

# Welcome to CTASS!

10/11/2019

Student group for...

**C**omputational

**T**ext

**A**nalysis in the

**S**ocial

**S**ciences

# What are we about?

Building an interdisciplinary community

Sharing knowledge and building a common vocabulary

Providing support

Expanding the acceptance and legitimacy of text analysis methods in the social sciences

# What to expect from...

**Sessions** - Some background information and lively discussion

**Workshops** - More guided exposure to common tools and a chance to work with other CTASS members

**Virtual Poster Sessions** - Opportunity to share your work and see what everyone else is up to

**Faculty Speakers** - Broad exposure to actual uses of text analysis across the social sciences

**Social Half-Hours** - A casual opportunity to get to know other CTASS members

**Hackathons** - Group problem solving to develop new skills and refine old ones

# Austin van Loon

PhD Candidate in Sociology  
President of CTASS (Autumn)



- Political Polarization
- Organizational Culture
- Social Networks

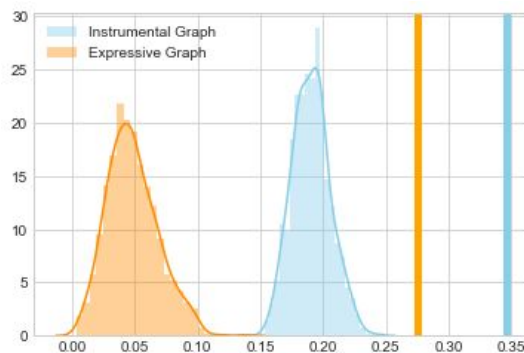
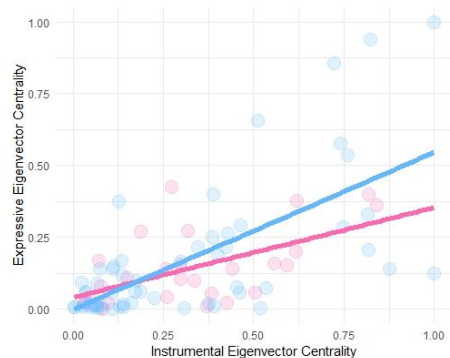
# Gender-based Allocative Discrimination in Organizations

(with Kata Mueller-Gastelle)

**RQ:** How do organizational processes combine with perceptual biases to result in gender-based allocative discrimination?

**Data:** Enron Email Corpus

**Hypothesis:** Women end up excluded in either work email networks or social email networks but not both



# Klint Kanopka

PhD Candidate in GSE  
President of CTASS (Winter)



- Educational Measurement
- Psychometrics
- Deep Learning

# The Fragility of Scoring Decisions in Automatic Essay Grading Software

**RQ:** How do automatic essay scoring (AES) systems make scoring decisions and can better understanding these decisions help develop “gaming” strategies?

**Data:** The Hewlett Foundation AES Kaggle Competition

**Preliminary Finding:** Seemingly high-performing AES systems can make scoring decisions based on single words and randomly inserting these words into low-scoring essays can result in better scores.





# Sheridan Stewart

PhD Candidate in Sociology  
President of CTASS (Spring)



- Meaning of work
- Modernity and cultural change
- Stratification

# “Getting” a Job: Social Position and the Experience and Meaning of Work

**RQ:** How do occupations “hang together” with respect to the meaning of work?

**Data:** ~3 million company reviews from Glassdoor, Inc.

**Preliminary analysis:** Features that predict the occupation of the reviewer

## Text with highlighted words

lot bright dedicated hard work people  
committed customer local management  
recognize reward top performer local  
management seem supportive internal move  
provide people challenging growth  
opportunity general local management  
supportive achieve good work life balance see  
downside

## Text with highlighted words

long term vision planning process  
improvement always talk seldom ever execute  
seem primarily due lack adequate resource  
support upper management abundance work  
local management supportive allow people  
pursue healthy work life balance amount  
work need get combine number people  
experience level make virtually impossible  
people achieve

Cashier

Software Engineer



# Coming up in the autumn quarter...

October 25th :

- CTASS Social Half-hour (10:00 AM - 10:30 AM)
- Amir Goldberg (Stanford GSB; 10:30 AM - 11:30 AM)

November 8th:

- Session: “From Theory to Measurement” (10:30 AM - 11:30 AM)

November 22nd:

- Workshop: “Web Scraping, Pre-Processing, and the Garden of Forking Paths” (10:30 AM - 11:30 AM)

December 6th:

- Virtual Poster Session: Greet and Treat (10:00 AM - 10:30 AM)
- Guest speaker: Brandon Stewart (Princeton Sociology; 10:30 AM - 11:30 AM)
- CTASS BBQ (4:00 PM - 6:00 PM)

# Coming up in the winter quarter...

Session: “Topic Models”

Session: “Word Embeddings”

Faculty Speaker + CTASS social half-hour

Faculty speaker: Jen Pan (Stanford Communication) + Yak and Snack (Virtual Poster Session)

Workshop: “The Twitter API and Common Tools for Sentiment Analysis, Topic Models, and Word Embeddings”

# Coming up in the spring quarter...

Session: “Deep Learning”

Faculty Speaker + Reveal and Meal (Virtual poster session)

Faculty Speaker + CTASS Social Half-Hour

CTASS Hackathon

(Potentially) CTASS Student Conference

# The Whens and Whys of Text Analysis

# Why NOT Text Analysis?

Language is noisy (“I made her duck”)

Language use is context-dependent

Might not be tied to any reality beyond subjective experience (“I’m upper-middle class”)

No unified model of language generation; requires assumptions that are difficult to validate or that we know/suspect are wrong

# Why Text Analysis?

Massive sources of text data increasingly accessible

- Historical materials (Google Books)
- Everyday communications (text, emails, etc.)
- Digital trace data (social media, web searches, etc.)

Social life (arguably increasingly so) is mediated through text

Very impressive momentum methodologically from computational linguistics and industry

Increased computational resources for complicated models and high-dimensional data



To NLP or not to NLP?  
That is NOT the question!

# When Text Analysis?

Is the text a (behavioral) measure, an outcome, or a thing to be studied in and of itself? How does it compare to alternative sources of data?

Consider volume of data vs. complexity of information (cf. human doing a deep read)

Can you validate conclusions you draw from the text (balancing against the unique contribution of the text) or can the text validate another measure?

How well-trodden is the proposed method, and what are its assumptions?

**<https://tinyurl.com/kss-ctass-10-11>**