RESEARCH

Mitigate Epidemic Spreading via Contact Blocking in Temporal Networks

Xunyi Zhao[1*] and John R.S. Smith[1,2]

*Correspondence: X.zhao-10@student.tudelft.nl
[1]Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Mekelweg 4, Delft, 2628 CD, The Netherlands
Full list of author information is available at the end of the article

Abstract

First part title: Text for this section.

Second part title: Text for this section.

Keywords: sample; article; author

Introduction

Human contact networks facilitate the spreading of Epidemic and information. Contact networks such as face-to-face contact networks are temporal networks, networks that evolve over time. In temporal networks, the activity (connected or not, having a contact or not) of a link between two nodes could change over time. This is in contract to static networks where the connection of each link is static. Epidemic/information spreading can be mitigated via the blocking physical contacts. Covid-19 measures like curfew, working at home, social distancing all aim to reduce physical contacts. These measures treat at least a subgroup of the population in the same way. In this work, we address the further question how to mitigate the epidemic spreading more effectively via selecting the contacts to block heterogeneously and strategically. We propose to develop contact removal strategies utilizing the temporal network properties of contacts.

A temporal network $\mathcal{G} = (\mathcal{N}, \mathcal{C})$ observed within a time window $[0, T]$ can be represented by its set of contacts $\mathcal{C} = \{c(i,j,t), t \in [0,T], i,j \in \mathcal{N}\}$ among its set $\mathcal{N}$ of N nodes, where contact $c(i,j,t)$ occurs between node i and j at time step t. We consider the Susceptible-Infected-Recovered (SIR) epidemic spreading model on a temporal network: Initially, a seed node is randomly selected and infected at $t = 0$, whereas all the other nodes are susceptible; An infected node infects a susceptible node with a probability β when a contact happens between the two nodes; Each infected node could become recovered with a probability γ at each time step. An recovered node will neither be infected nor infect any other node. We assume that the aggregated network $\mathcal{G}_\mathcal{W}$ that corresponds to the temporal network is given, where two nodes i and j are connected by a link $l(i,j)$ if the two nodes have at least one contact and the link is associated with a weight representing the number of contacts in between. The objective is to mitigate the epidemic spreading via blocking a given percentage ϕ of contacts based on the aggregated network. The fraction ϕ of contacts removed corresponds to cost of the mitigation. We propose probabilistic contact removal strategies. Namely, the probability that a contact $c(i,j,t)$ is removed is a generic function of a centrality metric of link $l(i,j)$ in the aggregated network and the time t of the contact. Each centrality metric leads to a unique mitigation strategy in contact removal. The impact of an SIR epidemic spreading can evaluated via the following performance measures, which will be used to evaluate the mitigation strategies: the average prevalence over time, where the prevalence at a time step is the number of infected nodes; the maximal prevalence, so called peak height, which suggests the maximal demand in e.g. hospital; the time to reach the peak prevalence, so called peak time, which indicates the time to prepare the medical resources for the peak demand.

The mitigation strategies that we have proposed are evaluated in 6 real-world temporal networks. We find that the epidemic can be better mitigated when contacts between node pairs that have fewer contacts are removed with a higher probability. Removing contacts that happen earlier in time could further enhance the mitigation effect. The number of contacts between a node pair is heterogeneous. It seems that the mitigation effect is better if the average number of contacts removed per node pair varies less. Static network studies have shown that a weighted network tends to be more robust against epidemic spreading i.e. having a higher epidemic threshold if its largest eigenvalue is smaller. The resultant aggregated network after contact removal, however, may have a lower prevalence if its largest eigenvalue is larger. This implies that the underlying temporal network may lead to new phenomena in epidemic spreading that differ from what we have learned from static networks.

The influence of temporal networks on dynamic processes has been widely studied [1, 2]. Gemmetto et. al have investigated the epidemic mitigation via excluding a sub-group of nodes in a temporal network [3]. Link blocking strategies using link centrality metrics to suppress information diffusion has been explored in [4]. The links to block are selected from the aggregated network. When a link is blocked, all contacts associated with the links are all removed. In this work, we investigate more in-depth at contact level, i.e. how to select a given number of contacts to block. Moreover, the consideration of the time of a contact in contact removal strategies may inspire the decision over when a mitigation should be introduced. Our previous [5] has addressed the same question, however, was confined to Susceptible-Infected (SI) model, which is a special case of SIR model. The SIR model considered in this work brings in not only the complexity in the dynamics of epidemic spreading, but also requires more comprehensive evaluation of mitigation

Methods

Firstly, we propose to select the contacts to block based on centrality metrics and the time of each contact. Specifically, the probability that a contact $c(i, j, t)$ is removed is defined as a function of a given centrality metric of the corresponding link $l(i, j)$ in the aggregated network and the time t of the contact. This function also ensures that a fraction ϕ of contacts are removed on average. Afterwards, we will introduce the real-world temporal networks and simulations that will be used to simulate the epidemic spreading process and further to evaluate the effect of the mitigation strategies.

Link centrality metrics

Since the probability that that a contact (i, j, k) is removed is based on a given centrality metric of the link (i, j) in the aggregated network $\mathcal{G}_{\mathcal{W}}$, we propose firstly a set of link centrality metrics in $\mathcal{G}_{\mathcal{W}}$. Each centrality metric will lead afterwards to a unique mitigation strategy.

An aggregated network $\mathcal{G}_{\mathcal{W}}$ can be derived from a given temporal network $\mathcal{G}$ observed within time window $[1, T]$. The weight of each link in the aggregated network represents the number of contacts between the two corresponding nodes in the temporal network. We consider the following link centrality metrics based on the weighted aggregated network:

- *Degree product* is the product of the degrees of the two end nodes of a link, where the degree of a node is the number of links incident to the node.
- *Strength product* is the product of the strengths of the two end nodes of a link, where the strength of a node is the total weights of all the links incident to the node. Equivalently, the strength of a node tells the total number of contacts the node involves in the temporal network.
- *Betweenness* is the number of shortest paths in the unweighted aggregated network that traverse the link between all possibly node pairs.
- *Link weight* is the weight of a link in the aggregated network. It is the total number of contacts between the two end nodes.
- *Weighted eigenvector component product* is the product of the principal eigenvector components of the two end nodes, where the principal eigenvector is the eigenvector corresponds to the largest eigenvalue of the weighted aggregated network.
- *Unweighted eigenvector component product* is the product of the principal eigenvector components of the two end nodes, where the principal eigenvector is the eigenvector corresponds to the largest eigenvalue of the unweighted aggregated network.
- *Random* sets 1 for all links. This metric is called random because this metric will lead to the strategy where the probability that a contact $c(i, j, t)$ is removed is independent of the centrality of (i, j).

Contact removal probability

Given a link centrality metric m, we can compute the centrality for m_{ij} for each link $l(i, j)$ in the aggregated network. Consider the simple case where the probability p_{ij} that a contact $c(i, j, t)$ between i and j is removed is independent of the time t. In this case, we propose

$$p_{ij} = m_{ij} \frac{\phi \sum_{ij} w_{ij}}{\sum_{ij} (w_{ij} m_{ij})} \tag{1}$$

where w_{ij} is the weight of link $l(i, j)$ in the aggregated network, and the normalization ensures that on average a fraction ϕ of contacts will be removed. The probability that a contact is removed is assumed to be proportional to the centrality m_{ij} of the corresponding link $l(i, j)$.

We find that some centrality metrics are highly heterogeneous. It is possible that the removal probability calculated by (1) is larger than 1 for contacts whose associated link $l(i, j)$ has an extremely large centrality m_{ij}. In such cases, the actual fraction of contacts removed can be lower than the expected ϕ, if all contacts with

Datasets	Nodes	Links	Contacts	Duration
HighSchool11	126	1709	28561	3.15
HighSchool12	180	2220	45047	8.44
WorkPlace13	92	755	9827	11.43
WorkPlace15	217	4274	78249	11.50
MIT1	74	355	29107	6.99
MIT2	45	200	22714	6.99
MIT	96	5078	1086404	232.30

Table 1 Basic properties of the temporal networks: the number of nodes, links and contacts. The duration is the duration of the observation time window [1,T] measured in days, thus T times the duration per discrete time step.

removal probability larger than 1 are removed. Therefore, we set the removal probabilities of those contacts to 1 and re-normalize the removal probability among the rest contacts. This process is repeated until the removal probabilities of all remaining contacts are between 0 and 1, while the actual fraction of contacts removed is the same as expected ϕ .

We further generalize the definition of the contact removal probability p_{ij} as

$$p^*_{ij} = m^\alpha_{ij} \frac{\phi \sum_{ij} w_{ij}}{\sum_{ij}(w_{ij} m^\alpha_{ij})} \tag{2}$$

The removal probability of a contact $c(i,j,t)$ is proportional to a polynomial function of m_{ij}. Our choice in (1) corresponds to the case when $\alpha = 1$. The random strategy, i.e. all contacts have the same probability of being removed, corresponds to the case when $\alpha = 0$. Consider (1) and the reciprocal metric $\frac{1}{m_{ij}}$ as a new centrality metric. This strategy is equal to the general definition (2) where metric m_{ij} is considered and $\alpha = -1$.

Hence, we consider removal probability (1) using the list of centrality metrics proposed and their reciprocals as well as the random strategy, which correspond to the general definition of (2) where $\alpha = 1, -1, 0$, respectively.

Furthermore, we generalize our strategy by considering the timestamps of the contacts. This is motivated by the intuition that early intervention, e.g. blocking early contacts, could be possibly more effective. We propose the probability $p_{ij}(t)$ that a contact $c(i,j,t)$ between i and j at t is removed as

$$p_{ij}(t) = m_{ij} f(t) \frac{\phi \sum_{ij} w_{ij}}{\sum_{ij}(w_{ij} m_{ij} f(t))} \tag{3}$$

where $f(t)$ implies the preference to block contacts at specific period. The probability that $c(i,j,t)$ is removed is proportional to $m_{ij} \cdot f(t)$.

As a simple start, we consider $f(t) = 4 \cdot 1_{t \leq T/2} + 1_{t > T/2}$, $f(t) = 1_{t \leq T/2} + 4 \cdot 1_{t > T/2}$ and $f(t) = 1$, where the indicator function 1_y is one is the condition y is true, and otherwise it is 0. They correspond to the preference for removing contacts happening in $[1, T/2]$, in $(T/2, T]$ and no preference for the timestamps of the contacts, respectively.

Datasets

We consider the following real-world temporal physical contact networks:

- HighSchool11&12[6] record the physical contacts between students in a high school in Marseilles, France. The two datasets consider two different groups of students.
- WorkPlace13&15[7] capture the contacts between individuals in an office building in France. the two datasets are measured from different groups of individuals respectively.
- MIT[8] are human contact network among students of the Massachusetts Institute of Technology. The MIT dataset has been measured for about 8 months.

All networks are undirected. Their basic properties are given in Table 1. The duration of each time step is 1 second in all the networks.

For the MIT dataset, we choose randomly two observation period, each of about one-week time. The temporal networks corresponding to these two periods are called MIT1 and MIT2. In this way, all the six temporal networks (HighSchool11&12, WorkPlace13&15, MIT1&2) are comparable in observation window. They will be used to study the impact of the mitigation strategies on the average prevalence over time, the focus of this work.

However, all the six networks have a short duration of the observation window, within 12 days. In order to observe the peak prevalence in the SIR process, the observation window of a temporal network needs to be long in duration. When we study the performance measure like peak height/prevalence and peak time, we repeat each of the temporal network HighSchool11&12, WorkPlace13&15 respectively for 5 times and consider the MIT data. The constructed networks, *HighSchool11&12, *WorkPlace13&15 which repeats one temporal network periodically are also called periodic networks [XXX add ref: Zhang YQ, Li X, Vasilakos AV. Spectral Analysis of Epidemic Thresholds of Temporal Networks. IEEE Trans Cybern. 2020 May;50(5):1965-1977.] Each constructed network has a duration five times as large as the orginal network. We consider 4 constructed network *HighSchool11&12, *WorkPlace13&15 and the MIT dataset to study the performance of the strategies in terms of peak prevalence and peak time.

Simulation

In this subsection, we will introduce the simulation of the SIR spreading process and the choice of parameters. The performance measures to evaluate the mitigation strategies will be discussed in the next section.

We consider the following discrete time SIR spreading process: a seed node is chosen as infected at $t = 0$, while are the other nodes are susceptible. Each contact between an infected node and a susceptible node could lead to an infection with probability β. At each time step, each infected node recovers with a recovery probability γ. We consider as an example the infection probability $\beta = 0.01$, which leads to a prevalence around the order of 10% by the end of the time window in the six temporal networks, when the recovery probability $\gamma = 0$. Furthermore, we consider the recover probability per time step $\gamma = 0.82 * 10^{-6}$ or $\gamma = 0$. The former, $\gamma = 0.82 * 10^{-6}$ leads approximately to a recovery probability 10% per day.

In the simulation, we simulate the exact infection and recovery process except the following approximation in the recovery process. If there is no contact in the whole network for the period $t_0, t_0 + t$, we update the state of each node only at the end of this time window $t_0 + t$ instead of at each of the t time steps and we approximate the probability for a node that is infected at t_0 to become recovered at $t_0 + t$ as $1-(1-\gamma)^t = 1 - e^{tlog(1-\gamma)} \approx 1 - e^{-t\gamma} \approx t * \gamma$. This approximation is motivated by that fact that γ is small. In the datasets we have considered, the longest gap that no contact happens is around one day. Correspondingly, the average prevalence is the number of infected nodes over the time steps when at least one contact happens in the network.

For each temporal network, we select each node as a possible seed node. For each seed node, we iterate the following for five times. In each iteration, the fraction ϕ of contacts to be removed are selected according to the probability (1) using the given link centrality metric; The SIR process starting from the given seed is performed on the temporal network where the selected contacts are removed; the prevalence ρ is recorded at each time step when there is a contact in the network. For each network and centrality metric, we could obtain the average prevalence at a time step as the average over all possible seed nodes and the five iterations per seed node. The fraction ϕ of contacts to be removed is a control parameter and $\phi = 10\%$ and $\phi = 30\%$ are considered. Simulations are performed in the same way when the time factor $f(t)$ are taken into account via the contact removal probability defined in (3).

Results

In this section, we evaluate our contact removal strategies via three performance measures: the average prevalence and the peak height (the maximal number of infected at a time step) and the peak time (the time to reach the maximal number of infected.

Average prevalence

Firstly, we evaluate the strategies as defined in (1) where the probability that a contact $c(i, j, t)$ is removed is independent of the time of the contact t but do depend a centrality metric of the link $l(i, j)$ in the aggregated network. In total, 13 strategies are considered that correspond to the aforementioned centrality metrics and their reciprocals. Figure 2 demonstrates the prevalence $\rho(t)$ over time in two periodic networks *HighSchool12 and *WorkPlace15 as examples when each of the 13 strategies is performed and 10% contacts are removed. The long duration of the observation window $5T$ of a periodic network allows us to observe the peak prevalence. The ordering of the prevalence $\rho(t)$ at each time step for the 13 strategies are relatively stable. This suggests that relative performance of the mitigation strategies in terms of average prevalence is less sensitive to duration of the observation time window.

We use the original network HighSchool11&12, WorkPlace13&15, MIT1&2 to evaluate the blocking strategies with respect to the average prevalence. These networks are comparable in duration of the observation time window, i.e. within 12 days.The average prevalence $E[\rho]$ is the average fraction of infected nodes at each time step when there is at least one contact in the network. It implies the average heath care resources needed over time.

We start with the simple case when the recovery rate $\gamma = 0$. In this case, the SIR model is equal to the Susceptible-Infected (SI) model. The performance of the

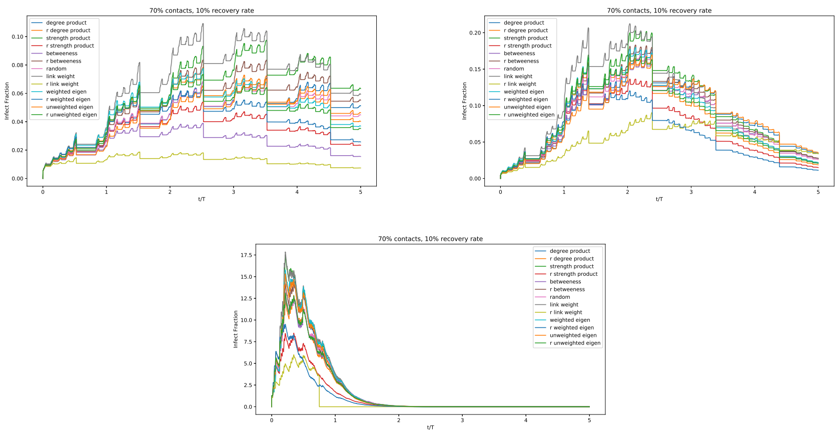

Figure 2 Left: *HighSchool12. Right: *WorkPlace15. The prevalence ρ of the SIR model over time on two periodic networks respectively mitigated via 13 contact blocking strategies defined by (1). The infection rate is $\beta = 0.01$, the recovery rate is $\gamma = 0.82 * 10^{-6}$, approximately when the recovery probability is 10% per day and 30% of the contacts are removed.

Metrics	HighSchool11	HighSchool12	WorkPlace13	WorkPlace15	MIT1	MIT2
degree product	0.041	0.036	0.028	0.094	0.124	0.191
1/degree product	0.050	0.043	0.028	0.107	0.111	0.178
strength product	0.045	0.038	0.028	0.099	0.120	0.200
1/strength product	0.043	0.036	0.027	0.091	0.098	**0.160**
betweeness	0.045	0.031	0.029	0.098	0.099	0.203
1/betweeness	0.044	0.038	0.028	0.101	0.121	0.208
random	0.049	0.036	0.024	0.091	0.125	0.183
link weight	0.050	0.041	0.028	0.109	0.126	0.181
1/link weight	**0.039**	**0.031**	**0.027**	**0.079**	**0.096**	0.187
weighted eigen	0.044	0.037	0.027	0.091	0.125	0.238
1/weighted eigen	0.051	0.036	0.029	0.097	0.099	0.169
unweighted eigen	0.042	0.040	0.027	0.098	0.116	0.206
1/unweighted eigen	0.041	0.035	0.028	0.107	0.093	0.193

Table 4 The average prevalence $E[\rho]$ when the recovery rate is $\gamma = 0\%$ per step, and $\phi = 10\%$ of the contacts are removed from each temporal network using removal probability (1) based on each centrality metric.

Additional file 2 — Sample additional file title
Additional file descriptions text.

Metrics	HighSchool11	HighSchool12	WorkPlace13	WorkPlace15	MIT1	MIT2
degree product	0.025	0.025	0.021	0.050	0.109	0.170
1/degree product	0.036	0.027	0.022	0.066	0.080	0.156
strength product	0.037	0.028	0.022	0.062	0.111	0.185
1/strength product	0.030	0.027	0.023	0.057	0.070	0.109
betweeness	0.033	0.024	0.024	0.058	0.080	0.160
1/betweeness	0.035	0.028	0.023	0.060	0.102	0.170
random	0.034	0.026	0.021	0.059	0.091	0.171
link weight	0.043	0.035	0.027	0.089	0.103	0.176
1/link weight	**0.020**	**0.018**	**0.021**	**0.036**	**0.056**	0.113
weighted eigen	0.032	0.028	0.024	0.063	0.109	0.181
1/weighted eigen	0.039	0.028	0.023	0.065	0.065	**0.095**
unweighted eigen	0.024	0.026	0.023	0.055	0.089	0.176
1/unweighted eigen	0.038	0.027	0.021	0.064	0.087	0.132

Table 5 The average prevalence $[\rho]$ when the recovery rate is 0% per step, and $\phi = 30\%$ of the contacts are removed from each temporal network using removal probability (1) based on each centrality metric.

Metrics	HighSchool11	HighSchool12	WorkPlace13	WorkPlace15	MIT1	MIT2
degree product	0.032	0.022	0.013	0.043	0.076	0.140
1/degree product	0.035	0.023	0.014	0.050	0.075	0.130
strength product	0.037	0.024	0.014	0.045	0.084	0.145
1/strength product	0.036	0.022	0.013	0.046	0.064	0.116
betweeness	0.036	0.021	0.013	0.045	0.072	0.125
1/betweeness	0.035	0.021	0.013	0.047	0.082	0.130
random	0.036	0.022	0.014	0.048	0.073	0.135
link weight	0.041	0.025	0.015	0.052	0.073	0.138
1/link weight	**0.031**	**0.019**	0.014	**0.037**	**0.056**	**0.111**
weighted eigen	0.038	0.024	0.014	0.050	0.076	0.146
1/weighted eigen	0.040	0.024	**0.013**	0.047	0.069	0.115
unweighted eigen	0.034	0.021	0.013	0.045	0.079	0.131
1/unweighted eigen	0.039	0.022	0.013	0.048	0.077	0.127

Table 6 The average prevalence $[\rho]$ when the recovery rate is $\gamma = 0.82 * 10^{-6}$ per step, and $\phi = 30\%$ of the contacts are removed from each temporal network using removal probability (1) and $f(t) = 1_{t \le T/2} + 4 \cdot 1_{t > T/2}$ based on each centrality metric. Contacts happening late i.e. $t > T/2$ in time are more likely to be removed.