

Twitter Data

Gilad Mishne - @gilad

Twitter Search

Info 290 - Analyzing Big Data With Twitter

UC Berkeley Information School

August 2012

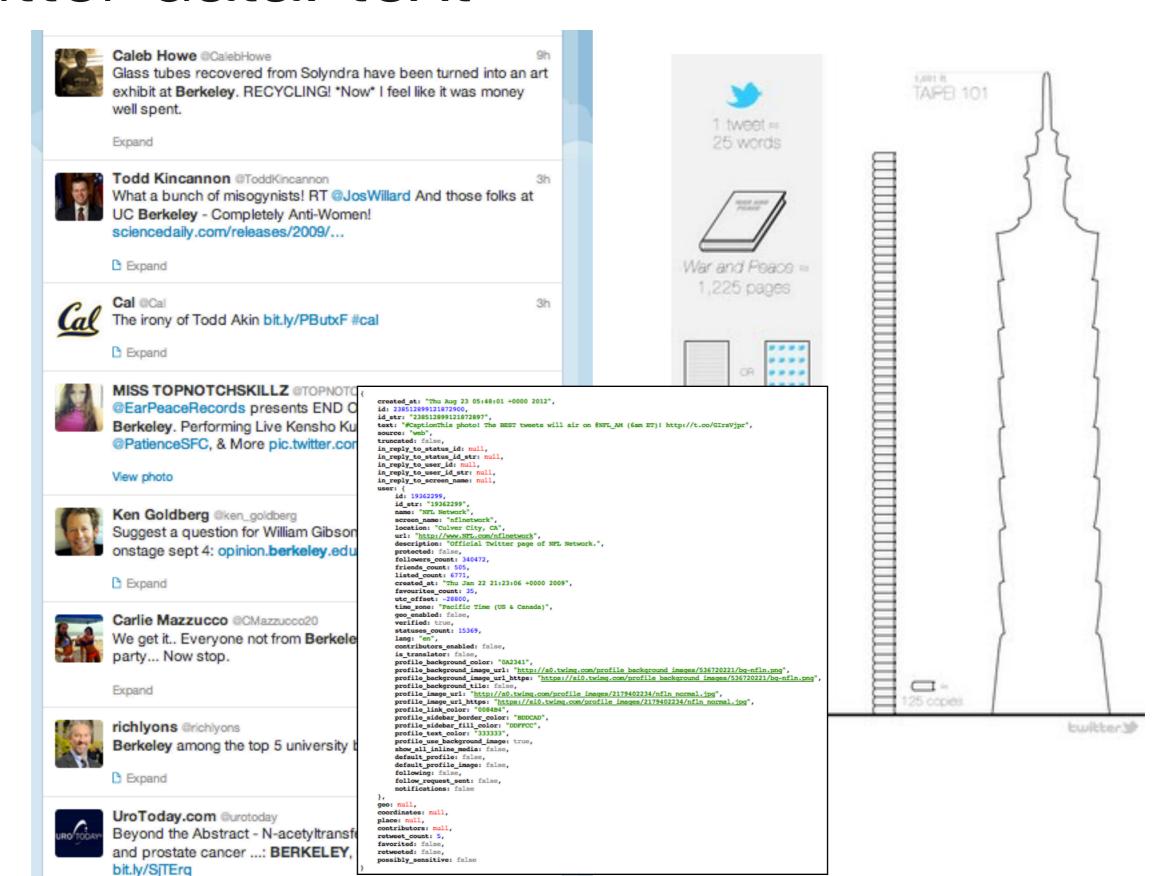
About Twitter

The fastest, simplest way to communicate

- More than 140M active users
 - Majority (also) mobile; 60% out of U.S.
- More than 400M twitter.com visitors
- More than 400M tweets/day (peak: 25K/sec)

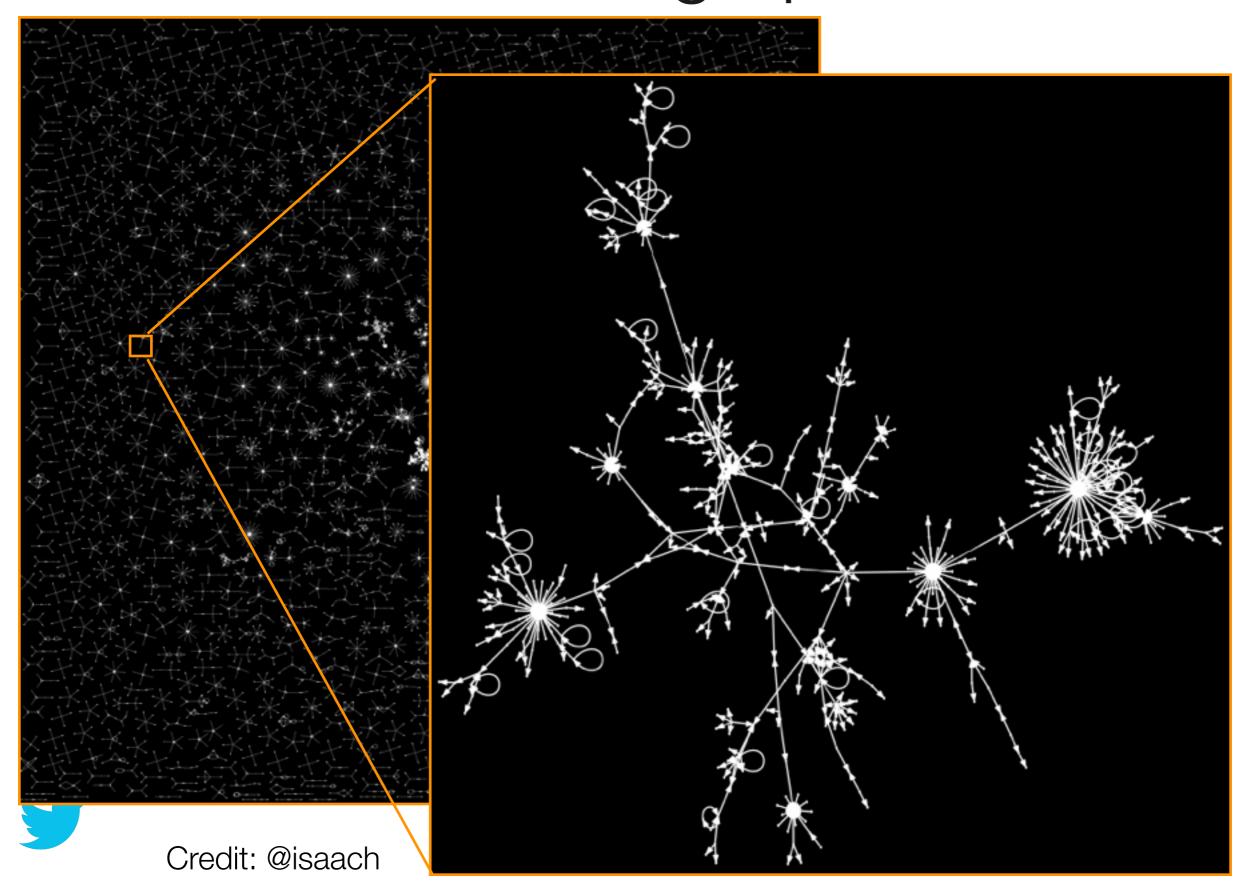
- 1,000 employees (majority in San Francisco)
 - 50% engineers

Twitter data: text

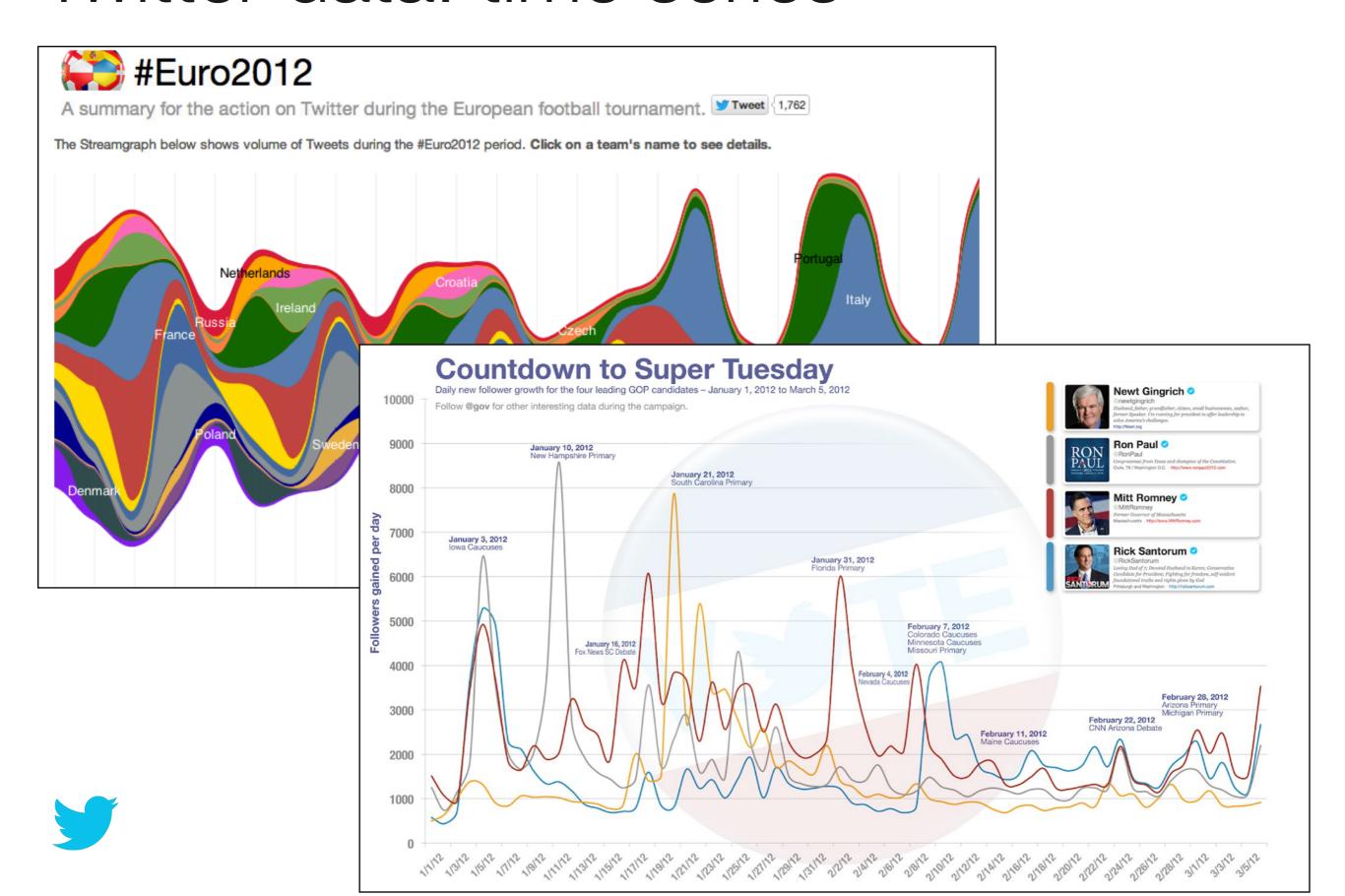




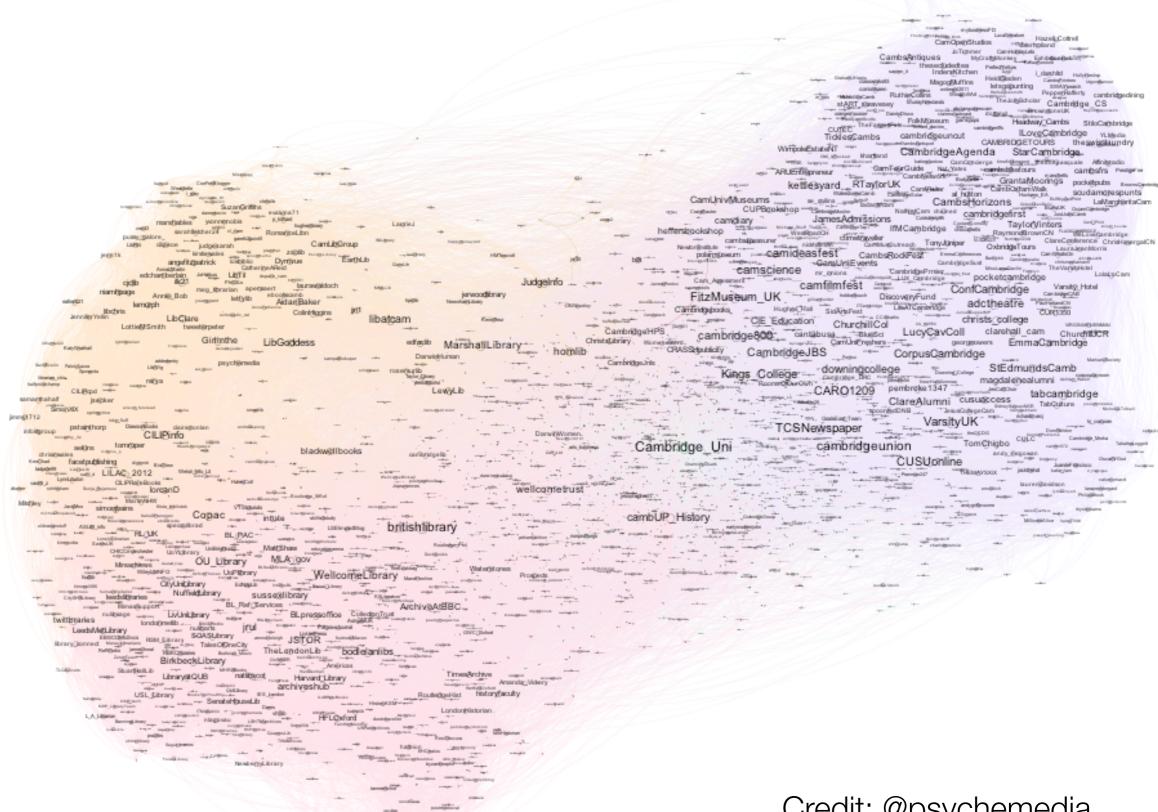
Twitter data: social graph



Twitter data: time series



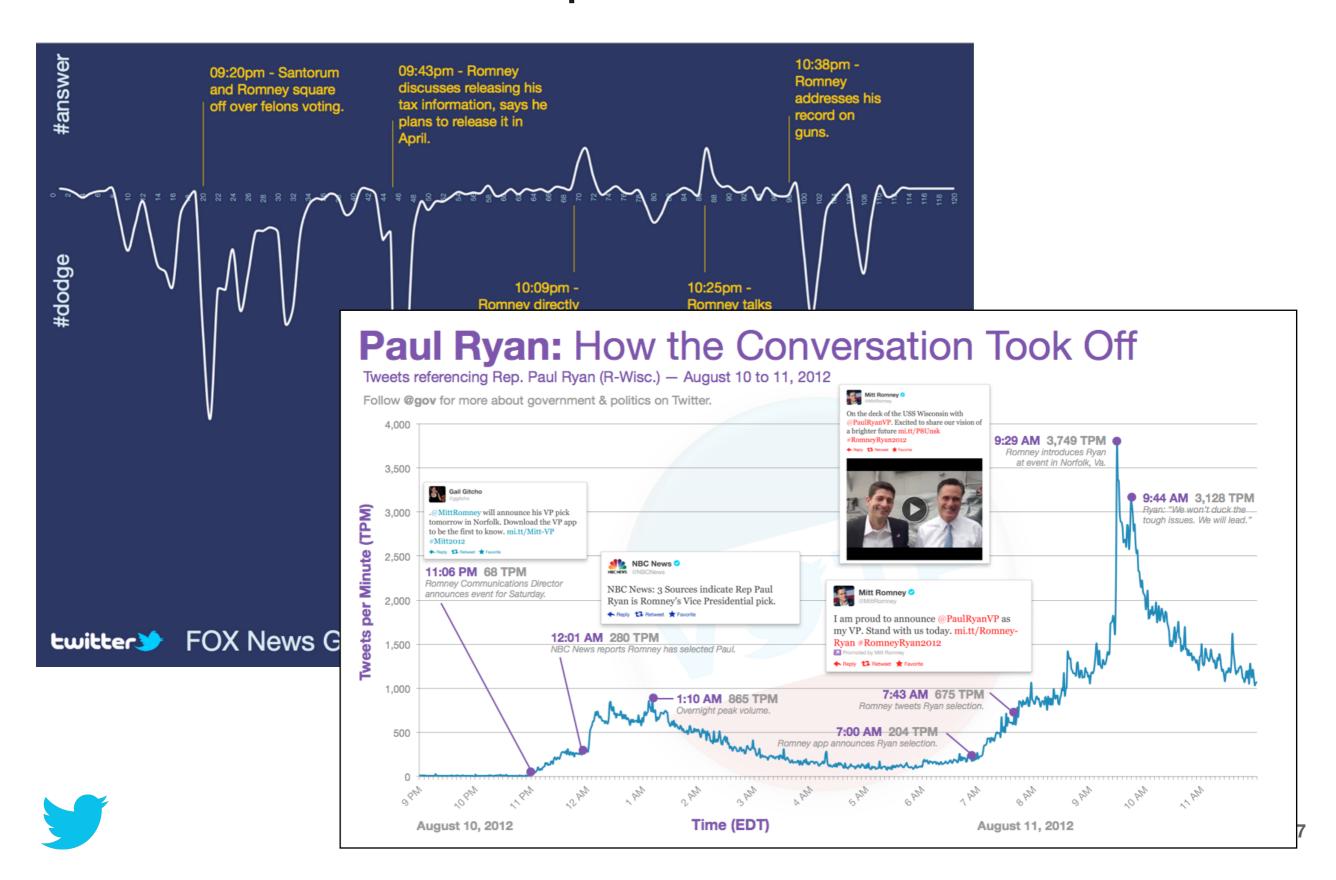
Twitter data: interest graph





Credit: @psychemedia

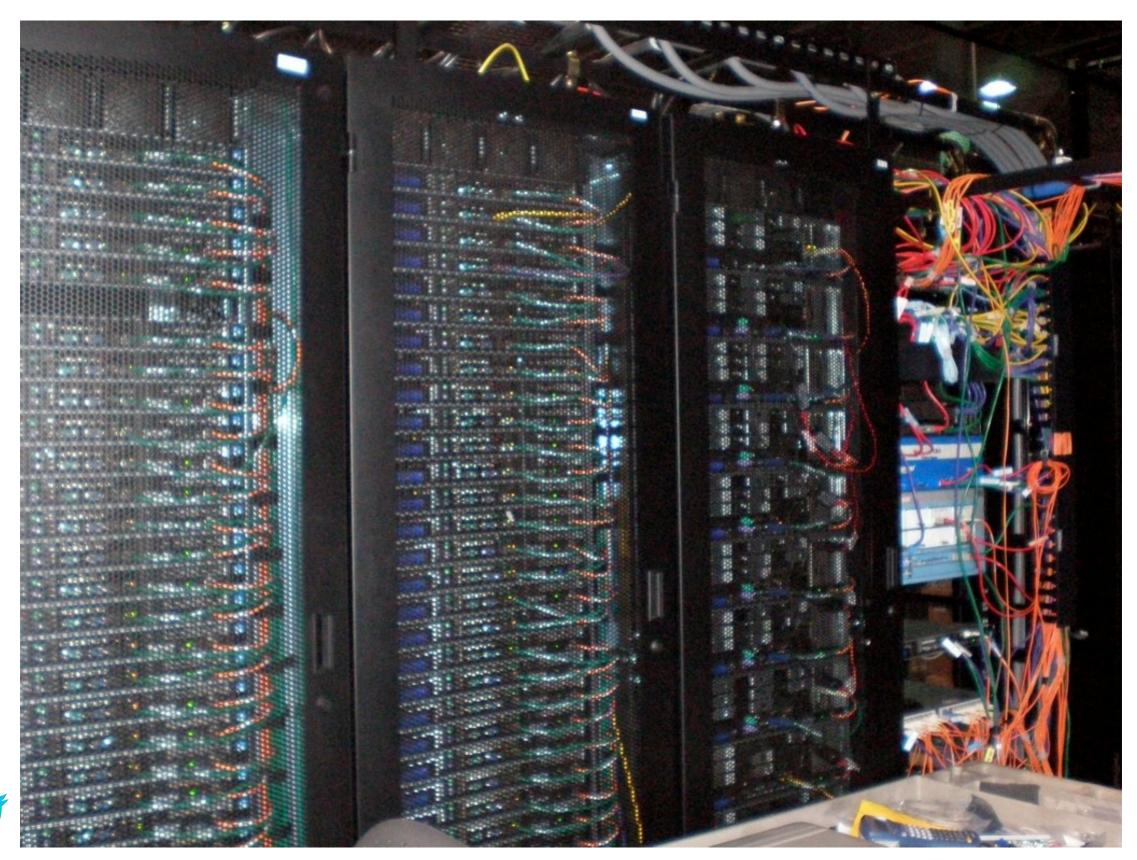
Combined: the pulse of the world



What we do with large data

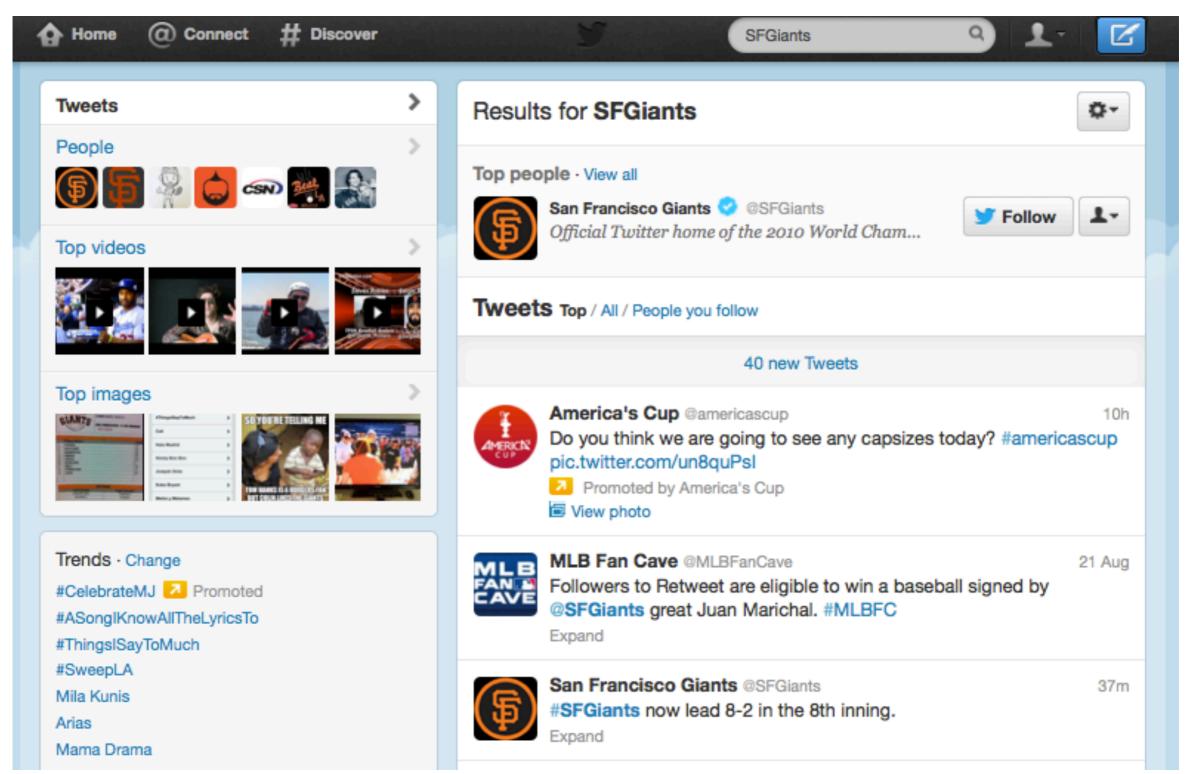


Scale



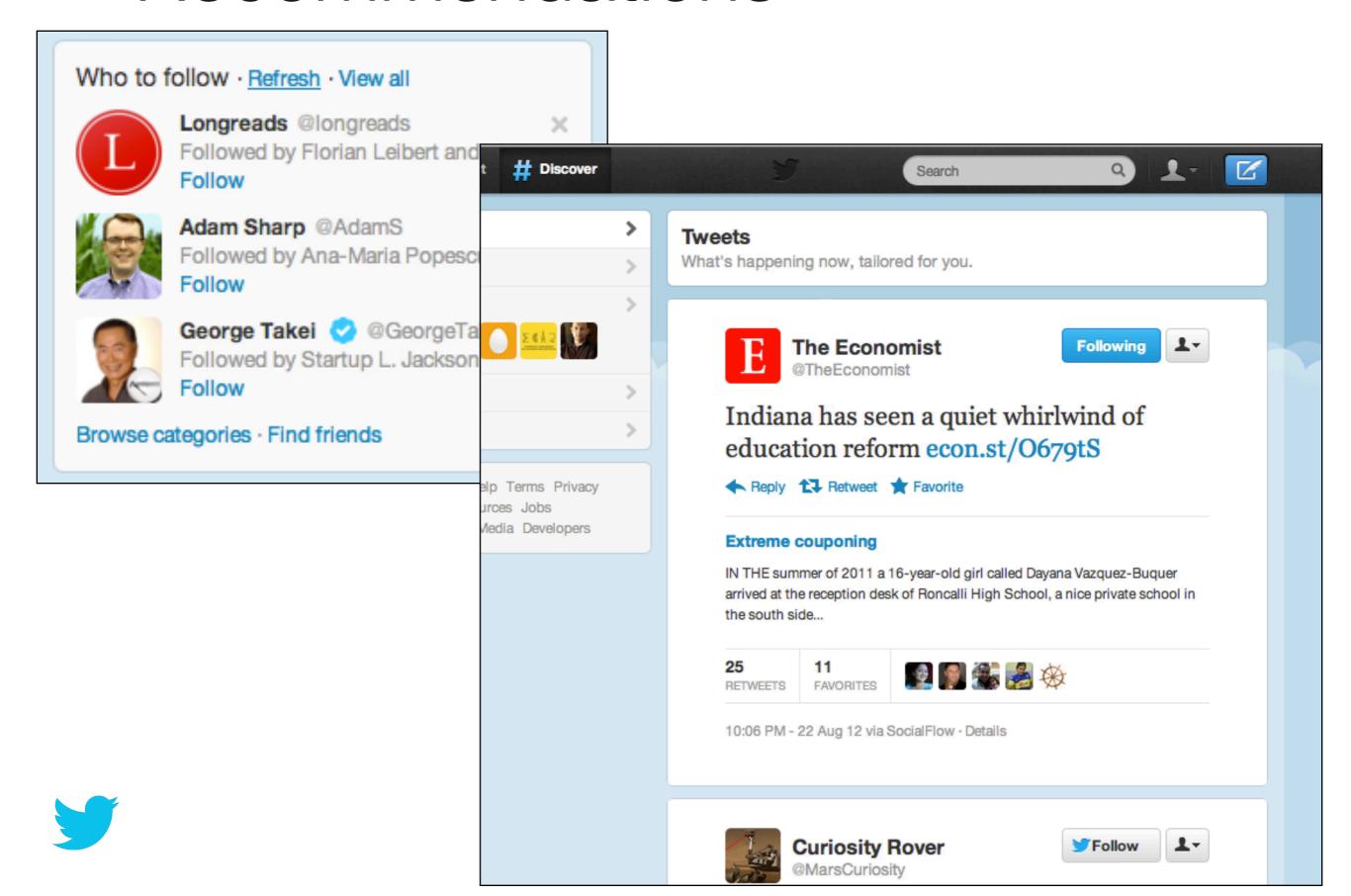


Search





Recommendations



Ads



perspective. http://globalizeyourthinking.com

Promoted



Anti-Spam





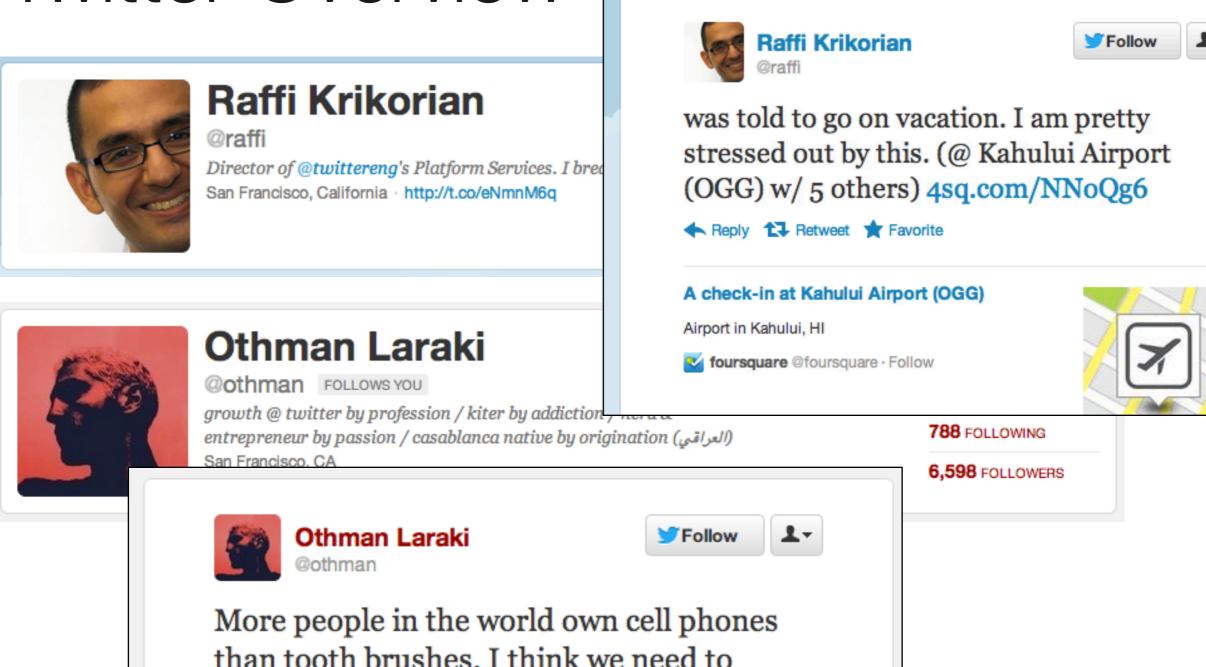
The Speakers

(Plus, some Twitter features)



Twitter Overview

Expanded Tweets





rethink Maslow's hierarchy of needs.

mindjumpers.com/blog/2012/01/s...

Reply Retweet * Favorite

Hadoop/Pig



Trends/Streaming





Search





Retweets, Favorites

Graphs, Recommendations, Relevance





Security, anomaly detection



Kurt Thomas

@inwyrd 🗎

http://t.co/LepOqQn



Delip Rao

@deliprao

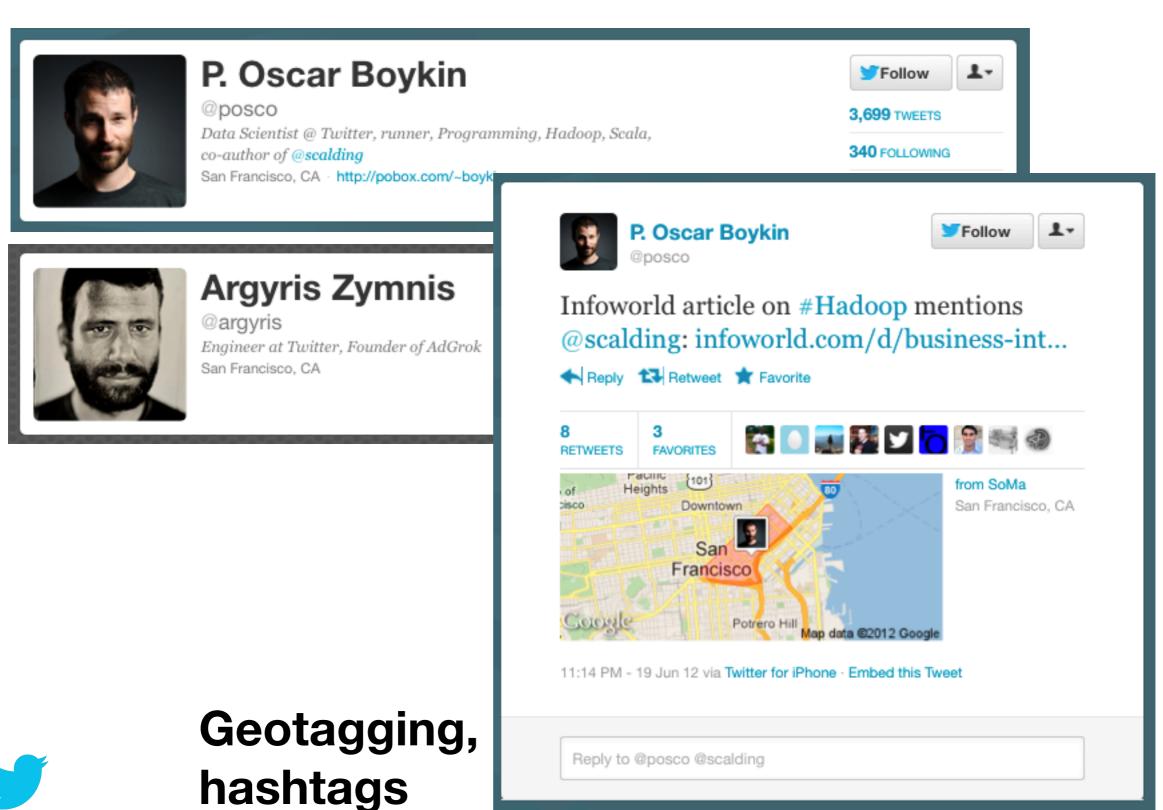
Twitter Number Crunching. Interests: Language Processing, Machine Learning Ultras, Marathons. My tweets are my ou San Francisco http://www.cs.jhu.edu/~del







Scalding





Goals

- Work with real data, on real problems
- Learn how it is to work in a place like Twitter
- Build something useful
- Have a good time!



Questions?

Follow me: @gilad

