Optimizing testing strategies for early detection of disease outbreaks in animal trade networks via MCMC

Sara Ansari, Jobst Heitzig, Mohammad R. Moosavi

Originally published as

Ansari, S., Heitzig, J., Moosavi, M.R. (2023): Optimizing testing strategies for early detection of disease outbreaks in animal trade networks via MCMC. - Chaos, 33, 4, 043144.
https://doi.org/10.1063/5.0125434

Chaos
An Interdisciplinary Journal
of Nonlinear Science

AIP Publishing

RESEARCH ARTICLE | APRIL 28 2023

Optimizing testing strategies for early detection of disease outbreaks in animal trade networks via MCMC

Sara Ansari ✉ ; Jobst Heitzig ; Mohammad R. Moosavi

Chaos 33, 043144 (2023)
https://doi.org/10.1063/5.0125434

Chaos ARTICLE scitation.org/journal/cha

Optimizing testing strategies for early detection of disease outbreaks in animal trade networks via MCMC

Cite as: Chaos **33**, 043144 (2023); doi: 10.1063/5.0125434
Submitted: 12 September 2022 · Accepted: 30 March 2023 ·
Published Online: 28 April 2023

Sara Ansari,[1,2,a)] Jobst Heitzig,[2] and Mohammad R. Moosavi[1]

AFFILIATIONS

[1] Department of Computer Science and Engineering, School of Electrical and Computer Engineering, Shiraz University, 7193616548 Shiraz, Iran
[2] FutureLab on Game Theory and Networks of Interacting Agents, Potsdam Institute for Climate Impact Research, 14473 Potsdam, Germany

[a)] **Author to whom correspondence should be addressed:** Ansari@pik-potsdam.de

ABSTRACT

The animal trades between farms and other livestock holdings form a complex livestock trade network. The movement of animals between trade actors plays an important role in the spread of infectious diseases among premises. Particularly, the outbreak of silent diseases that have no clinically obvious symptoms in the animal trade system should be diagnosed by taking special tests. In practice, the authorities regularly conduct examinations on a random number of farms to make sure that there was no outbreak in the system. However, these actions, which aim to discover and block a disease cascade, are yet far from the effective and optimum solution and often fail to prevent epidemics. A testing strategy is defined as making decisions about distributing the fixed testing budget N between farms/nodes in the network. In this paper, first, we apply different heuristics for selecting sentinel farms on real and synthetic pig-trade networks and evaluate them by simulating disease spreading via the SI epidemic model. Later, we propose a Markov chain Monte Carlo (MCMC) based testing strategy with the aim of early detection of outbreaks. The experimental results show that the proposed method can reasonably well decrease the size of the outbreak on both the realistic synthetic and real trade data. A targeted selection of an $N/52$ fraction of nodes in the real pig-trade network based on the MCMC or simulated annealing can improve the performance of a baseline strategy by 89%. The best heuristic-based testing strategy results in a 75% reduction in the average size of the outbreak compared to that of the baseline testing strategy.

This research investigates different testing strategies, which are worth to be implemented by authorities to detect the outbreaks as early as possible. The final goal is finding the optimal testing strategy in terms of the outbreak size, considering a limited overall testing budget N. Compared to testing a large number of randomly selected nodes once a year, targeted and much more frequent testing of a sensible selection of much fewer nodes based on the local network measures or the farm characteristic is substantially more effective in reducing the average number of infected animals at the time of outbreak detection. After discussing several heuristics for choosing these nodes, we study a Markov chain Monte Carlo (MCMC) method to determine those sets of nodes that are influential to be targeted as regular testing points to reduce the expected number of infected cases. We find that it outperforms heuristics significantly and can be improved further by incorporating insights from heuristics as well as simulated annealing. On sampled real-world networks and realistic synthetic networks, both these further improvements perform comparably well.

I. INTRODUCTION

Large-scale outbreaks of animal diseases have been a major concern in many countries in the last few decades, for example, foot and mouth disease in the UK in 2001[1] and the swine fever epidemics in the Netherlands and Germany in the 1990s and 2003.[2] The African swine fever (ASF) was first endemic in most of Africa, then jumped to Georgia and Russia, became a major outbreak in Eastern

Europe[3] in 2007 and later on in China in 2018–2021.[4] China's outbreak of ASF in 2018–2021 was more devastating than expected.[5,6] This virus reappeared after 40 years again in the U.S. in 2021 and in 15 countries in Asia. Therefore, studies of such spreading dynamics are crucial for both public health and economic reasons.

The translocation of both domestic and wildlife animals plays a significant role in the emergence of new diseases and the spread of fatal viruses, which threaten not only animals but also human health.[7–15] Trading of susceptible host animals is an important common theme in the introduction of exotic animal diseases or zoonotic human pathogens, such as monkey pox, SARS, Ebola, MERS, and probably COVID-19.[12,16–24] All these can lead to irreparable impacts on society, the global economy, and public health.[2,5,25–34] Therefore, limiting the damages caused by infectious diseases is an important challenge for governments. Early detection of an outbreak is very important and gives sufficient time to public health officials for disease management and implementation of countermeasures on a smaller scale of the population, which is not only easier but also more cost-efficient[35] and is the goal of this study.

Our case study is the pig-trade system in Germany. Some rare disease outbreaks might happen in this system. These diseases have a silent spreading phase, which means that there is considerable time between catching the disease and showing sufficiently clear symptoms to be diagnosed (this is called the "inconspicuous time"); hence, it should be identified by testing.

Therefore, active surveillance is crucial in this system for discovering the propagation of this kind of disease in advance. For this reason, authorities regularly conduct examinations on a random number of farms to make sure that the system is safe, but this random testing strategy as we show in the next parts is far from the optimum solution and, therefore, fails to prevent epidemics. Therefore, the European Union (EU) recently urged states to use more intelligent strategies and what they call a "risk-based" test design. The goal of this study is to find a near-optimal testing scheme for the early detection of disease outbreaks in animal trade networks.

Testing strategy is defined as making decisions for assigning the testing budget to farms. This decision can be described as a vector, including n entries; each one shows the amount of budget assigned to one farm. The overall testing budget N describes the maximum number of tests that can be carried out during the observation period (usually one year). In this paper, we investigate different testing strategies for finding the optimal set of farms as the sentinel set, which is defined as

$$S = \{v_0, v_1 \cdots v_k\}_{v_i \in V}. \tag{1}$$

v_i shows the ith sentinel node, a farm that is selected by a testing strategy as a regular testing point. k is the number of sentinels. Both v_i and k are outputs of a testing strategy. Therefore, the testing strategy consists of decisions about how many farms, which farms, and how often each farm (number of times per year) should be tested. In current practice, all farms are tested once a year and the testing time is completely random; we call this the "baseline" strategy. Therefore, the overall testing budget N is equal to the number of farms n. Instead of testing each farm once a year, we aim to test a small subset of so-called sentinel farms S more frequently, which is more rational based on the budget at hand.

The size of an outbreak at the time it is detected is considered the objective function and the selection of sentinels as the control variable of our optimization problem. For estimation of the outbreak size, the susceptible–infected (SI) epidemic model is employed in the simulation of the spreading dynamics. Ignoring later epidemic state transitions, such as the removal of animals due to death or recovery in our model, can be justified since we are interested in diseases where the initial spreading is much faster than these later transitions.

The rest of this paper is organized as follows: Sec. II describes the datasets that are used in this paper. Various heuristics for finding sentinels are discussed in Sec. III. In Sec. V, we develop a sentinel selection strategy based on Markov chain Monte Carlo (MCMC) and simulated annealing (SA) and show that it outperforms the studied heuristics. Finally, the paper is concluded in Sec. V.

II. MATERIALS AND METHODS

A. Real dataset

HI-Tier is a comprehensive database that provides the daily description of pig movements in the pig-trade system of Germany since 2006.[36] In this study, we use an extract of this database, which contains all trades among pig farms/holdings in Germany between January 1, 2011 and December 31, 2014. This dataset includes a set of transmissions that comprise information about the seller premise's identifier as the source of movement, the buyer premise's identifier as the target of movement, the number of animals (i.e., the *batch size*), and the date of movement. Our dataset contains more than 6.2×10^6 of such transmissions between 57 590 premises.

1. Network representation

The pig-trade system can be represented as a transmission network $\mathcal{G} = (V, \mathcal{E})$, consisting of a set V of farms as the nodes of the network and a set of movements between farms as edges $\mathcal{E}$, where every edge (i, j, t, w) is a directed weighted link denoting a transmission of w animals from farm i to j on day t.

Therefore, the pig-trade network is a large weighted directed *transmission network.* In addition to the edge weights w, also, the nodes can be weighted by their "capacity," representing the maximum number of animals that can reside there at any time, and this information can be useful in selecting sentinels. The farm capacity is a very important factor in the outbreak detection and also in the system modeling. Since these data were not directly available in the dataset, we estimate the shape of the distribution from the cumulative trading balance (incoming minus outgoing pig) in the available data as described in Ref. 35. Table I shows some properties of the German pig-trade network.

B. Realistic synthetic dataset

Access to animal trade data is essential for research in the field of animal epidemiology, but in most cases, they are confidential or unavailable. In an earlier study, we proposed a temporal network model for animal trade systems that generates synthetic data sets with realistic features.[35] In addition to the only real data set that we had access to, 100 generated realistic synthetic networks with $n = 1040$ and $e = 13\,500$ from that model are used in our experiments.

TABLE I. Standard network properties of the temporal German pig-trade network. GSCC is the giant strongly connected component, GIC is the giant-in component, and GOC shows the giant-out component.

Property	Value
Number of nodes	57 590
Number of edges	6 241 634
Edge density	108.381
Modularity	0.70
Average path length	5.65
Diameter	19
Size of GSCC	46%
Size of GIC	34%
Size of GOC	10%

The relevant topological features and the capacity distribution of both the real and realistic synthetic networks that distinguish them from standard network ensembles, such as random, random geometric, small-world, or scale-free networks, are described in Refs. 14 and 35.

III. HEURISTICS METHODS

As mentioned in Sec. I, we aim to select a small subset of important farms from all animal holdings as sentinel set S (i.e., testing points).

Since our data are modeled as a network, one idea is to look at network measures to identify important nodes in the network. There are many studies available on finding important nodes in the networks, mainly, degree, closeness centrality,[37] k-shell centrality,[38] and betweenness.[39] In a directed network, node degree measures the number of incoming and outgoing links attached to each node. Closeness centrality ranks the nodes based on their closeness to all other nodes in the network and betweenness measures the number of shortest paths that pass through each node. At first, we examined all of these methods for finding important nodes on a small realistic synthetic network of 120 nodes.

Our preliminary experiments show that the k-shell is not a good choice for our purpose in comparison with other measures. K-shell centrality assumes that nodes in higher shells are more important than nodes in lower shells. While this may be true in some cases, it may not always hold in disease spreading scenarios. For example, a node that is located at the periphery of the network may be less connected overall but may be a key link between different parts of the network and, thus, have a significant impact on disease spreading.[40,41] Overall, closeness provides the best results. Since computing the closeness of all nodes in a network is highly expensive, especially for a large network, we have used degree as one of the heuristics in later experiments. Using the degree is a simple method with comparable performance to closeness centrality and can be easily applied on very large networks.

In this section, we consider three heuristics for selecting a certain number $k > 0$ of sentinel nodes:

- the k nodes with the highest degree,
- the k nodes with the highest number of incoming transmissions (HIT), and
- the k nodes with the highest capacity.

At first, we apply each of these heuristics to the input network to find S by selecting k nodes as v_i in S. Next, the SI simulator is used to evaluate the performance of each heuristic in terms of the outbreak size, estimated by the expected total capacity of all nodes that are infected at the time the outbreak is detected either by a test or by the end of the inconspicuous time. In the SI simulation, initially, all nodes are in the susceptible state but the one randomly selected as the initially infected node. This infected node can transmit the disease via the transmission links with an infection probability of $\beta > 0$, which is set to 0.5 in our experiments.

How does our SI simulator work? The spreading process here is dynamic: We suppose that the outbreak initiates from a single node in an arbitrary time step. Assume that the node i is infected at time step t_i. Each neighbor j of this infected node potentially has the chance of being infected if their trading activity occurs after the infection time of i [i.e., there exists a link (i, j, t', w) such that $t' >= t_i$].

Our network is temporal (each trading link has a time property). Therefore, the simulator compares the time of link/trade with the infection time of source of that link to see whether this link could carry the pathogen to the destination or not. Therefore, both causality and the infection probability are considered in the spreading process of a disease in the SI simulator. The spreading process continues until reaching one of the pre-selected sentinel nodes in S. At this step, one test is performed if the test result is positive and the outbreak is detected at this node and the simulation terminates. Then, the total number of animals in all infected nodes is calculated. The objective function is then estimated by taking the average of this quantity over 10 000 independent SI simulations with the same selection of sentinels. If the test result is negative, then the simulation continues until the end of inconspicuous time, which will be discussed in Sec. III A.

For simplicity, we suppose that all animals in an infected farm will get the infection. The tests for infection can produce false negative results, meaning that even a test at an infected node might not detect the disease. Therefore, we consider false negative rate = 0.1 in the simulations. Again, for simplicity, we assume that this probability is equal for all premises and constant over time. Of course, the tests might also produce false positive results, mistakenly declaring susceptible nodes as infected. However, as we want to focus on detection, errors of this type will be disregarded. To validate the performance of a sentinel selection strategy, all results are reported on 100 networks. The performance of a sentinel selection strategy is then assessed by the average value of this estimate over all considered networks using for each network the concrete sentinel set produced by the considered strategy. The external parameters of the SI simulation are the infection probability β, the number of sentinels k, and the inconspicuous time of the disease, which will be discussed in Secs. III A–III C.

A. The impact of disease inconspicuous time

Inconspicuous time refers to the time period between the beginning of the infection and the appearance of clear symptoms that make the infection obvious and, hence, immediate detection without testing. Since various silent diseases have different inconspicuous times, we studied their impact on the outbreak size by considering a range of values in our simulations to provide a comprehensive evaluation. Figure 1 shows that the larger the inconspicuous time, the more the heuristics outperform the baseline (testing each farm once a year).

B. The impact of the number of sentinels

A key issue affecting the test strategy is the number of sentinels, *k*. Considering a fixed and given budget *N*, which leads to a fixed number of tests, our experiments in the latter parts indicate that it is better to have a smaller number of sentinels while testing them more often (see Fig. 4). The figure shows that picking a small set of 40 nodes out of 1040 as sentinel and testing each of them biweekly (resulting in a total number of tests $N = 1040$ per year equal to the baseline) provides the best results. Precisely, regardless of the testing strategy that is used, choosing a smaller or larger number of nodes increases the number of infected cases.

C. The effect of infection probability

Another important parameter in the SI simulation is the infection rate β, which means the probability of disease transmission from one node to its neighbors in the network. Our simulations validate the obvious expectation that a larger infection rate increases the number of infected cases. More interestingly, the higher the infection rate, the more our heuristic selection strategies outperform the baseline.

Therefore, for highly contagious disease testing, a number of randomly selected farms can lead to adverse results, such as unexpected growth in the number of new cases should be controlled in proper time or otherwise, the outbreak will cause an epidemic.

IV. APPLYING AN MCMC METROPOLIS–HASTINGS ALGORITHM

In a different epidemiological optimization problem (that of removing links from a contact network between people in order to increase the epidemic threshold), we have recently successfully applied a Markov chain Monte Carlo (MCMC) Metropolis–Hastings algorithm[42] as a form of an (approximate) optimization algorithm to select positions in a network relevant for counteracting epidemics (in that case, selecting edges for removal from a contact network to prevent further spreading, rather than selecting nodes for surveillance as in our current setting). This motivated us to examine the MCMC-based optimization method also in this context for finding a good set of sentinels. The input of the MCMC algorithm in our case consists of a network, the number of sentinels *k*, a strategy for generating a candidate proposal in each MCMC step, the parameters of the target function that is used for evaluating candidate proposals, a probability function for turning this evaluation into an acceptance probability, and the number of MCMC steps/iterations. The following probability distribution function was employed in our MCMC simulation:

$$p_G(c) = e^{-\nu I_G(c)}. \tag{2}$$

Here, G is the input network, ν is the inverse of thermodynamic temperature as defined in the Maxwell–Boltzmann distribution,[43] c is the candidate/proposal set for S in each MCMC step, and $I(c)$ is

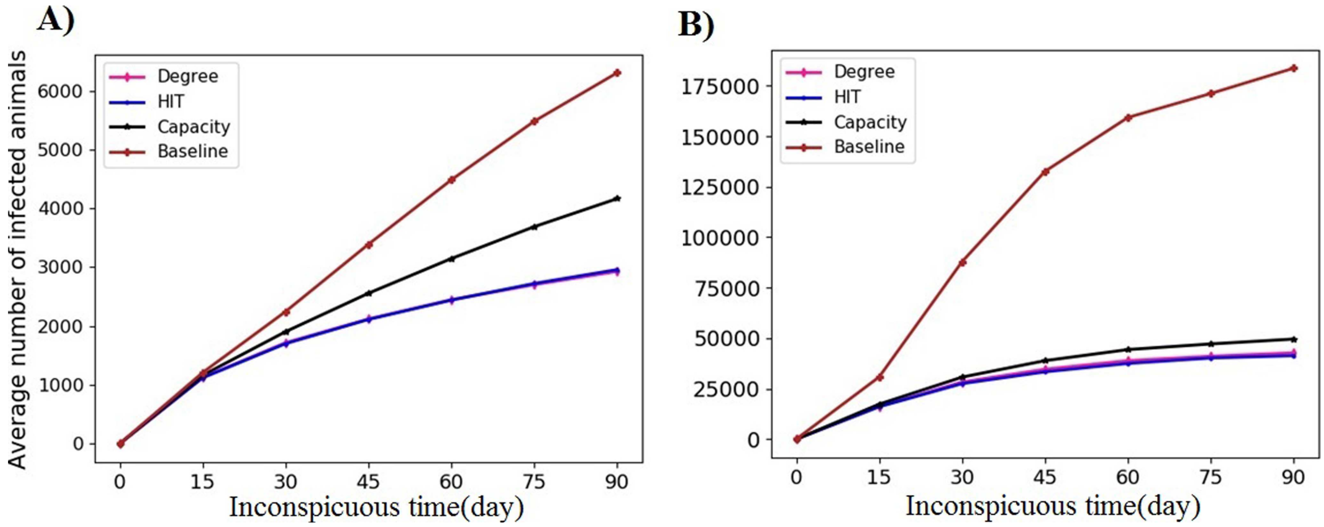

FIG. 1. The impact of disease inconspicuous time on the size of the outbreak, which is the average number of infected animals at the time the outbreak is detected by various heuristic testing strategies. It shows that in addition to increasing the outbreak size, larger inconspicuous times lead to much larger differences in the performance of heuristics and the baseline strategy in both realistic synthetic and real networks. (a) Result of SI simulation on an ensemble of 100 realistic synthetic networks for each network $n = 1040$ and $e = 13\,500$. (b) The SI result on an ensemble of 100 real networks, $n = 57\,590$, $e = 810\,000$.

the value of the objective function, which is the portion of population getting infected at the time of infection detected by one of the sentinel nodes in c, averaged over 10 000 SI simulation runs.

A. MCMC algorithm

Initialization. We set $\nu = 100$, $mcmc - step = 1\,000\,000$, number of nodes $n = 120$ and $n = 1040$ for realistic synthetic and $n = 57\,590$ for real networks. The number of sentinels k varies depending on the budget or the number of tests; for example, it could be once a year for every farm, and therefore, the total number of tests is the same as the number of nodes/farms. We consider a fixed budget $N = n$ and a different number of sentinels, such as $k = N/52, N/26, N/12$, and N.

1. The algorithm starts from a set of k randomly chosen sentinel nodes as initial state c_0 in the search space.
2. In each step t of the MCMC algorithm:
    (a) Generate a new candidate c': by swapping one node between the current sentinel set c_t and the remaining nodes (non-sentinels) in the network.
    (b) Use the temporal-SI simulator to evaluate this candidate (by averaging over 10^3 independent SI runs) and compute the average number of infected animals over all runs.
    (c) Calculate the acceptance rate α with the probability distribution described above:
$$\alpha = \min(1, p(c')/p(c_t)). \tag{3}$$
    (d) Accept or reject:
        - Generate a uniform random number u.
        - If $u \leq \alpha$, then accept the candidate by setting $c_{t+1} = c'$ and this is the new state in the Markov chain.
        - If $u > \alpha$, then reject the candidate by setting $c_{t+1} = c_t$; therefore, no movement in the search space is done.
3. Repeat step 2 MCMC-step times.

The number of iterations or essential MCMC steps for approximate optimization is an external parameter of the method that must be adjusted for the particular problem at hand. We started with 1000 steps and increased it gradually to 10 000 and then to 1 000 000 to see when it reaches a low steady state. The black curve in Fig. 2 shows how standard MCMC decreases the average size of an outbreak at the time of detection on an ensemble of 10 realistic synthetic networks, each of size 120 nodes. We can see that the average size of the outbreak by selecting the best sentinel set found in the standard MCMC is 2.03.

We applied heuristics and baseline on the same network to have a better understanding of the ability of the MCMC algorithm to find good sentinel sets. The value of the objective function for the best heuristic is 2.70 and for the baseline is 3.20. Therefore, the normal MCMC can improve the best heuristic by 25%, but still, the number of infected farms is above 2. Therefore, in Sec. IV B, we try to decrease this value by improving MCMC.

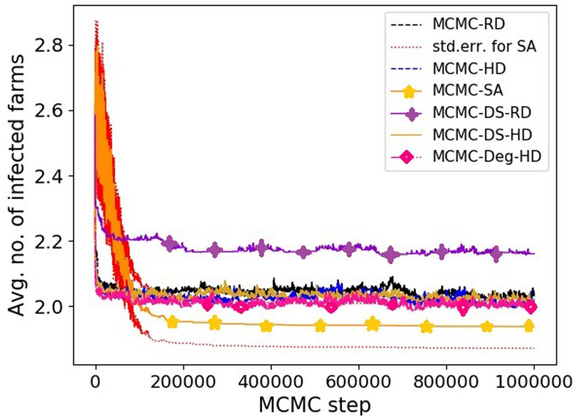

FIG. 2. Different *ad hoc* methods for improving MCMC performance on an ensemble of 10 realistic synthetic networks, each with 120 nodes. The MCMC-based optimization starts with either k random nodes (RD) or the k highest-degree nodes (HD) as sentinels. To generate the next candidate, one sentinel is replaced by a random node, with a uniform probability, or probability proportional to its distance to the other sentinels (DS) or to its degree (Deg). SA, simulated annealing. The objective function is here the expected number of infected farms (rather than animals). The HD version outperforms the RD one. Interestingly, simple simulated annealing without the other features seems to outperform the other strategies here. Still, the standard error of estimating the objective function by the average number of infected farms over 1000 SI simulations in the SA strategy (dotted red curve) shows also that MCMC-Deg-HD and some other strategies give comparable results.

B. Improving the MCMC performance

In this section, we introduce several *ad hoc* methods for enhancing the MCMC performance. In simple MCMC, the temperature parameter is fixed during the algorithm, resulting only in approximate optimization. A related problem is that temperature-induced exploration, which is helpful in initial steps to avoid trapping into local optima later, prevents the fine-tuning needed for precisely locating the optimum.

The standard solution for this problem is using simulated annealing (SA), which decreases the temperature gradually from an initial positive value to zero, similar to reducing the exploration rate gradually in other machine learning approaches. The yellow curve in Fig. 2 shows that SA (orange curve) improves the performance of MCMC and decreases the average size of outbreaks in the realistic synthetic data compared to normal MCMC (black curve).

Another idea is to start MCMC from the set of highest-degree nodes instead of a random set. This can help to avoid over-sampling of very low probability regions, and by doing this, we trim off the front of the Markov chain and shorten the path to settle into the equilibrium state (blue curve).

The method for generating proposals is another important part of MCMC. We consider different ways of sampling to not go far away from the high probability regions. In our basic MCMC algorithm, we find new samples by swapping randomly one node between the current set of sentinels and the rest of the nodes. We consider two alternative methods:

- selecting the next sentinel node with a probability proportional to the average distance of that node to the current sentinel set. This favors nodes that can detect outbreaks that the other sentinels would detect only much later.

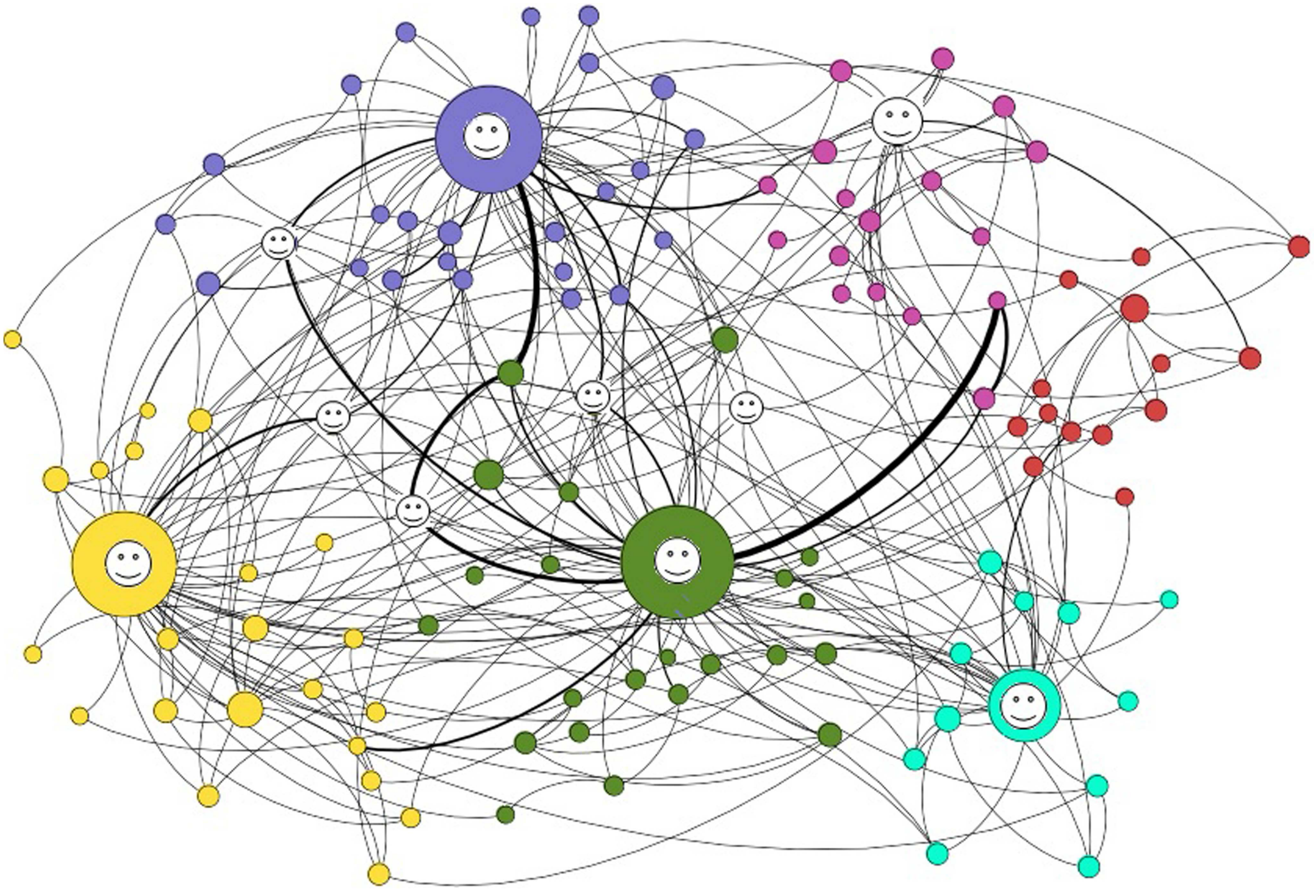

FIG. 3. Sentinel points found by MCMC on a realistic synthetic network $n = 120$; each color shows one community in the network and the smiling faces mark the optimal set of sentinels as selected by the algorithm. Most community centers are chosen as sentinels. The other sentinels chosen form bridges between communities. These nodes are linked to the centers of three or four communities.

- Selecting the next sentinel node with a probability proportional to the degree of that node. Therefore, the node with higher connections has more chance to be in the new sample set.

To find the farthest node to the current sentinel nodes, the average distance between each of the non-sentinel nodes and the current set of sentinel nodes is computed. Then, with a probability proportional to this distance, one node is selected as new sentinel and added to the current set and one of the current sentinels is randomly dropped out. Therefore, the farthest node has a higher chance to be chosen. It should be noted that this method is not time-consuming since the distances between all pairs of nodes are once calculated and stored before execution of the algorithm. In the heuristic section, we observed that higher degree nodes are good candidates for sentinel points. Therefore, our second method benefits from this insight in the sampling stage. We played around with these two methods and the combination of them with the idea of simulated annealing and changing the initial state in MCMC as depicted in Fig. 2. The results show that among them, applying simulated annealing (MCMC-SA) and a combination of starting from higher degree nodes and degree-based sampling (MCMC-Deg-HD) obtain better results. Therefore, we selected these *ad hocs* and apply them to the rest of the experiments.

In addition to the performance of the testing strategy, another output of the simulation is the nodes selected by that strategy. Therefore, we do some analysis on the characteristics of these nodes and also their positions in the network. Figure 3 shows the location of sentinel nodes on a small realistic synthetic network of 120 nodes. The network consists of six communities depicted in different colors, as detected by the community-detection algorithm in Gephi.[44] The ten sentinel points, which are found by the MCMC-Deg-HD method, are marked by a smiling face.

Half of the sentinel nodes are centers of some community except the center of the red community, which is a slaughterhouse. We disregarded slaughterhouses as potential sentinels in our

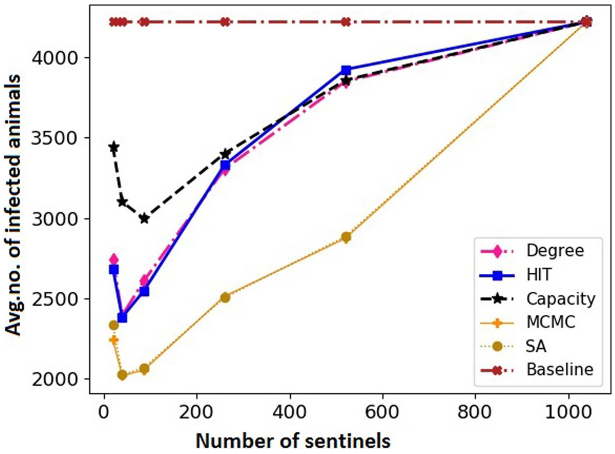

FIG. 4. The performance of all methods on an ensemble of 100 realistic synthetic networks with 1040 nodes and 13 500 edges. Targeted selection of 40 nodes using any of heuristics, MCMC or simulated annealing (SA), which means that testing these nodes every two weeks results in the minimum number of infected cases. Selecting more or less than 40 nodes increases the average size of the outbreak. Additionally, testing strategies based on MCMC or SA outperform all other strategies.

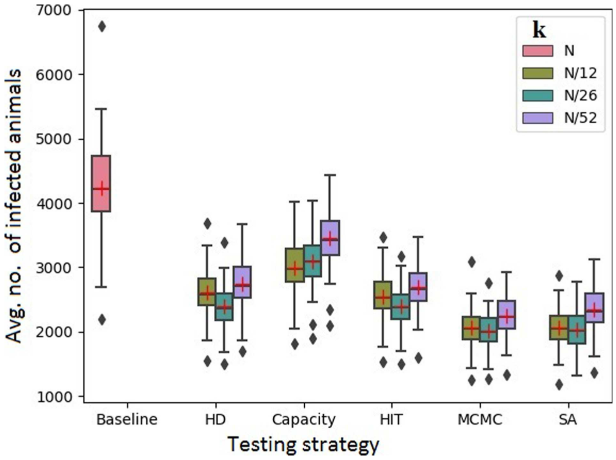

FIG. 5. Performance of the studied testing strategies for different numbers of sentinel nodes. Each box-plot shows how the expected outbreak size at the time of detection varies in an ensemble of 100 realistic synthetic networks with 1040 nodes and 13 500 edges. The expected outbreak size for each single network is estimated by running 10 000 SI simulations and summing up the capacities of nodes infected at the time of detection. The red marker shows the mean for each category.

analysis since in a slaughterhouse, the pathogen cannot spread because animals are killed upon arrival, and therefore, testing only a few animals at a slaughterhouse would not reveal an infection of any of the other animals killed there. Further analysis shows that any outbreak from this community is immediately detected by one of the sentinels in the neighboring communities, as 12 out of 14 nodes in this community with only one link are connected to the centers of other communities or close-by sentinels. A close look into the locations of sentinels shows that they are almost evenly distributed over the network, and five remained sentinels are those nodes that form connections between communities.

One important question is which farm type is better to be targeted as a testing point? We do further analysis to answer this question.

There are different types of farms in the pig-trade network, consisting of breeding farms, fattening farms, traders, and slaughterhouses. The pork production chain begins with breeding and goes through traders, fattening, and traders and ends with a slaughterhouse. Therefore, traders are between different pig growth stages, and most of the transactions in this area are done through them. After checking the type of sentinels, we noticed that they are either traders or fattening farms.

Therefore, in designing a test strategy, health policymakers should consider that between a breeding farm with a higher number of trade connections/transactions and a neighboring fattening farm with a smaller connection, it is more appropriate to pick the fattening farm as a testing point. The reason is that testing a breeding farm can only detect the outbreaks that start from that farm not more since it is placed at the beginning of the production chain but testing a fattening farm increases the chance of detection for all outbreaks that start from any breeding or trader farms connected to this farm.

After successfully applying MCMC on small realistic synthetic networks, we choose simulated annealing and the improved version of MCMC (MCMC-Deg-HD) and apply them on a larger network. The efficiency of each testing scheme in limiting the

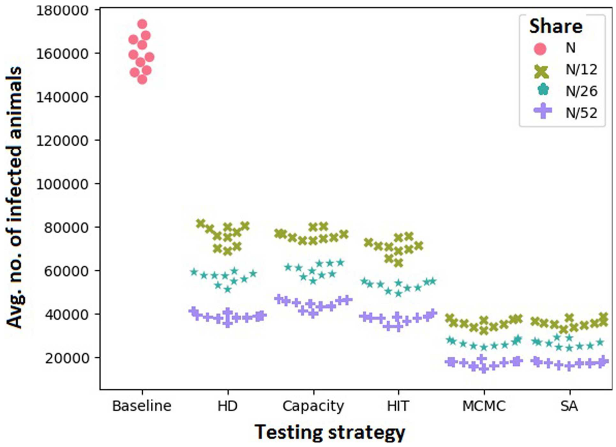

FIG. 6. Performance of the studied testing strategies on an ensemble of 10 real pig-trade networks with 57 590 nodes and 810 000 edges. Each point shows the estimated expected number of infected animals at the time of detection for one real network. The different colors show the influence of the number of nodes selected as sentinels ($k = N/12$, $N/26$, or $N/52$). In all studied methods, selecting only 1/52 of nodes and testing them weekly performed significantly better than, e.g., selecting 1/26 of nodes and testing them biweekly. MCMC and SA clearly outperform all other strategies. The standard error in the averaging of infected cases over all 1000 SI runs and all 10 networks is on average 500 for heuristics and 1200 for MCMC and SA strategies.

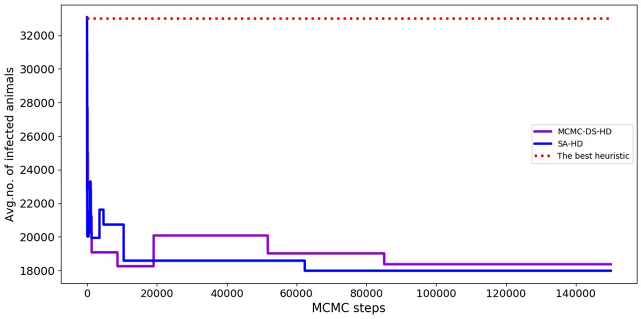

FIG. 7. MCMC and simulated annealing substantially decrease the number of infected animals obtained by best heuristics on real pig-trade data.

outbreak size is depicted in Fig. 4 and as a grouped box-plot in Fig. 5.

The performance of HIT has a range very similar to the degree strategy but generally choosing farms that have more purchasing transactions (HIT) leads to fewer infected cases. The testing strategy based on HIT makes a 44% reduction in the outbreak size compared to the baseline method. Applying the MCMC and SA strategies results in an even more reduction of 55%.

C. MCMC on the real network

Finally, we conduct the experiments on a real pig-trade dataset to find the best testing scheme in a real-world application. To assess the performance in terms of accuracy, similar to the experiments on realistic synthetic networks, we used an ensemble of 10 networks extracted from the real dataset, each containing a different non-overlapped 180-day time window of the transmissions.

Figure 6 shows the estimated average number of infected animals at the time of detection of an outbreak for each of these real networks for different testing strategies and numbers of selected sentinels. At first glance, the figure reveals that all testing strategies outperform the baseline much higher than the realistic synthetic dataset. Also, there is clear discrimination between the results of different sizes of *S* compared to that of a realistic synthetic network in Fig. 5.

As it was mentioned in Sec. III B, one tunable parameter is the fraction of nodes, which is important and necessary for surveillance goals. We infer from the plot that the choice of

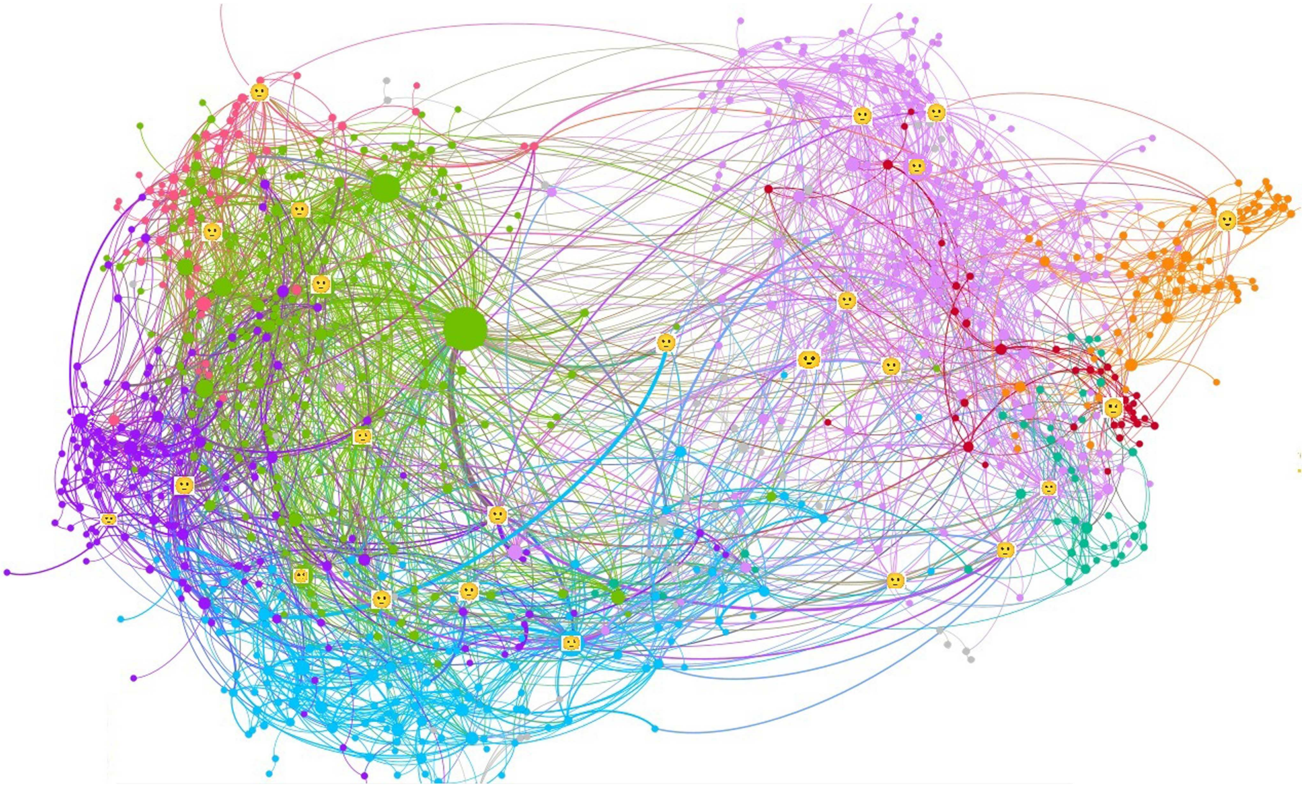

FIG. 8. Sentinel nodes found by MCMC on the real network after applying some filtering in Gephi. Each color shows one community (federal state) in the network, and the smiling faces mark the optimal set of sentinels as selected by the algorithm. Most community centers are chosen as sentinels except those that are of type slaughterhouse (for example, the two big green nodes) or breeding. These farms are at the beginning or at the end of the pork production chain and, therefore, are not good for monitoring purposes. The other sentinels are chosen from bridges between communities. It turns out that they are all fattening farms, i.e., in the middle of the pork production chain.

$k = N/52$ sentinel nodes results in the smallest outbreak size for a real dataset. Choosing an even smaller number of nodes than $N/52$ would mean testing them more often than weekly, which is hardly feasible in practice because of time and resource constraints. Therefore, we do not consider smaller numbers of sentinels in our simulations. For small realistic synthetic networks, the best choice had appeared to be $k = N/26$, while for the large real networks, it is $k = N/52$. If $k = N/52$, nodes are selected and tested weekly rather than all nodes tested yearly, and MCMC and SA reduce the average outbreak size by 89%. In the exemplary MCMC-SA runs shown in Fig. 7, we see that at certain points during the evolution, the algorithm finds innovative candidates that improve the objective considerably in just a single step. In view of the example result shown in Fig. 8, we speculate that these improvements correspond to adding a sentinel node that can detect outbreaks originating in a region of the network not well covered before, such as a center of a network community or a bridge node between communities. The second best method is the "highest incoming transmission" (HIT) heuristic with a 75% decrease.

Another insight from these experiments is that, at least in the German pig production system, a good option, if there were no data on trade at all, would be to simply choose the $N/52$ largest farms (as done by our strategy "Capacity"), which performs only about 4% worse than HIT.

V. CONCLUSION

The outbreak of silent diseases among animal holdings can lead to huge economic loss and threatens public health. Regular testing of farms is one solution that authorities follow in practice to detect outbreaks earlier, which then gives them enough time for implementing the appropriate countermeasures. Recently, the EU urged member countries and states to perform this testing in a more intelligent, "risk-oriented" way considering the constraints on budget, human resources, and medical facilities.[45] In this study, we investigated different testing strategies based on the network measures, farm capacity, and MCMC on both the realistic synthetic and real data. The results confirm that MCMC and simulated annealing outperform all other strategies in terms of the estimated average number of infected animals at the time of detection ("outbreak size"). While simulated annealing provides a smaller outbreak size compared to degree-based MCMC on small networks, both have nearly equal performance on larger networks. Another advantage of simulated annealing is that it converges very fast compared to the MCMC. However, both simulations are numerically sophisticated and computationally expensive methods, which make them infeasible to apply on very large networks. Therefore, a simple and practical strategy with acceptable and comparable performance is preferable to these numerical techniques. Among different heuristics, which are studied in this paper, targeting farms with the highest number of incoming transmissions (the number of times a farm receives shipping of animals) as testing points and testing them every one or two weeks can reduce the damage of outbreaks almost as much as the optimized selections produced by the MCMC.

ACKNOWLEDGMENTS

This work was supported by the German Bundesministerium für Bildung und Forschung (No. FKZ 01KI1812) as part of the Forschungsnetz Zoonosen. We thank Mehrnaz Anvari, Jürgen Kurths, Jacob Löbkens, Nora Molkenthin, and Malte Roloff for helpful discussions and valuable comments.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Sara Ansari: Conceptualization (equal); Data curation (equal); Formal analysis (lead); Investigation (lead); Methodology (equal); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). **Jobst Heitzig:** Conceptualization (equal); Data curation (equal); Formal analysis (supporting); Funding acquisition (lead); Methodology (equal); Project administration (lead); Resources (lead); Supervision (lead); Validation (supporting); Writing – review & editing (equal). **Mohammad R. Moosavi:** Supervision (supporting); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are openly available in HI-Tier at https://www.hi-tier.de/info03.html.

REFERENCES

[1] N. M. Ferguson, C. A. Donnelly, and R. M. Anderson, "The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions," Science **292**, 1155–1160 (2001).

[2] J. Fritzemeier, J. Teuffert, I. Greiser-Wilke, C. Staubach, H. Schlüter, and V. Moennig, "Epidemiology of classical swine fever in Germany in the 1990s," Vet. Microbiol. 77, 29–41 (2000).

[3] A.N.S.V.S.A., "Updated situation of African swine fever in Romania," (2018), https://food.ec.europa.eu/system/files/2018-04/reg-com_ahw_20180418_pres_asf_rou.pdf.

[4] FAO, "ASF situation in Asia update," (2020), http://www.fao.org/ag/againfo/programmes/en/empres/ASF/situation_update.html.

[5] Forbes, "China's swine fever crisis will impact global trade well into 2020," (2019), https://www.forbes.com/sites/siminamistreanu/2019/12/28/chinas-swine-fever-crisis-will-impact-global-trade-well-into-2020.

[6] S. Ito, J. Bosch, M. Martínez-Avilés, and J. M. Sánchez-Vizcaíno, "The evolution of African swine fever in China: A global threat?," Front. Vet. Sci. **9**, 828498 (2022).

[7] M. H. Woodford and P. B. Rossiter, "Disease risks associated with wildlife translocation projects," in *Creative Conservation* (Springer, 1994), pp. 178–200.

[8] A. A. Cunningham, "Disease risks of wildlife translocations," Conserv. Biol. **10**, 349–353 (1996).

[9] L. H. Taylor, S. M. Latham, and M. E. J. Woolhouse, "Risk factors for human disease emergence," Philos. Trans. R. Soc. Lond. B Biol. Sci. **356**, 983–989 (2001).

[10] E. M. Fèvre, B. M. D. C. Bronsvoort, K. A. Hamilton, and S. Cleaveland, "Animal movements and the spread of infectious diseases," Trends Microbiol. **14**, 125–131 (2006).

[11] C. M. Aiello, K. E. Nussear, A. D. Walde, T. C. Esque, P. G. Emblidge, P. Sah, S. Bansal, and P. J. Hudson, "Disease dynamics during wildlife translocations: Disruptions to the host population and potential consequences for transmission in desert tortoise contact networks," Anim. Conserv. **17**, 27–39 (2014).

[12] M. E. Craft, "Infectious disease transmission and contact networks in wildlife and livestock," Philos. Trans. R. Soc. Lond. B Biol. Sci. **370**, 20140107 (2015).

[13] K. Lebl, H. H. K. Lentz, B. Pinior, and T. Selhorst, "Impact of network activity on the spread of infectious diseases through the German pig trade network," Front. Vet. Sci. **3**, 48 (2016).

[14] H. H. K. Lentz, A. Koher, P. Hövel, J. Gethmann, C. Sauter-Louis, T. Selhorst, and F. J. Conraths, "Disease spread through animal movements: A static and temporal network analysis of pig trade in Germany," PLoS One **11**, e0155196 (2016).

[15] K. VanderWaal, A. Perez, M. Torremorrell, R. M. Morrison, and M. Craft, "Role of animal movement and indirect contact among farms in transmission of porcine epidemic diarrhea virus," Epidemics **24**, 67–75 (2018).

[16] E. De Wit, N. Van Doremalen, D. Falzarano, and V. J. Munster, "SARS and MERS: Recent insights into emerging coronaviruses," Nat. Rev. Microbiol. **14**, 523–534 (2016).

[17] E. I. Azhar, S. A. El-Kafrawy, S. A. Farraj, A. M. Hassan, M. S. Al-Saeed, A. M. Hashem, and T. A. Madani, "Evidence for camel-to-human transmission of MERS coronavirus," N. Engl. J. Med. **370**, 2499–2505 (2014).

[18] I. M. Mackay and K. E. Arden, "MERS coronavirus: Diagnostics, epidemiology and transmission," Virol. J. **12**, 222 (2015).

[19] S. Lytras, W. Xia, J. Hughes, X. Jiang, and D. L. Robertson, "The animal origin of SARS-CoV-2," Science **373**, 968–970 (2021).

[20] Z.-W. Ye, S. Yuan, K.-S. Yuen, S.-Y. Fung, C.-P. Chan, and D.-Y. Jin, "Zoonotic origins of human coronaviruses," Int. J. Biol. Sci. **16**, 1686 (2020).

[21] M. Rahman, M. Sobur, M. Islam, S. Ievy, M. Hossain, M. E. El Zowalaty, A. Rahman, and H. M. Ashour, "Zoonotic diseases: Etiology, impact, and control," Microorganisms **8**, 1405 (2020).

[22] A. Hassani and G. Khan, "Human-animal interaction and the emergence of SARS-CoV-2," JMIR Public Health Surveill. **6**, e22117 (2020).

[23] J. Huynh, S. Li, B. Yount, A. Smith, L. Sturges, J. C. Olsen, J. Nagel, J. B. Johnson, S. Agnihothram, J. E. Gates, and M. B. Frieman, "Evidence supporting a zoonotic origin of human coronavirus strain NL63," J. Virol. **86**, 12816–12825 (2012).

[24] G. Wong, Y.-H. Bi, Q.-H. Wang, X.-W. Chen, Z.-G. Zhang, and Y.-G. Yao, "Zoonotic origins of human coronavirus 2019 (HCoV-19/SARS-CoV-2): Why is this work important?," Zool. Res. **41**, 213 (2020).

[25] J. G. Pritchett, D. D. Thilmany, and K. K. Johnson, "Animal disease economic impacts: A survey of literature and typology of research approaches," Int. Food Agribus. Manag. Rev. **8**, 23–45 (2005).

[26] T. E. Carpenter, J. M. O'Brien, A. D. Hagerman, and B. A. McCarl, "Epidemic and economic impacts of delayed detection of foot-and-mouth disease: A case study of a simulated outbreak in California," J. Vet. Diagn. Invest. **23**, 26–33 (2011).

[27] K. M. Rich and B. D. Perry, "The economic and poverty impacts of animal diseases in developing countries: New roles, new demands for economics and epidemiology," Prev. Vet. Med. **101**, 133–147 (2011).

[28] P. Carlsson-Szlezak, M. Reeves, and P. Swartz, "What coronavirus could mean for the global economy," Harv. Bus. Rev. **3**, 1–10 (2020).

[29] N. Madhav, B. Oppenheim, M. Gallivan, P. Mulembakani, E. Rubin, and N. Wolfe, "Pandemics: Risks, impacts, and mitigation," in *Disease Control Priorities: Improving Health and Reducing Poverty*, edited by D. T. Jamison, H. Gelband, S. Horton, P. Jha, R. Laxminarayan, C. N. Mock, and R. Nugent (World Bank Group, 2017).

[30] A. Delivorias and N. Scholz, "Economic impact of epidemics and pandemics," Report No. EPRS_BRI(2020)646195 (European Parliamentary Research Service, 2020), https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2020)646195.

[31] A. Ahani and M. Nilashi, "Coronavirus outbreak and its impacts on global economy: The role of social network sites," J. Soft Comput. Decis. Support Syst. **7**, 19–22 (2020).

[32] H. Yilmazkuday, "Coronavirus disease 2019 and the global economy," Transp. Policy **120**, 40–46 (2022).

[33] Y. Shang, H. Li, and R. Zhang, "Effects of pandemic outbreak on economies: Evidence from business history context," Front. Public Health **9**, 632043 (2021).

[34] C. Esseau-Thomas, O. Galarraga, and S. Khalifa, "Epidemics, pandemics and income inequality," Health Econ. Rev. **12**, 7 (2022).

[35] S. Ansari, J. Heitzig, L. Brzoska, H. Lentz, J. Mihatsch, J. Fritzemeier, and M. Moosavi, "A temporal network model for livestock trade systems," Front. Vet. Sci. **8**, 766547 (2021).

[36] StMELF, "Herkunftssicherungs und Informationssystem für Tiere," (2019) https://www.hi-tier.de/info03.html, viewed 26.03.19.

[37] G. Sabidussi, "The centrality index of a graph," Psychometrika **31**, 581–603 (1966).

[38] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," Nat. Phys. **6**, 888–893 (2010).

[39] L. C. Freeman, "A set of measures of centrality based on betweenness," Sociometry **40**, 35–41 (1977).

[40] J. L. Herrera-Diestra, M. Tildesley, K. Shea, and M. J. Ferrari, "Cattle transport network predicts endemic and epidemic foot-and-mouth disease risk on farms in Turkey," PLoS Comput. Biol. **18**, e1010354 (2022).

[41] Y. Liu, M. Tang, T. Zhou, and Y. Do, "Improving the accuracy of the k-shell method by removing redundant links: From a perspective of spreading dynamics," Sci. Rep. **5**, 13172 (2015).

[42] S. Ansari, M. Anvari, O. Pfeffer, N. Molkenthin, M. R. Moosavi, F. Hellmann, J. Heitzig, and J. Kurths, "Moving the epidemic tipping point through topologically targeted social distancing," Eur. Phys. J. Spec. Top. **230**, 3273–3280 (2021).

[43] J. C. Maxwell, "V. Illustrations of the dynamical theory of gases.—Part I. On the motions and collisions of perfectly elastic spheres," Lond., Edinb., Dublin Philos. Mag. J. Sci. **19**, 19–32 (1860).

[44] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," J. Stat. Mech.: Theory Exp. **2008**, P10008.

[45] European Union law, Document 32020R0689, http://data.europa.eu/eli/reg_del/2020/689/oj, paragraph (8).