Sophie Thesis Update 06-10-16

Single Candidate Followers, Connecting Parts I & II, Stance & Virality

1. **What do Single Candidate Followers Look Like?**

Previously:

Users that follow at least one candidate:
13,472

Users that follow *only* one candidate:
6,567

Number of Tweets by Single-Candidate Followers:
24,470

Since we are using this categorization in my thesis (potentially as a proxy for a twitter user's candidate loyalty), I took a qualitative look at the top-10 tweeters in each single-candidate camp (followers of Clinton only, Trump only, etc.) plus a random sample of 10 users overall.

The good news is that all the accounts that were sampled in the top 10 and random 10 that look like normal user accounts (aka a person, not an organization or bot) seem to be reflective of a user's candidate loyalty.

Examples:

@TLW3 (Clinton follower, top 10 tweeter)
#News, #Politics, #Sarcasm, #Atheist, #Writer, #Liberal, #Irony, #Tech -

@DavidKGather (Clinton follower)
I write about climate change, Nikola Tesla, and Abraham Lincoln. Oh, and I'm an aquarium nut. Go figure.

@GKMTNtwits (Clinton follower)
Media Critic/Journalist/Hillary 4 Am. Vol. Leader ~Prolific!!! Bad Manners Blocked! Blogs: https://t.co/KraI7CJNlD & https://t.co/WRcDx0bP63 #PDMFNB #UniteBlue

@EileenDiana (Trump follower)
Trump supporter, free-lance writer, loves life, liberty & the pursuit of happiness

@leatherneck111 (Trump follower)
I am a former Marine (E-5 Sergeant.)I am a proponent of eradicating a welfare state. I am appalled by the degradation of society by the liberal establishment

@macrossfan (Trump follower, top 10 tweeter)
Geeky whitecollar southpaw who's family is priority. Anime Fan, foodie, arm-chair historian, and tech lover. I RT what I find intriguiging & most #news.

@poohsbrana (Sanders follower, top 10 tweeter)
#FeeltheBern #WeStandTogether

@farm_mom (Cruz follower, top 10 tweeter)
Conservative, Christian, Farm wife to @1861_again, Stay at home mom of 3 boys, Gardner, Altrium health suppliment dealer, Strict Constitutionalist, Former Nurse

@Jordan_Arras (Cruz follower, top 10 tweeter)
Proud citizen of Texas. Constitutional Conservative. Future Congressman. I believe in freedom, peace, and prosperity. Do you ?

Caveat: non-person accounts
There's quite a lot of political news accounts in the top-volume tweeters, which makes sense-- if you are sharing story links frequently, then it would be appropriate behavior for a newsfeed account.

These accounts do still seem to show party/candidate loyalty, but they just aren't individuals.

For example, we have:

@LettersFromUs (Clinton follower, top 10 tweeter)
Love a little political humor? Then Letters From Us is the site for you. Check us out for new daily blogs and comics to feed your left-brain.

@TheRReport (Sanders follower, top 10 tweeter)
Cutting a path through politics, with news updates, opinion and errant thoughts. Unapologetic progressive. Global warming. Water. Income inequality.

@LiberalNews2012 (Sanders follower, top 10 tweeter)
Democrats | Democratic Party | Barack Obama | Obama 2012 | Dems | TopProg | p2 | Election 2012 | Politics | Change | #TeamFollowBack | Decision 2012

One funny standout in the top 10 tweeters by volume who are in the Clinton camp is **@WSJopinion.** Apparently they only follow Clinton.

We also have businesses and organizations, again these do seem telling of party/candidate loyalty.

For example,

@DareSomething (Cruz follower, top 10 tweeter)
Info Central For Political Activism, Conservative Issues, Republican, Libertarian, jschulmansr, ResistNet.Com, Tea Parties, my blogs, sites and much more...

@Xtrememusicmix (Trump follower)
Northwest Iowa's Only Alternative Station!

---

2. Connecting Data from Part 1 of my thesis with Part 2 of my Thesis

From the first part of my thesis, I performed two studies. In both of them, I surveyed people about their political affiliations and whether or not they found a news article to be trustworthy. The first study include a larger amount of stories (117) with 3-way annotations; the second study looked specifically at effects of political affiliation and publication (I pretended the same story was from 3 different outlets plus unlabeled to measure effects) and thus only contained 8 stories.

No stories from study 2 were found in our tweets dataset.
From study 1, 52/117 (44.4%) stories were found to be tweeted about in our dataset, with 1,559 tweets total about them. (Recall we have 137,986 tweets total.)

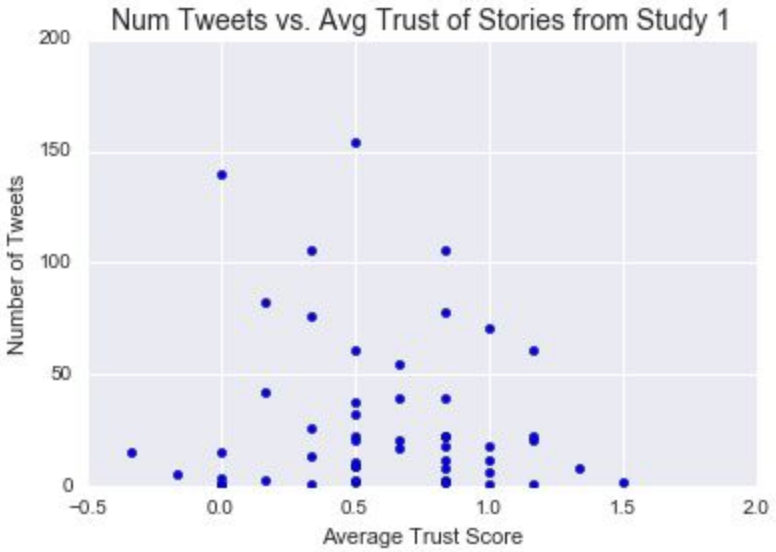
Only looking at stories with 10+ tweets about them, 33 stories from the first part of my thesis are in our dataset, with 1,488 tweets about them.

Do we see any kind of pattern for trust & virality? (tweets, mean trust score)

Story	# Tweets	Mean Trust
Jeb Bush on Donald Trump: 'The guy needs therapy'	154	0.5
Why Cruz is in trouble in Texas	140	0
Sanders a prolific Dem fundraiser	106	0.83
Ted Cruz's conventional campaign to win Iowa	106	0.33
Who's funding this pro-Ted Cruz super PAC?	82	0.17
CNN/WMUR Poll: Cruz climbs in NH	78	0.67
Sanders surpasses 2015 fundraising goal with sizable haul	76	0.33
Cruz rules out 'deportation force' to boot undocumented immigrants	71	1
Ted Cruz worry: Big South Carolina loss equals bad Super Tuesday	61	0.5
Hillary Clinton hopes for strong Latino turnout in Nevada	61	1.17
Evangelical leader sub-tweets Donald Trump	55	0.67
Political fallout for Clinton after North Korea's nuke test	42	0.17
Ted Cruz tries to seize Rand Paul's libertarian mantle	39	0.83
Takeaways from Ted Cruz's Iowa road trip	39	0.67
Obama hopes to pave way for Clinton with SOTU	38	0.5
Cruz to tea party supporters: Don't get 'burned' by fake conservatives	32	0.5
Clinton: 'I know what it is like to come from behind and win in New Hampshire'	26	0.33
Clinton to Planned Parenthood: 'I will always have your back'	22	1.17
Graham: Cruz worse than Obama	22	0.83

Average: 0.60
Average trust across entire study : 0.55

Correlation between Trust and # Tweets?
Using Pearson correlation, no significant correlation found.



However, since we also have rich reader (who they're voting for, political party, demographics) information for these 33 stories, we could still use this dataset to back up some of our findings with survey data.

---

3. Stance & Virality

The hypotheses that emerged from our last meeting that was most interesting to me was the idea that negative content (hate-linking etc.) was more likely to become viral (or at least have more volume) than positive content. Or if we found the reverse to be true, that's also interesting. (Wharton School study found that on the whole, *positive* content went *more* viral than negative content; however within the categories of positive and negative, those emotions that caused more arousal/activation were more likely to spread.) If we find that *negative* content on the whole is more likely to become viral if it's political, that's a cool finding.

Some comments on crowdsourcing the annotations:

A) Does it make sense to annotate all the stories for positivity/negativity/stance in the narrowed dataset?That's about 2.6K stories, on par with the Wharton School

study (subset of 2,566 (out of 6,956) manually coded). We would want at least 2 (best 3) annotators per story.

B) But will this work on our timeline? For instance, my first study took about 24 hours to complete, and that was ~120 stories of 3 annotations each. It is a pretty computationally intensive task to ask someone to read a full news story, so if we scale directly that's something like 21 days to complete. However I found there to be significant die-off in task completion over time-- that's because when you first put the job out there you can expect a pool of type A workers to respond ASAP. But once you go through that pool, recruiting starts to trickle. Not sure if that works in our timeline.
C) It is not feasible to code all stories in the Electome (22.9K).

As a workaround, I coded the top 10 most/least tweeted stories for each camp using the LIWC word dictionaries on two dimensions (positivity, emotionality*) from Wharton school study (see below). From this test dataset, the stories more likely to be tweeted about have significantly higher emotionality than those that don't.

Next steps:

I plan to automatically code all 2.6 K stories for those two dimensions over the weekend to see if there's a significant correlation b/t emotionality and tweet volume in the dataset.

If this is promising, then maybe we can use human annotators to code a subset.

***From Wharton School study:**

A computer program (LIWC) counted the number of positive and negative words in each article using a list of 7630 words classified as positive or negative by human readers (Pennebaker, Booth, and Francis 2007). We quantified positivity as the difference between the percentage of positive and negative words in an article. We quantified emotionality as the percentage of words that were classified as either positive or negative.