

# Social Computing (CS60017)

Instructors: Dr. Pawan Goyal

Dr. Animesh Mukherjee

# Course Details

- Teaching Assistants:
  - Abhijnan Chakraborty (chakraborty.abhijnan@gmail)
  - Sandipan Sikdar (sikdarsandipan99@gmail)
- Timing:
  - Slot E → WED(12-1) , THU(11-12) , FRI(9-10)
- Location: CSE 107
- Course Website: <http://cse.iitkgp.ac.in/~pawang/courses/SC16.html>
- Marks
  - Midsem: 20, Endsem: 40, Term project/Class test: 40

# TA Component

- Option between
  - Term Project (40): If you opt for this you have to self-declare that you are willing to take the load and that you will not complain that this a source of bottleneck for other subjects.
  - Pros: Hands-on experience of the course, teamwork, scope for research publications
- OR
- Class test (40)
- Cons: Routine test somewhere between midsem and endsem. The level of the exam will be equivalent to endsem.

# Term Projects

- Group of 3-4 people
- Open research problems from mentors
- Task includes: data collection, measurement, modeling, drawing inferences
- Midsem Eval – 40% results should be produced.
- Endsem Eval – Rest 60% should be produced.

# Broad Topics

- Text processing for Social Media
- Entity linking and Entity resolution
- Information diffusion
- Basic experimental studies on social media
- Advanced experimental studies
- Supervised random walks
- Heterogeneous networks and metapaths
- Topic models
- Recommendation and link prediction
- Correlation versus causality

# Online Social Networks

- Facebook – Friendship
- Twitter – Microblogging
- LinkedIn – Professional
- FourSquare – Location
- Google+ – Friendship+Blog
- ....
- ....

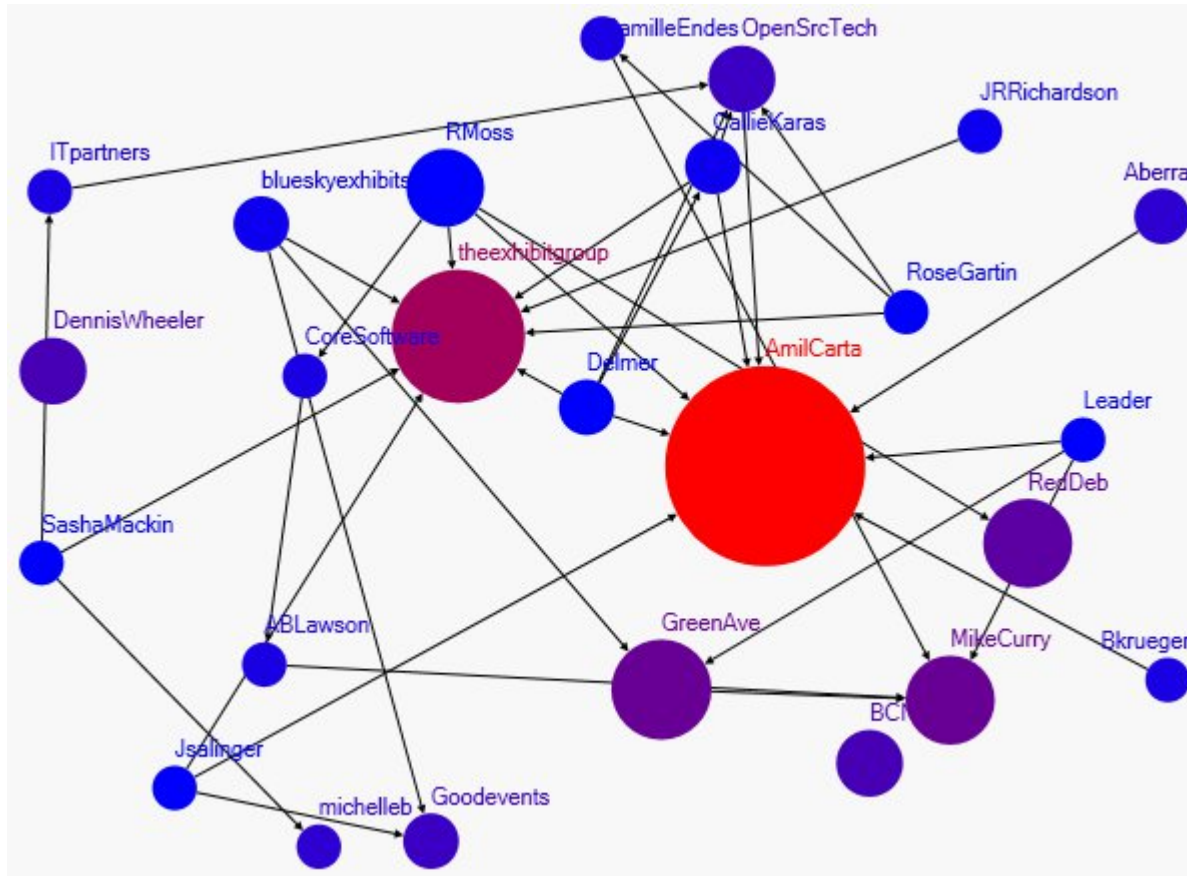


# Twitter

- Twitter is an online social networking website and microblogging service that allows users to post and read text-based messages of up to 140 characters, known as “tweets”.

Twitter Company Statistics	Data
Total number of active registered Twitter users	645,750,000
Number of new Twitter users signing up everyday	135,000
Number of unique Twitter site visitors every month	190 million
Average number of tweets per day	58 million
Number of Twitter search engine queries every day	2.1 billion
Percent of Twitter users who use their phone to tweet	43 %
Percent of tweets that come from third party applicants	60%
Number of people that are employed by Twitter	2,500
Number of active Twitter users every month	115 million
Percent of Twitters who don't tweet but watch other people tweet	40%
Number of days it takes for 1 billion tweets	5 days
Number of tweets that happen every second	9,100
Twitter Annual Advertising Revenue	Revenue
2013	\$405,500,000
2012	\$259,000,000
2011	\$139,000,000
2010	\$45,000,000

# The follower-follower network



Your tweets are visible only to your followers



# Important Entities

- **#hashtags** – metadata tag used to organize tweets
  - New artists announced for [#SXSW2014MusicFestival](#)
- **@mentions** – Directing a public post to a particular user (especially to reply to another user's post: @janedoe)
- **URLs** – short forms of web links (tiny URL)

# Facebook

Facebook Statistics	Data
Total number of monthly active Facebook users	1,310,000,000
Total number of mobile Facebook users	680,000,000
Increase in Facebook users from 2012 to 2013	22 %
Total number of minutes spent on Facebook each month	640,000,000
Percent of all Facebook users who log on in any given day	48 %
Average time spent on Facebook per visit	18 minutes
Total number of Facebook pages	54,200,000
Facebook Demographics	Data
Percent of 18-34 year olds who check Facebook when they wake up	48 %
Percent of 18-34 year olds who check Facebook before they get out of bed	28 %
Average number of friends per facebook user	130
Average number of pages, groups, and events a user is connected to	80
Average number of photos uploaded per day	205
Number of fake Facebook profiles	81,000,000
Global Facebook Reach Statistics	Data
Number of languages available on the Facebook site	70
Percent of Facebook users who are outside the United States	75 %
Number of users who helped translate Facebook	300,000
Facebook Platform Statistics	Data
Average number of aps installed on Facebook each day	20 million
Total number of apps and websites integrated with Facebook	7 million
Every 20 Minutes on Facebook	Data
Links shared	1 million
Friends requested	2 million
Messages sent	3 million
Facebook Company Statistics	Data
Total number of Facebook employees	4,619
Total 2012 Facebook revenue	\$5,090,000,000
Total 2013 Facebook revenue	\$6,150,000,000



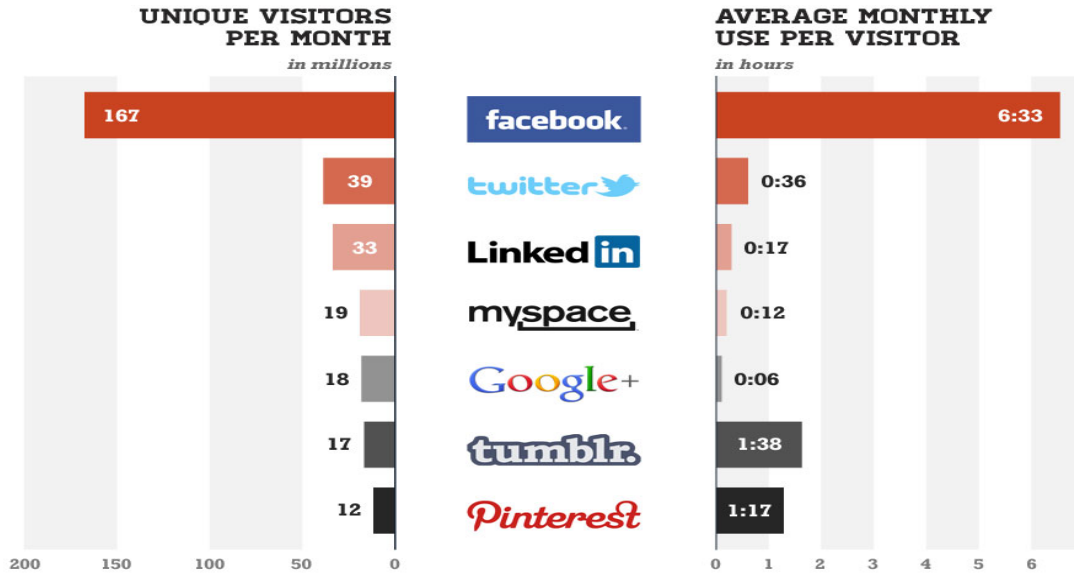
# Important Entities

- Likes – A way to vote for something you like
- Comment – A forum for discussion
- Share – Means to share and spread information
- Status – The current state of “mind”
- Photo/Video – Means to share photo/video
- ...
- ...

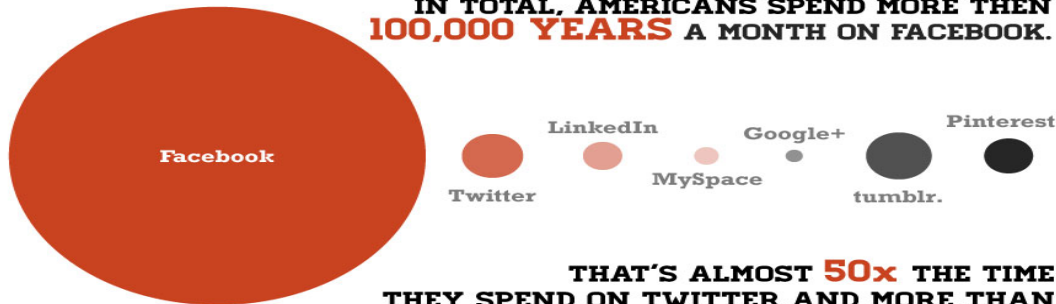
# Time is “worth this course”

Time is money

Social network use in the United States



IN TOTAL, AMERICANS SPEND MORE THEN **100,000 YEARS** A MONTH ON FACEBOOK.



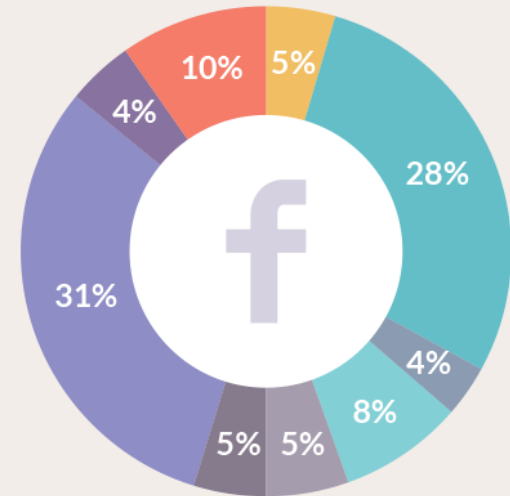
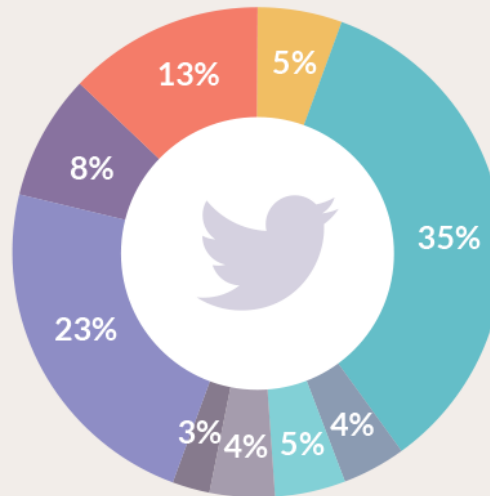
THAT'S ALMOST **50x** THE TIME THEY SPEND ON TWITTER AND MORE THAN **600x** THE TIME THEY SPEND ON GOOGLE+.

All numbers are estimates

# Points of Difference

Topic Engagement  
(including top 10)

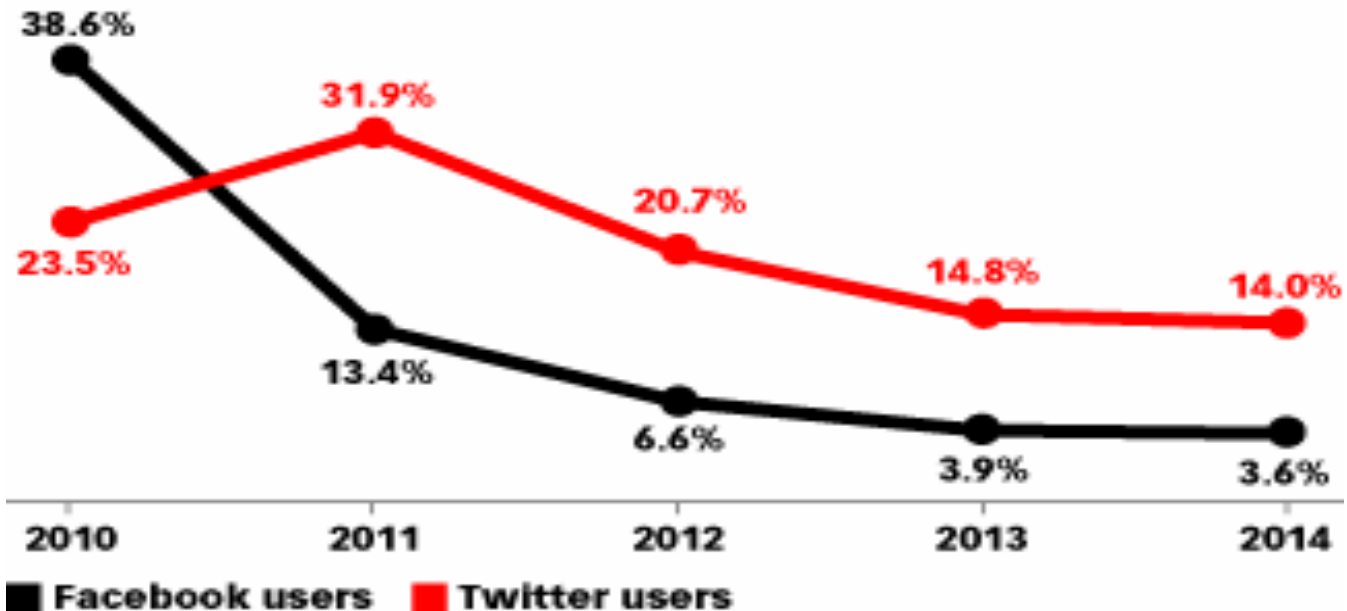
- Business
- Entertainment
- Fashion
- Food & Beverage
- Government & Politics
- Health & Wellness
- Lifestyle
- Sports & Recreation
- Technology



# In recent years

## US Facebook vs. Twitter User Growth, 2010-2014

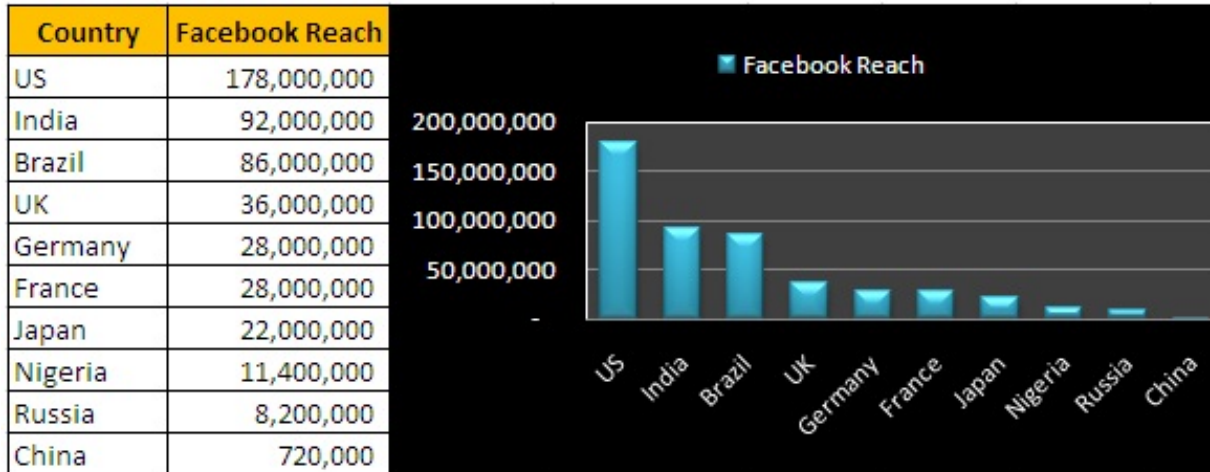
% change



*Note: internet users who access their Twitter or Facebook account via any device at least once per month*

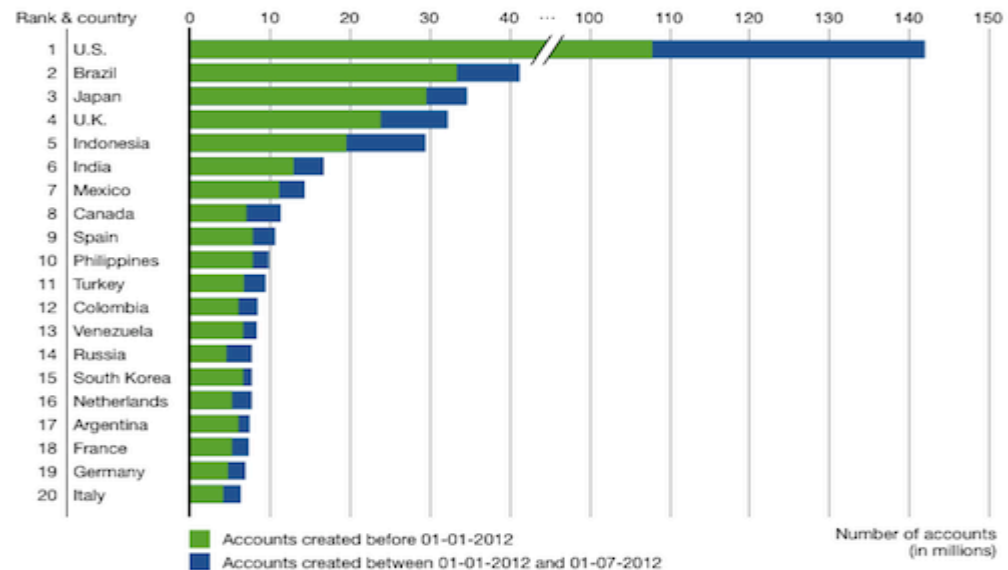
*Source: eMarketer, March 2012*

# Geography



## Top 20 countries in terms of Twitter accounts

(accounts created before 01-07-2012)

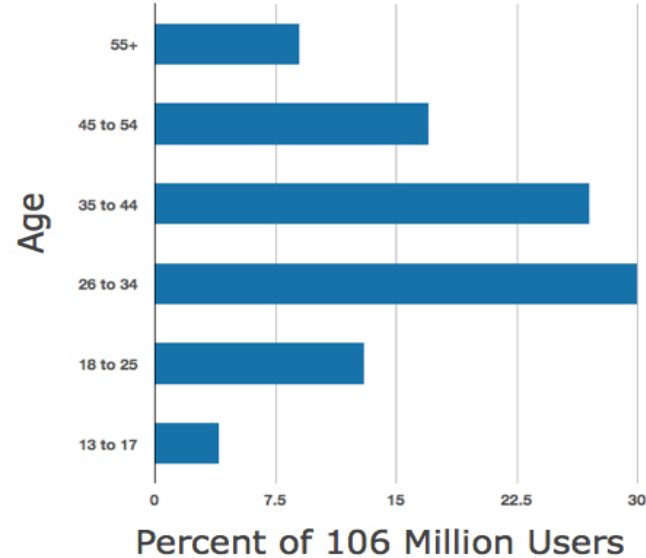
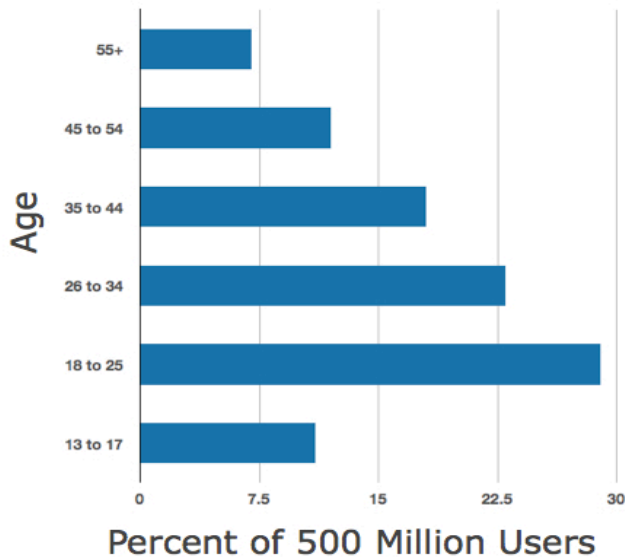




# Demography



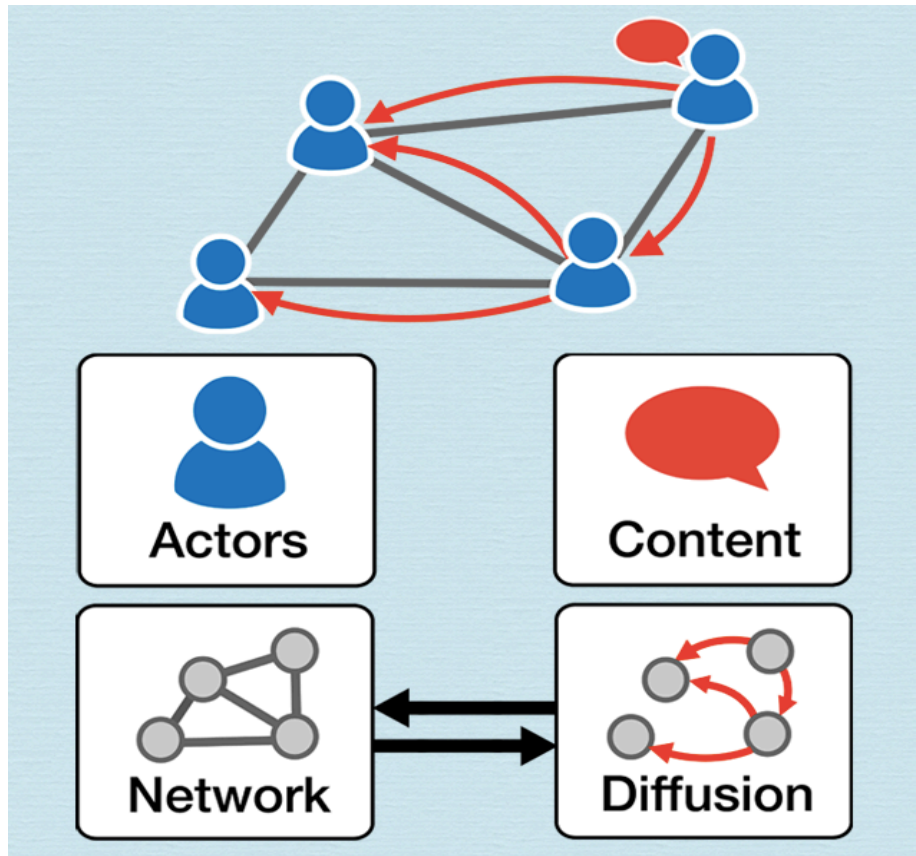
## Who's On by Age?



Source: Jackson, 2011, Barracuda Networks, Razorfish

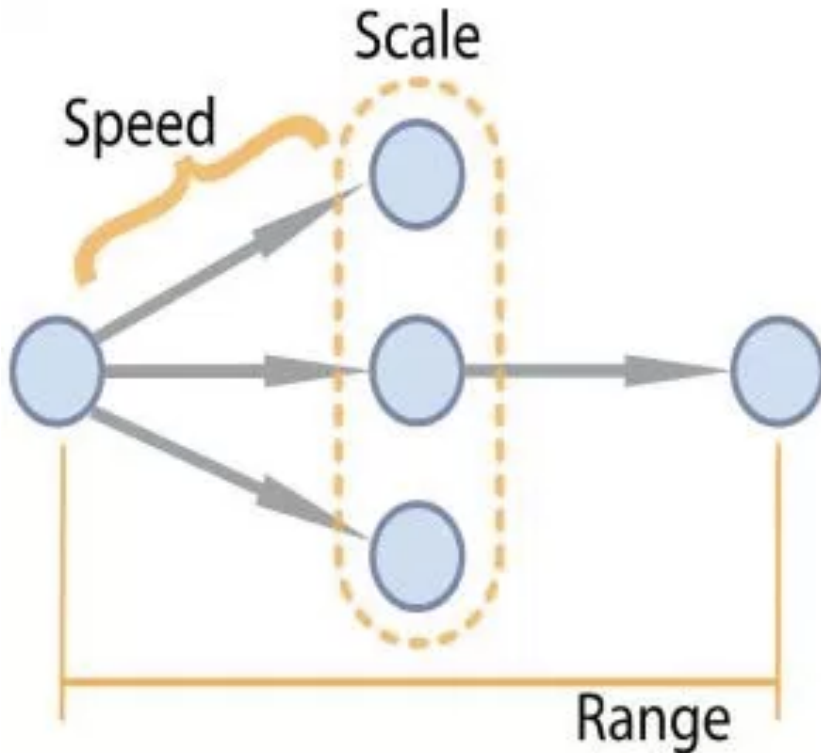
© Coughlin 2011  
[www.disruptivedemographics.com](http://www.disruptivedemographics.com)

# Information diffusion



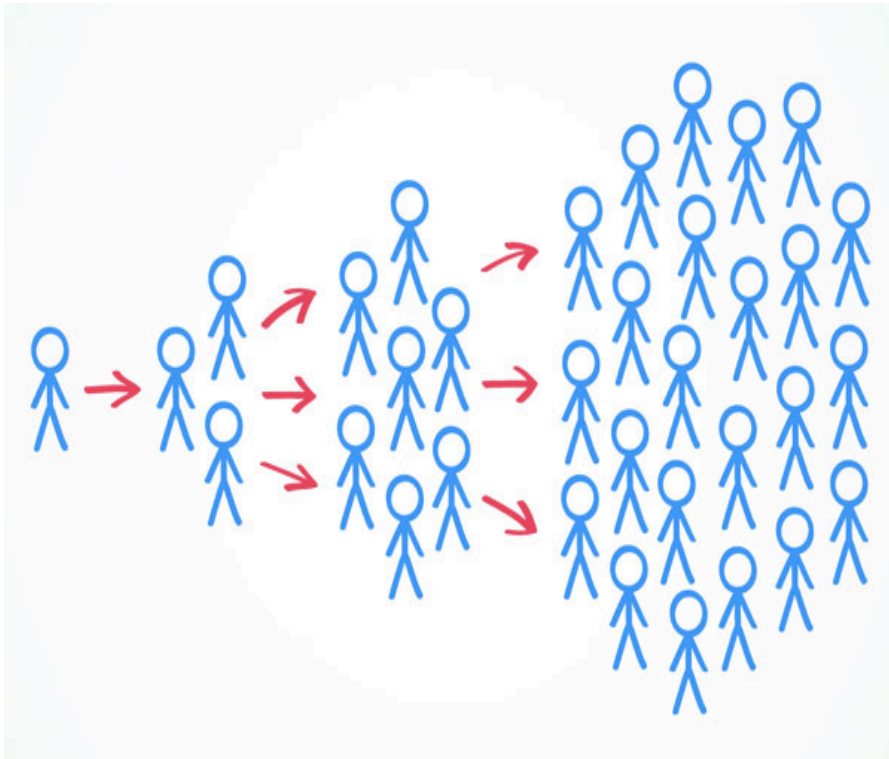
- People who share information;
- Features of spreading content;
- The interplays between network structure and diffusion process.

# Observables in diffusion



- **Speed** (how quickly does information travel through the network);
- **Scale** (how many nodes at the first degree are affected), and
- **Range** (how many hops in the network)

# Information cascades



- **Activation sequence:** An ordered set of nodes capturing the order in which the nodes of the network adopted a piece of information.
- **Spreading cascade:** A directed tree having as a root the first node of the activation sequence. The tree captures the influence between nodes (branches represent who transmitted the information to whom) and unfolds in the same order as the activation sequence.

# Social Data Analytics (fundamentals)

- **Social Media Data Analytics**
  - sophisticated analysis of the data in OSNs, in many cases requiring real-time (or near real-time) data analytics, measurements which understand and appropriately weigh factors such as influence, reach, and relevancy
- **Instances**
  - understand how sentiment evolved and impacted the Arab Spring movements sweeping the Middle East
  - understand how sentiment might be able to predict box office potential for yet-to-be-released movies
  - predict election outcomes, public sentiments about politicians

# India Election Game (MSR Initiative)

Predictions made by 4000 participants

Delhi (7 Seats)

[More Info](#)

Create your prediction ?

Bharatiya Janata Party(BJP) (4-5) [More Info](#)

0 1 2 3 4 5 6 7

Indian National Congress(INC) (0-1) [More Info](#)

0 1 2 3 4 5 6 7

Aam Admi Party(AAP) [More Info](#)

*The small state/UT's individually*

States	Seats	BJP	NDA Allies	Congress	UPA Allies	Third Front	Others
Andaman and Nicobar	1	1		0			
Arunachal Pradesh	2	1		1			
Chandigarh	1	1		0			
Dadra and Nagar Havel	1	1		0			
Daman and Diu	1	1		0			
Goa	2	2		0			
Lakshadweep	1	0		1			
Manipur	2	0	MPP 1	1			
Meghalaya	2	0	NPP 1	1			
Mizoram	1	0		1			
Nagaland	1	0		1			
Puducherry	1	0		0	AIADMK 1		
Sikkim	1	0		0	SDF 1		
Tripura	2	0		0			CPIM 2
<b>Total</b>	<b>19</b>	<b>7</b>	<b>2</b>	<b>6</b>	<b>1</b>	<b>1</b>	<b>2</b>

## Indian Elections 2014 Forecast

Generated on 12-May-2014, 6:30 pm by Indian Elections Game (<http://msrindianelections.com>)

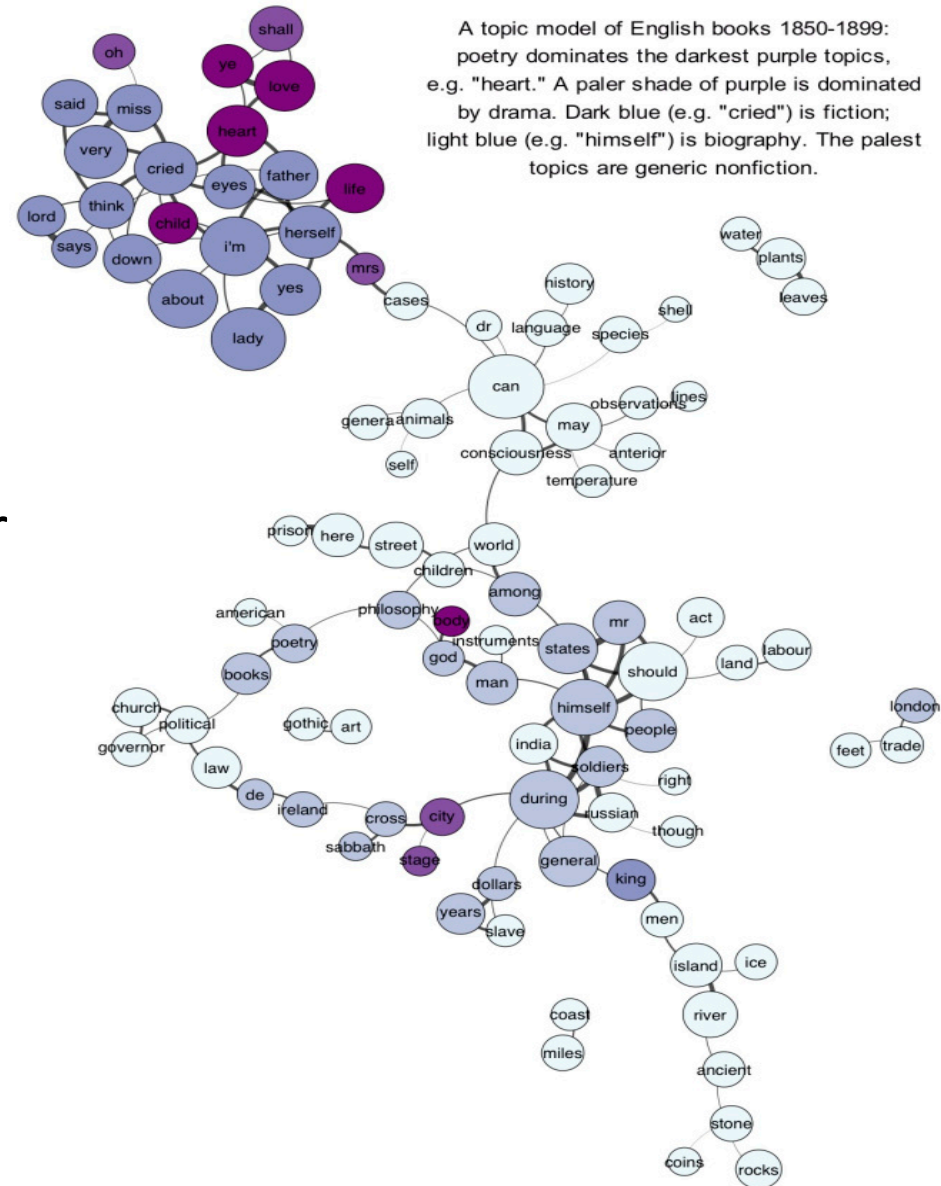
States	Seats	NDA		UPA		Third Front	Others	Error %
		BJP	NDA Allies	Congress	UPA Allies			
Andhra Pradesh	42	3	TDP 16	6			TRS 7 YSR 10	17%
Assam	14	5		6	AIUDF 1	AGP 1	BPF 1	23%
Bihar	40	20		3	RJD 9	JDU 8		14%
Chhattisgarh	11	8		3				11%
Delhi	7	4		1			AAP 2	20%
Gujarat	26	23		3				7%
Haryana	10	5	HJC 1	2			AAP 1 INLD 1	19%
Himachal Pradesh	4	3		1				15%
Jammu and Kashmir	6	2		1	JKNC 2		JKPDP 1	20%
Jharkhand	14	9		2	JMM 2		JVMP 1	18%
Karnataka	28	14		12		JDS 2		11%
Kerala	20	1		8	IUML 2	Left 9		17%
Madhya Pradesh	29	25		4				7%
Maharashtra	48	18	SS 14	9	NCP 6		AAP 1	11%
Odisha	21	6		2		BJD 12	AAP 1	14%
Punjab	13	3	SAD 5	4			AAP 1	23%
Rajasthan	25	22		3				7%
Tamil Nadu	39		BJP & 5-party 4	1		AIADMK 21	DMK 13	11%
Uttar Pradesh	80	44		5		SP 15	AAP 1 BSP 15	9%
Uttarakhand	5	4		1				16%
West Bengal	42	2		4		Left 11	AITC 25	11%
Small States/UTs	19	7	MPP 1 NPP 1	6	SDF 1	AIADMK 1	CPIM 2	
<b>Total</b>	<b>543</b>	<b>228</b>	<b>42</b>	<b>87</b>	<b>23</b>	<b>80</b>	<b>83</b>	<b>10%</b>



# Social Data Analytics (fundamentals)

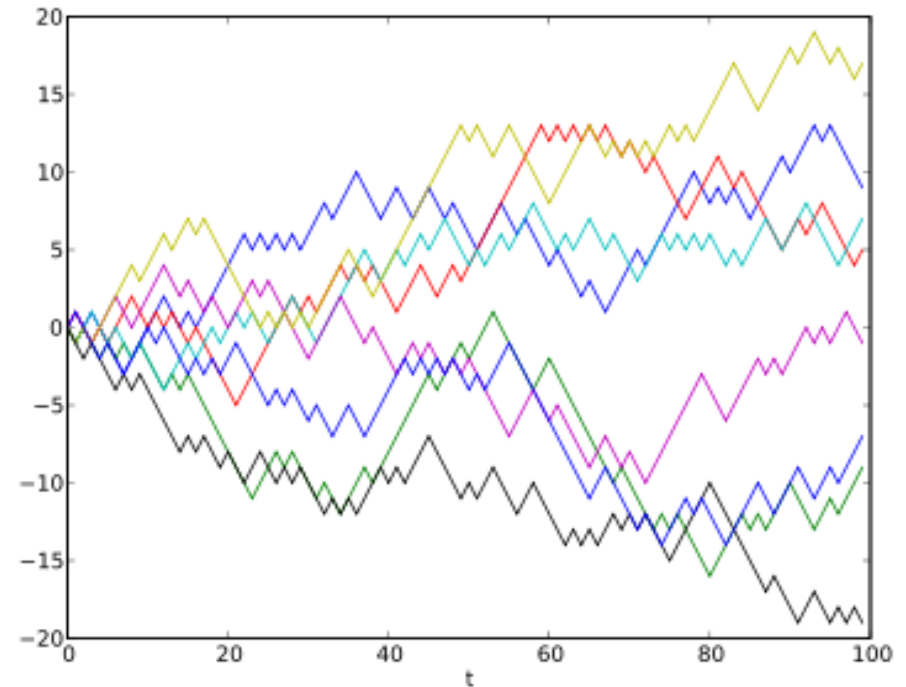
- Topic Model

- discovering the abstract "topics" that occur in a collection of documents
- "dog" and "bone" will appear more often in documents about dogs, "cat" and "meow" will appear in documents about cats, and "the" and "is" will appear equally in both
- LDA, PLSI



# Social Data Analytics (fundamentals)

- Random Walk
  - Markov Chains
  - Markov Processes
- Applications
  - Search the web
  - Sample OSNs
  - Mobility model
  - Image labeling and segmentation

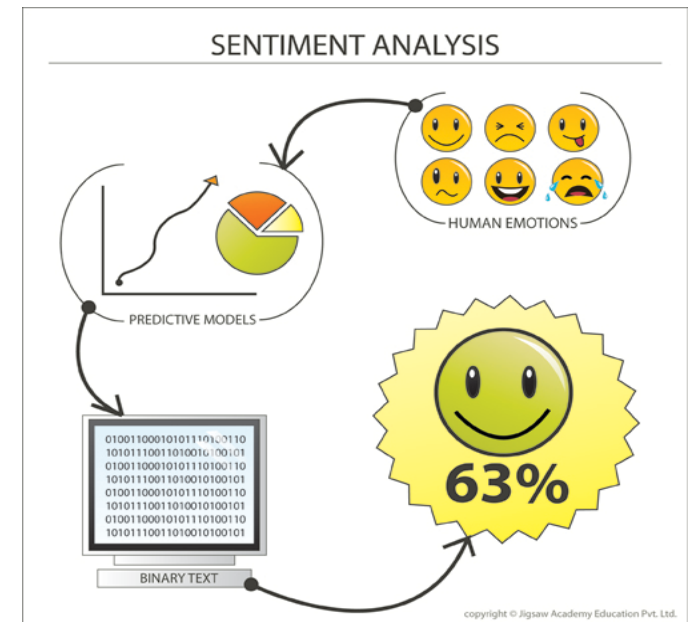


8 random walkers in 1D starting from 0



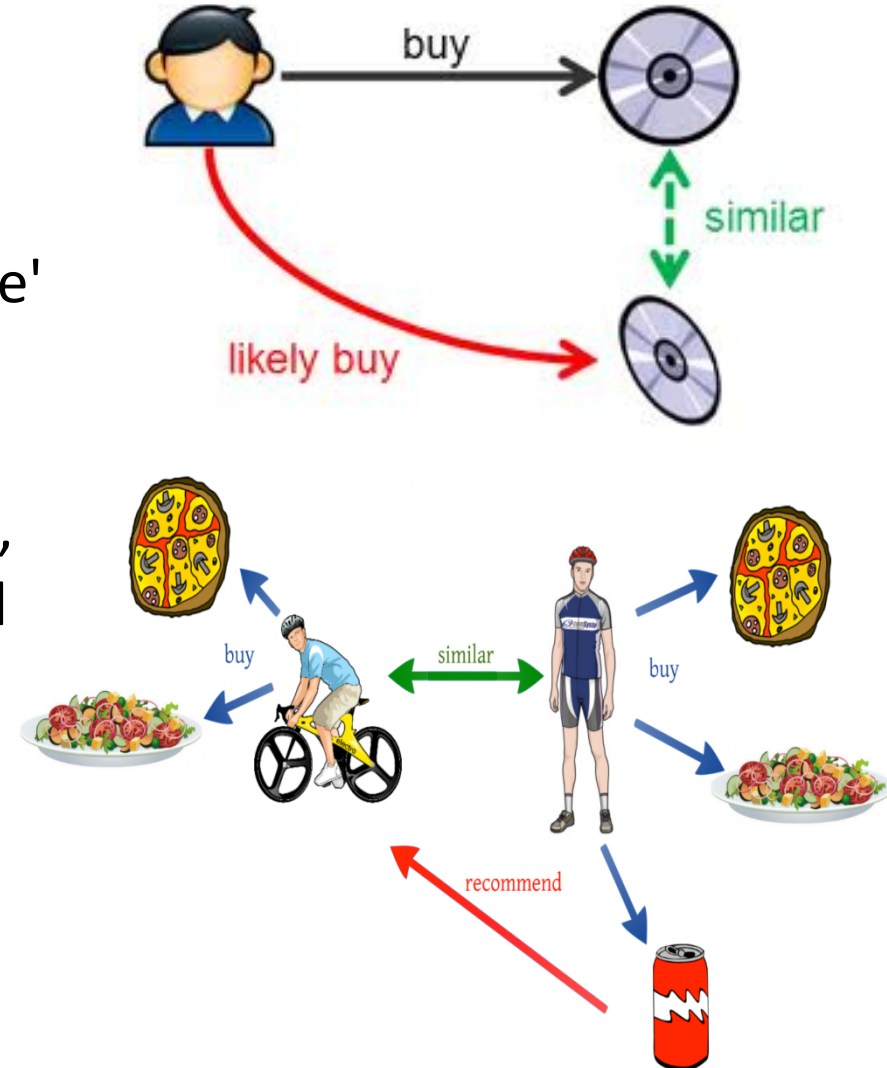
# Social Data Analytics (applied)

- Opinion mining and sentiment analysis
  - sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document



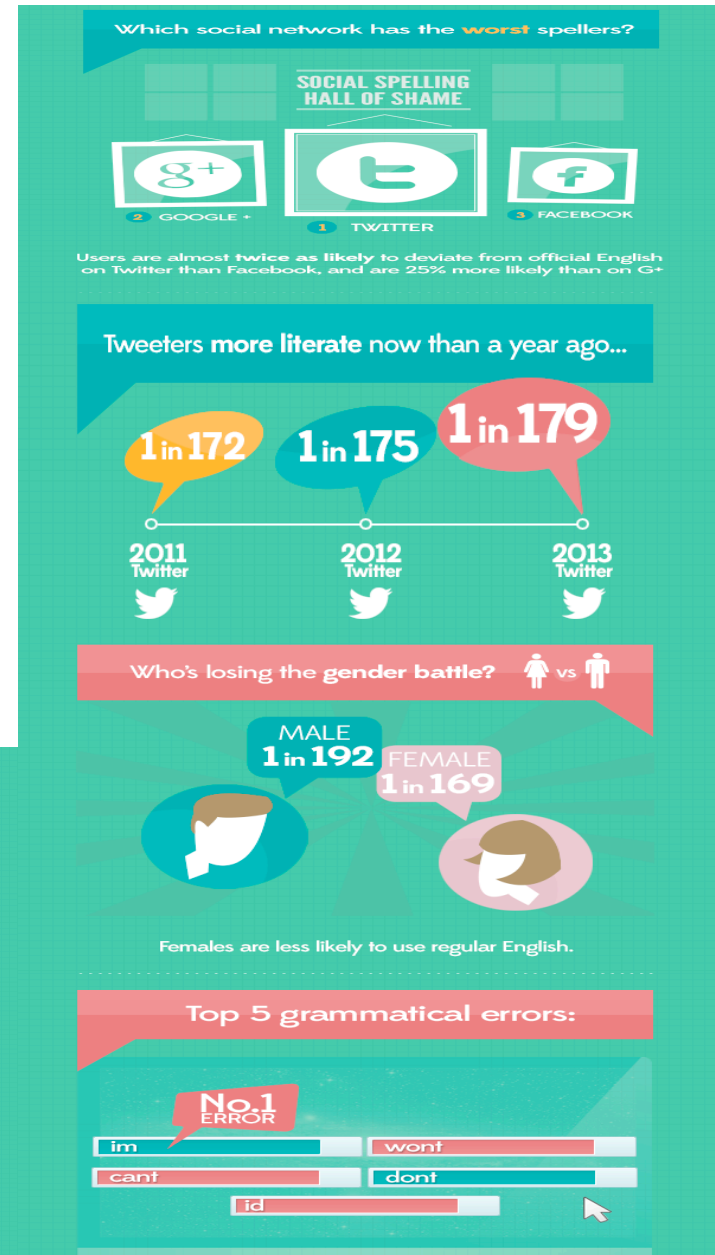
# Social Data Analytics (applied)

- Recommendation Systems
  - are a subclass of information filtering system that seek to predict the 'rating' or 'preference' that user would give to an item
  - movies, music, news, books, research articles, search queries, social tags, restaurants, financial services, life insurance, persons (online dating), and twitter followers



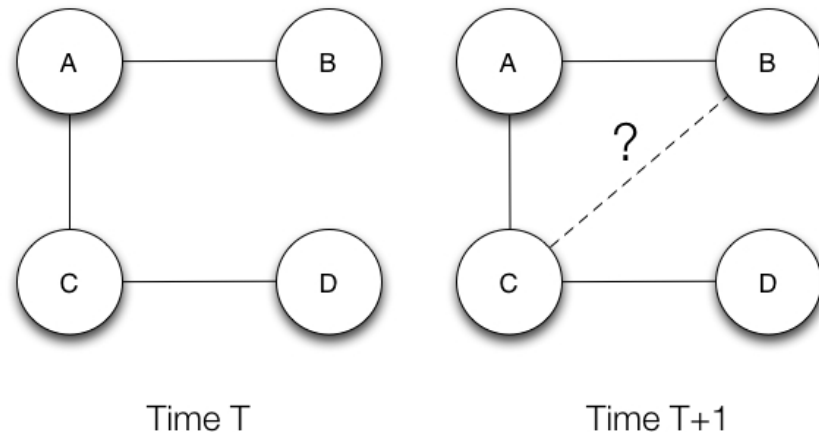
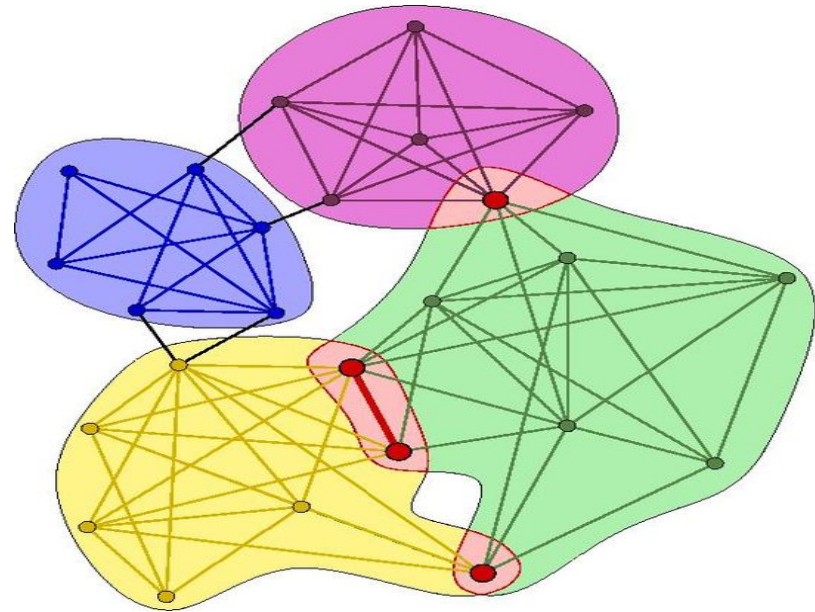
# Social Data Analytics (applied)

- Linguistics in online communities
  - Formal/Informal
  - Curse words/bad languages
  - Abbreviations
  - Slangs



# Social Data Analytics (applied)

- **Community Detection/ Link Prediction**
  - a network is said to have **community structure** if the nodes of the network can be easily grouped into (potentially overlapping) sets of nodes such that each set of nodes is densely connected internally
  - Given a snapshot of a social network, can one infer which new interactions among its members are likely to occur in the near future – **link prediction problem**



# Facebook Data Science Initiative



**Facebook Data Science** ✓  
Research Service

Timeline About Papers Reviews More ▾

Liked ▾ Following Message ⋮



## PEOPLE >



309,644 likes

5,547 visits



Facebook Data Science shared a note.

July 17 🌐

### Checking in at the World Cup and Leaving with Friends

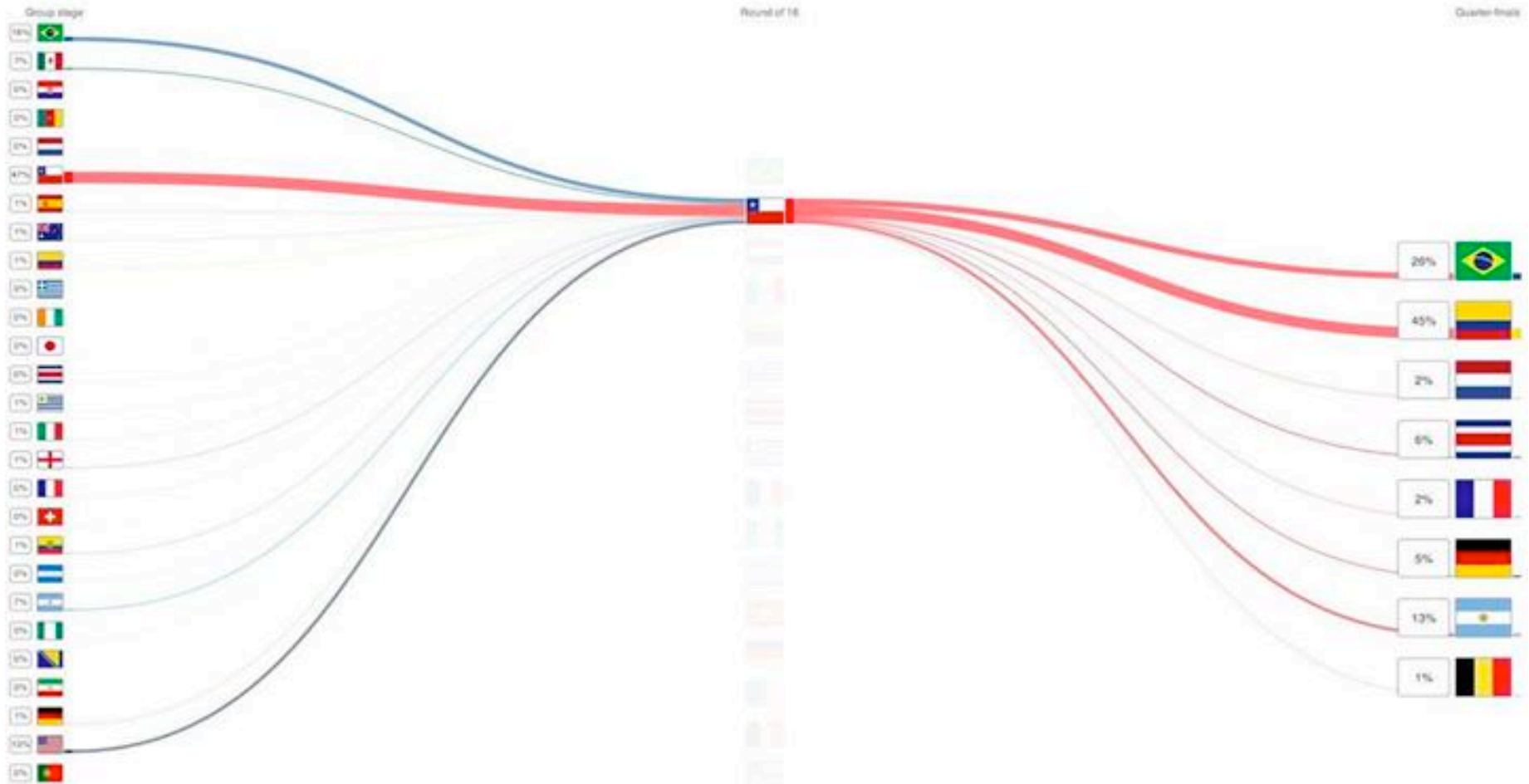


Hundreds of thousands of people from around the world traveled to the 2014 World Cup and we wanted to take a closer look at those who used Facebook to



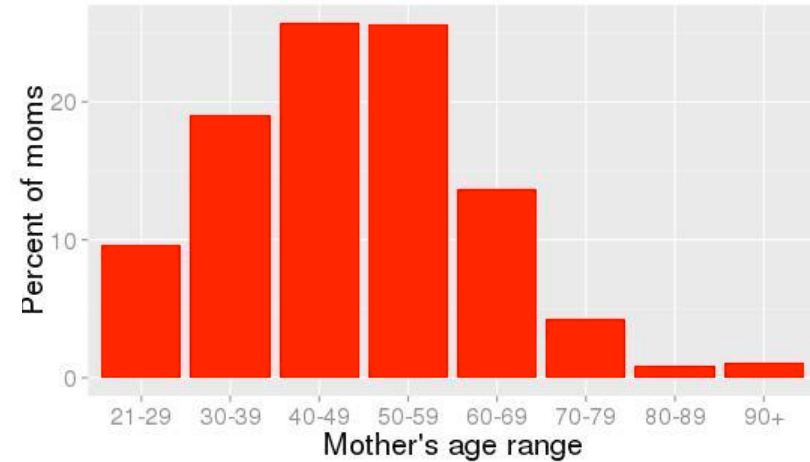
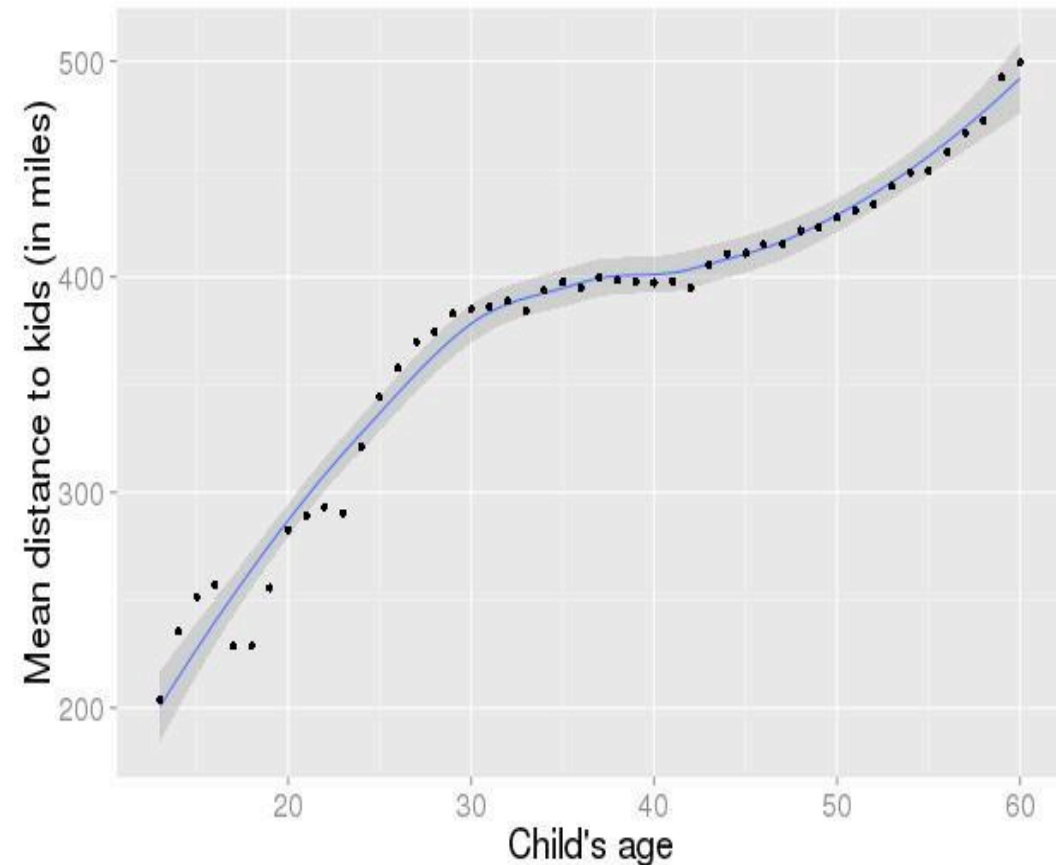
# Some Revelations

- World Cup Fan Migration



# Some Revelations

- Mother's Day 2014







# Some Revelations

- Springing Forward

