Co-reference networks for dramatic texts

Network analysis of German dramas based on co-referential information

Janis Pagel

`janis.pagel@uni-koeln.de`

Department for Digital Humanities, University of Cologne

Wednesday 27th July, 2022

Universität zu Köln   VolkswagenStiftung   DFG Deutsche Forschungsgemeinschaft

Overview

- ▶ Co-presence networks
- ▶ Co-reference networks
- ▶ Comparison of the different types of networks

Social Network Analysis in Computational Drama Analysis

Co-presence networks

- ▶ Node represents character
- ▶ Edges represent co-presence
  - ▶ How often do characters occur together on stage
  - ▶ Usually based on scenes
- ▶ Number of co-occurences can be used to weight edges

Example for co-presence network from DraCor (Fischer et al. 2019)
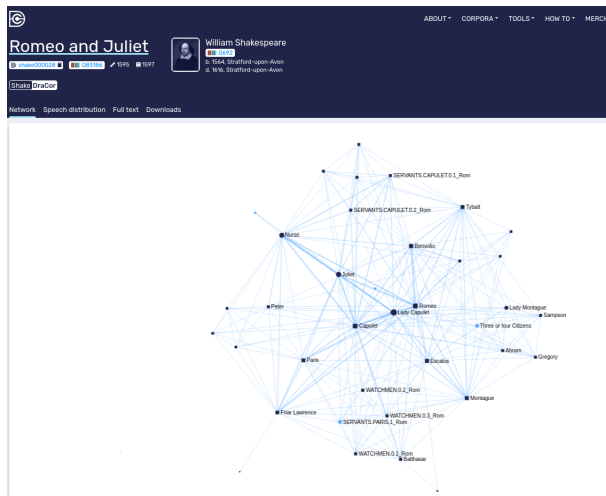

Figure: `https://dracor.org/shake/romeo-and-juliet`

Use information other than co-presence?

- ▶ For prose texts, several measures to base edges on have been proposed, e.g.
  - ▶ Adjacent quoted speech (Elson, Dames, and McKeown 2010)
  - ▶ Topic frequencies (Celikyilmaz et al. 2010)
  - ▶ Social events (Agarwal et al. 2012)
  - ▶ Similarities of word embeddings (Wohlgenannt, Chernyak, and Ilvovsky 2016)
- ▶ For computational drama analysis, co-presence networks are the standard

Co-reference networks

Co-reference networks

- Based on co-referential mentions of characters
- What is co-reference?
  - All mentions that refer to the same entity are co-referent
  - Noun phrases, pronouns
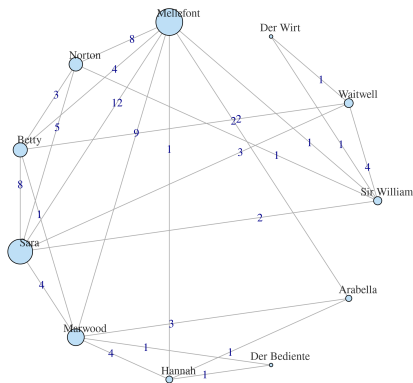- Edges are based on the information if a character mentions another character in a scene (or segment)

Co-reference network: Miss Sara Sampson

Lessing, Gotthold Ephraim: Miß Sara Sampson (1755)
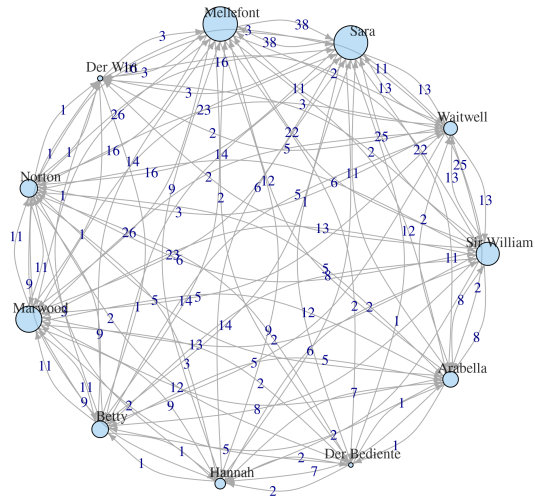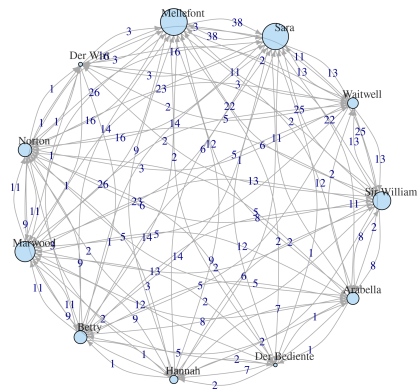
Comparison of co-presence and co-reference networks

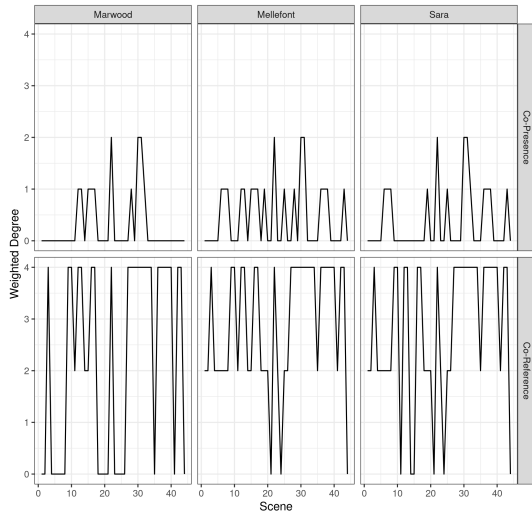Lessing, Gotthold Ephraim: Miß Sara Sampson (1755)

Lessing, Gotthold Ephraim: Miß Sara Sampson (1755)

Progression of degree over scenes: Miss Sara Sampson

Lessing, Gotthold Ephraim: Miß Sara Sampson (1755)

Analyses on a corpus of German plays

Corpus

- 31 texts from DraCor (https://dracor.org)
- Annotated for co-reference (Pagel and Reiter 2020)
- ca. 300 000 tokens, 60 000 mentions and 5500 entities

Comparison of centrality measures on corpus

	Centrality Measure	Mean	SD
Co-presence	Degree	0.48	0.28
	Betweenness	0.04	0.08
	Closeness	0.44	0.18
	Eigenvector	0.41	0.31
Co-reference	Degree	0.55	0.30
	Betweenness	0.05	0.09
	Closeness	0.43	0.18
	Eigenvector	0.47	0.33

Progression of degree for all characters
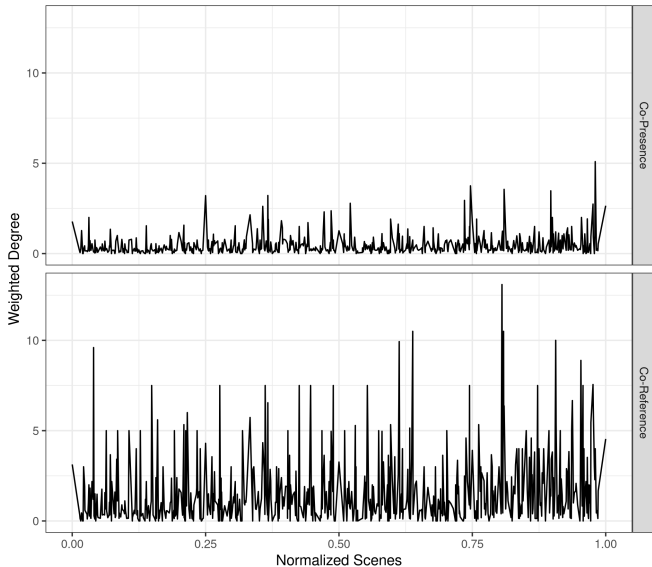
Correlation for degree of title characters and not title characters

- ▶ 10 title characters in the 31 plays
- ▶ Spearman's ρ between degree centrality and title characters and not title characters (encoded as 1 and 0)

Network type	Correlation
Co-presence	0.1249
Co-reference	0.24479

Conclusions

Conclusions

- New type of network: co-reference based networks
- Richer information about relationship of characters
- Useful for corpus-wide quantitative analysis
- Future work
    - Test usefulness of co-reference networks on larger corpora and bigger tasks
    - Test if automatically annotated co-reference information can improve upstream tasks as well
    - Explore injection of richer information into the networks (e.g. number of times a character mentioned)
    - How to compare (and potentially combine) networks containing different types of information

References I

Agarwal, Apporv, Augusto Corvalan, Jacob Jensen, and Owen Rambow (June 2012). "Social Network Analysis of Alice in Wonderland". In: *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*. Montréal, Canada, pp. 88–96. URL: https://aclanthology.org/W12-2513.

Celikyilmaz, Asli, Dilek Hakkani-Tur, Hua He, Greg Kondrak, and Denilson Barbosa (Dec. 2010). "The Actor-Topic Model for Extracting Social Networks in Literary Narrative". In: *Proceedings of the NIPS 2010 Workshop – Machine Learning for Social Computing*. Whistler, BC, Canada. URL: https://webdocs.cs.ualberta.ca/~denilson/files/publications/nips2010.pdf.

Elson, David K., Nicholas Dames, and Kathleen R. McKeown (July 2010). "Extracting Social Networks from Literary Fiction". In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*. Uppsala, Sweden, pp. 138–147. URL: https://aclanthology.org/P10-1015.

References II

Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling, and Peer Trilcke (June 2019). "Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama". In: *Proceedings of DH2019: "Complexities"*. Utrecht, The Netherlands. DOI: 10.5281/zenodo.4284002.

Pagel, Janis and Nils Reiter (May 2020). "GerDraCor-Coref: A Coreference Corpus for Dramatic Texts in German". In: *Proceedings of the Language Resources and Evaluation Conference (LREC)*. Marseille, France, pp. 55–64. URL: http://www.lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.7.pdf.

Wohlgenannt, Gerhard, Ekaterina Chernyak, and Dmitry Ilvovsky (Dec. 2016). "Extracting Social Networks from Literary Text with Word Embedding Tools". In: *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*. Osaka, Japan, pp. 18–25. URL: https://www.aclweb.org/anthology/W16-4004.