \in Z(i)$ for a vertex $i \in V$, then the vertex i will heal at time s.

Once we have chosen this function Z, we can use it to trace the evolution of the state of the vertices in our SIS model: suppose at time 0 the state of all vertices is given by a vector $X(0) = (X_1(0), \dots, X_N(0)) \in \{0,1\}^N$. Now we wish to know what is the state of all vertices at time s, given the state just before time s. We consider the collection of all times of all events:

$$T = (\cup_{i \in V} Z(i)) \bigcup (\cup_{e \in F} Z(e)).$$

This is a finite collection of times, and we can order them: there exists some $K \in \mathbb{N}$ (possibly 0) with

$$T = \{t_1, t_2, \dots, t_K\} \quad \text{with} \quad t_1 < t_2 < \cdots < t_K.$$

Define $t_0 = 0$ and $t_{K+1} = t$ (these two time points are not elements of T, with probability 1). If our time s satisfies $t_k \leqslant s < t_{k+1}$ for $0 \leqslant k \leqslant K$, or if $s = t_k$ for $k = K$, then we simply put $X(s) = X(t_k)$. Now suppose $s \in T$. Then either there exists a vertex i such that $s \in Z(i)$, or there exists an edge $e = (i,j)$ such that $s \in Z(e)$. If $s \in Z(i)$, then $X(s) = X(t_{k-1})$, except that $X_i(s) = 0$ (vertex i heals at time s). If $s \in Z(e)$, then $X(s) = X(t_{k-1})$, except that $X_j(s) = \max(X_i(t_{k-1}), X_j(t_{k-1}))$ (vertex j is either already infected, or it gets infected if vertex i is infected).

For the SIR model, we could use the same function Z: the only difference is that if an infected vertex heals at time s, its state is changed to the R state: resistant or recovered or removed, in which the node stays for the rest of the time: it neither gets infected nor infects nor heals after time s. The description also provides a natural way of coupling the SIS and SIR models, exhibiting the nice property that, if initially the infected nodes in the SIR epidemics are a subset of the infected nodes in the SIS epidemics, then they remain a subset for all times. We will concentrate on SIS in this paper, but this remark shows that our conclusion is also valid for SIR.

So far, we have described the entire evolution of the disease through all times in $[0,t]$. In the sequel, we need to prescribe the distribution of the random function Z,

$$Z \in \mathcal{X} := \mathcal{M}^W,$$

which requires us to define a probability measure on the space $\mathcal{X}$. Our choice is somehow obvious: all infecting and healing events are independent of each other, and we have the following marginal distributions:

$\forall\, i \in V : Z(i)$ is a Poisson process on $[0,t]$ with intensity δ_i

and

$\forall\, e = (i,j) \in F : Z(e)$ is a Poisson process on $[0,t]$

with intensity β_e.

This defines a probability measure μ on $\mathcal{X}$, which captures all the randomness of the evolution of the SIS process.

B. Discretization of the SIS process

To avoid technicalities (we would need to extend the FKG inequalities to nonfinite partially ordered spaces, for example), and to use a specific kind of ordering, we want to work on finite spaces. Therefore, it is necessary that we discretize time, so we choose n equally spaced time points $t_1, \ldots, t_n$ in the interval $[0,t]$ and the kth time point is $t_k = kt/n$. We will define a discrete SIS process that will tend to the true process for $n \to \infty$.

In the continuous-time description in Sec. II A, we defined a function Z that assigned a Poisson process on $[0,t]$ of healing events to each vertex i and a Poisson process of infecting events to each directed edge e. A natural way to discretize is the following: we define a function Z_n such that for each vertex i and time t_k we have

$$Z_n(i,t_k) = \begin{cases} 1 & \text{if } Z(i) \cap (t_{k-1},t_k] = \emptyset, \\ 0 & \text{if } Z(i) \cap (t_{k-1},t_k] \neq \emptyset. \end{cases}$$

Furthermore, for each directed edge e we define

$$Z_n(e,t_k) = \begin{cases} 0 & \text{if } Z(e) \cap (t_{k-1},t_k] = \emptyset, \\ 1 & \text{if } Z(i) \cap (t_{k-1},t_k] \neq \emptyset. \end{cases}$$

Note that the coding for a healing event $[Z(i,t_k) = 0$ if there is an healing event at time t_k at vertex $i]$ is different from the coding for an infection event $[Z(i,t_k) = 1$ if there is an infecting event at time t_k at vertex $i]$. The reason is that infecting leads to *more* infected nodes, whereas healing leads to *less* infected nodes; we will return to this difference soon.

For each time t_k, the function

$$Z_n : W \times \{t_1, \ldots, t_n\} \mapsto \{0,1\}$$

determines for each directed edge e whether there is an infection event or not at time t_k, and for each vertex i whether there is an healing event at time t_k or not. If there are two or more events for the continuous-time process in a small interval $(t_{k-1},t_k]$, then we replace these multiple events in the discrete version by one event. There are two justifications why this replacement is a sensible approximation. First, as $n \to \infty$, the probability that there exist more than one event in a small interval decreases as $1/n$. Furthermore, two infection events on the same edge in a small time interval lead to the same configuration as only one infecting event (with high probability), and the same is true for healing events.

We describe the state at time t_k by a vector $X^{(n)}(t_k) \in \{0,1\}^N$. The function Z_n tells us which vertices heal at time t_k and which vertices are trying to infect a particular set of other vertices at time t_k. The only problem is that we have to decide what happen first, the healing or the infecting. This problem does not occur in the continuous version, since with probability 1 all events happen at distinct times. We choose to start with the healing: define the intermediate state vector $Y^{(n)}(t_k) \in \{0,1\}^N$ as equal to $X^{(n)}(t_{k-1})$, except for all vertices i with $Z(i,t_k) = 0$, where we have that $Y_i^{(n)}(t_k) = 0$. The governing equation is

$$\forall 1 \leqslant i \leqslant N : Y_i^{(n)}(t_k) = X_i^{(n)}(t_{k-1})Z_n(i,t_k). \quad (3)$$

Then we deal with the infecting events: we define $X(t_k)$ equal to $Y(t_k)$, except for vertices which are infected at time t_k:

$$\forall 1 \leqslant i \leqslant N : X_i^{(n)}(t_k) = Y_i^{(n)}(t_k)$$
$$+ \left[1 - Y_i^{(n)}(t_k)\right] \max_{j:(j,i)\in F} Y_j^{(n)}(t_k)Z_n((i,j),t_k). \quad (4)$$

Equation (4) guarantees that i is infected at time t_k if either i was still infected after the healing events at time t_k $[Y_i^{(n)}(t_k) = 1]$, or one of the neighbors j of i was still infected after the healing events at time t_k and $Z((j,i),t_k) = 1$, which means that there is an infection attempt at time t_k from j to i. We assume that a vertex which becomes infected at time t_k cannot infect another site at the same time.

In this way we have described the entire evolution of the disease through all times $t_1, \ldots, t_n$. Just as in the continuous-time case above, we need to describe the distribution of the random function Z_n. We define

$$W_n = W \times \{t_1, \ldots, t_n\},$$

so

$$Z \in \mathcal{X}_n = \{0,1\}^{W_n}.$$

Therefore, we need to define a probability measure on the space $\mathcal{X}_n$. Our choice is fixed by the Poisson processes of the continuous-time model: all infecting and healing events are independent of each other, and we have the following marginal probabilities:

$$\forall i \in V, k \in \{1, \ldots, n\} : \mathbb{P}(Z(i,t_k) = 0) = 1 - e^{-\delta_i t/n}$$

and

$$\forall e = (i,j) \in F, k \in \{1, \ldots, n\} : \mathbb{P}(Z(e,t_k) = 1) = 1 - e^{-\beta_e t/n}.$$

This defines a probability measure μ_n on $\mathcal{X}_n$. When $n \to \infty$, the healing events on a given vertex i become a Poisson process in time with intensity δ_i, and the infection events on the directed edge e become a Poisson process in time with intensity β_e, just as in the continuous-time SIS model.

After this rather lengthy preparation, we are ready to point out the importance of the above discrete-time description of the SIS model. The space $\mathcal{X}_n$, on which we have defined our probability measure μ_n, has a natural partial ordering: if $Z_1, Z_2 \in \mathcal{X}_n$, so Z_1 and Z_2 are both functions from W_n to $\{0,1\}$, then we say that $Z_1 \leqslant Z_2$ if

$$\forall i \in V, k \in \{1, \ldots, n\} : Z_1(i,t_k) \leqslant Z_2(i,t_k)$$

and

$$\forall e \in F, k \in \{1, \ldots, n\} : Z_1(e,t_k) \leqslant Z_2(e,t_k).$$

This is just the pointwise ordering of two functions, which has the following special property. Suppose we fix an initial condition $X^{(n)}(0)$. Then the two functions $Z_1 \leqslant Z_2$ both lead to different evolutions of the SIS process, in other words, to different values of the state at time $t_n = t$. To make the dependence on Z explicit, we introduce the slightly cumbersome notation $X^{(n),Z_1}(t_k)$ and $X^{(n),Z_2}(t_k)$. Due to our choice of coding, we can now conclude that for any $i \in V$, we have

$$X_i^{(n),Z_1}(t) \leqslant X_i^{(n),Z_2}(t).$$

Indeed, suppose $Z_1 \leqslant Z_2$ and $X^{(n),Z_1}(t_{k-1}) \leqslant X^{(n),Z_2}(t_{k-1})$ (note that this is true for $k = 1$, defining $t_0 = 0$). For each vertex i, we then know that $X_i^{(n),Z_1}(t_{k-1}) \leqslant X_i^{(n),Z_2}(t_{k-1})$ and $Z_1(i,t_k) \leqslant Z_2(i,t_k)$. By using (3), we see that

$$Y_i^{(n),Z_1}(t_k) = X_i^{(n),Z_1}(t_{k-1})Z_1(i,t_k) \leqslant X_i^{(n),Z_2}(t_{k-1})Z_2(i,t_k)$$
$$= Y_i^{(n),Z_2}(t_k).$$

For each directed edge $e = (i,j)$ we also know that $Z_1(e,t_k) \leqslant Z_2(e,t_k)$. From (4), it follows that

$$X_i^{(n),Z_1}(t_k) = Y_i^{(n),Z_1}(t_k) + \left[1 - Y_i^{(n),Z_1}(t_k)\right]$$
$$\times \max_{j:(j,i)\in F} Y_j^{(n),Z_1}(t_k)Z_1((i,j),t_k).$$

If $Y_i^{(n),Z_1}(t_k) = 1$, then we have seen that also $Y_i^{(n),Z_2}(t_k) = 1$, so $X_i^{(n),Z_1}(t_k) = X_i^{(n),Z_2}(t_k) = 1$. Now suppose that $Y_i^{(n),Z_1}(t_k) = 0$. If $Y_i^{(n),Z_2}(t_k) = 1$, then $X_i^{(n),Z_2}(t_k) = 1 \geqslant X_i^{(n),Z_1}(t_k)$. If also $Y_i^{(n),Z_2}(t_k) = 0$, then

$$X_i^{(n),Z_1}(t_k) = \max_{j:(j,i)\in F} Y_j^{(n),Z_1}(t_k)Z_1((i,j),t_k)$$
$$\leqslant \max_{j:(j,i)\in F} Y_j^{(n),Z_2}(t_k)Z_2((i,j),t_k) = X_i^{(n),Z_2}(t_k).$$

This proves that $X^{(n),Z_1}(t_k) \leqslant X^{(n),Z_2}(t_k)$, for all $1 \leqslant k \leqslant n$. In short, $X_i^{(n),Z}(t)$ and $X_j^{(n),Z}(t)$, regarded as functions of Z, are increasing functions on the partially ordered set $\mathcal{X}_n$.

III. THE FKG INEQUALITY

To introduce the FKG inequality, we will need some concepts from order theory. A partially ordered space $\mathcal{X}$ is called a *lattice* if each pair $x,y \in \mathcal{X}$ has a supremum $(x \vee y)$ and an infimum $(x \wedge y)$. The supremum $x \vee y$ is defined by the property that if $x \leqslant z$ and $y \leqslant z$, then we must have $x \vee y \leqslant z$; the infimum has an analogous definition. Our partially ordered space $\mathcal{X}_n$, introduced in Sec. II B, satisfies the lattice condition. Indeed, if $Z_1, Z_2 \in \mathcal{X}_n$, then both are functions of W_n to $\{0,1\}$. Now take the pointwise maximum and minimum of these two functions on W_n to find $Z_1 \vee Z_2$ and $Z_1 \wedge Z_2$. Now consider a lattice $\mathcal{X}$ and a positive function μ on $\mathcal{X}$ such that

$$\forall\, x,y \in \mathcal{X}\,:\, \mu(x \vee y)\mu(x \wedge y) \geqslant \mu(x)\mu(y). \tag{5}$$

We will show that our probability measure μ_n on $\mathcal{X}_n$ satisfies this condition. The FKG inequality then states that for any two increasing functions f and g on $\mathcal{X}$, we have

$$\left[\sum_{x\in\mathcal{X}} f(x)g(x)\mu(x)\right]\left[\sum_{x\in\mathcal{X}}\mu(x)\right]$$
$$\geqslant \left[\sum_{x\in\mathcal{X}} f(x)\mu(x)\right]\left[\sum_{x\in\mathcal{X}} g(x)\mu(x)\right]. \tag{6}$$

If μ is actually a probability measure on $\mathcal{X}$, then inequality (6) can be rewritten as

$$E_\mu[f(X)g(X)] \geqslant E_\mu[f(X)]E_\mu[g(X)]. \tag{7}$$

In other words, the two increasing functions f and g must be non-negatively correlated. Applying (7) to the two increasing functions $X_i^{(n),Z}(t)$ and $X_j^{(n),Z}(t)$ then shows that these are

non-negatively correlated. Taking the limit $n \to \infty$ finishes the proof of (1) that for any finite graph and any fixed initial condition, we have that $X_i(t)$ and $X_j(t)$ are non-negatively correlated at time t.

All that remains is to check condition (5) for our measure μ_n. We can use the independence to see that

$$\mu_n(Z) = \prod_{i\in V}\prod_{k=1}^n \mu_{n,i}^H(Z(i,t_k))\prod_{e\in F}\prod_{k=1}^n \mu_{n,e}^I(Z(e,t_k)).$$

Here, the measure $\mu_{n,i}^H$ corresponds to the healing probability of vertex i, so

$$\mu_{n,i}^H(0) = 1 - e^{-\delta_i t/n} \quad \text{and} \quad \mu_n^H(1) = e^{-\delta_i t/n},$$

whereas the measure $\mu_{n,e}^I$ corresponds to the infecting probability of the directed edge e, so

$$\mu_{n,e}^I(1) = 1 - e^{-\beta_e t/n} \quad \text{and} \quad \mu_n^I(0) = e^{-\beta_e t/n}.$$

Now consider two functions Z_1 and Z_2. For any node $i \in V$ and $k \in \{1,\ldots,n\}$, we have

$$(Z_1 \vee Z_2)(i,t_k) = \max(Z_1(i,t_k),Z_2(i,t_k))$$

and

$$(Z_1 \wedge Z_2)(i,t_k) = \min(Z_1(i,t_k),Z_2(i,t_k)),$$

and

$$\prod_{i\in V}\prod_{k=1}^n \mu_{n,i}^H((Z_1 \vee Z_2)(i,t_k))\prod_{i\in V}\prod_{k=1}^n \mu_{n,i}^H((Z_1 \wedge Z_2)(i,t_k))$$
$$= \prod_{i\in V}\prod_{k=1}^n \mu_{n,i}^H(Z_1(i,t_k))\prod_{i\in V}\prod_{k=1}^n \mu_{n,i}^H(Z_2(i,t_k)),$$

simply because the same factors appear in both products. By a similar argument, we have

$$\prod_{e\in F}\prod_{k=1}^n \mu_{n,e}^I((Z_1 \vee Z_2)(e,t_k))\prod_{e\in F}\prod_{k=1}^n \mu_{n,e}^I((Z_1 \wedge Z_2)(e,t_k))$$
$$= \prod_{e\in F}\prod_{k=1}^n \mu_{n,e}^I(Z_1(e,t_k))\prod_{e\in F}\prod_{k=1}^{n,e} \mu_{n,e}^I(Z_2(e,t_k)),$$

from which we can indeed conclude that

$$\mu_n(Z_1 \vee Z_2)\mu_n(Z_1 \wedge Z_2) = \mu_n(Z_1)\mu_n(Z_2),$$

proving that μ_n satisfies (5).

IV. COUNTEREXAMPLES OF INEQUALITY (1)

A. When events are *not* generated by a Poisson process

The SIS model can be extended to the case where the healing events on each vertex and the infecting events on each directed edge are *arbitrary* processes on $[0,t]$, instead of Poisson processes. We will construct an example, due to Henk Don, where these processes are renewal processes, but the correlation between $X_1(t)$ and $X_2(t)$ is in fact negative, for two vertices 1 and 2.

Consider the graph with two vertices 1 and 2, and one edge between them. On both vertices, we define the following

renewal process: for each vertex $i \in \{1,2\}$, consider an i.i.d. sequence $T_1^{(i)}, T_2^{(i)}, \ldots$, where

$$\Pr\left[T_1^{(i)} = 0.6\right] = 0.5 \quad \text{and} \quad \Pr\left[T_1^{(i)} = 0.8\right] = 0.5,$$

which means that the healing of node i occurs either at time 0.6 or at time 0.8, but definitely at one of these times. For each vertex i, the healing events occur at times $T_1^{(i)}, T_1^{(i)} + T_2^{(i)}, \ldots$. Thus, there will always be exactly one healing event on $[0,1]$, for both vertices.

Now, we define the process of infecting events on the directed edges $e_1 = (1,2)$ and $e_2 = (2,1)$. For each directed edge e, we consider the i.i.d. sequences $S_1^{(e)}, S_2^{(e)}, \ldots$, where

$$S_1^{(e)} \sim U([0.65, 0.75]),$$

and where $U([a,b])$ denotes the uniform distribution on $[a,b]$. The infecting events for each directed edge occur at times $S_1^{(e)}, S_1^{(e)} + S_2^{(e)}, \ldots$. Our construction guarantees that there will be exactly one infection event on a directed edge e_1 (and similarly e_2) in the interval $[0,1]$, provided node 1 (and similarly node 2) is infected before time 0.65. More precisely, this infection event will be in the time interval $(0.65, 0.75)$, hence, always in between the two possible healing events, either at time 0.6 or at time 0.8.

As initial situation, we choose $X_1(0) = X_2(0) = 1$, so that both vertices are infected at time 0. Let us evaluate whether both nodes can be still infected at time 1, in other words, whether the event $X_1(1) = X_2(1) = 1$ is possible. This would be only possible if both vertices are not healed at time 0.8. However, not-healing at time 0.8 implies that both are healed at time 0.6 and that the infection is eradicated from the network, so at time 1 both vertices are still healthy. Hence, we arrive at $\Pr[X_1(1) = X_2(1) = 1] = 0$, which is equivalent to

$$E[X_1(1)X_2(1)] = 0.$$

We can conclude that

$$E[X_1(1)X_2(1)] - E[X_1(1)]E[X_2(1)] < 0,$$

if we can show that $E[X_1(1)] = \Pr[X_1(1) = 1] > 0$ (since by symmetry this would imply that also $E[X_2(1)] = \Pr[X_2(1) = 1] > 0$). Being infected at time 1 is only possible if there is no healing at time 0.8, which implies that there must be a healing at time 0.6, followed by an infection during the time interval $(0.65, 0.85)$ from the other infected node. Now, if $T_1^{(1)} = 0.6$ and $T_1^{(2)} = 0.8$, then vertex 1 will first be healed at time 0.6, but then will be infected by vertex 2 in the time interval $(0.65, 0.85)$, and therefore will still be infected at time

$t = 1$, thus

$$\Pr[X_1(1) = 1] = \Pr\left[T_1^{(1)} = 0.6\right] \Pr\left[T_1^{(2)} = 0.8\right] = \tfrac{1}{4} > 0.$$

The example shows that even for this simple graph, negative correlations cannot be excluded for non-Poisson processes. However, the example requires a very specific construction; in general, we do expect positive correlations.

B. Infection and curing rates are *not* constant

Another type of example, where the inequality (1) may be false, occurs in epidemic processes in which the infection rate and curing rate are themselves random. When the infection rate is node dependent and negatively correlated with neighboring infection rates, then negative correlations of site infections are possible as shown in [24], even with Poisson infection events. This negative correlation occurs when we do not know the realized infection rates: seeing an infected site can imply a larger infection rate, which in turn implies a lower infection rate at a neighboring site (due to the built-in negative correlation), which in turn can imply a healthy neighboring site. In our SIR and SIS setting, all rates (i.e., the infection rate β_e of link e and the curing rate δ_i of node i) are supposed to be fixed beforehand, and not random.

V. CONCLUSION

Although the inequalities (1) and (2) are natural and intuitive, their correctness for Markovian SIS (and SIR) epidemics in any graph and at any time is rigorously demonstrated. We consider the truth of those inequalities important, because a deeper understanding of the validity and accuracy of widely adopted mean-field approximations for epidemics on networks, that assume equality in (1) and (2), necessitates the involvement of correlations. Recent second-order mean-field approximations [16–18], that improve first-order mean-field approximations by incorporating the epidemic pair correlation $E[X_i X_j]$ after assuming a choice of closure, may be improved applying the non-negative correlation property (1). For non-Markovian epidemics, where infection and/or healing are still independent, but not Poisson processes anymore, the inequalities (1) and (2) can be violated, emphasizing the special role of Poissonian infection and healing processes in epidemics.

[1] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1991).

[2] O. Diekmann, H. Heesterbeek, and T. Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics* (Princeton University Press, Princeton, 2012).

[3] A. Ganesh, L. Massoulié, and D. Towsley, in *IEEE INFOCOM 2005*, edited by K. Makki and E. Knightly, Vol. 2 (IEEE, Piscataway, NJ, 2005), p. 1455.

[4] P. Van Mieghem, J. Omic, and R. E. Kooij, IEEE/ACM Trans. Netw. **17**, 1 (2009).

[5] C. J. Stam and J. C. Reijneveld, Nonlinear Biomed. Phys. **1**, 1 (2007).

[6] H. Wang, E. van Boven, A. Krishnakumar, M. Hosseini, T. Takema, H. van Hooff, N. Baken, and P. Van Mieghem, J. Adv. Complex Syst. **14**, 691 (2011).

[7] J. H. Fowler, J. E. Settle, and N. A. Christakis, Proc. Natl. Acad. Sci. USA **108**, 1993 (2011).

[8] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani (unpublished).

[9] P. Van Mieghem, Exact Markovian SIR and SIS Epidemics on Networks and an Upper Bound for the Epidemic Threshold, http://www.nas.ewi.tudelft.nl/people/Piet/TUDelftReports; arXiv:1402.1731.

[10] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).

[11] P. Van Mieghem, Computing **93**, 147 (2011).

[12] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 066117 (2001).

[13] P. Van Mieghem, D. Stevanović, F. A. Kuipers, C. Li, R. van de Bovenkamp, D. Liu, and H. Wang, Phys. Rev. E **84**, 016101 (2011).

[14] J. Omic, P. Van Mieghem, and A. Orda, in *IEEE INFOCOM 2009* (IEEE, Piscataway, NJ, 2009).

[15] E. Gourdin, J. Omic, and P. Van Mieghem, in *8th International Workshop on Design of Reliable Communication Networks (DRCN 2011), Krakow, Poland, October 10-12, 2011* (IEEE, Piscataway, NJ, 2011).

[16] J. P. Gleeson, Phys. Rev. Lett. **107**, 068701 (2011).

[17] E. Cator and P. Van Mieghem, Phys. Rev. E **85**, 056111 (2012).

[18] A. S. Mata and S. C. Ferreira, Europhys. Lett. **103**, 48003 (2013).

[19] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. USA **101**, 3747 (2004).

[20] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre, Commun. Math. Phys. **22**, 89 (1971).

[21] P. Van Mieghem and R. van de Bovenkamp, Phys. Rev. Lett. **110**, 108701 (2013).

[22] E. Cator, R. van de Bovenkamp, and P. Van Mieghem, Phys. Rev. E **87**, 062816 (2013).

[23] T. E. Harris, Ann. Probab. **6**, 355 (1978).

[24] T. P. Handford, F. J. Pérez-Reche, S. N. Taraskin, L. da F. Costa, M. Miazaki, F. M. Neri, and C. A. Gilligan, J. R. Soc. Interface **8**, 423 (2011).