

N11 Diesner, J., & Carley, K.M. (2010). Computational integration of network theory and topic modeling for investigating the relationship between socio-technical networks and innovation in the EU.

When text data pertaining to socio-technical networks are available, these texts are often either analyzed separately from the network data, or are reduced to the fact and frequency of the flow of data or objects between nodes. Examples for the joint availability of text data and network data include answers to open questions in classical network surveys, social media such as emails, blogs, and wikis, and the semantic web. Previous research on the relationship between language and networks suggests an impact of the position of individuals in the network on their motivation and ability to induce innovation and change in socio-technical networks. We present our findings from a study in which we empirically tested this relationship for the case of research proposal that were granted funding by the European Union under the Framework Programmes and a methodology that we developed in order to facilitate this type of studies. This methodology computationally integrates network theory and topic modeling, an unsupervised machine learning technique that reduces the dimensionality of text data to sets of semantically related words, such that network data are enriched through information from text data and vice versa. Our approach is based on prior work that assumes not only texts, but also authors and other types of entities and metadata to have probability distributions over topics (Mimno & McCallum 2008). We extend this notion by abstracting away from the level of individual authors and collaborators to the structural role level, where the actual role is defined by network theory.