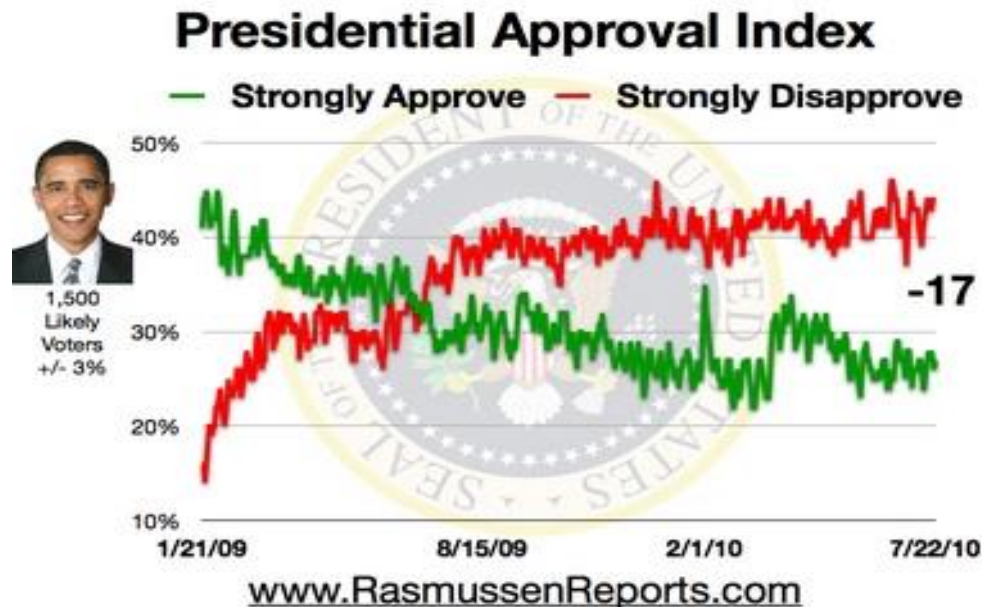


# Identifying Breakpoints in Public Opinion

**Cuneyt Gurcan Akcora**, Murat Ali Bayir, Murat Demirbas, \*Hakan Ferhatosmanoglu  
State University of New York at Buffalo  
\*Ohio State University

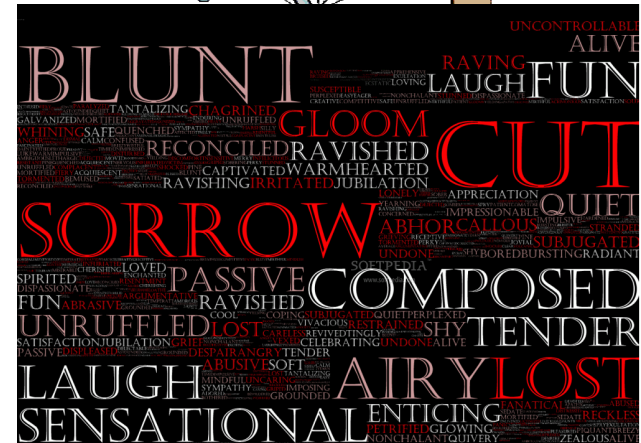
# Motivation

- Polling enables us to understand what public thinks about a certain topic.
  - Yet polling does not provide clues about how public opinion changes through time, and **why**?



# Problem Definition

- Can we use a microblogging tool (Twitter) to observe public opinion through time?



- Can we find the events that changed the public opinion?

# Observations: What signals Opinion Change?

- 1- An opinion changing event brings about a new **emotion pattern** in tweets. This new emotion pattern is repeated in the following time period(s).
  - 2- In the time periods following the event, public still post about the issue with quite **similar wording**.
- To claim a change in the public opinion, the emotion pattern and the word pattern must change according to these observations.

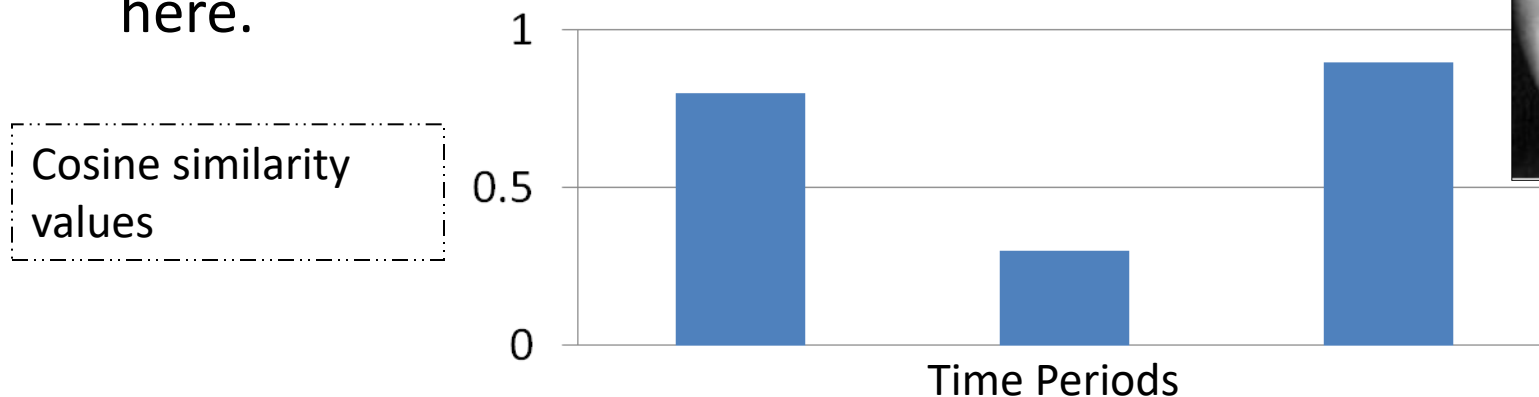
# Finding emotion(s) in Tweets

- Mining each tweet for emotions.
- Employing multi-grained sentiment analysis with eight emotion classes:
  - Anger, Fear, Sadness, Love, Joy, Shame, Surprise, Disgust
- Mapping a tweet to an eight dimensional emotion vector:

**Tweet:** *I was on main street in Norfolk when I heard about tiger woods updates and it made me feel angry, on 2009-12-11.*  
**Emotion vector:** *(1, 0, 0, 0, 0, 0, 0, 0).*
- A tweet can have more than one emotion.

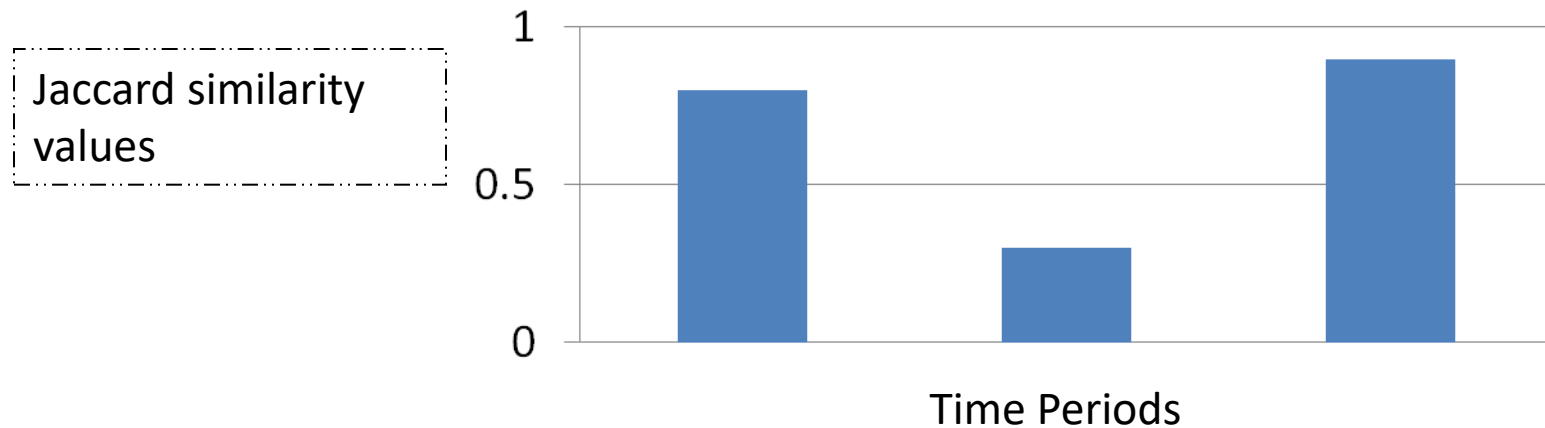
# Cosine Similarity for detecting emotion pattern change

- We will define time periods, and compare the emotion patterns of two successive time periods.
- For every  $n=12$  hour period, find a **centroid of the emotion vectors from tweets**.
- For every two successive periods, calculate cosine similarity of **centroids**.
  - The newly emerging **emotion pattern** is detected here.



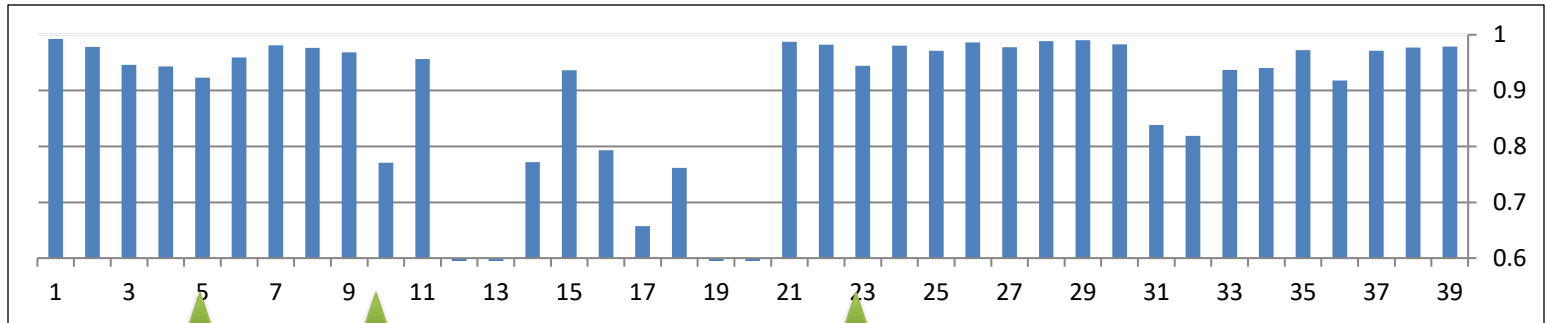
# Jaccard Similarity for detecting word pattern change

- For every  $n=12$  hour period, get a bag of words from tweets.
- For every two successive periods, calculate the Jaccard similarity.
  - The newly emerging **word pattern** is detected here.



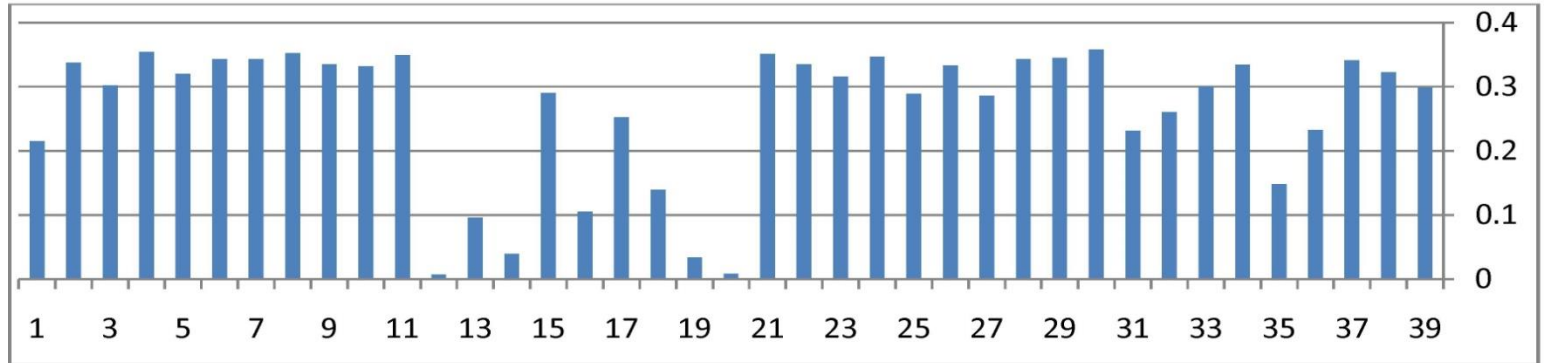
# Combining the two similarity methods

Cosine Similarity



▲ Breakpoint  
▲ False Positive

Jaccard Similarity

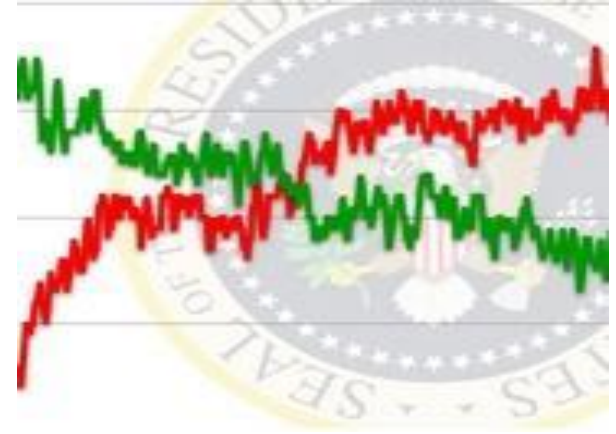


Similarity graphs for Tiger Woods case



# How do we account for the breakpoints?

- Can we summarize details of the event(s) that brought about that change?
- Breakpoint Representation: Extracting keywords to describe breakpoints.



# Breakpoint Representation

- We are detecting  $p=5$  prominent words to summarize the breakpoint.
- The Streaming TfIdf Algorithm

$$Tf_{x,D_n} = \frac{Count_{x,D_n}}{\sum_m Count_{m,D_n}}$$

$$Idf_{x,D_n} = \log\left(\frac{n}{|\{\forall k, k \leq n : x \in D_k\}|}\right)$$

- $n$  is increasing with every new period.

# Breakpoint Representation

- Document Update Phase:

$$tf_{x,D_o} = tf_{x,D_o} + F(D_o, D_n) \times tf_{x,D_n}$$

$$F(D_o, D_n) = 1/(n - o)$$

- Sort all words of a period according to their TfIdf values, and set top-5 words as the prominent words of that document.

# Breakpoint Representation

- Prominent words for the Tiger Woods case:

Period	Prominent words
1	crash, florida, injur, golf, accident
6	crash, wife, accident, mistress, golf
11	voicemail, wife, f***, golf, cheat
24	drop, stop, santa, claus, gatorade

# Explaining Breakpoints

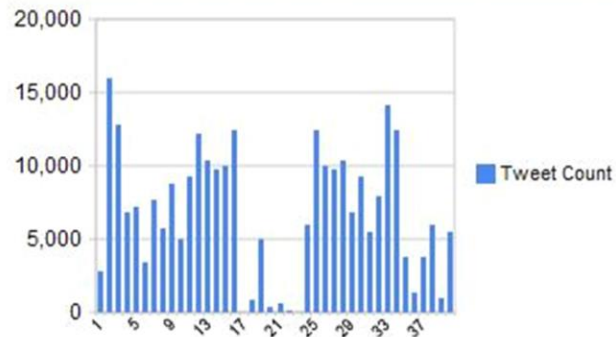
## Upinion Project

UB Computer Science and Engineering

[About Upinion](#) | [Contact](#) | [Supervisor](#)

[Home](#) | [Statistics](#) | [Add a Trendy Topic](#) | [Phases](#) | [Corpus](#) | [Team](#)

### Upinion Event Tracker for Tiger Woods Case



Prominent Words for 12/2/2009 : women, wife, mistress, golf, cheat

#### Breakpoints

[1st Period](#) | [6th Period](#) | [11th Period](#) | [24th Period](#)

tiger woods women, wi Search x

powered by Google™

Web

#### [News - Hear Tiger Panic to Mistress: "My Wife](#)

Dec 2, 2009 ... The other **woman**: hear what alleged **Tiger Woods mistress** Rachel Uchitel .... girl **Golf** has a 4 letter term beginning with the letter "f" called "fore". .... It takes two to **cheat** and tiger agreed to **cheat** so hes not so ... [www.usmagazine.com](#)

#### [Tiger Woods Mistresses : Pictures, Videos,](#)

Jul 22, 2010 ... Big News on **Tiger Woods Mistresses**. Includes blogs, news ... **Tiger Woods** Played His Most Important Round of **Golf** Ever at the Masters ... A man **cheats** on his **wife** or girlfriend and she decides to give him another chance at fidelity. Yet in relationships where the **woman cheats**, the relationship often ... [www.huffingtonpost.com](#)

#### [Tiger Woods Mistresses: Full List of Women &](#)

Dec 10, 2009 ... **Tiger Woods** has so many