To Be or Not to Be Central

On the Stability of Network Centrality Measures in Shakespeare's "Hamlet"

Anastasia Klimashevskaia
Graz University of Technology, Graz

Bernhard C. Geiger
Know-Center GmbH, Graz

Martin Hagmüller
Graz University of Technology, Graz

Denis Helic
Graz University of Technology, Graz

Frank Fischer
Higher School of Economics, Moscow

Introduction

Centrality measures derived from character networks can be used to detect the main characters in a play. For example, previous research has shown that characters with high network centrality typically perform the majority of speech acts and appear in most of the scenes (Fischer, Trilcke, Kittel, Milling, & Skorinkin, 2018). However, one can extract character networks from plays in various ways: Close reading may omit minor characters like attendants or servants, e.g., (Moretti, 2011), while distant reading (e.g., parsing an XML file) may include aggregate characters like "All", "Both Lords", or similar. Furthermore, the networks may display either implicit or explicit connections, depending on whether we connect characters because they appear in the same scene or because they are directly addressing each other, respectively. Thus, as adding more characters or connections to the network affects centralities and other network measures, the interpretation of both qualitative and quantitative aspects of character networks depends on the extraction method.

In this work we are concerned with the specific question whether details of the textual source and the extraction method, such as adding minor or aggregate characters, make main characters less "central". Answering this question will allow us to better assess the validity of automated literary network analysis.

Approach

We analyse six versions of the character network of Shakespeare's "Hamlet". All networks were extracted via close or distant reading from different XML or text sources (see Table 1) and analysed with NetworkX (Python). For each network, we compute four different centrality measures (closeness, betweenness, degree, and eigenvector centrality). Subsequently, for each centrality measure, we rank the 26 characters common in all networks and compare character ranks in different networks by computing their Spearman rank correlation.

Dataset	Implicit/Explicit Close/Distant	# of characters	# of Edges	Avg. DEGREE	Diam.	Avg. Path Length
DraCor (Fischer, Trilcke, & Orekhov, Drama Corpora Project, 2019)	IMPLICIT/DISTANT	38	220	11.57	3	1.74
Ezlinavis (Milling & Fischer, 2019)	IMPLICIT/DISTANT	34	192	11.29	3	1.70
Grandjean (Grandjean, 2015)	IMPLICIT/DISTANT	37	263	14.22	3	1.63
Haworth (Haworth, 2019)	IMPLICIT/DISTANT	35	217	12.4	3	1.67
Own XML Parser	EXPLICIT/DISTANT	35	74	4.23	4	2.3
Moretti (Moretti, 2011)	EXPLICIT/CLOSE	28	54	3.86	4	2.29

Table 1. Basic statistics of the character networks

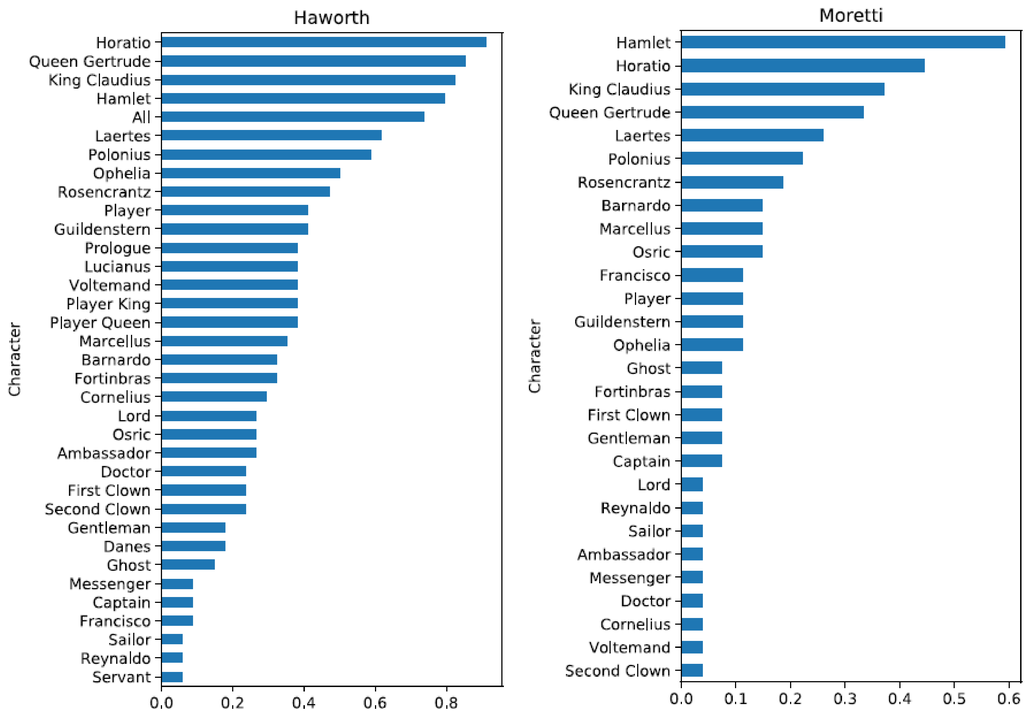

Figure 1. Degree centralities for networks extracted by Haworth (implicit connections, distant reading) and Moretti (explicit connections, close reading)

Observations

The networks including implicit connections are denser than those with only explicit connections (cf. Table 1). This yields different centrality ranks including the most important characters (cf. Figure 1). For example, Horatio has many more implicit connections and connections to minor characters, which makes him the character with the highest degree centrality in the Haworth network. In the Moretti network, which contains only explicit connections, Hamlet has the highest degree centrality.

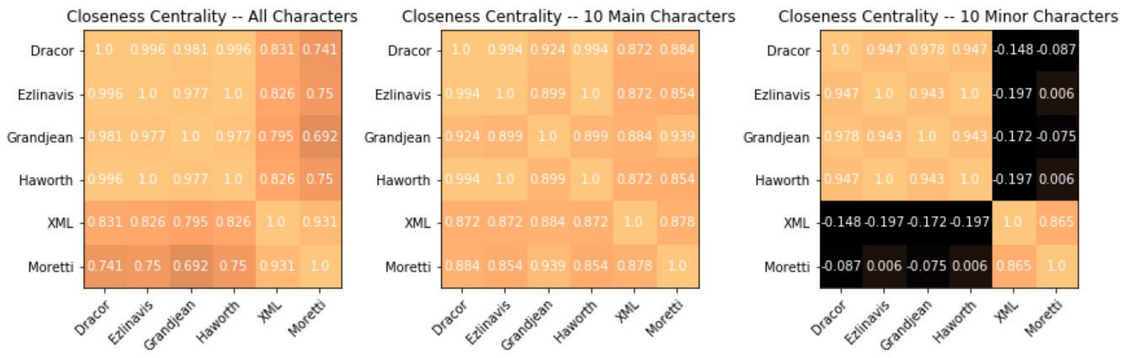

Figure 2. Heatmaps depicting the rank correlation between closeness centralities derived from different networks for all, the 10 most important, and the 10 least important characters

Conclusion and Outlook

Despite such individual differences, the groups of main characters derived from different networks exhibit relatively stable rankings, cf. (Fischer, Trilcke, Kittel, Milling, & Skorinkin, 2018). In contrast, rankings for minor characters tend to differ significantly (see Figure 2). Therefore, for detecting the group of main characters, the details of the network extraction method do not have a significant effect, at least in the datasets we consider. In future work we aim to validate the generality of this claim by considering larger corpora of dramatic plays.

Acknowledgements

The authors gratefully acknowledge permissions to use material from Martin Grandjean and Roger Haworth. The work was funded by the HRSM project "KONDE – Kompetenznetzwerk Digitale Edition". The work of Bernhard C. Geiger was partially funded by the Austrian Academy of Sciences within the go!digital Next Generation project "DiSpecs" (GDNG_2018-046_DiSpecs). The Know-Center is funded within the Austrian COMET Program – Competence Centers for Excellent Technologies – under die auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Digital and Economic Affairs, and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

References

Fischer, F., Trilcke, P., & Orekhov, B. (2019). Drama Corpora Project. Retrieved from github.com/dracor-org/shakedracor

Fischer, F., Trilcke, P., Kittel, C., Milling, C., & Skorinkin, D. (2018). To Catch a Protagonist: Quantitative Dominance Relations in German Language Drama (1730–1930). Digital Humanities 2018 (Puentes–Bridges). Book of Abstracts, 193–201.

Grandjean, M. (2015). Network visualization: mapping Shakespeare's tragedies. Retrieved from martingrandjean.ch/network-visualization-shakespeare/

Haworth, R. W. (2019). Who appears where matrices and reader's versions for Shakespeare plays. Retrieved from http://rhaworth.net/shakes/shaxref.php?csv=on&targ=on&dnl=on&p=menu#dummy

Milling, C., & Fischer, F. (2019). Easy Linavis. Retrieved from https://ezlinavis.dracor.org/

Moretti, F. (2011, March). Network Theory, Plot Analysis. New Left Review (68), pp. 80–102.