

Project Report

Final Software Project Report for I590 Big Data Application & Analytics, 15 Fall

Student Name: Ji Ma

1. Purpose of the Project

Board interlock refers to the practice of members of a corporate board of directors serving on the boards of multiple corporations. It is a critical topic for understanding the relationships among corporates. By using methodologies from network analysis, this software project will analyze the network among Chinese foundations using Python and a comprehensive dataset of foundations' board members. Specifically, I take the following steps to examine the research question.

1.1 Construct the network by matching the board member.

For the organizational network of foundations as nodes, a link between two nodes (i.e., foundations) means they have one or more board members in common, and the link is weighted by the number of board members they share. The board member is matched according to the name, gender, and date of birth.

1.2 Visualize the overall network.

According to the registration policy, there are two types of Chinese foundations: public foundations (PUFs) which are allowed to fundraise from the general public, and private foundations (PRFs) which can only receive funds from private sources. In the visualization, PUFs are colorized as yellow, PRFs are colorized as blue.

1.3 Rich-club phenomenon in the networks.

The 'rich-club' phenomenon refers to the trend of nodes with high importance to form tightly interconnected communities, and it is one of the crucial properties accounting for the formation of dominant communities in both computer and social sciences. I will examine the presence of rich-club in the overall network, PUFs network, and PRFs network.

1.4 Which kind of foundations is controlling the network?

Scholars have agreed that the PUFs (public foundations) are controlled by the government, and through the PUFs, the government can control the network of the overall foundations. Therefore, I would like to examine whether PUFs hold more important positions than PRFs (private foundations) in the network.

"Important position" is operationalized by four types of centrality values, i.e., degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality. *Degree centrality* measures a given node's direct connection with other nodes. *Closeness centrality* measures the sum of geodesic distances from a given node to all the other nodes in the network. *Betweenness centrality* measures how often a given node falls along the shortest path between two other nodes. *Eigenvector centrality* computes the centrality for a node based on the centrality of its neighbors. In order to control the network, PUF should have higher centrality values than PRF according to statistical analysis.

1.5 The maximal cliques and PUF's percentage in each clique.

A maximal clique is a complete subgraph that cannot be extended by including one more adjacent vertex. Each node connects with all the other nodes in a maximal clique, indicating that "everyone knows everyone." It is the most cohesive community structure in the network. In order to control the cliques, PUF should be dominant in the cliques (i.e., percentage of PUF should be more than 50%). This percentage value will be examined by one-sample t test.

2. Dataset and Software

The dataset, provided by the Research Infrastructure of China Foundations (<http://ssrn.com/abstract=2673879>), is comprehensive and verifiable. It consists of 3,344 Chinese foundations' basic information (from *Basic* table) and 34,395 board member records from 2,847 foundations (from *Board* table).

Python NetworkX library and SPSS 22 are used for the network and statistical analysis respectively.

3. Results Obtained

3.1 Visualize the overall network.

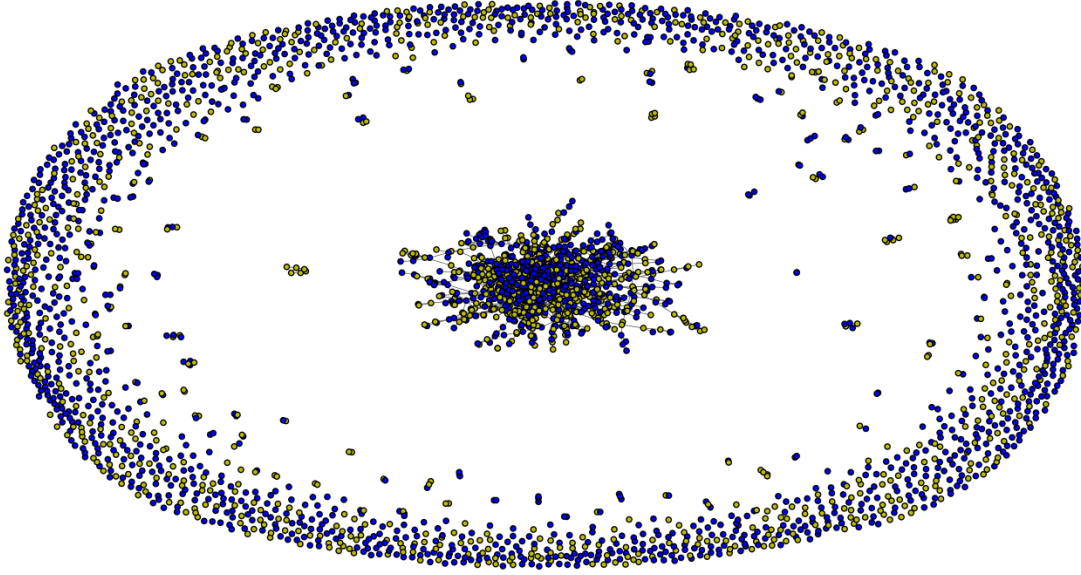


Figure 1: Visualization of Overall Network.

The overall network of Chinese foundations consists of 3,344 nodes and 1,880 edges. The nodes are colorized by their types, i.e., PUF as yellow and PRF as blue. The overall network is messy and has lots of isolated nodes which form a circle.

3.2 Rich-club phenomenon in the networks.

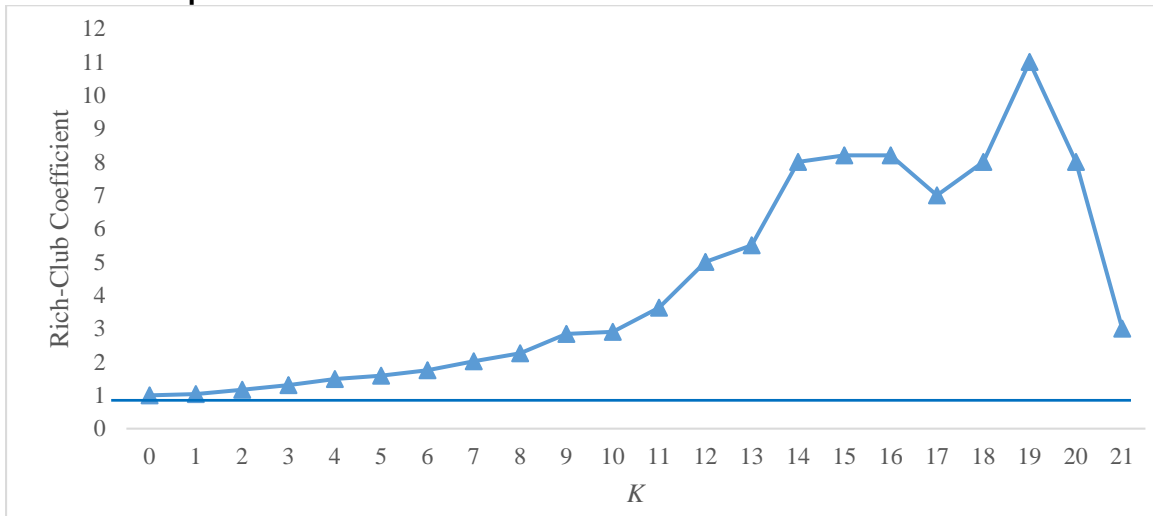


Figure 2: Rich-Club Coefficient in the Overall Network of Chinese Foundations.

Note: Coefficient is normalized; Times for random network experiment: 1,000.

The rich-club phenomenon is obvious in the overall network of Chinese foundations (characterized by the progressively increases of coefficient with the increasing degree k).

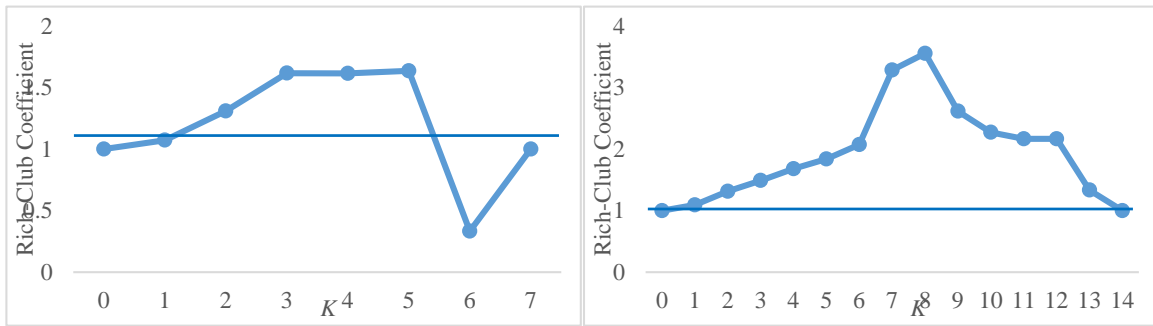


Figure 3: Rich-Club Coefficient in the Networks of Public Foundations (Left) and Private Foundations (Right). *Note:* Coefficient is normalized; Times for random network experiment: 1,000.

Rich Club is more salient in the network of PRFs than in the PUFs.

3.3 Which kind of foundations is controlling the network?

Table 1: Centralities by Foundation Type in the Overall Network.

Foundation	N	Degree		Closeness		Betweenness		Eigenvector	
		Mean(SD)*	p	Mean(SD)**	p	Mean(SD)*	p	Mean(SD)*	p
PUF	1,310	.41(.69)	.001	1.56(2.09)	.001	.23(.77)	.010	1.28(27.43)	.171
PRF	2,034	.29(.64)		1.15(1.93)		.17(.74)		.26(2.50)	

Note:

Bootstrap independent sample t-test, bootstrap sample = 1,000;
 PUF = Public Foundation, PRF = Private Foundation;
 * after multiply 1,000; ** after multiply 100.

Results of statistical analysis are showed in the Table 1. Positions held by PUFs in the network are more important than PRF when measured by degree centrality ($p = .001$), closeness centrality ($p = .001$), and betweenness centrality ($p = .010$).

3.4 The maximal cliques and PUF's percentage in each clique.

One thousand two hundred and twenty one (1,221) maximal cliques (clique size > 1) are found in the overall network. Statistical analysis shows the percentage of PUF in each clique does not significantly different from 50% ($p = .943$).

Results Remark

1. The overall network of Chinese foundations consists of 3,344 nodes and 1,880 edges. The overall network is messy and has lots of isolated nodes which form a circle.
2. The rich-club phenomenon is obvious in the overall network of Chinese foundations. It is more salient in the network of PRFs than in the PUFs.
3. PUFs hold more important positions in the network than the PRFs.
4. PUF does not dominate the maximal cliques.