Published in final edited form as:
IEEE Commun Lett. 2014 August ; 18(8): 1347–1350. doi:10.1109/LCOMM.2014.2325026.

Optimal Network-based Intervention in the Presence of Undetectable Viruses

Mina Youssef [Member, IEEE] and
Network Dynamics and Simulation Science Laboratory, Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA 24061 USA (myoussef@vbi.vt.edu)

Caterina Scoglio [Member, IEEE]
The Institute of Computational Comparative Medicine, College of Veterinary Medicine, and with Department of Electrical and Computer Engineering, Kansas State University, Manhattan, KS 66506 USA (caterina@ksu.edu).

Abstract

This letter presents an optimal control framework to reduce the spread of viruses in networks. The network is modeled as an undirected graph of nodes and weighted links. We consider the spread of viruses in a network as a system, and the total number of infected nodes as the state of the system, while the control function is the weight reduction leading to slow/reduce spread of viruses. Our epidemic model overcomes three assumptions that were extensively used in the literature and produced inaccurate results. We apply the optimal control formulation to crucial network structures. Numerical results show the dynamical weight reduction and reveal the role of the network structure and the epidemic model in reducing the infection size in the presence of indiscernible infected nodes.

Keywords

Spread of viruses; optimal control; networked-based model; asymptomatic infection

I. Introduction

Biological epidemiology has produced significant number of deterministic and stochastic models, which have been successful in providing insights and deep understanding of the epidemic process phenomenon leading to successful conclusions about prevention and prediction of spread of viruses. Epidemic modeling has been applied also to describe packet routing using epidemic models [3], [4] and to evaluate the spread of viruses in computer networks [1], [2]. However, in the literature, when intervention strategies are applied in the network, there are three primitive assumptions producing inaccurate modeling results. The first assumption is the homogeneous mixing assuming that all the nodes have the same expected communication degree neglecting the connectivity variation among the nodes [11], [8]. The second assumption is the static network connectivity. As a matter of fact, during the spread of a virus, connectivity patterns among healthy nodes and infected nodes are altered, and these alterations in turn modify the epidemic evolution. The two intertwined dynamic processes, the one related to the epidemic spreading and the other related to the network

connectivity adaptation, act together and impact the infection size. The third assumption is the detection of infected nodes. Many approaches in the literature assumed that every infected node shows infection symptoms neglecting the fact that symptoms can be indiscernible in some infected nodes. In the literature, some models fall into the three assumptions such as [5], [6], [9], [10], [12]. Therefore, a major challenge regarding the problem of optimal contact reduction is the assumption of observability of the infection status of neighbors. Not only privacy constraints, but also the lack of detectable symptoms leads to increase the infection size due to the indiscernible infection force from the asymptomatic infection, i.e. without showing infection symptoms. Thus, the assumption that the infection is always detectable can lead to extra contact reduction between the infected node and the susceptible neighbor node and in turns the incurred contact reduction cost increases. Therefore, quantifying the effect of the assumption that all the infected nodes are always detected given the real scenario that some infected cases can not be detected becomes a mandatory task. In this letter, we present an intervention technique for the spread of epidemics in networks that avoid the three assumptions. The new intervention is based on the optimal control theory which is applied to *Susceptible-Infected-Recovered* SIR epidemiological model. In the SIR model, a susceptible node becomes infected due to the contact with infected neighbor(s). To address the hidden infection, an infected node becomes either symptomatic infected - infection symptoms are observable - or asymptomatic infected - infection symptoms are indiscernible- with asymptomatic probability. After contracting the infection, infected nodes permanently recover from the infection with a given cure rate. The main objective is to minimize the infection cost and the intervention cost. In our case, the intervention represents the optimal reduction in the contact weight applied between susceptible nodes and infected nodes showing symptoms. The contact weight represents the contact frequency and link availability between two nodes. Unlike the trivial methods in which infected nodes are isolated due to intervention such as quarantine intervention, the new approach preserves a global minimum communication level for every contact/link. In addition, the optimal control formulation addresses the trade-off between minimization of total infected nodes and minimization of contact weights reduction. Differently from the literature, our approach considers the network structure when the optimal control theory is applied to the epidemic model. The epidemic model, the mean-field approximated model for the general Markovian connectivity and the dynamic rate reduction for intervention purpose have been considered in the literature. For instance, the application of classical optimal control theory for the networked epidemic model was first introduced in [6]. However, the modeling contribution is to integrate the mean-field approximated model with an asymptomatic state within an optimal control framework aiming to reduce the final expected number of infection. This model reflects the spread of viruses in computer networks where infected nodes with anti-viruses that detect the threats are symptomatic, while infected nodes with anti-viruses, which can not detect the threats, are asymptomatic. Thus, asymptomatic infection has longer lifetime than symptomatic infection. This model has many applications in epidemic routing scheme. It can be used to estimate the failure rate/malfunction of routers in the networks by estimating the asymptomatic infection parameters through collected data about packets routing and the time delay for a packet to reach the destination end system. It also can be used for multipath routing to guarantee certain quality-of-service [7] by tuning some nodes to act as

asymptomatic nodes. We numerically evaluate the new approach under three different scenarios: 1) non-intervention scenario, 2) intervention scenario assuming the infection is always observable and 3) intervention scenario only for the symptomatic infected nodes. The numerical evaluations show that the assumption that all infected nodes are detectable leads to extra reduction in contact weights, which is less investment because nodes extensively reduce the communication for longer time period.

II. Network-based SIR approach

The Networked SIR model [8] is mainly composed of susceptible, infected and recovered states. We divided the infected state into two states, symptomatic infected and asymptomatic infected as shown in Figure 1. The spreading mechanism is a stochastic process, which is modeled through continuous-time Markov chain. Given a network with N nodes, there are 4^N states that describe the network states during the spread of viruses; However, the mean field theory can be employed to reduce the complexity from exponential $O(4^N)$ to polynomial $O(N)$. Therefore, instead of considering the combinatorial states of the nodes in the network, we study each node separately [11] by decomposing the infinitesimal $Q_4{}^N \times 4^N$ matrix to N infinitesimal matrices, each with four states S_m, I_m^{as}, I_m^s and R_m as follows:

$$q_m(t) = \begin{bmatrix} -x & p_a x & (1-p_a)x & 0 \\ 0 & -\delta^{as} & 0 & \delta^{as} \\ 0 & 0 & -\delta^{s} & \delta^{s} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

where $x = \beta \sum_n w_{m,n}^o \left(r_a 1_{[i_n^{as}(t)=1]} + 1_{[i_n^s(t)=1]} \right)$. All the variables and parameters are summarized in table I. In this approach, we replace the actual event of a node to be susceptible $s_m(t) = 1$, asymptomatically infected $i_m^{as}(t) = 1$, symptomatically infected $i_m^s(t) = 1$, recovered $r_m(t) = 1$ with their effective probabilities $S_m(t) = p(s_m(t) = 1)$, $I_m^{as}(t) = p(i_m^{as}(t) = 1)$, $I_m^s(t) = p(i_m^s(t) = 1)$, and $R_m(t) = p(r_m(t) = 1)$, respectively. Replacing every event with its effective probability is basically a mean field approximation. Hence, $x^{eff} = \beta \sum_n w_{m,n}^o (r_a I_n^{as}(t) + I_n^s(t))$ in the effective $q_m^{eff}(t)$ infinitesimal matrix. For every node m, we derive a system of differential equations using the effective $q_m^{eff}(t)$ infinitesimal matrix $\frac{dState_m^t(t)}{dt} = State_m^T(t)\, q_m^{eff}(t)$ where $State_m^T(t) = [S_m(t) \quad I_m^{as}(t) \quad I_m^s(t) \quad R_m(t)]$ is the vector of the state probabilities of node m. The system of differential equations for node m is as follows:

$$\frac{dS_m(t)}{dt} = -S_m(t)\, \beta \sum_{n \in N} w_{m,n}^o (r_a I_n^{as}(t) + I_n^s(t)), \quad (1)$$

$$\frac{dI_m^{as}(t)}{dt} = S_m(t)\, p_a \beta \sum_{n \in N} w_{m,n}^o (r_a I_n^{as}(t) + I_n^s(t)) - \delta^{as} I_m^{as}(t), \quad (2)$$

$$\frac{dI_m^s(t)}{dt} = S_m(t)(1-p_a)\beta\sum_{n\in N} w_{m,n}^o\left(r_a I_n^{as}(t) + I_n^s(t)\right) - \delta^s I_m^s(t). \quad (3)$$

We only solve 3*N* simultaneous differential equations.

III. Optimal control problem

We formulate a continuous time optimal control problem to minimize the total infection size by optimally reducing the contact weights among nodes within a finite time interval $t \in [0, T_f]$. Initially, we assume that there is no weight reduction in the network $w_{m,n}(t=0) = w_{m.n}^o$ where $w_{m,n}^o$ is the normal weight value between node *m* and node *n*. When a virus invades a network at $t > 0$, weights are reduced from their initial values $w_{m,n}(t) \le w_{m,n}^o$. In particular, we impose two bounds on each link weight (*m*, *n*): a) $\alpha w_{m,n}^o$ and b) $w_{m,n}(t) \le w_{m,n}^o$, where is α global minimum connectivity coefficient. These constraints have direct implications for the network as follows: First, to preserve a minimum contact level among nodes even during the spread of viruses, we introduce a positive lower bound for the weights, $\alpha w_{m,n}^o \le w_{m,n}(t)$ where $0 < a < 1$. Second, the level of contact between two nodes cannot increase beyond the normal level $\left(w_{m,n}^o\right)$ whenever a virus invades the network. Weight reduction process only takes place between a symptomatic infected node and a susceptible node since there is no infection symptom that enforces the weight reduction between asymptomatic infected nodes and susceptible nodes. Thus, the contact weight between an asymptomatic infected node and a susceptible node remains constant with normal value $w_{m,n}^o$. Mathematically, for every node *m*, the susceptible probability $S_m(t)$, the symptomatic infection probability $I_m^s(t)$, and the asymptomatic infection probability $I_m^{as}(t)$ are the state variables, while the weight reduction $\left(w_{m,n}^o - m_{m,n}(t)\right)$ is the control function. The objective function is given by the sum of the weight reduction cost function, and the total expected new infection size as shown in the following equation:

$$\text{Minimize} \int_0^{T_f} \sum_{m,n\in N}\left[f\left(w_{m,n}^o - w_{m,n}(t)\right)\right] + \sum_{m\in N}\beta S_m(t)\sum_n\left[I_n^s(t)\,w_{m,n}(t) + I_n^{as}(t)\,r_a w_{m,n}^o\right]dt \quad (4)$$

where the first term *f*(.) is a non-negative strictly convex function representing the weight reduction cost, while the second term represents total expected new infection cases at time *t*. Moreover, the differential equations describing the state evolutions and the weight inequality are as follows:

$$\frac{dS_m(t)}{dt} = -S_m(t)\,\beta\sum_{n\in N}\left(w_{m,n}^o r_a I_n^{as}(t) + w_{m,n}(t)\,I_n^s(t)\right), \quad (5)$$

$$\frac{dI_m^{as}(t)}{dt} = S_m(t)\,p_a\beta\sum_{n\in N}\left(w_{m,n}^o r_a I_n^{as}(t) + w_{m,n}(t)\,I_n^s(t)\right) - \delta^{as} I_m^{as}(t), \quad (6)$$

$$\frac{dI_m^s(t)}{dt} = S_m(t)(1-p_a)\beta \sum_{n\in N}\left(w_{m,n}^o r_a I_n^{as}(t) + w_{m,n}(t) I_n^s(t)\right) - \delta^s I_m^s(t), \quad (7)$$

In differential equations 5-7, we emphasize that there is no weight reduction process when an asymptomatic infected node is in contact with susceptible node and the contact weight remains constant $w_{m,n}^o$. Hamiltonian methods and Pontryagin's minimum principle are applied to different optimization problems for epidemiological models to determine the explicit optimal control function and the optimal state variables. Therefore, we apply Pontryagin's minimum principle and we derive the Hamiltonian function *H* as follows:

$$\alpha w_{m,n}^o \le w_{m,n}(t) \le w_{m,n}^o \quad \forall \quad m,n \in N$$

The co-state equations are derived by evaluating $\frac{d\lambda_{S_m}}{dt}$, $\frac{d\lambda_{I_m^{as}}}{dt}$ and $\frac{d\lambda_{I_m^a}}{dt}$. The equations of the transversality conditions are equal to 0. Next, the optimality condition is $H\left(w_{m,n}^*, S_m^*, I_m^{as*}, I_m^{s*}\right) \le H\left(w_{m,n}, S_m^*, I_m^{as*}, I_m^{s*}\right)$ where $w^*_{m,n}$ is the optimal weight value at time *t* such that $H\left(w_{m,n}^*, S_m^*, I_m^{as*}, I_m^{s*}\right)$ is minimized. Let $y_{m,n} = w_{m,n}^o - w_{m,n}$ represents the weight reduction such that $0 \le y_{m,n} \le w_{m,n}^o(1-\alpha)$. After substituting the Hamiltonian Eq. (8) in the optimality condition, we obtain the following inequality:

$$\begin{aligned} H&\left(S_m, I_m^{as}, I_m^s, \lambda S_m, \lambda_{I_m^{as}}, \lambda_{I_m^s}\right) \\ &= \sum_{m,n\in N} f\left(w_{m,n}^o - w_{m,n}\right) \\ &+ \sum_{m\in N} \beta S_m \sum_n \left[I_m^s w_{m,n} + I_n^{as} r_a w_{m,m}^o\right] \\ &+ \sum_{m\in N}\left(\left(p_a\lambda_{I_m^{as}} + (1-p_a)\lambda_{I_m^s} - \lambda_{S_m}\right)\left(\beta S_m \sum_{n\in N} w_{m,n}^o r_a I_n^{as} + w_{m,n} I_n^s\right) - \delta^{as}\lambda_{I_m^{as}} I_m^{as} - \delta^s \lambda_{I_m^s} I_m^s\right) \end{aligned} \quad (8)$$

where $\psi_m = \beta S_m\left(1 - \lambda_{S_m} + p_a\lambda_{I_m^{as}} + (1-p_a)\lambda_{I_m^s}\right)$. Since $y_{m,n} = 0$ is an admissible point and $f(y_{m,n} = 0) = 0$, therefore, the following inequality holds for all time *t*:

$$\sum_{m,n\in N}\left(f\left(y_{m,n}^*\right)\right) - \sum_{m\in N}\psi_m^* \sum_{n\in N} y_{m,n}^* I_m^{s*} \le \sum_{m,n\in N}\left(f\left(y_{m,n}\right)\right) - \sum_{m\in N}\psi_m^* \sum_{n\in N} y_{m,n} I_m^{s*} \quad (9)$$

Theorem: For every link (*m*, *n*), the optimal dynamic weight reduction $y_{m,n}^*$ is as follows:

$$\sum_{m,n\in N}\left(f\left(y_{m,n}^*\right)\right) - \sum_{m\in N}\psi_m^* \sum_{n\in N} y_{m,n}^* I_m^{s*} \le 0. \quad (10)$$

Proof : For every link (*m*, *n*) in Eq. (10), $\frac{\partial\left(f(y_{m,n}) - \psi_m(t) I_m^s y_{m,n}\right)}{\partial y_{m,n}}\Big|_{y_{m,n}=y_{m,n}^*} = 0$ is evaluated to find the optimal $y_{m,n}^*$. Therefore, we obtain $y_{m,n}^* = \left(\frac{\partial f}{\partial y_{m,n}}\right)^{-1}(\psi_m^* I_m^s)$. Since inequality (10)

has to be preserved for all time t, $y^*_{m,n}$ is applied if and only if $\frac{\partial f}{\partial y_{m,n}}|_{y_{m,n}=0} \leq \psi_m I^s_m \leq \frac{\partial f}{\partial y_{m,n}}|_{y_{m,n}=(1-\alpha)_{w^o_{m,n}}}$ is true. Also, $y^*_{m,n}$ equals its upper bound $(1-\alpha)\, w^0_{m,n}$ if $\frac{\partial f}{\partial y_{m,n}}|_{y_{m,n}=(1-\alpha)} w^0_{m,n} < \psi_m$. Hence, $y^*_{m,n}$ and consequently $w^*_{m,n} = w^o_{m,n} - y^*_{m,n}$ are obtained.

IV. Numerical evaluation

We apply two different network structures for numerical evaluation. The network structures are a complete network and star network. Each network has five nodes. We numerically evaluate the final number of recovered nodes in each network. For each network, we simulate three intervention scenarios for the spread of viruses: 1) non-intervention scenario in Eqs. (1-3) with $w_{m,n}(t) = w^o_{m,n} \quad \forall t$, 2) intervention scenario without asymptomatic infection ($p_a = 0$) in Eqs. (5-7) to show the effect of neglecting the asymptomatic infection from the model and 3) intervention scenario only for the symptomatic infected nodes in Eqs. (5-7). For evaluation purpose, the values of the model parameters are set as follows: $r_a = 1$, $p_a = 0.33$, $\delta^{as} = 0.4$, $\delta^{as} \leq \delta^s \leq 1$, $\beta = \frac{2}{\rho}$, $a = 0.1$ and $w^o_{m,n} = 1$ where ρ is the maximum eigenvalue of the unweighted undirected network $w_{m,n}(t) = w^o_{m,n} = 1 \quad \forall t$. Also, we set the weight reduction cost function in Eq. (4) to be strictly convex $f(\cdot) = \left(w^o_{m,n} = 1 - w_{m,n}\right)^2$. The initial conditions are $S_m(0) = 0.99$, $I^{as}_m(0) = 0.05$, $I^s_m(0) = 0.05$ and $R_m(0) = 0\ \forall m \in N$. For all networks, number of recovered nodes in the absence of the controller $\left(w_{m,n}(t) = w^o_{m,n}\right)$ is higher than number of recovered nodes when the controller is dynamically changing over time. We evaluate both the model that considers the asymptomatic infection $p_a > 0$ and the one that does not consider the asymptomatic infection $p_a = 0$ (full knowledge of infection). We found that the link weight is more reduced in case of SIR model without asymptomatic infection because every infected node is assumed to be detected and the corresponding intervention is large e.g. more weight reduction. For SIR model with asymptomatic infection, only the symptomatic infected nodes are detected and the corresponding intervention is lower than the intervention for the SIR model without asymptomatic infection as shown in Figure 2. In addition, number of recovered nodes is lower for the SIR model without asymptomatic infection than for the SIR model with asymptomatic infection as shown in Table II. We also addressed the tradeoff between the total amount of weight reduction and number of infected nodes by evaluating the ratio between two quantities. For $\delta^{as} = \delta^s = 1$, the tradeoff values for the complete network in case of a realistic model and a fully knowledge model are 3.24 and 7.13, respectively. Similarly, the tradeoff values for the star network in case of realistic model and fully knowledge model are 1.33 and 2.84, respectively. It is not surprising that realistic model has less tradeoff value than the fully knowledge model, since the edge weights are not reduced for asymptomatic infected nodes in the realistic model. To clarify, the number of infected nodes when the SIR model without asymptomatic is considered is less due to the intensive contact weight reduction for longer time as shown in Figure 2. Recall that in the SIR model without asymptomatic infection, the model assumed that all infected cases are detectable which results in larger weight reduction. On the other hand, in the SIR model with asymptomatic

infection, the intervention is not applied between asymptomatic infected nodes and susceptible nodes because the infection is indiscernible and full contact rate is preserved. Finally, the influence of the network structure on the intervention is clearly observed with the star network which is highly heterogeneous in comparing with the complete network.

V. Conclusions and future work

In this letter, we presented the adaptive contact weighted networks to minimize a linear combination of the total number of infection cases and the weight reduction cost when viruses spread in computer networks. We briefly presented the networked SIR approach considering both detectable and undetectable infected nodes. The model reflects a realistic scenario that the intervention is not applied when a susceptible node has a contact with an asymptomatic infected node. Using the numerical evaluation, we found that the assumption that all infected nodes are detectable leads to intensive reduction in contact weights for longer time. Also, our approach captures the different dynamical weight reduction between homogeneous and heterogeneous networks. Our future work will be focused on proposing numerical methods to evaluate the optimal control on large complex networks and to analyze the sensitivity of the optimal solution to the model parameters.

Acknowledgments

This work has been partially supported by the Defense Threat Reduction Agency (DTRA) [award number HDTRA1-11-1-0016]; the National Institute of General Medical Sciences of the National Institutes of Health (NIH) [Models of Infectious Disease Agent Study (MIDAS) award number 2U01GM070694-09]; the National Science Foundation (NSF) [Network Science and Engineering Grant CNS-1011769]; and by the National Science Foundation (NSF) [award number DMS-1201427]. The content, any opinions, findings, and conclusions or recommendations expressed in this paper is solely the responsibility of the authors and does not necessarily represent the official views of DTRA, the NSF, or the NIH.

References

1. Kephart JO, White SR. Direct-graph epidemiological models of computer viruses. IEEE Computer Society Symposium on Research in Security and Privacy. 1991
2. Chen Z, Ji C. Spatial-Temporal Modeling of Malware Propagation in Networks. IEEE Transactions on neural networks. 16:2005.
3. Yoo J, Choi S, Kim C.-k. The capacity of epidemic routing in vehicular networks. IEEE Communications Letters. 2009; 13(6):459–461.
4. Zhang X, Neglia G, Kurose J, Towsley D. Performance modeling of epidemic routing. Computer Networks: The International Journal of Computer and Telecommunications Networking archive. Jul; 2007 51(10):2867–2891.
5. Lagorio C, et al. Quarantine Generated Phase Transition in Epidemic Spreading. Phys. Rev. E. 2011; 83:026102.
6. Youssef M, Scoglio C. Mitigation of epidemics in contact networks through optimal contact adaptation. Journal of Mathematical Biosciences and Engineering MBE. 2013; 10(4)
7. Tsai, J.; Moors, T. Proccedings of the ACoRN Early Career Researcher Workshop on Wireless Multihop Networking. Australia: 2006. A Review of Multipath Routing Protocols: From Wireless Ad Hoc to Mesh Networks.
8. Youssef, M.; Scoglio, C. JTB: Journal of Theoretical Biology. Vol. 283. Elsevier; 2011. An individual-based approach to SIR epidemics in contact networks; p. 136-144.
9. Prakash, B.; Tong, H.; Valler, N.; Faloutsos, M.; Faloutsos, C. Virus Propagation on Time-Varying Networks: Theory and Immunization Algorithms, ECML-PKDD 2010. Barcelona, Spain: 2010.

10. Zhang C, Yang X, Zhu Q, Liu W. Optimal control in a novel computer virus spread model. Journal of Information and Computational Science. 2011; 8:10, 19291938.

11. Van Mieghem P, Omic JS, Kooij RE. Virus Spread in Networks. IEEE/ACM Transaction on Networking. 17(1):1–14, 2009.

12. R Khouzani, MH.; Sarkar, S.; Altman, E. Proceedings of IEEE INFOCOM. China: 2011. Optimal Control of Epidemic Evolution.

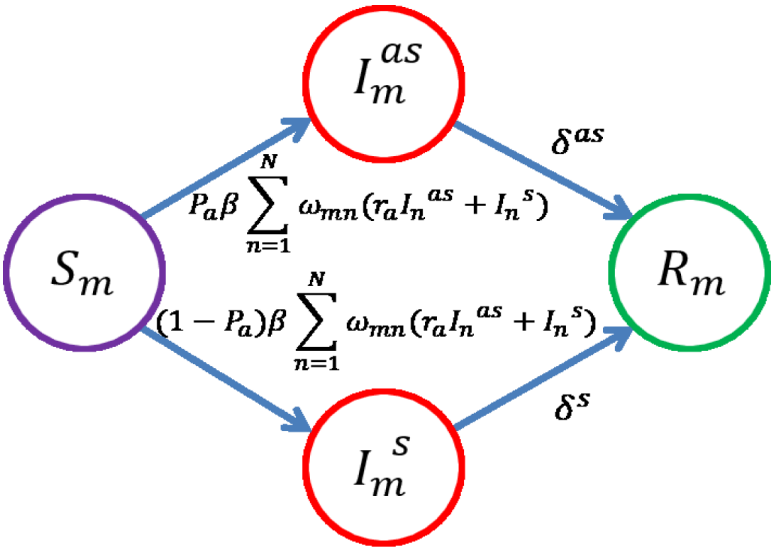

Fig. 1.
SIR model with symptomatic and asymptomatic infections

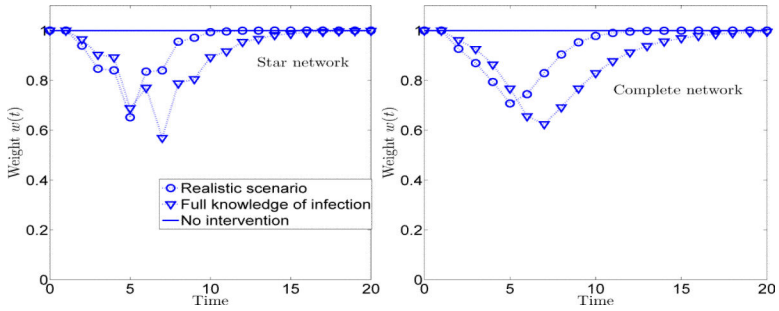

Fig. 2.

Contact weight $w(t)$ of the three scenarios given $\delta^{as} = \delta^{s} = 1$.

TABLE I

Definitions of parameters and variables

Parameters	**Definition**
T_f, β	Final time and infection rate
δ^s, δ^{as}	Cure rate for symptomatic and asymptomatic infections
α	Global minimum social contact coefficient
r_a	Infectivity of asymptomatic infection
P_a	Probability of a newly infected node is asymptomatic
Variables	Definition
$w_{m,n}$	Link weight at time t with initial value $w^{\circ}_{m,n}$
$S_m(t)$	The susceptible probability of node m
$I^{as}_m(t)$	The asymptomatic infection probability of node m
$I^{s}_m(t)$	The symptomatic infection probability of node m

TABLE II

Number of infected nodes with $\dagger^{as}$ =0.4

Networks and interventions	$\dagger^{s}$=0A	**0.6**	**0.8**	**1**
Complete net. no intervention	5	4.98	4.78	4.34
Complete net. realistic scenario	5	4.94	4.22	3.26
Complete net. infection full knowledge	4.9	4.73	3.74	2.89
Star net. no intervention	4.85	4.47	4.04	3.66
Star net. realistic scenario	4.75	4.18	3.52	2.72
Star net. infection full knowledge	4.68	4.12	3.37	2.49