

## Discussion

Fwd: Fw: RE: You have to see this paper!

Egami, Fong, Grimmer, Roberts, and Stewart  
Kim, Schein, Desmarais, and Wallach

Sarah B. Bouchat

bouchat@northwestern.edu

14 October 2017

# What We Learned

- (1) Life's Complicated: Think about text causally and relationally
  - (a) AISV
  - (b) Who, What, When (Why?)
- (2) Best Practices
  - (a) Training/Test
  - (b) GiR
- (3) Future research

# How to Make Causal Inferences Using Texts

- ▶ Text can be outcome or treatment in a causal analysis
- ▶ Text is high-dimensional, interpretation/low-dimension representation can result in AISV
- ▶ Function  $g$  maps text to treatment or outcome low-dimensional representations ( $g : T \rightarrow Z$  or  $g : Y \rightarrow Z$ )
- ▶ Separating data into training/test avoids using same data for defining categories as for measuring treatment effect

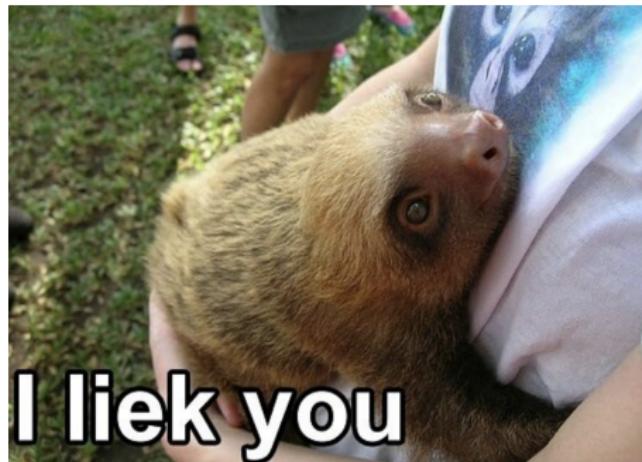


## Questions

- ▶ Should the approach differ depending on whether text is treatment vs. text is outcome?
- ▶ Examples/guidance of the train/test split
- ▶ Guidance about  $g$ 
  - ▶ Can subjects be involved in the specification of  $g$ ?
  - ▶ Could estimate  $g$  wrt attributes of documents or respondents
  - ▶ When would you be comfortable assuming a true  $g$ , if ever?
  - ▶ Evaluate how our sense of mapping comports with subjects'
- ▶ More discussion of handling compound treatment
  - ▶ Example with website design
  - ▶ Characteristics of text that transcend word choice/topics (CFPB example)
  - ▶ Validation with pre-defined topics/categories (e.g., complaint site categories)

## Going forward

- ▶ When can we do this with quasi-experiments?
- ▶ How well could this work as you relax control (e.g., documents of varying lengths)?
- ▶ Side-stepping text as treatment *and* outcome—but this is interesting!

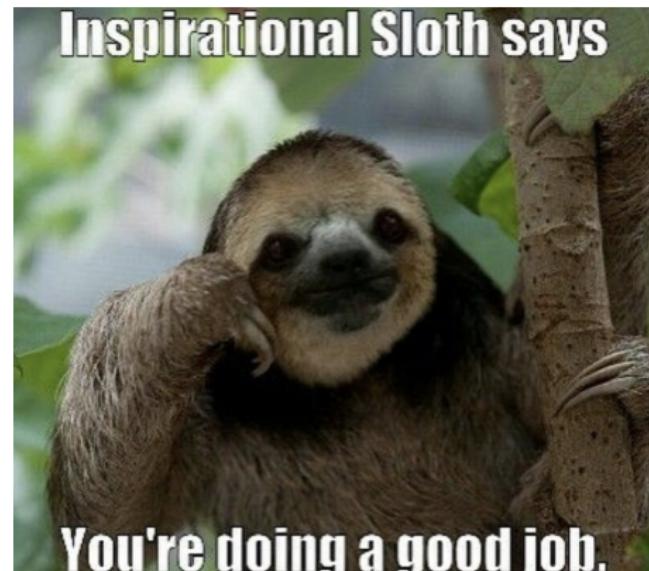


# A Network Model for Dynamic Textual Communications with Application to Government Email Corpora

- ▶ Introducing Interaction-Partitioned Topic Model (IPTM), combining LDA & ERGM
- ▶ Jointly modeling the who, what, when, and to whom of timed text-based relationships
- ▶ Network ties/relationships vary by topic/content

## Getting it Right: It works!

- ▶ Straightforward to use (sampling latent variables and data) and understand (forward and backward samples match)
- ▶ Should really be taken up elsewhere



## Emails

- ▶ Emails: inconsistent audiences? omitted information? missing communication? crisis vs. normal? validation?
- ▶ Competition logic needs a defense
- ▶ Ties & Timestamps

DEAR KEVIN,

I'M SORRY IT'S TAKEN ME TWO YEARS TO REPLY TO YOUR EMAIL. I'VE BUILT UP SO MUCH STRESS AND ANXIETY AROUND MY EMAIL INBOX; IT'S AN UNHEALTHY DYNAMIC WHICH IS MORE PSYCHOLOGICAL THAN TECHNICAL. I'VE TRIED ONE MAGICAL SOLUTION AFTER ANOTHER, AND AS EACH ONE HAS FAILED, DEEP DOWN I'VE GROWN MORE CERTAIN THAT THE PROBLEM ISN'T EMAIL—IT'S ME.

REGARDLESS, THESE ARE MY ISSUES, NOT YOURS; YOU'RE MY FRIEND, AND I OWE YOU THE BASIC COURTESY OF A RESPONSE. I APOLOGIZE FOR MY NEGLECT, AND I HOPE YOU HAVEN'T BEEN TOO HURT BY MY FAILURE TO REPLY.

ANYWAY, I APPRECIATE YOUR INVITATION TO JOIN YOUR PROFESSIONAL NETWORK ON LINKEDIN, BUT I'M AFRAID I MUST DECLINE...



## Tell me more

- ▶ What kinds of questions can we answer?
- ▶ Scale?
- ▶ Different applications
  - ▶ Citation and information propagation (#science!)
  - ▶ Social media shares/comments (mentions?) → works for Facebook but perhaps not Twitter?
- ▶ Network- vs. agent-level inference, extrapolation over time



Thanks!

