PNAS NEXUS

PNAS Nexus, 2026, **5**, pgag036

https://doi.org/10.1093/pnasnexus/pgag036
Advance access publication 23 February 2026
Research Report

Scenario-aware control of multipathway spread processes: Application to biological invasions

Prathyush Sambaturu [a], Manisha Sudhir [b], Hongze Chen [b], Anil Vullikanti [b,c], Rangaswamy Muniappan [d] and Abhijin Adiga [b,*]

[a]Department of Computer Science, University of Oxford, Oxford OX1 3QD, Oxfordshire, United Kingdom
[b]Biocomplexity Institute, University of Virginia, Charlottesville, VA 22903, USA
[c]Department of Computer Science, University of Virginia, Charlottesville, VA 22904, USA
[d]Center for International Research, Education, and Development, Virginia Tech, Blacksburg, VA 24061, USA
*To whom correspondence should be addressed: Email: abhijin@virginia.edu
Edited By Derek Abbott

Abstract

Optimal control of spread processes over networks is a challenging problem, even for simple diffusion models. Real-world processes—such as infectious disease outbreaks and biological invasions—often involve multiple spread pathways and time-varying network dynamics. In this work, we address the problem of region-wide interventions, where the goal is to select an optimal set of regions (groups of nodes) in a network to minimize spread, subject to budget constraints, intervention delays, and a given spread scenario which reflects prior knowledge of the process—such as initial infection locations, parameter estimates, and other context-specific assumptions. We present a general approach based on integer linear programming and sample average approximation, applicable across a broad class of diffusion models. We also establish theoretical performance guarantees for our method within the bicriteria approximation framework. To demonstrate its effectiveness, we apply the approach to model the spread of a representative agricultural pest. Our method yields near-optimal solutions and consistently outperforms standard baselines. The results emphasize the value of scenario-specific intervention strategies, showing that early action can significantly reduce spread under limited budgets and produce stable outcomes even under model uncertainty.

Keywords: combinatorial optimization, multipathway spread, interventions, biological invasions

Significance Statement

We present an optimization framework for designing scenario-aware interventions in network-based spread processes as applicable to a variety of epidemiological processes such as biological invasions, infectious diseases, and social contagions. Our method identifies near-optimal intervention solutions under constraints such as limited budgets, delays, and uncertainty in model parameters accounting for any prior knowledge of the contagion process. Applied in the context of biological invasion, the method consistently outperforms existing baselines and highlights the benefit of early, scenario-informed interventions. This framework enables principled planning for containment in diverse applications, from agriculture to epidemiology.

Introduction

Many real-world spreading phenomena, such as the transmission of infectious diseases, invasive species, memes, and malware, can be naturally represented as propagation processes over networks (1–4). Some examples include the spread of infectious diseases, invasive species, memes, and malware. Mitigating such spread processes is well-studied from a network perspective and corresponds to removing targeted nodes or edges in order to minimize the spread (4, 5). There are practical constraints, such as limited availability of resources, costs incurred, delay in discovery of the outbreaks, and the inability to respond promptly, that limit how many nodes can be intervened at and how quickly one can achieve this (6–9).

However, controlling network propagation processes is computationally very challenging due to the cascading effects (10). Most of the existing solutions are entirely based on structural properties of the underlying network that are known to influence the spread such as centrality measures or spectral properties (11–15). The few recent works that do account for the dynamics are only applicable to simple epidemiological models (13, 16–19). In most practical settings, the high complexity of the stochastic spread processes manifest in their models. One has to rely on in silico studies to understand spread dynamics as closed-form analytical solutions are seldom possible (20–29). In addition, these approaches for control are not typically equipped to incorporate priors such as possible introduction scenarios and response

Competing Interest: The authors declare no competing interests.
Received: June 18, 2025. **Accepted:** January 8, 2026

capabilities. Our objective is to bridge this gap. We develop a generic simulation-driven approach for scenario-specific control of network propagation processes with provable guarantees and apply it in the context of biological invasions.

Pests and pathogens (P&P) are a major threat to the environment, agriculture, and health (30–34). The spread of P&P is a multipathway phenomenon driven by various natural and anthropogenic factors (35), which can be modeled as a network propagation process, (6, 36–45), where nodes represent spatial regions (eg grid cells, farms, counties, and hosts) and edges represent flows between regions (eg through wind, trade of crops, or animal movement) through which the organism spreads over the network. An example of such a network is provided in Fig. 1b.

Modeling has played an important role in preparing for an invasion as well as guiding surveillance and interventions during an outbreak. Risk assessment tools are regularly used to assess emerging threats by identifying pathways of introduction and spread, determining the spatial distribution of establishment potential, and evaluating management methods (46, 47). There is extensive work on evaluating reactive strategies such as host removal and ring vaccination (livestock) during an outbreak (21–24, 27, 29, 48). However, in the context of preparing for an impending invasion, these works do not directly address the problem of where to intervene when resources are limited. Existing methods provide solutions based on structural properties of the underlying network without considering probable introduction scenarios and intervention delays (43, 44, 49, 50). Developing practical scenario-specific intervention strategies at the appropriate scale that account for model complexity, heterogeneity, and resource limitations is an important requirement in this domain (51).

In this work, we introduce a new budgeted group intervention problem with time delays. Let $G(V, E)$ be a graph with a collection of mutually disjoint subsets of nodes called *groups*. While in general, a group can be an arbitrary subset of nodes, from the domain perspective, it is a patch of spatially contiguous nodes (such as a city or a county, see Fig. 1b). A spread scenario $\mathcal{S}$ encapsulates prior knowledge, including the diffusion model, its parameters (or more generally, their distributions), and the potential seeding of the contagion. Intervening at a group corresponds to removing all nodes belonging to that group, thus controlling the spread. The objective is to determine the optimal set of regions to intervene, satisfying the budget constraints, so that the spread is minimized under the given scenario. The formal definition is as follows. Intervening at a group corresponds to removing all nodes belonging to the group, thus controlling the spread.

IASControl problem: *Let $G = (V, E)$ be a temporal edge-weighted and edge-labeled directed graph with a set of groups $\mathcal{Q}$ and $\mathcal{S}$ be a spread scenario specifying priors corresponding to the stochastic spread process and potential introduction of the contagion. Given a budget B on the number of groups that can be intervened, an intervention delay τ_d (the number of time-steps after revealing the source at which the intervention is performed), the objective is to select B groups to intervene at time step τ_d so as to minimize the spread for scenario $\mathcal{S}$.*

Node-level intervention, which is very well studied, is a special case where each node belongs to its own group. The interventions considered in this work are nonadaptive, ie the decision to intervene is not made by observing the system state at time $\tau_d - 1$ or before, unlike reactive control. Instead, it is based on the expected state of the system at τ_d. It corresponds to the delay anticipated by the modeler between the start of the spread and the implementation of the intervention, which could be due to various reasons such as the late discovery of the spread due to insufficient surveillance or the lack of infrastructure for prompt control of spread.

Contributions

This work addresses the IASControl problem for a broad class of discrete-time network diffusion models. We show that for even very restricted settings, the IASControl problem is computationally hard. Our approach to solve this problem involves sample average approximation (SAA) (52–54) and rounding linear programs constructed from a (small) user-generated scenario-specific sample of simulation instances (henceforth referred to as *cascades*). Figure 1a provides an outline of the process). This general framework enables a user to specify various (often complex) what-if counterfactuals and better prepare for an impending or ongoing invasion. Rigorous theoretical guarantees with respect to the optimal solution and extensive experimentation on realistic networks demonstrate the effectiveness as well as feasibility of the method. We apply our approach to study the spread of the South American tomato leafminer (55) (*Phthorimaea absoluta*), a representative pest that has rapidly spread worldwide. We demonstrate the superior performance of our approach by comparing it with multiple baselines. We extensively analyze the solutions with respect to the quality of approximation, stability with respect to cascade sample size, and model uncertainty. Our experiments emphasize the need for effective surveillance towards early discovery of P&P that enables effective mitigation under various uncertainties, while delays in intervening can lead to larger budget requirements and variance in solutions.

Results

Our work is applicable to the general class of SEI(RS) discrete-time network diffusion processes, where **S**, **E**, **I**, and **R** stand for node states Susceptible, Exposed, Infectious, and Recovered, respectively. In this work, we will focus on the MultiPath model of McNitt et al. (43) in the context of invasive species spread over a geographic landscape. It is an SEI process defined over a multiscale directed weighted network. All these models are defined in the Methods.

Computational hardness

We show that the group-scale intervention problem for MultiPath is NP-hard even when the graph G is a tree. Further, while the approximation hardness for IASControl is still open, we show that a variation—where a target bound K on the number of infections is given and the goal is to minimize the budget so that the expected number of infections is at most K—is very hard to approximate due to the group-level decisions. Formal statements are provided in the Supporting information (SI).

SpreadBlocking framework

We present a framework to solve the IASControl problem (see Fig. 1). Then, we argue that this approach can be generalized to other complex diffusion models that follow SEIR-class dynamics such as SI, SIS, SEI, SEIR, etc. Our framework has the following two main steps (see Fig. 1a): (Step 1): Represent the dynamics of the complex epidemic model as an Susceptible-Infectious-Removed (SIR) process (overview of dynamics of SIR model are presented in Methods). This can be achieved by the notion of auxiliary graphs such as *time-expanded networks*. (Step 2): Solve the control problem corresponding to SIR process on the auxiliary graph. We use the SAA technique from stochastic optimization and the linear programming (LP) relaxation and rounding techniques to solve the control problem (16, 18).

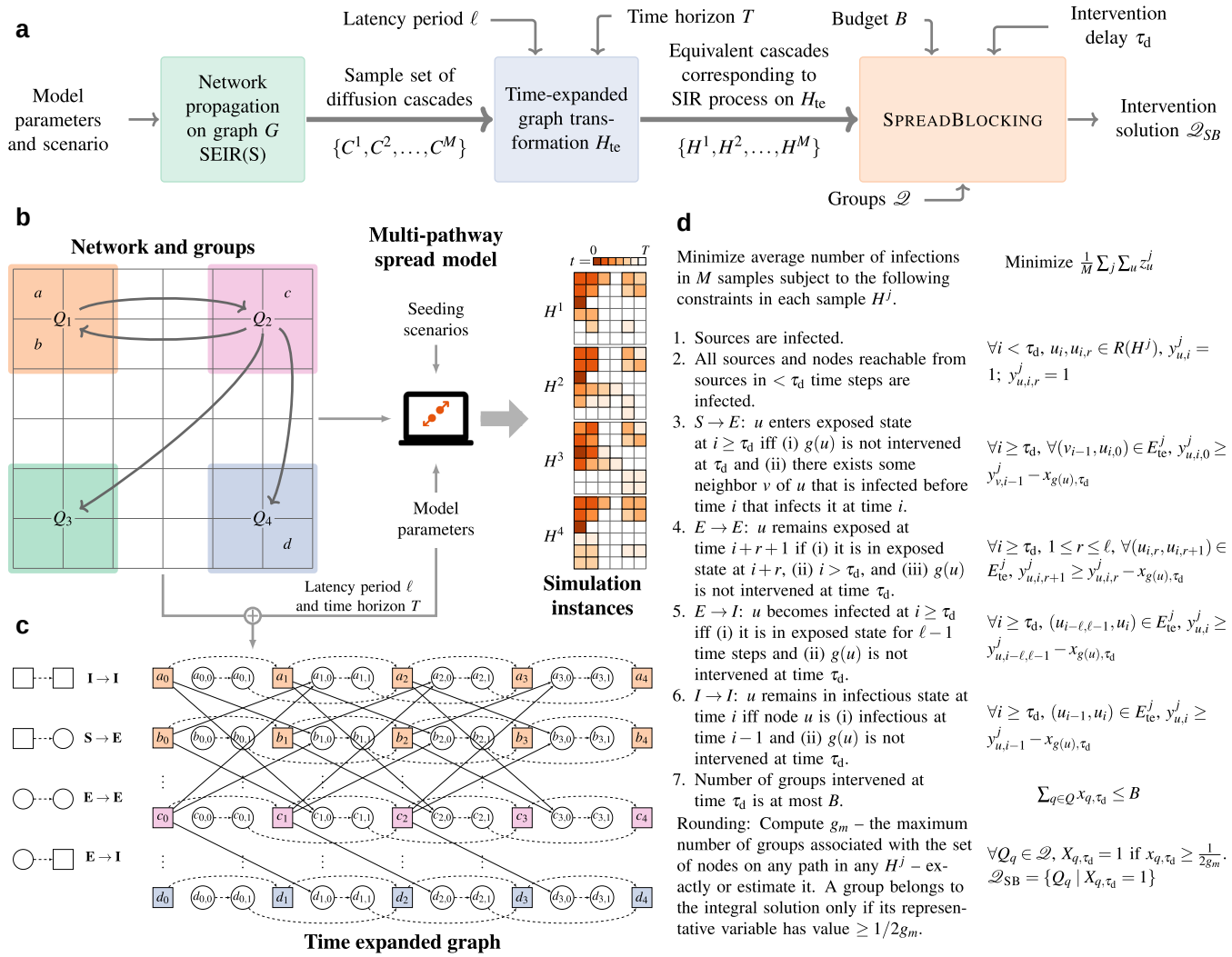

Fig. 1. SpreadBlocking framework. a) The workflow comprising of simulation of the user-defined scenario followed by transformation of the cascades to a standard time-expanded graph form, and the application of the SpreadBlocking algorithm on the transformed cascades. b) An example spatial network G with nodes (grid cells such as a and b) and groups of nodes Q_i. Cascades H^j corresponding to a scenario are depicted. c) The time-expanded graph H_{te} corresponding to G for the given latency period ℓ (=2) and time horizon T (=4) is depicted. Also shown are the different types of nodes and the corresponding transitions. d) The simplified form of the SpreadBlocking integer linear program (ILP), where all variables are Boolean. The ILP is relaxed and solved. The final solution $\mathcal{Q}_{SB}$ is obtained by rounding the solution of the relaxed program. In some of our analysis, we also apply the heuristic where the B groups with the highest group variable values are selected as the solution.

(Step 1) Representing MultiPath model as an SIR process

We represent the MultiPath model as an SIR process (instead of the SEI process) on an auxiliary network called the *time-expanded network*. Let $H_{te}(V_{te}, E_{te})$ denote the time-expanded network corresponding to the multipathway model on $G(V, E)$. See Fig. 1c for a time-expanded graph of G in Fig. 1b. The key idea is to treat every node u at each time step as a distinct node, ie we have $T+1$ copies $\{u_0, \ldots, u_T\}$ of u, where u_i represents the copy of u at time step i. Here, T is the *time horizon* of the spread process. To incorporate the exposed state in the underlying SEI process of the multipathway model, we have ℓ additional copies $\{u_{i,0}, \ldots, u_{i,\ell-1}\}$, corresponding to each u_i, where ℓ is the latency period (time to transition from exposed **E** to infectious **I**). The edge set E_{te} consists of exactly the following four types of edges which corresponds to different events in a SEI process: $\mathbf{S} \to \mathbf{E}$, $\mathbf{E} \to \mathbf{E}$, $\mathbf{E} \to \mathbf{I}$, and $\mathbf{I} \to \mathbf{I}$. Figure 1c shows these transitions. The details linking the edges of H_{te} to G are in the SI.

Let $\mathcal{O}_G$ denote a stochastic disease outcome of the SEI model on G or simply the cascade—this specifies the state $\sigma_G(v, t)$ for each (v, t), and set of the edges (u, v, λ, t) such that node u infects v at time t through pathway λ. Similarly, let $\mathcal{O}_{H_{te}}$ denote a cascade in the SIR model on H_{te}. The following result establishes the equivalence between the MultiPath on G and the SIR process on H_{te}. The proof is in the SI.

Theorem 1 Consider the multipathway diffusion process on $G(V, E)$ for T time steps with a latency period $\ell \geq 0$ and the SIR process on the corresponding time-expanded graph $H_{te}(V_{te}, E_{te})$. Then, for any cascade $\mathcal{O}_G$ and a consistent cascade $\mathcal{O}_{H_{te}}$, the probability that $\mathcal{O}_G$ is the cascade in the multipathway process on G is equal to the probability that $\mathcal{O}_{H_{te}}$ is the cascade in the SIR process on H_{te}.

(Step 2) Intervention Algorithm

In Figure 1a, we provide the outline of the SpreadBlocking and in Fig. 1d, a simpler form of the algorithm is provided. A more detailed version of the algorithm is in the SI. Here, we extend the

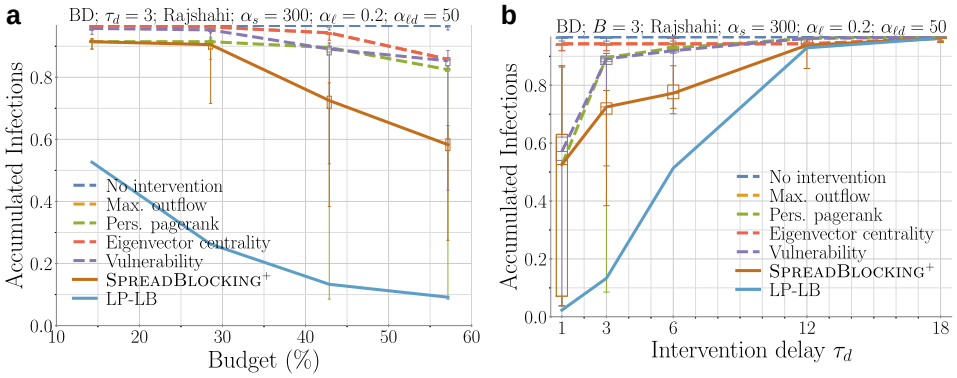

Fig. 2. Comparison of accumulated infections resulting after employing the algorithm(s) with respect to (a) budget and (b) intervention delay for one network and one scenario. The lines correspond to median values. These plots are for the rounding threshold of $1/2g_m$. The titles contain the following information in the order in which they are mentioned: network, budget/delay, seeding scenario, and pathway parameters α_s and $\alpha_{\ell d}$ (short-distance and long-distance parameters defined in Methods in the description of the MultiPath. Plots for other networks is available in the SI.

SAA-based techniques developed for node-level and edge-level intervention problems (16, 18). Firstly, M simulation outcomes or cascades $\{C^1, \ldots, C^M\}$ are generated for the scenario $\mathcal{S}$. These cascades from the MultiPath process are converted to cascades corresponding to the SIR process on the time-expanded graph $\{H^1, \ldots, H^M\}$. An ILP is constructed given these cascades, intervention delay τ_d, and budget B (see Fig. 1d). This ILP is relaxed to a linear program LP_{τ_d}, which is solved (in polynomial time) to obtain an optimal fractional solution. The fractional solution is rounded using a simple threshold $\frac{1}{2g_m}$ for each variable x_{q,τ_d} corresponding to group q. Here, g_m is the maximum number of groups associated with the set of nodes on any path in any cascade H^j, which can be computed efficiently (details in the SI). Groups whose variables were rounded to 1 correspond to the solution $\mathcal{Q}_{SB}$ selected for intervention. We provide theoretical performance guarantees for SpreadBlocking with respect to the optimal solution for the IAScontrol. In particular, we show that SpreadBlocking is a bicriteria approximation algorithm (formally stated in the SI). Our main result is as follows (proof in the SI).

Theorem 2 Given budget B and any positive real number $\epsilon < 1$, let $M \geq (6B + 3)\epsilon^{-2}n$ be the number of cascades. Let $\mathcal{Q}_{SB}$ be the intervention set computed by SpreadBlocking algorithm. Then with probability $1 - \frac{1}{|\mathcal{Q}|}$, (i) $inf_T(V(\mathcal{Q}_{SB})) \leq 2(\frac{1+\epsilon}{1-\epsilon})inf_T(V(\mathcal{Q}^*))$, where $\mathcal{Q}^* \subseteq \mathcal{Q}$ is an optimal solution for the given instance of IAScontrol and (ii) $|\mathcal{Q}_{SB}| \leq 2g_mB$.

Note that the approximation is at two levels: (i) Infection size, which is within a multiplicative factor of the infection size corresponding to an optimal solution and (ii) Budget violation: the number of groups intervened is within a multiplicative factor of budget B. Note that these are worst-case bounds. In practice, even for very small sample sets of cascades, the algorithm achieves much smaller multiplicative factors as demonstrated in the following discussion. We propose a simple method to provide solutions that satisfy the budget constraint exactly. First, SpreadBlocking is called for multiple budget values. Then, we construct a solution by taking the largest available solution that is within the budget and add to it groups of high marginal importance from outside this set. A precise description is provided in the Methods. The resulting algorithm is denoted as SpreadBlocking+.

Performance evaluation

We compared SpreadBlocking+ to several baselines for different B and τ_d values by applying it on five realistic networks each corresponding to a country relevant to the spread of *P. absoluta*: Bangladesh (BD), Indonesia (ID), Philippines (PH), Thailand (TH), and Vietnam (VN) (details in Methods and the SI). In each case, the average fraction of nodes infected and budget violation was used as the metric for evaluation. We used four baselines, which are described in the methods. The baselines *Maximum outflow* and *Eigenvector centrality* choose candidates based only on their position in the network, while the *Vulnerability* baseline chooses candidates based on their role in the diffusion process. Only the baseline *Personalized pagerank* is seeding-scenario-specific but does not consider the diffusion model. We use the objective value from the relaxed program to compare with the integral solution: LP-LB (LB for lower bound) is the objective value corresponding to budget B'. It serves as a lower bound for expected number of infections for the integral solution, which uses a budget of B'.

Efficacy

Representative results for two networks are in Figure 2, one corresponding to increasing B for a fixed τ_d and the other corresponding to increasing τ_d for a fixed B. Summary results across networks, scenarios, B, and τ_d are shown in Figure 3a. In general, the performance of SpreadBlocking is superior compared to the baselines. We note that for lower B and τ_d, the performance is better suggesting that the solutions provided by SpreadBlocking for early intervention and limited resources scenarios are significantly better than other schemes. We note that SpreadBlocking has much better approximation guarantees in practice compared to the bound in Theorem 2. For most cases, the factor is <1.5. Even when comparing with the lower bound LP-LB (Fig. 3b), we notice that for almost all cases, the approximation factor is around 1.6 indicating that the performance is near-optimal for the considered networks and scenarios.

Budget violation

Figure 3c corresponds to budget violation given by the ratio of budget of the solution provided by the algorithm to the given budget with respect to a specific threshold used for rounding ($1/2g_m$ in the figure). In most cases, the budget violation is at most 2. This ratio goes down further with increase in budget. Occasionally, we also see that the given budget is higher than what is required, leading to a solution with fewer groups to intervene at. We observe that the threshold $1/2g_m$ is a conservative choice and leads to budget violations that are higher than necessary. We experimented with higher thresholds as well (see SI) and

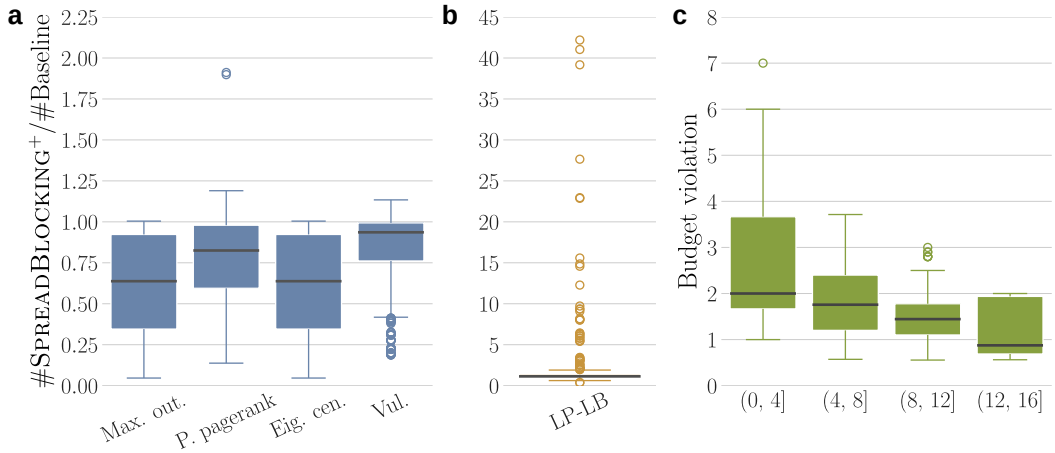

Fig. 3. Evaluating the results with respect to the number of infections and budget violation. This is a summary evaluation across networks, model parameters, seeding scenarios, budget and intervention delay. a) Performance evaluation of SpreadBlocking+ with respect to the accumulated infections: We plot the ratio of the number of infections given the solution of SpreadBlocking+ (#SpreadBlocking+) to the number of infections given the solution of the baseline (#Baseline) mentioned on the x-axis. Lower the better. b) Performance evaluation of SpreadBlocking+ with respect to the LP lower bound, and c) Budget violation with respect to user-given budget B. We plot the ratio of the SpreadBlocking budget and the user-given budget. Lower the better. Results for other thresholds are provided in the SI.

observed that the budget violation can be reduced significantly with a small impact on the number of infections.

Stability of solutions across sample sets

Since the solution of $\mathscr{S}pread\mathscr{B}locking$ depends on the sampled cascades, it is possible that the solutions vary across different sampled sets corresponding to the same simulation scenario. From an end-user perspective, this is important as a large variation in the solutions across different sample sets could lead to less confidence in decision making particularly if the solutions differ in efficacy and budget violation. To analyze the variability in the solution set, we generated multiple sets of cascades with the scenario and number of cascades per set (M) fixed. Then, we computed Jaccard index for each pair of the solutions to quantify the similarity between them. The first row in Fig. 4 shows the results for two networks. More results are in the SI. We observe that variation in solutions depends mainly on M. The greater the M, the lesser the variability. Consistently across networks, the variability is low for $M \geq 100$. We note that, to a lesser extent, budget and intervention delay also play a role. Higher budget and higher intervention delay mean more candidates for intervention sets, and therefore, a higher chance of variance. The second row in Fig. 4 shows the distribution of the objective value (infection size) corresponding to each solution. As M increases, even if several solution sets are possible, we observe that they have very similar effectiveness in controlling the spread.

Analysis of the scenario-specific solutions

Intervention delay and model uncertainty

We studied the stability of solutions with increasing intervention delay under model uncertainty. Due to uncertainty in model parameters, there can be multiple intervention solutions for a given spread scenario, budget, and intervention delay. To analyze these solutions under model uncertainty and thereby, provide robust solutions, we used a ranking-based approach. For a given seeding scenario and intervention delay τ_d, we gathered all solutions corresponding to the various model parameters (see Experiment setup in Methods) and budget values. We ordered the groups by their frequency of occurrence in these solutions with the interpretation that, higher the frequency, the more important is the group. Some of these plots are shown in Fig. 5 (last column) for one network with different seeding scenarios. We note that as the intervention delay increases, the ranking of groups changes significantly. Major production areas continue to be important with increasing τ_d. However, groups that are not production areas but are closer to the seeded location are important in the beginning, but lose their importance later. Even major production areas can lose their importance when all their neighbors are infected. Groups that were far away from seed groups to begin with will gain importance as the spread front approaches them. We also observed that with increasing τ_d, since the invasion is more widespread, the number of unique groups appearing in different solutions increases, and it becomes difficult to choose one group over the other. This underscores the need for *early discovery of the P&P and speed of intervention*.

The effect of introduction scenarios

To study the dependence of the intervention solution on the seeding location(s), we considered different scenarios based on the pathway analysis in McNitt et al. (43). In Fig. 5, two seeding scenarios are shown. We note that the solutions and their evolution with respect to delay τ_d can be very specific to seeding scenarios. Under uncertainty in the seeding scenarios, the rank plots help identify common groups (if any) that can be intervened at for different τ_d.

Temporality of the pathway network and time of introduction

Since the network is time-varying, the rate of spread varies over the time. During peak growing season, high volumes of the host crops are traded facilitating faster spread. At the same time, during off-season, many nodes do not have any production, and therefore, are not suitable for establishment of the pest. Therefore, it is important to study interventions with respect to time of introduction. We considered different months of the year as the time of introduction keeping the budget and intervention delay constant. The results for one of the networks is in Fig. 6. We observe that significance of timing of interventions depends on the peak growing season. For example, intervening early is crucial at start month 3 due to peak production in many localities and high volume of trade, but not as crucial when the start month is 7.

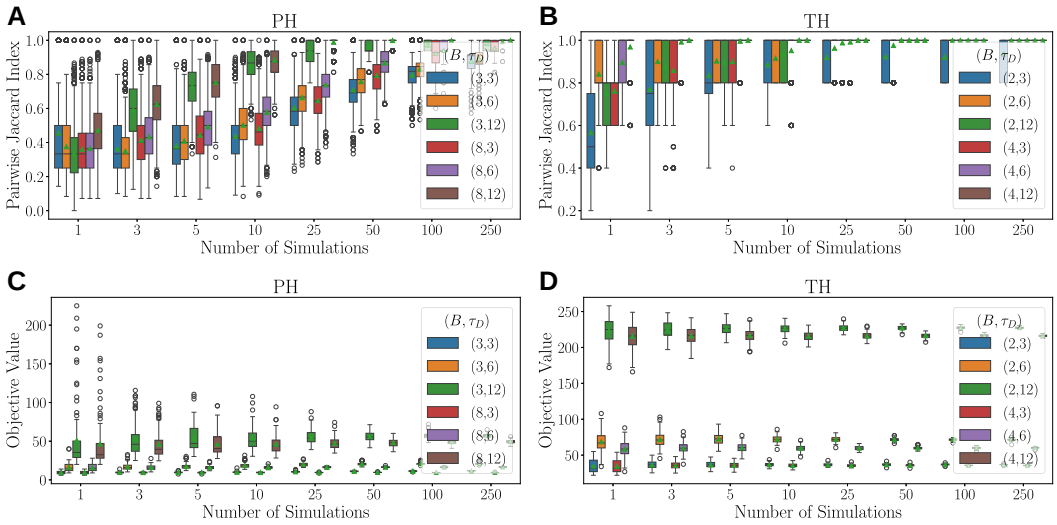

Fig. 4. Solution stability with respect to number of simulation instances M: Keeping the network and spread model parameters fixed, the number of simulation instances M was varied for several budget–intervention delay pairs. For each M, solutions for 100 independent sample sets were generated. The plots show the distribution of Jaccard index for each pair of solutions corresponding to two networks. More results are in the SI. a) and b) correspond to pairwise Jaccard index as the metric for solution similarity, while c) and d) correspond to variation in the objective value across different solutions.

Such knowledge can help in planning interventions such as biological control and setting of traps whose efficacy can be time dependent.

The influence of pathways on the efficacy of interventions

Importance of pathways has been considered from the perspective of extent of spread and damage (38, 43). Here, the focus is on the efficacy of the intervention. In the MultiPath model, three types of pathways (short-distance or natural, local human-mediated dispersal, and long-distance human-mediated dispersal) are considered. For a fixed network and introduction scenario, we varied the short-distance pathway (parameter α_s) and long-distance human-mediated ($\alpha_{\ell d}$) pathway parameters. For the resulting models, SpreadBlocking was applied to compute the intervention solutions for various combinations of budget and intervention delay. Our results are in Fig. 7, where the efficacy of the intervention is evaluated based on the number of infections. To compare the solutions corresponding to different (α_s, $\alpha_{\ell d}$) pairs, we used the number of infections corresponding to LP-LB—the objective value of the relaxed program corresponding to budget B—as the actual solutions of the SpreadBlocking algorithm might violate budget differently. When the intervention delay (τ_d) is small, with a small budget, long-distance spread can be effectively mitigated. In this case, the spread is controlled even for large values of α_s. However, for a larger τ_d (12), we observe that even a large budget is not effective in preventing the spread even for low values of α_s and $\alpha_{\ell d}$. This behavior highlights the importance of multipathway spread in the context of intervention efficacy. Network structure and the introduction scenario play an important role as well.

Network-specific observations

In networks with multiple interconnected localities that are large production centers (like VN and BD), the spread is difficult to control. However, if the production centers are far apart, it is easier to isolate the affected localities and prevent further spread. Our experiments show that the Bangladesh network (BD) presents the greatest control challenge. This difficulty stems from the country's relatively small area (it is half the size of the next smallest country among those studied), dense internal connectivity with regard to commodity flow, and constant importation from neighboring regions, all of which facilitate rapid propagation. Also, considering the performance gap between the LP-LB and SpreadBlocking solutions (see Fig. 2 and SI), BD would benefit most from rapid, country-wide interventions, as opposed to highly targeted approaches. Conversely, for the PH and ID, targeted interventions prove effective due to their more localized production structures. For Vietnam (VN network), production and demand are clustered into two distinct northern and southern regions. Since *P. absoluta* was first detected in the north in 2019 and is yet to be reported in other parts. Our experiments are inline with this observation as they show that it is relatively easy to isolate or delay the spread to the south (Fig. 5, first column).

Discussion

The importance of scenario-specific interventions

Our analysis of solutions for various budget–delay combinations highlight the importance of early discovery and prompt response. In the case of *P. absoluta*, this can be attributed to how our food systems are structured; concentrated production centers that are well-connected by human activities. It's ability to spread through multiple pathways enables rapid range expansion as it jumps from one locality to another in one or two production cycles, which has been observed in multiple regions globally (55, 56). In most countries, by the time the pest was discovered and acknowledged as a threat, it had already spread to multiple localities, and without coordinated efforts to control it, it spread further in the subsequent growing seasons (56–59). Our experiments provide useful insights for preparing for and controlling in regions such as North America, where *P. absoluta* is not yet reported. In the United States, for example, there is a large influx of tomato imports from Mexico and Canada. The production regions in the United States are concentrated in a few areas such as California and Florida. Our experiments suggest that early detection and rapid intervention in the production regions close to the points of entry can help prevent or delay the spread to the rest of the country. Also, it is critical to estimate the time it takes for the pest to spread from one production region to another.

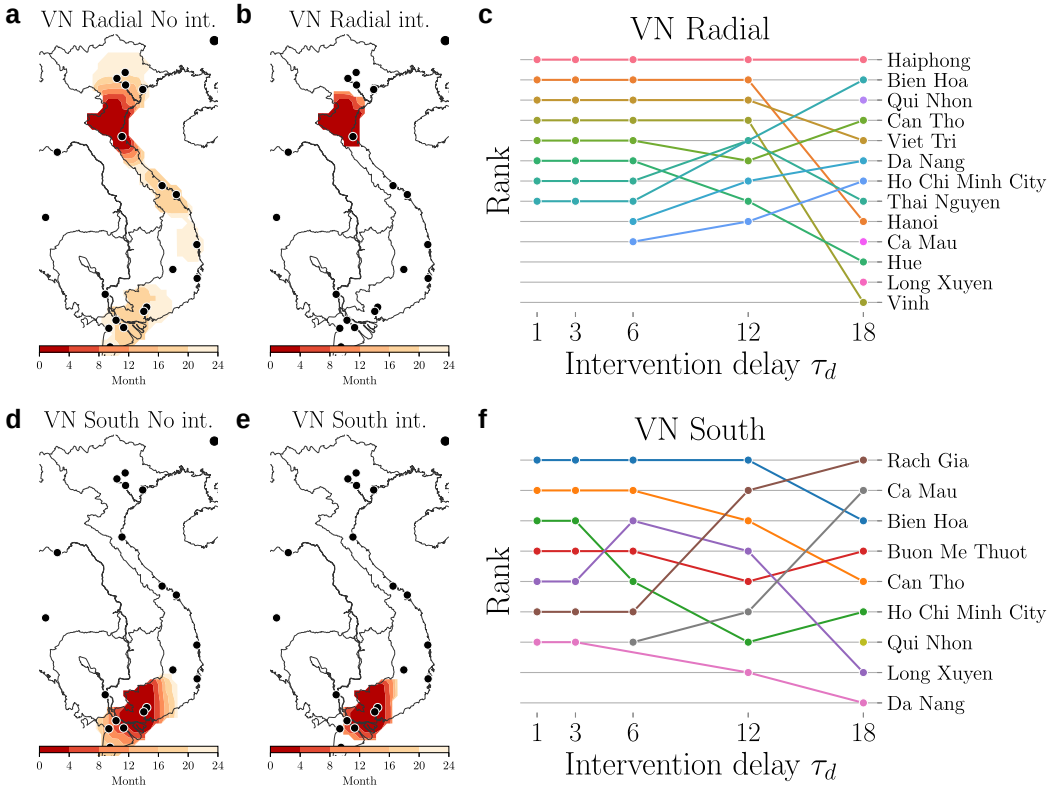

Fig. 5. Spread patterns and solution sets under different seeding scenarios. Each row corresponds to a seeding scenario. For a set of model parameters, plots (a) and (d) show the pattern of unmitigated spread, plots (b) and (e) show spread with intervention at $\tau_d = 12$ and $B = 4$ and 6, respectively, and plots (c) and (f) show ranking of groups for different B and τ_d values. The black dots are the localities. More results are provided in the SI.

Interventions literature

The control of epidemic processes on networks is an extensively studied topic. In ordinary differential equation models, interventions can be computed optimally, eg Medlock and Galvani (60). However, optimizing individual-based interventions in network SEIR models is much harder (5, 13, 16–18) (as also seen from our hardness results). Prior approaches do not immediately translate to solutions for the P&P context. Firstly, the multipathway models like those considered in this work (38, 43, 61) are different from simple diffusion models for which the strategies are developed. Secondly, the goal is to minimize the expected number of work (13, 39, 45). In order to address these gaps, our work significantly extends the techniques developed in earlier works (16, 18) in several aspects. Using the notion of time-expanded graphs to represent the simulation outcomes of the MultiPath model, we were able to apply the SAA approach to a broader class of SEIR(S) models. This approach naturally works for temporal networks, for which there are relatively very few works that address interventions (4). Also, to the best of our knowledge, group-scale interventions of the form proposed in this work have not been considered before.

P&P literature

Previous works on control and surveillance in the spread of invasive species have mostly relied on heuristics such as degree centrality and risk-based criteria (43, 44, 49, 50). In the context agricultural pests and pathogens, areas of large production of host crops (or livestock) are natural candidates according to this strategy. Our results show that while these characteristics are important, the efficacy of such mechanisms critically depends on factors such as the introduction scenario and the delay in intervention. For example, if the production areas are far apart, at any given time not every production area needs to be intervened at. However, when there is uncertainty in the introduction scenario (both space and time), the challenge is to come up with solutions that are effective for each of the possible scenarios. If prior probability distributions of scenarios can be estimated, it is straightforward to extend this work to account for uncertainty in model and introduction events by sampling cascade ensembles from the respective scenarios according to a prior probability distribution. In this context, notions of robustness can be incorporated into the framework. For example, one natural objective would be to minimize the worst-case expected size of infection across the different scenarios (62). In the SI, we provide a modified version of SpreadBlocking ILP that achieves this objective. Chades et al. (6) use a simple spatial networked representation of metapopulation of the target species to come up with simple rules for controlling spread. They develop strategies for various network motifs that are present in these representations.

Reactive control

Reactive control methods (21–24, 27, 29, 48) consider the setting of ongoing epidemics, deciding where and when to intervene based on observations of the current state of the system. Our work, on the other hand, is applicable to impending invasions or situations where the IAS has been recently introduced (accounting for intervention delays). Nevertheless, our results agree with the insights derived from this body of work. Firstly, these works stress on controlling at the frontier of invasion for optimal performance (21).

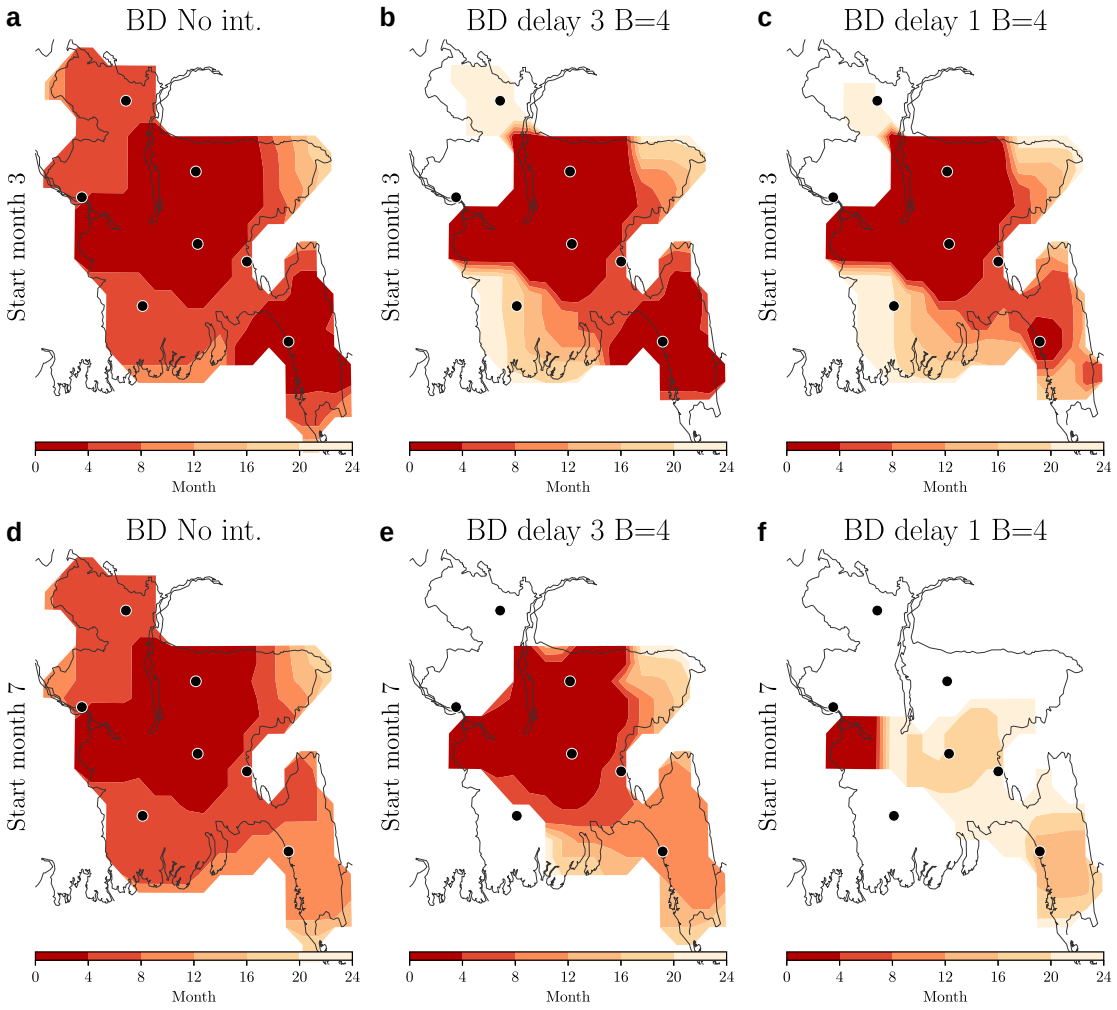

Fig. 6. Interventions under two seeding scenarios, which differ in the time of introduction of the P&P, but are identical otherwise. Plots (a) and (d) depict the patterns of unmitigated spread for one set of model parameters, plots (b) and (e) correspond to delay 3 and plots (c) and (f) correspond to delay 1. τ_d. Model parameters are $\alpha_s = 300$, $\alpha_\ell = 0.2$, and $\alpha_{\ell d} = 50$.

We arrive at a similar conclusion upon analysis of our solution sets, though, in a multipathway spread, the invasion wave front is not necessarily radial due to heterogeneity and multiscale nature of the spread (43, 63). Secondly, these works highlight the importance of delay in discovery and the need for a large budget if the control is delayed (22). Coupling with surveillance strategies and adaptive feedback-based control (64, 65) is an important direction to pursue.

Limitations and future work

Our work can be significantly extended in many ways. Fairness constraints can be included to prioritize the protection of certain groups or nodes (like production areas). We can easily extend the proposed work to incorporate jurisdictional restrictions by removing groups not under the jurisdiction from the candidate set of groups to intervene. Our experimental results highlight several implementation related challenges in addressing some of these extensions as the framework will be required to handle many cascades across scenarios. Heuristics such as pruning low vulnerability nodes can help mitigate this problem (18). The proposed method can be easily extended to multiple types of interventions that affect the spread at different spatio-temporal scales as long as these interventions correspond to removing a subset of nodes or edges. Our work does not consider efficacy of interventions. It is possible that not all nodes of a targeted group are intervened or still contain the pest due to reasons such as lack of compliance or deficiencies in the intervention method. In the setting of spread in human populations, Zhang et al. (66) considered group vaccination where resources are allocated to groups, but the principal does not have control over which individuals in the group get vaccinated. Extending our work to account for such scenarios is an important direction for future work. Despite these limitations, this work significantly contributes to the emerging body of literature (20, 25, 28, 67, 68) on applying high-resolution simulation models to design effective surveillance and control policies for complex spreading processes.

Methods

SIR model

The network-based discrete-time SIR (69) diffusion process is defined as follows. Each node is in one of the following states: **S**, **I**, or **R**. At any discrete time step t, each node u in state **I** infects its susceptible neighbor v with probability equal to the weight $w(u, v, \lambda, t)$ via the labeled edge (u, v, λ). Then, u transitions to state **R** at time $t + 1$.

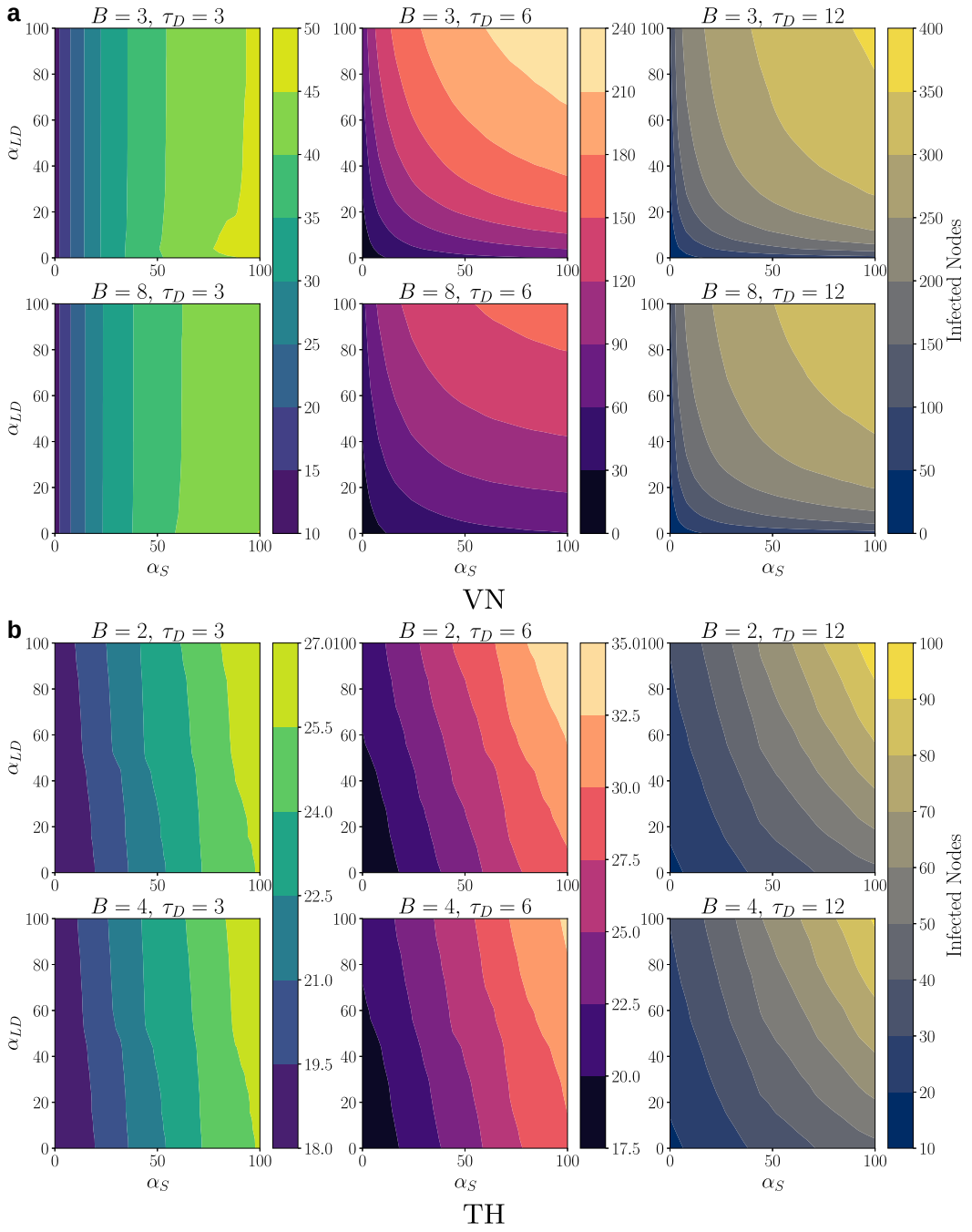

Fig. 7. The influence of the different pathways on the efficacy of intervention solutions. Here, the number of infections corresponds to the LP relaxation, LP-LB. Results for two networks are provided: (a) VN and (b) TH. More results are in the SI.

The MultiPath model

The model developed in McNitt et al. (43) referred to as the MultiPath model is as follows. The study region is divided into cells, which correspond to the set of nodes V of the spatial network. There are groups of spatially contiguous nodes (called localities), which represent regions of major supply of host crops and demand. Many nodes do not belong to any locality. These are not candidates for group-scale intervention. There are three pathways of spread. Self-mediated dispersal corresponds to diffusion from one cell to its adjacent cells, local human-mediated dispersal is diffusion within a group (farmer-market interactions), and long-distance dispersal corresponds to diffusion from cells from one group to another (trade). The diffusion model is a discrete-time SEI process where a node transitions from **E** to **I** after ℓ time steps, where ℓ is the latency period. A node has two periodic time-varying attributes, suitability $\epsilon(v, t)$ for pest establishment and infectivity $\rho(v, t)$. Edges are directed and labeled, and have time-varying weights that determine the probability of infection from source to target. We use the notation $w(u, v, \lambda, t)$ to denote the weight of the edge from u to v at time t corresponding to the pathway λ. The probability that a node can be infected ($\mathbf{S} \rightarrow \mathbf{E}$) through a pathway is modeled as a negative exponential function of infectivity, pathway parameters, and edge weights. These details are provided in the SI.

The SpreadBlocking$^+$ heuristic

First, we compute solutions to SpreadBlocking for the desired set of budgets $\mathcal{B} = \{B_{\min}, \ldots, B_{\max}\}$ for a fixed delay τ_d. We denote by $\mathcal{Q}_{SB}(B)$ the solution obtained for an input budget of B. Note that $|\mathcal{Q}_{SB}(B)| \geq B$ due to the bicriteria approximation. Next, we compute the *marginal importance* of each group q by computing a score $\sum_{B' \in \mathcal{B}} \frac{x_{q,\tau_d}(B')}{B'}$, where $x_{q,\tau_d}(B')$ is the value of the variable corresponding to q from the LP solution for budget B'. For a target budget B, we choose the largest solution whose size is at most B. If no such solution exists, we set it to the empty set. Then, we add the remaining groups in the decreasing order of their marginal importance until the budget is met. This is denoted as $\mathcal{Q}^+_{SB}(B)$. The initial set for intervention is chosen as the solution obtained for a smaller budget, as this provides a strong preliminary combination of target groups. This baseline is then refined by adding marginally important groups from the remaining set to meet the exact budget constraint.

Datasets

The networks along with their properties are listed in the SI. These were constructed by McNitt et al. (43) and are publicly available. There are several versions of the networks depending on the gravity model parameters. We used the values 2 and 500 for the distance function exponent and cut-off, respectively. These are among the best model parameters obtained after calibration in their work. Each network has groups containing on an average 20–30 nodes capturing key urban and producing areas. For most countries, a significant portion of the nodes do not belong to any group. However, these nodes together cover <20% of the total production and population in each country. The list of networks and their attributes are provided in a Table in the SI.

Experimental setup

The range of values for each parameter of the multipathway model were chosen to cover the best models with highest fit to ground truth (43). We used a full factorial design for each network with pathway parameters $\alpha_s \in [300, 500]$, $\alpha_\ell \in [0, 0.2]$, $\alpha_{\ell d} \in [50, 200]$, Moore range $r_M = 1$, start month = 5 and 100 replicates per simulation (unless explicitly specified). For the comparison with baselines, each simulation was run for $T = 24$ time steps corresponding to a time horizon of 2 years. The seeding scenarios were picked from McNitt et al. (43). We also added a few more for counterfactual experiments. For performance evaluation, we used mean number of infections across simulation instances as the metric for the time horizon. More information, including our computing environment, is given in the SI.

Baselines

Here, we briefly describe the baselines used in the experimentation. The *eigenvector centrality* algorithm computes the influence of nodes in the network—high score of a node indicates its connection to many nodes that have high scores. The algorithm takes the network G and returns the top B nodes with highest eigencentrality scores (14). The *personalized pagerank algorithm* (70–72) takes a network G, the seeds of infection S, and budget B as input and returns the set of B nodes with highest pagerank score (personalized for the seeds S). The *maximum outflow* corresponds to ordering groups by *outflow* in the group-to-group network (similar to degree-based method for undirected graphs), and picking the top B nodes. This method is often applied in the invasive species literature (43, 44). In this case, we considered the annual outflow by aggregating the outflows across different months. The second method corresponds to ordering groups by *vulnerability* (18, 73). We observed number of nodes infected in each group at time step 12 and picked B most vulnerable groups.

Rounding scheme

In SpreadBlocking algorithm, the x_{q,τ_d} variables in the fractional optimal solution to LP_{τ_d} are scaled by a factor $2g_m$, and rounded to 1 if the result after scaling is at least 1. The term g_m corresponds to the maximum number of groups associated with any path in any H^j. To obtain the value of g_m, we use a dynamic programming approach leveraging the fact that each cascade is a directed acyclic graph. The algorithm is provided in the SI along with a computational complexity analysis.

Computation time and scalability

The size of the ILP instance depends on the number of simulations and cascade size, which in turn depends on network size, model parameters, and the time horizon. We also note that the value of the budget—intervention delay adds to the complexity as more solution sets are possible. We used a high performance computing environment for the pipeline, where different scenario instances were run in parallel. More details on the computing environment and implementation are presented in the SI.

Acknowledgments

We sincerely thank the reviewers for their insightful comments. The suggestions made by the reviewers have substantially improved the article. Their comments led to a more practical version of the algorithm, and a more comprehensive analysis of the results. We thank members of the Biocomplexity Institute and UVA Research Computing for their support.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

This work was supported in part by United States Department of Agriculture, National Institute of Food and Agriculture, grant no. 2019-67021-29933 and grant no. 2021-67021-35344, United States Agency for International Development, under the Cooperative Agreement no. AID-OAA-L-15-00001, UVA Strategic Investment Fund SIF160, NSF Expeditions in Computing Grant CCF-1918656, and OAC-1916805 (CINES).

Author Contributions

A.A. and A.V. defined the scope of the research. A.V., P.S., and A.A. performed the theoretical analysis. A.A., P.S., M.S., H.C., and A.V. conceived and designed the experiments. P.S., M.S., H.C., and A.A. performed the analysis. M.R. and A.V. provided assistance in interpreting the results. A.A., P.S., M.S., H.C., and A.V. wrote the article with significant input from M.R. A.A. and A.V. supervised the research. All authors discussed the results and commented on the manuscript.

Previous Presentation

These results have not been previously presented elsewhere.

Data Availability

The data and code underlying this article are made available in https://github.com/NSSAC/SPREAD_interventions_public.

References

1 Easley D, Kleinberg J. *Networks, crowds, and markets: reasoning about a highly connected world*. Vol. 1. Cambridge University Press, 2010.
2 Keeling MJ, Eames KTD. 2005. Networks and epidemic models. *J R Soc Interface*. 2(4):295–307.
3 Newman MEJ. 2003. The structure and function of complex networks. *SIAM Rev*. 45(2):167–256.
4 Zino L, Cao M. 2021. Analysis, prediction, and control of epidemics: a survey from scalar to dynamic network models. *IEEE Circuits Syst Mag*. 21(4):4–23.
5 Nowzari C, Preciado VM, Pappas GJ. 2016. Analysis and control of epidemics: a survey of spreading processes on complex networks. *IEEE Control Syst Mag*. 36(1):26–46.
6 Chadès I, *et al.* 2011. General rules for managing and surveying networks of pests, diseases, and endangered species. *Proc Natl Acad Sci U S A*. 108(20):8323–8328.
7 Epanchin-Niell RS. 2017. Economics of invasive species policy and management. *Biol Invasions*. 19(11):3333–3354.
8 Eppinga MB, *et al.* 2021. Spatially explicit removal strategies increase the efficiency of invasive plant species control. *Ecol Appl*. 31(3):e02257.
9 Giljohann KM, Hauser CE, Williams NSG, Moore JL. 2011. Optimizing invasive species control across space: willow invasion management in the Australian Alps. *J Appl Ecol*. 48(5):1286–1294.
10 Hayrapetyan A, Kempe D, Pál M, Svitkina Z. Unbalanced graph cuts. In: *Proceedings of the 13th Annual European Conference on Algorithms, ESA'05*. Springer, Berlin, Heidelberg; 2005. p. 191–202.
11 Li D, Eliassi-Rad T, Zhang HR. Optimal intervention on weighted networks via edge centrality. In: *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*. SIAM; 2023. p. 424–432.
12 Miller JC, Hyman JM. 2007. Effective vaccination strategies for realistic social networks. *Phys Stat Mech Appl*. 386(2):780–785.
13 Saha S, Adiga A, Prakash BA, Vullikanti AKS. Approximation algorithms for reducing the spectral radius to control epidemic spread. In: *Proceedings of the 2015 SIAM International Conference on Data Mining*. SIAM; 2015. p. 568–576.
14 Tong H, Prakash BA, Eliassi-Rad T, Faloutsos M, Faloutsos C. Gelling, and melting, large graphs by edge manipulation. In: *Proceedings of the 21st ACM International Conference on Information and knowledge Management*. ACM; 2012. p. 245–254.
15 Van Mieghem P, *et al.* 2011. Decreasing the spectral radius of a graph by link removals. *Phys Rev E Stat Nonlin Soft Matter Phys*. 84(1):016101.
16 Babay AE, Dinitz M, Srinivasan A, Tsepenekas L, Vullikanti A. Controlling epidemic spread using probabilistic diffusion models on networks. In: *International Conference on Artificial Intelligence and Statistics*. PMLR; 2022. p. 11641–11654.
17 Borgs C, Chayes J, Ganesh A, Saberi A. 2010. How to distribute antidote to control epidemics. *Random Struct Algorithms*. 37(2):204–222.
18 Sambaturu P, Adhikari B, Prakash BA, Venkatramanan S, Vullikanti A. Designing effective and practical interventions to contain epidemics. In: *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. IFAAMAS; 2020. p. 1187–1195.
19 Wilder B, Suen S-C, Tambe M. Preventing infectious disease in dynamic populations under uncertainty. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. AAAI Press; 2018.
20 Aleta A, *et al.* 2020. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat Hum Behav*. 4(9):964–971.
21 Cunniffe NJ, Cobb RC, Meentemeyer RK, Rizzo DM, Gilligan CA. 2016. Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in California. *Proc Natl Acad Sci U S A*. 113(20):5640–5645.
22 Cunniffe NJ, Stutt ROJH, DeSimone RE, Gottwald TR, Gilligan CA. 2015. Optimising and communicating options for the control of invasive plant disease when there is epidemiological uncertainty. *PLoS Comput Biol*. 11(4):e1004211.
23 Fabre F, Coville J, Cunniffe NJ. Optimising reactive disease management using spatially explicit models at the landscape scale. In: *Plant diseases and food security in the 21st century*. Springer; 2021. p. 47–72.
24 Fang H, Caton BP, Manoukis NC, Pallipparambil GR. 2022. Simulation-based evaluation of two insect trapping grids for delimitation surveys. *Sci Rep*. 12(1):11089.
25 Ferretti L, *et al.* 2020. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*. 368(6491):eabb6936.
26 Hoops S, *et al.* High performance agent-based modeling to study realistic contact tracing protocols. In: *2021 Winter Simulation Conference (WSC)*. IEEE; 2021. p. 1–12.
27 Keeling MJ, Woolhouse MEJ, May RM, Davies G, Grenfell BT. 2003. Modelling vaccination strategies against foot-and-mouth disease. *Nature*. 421(6919):136–142.
28 Kerr CC, *et al.* 2021. Covasim: an agent-based model of COVID-19 dynamics and interventions. *PLoS Comput Biol*. 17(7):e1009149.
29 Tildesley MJ, *et al.* 2006. Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. *Nature*. 440(7080):83–86.
30 Perry B, Grace D. 2009. The impacts of livestock diseases and their control on growth and development processes that are propoor. *Philos Trans R Soc Lond B Biol Sci*. 364(1530):2643–2655.
31 Pimentel D. *Biological invasions: economic and environmental costs of alien plant, animal, and microbe species*. CRC Press, USA, 2011.
32 Pimentel D, Zuniga R, Morrison D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol Econ*. 52(3):273–288.
33 Roy HE, *et al.* 2023. IPBES invasive alien species assessment: summary for policymakers.
34 Savary S, *et al.* 2019. The global burden of pathogens and pests on major food crops. *Nat Ecol Evol*. 3(3):430–439.
35 Hulme PE. 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J Appl Ecol*. 46(1):10–18.
36 Bajardi P, Barrat A, Savini L, Colizza V. 2012. Optimizing surveillance for livestock disease spreading through animal movements. *J R Soc Interface*. 9(76):2814–2825.
37 Cabezas AH, Sanderson MW, Lockhart CY, Riley KA, Hanthorn CJ. 2021. Spatial and network analysis of US livestock movements based on interstate certificates of veterinary inspection. *Prev Vet Med*. 193:105391.
38 Carrasco LR, *et al.* 2010. Unveiling human-assisted dispersal mechanisms in invasive alien insects: integration of spatial stochastic simulation and phenology models. *Ecol Modell*. 221(17):2068–2075.
39 Ercsey-Ravasz M, Toroczkai Z, Lakner Z, Baranyi J. 2012. Complexity of the international agro-food trade network and its impact on food safety. *PLoS One*. 7(5):e37810.

40 Galvis JA, Jones CM, Prada JM, Corzo CA, Machado G. 2022. The between-farm transmission dynamics of porcine epidemic diarrhoea virus: a short-term forecast modelling comparison and the effectiveness of control strategies. *Transbound Emerg Dis*. 69(2): 396–412.

41 Kim Y, *et al.* 2018. Livestock trade network: potential for disease transmission and implications for risk-based surveillance on the island of Mayotte. *Sci Rep*. 8(1):1–10.

42 Marquetoux N, Stevenson MA, Wilson P, Ridler A, Heuer C. 2016. Using social network analysis to inform disease control interventions. *Prev Vet Med*. 126(5721):94–104.

43 McNitt J, *et al.* 2019. Assessing the multi-pathway threat from an invasive agricultural pest: *Tuta absoluta* in Asia. *Proc Biol Sci*. 286(1913):20191159.

44 Nopsa JFH, *et al.* 2015. Ecological networks in stored grain: key postharvest nodes for emerging pests, pathogens, and mycotoxins. *Bioscience*. 65(10):biv122.

45 Tatem AJ. 2009. The worldwide airline network and the dispersal of exotic species: 2007–2010. *Ecography*. 32(1):94–102.

46 De Vos CJ, *et al.* 2020. Cross-validation of generic risk assessment tools for animal disease incursion based on a case study for African swine fever. *Front Vet Sci*. 7:56.

47 ISPM 2. Framework for pest risk analysis. In: *International plant protection convention*. FAO; 2007. p. 8–20.

48 Ferguson NM, Donnelly CA, Anderson RM. 2001. The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science*. 292(5519):1155–1160.

49 Hyatt-Twynam SR, *et al.* 2017. Risk-based management of invading plant disease. *New Phytol*. 214(3):1317–1329.

50 Sutrave S, Scoglio C, Isard SA, Hutchinson JMS, Garrett KA. 2012. Identifying highly connected counties compensates for resource limitations when evaluating national spread of an invasive pathogen. *PLoS One*. 7(6):e37793.

51 Cunniffe NJ, *et al.* 2015. Thirteen challenges in modelling plant diseases. *Epidemics*. 10(Suppl. 5):6–10.

52 Kim S, Pasupathy R, Henderson SG. A guide to sample average approximation. In: *Handbook of simulation optimization*. Springer; 2015. p. 207–243.

53 Shapiro A. Monte Carlo sampling methods. In: *Handbooks in operations research and management science*. Vol. 10(2). Elsevier; 2003. p. 353–425.

54 Shapiro A, Dentcheva D, Ruszczynski A. *Lectures on stochastic programming: modeling and theory*. SIAM, 2021.

55 Biondi A, Guedes RNC, Wan F-H, Desneux N. 2018. Ecology, worldwide spread, and management of the invasive South American tomato pinworm, *Tuta absoluta*: past, present, and future. *Annu Rev Entomol*. 63(1):239–258.

56 Campos MR, Biondi A, Adiga A, Guedes RNC, Desneux N. 2017. From the Western Palaearctic region to beyond: *Tuta absoluta* 10 years after invading Europe. *J Pest Sci*. 90(3):787–796.

57 Bajracharya ASR, *et al.* 2016. The first record of South American tomato leaf miner, *Tuta absoluta* (Meyrick 1917) (Lepidoptera: Gelechiidae) in Nepal. *J Entomol Zool Stud*. 4(4):1359–1363.

58 Brévault T, Bernadas G, Sylla S, Diatte M, Diarra K. 2014. *Tuta absoluta* Meyrick (Lepidoptera: Gelechiidae): a new threat to tomato production in sub-Saharan Africas. *Afr Entomol*. 22(2):441–444.

59 Karadjova O, Ilieva Z, Krumov V, Petrova E, Ventsislavov V. 2013. *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae): potential for entry, establishment and spread in Bulgaria. *Bulg J Agric Sci*. 19(3):563–571.

60 Medlock J, Galvani AP. 2009. Optimizing influenza vaccine distribution. *Science*. 325(5948):1705–1708.

61 Stanaway MA, Reeves R, Mengersen KL. 2011. Hierarchical Bayesian modelling of plant pest invasions with human-mediated dispersal. *Ecol Modell*. 222(19):3531–3540.

62 Ben-Tal A, Ghaoui LE, Nemirovski A. *Robust optimization*. Vol. 28. Princeton University Press, 2009.

63 Brockmann D, Helbing D. 2013. The hidden geometry of complex, network-driven contagion phenomena. *Science*. 342(6164): 1337–1342.

64 Köhler J, *et al.* 2021. Robust and optimal predictive control of the COVID-19 outbreak. *Annu Rev Control*. 51(17):525–539.

65 Probert WJM, *et al.* 2019. Context matters: using reinforcement learning to develop human-readable, state-dependent outbreak response policies. *Philos Trans R Soc B*. 374(1776): 20180277.

66 Zhang Y, Adiga A, Saha S, Vullikanti A, Prakash BA. 2016. Near-optimal algorithms for controlling propagation at group scale on networks. *IEEE Trans Knowl Data Eng*. 28(12):3339–3352.

67 Bhattacharya P, *et al.* 2024. Novel multi-cluster workflow system to support real-time HPC-enabled epidemic science: investigating the impact of vaccine acceptance on covid-19 spread. *J Parallel Distrib Comput*. 191:104899.

68 Bussell EH, Dangerfield CE, Gilligan CA, Cunniffe NJ. 2019. Applying optimal control theory to complex epidemiological models to inform real-world disease management. *Philos Trans R Soc B*. 374(1776):20180284.

69 Marathe M, Vullikanti AKS. 2013. Computational epidemiology. *Commun ACM*. 56(7):88–96.

70 Kamvar SD, Haveliwala TH, Manning CD, Golub GH. Extrapolation methods for accelerating pagerank computations. In: *Proceedings of the 12th International Conference on World Wide Web*. ACM; 2003. p. 261–270.

71 Page L, Brin S, Motwani R, Winograd T. *The PageRank citation ranking: bringing order to the web*. Stanford infolab, 1999.

72 Park S, Lee W, Choe B, Lee S-G. 2019. A survey on personalized pagerank computation algorithms. *IEEE Access*. 7:163049–163062.

73 Qu B, Li C, Van Mieghem P, Wang H. 2017. Ranking of nodal infection probability in susceptible-infected-susceptible epidemic. *Sci Rep*. 7(1):9233.