Networks and Plant Disease Management: Concepts and Applications

M.W. Shaw[1,*] and M. Pautasso[2]

[1]School of Agriculture, Policy and Development, University of Reading, Whiteknights, Reading RG6 6AR, United Kingdom; email: m.w.shaw@reading.ac.uk

[2]Forest Pathology and Dendrology, Institute of Integrative Biology, ETH Zurich, CH-8092 Zurich, Switzerland; email: marpauta@gmail.com

Annu. Rev. Phytopathol. 2014. 52:477–93

First published online as a Review in Advance on June 23, 2014

The *Annual Review of Phytopathology* is online at phyto.annualreviews.org

This article's doi: 10.1146/annurev-phyto-102313-050229

*Corresponding author

Keywords

biosecurity, invasion, model, spatial heterogeneity, threshold, trade

Abstract

A network is a natural structure with which to describe many aspects of a plant pathosystem. The article seeks to set out in a nonmathematical way some of the network concepts that promise to be useful in managing plant disease. The field has been stimulated by developments designed to help understand and manage animal and human disease, and by technical infrastructures, such as the internet. It overlaps partly with landscape ecology. The study of networks has helped identify likely ways to reduce the flow of disease in traded plants, to find the best sites to monitor as warning sites for annually reinvading diseases, and to understand the fundamentals of how a pathogen spreads in different structures. A tension between the free flow of goods or species down communication channels and free flow of pathogens down the same pathways is highlighted.

INTRODUCTION: WHY THE INTEREST?

Many new types of networks have developed in the built and social environment over the past few decades. This has stimulated cross-fertilization between network-based approaches in a number of specialized areas of science. In particular, the development of the internet and then of computer-assisted social networking, the spread of HIV and other diseases, and the problems in biological conservation have all stimulated new insights into the general characteristics of abstract networks.

There have been many recent phytopathogen invasions of public concern. Thinking about these in terms of networks has been very valuable. The relevant networks may describe the interactions between individuals, populations, or regions. Network theory has direct relevance to trade and biosecurity, agricultural policy, and biological conservation, and also highlights some conflicts between public policy goals, which we discuss below. Network ideas are also beginning to be useful in managing established and endemic diseases. Previous reviews have summarized the application of network ideas for some areas of application in plant epidemiology (12, 18, 27, 28, 39, 49). The aim of this review is to explain more fully, although informally, some of the key ideas of network theory, with examples of their application in a plant pathology context, and relate them to other approaches. Where appropriate, we refer to examples from animal or human disease, but we have tried to keep the focus on questions relevant to plant disease and omit all mathematical detail. An excellent full introduction to network theory—although with little emphasis on plant disease—can be found in Reference 44; a brief but thoughtful review of the ideas across many disciplines can be found in Reference 36.

Network theory is also transformative in cell biology, where the connections are transformations or signals and the nodes are quantities or forms of substance. The relevance to molecular plant pathology is clear. However, this topic is beyond the scope of the present review; interested readers are referred to, for example, Chaiboonchoe et al. (9).

DEVELOPMENT OF NETWORK IDEAS

An abstract network consists of a collection of nodes (e.g., fields, plants, cells, individual people, internet routers, power stations, and cities; any distinct entities) and connections between the nodes (e.g., wind, tractors, aphids, chemical signals, conversation, optical fibers, electrical power grids, railways, etc.) (see sidebar, Many Names, One Concept). The connections may be physical flows between different locations, but they can also be abstract: eating or being eaten, infecting or being infected, for example. Network thinking in the population biology of plant pathogens has been implicit since the subject emerged. The changes over the past few decades have at least

MANY NAMES, ONE CONCEPT

Because the theory of networks has arisen in many fields, there are varying terms for the entities making up a network. They may be described abstractly as nodes or vertices and according to the application may be individuals, small populations, farm-holdings, areas of crop or forest, and so on; the connections may be called links, channels, or edges. The vertex and edge descriptions arise from graph theory and visualize the network as a generalization of a polyhedron, such as a cube or triangle. The node-link description is a natural way to think of the internet or a road or trade network; channels, pathways, connections, or contacts may be clearer ways to describe edges or links in population biology or epidemiology contexts.

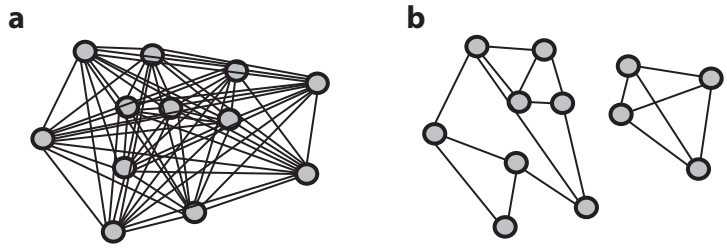


Figure 1

Networks equivalent to homogeneous mixing. (*a*) Every node connected to every other node. (*b*) Nodes are connected at random to several other nodes chosen at random from the whole population. In a small network, this may lead to more than one component (disconnected set of nodes) forming, as shown.

three aspects. First is an expansion of the ambition of ecology and subfields such as plant disease management to much larger spatial and temporal scales (24, 25, 37, 57). Second is the realization that networks may have counterintuitive properties that are not obvious from the properties of the constituent nodes alone, and that seemingly minor aspects of the ways the nodes are connected can have profound and unexpected consequences for the spread and severity of disease. Third is the concern over the inability of quarantine measures to prevent increasing numbers of pathogen invasions (4). These features are shared between plant and animal (including human) epidemiology and have much in common with the ecology of other invasive species. Differences in emphasis and language arise because practical priorities differ between systems and because of biological differences in host-pathogen interactions.

To understand how using a network viewpoint develops our understanding of plant disease, it is helpful to map ideas about networks onto well-established existing ways of looking at plant disease. The familiar introductory model is near-logistic growth in a single population, which results from a number of models of infection between individual plants; if infected units can recover or can be replaced by a newly planted susceptible, the model is known as SIS (susceptible→infected→susceptible). A substantial disease outbreak, with the proportion of hosts infected increasing, at least at first, takes place only if new infections occur faster than death or recovery. No account is taken of space: It is assumed that each infected unit (plant, portion of leaf, etc.) can infect any other with equal probability. The implicit network therefore has every individual connected to every other (**Figure 1*a***). This is sometimes referred to as homogeneous mixing, and the model and its solution are at least a century old.

Homogeneous mixing is, of course, rarely realistic. In models in which each infected individual is instead connected to a number of others chosen at random—a random network—all individuals are equivalent, and each individual is equally likely to be connected to any other. An SI (susceptible→infected; i.e., a susceptible plant may pass to the infected state but then stays infectious), SIR (susceptible→infected→removed or resistant), or SIS epidemic in such a network would cause the numbers of infected or removed plants to follow a logistic curve over time, and the threshold infectiousness for the pathogen to spread indefinitely—the epidemic threshold—is the same as in a fully connected network. If the total number of nodes is small, or the average number of links per node is small, a random network may have subnetworks (components) that are not connected together.

The practical need to develop this model to describe invasions and understand heterogeneous arrangements of hosts has led to much interest in the spatial spread of disease, as either diffusive or nondiffusive dispersal over a continuous landscape (56, 60, 61) or as dispersal from cell to cell

over a grid (19, 42, 43). These are closely related and to some extent complementary approaches. A grid with small enough cells approximates a continuous landscape and is frequently used in simulations to study problems initially formulated as problems in a continuous space. A spatial grid with equally weighted connections among all nearest neighbors (a lattice) is, of course, one type of network. It is a rather exact description of some types of plantation but a less accurate description of the interactions of pathogens and hosts in most real landscapes (17, 20). A two-dimensional spatial grid is rarely an approximation of modern trade or travel networks, although it probably was when land transport costs were very high—maybe more than two or three centuries ago, before the development of canal and rail transport.

Real organisms exist at variable densities in different places. A fertile recent way to conceptualize this has been metapopulation theory. The basic unit of a metapopulation is a group of individuals that are homogeneously mixed. These base populations are then linked together by immigration and emigration. The resulting structure is again a network, with properties distinct from but related to the properties of the individual nodes (base populations) (e.g., see 47).

The theory of processes operating on networks developed in the branch of mathematics known as graph theory and in several areas of theoretical physics, particularly percolation theory. Percolation theory was motivated by flows of fluids through a network of channels in a substrate; as channels are enlarged and lengthened, a previously impervious object will allow percolation of the fluid right through the substrate. This has found application in the study of soilborne disease: To survive, a fungal mycelium or other pathogen must be able to move between soil particles to the next host (46).

KEY CONCEPTS IN NETWORK THEORY

In a general network, the pattern, direction, and strength (weight) of links between nodes can be specified as a matrix (the adjacency matrix) in which rows and columns both correspond to the same list of the components of the network, and the entries indicate the existence or nonexistence, direction, and strength of a link between the nodes. Mathematical methods for working with matrices as single entities are well developed, and the representation can deal with completely arbitrary patterns of connections between nodes, and therefore with connections estimated from empirical data. Computational methods for dealing with matrices with a large proportion of zero entries are also well developed. The challenge in plant pathology, as elsewhere (11), is to estimate the actual network structure associated with a pathogen threat; omitting a type of linkage can lead to serious mistakes.

A one-dimensional arrangement of individuals or subpopulations, such as along a section of riverbank, would be represented by a matrix in which the central few diagonals were occupied but all other connections were zero. If individuals are connected at random without regard to space, the matrix will have a random pattern of entries. An arrangement in which there are occasional long-distance connections but most connections are spatially close has discernible diagonals but occasional nonzero entries scattered over the matrix (**Figure 2**). Such an arrangement, of course, is not restricted to the same number of connections to and from all nodes.

A general network may be made up of a several sets of nodes that have no links between the sets. This remains a useful concept because a single new link might merge two sets (**Figure 1**). A set of connected nodes is described as a component of the network.

The most mathematically tractable networks are those in which all links are of equal importance (unweighted), allowing flows in both directions (undirected). This means that the matrix describing the network has only ones and zeroes and is symmetrical around the diagonal line that shows the connections of nodes to themselves. For many epidemiological purposes, flows are very different in

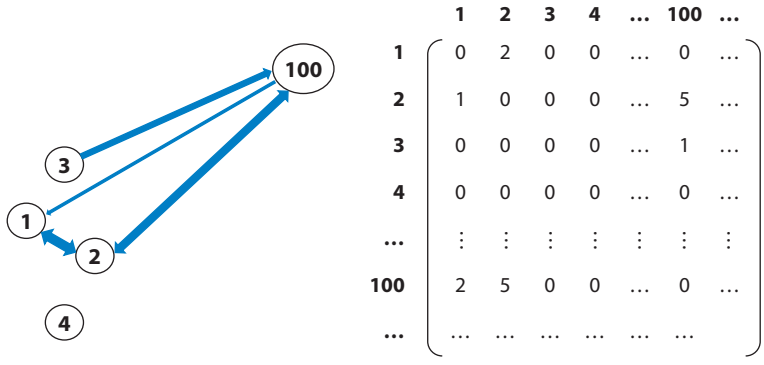


Figure 2

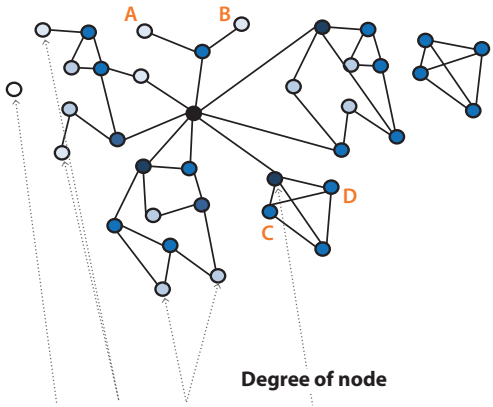
A network and its matrix representation. Node 4 is not connected or linked to any other in the selection shown: All entries in its row and column are zero. Node 3 has an out-link to node 100, and in the third row the column for node 100 contains the corresponding in-link. Nodes 2 and 100 have a heavily weighted symmetrical link, despite their physical separation, and there are corresponding entries in row 2, column 100 and row 100, column 2. In the application that we imagine the network to be used in, it is not meaningful to have a connection from a node to itself, so all the diagonal entries are zero.

opposite directions, e.g., the Netherlands exports very large numbers of flower bulbs but does not import very many. This means that matrices describing the networks relevant to plant pathology should usually be both directed (i.e., asymmetric) and weighted, with entries proportional to the flow between the nodes. Not all general analytical results apply to such asymmetric weighted networks, although many do. In any real application, the weights given to links need careful consideration: Ellis et al. (14) and Meentemeyer et al. (37)—considering, especially, *Phytophthora ramorum*—stress the need to incorporate all obstacles and routes of movement when considering the weight to be given to links between nodes representing populations, given that 200 meters over a mountain may allow much less movement than 20 kilometers down a river.

DESCRIBING NETWORKS

Much effort has gone into finding ways to describe large and complicated networks that allow common features to be extracted and seemingly diverse structures to be compared with a manageable amount of information. The degree distribution of a network is the complete frequency distribution of the number of connections per node (**Figure 3**), i.e., the table detailing the proportion of nodes having particular numbers of connections. For example, a two-dimensional spatial grid (lattice) usually has 4 connections per node except at the edges, so the frequency of nodes with degree 0, 1, or >5 (i.e., zero, one, or five or more connections) is zero; only nodes at the edges have degree 3 (three connections), and only the four corner nodes have degree 2. General networks can have nodes with degrees that vary widely. The degree distribution can be summarized by statistics such as the mean degree (number of connections per node) and variance in degree between nodes. For an asymmetric network—e.g., anything in which the pathogen vector is a trade flow—there are two degree distributions: one describing flows into a node, the other describing flows out of a node.

A degree distribution is, of course, far from a complete characterization of a network, and many other descriptors have been developed. An important feature in the flow of disease through a network is the extent of clustering. In a highly clustered network, nodes connected via a neighbor



Degree of node	0	1	2	3	4	5	6	7	8
Number of nodes	1	4	10	17	4	0	0	0	1
Frequency (%)	3	11	27	46	11	0	0	0	3

Figure 3

Network descriptors. A small, three-component network is shown along with its degree distribution. Node shading indicates degree, from lightest (degree 0) to darkest (degree 8). Dotted arrows indicate typical nodes with particular degrees. The node with degree 8 links together much of the network. Nodes A and B are not clustered; nodes C and D are.

also have a direct connection—my friends' friends are my friends too. This feature can be characterized by various types of clustering coefficients; it is important because infected individuals' contacts tend to already be infected, so the pathogen has a lower effective rate of reproduction. Also important is the mean and variance of the (weighted) distance between randomly selected nodes, which, in part, captures the notion of how far the network is from a fully mixed population.

In a conservation perspective, the total area accessible to a threatened population is important, and the parameters needed to describe a habitat network may differ from those most used in other areas of study. Baranyi et al. (3) compare a large set of proposed indices that provide weights for different landscape elements and also provide software to calculate these indices on the basis of pixel-based maps. What appears to be an unsolved problem is the optimization of a landscape so that individual host plant or animal populations are least susceptible to extinction because of demographic stochasticity or inbreeding while simultaneously being least vulnerable to invasions of novel diseases whose characteristics do not allow an endemic persistence.

A number of features relevant to epidemiology can be derived directly from the adjacency matrix or its representation in terms of eigenvalues and eigenvectors. The eigenvectors represent particular distributions of whatever the network exchanges over the nodes, which are so arranged that flows over the network do not change the relative amounts at each node—what has flowed out is balanced by what flows in, except for a general increase or decrease over the entire network. The eigenvalues are the overall multipliers applicable to the particular eigenvector. It is fairly natural from this description that—at least in some models—the value of the largest eigenvalue can determine whether a pathogen can increase in a particular network (see sidebar, Can an Epidemic Develop Within a Network of Hosts). The eigenvector associated with the largest eigenvalue can also be used as an indicator of how much network flow travels via each node, i.e., its centrality. Several other measures of centrality are available, but the key idea is that the importance of nodes to the behavior of the network can be characterized and ranked.

CAN AN EPIDEMIC DEVELOP WITHIN A NETWORK OF HOSTS?

Both Wang et al. (62) and Moslonka-Lefebvre et al. (40) present proof that the minimum infectivity required for an SIS epidemic to develop in an unweighted network is simply related to the largest eigenvalue of the association matrix of the network. Wang et al. assume a symmetric network structure; Moslonka-Lefebvre et al. allow asymmetry. In fact, both proofs carry through in essence with a weighted network structure, provided the weights have been normalized to represent their effect on the average rate at which an infected node infects a susceptible node. Thus, the effect of altering a network structure can be characterized by the effect the change has on the largest eigenvalue of the association matrix.

NETWORK STRUCTURES

Naturally arising networks can have surprising properties. Informally, the property that has drawn most attention is the small-world phenomenon, where in a huge population—even up to billions—it is possible to find a path that links node to node that is only a few steps long. This insight is usually made concrete by the phrase "six degrees of separation:" It is said that only approximately six acquaintanceships separate any two people, anywhere in the world. The implications of this for disease spread in a trade network are profound, given that modern trade and travel networks have the small-world property (35).

The feature of a network leading to the small-world phenomenon is the presence of a proportion of nodes with connections into spatially or topologically distant parts of the rest of the network. If infected, these nodes can spread this infection to parts of the network that would otherwise be safe from infection; from these areas, the pathogen may in turn reach other parts of the network. For plant pathogens spreading mostly to nearby hosts, the means by which long-distance links are generated include human or bird transport of spores or wind transmission (with or without an insect vector) (7, 26).

A further focus of interest has been the idea of a scale-free network. These are networks that have degree distributions in which the frequency of nodes of greater and greater degree declines in inverse proportion to a small power of the degree. For example, nodes of degree 2 might be four times as common as nodes of degree 4 and nodes of degree 20 four times as common as nodes of degree 40. This means that there is a very small—but nonzero—proportion of nodes with very, very large numbers of connections. In an infinite network like this, there would be a tiny proportion of nodes connected to any number, however large, of other nodes. Real networks are finite and cannot be strictly scale free, but the model has been proposed as useful in a number of contexts.

INVASION AND PERCOLATION THRESHOLDS

The literature on epidemiology using networks is dominated by considerations of the pattern and possibility of invasions of newly emerging diseases. This is clearly very important, but it might also be useful to set more routine disease management in a network context. For example, questions about the relationship between the design of plantations or crop landscapes and (endemic) disease susceptibility (e.g., see 59) might find useful answers if looked at from a network viewpoint. Even in a regular lattice network, such as an array of plants growing on a grid, Handford et al. (22) showed that heterogeneity in real root systems, and competition leading to negative correlations in the root extent of neighbors, made it (slightly) harder for a root pathogen to invade the planting.

The notion of a giant connected component in a network is of considerable heuristic use in understanding invasion and spread in a network. A giant component is one that grows in size in proportion to the total network size. The meaning is most easily visualized by imagining a network developing among many previously unconnected nodes. For example, as seed of a previously local or garden crop is swapped or sold more widely, new links are set up between previously isolated host populations. To begin with, small groups of nodes form; as new nodes are added, new small groups form, and some of the existing groups enlarge. As the number of nodes (gardens) in the network increases, the size of the groups increases moderately, with new groups forming to absorb the growth. At some (precisely defined, although depending on the detail) average number of links, a giant component starts to form. At the threshold number of links, this is not particularly large. However, if the number of links per node is increased further, more and more of the small groups get joined to the giant component, so it becomes more and more distinct from the other groups, beginning to deserve the name giant. The likelihood of having a giant component composed of a large proportion of the nodes increases with the average connectance level of the network.

The giant component is not of abstract interest only. If a pathogen enters a highly connected network with a giant component composed of a large proportion of the nodes, it is likely to enter the giant component. If it does so, most of the giant component, which is most of the network, becomes infected. The issue is further complicated in a directed network, such as a trade network, because linkages out do not coincide with linkages in, and therefore a network component may have portions linked only as recipients (for example, private gardens for an exotic), portions linked only as donors, and portions linked both ways (strongly linked).

Initial theoretical work on small-world networks formed from lattices with a random fraction of long-distance connections gave the shortcuts an equal probability of starting anywhere and ending anywhere on the lattice. Newman et al. (45) studied a SIR disease model where susceptible plants become infected and are then removed (without being replaced). Using ideas from statistical physics, they showed that a spreading epidemic could form in conditions where it would otherwise not be possible. This happened when the combined probability of spreading to a neighbor and connecting to a distant site exceeded the average number of nodes infected by a single outside infection that eventually died out. That is, if an infection that infects a few others but dies out has a sufficient probability of jumping across the lattice, it may spread to a large proportion of the nodes.

A more startling discovery was that in networks with a scale-free degree distribution, there is no threshold infectivity for a disease to invade a connected portion of the network (31, 48). More generally, high variance between hosts in the number of contacts they have tends to increase the likelihood of epidemics (e.g., see 63). This arises for the following reason. If an infection reaches a very highly connected node in an approximately scale-free network, it is likely to infect many other nodes. For example, suppose that on average an infected host infected 0.005 of the hosts connected to it. If each individual is on average connected to 5 others, an infected host leads on average to only 0.025 further infections, and the pathogen quickly becomes extinct. But if in a scale-free network the pathogen starts at or reaches a node connected to 1,000 others, it leaves 5 descendent infections; and if it starts at or reaches a node (perhaps a hundred times rarer but—because of the scale-free property—not nonexistent) with 10,000 connections, it leaves 50 descendants. Many of these are on highly connected nodes because those are the ones to which any node (including highly connected ones) is likely to be connected.

From the point of view of disease, this is a very unfortunate network structure because a rather weakly infectious pathogen can become a serious problem; but from the point of view of an airline, an airline passenger, or an internet search it is ideal: The whole world can be reached in a few steps. Notice that for a trade network in which transport costs are very low, an approximately

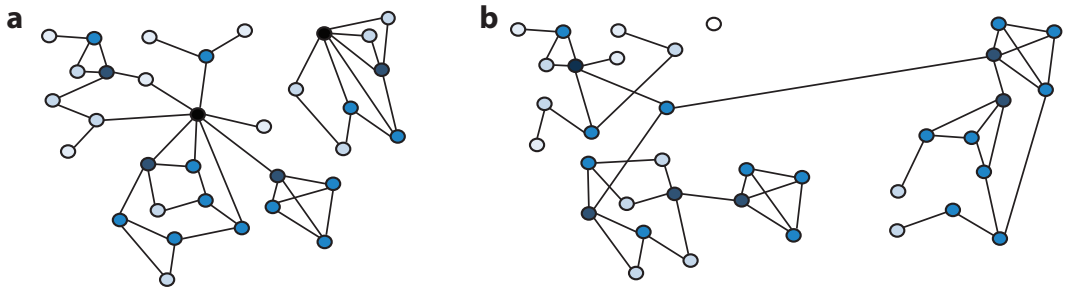


Figure 4

Simplified versions of network fragments to illustrate the difference between (*a*) scale-free and (*b*) small-world properties. In panel *a*, there are two components, within which there are hubs; but the components might remain separate if geographically or economically (e.g., different corporations) separate. In panel *b*, the most connected nodes (*darkest shading*) are not necessarily connected over long distances, but a proportion of links from some nodes, not necessarily very highly connected, cover long distances or connect otherwise disconnected components.

scale-free network is also small-world; but it is, in principle, possible that a scale-free network on an underlying geographic surface could encompass a number of disconnected components. Small-world and scale-free are not synonymous (**Figure 4**) (28).

The deduction that there was no threshold infectiousness for disease to spread and persist in a scale-free network has been used to argue that the spread of some human pathogens, such as HIV, cannot realistically be prevented. However, real networks are finite. Furthermore, if degree distributions of contacts among humans are correctly estimated, the actual distributions are not scale-free and thresholds for invasion of sexually transmitted pathogens exist (29).

The situation is further complicated for the finite-sized asymmetric networks typical of trade in plants. Consider two networks with the same average connectivity but one of which has an approximately scale-free degree distribution. The threshold infectivity that allows a pathogen to invade is lower in the scale-free network, but only if it also has a positive correlation between the numbers of incoming and outgoing links to each node (41). This result could have practical implications. A network with a positive correlation between incoming and outgoing links could be made more resistant to pathogen invasion by altering the structure so as to make the correlation negative. For example, the number of sources from which plant material is obtained by retailers with many customers could be reduced. Clearly, this has other implications and effects, but it does show how the use of network theory can clarify our understanding of complex situations.

The aim of many human networks—banking, the internet, air traffic, scientific publication—is, first, to move information or goods quickly and, second, to continue to function as a whole despite accidents or attacks that destroy links or nodes. Information flow, of course, includes information about how to manage plant disease (18, 52). Likewise, conservation aims to maintain as much biodiversity as possible in the face of destruction and isolation of habitats. The aim of disease management is, with rare exceptions, the precise opposite. We want problems to propagate slowly around the network, and we want cheap targeted interventions that restrict disease outbreaks to as little of the network as possible. From the previous paragraph, we want to identify nodes with very large numbers of connections (both incoming and outgoing) or with distant connections, and to the greatest extent possible weaken the number, remoteness, and capacity of the links they have. Unfortunately, commercial pressures and the need to ensure reliable flows of food and other goods require the reverse. In a resilient economy, goods (including food) should be able to reach any part of the world in a predictable and uninterruptible way. The same contradiction arises in

conservation biology: If the landscape is managed so that desirable host species are able to move freely between patches of habitat (8, 54, 55), diseases and pests are also likely to move freely. Similarly, assisted migration to help plant populations cope with rapid climate warming (21) also makes more likely the translocation of their pests and pathogens. The challenge of designing networks that simultaneously make some things available everywhere regardless of any damage to the network but that maintain local features safe from destructive forces arising elsewhere, is one that arises in both economics and ecology (36).

A real power-law or small-world network still conforms to the intuition that a pathogen needs a certain level of infectiveness before it can spread. Lack of such threshold infectivity could only be completely realized in an infinitely large network. Given that no real network is infinite, the theoretical lack of a threshold is instead a prediction that networks with degree distributions that include very highly connected nodes are vulnerable to weakly pathogenic organisms. In particular, vegetatively reproduced crops distributed from improvement centers can have a node (the improvement center) connected to almost every other node. The value of the abstraction is that it points to the need to consider the actual network, and the nodes exposed to risk of infection, rather than just the pathogen, when dealing with a disease management problem; it also emphasizes that there is often more to disease management than seeking a threshold for pathogen increase. For example, bottom-up localized plant breeding programs are less likely to result in hubs with many long-distance connections. It is usually the case that a successful pathogen is endemic where conditions are suitable for it. Because elimination is rarely possible with realistic levels of effort, the aim of management is to keep the effects of a pathogen primarily predictable and secondarily as low as possible.

APPLICATIONS IN PLANT DISEASE EPIDEMIOLOGY AND MANAGEMENT

The use of networks in plant disease epidemiology can be seen as developing in two directions. First, there is an interest in re-examining the general rules and ideas developed from analytically soluble or numerically tractable models in the light of network theory, and in examining the commonalities between animal disease models, conservation models, and problems in plant disease management. Second, there is greatly increased effort to characterize the structure of real networks through which pathogen infection can flow. This has two primary aspects. We can seek to design real landscapes so as to be less permeable to disease. Alternatively, we can seek to discover where to place effort in a trade network so as to reduce disease flow in a more effective and efficient way.

Hubs, or highly connected nodes, with distant connections are responsible for the scale-free phenomenon, and an obvious strategy for disease management is to remove them from the network or limit their geographic connectedness to try to alter the structure so as to slow and limit pathogen spread. Unfortunately, the countervailing economic forces often make this option unavailable. In this case, the obvious strategy appears to be to concentrate resources on managing disease in hubs and monitoring their links to prevent disease spread. However, this is not necessarily the most efficient way to minimize the impact of disease. In conservation biology, Chadès et al. (8) show that with limited resources, it is better to work on the easier problem of controlling disease (or an invasive species) in less connected nodes because the pressure of infection on them is lower. Once more isolated nodes are free of the undesirable species—for example, a pathogen—then the problem of managing highly connected nodes becomes easier because their high connectedness does not lead to continual reinfection; similarly, in a simplified two-patch model with limited management resource (53), it is more efficient to control the less-infected patch first, because it does not act as a source for the more-infected patch.

Margosian et al. (34) sought to understand the network of connections within the crop growing areas of the continental United States. US agricultural data are gathered at the scale of counties—units of approximately 1,000–10,000 km^2—of which there are around 3,000, and Margosian et al. (34) used these as nodes in a representation of the continental United States as the network of crop hosts seen from the viewpoint of a host-specific pathogen. They used the proportion of land sown to the crop, together with the distances between county centers, to measure the ease with which pathogens can move between nodes, describing this as the relative permeability of the landscape to the pathogen. For cotton and maize, there were large contiguous areas that split into disconnected components if the pathogen was assumed to need a higher and higher density of crop in order to move between counties. For wheat, there were three major areas, apparently because of low levels of connection afforded by occasional crops. The soybean network was subsequently used in an elegant study to reduce sampling effort during the year to guide protective spraying against soybean rust, *Phakopsora pachyrhizi*. Sutrave et al. (58) showed that by restricting monitoring of unsprayed sentinel plots to plots that were both highly connected and, based on previous seasons' infection, at high risk, very large reductions in monitoring effort could be achieved with minimal loss of effectiveness.

Two possible general policy options can also be envisaged from Margosian et al.'s analysis. First, it might be possible to discourage growth of certain crops in marginal areas that afford little profit but much opportunity for disease flow so as to make key production areas more independent. This involves a degree of central regulation or joint agreed action by geographically distant farmers, which could be difficult in some countries or among different countries. However, a policy similar to this has been effective in reducing the impact of *Puccinia striiformis*, yellow rust of wheat, in some regions of China (32) and has been suggested as a way to slow or halt the advance of the invasive and damaging cane toad (*Rhinella marina*) in northern Australia [by fencing water troughs fed from tube wells (16)]. Second, in the event of invasion of a large region by a new disease, it might be possible to concentrate management resources not on the major growing areas but instead on areas that function as connections between more important production zones. This is, of course, only possible where the pathogen does not have long-distance dispersal mechanisms or where long-distance dispersal mechanisms occur very occasionally. If it does, Minor & Gardner (38) suggest the best balance of resources is to concentrate on the largest patches. Again, large-scale cooperative decision-making methods are needed. Network theory can be instrumental in improving the dissemination of these innovative approaches to large-scale crop disease management.

Trade in plants and plant products across the world and within countries has grown enormously in the past few decades. A full description of what is occurring is difficult to obtain because commercial confidentiality limits the statistics that can be collected and published. Furthermore, the administrative regions in which data are collected are frequently inappropriate from an epidemiological point of view—a province in some jurisdictions may be larger than an entire country in others. However, aggregated statistics are widely available and provide useful information (**Figure 5**) (15). In plant biosecurity, the concept of pathway analysis (10) has now become part of the mainstream, thereby enabling the identification of high-risk commodities (e.g., firewood, plants for planting, seed) common to many different pests and pathogens.

Recently, Brockmann & Helbing (5) introduced a powerful representation of the airline passenger network as an ordinary continuous space with distances between nodes set inversely by a carefully chosen measure of passenger numbers. Human disease epidemics then spread on this equivalent space as a simple wave of advance; by mapping a developing outbreak onto maps of this space centered on different origins, they were able to identify the origin of several epidemics from single time data on prevalence. The method is elegant and could be applied to trade networks if trade is the dominant method of dispersal of a pathogen; it is less obvious how it could be applied

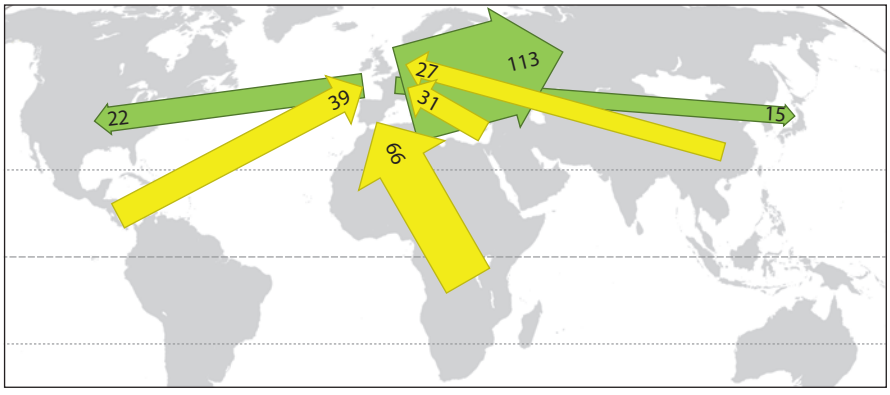


Figure 5

Partial representation of 2012 trade flows in potted plants to (*yellow*) and from (*green*) the European Union, in thousands of metric tons (15). As is typical of trade flows in particular products, the in-links and out-links are almost entirely different; this provides particularly effective shortcutting for pathogen movement in some directions but slows spread in others.

where long-distance transport is stochastic, as in the case of wind-blown pathogens (7). On much longer time-scales and whether flow is predictable or sporadic and stochastic, the historical pattern of spread of organisms over a host network can now often be inferred from phylogenetic data (1); in this case, it is not obvious that a distance mapping turning the spread into a regular wave could be found that was independent of the actual spread pattern.

A generalized network analysis (12, 27, 39) suggests, broadly, that it would be sensible to place quarantine efforts on hubs or on connections between major hubs. However, wherever effort is placed, a disproportionate increase in quarantine effort is needed to keep the rate of flow of pathogens across trade links constant as the trade through links increases. For example, suppose we are currently able to control the flow of pathogens so that one incursion per year occurs by an inspection regime that detects 80% of contaminated shipments, and 100 shipments are made annually; that is, one in 20 contaminated shipments causes an incursion of disease. If trade increases by a factor of 10, we need to detect 98% of contaminated shipments to keep the incursion rate constant. In the face of a tenfold increase in volume, it is likely to be hard to achieve such an increase in inspection efficiency. Switching from a random sampling strategy, where this is used, to a strategy focusing on links to and from hubs would increase efficiency, but there remains a daunting practical problem that does not seem to be of high political priority.

Use of a network may improve our ability to describe patterns of spread of disease, but substantial idealizations remain necessary. This is both because data are not perfect and rarely designed for the purpose, and because computer power remains limited. As with landscape descriptions of pathogen spread, problems of scale are severe. Do we attempt to simulate every plant? If not, how big should a population be to be counted, and how should it be defined? If a grid of occupied cells is used, how big should they be? It would help if there was a single structure, but real landscapes, although not fractal (similar regardless of the scale at which they are examined), are structured on many scales, and it can be awkward and specific to a particular problem to keep detail in the right places. Brooks (6) showed a hierarchical pattern in the *Silene-Microbotryum* pathosystem, but the range of scales investigated (1–1,000 m) was quite narrow. Harwood et al. (23) used a dual structure to model *P. ramorum* spread in both the natural landscape and the horticultural nursery trade, with an explicit network connecting geographically located trading centers of various kinds, and a

geographic grid with distance-based connections between grid cells implicit in the grid structure rather than listed in an explicit network. Feedback between the two components was important, and occasional long-distance dispersal within the natural landscape by wind, soil, etc. was responsible for epidemic development if horticultural control was effective. This parallels the result of Balcan et al. (2) who showed, reasonably enough, that air travel was largely responsible for global spread of influenza but that local commuting patterns controlled local intensification; removing local commuting had little effect on spread over the globe by transport of infection to the vicinity of new airports. The problem of working on multiple scales occurs throughout the literature on epidemiology of networks. Machens et al. (33) studied infection patterns in an extremely detailed set of data on contacts in a hospital and regrained the data without much loss by working with classes of individuals rather than the raw data. Although this is an interesting study, it does not really tackle the problems posed in systems containing millions of data cells and geographic scales spanning three to five orders of magnitude that are relevant to the spread of plant disease. However, it does seem that the computer power to deal with the problem is no longer an insuperable obstacle.

In a more general model of the horticultural industry, Pautasso et al. (51) found that structural changes directly linking producers and retailers, thereby reducing the numbers of wholesalers, would render the system more resistant to spread of disease. This is an interesting observation given that the internet renders it more practical to directly connect producers and retailers, although it also provides unprecedented opportunities for hubs such as Amazon and Wal-Mart. Trade and travel networks tend to be more efficient as hubs emerge, as does the movement of pathogens. Because epidemic final size in a directed trade network is well correlated with the number of outgoing links of the node first infected, producers selling to many wholesalers can also represent a high risk (50). This allows a prioritization of biosecurity effort toward the riskiest areas.

The introduction of economics explicitly into management strategies highlights differences between animal and plant systems. Kleczkowski et al. (30), following Dybiec et al. (13), analyzed three strategies for managing individuals in a human, animal, or plant epidemic that spread on a spatial lattice or a lattice with some long-distance links added so as to give it small-world properties; they explicitly treated the bacterial disease citrus canker (caused by *Xanthomonas axonopodis*). Epidemiological details specific to certain diseases determine whether it is economically more effective to eradicate infected individuals and those individuals most likely to be cryptically infected in the local area to allow the epidemic to run to equilibrium, or to apply some treatment to all individuals. A small-world structure only rendered local treatment ineffective if the density of long-distance links was high; otherwise, it was simply necessary to undertake more rigorous eradication of latently infected hosts.

OVERVIEW

Possibly the greatest impact of network thinking on plant disease epidemiology lies in focusing attention on the whole range of scales at which management action might be taken and on bringing trade and long-distance movement inside system boundaries. It is also clear that understanding and managing the consequences of long-distance trade and travel are crucial to future plant health, and this can only be done by merging a network-based approach with established plant epidemic models. With airborne pathogens, long-distance movement by wind already made traditional descriptions in terms of waves of advance a very partial description of pathogen populations; for soilborne pathogens, the addition of human trade pathways creates an unprecedented situation. Much work has focused on calculating invasion thresholds, but in many systems exclusion of disease is an unrealistic goal. Policy and practical advice leading to the lowest and most stable levels of endemicity are needed.

SUMMARY POINTS

1. Models of plant pathosystems as networks are very flexible and make it easy to include epidemiological features within the model that would otherwise be outside the system boundary.
2. The nature of pathogen invasion and persistence depends profoundly on the network structure within which hosts are linked.
3. Plant trade and human travel networks have properties that allow rapid global invasion of plant populations by pathogens and the persistence of less-infective pathogens than would otherwise be the case.
4. Effective biosecurity and quarantine should be adapted to the pathways present in the trade network being regulated.
5. We lack good data on many aspects of network structure for most pathosystems.
6. In many systems, exclusion of disease is an unrealistic goal. Properties of networks that minimize damage due to persistent pathogens deserve attention.

FUTURE ISSUES

1. Climate change will alter host susceptibility and distribution, environmental suitability for diverse pathogens, and network contact structures. Can network theory offer guidance as to how to prepare for these changes?
2. What insights can be gained by explicitly modeling changes, for example, in commercial trade patterns or agricultural structure, in the linkage pattern of a network over time?
3. Most trade flow and disease incidence data are collected according to political boundaries, which correspond to jurisdictions but not to natural biological units. How can we reconcile the need for biologically relevant data with arbitrarily sized and boundaried political units?
4. Vaccination or isolation and cure are effective ways to eliminate some human diseases but are often inappropriate for plant disease. A focus on threshold infectivity is therefore not always suitable. What can network theory say about the best ways to limit endemic disease levels and minimize fluctuations in disease levels?
5. How can we organize and lobby for politically acceptable changes in agricultural landscape linkages and trade?
6. What practical interventions best reach a balance between avoiding host fragmentation and permitting pathogen flow in a fragmented landscape?
7. Real pathosystem networks may have millions of nodes and more connections. Modern computing power can probably handle this, but how do we best organize research in pathology to fund, build, and study these models?
8. Is there a reliable way to know when we have identified the links between host individuals or populations so that we are working with a correctly identified network?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Karen Garrett, Tom Harwood, Ottmar Holdenrieder, Mike Jeger, and Mathieu Moslonka-Lefebvre for their long hours of discussion. We apologize to all those whose ideas did not make it into the limited space available.

LITERATURE CITED

1. Ali S, Gladieux P, Leconte M, Gautier A, Justesen AF, et al. 2014. Origin, migration routes and worldwide population genetic structure of the wheat yellow rust pathogen *Puccinia striiformis* f. sp. *tritici*. *PLoS Pathog.* 10:e1003903
2. Balcan D, Colizza V, Gonçalves B, Hud H, Ramasco JJ, Vespignani A. 2009. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. USA* 106:21484–89
3. Baranyi G, Saura S, Podani J, Jordán F. 2011. Contribution of habitat patches to network connectivity: redundancy and uniqueness of topological indices. *Ecol. Indic.* 11:1301–10
4. Brasier CM. 2008. The biosecurity threat to the UK and global environment from international trade in plants. *Plant Pathol.* 57:792–808
5. Brockmann D, Helbing D. 2013. The hidden geometry of complex, network-driven contagion phenomena. *Science* 342:1337–42
6. Brooks CP. 2006. Quantifying population substructure: extending the graph-theoretic approach. *Ecology* 87:864–72
7. Brown JKM, Hovmøller MS. 2002. Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* 297:537–41
8. Chadès I, Martin TG, Nicol S, Burgman MA, Possingham HP, Buckley YM. 2011. General rules for managing and surveying networks of pests, diseases, and endangered species. *Proc. Natl. Acad. Sci. USA* 108:8323–28
9. Chaiboonchoe A, Jurkowski W, Pellet J, Glaab E, Kolodkin A, et al. 2013. On different aspects of network analysis in systems biology. In *Systems Biology*, ed. A Prokop, B Csukás, pp. 181–207. Dordrecht, The Neth.: Springer
10. Colunga-Garcia M, Haack RA, Magarey RD, Borchert DM. 2013. Understanding trade pathways to target biosecurity surveillance. *NeoBiota* 18:103–18
11. Danon L, Read JM, House TA, Vernon MC, Keeling MJ. 2013. Social encounter networks: characterizing Great Britain. *Proc. R. Soc. B* 280:20131037
12. Dehnen-Schmutz K, Holdenrieder O, Jeger MJ, Pautasso M. 2010. Structural change in the international horticultural industry: some implications for plant health. *Sci. Hortic.* 125:1–15
13. Dybiec B, Kleczkowski A, Gilligan CA. 2009. Modelling control of epidemics spreading by long-range interactions. *J. R. Soc. Interface* 6:941–50
14. Ellis AM, Václavík T, Meentemeyer RK. 2010. When is connectivity important? A case study of the spatial pattern of sudden oak death. *Oikos* 119:485–93
15. European Commission. 2013. *Working Document: Flowers and Ornamental Plants*. Brussels, Belgium: DG Agri. **http://ec.europa.eu/agriculture/fruit-and-vegetables/product-reports/flowers/statistics-2013_en.pdf**
16. Florance D, Webb JK, Dempster T, Kearney MR, Worthing A, Letnic M. 2011. Excluding access to invasion hubs can contain the spread of an invasive vertebrate. *Proc. R. Soc. B* 278:2900–8
17. Foltête J-C, Clauzel C, Vuidel G. 2012. A software tool dedicated to the modelling of landscape networks. *Environ. Model. Softw.* 38:316–27
18. Garrett KA. 2012. Information networks for disease: commonalities in human management networks and within-host signalling networks. *Eur. J. Plant Pathol.* 133:75–88

19. Gosme M. 2008. Comment analyser la structure spatiale et modéliser le développement spatio-temporel des épiphyties? [How can the spatial structure and spatio-temporal mode of development of epidemics be analysed?]. *Can. J. Plant Pathol.* 30:4–23
20. Gosme M, Lucas P. 2009. Cascade: an epidemiological model to simulate disease spread and aggregation across multiple scales in a spatial hierarchy. *Phytopathology* 99:823–32
21. Gray LK, Gylander T, Mbogga MS, Chen P-Y, Hamann A. 2010. Assisted migration to address climate change: recommendations for aspen reforestation in western Canada. *Ecol. Appl.* 21:1591–603
22. Handford TP, Perez-Reche FJ, Taraskin SN, Costa LF, Miazaki M, et al. 2011. Epidemics in networks of spatially correlated three-dimensional root-branching structures. *J. R. Soc. Interface* 8:423–34
23. Harwood TD, Xu X, Pautasso M, Jeger MJ, Shaw MW. 2009. Epidemiological risk assessment using linked network and grid based modelling: *Phytophthora ramorum* and *Phytophthora kernoviae* in the UK. *Ecol. Model.* 220:3353–61
24. Heesterbeek JAP, Zadoks JC. 1987. Modelling pandemics of quarantine pests and diseases: problems and perspectives. *Crop Prot.* 6:211–21
25. Holdenrieder O, Pautasso M, Weisberg PJ, Lonsdale D. 2004. Tree diseases and landscape processes: the challenge of landscape pathology. *Trends Ecol. Evol.* 19:446–52
26. Irwin ME, Thresh JM. 1988. Long-range aerial dispersal of cereal aphids as virus vectors in North America. *Philos. Trans. R. Soc. B* 321:421–46
27. Jeger M, Pautasso M, Stack J. 2012. Climate, globalization, and trade: impacts on dispersal and invasion of fungal plant pathogens. In *Fungal Diseases: An Emerging Threat to Human, Animal, and Plant Health: Workshop Summary*, pp. 273–96. Washington, DC: Natl. Acad. Sci.
28. Jeger MJ, Pautasso M, Holdenrieder O, Shaw MW. 2007. Modelling disease spread and control in networks: implications for plant sciences. *New Phytol.* 174:279–97
29. Jones JH, Handcock MS. 2003. An assessment of preferential attachment as a mechanism for human sexual network formation. *Proc. R. Soc. B* 270:1123–28
30. Kleczkowski A, Oles K, Gudowska-Nowak E, Gilligan CA. 2012. Searching for the most cost-effective strategy for controlling epidemics spreading on regular and small-world networks. *J. R. Soc. Interface* 9:158–69
31. Lloyd AL, May RM. 2001. How viruses spread among computers and people. *Science* 292:1316–17
32. Lu N, Wang J, Chen X, Zhan G, Chen C, et al. 2011. Spatial genetic diversity and interregional spread of *Puccinia striiformis* f. sp. *tritici* in northwest China. *Eur. J. Plant Pathol.* 131:685–93
33. Machens A, Gesualdo F, Rizzo C, Tozzi AE, Barrat A, Cattuto C. 2013. An infectious disease model on empirical networks of human contact: bridging the gap between dynamic network data and contact matrices. *BMC Infect. Dis.* 13:185
34. Margosian ML, Garrett KA, Shawn Hutchinson JM, With KA. 2009. Connectivity of the American agricultural landscape: assessing the national risk of crop pest and disease spread. *BioScience* 59:141–51
35. Marvel SA, Martin T, Doering CR, Lusseau D, Newman MEJ. 2013. The small-world effect is a modern phenomenon. arXiv:1310.2636v1
36. May RM. 2013. Networks and webs in ecosystems and financial systems. *Philos. Trans. R. Soc. A* 371:20120376
37. Meentemeyer RK, Haas SE, Václavík T. 2012. Landscape epidemiology of emerging infectious diseases in natural and human-altered ecosystems. *Annu. Rev. Phytopathol.* 50:379–402
38. Minor ES, Gardner RH. 2011. Landscape connectivity and seed dispersal characteristics inform the best management strategy for exotic plants. *Ecol. Appl.* 21:739–49
39. Moslonka-Lefebvre M, Finley A, Dorigatti I, Dehnen-Schmutz K, Harwood T, et al. 2011. Networks in plant epidemiology: from genes to landscapes, countries, and continents. *Phytopathology* 101:392–403
40. Moslonka-Lefebvre M, Harwood T, Jeger MJ, Pautasso M. 2012. SIS along a continuum (SISc) epidemiological modelling and control of diseases on directed trade networks. *Math. Biosci.* 236:44–52
41. Moslonka-Lefebvre M, Pautasso M, Jeger MJ. 2009. Disease spread in small-size directed networks: epidemic threshold, correlation between links to and from nodes, and clustering. *J. Theor. Biol.* 260:402–11
42. Mundt CC, Browning JA. 1985. Development of crown rust epidemics in genetically diverse oat populations: effect of genotype unit area. *Phytopathology* 75:607–10

43. Mundt CC, Leonard KJ, Thal WM, Fulton JH. 1986. Computerised simulation of crown rust epidemics in mixtures of immune and susceptible oat plants with different genotype unit areas and spatial distributions of initial disease. *Phytopathology* 76:590–98
44. Newman MEJ. 2010. *Networks: An Introduction*. Oxford, UK: Oxford Univ. Press. 772 pp.
45. Newman MEJ, Jensen I, Ziff RM. 2002. Percolation and epidemics in a two-dimensional small world. *Phys. Rev. E* 65:021904
46. Otten W, Bailey DJ, Gilligan CA. 2004. Empirical evidence of spatial thresholds to control invasion of fungal parasites and saprotrophs. *New Phytol.* 163:125–32
47. Park AW, Gubbins S, Gilligan CA. 2001. Invasion and persistence of plant parasites in a spatially structured host population. *Oikos* 94:162–74
48. Pastor-Satorras R, Vespignani A. 2001. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86:3200–3
49. Pautasso M, Aistara G, Barnaud A, Caillon S, Clouvel P, et al. 2013. Seed exchange networks for agrobiodiversity conservation. A review. *Agron. Sustain. Dev.* 33:151–75
50. Pautasso M, Moslonka-Lefebvre M, Jeger MJ. 2010. The number of links to and from the starting node as a predictor of epidemic size in small-size directed networks. *Ecol. Complexity* 7:424–32
51. Pautasso M, Xu X, Jeger MJ, Harwood TD, Moslonka-Lefebvre M, Pellis L. 2010. Disease spread in small-size directed trade networks: the role of hierarchical categories. *J. Appl. Ecol.* 47:1300–9
52. Rebaudo F, Dangles O. 2011. Coupled information diffusion: pest dynamics models predict delayed benefits of farmer cooperation in pest management programs. *PLoS Comput. Biol.* 7:e1002222
53. Rowthorn RE, Laxminarayan R, Gilligan CA. 2009. Optimal control of epidemics in metapopulations. *J. R. Soc. Interface* 6:1135–44
54. Rubio L, Saura S. 2012. Assessing the importance of individual habitat patches as irreplaceable connecting elements: an analysis of simulated and real landscape data. *Ecol. Complex.* 11:28–37
55. Saura S, Vogt P, Velázquez J, Hernando A, Tejera R. 2011. Key structural forest connectors can be identified by combining landscape spatial pattern and network analyses. *Forest Ecol. Manag.* 262:150–60
56. Shaw MW. 1995. Simulation of population expansion and spatial pattern when individual dispersal distributions do not decline exponentially with distance. *Proc. R. Soc. B* 259:249–57
57. Skelsey P, van der Werf W, Kessel GJT, Rossing WAH, Holtslag AAM. 2007. Multi-scale modelling of infection pressure from *Phytophthora infestans*. *EPPO Bull.* 37:313–16
58. 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:e37793
59. Suzuki SU, Sasaki A. 2011. How does the resistance threshold in spatially explicit epidemic dynamics depend on the basic reproductive ratio and spatial correlation of crop genotypes? *J. Theor. Biol.* 276:117–25
60. van den Bosch F, Metz JAJ, Zadoks JC. 1999. Pandemics of focal plant disease, a model. *Phytopathology* 89:495–505
61. van den Bosch F, Zadoks JC, Metz JAJ. 1988. Focus expansion in plant disease. 2: Realistic parameter-sparse models. *Phytopathology* 78:59–64
62. Wang Y, Chakrabarti D, Wang C, Faloutsos C. 2003. Epidemic spreading in real networks: an eigenvalue viewpoint. *Proc. Symp. Reliab. Distrib. Comput.*, *22nd*, *Florence*, *Oct. 6–18*, pp. 25–34. Washington, DC: IEEE
63. Woolhouse MEJ, Dye C, Etard J-F, Smith T, Charlwood JD, et al. 1997. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl. Acad. Sci. USA* 94:338–42

RELATED RESOURCES

Xue M, Wan Y, Roy S, Saberi A. 2008. Dynamical Network Design for Controlling Virus Spread. In *Wolfram Demonstrations Project*. **http://demonstrations.wolfram.com/DynamicalNetworkDesignForControllingVirusSpread/**