The Myth of Voters' Sentiment Change
Analyzing Twitter Sentiment during 2016 First Presidential Debate*

Haju Kim, Kaiping Chen, Michael Fairley, Yifei Wu

September 12, 2017

1 Research Question

What explains the sentiment *spikes* among voters during the presidential debate? Are voters' sentiment spikes driven by the topics they resonate with or by politicians' communication tactics during the debate?

This research question is important in both theory and practice. Theoretically, it offers a response to the long debated topic in the field of political science on how voters' make decisions. Practically, it offers implications for politicians and practitioners working for campaigns by revealing what matters for voters' sentiment change and thus guides the design of campaign strategy.

2 Related Literature

The topic of what drives voters' choice and sentiment has been a key topic in the field of political science. One school argues that voter's choice is driven by the demand side, i.e., voter's attributes, including voters' party ID (Miller 1991), voters' evaluation of the incumbent's previous performance (Fiorina 1978), candidate's personal traits(Caprara et al. 2006),and voters' interest in the topics[1]. Therefore, the demand side school holds that it is very rare for political campaigns to change voters' minds and campaigns and political debates at most reinforce preexisting beliefs (Iyengar and Simon 2000). Thus, the effect of the campaign is to arise resonance from voters on their topics of interest and on voters' anticipation toward a candidate.

Nevertheless, the other school argues that voter's choice is driven by the supply side, including the frame of campaign messages, communication style from politicians and media portrayal of candidates (Rosenberg et al. 1986;Popkin 1994. Among the scholars in this school, one popular strategy recommended for the politicians is to talk about issues that they have the ownership (Petrocik 1996). The theory of "Issue Ownership" notes that Republican party and Democratic party have their own competitive advantages on certain issues. For instance, Republican candidates are more trustworthy among voters on national security issues while Democratic candidates are more trustworthy on issues of race and gender. Thus, talking about owned issues make it easier to win popularity among voters.

Therefore, whether voters' sentiments are driven by the demand side (the topic they resonate with) or driven by the supply side (the issue owned by the politicians) is still a myth in the field.

Moreover, although there has been extensive study on how political campaigns influence voter choice, there are fewer studies on exploring the effect of **presidential debate** on voters' sentiments and there are even fewer studies on exploring the potential voters on **online public forums**(Diakopoulos and Shamma 2010).

To respond to the supply-demand side debate and to extend the study of campaign strategies to presidential debates on online users, our paper examine Twitter users' sentiment during the first presidential debate in 2016 and investigates what the features of those sentiment spike moments are during the debate.

*We would love to sincerely thank Professor Sharad Goel and Imanol Arrieta Ibarra for offering us valuable suggestions during the whole project process.

[1] http://graphics.wsj.com/house-midterm-elections-facebook/; http://study.com/academy/lesson/factors-that-influence-voters-during-presidential-elections.html

3 Method

3.1 Data Source

We obtained two separate data sources: a C-SPAN transcription of the debate and tweets collected between Sep. 25th 2016 8:00PM ET – Sep. 27th 2016 7:59PM ET containing any of the following keywords: *#debate, #debates, #debates2016, #debatenight, debate, hillary, trump, election* (Bencina 2016). There were 25,064,377 tweets in total, with about 70% of these being retweets. The transcript was broken into segments with a median duration of 11 seconds each. Each segment identified the speaker and the UTC time at which the segment began. In total there were 486 segments during the 90 minute debate. These data were stored on Google Cloud's Big Query, which we queried to obtain a local dataset that we could process.

3.2 Data Processing

To turn the unstructured raw data into meaningful structured data, we filtered the tweets by removing retweets, removing non-ASCII characters from the text, and using a language detector to select only English tweets (*langdetect 1.0.7* 2016). We also used the Python nltk package to calculate a sentiment score for each tweet text (*NLTK* 2016). Due to the large number of tweets, we accomplished this using MapReduce on Amazon EMR. The tweets were inputted in json format and outputted as a .tsv format with each line containing the timestamp, user id, hashtags, urls, text and geographical coordinates (if available) for a tweet. This reduced the total number of tweets down to 6,717,333, so we could do further analysis on our local machine.

3.2.1 Bot Detection

We defined highly automated accounts as accounts which post at least 50 tweets a day (excluding retweets). These accounts are likely to be social media bots, or algorithm-driven entities that are known to be active in Twitter on political discussions. Among 2,039,535 accounts that posted tweet(s) during the first presidential debate, 4,864 accounts were identified as highly automated, and these accounts generated about 7.46% of the tweets. Since our population of interest is actual humans, tweets that were posted by these accounts were excluded from further analysis.

3.2.2 Political Orientation Identification

We used simple heuristics to identify the political orientation of the tweets in our dataset. Each tweet was coded and counted if it contained at least one of 115 pro-Trump (or anti-Hillary) hashtags, 145 pro-Hillary (or anti-Trump) hashtags, or 16 neutral hashtags. The hashtag list was based on the previous analyses by Howard et al and manual identification conducted for non-identified hashtags that had more than 200 instances in our dataset (Howard and Woolley 2016). Each political faction was coded as a binary variable. If a tweet contained multiple hashtags that support one particular candidate, this method counted this tweet only once. If a tweet contained hashtags for more than one political faction, the tweet was assigned to all relevant hashtag categories. Then we assigned tweets that contain only (1) hashtags for supporting a particular candidate or (2) hashtags for supporting a particular candidate and neutral hashtags to that political faction (pro-Trump or pro-Hillary). Table 1 summarizes the distribution of tweets in our dataset. This is a very strict and conservative assignment of sentiments, less prone to high misclassification error that we observed when we initially attempted to use NLTK sentiment score to code each tweet's sentiment into pro-Hillary and pro-Trump groups. Tweets that contained mixed hashtags for both candidates, which consist of about 1.8% of the dataset, were assumed to have ambiguous sentiment and were excluded from the analysis.

Table 1: Twitter Activity during the First Presidential Debate

	All Tweets in Sample	
	N	%
Neutral	2322648	84.4
Pro-Hillary	167942	6.1
Pro-Trump	212580	7.7
Trump-Hillary	21237	0.8
Trump-Hillary-Neutral	26574	1.0
Total	2750981	100.0

3.3 Analysis Method

We decided to use the proportion of pro-candidate hashtags as a proxy for the sentiment towards the candidate. Identifying the sentiment of tweets based on hashtags has an advantage of capturing the candidate to which the tweet is referring to and the sentiment towards the candidate. That less than 1.8% of tweets contained mixed hashtags suggests that pro-candidate (or anti-candidate) hashtags can serve as a reliable proxy for the sentiment. In order to compute the relative sentiment towards the candidate for different time periods, we combined tweets over 1-minute intervals and then measured the positive sentiment toward a candidate by computing the proportion of pro-Candidate tweets for each interval. For example, pro-Trump sentiments for the 1-minute interval starting at 2016-09-26 21:05 EDT is measured by dividing the total pro-Trump tweets by the total number of pro-Trump, pro-Hillary, and neutral tweets generated between 2016-09-26 21:05 EDT and 2016-09-26 21:06 EDT (Figure 1). Then we measured the relative advantage of the candidate for each time interval by computing the difference between the two proportions ([proportion of pro-Clinton tweets] - [proportion of pro-Trump tweets]). When the proportion difference is greater than 0, for example, tweets about Hillary reflected a more positive sentiment than tweets about Trump (Figure 6)

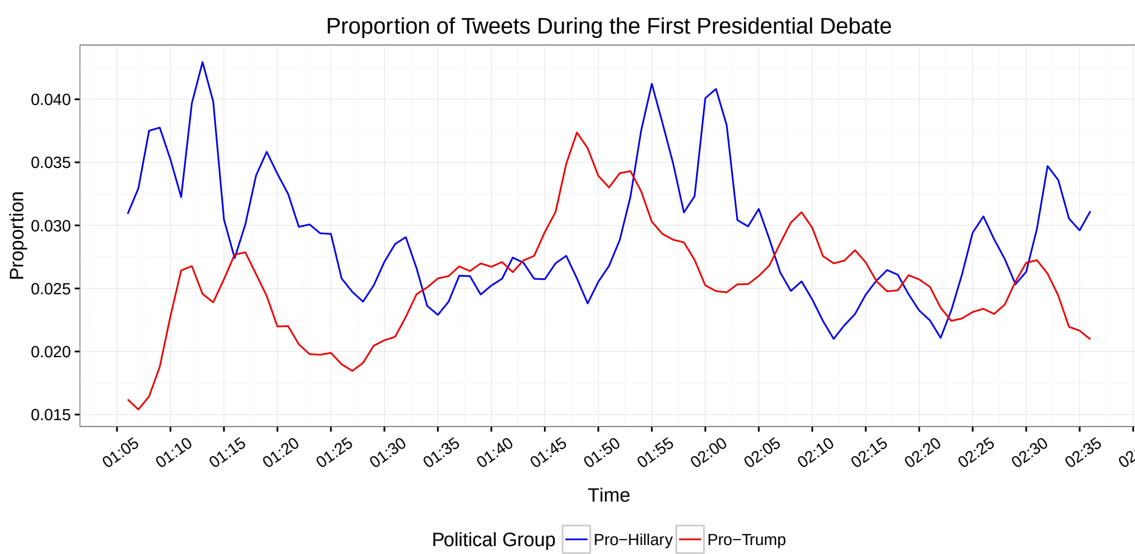

Figure 1: Proportion of Tweets During the First Presidential Debate

We identified the sentiment spikes by (1) looking at the moments when the relative proportion of candidate-supporting tweets were the highest for each candidate and (2) finding the moments when the relative sentiment towards each candidate differed the most in terms of absolute value.

Since our unit of analysis is each minute during the debate, before performing the topic model, we decomposed all of our tweets into the tweets by minute. Then, we utilized MALLET classification (with hyperparameter optimization) to generate the top 10 topics for all of our tweets and to compute the proportion of each of the top 10 topics for each minute of debate tweets. Therefore, we know what are the key words for the top ten topics. We also know the change in proportion of each of the top 10 topics during each minute of the debate.

4 Results

4.1 Basic Statistics

During the first presidential debate, there were 3,164,350 tweets generated by non highly-automated accounts and 1,886,792 tweets (59.6%) for which the political sentiment could be determined. The distribution of the tweets by user group over the time of the debate is shown in Figure 4 Table 2 shows that among 3,164,350 tweets, about about 62.6 % of the tweets contained at least one hashtag of any kind. The overall volume for the neutral group was much greater than the overall volume for either pro-Hillary and pro-Trump groups. While it would have been interesting to analyze the data geographically, not enough users included geotag information in their profiles. Less than 0.1% of the tweets contained geographic information.

Table 2: Twitter Activity during the First Presidential Debate

Variable	**Levels**	**n**	**%**
Hashtag Inclusion	TRUE	1980620	62.6
	FALSE	1183730	37.4
Group (Identified Only)	Neutral	1698185	90.0
	Pro-Hillary	91012	4.8
	Pro-Trump	80648	4.3
	Trump-Hillary	4461	0.2
	Trump-Hillary-Neutral	12486	0.7
Geographic Information	TRUE	2764	0.1
	FALSE	3161586	99.9

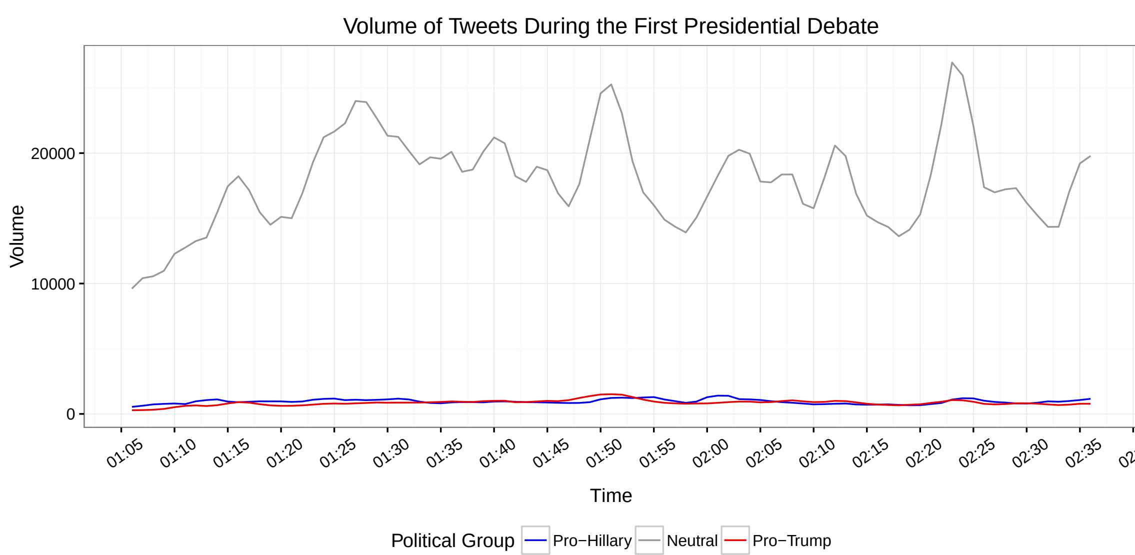

Figure 2: Volume of Tweets Per Minute During the First Presidential Debate

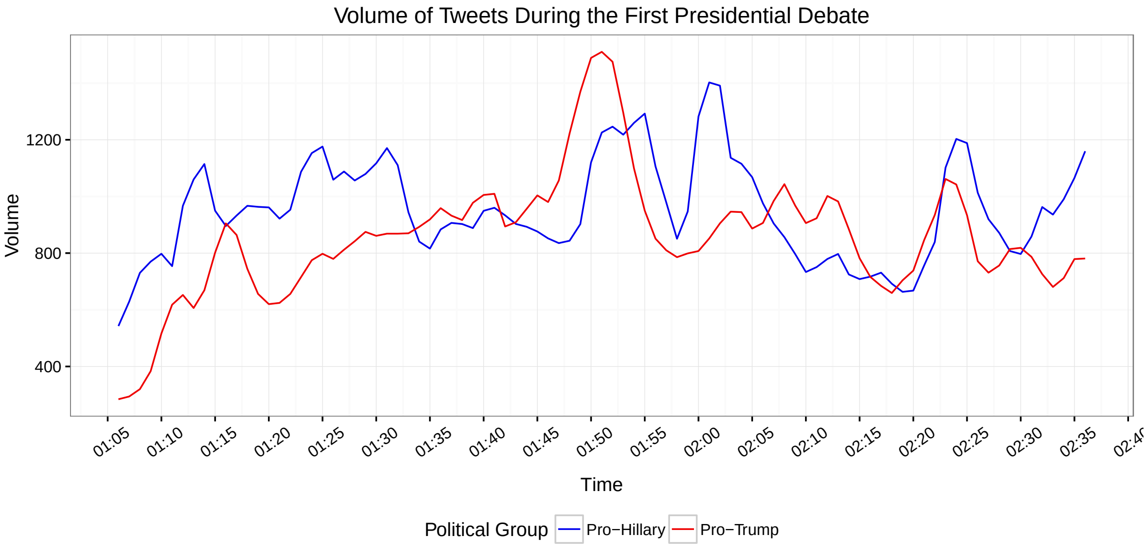

Figure 3: Volume of Tweets Per Minute During the First Presidential Debate

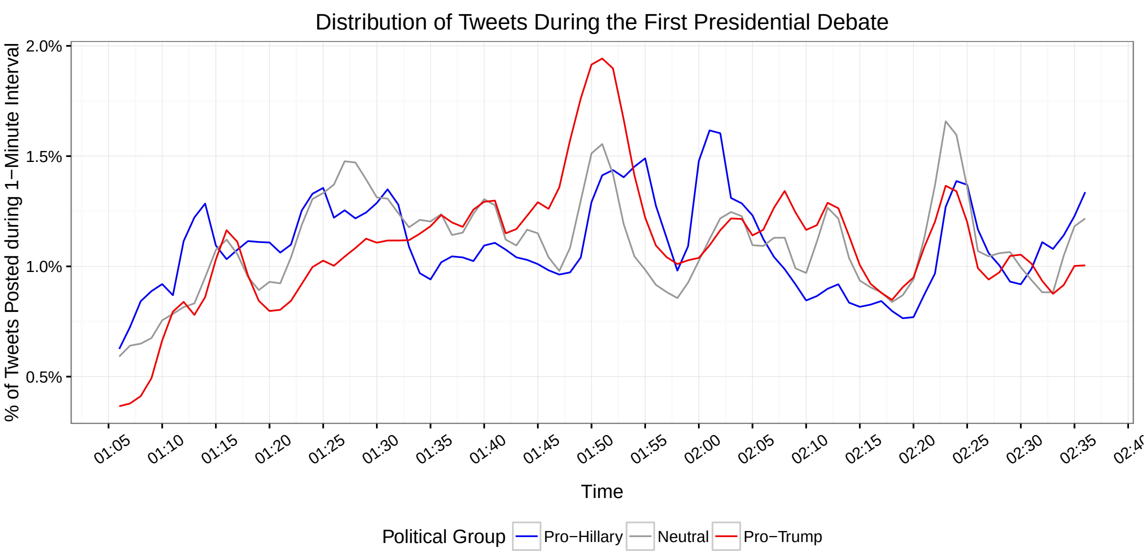

Figure 4: Distribution over Time of Tweets During the First Presidential Debate

4.2 Key Findings

4.2.1 Conceptualize Key Topics

Using the MALLET document classification algorithm, we obtained the top 10 topics and their related key words for all the tweets during the debate. Table 3 reports the key words under the top ten topics. Figure 5 shows the change in proportion of each top 10 topics during the debate (by minute).

Table 3: Top 10 Topics from Tweets During the Debate

Topic ID	Keywords
0	tax taxes emails returns release makes country smart pay world paying trump he's airports federal money talking audit income debatenight
1	debate lester hillary isis debatenight reality plan live interrupting stop talking moderator fighting he's tax donald fact facts book life
2	jobs debatenight mexico trump question china small hillary loan donald business back answer red talking trickle million debate watch trumped
3	vote winning nuclear nbc cast president nbcwashington told words china plan matter secret lester iran world talking tweet deal korea
4	stamina hillary nice doesn't experience vote lester bad women nbc cast winning told rosie nbcwashington question it's ads talking she's
5	cyber isis vote nbc cast nbcwashington told winning security endorsed russia oil son talking wrong bed fat question iraq hackers
6	temperament winning war hannity sean iraq nato told vote nbc nbcwashington cast lester call talking moderator asset he's record howard
7	trump debatenight debates debate hillary donald clinton realdonaldtrump i'm hillaryclinton don't it's watching time can't trump's people good https://t.co presidential
8	stop law order frisk race people police black trump guns amp talking gun chicago debatenight unconstitutional racial question back wrong
9	prepared president racist birth obama birther certificate black vote nbc winning cast told question nbcwashington talking african community answer good

Taking a first glance at the key words under each topic, we found some apparent key topic issues and, mentioning of the country and International organization names and candidate scandal issues such as emails and tax returns. To make sense of the key words under each topic, we further performed two things: we examined a sample of tweets under each of the top ten topics to understand what the key words are really talking about in details and their relationships; we searched these key words in the debate transcript to understand what candidates were talking about in terms of these key words.

The first presidential debate is organized along four topic areas: the prosperity of the country regarding jobs and business, the tax issue, the national security issue regarding cyber security, border security and terrorist attacks, and the race issue. We coded each speech act in the debate transcript along these four topics based on moderator Lester's signaling that the next question is about what topic. Figure 5 shows that for some top topics, they reached the highest proportion of certain debate topics. Topic 3 and topic 8 are popular contents among twitter users during the whole debate.

We labeled the category of each of the top ten topics to see whether the content of the tweets are similar to what is talked about during the debate. Table 4 indicates that the content of the tweets (as shown in the column topic_category) resonate with the debate topics which are jobs, tax, race, and security. Twitter users had a heated discussion about tax, fighting ISIS, the job problem facing the nation, and race issue regarding the black communities. These are issues closely related to people's prosperity and security in their daily lives. It is worth noting that topic_5 is not an issue about prosperity and security that closely matters to people's lives. Rather, it is an issue about politicians' experience. This topic category identified in the tweets also resonated with the nature of the debate, where politicians not only talk about policy issues but also question and attack each others' credentials for leading the nation.

We also found that some countries are specifically mentioned with the topic. For instance, when tweets

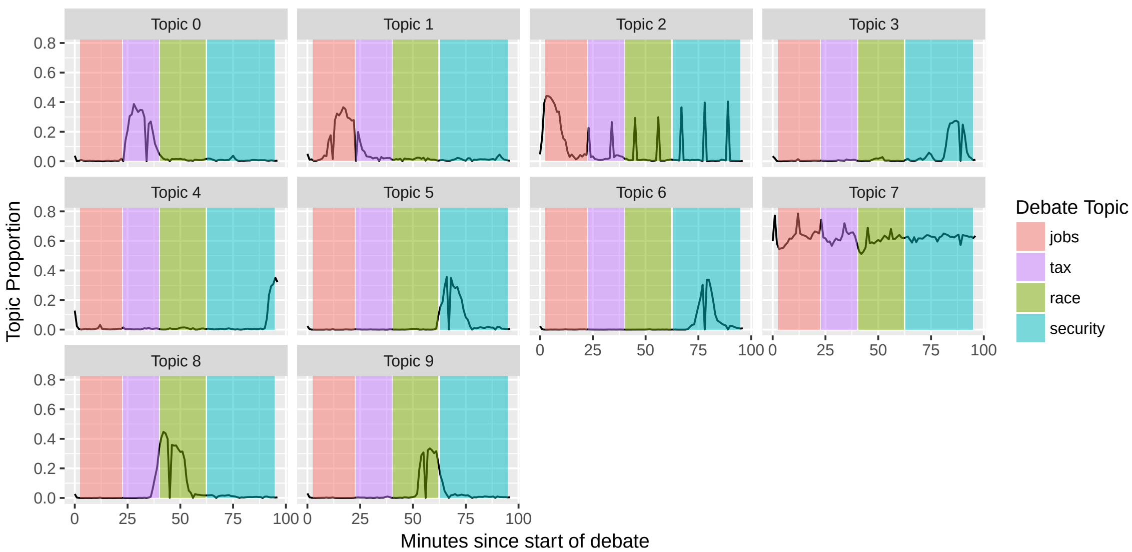

Figure 5: Proportion of each topic during debate

Table 4: Analyzing Topic Features

Topic ID	Domain	Country/Organization	Candidate	Personal Issue
0	Government Tax and Spend	world	trump	emails; tax form release
1	ISIS		trump	plan
2	Job	Mexico; China	trump;hillary	
3	Nuclear Issue	China;Iran		
4	Experience of Politicians		hillary	experience
5	Domestic Security	Iraq and Russia		hackers
6	International Relationship	Iraq, NATO		
7	Debate in General		trump;hillary	
8	Race Issue (policy-oriented)		trump	
9	Race Issue (racist-oriented)		obama	birth-certificate

are about job issue, many of them mentioned Mexico and China. Searching the words Mexico and China in the debate transcripts, we found that it was the time when Trump stressed that American's jobs and business had fled to Mexico and China. Again, the tweets contents resonated with what the politicians talked about and even with what countries the politicians mentioned during the debate. In a similar way, certain countries were associated with other topics: nuclear issues (Iran), domestic security (Russia and Iraq) and international relationships (NATO).

Besides identifying the dimension of topic category and country mentions, we also identified the dimension of personal issues from the tweets. We found that the content of the tweets are not just about domain issues such as jobs, tax, security, users also extensively discussed the personal issues of the candidates. For instance, the word "release tax forms" appears in topic_0, the word "plan" appears together with the topic of ISIS in topic_1, and the word "hackers" appears together with domestic security in topic_5. These words are closely related to politicians' scandals and their credentials to lead the nation well. For instance, topic_1 is about ISIS and Trump is mentioned frequently with this topic and the word in the personal issue column is "plan". Searching the key word ISIS in the debate transcript, the speech acts are about Hillary severely criticizing Trump about his vague plan to deal with ISIS. Differently, topic_4 is about Experience of Politicians and Hillary is mentioned frequently with this topic and the word in the personal issue column is indeed "experience". Searching the word experience in the debate transcript, the speech acts are about Trump criticizing severely about Hillary's bad experiences in her political career.

Therefore, we see that the contents of the tweets not only resonated with the specific domain topics in the debate, but also the countries and the politicians mentioned, and even the personal attacks during the debate.

4.2.2 Examining Topic Pattern in Key Debate Moments

We not only examined whether the content of the tweets resonated with what politicians talked about during the debate, we also explored how twitter users' sentiment toward the candidates resonates with the debate. We investigated whether there are some topic patterns during those spiking sentiment moments and what actually happened in the debate during those spiking sentiment moments.

As we mentioned in the Method section, we calculated the volume of different types of hashtags during each minute in the debate. The magnitude of the volume allows us to identify the sentiment spiking moments during the debate. By spiking, we mean both the positive and the negative spike, respectively for each candidate. Since politics is a win or lose game, one crucial thing to measure is a candidate's relative performance to the other. Therefore, we also identified the spiking moments when there were huge contrasts (differences) in the direction of the spikes for the two candidate(Figure 1).

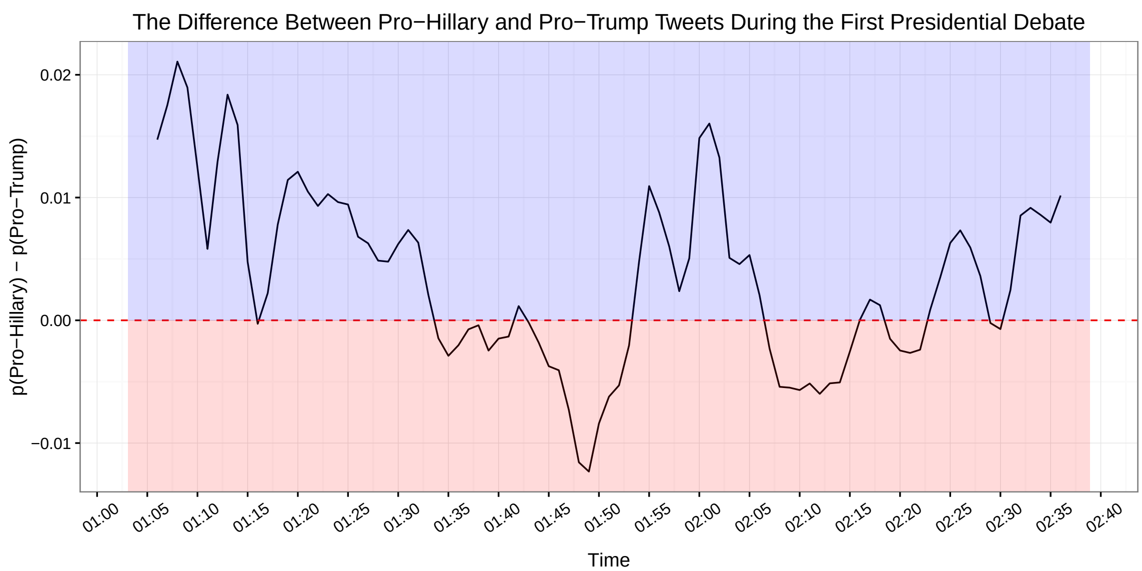

Figure 6: The Difference Between Pro-Hillary and Pro-Trump Tweets During the First Presidential Debate

Table 5 analyzes and reports the spiking moments, the spiking types and the two highest proportion topics in those moments. We have three main findings:

First, we found that there is a concentration of certain topics during those spiking moments. Topic 3, 8, and 9 have the highest proportion in the spiking moments. As known from Table 4, topic 8 and topic 9 are issues about race, with topic 8 focusing more on the policy to treat black communities fair and topic 9 focusing more on racial bias. Topic 3 are issues about nuclear weapons. According to the issue ownership theory(Petrocik 1996), Republican candidates have the advantage on issues of national security and they are regarded as more competent in dealing with these issues. Democratic candidates have the advantage on issues of race and they are regarded as more competent in tackling with these issues. The topic concentration during those spiking moments imply that those comparative advantaged issues receive more sentimental feedback from the Twitter users.

However, the results from our data(Table 5 indicate that it is not always true that candidates are favored in their comparatively advantageous issues. As we see from Table 5, for the race issue (in particular, the race policy), sometimes Hillary won over Trump and sometimes Trump won over Hillary. Also, the volume of pro-Hillary tweets reached peaks in both directions on the race policy issues. The volume of pro-Trump tweets reached peaks also in both directions on the race policy issues. This implies that for the issue of race, democratic candidates such as Hillary do not have the expected advantage predicted by the theory. This

Table 5: Spike Moment, Spike Types and High Proportion Topics

Spiking Minute	Spike Type	Top with highest proportion	Top with second highest proportion
9/27/2016 1:08:00	hillary_beat_trump	topic 8	topic 3
9/27/2016 1:09:00	hillary_beat_trump	topic 8	topic 3
9/27/2016 1:13:00	hillary_beat_trump	topic 8	topic 3
9/27/2016 1:49:00	trump_beat_hillary	topic 8	topic 9
9/27/2016 1:48:00	trump_beat_hillary	topic 8	topic 3
9/27/2016 1:50:00	trump_beat_hillary	topic 8	topic 9
9/27/2016 1:48:00	pro_trump_greatest	topic 8	topic 3
9/27/2016 1:49:00	pro_trump_greatest	topic 8	topic 9
9/27/2016 1:47:00	pro_trump_greatest	topic 8	topic 9
9/27/2016 1:07:00	pro_trump_least	topic 8	topic 3
9/27/2016 1:06:00	pro_trump_least	topic 8	topic 3
9/27/2016 1:08:00	pro_trump_least	topic 8	topic 3
9/27/2016 1:13:00	pro_hillary_greatest	topic 8	topic 3
9/27/2016 1:55:00	pro_hillary_greatest	topic 8	topic 9
9/27/2016 2:01:00	pro_hillary_greatest	topic 8	topic 10
9/27/2016 2:12:00	pro_hillary_least	topic 8	topic 6
9/27/2016 2:22:00	pro_hillary_least	topic 8	topic 7
9/27/2016 2:13:00	pro_hillary_least	topic 8	topic 6

can be explained by the complexities of the topics under the race issue. Examining the key words under the race topic and the speech acts on race in the debate transcript, we found that there are two main issues under race: one is the stop and frisk policy and the other is gun policy. The nature of these issues are very controversial and it is not surprising to observe that sentiment toward candidate on these issues vary.

Although topic 8 has the highest proportion across all spiking moments, the column of number_2_proportion_topic has some variations. For the moments when Hillary beat Trump and pro-Hillary (pro-Trump) volume reached positive (negative) peaks, the majority of the topic is topic 3, which is the nuclear issue. Searching all the speech acts that contain the word nuclear, we found that they are all speech acts Hillary made telling voters that when she was Secretary of State, she made Iran reduce its nuclear material stockpile and formed a coalition with Russia and China to put sanctions on Iran. This implies that Twitter users gave Hillary a high credit for her performance on this issue. For the moments when Trump beat Hillary and pro-Trump (pro-Hillary) volume reached positive (Negative) peaks, the majority of the topics are topic 9 and topic 6, which are issues about International Relationship and racial bias. Searching the words of Iraq, NATO, we found that the speech acts were Trump's criticisms of Obama and democratic party's policy on taking troops from Iraq, which created a vacuum for the growth of ISIS. Instead, America should put more military forces into the regions where ISIS is active. Trump's winning over Hillary on this issue among Twitter users implies that people prefer a strong military force to cope with ISIS, in particular with the rising attacks on American citizens from ISIS in recent years. What is unexpected is that Trump won over Hillary and received positive spikes for topic 9 which is about the racial bias issue. The key words under this topic is Obama's birth certificate. Searching this word in the debate transcript, the speech acts were Trump defending himself that he did a good job to urge Obama to produce his birth certificate and Hillary attacking Trump being a racist. Although Trump is regarded as a "racist", at least for twitter users, they do not punish Trump on this issue. This make us question who are the people on Twitter that tweet during the debate. Evidence has shown that Trump has won the Twitter/Facebook platform battle over Hillary by a large margin[2]. If the users who tweet during the are majority Trump-supporters or at least Hillary-opponents, then Trump's winning over Hillary even on the racial bias issue is not surprising.

The analyses on the topic patterns during the spiking moments and its relationship to the debate indicate that Twitter users do have more sharp sentiment on certain topics than others. However, the sentiment is not driven by whether a candidate owns the topic or not, but driven by how controversial the issue is and

[2] http://www.usatoday.com/story/tech/news/2016/08/04/trump-clinton-social-media-twitter-facobook-youtube-snapchat/87974630/

the ideology of the Twitter users.

5 Conclusion

Through identifying the key topics in Twitter users' sentiment spike moments and relating the topics to the debate transcript, we found that the key topics discussed among the twitter users do resonate with the debate topics.

Yet, the sentiments toward candidates among Twitter users are not driven by the supply side such as what the theory of issue ownership (Petrocik 1996) argues. Hillary won over Trump on national security issue, the issue that is conventionally owned by the Republican party. Similarly, Trump won over Hillary on racial bias issue, the issue that is conventionally owned by the Democratic party. Rather, we found that most sentiment spiking moments concentrate on the controversial topics -race policy, including gun control and stop-frisk policy. These are issues that matter closely to Twitter users' daily lives and therefore people are more likely to pay more attention to them and express stronger opinions on them.

It is important to that we did not study what **influenced** Twitter users' sentiment but studied the topic patterns during those spike moments. The reasons are several folds.

First, it is impossible to study sentiment/attitude change without a controlled experiment. There are so many confounding variables occurring during the debate time that can influence voters' sentiments. Even if we can identify the ideology of each user and focus on studying one group alone, we still cannot answer the question of "influence" because even though debate happens at the same time as when the Twitter users tweet, it does not mean that debate should have an influence on the sentiment.

Second, methodologically, different Twitter users have different reaction speed to the debate, and it is impossible to use a fixed lagging window for all the Twitter users. This is why given the big observational dataset we have, we can only examine the topic patterns during those spiking moments.

The ideal method to study this research question is to conduct a controlled experiment. We can balance the attributes of our samples and randomly assign voters into receiving different types of debate messages. We can ask voters' attitude before and after the experiment to study how debate messages influence their sentiment changes.

References

Bencina, John (2016). *25 Million Presidential Debate Tweets in Google BigQuery*. URL: http://www.jbencina.com/blog/2016/10/06/25-million-presidential-debate-tweets-in-google-big-query/.

Caprara, Gian Vittorio et al. (2006). "Personality and politics: Values, traits, and political choice". In: *Political psychology* 27.1, pp. 1–28.

Diakopoulos, Nicholas A and David A Shamma (2010). "Characterizing Debate Performance via Aggregated Twitter Sentiment". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. New York, NY, USA: ACM, pp. 1195–1198. ISBN: 978-1-60558-929-9. DOI: 10.1145/1753326.1753504. URL: http://doi.acm.org/10.1145/1753326.1753504.

Fiorina, Morris P (1978). "Economic retrospective voting in American national elections: A micro-analysis". In: *American Journal of Political Science*, pp. 426–443.

Howard, Philip N and Samuel C Woolley (2016). "Bots and Automation over Twitter during the First U.S. Presidential Debate". In: 1, pp. 1–4.

Iyengar, Shanto and Adam F Simon (2000). "New perspectives and evidence on political communication and campaign effects". In: *Annual review of psychology* 51.1, pp. 149–169.

langdetect 1.0.7 (2016). URL: https://pypi.python.org/pypi/langdetect.

Miller, Warren E (1991). "Party identification, realignment, and party voting: back to the basics." In: *American Political Science Review* 85.02, pp. 557–568.

NLTK (2016). URL: http://www.nltk.org.

Petrocik, J.R. (1996). "Issue Ownership in Presidential Elections, with a 1980 Case Study". In: 40.3, pp. 825–850. ISSN: 1540-5907.

Popkin, Samuel L (1994). *The reasoning voter: Communication and persuasion in presidential campaigns.* University of Chicago Press.

Rosenberg, Shawn W et al. (1986). "The image and the vote: The effect of candidate presentation on voter preference". In: *American Journal of Political Science*, pp. 108–127.