

N12 Diesner, J., & Carley, K.M. (2011). Extraction and validation of socio-technical network data about the Sudan from text corpora.

In 2005, the Government of Sudan and the Sudan People's Liberation Movement (SPLM) signed a Comprehensive Peace Agreement. In 2011, Southern Sudan will hold a referendum regarding its independence from Northern Sudan. Network data representing interactions in this socio-technical system during this six year time period can help us to understand the development of culture and conflicts in this region. Since such data is hard to collect through classic methods such as surveys, we use Relation Extraction methods to approximate network data from publically available news coverage on Sudan. We will report on how we utilize theoretically grounded, lexicalized features and feedback loops with subject matter experts to adjust our relation extraction technology to this domain. This technology uses a model that we trained via supervised machine learning, the classifier used for that are Conditional Random Fields. We will present our results from analyzing the retrieved socio-technical network that comprises tribes, issues and resources.