MODELING STRUCTURE AND CONTENT: SOCIO-SEMANTIC NETWORK ANALYSIS OF THE *MAHĀBHĀRATA*

Frederik Elwert, Center for Religious Studies, Ruhr-University Bochum, Germany. Email: <frederik.elwert@rub.de>.

Sven Sellmer, Center for Religious Studies, Ruhr-University Bochum, Germany. Email: <sven.sellmer@rub.de>.

See <www.mitpressjournals.org/toc/leon/50/5> for supplemental files associated with this issue.

Submitted: 9 November 2015

Abstract

There is a demand to incorporate content information into social networks. The authors constructed and visualized a network of the most important gods and heroes in the Sanskrit epic *Mahābhārata*. The network includes semantic information about the actors and their relationships. These two types of information were collected automatically with the help of the Nubbi topic modeling algorithm, which assigns separate sets of topics to both persons and their relations. The visualization of such a network provides intuitive access to a high density of information, like the topic distribution for each actor and the predominant topic for each relation.

This paper focuses on the methodological problem of the relationship between (social) structure and (semantic) content in humanities network research. Network analysis as a method traditionally leans toward the structural side. The social structure of fictional texts has been studied, e.g. by Moretti in his work on Shakespearean drama [1] or by Mac Carron and Kenna dealing with Icelandic sagas [2]. These approaches largely ignore the content of their texts. Here, we propose a method to jointly analyze structure and content.

We are interested in capturing the social structure but also including the semantics that are embedded within it. This kind of approach has been called “semantic social network analysis” [3,4]. Recent research in this area developed from “Semantic Web” principles. Social networks are extended with information from given ontologies to model semantically different types of nodes and edges. The outcomes are multi-modal and multiplex networks. The downside of this approach is that it requires the researcher to specify a semantic content model in the form of an ontology, instead of empirically building such a content model from the data. Instead, we pursue an inductive approach to uncover the internal qualities of the textual sources under study.

To this end, we apply techniques from topic modeling, which has the advantage of empirically detecting semantic clusters of words with little input from the researcher. One fitting application of topic modeling to the domain of network analysis has been developed and published under the name Nubbi, which is short for Networks Uncovered by Bayesian Inference [5]. This method performs topic modeling of context in a social network setting. The results are topics, or semantic word clusters, for both actors and relations in a social network. This allows the discovery of node classes and relationship types inductively. Nubbi applies a dual data model, which makes use of the fact that the text under study contextualizes entities and their relations. Entities mentioned in the text are extracted as a social network. Additionally, words around entities and entity pairs are assigned to nodes and edges as context documents. The topic modeling process then calculates topics for both nodes and edges. It assumes that words around single entities contribute only to entity topics, while words around entity pairs contribute to either entity topics or pair topics.

The results of this process are topics that describe entities (node classes) and topics that describe entity relations (edge types). There is an implementation of this algorithm available for R as part of the “lda” package [6].

The text to which we applied these methods is the *Mahābhārata*, the Indian national epic, in its original Sanskrit version (1M+ lexical entities). We added the topic distribution inferred by Nubbi to a social network of the 370 most frequent persons, based on the co-occurrence of two of them in one verse. Then, we visualized a subgraph containing 20 central gods and heroes as an arc diagram (Fig. 1) [7].

The graph displays the characteristic distribution pattern of entity topics for the single actors. The Nubbi algorithm e.g. distinguishes pure warriors from other persons with a more variegated profile. Relations between actors are represented as colored arcs whose thickness is proportional to the number of common contexts. The color for each pair is chosen according to the most frequently occurring pair topic between the two entities. The arc color represents dominant edge topics like “Religion” and “Fighting.” E.g. the persons connected by red and brown arcs on the right side of the graph are the most active participants in the great battle whose description makes up about 1/3 of the entire text.

To conclude, we found that Nubbi is a viable solution to the problem of analyzing structure and content in an integrated model. The arc diagram usefully highlights the most central connections but only works well with comparably few nodes.

Fig. 1. Semantic Social Network of the *Mahābhārata*.
(© Sven Sellmer)

References and Notes

This paper was presented as a contributed talk at Arts, Humanities, and Complex Networks—6th Leonardo satellite symposium at NetSci2015. See <<http://artshumanities.netsci2015.net>>. The SeNeReKo project is funded by the German Federal Ministry of Education and Research under the project number 01UG1242A. The authors of this paper are responsible for its content.

1. F. Moretti, “Network Theory, Plot Analysis,” *NLR*, No. 68 (2011) pp. 80–102.
2. P. Mac Carron and R. Kenna, “Network analysis of the Íslendinga sögur – the Sagas of Icelanders” (2013), <<http://arxiv.org/abs/1309.6134>>.
3. G. Erétéo, *Semantic Social Network Analysis* (2011), <http://www-sop.inria.fr/members/Guillaume.Ereteo/PhD_thesis_Semantic_Social_Network_Analysis.pdf>.
4. C. Thovex and F. Trichet, “Semantic social networks analysis,” *Social Network Analysis and Mining* 3, No. 1, pp. 1–15 (2012).
5. J. Chang, J. Boyd-Graber and D.M. Blei, “Connections between the lines: augmenting social networks with text,” in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (Paris: ACM, 2009) pp. 169–178.
6. J. Chang, “lda: Collapsed Gibbs sampling methods for topic models” (2012), <<http://CRAN.R-project.org/package=lda>>.
7. Our visualization is inspired by Steinweber’s “Similar Diversity” project. The code builds on the R implementation by G. Sanchez, “Star Wars Arc Diagram” (2012), <<http://gastonsanchez.com/work/starwars/>>.