FINDING AND ORGANIZING TWEETS

Analyzing Big Data With Twitter
Marti Hearst
November 27, 2012
WHY ARE THREE PROJECTS ABOUT ORGANIZING TWEETS?
TWITTER CATEGORIES AND LISTS
Browse categories
Select topics you’re interested in. Follow people you want to hear from.

Search for a topic, full name, or @username

Music 106 suggestions

Sports 78 suggestions

Entertainment 79 suggestions

Twitter 52 suggestions
# The 100 most popular Twitter lists

Ranked by followers count and refreshed daily.

<table>
<thead>
<tr>
<th>Rank</th>
<th>List Name</th>
<th>Followers</th>
<th>Following</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Team by Twitter</td>
<td>77,654</td>
<td>909</td>
</tr>
<tr>
<td>2</td>
<td>justinfollowplease by Justin Bieber Army</td>
<td>55,259</td>
<td>98</td>
</tr>
<tr>
<td>3</td>
<td>My Team by Justin Bieber</td>
<td>25,402</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>BlackBerry on Twitter by Research In Motion</td>
<td>23,634</td>
<td>33</td>
</tr>
<tr>
<td>5</td>
<td>World Leaders by Verified Accounts</td>
<td>15,529</td>
<td>38</td>
</tr>
</tbody>
</table>
Climate People & Lists

TOP PEOPLE TAGGED CLIMATE

tamron hall
Tamron Hall anchors MSNBC weekdays at 2pm ET. She also fills-in as a newscaster on Today and Weekend Today.
62,318 FOLLOWERS | FOLLOW

NASA Climate
Rocket science isn't enough; we're climate scientists, too. Follow the latest in climate change at http://t.co/ckatmyrf and on Facebook at http://t.co/mc5ifJ37.
49,250 FOLLOWERS | FOLLOW

Climate Reality
Reality. It's not an opinion.
99,213 FOLLOWERS | FOLLOW

Earthjustice
The earth needs a good lawyer
22,900 FOLLOWERS | FOLLOW

TRENDING IN CLIMATE

Links
Behind 'Rising India' lies the surrender of national dignity | Pankaj Mishra | Comment is free | The Guardian [guardian.co.uk]
Yahoo!
[ldd.yahooapis.com]
Live webcast | CBA Conference 2011 [oneworldgroup.org]
Sen. Stabenow jumps on climate denial train | Grist [grist.org]
Are Nuclear Plants Safe? Environmentalists Are Split : NPR [npr.org]
T - 3: More Ready Than Ready - Roz Savage, Ocean Rower [roz savage.com]
GreenBizDaily [feeds.feedburner.com]

People
@greenpeace @TheGreenHome @Katops

Top Tags
activism        eco        news
activists       environment progressive
art            green       socialjustice
artists        humanrights socialmedia
business       indigenous sustainable
climate        music       writers
children       media

More top tags →

The Listorious 140

The Top 140 Lists on Twitter →
The Top 140 Most Listed People on Twitter →
The 140 Oldest Accounts on Twitter →
The 140 Most Followed People on Twitter →

Email Updates
Subscribe to the latest lists on Twitter and climate...
Introducing the Ability to Follow Thought Leaders on LinkedIn

Ryan Roslansky, October 2, 2012

Everyday we are focused on helping our members to be great at what they do and today marks another exciting step towards making this possible for millions of professionals. For some time, you’ve been able to follow news by industry and sources, companies, and groups — these updates have seamlessly become part of the discussions you’re having everyday on LinkedIn with your peers. And now, you can follow other professionals on LinkedIn.

Not just any professionals, but 150 of the most influential thought leaders on LinkedIn who will be sharing unique knowledge and professional insights. Starting today, you can follow the likes of Richard Branson, Tony Robbins, Caterina Fake, Craig Newmark, President Barack Obama, Governor Mitt Romney, and many more. Read what they are saying, like and comment directly on their posts, and share with your network.
When we looked at k-core friend graphs of well-known users, they tended to produce sets of people within a topic area.

- LBNLcs, USGS, LosAlamosNatLab, arstechnica, NatGeo, nytimes, TheScienceGuy, Discovery, alexismadrigal
- RichardDawkins, SamHarrisOrg, robinince, michaelshermer, pzmyers, pennjillette
- bretlee, juniorbachchan, uberdoc, TheFarahKhan, taran_adarsh,
OBSERVATIONS ON TWITTER FOLLOWING

- Classify users into “elite” and “ordinary”
- Find that although audience attention is highly concentrated on a minority of elite users, much of the information they produce reaches the masses indirectly via a large population of intermediaries.
- Find support for the two-step flow of communications, where the critical intermediate layer is occupied by a category of media-savvy individuals called opinion leaders.
- Opinion leaders exist at every level of society (online as well as off).

- Find support for the two-step flow of communications, where the critical intermediate layer is occupied by a category of media-savvy individuals called opinion leaders.

- Found the population comprises two types:
  - those who receive essentially all of their media-originating information via two-step flows
  - and those who receive virtually all of it directly from the media.

- In total, the population of intermediaries is smaller than that of the users who rely on them, 99% are classified as ordinary users.

“Elite” users account for a substantial portion of attention on twitter, but it is allocated differently to different groups.

There is mainly homophily, except organizations pay more attention to bloggers.

<table>
<thead>
<tr>
<th>Table 3: Top 5 users in each category</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Celebrity</strong></td>
</tr>
<tr>
<td>aplusk</td>
</tr>
<tr>
<td>ladygaga</td>
</tr>
<tr>
<td>TheEllenShow</td>
</tr>
<tr>
<td>taylorswift13</td>
</tr>
<tr>
<td>Oprah</td>
</tr>
</tbody>
</table>
Figure 3: Share of tweets received among elite categories
OBSERVATIONS ON TWITTER FOLLOWING

Figure 4: RT behavior among elite categories
WHAT SHOULD TWITTER SEARCH LOOK LIKE?
TOPIC-BASED BROWSING OF TWITTER FEEDS

Eddi: interactive topic-based browsing of social status streams,
Bernstein et al., UIST 2010, NY NY

TOPIC-BASED BROWSING OF TWITTER FEEDS

Eddi: interactive topic-based browsing of social status streams
Bernstein et al., UIST 2010, NY NY
Table I. An example set of categories and tweets in those categories, as classified by TweeTopic.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple</td>
<td>W00t! Snow Leopard gave me 10 gigs back!</td>
</tr>
<tr>
<td></td>
<td>RT @username: gmail is down, but the imap connection on my iphone still works (fingers crossed!)</td>
</tr>
<tr>
<td></td>
<td>My iPhone 3GS cracked-on-a-rock, @username’s swam in a toilet, both repaired/replaced in 20 min @ Boylston Apple Store. Total cost: $0.</td>
</tr>
<tr>
<td>Obama</td>
<td>I think the most striking thing about Obama’s speech + GOP response for casual listeners would be how much agreement there was.</td>
</tr>
<tr>
<td></td>
<td>RT @username: The fastest way to prove you are an idiot is to call the President a liar on live TV</td>
</tr>
<tr>
<td>Research</td>
<td>@username Congratulations on the CSCW best paper nomination!</td>
</tr>
<tr>
<td></td>
<td>CORRECTION: the deadline for submissions to the Graduate Student Consortium for TEI ’09 is October 2</td>
</tr>
</tbody>
</table>
TWEETOPIC ALGORITHM

- Idea: make a tweet into a search engine query
  - Given a tweet, output ranked list of terms
    - noun phrases (not needed); remove twitter markup, stopwords
  - Query a twitter search engine (Y!BOSS)
  - Identify popular terms in the results (using TF.IDF weighting); use these to classify the tweet
  - Show the most popular terms in the interface; assign tweets to terms
  - Promote topics to the top that match what the user tweets about.
- Greatly outperformed LDA in classifying topic of tweet.
USABILITY STUDY

- Compared to a control vertical column view in reverse chron. order.
- Within-participants design (different 24 hour period); 3 trials of each.
- Task: browse your feed, find out as much of what happened in the 24 hour time period as possible in 3 minutes time.
USABILITY STUDY

- Subjective Results:
  - Eddi more efficient, enjoyable, easier to browse, less overwhelming
  - Eddi felt less comprehensive in coverage
  - People viewed half as many tweets with Eddi but felt more efficient
USABILITY STUDY: SUBJECTIVE RESULTS

Figure 6. Eddi fared better on all metrics except confidence in seeing everything desired in the feed.
Objective Results:

- Experimenters recorded which tweets the participants looked at when using each interface for the 3 minute tests.

- After the fact, they had them rate 300 tweets for interestingness.
  
  - 150 from using the control, 150 from using Eddi
  
  - Eddi: ~1 in 4 tweets seen later rated as interesting
  
  - Control: ~1 in 7

- (Recall was about the same in both)
NEXT CLASS

- Course Wrap-up
- Course Evaluations