Discovering Structure in Social Networks of 19th Century Fiction

Siobhán Grayson, Jennie Rothwell, Maria Mulvany, Karen Wade, Gerardine Meaney, and Derek Greene

Motivation
Aim: To apply community detection to networks constructed from 19th century literary texts by Jane Austen and Charles Dickens.

For years organisations like Project Gutenberg, Google Books, and Open Library have been digitising and archiving cultural works, transforming the literary landscape and the way researchers engage with it.

Social Network Analysis provides researchers with a unique level of abstraction. In the context of a novel, nodes represent characters and edges co-occurrences of characters.

By investigating the structure of character networks using overlapping community detection and in collaboration with literary scholars, we find the literary technique of the authors is revealed by the resulting communities.

Methodology
Character Networks
1. The full text of each chapter of the novel was annotated by a literary scholar to identify all characters and their aliases.
2. Using a collinear strategy, consecutive pairs of characters co-occurrences that appear within a sliding window of length w words are counted.
3. We then created a weighted character network for the chapter, where an edge is weighted such that it corresponds to the number of co-occurrences between the pair.
4. Finally, we construct an overall character network for the novel by aggregating the networks from all chapters.

Community Detection
OSLOM is used to for community detection. It works by measuring the quality of potential communities using a fitness function based on their statistical significance. Two parameters govern the nature of the communities detected:

- Coverage Parameter: Determines whether communities should be merged or not.
- P-Threshold: Alters the threshold at which a community is deemed to be significant.

Results and Analysis
The fictional societies constructed by Austen are generally more compact and closely-knit than those of Dickens.

Discussion
Many communities found in Oliver Twist can be linked to "micro-narratives".

In contrast, Austen’s communities appear to centre more on the social circles of specific individuals.

Conclusion
Community detection allowed us to observe micro-plots, character social circles, and identify overlapping characters.