N13 Diesner, J., & Carley, K.M. (2012). Impact of methods for relation extraction from text data on network analysis results.

Several methods are available for extracting relational data from natural language text data. While in prior research these methods have been applied across corpora and domains, there is a lack of understanding of the differences in network structure and network analysis results that are due to these different methods. We apply four different relation extraction methods to a corpus of newswire articles about a geopolitical entity (Sudan) and a corpus of funded research proposals, and compare the resulting networks in terms of structural properties and key entities. The following relation extraction methods aroused: First, the data to model approach to text coding in AutoMap. The key component of this process is a thesaurus, which maps text terms to concepts, and requires substantial human effort for creation and refinement. Second, we use a machine-learning based technique to automatically suggest a thesaurus, and repeat the first process with using this thesaurus. Third, we construct network data from meta-data about the texts; disregarding the bodies of the actual texts. Fourth, we have collaborated with subject matter experts on creating validated ground truth networks about tribal networks in the Sudan. Our findings show that there is little overlap in the networks identified by each method. The ground truth data are partially resembled by analyzing the content of text bodies (53% of the nodes and20% of the links), but not at all by relying on meta-data. We summarize the types of different views on networks that the employed relation extraction methods can provide.