NLP4CSS Write-up

Sangmin Woo

Problem 3b) In my attempts to improve the Z coefficient, I varied the number of topics and also tried an implementation using LDA to recover the topics instead of NMF as before. The results are as follows:

NMF

10-topic: 0.5393

20-topic: 0.4619

22-topic: 0.4692

23-topic: 0.4571

25-topic: 0.4668

30-topic: 0.4670

40-topic: 0.4707

50-topic: 0.4730

LDA

10-topic: 0.7025

20-topic: 0.5949

22-topic: 0.6969

23-topic: 0.6673

25-topic: 0.6720

30-topic: 0.6300

40-topic: 0.7390

50-topic: 0.7212

The closest coefficient came with the NMF setting on 23 topics.

Problem 5)

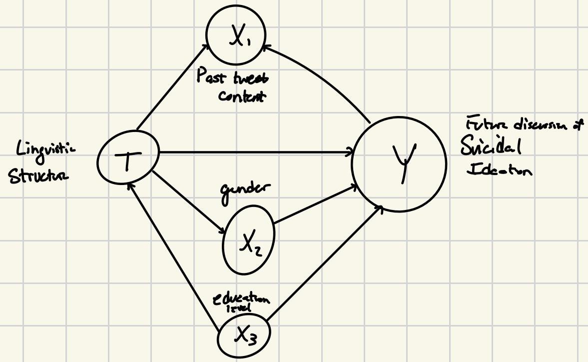
1. Treatment: Linguistic structures/patterns used in tweets

   Outcome: Future discussion of suicidal ideation

   Confounders: Content of past tweets, gender of user, level of education

The population of users "tweeting about mental health" is very vague – what is it that they are tweeting? Someone who is tweeting about mental health advocacy/education and someone tweeting about their poor mental health are likely to differ in rates of future suicidal ideation. Furthermore, suicide rate varies between gender and other demographics, like education level, which also affects linguistic structure in tweets.

Causal Graph:



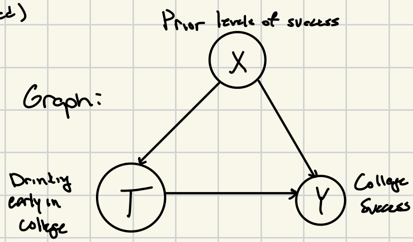
2. Treatment: Drinking early in college

   Outcome: College success (habits, relationships, criminal activity)

   Confounder: Levels of success prior to drinking

Someone who has poor habits and social relationships to begin with may be more likely to turn to drinking as a coping mechanism (and maintain low levels of "success" as defined)

Causal Graph:



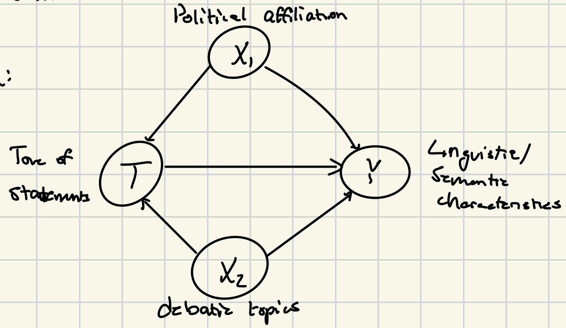
3. Treatment: Tones of statements made during politically charged debates

Outcome: changes in linguistic/semantic characteristics in subsequent responses

Confounders: Political affiliation, debate topics

Tone and linguistic/semantic characteristic can differ across parties, and much moreso when considering the affiliations of the debaters (and if they differ). Furthermore, some topics can be much more sensitive and important, affecting both tone and linguistic/semantic characteristics as well.

Causal Graph:



4. Treatment: AI article author gender

Outcome: Citation count

Confounder: Article quality/relevance

How well-written, and more importantly how relevant an article is will definitely impact citation count, e.g. more articles are written about more relevant topics, which cite other articles in the same field.

Causal Graph: