UE20CS344 – Network Analysis and Mining
Course Project

Project Title : Information spread and Influence Determination in social networks and complete network analysis(GOT)

Team member 1 SRN and name: PES2UG20CS139 HEMACHANDRAN S
Team member 2 SRN and name: PES2UG20CS181 M MANONMANA UDUPA
Team member 3 SRN and name: PES2UG20CS180 LOHIT SIRIKI
Team member 4 SRN and name: PES2UG20CS162 KHUSHI SATISH
(SECTION C)

Topic

- Problem statement

With the increasing popularity of social media platforms, it has become easier for information to spread rapidly, which can have both positive and negative effects. However, it is not well understood how information spreads in social networks and how it influences people's behavior. This project aims to address this gap by analyzing large-scale social network data to identify the patterns and mechanisms of information dissemination. The project will also investigate how user behavior and community structures contribute to the spread and influence of information, which can have significant implications for understanding and managing the impact of social media on society.

Uniqueness

Why you think this project is interesting or unique

This project is interesting and unique because it combines multiple fields such as network science, data analysis, and social psychology to study the complex phenomena of information spread and influence in social networks. By examining large-scale social network data, the project aims to provide insights into the mechanisms of how information spreads and influences people's behavior, which can have significant implications for areas such as marketing, public health, and politics. Furthermore, the project's findings can help develop strategies to manage and control the spread of misinformation and disinformation, which is becoming increasingly prevalent in the digital age. Overall, this project has the potential to contribute significantly to the understanding of the interplay between social networks, information diffusion, and human behavior.

Uniqueness

Importance: Social networks play a significant role in our daily lives, and the way information spreads through them can have a profound impact on society.

Relevance: With the proliferation of social media platforms and the ease with which information can spread, the issue of information dissemination has become more important than ever

Innovation: The project's innovative use of large-scale social network data and analysis techniques can provide new insights into information dissemination, which can potentially be used to develop more effective strategies for managing information flow on social networks.

Dataset

Data: https://github.com/Manonmana/NAM-Project

(Link has been provided in 11.txt file because the data was too huge to compress into a zip file)

The data set book1-edges.csv contains the edges (i.e., connections) between characters in the book "A Song of Ice and Fire" by George R. R. Martin. The attributes of this data set are:

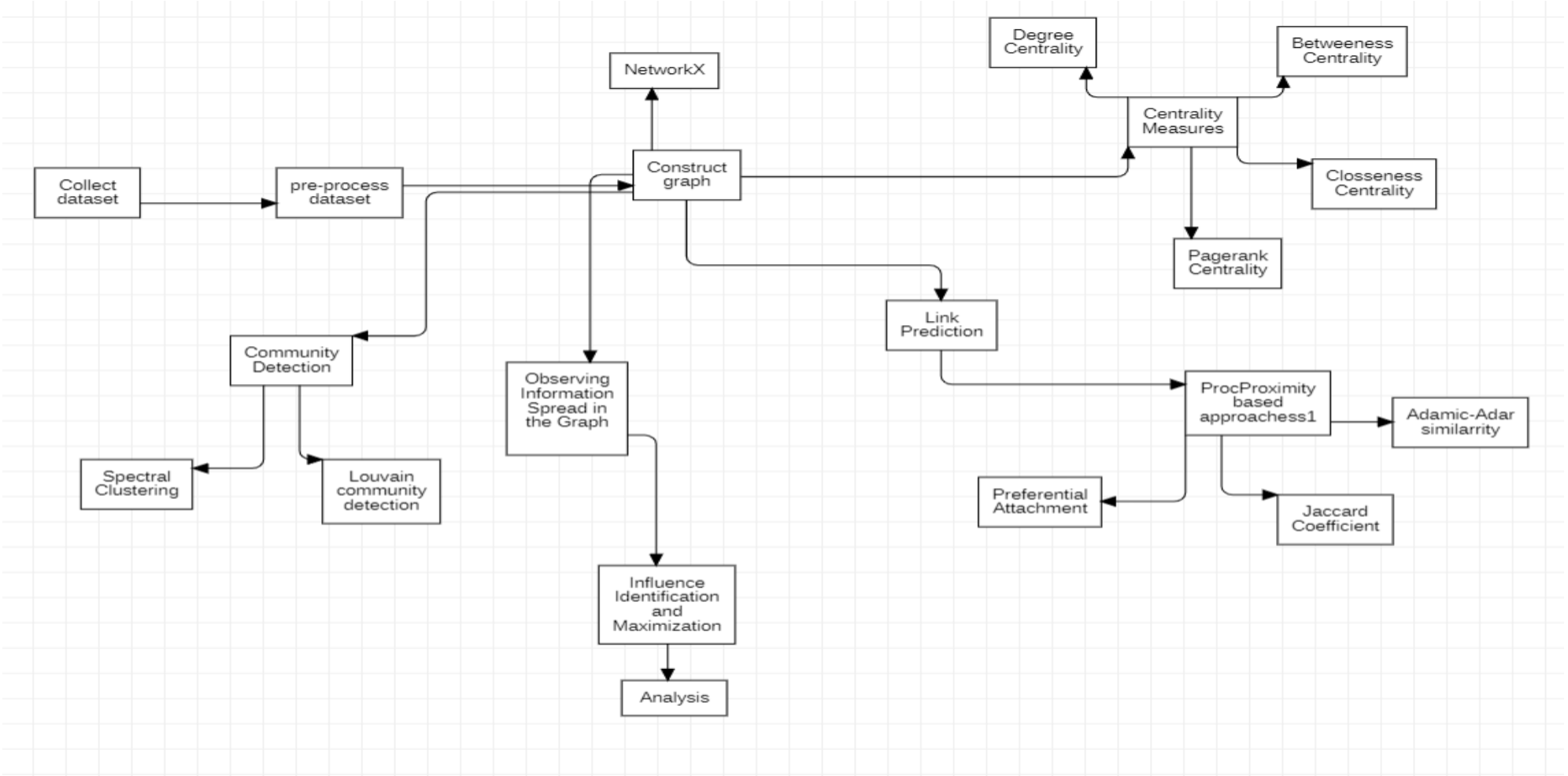
Source: the name of the character who is the source of the connection (edge)
Target: the name of the character who is the target of the connection (edge)
Weight: the weight of the edge, which represents the number of interactions or mentions between the characters in the book
Type: the type of interaction between the characters, which can be either "undirected" or "directed" depending on the nature of the relationship between the characters
Book: the name of the book in the series in which the interaction occurred (in this case, it is book 1)
POV: the point of view character from whose perspective the interaction was witnessed (if any)
These attributes provide important information about the relationships and interactions between the characters in the book, which can be used for network analysis and visualization.

Dataset

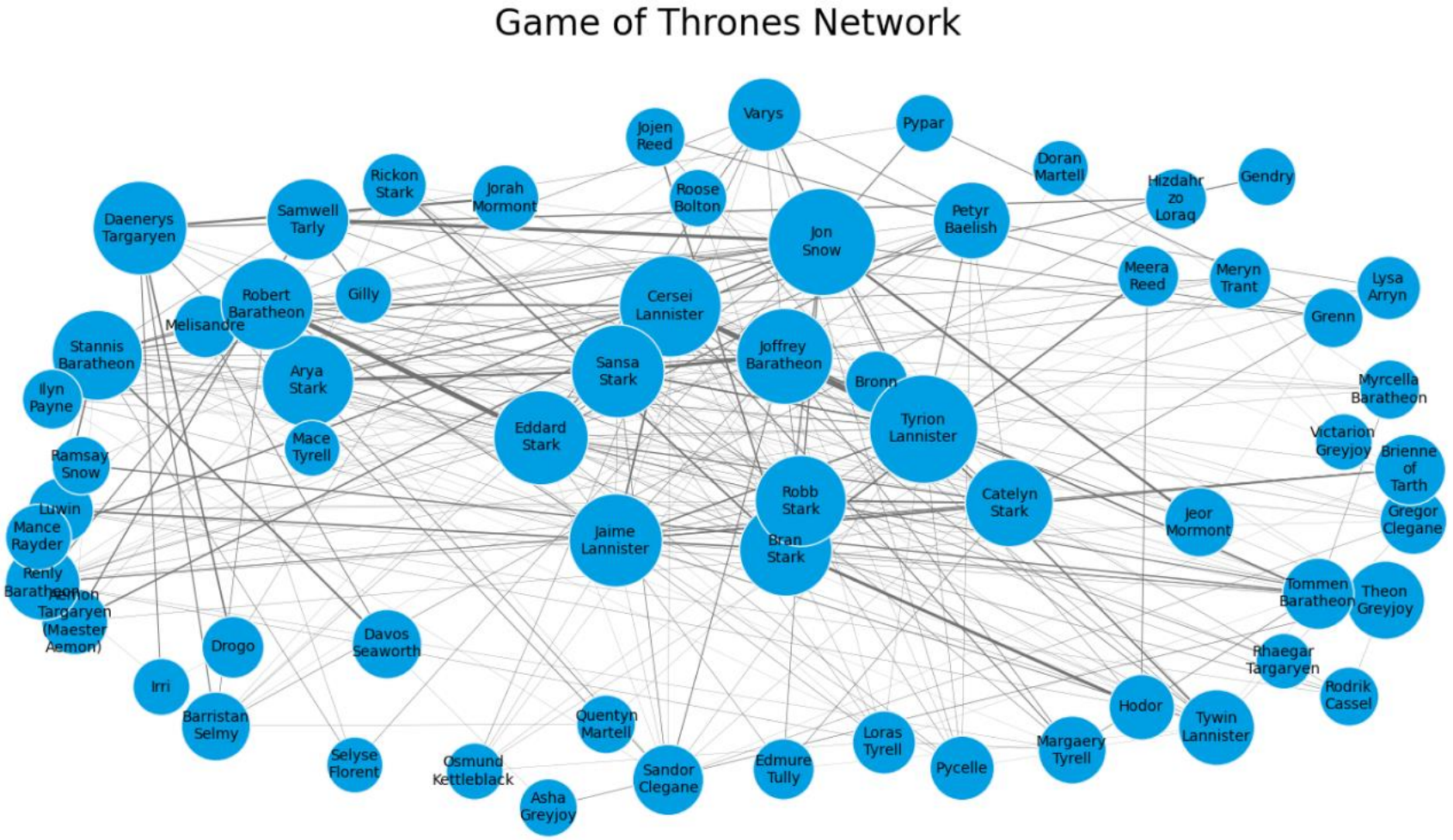
The data set book1-nodes.csv contains the following attributes:

Source: This represents the source node or character in the relationship between characters.
Target: This represents the target node or character in the relationship between characters.
Type: This specifies the type of relationship between the source and target characters, such as allies, family, or adversaries.
Weight: This indicates the strength or intensity of the relationship between the source and target characters. It is based on the number of interactions or mentions of the characters together in the text.
Book: This specifies the book in which the relationship between the source and target characters is mentioned.
book_count: This represents the number of books in which the relationship between the source and target characters is mentioned.
Overall, the data set provides information about the relationships between characters in the popular book series "A Song of Ice and Fire" by George R.R. Martin. The data set can be used to analyze the social network of characters in the book series, to identify communities of characters, and to understand the dynamics of character relationships throughout the series.

Overall design or approach in a free hand diagram



Graph Structure



Final results so far

- Any metric ?
  - ROC graph has been used to identify the best method for link prediction.

- Any visualisation ?
  - Graphs have been plotted and information flow through the graph is visualized.
  - Different communities in the graph have been visualized.
  - Evolution of the characters throughout the story has been visualized.

Quantity and quality of work

no	Code functionality	% Complete	Runs without problem (Y/N)	If there are minor issues, indicate	Individual Contribution(SRN)
1.	Load required libraries , dataset , install the libraries and construction of graph .	100 %	Yes	None	PES2UG20CS180
2.	Community detection and different centrality measures.	100%	Yes	None	PES2UG20CS162
3.	Influence identification and maximisation . Observing information spread in graph and link prediction.	100%	Yes	None	PES2UG20CS181
4.	Analysis , answering questions related to the story and correlation between different measures .	100%	Yes	None	PES2UG20CS139

Top few learning

Serial No	Top learning in this project
1.	Network analysis can help identify influential nodes: Analysing network data can help identify individuals with the highest potential to influence others. Various metrics, such as centrality measures and clustering coefficients, can be used to determine the most critical nodes in a network .
2.	Influence is not always correlated with popularity: In social networks, some individuals may have a large number of followers or friends but may not necessarily have significant influence **Checking evolution of a character throughout the story** .
3.	Understanding user behaviour is key to designing effective interventions: To create interventions that can curb the spread of misinformation or promote the spread of accurate information, it is essential to understand how users interact with information in a network. Identifying common patterns of behavior can help design targeted interventions that are more likely to be effective

Top unresolved challenges

Serial No	Brief description of unresolved challenges	Type of challenge (scope/data/design/implementation / others)
1.	Node classification - Node classification is only as effective as the information available about the nodes. If the dataset is biased then may not provide accurate results . The classification of nodes can be context-dependent. Nodes may be classified differently based on the network structure	Data set is biased and is context- dependent . So couldn't implement node classification .
2	Deciding fate of an individual in the story .	Inadequate information given in the dataset .

Reference papers

No	Paper Title	Authors
1.	Virality prediction and community structure in social networks .	Weng, L., Menczer, F., and Ahn, Y.Y. (2013). Scientific Reports, Vol. 3, Article 2522.
2.	An Empirical Study of the Spread of News on Digg and Twitter Social Networks	Lerman, K., and Ghosh, R. (2010). Information Contagion: Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM '10), pp. 90-97.
3.	The Role of Social Networks in Information Diffusion.	Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. (2012). Proceedings of the 21st International Conference on World Wide Web (WWW '12), pp. 519-528.
4.	Small but Slow World: How Network Topology and Burstiness Slow Down Spreading.	Karsai, M., Kivelä, M., Pan, R.K., Kaski, K., Kertész, J., Barabási, A.L., and Saramäki, J. (2016). Physical Review E, Vol. 93, No. 3, Article 032315.