N22 Diesner, J., Jiang, M. & Narisetti, S. (2018). Reliable construction of semantic networks based on text data and measurement of effects in text-based networks.

Semantic networks have been widely used, e.g., to represent and study discourse, communication, frames, and information organization in the brain, to analyze organizational behavior, to extract meaning from texts, and to represent text-based information in relational form. However, generating semantic networks based on text data is limited by our insufficient understanding of the relationship between a) mechanisms for constructing or extracting networks and b) suitable methods and metrics for analyzing them.

To address issue a), we use benchmark text corpora that had been annotated for entities and relations (ground truth data), and apply Natural Language Processing (NLP) techniques to the non-annotated text bodies to identify nodes and edges based on lexical, semantic, syntactic, and statistical features. Comparing the resulting data to the ground truth data (in terms of standard network metrics and top n nodes per centrality metric), we perform a comprehensive accuracy assessment of node and edge detection techniques for semantic network construction. We conduct an error analysis to examine false positives and false negatives. The ultimate goal with this work is to identify best practices for generating reliable text-based network data while also considering context, e.g., genre and domain (given in ground truth data). This work identifies the sensitivity of graphs to NLP methods in a systematic and comprehensive way; thereby improving the rigor of research involving semantic networks and related practical applications.

Once text-based networks have been built, standard network metrics are often used to analyze them. The assumption that these metrics are applicable and meaningful for analyzing semantic networks is insufficiently verified. This might be partially because common network metrics were designed to analyze groups of social agents (with some exceptions, e.g., page rank). Texts are different from people, e.g., they are sequential and refer back to previously seen entities, but are also similar to social structure, e.g., they follow norms that impact structure, such as grammar. We will outline these commonalities and differences. Prior work has associated values of several network metrics with categories of words, and has begun to consider text structure to define network metrics. We expand this work and address the aforementioned issue b) by exploring the construction of network metrics that account for the method of network extraction from text data and related network properties. This research can mature and advance semantic network analysis, and is extensible to measure linguistic effects in social networks.