Introduction to Social Computing

Social Computing

Department of Computer Science
University of Massachusetts, Lowell

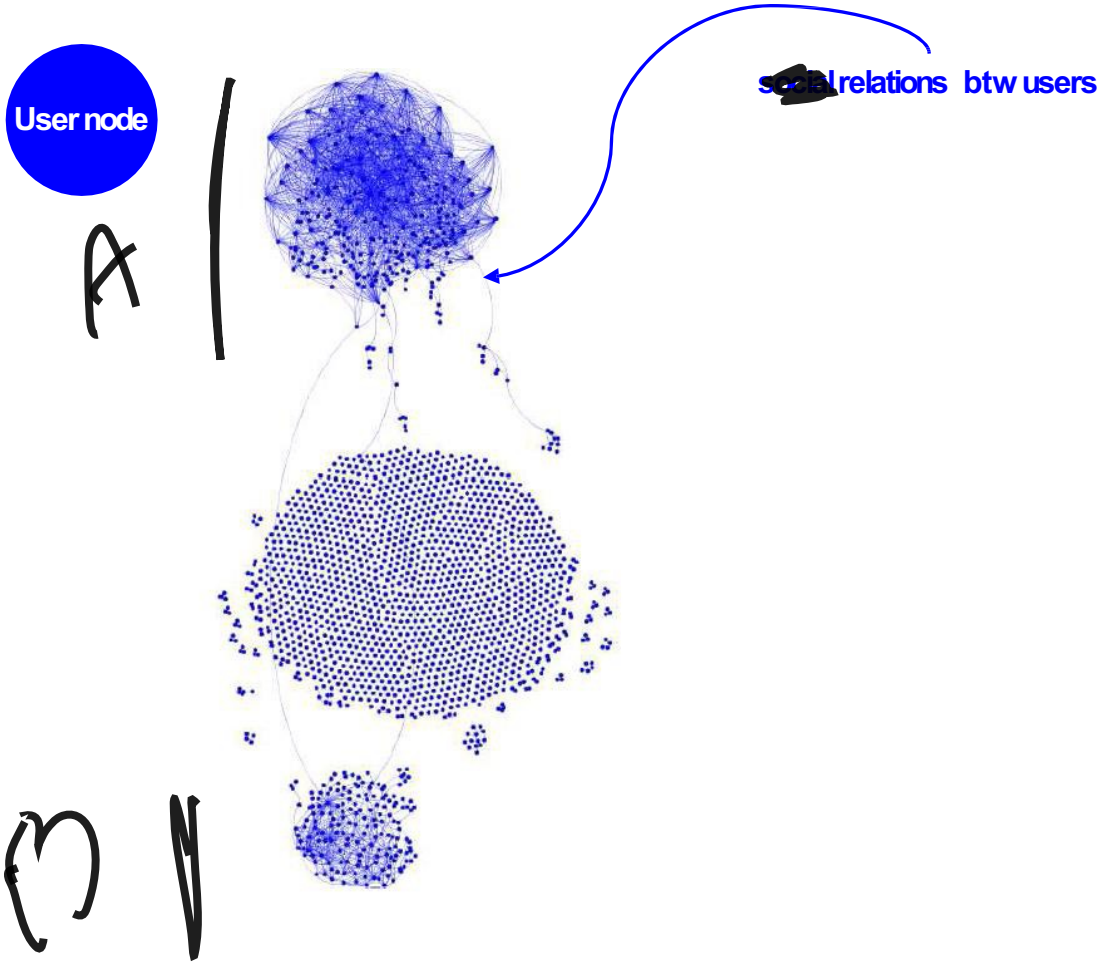
Hadi Amiri
hadi@cs.uml.edu

What's This Course about?

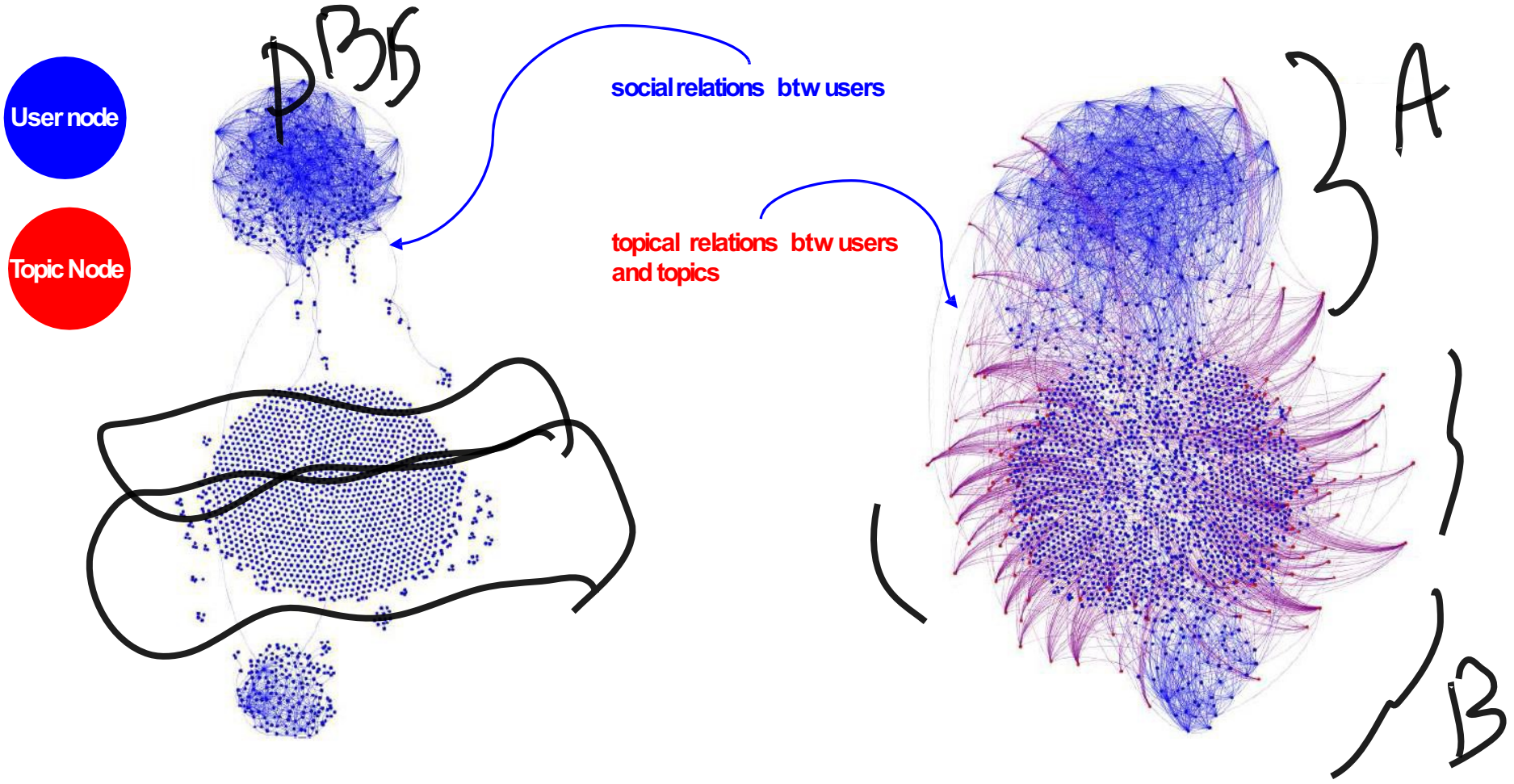
- Understanding various social phenomena through studying:
  - User-generated Content in Social Networks
    - deal with various user generated content and their propagation in networks.
  - Properties, design principles, and models!

Data Proliferation

Net & Content Interactions



Net & Content Interactions



What Are Networks?

- **Communication Networks**
  - Telco Nets
  - Messenger Nets
- **Friendship Networks**
  - Facebook
- **Microblogs**
  - Twitter
- **Information Networks**
  - Web!

Examples


vimeo
Zimbra
Renkoo
standpoint
SHOUTWIRE
gOFFICE
blogbeat
vSocial
wink
FeedBurner
Bloglines
Yub.com
ourmedia
Gcast
openomy
ajchat
ClipShack
chatsum
PANDORA
looklater
PROTOPAGE
squishr
WebJay
Noodly
box
Jots
Xdrive
vizu
digg
del.icio.us
AlmondRocks
Tagyu
writely
Simpy
Gtalkr
TRUVEO
egoSurf
Quimble
pegasus
remember the milk
SQUIDOO
picturecloud
newsvine
Clipfire
mozy
Basecamp
Netvibes
My Video karaoke
Lexxe alpha
powered by advanced natural language technology
facebook
PXN8
photos made easy
43 Things
Technorati
Backpack
Gmail
yelp
Real People. Real Reviews.
Smarkets
magnolia
ShoZu
feedmarker
filangy
mercora
MusicSearch
Meet With Approval.com
HomePortals
Spin spy
(beta)

Sample 1.

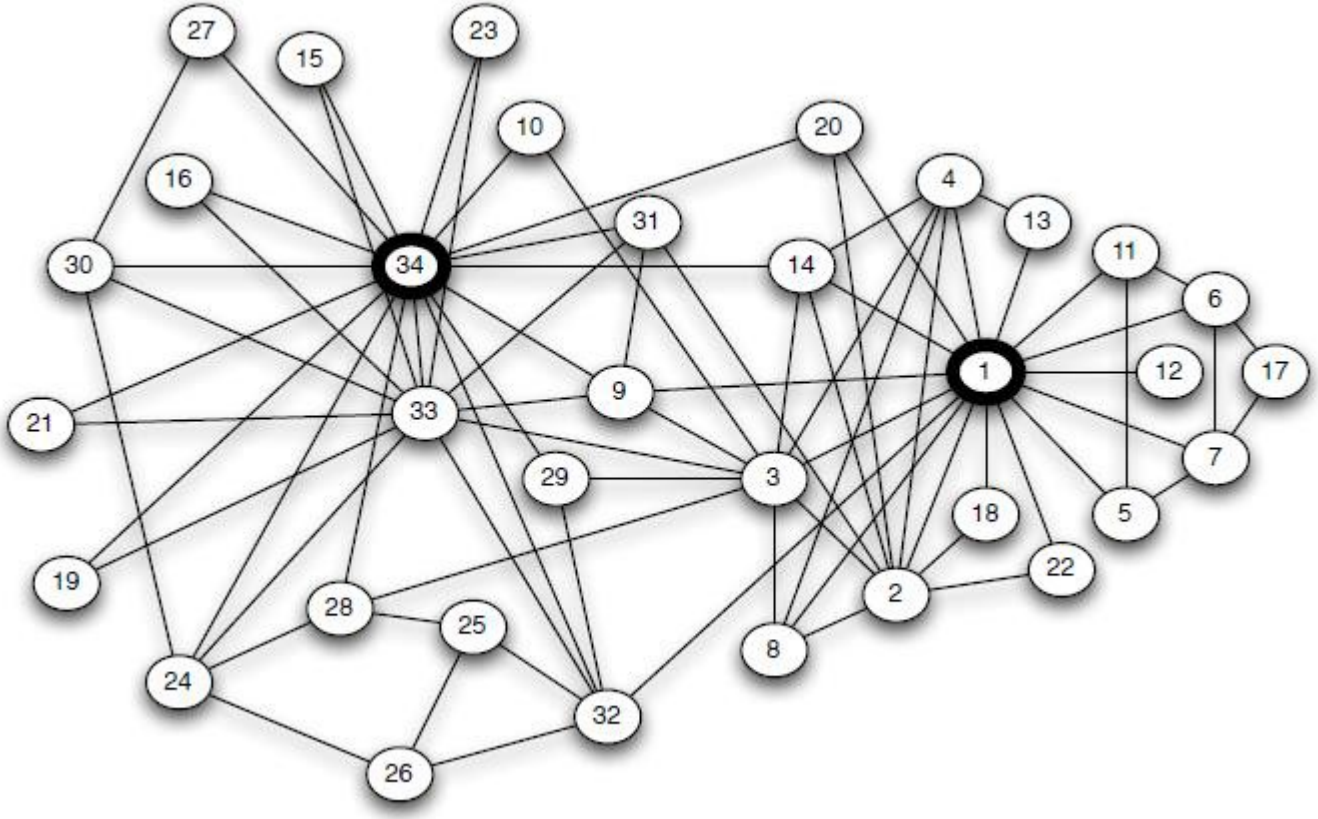


Figure 1.1: The social network of friendships within a 34-person karate club [421].

Sample 2.

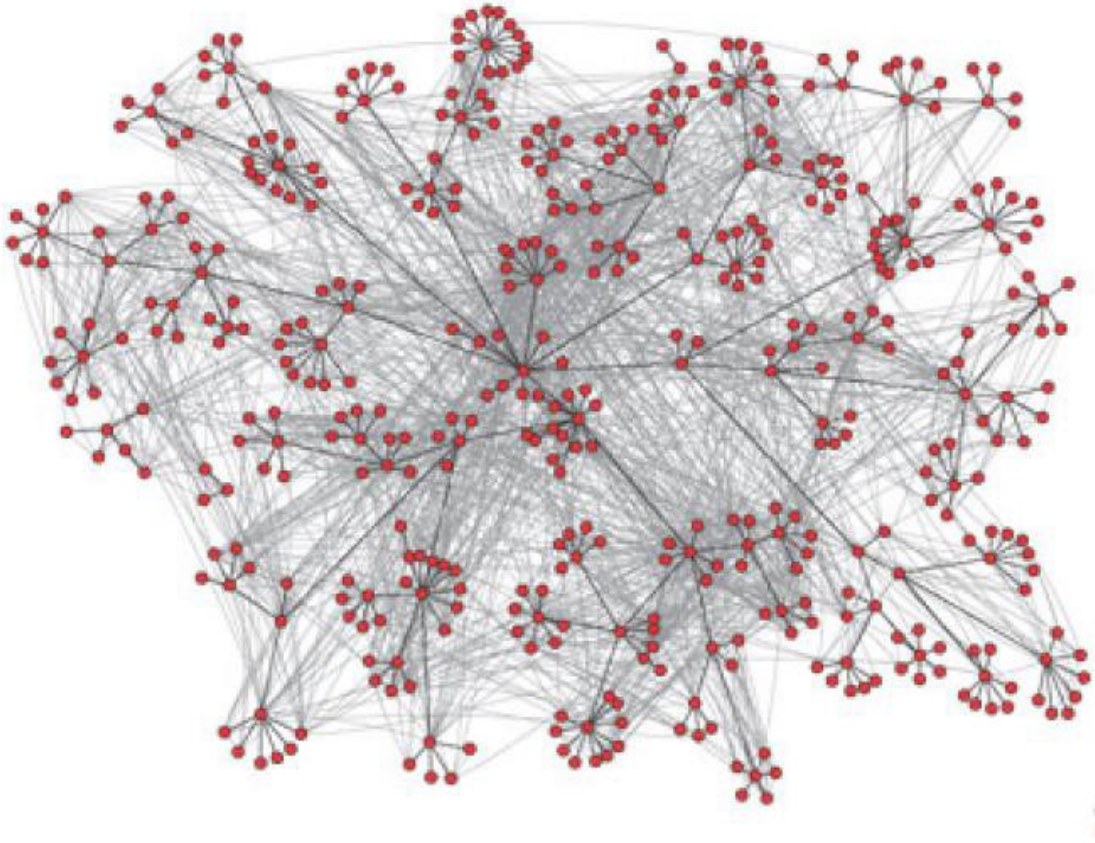


Figure 1.2: Social networks based on communication and interaction can also be constructed from the traces left by on-line data. In this case, the pattern of e-mail communication among 436 employees of Hewlett Packard Research Lab is superimposed on the official organizational hierarchy [6]. (Image from http://www-personal.umich.edu/ ladamic/img/hplabsemailhierarchy.jpg)

Sample 3.

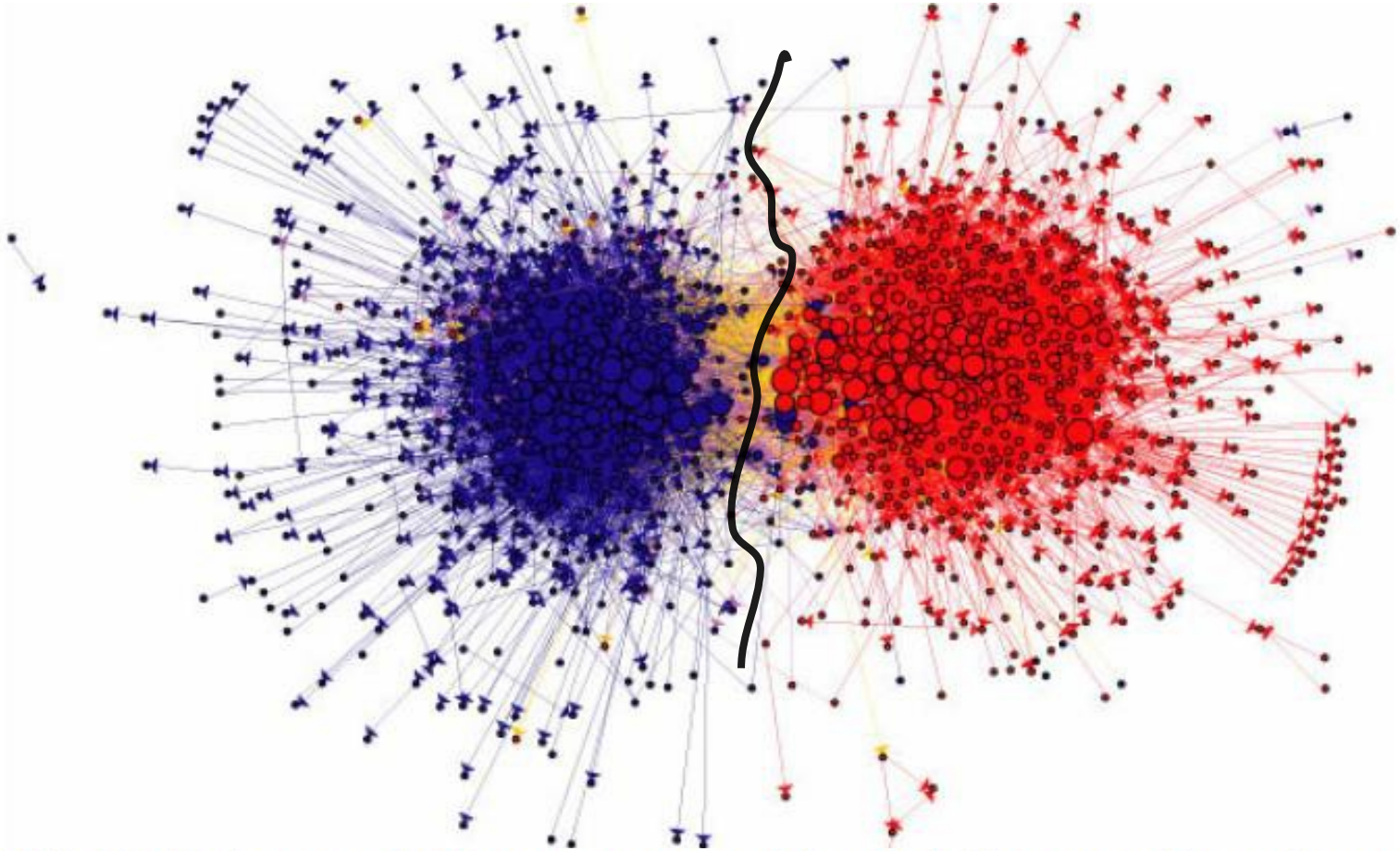


Figure 1.4: The links among Web pages can reveal densely-knit communities and prominent sites. In this case, the network structure of political blogs prior to the 2004 U.S. Presidential election reveals two natural and well-separated clusters [5]. (Image from http://www-personal.umich.edu/ ladamic/img/politicalblogs.jpg)

Source: [NCM] Easley and Kleinberg. 2010. Networks, Crowds, and Markets: Reasoning about a Highly Connected World.

Sample 4.

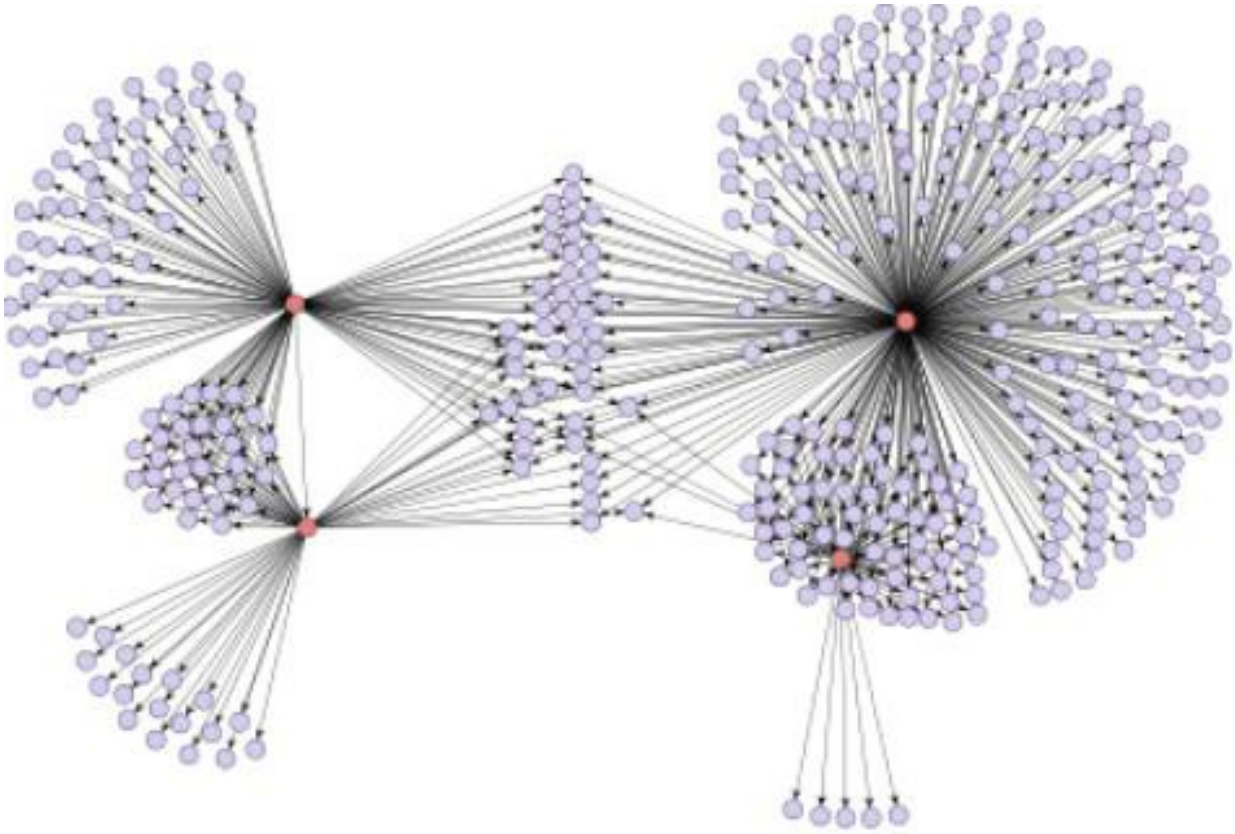


Figure 1.11: When people are influenced by the behaviors their neighbors in the network, the adoption of a new product or innovation can cascade through the network structure. Here, e-mail recommendations for a Japanese graphic novel spread in a kind of informational or social contagion. (Image from Leskovec et al. [271].)

Source: [NCM] Easley and Kleinberg. 2010. Networks, Crowds, and Markets: Reasoning about a Highly Connected World.

Sample 5.

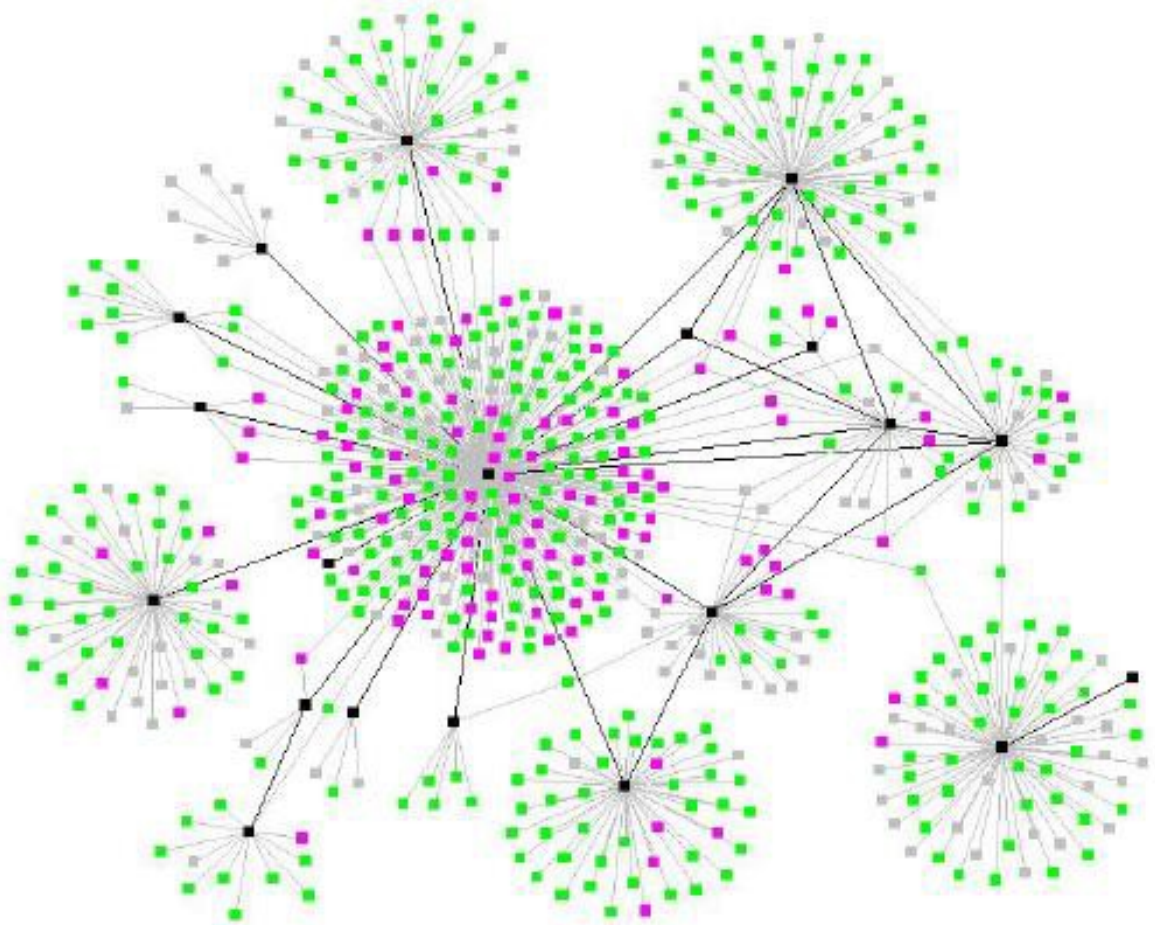
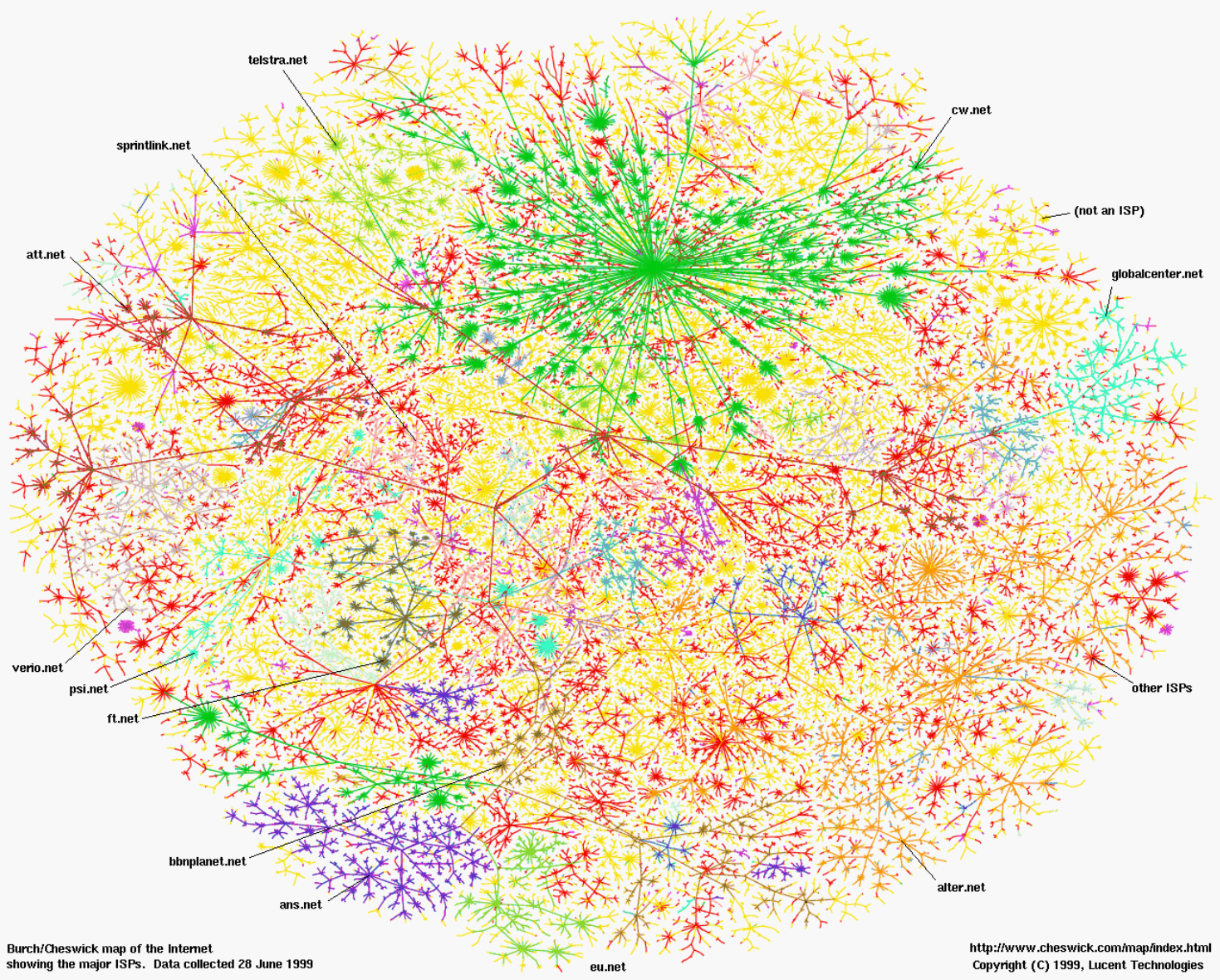


Figure 1.12: The spread of an epidemic disease (such as the tuberculosis outbreak shown here) is another form of cascading behavior in a network. The similarities and contrasts between biological and social contagion lead to interesting research questions. (Image from Andre et al. [16].)

Sample 6.

Network of Major ISPs.
1999



Source: http://www.cheswick.com/ches/map/gallery/index.html

Sample 7.

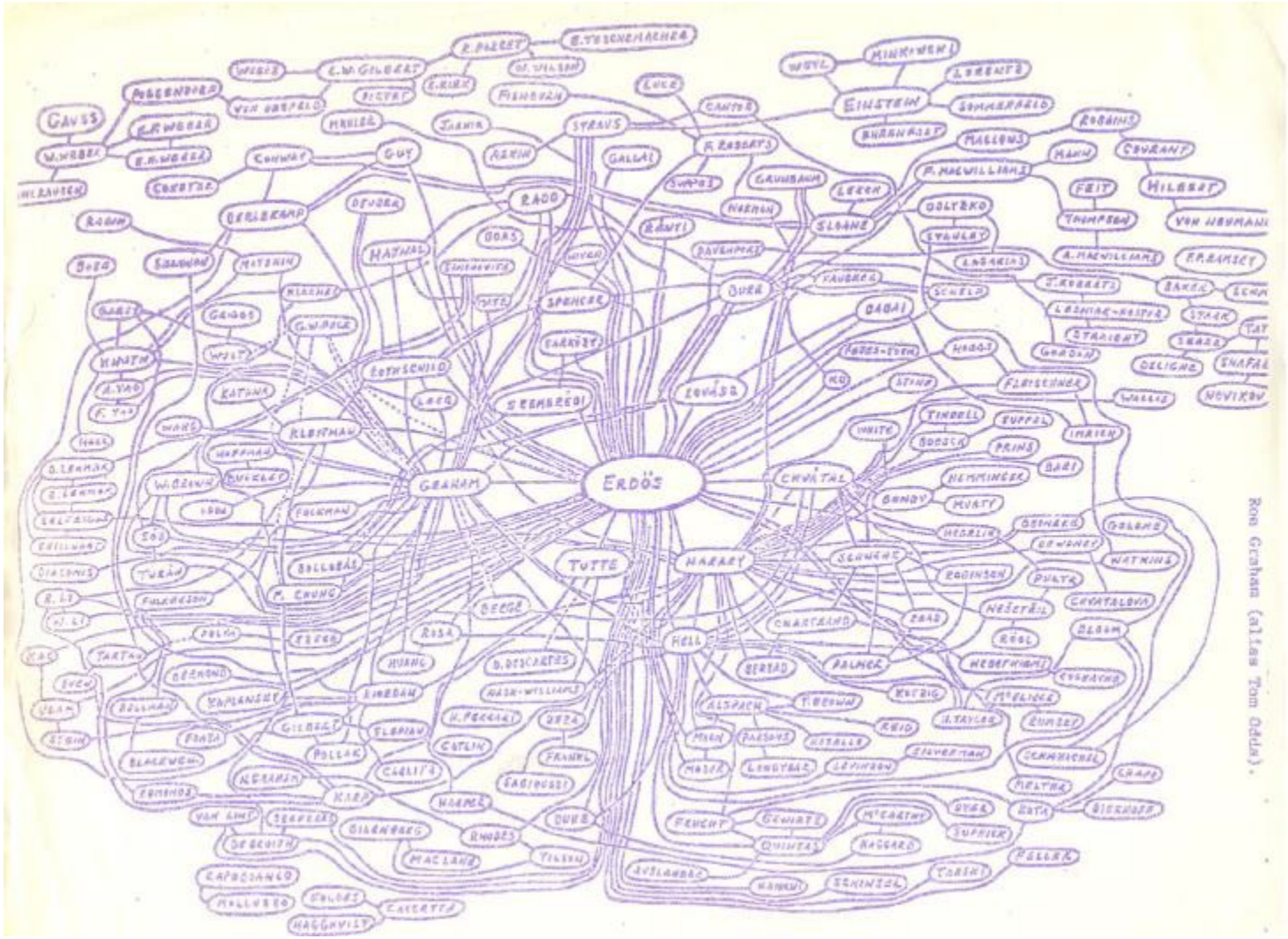


Figure 2.12: Ron Graham's hand-drawn picture of a part of the mathematics collaboration graph, centered on Paul Erdös [189]. (Image from http://www.oakland.edu/enp/cgraph.jpg)

Sample 8.

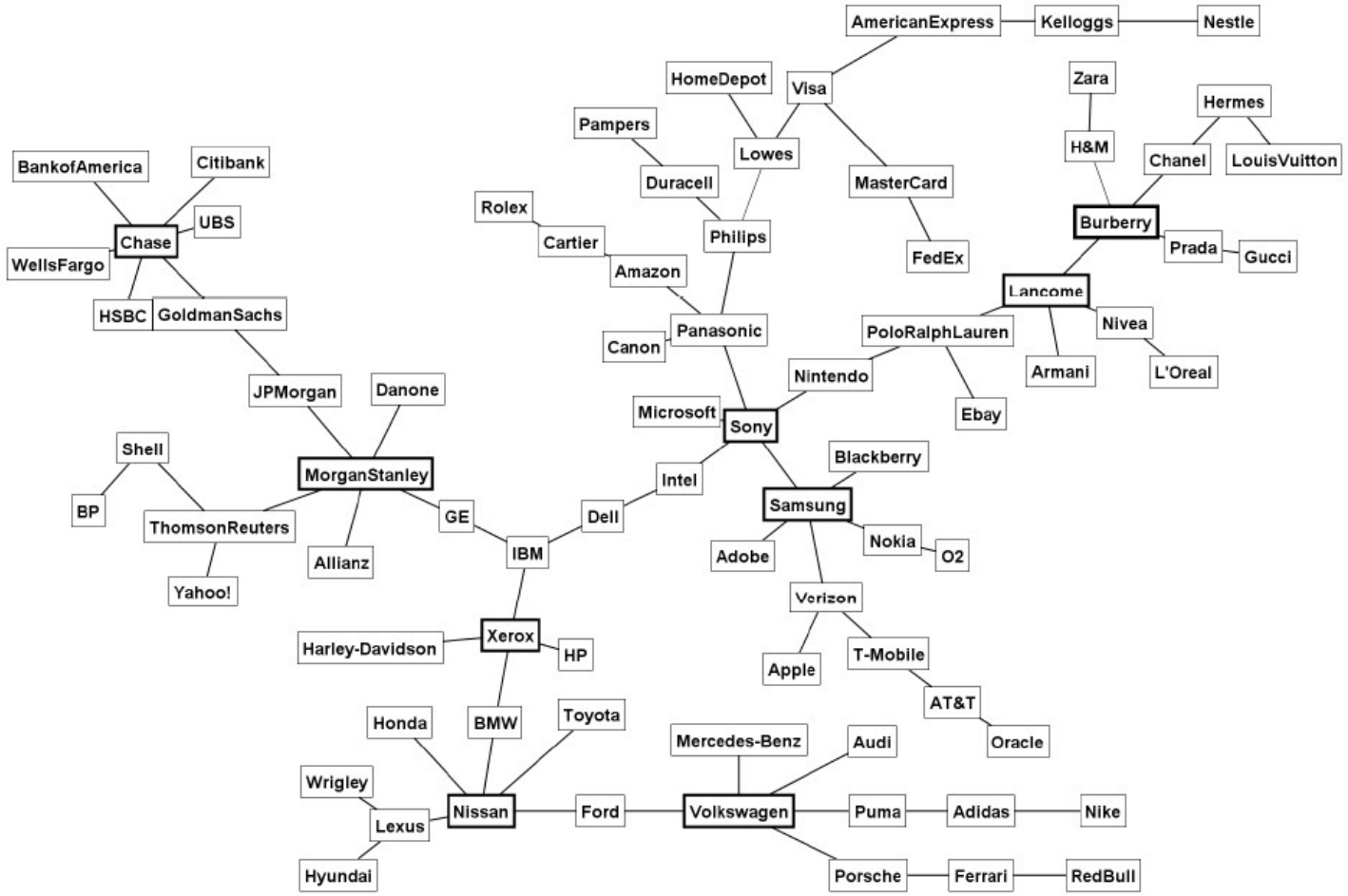


Figure 3. Minimum spanning tree (MST) of the most valued global brands. The MST of the brand network is the subset of edges that forms a tree reaching every brand such that the total length of all the edges is minimized. It is readily apparent that certain brands stand out prominently as hubs with connections to other brands radiating out from them. These hubs are generally the centers of well-formed market category groupings.

Source: Inferring brand proximities from user-generated content. Paul Dwyer. J of Brand Management 19, 2012.

Sample 9.

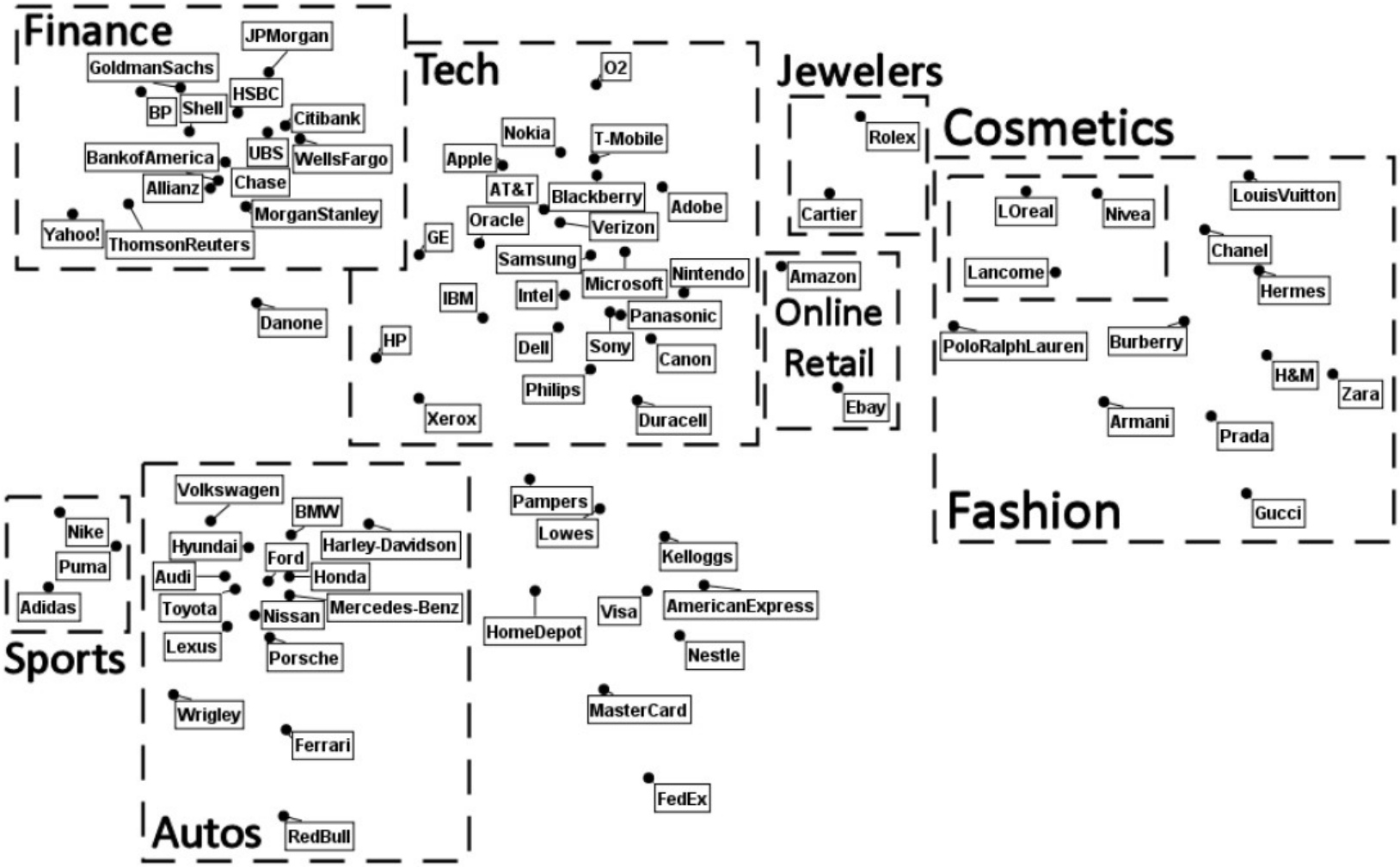
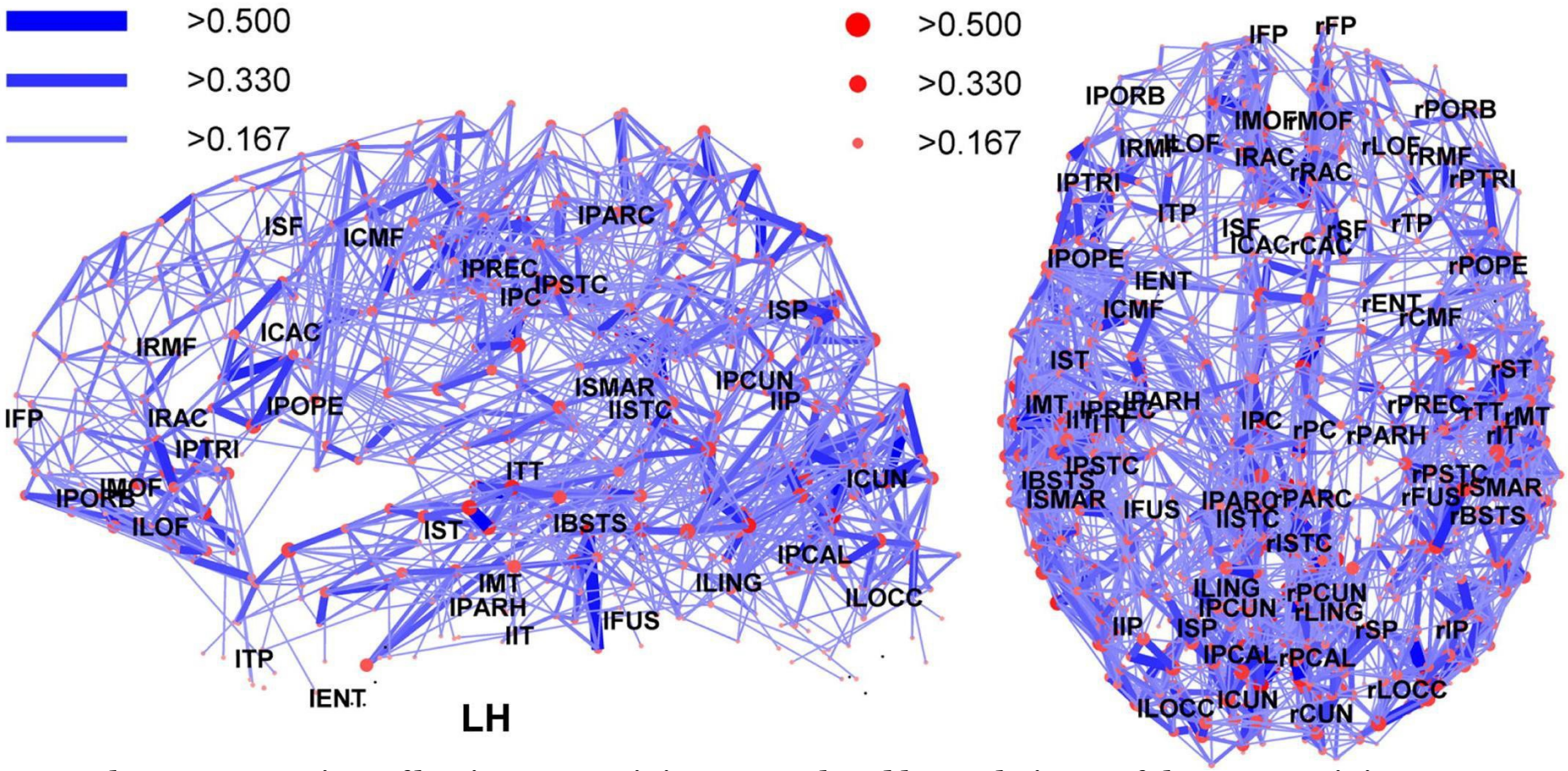


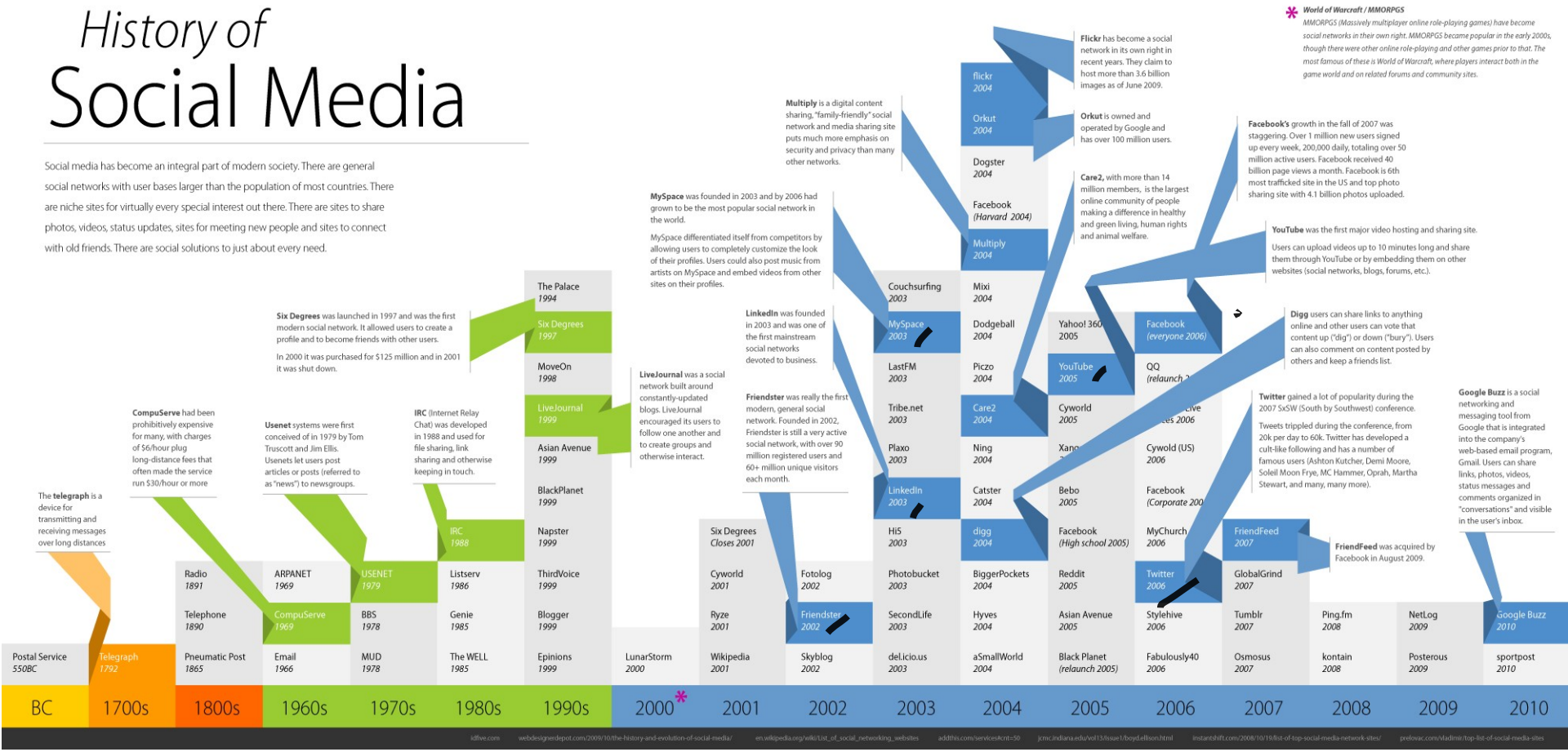
Figure 4. Map of brands. The minimum spanning tree augmented by triangulating each brand location from their nearest neighbors with forced-based layout yields a map high in face validity. Note the eight strong market category groupings outlined with broken lines.

Sample 10.

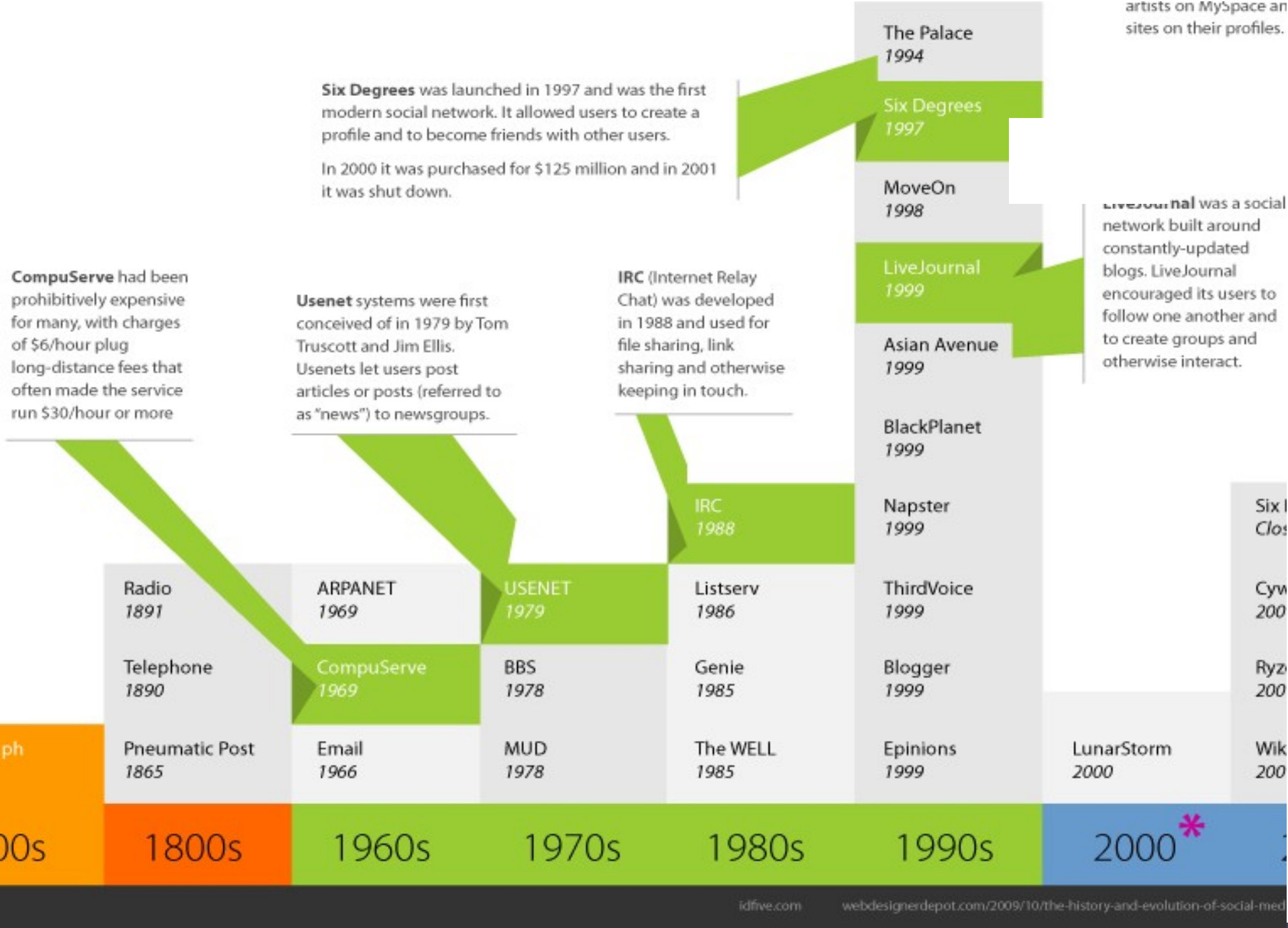


Network representation of brain connectivity: Dorsal and lateral views of the connectivity backbone of human brain. Labels indicating anatomical subregions are placed at their respective centers of mass. Nodes (individual ROIs) are coded according to strength and edges are coded according to connection weight.

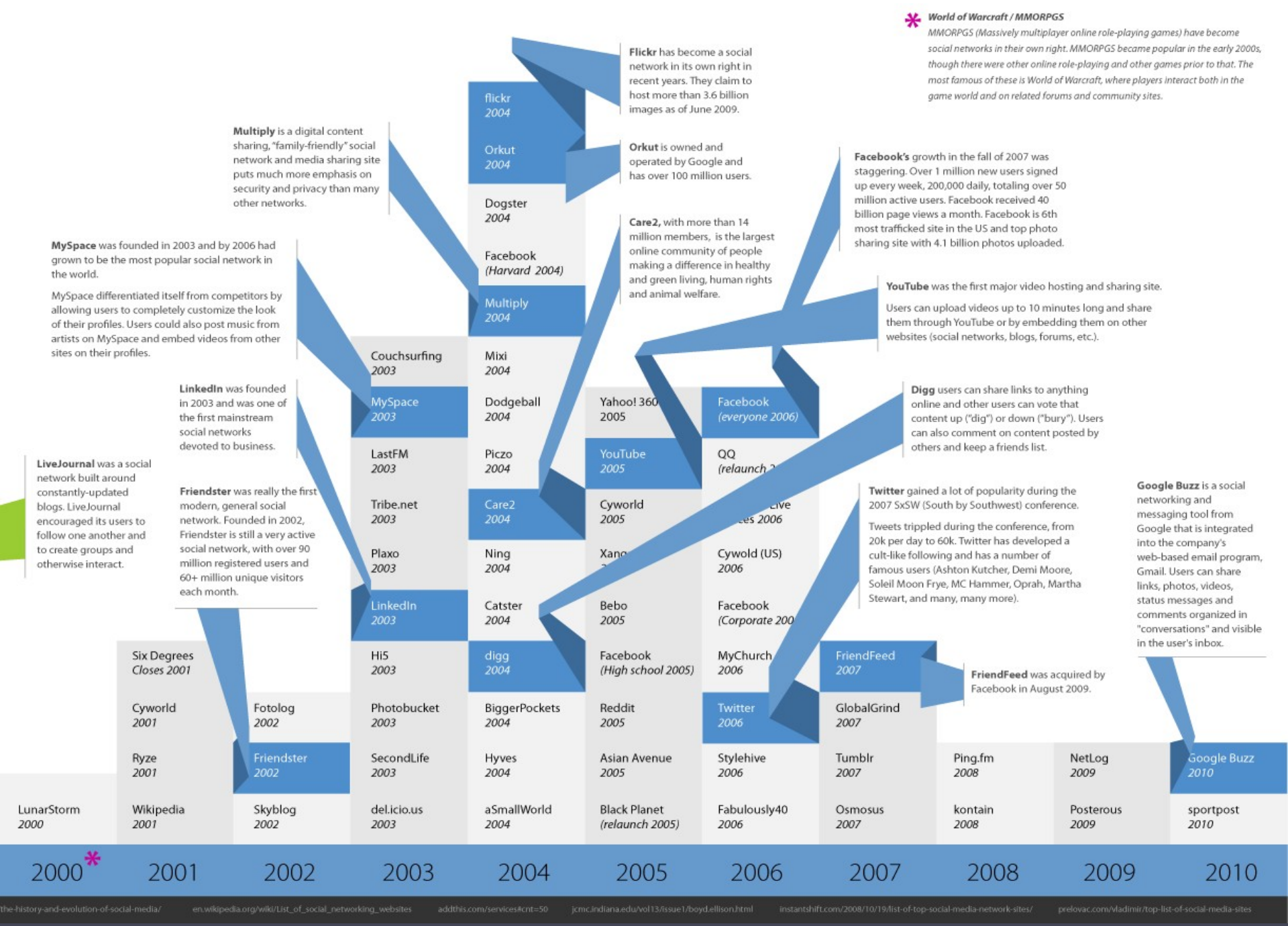
How Long They've Been Around?



Source: http://dustn.tv/stay-on-the-cutting-edge/



Source: http://dustn.tv/stay-on-the-cutting-edge/



Source: http://dustn.tv/stay-on-the-cutting-edge/

Why Should We Study Them?

- Social Networks provide cheap and high resolution views into population behavior:
  - Spread of news or diseases
  - Evolution of science
  - Stock market dynamics
  - Consumer response to products

Why Should We Study Them? Cnt.

- Computer Scientists
  - Algorithms and models
  - Computational challenges

TWEETS 779K | FOLLOWING 16.5K | FOLLOWERS 109K | Follow

VZW Support

@VZWSupport FOLLOWS YOU

Customer Support for Verizon Wireless. ?'s about your wireless service, device, features, etc. we're here to assist. 7 days a week from 7am - 2am CST

community.verizonwireless.com

TWEETS 599K | FOLLOWING 45.1K | FOLLOWERS 870K | Follow

American Airlines

@AmericanAir

Thanks for checking in! We're here to offer advice and inspiration for your trip on American. Please click here if you require a formal response to a complaint:

bit.ly/AACR1

TWEETS 37.6K | FOLLOWING 1,866 | FOLLOWERS 5,191 | Follow

McD Customer Service

@Reachout_mcd

McDonald's U.S. Customer Service. Here to listen, help or answer any questions you have 7 days a week 7:00am to 7:00pm CST

Oak Brook, IL · mcd.to/ULtdKh

TWEETS 211K | FOLLOWING 7,234 | FOLLOWERS 63K | Follow

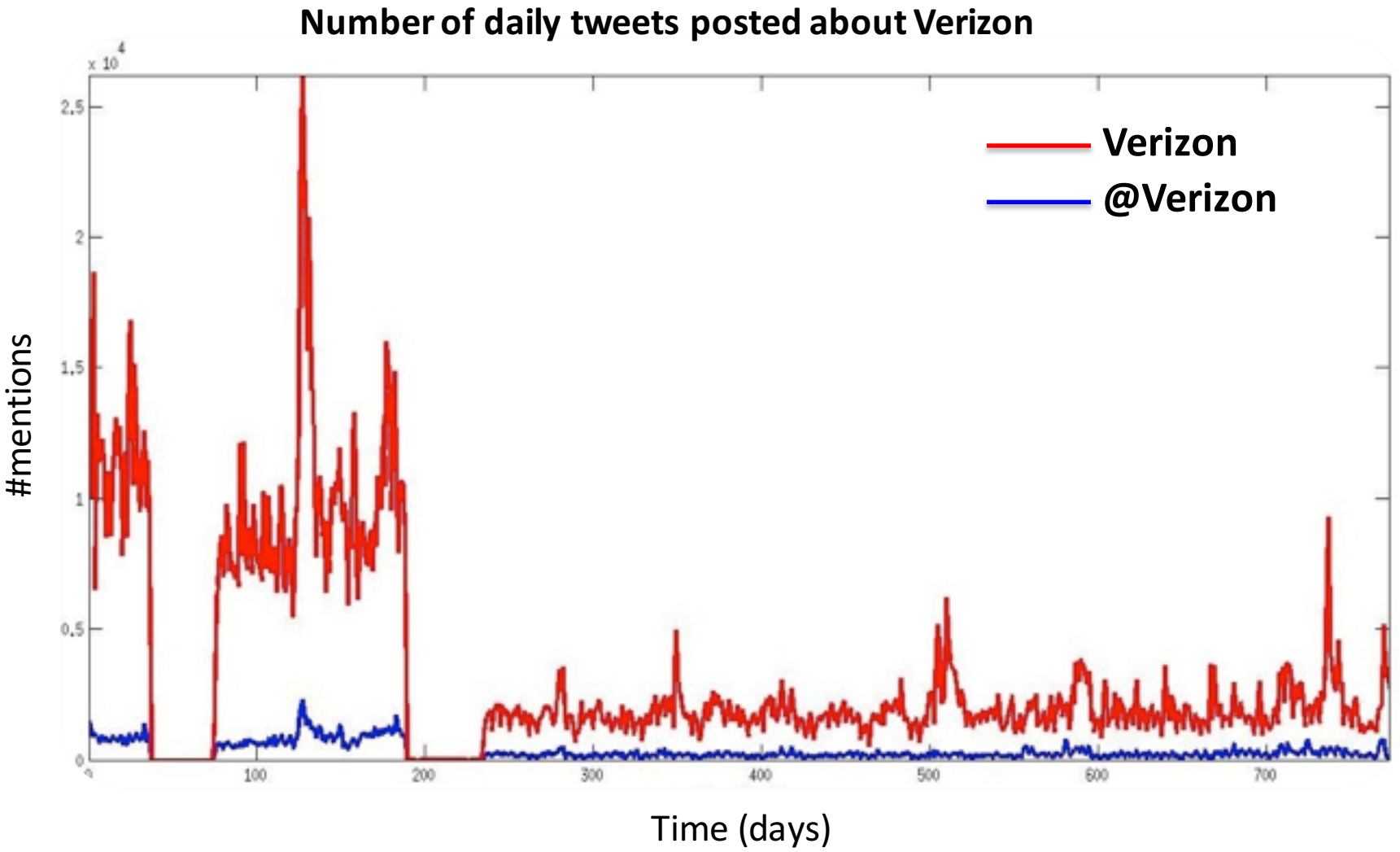
Telstra

@Telstra

We're here 24x7 to provide customer support and answer any Telstra questions you might have. Last week our average response time was 20 minutes

Australia · telstra.com.au

Why Should We Study Them? Cnt.



Why Should We Study Them? Cnt.

Figure 25: Brand and Reputation Monitoring of SMNs

Overall 465 respondents, LOB=107, EMEA=168

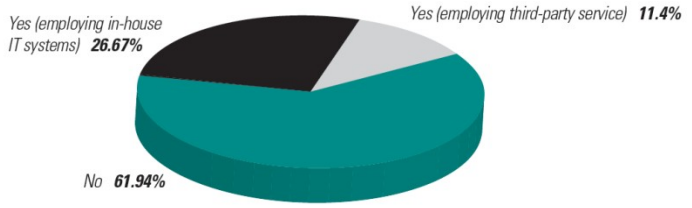


Figure 27a: Organizational Plans to Leverage Social Media Metrics Into Business Processes

Overall 459 respondents, LOB=107, EMEA=167

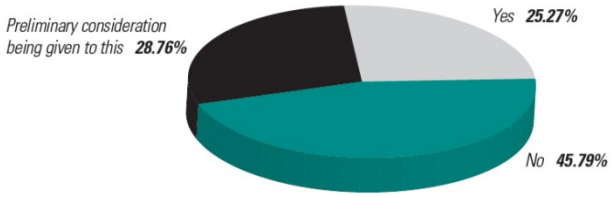
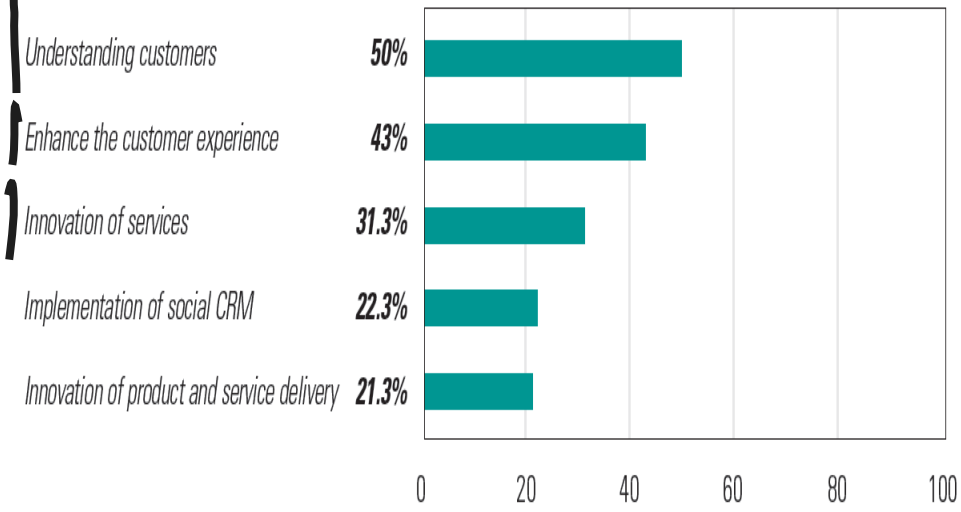


Figure 33: Top Business Processes Leveraging Social Media Data

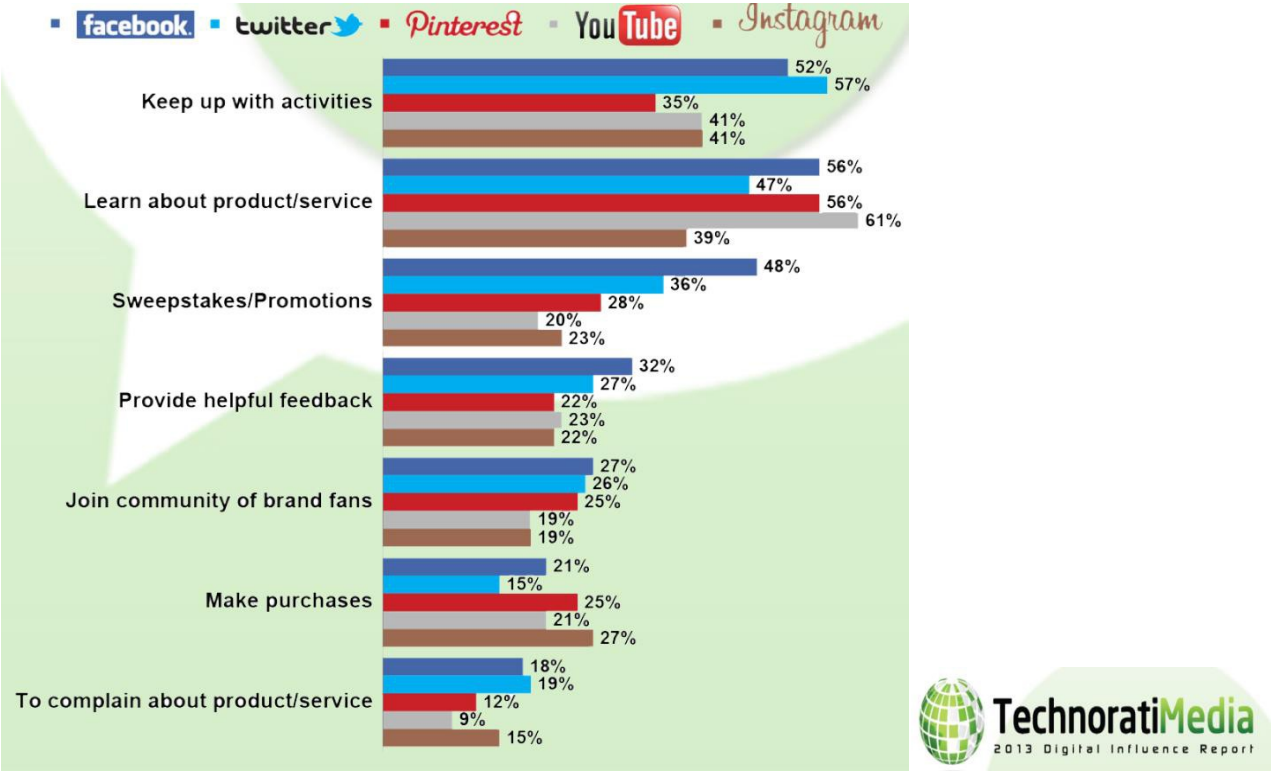
Overall 300 respondents, LOB=79, EMEA=105



Sponsored by IBM MARIST

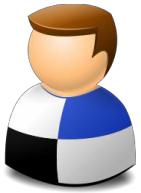
Produced by UNISPHERE RESEARCH Delivering Certainty
Thomas J. Wilson, President

Why Should We Study Them? Cnt

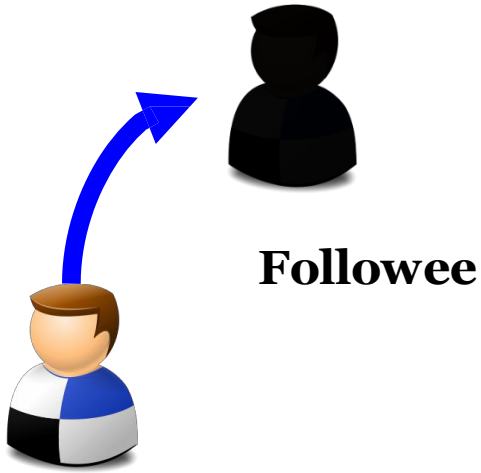


Let's Take a Closer Look at Twitter

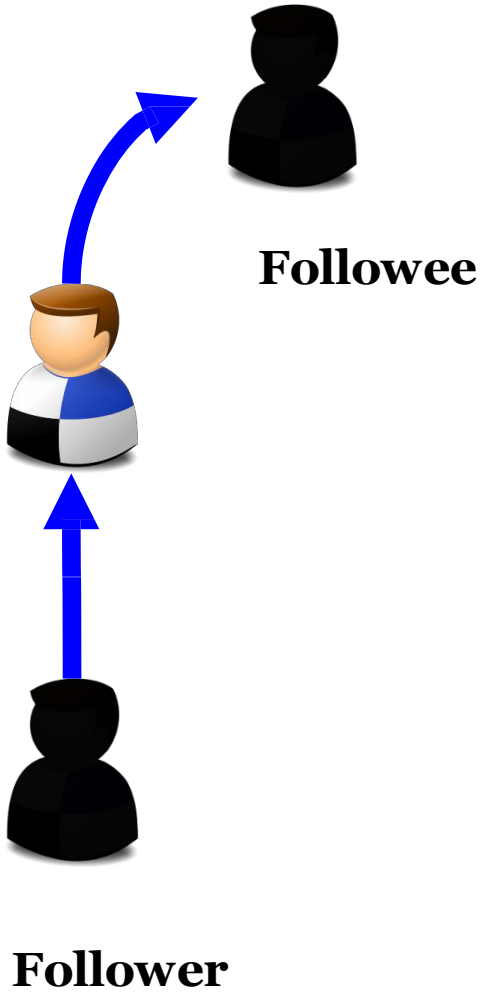
- Simple Structure

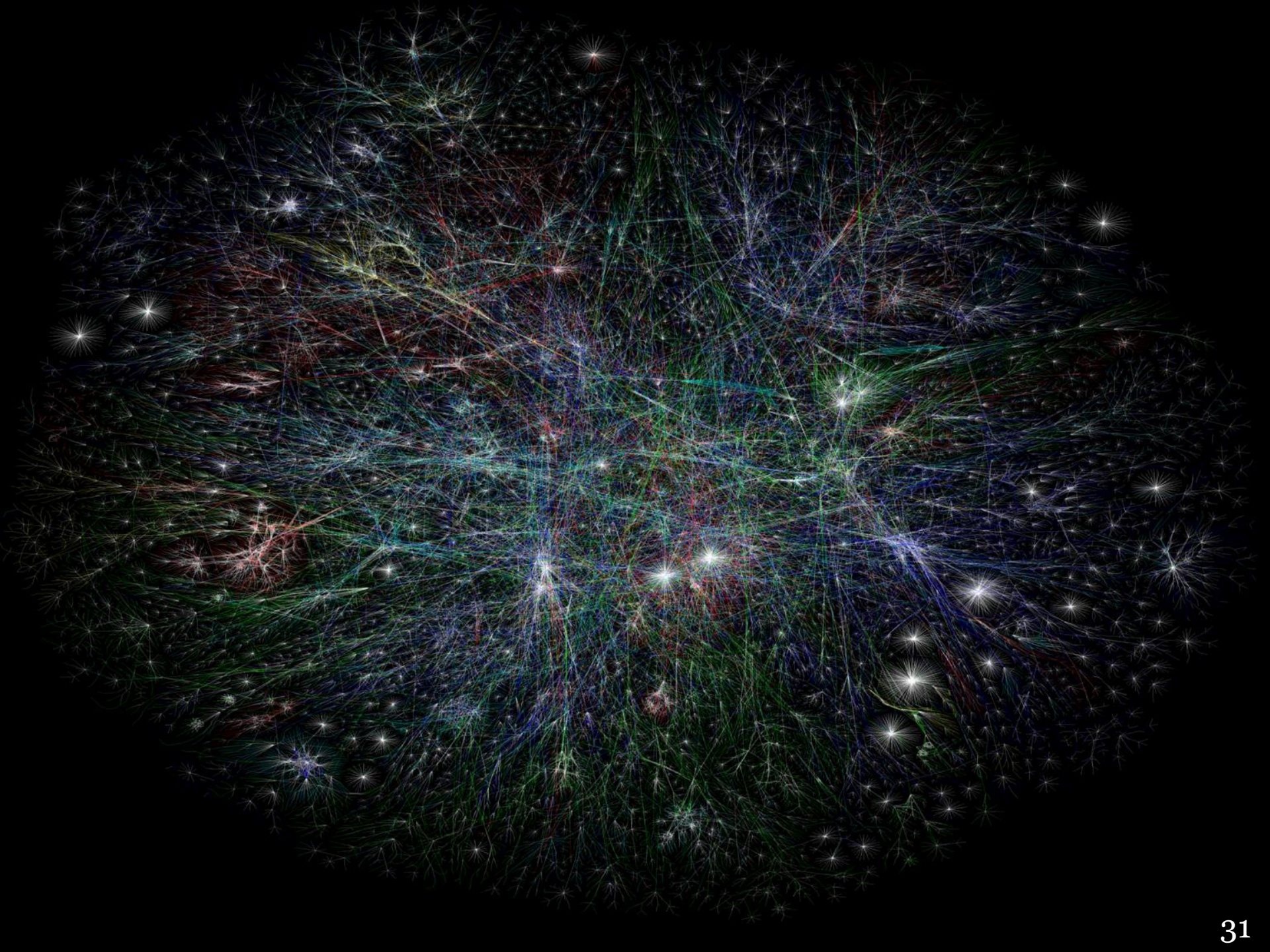


- Simple Structure
- Following
  - To subscribe to other people's posts



- Simple Structure
- Following
  - To subscribe to other people's posts

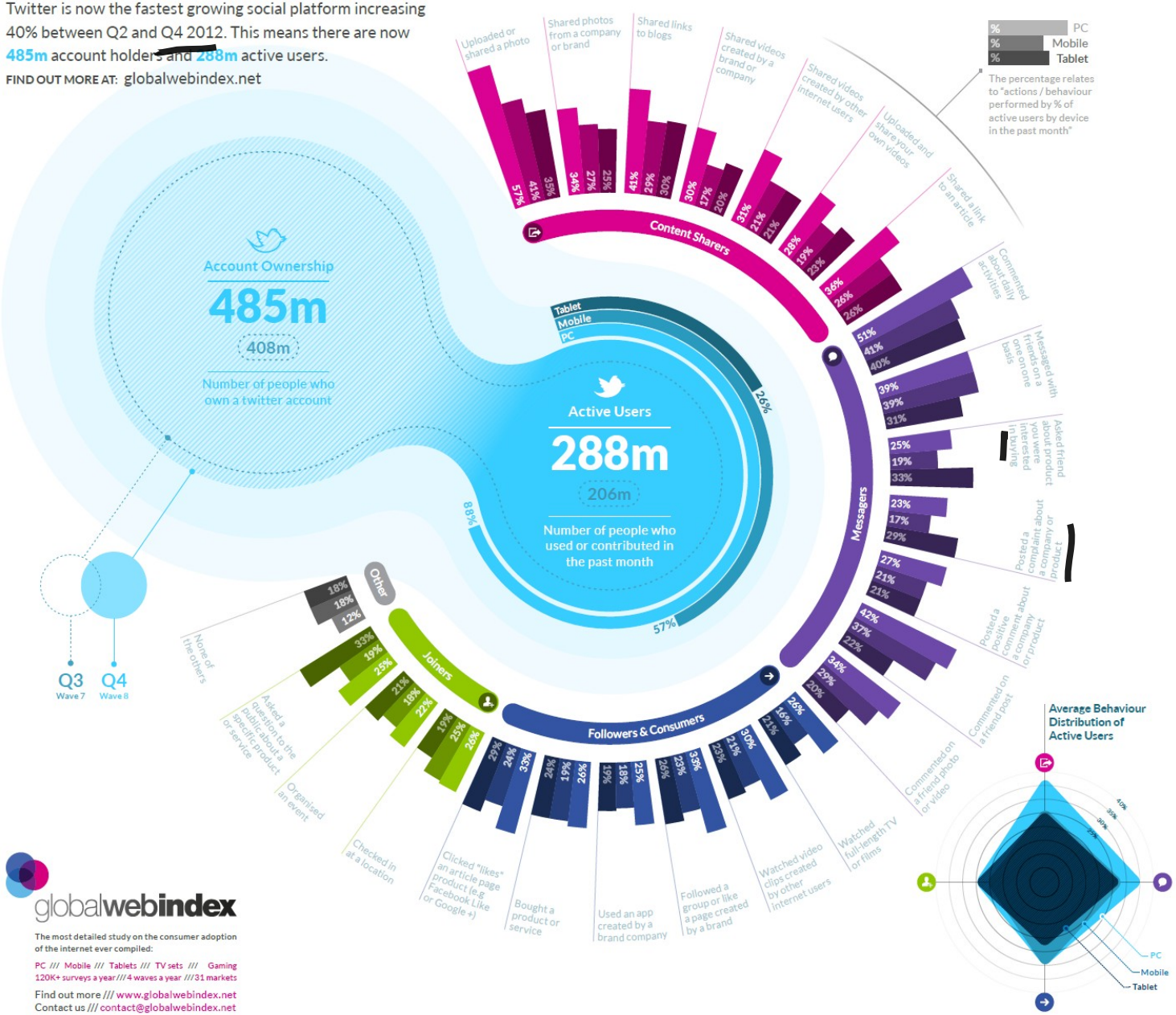




TWITTER The Fastest Growing Social Platform

Twitter is now the fastest growing social platform increasing 40% between Q2 and Q4 2012. This means there are now **485m** account holders and **288m** active users.

FIND OUT MORE AT: globalwebindex.net

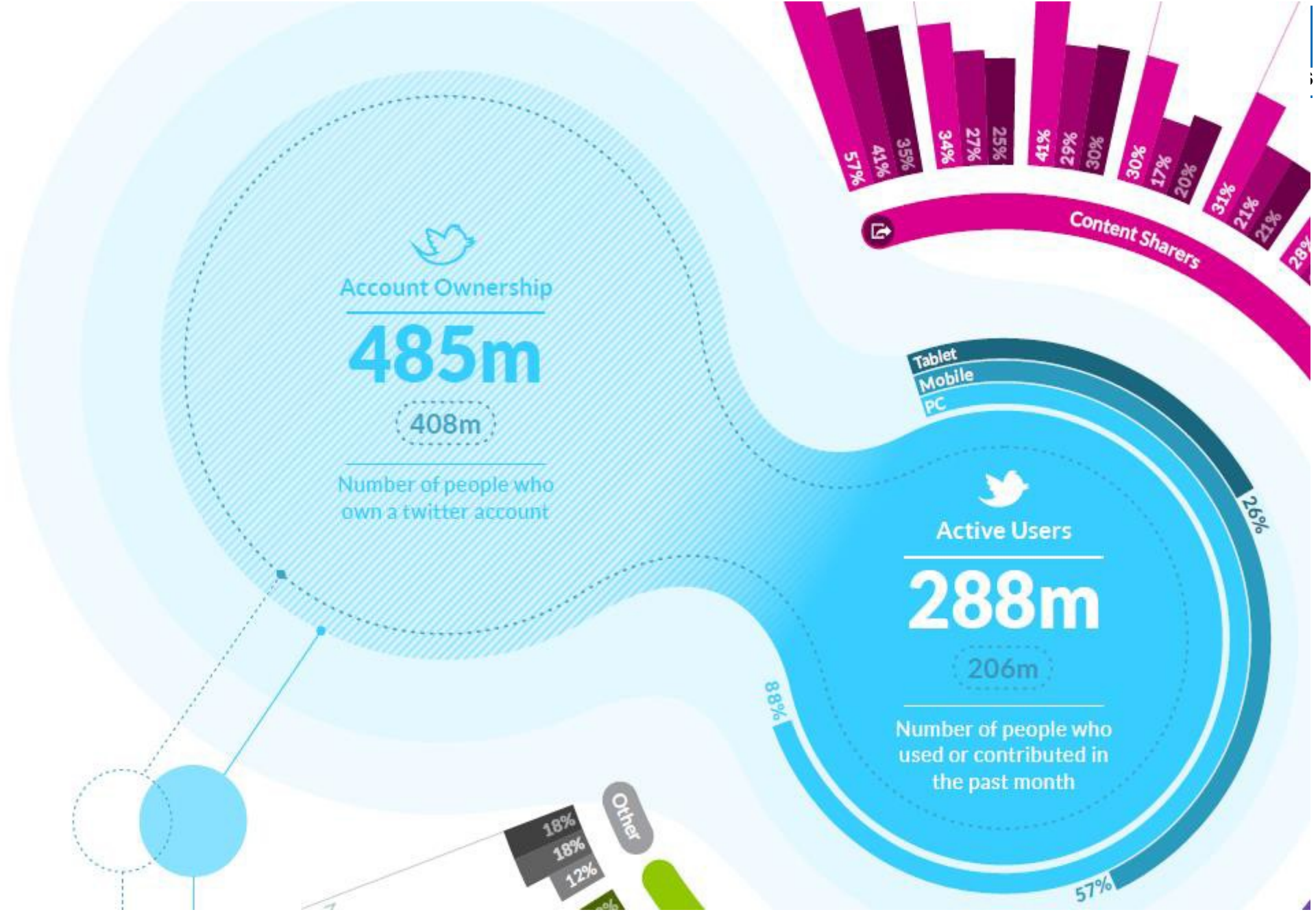


globalwebindex

The most detailed study on the consumer adoption of the internet ever compiled:

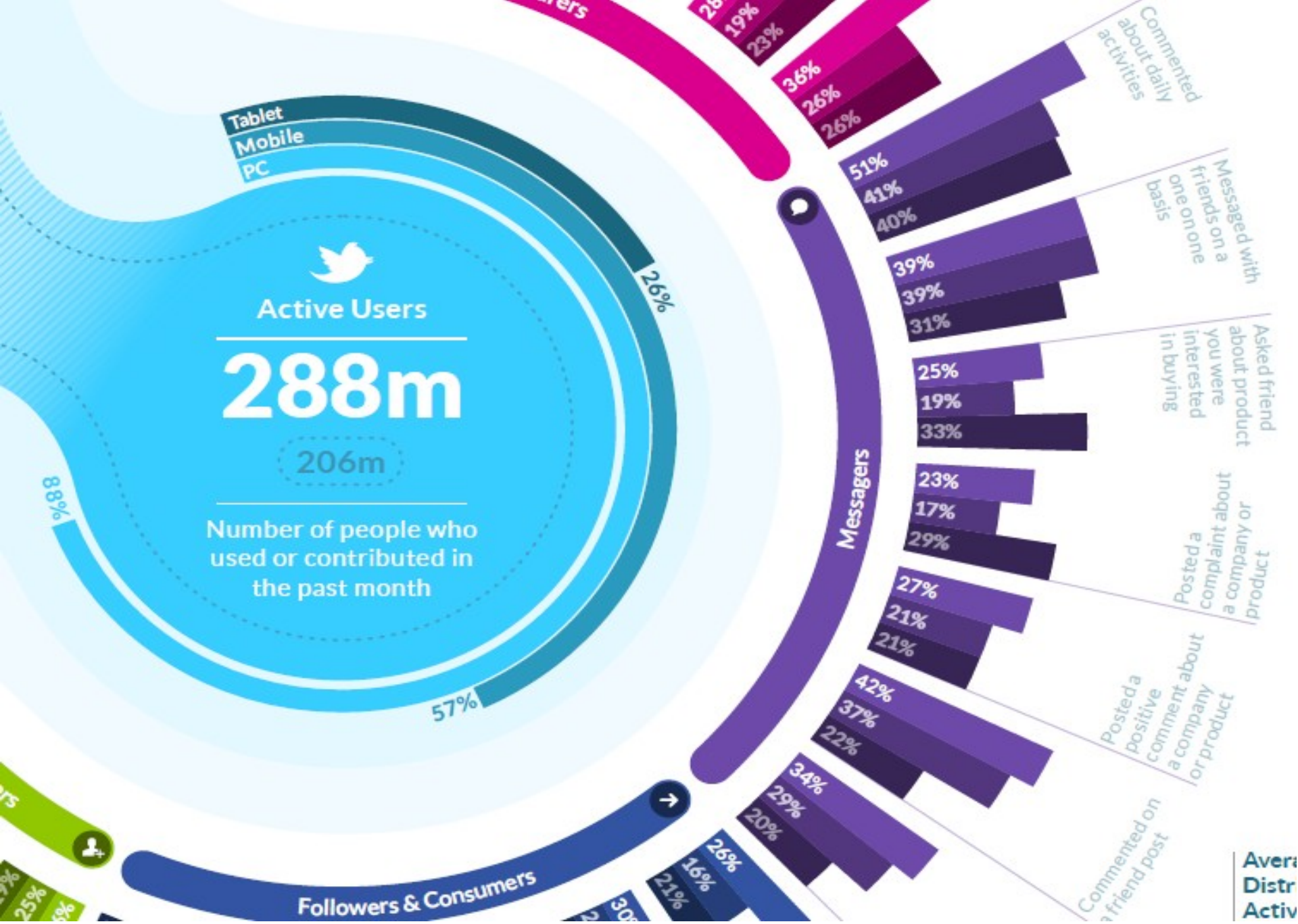
PC /// Mobile /// Tablets /// TV sets /// Gaming
120K+ surveys a year /// 4 waves a year /// 31 markets

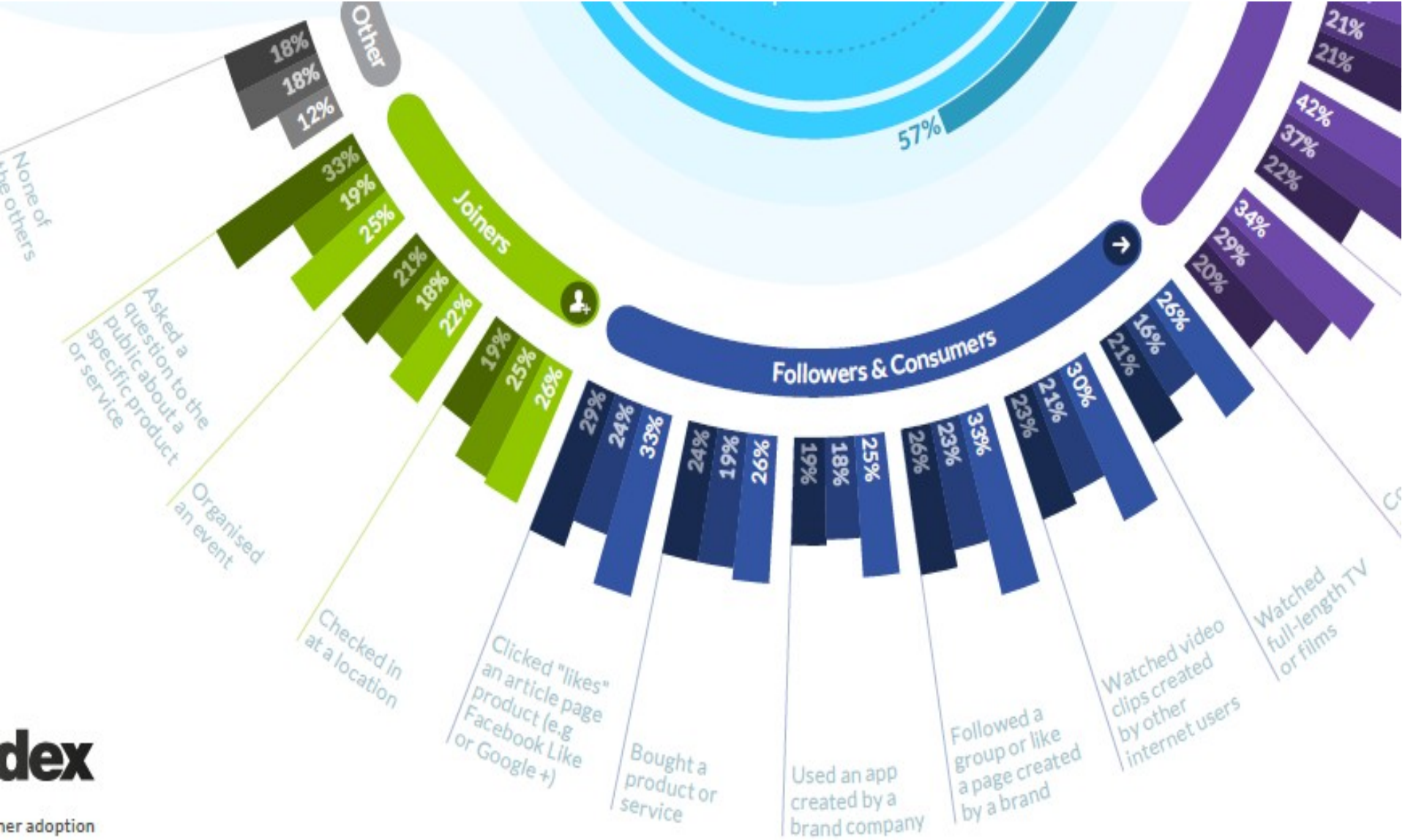
Find out more /// www.globalwebindex.net
Contact us /// contact@globalwebindex.net


Account Ownership
485m
408m
Number of people who own a twitter account
Active Users
288m
206m
Number of people who used or contributed in the past month
Tablet
Mobile
PC
26%
88%
57%
Content Sharers
57%
41%
35%
34%
27%
25%
41%
29%
30%
30%
17%
20%
31%
21%
21%
Other
18%
18%
12%

al Platform


Uploaded or shared a photo
57%
41%
35%
Shared photos from a company or brand
34%
27%
25%
Shared links to blogs
41%
29%
30%
Shared videos created by a brand or company
30%
17%
20%
Shared videos created by other internet users
31%
21%
21%
Uploaded and share your own videos
28%
19%
23%
Shared a link to an article
36%
26%
26%
Commented about daily activities
51%
Content Sharers
Tablet
Mobile
PC
% PC
% Mobile
% Tablet
The percentage relates to "actions / behaviour performed by % of active users by device in the past month"


Tablet
Mobile
PC
26%
Active Users
288m
206m
Number of people who used or contributed in the past month
88%
57%
Messagers
Followers & Consumers
Commented about daily activities
36%
26%
26%
Messaged with friends on a one on one basis
51%
41%
40%
Asked friend about product you were interested in buying
39%
39%
31%
Posted a complaint about a company or product
25%
19%
33%
Posted a positive comment about a company or product
23%
17%
29%
Commented on a friend post
27%
21%
21%
42%
37%
22%
34%
29%
20%
26%
16%
21%
19%
23%



dex

her adoption

Twitter User Activity

33% asked friend's **opinion** about a product

26% **bought** a product or service

29% **complained** about a brand or product

27% **recommended** a brand or product

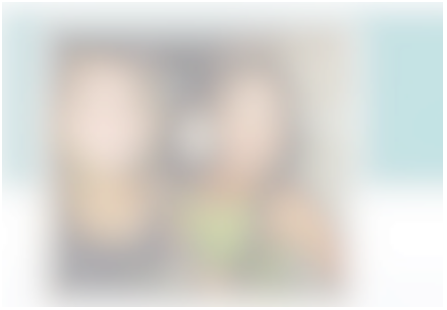
33% **followed** a group created by a brand

25% **used an app** created by a brand

30% **shared photos** about a brand

30% **shared video** created by a brand

twitter



...

...

Joined July 2010

TWEETS	PHOTOS/VIDEOS	FOLLOWING	FOLLOWERS	FAVORITES
477K	215	600	1,219	368


```
▼ object {21}
    created_at : Thu May 01 18:01:19 +0000 2014
    id : 461928366862376960
    id_str : 461928366862376960
    text : Debating if I should switch services with my family or if I should just stay on my
           own because I reallyyyyy don't want to leave Verizon..
    truncated : false
    in_reply_to_status_id : null
    in_reply_to_status_id_str : null
    in_reply_to_user_id : null
    in_reply_to_user_id_str : null
    in_reply_to_screen_name : null
  ▶ user {40}
    geo : null
    coordinates : null
    place : null
    contributors : null
    retweet_count : 0
    favorite_count : 0
  ▼ entities {4}
    ▶ hashtags [0]
    ▶ symbols [0]
    ▶ urls [0]
    ▶ user_mentions [0]
    favorited : false
    retweeted : false
    lang : en
```

Characteristics

- Very dynamic network structure:
  - Network relations are always changing.

- Content:
  - high prevelance of user-generated/urban words,
  - often short, context-less, and very noisy, and
  - of streaming type!

What Do We Learn?

- Data Crawling and Processing
- Social Relations and their related phenomena
- Power Laws & Popularity
- Information Diffusion & Cascading
- Noisy Text Processing and Insight Extraction
- Applications is language, healthcare, search & factuality, etc.

What Do We Learn? Cnt.

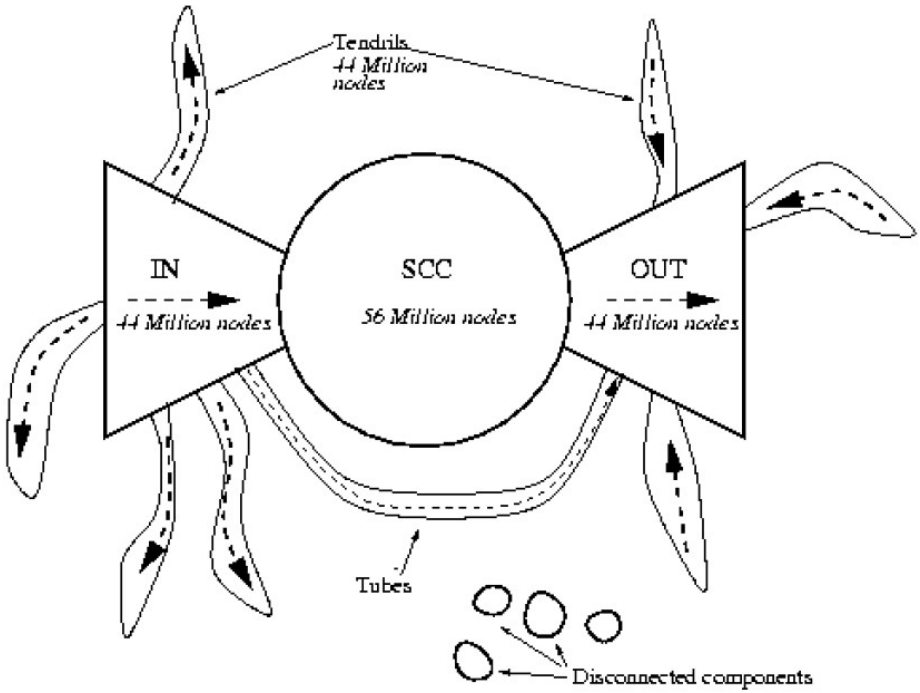
- The Structure of the Web
  - The Web contains a giant Strongly Connected Component

IN nodes:
can reach SCC but cannot be reached from it.

OUT nodes:
can be reached from SCC but cannot reach it.

Tendrils nodes:
(a) reachable from IN but cannot reach SCC,
(b) can reach OUT but cannot be reached from SCC.
Tendrils nodes satisfying both (a) and (b), travel in "tube" from IN to OUT without touching SCC.

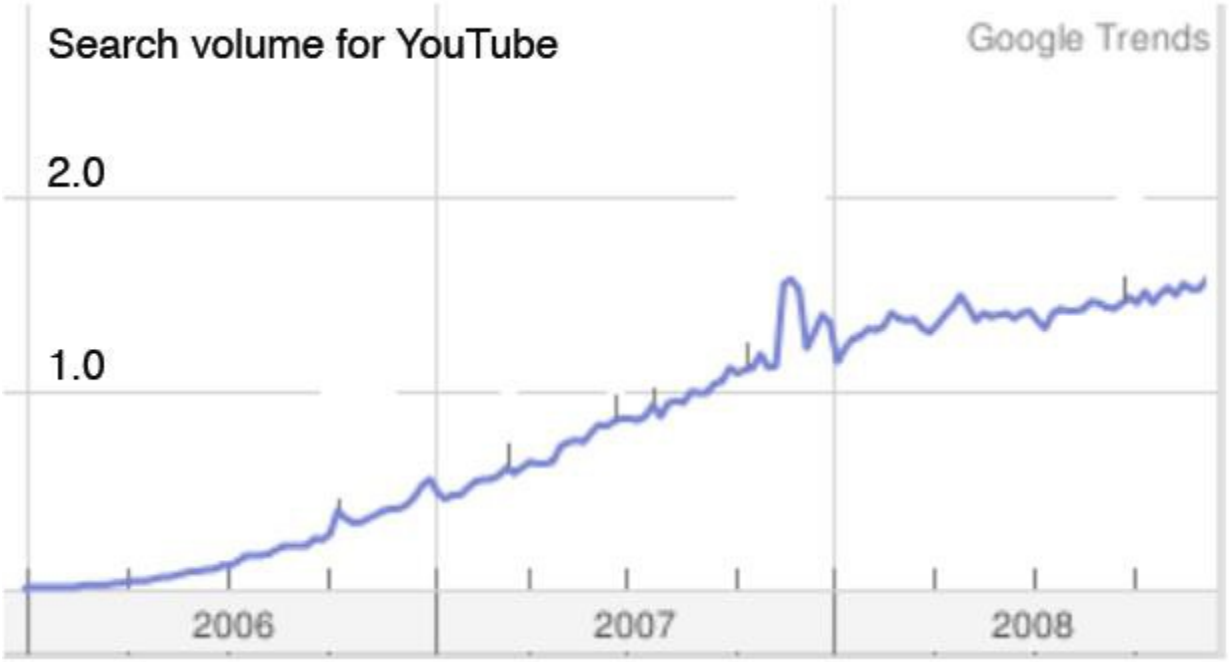
Disconnected nodes:
have no path to SCC ignoring directions



99.91% of individuals on FB belong to a single giant connected component

What Do We Learn? Cnt.

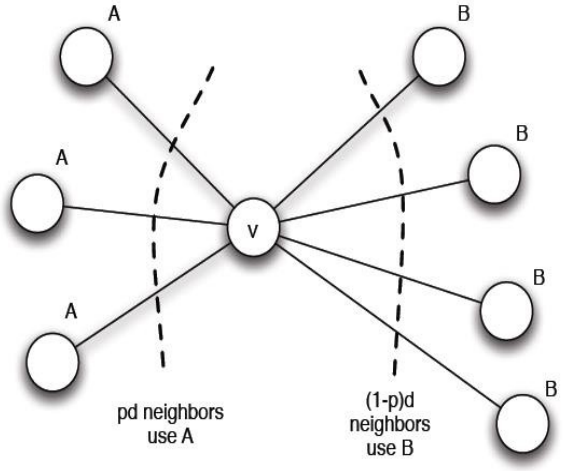
- Popularity in Networks: Rich Get Richer



Is it that the rich always get richer? new ideas always get attention and become viral?

What Do We Learn? Cnt.

- Information Cascading
  - Let's say you're at a dance class!
  - Some good-looking guy asks the woman next to you to dance.
    - She says **NO**.
  - He then asks another woman next to you to dance.
    - She says **NO**.
  - Now he asks you to dance. You say ???



What Do We Learn? Cnt.

- Small World Phenomenon

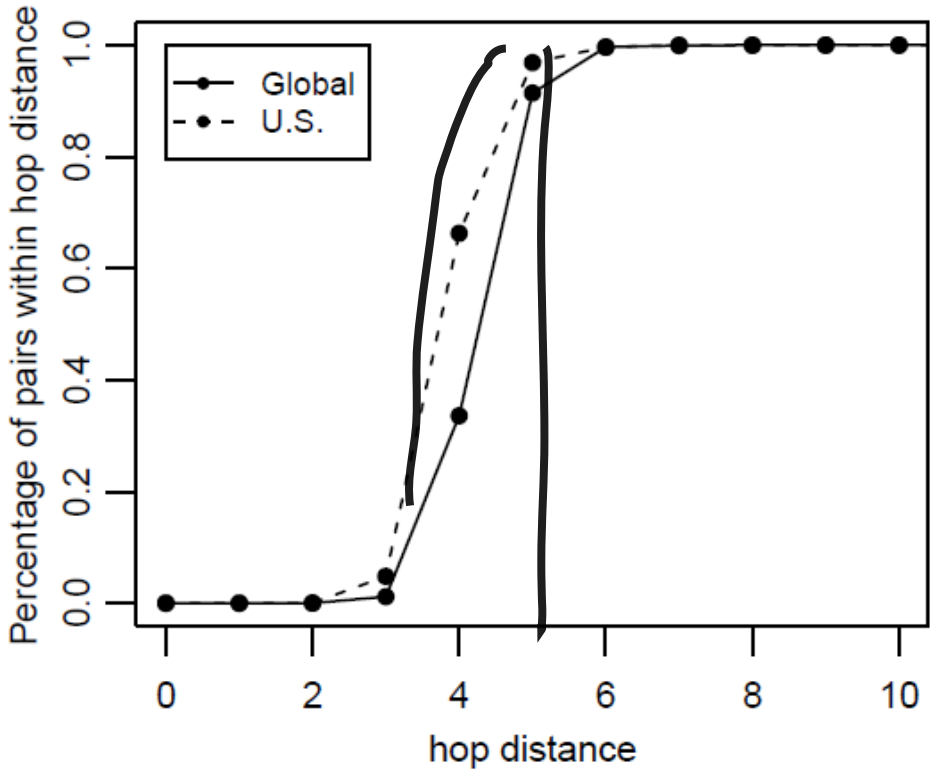


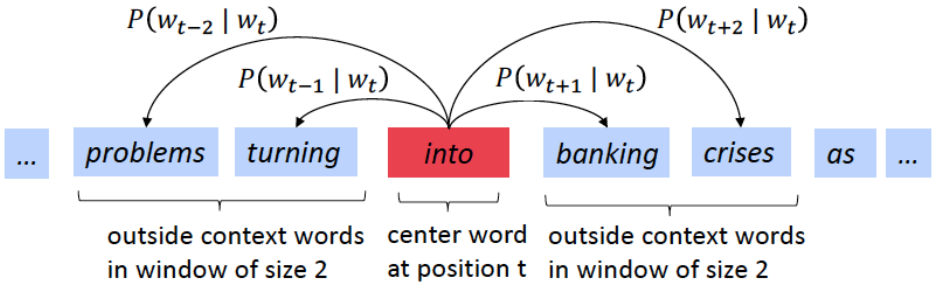
Figure 2. Diameter. The neighborhood function $N(h)$ showing the percentage of user pairs that are within h hops of each other. The average distance between users on Facebook in May 2011 was 4.7, while the average distance within the U.S. at the same time was 4.3.

Global
92.0%: within 5 degrees,
99.6%: within six degrees.

U.S. only
96.0%: within 5 degrees,
99.7%: within six degrees.

What Do We Learn? Cnt.

- Noisy Text Processing



$$P(o|c) = \frac{\exp(u_o^T v_c)}{\sum_{w \in V} \exp(u_w^T v_c)}$$

- Update vectors so you can predict well

Nearest words to frog:

1. frogs
2. toad
3. litoria
4. leptodactylidae
5. rana
6. lizard
7. eleutherodactylus

litoria

leptodactylidae

rana

eleutherodactylus

What Do We Learn? Cnt.

- Applications (*mainly given guest lectures*)
  - Language Analysis
  - Health Informatics
  - Search and Factuality
  - Moment Retrieval

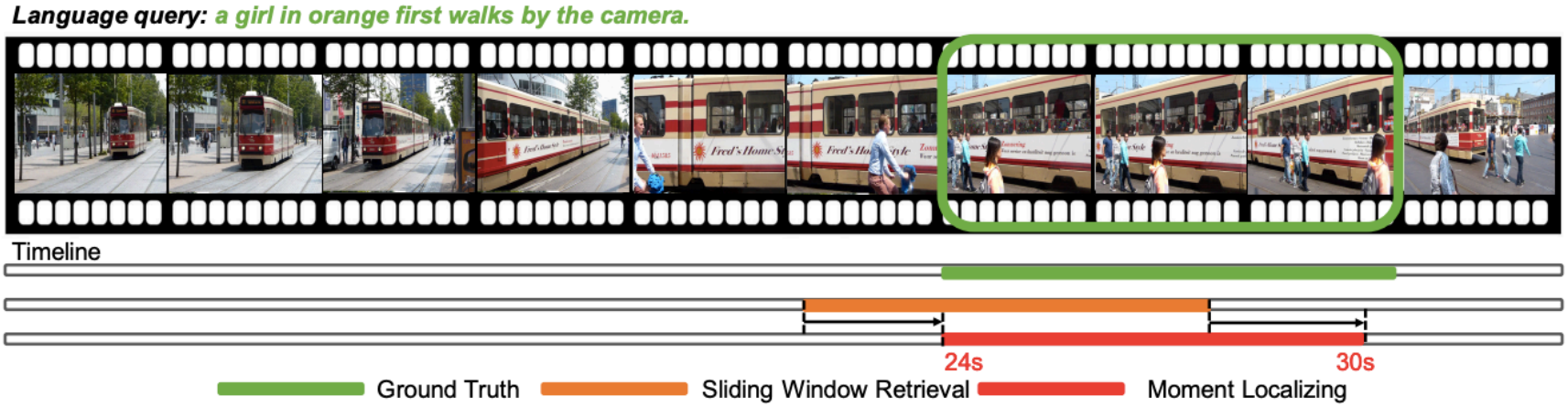
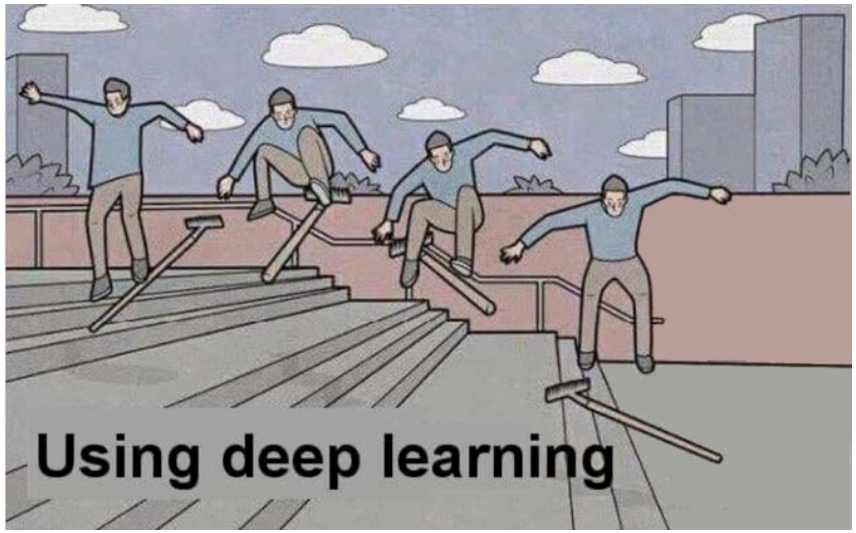
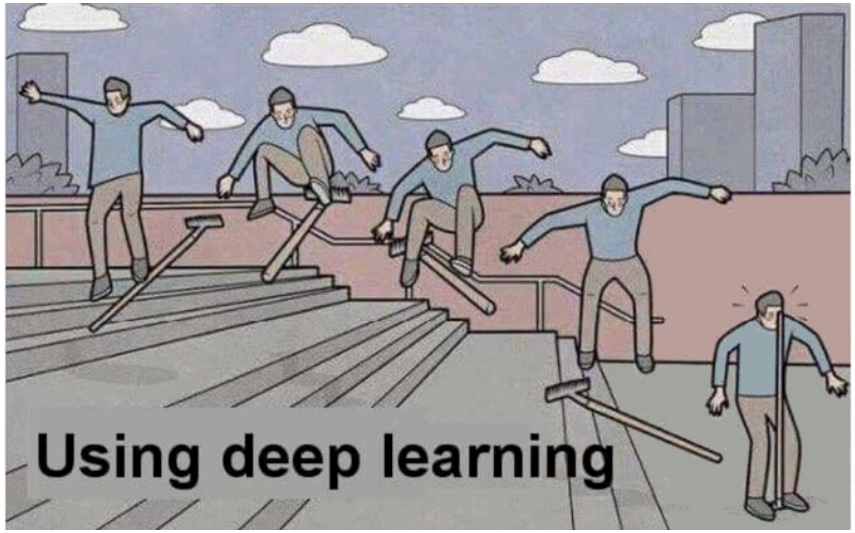


Figure 1: Temporal video moment localization is designed to localize a moment (the red bar) with a start point (24th s) and an end point (30th s) in the video according to the given language query. Here the green bar denotes the ground truth, the orange bar stands for the result of sliding window moment retrieval, and the red bar refers to the localizing result.

Techniques - Assumption



Techniques - Reality ¯_(ツ)_/¯



Reading

- Ch.01 Overview [NCM]

- Watch this 30 min TED talk by Deb Roy @ MIT:
  - From Gaga to Water: http://bit.ly/12fIOeR