International Food and Agribusiness Management Review
Special Issue - Volume 19 Issue A, 2016

A Network-Science Support System for Food Chain Safety: A Case from Hungarian Cattle Production

Ákos Jóźwiak [a], Mátyás Milkovics[b], and Zoltan Lakner[ⓣc]

[a] *DVM, Vice-Director National Food Chain Safety Office, System Management and Supervision Directorate, Keleti Károly u. 24, Budapest, Hungary*

[b] *MSC, National Food Chain Safety Office, System Management and Supervision Directorate, Keleti Károly u. 24, Budapest, Hungary*

[c] *Professor, Szent István University, Faculty of Food Science, Villányi út 35-43, Budapest, Hungary*

Abstract

In a risk analysis framework, food chain safety measures should be objective and scientifically based. Network science – as a decision support tool – may have an important role in bringing safety to the food supply.

The aim of the present work is to develop a network-based assessment methodology for Hungarian cattle holdings. The criteria of which is (1) suitable for risk-based planning in order to put resources into the most critical elements of the cattle production network; (2) should be capable of simulating different epidemiological situations in order to increase preparedness for real epidemics.

Keywords: big data, centrality measures, network analysis, veterinary inspection, risk-based planning

ⓣCorresponding author: Tel: + 36305251896
Email: Z. Lakner: zoltan.lakner@etk.szie.hu
A. Jozwiak: jozwiaka@nebih.gov.hu
M. M. Milkovics: milkovicsm@nebih.gov.hu

Introduction

Cattle breeding and trade have been an important part of economic life in the Carpathian basin since the late Iron Age (Bökönyi 1971). Hungarian landholders were important suppliers of animals to markets in Italy and Poland in the middle ages (Sugar et al. 1994). During the Communist regime, which lasted from after the Second World War to 1989, the overwhelming majority of cattle production was concentrated in large-scale production cooperatives and state farms (Csizmadia 1974). While these state and cooperative farms decreased efficiency and productivity considerably, they made veterinary inspection of bovine herds relatively simple. After the system change in 1989, as a result of the agricultural transition and privatization, the number of bovine herds increased but the professional quality of management remained relatively unchanged (Csáki 1990).

Food chain safety "from farm to fork"—together with its elements such as animal health or food safety—are the focus of both the agri-food industry and the control authorities. With the increasing volume and complexity of international trade, traceability issues have become more important than ever. Additionally, bovine-related veterinary problems (in particular BSE, foot and mouth disease) have increased the importance of veterinary management and inspection of herds worldwide (Nikiforuk 2008). The cattle passport system of the EU, as well as national animal movement detection systems (Dubé et al. 2009), offer the possibility of tracing animal movements.

It is well documented that the herd epidemiology is considerably influenced by the mobility of animals (Kao 2002; Kao et al. 2007). The arrival of new, infected animals on non-infected herds increase the probability of disease transmission. The EU animal health strategy highlights the importance of individual animal identification, supporting legal and financial issues necessary for data collection on animal transportation.

The increasing integration of the Hungarian agri-food system into the EU—as a consequence of both trade liberalization and EU membership—have made the situation even more difficult (Bojnec and Fertő 2009; Coulombier and Takkinen 2013). The food chain safety authority is faced with a mission which is practically impossible to implement using traditional methods: i.e. increasing the effectiveness and reliability of food chain control (veterinary inspection of herds in particular) while at the same time given declining resources (Luning et al. 2015).

The Risk Analysis Framework

It is essential to maintain the health of plants, animals and humans to ensure the chemical and microbiological safety of our food ('food chain safety'), while maintaining the sustainability of agri-food production and trade ('food security'). Food chain industry stakeholders—who have primary responsibility for ensuring safety—need to apply a hazard analysis framework to ensure a process-based, preventative, effective operation. Food chain control authorities, when making decisions on control and intervention issues, must use the risk analysis framework as defined by FAO/WHO (2007).

Risk analysis is used to develop an estimate of the risks to human health and safety (risk assessment); identify and implement appropriate measures to control the risks (risk management); and communicate with stakeholders about the risks and measures applied (risk communication). Risk is defined in this context as a product of the severity of the hazard and the probability of its occurrence. Based on FAO/WHO guidelines, EU member states have to apply a risk-based approach managing food chain safety risks: Article 3 of Regulation 882/2004/EC (European Parliament and the Council 2004) states that 'Member States shall ensure that official controls are carried out regularly, on a risk basis, and with appropriate frequency'. This means that risk managers should focus their resources on high risk entities: business operators, foodstuffs or particular hazards. During the risk-based planning of official controls, competent authorities have to take into account all of the objective evidence contributing to better decision making in the risk analysis framework. In this context, as a part of planning the most effective risk management options, different risk assessment and risk ranking methods are available, along with different planning techniques. Authorities have to choose whichever methods best fit their needs and resources. Continuous improvement and new methodologies are in the forefront of research.

Big Data and Network Science in the Field of Food Chain Safety

The need for handling, analysis and interpretation of large, interrelated datasets in various scientific fields, together with the rapid development of information-technology tools, have resulted in newly emerging data-related scientific fields. Their common characteristic is that with the use of computational science tools such rules or patterns could be identified which would otherwise be very hard or impossible using smaller datasets (Baranyi et al. 2013). Globalization, particularly its sociological and commercial aspects, started research of complex networks in the late 90's (Anderson and Marcouiller 2002). It quickly became evident that the structure and evolution of the networks showed many similarities regardless of what they represent (Baranyi et al. 2013). This phenomenon boosted research in different scientific fields which, after a short initial phase, network analysis methods found an application in many areas. It is used in sociology for the representation of the individuals and their relationships (Stanley and Katherine 1994; Salathé and Jones 2010), for mapping genes, proteins and their interactions with each other in molecular biology (Barabási and Albert 1999), and helps in the identification of business relationship of companies in different economical analysis (David and Douglas 1992).

As a definition (Börner et al. 2007) network science concerns itself with the study of different networks, be they social, biological, technological or scholarly networks. Its goal is to contrast, compare and integrate techniques and algorithms developed for a wide range of disciplines, primarily mathematics and statistics. Barabási (1999) compares the emergence of this science with sweeping developments in quantum mechanics in the 20th century. In his opinion, network sciences are building a theoretical and algorithmic framework which is energizing many research fields. "Born at the twilight of the twentieth century, network theory aims to understand the origins and characteristics of networks which hold together the components of various complex systems."

Data science, particularly network science, has an important role in food science enhancing security and safety of the food supply as well. Analysis tools based on network theory can be used in the risk-based control and monitoring systems of food business operators by analyzing their commercial relations with each other (Chmiel et al. 2007).

Borgatti et al. (2009) highlight that, over the past decades, network theory has supplied a valuable tool in explaining different social phenomena. In management science it has been widely used in supply chain management (Lazzarini et al. 2001), international trade analysis (Smith and White 1992), organizational development (Wasserman and Faust 1994) and policy analysis (Wagner and Leydesdorff 2005).

In the opinion of Fritz and Schiefer (2008), management challenges in closely co-operating enterprises, as well as the mutual dependence of all participants in the food chain, necessitate the application of network science in this area. The application of network science at the inter-firm level in agribusiness management is highlighted by Ng and Siebert (2009). In the most recent literature there are numerous examples of the successful application of network-theory approaches in the development of agribusiness systems, from the development of agricultural extension programs (Lehmann et al. 2012) to supply chain management (Farhat 2012).

Besides management science, network analysis is claimed to be an effective tool in food chain safety analysis as well. The first applications of network science in the field of food chain safety were aimed at mapping connections between countries or businesses. Petróczi et al. (2010, 2011) analyzed the notification data of the Rapid Alert System for Food and Feed (RASFF) of the European Commission. They identified European trade and notification patterns using network science methodology, using the model to forecast as well. Not only countries but different businesses were later analyzed for epidemiological purposes: Lentz et al. (2011) explored pig transport routes in Germany, showing hubs where cross infection was more likely. Ercsey-Ravasz et al. (2012) identified the most critical agri-food trade routes based on publicly available trade data. They drew attention to the fact that every second food batch produced is exported—and this proportion is increasing—providing proof of continuously growing international trade and an increasing need for the application of complex sciences.

The application of network theory for the analysis of animal migration has some decades-long tradition (Rommel et al. 1973; Harris 1979), but the conscious application of animal transportation data for the prevention of epidemiological problems is relatively new. This process is boosted by the rapid development of cattle identification systems. The comprehensive review of the New Zealand Ministry of Agriculture and Forestry offers a general overview on selected cattle identification and tracking systems worldwide (MAF 2009). The review proves, in which (1) all of the reviewed systems are implementing individual cattle identification requirements; (2) there is an increasing tendency to apply RFID technology; (3) most of them are mandatory; and (4) these systems are administered by governments or under industry-government partnerships. In the opinion of Schroeder and Tonsor (2012), cattle identification and traceability is becoming a necessary pre-condition for the international competitiveness of cattle and the cattle-product export market. In the last few years Dubé et al. (2008) has applied the network analysis approach to analyze and prevent foot and mouth disease. Martinez-López et al. (2009) have analyzed the trans-boundary flow of animals with the purpose of implementing disease prevention measures. Bajardi et al. (2011, 2012) mapped the Italian cattle trade network and made great progress in analyzing dynamic patterns, using network science tools to optimize cattle farm surveillance.

Motivation

When analyzing risk, food chain safety measures should be based on objective and scientifically based evidence. In most cases authorities are already using existing international risk assessments, risk ranking, risk-based priority setting tools, models, studies, and literature data. However, the data needed for substantiated risk assessment are in many cases not available. The lack of data or possible delays in providing updated records may hinder their use, especially for time-varying patterns (Valdano et al. 2015).

Our experience obtained in Hungarian and European food chain safety control planning systems show that, conventionally, risk is determined by the size of the herd as the frequency of official controls is determined by the number of animals present at any given time (taking into consideration other risk factors such as production type (dairy, meat, etc.) and the results of previous inspections). The calculations are based on the conventional risk approach as a product of severity of the hazard and the probability of occurrence. However, the effectiveness of this targeting mechanism can be questioned in many cases (Van Asselt et al. 2012).

One of the most important problems of this approach is that it doesn't take into account the network flow and the dynamics of the network; just the pure output or production data. The flow of animals denotes the animals transported from one node of the network to another during a given time period. A dynamic network is defined as a network where one or more of its relevant parameters (e.g. size of nodes, flow, etc.) changes as a function of time (Friesz et al. 1993).

The other drawback of the traditional risk based planning procedure is that it serves to set control priorities but is not suitable for epidemiological simulation exercises as the picture it captures is very static. Furthermore, the risk-based planning procedures of different member states are not cross-compatible, making international assessment very difficult or even impossible, and resulting in high coordination costs and significant delays when managing cross-border food chain incidents.

The cattle network consists of numerous closely cooperating holdings under the influence of natural (biological) and socio-economic factors, forming a network where the hubs of the network are the economic entities (e.g. farms, slaughterhouses, etc.) and the edges are the cattle-movements. The size of the hubs and edges can be considered as stochastic variables because the economic activities of the different entities show a considerable fluctuation. To minimize the risk of problems we have to understand the immanent structure of the network on the basis of network science. This will serve to fine-tune the strategy of decreasing risks of an epidemiological nature.

Network analysis is capable of capturing the time-dependent characteristics of the trade flow as well as selecting the highest risk nodes by their network characteristics. Furthermore, it is able to serve as a basis for epidemiological simulation exercises. Our motivation is to find a risk ranking tool which is able to capture those aspects of a functioning trade network.

Objectives

The aim of the present work is to develop a network-based assessment methodology, which is (1) suitable for the risk based planning of official controls (setting priorities based on network science) in order to place resources on the most critical elements of the cattle production network; (2) capable of simulating different epidemiological situations to increase preparedness for real epidemics using network-based spreading models. A majority of these models are based on system dynamics (Bagni et al. 2002) and in the last years agent-based simulation approach (Dion 2011), although there is a rapid development of Bayesian geo-statistical methods as well (Jewell et al. 2013; Ward et al. 2013). However, a critical point of all of these models is the quality of input data. Our results will shed light on how to prepare and interpret data for analysis. Finally, (3) to share the analysis methodology and algorithms with the network science and food chain safety community to enhance cross-compatibility of methods, making it possible to expand simulation exercises and risk-based planning processes across borders (since real world risks don't respect borders).

All of those objectives contribute to a methodology of letting decision-makers elaborate an optimized strategy for the inspection of cattle herds, control of cattle-traffic, and strategies for epidemiological crisis situations.

The hypotheses of the present work are (1) the Hungarian cattle network can be characterized as a scale free network; (2) the vulnerability of the network can be analyzed on the basis of the 'centrality' characteristics of different hubs of the network. The most vulnerable parts of the network are not necessarily the largest hubs, rather the ones which can be considered central parts of the network; (3) one centrality measure is not necessarily enough to characterize the centrality position of a given vertex, because the different centrality indicators reflect differently from every other aspect of the vertices (Friedkin 1991; Marsden 2002); (4) the Hungarian cattle network is a dynamic one. This means that the network size, the flow intensity of animals and other network properties, including centrality, can be characterized as considerably time-dependent.

We have applied wide ranging network analysis tools to determine the characteristic features of different nodes of the network and their time-variance. In this way we were able to characterize the Hungarian cattle trade network, improving the current control strategy and preparing for a possible crisis situation. This application of network science can be considered a relatively novel one as this paper shows a practical application of network theory by a food chain safety authority. This example of the application of a network science approach, based on big data analysis, can be considered a possible solution to a heretofore intractable big data-related problem in the food chain safety field.

Our aim was to shed light on the characteristic features of a given trade network (using the Hungarian cattle trade network as an example) for the purposes of increasing the effectiveness of food chain safety control and preparation for a possible outbreak. Lists of the most risky holdings obtained through the network analysis are used by risk managers while planning their annual control plans. Our model contributes to greater Hungarian food chain preparedness in a critical situation as it demonstrates a methodology which suitably determines the most critical parts of the network. It is not possible to offer a more concise and intelligible solution, because—as we will demonstrate – the actual features of the network are time-dependent variables.

Methods

Data Source

The cattle trade network is obtained using the database of the national cattle identification system (ENAR). This system is able to follow the animals along their whole life cycle from birth to slaughterhouse or from entering the territory of Hungary to their export. It has a legal background based on Regulation 1760/2000/EC on animal identification (European Parliament and the Council 2000), which makes the use of the system obligatory. In this way a continuous dataflow is generated, supplying more than 1000 lines of raw data each day. Each line represents an animal movement between two nodes. Each movement record reports the unique identifier of the animal, the codes of the holdings of origin and destination and the date of the movement.

To understand the structure and the characteristic features of this data flow an application of specific methods and approaches is needed. We have applied network analysis which has served as a tool for managing large amount of data.

Network analysis, as an interdisciplinary field of science, considers the relationship between organizations as a graph (Albert and Barabási 2002; Barabási 2002). The graph consists of a set of vertices and a set of edges (Tichy et al. 1979). In this case the vertices (or nodes) are (1) cattle-exporters to Hungary; (2) importers buying living cattle from Hungary and (3) various economic organizations—so called holdings—including farms, slaughterhouses, logistics/distribution centers, markets, artificial insemination stations, incinerators, fairs and animal health institutions. Movements are the transportation events of living cattle between different nodes. These are represented in our model as edges between nodes. These edges are called flows in graph theory when analyzing transportation processes (Wen and Arcak 2004).

The data inclusion criteria were: (1) time period between 01.01.2012.–31.12.2014.; (2) operating holdings with legal succession as well; (3) no limitation on age or birth of the animals (i.e. they didn't have to be born before 01.01.2012); (4) animals can die during the time period investigated; (5) all animals from the database.

The animal movements taken into consideration involved about 50,000 premises. The reason for the three-year time frame was to adequately characterize the Hungarian cattle-network for risk-based planning purposes yet not so much as to become outdated. The abovementioned trade routes represent approximately half a million movements a year. In the network, nodes may be active or inactive depending on whether farms sell or buy cattle in any given time frame.

The original raw data consisted of 4,667,479 lines, having 42,928,175 pieces of data altogether on animals and 713,482 on holdings. This static raw data was then cleaned and transformed through several steps into a static source-target matrix, containing data on 1,553,683 movements and 52,618 nodes. This static network was broken down into annual and monthly representations to analyze the behavior of the network over time. The basic network parameters were calculated each month resulting in a dynamic network containing 54,933,192 pieces of data attributed to nodes and 1,638,000 to edges.

This data-set can be described as a large-volume, complex, growing dataset concerning multiple, relatively autonomous parts. That's why it can be considered "big data" as defined by Wu et al. (2014), Power (2014), and Sonka (2014). The dataset satisfies the definition of the NIST group (2015) because it 'exceeds the capacity or capability of current or conventional methods or systems'. In the opinion of Ward and Barker, big data is not a set of data but 'a term, describing the storage and analysis of large and complex data steps using a series of techniques' (Ward and Barker 2013). The process of extracting insights from big data consists of five steps: (a) acquisition and recording, (b) extraction, cleaning and annotation, (c) integration, aggregation and representation, (d) modelling and analysis, (e) interpretation. Our current work contains all of these elements and lays down the basis for further modelling work.

Network Analysis Methods

The network of cattle holdings and movements were first analyzed to investigate the structure of the network and to calculate the main parameters. For each node the following measures were calculated:

The node ***degree*** is the number of relations (edges) of the nodes. However, in the case of the directed networks, the ***in-degree*** (incoming connections) and the ***out-degree*** (outgoing connections) values are important as well. Degree has generally been extended to the sum of weights when analyzing weighted networks and labelled node strength, so the ***weighted degree*** and the ***weighted in-*** and ***out-degree*** was calculated (Barrat et al. 2004; Newman 2001; Opsahl et al. 2010). These parameters offer an important piece of information on the intensity of relations between nodes and their environment. A high in-degree indicates that the node can be characterized as prominent; it receives animals from numerous farms. A high out-degree indicates that the node is influential because it has extensive connections with other farms. The same applies for weighted degrees but here the indicator shows not the number of connecting businesses, but the number of animals transported in and out.

In certain networks the nodes with the most important roles are the high degree nodes. However, this is a quite simplistic approach and, if the network has a strongly inhomogeneous structure (containing many clusters), it is certainly false. Low degree nodes connecting clusters in many cases play an important role in the network (Kleingberg 1999).

To understand the relative importance of different hubs in cattle flow besides the usual network metrics (e.g. in- and out-degrees, weighted degrees, etc.), we had to apply the centrality concepts of network analysis. Despite considerable research efforts invested into studying the centrality concept in network science, centrality is still an elusive concept which may be approximated from different perspectives where different centrality measures are available (Abbasi et al. 2012). We have analyzed betweenness centrality (Kim et al. 2012); closeness centrality (Freeman 1979); the two so-called prestige measures of centrality (Faust and Wasserman 1992): the hub centrality and the authority centrality (Kleinberg 1999, 2000), calculated using the HITS algorithm. There is a considerable difference between those centralities: In the case of the authority and hub centrality, a central node can be any node in the network, while in the case of betweenness and closeness centralities (as the names indicate) the central nodes cannot be the source-vertex or sink-vertex (Okoth and Wagner 2009). As defined by Newman (2005), a source vertex is a node with an in-degree zero while a sink vertex is a node without-degree zero.

The HITS algorithm was developed by Kleinberg (1999). This algorithm is a link analysis algorithm which helps in identifying the essential nodes in a graph. It consists of two scores, a ***hub score*** and an ***authority score***. The authority score of a node is a measure of the amount of valuable information that this node holds. The hub score of a node shows how many highly informative nodes or authoritative nodes this node points to. So a node with a high hub score shows that this node is pointing to many other authoritative nodes. On the other hand, a node with a high authoritative score shows that it is pointing to a large number of nodes, and as such, serves as a node of useful information in the network.

Betweenness centrality is an even more important statistical property of a network. This property is applied to a lot of real-world problems such as finding influential people in a social network, finding crucial hubs in a computer network, finding border crossing points which have the largest traffic or trade flow. The betweenness centrality of a node is an indicator of its centrality or importance in the network. It is described as the number of shortest paths from all the vertices to all the other vertices in the network that pass through the node in consideration (Brandes 2001).

Closeness centrality indicates how long it will take for information from a given node to reach other nodes in the network. The smaller the value, the more central role the node plays in the network.

We used Gephi open-source software for network visualization and analysis, making possible to use more than thirty algorithms and models. There are more than 100 plugins to the software, increasing the number of statistical tools (Devangana 2015). However, it was not necessary to apply these additional tools in our research. Further statistical analysis was made using Microsoft Excel software.

Results

Based upon network analysis it was possible to determine the most important (highest risk) flows in the system and construct different models for the cattle-network. On the basis of these models we have been able to determine the most important centers of the network which is extremely important because it is well-documented that the most vulnerable points of a network are not necessarily the largest hubs (Agarval et al. 2014; Wang et al. 2006; Wang et al. 2014).

The data-stream offered a possibility to determine the stability of different centralities of the system, as well as to analyze the stochastic relationships between these centrality indicators. As earlier stated, there is a considerable difference of importance between nodes according to their position in the network. The various statistical algorithms of the software tool provided a characterization of the cattle movement system, exploring both its structural and dynamical properties. There was an opportunity to compare these calculated values for each month and with this the central farms, logistics centers, slaughterhouses, and the peripheral holdings could be unveiled as well.

Network Structure

Mapping the Hungarian cattle holdings network, it was possible to calculate the basic metrics of the network. In network-related literature there is a wide range of indicators used to characterize a given network. Some of them aim to determine the position, sets and clusters of nodes and their connections. Another group of indicators describe the centrality of different nodes or offers information on network density. Other measures help to characterize the components, cores and cliques in the network. All of these pieces of information could furnish valuable insight into the network analyzed but we had to limit ourselves to simple characteristic features of the network.

A key property of each node (in this case, holdings) is its *degree*, representing the number of links it has to other nodes. In the cattle network it means the number of business partners. The *degree distribution*, p_k, provides the probability that a randomly selected node in the network has degree k. For a network with N nodes the degree distribution is given by the equation:

$$1)\qquad p_k = \frac{N_k}{N}$$

where N_k is the number of degree-k nodes. The degree distribution has a very important role in network theory following the discovery of scale-free networks (Barabási and Albert 1999). The degree distribution of the Hungarian cattle holdings network (Figure 1**)** shows a very characteristic heavy-tailed distribution specific to *scale-free* networks.

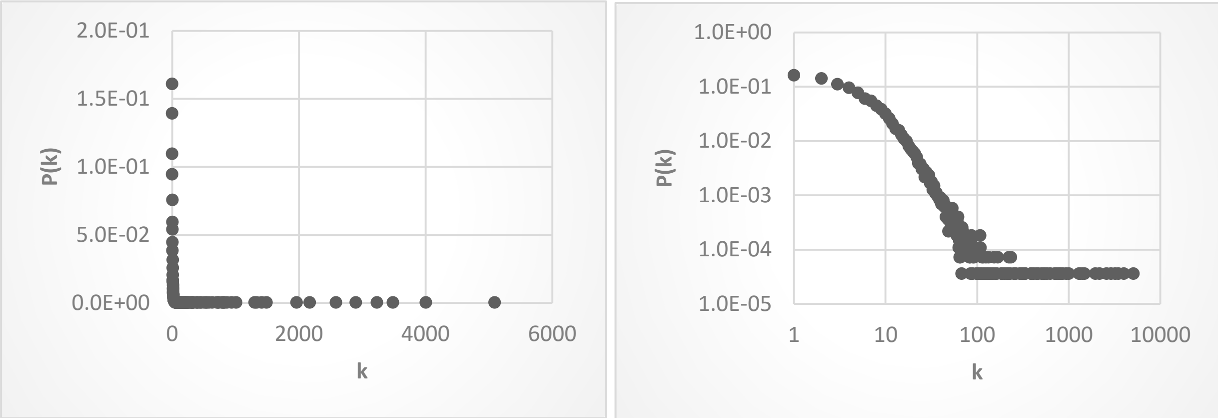

Figure 1. Different plots of the degree distribution of the Hungarian cattle holdings network (timeframe: 2012–2014); linear plot (left), log-log plot (right).

This heavy-tailed distribution shows that there are many small nodes (with few connections), and there are few very large nodes (with a lot of connections). The scale-free networks are networks whose degree distribution follows a power law. To prove the power law distribution and to obtain the degree exponent (γ), which is important for further analysis, a cumulative distribution was plotted and then a power law curve was fitted (Figure 2).

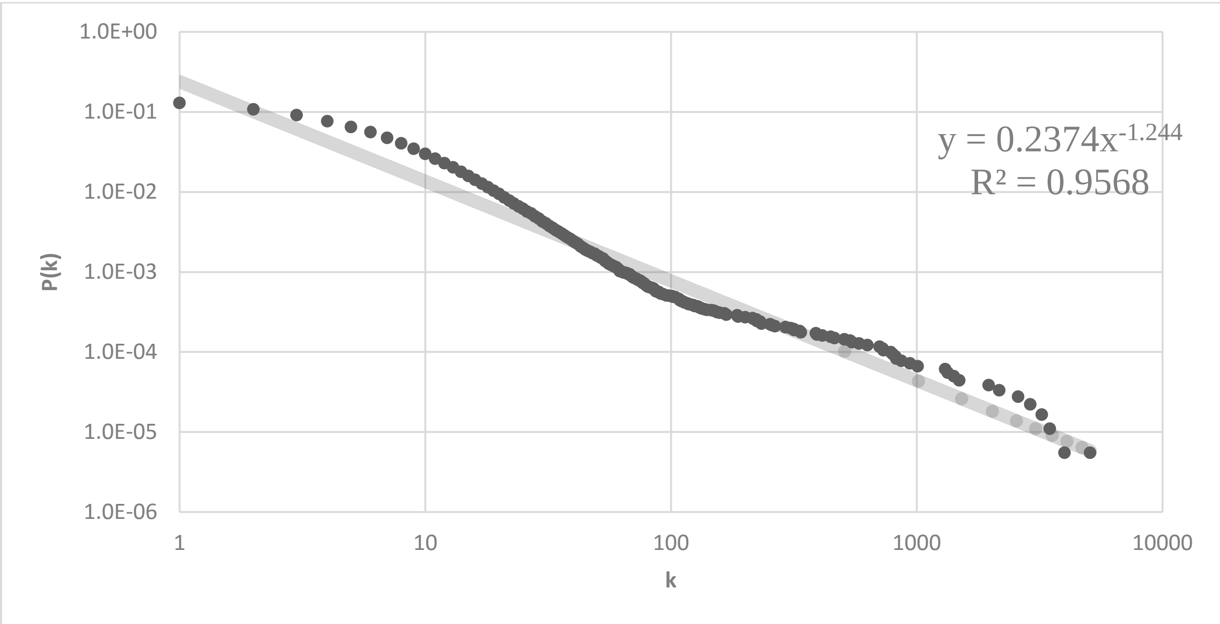

Figure 2. The degree distribution of the Hungarian cattle holdings network presented as a cumulative log-log plot from 2012–2014.

2) $$P_k = \sum_{q=k+1}^{\infty} p_q$$

In case of power law the cumulative distribution scales as

3) $$P_k \sim k^{-\gamma+1}$$

The degree exponent for the Hungarian cattle holdings network is **2.24**. As this network is a directed network, the scale-free property applies separately to the in- and the out-degrees (Figure 3).

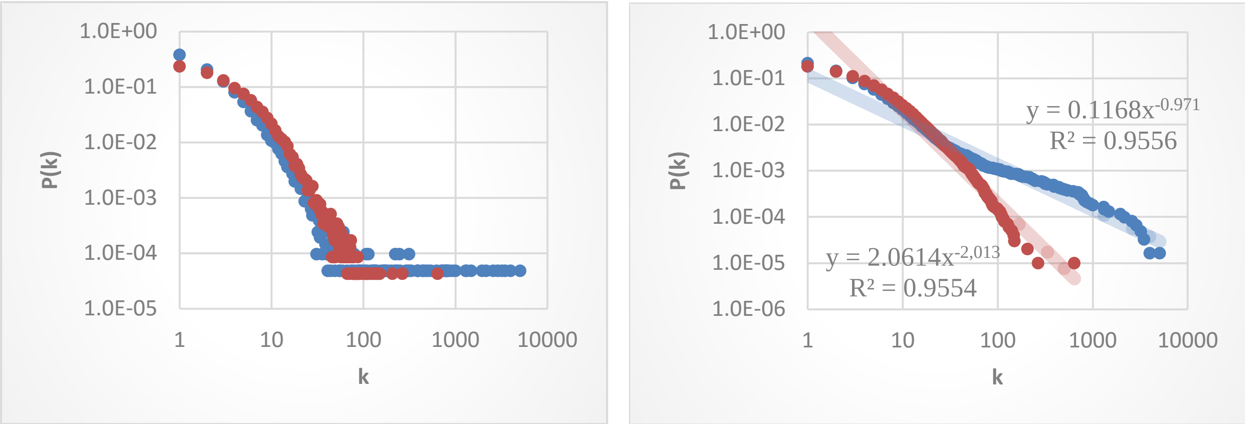

Figure 3. The in-degree and out-degree distributions of the Hungarian cattle holdings network from 2012–2014.

Note. Represented as a log-log plot (left) and a cumulative log-log plot (right). The in-degree is marked with blue and the out-degree is marked with red dots.

As it can be observed, the degree exponents are different for in-degree (γ = 1.97) and out-degree (γ = 3.01), showing a substantial difference between the two. This is attributable to the specific nodes with a very high in-degree e.g. slaughterhouses.

Similarly, the degree distribution can be calculated for the weighted degrees as well. Weighted degrees represent the size of the traffic going through a node; in this case, number of animals transported to and from the holdings (Figure 4.). The traffic size is important for food chain safety reasons because any epidemiological problem in a herd with intense traffic can be proliferated in the network extremely rapidly. That's why this piece of information helps risk management. The degree exponents can be calculated from the cumulative log-log plots of the weighted degree distribution (Figure 5).

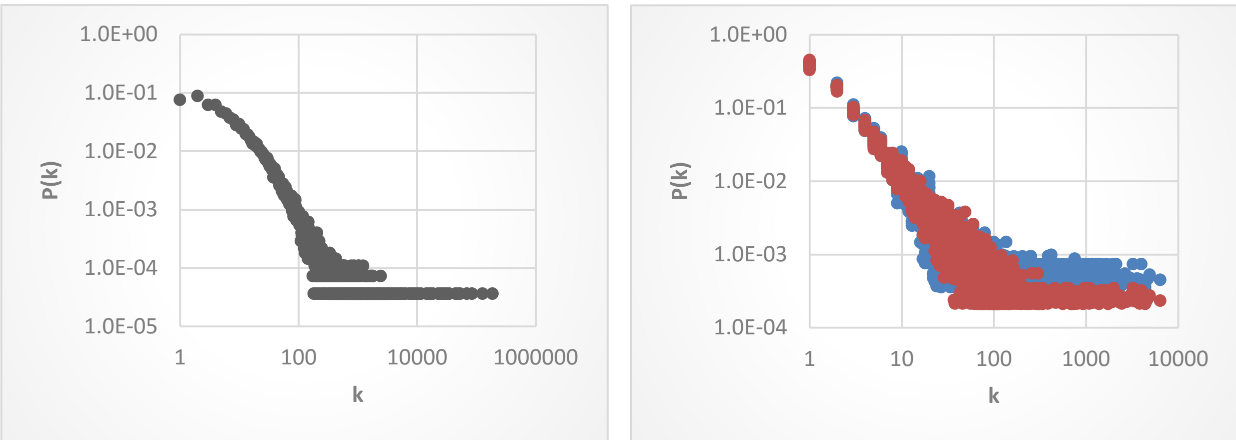

Figure 4. The weighted degree distributions of the Hungarian cattle holdings network represented as log-log plots from 2012–2014.

Note. The graphs show weighted degree (left) and a weighted in- and out-degree (right) distributions. The weighted in-degree is marked with blue and the out-degree is marked with red dots.

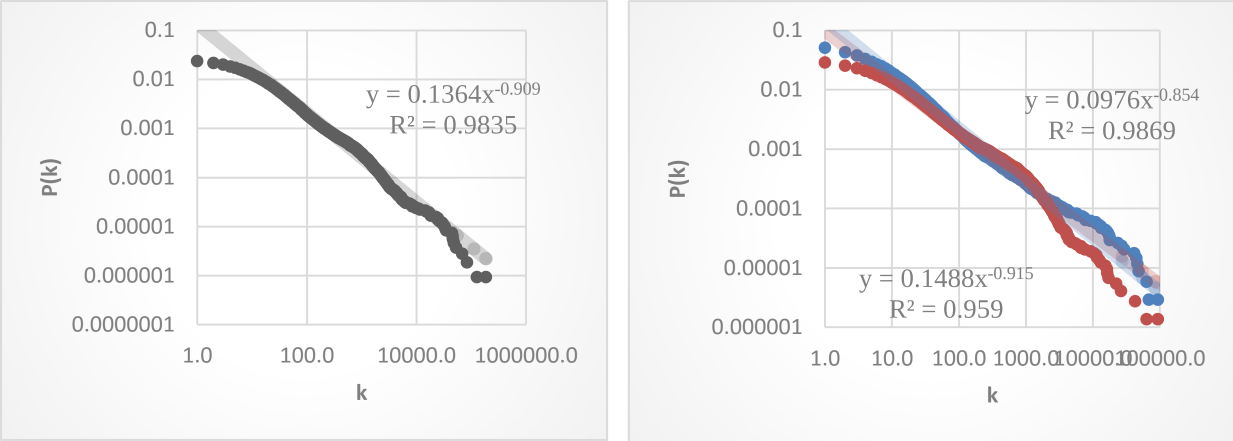

Figure 5. Weighted degree distributions of the Hungarian cattle holdings network represented as cumulative log-log plots from 2012–2014.

Note. The graphs show a weighted degree (left) and a weighted in- and out-degree (right) distributions. The weighted in-degree is marked with blue and the out-degree is marked with red dots.

Interestingly, degree exponents of the weighted degree, weighted in- and out-degree distributions are very similar to each other and fall in the range of $\gamma \sim 1.9$. The reason behind it is that the large weighted degree nodes are typically logistic centers where the incoming and outgoing flows are identical.

The results above show that this network has all the intrinsic properties of other scale-free networks, highlighting the fact that some holdings have a critical role in the network. Identifying those, we make a step towards controlling them. The scale-free name captures the lack of an internal scale, a consequence of the fact that nodes with widely different degrees co-exist in the same network. This feature distinguishes scale-free networks from lattices, in which all nodes have exactly the same degree ($\sigma = 0$), or from random networks whose degrees vary in a narrow range ($\sigma = \langle k \rangle^{1/2}$). This divergence is the origin of some of the most intriguing properties of scale-free networks, from their robustness to random failures to the anomalous spread of viruses (Barabási 2015).

This means that this network is quite robust against random failures but vulnerable in case of targeted attacks.

Having a degree exponent between two and three means this network also shows small world properties—meaning that only a few steps are needed to get from a random point to another random point—having an important implication in the case of spreading diseases. The average path length (steps needed to reach any random node from any other random node) for the Hungarian cattle holdings network is 6.92 for the three-year period.

Basic Network Properties

Devising the network's basic structural properties, other valuable information could be extracted from other network measures or indicators. As specified earlier, the degree, in-degree, out-degree, weighted degree, weighted in-degree and weighted out-degree of the nodes was calculated for different holdings, having an objective to set priority lists for different control purposes. The top five nodes (highest risk nodes) listed according to various properties are presented in Table 1.

Table 1. Top five nodes based on degree, in-degree, out-degree, weighted degree, weighted in-degree and weighted out-degree (timeframe: 2012–2014).

	ID	**Type**	**Degree**		**In-Degree**		**Out-Degree**		**Weighted Degree**		**Weighted In-Degree**		**Weighted Out-Degree**	
			Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value
Top 5 Degree	528781	L	1	5093	1	5091	16781	2	2	127196	3	63601	2	63595
	720342	S	2	4007	2	4007	49104	0	6	48419	4	48419	49104	0
	919273	L	3	3489	3	3483	6779	6	1	188300	1	94153	1	94147
	242344	S	4	3235	4	3233	16911	2	9	44956	6	44951	14995	5
	257710	S	5	2902	5	2901	20045	1	14	30050	8	30046	16035	4
Top 5 In-Degree	528781	L	1	5093	1	5091	16781	2	2	127196	3	63601	2	63595
	720342	S	2	4007	2	4007	49104	0	6	48419	4	48419	49104	0
	919273	L	3	3489	3	3483	6779	6	1	188300	1	94153	1	94147
	242344	S	4	3235	4	3233	16911	2	9	44956	6	44951	14995	5
	257710	S	5	2902	5	2901	20045	1	14	30050	8	30046	16035	4
Top 5 Out-Degree	230129	M	10	1417	19	774	1	643	55	4301	46	2503	112	1798
	456485	M	24	583	34	316	2	267	241	1362	114	764	392	598
	357532	F	48	217	1342	10	3	207	78	3099	1945	38	43	3061
	355807	M	35	318	52	167	4	151	370	890	149	521	556	369
	145593	F	60	150	9323	2	5	148	215	1472	2575	29	162	1443
Top 5 Weighted Degree	919273	L	3	3489	3	3483	6779	6	1	188300	1	94153	1	94147
	528781	L	1	5093	1	5091	16781	2	2	127196	3	63601	2	63595
	860928	F	20	734	21	725	3444	9	3	83254	7	39662	3	43592
	490540	I	7	2172	7	2168	10228	4	4	68908	2	68904	16316	4
	217330	F	39	265	39	247	930	18	5	54681	9	27528	4	27153
Top 5 Weighted In-Degree	919273	L	3	3489	3	3483	6779	6	1	188300	1	94153	1	94147
	490540	I	7	2172	7	2168	10228	4	4	68908	2	68904	16316	4
	528781	L	1	5093	1	5091	16781	2	2	127196	3	63601	2	63595
	720342	S	2	4007	2	4007	49104	0	6	48419	4	48419	49104	0
	582987	S	6	2590	6	2588	16542	2	8	45170	5	45153	6869	17
Top 5 Weighted Out-Degree	919273	L	3	3489	3	3483	6779	6	1	188300	1	94153	1	94147
	528781	L	1	5093	1	5091	16781	2	2	127196	3	63601	2	63595
	860928	F	20	734	21	725	3444	9	3	83254	7	39662	3	43592
	217330	F	39	265	39	247	930	18	5	54681	9	27528	4	27153
	941088	F	52	187	50	170	1077	17	7	47053	10	23801	5	23252

Note. The holding IDs are anonymized. L = logistics/distribution center; S = slaughterhouse, M = animal market; F = farm; I = incinerator.

As can be derived from the results presented in the table, the different network properties have different meaning from a real-life control perspective. Degree shows the connections between different holdings. The stability of these connections have implications on risk, in line with human epidemiology (e.g. in case of sexually transmitted diseases spread). The holdings tending to be more loyal to their business partners have lower risks compared to those switching their partners over time (Valdano et al. 2015).

The considerable (one order of magnitude) differences between the in- and out-degree values and distributions hold important information as well: nodes with the highest in-degrees (and weighted in-degrees) are slaughterhouses. They don't release animals so their out-degree is (virtually) 0; they are so-called sink vertices. In the case of hygiene, documentation, or traceability controls they are very important hubs to control and, in the case of some diseases, large slaughterhouses may be places where cross-contamination or cross-infection occurs. However, in case of other diseases they don't represent a real high risk vertex and they should be excluded from analyses since this is an end-point to animals. This implies that, depending on the actual control objective and the characteristics of the causative agent, those nodes should be included or excluded from analysis on a case-by-case basis.

Similarly, from a control perspective, it is important to observe nodes with low degree - high weighted degree (high trade activity with limited number of business partners), or high-degree, relatively low weighted degrees (high number of business partners, but limited trade with each of them), usually being markets or trans-loading stations.

As it is shown in the table, the maximum in-degree value of the network is 5091 and it belongs to a slaughterhouse; while the maximum out-degree value is only 643 (a market), confirming the phenomenon observed during degree distribution analysis, resulting in different degree exponents for in- and out-degrees. In contrary, the maximum values of the weighted in- and out-degrees are similar: 93,663 and 93,657, respectively (logistic center).

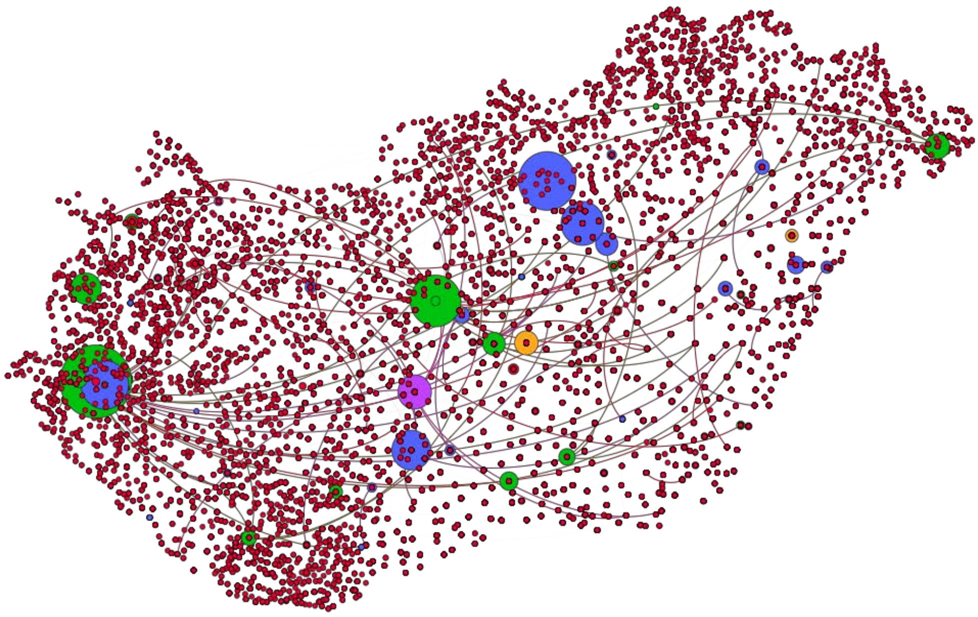

Figure 6. Geographical representation of the Hungarian cattle holdings network.

Note. The size of the nodes is influenced by degree, the color depends on the production type (red = livestock farm; blue = slaughterhouse; green = logistics/distribution center; orange = market; purple = incinerator). The color of the edges is influenced by source node, and the weight is limited to minimum 36 (at least one movement per month on average).

Pay attention to the fact, that 'export' was part of the dataset as one single node (since there is no information about the exact recipient holding), and the connections contributed to the degree values of the nodes but it was excluded from the ranking exercise. If we knew about the actual destination of the exported cattle, the out-degree would increase (with unchanged weighted out-degree). Similarly, the source of imported animals was present in the analysis as single nodes for each exporting country. Having information about the exact source holdings, the in-degree values would increase (with unchanged weighted in-degree). The export activity is far larger than import (207,094 and 55,240 movements in the three-year period, respectively), meaning if more precise data on holdings outside Hungary were present the difference between in- and out-degree distribution would decrease. The geo-layout of the Hungarian cattle holdings network is presented on Figure 6.

Centrality Measures

The most vulnerable points of a network are not necessarily their largest hubs, as discussed previously. To extract information on the nodes playing a central role in the network, different centrality measures were calculated: betweenness, closeness, authority and hub centralities were determined. The vertices of high betweenness centrality value are usually logistic centers, transloading places or major livestock farms. These nodes have an important role in epidemiological investigations because of the high risk of cross-infections. The top five nodes sorted according to different centrality values are presented in Table 2.

Table 2. Top five nodes based on betweenness centrality, closeness centrality, authority and hub centrality (timeframe: 2012-2014).

	ID	Type	Betweenness centrality		Closeness centrality		Authority		Hub centrality	
			Rank	Value	Rank	Value	Rank	Value	Rank	Value
Top 5 Betweenness centrality	230129	M	1	124525786	1	4,353	3	0,00529509	3	0,00658296
	769670	F	2	87082032	21	4,971	8	0,00267829	8	0,00293686
	456485	M	3	74391523	2	4,481	9	0,00216586	9	0,00255726
	919273	L	4	44847391	1893	6,118	1	0,02380399	1	0,02744064
	447999	F	5	34191545	252	5,285	14	0,00121616	12	0,00143846
Top 5 Closeness centrality	230129	M	1	124525786	1	4,353	3	0,00529509	3	0,00658296
	456485	M	3	74391523	2	4,481	9	0,00216586	9	0,00255726
	448931	F	40	3803558	3	4,640	957	0,00008199	638	0,00011987
	583058	F	39	3893053	4	4,761	3496	0,00003416	2908	0,00004995
	806192	F	87	2191809	5	4,817	3499	0,00003416	2911	0,00004995
Top 5 Authority	919273	L	4	44847391	1893	6,118	1	0,02380399	1	0,02744064
	490540	I	6	29500637	3143	6,530	2	0,01481942	2	0,01512382
	230129	M	1	124525786	1	4,353	3	0,00529509	3	0,00658296
	860928	F	44	3660243	2597	6,362	4	0,00494664	4	0,00569391
	431898	F	9	18940054	249	5,277	5	0,00335470	5	0,00379594
Top 5 Hub centrality	919273	L	4	44847391	1893	6,118	1	0,02380399	1	0,02744064
	490540	I	6	29500637	3143	6,530	2	0,01481942	2	0,01512382
	230129	M	1	124525786	1	4,353	3	0,00529509	3	0,00658296
	860928	F	44	3660243	2597	6,362	4	0,00494664	4	0,00569391
	431898	F	9	18940054	249	5,277	5	0,00335470	5	0,00379594

Note. The holding IDs are anonymized. L = logistics/distribution center; S = slaughterhouse, M = animal market; F = farm; I = incinerator.

On the basis of the centrality measure other extremities–the "peripheral holdings" could be defined. These entities are not regular participants of the global cattle network. Their role is marginal in the network as a whole, but—taking into consideration their often low technological level—it is important to include their activity because they can be sources of epidemiological problems.

As described earlier, different centrality concepts capture different aspects of a central node. It is an important question from a food chain safety perspective which concept is the most useful from a risk analysis point of view. It is out of the scope of this paper to answer this question. However, with the help of the large amount of data it was possible to analyze the stochastic relationships between the centrality indicators. We decided to filter our analysis since, given all the nodes from the network, correlation figures are largely biased due to sink vertices, export (represented as one single node) and import (source countries as nodes) data and the holdings characterized by small throughput. Therefore, the nodes (and the corresponding edges) outside Hungary were excluded from the calculation of centrality values, then the nodes with <3 in-degree and out-degree (at least one in and out connection a year) were excluded from the correlation analysis as well as nodes with betweenness centrality value of 0 and closeness centrality value of 1 (nodes with a small number of connections, not being part of the giant component of the network). Then the correlation between the different centrality results was calculated (Table 3**).** To understand the relationship between different centrality indicators, we have applied regression analysis. This is an extremely important step because on this basis we will be able to understand whether there is a possibility to decrease the number of centrality indicators to judge the position of a given vertex of not.

As seen from the results, there is a weak negative correlation between closeness and betweenness centralities, and a stronger correlation between betweenness centrality and hub centrality and authority. The strong relationship between hub and authority centrality can be explained by their similar role: in the opinion of Kleinberg (1999) hubs and authorities stand in a mutually reinforcing relationship. Valente et al. (2008) showed in their work a slight correlation between betweenness and closeness centrality, indicating that these measures are distinct, yet conceptually related.

Table 3. Stochastic relationship between betweenness centrality, closeness centrality, authority and hub centrality in case of the Hungarian cattle holdings network.

	Betweenness centrality	**Closeness centrality**	**Authority**
Betweenness centrality	1		
Closeness centrality	-0.1263	1	
Authority	0.4932	-0.0619	1
Hub centrality	0.5054	-0.0634	0.9975

Note. Time period: 2012–2014; directed network

Dynamic Patterns

The dataset offered a possibility to analyze the dynamic patterns of the network, to observe and draw conclusions on the time-dependent features which may have an influence on the planning of control activities. To that aim, monthly, annual and the whole dataset for three years were compared in this section.

The simplest approach is to observe the number of animals moving per month. The results (Figure 7) indicated that the trade becomes very active in June–July with a peak of activity at the end of the year. The trend of increasing activity in the second half of the year seems to be stable. This should have an impact on the control time schedules, assigning increased control frequencies to those periods.

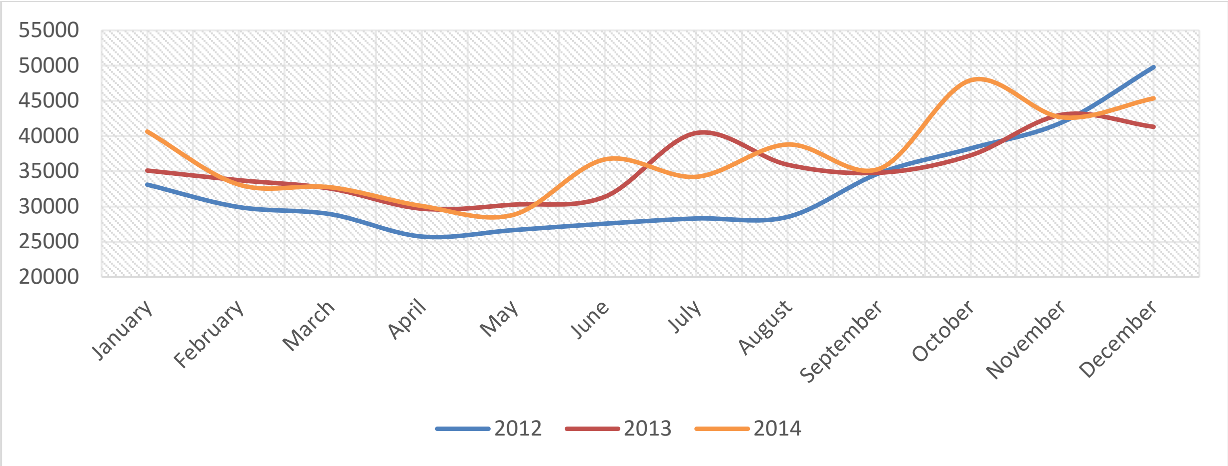

Figure 7. Monthly movement of the animals in the Hungarian cattle holdings network.

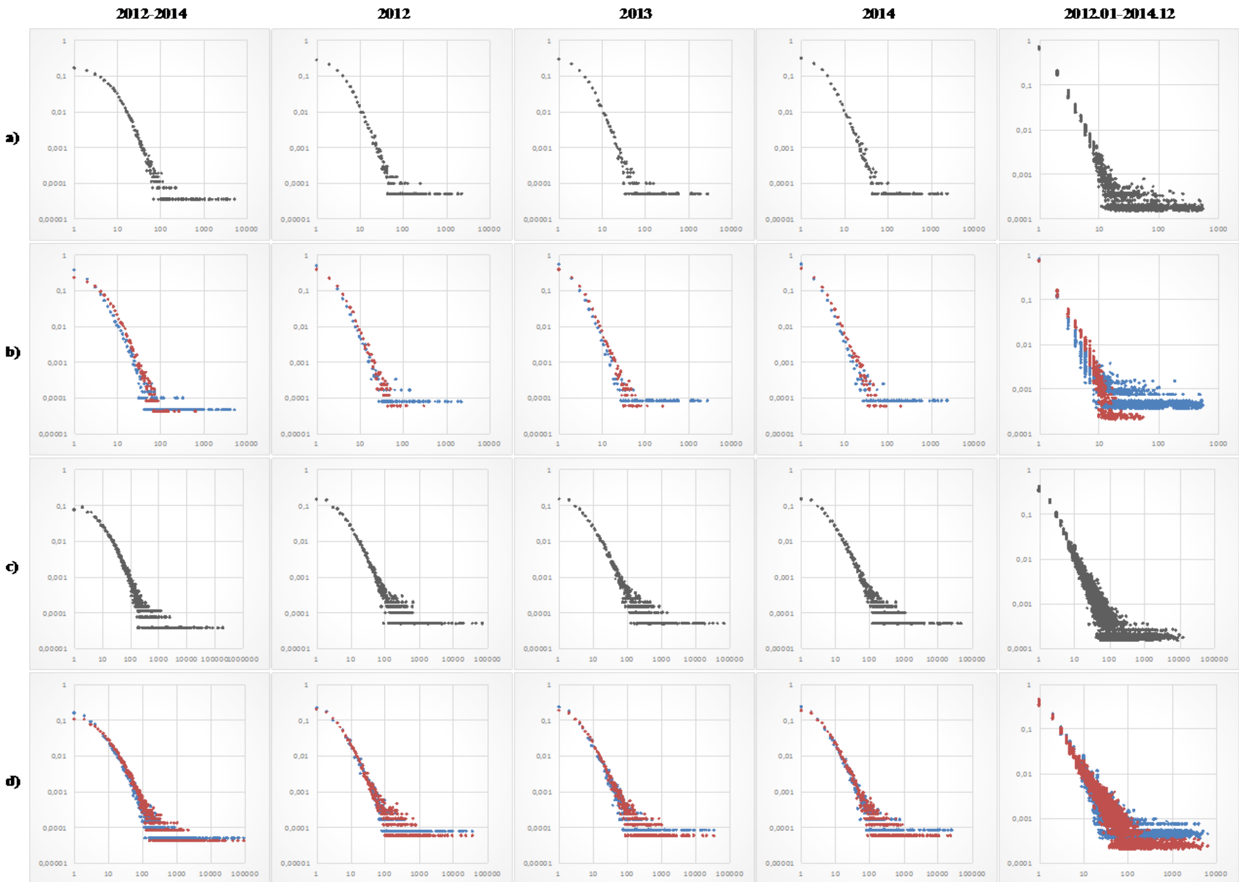

Figure 8. Dynamic patterns of the Hungarian cattle holdings network.

Note. a) Degree; b) In-Degree (blue) and Out-Degree (red); c) Weighted Degree; d) Weighted In-Degree (blue) and Weighted Out-Degree (red) distributions presented in 3-year, annual and monthly breakdowns.

During analysis, the changes in the activity of the holdings were recorded along with the analysis of the dynamic patterns of the entire network properties (Figure 8).

It can be derived from the results that, apart from small differences, network characteristics are quite stable over time, allowing for predictions at the overall network level. We selected the top five nodes on the betweenness centrality rank list and plotted the monthly betweenness centrality values to see when the nodes played a central role in the network over the three year period. The results (Figure 9) show a very volatile nature (differences of many orders of magnitude between months) of the holdings in relation to betweenness centrality values. It also shows that the analysis of the dynamic patterns is valuable especially in case of single holding analysis: performing time-dependent assessments, the results could be used for effective targeting of control, or prediction purposes as well.

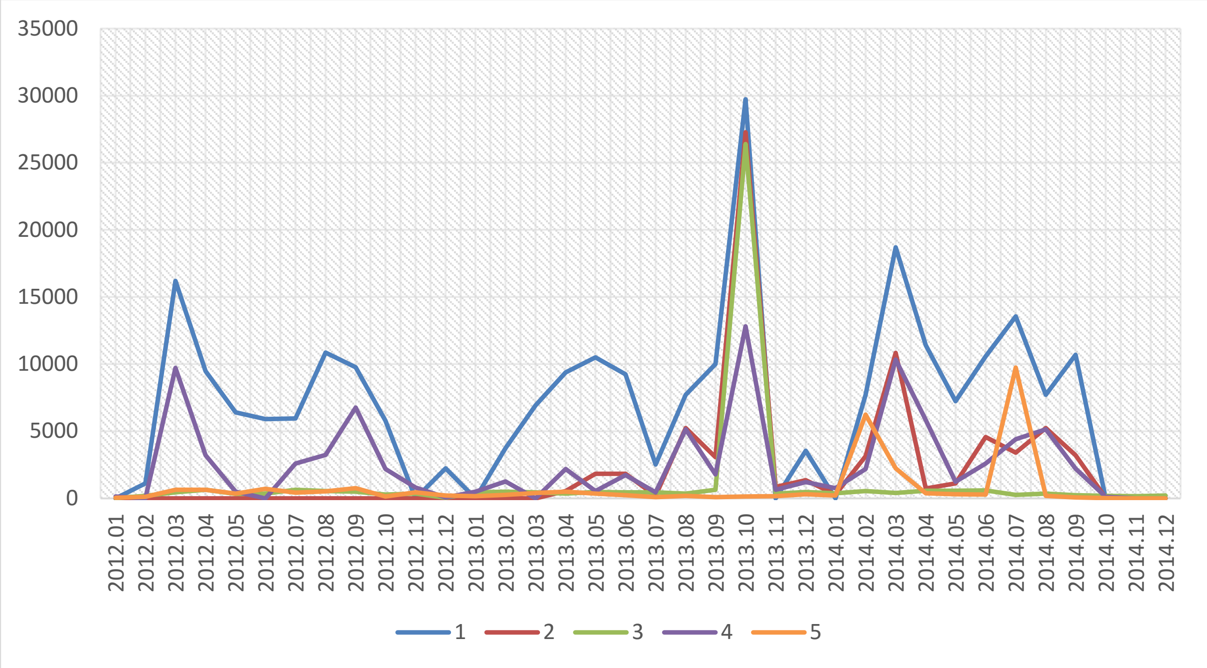

Figure 9. Changes in betweenness centrality values of the top five holdings of the Hungarian cattle holdings network during a thirty-six- month period.

Application of the Results in Practice

The results presented above all contribute to network analysis based, risk based control plans. The outputs of the analysis served as valuable input information in the planning process of official control plans. As an output, 100 highest risk cattle holdings were selected for food chain safety control based on the basic network properties, as well as on centrality measures. Those holdings are controlled for biosecurity measures, hygiene, animal welfare rules, safety assurance systems, documentation, etc. Furthermore, 100 highest risk holdings were selected for animal identification control. Those holdings are controlled for the identification and traceability rules.

The analyses performed provide information on the source and routes of possible infections so that preventive and control measures can be applied, increasing preparedness of the food chain stakeholders. In case of an outbreak, the mapped network makes a rapid traceability and epidemic spreading prediction possible, allowing for effective risk management.

Implications

Globalization, the data explosion, fast changing trade routes and food technologies are the important drivers which inspire us to develop new analysis and assessment methods in the field of food chain safety. There is a strong need for an interdisciplinary approach to monitor, understand, and control the trade-flow in the food chain.

The data on movements of cattle are increasingly becoming available thanks to identification and tracing systems put in place in the European Union. By using the approaches and techniques of network science it is possible to analyze the dynamic system of cattle movements, going beyond static and simple approximations (Bajardi et al. 2011; Natale et al. 2009).

During the analysis of the Hungarian cattle holdings network the basic structure of the network was revealed, showing scale-free properties, thus having serious implications from a food chain safety control perspective. This network has small world properties meaning that – because of the hubs and high centrality nodes—there is a small distance between any random holdings, potentially resulting in a rapid spread of epidemic as the spread of a pathogen on a scale-free network is instantaneous (Barabási 2015). This should be taken into account during the preventative measures (including e.g. vaccination strategies) at a business level as well as at an official control planning level. Furthermore, this phenomenon has very important implications for the forecasting and risk or crisis management in case of an actual outbreak, showing the possible advantages of using network spreading models in conjunction with traditional epidemiological modelling (Pastor-Satorras and Vespignani 2001).

The other consequence of having scale-free properties is the vulnerability to intentional attacks against the hubs or central nodes, showing the growing need for the control and preparedness of that critical infrastructure. It is important to emphasize that intentional attacks on the network behave differently and need a slightly different analytical and risk management approach compared to unintentional events. Epidemics follow the rules set by the characteristics of the infectious agents, while in case of intentional attacks, different spreading models should be used, based on socio-psychological and economic analysis.

On this basis, suggestions have been formulated for the food chain safety authority determining which farms should be the focus of their control activity. The list of the highest risk holdings obtained by network analysis is used directly by risk managers when outlining their annual control plans. Should any epidemiological problem occur, the easily updatable database on network characteristics offers essential input for further optimization of the control strategy. The tool used is suitable for a rapid assessment of a huge and complex system within minutes after data cleaning. It is possible to give a very informative graphical representation of the cattle holding network, making possible to easily choose control or audit targets. During the analysis of the dynamic properties of the network we revealed further possibilities to explore, hence making network based epidemiological simulations the next item on our research agenda.

As it could be seen from the difference of the ‘real life’ meaning of various network properties, and considering the implications of those, critical thinking during the application of the results is essential. Substantial knowledge of the food chain safety science is needed for the correct interpretation of the network analysis results and advanced skills in computational science are important in extracting valuable information from the underlying network data. This inter- and multidisciplinary field of science calls for such experts and the need for capacity building.

In addition, the software tools available are not by themselves suitable alone network-based food chain safety analyses. The data cleaning, transformation and enrichment steps needed to obtain a dynamic dataset suitable for network-analysis (as demanded by the software used) require many steps and, after the usual network analysis, many calculations are done over using other analysis tools. This calls for dedicated software development in the future, to decrease the time needed between receiving raw data and delivering pertinent information to decision makers. In epidemic situations, time is of utmost importance.

An important aim of this ongoing research is to share the methodology and algorithms with the network science and food chain safety community, in order to enhance the capacity building process and to improve the cross-compatibility of the methods. This makes it possible to expand simulation exercises and risk based planning processes across borders, as real world situations don't respect borders either. For that reason, the anonymized raw data, the data cleaning process, the analysis algorithms and the Gephi software settings used are published on the website of Hungarian National Food Chain Safety Authority (NÉBIH).[1] Furthermore, for the sake of better illustration, particularly for educational purposes, the key issues of the article (graph-dynamics) are illustrated in a Prezi, based on a series of Gephi files on the same site.

This study opens the road to future work in several directions. This work contributes to 1) determining the most vulnerable parts of a cattle holding network; 2) increasing the effectiveness of the control of the cattle-flow; 3) revealing the interdependencies; 4) helping to work out an optimized strategy for the inspection of herds; 5) increasing the preparedness against outbreaks and intentional attacks; 6) enhancing epidemiological modelling simulations; 7) providing information on the source of possible infections so that preventive and control measures can be applied; and finally 8) serving the food chain safety and network science community with analyzable data and helpful descriptions of the methodology to enhance cross-border co-operation.

References

Abbasi, A., L. Hossain, and L. Leydesdorff. 2012. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics* 6 (3): 403–412.

Agarwal, P. K., S. Har-Peled, H. Kaplan, and M. Sharir. 2014. Union of random Minkowski sums and network vulnerability analysis. *Discrete & Computational Geometry* 52 (3): 551–582.

Albert, R. and A.-L. Barabási. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74 (1): 47–91.

Anderson, J. E. and D. Marcouiller. 2002. Insecurity and the Pattern of Trade: An Empirical Investigation. *Review of Economics and Statistics* 84 (2): 342–352.

Bagni, R., R. Berchi, P. Cariello, 2002. A comparison of simulation models applied to epidemics. *Journal of Artificial Societies and Social Simulation* 5 (3).

[1] http://portal.nebih.gov.hu/web/guest/-/network-science-based-decision-support-in-food-chain-safety-systems.

Bajardi, P., A. Barrat, F. Natale, L. Savini, and V. Colizza. 2011. Dynamical Patterns of Cattle Trade Movements. *PLoS ONE* 6 (5): e19869.

Bajardi, P., A. Barrat, L. Savini and V. Colizza. 2012. Optimizing Surveillance for Livestock Disease Spreading Through Animal Movements. *Journal of the Royal Society Interface* 9 (76): 2814–2825.

Barabási, A.-L. 2002. *Linked: The New Science of Networks*. Perseus Books Group, USA. ISBN 0-452-28439-2.

Barabási, A.-L. 2015. Network Science. http://barabasi.com/networksciencebook/ [accessed October 20, 2015]

Barabási, A.-L. and R. Albert. 1999. Emergence of Scaling in Random Networks. *Science* 286 (5439): 509–512.

Baranyi, J., A. Jóźwiak, L.Varga, M. Mézes, J. Beczner, and J. Farkas. 2013. [Application Potentials of Network Science, Bioinformatics and Systems Biology to Food Science]. *Hungarian Science* 174 (9): 1094–1102.

Barrat, A., M. Barthélémy, R. Pastor-Satorras, and A.Vespignani, 2004. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences* 101 (11): 3747–3752.

Bojnec, S. and I. Fertő. 2009. Agro-food trade competitiveness of Central European and Balkan countries. *Food Policy* 34 (5): 417–425.

Bökönyi, S. 1971. The Development and History of Domestic Animals in Hungary: The Neolithic Through the Middle Ages. *American Anthropologist* 73 (3): 640–674.

Borgatti, S. P., A. Mehra, D. J. Brass, G. Labianca. 2009. Network Analysis in the Social Sciences. *Science* 323 (5916): 892-895.

Börner, K., S. Sanyal, A. Vespignani. 2007. Network science. *Annual Review of Information Science and Technology* 41 (1): 537-607.

Brandes, U. 2001. A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology* 25 (2): 163–177.

Chmiel, A. M., J. Sienkiewicz, K. Suchecki and J. A. Hołyst. 2007. Networks of companies and branches in Poland. *Physica A: Statistical Mechanics and its Applications* 383 (1): 134–138.

Coulombier, D. and J. Takkinen. 2013. From national to international challenges in cross-border multi-country, multi-vehicle foodborne outbreak investigations. *Euro surveillance: bulletin Européen sur les maladies transmissibles = European communicable disease bulletin* 18 (11): 20423.

Csáki, C. 1990. Agricultural Changes in Eastern Europe at the Beginning of the 1990s. *American Journal of Agricultural Economics* 72 (5): 1233–1242.

Csizmadia, E. 1974. New features in Hungarian agriculture in the 1970s. *Acta Oeconomica* 12 (3-4): 349–365.

David, A. S. and Douglas, R. W. 1992. Structure and Dynamics of the Global Economy: Network Analysis of International Trade 1965–1980. *Social Forces* 70 (4): 857–893.

Devangana, K. 2015. *Gephi Cookbook*. Packt Publishing. Birmingham, UK. ISBN: 978-1-78398-741-2.

Dion, E., L. VanSchalkwyk, E.F. Lambin, 2011. The landscape epidemiology of foot-and-mouth disease in South Africa: A spatially explicit multi-agent simulation. *Ecological Modelling* 222 (13): 2059–2072.

Dubé, C., C. Ribble, D. Kelton, and B. McNab. 2008. Comparing network analysis measures to determine potential epidemic size of highly contagious exotic diseases in fragmented monthly networks of dairy cattle movements in Ontario, Canada. *Transboundary and Emerging Diseases* 55 (9-10): 382–392.

Dubé, C., C. Ribble, D. Kelton, and B. McNab. 2009. A Review of Network Analysis Terminology and its Application to Foot-and-Mouth Disease Modelling and Policy Development. *Transboundary and Emerging Diseases* 56 (3): 73–85.

Ercsey-Ravasz, M., Z. Toroczkai, Z. Lakner, and J. Baranyi. 2012. Complexity of the International Agro-Food Trade Network and Its Impact on Food Safety. *PLoS ONE* 7 (5): e37810.

FAO/WHO. 2007. *Food safety risk analysis. A guide for national food safety authorities.* Food and Agriculture Organization of the United Nations, Rome, Italy. http://www.fao.org/docrep/012/a0822e/a0822e00.htm [accessed October 5, 2015]

Farhat, R., 2012. Management of logistics and ICT in food supply chain network: a conceptual framework. *International Journal of Logistics Economics and Globalisation* 4 (3): 162-178.

Faust, K. and S. Wasserman. 1992. Centrality and prestige: A review and synthesis. *Journal of Quantitative Anthropology* 4 (1): 23–78.

Freeman, L. C. 1979. Centrality in social networks conceptual clarification. *Social Networks* 1 (3): 215–239.

Friedkin, N. E. 1991. Theoretical foundations for centrality measures. *American Journal of Sociology* 96 (6): 1478–1504.

Friesz, T. L., D. Bernstein, T. E. Smith, R. L. Tobin, B. W. Wie. 1993. A variational inequality formulation of the dynamic network user equilibrium problem. *Operations Research* 41 (1): 179–191.

Fritz, M., and G. Schiefer. 2008. Food chain management for sustainable food system development: a European research agenda. *Agribusiness* 24 (4): 440–452.

Harris, D. R. 1979. Continuities and change in tropical savanna environments. *Current Anthropology* 20 (2): 394–397.

Jewell, C., J. Brown, M. Keeling, L. Green, G. Roberts, K.I.P. Verheyen, C. Fourichon. 2013. Bayesian epidemic risk prediction-knowledge transfer and usability at all levels. In, Society for Veterinary Epidemiology and Preventive Medicine. Proceedings of a meeting held in Madrid, Spain, 20-22 March 2013. *Society for Veterinary Epidemiology and Preventive Medicine* 127–142.

Kao, R. R. 2002. The role of mathematical modelling in the control of the 2001 FMD epidemic in the UK. *Trends in Microbiology* 10 (6): 279–286.

Kao, R. R., D. M. Green, J. Johnson, and I. Z. Kiss. 2007. Disease dynamics over very different time-scales: foot-and-mouth disease and scrapie on the network of livestock movements in the UK. *Journal of The Royal Society Interface* 4 (16): 907–916.

Kim, H. and R. Anderson. 2012. Temporal node centrality in complex networks. *Physical Review E* 85 (2): 026107.

Kleinberg, J. M. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM* 46 (5): 604–632.

Kleinberg, J. M. 2000. Navigation in a small world. *Nature* 406 (6798): 845–845.

Lazzarini, S., F. Chaddad, M. Cook. 2001. Integrating supply chain and network analyses: the study of netchains. *Journal on Chain and Network Science* 1(1): 7–22.

Lehmann, R.J., R. Reiche, G. Schiefer. 2012. Future internet and the agri-food sector: State-of-the-art in literature and research. Computers and Electronics in Agriculture 89 (2): 158–174.

Lentz, H. H. K., M. Konschake, K. Teske, M. Kasper, B. Rother, R. Carmanns, B. Petersen, F. J. Conraths, and T. Selhorst. 2011. Trade communities and their spatial patterns in the German pork production network. *Preventive Veterinary Medicine* 98 (2–3): 176–181.

Luning, P. A., K. Kirezieva, G. Hagelaar, J. Rovira, M. Uyttendaele, and L. Jacxsens. 2015. Performance assessment of food safety management systems in animal-based food companies in view of their context characteristics: A European study. *Food Control* 49: 11–22.

MAF. 2009. Review of Selected Cattle Identification and Tracing Systems Worldwide. Lessons for the New Zealand National Animal Identification and Tracing (NAIT) Project. MAF Biosecurity New Zealand Information Paper No: 2009/3. Wellington, New Zealand.

Marsden, Peter V. 2002. Egocentric and Sociocentric Measures of Network Centrality. *Social Networks* 24 (4): 407–422.

Martínez-López, B., A. M. Perez, and J. M. Sánchez-Vizcaíno. 2009. Social Network Analysis. Review of General Concepts and Use in Preventive Veterinary Medicine. *Transboundary Emerging Diseases* 56 (4): 109–120.

Natale, F., A. Giovannini, L. Savini, D. Palma, L. Possenti, G. Fiore, and P. Calistri. 2009. Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. *Preventive Veterinary Medicine* 92 (4): 341–350.

National Institute of Standards and Technology-Big Data Working Group. 2013. http://bigdatawg.nist.gov/home.php. [accessed October 2, 2015]

Newman, M. E. J. 2001. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E.* 64 (1): 1–7.

Newman, M. E. J., 2005. A measure of betweenness centrality based on random walks. *Social Networks* 27 (1): 39–54.

Ng, D., Siebert, J. W. 2009. Toward better defining the field of agribusiness management. *International Food and Agribusiness Management Review* 12 (4):123–142.

Nikiforuk, A. 2008. Pandemonium: Bird Flu Mad Cow and Other Biological Plagues of the 21st Century. Penguin, Canada.

Okoth, I. O. and S. Wagner. 2009. Locally Oriented Noncrossing Trees. *The Electronic Journal of Combinatorics* 22 (3): P3.36.

Opsahl, T., F. Agneessens and J. Skvoretz. 2010. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks* 32 (3): 245–251.

Pastor-Satorras, R. and A. Vespignani. 2001. Epidemic spreading in scale–free networks. *Physical Review Letters* 86 (14): 3200–3203.

Petróczi, A., T. Nepusz, G. Taylor, and D. P. Naughton. 2011. Network analysis of the RASFF database: a mycotoxin perspective. *World Mycotoxin Journal* 4(3): 329–338.

Petróczi, A., G. Taylor, T. Nepusz, and D. P. Naughton. 2010. Gate Keepers of EU Food Safety: Four States Lead on Notification Patterns and Effectiveness. *Food and Chemical Toxicology* 48 (7): 1957–1964.

Power, D. J. 2014. Using 'Big Data' for analytics and decision support. *Journal of Decision Systems* 23 (2): 222–228.

Regulation (EC) No 1760/2000 of the European Parliament and of the Council of 17 July 2000 establishing a system for the identification and registration of bovine animals and regarding the labelling of beef and beef products and repealing Council Regulation (EC) No 820/97. *Official Journal of the European Communities* L 204, 11-August 2000.

Regulation (EC) No 882/2004 of the European Parliament and of the Council of 29 April 2004 on official controls performed to ensure the verification of compliance with feed and food law, animal health and animal welfare rules. *Official Journal of the European Union*, L 165, April 30, 2004.

Rommel, S. A. and J. D. McCleave. 1973. Prediction of oceanic electric fields in relation to fish migration. *Journal du Conseil* 35 (1): 27–31.

Salathé, M. and Jones, J. H. 2010. Dynamics and Control of Diseases in Networks with Community Structure. *PLOS Computational Biology* 6 (4): e1000736.

Schroeder, T. C. and G. T. Tonsor. 2012. International Cattle ID and Traceability: Competitive Implications for the US. *Food Policy* 31: 31–40.

Smith, D.A., D.R. White. 1992. Structure and dynamics of the global economy: network analysis of international trade 1965–1980. *Social forces* 70 (4): 857–893.

Sonka, S. 2014. Big Data and the Ag Sector: More than Lots of Numbers. *International Food and Agribusiness Management Review* 17 (1): 1–20.

Stanley, W. and F. Katherine, 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, UK.

Sugar, P. F., P. Hanak, and T. Frank 1994. *A History of Hungary*. Indiana University Press, Bloomington, Indiana, USA.

Tichy, N. M., L.T. Michael, and F. Charles. 1979. Social Network Analysis for Organizations. *Academy of Management* 4 (4): 507–520.

Valdano, E., C. Poletto, A. Giovannini, D. Palma, L. Savini, and V. Colizza. 2015. Predicting Epidemic Risk from Past Temporal Contact Data. *PLOS Computational Biology* 11 (3): e1004152.

Valente, T. W., K. Coronges, C. Lakon, and E. Costenbader, 2008. How Correlated Are Network Centrality Measures? *Connect (Tor)* 28 (1): 16–26.

Van Asselt, E. D., P. Sterrenburg, M.Y. Noordam, and H. J. Van der Fels-Klerx. 2012. Overview of available methods for risk based control within the European Union. *Trends in Food Science & Technology* 23 (1): 51–58.

Wagner, C. S., and L. Leydesdorff. 2005. Network structure, self-organization, and the growth of international collaboration in science. *Research Policy* 34 (10): 1608–1618.

Wang, L., S. Jajodia, A. Singhal, P. Cheng, and S. Noel. 2014. k-Zero Day Safety: A Network Security Metric for Measuring the Risk of Unknown Vulnerabilities. *IEEE Transactions on Dependable and Secure Computing* 11 (1): 30–44.

Wang, L., S. Noel, and S. Jajodia. 2006. Minimum-cost network hardening using attack graphs. *Computer Communications* 29 (18): 3812–3824.

Ward, J.S., and A. Barker. 2013. Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821.

Wasserman, S., and K. Faust. 1994. Social network analysis: methods and applications. Cambridge: Cambridge University Press.

Wen, J.T., and M. Arcak. 2004. A unifying passivity framework for network flow control. *IEEE Transactions on Automatic Control* 49 (2): 162–174.

Wu, X., Zhu, X. Wu, G.-Q. and Ding, W. 2014. Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering* 26 (1): 97–107.

