

# Introduction to Social Media Computing

CMSC 498J: Social Media Computing

Department of Computer Science  
University of Maryland  
Spring 2015

Hadi Amiri  
[hadi@umd.edu](mailto:hadi@umd.edu)



# Lecturer

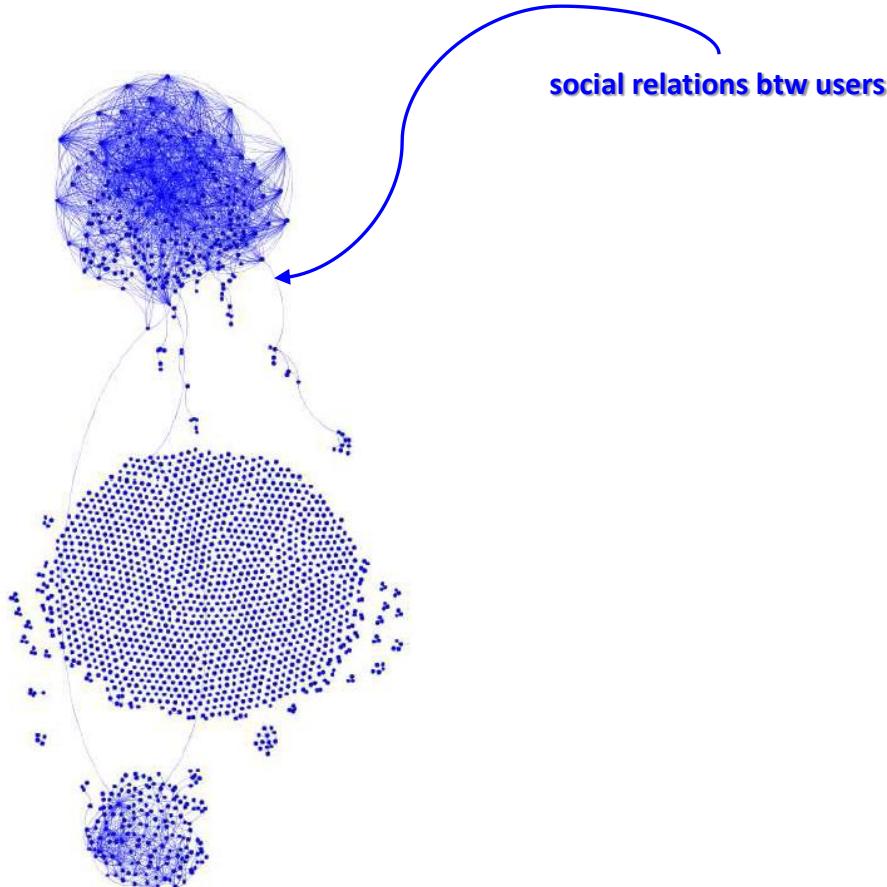
- Hadi Amiri
  - [hadi@umd.edu](mailto:hadi@umd.edu)
  - [www.umiacs.umd.edu/~hadi/](http://www.umiacs.umd.edu/~hadi/)
  - Office: AVW 3161
    - Always get lost in AVW?
      - <http://eit.umd.edu/maps/avwilliams>
      - [www.umiacs.umd.edu/~hadi/contact.html](http://www.umiacs.umd.edu/~hadi/contact.html)
  - Office hours: Mons 3:30-4:30
  - Hobbies: I enjoy running in marathons among other sports.



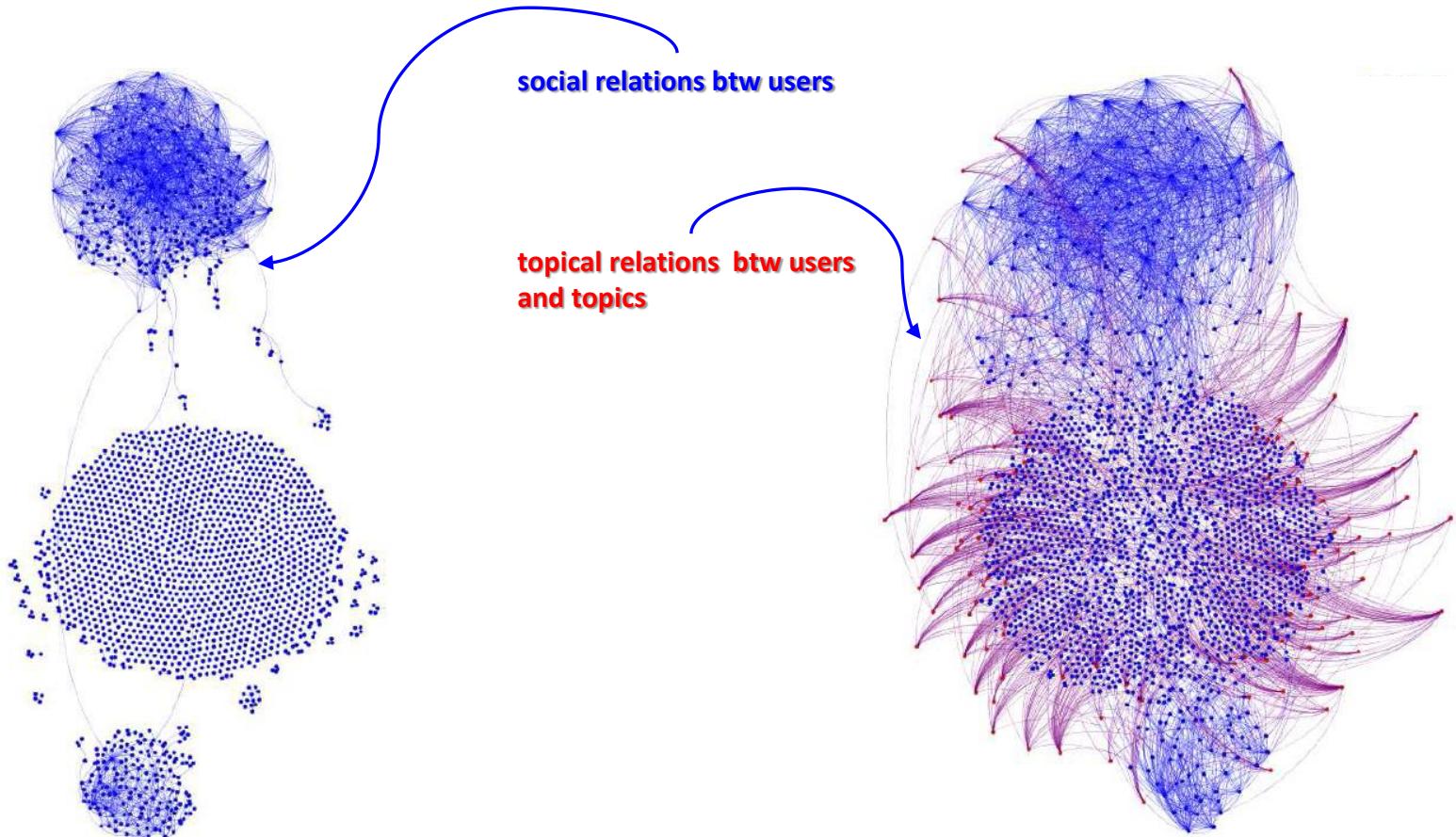
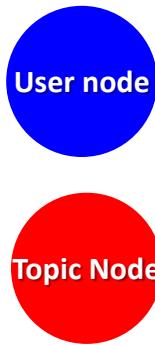
# What Is This Course about?

- Understanding various social phenomena through studying:
  - Networks
    - a pattern of inter-connections among a set of things!
    - deal with structure
  - Ways contents are perceived in them
    - deal with various user generated contents and their propagation in networks.
- We aim to understand networks, contents, and the interaction between the two.
  - Properties, design principles, and models!

# What Does that Mean?



# What Does that Mean? Cnt.





# What Are Social Networks?

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

# Let's See Some Samples!



# Sample 1.

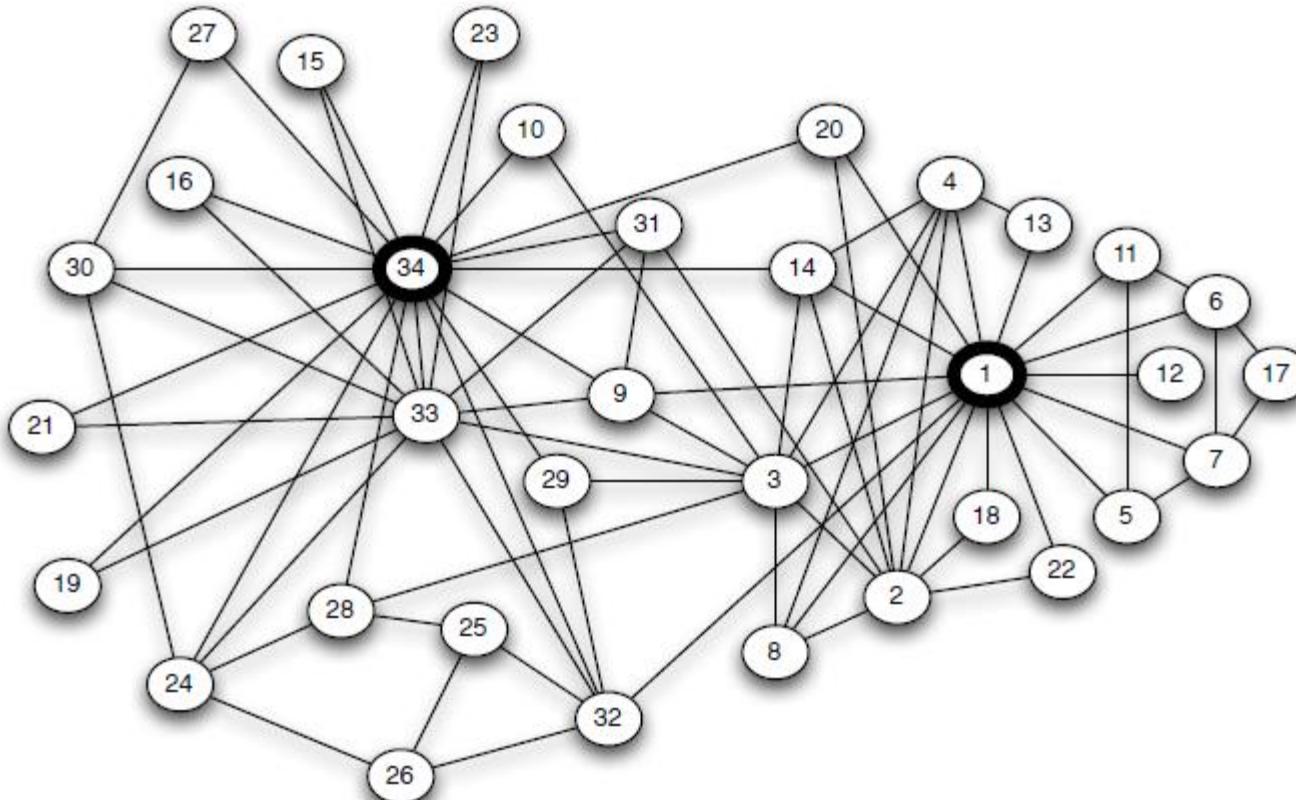


Figure 1.1: The social network of friendships within a 34-person karate club [421].

# Sample 2.

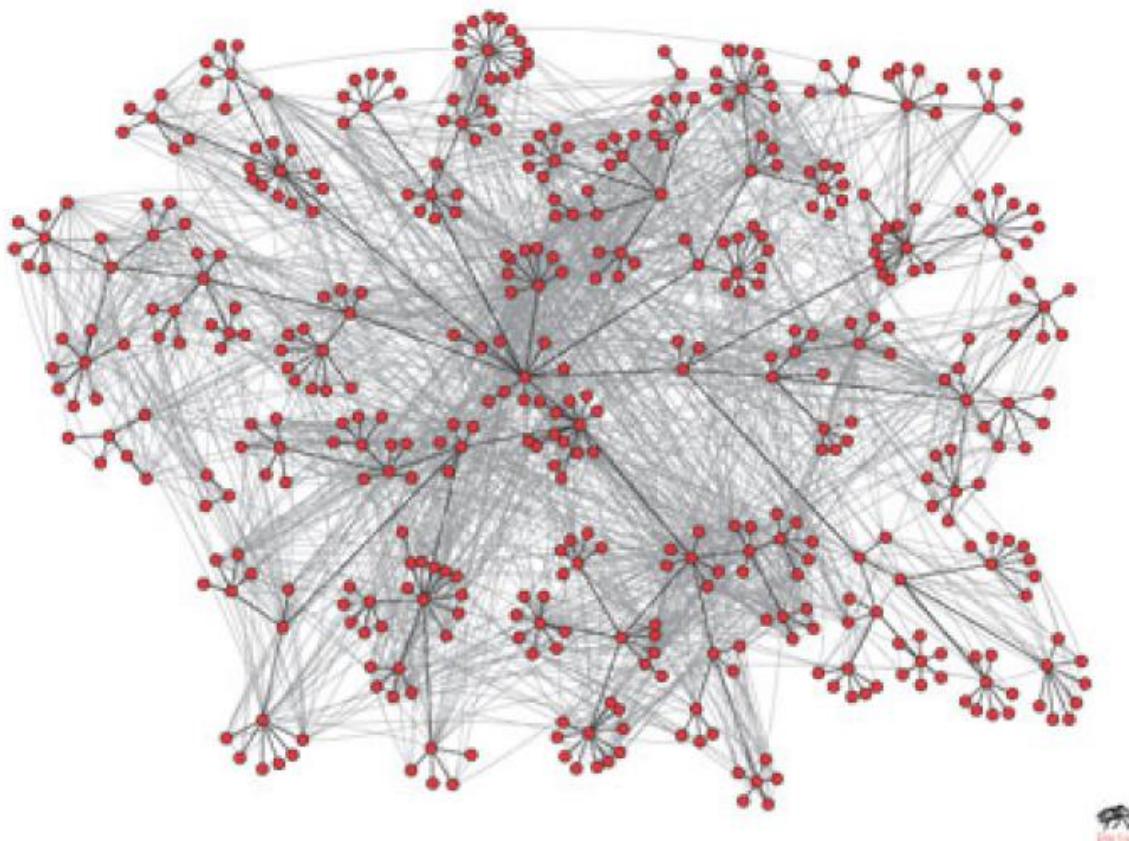


Figure 1.2: Social networks based on communication and interaction can also be constructed from the traces left by on-line data. In this case, the pattern of e-mail communication among 436 employees of Hewlett Packard Research Lab is superimposed on the official organizational hierarchy [6]. (Image from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

# Sample 3.

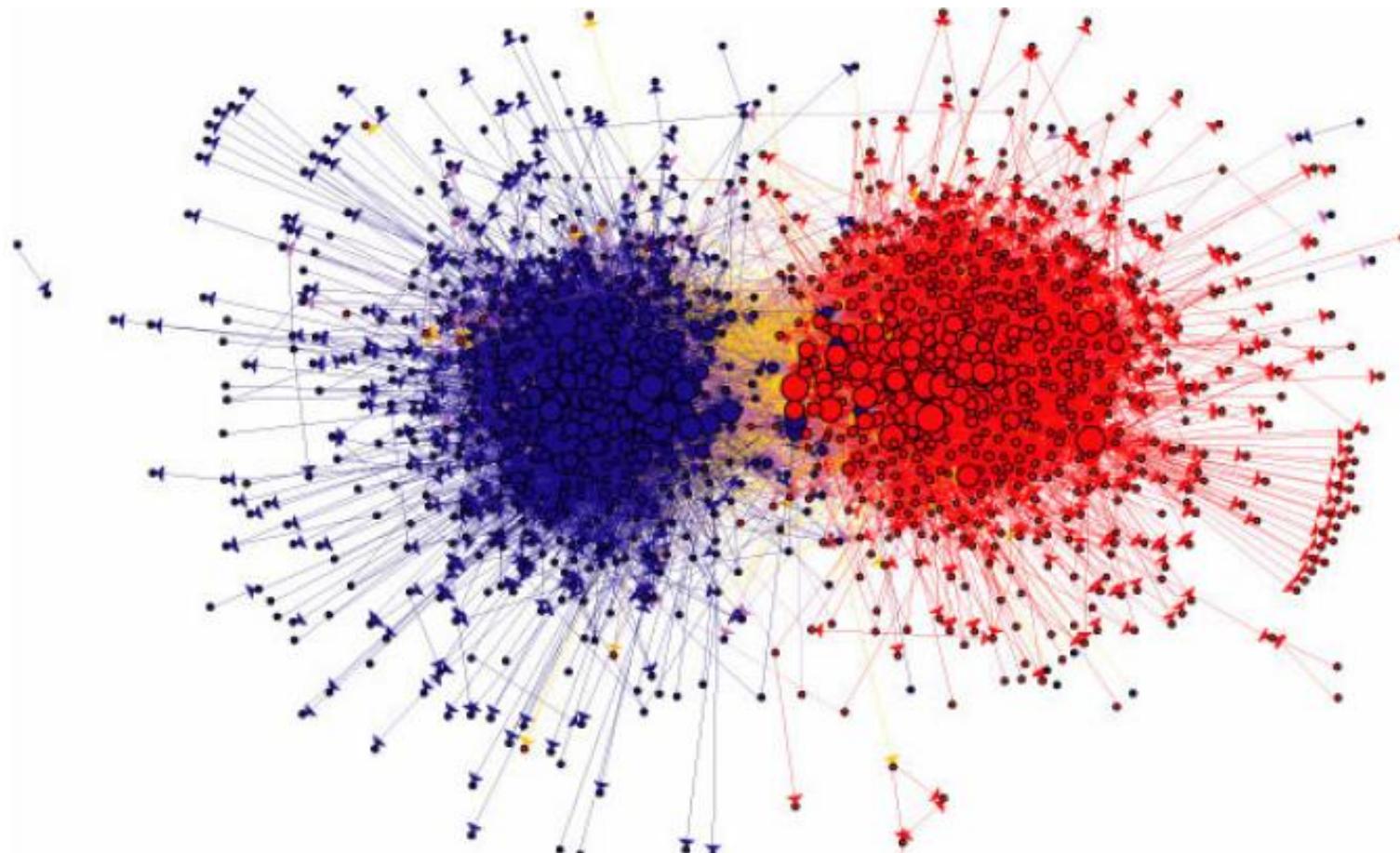


Figure 1.4: The links among Web pages can reveal densely-knit communities and prominent sites. In this case, the network structure of political blogs prior to the 2004 U.S. Presidential election reveals two natural and well-separated clusters [5]. (Image from <http://www-personal.umich.edu/~ladamic/img/politicalblogs.jpg>)

# Sample 4.

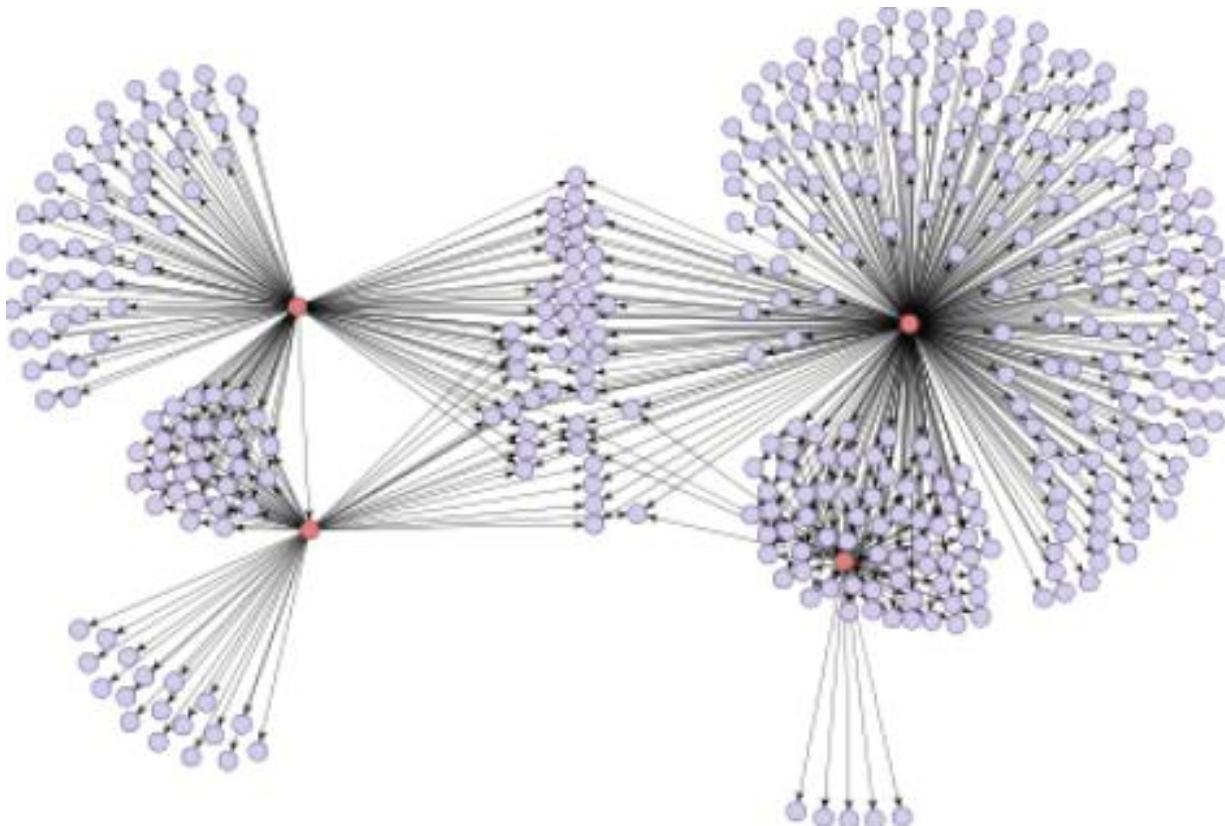


Figure 1.11: When people are influenced by the behaviors their neighbors in the network, the adoption of a new product or innovation can cascade through the network structure. Here, e-mail recommendations for a Japanese graphic novel spread in a kind of informational or social contagion. (Image from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

# Sample 5.

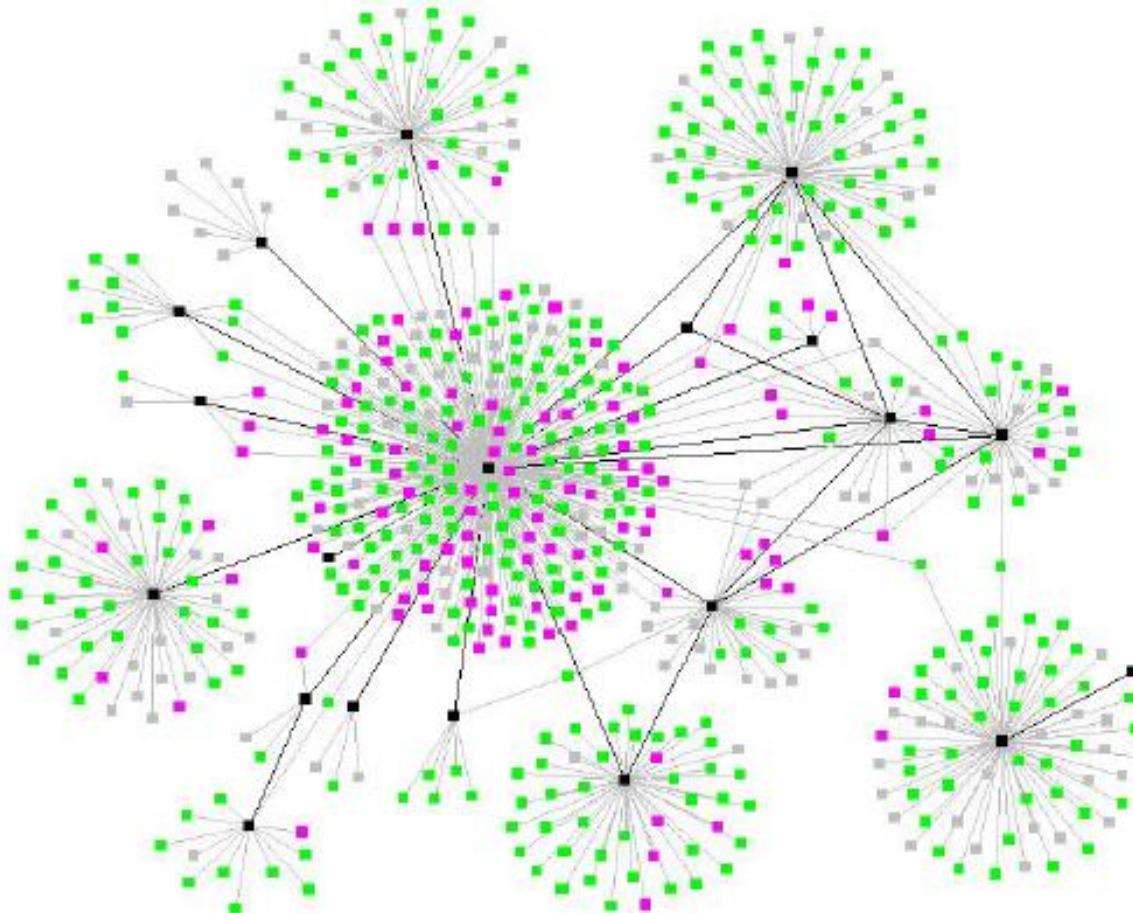
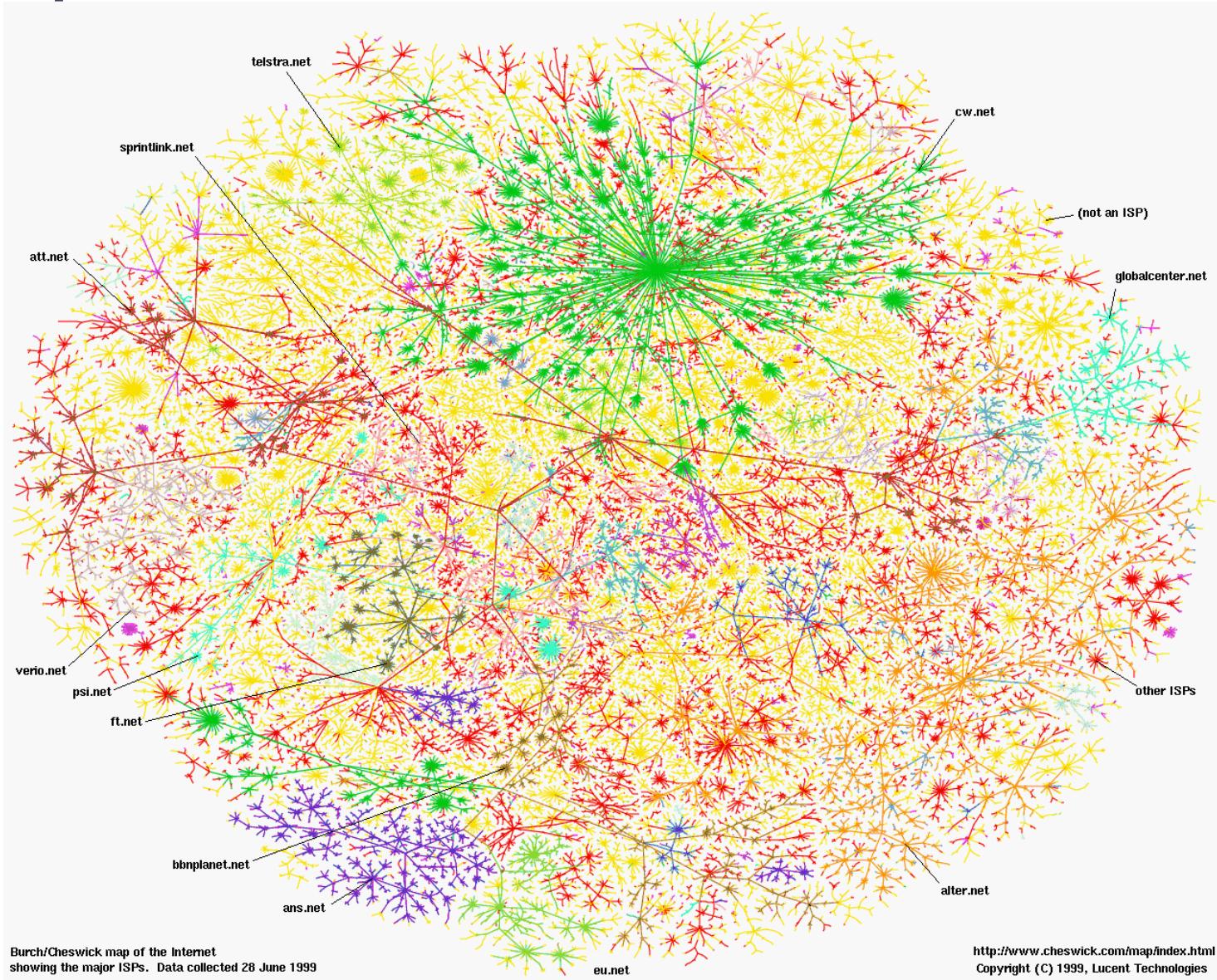


Figure 1.12: The spread of an epidemic disease (such as the tuberculosis outbreak shown here) is another form of cascading behavior in a network. The similarities and contrasts between biological and social contagion lead to interesting research questions. (Image from Andre et al. [16].)

# Sample 6.

Network of Major ISPs.  
1999



# Sample 7.

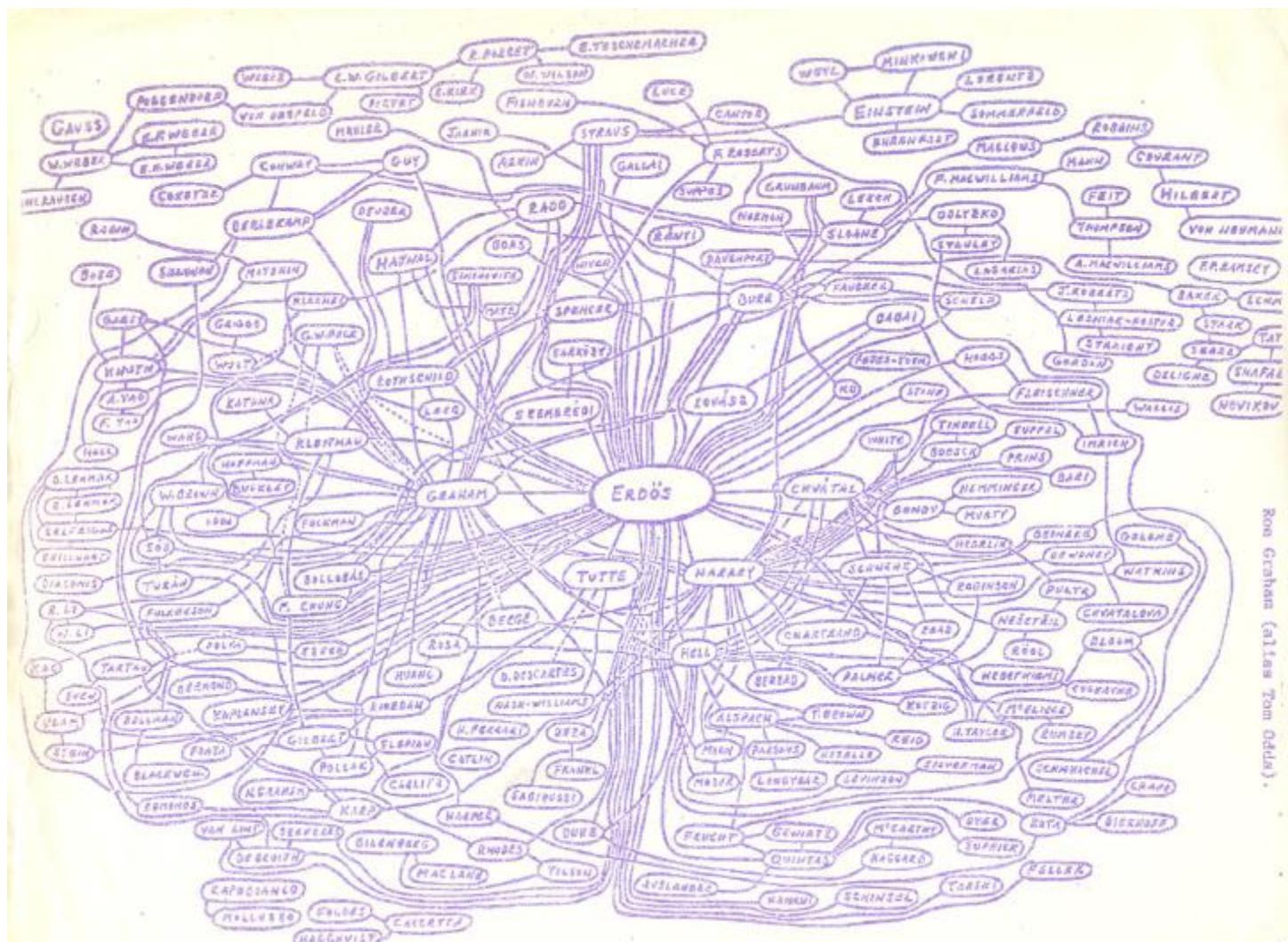


Figure 2.12: Ron Graham's hand-drawn picture of a part of the mathematics collaboration graph, centered on Paul Erdős [189]. (Image from <http://www.oakland.edu/enp/cgraph.jpg>)

# Sample 8.

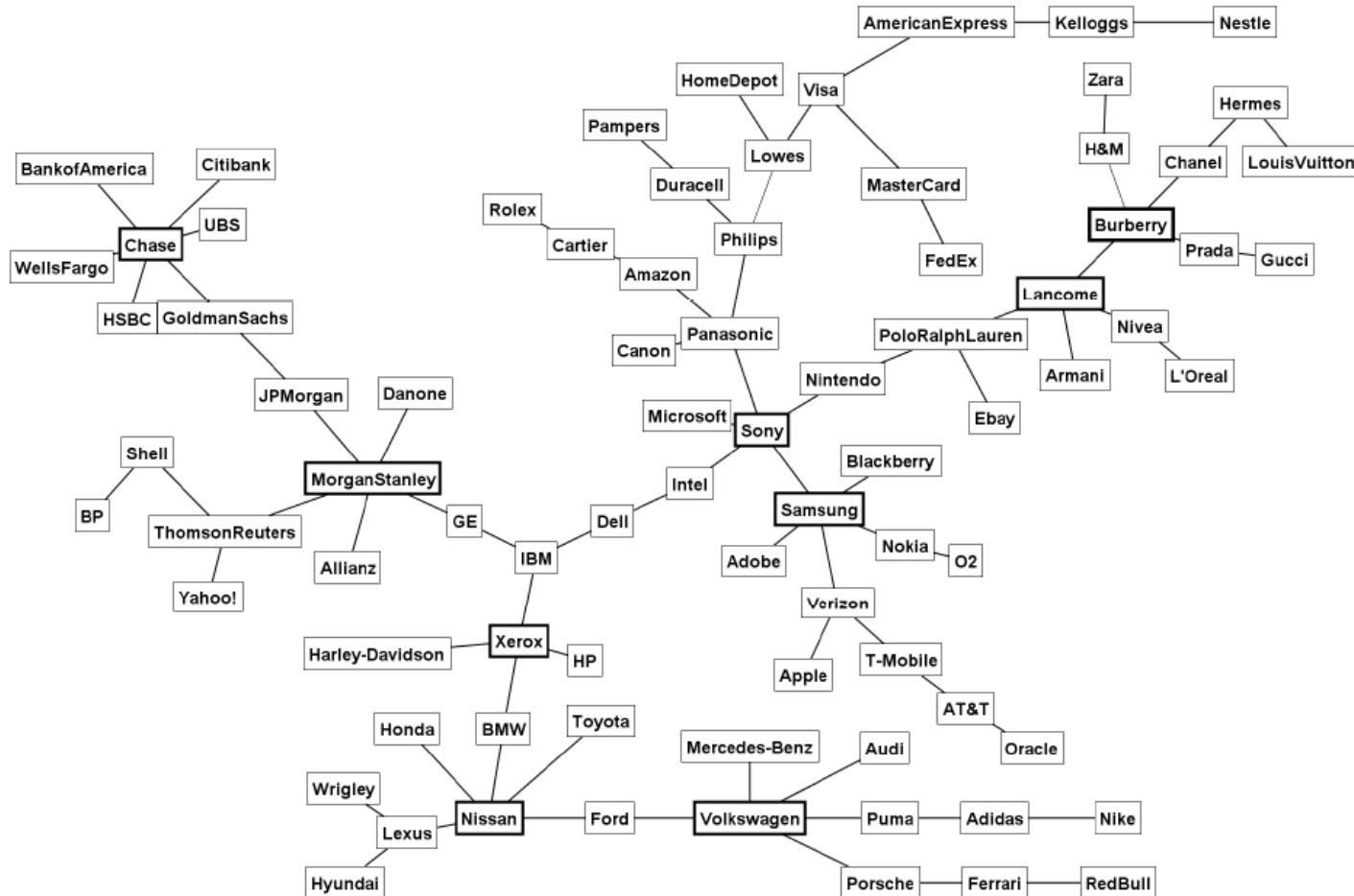


Figure 3. Minimum spanning tree (MST) of the most valued global brands. The MST of the brand network is the subset of edges that forms a tree reaching every brand such that the total length of all the edges is minimized. It is readily apparent that certain brands stand out prominently as hubs with connections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

# Sample 9.

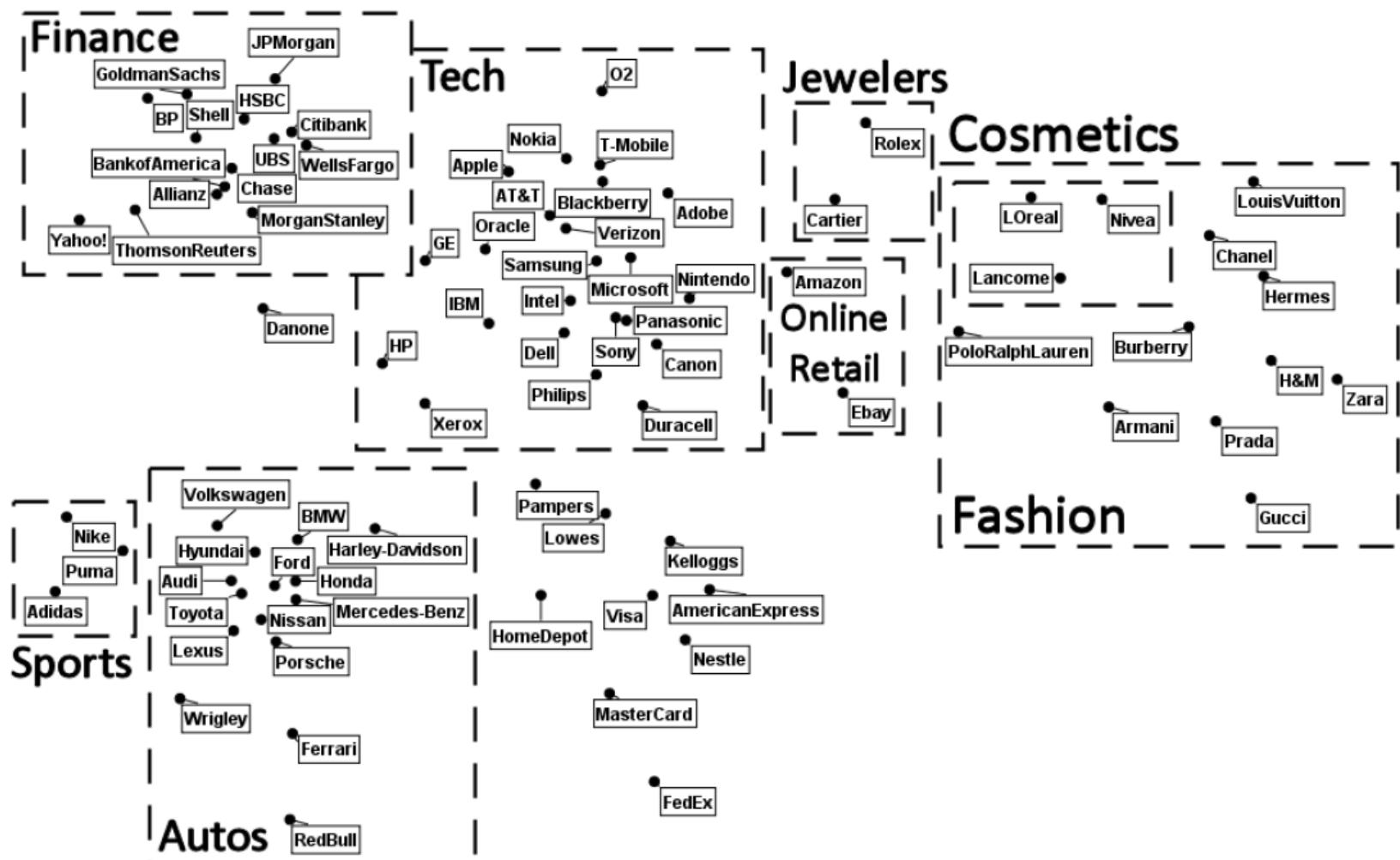
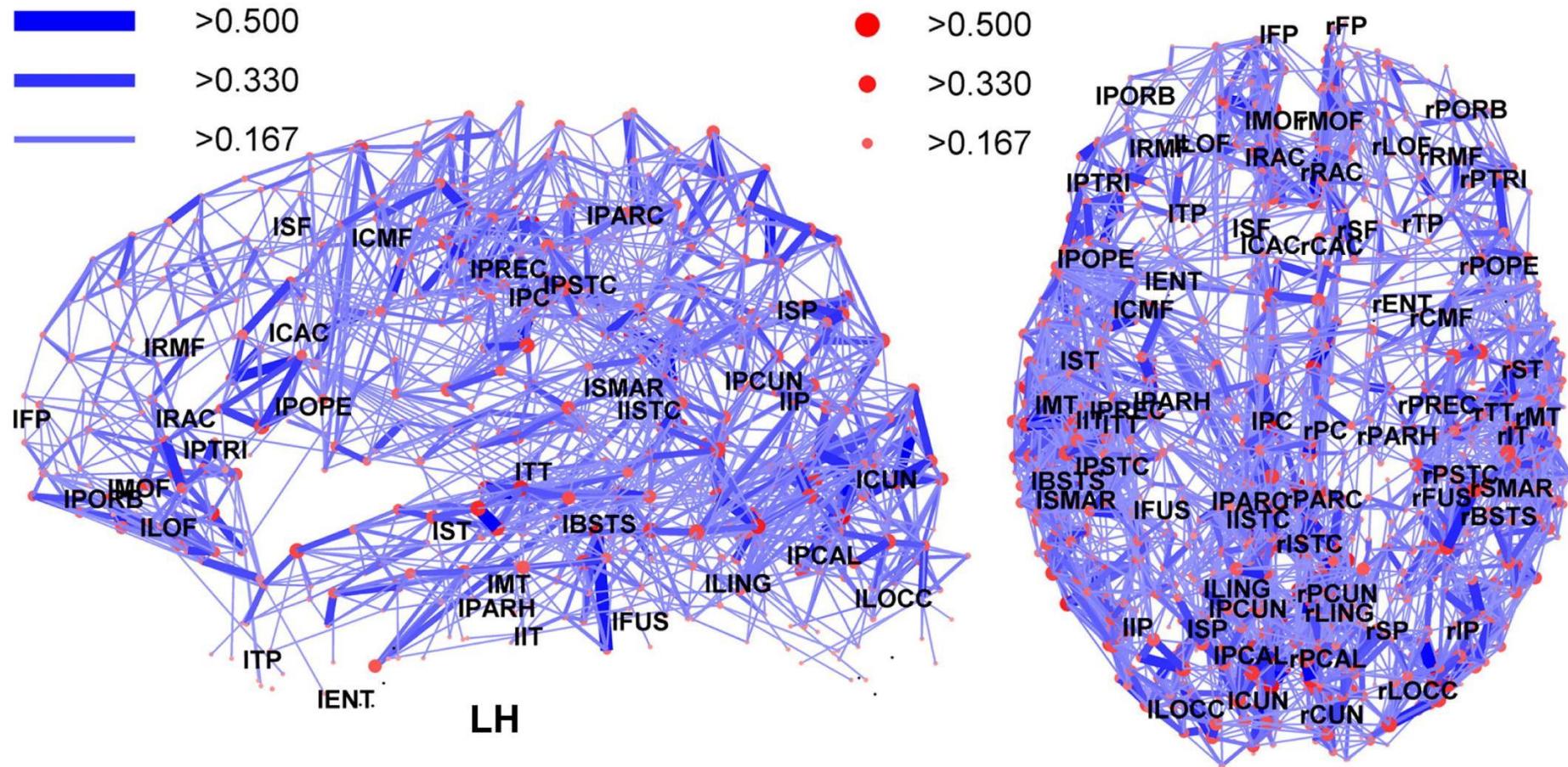


Figure 4. Map of brands. The minimum spanning tree augmented by triangulating each brand location from their nearest neighbors with forced-based layout yields a map high in face validity. Note the eight strong market category groupings outlined with broken lines.

# Sample 10.

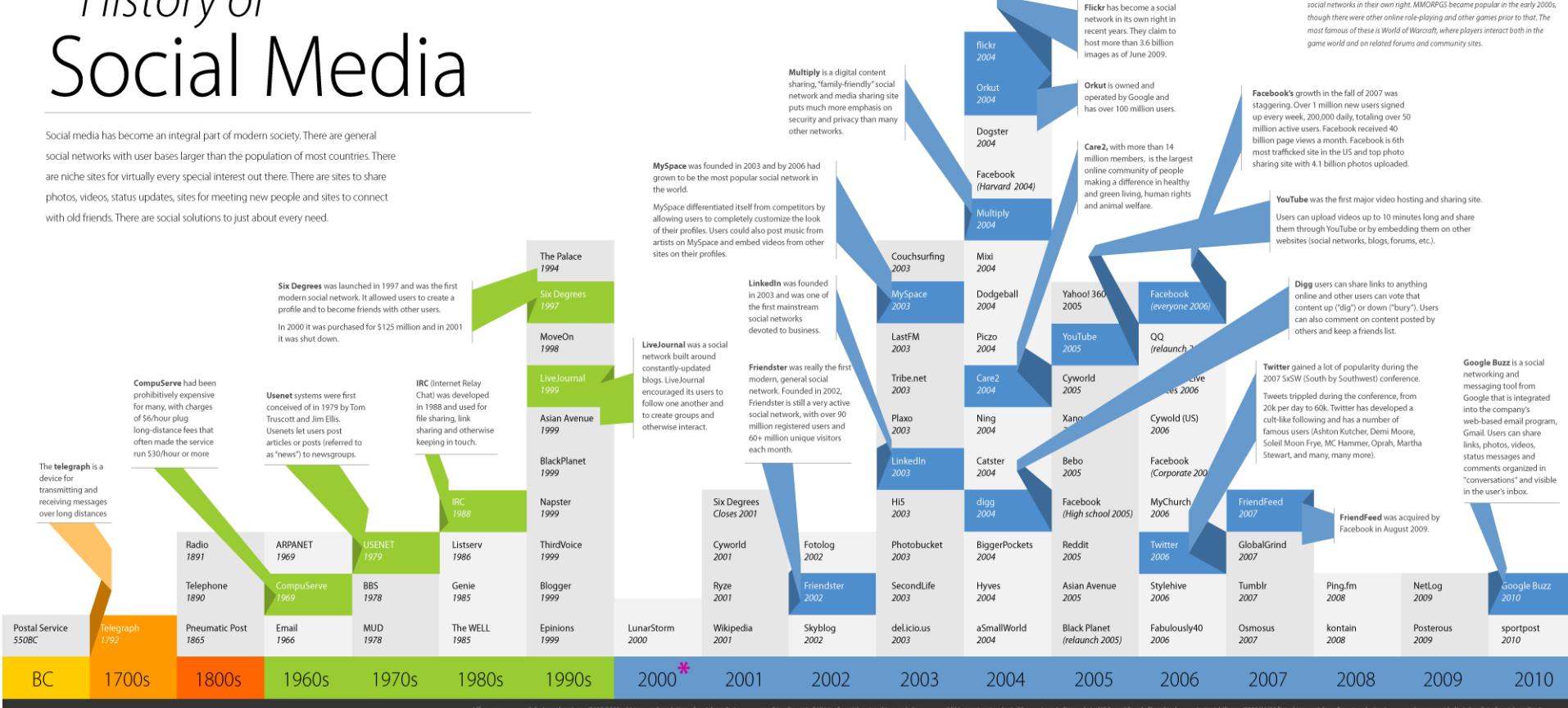


Network representation of brain connectivity: Dorsal and lateral views of the connectivity backbone of human brain. Labels indicating anatomical subregions are placed at their respective centers of mass. Nodes (individual ROIs) are coded according to strength and edges are coded according to connection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

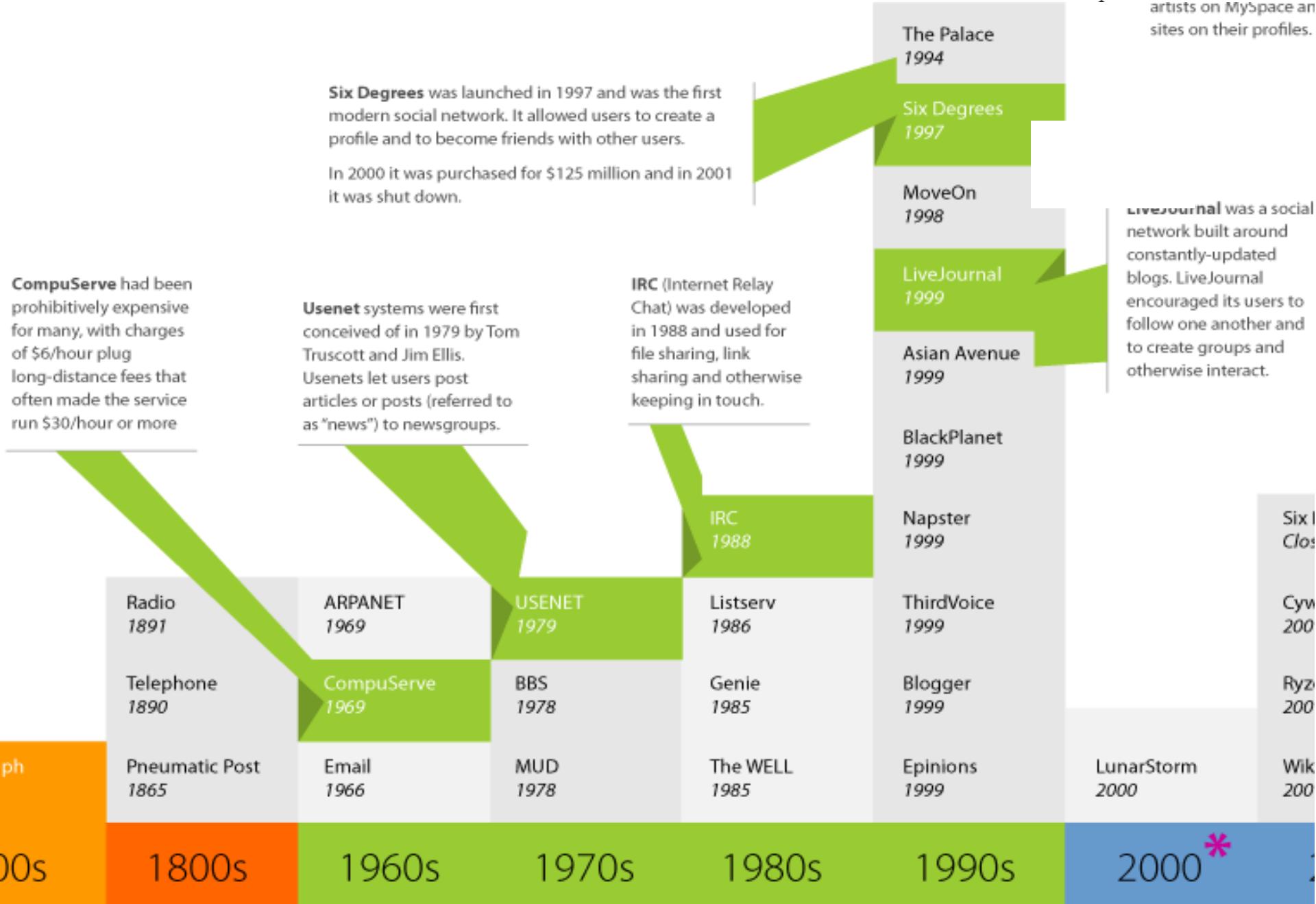
# How Long They've Been Around?

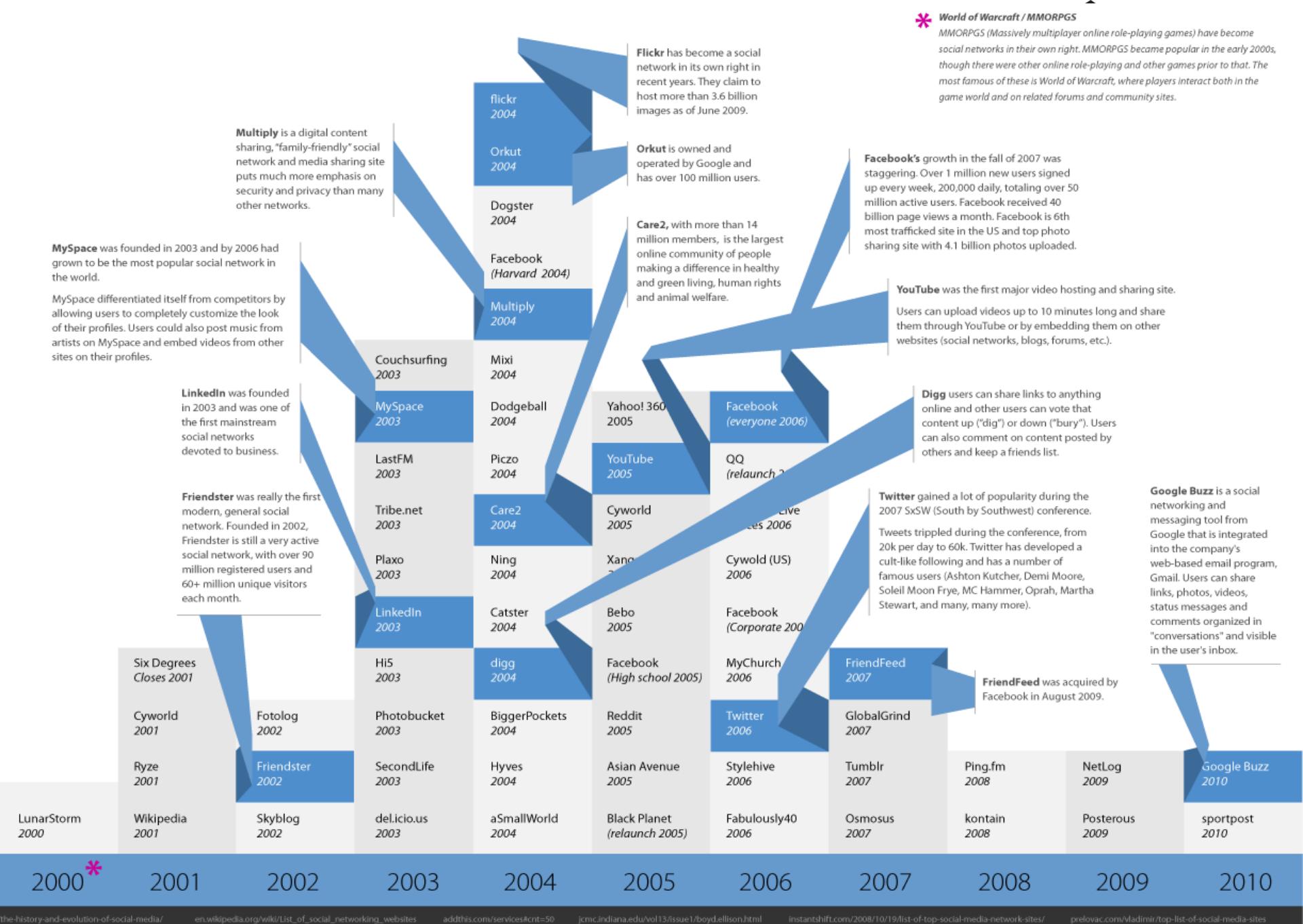
## History of Social Media

Social media has become an integral part of modern society. There are general social networks with user bases larger than the population of most countries. There are niche sites for virtually every special interest out there. There are sites to share photos, videos, status updates, sites for meeting new people and sites to connect with old friends. There are social solutions to just about every need.



artists on MySpace and sites on their profiles.





# Why Should We Study Them?

- Social Nets provide powerful ways of looking at complex data and systems:
  - Spread of news or diseases
  - Evolution of science
  - Structure of the Web
  - Markets & models of trades
- Networks help to understand if a principle holds across many settings and fields, and
- We have lots of them!

# Why Should We Study Them? Cnt.

- Computer Scientists
  - Algorithms and models
  - computational challenges

# Got something TO ASK US?

We're happy to help.



@VZWWIRELESS @VZNW35 @VZNSUPPORT @VZNSMALLBIZ @VERIZONLATINO

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](http://community.verizonwireless.com)

## McD Customer Service



### Celebrate the sweet life.



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](http://mcd.to/ULtdKh)



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/AACRI](http://bit.ly/AACRI)



TWEETS  
211K

FOLLOWING  
7,234

FOLLOWERS  
63K



[Follow](#)

## Telstra



@Telstra

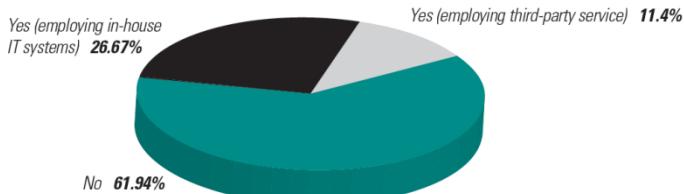
We're here 24x7 to provide customer support and answer any Telstra questions you might have. Last week our average response time was 20 minutes

Australia · [telstra.com.au](http://telstra.com.au)

# Why Should We Study Them? Cnt.

**Figure 25: Brand and Reputation Monitoring of SMNs**

Overall 465 respondents, LOB=107, EMEA=168



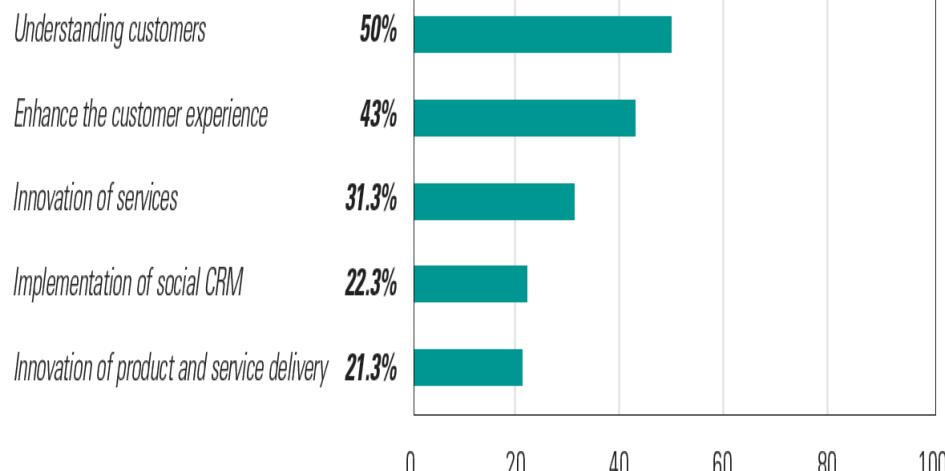
**Figure 27a: Organizational Plans to Leverage Social Media Metrics Into Business Processes**

Overall 459 respondents, LOB=107, EMEA=167



**Figure 33: Top Business Processes Leveraging Social Media Data**

Overall 300 respondents, LOB=79, EMEA=105

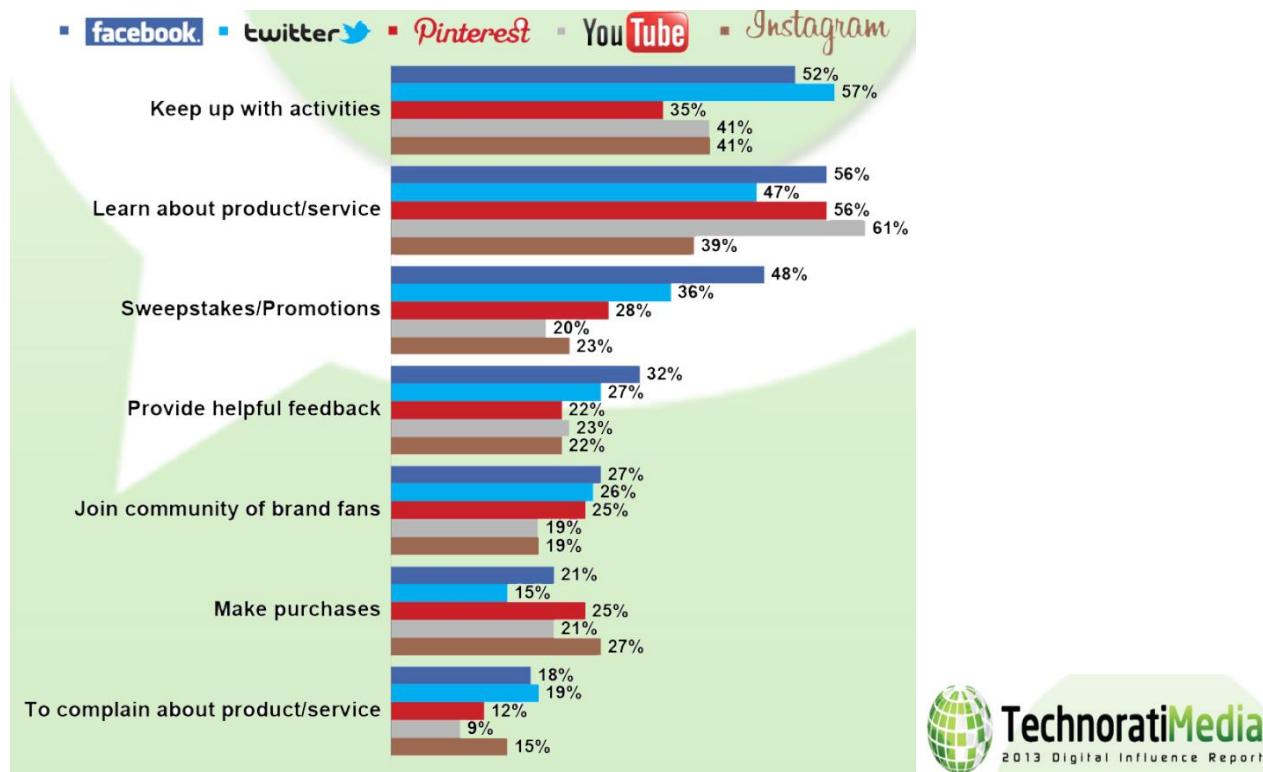


Sponsored by



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

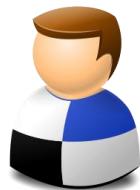
# Why Should We Study Them? Cnt



# Let's Take a Closer Look at Twitter



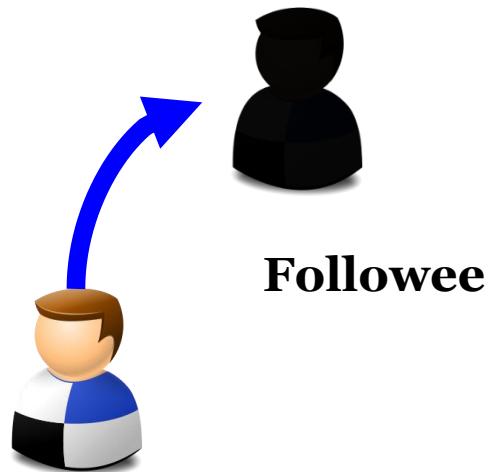
- Simple Structure



- Simple Structure

- Following

- To subscribe to other people's posts

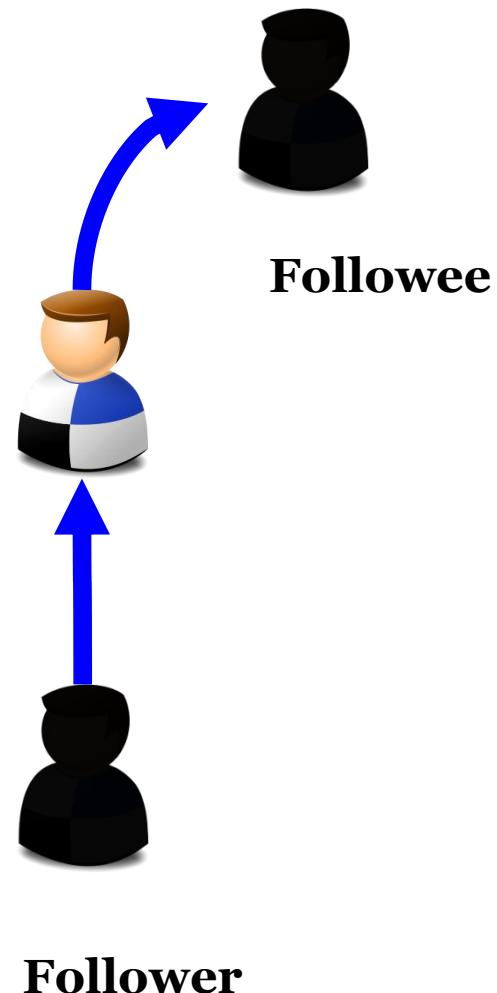


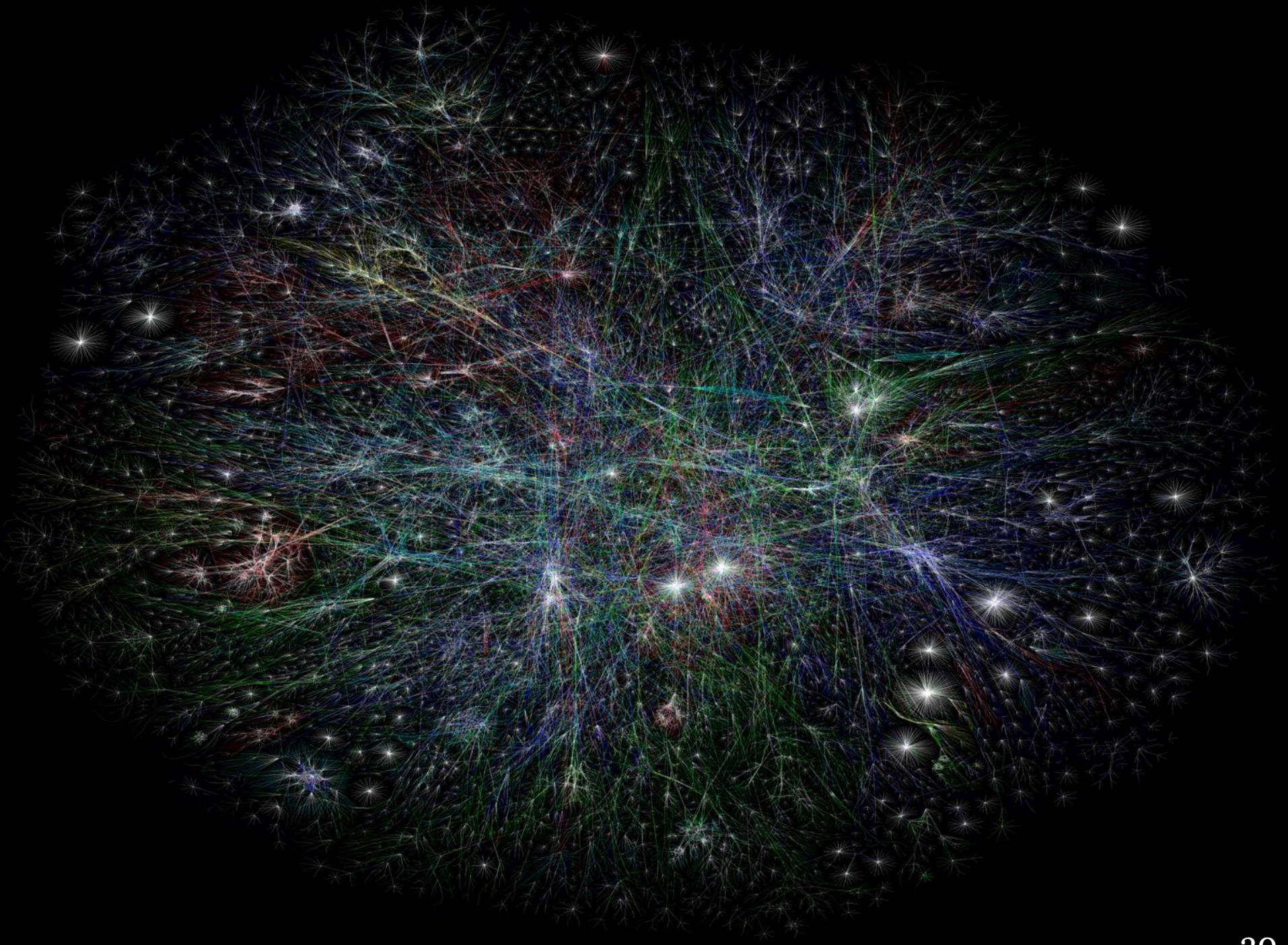
**Followee**

- Simple Structure

- Following

- To subscribe to other people's posts

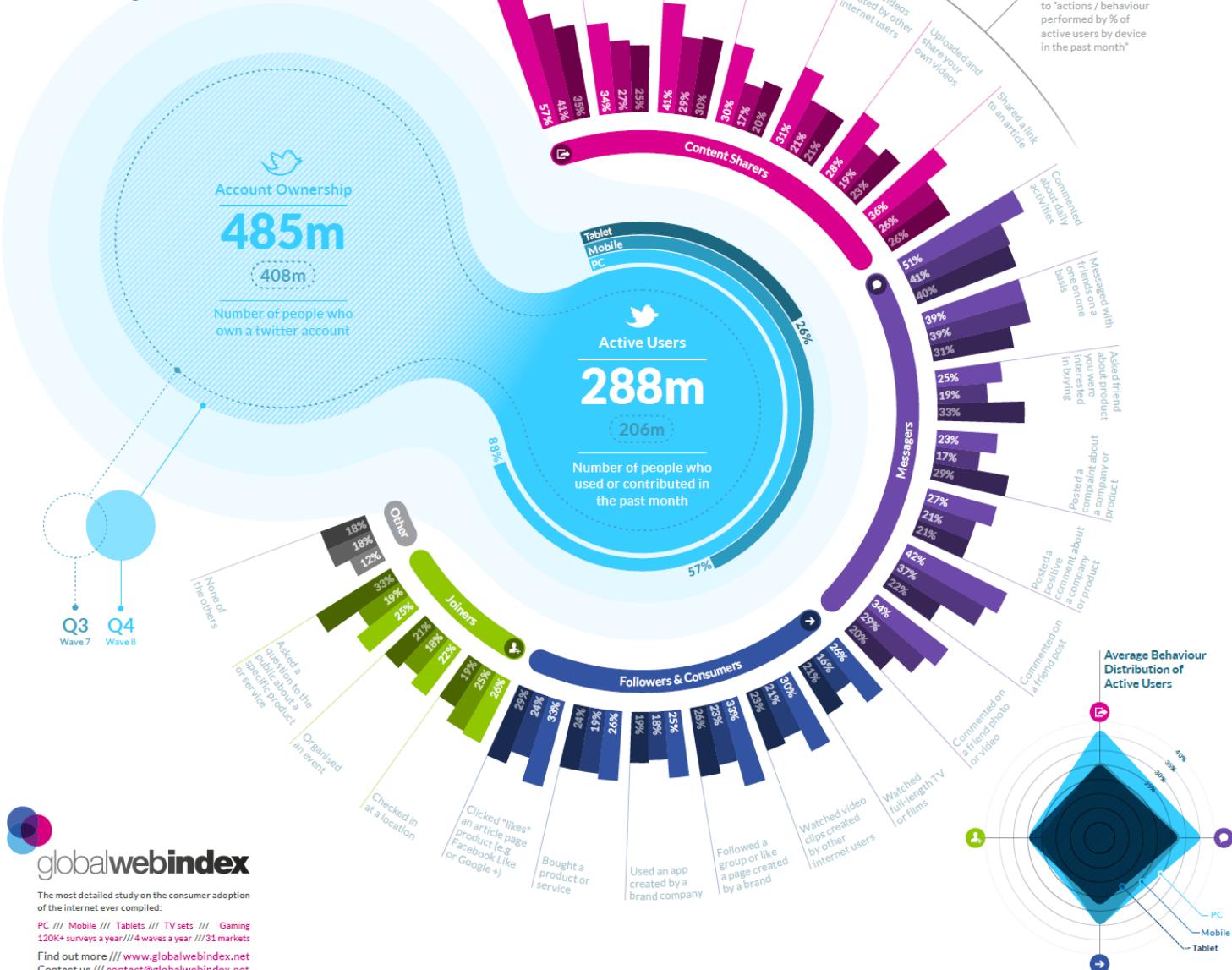




# TWITTER The Fastest Growing Social Platform

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](http://globalwebindex.net)



### Account Ownership

**485m**

408m

Number of people who own a twitter account

18%  
18%  
12%

Other

88%

### Active Users

**288m**

206m

Number of people who used or contributed in the past month

57%

57%

41%

35%

34%

27%

25%

41%

29%

30%

30%

17%

20%

28%

Content Sharers

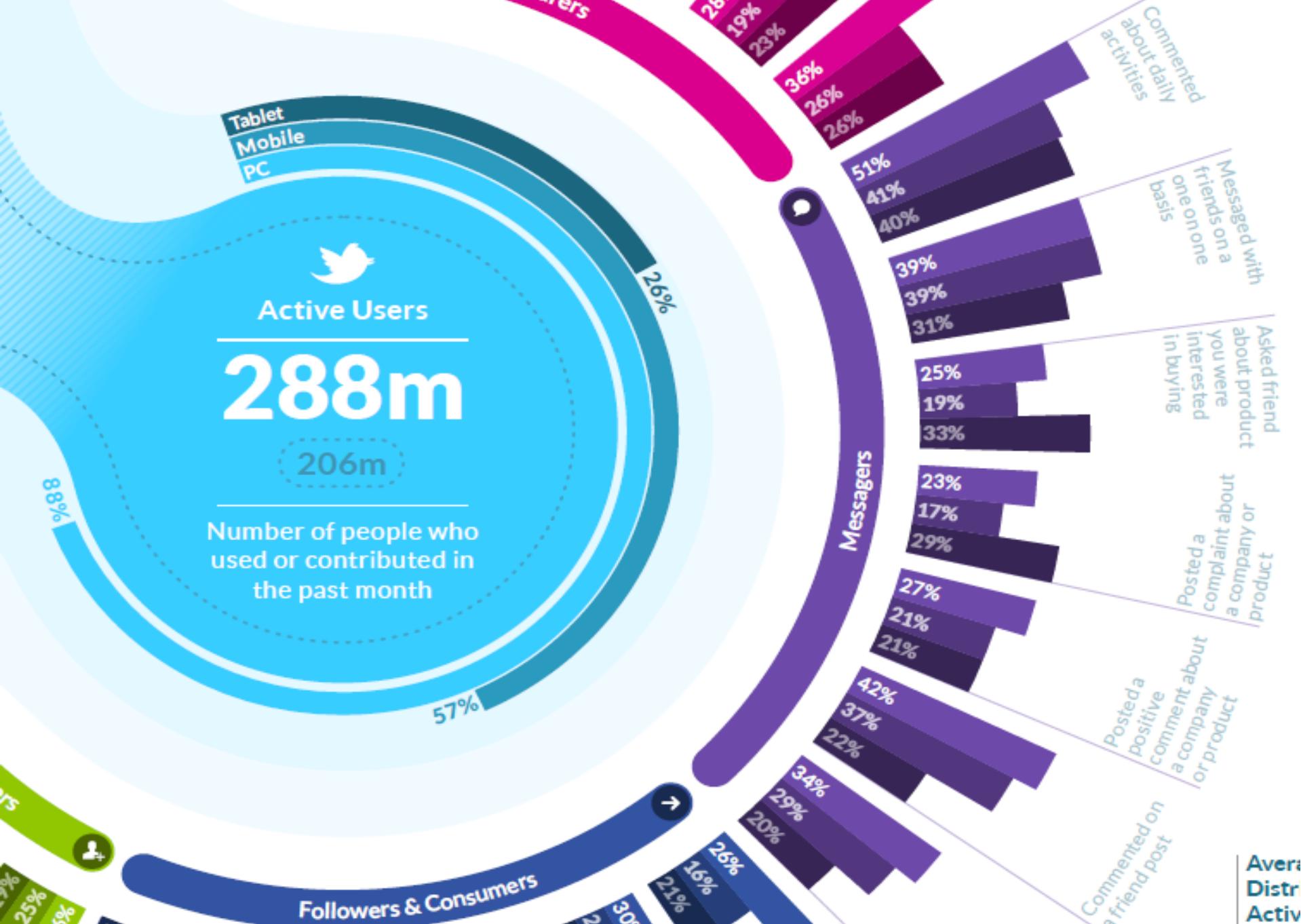
Tablet  
Mobile  
PC

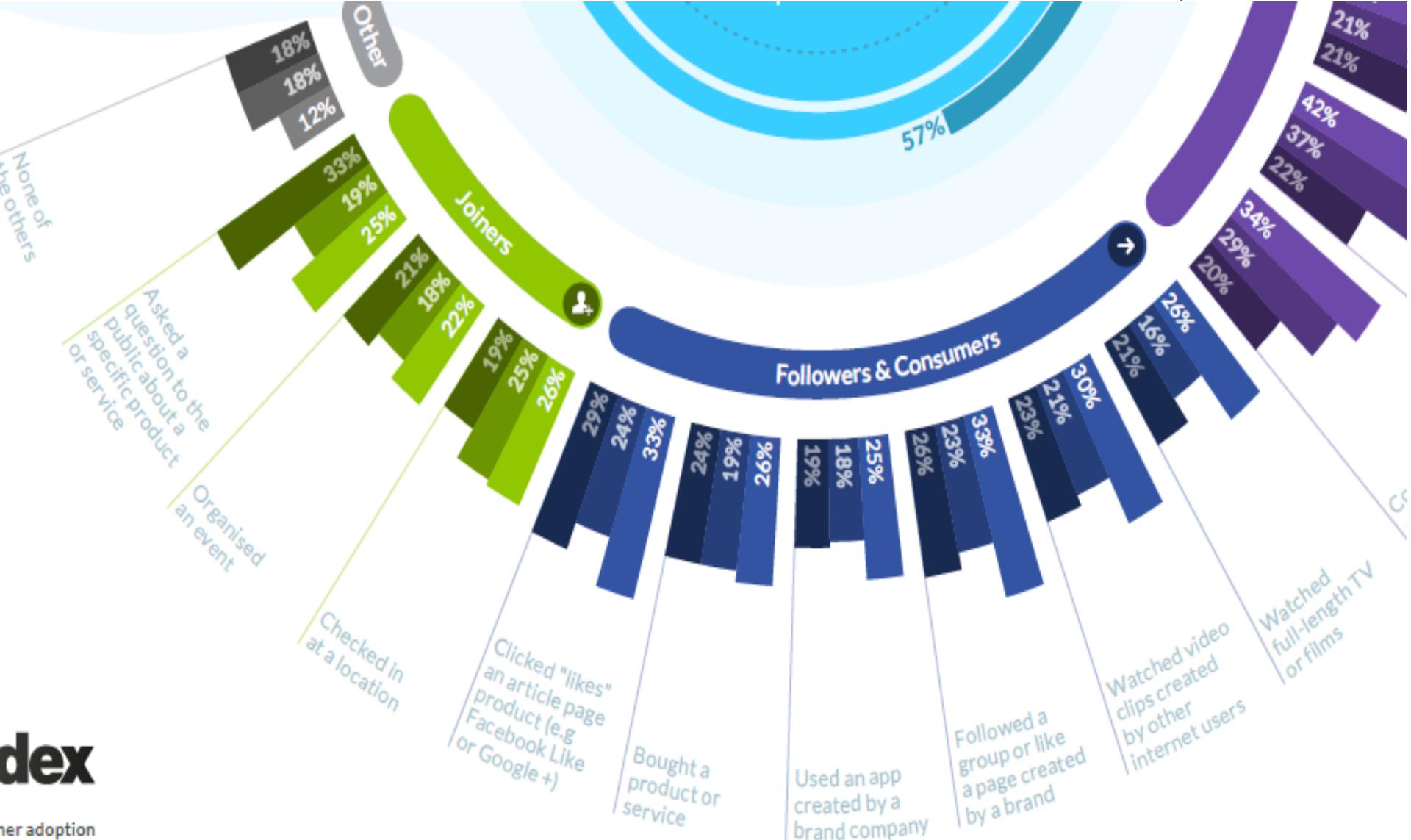
26%

57%

# al Platform









## Savannah Castillo

@Savann\_uhh

Jeremiah 29:11 ❤ God. Family. NMSU.  
AlphaXiDelta. :)

📍 #TB @toniooh\_

⌚ Joined July 2010

TWEETS PHOTOS/VIDEOS FOLLOWING FOLLOWERS FAVORITES  
47.7K 215 600 1,219 368

```
object {41}
  id : 170038954
  id_str : 170038954
  name : Savannah Castillo
  screen_name : Savann_uhh
  location : #TB @toniooh_
  description : Jeremiah 29:11 ♥ God. Family. NMSU. AlphaXiDelta. :]
  url : null
  entities {1}
    protected : false
  followers_count : 1219
  friends_count : 600
  listed_count : 10
  created_at : Fri Jul 23 20:03:09 +0000 2010
  favourites_count : 368
  utc_offset : -21600
  time_zone : Mountain Time (US & Canada)
  geo_enabled : true
  verified : false
  statuses_count : 47710
  lang : en
  status {20}
    contributors_enabled : false
    is_translator : false
    is_translation_enabled : false
    profile_background_color : F253B0
    profile_background_image_url : http://pbs.twimg.com/profile_background_images/540368237/kenny.png
    profile_background_image_url_https : https://pbs.twimg.com/profile_background_images/540368237/kenny.png
    profile_background_tile : true
    profile_image_url : http://pbs.twimg.com/profile_images/483389022408957952/jTaCTiEn_normal.jpeg
    profile_image_url_https : https://pbs.twimg.com/profile_images/483389022408957952/jTaCTiEn_normal.jpeg
    profile_banner_url : https://pbs.twimg.com/profile_banners/170038954/1380910058
    profile_link_color : FF006A
    profile_sidebar_border_color : D91175
    profile_sidebar_fill_color : FF428E
    profile_text_color : 747573
    profile_use_background_image : true
    default_profile : false
```

 Savannah Castillo  
@Savann\_uhh Follow

Debating if I should switch services with my family or if I should just stay on my own because I reallyyyyyy don't want to leave Verizon..

 Reply  Retweet  Favorite  More

10:01 AM - 1 May 2014

```
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 reallyyyyyy 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-wise, social contents are:
  - full of user-generated / urban words,
  - often short, context-less, and very noisy, and
  - are of streaming type!

# What Do We Learn?

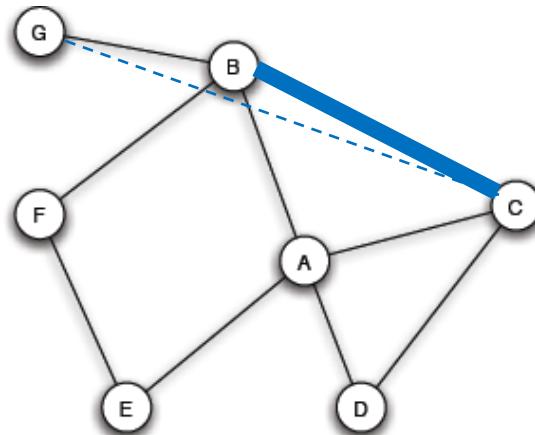
- Strong and Weak Ties
- Balanced and Unbalanced Networks
- Fun topics: ideas, tools and datasets, vw tutorial, etc.
- Web Graph and Network Popularity
- Link Analysis
- Information Cascading
- Small World Phenomenon
- Node Similarity and Homophily
- Mining Data Streams
- Social Search
- Trend Detection and Tracking
- etc.

# What Do We Learn? Cnt.

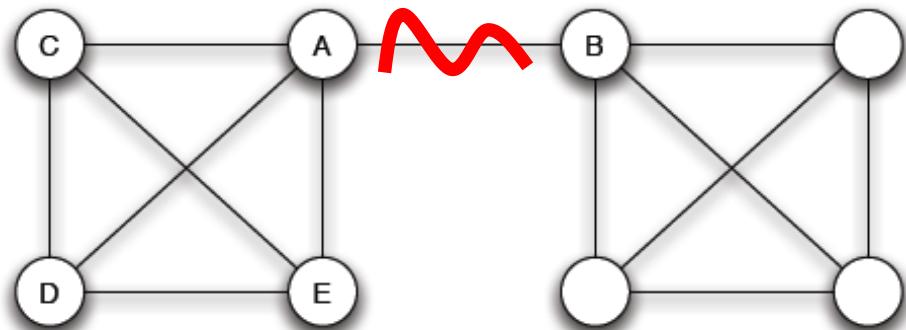
- Strong and Weak Ties
  - In 1960s, Mark Granovetter studied "How people find out about new jobs?"
    - People find the information through personal contacts
    - But: contacts were often *acquaintances* (weak ties) rather than *close friends* (strong ties)!
  - This is surprising as one would expect your close friends to help you more than acquaintances!
  - We study structural properties of networks to understand why acquaintances are most helpful?

# What Do We Learn? Cnt.

- Strong and Weak Ties



C-B is more likely to form or C-G?

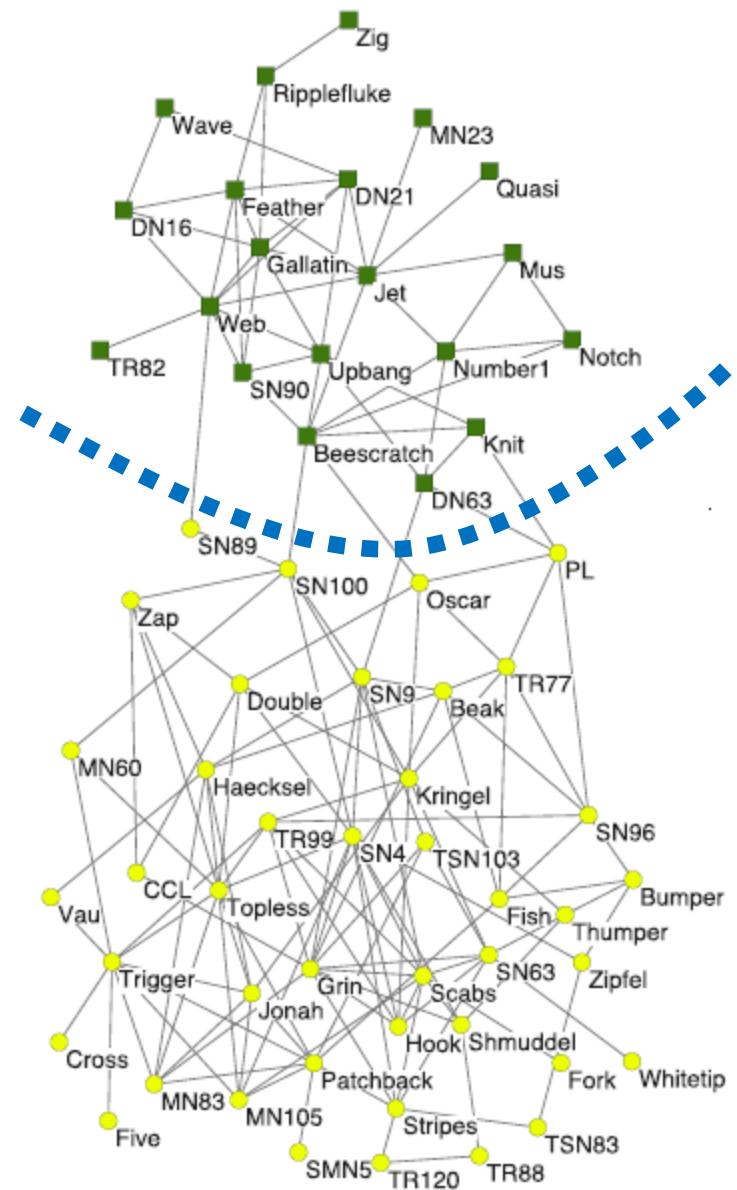
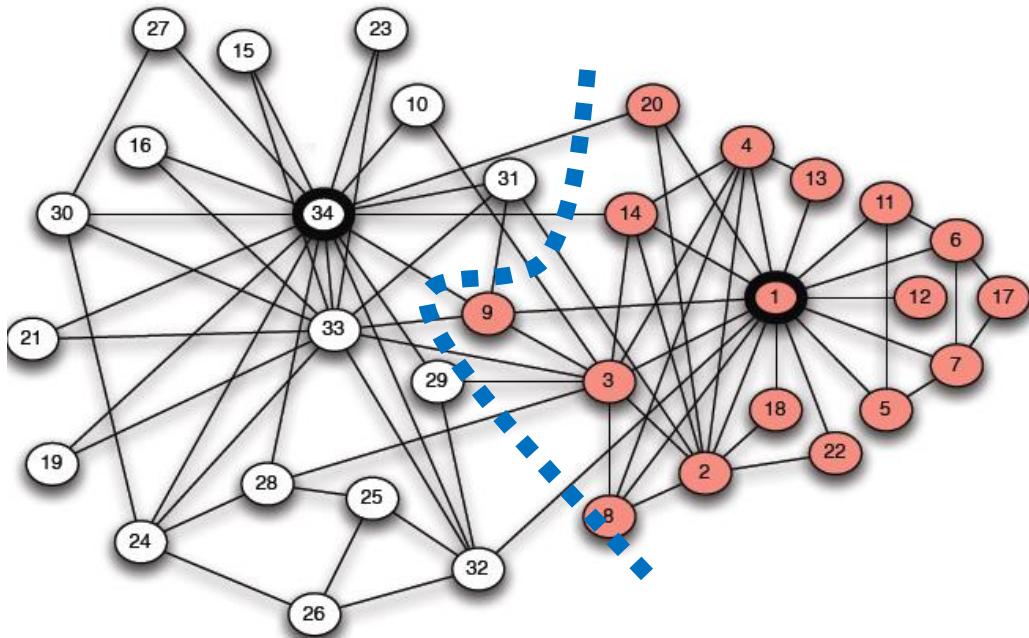


Which link provides access to parts of the net that are unreachable by other means?

**Are some nodes more important due to their position in networks?**

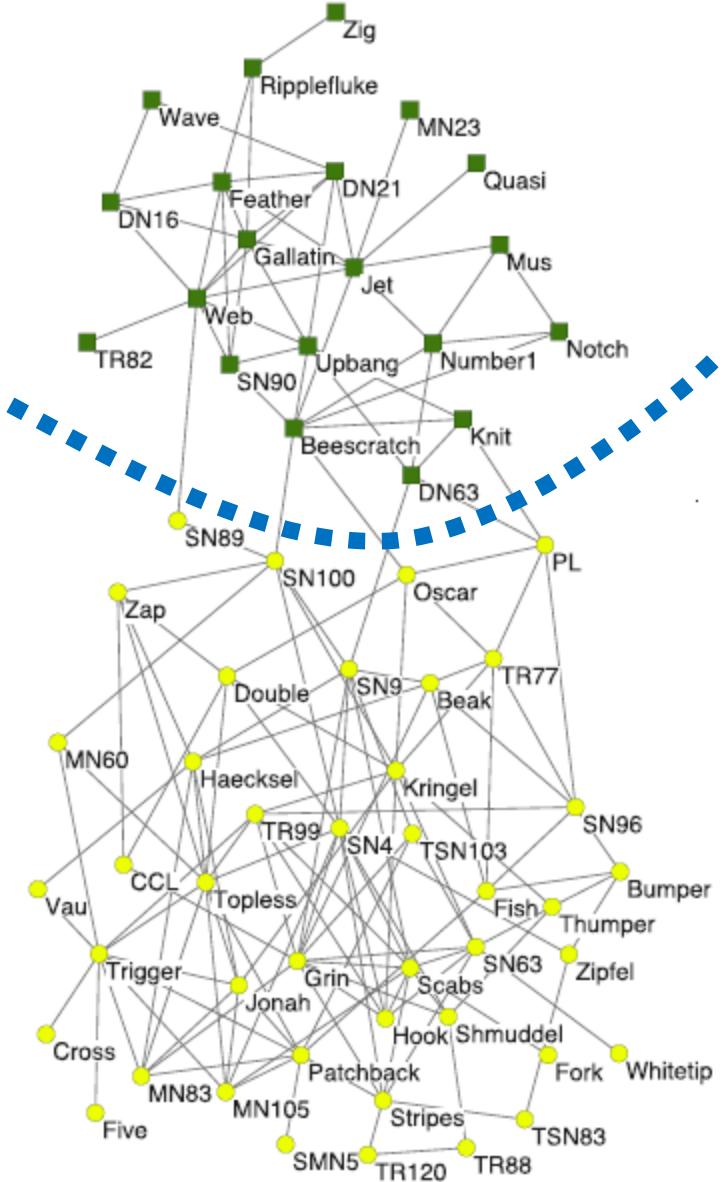
# What Do We Learn? Cnt.

- Social Network Clustering
  - Clusters / Communities
  - Clustering Coefficient



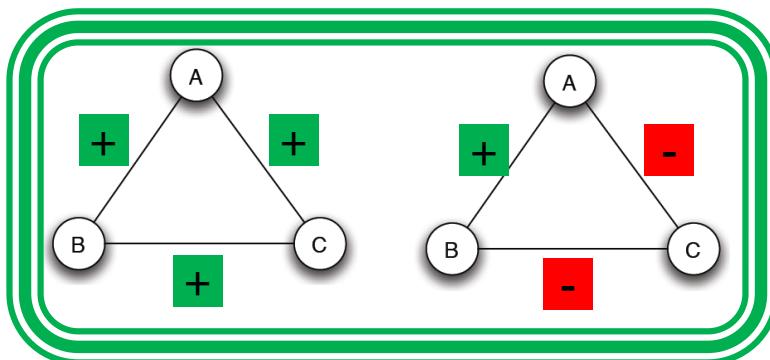
# What Do We Learn? Cnt.

- Social Network Clustering
  - Given a notion of similarity / distance, find groups such that:
    - Members of the same groups are similar to each other
    - Members of different groups are dissimilar!
  - Aim:
    - Maximizing intra-similarity
    - Minimizing inter-similarity

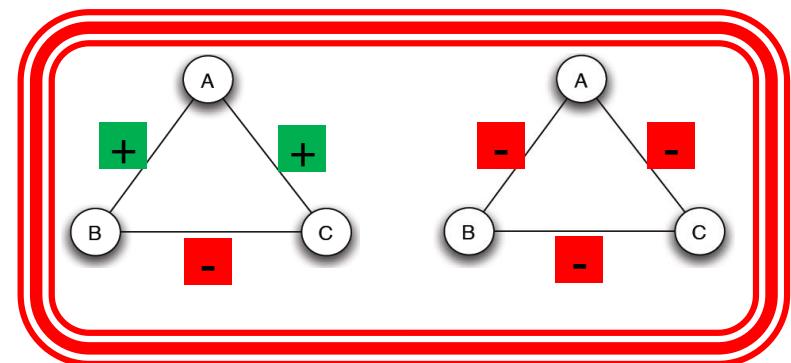


# What Do We Learn? Cnt.

- Structural Balance
  - Take a network and annotate its links with
    - + sign representing friendship, and
    - - sign representing antagonism
  - How should we reason about such networks?
    - Say to understand the *tension* between these two forces!



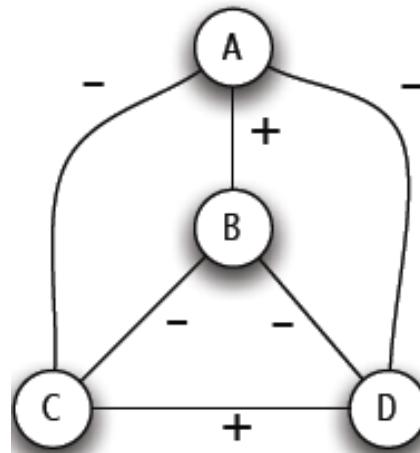
**Balanced**



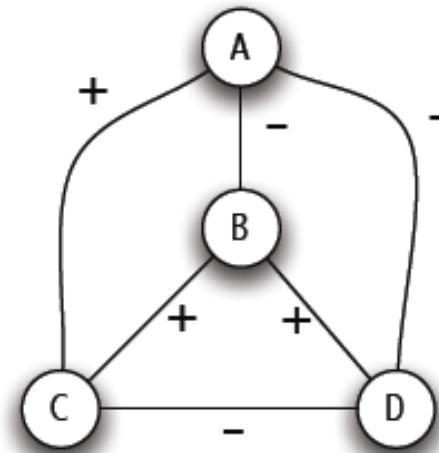
**Unbalanced** Psychologically unstable?

# What Do We Learn? Cnt.

- Structural Balance
  - Take a network and annotate its links with
    - + sign representing friendship, and
    - - sign representing antagonism
  - How should we reason about such networks?
    - Say to understand the *tension* between these two forces!



balanced



not balanced

# 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.

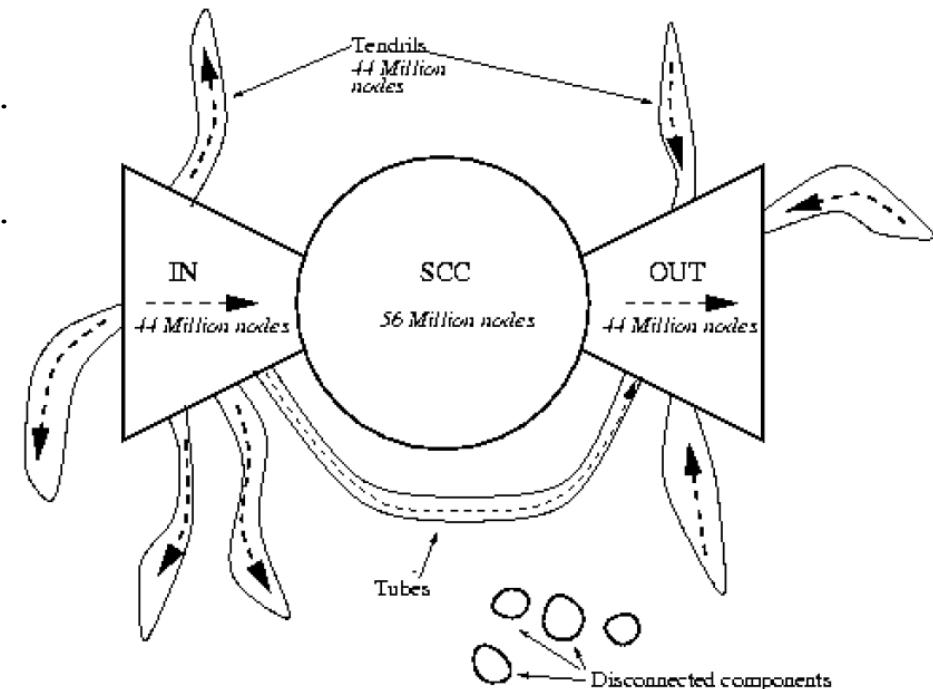
**Tendrils nodes:**

- (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:**

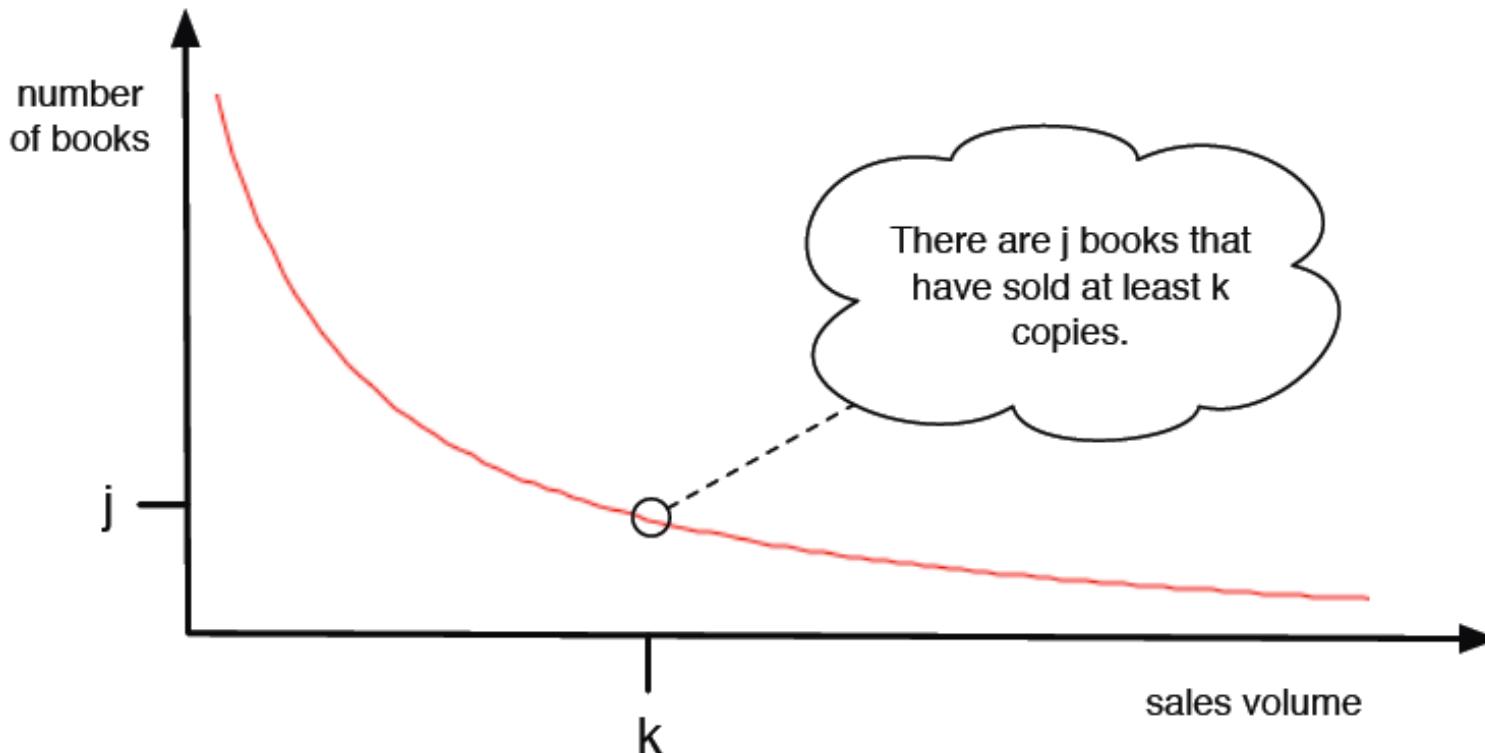
have no path to SCC ignoring directions



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

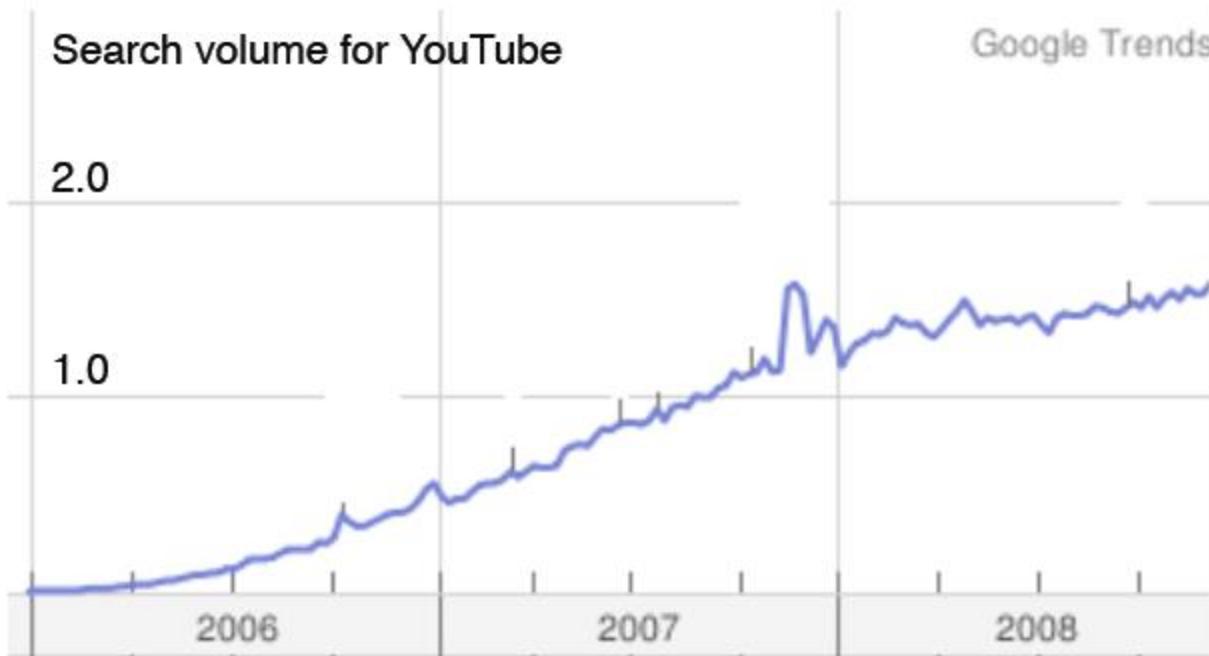
# What Do We Learn? Cnt.

- Popularity in Networks: Power Laws



# What Do We Learn? Cnt.

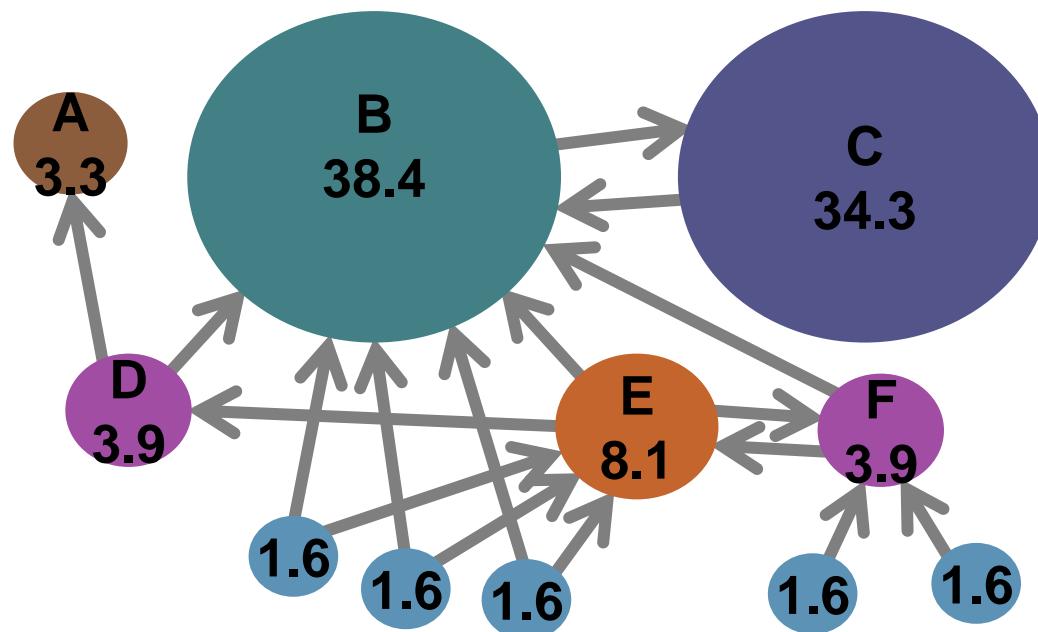
- Popularity in Networks: Rich Get Richer



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

# What Do We Learn? Cnt.

- Link Analysis and Web Search
  - Networks, e.g. the Web, are often huge while our information need is often small.
  - Nodes are not equally important.
    - So, we need a good way to rank them!



# 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 ???



**Information cascade is just behaviors that cascade from one node to another like an epidemic !**

# What Do We Learn? Cnt.

- Information Cascading

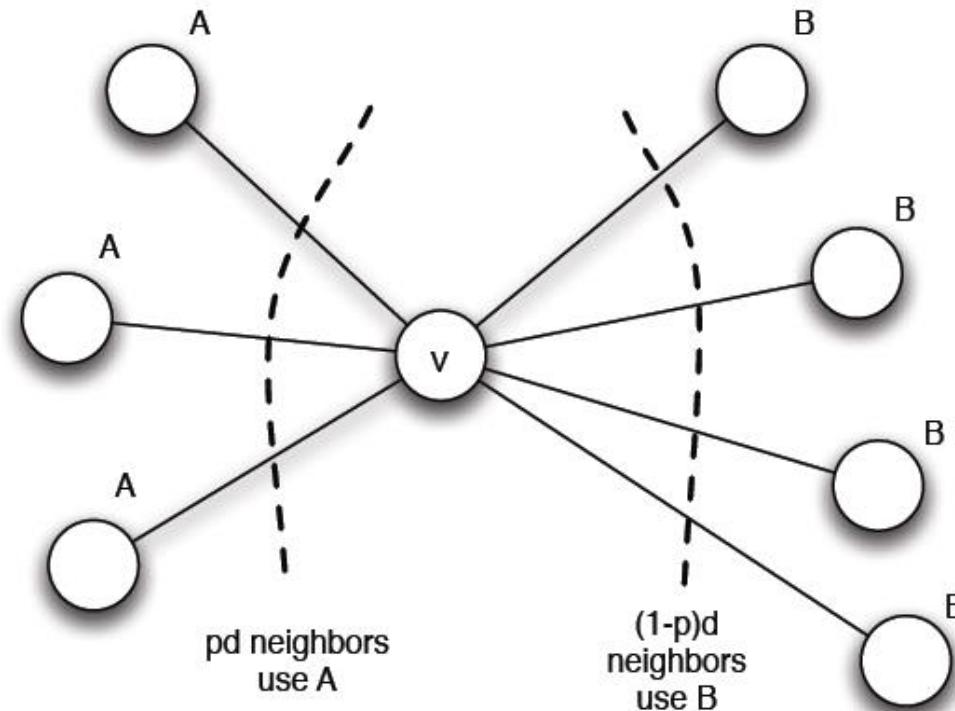
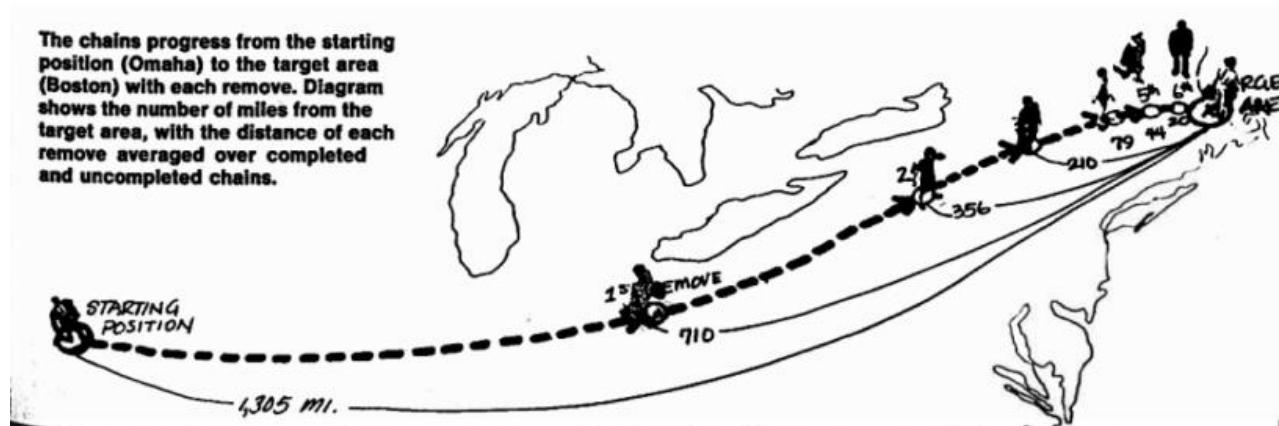


Figure 19.2:  $v$  must choose between behavior  $A$  and behavior  $B$ , based on what its neighbors are doing.

# What Do We Learn? Cnt.

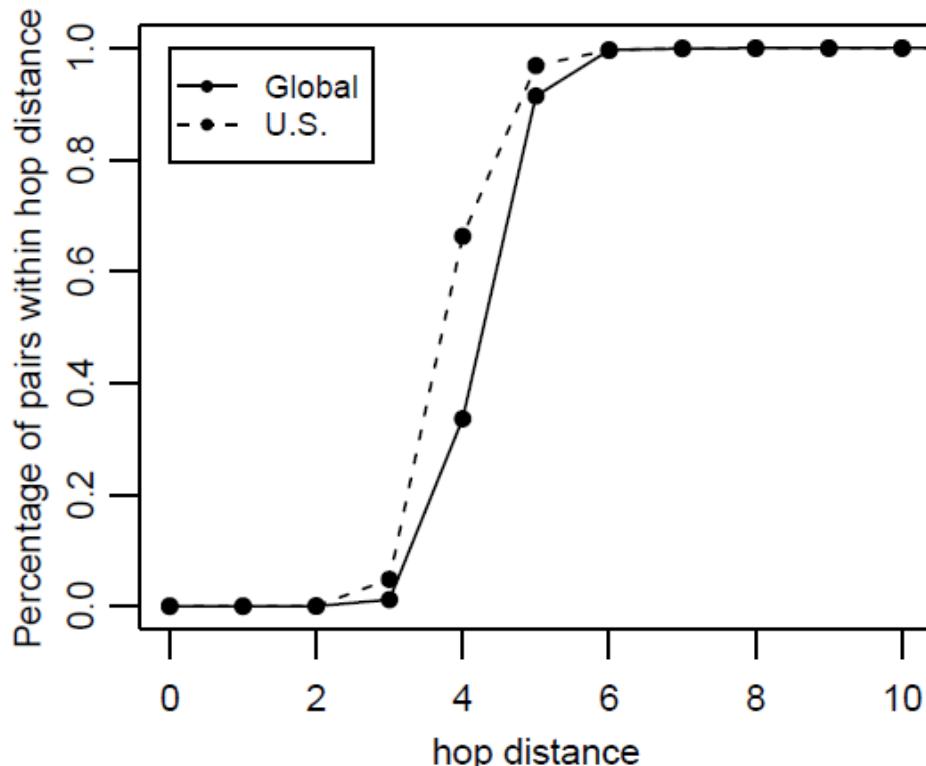
- Small World Phenomenon
  - Stanley Milgram (1967):
    - People were asked to send a letter to a target person in Sharon, MA!
      - They have target's name, address, occupation, etc.
    - But, each sender could only pass on based on first name basis
    - 33% of the letters arrived, Median: 6 steps



# What Do We Learn? Cnt.



- Small World Phenomenon



## Global

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

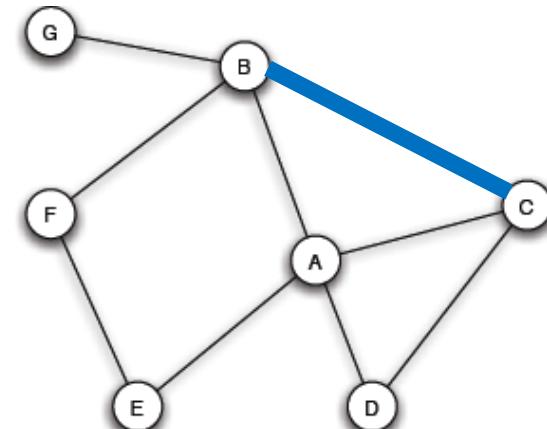
## U.S. only

96.0%: within 5 degrees,  
99.7%: within six degrees.

**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.

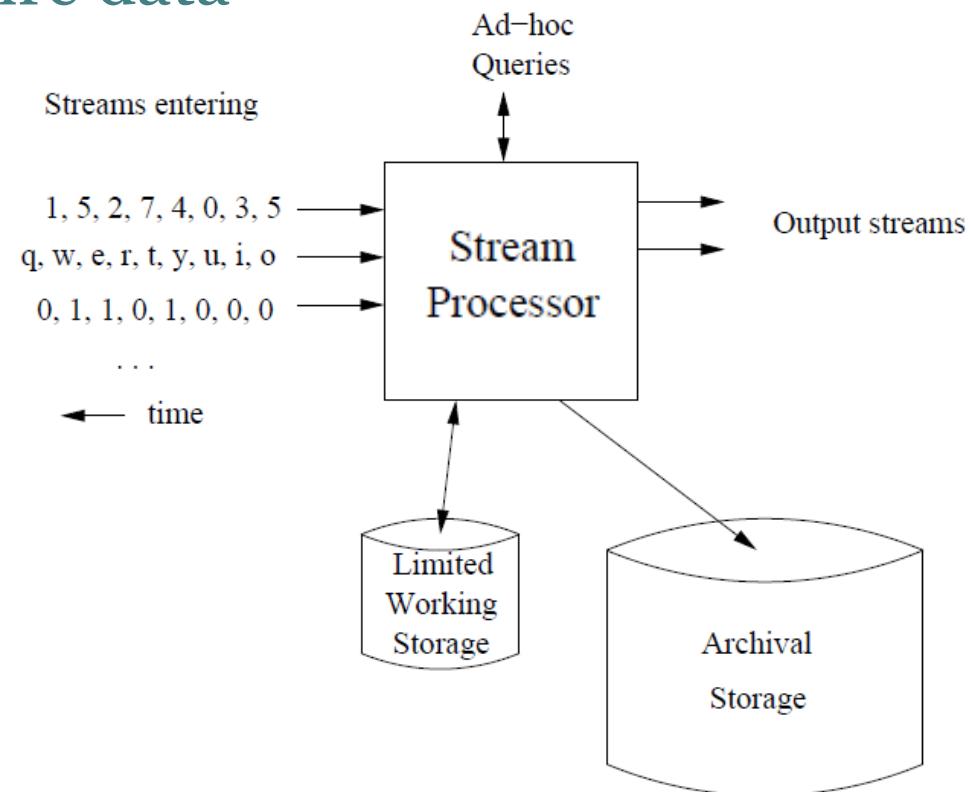
# What Do We Learn? Cnt.

- Homophily and Link Prediction
  - Homophily: we tend to have similar characteristics with our friends!
    - How can we test if a network exhibits homophily?
    - How can we predict the likelihood of the existence of a link between two nodes?
      - Links btw words and documents
      - Links btw Individuals, etc.



# What Do We Learn? Cnt.

- Stream Processing
  - Data streams in at a rapid rate from different sources
  - System cannot store entire data
    - limited memory
  - If data is not processed immediately, it is lost forever.
  - How do we respond queries about:
    - data, or
    - summaries of data?



# What Do We Learn? Cnt.

- Microblog Search

Recency important

Results for "Super Bowl"

Save

Top / All

17 new results

MIT follows  
**Business Insider** @businessinsider · 5h  
 A few over the top NYC **Super Bowl** parties we've stumbled upon  
[read.bi/1xKXM50](#)

 View photo

Mike Greenberg @HarperCollins · 5h  
 Looks like NFL hired an investigator. Which conveniently pushes any discoveries well beyond the **Super Bowl**. That was never going to happen.

 View photo

Stephanie Haberman @StephLauren · 5h  
 Please don't choose recipe #49 on our **Super Bowl** food poll  
[on.si.com/49Recipes](#)

 View photo

KPRC Local 2 Houston @KPRCLocal2 · 5h  
 Sports update: **Super Bowl** XLIX commercials [bit.ly/1Cvzgvh](#) #KPRC  
#Local2Sports

 View summary

Digg @digg · 5h  
 A cat's advice to a kitten about the **Super Bowl**: [di.gg/1xJ7ENE](#)

 View summary



"Super Bowl"

Web Images Videos Maps News More

14,600,000 RESULTS Any time ▾

**Super Bowl - Wikipedia, the free encyclopedia**  
[en.wikipedia.org/wiki/Super\\_Bowl](#) ▾  
 The **Super Bowl** is the annual championship game of the National Football League, the highest level of professional football in the United States, culminating a ...  
 Origin · Date · Game history · Television coverage ... · Entertainment · Venue

**Super Bowl XLVIII - Official Site**  
[www.nfl.com/superbowl/48](#) ▾  
 2014 Super Bowl - Super Bowl XLVIII - Official Site of the National Football League  
 Super Bowl 48

**Super Bowl 2015 - CBSSports.com**  
[www.cbssports.com/nfl/superbowl](#) ▾  
 Get the latest picks, news and analysis for the 2015 **Super Bowl** at CBSSports.com

**News about "Super Bowl"**  
[bing.com/news](#)

 **Man loses Super Bowl pass because of Facebook post**  
 FOX 8 · 1 minute ago  
 PHOENIX — It was a dream job for longtime sports fan Russ Knight – a chance to work at the **Super Bowl**. But that dream turned into nightmare, because of a...

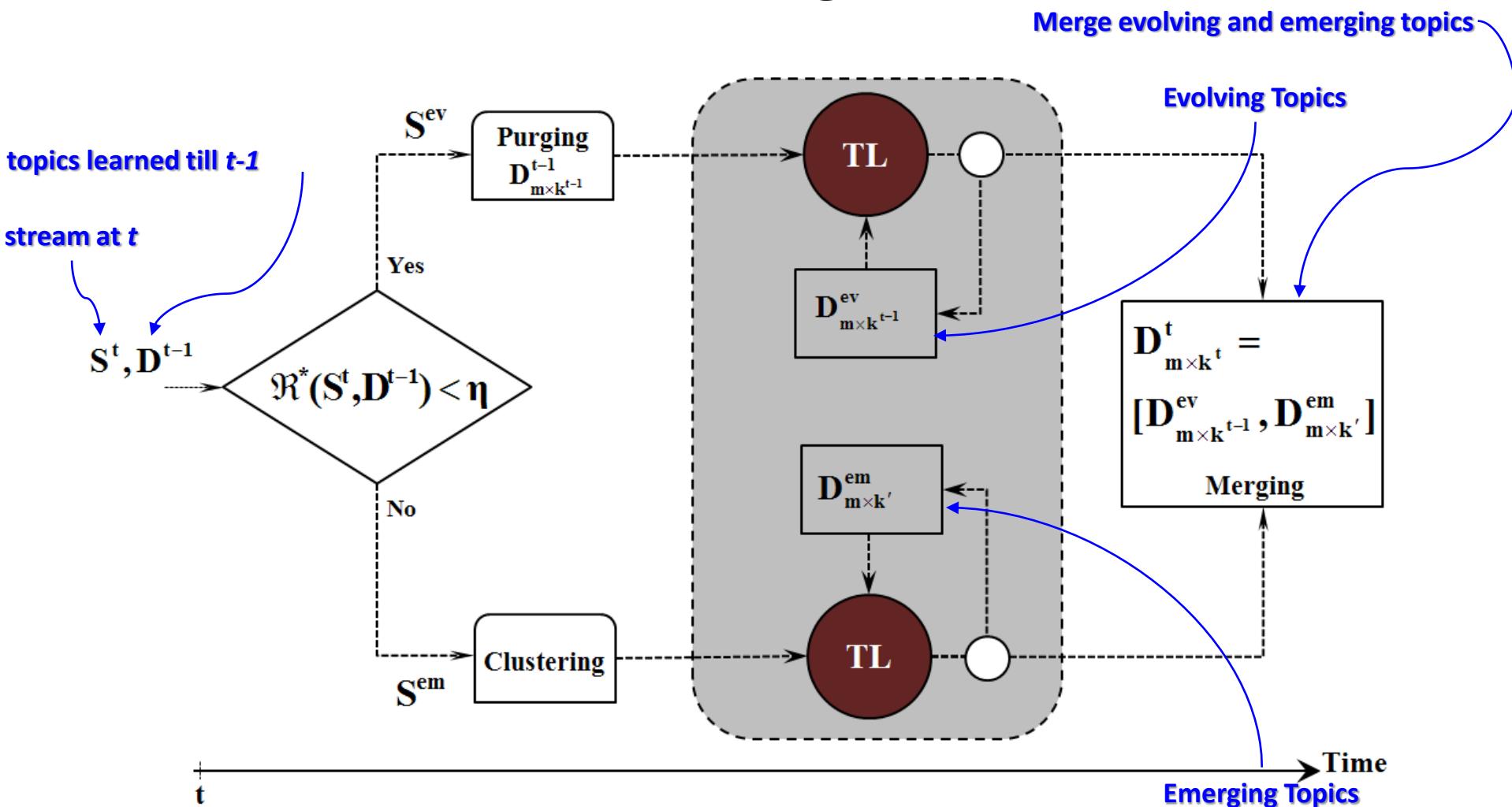
**Super Bowl prop bets: Idina Menzel's anthem among wagers**



Relevance important

# What Do We Learn? Cnt.

- Trend Detection & Tracking



# What Do We Learn?

- Strong and Weak Ties
- Balanced and Unbalanced Networks
- Fun topics: ideas, tools and datasets, vw tutorial, etc.
- Web Graph and Network Popularity
- Link Analysis
- Information Cascading
- Small World Phenomenon
- Node Similarity and Homophily
- Mining Data Streams
- Social Search
- Trend Detection and Tracking
- etc.

# Questions?

# Reading

- Ch.01 Overview [NCM]
- Ch.10.1 Social Networks as Graphs [MMD]
- What is Twitter:
  - What is Twitter, a social network or a news media?  
Kwak et al. WWW 2010.
- Who uses Twitter:
  - Understanding the Demographics of Twitter Users.  
Mislove, Alan, et al. ICWSM 2011.
- What do they Tweet about:
  - Why we twitter: understanding microblogging usage and communities. Java, Akshay, et al. *WebKDD and SNA-KDD 2007*.

# Announcements!

- No Class on Wed Jan 28! Instead:
- Watch this 30 min TED talk by Deb Roy @ MIT:
  - From Gaga to Water: <http://bit.ly/12fIOeR>
- Go to Talk Local (start-up co-founded by UMD Alumni):
  - Wed. 4:00 - 5:00 p.m. <http://bit.ly/1y32Q4F>

