

Social Media Mining

Alok Choudhary

Henry and Isabel Dever Professor

EECS and Kellogg School of Management
Northwestern University

choudhar@eecs.northwestern.edu

Founder and Chief Scientist

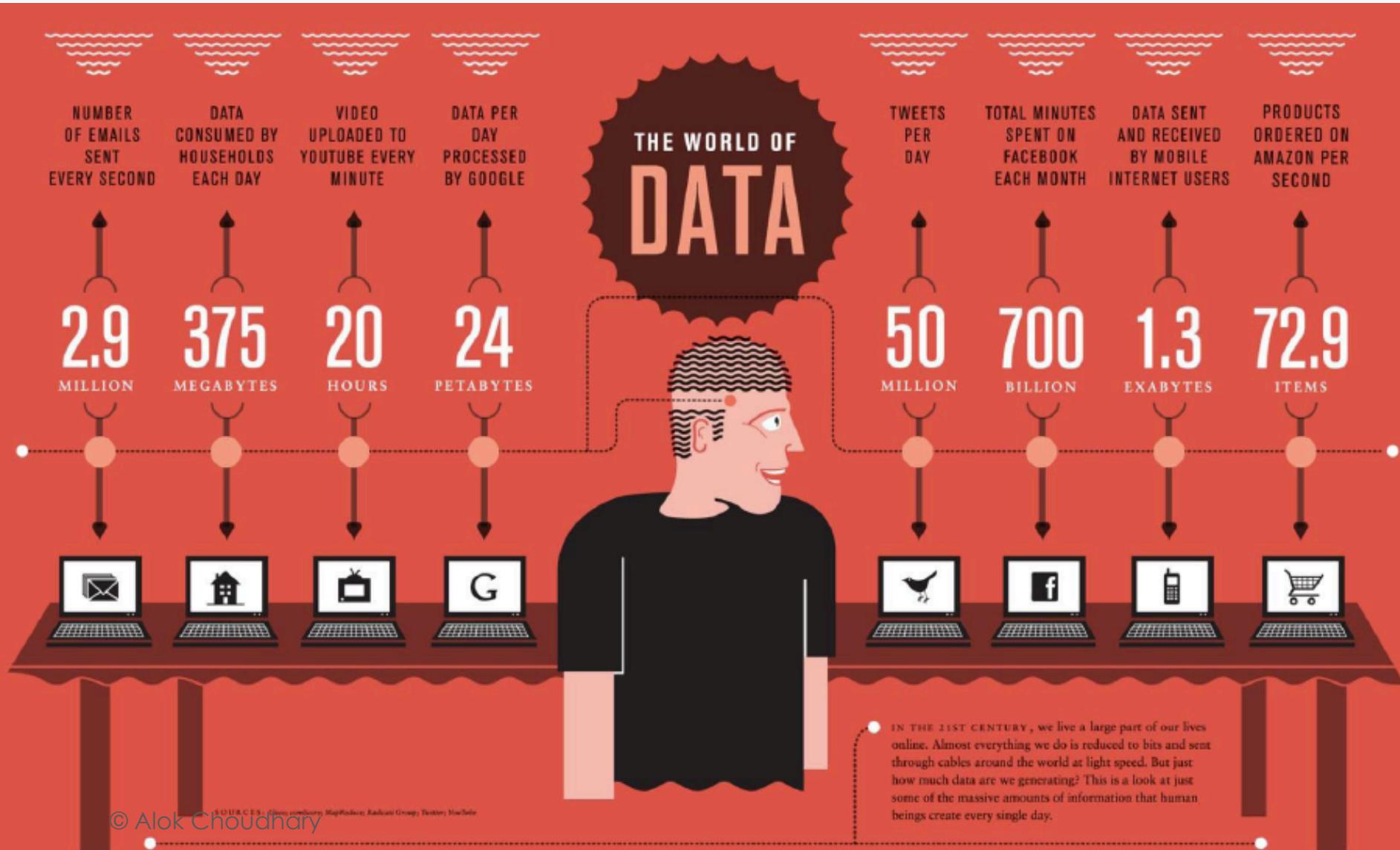
4C: A Big Data Science Company

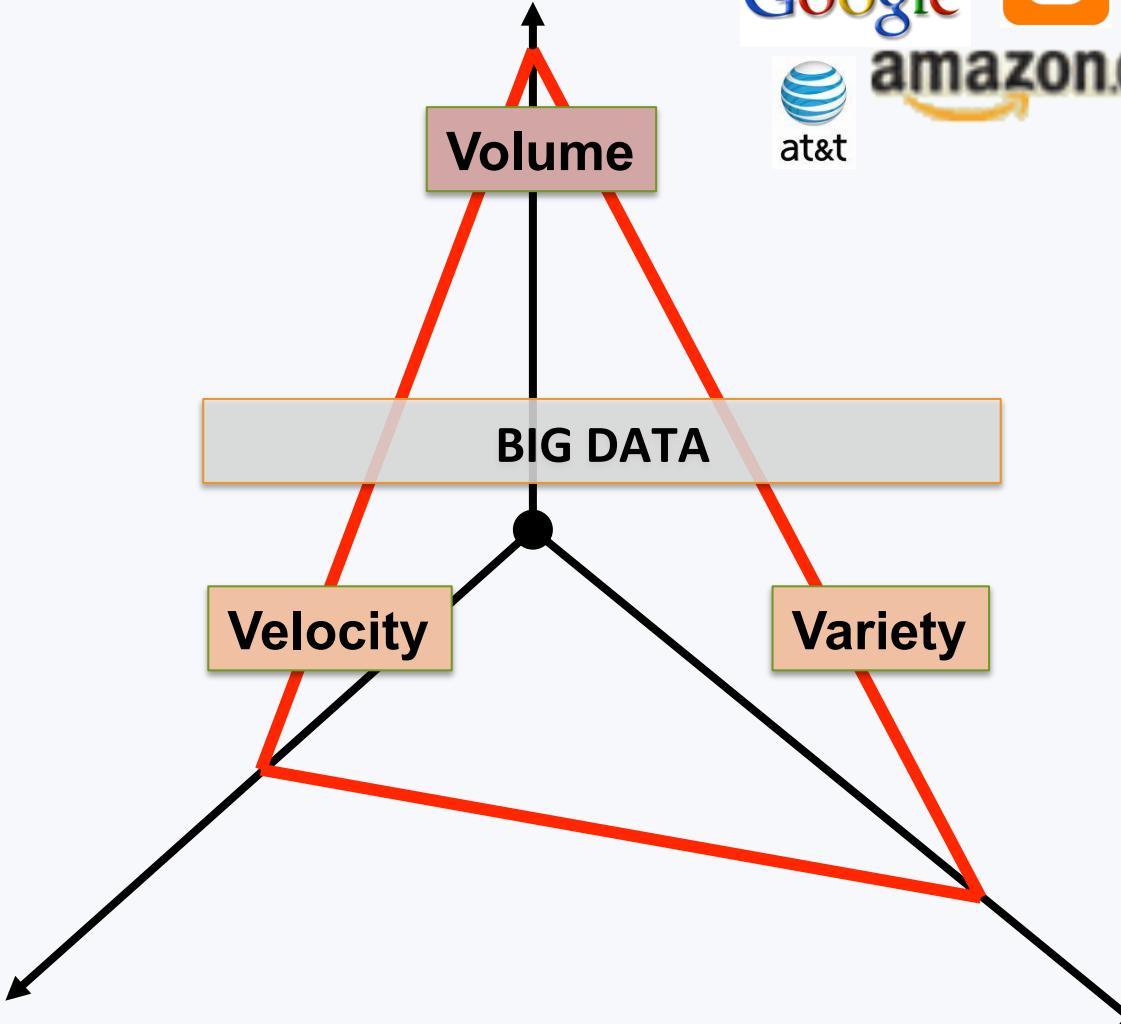
+1 312 515 2562

alok@4Cinsights.com

Team members for this class: Prof. Ankit Agrawal, Rosanne Liu (TA), Amar Krishna, Reda Al-Bahrani, Dianwei Han, Arindam Paul

Big Data ...Popular View.. Streaming..





Social Media

Definition

Social Media is the use of communication, Internet and tools for the purpose of sharing and discussing information and experiences in a scalable and efficient way.



Facebook

facebook    



[Home](#) [Profile](#) [Account](#) 



Mark Zuckerberg

Has worked at Facebook • Studied Computer Science at Harvard University • Lives in Palo Alto, California • From Dobbs Ferry, New York • Born on May 14, 1984



Education and Work

Employers

 **Facebook**
Feb 2004 to present • Palo Alto, California
[• FBX Profile](#)

College

 **Harvard University**
Computer Science · Psychology

- CS182. Intelligent Machines with Andrew Bosworth
- CS121. Introduction to Computational Theory with James Wang and Kang-Xing Jin

High School

 **Ardsley High School**

 **Phillips Exeter Academy**
Class of 2002

Philosophy

Favorite Quotes

"All children are artists. The problem is how to remain an artist once he grows up."

You and Mark



3 Mutual Friends

Sponsored [Create an Ad](#)

Police Auctions [gsaauctions.gov](#) 

Like a slick deal? Now you can get up to 90% retail with police seized auctions. Get in on the action.



SF Bucket List [partners.livingsocial.com](#) 

Things to do in San Francisco before you die. One huge coupon emailed daily.



Stay close to your team  

Check the score and see highlights from the big game with AT&T High Speed Internet for only \$14.95/mo.

Craft Beer Attorney  

Need legal assistance with your California craft beer?

Twitter

Gmail - Shipment Notification × www.fedex.com/Tracking?ac × Twitter / Who to follow / Interests × Shorten & Share | bit.ly | a si × Verizon details additional 49 ×

← → C twitter.com/#!/who_to_follow/interests/technology

Apple WordPress This Google Maps YouTube Wikipedia VEVO | Music Facebook Google News New Tab Other Bookmarks

twitter Search Home Profile Messages Who To Follow dgreatwood

Who to follow

View Suggestions **Browse Interests** Find Friends

View Technology (60)

Art & Design · Books · Business · Charity · Deals · Entertainment · Family · Fashion · Food & Drink · Funny · Health · Music · News · Politics · Science · Sports · Staff Picks · Staff Picks: NFL Playoffs · Technology · Travel · Twitter

guardiantech Guardian Tech
News and comment from the Guardian's technology team [Follow](#)

kaifulee Kai-Fu Lee I am the CEO of Innovation Works (www.innovation-works.com). I used to work at Microsoft, SGI, Apple, and Google. I have written several Chinese books. [Follow](#)

sacca Chris Sacca I'm an investor in Twitter & other startups. Learn more at www.lowercasellc.com. Just finished bicycling across the USA and am thinking about my next adventure. [Follow](#)

kevinrose Kevin Rose Tech angel investor. Founder of digg.com, wefollow.com. Random ideas, entrepreneur, climber of rocks, video blogger, & tea drinker. [Follow](#)

arsTechnica Ars Technica The Art of Technology [Follow](#)

lessig Lessig law prof, reformer. [Follow](#)

leolaporte Leo Laporte Podcaster, broadcaster and tech pundit. The Tech [Follow](#)

You recently followed · [View All →](#)

GlobalFundWomen Global Fund forWomen · Follow
Largest global grantmaker exclusively funding women's rights. Committed to a world of equality & social justice.

nockee Chris · Follow
Ultracyclist and endurance junkie. Did Race Across Oregon and Hoodoo 500 Voyager in 2009. Doing RAW solo in 2011

You recently viewed · [Clear](#)

StartupReport · Follow Kevin Steger

pierre · Follow Pierre Omidyar

cote · Follow cote

guardiantech · Follow Guardian Tech

spoonen · Follow Sanjay Poonen

Stratalux · Follow Stratalux

Invite Friends

Not finding who you're looking for? Invite friends to Twitter via email. [See what you'll send them.](#)

your friend's email address [Invite](#)
Lots of people to invite? Separate multiple email addresses with commas.

Related Services

[Microsoft](#) · Bing Social's Recommendations for Influential Twitter Users.

[LinkedIn](#) · Find your LinkedIn colleagues who use Twitter.

About · Help · Blog · Status · Jobs · Terms · Privacy · Shortcuts
Advertisers · Businesses · Media · Developers · Resources · © 2011 Twitter

Pinterest



high

heels

women



Stilettos

Christian
Louboutin

Black

Products

Fashion

Red Bottoms

Pumps

Christmas
Gifts

All Pins

Your Pins

Pinners

Boards



from PlayBuzz

10<sup>ful Hacks To Make Your
More Comfortable</sup>

N@b£|¥ omg these heels
need to get in my closet, right now !
Black and beige classic high heels.



from Sole Society

Avalon t-strap pump -



@Ralph & Russo high heels

from Instagram

1521 153

Theresa D. Smith
Shoes and Bags



Instagram

Instagram



3DKS2-VgzKu/

YouTube

YouTube

Upload

Home

My Channel

Trending

Subscriptions

History

Watch Later

Add channels

Popular on YouTube

Music

Sports

Gaming

Browse channels

Manage subscriptions

YouTube Red

Home

Trending

Subscriptions

Show ad ▲

Recommended



Meet the brains that power BuzzFeed's tech and the Ne...

Re/code

659 views • 3 weeks ago



CNN Election Center Cold Open - SNL

Saturday Night Live

2,709,844 views • 2 weeks ago



Comedy Movie from Bollywood - Best Comedy ...

Shemaroo

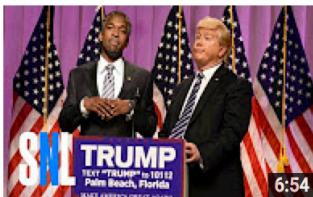
22,604,382 views • 4 years ago



Christian T. Petersen Interview | Founder/CEO ...

Team Covenant

5,810 views • 2 months ago



Carson Endorsement Cold Open - SNL

Saturday Night Live

2,056,082 views • 1 week ago



When Raju Shrivastav mimicked Lalu Yadav in front of ...

Headlines Today

5,713,599 views • 3 years ago



MLB Media's Bob Bowman at Code/Media 2016 (full)

Re/code

219 views • 3 weeks ago



Superagent Casey Wasserman at Code/Media ...

Re/code

724 views • 3 weeks ago

Show more

Disco Recommended videos for you

Subscribe 81,066

Amazon

Amazon.com: Online Shopping

www.amazon.com

amazon Your Amazon.com | Today's Deals | Gift Cards | Help

Shop by Department ▾ Search All ▾ Go

Hello, Sign In Your Account ▾ Cart ▾ Wish List ▾

FREE Two-Day Shipping Join Amazon Prime Today

Amazon Instant Video: The latest TV shows, available instantly. EXPAND TO LEARN MORE

Instant Video MP3 Store Cloud Player Kindle Cloud Drive Appstore for Android Digital Games & Software Audible Audiobooks

The All-New kindle fire HD The ultimate HD experience From \$199 >[Shop now](#)

Introducing kindle paperwhite The world's most advanced e-reader From \$119 >[Shop now](#)

Bikes with Street Cred Clothing Trends Amazon Prime

THE AMAZON CLOTHING STORE DRESS SHOP

The season's biggest trends from Calvin Klein, BCBGMAXAZRIA, Tracy Reese, and more.

>[Shop Dresses](#) >[Shop All Clothing](#)

Friends & Family Gifting

Gift Connections Made Easy [Learn more](#)

\$0 Hundreds of Free Songs Play on the Web, Android, and Kindle Fire

>[Shop now](#) amazon MP3

Advertisement

Gold Box Deal of the Day

\$1.99 Kindle Inspirational Memoirs and More



Canon gets it right with this mid-range SD series camera.
Small and versatile

By Anjana Nigam

Last year, Canon SD870IS camera was a winner in the SD series and the SD880IS manages to offer even more at a lower starting price (than that of SD870IS last year). **GENERAL POSITIVE COMPARATIVE**

A BIGGER CCD - 1/2.3" up from the previous 1/2.5" in the SD870 which allows for increase in MPs to 10MP (up from 8MP of SD870).

FEATURE: RESOLUTION , POSITIVE COMPARATIVE

- 4X Zoom (up from 3.8X on SD870) and a wide angle which really helps with shots at close quarters (in a car/bus, of a large group etc.

FEATURE: LENS , POSITIVE COMPARATIVE

- The huge 3 inch LCD screen is the same as the SD870 but slightly improved technology gives you brighter playback in low-light which makes it easier to see the shots in replay.

FEATURE: LCD , POSITIVE COMPARATIVE

It does not have a viewfinder but I have never missed the viewfinder on my SD870 since Canon made the screen anti-glare. I'd take a bigger screen any day over a 2.5 inch screen with a viewfinder, but if you like to use the viewfinder to compose your shots you might need to look at the SD990 IS and drop another 100 bucks.

FEATURE: LCD , NEGATIVE COMPARATIVE

amazon.com Hello, Alok N Choudhary. We have recommendations for you. (See more) Alok's Amazon.com | Today's Deals | Gifts & Wish List

Shop All Departments Search Camera & Photo

Customer Reviews Flip UltraHD Camcorder, 120 Minutes (Black) by Flip Video

Average Customer Rating ★★★★★ (530 customer reviews)

Star Rating	Count
5-star	(242)
4-star	(148)
3-star	(66)
2-star	(58)
1-star	(73)

Create your own review

The Most Helpful Reviews

The most helpful favorable review 1,124 of 1,134 people found the following review helpful:
Helpful The camcorder SHOWDOWN: I've done the product comparisons for you. If I'm going to spend more than a hundred bucks on an item or somewhere around there, I do extensive research first to know I get the best bang-for-the-buck and, consequently, dodge any potential future buyer's remorse.

I've realized that the time I spend doing my product comparisons is often time that others don't have so I may as well share what I can... [Read the full review](#). Published 4 months ago by A. Chandler

[See more 5_star, 5_star reviews](#)

The most helpful critical review 1,178 of 1,112 people found the following review helpful:
Helpful Flip Ultra HD vs. Flip Minidv I previously owned the Flip Minidv as this review is based on a comparison between the Flip Minidv and the new Flip UltraHD. After reading about the Flip UltraHD's addition of an HDMI output, larger screen, and longer battery life I decided to sell my 4 month old Flip Minidv and purchase the Flip UltraHD. Since the Minidv is very sleek and diminutive in...

[Read the full review](#). Published 10 months ago by Jeff

[See more 3_star, 3_star, 3_star reviews](#)



Online Social Networking

Online Social Networks are web-based services that allow individuals and communities to connect with real world friends and acquaintances online

- Interactions
 - Friendship interaction
 - Friends, like, comments, ...
 - Media Sharing
 - Sending and receiving messages

- Examples
 - Facebook.com
 - Twitter.com
 - Instagram



Blogging

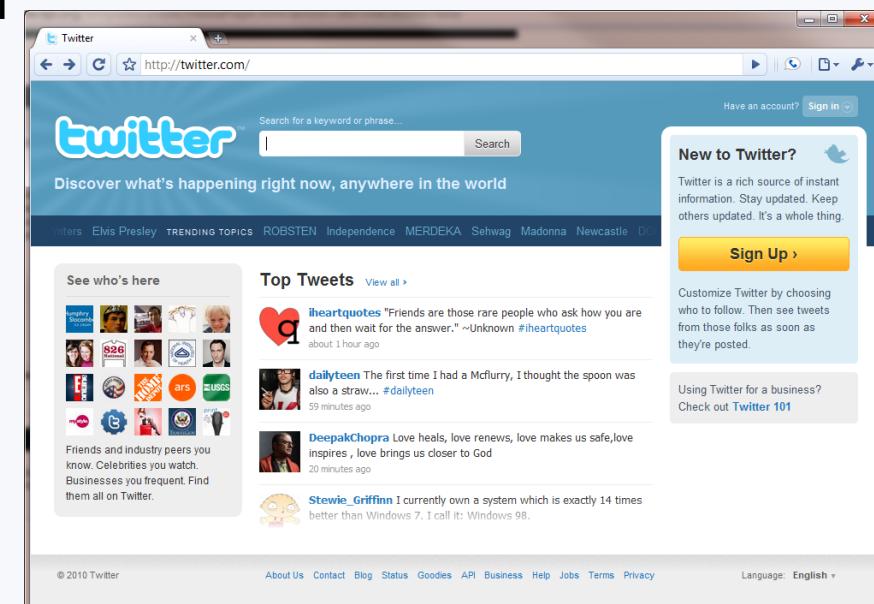
A blog is a journal-like website for users, a.k.a. bloggers, to contribute textual and multimedia content, arranged in reverse chronological order

- Maintained both individually or by a community Usages:
 - Sharing information and opinions with friends and strangers
 - Disseminating subject-specific content

Microblogging

Microblogging is a relatively new phenomenon that can be considered as a counterpart to blogging, but with limited content

- Usage
 - communication medium
 - social interaction
 - citizen journalism
- Service Providers:
 - Twitter
 - Tumblr



Wiki

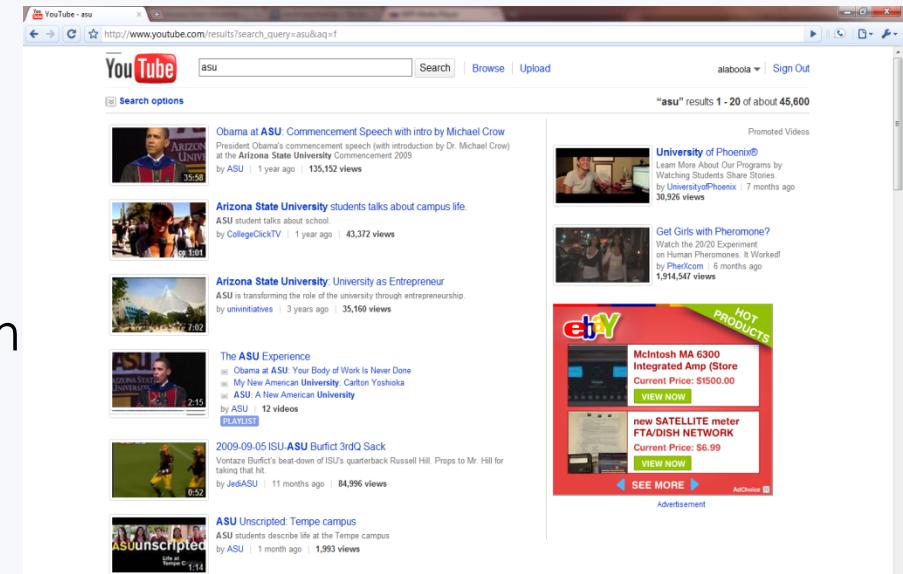
A wiki is a collaborative editing environment that allows users to develop Web pages using a simplified markup language

- Wikipedia allows interested individuals to collaboratively develop articles on a variety of subjects.
- Using the wisdom of crowds effectively, it has become a comprehensive repository of information useful to a variety of individuals

Media Sharing

Media sharing is an umbrella term that refers to the sharing of a variety of media on the web. Users share such multimedia content of possible interest to others

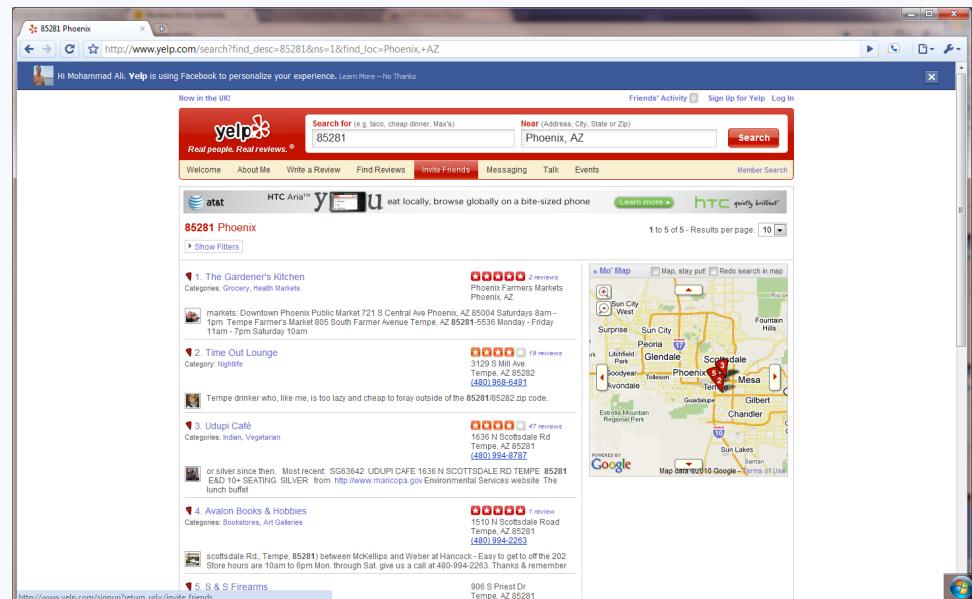
- Samples:
 - Video Sharing:
 - YouTube.com
 - Photo Sharing:
 - Flickr.com, instagram
 - Document Sharing:
 - Scribd.com, Slideshare.com
 - Livecasting:
 - Justin.tv, Ustream.com



Opinion, Review, and Ratings Websites

Opinion, review, and ratings websites are websites whose primary function is to collect and publish user-submitted content in the form of subjective commentary on existing products, services, entertainment, businesses, places, etc. Some commercial sites may serve a secondary purpose as review sites by publishing product reviews submitted by customers.

- Examples
 - Cnet.com
 - Epinions.com
 - yelp.com
 - tripadvisor.com

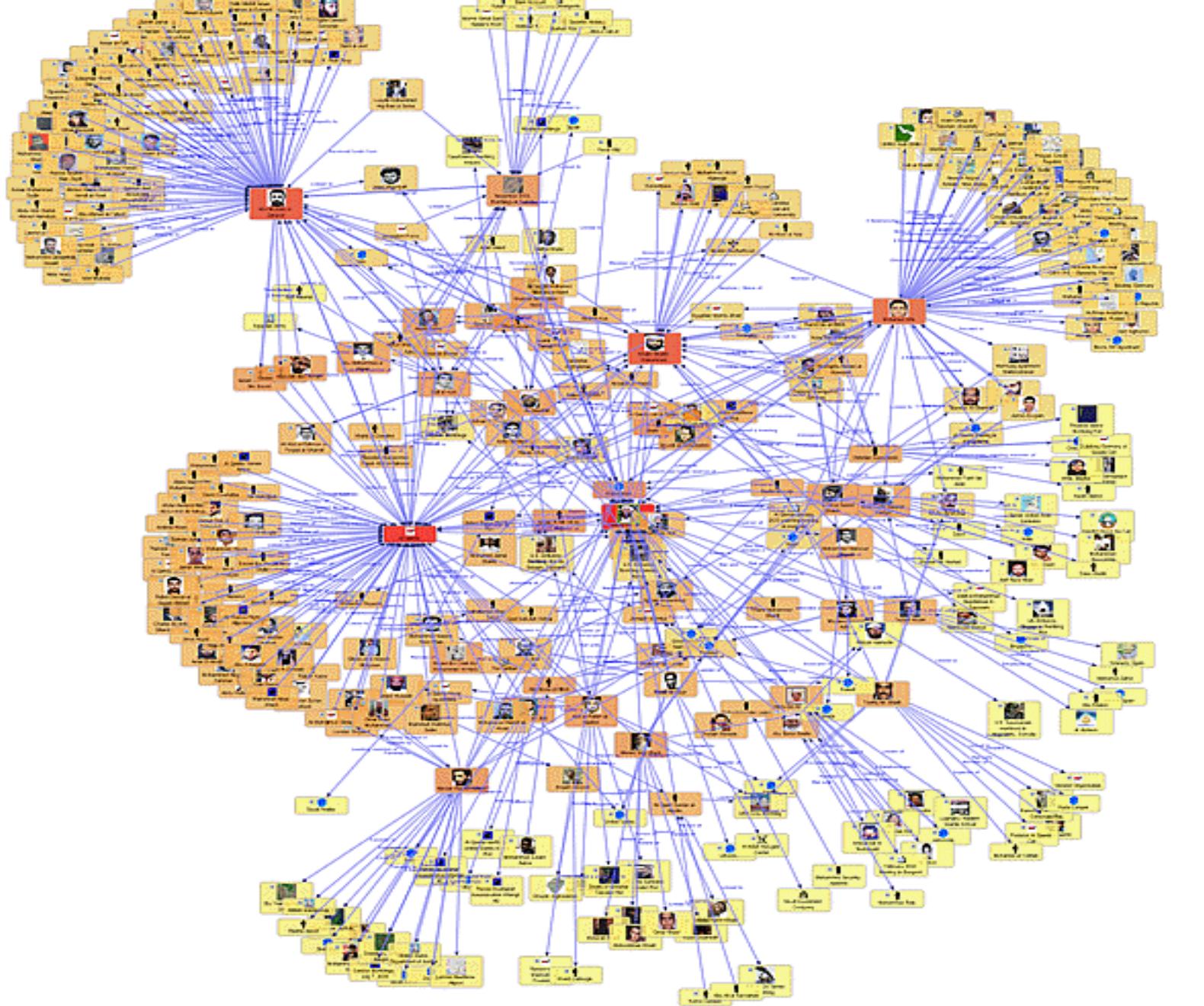


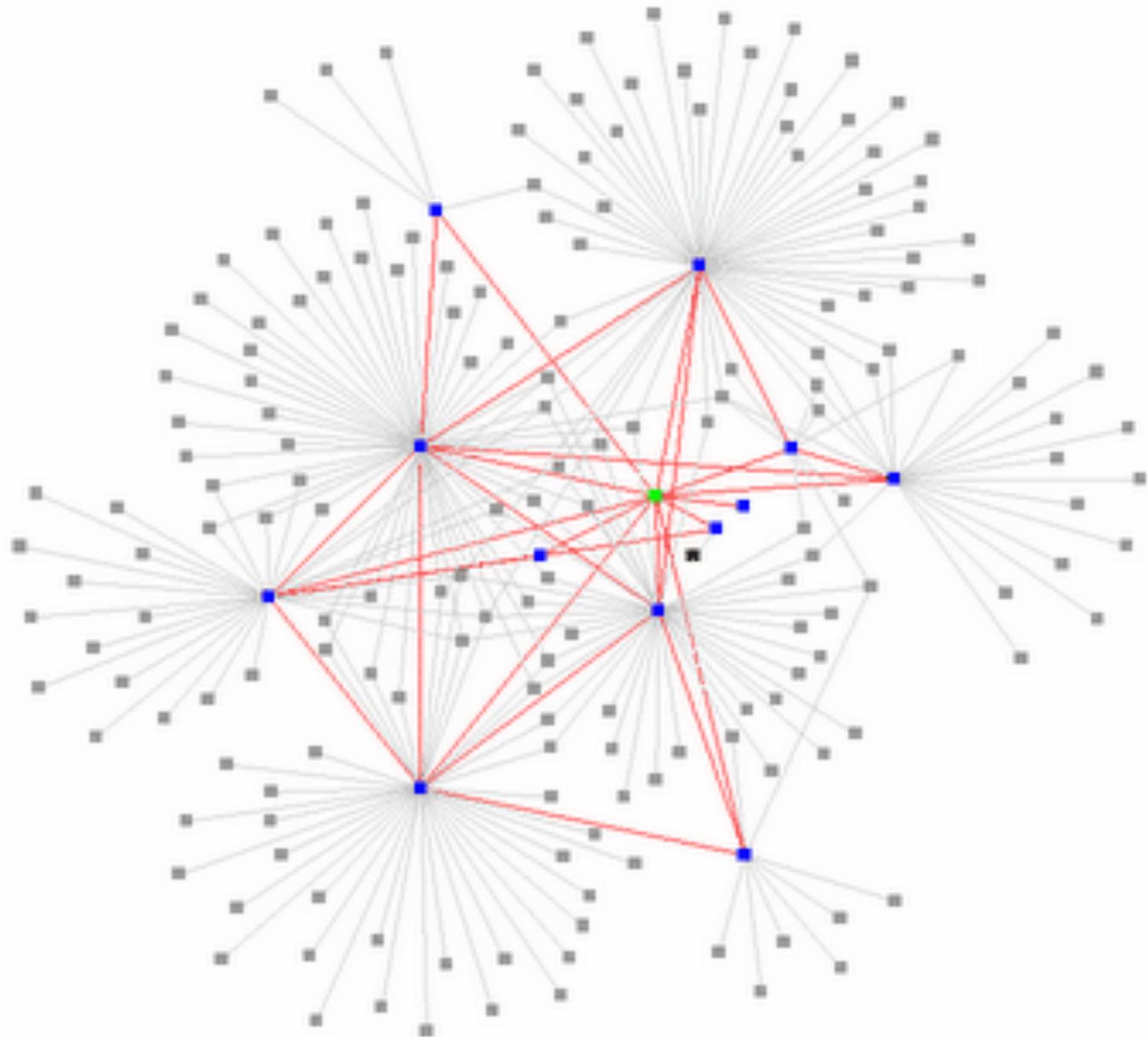
Main Characteristics

- Participation
 - social media encourages contributions and feedback from everyone who is interested. It blurs the line between media and audience.
- Openness
 - most social media services are open to feedback and participation. They encourage voting, comments and the sharing of information. There are rarely any barriers to accessing and making use of content – password-protected content is frowned on.
- Conversation
 - whereas traditional media is about “broadcast” (content transmitted or distributed to an audience) social media is better seen as a two-way conversation.
- Community
 - social media allows communities to form quickly and communicate effectively. Communities share common interests, such as a love of photography, a political issue or a favorite TV show.
- Connectedness
 - Most kinds of social media thrive on their connectedness, making use of links to other sites, resources and people.

Social Media Mining is the process of representing, analyzing, and extracting meaningful patterns and insights from social media data

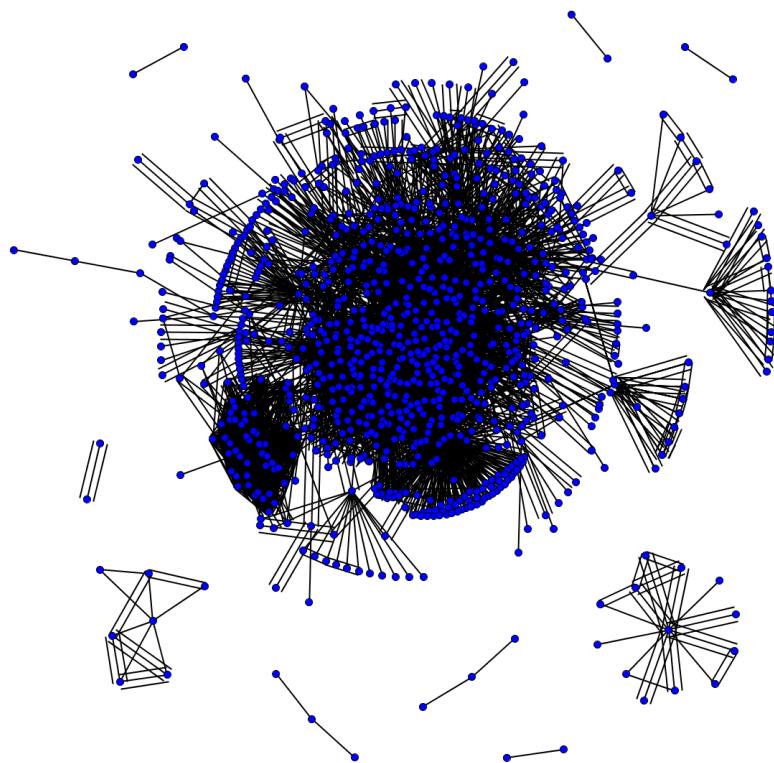




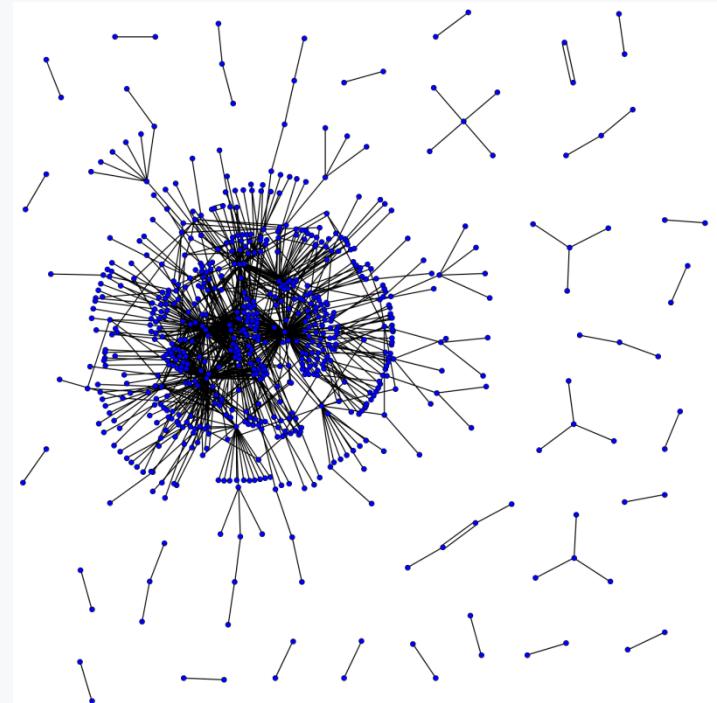


Slow and Fast Changing Dynamics in a Network

Static (Slowly Changing)
Social Network(1000 users)



Dynamic Topic based
response network



Brands on Facebook..

Brand: Any social entity, e.g. presidential candidates, movies/TV Shows, doctors, companies, athletes, teams, business schools, products(cameras, drinks, cars..). <https://www.facebook.com/barackobama>



Number of fans
(this is the number of fans who have ever liked this page)



https://www.facebook.com/CocaColaUnitedStates?brand_redir=1

Brand Active Engagers (who post, like posts and comments, like posts..., tweet, retweet, fav, etc.. That is, interact with a Brand)

- Where do you think Facebook's data can be used?

POST



POST LIKE

COMMENT



The Class

• • •

Everything on Canvas

Office Hours: By appointment T or TH

Class Outline (Lectures)

1. Introduction to Course: What is Social Media Mining?
2. Graphs: Graph Basics, Graph Representation, Special Graphs
3. Link prediction, who-follows-who problem in social networks
4. Data Mining Essentials 1: Data Mining Basics, Data, Data Preprocessing
5. Data Mining Essentials 2: Tutorial Data Mining in Python, IPython Notebook
6. Tools: Weka for Data Mining
7. Data Mining Essentials 3: Introduction to large scale data mining: Map Reduce, Spark, and Storm

Class Outline (Lectures)

8. NLP, Text Mining, document classification with [MALLET](#)
9. Influence and Applications, Topic Modeling
10. Introduction to Deep Learning, with demos of theano, torch, tensorflow, keras
11. Part 1: continue with deep learning demos; Part 2: Social applications of deep learning
12. Social media mining in practice: web scraping and the “Like”
13. Big Data Science and Social Media Business Applications
14. Lecture 14-19 paper presentations

Student paper presentations and discussion

- Guideline on [How to Read a Paper](#)
- A team of two students will present both papers in the topic categories listed below.
- Each student (other than the presenters) should submit a two-paragraph summary of the **best** paper **PRIOR TO** the class time in which the topic is being discussed.
- The paper summaries should cover the following:
- A brief summary of the paper and its contributions
 - Problem solving approach (concepts may or may not have covered in class).
 - At least one area for improvement in the paper. A suggestion for "next steps" in the same direction that would make interesting future work.
 - A brief assessment of whether you liked the paper, overall.

The Course Requirements

- The course is composed of readings, assignments and a class project.
- Group Presentation and class participation 40%
 - 10% is for paper summaries
- Project 60%
 - 10% is for preliminary report

Project

- **Project proposals** are due lecture 7 at 11:59PM, and should be about 1 page in length (single spaced). It should contain title, team member names, project goals and motivation, and related work.
- **Preliminary Reports** are due lecture 13 at 11:59 PM. Each group must submit a two-page summary of their project progress by discussing 1) Steps you have completed, and any results you have obtained so far 2) The key remaining steps you plan to complete before the end of the quarter 3) Any questions or concerns you have regarding the project.
- **Final Reports** are due June 9th at 11:59PM. The final report should be about 4-10 pages in length (single spaced) and should include your project goals and motivation, along with a concise and clear statement of what results you obtained. Also mention which aspects of future work would be most interesting. Clarity in your report and presentation will contribute significantly to your grade.
- **Software:** Software source with an ability to test results

Software

- Tableau (<http://www.tableau.com>)
- Weka (
<http://www.cs.waikato.ac.nz/ml/weka/>)
- iPython notebook
- Python libraries: numpy, scipy, keras, tensorflow, scikit-learn, nltk, matplotlib

Text and Reference Books

references:

- Social Media Mining, Reza Zafarani, Mohammad-Ali Abbasi, Huan Liu, Cambridge University Press,
- Networks: An Introduction, Mark Newman, Oxford University Press, 2010.
- Community Detection and Mining in Social Media, [Morgan & Claypool](#), 2010.
Networks, Crowds, and Markets: Reasoning about a Highly Connected World,
([available in pdf](#)), David Easley and Jon Kleinberg, Cambridge University Press 2010.
- [Exploratory Social Network Analysis with Pajek](#), Wouter de Nooy, Andrej Mrvar, Vladimir Batagelj, Cambridge, 2005, [Package and Manual](#)
- [Social Network Analysis](#) - Methods and Applications, Stanley Wasserman and Katherine Faust, Cambridge, 1994
- [Modeling and Data Mining in Blogosphere](#), Morgan & Claypool, 2009
- [Web Data Mining](#) - Exploring Hyperlinks, Contents, and Usage Data, Bing Liu, Springer, 2011 (2nd Edition)
- [Mining the Web](#) - Discovering Knowledge from Hypertext Data, Soumen Chakrabarti, Morgan Kaufmann, 2003

Next Week: Graphs and Networks

• • •