This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3108075, IEEE Transactions on Network Science and Engineering

Controlling Metastable Infection Patterns in Multilayer Networks via Interlink Design

Srinjoy Chattopadhyay, *Student Member, IEEE,* Huaiyu Dai, *Fellow, IEEE,* and Do Young Eun, *Senior Member, IEEE*

Abstract—Recent research on epidemic spreading in networks has uncovered the phenomena of metastable infection patterns, where epidemics can be sustained in localized regions of activity, in contrast to the classical dichotomy between a quick extinction of infections and a network-wide global infection. Our objective in this work is to leverage this localized infection state to achieve *controlled* spreading in multilayer networks via intelligent design of the interlink structure between the network layers. Following the approach in recent works, the dynamic contact process is approximated by studying the dynamics in local regions around the hubs of the network. This allows us to approximately track the contact process in the near-threshold regime and estimate the mean metastable infection size over the lifetime of the infection. Furthermore, interlinking strategies are devised that can achieve a desired infection size under certain conditions. Theoretically optimal interlink structures can be derived under special cases, whereas greedy strategies are proposed for the general case. We compare the interlinking strategies developed in this work to some popular heuristics and demonstrate their superiority by extensive simulation experiments on both synthetic and real-world networks.

Index Terms—Interlink design, localized epidemics, multilayer networks, susceptible-infected-susceptible model.

1 Introduction

The study of dynamic contact processes in complex networks has emerged as an area of broad and current interest in network science, due to its wide applicability in many disciplines. Traditionally, the research in this area was limited to studying infectious diseases spreading through contact networks. In recent times, the scope has broadened significantly to include diverse topics, like the spread of virus or malware through technological networks [1], adoption of new technologies or products in recommender systems [2], spread of news through social media [3], information spreading through communication networks [4], and the alignment of public emotion and opinion [5]. Here, we use *infection* as the broad term denoting these different types of information (contagious disease, malware, rumor, messages, public opinion, etc.) that propagate through networks. The Susceptible-infected-susceptible (SIS) [6, 7, 8, 9, 10] is one of the simplest and most widely used model for studying infection propagation in networks, where components of the network can belong to either of two infection states: healthy (susceptible) or infected. In this work, our focus is to study the recently uncovered phenomena of localized metastable infection patterns [11, 12, 13] and explore the scope of controlling these patterns in multilayer networks by appropriate design of the inter-layer connections.

1.1 Motivation

The majority of research on infection spreading focuses on the determination of the critical infection rate λ_c [14, 15, 16, 17], indicating a phase transition between the healthy and infected phases. λ_c specifies the infection rate separating abrupt outbreak of contagion in the infected (super-critical) phase, where an initial infection spreads throughout the network and affects a finite fraction of its constituents in the steady state, from the healthy phase (sub-critical), where any initial infection quickly converges to the all-healthy absorbing state. Later works [12, 13] took a closer look into the epidemic dynamics near λ_c and discovered the phenomenon of localized infection states, where localized epidemics can be sustained in the neighborhood of large-degree nodes (hubs). The infection patterns in this near-threshold regime in fact constitute a new phase, where the infection neither propagates globally nor dies out quickly. This infection state is termed "metastable", where the metastability refers to the fact that the infections, while being restricted to a localized region in the network, are maintained in quasi-stable states [18], albeit dying out eventually in the steady state.

This near-threshold behavior of dynamical processes is of great interest, as it is (empirically) observed that many real-world complex systems from diverse fields operate near the phase transition point, like gene expression patterns [19], optimal cell growth [20], dynamics of flock of birds [21], and also the neural connection structure of the human brain [22]. For the scenario considered here, working under the near-threshold conditions gives us the rare capability to induce and control the spread of infections in desired regions of the multilayer network by leveraging these metastable infection states. Note that this is in contrast to the classical understanding of the dichotomy around λ_c, where in the sub-critical regime no long lasting infection patterns are observed, whereas in the super-critical regime the infection diverges throughout the network, precluding any localized control on the infection patterns.

By modeling the infection propagation in finite networks as a discrete-time Markov chain, SIS infections are guaranteed to die out eventually, due to the presence of the unique absorbing state: the all-healthy state. As a result, steady state analysis of infection spreading processes on finite networks

The authors are with North Carolina State University, Raleigh, NC.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3108075, IEEE Transactions on Network Science and Engineering

is not very meaningful, especially in the near-threshold regime due to the existence of the metastable phase. This highlights the importance of studying the transient properties of these dynamical processes on finite-sized networks in the near-threshold regime. These transient properties have been studied in several mathematical works like [11, 23, 24] albeit under different settings. Following these works, our objective is to characterize the mean *metastable infection size*, measuring the average number of infected nodes as the epidemic progresses from the initial infection seeds to eventual dying out in the steady state. Specifically, we show that in the regime near the epidemic threshold λ_c, intelligent design of the interlinking structure can allow us to achieve a desired infection size, subject to certain constraints.

Under the classical notion of the spreading of a contagious disease through a population, the target is to always minimize the spread. For other domains, like characterizing the impact of an advertisement campaign, the goal can be to promote the propagation of infections, which in this case refers to the knowledge of the advertised product. Here, we pose our problem in a more sophisticated setting, where the objective of the network designer is to induce a localized infection pattern of a *desired* size by optimally interlinking the network layers.

1.2 Related Works

Several mathematical works [11, 23, 24], investigating the metastable infection patterns near the infection threshold λ_c for SIS epidemics, have estimated the metastable infection sizes for various theoretical graph topologies, such as scale-free [11, 23] and random geometric graphs [24]. These works are defined in the thermodynamic limit, where the infection rate is close to zero and the network size is infinite. In this setting, first order information about the network topology, i.e. knowledge of the degrees of hub nodes and inter-hub distances in the network, is enough to *exactly* predict the infection patterns. These results are not applicable to our problem as we consider finite-sized networks with arbitrary topologies and additionally pose our problem in a multi-layer setting. Nonetheless, we empirically show that this first order information about the network geometry can be utilized even in finite parameter ranges, as long as the infection spread is localized in nature. Researchers in Physics [25, 26] have also looked at the phenomena of metastable epidemics from the perspective of characterizing the infection threshold λ_c. However, they do not focus on studying the size of the metastable infection patterns. Although the results of these works in Mathematics and Physics cannot be directly applied to our problem, we show that the fundamental principle behind them is still applicable – near λ_c, the global epidemic dynamics can be approximated by studying the local dynamics around hubs and characterizing the interaction among these localized regions of activity.

For multilayer networks, several works [27, 28] have studied epidemic dynamics, focusing on the determination of the epidemic threshold. More recently, the localization phenomenon of SIS epidemics has also been studied on multilayer networks [29, 30]. However, the numerical analysis in [29, 30] involves the eigen-decomposition of the adjacency matrix of the entire network, which becomes intractable when the network is very large and global information about the exact network topology is unavailable. Furthermore, these works focus on multiplex networks, which is a special case of multi-layer networks where the interlink structure trivially connects copies of the same node in the different network layers. Thus, the notion of interlink design is not valid in multiplex networks. Other works [31, 32] in the domain of interlinking design in multilayer networks have studied the impact of interlinking structures on the steady state infections. Our work is different from these in the sense that we focus on localized infection patterns in the near-threshold regime, where although the infections will die out in the steady state, they are maintained in metastable infection states for a long time. Our goal here is to *control* these metastable infection states in multi-layer networks by intelligent design of the interlinks.

1.3 Our contribution

Due to the inherent complexity of studying epidemic dynamics on arbitrary multilayer network topologies, we study the local (microscopic) characteristics of the dynamics around hubs and approximate the global dynamics (macroscopic) by the interaction between the different hubs. Specifically, we prove three characteristics of the epidemic dynamics on the local neighborhood around hubs, which provide the crucial building blocks to design the interlink structure to achieve controlled infection spreading. These three characteristics are: i) quantifying the definition of "hubs" in a multilayer setting as a function of the infection rate, where nodes capable of sustaining long lasting epidemics in their neighborhood are referred to as hubs; ii) approximating the mean metastable infection size in the intra-layer neighborhood of hubs; and iii) characterizing the probability of infection propagation between two interlinked hubs. We leverage this knowledge of the localized dynamics to estimate the average size of the infection patterns for a candidate interlink structure.

Our ultimate objective is to control the metastable infection size in a multi-layer network by appropriate design of the interlink structure. We are interested in developing design principles for constructing the inter-layer link structure, which can induce a metastable infection pattern of a desired size in the network, while utilizing a minimum number of interlinks. The interlink structure is obtained by repeated iterations of a two-step process. In the first step, we obtain the maximum achievable infection size for a small number of available interlinks. Under certain homogeneous assumptions about the hub degrees, we mathematically derive the optimal interlink structure maximizing the infection size for any given number of interlinks. In the general case however, the problem defies a possible analytical solution and we instead develop a greedy approach for interlink design. In the second step, we consider additional interlinks if the desired infection size is not achieved. This is equivalent to obtaining the interlink strategy with the minimum number of interlinks that is capable of sustaining the desired infection size. This is because every additional interlink is optimally utilized to maximize the resulting infection size. We present some approximate bounds on the metastable infection size that is achievable via this approach and also discuss some techniques for speeding up the interlink design algorithm,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3108075, IEEE Transactions on Network Science and Engineering

where instead of starting with a single interlink we can start from a conservative estimate of the number of required interlinks.

Due to the lack of comparable interlink structures for inducing localized metastable infection states in relevant literature, the developed strategies are compared to popular heuristic strategies. We present extensive simulation experiments on several network topologies, including a real-world scenario of multi-layer network design. Our work shows that in the near-threshold regime, global metastable infection patterns can be reliably approximated by studying the local characteristics of the dynamics. Additionally, through extensive simulation experiments, we illustrate the superior performance of our proposed interlink structures to popular multi-layer network design heuristics.

The remainder of this paper is organized as follows. Section 2 specifies the system model and defines the problem tackled in this work. In Section 3, we derive the local properties of the epidemic dynamics and utilize them to estimate the metastable infection size. The interlink structure is derived in Section 4 as a solution to a constrained optimization problem maximizing the estimated infection size. Section 5 combines these components to develop algorithms for engineering the desired metastable infection state. The relevant simulation experiments are presented in Section 6. Finally, we conclude our work in Section 7 and indicate future avenues of research.

2 System Model and Problem Statement

We represent a network layer by $G = (V, E)$, where V and $E \subseteq V \times V$ are the set of nodes and edges, respectively. Given two isolated network layers $G_a = (V_a, E_a)$ and $G_b = (V_b, E_b)$, where $|V_a| = m$ and $|V_b| = n$, our aim is to design the interlink structure $E_{ab} \in \{0,1\}^{m \times n}$ between G_a and G_b, with the objective of controlling the metastable infection size on the multilayer network $\mathcal{G} = (G_a, G_b, E_{ab}, E_{ba} = E_{ab}^T)$. We consider epidemics following the SIS dynamics, where the nodes can be susceptible (S) or infected (I). Infected nodes recover spontaneously at a rate of μ, which can be set to unity without loss of generality. Thus, the unit of time is the average duration of the recovery event. Infections propagate from infected nodes to their susceptible neighbors via edges at rates represented by $\boldsymbol{\lambda} = \{\lambda_a, \lambda_b, \tilde{\lambda}\}$, where the different rates specify the propagation through the intra-layer edges of layer A and B, and the inter-layer edges, respectively. Our objective is to design the interlink structure $E_{ab} \subseteq D_a \times D_b$, so as to sustain a metastable infection state in desired regions of $\mathcal{G}$, represented by the sub-graphs $D_a \subseteq G_a$ and $D_b \subseteq G_b$, as depicted in Fig. 1. We assume that it is feasible to construct an interlink between any two nodes in D_a and D_b. Our goal is to induce a metastable infection pattern of a desired infection density $\Gamma = (\gamma_a, \gamma_b)$, which implies that the metastable infection size in D_a and D_b should exceed $\xi_a^{\text{des}} \triangleq \gamma_a |D_a|$ and $\xi_b^{\text{des}} \triangleq \gamma_b |D_b|$, where $|D_l|$ denotes the number of nodes in the sub-graph D_l of network layer l. Our problem statement can be mathematically stated as:

Problem 2.1. *Given two isolated network layers $\{G_a, G_b\}$ and SIS epidemic dynamics with infection parameters $\boldsymbol{\lambda} = \{\lambda_a, \lambda_b, \tilde{\lambda}\}$, design an interlink structure to induce a metastable infection state of density $\Gamma = (\gamma_a, \gamma_b)$ in the desired subgraphs $D_a \subseteq G_a$ and $D_b \subseteq G_b$.*

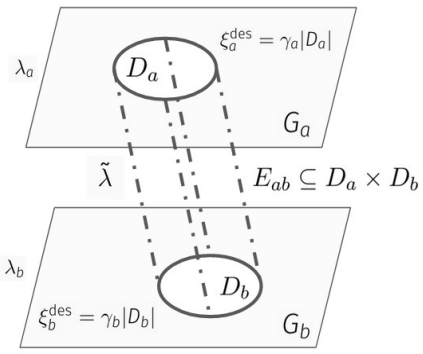


Figure 1: Description of the problem – infection propagation rates in G_a, G_b, and along the inter-layer links are λ_a, λ_b, and $\tilde{\lambda}$, respectively. The objective of the network designer is to design $E_{ab} \subseteq D_a \times D_b$, the interlink structure that induces a metastable infection pattern of size exceeding $(\xi_a^{\text{des}}, \xi_b^{\text{des}})$.

It should not be difficult to see that any Γ cannot be achieved for an arbitrary network topology and arbitrary $\boldsymbol{\lambda}$. In fact, developing an understanding of the feasibility regions for which Problem 2.1 can be solved is a key contribution of this work. The fundamental concept behind our work is the existence of metastable infection patterns in the local neighborhood of hubs [11, 23]. For a specific infection propagation rate ($\boldsymbol{\lambda}$), a node that can sustain an infection pattern in its local neighborhood for a sufficient amount of time is defined as a hub. This definition becomes exact in the thermodynamic limit (as network size approaches ∞ and infection propagation rate approaches 0), where a hub can be defined as a node which can sustain infection patterns in its local neighborhood for a time exponential in the degree of the node, i.e. h is a hub if $\mathbb{E}_{\boldsymbol{\lambda}}(\tau_h) \geq \exp\{c|h|\}$, where τ_h represents the extinction time of an infection in the neighborhood of node h and $|h|$ represents the degree of h [23, 33]. Note that hubs are defined with respect to (w.r.t.) the infection propagation rate $\boldsymbol{\lambda}$. In the finite network setting, such dependence is derived in the following section in Lemma 3.1. This is a crucial idea in this work, where the criteria for selecting hubs becomes more stringent as the infection rate decreases. This is intuitive as infections with lower $\boldsymbol{\lambda}$ are generally more quickly to die out and consequently larger neighborhood sizes are necessary to ensure exponential extinction times (τ_h). The metastable infection patterns around hubs are mathematically tractable in the near-threshold regime, thus allowing *controlled* epidemic spreading. Our key strategy is to design the interlink structure to enforce interaction between different hubs in D_a and D_b, so as to artificially create a composite metastable infection state spanning the local neighborhoods of multiple hubs.

3 Analysis of Epidemic Dynamics

In this section, we develop the mathematical model for tracking the epidemic dynamics. The infection state of a node i in layer l at time t is specified by a Bernoulli random variable $X_i^{(l)}(t)$, where $X_i^{(l)}(t) = 0$ for a healthy

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3108075, IEEE Transactions on Network Science and Engineering

node and $X_i^{(l)}(t) = 1$ for an infected node. We track the epidemic dynamics in $\mathcal{G}$ using the node infection probability $\rho_i^{(l)}(t)$, defined as the probability that node i is infected at time t, i.e. $\rho_i^{(l)}(t) \triangleq \Pr[X_i^{(l)}(t) = 1]$. Let $\boldsymbol{\rho}_t \equiv [\rho_1^{(a)}(t) \cdots \rho_m^{(a)}(t) \rho_1^{(b)}(t) \cdots \rho_n^{(b)}(t)]^T$ be the vectorized infection probability of all nodes in $\mathcal{G}$. The time series $(\boldsymbol{\rho}_t)_{t\geq 0}$, where $\boldsymbol{\rho}_t \in \mathbb{R}^{m+n}$, models the dynamics of the infection spreading process. The initial seeds of infection are represented by $\boldsymbol{\rho}_0$ and the progression of the time series is governed by the SIS model.

Despite the simplicity of the SIS model, a general analytic treatment is not possible except for very special cases [34]. Most works analyze the dynamics approximately by using mean-field approaches; classical examples include the heterogeneous mean field (HMF) model [6] and the quenched mean field (QMF) model [34]. The HMF model is agnostic to both dynamical and topological correlations between the infection states of nodes. QMF presents an improvement over HMF by considering the explicit network topology, albeit neglecting the dynamical correlations. However, it has been shown in [8, 35] that the threshold predicted by QMF is not very accurate for network topologies with high heterogeneity. In this work, we approximate the global epidemic dynamics by studying localized epidemics in the neighborhoods of hubs. These local regions have a highly heterogeneous topology since the degree of the central hub is much larger than others. Due to this reason, we move away from QMF model for tracking the infection propagation. In fact in Section 6, we present results depicting the shortcomings of QMF for our scenario.

3.1 The coarse-grained model

In this work, we employ the modified SIS dynamics valid over coarse-grained time scales introduced in [25] as it can account for dynamical correlation in the infection states of distant nodes. This model characterizes the epidemic dynamics at a larger time-scale encompassing multiple infection and recovery events. This allows us to efficiently capture the correlation in infection states among nearby nodes. This model is particularly useful in studying the epidemic dynamics near the critical threshold and has been utilized in [25, 26] to uncover new results about the threshold in complex networks. Here we extend the coarse-grained (CG) model to multilayer networks, where the governing system of differential equations can be written as:

$$\frac{d\boldsymbol{\rho}(t)}{dt} = -\boldsymbol{\Delta}\boldsymbol{\rho}(t) + (\mathbf{1} - \boldsymbol{\rho}(t)) \circ \mathbf{M}\boldsymbol{\rho}(t), \tag{1}$$

where $\boldsymbol{\Delta} \equiv \text{diag}[\bar{\delta}_a(1) \cdots \bar{\delta}_a(m) \bar{\delta}_b(1) \cdots \bar{\delta}_b(n)]$ represents the diagonal matrix of the effective recovery rates, $\mathbf{M} \in \mathbb{R}^{(m+n)\times(m+n)}$ represents the infection propagation matrix comprising the effective infection rate between all pairs of nodes, and $\circ$ denotes element-wise multiplication or the Hadamard product operator. The elements of $\mathbf{M}$ are:

$$\mathbf{M}_{ij} = \begin{cases} 0 & \text{if } i = j, \\ \bar{\lambda}_{aa}(i,j) & \text{if } i,j \in [1,m], \\ \bar{\lambda}_{ab}(i,j-m) & \text{if } i \in [1,m] \text{ and } j \in [m+1,m+n], \\ \bar{\lambda}_{ab}(j,i-m) & \text{if } i \in [m+1,m+n] \text{ and } j \in [1,m], \\ \bar{\lambda}_{bb}(i-m,j-m) & \text{if } i,j \in [m+1,m+n], \end{cases}$$

where $\bar{\lambda}_{pq}(r,s)$ denotes the effective infection propagation rate between node r in layer p and node s in layer q. (1) is similar in structure to the standard QMF model with the important difference that *effective* rates of recovery ($\bar{\delta}$) and infection ($\bar{\lambda}$) are considered in the CG model that quantifies the impact of the local neighborhood on the infection and the recovery process. Note that the CG model is just an effective choice that enables us to satisfy the objectives of this work: realizing controlled infection spreading in multilayer networks.

Next, we briefly discuss the notion of the effective rates, where we generalize the arguments in [25, 26] to the case of multi-layer networks. The mathematical details of the multi-layer CG model are presented in Appendix A of the supplemental material. The coarse-grained recovery rate $\bar{\delta}(i)$ can be thought of as the effective recovery rate for infection at a node i, accounting for the interaction in the local neighborhood of i. For low infection rates, it can be approximated as:

$$\bar{\delta}_l(i) = \exp(-k_i\lambda_l^2 - \tilde{k}_i\tilde{\lambda}^2), \tag{2}$$

where k_i and $\tilde{k}_i$ denote the intra-layer and inter-layer degree of node i, respectively.

The coarse-grained infection rate $\bar{\lambda}(i,j)$ represents the effective infection propagation rate between distinct nodes i and j. It is defined as:

$$\bar{\lambda}(i,j) = \max_{p\in\mathcal{P}} \bar{\lambda}_p(i,j), \tag{3}$$

where p denotes a particular path between i and j, $\mathcal{P}$ denotes the set of all such paths, and $\bar{\lambda}_p(i,j)$ is defined as the effective infection propagation rate along p. For $i \in D_a$, if p is composed of d_a edges in layer A, $\tilde{d}$ intra-layer edges and d_b edges in layer B, we have:

$$\bar{\lambda}_p(i,j) = \lambda_a \left(\frac{\lambda_a}{\lambda_a+1}\right)^{d_a-1} \left(\frac{\tilde{\lambda}}{\tilde{\lambda}+1}\right)^{\tilde{d}} \left(\frac{\lambda_b}{\lambda_b+1}\right)^{d_b}, \tag{4}$$

where $\frac{\lambda}{\lambda+1}$ denotes the probability of an infection propagation event to occur via a link. Recall that propagation occurs at rate λ while the recovery occurs at rate 1; interested readers can refer to Lemma A.2 in the supplemental material for the details.

Note that the vectorized system equation (1) is a nonlinear differential equation of the first order for which closed form solutions are not possible in the general case. This showcases the main roadblock preventing an exact analysis of global epidemic dynamics in networks. Most practical problems involve large network sizes with arbitrary and partially known topologies, which precludes both rigorous theoretical analysis as well as numerical solutions. Due to this reason, we approximate the global epidemic dynamics by studying the local dynamics in the neighborhood of hubs and characterize the interaction between local regions following [25, 26, 23, 11].

3.2 Approximating global dynamics

We study the global epidemic dynamics by de-coupling the role of the intra-layer and the inter-layer links as shown in Fig. 2. Firstly, we study the epidemic dynamics in the local neighborhood of interlinked hubs (Lemma 3.1). Our target is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3108075, IEEE Transactions on Network Science and Engineering

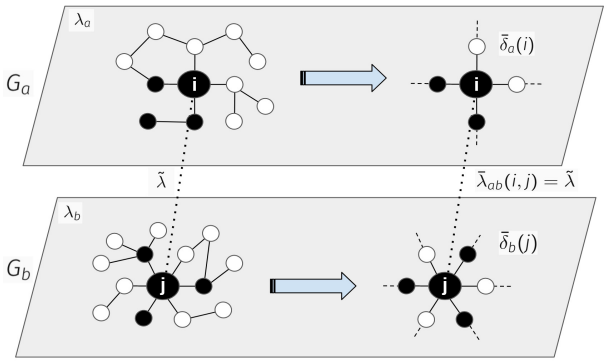


Figure 2: Global dynamics are approximated by studying the epidemics in local regions: the local neighborhoods of interlinked hubs $i \in V_a$ and $j \in V_b$, with $(k_i, k_j) = (4, 6)$. The arbitrary network topology (left) is approximated (exact in the thermodynamic limit) by the extended star topology (right). Infected nodes are colored dark. In the near threshold regime, the (untractable) dynamics in the real network are closely approximated by the (tractable) dynamics in the simplified topology defined w.r.t. the effective propagation rates $\bar{\lambda}_{aa}(i,\cdot)$, $\bar{\lambda}_{ab}(i,j)$ and $\bar{\lambda}_{bb}(j,\cdot)$ and the effective recovery rates $\bar{\delta}_a(i)$ and $\bar{\delta}_b(j)$.

to characterize the conditions under which two interlinked nodes can function as "hubs", where they can sustain a metastable infection state in their local neighborhood. Secondly, we estimate the size of localized infection patterns around each hub by simplifying the intra-layer topology around a hub as an extended star, where all outgoing edges from the hub form line graphs (Lemma 3.2). This simplification makes sense in the low infection rate regime as the infection is unable to propagate beyond a few hops of the hub due to the exponential decrease in rate according to (4). Thirdly, we characterize the re-infection between different hubs through the inter-layer links (Lemma 3.3). Re-infection between hubs is vital for maintaining the metastable infection patterns; when the localized infection patterns in the neighborhood of a particular hub dies out, the hub can be re-infected by an interlinked infected neighbor. In this way, the set of interlinked hubs can sustain long lasting infection patterns of the desired size.

Lemma 3.1. *Consider SIS epidemics with parameters $\boldsymbol{\lambda} = \{\lambda_a, \lambda_b, \tilde{\lambda}\}$ on the multi-layer network with interlinked hubs $i \in V_a$ and $j \in V_b$. A sufficient condition for the existence of a metastable infection state in this subgraph is:*

$$\bar{\delta}_a(i) \cdot \bar{\delta}_b(j) < \tilde{\lambda}^2, \tag{5}$$

where $\bar{\delta}_l(i)$ denotes the effective recovery rate for i in layer l.

Proof. The existence of a metastable infection state is contingent on the instability of the system equations (1), linearized around the absorbing state $\boldsymbol{\rho} = 0$, which can be written as:

$$\frac{d\boldsymbol{\rho}}{dt} = (\mathbf{M} - \boldsymbol{\Delta})\boldsymbol{\rho} = \mathbf{S}\boldsymbol{\rho}, \tag{6}$$

where $\mathbf{S}$ represents the interaction matrix $\mathbf{M} - \boldsymbol{\Delta}$. The instability of this fixed point is guaranteed if $\mathbf{S}$ has a positive spectral radius leading to the above condition; the details of which can be obtained from Lemma B.1 of the supplemental material. ■

The above lemma characterizes the requirements for a node to be classified as a hub. Referring to Fig. 2, the condition (5) obtained in Lemma 3.1 can be understood as follows: two interlinked nodes i and j can be classified as hubs if the product of the (effective) recovery rates at both nodes, given by $\bar{\delta}_a(i) \cdot \bar{\delta}_b(j)$, is lower than the product of the (effective) propagation rates between them, given by $\bar{\lambda}_{ab}(i,j) \cdot \bar{\lambda}_{ba}(j,i) = \tilde{\lambda}^2$. In other words, infection patterns are sustained in the neighborhood of nodes (hubs) if the infection propagates at a higher rate than it recovers. Let us consider the simplification $\lambda_a = \lambda_b = \tilde{\lambda} = \lambda$, for which the effective recovery rate can be approximated as $\bar{\delta}_l(i) \sim \exp(-(k_i + 1)\lambda^2)$, where i has k_i intra-layer and 1 inter-layer link. Plugging this into (5) reveals that a metastable infection state exists when:

$$k_i + k_j > \frac{2}{\lambda^2} \log\left(\frac{1}{\lambda}\right) - 2. \tag{7}$$

The above expression presents a simplified mathematical definition of hubs as a function of the infection propagation rate based on the microscopic (localized) analysis of epidemics. This implies that metastable states can exist even for arbitrarily small infection rates as long as the network contains hubs of sufficiently high degrees. This insight allows us to appreciate the results [11, 23], which state that metastable states exist for any infection rate in infinite sized power law graphs. For the finite sized networks considered in this work, this condition calibrates the definition of "hubs" as a function of the infection rate. Interested readers can refer to Lemma 3.2 in [23] for the corresponding calibration of hubs for the case of scale-free networks in the thermodynamic limit.

Lemma 3.2. *Consider a single-layer SIS dynamic with infection rate λ on an extended star topology, where the hub is permanently infected. The average number of infected nodes around the hub of degree k is approximated by $\lambda k(1 + \lambda)$.*

Proof. The probability of a node at a distance d from the hub to be infected is given by $\lambda(\frac{\lambda}{\lambda+1})^{d-1}$. The number of infected nodes at a distance d from the hub is just k times this probability. Let us represent the length of the paths attached to the hub by l. Thus, the total number of infected nodes (T) in the local neighborhood of the hub can be computed as:

$$T(l) = \sum_{d=1}^{l} \lambda k \left(\frac{\lambda}{1+\lambda}\right)^{d-1} = \lambda k \frac{1-p^l}{1-p}, \tag{8}$$

where $p = \frac{\lambda}{1+\lambda}$. As $l \to \infty$, $T(l) \to \lambda k(1 + \lambda)$. At low infection rates near the epidemic threshold, the error of approximating $T(l)$ by $\lambda k(1+\lambda)$ is very small even for finite values of l. ■

The above result characterizes the extent of infection propagation around an active hub. Note that in reality, the estimated infection size around the hub can vary since the different paths emanating from it can interact with each other. However, under the low infection rate regime, the interaction between different paths is minimal and the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3108075, IEEE Transactions on Network Science and Engineering

infection size estimated by Lemma 3.2 provides a good approximation. We substantiate this argument via simulation experiments in Section 6. Lemma 3.2 states that hubs roughly infect $\lambda k(1+\lambda)$ nodes in their vicinity, which will not be sufficient to satisfy the infection density requirements specified by Γ in Problem 2.1 for most practical problems. Thus, it is clear that although a single interlink directly connecting two hubs is capable of inducing a metastable state, the resulting infection size is quite small. For most practical problems, we need to consider multiple hubs constituting a composite metastable state; each hub infects a fraction of its neighborhood, collectively satisfying the infection density requirements of our problem.

Lemma 3.3. *Consider SIS epidemic dynamics with parameters $\boldsymbol{\lambda} = \{\lambda_a, \lambda_b, \tilde{\lambda}\}$ on the multi-layer network with interlinked hubs $i \in V_a$ and $j \in V_b$ with i infected initially. The probability (η_i) of the infection propagating from hub i to the interlinked neighbor j can be written as:*

$$\eta_i = \frac{\tilde{p}}{\tilde{p} + (1-\tilde{p})(1-p_a^2)^{k_i}}, \tag{9}$$

where $\tilde{p} \equiv 1 - \exp(-\tilde{\lambda})$, $p_a \equiv 1 - \exp(-\lambda_a)$, and k_i represents the intra-layer degree of i.

Proof. An infected node propagates its infection to a fraction of its neighborhood. Thus, it should be easy to see that nodes of high intra-layer degrees are very likely to sustain infection for a long time via multiple re-infection cycles, where upon recovery, hub i is re-infected by one of its infected neighbors. Node j interlinked to i can only be infected via propagation along the interlink. On average, i remain infected for unit time and thus, the probability of infection propagation via the interlink is given by [36] $\tilde{p} = 1 - \exp(-\tilde{\lambda})$. Node i can get multiple opportunities of propagating infection to j, since every time it recovers, i can get re-infected by its k_i intra-layer neighbors. The probability that i is re-infected once can be computed as $X \triangleq 1 - (1-p_a^2)^{k_i}$, where $p_a = 1 - \exp(-\lambda_a)$. This follows from the characterization of $\bar{\delta}(i)$ presented in Lemma A.1 in the supplemental material. The probability that j is not infected at the end of m rounds of re-infection is given by $X^m(1-\tilde{p})^m$. As a result, the total probability that i successfully propagates infection to j is given by:

$$\eta_i = \sum_{m=0}^{\infty} \tilde{p}[1-(1-p_a^2)^{k_i}]^m (1-\tilde{p})^m, \tag{10}$$

which yields the desired expression. ∎

Note that we ignore the presence of additional inter-layer links of i in the computation of the re-infection probability. This is because the inherent assumption is that construction of interlinks is a costly operation and the number of interlinks associated with a hub node is expected to be insignificant in comparison to the intra-layer links. In the presence of additional inter-links, η_i represents a lower bound to the actual probability.

3.3 Estimation of infection size

We represent the set of hubs by $H_a \subseteq D_a$ and $H_b \subseteq D_b$, respectively, where cardinality of the two sets is given by n_a and n_b. Here, we estimate the average infection size resulting for a candidate interlink structure $\sigma \equiv \{\sigma_a(1), \cdots, \sigma_a(n_a), \sigma_b(1), \cdots, \sigma_b(n_b)\}$, where $\sigma_a(i)$ (or $\sigma_b(i)$) represents the set of hubs in layer B (or A) interlinked to the ith hub in layer A (or B). We focus on the intra-layer and inter-layer neighborhood of a particular hub i in network layer A. When the infection in the local intra-layer neighborhood of i dies out, hub i can be re-infected by any of its inter-layer neighbors, represented by the set $\sigma_a(i)$. The probability that a particular hub $j \in \sigma_a(i)$ propagates its infection to i is given by η_j, according to Lemma 3.3. Thus, the total infection probability of i is given by: $1 - \prod_{j\in\sigma(i)}(1-\eta_j)$. This expression can be thought of as the lower bound for the infection probability as only the contribution of hubs directly interlinked to i is considered here. In most practical cases, the effect of additional hubs is marginal since $\bar{\lambda}_{ij}$ decreases exponentially with the distance between i and j. Combining this with Lemma 3.2, the average infection size around i is given by:

$$f(\sigma_a(i)) = \lambda_a k_i (1+\lambda_a)\Big[1 - \prod_{j\in\sigma_a(i)} (1-\eta_j)\Big], \tag{11}$$

where $f(\sigma_a(i))$ denotes the infection size around i. We remind that our focus is on the low infection rate regime, where the epidemics are localized around nodes of high degrees. Consequentially, the total infection size around the different hubs can be approximated by the sum of the individual infection sizes. Note that this approximation is exact [11, 23] in the thermodynamic limit. Although we focus on finite sized networks in this work, we verify in Section 6 that the estimates of infection sizes remain largely consistent with the simulations, for extended ranges of parameters. For an interlink structure σ, the average infection size around the hubs, denoted by $f(\sigma)$, is given by:

$$f(\sigma) = \sum_{i\in H_a} \lambda_a k_i (1+\lambda_a)\Big[1 - \prod_{j\in\sigma_a(i)} q_j\Big] + \sum_{i\in H_b} \lambda_b k_i (1+\lambda_b)\Big[1 - \prod_{j\in\sigma_b(i)} q_j\Big], \tag{12}$$

where the summations index over the hub nodes in the network layers, and $q_j \triangleq (1-\eta_j)$.

4 Optimizing the Interlink Structure

The fundamental purpose of the interlink structure is to maintain interaction between different hubs so that if the infection in the local neighborhood of a hub dies out, it can get re-infected by other inter-linked hubs. Thus, it is easy to see that the inter-layer links should form a connected graph so that all hubs can re-infect each other. Since the minimally connected graph spanning n vertices is a tree, the minimum number of interlinks required to maintain connectivity between $n_a + n_b$ hubs is $n_a + n_b - 1$, where n_a and n_b represent the cardinalities of the set of hubs H_a and H_b, respectively. The choice of $H_a \subset D_a$ and $H_b \subset D_b$ is discussed in the next section. In this section, we impose this minimum constraint on the interlink structure σ and obtain $\sigma^\star$ maximizing the average infection size $f(\sigma)$ given by (12). The maximization of the infection size guarantees

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3108075, IEEE Transactions on Network Science and Engineering

optimal utilization of the interlinks. The optimal interlink structure $\sigma^\star$ is obtained by solving:

$$\max_{\sigma} \quad f(\sigma) \tag{13a}$$

$$\text{subject to} \quad \sum_{i\in H_a}|\sigma_a(i)| = \sum_{i\in H_b}|\sigma_b(i)| = n_a + n_b - 1, \tag{13b}$$

$$\bigcup_{i\in H_a}\sigma_a(i) = H_b,\ \bigcup_{i\in H_b}\sigma_b(i) = H_a. \tag{13c}$$

The first constraint enforces that the minimum number of interlinks are used, while the second one ensures that all hubs in H_a and H_b are supported via interlinks. In the general case, (13) is a discrete set optimization problem which is computationally prohibitive for solving for large number of hubs. However, under the special case of homogeneous hub degrees, the problem becomes mathematically tractable and we can obtain the theoretically optimal $\sigma^\star$. Although real-world problems are unlikely to exhibit identical hub degrees, we show that the properties of this special case solution can be leveraged to develop interlink design strategies in the general case.

4.1 Interlink structure for identical hub degrees

Under the assumption of identical hub degrees, the interlink structure σ can be simplified to the allocation of interlinks $\boldsymbol{b} \equiv [b_1 \cdots b_{n_a} b_{n_a+1} \cdots b_{n_a+n_b}]$, where b_i specifies the number of interlinks allocated to hub i. The first n_a indices of $\boldsymbol{b}$ define the interlink allocation for H_a, while the last n_b indices define the allocations for H_b. This simplification results from the fact that the identical hub degree assumption obfuscates the identity of the nodes in H_a and H_b. As a result, the interlink structure σ is completely specified by the number of interlinks allocated to each hub. This simplification does not hold in the general case, where hubs in H_a and H_b have distinct degrees. However, we show that a similar interlink structure continues to be optimal even in the general case except for some special scenarios. Thus, although the identical hub assumption does not hold true in practice, the resulting interlink structure has practical importance. For identical hub degrees, the optimization problem (13) is simplified as:

$$\max_{\boldsymbol{b}} \quad \sum_{i\in H_a}\xi_a\left(1-{q_a}^{b_i}\right) + \sum_{i\in H_b}\xi_b\left(1-q_b^{b_i}\right) \tag{14a}$$

$$\text{subject to} \quad \sum_{i\in H_a} b_i = \sum_{i\in H_b} b_i = n_a + n_b - 1, \tag{14b}$$

$$b_i \geq 1,\ \ \forall i \in H_a \cup H_b, \tag{14c}$$

where $\xi_a \triangleq \lambda_a k_a(1+\lambda_a)$ and $\xi_b \triangleq \lambda_b k_b(1+\lambda_b)$ are defined as the expected infection size around active hubs in both layers, as specified by Lemma 3.2; q_a (or q_b) denotes the probability of infection not propagating from a hub in H_a (or H_b) to an inter-layer neighbor in H_b (or H_a); and $b_i \triangleq |\sigma(i)|$ denotes the number of inter-layer neighbors of hub i.

Let us separate the interlink allocation vector into two components $\boldsymbol{b}_{H_a}$ and $\boldsymbol{b}_{H_b}$, representing interlink allocation strategies for the hubs in H_a and H_b, respectively. It can be observed that the maximization of the objective $f(\boldsymbol{b})$ over the space of interlink allocation $\boldsymbol{b}$ is equivalent to independently maximizing the two summations in (14a) over the layer-wise interlink allocation structures $\boldsymbol{b}_{H_a}$ and $\boldsymbol{b}_{H_b}$. This is because the two summations in the objective function in (14) are only dependent on the layer-wise interlink allocations, i.e. variation in $\boldsymbol{b}_{H_a}$ does not affect the second term in the objective function (14a) and vice versa. Note that the independence of the two summations only holds true for the identical hub degree case. We use the following result in determining the optimal interlink structure maximizing the objective in (14), whose proof is presented in the supplemental material in Lemma B.2.

Lemma 4.1. *The function $f : \mathbb{R}^d \to \mathbb{R}$ defined by $f(\boldsymbol{x}) = \sum_i c_1(1-c_2^{x_i})$ is Schur-concave for $c_1 \geq 0$ and $0 \geq c_2 \geq 1$, where $\boldsymbol{x} = [x_1 \cdots x_d]^T$.*

Definition 4.2. *An interlink allocation strategy $\boldsymbol{b} \equiv [b_1 \cdots b_{n_a} b_{n_a+1} \cdots b_{n_a+n_b}]$ is defined to be maximally uniform, if for both $\boldsymbol{b}_{H_a} \equiv [b_1 \cdots b_{n_a}]$ and $\boldsymbol{b}_{H_b} \equiv [b_{n_a+1} \cdots b_{n_a+n_b}]$, the variance of the number of interlinks allocated to each hub in H_a and H_b is minimum, i.e.,*

$$\max_{i\neq j} |b_i - b_j| \leq 1,\ \forall i,j \in H_a (or\ H_b). \tag{15}$$

With the Schur-concavity of f in Lemma 4.1 and the definition of maximally uniform interlink allocation in hand, we can state the result for the optimal interlink structure.

Theorem 4.3. *Under a homogeneous assumption on the hub degrees, the maximally uniform interlink allocation strategy maximizes the expected infection size in the network.*

Proof. As discussed earlier, under the uniform degree assumption of the hub nodes in H_a and H_b, the maximization of the objective function in the optimization problem (14) is equivalent to two independent optimization problems, for each layer of the network, which for layer A can be written as:

$$\max_{\boldsymbol{b}} \quad \sum_{i\in H_a}\xi_a\left(1-{q_a}^{b_i}\right) \tag{16a}$$

$$\text{subject to} \quad \sum_{i\in H_a} b_i = n_a + n_b - 1, \tag{16b}$$

$$b_i \geq 1,\ \ \forall i \in H_a. \tag{16c}$$

Recall that $q_a \triangleq 1-\eta_a$, where η_a is defined in Lemma 3.3 as the probability of infection propagating from a hub node in layer A to an interlinked hub in layer B. ξ_a is defined as the average infection size in the local intra-layer neighborhood of an active hub in layer A and it can be expressed as $\xi_a = \lambda_a k_i(1+\lambda_a)$, where k_a is the intra-layer degree of the hubs in layer A and λ_a is the intra-layer infection propagation rate in layer A. It is easy to verify from Lemma 4.1 that (16) is Schur-concave. As a consequence, for all interlink allocation strategies $\boldsymbol{b}_x, \boldsymbol{b}_y \in \mathbb{R}^{n_a}$, where $\boldsymbol{b}_x$ is majorized by $\boldsymbol{b}_y$, the value of the objective function corresponding to $\boldsymbol{b}_x$ will be greater. As the interlink allocation strategy becomes more uniform, the value of the objective function progressively increases.

We are interested in obtaining the allocation strategy for $(n_a + n_b - 1)$ interlinks to n_a hubs in layer A. Certain values of (n_a, n_b), say $(3, 5)$, preclude the existence of a uniform solution. However, as an outcome of the Schur-concavity of the objective function, we can conclude that the maximally uniform allocation strategy maximizes the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3108075, IEEE Transactions on Network Science and Engineering

objective function. The same arguments can be made for the allocation of $(n_a + n_b - 1)$ interlinks to the n_b hubs in layer B, thus proving the theorem. ■

4.2 Interlink structure for the general case

In the general case with non-identical hub degrees, the Schur-concavity of the objective function is lost and a theoretically optimal interlink structure cannot be obtained. However, we present a conjecture about the interlink allocation strategy which holds in most realistic cases. In particular, we posit that the uniform allocation of interlinks continues to be a good strategy even for the general case. Thereafter, we propose a greedy strategy for the construction of the inter-layer links between H_a and H_b. Although both the interlink allocation strategy and the greedy design of the interlinks are not theoretically guaranteed to be optimal, the superiority of these designs in comparison with the randomized and heuristic strategies is clearly showcased via extensive simulation experiments in Section 6.

Conjecture 4.4. *Consider an interlink allocation strategy, where two arbitrary hubs y and z in the same layer (say layer A) of the network are interlinked to the set of hubs $\sigma_a(y)$ and $\sigma_a(z)$, respectively, where $b_y \triangleq |\sigma_a(y)|$ and $b_z \triangleq |\sigma_a(z)|$. If $b_y > b_z + 1$, re-allocating an interlink from $\sigma_a(y)$ to $\sigma_a(z)$, to make the interlink allocation more uniform, leads to an improvement in the objective function $f(\sigma)$ in (13) with high probability.*

We focus on the limiting case, where $b_y = b_z + 2$. For a larger difference in the allocation of the interlinks, the arguments can be applied repeatedly. Let the sets of nodes interlinked to y and z before the exchange be given by $\sigma_a(y) = \{h_1, \cdots, h_{b_y}\}$ and $\sigma_a(z) = \{h_{b_y+1}, \cdots, h_{b_y+b_z}\}$. We are interested in examining the change in the value of the objective function when the interlink allocation is made more uniform, i.e. when any hub from $\sigma_a(y)$ is moved to $\sigma_a(z)$. Let $f(\sigma_a(y), \sigma_a(z))$ and $f(\sigma'_a(y), \sigma'_a(z))$ denote the average infection sizes around the two hubs y and z before and after the exchange of the interlinks. Thus, we can write:

$$f(\sigma_{a(y)}, \sigma_{a(z)}) = \xi_y \left(1 - \prod_{j=1}^{b_y} q_j\right) + \xi_z \left(1 - \prod_{j=b_y+1}^{b_y+b_z} q_j\right), \quad (17)$$

$$f(\sigma'_{a(y)}, \sigma'_{a(z)}) = \xi_y \left(1 - \prod_{j=1}^{b_y-1} q_j\right) + \xi_z \left(1 - \prod_{j=b_y}^{b_y+b_z} q_j\right), \quad (18)$$

where we assume that the hub h_{b_y} (arbitrary choice) is interlinked to z after the exchange. The difference between the infection sizes after the exchange, $\Delta f \triangleq f(\sigma') - f(\sigma)$ can be written as:

$$\Delta f = (1 - q_{b_y}) \left(\xi_z \prod_{j=b_y+1}^{b_y+b_z} q_j - \xi_y \prod_{j=1}^{b_y-1} q_j\right). \quad (19)$$

The two products in the second parenthesis contain b_z and $b_z + 1$ elements, respectively. Recall that q_j defines the probability that an infection from hub h_j does not propagate to an interlinked hub. For nodes of high intra-layer degrees (hubs), $q_j \equiv 1 - \eta_j$ is expected to be a small value near 0, as η_j is close to 1 according to Lemma 3.3. As a result, the product of b_z terms is significantly higher than the product of $b_z + 1$ terms. In comparison, the variation in the $\xi \triangleq \lambda k(1 + \lambda)$ values for different hubs is expected to be much lower since only nodes of high intra-layer degrees qualify as hubs. In most realistic cases, the hub degrees do not vary by orders of magnitude. Thus, Δf in (19) is positive in most cases, satisfying our conjecture. However, it is possible to design counter-examples with one super-hub, i.e. $k_y \gg k_z$, where Δf can be negative due to the huge discrepancy between ξ_z and ξ_y. Since such cases will rarely exist in practice, our conjecture is expected to hold in most practical problems.

The above discussion pertains to the case of interlink allocation among the two sets of hubs. It is important to note that under the general case of non-identical intra-layer degrees of the hubs in H_a and H_b, the interlink allocation strategy cannot solely determine the structure. Unlike the previous case, here we need to devise an interlinking pattern in addition to the allocation strategy. Thus, we are interested in the problem of optimizing the interlink structure given the maximally uniform interlink allocation, which was shown to be a good strategy in Conjecture 4.4. Unfortunately, even for this case a mathematically rigorous optimal structure cannot be obtained. Interestingly, for a maximally uniform allocation we can rigorously obtain the interlink structure *minimizing* the objective function in (13) but a strategy for maximizing the objective cannot be obtained. This stems from the peculiar result arising from the Generalized Rearrangement Inequality, which we developed in a previous work [37] and present as Lemma B.3 in the supplemental material. As a consequence, we can show that the monotonic arrangement of interlinks, where the hubs in H_a and H_b are interlinked in the order of their intra-layer degrees, is the worst; however, the arrangement that maximizes the average infection size, remains open. In this work, we propose the construction of the interlinks based on the anti-monotonic arrangement, wherein the hubs with the highest intra-layer degree in layer A are coupled to the hubs with the lowest intra-layer degree in layer B. Although this strategy is sub-optimal, simulation results show that it outperforms other randomized strategies. Interestingly, our simulation experiments reveal that if the uniform interlink allocation is followed, variation of the interlink structure does not have a major impact on the infection size. Thus, we can infer that for the general case of non-identical hub degrees, the interlink allocation strategy plays a more important role than the specific arrangement and the uniform allocation continues to be a good strategy.

5 Engineering the Metastable State

In this section, we describe the methodology for constructing the composite metastable infection state, $H_a \subset D_a$ and $H_b \subset D_b$, which can satisfy the infection density requirements of Problem 2.1.

5.1 The core and supplementary groups

We start our discussion by defining two groups of nodes: the *core* group and the *supplementary* group. The core and the supplementary groups define the set of nodes in D_a and D_b which can support a metastable infection state by using a single interlink. A core group is defined as a set of nodes $\chi_a^0 \cup \chi_b^0$, where an interlink between any $a \in \chi_a^0$ and any $b \in \chi_b^0$ satisfies Lemma 3.1 and can sustain a metastable

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3108075, IEEE Transactions on Network Science and Engineering

infection state in the neighborhood of its endpoints. On the other hand, the supplementary group, represented by $\chi_a^1 \cup \chi_b^1$, is defined as the set of nodes in D_a and D_b which need the support of specific nodes from the core group. As a consequence, any interlink between χ_a^0 and χ_b^0 can induce metastable infection states, whereas hubs in χ_a^1 (or χ_b^1) can sustain infections only by interlinking to specific hubs in χ_b^0 (or χ_a^0). The core group provides us with increased flexibility in designing the interlink structure since by definition, any interlink between the constituents is valid, whereas the supplementary group puts additional constraints on the interlink design.

Let us represent the nodes with the highest intra-layer degree in D_a and D_b by a_1 and b_1, respectively. Let ϕ_a (or ϕ_b) denote the set of nodes in D_a (or D_b), which satisfy the constraints in Lemma 3.1 when interlinked to $b_1 \in D_b$ (or $a_1 \in D_a$). Mathematically, this can be written as:

$$\phi_a = \{v \in D_a | \bar{\delta}_a(v) \cdot \bar{\delta}_b(b_1) < \tilde{\lambda}^2\}, \tag{20}$$

$$\phi_b = \{v \in D_b | \bar{\delta}_b(v) \cdot \bar{\delta}_a(a_1) < \tilde{\lambda}^2\}. \tag{21}$$

where $\bar{\delta}_l(i)$ denotes the effective recovery rate for node i in layer l. Since ϕ_a and ϕ_b are defined w.r.t. the highest degree nodes in the other layer, these two sets constitute all nodes in D_a and D_b which can possibly sustain a metastable infection state in their neighborhood by using a single interlink. Recall that the effective recovery rate $\bar{\delta}(i)$ decreases with the hub degree following (2). Thus, for a low enough infection parameter $\boldsymbol{\lambda} = (\lambda_a, \lambda_b, \tilde{\lambda})$, the sets ϕ_a and ϕ_b can be empty. ϕ_a and ϕ_b can also be empty if the desired subgraphs D_a and D_b comprises nodes of low intra-layer degrees. The emptiness of the ϕ's imply that the specifications in Problem 2.1 are not sufficient to sustain metastable infection states. However, in the supplementary material we discuss some techniques for augmenting the metastable infection state beyond the core and the supplementary hubs. Adopting those approaches can induce metastable infection states even for empty ϕ's in some cases albeit using a very large number of interlinks. In the following discussion, we focus on scenarios with non-empty ϕ_a and ϕ_b, which constitutes the feasible region for solving Problem 2.1. The core hub group can be defined as:

$$\chi_a^0 = \{v \in D_a | \bar{\delta}_a(v) < \tilde{\lambda}\}, \tag{22}$$

$$\chi_b^0 = \{v \in D_b | \bar{\delta}_b(v) < \tilde{\lambda}\}. \tag{23}$$

It can be easily verified that an interlink between any two nodes, $i \in \chi_a^0$ and $j \in \chi_b^0$, satisfies Lemma 3.1. Additionally, ϕ_a and ϕ_b defined in (20)-(21) comprise the set of hubs which can possibly sustain a metastable infection state by interlinking to a hub in the other layer. Thus the set of nodes defined by $\chi_a^1 = \phi_a - \chi_a^0$ and $\chi_b^1 = \phi_b - \chi_b^0$ comprise hubs which can satisfy the constraints in Lemma 3.1 for specific inter-layer connections. We refer to these set of nodes, i.e. χ_a^1 and χ_b^1 as the supplementary group of hubs. The hubs in the core group can induce a metastable infection state by interlinking to any hub in the core group in the other layer, whereas the hubs in the supplementary group can do so by interlinking to specific hubs in the core group. For each node $i \in \chi_a^1$, we can define the candidate set:

$$\kappa_a(i) = \{j \in \chi_b^0 | \bar{\delta}_a(i)\bar{\delta}_b(j) < \tilde{\lambda}^2\}. \tag{24}$$

The candidate set for $i \in \chi_b^1$ can be obtained similarly. By Lemma 3.1, we know that any interlink between i and an element of the candidate set is able to sustain a metastable infection state. The candidate set for the core group is nothing but the entire core group in the other layer. The construction of the interlink structures among these two group of hubs follows the two principles put forward in the previous section: i) uniform allocation of interlinks, and ii) anti-monotonic pattern of interlinks. For brevity, the details of the algorithms are presented in the Appendix D of the supplemental material.

5.2 Constructing the hub sets

The scheme of achieving controlled infection spreading in the network is to incrementally assemble a composite metastable infection state (H_a, H_b) in the desired region of the network, comprising multiple interlinked hubs. Each hub is capable of sustaining a metastable infection in its local neighborhood. The interlink structure between these hubs maintains a reliable interaction between them so that a long lasting infection pattern is induced around these hubs.

In order to construct H_a and H_b, we formulate a subroutine called GENOPTINTER that constructs the interlink structure between two given sets of hubs, the details of which is presented in Appendix D of the supplemental material. It is important to note that the problem of estimating the number of hubs cannot be decoupled from the problem of designing of the interlink structure. Due to this reason, one effective way to ensure the use of a minimum number of interlinks is to sequentially grow the metastable infection state by adding hubs to H_a and H_b, until the infection density requirements are achieved. Following this principle, the hub selection and interlinking strategy is presented in

Algorithm 1 Constructing the metastable infection state

procedure CONSTRUCTMETA($\xi^a_{\text{des}}, \xi^b_{\text{des}}, \boldsymbol{\lambda}$)
  Compute $\chi_a^0, \chi_a^1, \chi_b^0, \chi_b^1$ for $\boldsymbol{\lambda}$
  $H_a^\infty \leftarrow \{\chi_a^0, \chi_a^1\}$ indexed by degrees.
  $H_b^\infty \leftarrow \{\chi_b^0, \chi_b^1\}$ indexed by degrees.
  $(i_a, i_b) \leftarrow (1, 1)$.
  $H_a \leftarrow \{\chi_a^0[i_a]\}$, $H_b \leftarrow \{\chi_b^0[i_b]\}$.
  $\sigma \leftarrow$GENOPTINTER$(H_a, H_b, \boldsymbol{\lambda})$
  $(\xi^a_{\text{curr}}, \xi^b_{\text{curr}}) = f(\sigma)$ computed using (12).
  while $\xi^a_{\text{curr}} < \xi^a_{\text{des}}$ or $\xi^b_{\text{curr}} < \xi^b_{\text{des}}$ **do**
    if $\xi^a_{\text{curr}} < \xi^a_{\text{des}}$ **then**
      if $H_a == H_a^\infty$ **then**
        return False
      $H_a = H_a \cup H_a^\infty[i_a]$
      $\sigma =$ GETOPTINTER(H_a, H_b)
      $\xi^a_{\text{curr}} = f_a(\sigma)$ computed using (12)
      $i_a = i_a + 1$
    if $\xi^b_{\text{curr}} < \xi^b_{\text{des}}$ **then**
      if $H_b == H_b^\infty$ **then**
        return False
      $H_b = H_b \cup H_b^\infty[i_b]$
      $\sigma =$ GETOPTINTER(H_a, H_b)
      $\xi^b_{\text{curr}} = f_b(\sigma)$ computed using (12)
      $i_b = i_b + 1$
  return σ

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3108075, IEEE Transactions on Network Science and Engineering

Algorithm 1. The input to Algorithm 1 is the topology of the desired regions D_a and D_b, the desired infection sizes $\xi^a_{\text{des}} \triangleq \gamma_a|D_a|$ and $\xi^b_{\text{des}} \triangleq \gamma_b|D_b|$, and the infection propagation parameters $\boldsymbol{\lambda}$. The algorithm first computes the core and the supplementary node groups and then constructs the ordered sets $H^\infty_a \subset D_a$ and $H^\infty_b \subset D_b$, comprising the core and supplementary hubs arranged in the decreasing order of the intra-layer degrees. H^∞_a and H^∞_b describe the order of the nodes from the desired region which are used to construct the required composite metastable infection state. The algorithm proceeds by incrementally allocating new hubs to H_a and H_b from H^∞_a and H^∞_b, respectively, until either the desired infection sizes is achieved or all candidate hubs in H^∞_a or H^∞_b are used up. In the former case, the algorithm outputs the resulting interlink structure σ. The latter case falls in the domain of infeasible region, where the problem settings $(D_a, D_b, \boldsymbol{\lambda})$ are inadequate to induce the desired infection size. The intermediate infection sizes corresponding to the current choices of H_a and H_b are denoted by ξ^a_{curr} and ξ^b_{curr} in Algorithm 1; these are computed following (12).

Thus, we can see that Algorithm 1 essentially achieves the controlled infection spreading parameterized by γ_a and γ_b in the desired region D_a and D_b; it returns False when the problem settings are infeasible and returns the required interlink structure if the desired infection sizes can be achieved via the core and supplementary group of hubs. Note that the control of the infection density is discrete in nature. As additional nodes are interlinked to the metastable infection state, a discrete increment in the infection size occurs. Furthermore, the optimal interlink structure is recomputed every time additional hubs are included. Interestingly, simulation experiments in Section 6 reveal that the dependence of the infection size on the interlink structure is not pronounced as long as the interlink allocation is maximally uniform. For future extensions to this work, we are exploring the possibility of utilizing this idea to optimize the algorithm so as to update the interlink structure only in intermediate steps. Such optimizations are beyond the scope of the current work.

5.3 Limits of achievable infection densities

We end this section by discussing the limits of the achievable infection densities in the desired subgraphs $D_a \subseteq G_a$ and $D_b \subseteq G_b$. It should be intuitive that the entire range $0 \leq \gamma_a, \gamma_b \leq 1$ is not achievable for any infection parameter $\boldsymbol{\lambda}$. The exact computation of the infection size achievable by these two groups of hubs is given by (12), which is dependent on the interlink structure σ. To avoid the dependence on σ, a simple upper bound can be obtained by neglecting the q_j terms:

$$\gamma^{\text{cs}}_a = \lambda_a(1+\lambda_a) \sum_{i \in \chi^0_a \cup \chi^1_a} k_i, \tag{25}$$

$$\gamma^{\text{cs}}_b = \lambda_b(1+\lambda_b) \sum_{i \in \chi^0_b \cup \chi^1_b} k_i. \tag{26}$$

The above thresholds serve as rough upper bounds and are not achievable in practice owing to the non-zero q_j terms. However, in practice q_j terms are very small for hubs with high degree and thus $\Gamma^{\text{cs}} = (\gamma^{\text{cs}}_a, \gamma^{\text{cs}}_b)$ provides a crude and easily to compute upper bound of the achievable infection size. Although we discuss techniques for augmenting the metastable infection state in Appendix C in the supplemental material, these techniques are not very reliable in pushing the achievable infection density far beyond Γ^{cs}. Thus, for cases where the desired infection density (Γ) is much higher than Γ^{cs}, a sensible strategy is to strive for increasing $\boldsymbol{\lambda} = (\lambda_a, \lambda_b, \tilde{\lambda})$ so as to include more nodes in ϕ_a and ϕ_b, which can lead to an increment in the achievable infection density Γ^{cs}.

6 Simulation Results

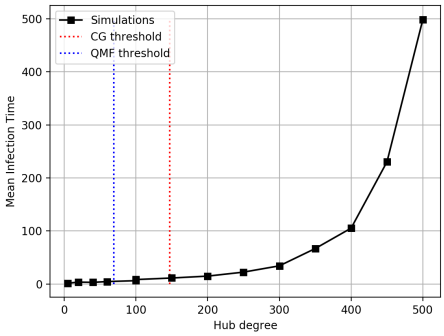
In this section, we discuss the simulation experiments validating our theoretical results on both synthetic and real-world networks. We simulate the SIS epidemics on networks following the popularly used Gillespie algorithm [38], the details of which are presented in the supplemental material in Appendix E. We use Python and its associated libraries (NetworkX, NumPy) to build our simulation framework. Due to the stochastic nature of the epidemics, we plot the averaged results over multiple independent Monte Carlo (MC) runs. We start this section by verifying the theoretical results presented in the Lemmas by testing them on complex network topologies. Thereafter, we compare our interlink design strategy with other heuristics to observe the variation in the resulting metastable infection size. Finally, we apply the principles developed in this work to a real world network design problem.

6.1 Verification of the Lemmas

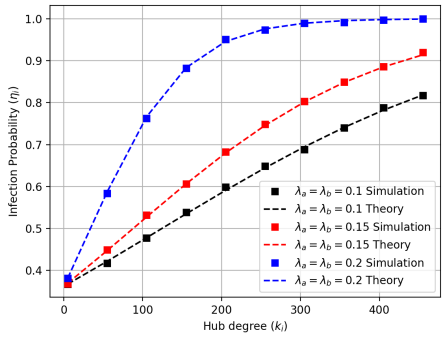
For the verification of Lemma 3.1, we simulate the infection times of epidemics on multilayer networks with a single interlink, as shown in the left subfigures in Fig. 2, where two hubs (i, j) with intra-layer degrees (k_i, k_j) are coupled to each other. The infection time is defined as the time duration between the initial infection configuration with $\{i, j\}$ infected and the final all-healthy state, averaged over 500 MC runs. The infection recovery rate is taken to be 1 and thus, the time in our simulations can be understood as the number of rounds of an average infection-recovery event. We generate G_a and G_b as Barabasi-Albert (BA) graphs with parameter $m = 1$ of size $n = 5000$. A node of degree 100 from G_b is taken to be the fixed end-point j. For different choices of i from G_a, the infection time of the epidemics with propagation rates $\boldsymbol{\lambda} = \{0.1, 0.1, 0.1\}$ is plotted w.r.t. the hub degree k_i in Fig. 3a. The vertical lines in Figure 3a represent the thresholds predicted by the coarse-grained (CG) model [25] and the quenched mean field (QMF) model [34]. We observe that the infection time increases exponentially when the hub degree (x axis) goes beyond the CG threshold, computed using (5). This verifies Lemma 3.1, which derives the condition for the existence of long lasting (varying exponentially with hub degree) infection patterns in the neighborhood of an interlink. The corresponding QMF threshold, computed as the inverse of the spectral radius of the multilayer topology comprising G_a, G_b and the interlink $i \leftrightarrow j$, is also indicated in Figure 3a.

In Figure 3b, we verify the predictions in Lemma 3.3 by simulating epidemics in the same multi-layer structure as

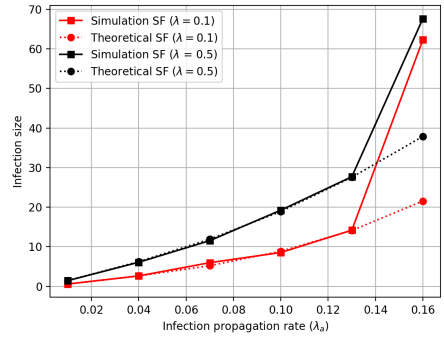
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3108075, IEEE Transactions on Network Science and Engineering



(a)



(b)



(c)

Figure 3: Verifying the epidemic properties on interlink subgraphs: a) variation of extinction time of the infections with hub degrees, b) variation of probability of propagation via interlink with hub degrees, and c) variation of mean infection size with propagation rates. All figures compare our theoretical predictions to the simulated epidemics on multilayer networks with a single interlink.

above but with the initial infection configuration $\{i\}$, i.e. only one of the hubs is infected. We want to characterize the probability that an infection from hub i propagates to hub j via the interlink. We consider the same setting as above, with fixed k_j and variable k_i. We perform 10000 MC runs and the propagation probability is computed as the fraction of runs leading to successful propagation to j. It can be observed from Figure 3b that the simulated probabilities closely follow the theoretical predictions for the three cases, corresponding to different choices of the inter-layer infection propagation rate $\tilde{\lambda} = 0.05, 0.1, 0.15$, corresponding to the black, red and blue plots, respectively.

Finally, in Figure 3c, we compare the theoretical estimation of the infection size presented in (12) with the simulations. Recall that the network topology considered for the verification of the lemmas comprises two hubs, one in each layer. As a consequence of Lemma 3.2 and Lemma 3.3, the metastable infection size in G_a can be approximated by $\lambda_a k_i (1 + \lambda_a)\eta_j$, where η_j denotes the probability of the infection propagating to i from j. The epidemic dynamics are simulated on the multi-layer structure with $k_i = k_j = 100$ and the metastable infection size in G_a with varying infection rate ($\lambda_a{=}\lambda_b{=}\tilde{\lambda}{=}\lambda$) is plotted. This figure illustrates the basic idea behind our claim of controlled infection spreading in networks. Figure 3c shows that the infection sizes closely follow the theoretical predictions for an extended range of λ. At high infection propagation rates, the simulated infection size blows up due to the divergent epidemic dynamics that spreads throughout the network instead of being localized. This illustrates the fact that the controlled infection spreading cannot be achieved at high propagation rates due to the onset of global epidemics affecting a finite fraction of the entire network.

6.2 Comparison of interlink strategies

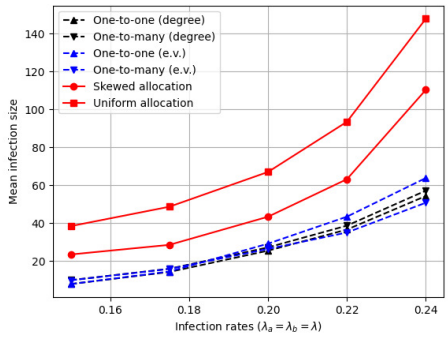
To the best of our knowledge, ours is one of the first works in optimizing the network structure to achieve a *desired* infection size. Most relevant works [31, 32, 39] in the domain of multi-layer interlink design focus on extremizing (maximize or minimize) the spread of epidemics. Interlinking w.r.t. different centrality measures, like degree centrality, eigenvector centrality or other sophisticated measures like Katz centrality [31], are commonly suggested in these works. Note that our focus here is to control the metastable infection size as opposed to these works which focus on the steady state infection size. Nonetheless, due to lack of relevant interlinking algorithms in literature focusing on the metastable infection patterns, we compare our proposed strategy with these popular approaches defined w.r.t. the degree and eigenvector centralities. Note that sophisticated measures like eigenvector centralities are computationally expensive for large network sizes and require the exact knowledge of the network topology, which might not be available in many practical scenarios. This further demonstrates the effectiveness of our algorithms, which only require degree information of the hub nodes and the relative distance between these hubs. We consider two isolated network layers (G_a, G_b), generated as BA graphs of size 10000 with $m = 1$. The desired regions in these layers (D_a, D_b) are specified by the three hop neighborhood around two randomly chosen points a_{rdn} and b_{rdn}. We construct the interlink structure between D_a and D_b following different strategies discussed below and plot the average infection size over 500 MC runs in G_a as a function of the intra-layer infection propagation rate $\lambda_a = \lambda_b = \lambda$; $\tilde{\lambda}$ is fixed at 0.2.

In Figure 4a, we generate the interlinks using the following six strategies: i) one-to-one interlinking w.r.t. degrees, where the nodes in D_a are coupled to nodes in D_b in the decreasing order of their intra-layer degrees, ii) one-to-many interlinking w.r.t. degrees, where the highest degree nodes in D_a are coupled to four random nodes in D_b, iii) one-to-one interlinking w.r.t. eigenvectors [32], iv) one-to-many interlinking w.r.t. eigenvectors, v) skewed interlink

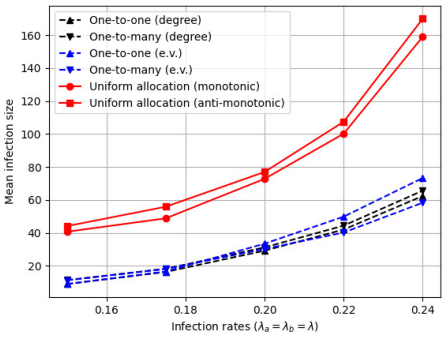
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3108075, IEEE Transactions on Network Science and Engineering

allocation, where a randomized collection of non-uniform interlink allocation strategies are used with the four highest degree hubs in D_a (H_a) and the six highest degree hubs in D_b (H_b) and the average performance is plotted, and vi) uniform interlink allocation, where the same number of hubs as v), i.e. four and six, respectively, are interlinked following the maximally uniform allocation strategy described in Theorem 4.3 with the randomized interlink structure. The first four strategies are baseline algorithms proposed in the relevant literature and the last two are variations of our algorithms for interlink design in multi-layer networks. Degree centrality based interlinking strategies like i) and ii) are by far the most popular heuristic. Eigenvector centrality based interlinking has been shown in recent works [31, 32] to maximize the steady state prevalence of the infection patterns. Note that this is different from the objective considered in this work, where we study the metastable and localized infection patterns which eventually die out in the steady state. Steady state infection patterns are observed for larger infection propagation rates beyond the critical infection threshold λ_c, whereas our focus is on metastable infection patterns which are observed in the neighborhood of λ_c. The specific strategy iii) is proposed in [32] where the nodes ranked by the eigenvector centralites are interlinked to each other in a one-to-one fashion. iv) is a modified version of this strategy in the same line as ii), where nodes with the highest eigenvector centralities in D_a are coupled to four random nodes in D_b.

In our simulations, the baseline strategies are constructed with orders of magnitude more interlinks than our proposed methods; specifically, more than 100 interlinks (in a network region spanning around 3500 nodes) are used in i) through iv) whereas our algorithms v) and vi) use less than 10 interlinks. The number of interlinks used in the last two strategies are $(4 + 6 - 1) = 9$, due to the number of hubs ($|H_a| = 4, |H_b| = 6$) and the minimally connected graph spanning these hubs. The maximally uniform interlink allocation w.r.t. the four (or six) hubs in layer A (or B) is given by $\{3, 2, 2, 2\}$ (or $\{2, 2, 2, 1, 1, 1\}$) or any permutation of this set. Any other allocation of the 9 interlinks constitutes the skewed interlink allocation strategy. Following Conjecture 4.4, the anti-monotonic arrangement is adopted as the interlink strategy for both v) and vi), where the highest degree hubs in H_a are interlinked to the lowest degree hubs in H_b. The initial infection configuration of the network is taken to be the set of interlinked nodes, i.e. the infection seeds of the epidemic are the nodes in D_a and D_b that are interlinked to the other network layer. It can be observed from Figure 4a that the resulting infection sizes of the last two strategies is much higher than the first four strategies, even though only a fraction of the interlinks are used. Among the baseline interlinking structures i) through iv), one-to-one interlinking w.r.t. the eigenvector centrality performs marginally better than the other strategies. It is important to note that at low infection rates near the critical threshold, which is the regime we are interested in, it is known [31] that interlinking strategies defined w.r.t. various centrality measures have comparable performances. Our simulations largely verify this phenomenon, while establishing the superiority of the proposed strategies over the baseline approaches. This clearly shows that enforcing collective interaction between different hubs is a good strategy and centrality based interlink designs popular in literature can perform extremely poorly in the context of localized metastable infection states. Furthermore, the difference between the uniform and the non-uniform allocation of the interlinks is also significant, which illustrates the fact that a uniform allocation is clearly superior to skewed allocation of interlinks. Under certain constraints, we prove that this *maximally uniform* interlink allocation is optimal. Figure 4a provides empirical evidence for Conjecture 4.4, which states that uniform allocation continues to perform well in the general case.



(a)



(b)

Figure 4: Comparing infection sizes under different interlink strategies: a) significant difference between strategies with uniform and skewed allocation, b) limited difference between anti-monotonic and monotonic interlink structures.

Next, we want to observe the impact of different designs of the interlink structure, while adhering to the maximally uniform allocation strategy. The monotonic arrangement interlinks the best (highest degree) hubs in H_a with the best hubs in H_b till a maximally uniform allocation is obtained. The anti-monotonic arrangement interlinks the best hubs in one layer to the worst ones in the other. In Figure 4b, the baseline strategies are the same as Figure 4a, but v) and vi) are replaced by the monotonic and anti-monotonic arrangement of the maximally uniform allocation strategy. Recall that we hypothesize the superiority of the anti-monotonic arrangement, although we were not able to obtain a rigorous proof in the general case. However, we can observe from Figure 4b that the specific interlinking scheme has a limited effect on the infection size, as long as maximally uniform allocation is followed.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSE.2021.3108075, IEEE Transactions on Network Science and Engineering

These results validate our theoretical predictions showing that the local properties of the epidemics derived in Section 3 for the extended star topology hold true for complex topologies in the regime we focus on. This confirms that the global epidemic dynamics on complex networks can be approximated by studying the local dynamics around hubs and quantifying the inter-relationship between these local regions of activity. Furthermore, the comparison of infection sizes for different interlink strategies also shows the superiority of the uniform allocation and the limited impact of the specific interlink structure under a maximally uniform allocation strategy.

6.3 Case Study

We consider the example of promoting inter-disciplinary research among two groups of researchers, representing the isolated network layers. It is important to note that a rigorous modeling of the human interactive patterns in the context of the spread of infectious diseases is an active area of research in itself [40, 41], where significantly complicated architectures are employed to emulate the interaction between the disease-prone agents. For our case study, we consider a simplified interaction model, where a static multi-layer graph is taken to simulate the interaction between the two groups of the disease-prone agents. We consider the case where the two network layers represent the collaboration network between researchers in the areas of Condensed Matter Physics (G_a) and Astrophysics (G_b) [42] , where $|G_a| = 23312$ and $|G_b| = 18872$. An intra-layer link between two nodes $(i, j) \in G_a$ or G_b indicates that the two researchers (nodes) have co-authored a paper in the past. The infection in this case can be thought of as active participation in inter-disciplinary research in terms of publications or submissions of research proposals. The network designer can be imagined to be a funding agency which is interested in devising strategies for allocating funds to pairs of researchers from the two areas (G_a, G_b) with the objective of sustaining inter-disciplinary research activity in a desired region, say in North America $(D_a \cup D_b)$. The allocation of funds to a node pair $(i, j) \in (G_a, G_b)$ is equivalent to the construction of an interlink between these two nodes. Similar to the previous case of comparing different interlinking strategies, the initial infection configuration of the network is taken to be all inter-linked nodes, i.e., the infection seeds comprise the funded researchers engaging in active research efforts in the inter-disciplinary area. This activity of the interlinked nodes can influence their neighbors to engage in this new inter-disciplinary research area; we model this as the propagation of infection from infected nodes to their susceptible neighbors. Thus, the infection can be sustained in the neighborhood of the infection seeds via propagation of infection among inter-linked and intra-linked neighbors. The specific goal of the network designer in our case is to design the interlink structure to realize metastable infection patterns of a desired infection size in the multi-layer network, i.e., the funding agency will give grants to researchers in the two areas with the objective of inducing sustained inter-disciplinary research activity of a desired metastable infection size utilizing a minimum number of grants (interlinks) awarded to pairs of researchers. It goes without saying that a practical model for this problem will involve much more intricate details than those considered here. However, we feel that case studies like these illustrate the wide applicability of the network design problems considered here.

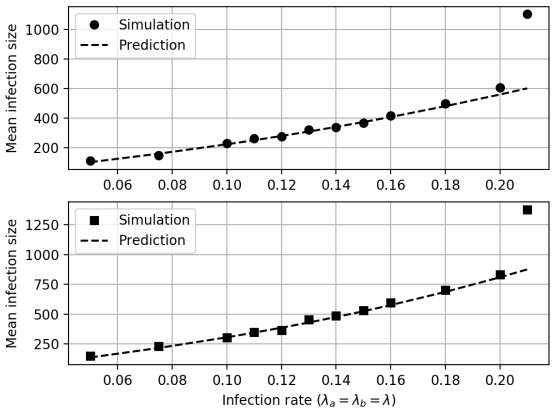


Figure 5: Predicted vs simulated infection size in the two network layers, D_a (top) and D_b (bottom), with varying infection propagation rate for Case 1.

Defining the problem more specifically, the network designer is interested in constructing the inter-layer links between $D_a \subseteq G_a$ and $D_b \subseteq G_b$ so as to ensure that, say, $\gamma_a = 0.05$ fraction of Condensed Matter physicists and $\gamma_b = 0.1$ fraction of Astrophysicists in North America continue to publish interdisciplinary research. Let the infection parameters be given by $\boldsymbol{\lambda} = \{0.1, 0.1, 0.2\}$. We specify the desired regions by the three hop neighborhood around two random points $a_{\text{pos}} \in G_a$ and $b_{\text{pos}} \in G_b$ and obtain the desired regions D_a and D_b with sizes 3842 and 2799, respectively. The desired infection sizes yield: $\xi^a_{\text{des}} \approx 193$ and $\xi^b_{\text{des}} \approx 280$. The required design of the interlink structure can be obtained by Algorithm 1 by calling the method ConstructMeta with the required arguments. Algorithm 1 sequentially grows the metastable infection state until the desired infection sizes are achieved. For this case, we find out that the desired infection sizes are satisfied by considering 5 hubs in D_a and 7 hubs in D_b, all of which belong to the core group of hubs, requiring $11\ (= 5 + 7 - 1)$ interlinks. In Fig. 5, we plot the average infection size corresponding to the proposed interlink structure comprising 11 interlinks for different intra-layer infection propagation rates ($\lambda_a = \lambda_b = \lambda$) and a fixed inter-layer rate $\tilde{\lambda} = 0.2$. The simulated average infection sizes are $(\xi^a, \xi^b) = (229.8, 304.4)$, which satisfies the infection density requirements of this problem. Furthermore, it can be clearly observed from the figure that the theoretically predicted infection sizes agree very well with the simulations for an extended range of λ. This clearly shows that controlled infection spreading is realized here. Note that at higher values of λ, the simulated dynamics do not follow the predictions. This illustrates the limitation of our approach. Our mathematical model for tracking the epidemic dynamics performs well in the near-threshold regime and can accurately predict localized infection patterns. However, the approach does not generalize to higher infection rates characterized by global infection patterns.

7 Conclusion and future works

Recent works [11, 23, 25, 26] study epidemic dynamics on reduced network topologies, capturing the hubs and the inter-relationship among them. We extend this approach to multilayer networks and show that the interlink structure can be designed to realized *controlled* infection spreading. For a wide range of infection propagation rates, we show that the metastable infection size can be theoretically approximated. It is important to remember that epidemics in finite networks continue to exhibit metastability even at higher infection propagation rates. However, as shown in Fig. 3c, the approximations for the metastable infection size developed in this work are not accurate for such cases and novel approaches that can track the epidemic dynamics in this regime are required. Additionally, note that the ideas presented in this work can be applied with certain modifications to single layer networks. However, we adhere to the multi-layer model since it provides a clean distinction between the existing network (intra-layer topology) and the additional manipulations (inter-layer topology), thereby allowing us to better unravel our contributions.

Possible extensions to this work include generalization of the framework to a cost constrained setting similar to [39, 43], where differential costs are associated with the construction of different interlinks. This amounts to the inclusion of additional constraints in the optimization problem (13) to reflect the cost structure. Another important direction for future work is to study interlink design problems in settings with more than two layers, where significant increase in the complexity of the mathematical models tracking the multilayer epidemic dynamics is expected. A fundamental yet crucial idea that is verified through our study is that the dynamics over finite multilayer networks can be reliably approximated by the first order information about the network topology in the near-threshold regime. Quantifying the impact of higher order topological information is another interesting research direction. Such studies will help us to accurately characterize the error of studying epidemic dynamics on approximate network topologies.

8 Acknowledgments

This work was supported in part by the US National Science Foundation under grant CNS-1824518, and in part by the US Army Research Office under Grant W911NF-17-1-0087.

References

[1] S. Hosseinalipour, J. Wang, Y. Tian, and H. Dai. "Infection Analysis on Irregular Networks through Graph Signal Processing". In: *IEEE Transactions on Network Science and Engineering* 7.3 (2019), pp. 1939–1952.

[2] M. Mital et al. "Adoption of Internet of Things in India: A test of competing models using a structured equation modeling approach". In: *Technological Forecasting and Social Change* 136 (2018), pp. 339–346.

[3] C. Hens et al. "Spatiotemporal signal propagation in complex networks". In: *Nature Physics* 15.4 (2019), pp. 403–412.

[4] A. Medvedev and J. Kertesz. "Empirical study of the role of the topology in spreading on communication networks". In: *Physica A: Statistical Mechanics and its Applications* 470 (2017), pp. 12–19.

[5] J. Woo and H. Chen. "Epidemic model for information diffusion in web forums: experiments in marketing exchange and political dialog". In: *SpringerPlus* 5.1 (2016), p. 66.

[6] R. Pastor-Satorras and A. Vespignani. "Epidemic dynamics and endemic states in complex networks". In: *Physical Review E* (2001), p. 066117.

[7] P. V. Mieghem. "The N-intertwined SIS epidemic network model". In: *Computing* 93.2-4 (2011), pp. 147–169.

[8] E. Cator and P. V. Mieghem. "Susceptible-infected-susceptible epidemics on the complete graph and the star graph: Exact analysis". In: *Physical Review E* (2013), p. 012811.

[9] R. Pastor-Satorras et al. "Epidemic processes in complex networks". In: *Reviews of modern physics* (2015), p. 925.

[10] B. Qu and H. Wang. "SIS epidemic spreading with heterogeneous infection rates". In: *IEEE Transactions on Network Science and Engineering* 4.3 (2017), pp. 177–186.

[11] S. Chatterjee and R. Durrett. "Contact processes on random graphs with power law degree distributions have critical value 0". In: *The Annals of Probability* (2009), pp. 2332–2356.

[12] A. V. Goltsev et al. "Localization and spreading of diseases in complex networks". In: *Physical review letters* (2012), p. 128702.

[13] H. K. Lee, P. Shim, and J. D. Noh. "Epidemic threshold of the susceptible-infected-susceptible model on complex networks". In: *Physical Review E* (2013), p. 062812.

[14] N. Berger et al. "On the spread of viruses on the internet". In: *Soda*. 2005, pp. 301–310.

[15] A. Ganesh et al. "The effect of network topology on the spread of epidemics". In: *Proceedings of the IEEE Computer and Communications Societies*. 2005, pp. 1455–1466.

[16] C. Castellano and R. Pastor-Satorras. "Thresholds for epidemic spreading in networks". In: *Physical review letters* (2010), p. 218701.

[17] Y. Zhang, X. Li, and A. V. Vasilakos. "Spectral analysis of epidemic thresholds of temporal networks". In: *IEEE transactions on cybernetics* 50.5 (2017), pp. 1965–1977.

[18] R. H. Schonmann. "Metastability for the contact process". In: *Journal of statistical physics* (1985), pp. 445–464.

[19] M. Nykter et al. "Gene expression dynamics in the macrophage exhibit criticality". In: *Proceedings of the National Academy of Sciences* 105.6 (2008), pp. 1897–1900.

[20] C. Furusawa and K. Kaneko. "Adaptation to optimal cell growth through self-organized criticality". In: *Physical review letters* 108.20 (2012), p. 208103.

[21] W. Bialek et al. "Statistical mechanics for natural flocks of birds". In: *Proceedings of the National Academy of Sciences* 109.13 (2012), pp. 4786–4791.

[22] P. Moretti and M. A. Muñoz. "Griffiths phases and the stretching of criticality in brain networks". In: *Nature communications* 4.1 (2013), pp. 1–10.

[23] T. Mountford et al. "Metastable densities for the contact process on power law random graphs". In: *Electronic Journal of Probability* (2013).

[24] V. H. Can. "Super-exponential extinction time of the contact process on random geometric graphs". In: *Combinatorics, Probability and Computing* (2018), pp. 162–185.

[25] M. Boguná, C. Castellano, and R. Pastor-Satorras. "Nature of the epidemic threshold for the susceptible-infected-susceptible dynamics in networks". In: *Physical review letters* (2013), p. 068701.

[26] S. C. Ferreira, R. S. Sander, and R. Pastor-Satorras. "Collective versus hub activation of epidemic phases on networks". In: *Physical Review E* (2016), p. 032314.

[27] M. Zheng et al. "Multiple peaks patterns of epidemic spreading in multi-layer networks". In: *Chaos, Solitons & Fractals* (2018), pp. 135–142.

[28] A. Saumell-Mendiola et al. "Epidemic spreading on interconnected networks". In: *Physical Review E* (2012), p. 026106.

[29] G. F. de Arruda et al. "Disease localization in multilayer networks". In: *Physical Review X* (2017), p. 011014.

[30] G. F. de Arruda et al. "Universality of eigenvector delocalization and the nature of the SIS phase transition in multiplex networks". In: *arXiv preprint arXiv:2005.08074* (2020).

[31] L. Pan, W. Wang, S. Cai, and T. Zhou. "Optimal interlayer structure for promoting spreading of the susceptible-infected-susceptible model in two-layer networks". In: *Physical Review E* 100.2 (2019), p. 022316.

[32] L. Pan, W. Wang, S. Cai, and T. Zhou. "Optimizing spreading dynamics in interconnected networks". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.10 (2019), p. 103106.

[33] B. Schapira and D. Valesin. "Extinction time for the contact process on general graphs". In: *Probability Theory and Related Fields* 169.3 (2017), pp. 871–899.

[34] P. V. Mieghem et al. "Virus spread in networks". In: *IEEE/ACM Transactions On Networking* (2008), pp. 1–14.

[35] N. A. Ruhi, C. Thrampoulidis, and B. Hassibi. "Improved bounds on the epidemic threshold of exact SIS models on complex networks". In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE. 2016, pp. 3560–3565.

[36] M. Newman. *Networks*. Oxford university press, 2018.

[37] S. Chattopadhyay, H. Dai, S. Hosseinalipour, and D. Y. Eun. "Designing Optimal Interlink Patterns to Maximize Robustness of Interdependent Networks against Cascading Failures". In: *IEEE Transactions on Communications* (2017).

[38] D. T. Gillespie. "Exact stochastic simulation of coupled chemical reactions". In: *The journal of physical chemistry* (1977), pp. 2340–2361.

[39] A. Tavasoli, E. Ardjmand, and H. Shakeri. "Maximizing the algebraic connectivity in multilayer networks with arbitrary interconnections". In: *arXiv preprint arXiv:2008.13036* (2020).

[40] Yi-Qing Zhang, Xiang Li, Jian Xu, and Athanasios V Vasilakos. "Human interactive patterns in temporal networks". In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 45.2 (2014), pp. 214–222.

[41] Haimiao Mo, Shuai Ding, Shanlin Yang, Xi Zheng, and Athanasios V Vasilakos. "The Role of Edge Robotics As-a-Service in Monitoring COVID-19 Infection". In: *arXiv preprint arXiv:2011.08482* (2020).

[42] J. Leskovec et al. "Graph evolution: Densification and shrinking diameters". In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1.1 (2007), p. 2.

[43] S. Chattopadhyay, H. Dai, and D. Y. Eun. "Maximization of robustness of interdependent networks under budget constraints". In: *IEEE Transactions on Network Science and Engineering* 7.3 (2019), pp. 1441–1452.

Srinjoy Chattopadhyay (S'15) received the B.Tech. degree in electronics and electrical communication engineering and the M.Tech. degree in telecommunication systems engineering from IIT Kharagpur, India, in 2013. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, North Carolina State University, USA. His current research interests include wireless communications, computer networks, spectral theory of graphs, and network design problems on multilayer interdependent networks.

Huaiyu Dai (F'17) received the B.E. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1996 and 1998, respectively, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ in 2002.

He was with Bell Labs, Lucent Technologies, Holmdel, NJ, in summer 2000, and with AT&T Labs-Research, Middletown, NJ, in summer 2001. He is currently a Professor of Electrical and Computer Engineering with NC State University, Raleigh, holding the title of University Faculty Scholar. His research interests are in the general areas of communications, signal processing, networking, and computing. His current research focuses on machine learning and artificial intelligence for communications and networking, multilayer and interdependent networks, dynamic spectrum access and sharing, as well as security and privacy issues in the above systems.

He has served as an editor for IEEE Transactions on Communications, IEEE Transactions on Signal Processing, and IEEE Transactions on Wireless Communications. Currently he is an Area Editor in charge of wireless communications for IEEE Transactions on Communications, and a member of the Executive Editorial Committee for IEEE Transactions on Wireless Communications. He was a co- recipient of best paper awards at 2010 IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS 2010), 2016 IEEE INFOCOM BIGSECURITY Workshop, and 2017 IEEE International Conference on Communications (ICC 2017).

Do Young Eun (M'03–SM'15) received the B.S. and M.S. degrees in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1995 and 1997, respectively, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2003. Since 2003, he has been with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA, where he is currently a Professor. His research interests include network modeling and performance analysis, mobile ad hoc/sensor networks, mobility modeling, social networks, and graph sampling. Dr. Eun has been a member of the Technical Program Committee of various conferences including IEEE INFOCOM, ICC, GLOBECOM, ACM MobiHoc, and ACM Sigmetrics. He served on the Editorial Board of the IEEE/ACM Transactions on Networking and Computer Communications, and was TPC Co-Chair of WASA 2011. He received the Best Paper awards in the IEEE ICCCN 2005, IEEE IPCCC 2006, and IEEE NetSciCom 2015, and the National Science Foundation CAREER Award 2006. He supervised and coauthored a paper that received the Best Student Paper Award in ACM MobiCom 2007.