Network Models and Simulation Analytics for Multi-scale Dynamics of Biological Invasions

Abhijin Adiga[1]*, Nicholas Palmer[1], Young Yun Baek[1], Henning Mortveit[1,2] and S. S. Ravi[1,3]

[1] Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, VA, United States, [2] Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, United States, [3] Department of Computer Science, University at Albany—SUNY, Albany, NY, United States

Globalization and climate change facilitate the spread and establishment of invasive species throughout the world via multiple pathways. These spread mechanisms can be effectively represented as diffusion processes on multi-scale, spatial networks. Such network-based modeling and simulation approaches are being increasingly applied in this domain. However, these works tend to be largely domain-specific, lacking any graph theoretic formalisms, and do not take advantage of more recent developments in network science. This work is aimed toward filling some of these gaps. We develop a generic multi-scale spatial network framework that is applicable to a wide range of models developed in the literature on biological invasions. A key question we address is the following: how do individual pathways and their combinations influence the rate and pattern of spread? The analytical complexity arises more from the multi-scale nature and complex functional components of the networks rather than from the sizes of the networks. We present theoretical bounds on the spectral radius and the diameter of multi-scale networks. These two structural graph parameters have established connections to diffusion processes. Specifically, we study how network properties, such as spectral radius and diameter are influenced by model parameters. Further, we analyze a multi-pathway diffusion model from the literature by conducting simulations on synthetic and real-world networks and then use regression tree analysis to identify the important network and diffusion model parameters that influence the dynamics.

Keywords: invasive species, network models, diffusion processes, multi-scale networks, simulation analytics, spatial networks

1. INTRODUCTION

1.1. Background and Motivation

Many natural- and engineered complex systems can be represented as systems of interconnected networks (Buldyrev et al., 2010; Barrett et al., 2013; Bashan et al., 2013). Such representations are very effective at capturing relationships of different types and at different scales between the entities comprising the complex system. There are several classes of such networks depending on the application, including multi-layer networks, multiplex networks, and network of networks, to name just a few. We refer the reader to Kivelä et al. (2014) for a comprehensive list. Here, we study modeling of complex networks arising in the context of multi-scale spatial dynamical processes,

such as biological invasions, spread of infectious diseases, and built infrastructure (Balcan et al., 2009; Serrano et al., 2009; Walpole et al., 2013), our focus mainly being on the properties of dynamical processes over such networks.

In recent years, there has been a big thrust toward developing cyberinfrastructures to address problems related to resilient and sustainable agriculture (USDA-NSF, 2017, 2021; Microsoft, 2020). One of the greatest threats to biodiversity and food security is the increasing rate of biological invasions (Pimentel et al., 2005; Crowl et al., 2008; Pyšek and Richardson, 2010). The annual economic costs arising from environmental damages and the losses caused by invasive species run into billions of dollars (Diagne et al., 2021). Understanding the role of natural processes and human activities in the spatio-temporal spread of biological invasions is of utmost importance (Cunniffe et al., 2015).

A variety of models for biological invasions have been proposed with the goal of studying these complex phenomena at appropriate levels of resolution and fidelity (Pitt et al., 2009; Carrasco et al., 2010; Smolik et al., 2010; Fitzpatrick et al., 2012; Robinet et al., 2012; Ferrari et al., 2014; Chapman et al., 2015; Douma et al., 2016; McNitt et al., 2019; Venkatramanan et al., 2020). In many of these works, biological invasions have been viewed as propagation processes over multi-scale networks. As depicted in **Figure 1**, there are various pathways or modes of dispersal. The spreading ability of the invasive organism, environmental factors (e.g., winds, ocean currents, and suitable environment), and anthropogenic factors (e.g., production and trade of host crops, and tourism) are among the primary pathways. These pathways affect the spread at multiple spatial and temporal scales. For example, the rate and pattern of self-mediated spread can be significantly different from that of human-assisted spread (Davis et al., 2001; Hoffmann and Courchamp, 2016). The functions or mechanisms that govern the process can vary within and across scales; the flying ability of a pest determines how far it can naturally spread, while trade of host crops and plant material can facilitate long distance spread. Accordingly, there are potentially several ways to control the phenomenon ranging from application of pesticides (farm-level) to trade restrictions (administrative-level).

While the models referenced above have been designed and analyzed from the perspective of biological invasions, there has been very little work on exploring the connection to well-known formal frameworks developed in network science. A typical modeling effort uses networks to represent movements of species based on data or assumptions regarding the processes involved (or both). The resulting networks are analyzed using basic structural properties such as degree distribution and betweens centrality. Further, suitable diffusion models are used to simulate the spread over this network. It is important to understand these models from a network science perspective in order to (i) make robust predictions, (ii) rigorously address key aspects such as calibration, validation, verification, sensitivity analysis, and uncertainty quantification, and (iii) leverage state-of-the-art algorithms to address important problems pertaining to control, monitoring, and various inference problems in the field of network dynamical systems. The aim of this work is to

FIGURE 1 | A multi-pathway network model of a spatial diffusion process in the context of biological invasions. The spread takes place at multiple spatial scales, such as (i) one cell to adjacent cells due to self-mediated spread, (ii) within a locality of high human activity, and (iii) long distance jumps facilitated by trade and travel.
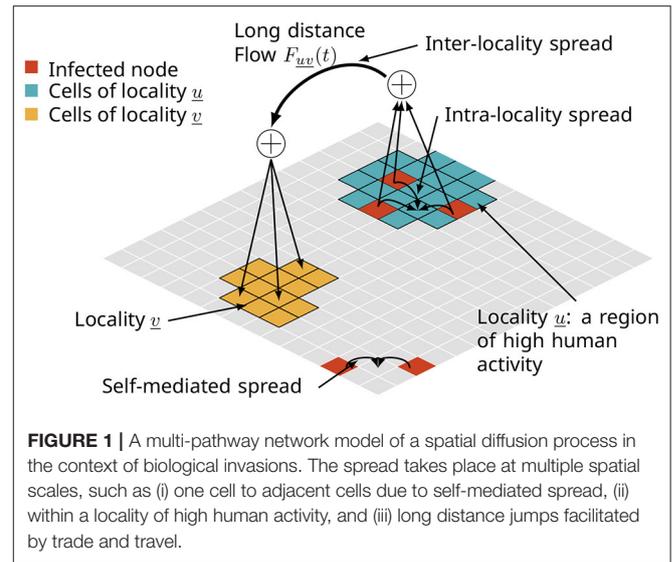
provide the first steps in developing the foundations for these models. Through such formal grounding and abstractions one may also be able to take advantage of existing mathematical and computational theory, including computational paradigms facilitating efficient mapping of models onto high-performance software implementations that exploit modern hardware.

In this paper, we develop an abstract model for multi-pathway spatial processes. For this model, we characterize complex spatial diffusion processes using structural properties of the underlying network. Next, we provide an analysis of the roles of the different pathways on the rate and pattern of spread by applying a complex multi-pathway diffusion model proposed in McNitt et al. (2019), which we refer to as SPREAD (which stands for Simulation tools for Pest Risk analysis accounting for Ecological and Anthropogenic Drivers). Finally, we provide analyses of structural and dynamical properties of networks through the lens of machine learning algorithms to identify the primary drivers of spread. A more detailed description of our contributions is as follows.

1.2. Contributions
1.2.1. A Model for the Multi-Pathway Spatial Network
We define a grid-based model called MPSN for synthetic spatial networks which captures the multi-scale nature of the spread process. The synthetic network is composed of nodes of a grid and metanodes called localities, each of which consists of a unique set of contiguous grid nodes. Its edge set is the union of the edge sets of three different graphs, each corresponding to a different pathway of spread as in **Figure 1**. The main purpose of proposing such a model is to bridge the gap between simple static network models like Erdős-Rényi or Chung-Lu graphs, and complex real-world networks that describe the interactions in the biological invasion process. These networks, while being significantly more complex than standard models, are still amenable to theoretical and experimental analyses as we demonstrate in this work.

1.2.2. Theoretical Results on Spectral Radius and Diameter

We derive bounds on the spectral radius and diameter of MPSN model. These two graph invariants are important indicators of the diffusion dynamics. The bounds highlight the importance of inter-locality and intra-locality components of the network, which represent the human factors in the spread.

1.2.3. Experimental Analysis of Synthetic and Real-World Networks

We report results from extensive experiments conducted on hundreds of synthetic networks using the MPSN model and several real-world networks. First, we study how the MPSN model parameters influence the spectral radius and diameter. This is followed by running simulations for varying pathway probabilities. In both analyses, to understand the influence of model parameters (i.e., parameters of the network and of the diffusion process) on the extent of spread, we view it as a supervised learning problem. Here the feature vector is comprised of model parameters and the observed variable is a graph invariant (in the structural analysis) or a simulation output (in the dynamical analysis), and we use regression trees and random forests to identify the primary drivers of diffusion process. Our experimental analyses help to identify the regimes where long distance edges and locality size are highly influential in increasing the spread.

2. RELATED WORK

2.1. Networked Representations of Invasive Species Spread

The current state-of-the-art for modeling invasive species involves developing risk maps using ecological niche models (Venette et al., 2010). Such models account for climate and biology of the invasive species and its hosts to map the long-term establishment potential. They do not provide a causal explanation to the extent and dispersion of spread or explicitly account for human-mediated pathways. However, in recent years, network diffusion models are being increasingly applied to model the spread dynamics of invasive species in order to account for their long distance spread. Hernandez Nopsa et al. (2015) studied the structure of rail networks for grain transport in the United States and Eastern Australia to identify the shortest paths for the anthropogenic dispersal of pests and mycotoxins, as well as the major sources, sinks, and bridges for movement. Sutrave et al. (2012) used an SI model (Easley and Kleinberg, 2010) to county-to-county network to model wind speed and direction, and host density to identify locations to monitor soybean rust, a pathogen. Koch et al. (2014) assess the risk of forest pests due to camping activities. Venkatramanan et al. (2020) used a Bayesian inference method to identify the most likely spread pattern of an invasive pest by modeling the spread as a diffusion process on a time varying network.

2.2. Complex Multi-Pathway Dispersal Models

Carrasco et al. (2010) considered both local and long distance spread in a process-based spatially explicit simulation model to study pest of the maize crop. They use phenology models to estimate pest population size, a negative power law kernel for self-mediated spread, and a gravity model representation of long distance edges. Our work is based on a similar approach by McNitt et al. (2019) who model the multi-pathway spread by accounting for self-mediated spread and spread within and between areas of high human activity (e.g., urban areas). Both works account for suitability of establishment and the distribution of host crop production. Similar modeling approaches have been applied to study infectious diseases in humans and livestock (Ajelli et al., 2010; Kim et al., 2018; Venkatramanan et al., 2021).

2.3. Spectral Characterization of Network Dynamics

Several structural properties of networks have been used to understand the progression of a diffusion process. These include basic properties such as degree distribution or clustering coefficient to other properties such as graph spectrum, diameter, and degeneracy, to name a few. The **spectrum** of a graph is the set of eigenvalues of its adjacency matrix. There are several works that relate spectrum, particularly the first eigenvalue or *spectral radius* $\lambda_1(G)$ of the adjacency matrix of a graph G, to disease spread in SEIR-like epidemic models (Ganesh et al., 2005; Prakash et al., 2012). A well known result that highlights the impact of the network structure on the dynamics is the following: an epidemic dies out "quickly" if $\lambda_1(G) \leq T$, where T is a threshold that depends on the disease model. This relationship has motivated a number of works on epidemic control where the objective is to find an optimal set of nodes (or edges) to remove from the network that leads to the maximum reduction in its spectral radius (Van Mieghem et al., 2011; Saha et al., 2015; Zhang et al., 2016; Chen et al., 2021).

2.4. Diameter and Network Dynamics

The diameter of real-world networks is an important structural parameter used to characterize epidemics on real-world graphs (Holme, 2013; Pastor-Satorras et al., 2015). In the literature, average path lengths between pairs of nodes and diameter of the network have been observed to have an effect on the rate of diffusion. In particular, the lower the diameter, the higher is the diffusion rate (Banos et al., 2015; Taghvaei et al., 2020; Kamra et al., 2021). In spatial networks, the diameter tends to be large when compared to social networks that exhibit the small-world effect (Watts and Strogatz, 1998), and long distance edges can be responsible for bringing the diameter down (Barthélemy, 2011).

2.5. Machine Learning and Simulation Systems

From the works described above, it is clear that even simple SIR-like processes (Easley and Kleinberg, 2010) on arbitrary static networks are difficult to characterize. In recent years, there have been several studies that use a combination of extensive simulations and machine learning algorithms to understand the phase space of complex models. Fox et al. (2019) explored multiple ways in which such a nexus of the two modeling approaches can be used to understand complex systems. Lamperti et al. (2018) used extreme gradient boosted trees for the purpose of phase space exploration of a complex model. Our approach to apply classification and regression trees (CART) and random forests (Breiman, 2001, 2017) is motivated by this work. Other methods based on Gaussian emulators have been used for calibrating agent-based models (Fadikar et al., 2018).

3. PRELIMINARIES

3.1. Graph Theoretic Concepts

We begin with the definitions of several basic graph theoretic concepts. Additional information regarding these concepts can be found in West (2006); Brouwer and Haemers (2012). We assume that graphs are simple (i.e., they have no multi-edges or self loops).

Given an undirected graph $G(V, E)$, where $V = \{v_1, v_2, \ldots, v_n\}$, the **adjacency matrix** A_G is an $n \times n$ (symmetric) matrix defined as follows.

$$A_G[i,j] = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } v_i \text{ and } v_j \text{ are adjacent and} \\ 0 & \text{otherwise.} \end{cases}$$

We use $\text{degree}(v_i)$ to denote the degree of node v_i. Also, we use $\delta(G)$ and $\Delta(G)$, respectively, to denote the minimum and maximum node degrees in G.

Consider a connected undirected graph $G(V, E)$ with a non-negative weight $w(e)$ on each edge $e \in E$. When $w(e) = 1$ for each $e \in E$, we sometimes refer to G as an *unweighted* graph. A **shortest path** between any pair of nodes v_i and v_j is a path such that the total weight of all the edges in the path is a minimum among all the paths between v_i and v_j. Let $d_G(v_i, v_j)$ denote the length of a shortest path between v_i and v_j. Then, the **diameter** of G is defined by

$$\text{diam}(G) = \max\{d_G(v_i, v_j) : v_i, v_j \in V\}.$$

If G is not connected, then by convention, $\text{diam}(G)$ is taken as ∞.

Given an undirected graph $G(V, E)$ and an integer $t \geq 1$, the **t-th** power of G, denoted by G^t, is an undirected graph $G^t(V, E^t)$, where $\{v_i, v_j\} \in E^t$ iff there is a path with at most t edges between v_i and v_j in G. Given two graphs $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$, the **Kronecker product** of G_1 and G_2, denoted by $G_1 \times G_2$, is the graph $G'(V', E')$, where $V' = V_1 \times V_2$ and an edge $\{(a, b), (c, d)\}$ is in E' iff $\{a, c\} \in E_1$ and $\{b, d\} \in E_2$.

3.2. Matrix Concepts

For any $n \times n$ symmetric matrix M with real entries, it is known that all the n eigen values, denoted by $\lambda_1, \lambda_2, \ldots, \lambda_n$, are real (Brouwer and Haemers, 2012). Since the $n \times n$ adjacency matrix of an undirected graph with n nodes is symmetric and all its entries are from {0,1}, it follows that all the eigen values of the adjacency matrix are real. We will assume without loss of generality that these n eigen values are ordered so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Thus, λ_1 will be referred to as the **first** eigen value or the **spectral radius**.

3.3. Geometric Concepts

For any pair of points $a = \langle a_x, a_y \rangle$ and $b = \langle b_x, b_y \rangle$ in the 2D-plane, let $d_{\text{Euc}}(a, b) = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2}$ denote the Euclidean distance between a and b. It is well known that the Euclidean distance satisfies the **triangle inequality**: for any three points a, b, and c, $d_{\text{Euc}}(a, b) + d_{\text{Euc}}(b, c) \geq d_{\text{Euc}}(a, c)$. As a generalization of this inequality, we have the following fact: if a_1, $a_2, \ldots, a_n$ are $n \geq 3$ points in the plane, then $\sum_{i=1}^{n-1} d_{\text{Euc}}(a_i, a_{i+1}) \geq d_{\text{Euc}}(a_1, a_n)$. We will use this inequality in Section 6.

4. MULTI-PATHWAY SPATIAL NETWORK

4.1. Motivation

A number of models have been proposed for spatial networks, where nodes and edges are embedded in a d-dimensional space with some geometric constraints. Such a network model acts as a reference model for comparison with real-world networks, and is amenable to theoretical analysis and controlled experimentation. Network models are also a principled approach to approximate partially-known real-world networks and to capture their variability (Gutfraind et al., 2015). Barthélemy (2011) provides a comprehensive review of such models. Lattice graphs and random geometric graphs are examples of some of the simplest spatial networks. More complex models that exhibit properties found in real-world spatial networks have also been proposed. Most of these models have been derived by spatial generalizations of well-known random graph models, such as Erdős-Rényi graphs, the Watts-Strogatz model, and the preferential attachment model Easley and Kleinberg (2010). However, to the best of our knowledge, none of these models is representative of networks arising out of biological spread processes like epidemics of invasive species or infectious diseases. While these network models have a base component that is grid- or lattice-like, they are missing a multi-scale component due to human-mediated pathways of spread. At the same time, the domain specific studies are limited to simulations on a single instance of the network that is usually constructed using a combination of available data and spatial interaction models such as the gravity model. Uncertainty quantification and sensitivity analyses are typically limited to diffusion model parameters, not network structure. Here, we address these gaps by developing a model that is a realistic representation of such processes, and analyzing its structural parameters such as spectral radius and diameter that are well-studied in the context of diffusion processes on networks. Later, in Sections 5 and 6, we will

further discuss how the network model parameters affect these structural properties.

4.2. General Structure

At a high level, some aspects of this spatial network are captured in **Figure 1**. Here, we will describe the graph class $G(V, E)$ of multi-pathway networks where V is a set of n vertices embedded in some suitable, 2-dimensional space M (e.g., Euclidean space $\mathbb{R}^2$, the sphere S^2). The graph $G = G(V, E)$ is composed of three graphs G_S, G_L, and G_{LD}, where the subscripts S, L, and LD denote corresponding pathways. With respect to **Figure 1**, G_S corresponds to the self-mediated spread, G_L is the spread within a locality, and G_{LD} is the inter-locality spread or long distance jumps. In general, G will be a loop-free, edge-labeled, and directed multi-graph. Edges are labeled by the pathway to which they belong with (u, v, π) encoding the edge from u to v due to pathway π. In the case where G is undirected we will use the notation $(\{u, v, \}, \pi)$ for undirected edges, or simply $\{u, v\}$ when the pathway is clear from the context. Let r denote the dispersal range. The graph G_S has vertex set V and an edge between each pair of vertices $v, v' \in V$ for which $d(v, v') \leq r$ where d is some suitable metric on the embedding space M. Next, let $\mathcal{L} = \{L_1, L_2, \ldots\}$ denote a collection of n_L mutually disjoint subsets of V. For each $L_i \in \mathcal{L}$, let E_i be some edge set over L_i, and let $G_i(L_i, E_i)$ denote the corresponding *locality graph*. The graph $G_L = G_L(V_L, E_L)$ is the (disjoint) union of the graphs G_i. Let F_{LD} be a graph with vertex set $\mathcal{L}$. We use a mapping f to define the edge set E_{LD} of G_{LD} via F_{LD} by associating to each edge (ℓ, ℓ') of F_{LD} a set of edges $E_f(\ell, \ell') \subset V \times V$ subject to the constraint that whenever $(v, v') \in E(\ell, \ell')$ we have $v \in \ell$ and $v' \in \ell'$. The graph G_{LD} has vertex set V and its edge set is the union of the sets $E_f(\ell, \ell')$ over all edges (ℓ, ℓ') of F_{LD}.

4.3. A Multi-Pathway Spatial Network Model

Here, we define the graph model class called *multi-pathway spatial network* (MPSN). Its vertex set consists of points on a $\sqrt{n} \times \sqrt{n}$ square lattice with integer coordinates (i, j), where $i, j \in \{0, 1, 2, \ldots, \sqrt{n} - 1\}$. We will assume throughout that n is a perfect square. Let k be a positive integer satisfying the following conditions: (i) it is a perfect square and (ii) $\sqrt{k}$ is a factor of $\sqrt{n}$. The vertex set is partitioned into k *regions*, each inducing a $\sqrt{n/k} \times \sqrt{n/k}$ subgrid. Each region consists of one *locality* with s vertices, where s has the same parity (odd or even) as n/k. Each locality is again a square grid of size $\sqrt{s} \times \sqrt{s}$ at the center of that region. For a locality v, let L_v denote the set of points on its square grid. Let $\mathcal{L}$ denote the set of localities.

As described above, we have three component graphs G_S, G_L, and G_{LD}. The edge set of G_S is determined by the *range* parameter r. Any two lattice points are adjacent in G_S if they are at a *Euclidean distance* of at most r. The edge set of G_L is determined by a template graph H_L, referred to as the *intra-locality graph*, on a $\sqrt{s} \times \sqrt{s}$ grid. For each locality v, let G_v denote the graph on the vertex set L_v. Consider the bijection between the vertex sets of G_v and H_L induced by the natural ordering of their vertices on the $\sqrt{s} \times \sqrt{s}$ grid. Each G_v is isomorphic to H_L under this bijection. The graph G_L is the union of all graphs G_v. Let F_{LD} be

a graph defined on the localities, referred to as the *inter-locality graph*. The long distance human-mediated pathway graph G_{LD} is defined as follows: $(a, b) \in E(G_{LD}) \iff a \in L_u$, $b \in L_v$ and $(\underline{u}, \underline{v}) \in E(F_{LD})$.

We will henceforth denote this model by $\text{MPSN}(n, r, k, s, H_L, F_{LD})$. In the following discussion, we will consider scenarios where F_{LD} is sampled from some random graph model denoted by $\mathcal{G}$. Examples of such models are Erdős-Rényi model and the preferential attachment model (Easley and Kleinberg, 2010). In such cases, F_{LD} is replaced by $\mathcal{G}$, and the resulting notation is $\text{MPSN}(n, r, k, s, H_L, \mathcal{G})$.

4.3.1. A Note on Directed Graphs

For simplicity, we only consider the undirected version of the multi-pathway network model. However, in real-world applications, these are typically directed weighted graphs that do not exhibit symmetry with respect to edge weights [see for example the networks in McNitt et al. (2019) or Carrasco et al. (2010)]. A natural directed version of the MPSN would correspond to G_S and G_L graphs with every undirected edge replaced by a pair of bidirectional edges and F_{LD} being a directed graph. Note that the resulting graph will be strongly connected.

5. SPECTRAL CHARACTERIZATION

Here, we provide bounds for the spectral radius of an MPSN instance in terms of the spectral radii of the component networks representing the different pathways. The main objective is to understand the importance of each pathway in a diffusion process through the lens of graph spectra. Our main result is Theorem 5.1, where lower and upper bounds on the spectral radius of the multi-pathway network are expressed in terms of the parameters of the MPSN model.

Theorem 5.1. *Let* $G = \text{MPSN}(n, r, k, s, H_L, F_{LD})$ *be a multi-pathway spatial network. The spectral radius of G can be bounded as follows:*

$$\max\left\{\frac{r^2}{2}, \lambda_1(H_L), \lambda_1(F_{LD})(s-1)\right\} \leq \lambda_1(G) \leq 4r^2 + \lambda_1(H_L) + \lambda_1(F_{LD})(s-1)$$

The first term of both bounds in the above result corresponds to self-mediated spread (G_S). We note that this is fully determined by the range parameter r and not by the size of the graph. Also, it increases as the square of the range. From an invasive species perspective, this suggests that a species with strong flying capability can rapidly expand its range. An example of such a recent global invasion is that of the fall armyworm (Westbrook et al., 2016; Day et al., 2017), which is known for its long distance migration. The second term in the bounds corresponds to intra-locality spread. This depends on the size and structure of the localities. A more significant contribution comes from the third term corresponding to the inter-locality spread pathway. It is proportional to the spectral radius of the inter-locality graph F_{LD}. The denser the graph, the greater its contribution. More importantly, it is amplified by $(s-1)$, the size of the locality. The main implication of this is that the higher the number of

areas are that are suitable for establishment of the pest in a locality, the greater the influence of the inter-locality pathway.

The rest of the section is devoted to a proof of the above theorem. Throughout these proofs, we will denote the edge set of a graph G by $E(G)$.

Lemma 5.2. *The spectral radius of the short-distance pathway network G_S with n nodes and range r is $\Theta(r^2)$.*

Proof: Note that G_S is a network induced on a set of points located on an $\sqrt{n} \times \sqrt{n}$ grid with integer coordinates and range parameter r. We use the notation $u = \langle i,j \rangle$ for each vertex, where $i,j \in \{0,\ldots,\sqrt{n}-1\}$ are the coordinates. Let H_1 denote the grid graph where each $\langle i,j \rangle$ is adjacent to $\langle i \pm 1, j \rangle$ and $\langle i, j \pm 1 \rangle$. Let H_2 denote the graph where each $\langle i,j \rangle$ is adjacent to $\langle i \pm 1, j \rangle$, $\langle i, j \pm 1 \rangle$, and $\langle i \pm 1, j \pm 1 \rangle$. In the definitions of both H_1 and H_2, edges to vertices whose coordinates do not exist within the $\sqrt{n} \times \sqrt{n}$ grid are not added.

Claim 1: $E(H_1^{\lfloor r \rfloor}) \subseteq E(G_S) \subseteq E(H_2^{\lfloor r \rfloor})$, where $H_1^{\lfloor r \rfloor}$ and $H_2^{\lfloor r \rfloor}$ are the $\lfloor r \rfloor$th powers of H_1 and H_2, respectively.

Proof of Claim 1: We recall that $d_H(u,v)$ is the distance between u and v in graph H and $d_{\text{Euc}}(u,v)$ is the Euclidean distance between them. First, we will show that for any $(u,v) \in E(H_1^{\lfloor r \rfloor})$, $d_{\text{Euc}}(u,v) \leq r$. To this end, it is enough to prove this for $u = \langle 0,0 \rangle$ and $v = \langle x,y \rangle$ since both distances are shift invariant. For this case, $d_{\text{Euc}}(u,v)^2 = x^2 + y^2$. Since u and v are adjacent in $H_1^{\lfloor r \rfloor}$, v is reachable from u in at most $\lfloor r \rfloor$ hops in H_1. This means that $x + y \leq r$, with x and y being positive. Therefore, $x^2 + y^2 \leq x^2 + (r-x)^2 \leq r^2 + 2x(x-r) \leq r^2$, and $E(H_1^{\lfloor r \rfloor}) \subseteq E(G_S)$. Now, we will show that if $x^2 + y^2 \leq r^2$, then $(u,v) \in E(H_2^{\lfloor r \rfloor})$. Since $x^2 + y^2 \leq r^2$ and x and y are integers, we have $x,y \leq \lfloor r \rfloor$. Note that in H_2, $d_{H_2}(\langle 0,0 \rangle, \langle x,y \rangle) = \max(x,y) \leq \lfloor r \rfloor$. This is because, assuming without loss of generality, that $x \leq y$, $\langle x,y \rangle$ is reachable from $\langle 0,0 \rangle$ by the path $\langle 0,0 \rangle \langle 1,1 \rangle \cdots \langle x,x \rangle \langle x, x+1 \rangle \cdots \langle x,y \rangle$, hence $E(G_S) \subseteq E(H_2^{\lfloor r \rfloor})$, completing our proof of Claim 1.

We now recall a result from spectral graph theory (Brouwer and Haemers, 2012).

Lemma 5.3 (Brouwer and Haemers, 2012). *For any undirected graph G, $\delta(G) \leq \lambda_1(G) \leq \Delta(G)$, where $\delta(G)$ and $\Delta(G)$ denote the minimum and maximum node degree in G.*

Claim 2: (i) $\delta(H_1^{\lfloor r \rfloor}) \geq \lfloor r \rfloor (\lfloor r \rfloor + 1)/2$. (ii) $\Delta(H_2^{\lfloor r \rfloor}) \leq 4(\lfloor r \rfloor - 1)^2 + 4(\lfloor r \rfloor)$.

Proof of Claim 2: To prove Part (i), note that in $H_1^{\lfloor r \rfloor}$, the nodes with minimum degree are the corner vertices of the grid. Recalling that $\langle 0,0 \rangle$ is a corner vertex, and any neighbor $\langle x,y \rangle$ satisfies $x + y \leq \lfloor r \rfloor$, its degree is $\lfloor r \rfloor (\lfloor r \rfloor + 1)/2$.

To prove Part (ii), note that in $H_2^{\lfloor r \rfloor}$, if $\max(x,y) \leq \lfloor r \rfloor$, then $\langle x,y \rangle$ is a neighbor of $\langle 0,0 \rangle$. There are $\lfloor r \rfloor^2 - 1$ such vertices. For sufficiently large n in comparison to r, a non-corner vertex can have a degree of up to $4(\lfloor r \rfloor - 1)^2 + 4(\lfloor r \rfloor)$. This completes our proof of Claim 2.

We now continue with our proof of Lemma 5.2. As argued earlier, $E(H_1^{\lfloor r \rfloor}) \subseteq E(G_S) \subseteq E(H_2^{\lfloor r \rfloor})$. Thus, $\delta(H_1^{\lfloor r \rfloor}) \leq \delta(G_S)$ and

so from Lemma 5.3 and Part (i) of Claim 2, we have $\lambda_1(G_S) \geq \delta(G_S) \geq \lfloor r \rfloor (\lfloor r \rfloor + 1)/2 = \Omega(r^2)$. Likewise, $\Delta(G_S) \leq \Delta(H_2^{\lfloor r \rfloor})$. So from Lemma 5.3 and Part (ii) of Claim 2, we have $\lambda_1(G_S) \leq \Delta(G_S) \leq 4(\lfloor r \rfloor - 1)^2 + 4(\lfloor r \rfloor) = O(r^2)$. We thus conclude that $\lambda_1(G) = \Theta(r^2)$. $\square$

Lemma 5.4. *Consider the local human-mediated spread pathway network G_L with n nodes, number of regions k and locality size s. Let H_L be the intra-locality graph. The set of eigenvalues of the adjacency matrix G_L is identical to the set of eigenvalues of the adjacency matrix of H_L.*

Proof: By design, G_L induces multiple connected components, with each component being isomorphic to H_L. The result follows. $\square$

Lemma 5.5. *Consider the local human-mediated spread pathway network G_{LD} with n nodes, number of regions k and locality size s. Let F_{LD} be the inter-locality graph. The spectral radius of G_{LD} equals $(s-1)\lambda_1(F_{LD})$.*

Proof: We will use the definition of localities from Section 4.3 and the definition of Kronecker product of graphs from Section 3.

By definition, $(a,b) \in E(G_{LD}) \iff a \in L_{\underline{u}}, b \in L_{\underline{v}}$ and $(\underline{u},\underline{v}) \in E(F_{LD})$. Also, each $L_{\underline{u}}$ is of the same size s. Let G_K denote a graph on s vertices which induces a complete graph. Now we show that the graph G_{LD} is equivalent to the graph H obtained as the Kronecker product of F_{LD} and G_K. Since G_K has s nodes, we can define a bijection between $V(G_K)$ and $V(L)$ for any locality L. Let this bijection be denoted by $\pi_{\underline{u}}: V(G_K) \to L_{\underline{u}}$. Let $(\underline{u},x),(\underline{v},y) \in V(H)$ where $\underline{u},\underline{v} \in V(F_{LD})$ and $x,y \in V(G_K)$.

Suppose $(\underline{u},x)$ and $(\underline{v},y)$ are two vertices in H, where $\underline{u} \neq \underline{v}$, $a = \pi_{\underline{u}}(x)$ and $b = \pi_{\underline{v}}(y)$. We will now show that $\{(\underline{u},x),(\underline{v},y)\}$ is an edge in H iff a is adjacent to b in G_{LD}. Suppose $\{(\underline{u},x),(\underline{v},y)\}$ is an edge in H. Since $(\underline{u},x)$ is adjacent to $(\underline{v},y)$, we have that $\underline{u}$ is adjacent to $\underline{v}$ in F_{LD} and x is adjacent to y in G_K. Note that since G_K is a complete graph, the latter condition will always be true. Since a belongs to locality $\underline{u}$ and b belongs to locality $\underline{v}$, it follows by the definition of G_{LD} that a and b are adjacent as well. Now suppose $\{a,b\} \in G_{LD}$. Since the two vertices belong to different localities, they can be adjacent only because $\underline{u}$ is adjacent to $\underline{v}$ in F_{LD}. Again noting that G_K is a clique, $x = \pi_{\underline{u}}^{-1}$ and $y = \pi_{\underline{v}}^{-1}$ are adjacent in G_K. Therefore, $\{(\underline{u},x),(\underline{v},y)\}$ is an edge in H.

Since G_K is a clique, its spectral radius is $s-1$ (Brouwer and Haemers, 2012). By above arguments, $\lambda_1(H) = \lambda_1(G_{LD})$. Since H is a Kronecker product of F_{LD} and G_K, its spectral radius is $\lambda_1(F_{LD})\lambda_1(G_K)$ (Brouwer and Haemers, 2012). Since $\lambda(G_K) = s-1$, it follows that $\lambda_1(H) = (s-1)\lambda_1(F_{LD})$. This completes our proof of Lemma 5.5. $\square$

From Lemmata 5.2, 5.4, and 5.5, the proof of Theorem 5.1 follows by (i) noting that the spectral radius of any graph is at least as large as the spectral radius of any of its subgraphs, and (ii) for any two Hermitian matrices A and B (such as the adjacency matrices of graphs), $\lambda_1(A+B) \leq \lambda_1(A) + \lambda_1(B)$.

6. DIAMETER

In this section, we derive an upper bound on the diameter of an MPSN instance, where the inter-locality graph is drawn from an Erdős-Rényi model. In particular, we show that the addition of a few long distance edges corresponding to the inter-locality graph can lead to a significant decrease in the diameter. Our main result is as follows.

Theorem 6.1. *Let $G \in \mathrm{MPSN}(n, r, k, s, H_L, \mathcal{G})$ be an instance of the multi-pathway network model where $F_{LD} \in \mathcal{G}$ is an instance of the Erdős-Rényi random graph model $\mathcal{G} = H(k, \epsilon/k)$ for some $\epsilon > 0$. (i) The diameter of the graph G_S (i.e., G without the inter- and intra-locality edges) is $\Omega(\sqrt{n}/r)$. (ii) The diameter of G is asymptotically almost surely $O\left(\sqrt{\frac{n}{k}} \log k\right)$.*

This theorem implies that, for a sufficiently large k (number of localities), even when the range r is low, a small number of long distance edges can reduce the diameter by a significant amount. In practice, this reduction in diameter is significant since the value of r is small while k is a much larger integer. From an application perspective, this suggests that even organisms with limited flying ability can cover great distances in a few hops due to long distance trade or travel links.

To prove Theorem 6.1, we use the following result on the diameter of randomly perturbed graphs due to Krivelevich, Reichman and Samotij (Krivelevich et al., 2015). Their main observation is that if for some small $\epsilon > 0$, approximately ϵn random edges are added to any n-node graph, then asymptotically almost surely, the diameter of the resulting graph is $O(\log n)$. Formally, their result is stated as follows.

Lemma 6.2 (Krivelevich, Reichman and Samotij (Krivelevich et al., 2015)). *For every $\epsilon > 0$, there exists $C > 0$ such that the following holds. Let G be an n-vertex connected graph, choose $R \sim G(n, \frac{\epsilon}{n})$ and let $G^* = G \cup R$. Then, asymptotically almost surely, the diameter of G^* is at most $C \log n$.*

Proof of Theorem 6.1: **Part (i):** First, we will prove the lower bound on the diameter of G_S. On the grid over which G_S is defined, consider the two nodes $x = \langle 0, 0 \rangle$ and $y = \langle \sqrt{n} - 1, \sqrt{n} - 1 \rangle$. The Euclidean distance between these two nodes is $d_{\mathrm{Euc}}(x, y) = \sqrt{2(n - 2\sqrt{n} + 2)}$. It can be verified that for $n \geq 9$, $d_{\mathrm{Euc}}(x, y) \geq \sqrt{n}$. Suppose the length of a shortest path P between x and y in G_S has ξ edges. By the definition of G_S, each of these edges has a geometric length of at most r. Therefore, the total geometric length of all these edges in P is at most $r\xi$. Thus, by the generalized triangle inequality (Section 3), the distance between x and y is at most $r\xi$. As noted above, for $n \geq 9$, this distance is at least $\sqrt{n}$. Thus, $r\xi \geq \sqrt{n}$ or $\xi \geq \frac{\sqrt{n}}{r} = \Omega(\sqrt{n}/r)$. In other words, there is at least one pair of nodes in G_S for which the shortest path uses $\Omega(\sqrt{n}/r)$ edges. Therefore, the diameter of G is $\Omega(\sqrt{n}/r)$, and this completes our proof of Part (i).

Part (ii): Recall that $\mathcal{L}$ is the set of localities. By definition, there are k localities placed on a $\sqrt{k} \times \sqrt{k}$ grid (see **Figure 2**). We will first create a base graph F_B on $\mathcal{L}$ by adding edges between neighbors on the grid. More formally, we will assume

that the coordinates of each locality $\langle x, y \rangle$ satisfy $x, y \in \{0, 1, 2, \ldots, \sqrt{k} - 1\}$. In F_B, a locality with coordinates $\langle x, y \rangle$ is adjacent to $\langle x \pm 1, y \pm 1 \rangle$ (if nodes with those coordinates are in the same locality). Noting that F_B is connected and by assumption, $F_{LD} \in H(k, \epsilon/k)$, from Lemma 6.2, it follows that the diameter of the graph $F_{LD} \cup F_B$ obtained by adding the edges of F_{LD} to F_B is asymptotically almost surely at most $c \log k$ for some positive constant c.

Now, we will prove the upper bound on the diameter of G. Let $u, v \in V$ be two distinct nodes on the grid. Let $L_{\underline{u}}$ and $L_{\underline{v}}$ be the localities closest to u and v respectively. The distance from u to the closest vertex $u' \in L_{\underline{u}}$ is at most most $2(\sqrt{n/k} - s)/\left\lceil \frac{r}{\sqrt{2}} \right\rceil$. The same holds for v. As proven above, there exists a path P of length at most $c \log k$ almost surely from $\underline{u}$ to $\underline{v}$ in $F_{LD} \cup F_B$. Let $P = \underline{u} - \underline{u}_1 - \underline{u}_2 - \cdots - \underline{u}_t - \underline{v}$. Some of the edges in this path belong to F_B and the rest to F_{LD}. If $\underline{u}_i \underline{u}_{i+1}$ belongs to F_{LD}, then by definition of G_{LD}, there exist corresponding nodes $u_i \in L_{\underline{u}_i}$ and $u_{i+1} \in L_{\underline{u}_{i+1}}$ such that u_1 and u_2 are adjacent in G_{LD} and, therefore, are adjacent in G. If on the other hand, $\underline{u}_i$ and $\underline{u}_{i+1}$ are adjacent in F_B, then, any two vertices $u_i \in L_{\underline{u}_i}$ $u_{i+1} \in L_{\underline{u}_{i+1}}$ are at a distance at most $2(\sqrt{n/k} - s)/\left\lceil \frac{r}{\sqrt{2}} \right\rceil + 2 \operatorname{diam}(G_L)$. The first term, which corresponds to inter-locality distance on the grid, is clearly $O(\sqrt{n/k})$. The second term corresponds to node-to-node distance within a locality. Note that $\operatorname{diam}(G_L) \leq s/\left\lceil \frac{r}{\sqrt{2}} \right\rceil$. Noting that $s \leq \sqrt{k}$ and $k \leq \sqrt{n}$, it can be verified that $\operatorname{diam}(G_L) = O(\sqrt{n/k})$. Thus, the length of each edge in P is $O(\sqrt{n/k})$ and the number of edges in P is asymptotically almost surely at most $c \log k$. It follows that the length of path P from u to v is asymptotically almost surely $O\left(\sqrt{\frac{n}{k}} \log k\right)$. This completes our proof of Part (ii). □

7. A MULTI-PATHWAY DIFFUSION MODEL

We now describe briefly the model that is based on the approach of McNitt et al. (2019), referred to as the SPREAD model. We refer to **Figure 1** and Section 4 for the network representation on which the diffusion process is based. There are three pathways of spread. Self-mediated dispersal corresponds to diffusion from one node to its neighboring nodes, intra-locality dispersal is diffusion within a locality (farmer-market interactions), and inter-locality diffusion corresponds to long distance dispersal from one cell to another (trade). The diffusion model is a discrete-time SEI process where a node transitions from **E** to **I** after ℓ time steps. The transition from **S** to **E** is described below.

Each node v has a periodic time-varying attribute, namely its infectivity $\rho(v, t)$. The probability that a node can be infected through a pathway is modeled as a negative exponential function of infectivity and pathway parameters, which can be expressed as edge weights between two nodes as follows. For the short-distance dispersal, the probability that node v is infected by any "neighbor" v' that is at a distance at most r (range) is given by

$$w(v', v, P_s, t) = 1 - \exp(-\alpha_s \rho(v', t)), \quad (1)$$

FIGURE 2 | An instance of the multi-pathway spatial network model $MPSN(n, r, k, s, H_L, F_{LD})$. Here, number of nodes $n = 144$, number of regions $k = 9$, and locality size $s = 4$. The range $r = 1$ resulting in a grid graph for G_S. Here, the intra-locality graph H_L is a clique on four vertices.
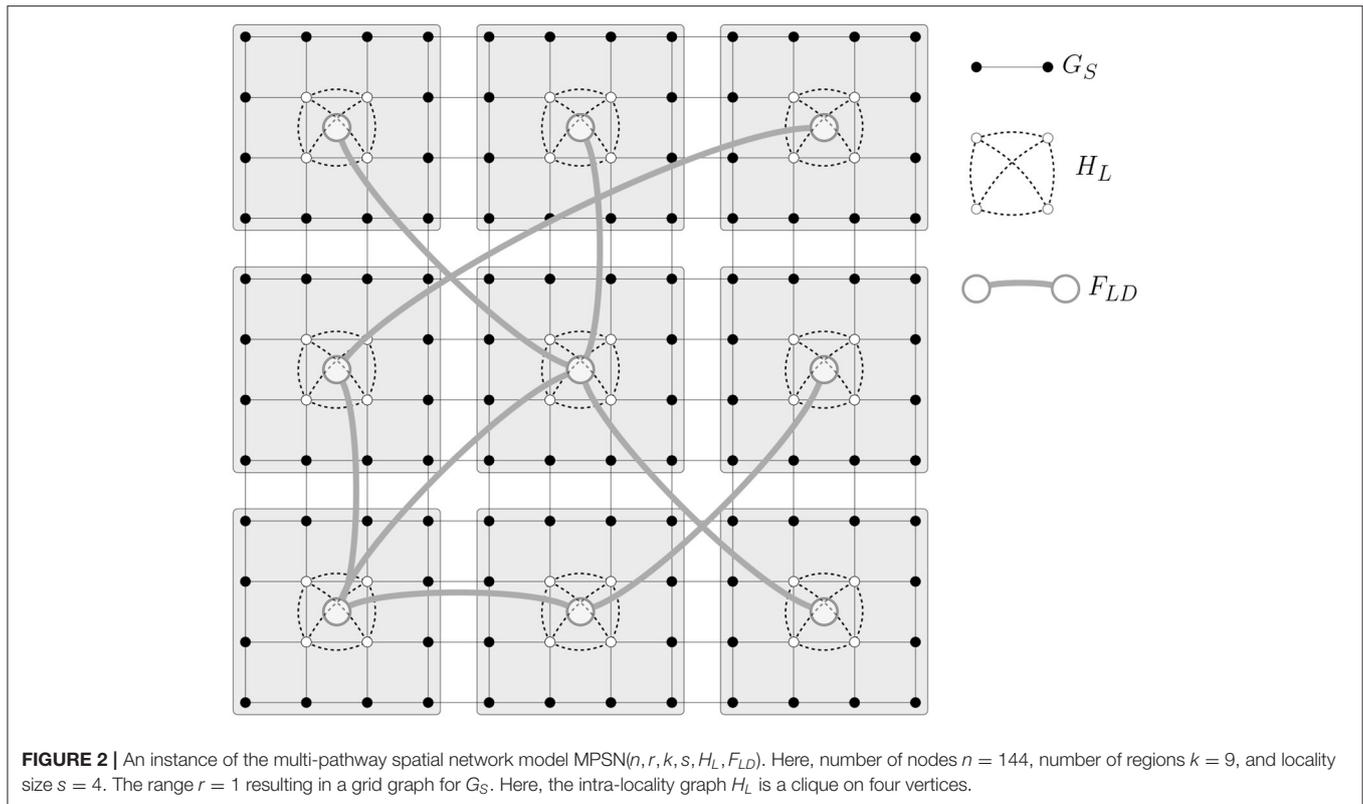
where P_s is the edge label corresponding to the pathway and α_s is a tunable pathway parameter. For two nodes v and v' within a locality, the probability of within intra-locality transmission from v' to v is given by

$$w(v', v, P_\ell, t) = 1 - \exp(-\alpha_\ell \rho(v', t)), \qquad (2)$$

where P_ℓ is the pathway label and α_ℓ is the tunable pathway parameter. For the inter-locality transmission, the weights on the flow network F_{LD} influence the probability of transmission. Suppose $v \in L_v$ and $v' \in L_{v'}$; then the probability that v is infected by v' through this pathway is given by

$$w(v', v, P_{\ell d}, t) = 1 - \exp(-\alpha_{\ell d} F_{v'v} \rho(v', t)), \qquad (3)$$

where $P_{\ell d}$ is the pathway label and $\alpha_{\ell d}$ is the tunable pathway parameter.

As mentioned earlier, the dynamics of the SPREAD model follow the Susceptible-Exposed-Infectious (SEI) (Marathe and Vullikanti, 2013), as defined below. Each node is in one of the following states: susceptible (**S**), exposed (**E**), or infectious (**I**). Let $S \subseteq V$ denote the initial set of nodes in state **I**. The nodes in S serve as the seeds for the diffusion process. At any time t, each node u in state **I** infects its susceptible neighbor v with probability equal to the weight $w(u, v, P, t)$ via the labeled edge (u, v, P). An exposed node (i.e., a node in state **E**) transitions to the infectious state after ℓ time steps, where ℓ is the latency period. It corresponds to the time taken for the node to transition from being exposed to being infectious.

8. EXPERIMENTS

8.1. Outline

The extent and pattern of spread in a network depend on both the network structure and the diffusion model properties. Our goal here is to identify the drivers of the diffusion process. We will study how combinations of these properties affect the spread in the network. Specific studies are as follows.

1. We study how MPSN model parameters such as range, number of localities and the density of the intra- and inter-locality graphs affect the structure of the network from the perspective of dynamics. Specifically, we analyze the evolution of the spectral radius ($\lambda_1(G)$) and diameter ($\text{diam}(G)$) with respect to model parameters.

2. We analyze the networks considered above with respect to the SPREAD diffusion model. Here, the objective is to study how combinations of the pathway probabilities determined by α_s, α_ℓ, and $\alpha_{\ell d}$ and network model parameters affect the spread. We also study how combinations of structural and dynamical properties influence the spread. Further, we investigate how representative network parameters (such as spectral radius and diameter) are in characterizing SEI-like diffusion processes especially when non-uniform edge probabilities are involved.

3. Based on the insights derived from the above two studies, we analyze the properties of several real-world commodity flow networks (McNitt et al., 2019). These networks are temporal and edge-weighted unlike MPSN considered in the two studies

TABLE 1 | List of model parameters used in the experiments for the MPSN model.

Parameters	Values
Number of nodes n	4096
Number of regions k	4, 16, 64
Range r	1, 1.5, 2, . . ., 4
Locality size s	4, 16, 64
Intra-locality graph H_L	Complete graph, star graph
Inter-locality graph F_{LD}	$G(n, p)$
ϵ: (inter-locality graph edge probability ϵ/k)	0.1, 0.5, 1, 5, 10

TABLE 2 | List of real-world networks used and their attributes.

Net.	Name	#Nodes	#Edges	#Groups	#Group edges
BD	Bangladesh	211	6846	7	141
ID	Indonesia	3296	110640	35	2181
PH	Philippines	673	20108	16	450
TH	Thailand	738	27666	5	48
VN	Vietnam	503	16746	15	426

Each of these networks is constructed using various datasets on production, trade, and consumption of tomatoes (McNitt et al., 2019). These networks are edge-weighted.

TABLE 3 | List of network and SPREAD model parameters used in the experiments for the MPSN model.

Parameters	Values
Number of nodes n	4096
Number of regions k	16
Range r	1, 2, 3
Locality size s	4, 16
Intra-locality graph H_L	Complete graph
Inter-locality graph F_{LD}	$G(n, p)$
ϵ: (inter-locality graph edge probability ϵ/k)	0.1, 1, 10
α_s: grid pathway parameter	0, 0.001, 0.005, 0.01, 0.015, 0.01
α_ℓ: intra-locality pathway parameter	0, 0.001, 0.005, 0.01, 0.015, 0.01
$\alpha_{\ell d}$: inter-locality pathway parameter	0, 0.001, 0.005, 0.01, 0.015, 0.01
ρ: Infectivity of a node	1 for all nodes
Edge weights	1 for all edges
ℓ: exposure delay	0
Time horizon T	6, 12, 18, 24

above. We analyze the structural properties of the temporal snapshots of these networks.

8.2. Networks and Experiment Design
All the code used to generate the results is made publicly available (Adiga, 2021). The parameters used to generate synthetic networks corresponding to the MPSN model and the characteristics of real-world datasets corresponding to domestic trade of tomatoes are provided in **Tables 1**, **2**, respectively. For the synthetic networks, we considered a 64 × 64 grid of nodes. Using a full factorial design, networks were constructed for combinations of parameters in **Table 1**. For the inter-locality

graph F_{LD}, we used the Erdős-Rényi random graph model where the edge probabilities were chosen to be functions of the number of nodes of F_{LD}, which in turn is equal to the number of regions k. However, not all combinations of parameters are valid. Since these networks have a random graph component, for each valid combination of parameters, we constructed 10 replicates. Thus, over 7,000 networks were constructed. We note that while the synthetic networks considered here are undirected and unweighted, the real-world networks considered (**Table 2**) are directed and edge-weighted. In addition, these temporal networks are also periodic: there are 12 snapshots of networks, one for every month of the year induced by the seasonal production and flow of commodity around the year. The details of network construction are given in McNitt et al. (2019).

Structural properties were computed for all the synthetic networks. A subset of these networks was used for dynamical analysis using the diffusion model presented in Section 7. For this, a range of parameter values corresponding to the SPREAD diffusion model of Section 7 were used. This is listed in **Table 3**. The number of simulation instances in each case was 100. Since diffusion is modeled as an SEI process, assuming that the network is strongly connected, the fixed points or the equilibrium points of the system are either all nodes being in the infected state or all nodes being in the susceptible state. Therefore, we are interested in the state of the system at specific time steps. The observed variable is the mean number of nodes infected by a given *time horizon* or by *time step T*. For the initial seeding, five percent of the nodes were chosen uniformly at random and set to the infected state.

8.3. Structural Analysis of MPSN Model
The decision tree analysis of spectral radius $\lambda_1(G)$ in **Figures 3**, **4** indicates that the primary drivers of its value are locality size s and the inter-locality edge probability factor ϵ. The significance of the former can be attributed to the model. We recall that the dominating term in the bounds provided in Theorem 5.1 is $\lambda_1(G_{LD}) = \lambda_1(F_{LD})(s - 1)$ consisting of the locality size and ϵ that influences $\lambda_1(F_{LD})$. For Erdős-Rényi graphs, the spectral radius asymptotically tends to the maximum of ϵ and the square root of the maximum degree of the graph (Krivelevich and Sudakov, 2003). We note that even though range r contributes to the value of $\lambda_1(G)$, its influence is quite small compared to the other parameters. Here, we recall that locality size s, $\lambda_1(H_L)$ and $\lambda_1(F_{LD})$ are factors related to human-mediated dispersal. Under the assumption that the spectral radius is an indicator of rate of dispersal, we can see the significance of these components with respect to the diffusion process. We also note that for larger values of locality size and ϵ, the number of regions k contributes to the value of the spectral radius as the maximum degree of F_{LD} increases with k.

While diameter, like spectral radius, is closely related to dynamics, we observe in **Figures 3**, **4** that the most significant influence on its value is due to the range r; this is closely followed by the number of regions k and ϵ. We note that doubling the range reduces the diameter by approximately half. Greater the number of regions and ϵ, the greater is the number of long distance edges. Hence, the diameter goes down. We note that

FIGURE 3 | Regression tree analysis of network properties with respect to model parameters for the MPSN model. The properties were computed for various networks given in **Table 1**. **(A)** CART analysis: Spectral radius λ1(G). **(B)** CART analysis: Diameter diam (G).
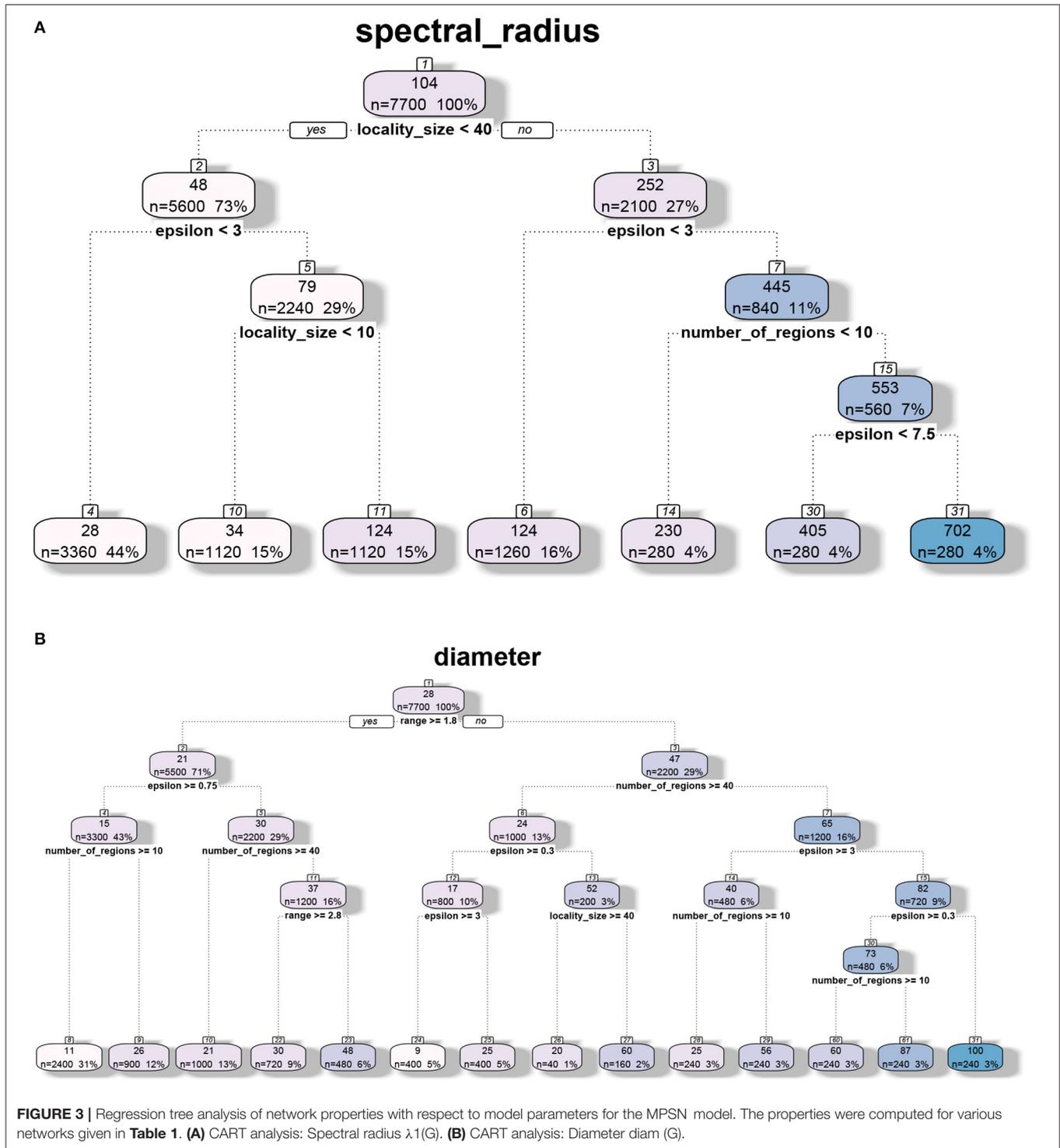
unlike the case of spectral radius, locality size is not an important factor with respect to diameter.

8.4. Dynamical Analysis of MPSN Model

These experiments correspond to diffusion using the SPREAD model on different MPSN instances and different combinations of pathway parameters α_s, α_ℓ, and $\alpha_{\ell d}$. The observed variable is the mean number of infections that have occurred by a given time T. The results of our experiments are in **Figures 5**, **6**, where the progressions of the diffusion process under different conditions are plotted. A general observation is that, when the number of infected nodes is much greater than $\sqrt{n}$, the rate of

FIGURE 4 | Parameter importance using random forest analysis of network properties with respect to model parameters for the MPSN model. The properties were computed for various networks given in **Table 1**. **(A)** Parameter sensitivity: Spectral radius λ1(G). **(B)** Parameter sensitivity: Diameter diam(G).

FIGURE 5 | Unraveling pathways: Here, we choose a positive value $0 \leq c < 1$. All pathway parameters α_s, α_ℓ, and $\alpha_{\ell d}$ are either set to 0 or c. In the experiments, we incrementally unravel the pathways starting the edges from the grid ($\alpha_s = c$, $\alpha_\ell = 0$, $\alpha_{\ell d} = 0$), followed by intra-locality edges ($\alpha_s = c$, $\alpha_\ell = c$, $\alpha_{\ell d} = 0$), and finally, inter-locality edges ($\alpha_s = c$, $\alpha_\ell = c$, $\alpha_{\ell d} = c$). For all the cases where $\alpha_{\ell d} \neq 0$, the variance is higher compared to the remaining cases as these also account for 10 replicates of the networks.

spread is linear. This is because, for any infected vertex, most or all of its neighbors are infected. In each time step, at the forefront of the spread process, at most a constant times $\sqrt{n}$ vertices are available to be infected in the next time step.

8.4.1. Progression of the Diffusion Process
Our first set of experiments concern the importance of pathway parameters given a network. Here, we fix a transmission probability on each edge. Using the pathway parameters, we first activate only the edges of G_S. This is the baseline process of diffusion over just the grid. From the domain perspective, this corresponds to natural or self-mediated spread alone. Then,

we unravel the edges corresponding to G_L, followed by the edges of G_{LD}. In **Figure 5**, the top left plot is the reference plot corresponding to MPSN with $s = 16$, $\epsilon = 1$, and $r = 2$. We see that when only $\alpha_s > 0$ or in other words, only the grid edges are activated, the spread is slow in the beginning followed by an increase in the rate before it saturates. A similar phenomenon is observed even when the intra-locality and inter-locality edges are activated. We note that the rate of increase is higher when the inter-locality edges are introduced; this highlights the importance of long distance edges in increasing the rate of spread. We compared the reference plot with diffusion on networks where $\epsilon = 0.1$ (top middle in **Figure 5**) and $\epsilon =$

FIGURE 6 | Non-uniform pathway probabilities: In each figure, there are three sets of plots. In each set, two pathway parameters are fixed while the remaining pathway parameter value is varied from 0 to 0.01.
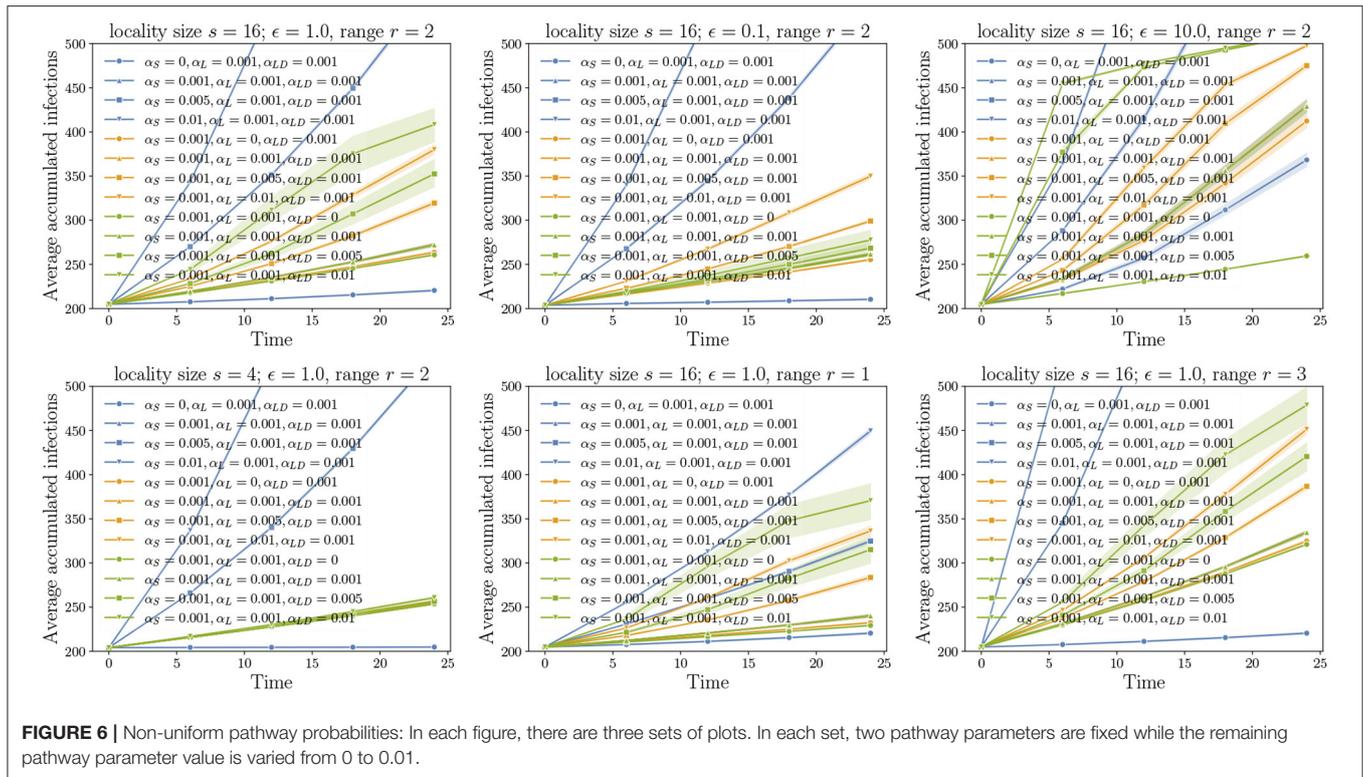
10.0 (top right). For $\epsilon = 0.1$, the inter-locality edges do not contribute to the infections, while for $\epsilon = 10$, this pathway is dominant. In addition, the progression of the process is very fast in the beginning and saturates quickly, after which the increase is almost linear. This indicates that the long distance edges contribute to fast short-term spread, an important factor to account for in preparing for interventions.

8.4.2. Importance of Locality Size and Range
The bottom left plot in **Figure 5** corresponds to a smaller locality size. We see that locality size significantly affects both intra- and inter-locality spread. Reduction in locality size to 4 makes these pathways insignificant. With increase in range (bottom right), the diffusion through short distance pathway is rapid, and, therefore, the process reaches saturation before either the intra- or inter-locality pathways can make a difference.

8.4.3. Non-Uniform Probabilities
In **Figure 6**, we fix two pathway parameters while the third parameter is varied. We observe that the highest increase in the rate of spread corresponds to the short-distance parameter α_s and the least is α_ℓ, the intra-locality parameter. We also note that larger the ϵ, the greater the number of long distance edges and, therefore, the higher the increase in the rate of spread as $\alpha_{\ell d}$ is increased (see top row of **Figure 6**).

8.4.4. Regression-Tree Analysis
We applied regression tree and random forest algorithms on the simulation data. The first objective was to find the network and model parameters that drive the infection. The second objective

is to determine how structural parameters such as spectral radius and diameter influence the spread. Accordingly, we have two sets of results in **Figures 7**, **8**, respectively. The regression-tree analysis succinctly captures some aspects of the results in **Figures 5**, **6**. As observed in the earlier plots, the rate of spread is determined primarily by α_s and range r. For smaller time steps, the inter-locality spread is significant, particularly when α_s and r are small (**Figure 7**). Therefore, ϵ and locality size s are significant predictors of spread. However, in the later time steps, due to saturation, only the parameters that affect short-distance spread are good predictors of spread.

8.4.5. Spectral Radius, Diameter, and Number of Infections
In **Figure 8**, we note that spectral radius features only in the first plot. This is again because spectral radius is determined primarily by inter-locality parameters. However, for later time steps, only diameter and α_s determine the mean number of infections. Again, this just reflects the fact that diameter and range are correlated. Therefore, spectral radius is a good predictor of spread during the initial phase (either first few time steps or when the number of infections is small).

8.5. Robustness of Regression Tree Analysis
We note that the outputs of the random forest algorithm are influenced by hyper parameters, such as number of estimators (or trees), maximum depth, minimum number of samples required to split an internal node, etc. In **Figure 9**, we have plotted the

FIGURE 7 | Regression tree analysis of simulation results where the independent variables are the MPSN parameters and pathway parameters of the SPREAD model and the observed variable is the mean number of infections at different time steps T.

parameter importance based on increase in node purity (Gini index) and increase in mean square error for node size and number of trees. We note that the results are consistent.

8.6. Analysis of Real-World Commodity Flow Networks

For the real-world networks, we computed the spectral radius and diameter for the following graph. We first decomposed the network into 12 components, one for each month of the

year. In the SPREAD model of Section 7, we note that in Equations (1)–(3) the probability of transmission is dependent on infectivity $\rho(v, t)$. For the short-distance graphs G_S and G_L, every outgoing edge was assigned a weight of $\rho(u, t)$, where u is the source node. For G_{LD}, every outgoing edge was assigned the weight of $\rho(u, t)F_{\underline{uv}}$, where $\underline{u}$ is the locality to which the source u belongs and $F_{\underline{uv}}$ is the edge weight on the edge $\{\underline{u}, \underline{v}\}$ in the inter-locality graph F_{LD}. The (weighted) adjacency matrices of these graphs were added together and the

FIGURE 8 | Regression tree analysis of simulation results where the independent variables are the spectral radius and diameter of the networks and the observed variable is the mean number of infections at different time steps *T*.

spectral radius of the resulting matrix is computed. Note that the spectral radius is a real number as the matrix corresponds to a strongly connected directed graph. This follows from the Perron-Frobenius theorem (Brouwer and Haemers, 2012). We call this the *weighted spectral radius*. We also computed the spectral radius of the graph without the weights (i.e., the weight is 1 iff the *ij*th entry of the adjacency matrix is non-zero). This is referred to as the *unweighted spectral radius*. The diameter was computed for the unweighted graph.

The results are in **Figure 10** where the three structural properties are plotted for different months and two range values, namely 1 and 2. We observe that the weighted spectral radius is very high for BD compared to other networks. In all networks, we see variation in the spectral radius. This is due to seasonal fluctuations in the production (which affects ρ) and, therefore, in trade (which affects the long distance edges). The PH network shows highest variation in the case of unweighted spectral radius. This parameter is sensitive to the presence or absence of edges.

FIGURE 9 | Parameter importance using random forest analysis. The analysis is performed for different hyperparameter values of the random forest model.

FIGURE 10 | Structural properties of real-world datasets from **Table 2**. Since these are temporal networks, the properties are plotted for 12 snapshots, each representing a month of the year by natural order. In the top right plot, TH overlaps with VN.

During offseason, the absence of edges leads to a reduction in the unweighted spectral radius. However, in other networks, during offseason, only the weights decrease. We note that spectral radii increase significantly when the range is increased.

The diameter remains constant and is very high unlike what was observed in MPSN. The primary reason for this is that the trade flows do not follow a random graph pattern. The presence of a few areas of production implies many outgoing edges from a small number of localities to other localities that are major areas of consumption. This implies that the rate of spread could be highly dependent on the seed nodes. Thus, if the seed nodes correspond to high production areas, then the spread will be rapid.

9. SUMMARY AND FUTURE WORK

Motivated by the need to obtain a better understanding of the spread of biological invasions, we proposed an abstract multi-pathway diffusion model. Using structural properties of the underlying networks, we characterized complex spatial diffusion processes over networks. We also investigated the role of the different pathways in determining the rate and pattern of spread. In addition, we analyzed the structural and dynamical properties of networks using data mining algorithms such as regression trees.

Our methods provide just the first step in understanding how multiple pathways and diffusion parameters determine the extent and pattern of the spread of biological invasions. There

are several additional directions for future research. First, one can investigate the role of other structural parameters of the graphs generated by multi-pathway model (e.g., clustering coefficient, k-core sizes, closeness centrality (Easley and Kleinberg, 2010)) in determining properties of the spread process. A second direction is to investigate the suitability of other epidemic models proposed in the literature (Marathe and Vullikanti, 2013) for biological invasions. A third direction is to incorporate more realistic models like the gravity model and scale-free graphs for inter-locality spread. Finally, it may be of interest to refine the abstract multi-pathway model to allow the generation of temporal networks which play an important role in practice.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here at: https://github.com/rmuniappan/SPREAD_model.

AUTHOR CONTRIBUTIONS

AA defined the scope of the research. AA, NP, and YB designed and conducted the experiments and analyzed the results. AA, SR, and HM contributed to the theoretical formulation and results and wrote the paper. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Adiga, A. (2021). *Software for "Network Models and Simulation Analytics for Multi-Scale Dynamics of Biological Invasion.* Available online at: https://github.com/abhijin/SPREAD_frontiers_bigdata_21_published_code

Ajelli, M., Gonçalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J. J., et al. (2010). Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infect. Dis.* 10, 1–13. doi: 10.1186/1471-2334-10-190

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., and Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci.* 106, 21484–21489. doi: 10.1073/pnas.0906910106

Banos, A., Corson, N., Gaudou, B., Laperrière, V., and Coyrehourcq, S. R. (2015). The importance of being hybrid for spatial epidemic models: a multi-scale approach. *Systems* 3, 309–329. doi: 10.3390/systems3040309

Barrett, C., Bisset, K., Chandan, S., Chen, J., Chungbaek, Y., Eubank, S., et al. (2013). "Planning and response in the aftermath of a large crisis: an agent-based informatics framework," in *Proceedings of the 2013 Winter Simulation Conference* (Washington, DC), 1515–1526.

Barthélemy, M. (2011). Spatial networks. *Phys. Rep.* 499, 1–101. doi: 10.1016/j.physrep.2010.11.002

Bashan, A., Berezin, Y., Buldyrev, S. V., and Havlin, S. (2013). The extreme vulnerability of interdependent spatially embedded networks. *Nat. Phys.* 9, 667–672. doi: 10.1038/nphys2727

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Breiman, L. (2017). *Classification and Regression Trees.* Boca Raton, FL: Routledge. doi: 10.1201/9781315139470

Brouwer, A. E., and Haemers, W. H. (2012). *Spectra of Graphs.* New York, NY: Springer.

Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., and Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. *Nature* 464, 1025–1028. doi: 10.1038/nature08932

Carrasco, L. R., Mumford, J., MacLeod, A., Harwood, T., Grabenweger, G., Leach, A., et al. (2010). Unveiling human-assisted dispersal mechanisms in invasive alien insects: integration of spatial stochastic simulation and phenology models. *Ecol. Model.* 221, 2068–2075. doi: 10.1016/j.ecolmodel.2010.05.012

Chapman, D. S., White, S. M., Hooftman, D. A., and Bullock, J. M. (2015). Inventory and review of quantitative models for spread of plant pests for use in pest risk assessment for the EU territory. *EFSA Support. Publ.* 12, 1–190. doi: 10.2903/sp.efsa.2015.EN-795

Chen, C., Peng, R., Ying, L., and Tong, H. (2021). Fast connectivity minimization on large-scale networks. *ACM Trans. Knowl. Disc. Data (TKDD)* 15, 1–25. doi: 10.1145/3442342

Crowl, T. A., Crist, T. O., Parmenter, R. R., Belovsky, G., and Lugo, A. E. (2008). The spread of invasive species and infectious disease as drivers of ecosystem change. *Front. Ecol. Environ.* 6, 238–246. doi: 10.1890/070151

Cunniffe, N. J., Koskella, B., Metcalf, C. J. E., Parnell, S., Gottwald, T. R., and Gilligan, C. A. (2015). Thirteen challenges in modelling plant diseases. *Epidemics* 10, 6–10. doi: 10.1016/j.epidem.2014.06.002

Davis, M. A., Thompson, K., and Grime, J. P. (2001). Charles S Elton and the dissociation of invasion ecology from the rest of ecology. *Divers. Distrib.* 7, 97–102. doi: 10.1046/j.1472-4642.2001.00099.x

Day, R., Abrahams, P., Bateman, M., Beale, T., Clottey, V., Cock, M., et al. (2017). Fall armyworm: impacts and implications for Africa. *Outlooks Pest Manag.* 28, 196–201. doi: 10.1564/v28_oct_02

Diagne, C., Leroy, B., Vaissière, A.-C., Gozlan, R. E., Roiz, D., Jarić, I., et al. (2021). High and rising economic costs of biological invasions worldwide. *Nature* 592, 571–576. doi: 10.1038/s41586-021-03405-6

Douma, J., Pautasso, M., Venette, R. C., Robinet, C., Hemerik, L., Mourits, M. C., et al. (2016). Pathway models for analysing and managing the introduction of alien plant pests: an overview and categorization. *Ecol. Model.* 339, 58–67. doi: 10.1016/j.ecolmodel.2016.08.009

Easley, D., and Kleinberg, J. (2010). *Networks, Crowds and Markets: Reasoning About a Highly Connected World* (New York, NY: Cambridge University Press).

Fadikar, A., Higdon, D., Chen, J., Lewis, B., Venkatramanan, S., and Marathe, M. (2018). Calibrating a stochastic, agent-based model using quantile-based emulation. *SIAM/ASA J. Uncertainty Quantification* 6, 1685–1706. doi: 10.1137/17M1161233

Ferrari, J. R., Preisser, E. L., and Fitzpatrick, M. C. (2014). Modeling the spread of invasive species using dynamic network models. *Biol. Invasions* 16, 949–960. doi: 10.1007/s10530-013-0552-6

Fitzpatrick, M. C., Preisser, E. L., Porter, A., Elkinton, J., and Ellison, A. M. (2012). Modeling range dynamics in heterogeneous landscapes: invasion of the hemlock woolly adelgid in eastern North America. *Ecol. Appl.* 22, 472–486. doi: 10.1890/11-0009.1

Fox, G., Glazier, J. A., Kadupitiya, J., Jadhao, V., Kim, M., Qiu, J., et al. (2019). "Learning everywhere: pervasive machine learning for effective

high-performance computation," in *2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)* (Rio de Janeiro: IEEE), 422–429.

Ganesh, A., Massoulié, L., and Towsley, D. (2005). "The effect of network topology on the spread of epidemics," in *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, Vol. 2 (Miami, FL: IEEE), 1455–1466.

Gutfraind, A., Safro, I., and Meyers, L. A. (2015). "Multiscale network generation," in *2015 18th International Conference on Information Fusion (Fusion)* (Washington, DC: IEEE), 158–165.

Hernandez Nopsa, J. F., Daglish, G. J., Hagstrum, D. W., Leslie, J. F., Phillips, T. W., Scoglio, C., et al. (2015). Ecological networks in stored grain: key postharvest nodes for emerging pests, pathogens, and mycotoxins. *BioScience* 65, 985–1002. doi: 10.1093/biosci/biv122

Hoffmann, B. D., and Courchamp, F. (2016). Biological invasions and natural colonisations: are they that different? *NeoBiota* 29, 1. doi: 10.3897/neobiota.29.6959

Holme, P. (2013). Extinction times of epidemic outbreaks in networks. *PLoS ONE* 8:e84429. doi: 10.1371/journal.pone.0084429

Kamra, N., Zhang, Y., Rambhatla, S., Meng, C., and Liu, Y. (2021). PolSIRD: modeling epidemic spread under intervention policies. *J. Healthcare Inf. Res.* 5, 1–18. doi: 10.1007/s41666-021-00099-3

Kim, Y., Dommergues, L., M'sa, A. B., Mérot, P., Cardinale, E., Edmunds, J., et al. (2018). Livestock trade network: potential for disease transmission and implications for risk-based surveillance on the island of Mayotte. *Sci. Rep.* 8, 1–10. doi: 10.1038/s41598-018-29999-y

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *J. Complex Netw.* 2, 203–271. doi: 10.1093/comnet/cnu016

Koch, F. H., Yemshanov, D., Haack, R. A., and Magarey, R. D. (2014). Using a network model to assess risk of forest pest spread via recreational travel. *PLoS ONE* 9:e102105. doi: 10.1371/journal.pone.0102105

Krivelevich, M., Reichman, D., and Samotij, W. (2015). Smoothed analysis on connected graphs. *SIAM J. Disc. Math.* 29, 1654–1669. doi: 10.1137/151002496

Krivelevich, M., and Sudakov, B. (2003). The largest eigenvalue of sparse random graphs. *Combinatorics Probabil. Comput.* 12, 61–72. doi: 10.1017/S0963548302005424

Lamperti, F., Roventini, A., and Sani, A. (2018). Agent-based model calibration using machine learning surrogates. *J. Econ. Dyn. Control* 90, 366–389. doi: 10.1016/j.jedc.2018.03.011

Marathe, M., and Vullikanti, A. K. S. (2013). Computational epidemiology. *Commun. ACM* 56, 88–96. doi: 10.1145/2483852.2483871

McNitt, J., Chungbaek, Y. Y., Mortveit, H., Marathe, M., Campos, M. R., Desneux, N., et al. (2019). Assessing the multi-pathway threat from an invasive agricultural pest: *Tuta absoluta* in Asia. *Proc. R. Soc. B* 286:20191159. doi: 10.1098/rspb.2019.1159

Microsoft (2020). *AI for Earth*. Available online at: https://www.microsoft.com/enus/ai/ai-for-earth (accessed January 21, 2022).

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., and Vespignani, A. (2015). Epidemic processes in complex networks. *Rev. Mod. Phys.* 87, 925. doi: 10.1103/RevModPhys.87.925

Pimentel, D., Zuniga, R., and Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol. Econ.* 52, 273–288. doi: 10.1016/j.ecolecon.2004.10.002

Pitt, J. P., Worner, S. P., and Suarez, A. V. (2009). Predicting argentine ant spread over the heterogeneous landscape using a spatially explicit stochastic model. *Ecol. Appl.* 19, 1176–1186. doi: 10.1890/08-1777.1

Prakash, B. A., Chakrabarti, D., Valler, N. C., Faloutsos, M., and Faloutsos, C. (2012). Threshold conditions for arbitrary cascade models on arbitrary networks. *Knowl. Inf. Syst.* 33, 549–575. doi: 10.1007/s10115-012-0520-y

Pyšek, P., and Richardson, D. M. (2010). Invasive species, environmental change and management, and health. *Ann. Rev. Environ. Resour.* 35, 25–55. doi: 10.1146/annurev-environ-033009-095548

Robinet, C., Kehlenbeck, H., Kriticos, D. J., Baker, R. H., Battisti, A., Brunel, S., et al. (2012). A suite of models to support the quantitative assessment of spread in pest risk analysis. *PLoS ONE* 7:e43366. doi: 10.1371/journal.pone.0043366

Saha, S., Adiga, A., Prakash, B. A., and Vullikanti, A. K. S. (2015). "Approximation algorithms for reducing the spectral radius to control epidemic spread,"

in *Proceedings of the 2015 SIAM International Conference on Data Mining* (Vancouver, BC: SIAM), 568–576. doi: 10.1137/1.9781611974010.64

Serrano, M. Á., Boguná, M., and Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci.* 106, 6483–6488. doi: 10.1073/pnas.0808904106

Smolik, M., Dullinger, S., Essl, F., Kleinbauer, I., Leitner, M., Peterseil, J., et al. (2010). Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. *J. Biogeography* 37, 411–422. doi: 10.1111/j.1365-2699.2009.02227.x

Sutrave, S., Scoglio, C., Isard, S. A., Hutchinson, J. S., and Garrett, K. A. (2012). Identifying highly connected counties compensates for resource limitations when evaluating national spread of an invasive pathogen. *PLoS ONE* 7:e37793. doi: 10.1371/journal.pone.0037793

Taghvaei, A., Georgiou, T. T., Norton, L., and Tannenbaum, A. (2020). Fractional SIR epidemiological models. *Sci. Rep.* 10, 1–15. doi: 10.1038/s41598-020-77849-7

USDA-NSF (2017). *Innovations at the Nexus of Food, Energy and Water Systems (INFEWS)*. National Science Foundation. Available online at: https://www.nsf.gov/pubs/2017/nsf17530/nsf17530.htm (accessed January 21, 2022)

USDA-NSF (2021). *AI Research Institutes*. National Science Foundation. Available online at: https://nifa.usda.gov/pressrelease/usda-nifa-and-nsf-invest-220m-artificialintelligence-research-institutes (accessed January 21, 2022).

Van Mieghem, P., Stevanović, D., Kuipers, F., Li, C., Van De Bovenkamp, R., Liu, D., et al. (2011). Decreasing the spectral radius of a graph by link removals. *Phys. Rev. E* 84:016101. doi: 10.1103/PhysRevE.84.016101

Venette, R. C., Kriticos, D. J., Magarey, R. D., Koch, F. H., Baker, R. H., Worner, S. P., et al. (2010). Pest risk maps for invasive alien species: a roadmap for improvement. *BioScience* 60, 349–362. doi: 10.1525/bio.2010.60.5.5

Venkatramanan, S., Sadilek, A., Fadikar, A., Barrett, C. L., Biggerstaff, M., Chen, J., et al. (2021). Forecasting influenza activity using machine-learned mobility map. *Nat. Commun.* 12, 1–12. doi: 10.1038/s41467-021-21018-5

Venkatramanan, S., Wu, S., Shi, B., Marathe, A., Marathe, M., Eubank, S., et al. (2020). Modeling commodity flow in the context of invasive species spread: Study of *Tuta absoluta* in Nepal. *Crop Protect.* 135:104736. doi: 10.1016/j.cropro.2019.02.012

Walpole, J., Papin, J. A., and Peirce, S. M. (2013). Multiscale computational models of complex biological systems. *Ann. Rev. Biomed. Eng.* 15, 137–154.

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393, 440–442. doi: 10.1038/30918

West, D. B. (2006). *Introduction to Graph Theory*. (Upper Saddle River, NJ: Prentice-Hall International).

Westbrook, J., Nagoshi, R., Meagher, R., Fleischer, S., and Jairam, S. (2016). Modeling seasonal migration of fall armyworm moths. *Int. J. Biometeorol.* 60, 255–267. doi: 10.1007/s00484-015-1022-x

Zhang, Y., Adiga, A., Saha, S., Vullikanti, A., and Prakash, B. A. (2016). Near-optimal algorithms for controlling propagation at group scale on networks. *IEEE Trans. Knowl. Data Eng.* 28, 3339–3352. doi: 10.1109/TKDE.2016.2605088