Network analysis of swine shipments in Ontario, Canada, to support disease spread modelling and risk-based disease management

S.Dorjee[1], C. W. Revie[1], Z. Poljak[2], W.B. McNab[3], J. Sanchez[1]

[1]CVER, Department of Health Management, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, Prince Edward Island, Canada.

[2]Department of Population Medicine, Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada.

[3] Animal Health & Welfare Branch, Ontario Ministry of Agriculture, Food and Rural Affairs, Guelph, Ontario, Canada.

Correspondence:

S. Dorjee

Department of Health Management, Atlantic Veterinary College,

University of Prince Edward Island

550 University Avenue, Charlottetown, PE, Canada C1A 4P3

Tel.: +1 902 566 0969; Fax: +32 9 264 7534

E-mail: sdorjee@upei.ca

Abstract

Understanding contact networks are important for modelling and managing the spread and control of communicable diseases in populations. This study characterizes the swine shipment network of a multi-site production system in southwestern Ontario, Canada. Data were extracted from a company's database listing swine shipments among 251 swine farms, including 20 sow, 69 nursery and 162 finishing farms, for the two-year period of 2006 to 2007. Several network metrics were generated. The number of shipments per week between pairs of farms ranged from 1 to 6. The medians (and ranges) of out-degree were: sow 6 (1–21), nursery 8 (0–25), and finishing 0 (0–4), over the entire two-year study period. Corresponding estimates for in-degree of nursery and finishing farms were 3 (0–9) and 3 (0–12) respectively. Outgoing and incoming infection chains (OIC and IIC), were also measured. The medians (ranges) of the monthly OIC and IIC were 0 (0–8) and 0 (0–6) respectively, with very similar measures observed for intervals of two-weeks. Nursery farms exhibited high measures of centrality. This indicates that they pose greater risks of disease spread in the network. Therefore, they should be given a high priority for disease prevention and control measures affecting all age groups alike. The network demonstrated scale-free and small-world topologies as observed in other livestock shipment studies. This heterogeneity in contacts among farm types and network topologies should be incorporated in simulation models to improve their validity. In conclusion, this study provided useful epidemiological information and parameters for the control and modelling of disease spread among swine farms, for the first time from Ontario, Canada.

Keywords: Network analysis; swine; pigs; shipments/movement; modelling parameters

Introduction

The shipment of animals between farms is one of the main mechanisms of spread of infectious disease among livestock (Gibbens et al., 2001; Mansley et al., 2003; Kiss et al., 2006a; Ortiz-Pelaez et al., 2006; Poljak et al., 2008). Recently, network analyses have been applied to describe and quantify animal shipments between farms, and to study their implications for disease spread and control (Bigras-Poulin et al., 2006; Kiss et al., 2006b; Ortiz-Pelaez et al., 2006; Bigras-Poulin et al., 2007; Dubé et al., 2008; Kiss et al., 2008; Martinez-Lopez et al., 2009a; Vernon and Keeling, 2009; Lockhart et al., 2010; Volkova et al., 2010; Nöremark et al., 2011). Analyses of animal shipment networks help identify heterogeneity in contact frequency and the connectedness of the network. They also identify whether these contacts are permanent (for example, through trade agreements between farmers) or differ over time or distance. Knowledge of these network characteristics is useful to assess disease spread in a population (Kiss et al., 2006a; Martinez-Lopez et al., 2009b; Dubé et al., 2011; Nöremark et al., 2011). Furthermore, it can provide guidance for targeted tracing, surveillance and disease control measures. Farms with high *out-degree* (number of off-farm links) can act as a key source of disease spread, whereas farms with high *in-degree* (number of onto-farm links) are at relatively greater risk of disease introduction since they receive shipments from many farms. Farms with both high out-degree and in-degree can act as hubs for disease spread in a network. For example, once a disease infects so called hub farms, it can spread rapidly in a scale-free network (Shirley and Rushton, 2005b; Ortiz-Pelaez et al., 2006). Strategically applying control measures at such hubs (including quarantine, vaccination or stamping out) will be more effective than similar control measures applied randomly. For risk-based surveillance, farms with low in-degree but high out-degree (low probability but high consequences) can be targeted to mitigate high impact of exotic disease introduction (Cameron, 2012). Whereas farms with high in-degree (high probability of infection) can be targeted for high surveillance sensitivity and demonstration of freedom from disease (Cameron, 2012; Frossling et al., 2012). Disease spread may be relatively slower in a small-world network, but it can spread to topologically more distant clusters and facilitate more persistent infection in the population, than in random or scale-free networks (Rahmandad and Sterman, 2008; Rahmandad et al., 2011).

Network *component* analyses (where components are more highly connected sub-regions of a network), have been used to estimate the likely lower and upper limits of epidemic-sizes (Bigras-Poulin et al., 2006; Kao et al., 2006; Kiss et al., 2006b; Robinson et al., 2007). However, Dubé et al. (2008) argued that "infection chain" measures (which take into consideration the temporal sequence of livestock shipments between farms) may provide better estimates of a potential epidemic size. Furthermore, understanding animal movement networks can support regionalization and compartmentalization approaches to disease control, as encouraged by the World Organization for Animal Health (OIE). This would allow the identification of groups of regions or farms that behave as a single epidemiological unit (referred to as a compartment) in terms of risk for disease transmission, as well as risk-free groups with no links to infected groups. This can facilitate resumption or continuation of trade from disease-free regions, or even from risk-free herds within an infected region (Bigras-Poulin et al., 2007; Martinez-Lopez et al., 2009b).

Knowledge of contact networks facilitates more realism in simulation modelling of the spread and control of disease in human (Eubank et al., 2004; Longini et al., 2005; Aparicio and Pascual, 2007; Hsu and Shih, 2010; Rahmandad et al., 2011), and animal populations (Green et al., 2006; Kiss et al., 2006a, 2008; Sharkey et al., 2008; Vernon and Keeling, 2009; Dubé et al., 2011). Examples of input parameters required for such models include: the number of contacts each farm has, topologies of contact networks, frequency and size of shipments, as well as movement distances between farms (Morris et al., 2001; Garner and Beckett, 2005; Harvey et al., 2007).

Pork production is a major industry in Canada. Accordingly, an outbreak of a highly contagious disease among Canadian swine would have a significant negative socioeconomic impact. Analyses of swine shipment networks have been reported from several jurisdictions, but not yet from Canada (Bigras-Poulin et al., 2007; Martinez-Lopez et al., 2009a; Nöremark et al., 2011; Rautureau et al., 2012; Buttner et al., 2013). Therefore, the objectives of this study were to generate farm and network level contact measures, from the swine industry of the Province of Ontario, Canada, that are important for the development of simulation models and to support disease management strategies.

Materials and Methods

Study population

Data used in this study described shipments of pigs among 251 farms (20 sow, 69 nursery and 162 finishing farms), from January 2006 to December 2007. The study data were provided by a large swine health and management company, which serviced the above described farms, located across 15 counties in southwestern Ontario, Canada. Sow farms comprised a mixture of farrow-to-wean and farrow-to-finish operations. Nursery farms consisted of a mixture of farms rearing pigs of approximately three to 10 weeks of age (nursery) and approximately three weeks to market age (wean-to-finish). Finishing farms housed pigs at approximately 10 weeks to market age. Information recorded in the database included unique identity numbers (IDs) of the farms, owners, managers, and of batches of swine managed and moved as groups. In addition, shipment dates, IDs of source and destination farms, production type, number of animals on the farms, the number of animals in shipments, and types of animals shipped were also recorded in the database. The management company did not own any of the farms but provided health and management services, including the maintenance of swine shipment data for the swine producers. The farms were owned by 33 different owners with some owners owning multiple farms. Each farm was located on a unique land parcel and was managed by a different farm manager. Shipments of pigs from one farm to another were not limited to farms belonging to a same owner. Not all farms located in the area were serviced by this one company. Therefore, the data did not include all swine shipments during the study period, among all farms in the 15 counties. However, the data were considered to be representative of the Ontario swine industry because the vast majority of the swine farms in the province employ intensive multi-site production systems, similar to the farms managed by this company. The emphasis of this study was the network of swine movements influencing spread of disease between farms. Accordingly, a detailed analysis of the network of swine shipments from farms to abattoirs was not included.

Descriptive statistics

Statistics were generated describing the number of swine-shipments per week between different farm types, for the two-year period of 2006 and 2007. Also, distributions of the sizes of farms and shipments were generated. Farm size represented the number of sows (excluding piglets), nursery or finishing pigs that a farm can accommodate. The median value for shipment size was used in cases of multiple shipments between any specific pair of study farms.

Network analysis

Patterns of swine shipments among farms were characterized using network analysis (Wasserman and Faust, 1994; Newman, 2003). A network consists of collections of *nodes* or *vertices* (the unit of interest) which may or may not be connected to others by *links* (also known as *edges or arcs).* In this study, nodes represented farms and links represented shipments of pigs between source and destination farms. Links can be treated as weighted or binary depending on whether or not multiple shipments between pairs of farms are taken into consideration. Each shipment event included a batch of one or more pigs from a source to a destination farms. All links were treated as directed (*arcs*), involving directional shipments of pigs between farms.

Contacts (shipments) between the 20 sow, 68 nursery and 157 finishing farms were analyzed using a one-mode directed binary network matrix. One nursery and five finishing farms that received shipments before January 2006 but shipped pigs to abattoirs only during the study period were excluded from this network. A one-mode network represents the links between the same set of entities (the row and column elements in the matrix relate to the same set of farms). In the binary network links were assigned a value of one or zero based on whether or not there was at least one shipment of pigs between a pair of farms. The intensity of off-farm (out-degree) and onto-farm (in-degree) shipments between pairs of farms were assessed by treating links as weighted, taking into consideration the number of shipments of animals between pairs of farms.

The network was characterized in terms of: (i) Network size: number of nodes, number of directed links; (ii) Farm-level centrality measures: out-degree, in-degree, betweenness, eigenvector, including normalized indices of these measures; (iii) Network-level measures: density, centralization indices of out-degree, in-degree, and betweenness, diameter, geodesic distance, clustering coefficient and fragmentation index (Wasserman and Faust, 1994). Descriptions of the network terminology used in this paper are outlined in Table S1 of supplementary tables, and were based on the definitions adopted by Dubé et al., (2009), Martinez-Lopez et al., (2009b) and Newman, (2003).

Infection chain analysis

It has been proposed that measures referred to as the *outgoing infection chain (*OIC*)* and the *incoming infection chain (*IIC*)* are particularly relevant when considering the epidemiology of disease spread and control measures (Dubé *et al.*, 2008; Nöremark *et al.*, 2011). Therefore these measures were assessed in this study. OIC and IIC count all direct and indirect (contacts through other farms) off-farm and onto-farm contacts respectively, accounting for the time sequence of shipments. These measures were generated for each two-week and one month periods, along with corresponding in-degree and out-degree measures for the same time intervals. These intervals were chosen based on the plausible range of infectious periods of most common infectious diseases and plausible lengths of a “silent phase”, after an outbreak has started but before official detection occurs. For instance, an outbreak detection delay up to three weeks for foot-and-mouth disease has been reported in the literature (Keeling et al., 2001). These measures were generated using the EpiContactTrace package (version 0.6.9, http://cran.r-project.org/web/packages/EpiContactTrace) of R software version 2.15.1 (R Foundation; http://www.r-project.org).

Scale-free and small-world topologies

Since many networks of livestock shipments have been observed to exhibit scale-free or 'small-world' topologies (Christley and French, 2003; Webb, 2005; Bigras-Poulin et al., 2006; Kiss et al., 2006b; Webb, 2006; Bigras-Poulin et al., 2007; Lockhart et al., 2010), the network in this study was also examined for these properties. A scale-free network is characterized by a right skewed, long-tailed, power-law distribution of the number of links (degrees) to nodes, where a large number of nodes have a few links, but a few nodes have relatively large numbers of links. Such networks are hypothesized to evolve through preferential attachment (Barabasi and Albert, 1999; Barabasi and Bonabeau, 2003). Accordingly, power-law ($P(k) \sim k^{-\gamma}$) distributions were fitted to out-degree, in-degree and total degree data, using statistical approaches described by Clauset et al., (2009). Specifically, γ and k were estimated by fitting the power law distributions to the data using maximum likelihood methods and goodness-of-fit tests based on the Kolmogorov-Smirnov statistics, where k was based on a k_{min} recommended by Clauset et al. (2009). The power-law distribution fitting and hypothesis testing were conducted using R software, including the routines: plfit.R, ConfidenceIntervals.R and GoodnessOfFit.R (available at: https://sites.google.com/site/beyondmicrofoundationscoderepo/home/r-code-repository/power-laws).

Small-world networks are characterized by clusters of nodes that are connected to each other through a few long range links. They tend to have a relatively high clustering coefficient and shorter geodesic distance than an equivalent size random network. As there is no formal statistical test for detecting small-world characteristics in a network, the clustering coefficient and geodesic distance of the observed network were compared with a set of randomly generated networks using the equivalent number of nodes and links (Newman, 2003; Opsahl and Panzarasa, 2009). The observed network was compared with 500 Erdös-Renyi random graphs generated using the same numbers of farms and links. In the literature, a network demonstrating at least a 6-fold increase in the clustering coefficient, in comparison to the analogous random network, has been classified as small-world (Watts and Strogatz, 1998; Lockhart et al., 2010).

Statistical analyses

Differences in the sizes of farms and shipments among three farm types were compared using a Kruskal-Wallis (KW test) with Bonferroni *post hoc* adjustment for multiple comparisons. A critical value of the *post hoc* adjustment was calculated as $\alpha/(k*(k-1))$, where $\alpha=0.05$ and k is the number of groups compared. Associations between means of various network measures, including out-degree, in-degree, betweenness, and eigenvector with farm type, farm or shipment sizes, were compared using a network adapter ANOVA and t-test (for pairwise comparisons), that are based on bootstrapping and permutation approaches (Hanneman and M, 2005). These methods were used because the network measures violate the random sampling and independence assumptions, and are robust to violation of assumptions of normality and equal variance of residuals. For comparisons, farm sizes were categorized into four groups (A = 250 to 999; B = 1,000 to 1299; C = 1,300 to 2,099; and D = 2,100 to 4,000) based on 25th, 50th, and 75th percentile distribution. The records of farm sizes were missing for three sow farms and two each of the nursery and finishing farms. They were estimated for these farms using a multivariate normal regression imputation method based on their respective median shipment size, out-degree, in-degree, and betweenness values. The median of 100 imputed values was used for each of these seven farms. Similarly, sizes of shipments were categorized into four groups as: A = 1 to 54; B = 55 to 100; C = 101 to 208; and D = 209 to 2,598. The numbers of relevant links identified through infection chain analysis, as compared to the number of links identified by simple degree counts, were assessed using the same two-week and monthly time intervals applied to the infection chain analysis. Differences in two-week and monthly OIC vs. out-degree and IIC vs. in-degree measures were compared among the three farm types as well as the four groups of farm size or shipment size, using the KW test with Bonferroni *post hoc* adjustment described above. In addition, the distributions of both OIC and out-degree, and IIC and in-degree, were compared (for each of the two-week and monthly time intervals), using the KS goodness-of-fit test, because this test takes into consideration the whole distribution.

Cluster analysis

Cluster analysis was used to examine whether there were distinct groups of farms based on the farm and shipment sizes, together with key network measures (out-degree, in-degree, betweenness and eigenvector measures). Agglomerative hierarchical cluster analysis using the weighted average linkage method with a stopping rule based on the Duda and Hart index as described by Everitt et al., (2011) was used for this analysis. A Euclidean distance measure with all variables rank standardized was used for the analysis.

All network and statistical analyses, unless specified previously, were carried out using UCInet v6.360 for Windows (Borgatti et al., 2002) and Stata version 11.2 (StataCorp. 2009. Stata

Statistical Software: Release 11. College Station, TX: StataCorp LP), respectively. For multiple comparisons using Kruskal-Wallis test, a kwallis2 package in Stata (http://fmwww.bc.edu/RePEc/bocode/k) was used. The significance of all statistical tests was assessed at the 5% significance level.

Results

Descriptive statistics

Descriptive statistics of swine shipments between farms are summarized in Table 1. Shipments from sow to nursery farms accounted for 64% (3,690/5,760) , while shipments from nursery to finishing farms accounted for 27% (1,527/5,760) of movements. Each farm shipped pigs to 1–4 other individual farms per week. The number of shipments per week between pairs of farms ranged from 1–6. Limited number of shipments between farms of the same type (nursery or finishing) were also observed. Considerable variations in the frequency of weekly and monthly shipments were noted within each year and between the two years (Figure S1 of supplementary materials). The medians (ranges) of farm sizes by farm type were: sow 1,200 (500–2,700), nursery 2,100 (450–4,000) and finishing 1,060 (250–2760). The farm sizes varied significantly by farm types ($P<0.001$; Figure S2). The distribution of shipment size also varied widely with an overall median (range) of 101 (1–2,930) pigs per shipment (Table S2). A significant difference in the size of shipments among the three farm types was observed with larger size shipments from nursery to finishing farms ($P<0.001$). Furthermore, a positive association between farm size groups and shipment size groups was evident (Pearson’s chi-squared test P value of <0.001).

Network of swine shipments among farms

The network consisted of 245 farms (nodes) with 810 links. A total of 5,760 shipments was observed from 147 individual sources to 211 individual destination farms. Detailed farm and network level measures are presented in Table 2. The overall difference in the out-degree among the three farm types was significant (*P*<0.001). Sow and nursery farms shipped pigs to a median of six nursery and eight finishing farms respectively (with the exception of limited shipments to other nursery farms) (Table 2). The difference in out-degrees between these two farm types was not statistically significant (when assessed against a *post hoc* test's critical value). Except for the 63 finishing farms that shipped pigs to 1–4 other finishing farms, most finishing farms (94) shipped pigs only to slaughterhouses for processing. Since those shipments to slaughterhouses were excluded from the analysis, most finishing farms had out-degrees of zero in this network of farm-to-farm animal shipments. Similarly, three nursery farms were wean-to-finish farms (sent shipments off-farms only to abattoirs) and therefore had out-degrees of zero for this network. Significant positive associations between out-degree and farm size groups and shipment size groups were evident (*P*<0.001; Figure S3).

The difference of in-degree distributions between nursery and finishing farms was not significant (when assessed against a *post hoc* test's critical value). In-degree values presented a borderline signficant association with farm size groups (*P* = 0.055), but they had a significant association with the shipment size groups (*P* <0.001; Figure S4). None of the sow, eight nursery and six finishing farms in this data received any shipments (in-degress of zero). Most of these farms sourced replacements from within their own farms, but some may have had other sources who were not clients of the swine management company involved in this study and were therefore not included in the study data.

Yearly network measures were also estimated (but not provided in Table 2). The overall median (range) of out-degree values for both the yearly networks was 0 (0–15), and the in-degree values were 2 (0–7) and 2 (0–8) for 2006 and 2007, respectively. Sow and nursery farms had similar values of out-degree with an overall median (range) of 4 (0–5) for the yearly networks (not including shipments from finishing farms to abattoirs). The in-degree values for the nursery and finishing farms were similar, with an overall median (range) of 2 (0–8). The medians (ranges) of the overall total-degree of 2006 and 2007 yearly networks were 4 (1–17) and 4 (1–18), respectively.

Only nursery and finishing farms had betweenness centrality scores greater than zero. In particular, four nursery farms had the highest betweenness score of 93 (farm K), 102 (farm J), 104 (farm D), and 106 (farm G) (Figure 1(b)). These farms also had relatively high out-degree, in-degree and eigenvector scores (Figure 1(a) and (b)). The betweenness scores of nursery farms were significantly higher than for finishing farms ($P<0.001$). Betweenness scores were significantly associated with farm size and shipment size groups ($P<0.001$; Figure S5).

No significant differences in eigenvector scores among farm types or farm size groups were observed (P value of 0.121 and 0.163 respectively; Figure S6(a)). However, eigenvector scores differed significantly among shipment size groups ($P<0.001$; Figure S6(b)). A plot of eigenvector and betweenness scores shown in Figure 1(b) identified key farms with high scores of these measures, indicating their influential role in the flow of swine shipments among the farms.

Cluster analysis

Cluster analysis identified five clusters (Table S3). Cluster #1 included a range of farm characteristics. Cluster #2 consisted primarily of finishing farms with high in-degree, small farm and shipment sizes, and the lowest out-degree, betweenness, and eigenvector scores. Cluster #3 consisted of an equal proportion of sow and nursery farms with large farm size, intermediate values of shipment size, out-degree, betweenness and eigenvector scores, and lowest in-degree. Cluster #4 consisted of only nursery farms with the highest out-degree, betweenness and eigenvector scores and largest farm and shipment sizes, and relatively high in-degree (see Figure 1(a) and (b)). Cluster #5 consisted of a single nursery farm with the largest shipment size and larger farm sizes, but with lowest network measures.

Infection chain analysis

The medians for each of the OIC, IIC, out-degree and in-degree measures, were not significantly different within each measure-type, based on two-week vs. one-month time intervals. However, the maximum values of out-degree, IIC and OIC, were higher for monthly vs. two-week intervals. Accordingly, for brevity, only results for the monthly network measures are presented here. The overall medians (ranges) of monthly measures were: out-degree 0 (0–7), OIC 0 (0–8), in-degree 0 (0–5), and IIC 0 (0–6). These measures differed significantly among farm types ($P<0.001$). Sow farms had relatively higher out-degree and OIC with median (range) values of 2 (0–7) and 3 (0–8) respectively, and zero in-degree and IIC. The median values of all these measures were zero for both nursery and finishing farms. However, the nursery farms had relatively higher ranges of out-degree (0–5) and the OIC (0–6) than the finishing farms (out-degree of 0–2; OIC of 0–3). While the nurseries had higher in-degree range (0–5) than finishing farms (0–3), IIC range was lower for nursery (0–5) than finishing farms (0–6). The percentages of three farm types likely to have ranges of monthly OIC and IIC are shown in Figure 2(a) and (b) respectively. It shows that 88% of the sow farms, 36% of the nursery and 4% of the finishing farms shipped pigs to at least one farm in a month (Figure 2(a)). Similarly, 39% of the nursery and 21% of the finishing farms received pigs from at least one farm in a month (Figure 3(b)). No differences in these measures were observed between 2006 and 2007 years (($P>0.05$). Significant positive associations between monthly OICs or IIC measures with farm size groups or shipment size groups were evident (all $P<0.0001$).

Scale-free topology

The out-degree and in-degree distributions had power law scaling parameters (γ) of 1.97 (95% CI 1.82 – 2.14) and 4.34 (95% CI 3.65–5.15) for farms with the out- and in-degree >2 and >6, respectively. There were 147 and 79 farms above these threshold limits, respectively. The total degree distribution had a power law scaling exponent (γ) of 2.69 (95% CI 2.40–3.02) for farms with the degree >6 (consisted of 116 farms). The KS test failed to reject the power law model as a plausible model for out-degree and total degree ($P>0.05$). However this test was significant for the in-degree distribution ($P<0.001$).

Network-level measures

In general, the network density and clustering coefficient were low. The high fragmentation index indicated that the proportion of unreachable pairs of farms was high in the network. The network's out-degree, in-degree and betweenness centralization indices were also low, illustrating a relatively low reliance or concentration of off- and onto-farms shipments from/to a few nodal farms at the macro-level of the entire network. A median geodesic distance of two indicates the presence of only one farm on the most efficient pathway between any two farms in general.

Small-world topology

The overall median geodesic distance of 500 simulated random graphs of equivalent network size of the observed network was five with the overall median (5th and 95th percentiles) of clustering coefficient of 0.013 (0.011–0.012). The relatively shorter geodesic distance (two) and larger clustering coefficient (0.09) of the observed network compared with random graphs of equivalent size demonstrated small-world topology of the swine shipment network. A network graph showing farms, where the size of the nodes was made proportional to the total degree is shown in S7.

Discussion

This study described the network characteristics of swine shipments among 245 farms, in 15 counties, in the province of Ontario, Canada. As a non-random sample, the study data were not fully representative of all swine farms in Ontario. However, it did include swine shipment events over a two-year period, from 6.2% (251/4070) of all swine farms in Ontario (Statistics Canada, 2007). The majority of Ontario swine farms participate in multi-site production systems, where individual farms specialize in a specific production type (breeder, farrow-to-wean, nursery, or finishing).. Most of these farms also operate under a single integrator or on the basis of fairly permanent trade partnerships, where a farrowing farm supplies a specific, relatively fixed set of nursery farms, etc. Thus data obtained for this study described a sub-network, within the overall network of swine shipments in Ontario. Considering this farming structure, we can assume that estimates of farm-level centrality measures of this network represent reasonable estimates of the typical commercial multi-site production systems in Ontario. Furthermore, since these data captured all swine shipments of participating farms, including those shipments to and from farms outside the enrolment of this company (except for the supply of replacement sows), the magnitude of bias in the farm-level measures are likely small. However, network level measures observed in this study, such as density and fragmentation indices will be subject to some degree of bias, because such measures depend on the total number of farms present in the network.

This study provided useful preliminary information on the characteristics of multi-site commercial swine shipment networks in Ontario. It illustrated the types of farms that tend to be more centrally connected and could thus guide the prioritization of control measures, and provide some indication of the potential size of epidemics in such networks. It also provided preliminary parameters required for modelling disease spread among swine farms, such as contact frequencies, the number of links between farms, and the structure of network topologies. As a caution, the network statistics observed during normal trading may change once quarantines or movement controls are implemented during the outbreak response phase (Shirley and Rushton, 2005a).

The frequency of off-farm shipments per week (ranging from 1 to 6, Table 1), can be used as input parameter for the contact rate in simulation models. However, such simple statistics are much more informative when combined with the network information such as: the number of links over which those shipments occur, which types of nodes are involved, or the contact structure of the network.

The high and similar out-degree distributions observed for sow and nursery farms compared with finishing farms illustrated the similar risk these two farm-types present as disease sources, once they become infected. However, when one also considers the higher in-degree of nursery farms (relative to sow farms), it is clear that nursery farms present a greater overall risk through their ability to act as hubs in the spread of disease. In general, finishing farms may present a lower risk to the spread of disease from farm-to-farm by animal shipments. The similar in-degree distributions of nursery and finishing farms implied similar vulnerability to, or risk of, introduction (receipt) of diseased animals. However once again, considering the higher out-degree of nursery farms, they will likely be more influential in the spread of disease in the network. The out-degree measures of sow and finishing farms, and in-degree measure of finishing farms in this study were similar to those observed by Buttner et al., (2013).

Their higher betweenness scores means that targeting nursery farms for disease control measures can be more effective and efficient at the network population level (although this may be infection-specific and nursery-specific) than similar controls applied to sow or finishing farms. The eigenvector scores for individual farms provided additional insights into their respective centrality, because this score is weighted by the centrality scores of farms connected to the farm being assessed. Farms with both high betweennesss and high eigenvector scores are certainly hubs that will likely act as ‘super-spreaders’ if infected. In this study, all such farms were observed to be nursery farms (Fig 1 (b)) and belonged to a single cluster. Farms with high betweenness and low eigenvector scores act as central bridges connecting farms to the core of a network, which would otherwise be isolated. Removing these farms from the network increases the fragmentation index of a network. Farms with low betweenness scores but high eigenvector are well-connected and lie at the core of the network, but they have very few connections outside the core region. Identifying such farms can be useful for disease surveillance and disease outbreak management.

The positive associations observed for farm-size and shipment-size, with out-degree, and measures of centrality indicated that, in addition to large farms receiving and making more shipments to more farms, the probability of at least one infected animal being included in any of those shipments is greater for any given prevalence of disease within the source herd at the time. Some of the disease spread modelling tools explicitly model the disease transmission probability per shipment as a function of farm size, within-farm prevalence of an infection, and shipment size (Morris et al., 2001; *NAADSM* Development Team, 2011). It should be cautioned that other factors such as the biology of the disease agent, biosecurity and management practices of farms, and other factors governing between-farm linkages should also be considered in estimating the overall risk of disease transmission.

Cluster analysis provided a useful holistic picture of important aspects of the network. Most notably, it identified a group of specific nursery farms (cluster #4) with combined characteristics of large farm and shipment sizes and high measures of centrality. Given that individual farm characteristics may change over time (e.g. farm size, shipment size, frequency or degrees); if contemporary farm-specific data of such characteristics are not available at the time of an outbreak, then a reasonable precautionary policy would be to give higher priority to nursery farms in general, as targets for control measures.

Normally infection chains analyses (OIC and IIC) are used for contact tracing during response to an outbreak of an emerging disease .. However, knowledge of the characteristics of infection chains in a given network, before an outbreak, is also useful. Such knowledge can be used to prepare for and mitigate an outbreak, as well as estimate its likely magnitude, before an outbreak even starts. This is because infection chain analyses take into account direct and series contacts (through other farms), and the temporal sequence of these contacts (Dubé et al., 2008; Nöremark et al., 2011). Estimates of the OIC and IIC of two-week and monthly interval networks were similar to out-degree and in-degree estimates in this study. This suggested that estimates based simply on out-degree and in-degree provide an equivalent estimate of the likely magnitude of disease spread for a duration of up to one month. Since sow farms had significantly higher OIC than other farm types, the consequence component of the risk of disease spread from sow farms was higher than for other farm types. Similarly, since finishing farms had significantly higher IIC, the risk of them becoming infected was greater than for other farm types. Thus, farms with high IIC can be targeted for surveillance to increase the probability of detecting the infection, relative to random surveillance sampling of the population (Nöremark et al., 2011; Frossling et al., 2012; Rautureau et al., 2012; Buttner et al., 2013).

The scale-free topology of the network observed here was consistent with the findings of other studies of livestock shipments (Bigras-Poulin et al., 2007; Dubé et al., 2008; Lockhart et al., 2010; Nöremark et al., 2011; Rautureau et al., 2012; Buttner et al., 2013). This scale-free topology indicated that the degree distributions of farms were heterogeneous and that hubs (highly connected farms) were present.. Knowledge of the scale-free topology and highly connected farms is useful for prioritizing risk-based surveillance, tracing and control measures (Martinez-Lopez et al., 2009b; Dubé et al., 2011). It is particularly important because the speed of disease spread is faster in scale-free than in non-scale-free networks due to the presence of ‘super-spreaders’ (Shirley and Rushton, 2005b; Rahmandad and Sterman, 2008). Also, the speed of detection and control is faster if one takes into account the scale-free nature of the network by targeting hubs, than if only random surveillance and routine contact tracing is used to control an outbreak in a scale-free network (Shirley and Rushton, 2005b; Kiss et al., 2006a; Dube et al., 2011). The power-law scaling exponent for the total degree in this study was similar to other studies (Bigras-Poulin et al., 2007; Rautureau et al., 2012; Buttner et al., 2013). The out-degree and the scaling exponent are important parameters required for constructing a realistic scale-free network structure for network-based disease spread simulation models.

Low density and low clustering coefficient, and high fragmentation index of swine shipments networks in general indicate that the likelihood of a disease spreading to every farm may be relatively low. The small-world properties observed in this network indicated the presence of localized clusters that were connected to topologically distant clusters through a few long range connections. This kind of topology makes it possible for a disease to spread between distant clusters in the network. Small-world properties observed in this network were consistent with those reported in swine (Nöremark et al., 2011; Rautureau et al., 2012) and other livestock species (Leon et al., 2006; Dubé et al., 2008; Lockhart et al., 2010).

General comments

These results should be extrapolated cautiously. Nevertheless, this study provided epidemiologically relevant and reasonable estimates of farm-level measures, particularly for southwestern counties of Ontario. Future studies should consider incorporating such indirect contacts and shipment distances between farms as these are important parameters to consider in spatially-explicit diseases spread models (Morris et al., 2001; Garner and Beckett, 2005; Green et al., 2006; Harvey et al., 2007).

Conclusion

This study provided insights into a large swine shipment contact network in southwestern Ontario in Canada. Nursery farms are the high-risk farms in terms of risk of disease introduction as well as a potential source of spread to other farms (for infectious diseases that affect all age groups alike), and should be accorded high priority for disease prevention and control measures. The study found heterogeneity in the frequency of contacts among farm types. Also the network demonstrated small-world and scale-free topologies, consistent with livestock shipment networks reported in several other countries. Any attempt to incorporate disease transmission dynamics within simulation models should take these into consideration.

Conflict of interest statement

None declared

Acknowledgement

The authors would like to gratefully acknowledge OMFRA-AHSI (Animal Health Strategic Investment project of Ontario Ministry of Agriculture and Rural Affairs, Ontario) and CFIA (Canadian Food Inspection Agency) of Canada for funding this research and Atlantic Veterinary College of University of Prince Edward Island for stipend support to the first author. We would like to extend our appreciation to an Ontario animal health and management company for providing us the data and for their support. We also acknowledge three anonymous reviewers and the section editor for providing numerous valuable comments.

References

Aparicio, J.P., Pascual, M., 2007. Building epidemiological models from R0: an implicit treatment of transmission in networks. Proc Biol Sci 274, 505-512.

Barabasi, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509-512.

Barabasi, A.L., Bonabeau, E., 2003. Scale-free networks. Scientific American 288, 60-69.

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.

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.

Borgatti, S.P., Everett, M.G., Freeman, L.C., 2002. UCInet 6.0 Version 3.60 for Windows: Software for Social Network Analysis. Havard: Analytic Technologies.

Buttner, K., Krieter, J., Traulsen, A., Traulsen, I., 2013. Static network analysis of a pork supply chain in Northern Germany-Characterisation of the potential spread of infectious diseases via animal movements. Prev Vet Med 110, 418-428.

Cameron, A.R., 2012. The consequences of risk-based surveillance: Developing output-based standards for surveillance to demonstrate freedom from disease. Prev Vet Med 105, 280-286.

Christley, R.M., French, N.P., 2003. Small-world topology of UK racing: the potential for rapid spread of infectious agents. Equine Vet J 35, 586-589.

Clauset, A., Shalizi, C., Newman, M., 2009. Power-law distributions in empirical data. SIAM Review 51, 661-703.

Dube, C., Ribble, C., Kelton, D., McNab, B., 2011. Estimating potential epidemic size following introduction of a long-incubation disease in scale-free connected networks of milking-cow movements in Ontario, Canada. Prev Vet Med 99, 102-111.

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.

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.

Dubé, C., Ribble, C., Kelton, D., McNab, B., 2011. Introduction to network analysis and its implications for animal disease modelling. Rev Sci Tech 30, 425-436.

Eubank, S., Guclu, H., Kumar, V.S., Marathe, M.V., Srinivasan, A., Toroczkai, Z., Wang, N., 2004. Modelling disease outbreaks in realistic urban social networks. Nature 429, 180-184.

Everitt, B.S., Landau, S., Leese, M., Stahl, D., 2011. Cluster Analysis. John Wiley & Sons Ltd Chichester, West Sussex, UK.

Frossling, J., Ohlson, A., Bjorkman, C., Hakansson, N., Noremark, M., 2012. Application of network analysis parameters in risk-based surveillance - examples based on cattle trade data and bovine infections in Sweden. Prev Vet Med 105, 202-208.

Garner, M.G., Beckett, S.D., 2005. Modelling the spread of foot-and-mouth disease in Australia. Aust Vet J 83, 758-766.

Gibbens, J.C., Sharpe, C.E., Wilesmith, J.W., Mansley, L.M., Michalopoulou, E., Ryan, J.B., Hudson, M., 2001. Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. Vet Rec 149, 729-743.

Green, D.M., Kiss, I.Z., Kao, R.R., 2006. Modelling the initial spread of foot-and-mouth disease through animal movements. Proc Biol Sci 273, 2729-2735.

Hanneman, R., M, R., 2005. Introduction to social network methods. Published in digital form at http://faculty.ucr.edu/~hanneman/. Riverside, CA: University of California (accessed on 10 July 2012).

Harvey, N., Reeves, A., Schoenbaum, M.A., Zagmutt-Vergara, F.J., Dube, 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.

Hsu, C.-I., Shih, H.-H., 2010. Transmission and control of an emerging influenza pandemic in a small-world airline network. Accident Analysis & Prevention 42, 93-100.

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 Biol Sci 273, 1999-2007.

Keeling, M.J., Woolhouse, M.E., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J., Grenfell, B.T., 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science 294, 813-817.

Kiss, I.Z., Green, D.M., Kao, R.R., 2006a. Infectious disease control using contact tracing in random and scale-free networks. J R Soc Interface 3, 55-62.

Kiss, I.Z., Green, D.M., Kao, R.R., 2006b. The network of sheep movements within Great Britain: Network properties and their implications for infectious disease spread. J R Soc Interface 3, 669-677.

Kiss, I.Z., Green, D.M., Kao, R.R., 2008. The effect of network mixing patterns on epidemic dynamics and the efficacy of disease contact tracing. J R Soc Interface 5, 791-799.

Leon, E.A., Stevenson, M.A., Duffy, S.J., Ledesma, M., Morris, R.S., 2006. A description of cattle movements in two departments of Buenos Aires province, Argentina. Prev Vet Med 76, 109-120.

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.

Longini, I.M., Nizam, A., Xu, S., Ungchusak, K., Hanshaoworakul, W., Cummings, D.A.T., Halloran, M.E., 2005. Containing Pandemic Influenza at the Source. Science 309, 1083-1087.

Mansley, L.M., Dunlop, P.J., Whiteside, S.M., Smith, R.G., 2003. Early dissemination of foot-and-mouth disease virus through sheep marketing in February 2001. Vet Rec 153, 43-50.

Martinez-Lopez, B., Perez, A.M., Sanchez-Vizcaino, J.M., 2009a. Combined application of social network and cluster detection analyses for temporal-spatial characterization of animal movements in Salamanca, Spain. Prev Vet Med 91, 29-38.

Martinez-Lopez, B., Perez, A.M., Sanchez-Vizcaino, J.M., 2009b. Social network analysis. Review of general concepts and use in preventive veterinary medicine. Transbound Emerg Dis 56, 109-120.

Morris, R.S., Wilesmith, J.W., Stern, M.W., Sanson, R.L., Stevenson, M.A., 2001. Predictive spatial modelling of alternative control strategies for the foot-and-mouth disease epidemic in Great Britain, 2001. Vet Rec 149, 137-144.

NAADSM Development Team, 2011. NAADSM version number 4.0.11. Free program distributed via the Internet at http://www.naadsm.org.

Newman, M.E.J., 2003. The Structure and Function of Complex Networks. SIAM Review 45, 167-256.

Nöremark, M., Hakansson, N., Lewerin, S.S., Lindberg, A., Jonsson, A., 2011. Network analysis of cattle and pig movements in Sweden: measures relevant for disease control and risk based surveillance. Prev Vet Med 99, 78-90.

Opsahl, T., Panzarasa, P., 2009. Clustering in weighted networks. Social Networks 31, 155-163.

Ortiz-Pelaez, A., Pfeiffer, D.U., Soares-Magalhaes, 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.

Poljak, Z., Dewey, C.E., Martin, S.W., Christensen, J., Carman, S., Friendship, R.M., 2008. Prevalence of and risk factors for influenza in southern Ontario swine herds in 2001 and 2003. Can J Vet Res 72, 7-17.

Rahmandad, H., Hu, K., Duintjer Tebbens, R.J., Thompson, K.M., 2011. Development of an individual-based model for polioviruses: implications of the selection of network type and outcome metrics. Epidemiol Infect 139, 836-848.

Rahmandad, H., Sterman, J., 2008. Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models. Management Science 54, 998–1014.

Rautureau, S., Dufour, B., Durand, B., 2012. Structural vulnerability of the French swine industry trade network to the spread of infectious diseases. Animal 6, 1152-1162.

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 R Soc Interface 4, 669-674.

Sharkey, K.J., Bowers, R.G., Morgan, K.L., Robinson, S.E., Christley, R.M., 2008. Epidemiological consequences of an incursion of highly pathogenic H5N1 avian influenza into the British poultry flock. Proc Biol Sci 275, 19-28.

Shirley, M.D., Rushton, S.P., 2005a. Where diseases and networks collide: lessons to be learnt from a study of the 2001 foot-and-mouth disease epidemic. Epidemiol Infect 133, 1023-1032.

Shirley, M.D.F., Rushton, S.P., 2005b. The impacts of network topology on disease spread. Ecological Complexity 2, 287-299.

Statistics Canada, 2007. Farm operators by farm type and province (2001 and 2006 Censuses of Agriculture) (Canada) (Summary Tables). Last updated May 16, 2007. Statistics Canada:http://www40.statcan.ca/l01/cst01/agrc22a.htm?sdi=farm%20operators (Accessed on 28 March 2010).

Statistics Canada, 2012. 2011 Census of Agriculture: Farm and farm operator data. Statistics Canada:http://www.statcan.gc.ca/ca-ra2011/index-eng.htm (Accessed on 9 May 2013).

Vernon, M.C., Keeling, M.J., 2009. Representing the UK's cattle herd as static and dynamic networks. Proc Biol Sci 276, 469-476.

Volkova, V.V., Howey, R., Savill, N.J., Woolhouse, M.E., 2010. Sheep movement networks and the transmission of infectious diseases. PLoS One 5, e11185.

Wasserman, S., Faust, K., 1994. Social Network Analysis, Methods and Applications. Cambridge University Press New York.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440-442.

Webb, C.R., 2005. Farm animal networks: unraveling the contact structure of the British sheep population. Prev Vet Med 68, 3-17.

Webb, C.R., 2006. Investigating the potential spread of infectious diseases of sheep via agricultural shows in Great Britain. Epidemiol Infect 134, 31-40.

Table 1: Descriptive statistics of swine shipments among three farm types for 2006 and 2007 in southwestern Ontario, Canada.

Farm type		**No. of source farms**		**No. of destination farms**		**Total no. of links**		**Total no. of shipments**		**Median (range) of weekly off-farm links**	**Median (range) of weekly shipments/ pair**
Source	**Destination**	**2006**	**2007**	**2006**	**2007**	**2006**	**2007**	**2006**	**2007**		
Sow	Nursery	19	20	47	47	121	113	1787	1903	1 (1–3)	2 (1–6)
Nursery	Nursery	21	39	20	36	23	51	26	82	1 (1)	1 (1)
Nursery	Finishing	46	46	107	110	295	294	696	831	1 (1–4)	1 (1–3)
Finishing	Finishing	49	57	48	61	61	89	232	203	1 (1–2)	1 (1–6)

Table 2: Descriptive network metrics of swine shipments among three farm types for the 2-year period (2006–2007) in southwestern Ontario, Canada.

Network metrics	**Overall**	**Sow**	**Nursery**	**Finishing**
A. Network size				
i) Number of nodes (farms)	245	20	68	157
ii) Number of directed links	810	-	-	-
iii) Total number of shipments	5,760	-	-	-
iv) Network size (all possible pair-wise links)	59,780	-	-	-
B. Node level centrality measures - values reported are median (minimum and maximum range)				
i) Out-degree	1 (0–25)	6 (1–21)	8 (0–25)	0 (0–4)
ii) Normalised out-degree	0.41 (0.00–10.25)	2.46 (0.41–8.61)	3.28 (0.00–10.25)	0.00 (0.00–1.64)
iii) Frequency of off-farm shipments	1 (0–409)	203 (9–409)	16 (0–279)	0 (0–260)
iv) In-degree	3 (0–12)	-	3 (0–9)	3 (0–12)
v) Normalised median in-degree	1.23 (0.00–4.92)	-	1.03 (0.00–3.69)	1.23 (0.00–4.92)
vi) Frequency of onto-farm shipments	8 (0–363)	-	48 (0–215)	7 (0–363)
vii) Total degree	5 (1–29)	6 (1–21)	9 (1–29)	4 (1–13)
viii) Betweenness score	0 (0–106)	-	8 (0–106)	0 (0–52)
ix) Normalised betweenness score	0.00(0.00–0.18)	-	0.01 (0.00–0.18)	0.00 (0.00–0.09)
x) Eigenvector score	0.01 (0.00–0.23)	0.03 (0.00–0.18)	0.02 (0.00–0.23)	0.01 (0.00–0.14)
xi) Normalised Eigenvector score	1.72 (0.00–32.39)	3.94 (0.00–24.68)	2.74 (0.00–32.39)	1.13 (0.00–20.01)
C. Network level measures				
i) Density	0.014	-	-	-
ii) Out-degree centralisation index	8.9%	-	-	-
iii) In-degree centralisation index	3.6%	-	-	-
iv) Betweenness centralization index	0.17%	-	-	-
v) Diameter	5.0	-	-	-
vi) Median geodesic distance (mode)	2.0 (2.0)	-	-	-
vii) Clustering coefficient	0.09	-	-	-
viii) Fragmentation	0.96	-	-	-

Figure 1: Scatter plot distributions of: (a) in-degree versus out-degree of nursery (red circle) and finishing (blue triangle) farms where size of marker is proportional to the betweenness centrality scores (sow farms are not shown as they have zero in-degree and betweenness score); and (b) Eigenvector versus betweenness scores of sow (green diamond), nursery (red circle) and finishing (blue triangle) farms involved in the swine shipment network in southwestern Ontario. Labeled farms belonged to cluster 4 groups that may play a central role in terms of vulnerability to disease introduction and spread to other farms as they have high key network measures; particularly farms K, J, D, and G with highest betweenness centrality measure.

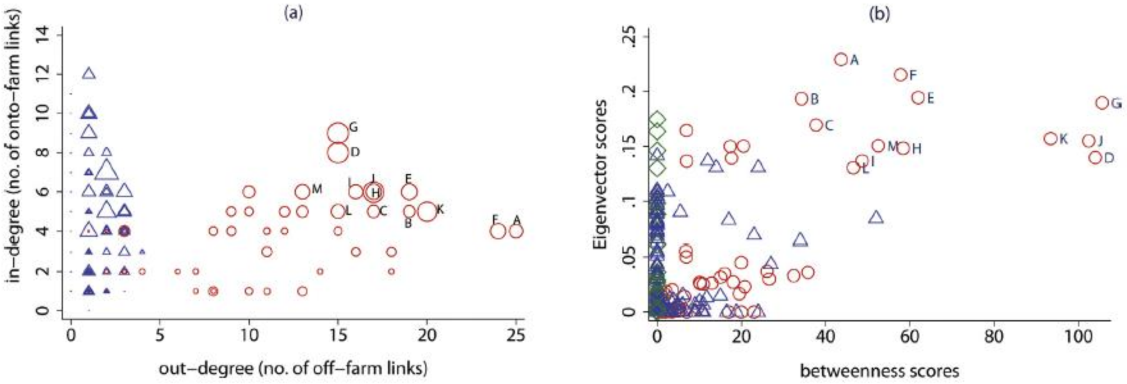

Figure 2: Percentage of farms with a given number of monthly: (a) *outgoing infection chain*, and (b) *incoming infection chain* values for each of the three production types described in the network analysis of swine shipments in southwestern, Ontario, Canada.

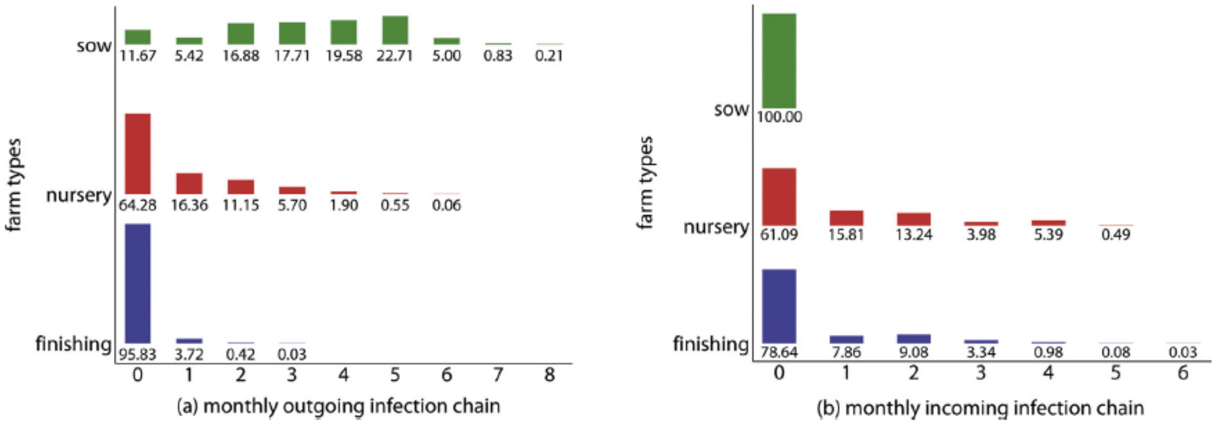

Table S1: Description of network terminology used in the main text.

Terminology	**Description**
Betweenness	The number of times that a node falls on the shortest path between other pairs of nodes in the network.
Betweenness centralization index	The sum of differences between largest betweenness of most central node with all other nodes divided by the maximum possible values, and expressed in percentage terms.
Clustering coefficient	The proportion of closed triplets out of the total number of triplets (both open and closed) present in the network. A triplet is formed when three nodes are connected by either two (open triplets) or three (closed triplets) undirected links [1].
Density	Density is the proportion of links (L) present out of all possible links in the network. For a directed network density is equal to L/(N(N-1)), where N is the number of nodes present.
Diameter	The longest geodesic distance between any pair of nodes of a network.
Eigenvector	A relative centrality score of a node assigned as an increasing function of the sum of all the centralities of all the nodes to which it is connected. A node connected to other nodes with high centrality scores will have higher Eigenvector score than another node with equal connections, but connected to nodes with low centrality scores. Therefore this measure not only accounts for a node’s degree but also for the degree of the other nodes to which it is connected.
Fragmentation	The proportion of pairs of nodes in a network that are unreachable either through direct or indirect pathways and provide measure of network’s disconnectedness.
Geodesic distance	The shortest possible path between two nodes and is the most efficient connection between them.
In-degree	In the directed network, in-degree is the number of nodes from which a particular node received shipments of animals. For weighted directed network, in-degree represents number of shipments received by a node irrespective of the node from which the shipment originated.
In-degree centralization index	The sum of the differences between largest in-degree of most central node with all other nodes divided by the maximum possible values and expressed in percentage.
Links	The connection between pairs of nodes established through shipments of animals.
Network size	The total number of possible unique pairs of nodes. If N is equal to the number of nodes, then network size is equal to N (N-1).
Normalized betweenness	The proportion of betweenness of a node to a maximum possible betweenness of the network and expressed in percentage.
Normalized in-degree	It is the in-degree of a node divided by the maximum possible degree of the network and expressed in percentage.
Normalized out-degree	It is the out-degree of a node divided by the maximum possible degree of the network and expressed in percentage.
Nodes/vertices	Premises such as farms and PUs in the network.
Out-degree	In the directed network, out-degree of a node is the number of out-going links to other nodes (number of premises to which a farm sent its swine shipments). In a weighted directed network, out-degree of a node is the number of shipments sent to other premises.
Out-degree centralization index	The sum of the differences between largest out-degree of most central node with other nodes divided by the maximum possible values, and expressed in percentage.

Table S2: Distribution of the number of pigs shipped by farm types for the period 2006 to 2007 in southwestern Ontario, Canada.

Premises type		Median (5th & 95th percentiles) of shipment size	Shipment size (minima – maxima)
Source	Destination		
Sow	Nursery	203 (64–590)	9–1190
Nursery	Nursery	28 (1–218)	1–1580
Nursery	Finishing	243 (30–1943)	1–2355
Finishing	Finishing	15 (1–274)	1–2598

Table S3: Clusters identified by hierarchical cluster analysis using weighted-average linkage method based on farm-level variables and measures of centrality of the swine shipments network analysis in southwestern Ontario, Canada.

Variable	**Cluster 1 (n = 159)**	**Cluster 2 (n =62)**	**Cluster 3 (n = 10)**	**Cluster 4 n=13)**	**Cluster 5 (n=1)**
Farm size[a]	1,406 (250–2,700)	1,060 (360–2,760)	1,965 (1,000–4,000)	2,000(1,235– 2,540)	1,950
Shipment size[a]	39 (0–1,166)	0 (0–145)	346 (124–232)	467 (232–600)	2,352
Out-degree[a]	1 (0–15)	0 (0–3)	15 (12–21)	17 (13–25)	1
In-degree[a]	2 (0–6)	7 (4–12)	1 (0–3)	6 (4–9)	1
Betweenness[a]	0 (0–36)	0 (0–52)	4 (0–21)	58 (34–106)	0
Eigenvector[a]	0 (0–0.09)	0.01 (0– 0.14)	0.15 (0.10– 0.18)	0.16 (0.13–0.22)	0
Sow (%)	9.4	0	50.0	0	0
Nursery (%)	30.2	1.6	50.0	100	100
Finishing (%)	60.4	98.4	0	0	0

[a] *Values reported are median and interquartile range.*

Reference

1. Opsahl T., Panzarasa P. 2009 Clustering in weighted networks. Social Networks 31, 155-163.

Supplementary figures

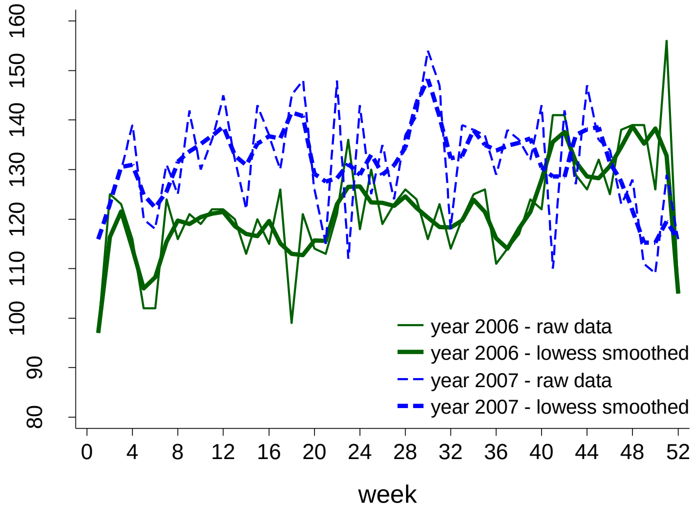

Figure S1: Frequency of weekly shipments of swine for 2006 and 2007 in southwestern Ontario, Canada. A lowess smoothing line was generated using locally weighted regression of frequency of shipments on weeks of each year with a bandwidth of 0.095.

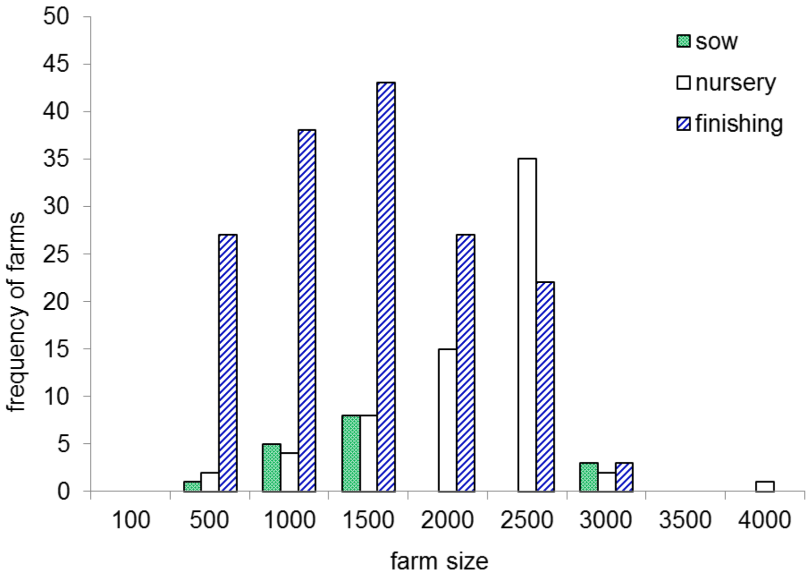

Figure S2: Distribution of the number of farms according to the total number of animals recorded in the database for the period 2006 to 2007 by farm type in southwestern Ontario, Canada.

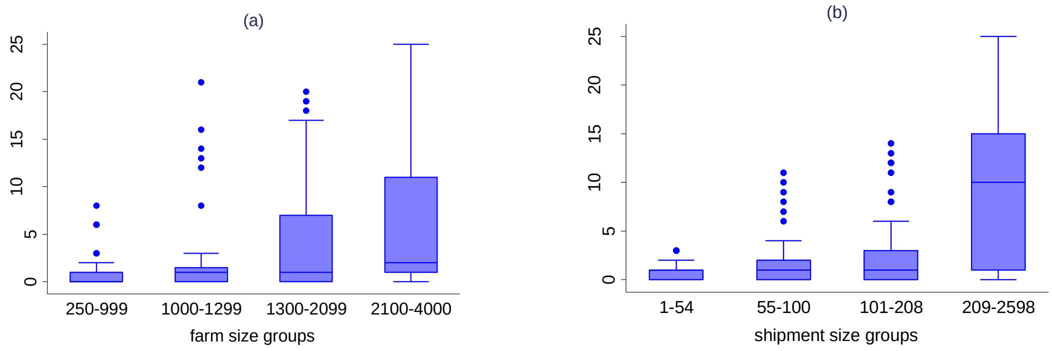

Figure S3: Out-degree distributions by: (a) farm size groups and (b) shipment size groups of a swine shipment network in southwestern Ontario, Canada.

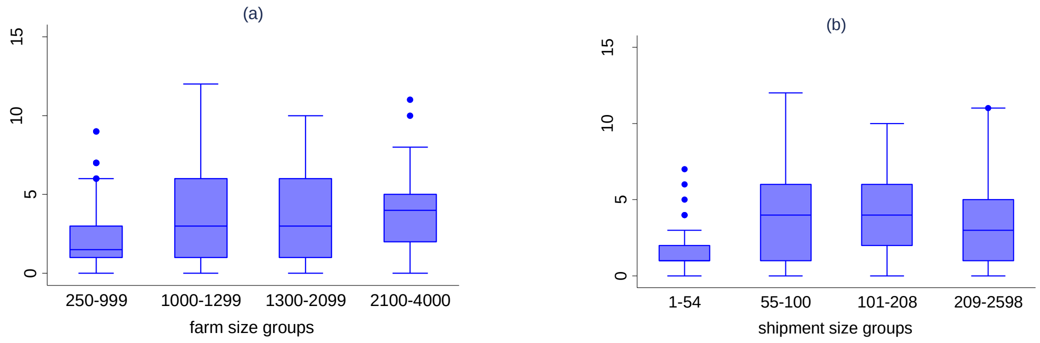

Figure S4: In-degree distributions by: (a) farm size groups and (b) shipment size groups of a swine shipment network in southwestern Ontario, Canada.

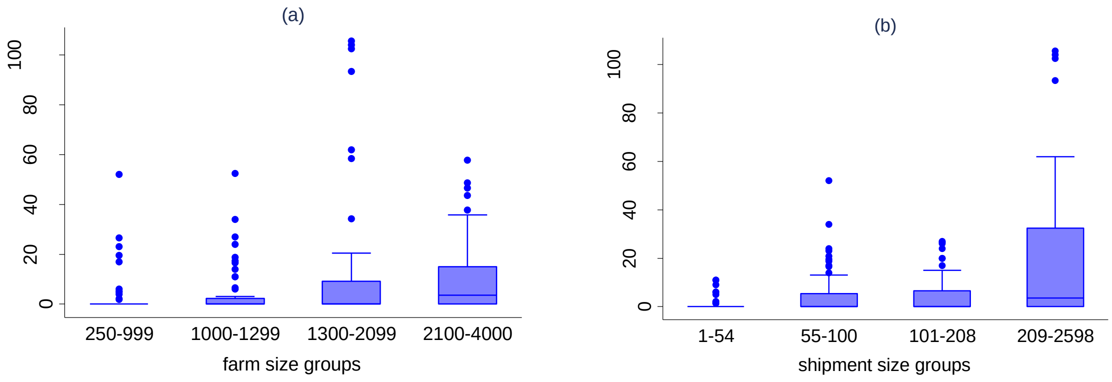

Figure S5: Betweenness centrality score distributions by: (a) farm size groups and (b) shipment size groups of a swine shipment network in southwestern Ontario, Canada.

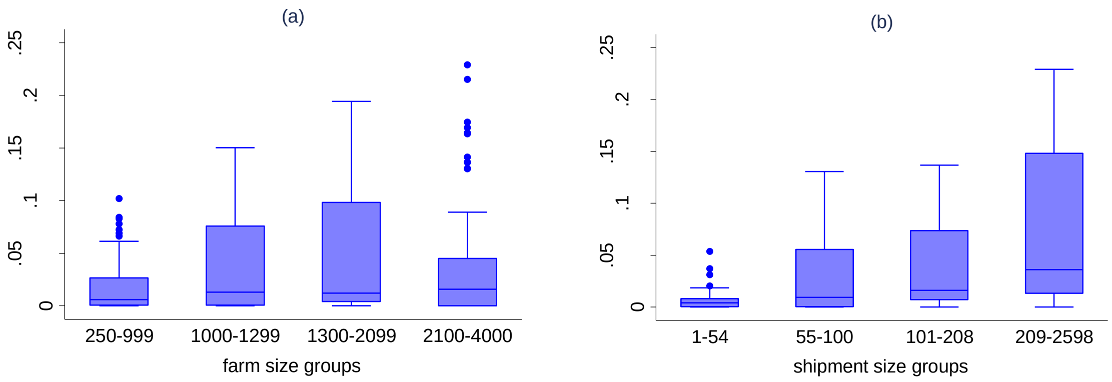

Figure S6: Eigenvector centrality distributions by: (a) farm size groups and (b) shipment size groups of a swine shipment network in southwestern Ontario, Canada.

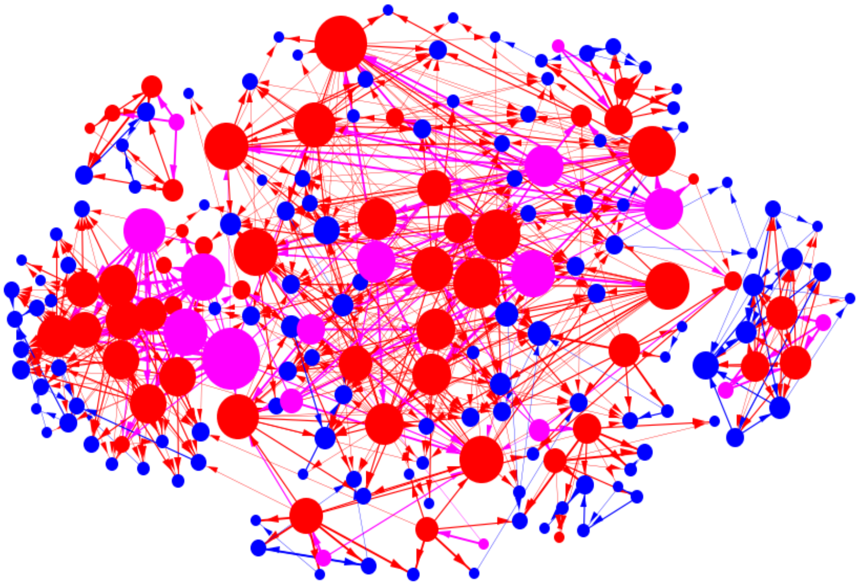

Figure S7: Network of swine shipment among three farm types in southwestern Ontario, Canada. Node sizes and width of links are proportional to the total degree and three categories of frequency of shipments (1, 2–10 and 11–258). Key: sow herd (pink); nursery (red); finishing (blue).