Modeling Epidemics in Seed Systems: The Case of Sweetpotato in Northern Uganda

Kelsey F. Andersen[1,2,3], Chris E. Buddenhagen[1,2,3], Paul Rachkara[4], Richard Gibson[5], Stephen Kalule[4], David Phillips[5], Karen A. Garrett[1,2,3]

[1]Plant Pathology Department, University of Florida, Gainesville, FL 32611-0680, United States

[2]Institute for Sustainable Food Systems, University of Florida, Gainesville, FL 32611-0680, United States

[3]Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611-0680, United States

[4]Department of Rural Development and Agribusiness, Gulu University, Gulu, Uganda

[5]Natural Resource Institute, University of Greenwich, United Kingdom

Corresponding authors: KA Garrett (karengarrett@ufl.edu) and KF Andersen (andersenk@ufl.edu)

Mailing address: Plant Pathology Department, University of Florida, Gainesville, FL 32611-0680, United States

Abstract

Seed systems are critical for deployment of improved varieties, and serve as major conduits for the spread of seed-borne pathogens. Vegetatively propagated crops in low-income countries are particularly vulnerable to seed degeneration, where yield is lowered through successive cycles of propagation because of pathogen accumulation. We evaluated the structure of an informal sweetpotato seed system for its vulnerability to the spread of epidemics, and its utility for disseminating improved varieties. During the 2014 growing season (April-Oct), vine sellers were surveyed weekly in the Gulu Region of Northern Uganda. Our analysis draws on tools from network theory to evaluate the potential for epidemic spread in this region. Utilizing empirical 2014 seed transaction data and estimated spatial spread, we constructed a seed transaction network, which was used to simulate the introduction of a pathogen, and evaluated the influence of both epidemic starting point and quarantine treatments on epidemic progress. Results indicate that the starting position in the network is critical for epidemic progress and final epidemic outcomes. Quarantine of 30 villages lowered epidemic progress up to 65.7%, when compared to the control (no quarantine), over 20 timesteps in 500 realizations. The percent control conferred by quarantine treatments was also influenced by the epidemic starting point. Considering equal likelihood of any node being an introduction point for a new epidemic, villages of particular utility for disease monitoring were also identified. Sensitivity analysis identified important parameters and priorities for future data collection. We compared the utility of node degree, betweenness, closeness, and eigenvector centrality for selecting quarantine locations, finding that betweenness had more limited utility. This analysis pipeline can be applied to provide recommendations for a wide variety of seed systems.

Introduction

Seed systems, both formal and informal, are a critical component of global food security. Even so, efforts to implement seed systems that work better for smallholder farmers in low-income countries have often been unsuccessful (Gibson and Kreuze 2015; Thomas-Sharma et al. 2016). Improving seed security – defined as timely access to quality planting material by all, at a fair price (Almekinders et al. 1994; Gibson et al. 2011; McGuire and Sperling 2013; Sperling 2008) – is vital for improved livelihoods, particularly for smallholder farmers. In most low-income countries, seed systems are local, with most farmers keeping seed from previous seasons, or obtaining seeds from neighbors, local traders, or local markets of unknown quality, with some instances of long distance trade (Gildemacher et al. 2009). Having robust analytic models of seed systems supports risk assessment efforts and policy development in times of system disruption caused by system shocks such as novel pest introduction, climate change, or political unrest.

Network analysis is a powerful analytic tool, used across many disciplines, with recent advances in its utility for analyzing epidemic spread in human, animal, and plant systems (Shaw and Pautasso 2014; Silk et al. 2017). Seed systems are amenable to network analysis because they are inherently networks with a suite of actors (network nodes) that move both genetic material and information through space and time (dynamic or static network links) (Pautasso 2015). Until recent years, plant disease epidemiologists have given relatively little attention to the study of seed system (and plant trade) networks, although the movement of planting material plays a fundamental role in the spread of plant disease and the persistence of epidemics (Buddenhagen et al. 2017; Pautasso 2015; Pautasso and Jeger 2008). Increasing availability of computational tools, coupled with advances in network analysis in the medical and social

sciences, make the implementation of network analysis for plant disease epidemiology more obtainable.

Network analysis can be applied to study biophysical and information networks (Garrett 2012, 2017; Pautasso and Jeger 2008). Network statistics, such as measures of node betweeness, closeness, and degree can be calculated to define key nodes and actors in a system, provide diagnostic metrics, and forecast the risk of pathogen introduction, pathogen spread, or technology adoption on a cropping system (Garrett 2012; Harwood et al. 2009; Moslonka-Lefebvre et al. 2011; Pautasso 2015; Pautasso and Jeger 2008; Sanatkar et al. 2015). There are a number of statistics available that are likely to be associated with the importance of a node in an epidemic, but it is an open question which statistics are most important for prioritizing nodes for sampling and mitigation. Understanding the structure and key nodes of seed systems can inform policy and intervention strategies, where intervention may be particularly important after recent introduction of invasive pests or pathogens (Chades et al. 2011; Nelson and Bone 2015)

Network analysis can also be used to evaluate nodes important for sampling and mitigation of the movement of pathogens or other contaminants through networks (Sutrave et al. 2012). In a study of wheat grain movement in the United States and Australia, network analysis was used to identify key locations that could be strategically targeted for sampling and management of mycotoxins (Hernandez Nopsa et al. 2015). Network modeling is increasingly being used to evaluate likely disease spread, and can provide insights into the utility of forecasting, as shown by its use for soybean rust forecasting in the United States (Sanatkar et al. 2015).

Previous studies analyzing seed system networks have focused on understanding the effects of social ties on seed system structure, and how well seed system networks may conserve

variety diversity in the landscape (Pautasso 2015; Pautasso et al. 2013). Abay et. al (2011) applied network analysis in a study of informal barley seed flows in the Tigray, Ethiopia region with the goal of informing breeding or technology deployment strategies. Network properties, such as betweenness and degree centrality, were used to characterize key nodes and their role in connecting the seed system. Each of these metrics measures the importance of a node in the network in terms of the number of connections it has, or the number of paths across the network of which it is a part. Pautasso et al. 2015 further analyzed this empirical seed network data, finding that the degree distribution of the network, and particularly the out-degree of the starting node, influenced the percentage of the network that could be reached by a new variety.

The current study draws on concepts previously developed for hypothetical and empirical large- and small-scale plant trade network epidemic spread (Buddenhagen et al. 2017; Moslonka-Lefebvre et al. 2011; Nelson and Bone 2015; Pautasso 2015; Pautasso and Jeger 2008; Pautasso et al. 2010). We expand on previously developed theory about the influence of node in- and out-degree on variety dissemination (and pathogen transmission) in small-scale, real-world seed networks (Pautasso 2015). We also apply concepts developed to assess quarantine influence on disease spread in hypothetical trade networks (Nelson and Bone 2015) in a new set of simulation experiments. This study builds on seed system epidemiology by integrating these methods and adding further analytical tools, such as using geo-spatial proximity to model village to village transactions and incorporating node attributes to prescribe control strategies. These tools are particularly important for understanding the epidemiology of pathogen spread in vegetatively propagated crops, a major limitation to yields in low- and middle-income countries.

Understanding seed system epidemic dynamics is vital because viruses and other seed-transmitted diseases are important risks to yield and quality degeneration within seed systems, particularly those of vegetatively propagated crops (Thomas-Sharma et al. 2016). Informal seed systems without healthy seed certification protocols are often at higher risk. Newly introduced viruses can be particularly severe, as methods for detection may be limited or unavailable, and host resistance may be unattainable for several years. It is important, therefore, that the risk of novel pathogen introduction into a seed system be understood for the swift recommendation of intervention (such as sampling, quarantine, variety deployment, and education). This problem was illustrated in 2011 when maize lethal necrosis (MLN) was first reported in Kenya (Wangai et al. 2012) and soon was detected in several other Sub-Saharan African countries. MLN symptom expression results from coinfection with *Maize chlorotic mottle virus* (MCMV) and a potyvirus (Mahuku et al. 2015). Since its introduction, MLN has been detected in several East African countries including Ethiopia, Uganda, South Sudan, Tanzania, DRC, and Rwanda, and up to a 22% yearly yield loss has been reported (De Groote et al. 2016; Hilker et al. 2017; Mahuku et al. 2015). Characterization of seed systems using network analysis may inform strategic intervention, such as quarantine and sampling, in these scenarios. Similarly, in 2014 cassava mosaic disease (CMD) was first detected in Southeast Asia and efforts are still ongoing to mitigate spread and deploy resistance (Graziosi et al. 2016; Wang et al. 2016).

Seed system networks are comprised of a suite of actors, or nodes, including farmers, multipliers, traders, NGOs, seed companies, breeding organizations, and communities. The connections between these nodes (links), representing formal and informal interactions, are complex and require analyses that address this complexity. Network analysis allows us to simulate multiple scenarios in known systems, including the impact of epidemic spread or

management deployment. In this study we propose a framework for analyzing such networks that can be transferred to a broad range of seed systems. In this paper we, i) analyze, as a case study, key network properties within a seed system important to regional food security; ii) evaluate variety dissemination within the network, comparing the distribution of higher-nutrient introduced varieties and landraces; iii) model the progress of a potential seed-borne pathogen introduced into the network and compare the impact of quarantine scenarios devised based on different network statistics as selection criteria for quarantine nodes; and iv) perform sensitivity analysis to determine the influence on epidemic outcomes of the parameters used to construct the village-to-village transaction networks.

Methods

Study System: Sweetpotato in Northern Uganda

This study examines sweetpotato vine transactions in Northern Uganda. Sweetpotato is a major staple food crop in many African countries, and Uganda is the second largest producer in Africa, fourth globally (FAOSTAT 2013). Sweetpotato is generally grown by women in Uganda in small plots of land, close to the household, and is important for household food security (Behrman 2011; Johnson and Gurr 2016). In the last decade, sweetpotato has increased in importance due to the introduction of a betacarotene-rich biofortified crop known as Orange-fleshed Sweetpotato (OFSP) by HarvestPlus, part of the CGIAR Research Program on Agriculture for Nutrition and Health. OFSP varieties were introduced with the goal of addressing vitamin A deficiency in women and children in this region (Behrman 2011).

Viral diseases are major biotic limiters to sweetpotato production in Uganda and throughout Sub-Saharan Africa, with the most yield-limiting being sweet potato virus disease (SPVD), which occurs when a plant is co-infected with *Sweet potato feathery mottle virus* (SPFMV) and *Sweet potato chlorotic stunt virus* (SPCSV) (Karyeija et al. 1998). Seed degeneration is the successive loss in yield over generations of vegetatively-propagated seed material due to the accumulation of viruses and other seed borne pathogens (Thomas-Sharma et al. 2016). Degeneration is highly problematic in informal seed systems where farmers tend to save seed season-to-season and where certified seed sources are rare or non-existent. Both SPFMV and SPCSV can be transmitted through vegetatively-propagated material, with evidence of yield degeneration over five generations in high pressure fields in Uganda (Adikini et al. 2015). SPVD has not yet been reported in Northern Uganda, likely because the extended dry season in this region is unfavorable for the whitefly vector (Richard Gibson, *personal observation*). Changes in climate patterns or vector range, however, could potentially extend the range of this disease into this region. The prospect of novel pathogen introduction makes it important to model epidemic scenarios to inform intervention strategies.

In Northern Uganda, sweetpotato seed material is sold in small bundles of vine cuttings. In this region, sweetpotato vine distribution is largely informal, consisting of smallholder farmers who have access to fields with adequate moisture to produce roots and vines through the extended dry season, which typically lasts from December to April (Gibson 2013). These counter-season multipliers generally produce local landraces which tend to be well-adapted white-fleshed cultivars. (Gibson 2013). Vine cuttings are not easily stored, and because of a single, extended dry season in Northern Uganda, vines need to be obtained by farmers at the beginning of each season (Gibson et al. 2011). There are also several formal institutions involved

in sweetpotato breeding and distribution in Uganda, which include the National Sweetpotato Program (NSP), the International Potato Center (CIP), private sector enterprises, and NGOs (Gibson 2013).

Survey Methods

A survey of vine multipliers and sellers was conducted in 2013-2014 in the Gulu Region of Northern Uganda. Survey methods are fully described by Rachkara et al. 2017. In the first year of the study (2013), the transactions of a small number of local multipliers were tracked throughout the season. A more complete cohort of multipliers and sellers were surveyed in 2014, the focus of our analysis. All seller names have been anonymized to protect the identity of individuals who participated in this survey. Each seller was visited weekly from the start of the growing season (April) and surveyed twice per week to record all transactions that occurred in the period since they were last visited until the end of the season (August). Volume of transaction (number of bundles), price, variety, origin of buyer, and buyer type (farmer or seller) were recorded. In this study, a small bundle refers to 50 vines cut to 40 cm in length. Large bundles are equal to 20 small bundles. For consistency, large bundles will be used to describe volume in this paper. Because of the high volume of transactions, names of individual buyers were not collected and therefore sales transactions were summarized by the village from which the buyer originated.

Seed Network Analysis

Nodes in this analysis include sellers and villages, with one set of directed links representing vine sales from an individual seller to an individual village. Villages in this region

of Uganda typically are composed of 40-60 households. The categorization of the network into two distinct types (village and seller), produces a single, bipartite graph (*g*). Although several transactions were recorded on a weekly basis, transactions were aggregated for this analysis so that links in graph *g* represent the existence of at least one transaction over the course of the season. Seller-village links in *g* are based entirely on the data from (Rachkara et al. 2017). Network properties, such as number of nodes and network density, were calculated for the 2014 season. Node measures – such as degree, closeness, eigenvector, and betweenness centrality – were evaluated for both villages and sellers (Table 1). All analyses were conducted in the R programing environment (R Core Team 2017) using custom code along with software packages including igraph (Csárdi and Nepusz 2006).

Incorporating village-to-village spread

To incorporate epidemic risk in the seed system outside spread through sellers, a second set of directed links was evaluated representing village-to-village spread through informal exchange of planting materials or vector movement, composing a new graph (*e*). These village-village links were modeled using a power law function, and combined with transaction links to form an expanded adjacency matrix (Fig. 1). The power-law function captures the tendency for seed exchange to be more likely between villages (*V*) that are geospatially proximate. That is, farmers have a higher chance of exchanging seed with farmers in neighboring villages than with farmers in villages that are far away. A link represents a seed movement event (transaction) or movement of vectors that may be viruliferous. The power law equation used here is $y = ad^{-\beta}$, where d is the Euclidean distance between a pair of villages, y is proportional to the probability of movement between the villages, and a is a scaling factor. As default values, we used a = 1 and $\beta = 1.5$. The seed transaction threshold (*t*) was set to 0.01, such that if $y_{(Vx, Vy)} > 0.01$ a link was

created between villages x and y. Village to village links were established once and the same underlying adjacency matrix was used across simulation experiments for a given parameter combination. In individual realizations, the underlying adjacency matrix was modified as links were maintained with a fixed probability, as described below. A sensitivity analysis was later performed to determine the influence of the parameters β and *t* on the analysis. It is important to note that only villages of farmers who purchased vines from sellers surveyed in this study were included in this analysis, and there may be villages in this region that contribute to risk, but were not included.

Modeling Epidemic Spread

This study draws on tools from graph theory to better understand the potential for epidemic spread within a real-world seed network. This analysis utilized both empirical 2014 seed transaction data and estimated spatial transactions to simulate the invasion of a hypothetical seed-borne pathogen. We performed simulation experiments to address the following questions. (1) What are the optimum sampling locations to detect disease spread, given equal likelihood of disease introduction at each node? (2) How does the risk of epidemic spread differ between sellers who sell landraces and sellers who sell orange-fleshed sweetpotatoes? (3) What is the influence of quarantine, where nodes are selected to have their role in epidemic spread removed? In each experiment, simulations were conducted over 20 timesteps and repeated 500 times. The probability of pathogen transmission (P_t) occurring in each of the generated links was set to 0.10. The probability of persistence (P_p), or the diagonals of the adjacency matrix, were fixed at 1.

The value of villages as effective monitoring locations for future epidemics

We evaluated the importance of each village as a pathogen monitoring location. In this scenario analysis, all nodes (sellers and villages) were assigned an equal likelihood of being the starting point for epidemic introduction. For each possible combination of an epidemic starting node and a sampling node, we determined the number of nodes that could become infected by the time the pathogen is detected at the sampling node, as in Buddenhagen et al. (2017). In this analysis we compare only villages, not sellers, for their relative value as locations for epidemic monitoring. This analysis was repeated over 500 simulations, and the mean and variance of the number of nodes infected by the time the pathogen was detected in that village were evaluated. Summarizing over all the potential starting nodes allows for the comparison of rank of importance of each village as a potential location for monitoring, given an unknown starting location with each potential starting location equally likely. Simulations were implemented using custom R code (R core team 2017).

Comparing Epidemic Spread Between Varieties

The second simulation experiment evaluated the potential spread through the above described seed network, where the epidemic starts with a single seller that 1) sells only landraces (86% of sellers, Fig. 3) or 2) sells only orange fleshed sweet potato varieties (14% of sellers, Fig. 3). The two sellers with the highest out-degree in each category (OFSP and landrace seller) were chosen for comparison in this analysis, as representing the greatest risk for each type. It has been previously shown that the out-degree of starting node is strongly positively correlated with final epidemic outcome (Pautasso 2015; Pautasso et al. 2010). At the start of each realization (Time 1), the "starting seller" was set with probability $P_t = 0.01$ of transmitting the virus across the links established in the complete network, as described above. The starting seller maintained this

infection status throughout the realizations, and all other sellers remained uninfected. Villages that became infected after a single round of simulations maintained infected status in the subsequent time step (Time t+1) and could thus infect villages to which they had links with the same probability (P_t = 0.01) in the subsequent timestep (Time t+2). For this simulation, once infected, a node stayed infected and infectious throughout a realization (probability of infection persistence, P_p = 1; see Supplemental Figure 1). Infection was evaluated across 20 timesteps in 500 realizations, to evaluate the frequency distribution of outcomes. Simulations were carried out using custom R code (R core team 2017). To evaluate the differences in disease progress between starting nodes, the area under the disease progress curve (AUDPC) was calculated by summing the trapezoids between timesteps under the curve using the r package agricolae (de Mendiburu 2017).

The influence of quarantine on infection dynamics

In the third simulation experiment, we evaluated the influence of quarantine on disease spread over time. Quarantine was defined here as the removal of the potential for a node to transmit or acquire infection in each time step. The influence of 10, 15, 20, 25 and 30 quarantined nodes on infection spread over time was compared with the control, a scenario with no quarantine treatment. Quarantined nodes (villages only) were selected based on a trait such as node-degree, with villages with the highest node degree selected first. Again, links from sellers to villages are based on known transactions in 2014 and the probability of a link existing between villages is calculated based on a power law function of distance, as described above. To maintain consistency, the same two sellers (S_25 and S_15) that were chosen to start the epidemic in the previous experiment were used in this analysis. Simulations were implemented

as described previously, over 20 timesteps in 500 realizations. AUDPC values were calculated for each realization (as described above) compared for each treatment.

To evaluate the utility of node degree in comparison to other key network node properties for the selection of nodes as quarantine candidates, the above analysis was also performed with nodes selected based on their node betweenness centrality, closeness centrality, and eigenvector centrality, in comparison to a scenario where 30 villages were drawn at random (Table 1). The value of each villages as a monitoring location, calculated for the previously described evaluation of the utility of nodes as sampling sites, was also compared for its utility to select quarantine candidates. For each separate method for ranking the likely value of nodes for quarantine, nodes were ranked based on that method and the top ranked 15, 20, 25 or 30 nodes were selected for quarantine.

Sensitivity analysis and uncertainty quantification

To evaluate the influence of parameter choice for system-specific and often unknown features such as the dispersal kernel, we evaluated the effects of varying parameter values on the epidemic outcome (summation of infection status of all nodes at the end of 20 timesteps) for each of the parameter combinations described in Table 1. Selected parameters β and t influence the number of static links formed between villages, ultimately influencing the likelihood of transmission of a pathogen at each timestep. The probability of transmission (P_t) determines the probability of infection in each timestep. These analyses were done for both infection starting at S_15 (the most active landrace seller) and S_25 (the most active OFSP seller).

Results

Network Properties

In 2014, 27 sellers were tracked, resulting in a total of 878 individual vine sales to buyers from 99 distinct villages. Using an adjacency matrix based on aggregated transactions from sellers to villages (where the presence of at least one transaction was equal to 1 and the absence of any transaction was equal to 0), a network graph, *g*, was constructed (Figure 2a). It is important to note here that although transactions were collected over time, the presence of a single, directed link here represents at least one transaction in the 2014 season. This graph, *g*, has a total of 126 nodes and 205 links (link density = 0.013). After the addition of village-to-village links, graph *e* (link density = 0.047, Figure 2b), the number of links increased to 743, roughly representing the number of potential vine transactions in each time step in simulation experiments. Node degree, or the number of incoming (in-degree) and outgoing (out-degree) links was calculated for both seller nodes (μ = 7.6, min =1, max =42) and village nodes (μ = 12.9, min = 1, max = 60; Supplemental Fig. 2). The 30 villages with the highest node degree (both in- and out-degree) were selected as candidates for quarantine in subsequent analyses. Betweenness, closeness, and eigenvector centrality were also measured for all villages in this analysis (Table 1) and used in subsequent analysis. These four measures of centrality were positively correlated based on Pearson's correlation coefficient (Supplemental Fig. 3).

Variety Dissemination

In this data, a total of 15 cultivars were sold during the 2014 season. These cultivars were a mix of landraces and varieties introduced by the national breeding program (Fig. 3). Six of these cultivars were OFSP cultivars, and were disseminated by only two sellers, in many

individual transactions (Fig. 3). In comparison, the white-fleshed landrace, Ladwe Aryo, was sold by 25 distinct sellers in hundreds of transactions throughout the season. When the network is examined separately by variety, disaggregation becomes apparent (Fig. 4). Evaluating networks representing the distribution of the top eight varieties to villages in the network (Fig. 4) indicates that only a small number of sellers and villages are exchanging orange-fleshed varieties. It appears, from this survey, that most individuals from a single village only buy a single variety, even when they have access to multiple sellers.

The value of villages as monitoring locations

Under the scenario where each node is an equally likely starting point for an epidemic, we evaluated the value of each village as a monitoring location, so that a village was considered a more effective location if only a small proportion of other nodes were colonized before the pathogen could be detected at that village. The village with the highest value for monitoring was V_58, with the lowest mean number of villages affected by the time the pathogen would be detected there (Fig. 5).

Modeling Epidemic Progress with Variable Starting Sellers

The two sellers with the greatest out-degree for their product, S_25 (OFSP seller) and S_15 (landrace seller), were compared as starting nodes for an introduced epidemic simulated over 20 timesteps in 500 realizations (Fig. 6). We were interested in the influence of these two important types of sellers on progress through the networks. These sellers had an out-degree of 42 and 19, respectively. Progression of the infection was variable between the two starting nodes. Infection starting with the landrace seller (S_15) approached epidemic saturation (where

the maximum number of villages that can be reached, have been reached) around 50 infected nodes (or 50.5% of the 99 villages in the survey). The OFSP seller appeared to approach saturation around 70 nodes (or ~71% of the 99 villages in the survey). Overall, the mean number of nodes infected by time 20 was lower (μ = 50.19, sd = 2.11) when infection started with the landrace seller than for the OFSP seller (μ = 69.40, sd = 2.49). Disease progress was, therefore, variable between starting nodes as well with AUDPC being higher for the infection starting with the OFSP seller (μ = 977.05, sd = 54.63) than for the landrace seller (μ = 697.025, sd = 57.21).

Influence of Quarantine Scenarios on Disease Spread

Another important question when considering an emerging epidemic is the influence of timely intervention or management deployment. In this study we evaluated the influence of quarantining villages so they cannot become infected or transmit infections to other villages. This attempts to simulate the common practical scenario of phytosanitary quarantine by regulatory agencies after the detection of pathogens in new regions. Nodes were selected based on node degree rank (with comparison to other methods of selecting nodes in the next step). Node degree for the 30 nodes selected as quarantine candidates ranged from 60-17 (with the highest selected first). Epidemic simulations, as previously described, were repeated for both seller (S_25 and S_15), for each quarantine scenario based on degree (Fig 7). For infection beginning with a landrace seller (Fig 7a) the mean number of nodes infected by the end of the simulation (timestep 20) decreased by 7.03%, 17.06%, 29.34%, 47.81%, and 65.74%, for each quarantine treatment (quarantine of 10, 15, 20, 25, and 30 nodes, respectively) when compared to the control of no quarantine. A similar trend in decrease was observed for AUDPC values calculated for each realization of each scenario (Fig 7b).

For the OFSP seller, the mean number of nodes infected at the end of time 20 decreased by 4.51%, 11.89%, 20.23%, 31.17%, and 41.20% for quarantine of 10, 15, 20, 25, and 30 nodes, respectively (Fig 7c). Across treatments, final epidemic outcome was higher for the case of infection starting with the OFSP seller (S_25), and although a similar effect of quarantine was observed, the percent control imposed by the treatment appeared to be less than the percent control conferred to the infection starting with the landrace seller (S_17).

The influence of different centrality measures on quarantine outcomes

To test the effect of the method chosen for selection of quarantine candidates on epidemic outcomes, we compared node degree centrality (used in analysis described above) to closeness, betweenness, eigenvector centrality, the estimated value of the location for monitoring, and villages drawn at random, on mean AUDPC outcomes across realizations (Figure 8). It was not surprising that, after quarantining at least 15 nodes , the “smart” selection criteria (centrality measures and monitoring efficiency) outperformed the “naive” selection criteria (randomly selecting villages to quarantine) (Supplemental Figs. 5 and 6). It was somewhat surprising, however, that this trend was not clear for the 10 node quarantine treatment, indicating that the invasion may be able to “overcome” quarantine of less than 10 nodes, even if they are those with the highest centrality scores.

In this system, it appears that closeness, eigenvector, and degree centrality are equally good predictors of quarantine hubs and confer a similar reduction to AUDPC (Figure 8). This is consistent for each of the starting nodes. Interestingly, independent of starting node, betweenness centrality consistently resulted in a higher mean AUDPC when 20 and 25 nodes were quarantined.

Sensitivity analysis and Uncertainty Quantification

Sensitivity analysis was performed to examine the influence of key parameters, used to model the network and epidemic spread, on final epidemic outcomes, over 100 simulations. Both the exponent of the power law β and the threshold for link formation t influence the epidemic progress curve (Fig. 9). As β and t decrease, we see an increase in the number of nodes infected by the end of the epidemic, time 20 (Fig. 9). At time 20, with $\beta = 0.5$ and $t = 0.001$ all villages are infected, in every realization. Alternatively, when $\beta = 2.5$ and $t = 0.20$ little or no epidemic progress is made. These dramatic differences are likely due to the number of links formed between villages with each parameter set. With the network highly saturated with links, as is the case with $\beta = 0.5$ and $t = 0.01$, the epidemic can move across the system more rapidly.

Discussion

Seed system assessment provides guidance for strategic intervention in times of need. Understanding key system properties prior to emergencies is essential for quality development aid or plans for recovery from a new epidemic. The analytic framework used in this study can be used as a template to characterize and model other seed systems, particularly those of vegetatively-propagated crops with high potential for seed-borne pathogen introduction and spread. It is clear from this study that sweetpotato vines in Northern Uganda are sold in a highly connected informal seed system. There exists a trade-off of centrality within this seed network. That is, it is favorable to be a village with high in-degree (many links) because of increased access to a diversity of cultivars (such as OFSP) and sellers. However, this high connectivity may make a village or individual more susceptible to pathogen introduction.

This study used scenario analysis to characterize villages for their utility as pathogen sampling hubs based on a simulation where all nodes had an equal chance of being the point of epidemic start. Based on this simulation, villages were identified as potential monitoring targets because of the high likelihood of pathogen detection in these locations before the rest of the network becomes colonized. This method can be used to identify sentinel villages (nodes) to focus sampling efforts, and can be particularly useful in informal systems where production and distribution are not centralized, and therefore are constructed due to a number of social and economic factors. This method can serve as a complement to new diagnostic technologies, such as loop-mediated amplification (LAMP) assays, which are becoming increasingly available to practitioners in the field and have the potential for rapid on-site detection of viruses and other pathogens.

When examining vine distribution by variety in this network, it is clear that most surveyed cultivars are not well disseminated throughout the network. Although much effort has been made to promote modern varieties, in a 2015 study, 89% of 51 fields surveyed grew local landraces (Obong et al. 2017). Based on the observed data from this study, we cannot be sure if this is because of preference or availability of planting material, or a combination of the two. Understanding the reason behind this is important, because the adoption of OFSP is closely tied with human nutritional benefits. Similar methods to those described here may be utilized to target villages for development projects that aim to disseminate varieties to key hubs and maximize their distribution.

This study is consistent with previous results (Pautasso 2015; Pautasso et al. 2010), which found that sellers with high node degree have a greater influence on final epidemic outcomes. Additionally, we found that the starting seller greatly influences the progress of

epidemics and also, through quarantine scenario analysis, epidemic control. This is an important consideration as pathogen dissemination through vegetatively propagated seed material is closely tied to variety dissemination, indicating that there is less chance that an introduced variety (such as the OFSP cultivars that were introduced in this region) will reach all farmers that need them. This suggests that agencies should consider node degree when deploying new varieties. Quarantine appeared effective at slowing epidemic progress, when sufficient nodes were included, but not sufficient for halting epidemic spread under any scenario. Future research should consider incorporating the influence of the deployment of other components of an integrated seed health strategy (Thomas-Sharma et al. 2017), such as positive selection, resistant varieties, clean seed, and education on disease progress.

It is important to note that this study was limited to data related to sellers that participated in weekly monitoring in the Gulu region of Northern Uganda. Although this dataset was expansive, there is a potential that there were other sources of vines, or sellers, that were not captured. In addition, there are other villages with sweetpotato cropland in the region of study that were not included in in this analysis because they were not associated with a buyer in the dataset. Furthermore, links in this study were unweighted and fixed through time. It may be valuable to examine the influence of a weighted, dynamic contact structure on epidemic progress in seed networks. To our knowledge, this has not been examined in the literature. The model used here is a case of an SI model, where once nodes become infected, they remain infected (and infectious) through the duration of timesteps. In human epidemiology, diseases that follow an SI model are difficult to control. Because there may be options for epidemic recovery in this system (through reversion, rouging, or positive selection), future studies may include a recovery term in the model.

Understanding the network structure of seed systems provides a unique lens for understanding variety distribution and pathogen risk. Results from these studies can be strategically utilized to deploy sampling efforts and to disseminate new varieties in informal seed systems. Future surveys that include questions about social ties and the movement of information among farmers would support better models of variety adoption and distribution in this system. Next research steps will include more finely parameterizing transmission patterns, including the impact of variety resistance, and modeling system adaptation to sustained exogenous shocks and stressors. There is the potential to include data about known yield degeneration rates and known environmental conditions to predict regional yield loss in the case of pathogen introduction. Understanding these system components supports better strategies for seed system development.

Acknowledgements

This work received institutional review board approval through the University of Florida IRB-02 #IRB201700024. We confirm farmer participation was voluntary and personal and demographic information was protected. This research was undertaken as part of, and funded by, the CGIAR Research Program on Roots, Tubers and Bananas (RTB) and supported by CGIAR Fund Donors http://www.cgiar.org/about-us/governing-2010-june-2016/cgiar-fund/fund-donors-2/, The Bill and Melinda Gates Foundation (OPP1080975), US NSF Grant EF-0525712 as part of the joint NSF-NIH Ecology of Infectious Disease program, US NSF Grant DEB-0516046, and the University of Florida. The authors confirm they have no conflicts of interest.

References

Abay, F., de Boef, W., and Bjørnstad, Å. 2011. Network analysis of barley seed flows in Tigray, Ethiopia: supporting the design of strategies that contribute to on-farm management of plant genetic resources. Plant Genetic Resources 9:495-505.

Adikini, S., Mukasa, S. B., Mwanga, R. O. M., and Gibson, R. W. 2015. Sweet potato cultivar degeneration rate under high and low sweet potato virus disease pressure zones in Uganda. Canadian Journal of Plant Pathology 37:136-147.

Almekinders, C. J. M., Louwaars, N. P., and Debruijn, G. H. 1994. Local seed systems and their importance for an improved seed supply in developing countries. Euphytica 78:207-216.

Behrman, J. 2011. HarvestPlus Reaching End Users (REU) Orange-Fleshed Sweet Potato (OFSP) Project: Report of Qualitive Findings from Uganda. International Food Policy Research Institute 1-29.

Buddenhagen, C. E., Hernandez Nopsa, J. F., Andersen, K. F., Andrade-Piedra, J., Forbes, G. A., Kromann, P., Thomas-Sharma, S., Useche, P., and Garrett, K. A. 2017. Epidemic Network Analysis for Mitigation of Invasive Pathogens in Seed Systems: Potato in Ecuador. Phytopathology 107:1209-1218.

Chades, I., Martin, T. G., Nicol, S., Burgman, M. A., Possingham, H. P., and Buckley, Y. M. 2011. General rules for managing and surveying networks of pests, diseases, and endangered species. Proc Natl Acad Sci U S A 108:8323-8328.

Csárdi, G., and Nepusz, T. 2006. The igraph software package for complex network research. InterJournal, Complex Systems 1695:1-9.

De Groote, H., Oloo, F., Tongruksawattana, S., and Das, B. 2016. Community-survey based assessment of the geographic distribution and impact of maize lethal necrosis (MLN) disease in Kenya. Crop Protection 82:30-35.

Garrett, K. A. 2012. Information networks for disease: commonalities in human management networks and within-host signalling networks. European Journal of Plant Pathology 133:75-88.

Garrett, K. A. 2017. Impact Network Analysis: a framework for evaluating the effects of information and other technologies through linked socioeconomic and biophysical networks. bioRxiv.

Gibson, R. 2013. How sweet potato varieties are distributed in Uganda: actors, constraints and opportunities. Food Security 5:781-791.

Gibson, R. W., and Kreuze, J. F. 2015. Degeneration in sweetpotato due to viruses, virus-cleaned planting material and reversion: a review. Plant Pathology 64:1-15.

Gibson, R. W., Namanda, S., and Sindi, K. 2011. Sweetpotato seed systems in East Africa. African Crop Science Conference Proceedings 10:449-451.

Graziosi, I., Minato, N., Alvarez, E., Ngo, D. T., Hoat, T. X., Aye, T. M., Pardo, J. M., Wongtiem, P., and Wyckhuys, K. A. 2016. Emerging pests and diseases of South-east Asian cassava: a comprehensive evaluation of geographic priorities, management options and research needs. Pest Management Science 72:1071-1089.

Harwood, T. D., Xu, X., Pautasso, M., Jeger, M. J., and Shaw, M. W. 2009. Epidemiological risk assessment using linked network and grid based modelling: Phytophthora ramorum and Phytophthora kernoviae in the UK. Ecological Modeling 220:3353-3361.

Hernandez Nopsa, J. F., Daglish, G. J., Hagstrum, D. W., Leslie, J. F., Phillips, T. W., Scoglio, C., Thomas-Sharma, S., Walter, G. H., and Garrett, K. A. 2015. Ecological networks in

stored grain: Key postharvest nodes for emerging pests, pathogens, and mycotoxins. BioScience 65:985-1002.

Hilker, F. M., Allen, L. J. S., Bokil, V. A., Briggs, C. J., Feng, Z., Garrett, K. A., Gross, L. J., Hamelin, F. M., Jeger, M. J., Manore, C. A., Power, A. G., Redinbaugh, M. G., Rua, M. A., and Cunniffe, N. J. 2017. Modeling Virus Coinfection to Inform Management of Maize Lethal Necrosis in Kenya. Phytopathology 107:1095-1108.

Johnson, A. C., and Gurr, G. M. 2016. Invertebrate pests and diseases of sweetpotato (Ipomoea batatas): a review and identification of research priorities for smallholder production. Annals of Applied Biology 168:291-320.

Karyeija, R. F., Gibson, R. W., and Valkonen, J. P. T. 1998. The significance of sweet potato feathery mottle virus in subsistence sweet potato production in Africa. Plant Disease 82:4-15.

Mahuku, G., Lockhart, B. E., Wanjala, B., Jones, M. W., Kimunye, J. N., Stewart, L. R., Cassone, B. J., Sevgan, S., Nyasani, J. O., Kusia, E., Kumar, P. L., Niblett, C. L., Kiggundu, A., Asea, G., Pappu, H. R., Wangai, A., Prasanna, B. M., and Redinbaugh, M. G. 2015. Maize Lethal Necrosis (MLN), an emerging threat to maize-based food security in Sub-Saharan Africa. Phytopathology 105:956-965.

McGuire, S., and Sperling, L. 2013. Making seed systems more resilient to stress. Global Environmental Change 23:644-653.

Moslonka-Lefebvre, M., Ann Finley, I. D., Dehnen-Schmutz, K., Harwood, T., Jeger, M. J., Xu, X., Holdenrieder, O., and Pautasso, M. 2011. Networks in plant epidemiology: from genes to landscapes, countries, and continents. Phytopathology 101:392-403.

Nelson, M. F., and Bone, C. E. 2015. Effectiveness of dynamic quarantines against pathogen spread in models of the horticultural trade network. Ecological Complexity 24:14-28.

Obong, Y., Omony, T., Rachkara, P., and Gibson, R. W. 2017. Disseminating modern sweetpotato varieties using participatory variety demonstration trials on informal nodal multipliers' fields at hub locations. Journal of Crop Improvement:1-18.

Pautasso, M. 2015. Network simulations to study seed exchange for agrobiodiversity conservation. Agron. Sustain. Dev. 35:145-150.

Pautasso, M., and Jeger, M. J. 2008. Epidemic threshold and network structure: The interplay of probability of transmission and of persistence in small-size directed networks. Ecological Complexity 5:1-8.

Pautasso, M., Moslonka-Lefebvre, M., and Jeger, M. J. 2010. The number of links to and from the starting node as a predictor of epidemic size in small-size directed networks. Ecological Complexity 7:424-432.

Pautasso, M., Aistara, G., Barnaud, A., Caillon, S., Clouvel, P., Coomes, O. T., Delêtre, M., Demeulenaere, E., De Santis, P., Döring, T., Eloy, L., Emperaire, L., Garine, E., Goldringer, I., Jarvis, D., Joly, H. I., Leclerc, C., Louafi, S., Martin, P., Massol, F., McGuire, S., McKey, D., Padoch, C., Soler, C., Thomas, M., and Tramontini, S. 2013. Seed exchange networks for agrobiodiversity conservation. A review. Agron. Sustain. Dev. 33:151-175.

Rachkara, P., David P Phillips, P., Kalule, S., and Gibson, R. 2017. Innovative and beneficial informal sweetpotato seed private enterprise in northern Uganda. *in press*.

Sanatkar, M. R., Scoglio, C., Natarajan, B., Isard, S. A., and Garrett, K. A. 2015. History, epidemic evolution, and model burn-in for a network of annual invasion: Soybean rust. Phytopathology 105:947-955.

Shaw, M. W., and Pautasso, M. 2014. Networks and plant disease management: concepts and applications. Annu Rev Phytopathol 52:477-493.

Silk, M. J., Croft, D. P., Delahay, R. J., Hodgson, D. J., Boots, M., Weber, N., and McDonald, R. A. 2017. Using social network measures in wildlife disease ecology, epidemiology, and management. Bioscience 67:245-257.

Sperling, L. 2008. When disaster strikes: a guide to assessing seed system security.

Sutrave, S., Scoglio, C., Isard, S. A., Hutchinson, J. M. 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.

Thomas-Sharma, S., Andrade-Piedra, J., Carvajal Yepes, M., Hernandez Nopsa, J., Jeger, M., Jones, R., Kromann, P., Legg, J., Yuen, J., Forbes, G., and Garrett, K. A. 2017. A risk assessment framework for seed degeneration: Informing an integrated seed health strategy for vegetatively-propagated crops. Phytopathology 107:1123-1135.

Thomas-Sharma, S., Abdurahman, A., Alic, S., Andrade-Piedrad, J. L., Bao, S., Charkowski, A. O., Crook, D., Kadian, M., Kromann, P., Struik, P. C., Torrance, L., Garrett, K. A., and Forbes, G. A. 2016. Seed degeneration in potato: the need for an integrated seed health strategy to mitigate the problem in developing countries. Plant Pathology 65:3-16.

Wang, H., Cui, X., Wang, X., Liu, S., and Zhou, X. 2016. First Report of *Sri Lankan cassava mosaic virus Infecting Cassava in Cambodia*. Plant Disease 100:1029.

Wangai, A. W., Redinbaugh, M. G., Kinyua, M., Miano, D. W., Leley, P. K., Kasina, M., Mahuku, G., Scheets, K., and Jeffers, D. 2012. First Report of Maize chlorotic mottle virus and Maize Lethal Necrosis in Kenya. Plant Disease 96:1582.

Table 1

Common Network Node Centrality Measures

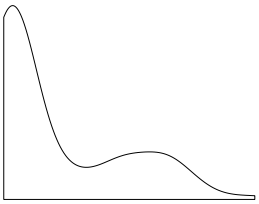
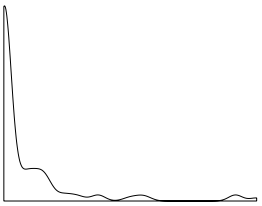
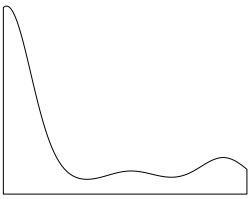
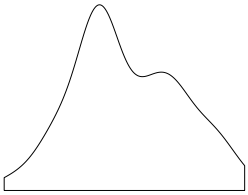
Measure	Definition	Relevance for Epidemics in Seed Systems	Density plot for distribution of each measure for villages in this network
Degree Centrality	The number of links a node has to other nodes in the network (both incoming and outgoing)	Node degree of epidemic starting point (number of links) has been shown to influence epidemic outcomes. Those with high degree may also be "super spreaders" once infected	
Betweenness centrality	The number of shortest paths through the network of which a node is a part	A measure of how much a node serves as a "bridge" to new nodes. Removal of nodes with high betweenness may contain an epidemic within a region	
Eigenvector centrality	A weighted sum including both direct links to a node (degree) and the link degree of neighbors	If a node does not itself have a high node degree, but is connected to nodes with high degree, it may be at increased risk of infection and spreading infection	
Closeness Centrality	The inverse of the average length of the shortest path to/from all the other nodes in the network	Efficiency of the node to spread disease to any other node in the network	

Table 2. Parameters describing dispersal probabilities in a seed system network

Parameter	Definition	Values	Comments
β	Exponent of the power law, where $\beta > 0$. Influences the shape of decay	0.5-2 (0.5, 1.0, 1.5*, 2.0, 2.5)	Larger values of β result in steeper decay, and fewer village to village links (dependent on *t*).
t	Seed transaction threshold. Applied to power law transformed distance data.	0.001-0.15 (0.001, 0.005, 0.01*, 0.015, 0.02)	Influences how close neighboring villages need to be to have a link formed in the village-to-village network.
P_t	Probability of transmission of infection at each timestep, over an established link	0.05-0.15 (0.05, 0.01*, 0.015)	Influences the probability that infection will be transmitted along a network link in each time step, from seller to village, or village to village

**denotes the value that was used as the default in simulation experiments*

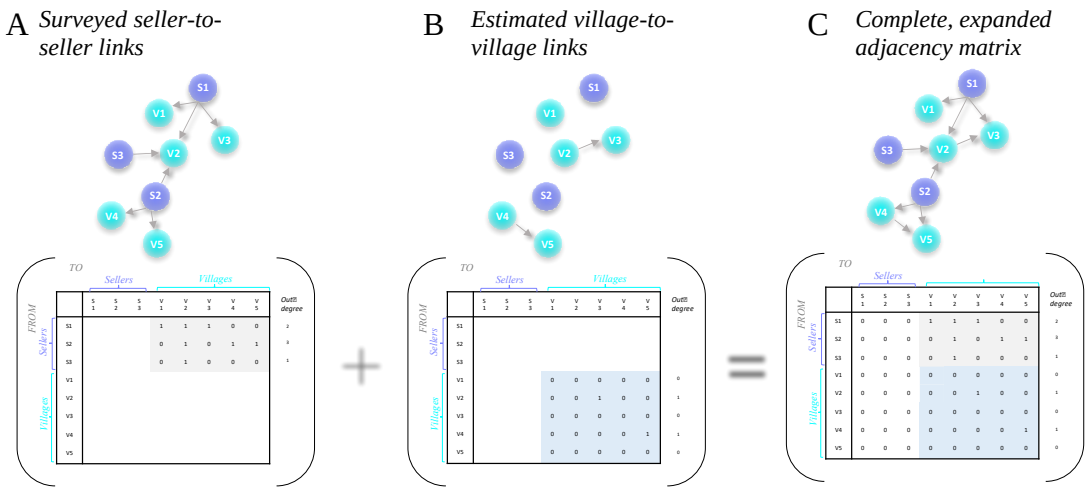
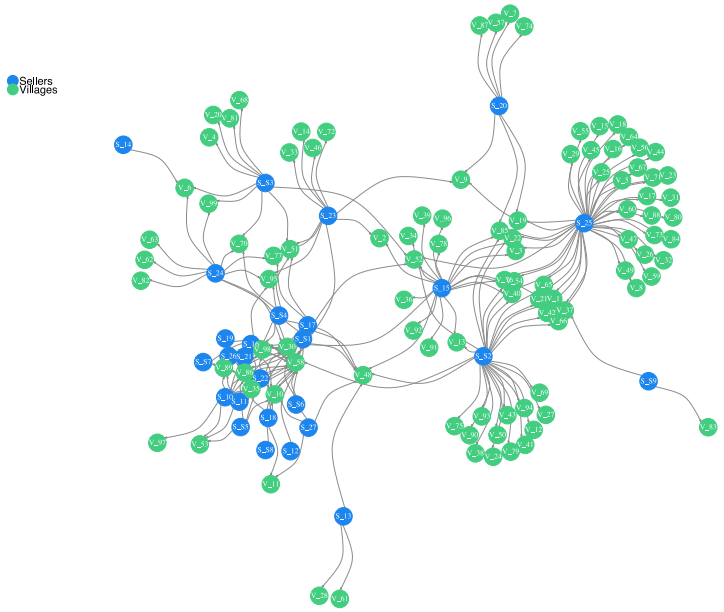


Figure 1

A schematic illustrating the method used to combine adjacency matrices of A) links observed in the 2014 vine seller survey, and B) links between villages, estimated using the power law function to describe how the probability of movement decreases with increasing distance. This example used a hypothetical network of three sellers and five villages, each represented as a node in the network. A link between nodes is represented as a 1 in the matrix, and no link is indicated by a 0. Matrices were combined in C) a "complete network" with transactions from sellers to villages, and also between villages. Note that in this study all seller-to-seller links and village-to-seller links were set to zero, with no transactions taking place in this direction. The "complete network" was used in the simulation experiments presented in this study.

A



B

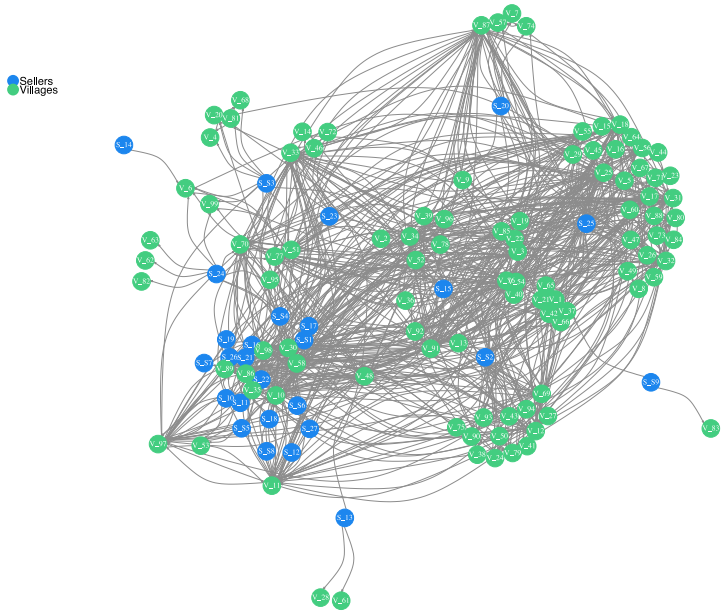


Figure 2

The network structure of sweetpotato vine transactions occurring in the 2014 growing season in Northern Uganda with both sellers (blue nodes) and villages (green nodes). Links represent the occurrence of at least one transaction in the 2014 growing season. Note that the graph layout is

based on the Fruchterman-Reingold force directed algorithm, which locates nodes with links closer together and those without links further apart, and not the geographic location of villages. Plot A) represents empirically sampled seller-to-village links and, B) represents village-to-village links inferred as a function of inter-village distance.

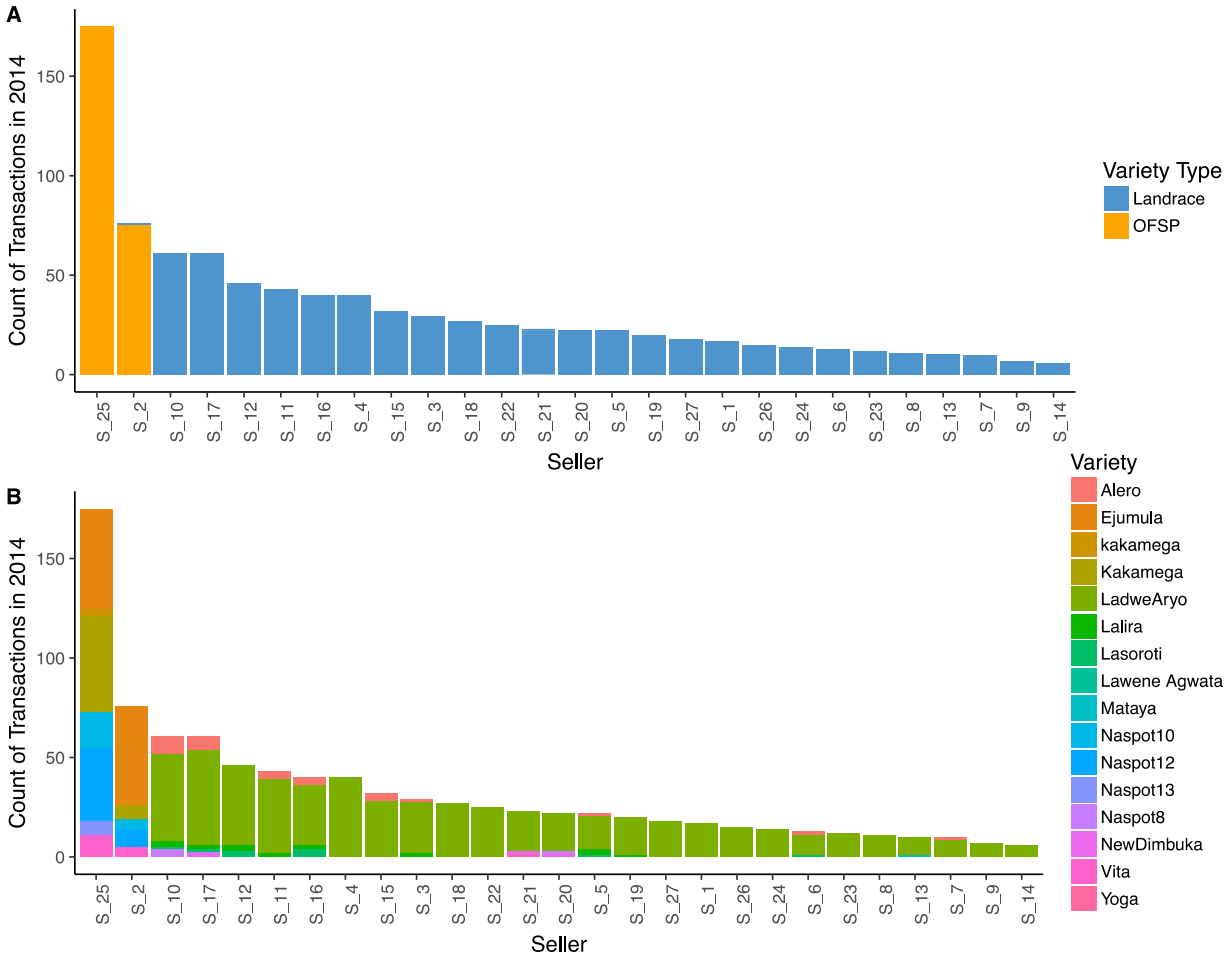


Figure 3

Bar chart indicating the reported transaction count per seller for A) varieties grouped as landrace or OFSP varieties, and B) all individual varieties that were sold in 2014. Sellers were visited once per week during the growing season to obtain their reported transaction history for the previous week. Two sellers in the seed network sell OFSPs, with the rest selling only landraces. The landrace variety Ladwe Aryo is sold across 93% of sellers surveyed, with the highest number of total transactions in 2014.

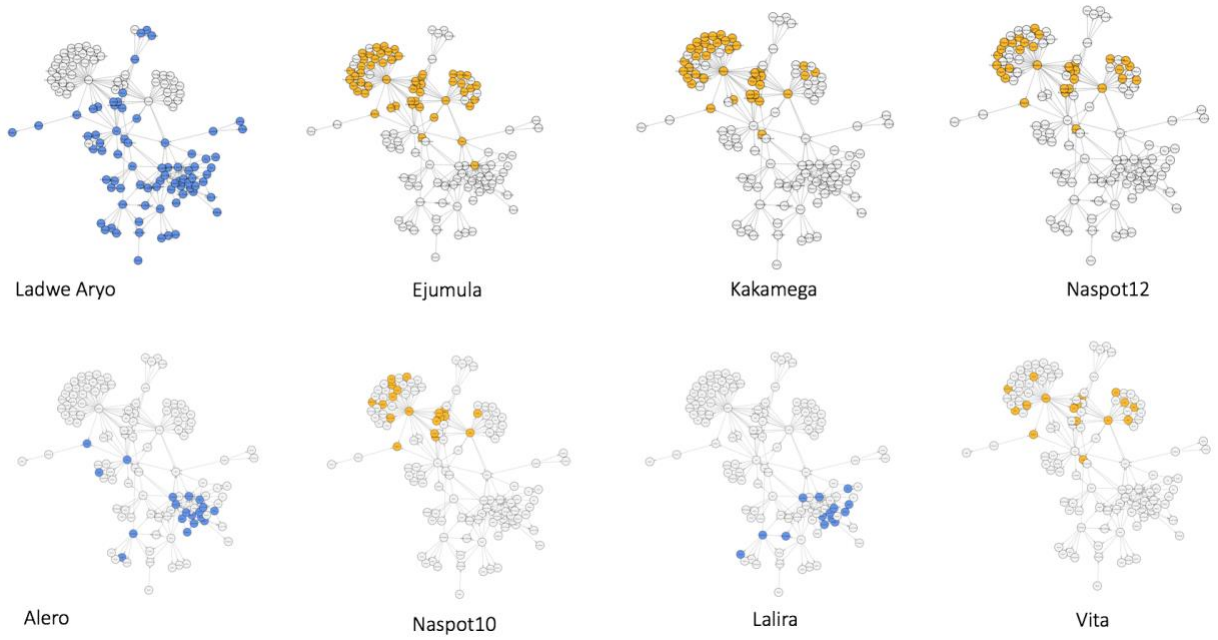


Figure 4

Networks of dissemination of the top eight cultivars sold during the 2014 growing season, in terms of number of transactions reported. All sellers and villages surveyed are represented in each network, while colored nodes represent sellers and villages involved in the sale or purchase of the specified cultivar. White villages did not access a given variety in 2014 through this surveyed seed network. Node color indicates white-fleshed landraces (blue) and OFSP (orange) cultivars.

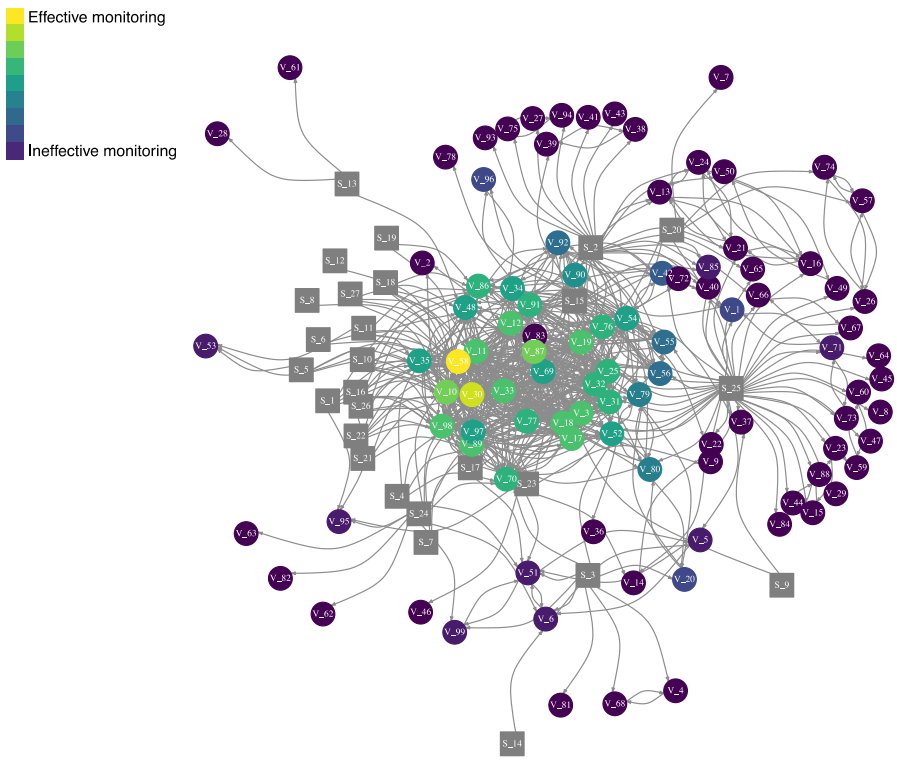


Figure 5

Network resulting from a simulation (500 realizations) where each node had equal likelihood of being the introduction point of an invasive pathogen. Node color indicates the total number of nodes that would be reached in the network before detection at that node. If few nodes were reached (yellow) the node was considered an effective monitoring location, while if many nodes were reached (purple) the node was considered ineffective for monitoring. Nodes labeled with an "S" and depicted with gray boxes are sellers, and circles ("V") are villages. Only villages were considered in this analysis for their efficacy as sampling locations.

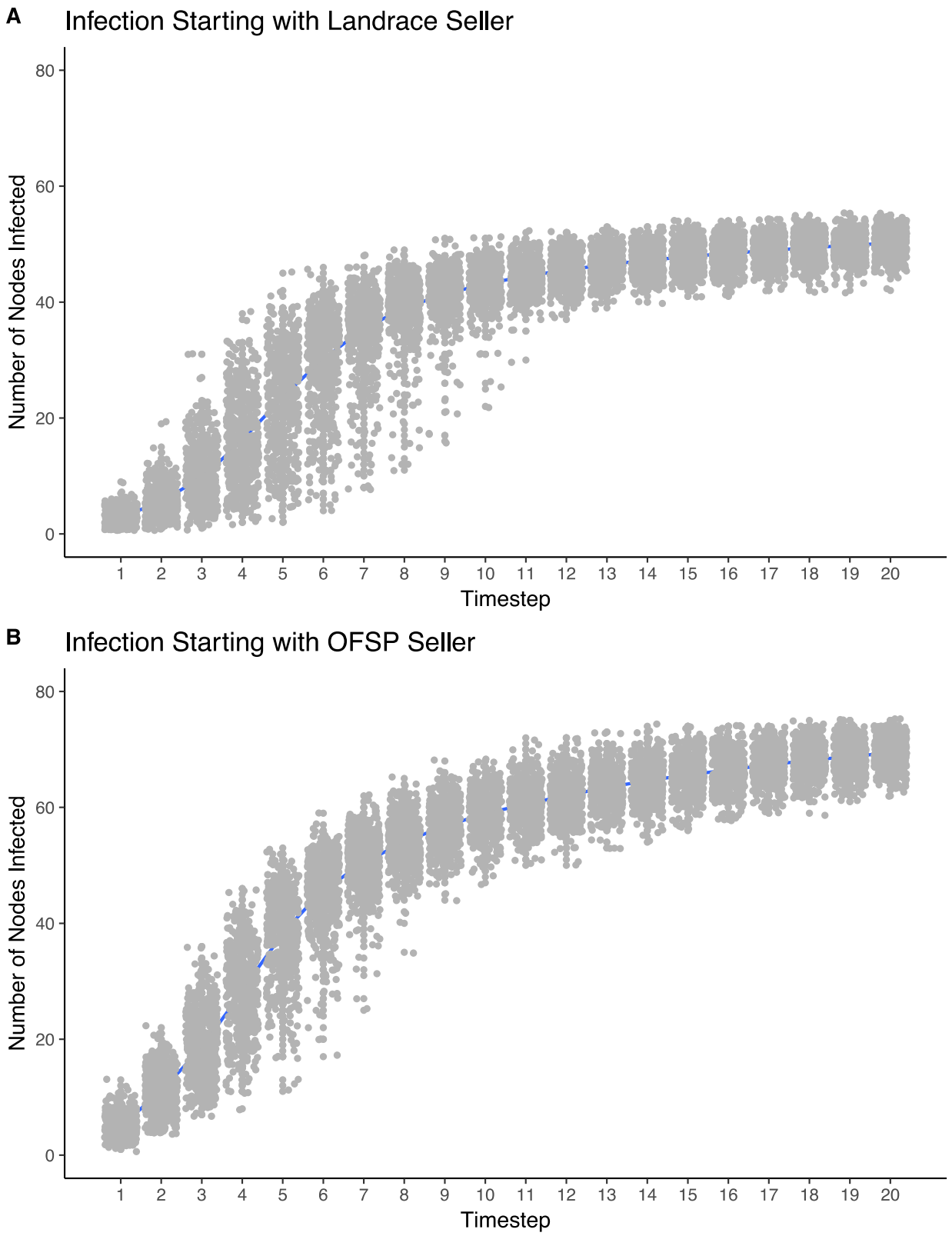

A
Infection Starting with Landrace Seller
Number of Nodes Infected
0
20
40
60
80
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Timestep
B
Infection Starting with OFSP Seller
Number of Nodes Infected
0
20
40
60
80
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Timestep

Figure 6

Epidemic progress through the seed network over 20 hypothetical timesteps in 500 realizations (each realization indicated as a gray point) with the blue line indicating smoothed conditional mean (local regression) of disease progress across realizations. Epidemics were modeled with two starting sellers, A) the seller with the highest out-degree of landrace sales (S_15), and b) the seller with the highest out-degree of OFSP cultivar sales (S_25). Number of nodes infected represents the sum of the infection statuses of all nodes at each time step.

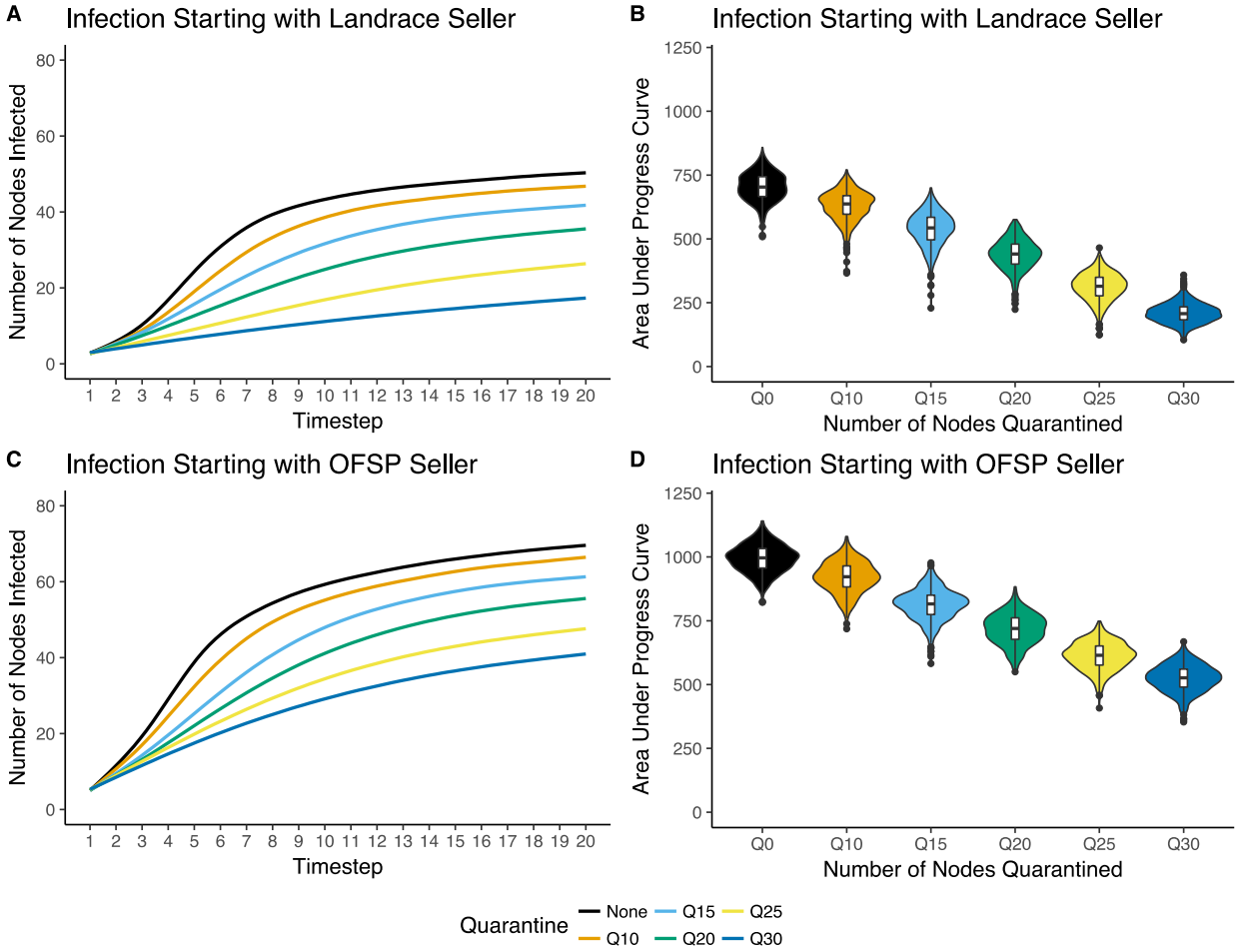


Figure 7

Disease progress over 20 time steps under different quarantine regimes (None = no quarantine, and Q10, Q15, Q20, Q25, and Q30 indicate 10, 15, 20, 25, and 30 nodes quarantined, respectively). A) and C) Disease progress, for each quarantine status, with infection starting with the seller with the highest out-degree of landrace sales (S_15) in A, or the seller with highest out-degree of OFSP sales (S_25) in C. Lines represent smoothed conditional means. B) and D) Area Under the Disease Progress Curve boxplot and distribution (violin plot) for each of the 500 realizations, in A and C, respectively.

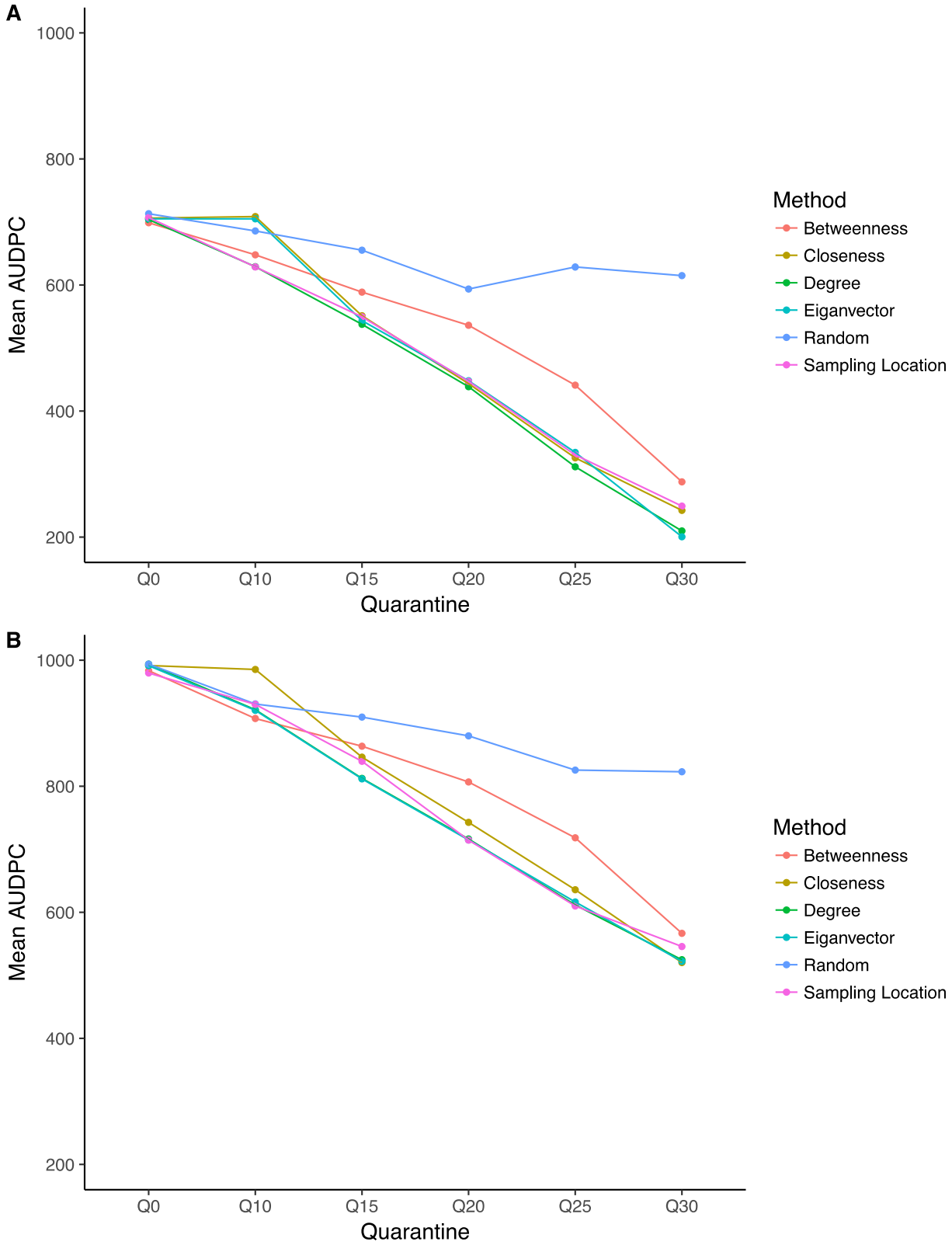

A
1000
800
600
400
200
Mean AUDPC
Q0
Q10
Q15
Q20
Q25
Q30
Quarantine
Method
Betweenness
Closeness
Degree
Eiganvector
Random
Sampling Location
B
1000
800
600
400
200
Mean AUDPC
Q0
Q10
Q15
Q20
Q25
Q30
Quarantine
Method
Betweenness
Closeness
Degree
Eiganvector
Random
Sampling Location

Figure 8

Mean AUDPC values for six quarantine scenarios (Q0 = no quarantine, and Q10, Q15, Q20, Q25, and Q30 indicate 10, 15, 20, 25, and 30 nodes quarantined, respectively) for infection starting with A) the seller with the highest out-degree of landrace sales (S_15), and B) the seller with the highest out-degree of OFSP sales (S_25). Four different centrality measures were compared for their utility in selecting quarantine locations (betweenness, closeness, degree, and eigenvector). These were also compared to a scenario where nodes were selected at random for quarantine, and nodes that were selected as effective monitoring locations ("Sampling Location") in the previously described simulation experiment (Figure 5).

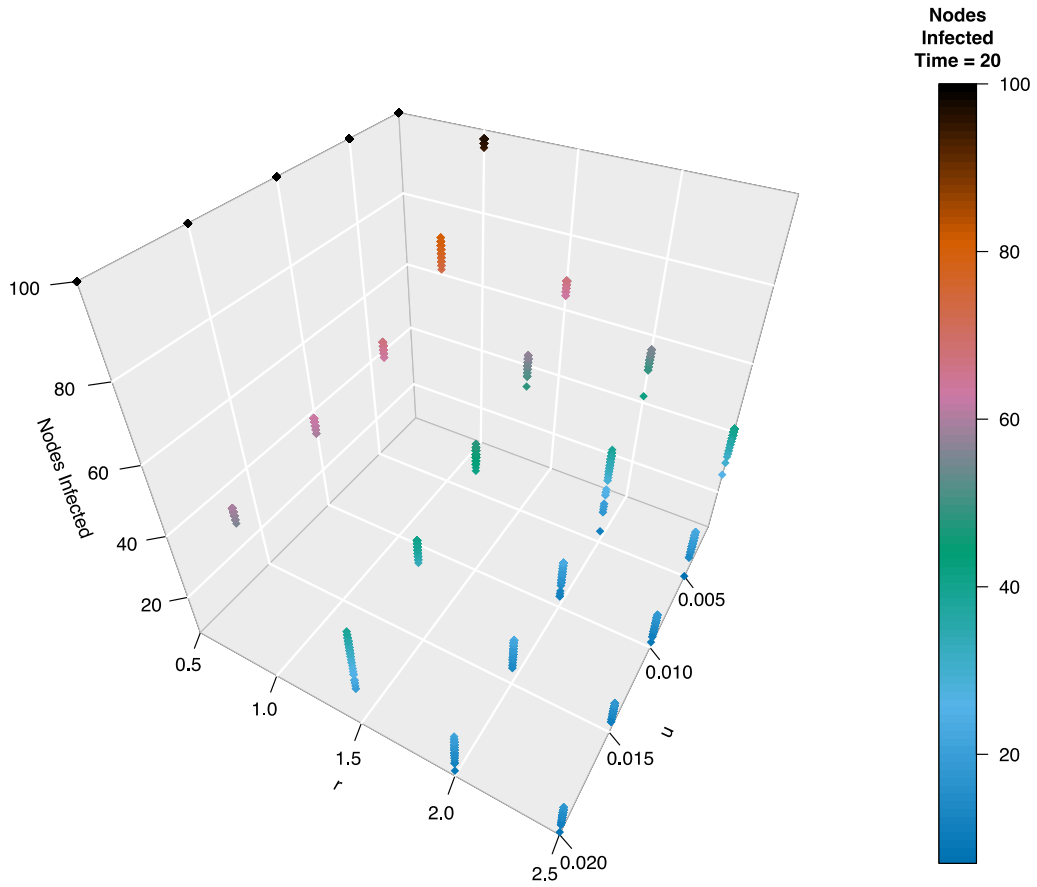


Figure 9.

Sensitivity analysis where the number of nodes infected at timestep 20, for each of 100 simulations, is compared across a range of values for parameters β (0.5-2.5) and *t* (.001-.02). The parameters chosen here were used to create seed transaction links between villages, and therefore influenced disease spread.