

Exploring the relationships among characters in a novel

Ali Kaya
ali.kaya@abo.fi

Problem statement

Network analysis is an invaluable tool, even in the context of complex systems like literature. The relationships between characters within a novel are a fundamental narrative element. Exceptional novels such as Harry Potter often boast intricate character networks due to their rich cast of characters and intricate plotlines. Whether readers are familiar with the series or not, delving into and visualizing the intricate character relationships provides a unique opportunity to gain insights into this famous series.

Problem importance

In my personal experience, I seldom have the opportunity to immerse myself in reading such novels or watching their film adaptations. However, the chance to utilize network analysis methods and techniques in Data Science Lab 3 has genuinely sparked my curiosity. I am enthusiastic about delving into the series from a data scientist's perspective, scrutinizing the intricate web of character relationships. This endeavor promises to become a valuable resource for my future leisurely reading, guaranteeing a seamless experience when I decide to delve into the series.

Proposed approach

- Seeking a book or a series of books for character extraction: **Harry Potter 1-7**
- Collecting all characters from the books as a dictionary: [Json Files](#)
- Employing Named Entity Recognition tools such as Spacy NER to extract characters from the content of the books.
- Utilizing the character data obtained in step 2 as a dictionary to remove inaccuracies introduced by the NER process in last step.
- Establishing connections between pairs of characters through various strategies, such as their co-occurrence in the same sentence or paragraph.
- Visualizing the networks representing the connections calculated in last step.
- Applying different centrality measures to assess the importance of each character.
- Identifying character communities using a or some partitioning strategy.
- Analyzing the progression and development of characters throughout the books.

Expected outcome

- Co-occurrence Analysis: the outcome includes the analysis of the frequency and patterns of characters co-occurrence in Harry Potter
- Character Networks: Create a network where characters are nodes, and connections between them represent interactions

- Community Detection: Use network analysis algorithms to detect communities or groups of characters that frequently interact with each other.
- Centrality Measures: Calculate centrality metrics (e.g., degree centrality, between centrality) to determine the most influential or important characters in the series.
- Evolution of Characters: Track the changes in character relationships and interactions across different books in Harry Potter.