Term Projects

- 2 Basic Requirements:
  - Your term project must deal with **social media data**, which could be text data, network data, or some other data type that is of interest to you (e.g. images/videos), or any combination of them.
  - Your term project must ‘**mine**’ or ‘**analyze**’ your social media data to find/discover interesting patterns/knowledge.
- Form a team of 3 or 4 people.

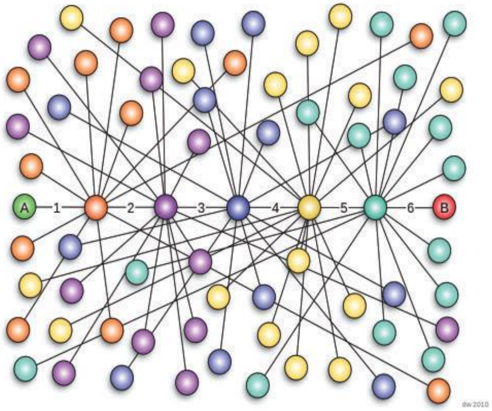
Term Project Proposal & Deliverables

- Term project proposals/description due **March 30.** At a minimum, 2 pages describing:
  - Your project idea and features
  - Significance of your idea (Why do you think that's a good idea?)
  - Work plan: technical tasks, their start/end datess, and who is responsible for which task
  - Will be graded mainly by 3 criteria: **creativity, completeness and clarity**

Term Project Proposal & Deliverables

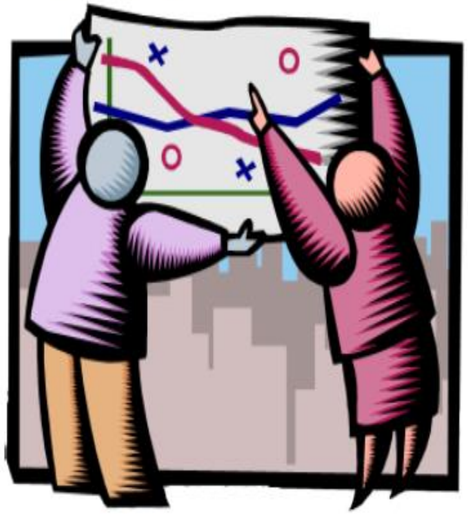
- Term Project Presentations: **April 24 & 26** (More details to come)
- Final project deliverables due: **May 6** (See the syllabus for the list of term project deliverables)
- Note: The due date (May 7) in the syllabus is wrong. It should be May 6.

Area 1: Network Modeling



- Large Networks demonstrate statistical patterns
  - Small-world effects (e.g., 6 degrees of separation)
  - Power-law distribution (a.k.a. scale-free distribution)

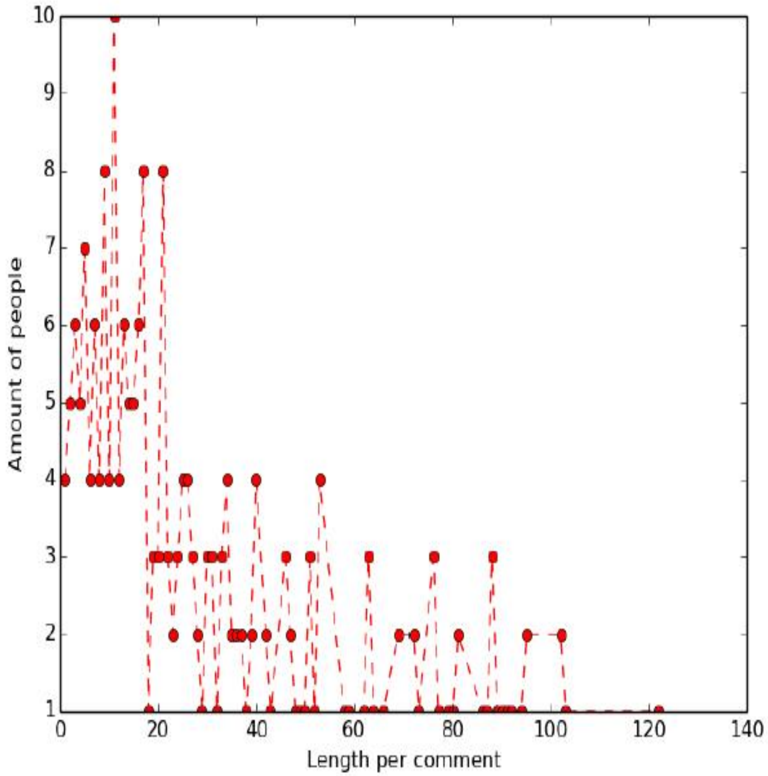
    $f(x) = ax^k$ $\qquad$ $f(cx) = a(cx)^k = c^k f(x) \propto f(x).$
  - Community structure (high clustering coefficient)
- Model the network dynamics
  - Find a mechanism such that the statistical patterns observed in large-scale networks can be reproduced.
  - Examples: random graph, preferential attachment process, Watts and Strogatz model
- Used for simulation to understand network properties
  - Thomas Shelling's famous simulation: What could cause the segregation
  - Network robustness under attack
  - Information diffusion within a given network structure



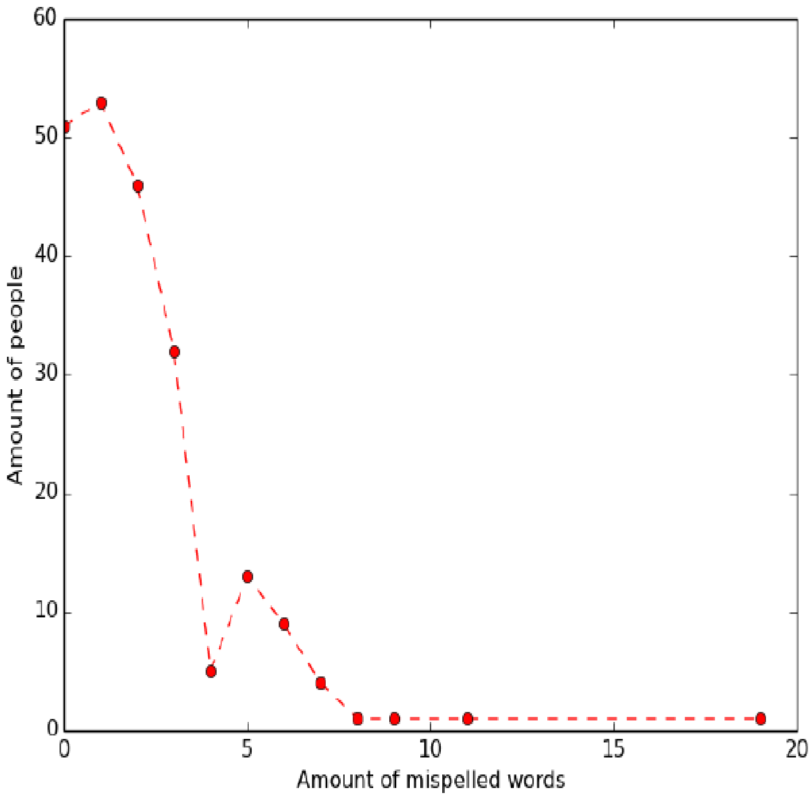
Comment Bank

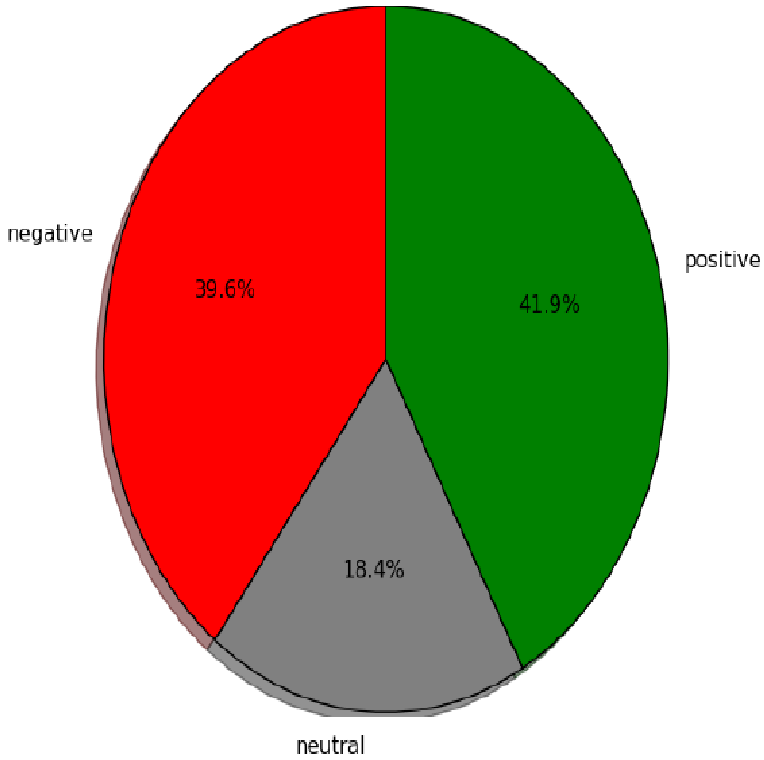
B. Hubbell && S. Shum

Sample - QFkgafvlRul



Sample - QFkgafvlRul





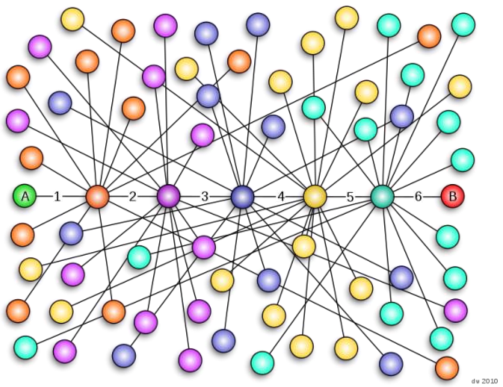
Analyze the Navigation Capability on Facebook Social Network

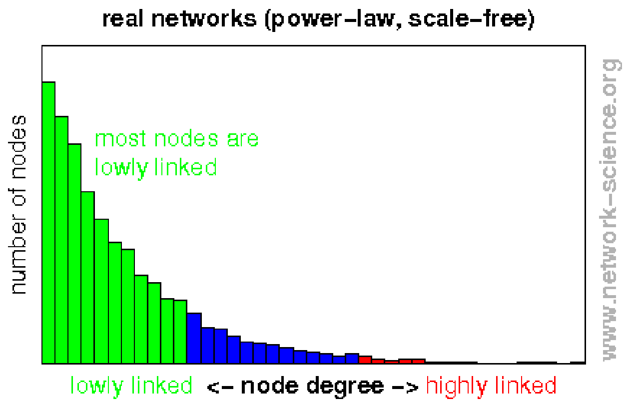
XIAO Zhang

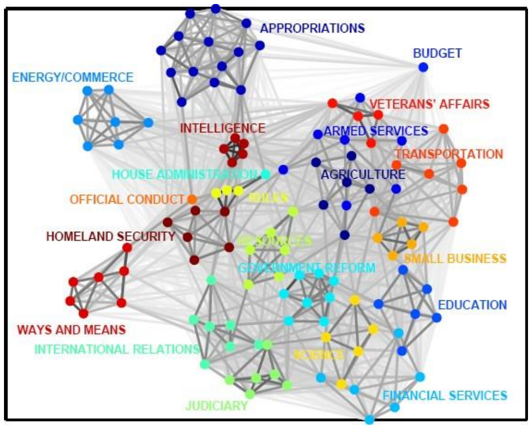
12/04/2013

Motivation

- Small-world Effects
- Power-law Distribution
- Community Structure

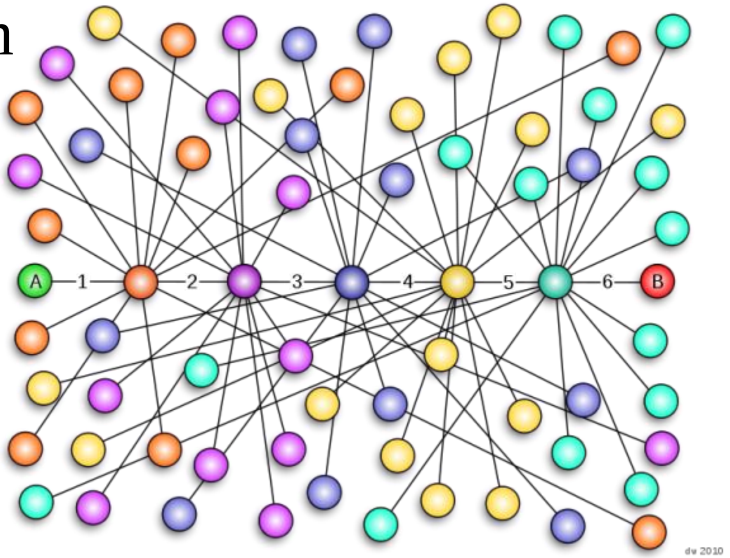






Motivation

- Small-world Effects
  - Six-degree of separation
  - Global structure information needed
  - How to navigate with local network information?
    - Degree-based Greedy Algorithm
    - MDS-based Greedy Algorithm



Area 2: Centrality Analysis

- Centrality Analysis
  - Identify the most important actors (leaders) or edges
    - The importance of a node is determined by the number of nodes adjacent to it
    - The larger the degree, the more import the node is
    - Only a small number of nodes have high degrees in many real-life networks
- Influence modeling:
  - Understand how people influence each other
  - Information diffusion
  - Viral marketing: word-of-mouth effect
  - Influence maximization: which k nodes, in a given network, should be activated (or targeted) first to maximize the number of active nodes at the end?

Quantifying Twitter Users' Sphere of Influence*

Aleksey Panasyuk
Dept. of Electrical Engineering &
Computer Science
Syracuse University, New York, USA
apanasyu@syr.edu

Edmund Szu-Li Yu
Dept. of Electrical Engineering &
Computer Science
Syracuse University, New York, USA
esyu@syr.edu

Kishan G. Mehrotra
Dept. of Electrical Engineering &
Computer Science
Syracuse University, New York, USA
mehrotra@syr.edu

ABSTRACT

In this paper, we attempt to quantify a user's sphere of influence on Twitter. We illustrate a national correlation measure that can identify whether user's influence is at local or national level. For users that are local we utilize user's followers' self-defined locations to pinpoint user's central location and user's radius. Radius is the average distance to the central location of user's followers' locations that have above average frequency. User's sphere of influence thus consists of (i) whether a user's influence is at local or national level and if the user has local influence (ii) the user's central location and radius. We propose novel ways for estimating location distribution, quantifying influence using location distribution, and visualizing this influence on a map. The approach is verified against known local and national accounts and accounts with a known central location.

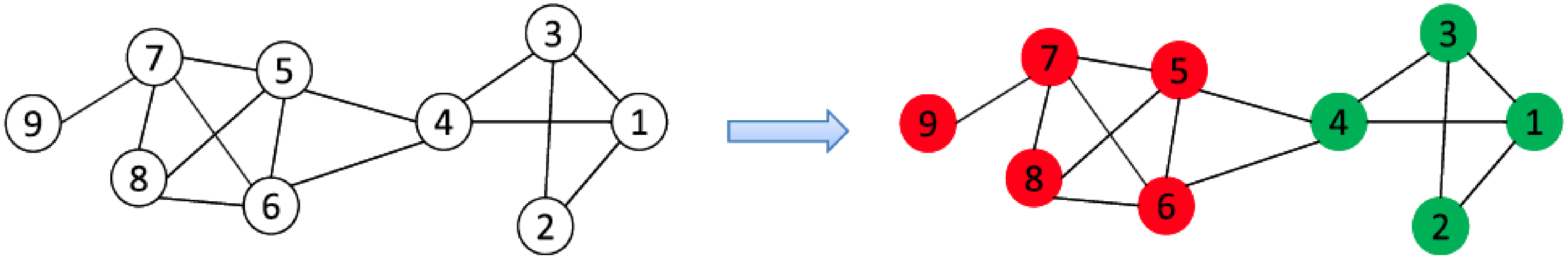
CCS CONCEPTS

There are many different flavors of what could be considered a national user, for example, the user might have a strong following along a particular coast, might have a presence across many states along political lines, and so on. A local user, on the other hand, has a specific pattern where there is an epicenter around which most of their followers are clustered. For users that are local we utilize user's followers' self-defined locations to pinpoint the user's central location and the radius size, where the radius is the average distance to the central location of user's followers' locations that have above average frequency. Hence, in our definition, user's `sphere of influence' can be national or local; for a local user it includes the user's central location and radius.

The rest of the paper is organized as follows: in section 2, we describe research related to quantifying the most influential users and how other researchers have utilized user-defined location. Section 3 describes the data used for location determination and

Area 3: Community Detection

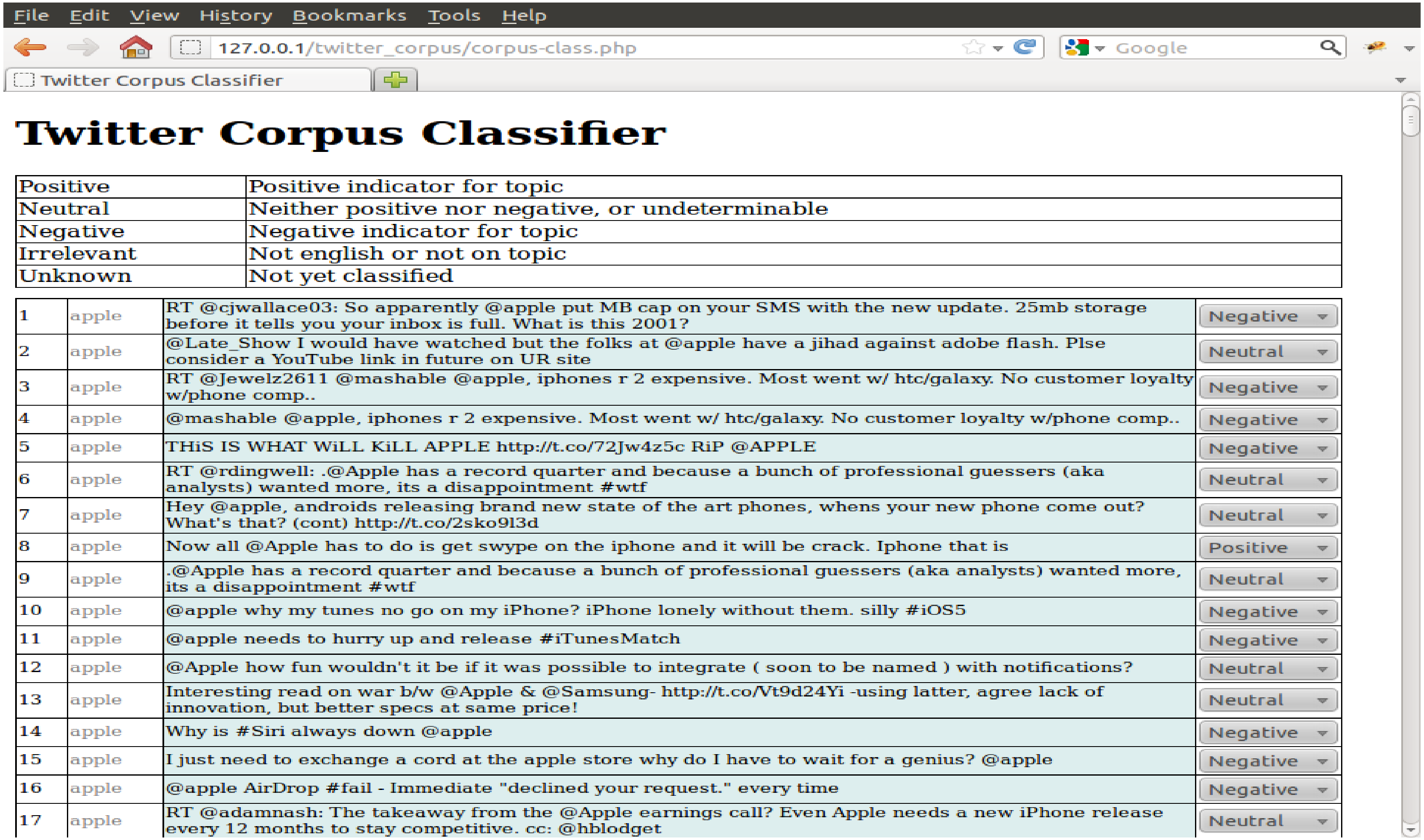
- A community is a set of nodes between which the interactions are (relatively) frequent
  - A.k.a., group, cluster, cohesive subgroup, module



- Applications: Recommendation based communities, Network Compression, Visualization of a huge network
- New lines of research in social media
  - Community Detection in Heterogeneous Networks
  - Community Evolution in Dynamic Networks
  - Scalable Community Detection in Large-Scale Networks

Area 4: Classification

❖ Content-based classification (Sentiment Analysis)



Twitter Corpus Classifier

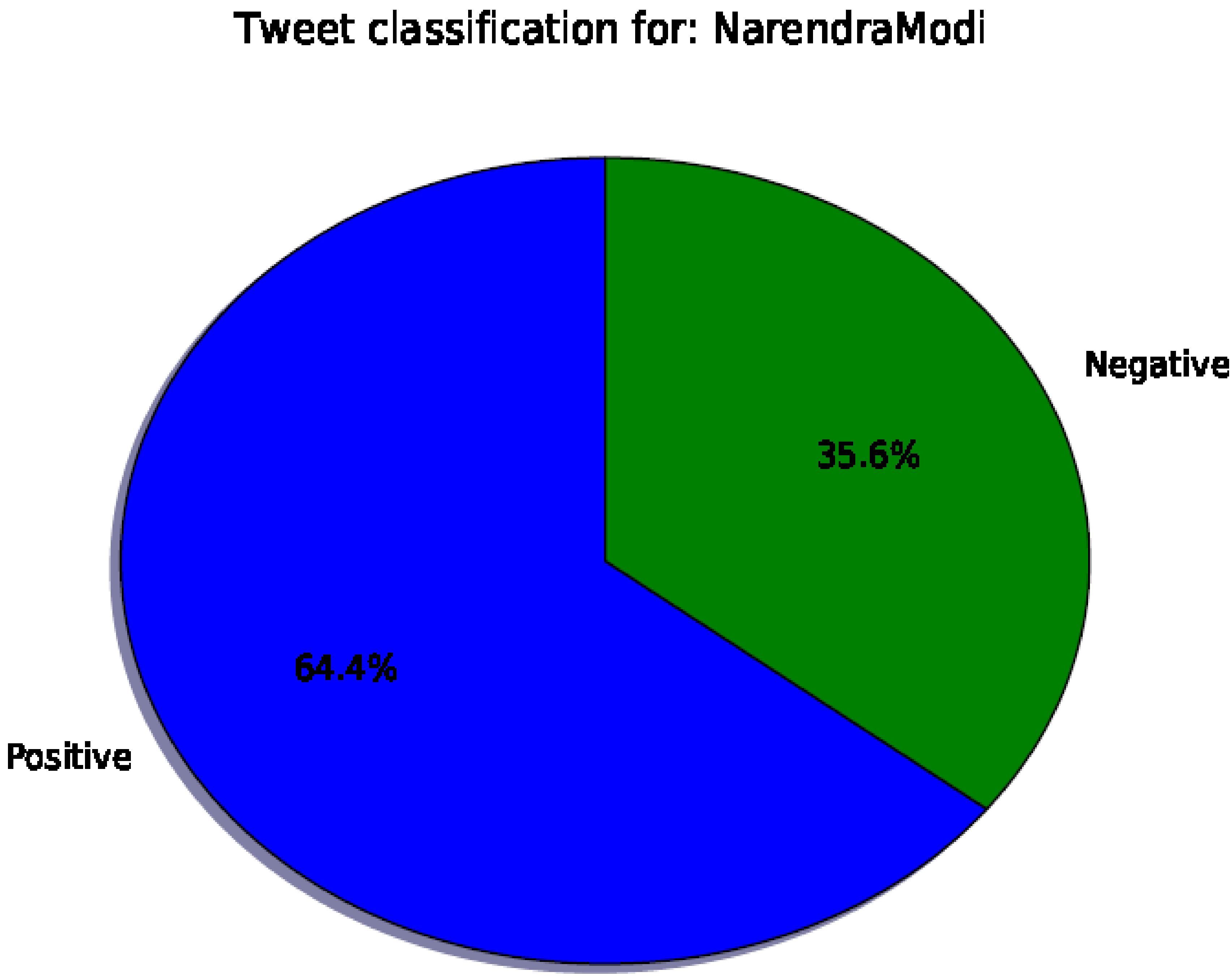
Positive	Positive indicator for topic
Neutral	Neither positive nor negative, or undeterminable
Negative	Negative indicator for topic
Irrelevant	Not english or not on topic
Unknown	Not yet classified

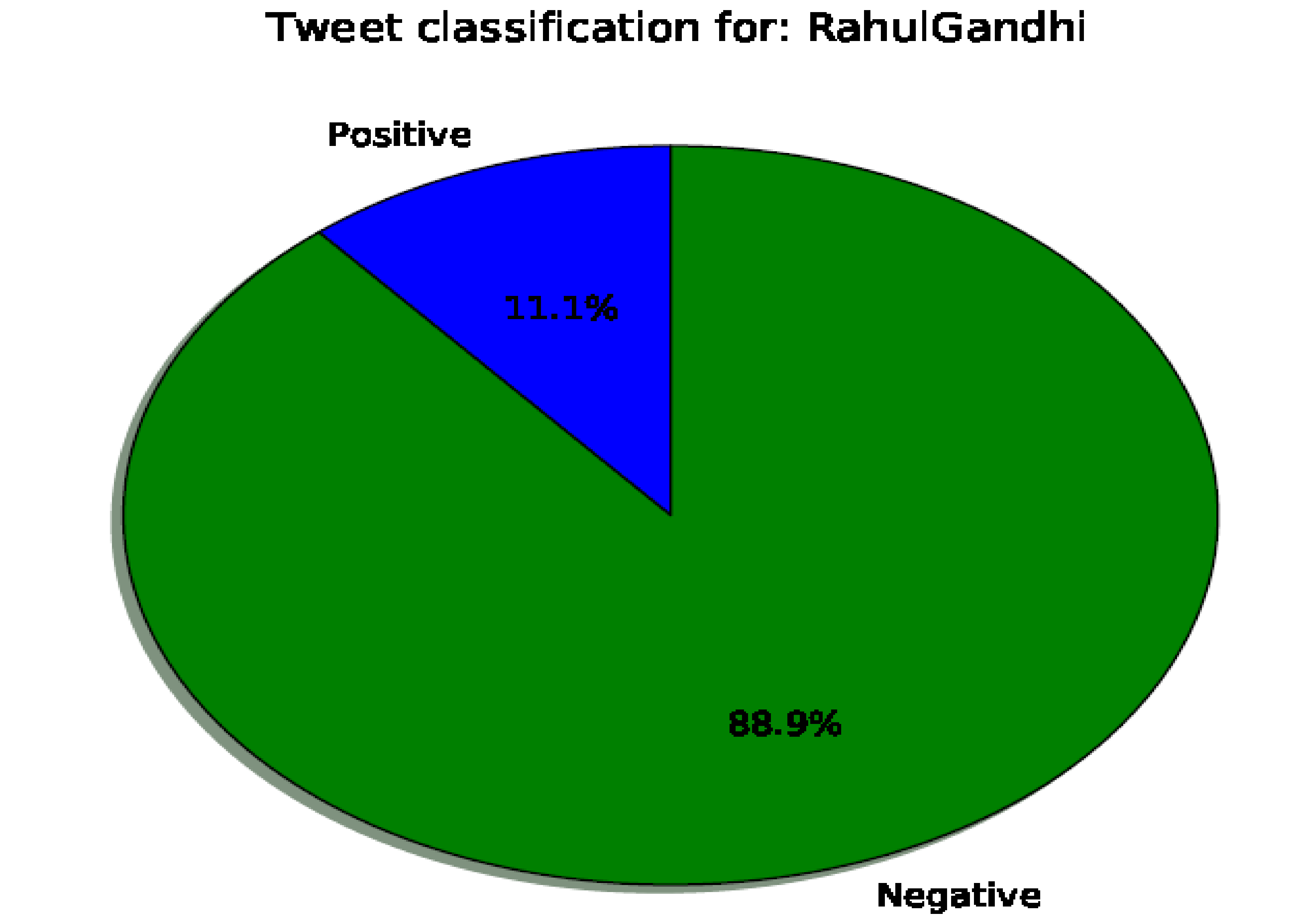
1	apple	RT @cjwallace03: So apparently @apple put MB cap on your SMS with the new update. 25mb storage before it tells you your inbox is full. What is this 2001?	Negative
2	apple	@Late_Show I would have watched but the folks at @apple have a jihad against adobe flash. Plse consider a YouTube link in future on UR site	Neutral
3	apple	RT @Jewelz2611 @mashable @apple, iphones r 2 expensive. Most went w/ htc/galaxy. No customer loyalty w/phone comp..	Negative
4	apple	@mashable @apple, iphones r 2 expensive. Most went w/ htc/galaxy. No customer loyalty w/phone comp..	Negative
5	apple	THiS IS WHAT WiLL KiLL APPLE http://t.co/72Jw4z5c RiP @APPLE	Negative
6	apple	RT @rdingwell: .@Apple has a record quarter and because a bunch of professional guessers (aka analysts) wanted more, its a disappointment #wtf	Neutral
7	apple	Hey @apple, androids releasing brand new state of the art phones, whens your new phone come out? What's that? (cont) http://t.co/2sko9l3d	Neutral
8	apple	Now all @Apple has to do is get swype on the iphone and it will be crack. Iphone that is	Positive
9	apple	.@Apple has a record quarter and because a bunch of professional guessers (aka analysts) wanted more, its a disappointment #wtf	Neutral
10	apple	@apple why my tunes no go on my iPhone? iPhone lonely without them. silly #iOS5	Negative
11	apple	@apple needs to hurry up and release #iTunesMatch	Negative
12	apple	@Apple how fun wouldn't it be if it was possible to integrate (soon to be named) with notifications?	Neutral
13	apple	Interesting read on war b/w @Apple & @Samsung- http://t.co/Vt9d24Yi -using latter, agree lack of innovation, but better specs at same price!	Neutral
14	apple	Why is #Siri always down @apple	Negative
15	apple	I just need to exchange a cord at the apple store why do I have to wait for a genius? @apple	Negative
16	apple	@apple AirDrop #fail - Immediate "declined your request." every time	Negative
17	apple	RT @adamnash: The takeaway from the @Apple earnings call? Even Apple needs a new iPhone release every 12 months to stay competitive. cc: @hblodget	Neutral

Who will be the next Prime Minister of India?

Use Sentiment Analysis techniques on twitter to gauge sentiments for next General Elections of India

Prasoon Dilip Pandya


Tweet classification for: NarendraModi
Negative
35.6%
64.4%
Positive


Tweet classification for: RahulGandhi
Positive
11.1%
88.9%
Negative

Complex Adaptive Systems, Publication 4
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2014- Philadelphia, PA

Controversial Topic Discovery on Members of Congress with Twitter

Aleksey Panasyuk, Edmund Szu-Li Yu, Kishan G. Mehrotra*

Dept. of Electrical Engineering & Computer Science
Syracuse University, New York, USA

Abstract

This paper addresses how Twitter can be used for identifying conflict between communities of users. We aggregate documents by topic and by community and perform sentiment analysis, which allows us to analyze the overall opinion of each community about each topic. We rank the topics with opposing views (negative for one community and positive for the other). For illustration of the proposed methodology we chose a problem whose results can be evaluated using traditional news articles. We look at tweets for republican and democrat congress members for the 112th House of Representatives from September to December 2013 and demonstrate that our approach is successful by comparing against traditional news media.

Keywords: Twitter; Latent Dirichlet Allocation; Topic Modeling; Polarizing Topics; Semantic Extraction; Social Media Mining

1. Introduction

Twitter has become an important social media site since its inception in 2006. It is a micro blogging service, which allows users to post messages up to 140 characters known as tweets. Twitter users are followed and are themselves following others, thus creating a social network. This social network can be used to identify

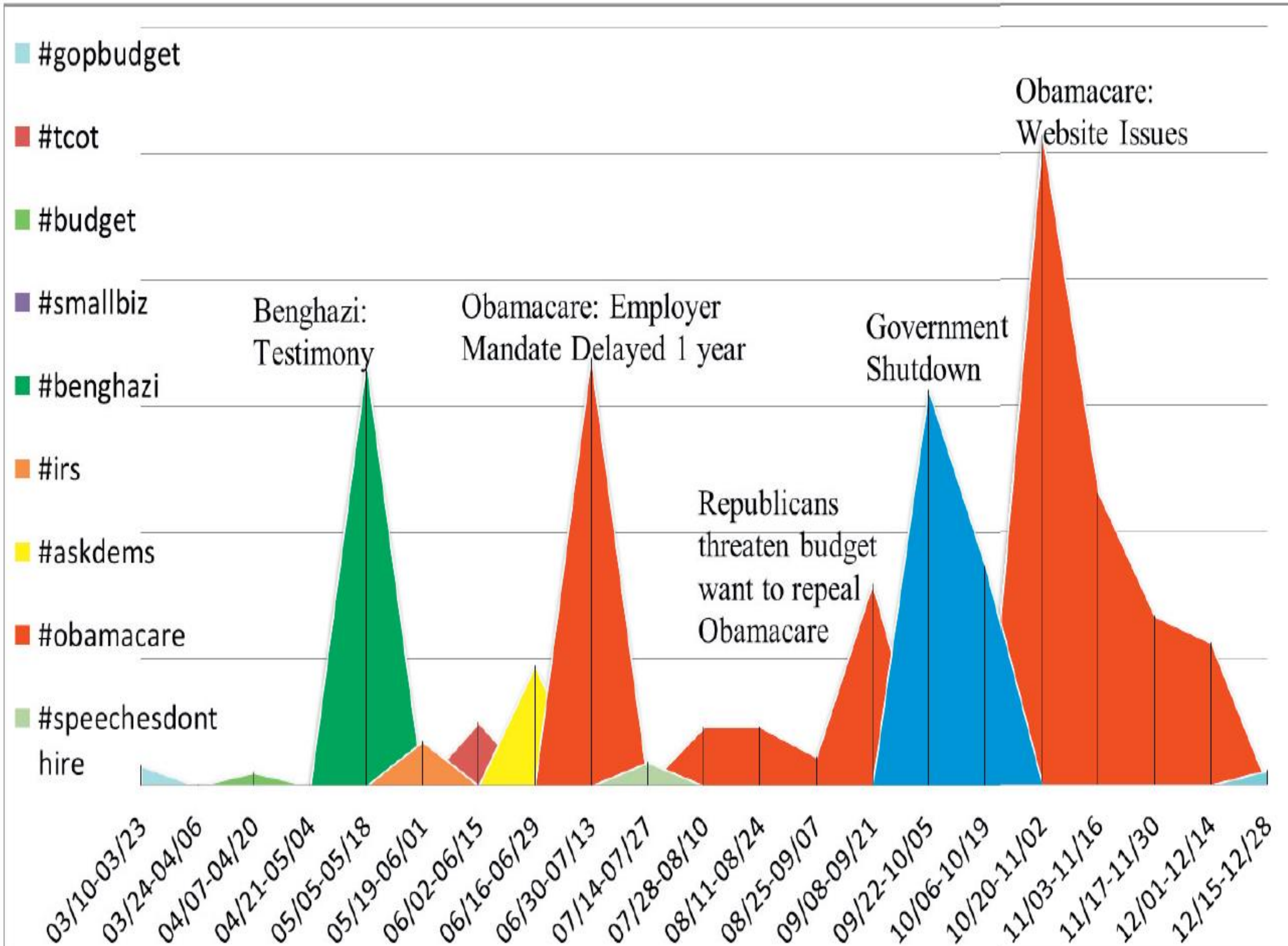


Figure 1. Time line of polarity scores, for top topics, over two week periods 2013-03-10 to 2013-12-28.

Sentiment Analysis in the tweets related to the impeachment of the president of Brazil

Heloisa Carbone

Introduction

- Dilma Rousseff, the president
- Corruption Scandals
- Impeachment crisis
- Lawmakers in the Brazilian parliament’s lower house voted to impeach Dilma Rousseff
- What happens next: The impeachment motion will next go to the country’s Senate

Project Overview

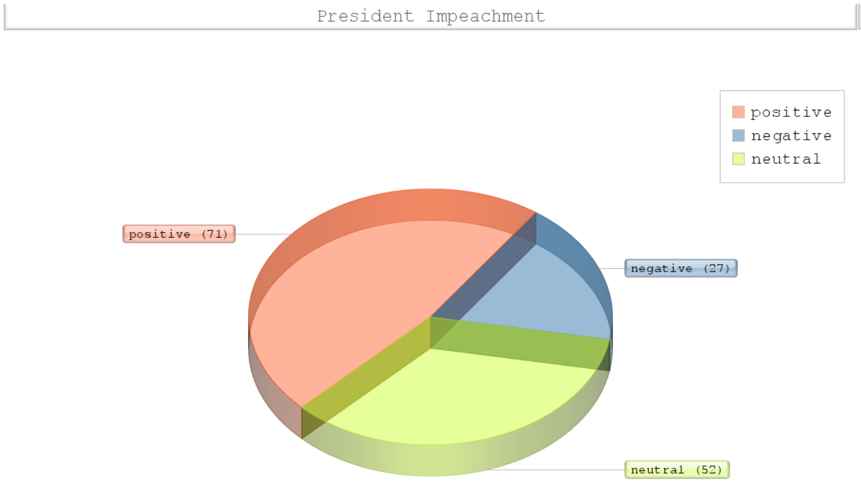
- Tweet analyze to see the proportion of people that are in favor or against the impeachment.
- Compare the proportions before the first voting in the parliament, which occurred in 04/17/16 and the subsequent tweets.
- The country from which the tweet came from is also used to comparisons.

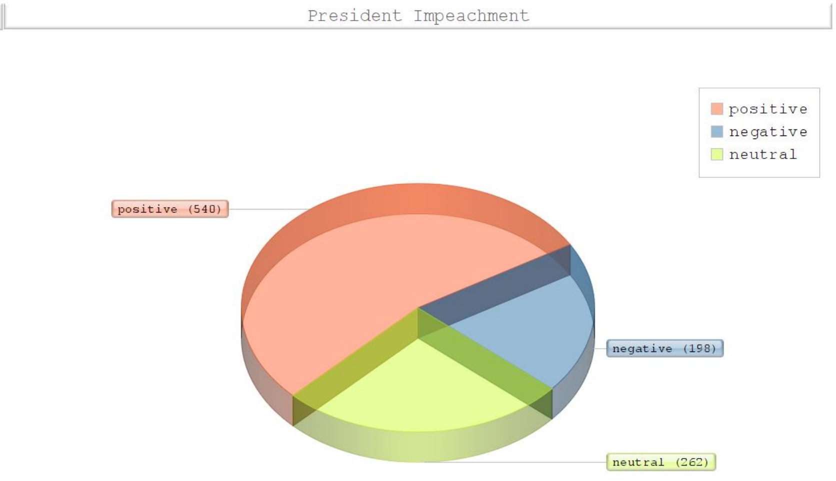
Results

- Using 160 classified tweets

150 tweets as test set

1000 tweets as test set

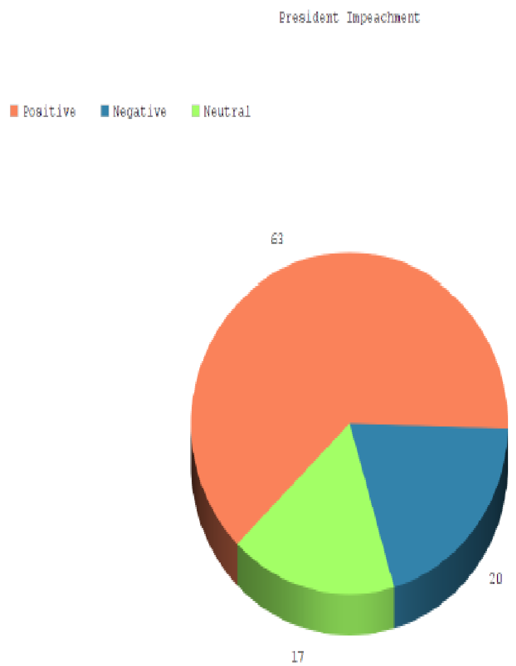




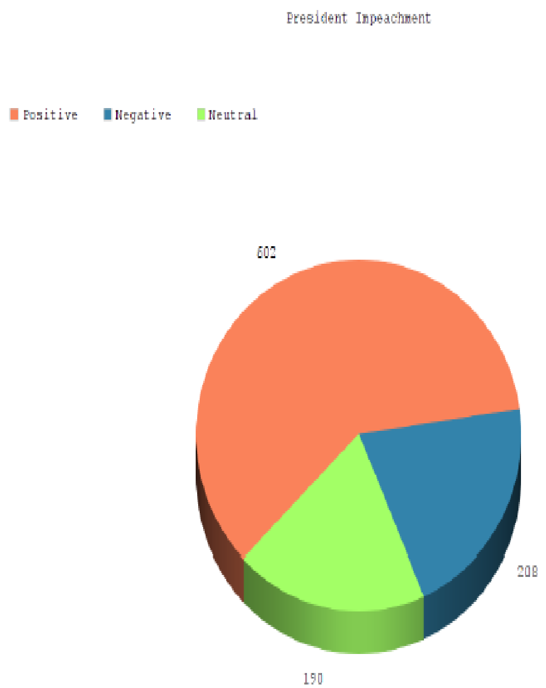
Tweets Before the first act of impeachment (after 04/17/2016)

b) Using 160 Classified Tweets (exactly the same quantity for each class

100 Tweets



1000 Tweets



Tweets After the first act of impeachment (after 04/17/2016)

As we can see, the number of people in favor of the impeachment has increased around the world as well, even taking into consideration the accuracy of the algorithm in the world collection, that was described in the world part before the act of impeachment.

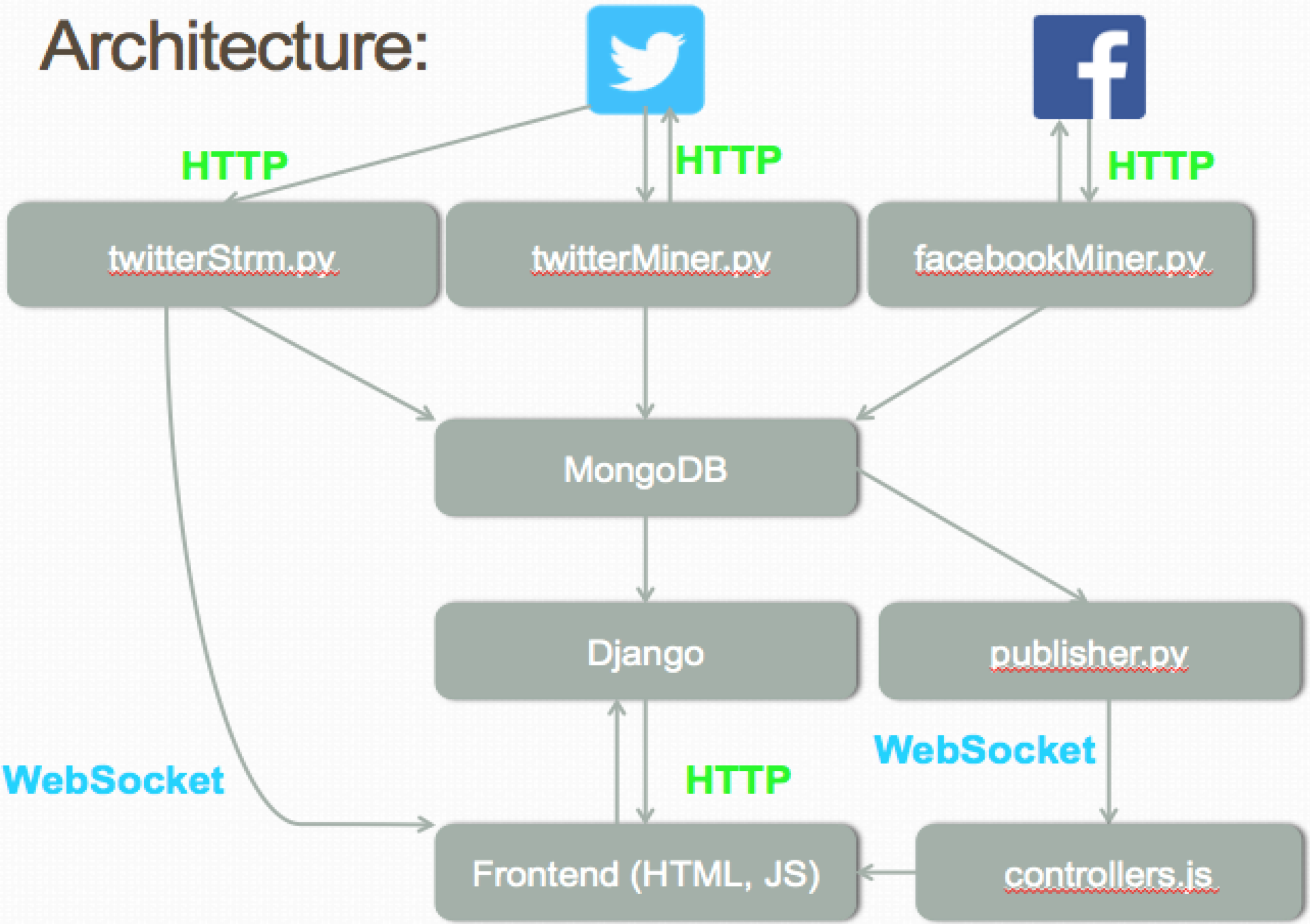
Real-Time Political Sentiment Analysis

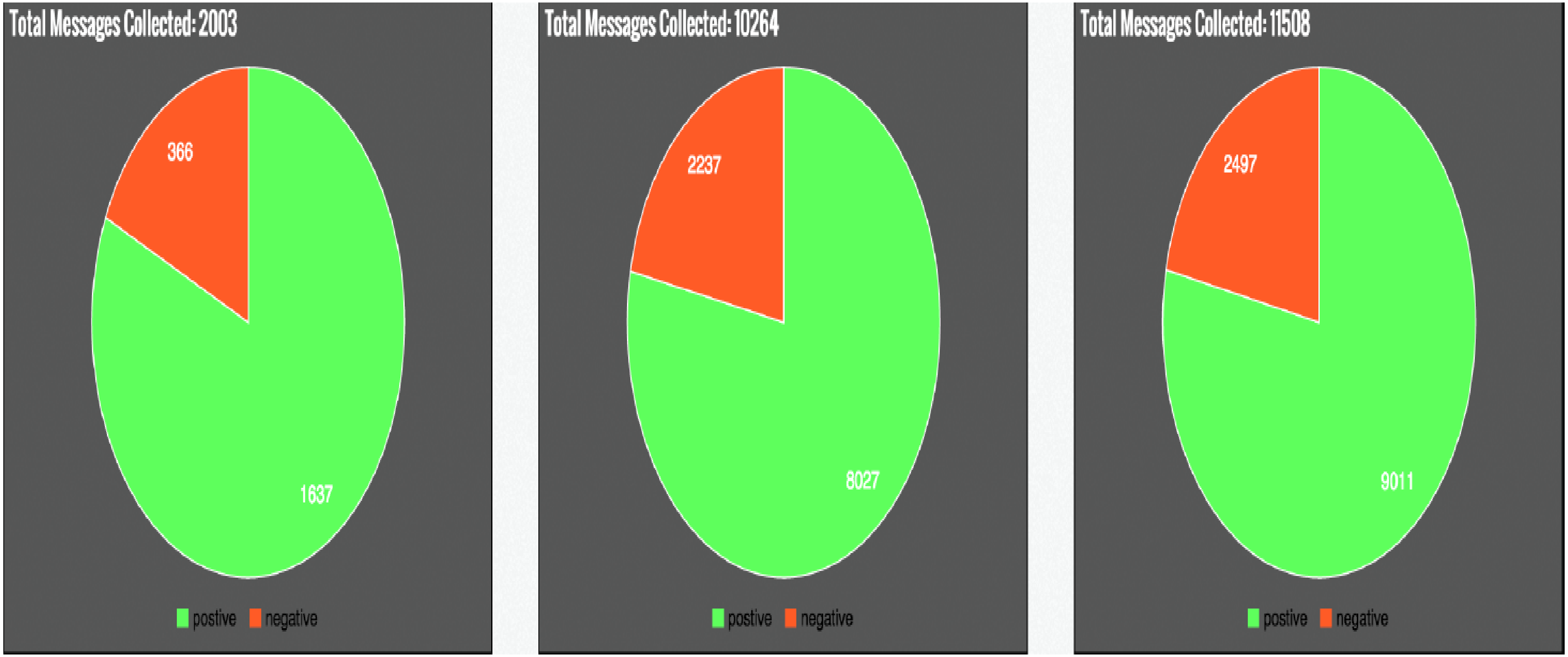
Gabriel Smadi
CSE 400 & CSE489 Term Project
Prof. Yu

Purpose:

- Predict to some extent who will be the next Governor of Puerto Rico
- Use sentiment analysis to find relative support for each candidate

Architecture:





The first chart is for Lugaro, the second for Rossello and the third for Padilla. Total messages gathered from all three politicians in total are about 23,775. Where this is the sum of all tweets, posts and comments gathered. Currently, Lugaro leads with about 81.7% of positive feedback, second comes Padilla with a 78.3% positive feedback and lastly Rossello with a 78.0% positive feedback.

Cuse Pulse

by Carl Poole and Terence Nip

Where do people tweet from the most in the SU area?

What is the current mood of the SU network?

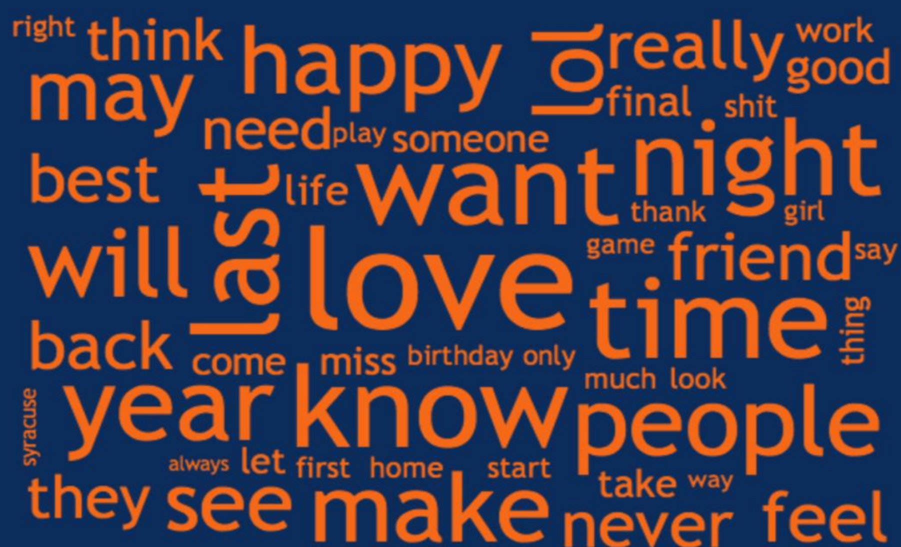
What are the hottest topics in the SU network?

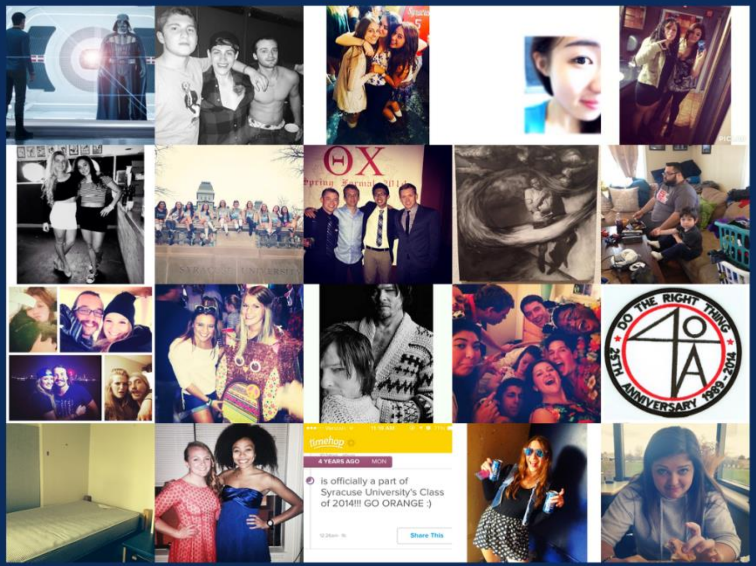
Does weather in Syracuse drive the most tweets?

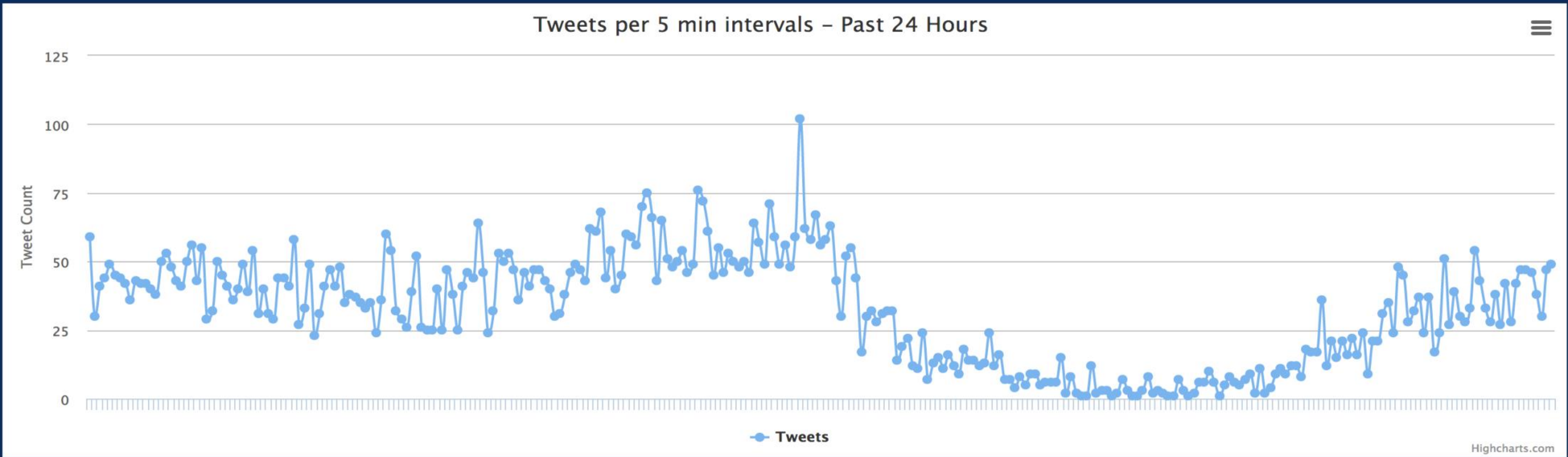
Do people tweet more during or outside of class?

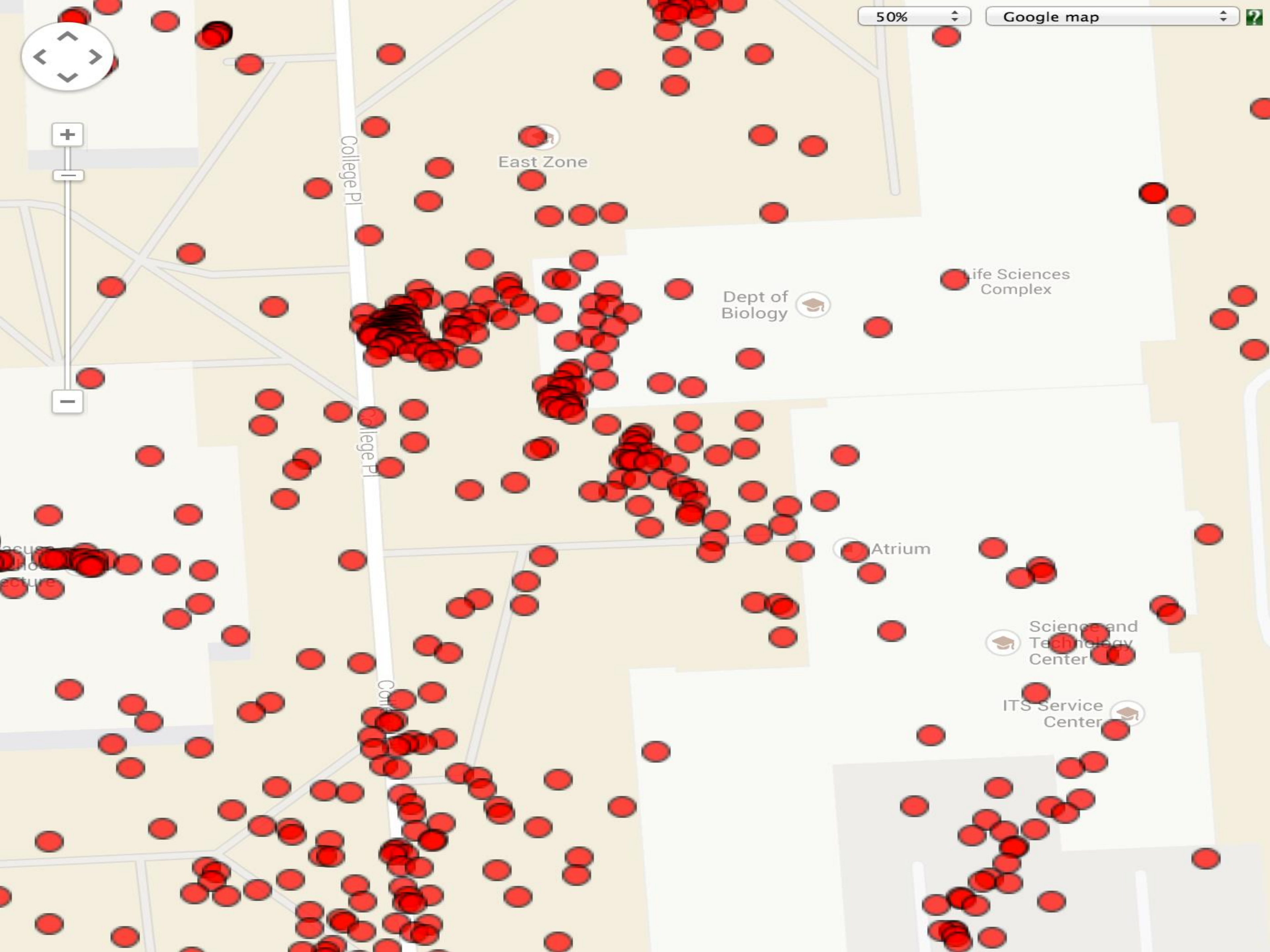
Using Twitter to Mine the Syracuse University social network

Tweets Last Hour








50%
Google map
College Pl
East Zone
Dept of Biology
Complex
Atrium
ITS Service Center

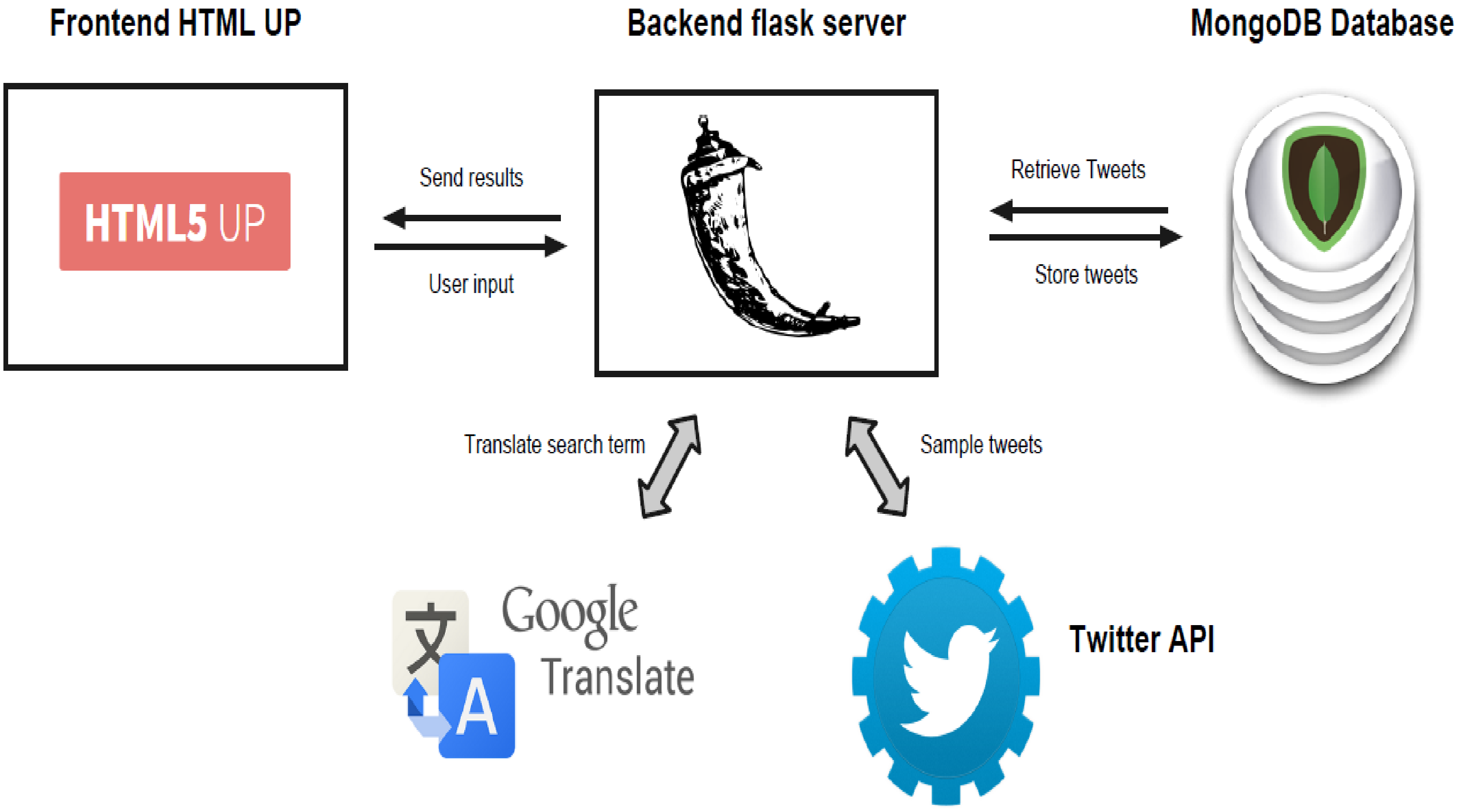
DGD - Data Gathering Determiner

Derrick, Grant, Daniel

Application idea

- **Purpose:**
  1. Web server that allows users to input a search
  2. Pull a sample of tweets with that search term
  3. Store tweets in a database
  4. Perform sentiment analysis on stored tweets
  5. Return analysis results to frontend

Application Architecture



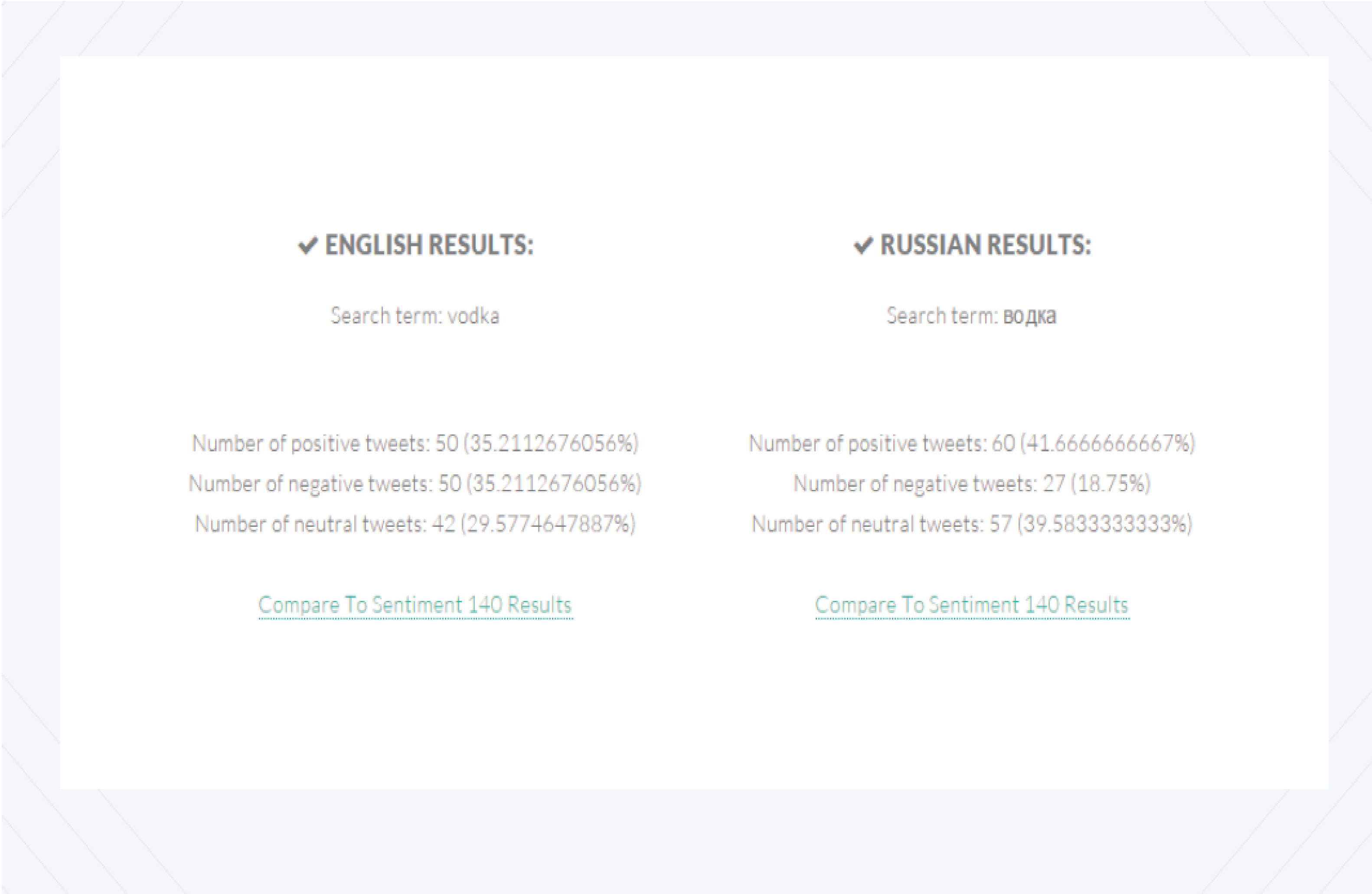
DGD BY DANIEL, GRANT, AND DERRICK

DATA GATHERING DETERMINER

THIS IS **DATA GATHERING DETERMINER (DGD)**, A FREE TWITTER SENTIMENT ANALYSIS PLATFORM BY DANIEL, DERRICK, AND GRANT.

GET STARTED

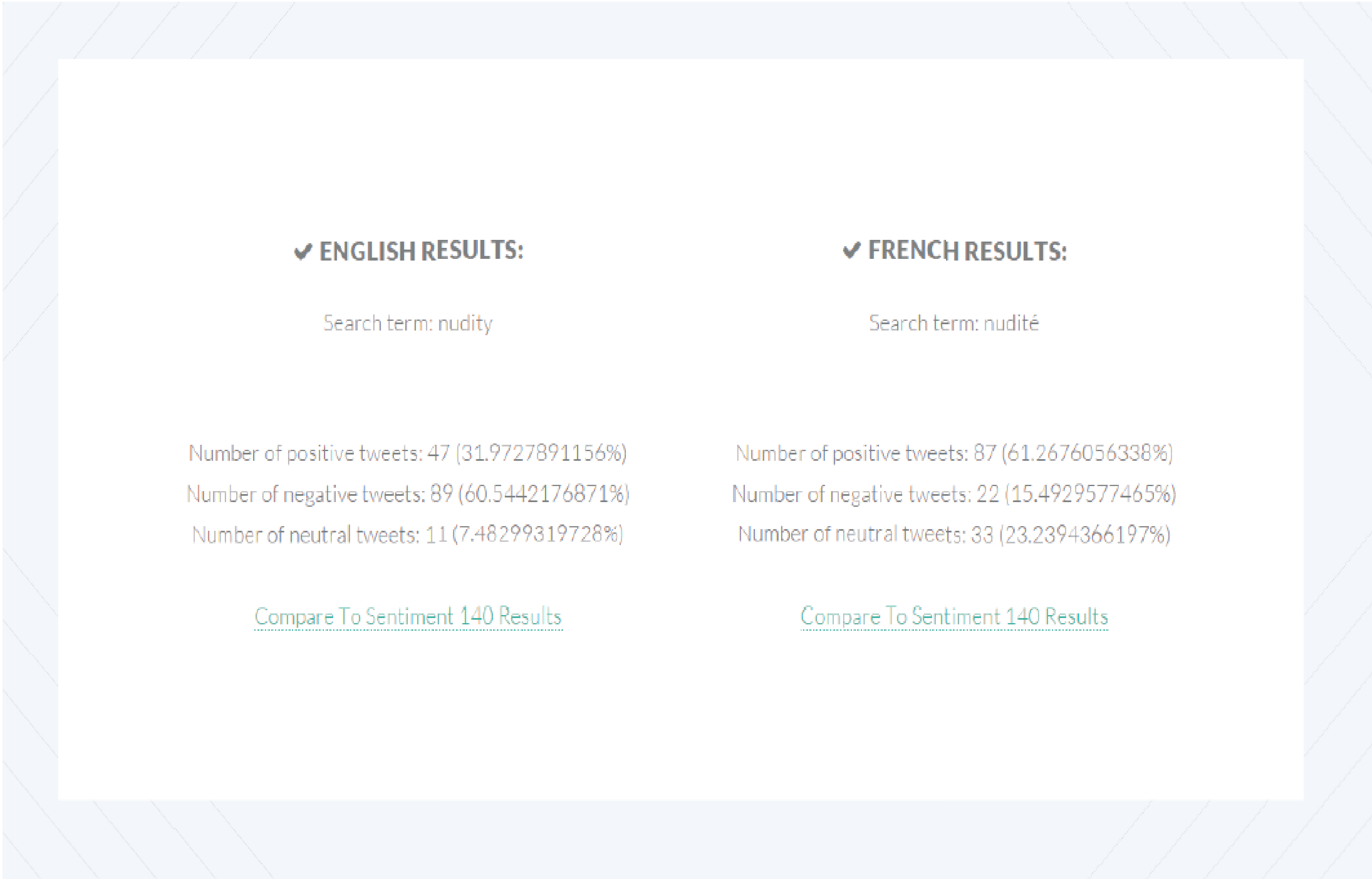
Results page



Prediction: Russian people generally are more tolerant of vodka, so they should have more positive tweets than negative.

Result: Mostly met expectations, though not too extremely.

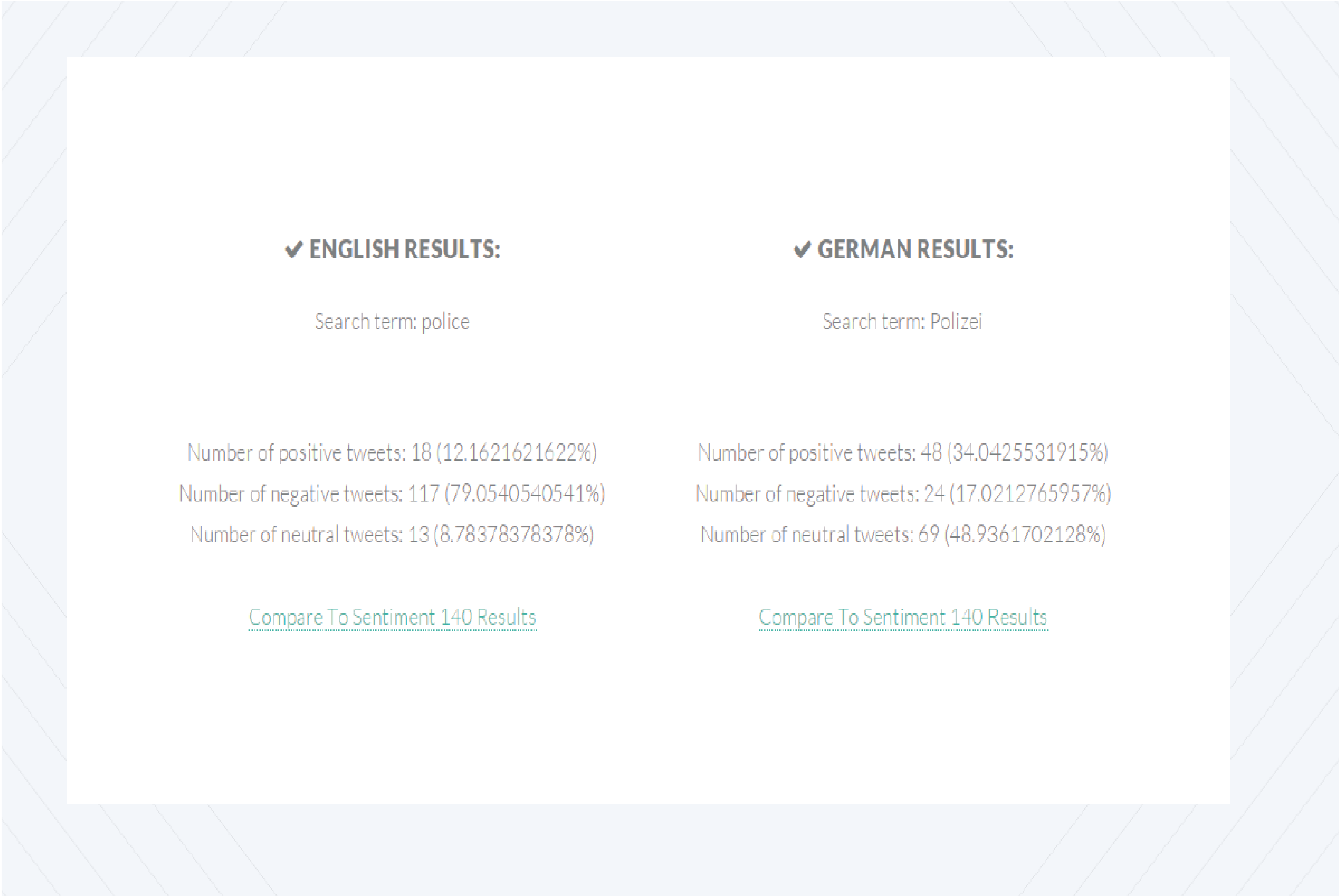
Results page



Prediction: Nudity is seen negatively in english speaking countries, whereas french speakers are likely to be more okay with nudity due to cultural reasons.

Result: Matches our prediction

Results page

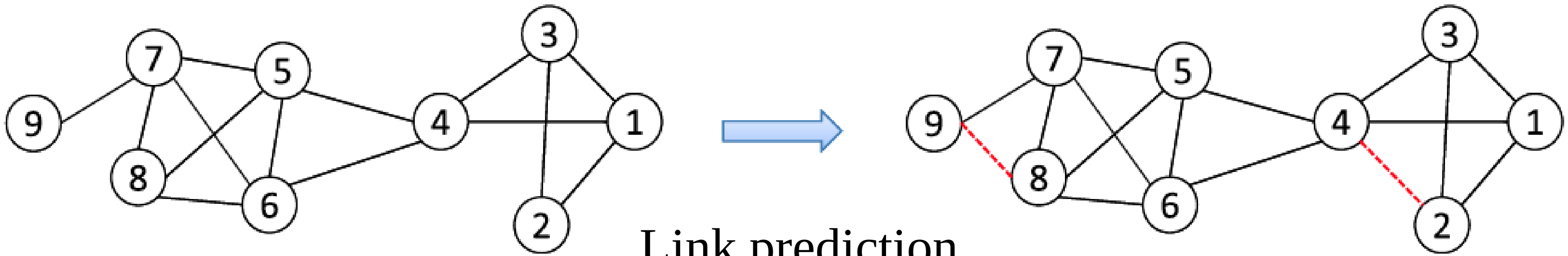


Prediction: Police has had a very poor image in the english media lately due to recent incidents. There will be a lot of negative tweets.

Result: Matches our prediction

Area 5: Recommendation

- Very common in social media applications
  - Tag, Friend, Group, Media, Link Recommendations



Link prediction

TRECT: A Hashtag Recommendation System for Twitter *

Mahmuda Rahman
mrahma01@syr.edu

Qinyun Zhu
qzhu02@syr.edu

Edmund Szu-Li Yu
esyu@syr.edu

Department of Electrical Engineering and Computer Science
Syracuse University
Syracuse, NY 13210

ABSTRACT

Hashtags appearing in the same status message in Twitter implies that they have some kind of relationship in their corresponding contextual domains. A "Retweet" of such a message contributes to the credibility and propagation of that message throughout the network. Also, "Reply" to a message has a high tendency to use the hashtags appeared in the original message as well as some more from the same context. In TRECT we considered the impacts reflected by the frequency of hashtags appearing together in a Twitter message to generate the score to rank a hashtag to recommend it to be used with relevant ones.

Categories and Subject Descriptors

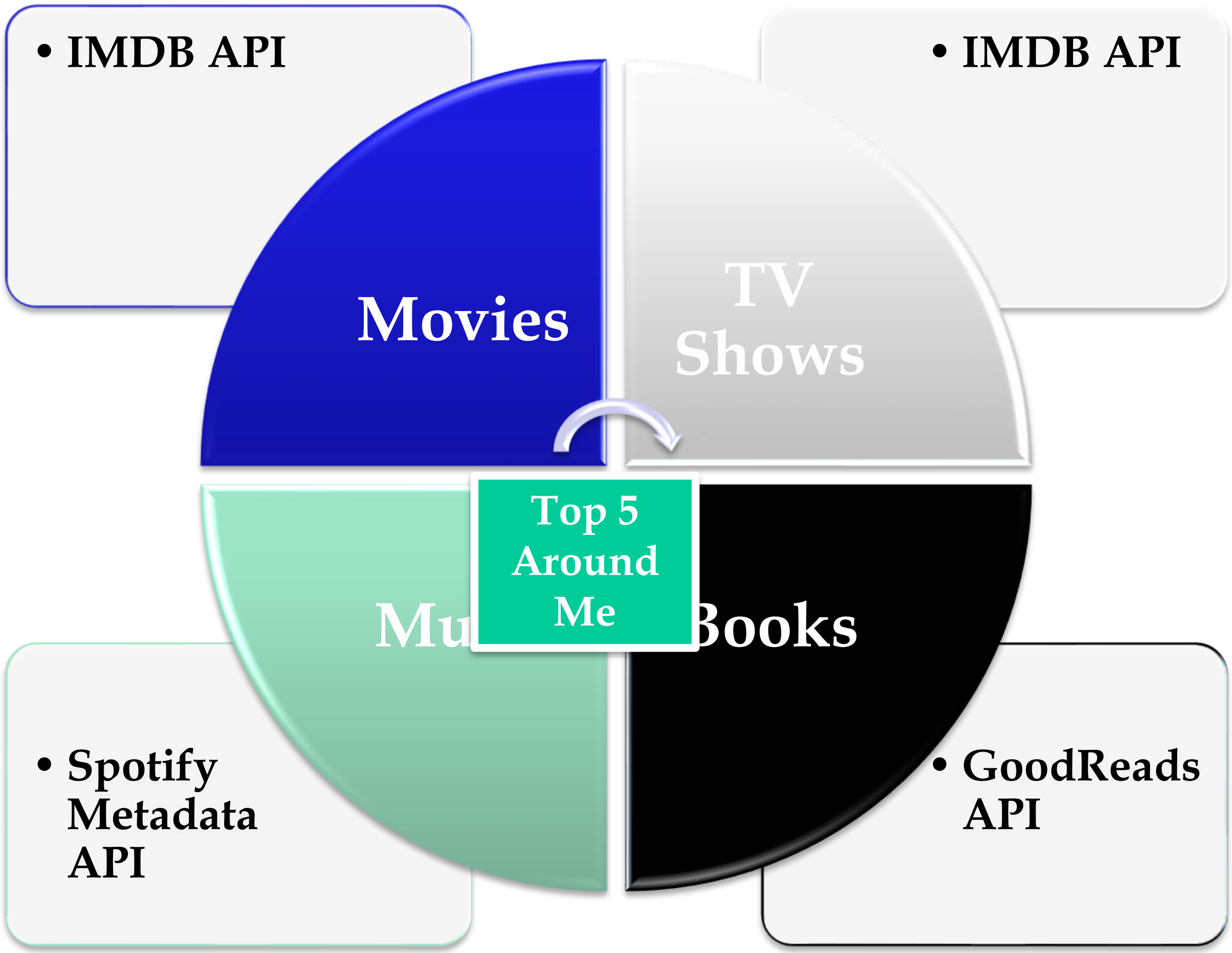
H.3.3 [**Information Search and Retrieval**]: Information filtering

some other related topics are very likely to emerge from the same context. But an single hashtag often fails to convey such context.

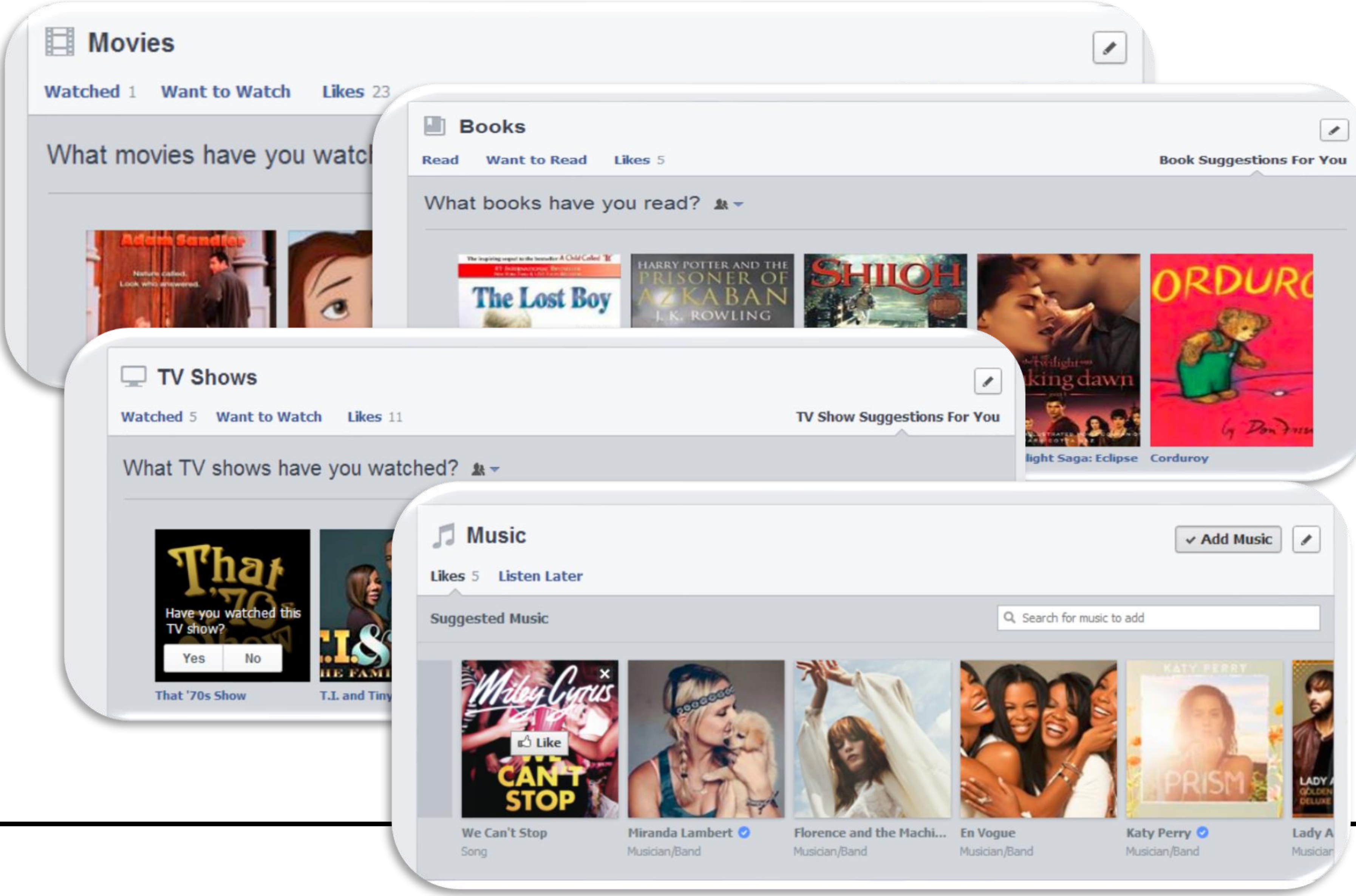
Therefore, we propose an effective suggestion mechanism for hashtags from the pool of existing ones to the those which are of the same interest. Contributions of this work is composed of following:

- We constructed "Hashtag Graph" by gathering all the status messages containing at least two hashtags for a small time slice using streaming API provided by Twitter. Same set of hashtags can repeatedly appear in different messages including retweets and replies. This develops a multi-graph where each node stands for a number of occurrence of that specific pair of hashtags and corresponding edges represent each such occurrence.

Interest Circle: A Term Project



Interest Circle on Facebook



Project Specs

- **Application -** A website with integrated Facebook login
- **Technology -** HTML5, CSS, Python/JavaScript, AJAX
- **External Web API**
  - Facebook Graph API
  - IMDB API
  - Spotify Metadata API
  - Goodreads API

FACEBOOK MATCH MAKER: A Term Project

- Data mining using facebook API
- Facebook based compatibility analysis
- The app would return a percentage score of compatibility

FACEBOOK MATCH MAKER: A Term Project

- Compatibility analysis of facebook friends based on:
    - Likes, interests, activities
    - Name Compatibility
    - Numerology, Zodiac sign compatibility
    - https://github.com/amolpatil8187/facebook_mining