

Social Science Data Analytics Guest Speaker Series

Sociolinguistic Structure Induction: Language and Social Networks

Friday, January 29, 2016

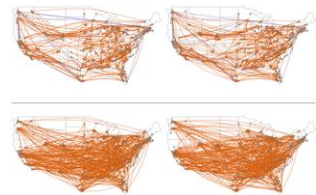
10:00 am

104 South Kedzie Hall



Jacob Eisenstein
Assistant Professor
School of Interactive
Computing
Georgia Tech

Language interacts with social structures ranging from local social networks to large-scale social categories. Computational modeling of language's social dimension offers two potential benefits. First, with the rise of text data sources that include rich social metadata, there is the potential to build a new generation of computational linguistic methodologies for social scientific analysis. I will describe two such efforts: (1) modeling local signed social networks in relation to formality of address; (2) inducing large-scale pathways of sociocultural influence from the spread of new linguistic styles in social media. Methodologically, these projects employ probabilistic factor-graph models, which provide a unified framework for treating social and linguistic data. Second, contemporary language technology is bedeviled by sociolinguistic variation, which complicates efforts to search, mine, and translate text. I will present our recent work on leveraging social network community detection to make document classification more robust to sociolinguistic variation, building on recent progress on task-specific word embeddings.



The Social Science Data Analytics (SSDA) initiative facilitates research and training that harness computational advances to better understand dynamic multi-scale social problems.