Games of Thrones

Mattia Buzzoni, Artificial Intelligence Master Degree, ID number (0001145667)
Leonardo Mannini, Artificial Intelligence Master Degree, ID number (matricola)
Mirko Mornelli, Artificial Intelligence Master Degree, 0001113084
Riccardo Romeo , Artificial Intelligence Master Degree, ID number (matricola)
Diego Rossi, Artificial Intelligence Master Degree, 0001138652

1 Introduction

Literature is a broad and multifaceted field that reflects the complexity of human society and imagination. Among its many subdomains, narrative fiction—especially epic fantasy—offers rich ground for structural and character-based analysis. The series A Song of Ice and Fire, written by George R. R. Martin, represents one of the most intricate examples of contemporary fantasy literature. Its dense network of characters, shifting alliances, and multilayered plots makes it a fascinating subject of both literary and computational investigation. This project falls within the scope of digital literary studies, and more specifically in the intersection between narrative analysis and network science. The work focuses on the social structures embedded in A Song of Ice and Fire, analyzing character interactions across the five published books of the saga. These volumes present an ever-expanding universe populated by hundreds of characters, each with a role in a larger political and personal drama.
To uncover the underlying structure of these character interactions, we apply Social Network Analysis (SNA) techniques, which allow us to model the narrative as a graph where characters are nodes and their interactions are edges. By examining these networks using computational tools, we aim to gain quantitative insights into the narrative architecture of the saga. In particular, we investigate the evolution of character importance, community structures, and connectivity patterns over the course of the five books.
Through this interdisciplinary approach, the project seeks not only to explore the rich world of these books from a novel analytical perspective but also to validate the use of network-based methodologies in the study of complex literary works.

2 Problem and Motivation

This project investigates the underlying structure of character interactions within the fantasy saga *A Song of Ice and Fire*, focusing on the first five books of the series. The central aim is to explore how characters are positioned within the social landscape of the narrative and to assess their importance across different stages of the story. Unlike many traditional narratives with a single protagonist, this saga features a complex and distributed set of characters whose relevance varies considerably over time and context.
The problem we seek to address is the difficulty of identifying and evaluating character centrality in such a fragmented and multi-perspective storyline. In a literary universe composed

of hundreds of characters, each with their own alliances, roles, and trajectories, determining who the key figures are—and how these figures influence or connect with others—presents both a narrative and analytical challenge. This is particularly important in a work like *A Song of Ice and Fire*, where shifts in power, betrayal, and character deaths play a crucial role in the progression of the plot.

The motivation for this work lies in the potential of structural analysis to offer a deeper and more systematic understanding of narrative dynamics. By investigating which characters act as central nodes, which form cohesive groups, and how the overall structure evolves from one book to the next, we aim to provide insights that complement traditional literary interpretation. This approach enables a more objective reflection on character importance, the density of relationships, and the organization of fictional societies.

The main contribution of the project is to map and analyze the evolving network of character interactions across the five books, highlighting patterns of centrality, influence, and group formation. Through this, we hope to shed light on the complex mechanisms of storytelling in epic fantasy and to demonstrate how structural perspectives can enrich our understanding of literary worlds.

3 Datasets

The dataset used in this project is a publicly available resource from Kaggle, titled Game of Thrones Network. It provides a structured representation of character interactions across the first five books of the series. Each file corresponds to a specific book and contains a list of interactions between characters, represented as pairs of names. An interaction is defined as the co-occurrence of two characters within a 15-word window in the narrative text. While this methodological choice may appear somewhat arbitrary, it is a widely adopted approach in co-occurrence analyses and allows for the inclusion of a broad range of both direct and indirect interactions.

	Book 1	**Book 2**	**Book 3**	**Book 4**	**Book 5**	**All Books**
Nodes	187	259	303	274	317	796
Edges	684	775	1008	682	760	2823

For each book, a social network was constructed using the Python library networkx, with preprocessing steps applied to standardize character names and consolidate interactions. The resulting networks are undirected and unweighted graphs, where nodes represent characters and edges represent their narrative co-occurrences.

Although the dataset was created by a community user rather than an official source, it exhibits a consistent and well-structured format. The methodology employed aligns with common practices in computational literary analysis and effectively captures the complex network of relationships present in the books. Nevertheless, it is important to acknowledge that the adopted criteria may occasionally include interactions that are not narratively significant. To address this, we applied filtering and qualitative validation steps during our analysis to ensure that the extracted relationships align with the narrative structure and plot development of the original text.

4 Validity and Reliability

Validity, crucial in measuring how accurately the data represent the plot of the five books, cannot be attested with complete certainty, as the dataset was built and uploaded by a Kaggle user. In fact, the methodology followed when collecting the information can be questioned as the choice for what an interaction between two characters is determined by can be viewed as quite arbitrary. To be exact, two characters are deemed as interacting in the dataset when they both appear in fifteen words of the narration. Nevertheless, since we made known that the dataset is sourced from Kaggle (a platform known for its reliable and reviewed datasets) a certain level of accuracy can be assumed.
The decision to proceed with this dataset was pursued as correspondences to the books were observed, not only in our early findings during the network analysis but even when implementing more specific metrics. A high-level of consistency was found between the relationship established by the characters in the books and the way the same characters are represented in the graphs. Therefore, this qualitative control is helpful to validate how the database is built. Some examples of metrics we have used as implements to support validity are cliques and k-cores, for which the resulting subgraphs are easily interpretable by experts as coherent with the narrative progression of the books. A specific setting in which affinity to the plot development can be observed is found analyzing the 11-core for the second book, in which all the characters that are part of the Kingsguard are present.

11-Core for the second book: Amory Lorch, Tickler, Cersei Lannister, Polliver, Chiswyck, Joffrey Baratheon, Gregor Clegane, Meryn Trant, Rafford, Sandor-Clegane, Ilyn Payne, Dunsen.

The reliability, hence the reproducibility of our study, can instead be considered high, since the dataset is available on the Kaggle platform and easy to download in .csv format.
The data preprocessing steps executed to manipulate and reorganize data should be easily reproducible, as we have exploited some of the most known Python libraries for network analysis. In addition, the Python notebook developed for the network analysis is publicly available on GitHub, for a further explanation of how we have implemented each of the functions for the study and in-depth data visualization.

5 Measures and Results

(Centrality metrics: tengono conto di quante volte un nodo e' connesso con gli altri. **Eigen vector centrality** tiene conto di quanto un nodo e' connesso ad altri nodi che contano quindi riesce a identificare i personaggi piu' importanti, pero' simile alla degree centrality nel nostro caso perche' abbiamo diversi personaggi centrali che rappresentano i protagonisti, nel nostro caso non esiste un main characters. Kathz centrality risolve un problema della eigen vector centrality per i grafi diretti ma noi non abbiamo grafi diretti quindi ha quasi le stesse caratteristiche della eigen vector centrality [si puo' inserire solo con informazione del fatto che a noi non serve ma e' stata calcolata]. **Closeness centrality**: cerca di identificare un main character perche' un protagonista e' piu' probabile che sia collegato ai diversi personaggi secondari [simila a eigen vector centrality]. **Betweeness centrality**: collega diversi punti di controllo, personaggio intermedio tra altri, quindi in contatto con piu' personaggi anche se non principale. Cluster:

Cliques sono il set di personaggi che sono tutti cooccorrenti tra loro,) What measures did you apply (brief explanation of how they work)? How do they relate to the intent of the study? Why

are they relevant? What is the connection among the gathered data, the applied measures, and the properties found?

6 Conclusion

Qualitative analysis of the quantitative findings of the study.

7 Critique

Do you think your work solves the problem presented above? To which extent (completely, what parts)? Why? What could you have done differently to answer your research problems (e.g., gather data with additional information, build your model differently, apply alternative measures)?