Assignment

Mingxi Li (), Xiaoyu Zhou (4646789)

March 2018

1 Introduction

Consider that a high school Face-to-Face temporal network as G_{data}, with the data are given. This network is sampled once every 20 seconds.

2 Results and Analysis

2.1 Part A: The topological features of the network G

The network properties for G are as follows:

All metrics computed for the network G are put into a table in the following figure 1.

N	L	p	$E[D]$	$Var[D]$	p_D	C	$E[H]$	$H_{\max}$	λ_1	μ_{N-1}
328	5818	0. 108	35. 476	185. 475	0. 033	0. 444	2. 159	4	41. 23	1. 93

Figure 1: The Metrics for Network G

Question 2

The degree distribution of network G are shown in Figure 2:

As shown in figure 1, k represents the number of degree of each node, and p(k) represents the percentage of number of nodes with degree k. It is obvious that the degree distribution is approximately a binomial distribution. While in a scale-free network, the degree distribution is not binomial, but has a power-law tail: Pr[d=k] $k^{-\gamma}$. Thus, ER random graph could better model this network.

Question 3

The degree correlation (Assortativity) $p_D \approx 0.033$.

And the physical meaning is that The assortativity represents to what extent nodes in a network associate with other nodes in the network. And it refers to the tendency of network nodes to connect preferentially to other nodes with either similar or opposite properties.

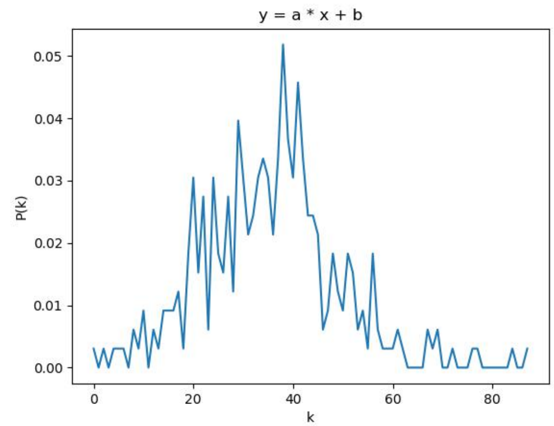

Figure 2: The Degree Distribution

Question 6

This network does not have the small-world property.

A small-world graph mainly refers to two properties: (1) the average short path E[H] is small, like in an ER random graph; and (2) the clustering coefficient C is much higher. While in this network, it is obvious that the average path E[H] ≈ 2.159. The clustering coefficient C ≈ 0.444. Therefore, this network does not have the properties of small-world graph.

2.2 Part B: Information spreading on a temporal network

Considering an information spreading process on the given temporal network G_{data} for N iteration. Each iteration starts with a different seed node infected at t = 0 and ends at t = T = 7375 the last time step that the network is measured.

Question 9

Taking all the N iterations into account, the figure 3 shows that the average number of infected nodes E[I(t)] together with it error (standard deviation $\sqrt{Var[I(t)]}$) as a function of time step t is in the following.

Question 10

With the vector R = [1,21,37,43,44,......,24]. The the most influential node that infects 80% nodes in the shortest time R_1 is node 1.

Question 11

The figure 4 shows that $r_{RD}(f)$ and $r_{RC}(f)$ are as a function of f, where f = 0.05, 0.1, 0.15, ..., 0.5.

In figure 4, the blue line represents $r_{RD}(f)$ and the red line represents $r_{RC}(f)$ in the following. And x axis represents f, y axis represents the recognition rate. It is obvious that the clustering coefficient could better predict the influence of the nodes because that the recognition rate $r_{RC}(f)$ has a higher value than the

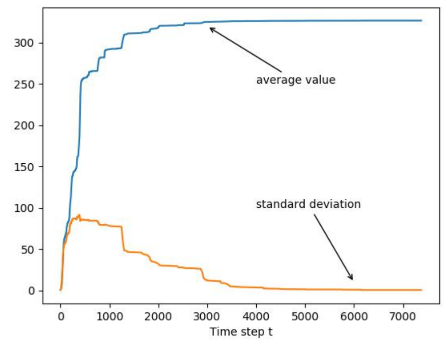

Figure 3: Functions of time steps

$r_{RD}(f)$ with the increase of f. And $r_{RC}(f)$ is approximately equal to the value of f, which indicates the clustering coefficient could better predict the influence of nodes with high accuracy.

Question 12

Another two nodal/centrality features proposed to predict nodes' influence are closeness centrality and betweenness centrality, which are abbreviated as B and C_c. The figure 5 shows that $r_{RB}(f)$ and $r_{RC_c}(f)$ are as a function of f, where f = 0.05, 0.1, 0.15, ..., 0.5.

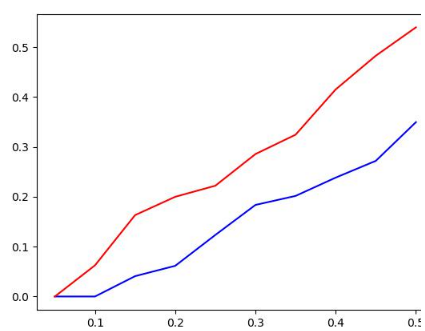

Figure 4: $r_{RD}(f)$ and $r_{RC}(f)$ are as a function of f

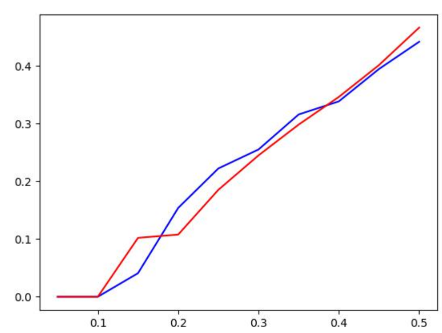

Figure 5: $r_{RB}(f)$ and $r_{RCc}(f)$ are as a function of f

In figure 5, the blue line represents $r_{RB}(f)$ and the red line represents $r_{RCc}(f)$ in the following. And x axis represents f, y axis represents the recognition rate. Thus, it indicates that both the closeness centrality and betweenness centrality could predict the influence of the nodes efficiently with similar recognition rate at same f. But both of them could not better than the clustering coefficient because the recognition rate of them are smaller than the clustering coefficient.

Question 13

Rather than ranking in the top f fraction, the aforementioned analysis method could be improved by ranking in nodal temporal features. In this way, the prediction results would much accurately.

2.3 Part C: Influence of temporal network features on information spreading

Construct three temporal networks: G_{data}, G_2, G_3

Question 14 & 15

the figure 6 & 7 show that the average number of infected nodes E[I(t)] standard deviation $\sqrt{Var[I(t)]}$) as a function of time step t for G_2 and G_3 respectively in the following.

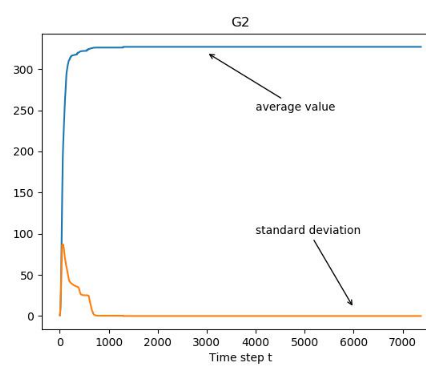

Figure 6: G_2 with R:(302, 42)

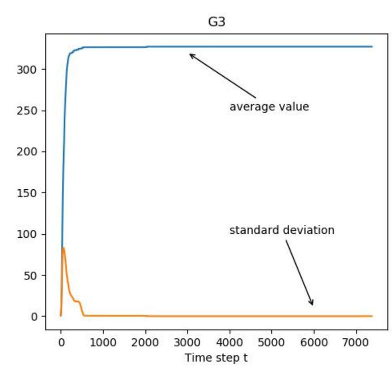

Figure 7: G_3 with R:(302, 43)

By comparing figure 3,6 &7, it shows the prevalence of three networks are finally same. What's more, the average number of infected nodes in G_3 could approach the peak quickly by approximately 550 time steps, while it needs about 1000 time steps in G_2 and around 3500 time steps in G_{data}. Therefore, it is obvious that the information could spread quickly in G_3 by spending the least time steps. Conversely, the information spreading slowly in G_{data} by costing the most time steps among these three networks. The reason is that the time stamps are randomly reassigned to the temporal links in G_2 & G_3 and random two nodes in G_3 can even contact more than once, which could improve the speed of information spreading. Thus, the information spreading performance of G_3 is the best.