

### **Which tasks have been completed?**

I am working on reproducing the Causal Topics Research Paper. The tasks that I have completed so far are:

1. Prepping the NYT data and IEM Prices for processing
  - a. I filtered the documents from 5/2000-8/2000 and saved each document as a line in a txt file. From each document, I only considered paragraphs with Bush and Gore
  - b. For the IEM data, I normalized the values to  $\text{Dem Price}/(\text{Dem Price} + \text{Rep Price})$
2. Refactoring PLSA implementation from MP3 to factor the prior in calculating  $p(w|\theta_j)$  in the M-step
3. Implemented a for loop to run the PLSA model for each date
4. Researched different Granger and Pearson libraries

### **Which tasks are pending?**

1. Implementing topic and word-level causality
  - a. For topic causality, I will use Granger test, with the varying  $\mu$  and 5 day lag
  - b. For word level causality, I will use F measure (for +/- correlation) and Granger test for significance value
2. Implementing Prior Generation
  - a. Have to implement cutoff of words for each topic (making the **TM** list), based on probM
  - b. Need to implement calculating the topic prior
3. Testing and comparing my results to the experiment results

### **Are you facing any challenges?**

1. I am not sure if I should use the background model for the PLSA algorithm
2. I am not sure what value to use for probM