Modeling and Data Analysis in Complex Network Report

Tian Gan 5028531, Yanbin He 5037751

March 10, 2020

1 Introduction

In this report, we answered the questions under problem A, B and C.

2 Problem A

We will first list all the results in a table which is shown below. And all the graphs and analysis are shown as follow [1]. To give an overview of this temporal network, we illustrate this network shown in Fig 1 [1]. The upper part is a time slice of this temporal network and the other part shows the links and their appearing time.

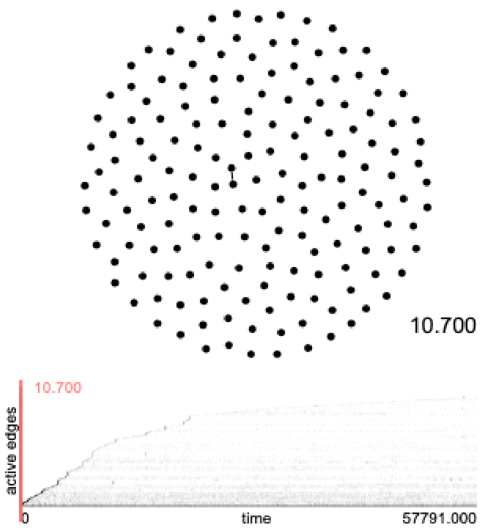

Figure 1: Network

Table 1: Results

N	167
L	3250
p	0.2344
$E[D]$	38.9222
$Var[D]$	999.8312
ρ_D	-0.2952
C	0.5413
$E[H]$	1.9674
H_{max}	5
λ_1	60.6393
μ_{N-1}	0.3811

[1]https://github.com/benmaier/tacoma

2) The degree distribution is shown in Fig 2. This network could be better modelled by scale-free network model since the degree distribution is a line in the log-log scale.

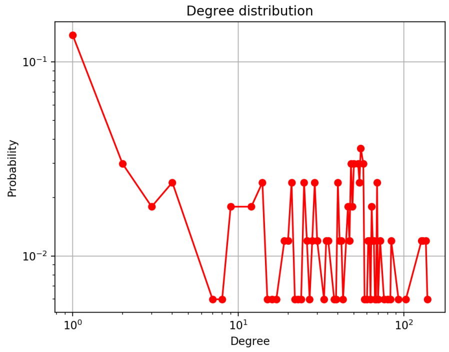

Figure 2: Degree Distribution

3) The value of assortativity ρ_D has been listed in the Table 1. Since the value $\rho_D < 0$, this network is degree disassortative which means the degree of each node of links is uncorrelated.

6) Small-world property? The values of $\frac{L(p)}{L(0)}$ and $\frac{C(p)}{C(0)}$ are 1.1108 and 2.4696 respectively. According to the Fig 3, we can find the difference between $\frac{C(p)}{C(0)}$ and $\frac{L(p)}{L(0)}$ is 1.3588 which satisfies the property of small world. Thus, this network is a small-world network.

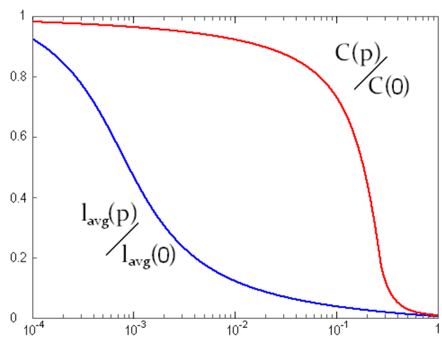

Figure 3: Small world property

3 Problem B

The answer of each problem is written below.

9) The plotting answer is shown in Fig 4. As we can see, the number of the infected nodes increases as

time changes. When starting from different nodes, the Susceptible-Infected processes at first are different, for the variance is quite large. As the process goes, the number of the infected nodes are somewhat synchronized.

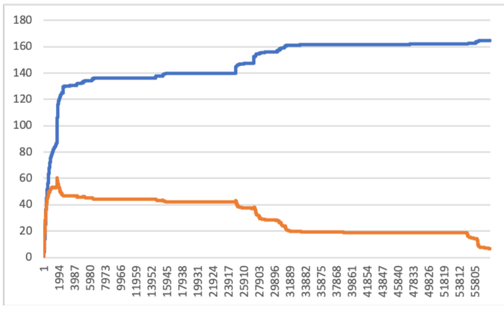

Figure 4: $E[I(t)]$: Blue and $\sqrt{Var[I(t)]}$: Red

10-11)In the Fig 5, we plotted the $r_{RD}(f)$ and $r_{RC}(f)$ as a function of f from 0.05 to 0.5. There is a thing that need to be clarified is that in our simulation, we choose the time required to reach 60% nodes as the measuring metric, for there is little difference among the results when choosing 80%. As we can see, the $r_{RD}(f)$ is always larger than $r_{RC}(f)$, meaning that the number of elements in the intersection of R and D is more than that in the intersection of R and C. Thus, in this case, degree can better describe how influential a node is than clustering coefficient does. This is because more neighbours a node has, more nodes it can affect. However, from the results we can see that when f is small, the intersection is also small, meaning that this type of prediction is quite rough. Although degree is better than clustering coefficient, both of them still perform not well.

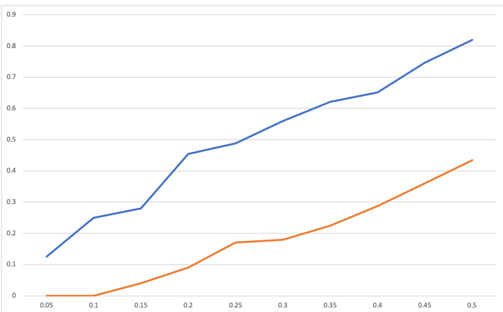

Figure 5: R-D: Blue and R-C:Red versus f

12)In this question, we choose the betweenness (B) from aggregated network features and temporal degree centrality (DC) from temporal network features [2], which is shown in Fig 6. As we can see, temporal degree centrality performs the worst in all the metrics, meaning that degree perhaps is not suitable for predicting how influential a node is.

13)In Fig 7, we illustrate the results of $r_{R'D}(f)$, $r_{R'C}(f)$ and $r_{R'R}(f)$ as function of f. It is obvious that

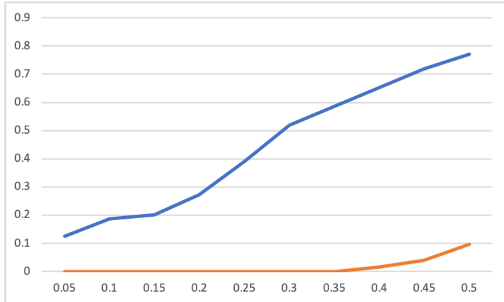

Figure 6: R-B: Blue and R-DC:Red versus f

$r_{R'R}(f)$ performs much better than the other two metrics. This is probably because when the information spreads faster, the average time of arriving infected nodes is lower. $r_{R'D}(f)$ and $r_{R'C}(f)$ are quite similar with the previous one, since R' and R have strong correlation.

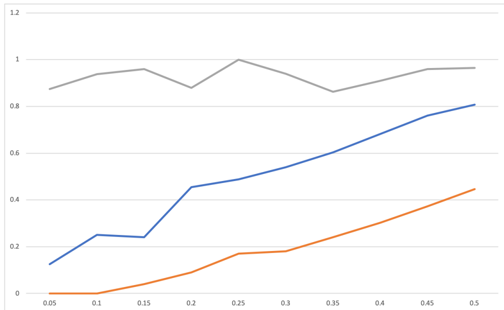

Figure 7: R'-R: Grey, R'-D: Blue and R'-C:Red versus f

4 Problem C

The answer of each problem is shown as follows.

14) The figures of probability density function of the inter-arrival time of two consecutive contacts between a node pair of three networks G_{data}, G_2 and G_3 are shown below.

In G_{data}, we can find there is a straight line in log-log scale, which means it follows power-law distribution. However, with the increase of the randomness, the inter-arrival distribution has different phenomenon.

In G_2, we reshuffle the timestamp of G_{data}, but it also has the same link set of G_{data}.

15) we simulate the spreading process on G_2 and G_3 and plot $E[I(t)]$ and $\sqrt{Var[I(t)]}$ of each network respectively. The figures of G_2 and G_3 are shown below and the figure of G_{data} is shown in previous section. As we can see, the information spreading processes go fast at first on all of these graphs. But for G_{data}, there is a time that this process goes very slow while the other two are not. Besides, the spreading process performs best on G_3 than on G_2 and G_{data}. This is possibly because the average contact inter-arrival time on G_3 is the smallest, as shown in the inter-arrival distribution. In that figure, more points with small inter-arrival time appears, which reduces the average time.

[2]https://github.com/wiheto/teneto

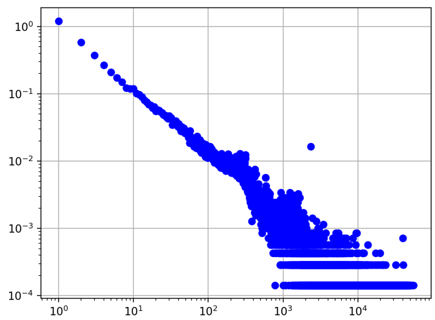

Figure 8: G_{data}

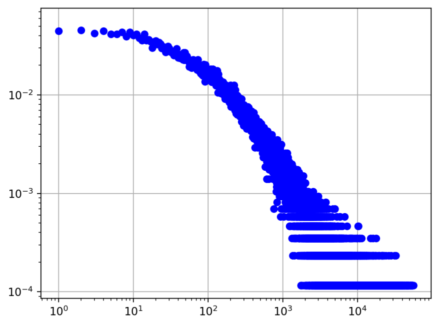

Figure 9: G_2

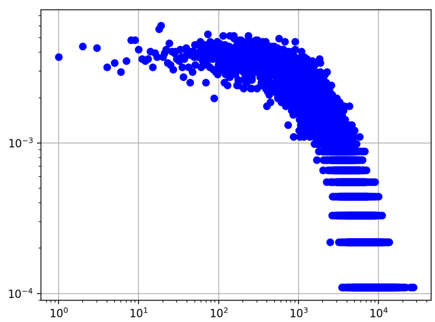

Figure 10: G_3

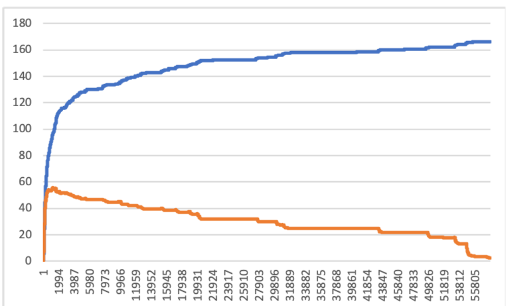

Figure 11: G_2

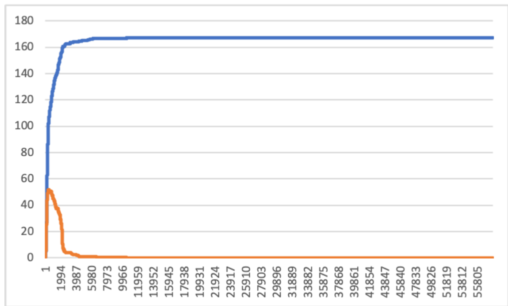

Figure 12: G_3

References

[1] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.