

---

# CSCE 5543: Final Project

---

**April Walker**  
adw027@uark.edu

University of Arkansas,  
Fayetteville, AR, 72701, USA

## Social Networks in the Game of Thrones Universe

### Introduction

Social Network Analysis (SNA) is utilized across various disciplines in order to provide insights and predictions about various social activities. For this particular project, I will develop weighted social network graphs and calculate various social network metrics for the Game of Thrones universe.

As a more exploratory measure (which you should only vaguely consider a part of this project) I also measured the cosine similarity of the vocabularies of each "top character" in the television series. My assumption is that characters which intuitively or quantitatively interact more will have a more similar vocabulary.

### Methodology

For this section, only communication between the top 10 characters was considered. The top characters were determined by the number of sentences they spoke over the period of the show considered (either a season, a few seasons, or the whole show).

In order to determine the social network, I will make a assumption similar to that in E. Nalisnick and H. Baird's 2013 paper - whichever character is speaking is directing their statement to the character who spoke immediately prior. While this is not always true, it allows a much more simplistic algorithm to calculate the weighted social network. Also, since this project explores a TV series rather than a play, there wasn't an option to assume everyone in a given scene is connected such as in Shukla, Gauch, and Evalyn's 2018 paper.

My methodology results in a collection of one-way interaction "weights", however, since I wanted to only consider the combined social network for graph simplicity, I combine the one-way interactions to get the total weight for each edge of my social network graph. To hold my "weights" I created a list of hashtables (well, it's literally a hashtable of hashtables) called `edge_weights` such that `edge_weights[name1][name2]` provides the "weight" of character `name1`'s interactions with `name2` as the value in the key/value pair. Thus the total weight is calculated as:

```
total_weight = edge_weights[name1][name2] + edge_weights[name2][name1]
```

For the cosine similarity segment, I recycled code from previous assignments to create an empty hashtable of unigrams of the combined vocabulary of the top characters. Then, this hashtable was filled in with each characters particular vocabulary then flattened into an array. From this the cosine similarity between a vocabulary  $A$  and  $B$  could be measured using the equation:

$$\cos\_sim(A, B) = \frac{A \cdot B}{||A|| \times ||B||}$$

## Results and Discussion

For this project I explored social dynamics for the entirety of Game of Thrones. In addition, I individually explored Season 1, Season 1-3, 4-6, and Season 7. These graphs are available in the index. For now, let's cover the overarching social network of Game of Thrones. Below is the social network graph and a table with the social network metrics for the entirety of Game of Thrones. Note that the order of the characters in the table is determined by the number of sentences spoken by the character over the course of the series, implying Tyrion was the most vocal. This could also be considered a social metric of its own.

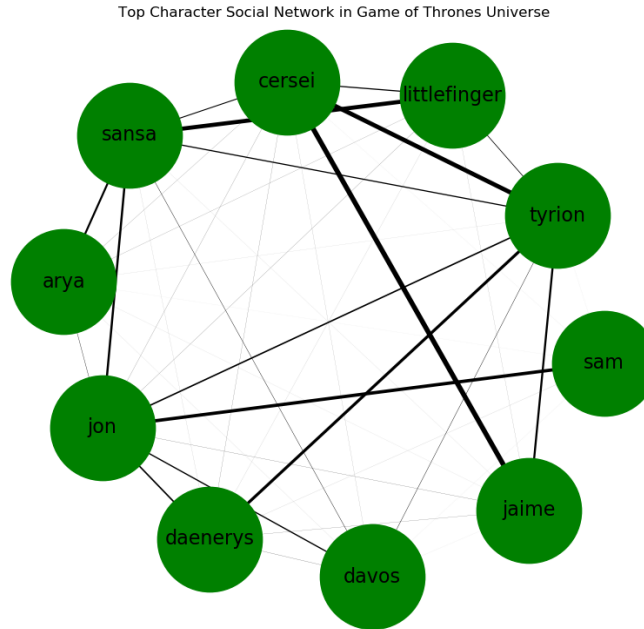


Table 1: Social Network Metrics for Game of Thrones Universe

	degree	eigenvector centrality	closeness centrality	harmonic centrality	eccentricity
tyrion	798	0.5098	73.6119	9.0	1
cersei	723	0.5289	66.4392	9.0	1
jon	633	0.2522	83.3513	9.0	1
daenerys	301	0.2174	61.4714	8.5	2
jaime	451	0.4049	55.5749	8.5	2
sansa	663	0.3158	73.4026	8.5	2
arya	158	0.0875	47.5630	8.5	2
sam	209	0.0934	60.9917	7.5	2
littlefinger	385	0.2528	63.8295	8.0	2
davos	137	0.0744	42.2334	8.5	2

In this case, degree is weighted, so rather than referring to the total number of connected nodes it refers to the total number of interactions with other "top characters" in the Game of Thrones universe. In each of these cases, centrality aims to identify the central character. In each case a higher value implies the character is more central. There's some discrepancies between the metrics on who the main character is. For example, Jon has the highest closeness centrality, is tied for the highest harmonic centrality, and has a reasonably high degree; however, his eigenvector centrality slips. I explored shifting the number of top characters from 10 to 15 to 20 and this irregularity did not shift. Tyrion, Cersei, and Jon are the only characters who spoke to each of the 10 top characters given their

centrality is 1.

Below in Figure 1 the cosine similarity graph for the vocabularies of the top characters and in Figure 2 is the cosine similarity graph for the filtered vocabularies of the top characters. The filtering in this project simply included the removal of stop words. The index will only include the unfiltered edition, as my intuition suggested stop words are relevant - the frequency and choice in stop words may differ depending on who you interact with and where you come from. On the other hand, a more heavily filtered vocabulary might be a more reasonable approach as one might be able to pick up on regional differences in speech, or references to particular characters, etc. Specifically

Figure 1: Cosine Similarity of Vocabularies in Game of Thrones

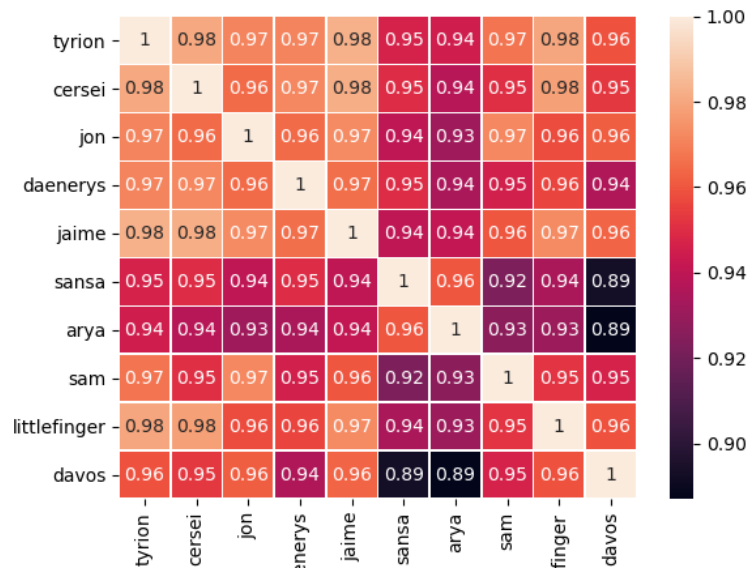
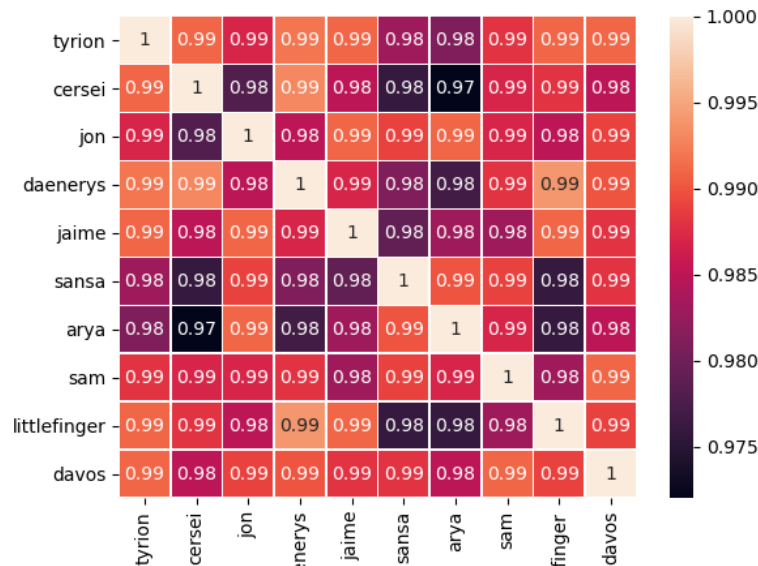


Figure 2: Cosine Similarity of Filtered Vocabularies in Game of Thrones



focusing on the unfiltered graph, I did find some results aligning with expectations. Tyrion, Jaime, and Cersei who are all siblings consistently have the highest cosine similarity. Similarly the sisters Sansa and Arya have the most similar vocabularies. While it may be shocking that Daenerys has the second highest cosine similarity with the bunch, most of her interactions are with Tyrion. While this is more my intuition, I'd also argue some character differences are at play. Both Daenerys and Cersei are queens and act like queens. A last observation is that while the unfiltered graph shows Jon having a rather dis-similar vocabulary from his sisters, the filtered graph seems to pick up a stronger similarity. Over the course of the show, Jon Stark rarely gets to interact with his sisters, but the filtered vocabulary might pick up the regional distinction of their families vocabulary from being born in the same region (the North).

## **Conclusion**

This project allowed me to explore various libraries and methodologies in social network analysis. While many of the specific findings could not be discussed in a reasonable amount of time for this project, the work was interesting both due to my personal interest in Game of Thrones and my academic and professional interest in social network analysis.

## Index

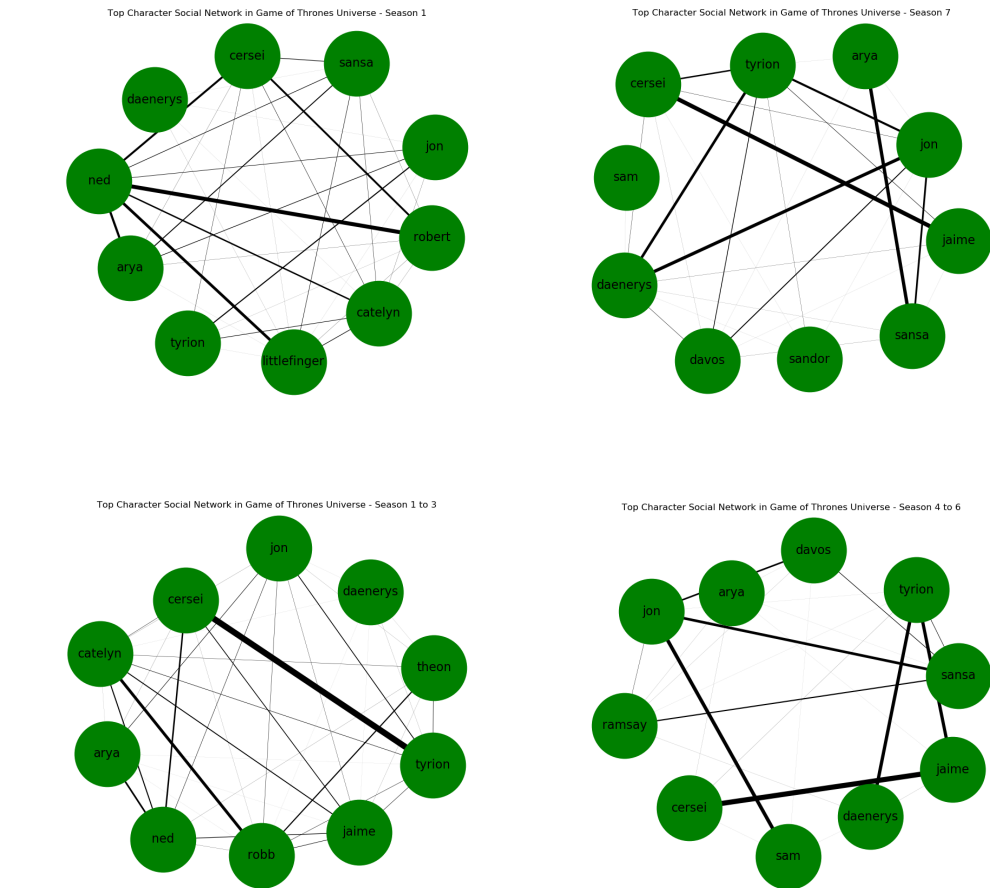


Figure 3: Cosine Similarity for Season 1-3 (left) and Season 4-6 (right)

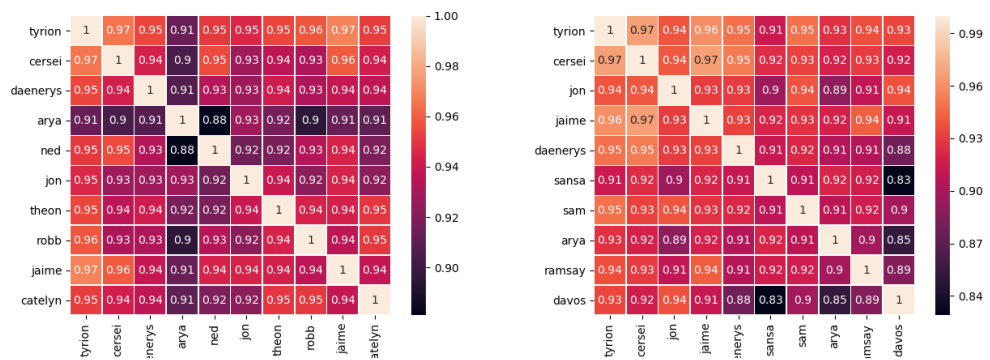


Table 2: Social Network Metrics for Game of Thrones Season 1-3

	<b>degree</b>	<b>closeness centrality</b>	<b>harmonic centrality</b>	<b>eccentricity</b>
<b>tyrion</b>	316	14.3197	8.5	2
<b>cersei</b>	326	14.5260	7.5	2
<b>daenerys</b>	9	2.8045	7.5	2
<b>arya</b>	88	12.2820	8.0	2
<b>ned</b>	202	14.5443	8.0	2
<b>jon</b>	89	10.2661	8.5	2
<b>theon</b>	79	10.4550	8.0	2
<b>robb</b>	200	12.1970	9.0	1
<b>jaime</b>	133	12.2633	9.0	1
<b>catelyn</b>	224	13.1864	9.0	1

Table 3: Social Network Metrics for Game of Thrones Season 4-6

	<b>degree</b>	<b>close_central</b>	<b>harmonic_central</b>	<b>eccentricity</b>
<b>tyrion</b>	221	9.4351	7.5	2
<b>cersei</b>	163	8.7172	6.5	2
<b>jon</b>	249	9.4974	7.0	2
<b>jaime</b>	259	9.0520	6.5	2
<b>daenerys</b>	107	8.7189	6.5	2
<b>sansa</b>	152	10.0122	7.0	2
<b>sam</b>	112	8.8153	6.5	2
<b>arya</b>	7	1.9948	7.0	2
<b>ramsay</b>	51	8.4498	7.5	2
<b>davos</b>	71	8.1040	7.0	2