STUDYING DELETED TWEETS IN TWITTER

Aakash Anuj	10CS30043
Amrith Krishna	13CS60R12
Yetesh Chaudhary	10CS30044

Mentor: Parantapa Bhattacharya

Supervisor: Prof. Animesh Mukherjee

Roadmap

- MIDSEM EVAL:
  - Had a smaller dataset (≈64K)
  - Were missing concrete distinctions between deleted tweets and undeleted tweets
- NOW:
  - We have a much larger dataset (≈ 8M)
  - We have tried to make the best possible use of the random sample that we have !

Dataset

1% random sample – Spritzer API

Technicalities / Technical Challenges

- Non-English Tweets
  - Translatation of all tweets to English – GoSlate Library (Google API workaround for rate limit)
- Prediction of gender from first name using a Naïve Bayes Classifier
  - Source: http://stephenholiday.com/articles/2011/gender-prediction-with-python/
- POS Tagger for Twitter
  - CMU ARK (Used in our work) Vs. GATE PoS Tagger
- Wordnet for lexical analysis of tweets
- Latent Dirichlet Allocation (LDA) for finding out topics
  - Gibbs LDA

RESULTS

Statistics

	Deleted	*Undeleted*
Unique Users	*3.6 M*	*4.7 M*
% of deleted tweets containing links	*19.26*	*12.92*
% sensitive links in deleted tweets	*4.09*	*3.21*

More Statistics

	Deleted	Undeleted
% of verified users	0.059	0.13
Average number of followers	5794	1571
Average number of friends	1636	724

More Statistics

- We have a sufficient number of verified users in both deleted and undeleted tweets
  - Verified users are people whom users tend to follow a lot !
  - We can't say which of verified or unverified users delete more simply based on these counts
  - But we can definitely perform a lexical analysis on how their tweets differ in content

	Deleted	*Undeleted*
Verified	*4797*	*10978*
Unverified	*8014355*	*7989022*

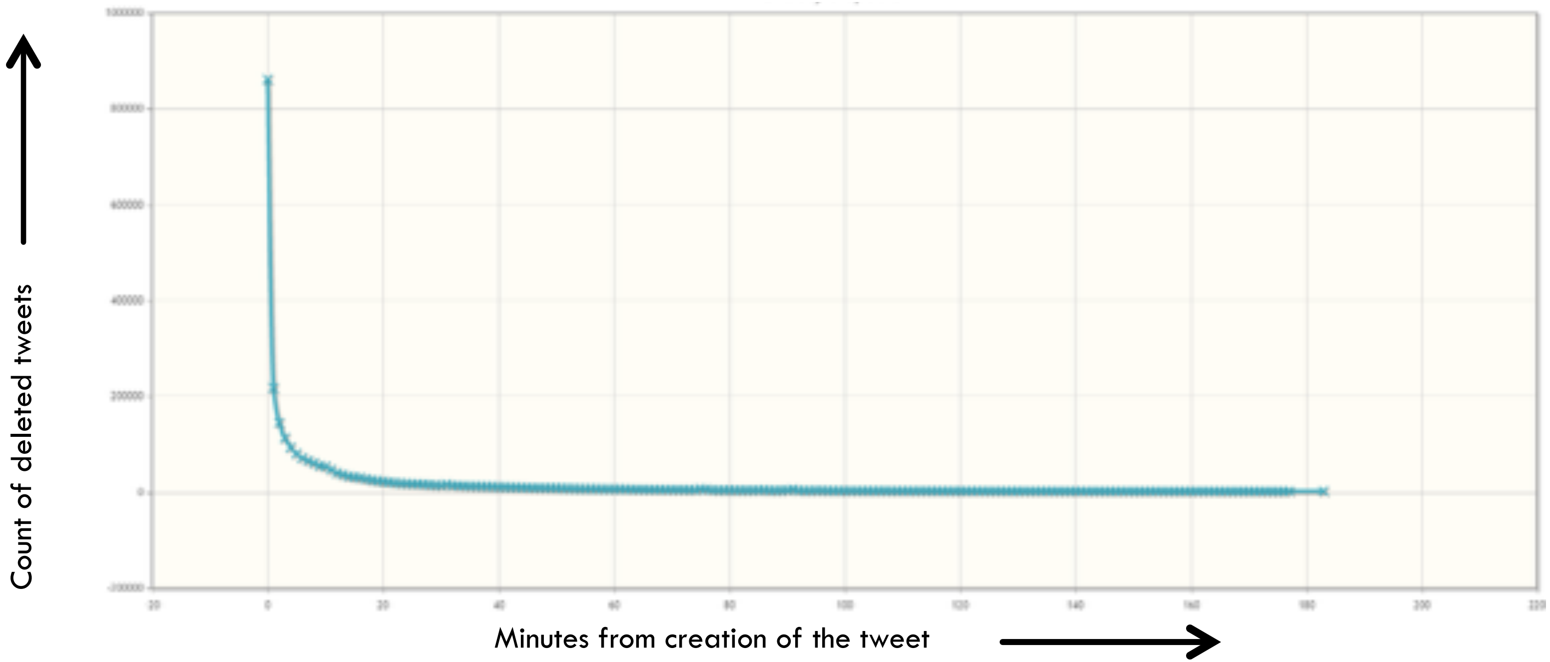
Breakup of Tweets

- Here is the breakup of tweets in terms of how many deleted tweets are status updates, replies or mentions

	Deleted	*Undeleted*
Status updates (%)	*44.74*	*44.96*
Replies (%)	*16.147*	*20.68*
Mentions (%)	*39.37*	*34.34*

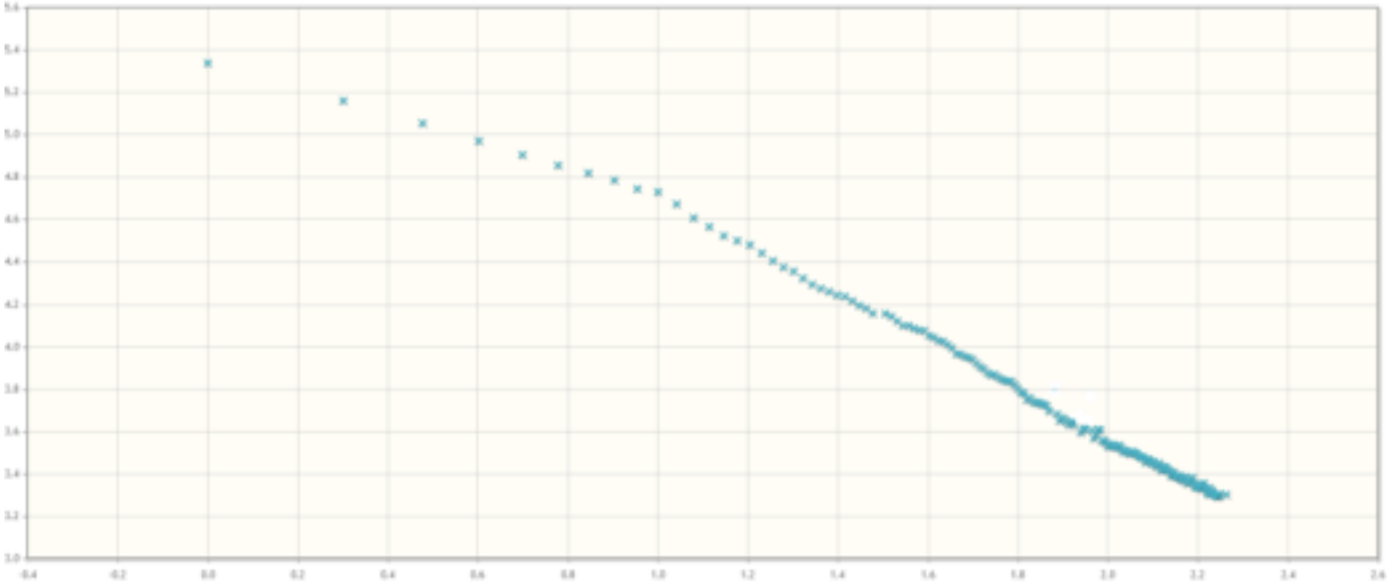
How fast is a tweet deleted ?

- **Follows POWER LAW !**



How fast is a tweet deleted ?

- **In log –log scale**



Count of deleted tweets

Minutes from creation of the tweet

Topic Comparison

- Now we take equal number of random samples from 2 ranges
  - Short range ($\leq$12 hours)
  - Long range (>5 days)

Now we apply **LDA** to compare topics in each range

Topic Comparison

≤ 12 hours	> 5 days

More of SPAM !

Less of SPAM

Gender Based Cursing

- We predict the gender using a **Naïve Bayes Classifier**

Sender	*Recipient*	*Total # of tweets in this category (deleted)*	*#Cursing Tweets (deleted)*	*Cursing Ratio (deleted)*	*Total # of tweets in this category (undeleted)*	*#Cursing Tweets (undeleted)*	*Cursing Ratio (undeleted)*
F	**M**	140091	7355	5.25	160252	7247	4.52
F	***F***	***55826***	***2986***	***5.35***	***64048***	***3041***	***4.75***
M	**F**	38849	2137	5.51	39221	1949	4.97
M	***M***	***132295***	***7399***	***5.95***	***135911***	***7390***	***5.44***

Verified vs. Unverified Users

Unverified

Verified

Words occurring in unverified topics, but not in verified

A topic found in unverified user's tweets

No such topic exists !

Regretted Content and Its Deletion

- To bring out the plausible relation between regr a ted tweets
- Done on 8M deleted and undeleted tweets
- We select 4 regrettable topics:
  - Alcohol and Drug abuse
  - Vulgar content
  - Religion and politics
  - Offensive comments

Alcohol and Drug abuse

Vulgar content

Offensive comments

Religion and politics

Reference for choosing topics :Tweets Are Forever: A Large-Scale Quantitative Analysis of Deleted Tweets, Almuhimedi et al.,CSCW '13

Regretted Content and Its Deletion

- The tweet is assigned to a regrettable topic if it contains at least one word from the topic word/collocation list

Regrettable topics	***Source***	***Keyword Count***	***Deleted (%)***	***Undeleted (%)***
Alcohol & Drug abuse	***Wordnet***	***62***	***0.34***	***0.37***
Vulgar content	***Wordnet***	***59***	***3.57***	***3.34***
Religion and Politics	***Wordnet***	***63***	***0.37***	***0.52***
Offensive comments	***Github repository***	***419***	***7.25***	***5.99***

Topic Comparison

- Now we categorize the deleted tweets of verified and unverified users into these 4 regrettable topics

	Verified(%)	Unverified(%)
Alcohol & Drug abuse	0.33	0.28
Vulgar content	2.65	3.41
Religion and Politics	0.63	1.01
Offensive comments	3.62	5.21

Geographical analysis

- We took the ratio of the presence of countries in deleted tweets to undeleted tweets
- Compare topics in countries having a high ratio to that having a low ratio using LDA

Country	Ratio
Turkey	1.73
Norway	1.42
United States	1.06
Japan	0.95
Germany	0.88

Country	Ratio
Indonesia	0.59
Argentina	0.57
Portugal	0.53
Malaysia	0.50
South Africa	0.42

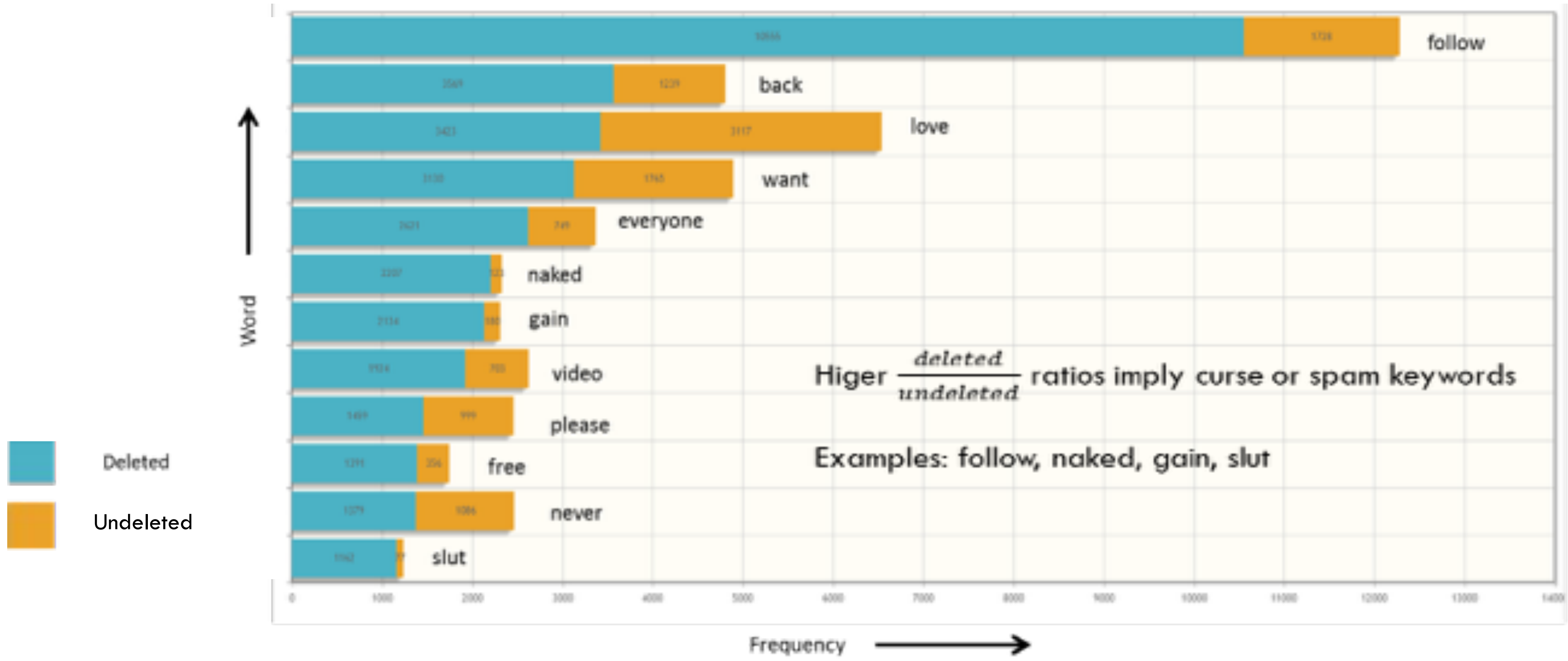
Geographical analysis

- Comparing topics between the two classes of countries using LDA gives some interesting results

A SPAM topic in case of countries having a high ratio

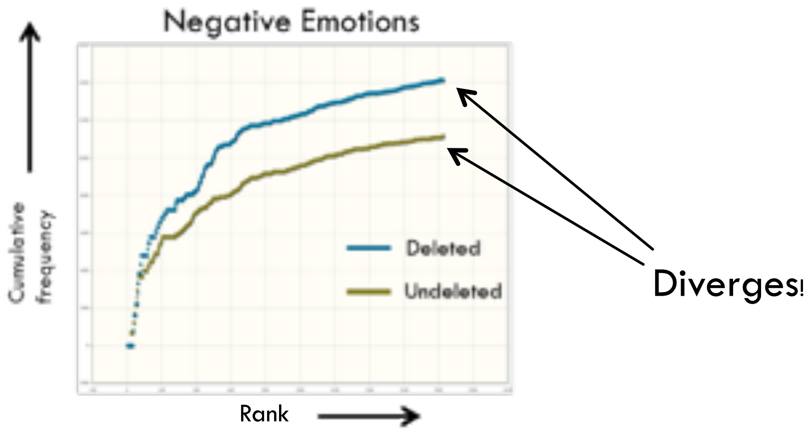
No such topic exists in the latter group !

Most frequently used terms



Positive and Negative emotions (AFINN)

- Cumulative frequency of positive and negative words for both deleted and undeleted tweets
- AFINN is a list of English words used in social networks **rated for valence** with an integer between -5 (negative) and +5 (positive)



Part of Speech (POS) distribution

- We analyze the POS tag distributions for both deleted tweets and undeleted tweets
- We see that the two categories have a significant difference in some POS's

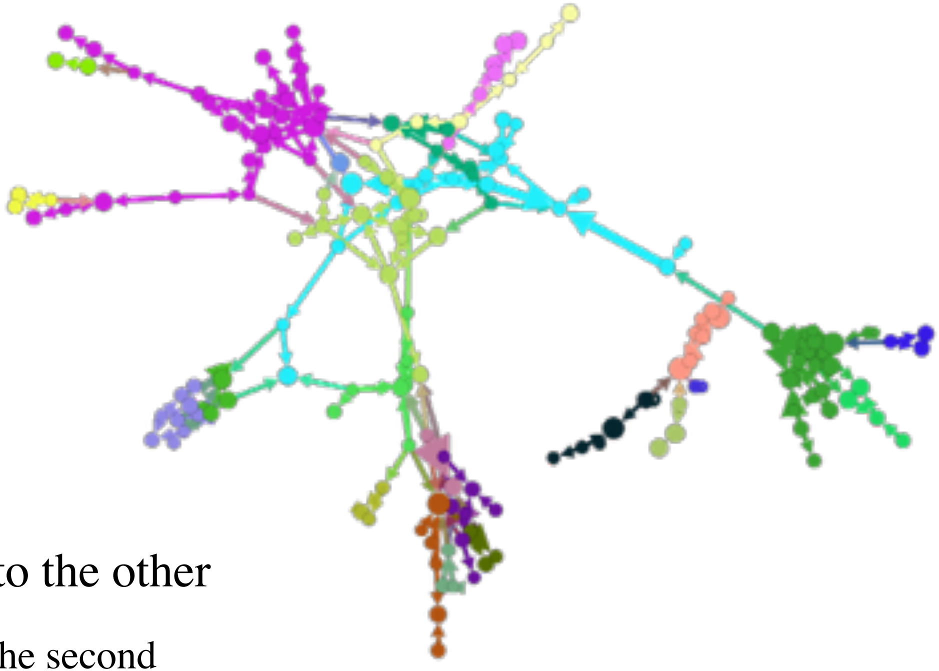
POS	N	^	S	Z	V	A	R	!	#	@	U	~
Ratio (Deleted/undeleted)	*0.99*	*0.92*	*0.77*	*0.70*	*1.01*	*0.99*	*1.02*	*0.93*	*0.86*	*0.94*	*1.13*	*1.01*
	↓	↓	↓	↓	↑	↓	↑	↓	↓	↓	↑	↑
P-value (Chi square test)	*<0.01*	*<0.01*	*<0.01*	*<0.01*	*<0.01*	*>0.05*	*<0.01*	*<0.01*	*<0.01*	*<0.01*	*<0.01*	*>0.05*

N: common noun
^ : proper noun
S : nominal + possessive
Z : proper noun + possessive

V : verb
A : adjective
R : adverb
! : interjection

U : URL / email
@ : at-mention
# : hashtag
~ : discourse marker

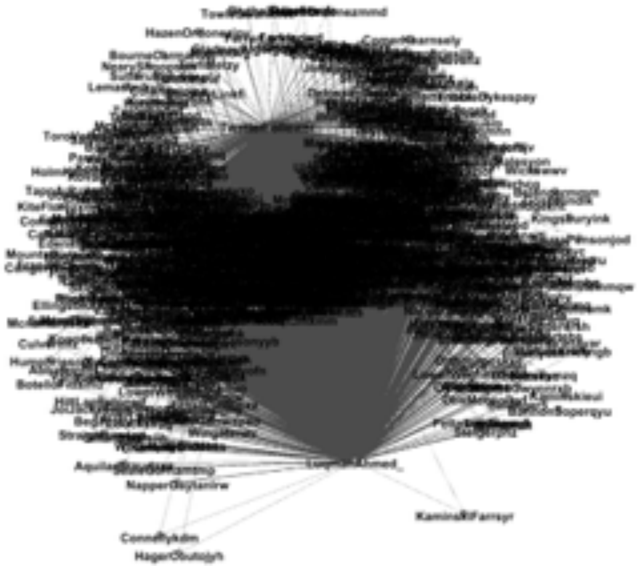
Network of mentions

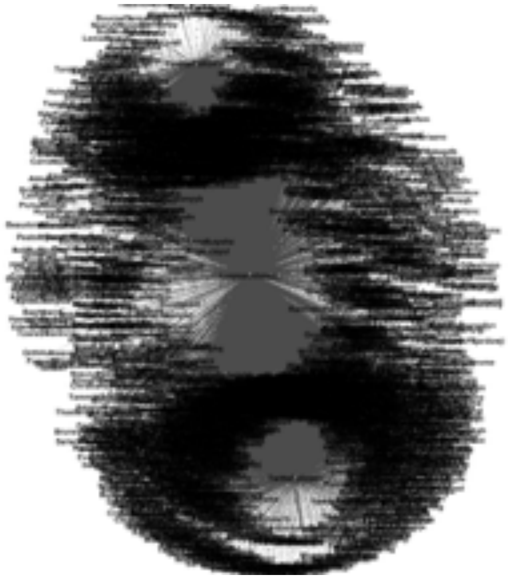


- Nodes: unique users
- Edge from one user to the other
  - First user mentions the second

EGO center in the mentions graph

- Ego centric graph of the nodes with highest in-degree (>4000)





EGO center in the mentions graph

- In case of undeleted tweets, the maximum degree in-degree was found out to be just 361

Conclusion

- The deletion time of tweets follows a power law; tweets getting deleted quickly containing more of spam
- Deleted tweets contain more curse words than undeleted tweets, with intra-gender cursing a lot more than inter-gender cursing
- Verified users are more decent in their tweeting content
- Countries with a high ratio of deleted to undeleted tweets spam more
- Adjectives do not differ much in the two streams, but all the other POS do
- Tweets containing mentions tend to be deleted more

References

- **Research papers:**

  - Self-Censorship on Facebook, Sauvik Das and Adam Kramer
  - Tweets Are Forever: A Large-Scale Quantitative Analysis of Deleted Tweets, Almuhimedi et al.,CSCW '13
  - I Wish I Didn't Say That! Analyzing and Predicting Deleted Messages in Twitter, Petrovic et al., CoRR, May 2013
  - Cursing in English on Twitter, Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, Amit Sheth, ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW 2014)

Thank you

Questions / Comments?