DS 785 Data Science Capstone

Capstone Activities Report

Name: Spencer Swartz

Project Title: Social Media Content Optimization and Data Driven Brand Advocate Identification

Activity Report #3

Agenda/goals:

At the completion of this activity period to remain on track for the project it is expected that I was able to complete phase 1 of the project (A visualization answering basic questions related to BA's twitter campaigns) and the identification of a model in which to use to identify brand advocates within the platform.

Contacts Made/Method of Contact:

I have contacted Bart Watson (BA) via email with a similar update notifying him of the progress made in the data analysis period on 3/19/18.

Resources and Investigation Methods:

As stated in the previous report, Tableau products played a big factor in this portion of the project. I will not go into detail again about the ins and out of the Tableau products but would advise the reader to look back on Activity Report #2 for those interested or in need of a reminder. Instead I will focus on the research done to identify an algorithm for brand advocacy.

Over the week I was able to identify that there had been some research related to the problem proposed in phase 2 of this project (identifying brand advocates). A large majority of that research seemed to have a common node within Dr. Amit Goyal from The University of British Columbia, Two articles including him as an author were identified as informative and useful (Learning Influence Probabilities In Social Networks, and A Data-Based Approach to Social Influence Maximization). The final modeling will likely include subsets of both as well as basic network statistics and page rank as inputs.

Progress:

The project progress may be a little behind at the moment, this is because the retrieval of additional data has taken longer than expected. That being said phase 1 was completed on schedule and the corresponding dashboard was provided to BA. [The dashboard can be viewed at this link](#).

To go into more detail as to why progress has slowed one would have to look into the documentation related to the Twitter API. To move forward into looking at brand advocacy additional data was needed to be pulled. Specifically, it was needed to look at who followed individuals that retweeted BA tweets as well as who their followers where. This is because some of BA's network may be closely knit or connected, if this is the case then many of the retweeter's of BA's network likely are not spreading there message far. So if BA can identify those who are more desperate in their followers the there is a greater likelihood in casting a wider net.

What has been the limiting factor in this part of the project is collecting the data as stated before. There are API calls that return profiles of followers of a given users, but only at 100 profiles at a time, and only 15 calls per minute. This has proven to be difficult as there are in the range of 30,000 unique individuals who have varying number of followers. In a somewhat lenient estimation at this level access and calls to the API I would likely take more than 20 complete days to collect all of the data. This is simply not possible in the scope of this project. Because of this I will begin to treat this part of the project as a "proof of concept" and will collect as much real data as possible while realizing that any insights identified will be incomplete. BA is aware of this situation.

Achievements:

In this reporting period I have completed phase 1 of the project and delivered a dashboard that helps to answer the basic questions outlined in the project description. I have also identified specific aspects of a potential model to be implemented in phase 2 of the project. Data collection has also begun under the pretense of the above information.

Questions:

In reference to the news above, I would ask if the professor agrees that the strategy going forward with part two is an acceptable deviation from the initial plan and scope of the capstone project.

Next step:

The next reporting period will be focused on continuing to gather data related to phase 2 as well as implementing the brad advocate model.

Additionally writing will begin outlining the project to date for use in the final documentation.