WHAT IS TWITTER?
TWITTER AS CULTURAL PHENOMENON

Ignoring Deadline to Quit, G.O.P. Senate Candidate Defies His Party Leaders
BY JONATHAN WEISMAN
Representative Todd Akin said definitively Tuesday that he would not step aside. After his comments on rape, fellow Republicans, including Mitt Romney, asked him to drop out of the Missouri Senate race.

Missouri Controversy May Endanger Republican Chances in the Fall
BY JENNIFER STEINHAUER
Representative Todd Akin’s remarks on rape have focused attention on the party’s agenda on restricting abortion rights, a politically volatile topic for Mitt Romney and other candidates.

In Toll of 2,000, New Portrait of Afghan War
BY JAMES DAO and ANDREW W. LEHREN
As the war in Afghanistan wears on, an analysis of troop casualties reflects the effect of the 2010 troop surge and points to the growing danger from an ostensible ally.

Dispute Over Islands Reflects Japanese Fear of China’s Rise
BY MARTIN FACKLER
A chain of islands at the center of a territorial dispute have Japanese nationalists pushing their country to boldly counter the rising power of China and North Korea.

The New Rise of a Summer Hit: Tweet It Maybe
BY BEN SISARIO
This summer’s pop hit, Carly Rae Jepsen’s cheerfully flirtatious “Call Me Maybe,” shows how much the hitmaking machine, and the music industry, has been upended by social media.
Findings

Some 15% of online adults use Twitter as of February 2012, and 8% do so on a typical day. Although overall Twitter usage has nearly doubled since the Pew Research Center’s Internet & American Life Project first asked a standalone Twitter question in November 2010, the 15% of online adults who use Twitter as of early 2012 is similar to the 13% of such adults who did so in May 2011. At the same time, the proportion of online adults who use Twitter on a typical day has doubled since May 2011 and has quadrupled since late 2010—at that point just 2% of online adults used Twitter on a typical day. The rise of smartphones might account for some of the uptick in usage because smartphone users are particularly likely to be using Twitter.

Twitter usage over time

% of internet users who use Twitter

<table>
<thead>
<tr>
<th>Month</th>
<th>Typical day</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>November 2010</td>
<td>2%</td>
<td>8%</td>
</tr>
<tr>
<td>May 2011</td>
<td>4%</td>
<td>13%</td>
</tr>
<tr>
<td>August 2011</td>
<td>5%</td>
<td>12%</td>
</tr>
<tr>
<td>February 2012</td>
<td>8%</td>
<td>15%</td>
</tr>
</tbody>
</table>

Source: Pew Research Center’s Internet & American Life Project Winter 2012 Tracking Survey, January 20-February 19, 2012. N=2,253 adults age 18 and older, including 901 cell phone interviews. Interviews conducted in English and Spanish. Margin of error is +/-2.7 percentage points for internet users (n=1,729).
## Who uses Twitter?

% of internet users within each group who use Twitter

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All adult internet users (n=1729)</strong></td>
<td>15%</td>
</tr>
<tr>
<td>Men (n=804)</td>
<td>14</td>
</tr>
<tr>
<td>Women (n=925)</td>
<td>15</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
</tr>
<tr>
<td>18-29 (n=316)</td>
<td>26**</td>
</tr>
<tr>
<td>30-49 (n=532)</td>
<td>14</td>
</tr>
<tr>
<td>50-64 (n=521)</td>
<td>9</td>
</tr>
<tr>
<td>65+ (n=320)</td>
<td>4</td>
</tr>
<tr>
<td><strong>Race/ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>White, Non-Hispanic (n=1229)</td>
<td>12</td>
</tr>
<tr>
<td>Black, Non-Hispanic (n=172)</td>
<td>28**</td>
</tr>
<tr>
<td>Hispanic (n=184)</td>
<td>14</td>
</tr>
<tr>
<td><strong>Annual household income</strong></td>
<td></td>
</tr>
<tr>
<td>Less than $30,000/yr (n=390)</td>
<td>19</td>
</tr>
<tr>
<td>$30,000-$49,999 (n=290)</td>
<td>12</td>
</tr>
<tr>
<td>$50,000-$74,999 (n=250)</td>
<td>14</td>
</tr>
<tr>
<td>$75,000+ (n=523)</td>
<td>17</td>
</tr>
<tr>
<td><strong>Education level</strong></td>
<td></td>
</tr>
<tr>
<td>No high school diploma (n=108)</td>
<td>22</td>
</tr>
<tr>
<td>High school grad (n=465)</td>
<td>12</td>
</tr>
<tr>
<td>Some College (n=447)</td>
<td>14</td>
</tr>
<tr>
<td>College + (n=698)</td>
<td>17</td>
</tr>
<tr>
<td><strong>Geographic location</strong></td>
<td></td>
</tr>
<tr>
<td>Urban (n=520)</td>
<td>19**</td>
</tr>
<tr>
<td>Suburban (n=842)</td>
<td>14**</td>
</tr>
<tr>
<td>Rural (n=280)</td>
<td>8</td>
</tr>
</tbody>
</table>

**Source:** Pew Research Center’s Internet & American Life Project Winter 2012 Tracking Survey, January 20-February 19, 2012. N=2,253 adults age 18 and older, including 901 cell phone interviews. Interviews conducted in English and Spanish. The margin of error is +/−2.7 percentage points for internet users. **Represents significant difference compared with all other rows in group.**
Twitter use among 18-24 year olds increased dramatically between May 2011 and February 2012, both overall and on a "typical day" basis. Twitter use within the overall population remained steady over the last year, and usage rates within most major demographic groups changed little over the same time period. The youngest adults (those between the ages of 18 and 24) are the primary exception—nearly one third of internet users in this age group now use Twitter, up from 18% in May of 2011 and 16% in late 2010.

Twitter use by those in their mid-20s to mid-40s largely leveled off in the last year after roughly doubling between late 2010 and mid 2011.

Twitter adoption by age, 2010-2012
% of internet users in each group who use Twitter

<table>
<thead>
<tr>
<th>Age Group</th>
<th>November 2010</th>
<th>May 2011</th>
<th>February 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Adults</td>
<td>8%</td>
<td>13%</td>
<td>15%</td>
</tr>
<tr>
<td>18-24</td>
<td>16</td>
<td>18</td>
<td>31</td>
</tr>
<tr>
<td>25-34</td>
<td>9</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td>35-44</td>
<td>8</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>45-54</td>
<td>7</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>55-64</td>
<td>4</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>65+</td>
<td>4</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

Sources: Pew Research Center’s Internet & American Life Project tracking surveys. 2012 data based on January 20-February 19, 2012 Tracking Survey. N=2,253 adults age 18 and older, including 901 cell phone interviews, margin of error is +/-2.7 percentage points based on internet users (n=1729).
Evan Williams on the genesis of Twitter, ICWSM, April 2007:

- A side project started on a whim; Jack Dorsey’s idea; launched Oct, 2006.
- Wanted a ubiquitous status message.
- A community of people answering the question “what are you doing?”
- Exploded at SXSW; good for collective backchanneling; SF earthquakes.
- “Ambient intimacy”
- Huge API usage was unexpected, as was the rise of the @ sign for replies

TWITTER: A PLATFORM FOR RESEARCH
TWITTER:
AN INSPIRATION FOR RESEARCH
TWITTER: AN INSPIRATION FOR RESEARCH

**Publications (1215)**

*Why we twitter: understanding microblogging usage and communities* (Citations: 162)

Akshay Java, Xiaodan Song, Tim Finin, Belle Tseng

...phones, email or the web. **Twitter**, a popular microblogging tool has...topological and geographical properties of **twitter**'s social network. We find...

Published in 2007.

*A few chirps about **twitter*** (Citations: 100)

Balachander Krishnamurthy, Phillipa Gill, Martin Arlitt

...smaller network applications such as **twitter** and dodgeball. We present a detailed characterization of **twitter**, an application that allows users...including constrained crawls of the **twitter** network using two different methodologies, and a
TWITTER: AN INSPIRATION FOR RESEARCH

- **What is Twitter?** (Kwan et al, “What is Twitter; a social network or a news media?”, WWW, 2009)

- **Who uses Twitter?** (Mislov et al., “Understanding the Demographics of Twitter Users,” IWCSM, 2011)

TWITTER: AN INSPIRATION FOR RESEARCH

- **Who is influential?** (Weng et al., “Twitterrank: finding topic-sensitive influential twitterers” WSDM, 2010)

- **Usage in large-scale emergencies** (Krishnamurthy et al., “A few chirps about Twitter,” ACM workshop on Online Social Networks, 2008)

- **Reciprocal social relationships** (Huberman et al., “Social Networks that Matter; Twitter under the Microscope,” First Monday 14(1-5), 2009)
TWITTER: AN INSPIRATION FOR RESEARCH

• Microblogging at work (Zhao & Rosson, “How and why people Twitter; the role of that micro-blogging plays in informal communication at work,” ACM Group, 2009)

• Use of the @ sign (Honey & Herring, “Beyond Microblogging: conversation and collaboration via Twitter,” HICSS 2009)

• Monitoring disease outbreaks (Chew et al., “Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak,” PloS One, 5(11), 2010)
TWITTER BIG DATA CASE STUDY: MOOD SWINGS

Golder & Macy, "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures", Science, 30 September 2011: Vol. 333 no. 6051 pp. 1878-1881

• Goal: obtain data about people’s moods throughout the day and across the globe to see if and how they correspond to time of day.

• This is in contrast to self-report:
  • not prompted by an experimenter
  • not reported after the fact
  • not a small sample of undergraduates.
TWITTER BIG DATA CASE STUDY: MOOD SWINGS

• Background:

  • Individual mood is an affective state, influenced by:
    • Neurochemicals and hormones, and
    • Social activity including daily routines
  • Positive and negative affect are independent dimensions
    • positive (PA): enthusiasm, delight, alertness
    • negative (NA): distress, fear, anger, guilt
  • Research suggests that low PA indicates the absence of positive feelings, not the presence of negative ones.

TWITTER BIG DATA CASE STUDY: MOOD SWINGS

• 509 M posts
• 2.4 M users’ posts
• >25 and < 400 posts/user
• Feb 2008 - Jan 2010
• International, but English only

TWITTER BIG DATA CASE STUDY: MOOD SWINGS

• Text of tweets was analyzed for positive and negative affect

• Used LIWC, a common content analysis tool

Golder & Macy, "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures", Science, 30 September 2011: Vol. 333 no. 6051 pp. 1878-1881
Golder & Macy, "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures", Science, 30 September 2011: Vol. 333 no. 6051 pp. 1878-1881
Figure S3. Hourly changes in within-individual affect (y-axis) over a 24-hour cycle (x-axis), broken down by chronotype, for PA (top) and NA (bottom). Each series shows the mean affect (black lines).

Golder & Macy, "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures", Science, 30 September 2011: Vol. 333 no. 6051 pp. 1878-1881
DISCUSSION

• What are some of the assumptions behind this work?
• What are some of the computational challenges?
WHAT IS BIG DATA?

WHY IS IT CHALLENGING?
TWITTER BIG DATA CASE STUDY:
HOW DO PEOPLE USE TWITTER?

Java et al., "Why We Twitter: Understanding Microblogging Usage and Communities," Joint 9th WebKDD and 1st SNA-KDD Workshop '07, Aug 12, 2007, San Jose, CA.

• Goal: gain an understanding of Twitter usage, mainly through quantitative techniques.

• Approached the analysis in several interesting, varying ways.

• Questions to consider:
  • What are the data analysis problems?
  • This was a very early paper; how well do is results stand up?
TWITTER BIG DATA CASE STUDY:
HOW DO PEOPLE USE TWITTER?


• Data:
  • April 1, 2007 - May 30, 2007
  • 1.3M posts
  • 76K users
TWITTER BIG DATA CASE STUDY:
HOW DO PEOPLE USE TWITTER?


• Network Properties:

Figure 6: Scatter plot showing the degree correlation of Twitter social network. A high degree correlation signifies that users who are followed by many people also have large number of friends.

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>total nodes</td>
<td>88K</td>
</tr>
<tr>
<td>total links</td>
<td>830K</td>
</tr>
<tr>
<td>avg degree</td>
<td>19</td>
</tr>
<tr>
<td>degree corr.</td>
<td>0.6</td>
</tr>
<tr>
<td>reciprocity</td>
<td>0.6</td>
</tr>
</tbody>
</table>
**TWITTER BIG DATA CASE STUDY:**

**HOW DO PEOPLE USE TWITTER?**


- Geographic Distribution

![Distribution of Twitter Users Across the World](image)

**Figure 7:** Figure shows the global distribution of Twitter users. Though initially launched in US Twitter is popular across the world.

<table>
<thead>
<tr>
<th>From-to</th>
<th>Asia</th>
<th>Europe</th>
<th>Oceana</th>
<th>N.A</th>
<th>S.A</th>
<th>Africa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td>13.45</td>
<td>0.64</td>
<td>0.10</td>
<td>5.97</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>Europe</td>
<td>0.53</td>
<td>9.48</td>
<td>0.25</td>
<td>6.16</td>
<td>0.17</td>
<td>0.02</td>
</tr>
<tr>
<td>Oceana</td>
<td>0.13</td>
<td>0.40</td>
<td>0.60</td>
<td>1.92</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>N.A</td>
<td>5.19</td>
<td>5.46</td>
<td>1.23</td>
<td>45.60</td>
<td>0.60</td>
<td>0.10</td>
</tr>
<tr>
<td>S.A</td>
<td>0.06</td>
<td>0.26</td>
<td>0.02</td>
<td>0.75</td>
<td>0.62</td>
<td>0.00</td>
</tr>
<tr>
<td>Africa</td>
<td>0.01</td>
<td>0.03</td>
<td>0.00</td>
<td>0.11</td>
<td>0.00</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 3: Table shows the distribution of Twitter social network links across continents. Most of the social network lies within North America. (N.A = North America, S.A = South America)

<table>
<thead>
<tr>
<th>Property</th>
<th>N.A</th>
<th>Europe</th>
<th>Asia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Nodes</td>
<td>16,998</td>
<td>5201</td>
<td>4886</td>
</tr>
<tr>
<td>Total Edges</td>
<td>205,197</td>
<td>42,664</td>
<td>60519</td>
</tr>
<tr>
<td>Average Degree</td>
<td>24.15</td>
<td>16.42</td>
<td>24.77</td>
</tr>
<tr>
<td>Degree Correlation</td>
<td>0.62</td>
<td>0.78</td>
<td>0.92</td>
</tr>
<tr>
<td>Clustering Coefficient</td>
<td>0.147</td>
<td>0.54</td>
<td>0.18</td>
</tr>
<tr>
<td>Percent Reciprocity</td>
<td>62.64</td>
<td>71.62</td>
<td>81.40</td>
</tr>
</tbody>
</table>

Table 4: Network properties of social networks within a continent. Europe and Asia have a higher reciprocity indicating closer ties in these social networks. (N.A = North America)
**TWITTER BIG DATA CASE STUDY:**

**HOW DO PEOPLE USE TWITTER?**


• **User Intention**
  
  • First find “communities” based on link structure, using 2 algorithms:
    
    • Hubs & Authorities (Kleinberg ’99) to find top authorities
    
    • Clique Percolation Method (Palla et al ’05) to find groups
  
  • Key terms computed with log likelihood ratios based on when they appear and where in network they appear (Rayson & Garside ’00)

![Figure 8: An example of a “gaming” community who also share daily experiences.](image)
TWITTER BIG DATA CASE STUDY:
HOW DO PEOPLE USE TWITTER?

Java et al., "Why We Twitter: Understanding Microblogging Usage and Communities," Joint 9th WebKDD and 1st SNA-KDD Workshop '07, Aug 12, 2007, San Jose, CA.

Figure 10: Example Communities in Twitter Social Network. Key terms indicate that these communities are talking mostly about technology. The user Scobiezer connects multiple communities in the network.
TWITTER BIG DATA CASE STUDY:
HOW DO PEOPLE USE TWITTER?


- Final step: manual classification of messages (how many not stated)
- Main intentions:
  - Daily chatter
  - Conversations
  - Sharing information / URLs
- Main categories of users:
  - Information source (hub with many followers)
  - Friends
  - Information seeker
THIS COURSE

• Focusing on software methods for handling big data

• **Not** focusing on machine learning, **not** data mining, **not** natural language processing, although we will touch on these

• Fast-paced, self-directed learning

• 3 or 4 programming assignments

• Class project
CLASS PROJECT

• Project Mentor from Twitter

• Teams of 3 **REQUIRED**

• Presentations at both UCB and Twitter

• Wide range of topics:
  • Infrastructure, apps, analysis ...
Twitter Status Calendar by @alan_ritter
A demonstration of NLP on twitter showing the most prominent events mentioned in association with each day.

What is StatusCalendar?
Users of Social Networking sites frequently discuss events which will occur in the near future. By annotating Named Entities and resolving temporal expressions (for example "next Friday"), we are able to automatically extract a calendar of popular events occurring in the near future from Twitter.

Related Papers
Open Domain Event Extraction from Twitter
Named Entity Recognition in Tweets: An Experimental Study

Team
Alan Ritter, Mausam, Oren Etzioni, Sam Clark, Nikki Lee.

http://statuscalendar.cs.washington.edu/

August 2012

Mon | Tue | Wed | Thu | Fri | Sat | Sun
---|---|---|---|---|---|---
20 | 21 | 22 (today) | 23 | 24 | 25 | 26
school: starts, start, excited
demi: birthday, turning, tweet
evorton: debut, game, season
eid: celebrating, celebrate, celebrated
rvp: start, game, see more...
trey songz: album, comes out, come out
school: starts, excited, sleep
kiss the ring: album, in stores, #countdown
rob: cards, free, use
chelsea: reading, game, playing
open house: see, school, meet
more... | more... | more... | gcs: results, getting, nervous
leeds: fest, excited, sleep
waterloo road: starts, back on, excited
leeds: fest, excited, packed
good luck: results, getting, gcse results
barcelona: game, win, watch
more... | keith lemon: film, movie, see
knights of columbus: sounds by, party, come out
russell wilson: start, starting, plays at
tax: party, free, #thecrewsparty
north houston avenue: party, #thecrewsparty, admission
more... | loveuariana: birthday
roger clemens: start, pitch, signed
5503 richard: hosted, bash, bday
dj kali: hosted, bash, bday
visions: hosted, bash, free
more... | tshirt smtown: berapa, help
zhongzhou: attend, perform
get arena: hold, win
sm marilao: a @imdanieldpaco
gabby douglas: interview, watch
more... | school: starts, start, is next
national heroes day: classes, holiday, advisory
ningory aquino day: more...
TO DO

• Readings for next week

• Get set up with Hadoop and Pig