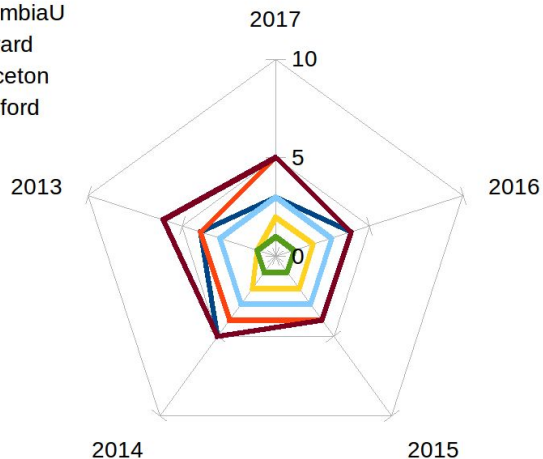


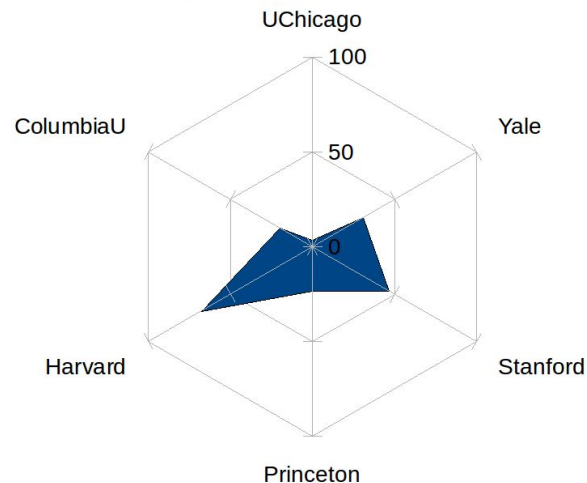
# UChicago: Capacity Building Strategies in Twitter

— UChicago  
— ColumbiaU  
— Harvard  
— Princeton  
— Stanford  
— Yale



US News National University Ranking (2013-2017)

■ Followers ( $10^4$ )



Followers of Official Twitter Account

by Yuxiao Sun

# Executive Summary

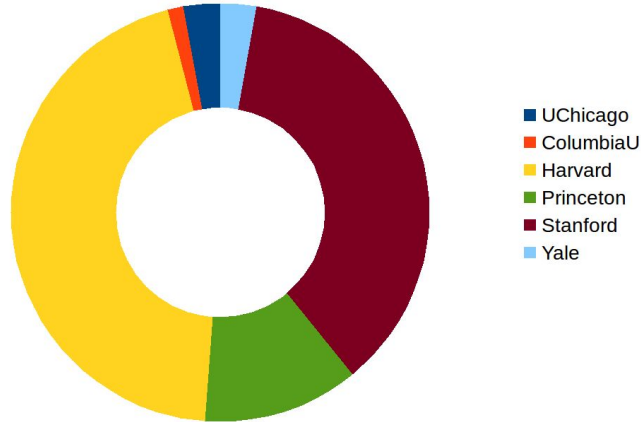
1. The university of Chicago (UChicago) is consistently ranked as a top 5 institutions (**University of Chicago, Columbia University, Harvard University, Princeton University, Stanford University & Yale University**) in the United States, according to U.S. News & World Report.
2. UChicago's impact in twitter is incompatible with its prestige: as a top university in US, the followers of its official twitter account is about **1/10** of its peer schools; the total number of tweets related to UChicago during our observation period is **1/7** of peer average; and the retweets rate for UChicago related tweets is about **1/2** of its peer average.
3. Recommendations for UChicago include: 1) diversify topics to augment audience space; 2) popularize information based on users' needs to increase retweets; 3) personalize its relationship with users for deeper bonding; and 4) apply text-analytics techniques to better understand social media.

# Data and Methodology

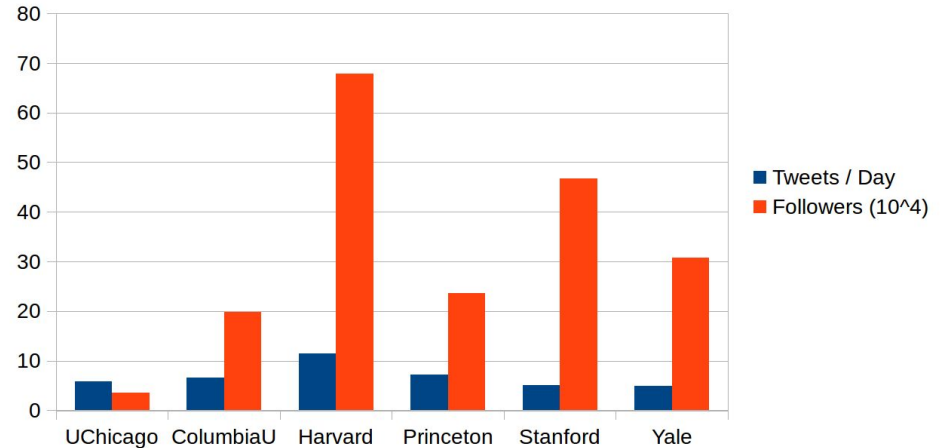
- Data source: 1.2 million tweets related to six top universities (UChicago, Columbia, Harvard, Princeton, Stanford, Yale) created between Oct. 2015 and Mar. 2016.
- Methodology: use Apache Pig (Elephant Bird) to retrieve tweets relevant to the top universities; use Sentiment Analysis, Topic Modelling (Latent Dirichlet Allocation) and Classification models (linear svm and multinomial naive bayes) for analysis.
- Pipeline:  
[21 Million Records]---(retrieve and filter)--->  
[1.2 Million Tweets Related to Top Universities]---(classify)--->  
[626,226 Tweets and 608,872 Retweets]---(model)-->[Recommendations]

# UChicago in Twitter: Impact

Comparison of Top 5 Universities in US: Number of Tweets (%)



Comparison of Top 5 Universities in US: Official Account



- The number of tweets related to UChicago (including University of Chicago) is **1/7** of its cohort average.
- The followers of official twitter account @UChicago is less than **1/10** of its cohort average, although it does produce a cohort average amount of tweets per day since its launch.
- **Harvard** and **Stanford** outperform other top universities both in terms of the number of mentions and followers; **Yale** has a large number of followers, but its tweets mentions are relatively low
- This project defines impact by the number of tweets mentions (over a period of time). And as it is shown above, in many cases this number concurs with the number of followers of official twitter account.

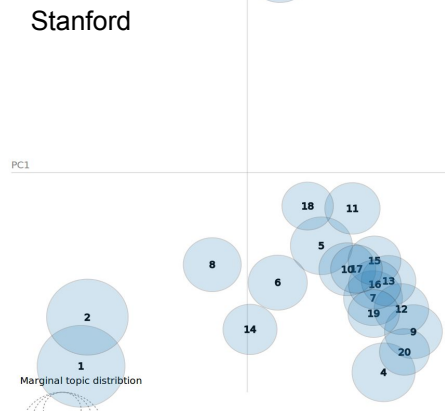
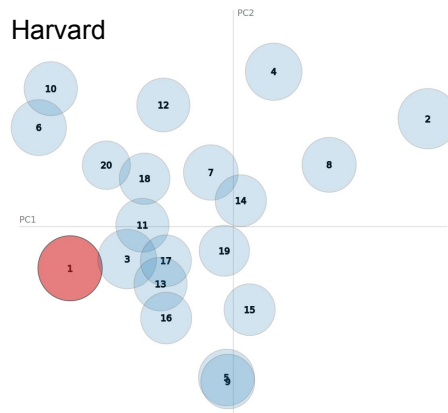
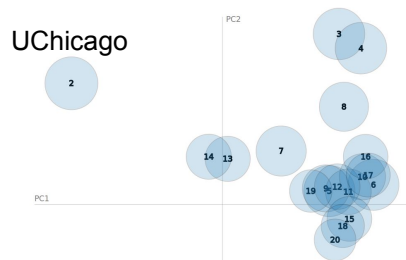
# UChicago in Twitter: Keywords

- **Goal:** find keywords in UChicago tweets among top universities between 10-2015 and 03-2016
- **Data:** 48,000 randomly selected tweets (8,000 for each university)
- **Model:** Multinomial Naive Bayes and linear SVM classifier. Train on 48,000 tweets (words and bigrams) and predict on 8,000 tweets. Average precision of SVM = 0.76; Average recall= 0.76

Results	UChicago	Harvard	Princeton	Sanford	Yale	ColumbiaU
1	bernie	business	review	usc	art gallery	virology
2	trauma	business review	arrest	cardinal	halloween	mailman
3	bernie sanders	referencing	bitcoin	iowa	hiring job	journalism
4	booth	encryption	brush	mccaffrey	racism	Exxon
5	medicine	bomb	imani perry	Pac 12	bulldog	frontier science

- As expected, the top keywords are usually **geo-related**. The table above shows the top 5 keywords that are not about geography. These can be viewed as hot spots for online discussion between 10-2015 and 03-2016.
- The advantage of a classification model over a word frequency counts model is that it can partial out common words/phrases and only leave entity-specific traits. However like the word counts model it cannot capture the **layers and structure** of tweets, which requires topic modelling techniques presented in the next slide.

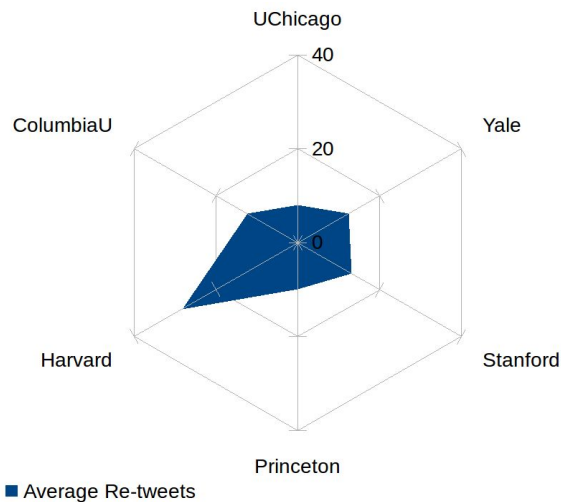
# Recommendation for Topics: Diversify



Inter-topic Differences: LDA models, topics = 20, PCA = 2

- **Goal:** detect topics in tweets.
- **Data:** 13,113 tweets related to UChicago
- **Model:** Latent Dirichlet Allocation.
- **Results:** LDA models each tweet as a combination of topics defined by a probability distribution of words. The focus here, however is the inter-topic differences measured by Jensen-Shannon divergence. From the figures above we can see that Harvard in particular have a very scattered distribution of topics. The average inter-topic difference for Yale, on the other hand, is the lowest among the four.
- **Recommendation:** Yale's low topics diversity might explain why with such a high followers population comparable to that of Stanford and Harvard, its total tweets mentions is less than 1/10 of the two. The lesson for UChicago, in addition to attract more followers, is the importance of topic diversity to increase brand awareness over a broader audience space.

# Recommendation for Retweets: Popularize

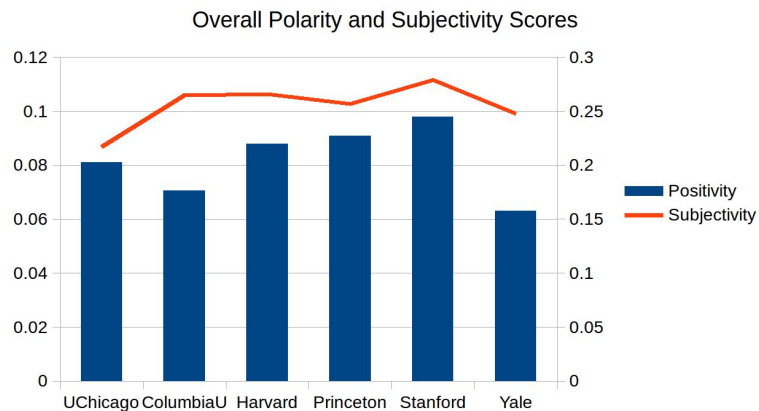


- **Recommendation:** share 1) important campus events; 2) interesting academic studies of its students and faculties; 3) admission and program information; 4) stories and anecdotes of famous alumni. All in all, for universities to get more retweets it is important to tailor the language and content of their tweets for a broader audience.

- **Goal:** find text features (words and bigrams) that attract high retweets.
- **Data:** 608,872 retweets->92,177 unique retweets
- **Model:** Naive Bayes and linear SVM classifier. Train on 80,000 retweets (words and bigrams) and predict on 12,177. Precision of SVM = 0.76; Recall = 0.80.
- **Results:** 1) UChicago tweets have the lowest average retweets rate; 2) top features for high retweets:

Type	Words/Bigrams
Academic	receive offer, grade intensive, study program, intensive reading, ecology
Events	entrepreneurial thought, leader seminar
Celebrity Stories	gate drop (Bill Gates); yeah dropped, dropped 11th

# Recommendation for Semantics: Personalize



- **Goal:** identify sentiments in the tweets
- **Data:** 626,226 tweets related to the top universities
- **Model:** Pattern Analyzer (TextBlob). Average the sentiment scores of word within a text.

- **Results:** 1) overall speaking, UChicago scores average for the polarity metrics but lowest for the subjectivity metrics; 2) although not perfect, the polarity metric in subareas seems to be a good reflection of the relative strength of universities in the area. Among the top universities, UChicago receives most positive sentiments in Economics related tweets, above average polarity scores in campus and belong average in tuition.

- **Recommendation:** UChicago's tweets have the lowest subjectivity score. This can be a sign of lack of personal bondings between users and the university. As suggested by the close relationship between subjectivity score and tweets popularity, it is important to encourage users to interact with the university community and share their personal feelings.

