Name: James McAuliffe

Professor Dennis Feehan
November 20, 2024

University of California, Berkeley
Department of Demography

Demography 180: Social Networks
Mini Project Results

PART A (10 points)

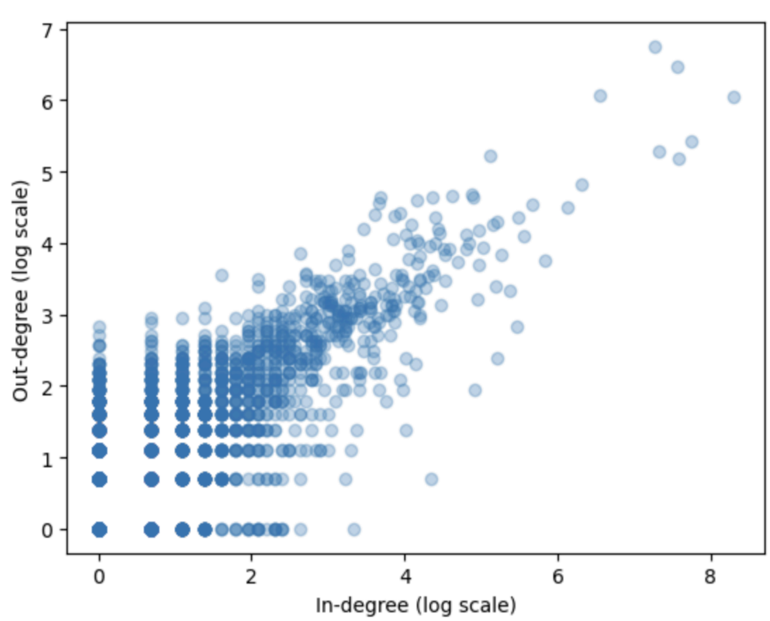
We'll start by investigating the Tudor State Papers as a directed network. In this directed network, there is an edge from person A to person B when A has sent a letter to B. (Note that you would typically add a weight for each edge corresponding to the number of letters that A has sent to B; we won't use that weight in this assignment, so you can create it or not - up to you.)

Treating the correspondence network as a directed network, please answer the following questions:

1. [5 points] What is the average in-degree?

1.9511186770428015

2. [5 points] Please make a scatterplot that shows the relationship between a node's in-degree (x-axis) and out-degree (y-axis). Each point in the scatterplot should be one of the nodes in the network. Please use log scales for the x- and y- axes. Paste the scatterplot to the space below.



PART B (10 points)

To answer the following questions, please treat the correspondence network as an undirected network. In this undirected network, there is an edge from person A to person B when A has sent a letter to B. The weight of the edge should correspond to the number of letters that A has sent to B.

1. [2 points] Please **choose one** of the following claims by **checking one of the boxes** below and conduct a hypothesis test to quantify the strength of the evidence in support of it. Please

- use the configuration model to construct a null model
- focus on the undirected version of the network

■ **(Claim 1)** Is there evidence of triadic closure? Test this by investigating whether the average clustering coefficient in the network is higher than we would expect if connections were formed at random.

☐ **(Claim 2)** Is there evidence of assortativity by degree? In other words, does it appear to be the case that nodes with high degrees are more likely to be connected to other nodes of high degrees than we would expect by chance?

2. [8 points] Report your results. First, clearly state your conclusion in one or two sentences. Second, justify your conclusion by presenting and describing your supporting evidence (hypothesis testing statistics, network metrics, plots, etc.) in a paragraph. Plots can be pasted to the next page.

There is significant evidence of traidic closure in the Tudor State Papers network. By constructing a configuration network, I created a model “that can generate networks that are random, but whose degree distributions match what we see in some real-world network.” Using this model, I calculated the average clustering coefficient in the Tudor network, with a value of 0.1747101288269982, or:

tudor_cc = 0.175.

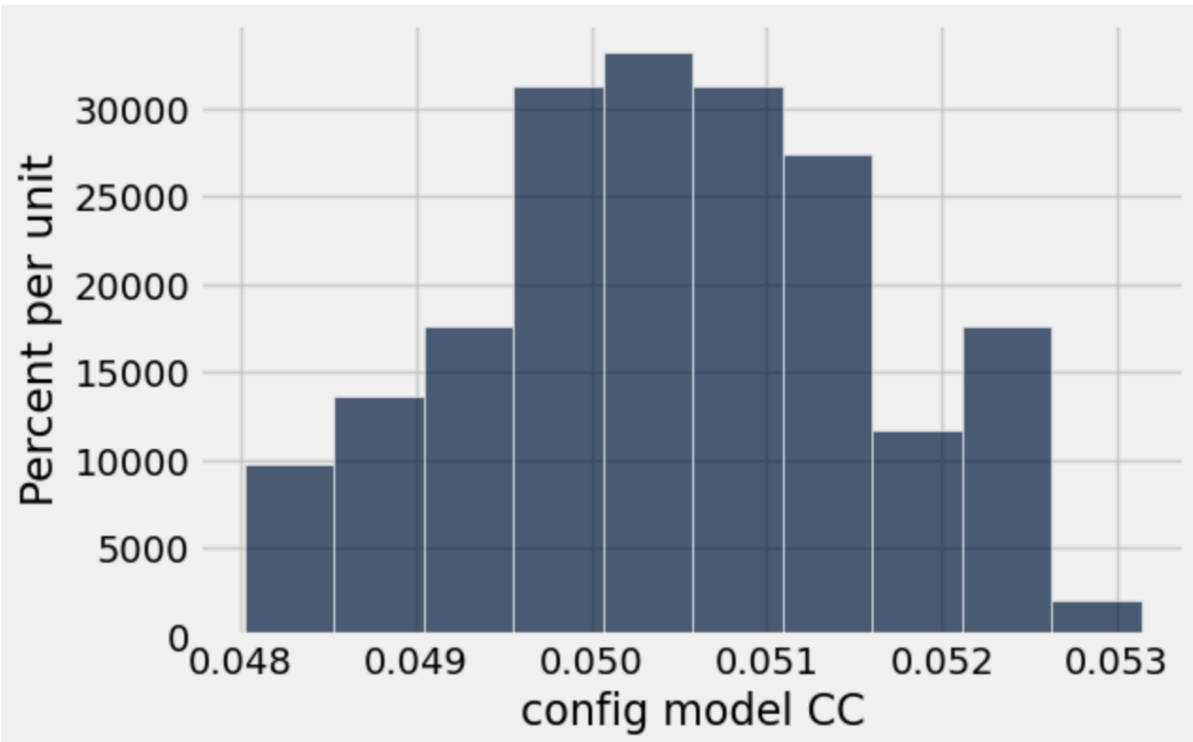
Next, I used the configuration network to run 100 simulations on random networks that match the degree distribution of the Tudor network. Ideally, I would have run a larger number of simulations, perhaps 1,000 or even 10,000, but given the limited computing power, 100 simulations will have to do. (That being said, the results are so clear at 100 simulation that they are unlikely to have significant change given a larger number of simulations). With the output of those simulations, I calculated the minimum and maximum values, as well as the 25th, 50th, and 75th percentiles:

Minimum = 0.048021139456 → 0.049
25th percentile = 0.0496866167036 → 0.050

50th percentile (median) = 0.0505104709067 → 0.051
75th percentile = 0.051296055866 → 0.051
Maximum = 0.0531362150473 → 0.053

Analyzing these values, it is clear that even the maximum clustering coefficient value (0.053) generated by the simulations on the configuration model, falls massively short of the clustering coefficient observed in the Tudor network (0.175). Given this information, we can safely conclude that the clustering coefficient in the Tudor network is significantly greater than we would expect thanks to chance alone, meaning that the Tudor network experienced more clustering, very possibly due to the principle of triadic closure.

Below is a histogram of the clustering coefficient values of the 100 simulations, followed by the computation of the minimum, maximum, and percentile values.




```
print(min(cc_table.column(0)))
print(np.percentile(cc_table.column(0), 25))
print(np.percentile(cc_table.column(0), 50))
print(np.percentile(cc_table.column(0), 75))
print(max(cc_table.column(0)))
```

```
0.048021139456
0.0496866167036
0.0505104709067
0.051296055866
0.0531362150473
```

PART C (10 points extra credit)

There are many famous people in this dataset. Here are a few (note: these are the spellings of these names found in the dataset):

a) Henry VIII
b) Katharine of Arragon
c) Thomas Cromwell, Earl of Essex
d) Emery Molyneux
e) Desiderius Erasmus

1. [4 points] For each of these people, please report

(i) their in-degree

Henry VII: **704**
Katharine of Arragon: **22**
Thomas Cromwell, Eral of Essex: **1960**
Emery Molyneux: **0**
Desiderius Erasmus: **39**

(ii) their out-degree

Henry VII: **430**
Katharine of Arragon: **22**
Thomas Cromwell, Eral of Essex: **177**
Emery Molyneux: **2**
Desiderius Erasmus: **96**

(iii) their clustering coefficient

Henry VII: **0.006184422848047197**
Katharine of Arragon: **0.10661268556005399**
Thomas Cromwell, Eral of Essex: **0.0011841127076401838**
Emery Molyneux: **1.0**
Desiderius Erasmus: **0.02503967554223241**

(iv) their eigenvector centrality

Henry VII: **0.04153751662262397**
Katharine of Arragon: **0.0026266482807409987**
Thomas Cromwell, Eral of Essex: **0.05018067799421277 5**
Emery Molyneux: **0.00900057048223011**
Desiderius Erasmus: **0.003254850080722831**

2. [6 points] Using these metrics, what would you hypothesize about the position each of these people occupies in the network? (You don't have to do any additional coding for this part - the question is asking you to interpret these quantities, perhaps in conjunction with a little bit of background on who these people were.)

As King of England and the man who initiated the English Reformation, Henry VIII received (704) many more letters than he sent (430). Being such an important and influential figure, however, means both of those numbers are quite large. With so many connections, his clustering coefficient is rather small, as expected. His eigenvector centrality is on the larger side, at least compared to the others in this selection. This make sense, as he would've been a very central player of the era.

Katharine of Arragon was a prominent political figure herself, having been married to Henry VII. However, her in- and out- degrees are extremely low, perhaps due to the role of women in society at the time. With her smaller degree comes a large clustering coefficient, but her eigenvector centrality is tiny at 0.002, meaning she did not play a "central" role in this network of letters.

Thomas Cromwell was a lawyer who worked closely under Henry VIII and was a chief proponent of the English Reformation. He sent around a third of the letters that Henry VII did, but received an absurd amount, nearly 2000. This was likely due to his position as chief minister to the King, so many letters were probably directed to Cromwell first. With that comes an almost negligible clustering coefficient, but the highest eigenvector centrality of the group at 0.05.

Emery Molyneux was a mathematician and globe-maker. He became acquainted with many of the prominent men of the day, with his globes acquired by royalty and noblemen. However, Emery did not seem to be big on the writing front, with no letters received and only two sent. The two people who received his letters were also friends with each other, resulting in a clustering coefficient of 1! His clustering coefficient, however, was quite small, which is expected given his miniscule degree.

Desdiderus Erasmus, was a priest and theologian who translated a great amount of literature. I assume he is also the namesake for the European study abroad program, which students refer to simply as Erasmus. He received a moderate amount of letters, 39, and sent more than twice as many. A clustering coefficient of 0.25 is relatively large considering his degree, which means many of his connection were also connections with each other. Perhaps this was his academic circle. His eigenvector centrality is extremely small, 0.003, which further strengths my theory of his academic circle.