

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.

Music 106 suggestions



Sports 78 suggestions



Entertainment 79 suggestions



Twitter 52 suggestions



Twitter Counter tracks statistics of more than 67 million Twitter users

I am on Twitter, [show me](#) my Twitter stats

Show Twitter users

Search

Featured Twitter Users



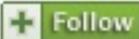
JenniferBeinke

Officially signed a new warehouse lease for KMM! We will now be located in both Santa Barbara and San Diego! Dreaming of NY soon?



WJeffreyBrown

My daily stats: 1 new followers, 3 new unfollowers via <http://t.co/ZdwHrpCx>



AffordableSvcs

Nutrition for Everyone: How Many Fruits and Vegetables Do You Need? Via @CDCgov <http://t.co/RmNyCJez> #health



ArianaGrande

#NewVideo! The Question Game by Liz Gillies & Ariana Grande via @YouTube: <http://t.co/azdglenT> @LizGillies Please enjoy & share RT



Get Twitter Followers



The 100 most popular Twitter lists

Ranked by followers count and refreshed daily.

1 Team by Twitter



77,654
followers

909
following

2 justinfollowplease by Justin Bieber Army



Beliebers that would like to be followed by Justin.

55,259
followers

98
following

3 My Team by Justin Bieber

#WelcomeToTwitter



25,402
followers

12
following

4 BlackBerry on Twitter by Research In Motion



A complete list of all Research In Motion Twitter accounts.

23,634
followers

33
following

5 World Leaders by Verified Accounts



15,529
followers

38
following

Listorious

Search over 2 million top Twitter users

Search

activism charity entertainment media progressive technology
art children environment music science travel
business climate food news socialmedia twitter
celebrities education health politics sports writers

[Tags](#) [Add Yourself](#) [Add a Twitter List](#) [Inbox](#) [Sign in with Twitter](#) [About](#)

A Sawhorse Media Production. [Terms](#). [Privacy](#).

Climate People & Lists

[Tweet](#) 5

TOP PEOPLE TAGGED CLIMATE

[tamron hall](#)

Tamron Hall anchors MSNBC weekdays at 2pm ET. She also fills-in as a newsreader 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/ckatmryf> 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!](#)

[lrd.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](#)

[rozsavage.com]

[GreenBizDaily](#)

[feeds.feedburner.com]

People

[@greenpeace](#)[@TheGreenHome](#)[@Katy...](#)

Top Tags

activism	eco	news
activists	environment	progressive
art	green	socialjustice
artists	humanrights	socialmedia
arts	indigenous	sustainable
business	liberal	twitter
children	media	women
climate	music	writers

[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



Introducing the Ability to Follow Thought Leaders on LinkedIn

Ryan Roslansky, October 2, 2012

1,807

Share

658

Tweet

599

Like

103

+1

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.

The screenshot shows the LinkedIn homepage for user Melinda Ashcroft. At the top, there's a navigation bar with 'Home', 'Profile', 'Contacts', 'Groups', 'Jobs', 'Inbox', 'Companies', 'News', and 'More'. A search bar is on the right. Below the navigation, a banner for 'Top Producing Realtor - Over \$48M Sold in 2011. Ask me about off market listings.' is visible. The main content area features a post from Richard Branson, Founder at Virgin Group, titled 'Five top tips to starting a successful business'. The post includes a photo of Branson and a quote: 'We have two ears and one mouth, using them in proportion is not a bad idea! To be a good leader you have to be a great listener. Brilliant ideas can spring...'. Below the post is a 'Follow Richard' button and a 'See more people to follow' link. On the right side, there's a 'PEOPLE YOU MAY KNOW' sidebar with three suggestions: Sandra Kelley (Software Engineer at Danberley Corp), Monica Reynolds (Director, Product Management at Claris), and Brad Newman (Senior Engineer at Ariba Corp). Each suggestion includes a profile picture, name, title, company, and a 'Connect' button.

OBSERVATION FROM ASSIGNMENT 3

- 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

- “Who says what to whom on twitter,” S Wu, JM Hofman, WA Mason, DJ Watts - Proceedings of the WWW 2011.
 - 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).

OBSERVATIONS ON TWITTER FOLLOWING

- “Who says what to whom on twitter,” S Wu, JM Hofman, WA Mason, DJ Watts - Proceedings of the WWW 2011.
- 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.

OBSERVATIONS ON TWITTER FOLLOWING

- “Who says what to whom on twitter,” S Wu, JM Hofman, WA Mason, DJ Watts - Proceedings of the WWW 2011.
- “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 3: Top 5 users in each category

<i>Celebrity</i>	<i>Media</i>	<i>Org</i>	<i>Blog</i>
aplusk ladygaga TheEllenShow taylorswift13 Oprah	cnnbrk nytimes asahi BreakingNews TIME	google Starbucks twitter joinred ollehkt	mashable prologger kibeloco naosalvo dooce

OBSERVATIONS ON TWITTER FOLLOWING

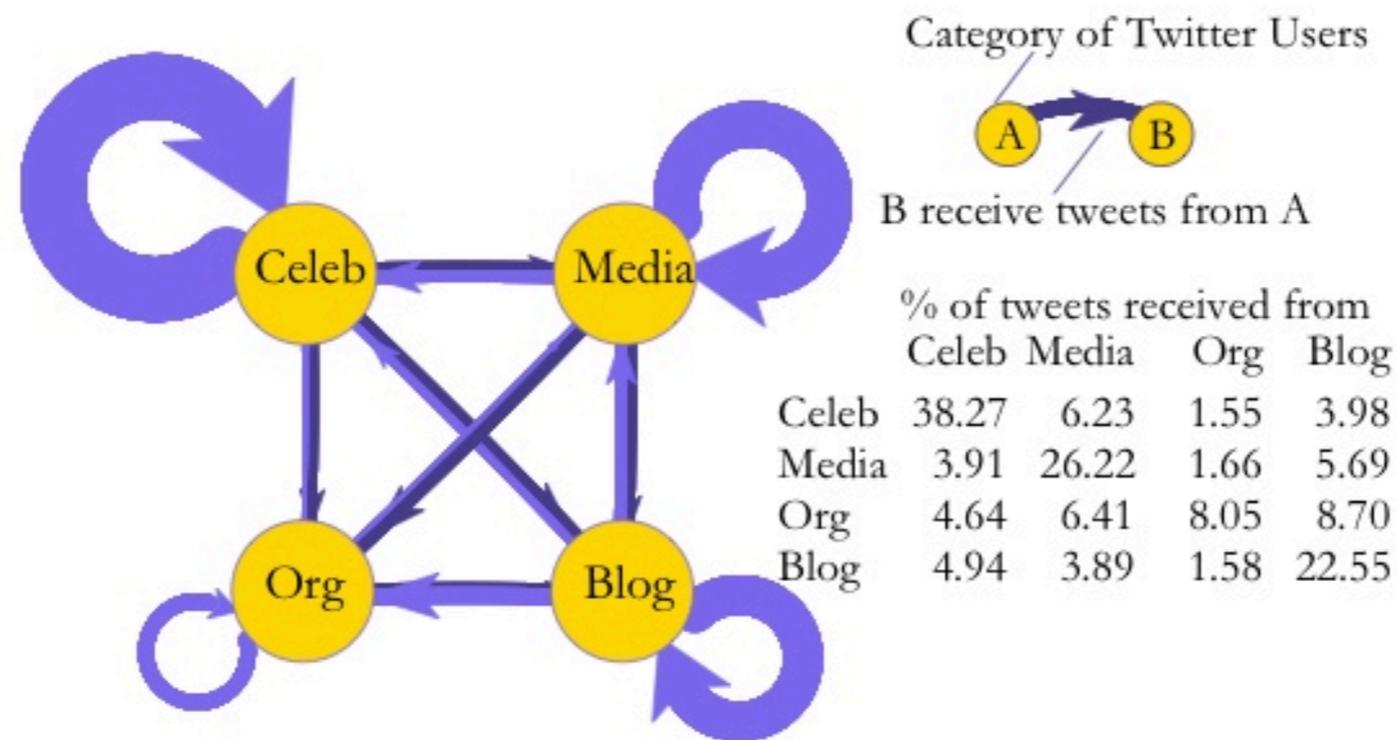


Figure 3: Share of tweets received among elite categories

OBSERVATIONS ON TWITTER FOLLOWING

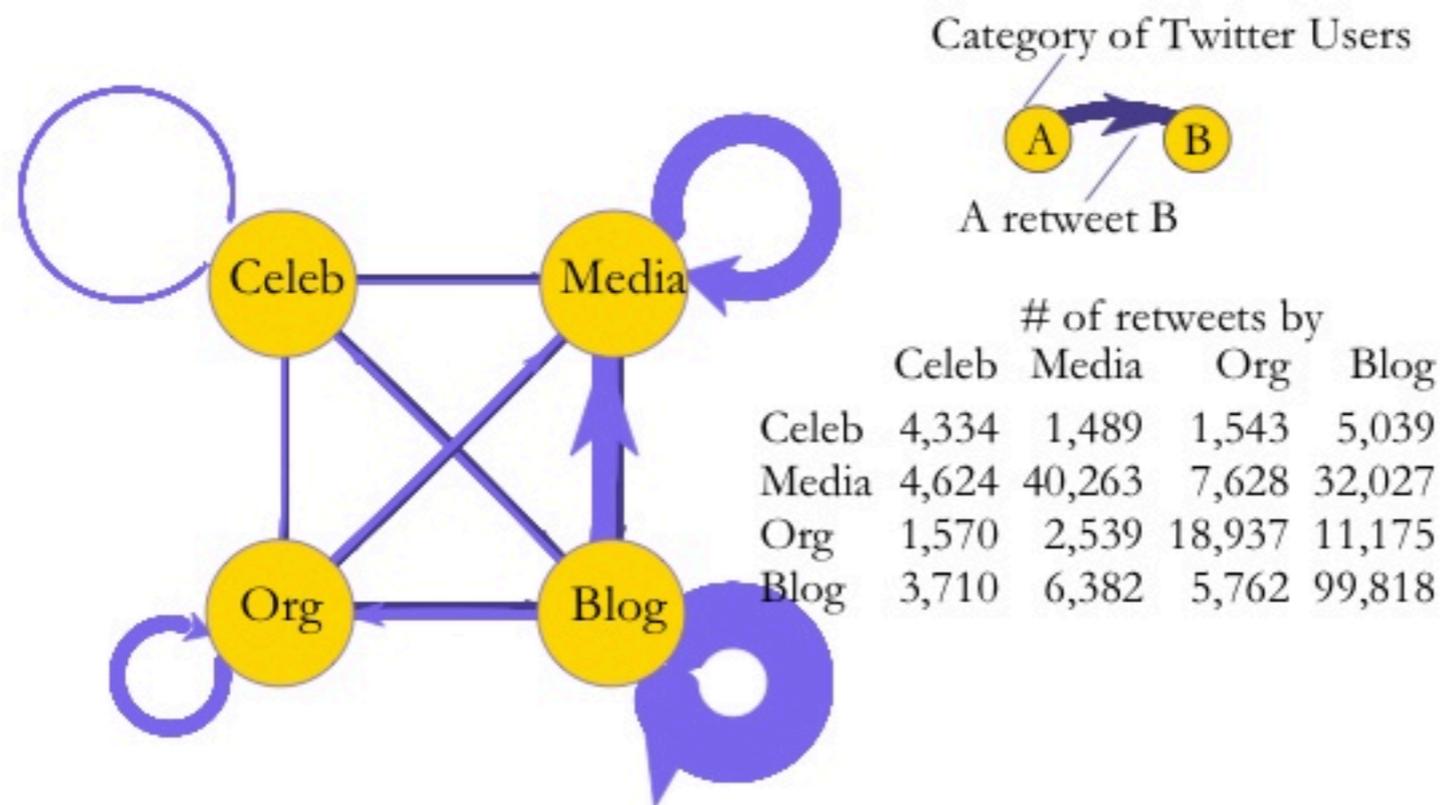


Figure 4: RT behavior among elite categories

WHAT SHOULD TWITTER
SEARCH LOOK LIKE?

TOPIC-BASED BROWSING OF TWITTER FEEDS

The screenshot displays the 'eddi' application interface. At the top left is the 'eddi' logo. Below it, there is a search bar and a date selector set to '09/09/2009'. A search results list on the left side shows various categories and their counts, such as '77 twitter', '24 video', '21 facebook', etc. The main area features a 'Tag Cloud' and a 'Timeline' tab. The tag cloud contains numerous words of varying sizes, with 'facebook', 'twitter', 'google', 'music', 'video', and 'social media' being prominent. Below the tag cloud, there are sections for 'Tweets you might be interested in:', 'BOSTON (3)', 'BIKE (4)', 'TWITTER (77)', and 'TEA (2)', each containing a list of tweets with user avatars and text.

<http://people.csail.mit.edu/msbernst/videos/eddi-uist2010.mp4>

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

TOPIC-BASED BROWSING OF TWITTER FEEDS



The image shows a screenshot of a Twitter feed filtered by the topic "google". The header of the feed is a dark blue bar with the word "google" in white lowercase letters. Below the header, there are four tweets, each with a profile picture, a name, a timestamp, and the text of the tweet.

- patelpatel** (1 week, 3 days ago): Using the XMPP service - Google App Engine. <http://bit.ly/jUi5M>
- namevoyager** (1 week, 3 days ago): New milestone in online journalism: Google News just put an Onion article on its front page.
- trebaknow** (1 week, 4 days ago): email, it consumes my life. i have gmail on intravenous now, so i actually almost died during tuesday's outage.
- smackipoo** (1 week, 4 days ago): Android phone users - please download "implicit authorization" app for study (<http://bit.ly/lCnei>) to win \$150 gift certificate

Below the tweets, there is a section titled "Possibly related to google:" with three entries, each featuring a star icon, a profile picture, and text:

- HCIR_JohnD RT: @nahumg: How to Restrict Google Results to the Past Minutes or Seconds <http://post.ly/4sp>
- fitzroy @disqus Having trouble? I'm getting a 503 error on my blog when I try to reply. <http://bit.ly/ilzcp>
- HCIR_GeneG Thanks! This seems like an interesting (if hard) topic for HCIR research

TOPIC-BASED BROWSING OF TWITTER FEEDS

Apple



W00t! Snow Leopard gave me 10 gigs back!



RT @username: gmail is down, but the imap connection on my iphone still works (fingers crossed!)



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.

Obama



I think the most striking thing about Obama's speech + GOP response for casual listeners would be how much agreement there was.



RT @username: The fastest way to prove you are an idiot is to call the President a liar on live TV

Research



@username Congratulations on the CSCW best paper nomination!



Stanford scientists turn liposuction leftovers into embryonic-like stem cells: <http://bit.ly/3GHsw9>



CORRECTION: the deadline for submissions to the Graduate Student Consortium for TEI '09 is October 2
<http://bit.ly/15D8Mv>

Table I. An example set of categories and tweets in those categories, as classified by TweepTopic.

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.

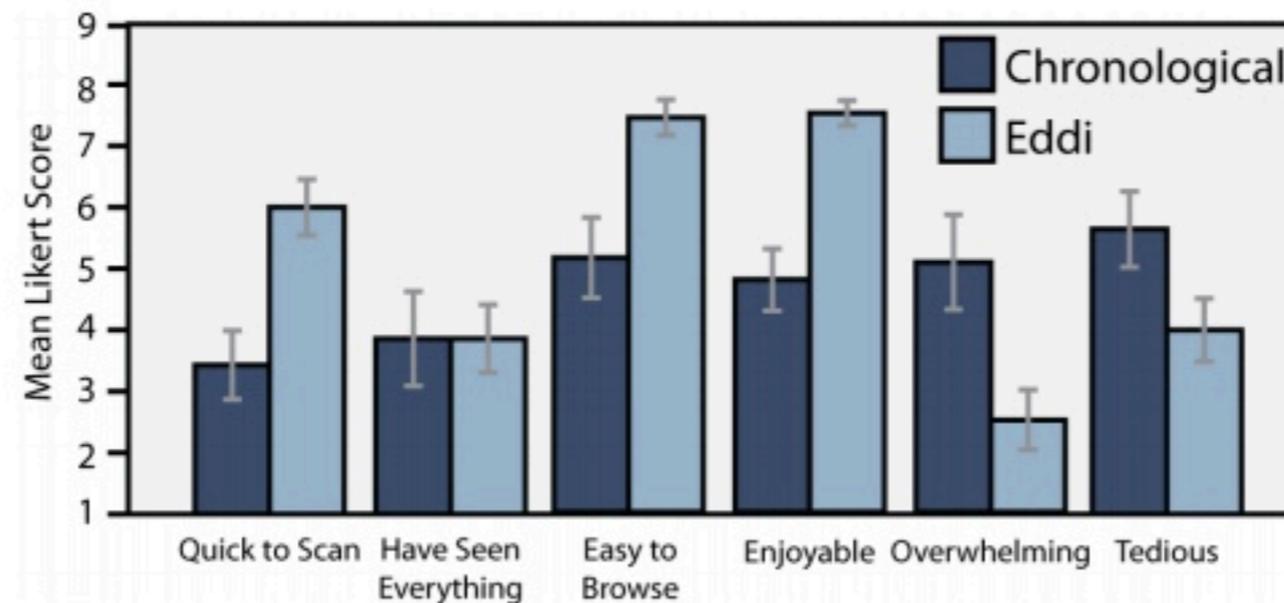
USABILITY STUDY

- 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



Metric	μ (Eddi)	μ (Chronological)	p-value
Quick to Scan	6.0	3.4	< .01
Have Seen Everything	3.9	3.9	.97
Easy to Browse	7.4	5.1	< .01
Enjoyable	7.5	4.8	< .01
Overwhelming	2.5	5.1	< .05
Tedious	4.0	5.6	< .05

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

USABILITY STUDY

- 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