Visualizing Twitter Trends
by Ahmed Al-Asadi
Mentors: James Abello and Yifan Hu
Data

- tweets collected from the Twitter API
- Only 140 characters long
- Contains slang, abbreviations, misspelled words, and poor grammar
Challenges

- Creating a tokenizer to extract useful information from tweets
- Finding a method to visualize twitter data
- Making the clusters viewable by removing any “useless” tweets (i.e. that don’t help in predicting trends)
Sample tweet:

@kwong @mpop @heftyleft @neel  Federer won his 7th Wimbledon title. Federer is the best tennis player in history: on.wsj.com/pcrqiz #bbctennis

Tokens: federer won wimbledon title tennis player history
Tokenizing

• Removes:

1. usernames, URLs, hashtags, duplicate words
2. very common words and phrases
3. numbers, and words with non-latin characters

• Incorporates a language detector to keep only English tweets (created by Nakatani Shuoyo)
• Shortens elongations: coooooool -> cool
Graph

- Each node is a tweet
- Every tweet contains a vector of term frequency scores for every word in the total list of tweets
Graph

Tokenized tweet (1) obama healthcare
Tokenized tweet (2) college graduation
Tokenized tweet (3) obama speaker graduation

Create vectors with dimension \( n \), where \( n \) is the number of unique words in the corpus after tokenization

\(<\text{obama, healthcare, college, graduation, speaker}>\>

The value in each position is the frequency of the word in the corpus, but 0 if that word is not contained in the tweet

Vector for (1): <2, 1, 0, 0, 0>
Vector for (2): <0, 0, 1, 2, 0>
Vector for (3): <2, 0, 0, 2, 1>
Graph

- The edge weight between (tweet1, tweet2) is the cosine similarity of the tweet’s term frequency vectors.

\[
\text{cosine similarity} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}
\]

Note that \(0 \leq \text{cosine similarity} \leq 1\), where 1 means the tweets are the same and 0 means the tweets are very different (i.e. no words in common).
Graph

- We add an additional criterion: if a pair of tweets don’t have at least two words in common, we give the edge weight between the tweets a value of 0.
- Afterwards, we prune all edges with edge weight 0 and additionally we remove all isolated nodes.
- In the end, we keep all the nodes that have degrees that are more than two standard deviations away from the mean, and prune the rest.
Visualization

- We input the graph into graphviz, a graph visualization program, and cluster the graph
- Clustering groups sets of related nodes
- We then input the resulting graph into gvmap, which emphasizes clustering relationships on a map
Results

- This is a map of ~40,000 tweets from July 8\textsuperscript{th}, 2012. Each color corresponds to a different cluster.

- In the following slides, there is a more close-up picture of each cluster.
RT @annabellegrant: I don't understand how people can think Nicki Minaj is attractive. Is she even buy tickets to see her?

RT @tr4ngpak: wait is nicki minaj shitting into the microphone or is that what she sounds like live?!

Nicki minaj is amazing can't wait to go and see her :-) #excited

I havnt been so jelouse this weekend as i am right now watchin nicki minaj !!!

RT @HazzaP_10: Sorry but Nicki Minaj is so rotten #howler

Stop laughing hahaha

Talving nicki minaj at T in the park On the telly has got me so excited to see her live!!!

Gin: ..... RT @He_CuteeAsFuck: Yoo WTF.??? :O RT @You_Fuckface: Holly shittt Nicki Minaj :O!!!! http://t.co/ ...
Results

Greatly saddened by the news of the passing of actor Ernest Borgnine. R.I.P. to the legend that is Ernest Borgnine.

Wow that's old-very cool. RT @ABCWorldNews: BREAKING: Oscar-winning film star Ernest Borgnine dies at 95, AP reports http://t.co/6I5WSMo

RT @msnbc_breaking: Actor Ernest Borgnine dies at 95, AP reports http://t.co/6I5WSMo

RT @GMA: RT @ABCWorldNews #breakingnews: Ernest Borgnine has died

Ernest Borgnine - A legendary actor. #RIP

RIP Ernest Borgnine *Mouse Ears* #Marine

RT @CBSLA: #BREAKINGNEWS A spokesman for Ernest Borgnine says the Academy Award winner passed away at his home in Los Angeles.

RT @swirsk054: Â¬ @BreakingNews: Spokesman: Oscar-winning film star Ernest Borgnine has died in Los Angeles at age 95 - @APâMcHales Navy....

Ernest Borgnine, the legendary actor perh... http://t.co/3MVYLulg #celebrity #gossip

I miss the Great Circus Parade, btw. #Milwaukee  Loved watching Ernest Borgnine. ...

RT @washingtonpost: #BREAKINGNEWS: Oscar-winning actor Ernest Borgnine dies at 95 http://t.co/GkvNWvBQ
Actor Ernest Borgnine dead at 95
By the CNN Wire Staff
updated 7:20 PM EDT, Sun July 8, 2012

Ernest Borgnine died on Sunday, July 8, at age 95. He's pictured here in 1973. In his greatest acting achievement, he won an Academy Award for his role in 1955's "Marty."

Borgnine through the years
Results

One Direction. Favorite for The Wanted.

if you believe one direction have split you're silly, very silly.

LOL MAYBE THE USA ONE DIRECTION SPLITTED UP LOLOLOLOL LAFFIN 4 EFA

RT @teenagedirrtbag: looks like they're all going their separate ways. so much for One Direction

@1DCentral_ @1Dnews @ @The1DScene ONE DIRECTION ARE SCHEDULED TO BE ON THE ELLEN SHOW ON JULY 16

litting up... there was obviously more than one direction then #haha
Results

Perez Hilton: One Direction Will Break Up

07 Jul 12 By: laurencf Liked it? SQUEEZE IT

Like 124 Tweet +1 0
Results

Frank Ocean On Full Blast! #CantStopWontStop!

RT @AmHallo: Frank Ocean And His BoyFriendâ€¦! Awww â€“ >>> http://t.co/OtS5yPTx

RT @kisssmyapash: Frank Ocean and The Weeknd should

RT @Hot_Dxmn: np Frank Ocean

RT @lois_wilkesss: Frank Ocean com

Legend: #10FavoriteVocalists Frank Fucking Ocean.

I canâ€™t even sing frank ocean â€” thinking about youâ€” no more. â‚¬

Ash RT fake Frank Ocean but that was that real though I must say.

Real Niggas listen to Frank Ocean http://t.co/w8L6yG9h

#4FavoriteVocalists adele, beyonce, frank ocean, melanie fiona, i canâ€™t name 10

comes out with a song called sweet life lol double entendre and irony at the same dame time

All @DrunkBarney tweets be retweets of that fake ass Frank Ocean

Bonjour_ImDejjj: People Be Making Up Some Wack Ass Jokes About Frank Ocean Being Gay .

i only like 3 songs by frank ocean: novacane, thinking about you & we all try.
Results

R&B star Frank Ocean makes media waves with admission he’s gay

By THOMAS CONNER tconner@suntimes.com July 11, 2012 5:28PM
I'm pretty sure Harry potter can do more with his wand than Brian can. Lol

Harry Potter sucks

RT @tonksx7: If you've never cried over a fictional character I can only assume that you've never read Harry Potter.

RT @grace_clayne: if you don't love Harry Potter there is something seriously wrong with you.. #tbh

Harry Potter is complete.

re-reading all the harry potter books..... #nolife

Hmmm, can't be bothered to go to bed. Harry Potter here I come.

a Harry Potter marathon very soon. 17 & and a half hours of harry Potter, yesss

RT @alliemackie_: on the subject of harry potter, my chemistry class is watching harry potter and the deathly hallows. #nolife

Me & mitch held up twilight & harry potter to see which movie teddy would want to see. Twilight won.

http://t.co/kBpJ1VNK

RT @playsthetart: @_JessMelendez That's why the evil cat in Harry Potter is named for her! :-) 

Scotts talking about harry potter on Big Brother. He is now my favourite! #bbuk
Results

2:00 pm
Harry Potter and the Goblet of Fire
TV-PG | D, L, S, V
On Air Only

3:45 pm

5:30 pm
Harry Potter and the Order of the Phoenix
TV-PG | V
When a new professor's bureaucratic methods leave Hogwarts unprepared, a dedicated group of students teach how to defend against the dark arts. Based on a novel.
On Air Only

8:30 pm
Harry Potter and the Half-Blood Prince
TV-PG | L, V
Dumbledore tries to prepare Harry for the final battle with Voldemort, while the Wizard world turns and tragedy looms on the horizon. Based on a novel.
Results

Federer beats Andy Murray to win men's final http://t.co/88iSwfDp

Murray, maybe if you had kissed Luke you might have won

National pride soaring thanks to Andy Murray and Wimbledon http://t.co/3nTBCVGv via @p

New post: Roger Federer Defeats Andy Murray For His 7th Wimbledon Championship
Results

Roger Federer wins 7th Wimbledon title

Roger Federer of Switzerland celebrates with the trophy after winning the men's singles final against Andy Murray of Britain at the All England Lawn Tennis Championships at Wimbledon, England, Sunday, July 8, 2012.

Chris Lehourites, Associated Press
Sunday, July 8, 2012 | 11:16 a.m.
Results

An amazing new weight loss product! Lose 10lbs a week, I didn't even change my diet! http://t.co/gwa6FJta

An amazing new weight loss product! Lose 10lbs a week, I didn't even change my diet! http://t.co/PKLzI9Jj

An amazing new weight loss product! Lose 10lbs a week, I didn't even change my diet! http://t.co/1cY1Hmoh

An amazing new weight loss product! Lose 10lbs a week, I didn't even change my diet! http://t.co/G1gqicFE
Twitter hit by amazing weight loss spam attack [VIDEO]

Has your Twitter account been spewing out messages about an amazing weight loss product?

Many users' accounts have been posting messages like this, pointing to a website promoting a "miracle" Acai Berry diet:

An amazing new weight loss product! Lose 10lbs a week, I didn't even change my diet!
Given a cluster of events, can you tell what day it was?
Given a cluster of events, can you tell what day it was?

- Marissa Meyer becomes CEO of Yahoo -> July 16th
- Love and Hip Hop airs that day -> July 16th
- Jon Lord dies -> July 16th
- Ray Rice gets signed to the Ravens -> July 16th

Easy to tell!
Complexity

- The complexity is $O(n^2)$, because we consider all possible edges between tweets in the beginning.
Future work

- How do we name the clusters so that it best represents what people are talking about in that cluster?
- How can we see what trends change with time?
- Compare different ways of clustering
Acknowledgements

- Thanks to:

  James Abello, Yifan Hu, Neel Parikh, Michael Poplavski, and Kevin Wong for their help
References

- Nakatani Shuoyo, author of the language detector [http://shuyo.wordpress.com/](http://shuyo.wordpress.com/)
- [http://beta.abcfamily.go.com/schedule](http://beta.abcfamily.go.com/schedule)