

NETWORK ANALYSIS ON THE AUSTRIAN MEDIA CORPUS

Gabriel Grill^a, Julia Neidhardt ^a, Hannes Werthner ^a

^a E188 - Institute of Software Technology and Interactive Systems

INTRODUCTION

The role and influence of media in society has been a much debated topic in the last several months, in particular with respect to populist political movements. On the one hand they supposedly benefit from sensational media reporting ^[1], while on the other hand claiming that media outlets report strongly biased against them, a technique for instance implemented by the Pegida movement ^[2]. In order to contribute to the study this kind of complex reciprocal interplay, this work aims to develop a theoretical framework that enables scholars to examine media reporting using network-based methods. Thereby revealing central entities, topics and statistical connections between extracted entities of interest in news articles. The comparison of different media outlets based on these results may yield insights into the study of media bias ^[3]. This work may encourage creators of media to reflect on their reporting and the public to be more informed on underlying systematics of news coverage. The framework is evaluated based on a case study focusing on last year's Austrian presidential election. This is the first research effort applying a network-based approach onto the Austrian Media Corpus ^[4], a complete and unique collection of over 33 million articles encompassing the last three decades of Austrian media coverage. The project is conducted in cooperation between the Electronic Commerce Group, Institute of Software Technology and Interactive Systems, TU Wien and the Austrian Centre for Digital Humanities, Austrian Academy of Sciences. Work on semantic annotation of entities and relations as well as network analysis is carried out in an internship at the National Institute of Informatics in Tokyo, Japan.

FUNDAMENTAL OF THE PROBLEM

This work specifically focuses on a case study encompassing reporting on the Austrian presidential elections in 2016. We expect insights related to the following research questions:

- Are we able to detect central entities, topics or key words in the extracted networks?
- What can we infer about structural properties of the networks and are we able to identify communities?
- How can we capture differences in reporting in respect to variant news publishers and time spans?

METHODOLOGY

In order to generate, analyse and visualize different types of networks, methods from the fields Natural Language Processing, Network Analysis and Data Mining are applied. First, an expressive subset of stemmed nouns, verbs and adjectives as well as co-occurrence-based relations between them are extracted from news articles. All words are stored with corresponding metadata. This includes a reference to the container sentence, paragraph and article as well as information on the article such as the

publishing date, news outlet and associated department. Nouns are also assigned different class labels (e.g. person, location) using a Named Entity Recognition framework. Networks are made of nodes representing selected nouns, referred to as entities in this work, and edges modelling relationships between them. Co-occurrence Networks consist of edges representing co-occurrence of entities in a sentence, paragraph or document. Emotion Networks contain up to eight types of edges representing either one of seven emotions or neutrality. Emotion labels are assigned based on the amount of words, included in an emotions dictionary^[5], associated with a co-occurrence relation between two entities. Sentiment Networks consist of edges between co-occurring entities and can either indicate a positive, negative or neutral valence. Sentiment labels are either assigned based on the strength of the valence and amount of words associated with a co-occurrence, or the valence of a verb between a subject and an object in a sentence^[6]. Valence is determined based on a dictionary^[7]. Social Networks are Sentiment Networks comprised only of entities classified as person or organisation.

EXPECTED RESULTS

The expected results of this work belong to two different levels. By studying network models and measures, we gain insights to develop a theoretical framework for different types of co-occurrence relationships. By conducting empirical analyses, we obtain concrete statements about the Austrian presidential election in 2016. Concrete results may open up discussion in the public or other fields. Results will be presented at the Vienna young Scientific Symposium in more detail.

CONCLUSION

In this work, a network-based approach is applied on the Austrian Media Corpus. This will help to determine whether such an approach is indeed well suited for the study of Austrian news reporting and whether statistical connections can be revealed that otherwise would not be easy to detect. The case study on the Austrian presidential election in 2016 will provide insights whether there are differences in reporting across news publishers, time spans and ballots. In a next step the classification of extracted entities and relations will be enriched using available semantic databases.

REFERENCES

- [1] Robert G. Picard, "Are Journalists Reporting - Or Publicising - Populism?", European Journalism Observatory - EJO, 28-Oct-2016.
- [2] K. Holt and A. Haller, "The Populist Communication Paradox of PEGDIA: Between 'Lying Press' and Journalistic Sources", 66th annual ICA conference "Communicating with power". Fukuoka, Japan, 9-13 June. Preconference: Populism in, by, and Against the Media, 2016.
- [3] W. L. Bennett, "News: The Politics of Illusion", Tenth Edition. University of Chicago Press, 2016.
- [4] M. Durco, K. Moerth, H. Pirker, and J. Ransmayr, "Austrian Media Corpus 2.0", 2014.
- [5] R. Klinger, S. S. Suliya, and N. Reiter, "Automatic Emotion Detection for Quantitative Literary Studies - A case study based on Franz Kafkas 'Das Schloss' and 'Amerika' ", Digital Humanities 2016: Conference Abstracts, Krakw, Poland, 2016, pp. 826-828.
- [6] S. Sudhahar, G. A. Veltri, and N. Cristianini, "Automated analysis of the US presidential elections using Big Data and network analysis", Big Data Soc., vol. 2, no. 1, Feb. 2015.
- [7] Remus, Robert, Uwe Quasthoff, and Gerhard Heyer. "SentiWS-A Publicly Available German-language Resource for Sentiment Analysis", LREC. 2010.