Network analysis and optimization of animal transports

Nina Håkansson

Department of Physics, Chemistry and Biology
Theoretical Biology
Linköping University
SE-581 83 Linköping, Sweden

Linköping, April 2012
Cover page shows example of route coordination with heuristics used in the thesis. On front routes are coordinated with Clarke & Wright heuristic (right) and good choice heuristic (left). On the back routes are coordinated by good choice heuristic with good parameters (right) and bad parameters (left). In the upper corner on back page routes do not end up at closest destination node. There are many reasons to why my thesis is pink, one is that pink books are often nice and easy to read and that is something to aim for with a thesis.
Abstract

This thesis is about animal transports and their effect on animal welfare. Transports are needed in today’s system of livestock farming. Long transports are stressful for animals and infectious diseases can spread via animal transports. With optimization methods transport times can be minimized, but there is a trade-off between short distances for the animals and short distances for the trucks. The risk of disease spread in the transport system and disease occurrence at farms can be studied with models and network analysis.

The animal transport data and the quality of the data in the Swedish national database of cattle and pig transports are investigated in the thesis. The data is analyzed regarding number of transports, number of farms, seasonality, geographical properties, transport distances, network measures of individual farms and network measures of the system. The data can be used as input parameters in epidemic models.

Cattle purchase reports are double reported and we found that there are incorrect and missing reports in the database. The quality is improving over the years i.e. 5% of cattle purchase reports were not correctly double reported in 2006, 3% in 2007 and 1% in 2008. In the reports of births and deaths of cattle we detected date preferences; more cattle births and deaths are reported on the 1st, 10th and 20th each month. This is because when we humans don’t remember the exact number we tend to pick nice numbers (like 1, 10 and 20). This implies that the correct date is not always reported.

Network analysis and network measures are suggested as tools to estimate risk for disease spread in transport systems and risk of disease introduction to individual holdings. Network generation algorithms can be used together with epidemic models to test the ability of network measures to predict disease risks. I have developed, and improved, a network generation algorithm that generates a large variety of structures.
In my thesis I also suggest a method, the *good choice* heuristic, for generating non-optimal routes. Today coordination of animal transports is neither optimal nor random. In epidemic simulations we need to model routes as close to the actual driven routes as possible and the *good choice* heuristic can model that. The heuristic is tuned by two parameters and creates coordination of routes from completely random to almost as good as the Clarke and Wright heuristic. I also used the method to make the rough estimate that transport distances for cattle can be reduced by 2-24% with route-coordination optimization of transports-to-slaughter.

Different optimization methods can be used to minimize the transport times for animal-transport in Sweden. For transports-to-slaughter the strategic planning of “which animals to send where” is the first step to optimize. I investigated data from 2008 and found that with strategic planning, given the slaughterhouse capacity, transport distances can be decreased by about 25% for pigs and 40% for cattle. The slaughterhouse capacity and placement are limiting the possibility to minimize transport times for the animals. The transport distances could be decreased by 60% if all animals were sent to the closest slaughterhouse 2008. Small-scale and mobile slaughterhouses have small effect on total transport work (total transport distance for all the animals) but are important for the transport distances of the animals that travel the longest.
Med dagens djuruppfödningssystem transporteras kor och grisar mellan gårdar och till slakterier Sverige. Transporterna påverkar djurens välfärd och kan öka smittspridningen vid eventuella epidemic. Nästan alla djur transportereras minst en gång under sin livstid, till slakteriet. Det finns ungefär 1700 aktiva gårdar i Sverige som har grisar, 21 600 gårdar med kor och ca 60 slakterier. I Sverige slaktas årligen 0.5 miljoner kor och 3 miljoner grisar.


kan ta prover på alla gårdar för att leta efter infekterade djur. Genom att använda ett riskbaserat urval kan provtagning fokuseras till de besättningar som har högre risk att sjukdomen finns i besättningen och där kan nätverksmätt vara användbara.


Acknowledgements

Jag vill tacka min handledare för möjligheten att gå en forskarutbildning och vara delaktig i många olika forskningsprojekt. Du är alltid positiv och full av nya forskningsidéer och du har insikten att livet utanför jobbet också är viktigt. Tack Uno!

Det har varit roligt att vara en del av forskningsmiljön Systems Biology Research Centre på Högskolan i Skövde. Annie har varit min handledare i Skövde och förutom att vara glad och engagerad har du alltid gett mig precis den återkoppling jag behövtt på skriven text. Tack Annie!


Stort tack till de som tillsammans med mig har skrivit artiklarna till avhandlingen: Maria, Sanna, Patrik, Mikael, Mathias, Tom, Ann, Jenny, Bo, Annie och Uno. Speciellt till Maria som gjorde ett fantastisikt jobb med manuset till två av artiklarna. Ett stort tack till Frida och Linda för värdefulla kommentarer på kappan, och för hjälp med allt administrativt småpyssel inför disputationen.

Jag vill tacka alla kollegor (ingen nämnd och ingen glömd) i Linköping och i Skövde för alla trevliga samtal över lunch, fika och lyxika om forskning och om livet i stort. Tack till alla er som jag under åren diskuterat vad forskning är och vad det innebär att vara doktorand. Tack och tänk att det blev en avhandling till slut!

Tack till Peter som lånade ut sin laptop med Cygwin när min dator gav upp dagarna innan avhandlingen skulle till tryck. Tack till IFM support som lagade datorn. Jag vill tacka min bror Andreas som förutom att gärna diskuterar matematik och datorer, påpekat att inget man behöver ska sparas på härddisken. Det blir mer äventyr och mindre katastrof att datorn kraschar när alla filer är trygga på Home.

Rustan, tack för att du läst, kommenterat, diskuterat och hjälpit mig förbättra min avhandling ☺.
Publications included in thesis


1Submitted to PLoS ONE
My contributions to the papers

Main responsible for the writing of paper I, IV and VI. Implemented all code and ran all simulations for paper I and IV. Main responsible for data editing for paper II, III and VI. Main responsible for calculating network measures to paper III and for implementing the analysis in paper II. Main responsible for the implementation of the SpecNet algorithm in paper IV and V. No part in implementing the facility location model in paper VI or the CM algorithm in paper V. Participated in choosing analysis and simulations methods for all papers. Participated in writing all papers and analyzing the results for all papers.

Publications not included in thesis


# Contents

Abstract i  
Populärvetenskaplig sammanfattning iii  
Acknowledgements v  
Publications included in thesis vi  
Contents viii  
List of Figures ix  
List of Tables ix  

1 Aims 1  

2 Introduction 3  
2.1 Network analysis of the databases of animal movements 4  
2.2 Network generation algorithms 5  
2.3 Modeling and optimizing animal transports 6  

3 Methods and data 9  
3.1 Point patterns from 2-dimensional 1/f-noise 9  
3.2 Probability distribution with varying kurtosis and variance 10  
3.3 Measuring aggregation 10  
3.4 Network measures 11  
3.5 SpecNet and its extension 13  
3.6 Coordinate, road and transport databases 14
List of Figures

3.1 Point patterns ............................................. 9
3.2 Infection chain illustration ................................ 11
3.3 SpecNet networks ........................................... 13
3.4 Farms and slaughterhouses in Sweden .................. 15
3.5 Route coordination heuristics ............................. 16

List of Tables

3.1 Network measures used in this thesis .................... 12
5.1 Transport distances in Sweden 2008 ....................... 26
1 Aims

My overall aim was to analyze the animal transport-system in Sweden and its effect on animal welfare. The aim was to investigate methods to minimize the negative effect of animal transports on animal welfare; both the direct effect of transport routes and long transport times but also the indirect effect of the risk for spread of diseases via the animal transports. In more detail my aims were to:

1. *Investigate the Swedish animal transport data and the quality of the data.*

2. *Develop a network generation algorithm suitable for simulations to thoroughly test network measure’s ability to predict risk of disease spread in animal transport networks.*

3. *Develop a method to model animal transport routes as they are coordinated today.*

4. *Apply optimization tools to minimize transport times by considering where to send the animals and how to coordinate transports and also estimate the potential economic gain of transport coordination.*

5. *Investigate the effect of geographic limitation of today’s slaughterhouse capacity on transport times for animals.*
2 Introduction

The structure of livestock farming in Sweden today requires many animal transports. There are about 21,600 farms with cattle and 1,700 with pigs in Sweden [3]. Animals are transported between farms and to slaughterhouses. Especially pigs are transported between holdings because of the pig-farming system with sow pools and nucleus. About 0.5 million cattle and 3 million pigs are slaughtered at the about 60 slaughterhouses in Sweden every year. The livestock farming structure has changed over time; slaughterhouses and farms are centralized because of the benefits with larger facilities [37]. Animals are also to some extent moved closer to the slaughterhouses capacity [37]. Centralized slaughterhouses mean that there is a conflict between having livestock farming in the whole country and having short animal transports [37].

Animal transports affect animal welfare directly because the animals are stressed during transport [29]. The level of stress correlates with transport time and transports longer than 6h are extra stressful [29; 34]. The reason that welfare decreases with transport time can be that food and water are not supplied to the animals [34]. Studies have shown that more animals die during long transports (over 100 km) [42; 49]. Queues at slaughterhouses affect the unloading times and can create stressful environments for both staff and animals [28]. It is stressful for the animals to be kept overnight at the slaughterhouse [1]. Better planning of transport could reduce the need to have many animals already at the slaughterhouse in the morning [1]. Animal transports also affect animal welfare indirectly because they might increase the risk for spread of contagious infectious diseases [31; 34]. Throughout the thesis when I use the word disease it is always contagious infectious disease that is meant.

An economic reason to maintain high animal welfare is that slaughter of stressed animals produce meat of lower quality, with more PSE (Pale Soft Exudative) and DFD (Dark Firm Dry) meat [4]. Animals are especially stressed when mixed with
2. Introduction

unfamiliar animals during transport [41; 4]. No transport coordination is the best strategy for the animals (shortest distances and no mixing), but that strategy is not economically viable or good for the environment regarding CO$_2$-emissions. Optimized coordination of animal transports decrease the transport distances and therefore lower the amount of CO$_2$-emissions. There is a trade-off between short transport distances for the animals and short driving distances for the trucks; and it needs to be considered when optimization software for animal transport is developed. Slaughtering at local and mobile slaughterhouses might be economically inefficient but consumers are willing to pay extra for meat (pigs and cattle but not chicken) slaughtered at local or mobile slaughterhouses [2; 17].

2.1 Network analysis of the databases of animal movements

There is legislation in EU since 2002 that cattle and pig transports must be reported and registered. These national databases have been studied with descriptive statistics and network analysis. We have studied the database in Sweden (paper II and III) and the corresponding databases have been studied in Belgium [61], UK [62; 63; 59; 52], Italy [8], Spain [50], Argentina [5], France [60] and Denmark [14; 13].

Network analysis is suggested to be used to predict size of epidemics and to locate risk holdings, see review articles [24; 51; 39; 21]. In network analysis networks consisting of nodes (vertices) and links (arcs, edges or contacts) between the nodes are studied. In our case the nodes are animal holdings and the links are animal transports between the holdings. Network models of disease spread give more accurate results compared to homogeneous-mixing models [9]. However, broad use of network analysis might be impossible because of the large amount of data [51]. Some network measures are much more time and computer demanding to calculate and that will effect when, how and if they are used at all.

Social network analysis was first used to model disease spread in human contact networks; and the most common application there is to use modeled networks of contact structure together with transition models to model the spread of disease [51]. Often networks are modeled as random, scale-free, small-world or lattice networks, see for example [39] for a comparison between simulations in the mentioned network types plus spatial networks. The networks used in veterinary medicine models are often static and non-temporal [8]. It is important to remember that the risk for spread of diseases via animal transport is both directional and isolated in time and it therefore behave different from disease spread in undirected contact
networks [74]. Consequently temporal aspects are important [8; 36] and dynamic networks are more appropriate to model spread of diseases via animal transports [71; 8].

Epidemic modeling of disease spread via animal transports has traditionally been done in real-world or modeled static networks [8] and different network measures have been found as candidates to predict epidemic sizes and disease risks of holdings and systems. Christley et al [18] showed that centrality measures of nodes, the contribution of single nodes to the entire network, are important for spread of diseases. They found that the degree (number of connections) is at least as good to predict infection risk as the more complicated measures of centrality: random-walk betweenness, shortest-path betweenness and farness [18]. Size of largest strong component (directed measure) and weak component (undirected measure) has been used to predict epidemic size [19], but also degree assortativity (measures if nodes with similar or different degree are connected) and clustering coefficient (measures triangles, if neighbours to a node are likely also neighbours) have been suggested [6]. With an SI (susceptible-infected) model and a static network of the Italian transport data the centrality measures closeness and eigenvector of holdings were shown to be correlated to the epidemic size [55]. The risk for spread of diseases has been shown to decrease (disappearing giant strongly connected component) when nodes with high betweenness centrality are deleted from the network [60]. A way to consider the important time aspects of animal transports is to use temporal network measures. Infection chain (a temporal and directed measure) predicts epidemic size better than strong and weak components [23]. In paper III we calculate out-going infection chain and in-going infection chain among other network measures for animal transport networks in Sweden.

2.2 Network generation algorithms

Many network generation algorithms have been developed to generate networks of different types like random, scale-free, small-world or spatial (see for example the review articles [68; 56; 35; 11]). Gilbert [30] introduced the theory of random graphs. In random graphs a specified number of nodes are randomly connected with links and the probability of a link between every pair of nodes is independent and equal to $p$. Small-world networks were introduced by Watts [72] and they are highly clustered like regular lattices and have small path lengths. Scale-free networks were introduced by Barabási and Albert [10] and refers to networks that have node degree distributions that follow a negative power law $p(k) = k^{-\alpha}$. For large
2. Introduction

positive $\alpha$ this means that nodes with many connections are rare and nodes with few connections are more numerous. Many methods to generate scale-free networks exist and one example is preferential attachment (node has high probability to connect to other nodes that already have many connections). The importance of the scale-free property is debated [40; 43] and some of the biological networks reported as scale-free are not [43]. The trend is not linear just because a line is fitted to data with regression and the goodness-of-fit is often not considered [43].

The spatial SpecNet algorithm (developed in paper V and IV) is a variant of the RG1 model introduced by Waxman [73] and it is similar to Keeling’s model [38]. In Waxman’s algorithm nodes are uniformly distributed in space, and an exponential distribution controls link forming dependent on the Euclidean distance between nodes. In Keeling’s algorithm focal nodes are used to create aggregated spatial distributions of farms and a Gaussian distribution controls link forming. We use spectral methods to create the more or less aggregated point-patterns of nodes and a generalized normal distribution for link-forming.

The animal transportation networks in Sweden have negative assortativity and clustering coefficients close to zero (calculated in paper III and V) and are therefore not best modeled as small-world networks. Farms and animal transports exist in a spatial context and the probabilities for transports depend on distances. In paper II we found that most transports are short range. Therefore spatial networks, with distance-dependent linking, seem a logic way to model animal transport networks. The network generation algorithm configuration model (CM) focuses on the degree distribution of the nodes and was introduced by Bender [12], (see more in [56]). It seems a promising method to construct animal transport networks, because it can be used to reconstruct networks with predefined degree distributions. In paper V a variant of the configuration model algorithm is compared to our spatial network algorithm SpecNet.

2.3 Modeling and optimizing animal transports

There are several different types of optimization that can be made for animal transports: strategic planning of where to send the animals, analysis of placement of slaughterhouses and tactical planning of how to coordinate routes. The final step is operational optimization that can be used in real time.

We have investigated the strategic planning of transports-to-slaughter as a location problem using a facility location model in paper VI. We minimize the distance that animals travel, the transport work, measured in animal*km. The ani-
2. Introduction

Animal transport-to-slaughter can be more completely analyzed as a location routing problem (LRP), see [54] for a review of location-routing problems. In Bavaria the problem of determining optimum size and location of slaughterhouses to minimize transportation and distribution costs has been investigated [26]. In Norway the problem of locating a slaughterhouse considering uncertainty in demand and run costs has been studied [66]. The location problems of slaughterhouses in Norway and Bavaria have been studied with focus on minimizing transport and slaughterhouse costs [16; 26]. In paper VI we minimize transport distances for the animals and have used detailed real data regarding transports made 2008, geographical coordinates of farms and slaughterhouses and actual distances between farms and slaughterhouses.

The rich vehicle routing problem (VRP) considers how to coordinate transports. The livestock collection problem (LCP) is a VRP extended with inventory constraints so that there are always animals in the slaughterhouse, because it is costly to stop slaughter and wait for animals [57; 58]. Oppen et al [57; 58] study LCP in Norway and use real world constraints on the routes, for example transport times of max 8h and heterogeneous vehicle fleet. They also include the possibility for health status of farms to affect pick-up order, for example protected farms need to go first (clean vehicle) and risk farms last (disease will not be spread further). Oppen et al [58] use column generation to solve the problem to optimum and handles up to 25 farms. Sigured et al [67] have studied VRPs using time windows and precedence constraints. They divide the VRP into a master problem and a sub problem and VRPs with up to 580 farms are solved to optimality.

We suggest a way to model animal transport coordination using distance dependent probability for transport coordination in paper I. This good choice heuristic will generate route coordination between optimal routes and random routes. No databases over route coordination exist and we need a way to model them. We compare the routes generated by the good choice heuristic to routes optimized with the Clarke & Wright heuristic.
3 Methods and data

3.1 Point patterns from 2-dimensional $1/f$-noise

In paper I, IV and V spectral methods and $1/f^\gamma$-noise are used to arrange nodes in networks and place farms in landscapes. This is done on a scale from random to very aggregated. The parameter $f$ is frequency and $\gamma$ is the parameter that regulates aggregation. Signals can be split into sine and cosine components with Fourier transform. In a periodogram the square of amplitudes for each frequency are con-

![Figure 3.1: Example of point patterns (bottom row) and corresponding periodogram with amplitudes plotted against frequency on a log-log axis (top row). The point patterns are more aggregated for higher $\gamma$, to the right.](image-url)
3. Methods and data

To generate aggregated point patterns a with noise signal is scaled by dividing the amplitude of the components with $f^{0.5}$. The periodogram of the scaled signal is proportional to $1/f^{\gamma}$. High $\gamma$ means that all high-frequency waves become very small compared to the low frequency waves, and the low frequency waves therefore dominate the signal. The scaled signal is inverse transformed into a 2-dimensional $1/f^{\gamma}$-noise matrix.

The method is similar to the one presented in appendix in [33] and we have used a variant of it to generate lattice landscapes [45]. To get a point pattern the row and column index of the matrix cells with highest values are chosen as coordinates for the nodes. See figure 3.1 for examples of point patterns generated with the method for different $\gamma$.

3.2 Probability distribution with varying kurtosis and variance

The distance between nodes or farms were used to calculate the probability for links or coordination of routes, in papers I, IV and V. The generalized normal distribution $P(d) = V^{-1}exp(-d^b/a^b)$ was used as probability distribution. The parameter $V$ is volume, $d$ is distance and the parameters $a$ and $b$ regulate the kurtosis and variance of the distribution. By changing only the parameter $b$ the probability distribution goes from uniform ($b = \infty$), via normal ($b = 2$), to exponential ($b = 1$). Variance and kurtosis can be estimated from for example dispersal data [20] or animal transport data [46].

3.3 Measuring aggregation

The structure of the spatial distribution of farms in Sweden and in parts of Sweden were needed in papers I and V. The method described in [53] was used to calculate the periodogram and the r-spectrum of the point pattern. We used the same method to measure the spatial aggregation of trees [44]. The mean amplitude and the mean frequency of the periodogram were calculated for the frequency intervals $[n, n+1]$ for all ($n = 0, 1, 2, \ldots$). Then a regression line was fitted to the points to find the slope, $\gamma$. See figure 3.1 (top row) for examples of the regression line fitted to the periodogram with this method. The point pattern of farms does not need to be fractal or pure $1/f^{\gamma}$-noise. The only goal is to find the best $\gamma$ to reconstruct the farm pattern when one has decided to use $1/f^{\gamma}$-noise.
3. Methods and data

3.4 Network measures

Network measures are used in paper III, IV and V. We used the network measures mean degree and link density to measure the amount of links or transports in the network. The degree of a node is the number of links connected to it. The link density is the number of links in the network compared to the maximum theoretical possible number, given the number of nodes. Clustering coefficient measures the amount of pairs of neighbors that are directly connected. Betweenness centralization index measures if nodes in the network have similar or different betweenness. The betweenness of a node measures how many shortest paths that pass the node. Fragmentation index measures the fragmentation of the network as the proportion of node-pairs that are not connected (neither directly nor indirectly). Degree assortativity measures if nodes with similar degree are likely to be connected or not and if the probably for links between nodes are independent of node degrees. See table 3.1 for an overview of network measures used in the thesis.

The measures in-going and out-going infection chain are both directed and temporal. They were calculated for animal transport networks in Sweden in paper III.
3. Methods and data

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>path</td>
<td>A path in a network is list of nodes with links between each node and the next node in the list.</td>
</tr>
<tr>
<td>shortest path</td>
<td>The shortest path(s) between ( n_1 ) and ( n_2 ) is the path(s) with minimum number of nodes.</td>
</tr>
</tbody>
</table>

**non-directed node properties**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>degree (deg)</td>
<td>Number of links connected to a node.</td>
</tr>
<tr>
<td>betweenness centrality (BC)</td>
<td>A node with high betweenness (BC near 1) lies on the shortest path(s) between many other pairs of nodes. Range ([0, 1]).</td>
</tr>
<tr>
<td>clustering</td>
<td>The part of a nodes neighbours that are also connected.</td>
</tr>
</tbody>
</table>

**non-directed network properties**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>order</td>
<td>Number of nodes in the network. Number of links in the network.</td>
</tr>
<tr>
<td>size</td>
<td>Number of links in the network. Number of links in the network.</td>
</tr>
<tr>
<td>mean degree</td>
<td>Mean number of links per node in the network.</td>
</tr>
<tr>
<td>clustering coefficient</td>
<td>Average clustering of the nodes. Range ([0, 1]).</td>
</tr>
<tr>
<td>link density</td>
<td>Number of links divided by the theoretical number of links. Range ([0, 1]).</td>
</tr>
<tr>
<td>group betweenness central-</td>
<td>Measures if nodes have similar (GBCI near 0) or different (GBCI near 1) betweenness. Range ([0, 1]).</td>
</tr>
<tr>
<td>ization index (GBCI)</td>
<td></td>
</tr>
<tr>
<td>fragmentation index (F)</td>
<td>Measures if the network is connected (F near 0) or not (F near 1). It measures the amount of node pairs that are not connected. Range ([0, 1]).</td>
</tr>
<tr>
<td>degree assortativity (r)</td>
<td>Measures if nodes with similar degrees (r near 1) or different degrees (r near -1) are connected. If the probability for a link is independent of node degrees (r=0). Range ([-1, 1]).</td>
</tr>
</tbody>
</table>

**directed node properties, links have direction**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>in-degree or support</td>
<td>The number of links going into a node.</td>
</tr>
<tr>
<td>out-degree or influence</td>
<td>The number of links going out from a node.</td>
</tr>
</tbody>
</table>

**directed and temporal node properties, links have both time and direction**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>in-going infection chain</td>
<td>The number nodes directly and indirectly connected into the node.</td>
</tr>
<tr>
<td>out-going infection chain</td>
<td>The number nodes directly and indirectly connected out from the node.</td>
</tr>
</tbody>
</table>

Table 3.1: Network measures used in this thesis.
The out-going infection chain of a specific node is the number of other nodes that are directly and indirectly connected from that node. It is the number of other farms that can be infected (via transports) from a disease starting at the specific farm. In-going infection chain of a specific node is the number of other nodes directly and indirectly connected in-to that node. It is the number of farms the disease could have started in if it spread with animal transports and ended up at the specific farm. See figure 3.2 for infection chains calculated for an example network.

### 3.5 SpecNet and its extension

Spectral methods were used to generate spatial networks in paper IV (first version) and V (extended version). Nodes were placed in a spatial context aggregated or random using 2-dimensional $1/\gamma$-noise. The probabilities for links between nodes depended on the distance between them. The method has five parameters: number of nodes, link density, node aggregation ($\gamma$) and variance and kurtosis of the probability distribution. With these parameters the method can create networks with different structures regarding clustering coefficient, degree assortativity, fragmentation index and GBCI (group betweenness centralization index). The first version of the method could not generate networks with negative degree assortativity. To be able to generate networks with negative degree assortativity we extended the algorithm and added a step where nodes were divided into two groups: regular and focal. The probability for a link between two nodes from different groups was mul-
3. Methods and data

tiplied with a constant (focal scale factor). This added two more parameters to the algorithm: the number of focal nodes and the focal scale factor. See examples of networks generated with SpecNet in figure 3.3.

3.6 Coordinate, road and transport databases

The database of animal movements is analyzed in papers II, III, V and VI. The database of animal movements, cattle births and cattle deaths was obtained from the Swedish board of Agriculture. For cattle the movements are reported on an individual level, and each cattle has an identification number. PPNs (production place numbers) are used to identify farms and slaughterhouses. A movement report for cattle contains the identification number of the animal, date, sending PPN, receiving PPN, type of movement and when relevant also slaughtering PPN. Purchase and slaughter movements are double reported for cattle; both sending and receiving PPN makes a report. This allows us to evaluate the quality of the data by checking the number of reports missing their corresponding report. Pig movements are single reported on group level. A movement report for pigs contains date, number of animals, sending PPN and receiving PPN.

The database of animal holdings from the Swedish board of Agriculture is used in papers I, II, III, V and VI. The holding database contains coordinates, postal codes, county and production type (for example slaughter for a slaughterhouse) for the PPNs. Active farms and slaughterhouses for 2008 are shown in figure 3.4.

From SLV (the National Food Administration) we received data over number of slaughtered cattle and pigs for all slaughterhouses in Sweden for the years 1998-2007. At SLV slaughterhouses are identified by KS-number. We also received a list of addresses for the slaughterhouses. We matched the list of slaughterhouses from SLV to the slaughterhouses in the holding database using slaughter capacities, postal codes and geographical coordinates. Slaughter capacities were calculated from the transport database and compared with the SLV data. For slaughterhouses without coordinates in the holding database we used the address of their match to find approximate coordinates via the address search engines eniro.se and hitta.se.

The Forestry National Road database (SNVDB) together with the Krönt vägval system [25] were used to find the likely transport times and distances between farms and slaughterhouses in paper VI. SNVDB contains detailed information on all Swedish roads: state road network, the municipal road and street network and the private road network. The Krönt vägval system considers that drivers are likely to chose high quality roads and not necessarily the shortest ones.
Figure 3.4: Active farms (left, dots) and slaughterhouses (right, squares) in Sweden 2008. ArcView-shape files for plotting contours of counties in Sweden were provided by SCB.
3. Methods and data

3.7 Strategic planning with facility location model

Animal transports to slaughter were studied as a variant of a facility location problem in paper VI (for more information on facility location problems see for example [48]). In facility location problems customers (farms) and potential facilities (slaughterhouses) are located in space. There are fixed costs associated with using a facility at all. The problem is to decide which facilities to use and which specific facility each customer should use. The combination of distances to the facilities for the customers and the fixed costs are minimized. We minimized the transport distances for the animals and the slaughterhouses had limited capacities.

3.8 Route optimization heuristics

![Figure 3.5: Examples of route generation heuristics. Routes are more optimized to the left and more random to the right. To the very left Clarke and Wright heuristic and the three to the right are the good choice heuristic.](image)

Two heuristics were used to coordinate routes for transports-to-slaughter in paper I. The Clark & Wright heuristic starts by coordinating no routes and therefore placing all farms on their own separate route. It then iteratively coordinates the two routes that will decrease transport distance for the trucks the most. This means that the animals on two farms close to each other and far from the slaughterhouse are likely to end up on the same truck.

The other heuristic, which we refer to as the good choice heuristic, also starts by giving all farms their own route. It then iterates through the farms and connects the farm to up to two other farms. In each step if the farm is already connected to two farms it is in the middle of a route and nothing is done. If not, the next or previous stop in the route is picked, or both if the farm is still the only farm on the route. The distances from the specific farm to the slaughterhouse and to all farms in ends
of routes (with enough space in the truck) are calculated. The generalized normal distribution and the distances are then used to calculate the probabilities for the farms and slaughterhouse to be picked as next or previous stop on the route. If all probabilities are zero, the farm will keep its connections to the slaughterhouse. The behavior of the heuristic is controlled by the variance and kurtosis of the probability distribution. If low variance and $\textit{kurtosis} > 2$ are used the behavior will be more similar to the Clarke and Wright heuristic. If high variance and $\textit{kurtosis} = 4/3$ are used the coordination of farms will be completely random. See figure 3.5 for examples of routes generated with the two different heuristics.
4 Summary of included papers

I. Route planning reduces the costs of animal transportation: Animal welfare versus economics

We developed a method, *good choice*, to model coordination of transports on a scale from random to optimal using a probability distribution and compared it to the Clarke & Wright heuristic. The different transport strategies were tested in simulated landscapes. We also used the actual geographical coordinates of 22,657 farms and 68 slaughterhouses in Sweden and Euclidean distances between the facilities. All transport strategies were compared to the scenario of no transport coordination. We showed that Clarke & Wright heuristic could lower the transport distances for trucks by about 40% while increasing the transport distances for the animals by only 5% when coordinating in mean 1.8 farms per route. The Clarke and Wright heuristic gave 2-24% lower transport distances for the animals compared to the *good choice* algorithm. Our result also indicated that the small-scale slaughterhouses are very important for the animals travelling the longest.

II. Spatial and temporal investigations of reported movements births and deaths of cattle and pigs in Sweden

We investigated the national Swedish databases of pig and cattle movements and cattle births and deaths in Sweden from July 2005 to June 2006. Results can be used in modeling of disease spread, in design of disease control programs and for contingency planning. The maximum number of farms within 3km from a farm was 45 for cattle-farms and 23 for pig-farms. We found that most cattle and pig transports in Sweden are shorter than 100km. But long transports do occur and this means that when a disease is detected in one part of Sweden it could already have
4. Summary of included papers

spread to other parts of the country. We detected date preferences in the data, i.e. more cattle are reported to be born at the 1st, 10th and 20th each month and die at the 1st, 10th, 15th and 20th each month. Cattle purchase movement data are double reported, and 6.8% of the 515,572 reports were not correctly double reported.

III. Network analysis of cattle and pig movements in Sweden: Measures relevant for disease control and risk based surveillance

We calculated network measures for the transport networks in Sweden and introduced a new measure within veterinary epidemiology, the in-going infection chain, which we suggested as potentially useful to identify farms with high risk of getting disease via animal transports. Transport data from year 2006, 2007 and 2008 was used. The quality of data in the database is improving: 5.45% of purchase reports were not correctly double reported 2006, 2.86% in 2007 and only 1.24% in year 2008. Farms with low in-degree and high in-going infection-chain exist and such farms have many indirect contacts and few direct contacts. They could have high probabilities to be infected despite that the few indirect contacts suggest low risk for infection.

IV. Generating structure specific networks

We developed a method to generate networks of varying structures regarding the network measures mean degree, degree assortativity, clustering coefficient, fragmentation index and betweenness centralization index. Spectral methods were used to place nodes random or aggregated in a spatial landscape. Links were formed using a probability distribution with varying kurtosis and standard deviation. Unfortunately networks with negative degree assortativity could not be constructed.

V. SpecNet: a spatial network algorithm that generates a wide range of specific structures

The network generation algorithm developed in paper IV was extended to also be able to create networks with negative degree assortativity. This was done by dividing the nodes in two groups: focal nodes and regular nodes. The probability for links between nodes form different groups was increased. We compared the range of networks possible to generate with this algorithm to a variant of the non-spatial
CM (configuration model) network generating algorithm, and SpecNet produced a larger span of structures. Networks with low degree assortativity can be generated with the extended version of the algorithm. Two transportation networks for pigs in Sweden were reproduced with good results by SpecNet.

VI. A strategic analysis of slaughterhouses and animal transportation in Sweden

We did a strategic analysis of slaughterhouses in Sweden using a facility location model and transport and geographical data for 2008. The SNVDB (the Forestry National Road database) was used to calculate actual distance between farms and slaughterhouses. We minimized total transport work (total distance traveled by all animals measured in animal*km). We showed that strategic planning on national level could be used to reduce the total transport work by 40% for cattle and 25% for pigs in Sweden. The number of slaughterhouses was found to be not important for the total transport distance of animals in Sweden and about 50% of the slaughterhouses in Sweden could be closed with almost no effect on total transport work. The number of slaughterhouses is however very important for the animals that travel longest. The amount of animals traveling long distances could further be reduced with more mobile or regular slaughterhouses and by including animal welfare weighting in the objective function. The total transport work would decrease by more than 60% if animals were sent to the closest slaughterhouse compared to where they were sent 2008.
5 Discussion

In my thesis the national Swedish transport databases were studied, an algorithm for generation of networks was developed and optimization and models of animal transports were made.

The Swedish national animal transport database was investigated in paper II and III. The information obtained can be used to find realistic parameters as input to epidemic models. Similar investigations have been made of databases in other countries and there are now much data available, [61; 62; 63; 59; 52; 8; 50; 5; 14; 13; 60], to analyze and use. It is interesting to compare the transport data and models between countries as in [70] and relate it to data of disease occurrences as in [27; 69].

We found that the quality of the cattle movement database improved over the years 2006-2008 (paper III). Improvement over the years was also found for the Cattle tracing system in Great Britain [32]; they found a marked improvement 2001 when reporting became mandatory. In Sweden it has been mandatory to report since 1998 (cattle) and 2002 (pigs) and the fee for reports was removed 2008.

We found digit preferences in the cattle birth and death data, more cattle were reported to die and be born at the 1st, 10th and 20th and Robinson et al, [62], found the same in the corresponding data in Britain for the same dates. This means that all dates are not correct in the databases.

Various node and network measures have been suggested to be important for risk of spread of diseases [51]. Network measures to identify risk holdings and to find transport patterns with high risk of spread of diseases are interesting. However most of the epidemic simulations are made on static networks, not considering the time of the links [8]. Using dynamic networks or temporal network measures are two different methods to address same thing: the time of the transports is important. Directed networks and measures can also be more relevant than undirected ones because transports are directed. The directed measures in-degree and out-degree are
5. Discussion

often considered in network analysis of animal transports [51]. Weighted network measures accounting for the number of animals transported between nodes might also be interesting. Epidemics simulations in weighted networks has been shown to behave different compared to simulations in un-weighted networks [22; 15].

In paper III we suggest in-going infection chain as network measures that capture the risk for having an infection at a farm. In- and out-going infection chain considers both the direction and time of the links. Out-going infection chain can be used to estimate epidemic size [23]. In a recent paper [27] we showed that using in-going infection chain and in-degree to select farms resulted in finding more of the infected farms compared to random selection. Risk based selection is useful when the goal is to locate as many as possible of the infected farms.

Restrictions on the transport system affect the system itself [64]. For example using centrality measures and stop transports from or to holdings with high centrality can change the structure of the transport network and other holdings could take over the same centrality properties as the nodes cut-out. Studies on how to use the relevant network measures are needed in the future. More comparison with real data is also needed as well as comparison between epidemic simulations in static and dynamic networks to verify which of the more complex network measures are best to use.

We have developed a spatial network generation algorithm SpecNet (paper IV and V) that generates a large variety of networks. SpecNet can generate networks on a wide scale regarding the network measures: fragmentation index [0, 0.96], clustering coefficient [0, 0.81], degree assortativity [-0.98, 0.88] and mean degree, and was more successful than the version of CM algorithm we compared it to in paper V. Badham and Stocker present a variant of Keeling’s algorithm with the ability to construct networks with specified clustering coefficient and degree assortativity [7]. Short intervals of assortativity [0, 0.4] and clustering [0, 0.4] are considered but the possibility to also affect degree distribution is included [7].

SpecNet can be used together with SI (susceptible-infected) and SIR (susceptible-infected-recovered) models to test the effect of network measures as predictors of disease risks, both for the individual farms and for the system. SpecNet can also be used to replicate real transport networks, and two pig transport networks were replicated in paper V. Some of the parameters used in the algorithm can be estimated from real data; the kurtosis and variance can be estimated from transport data [46]. Using the $\gamma$ measured for the spatial distribution of farms in Sweden gave the best reconstructed real networks in paper V.

A method to model animal transports as they are carried out today (without optimization) is presented in paper I. Simulations showed that optimized transports
give 2-24% shorter transport distances for animals and 6-30% lower transport distances for trucks than the ones generated with the good choice heuristic. This is in agreement with the potential saving of up to 20% of transport distances found in an investigation of a slaughterhouse in Sweden [47]. The good choice heuristic is something to use when the routes we want to model are neither random nor optimized.

Paper I considers the trade-off between animal welfare and costs. With Clarke & Wright heuristic distance for the trucks is lowered by 40% and distances for the animals are increased by about 5% compared to all farms getting separate routes. These are very rough estimates. Geographical coordinates for farms in Sweden were used in the analysis, but the data was not complete. The number of animals on each farm was randomly drawn from a uniform distribution, because we had no transport data. In the year 2008 23% of the cattle transports to slaughter was just one animal, 20% two animals and 14% three animals; this affects the possibility for the transports to be coordinated.

I have rerun the heuristics (from paper I) on the entire transport data for 2008. For each day transports were coordinated with both heuristics and total distances for trucks and animals were calculated, see results and details in table 5.1. For cattle transports-to-slaughter the Clarke & Wright heuristic decreases truck distances by 61% (to 8 million km) and increases distances for the animals by 17% (to 55 million animal*km) compared to all farms on separate routes. The lower results in paper I might be caused by the distribution of number-of-animals used there.

Transports were also coordinated each week with the Clarke & Wright heuristic, see table 5.1. This means that animals transported on a Monday could be transported in the same route as animals transported on a Friday the same week. The differences in transport distances for day and week give an estimate of the potential gain with route optimization. The gains for cattle transports-to-slaughter are 12% shorter distances for trucks and 5% shorter distances for the animals. These estimates are within the intervals found in paper I. The routes actually made 2008 were not optimized and the time window for picking up animals at a farm is three weeks [Krister Svensson personal communication 2010]. This means that the estimated 12% decrease of distance for trucks is a lower bound.

Revisiting the 2008 transport data used paper in III and VI, I estimated mean transport distances for the different types of transports. I used Euclidean distances between farms and slaughterhouses multiplied with 1.23. The estimated mean transport distances for cattle were 16 km to pasture, 110 km to slaughter and 65 km for purchase transports. Estimated mean transport distances for pigs were 120 km to slaughter and 77 km for livestock transports.
5. Discussion

<table>
<thead>
<tr>
<th>Transport distances for trucks in Mkm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>no coordination</td>
</tr>
<tr>
<td>day&amp;week</td>
</tr>
<tr>
<td>Cattle purchase</td>
</tr>
<tr>
<td>Cattle slaughter</td>
</tr>
<tr>
<td>Pig livestock</td>
</tr>
<tr>
<td>Pig slaughter</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transport distances for animals in M animal*km</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>no coordination</td>
</tr>
<tr>
<td>day&amp;week</td>
</tr>
<tr>
<td>Cattle purchase</td>
</tr>
<tr>
<td>Cattle slaughter</td>
</tr>
<tr>
<td>Pig livestock</td>
</tr>
<tr>
<td>Pig slaughter</td>
</tr>
</tbody>
</table>

Table 5.1: Total transport distances for animals and trucks in Sweden 2008 for different coordination heuristics. The transport data from 2008 of from-were and to-were and number of animals was used. Distances between farms and to slaughterhouses were calculated as 1.23 times the Euclidean distances. Farms missing coordinates were assigned the same coordinates as a random farm in the same county. This concerned 1% of the pigs and 4% of the cattle transported to slaughter and 15% of the cattle in purchase transports. Routes were coordinated each week using Clarke & Wright and each day using Clarke & Wright, the good choice ($\sigma = 30$ km and kurtosis $= 10^3$) and the bad choice ($\sigma = 200$ km and kurtosis $= 1.5$). Coordinated routes could not be longer than 640 km ($8h \cdot 80$ km/h) for the animals. At most 32 cattle or 265 pigs could be transported on one route.

Paper VI considers geographic aspects of slaughterhouse capacities. Detailed real data regarding geographic coordinates, transport data from 2008 and distances between farms and slaughterhouses were used in the analysis. Strategic planning can decrease transport distances and as a consequence also transport times for the animals. We minimize the total transport work (total distance traveled by the animals measured in animal*km). Optimization can be done to minimize costs [16] instead of transport distances for the animals. Animals might travel longer in that type of approach, especially small slaughterhouses can be non-efficient from a cost perspective and suggested to be closed.
Transport optimization with animal welfare weights in the objective function is one way to especially decrease the number of long transports (done in paper VI). We show that it is possible to limit the amount of long transports and still get low total transport work. From animal welfare perspective there is much to be gained; with strategic planning transport-to-slaughter distances can be reduced by 25-40% for pigs and cattle. Mobile and small-scale slaughterhouses are important to decrease the number of animals that otherwise travel the longest, this is found in paper I and VI. The total transport work is not affected by the mobile and small-scale slaughterhouses because it is only a small amount of the animals that are and can be slaughtered at these facilities.

Conclusions

The national database of animal movements makes it possible to calculate and use network measures in the preventive veterinary work. The new measure in-going infection chain can be used to identify farms with many indirect contacts and high risk of getting an infection.

The SpecNet algorithm can be used to test the efficiency of network measures to predict risk for spread of diseases. More simulations and validation through comparison with real disease data are needed to decide which of the more complex network measures to use and how to use them.

There is a large potential to reduce transport distances for animals and trucks in Sweden using strategic and tactic planning. Software that considers queues at the slaughterhouses and can be used in real-time needs to be developed and applied.
References


REFERENCES


REFERENCES


REFERENCES


REFERENCES


