N16 Wilner, S., Aleyasen, A., Mishra, S., Soltani, K., & Diesner, J. (2014). Comparison of network data constructed from text data and textual meta-data.

Coding texts as networks – a process also known as relation extraction – allows for constructing or supplementing network data. This approach is particularly useful when alternative methods for collecting network data fail, e.g. in the case of covert or historic networks, and when jointly considering social network data plus information produced or shared by network participants. Relation extraction techniques require humans – developers as well as end-users - to make choices about appropriate methods and parameter settings. The impact of these choices on the resulting data and findings can be strong, but is insufficiently understood. We have been addressing this problem by conducting a series of controlled, empirical experiments that we will report on in our talk. More specifically, we are extracting network data from the substance of the text data collections as well as from meta-data pertaining to these text corpora. We use co-occurrence based and syntax based relation extraction techniques; two of the most common relation extraction techniques. The results are compared against ground truth data and each other. We use Context, a tool that supports relation extraction and the joint analysis of text data and network data, for this purpose. Observed trends in the similarities and differences in network structure and behavior depending on the employed method will be discussed. Finally, we show how these methods can be combined to gain a more robust and comprehensive understanding of a network.