

Game of Thrones Network Analysis

Federico Ceccato

Università Cà Foscari Venezia



Agenda

1 Introduction

2 EDA

3 Analysis

4 Coclusions

Introduction

In this project I tried to understand if it's possible to know the importance and the evolution of a character based on the number of interactions that he/she has in the books of Game of Thrones.

We have an interaction between two characters if their names are between 15 words starting from the first name that appears. A purpose of this project is also to understand if this way to calculate interaction could be fine or not.

Dataset

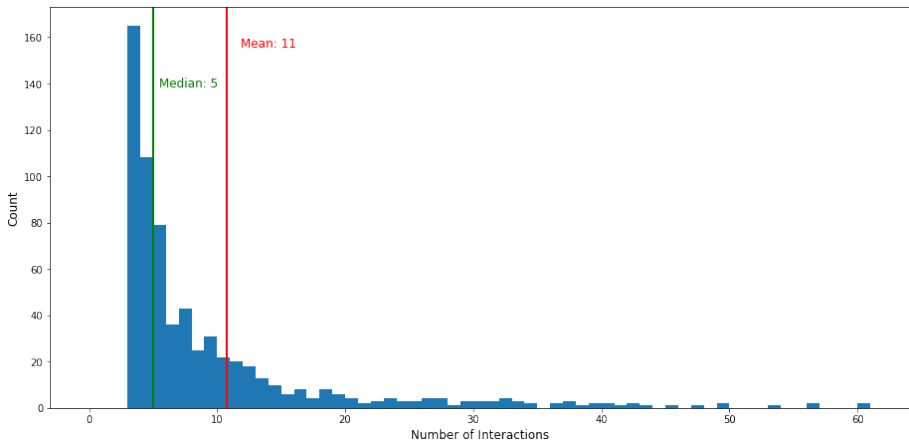
The dataset informations are:

- Source: the first character name that appears
- Target: the second character name that appears
- Weight: the number of interactions that the two characters have
- Book: the book number

Results First Book

In the first book we have 139 unique values for source and 143 for target. The distribution is right skewed with a mean of 11, a median of 5 and the majority of values lay between 0 and 15.

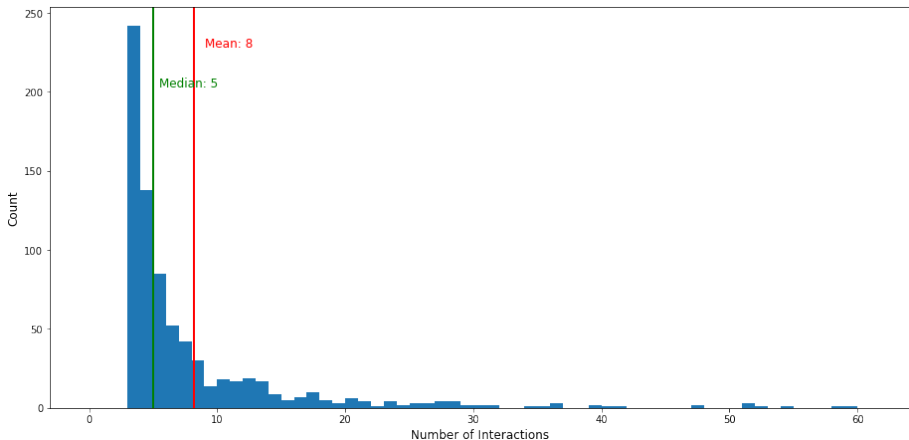
Interactions Distribution



Results Second Book

In the second book we have 193 unique values for sources and 186 for target. The distribution is right skewed with a mean of 8, a median of 5 and the majority of values lay between 0 and 12.

Interactions Distribution



First Network

The basic stats of the network are:

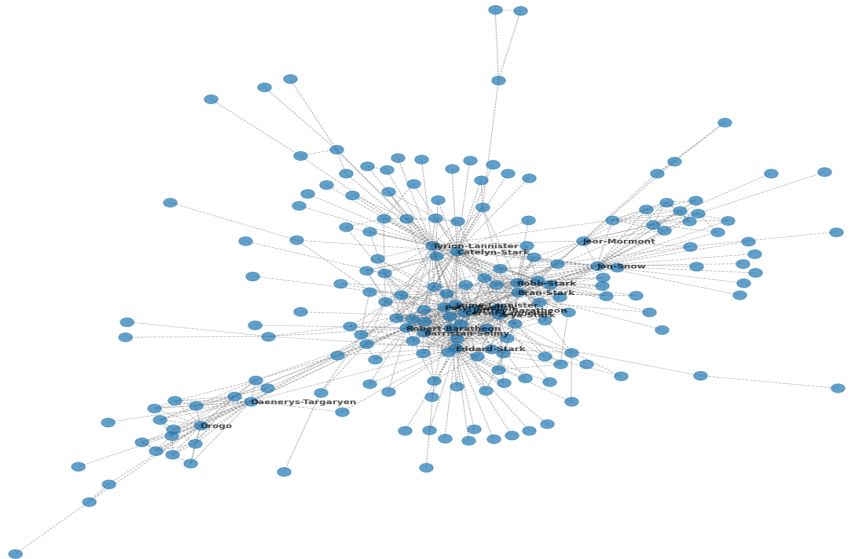
- Number of nodes: 187
- Number of edges: 684
- Average degree: 7.3155
- Network density: 0.03933

Second Network

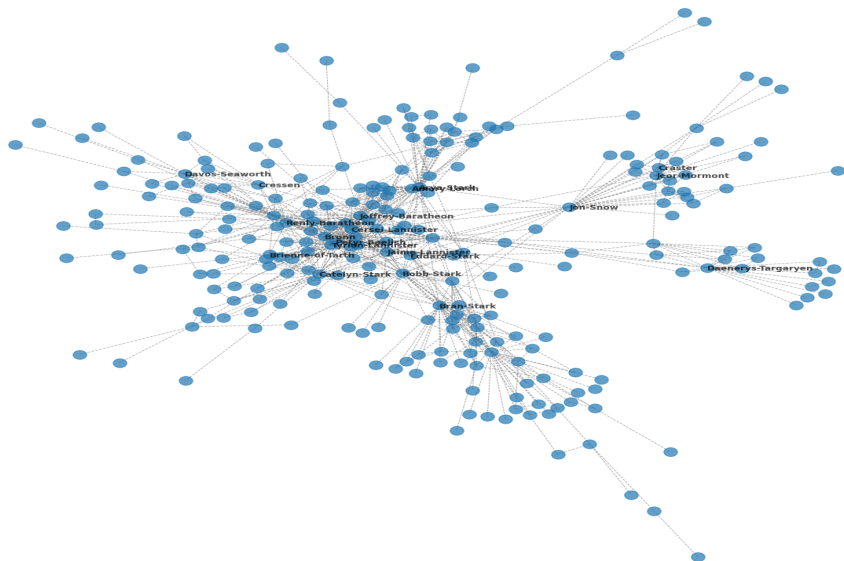
The basic stats of the network are:

- Number of nodes: 259
- Number of edges: 775
- Average degree: 5.9846
- Network density: 0.02319

Network Graph 1



Network Graph 2



Metrics

I decided to use some metrics to understand the importance of the characters.

- Degree Centrality: which is defined as the number of links incident upon a node.
- Betweenness Centrality: is the average length of the shortest path between the node and all other nodes in the graph.
- Eigenvector Centrality: is a measure of the influence of a node in a network.
- Closeness Centrality: quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.

Metrics First Book

Degree Centrality:

- 'Eddard Stark', 66
- 'Robert Baratheon', 50
- 'Tyrion Lannister', 46

Betweenness Centrality:

- 'Eddard Stark', 0.26960
- 'Robert Baratheon', 0.21403
- 'Tyrion Lannister', 0.19021

Eigenvector Centrality:

- 'Eddard Stark', 0.29640
- 'Robert Baratheon', 0.26948
- 'Sansa Stark', 0.23155

Closeness Centrality:

- 'Eddard Stark', 0.56363
- 'Robert Baratheon', 0.54545
- 'Tyrion Lannister', 0.51098

Metrics Second Book

Degree Centrality:

- 'Tyrion-Lannister', 53
- 'Joffrey-Baratheon', 47
- 'Cersei-Lannister', 43

Betweenness Centrality:

- 'Arya Stark', 0.18811
- 'Jon Snow', 0.17443
- 'Robb Stark', 0.16494

Eigenvector Centrality:

- 'Joffrey Baratheon', 0.30736
- 'Cersei Lannister', 0.29520
- 'Tyrion Lannister', 0.28264

Closeness Centrality:

- 'Robb Stark', 0.47777
- 'Eddard Stark', 0.45744
- 'Robert Baratheon', 0.44869

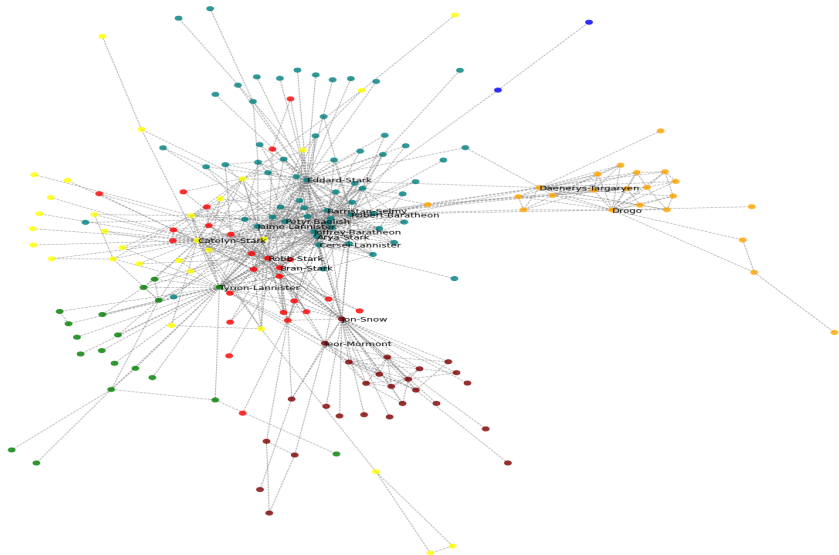
Community Detection

Knowing the structure of the map of the Game Of Thrones world I thought that will be interesting try to detect possible communities in the network. To do this I decided to apply different methods:

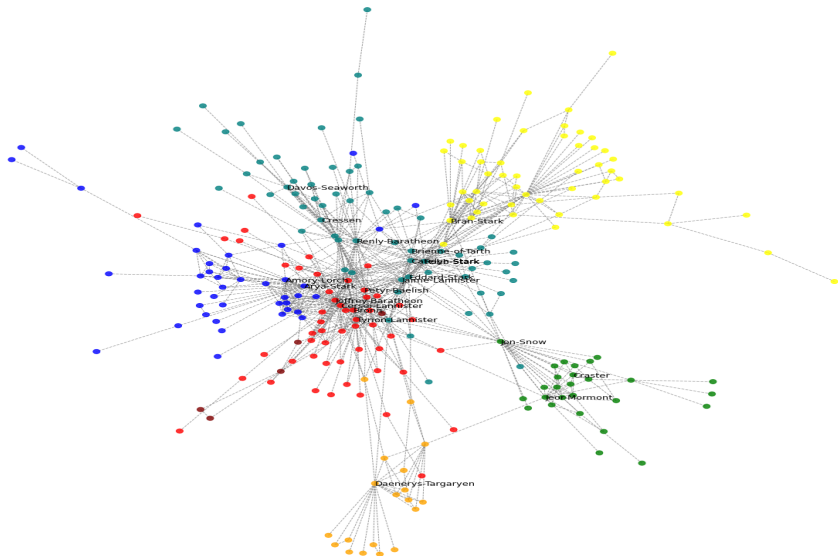
- Louvain
- Girvan Newman
- K-Clique
- Label Propagation

In the next slides we'll see the results of these algorithms for both books.

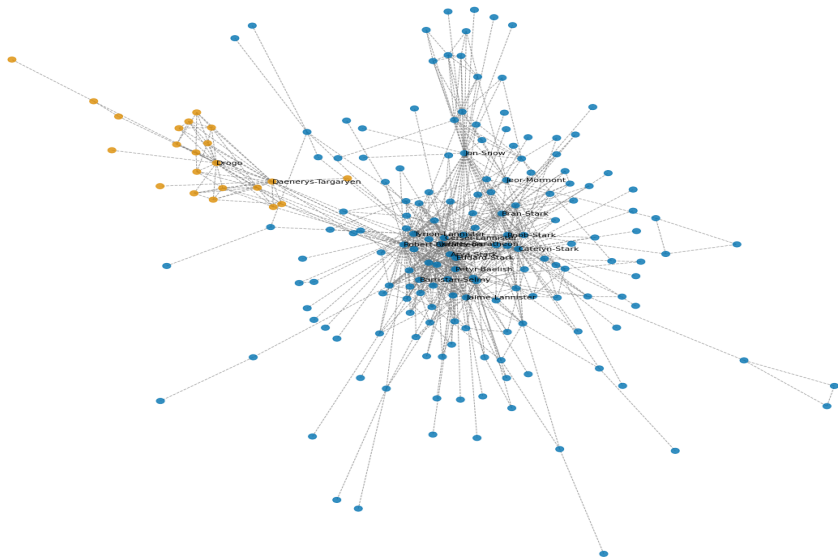
Louvain for First Book



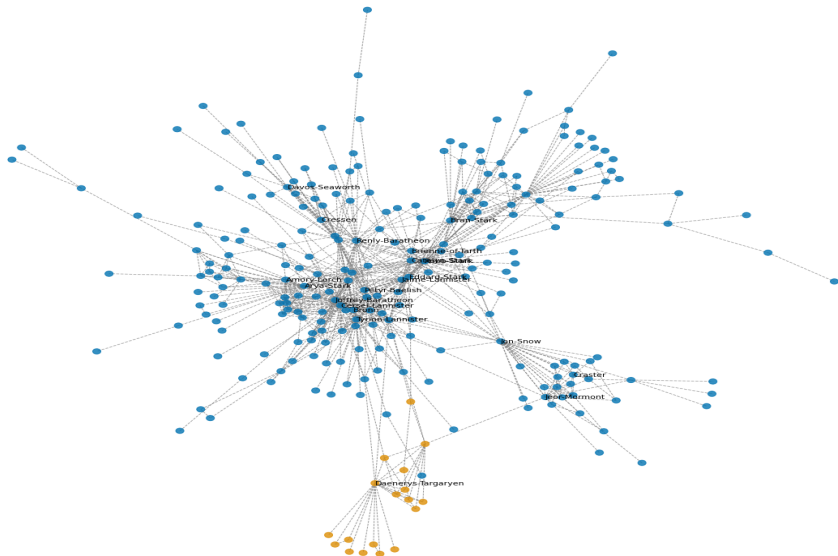
Louvain for Second Book



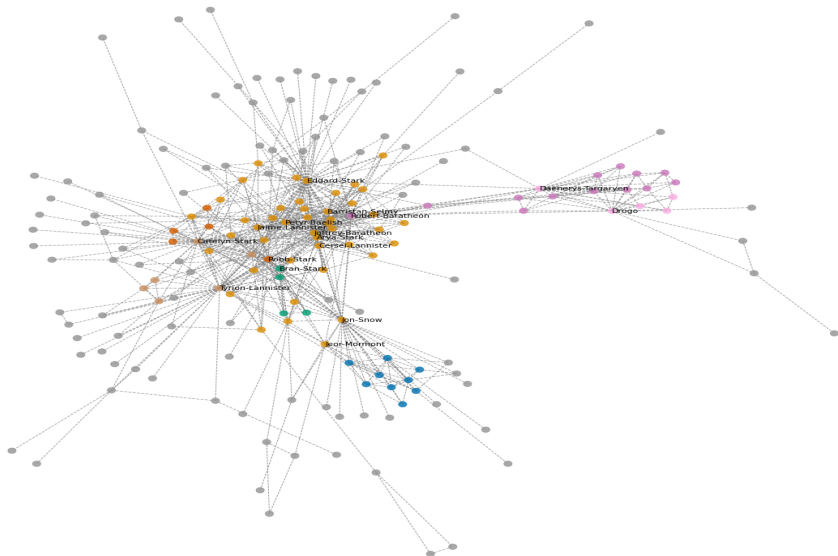
Girvan Newman for First Book



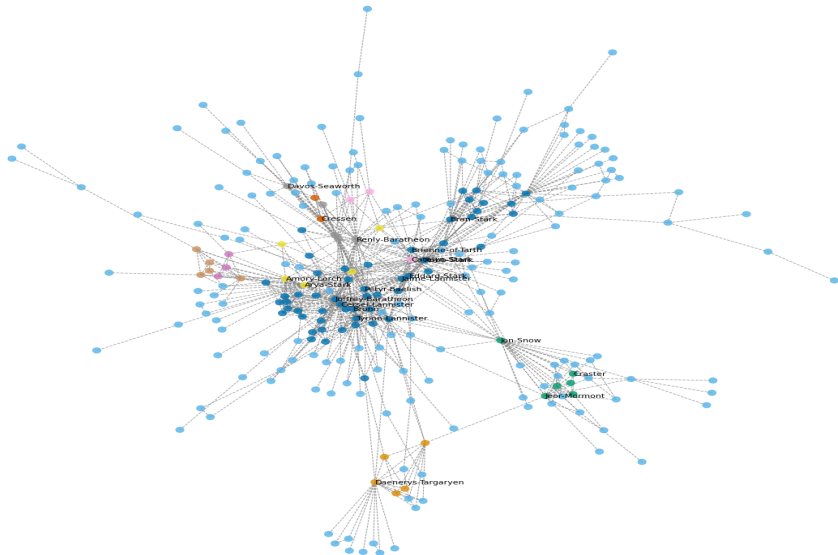
Girvan Newman for Second Book



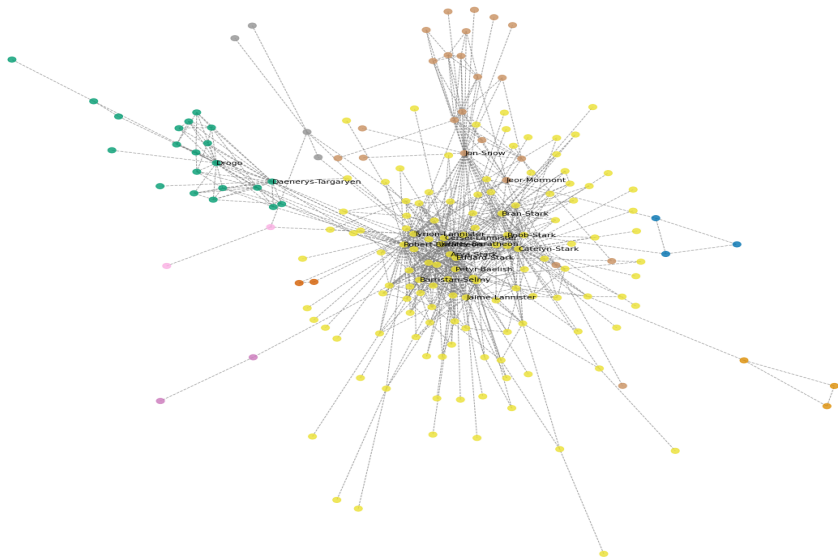
5-Clique for First Book



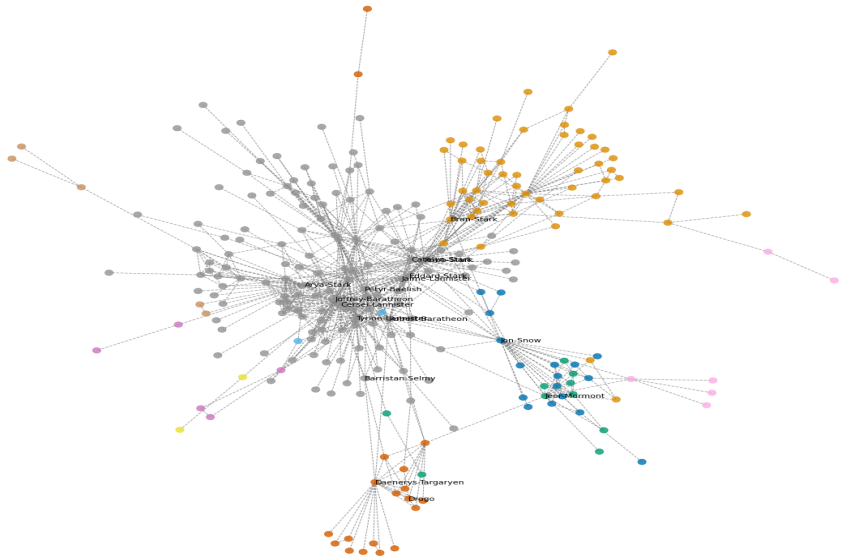
5-Clique for Second Book



Label Propagation for First Book



Label Propagation for Second Book



Conclusions

In the end we can conclude that the method for calculating interactions is quite reliable, even if we have a lot of interactions that don't exist in the reality. In fact some people that are dead or not present during a dialogue can be count as interactions. One possible explanation could be that we are in the presence of flashbacks but I think there are too much interactions to explain this. So I think that we are able to understand the importance and the evolution of the characters with these analyses but we could implement the way we calculate the interactions adding the column and the quotation marks to isolate a direct speech.