Project Proposal for Team JEM

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.

   Team JEM is made up of the following three members. The team coordinator/leader is Zhangzhou Yu (zy37@illinois.edu).
   - Zhangzhou Yu (zy37@illinois.edu),
   - Matthew McCarty (mdm12@illinois.edu),
   - Jack Ma (jma46@illinois.edu)

2. Which paper have you chosen?

   Causal topic modeling
   - Hyun Duk Kim, Malu Castellanos, Meichun Hsu, ChengXiang Zhai, Thomas Rietz, and Daniel Diermeier. 2013. Mining causal topics in text data: Iterative topic modeling with time series feedback. In Proceedings of the 22nd ACM international conference on information & knowledge management (CIKM 2013). ACM, New York, NY, USA, 885-890. DOI=10.1145/2505515.2505612

3. Which programming language do you plan to use?

   Python (as we found a package to do granger causality test).

4. Can you obtain the datasets used in the paper for evaluation?

   We are in the process of gaining access to the text datasets on Linguistic Data Consortium for the New York Times corpus from May through October 2000. The access is pending upon approval by the UIUC admin access, where the professor is actively engaged on this issue. Currently, we are able to access prices from the Iowa Electronic Markets (IEM) 2000 Presidential Winner-Takes-All Market as well as the stock prices of Apple and American Airlines during that time period.

5. If you answer “no” to Question 4, can you obtain a similar dataset (e.g. a more recent version of the same dataset, or another dataset that is similar in nature)?

   If we are unable to access the text data on the Linguistic Data Consortium for the New York Times Corpus for the 2000 presidential election, we will investigate using the online New York Times archives (or some other accessible newspaper archive) for the 2016 presidential election articles along with stock prices of American Airlines and Apple of the same time period.

6. If you answer “no” to Questions 4 & 5, how are you going to demonstrate that you have successfully reproduced the method introduced in the paper?

   To demonstrate that we have successfully reproduced the method introduced in the paper, we will output significant topic lists that affected the 2016 presidential election and stock

prices of American Airlines and Apple in 2016 respectively. We will check the results of the top topic lists against what truly happened in that time frame.