Background and Research Question

The literary classic Romance of the Three Kingdoms by author, Luo Guanzhong, is one of the four great pillars of Chinese literature. The book is set in the third century and draws upon China's history and folklore to create a story showing the era's politics and social affairs. The book follows the ascent of Sima Yi who is part of the Han Dynasty. They identified themselves as China's next rulers but have brought the empire to the brink war.

Through strategies, conversations and fights for power we can trace the relationship between the 70 characters present throughout the book. From the various relationships we are curious to explore questions like: which characters like or dislike each other throughout the story? Would sentimental analysis identify certain relationships between characters from different countries? Would the relationship network change over time?

Data Understanding

The original book is written in Chinese. And there are vast amounts of English translations of the book. For our analysis, we collected our data from one of the existing translation on Github[1]. And the Github page also provides a list of 70 major character names. To provide detailed analysis for characters from different countries, we labelled each character for which country they are in based on history fact.

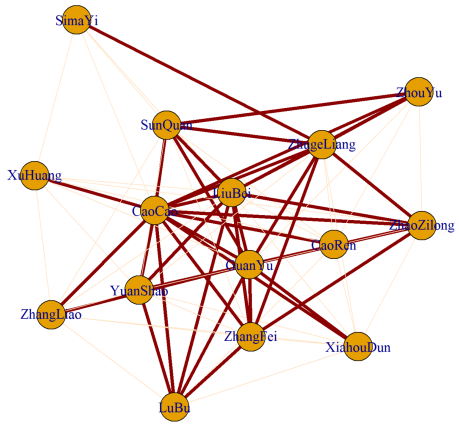
The primary approach we are using to define a relationship exists is by applying textual analysis. We first divide the sentences into tuples of three words. Then we created a function to compare the distance between two character names. If that distance is above a certain threshold, we define that relationship exists. This approach is intuitive to discern undirected relationships. We tried to build directed networks to better understand interactions between characters. But

[1] https://github.com/dmanolidis/three-kingdoms/blob/master/Text%20Files/fulltext.txt

considering that it's challenging to identify directed relationships based on verbs for subjunctive verbs, we decided to proceed our analysis based on undirected networks.

Exploratory Data Analysis

First, we wanted to examine if our method on retrieving data had worked properly. To do so, we constructed an igraph containing only the top 15 percentile centrality score (graphed separately for betweenness and degree). The names that appeared on the graph correspond to well known and influential characters of the book. In the graph, we can discover that CaoCao, LiuBei, and SunQuan, who is the king of the three kingdoms, are in the middle of the most important characters. Furthermore, the rest of the characters have a high hierarchy in the kingdom. It makes sense that people with higher hierarchy have more connection to other people. This confirmed that our approach has worked and we can then proceed to further analysis on the network.



What else can we do to study the relationships between characters within this book? We believed that we can study how the network changed overtime and see how the power dynamic

changed throughout the book. Therefore, we went back and performed the same procedure as before, only chapter by chapter this time.

Since the data we extracted comes in the form of one cell per chapter and has the node, edge, weight and chapter number all listed together. We had to first clean out punctuation characters such as "(", "[", and "/". We then realized that R does not recognize some characters as the dame due to empty spaces in front that string split has caused. We then got rid of any spaces for all columns to ensure this doesn't become a problem again. Next, we thoroughly examined the cleaned data set and found out there were duplicates in the data with two people in the opposite position but the same weight. Therefore, we have filtered the data again to where every entry is unique.

Lastly, because creating 120 igraphs is unnecessary and chapter by chapter might not show much progression, we decided to divide the whole book into 12 sections, containing 10 chapters each. Therefore, we had to label each relationship by section number. For each section, we then added all the weights of a relationship between the same two people across chapters.

We now have a cleaned dataset to work with, containing 5 columns: "Person 1", "Person 2", "Weight", "Chapter", and "Section".

Network Exploration and Analysis

For each section, we created a graph and computed the degree, betweenness, and closeness. We then saved this information into a dataframe with corresponding character names next to it for further visual analysis on Tableau.

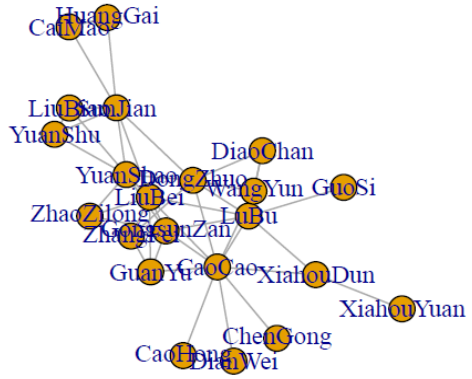
At the beginning, the characters often form a big cluster. As time progresses, the network starts to become denser and denser, then starts to decay as it reaches the end. More importantly,

in the later section, there are some small clusters and one on one relationships shown in the network. We can actually match the network with plots of the book that involve those characters at that time.

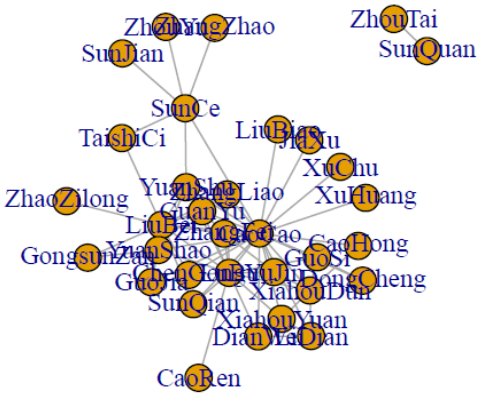
There are some interesting observations. In many sections, CaoCao served as a connector in the network with lots of connection with people within his country. This is corresponding to the history fact, because Caocao is the ruler of wei, and all the nodes connected to Caocao are his military generals. In section 9, we can find a 2-plex within that network. Guanxing is Guanyu's son. And Zhang Bao is Zhang Fei's son. Since Guanyu, ZhangFei and Liubei shared a brotherly relationship with each other, it makes perfect sense that their child also knows each other as well. In section 10, we can see there is a bridge in that network. If we remove the relationship between Sun Quan and Zhuge Liang, Zhang Zhao and Lu Xun will be disconnected with the rest of the network.

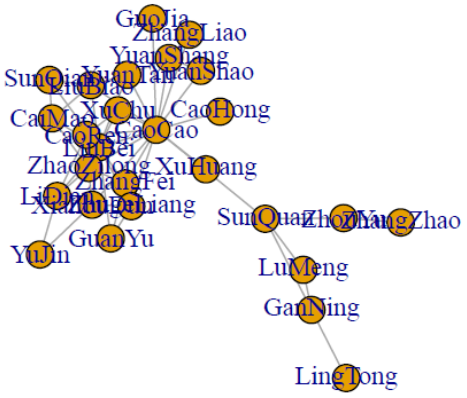
Moving on, in section 11, we noticed a small cluster consisting of Guan Yu and Lu Meng, which corresponds to the plot that Lu Meng killed Guan Yu. In section 12, there is a small cluster consisting of Cao Pi and Cao Cao, which correspond to their father and son relationship. We also discovered that the remaining character is all about the next generation in the last section. Moreover, many people have faded out of the network. Those character probably died sometimes in between the last few chapters.

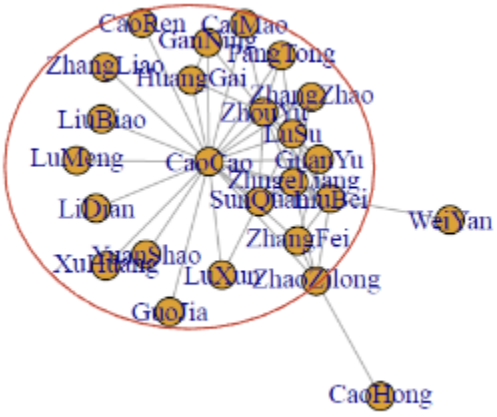
Chapter Section 1

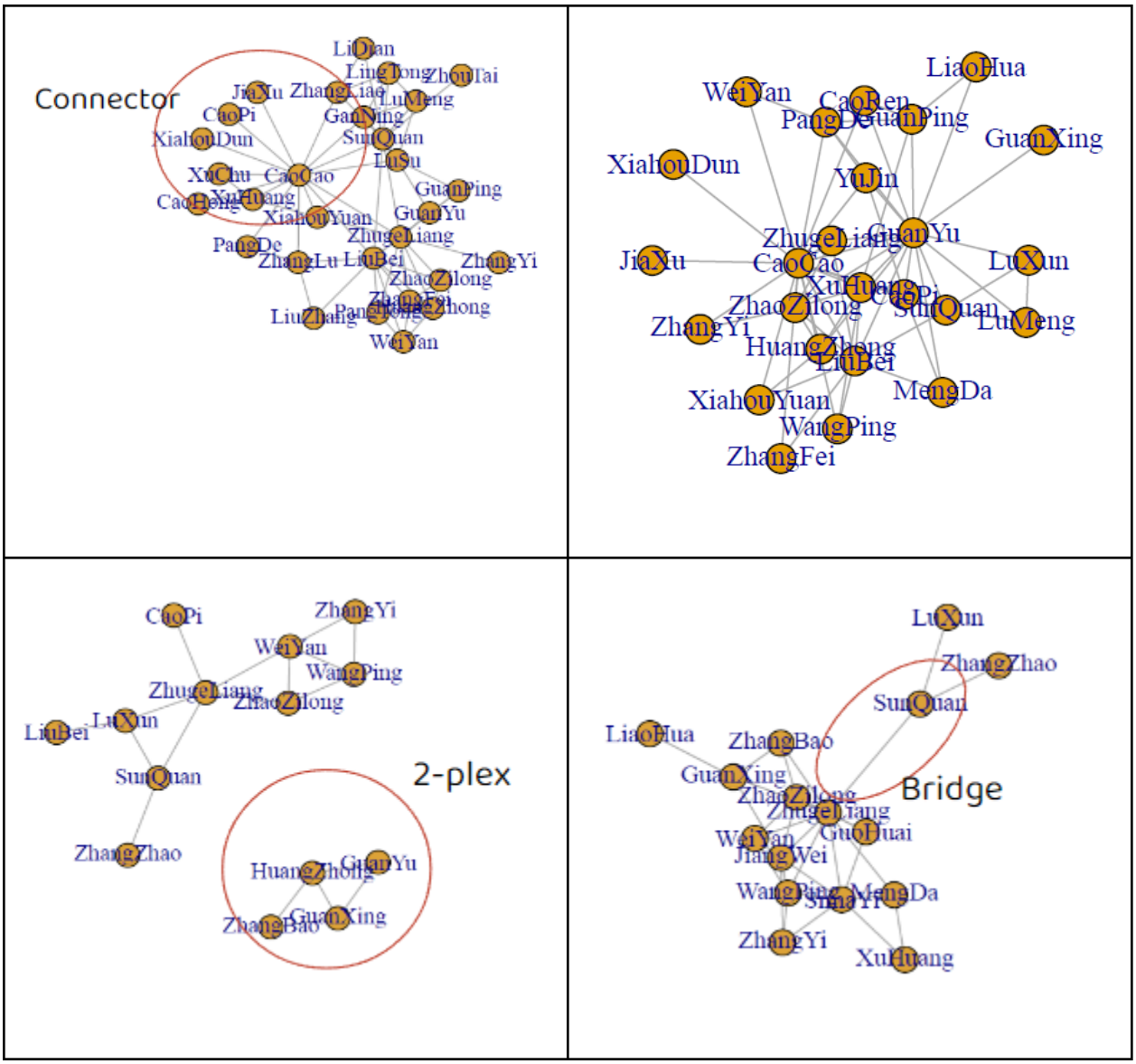


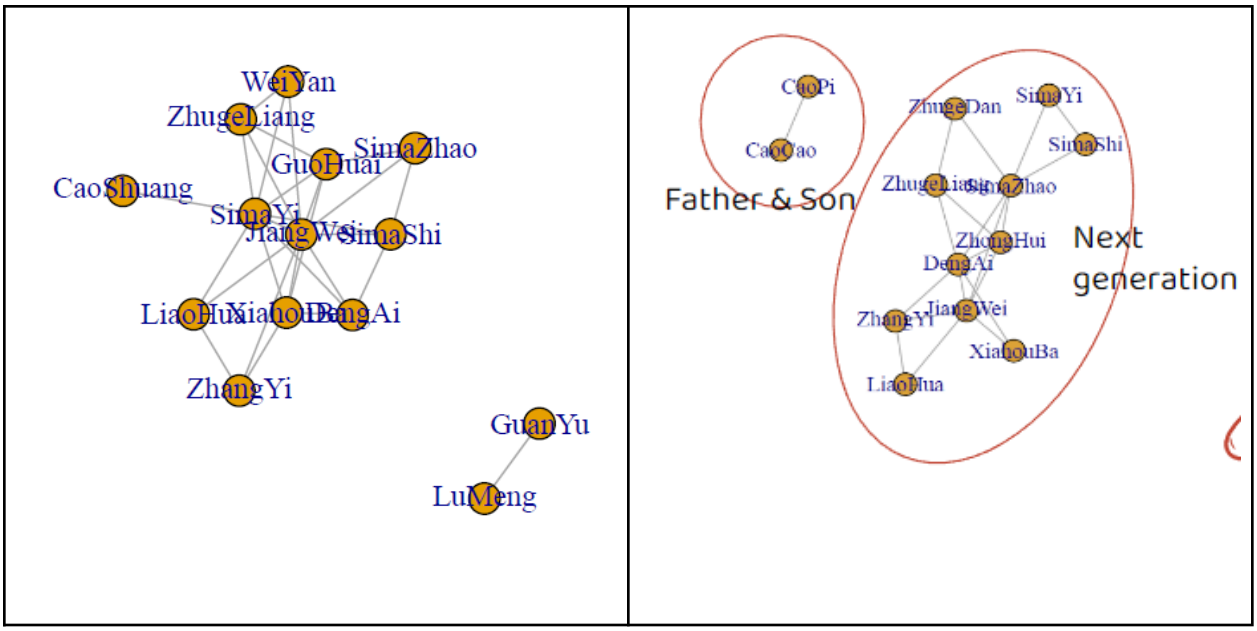
Chapter Section 2








Connector
LiuBei
ZhangFei
ZhaoZilong
ZhugeLiang
GuanYu
ZhangYi
XiahouYuan
WeiYan
LiaoHua
YuJin
GuanXing
XiahouDun
JiaXu
CaoCao
LuXun
SunQuan
LuMeng
MengDa
WangPing
ZhangZhao
2-plex
HuangZhong
Bridge



Sentiment Analysis

While it's incredibly important to measure the classic network parameters we're familiar with as observed above, this analysis covers only a piece of the overarching puzzle that is character behaviours, motivations, and interactions in this layered novel. To truly understand what these relationships mean for the extent of these character actions, we need to understand the positivity or negativity of the associated language for each character. This is the basis for our exploration into sentiment analysis.

Our sentiment analysis is relatively simple – we sought to measure the relative positive or negative sentiment of sentences for each line in the book and then parsed for each character name in those respective sentences. From there, we iterated over each sentence and character so that we could find a cumulative total for each character over the entire book. To accomplish this, we utilised the VADER lexicon toolkit and selected the Sentiment Intensity Analyser, which we dubbed as 'SID' (sentiment intensity device). We then created a tuple of all 70 characters as well

as a subset of the entire novel (chapter 1) to test out our analyzer and ensure it's working as validity was easier to guarantee on a smaller sample size.

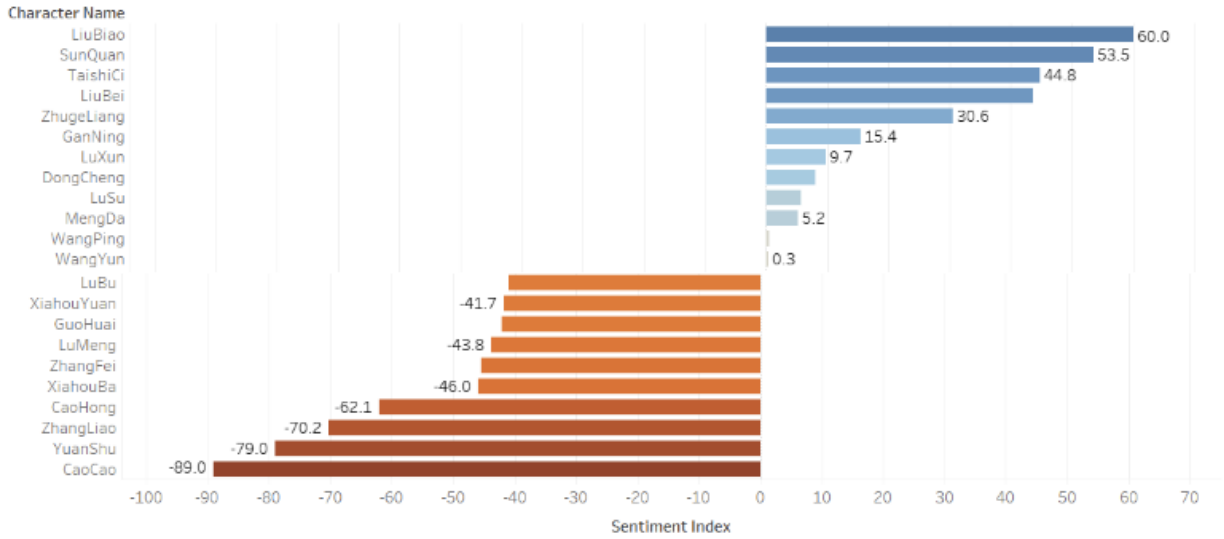
Upon guaranteeing its validity, we ran the sentiment analyzer through the entire novel and below are the results of that analysis.

Most Liked Characters

Character Name	
liu biao	59.98
sun quan	53.49
taishi ci	44.75
liu bei	43.56
zhuge liang	30.59
gan ning	15.44
lu xun	9.69
dong cheng	7.98
lu su	5.63
meng da	5.20

Most Disliked Characters

Character Name	
cao cao	-89.00
yuan shu	-78.97
zhang liao	-70.21
cao hong	-62.08
xiahou ba	-45.95
zhang fei	-45.42
lu meng	-43.77
guo huai	-42.02
xiahou yuan	-41.73
lu bu	-40.98



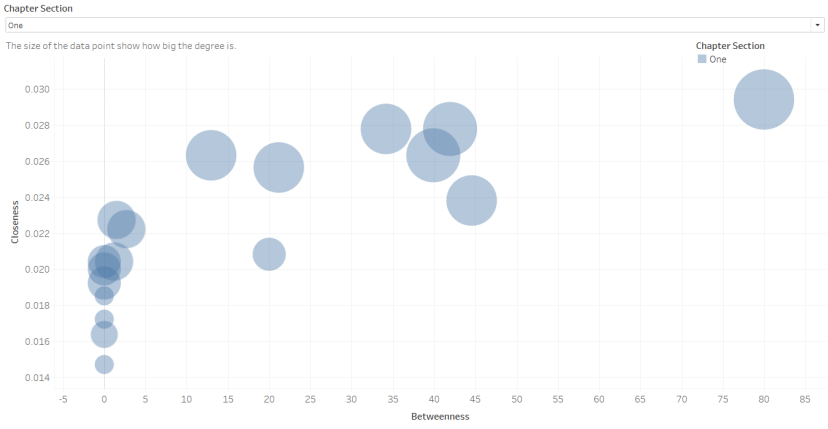
There are a couple of main takeaways. The first is that, on average, the sentiment of characters are actually negative, as there are significantly more negative values than positive

ones and the overall intensity of negative values is higher as there's higher values on average. Second is the few standouts for significant positive sentiment. Although all of the top 10 negative sentiment characters are relatively large, only the top 5 positive sentiment characters have any meaningful positive values. These takeaways make sense given the novel focuses on the aspects of war, so there would be plenty of reasons to have negative sentiment scores given the death and destruction associated with it.

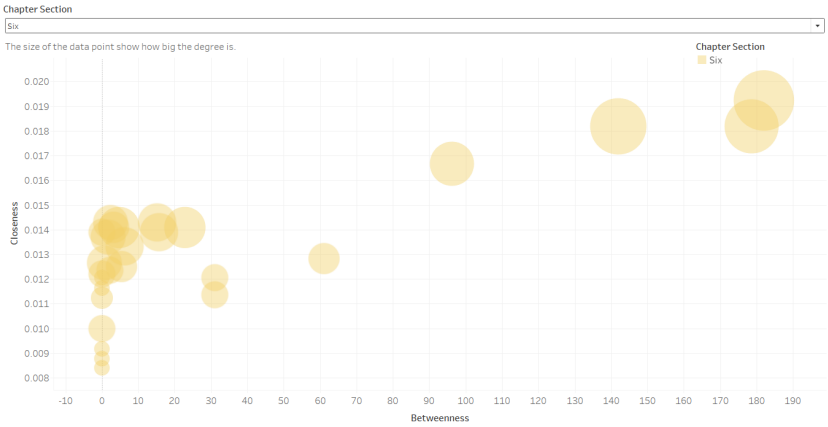
Lastly, we were intrigued with the fact that our protagonist, Liu Bei, was only 4th on our list rather than higher. This led us to believe that, since Liu Bei is involved in the main plotline and has to deal with the negative aspects of war, those sentences dragged down his cumulative score relative to someone that would serve a secondary role of mentor, gifter, or other role such as Liu Biao.

Network Insights

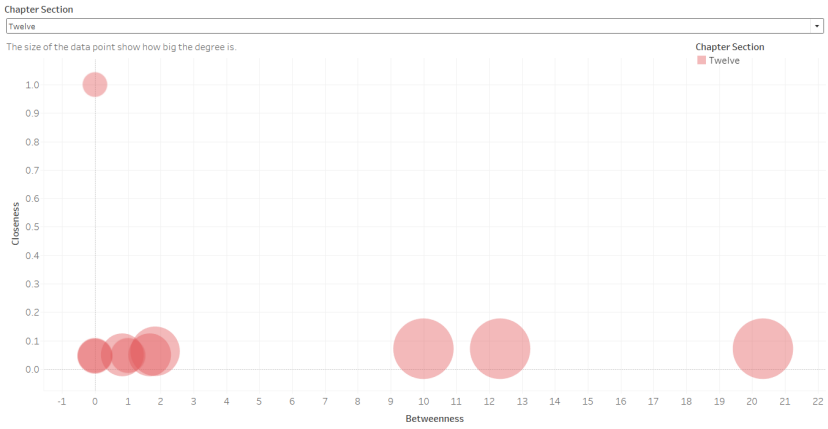
From our network data we explore the relationship through degree centrality, betweenness and closeness score. As we split the book into 12 sections we were able to see the change over time of each character's scores and the introduction of new characters throughout the book. We see that at the start of the book there were around 20 characters who were introduced with CaoCao having the highest degree, betweenness and closeness score. This means that CaoCao is the most connected, has the most paths, and is connected with others who have large connections.



By the middle of the book we see that most of the characters have appeared in the book and some have been killed due to war going on as their score disappears. At this time we can see the change in power and influence as Liu Bei, Zhou Yu and Cao Cao scores were in the top three throughout the middle section of the book. This also aligns with the context as Cao Cao is the ruler of Wei, Zhou Yu is the chancellor of Wu and Liu Bei is the Ruler of Shu, which are the 3 Kingdoms in the story.



By the end of the story we see that only around 10 characters remain with new characters we didn't see at the start tops the degree and betweenness. This shows the change in power and influence. We see that Sima Zhao, Jiang Wei and Deng Ai are at the top and signifies their rise to power by the end of the book. However, we see Cao Cao ended up last in the degree centrality and betweenness score as the character loses power, but interestingly Cao Cao still remains at the top for closeness. This suggests that they are still the person who connects everyone to another character.



When exploring the network score in a time series we were able to see the movement of each character betweeness, closeness and degree centrality, which helps indicate how connected

and influential they are at that point in time of the book. In the time series, it also allow us to easily identify when a character have been killed or when the chracter influence increas or decrease.

Exploring the top liked (Liu Bei, Sun Quan, Taishi Ci, Liu Bei and Zhuge Liang) and bottom dislike 5 character (Cao Cao, Yuan Shu, Zhang Liao, Cao Hong, Xiahou Ba), which we were able to determined through our sentiment analysis, network scores over time. From what we know, we see that being the most connected like Cao Cao does not mean that you will be the most like. This goes inline with the idea of the most power and influence you have the less people like you. From the visualizations we see that many characters network scores increase and decreases throughout the story as they go in and out of power and influence. The time series help us tell a story of each character and how their rise or fall from power happens throughout the book. Through the graph we also are able to see when each character appears in the book and where they die by looking at the start and the ending of the line.

Overall, the findings for our network analysis mostly correspond to historical facts and the book context. We were able to see where each character dies, where they were the most influential and their relationship between each character. Although the rank of closeness measure might be somewhat unexpected, we believe that our analysis is still reasonable. The counter intuition might because the text is conveying underlying messages from writers that readers may often overlook.

Future Exploration

We believe that our network analysis could be applied to any other novels. The analysis on local structures, in addition to closeness measures, provides an additional lens for both writers and readers to understand different aspects of characters. Combining with sentimental analysis, our network analysis would be beneficial for theme writers to create more engaging stories.