Annie Wu

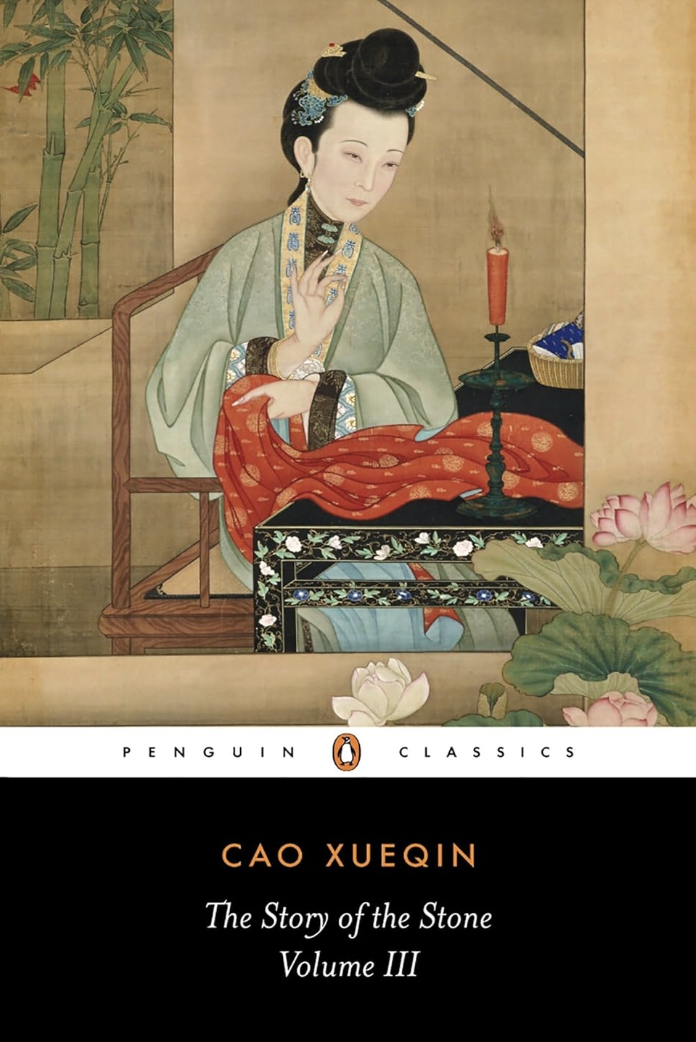
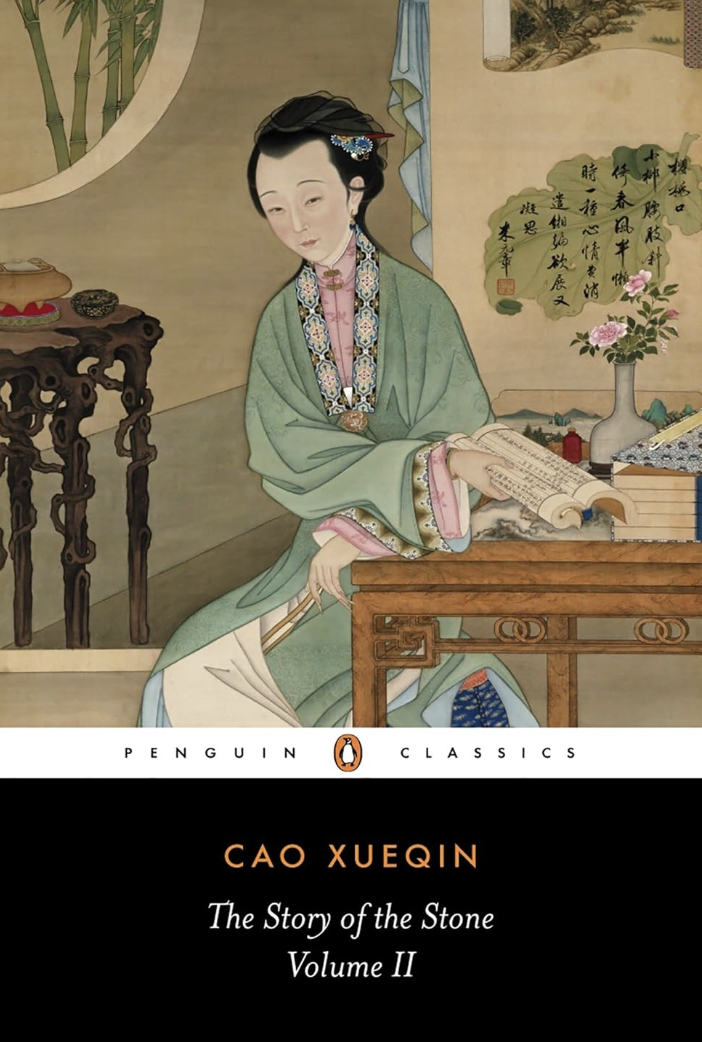
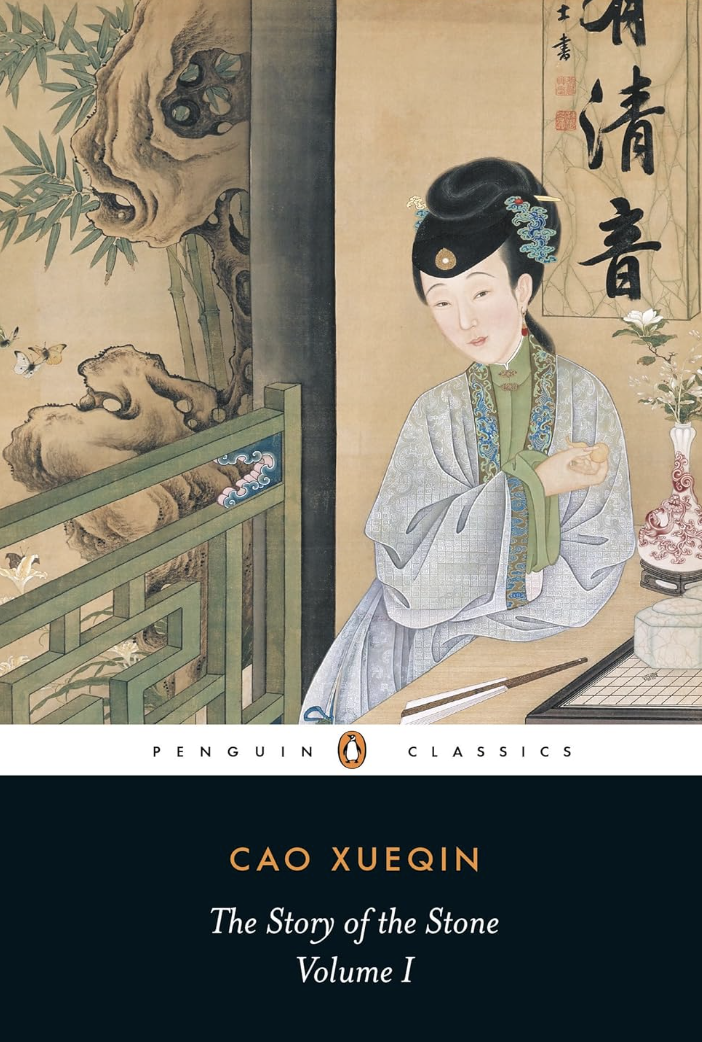
Prof. Brooke Luetgert

DIGS 20004: Data Visualization for the Humanities

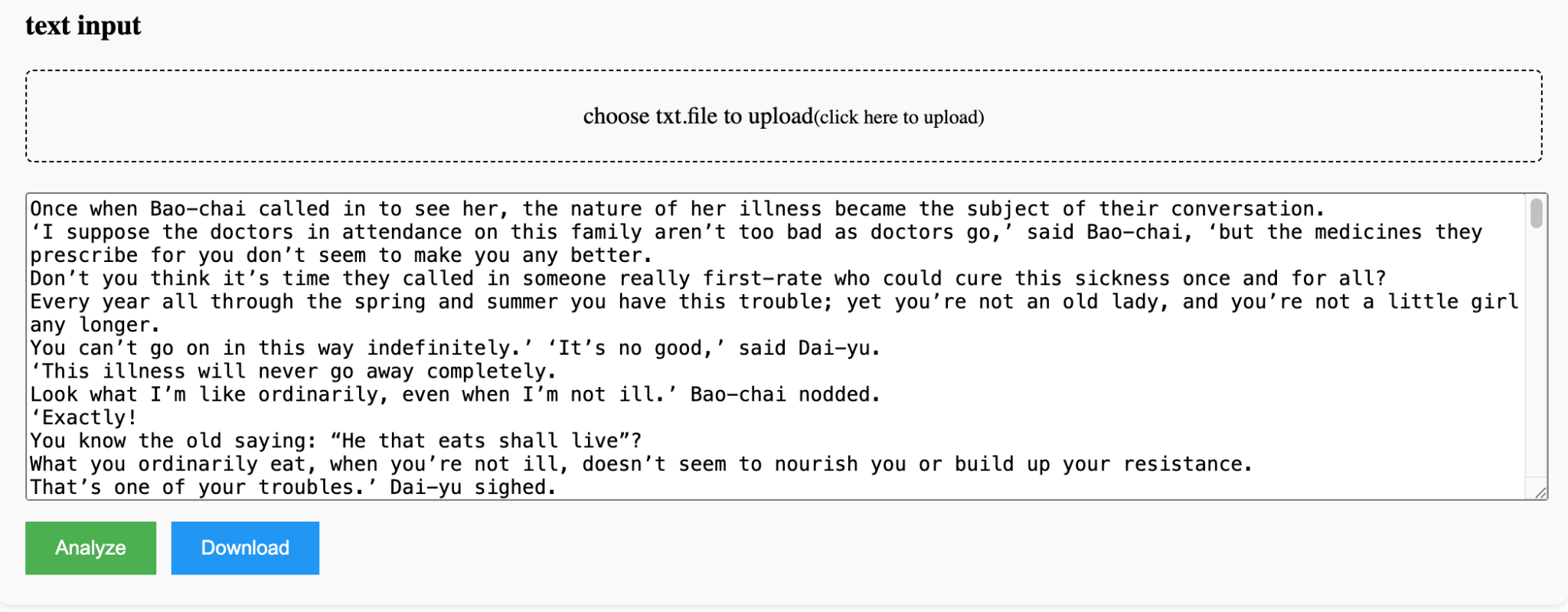
March 14th, 2025

Text & Network Analysis for *Dream of the Red Chamber*

In this project, I chose to visualize the sentiment analysis and network analysis of the classical Chinese text *Dream of the Red Chamber* from the 18th-century Qing Dynasty. For the sentiment analysis, I want to analyze the direct interaction between the two female characters Bao-chai and Dai-yu. However, as I was reading through this literary work, it seems that there’s only a few chapters that actually involve direct interactions (verbal) between the two characters, even though they appear together along with other characters simultaneously multiple times. In this case, I extracted their direct interactions out of the first eight chapters into TXT files for sentiment analysis. And I used David Hawke’s English translation *Story of the Stone* published in the 1970s.[[1]](#footnote-0) For the network analysis, I originally planned to manually input characters’ relations and descriptions or use generative AI to build a CSV file for analysis. Then I realized that neither myself or generative AI could digest and formulate such large numbers of total 242 characters within the timeframe. Therefore I decided to use preexisting data of character relations collected by previous research in Chinese[[2]](#footnote-1) and translated them by google translate in python and my own translation using OpenRefine.

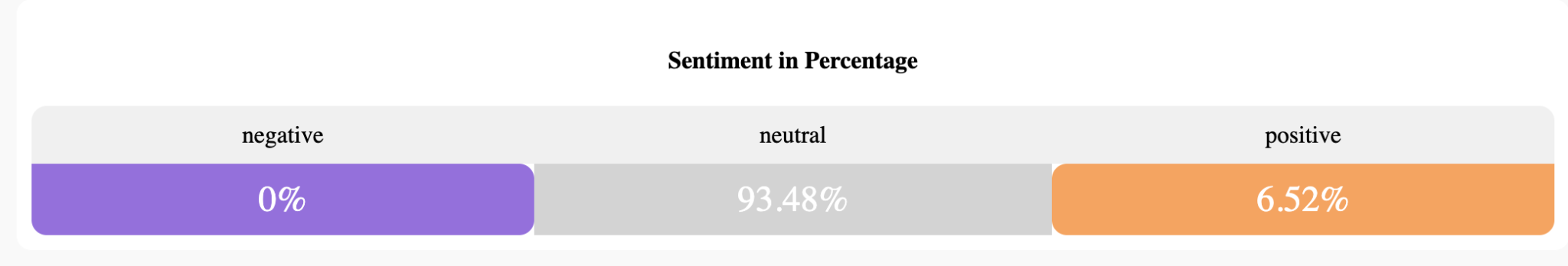


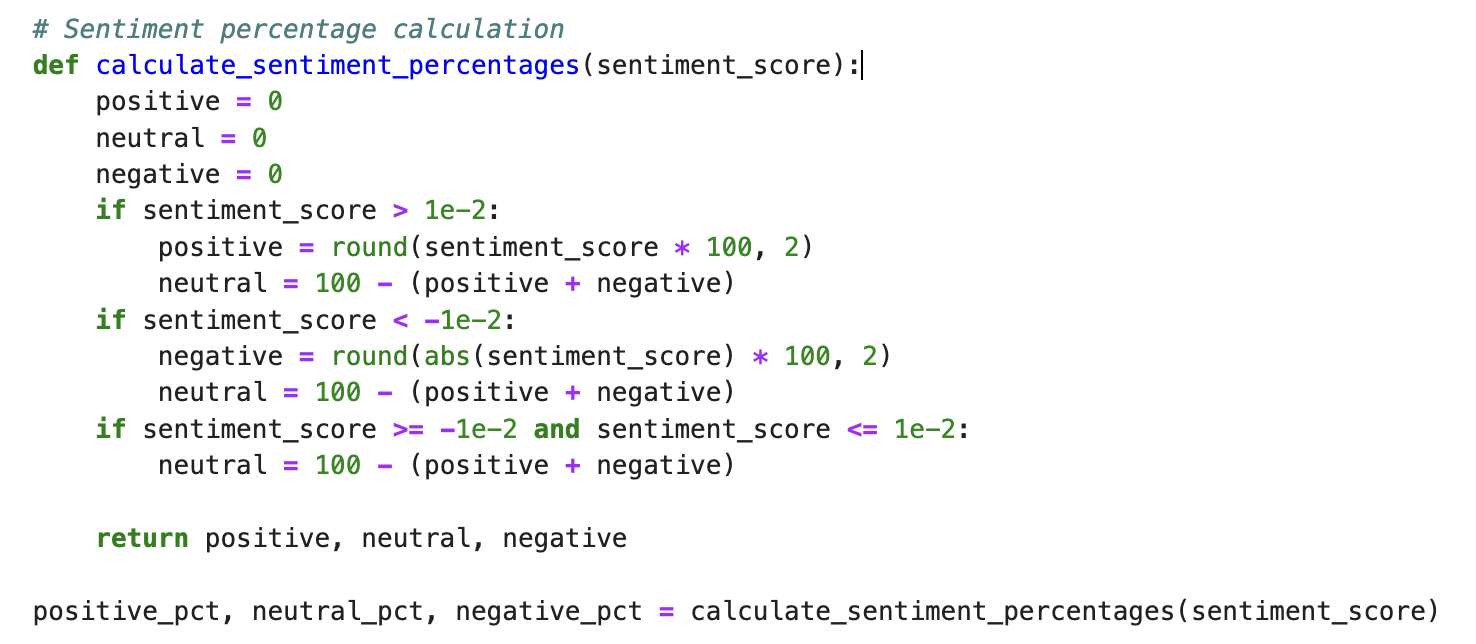
For the visualization of sentiment analysis, based on the code from week 9 of “w9\_textex,” I added several features to make the visualization more interactive and accessible. I first added a dropbox for text input, so instead of copying and pasting the text in the possibility of losing some of the content, the user can directly drop their text as a TXT file. The only note is that only TXT file is readable in this code, other formats like word document, pdf, CSV are not supported. If the user scrapes the text from somewhere else and wants to convert the text input as a TXT file, they can also use the “Download” button to accomplish that with the timestamp. In addition, the sentences in the TXT file must be separated by “/n” beforehand for it to be recognized as different sentences.



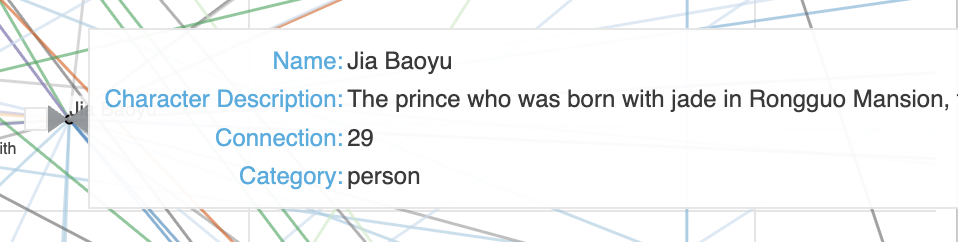


The second change I made is the color scheme and image size for word frequency. The color in the original code is blue and the word cloud is much smaller and disproportionate with the top15 words graph. I adjusted the color into red color scheme because the book title is *Dream of the Red Chamber*. I also changed the color scheme for the sentiment analysis section, the edgelines for sentence distribution to make it more aesthetically visually better and readable. I also added a little section on the percentage of different sentiments using textblob to make the sentiment score clear and easily understandable.

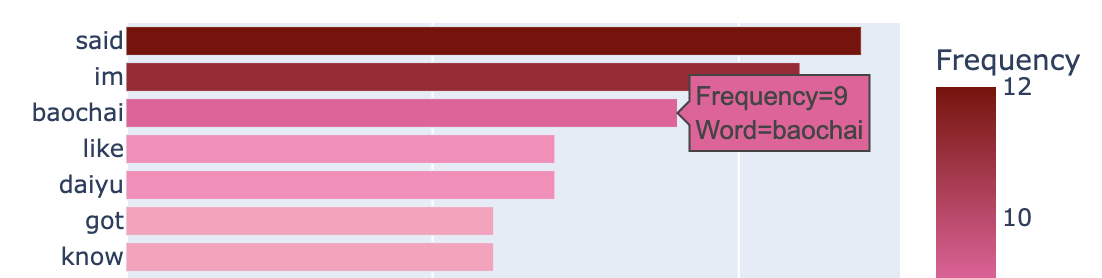




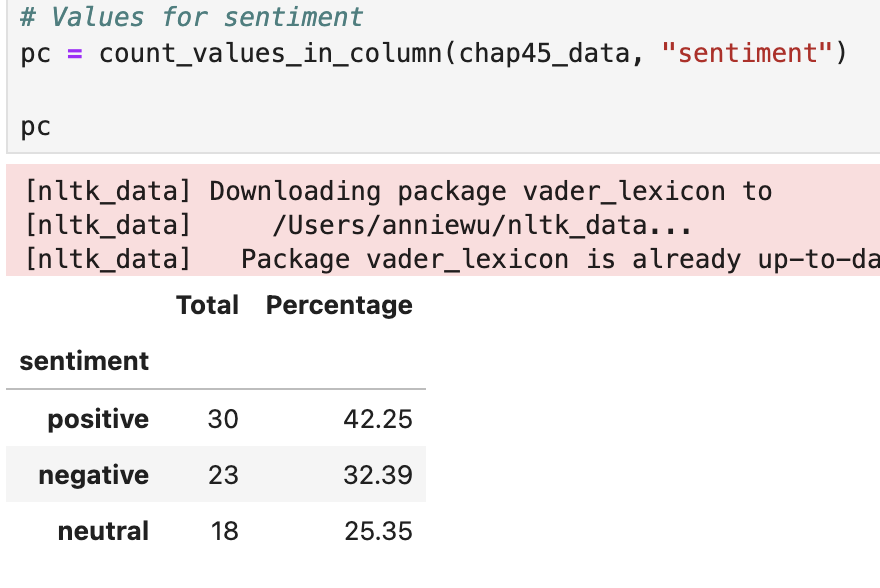
For the network analysis, I used different colors to highlight the different relationships between characters. For example, for brothers and fathers, or male kinships, the colors are highlighted in blues. Servants are highlighted in light green and friends in orange. And all edges are also labeled with relation categories as well. For nodes, each node represents a character appeared in the book. By clicking the node, the user can get the character’s name, detailed description of character’s personality and social status, as well as how many connections they have with other characters. This network could also stretch into different shapes or sizes depending on which character they want to take a closer look.



I do think that Python could provide a very general, introductory level of displaying the character’s relationships and emotions, yet I do think that it requires deeper coding and literary interpretive skills to structure a refined model to conduct quantitative textual analysis of pre-modern literary pieces. For example, the word cloud and the word frequency generator definitely catch Bao-chai and Dai-yu’s names as the focus of the sentences. While it does not necessarily provide a precise understanding on sentiment of their verbal interactions, it still gives a preliminary definition of their interaction as 93% neutral and 6% positive as presented earlier.



There are also several aspects that I believe have potential for improvements. For the network analysis, it could have a dashboard or button to filter out different types of edges. However, it seems that it must use Javascript in standalone output with Bokeh, which is something I’m not familiar with. For the sentiment analysis, I also tried with the other sentiment test with codes from week 5 on the twitter data, and the results using NLTK and ​​SentimentIntensityAnalyzer (of sentence neutral, positive, and negative) is quite different from textblob that’s implemented right now. So if possible, it would also have the potential to convert the TXT file into a CSV file to conduct a different sentiment test on the literary piece. In addition, the original w9\_textext code does have a dashboard that generates all different analyses into one page or block. I changed it into different blocks instead as I think the scrolling style/function provides a more detailed viewing and using experiences than putting all figures together, which diminishes the values of each specific part.



In terms of data management, I mostly use jupyter notebook in the output for tracing and adjustment accordingly. The codes are also available in py.file format. The TXT and CSV files are also available in named folders in this repository.

1. Cao, Xueqin. *The Story of the Stone*. Translated by David Hawkes, vol. 1–3, Penguin Classics, 1973–1980. [↑](#footnote-ref-0)
2. Fan, Chao. "Research on Relationships of Characters in The Dream of the Red Chamber Based on Co-Word Analysis." ICIC Express Letters, Part B: Applications, vol. 11, no. 5, 2020, pp. 493–500. ICIC International. [↑](#footnote-ref-1)