Exact Network Reconstruction from Complete SIS Nodal State Infection Information Seems Infeasible

Prasse, Bastian; Van Mieghem, Piet

DOI
10.1109/TNSE.2018.2872511

Publication date
2019

Document Version
Accepted author manuscript

Published in
IEEE Transactions on Network Science and Engineering

Citation (APA)
Prasse, B., & Van Mieghem, P. (2019). Exact Network Reconstruction from Complete SIS Nodal State Infection Information Seems Infeasible. *IEEE Transactions on Network Science and Engineering*, *6*(4), 748-759. Article 8476203. https://doi.org/10.1109/TNSE.2018.2872511

Exact Network Reconstruction from Complete SIS Nodal State Infection Information Seems Infeasible

Bastian Prasse and Piet Van Mieghem

Abstract—The SIS dynamics of the spread of a virus crucially depend on both the network topology and the spreading parameters. Since neither the topology nor the spreading parameters are known for the majority of applications, they have to be inferred from observations of the viral spread. We propose an inference method for both topology and spreading parameters based on a maximum-a-posteriori estimation approach for the sampled-time Markov chain of an SIS process. The resulting estimation problem, given by a mixed-integer optimisation problem, results in exponential computational time if a brute-force approach is employed. By introducing an efficient and accurate, polynomial-time heuristic, the topology of the network can almost always be exactly reconstructed. Notwithstanding, reconstructing the network with a reasonably high accuracy requires a subexponentially increasing number of observations and an exponentially increasing computation time with respect to the number of nodes N. Such long observation periods are hardly realistic, which justifies the claim in the title.

Index Terms—SIS Process, Network Reconstruction, Spreading Parameter Estimation, Bayesian Estimation

1 Introduction

Epidemics on networks received much attention in recent years [1]. Modern epidemics is a genuinely interdisciplinary field and incorporates epidemiology, social and computer sciences, mathematics and physics. Epidemic models consist of two intertwined parts. The first part is the network, which is characterised by the number N of nodes and the L links between the nodes, specified by the adjacency matrix $A \in \mathbb{R}^{N \times N}$. The second part of epidemic models is the dynamic behaviour of the viral spread, which is mostly described by differential equations.

In the fundamental susceptible-infected-susceptible (SIS) model, each node is either in a susceptible or an infected state, which is specified by a Bernoulli random variable $x_i(t) \in \{0, 1\}$. For a node i in the SIS model, the susceptible state at time t is denoted by $x_i(t) = 0$ and the infected state at time t is denoted by $x_i(t) = 1$. Thus, for any node there are two transitions possible in the SIS model, from susceptible to infected and vice versa. The SIS model is formulated in continuous time $t \in \mathbb{R}^+$, starting at $t = 0$. The SIS model assumes that the curing process per node i is a Poisson process with *curing rate* δ and that the infection rate per link is a Poisson process with *infection rate* β. An extension is the $\epsilon-$SIS model, where a susceptible node also suffers from self-infections, an event that is independent of the number of infectious neighbours of the respective node and characterised by the additive *self-infection* rate ϵ. The knowledge of the underlying topology and of the curing and infection rates is decisive for the prediction of the viral spread and for the design of control strategies which aim for steering the network towards a desired state.

• *B. Prasse and P. Van Mieghem are with the faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, The Netherlands.*
E-mail: {B.Prasse, P.F.A.VanMieghem}@tudelft.nl

This work considers the inverse problem of estimating both the adjacency matrix A and the *spreading parameters* β, δ and ϵ, given the knowledge of the viral states $x_i(t)$ of all nodes $i = 1, ..., N$ of the sampled-time Markov chain of an ϵ-SIS process, described in Section 5, over a sequence of n time slots.

Section 2 reviews related work. The nomenclature is introduced in Section 3. Section 4 states our assumptions. The sampled-time Markov chain of the ϵ-SIS process is described in Section 5. The estimation problem is formulated in a Maximum-A-Posteriori (MAP) sense in Section 6, which gives rise to a mixed-integer programming problem. A brute-force approach would require a computation time of $\mathcal{O}(2^{N^2/2})$. In order to find a close-to-optimal point of the mixed-integer programming problem in a feasible computation time for larger networks, an efficient heuristic based on solving multiple convex problems is given in Section 7. Numerical evaluations of the heuristic and the brute-force approach are given in Section 8. The numerical experiments show that a reconstruction of the network and, up to a small error, an estimation of the spreading parameters is possible for the SIS process. The crucial negative result of our work is that the estimation requires an infeasibly high amount of data, i.e. observations of the viral states $x_i(t)$ of all nodes, for large networks.

2 Related Work

Various methods for the network reconstruction were proposed for the susceptible-infected (SI) and the susceptible-infected-recovered (SIR) models, whereby infected nodes remain infected, and hence contagious, or can be recovered, respectively. Since in these models the nodes change their state only once from susceptible to infected (and possibly to recovered), it is usually not possible to infer the network

topology from observing the nodal states over time of only one outbreak. To overcome this shortage of available information for the network reconstruction for SI and SIR models, a commonly employed setting is the observation of multiple, independent outbreaks or cascades [2]. The state-of-the-art network reconstruction methods for SI and SIR models confine to a maximum-likelihood formulation in discrete time [3], [4], [5], [6], [7], [8]. Besides network reconstruction methods based on observing viral dynamics, current research also focusses on other dynamical processes on networks. For general dynamical systems, Peng *et al.* [9] propose a parameter estimation method by a chaotic ant swarm algorithm.

Substantially less attention has been drawn to the network reconstruction for the SIS epidemics. For a discrete-time SIS process in which all nodes change their viral state simultaneously, Shen *et al.* [10] reconstruct the network topology by observing the states of all nodes and employ a heuristic which allows for compressed sensing. Vajdi and Scoglio [11] formulate the network reconstruction for the continuous-time SIS model in a Bayesian sense by utilising the time intervals between infections as observations, derive the closed-form expression for the posteriori distribution of the infection rates and propose a Gibbs sampling approach for a large number of uncertain links. Since in their work the posterior probability densities of the infection rates instead of the existence of links is estimated, they assume the transmission rate for an existing link to be known in order to deduce the existence of a link from the respective infection rate. Under the assumption that the adjacency matrix A is exactly known, Paré *et al.* [12] estimate the spreading parameters β and δ of the N-Intertwined Mean-Field Approximation (NIMFA) [13], [14], [15] of the SIS process in discrete-time by observing the viral states $x_i(t)$ of all nodes i over time t.

In this work, we assume the more general ϵ-SIS model [16] in discrete-time *without* information on the network topology A nor the spreading parameters β, δ and ϵ. Our contribution consists of rigorously stating the estimation problem for both the network topology and the spreading parameters in a MAP, or Bayesian, sense, taking into account every transition of the sampled-time Markov chain of the SIS process. Furthermore, by solving the reconstruction problem, we conclude that the topology and spreading parameter estimation is hardly feasible in practice.

3 Nomenclature

We denote the number of nodes in the graph by N. An adjacency matrix is denoted by A and the elements of A by a_{ij}. The set of all unweighted adjacency matrices A is denoted by $\mathcal{A}$. The probability of an adjacency matrix A is denoted by $\Pr[A]$. The probability density functions of the continuous spreading parameters are given by $f_{\text{beta}}(\beta)$, $f_{\text{delta}}(\delta)$ and $f_{\text{epsilon}}(\epsilon)$. Furthermore, the viral state of all nodes at time k is the $N \times 1$ vector $x[k]$. The infected state of node i at time k is indicated by $x_i[k] = 1$ and the susceptible state by $x_i[k] = 0$. We denote all observations until discrete time k by the $N \times k$ matrix $X[k] = (x[k], x[k-1], ..., x[1])$. The parameters which are to be estimated are contained in the tuple $\theta = (A, \beta, \delta, \epsilon)$. The parameters are given by the random variable Θ. The conditional probability density function of the realisation θ of the parameters given the observations $X[k]$ is denoted by $f_{\Theta|X[k]}(\theta)$ and the unconditional density by $f_{\Theta}(\theta)$.

4 Assumptions

The sampling time T of the sampled-time Markov chain of the SIS process needs to be "small enough", which is stated more precisely by the following two assumptions.

1) The number of changes of nodal states per time depends on the value of the spreading parameters β, δ and ϵ, and the greater their value the more changes of the nodal states per time. Since the exact values of the spreading parameters are unknown, we cannot directly state an upper bound on the sampling time T. Instead, we assume that arbitrary but finite upper bounds on the spreading parameters are available, i.e.
$$\beta \leq \beta_{\max}, \quad \delta \leq \delta_{\max} \quad \text{and} \quad \epsilon \leq \epsilon_{\max}$$
In Section 5, we derive an upper bound on the sampling time T using the upper bounds on the spreading parameters above.
2) The sampling time T of the sampled-time Markov chain of the SIS process, introduced in Section 5, is small enough such that describing the SIS process by the first-order terms of the Taylor expansion of the transition probabilities of the sampled-time Markov chain of the SIS process is "sufficiently accurate". This assumption translates into two requirements for the sampling time T: Firstly, each transition is observed, i.e. there is at most one transition in the interval $[kT, (k+1)T]$ for $k \in \mathbb{N}$. Secondly, the MAP estimate of the adjacency matrix A for the continuous-time SIS process coincides with the MAP estimate of the adjacency matrix A for the sampled-time Markov chain of the SIS process.

For the estimation of the adjacency matrix A and the spreading parameters β, δ and ϵ, the following assumptions are made.

3) There is an arbitrary, but positive lower bound available for β, i.e. $0 < \beta_{\min} < \beta$.
4) A-priori, the adjacency matrix A and the rates β, δ and ϵ are stochastically independent distributed. Furthermore, they are stochastically independent of the first observation $x[1]$.
5) The prior distribution $\Pr[A]$ of $A \in \mathcal{A}$ is logarithmically concave when extending the range of values to the convex hull of $\mathcal{A}$, i.e. $\log(\Pr[A])$ is concave when the range of elements of A is extended from $a_{ij} \in \{0, 1\}$ to $a_{ij} \in [0, 1]$. Every nonnegative concave function is logarithmically concave, but not vice versa [17].
6) The unknown spreading parameters β, δ and ϵ are assumed to be uniformly distributed in the intervals described above, which complies with the maximum-entropy principle [18].

TABLE 1
Nomenclature

$\mathcal{A}$	Set of all adjacency matrices of simple graphs, $\mathcal{A} = \{A \in \{0,1\}^{N\times N} \vert a_{ij} = a_{ji}, a_{ii} = 0,\ \forall i,j\}$
β and β_T	Infection rate β and sampled-time infection intensity $\beta_T = T\beta$
δ and δ_T	Curing rate δ and sampled-time curing intensity $\delta_T = T\delta$
ϵ and ϵ_T	Self-infection ϵ rate and sampled-time self-infection intensity $\epsilon_T = T\epsilon$
N	Number of nodes
n	Number of time instants which were observed, $n \in \mathbb{N}$
$\mathbb{N}_N$	Set of natural numbers not greater than N, i.e. $\mathbb{N}_N = \{1, 2, ..., N\}$
S_θ	Set of all possible parameter tuples θ
T	Sampling time of the sampled-time Markov chain
θ	Parameters which are to be estimated: $\theta = (A, \beta, \delta, \epsilon)$
$\tilde{\theta}$	Parameters θ after change of variables: $\tilde{\beta} = \beta_T^{-1}, \tilde{\delta} = \delta_T/\beta_T, \tilde{\epsilon} = \epsilon_T/\beta_T$
$\tilde{\theta}_{\text{cvx},l}$	Solution to the optimisation problem based on convex relaxation and line segment l
$\theta_{\text{heur},l}$	The heuristic estimate for θ based on the l-th line segment
θ_{heur}	The heuristic $\theta_{\text{heur},l}$ which results in the minimum value of the objective function
u	All-one vector $u = (1, ..., 1)^T \in \mathbb{R}^N$
$x[k]$	At time k: $x_i[k] = 0$ denotes the susceptible and $x_i[k] = 1$ the infected state of node i
$X[k]$	All observations at time k and before, $X[k] = (x[k], x[k-1], ..., x[1])$
w	Number of line segments of the piecewise-linear approximation

Assumptions 4-6 are required for the MAP approach in Section 6. All these three assumptions could be omitted and a maximum-likelihood approach be employed instead, which follows straightforwardly from the presented MAP method by omitting the additive term of the prior distribution of the parameters θ. For the numerical experiments in Section 8, we choose to generate the network topology A and spreading parameters β, δ and ϵ such that assumptions 4-6 do hold. In Section 8, the MAP estimate is thus at least as accurate as the maximum-likelihood estimate, and the performance of the MAP estimate serves as a best-case scenario in view of the difficulty of the estimation problem.

The Assumptions 1-3 are satisfied for continuous-time SIS processes on real-world networks, which justifies the application of the maximum-likelihood approach as introduced in this work. On the other hand, Assumptions 4-6 may not be not satisfied for the majority of real-world networks. Hence, the introduced MAP procedure cannot be used in a straightforward manner instead of the maximum-likelihood approach for real-world networks.

5 Sampled-Time ϵ-SIS Process

The inverse problem of the reconstruction of the network topology and the estimation of the spreading parameters, given a number of measurements, is best described in discrete-time. For epidemic processes on computer systems, such as the spread of opinions on social media, the dynamics are inherently in discrete-time due to the digital design of hardware and software. For other epidemic processes, such as the spread of a disease, it is reasonable to assume that there is a limit to the temporal resolution of the empirical measurements. In the following, the sampled-time Markov chain for the SIS process is stated.

We denote the transition probability of the continuous-time Markov chain of the SIS process from state i at time τ to state j at time $t+\tau$ by $P_{ij}(t)$, which is independent of τ since the SIS process is stationary. The sampled-time Markov chain with sampling time T is a discrete-time Markov chain [19], where the transition probabilities P_{ij} from state i to state j are given by the first-order Taylor expansion of $P_{ij}(t)$,

$$P_{ij} = P'_{ij}(0)T,$$

and the transition probabilities from state i to state i are given by

$$P_{ii} = 1 - \sum_{l=1}^{N} P_{il}$$

The transition probabilities depend on the adjacency matrix A and the spreading parameters β, δ, ϵ, which comprise the compound parameter tuple $\theta = (A, \beta, \delta, \epsilon)$. Due to Assumption 2, there are three transitions possible in the sampled-time Markov chain of the ϵ-SIS process. These transitions are listed below and their probabilities are inferred from the continuous-time SIS equations. For all transitions, we state upper bounds on the sampling time T, such that the corresponding transition probabilities are in $[0, 1]$.

1) A single node i changes from the infected state at discrete time k to the susceptible state at discrete time $k+1$. The probability of this transition is

$$\Pr\Big[x_i[k+1] = 0 \vert x_i[k] = 1, x[k], \theta\Big] = \delta T, \quad (1)$$

which needs to be smaller than one. Since $\delta T = \delta_T \leq T\delta_{\max}$, T must obey

$$T \leq \frac{1}{\delta_{\max}} \quad (2)$$

2) A single node i changes from the susceptible state at time instant k to the infected state at time instant $k+1$ with the probability

$$\Pr\Big[x_i[k+1] = 1 \Big\vert x_i[k] = 0, x[k], \theta\Big] = (\beta N_i(A, k) + \epsilon)T,$$

where $N_i(A,k)$ is the number of infected nodes adjacent to node i in A at time k. Since $\beta_T = \beta T$, $\epsilon_T = \epsilon T$ and

$$N_i(A,k) = \sum_{j=1}^{N} a_{ij} x_j[k], \tag{3}$$

we obtain

$$\Pr\Big[x_i[k+1] = 1 \Big| x_i[k] = 0, x[k], \theta\Big] = \epsilon_T + \beta_T \sum_{j=1}^{N} x_j[k] a_{ij} \tag{4}$$

In order to ensure that $\Pr[x_i[k+1] = 1 | x_i[k] = 0, x[k], \theta]$ is not greater than one, we consider the upper bound

$$\Pr\Big[x_i[k+1] = 1 \Big| x_i[k] = 0, x[k], \theta\Big] \leq \epsilon_T + \beta_T N$$

Since $\epsilon_T \leq \epsilon_{\max} T$ and $\beta_T \leq \beta_{\max} T$, we obtain that T must obey

$$T \leq \frac{1}{\epsilon_{\max} + N\beta_{\max}} \tag{5}$$

3) The state of no node changes from time k to time $k+1$. Denote the susceptible and infected nodes, respectively, at time instant k by

$$M_0[k] = \{j \in \mathbb{N}_N | x_j[k] = 0\}$$

and

$$M_1[k] = \{j \in \mathbb{N}_N | x_j[k] = 1\}$$

By denoting the all-one vector as $u = (1, ..., 1)^T$, it holds

$$|M_0[k]| = u^T(u - x[k]) \tag{6}$$
$$|M_1[k]| = u^T x[k] \tag{7}$$

Then, the probability of no change from time k to $k+1$ can be written as

$$\Pr\Big[x[k+1] = x[k] \Big| x[k], \theta\Big] = 1 - \sum_{j \in M_1[k]} \delta_T - \sum_{i \in M_0[k]} (\beta_T N_i(A,k) + \epsilon_T)$$

Using (3) and (7) gives

$$\Pr\Big[x[k+1] = x[k] \Big| x[k], \theta\Big] = 1 - \delta_T u^T x[k] - \sum_{i \in M_0[k]} \left(\epsilon_T + \beta_T \sum_{j=1}^{N} a_{ij} x_j[k]\right)$$

With (6), we obtain

$$\Pr\Big[x[k+1] = x[k] \Big| x[k], \theta\Big] = 1 - \delta_T u^T x[k] - u^T(u - x[k])\epsilon_T - \sum_{i \in M_0[k]} \sum_{j=1}^{N} \beta_T a_{ij} x_j[k]$$

Finally, since $\sum_{i \in M_0[k]} a_{ij} = \sum_{i=1}^{N} (1 - x_i[k]) a_{ij}$ and $u^T u = N$,

$$\Pr\Big[x[k+1] = x[k] \Big| x[k], \theta\Big] = 1 - N\epsilon_T + (\epsilon_T - \delta_T) u^T x[k] - \sum_{j=1}^{N} \beta_T x_j[k] \sum_{i=1}^{N} (1 - x_i[k]) a_{ij} \tag{8}$$

In order to ensure that the expression for $\Pr[x[k+1] = x[k] | x[k], \theta]$ does not exceed one, we consider the upper bound of (8)

$$\Pr\Big[x[k+1] = x[k] \Big| x[k], \theta\Big] \leq 1 - N\epsilon_T + (\epsilon_T - \delta_T) u^T x[k], \tag{9}$$

which follows from the fact that the sum in equation (8) is not negative. If the upper bound (9) does not exceed one, then also the transition probability $\Pr\Big[x[k+1] = x[k] \Big| x[k], \theta\Big]$ is bounded by one. We consider two cases, depending on the values the self-infection rate ϵ_T and the curing rate δ_T. If $\epsilon_T \geq \delta_T$, then we apply $u^T x[k] \leq N$ and obtain the following upper bound on the transition probability

$$\Pr\Big[x[k+1] = x[k] \Big| x[k], \theta\Big] \leq 1 - \delta_T N \tag{10}$$

On the other hand, if $\epsilon_T < \delta_T$, the transition probability is upper bounded by

$$\Pr\Big[x[k+1] = x[k] \Big| x[k], \theta\Big] \leq 1 - N\epsilon_T, \tag{11}$$

due to $u^T x[k] \geq 0$. Since the curing rate δ_T as well as the self-infection rate ϵ_T are not negative, both upper bounds (10) and (11) are smaller than or equal to one. Thus, also the transition probability $\Pr\Big[x[k+1] = x[k] \Big| x[k], \theta\Big]$ does not exceed one.

To ensure that the expression for $\Pr[x[k+1] = x[k] | x[k], \theta]$ is not negative, we deduce from (8)

$$\Pr\Big[x[k+1] = x[k] \Big| x[k], \theta\Big] \geq 1 - N\epsilon_T + (\epsilon_T - \delta_T) u^T x[k] - \beta_T \sum_{j=1}^{N} x_j[k] \sum_{i=1}^{N} (1 - x_i[k]),$$

since $a_{ij} \geq 0$. Furthermore, it holds

$$\Pr\Big[x[k+1] = x[k] \Big| x[k], \theta\Big] \geq 1 - N\epsilon_T + (\epsilon_T - \delta_T) u^T x[k] - \beta_T \frac{N^2}{4}, \tag{12}$$

which follows after minimisation with respect to $\xi = \sum_{j=1}^{N} x_j[k]$, and the minimum is attained at $\xi = \frac{N}{2}$. Assumption 1 states the bounds on the infection and curing rates, namely $\beta_T = \beta T \leq \beta_{\max} T$, $\delta_T \leq \delta_{\max} T$ and $\epsilon_T \leq \epsilon_{\max} T$. Using these bounds on the spreading parameters and the bound (12), we obtain that the transition probability $\Pr\Big[x[k+1] =$

$x[k]\Big|x[k],\theta\Big]$ is not negative if the sampling time T satisfies

$$T \leq \frac{4}{N^2\beta_{\max} + 4N\max\{\epsilon_{\max}, \delta_{\max}\}} \tag{13}$$

The upper bound on T in (13) is smaller than both bounds in (2) and (5) and, hence, condition (13) is a sufficient condition for all transition probabilities of the sampled-time Markov chain to lie in the interval $[0, 1]$.

6 Maximum-a-Posteriori Formulation

The estimation problem is stated in the MAP sense. Given all measurements $X[n] = (x[n], ..., x[1])$, we aim to find the parameter tuple θ_{MAP}, which maximises the posterior:

$$\theta_{\text{MAP}} = \arg\max_{\theta \in S_\theta} f_{\Theta|X[n]}(\theta), \tag{14}$$

where S_θ denotes the set of feasible solutions for the parameter tuple θ according to the assumptions in Section 4, i.e.

$$S_\theta = \mathcal{A} \times [\beta_{\min}, \beta_{\max}] \times [0, \delta_{\max}] \times [0, \epsilon_{\max}],$$

where $S_1 \times S_2$ denotes the Cartesian product of set S_1 and set S_2.

Applying the MAP estimation method is motivated especially by two facts. Firstly, the estimation problem is translated into an optimisation problem, which often allows for an efficient computation due to the advances in modern optimisation theory and the availability of high performance computers [20], [21].

Secondly, the MAP method exhibits two important accuracy properties for continuous parameter estimation, namely the MAP estimation gives an *unbiased* and *efficient* estimator. In order to introduce these accuracy properties, we denote an arbitrary estimator of the true parameters θ given n observations by $\hat{\theta}(n)$. Since the observations are generated by a random process, e.g. the ϵ-SIS process, and the estimator $\hat{\theta}(n)$ is a non-trivial mapping of these observations, the estimator $\hat{\theta}(n)$ is a random variable.

The first accuracy property of the MAP estimator is that, under mild conditions, the MAP estimator is (unconditionally) unbiased, which means that its expectation $\mathrm{E}[\hat{\theta}(n)]$ equals the true parameters, if $\hat{\theta}(n)$ equals the MAP estimator [22, Theorem 4.16].

In order to state the second property of the MAP estimator, we refer to the bound on the highest attainable accuracy of estimators, which was discovered independently by Rao [23] and Cramér [24], in 1945 and 1946, respectively. Under mild conditions, the Cramér-Rao inequality [22] gives a lower bound on the mean square error for any estimator $\hat{\theta}(n)$,

$$\mathrm{E}[(\hat{\theta}(n) - \theta)(\hat{\theta}(n) - \theta)^T] \succeq M(n, X[n], f_\Theta), \tag{15}$$

where for matrices A, B, we denote by $A \succeq B$ that $A - B$ is a positive semidefinite matrix, $M(n, X[n], f_\Theta)$ is a matrix depending on the number of observations n, the observations $X[n]$ and the prior distribution f_Θ of the parameters θ. An estimator $\hat{\theta}(n)$ is efficient if equality in (15) holds. The importance of the MAP estimator stems from the fact that if an unbiased efficient estimator exists, then it coincides with the MAP estimator[1]. We emphasise that the bound (15) solely depends on the estimation problem and not on the specific estimator $\hat{\theta}(n)$. Hence, the bound, together with the equality-achieving MAP estimator, gives a measure of difficulty of the respective estimation problem.

The Cramér-Rao bound requires the parameters θ to be continuous. For the ϵ-SIS process considered in this work, the parameters are not continuous due to the binary-valued adjacency matrix A. Only recently, attention has been drawn to establishing accuracy bounds on estimators for discrete parameter settings [25]. Motivated by translating the estimation into an optimisation problem and the strength for completely continuous parameter estimation, the vast majority of approaches to network reconstruction rely on maximum-likelihood or MAP estimation methods [3], [4], [5], [6], [7], [8].

The optimisation problem (14) is translated into a mixed-integer program. Bayes' theorem gives

$$f_{\Theta|X[n]}(\theta) = \frac{\Pr[X[n]|\theta]\, f_\Theta(\theta)}{\Pr[X[n]]},$$

and, since the SIS process is Markovian [19, Chapter 9],

$$f_{\Theta|X[n]}(\theta) = \frac{\Pr[x[1]|\theta]}{\Pr[X[n]]} f_\Theta(\theta) \prod_{k=2}^{n} \Pr[x[k]|x[k-1], \theta] \tag{16}$$

The term $\Pr[X[n]]$ is not a function of θ and neither is $\Pr[x[1]|\theta] = \Pr[x[1]]$, since θ and $x[1]$ are stochastically independent by Assumption 4 in Section 4. Hence, the first factor of (16) can be neglected and the estimation problem (14) becomes

$$\theta_{\text{MAP}} = \arg\max_{\theta \in S_\theta} \Pr[A] f_{\text{beta}}(\beta) f_{\text{delta}}(\delta) f_{\text{epsilon}}(\epsilon) \cdot \cdot \prod_{k=2}^{n} \Pr[x[k]|x[k-1], \theta],$$

since A, β, δ and ϵ are stochastically independent by Assumption 4. The spreading parameters β, δ and ϵ are uniformly distributed by Assumption 6, thus the last three factors can be neglected, and by taking the logarithm (preserving the same optimum), we obtain

$$\theta_{\text{MAP}} = \arg\max_{\theta \in S_\theta} \log(\Pr[A]) + \sum_{k=2}^{n} \log(\Pr[x[k]|x[k-1], \theta]) \tag{17}$$

We denote the set of the time instants of the infections of node i as

$$H_{01}[i] = \{k \in \mathbb{N}_n | x_i[k+1] = 1 \wedge x_i[k] = 0\}$$

1. For a finite number of observations n, the Cramér-Rao inequality can be stated in two versions, either neglecting the prior distribution f_Θ of the parameters θ [22, Theorem 4.13] or taking the prior distribution into account [22, Theorem 4.17]. The former bound is attained for maximum-likelihood estimators and the latter, and tighter, bound is attained for MAP estimators. The bounds coincide when the observation length n tends to infinity. In other words, considering the prior distribution of the parameters θ in the estimation procedure only has an impact on the accuracy for small observation lengths n.

Furthermore, we denote the set of time instants which correspond to the curing of a node and to a constant transition, respectively, as

$$H_{10} = \{k \in \mathbb{N}_n | \exists i \in \mathbb{N}_N : x_i[k+1] = 0 \wedge x_i[k] = 1\}$$
$$H_{\text{const}} = \{k \in \mathbb{N}_n | \forall i \in \mathbb{N}_N : x_i[k+1] = x_i[k]\}$$

The addends of the optimisation problem (17) correspond to the time instant of either one of the sets $H_{01}[i], H_{10}$ or H_{const}. We obtain

$$\theta_{\text{MAP}} = \arg\min_{\theta \in \Theta} f_{\text{obj}}(\theta),$$

where the objective function arises from the expressions for the transition probabilities (1), (4) and (8), and is explicitly given by

$$\begin{aligned} f_{\text{obj}}(\theta) = &- \log(\Pr[A]) - \sum_{k \in H_{10}} \log(\delta_T) \\ &- \sum_{i=1}^{N} \sum_{k \in H_{01}[i]} \log\left(\epsilon_T + \beta_T \sum_{j=1}^{N} x_j[k] a_{ij}\right) \\ &- \sum_{k \in H_{\text{const}}} \log\Big(1 - N\epsilon_T + (\epsilon_T - \delta_T) u^T x[k] \\ &\qquad + \beta_T \sum_{i,j} x_j[k](x_i[k] - 1) a_{ij}\Big) \end{aligned} \quad (18)$$

Since transitions can occur multiple times, i.e. $x[k_1] = x[k_2]$ and $x[k_1 - 1] = x[k_2 - 1]$ for $k_1 \neq k_2$, the objective function $f_{\text{obj}}(\theta)$ may contain the same addends multiple times and they can be replaced by a single addend weighted with the multiplicity of the addend. By formulating $\theta \in S_\theta$ as constraints, the MAP estimation is given by the following *mixed-integer programming* optimisation problem

$$\begin{aligned} \underset{\theta}{\text{minimise}} \quad & f_{\text{obj}}(\theta) \\ \text{subject to} \quad & a_{ij} \in \{0, 1\} \quad \forall i, j \\ & T\beta_{\min} \leq \beta_T \leq T\beta_{\max} \\ & 0 \leq \delta_T \leq \delta_{\max} T \\ & 0 \leq \epsilon_T \leq \epsilon_{\max} T \end{aligned} \quad (19)$$

The solution to the above optimisation problem (19) is denoted by θ_{MAP}. The optimisation problem (19) can be proved to be NP-hard [26], for *any* connected true adjacency matrix A, on which the SIS viral state sequence $x[1], ..., x[n]$ was generated. If the considered graph is simple, i.e. undirected and without self-loops, the constraints $a_{ij} = a_{ji}$ and $a_{ii} = 0$ can be added to the optimisation problem. For ease of exposition, these constraints are not explicitly stated in the following.

The optimum θ_{MAP} of problem (19) can be found by a brute-force algorithm, whose computation time is in $\mathcal{O}(2^{N(N-1)/2})$. Thereby, the minimum of the optimisation problem (19) is computed for every possible $N \times N$ adjacency matrix $A_1, A_2, ... \in \mathcal{A}$. For a fixed $A = A_m$, where $m \in \{1, ..., 2^{N(N-1)/2}\}$, the optimisation is performed with respect to the three spreading parameters $\beta_T, \delta_T, \epsilon_T$, and the objective function (18) is convex, since the objective function is a sum of composition of negative logarithms and linear functions. The brute-force approach yields $2^{N(N-1)/2}$ feasible points θ_m, one for each A_m, and the solution θ_{MAP} to the optimisation problem (19) is given by the feasible point which results in the minimal objective value. Since a computational complexity of $\mathcal{O}(2^{N(N-1)/2})$ is infeasible for large N, a heuristic based on convex relaxation and piecewise-linear approximation is presented in the Section 7.

7 Heuristic by Piecewise-Linear Approximation and Convex Relaxation

A common heuristic for solving mixed integer programming problems is based on the solution of a convex optimisation problem which results from relaxing the integer constraint [17], i.e. replacing $a_{ij} \in \{0, 1\}$ by $a_{ij} \in [0, 1]$. However, the objective function f_{obj} given by (18) contains the terms βa_{ij} which render f_{obj} non-convex also when applying the convex relaxation of the integer constraint.

The intuitive approach of introducing a new variable $\tilde{a}_{ij} = \beta_T a_{ij}$ and relaxing the binary constraint $\tilde{a}_{ij} \in \{0, \beta_T\}$ to $\tilde{a}_{ij} \in [0, \beta_T]$ cannot be straightforwardly employed. Firstly, there is no guarantee that expressing $\log(p[A])$ by $\tilde{a}_{ij}$ (and possibly by β_T) results in a convex function. Secondly, even if A is assumed to be uniformly distributed and the term $\log(p[A])$ can hence be omitted, replacing $\beta_T a_{ij}$ by $\tilde{a}_{ij}$ would erase β_T in the objective f_{obj}, and β_T would only appear in the constraints $\tilde{a}_{ij} \in [0, \beta_T]$ and $T\beta_{\min} \leq \beta_T \leq T\beta_{\max}$. Thus, setting $\beta_T = T\beta_{\max}$ would always be a solution to the corresponding convex optimisation problem since the constraint $[0, T\beta_{\max}]$ is the least restrictive interval for $\tilde{a}_{ij}$. It is not obvious how to infer an estimate β_T which does not equal $T\beta_{\max}$, and furthermore, how to deduce $a_{ij} = 0$ or $a_{ij} = 1$ from a solution $0 < \tilde{a}_{ij} < T\beta_{\max}$.

We propose a heuristic by translating the non-convex optimisation problem (19) into w convex optimisation problems with the solutions $\tilde{\theta}_{\text{cvx},l}$ for $l = 1, ..., w$. The translation is achieved by a transformation of the optimisation variables, i.e. $\tilde{\theta} = t(\theta)$ for a bijective function t, a piecewise-linear approximation of non-convex terms of the objective function with w line segments and by a convex relaxation of the binary constraint $a_{ij} \in \{0, 1\}$. The greater the number of line segments w, the more accurate is the piecewise-linear approximation. We refer the reader to the Appendix for details.

The solutions $\tilde{\theta}_{\text{cvx},l} = (A_{\text{cvx},l}, \tilde{\beta}_{\text{cvx},l}, \tilde{\delta}_{\text{cvx},l}, \tilde{\epsilon}_{\text{cvx},l})$, $l = 1..., w$, correspond to a non-binary estimates of the links, i.e. $(A_{\text{cvx},l})_{i,j}$ may neither be 0 nor 1. Considering that the links have to be binary-valued, a heuristic is employed which aims for finding an estimate that approximates the exact solution θ_{MAP} of the original optimisation problem (19). For each $\tilde{\theta}_{\text{cvx},l}$, the following two steps are performed.

1) The solution $\tilde{\theta}_{\text{cvx},l}$ of the optimisation problem (22) corresponds to a non-binary valued adjacency matrix $A_{\text{cvx},l}$. To obtain a binary-valued solution, we round the elements of $A_{\text{cvx},l}$ to the nearest integer, which results in the heuristic estimate denoted by

$$(A_{\text{heur},l})_{ij} = \begin{cases} 1 & \text{if } (A_{\text{cvx},l})_{ij} \geq \frac{1}{2} \\ 0 & \text{if } (A_{\text{cvx},l})_{ij} < \frac{1}{2} \end{cases} \quad (20)$$

2) The binary-valued heuristic estimates for the adjacency matrix $A_{\text{heur},l}$, together with the estimates $\tilde{\beta}_{\text{cvx},l}, \tilde{\delta}_{\text{cvx},l}, \tilde{\epsilon}_{\text{cvx},l}$ for the spreading parameters obtained by the convex problems (22), do realise a feasible point to the original optimisation (19). However, a better estimation of the three rates, for $A_{\text{heur},l}$ given, can be performed as follows.
When A is fixed to $A_{\text{heur},l}$, the objective of the original problem (19) becomes a function of the three rates, i.e.

$$f_{\text{obj},l}(\beta_T, \delta_T, \epsilon_T) = f_{\text{obj}}(A_{\text{heur},l}, \beta_T, \delta_T, \epsilon_T), \quad (21)$$

The function $f_{\text{obj},l}(\beta_T, \delta_T, \epsilon_T)$ is convex with respect to $\beta_T, \delta_T, \epsilon_T$. Hence, the function $f_{\text{obj},l}(\beta_T, \delta_T, \epsilon_T)$ can be efficiently minimised with respect to these three variables, whereby the constraints of the original optimisation problem (19) have to be considered (the constraints are also convex with respect to $\beta_T, \delta_T, \epsilon_T$). The reduced-size optimisation problem is also called *convex restriction* [27] of the original optimisation problem at the point $A_{\text{heur},l}$.
The refinement of the spreading parameters by convex optimisation results in an heuristic estimate, which is denoted by $\theta_{\text{heur},l} = (A_{\text{heur},l}, \beta_{\text{heur},l}, \delta_{\text{heur},l}, \epsilon_{\text{heur},l})$. As the heuristic is performed for every line segment $l = 1, ..., w$, we obtain w solution candidates $\theta_{\text{heur},l}$.

The final estimate of the presented heuristic estimation approach is denoted by θ_{heur} and is given by the candidate $\theta_{\text{heur},l}$ which results in the minimal value of the objective function (18), i.e.

$$\theta_{\text{heur}} = \text{ arg min } \left\{ f_{\text{obj}}\left(\theta_{\text{heur},l}\right) \middle| l = 1, ..., w \right\} \quad (22)$$

The corresponding value of the objective function is denoted by $f_{\text{heur}} = f_{\text{obj}}(\theta_{\text{heur}})$.

In pseudocode, the approach for determining a heuristic estimate θ_{heur} is given by Algorithm 1. The presented algorithm allows for parallelisation in a straightforward manner since $\theta_{\text{heur},l}$ can be obtained independently for the different line segments l.

Algorithm 1 Heuristic for SIS Network Reconstruction and Spreading Parameter Estimation

Input: Observations $X[n]$
Output: Heuristic for MAP estimate θ_{heur}
$f_{\text{heur}} \leftarrow \infty$
for $l = 1, ..., w$ **do** ▷ Piecewise-linear approximation with w segments
  Obtain $\tilde{\theta}_{\text{cvx},l} = (A_{\text{cvx},l}, \tilde{\beta}_{\text{cvx},l}, \tilde{\delta}_{\text{cvx},l}, \tilde{\epsilon}_{\text{cvx},l})$ by solving the convex optimisation problem (22)
  Obtain binary $A_{\text{heur},l}$ from non-binary $A_{\text{cvx},l}$ by rounding (20)
  Obtain $(\beta_{\text{heur},l}, \delta_{\text{heur},l}, \epsilon_{\text{heur},l})$ by minimising the convex $f_{\text{obj},l}(\beta_T, \delta_T, \epsilon_T)$ given by (21)
  if $f_{\text{obj}}\left(A_{\text{heur},l}, \beta_{\text{heur},l}, \delta_{\text{heur},l}, \epsilon_{\text{heur},l}\right) < f_{\text{heur}}$ **then**
    $\theta_{\text{heur}} \leftarrow (A_{\text{heur},l}, \beta_{\text{heur},l}, \delta_{\text{heur},l}, \epsilon_{\text{heur},l})$
    $f_{\text{heur}} \leftarrow f_{\text{obj}}(\theta_{\text{heur}})$
  end if
end for

8 Numerical Evaluation

Both the heuristic estimation approach, resulting from convex relaxation and piecewise-linear approximation, and the brute-force approach are numerically evaluated. Multiple Erdős-Rényi graphs are generated randomly. For each of these graphs, the nodal infection state matrix $X[n]$ is created by a random number generator according to the transition probabilities of the sampled-time ϵ-SIS process described in Section 5. The nodal states $X[n]$ are then given as input to the estimation procedures. A solution to the optimisation problems (22) of the heuristic is obtained by the Matlab command `fmincon`. The expression for the gradient of the respective objective function is provided to the solver `fmincon`. The resulting estimated graph and spreading parameters are compared with the true parameters.

We choose the link probability p of the Erdős-Rényi model such that the generated random graphs are connected with a high probability, which holds if p is significantly greater than the threshold $\log(N)/N$. By setting $p = 0.7$, we ensure $p > 2\log(N)/N$ for all networks considered in the numerical evaluation, which are of size $N = 4$ or greater. For Erdős-Rényi random graphs, the logarithm of the prior distribution of the adjacency matrix A is given by

$$\log\left(\Pr[A]\right) = \frac{1}{2}N(N-1)\log\left(1-p\right) + \log\left(\frac{p}{1-p}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}a_{ij} \quad (23)$$

Only the second addend depends on A and has to be considered for the optimisation. Alternatively, prior information on the link density could be considered by replacing $\log\left(\Pr[A]\right)$ by $\rho\sum_{ij}a_{ij}$, where ρ is the sparsity parameter [6].

The brute-force and the heuristic estimation of Section 7 are compared for small[2] network sizes $N = 4, 5, 6$ and for gradually increasing observation lengths from $n = 100$ to $n = 5000$. Furthermore, the heuristic estimation approach is numerically evaluated for larger networks up to $N = 24$ nodes and for gradually increasing observation lengths ranging from $n = 10^3$ to $n = 10^6$. For each pair of number of nodes N and observation length n, 10^3 networks are randomly generated according to the Erdős-Rényi random graph model with $p = 0.7$ - except for the comparison of brute-force and heuristic in Subsection 8.1, where $2 \cdot 10^3$ networks are created.

The spreading parameters are set to $\beta = 2/3$, $\delta = 1$ and $\epsilon = 0.01$. The upper bounds on the parameters are set to $\beta_{\text{max}} = 1$, $\delta_{\text{max}} = 1$ and $\epsilon_{\text{max}} = 1$. The lower bound on the infection rate is set to $\beta_{\text{min}} = 0.1\beta = 2/30$. The sampling time T is set as large as possible, considering the upper bound (13). Every node is set initially to the infected state, i.e. $x[1] = u$. For the heuristic approach presented in Section 7, the number of line segments for the piecewise-linear approximation is set to $w = 10$.

The accuracy of the resulting estimates is compared as follows. For the spreading parameters β, δ and ϵ, the error

2. The ratio of possible adjacency matrices grows exponentially with respect to N. For $N = 6$, there are approximately $3 \cdot 10^4$ possible adjacency matrices and for $N = 7$ there are approximately $2 \cdot 10^6$ possible matrices.

of the estimates is defined as the relative deviation, i.e. $(\beta - \hat{\beta})/\beta$, where $\hat{\beta}$ is the estimate and β the true value (the error of δ and ϵ is defined analogously). For the adjacency matrix, the error is defined as $(\hat{A} - A)/L$, where $\hat{A}$ and A are the estimated and true adjacency matrix, respectively, and $L = (N-1)N/2$ is the number of possible links.

8.1 Evaluation of the Heuristic Estimation Method

The heuristic based on convex relaxation and piecewise-linear approximation, as introduced in Section 7, and the exact brute-force algorithm are compared with respect to accuracy and computation time. Figure 1 depicts that for almost every randomly generated graph the heuristic and the brute-force estimates are identical. Figure 2 demonstrates the difference in computation time of the heuristic and exact brute-force approach.

We aim to justify the convex relaxation of the binary constraints of the MAP estimation problem (19), i.e. the replacement of the constraints $a_{ij} \in \{0,1\}$ by $a_{ij} \in [0,1]$. We set the spreading parameters to their true value and consider the estimation only with respect to the adjacency matrix A:

$$\begin{aligned} A_{\text{cvx}} = \arg\min_{A} \quad & f_{\text{obj}}(A, \beta_T, \delta_T, \epsilon_T) \\ \text{s. t.} \quad & a_{ij} \in [0,1] \quad \forall i,j \end{aligned} \tag{24}$$

Figure 3 illustrates the error of the non-binary estimate A_{cvx} and the rounded, binary estimate A_{heur}. The accuracy of the non-binary estimate A_{cvx} increases monotonically, which justifies the convex relaxation of the binary constraints of the MAP estimation problem (19).

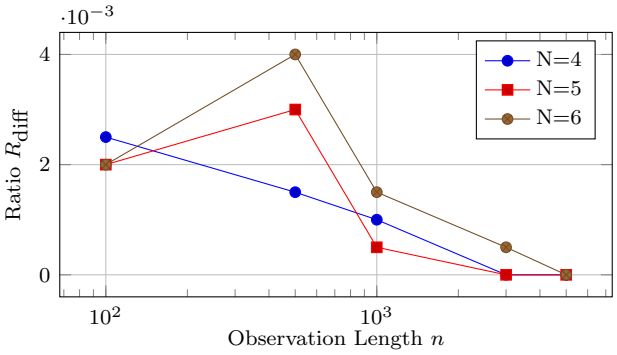


Fig. 1. Accuracy of heuristic: Fraction R_{diff} of the $2 \cdot 10^3$ randomly generated graphs for which the results of the heuristic and the exact brute-force methods do not coincide.

8.2 Accuracy of Estimation Depending on Observation Length

The dependency of the accuracy of the heuristic estimate θ_{heur}, given by equation (22), on the number of observation samples n is depicted in Figure 4. The figure shows that for a sufficiently large observation length, the adjacency matrix can be reconstructed with the heuristic approach almost always exactly. The network size N has not a great impact on the accuracy of the estimate for δ, but the network size N does have a considerable impact on the goodness of the estimates of both β and ϵ. The accuracy of the estimate of the

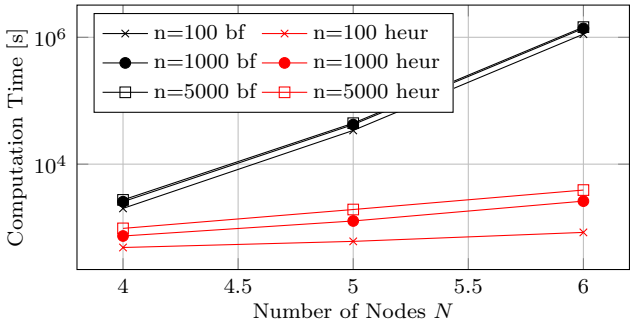


Fig. 2. Comparison of computation time of heuristic (heur) and brute-force (bf) approach in dependency of the number of nodes N.

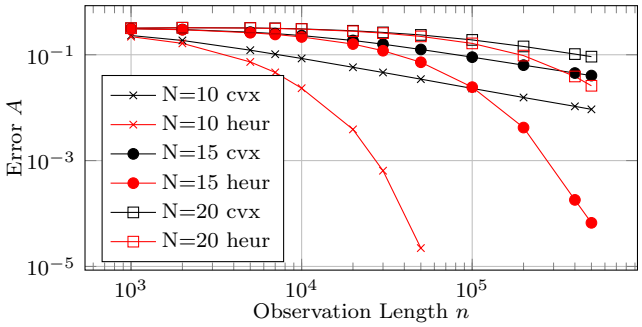


Fig. 3. Accuracy of the non-binary estimate A_{cvx}, given by (24), and the rounded estimate A_{heur} in dependency of the observation length n. Discontinued graphs in the logarithmically scaled plot refer to zero errors of the estimate of A.

self-infection rate ϵ is not monotonically increasing when the number of observations n is small and the adjacency matrix A is reconstructed very poorly.

To evaluate the introduced network reconstruction method for real-world networks, we consider the Zachary karate club [28] with $N = 34$ nodes and the network of windsurfers [29] with $N = 43$ nodes. We accessed the networks via the *Konect* network collection [30]. In both networks, the nodes refer to an individual, and the edges refer to a tie or interpersonal contact of the individuals, respectively. For the two networks, 100 different SIS viral state sequences $x[1], ..., x[n]$ were created. Since the prior distribution is not available, we perform a maximum-likelihood estimation by omitting the term $\log(\Pr[A])$ in the objective function (17). Figure 5 depicts the resulting accuracy of the estimates of the adjacency matrices A, averaged over the 100 different SIS viral state traces, in dependency of the observation length n. For both networks, the number of observations n is very large if a reasonable estimation accuracy is to be achieved.

8.3 Impact of Self-Infections

The ϵ-SIS model is more general than the SIS-model without self-infections ($\epsilon = 0$), and the network reconstruction method in this work is applicable in both cases. Without the presence of self-infections, the virus can die out, which means that $x_i[k] = 0$ for all nodes i and some time k. If the virus dies out at time k, then we reset the viral state to the all-one vector: $x[k+1] = u$. If the self-infection equals

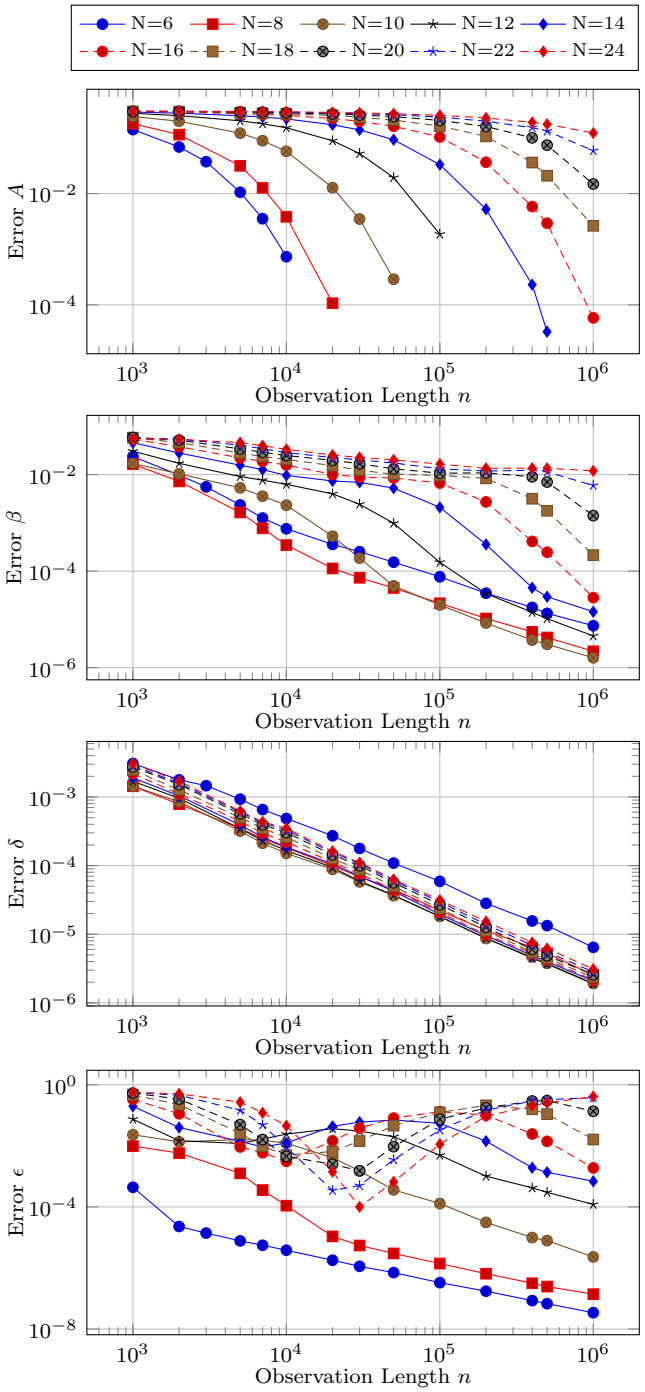


Fig. 4. Accuracy of reconstructed network and estimated parameters in dependency of the observation length n. Discontinued graphs in the first, logarithmically scaled plot refer to zero errors of the estimate of A.

$\epsilon = 0$, then only the infection rate β, the curing rate δ and the adjacency matrix A are estimated in the MAP problem (19). Figure (6) illustrates the impact of self-infections on the accuracy of the network reconstruction and estimation of the infection rate β. The accuracy of the network A changes to a negligible extend when self-infections do not occur. For larger observation lengths n, the accuracy of the estimation of the infection rate β deteriorates in the presence of self-infections.

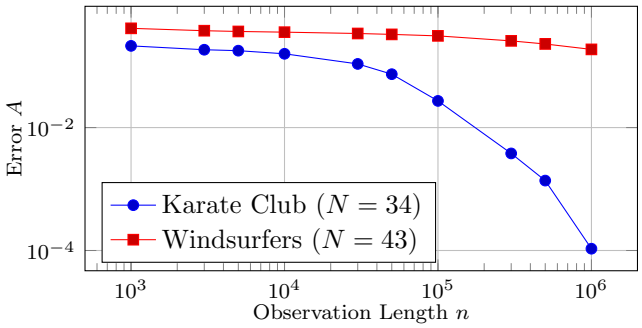


Fig. 5. Accuracy of reconstructed network and estimated parameters in dependency of the observation length n, for two real-world networks.

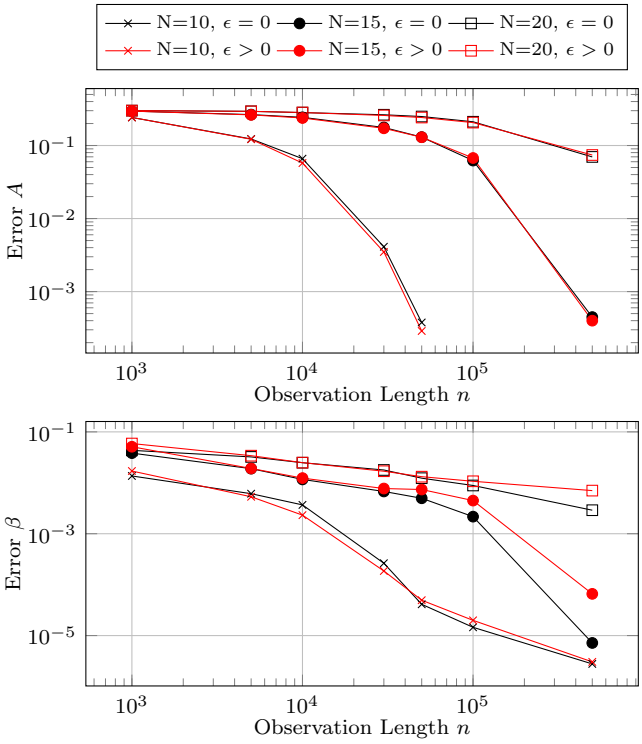


Fig. 6. Impact of self-infections: Comparison of accuracy of reconstructed network and estimated spreading parameters in dependency of the observation length n. Discontinued graphs in the logarithmically scaled plot refer to zero errors of the estimate of A.

8.4 Impact of Knowledge of Spreading Parameters

In some cases, the spreading parameters β, δ and ϵ may be available, and the estimation problem reduces to finding the adjacency matrix A. Figure (7) depicts the error of the reconstructed network for the two cases of known and unknown spreading parameters. If the spreading parameters are known, then the network reconstruction method performs better -especially for large observation lengths n- than if the spreading parameters are unknown.

8.5 Impact of the Value of the Infection-Rate

The greater the infection rate β, the more nodes are infected in the metastable state of the SIS process, which may have an impact on the accuracy of the network reconstruction.

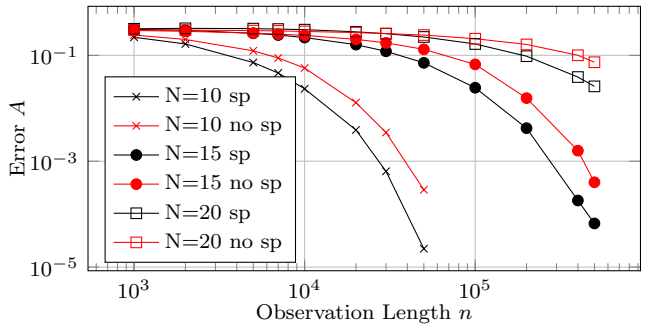


Fig. 7. Impact of knowledge of spreading parameters: Comparison of accuracy of reconstructed network with knowledge of spreading parameters (sp) and without knowledge of spreading parameters (no sp). Discontinued graphs in the logarithmically scaled plot refer to zero errors of the estimate of A.

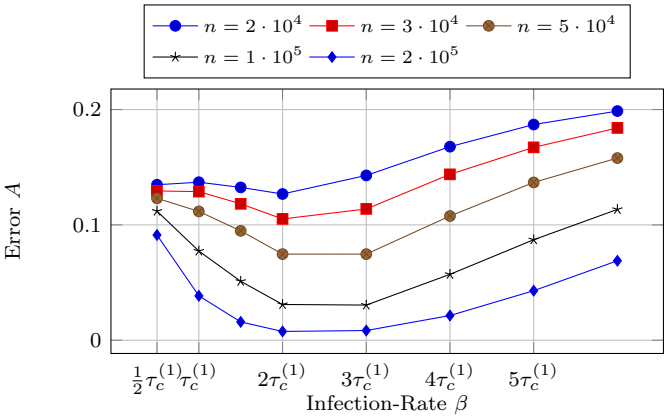


Fig. 8. Average error of estimates of A in dependency of the infection rate β.

We aim to choose the spreading parameters close to the epidemic threshold, which is decisive for the average number of infected nodes in the metastable state.

We generate adjacency matrices A with $N = 22$ nodes by the Erdős-Rényi model with the link probability $p = 0.1$, and keep 1000 connected adjacency matrices. Since we discard disconnected adjacency matrices, the prior distribution (23) is not correct, and we perform a maximum-likelihood estimation instead by omitting the term $\log(\Pr[A])$ in the objective function (17). For each generated adjacency matrix A, the infection rate β is set to a multiplicity $l \in \mathbb{R}$ of the lower bound on the epidemic threshold given by [19, Theorem 17.3.1],

$$\frac{\beta}{\delta} = l\tau_c^{(1)} = l\frac{1}{\lambda_1},$$

where λ_1 is the spectral radius of the adjacency matrix A on which the SIS process is run. The largest eigenvalue λ_1 of an $N \times N$ adjacency matrix is bounded by $\lambda_1 \leq N - 1$ and $\lambda_1 \geq 2L/N$, where L is the number of links of the adjacency matrix A. For Erdős-Rényi random graphs, the expected value of the lower bound on the eigenvalue λ_1 is $\mathrm{E}[2L/N] = (N-1)p$. For a given multiplicity l, we set the bounds of the infection rate to $\beta_{\min} = 0.1l(N-1)^{-1}$ and $\beta_{\max} = 10l(p(N-1))^{-1}$. We set the curing rate to $\delta = 1$ and the self-infection rate to $\epsilon = 0.01$. Figure 8 illustrates the network reconstruction error versus the infection rate β.

8.6 Impact of Knowledge on Prior Distribution

In Figure 9, the impact of knowledge of the prior distribution of the adjacency matrix A on the accuracy of the estimation is depicted. For the link probabilities $p = 0.7$ and $p = 0.9$, the heuristic estimation as presented in Section 7 is performed in two versions. The first version assumes that a-priori knowledge is available and sets the term of the a-priori distribution $\Pr[A]$ in the objective function (18) accordingly. The other version assumes that there is no a-priori knowledge available and the term $\Pr[A]$ in the objective function (18) is omitted, which is equivalent to assuming a uniform distribution of A. The plot shows that, if no a-priori knowledge is available, then the estimation accuracy for A is worse for small observation lengths n but the estimation accuracy converges to the accuracy of the case where a-priori knowledge is available if n increases. The observation is due to the fact that the ratio of $\log(\Pr[A])$ to $f_{\text{obj}}(\theta)$ in (18) converges to zero as the observation length n tends to infinity. For $p = 0.9$ the relative gap is larger than for $p = 0.7$, which is align with the intuition that a-priori knowledge of the distribution of the adjacency matrix A has more impact when A is randomly generated by a random graph model with higher entropy.

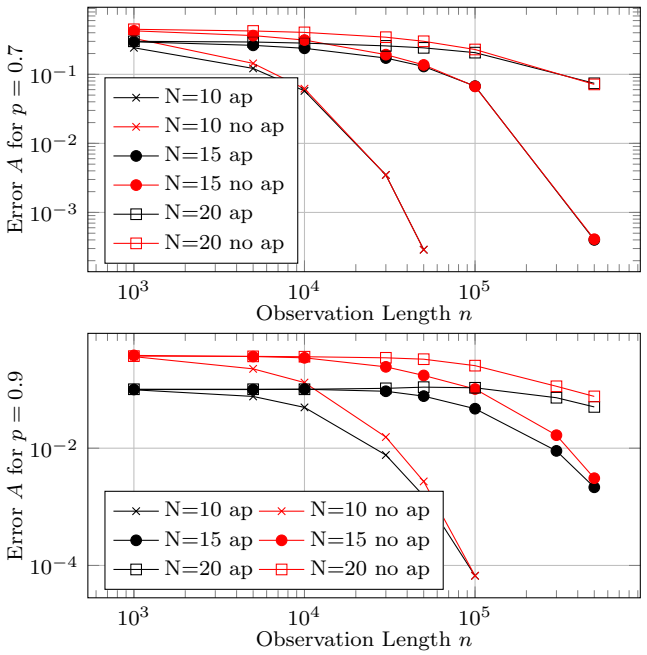


Fig. 9. Impact of a-priori knowledge: Comparison of accuracy of reconstructed network and estimated spreading parameters with a-priori knowledge (ap) and without a-priori knowledge (no ap) for the link probabilities $p = 0.7$ and $p = 0.9$. Discontinued graphs in the logarithmically scaled plots refer to zero errors of the estimate of A.

8.7 Required Observation Length and Computation Time

The number of observations and the computation time, which are required for a certain error tolerance of the reconstructed network, is a decisive indicator for the amount

TABLE 2
Parameters obtained by fitting subexponential and exponential functions to the number of observations n and the computation time T_{comp} in dependency of the number of nodes N, respectively.

Error A	0.2	0.15	0.1	0.05
α	0.5548	0.5596	0.5566	0.5549
b	-0.2259	0.0097	0.2985	0.5621
m	0.1380	0.1459	0.1551	0.1712
d	-1.5966	-1.5138	-1.4777	-1.5518

of information and time required for solving the estimation problem. In Figure 10 and Figure 11, the dependency of the observation length n and of the computation time, respectively, on the number of nodes N is depicted, if a certain error margin on the estimate of A is given. Since none of the data points in Figure 4 coincides exactly with either of the desired error on the estimate of A, a linear interpolation between these points is performed. The Figures 10 and 11 indicate that, given a desired average error on the adjacency matrix A, both the observation length and the computation time seem to grow exponentially with respect to the number of nodes N. Indeed, given the small range for N, we can approximately deduce a subexponential dependency [31] of the number of observations n on the number of nodes N, $\log_{10}(n) \approx N^{\alpha} + b$. Similarly, for the computation time T_{comp}, we approximately find an exponential dependency on the number of nodes N, $\log_{10}(T_{\text{comp}}) \approx mN + d$. The parameters α, b, m and d, which resulted from fitting the (sub)exponential functions to the interpolated data points, are given by Table 2.

Interior-point algorithms, such as the one implemented by the Matlab command `fmincon`, allow for a large number of decision variables. The number of decision variables in the optimisation problem (22) equals $(N(N-2)/2+3)$, which is not problematic for computing the solution of (22). Instead, the size N of the network, for which the network reconstruction can be performed, is in practice determined by the computation time, which is required for a reasonable estimation accuracy. For the Zachary karate club network [28] of size $N = 34$, Algorithm 1 took approximately 75 minutes on a *2.5GHz Intel Xeon Processor E5-2670 v2* in order to estimate the network with a reasonable accuracy (average fraction of erroneous links $\varepsilon_A \approx 10^{-4}$). As illustrated by Figure 11, the computation time for a desired average fraction of erroneous links ε_A grows exponentially with respect to the number of nodes N, which poses a severe practical constraint, even if the large number of observations n given by Figure 10 is available.

9 Conclusions

The problem of reconstructing the underlying graph and estimating the spreading parameters of the sampled-time SIS process is formulated in a Bayesian sense and an efficient, polynomial-time heuristic is proposed. Numerical evaluations indicate that the heuristic estimation procedure performs very well in comparison to the exact solution on small-scale networks. This observation motivates to use the

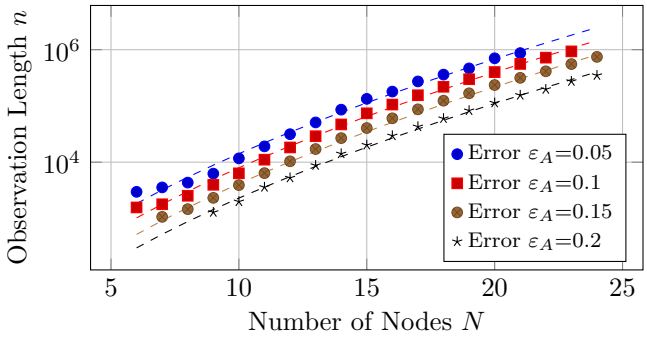


Fig. 10. Required observation length n for given average fraction ε_A of erroneous links of the estimate of A, in dependency of the number of nodes N. The points are obtained by interpolation and the fitted subexponential dependencies are given by the dashed graphs.

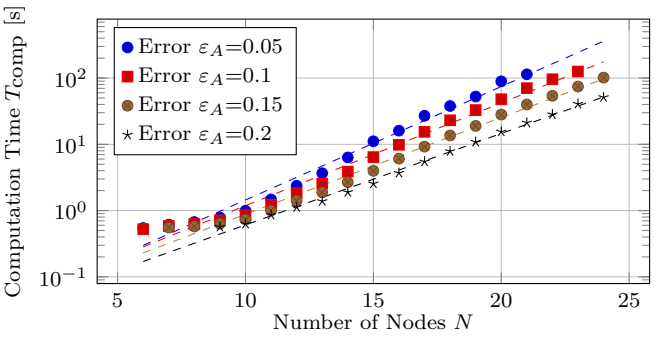


Fig. 11. Required computation time T_{comp} for given average fraction ε_A of erroneous links of the estimate of A, in dependency of the number of nodes N. The points are obtained by interpolation and the fitted subexponential dependencies are given by the dashed graphs.

heuristic for larger networks, where solving the estimation problem exactly becomes computationally infeasible. Indeed, we have proved in another work [26] that the maximum-likelihood estimation problem is NP-hard for *any* connected true adjacency matrix A, on which the SIS viral state sequence was generated.

Numerical evaluations demonstrate that also for larger networks the heuristic algorithm estimates the true spreading parameters up to a small error margin and that the true topology A is almost always inferred correctly for sufficiently many observations n. However, the number of observations n which is required for a sufficiently high accuracy of the network reconstruction grows rapidly with respect to the size N of the network, in particular $\log_{10}(n) \approx N^{\alpha} + b$, for $\alpha \approx 0.56$.

In practical applications, the underlying network is only available in exceptional cases and reconstructing the network is rendered infeasible due to the tremendous amount of required observations. The negative result will further deteriorate with incomplete information, when certain nodes or periods of time are not observable. Endeavours that aim for steering the viral spread towards a desired state, such as the mitigation of an outbreak of an infectious human disease or countermeasures against fake news on social media, are thus subject to the fundamental limit of uncertainty of the underlying network. This limit strongly opposes the *Big Data* belief, suggesting that the vast amount of available

data is sufficient to solve most problems. Since observing the viral states for a practicable number of observation does not allow for reconstructing the network with a viable accuracy, approaches for steering the viral spread necessarily have to incorporate the uncertainty of the underlying network.

REFERENCES

[1] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Reviews of Modern Physics*, vol. 87, no. 3, p. 925, 2015.

[2] M. Gomez Rodriguez, D. Balduzzi, and B. Schölkopf, "Uncovering the temporal dynamics of diffusion networks," *arXiv preprint arXiv:1105.0697*, 2011.

[3] P. Netrapalli and S. Sanghavi, "Learning the graph of epidemic cascades," in *ACM SIGMETRICS Performance Evaluation Review*, vol. 40, pp. 211–222, ACM, 2012.

[4] A. Braunstein and A. Ingrosso, "Network reconstruction from infection cascades," *arXiv preprint arXiv:1609.00432*, 2016.

[5] Q. Duong, M. P. Wellman, and S. Singh, "Modeling information diffusion in networks with unobserved links," in *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, pp. 362–369, IEEE, 2011.

[6] S. Myers and J. Leskovec, "On the convexity of latent social network inference," in *Advances in Neural Information Processing Systems*, pp. 1741–1749, 2010.

[7] E. Sefer and C. Kingsford, "Convex risk minimization to infer networks from probabilistic diffusion data at multiple scales," in *2015 IEEE 31st International Conference on Data Engineering (ICDE)*, pp. 663–674, IEEE, 2015.

[8] M. Gomez Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1019–1028, ACM, 2010.

[9] H. Peng, L. Li, Y. Yang, and F. Liu, "Parameter estimation of dynamical systems via a chaotic ant swarm," *Physical Review E*, vol. 81, no. 1, p. 016207, 2010.

[10] Z. Shen, W.-X. Wang, Y. Fan, Z. Di, and Y.-C. Lai, "Reconstructing propagation networks with natural diversity and identifying hidden sources," *Nature Communications*, vol. 5, 2014.

[11] A. Vajdi and C. Scoglio, "Missing Link Identification Using SIS Epidemic Traces," *arXiv preprint arXiv:1708.01631*, 2017.

[12] P. E. Pare, J. Liu, C. L. Beck, B. E. Kirwan, and T. Basar, "Analysis, Identification, and Validation of Discrete-Time Epidemic Processes," *arXiv preprint arXiv:1710.11149*, 2017.

[13] P. Van Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 17, no. 1, pp. 1–14, 2009.

[14] P. Van Mieghem, "The N-intertwined SIS epidemic network model," *Computing*, vol. 93, no. 2-4, pp. 147–169, 2011.

[15] K. Devriendt and P. Van Mieghem, "Unified mean-field framework for susceptible-infected-susceptible epidemics on networks, based on graph partitioning and the isoperimetric inequality," *Physical Review E*, vol. 96, no. 5, p. 052314, 2017.

[16] P. Van Mieghem and E. Cator, "Epidemics in networks with nodal self-infection and the epidemic threshold," *Physical Review E*, vol. 86, no. 1, p. 016116, 2012.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[18] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review*, vol. 106, no. 4, p. 620, 1957.

[19] P. Van Mieghem, *Performance Analysis of Complex Networks and Systems*. Cambridge University Press, 2014.

[20] P. J. Green, K. Łatuszyński, M. Pereyra, and C. P. Robert, "Bayesian computation: a summary of the current state, and samples backwards and forwards," *Statistics and Computing*, vol. 25, no. 4, pp. 835–862, 2015.

[21] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, 2016.

[22] L. Pronzato and A. Pßzman, "Asymptotic Properties of M, ML, and Maximum A Posteriori Estimators," in *Design of Experiments in Nonlinear Models*, pp. 79–104, Springer, 2013.

[23] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," in *Breakthroughs in Statistics*, pp. 235–247, Springer, 1992.

[24] H. Cramér, *Mathematical Methods of Statistics*, vol. 9. Princeton University Press, 2016.

[25] C. Choirat and R. Seri, "Estimation in discrete parameter models," *Statistical Science*, vol. 27, no. 2, pp. 278–293, 2012.

[26] B. Prasse and P. Van Mieghem, "Maximum-likelihood network reconstruction for SIS processes is NP-hard," *arXiv preprint arXiv:1807.08630*, 2018.

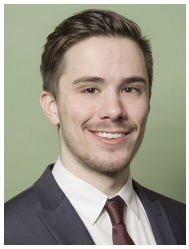
[27] S. Diamond, R. Takapoui, and S. Boyd, "A general system for heuristic minimization of convex functions over non-convex sets," *Optimization Methods and Software*, pp. 1–29, 2017.

[28] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of anthropological research*, vol. 33, no. 4, pp. 452–473, 1977.

[29] L. C. Freeman, S. C. Freeman, and A. G. Michaelson, "On human social intelligence," *Journal of Social and Biological Structures*, vol. 11, no. 4, pp. 415–425, 1988.

[30] J. Kunegis, "Konect: the koblenz network collection," in *Proceedings of the 22nd International Conference on World Wide Web*, pp. 1343–1350, ACM, 2013.

[31] R. Impagliazzo, R. Paturi, and F. Zane, "Which problems have strongly exponential complexity?," in *Proceedings 39th Annual Symposium on Foundations of Computer Science*, pp. 653–662, IEEE, 1998.



Bastian Prasse is pursuing his PhD degree since April 2017 at Delft University of Technology, The Netherlands. He obtained an M.Sc. degree in systems and control theory at the Royal Institute of Technology (KTH), Stockholm, Sweden, and an M.Sc. degree in computer engineering (*Dean's List*) at RWTH Aachen University, Aachen, Germany, both in 2015. In 2012, he obtained a B.Sc. degree in computer engineering (*Dean's List*) at RWTH Aachen University. His main research interests include network reconstruction and network control.

Piet Van Mieghem received the Masters (*magna cum laude*, 1987) and PhD (*summa cum laude*, 1991) degrees in electrical engineering from the K.U. Leuven, Leuven, Belgium. He is a Professor at the Delft University of Technology and Chairman of the section Network Architectures and Services (NAS) since 1998. His main research interests lie in modeling and analysis of complex networks and in new Internet-like architectures and algorithms for future communications networks. Before joining Delft, he worked at the Interuniversity Micro Electronic Center (IMEC) from 1987 to 1991. During 1993-1998, he was a member of the Alcatel Corporate Research Center in Antwerp, Belgium. He was a visiting scientist at MIT (1992-1993), a visiting professor at UCLA (2005), a visiting professor at Cornell University (2009), and at Stanford University (2015). He is the author of four books: Performance Analysis of Communications Networks and Systems (Cambridge Univ. Press, 2006), Data Communications Networking (Techne, 2011), Graph Spectra for Complex Networks (Cambridge Univ. Press, 2011), and Performance Analysis of Complex Networks and Systems (Cambridge Univ. Press, 2014). He was member of the editorial board of Computer Networks (2005-2006), the IEEE/ACM Transactions on Networking (2008-2012), the Journal of Discrete Mathematics (2012-2014) and Computer Communications (2012-2015). Currently, he serves on the editorial board of the OUP Journal of Complex Networks.