Near-Optimal Spectral Disease Mitigation in Healthcare Facilities

Masahiro Kiji, D. M. Hasibul Hasan, Alberto M. Segre, Sriram V. Pemmaraju, and Bijaya Adhikari

Dept. of Computer Science, The University of Iowa

Email: [masahiro-kiji,dmhasibul-hasan,alberto-segre,sriram-pemmaraju,bijaya-adhikari]@uiowa.edu

For the CDC MInD Healthcare Network

Abstract—Healthcare associated infections (HAIs) impose a substantial burden, both on patients and on the healthcare system. Designing effective strategies by using interventions such as vaccination, isolation, cleaning, mobility modification, etc., to reduce HAI spread is an important computational challenge. *Spectral approaches* are quite useful for modeling and solving problems of reducing disease spread over contact networks, but they have not been used for disease-spread models and contact networks that are specific for HAIs. Our main contribution in this paper is to close this gap. We make 3 specific contributions. (i) We present the first epidemic threshold results on *temporal bipartite networks*, i.e., a time-varying sequence of bipartite people-location network, for the Susceptible-Infected-Susceptible (SIS) model. (ii) We leverage our epidemic threshold result to pose the HAI mitigation problem as minimizing the spectral radius of the system matrix, while removing few nodes or edges. We present a scalable combinatorial algorithm that provides approximation guarantees. (iii) Through extensive experiments on actual healthcare contact networks derived from operations data from the University of Iowa Hospitals and Clinics, Carilion Clinic, and several other healthcare facilities, we show that our algorithm consistently outperforms a number of baselines (random, degree, top-k, eigen centrality) both in terms of reducing the spectral radius of the system matrix and in terms of reducing infections.

Index Terms—HAIs, epidemic threshold, SIS model, spectral radius

I. Introduction

Healthcare associated infections (HAIs), for example *C. diff* infections (CDI) and Methicillin-resistant *Staphylococcus aureus* (MRSA) infections, impose a substantial burden, both on patients that suffer from these infections and on the healthcare system. It is generally believed that these infections can be reduced by better hand hygiene by healthcare professionals (HCPs) [12], better room cleaning protocols [7], modifying the mobility patterns of HCPs and patient flow [23], [8], changing architectural layout of healthcare facilities [9], etc. HAIs such as CDI and MRSA infections are often spread through pathogen or spores deposited on and picked up from surfaces (e.g., door knobs, bed rails, keyboards, etc.). So it is critical that models for HAI spread take locations into account. In a healthcare setting (e.g., in an in-patient hospital unit) HCPs provide regular care to patients by visiting patient rooms. While these visits are crucial for the care of patients, they also serve as potential pathways for pathogen transmissions. Our goal is to model and solve the problem of reducing disease-spread in healthcare facilities using interventions such as isolation, cleaning, mobility modification, etc.

Spectral approaches have turned out to be quite useful for modeling and solving problems of reducing disease spread over contact networks. In this context, spectral approaches characterize the spread of disease over a network in terms of the spectrum (sequence of eigenvalues) of the adjacency matrix or the Laplacian of the underlying network. There is now a growing body of literature [6], [3], [13], [14], [16] on deriving the *epidemic threshold*, a quantity that is a function of disease parameters, such that (informally speaking) the disease will die out iff the largest eigenvalue of the network is below the threshold. Using epidemic thresholds, one can model problems of disease mitigation as problems of removing nodes or edges of the network so as to bring the largest eigenvalue below the threshold [20], [21], [19], [4], [17], [15]. Here node deletion could model vaccinating or isolating (for diseases which do not have vaccines) an individual and edge deletion could model modifying mobility so as to prevent some contacts.

While spectral approaches have been widely used for both static networks and for dynamic, time-varying networks [14], [16], they have not been used in the context of people-location networks that are important for modeling HAIs. Our main contribution in this paper is to apply spectral approaches to the disease mitigation problem in healthcare settings. More specifically, we make 3 main contributions in this paper:

Epidemic threshold. We model interactions in a healthcare facility as a *temporal bipartite network*, i.e., as a time-varying sequence of bipartite people-location networks. We then derive an *epidemic threshold* for the Susceptible-Infected-Susceptible (SIS) model for temporal bipartite networks while making realistic assumptions on healthcare practices. To the best of our knowledge, these are the first results on epidemic thresholds for temporal bipartite graphs.

Effective algorithm. Our epidemic threshold result allows us to pose problems of disease mitigation in healthcare facilities as problems of minimizing the spectral radius of the system matrix, while removing few nodes or edges. These problems are not only computationally hard [19] but are expensive to solve even via approximation algorithms or heuristics. We present a near optimal combinatorial approximation algorithm that avoids costly eigenvalue computation.

Extensive experiments. Through extensive experiments on

actual healthcare contact networks derived from operations data from the University of Iowa Hospitals and Clinics and several other healthcare facilities, we show that our algorithm consistently outperforms a number of baselines both in terms of reducing the spectral radius of the system matrix and in terms of reducing infections. This shows the applicability of our approach to HAI mitigation.

II. PRELIMINARIES

Temporal bipartite networks: We model mobility within a healthcare facility as a series of daily activity snapshots with a periodicity of τ. Specifically, we represent mobility as a temporal network $\mathcal{G} = \{G_1, G_2, G_3, \ldots, G_\tau\}$, where each $G_t(P, L, E_t) \in \mathcal{G}$ is an undirected, unweighted, bipartite network representing the interactions between people P and locations L at time t. Each edge $e(p, \ell) \in E_t$, $p \in P, \ell \in L$ indicates that a person p was at a location ℓ at time t. Without loss of generality, we assume every G_t has the same node set in both partitions P and L (otherwise just define $P = \cup_t P_t$ and $L = \cup_t L_t$). The periodicity of τ implies that $G_{\tau+1} = G_1$, $G_{\tau+2} = G_2$ and so on implying that we have an implicit infinite sequence of activity snapshots. This periodicity assumption is motivated by the fact that healthcare activities tend to be regular and schedules are often periodic. So τ could be 2 and G_1 and G_2 could represent day and night activity or τ could be 7 and each G_i could represent activities of a day of the week.

Epidemics on Networks: In the *Susceptible-Infected-Recovered (SIR)* disease-spread model over a static network, in each step a susceptible node gets infected by its infected neighbor with probability $0 \leq \beta \leq 1$ and an infected node recovers with probability $0 \leq \delta \leq 1$. Recovered nodes do not get re-infected. The SIS model, which we discuss further in this work, does not have the Recovered state and infected nodes become susceptible again after recovery. The SIS model is appropriate for our work because HAIs such as CDI are not known to confer immunity and, in fact, prior CDI is a significant risk factor for subsequent acquisition of CDI [5]. These models have been generalized to temporal networks such that an infected node at time t can only infect its neighbors in G_t [14].

III. PROBLEM FORMULATION

A. Informal Problems.

We ask two specific questions with an aim of developing preemptive measures against disease outbreaks in healthcare facilities. *(i) Given limited resources, which individuals should be isolated (or vaccinated) in order to reduce the likelihood of epidemic outbreaks? (ii) Given limited flexibility on schedule, which individual-location contacts should be prohibited in order to reduce the likelihood of epidemic outbreaks?* Note that isolating a high number of individuals or prohibiting arbitrarily high number of individual-location contact prevents healthcare workers from providing effecting patient care. These two questions naturally lead to node and edge deletion problems respectively on the bipartite temporal contact network $\mathcal{G}$ which

models mobility within healthcare facilities. Note that deleting a node in P (i.e., a person) removes the node from all G_t. Similarly, deleting an edge from E_t means that the edge is gone even when the activity repeats, after τ time units. In order to make our problems formal and then provide solutions to these problems, we take two next steps: **(1)** We describe the standard SIS disease-spread model in temporal bipartite networks representing mobility in healthcare facilities and **(2)** We characterize the vulnerability of $\mathcal{G}$ to outbreaks in terms of an epidemic threshold that we derive. The next two subsections describe these.

B. SIS model on Dynamic Bipartite Networks.

The standard SIS model has been extended to temporal networks [2], [14]. It also has been studied in the context of bipartite networks [24]. Here, we describe the SIS model in temporal bipartite networks in the context of HAIs. We assume the process begins at time $t = 1$ in G_1, where a set $S \subseteq P$ of people are infected. At each subsequent time t, an infected person $p \in P$ has an opportunity to contaminate all locations $l \in L$ they visit at time t. More precisely, an infected person p infects location l with probability β_p if $(p, l) \in E_t$. An infected person recovers and becomes susceptible again with probability δ_p. In the next time-stamp, $t + 1$, newly contaminated locations infect their visitors with probability β_l. Since we are modelling HAI spread, we model the stringent cleaning protocols often employed at healthcare facilities. To this end, we assume contaminated locations "recover" with probability $\delta_l = 1$ after they get their chance to infect the visitors. As mentioned earlier, at time steps $t > \tau$, the mobility is represented by $G_{((t-1) \bmod \tau)+1}$ and the process continues till the disease dies out.

C. Deriving an epidemic threshold for $\mathcal{G}$.

We begin by defining some notations. Let $\mathbf{B}_t$ be the $|P| \times |L|$ bi-adjacency matrix representing the graph G_t. Similarly, $\mathbf{p}_t$ be a binary vector of size $|P|$ indicating whether each individual in P is infected or not at time t. Let us define the *system matrix* of the bipartite temporal network $\mathcal{G}$ as follows:

$$\mathbf{S} = \prod_{t=1}^{\tau}(1 - \delta_p)\mathbf{I} + \beta_p\beta_l\mathbf{B}_t\mathbf{B}_{(t \bmod \tau)+1}^{\mathsf{T}} \qquad (1)$$

Intuitively, $\mathbf{S}$ captures progression of infection states over time by combining the recovery process (captured by the '$(1-\delta_p)\mathbf{I}$' term) and the process of infection spread (captured by the '$\beta_p\beta_l\mathbf{B}_t\mathbf{B}_{(t \bmod \tau)+1}^{\mathsf{T}}$' term). Now, let $\lambda_{\mathbf{S}}$ be the leading eigenvalue of $\mathbf{S}$. The epidemic threshold of $\mathcal{G}$ is characterized by the following theorem. The proofs of theorems/lemmas are deferred to the full version of the paper.

Theorem 1: If $\lambda_{\mathbf{S}} < 1$, then $\mathbf{p}_t$ is asymptotically stable at $\mathbf{0}$.

Note: Theorem 1 states that if $\lambda_{\mathbf{S}} < 1$ then the non-linear dynamical system representing pathogen propagation in $\mathcal{G}$ is stable when everybody is uninfected/healthy. It implies that under small perturbations, the system quickly returns to the state where everyone is healthy. Hence, regardless of initial
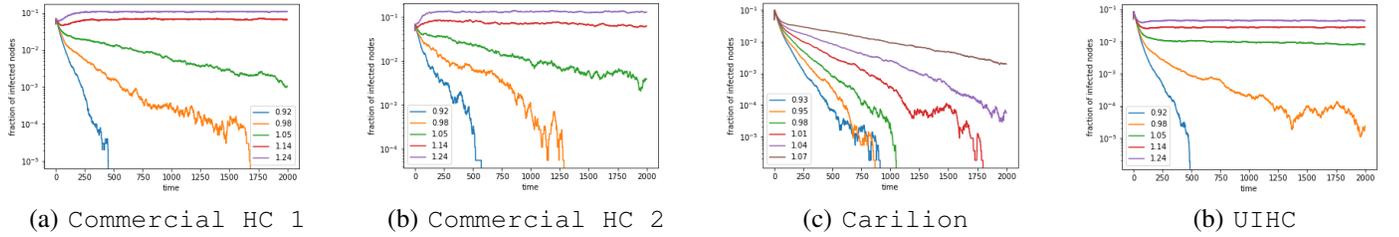
| (a) Commercial HC 1 | (b) Commercial HC 2 | (c) Carilion | (b) UIHC |

Fig. 1: Time (x-axis) vs number of infections (y-axis) for the SIS model on temporal bipartite graphs extracted from various healthcare facilities (see Section V for dataset description) for various values of $\lambda_\mathbf{S}$. As predicted by Theorem 1, infection dies out quickly when $\lambda_\mathbf{S} << 1$, survives for few time-stamps when $\lambda_\mathbf{S} \approx 1$ and continues on otherwise.

condition, $\mathbf{p}_t$ quickly converges to $\mathbf{0}$. On the other hand, if $\lambda_\mathbf{S} > 1$, no such conditions hold.

We empirically validated Theorem 1 on four different temporal bipartite networks extracted from various healthcare facilities. We varied $\lambda_\mathbf{S}$ and repeated the experiment for each healthcare facility (See Figure 1. As observed in the figure, the infection dies out quickly when $\lambda_\mathbf{S} << 1$, survives for few time-stamps when $\lambda_\mathbf{S} \approx 1$ and continues on otherwise as predicted by Theorem 1.

D. Formal Problem Statements.

Having precisely described the disease-spread model (Section III-B) and derived an epidemic threshold for temporal bipartite graphs for the SIS mode (Section III-C), we are now ready to state our problems formally.

Problem 1: (DYNAMIC BIPARTITE EDGE DELETION) Given a temporal bipartite network $\mathcal{G} = \{G_1, G_2, G_3, \ldots, G_\tau\}$ with $G_t(P, L, E_t) \forall_{t \in [1,\tau]}$ and target fraction $\alpha_E \in (0,1]$, find a smaller graph $\mathcal{G}^* = \{G_1^*, G_2^*, G_3^*, \ldots, G_\tau^*\}$ with $G_t^*(P^*, L, E_t^*) \forall_{t \in [1,\tau]}$ by removing edges from $\mathcal{G}$ such that $|\cup_t E_t^*| \geq (1 - \alpha_E) |\cup_t E_t|$ and $\mathcal{G}^* = \arg\min_{\mathcal{G}'} \lambda_{\mathbf{S}'}$.

The node version of the problem, DYNAMIC BIPARTITE NODE DELETION, is stated in a similar manner. In the problems above, $\mathcal{G}'$ are the temporal graphs obtained by removing edges/edges from $\mathcal{G}$ and $\lambda_{S'}$ is the spectral radius of the system matrix of $\mathcal{G}'$.

IV. OUR APPROACH

Both the edge and node versions of our problem are computationally challenging and in fact special cases of these problems are already NP-Complete [19]. Furthermore, even natural greedy heuristics are costly since they involve repeated eigenvalue computation. Approaches using matrix perturbation theory [18] for quickly estimating the change in $\lambda_\mathbf{S}$ due to an edge/node deletion seem difficult to use since $\mathbf{S}$ is a matrix product. We next present a greedy combinatorial algorithm which avoids eigenvalue computation completely.

We begin by noting that the relation between number of closed walks in a *static* graph and eigenvalues of its adjacency matrix is well known [15], [22]. We exploit this fact to design a a scalable approach with approximation guarantees. However, since our graph is dynamic in nature and it is the eigenvalue

of the system matrix (as opposed to the eigenvalue of the adjacency matrix for static graphs) that we want to reduce, we need to answer the question: "is there a graph we can derive from $\mathcal{G}$ such that the walks on this graph are related to the eigenvalues of the system matrix?". We answer this question first by creating a time-expanded graph $F(V_f, E_f)$ from the bipartite snapshots of $\mathcal{G}$ and then relating the number of closed walks in F to the eigenvalues of the system matrix. The static directed graph F is defined as follows.

Definition 1: **Time-Expanded Bipartite Network.** F is a time-expanded view of $\mathcal{G}$. It consists of $2 \times \tau$ layers of nodes, with odd numbered layers consisting of people nodes and even numbered layers corresponding to location nodes. The directed edges between layers i and $i+1$ for when i is an odd number is given by $\mathbf{B}_{(i+1)/2}$ and is given by $\mathbf{B}_{i/2}^\mathsf{T}$ when i is even. The outgoing edges in the last layer in F connects to the first.

For a special case of the SIS model, with $\delta_p = 1$, which assumes people get cured in one time-stamp, much like the popular Independent Cascade model, we can show that the leading eigenvalue of the system matrix and the number of closed walks of certain lengths in F are closely related. For such a model, the system matrix reduces to $\mathbf{S}_{\mathbf{IC}} = \prod_{t=1}^{\tau} \beta_p \beta_l \mathbf{B}_t \mathbf{B}_{(t \bmod \tau)+1}^\mathsf{T}$. Now, the relation between the leading eigenvalue of $\mathbf{S}_{\mathbf{IC}}$ and the number of closed walk in F is given by the following theorem.

Theorem 2: For any $k = 2\tau c$ for $c \in \mathbb{N}$, we have the following

$$\sum_{w \in W(F)} \text{nodes}(w) = \frac{k}{(\beta_p \cdot \beta_l)^{k/2}} \sum_i^n (\lambda_i(\mathbf{S}_{\mathbf{IC}}))^{k/2\tau}$$

where $W(F)$ is the set of closed paths of length k in F and nodes(w) is the number of unique nodes in walk w.

Theorem 2 states that the number of closed walks in F and the eigenvalues of the system matrix of $\mathcal{G}$ for a special case of the SIS model are closely related. Relying on this fact, we develop a greedy algorithms called TEMPORALEDGECOVER which repeatedly removes an edge participating in most closed walks. The node version of the algorithm TEMPORALNODECOVER does exactly the same except it removes nodes instead of edges. The pseudocode for TEMPORALEDGECOVER is presented in Algorithm 1.

It turns out, we can actually prove that TEMPORALEDGE-COVER is near-optimal. Let E_R be the edges selected by Al-

Algorithm 1 TEMPORALEDGECOVER

Require: $\mathcal{G} = \{G_1, G_2, G_3, \ldots, G_\tau\}$, $0 < \alpha_E < 1$, K
Ensure: Temporal graph $\mathcal{G}'$
 1: Compute the time-expanded graph F and $\mathbf{F}$ using Definition 1
 2: Let $E_U \leftarrow \cup_t E_t$
 3: $E_R \leftarrow \emptyset$
 4: **while** $|E_R| < \alpha_E |E_U|$ **do**
 5:     Compute $\mathbf{X} = \mathbf{F}^{K-1} \circ \mathbf{F}^\mathsf{T}$
 6:     **for** every edge $(i, j) \in E_U \setminus E_R$ **do**
 7:         Calculate W-Score$(i,j) = \sum_t \mathbf{X}(i,j,t)$, where $\mathbf{X}(i,j,t)$ the copy of edge (i,j) between layers t and $t+1$
 8:     $(i,j)^* = \arg\max$ W-Score(i,j)
 9:     Remove all copies of $(i,j)^*$ from F and $\mathbf{F}$
10:     $E_R \leftarrow E_R \cup E_r$
11:     **if** W-Score(i,j) == 0, for all edges **then**
12:         $R \leftarrow \alpha_E |E_U| - |E_R|$
13:         $E_R \leftarrow E_R \cup R$ best edges to remove based on W-Score(i,j) computed in the previous iteration
14: return $\mathcal{G}' \leftarrow \mathcal{G} \setminus E_R$

---

gorithm TEMPORALEDGECOVER to be removed. Removing some edges from G_t converts some of the 1's in $\mathbf{B}_t$ to 0's; let $\mathbf{B}_t \setminus E_R$ denote the bi-adjacency matrix obtained by replacing the 1's in $\mathbf{B}_t$ that correspond to time t edges in E_{TEC} by 0's. Let $\mathbf{S_{SI}} \setminus E_R$ denote the system matrix, but with each $\mathbf{B}_t$ and $\mathbf{B}_{t+1}^\mathsf{T}$ replaced by $\mathbf{B}_t \setminus E_R$ and $\mathbf{B}_{t+1} \setminus E_R$ respectively. Recall F is the directed graph defined in Definition 1. Let n be the shorthand for the number, $\tau |P \cup L|$ of nodes in F.

Lemma 1: Given any constant $\varepsilon > 0$, pick a smallest c such that $k = 2\tau c$ is an integer larger than $\frac{\log n}{\log((1+\varepsilon)(\beta_p \beta_\ell)^{-\tau})}$. Then, we have (a) $\lambda_1(\mathbf{S_{SI}} \setminus E_R) \leq (1+\varepsilon)$, and furthermore (b) $|E_R| = O(|E_{OPT}| \cdot \log^2 n)$ where E_{OPT} is the set of fewest edges in $\mathcal{G}$ such that $\lambda_1(\mathbf{S_{SI}} \setminus E_{TEC}) < 1$.

V. EXPERIMENTS

We describe our experimental setup next. Code (and one of our data along with documentation) is available for academic purposes[1]. Experiments were conducted on Intel(R) Xeon(R) machine with 528GB memory and 4 GPUs (GeForce GTX 1080 Ti).

Datasets: Our datasets consist of high resolution mobility logs collected from various healthcare facilities. UIHC consists of 7 consecutive days of healthcare mobility log collected from the University of Iowa Hospitals and Clinics. From this log, we extract who-goes-where graph between healthcare professionals (HCPs) and locations. There are 5474 HCPs, 4216 unique locations, and a total of 26754 visits in the dataset. Commercial HC is derived from proprietary data capturing mobility in different healthcare facilities. The dataset captures locations of HCPs at various times. There are a total of 44 million entry-exit logs at 20 different healthcare facilities of varying types and sizes. From these, we extract 7-day mobility graphs at 3 different facilities of varying sizes. Commercial HC 1 consists of 995 HCPs, 434 unique locations, and 10833 visits. Similarly, Commercial HC 2 consists of 184 HCPs,

85 unique locations, and 1686 visits. Carilion [10] consists daily of interactions between 3.1 K individuals and locations that took place in Carilion Clinic in Roanoke, VA. It also consists of weekly mobility graphs and has an average of 30K interactions per day.

Baselines: We compare performance of our approaches against several baselines. RANDOM is a naive baseline which removes nodes/edges in a random order. DEGREE deletes nodes (edges) based on the sum of their degrees (sum of the degree of its endpoints) at each time-stamp. TOPK is an extension of existing approaches [19] in static monopartite setting. It computes ranks of each node/edge based on the drop in the value of λ_S caused by its removal. Then it simply deletes the top-k nodes/edges. EIGVECCEN deletes the nodes/edges with the highest eigenvector centrality in $\mathbf{F}$. MONOPARTITE [14] models the interactions as monopartite temporal networks and removes nodes/edges to reduce the epidemic threshold.

A. Performance.

In our first experiment, we compare the performance of our approach against all the baselines for both the edge and node deletion problems as measured by their effectiveness in reducing λ_S. In each data, for each method we delete the edges/nodes chosen by the method to obtain $\mathcal{G}'$ from $\mathcal{G}$. We then compute the following.

$$R_S = \frac{\lambda_S - \lambda_{S'}}{\lambda_S - \lambda_{S^o}} \times 100\%$$

Recall that $\lambda_{S'}$ is the spectral radius of the system matrix of $\mathcal{G}'$. We define λ_{S^o} to be the spectral radius of system matrix of the graph obtained by removing all the edges (or all the nodes). $\lambda_{S^o} = 0$ for node deletion and $\lambda_{S^o} = (1 - \delta)^\tau$ for edge deletion. Note that R_S computes the percentage of drop in the value of λ_S as compared to the maximum possible drop.

We repeat our experiments for various values of α_E and α_N. The results for edge deletion are presented in the top row of Figure 2. As seen in the figure, TEMPORALEDGECOVER (black bar) is the best method in 14 out of 16 setups across all values of α_E in all datasets. Even in the setups where TEMPORALEDGECOVER is not the best ($\alpha_E = 0.1$ for Commercial HC 1 and $\alpha_E = 0.3$ for Carilion), it is very competitive. TEMPORALEDGECOVER's good performance over all the baselines can be attributed to the fact that it is provably near-optimal. Unsurprisingly, RANDOM has one of the worst performances consistently. It is slightly more surprising to see that DEGREE is the second best method ahead of TOPK and EIGVECCEN for the edge deletion problem. We hypothesize that in our dynamic setting the downstream effect of node/edge removal in one time-stamp to another is high. Moreover, because of the centralized nature of healthcare facilities there are natural 'hub' nodes such as nurses' desks, check-in counters, nurse practitioners. DEGREE does well by targeting such high degree nodes. TOPK and EIGVECCEN only quantify centrality once and hence they miss out on the capturing evolving centralities as a result of each subsequent edge removal.
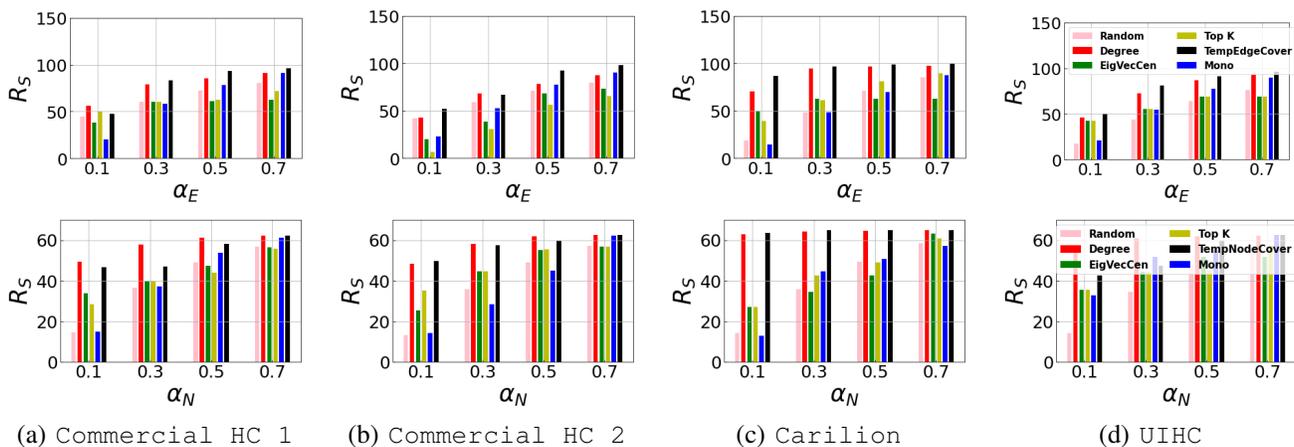
Fig. 2: The percentage drop in the leading eigenvalue of $\lambda_{\mathbf{S}}$ obtained by deleting edges (top row) and deleting nodes (bottom row) for various values of α_E and α_N respectively. Our approach (the last bar) outperforms all the baselines comprehensively. (Best viewed in color)
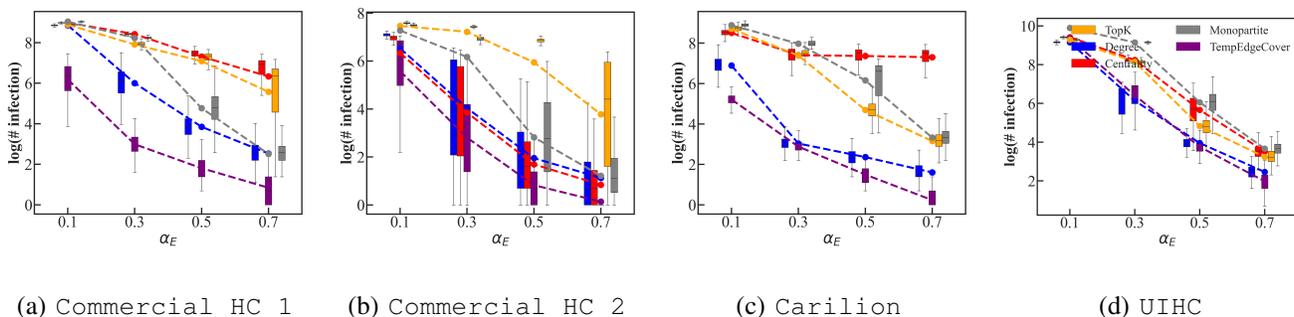

Fig. 3: The log of infection count (averaged over 1000 runs) for the SIS model post edge deletion on all datasets.

Finally, the poor performance of MONOPARTITE highlights the necessity of modeling pathogen spread in hospital as a process over temporal bipartite networks. The results for node deletion are presented in the bottom row of Figure 2. The story is similar. Our approach TEMPORALNODECOVER outperforms baselines in majority of the setting (11 out 16 settings) and remains competitive even when it is not the best.

B. Reduction in Infections for the SIS Model.

Next, we compare performance of our algorithms and the baselines in infection reduction. Here we first run all methods for both edge and node deletion problems to obtain $\mathcal{G}'$. Note that all methods remove equal number of edges (and nodes) from $\mathcal{G}$ to obtain $\mathcal{G}'$. We run 1000 replicates of the SIS simulations on $\mathcal{G}'$ starting from 10 random seeds for each method and keep track of the total number of infected nodes at the end. We then summarize the results using box plots showing variation in the total number of infections. We repeat this for each method and for multiple values of α_E and α_N. The result is summarized in Figure 3.

As observed in the figure, removing edges based on TEM-PORALEDGECOVER leads to fewest infections in $\mathcal{G}'$ in all five of our datasets. We see that the benefit of our approaches over the baselines are maximum for lower values of α_E, which are

easier to implement in real world. For $\alpha_E = 0.7$ we observe that all methods except EIGVECCEN and TOPK lead to near zero infection in Commercial HC 1. This is because these two methods were not able to reduce $\lambda_{\mathbf{S}}$ in Commercial HC 1 (See Figure 2 (a) top row). An interesting observation is that there is a clear correspondence between the infection count presented in Figure 3 and R_S presented in Figure 2. This experiment shows that reducing λ_S does lead to reduction in infection and also serves as an additional validation of the threshold we derived in Theorem 1.

The results show a similar pattern for the node deletion problem (Figure 3 bottom row). TEMPORALNODECOVER is the best methods overall, followed by DEGREE. Note that there are natural 'hub' nodes such as nurses' desks, check-in counters, and nurse practitioners in our datasets which favours DEGREE over other baselines.

VI. RELATED WORKS

Epidemic Thresholds: The papers by Ganesh et al. [6] and Chakrabarti et al. [3] are the earliest results on epidemic thresholds for arbitrary networks. These results assume a static network and apply to the SIS diease-spread model and the latter result is obtained by using the *nonlinear dynamical system (NLDS)* approach. Prakash et al. [13] generalize the

work of Chakrabarti et al. [3] to derive unified thresholds for a variety of disease-spread models. In a direction particularly relevant to this paper, Prakash et al. [14] generalize [3] to derive epidemic thresholds for the SIS model on periodic, time-varying dynamic networks. Sanatkar et al. [16] derive epidemic thresholds for the SIS model on time-varying networks where the adjacency matrix at each time step is chosen at random from a fixed set of matrices. All of these results and more are discussed in the survey of Leitch et al. [11].

Node/Edge Deletion to reduce infections: Tong et al. [20] propose the use of eigen-drop due to the deletion of a set S of k nodes, as a measure of the "shield-value" of S. Torres et al. [21] focused on the nodes that have the largest influence on the leading eigenvalue of a non-backtracking matrix of a graph and proposed two metrics: X-non-backtracking centrality and X-degree centrality, to identify the important nodes. The former metric considers aggregating non-backtracking centrality of each node after specific nodes are removed whereas the latter considers aggregating the degree of each node in prior of specific nodes are removed. Since edge deletion is more appropriate in some cases, Tong et al. [19] propose algorithms for the deletion of k edges (as opposed to nodes) in order to reduce infections. Chen et al. [4] propose algorithms for the same problem that take advantage of the relationship between eigenvector centrality and the leading eigenvalue. This method, however, considers unweighted graphs and works on a static network only. Song et al. [17] extend previous work by considering a setting where there are likely to be many outbreaks, as opposed to a single outbreak.

Epidemic Modelling of Hospital Acquired Infections: Two recent examples of epidemic modeling in healthcare settings appear in [9], [1]. Jang et al. [9] model the spread of MRSA whereas Adhikari et al. [1] model the spread of CDI.

VII. CONCLUSIONS

In this paper, we posed the problem of mitigating infections in healthcare facilities as edge and node deletion problems over temporal bipartite networks. To formulate the problems, we derived epidemic thresholds on temporal bipartite networks for the SIS model. We showed that the epidemic threshold depends on the leading eigenvalue of the system matrix. We proposed a novel combinatorial algorithm TEMPORALEDGECOVER which deletes nodes and edges from the network such that the maximum possible drop on the largest eigenvalue of the system matrix is achieved. We also demonstrated that TEMPORALEDGECOVER is near optimal. Our extensive experiments on mobility networks obtained from various healthcare facilities demonstrate the validity of our epidemic threshold and efficacy of our proposed algorithms.

REFERENCES

[1] B. Adhikari, B. Lewis, A. Vullikanti, J. M. Jiménez, and B. A. Prakash. Fast and near-optimal monitoring for healthcare acquired infection outbreaks. *PLoS computational biology*, 2019.

[2] B. Adhikari, Y. Zhang, A. Bharadwaj, and B. A. Prakash. Condensing temporal networks using propagation. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, pages 417–425. SIAM, 2017.

[3] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos. Epidemic thresholds in real networks. *ACM Trans. Inf. Syst. Secur.*, 2008.

[4] C. Chen, H. Tong, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos. Eigen-optimization on large graphs by edge manipulation. *TKDD*, 2016.

[5] E. R. Dubberke, K. A. Reske, S. Seiler, T. Hink, J. H. Kwon, and C.-A. D. Burnham. Risk factors for acquisition and loss of clostridium difficile colonization in hospitalized patients. *Antimicrobial Agents and Chemotherapy*, 2015.

[6] A. Ganesh, L. Massoulie, and D. Towsley. The effect of network topology on the spread of epidemics. In *IEEE INFOCOM*, 2005.

[7] J. H. Han, N. Sullivan, B. F. Leas, D. A. Pegues, J. L. Kaczmarek, and C. A. Umscheid. Cleaning hospital room surfaces to prevent health care-associated infections: A technical brief. *Annals of internal medicine*, 2015.

[8] D. M. H. Hasan, A. Rohwer, H. Jang, T. Herman, P. M. Polgreen, D. K. Sewell, B. Adhikari, and S. V. Pemmaraju. Modeling and evaluation of clustering patient care into bubbles. In *IEEE ICHI*, 2021.

[9] H. Jang, S. Justice, P. M. Polgreen, A. M. Segre, D. K. Sewell, and S. V. Pemmaraju. Evaluating architectural changes to alter pathogen dynamics in a dialysis unit. In *IEEE ASONAM*, 2019.

[10] J. M. Jiménez, B. Lewis, and S. Eubank. Hospitals as complex social systems: Agent-based simulations of hospital-acquired infections. In *International Conference on Complex Sciences*, pages 165–178. Springer, 2012.

[11] J. Leitch, K. A. Alexander, and S. Sengupta. Toward epidemic thresholds on temporal networks: a review and open questions. *Appl. Netw. Sci.*, 2019.

[12] M.-L. McLaws. The relationship between hand hygiene and health care-associated infection: it's complicated. *Infection and drug resistance*, 2015.

[13] B. A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, and C. Faloutsos. Threshold conditions for arbitrary cascade models on arbitrary networks. In *IEEE ICDM*, 2011.

[14] B. A. Prakash, H. Tong, N. Valler, M. Faloutsos, and C. Faloutsos. Virus propagation on time-varying networks: Theory and immunization algorithms. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2010.

[15] S. Saha, A. Adiga, B. A. Prakash, and A. K. S. Vullikanti. Approximation algorithms for reducing the spectral radius to control epidemic spread. In S. Venkatasubramanian and J. Ye, editors, *SIAM SDM*, 2015.

[16] M. R. Sanatkar, W. N. White, B. Natarajan, C. M. Scoglio, and K. A. Garrett. Epidemic threshold of an SIS model in dynamic switching networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2016.

[17] C. Song, W. Hsu, and M. L. Lee. Node immunization over infectious period. ACM CIKM, 2015.

[18] G. W. Stewart. Matrix perturbation theory. 1990.

[19] H. Tong, B. A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos. Gelling, and melting, large graphs by edge manipulation. In *ACM CIKM*, 2012.

[20] H. Tong, B. A. Prakash, C. Tsourakakis, T. Eliassi-Rad, C. Faloutsos, and D. H. Chau. On the vulnerability of large graphs. In *IEEE ICDM*, 2010.

[21] L. Torres, K. S. Chan, H. Tong, and T. Eliassi-Rad. Node immunization with non-backtracking eigenvalues. *arXiv preprint*, 2020.

[22] C. E. Tsourakakis. Fast counting of triangles in large real networks without counting: Algorithms and laws. In *IEEE ICDM*, 2008.

[23] E. Valdano, C. Poletto, P.-Y. Boëlle, and V. Colizza. Reorganization of nurse scheduling reduces the risk of healthcare associated infections. *Scientific Reports*, 2021.

[24] L. Wen and J. Zhong. Global asymptotic stability and a property of the sis model on bipartite networks. *Nonlinear Analysis: Real World Applications*, 13(2):967–976, 2012.