



# The Polarization of Information on the Web

Charles Dickens, Narayana Prasad Santhanam  
Dept. of Electrical Engineering, University of Hawaii at Manoa

## Introduction & Motivation

Many sources claim that the political climate is becoming increasingly polarized, but few attempt to quantify this feeling. Furthermore, the strict dichotomy of left vs. right is a simple and typically accepted narrative, but perhaps a more fine grained model which considers smaller schools of thought on a topic to topic basis may better capture what is truly happening on the web. With an objective understanding of the polarization of information on the web, researchers could pinpoint causes, and potentially come up with solutions to reverse this concerning trend and foster more fruitful conversation.

## Project Description

Popular microblogs such as Twitter have become primary sources of information for many citizens. Twitter encourages users to share their opinions and information about current issues and events, no matter how contrarian, and access to such a variety of viewpoints has the potential to foster awareness. However, it seems groups of society are forming increasingly polarized opinions on topics, leading to disagreements over even factual details. This concerning observation has been a common topic of conversation as of late, but an accepted method for quantifying the polarization between camps on a topic to topic basis has yet to be developed, leaving the dialogue and, as a result, the proposed solutions subjective and difficult to act upon.

The short term objectives of this project are:

1. Develop a method for comparing the polarization of existing communities of thought at the topic specific level
2. Improve methods for modeling social networks

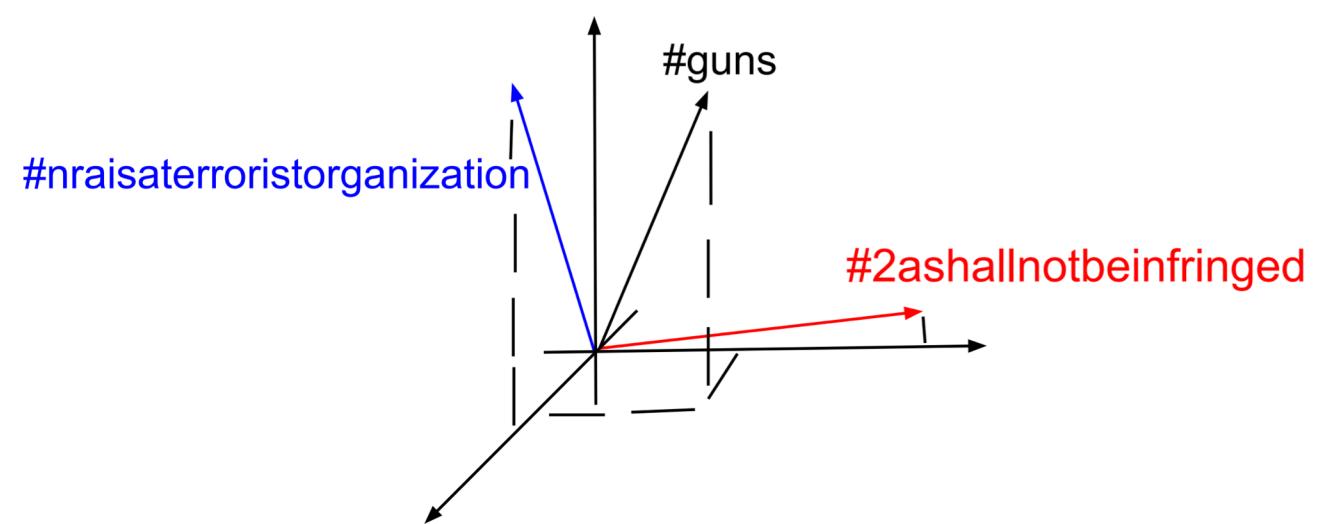
The contributions of this report is the framework for a workflow which

1. Models a network of tweets related to a common topic using a semi-supervised approach
2. Finds community structure in the model
3. Measures the polarity using graph conductance.

## Materials & Methods

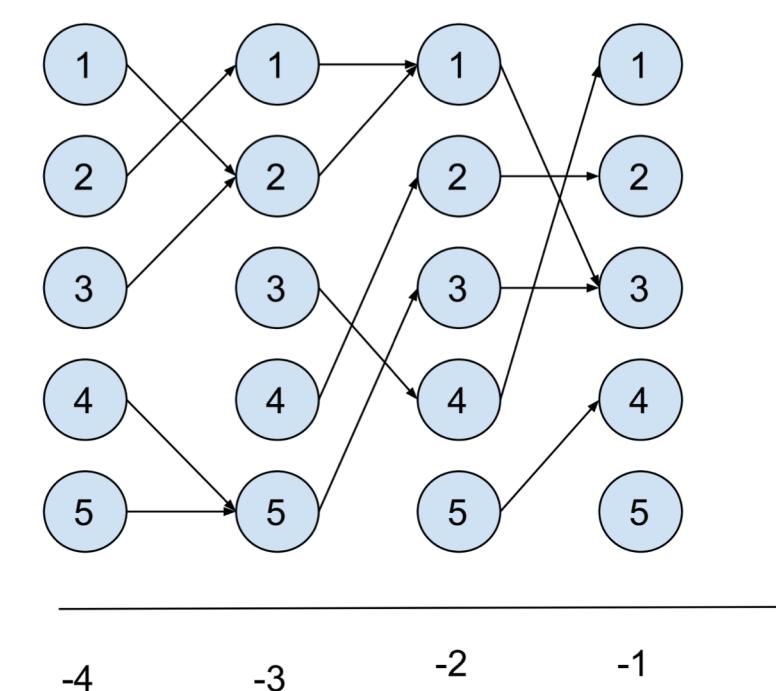
### Retrieving Relevant Information and Opinions:

Twitter granted this project access the premium search API sandbox, which allows for 25000 tweets to be collected each month.



### Community Detection:

- The community detection algorithm used in the system, backward path community detection, was developed at the University of Hawai'i Big Data Lab (Paravi and Santhanam, 2015)



Polarity Calculations: The polarity is measured at an individual community and network wide level for each topic using an estimate of the graph conductance metric. The conductance of a cut  $(S, \bar{S})$  of a graph  $G = (V, E)$  is defined as:

$$\varphi(S) = \frac{\sum_{i \in S, j \in \bar{S}} w_{ij}}{\min(w(S), w(\bar{S}))} \quad \text{where} \quad w(S) = \sum_{i \in S} \sum_{j \in V} w_{ij}$$

Then the graph conductance is defined as  $\phi(G) = \min_{S \subseteq V} \varphi(S)$ .

Table 1: Twitter API Search Queries by Topic

Topic	Queries
Gun Control	'guns'
Voting	'voting rights', 'voter turnout', 'voter fraud', 'voter suppression', 'vote'
Immigration	'immigration', 'illegal immigrants', 'trump wall', 'border security', 'asylum seekers'

### Modeling the Network:

- Tweets,  $t_i \in T$ , are modeled as nodes in the network and an edge between two nodes is weighted to reflect how similar the two tweets are.
- The set of all unique hashtags,  $H$ , is aggregated into a file and is manually labeled with a sentiment value existing in the set  $\{-1, -1/2, 0, 1/2, 1\}$

```
Algorithm 1 Update Network Edge Weights
for all  $t_i$  in  $T$  do
     $v_i \leftarrow \frac{h_{i,1}n_1(p_{1,1}, \dots, p_{1,d}) + \dots + h_{i,m}n_m(p_{m,1}, \dots, p_{m,d})}{h_{i,1}n_1 + \dots + h_{i,m}n_m}$ 
end for
for all  $t_i$  in  $T$  do
    for all  $t_j$  in  $T$  do
        if  $t_i = t_j$  then
             $w_{i,j} = 0$ 
        else
             $w_{i,j} = \phi(\frac{<t_i, t_j>}{||t_i|| ||t_j||})$ 
        end if
    end for
end for
return  $W = (w_{i,j})$ 
```

```
Algorithm 2 Build Network Model
for all  $t_i$  in  $T$  do
     $c_i = [t_i]$ 
end for
repeat
     $c'_k \leftarrow c_k$  for all  $k$ 
     $W = (w_{i,j}) \leftarrow$  Update Network Edge Weights
     $C = (c_k) \leftarrow$  Backward Path Community Detection
until  $c'_k = c_k$  for all  $k$ 
```

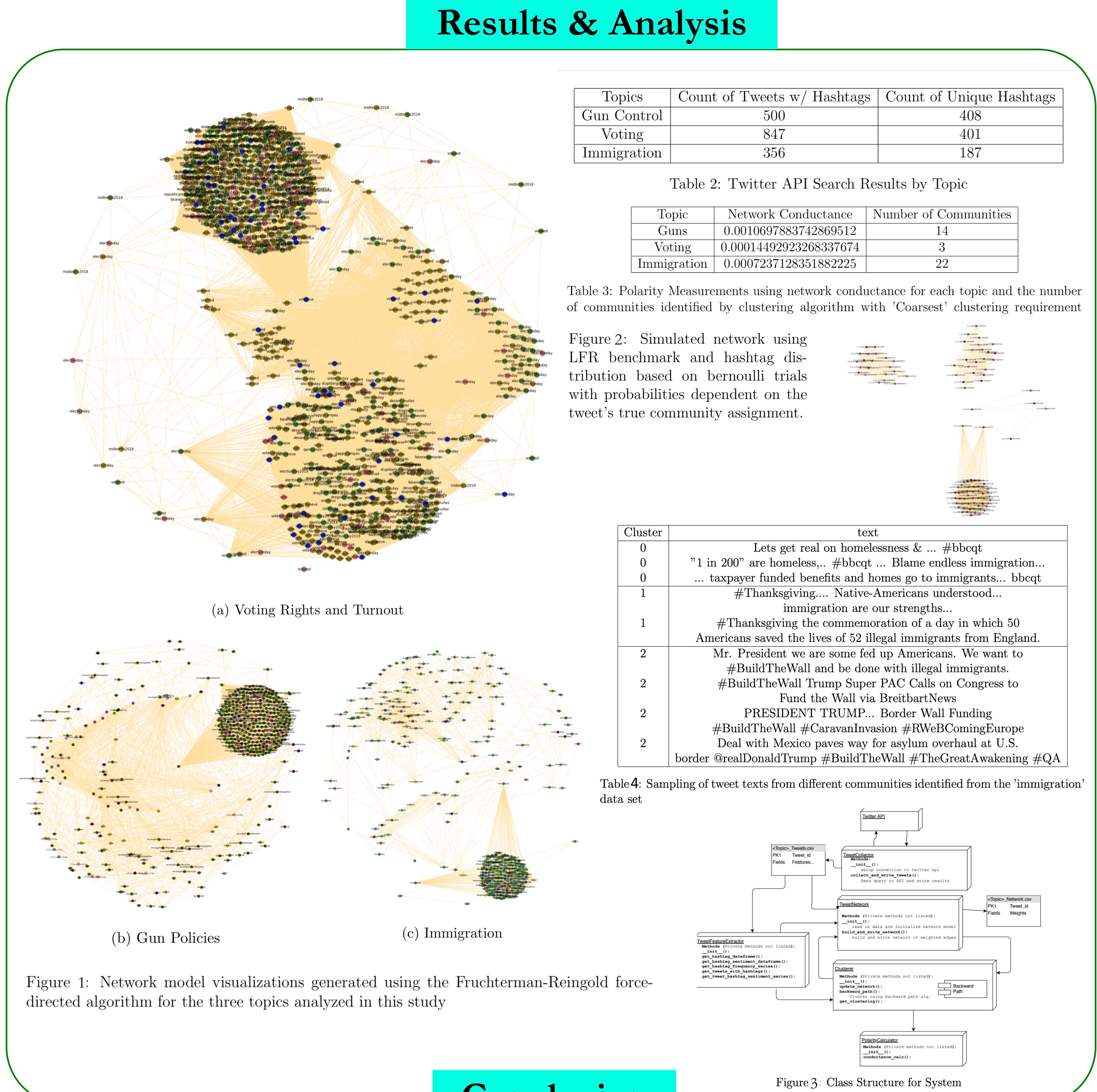


Figure 1: Network model visualizations generated using the Fruchterman-Reingold force-directed algorithm for the three topics analyzed in this study

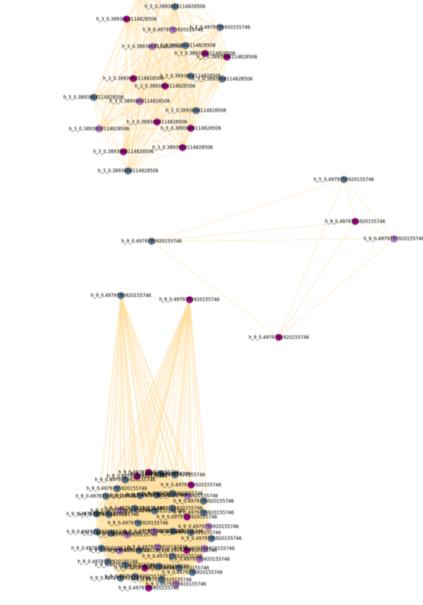
Topics	Count of Tweets w/ Hashtags	Count of Unique Hashtags
Gun Control	500	408
Voting	847	401
Immigration	356	187

Table 2: Twitter API Search Results by Topic

Topic	Network Conductance	Number of Communities
Guns	0.0010697883742869512	14
Voting	0.00014492923268337674	3
Immigration	0.0007237128351882225	22

Table 3: Polarity Measurements using network conductance for each topic and the number of communities identified by clustering algorithm with 'Coarsest' clustering requirement

Figure 2: Simulated network using LFR benchmark and hashtag distribution based on bernoulli trials with probabilities dependent on the tweet's true community assignment.



Cluster	text
0	Lets get real on homelessness & ... #bbcqt
0	"1 in 200" are homeless... #bbcqt ... Blame endless immigration... ... taxpayer funded benefits and homes go to immigrants... bbcqt
0	#Thanksgiving... Native-Americans understood...
1	immigration are our strengths...
1	#Thanksgiving the commemoration of a day in which 50 Americans saved the lives of 52 illegal immigrants from England.
2	Mr. President we are some fed up Americans. We want to #BuildTheWall and be done with illegal immigrants.
2	#BuildTheWall Trump Super PAC Calls on Congress to Fund the Wall via BreitbartNews
2	PRESIDENT TRUMP... Border Wall Funding #BuildTheWall #CaravanInvasion #RWBComingEurope Deal with Mexico paves way for asylum overhaul at U.S. border @realDonaldTrump #BuildTheWall #TheGreatAwakening #QA
2	

Table 4: Sampling of tweet texts from different communities identified from the 'immigration' data set

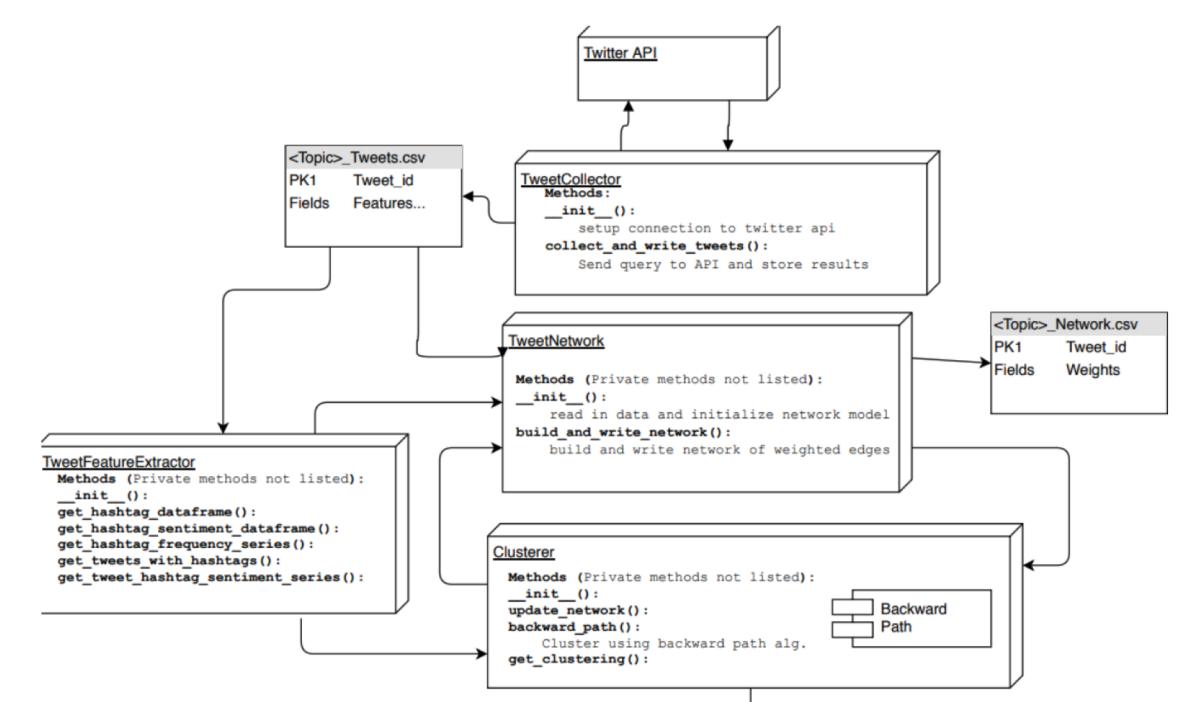


Figure 3: Class Structure for System

## Conclusion

Our system has analyzed multiple topics of discussion on Twitter and has identified groups that align with our intuitions. The network structure including polarity calculations can be successfully measured and compared across topics.

This implementation has been a proof of concept for the techniques used to build the network model, identify communities, and measure polarity. The system will be developed in future work to consider additional features such as tweet popularity, tweet text sentiment, and more. Moreover, the workflow for collecting the tweets can be modified to minimize bias introduced by using human intuition to create search queries.

## Acknowledgements

### Special thanks to

CSOI: Center for Science of Information (<https://www.soihub.org>) is a National Science Foundation Science and Technology Center made possible under grant NSF CCF- 0939370

