this lecture

introduction & definitions

a human-centric review of research on facebook

descriptive analysis of users
user motivations
user identity
social interaction among users
the real-name web: privacy & information disclosure
announcements

● today
  ● assignment #1
  ● I’ll be absent part of AM session and part of PM session

● next week
  ● assignment #2
  ● Skanda will be in charge of the AM and PM session
## facebook basic statistics (as of Dec 31, 2017)


<table>
<thead>
<tr>
<th>Statistics</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>monthly active users</td>
<td>1.23 B</td>
<td>1.39 B</td>
<td>1.59 B</td>
<td>2.13 B</td>
</tr>
<tr>
<td>mobile monthly active users</td>
<td>945 M</td>
<td>1.19 B</td>
<td>1.44 B</td>
<td>N/A</td>
</tr>
<tr>
<td>daily active users on average</td>
<td>757 M</td>
<td>890 M</td>
<td>1.04 B</td>
<td>1.4 B</td>
</tr>
<tr>
<td>daily active users outside U.S. &amp; Canada</td>
<td>81%</td>
<td>82.4%</td>
<td>83.6%</td>
<td>N/A</td>
</tr>
<tr>
<td>employees</td>
<td>6337</td>
<td>9199</td>
<td>12691</td>
<td>25105</td>
</tr>
</tbody>
</table>
“The top moments were determined by gathering the top keywords by volume mentioned in a single day between 01.01 and 01.11 2017. To identify topics unique to 2017, we compared these keywords to the previous year’s maximum single-day volume. Keywords like Mother’s Day and Halloween, which typically are top moments every year, do not repeatedly appear at the top of the list.”
Top Life Events
Life events people added to their Timeline most frequently in 2013

1. Added a relationship, got engaged or got married
2. Traveled
3. Moved
4. Ended a relationship
5. First met a friend
6. Added a family member, expecting a baby or had a baby
7. Got a pet
8. Lost a loved one
9. Got a piercing
10. Quit a habit
definitions: social network site

“web-based [and mobile] services that allow individuals to

(1) construct a public or semi-public profile within a bounded system

(2) articulate and make visible a list of other users with whom they shared a connection

(3) view and traverse their list of connections and those made by others within the system

the nature of these connections may vary from site to site… meeting strangers is possible but not necessarily the primary practice”

“...the questions, methods, and perspectives were so diverse and fragmented that it would be impossible to write a coherent summary of the literature.”

“...but we also realized that without summarizing the current trends the situation was unlikely to improve.”

412 articles (end 2011)
**Table 1. Areas of Facebook Research Identified in the Literature Review**

<table>
<thead>
<tr>
<th>Area of research</th>
<th>No. of articles</th>
<th>% of total</th>
<th>Associated research question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive analyses of users</td>
<td>97</td>
<td>24%</td>
<td>Who is using Facebook, and what are users doing while on Facebook?</td>
</tr>
<tr>
<td>Motivations for using Facebook</td>
<td>78</td>
<td>19%</td>
<td>Why do people use Facebook?</td>
</tr>
<tr>
<td>Identity presentation</td>
<td>50</td>
<td>12%</td>
<td>How are people presenting themselves on Facebook?</td>
</tr>
<tr>
<td>Role of Facebook in social interactions</td>
<td>112</td>
<td>27%</td>
<td>How is Facebook affecting relationships among groups and individuals?</td>
</tr>
<tr>
<td>Privacy and information disclosure</td>
<td>75</td>
<td>18%</td>
<td>Why are people disclosing personal information on Facebook despite potential risks?</td>
</tr>
<tr>
<td>Total</td>
<td>412</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>
facebook users & articles: growth over time

Fig. 1. Facebook users and articles: Cumulative totals by year
topic 1:
descriptive analysis of users who uses facebook and what do they do?
graphs as models of networks: quick basics

All materials of this section are taken from:
Figure 2.1: Two graphs: (a) an undirected graph, and (b) a directed graph.
Figure 2.2: A network depicting the sites on the Internet, then known as the Arpanet, in December 1970. (Image from F. Heart, A. McKenzie, J. McQuillian, and D. Walden [214]; on-line at http://som.csudh.edu/cis/lpress/history/arpamaps/.)

Figure 2.3: An alternate drawing of the 13-node Internet graph from December 1970.
**paths and cycles**

**Path**: sequence of nodes where each consecutive pair is connected by an edge

**Simple path**: paths that contain no repeated nodes

**Cycle**: a path with at least three edges, in which the first and last nodes are the same, but otherwise all nodes are distinct

Figure 2.3: An alternate drawing of the 13-node Internet graph from December 1970.
**connectivity**

**Connected component:** a subset of the nodes such that:
(i) every node in the subset has a path to every other;
(ii) the subset is not part of some larger set with the property that every node can reach every other.

(i) says that the component is indeed internally connected
(ii) says that it really is a free-standing piece of the graph

Figure 2.5: A graph with three connected components.
Giant component: connected component that contains a significant fraction of all the nodes in large networks.

When a network contains a giant component, it almost always contains only one.

Figure 2.7: A network in which the nodes are students in a large American high school, and an edge joins two who had a romantic relationship at some point during the 18-month period in which the study was conducted [49].
path length and distance between nodes

Path length: number of edges in the sequence that comprises it
Distance between two nodes: length of the shortest path between them

\[ l(MIT, BBN, RAND, UCLA) = 3 \]
\[ l(MIT, UTAH) = 1 \]
\[ d(LINC, SRI) = 3 \]

I(MIT, BBN, RAND, UCLA) = 3
I(MIT, UTAH) = 1
breadth-first search: a method to determine distances

It searches a graph from a starting node, reaching the closest nodes first.

It serves as a conceptual framework to organize a graph’s structure, arranging the nodes based on their distances from a fixed starting point.

Figure 2.9: The layers arising from a breadth-first of the December 1970 Arpanet, starting at the node MIT.
small-world phenomenon

six degrees of separation
  + any two people are separated by no more than six intermediate connections
  + the world looks “small” given these short paths
  + proposed by Frigyes Karinthy in short story (1929)
  + popularized by John Guare’s *Six Degrees of Separation* play (1990)

“I read somewhere that everybody on this planet is separated by only six other people. Six degrees of separation between us and everyone else on this planet.”
Stanley Milgram’s small-world experiment (1960s)

- 296 volunteers living in US midwest
- they sent a message to a person living in Boston suburbs
- volunteers could not send message directly to target (unless personal contact)
- they sent message to a personal contact who was likely to know the target
- 64 chains reached the target
- avg. # of intermediate persons = 5.2
  (distance: 6.2)

Figure 2.10: A histogram from Travers and Milgram’s paper on their small-world experiment [391].

people navigate short paths with success, even if entire network is not visible
the anatomy of facebook
(Backstrom et al 2011)

721 million facebook users
69 billion friendships
largest social network ever studied

research questions:
1. how many friends do people have?
2. how many degrees of separation between any two users?

L. Backstrom, Anatomy of Facebook, 21.11.2011

how many friends?
(Backstrom et al 2011)

cumulative degree distribution: % of people with less than X friends

10% of people: less than 10 friends
20% of people: less than 25 friends
50% (the median): less than 100 friends
average friend count = 190
4 degrees of separation (degrees = intermediaries) (Backstrom et al 2011)

99.6% of user pairs are connected by at most 5 degrees (6 edges)
92% connected by 4 degrees (5 edges)

average distance (2008) = 5.28
average distance (2011) = 4.74

within same country, connectivity is even higher
84% of all connections are between users in the same country
4 degrees of separation, revisited 2016
https://research.facebook.com/blog/three-and-a-half-degrees-of-separation/

average distance (2008) = 5.28
average distance (2011) = 4.74
average distance (2016) = 4.57

recomputed on entire FB graph: 1.59 B users
homophily in facebook
(Backstrom et al 2011)

users’ friends are most likely to be of similar age

homophily
“tendency of individuals to associate and bond with similar others” (wikipedia)

birds of a feather flock together
small-world phenomenon in MS instant messenger
(Leskovec et al 2008)

Figure 2.11: The distribution of distances in the graph of all active Microsoft Instant Messenger user accounts, with an edge joining two users if they communicated at least once during a month-long observation period [273].

topic 2: motivations
why do people use facebook?
top motivations why people use facebook

desire to keep in touch with friends (Ellison et al. 2006) (Lampe et al. 2006)

engage in social grooming
gossip & small talk (Tufekci, 2008) (Gosling, 2009)

develop social capital
benefits from relationships (Ellison et al. 2007) (Burke et al. 2010)

reduce boredom & loneliness (Burke et al. 2010) (Lampe et al. 2010)

participate
add content & join groups (Viswanath et al. 2009)

credit: sean macentee @ flickr (cc): http://www.flickr.com/photos/smemon/5684115572/
motivation #1:
social grooming
Robin Dunbar (1998)

**gossip, small talk, people-curiousity**

human equivalent of social grooming in primates
language allowed people to live in larger groups (~150)

“essential to forging bonds, affirming relationships, displaying bonds, learning about hierarchies and alliances”

“a way to figure out where we all stand in relation to each other”

“means to improve one’s reputation and access to resources and solidarity”

---


credit: lawrence murray @ flickr (cc): http://www.flickr.com/photos/lawmurray/3065124869/
how many people can anyone be friends with?

Dunbar’s number:

“The figure of 150 seems to represent the maximum number of individuals with whom we can have a genuinely social relationship, the kind of relationship that goes with knowing who they are and how they relate to us”

“It’s the number of people you would not feel embarrassed about joining uninvited for a drink if you happened to bump into them in a bar”

intimates -> close personal friends -> friends -> acquaintances -> strangers

R. Dunbar, "Neocortex size as a constraint on group size in primates". Journal of Human Evolution, 22 (6), 1992
http://en.wikipedia.org/wiki/Dunbar%27s_number
D. Bennett, The Dunbar number, from the guru of social networks, Jan. 2013
assignment #1: who are your Facebook friends?

1. Watch the talk by Robin Dunbar at ICWSM 2012 (1 hour)
   http://videolectures.net/icwsm2012_dunbar_facebook/

2. Go to your FB page, tab “All friends”, and manually categorize your own network, responding to the following questions
   - how many FB friends do you have?
   - how many are
     (1) intimate (e.g. partner, parents, best friend)
     (2) close personal friend
     (3) friend
     (4) acquaintance
     (5) stranger (people-you-don’t-really-know-but-somehow-got-in)
   - what is your FB Dunbar’s number?
   NOTE: if you have more than 250 FB friends, only categorize a random set of 250, but indicate the number you actually have.

3. Write a summary with the above information, adding any thoughts inspired by the above exercise and Dunbar’s talk (one-page max)

4. If you are not on Facebook: do step (1) above in class, and then your assignment will be to report and present the collected stats in class
assignment #1: logistics and deadlines

1. Rest of morning session: watch the talk by Robin Dunbar.

2. Afternoon session:
   - finish watching the talk
   - start working on the categorization task
   - if finished, start working on the one-page report

3. HARD DEADLINE TO SUBMIT REPORT: Wed 07.03.2018, midnight
   - send by email to:
     daniel.gatica-perez@epfl.ch
     skanda.muralidhar@epfl.ch
   - pdf format
   - late assignments will NOT be given any credit

4. People with no Facebook: talk to Skanda in the PM session to get organized:
   - presentation of aggregated results in class: Fri 16.03.2018
   - send presentation slides to both Daniel & Skanda on the same day
questions?

daniel.gatica-perez@epfl.ch