Is Social Media Really Helping People Making Informed Decision in the 2016 Election?

Ang Li, Keyang Zheng, Sara Daraei, Yang Sun
School of Information Science, University of Pittsburgh, Pittsburgh, PA

Motivation and Problem

Previous years, social media has played an important role for presidential candidates in elections. **However, how social media help people to make a better informed decision still remains a challenge.** Previous research has showed the existence of "filter bubble" effect in social media which means that content suggested to users are customized only based on users' taste. As a result, in the context of election, users start to receive information from their favorite side and will be isolated from the other side, which makes the user's decisions biased.

In this project, we aim to investigate **how much filter bubble in Twitter affects users decision during the election** and what are the strategy that social media can improve users' informed decision.
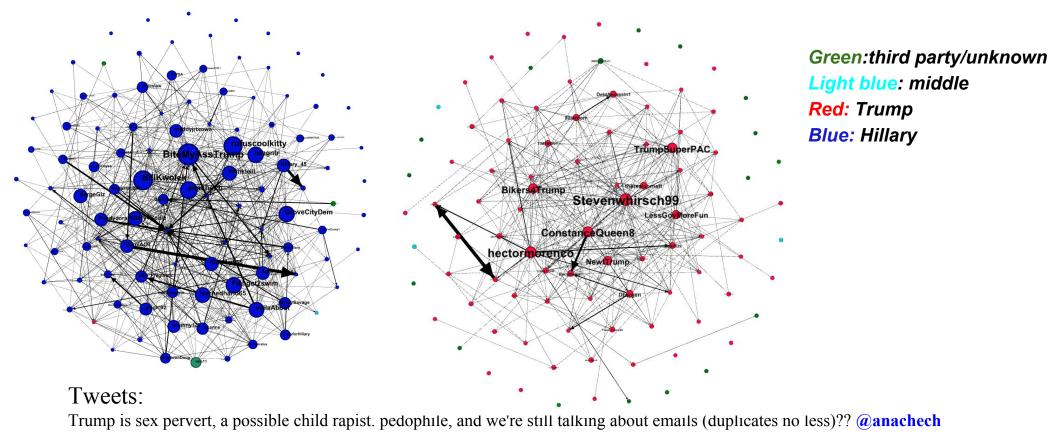
Method and Data Description

- Method
  - **Create Information networks**
    Nodes: Started with one seed user
      Randomly selected 200 users from the 2 levels friends(bidirectional)
      Nodes need to have tweets longer than 30 days
    Links: Retweet, reply, and mention
    Weight: Number of retweets, replies, and mentions
    Political side: based on user description and their past tweets
  - Identify whether there exist "bubble"
    Are friends all under same party? Are Tweets informed or biased?
  - How did information spread out (*network structure*)?
  - What can we do to reduce the bubble effect and spread out different perspective? (*Simulation by adding "bridging roles"*)
  - **Create Large Social Network:** to find middle ground or broker's role by looking at the whole network (**Link** as bidirectional follow)
- Data
  - Hillary's Side Information Network: 109 nodes and 622 links
  - Trump's Side Information Network: 97 nodes and 331 links
  - Large social network Data: 24 seed users and their friends

Result - Tweets Analysis

Finding Highlights
- Majority of the tweets are biased
- Re-tweets are generally more biased than original tweet
- Hillary's side: **10.56%** of tweets mentioned Trump, **30.47%** mentioned Trump
- Trump's side: **9.54%** of tweets mentioned Hillary, **23.703%** mentioned Trump

Network Represent by Out-degree Centrality

Green:third party/unknown
Light blue: middle
Red: Trump
Blue: Hillary

Tweets:
Trump is sex pervert, a possible child rapist. pedophile, and we're still talking about emails (duplicates no less)?? **@anachech**
@spooney35 @FBI @HillaryClinton @cspanwj Hope we beat Trump like a hebrew slave. **@Neo11**
#RememberWhenHillary created the Clinton Foundation to sell influence, intel, favors, weapons, and people to anyone willing to pay. **@hectormorenco**
RT @newtgingrich: So the core of the election is Trump "drain the swamp" versus Clinton "build a bigger swamp with foreign bribes"-clear ch???&?? **@NewTrump**

Result - Tweets Analysis

Network Represent by In-degree Centrality

Tweets
Trump is only +1 in GA. Good day to follow GA senior advisor @lewtml and GA political director @MrPeachState - crushing it every single day! **@AdamParkhomenko**

Let's be clear: purpose of private server was to hide Clinton Foundation activities-pay-to-play @CNN @MSNBC #MAGA??%?? https://t.co/kwYzuHZjqv **@Timetravels007**

Network represent by Betweenness Centrality

Tweets
Unacceptable. Young boy with cerebral palsy has wheelchair kicked at Trump rally. #ImWithher #StrongerTogether https://t.co/1FM88MvKys **@maramcewin**

RT @MaryHukill: Secret Service Detail: Hillary Clinton Was Drunk Morning of Campaign Rally https://t.co/vsTaUwSfgh **@Biker4Trump**

Hashtag Word Cloud

Trump Supporters Mentioned about Trump
"MAGA", "DrainTheSwamp"

Trump Supporters Mentioned about Hillary
"CrookedHillary", "NeverHillary"

Hillary Supporters Mentioned about Trump
"Never Trump", "Not My President"

Hillary Supporters Mentioned about Hillary
"ImWithHer", "StrongTogether"

Results - Information Network

Two Networks have no connection
- **Network Measurements**

	Hillary's Side	Trump's Side
Density	5.2%	3.6%
Diameter	34	88
Average Path Length	3.12	2.52
Average Reachability	40%	14%
Global Transitivity	0.19	0.24

- Hillary's net showed to have **better interaction** and **higher reachability** to other nodes
- Both networks have **low transitivity**, indicating a very low level of clusters or groups connected as triad
- Both have **short average path length**
- All indicate that both networks are "**Hierarchical Structured**"

Results - Information Network

Krackhardt's graph theoretical dimensions of hierarchy

	Hillary's Side	Trump's Side
Hierarchy	0.76	0.94
Connectedness	0.86	0.58
Efficiency	0.95	0.95
Lubness	0.35	0.19

Tree structured network:
- Information flow are **not reciprocated** (on direction)
- Some actors can **reach** most of the users (Hillary's side higher than Trump's side).
- Users usually have **single previous level** information sources.
- Those information sources are **different** though they all very biased and radical.

Centrality Correlations:
Tree Structure Information Flow
* Not significant if backcolor is blank

Hillary's Side — Trump's Side

Both sides:
Outdegree, closeness and *betweenness*:
Users that "talk" to the most others are the likeliest to serve as bridges between the groups, and in the very center of the group.

Indegree and outdegree:
"Talk" with others is NOT reciprocated

Hillary's side:
In-, out- degree and eigenvector:
Users that talk the most to others (or, relatedly, are talked to the most by others) are also the ones that are connected to other highly connected users

Information Diffusion in SIR Model

Original information diffusion: Non-recovery Vs. Recovery

SIR (susceptible-infected-recovered) Model:
Infection rate: 5
Recovery rate: 2
Simulate: 1500 times with every time randomly started with one node.

Information diffusion of adding a node to connect random 3 nodes from two parties

Original network information flow:
Only reach half of the nodes

Information diffusion of adding a node to connect 3 influential nodes from two parties

By adding a "Bot Node" to connect two sides we see the information spread wider and faster.

Summary

- **Bubble exists in both information networks:**
  - There are no connection between two parties.
  - Majority users inside the network are under the same party.
  - Tweets are very biased and radical.
- **Information Diffusion:**
  - Both networks showed that information spread in a hierarchical/tree structure
  - Hillary's side showed more interaction than Trump's side
  - Add "bridging" role to connect the high influential nodes will reach wider users with faster speed.

Results - Full Friends Network

Localized Social Network in Pittsburgh

Democrats	Republicans
@dmstoicovy	@beat_train
@IUPcollegeDems	@vonnyD412
@BenfordMatt	@RBrawdy
@szhamilton	@connienovak1972
@jenashburn	@Quitesmartin
@jbaarp1	@ben_karpinski
@AylaSchermer	@jpgendell
@CorreyLyn	@CJCoch
@CookieSheila	@carmanavenue
@calitopgh	@cblakes
@ashlynbrierre	@AllyMazzoni
@soph_englesberg	@ewc

Localized Social Network (two-way connection)

- Based on 24 local pittsburghers' Twitter network, two-way connection should be in both the following and follower group. Two-way connection ensures the user does not only receive the information, there is also a better chance for interactions between users.
- Total of 5285 different connections, 3129 of those originated from Democratic group while the other 2156 from Republicans.
- 3 of the original 24 pittsburghers are isolated from the network, two of them are Republicans or their supporters. All of them do not have more than 13 connections.
- There are total 52 intersection points in the network, which could be seen as the bridger. However, a deeper look at these nodes tells another story.
- In those 52 intersection points, only 3 of them are more in line with normal users, rather than celebrate, marketing purpose or local information accounts.
- In those 3 (@pghseth, @calipanthergrl and @hittecece) users, all of them have showed some form of political affiliations. @pghseth and @calipanthergrl show a strong feeling for the Democrats, while @hittecece on the other hand shows clear support for the Republicans.
- When removing all the other none-friend intersection nodes, it's more clear, two bubble is forming for each side.

Intersection Nodes

Modified Social Network (with only 3 brokers)

The connectivity of the network reduces greatly after removing other none-broker intersections.

There is a minor party supporter in the original 52 intersection points. However, his celebrate quality in social media makes him a very unlikely candidate for being the actual broker in the system.

- Also using just following (as Twitter puts it friends) information, we also built a directional social network.
- It's clear in both network, the Democrats are more active on social media, or more outgoing to express their thoughts.
- Aside from marketing and local information, the most common connection between 2 sides seems to be Sports and Music, which consist most of the intersection points in following social network. And it works on the original social network, too.

Social Network (using following)

Limitation and Future Work

Limitation:
- Relatively small sample size does not reflect the filter bubbling of information in a larger scale.
- Using Twitter as the data source also has its limitation, as generally people are not comfortable with sharing their political opinions with someone they don't know.
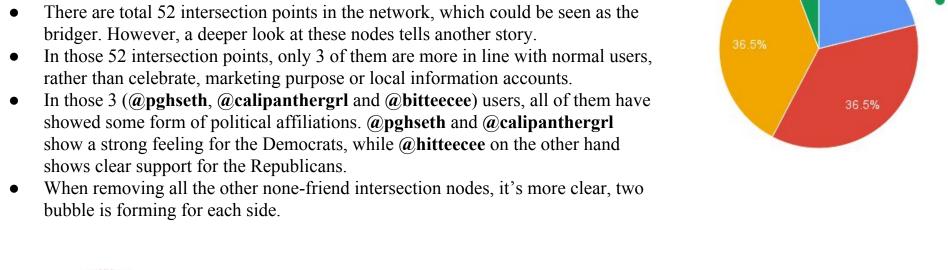
Future Work:
- Larger sample size with multiple data sources.
- Using more users as the ground point for building localized social network.
- Building information networks for localized social network (links defined as retweet, reply and mention).
- Research more close-relationship social media like Facebook, to see the difference between these two.
- While doing the analysis, we suggests also doing a simple test subjects study, to get a better sense of his or her political strand.