Angela Guevarra, Michelle Wan, Tara Chacko, Jessica Fu
ISOM 673: Social Network Analytics
Final Project

Six Degrees of Francis Bacon

Introduction

Exploring the Six Degrees of Francis Bacon datasets, we were interested in how the network of early modern figures in England evolved over time. Looking into the broader network scope and then narrowing our focus to relationships among individuals and professions, we broke this topic down into six questions. First, we set up the network to determine whether the shape had a core-periphery structure. Next, we looked into the most central individuals and their occupations. Likewise, we were also interested in confirming whether Francis Bacon was the most central figure and if everyone in the network was in fact within six degrees of him. Other questions we raised include exploring which professions showed up the most out of the high-confidence ties, what kind of profession are two people most likely to have out of the low-confidence ties, what attributes influenced a person's centrality and what are the relationships between different groups in the network. Answering these questions provides a holistic analysis into the social network of early modern England.

Background on the Data

We used three data files to complete the project. The first provides information about every person in the network, stating a unique person ID, their gender, their historical significance, and their birth year. We believe their historical significance is the profession that each person was most known for. The second provides information on different groups. It contains a list of members and the start and end year of each group. The third provides information about different relationships in the network. Each observation lists the two people in the relationship, the confidence that the relationship existed and the start and end year of the relationship.

Question 1

We were interested in determining which individuals were the most influential; based on these individuals, we also examined patterns in their professions. We are essentially calculating eigencentrality.

First, we took the relationship ties between two individuals to make an edge-list. Next, we transformed the data table into a graph object in order to calculate eigencentrality. Once we calculated the eigencentrality and sorted the table in descending order, we merged the information with the individuals' names and professions.

Name	**Profession**	**Eigencentrality**
King Charles II	King	1.00000
King Charles I	King	0.9356996
King James I and VI	King	0.7709230
King James II and VII	King	0.5422066
Francis Bacon	Lord Chancellor, politician, philosopher	0.5265150
Elizabeth Tudor	Queen	0.4424713
Oliver Cromwell	Lord Protector	0.4265634
King William III and II	King	0.4071563

Based on the results, it's clear that individuals whose professions were related to royalty were the most influential. It's interesting to note that although Francis Bacon wasn't the most influential, he is one the only figures in the top list that isn't a king or queen. In fact, kings make up five out of the eight figures with the highest eigencentrality.

Question 2

As the premise of the network is that most people were within six degrees of Francis Bacon, we wanted to find out how many people were exceptions to this and how Francis Bacon's centrality compared to other central figures in the network. To do this, we created an adjacency matrix of all relationships in the network and then checked which column in the matrix represented Francis Bacon. We then took the first to sixth power of the matrix and for each power, we added the column representing Francis Bacon to a data table. The data table thus showed the number of walks of length 1 to 6 between

Francis Bacon and each node in the network. We then checked for rows that only had 0s to find the nodes that were not within six degrees of Francis Bacon. By doing this, we found that there are 66 people in the network who are not within six degrees of Francis Bacon. The most central person in the network by all measures (degree, closeness, betweenness and PageRank) was King Charles II. However, Francis Bacon still ranked quite highly in all the centrality measures. The values for these measures as well as his ranking for each is shown in the table below.

Centrality measure	**Francis Bacon Value**	**Francis Bacon Ranking**
Degree	662	15
Closeness	1.12e-06	6
Betweenness	1,657,620	10
PageRank	0.00156	16

Question 3

We wanted to analyze whether a person's profession has a significant correlation with how strong or confident the creators of this dataset were that the relationship given existed. Without any testing, we assumed that the more prominent and highly relational professions would have higher confidence in the classifying for their relationships than those with professions that were more individual and less public interaction.

To test our hypothesis, we first had to classify ties as being strong, regular or weak ties based on the confidence level given that the tie existed. We plotted a histogram of the confidence levels to decide the threshold values and after looking at the score distribution, we decided that ties given a confidence level of 60% or greater would be classified as strong ties and any 15% or less would be classified as weak ties. Those in between are just considered to be regular ties. From there, we subsetted the relationships for those that were strong ties and then merged the people ids from the subsetted data to the people dataset to

connect the person with their profession. Once we had this merged dataset, we were able to count the amount of people for each type of profession with basic data table aggregation. Then, we calculated the probability for each type of profession by taking the total count for each profession divided by the total number of people that have strong ties. We then repeated this same process for the weak confident ties. Our results are displayed in the tables below:

Top professions of high confident ties

Final <chr>	count <int>	prob <dbl>
politician	356	0.05757723
Church of England clergyman	318	0.05143134
courtier	169	0.02733301
poet	168	0.02717128
army officer	148	0.02393660
physician	144	0.02328967
judge	130	0.02102539
lawyer	115	0.01859939

8 rows

Top professions of low confident ties

Final <chr>	count <int>	prob <dbl>
Church of England clergyman	427	0.06269270
politician	413	0.06063720
poet	203	0.02980473
judge	164	0.02407870
courtier	161	0.02363823
physician	157	0.02305095
army officer	155	0.02275730
merchant	125	0.01835266

8 rows

We were surprised that most of the professions were the same for both the high and low confident relationships. We can assume that the profession did not have much correlation with the level of confidence the relationship ties were given. The only difference between the two groups was that lawyers were in the top for strong ties and merchants for weak ties. This makes sense to us because lawyers would be in more formally documented relationships where you can be very sure of the existence of the relationship even years later. On the other hand, merchants have lots of interactions which makes sense for how they were able to capture the relationships for this dataset but they would be found in more informal, casual documents where you might not be as sure of the validity when trying to collect this data years later.

Question 4

We wanted to know if attributes like a person's gender, whether they are royalty or not, and their profession could be good predictors of a person's centrality level within the network. The first step we had to do was to get each person's centrality score which we calculated in question 2. This will be the

variable we predict within our model. Then, we merged the centrality scores with the people's information dataset to have the gender and profession of each person. With this dataset, we then calculated the predictor variables in a numerical format we can run in the model. By using a simple ifelse function, we created a binary dummy variable for gender. We then created a new dummy variable for if a person was royalty or not. We classified this by checking through regular expressions whether the person's profession had king, queen, prince, princess, duke, or duchess in it. Lastly, we turned the profession column into a categorical variable using the factor function so every unique profession would be represented by a corresponding number.

The results of our linear regression model

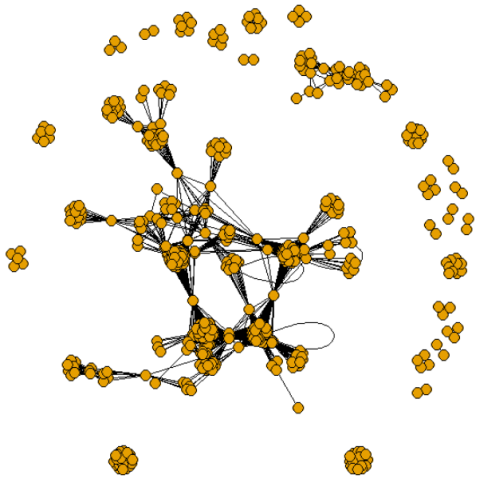
```
Call:
lm(formula = V1.x ~ royalty + GenderBinary + royalty + profession,
    data = eigenNames)

Residuals:
     Min       1Q   Median       3Q      Max
-0.10706 -0.01461 -0.00671  0.00667  0.89345

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.932e-03  9.482e-04  10.474   <2e-16 ***
royalty      8.711e-02  3.336e-03  26.115   <2e-16 ***
GenderBinary 8.721e-03  9.083e-04   9.601   <2e-16 ***
profession   8.927e-07  4.587e-07   1.946   0.0516 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02701 on 13031 degrees of freedom
Multiple R-squared:  0.05318,   Adjusted R-squared:  0.05297
F-statistic:   244 on 3 and 13031 DF,  p-value: < 2.2e-16
```

As you can see from the model results, while the predictors are all statistically significant, none of them are very strong predictors of centrality within this network, especially with an r-squared score of 5%. Royalty out of the three is the strongest predictor with the highest coefficient of .087. It makes sense that royalty is strongest since being royalty completely changes the types of relationships and the quantity of ties a person has.

Question 5

Lastly, we wanted to analyze what the network looked like between people who were part of groups during that time in England. In order to do this, we took the groups dataset and split the members column

up so that there was only one member in each column. The motivation in doing this is so we can create a list of all the different relationships that consist of a pair combination of people in the same group. In R, this is done by using the combn() function. In order to do this, we had to remove groups that had only one person in them because relationships cannot be formed with less than two people. After creating that edge list of relationships between people that belong in the same group, we created an igraph object that allowed us to visualize that network that resulted in the following:



As can be seen in the network figure above, the network formed amongst people that belonged in a group also formed a core periphery network. However, in the periphery, we see small clusters that have formed groups of their own as well but do not have that bridge to tie it to the main core of the network. Therefore, we decided to take a deeper dive into the differences between the groups that are in the core and those that are not. First, we wanted to determine what proportion of the nodes/people are in the core versus who aren't. We did this by taking the eigen centrality scores of each person to be able to retrieve the concentration correlations to determine the partition between the core and the periphery. Then we took the maximum correlation to get the number of nodes that are in the periphery and divided it by the length of all correlations to get the proportion. From these calculations, we found there to be 428 nodes in the core which is about 83% of all the nodes. We then took a closer look at those people in the periphery and found that the people in the periphery were those that belonged in groups that had a small number of members

(2-4 members). However, there were a few outliers where big groups never made ties that connected them to the core. These groups were: Sidney Circle, The King's Musik, Midwives, and Stationer's Company.

Conclusion

After conducting our analysis, we drew many insights and conclusions regarding the shape of the network of early modern figures in England as well as how that network changed over time and which figures were influential within that network. Because of its core-periphery shape, it makes sense that most people within this network are within six degrees of Francis Bacon and it is rare to find the people that sit in that periphery given how many people consist of the network. However, it misleads the audience into thinking that Francis Bacon is the most central person in this network, when he is in fact not. We see a pattern that the most central people are royalty which indeed makes a lot of sense. This led us to determine what factors, other than being royalty, contribute to one's influence in the network and found that none are as strong as the person being of royalty.

The extent of our analysis leads us to wanting more data and characteristics/traits of people in this network in order to dive deeper into further analysis. For example, having some way of sorting each person's relationship types with each other using the given relationship type data table. Additionally, it would be interesting to extend our analysis of how the network changed over time and relate it to a historical event that happened in that year in the same region of England.