Social network analysis of communities in Literature.

Ana Castillo, Víctor Buendía, Joan Losa, Luca Marconi, Álex Molas, Alejandro Morán, Gianmarco G. Pisano, Joan Pont, Patrick Sánchez, Miguel A. Trigo, Eduardo Varela

IFISC (CSIC-UIB) Palma de Mallorca – Spain.

Abstract

In this work we provide the **interaction networks between characters** from three widely known novels: The Lord of the Rings, Harry Potter and Dune. We analyze the **structure in communities** of them and also test the accuracy of the **community detection algorithm** by directly determining the relationships between characters, as well as the limitations of the algorithm. We have also computed the **betweenness centrality** and shown that the higher value of this measure allow us to distinguish a main character, on whom falls the weight of the plot, clearly differentiated from the rest of the roles. Finally, by computing some characteristic magnitudes, such as the **average path length** and the **clustering coefficient** of the network, we have determine if social networks described in fiction are realistic and could be representative of a real society.

Network construction

We parsed the ebook, incrementing the edge weight between two characters when their names (or nicknames) appeared within 15 words of one another. However is important to notice that an edge between two characters doesn't mean that they are friends or enemies; it simply means that they are mentioned together.

Small-worldness measure

In order to determine if the book's networks have small-world properties we have measured their average path length (*l)* and clustering coefficient (C_i) given by

$$\ell = \frac{1}{N(N-1)} \sum_{i \neq j} d(i,j) \qquad \mathcal{C}_i = \frac{2E_i}{k_i(k_i-1)},$$

where *N* is the number of of vertices in the graph, *d(i, j)* the shortest distance between the nodes *i* and *j*, E_i and k_i the number of edges of the nearest neighbors and numbers of edges of the node *i.*

Community detection

The community detection has been developed by implementing a Louvain algorithm to maximize modularity. The main steps of the algorithm are:

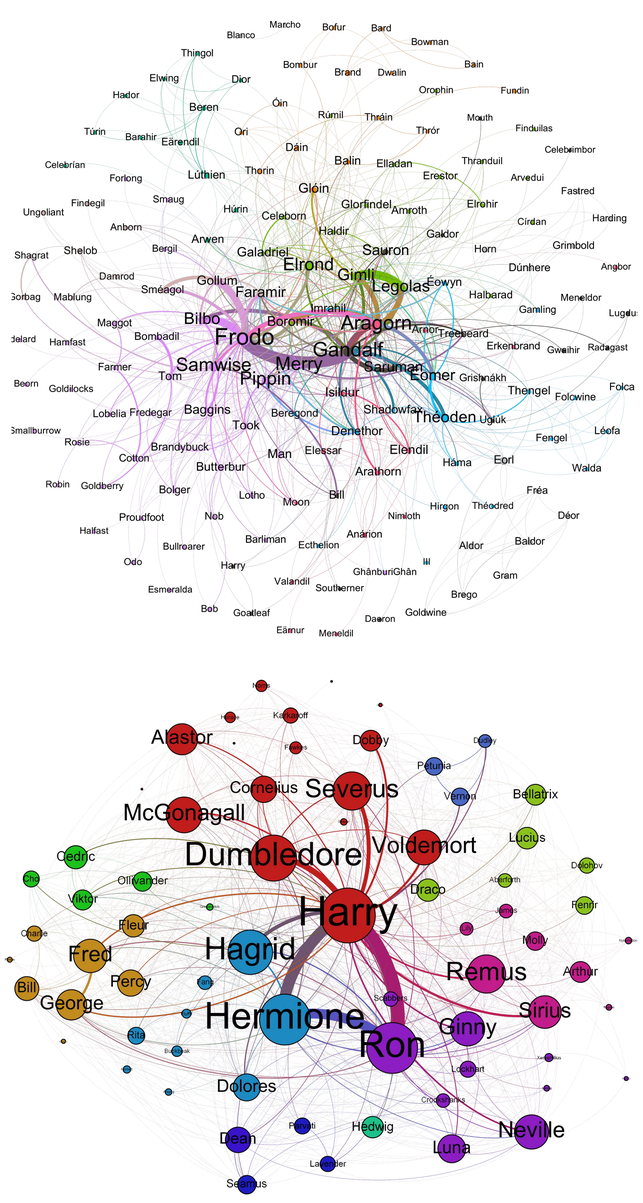
1. Each node is considered to be a community in itself.
2. Remove the node and compute the change in modularity (*Q)*.
3. Add the node to the neighbour's community with highest new modularity.
4. Use the communities as nodes of a new network. Links are weighted using the number of links between nodes of the first network.

Modularity has been computed as,

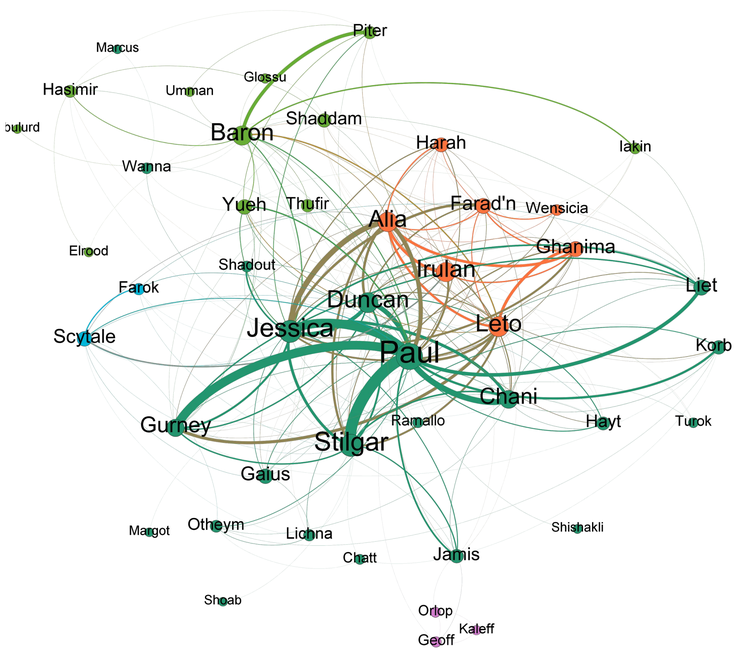
$$\mathcal{Q} = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

where *A* is the adjacency matrix, c_i the community where *i* is assigned and *m* is one half of the sum of all the adjacency matrix entries.

Results



Right: network of the first three novels of **Dune**. Paul is the most central character, with a very high betweenness centrality, *g=0.32*. In dark green, Atreides family and Fremen, both directly related with Paul, are detected. In orange, descendants and related of Atreides. In light green, Harkonnen family. Here most important node is the Baron, antagonist of the novels. We obtain a clustering of *c=0.70* and a mean path length of *l=2.0.* The high clustering and small path length says that this network has small-world properties.



Left up: network of **The Lord of the Rings**. Central characters can be seen in the center, in fact Frodo has the highest value of the betweenness centrality with a value of *g=0.18.* Communities coincide pretty well with countries in the novel: for example, light blue corresponds to Rohan and Gondor, purple to the Shire, light green to elves, orange to dwarfs, and pink to Aragorn's ascendants. The clustering coefficient of the network gives *c=0.70,* with an average path length of *l=2.7*, again in consonance with a small-world network.

Left down: network of the **Harry Potter** novels. The main character, Harry, is accurately predicted by measuring the betweenness centrality of the network, with *g=0.26*. Different communities are detected, of which the most important are: Hogwarts professors in red and two communities of students in purple and blue. A representation of Death Eaters is painted in light green and in blue grey we observe a muggle group, highly disconnected from the rest of the network. However, as in this novel characters are well integrated in different important groups, this differentiation breaks other important communities, i.e. The Order of the Phoenix and the Weasley family.
The average path length and the cluster coefficient computed for this network are respectively *l=1.9* and *c=0.72.* This results indicate a small-world topology in the network which is coherent with results obtained in real-world social networks.

References

-A. Beveridge and J. Shan, *"Network of Thrones"* Math Horizons Magazine , **Vol. 23**, No. 4, pp. 18-22 (**2016**)
-Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre. *"Fast unfolding of communities in large networks"*, (**2008**)
-R. Albert, A. L. Barabási. *"Statistical mechanics of complex networks"*. Reviews of Modern Physics. 74 (1): pp. 47-97, (**2002**)
-L. Muchnik, S. Pei, L. C. Parra, S. D. S. Reis, J. S. Andrade, S. Havlin, H. A. Makse, *"Origins of power-law degree distribution in the heterogeneity of human activity in social networks",* Scientific Reports 3, Article number: 1783 (2013)
-D. J. Watts, S. H. Strogatz. *"Collective dynamics of small-world networks"*, Nature **393**, 440-442 (**1998**)
-A. Clauset, C. Rohilla Shalizi and M. E. J. Newman. *"Power-law distributions in empirical data"*, SIAM Rev., **51(4)**, 661–703.