

## Abstract

With the number of posts on social media platforms being orders of magnitude larger than survey responses collected by sociologists, Empirical Sociology faces increasing challenges. Our goal with this work is to meet these challenges by realizing the opportunities which digital media presents. In contrast to survey responses, digital media is available in large quantities and collectable at relatively low costs. In three application scenarios we develop probabilistic models and inference schemas to distill latent semantic frames, providing a macroscopic view on individual processes that would have remained hidden to conventional methods. The first application scenario addresses the question whether Bayesian models are a valid assumption on human social interaction at the example of opinion formation in social networks. In the next application scenario, inference strategies for the learning of semantically meaningful representations from large collections of news articles and scholarly publications are at the center. The final application scenario extends these inference strategies to grouped and temporally ordered microblog entries. In each of these application scenarios, we build on formal model definitions and follow the rules of probability to reason within them, providing a new rigor to the study of social phenomena.