Name: Assmaa Aziz **Partner Name (if applicable):**

Professor Dennis Feehan
November 12, 2024

University of California, Berkeley
Department of Demography

Demography 180: Social Networks
Mini Project Results

PART A (10 points)

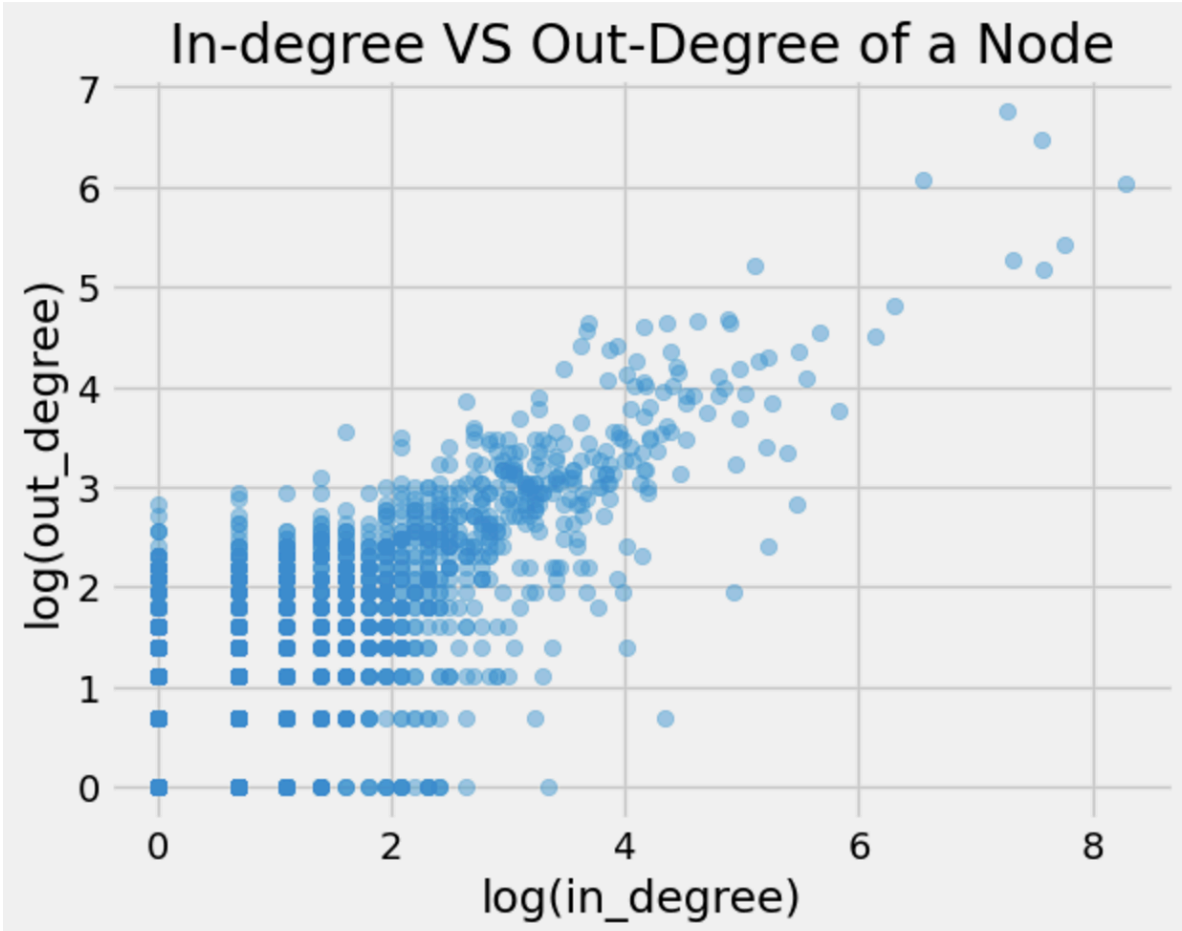
We'll start by investigating the Tudor State Papers as a directed network. In this directed network, there is an edge from person A to person B when A has sent a letter to B. (Note that you would typically add a weight for each edge corresponding to the number of letters that A has sent to B; we won't 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?

The average in-degree is 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

X	**(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? Test this by investigating whether degree assortativity in the network is higher than we would expect if connections were formed at random.

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.

H0: The average clustering coefficient in the network is equal to what we would expect if connections were formed at random. There is no evidence of triadic closure.

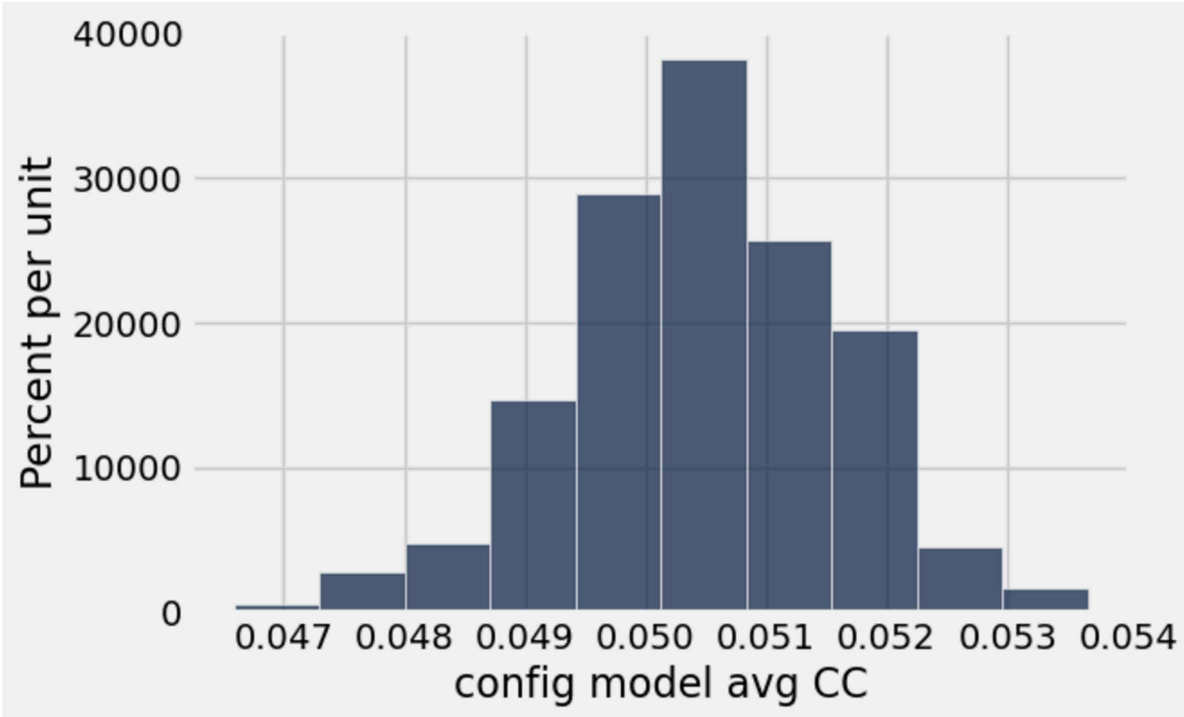
H1: The average clustering coefficient in the network is higher than we would expect if connections were formed at random. There is evidence of triadic closure.

Conclusion:
We **reject the null hypothesis** that states there is no evidence of triadic closure. From our test, we can conclude that the average clustering coefficient in the network is higher than we would expect if connections were formed at random.

Justification:
Our test statistics are the average clustering coefficients obtained from our configuration null model, which are stored in our table below.

Observed_cc = 0.1747101288269982

Based on the distribution of our null model, we can see that the average clustering coefficient in many random configuration models is centered around ~0.05.
When comparing the average clustering coefficient of our observed network to the distribution of the null model, we can see that our observed value is unlikely under the null model.
Therefore, we reject the null hypothesis that states there is no evidence of triadic closure.
From our test, we can conclude that the average clustering coefficient in the network is higher than we would expect if connections were formed at random. There is evidence of triadic closure.



config model avg CC
0.050772
0.050131
0.0515326
0.0510749
0.0483908
0.0508766
0.0507968
0.0512113
0.0518613
0.0510362

... (490 rows omitted)

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

(ii) their out-degree

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

(iii) their clustering coefficient

Henry VIII: 0.007248701306790577
Katharine of Arragon: 0.10661268556005399
Thomas Cromwell, Eral of Essex: 0.0013874739561082375
Emery Molyneux: 1
Desiderius Erasmus: 0.022069424420538983

(iv) their eigenvector centrality

Henry VIII: 0.08961102002782177
Katharine of Arragon: 0.004076069220006484
Thomas Cromwell, Eral of Essex: 0.08399252390052572
Emery Molyneux: 9.418064404341942e-32

Desiderius Erasmus: 0.004448673389188995

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.)

Clustering coefficient tells us where there are tightly knit groups within a network. In this context, a high clustering coefficient means the probability that a node's neighbors are sending letters to one another is high. This means there is a group of nodes where letters are exchanged regularly between them. In eigenvector centrality, more central nodes count more than other nodes, where a node benefits from being connected to important nodes. In this context, large eigenvector centrality would mean that a node is exchanging letters with influential individuals and could be an important individual as well.

Henry VIII

Henry VIII is second in receiving the most letters and has the largest eigenvector centrality. He has the second lowest clustering coefficient. He sends out the most letters but receives more letters than he sends letters. From these metrics, it appears Henry VIII is a **central, influential node** whose neighbors are not highly connected to one another. This means Henry VIII may be in a local bridge and can serve as a **gatekeeper of information.** This all makes sense because Henry VIII was the king of England, so he would have strong influence and would be in communication with many other important individuals. He would be connected to many unconnected individuals, like nobles, other foreign leaders,etc. He could maintain control since he was the main connection people had to those other individuals.

Katharine of Arragon

Katharine has less connections than Henry VIII, but sends and receives an equal number of letters. She has the second largest clustering coefficient, which means her neighbors are highly connected and that she's part of a tight knit group. She also has lower eigenvector centrality, so she is not a central node like Henry VIII. Katharine of Arragon was Henry VIII's first wife and they eventually got divorced. It makes sense that she would have less influence and play less of a part in government than her husband and those who worked alongside him. **Her role in the network is not highly influential.** She could play the role of a **connector**, such as being the reason the others in the group are connected to each other, such as. people she met while being married who maintained connection with her after the divorce.

Thomas Cromwell, Eral of Essex

Thomas Cromwell receives the most letters and has the second largest eigenvector centrality. He has the lowest clustering coefficient. Similar to Henry VIII, Thomas Cromwell is a **central, influential node** whose neighbors aren't very connected to one another. He would have a similar position as Henry VIII, where he can link unconnected nodes to one another and can serve as a **gatekeeper of information.** Thomas Cromwell was the chief minister to Henry

VIII, which means he was essentially running the government and working under Henry VIII, so it makes sense that he had high influence and received many letters.

Emery Molyneux

Emery received 0 letters and sent out 2. He has the largest clustering coefficient and the lowest eigenvector centrality. These metrics imply he had little influence and hardly any interactions with others. Those he did interact with were connected. Emery Molyneux was a maker of globes, mathematical instruments, and weapons. Emery likely only communicated with other professionals in his industry and **lived a more isolated life.** He would hold **a weak, uninfluential position within the network.**

Desiderius Erasmus

Desiderius was sending and receiving letters more often than Katharine and Emery but less often than Henry VIII and Thomas. We can see the similar pattern that his various metrics are in between that of the other people we've analyzed. His clustering coefficient is smaller than the people who play a weak role in the network but it's larger than those with strong roles. His eigenvector centrality is larger than those who play weak roles but it's smaller than those who play strong roles. Based on our previous analysis, we can expect Desiderius to be more influential in the network than Katharine and Emery but less influential than Thomas and Henry VIII. Desiderius was a Catholic priest, theologian, satirist, and philosopher. Being a scholar makes sense that he would be connected to many but not as much as King Henry VIII and those working for him. Desiderius would have **medium influence on the network.** He could play the role of a **mediator** since he is a priest and highly educated. These traits could cause others to trust him and his opinion.