

Degree Thesis

Master's Level

Modelling of Social Networks, Analysis of Community Structures and Disease Simulation on Dairy Cattle

The aim of this study is to model a social network of dairy cows, detect communities and analyse the influence of these substructures on the spread of a contagious disease. By using this network, the transmission of a contagious disease will be simulated through a theoretical simulation function.

Author: Juliana Macedo and Mira Czarnetzki

Supervisor: Lars Rönnegård

Examiner: Siril Yella

Subject/main field of study: Microdata Analysis

Course code: MI4001

Credits: 30 ECTS

Date of examination: 2022-01-20

At Dalarna University it is possible to publish the student thesis in full text in DiVA. The publishing is open access, which means the work will be freely accessible to read and download on the internet. This will significantly increase the dissemination and visibility of the student thesis.

Open access is becoming the standard route for spreading scientific and academic information on the internet. Dalarna University recommends that both researchers as well as students publish their work open access.

I give my/we give our consent for full text publishing (freely accessible on the internet, open access):

Yes

No

Dalarna University – SE-791 88 Falun – Phone

+4623-77 80 00

Abstract

Past research has demonstrated that social networks and community structures have a strong effect on the dynamics of how an infectious disease spreads, although many have focused more intensively on the impact of the number of connections an individual has. This thesis investigated how disease transmission is affected by the social networks' structures such as communities, centrality measures, cliques, and diameter. For this purpose, a farm located in Sweden with around 200 cows will be used and two social networks were built based on spatial data from a period of 24h. A hypothetical model was built to simulate disease transmission with two scenarios: one when there was immunity and another without immunity. The main hypothesis presented here is that if the focus of infections starts on an individual located in the largest community, it would result in faster and larger spread of the disease, the betweenness centrality was used in these communities to choose the individual to start the focus of infection. It is also investigated whether there was any preferential aggregation due to parity, which is related to the number of calves a cow has had. It was found that the size of the community did not have a high influence in the rate of infections, meanwhile the overall centrality in the whole networks and the presence of certain individuals in the cliques seemed to play a bigger role. The results indicated some preferential relationship due to parity, although it is unclear if there is a specific parity that tends to aggregate more than others.

The two networks had different structures, making it difficult to generalize results and make a recommendation to farmers.

Keywords

Community Detection

Dairy cattle

Disease simulation

Social Network

Table of Contents

Abstract.....	1
Keywords.....	1
1 Introduction	4
2 Background overview.....	7
2.1 A Brief Introduction on Social Networks.....	7
2.1.2 Social Networks: Animals & Cattle.....	9
2.2 The importance of Parity in disease transmission.....	10
2.3 Background Overview of Contagious Diseases in Cows.....	11
2.3.2 Contagious Disease Simulation in Cows	12
3 Data	14
3.1 Description of Simulated data	16
4 Methodology.....	18
4.1 Network Modelling	18
4.1.1 Pre-processing.....	18
4.1.2 Definition of thresholds	18
4.1.3 Construction of the networks	19
4.1.4 Community Detection	19
4.1.5 Estimated Metrics	19
4.1.6 Secondary Hypothesis.....	19
4.1.7 ERGM	20
4.2 Disease Simulation	21
4.2.1 Without Immunity.....	21
4.2.2 With Immunity of one cow	24
5 Results.....	25
5.1 Social Networks.....	25
5.1.1 Descriptive characteristics of the networks.....	25
5.1.2 ERGM analytical results	27
5.2 Disease Simulation	29
5.2.1 Results Disease Simulation without Immunity	29
5.2.2 Results of Disease Simulation with Immunity.....	30
5.2.3 Results comparison to right side of the barn.....	32
6 Discussion & Conclusion	34
References	38
Appendix A.....	42
Figures related to Disease Simulation	42

Appendix B	46
Diameter & Neighbors	46
Neighbors	46
Appendix C	46
Outcomes from ERGM	46

1 Introduction

The dairy industry is very important in Sweden. Measured by the total production value, animal production in the dairy industry accounts for around one third of the animal production in Sweden. Between 2012-2016, the dairy and egg export made up an average of approximately 9% of the total food exports. Swedish milk production is the most efficient one in Europe, with the highest milk yield in the whole of Europe. Each cow produces around 9,500 kg of milk each year, which is around 30 kg each day. This allows for the approximate 3,500 farms that have an average of 92 cows, to produce this substantial amount of dairy. A combination of specialized breeding, but also improved animal welfare and health has allowed the Swedish Dairy Industry to reach this high level of efficiency (Jordbruksverket, 2017; Lassen, 2013).

At the same time, Sweden has some of the highest regulations surrounding animal welfare, climate conditions and feeding regarding dairy cattle cows. Sweden distinguishes itself in the farming industry with these high standards for animal welfare and health. As such, Sweden has been able to eliminate several animal diseases that are still common in Europe. Swedish animals are amongst the healthiest, all while Swedish Farms are using the least amount of antibiotics compared to other farms in Europe. However, Sweden has not been able to eliminate the disease with the highest negative economic impact in the European dairy industry, known as Mastitis (Jordbruksverket, 2017; Lassen, 2013).

Mastitis is an inflammation of the udder and it directly affects farmers economy as milk is discarded as a result and cows can be retrieved for treatment for several days, depending on the severity. If the severity is not high, cows are usually treated with antibiotics and continue within the group, elevating the chances to infect those who are healthy. It can spread through both contagious and environmental pathogens, but the exact way in which it can be transmitted from one cow to another remains unknown (Lundberg, Nyman, Aspán, Börjesson, Unnerstad, & Waller; 2016).

Instead of focusing solely on Mastitis, this study aims at investigating how a hypothetical contagious disease spreads based on the proximity of individuals. Even though Mastitis is the disease with the biggest economic impact in the dairy industry, still little data on the infection probabilities exist. Therefore, the executive decision was made to base the simulation on a hypothetical disease for which the parameters can be easily adjusted, so that the simulation can be used by other researchers or farmers to their own case study. This allows the simulation to be more applicable in a wider range of cases. The expectation is that understanding it and considering that cows are gregarious animals, it would help the farmer to act in advance or maybe retrieve those cows with low severity from the group, avoiding further infections.

Similar to humans, cows are social animals. They form long lasting relationships that are stable over time. While the initial group dynamics and hierarchy needs to be established upon first contact, after that, the herd will follow a strict social structure. This social structure is formed by dominance, subordination and preferential relationships between the cows (Rocha, Terenius, Veissier, Meunier & Nielsen; 2020).

Considering this persistence in sociality among cows, it is opportunistic to use Social Network Analysis to better understand the contact structure and therefore, help prevent the transmission of infectious diseases.

Social networks have been used in a variety of fields: math, psychology, medicine, anthropology among other areas. They have been largely used in studies to help understand interaction among animals and the effects of those interactions on the dynamics of disease transmission (Chen, White, Sanderson, Amrine, Ilany, & Lanzas, 2014; De Freslon, Martínez-López, Belkhiria, Strappini, & Monti, 2019; Scott, 2011).

Independently of the specific way in which a disease might be transmitted, whether by direct contact or close proximity, social networks have proven to be a useful methodology in the identification of contact structures. As such this methodology allows for the identification of individuals who are more influential, also understood to be more central to a network structure. Therefore, influencing the infection of others to a larger degree in the case of an outbreak (Summer, McCabe, & Nunn, 2018).

Social network structures, such as node degree, clustering coefficients, centrality, modularity, and subgroup cohesion help understand the dynamic of disease transmission as it explains the structure of the network and the net of contacts in it (Summer et al., 2018). These concepts will be explored in detail in the next section.

Once a network is formed, it is possible to identify, using different algorithms, the formation of communities (those groups intentionally formed that hold some kind of cohesion through time). The degree to which contacts happen inside a community compared to the contacts among communities is called modularity. An increase in modularity is expected to reduce transmission possibilities, as individuals tend to limit their interaction to a smaller number, due to time constraints (Summer et al., 2018).

Therefore, this study aims to investigate the substructure of communities and its influence on pathogen transmission by infecting different individuals inside different communities at a certain time and observe the dynamic of spread, based on the network structures.

This Thesis will analyse data collected from a farm located in Sweden, which has two groups of approximately 100 cows each. Based on the position and activity recorded, per second, of each individual animal, two social networks will be constructed and its substructures extracted, such as communities. The networks will be built based on the daily cattle position, using the thresholds of distance as 1.25 meters and minimal time of contact as 30 minutes per day based on (Rocha et al., 2020).

By initiating the infection spread from different cows, the importance of centrality on the spread of the contagious disease is analysed. These simulations are built upon the constructed social network data identified in the first part of this paper. The type of contagious disease simulated follows a "Susceptible-Infectious-Recovered" disease structure common in epidemic simulation studies when distance and network structure are influential. To further assess the impact that both the nature of the disease and the network structure has on the dynamics of the spread, this thesis supported the investigation with different scenarios. These scenarios played around with the origination of the

disease, as well as assessing the role immunity of specific important cows in the network could have on the spread compared to no immunity (Mukhamadiarov, Deng, Serrano, Priyanka, Nandi, Hong Yao & Täuber; 2021).

The aim of this study is to model a social network of dairy cows, detect communities and analyse the influence of these substructures on the spread of a contagious disease. By using this network, the transmission of a contagious disease will be simulated through a theoretical simulation function.

The main hypothesis is that if the infection is originated in a larger community it will result in larger spread throughout the network.

2 Background overview

2.1 A Brief Introduction on Social Networks

As mentioned in the introduction, the theory of social networks has been used in different areas. The idea of social networks emerged around 1930 within the field of sociology as an approach to think about social structures. Jacob Moreno, who was a psychiatrist was the first to introduce the visualization of social structure through network diagrams with points and lines. Graph theory, which is the most used approach so far in social network analysis initiated in mathematical field. It proposes that individuals or groups, and their social interactions can be depicted through sociograms (Scott, 2011).

Graph theory can be explained by theorems that analyse the properties of these sociograms. This theory depicts the network using edges, which are the lines representing the connection between individuals, and nodes representing each individual. It uses undirected, directed, and valued data to present the measures that help understand how the connections are made. A network is undirected whenever the direction of contact is not relevant, the directed network in the other hand, takes into consideration the direction of contact, who initiates it and who receives it. Valued or weighted data contain extra information that clarifies the strength of the relationship between two points (Scott, 2011).

A social network can be depicted as a classic sociogram, where individuals and groups are represented by points and their social interactions by lines connecting these individuals and groups (Scott, 2011).

A fundamental point of social network analysis is the investigation of community structure. The formation of communities is a common and observable trait among humans, animals, and insects. Communities can be described as any type of aggregation among individuals with the objective to achieve a certain task (Scott, 2011; Missaoui & Sarr, 2014;2015).

In social networks, communities describe the formation of clusters as subgraphs. It is defined as the formation of groups where there are more edges connecting the nodes of the same group and fewer edges connecting those from different groups. Edges are known as the links and connections between a pair of nodes, and nodes are the representation of the members of the network (Missaoui & Sarr , 2014; 2015).

Some studies have demonstrated that understanding community formation is crucial when studying infectious diseases transmitted by close contact. Communities can affect infection rates as it influences the likelihood of infection: in communities with high modularity, once a member inside a community is infected, it is more likely this individual will infect the other members of the same community than the outsiders. The strength of community formation is measured by the modularity it presents. Modularity is given by the degree of contact within versus between communities (Salathé & Jones, 2010).

Through social network analysis it is possible to identify the path of connections, those individuals that have more connections, those that could influence the most, formation of subgroups such as the communities and largest cliques. All of this can help understanding the dynamics of disease

spread, since it gives a map of the contacts made, an indication of the flow of the spread through the net of contacts once it is possible to identify who is connected to who and those who, for example, make the bridges between one community and the other.

There are different metrics that can be taken from a network which helps understanding its structure. Centrality measures, for example, are essential in understanding the dynamics inside a social network as they indicate power and influential relationships. In this study, the betweenness centrality is used, which is the degree of shortest paths that goes through a node, to analyse those individuals most central in their communities. (Scott, 2011).

The concept of cliques is also being analysed, more specifically the largest cliques, which are subgroups inside a graph where each node is connected to all other nodes inside that subgroup. This also has relevance when studying disease transmission through networks, a study made on the influence of contact structure of networks and transmission rates highlights that in networks with many cliques, diseases could have high transmission rates and shorter periods of infections (Read & Keeling , 2003).

The degree in which the nodes associate based on their connections is called assortativity and it is also an important structural metric. It calculates the probability of a connection between two individuals nodes happening based on the number of connections they have. It indicates whether those with many connections tend to associate with others also with a higher number of connections. Understanding this, can help prevent outbreaks, some studies have found that if assortativity is high, it increases the probability of an epidemic to happen (Badham & Stocker , 2010).

Equally, transitivity is also a metric that helps understand the impact of these social structures in disease transmission. This is also known as clustering coefficient and measures the degree in which neighbours of a node are also neighbours of each other. High transitivity indicates that networks have groups that are densely connected internally, this metric ranges from 0 to 1. Although research on the impact of this metric on epidemics is limited, some studies have found that when transitivity is high, it is less likely that an epidemic will occur (Badham & Stocker , 2010).

The path of these connections inside a network is another structure that will be investigated. One measurement of these paths is called the diameter, which is the longest path length. In other words, the path that connects one side of the network to the other side. In networks with smaller diameters, the members are connected through fewer intermediates, making the transfer of a disease potentially faster than in groups with bigger diameters (Wey, Blumstein, Shen, & Jordán, 2008).

For better understanding, a summary of the basic concepts that are analysed in this work is presented in Table 1.

Table 1*Summary of metrics definitions*

Metric	Definition
Degree centrality	Number of ties a node has
Centralization	The degree in which the network is centralized around one node
Betweenness centrality	Number of times a node lies on the shortest path between other nodes
Density	How much of all possible ties are present in the network
Cliques	A subgroup of a network where all members are connected to each other
Homophily (Assortativity)	Correlation of nodes with similar degree or attribute. It ranges from -1 to 1, negative values indicate that the relationship between nodes of different characteristics is stronger, positive values indicate relationship between nodes with similar characteristics
Transitivity	The degree in which neighbours of a node are also neighbours of each other
Articulation point	Is that node if removed would disconnect the whole network
Diameter	The path that connects one side of the network to the other side
Modularity	Measure how much the individuals tend to interact with their subgroup members proportional to the entire network
Community	Subgroups of a graph

Individual traits can also influence this dynamic as it could indicate whether certain individuals in a group tend to group together based on their personal attributes. To analyse this, the Exponential Random Graph Model (ERGM) is used, which is a probability model similar to a logistic regression. The idea behind using this model is that it will allow us to estimate the probability of a contact to happen given an individual attribute (de Freslon et al., 2019).

2.1.2 Social Networks: Animals & Cattle

Group associations are known to be positive within different animal species. It protects the groups from predators and improves the defence of food resources, for example. On the other hand, life in groups can increase the risk of disease transmission through closer or more frequent contact among the members of a group (Sueur, Jacobs, Amblard, Petit, & King, 2011).

In almost all species in which individuals live in flocks or are organized in communities, the aggregation between them is known to be purposed, intentional, not random. Cattle is a common type of large domestic animals that have proven to associate through certain characteristics as dominance and subordination, and preferential relationships (Suer et al., 2011; Rocha et al., 2020).

The level of social interaction among animals is defined as the relative preference of an individual to stay in close contact with a member of the same species. In cattle, this behaviour is described by a tendency to form and maintain groups. Gibbons, Lawrence & Haskell (2010) made an experiment where they measured sociability of dairy cows on commercial farms by comparing the response of cows in a runaway test to those from observations of spontaneous behaviour. The overall conclusion of the study is that there was a cohesion interaction that was repeated through the experiment, showing that aggregation among cows were not random. They also defined thresholds

for the distance between cows in the runaway test and found out there was a tendency of the animals to group.

Rocha et al. (2020) conducted research with cattle in order to understand whether the sociality factor depended on individual characteristics or the composition of groups. One of the elements pointed out in the study is that for clusters where there were no changes of individuals inside the group, the sociality was preserved. It suggests that cows tend to establish connections with specific partners when the population is fixed.

Summer et al. (2018) have evaluated some substructures of social networks such as modularity and subgroup cohesion in relation to the network size and its effect on pathogen transmission. They tested seven different algorithms for community detection in small-scale primate social networks. It was concluded that as the size of a social network grows, the subgroup cohesion becomes stronger as the individuals tend to limit their interactions. The direct impact observed of this on pathogen transmission is that when simulating the transmission, they could observe that in those networks where the modularity was higher the transmission rates were lower. Hence, there were more communities or groups formed, and the subgroup cohesion was high, meaning that individuals would not change groups.

Another study using social networks conducted on dairy cows aimed to understand whether contact patterns were influenced by individual traits. The network was constructed based on direct observation of contact among cows, and estimations for disease transmission were made using ERGM. They highlighted that there was a high level of interaction between cows in oestrus and male calves, which could play an important role in an outbreak. They also mentioned that cows interacted mainly with those of the same age, pointing out the importance of including individual traits in social network analysis (de Freslon et al., 2019).

Chen et al. (2015) presented a research also conducted with cattle, but this time using a spatial-temporal dynamic social network. Their proposal was to analyse cattle's position gathered by using a real-time location system at an hourly level. They constructed different networks divided by location inside the barn and by the hour of the day. In this study no simulation of disease transmission was made. General findings of this study were that network structures changed substantially depending on the time of the day and the location. And that proximity among cows at a dyadic level (i.e. for pairs of individuals) was more consistent in certain areas such as where they fed themselves or in the hay.

It has been proven that strong social networks in dairy cattle cows have the ability to reduce stress and speed up recovery from diseases. Therefore, not only is the sociality important from an animal welfare perspective, but also from an economic standpoint. With swifter recovery come reduced costs to the farmer, as less treatment costs are encountered, and milk discarded (Rocha et al., 2020).

2.2 The importance of Parity in disease transmission

This study will differ from others in the sense that it will focus on the substructures of the network and only one individual trait which is the parity number. This Thesis aims to

understand whether this trait influences somehow in the way the analysed group forms its communities.

Parity is the number of times a female cow has had a calf. This is an important trait, because it directly affects milk production. The period between one calving and the next is known as the lactation cycle, and during this period the yield of milk increases. In terms of milk production, a cow should ideally calve every 12 months for optimal milk yield (Global AG Media, 2015).

Some diseases are known to be connected to the calving periods, such as mastitis, lameness, and other metabolic disorders (Sehested, J., Gaillard, C., Lehmann, J., Maciel, G., Vestergaard, M., Weisbjerg, M., . . . Kristensen, T. , 2019). Cows are at highest risk to be infected with clinical mastitis, which is a contagious disease, around calving and in early lactation. In a recent study made in Sweden, it was found that in herds with mastitis problems, some pathogens that cause the disease were mostly commonly found just after calving (Lundberg, 2015).

2.3 Background Overview of Contagious Diseases in Cows

2.3.1 *Why network structures are important in disease transmission*

Social structures of individuals and animal populations are at the centre of disease transmission. Even more so for communicable diseases, such as contagious diseases, which have a higher probability of infection from a prolonged period of contact or closer contact to an infected individual (Rossi, De Leo, Pongolini, Natalini, Zarengi, Ricchi, & Bolzoni, 2017).

The dynamics of how an infectious disease will spread directly depends on the underlying contact network of the host. Understanding the influence that the community structure plays on the epidemic spread of a disease, may help researchers provide insights on how changes to the network can influence the spread of the disease. These changes may result from intervention strategies such as isolation of the infected individual, antibiotic treatment or other pharmaceutical or non-pharmaceutical interventions. Research has shown that these interventions are more effective in a strong community structure compared to only targeting a single highly connected individual (Salathe & Jones, 2010).

As outlined earlier, the contact structure of a population depends on behaviours, which include the proximity, time, animal movement and social interactions. These social contacts are unavoidable due to the nature of the spatial and temporal concentration within the barn, but also considering the engrained DNA calling for social contacts between livestock. For group-living species, such as the cattle cows the researchers are considering, the transmission of diseases through these social connections embodies the cost of living in groups (Craft, 2015).

Using the constructed network data for network modelling of diseases or epidemics, first gained traction when understanding sexually transmitted diseases. Later, the Severe Acute Respiratory Syndrome (SARS) outbreak in 2003 led researchers to understand the importance of this methodology in analysing the spread of a disease. Network modelling has crystalized itself to be especially important when considering targeted interventions, as as simulated in this study (Chaters, Johnson, Cleaveland, Crispell, de Glanville, Doherty, Kao, 2019; Craft 2015).

Highly contagious infectious diseases are defined by having a high impact on morbidity and mortality of both domestic and international animals. As such they are a huge economic burden on farmers and society, as they result in treatment costs and loss in production. They also present a large risk to the animal welfare (Saatkamp, Mourits & Howe, 2016).

The European Union aims to ensure a high animal welfare. As such, they have implemented a strategy of non-vaccinations for most contagious livestock diseases, which has resulted in a highly susceptible livestock. The risks of introducing such a susceptible disease are directly determined by the direct and indirect contacts of the cows. Therefore, influential factors on the disease dynamics are not only determined by the disease directly, but also features related to the livestock structure such as the farm size, the number of farms in proximity, livestock movement and many more. These factors will influence the magnitude of the disease spread and thus the economic impact. A long time frame will increase the uncertainty in respect of the driving forces and impact that these factors have on the disease simulation, while a short time frame will be insufficient in illustrating the impact the driving forces have on the simulations. Keeping this in mind, researchers have used time periods between three to nine weeks in the past (Boender, Nodelijk, Hagenaars, Elbers, de Jong, 2008; Hop, Mourits, Oude & Saatkamp, 2014).

Research has shown that the spread and simulation of contagious diseases is complex. Various internal and external, controllable, and non-controllable factors influence the dynamics of the spread. Nonetheless, a higher disease spread has been identified when cows are more densely concentrated in certain areas. In these situations, the rate of the spread is faster than the rate at which animals can develop immunity, on a natural way or through interventionary methods, such as vaccines or removal of the cow from the herd (Hop et al., 2014).

Understanding how the infectious disease spreads through the population is at the center of controlling them. Computational simulations, like those used in this paper, provide a key element in the study of these disease dynamics. Contagious diseases need to be identified and controlled quickly to limit the spread to unaffected areas in the network (Hop et al., 2014; Salathe & Jones, 2010).

2.3.2 Contagious Disease Simulation in Cows

The use of network data in understanding the transmission dynamics of contagious diseases is a common practice used for different livestock. It allows researchers to identify the importance of risk-based surveillance in the interest of both the animal welfare and economic outcomes (Craft, 2015).

In dairy cattle, Mastitis is the most common disease not only in Europe, but worldwide. As a Mastitis infection results in the milk being discarded and, depending on the severity of the infection, the cow needing antibiotic treatment, it also established itself as the costliest disease to farmers worldwide. Due to Mastitis being the costliest diseases in the dairy industry worldwide, it is of high interest to reduce infections both from an economic, but also from an animal welfare perspective. The severity of Mastitis can be subdivided into two categories:

I. Clinical Mastitis (CM) – clear visual signs of inflammation in the milk and udder

II.Subclinical Mastitis (SCM) – inflammation without visual signs present (Pettersson-Wolfe, Mullarky, & Jones, 2019)

Mastitis can be caused by a large variety of different pathogens and bacteria. These include both environmental and contagious pathogens, some can even spread through both methods (Burvenich, Van Merris, Mehrzad, Diez-Fraile & Duchateau, 2003).

While the initial aim of this study was to provide a simulation solely on the spread of the Mastitis disease, a lack of available research on the influential parameters causing the spread of the disease resulted in a generalized approach on the simulation of a contagious disease amongst the social network of dairy cattle cows. However, the nature of the simulation thus allows for a wider application of the model, as farmers and researchers can input their own parameters into the model for an individualized simulation. The connection of this contagious disease simulation with the Mastitis disease further highlights the relevance of the simulation in the real-world application.

The rate of inter-herd transmission of contagious diseases differs. Not only on the contagious disease itself, but many complex factors some controllable and measurable, others non-controllable and at parts non-measurable. The probability of infection is influenced, but not limited, to the following factors (Burvenich et al., 2003; Craft, 2015; Hop et al., 2014; Pettersson-Wolfe, Mullarky & Jones, 2019):

- Hygiene factors implemented by the individual Farm
- Contact time and frequency of the cows with the personal of the Farm
- Cleanliness of the workers
- Temperature of the surrounding within and outside the Barn
- Days in milk of the cow
- Nutrition and general health of the cow
- Stress of the cow
- Inter- and between-farm movement of cows
- Farm concentration level within the area
- Contact time
- Distance between the susceptible and infected individual

This exemplifies the difficulties encountered by others researches and ourselves in defining parameters surrounding the disease simulation. Nonetheless, best attempts were made at basing our simulation parameters on the previous research available. However, the specific way in which the parameters are decided upon for the simulation model are explained in detail in the methodology section.

3 Data

In response to the Bovine Spongiform Encephalopathy, more commonly known as the mad cow disease crisis, in the late 1980s, the European Union has implemented more elaborate controls within farms of dairy cattle cows. An element of this is the extensive capturing of movement data through location systems allowing for the reveal of existing social network data (Rossi et al., 2017).

This type of location data is also what this study is based upon. This dataset is comprised of location coordinates of each individual cow in the barn. It collects and stores data for every second of the day.

Data used in this study was collected as part of another research project within the Swedish University of Agricultural Sciences (SLU), which has been gathering data since November 2019 (Swedish University of Agricultural Sciences, 2021). For this study, only one day will be used for the social network analysis: 16/10/2020. The reason why this date was chosen is because it is also part of a research period conducted by a researcher from SLU, and it could provide interesting information as basis for comparison to her work. This researcher is analysing a period of 14 days, while this project will analyse just the first day of that period.

The collection of data was made through an ultra-wideband system (UWB), which is known as being a reliable and accurate technology of indoor positioning. The investigated farm uses a system called GEA CowView, which is an automatic indoor localization system for dairy cattle that provides data on position and activity of individual animals per second (Ren, Nielsen, Alam, & Rönnegård, 2021). In order to get the location of each animal, the system collects information through different sensors that transmit the information using radio signals. These signals are collected through fixed placed sensor and those carried by individual animals. Each animal carries an accelerator and a gyroscope that detects their movement, called the collar tags. The system then calibrates the real-time position measured by these two types of sensors and calculates the precise position of the animal tag (U.S. Patentnr 10,234,535, 2019).

In the case of the studied barn, the collar tags transmit the individual movement signals to 8 different static sensors located in the ceiling of the barn (see anchors in Figure 3(a)). Data then, is collected and pre-processed through a chain of data-processing modules in the system and transmitted to a GEA database. It generates a file with a tag id, timestamp and the (x,y) coordinates for each individual (Ren et al., 2021).

Figure 3 (a) illustrates the structure of the barn from which the data of this paper is used. The farm is located in Sweden, it is non-insulated free-stall and has a rectangular plan of 74m x 33m. There is a feeding area on both sides of the barn towards the outer walls. The cows are kept in two separated groups, each one with access to one of these feeding alleys. There is a milking parlour inside the barn and the cows are milked twice a day. Food is delivered 12 times a day and is available at any time. The structure in the middle separates the two groups stopping them from mixing or coming into close contact. Therefore, for the purpose of the analysis, two separated networks are built (Ren et al., 2021).

The cows are purebred Holstein Friesian or Swedish Red and crossbreds. From each cow, the following records were collected: the animal's identification number given by the tag, their location

per second through the period of 24h, and the parity number. In total, there were 203 cows, 94 cows on the left side and 109 on the right side.

The parity number in the dataset is characterized by numbers ranging from 1-7, meaning that a cow with parity number 1 has had only one offspring, a cow with parity number 2 has had two and so on. Since the majority of them are in between parity 1 and 4 (see Fig.1 and Fig. 2), it was decided to join all parity numbers above 4 in one category, which is referred to as number 4 in this study to facilitate some of the analyses. In other words, the parity number 4 will contain the sum of the number of cows with parity between 4-7.

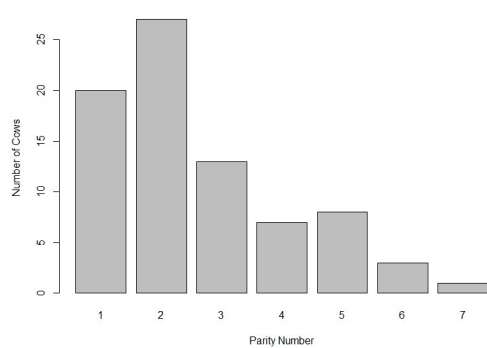


Figure 1

Frequency of Parity numbers – left side

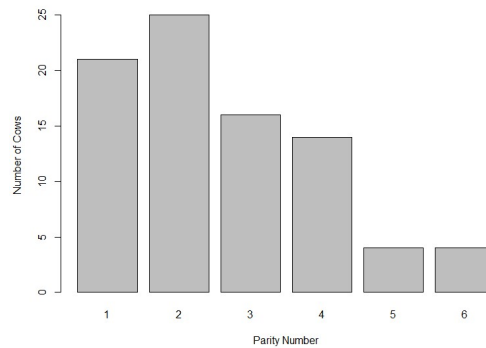


Figure 2

Frequency of Parity numbers – right side

The area on the lower part of the barn, which is a rectangular plan of 26m x 33m, is used to keep separated from the group those cows who are in the dry-off period, until they have a new calve. According to research, a dry period is necessary for milk production. This period is characterized as a pause in the milking process and once they enter this period they need to be removed from the herd. This period can vary between 50-70 days before calving (Amaral-Phillips & Franklin, 1999). There was a total of 11 cows in dry period which were excluded from analysis.

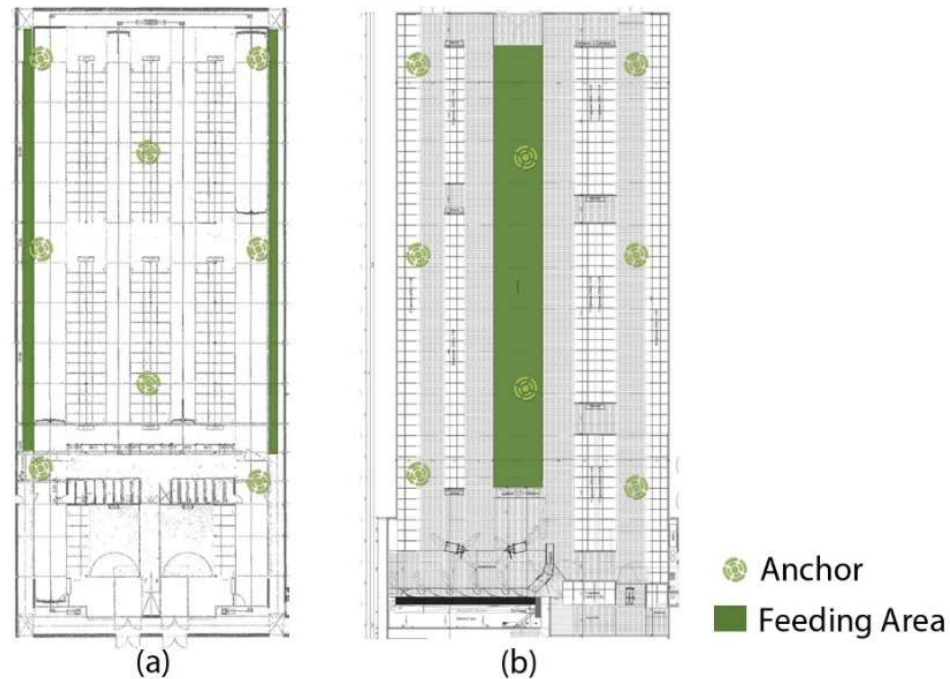


Figure 3
The Layout of the Barn from which the data was collected (Ren et al., 2021)

Note: On the left side is the Swedish barn, which is object of this study. On the right side is the Dutch barn which is not part of this study, but it is part of the SLU research.

3.1 Description of Simulated data

For the second half of this paper, in which a disease transmission is simulated, the network data produced in the earlier half of this paper was used. The igraph data frame that was constructed using the methodologies explained above, was used to run the disease spread simulation. This network dataset stored information about which cow was connected to which other individual, their degree, the eigen and the betweenness centrality on an individual ID level. As such, this was used to build an adjacency matrix of the connections between the cattle. The constructed matrix allowed for the disease simulation to be implemented on a static network. As such, the simulation model assumed that no births, deaths or migrations occurred within the social network during the simulated disease spreading period. The network was constructed based on one day of data, as previously explained. However, the simulation is extrapolating this data onto a 50-day simulation, as reasoned from the background research and further discussed in the methodology. For each scenario, the simulation was run 500 times and the average data per day was used for the analysis between the different scenarios. Each scenario constructed can be followed in Figure 4. It illustrates for which identified cows scenarios with and without immunity were analysed.

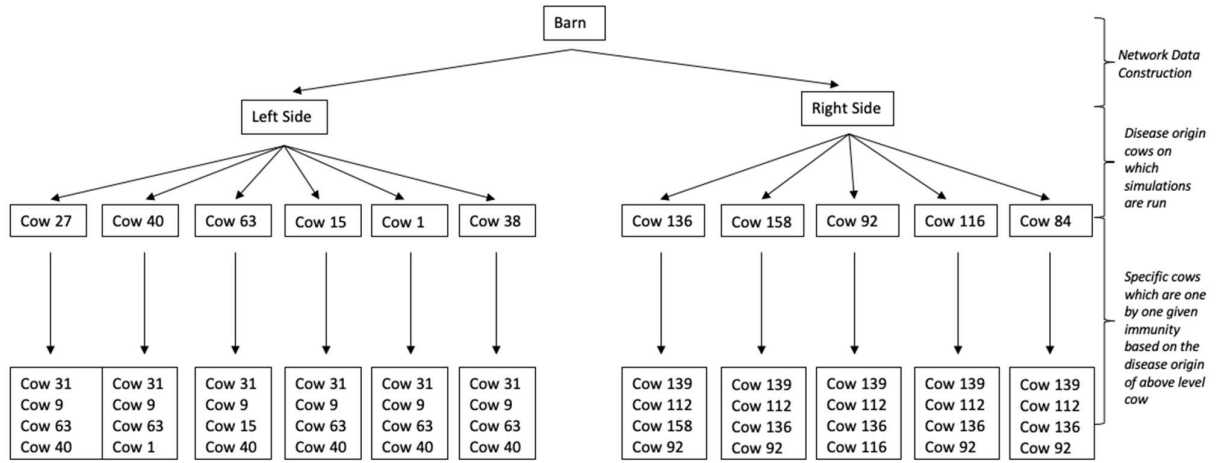


Figure 4
Overview of all Simulations Performed and Analyzed

4 Methodology

4.1 Network Modelling

A graph of a social network can be denoted as $G = (V, E)$, where V is a fixed set of vertices or nodes, and E the edges or connections between these nodes, represented by lines. The edges are a representation of a set of pairs (i, j) , where i and j are both indices between 1 and the number of elements. An undirected graph is a set of unordered pairs of vertices (Adhikari & Dabbs, 2018). In this study, two social networks are built, both undirected. Undirected graphs were chosen because information concerning which cow that takes the initiative in a contact was not available.

The `igraph` package in R (Csardi & Nepusz, 2006) was used to build the networks and the ERGM package (Handcock, Hunter, Butts, Goodreau, Krivitsky & Morris, 2021) present in the Statnet suite (Krivitsky, 2021) was used to analyse the influence of individual attributes in aggregation.

4.1.1 Pre-processing

Before starting to work on the networks, a program built as part of the research within SLU, mentioned earlier, was used and extended, in which data has been pre-processed (Otrante-Chardonnet, 2021). This code was built in python and it allows the creation of data frames and sub-data frames given the original csv file (raw data) and the conditions the user wants: such as the date, time, and location of a given pair of cows. In summary, it does the following steps to pre-process data: 1. Reads the raw data file and drops all inactive tags, those from which movement cannot be detected. 2. Fills in missing data by interpolation using `Akima1DInterpolator`, present in the `SciPy` package in python (Virtanen, Gommers, Oliphant, Haberland, Reddy, Cournapeau, & SciPy 1.0 Contributors, 2020). 3. Calculates the distances between every pair of cows based on their position and the barn coordinates. 4. Creates a data frame with a rounded time for each pair. 5. Calculates the aggregated distance for each pair of cows. 6. Calculates the position for each pair of cows, the time and distance based on the location of the barn (bedding, feeding areas, etc). 7. Create an interface that allows users to customize the outcome by area, time, and print a histogram for each pair.

To this program, a function that creates a subset for the left and another for the right side without subsetting it by area, was added. That means it considers all positions contained in the raw data for this specific day. I was also added another function that creates a data frame aggregating the distance for each pair of cows for every 0.5m. By using these two subsets a program in R was developed with `igraph` package to create the two networks (Macedo, 2022).

4.1.2 Definition of thresholds

In order to build the network, the social interaction among two individuals was quantified using the distance between them and the total time they spent in the company of each other in a period of 24h. For the distance, the threshold of 1.25 meter was applied, meaning that we excluded those interactions in which each cow was located being more than 1.25 meter apart from each other. For the time they spent together, the threshold of minimum 30 minutes was applied, which excluded those interactions that lasted for less than 30 minutes (summed up the total seconds for the whole period).

The thresholds were chosen based on the literature research. According to Gibbons, Lawrence & Haskell (2010) social contact among cows can be characterized if they tend to group in distances of at least 5m. Regarding the time, we tried to avoid those passing-by behaviours and looked for more long-lasting contacts during the day.

The distance is used to define the edges of the network and the strength of the connections is given by the total time, which is calculated as the sum of seconds they are in the company of each other during the whole period.

4.1.3 Construction of the networks

To create the network, a list with all nodes and a list of edges was initialized. The list of nodes is constructed by taking all unique tag ids, which initially sum up to 94 for the left side and 109 for the right side, excluding those with missing tag strings, which are used to identify the cows and match them with other individual attributes. This list is then matched with the parity number and those tags whose parity information was lacking were dropped. The individuals that are located in the area dedicated for dry period were identified from this list and exclude (those which the average location in the y-coordinates is below 2.595 cm for the whole day). It results in two undirected networks: on the left side, containing 79 nodes and 268 edges and on the right side, containing 81 nodes and 350 edges. The diagram of each network is showed in the results section.

To the nodes it was added an identification number on the left side ranging from 1-79 and on the right side ranging from 80-160, in order to facilitate the analysis.

4.1.4 Community Detection

To detect communities, the Louvain method was used through the available package in R (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008), which is a method focused on modularity optimization. After detecting the communities, subgraphs for each one of them were created and the betweenness centrality was calculated in each community among the participant nodes, in order to identify what we will call “The Most Important” cows, which are those with higher betweenness centrality in their communities. This measure will differ from the overall betweenness centrality of the graphs, as those will be taken over the whole graphs, meanwhile the one mentioned is going to be taken over each subgraph of community.

4.1.5 Estimated Metrics

In terms of metrics, the main network descriptive metrics such as the density, diameter, degree of connections, betweenness centrality, assortativity, transitivity, diameter, and cliques are estimated (concepts already presented in Table 1). All calculated with the igraph package. The network's measures and dimensions were also compared to the average result of 1000 random graphs, which are generated based on Erdos-Renyi model of random networks, in which every possible edge is made with same constant probability (Erdos & Renyi, 1959).

4.1.6 Secondary Hypothesis

The secondary hypothesis is that those cows with same number of parity tend to group together.

4.1.7 ERGM

In order to understand the influence of parity in the network, the analysis was extrapolated from assortativity using ERGM, since the last one can give a more detailed understanding of the factors. It uses the attributes of the members and characteristics of the network to explain or predict the probability of a tie.

Three different models were built with ERGM: the first one takes into consideration only the number of ties to evaluate the overall probability of a tie between nodes with same number of edges; The edges are added to the model as an intercept term and the model is estimated using the same maximum likelihood estimation methods used in a binary logistic regression (Harris, 2013).

The outcome of the model will show an estimate, which is the coefficient of each term. The resulting probability of a connection in this model will be the same as the density since the edges term indicates the number of links in the network, and the change statistics represents the change when an edge is added, hence being equal to 1 (Harris, 2013) The probability of the first model was calculated by using an anonymous function in R (Macedo, 2022), that takes the following format:

$$P(Y_{ij} = 1 | n \text{ actors}, Y_{ij}^c) = \text{logistic}(\theta_{edges} \delta_{edges}) \quad (1)$$

where θ_{edges} is the coefficient of the edges term and δ_{edges} represents the change statistic.

The second model evaluates the probability of a link between two individuals sharing the same parity. Therefore, the attribute of parity is added as an interaction term to the model using the expression "nodematch.parity". The resulting probability here will indicate how likely two individuals with the same parity number tend to have a connection, when the number of edges is increased by one unit.

The probability of the second model was calculated according to Equation 2.

$$P(Y_{ij} = 1 | n \text{ actors}, Y_{ij}^c) = \text{logistic}(\theta_{edges} \delta_{edges} + \theta_{nodematch.parity} \delta_{nodematch.parity}) \quad (2)$$

where $\delta_{nodematch.parity}$ equals to 1 if two individuals and have the same parity number.

On the third model, parity is added as a main effect predictor. Since parity is a categorical variable, the parity number is classified from one to four categories, the change statistic is either 0, 1 or 2. If none of the individuals in the dyad (as in the pair of nodes) has the characteristic under investigation, then there is no change statistic. If one of the individuals has it, then the change is equal to 1 and if both have it, then the change is equal to 2. This model differs from the previous one as it evaluates the characteristics of each individual member, while the previous model takes in account the attributes of both members of a dyad (Harris, 2013).

The probability of the third model was calculated according to Equation 3.

$$P(Y_{ij} = 1 | n \text{ actors}, Y_{ij}^c) = \text{logistic}(\theta_{edges} \delta_{edges} + \delta_{nodematch.factor}) \quad (3)$$

where $\delta_{nodematch.factor}$ equals to 2 if two individuals share the same parity number.

A summary of what is being evaluated in each model is provided in Table 2.

Table 2
Summary of description of the ERGM models

Model	Description
Model 1	Overall probability of aggregation based on number of edges
Model 2	Overall probability of aggregation between cows with same parity number
Model 3	Probability of aggregation based on specific parity number

4.2 Disease Simulation

4.2.1 Without Immunity

The cows that were identified to have the highest betweenness centrality were infected first. As the infection was started from cows of different centralities and in different communities of the network, this allowed for different scenarios to be assessed in terms of their impact on the disease spread. Given the background research, the working hypothesis here is that an infected cow with greater centrality and that is part of a larger community will result in a faster spread through the network than a cow again with highest betweenness centrality in their community, but which is part of a smaller community.

Two different disease scenarios were simulated. One simulation function did not allow for immunity after a recovery from the disease. The second simulation function allowed for one specific cow to be fully immune to the disease from the beginning until the end of the simulation. This allowed us to more clearly analyse the impact betweenness centrality and other parameters may have on the disease dynamics.

Our approach to the simulations is also known as a Susceptible – Infectious – Susceptible (SIS) model. SIS is commonly used to analyse close contact transmitted diseases or pathogens. In this type of model, each individual (node) can be in one of three stages. Either they are susceptible in receiving the disease, currently carry the disease and are infectious or they have just recovered from the disease and become susceptible again. No immunity is gained after this type of recovery, like it is also the case with Mastitis. Hence, the two types of infection models outlined in the paragraph above were able to be constructed using this framework (Sumner, McCabe, & Nunn, 2018).

The simulation was built in R based on the structure of the networks constructed previously. Using this dataset, an adjacency matrix of all the cows was constructed to save which cow was in contact with whom other. This was necessary to build the simulation function, which refers to the adjacency matrix in simulating the spread.

The simulation function constructed calculates which cows are infected as time passes using multiple *for* and *if* loops. The inputs needed for this simulation function include:

- How many cows should be infected at the start of the simulation
- The probability of becoming infected
- The probability of recovering from the disease and becoming susceptible again
- The timeframe for which to run the simulation

For the simulations in this thesis, the infection originated only from one cow at a time. This was done in an effort to control the variables and assess the impact the origin of the infection has on the spread of the disease throughout the network.

The probability of recovery considered different previous research done by multiple researchers. It marks the point when moving from susceptible to infectious. The infection probability was set to 0.03 or 3%. Given the complexity of this parameter, as previously raised, this threshold was also based upon a measurement of how frequently the disease would die out. At 3%, the disease died out in 50% of the repeated simulations and 50% of the times the disease spread until the end of the simulation period. While each of the different scenarios resulted in a different number, the average centred around this threshold (see Table A1).

The timeframe for the spread which was simulated in this thesis was 50 days, which was based upon the medium length timeframe recommendations of previous research outlined in the ‘*Background Research*’ section.

As discussed, the simulation function, constructed based on the methodology outlined above, was then run multiple times to identify the impact the infection origin has on spread of the disease. The most important cow in each of the communities was identified using the betweenness centrality measure explained earlier. This resulted in the cows listed in Table 3 and 4 to be identified.

Table 3
Overview of Social Communities from generated Network Dataset - Left side of Barn

Community	Number of Cows in Community	Cow ID
1	13	27
2	10	40
3	18	63
4	14	15
5	11	1
6	13	38

Table 4
Overview of Social Communities from generated Network Dataset - Right side of Barn

Community	Number of Cows in Community	Cow ID
1	27	136
2	21	158
3	9	92
4	10	116
5	14	84

Therefore, the simulation was run 6 times for the network data for the left side of the barn, with the infection originating from the different cow IDs identified in Table 1. It was also run 5 times on

the network data for the right side of the barn, with the infection starting at each of the different cow IDs identified by Table 2. Each of the simulations was repeated 500 times for a time period of 50 days. The outputs were saved in a matrix, from which the average number of daily infected cows in each scenario was calculated. The analysis between the different disease spreads in the varying simulations was based upon these daily averages.

4.2.2 With Immunity of one cow

To further our analysis, another parameter was added to the simulation function outlined above. The immunity of one cow in the network was added into the calculations. This was done in an effort to identify not only how the origin of the disease played a role, but also how the network dynamics and potential changes to this may influence the disease spread. It also allows us to analyse how an interruption in the disease dynamics would influence the spread of the disease.

The simulations that included the immune parameter were still run on the same cows as specified in Tables 3 and 4. However, this time the simulations for each different origin were repeated several times with different cows identified as immune. Table 5 below identifies the scenarios for which cows were made immune and why. Keeping in mind that in each simulation only one cow was made immune to enable variable control and more meaningful analysis.

Table 5
Cow IDs for which Immunity Simulations were Performed

Cow ID	Reason for Immunity
Left side of Barn	
31	Highest degree (most connections) in entire network
9	Articulation point ¹
63	Central to largest community
15	Central to second largest community ²
40	Central to smallest community
1	Central to second smallest community ³
Right side of Barn	
139	Highest degree centrality & eigen vector (most connections)
112	Highest betweenness centrality
136	Central to largest community
158	Central to second largest community ²
92	Central to smallest community
116	Central to second smallest community ³

A clearer overview of all the scenarios for which disease simulations were built and run can be seen in Figure 4 , which illustrates the Cows for which no immunity simulations were run first, followed by the specific immunity simulations run for each cow.

¹ The Articulation point is essential in ensuring the connectivity of the Network. The removal of this point will result in the collapse of the Network (Tian et al., 2017).

² Simulation of immunity on the cow central to the second largest community is only performed in the scenario where the infection originates from the cow central to the largest community.

³ Simulation of immunity on the cow central to the second smallest community is only performed in the scenario where the infection originates from the cow central to the smallest community.

5 Results

5.1 Social Networks

5.1.1 Descriptive characteristics of the networks

The resulting networks had different structures. The network from the left side had 23.5% less ties but a higher modularity, which indicates that in the second network cows grouped in larger groups, resulting in a lower number of communities and a higher number of connections per individual (Table 6) (Figs. 5 and 6).

Both networks had similar density, average path distance and diameter and none of these measures were high, which does corroborate with the literature research pointing out to the fact that cows have preferential relationships (Table 6).

The homophily of both networks was measured in terms of the number of degrees for each node and also the parity. There was a positive correlation in both characteristics although not high in both cases, meaning that there was some indication that cows that are highly gregarious tend to associate with others that are also gregarious, likewise that they tend to aggregate by parity. Differently from what would have been expected from a null model which indicates heterogeneity in terms of degree assortativity (Table 6).

These basic structures of both networks were compared to a null model, as described in methods, and there wasn't anything unexpected in most of the metrics, but the largest cliques and homophily as described above. On both sides, the number of largest cliques were much lower than those in the average random graphs, again demonstrating the selectiveness through which cows tend to aggregate (Table 6).

The hypothesis that the members of each clique would be in the same communities, was true in half of the cases on the left side and all cases on the right side. There were just two cows from the left side that were identified as the most important in their communities that were also present in the cliques, and only one of them was present in the diameter. Regarding the parity number both parity 2 and 3 were the most common in the cliques, with the same frequency (Tables 7 and 8).

In both networks there were indications that cows with higher centrality in the networks are also present in larger cliques. Opposite to the results found on the left side, where cows from the largest cliques were not part of the largest communities, on the right side all of them were part or the second largest community (Tables 7 and 8).

One articulation point was identified on the left network (Cow 9), but no articulation points were found on the right network.

The most frequent parity number on both sides was parity 2 (Figures 1 and 2). The network diagram on the left side reveals that cows with parity number 3 seems to be the ones with higher number of connections (Figure 7), while on the right side cows with parity 1 and 2 are the most connected (Figure 8).

The individuals present in the diameter on the right side were not found in the clique and only one scored the highest betweenness centrality in her community. Nor the cow with highest degree and eigenvector centralities was present in the diameter (Tables 7 and 8).

Table 6

Summary of characteristics of both social networks

Characteristics of the network	Left	Average 1000 Random Graphs	Right	Average 1000 Random Graphs
Number of cows	79	79	81	81
Number of edges	268	268	350	350
Mean degree	6.8	6.7	8.6	8.5
Mean distance	2.5 edges	2.4	2.3 edges	2.2
Centralization	0.08	0.08	0.07	0.09
Edge density	0.08	0.08	0.10	0.10
Number of Communities	6	6.6	5	6.5
Modularity	0.37	-	0.30	-
Homophily (assortativity degree)	0.08	-0.03	0.10	-0.02
Homophily – parity	0.13	-	0.20	-
Diameter	5	4.5	4	4
Largest cliques	4 with 4 members each	28	1 with 5 members	12
Transitivity	0.11	0.08	0.16	0.10

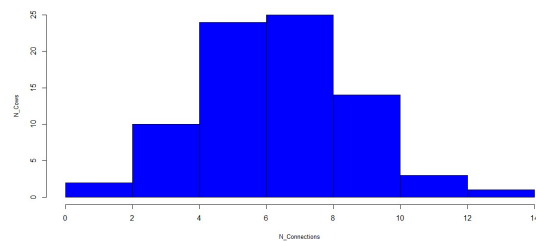


Figure 5
Histogram of the distribution of nodes degree (left side)

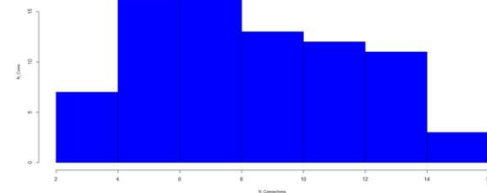


Figure 6
Histogram of the distribution of nodes degree (right side)

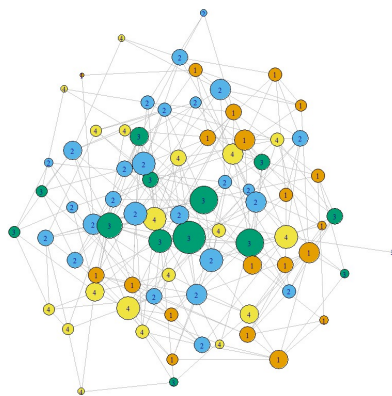


Figure 7
Social Network – Left Side

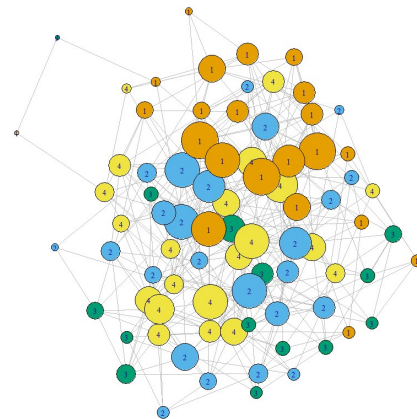


Figure 8
Social Network – Right Side

Note: the nodes are colored by parity number and the size of nodes is given as a factor of their degree of connections.

Table 7

Summary of characteristics of cliques – Left side. The column of Most Important identify those cows with highest betweenness centrality in their communities and the column Part of the diameter identify those that are included in the diameter (Y = Yes).

Cow ID	Parity	Community	Most Important	Part of the diameter?
Clique 1				
1	3	5	Y	Y
6	4	5	-	
35	2	5	-	
50	2	5	-	
Clique 2				
31	3	2	-	
36	3	2	-	
40	3	2	Y	
44	2	2	-	
Clique 3				
1	3	5	Y	
31	3	2	-	
50	2	5	-	
60	1	5	-	
Clique 4				
10	3	3	-	
12	2	1	-	
14	2	1	-	
17	2	1	-	

Note: Cow number 1 was identified as having the highest betweenness centrality in her community and it is also in the diameter. She is present in 2 of the cliques, as well as Cow 31 and Cow 50. Cow 31 is also the one most central to the network.

Table 8

Summary of characteristics of cliques – Right side. The column of Most Important identify those cows with highest betweenness centrality in their communities and the column Part of the diameter identify those that are included in the diameter (Y = Yes).

Cow ID	Parity	Community	Most Important	Part of the diameter?
Clique 1				
139	1	2	-	-
140	1	2	-	-
141	1	2	-	-
152	1	2	-	-
155	1	2	-	-

Note: Cow 139 scored higher in two centrality metrics, the degree and eigenvector.

5.1.2 ERGM analytical results

Results of all models for both sides are presented in Appendix B. The first model, with only the edges, indicated on both sides that an additional link influences the probability of formation of a tie, but it has not indicated a dense network, since the coefficients were negative (on the left: -2.35, on the right: -2.11. Negative coefficients will indicate that density is below 50% since an edge term of 0 would lead to .5 density.

The second model showed significant positive coefficients on both sides (on the left: 0.27 and p-value 0.05, on the right: 0.54 and p-value <1e-04), indicating that two cows that share the same parity number are likely to be connected, although this likelihood is not very high. On the left side they are 10% more likely to have a tie, while on the right side this likelihood goes up to 15%.

This correlation was further investigated on the third model in an attempt to understand whether there were any differences depending on the parity. On both sides, the only parity that showed relevant results was parity number 3 (on the left: 0.28 and p-value: 0.0389, on the right: -0.41 and p-value: 0.00195). On the left side, cows with parity number 3 were 10% more likely to aggregate while on the right side they were 5% less likely to aggregate when increasing the number of edges in one unit.

5.2 Disease Simulation

5.2.1 Results Disease Simulation without Immunity

Using the methodology described above, this study used those average number of infected cows per day values to assess the hypothesis that a disease originating from a cow in a large community has a higher impact on the spread than if the cow was central in a smaller community. The results from the simulations are illustrated in Figure 9 below.

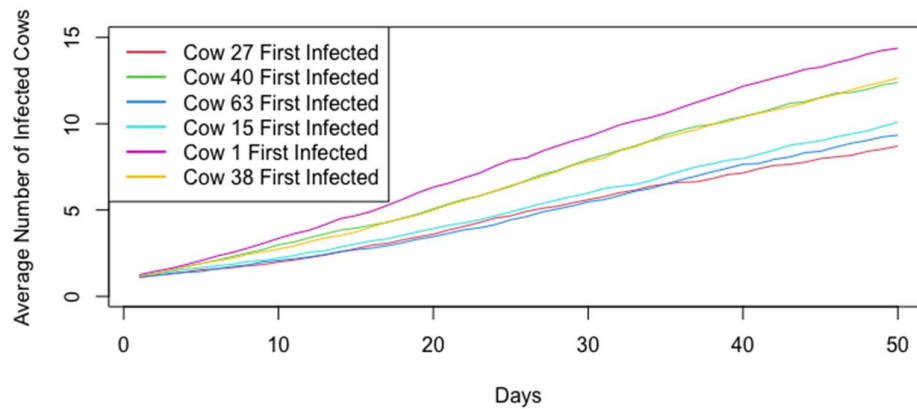


Figure 9
Average Number of Infected Cows per Day without Immunity – Left side of Barn

It is noticeable that cow 1 results in the largest spread, with cow 40 and cow 38 following in a very similar dynamic. Cow 27 results in the smallest spread over the time period. It is surprising that cow 1 results in the biggest spread, as it is only the most central cow in the second smallest community. Simulations originating from cow 38 from a medium sized community and cow 40 from the smallest community followed very similar disease dynamics. Again, it was against our hypothesis that these resulted in the second largest spread of the disease. In fact, the simulation from the cow central to the largest community, cow 63, resulted only in the second smallest spread of the simulation on average (Figure 9). Therefore, our hypothesis that a disease breakout in a large community would result in a larger spread of the disease is not supported by the results from the simulations.

The same methodology was applied to the network data of the right side of the barn to investigate if a similar inconclusiveness could be noted in that network, as seen in Figure 10 below.

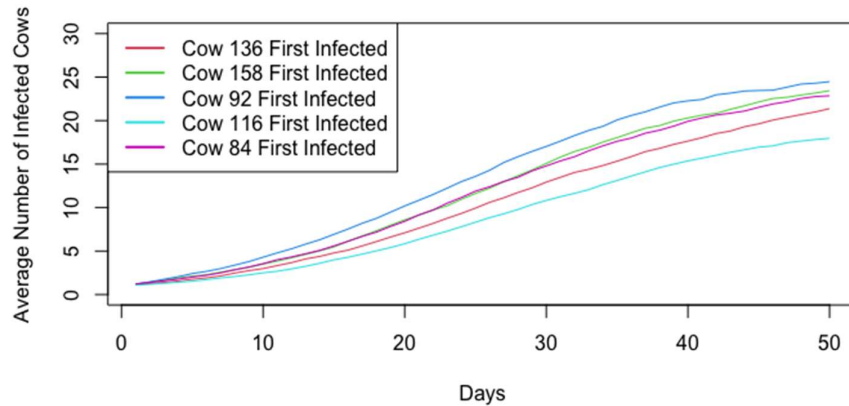


Figure 10
Average Number of Infected Cows per Day without Immunity – Right side of Barn

It can be noted that while the number of cows on each side of the barn is quite similar (79 cows on the left and 81 on the right side of the barn) starting an infection on the right side of the barn results in a slightly larger total spread of the disease on average. A different disease dynamic can also be noted as the slope of the average daily number of infection cows per day is slightly more shaped like an exponential graph while the slope on the left followed that of a regression graph (Figure 10).

On the right side, simulations where the disease originated from cow 116 resulted in the smallest spread of the disease, Simulations originating from cow 92, resulted in the largest average spread of the disease. When looking further, it can be observe that one of the direct neighbours of this cow is present in the clique (Table B2). Simulations from the two biggest communities, cow 136 and 158, also resulted in contradicting results. Simulations from cow 136, central to the by far the largest community, resulted in the second smallest spread of the disease. However, simulations where the disease originated from cow 158, which is central to the second largest community, resulted in the second largest spread of the disease. Interestingly, this individual also has a direct neighbour present in the clique (Table B2).

Therefore, comparing the impact that centrality has on the disease dynamics and rate of spread did not allow for a clear conclusion to be made upon the influence that the origin of the disease plays on the disease dynamics. Notable discrepancies between the two different individual networks on either side of the barn have also added to this inconclusiveness. Therefore, this thesis further investigated how a disruption to this network cohesiveness through an immune cow would influence the simulation results.

5.2.2 Results of Disease Simulation with Immunity

As outlined in the methodology section, different strategies were used to identify which cows to build immunity simulations for. The different results of each variation of immunity for the different origins of the disease are illustrated in the graphs below.

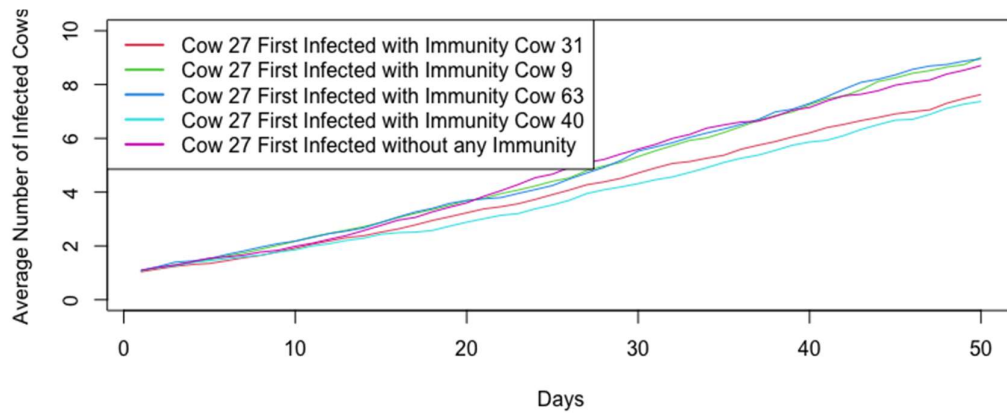


Figure 11
Average Number of Infected Cows per Day with vs. without Immunity Cow 27 – Left side of Barn

Figure 11 displays the average number of the of infected cows per day for the different scenarios with immunity of the specified cow and without immunity when the infection originates from cow 27. Cow 31 leads to the second smallest spread of the disease. In comparison, the immunity of cow 40, leads to the smallest spread of the disease (Figure 11). Unexpectedly the immunity of cow 9, which has been identified as the articulation point and is represented by the green line resulted in a slightly higher amount of cows becoming infected than the base level without any immunity parameter. Simulations where cow 63 was immune resulted in similar infection rates as those where immunity of cow 9 was present (Figure 11).

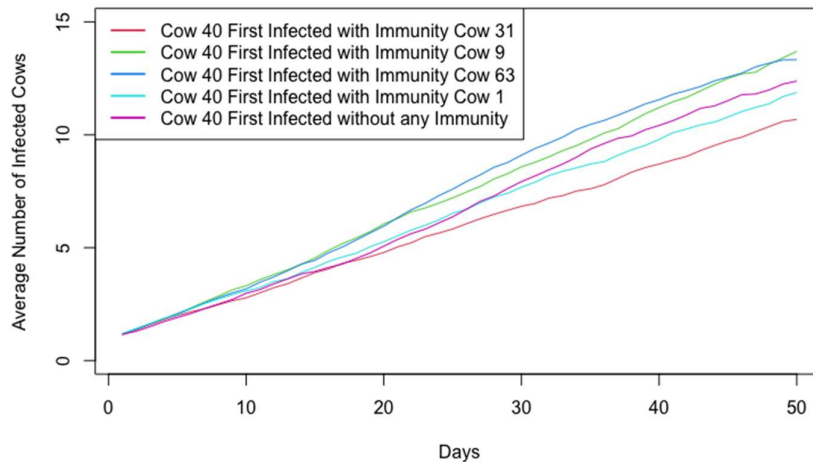


Figure 12
Average Number of Infected Cows per Day with vs. without Immunity Cow 40 – Left side of Barn

In Figure 12, scenarios for Cow 40, which is central to the smallest community, are displayed. Again in these simulation results it can be noted that the immunity of cow 31 resulted in the smallest spread of the disease. Over, the immunity of cow 9 and cow 63 resulted in a higher spread of the

disease than even the baseline simulation without any immunity. Similar to the simulations from cow 27, the immunity of cow 40 from the smallest community had the second largest impact on decreasing the spread.

A similar trend can be noted in all of the other disease origin scenarios (See Table A2-A5 for the other graphs of the infected cows). The immunity of cow 31 almost always results in the smallest spread of the disease. On the other hand the immunity of cow 9 and 63 do not seem to be impactful on the dynamics of the disease spread, as they result in similar infection rates as simulation results without any immunity. Due to the background knowledge that cow 9 is the articulation point in the network, the removal of which would have resulted in a collapse of the network, a large decline was expected in the spread of the disease when cow 9 was immune. However, this was not evident in the results gathered. To make a final conclusion the thesis compared the results to those of the right side of the barn.

5.2.3 Results comparison to right side of the barn

For the right side of the barn the immunity scenarios were slightly different. While still the cow with the highest degree centrality (the one with the most connections) and those out of the smallest and largest community were used to analyse different scenarios, no articulation point on the right side of the barn and its network was able to be identified. Instead it was run on the cow with the highest betweenness centrality (see Figure 3). This allowed for similar variable selection to compare to those of the left side, as identified in Table 5 earlier.

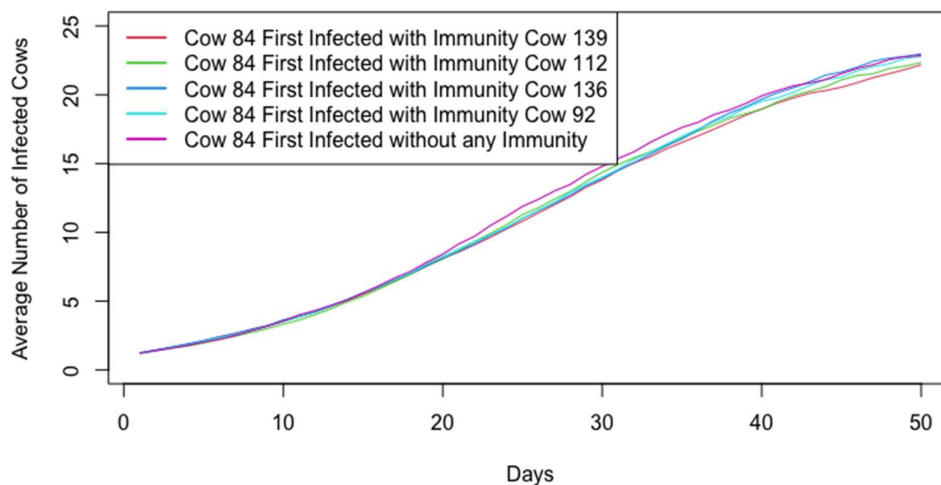


Figure 13
Average Number of Infected Cows per Day with vs. without Immunity Cow 84 – Right side of Barn

The different disease dynamics that can be noted from the simulations of disease origination from cow 84, seen in Figure 13, are a good example of the general trend noted by the network found on the right side of the barn. The distribution of infection rates is closer together when compared to the left side. It is also noted that a simulation without any immunity in all but one case result in the highest average amount of infected cows per day at the end of the simulation period. This is more along the lines of what would be expected based on the theoretical background, compared to the left side simulations. If no immunity is given, the disease should spread more no matter the immunity parameters (Sumner, McCabe, & Nunn, 2018).

The red line represents the simulations where cow 139, the one with the highest degree and eigen vector centrality, is immune, as well as the green line with cow 112, the cow with the highest betweenness centrality, is immune. Since both of these lines are at the bottom of the figure, it can be said that an interference with the network at those points slows the spread of the disease down the most. However, the scale is important to note here. Unlike in the simulations performed on the network of the left side of the barn, the simulation results on the right side are significantly closer together. These trends are noticeable throughout all of the five different community simulations (*refer to A6-A9*).

6 Discussion & Conclusion

6.1 Conclusions

The objective of this paper was to build a social network based on spatial data from dairy cattle to use for simulations of contagious disease transmission in the attempt to identify influential network factors on the rate of infections. Implications of this include better cattle management for farmers, animal welfare improvements and potential reduction in economic impacts that results from disease infections.

The simulation of a contagious disease based on social behaviour is complex. Different distributions of possible outcomes, a range of possible probabilities and parameters are just a few of the elements that add to this complexity. The impact that centrality has on the spread of a contagious disease was investigated, by changing the origin of the disease and keeping other factors constant.

It was hypothesized that the size of the community would influence the dynamics of the spread of a disease, the results did not corroborate with our initial hypothesis. In all simulations the size of the community did not have such a high impact, neither the centrality inside the communities played an important role, what showed to be of more influence was the centrality of a node to the whole network, those individuals who had higher score in most of the centrality measures (degree centrality, betweenness centrality and eigen vector centrality) when considered immune in the simulations, resulted in slowing down the rate of infections. Also, some relationship to the cliques in both networks from key individuals could be observed.

Although these are not conclusive results as this cannot generate an overall recommendation to farmers, it does highlight the importance in understanding the substructures of the networks in the occurrence of an outbreak. It was observed that in one of the networks the overall centrality and presence in the diameter and the cliques played a big role, the individual with these characteristics was the one, when identified as being the focus for initial infection, resulted in a higher disease spread. Although on the right side, the individuals who resulted in higher spread did not have any direct connection to these structures, when looking further, a direct connection to two other individuals inside the clique was found, showing some evidence that these substructures can indeed have an impact. It does require further investigation for conclusiveness though, as this analysis is limited to only 2 networks observed during a short period of time.

Overall, there was an indication of preferential relationship on both networks due to parity, being parity number 3 the one with most significant results.

6.2 Future Perspectives

As discussed previously, the networks constructed in this study and used in the disease simulation were built upon a period of only 24h, thus assuming a fixed network. Although, research in this field has shown social networks in dairy cattle cows are relatively stable, it would be interesting to extend this research for larger periods and compare the results. Similarly, other thresholds for the distance can be tested, especially when looking into specific areas (Rocha et al., 2020).

There are several measurements that could be taken to illustrate and understand a social network. For example, this thesis touched upon the concepts of cliques, but due to limitations of time it was unable to explore this further. A possible improvement for future research would be to try to

identify the individuals that are directly infected after the first individual is infected and compare whether these individuals are inside the same cliques. This could indicate a clear route for the spread of a disease and help farmers minimize the impact in an outbreak. The deep dive into the specific disease distribution could perhaps allow for a more conclusive recommendation to be drawn, as deeper trends may be found within the data.

Another area of improvement regarding the social network is the community detection algorithm. There are several possibilities one could work with: Optimal, Infomap, Walktrap, Edge Betweenness, Label Propagation, Leading Eigenvector. Different community detection algorithm can influence the outcomes of the structures of the network, changing the scenarios for simulating disease transmission. This study focused on the Louvain method, but future research could explore the use of other algorithms. Also, testing different thresholds for modularity would be a recommendation, as Summer et al. (2018) have found that for the Louvain method, for example, modularity only had an effect on simulating disease spread over $Q_{rel} > 0.60$ in their study.

Furthermore, the time two individuals spend together could influence in the dynamics of disease diffusion. Some diseases have different time of incubation period and different times when they are more contagious. In this thesis, this was not taken into consideration as our intention was to model a hypothetical disease with similar characteristics to a Mastitis, and in that case there is still a gap in research in terms of identifying the period when the disease has higher risk of spreading.

6.3 Limitations

The biggest limitation for the disease spread simulation was the lack of available data and research on the parameters surrounding the Mastitis disease or that being said any contagious disease which was tried to be simulated. While extensive research was done this topic, this large research gap became more and more apparent. As previous research data was found to be unavailable on the complex topic of disease dynamics, more general parameters needed to be used for the construction of the simulation function. The availability of this data surrounding disease dynamics would have substantially improved the simulations of this study and thus allowed for stronger conclusions to be drawn.

For the simulation of disease transmission, this thesis based its methodology on the static network data. However, due to these the temporal connections during the simulation period are disregarded. As a result, the impact of centrality on disease transmission is overestimated. This clearly marks an area of improvement, which would impact the strength of our conclusions drastically (Rossi et al., 2017). In addition to this, any movements during the simulation period, such as the movement of cattle from one side of the barn, addition of cattle to the network or the death of a cow during the timeframe of the simulation are ignored. While this does not influence the reliability of the constructed simulation, it does affect the applicability to a real-life scenario, where movement of cattle and thus a dynamic network present themselves as a more realistic interpretation.

While this study investigated substructures of a social network on the disease transmission rate, other factors influencing the spread of a disease are ignored or kept constant. However, as Rossi et al. (2017) investigate in their study, the impact workers can have upon the spread of a communicable disease shall also not be disregarded. While direct contact defines the contact between two cows directly, in this study, indirect contact include much more complex ways of

disease transmission, such as through the workers that have contact with the cattle, livestock movement between different farms or groups of cattle within a farm and many more alternative routes of disease transmission. Therefore, the impact that these indirect contacts may have had upon the disease simulation are ignored in the analysis entailed in this paper. However, future work may be taken up on this to improve upon the conclusions made in an impactful way (Rossi et al., 2017).

It has become apparent through our discussion of the limitations that the methodology used for the disease simulation is element to multiple drawbacks. These include, but are not limited, to the discussed parameters and use of stable network data. In addition, while effective, the simulation function itself may be overly simplistic in its construction. This is not only resulting from the lack of parameters, but also a result of the available time and resources.

Initially it was aimed to build the disease simulation using the EpiModel package in R studio. This tool allows for a mathematical modelling of disease simulation by analyzing infectious disease dynamics. The type of disease simulation of interest in our paper, SIR, is supported by this package. The base of the simulation is statistical framework or a network build using ERGM in the Statnet package in R studio. It was also discussed to make use of exponential random graph models (ERGMs) to analyse disease dynamics in conjunction with social networks constructed using ERGMs (Garry et al., 2006). While EpiModel simulations allow for more complexity, it is this complexity that takes a large amount of time and resources. In addition, our network was build using iGraph. Hence, future outlooks include simulations using different methodologies. Nonetheless, it is aware that both of these models perhaps would have been more successful in capturing the complexity of contagious diseases (Statnet Development Team, 2020).

The location tracking could also be extended to capture between-herd or market movement data, a more holistic understanding of the origin of diseases could be studied. While collecting this data is likely to be rather costly and needs evidential justification, it would allow for livestock traceability at a higher level. This would be particularly important in tracing epidemiological outbreaks (Chaters et al., 2019).

Unlike direct contacts, which are fairly easily to track and allow for a clear network to be built, the role that indirect transmission has on the spread of diseases throughout a barn is still a highly unknown area. This research gap unfolded itself on the one side due to the complexity and subtle nature of these indirect contacts, on the other side data privacy also influenced it to a large extent. This is because the tracking of livestock movement entails basically no data privacy issues, while the collection of data on the farm workers and personal throughs up a bigger data privacy consideration. Hence, by working out solutions for this privacy concern and methodology to capture the privacy concerns (Rossi et al., 2017).

Building on this, due to the complexity of diseases and thus their simulation, as outlined in the '*Data*' section, there are many different parameters adjustments and investigations that can further complement this work. Examples of these may include investigating if the stress and physical state of the cow influences their susceptibility and transmission probability. Biologists have hypothesized that a lactating cow is under higher stress and thus more susceptible to an infection of a disease (Rocha et al., 2019). In an effort to control variables, our analysis also ignored external factors on disease transmission, such as temperature of the barn and the outside. These can also have a

substantial impact on the incubation period of the infection, life of the disease outside the living organism, etc (Rossi et. al., 2017). These examples show that due to the vast complexity of diseases combined with the research gap in this field, many further studies in this area can find their potential.

It is evident that due to the research gap in both contagious disease parameter data itself, but especially in conjunction with disease simulations and network data many future opportunities present themselves. While some lack of data hindered the conclusiveness of this study, these present themselves as opportunities for further investigation in future works.

References

- Adhikari, S., & Dabbs, B. (2018). Social network analysis in R: a software review. *Journal of Educational and Behavioral Statistics*, 43(2), 225-253.
- Amaral-Phillips, D., & Franklin, S. (1999). Feeding and managing the far-off dry cow. University of Kentucky, College of Agriculture, Cooperative Extension Service.
- Ankinakatte, S., Norberg, E., Løvendahl, P., Edwards, D., & Højsgaard, S. (2013). Predicting mastitis in dairy cows using neural networks and generalized additive models: A comparison. *Computers and Electronics in Agriculture*, 99, 1-6. doi:<https://doi.org/10.1016/j.compag.2013.08.024>
- Badham, J., & Stocker, R. (2010). The impact of network clustering and assortativity on epidemic behaviour. *Theoretical Population Biology*, 77(1), 71-75. doi:<https://doi.org/10.1016/j.tpb.2009.11.003>
- Blondel, V., Guillaume, J., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10), P10008.
- Boyland, N., Mlynski, D., James, R., Brent, L., & Croft, D. (2016). The social network structure of a dynamic group of dairy cows: From individual to group level patterns. *Applied Animal Behaviour Science*, 174, 1-10. doi:<https://doi.org/10.1016/j.applanim.2015.11.016>
- Burvenich, B., Van Merris, V., Mehrzad, J., Diez-Fraile, A., & Duchateau, L. (2003). Severity of E. coli mastitis is mainly determined by cow factors. *Veterinary research (Paris)*, 34(5), 521-564. doi:<https://doi.org/10.1051/vetres:2003023>
- Chaters, G., Johnson, P., Cleaveland, S., Crispell, J., de Glanville, W., Doherty, T., . . . Kao, R. (2019). Analysing livestock network data for infectious disease control: an argument for routine data collection in emerging economies. *Philosophical transactions. Biological sciences*, 374(1776), 20180264-20180264. doi:<https://doi.org/10.1098/rstb.2018.0264>
- Chen, S., Ilany, A., White, B., & Sanderson, M. (2015). Spatial-temporal dynamics of high-resolution animal networks: What can we learn from domestic animals? *PLoS One*, 10(6), e0129253-e0129253. doi:<https://doi.org/10.1371/journal.pone.0129253>
- Chen, S., White, B., Sanderson, M., Amrine, D., Ilany, A., & Lanzas, C. (2014). Highly dynamic animal contact network and implications on disease transmission. *Scientific Reports*, 4(1), 4472-4472. doi:<https://doi.org/10.1038/srep04472>
- Craft, M. (2015, 05). Infectious disease transmission and contact networks in wildlife and livestock. *Philosophical transactions. Biological sciences*, 370(1669), 20140107. doi:<https://doi.org/10.1098/rstb.2014.0107>
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, complex systems*, 1695(5), 1-9.
- Czarnetzki, M. (2022, Jan). MasterThesis. Simulation of Disease Transmission of Dairy Cattle. Retrieved from <https://github.com/miraczarnetzki/MasterThesis>

- de Freslon, I., Martínez-López, B., Belkhiria, J., Strappini, A., & Monti, G. (2019). Use of social network analysis to improve the understanding of social behaviour in dairy cattle and its impact on disease transmission. *Applied Animal Behaviour Science*, *213*, 47-54.
doi:<https://doi.org/10.1016/j.applanim.2019.01.006>
- Down, P., Bradley, A., Breen, J., Browne, W., Kypraios, T., & Green, M. (2016). A bayesian micro-simulation to evaluate the cost-effectiveness of interventions for mastitis control during the dry period in UK dairy herds. *Preventive Veterinary Medicine*, *133*, 64-72.
doi:<https://doi.org/10.1016/j.prevetmed.2016.09.012>
- Erdos, P., & Renyi, A. (1959). On random graphs. *Publicationes Mathematicae*, *6*, 290–297.
- Garry, R., Pip, P., Yuval, K., & Dean, L. (2007). An introduction to exponential random graph (p^*) models for social networks. *Social Networks*, *29*(2), 173-191.
doi:<https://doi.org/10.1016/j.socnet.2006.08.002>
- Gibbons, J., Lawrence, A., & Haskell, M. (2010). Measuring sociability in dairy cows. *Applied Animal Behaviour Science*, *122*(2), 84-91. doi:<https://doi.org/10.1016/j.applanim.2009.11.011>
- Global AG Media. (2015, 05 18). *Managing Cow Lactation Cycles*. Retrieved from The Cattle Site:
<https://www.thecattlesite.com/articles/4248/managing-cow-lactation-cycles/>
- Gygax, L., Neisen, G., & Wechsler, B. (2010). Socio-spatial relationships in dairy cows. *Ethology*, *116*(1), 10-23. doi:<https://doi.org/10.1111/j.1439-0310.2009.01708.x>
- Handcock, M., Hunter, D., Butts, C., Goodreau, S., Krivitsky, P., & Morris, M. (2021). ergm: Fit, Simulate and Diagnose Exponential-Family Models for Networks. *The Statnet Project* (<https://statnet.org>), R package version 4.1.2. Retrieved from <https://CRAN.R-project.org/package=ergm>.
- Handcock, M. S. (2011). statnet: Software Tools for the Statistical Modeling of Network Data, version 2.0. Retrieved from statnet.org.
- Harris, J. (2013). *An introduction to exponential random graph modeling* (Vol. 173). Sage Publications.
- Jordbruksverket. (2017). *Swedish Food Production*. Jönköping: Jordbruksverket. Retrieved from <https://www2.jordbruksverket.se/download/18.5c2fc416166df50915ca0f0c/1541574402204/ovr462.pdf>
- Koene, P., & Ipema, B. (2014). Social networks and welfare in future animal management. *Animals (Basel)*, *4*(1), 93.
- Krivitsky, P. (2021). *statnet.common: Common R Scripts and Utilities Used by the Statnet Project Software*. The Statnet Project Software (<https://statnet.org>). R package version 4.5.0. Retrieved from <https://CRAN.R-project.org/package=statnet.common>
- Lassen, B. J. (2013, 1). *Dairy production in Sweden - impressions -*. Retrieved from <http://www.agribenchmark.org/>: http://www.agribenchmark.org/fileadmin/Dateiablage/B-Dairy/Country_Report/13_1_country_report_SE.pdf
- Lundberg, Å. (2015). *Mastitis in Dairy Cows: Genotypes, Spread, and Infection Outcome of Three Important Udder Pathogens*. Uppsala: Faculty of Veterinary Medicine and Animal Science Department of Clinical Sciences Uppsala and National Veterinary Institute Department of Animal Health and

- Antimicrobial Strategies Uppsala. Retrieved from https://pub.epsilon.slu.se/11975/1/lundberg_a_150306.pdf
- Lundberg, Å., Nyman, A., Aspán, A., Börjesson, S., Unnerstad, H., & Waller, K. (2016). Udder infections with staphylococcus aureus, streptococcus dysgalactiae, and streptococcus uberis at calving in dairy herds with suboptimal udder health. *Journal of Dairy Science*, 99(3), 2102.
- Macedo, J. (2022, Jan). Python and R code used in this thesis for network construction. Retrieved from <https://github.com/julianaabdias/Master-Thesis>
- Missaoui, R., & Sarr, I. (2014; 2015). *Social network analysis - community detection and evolution*. Cham: Springer.
- Otrante-Chardonnet, Y. (2021, Jun). CowDistance-Project. Retrieved from <https://github.com/CSI-DT/CowDistance-Project>
- Petersson-Wolfe, C., Mullarky, I., & Jones, G. (2019, 08 16). *Staphylococcus aureus Mastitis: Cause, Detection, and Control*. Retrieved from Dairexnet: <https://dairy-cattle.extension.org/staphylococcus-aureus-mastitis-cause-detection-and-control/>
- Read, J., & Keeling, M. (2003). Disease evolution on networks: The role of contact structure. *Proceedings of the Royal Society. B, Biological Sciences*, 270(1516), 699-708. doi:<https://doi.org/10.1098/rspb.2002.2305>
- Ren, K., Nielsen, P., Alam, M., & Rönnegård, L. (2021). Where do we find missing data in a commercial real-time location system? Evidence from 2 dairy farms. *ANIMAL NUTRITION AND FARM SYSTEMS RESEARCH TECHNICAL NOTE*, 2(6), 345-350. doi:<https://doi.org/10.3168/jdsc.2020-0064>
- Rocha, L., Terenius, O., Veissier, I., Meunier, B., & Nielsen, P. (2020). Persistence of sociality in group dynamics of dairy cattle. *Applied animal behaviour science*, 223, 104921. doi:<https://doi.org/10.1016/j.applanim.2019.104921>
- Rossi, G., De Leo, G., Pongolini, S., Natalini, S., Zarengi, L., Ricchi, M., & Bolzoni, L. (2017). The Potential Role of Direct and Indirect Contacts on Infection Spread in Dairy Farm Networks. *PLoS computational biology*, 13(1), e1005301-e1005301. doi:<https://doi.org/10.1371/journal.pcbi.1005301>
- Salathé, M., & Jones, J. (2010). Dynamics and control of diseases in networks with community structure. *PLoS Computational Biology*, 6(4), e1000736-e1000736. doi:<https://doi.org/10.1371/journal.pcbi.1000736>
- Scott, J. (2011). Social network analysis: Developments, advances, and prospects. *Social Network Analysis and Mining*, 1(1), 21-26. doi:<https://doi.org/10.1007/s13278-010-0012-6>
- Sehested, J., Gaillard, C., Lehmann, J., Maciel, G., Vestergaard, M., Weisbjerg, M., . . . Kristensen, T. (2019). Review: Extended lactation in dairy cattle. *Animal (Cambridge, England)*, 13(S1), s65-s74. doi:<https://doi.org/10.1017/S1751731119000806>
- Silva, N., Guimarães, F., Manzi, M., Júnior, A., Gómez-Sanz, E., Gómez, P., . . . Torres, C. (2014). Methicillin-resistant staphylococcus aureus of lineage ST398 as cause of mastitis in cows. *Letters in Applied Microbiology*(59), 665-669. doi:<https://doi.org/10.1111/lam.12329>
- Sloth, K., & Frederiksen, D. (2019). *U.S. Patent No. 10,234,535*.

- Sueur, C., Jacobs, A., Amblard, F., Petit, O., & King, A. (2011). How can social network analysis improve the study of primate behavior? *American Journal of Primatology*, 73(8), 703-719.
doi:<https://doi.org/10.1002/ajp.20915>
- Sumner, K., McCabe, C., & Nunn, C. (2018). Network size, structure, and pathogen transmission: A simulation study comparing different community detection algorithms. *Behaviour*, 155(7-9), 639-670. doi:<https://doi.org/10.1163/1568539X-00003508>
- Swedish University of Agricultural Sciences. (2021, Nov 08). *Precision livestock breeding – improving both health and production in dairy cattle*. Retrieved from SLU - Faculty of Veterinary Medicine and Animal Science: <https://www.slu.se/en/faculties/vh/research/forskingsprojekt/not/precision-livestock-breeding/>
- The SciPy community. (2008-2021). *scipy.interpolate.Akima1DInterpolator*. Retrieved from API reference: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.interpolate.Akima1DInterpolator.html>
- Virtanen, P., Gommers, R., Oliphant, T., Haberland, M., Reddy, T., Cournapeau, D., & ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261–272. doi:<https://doi.org/10.1038/s41592-019-0686-2>
- Wey, T., Blumstein, D., Shen, W., & Jordán, F. (2008). Social network analysis of animal behaviour: A promising tool for the study of sociality. *Animal Behaviour*, 75(2), 333-344.
doi:<https://doi.org/10.1016/j.anbehav.2007.06.020>

Appendix A

Figures related to Disease Simulation

Table A1

Number of Cases in which the disease dies out

Number of the Cow from which disease originated	Number of Cases (out of 500) in which the Disease dies out before the 50 days have been Simulated
Left Side	
27	303
40	267
63	284
15	269
1	231
38	243
Right Side	
136	206
158	202
92	214
116	211
84	276

Figure A1

Cow 63: Average Number of Infected Cows per Day with Different Immunity vs. No Immunity Scenarios - Left side of Barn

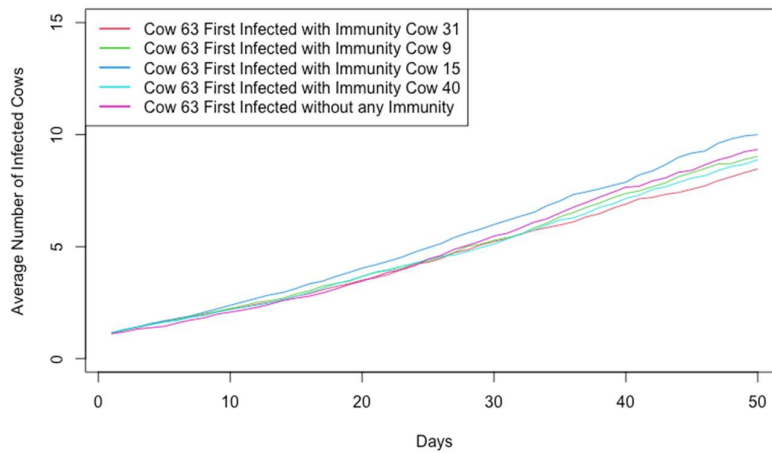


Figure A2

Cow 15: Average Number of Infected Cows per Day with Different Immunity vs. No Immunity Scenarios - Left side of Barn

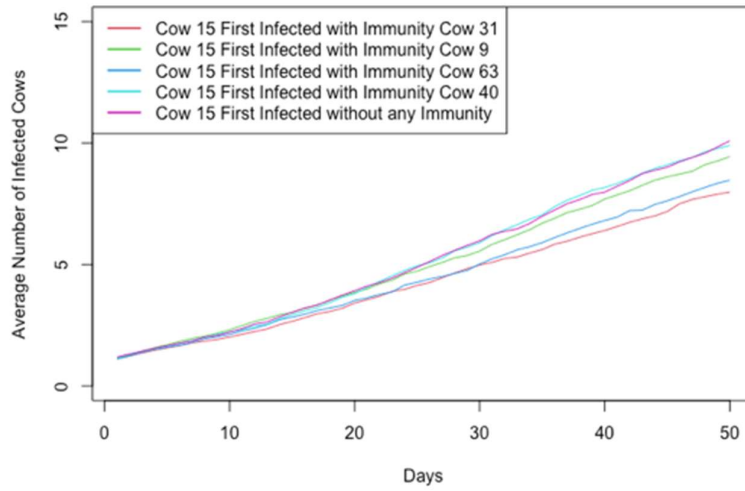


Figure A3

Cow 1: Average Number of Infected Cows per Day with Different Immunity vs. No Immunity Scenarios - Left side of Barn

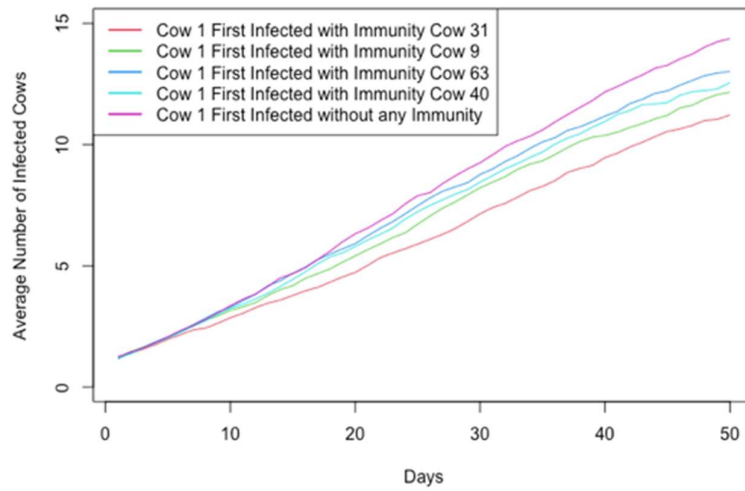


Figure A4

Cow 38: Average Number of Infected Cows per Day with Different Immunity vs. No Immunity Scenarios - Left side of Barn

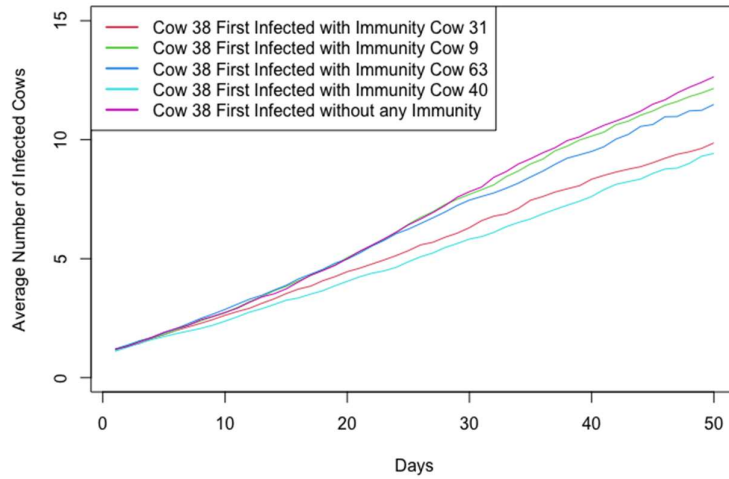


Figure A5

Cow 92: Average Number of Infected Cows per Day with Different Immunity vs. No Immunity Scenarios - Right side of Barn

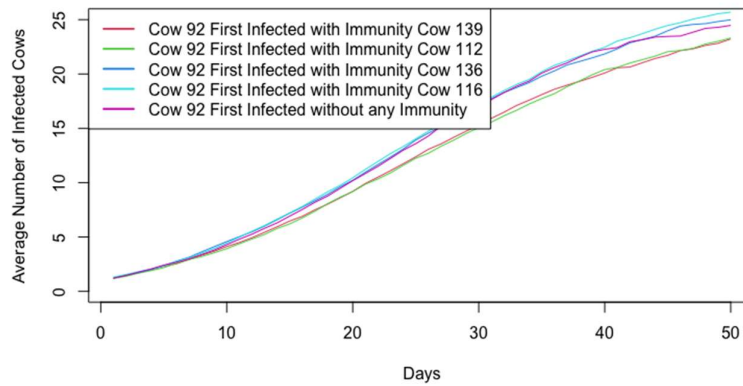


Figure A6

Cow 116: Average Number of Infected Cows per Day with Different Immunity vs. No Immunity Scenarios - Right side of Barn

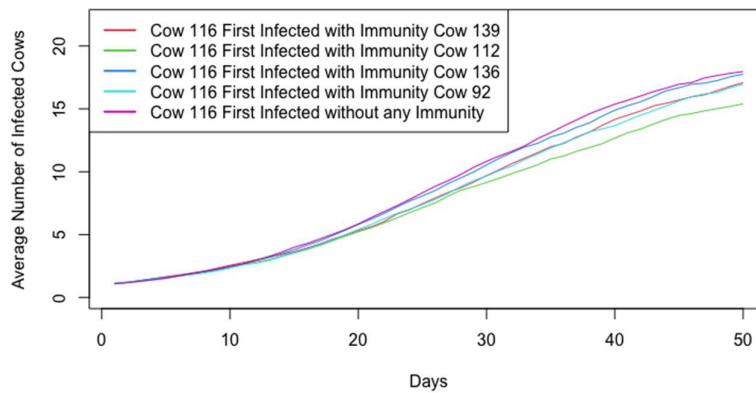


Figure A7

Cow 136: Average Number of Infected Cows per Day with Different Immunity vs. No Immunity Scenarios - Right side of Barn

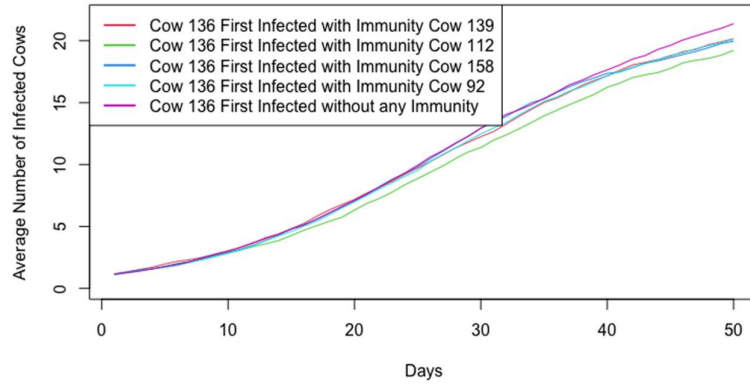
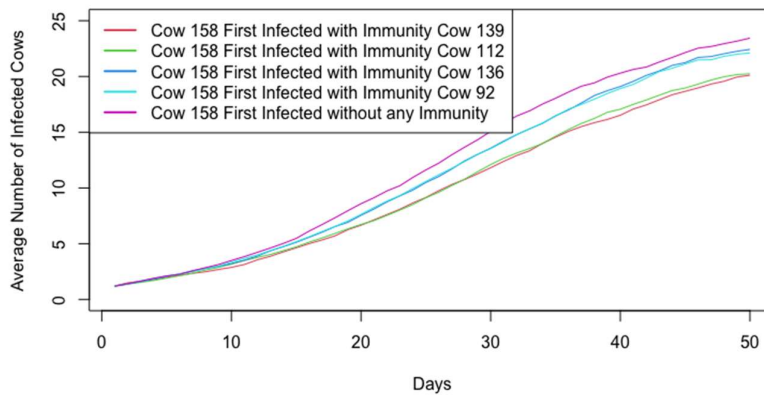


Figure A8

Cow 158: Average Number of Infected Cows per Day with Different Immunity vs. No Immunity Scenarios - Right side of Barn



Appendix B

Diameter & Neighbors

Table B1

Cow Id from those individuals part of the Diameter from both networks

	Left	Right
Cow ID	39, 9, 1, 2, 5, 57	80, 95, 89, 116, 85

Neighbors

Table B2

Cow Id from those individuals identified as neighbours of Cow 92 and 158 (right side). The column "Present in the Clique" identify which of the neighbors are in the clique.

Cow ID	Neighbors	Present in the Clique
92	81, 86, 87, 91, 98, 106, 122, 133, 135, 146, 148, 152	152
158	93, 122, 123, 137, 141, 143, 147, 148, 151, 154, 159,	141

Appendix C

Outcomes from ERGM

Table C1

ERGM results Model 1 – Left side

Call:	Estimate	Std. Error	MCMC %	z value	Pr(> z)
edges	-2.35102	0.06393	0	-36.78	<1e-04 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null Deviance: 4271 on 3081 degrees of freedom

Residual Deviance: 1821 on 3080 degrees of freedom

AIC: 1823 BIC: 1829 (Smaller is better.)

Table C2

ERGM results Model 2 – Left side

Call:	Estimate	Std. Error	MCMC %	z value	Pr(> z)
Edges	-2.42632	0.07649	0	-31.72	<1e-04 ***
nodematch.parity	0.27060	0.13950	0	1.94	0.0524 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null Deviance: 4271 on 3081 degrees of freedom

Residual Deviance: 1817 on 3079 degrees of freedom

AIC: 1821 BIC: 1833 (Smaller is better.)

Table C3

ERGM results Model 3 – Left side

Call:	Estimate	Std. Error	MCMC %	z value	Pr(> z)
edges	-2.515175	0.175525	0	-14.329	<1e-04 ***

nodefactor.parity.2	0.092937	0.121483	0	0.765	0.4443
nodefactor.parity.3	0.288308	0.139589	0	2.065	0.0389*
nodefactor.parity.4	-0.006461	0.134168	0	-0.048	0.9616

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null Deviance: 4271 on 3081 degrees of freedom
Residual Deviance: 1815 on 3077 degrees of freedom
AIC: 1823 BIC: 1848 (Smaller is better.)

Table C6

ERGM results Model 1 – Right side

	Estimate	Std. Error	MCMC %	z value	Pr(> z)
Edges	-2.11108	0.05659	0	-37.3	<1e-04 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null Deviance: 4492 on 3240 degrees of freedom
Residual Deviance: 2219 on 3239 degrees of freedom
AIC: 2221 BIC: 2227 (Smaller is better.)

Table C7

ERGM results Model 2 – Right side

	Estimate	Std. Error	MCMC %	z value	Pr(> z)
Edges	-2.26868	0.06957	0	-32.611	<1e-04 ***
nodematch.parity	0.54183	0.12039	0	4.501	<1e-04 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null Deviance: 4492 on 3240 degrees of freedom
Residual Deviance: 2199 on 3238 degrees of freedom
AIC: 2203 BIC: 2215 (Smaller is better.)

Table C8

ERGM results Model 3 – Right side

	Estimate	Std. Error	MCMC %	z value	Pr(> z)
Edges	-1.95886	0.14273	0	-13.725	< 1e-04 ***
nodefactor.parity.2	-0.06752	0.10662	0	-0.633	0.52652
nodefactor.parity.3	-0.41909	0.13526	0	-3.098	0.00195**
nodefactor.parity.4	0.02664	0.10704	0	0.249	0.80347

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null Deviance: 4492 on 3240 degrees of freedom
Residual Deviance: 2205 on 3236 degrees of freedom
AIC: 2213 BIC: 2238 (Smaller is better.)