A Data Study on Doctor Network

Yuting Lu

Department of Mathematics, Imperial College London

Introduction

In this project, a doctor network dataset is given, where nodes correspond to doctors having some attributes. The edges represent advice giving, discussions, friendships or random links between doctors. The aim of this project is to explore this social network of doctors and research their prescribing behavior.

Community Detection

To understand more about this system, we use community detection to organize vertices in clusters. Here the method of Newman's eigenvector is used. To compare the output communities with the doctor's city we use a confusion matrix, shown in the table below.

	Peoria	Bloomington	Quincy	Galesburg
C1	48	0	1	0
C2	1	48	0	1
C3	0	0	40	0
C4	68	0	1	0
C5	0	0	0	34

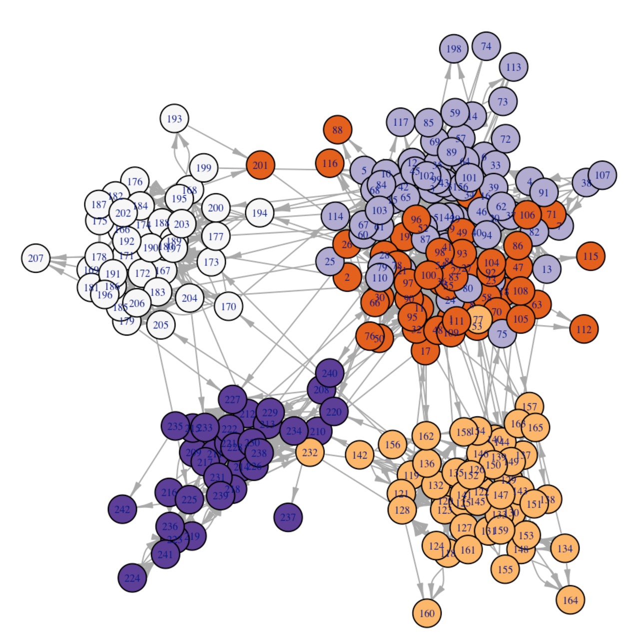

Figure 1: Results of community detection marked with vertex colors

From the plot it can be noticed that

- In the city of Peoria, there are two doctor communities.
- Doctors in the other cities they are usually part of the same community as other medics from their city
- The more central a node is in the community, the closer it is to its surrounding nodes.

Centrality

- Method:
  1) Create three lists with different centrality measurements (betweenness, eigenvector, and page rank).
  2) Arrange the lists in descending order of their respective measurement and extract the first 150 nodes in each list.
  3) Extract the intersection set of vertices and delete it.
- Results:
  1) 89 nodes were removed.
  2) Network now split into 1 big component with size 124 (dropped from 243) and 16 small size components.

Figure 2: Barplot of the size of the components

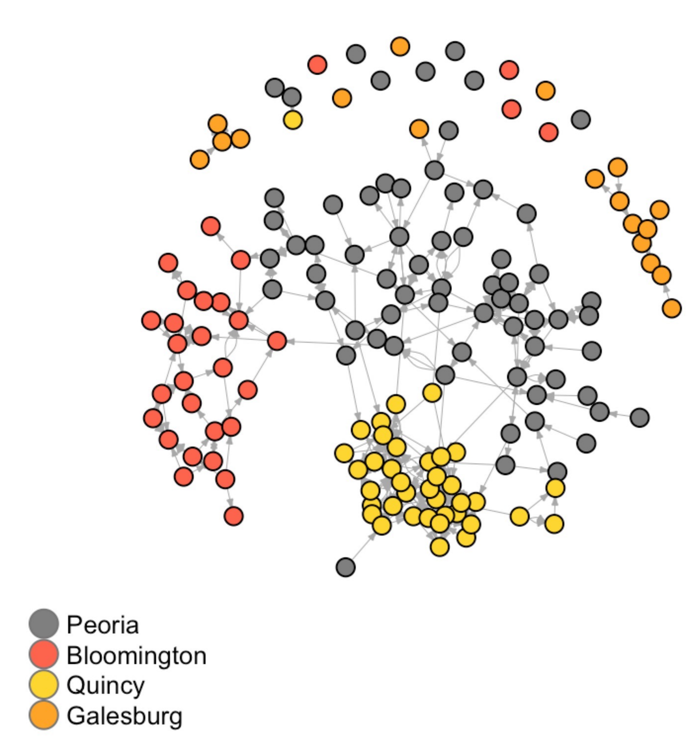

Figure 3: Plot of the network after some vertices were removed

Infection Time

- Looking for: types of nodes that tends to get exposed first and last.
  Process:
- Simulated the infection process 100 times with different starting points.
- Stored and added up each node's time to infection for each run of the simulation and averaged them.

Results:

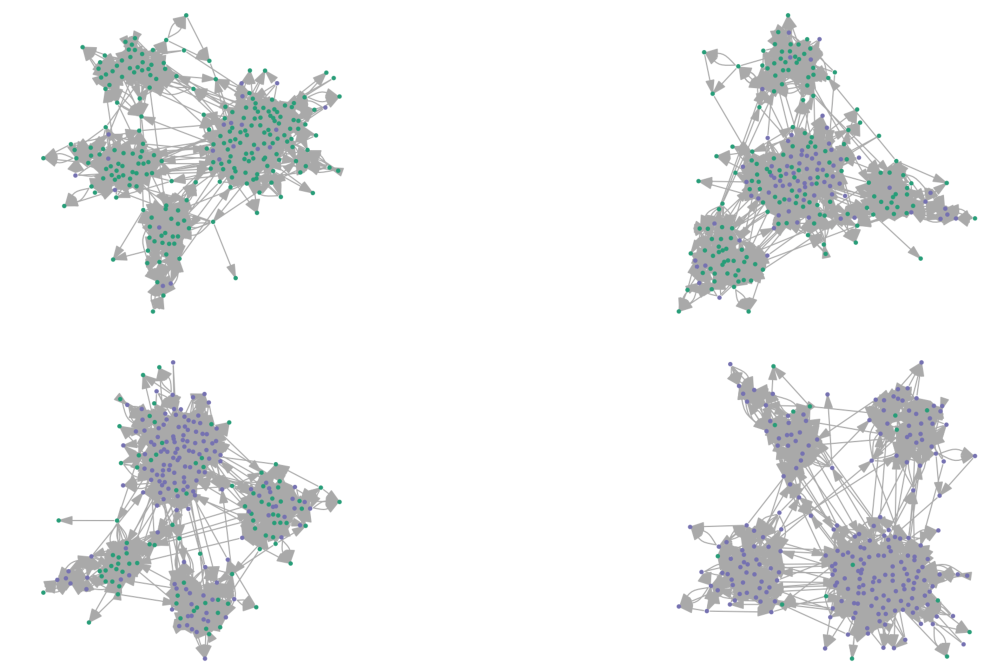

Figure 4: Simulation of how nodes get infected at times 1, 500, 1000, 2000

- For my analysis I compare the node's infection time with their degree.
- The first 10 nodes that get infected: 122, 141, 29, 36, 100, 46, 55, 212, 98, 11.
- The 20 nodes with the highest degree: 122, 29, 141, 36, 212, 100, 46, 189, 227, 213, 55, 31, 145, 228, 94, 180, 233, 238, 3, 11 (together with 23, 98, 174).
- Within the first 10 nodes, all of them have relatively high degree.
- The last 10 nodes that get infected (from the longest time): 237, 201, 193, 88, 242, 198, 112, 115, 74, 199.
- The 10 nodes with the lowest degree (from low to high): 237, 242, 198, 115, 112, 88, 74, 201, 199, 193.
- All of the last 10 nodes that get infected are included.
- Conclusion: The doctor corresponding to node 122 tend to get exposed to the new information first. On the other hand, nodes with low degree tend to get exposed last.

Doctor's prescribing behavior

- Why some doctors prescribed the drug early while others did not?
- Aspect 1: Study each node's individual attributes.
- Process: perform a linear regression with AdoptionDate set as the dependent variable.Then use summary table to see how the predictors are associated with the adoption date.

Result:

Table 1: Extract from the summary table with some significant predictors

	Coefficient	t-value
Degree	-0.08352	0.0468
City4	1.72278	0.0429
Meeting2	2.03096	0.0270
Clubs1	-2.36834	0.0145
Specialty4	10.43009	<0.0001

Interpretation:
1) Higher degree leads to earlier prescription.
2) Doctors in city 4 (Galesburg) tend to prescribe the drug later.
4) Attending specialty meetings leads to adopt it later.
6) Belonging to a club or having a hobby together with other doctors leads to earlier prescription.
7) Sharing building and office with other doctors leads to earlier prescription.
8) With the exception of GPs, internists and pediatricians, other categories of doctors tend to prescribe the drug later.

- Aspect 2: Study the doctor's social network.
- Process: Simulate the infection in different network and plot the number of infected nodes against time.

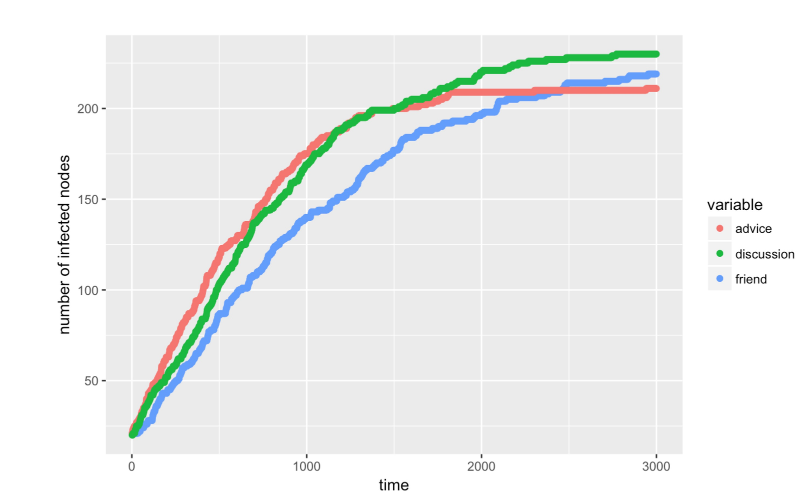

Figure 5: number of infected nodes vs. time for 3 networks

Conclusion

In conclusion, the data shows the doctor's individual attributes could affect their prescription behavior. Furthermore, there is a correlation between how doctors socialize with other doctors and their knowledge of the new drug.

The contents of this work and the associated code are my own unless otherwise stated.