CARLETON UNIVERSITY

SCHOOL OF
MATHEMATICS AND STATISTICS

HONOURS PROJECT

TITLE:	Using Network Analysis for Controlling Disease Outbreak for African Swine Fever Virus
AUTHOR:	Madeleine Davis
SUPERVISOR:	Dr. Patrick Farrell
DATE:	May 4th, 2020

TABLE OF CONTENTS

ABSTRACT

With an outbreak of African Swine Fever (ASF) overseas (Haley & Gale, 2020), Canadian pig farmers are rightfully concerned about a potential outbreak in North America. For a virus as deadly and contagious as ASF, Canada is using all of its resources to keep our pigs safe (Canadian Pork Council (CPC), 2020). Should Canada witness an outbreak of ASF in our pig farms, measures are in place to prevent spread and contain the disease in a timely manner (Canadian Food Inspection Agency (CFIA), 2019). However, at the centre of an outbreak, when resources are running dry, it is vital to have measures in effect to avoid the potentially devastating impact ASF could have on our pork industry.

This paper aims to explore the use of network analysis on pig movements around Canada to find a way of reducing spread of ASF should there be an outbreak. In Canada, livestock movement reporting is mandatory (CPC, 2020). Using a synthesized dataset based on real pig movement reports, we are able to analyse the network using five node-level metrics, namely degree, infection chain, betweenness, mean infection potential, and purchase-sales-balance. Based on these metrics, this paper also discusses a new metric, degree balance. Using these metrics, we can simulate an outbreak and calculate the impact of removing "super-spreaders" defined to be premises with high mean infection potential or high betweenness to great success. By removing only 10% of premises in the synthesized pig movement network, the number of premises infected can be reduced by up to 70%. With these kinds of results, it is clear to see that network analysis can drastically assist in the mitigation of an ASF outbreak.

1 INTRODUCTION

As defined by the World Organization for Animal Health (OIE), "African Swine Fever (ASF) is an infectious disease of domestic and wild pigs of all breeds and ages, caused by the ASF virus (ASFV)" (World Organization for Animal Health (OIE), 2019). Since the 1900's, the disease has survived in wild hogs in Africa (Canadian Food Inspection Agency (CFIA), 2019), although in recent years, ASF has been discovered in Europe and Asia. In 2007, ASF emerged in the Republic of Georgia, and subsequently traveled throughout Europe (Canadian Food Inspection Agency (CFIA), 2019).The most recent outbreak was detected in China in 2018 (CPC, n.d.), greatly impacting the country's economy (Pitts & Whitnall, 2019), and killing many of its hogs (both domestic and wild). ASF then spread to many other countries and is still infecting more pigs around the globe today.

African Swine Fever Virus is not only highly contagious (CPC, n.d.), it is extremely fatal. "(ASF) is arguably the most dangerous swine disease, which threatens wild boar and domestic [pig] populations worldwide. The mortality rate is approaching 100%" (Malogolovkin, Sereda, & Kolbasov, 2020, p. 27), making it all the more concerning that to this date, there is no known vaccine for ASFV.

In 2019, China lost 440 million pigs, either from pigs being infected or from culling animals as an attempt to limit the spread of the disease (Malogolovkin, Sereda, & Kolbasov, 2020). Historically, China has been the leading pig producer, accounting for 50% of pigs worldwide (Pitts & Whitnall, 2019). Since the outbreak of ASF in August of 2018, 40% of Chinese hogs have died (Haley & Gale, 2020), meaning that almost 25% of the world's hog population has died from the disease. In Europe, one million pigs have died since the Georgian outbreak of 2007 (CPC, n.d.).

It is clear that there is a need to prepare for a potential outbreak in Canada should one occur. This paper focuses on the use of network analysis as a potential means of modeling the control of the disease within Canada.

2 MATERIALS AND METHODS

2.1 DATA

The data available is based on hog movement data which has been provided by the Canadian Pork Council (CPC) to use for this project. The data, referred to as PigTrace, consists of both incoming and outgoing reports of hog movements reported by pig farmers across Canada from 2017 to the end of 2019. The reports, which farmers are required to report within seven days of a movement, contain the date of the event, the premise ID number and premise name of the reporting party, the premise ID number of the receiving party, the premise ID number of the sending party, the event type (more on this later), and the lot quantity (i.e. the number of pigs in the movement). For the majority of cases, pigs in Canada do not have identification tags, and so the number of pigs in a particular movement is used. The event type variable is defined as being one of the following: move-in, move-out, slaughter, import, export, or what is called "rendering or deadstock pickup". The latter refers to the event of a truck collecting pigs who died before slaughter, and all of the movements labelled as such have a lot quantity of zero. The dataset also contains other variables such as stakeholder information, licence plate of the truck used to move the pigs, and the date the event was reported. Due to the time constraint on this project, these variables were not used.

Although we had access to the PigTrace data, due to confidentiality concerns, we instead used a synthesized dataset created by our partners at CPC (2018). Unlike the PigTrace data, the dummy dataset only has record of one side of each movement. To assimilate the true dataset, the other half of the records were imputed. The date range on the synthesized dataset spanned over six years, so we transformed the date range to be within 2018 in order to have sufficient data to analyse a temporal network. The cleaning process (described later) was created for the true dataset but was then applied to the dummy dataset for completeness.

2.2 CLEANING STEP

After consulting with CPC and the Methodology Branch of Statistics Canada, it was decided that the cleaning process would consist of three main steps. The first step is to remove known duplicates. CPC informed us that one particular abattoir had been reporting movements twice – once as a "move-in" event and once as a "slaughter" event. This meant that for that particular premise, the number of incoming pigs counted would be doubled. To account for this, since there were more move-in events than slaughter events, the data was first sorted to identify duplicate rows, then all move-in events which matched a slaughter event were removed. Move-in events which did not have a match were left in the dataset as they were.

Next, we set aside import, export, and rendering and deadstock pickup movements. In terms of controlling a disease outbreak, imported movements might indicate where an outbreak occurred, but since the project is focused on finding potential Canadian hog premises which may be infected, the record detailing the import of pigs is not useful. In a similar vein, records of exported pigs, while useful in determining where infected pigs may have been sent, does not prove useful for the purposes of this project. The rendering and deadstock pickup movements are removed for a few reasons. First, since they are reporting deadstock and not live hogs, they are all marked as having a lot quantity of zero. Second, in terms of disease transmission, the rendering and deadstock pickup trucks are considered to be the end of the chain and are unlikely to pass disease on to other premises. Although some similar articles have considered trucks when it comes to possible transmission, in the interest of time, this paper only considers farm to farm movements as possible modes of transmission.

The final step of the cleaning process is to sum identical records' pig counts. We noticed by manually looking through the data that there are some movements where, for example, Premise A reports sending 60 pigs to Premise B, but B reports receiving two shipments of 30 pigs each. This is likely the result of a truck shipping the pigs in two trips on the same day. To account for this, we simply found records that matched exactly (i.e. date, truck licence plate, sending and receiving premise ID numbers, and event type all match) and created one row with the total

number of hogs. This is particularly useful in the next process of linking movements from the shipping and receiving ends of the transactions.

The matching process begins by looking at exact matches. This is defined as an outgoing record and an incoming record where both records have the same date, same sending and receiving premise ID's, and same lot quantity. In the PigTrace data from 2017 to 2019, this captures roughly 70% of cases. The next step in the matching process is to look at near perfect matches. It is worth noting that since the dummy dataset has entirely imputed record matches, 100% of the dummy records fall under the exact matches category. The following steps in the matching process are intended solely for the true dataset.

The PigTrace data has many cases where the shipping and receiving sides of the same movement may be different. While the lot quantity may differ due to a movement being split into two truckloads as mentioned earlier, there may also be a difference in incoming and outgoing lot quantities due to pigs dying in transport, or slight rounding of one of the reporting parties. From consulting with the industry, we learned that it is more likely for incoming movements to be exact, since they are paying for each live pig arriving in the truck. There are some instances in the data where the outgoing movement may report a lot quantity of say 200 pigs, but the incoming movement reports a lot quantity of 204 pigs. Based on the above, we may assume (provided there is no similar outgoing report of 204 pigs between the two premises in question) that the outgoing record rounded their lot quantity. In the event that the outgoing quantity is higher than the incoming quantity, we may continue to assume four pigs died in transit. Due to these reasons, when matching incoming and outgoing records, provided an exact match (as defined above) doesn't exist, the matching process looks for a record where the date, and sending and receiving premise ID numbers match, but the lot quantity may be plus or minus up to ten pigs. Note that the number ten was chosen to be something small enough to suggest that differences in pig numbers could logically be attributed to the above reasons. This number can be easily changed if there is reason to believe another number would be better. It is also worth noting that in the case where a particular movement seems to have a match with multiple movements, where each potential match has a lot quantity within plus or minus ten, preference

is given to movement pairs where the outgoing record has a higher lot quantity than the incoming record.

There may also be cases where a truck leaves on Monday and arrives on Tuesday after stopping overnight. Although rare, it is possible for this to occur in such a large country, and so the matching process looks for records with a difference in date of one, provided a better match doesn't exist. Given that farmers are permitted a seven-day latency period in reporting, it is expected that in some cases, the date may be off due to recall error, and so the process also looks for cases where the incoming event is reported one day before the outgoing event, giving the converse case priority. Also due to the fact that farmers have up to seven days to report, we can assume that there will always be cases where the other side of the record pair has not yet been reported. These records will not have a match and will require imputing the second side of the report and will be handled later on in the matching process.

After running the PigTrace data through the matching process, it was found that only 80% had a match. For the purposes of disease control, we imputed pairs for the remaining 20% of records. As mentioned previously however, we continue to use the synthesized dataset due to concerns about confidentiality.

2.3   NETWORK METRICS

One of the most common approaches to developing disease control strategies is to use livestock movement data to perform a network analysis (Noremark, et al., 2011). To do this, premises sending or receiving pigs are considered nodes (or vertices) of the network, and shipments of pigs are considered as edges connecting two nodes. The livestock movement network can then be described as an unweighted dynamic network (Kinsley, Perez, Craft, & Vanderwaal, 2019), meaning that dates of movements are considered. In the case of pig systems, since pigs in Canada do not require identification tags, movement records contain a total number

of hogs, which can then be interpreted as the weight of the edge, therefore the pork movement network can be described as a weighted dynamic network.

To describe the movement network, we can look at multiple different metrics. For the purposes of this paper, we will look into metrics shown in the literature to be important when analysing potential disease outbreaks. In this study, the following metrics were calculated for the synthesized dataset:

i) In-degree and out-degree (Wasserman & Faust, 1994): directed measure of the number of contacts with a particular premise. In-degree pertains to the number of premises selling pigs to a particular premise, whereas the out-degree pertains to the number of premises receiving pigs from a particular premise. An example of in- and out-degree is seen in Figure 1.

ii) Ingoing and outgoing infection chain (Dubé, et al., 2008; Noremark & Widgren, 2014): directed and temporal measure of the number of direct or indirect contacts with a particular premise. Ingoing infection chain includes all premises who may
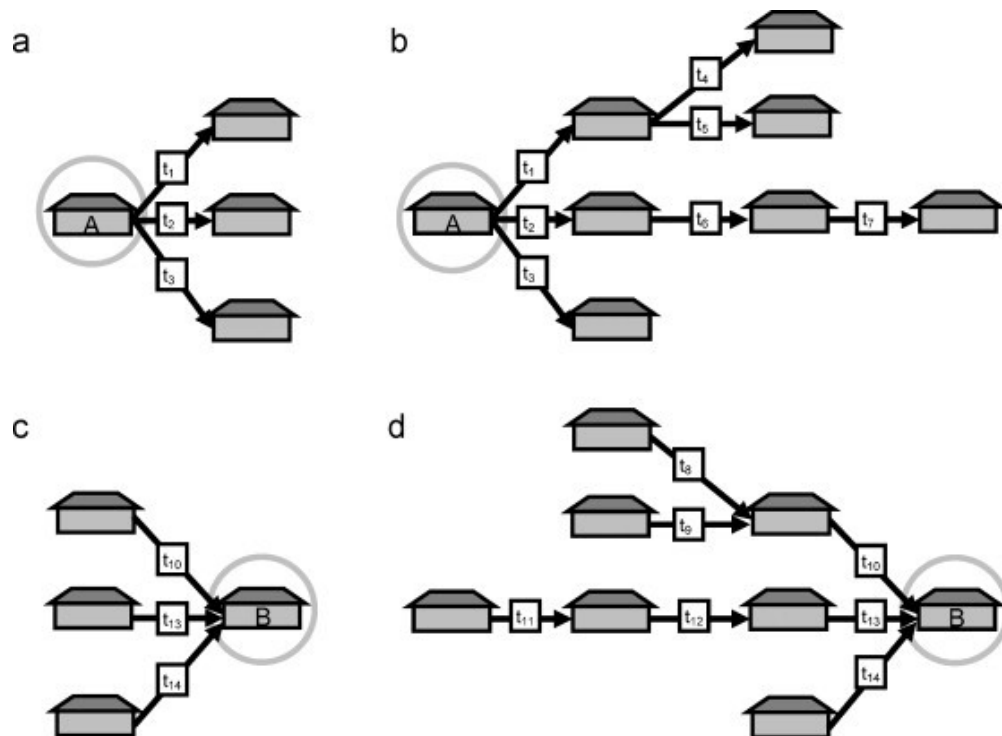

Figure 1 - Degree and Infection chain example from Noremark et al., 2011. Premise A has an out-degree of 3 and an outgoing infection chain of 7, whereas premise B has an in-degree of 3 and an ingoing infection chain of 7. Note that $t_1 \leq t_2 \leq \cdots \leq t_m$.

sell pigs directly or indirectly to a particular premise, whereas outgoing infection chain includes all premises who may receive pigs directly or indirectly from a particular premise. The temporal aspect ensures that for example in the case that premise A ships pigs to premise B after premise B ships pigs to premise C, premise A is not counted in the ingoing infection chain of premise C. An example of ingoing and outgoing infection chain can be seen in Figure 1.

iii) Mean infection potential (Rossi, et al., 2017; Kinsley, et al., 2019): a metric comparing both ingoing and outgoing infection chain over time as an average of the *infection potential*, defined as follows:

$$\rho_i(d) = \frac{IIC_i(d)}{N-1} OIC_i(d),$$

where $IIC_i(d)$ and $OIC_i(d)$ are ingoing and outgoing infection chains respectively for premise i on day d, and N is the total number of premises in the network. The *mean infection potential* is then defined as follows:

$$\rho_i = \frac{1}{m} \sum_{d=1}^{m} \rho_i(d),$$

where $\rho_i(d)$ is the infection potential of premise i defined above and m is the infection period of interest. Rossi suggests only considering the infection chain on a finite *infectious period* γ. For our purposes, we will be calculating infection chain on date d with an infectious period γ equal to 21 days, as this is the recommended length of quarantine from ASF (World Organization for Animal Health (OIE), 2019, Article 15.1.9)[1]. Our duration of interest m is set to be 284, given that the synthesized data covers 344 days in 2018 and an allowance of 30 days at the start and end of the data is needed to account for potential boundary effects (Rossi, 2017).

iv) Purchase-Sales-Balance (Koeppel, et al., 2018): a bounded measure referring to the ratio of pigs purchased to pigs sold, normalized to be between -1 and 1 as follows:

---

[1] Although the incubation period for African Swine Fever is 15 days (World Organization for Animal Health (OIE), 2019, Article 15.1.1), the infectious period is longer to ensure minimal risk.

$$B = \frac{P - S}{P + S}$$

where P is the total number of pigs purchased in the time frame of interest and S is the total number of pigs sold in the time frame of interest.

v) Betweenness (Wasserman & Faust, 1994): a metric describing the frequency that a premise is a link appearing on the shortest path between two nodes. In network analysis, the term "shortest path" refers to the minimum number of links connecting two nodes (Wasserman & Faust, 1994). Betweenness can be considered as either a node metric as defined above, or as a network metric called *group betweenness centralization index* (GBCI) (Wasserman & Faust, 1994), defined as "the variability in the betweenness of the [premises]" (Noremark, et al., 2011, p. 82). In particular, the GCBI can be calculated as follows:

$$GBCI = \frac{2}{(g - 1)^2(g - 2)} \sum_{i=1}^{g} \left(B(n^*) - B(n_i) \right),$$

where g is the number of nodes (n_i) in the network, $B(n)$ indicates the betweenness measure for a particular node, and $B(n^*)$ is the highest betweenness measure of all the nodes in the network (Wasserman & Faust, 1994). GBCI is then bounded by 0 and 1.

2.4 SOFTWARE

To analyse the above metrics, we used the packages "igraph" (Csardi & Nepusz, 2006) and "EpiContactTrace" (Noremark & Widgren, 2014) in R (R Core Team, 2020). Specifically, functions IngoingContactChain and OutgoingContactChain form the EpiContactTrace package were used to calculate the ingoing and outgoing infection chain and the mean infection potential. Betweenness and degree were calculated using the functions in igraph. Epidemic simulations were calculated with "epinet" (Groendyke & Welch, 2018). Details of the code can be found in Appendix A.

3 RESULTS & DISCUSSION

The synthesized dataset contained movements of pigs between thirty-nine premises over the span of 344 days in 2018. The average size of movement was 213.4 pigs, with a maximum of 6320 pigs in a movement. Of the 29 premises which send pigs, the average number of outgoing movements was 14.21 per premise. The premises sending pigs sold on average 3032 pigs over the course of the year. Of the 34 premises receiving pigs, the average number of ingoing movements was 12.12 per premise. The premises receiving pigs purchased on average 2586 pigs over the course of the year. Five node-level network metrics were calculated for each of the premises, including degree, infection chain, betweenness, mean infection potential and purchase-sales-balance. In this section, each metric will be discussed, with further details provided in Table 1. The network can be seen in Figure 2.
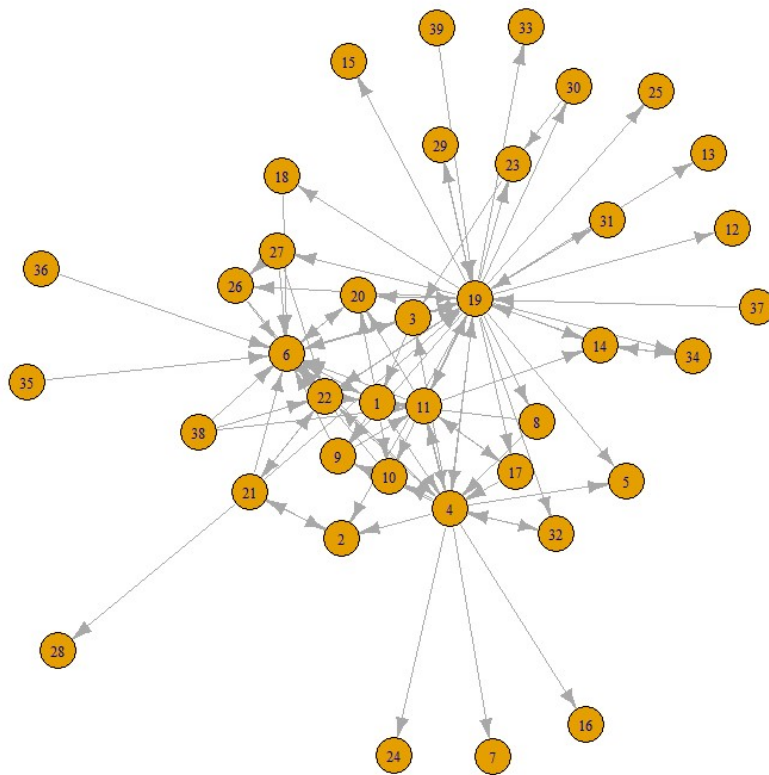

Figure 2 - Simplified view of the network being analysed. Nodes represent premises 1 through 39, and edges represent the existence of at least one shipment.

Measure	0%	25%	50%	Mean	75%	100%
In-Degree	0	1	1	2.59	3	16
Out-Degree	0	0.5	1	2.59	2.5	27
Ingoing Infection Chain	0	2	24	15.95	25	26
Outgoing Infection Chain	0	0.5	23	15.95	25	32
Betweenness	0	0	0	35.90	15.73	670.57
Mean Infection Potential	0	0	0	0.014	0.007	0.204
Purchase-Sales-Balance	-1	-0.473	0.612	0.233	0.997	1

Table 1 – Summary of Network Metrics

The in- and out-degree both averaged at 2.59. It makes sense that they both have the same average since the sum of in-degree and out-degree is the same. This is because for every link into a premise, there exists a premise with an outgoing link. The in-degree and out-degree are both highly skewed to the right, with maximums at 16 and 27 respectively. This pattern is also found by Noremark (2011) among both swine and cattle data. In terms of disease control, it is important that the in- and out-degrees are highly right skewed since it implies that very few premises are "hubs" at which diseases can spread to many different premises.

As with in- and out-degree, the ingoing and outgoing infection chains will have the same averages. The average infection chain is 15.95. Interestingly, the ingoing and outgoing infection chains are not right skewed. Although there are a decently large number of premises with low infection chains, the medians of the ingoing and outgoing infection chains are 24 and 23 respectively, implying that there is a significant number of premises with high infection chains. With only 39 premises in the network, these numbers represent premises who have direct or indirect connections with 61.5% and 59.0% of the premises in the network respectively. The maximum values for ingoing and outgoing infection chain are 26 and 32 respectively, associating with premises which have direct or indirect connections with 66.7% and 82.1% of the premises in the network. These results are contradictory to the swine and cattle networks in Noremark's study (2011), which are highly right skewed. This could have implications on the types of

operations in the network. In Canada, pig premises generally fall into different categories of operation types defined with the age of the pigs on a particular premise. For example, premises who farrow pigs and keep them until they are weaned are considered farrow-wean operations, whereas premises who keep pigs from farrow until they are sent off for slaughter are considered farrow-finish operations. In Noremark's study of pig populations in Sweden, operation types are listed as nucleus or multiplier, sow pool, farrow to grower, farrow to finish, and fattening. In Canada, operation types include farrow-wean, farrow-grow, nursery, wean-finish, grow-finish, farrow-finish, nucleus, gilt development, AI Stud, and abattoir (or slaughterhouse, which should be noted is not included in Noremark's study)[2]. It is possible that with set operation types, there is more movement along a "chain", thus allowing for premises to have high infection chain while maintaining a low degree, or put simply, a premise will have few direct connections but many indirect connections. It should also be noted however that the data being analysed here is simulated data and another study would have to be made to see if real Canadian data follows the same pattern.

The average premise betweenness was 35.9. However betweenness is highly right skewed, as the minimum and median are both 0. The maximum betweenness is 670.57. For a network with 39 nodes, a node can appear on a maximum of 703 paths (since there are 741 combinations of pairs from 39 nodes, and 38 pairs include the particular node of interest as either a source or a destination). Thus, the premise with the highest betweenness measure appears on 95.4% of the possible shortest paths. It is worth noting however that this extreme value is likely only to appear in the synthesized data and is not representative of the true pig network. It is obvious that premises with high betweenness measures are dangerous when it comes to the spread of disease, since they can potentially transfer the disease to other parts of the country or region which normally are not connected. In the synthesized network, the premise with the highest betweenness measure acts as a central node to many other premises which would otherwise not be connected, and thus acts as what is considered a "super-spreader" in terms of disease. The group betweenness centralization index (GBCI) of the network is very high at

---

[2] This information was provided to us from the Canadian Pork Council (CPC) via personal communication, May 2019

92.66%, indicating that there is one premise (i.e. the "super-spreader") which serves as a link between most other premises in the network.

Mean infection potential was calculated in R (R Core Team, 2020) using the code provided in Appendix A. The first function calculates the infection potential of each premise on a particular day. The second function then calculates the infection potential of each premise for each day in a set number of days and calculates an average. In Rossi's study (2017), it was mentioned that there could be boundary effects in the data, and so although a year of data was available, the infection potential was computed for 245 days. Rossi also used an infectious period γ of 14 days but considered the effects of choosing values from 3 to 28 days. In Canada, given that pork farmers are allowed to report up to seven days after an event[3], we allowed for more days to be included in our analysis, while still maintaining space to avoid the boundary effects. With 344 days available in the synthesized data, we were able to remove 30 days from each end, leaving an m of 284 days in the analysis. We chose γ to be 21 days based on the World Organization for Animal Health's recommended quarantine time for pigs with African Swine Fever Virus (OIE, 2019). The mean infection potential was highly skewed, with a minimum and a median of zero, a mean of 0.014 and a maximum of 0.204.

Many researchers have considered the premise operation type when comparing node-level metrics (Koeppel, et al., 2018; Kinsley, et al., 2019; Noremark, et al., 2011; Thakur, et al., 2016). Given that the operation type is not necessarily available in the PigTrace data, it is worth analysing the ratio of purchases to sales to determine basic operation types. Following the suggestion of Koeppel et al. (2018), the purchase-sales-balance defined in section 2.3 will be used. Koeppel et al. defined breeding farms as those with a balance of less than -0.5, slaughterhouses as premises with a balance of greater than 0.7, and fattening farms as those in between. With this approach, 17 out of the 39 premises would be considered slaughterhouses and 10 would be considered breeding farms. As seen in the histogram for the purchase-sales-balance in Appendix B, there are a large number of premises near the extremes (-1 and 1). This is likely unrealistic, as there are not as many slaughterhouses in the country as there are breeding

---

[3] This information was provided to us from the Canadian Pork Council (CPC) via personal communication, May 2019

farms. However, these results are likely attributable to the data being synthesized. Overall, the purchase-sales-balance had a mean of 0.23 and a median of 0.61.

While the purchase-sales-balance can be used to estimate the operation types of the premises, it can theoretically also be used in congruence with in- and out-degree to get a more accurate picture of the prevalence of a particular premise in the network as a whole. While in- and out-degree measure the number of premises a node is directly connected to, the purchase-sales-balance represents the ratio of pigs entering and leaving the premise. Thus, we will define a new metric, degree balance, in the same manner as purchase-sales-balance, as follows.

$$Degree\ Balance = \frac{InDegree - OutDegree}{InDegree + OutDegree}$$

This metric is a measure of connective traffic bounded between -1 and 1. A value of positive one suggests the premise only receives pigs, meaning it is likely to be a slaughterhouse. A degree balance of negative one suggests the premise only sells pigs, meaning it is likely to be a farrowing operation. Premises with a degree balance near zero indicate premises who have an equal number of shipping and receiving partners. Note as well that when a premise has a degree balance of 1 or -1, it will also have a purchase-sales-balance of 1 or -1. We can then compare degree balance with purchase-sales-balance as seen in Figure 3. The Pearson's product-moment correlation coefficient is 0.87, with a p-value of 6.2×10^{-13}. The degree balance of the network was much more evenly distributed than the purchase-sales-balance. The average degree balance was 0.16, and the median was 0.14, where the mean and median for purchase-sales-balance were 0.23 and 0.61, respectively. Histograms of both metrics can be found in Appendix B.

Degree is by definition less than or equivalent to infection chain since infection chain contains direct contacts as well as indirect contacts. While they are related, they are not necessarily indicative of each other. The Pearson's correlation coefficient between degree and infection chain[4] was only 46.53% with a p-value of 2.8×10^{-3}. When comparing the difference

---

[4] Tests for correlation between in-degree and ingoing infection chain gave the same results as the test between out-degree and outgoing infection chain to four decimal places.

Figure 3 – Comparing purchase-sales-balance and degree balance shows a near one-to-one correlation

between infection chain and degree (i.e. the number of indirect connections) with degree (i.e. the number of direct connections), the correlation coefficient was even lower, and did not disprove the alternate hypothesis that the correlation was not zero. The correlation coefficient between in-degree and indirect ingoing connections was 0.19 with a p-value of 0.25 and the correlation coefficient between out-degree and indirect outgoing connections was 0.08 with a p-value of 0.65. These findings are in agreeance with the findings of Noremark (2011).

Rossi et al. (2017) defines super-spreaders to be farms with a mean infection potential above the 95th percentile. It is of interest then to use the metrics described above to identify potential super-spreaders in the network. Kinsley et al. (2019, p. 1) mentions that based on their simulations, "by targeting farms based on their mean infection potential, we could reduce the potential spread of an infectious pathogen by 80% when removing approximately 25% of farms"[5]. With the potential uses for mean infection potential in mind, it is of interest to use this metric as

---

[5] Although it is worth noting that Kinsley et al.'s results for mean infection potential do not agree with neither Rossi et al.'s nor our numbers, since Kinsley et al.'s numbers range from 0 to 70, Rossi et al.'s numbers do not exceed 0.02, and ours do not exceed 0.2 in the extreme case.

well as the other metrics previously discussed to identify super-spreaders and remove them from the network in the event of a disease outbreak to mitigate spread early on. In particular, premises with high infection chain, high mean infection potential, high betweenness, and degree balance and purchase-sales-balance near zero would be very high-risk nodes in the network.

To try and capture the effect of removing some potential super-spreaders, we ran a simulation using the R-package "epinet" (Groendyke & Welch, 2018). From the 39 premises, the nodes with the highest 10% mean infection potential were removed from the network. The number of infected premises and the entire length of the epidemic was calculated for 50 simulations once for the full network and once for the network after removing identified super-spreaders. The number of infected premises dropped on average 30 percentage points when only 10% of premises were removed, and the entire epidemic lasted on average 7.4 days shorter. The 95% confidence interval size for the percent of premises infected in the reduced network was 3.87 percentage points, where the interval size for the percent of premises infected in the full model was 3.26, suggesting that a 30-percentage point difference is significant.

With such positive results from following the methods of Rossi et al. (2017) and Kinsley et al. (2019), it was of interest to try removing the premises within the top 10% of a different variable. For this, we compared the full network with the network after removing the premises with the top 10% betweenness measure. Again, the number of infected premises and the length of the epidemic was calculated for 50 simulations. By removing the premises with the highest betweenness measure, the number of infected premises dropped by 71-percentage points on average, and the length of the epidemic was on average 23.4 days shorter. This is clearly a much better metric to use to identify super-spreaders compared to using the mean infection potential. The 95% confidence interval size for the average percent of infected premises with the reduced network is 1.17, implying that a 71-percentage point difference is significant.

4   CONCLUSION

To summarise, this paper suggests that it is possible to drastically reduce the spread of a disease by removing a relatively small number of premises from the network by using network metrics to identify super-spreaders. In the event that African Swine Fever (ASF) virus enters Canada, removing super-spreaders early could determine whether our pork industry can survive an ASF outbreak, or whether such devastating effects seen in the pork industry of China (Haley & Gale, 2020) would happen here.

Moving forward, it is of interest to determine the impact of removing premises with high numbers for the other metrics to compare with the impact of using mean infection potential. Tests should also be done to determine if combinations of metrics can be used to identify super-spreaders. In the event of an outbreak, it is important to weigh the economic impact of shutting down 10% of premises with the economic results of an outbreak should the super-spreaders not be contained. This analysis has not been done for this project.

Given more time, it would also be of interest to research the impact of other parties in the network such as trucks, veterinarians, and feed producers. These larger networks were analysed in both Thakur et al. (2016) and Rossi et al. (2017) with the common conclusion that other networks, not just the farms networks, need to be considered in the event of an outbreak. As mentioned in section 2.1, the data includes a variable for licence plate of the truck used in the movement. Due to the way the data was cleaned (in particular the cleaning step of summing total movements between two premises in one day), this variable could not be used. By changing the method of matching ingoing and outgoing records, it may be possible to use the licence plate variable to identify trucks. It is worth noting however that this variable is susceptible to much more human error than other variables.

To continue the results found in this paper, it would be interesting to compare the node-level metrics among different operation types as has been done in the literature (Koeppel et al. (2018); Kinsley et al. (2019); Thakur et al. (2016); Noremark et al. (2011)). The reason this was not done in this paper was that the operation types would have to be estimated for the synthesized dataset, and any comparisons made would be unreliable. In the real PigTrace

dataset, the operation types are known for a sample of the premises. Using the sample data, it would be possible to compare node-level metrics among operation types with a higher level of accuracy.

It should be noted that the dataset used for this analysis was entirely fictional, and thus cannot definitively represent the real PigTrace data. Using the real data may present different challenges on top of the cleaning and matching processes. More research should be done with the real data before conclusions regarding an alternative method for controlling disease outbreak based on this project are reached. Rather, the purpose of this paper is to shed light on a possible supplemental approach to assist the current disease control measures.

The analysis done in this paper can be applied to more than just ASF in Canadian pigs. There is room to apply the results of this paper to other livestock diseases, such as Foot and Mouth Disease. Similar ideas discussed in this paper could have an impact on the tracing of viruses that affect humans. While there are no reports of human movement, there is growing interest in the industry in using cellphone data to track high-traffic areas (Kasson, 2020). It is possible that using this information, potential super-spreaders could be detected in a similar way to the methods discussed here.

BIBLIOGRAPHY

Canadian Food Inspection Agency (CFIA). (2019). *African Swine Fever Forum 2019 – Backgrounder Document.* Retrieved March 2020, from CFIA Web Site: https://inspection.gc.ca/DAM/DAM-animals-animaux/STAGING/text-texte/dis_africswine_event_asf_backgrounder_document_1568998449240_eng.pdf

Canadian Pork Council (CPC). (2018). Simulated Pig Movement Events 2018 [datafile].

Canadian Pork Council (CPC). (2020). *What are we doing to prevent African swine fever from infecting Canadian pigs?* Retrieved March 2020, from Canadian Pork Council Web Site: https://www.cpc-ccp.com/actions-to-prevent-asf

Canadian Pork Council (CPC). (n.d.). *African swine fever in other countries*. Retrieved March 2020, from Canadian Pork Council Web Site: https://www.cpc-ccp.com/asf-in-other-countries

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *Complex Systems*(R version 3.6.3), 1695. Retrieved from http://igraph.org

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

Groendyke, C., & Welch, D. (2018). epinet: An R Package to Analyze Epidemics Spread acorss Contact Networks. *Journal of Statistical Software, 83*(11), 1-22. doi:10.18637/jss.v083.i11

Haley, M., & Gale, F. (2020). *African Swine Fever Shrinks Pork Production in China, Swells Demand for Imported Pork*. Retrieved March 2020, from United States Department of Agriculture Web Site: https://www.ers.usda.gov/amber-waves/2020/february/african-swine-fever-shrinks-pork-production-in-china-swells-demand-for-imported-pork/

Kasson, P. M. (2020). Infectious Disease Research in the Era of Big Data. Annual Review of Biomedical Data Science, 3.

Kinsley, A., Perez, A., Craft, M., & Vanderwaal, K. (2019). Characterization of swine movements in the United States and implications for disease control. *Preventive Veterinary Medicine, 164*, 1-9.

Koeppel, L., Siems, T., Fischer, M., & Lentz, H. (2018). Automatic classification of farms and traders in the pig production chain. *Preventive Veterinary Medicine, 150*, 86-92.

Malogolovkin, A., Sereda, A., & Kolbasov, D. (2020). African Swine Fever Virus. In Y. Malik, M. Yadav, & R. Singh (Eds.), *Emerging and Transboundary Animal Viruses* (pp. 27-54). Singapore: Springer.

Noremark, M., & Widgren, S. (2014). EpiContactTrace: an R-package for contact tracing during livestock disease outbreaks and for risk-based surveillance. *BioMed Central Veterinary Research, 10*, 71.

Noremark, M., Hakansson, N., Sternberg Lewerin, S., Lindberg, A., & Jonsson, A. (2011). Network analysis of cattle and pig movements in Sweden: Measures relevant for disease control and risk based surveillance. *Preventive Veterinary Medicine, 99*, 78-90.

Pitts, N., & Whitnall, T. (2019). *Impact of African swine fever on global markets*. Retrieved March 2020, from Australian Government Department of Agriculture, Water and the Enviroment: https://www.agriculture.gov.au/abares/research-topics/agricultural-commodities/sep-2019/african-swine-fever

R Core Team. (2020). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Rossi, G., De Leo, G., Pongolini, S., Natalini, S., Zarenghi, L., Ricchi, M., & Bolzoni, L. (2017). The Potential Role of Direct and Indirect Contacts on Infection Spread in Dairy Farm Networks. (M. Ferrari, Ed.) *PLoS Comput Biol, 13*(1), e1005301. Retrieved from https://doi.org/10.1371/journal.pcbi.1005301

Thakur, K., Revie, C., Hurnik, D., Poljak, Z., & Sanchez, J. (2016). Analysis of Swine Movement in Four Canadian Regions: Network Structure and Implications for Disease Spread. *Transboundary and Emerging Diseases, 63*, e14-e26.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications.* Cambridge, UK: Cambridge University Press.

World Organization for Animal Health (OIE). (2019). Chapter 15.1. Infection with African Swine Fever Virus. In *Terrestrial Animal Health Code.*

World Organization for Animal Health (OIE). (2019). Chapter 3.8.1 African swine fever (infection with African swine fever virus). In *Manual of diagnostic tests and vaccines for terrestrial animals* (pp. 1-18).

```
#degree and betweenness calculated with igraph, where transfers_graph
# is an igraph object.
deg_in<-degree(transfers_graph,mode="in")
deg_out<-degree(transfers_graph,mode="out")
btwns<-betweenness(transfers_graph)


#startDate <- max(transfers$date)
#t.days<-as.integer(max(transfers$date) - min(transfers$date))
#PremList <- list of premises in the network
#n<- length(PremList)


#Ingoing and outgoing infection chain from Noremark & Widgren, 2014
OGCC<-OutgoingContactChain(transfers,PremList,startDate,t.days)
IGCC<- IngoingContactChain(transfers,PremList,startDate,t.days)


# mean infection potential based on Rossi et al., 2017:
rho_d<-function(transfers,d,N,gamma=21){
  #calculate infection potential for the ith farm on day d
  #-transfers is livestock movement dataframe
  #-N is the total number of premises in the network
  IIC_d<-IngoingContactChain(transfers,c(1:N),d,gamma)
  OIC_d<-OutgoingContactChain(transfers,c(1:N),d,gamma)


  rho_d<-IIC_d$ingoingContactChain*OIC_d$outgoingContactChain/(N-1)
}
```

```r
MIP<- function(transfers,m,first_d,N){
  #-transfers is livestock movement dataframe
  #-m is number of days of interest. (Used 284)
  #-first_d is the first day of interest.
  #-N is the total number of premises in the network
  v<-vector("list",m)
  for(i in 1:m){
    v[[i]]<-rho_d(transfers,first_d+(i-1),N=N)
  }
  return(v)
}


#purchase sales balance based on Koeppel et al., 2018:
PSB<-function(transfers){
  #-transfers is a dataframe of the movement data including:
  #  -destination: receiving premise ID
  #  -source:      sending premise ID
  #  -d:           date of shipment
  #  -n:           number of pigs in the shipment
  P<- transfers %>% group_by(destination) %>% summarise(pigs_in =
sum(n,na.rm=T))
  S<- transfers %>% group_by(source)      %>% summarise(pigs_out=
sum(n,na.rm=T))


  PSB<- (P - S)/(P + S)
}
```

In-Degree Distribution
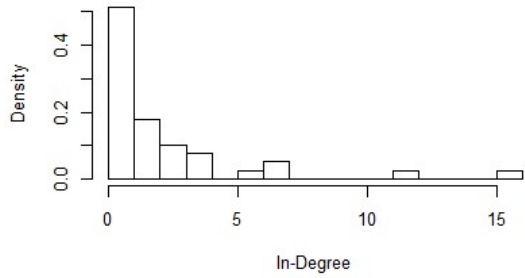
Out-Degree Distribution
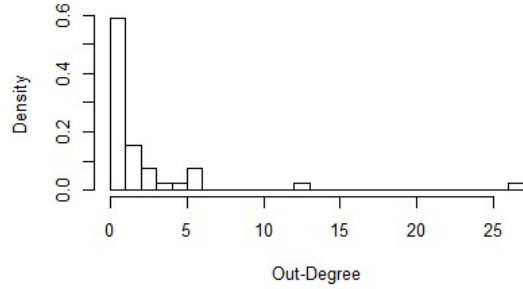
Ingoing Infection Chain Distribution
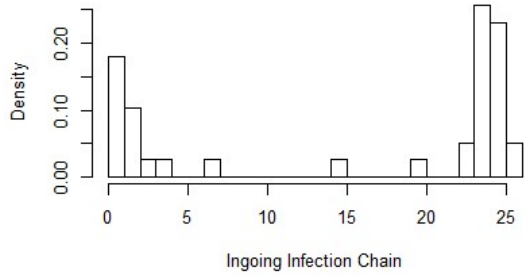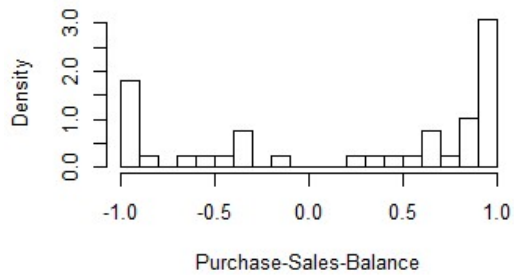
Outgoing Infection Chain Distribution
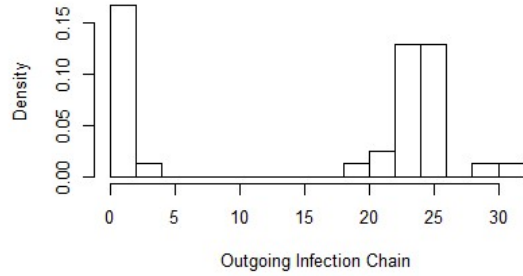
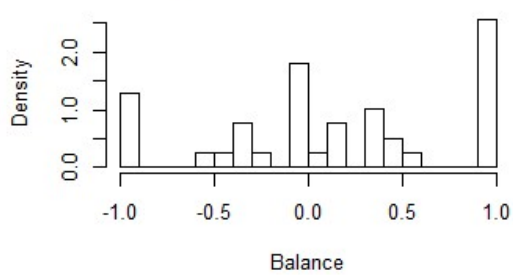Mean Infection Potential Distribution
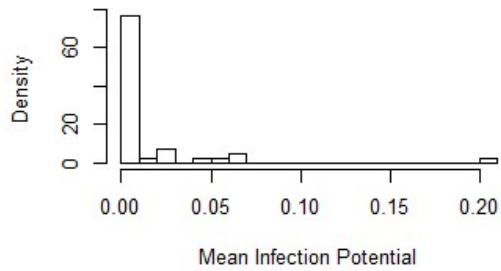
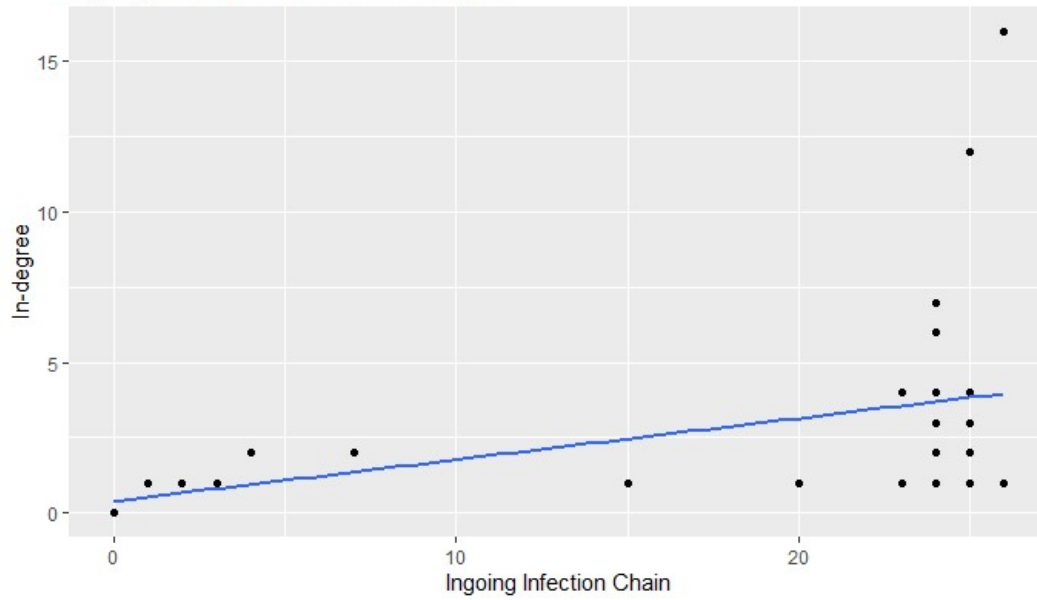Betweenness Distribution
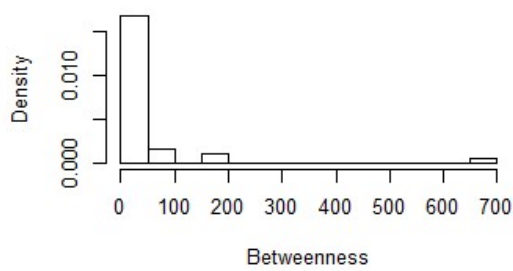
Purchase-Sales-Balance Distribution

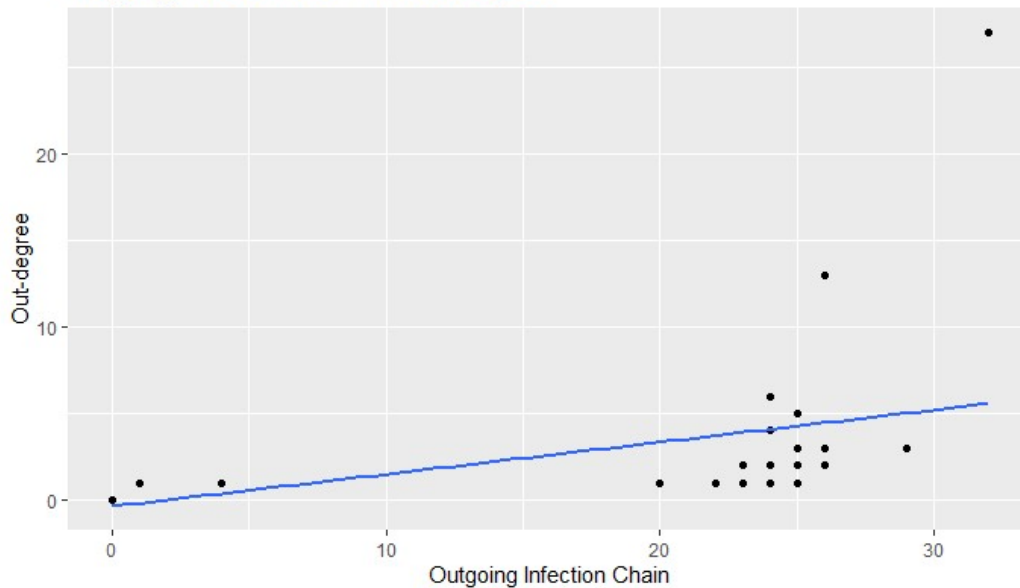Degree Balance Distribution

Ingoing infection chain vs. In-degree

```
        Pearson's product-moment correlation

t = 3.1976, df = 37, p-value = 0.002837
alternative hypothesis:          true correlation is not equal to 0
95 percent confidence interval:  0.1755717 0.6808684
sample estimates:                0.4653124
```
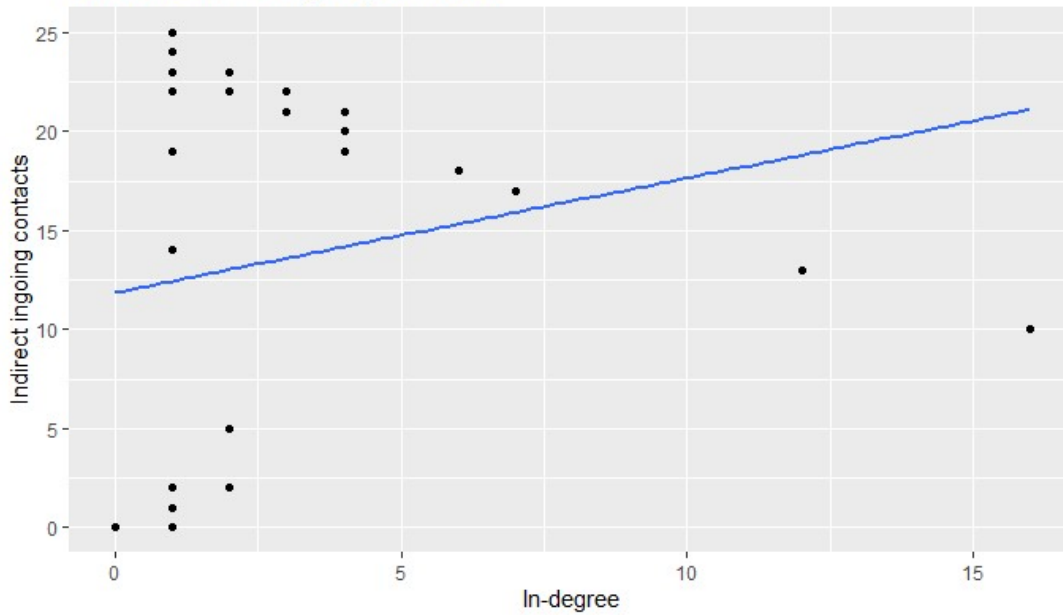

Outgoing infection chain vs. Out-degree

```
        Pearson's product-moment correlation

t = 3.1971, df = 37, p-value = 0.002841
alternative hypothesis:          true correlation is not equal to 0
95 percent confidence interval:  0.1754992 0.6808283
sample estimates:                0.4652538
```
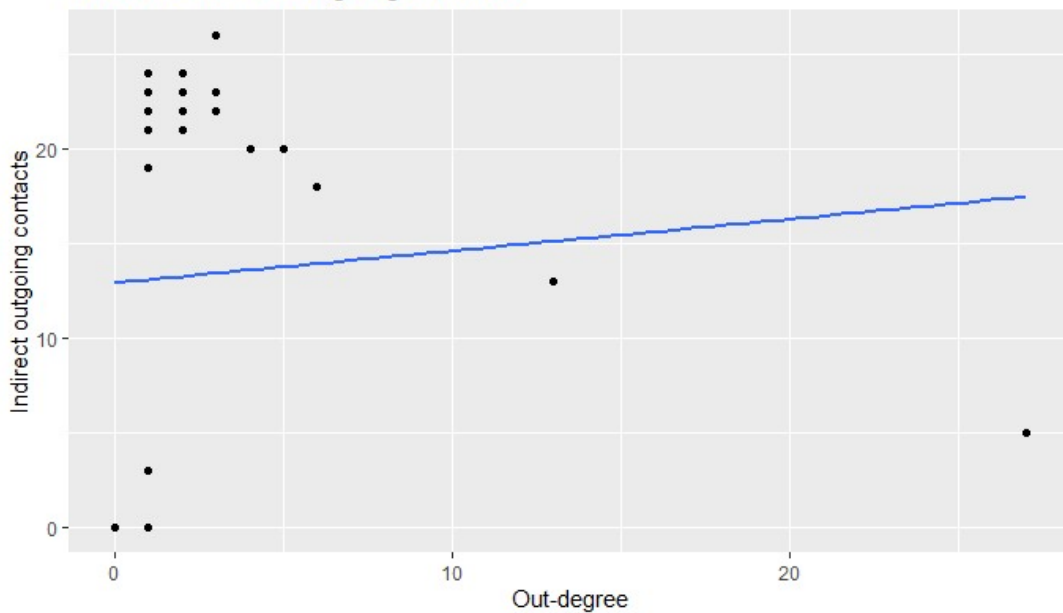
25

Direct vs Indirect Ingoing Contacts

Pearson's product-moment correlation

```
t = 1.1708, df = 37, p-value = 0.2492
alternative hypothesis:          true correlation is not equal to 0
95 percent confidence interval:  -0.1345253  0.4761360
sample estimates:                0.1890144
```


Direct vs Indirect Outgoing Contacts

Pearson's product-moment correlation

```
t = 0.45997, df = 37, p-value = 0.6482
alternative hypothesis:          true correlation is not equal to 0
95 percent confidence interval:  -0.2459659  0.3818357
sample estimates:                0.07540287
```