

IDS 564: FINAL PROJECT

GROUP 7

Member Name	UIN
Disha Pandey	676191227
Mehul Gupta	677991579
Shreya Appana	657488017

GAME OF THRONES SERIES NETWORK ANALYSIS



1. INTRODUCTION

Game of Thrones (GoT) is an American fantasy drama T.V. series by David Benioff and D. B. Weiss for HBO, a classic adaptation of A Song of Ice and Fire, a series of fantasy novels by George R. R. Martin. It premiered on HBO in the US on April 17, 2011, and concluded on May 19, 2019, with 73 episodes broadcasted over eight seasons.

Set on the fictional continents of Westeros and Essos, GoT has many characters and follows several stories that run simultaneously. The first significant storyline concerns the Iron throne and the politics that surround it. The second storyline focuses on the last descendent of the realm's dynasty, who has been exiled to Essos and is plotting to reclaim the throne. The third and final

storyline follows the Night's Watch, a military order defending the realm against threats beyond Westeros's northern border.

The colossal complex network serves as an exemplary model for Social Network Analysis. Apart from being an excellent show, Game of Thrones provides a broad scope for analysis, covering the copious pool of characters. It presents a preeminent dataset for Data Visualisation. We can glean some great insights from analysing the dataset. For our research, we have looked at centrality measures, interpreting each character's role and how it affects the network. Intrigued by the notion of friends and enemies among the seven kingdoms, we used the community detection process to find the allies and foes of the characters. Since the whole idea behind the show dwells on the Iron throne and who will be the last one standing to reign over the seven kingdoms, we were motivated to enlighten the GoT fans about the hidden statistics behind the web of thrones. And we want to use our analytical skills to understand complex networks better.

2. RELATED WORK

We are not the only ones who were intrigued and motivated to perform social network analysis on popular television dramas. We can find precedent work on the complex Game of Thrones network analysis. Fans have used different visualisation tools like Graphia and Gephi to visualise the network ^{[1][2]}. Many have worked on the television dataset, while some prioritised books for their predictions. Game of Thrones enthusiasts has used Natural Language Processing (NLP) and Machine Learning techniques on various datasets for predictive analysis. NLTK and word cloud are a few prominent packages used among the TV series aficionados. Machine learning predicts the character most likely to die in Game of thrones, an intriguing way to use predictive analysis by Logan Booker ^[3].

3. DATASET

While analysing from throughout eight seasons, we used sixteen files, eight files each of nodes and edges. Edges consist of 4 columns depicting Source, Target, Weight, and Season number, nodes connecting the character with their aliases. When we talk about nicknames or pseudonyms, Game of Thrones provides plentiful characters with these attributes. For example, Daenerys Targaryen is called Queen of Stormborn, Queen of Andals, and Dany by her lover Jon Snow. While data pre-processing, the crucial problem faced was connecting these alias's names to each character. The edges file was distributed over a total of 4,110 edges, and the nodes file was spread over a total of 968 nodes over eight seasons.

We have analysed and visualised a few findings using two different CSV files, which comprise all the books' dialogues. Thanks to Alben Tumanggor for providing the data for analysis purposes.

> nodes		
	Id	Label
1	ADDAM_MARBRAND	Addam
2	AEGON	Aegon
3	AERYS	Aerys
4	ALLISER_THORNE	Allister
5	ARYA	Arya
6	ASSASSIN	Assassin

Fig 1. Node Values

> edges				
	Source	Target	Weight	Season
1	NED	ROBERT	192	1
2	DAENERYS	JORAH	154	1
3	JON	SAM	121	1
4	LITTLEFINGER	NED	107	1
5	NED	VARYS	96	1
6	DAENERYS	DROGO	91	1

Fig 2. Edge Values

4. NETWORK STATISTICS

We initially visualise the overall books dataset to create a network graph. We are using Gephi for our visualisation and updating the graph. We removed edges with insignificant weights. The comprehensive network graph looks like the snippet below. Our visualisation uses Fruchterman Reingold Layout with parameters set for best visualisation.

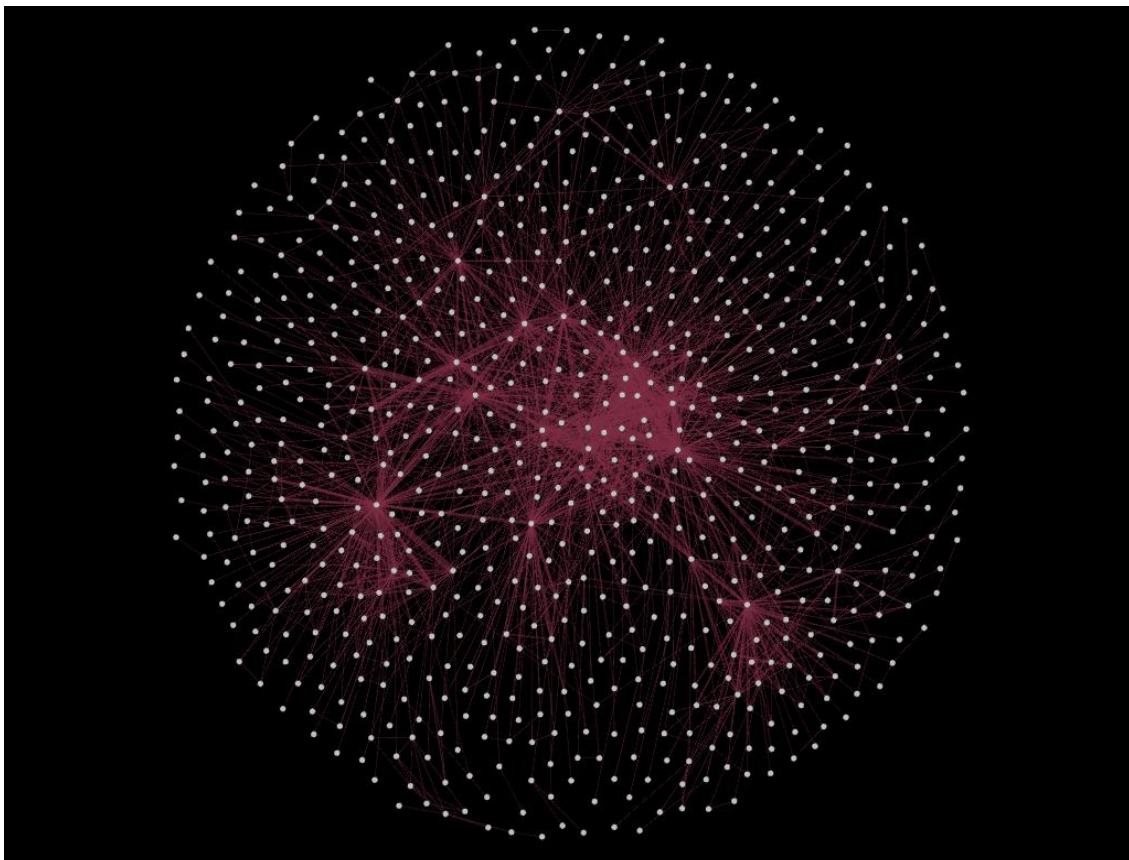


Fig 3. Network Graph using Fruchterman Reingold Layout

Our observations show that Jon Snow is one of the most crucial characters as many edges go in and out of the node. (Refer Fig 4.)

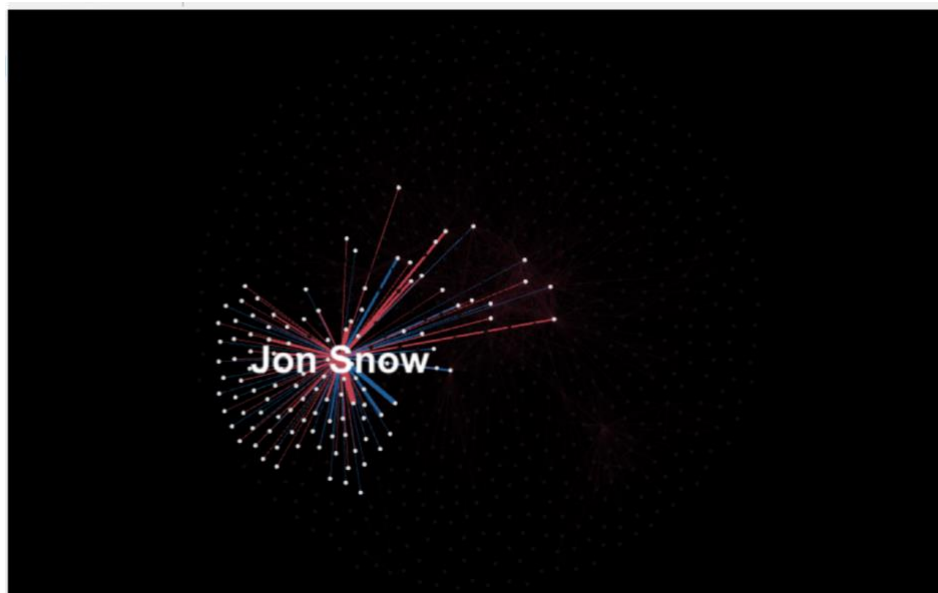


Fig 4. Network Graph for Most Crucial Character

Furthermore, giving us one of least essential characters, Rickon stark. (Refer Fig 5.)

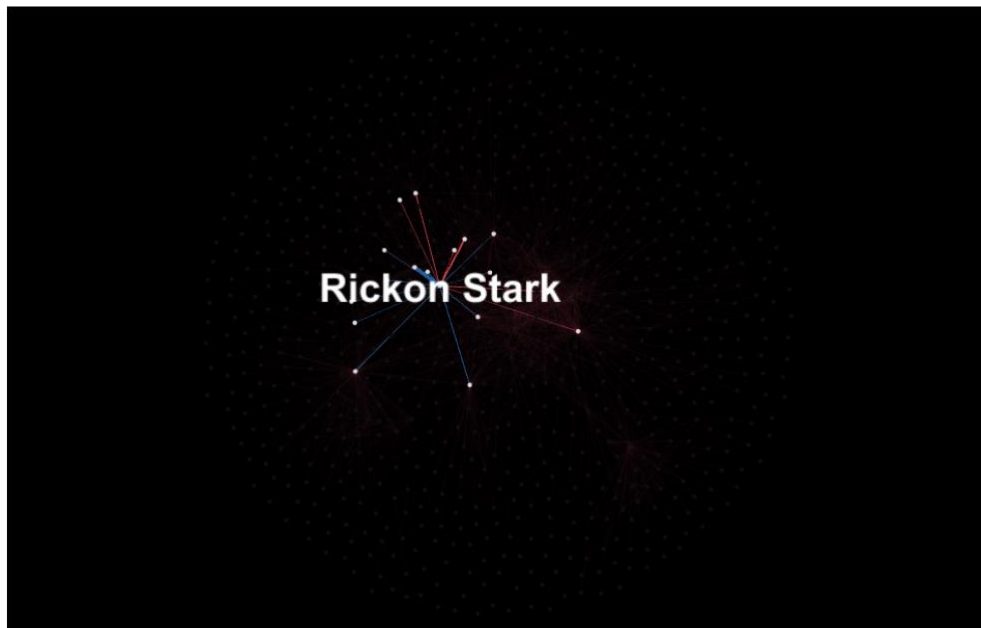


Fig 5. Network Graph for Least Crucial Character

Further, working on the Dplyr package for Exploratory Data Analysis, we discovered the most crucial house, which has been coined the most times. Our analysis (Refer Fig 6.) gleans that House Stark takes the pole position, and Lannister follows it to second place.

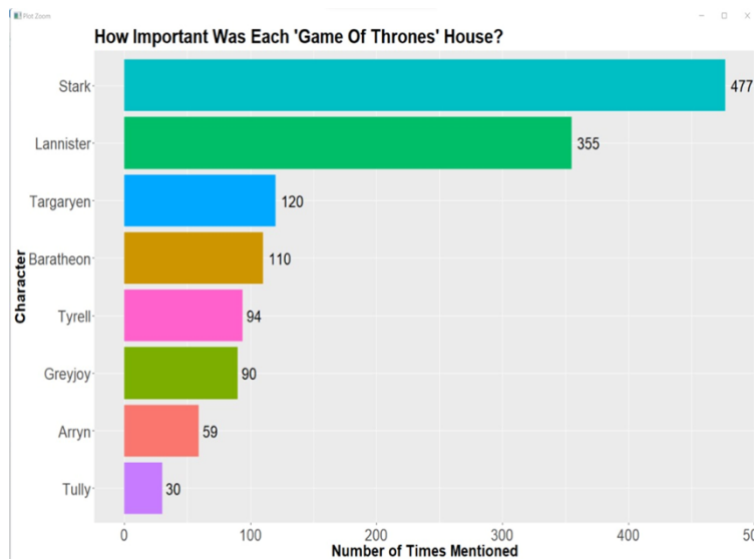


Fig 6. Bar Graph representing Important Houses

We were intrigued to find the crucial characters, using EDA to see the character that was mentioned numerous times throughout the seasons. (Refer Fig 7.) Results are not shocking; Jon Snow, with the most mentions, becomes the crucial character for this network, as the network is hugely affected by this character.

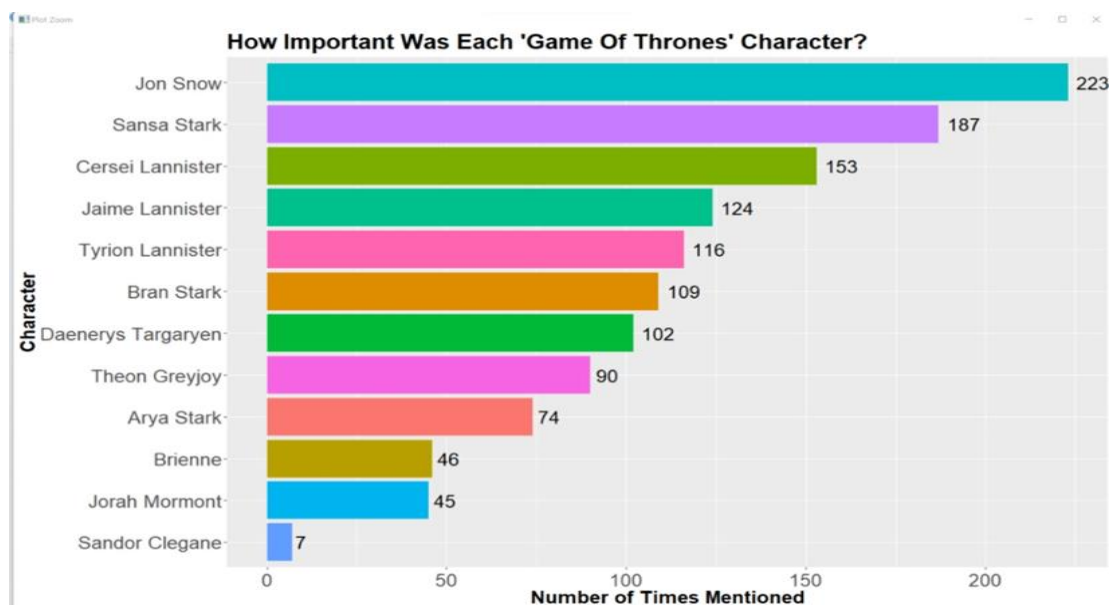


Fig 7. Bar Graph representing Important Characters

Fig 8. Summary Statistics

5. SUMMARY STATISTICS

Further deep diving into the statistics behind the network graph, we found the average degree to be 7.093, with a network diameter of 9.0 and an average path length of 3.416. An exciting study comes to light about the average cluster coefficient, values at 0.635. A value of 0.635 means that approximately 63.5% of connecting triples are closed to form triangles.

Filters	Statistics	×	—
Settings			
<input checked="" type="checkbox"/> Network Overview			
Average Degree	7.093	Run	?
Avg. Weighted Degree	81.982	Run	?
Network Diameter	9	Run	?
Graph Density		Run	●
HITS		Run	●
Modularity	0.567	Run	?

6. CENTRALITY MEASURES

Like any other television drama, Game of Thrones has many character developments. We have tried to align these developments with the centrality measures. Centrality measures are essential in analysing and extracting intuitive insights from the network. The most straightforward **Degree centrality** tells us about the network’s most impactful or central node. To our findings, in the overall network, Jon Snow, Daenerys Targaryen, Tyrion Lannister, and Arya Stark are among the top-rated. Here’s a snippet of the visualisation; the node sizes are based on degree centrality value.

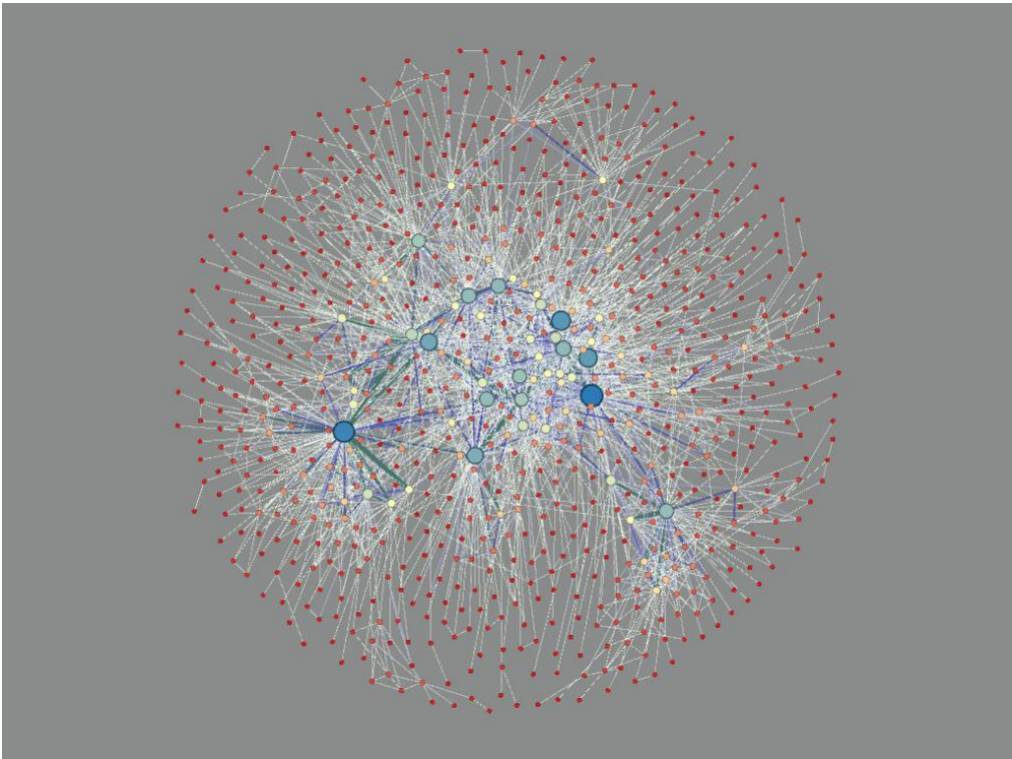


Fig 9. Degree Centrality



Fig 10. Degree Centrality of Complete Books

Furthermore, we have used our prowess to find some connections among characters throughout the eight seasons. Our results showed that Tyrion is the most impactful character throughout the show. His use of wit and intellect to overcome prejudices contributed a great deal to the storyline. Further, Ned has been the next impactful character from season 1 to season 4 despite dying in the first season. His posthumous mentions signify his impact throughout the series. Taken over by Sansa in season 4 and remained constant till season 8. The results match perfectly with the character development of Sansa from season 1 to season 8.

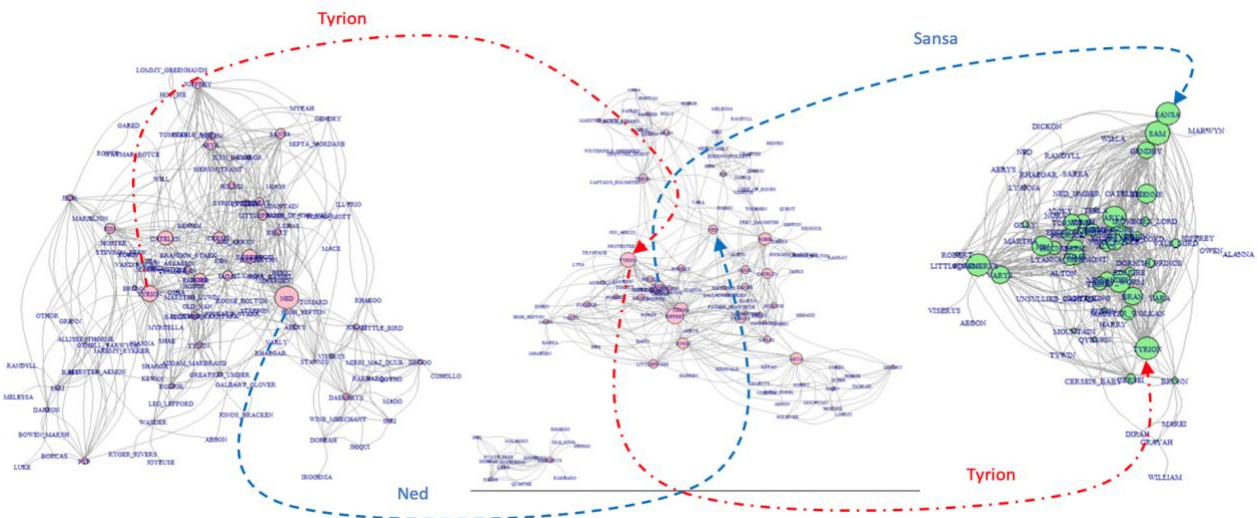


Fig 11. Degree Centrality across all 8 seasons

Second (Refer Fig 11.), we want to focus on the most potent or well-connected person or family. To calculate this, we use the **EigenVector** Centrality measure, which measures the transitive influence of nodes. We see **RED** nodes with the highest eigenvector scores for the overall dataset.

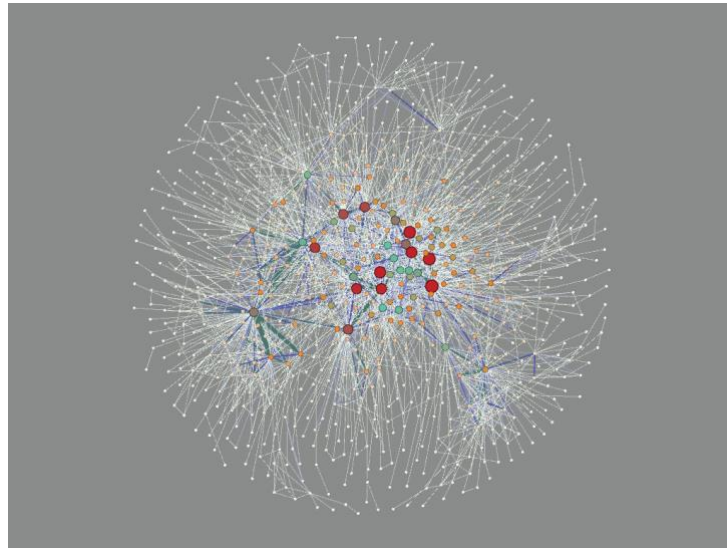


Fig 12. Eigen Vector Centrality of Complete Books

Cersei Lannister, one of the most feared characters in Game of Thrones, leads the charge with the most powerful connections in the land of the seven kingdoms. Her prominent relationships influence her powers over the land of Westeros and possibly explain why she was ruling and sitting on the Iron Throne for much of seasons 7 and 8.



Fig 13. Network Graph for Character with Most Powerful Connections

When we individually work on a per season dataset, we find the similar results and predict accurately that Cersei Lannister has been well connected throughout the series. As an exciting finding comes to light, Ned Stark, who died in the first season, has posthumous mentions even in season 4. He's been an integral part of George R. R. Martin's book, as he has seen and fought numerous battles, was named King of The North and was an advisor to the King of Westeros.

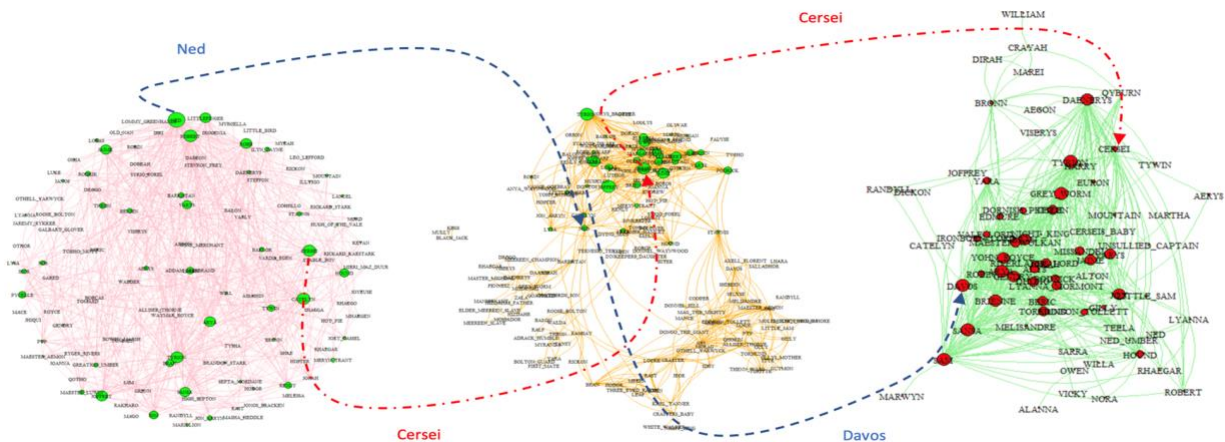


Fig 14. Eigen Vector across all 8 Seasons

Third, (Refer Fig 15.) our focus shifted to **Closeness** Centrality because we were hoping to find the most prevalent node, as the node that can spread information very efficiently through the graph. As expected, characters like Davos Seaworth, Petyr Baelish, and Lord Varys are among the few with excellent closeness scores, as their sources are readily available. And transferring information throughout the graph with efficiency is effortless for them.

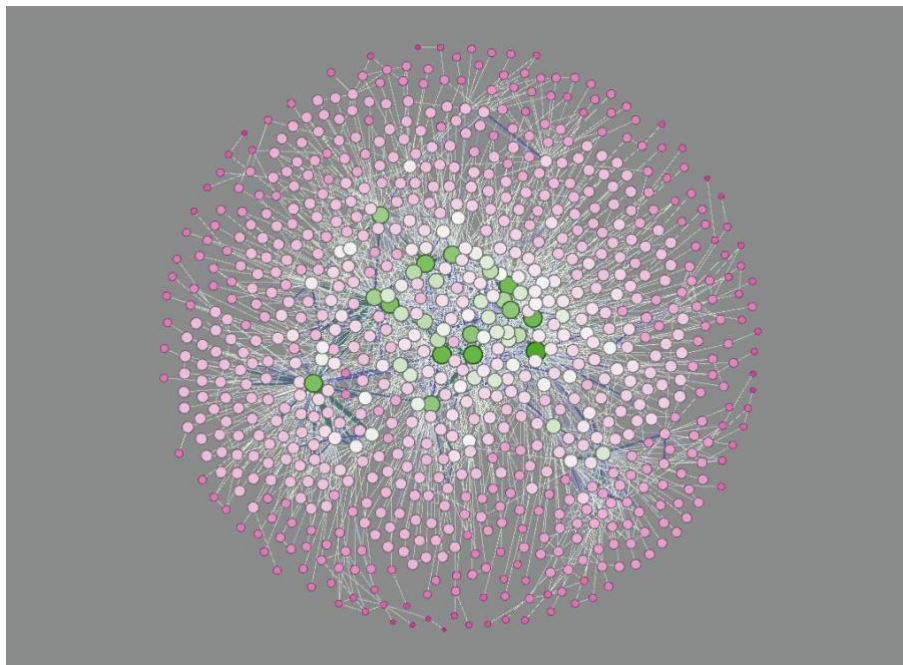


Fig 15. Closeness Centrality



Fig 16. Character with highest Closeness Centrality

From Fig 16., Lord Petyr Baelish, also known as Little Finger, has a fantastic ability to transfer information expeditiously.

While using the same centrality measure over the dataset created by the complete books. We find that Catelyn Stark has played an essential role in transferring information around the network over the first four seasons as she travelled all around the world of Game of Thrones. Look who we have here for the silver medal, Tyrion Lannister.

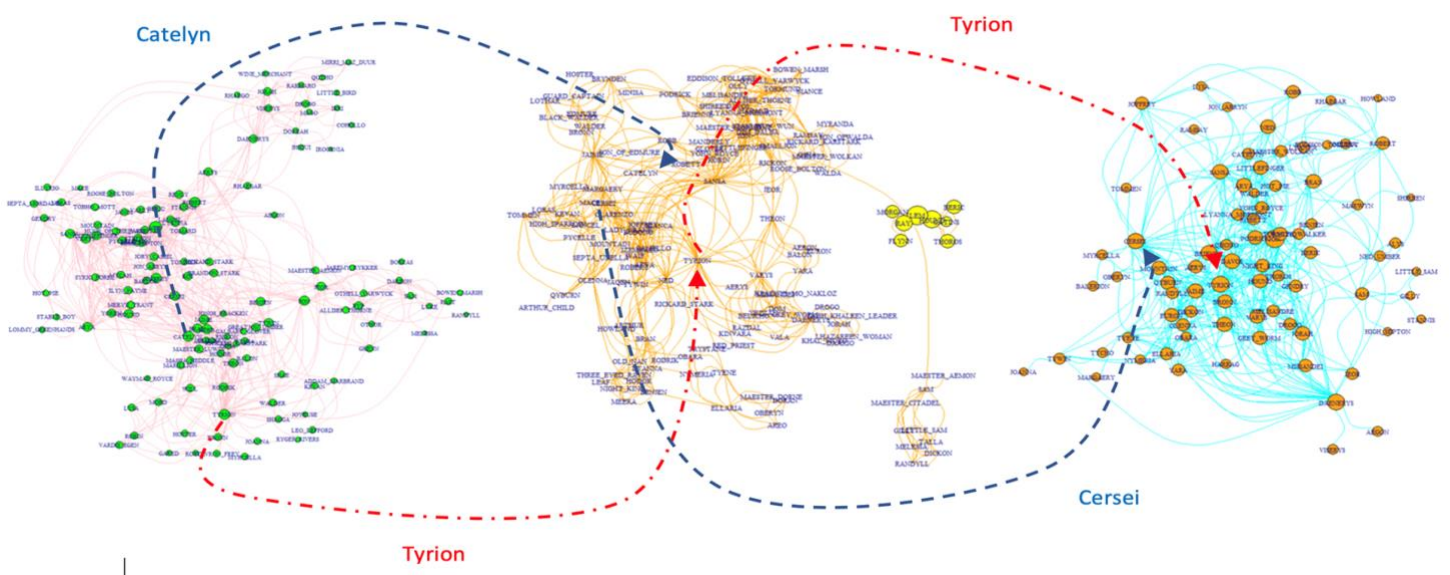


Fig 17. Closeness Centrality across all 8 Seasons

Fourth, we tend to look at the influence of each node on the flow of information in the network. **Betweenness** centrality plays a vital role in selecting these nodes—the higher the betweenness scores, the greater the influence of each character over the flow of information. Jon Snow and Tyrion Lannister are two red highlighted nodes with the highest betweenness scores.

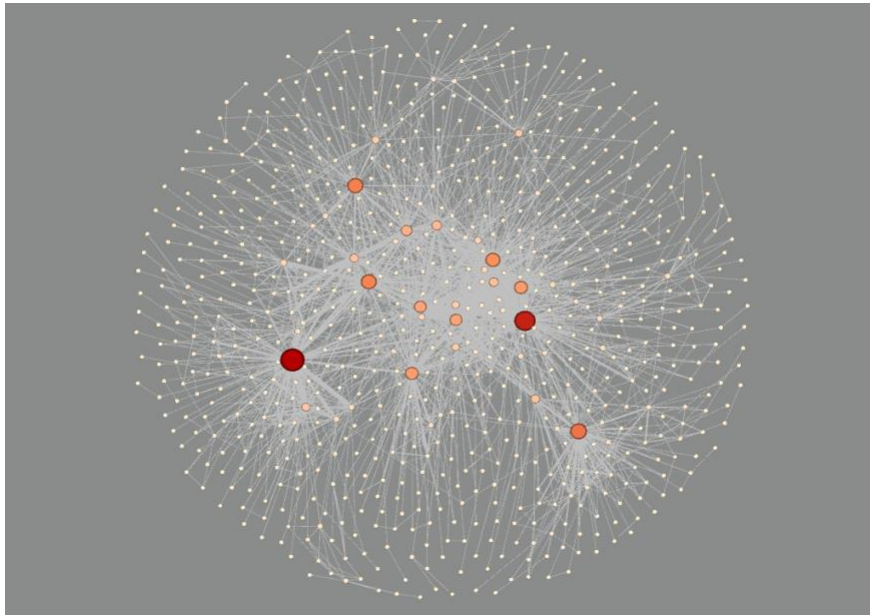


Fig 18. Betweenness Centrality

However, it was astonishing to see Theon Greyjoy with a notably high betweenness score. Despite not being categorised in the pool of major characters. He has a decent influence on the flow of information.



Fig 19. Character with highest influence on Information Flow

7. COMMUNITY DETECTION

In a complex network like Game of Thrones, several characters fight to be the last man or woman standing to reign over the seven kingdoms. And a true king knows there is no winning in war unless you have allies or comrades. We used community detection algorithms to identify these allies. Two algorithms that we considered are Fast Greedy and Louvain Cluster. The approach of the Fast Greedy algorithm works on modularity and creates a subgraph composed of links between highly connected nodes. As mentioned earlier, we compared the plots of seasons 1, 4, and 8.

Daenerys Targaryen has been a crucial part of various communities throughout eight seasons. She started her journey by escaping the lands of Westeros to Essos, as she was the last breathing Targaryen, and the former King Robert Baratheon was hunting down all Targaryen's. As the season progresses, she builds herself an army and moves to Westeros to attack its capital, King's Landing. Her community grows as we come to the end of season 8. Here is the visual representation of the Fast Greedy Community Detection from seasons 1-4-8.

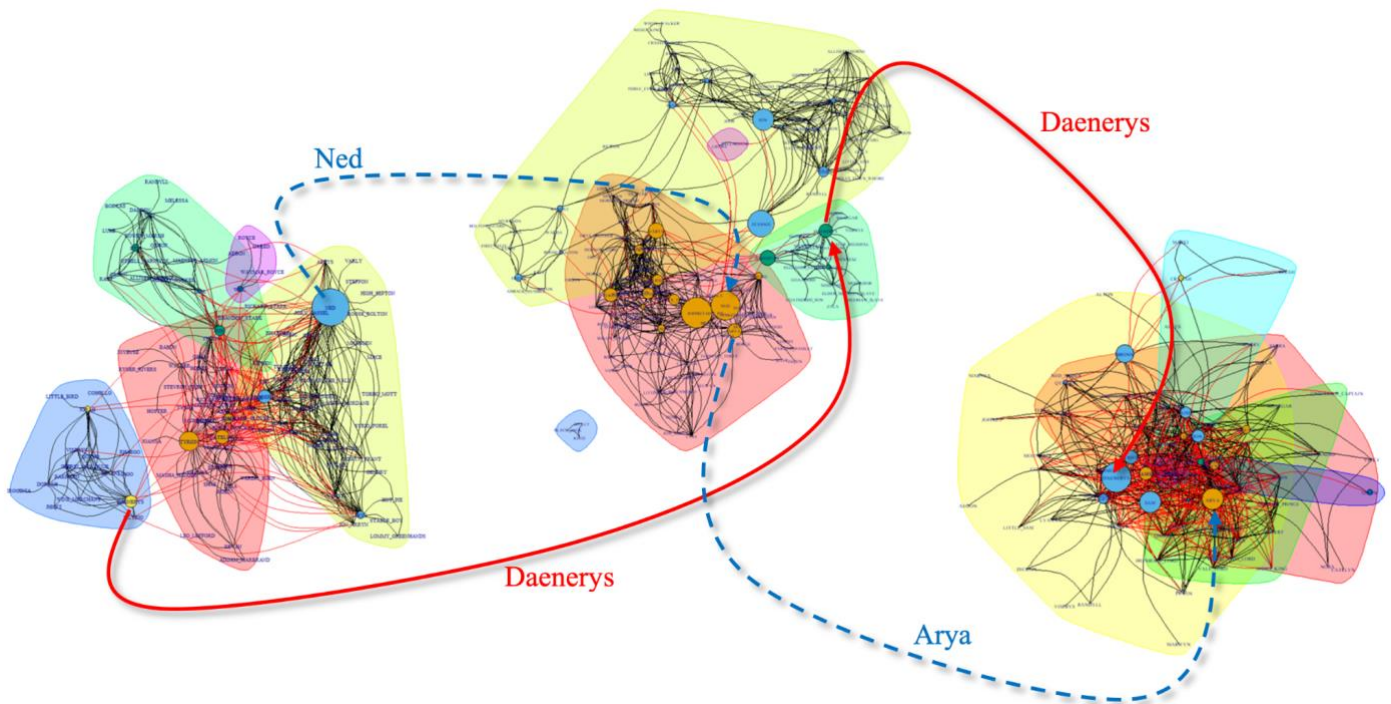


Fig 20. Fast Greedy Community Detection across Season 1,4,8

Similar results can be seen from the visual representation of Louvain Cluster Community Detection.

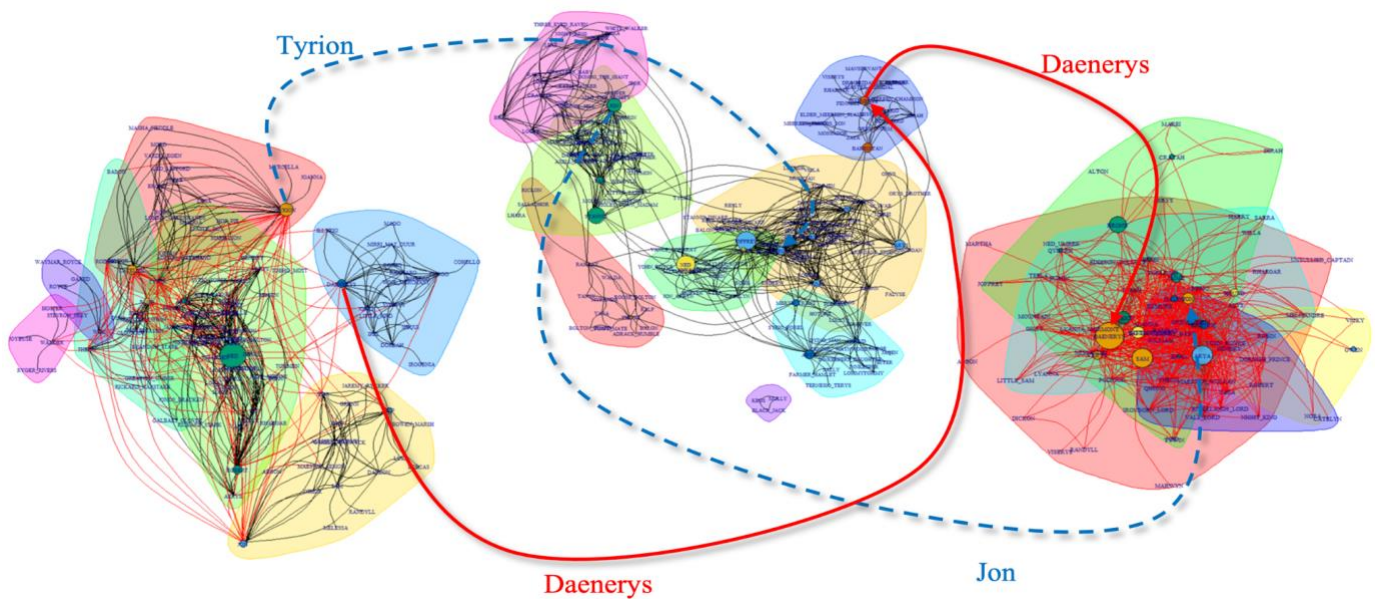


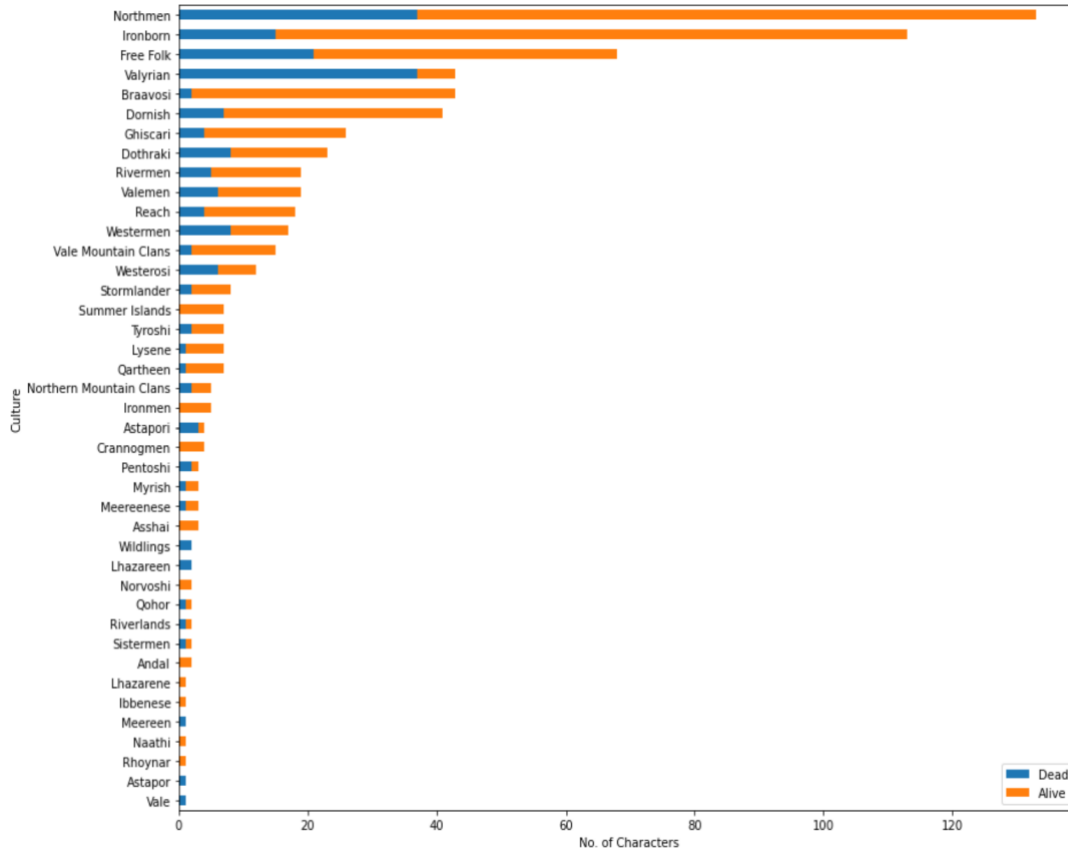
Fig 21. Louvain Cluster Community Detection across Season 1,4,8

8. ANALYSIS / FINDINGS

Several wars have been orchestrated in the lands of seven kingdoms. Some Kings wanted to conquer and reign over the whole land, and there were some Kings, like Mad Kings, who started wars because “Some Whispers in his head would ask him to do so”. But who are the winners here? Which region, part of Westeros, has the most survivors?

In our interest, we looked at the survivors and learned more about their culture and which land they represent. We used the dataset taken from Kaggle; the CSV files comprise information about the battles held, location, number of survivors, characters who died, the year of the wars, etc. From our analysis, (Refer Fig 22.) we figured out that men from the North (ruled by the Starks) have the highest survival rate and are more combat. Here is the visual representation to prove our findings.

Fig 22. Bar Graph for Survivors



9. CONCLUSION

Social Network Analysis provides many statistical tools and methods, and we tried to cover them to a great extent in our research. Centrality measures and Community detection have been our main topics of study. To conclude, our analysis comprehends a lot about the complex network of “Game of Thrones”. Moreover, various centrality measures aid us in interpreting who is the most impactful or most powerful character in the world of GoT. Tyrion Lannister, Jon Snow, Daenerys Targaryen, Ned Stark, and Cersei Lannister are among the pool of significant contributors with high scores in different centrality measures. After exploratory data analysis, we conclude that Starks are the predominant family in GoT, followed by the Lannister’s. And Jon Snow is the prominent character. Character development is a cliché in any television series, and as the season progresses, we find several characters have grown, and community detection can interpret this. Daenerys and Jon started with a few allies, and by the end, they became crucial characters and had numerous supporters.

We wanted to explore a question about how culturally different people in the various parts of Westeros and Essos are. We studied the survival rate from the remnants of several wars held in GoT and found that people from the North, ruled by House Stark, have the highest survival rate, representing better combat skills. We can elucidate ourselves by analysing more about the people’s cultures in GoT using NLP for future work.

10. REFERENCES

- [1]. https://medium.com/@sanchit_kumar/my-experience-with-gephi-visualizing-game-of-thrones-a-song-of-ice-and-fire-61587eefba9d
- [2]. <https://medium.com/mlearning-ai/combining-nlp-features-with-3d-network-visualisations-a-game-of-thrones-analysis-9f6a3105eba6>
- [3]. <https://www.gizmodo.com.au/2017/07/machine-learning-predicts-characters-most-likely-to-die-in-game-of-thrones/>
- [4]. <https://www.kaggle.com/datasets/mylesoneill/game-of-thrones>
- [5]. https://medium.com/@weatheranchor_43165/game-of-thrones-network-анализ-52b69e13377a
- [6]. <https://medium.com/@yangyuesongyys/network-analysis-on-game-of-thrones-8b8b7dd71945>
- [7]. <https://pub.towardsai.net/social-network-analysis-of-game-of-thrones-in-networkx-ff21ef65dc11>