computational social media

lecture 3: tweeting

part 1

daniel gatica-perez
this lecture

a human-centric view of twitter

1. introduction
2. twitter users & uses
3. understanding large-scale human behavior
4. inferring real-world events & trends
5. spreading information in the real world
twitter basic official statistics

mission
“to give everyone the power to create and share ideas and information instantly, without barriers”

<table>
<thead>
<tr>
<th></th>
<th>2014</th>
<th>2016</th>
<th>2018 &amp; later</th>
</tr>
</thead>
<tbody>
<tr>
<td>monthly active users</td>
<td>241M</td>
<td>320M</td>
<td>N/A</td>
</tr>
<tr>
<td>tweets sent per day</td>
<td>500M</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>active users on mobile</td>
<td>76%</td>
<td>80%</td>
<td>N/A</td>
</tr>
<tr>
<td>accounts outside the US</td>
<td>77%</td>
<td>79%</td>
<td>N/A</td>
</tr>
<tr>
<td>supported languages</td>
<td>35+</td>
<td>35+</td>
<td>N/A</td>
</tr>
<tr>
<td>employees</td>
<td>2700</td>
<td>3900</td>
<td>N/A</td>
</tr>
</tbody>
</table>
"an echo chamber of random chatter"

“140 characters: between an SMS (with larger audience) and a blog (but less cumbersome)"

280 characters (Fall 2017)

J. van Dijck The culture of connectivity, Oxford University Press, 2013
just setting up my twttr

https://twitter.com/jack/status/20

https://about.twitter.com/milestones
before twitter…

古池や蛙飛び込む水の音
ふるいけやかわずとびこむみずのおと

old pond . . .
a frog leaps in
water’s sound

Bashõ (17th century)

The Dinosaur

On waking, the dinosaur was still there.

Augusto Monterroso (20th century)

https://en.wikipedia.org/wiki/Haiku

https://es.wikipedia.org/wiki/Microrrelato
Dr. Enrico Fermi
Institute of Nuclear Studies
University of Chicago
Chicago, Illinois

RESERVATIONS MADE AT RITTENHOUSE CLUB, 1811 WALNUT STREET FOR BOTH EVENINGS.

HENRY B. ALLEN
THE FRANKLIN INSTITUTE
this lecture

a human-centric view of twitter

1. introduction
2. twitter users & uses
3. understanding large-scale human behavior
4. inferring real-world events & trends
5. spreading information in the real world
what is twitter made of?

@username

follow (2006)
users subscribe to other users’ tweets

#hashtag

hashtags # (2007, official 2009)
words articulating a topic or event
allow for search and clustering

retweet RT (2007, official 2009)
repost tweets towards one's followers
enables trends by retweeting

https://about.twitter.com/press/brand-assets
https://about.twitter.com/milestones
J. van Dijck The culture of connectivity, Oxford University Press, 2013
Hashtags

Link strangers into larger conversations

Facilitate impromptu interactions

Not directed communication but a stream

Enable the emergence of trending topics

---

https://twitter.com/SwissEmbassyUSA/status/1374086527328907270
https://twitter.com/WangenSwiss/status/1372877615472738308
https://twitter.com/MySwitzerland_e/status/1373205370017964035

how do you feel about using # (pound) for groups. As in #barcamp [msg]?

https://twitter.com/chrismessina/status/223115412

ReTweet: jmalthus @spin Yes! Web2.0 is about social media, and guess what people like to be social about? Themselves. Social Narcissism

https://twitter.com/ericrice
geolocation
users and usage

2006: older professional users in business and news
2009: shift to younger adults, then mainstream

from social network to information network

tool for communication
- people’s everyday small talk
- journalism
- political grassroots activism
- emergencies and disasters
- community participation
- misinformation

tool for self-promotion
- celebrities, stars
- politicians

“the impulse to make life a publicly annotated experience has blurred the distinction between advertising and self-expression, marketing and identity” (Hagan 2011)

J. van Dijck The culture of connectivity, Oxford University Press, 2013
CCDF: Complementary Cumulative Distribution Function

\[ \text{CCDF}(x) = 1 - F(x) = P(X>x) \]

Fewer than 10% of users have 100 or more followers

A tiny fraction of users have over 100,000 followers

\( x = 20 \): twitter recommended new users to follow 20 users to start with

\( x = 2000 \): before 2009 there was limit on the number of people one could follow (rule removed later)

following vs. friending

"connection with very low expectation": weak ties (Murthy, 2013)

low reciprocity, highly asymmetric links (Kwak, 2010)

"77.9% of user pairs with any link between them are connected one-way."

"67.6% of users are not followed by any of their followings. For these users Twitter is rather a source of information than a social networking site"

real names in twitter

Twitter does not require real names. Pseudonyms are valuable in information networks: "not real names but persistent identity with reputation attached." It works as identity service for individuals & entities whose long-time presence depends on being identified.

downside: fake user accounts

M. Ingram, Why Twitter doesn’t care what your real name is. GigaOM, Sep 2011
a human-centric view of twitter

1. introduction
2. twitter users & uses
3. understanding large-scale human behavior
4. inferring real-world events & trends
5. spreading information in the real world
case study: twitter and human mood

S. A. Golder and M. W. 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
mood

“a conscious state of mind or predominant emotion” (Merriam-Webster dictionary)

“a temporary state of mind or feeling” (Oxford dictionary)

Photo credit: Danielle MacInnes on Unsplash: https://unsplash.com/photos/IuLgi9PWETU
understanding mood expressed on Twitter

**positive affect (PA):**
- enthusiasm, delight, activeness, alertness

**negative affect (NA):**
- distress, fear, anger, guilt

PA and NA are independent dimensions
- low PA: absence of positive feelings, not presence of negative ones

**goal:** study variations in PA & NA over time of day, day of week, and world region using longitudinal twitter data

* 2.4 million twitter users worldwide
* 509 million tweets
* up to 400 public messages per user
* all users had at least 25 messages
* average: 212 tweets/user
* period: 02.2008 and 01.2010
* only english speakers
extraction of positive affect (PA) and negative affect (NA)

Vice President Kamala Harris 🔵 @VP • Mar 20

United States government official

Sending best wishes to @SuluhuSamia following her swearing in as Tanzania's new President - the first woman to hold the office. The United States stands ready to work with you to strengthen relations between our countries.

https://twitter.com/VP/status/1373341947184693252

Linguistic Inquiry & Word Count (LIWC)
Word categories related to psychological constructs and personal concerns
Word count per category

PA
NA
LIWC2015 is the gold standard in computerized text analysis. Learn how the words we use in everyday language reveal our thoughts, feelings, personality, and motivations. Based on years of scientific research, LIWC2015 is more accurate, easier to use, and provides a broader range of social and psychological insights compared to earlier LIWC versions. Check it out.


https://liwc.wpengine.com
LIWC

dictionary of 6,400 words (2015) (4,500 words in 2007)
each word belongs to one or more categories
“agree” is part of: affect, positive emotions and assent
over 60 word categories (excluding punctuation)
positive emotion category words
negative emotion category words
measurements

\[ PA_u(h) = \frac{||PAWORDS_u(h)||}{||WORDS_u(h)||} \]

where \( h \in H \) and \( H = \{0 \ldots 167\} \), or the 168 hours of the week (24 hours/day * 7 days). The measure for NA was computed similarly, as were the measures taken over 24 hours.

Between-individual variation captures how individuals differ from one another in their baseline affect regardless of the time of day or day of week. It is simply the individual's mean affect across all hours:

\[ BPA_u = \overline{PA_u} = \frac{1}{||H||} \sum_{h \in H} PA_u(h) \]

The within-individual PA score for a person-hour measures the signed difference between the person's score that hour and their baseline as defined in (2). Within-individual scores are comparable across people because individuals' baseline tendencies toward being upbeat or downbeat have been removed, leaving only the change over time that is within each individual:

\[ WPA_u(h) = PA_u(h) - BPA_u + \frac{1}{||UH||} \sum_{(u,h)\in UH} PA_u(h) \]

where \((u, h)\) pairs indicate user-hours and \( UH \) is the set of all such pairs in the dataset.
Fig. 1 Hourly changes in individual affect broken down by day of the week (top, PA; bottom, NA). Each series shows mean affect (black lines) and 95% confidence interval (colored regions).
Fig. 2 Hourly changes in individual affect in four English-speaking regions. Each series shows mean affect (black lines) and 95% confidence interval (colored regions).
A. Mislove et al., Pulse of the Nation: US mood throughout the day inferred from Twitter (2010)
http://www.ccs.neu.edu/home/amislove/twittermood
what to remember

twitter as an information network
brevity as a value
a network of weak links
low reciprocity, highly asymmetric

large-scale human behavior
traces of human states like mood
language in twitter is short yet informative
beware of biased data
questions?

daniel.gatica-perez@epfl.ch