Chapter 10
Surveillance for Outbreak Detection in Livestock-Trade Networks

Frederik Schirdewahn, Vittoria Colizza, Hartmut H. K. Lentz, Andreas Koher, Vitaly Belik, and Philipp Hövel

Abstract We analyze an empirical, temporal network of livestock trade and present numerical results of epidemiological dynamics. The considered network is the backbone of the pig trade in Germany, which forms a major route of disease spreading between agricultural premises. The network is comprised of farms that are connected by a link, if animals are traded between them. We propose a concept for epidemic surveillance, which is generally performed on a subset of the system due to limited resources. The goal is to identify agricultural holdings that are more likely to be infected during the early phase of an epidemic outbreak. These farms, which we call *sentinels*, are excellent candidates to monitor the whole network. To identify potential sentinel nodes, we determine most probable transmission routes by calculating functional clusters. These clusters are formed by nodes that – chosen as seed for an outbreak – have similar invasion paths. We find that it is indeed possible to group the German pig-trade network in such clusters. Furthermore, we select sentinels by choosing nodes out of every cluster. We argue that any epidemic outbreak can be reliably detected at an early stage by monitoring a small number

F. Schirdewahn (✉) • A. Koher • P. Hövel
Institute of Theoretical Physics, Technische Universität Berlin, Hardenbergstr. 36, 10623, Berlin, Germany
e-mail: frederik.schirdewahn@t-online.de; andreas.koher@campus.tu-berlin.de; phoevel@physik.tu-berlin.de

V. Colizza
Sorbonne Universités, UPMC Univ Paris 06, INSERM, Institut Pierre Louis d'épidémiologie et de Santé Publique (IPLESP UMRS 1136), 75012, Paris, France
e-mail: vittoria.colizza@inserm.fr

H.H.K. Lentz
Institute of Epidemiology, Friedrich-Loeffler-Institut, Südufer 10, 17493, Greifswald, Insel Riems, Germany
e-mail: hartmut.lentz@fli.de

V. Belik
System Modeling Group, Institute for Veterinary Epidemiology and Biostatistics, Freie Universität Berlin, Königsweg 67, 14163, Berlin, Germany
e-mail: vitaly.belik@fu-berlin.de

N. Masuda, P. Holme (eds.), *Temporal Network Epidemiology*, Theoretical Biology, DOI 10.1007/978-981-10-5287-3_10

of those sentinels. Considering a susceptible-infected-recovered model, we show that an outbreak can be detected with only 18 sentinels out of almost 100,000 farms with a probability of 65% in approximately 13 days after first infection. This finding can be further improved by including nodes with the largest in-component (highest vulnerability), which increases the detection probability to 86% within 8 days after first occurrence of the disease.

10.1 Introduction

Diseases in livestock holdings have been a major challenge in the industrial meat production and related economy in the last decades. For example, the foot-and-mouth disease (FMD), which broke out in Great Britain in 2001 in herds of cloven hoofed animals, caused estimated costs of 8 billion British Pound [1]. In rare occasions FMD could even pose a health risk to humans, which means that it becomes zoonotic, that is, it can be transferred from animals to humans. In general, outbreaks of animal-related diseases should be prevented for multiple reasons: They diminish animal well-being, reduce productivity, cause great economic losses, and might be transferable to human.

The study of spreading livestock diseases contributes to a better understanding of contagion processes in general [2]. To model an infection many mathematical models have been successfully investigated such as the SIR (susceptible-infected-recovered), SIS (susceptible-infected-susceptible) or SI (susceptible-infected) model [1, 3–5]. Major transmission routes of disease spread may be geographical proximity, where aerial transmission is the main carrier. In addition, arthropods (mosquitoes or ticks) can be vectors. We will focus on the trade of livestock, which was the main route due to direct transmission between animals, for instance, during a swine-fever outbreak in Germany in the 1990s [6]. The disease transmission between animal holdings takes place, if an infected animal is transported from one farm to the next. To model and analyze the impact of disease spread due to livestock trade, we use concepts from network science [7].

Since livestock-trade networks span tens of thousands of agricultural holdings, it is not possible to examine every single farm for an infection due to limited resources. Examinations should therefore focus on some premises with a high probability of being infected in case of an outbreak. In Ref. [8], Bajardi et al. analyzed the Italian cattle-trade network and presented a novel surveillance concept. We will apply the same framework to identify special nodes, the so called *sentinels*, that may be affected by a potential outbreak occurring in the system with a high probability. We will demonstrate that the number of sentinel nodes is several orders of magnitude smaller than the total number of animal holdings. For this purpose, we consider different selection protocols and show that surveillance can be made much more efficient by concentrating resources on a few nodes.

This chapter focuses on the data of the German pork industry, which is one of the largest in the world. Every year, five million tons of pork meat are produced and the rate is increasing.[1] Therefore, investigating efficient detection schemes on the underlying network is of great relevance.

The rest of this chapter is organized as follows: In Sect. 10.2, we will introduce the susceptible-infected-recovered model and some concepts from network science. We will show how an invasion path evolves on a temporal network allowing to define functional clusters. In Sect. 10.3, we describe the data under consideration and summarize the steps taken to analyze the network on a temporal basis. Furthermore, we apply strategies proposed in Ref. [8] to the network and discuss the possibility to identify sentinel nodes. Finally, we conclude with a summary in Sect. 10.4.

10.2 Theory

In the following, we will review basic aspects of the susceptible-infected-recovered (SIR) model (Sect. 10.2.1) and discuss how an epidemic can spread in a network via invasion paths (Sect. 10.2.2). We provide details on our numerical simulation in Sect. 10.2.3. The characterization of different nodes in the network according to their in- and out-components will be the topic of Sect. 10.2.4 and we will elaborate how clusters evolve from different invasion paths in Sect. 10.2.5.

10.2.1 Deterministic Susceptible-Infected-Recovered-Model

To describe the spreading of an infectious disease in a population, we need a model for its progression [1]. Let us assume that size of the population is constant and that it can be divided in susceptible (or healthy) S, infected (and therefore infectious) I and recovered (and hence immunized) individuals R. Following the transition scheme

$$S \xrightarrow{\alpha} I \xrightarrow{\beta} R$$

a susceptible individual becomes infected with a probability α upon contact with an infected. After an infectious period of β^{-1}, where β denotes the recovery rate, an infected individual turns into a recovered one. Note that this scheme does not account for births, deaths, or migration. In our study, we consider a deterministic version of the SIR model with a fixed recovery time and guaranteed infection upon

[1] *Agrarpolitischer Bericht der Bundesregierung* (2015). *Bundesministerium für Ernährung und Landwirtschaft* (BMEL), available as http://www.bmel.de/SharedDocs/Downloads/Broschueren/Agrarbericht2015.html

contact, that is, $\alpha = 1$ [8, 9]. Alternatively, the SIR dynamics can also be written as a set of differential equations [10].

Livestock diseases may spread directly between animals. Here, we model a corresponding contagion process on a broader perspective by considering the agricultural holding as epidemiological unit. Our main goal is not to investigate a detailed model for the local disease dynamics within a farm. Instead, we assume that every infected animal will transmit the disease immediately to the whole population, when it arrives at another farm. In the beginning of each simulation, all premises are considered as susceptible or disease free except for a single node [8, 11], which we call the *seed*. The infection is transmitted in each time step along outgoing links connected to susceptible neighbors, which then transmit the disease in the following time step further in the network via their susceptible neighbors and so on. In short, the considered model consists of two dynamical mechanisms [8]:

1. A susceptible farm will be infected with a probability $\alpha = 1$, if it receives an animal from an infected farm.
2. A farm stays infected for a duration of τ days, which we call the *infectious period*. We set this value to $\tau = \beta^{-1} = 7$ days. Afterwards, the farm recovers and cannot be infected again.

Note that the first mechanism implicitly accounts for directionality. Opposed to other mobility scenarios such as commuting, only the node at the end of an edge is at risk to become infected in a production chain. If a susceptible farm sells an animal to an infected one, it still maintains its disease-free status. The advantage of such a deterministic model is a significant reduction of computational effort. It allows us to consider all nodes as a possible starting point of an outbreak. In short, our numerical findings provide information in terms of a worst-case scenario. Bajardi et al. also obtained similar results using a stochastic modeling approach [8].

The next sections describe how an infection takes place on a temporal network and how the algorithm used in this study is implemented.

10.2.2 Temporal Networks

As Vernon and Keeling pointed out in Ref. [12], the spread of infectious diseases is only predicted correctly, if the chronology of contacts is accurately accounted for. For a realistic model of disease transmission, we therefore consider a directed temporal network, because typical trade connections take place on different timescales and a disease can only be transmitted along time-respecting paths.

Next, we will give a short introduction into the mathematical description of temporal networks [13–15]. We define $G = (V, E)$ as a directed, temporal graph consisting of a set of nodes V and time-stamped edges E connecting these nodes. For further reading, in particular connected to livestock-trade networks, we refer to [8, 9, 12, 16, 11, 17, 18].

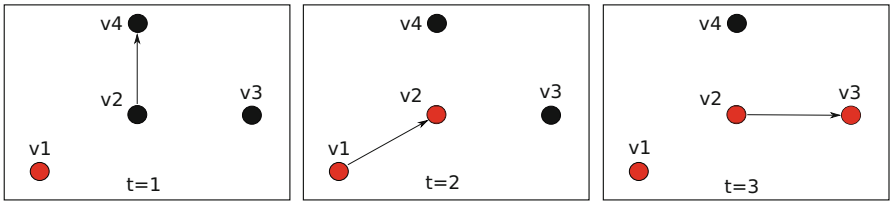

Fig. 10.1 Snapshot of a schematic network for three different times. Initially node v_1 is infected (indicated by the *red dot*) and the disease can spread to node v_3, which is susceptible (indicated by the *black dot*), via v_2

If an outbreak at a node v_i can reach a node v_j, there has to be either a direct link, that is, an edge, or an indirect connection. The latter case is described by a path from one to the other. Such a path P_{ij} consists of a sequence of edges via intermediate nodes v_k, where no node is visited twice. Therefore, a path is given by:

$$P_{ij} = \left[(v_i, v_1, t_0), (v_1, v_2, t_1), \ldots, \left(v_{n-1}, v_j, t_{n-1}\right) \right].$$

The length of the path is the number of edges n. Note that we introduce a time stamp to each edge of the path. Hence, a time-respecting path satisfies: $t_0 < t_1 < \cdots < t_{n-1}$. Between a pair of nodes, there might be a large number of paths of different lengths [19]. We stress that a path with the smallest number of edges might not be the fastest depending on the specific timing of its edges [20]. For a disease spread between two nodes, the earliest arrival time is of particular importance. We call the set of directed, time-respecting edges starting at a particular seed node *invasion path* Γ . In the considered deterministic SIR model, just the first contact with the disease infects the node. Recurrent infections will have no effect as repeated infections are not possible.

Figures 10.1, 10.2 and 10.3 provide different perspectives of a spreading process on a temporal network. The disease starts at a single infected node v_1. While Fig. 10.1 depicts a series of snapshots at different times, Fig. 10.2 shows an overlay of the snapshots, where the times, when an edge is active, are explicitly given. In this schematic example, an invasion path $\Gamma_{13} = [(v_1, v_2, t = 2), (v_2, v_3, t = 3)]$ exists between the initially infected seed node v_1 and node v_3 via v_2. Node v_4, however, cannot be infected, because the connection P_{24} takes place, before the outbreak reaches v_2. Hence, the path $P_{14} = [(v_1, v_2, t = 2), (v_2, v_4, t = 1)]$ is not time-respecting and violates causality. The notion of an invasion path includes the possibility of branching into tree-like transmission routes. A time-layered aspect is depicted in Fig. 10.3. Here, the number of nodes that are going to be infected in every time step, so called *incidences*, is easy to see.

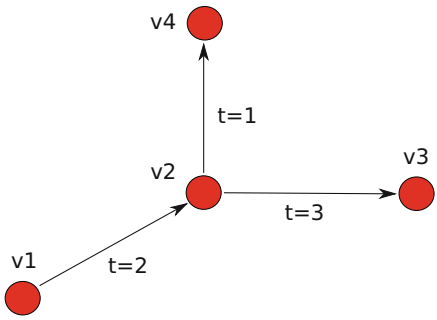

Fig. 10.2 Overlay of snapshots of a temporal network (cf. Fig. 10.1). A time-respecting path leads from node v_1 to v_3. If one aggregates the network over all times, however, a path between v_1 and v_4 emerges that does not exist in the temporal case

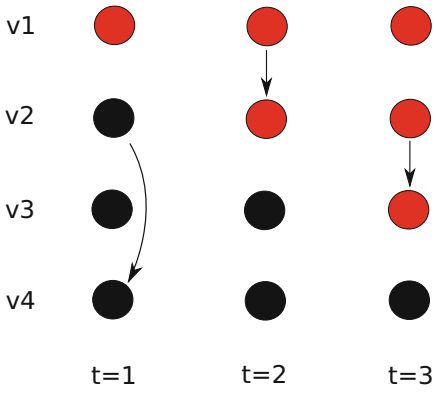

Fig. 10.3 The same temporal network as in Figs. 10.1 and 10.2, but in a layered representation. In time step t=1 two susceptible nodes have contact. Only in step t=2 and t=3 the disease can be transmitted

10.2.3 Modelling an Infection on the Network

To model the spread of an infectious disease on the network, we use an algorithm of breadth-first-search type to iteratively simulate the deterministic SIR dynamics introduced in Sect. 10.2.1. The main steps are the following: We start at a seed node $v_i \in V$ and mark it as infected at time t_0. In every time step t_n, we identify all edges (v_i, v_j, t_n) that start at the initially infected node v_i (or in further steps at nodes along the production chain originating from v_i) and lead to a susceptible node v_j. All nodes that can be reached this way are marked as infected, that is, we assume a transmissibility of 100%. A node can transmit the disease as long as it is infected. After having acquired an infection, the node stays infected and infectious for a fixed period, which we choose as $\tau = 7$ days. Subsequently, we iterate over all infected nodes v_i and mark those, which have recovered, as removed. In the next step the time t_n is incremented by one corresponding to the temporal resolution of the available data and the process will be repeated, until no more infected nodes are present. The time that it takes from the beginning of the outbreak to its termination is called the *outbreak duration*.

In the next section, we will summarize some measures of a temporal network, which help to characterize its structure.

10.2.4 Measure of Centrality

There is a large number of measures that quantify the centrality of nodes in a network [21–23]. For epidemiological purposes, central nodes may have a high chance to become infected or may transmit a disease to large parts of the network. In this section, we will focus on some of those measures that have a direct epidemiological relevance.

In network terminology, the *out-component* $c_{\text{out}}(v_i,\ \tau,\ t_0)$ of a node v_i is given by a set of nodes that can be reached from a primary infected node $v_i \in V$. The parameter τ is the finite infectious period introduced in Sect. 10.2.1 and t_0 denotes the starting time of the epidemic. In general, a large infectious period τ produces more secondary outbreaks and leads therefore to a greater probability to reach more nodes in the network [11]. $c_{\text{out}}(v_i,\ \tau,\ t_0)$ can be calculated as the union of the sets of nodes along all possible invasion paths originating from v_i at time t_0. This out-component corresponds to the final size of an epidemic, which is an important quantity in epidemiology. It indicates the accumulated number of all infected individuals during an epidemic. The impact of a node in terms of the size of its out-component can be interpreted as a measure of centrality.

Another important network property is the set of nodes, from which a particular node $v_j,\ \in V$ can be infected. This is called *in-component* $c_{\text{in}}(v_j,\ \tau,\ t_0)$. The size of the in-component can be used as a measure for the *vulnerability* of a node [11]. Furthermore, we define the *out-degree* k_i^{out} and *in-degree* k_i^{in} of node v_i as the number of edges, which leave a node (selling events) or arrive at a node (buying events) aggregated over the whole observation time, respectively.

After this brief excursion to notions from network science connected to epidemiology, we will introduce additional aspects such as seed clusters, which contain nodes with similar invasion paths and spreading behavior, in the next section.

10.2.5 Invasion Path and Seed Clusters

If we consider a node v as infectious and if it has contact with susceptible nodes during its infectious period over some directed links e, the disease will be transmitted in the framework of the considered deterministic SIR model. If this node is the origin of the disease, we call it a seed. All nodes, which will be infected as time goes on, are part of one of its invasion paths at least. As defined in Sect. 10.2.2, an invasion path of length $n \in \mathrm{N}$ consists of a set of directed edges $\{e_0, ..., e_{n-1}\} \subseteq E$ connecting a set of nodes $\{v_0, ..., v_n\} \subseteq V$ at times $t_0 < \cdots < t_{n-1}$.

The invasion paths depend strongly on the initial conditions given by the starting time t_0 and seeding node v_i. To explore the dependence of the spreading process on the initial conditions, we aim to identify similar spreading patterns. For this purpose, we use the unions Γ_1 and Γ_2 of invasion paths of two seeds at a fixed starting time t_0 to compute the similarity between them. We define the *Jaccard index* Θ_{12} as the relative overlap of the two sets measured by the number of their common nodes:

$$\Theta_{12} = \frac{|\Gamma_1 \cap \Gamma_2|}{|\Gamma_1 \cup \Gamma_2|}, \tag{10.1}$$

where $|\Gamma|$ denotes the number of nodes. In words, we calculate the fraction of the sizes of the intersection between the two node sets and their union. Consider Fig. 10.4, where a schematic example of two invasion paths $\Gamma_1 = [(v_1, v_3), (v_3, v_4), (v_4, v_6), (v_6, v_8)]$ and $\Gamma_2 = [(v_2, v_3), (v_3, v_4), (v_4, v_6), (v_6, v_9)]$ is shown in blue and red, respectively. We find a Jaccard index of $\Theta_{12} = |\Gamma_1 \cap \Gamma_2|/|\Gamma_1 \cup \Gamma_2| = 3/7$ as the relative overlap of the two paths.

Since the disease can in principle start from any node, we need to consider every node pair at a fixed starting time t_0 and evaluate the similarity of their invasion paths. If we calculate this overlap Θ_{ij} between every pair of potential seeds (v_i, v_j), it is possible to construct a weighted and undirected network, which is called the *initial-condition similarity network* (see Fig. 10.5). In that network, nodes refer to invasion paths, which are determined by their seed. The strength of a link between two invasion paths Γ_i and Γ_j is given by the overlap $\Theta_{ij} \in [0, 1]$. This gives rise to

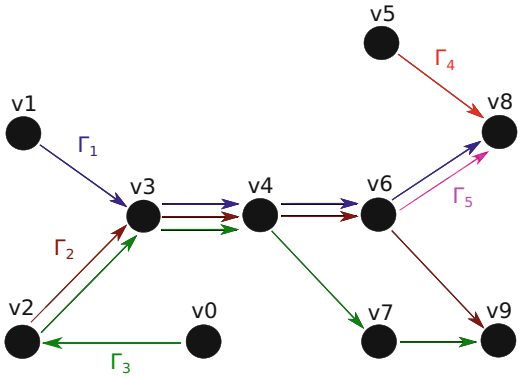

Fig. 10.4 Overlap between invasion paths $\Gamma_1 = [(v_1, v_3), (v_3, v_4), (v_4, v_6), (v_6, v_8)]$ (*blue*), $\Gamma_2 = [(v_2, v_3), (v_3, v_4), (v_4, v_6), (v_6, v_9)]$ (*red*), $\Gamma_3 = [(v_0, v_2), (v_2, v_3), (v_3, v_4), (v_4, v_7), (v_7, v_9)]$ (*green*), $\Gamma_4 = [(v_5, v_8)]$ (*orange*), and $\Gamma_5 = [(v_6, v_8)]$ (*pink*). The paths Γ_1 and Γ_2 have nodes v_3, v_4, and v_6 in common, which results in a Jaccard index $\Theta_{12} = 3/7$.The value $\Theta_{23} = 4/7$ is found for Γ_2 and Γ_3, but not for Γ_1 and Γ_3, which is $\Theta_{13} = 2/9$. The connection of two nodes v_5 and v_6 to the same final node v_8 with the paths Γ_4 and Γ_5 can be seen as a triadic motif for a relatively high Jaccard index of $\Theta_{45} = 1/3$. The Jaccard index between Γ_1 and Γ_4 is $\Theta_{14} = 1/6$

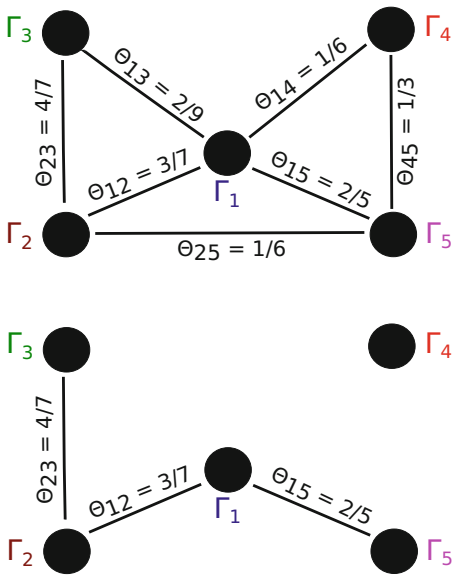

Fig. 10.5 *Top*: the undirected and weighted similarity network according Fig. 10.4 based on different initial conditions emerges out of the overlap of the respective invasion paths (only non-zero overlap shown). The network is weighted by the overlap (Jaccard index). *Bottom*: exemplary depiction of the emergence of cluster if a threshold of $\Theta_{\text{th}} \in (1/3,\ 2/5]$ is applied. In this example one cluster contains the seeding nodes v_1, v_2, and v_6 of the invasion paths Γ_1, Γ_2, Γ_3, and Γ_5. A second cluster refers to just one seeding node v_5 of the invasion path Γ_4

an all-to-all connected network. If we apply a threshold Θ_{th} to the edge weights in that network and disregard smaller ones, the resulting network disintegrates and we obtain subsets of nodes with similar invasion paths Γ. This thresholding can lead to disconnected subgraphs and we call their connected componenents *clusters*. The bottom panel of Fig. 10.5 depicts the two clusters obtained for invasions paths of Fig. 10.4 for a threshold of $1/3 < \Theta_{\text{th}} \leq 2/5$. We define the size of a cluster by the number of seed nodes at the origin of the invasion paths that lead to the formation of that cluster.

Note that it is not required that all nodes in the same cluster are connected with each other by an invasion path. If two nodes $v_i \subseteq V$ and $v_j \subseteq V$ belong to the same cluster, it simply means that there is a set of other nodes $\{v_1, v_2, ..., v_p\} \subseteq V$ that have an overlap $\Theta_{i1}, \Theta_{12}, \cdots, \Theta_{pj}$ greater than the threshold, but not necessarily that the overlap Θ_{ij} is greater than Θ_{th}. See, for instance, the Jaccard index for the two pairs of invasion paths (Γ_1, Γ_2) and (Γ_2, Γ_3) in Fig. 10.4. The respective overlaps are $\Theta_{12} = 3/7$ and $\Theta_{23} = 2/3$, although the Jaccard index between Γ_1 and Γ_3 is smaller: $\Theta_{13} = 1/4$. It is also important to note that these different clusters evolve over time. Invasion paths, from which clusters are computed, refer to the same starting time t_0. Since an invasion path depends on t_0, the clusters are time dependent, too. The robustness of the clusters will be the topic of Sect. 10.3.6.

Based on our numerical simulations, we measure the overlap of every possible pair of seeds to group nodes in clusters. Note that geographical proximity is not a necessary initial condition for this network-based procedure. Therefore, two nodes that have a great geographical distance can be part of the same cluster because of their similar invasion paths.

Many nodes considered as seeds for an outbreak lead to short infection paths [8, 19], but high Jaccard indexes. See, for instance, the triadic motif depicted in Fig. 10.4. Two premises (node v_5 and v_6) are just connected to the same dead end (node v_8), that is most likely, a slaughterhouse, which yields an overlap of 1/3. To avoid these misleading high values, we consider only infection paths that contain at least 10 nodes. Both of these restrictions still lead to the emergence of non-trivial clusters of initial conditions, that is, other than single, isolated nodes. Nodes with an out-component $|c_{\text{out}}| \geq 10$ nodes have a high spreading potential and usually belong to a part of the network that is called *giant in-component* (GIC) or *giant strongly connected component* (GSCC) [19]. The latter is defined as a set of nodes, in which any pair of nodes is connected by a directed, time-respecting path. The GIC consists of an additional set of nodes that are not part of the GSCC, but are connected to the GSCC via time-respecting paths.

In the next section, we will present the methodology to compare clusters obtained for different starting times. This will lead to the analysis of the robustness of the clusters.

10.2.6 Measurement of the Robustness of the Clusters Over Time

The method described in the last section leads to a partition $\{C_1(t_0),\ C_2(t_0),\ \cdots\}$ of different clusters based on the similarity of invasion paths with starting time t_0 [8]. We will consider only the M largest clusters in the following. To measure the robustness of a cluster $C_i(t_0)$ at a later time t, we compute the relative change of the cluster size in comparison to any of the M largest clusters:

$$\rho_{ij}\,(t_0, t) = \frac{\mid C_i\,(t_0) \cap C_j(t)\mid}{\mid C_i\,(t_0)\mid}. \tag{10.2}$$

This $M \times M$ matrix $\{\rho_{ij}\}$ represents in every row $\rho_i(t_0,\ t)$ the fraction of nodes of $C_i(t_0)$ present in the cluster $C_j\ (t)$, which is computed according to invasion paths starting at time t. If the cluster $C_i(t_0)$ persists or becomes part of one larger cluster, the row $\rho_i(t_0,\ t)$ will have one entry equal to 1, and all others will be zero. Similarly, when all nodes of $C_i(t_0)$ are redistributed over the M largest clusters, the sum over the i-th row will be unity. Following this intuition, we define a robustness measure by $\sigma_i\,(t_0, t) = \sum_{j=1}^{M} \rho_{ij}\,(t_0, t)$. This quantity will be smaller than 1, if some nodes of cluster $C_i(t_0)$ are not present in any of the M largest clusters at time t. Note that for $t_0 \neq t$, the matrix $\{\rho_{ij}\}$ does not need to be symmetric, because the M largest clusters might differ considerably in size and node set for different times.

For further quantitative analysis, we compute the conditional entropy of the i-th cluster defined as

$$H_i(t_0,t) = \frac{\sum_{j=1}^{M} \rho_{ij}(t_0,t) \log\left[\rho_{ij}(t_0,t)\right]}{\sigma_i(t_0,t) \log \frac{\sigma_i(t_0,t)}{M}}. \tag{10.3}$$

The entropy quantifies the redistribution among the M largest clusters at time t in comparison to an earlier time t_0. The entropy vanishes ($H_i = 0$), if $C_i(t_0)$ is also a cluster at time t. Apart from this extreme case of stationary clusters, the minimum entropy is given by $H_{\min,i}(t_0, t) = [1 - \log(M)/\log(\sigma_i)]^{-1}$, if all nodes of $C_i(t_0)$ are found in exactly one cluster $C_k(t)$ at time t except a fraction $(1 - \sigma_i)$ of them that do not belong to any of the M largest clusters anymore. This configuration yields: $\rho_{ik}(t_0, t) = \sigma_i(t_0, t)$ and $\rho_{ij}(t_0, t) = 0$ for $j \neq k$ and we find indeed

$$H_i(t_0,t) = \frac{1}{1 - \frac{\log(M)}{\log[\sigma_i(t_0,t)]}}. \tag{10.4}$$

In case that all nodes of $C_i(t_0)$ are equally distributed over the M largest clusters or if no node of $C_i(t_0)$ is anymore found in one of them, i.e., $\rho_{ij}(t_0, t) = 0$ and thus $\sigma_i(t_0, t) = 0$, we have $H_i = 1$ [8].

10.3 Results for the German Pig-Trade Network

This section provides an overview of the characteristics of the considered livestock-trade network in Sect. 10.3.1. Then, we will apply the deterministic SIR model to this particular time-varying network (cf. Sects. 10.2.1 and 10.2.3) and thereby calculate different seed clusters (Sect. 10.3.3). In Sect. 10.3.4, we present different ways to identify sentinel nodes and finally, we will exploit the underlying mechanism to design a detection scheme for possible outbreaks in Sect. 10.3.5.

10.3.1 From Data to Network

In the present study, anonymized data on pig-trade movements are analyzed in collaboration with the Friedrich-Loeffler-Institut. The dataset spans the period from January 1, 2011 to December 31, 2014 and is extracted from the *HI-Tier* database.[2] Within this 4-year period, each German pig holding recorded the number

[2] *Bayerisches Staatsministerium für Ernährung, Landwirtschaft und Forsten* (StMELF). *Herkunftssicherungs- und Informationssystem für Tiere*, available from: www.hi-tier.de

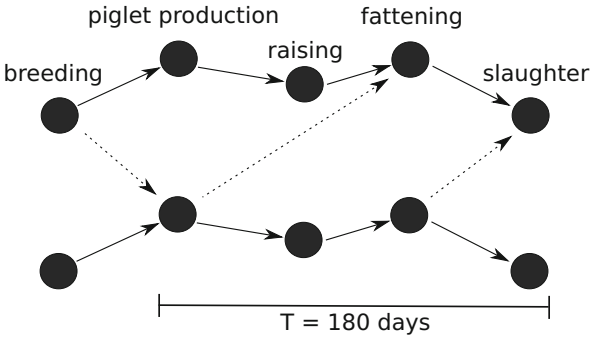

Fig. 10.6 Schematics of the production chain forming the pig trade [19]. The *dashed arrows* refer to deviations from this chain, which are present in the data, because the network contains more edges than the minimal production-chain forest

of pigs of every purchase so that we can infer the corresponding movements of livestock within Germany from the dataset. Note that only the aggregated trading volume (batches) is recorded in the database. The available resolution for this time-dependent network is 1 day. Farmers are required to register each transaction within 7 days, which sets the upper bound for the uncertainty of data accuracy. Every trade record includes the premises of origin and destination via anonymized IDs, the date, and the number of delivered pigs. From a graph-theoretical perspective, the dataset can be interpreted as a dynamical network, where nodes, directed edges, and edge weights correspond to farms, trading events, and the number of traded animals, respectively. For a detailed, time-resolved analysis of this dataset, see Ref. [19].

Figure 10.6 depicts an illustration of the production chain of the underlying farming system, which is composed of different farm types. Different stages of the production chain refer to breeding, piglet production, raising, fattening, and slaughter. In addition, trades can also be mediated by brokers. These are part of the recorded transaction in the database, but do not own a farm themselves. The lifetime of a pig is 180 days, which sets the timescale of the total production chain. Each farm has an anonymized ID from 0 to 97,980. The considered period of 4 years contains more than 6.3 million movements with a total trade volume of 615 million pigs. In the year 2014, 28 million pigs have been bred. This implies that each animal is traded roughly five times along the product chain indicating a high specialization and different farm types. Some basic characteristics of the time-aggregated network are summarized in Table 10.1.

Next, we will present the main results of our numerical simulations.

10.3.2 Outbreak Duration and Size

In our simulations, we consider all nodes as seed and choose the first Monday in each month as starting time t_0 or, if it is a holiday, we use the following working day. In previous studies, these days have been found to show the highest trade activity in the network and are therefore the days for which the largest outbreak size can be

Table 10.1 Standard network properties of the static, i.e., time-aggregated, German pig-trade network

Property	Value
Number of nodes	97,980
Number of edges	315,333
Edge density	3.2×10^{-5}
Size of GSCC	28 %
Diameter	18
Average shortest path length	5.5
Path density	0.24
Median and average trade volume of a premises	
on a day	32.0, 113.4
in a month	88.0, 355.0
in a year	280.0, 2587.6

Diameter and shortest path length are computed for the giant strongly connected component (GSCC)

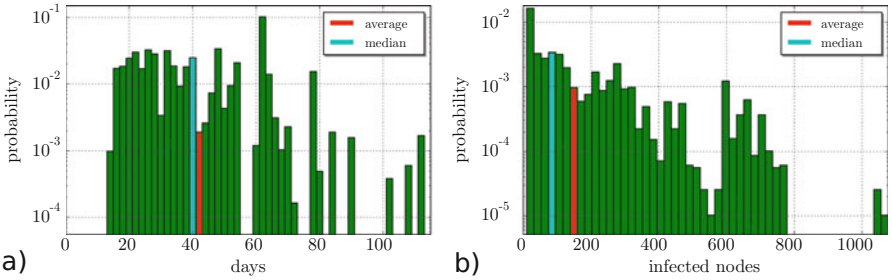

Fig. 10.7 Panel (**a**): Normalized distribution for the outbreak duration. Duration for a fixed starting time averaged over all possible seed nodes (*red*): 41.6 days; median (*cyan*) 40 days. Panel (**b**): Normalized distribution of outbreak size. Average size (*red*): 149 nodes; median (*cyan*): 84 nodes. All nodes with an out-component $|c_{\text{out}}| \geq 10$ are considered as seed. The starting times t_0 are chosen as the first Monday in each month or, if it is a holiday, we use the following working day

expected. In this sense, they cause the most harm to the network [19]. Since we are interested in nodes that can trigger outbreaks of a considerable size, we restrict the pool of potential sentinels to nodes with an out-component $|c_{\text{out}}| \geq 10$.

In Fig. 10.7a, one can see the normalized distribution that an outbreak lasts a certain number of days in the network. Panel (b) shows the normalized distribution of the size of an outbreak. Average and mean values are also indicated by red and cyan bars, respectively. We find that the average outbreak lasts 41.6 days, during which 149 nodes are infected.

Using a deterministic SIR model on a network to explore a worst-case scenario (cf. Sects. 10.2.1 and 10.2.3), we find that all outbreaks eventually come to an end in our simulations. As we show later in Sect. 10.3.3 we observe outbreak durations of around 60 days for the considered infectious period of 7 days. This

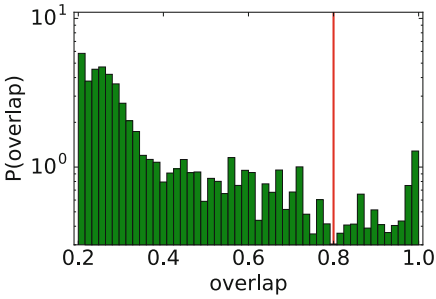

Fig. 10.8 Distribution of overlap of invasion paths calculated based on the Jaccard index. A minimum is found at a value of $\Theta = 0.8$ (*red line*), which we choose as a threshold to define clusters

is much shorter than the 4 years observation time of the network. In other words, we measure the complete out-components. Therefore, we conclude that we capture the entire dynamical process by the proposed modelling framework. See Ref. [13] for a discussion of finite observation periods for a temporal network.

In the next section, we demonstrate how clusters introduced in Sect. 10.2.5 can be constructed from the numerical results.

10.3.3 Seed Clusters

Our aim is to design a surveillance scheme that requires only a small number of nodes. For this purpose, we identify similar spreading patterns and partition the network in functional clusters. In our simulations, we consider every node in the network with $|c_{\text{out}}| \geq 10$ as starting point of an outbreak and consider different starting times as well. Next, however, we discuss the results obtained for the starting time $t_0 =$ January 3, 2011, as an example.

Figure 10.8 shows the distribution of the Jaccard index. As mentioned above, it describes the overlap of different invasion paths. Therefore, a matrix Θ with elements Θ_{ij} will be calculated out of invasion paths i and j. For the cluster calculation, we consider just overlaps greater than the threshold value $\Theta_{\text{th}} = 0.8$, which corresponds to the minimum in the distribution of overlaps (red line). Therefore, all overlaps with a larger Jaccard index are considered in the following. For further information on this subject see Refs. [19, 24]. This choice coincides with the threshold reported in Ref. [8].

Figure 10.9 shows a ranking of cluster sizes for this threshold (red dots) and the cumulative cluster-size distribution (blue triangles). We find that there are many small clusters. More than half of the clusters consist of at most ten seed nodes. The largest cluster is formed by 284 seed nodes. In the following, we consider only the largest 18 clusters. They contain at least 79 seed nodes (red horizontal line) and together cover 31.7% of all seed nodes that can be grouped in clusters (blue horizontal line).

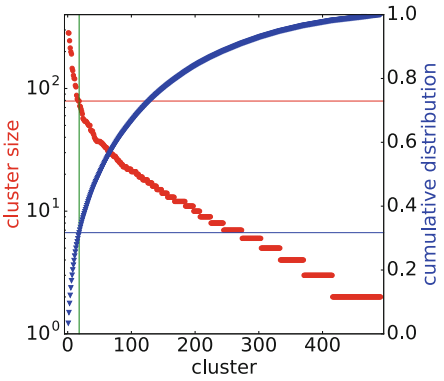

Fig. 10.9 Ranking of cluster size (*red dots*). For the initial time t_0 = January 3, 2011 and the threshold $\Theta_{\text{th}} = 0.8$, we find 491 clusters. The *blue* triangles refer to the cumulative distribution of cluster sizes. The *green* line marks the 18 largest clusters. The *blue* line marks the cumulative cluster distribution of the 18 largest cluster (*green vertical line*). The size of the 18th largest cluster is indicated by the *red line*. There are in total 8490 seed nodes in the observed 491 clusters

Next, we compute the outbreak size triggered from each cluster. It is given by the number of nodes, which can be reached by an infection starting at the seed nodes that form the respective cluster. We call the corresponding percentage *network coverage*. Furthermore, we calculate the power of each cluster to detect an outbreak. This is quantified by the percentage of outbreaks (*detection probability*) that involve any node of the respective cluster. Table 10.2 shows the cluster size, the network coverage (in %), and the detection probability (in %). In general, we find that the numbers fluctuate in both the network coverage and detection probability. For example, there are clusters whose invasion paths appear to be rather isolated in the network, which results in a small detection probability. Other clusters that do not necessarily consist of a large number of seed nodes have a much higher probability to detect an outbreak. For comparison to our findings, consider the results on the 18 largest cluster of the Italian cattle-trade network presented in Ref. [8].

Figure 10.10 depicts numerical results for the 18 largest clusters, which are computed via the Jaccard coefficient of all invasion paths starting at t_0 = January 3, 2011. For each cluster, the time series of the prevalence is shown for every node of the cluster considered as seed (red curves). The black curve refers to the average of all prevalence curves originating from the cluster. The blue curves correspond to the size of the epidemic measured by the number of recovered nodes and the black curve shows again the average.

In general, all time series exhibit a qualitatively similar behavior: an increasing number of infections leading to a peak, beyond which the curve decreases again and the outbreak eventually terminates. These qualitative features are in line with the expected dynamics of the SIR model. All premises within one cluster show a similar spreading pattern, which means that for a given initial condition of seed

Table 10.2 Cluster size, network coverage (in %), and detection probability (in %) of the 18 largest clusters

Cluster	Size	Network coverage	Network coverage of nodes with $\lvert c_{\text{out}} \rvert \geq 10$	Detection probability	Cumulative detection probability
1	284	0.5	3.2	38.2	38.2
2	283	0.9	5.8	3.4	38.7
3	245	1.6	9.9	13.8	43.9
4	214	1.3	8.1	10.8	47.7
5	199	0.7	4.4	10.0	49.6
6	191	0.4	2.8	4.0	50.2
7	146	0.6	3.9	4.5	50.5
8	140	0.5	3.0	25.2	53.7
9	128	0.7	4.7	19.1	56.0
10	121	0.8	4.9	25.4	57.9
11	120	0.4	2.8	23.2	58.4
12	106	0.8	5.0	14.7	59.1
13	103	0.3	1.8	0.9	59.2
14	88	0.2	1.2	43.4	61.2
15	82	0.2	0.9	26.0	65.0
16	81	0.2	1.1	0.5	65.1
17	79	0.2	1.0	0.2	65.1
18	79	0.3	1.7	0.001	65.1

Starting time $t_0 =$ January 3, 2011

and time $(v_i,\ t_0)$ the number of infected premises is roughly the same. We also find that the timing of the peak does not vary much between the different clusters. There are, however, considerable quantitative differences between prevalence curves of different clusters. Consider, for instance, the duration of an outbreak, the peak number of infected nodes (maximum prevalence), or the total number of infected nodes. The mean outbreak duration $\langle \delta_i \rangle$ in the i-th cluster and we obtain that it varies between 30 and 76 days. The average duration of infection for all 18 largest clusters is 55 days.

Recall that each cluster refers to a set of seed nodes. Since the out-component of a cluster is given by the nodes in the network that can be infected from its seeds, the out-component can be larger than the size of the cluster itself, that is, the number of its seed nodes. For some clusters (cf. cluster 3 or 4), even the peak of the prevalence is larger. In order to design an efficient surveillance protocol, we have to make sure that the infection will be detected very early before the outbreak reaches large parts of the potential out-component.

After the construction of clusters of similar invasion paths, we will show in the next section, how this can be used to select a small number of sentinel nodes for surveillance of the whole network.

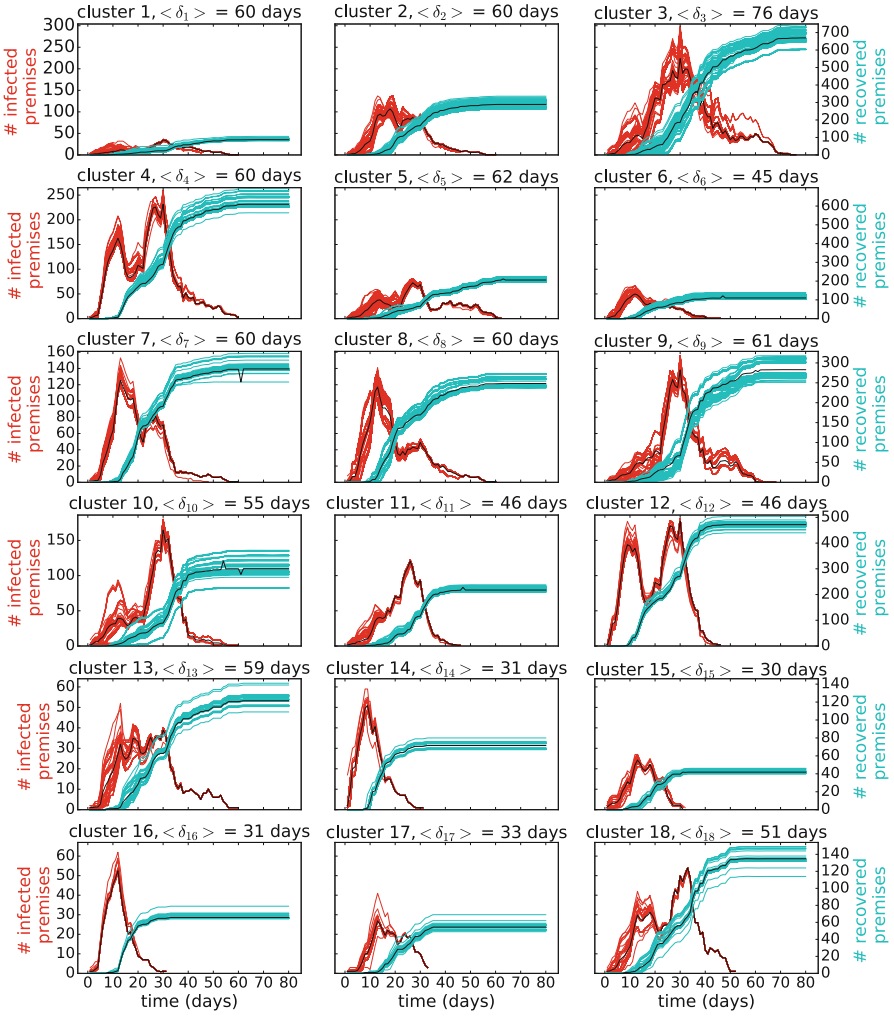

Fig. 10.10 SIR dynamics on the German pig-trade network for the 18 largest clusters. The *red* curves refer to the time series of the number of infected nodes for all nodes in the respective cluster taken as seed. The *blue* curves represent the number of recovered nodes over time. The *black* curves show their average of each cluster. δ_i is the mean duration of outbreaks in the i-th cluster. For the starting time $t_0 =$ January 3, 2011, the mean outbreak mean duration of all nodes in the 18 largest clusters is $\delta = 55$ days. Parameter: infectious period $\tau = 7$ days

10.3.4 Sentinel Nodes

For an identification of potential sentinel nodes, we propose two approaches and evaluate them in terms of detection probability, fast detection, and minimum number of infected nodes until detection. The selection of an optimal, that is, minimum, set is an open question related to set cover problems in combinatorial geometry and has recently been linked to optimal percolation. See Ref. [26] and references therein. This family of problems is known to be NP hard. The methods used here serve as heuristics for the exact problem.

The first protocol consists of the following strategy: Choose the node of largest or second-largest sum of in- and out-degree of each cluster. This results in 18 or 36 sentinel nodes, respectively. We conjecture that these hubs are good candidates for the following reason: Hubs are known to be infected at an early stage of outbreaks on scale-free networks and thus key players for the spreading [25]. The set of sentinel nodes will be most likely part of the GSCC, because they need to receive and send livestock from/to many different nodes to meet the selection criterion. Therefore, they are expected to have a large out-component.

Figure 10.11 shows, how in- and out-degree varies in the two largest clusters. Nodes with the largest sum of in- and out-degree can be found on the *upper*, *right side* in the figures. The candidate nodes to serve as sentinels (*red square* and *diamond*) are well separated from the rest of the seed nodes that form the respective cluster (*green dots*).

As a second approach, we apply the algorithm introduced in Sect. 10.2.3 to infect all nodes of the network at the starting time t_0 and then rank them according to how often each node appears in an invasion paths. This way, we exploit the size $|c_{\text{in}}|$ of the in-component, which is equivalent to the vulnerability of a node. The set of sentinel nodes is given by the top ranked nodes.

Following Ref. [8], we are interested in the nodes that are part of the outcomponent of a large number of nodes. These nodes will be hit by many epidemics

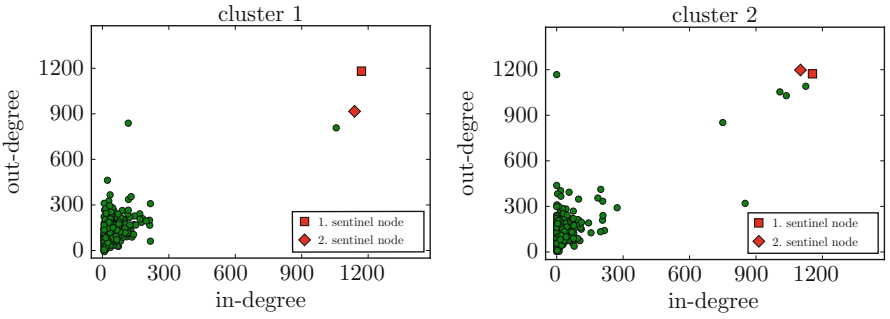

Fig. 10.11 In- and out-degree for all seed nodes of the two largest clusters. We choose sentinel nodes based on the largest sum of in- and out-degree. These nodes can be found in the upper, right part of the panels

starting at different nodes. Therefore, nodes that have a high $|c_{\text{in}}|$ are more vulnerable than nodes with smaller in-component. Some of these nodes, however, are slaughterhouses and are found at the end of the production chain. They are not suitable as sentinel nodes for early disease detection, because the damage of an outbreak would have been done already and could not be contained. These nodes can easily be excluded, because they have an out-degree $k_i^{\text{out}} = 0$. In addition, sentinel nodes should have a significant spreading potential. Therefore, we consider only nodes as sentinels that at the same time have an out-degree of $k_i^{\text{out}} \geq 5$. We choose 18 of these, which we call most infected nodes, and take those 18 most infected nodes together with the 18 nodes of the largest sum of in- and out-degree in each cluster to define the set of sentinel nodes. In an additional protocol, we also consider the 36 nodes with the largest in-component for comparison.

Next, we will investigate, how the different protocols to select sentinel nodes perform in terms of detection probability, detection time and how many nodes become infected until detection.

10.3.5 Disease Detection with Sentinel Nodes and Results

Applying different protocols to select sentinel nodes as introduced in Sect. 10.3.4, we calculate the probability to detect an outbreak for every starting day. See Fig. 10.12, where panel (a) depicts this detection probability based on 18 (blue

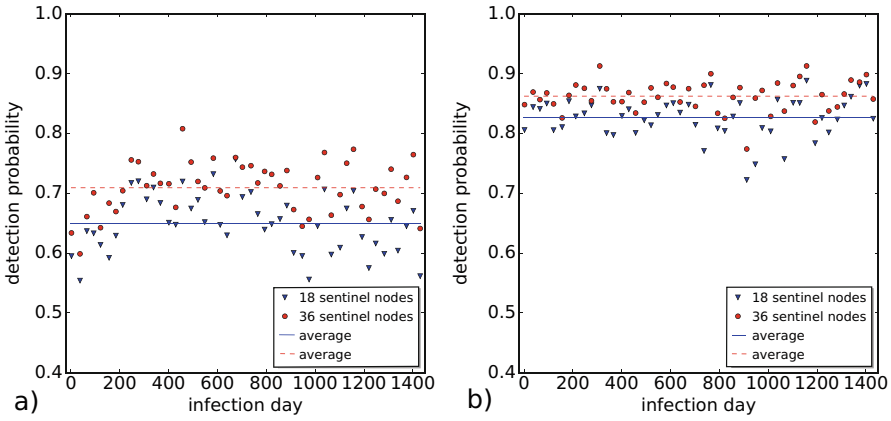

Fig. 10.12 Detection probability of 18 (*blue triangles*) and 36 (*red dots*) sentinel nodes based on (**a**) largest and second-largest sum of in- and out-degree out of each cluster, (**b**) 18 nodes based on highest vulnerability (*blue triangles*) and additionally 18 nodes out of each cluster with the largest sum of in- and out-degree (*red dots*). The mean value is depicted by the solid and dashed lines, respectively: (**a**) 65% and 70.9%; (**b**) 82.7% and 86.2%. Each *dot* refers to the starting time t_0 (day of initial infection), which is chosen as the first Monday in each month or, if it is a holiday, we use the following working day. All nodes with an out-component $|c_{\text{out}}| \geq 10$ are considered as seed

Table 10.3 Detection probability, time until detection, and number of infected nodes until detection for the considered selection protocols to determine the set of sentinel nodes

Protocol	Detection probability	Detection time / days	Outbreak size
18 nodes based on sum of in- and out-degree	65.1%	12.5	43.6
36 nodes based on sum of in- and out-degree	71.0%	10.5	31.2
18 nodes based on highest vulnerability	82.7%	9.0	22.2
36 nodes based on highest vulnerability	83.1%	8.9	21.4
18 nodes based on highest vulnerability and 18 nodes based on in- and out-degree	86.2%	7.8	15.4

triangles) and 36 (red dots) sentinel nodes with the largest sum of in- and out-degree in each cluster, respectively. The blue solid and red dashed lines represent the average probability of a disease detection, which is 65% and 70.9%, respectively. Similarly, Fig. 10.12b depicts the protocol, where 18 sentinel nodes are selected based on the highest vulnerability (blue triangles) or additional 18 nodes with the largest sum of in- and out-degree for each cluster (red dots). This results in average detection probabilities of 82.7% and 86.2%, respectively.

Table 10.3 provides an overview of the obtained results for all proposed selection schemes. Considering twice as many sentinel nodes improves all considered quantities: a higher detection probability, a shorter detection time, and a smaller number of infections until detection. An earlier detection by 2 days results in a reduction of the epidemiological impact by about 25%. This is in agreement with findings of Ref. [8]: The information provided by the sentinel nodes is meaningful as long as the detection occurs rather early during an outbreak. This result is not only important for surveillance, but also for identifying the initial outbreak location, because it enhances the chances to trace the invasion path back to the seed. An even stronger improvement can be obtained, if the selection of sentinels is based on the highest vulnerability. This advantage can be further improved in combination with nodes of largest in- and out-degree. Then, the detection probability is larger than 86% with an average detection time of 7.8 days and an average outbreak size of 15.4 nodes. This gives a larger benefit than choosing 36 nodes with highest vulnerability, for example.

10.3.6 Cluster Development in Time

In this section, we will investigate the temporal stability of the clusters given their importance in the identification of sentinel nodes. Consider a pair of seed nodes, which are a part of the same cluster at one instance in time. They might, however, not belong to the same or any other cluster at a later time. In detail, we consider the development of the 18 largest clusters. Based on two partitions of clusters at different times, that is, $P(t_0) = \{C_1(t_0), C_2(t_0), \cdots, C_{18}(t_0)\}$ and $P(t) = \{C_1(t), C_2(t), \cdots, C_{18}(t)\}$, we calculate the relative overlap via $\rho_{ij} = |C_i(t_0) \cap C_j(t)|/|C_i(t_0)| \in [0, 1]$. We expect $\rho_{ij} = 0$, if the clusters $C_i(t_0)$ and $C_j(t)$ do not have a single node in common, and unity, if clusters persist or expand.

Figure 10.13 shows the matrix $\{\rho_{ij}\}$ for different times. Trivially, we find the identity matrix for $t = t_0$ due to disjoint clusters corresponding to disconnected subgraphs. The clusters evolve and change their nodes over time. For subsequent times t, nodes belonging at t_0 to the same cluster can be redistributed in multiple clusters, which might consist of additional nodes, or might not be a part of any other subsequent cluster. One can see that for times $t = 7$, $t = 14$, and $t = 21$, there is no significant overlap anymore. This can also be seen in the bottom panels, which show a distribution of the overlap between the 18 initial clusters and the 18 subsequent clusters.

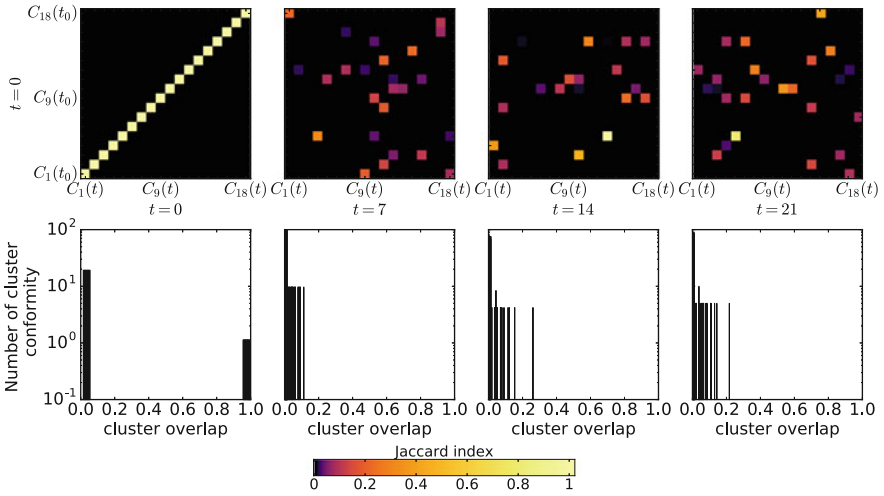

Fig. 10.13 Change of the cluster partitions. The color code refers to the relative overlap of the 18 largest clusters at different times in comparison with $t_0 = 0$ corresponding to January 3, 2011. The top left figure shows the comparison from the cluster $t_0 = 0$ with itself, that is, a trivial perfect overlap along the diagonal. The lower four figures show the distribution of the cluster overlap at respective times

How rapidly and to which extent the node set of the clusters changes can be calculated with the entropy function (cf. Sect. 10.2.6), which will be the topic of the next section.

10.3.7 Entropy of Clusters

In order to quantify the robustness of a cluster, we compute the conditional entropy $H_i(t_0, t)$ of each cluster $C_i(t_0)$ given by Eq. (10.3) comparing different times. This provides insight, how much the nodes of a cluster of time t_0 are redistributed among the M largest clusters at a later time t. Recall that $H_i(t_0, t)$ vanishes, if the set of seed nodes forming a cluster does not change over time. We have $H_i(t_0, t) = 1$, if no node is part of any of the M largest clusters at time t. For comparison, we also calculate the minimum entropy $H_{\min}$, which corresponds to the case that a fraction of nodes of a cluster still form a cluster and the rest does not belong to any of the M largest clusters.

Figure 10.14 depicts the entropy $H(t_0, t)$ (red dots), the minimum entropy $H_{\min}$ (blue circles), and the difference between them (yellow bar) for exemplary clusters 4 and 15. The difference $H(t_0, t) - H_{\min}$ can be interpreted as the robustness of the cluster. A cluster is more robust, if that difference is smaller (and the entropy is not equal to one as in cluster 15), because many nodes from the starting time are still found in one of the 18 largest clusters. Cluster 4, for instance, remains stable over the first 30 weeks.

In cluster 15 we can see that $H = 1$ at 11 different times due to the peculiarities of the cluster development. Cluster 15 has such a high entropy for many weeks, because its nodes do not belong to any of the 18 largest clusters at these times. In

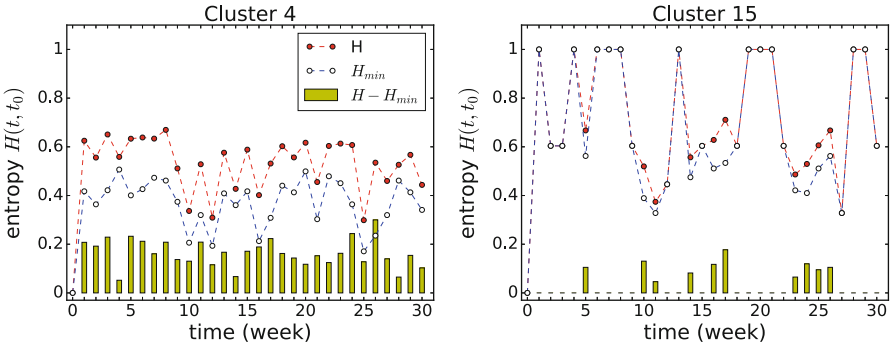

Fig. 10.14 Entropy $H(t_0, t)$ of cluster 4 and 15 over time (*red dots*), minimum entropy (*blue empty dots*), and their difference (*yellow bars*)

contrast to the fluctuating entropy of cluster 15, cluster 4 is quite stable over the first 30 weeks. The time-resolved entropy of the 16 largest clusters is added in the appendix as Figs. 10.15 and 10.16 for comparison.

10.4 Conclusion and Outlook

We have applied the concept of sentinel nodes proposed in Ref. [8] to the German pig-trade network. For this purpose, we have implemented a deterministic susceptible-infected-recovered model and computed invasion paths for different seed nodes and starting times. Our results have shown that the approach of seed clusters, which was initially applied to the Italian cattle-trade network, can indeed be transferred to the considered dataset. The clustering method can be used to design an optimized surveillance system and allows for rapid and efficient containment strategies.

Large delays between the start of the outbreak and its detection results in larger outbreak sizes. After a few days, the outbreak often reaches a number of nodes far greater than the size of the cluster (number of seed nodes identified to yield a similar outbreak pattern), where it started. Then, the disease is able to infect large fractions of the network. In addition, high temporal variability and the complex nature of the network make identification of the possible origin of the outbreak a particularly difficult task. Recently, some approaches using the concept of effective distance have been proposed [27, 28].

Following a network-based analysis, we have identified farms that are at a high risk of becoming infected and subsequently promote the spreading the disease further. We have conjectured that these farms are good candidates to detect an outbreak early in its evolution. Therefore, we have chosen one or two nodes with the largest sum of in- and out-degree for each cluster. In addition, we have also considered farms that have the largest in-component in the network. These nodes are very vulnerable, because they can be infected from a large number of outbreak origins. We have found out that these farms, when considered as sentinel nodes, have the highest detection probability and the shortest detection time. As a consequence, the outbreak size before detection can be considerably reduced. This can be further improved by combining both selection protocols.

Acknowledgements This work was supported by *Deutscher Akademischer Austauschdienst (DAAD)* within the PPP-PROCOPE scheme. FS, AK, and PH acknowledge funding by *Deutsche Forschungs- gemeinschaft* in the framework of Collaborative Research Center 910. The work is partially funded by the EC-ANIHWA Contract No. ANR-13-ANWA-0007-03 (LIVEepi) to VC.

A.1 Appendix

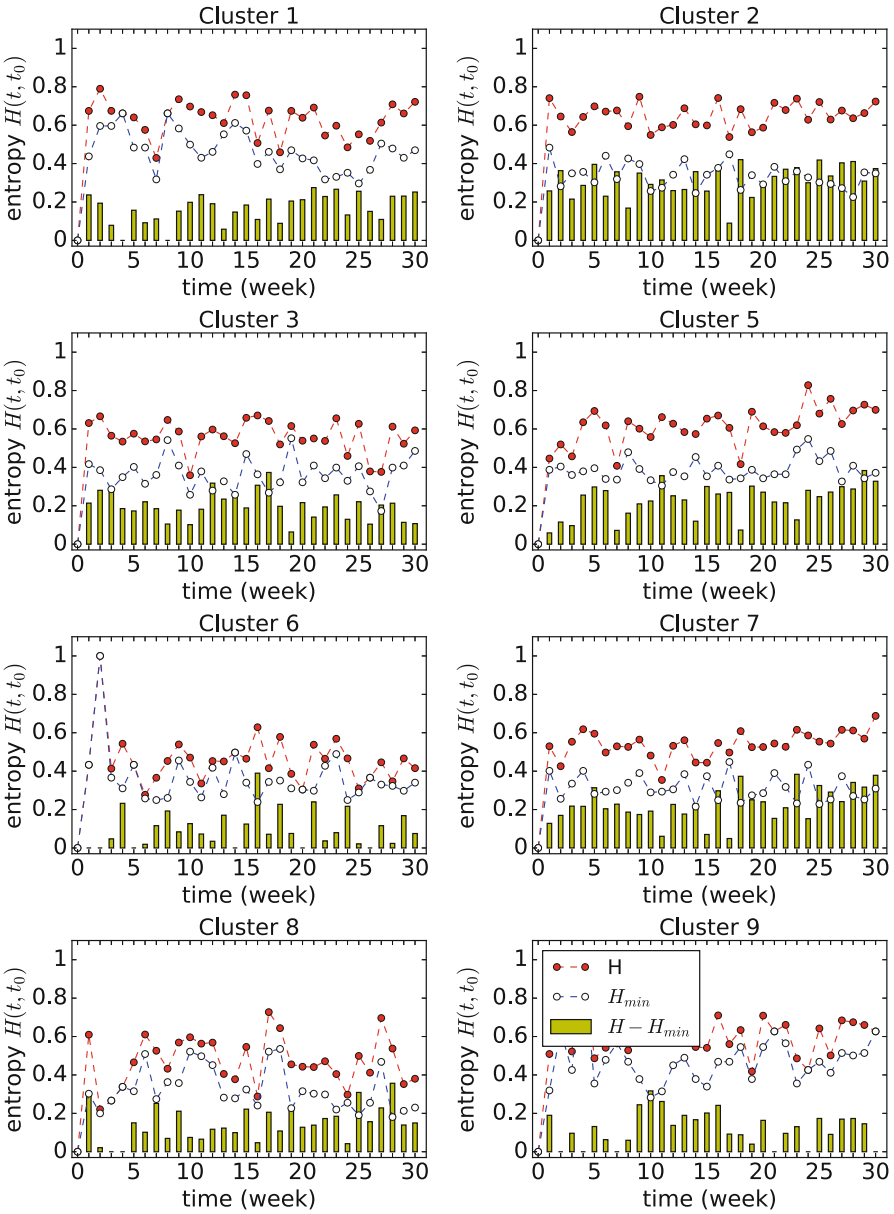

Fig. 10.15 Entropy $H(t_0, t)$ of the eight largest clusters not mentioned in the main text (for cluster 4 see Fig. 10.14) over time (*red dots*), minimum entropy (*blue empty dots*), and their difference (*yellow bars*)

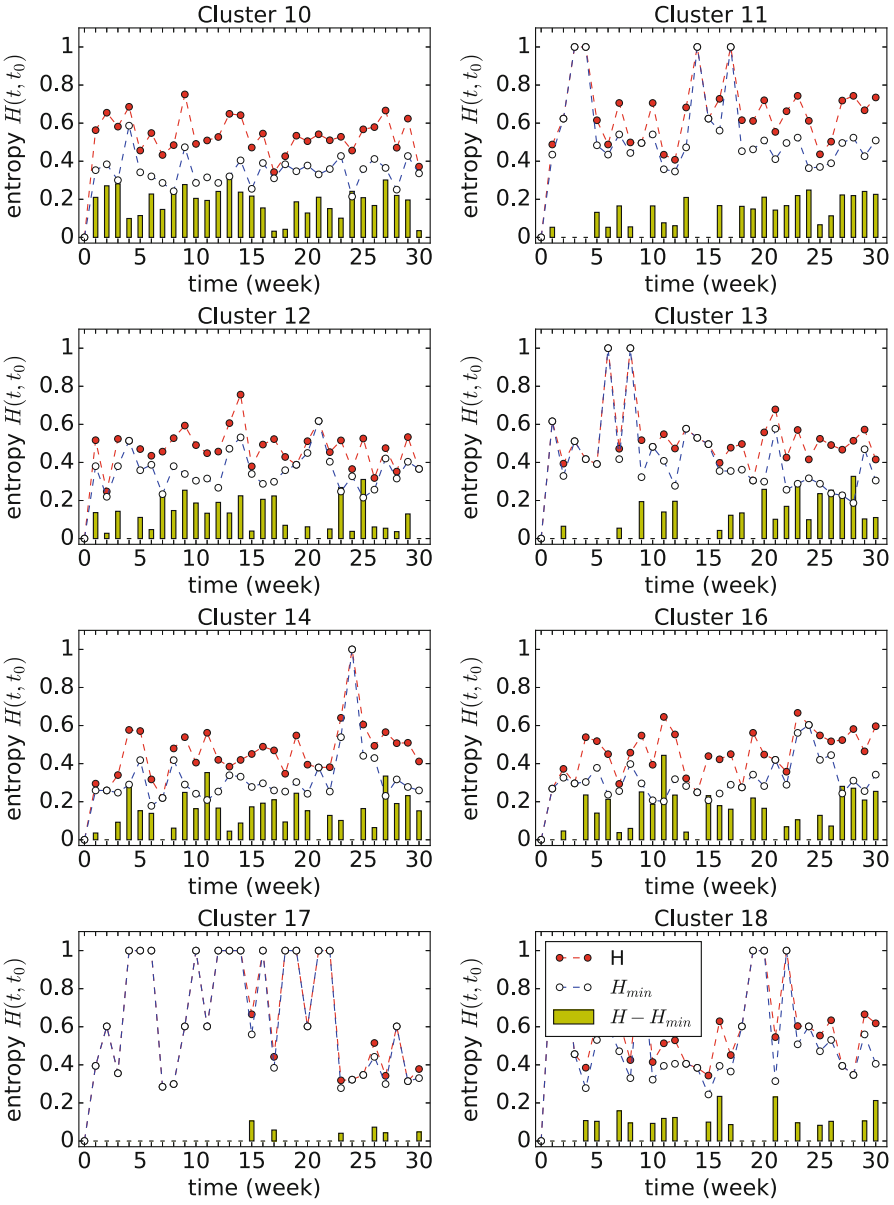

Fig. 10.16 Entropy $H(t_0, t)$ of the clusters 9–18 except for cluster 15, which is shown in Fig. 10.14, over time (*red dots*), minimum entropy (*blue empty dots*), and their difference (*yellow bars*)

References

1. Keeling, M.J., Rohani, P.: Modeling Infectious Diseases in Humans and Animals. Princeton University Press, Princeton (2008)
2. Funk, S., Gilad, E., Watkins, C., Jansen, V.A.: Proc. Natl. Acad. Sci. **106**, 6872 (2009)
3. Anderson, R.H., May, R.M.: Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford/New York (1992)
4. Murray, J.D.: Mathematical Biology: I. An Introduction Interdisciplinary Applied Mathematics. Springer, New York (2002)
5. Diekmann, O., Heesterbeek, H., Britton, T.: Mathematical Tools for Understanding Infectious Disease Dynamics. Princeton University Press, Princeton (2013)
6. Fritzemeier, J., Teuffert, J., Greiser-Wilke, I., Staubach, C., Schlüter, H., Moennig, V.: Vet. Microbiol. **77**, 29 (2000)
7. Hethcote, H.W.: SIAM Rev. **42**, 599 (2000)
8. Bajardi, P., Barrat, A., Savini, L., Colizza, V.: J. Roy. Soc. Interface. **9**, 2814 (2012)
9. Koher, A., Lentz, H.H.K., Hövel, P., Sokolov, I.: PLoS One. **11**, e0151209 (2016)
10. Newman, M.E.J.: Phys. Rev. E. **66**, 016128 (2002)
11. Konschake, M., Lentz, H.H.K., Conraths, F., Hövel, P., Selhorst, T.: PLoS One. **8**, e55223 (2013)
12. Vernon, M.C., Keeling, M.J.: Proc. R. Soc. Lond. B. Biol. Sci. **276**, 469 (2009)
13. Holme, P., Saramäki, J.: Phys. Rep. **519**, 97 (2012)
14. Casteigts, A., Flocchini, P., Quattrociocchi, W., Santoro, N.: Int. J. Parallel Emergent Distrib. Syst. **27**, 387 (2012)
15. Holme, P.: EPJ B. **88**, 1 (2015)
16. Bajardi, P., Barrat, A., Natale, F., Savini, L., Colizza, V.: PLoS One. **6**, e19869 (2011)
17. Rocha, L.E., Liljeros, F., Holme, P.: PLoS Comput. Biol. **7**, e1001109 (2011)
18. Valdano, E., Ferreri, L., Poletto, C., Colizza, V.: Phys. Rev. X. **5**, 021005 (2015)
19. Lentz, H.H.K., Koher, A., Hövel, P., Gethmann, J., Sauter-Louis, C., Selhorst, T., Conraths, F.: PLoS One. **11**, e0155196 (2016)
20. Wu, H., Cheng, J., Huang, S., Ke, Y., Lu, Y., Xu, Y.: Proc. VLDB Endowment. **7**, 721 (2014)
21. Newman, M.E.J.: Networks: An Introduction. Oxford University Press, Inc., New York (2010)
22. Barabasi, A.L.: Network Science. Cambridge University Press, Cambridge (2016)
23. Lü, L., Chen, D., Ren, X.-L., Zhang, Q.-M., Zhang, Y.-C., Zhou, T.: Phys. Rep. **650**, 1 (2016)
24. Dorogovtsev, S.N., Mendes, J.F.F., Samukhin, A.N.: Phys. Rev. E. **64**, 025101 (2001)
25. Pastor-Satorras, R., Vespignani, A.: Phys. Rev. Lett. **86**, 3200 (2001)
26. Morone, F., Makse, H.A.: Nature. **524**, 65 (2015)
27. Brockmann, D., Helbing, D.: Science. **342**, 1337–1342 (2013)
28. Iannelli, F., Koher, A., Brockmann, D., Hövel, P., Sokolov, I.M.: Phys. Rev. E. 95, 012313 (2017)