Infectious Disease Spread

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

Complex Networks 2017/2018

**Introdution**

A Graph or Network consists in a mathematical structure used to model relations between pairs. This model consists in a set of nodes and each node can be connected with several others by links.

The interesting thing in this mathematical model is that he can be applied to a wide range of problem. In our project we use these model to represent a human contact network in a typical school day. But why is it interesting to understand the way we contact with each other in a school day?

Infectious Diseases are normally passed via droplets during close proximity interactions thus the pandemic spread of an infectious disease is a big threat to society and has a high economic impact. By modelling the contact network in a typical school day we can understand the dynamics of an infectious disease spread in a school.

“Schools are particularly vulnerable to infectious disease spread because of the high frequency of close proximity interactions” citação - <http://sing.stanford.edu/pubs/PNAS-2010-1009094108.pdf>

With all that said our goal for this project was to analyse the dataset flu-data from http://sing.stanford.edu/flu/ and demonstrate how complex networks theory can be applied to understand and possibly solve this problem.

**The Dataset**

In order to record the CPI’s a wireless sensor network (TelosB motes) was used by 789 individuals (655 students, 73 teachers, 55 staff and 5 other persons) representing 94% of the entire school population. Each sensor had a beaconing frequency of 0.05 s^1 (20 seconds) thus every time the beacon is activated the device records all close proximities (who is in a range of 3 meters from that individual). As an example, if student A and student B were talking within 3 meters of one another, for one minute 3 beacons would be activated which would result in 3 close proximity records (CPRs).

An interaction is a continuous sequence (>= 1) of close proximities between the same two individuals. Thus, a contact exists between 2 individuals if there is at least one interaction and the contact size is the sum of all interactions.

The entire dataset consists in 762,868 interactions and 118,291 contacts.

For the official project’s webpage, see <http://sing.stanford.edu/flu/>.

**Network Representation**

The aggregate network for the entire day can be represented by a weighted undirected network.

The nodes represent individuals wearing mote sensors (students, teachers, staff and other).

The edges represent the contacts between the individuals. The edge weight represents the duration of the contact. There are four different types of edge and weight representation, explained in the next session.

Nodes = 788 (655 students, 73 teachers, 55 staff, and 5 other persons)

Edges = 118,291 or 762,868 depending on the model used.

**edge/contact REPRESENTATION MODELS**

In this dataset there were different models to represent the edges.

Recall an interaction between 2 individuals is defined by a continuous sequence of CPRs, and a contact is the sum of all interactions. Having said this, after recording for each individual all the CPRs how do we build the edges? And what weight do we assign in order to differentiate contacts with bigger duration from the ones with small durations?

To do these 4 strategies are used in this dataset:

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

ToDo

[RR] Average Path Length

ToDo

[JR] Clustering Coefficient

ToDo

[RC] Diameter

ToDo

**Node Metrics**

[RC] Degree Centrality

ToDo

[RC] Eigenvector Centrality

ToDo

[JR] Closeness Centrality

ToDo

[JR] Betweeness Centrality

ToDo

**Conclusion**

ToDo