Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Jiaxiang Gai Date

A Simulation Study on a Fly-healthy Network

By

Jiaxiang Gai
MSPH

Rollins School of Public Health
Department of Biostatistics

_____

Thesis Advisor: Vicki S. Hertzberg

_____

Reader: George A Cotsonis

A Simulation Study on a Fly-healthy Network

By

Jiaxiang Gai

B.E., Wuhan University, 2012
MSPH, Emory University
Rollins School of Public Health
2014

Thesis Committee Chair: Vicki S. Hertzberg, PHD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics
2014

A Simulation Study on a Fly-healthy Network

By Jiaxiang Gai

Abstract:

There are a number of cases reported that infectious diseases are transmitted during commercial air travel. Individuals on a commercial flight are connected together to form large social networks. Study of their movement and contacts provides a better understanding of the spread of infectious disease. Such understanding, in turn, helps us to build effective disease control strategies.

We conducted a simulation study based on our observed network and we simulated the spread of an infectious disease. Each individual was set to be infectious one at a time, and we calculate the resulting incidence proportion during the flight.

We found that, in general, we would place the sick passenger in the window seats and in the front or the back on the plane if we want low incidence of transmission. Also, aisle seats are considered more risky than window and middle seats.

A Simulation Study on a Fly-healthy Network

By

Jiaxiang Gai

B.E., Wuhan University, 2012
MSPH, Emory University
Rollins School of Public Health
2014

Thesis Committee Chair: Vicki S. Hertzberg, PHD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics
2014

Acknowledgements

I want to thank the faculty, advisors, and staff of the Biostatistics Department at Rollins School of Public Health for the dynamic two years of learning that I have had. This thesis is only a sample of the vast knowledge that was attained and applied through my two years here at Rollins. I would especially like to thank Professor Vicki S. Hertzberg for all of his advice and support to help me write this thesis. Also I would like to give a special thanks to George A Cotsonis for taking the time to read my thesis.

Table of Contents

Introduction

According to the International Civil Aviation Organization (ICAO), around 3 billion passengers travel by air annually. [1] Passengers are in a cramped space sharing recycled air with hundreds of other people. Consequently, people on board are at risk of disease transmission during commercial air travel. Although many airports are equipped with thermal scanners that detect abnormal temperature, there are cases reported disease transmission during flights or after flights. On March 1979, a commercial aircraft with 5 attendants and 49 passengers was delayed on the ground for 3 hours because of engine failure. Most passengers and attendants stayed on the airplane during the delay. Within 72 hours, it was found that 72 percent of the attendants and passengers developed influenza-like illness. It was determined that one passenger, the index case, had been ill when he board the plane. [2] In 2002, one 3-hour flight carrying 120 passengers traveling from Hong Kong to Beijing began a spreading event; 22 of the 120 people contracted SARS during this trip.[3] The two cases above show just how air travel can play an important role in the rapid spread of newly emerging infections. The study of social contacts can help us gain a better understanding of the concomitant risk of disease transmission.

Individuals on a commercial flight are connected together to form large social networks. In practice each individual has a finite set of contacts to whom they can pass infection; the ensemble of all such contacts forms a “contact network”. [4] Studying and conceptualizing their movement and contacts provides a better understanding of factors affecting the spread of infectious agents transmitted during close or intimate personal contact. [5] Such understanding, in turn, helps us to build effective disease control strategies.

The objectives of this study is to use simulation techniques to study the spread of a disease within social networks on a commercial flight. Given the fact that one individual is sick, we want to find places on the flight that give lowest as well as the highest incidence proportion of the disease. We also want to decide the places that have lowest and highest transmission risk so that we can arrange the susceptible individuals, such as the elderly and the very young.

Background

Data Collection

Movement data were collected on 5 round trips: Atlanta to Portland, Atlanta to San Diego, Atlanta to Los Angeles, Atlanta to San Francisco and Atlanta to Seattle. Ten graduate students were trained to collect and process the movements and contact during the flight for both passengers and flight attendants. The iPad 2 with an IOS application based on Numbers was used for recording movement of passengers and flight attendants("Crew"). During the flight, observers were required to record flight information, movements and passenger description. The plane was divided into 5 zones and two observers were responsible for each zone. Data were collected and consolidated by observers immediately after the landing of each flight.

The movement data in this study are from the flight of Atlanta to Los Angeles. There were a total of 158 passengers and 5 flight attendants on the plane. Passengers each has a unique row number from 19 to 45. There are usually 6 seats on each row from A to F except for row

19, which has 2 seats(D and E). Seat letters A and F designate window seats; B and E designate middle seats; C and D designate aisle seats. The movement data were then used to determine which pairs of individuals were in a contact. If two individuals are within 1 meter of one another, we say that this two individuals have a contact. It is believed that those people would have a high risk of becoming infected via respiratory tract.

Programming Software

Gephi

Gephi is an open-source software for visualizing and analyzing large network graphs. It is very useful to explore, analyze, spatialize, filter, cluster, manipulate and export all types of graphs. [6] The node and edge spreadsheets of movement data were first imported into Gephi. They contain all the contact information from the startting time point to the end time point. Gephi identifies the two subjects with contact as “Source” and “Target” and then totals the total contact time for each pair of subjects. These data were then exported into R for further analysis.

R

R is a powerful statistical computation, programming and graphics language that is available at no cost. [7] Most of the figures and graphs in this paper are generated from R. A network package in R called “igraph” is used for visualization of the network. The igraph library can handle large graphs efficiently, and in the meantime it can be embedded into a higher level program or programming language. [8]

Network Data Structure

The structure of network data is somewhat different from that of conventional statistical data. The conventional statistical data consist of a rectangular array of measurements. The rows of the array are the cases, subjects, or observations. The columns consist of variables (quantitative or qualitative) indicating attributes or measures. [9] The network data, on the other hand, are square arrays. The rows of the array are the cases, subjects, or observations. The columns of the array are the same set of cases, subjects, or observations. In each cell of the array, the value of 0 indicates no contacts or relationships between the two subjects; values other than 0 indicate contact or relationship between two subjects.

Another way to indicate these relationships is to use an edge list. An edge list is formed by a two-column matrix, with each row defining one contact between two nodes. An edge is drawn from each element in the first column to the corresponding element in the second. [10] A weighted edge list is created on the basis of an edge list. A third column usually named “weight” indicates how close the relationship is or how many contacts the two subjects have and the duration of a contact.

Network with all edges and Network without tribe edges

The tribe edges consist of connections of pairs of individuals while all are seated. The anaylsis are performed both on a network with all edges and a network with no tribe edges. A network with all edges consist of the contact of both dynamic connections as well as static connections. A network with no tribe edge consist of the contact for only dynamic connections.

Methodology

Network Analysis and Visualization

An R package “igraph” was used to visualize the network structure of the social contacts on the flight. The contact network was represented graphically by letting each passenger or flight attendent be a node and each contact be an edge between two nodes. Degree of a individual is calculated as the number of contacts that he has during the flight. Subjects with a high degree were drawn at the center; subjects with a low degree were displayed around them.

Simulation Study on Disease Transmission

The transmission probability between each pair of individuals in this study follows an exponential distribution.

$$Transmission\ probability\ P_{obs} = \int_0^T \lambda\, e^{-\lambda x} dx \quad (x\text{: contact time})$$

The rate parameter λ is determined by solving the equation for a fixed T and P_{obs} . For a influenza transmission, with a cumulative contact time of 10 minutes, the probability of the disease transmission is 0.004. [11]

Thus we have :

$$0.004 = \int_0^{10} \lambda\, e^{-\lambda x} dx$$

The rate parameter is then determined from the above formula and thus we get:

$\hat{\lambda}$ = 0.0004008

In the simulation, we consider each individual that the person is in contact with. We obtain the transmission probability (P_{obs}) to that person based on the observed contact time for the pair. We also generate a random number U_{rand} from a uniform distribution between 0 and 1. If U_{rand} is smaller than P_{obs}, then that person will be "infected". We do the same comparison 1000 times and do it with all individuals on the flight. A pseudo code is shown below and an example output for the 287th simulation is shown in Table 1. "Infect" indicates the status of the transmission for the current simulation: 0 means not infected and 1 means infected. "Count" indicates the total number of a target individual getting infected up to the 287th simulation.

Pseudo Code:

```
For(i in unique(individual){
this.individual = unique(individual)[i]
for(j in 1:1000){
infect =0 ;
R_rand j = uniform(0,1) ;
P_obs j = CDF_exp(T_obs ; λ) ;
If R_rand j < P_obs j then infect = 1;
Count _j = Count _j +infect ;
}
}
```

Table 1: Example Output for the 287th Simulation:

Source	Target	Label	Time (min)	Prob	Urand	Count	Infect
29-D	28-E	28-E-29-D	188.42	7.27E-02	0.80539	17	FALSE
29-D	28-F	28-F-29-D	192.82	7.44E-02	0.04300	22	TRUE
29-D	28-C	28-C-29-D	187.33	7.23E-02	0.28733	23	FALSE
29-D	28-D	28-D-29-D	194.28	7.49E-02	0.08552	24	FALSE
29-D	28-A	28-A-29-D	1.00	4.01E-04	0.83117	0	FALSE
29-D	28-B	28-B-29-D	1.72	6.88E-04	0.79647	1	FALSE
29-D	29-E	29-D-29-E	191.02	7.37E-02	0.39345	15	FALSE
29-D	29-F	29-D-29-F	192.82	7.44E-02	0.93653	23	FALSE
29-D	30-C	29-D-30-C	191.13	7.37E-02	0.21761	23	FALSE
29-D	30-D	29-D-30-D	184.58	7.13E-02	0.98279	25	FALSE
29-D	30-E	29-D-30-E	190.60	7.35E-02	0.29547	21	FALSE
29-D	30-F	29-D-30-F	192.82	7.44E-02	0.06345	16	TRUE

To summarize our simulation, we include two strategies for analysis.

First, we summed up the count number over “source” (the infected) and then divided the total count by the total number of simulations. This gave us the average transmission number for each individual.

$$\text{Mean Number of Transmission}_i = \text{Sum}(\text{Count}_i)/1000$$

$$\text{Incidence Proportion}_i = \text{Mean Number of Transmission}_i / (\text{Total Individuals} - 1)$$

If we sort all individuals by their corresponding incidence proportion, we can get the top ten most infectious persons as well as the bottom ten least infectious persons.

Second, we summed up the count number over "target" (to be infected) and then divided the total count by the total number of simulation. This gave us the average number of times that the individual was infected. If we assume only one person is sick on the plane, we further divide that number by the (total passenger -1). It gave us the chance that individual would get infected assuming there is only one infectious passenger on the plane.

$$\text{Transmission Risk}_i = \text{Sum}(\text{Count}_i)/(1000^* (\text{Total Individuals} - 1))$$

If we sort all individuals by their tranmission risk, we can get the top ten most at risk persons as well as the bottom ten least at risk persons on the flight.

Both strategies were performed on the network with all edges and the network with no tribe edges.

Results

Network Analysis and Visualization

The visualization of the contact networks are shown in figure 1 to figure 4. The width of the edges indicate the contact time that a pair of individuals have. The contact networks for both flight attendants and passengers are shown in figure 1(network with all edges) and 2(network without tribe edges). Most flight attendants are placed at the center while passengers with few contacts are placed at the margin. The contact networks for only passengers are shown in figure 3(network with all edges) and 4(network without tribe edges). Passengers with a high degree are placed at the center; most of them are from seat

C and D. Passengers with a low degree are placed at the margin; most of them are from seat A and F.

Summary Statistics of the Network

For the network with all edges, the summary statistics of the social contact network can be found in Table 2. There are a total of 163 nodes and 4004 edges in our network. The average degree is 49.13 and the median degree is 49.

For the network data without tribe edges, the summary statistics can be found in Table 3. There are a total of 163 nodes and 3702 edges in our network. The average degree is 45.42 and the median degree is 48.

Table 2. Summary statistics of the network (all edge)

Number of Nodes	163
Number of Edges	4004
Average Degree	49.13
Median Degree	49

Table 3. Summary statistics of the network (no tribe edge)

Number of Nodes	163
Number of Edges	3702
Average Degree	45.42
Median Degree	48

Degree Distribution

As can be seen in figure 5 and figure 6, most of the subjects have a degree from 10 to 70, and very few subjects have a high degree (≥120). For those with a high degree, they are all flight attendants who need to serve meals and drinks several times during the flight, accounting for contact with nearly every passengers. The average degree is around 50, which means that, on average, each passenger will have contact with approximately 50 other people on a flight.

Simulation Results

The simulation results of the network with all edges are shown in table 4 to table 7. The top ten nodes that give highest incidence proportion are shown in table 4. They are all from aisle seats, with a resulting incidence proportion from 0.0053 to 0.0056. The bottom ten nodes that give lowest incidence proportion are shown in table 5. Only one of the top ten passengers is seated in the aisle seats and others are from middle seats and window seats. Most (7 out of 10) are from either the first section or the last section of the compartment. If we assume one person is sick on the plane, the mean incidence proportion is 0.0042. The top ten passengers with the highest risk of getting infected are shown in table 6. These people can be considered the most at risk of getting the disease assuming that one person on the plane is sick. Again they are all from aisle seat. Similarly, the least ten at risk people can be found from table 7. These ten seats can be considered the ten safest seats on the plane. They are either from the front or the back of the plane.

For the network without tribe edges, the simulation results are shown in table 8 to table 11. As can be seen in table 8, similar findings are found for the top ten nodes that result in highest incidence proportion. Most of them are from aisle seats. For the nodes that result in least incidence proportion, as can be seen in table 9, they are all from window seats. If we assume one person is sick on the plane, the mean incidence proportion is 0.0021. As can be found in table 10, the top ten most at risk persons are mostly from the aisle, similar to the findings in the network with all edges. Table 11 shows that the ten least at risk persons are all from the window seats(A and F). These seats are not necessarily from the front or the back; they can be found in the middle of the plane.

Table 4. Top ten nodes ordered by incidence proportion (all edges)

Rank	Node	Incidence Proportion
1	26-C	0.00564
2	27-D	0.00554
3	34-C	0.00550
4	36-C	0.00548
5	28-C	0.00546
6	21-D	0.00543
7	44-D	0.00542
8	36-D	0.00543
9	38-C	0.00542
10	37-C	0.00536

Table 5. Bottom ten nodes ordered by incidence proportion (all edges)

Rank	Node	Incidence Proportion
1	19-E	0.00213
2	45-F	0.00217
3	45-A	0.00218
4	45-E	0.00233
5	45-B	0.00244
6	20-B	0.00296
7	36-B	0.00316
8	37-E	0.00318
9	20-E	0.00326
10	45-D	0.00332

Table 6. Top ten nodes ordered by transmission risk (all edges)

Rank	Node	Transmission Risk
1	31-D	0.00576
2	22-D	0.00574
3	22-C	0.00570
4	37-D	0.00569
5	43-D	0.00561
6	27-C	0.00553
7	35-C	0.00553

8	42-D	0.00553
9	33-D	0.00550
10	34-C	0.00546

Table 7. Bottom ten node ordered by transmission risk (all edges)

Rank	Node	Transmission Risk
1	19-E	0.00149
2	45-F	0.00218
3	20-B	0.00226
4	45-A	0.00238
5	45-B	0.00241
6	45-E	0.00244
7	20-A	0.00253
8	19-D	0.00261
9	20-E	0.00315
10	20-C	0.00317

Table 8. Top ten node ordered by incidence proportion (no tribe edge)

Rank	Node	Incidence Proportion
1	36-D	0.00071
2	39-D	0.00061

3	23-C	0.00056
4	24-D	0.00047
5	20-B	0.00045
6	37-C	0.00044
7	22-C	0.00042
8	21-D	0.00041
9	38-D	0.00041
10	20-F	0.00040

Table 9. Bottom ten nodes ordered by incidence proportion (all edges)

Rank	Node	Incidence Proportion
1	24-A	0.00000
2	33-F	0.00000
3	34-A	0.00000
4	34-F	0.00000
5	23-F	0.00004
6	27-F	0.00004
7	29-F	0.00004
8	36-A	0.00004
9	40-A	0.00004
10	45-F	0.00004

Table 10. Top ten nodes ordered by transmission risk(all edges)

Rank	Node	Transmission Risk
1	36-D	0.000743
2	39-D	0.000643
3	23-C	0.000525
4	22-C	0.000512
5	24-D	0.000493
6	23-D	0.000468
7	38-C	0.000456
8	20-B	0.000443
9	20-F	0.000425
10	22-D	0.000418

Table 11. Bottom ten node ordered by transmission risk(all edges)

Rank	Node	Transmission Risk
1	21-F	0.00E+00
2	27-A	0.00E+00
3	41-A	0.00E+00
4	20-A	6.25E-06
5	24-F	6.25E-06
6	25-F	6.25E-06

7	27-F	6.25E-06
8	29-A	6.25E-06
9	32-F	6.25E-06
10	33-A	6.25E-06

Summary and Discussion

In the visualization results of contact network for both flight attendents and passengers, most flight attendants are placed at the center, because they generally have contact with all the passengers and flight attendants on the plane. Passengers with a low degree, such as 20-A and 42-F, are drawn on the edge of the figure. They have few contacts with people on the plane; in most cases, they remain seated during the flight so that they only have contact with the passengers sitting next to them. In the visualization results of contact network for only passengers, one can find that most passengers with a low degree are those in window seats(A and F). They may prefer staying in their seat because they need to climb over others to get out. Passengers with high contact are those in the aisle seats (C and D). This can be partly explained by the fact that the aisle seats give more freedom of movement.

Based on the simulation results, we can have the following strategy for placement. For network with all edges, if we have some sick passengers and we want to place them somewhere on the plane with low resulting incidence proportion, we should put them in window or middle seats, in the front or back of the plane. When we need to arrange seats for susceptible individuals so that they get less chance of being infected, we should put them in window seats, in the front or back of the plane. For network without tribe edges, if we have some sick passengers and we want to put them somewhere on the plane with low resulting incidence proportion, we should put them in the window seats. When we need to arrange a seat to people in poor health so that they get less chance of being infected, then again window seats should be used for these passengers. Generally, based on the simulation

results, aisle seats are more risky than window seats and middle seats. Seats in the front or back of the plane are safer than the seats in the middle rows.

The simulation performed now is a naïve way of analysis the disease transmission. It can be considered as a first stage of analysis. Because we are now only using the total contact time for each edge in the network. We are also ignoring the different contact types as well as the locations of contact. In future work, the total contact time can be split by different behaviors and different locations. Different weights can be adjusted to different behaviors, based on how intimate the contact is. For example, a talking behavior may be treated as a closer contact than a passing behavior.

Reference

[1] International Civil Aviation Organization. (2012). Annual Passenger Total Approaches 3 Billion According to ICAO 2012 Air Transport Results. Retrieved from: http://www.icao.int/Newsroom/Pages/annual-passenger-total-approaches-3-billion-according-to-ICAO-2012-air-transport-results.aspx

[2] Moser, M. R., Bender, T. R., Margolis, H. S., Noble, G. R., Kendal, A. P., & Ritter, D. G. (1979). An outbreak of influenza aboard a commercial airliner. *American journal of epidemiology*, *110*(1), 1-6.

[3] Mangili, A., & Gendreau, M. A. (2005). Transmission of infectious diseases during commercial air travel. *The Lancet*, *365*(9463), 989-996.

[4] Keeling, M. J., & Eames, K. T. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*, *2*(4), 295-307.

[5] Keeling, M. J., & Eames, K. T. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*, *2*(4), 295-307.

[6] Learn how to use Gephi, Retried from https://gephi.org/users/

[7] R Programming/Network Analysis. Retrieved from http://en.wikibooks.org/wiki/R_Programming/Network_Analysis

[8] Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*(5).

[9] Hanneman, R. A., & Riddle, M. (2005). Introduction to social network methods.

[10] Klovdahl A S, Potterat J J, Woodhouse D E, et al. Social networks and infectious disease: The Colorado Springs study[J]. Social science & medicine, 1994, 38(1): 79-88.

[11] Potter, G. E., Handcock, M. S., Longini Jr, I. M., & Halloran, M. E. (2012). Estimating within-school contact networks to understand influenza transmission. *The annals of applied statistics*, *6*(1), 1.

Appendix

Figure 1: Contact Networks for Flight Attendents and Passengers (all edges)

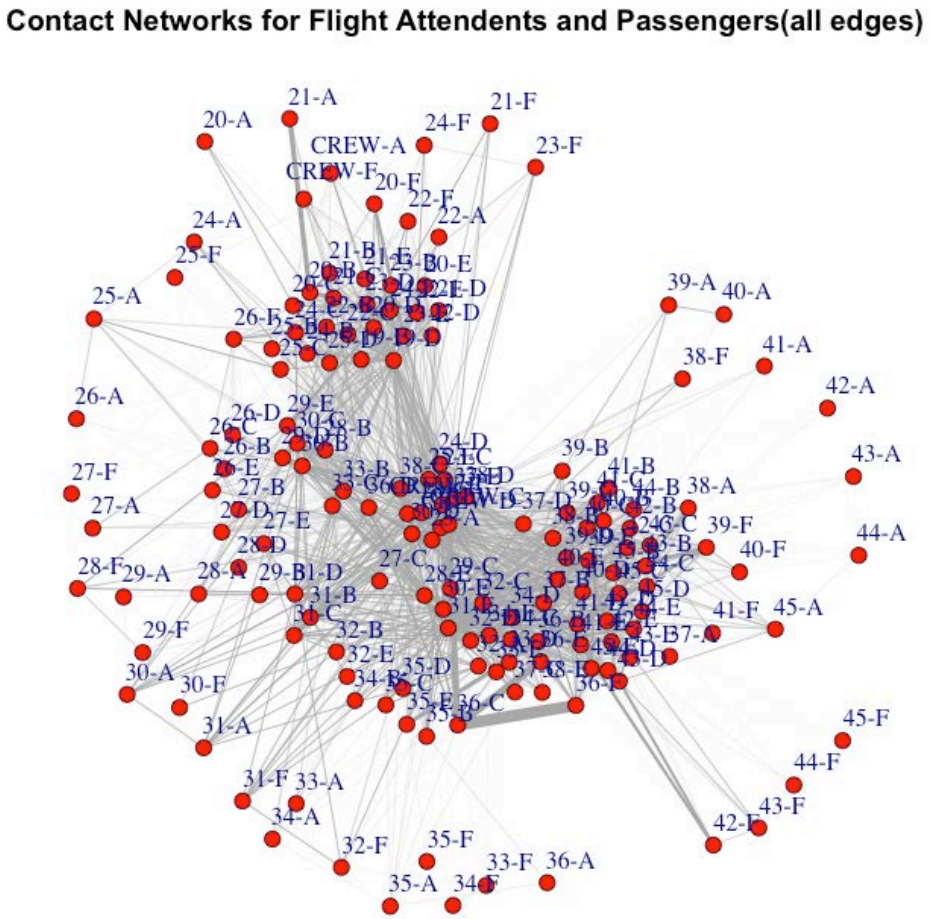

Figure 2: Contact Networks for Flight Attendents and Passengers (no tribe edges)

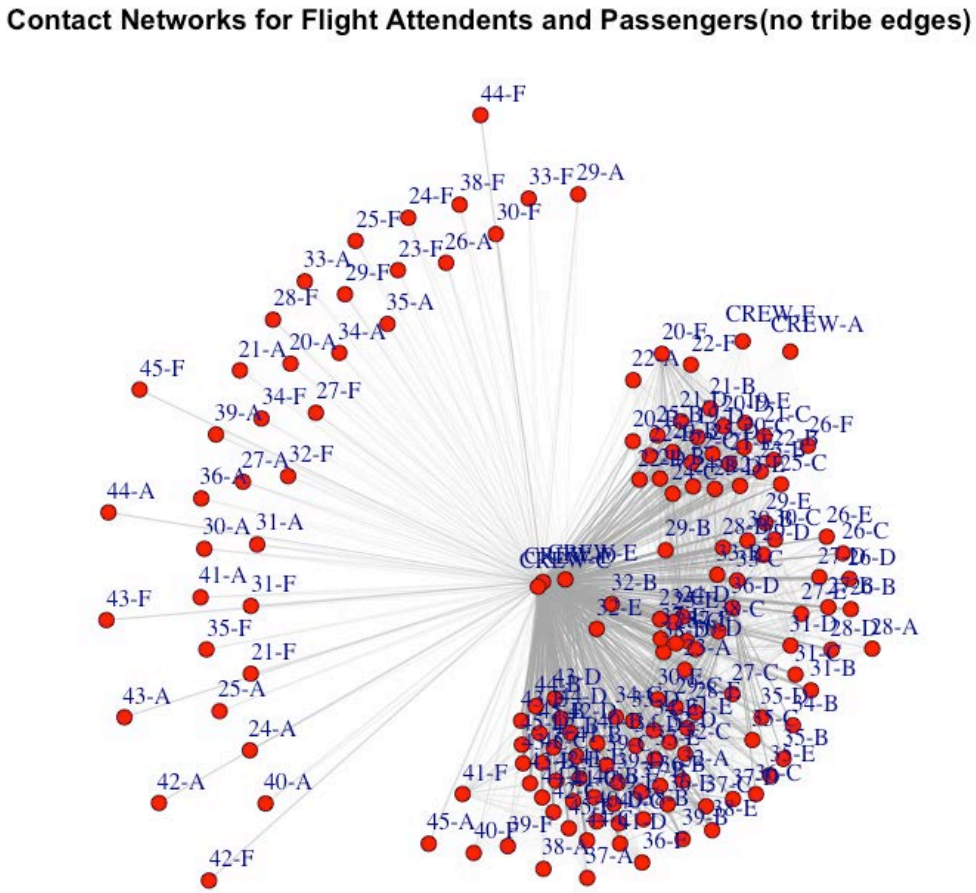

Figure 3: Contact Networks for Passengers (all edges)

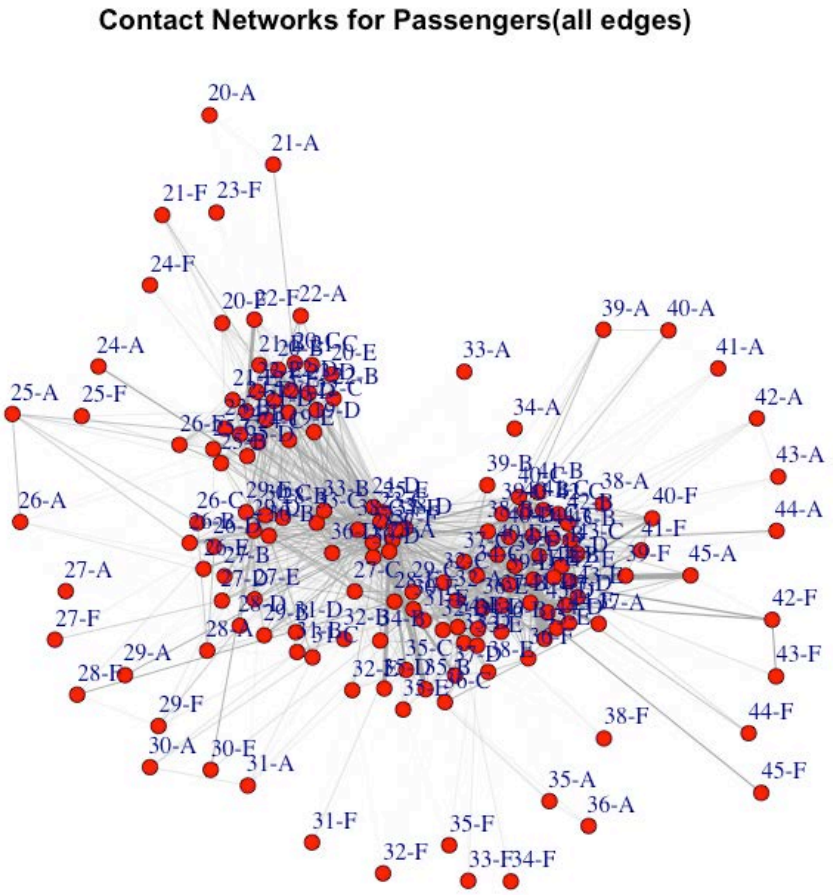

Figure 4: Contact Networks for Passengers (no tribe edges)

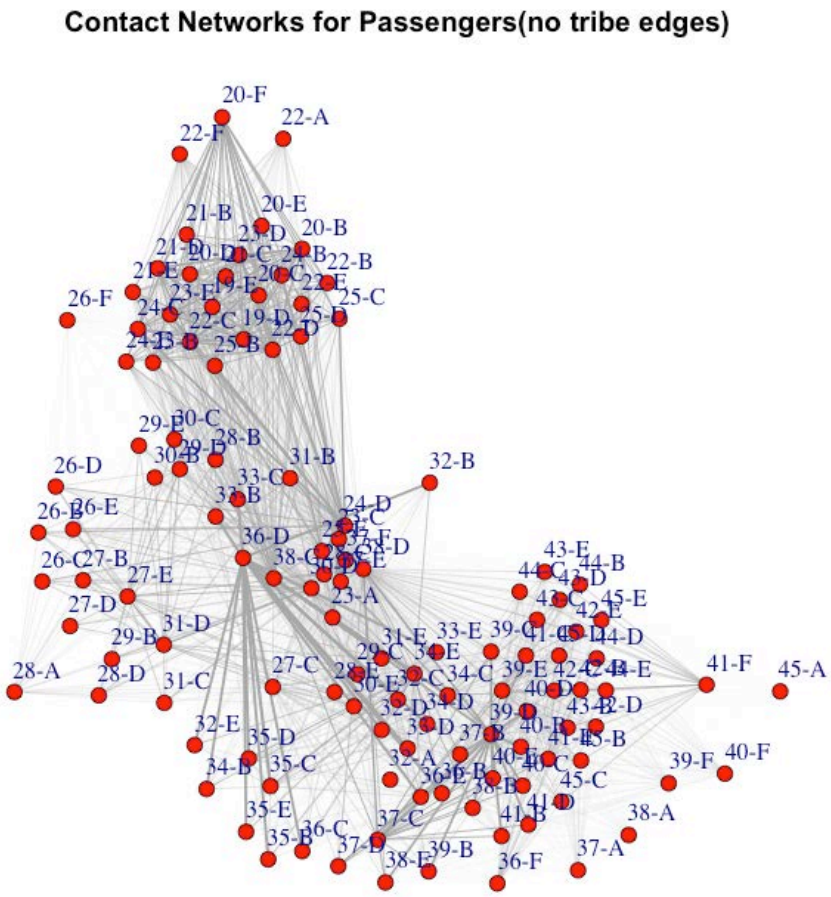

Figure 5: Histogram of Degree Distribution (all edges)

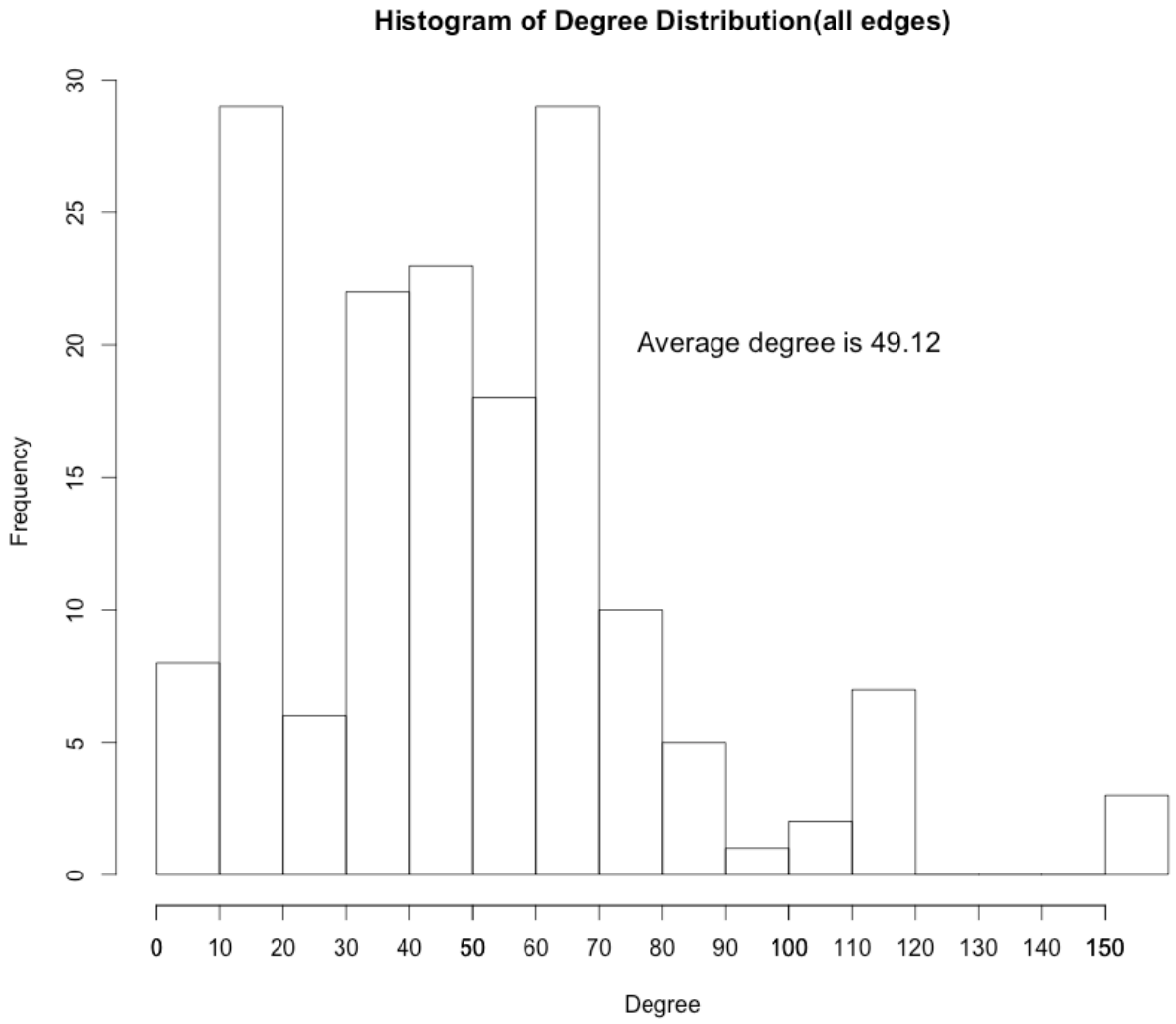

Figure 6: Histogram of Degree Distribution (no tribe edges)