Infectious Disease Spread

João Ribeiro (77209) | Ricardo Rei (78047) | Raquel Casteleiro (82027)

Complex Networks 2017/2018

**Introdution**

A complex network is a mathematical structure that models how a population of entities behaves with one another. This structure consists on a set of nodes, which represent the entities, and edges, which represent the relationships between nodes. Networks can be applied to almost anything in most study areas.

In this work, network science is used to model a human contact network related to the propagation of diseases in a population. It is based on a study done on Stanford University[1] (California, USA), where all credit for the previous analysis is due.

Infectious diseases are usually passed via droplets during the close proximity interactions, and thus the pandemic spread of an infectious disease poses a big threat to society in several ways.

This work aims to understand how a disease might spread across a population, so that us humans can understand how to fight it, by studying on a low-scale controled environment such as a highschool. Each individual was given a proximity sensor that registered it’s close proximity interactions. These devices have a radius of 3 meters and had a coverage of 94% of the total school’s interactions.

***“Schools are particularly vulnerable to infectious disease spread because of the high frequency of close proximity interactions”*[2]**

**The Dataset**

The dataset used consists of a single-day recording obtained from an American highschool. The network built is a weighted undirected network.

Nodes represent individuals (i.e., people) and edges represent close proximity interactions between said individuals.

There are 788 total nodes (655 students, 73 teachers, 55 staff, and 5 other persons). The edge representation is detailed in the next session.

The dataset can be found on the original’s study’s webpage[1]

**Edges – Contact Representation models**

Each edge represents a close proximity interactions between two individuals. An interaction between two individuals is defined by a continuous sequence of close proximity records, stored on the individual’s device. Therefore, there is a subtle issue of how to represent relevant details such as the duration and number of the interactions between two individuals.

The dataset already solved this issue for us by recording interactions in four different strategies:

1. Add-then-chop
2. Chop-Then-Add
3. Chop-then-count
4. Just-Chop

The first 3 make use of the “minimum duration” parameter that defines the minimum duration (in CPRs) for an interaction must be set and the last one makes use of the “drop-off” parameter that defines the minimum CPR gap to be filled (allows you to assume that the dataset might be missing CPRs).

The first strategy first adds all CPRs resisted by the interactions between two individuals to create the weight of the edge between the two, and then applies the minimum duration parameter, i.e. doesn’t consider edges with a weight less that the value of the minimum duration parameter. The second strategy first applies the minimum duration parameter to all interactions between two individuals and then adds the remaining interactions in order to create the weight between the edges of the 2. The third strategy returns the number of interactions ignoring the number of CPRs in each interaction. Finally, the last strategy uses the drop-off parameter in order to fill “gaps” between interactions, e.g. if drop-off = 1 and there is an interaction between A and B that registered 3 CPRs in time steps 1, 2 and 3 and there is another interaction between A and B that registered only one CPRs in time step 5 the gap between these 2 interactions will be filled in order to create a bigger interaction with 5 CPRs in time steps 1, 2, 3, 4 and 5.

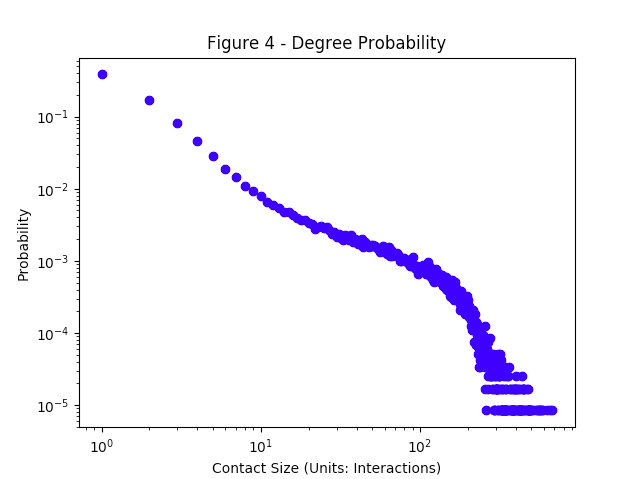
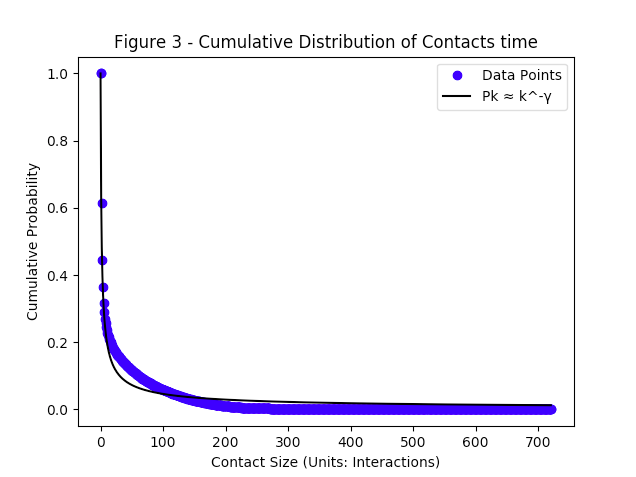
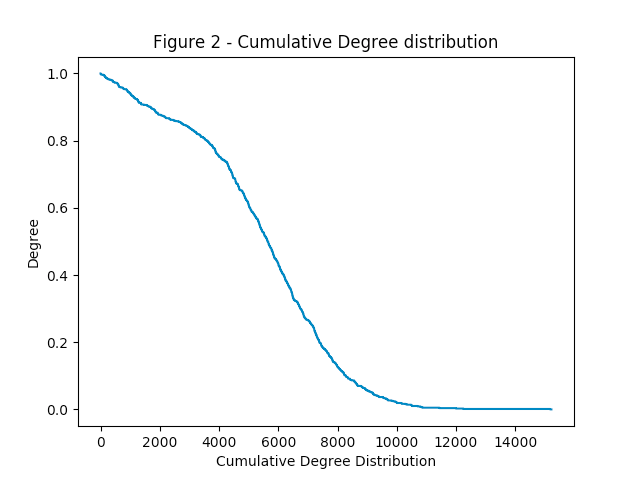
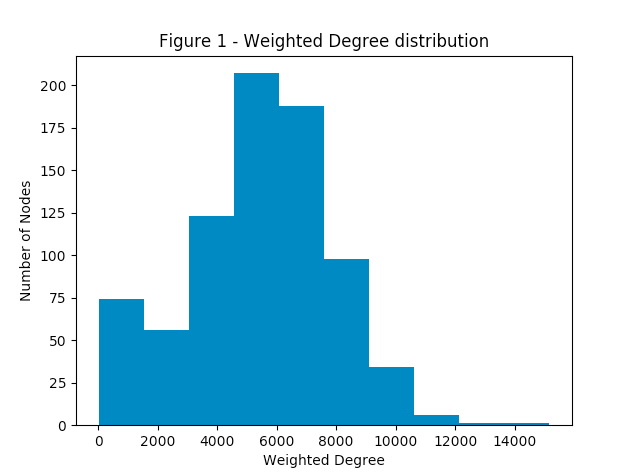
Colocamos algum exemplo?

The dataset edges lists all assume that minimum duration parameter value is 1 and the drop-off is 0.

For our analysis we used the Add-Then-Chop model because in our opinion is the simplest one to understand and represents well the difference between small contacts and big contacts

**Network Metrics**

[RR] Degree Study



The average degree of the network if we ignore the weights is approximately 300 which means that in a normal school day each person has in average 300 contacts. Note that not every contacts have the same duration and in terms of a infectious disease spreading if individual A is infected the longer the contact between another individual B the higher is the risk of B is infected.

Being that said its important to look for the degree metric considering weights in order to understand the size of the contacts. The weighted degree it’s based not only on the number of edges but pondered by the weights of each node. Figure 1 shows the histogram of the weighted degree and figure 2 shows the cumulative weighted degree distribution. We can see by these figures that the great majority of individuals have between 4000-8000 CPRs with an average of 5425 CPRs per day, recall that an CPRs is approximately 20s which is a lot of time in a 3m range from other individuals. Note that this value of a CPR considers overlapped time steps, e.g. if individual A stayed for 1 hour with 30 persons the degree of the node representing A will be something like 5425 ( (5425 \* 20s)/60s/60m ≈ 30h )

but in reality person A only stayed 1 hour in contact with other individuals.

Another interesting thing is to see the distribution of the contacts size or in another words the distributions of the weights in the graph. Figure 3 and 4 shows that the number of small contacts it’s a lot bigger than the number of longer contacts and inclusively the relation between the probability of having an contact with size k and the size of the contact itself is a power law with γ ≈ 0.667.

[RR] Average Path Length

The average path length is 1.6219 considering no weights. This means that in average the distance between 2 pairs of individuals is another person and “half”. Since there is no “half” persons the intuition is that person A had contact with half of the school directly (path length = 1), and to the other half, A had contact with someone that had contact (path length = 2) so in average we get ≈ 1.5 (0.5 \* 1 + 0.5\*2).

Considering weights, the average path length is 1.87 which is not much different from the average path length without weights because the great majority of contacts only have size 1 (as showed in figure 3).

[JR] Clustering Coefficient

Average Clustering Coefficient: 0.005622395491693588

Node with lowest Clustering Coefficient 374(Role Here) -> ( 0.0013795108060206942 )

Node with highest Clustering Coefficient 25(Role Here) -> ( 0.04722572403965249 )

[RC] Diameter

ToDo

**Node Metrics**

[RC] Degree Centrality

ToDo

[RC] Eigenvector Centrality

ToDo

[JR] Closeness Centrality

Average Closeness Centrality: 0.6207525375025098

Node with lowest Closeness Centrality 375(Role Here) -> ( 0.3705273069679849 )

Node with highest Closeness Centrality 170(Role Here) -> ( 0.7509541984732825 )

[JR] Betweeness Centrality

Average Betweeness Centrality: 0.0011027788727294

Node with lowest Betweeness Centrality 266(Role Here) -> ( 0.0 )

Node with highest Betweeness Centrality 15(Role Here) -> ( 0.006957799391153176 )

**Conclusion**

ToDo

**References**

[1] <http://sing.stanford.edu/flu>

[2] <http://sing.stanford.edu/pubs/PNAS-2010-1009094108.pdf>