Social Media Mining

Introduction

Dear instructors/users of these slides:

Please feel free to include these slides in your own material, or modify them as you see fit. If you decide to incorporate these slides into your presentations, please include the following note:

R. Zafarani, M. A. Abbasi, and H. Liu, *Social Media Mining: An Introduction*, Cambridge University Press, 2014.
Free book and slides at **http://socialmediamining.info/**

or include a link to the website:
http://socialmediamining.info/

Facebook

- **How does Facebook use your data?**
- **Where do you think Facebook can use your data?**

What about Amazon?

Yelp?

English

Friends' Activity 0 | Sign Up for Yelp | Log In

yelp
Real people. Real reviews.®

Search for (e.g. taco, cheap dinner, Max's): san francisco restaurants

Near (Address, Neighborhood, City, State or Zip): San Francisco, CA

Search

- San Francisco Community Acupuncture
- San Francisco Food Bank
- San Francisco Symphony
- san francisco restaurants
- San Francisco Movie Tours
- San Francisco Marriott Fishermans Wharf
- San Francisco Gay Wedding Video
- San Francisco Soup Company
- San Francisco Test Only Smog
- San Francisco CityPASS

Welcome | About Me | Write a | Talk | Events | Member Search

Try Yelp in... | Hide

Are You Looking For Yelp Berkel

Amsterdam, Atlanta, Austin, Berlin, Boston, Chicago, Dallas, Denver, Detroit, Dublin, Honolu, Houstc, Philadelphia, Portland, Saint Louis, San Diego, San Francisco, San Jose, Seattle, Toronto, Vancouver, Washington, DC, More Cities »

Yelp Berkeley

Yelp is the fun and easy way to find and talk about great (and not so great) local businesses | Sign Up Now

Best of yelp | More "Best Of" »

Restaurants 4505 reviewed

1. La Bedaine
2. Kingston 11 Cuisine
3. Vital Vittles
4. Cheese Board Pizza
5. Emilia's Pizzeria

...see more »

Shopping 4852 reviewed

1. Chestnut & Vine Floral...
2. Waterside Workshops
3. UniFormal & UniEleganza Tuxedo...
4. Lee's Florist & Nursery
5. Supple Integrative Skin Care

...see more »

Nightlife 881 reviewed

Beauty and Spas 2566 reviewed

Browse by Category
- Restaurants
- Food
- Nightlife
- Shopping
- Beauty and Spas
- Arts & Entertainment
- Event Planning & Services
- Health and Medical
- Active Life
- Hotels & Travel
- Automotive
- Home Services
- Local Services
- Local Flavor
- Pets
- Education

Get the Yelp app on your mobile phone

It's free and helps you find great, local businesses on the go!

Get it for Free Now

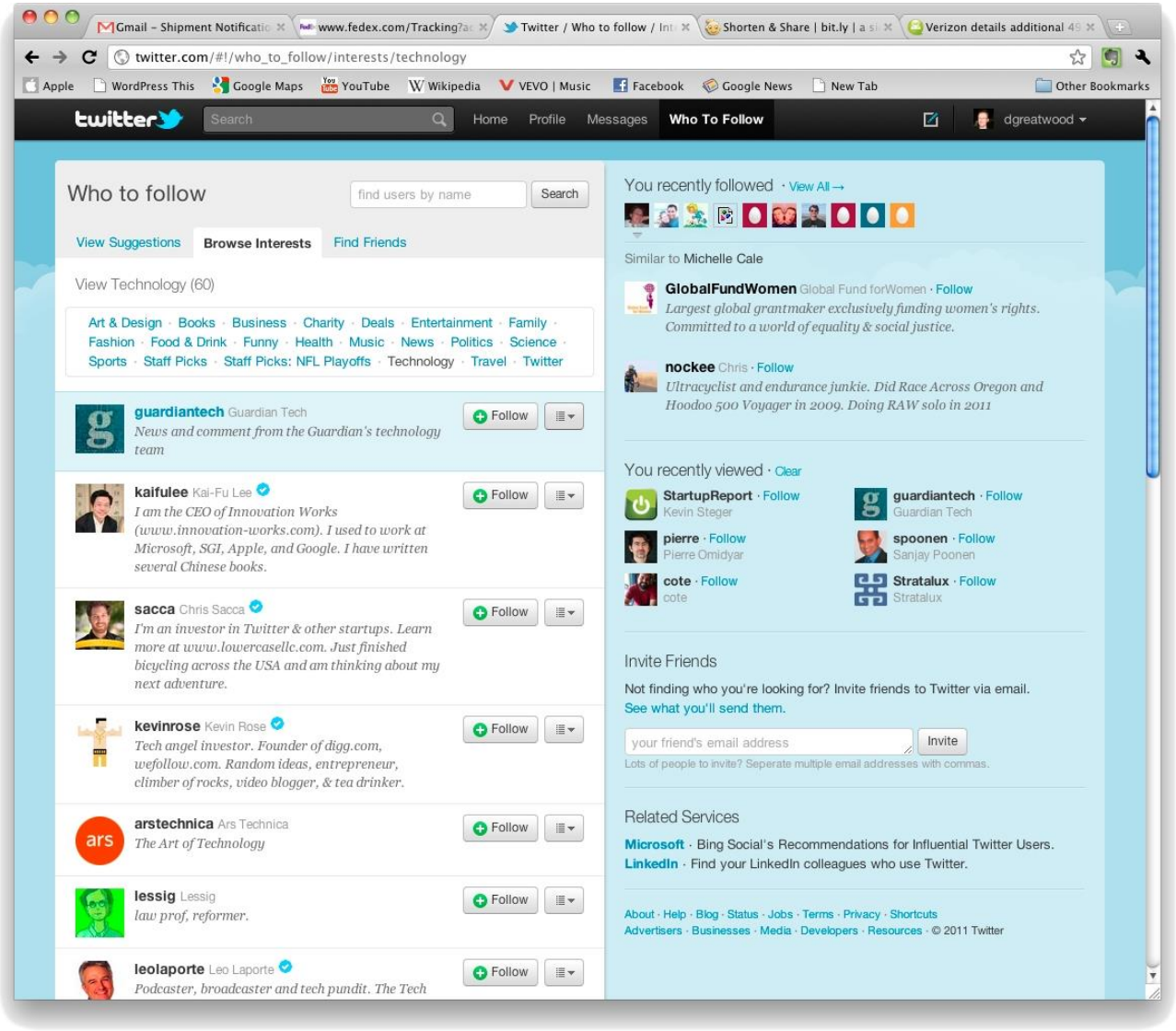
Review of the Day | Archive »

Voted by our members!

Dopo

*Cozy, romantic little neighborhood spot: Nice location right across from good ol' Fenton's. No reservations at this little place, so either come a little

Or Twitter?


twitter.com/#!/who_to_follow/interests/technology
twitter
Search
Home
Profile
Messages
Who To Follow
dgreatwood
Who to follow
find users by name
Search
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
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.
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.
kevinrose Kevin Rose
Tech angel investor. Founder of digg.com, wefollow.com. Random ideas, entrepreneur, climber of rocks, video blogger, & tea drinker.
arstechnica Ars Technica
The Art of Technology
lessig Lessig
law prof, reformer.
leolaporte Leo Laporte
Podcaster, broadcaster and tech pundit. The Tech
Follow
You recently followed · View All →
Similar to Michelle Cale
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
guardiantech · Follow
Guardian Tech
pierre · Follow
Pierre Omidyar
spoonen · Follow
Sanjay Poonen
cote · Follow
cote
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? Seperate 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

Course Information

Social Media Mining

- Line number:
  - CIS 700
- Priority for students that have taken CIS 787
  - See blackboard for more information / Follow the procedure to obtain my signature
- Classroom and Hours:
  - CST 3-216, MW 3:45 – 5:05 PM
- Blackboard:
  - Everything (slides/homeworks/projects/etc.)

Introduction

- **Instructor:**
  - Reza Zafarani (rzafaran@syr.edu)

- **Office Hours (Starting Jan 28th):**
  - Thursday 9:00 - 10:00am, CST 4-279
  - Other times: by appointment only

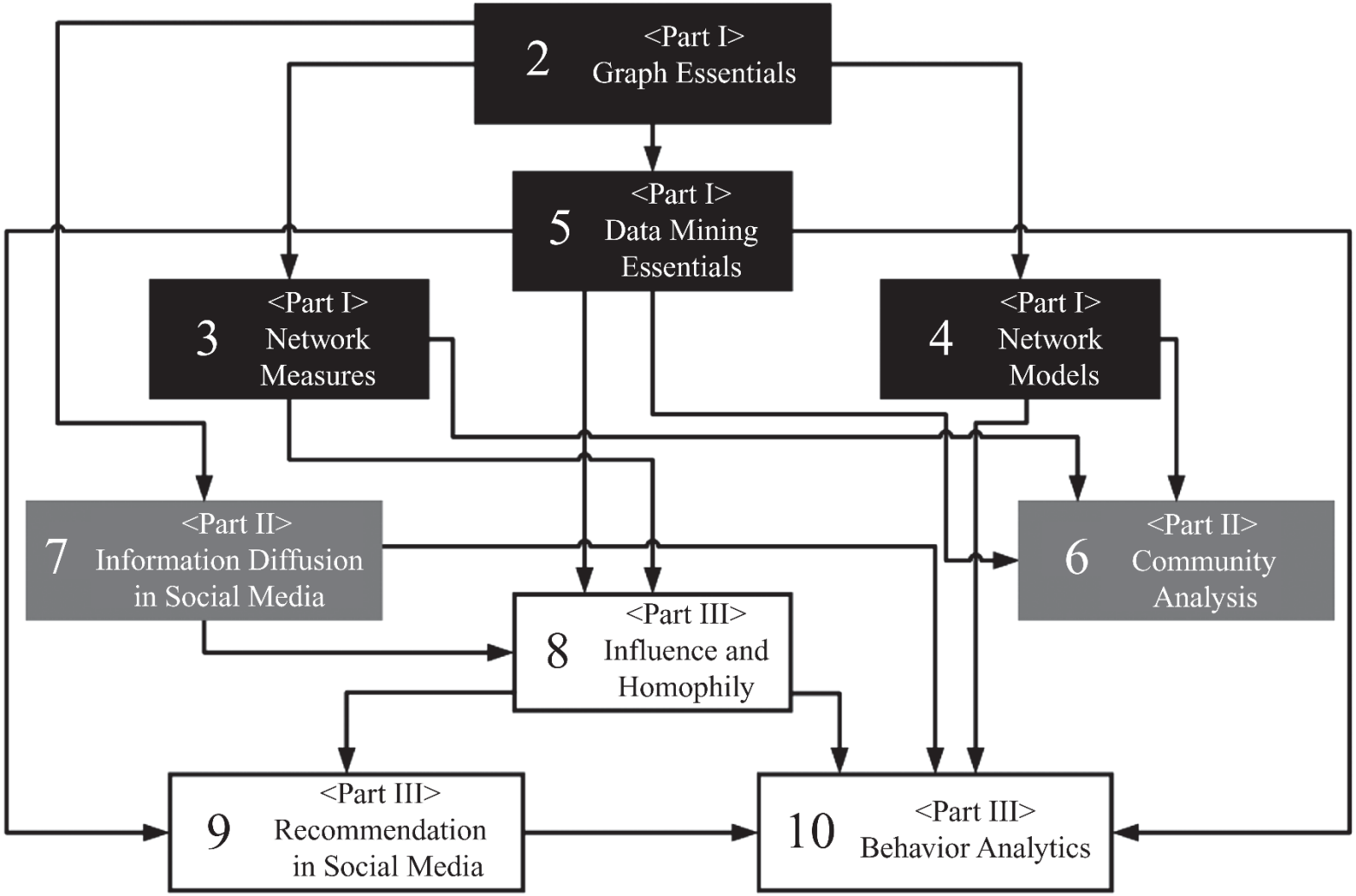
Objectives of Our Course

- Understand social aspects of the Web
  - Social Theories + Social media + Mining
  - Learn to collect, clean, and represent social media data
  - How to measure important properties of social media and simulate social media models
  - Find and analyze communities in social media
  - Understand how information propagates in social media
  - Understanding friendships in social media, perform recommendations, and analyze behavior
- Study or ask interesting research issues
  - e.g., start-up ideas / research challenges
- Learn representative algorithms and tools

Course Information

- Prerequisites:
  - CIS 787 – Analytical Data Mining
  - Data Structures and Algorithms
    - Search/Sort algorithms
    - Graphs
    - Graph Algorithms (Traversal, MST, shortest-paths)
    - Time/Space Complexity
  - Programming Skills: Java, basic understanding of MATLAB is a plus
    - E.g., Being able to crawl a website with Java
    - E.g., Computing eigenvalues of a matrix with MATLAB
  - Basic knowledge of probability, statistics, calculus, and linear algebra
    - Expectation, variance, standard deviation,
    - Eigenvalue computation, determinants, characteristic equation
    - Basic differentiation, integration, and differential equations

Overview – Dependency Graph



The weekly schedule will be available on blackboard

Course Workload and Evaluation

- **A lot** of work is expected from you! Think twice if it does not fit your schedule or match your expectation.
  - Lectures
    - Experienced researchers or practitioners may be invited as guest instructors for specific topics.
  - Homework assignments (15%) – 4 HWs
    - Conceptual; deep thinking required
  - Projects (20%)
    - Two projects
      - first project is individual, second is group project
  - 3 Exams (50%) – 3/6: highest, 2/6: medium, 1/6: lowest
  - Quizzes (15%) – 9-10 quizzes; after chapters/topics
  - Late penalty:
    - Exponential penalty: -50% first day, -75% second day, no points on or after the third day.

Academic Integrity

- Please see:
  - http://supolicies.syr.edu/ethics/acad_integrity.htm

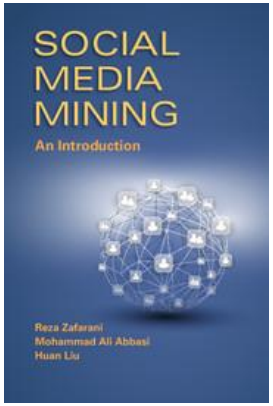
- You are encouraged to form groups to solve problems and coding assignments; however, when writing, write in your own words and provide your own solutions.

Text and Reference Books

Primary Reference:

Social Media Mining, Reza Zafarani, Mohammad-Ali Abbasi, Huan Liu, Cambridge University Press 2014,

- Available at http://socialmediamining.info or amazon.com



Communication Channels and Schedule

Me → You

- Announcements are made regularly on Blackboard
  - Check blackboard regularly
- Emails will be sent out on a need basis

You → me

- Office hours / Email (I don't check my voicemail)

Many ↔ Many

- Q & A: You can ask questions from the instructor and/or other students on blackboard.
  - After-class Blog (i.e., Discussion Board)

Feedback

- A class survey today
- One in the middle of the semester
- At the end of each of classes:

Topic	Yes	No
Topic 1	X	
Topic 2		X
Topic 3	X	
Topic 4		X
“the equation on slide 5”		X

- All surveys are anonymous

Social Media

Definition

Social Media is the use of electronic and Internet tools for the purpose of sharing and discussing information and experiences with other human beings in more efficient ways.

Social Media Landscape 2015

FredCavazza.net

Social Media: Examples

- A wiki article
- Web reviews and ratings of a popular pizza place in your city
  - E.g., Yelp.com
- An online social network of your professional contacts
  - E.g., Facebook.com, LinkedIn.com
- An iPhone application that informs you where parking is likely available
  - FasPark

Types of Social Media

- Online Social Networking
- Publishing
  - Blogging
  - Wiki
- Micro blogging
- Social News
- Social Bookmarking
- Media Sharing
  - Video Sharing
  - Photo Sharing
  - Podcast Sharing
- Opinion, Review, and Ratings Websites
- Answers
- Entertainment

Online Social Networking

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

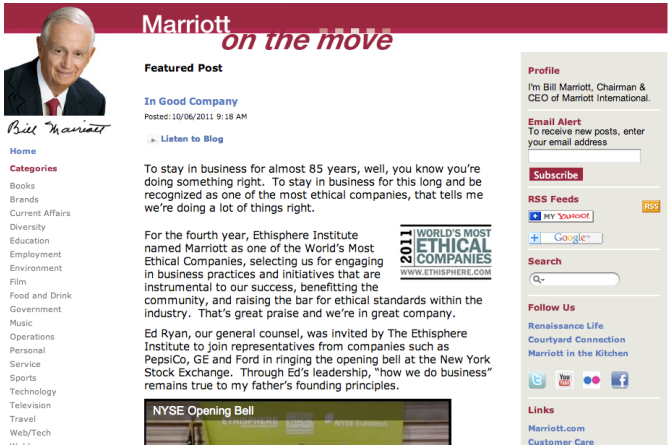
- Interactions
  - Friendship interaction
    - Friends, like, comments, …
  - Media Sharing
  - Sending and receiving messages

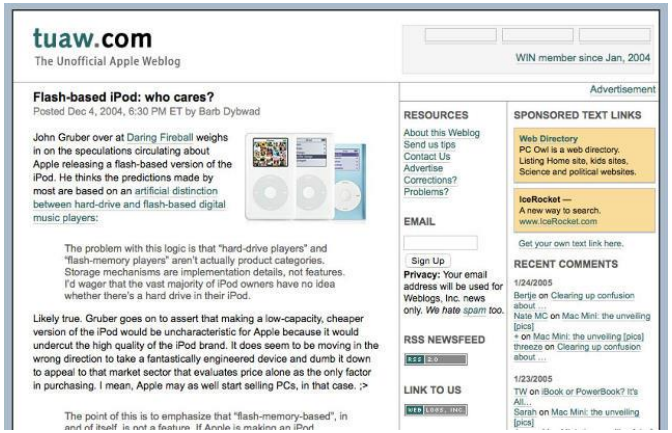
- Examples
  - Facebook.com
  - MySpace.com
  - Bebo.com
  - Orkut.com

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
  - See a tutorial at KDD http://videolectures.net/kdd08_liu_briat/
- Usages:
  - Sharing information and opinions with friends and strangers
  - Disseminating subject-specific content
  - Who is the influential http://videolectures.net/wsdm08_agarwal_iib/





Microblogging

Microblogging can be considered as a counterpart to blogging, but with limited content

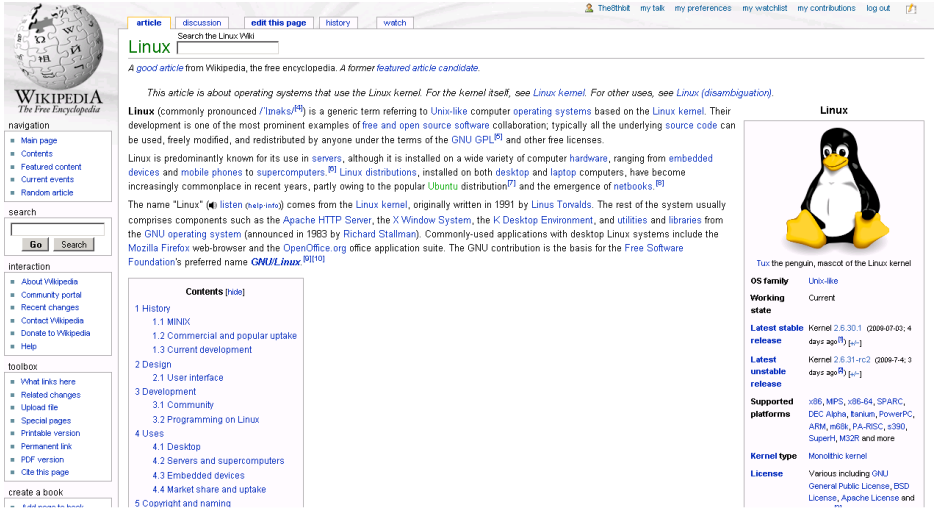
- Usage
  - communication medium
  - social interaction
  - citizen journalism
- Service Providers:
  - Twitter
  - Google buzz

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.

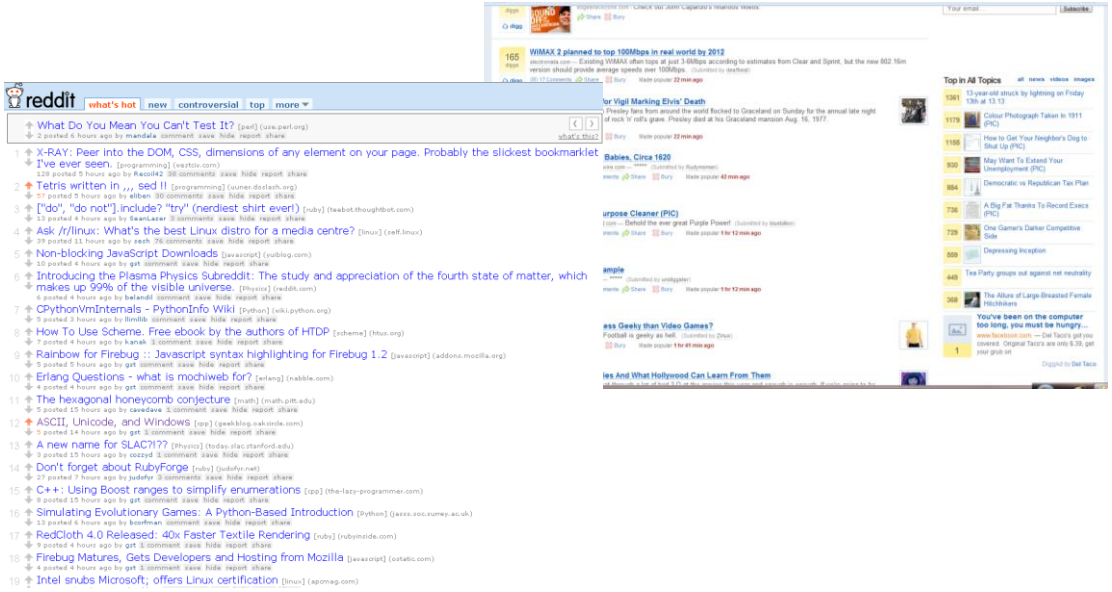
- Using the wisdom of crowds effectively, it has become a comprehensive repository of information useful to a variety of individuals



Social News

Social News refers to the sharing and selection of news stories and articles by a community of users.

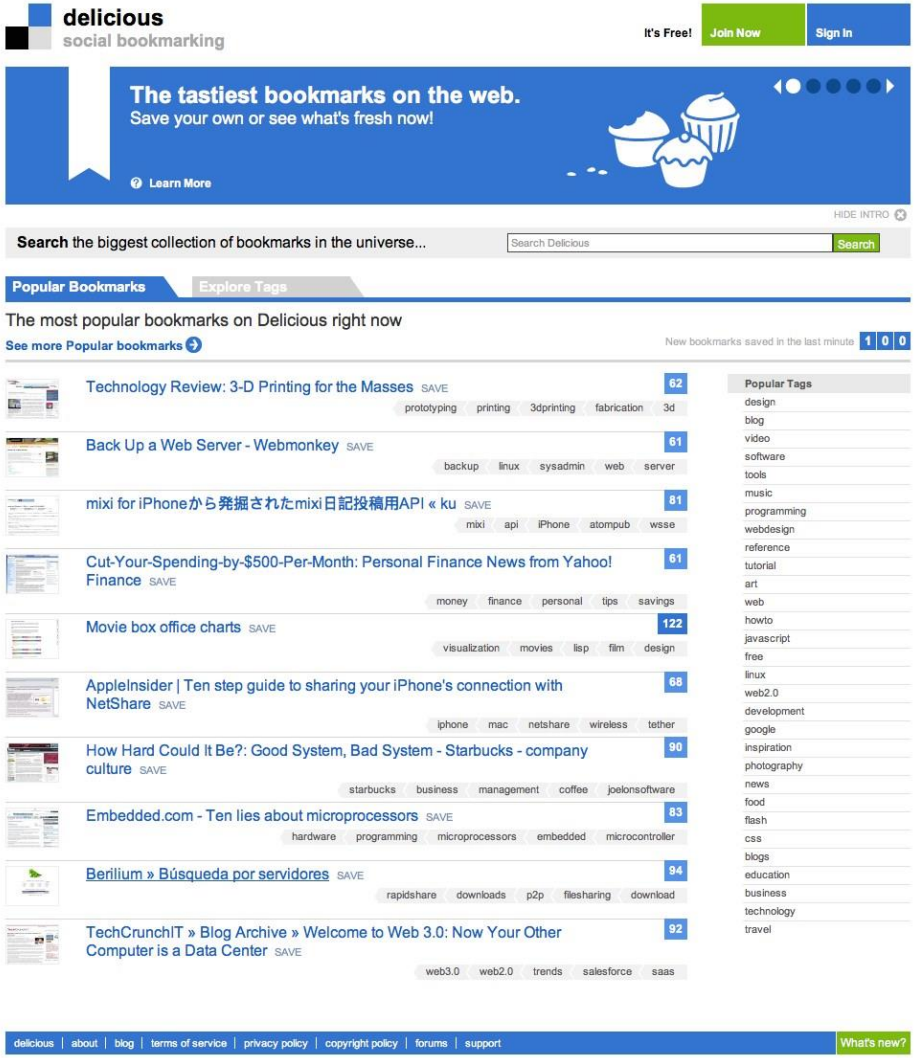
- Users can share articles that they believe would interest the community
- Samples:
  - Digg.com
  - Slashdot
  - Fark
  - Reddit



Social Bookmarking

Social Bookmarking sites allow users to bookmark web content for storage, organization and sharing.

- These bookmarks can be tagged with metadata to categorize and provide context to the shared content, allowing users to organize information making it easy to search and identify relevant information.
- Samples
  - Delicious.com
  - StumbleUpon.com

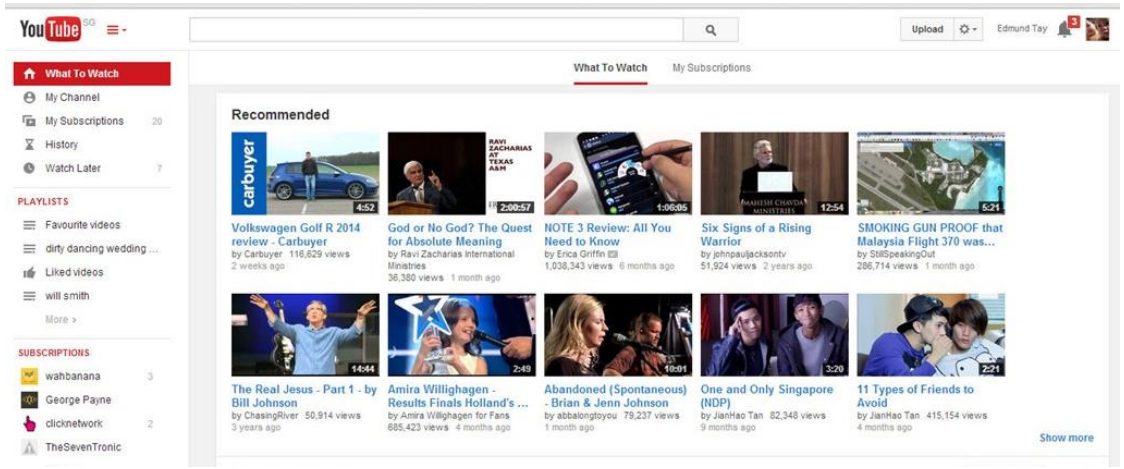


Media Sharing

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

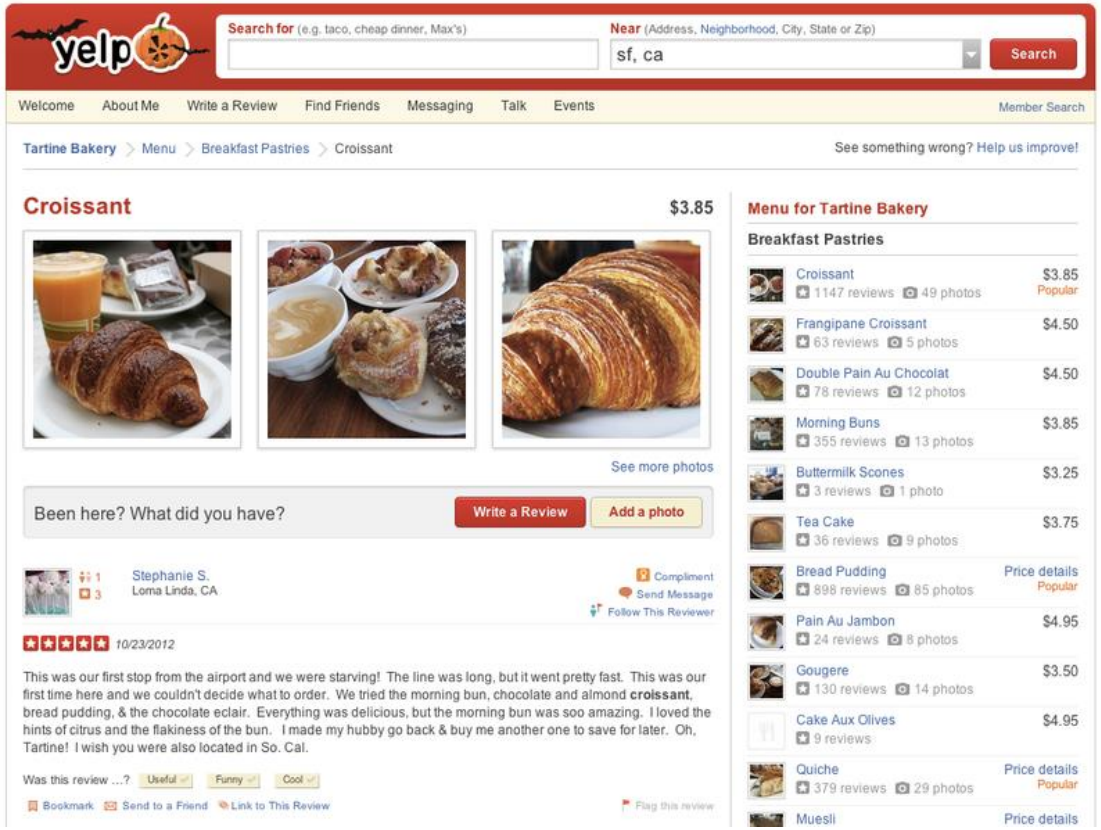
- Samples:
  - Video Sharing:
    - YouTube.com
  - Photo Sharing:
    - Flickr.com, picasa.com
  - Document Sharing:
    - Scribd.com, Slideshare.com
  - Livecasting:
    - Justin.tv, Ustream.com



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

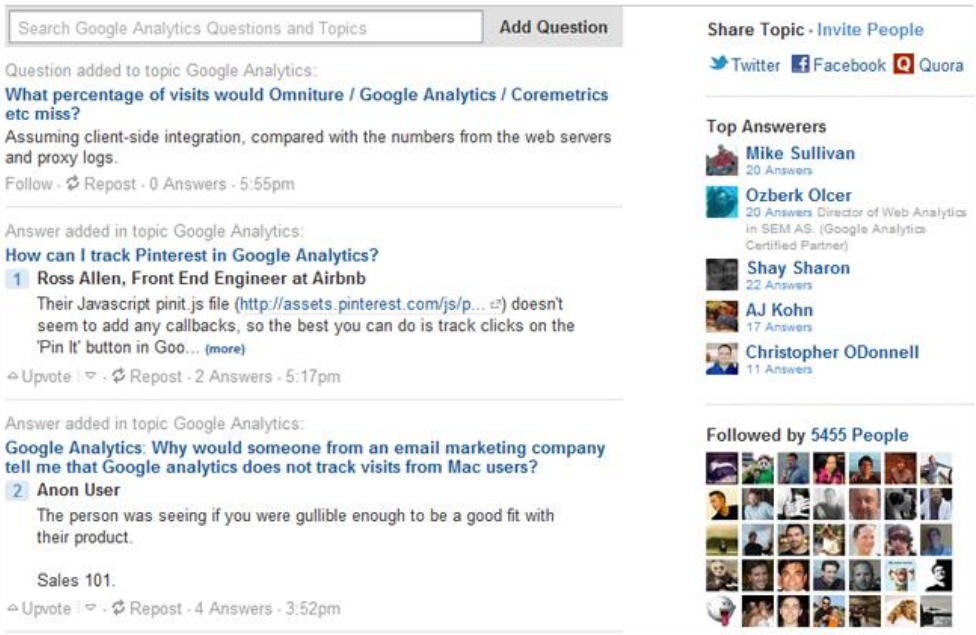


Socially-Provided Answers

In these sites, users who require certain guidance, advice or knowledge can ask questions. Other users from the community can answer these questions based on knowledge acquired from previous experiences, personal opinions or from relevant research.

- Unlike review and opinion sites, which contain self-motivated contribution of opinions, answer sites contain knowledge shared in response to a specific query.

- Samples:
  - WikiAnswers, Yahoo Answers, Quora



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 from social media data*

Social Media Mining Challenges

1. **Big Data Paradox**
   1. Social media data is big, yet not evenly distributed.
   2. Often little data is available for an individual

2. **Obtaining Sufficient Samples**
   1. Are our samples reliable representatives of the full data?

3. **Noise Removal Fallacy**
   1. Too much removal makes data more sparse
   2. Noise definition is relative and complicated and is task-dependent
4. **Evaluation Dilemma**
   1. When there is no ground truth, how can you evaluate?

TODO Items

- 3 To-do items for you:
  - Follow the process on blackboard to obtain my permission
  - Familiarize yourself with blackboard
  - Take the course survey