

Introduction to network analysis and its implications for animal disease modelling

C. Dubé⁽¹⁾, C. Ribble⁽²⁾, D. Kelton⁽³⁾ & B. McNab⁽⁴⁾

(1) Animal Health and Production Division, Canadian Food Inspection Agency, 59 Camelot Drive, Ottawa, Ontario, K1A 0Y9, Canada

(2) Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, TRW 2D22, 3330 Hospital Drive NW, Calgary, Alberta, T2N 4N1, Canada

(3) Department of Population Medicine, Clinical Research Bldg #174, Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada

(4) Animal Health and Welfare Branch, Ontario Ministry of Agriculture, Food and Rural Affairs, 1 Stone Rd, Guelph, Ontario, N1G 4Y2, Canada

Summary

Social networks analysis (SNA) has recently been used in veterinary epidemiology to study livestock movements. A network is obtained by considering livestock holdings as nodes in a network and movements among holdings as links among nodes. Social networks analysis enables the study of the network as a whole, exploring all the relationships among pairs of farms. Highly connected livestock holdings in the network can be identified, which can help surveillance and disease prevention activities. Observed livestock movement networks in various countries have shown an important level of contact heterogeneity and clustering (topological, not necessarily geographical or spatial) and understanding the architecture of these networks has provided a better understanding of how infections may spread. The findings of SNA studies of livestock movement should be used to build and parameterise epidemiological models of infection spread in order to improve the reliability of the outputs from these models.

Keywords

Epidemiological modelling – Infection spread – Livestock movements – Network analysis.

Introduction

Modelling the spread of animal diseases, particularly highly contagious diseases, requires knowledge of the patterns of animal movements. Animals moved from farm to farm, for example, represent a risk of propagation of an infectious disease agent. The knowledge of where and when animals move on and off premises is also a critical piece of information during an epidemic in order to trace potentially infected animals. As a result, a number of countries throughout the world have created livestock movement databases. The information contained in these databases is now available to researchers wanting to develop parameters for epidemiological modelling studies.

Until recently, the most common approach to studying animal movements was to obtain information on the frequency of movements on and off farms (11, 16, 51, 52).

With the availability of livestock movement data, it now becomes possible to build networks of livestock holdings connected to each other through the movement of animals. In these networks the unit of interest is the holding, or *node*, and the relationship is the movement, or *tie*, which produces *paths* on which infectious disease agents can spread. Social network analysis (SNA) provides the tools and methods to study these networks as a whole and to understand the role each holding plays in the network. This can lead to the identification of holdings that are central in the flow of animals in the population, perhaps due to their large number of commercial partners for example. These holdings could be targeted for surveillance activities in order to accelerate the detection of highly contagious diseases. Social network analysis can also help us understand the potential spread of an infectious disease agent during the silent spread phase, which is the time period between introduction of an infectious agent in a population and first detection.

Social network analysis is an approach that is based on the study of the relationships among social entities, and on the patterns and implications of these relationships (56). Two reviews have been published describing the terminology and applications of the technique in preventive medicine and epidemiological modelling (24, 41). The approach is based on graph theory, which is used to study pair-wise relationships between objects in the same collection. This approach has only recently been used for analysing livestock movements (12, 13, 18, 23, 37, 48, 49, 58, 59). The objective of this paper is to introduce the most commonly used concepts and measures of SNA in the study of animal movements and how these can be used to support animal disease modelling.

Network representation and description

Box 1, based on the table provided in the review paper by Dubé *et al.* (24), provides the definitions of SNA terms presented in this paper. Words in the box appear in italics the first time they are mentioned in the text. A network is a collection of units of interest that may or may not be connected. The units of interest are normally called *nodes* or *vertices* in physics and mathematics, while they are normally referred to as *actors* in the social sciences. Examples of networks include the world wide web (pages are nodes that can be connected to each other), social networks (who is friends with who), communications or power system networks, and transportation systems (airports are nodes linked by aircraft flights). In a population of farms, each farm would be a node in the network. Nodes may have attributes, such as the type of species they contain, their geographical location, and their size in terms of number of animals, which can be studied in the context of SNA. Nodes are linked to each other through a relationship of some sort. For example, animal movements from farm-to-farm link farms together in a network. When these links between farms are reciprocal or *undirected* they are called *edges*. When these links are unidirectional or *directed*, they are called *arcs* (56). How a relationship is defined between two nodes can allow arcs to be considered as edges. For example, the movements of animals from farm to farm may be viewed as arcs (directed) if we consider the directionality from a *source* farm to a recipient farm. But if we consider this relationship as being a business transaction, then arcs can become edges (reciprocal) in SNA terminology. Arcs and edges may be binary or valued, depending on the characteristics of the relationship under study.

When made of arcs, a network is described as directed, whereas it is considered undirected when made of edges (Fig. 1). As SNA has its roots in graph theory, networks can

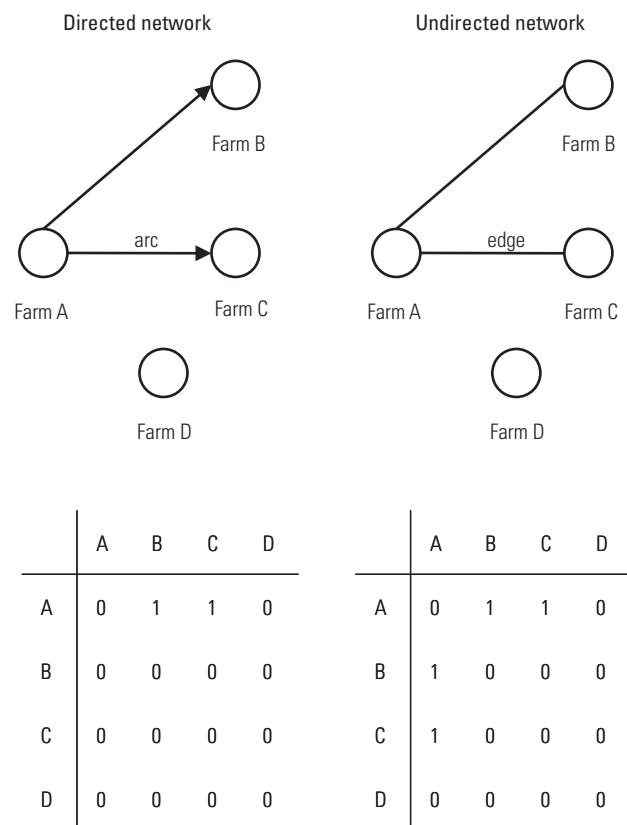


Fig. 1

Graphical representation of directed and undirected networks

In the directed network, farm A shipped animals to farms B and C.

In the undirected network, farm A has a mutual relationship with farms B and C. Below each network is the data it contains in matrix format

be represented either in matrix format or as graphs (Fig. 1). The matrix format allows the execution of a variety of measures and calculations used to provide descriptive statistics of the networks. The size of a network is represented by the total number of nodes present in the network.

The duration of the time period that is explored in order to build the network of relationships will be a critical factor and must be carefully considered at the start of the study. For example, studying livestock movements that took place over a year might yield a very different network than a monthly network. When building a livestock movement network, the unit of concern must be defined. In most cases, a livestock holding (market, farm, or dealer) will be a node, but in some situations an entire village may be a node in the network. Secondly, the description of the movements that occurred among the nodes in a defined time period is required. The data will be arranged so that it is possible to know the source and destination of the movement. Attribute information for each node can also be included in the analysis as well as any weighting given to the links among nodes.

Box 1**Glossary of social network analysis terms interpreted in the context of livestock movement** (adapted from 24)

Name	Definition
Accessible world	The total number of livestock holdings that can be reached directly (in one step) or indirectly (through another holding) (59). Also known as the 'output domain' in social networks analysis (21)
Arc	A directed link between two nodes (56)
Average path length	The shortest path, or geodesic, among two livestock holdings averaged over all pairs of livestock holdings in the network (57)
Betweenness	The frequency with which a livestock holding is in the shortest path between pairs of holdings in the network (27)
Clustering coefficient	If a neighbour is defined as a livestock holding in direct contact with the holding of interest, the clustering coefficient represents the proportion of one's neighbours who are also neighbours of one another (57)
Closeness	The mathematical inverse of farness or the inverse of the average distance between a holding and all the other holdings in a network (56)
Components	Maximally connected subregions of a network in which all pairs of livestock holdings are directly or indirectly linked (49)
Cut-points	Represents a livestock holding which if removed from the network leads to an increase in the number of components, therefore increasing the level of fragmentation in the network (48)
Density	Represents the proportion of links (C) among livestock holdings (k) in the network that are actually present using equation: $C / k(k-1)$ (56)
Directed network	A network in which ties among nodes are represented by arcs (unidirectional; 56)
Edge	An undirected, reciprocal, link between two livestock holdings (56)
Farness	The sum of the shortest distances (not geographical, but path length) from a source livestock holding to all other reachable holdings in the network (18)
Fragmentation	The proportion of pairs of livestock holdings that are unreachable in the network; a path does not exist between them (14)
Geodesic	The shortest path length between two livestock holdings (21)
Giant strong component	The largest strong component in the network (32)
Giant weak component	The largest weak component in the network (32)
Hub	A holding with high in- and out-degree values compared to other holdings in the network
Infection chain	The number of holdings that can be reached directly (one-step) or indirectly (through another holding) by a holding in the network, accounting for the sequence of movements in time (23)
In-degree	Number of individual sources providing animals to a specific livestock holding (56)
k-neighbours	The number of livestock holdings that can be reached in k steps from a specific holding (44)
Measures of centrality	Measures to identify the importance and role of individual livestock holdings in the network such as: degree distributions, betweenness, and farness (56)
Measures of cohesiveness	Measures to determine the level of connectivity in the network such as: density, clustering coefficient and average path length
Node	The unit of interest in social network analysis
Out-degree	Number of individual recipients obtaining animals from a specific livestock holding (56)
Path	A path between farm A and farm C (Fig. 1) is the number of steps required to travel from A to C. In this example, 1 step is required. In a path, livestock holdings (nodes) and links (arcs) cannot be repeated to move from a source to a recipient (21). Also see geodesic and average path length
Random homogeneous mixing	It is a network in which all individuals are equally likely to be infected by contact with an infected individual
Scale-free network	A network in which the out-degree and in-degree distributions fit a power law distribution (Fig. 2b). Allows the presence of hubs in the network (9)
Small-world network	A network characterised by high clustering and short path length (53). In such networks agents tend to spread more rapidly, but ultimately infect fewer individuals, compared to random mixing networks (43)
Source	Livestock holding that has zero in-degree but its out-degree is ≥ 1 . It does not receive shipments from anyone in the network (21)
Strong component	A directed network in which all livestock holdings are mutually accessible by following the direction of the links in the network (49)
Tie	Relationship among nodes in the network
Topology	Also known as network architecture. It is the study of the arrangement or graphing of the elements (livestock holdings, links) of a network and refers to topological properties such as small-world, clustering, path length, and scale-free properties
Undirected network	A network in which ties among nodes are represented by edges (bidirectional; 56)
Weak component	An undirected network in which all livestock holdings are linked, not taking into account the direction of the links (49). But not all farms are reachable from one another if we take into account the direction of the links

Measures of centrality: node-level metrics

One of the main objectives of SNA is to identify nodes that are central, or important in the network. Three measures are used to characterise nodes: node *degree*, *betweenness* and *farness* (56). Each of these measures will be presented as well as examples of how they have been applied in livestock movement studies.

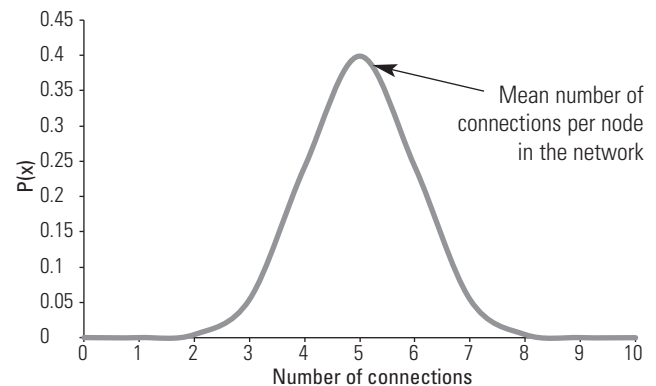
The node degree represents the number of contacts per node. In an *undirected network*, it would represent the number of nodes with which it shares a connection. For example, in Figure 1, farm A would have a degree of 2. In a *directed network*, the degree value is divided into *out-degree*, which is the number of arcs that originate from each node in the network, and *in-degree*, which is the number of arcs that each node receives in the network (56). In Figure 1, farm A would have an out-degree of 2 and an in-degree of 0. Livestock holdings with high out-degree and in-degree values could be considered *hubs* in the network and can be at risk of becoming infected and infecting a large number of other holdings in the network.

The betweenness of a node represents the frequency with which a node is on the shortest 'path' between pairs in the network. This means that a node with high betweenness is important in linking a high number of pairs in the network and its removal, through quarantine during an outbreak or depopulation, can lead to *fragmentation* of the network, which results in a network in which nodes are not as reachable.

Farness of a node expresses how distant topologically, not geographically, it is from all other nodes in the network. Mathematically, it is the inverse of *closeness*, which is a measure of how closely connected a node is to all other nodes in the network. Presumably, a node with a high value of farness should not represent a threat to other nodes in the network when considering the spread of an infectious disease agent.

All these measures help researchers gain an understanding of the importance of individual nodes in the network. Recent SNA studies of livestock movements (12, 18, 23, 35) explored the distributions of out-degree and in-degree values and noted an important level of heterogeneity or variation in the number of contacts per holding. These degree distributions are characterised by a power-law distribution and the networks are characterised as *scale-free* networks (3). The reason for calling these networks scale-free is that the power law distribution does not have a peak and it is characterised by a long tail which gives the high variation in the number of contacts per node (Fig. 2). Barabási (7) translates this into the absence of an average

a) Normal distribution



b) Power law distribution

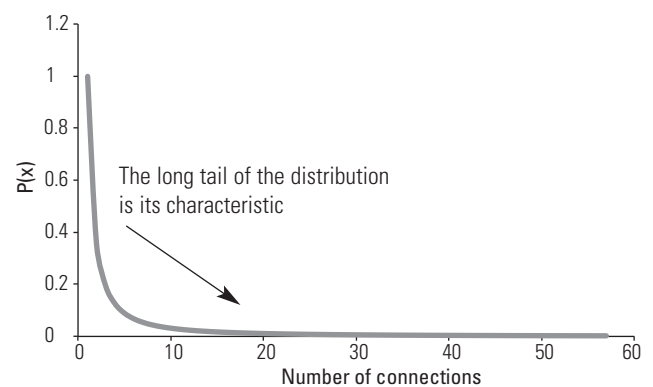


Fig. 2
Comparison of the normal and power law distributions
Based on Dubé *et al.* (24)

node in comparison to the average node found in the Poisson or normal distributions. Because of the presence of hubs in the network, the average node does not represent the typical node in scale-free networks. The presence of the tail in the distribution allows for the presence of hubs in a network, a concept that is not allowed in traditional Poisson-based random networks (2). Various observed networks in the world have been defined as scale-free: the internet (26), the world wide web (3), the power grid in the United States and the movie actors collaboration network (8), the citation network of scientific collaborations (47) and the web of sexual contacts (39). A review on scale-free networks has been published (2).

The implication of scale-free networks for the way in which infectious diseases might spread is important (9). These networks have been shown to display a high level of resiliency to random attacks. This is due to the presence of a very large number of weakly connected nodes in the network. In this situation, the probability of a random attack being directed at a hub is low. However, scale-free networks are susceptible to targeted attacks directed at hubs, which will seriously disturb the structure and functioning of the network. If an infectious disease agent such as the foot and mouth disease (FMD) virus is

introduced randomly in a population of farms, the consequences are not necessarily dramatic; however, if the virus finds its way to a livestock market, a hub in the network, as occurred in the 2001 outbreak in the United Kingdom (UK), a large number of farms can become infected before authorities are aware that the virus is circulating. This can make it hard for authorities to gain control of the spread of the virus and can lead to a large number of infected holdings.

Node-level measures are therefore important to assess the role of various livestock holdings in animal movement networks. Finding holdings with high betweenness will help us understand who can control the flow of animals from one part of the network to another. Robinson and Christley (48) studied the movement of cattle through auction markets in the UK between 2002 and 2004 and found that markets and some farms had high betweenness. The authors identified these holdings as *cut-points* which, if removed from the network, increase the number of *components*, or fragment the network. Identifying attributes of these farms and knowing their geographical location in advance could potentially reduce the spread of an infectious disease agent, because quarantine measures could be introduced on such premises in the event of an outbreak of a highly contagious disease.

Measures of cohesiveness: network-level metrics

These measures assess the level of connectedness of the network as a whole. They include *density*, *fragmentation*, *average path length* and *clustering coefficient (CC)*. The density of the network represents the proportion of links, out of all those that are possible, that are actually present. This value can range from 0 to 1, where 1 would mean that all nodes in a network are connected to each other directly. Fragmentation represents the proportion of pairs in the network that are unreachable, which means that a path does not exist to connect them. In Figure 1, the pair made of farms B and C is unreachable because a path does not link them. Ranging from 0 to 1, fragmentation gives an insight into how an infectious disease agent may spread: the more fragmented a network, the less easy it is for an infection to spread to a large number of holdings. The average path length refers to the average shortest distance (number of steps) among reachable pairs in the network. It has been used in the social sciences to determine how connected the world is, leading to the 'six degrees of separation' concept (42).

The clustering coefficient is an important measure in SNA as it has been used to characterise networks as *small-world*. Watts and Strogatz (57) have defined the CC as the

proportion of one's neighbours in the network who are also neighbours of one another. Small-world networks are characterised by high CCs and short average path length represented by a few long-distance connections that link topologically distant clusters in the network. The implication of small-world networks is that an infection will not only spread within clusters in the network, but it can also reach other clusters, topologically (not necessarily geographically) distant in the network. Therefore, large geographical distances between farms may not prove to be a barrier to infection spread in small-world networks. Various networks have been characterised as small-world: the neural network of the nematode worm *Caenorhabditis elegans*, the power grid in the United States (57), the cattle-movement network in the UK (18) and in Denmark (13), the poultry industry network in New Zealand (40), and the movement of adult milking cows in Ontario, Canada, as part of the Dairy Herd Improvement Program (23).

Finding cohesive sub-groups

In large complex networks, a number of sub-groups can be found which may be of interest to the researcher and there exist various rules to identify sub-groups in a network. Cohesive sub-groups or *components* are maximally connected sub-regions of a network in which all pairs of livestock holdings are directly or indirectly linked (49). Figure 3 provides a representation of components in a directed network such as a livestock movement network. In an undirected network, the component includes all nodes that are mutually reachable. In a directed network, the *strong component* includes all the nodes that are mutually accessible when accounting for the direction of

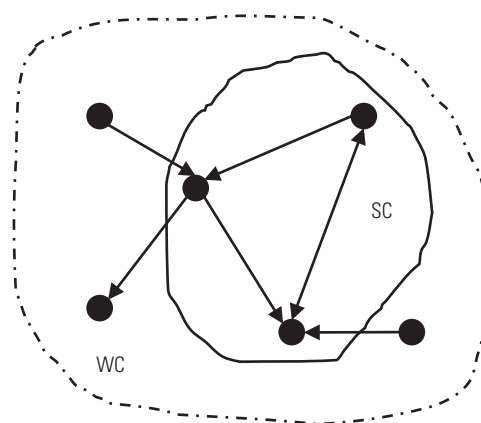


Fig. 3
Strong (SC) and weak components (WC) in a livestock movement network

The SC includes all holdings that can be reached from one another, respecting the direction of the connections. The WC includes all holdings that can be reached from one another disregarding the direction of the connections, considering them reciprocal

the arcs in the network. The *weak component* includes all nodes that are mutually accessible when links are considered undirected or reciprocal. In a livestock movement network this would mean that we would disregard the actual direction of the shipments. In large complex networks, strong and weak components of various sizes can be identified and, usually, single large *giant strong* (GSC) and *giant weak* (GWC) components will emerge.

The size of the components has been used to predict the potential size of epidemics, a topic that will be covered later in this paper. Components have also been used to describe the evolution of the cattle movement network in Great Britain that resulted from movement regulations that were put in place in 2001 (49). The analysis showed a linear increase in the size of the GSC in weekly networks between 2002 and 2004. This increased connectivity was attributed to increased communications among producers in order to maximise movements under the new regulations and an increased reliance on dealers and markets for these movements. Although the new movement regulations were put in place to reduce the potential for large epidemics, the results of the Robinson *et al.* study suggest the regulations had the opposite effect, leading to a self-organising system that has increased the potential size of infectious disease epidemics within the cattle industry.

Use of social network analysis in veterinary epidemiology

The application of SNA in veterinary epidemiology is only recent, with the technique being first presented at a conference in 2002 (60) and the first two papers being published in 2003 (17, 20). Most of the published work comes from the UK and the studies can be classified into three main categories:

- descriptive SNA studies
- retrospective analyses of epidemics
- studies of how the network structure impacts disease control measures and studies used to predict potential epidemic spread.

Most descriptive studies focus on *measures of centrality* and *measures of cohesiveness* (12, 13, 15, 18, 22, 23, 40, 48, 55, 58, 59, 60). Movement networks in these studies have been characterised as scale-free and small-world, suggesting universal properties that can help us understand how infectious disease agents may spread and evaluate how best to intervene in order to stop their spread. An important element of these descriptive studies was their use of the components to predict maximal potential epidemic size, as discussed later in the paper.

Two studies evaluated the networks of livestock movements in the initial phase of the 2001 FMD epidemic in the UK (44, 54). Both identified markets and dealers as key players in the initial transmission of the FMD virus, leading to long-distance transmission and scale-free-type architecture of the network. Ortiz-Pelaez *et al.* (44) also identified a few key farms with high betweenness that played a key role in the early widespread transmission of the virus. This suggests that livestock holdings with high betweenness and a high number of *k-neighbours* (where $k \leq 2$) should be targeted for disease control activities once primary actors such as markets and dealers have been contained. Shirley and Rushton (54) suggested that because of the high level of clustering of farms in the UK landscape, once movement restrictions are in place, a contagious spatially-clustered spread takes place. In order to be effective, control measures must be swift and total at the local level, simultaneously removing infected holdings and all of their possible contacts.

One of the main uses of SNA in veterinary epidemiology has been as a *post hoc* analysis tool for evaluating movements that have taken place and extrapolating to what might happen in the future. This allows us to evaluate the impact of disease control measures according to the structure of the network and to try to predict potential epidemic size following the introduction of a highly contagious disease. For example, the movements of sheep in Great Britain were analysed using SNA (58, 59, 60) in order to understand the impact of the six-day movement standstill implemented in the country following the 2001 FMD epidemic in the UK (59). The results of the study showed that the standstill was ineffective in fragmenting the sheep movement network because most agricultural shows were eight or more days apart. The results demonstrated that, in any future epidemic, restrictions of 7, 14 or 21 days would have the desired effect of fragmenting the network and creating disconnected components in the network, thus reducing the potential spread of an infectious disease agent (59).

The impact of contact tracing has also been explored in the context of movement networks. Tracing during an epidemic is used to identify livestock holdings that might have received infected animals in order to reduce the spread of an infection. Kiss *et al.* (36) evaluated the impact of contact tracing and removal of nodes in theoretical scale-free and random network models. Because an infection spreads faster in scale-free networks, due to the presence of hubs, contact tracing is not able to catch up to the disease. For this reason the authors recommend intelligent tracing, using prior knowledge of who may be most connected in the network and removing these highly connected nodes from the network before they are able to spread an infection. This approach reinforces the need to identify highly connected nodes beforehand and to have a

strategy in place should a highly contagious disease be introduced in the population.

Various measures have been used to estimate potential epidemic size using SNA. Christley *et al.* (18) proposed the size of the GSC and GWC as a measure of the population at risk in the network as a result of the introduction of an infectious disease agent. (18). This approach is based on the fact that components represent a region of maximal connectivity among livestock holdings in a network where all nodes are connected to each other. In theory, all holdings in a component could become infected if an infectious disease agent were introduced into such a region. Kao *et al.* (31) further developed the concept of using the GSC as a measure of the lower bound of maximal potential epidemic size and the GWC as the upper bound of maximal potential epidemic size, a concept that was then used by other researchers (32, 37, 49, 55).

However, Dubé *et al.* (23) suggested that components may not be the best measures to use for estimating maximal epidemic size because there are rules defining these components that may not be appropriate when thinking of livestock movements and infection spread. There is an implied requirement for the strong component that from any given livestock holding that is a source of livestock movements, there must be a path formed by movements in the network that leads back to that source holding. However, infection spread in an epidemic does not follow such a rule. Also, the weak component requires that all relationships between holdings be reciprocal, which is not necessarily true for livestock movements. Therefore, the size of the GSC and GWC may not reflect the size of actual *infection chains* or the *accessible world* of individual holdings that are contained within the network. In fact, Dubé *et al.*

(23) suggest that the infection chain provides the most biologically plausible estimate of potential maximal epidemic size because the technique takes into account the direction of the shipments and their order in time, two important concepts when considering infection spread. In a simple example to illustrate this point consider three farms (A, B and C) where A ships to B, which in turns ships animals to C. In order for farm A to be a risk of infection for farm C, it must have shipped to B before farm B shipped to farm C. This order in time of shipments is not considered in the strong component, yet, it is critical when considering the spread of infections.

Implications for epidemiological modelling

Infection spread will vary greatly depending on the *topology* of the network (53). Therefore, accounting for this topology is critical in order to improve the results of epidemiological modelling studies. A review of networks and epidemic models has been published by Keeling and Eames (33). Standard approaches to epidemiological modelling have been based on assumptions of *random homogeneous mixing* within the population (4, 5, 34), with the use of mass-action or differential equation models such as susceptible-infectious-recovered (SIR) or susceptible-infectious-susceptible (SIS) models, or discrete-time chain binomial models (Reed-Frost; 1). In such models connections are assigned randomly and all individuals have the same low probability of coming into contact with each other. These models may be represented by random graphs (6, 10). In random graphs (Fig. 4), connections

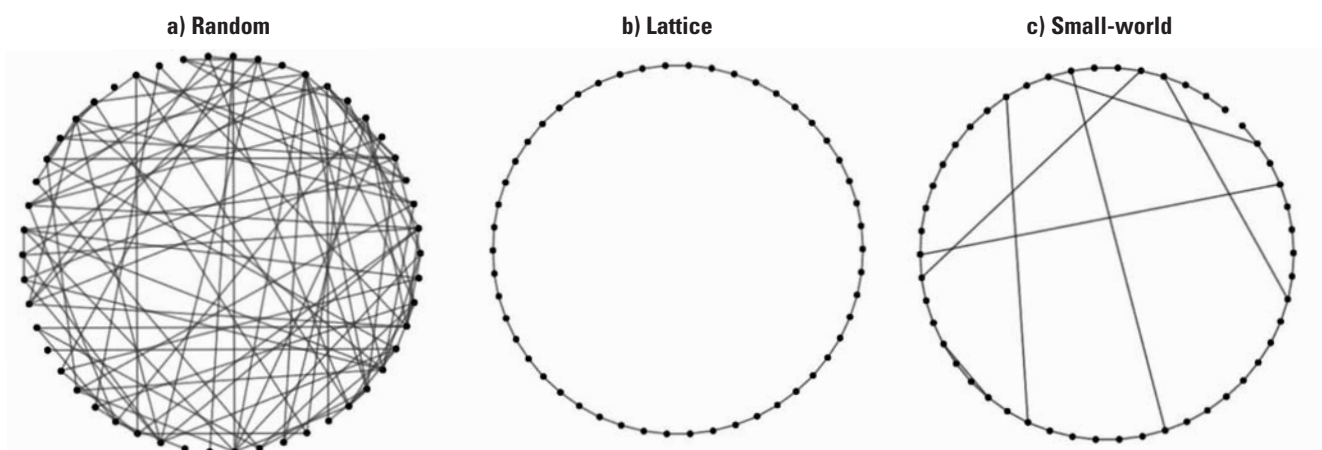


Fig. 4

Representation of some of the different network model structures

- a) The random network in which all connections are assigned randomly, based on the Poisson or normal distribution. It displays very little clustering and a short average path length
- b) The lattice network, which is highly clustered but has a long average path length. All nodes in this network are connected to their neighbours
- c) The small-world network which displays high clustering and a short average path length. Nodes are connected to their neighbours but a few links also connect different clusters in the network, shortening the path length compared to the lattice network

among individuals in the population are based on algorithms, developed by Erdős and Renyi (25), which assign connections randomly based on the average number of connections per node (or degree). The degree distribution in these networks is represented by a Poisson or normal distribution. It is rare to find nodes that have either a very low or a very high number of connections compared to the average node in these types of networks.

Random networks or graphs have been used because of the ease with which they can be generated and analysed. The spatial location of individuals in these networks is not relevant. Because of the random processes used to generate these networks they display very little clustering and the likelihood of being connected to one's neighbour's neighbour is no different than the likelihood of being connected to a random node in the network, they display very little clustering. But they do have a short average path length because connections within them are established randomly, to any part in the network, with the same probability (53). These random network models have been useful in a variety of situations in which accounting for contact heterogeneities is not critical. For example, this type of model is often well suited to within-herd transmission modelling (46).

Refinements and expansions of these standard techniques have been made to account for various heterogeneities as well as spatial and host characteristics. For example, large, state-transition, spatially explicit simulation models have been developed to study the spread of the FMD virus in populations (29, 50). Although these models are able to account for spatial clustering, stochasticity, and host-level factors that influence infection transmission, they do not represent the contact heterogeneity found in observed livestock movement networks.

Another method of network modelling has been to use highly regular lattice structures to connect the nodes in a network (Fig. 4). In such networks, the probability of being connected to one's neighbour's neighbour is much higher than the probability of being connected to a random node in the network. These types of networks have high clustering but a long average path length compared to random networks of similar size, because some regions of the network are topologically remote from others (53). They are often used to represent spatial relationships among nodes (19, 28).

Both random and lattice networks are forms of random homogeneous mixing in which a high level of homogeneity in the number of contacts per node exists; however, they differ in the way in which connections are established, which leads to differences in clustering and average path length (53). Recently, with the characterisation of various observed networks as scale-free and small-world, these models have become less useful for representing observed

networks that are non-homogeneous. The study of the impact of the topology of these networks on infection spread has shown that network characteristics such as clustering, average path length and degree distribution will influence the velocity of spread of an infection, the final size of epidemics, and the impact of control measures (30, 36, 53). The shorter the average path length, or the more connections among nodes in the network, the faster an epidemic can spread (53). Early on in an epidemic, because of the presence of hubs in theoretical scale-free networks, the disease will spread more quickly on this type of network than on random networks of similar size (36, 45). However, once hubs are infected and they have infected their partners, infection then spreads more slowly than in random networks (31). In some cases, the overall epidemic size will be larger in random networks than in scale-free networks because of the rapid depletion of potential contacts early on in epidemics in scale-free networks (35). Clustering (not necessarily geographical or spatial) has also been shown to reduce the size of an epidemic on theoretical highly clustered networks, but it lowers the epidemic threshold, making it easier for diseases to spread (43).

The results of SNA studies have provided us with insights into the topology of networks and the importance of accounting for this topology when considering infection spread. However, few epidemiological models currently account for contact heterogeneity at the time of writing this paper.

Conclusion

Social network analysis is a useful tool for the epidemiologist in understanding the broad consequences of livestock movements. No other approach to date has allowed the study of all the relationships among livestock holdings resulting from these movements. In the future it will be possible to add other types of contacts, such as the movements of dairy tankers or artificial inseminators, in order to increase the detail-level of contact networks on which infections can spread. Therefore, as more information becomes available, SNA will become increasingly useful for studying potential epidemics.

There are limitations to the use and usefulness of SNA. It does require high quality and complete data in order to provide an appropriate understanding of the true networks. In some regions, these data may not be available for financial, political or confidentiality reasons. In such situations, samples have been used and the issue of representativeness of the results, as with any other epidemiological study, must be considered. When data are available, network databases can be very large and can easily overwhelm generic SNA software. Martinez-López *et*

al. (41) suggest the aggregation of records into nodes that could represent municipalities, districts, or countries. These data could then be screened for the identification of areas that would be of interest for further study, as was done by León *et al.* (38).

Social network analysis is a tool for studying potential epidemic spread and identifying central and/or highly connected livestock holdings in a network and can therefore help in planning surveillance, disease control and eradication measures. It also provides data to build and parameterise epidemiological models of infection spread. To date, SNA studies have shown the importance of considering livestock movement network topology, for this topology can have an impact on how quickly and widely

an infection can spread. As a result, epidemiological modellers should assess how they can represent the levels of heterogeneity and network topology of observed livestock movement networks in their models. Keeling and Eames (33) suggest that the only appropriate approach to studying control measures for infectious diseases such as contact tracing and ring vaccination is to use network-based models, and that the ultimate goal should be to develop a set of robust network statistics that allow us to predict epidemic dynamics when the population structure deviates from the random homogeneous mixing framework. ■

Introduction à l'analyse des réseaux et à ses conséquences pour la modélisation de la santé animale

C. Dubé, C. Ribble, D. Kelton & B. McNab

Résumé

Des chercheurs en épidémiologie animale ont récemment appliqué l'analyse des réseaux sociaux à l'étude des mouvements de bétail. On construit un réseau en considérant chaque élevage comme un nœud du réseau et chaque déplacement entre élevages comme un lien entre nœuds. L'analyse des réseaux sociaux permet d'étudier un réseau comme un tout, en envisageant toutes les relations possibles entre les élevages pris par paires. Les élevages déployant le plus d'interactions au sein du réseau sont ainsi identifiés, ce qui s'avère très utile lors des opérations de surveillance et de prévention. D'après ces études, dans plusieurs pays les mouvements d'animaux domestiques tracent des configurations en réseaux qui révèlent une grande hétérogénéité en termes de contacts et de formation de grappes (au plan topologique, et non nécessairement géographique ou spatial) ; la prise en compte de l'architecture de ces réseaux a permis de mieux comprendre comment les infections se propagent. Il serait intéressant d'utiliser les résultats des analyses des réseaux sociaux appliquées aux mouvements d'animaux d'élevage pour construire des modèles épidémiologiques de la propagation des infections en établissant des paramètres appropriés, ce qui permettrait d'améliorer la fiabilité des projections obtenues au moyen de ces modèles.

Mots-clés

Analyse des réseaux – Modélisation épidémiologique – Mouvements d'animaux d'élevage – Propagation d'une infection. ■

Introducción al análisis de redes y sus consecuencias para la elaboración de modelos de enfermedades animales

C. Dubé, C. Ribble, D. Kelton & B. McNab

Resumen

Últimamente, en el ámbito de la epidemiología veterinaria se ha utilizado el análisis de redes sociales para estudiar los movimientos de ganado. Para definir una red se considera que sus nodos son las explotaciones ganaderas, y que los movimientos entre éstas son los vínculos entre los nodos. El análisis de redes sociales permite estudiar la red en su integridad, observando todas las relaciones entre elementos homólogos de las explotaciones. Es posible así detectar centros ganaderos con gran densidad de conexiones, lo que puede ser útil para las actividades de vigilancia y prevención de enfermedades. Las redes de movimientos de ganado observadas en varios países han puesto de relieve un importante nivel de heterogeneidad y agrupamiento (topológico, no necesariamente geográfico o espacial) de los contactos, y la comprensión de la arquitectura de esas redes ha ayudado a entender mejor la forma en que pueden propagarse las enfermedades. Para crear y parametrizar modelos epidemiológicos de la diseminación de enfermedades convendría utilizar los resultados de los análisis de redes sociales aplicados a los movimientos de ganado, lo que daría un mayor grado de fiabilidad a los resultados obtenidos con esos modelos.

Palabras clave

Análisis de redes – Elaboración de modelos epidemiológicos – Movimientos de ganado – Propagación de infecciones.



References

- Abbey H. (1952). – An examination of the Reed-Frost theory of epidemics. *Hum. Biol.*, **243**, 201-233.
- Albert R. & Barabási A.L. (2002). – Statistical mechanics of complex networks. *Rev. mod. Phys.*, **74**, 47-97.
- Albert R., Jeong H. & Barabási A.L. (1999). – Diameter of the World-Wide Web. *Nature*, **401**, 130-131.
- Anderson R.M. (1982). – The population dynamics of infectious diseases: theory and applications. Population and Community Biology. Chapman and Hall, London.
- Anderson R.M. (1991). – Populations and infectious diseases: ecology or epidemiology? *J. anim. Ecol.*, **60**, 1-50.
- Andersson H. (1998). – Limit theorems for a random graph epidemic model. *Ann. appl. Probab.*, **8**, 1331-1349.
- Barabási A.L. (2003). – Linked. Penguin Books, London.
- Barabási A.L. & Albert R. (1999). – Emergence of scaling in random networks. *Science*, **286**, 509-512.
- Barabási A.L. & Bonabeau E. (2003). – Scale-free networks. *Sci. Am.*, **288** (5), 60-69.
- Barbour A. & Mollison D. (1990). – Epidemics and random graphs. In Stochastic processes in epidemic theory (J.-P. Gabriel, C. Lefèvre & P. Picard, eds). Springer, New York, 86-89.
- Bates T.W., Thurmond M.C. & Carpenter T.E. (2001). – Direct and indirect contact rates among beef, dairy, goat, sheep and swine herds in three California counties, with reference to control of potential foot-and-mouth disease transmission. *Am. J. vet. Res.*, **62**, 1121-1129.
- Bigras-Poulin M., Barfod K., Mortensen S. & Greiner M. (2007). – Relationship of trade patterns of the Danish swine industry animal movements network to potential disease spread. *Prev. vet. Med.*, **80**, 143-165.

13. Bigras-Poulin M., Thompson R.A., Chriel M., Mortensen S. & Greiner M. (2006). – Network analysis of Danish cattle industry trade patterns as an evaluation of risk potential for disease spread. *Prev. vet. Med.*, **76**, 11-39.
14. Borgatti S.P., Everett M.G. & Freeman L.C. (1999). – UCINET 6.0 Version 6.17. Natick: Analytic Technologies.
15. Brennan M.L., Kemp R. & Christley R.M. (2008). – Direct and indirect contacts between cattle farms in north-west England. *Prev. vet. Med.*, **84**, 242-260.
16. Christensen J., McNab B., Stryhn H., Dohoo I., Hurnik D. & Kellar J. (2008). – Description of empirical movement data from Canadian swine herds with an application to a disease spread simulation model. *Prev. vet. Med.*, **83**, 170-185.
17. Christley R.M., Pinchbeck G.L., Bowers R.G., Clancy D., French N.P., Bennett R. & Turner J. (2003). – Infection in social networks: using network analysis to identify high-risk individuals. *Am. J. Epidemiol.*, **162**, 1024-1031.
18. Christley R.M., Robinson S.E., Lysons R. & French N.P. (2005). – Network analysis of cattle movement in Great Britain. In Proc. Meeting held in Nairn, Inverness, 30 March to 1 April (D.J. Mellor, A.M. Russell & J.L.N. Wood, eds). Society for Veterinary Epidemiology and Preventive Medicine, 234-244.
19. Cliff A.D. & Ord J.K. (1981). – Spatial processes: models and applications. Pion Press, London.
20. Corner L.A., Pfeiffer D. & Morris R.S. (2003). – Social-network analysis of *Mycobacterium bovis* transmission among captive brushtail possums (*Trichosurus vulpecula*). *Prev. vet. Med.*, **59**, 147-167.
21. De Nooy W., Mrvar A. & Batagelj V. (2005). – Exploratory social network analysis with Pajek. Cambridge University Press, New York.
22. Dent J.E., Kao R.R., Kiss I.Z., Hyder K. & Arnold M. (2008). – Contact structures in the poultry industry in Great Britain: exploring transmission routes for a potential avian influenza virus epidemic. *BMC vet. Res.*, **4**, 27.
23. Dubé C., Ribble C., Kelton D. & McNab B. (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. *Transbound. emerg. Dis.*, **55**, 382-392.
24. Dubé C., Ribble C., Kelton D. & McNab B. (2009). – A review of network analysis terminology and its application to foot-and-mouth disease modelling and policy development. *Transbound. emerg. Dis.*, **56**, 73-85.
25. Erdős P. & Renyi A. (1960). – On the evolution of random graphs. *Magy. Tud. Akad. Mat. Kut. Intéz. Kozl.*, **5**, 17-61.
26. Faloutsos M., Faloutsos P. & Faloutsos C. (1999). – On power-law relationships of the Internet topology. In Proc. Conference of the Special Interest Group on Data Communication: SIGCOMM '99. Applications, technologies, architectures, and protocols for computer communication. Association of Computing Machinery, New York, 251-262. Available at: doi.acm.org/10.1145/316188.316229.
27. Freeman L.C. (1978/1979). – Centrality in social networks: conceptual clarification. *Soc. Networks*, **1**, 215-239.
28. Hägerstrand T. (1969). – Innovation diffusion as a spatial process. University of Chicago Press, Chicago.
29. Harvey N., Reeves A., Schoenbaum M.A., Zagmutt-Vergara F.J., Dubé C., Hill A.E., Corso B.A., McNab W.B., Cartwright C.I. & Salman M.D. (2007). – The North American Animal Disease Spread Model: a simulation model to assist decision making in evaluating animal disease incursions. *Prev. vet. Med.*, **82**, 176-197.
30. Huerta R. & Tsimring L.S. (2002). – Contact tracing and epidemics control in social networks. *Phys. Rev. E*, **66**, 056115.
31. Kao R.R., Danon L., Green D.M. & Kiss I.Z. (2006). – Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. *Proc. roy. Soc. Lond., B, biol. Sci.*, **273**, 1999-2007.
32. Kao R.R., Green D.M., Johnson J. & Kiss I.Z. (2007). – Disease dynamics over very different time-scales: foot-and-mouth disease and scrapie on the network of livestock movements in the UK. *J. roy. Soc., Interface*, **4**, 907-916.
33. Keeling M.J. & Eames K.T. (2005). – Networks and epidemic models. *J. roy. Soc., Interface*, **2**, 295-307.
34. Kermack W.O. & McKendrick A.G. (1927). – A contribution to the mathematical theory of epidemics. *Proc. roy. Soc. Lond., B, biol. Sci.*, **115**, 700-721.
35. Kiss I.Z., Green D.M. & Kao R.R. (2006). – The effect of contact heterogeneity and multiple routes of transmission on final epidemic size. *Math. Biosci.*, **203**, 124-136.
36. Kiss I.Z., Green D.M. & Kao R.R. (2006). – Infectious disease control using contact tracing in random and scale-free networks. *J. roy. Soc., Interface*, **3**, 55-62.
37. Kiss I.Z., Green D.M. & Kao R.R. (2006). – The network of sheep movements within Great Britain: network properties and their implications for infectious disease spread. *J. roy. Soc., Interface*, **3**, 669-677.
38. León E.A., Stevenson M.A., Duffy S.J., Ledesma M. & Morris R.S. (2006). – A description of cattle movements in two departments in Buenos Aires province, Argentina. *Prev. vet. Med.*, **76**, 109-120.
39. Liljeros F., Edling C.R., Amaral L.A.N., Stanley H.E. & Aberg Y. (2001). – The web of human sexual contacts. *Nature*, **411**, 907-908.

40. Lockhart C.Y., Stevenson M.A., Rawdon T.G., Gerber N. & French N.P. (2010). – Patterns of contact within the New Zealand poultry industry. *Prev. vet. Med.*, **95**, 258-266.
 41. Martinez-Lopez B., Perez A.M. & Sánchez-Vizcaíno J.M. (2009). – Social network analysis. Review of general concepts and use in preventive veterinary medicine. *Transbound. emerg. Dis.*, **56**, 109-120.
 42. Milgram S. (1967). – The small world problem. *Psych. Today*, **2**, 60-67.
 43. Newman M.E.J. (2003). – Properties of highly clustered networks. *Phys. Rev. E*, **68**, 026121-1–026121-4.
 44. Ortiz-Pelaez A., Pfeiffer D.U., Soares-Magalhães R.J. & Guitian F.J. (2006). – Use of social network analysis to characterize the pattern of animal movements in the initial phases of the 2001 foot and mouth disease (FMD) epidemic in the UK. *Prev. vet. Med.*, **76**, 40-55.
 45. Pastor-Satorras R. & Vespignani A. (2002). – Epidemic dynamics in finite size scale-free networks. *Phys. Rev. E.*, **65**, 035108(R).
 46. Perez A.M., Ward M.P., Charmandarián A. & Ritacco V. (2002). – Simulation model of within-herd transmission of bovine tuberculosis in Argentina dairy herds. *Prev. vet. Med.*, **54**, 361-372.
 47. Redner S. (1998). – How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J., B*, **4**, 131-134.
 48. Robinson S.E. & Christley R.M. (2007). – Exploring the role of auction markets in cattle movements within Great Britain. *Prev. vet. Med.*, **14**, 21-37.
 49. Robinson S.E., Everett M.G. & Christley R.M. (2007). – Recent network evolution increases the potential for large epidemics in the British cattle population. *J. roy. Soc., Interface*, **4**, 587-762.
 50. Sanson R.L. (1993). – The development of a decision support system for an animal disease emergency, Unpublished PhD thesis, Massey University, Palmerston North, New Zealand.
 51. Sanson R.L. (2005). – A survey to investigate movements off sheep and cattle farms in New Zealand, with reference to the potential transmission of foot-and-mouth disease. *N.Z. vet. J.*, **53**, 223-233.
 52. Sanson R.L., Struthers G., King P., Weston J.F. & Morris R.S. (1993). – The potential extent of foot-and-mouth disease: a study of the movements of animals and materials in Southland, New Zealand. *N.Z. vet. J.*, **41**, 21-28.
 53. Shirley M.D.F. & Rushton S.P. (2005). – The impacts of network topology on disease spread. *Ecol. Complex.*, **2**, 287-299.
 54. Shirley M.D.F. & Rushton S.P. (2005). – Where diseases and networks collide: lessons to be learnt from a study of the 2001 foot-and-mouth epidemic. *Epidemiol. Infect.*, **133**, 1023-1032.
 55. Volkova V.V., Howey R., Savill N.J. & Woolhouse M.E.J. (2010). – Sheep movements and the transmission of infectious diseases. *PLoS ONE*, **5**, e11185.
 56. Wasserman S. & Faust K. (1994). – Social network analysis: methods and applications. Cambridge University Press, New York.
 57. Watts D.J. & Strogatz S.H. (1998). – Collective dynamics of 'small-world' networks. *Nature*, **393**, 440-442.
 58. Webb C.R. (2005). – Farm animal networks: unravelling the contact structure of the British sheep population. *Prev. vet. Med.*, **68**, 3-17.
 59. Webb C.R. (2006). – Investigating the potential spread of infectious diseases of sheep via agricultural shows in Great Britain. *Epidemiol. Infect.*, **134**, 31-40.
 60. Webb C.R. & Sauter-Louis C. (2002). – Investigations into the contact structure of the British Sheep population. *In Proc. Meeting of the Society for Veterinary Epidemiology and Preventive Medicine*, 3rd to 5th April, Cambridge, 10-20.
-