



# SOCIAL MEDIA MINING

## Introduction

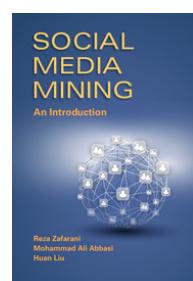


## Text and Reference Books

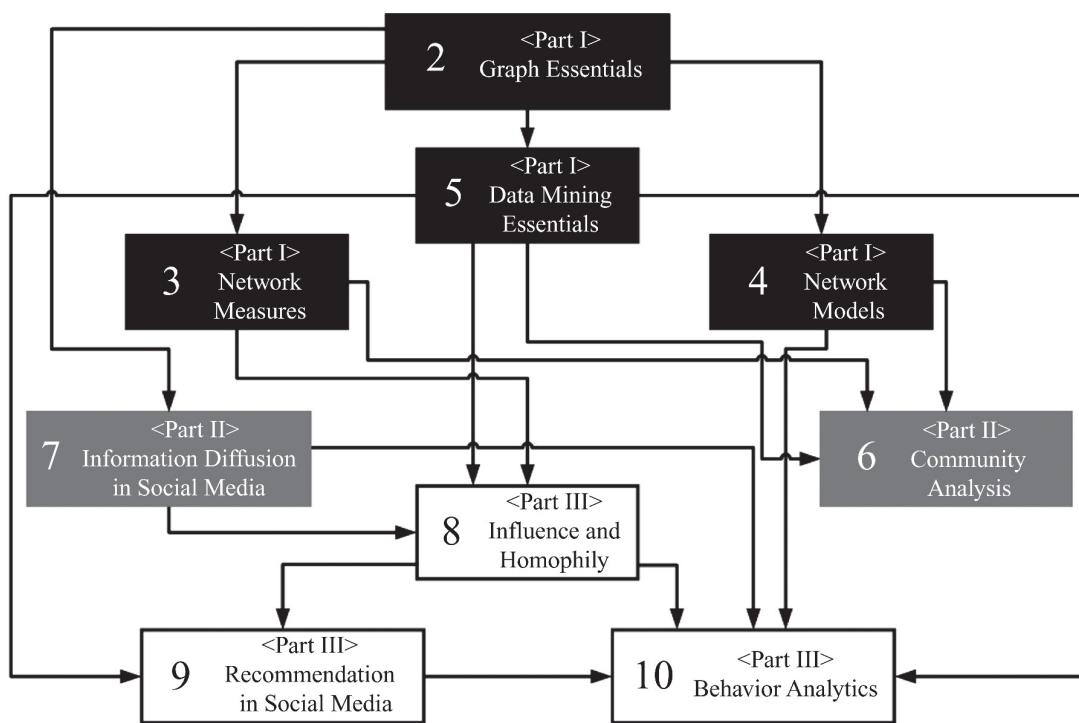
### Primary Reference:

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

– Available at <http://dmml.asu.edu/smm>



# Overview – Dependency Graph



# 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.

# Introduction

- 1<sup>st</sup> social media site
  - Geocities in 1994 which allows users to create their own homepages
- 1<sup>st</sup> social networking site
  - SixDegree.com in 1997
- Many other social media sites provide services to millions of people
  - Forms a virtual world where individuals (social atoms), entities(content, sites, networks), and interactions coexists
  - By understanding these social norms and models of human behaviors and combining them with the observations and measurements of this virtual world, one can systematically analyze and mine social media

## Social Media Landscape 2015



# 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

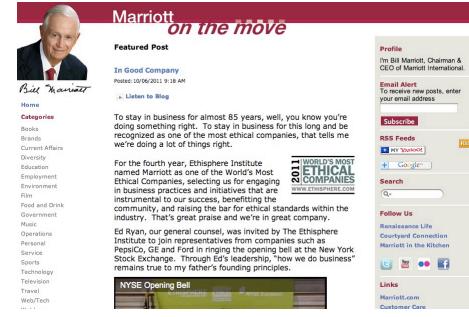
- Examples
  - Facebook.com
  - MySpace.com
  - Bebo.com
  - Orkut.com

A screenshot of a MySpace profile for a user named "Pei Pei". The profile includes a profile picture of a young woman, a bio about being from Seattle, United States, and having 208 profile views. It features a music player showing tracks by Green Day and Bruce Springsteen. The profile has 14 plays for "I Want You To Be Rich" and 14 plays for "I Like You". There are also links to her blog and upcoming shows. The sidebar shows news items and friend suggestions.

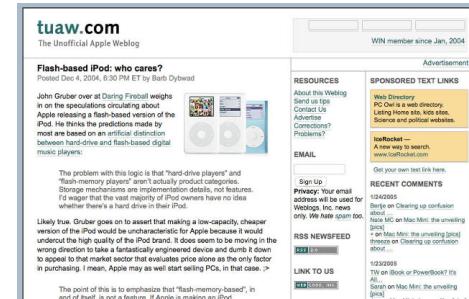
# Blogging

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/](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/](http://videolectures.net/wsdm08_agarwal_iib/)



The screenshot shows the homepage of the Marriott 'on the move' blog. It features a portrait of Bill Marriott, Chairman & CEO of Marriott International. The header includes the text 'Marriott on the move'. A 'Featured Post' section highlights an article titled 'In Good Company' posted on October 10, 2011, at 9:18 AM. The sidebar contains a 'Categories' list with links to Books, Brands, Current Affairs, Diversity, Education, Environment, Film, Food and Drink, Government, Music, Opinions, Personal, Politics, Science, Sports, Technology, Television, Travel, and Web/Tech. On the right, there's a sidebar with a 'Profile' section for Bill Marriott, a 'RSS Feeds' section, a 'Search' bar, and a 'Follow Us' section with links to various social media platforms.



The screenshot shows the homepage of tuaw.com, The Unofficial Apple Weblog. It features a post titled 'Flash-based iPod: who cares?' by Barb Dybdahl, posted on December 4, 2004, at 6:30 PM ET. The main content discusses the shift from hard-drive to flash-based iPods. The sidebar includes sections for 'RESOURCES' (About this Weblog, RSS feed, Contact Us, Advertise, Comments, Problems), 'EMAIL' (Sign Up, Privacy), 'SPONSORED TEXT LINKS' (Web Directory, PC Owl, Renaissance Life, Mac Mini, Science and political websites), and 'iPodRocket' (A new way to search). There are also links for 'Get your own iFixit link here.', 'RECENT COMMENTS' (10/25/05, 10/26/05, 10/27/05, 10/28/05, 10/29/05, 10/30/05, 10/31/05, 11/1/05, 11/2/05, 11/3/05), and 'RSS NEWSFEED' (RSS, XML).

# 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

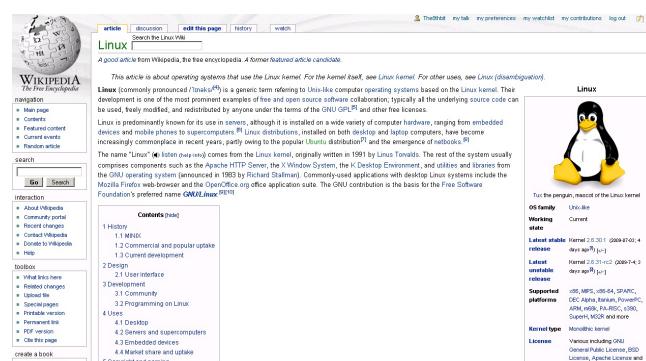


The screenshot shows a Twitter profile for Mario Armstrong (@marioarmstrong). The profile picture is a portrait of a man. The bio reads: 'Tweets at Fox News & MSNBC. People I敬爱. Shoe addict 1.5m followers at Socialcam.com/Mario Television and Online - http://marioarmstrong.com'. The stats show 12,936 tweets, 1,444 followers, and 12,128 following. The timeline displays several tweets from Mario Armstrong, including one from October 10, 2011, at 10:44 AM, and another from October 11, 2011, at 10:44 AM.

# Wiki

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

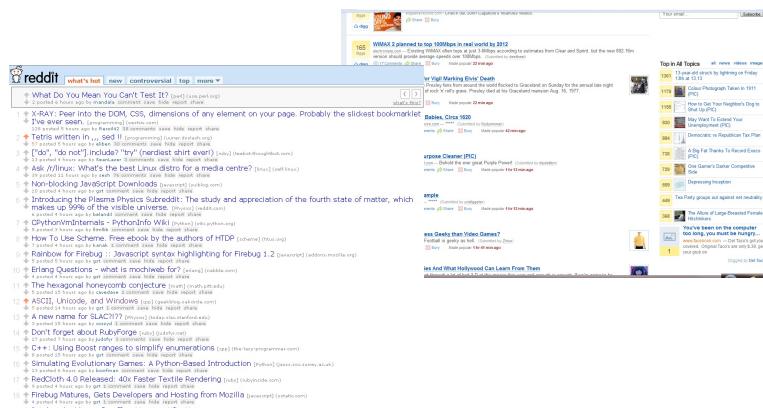
- Wikipedia allows interested individuals to collaboratively develop articles on a variety of subjects.
- Using the wisdom of crowds effectively, it has become a comprehensive repository of information useful to a variety of individuals



## 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

The screenshot shows the Delicious.com homepage. At the top, there's a banner with the text "The tastiest bookmarks on the web. Save your own or see what's fresh now!" and icons for a cupcake and a cookie. Below the banner, there's a search bar and a "Popular Bookmarks" section. This section lists several bookmarks with their titles, save counts, and tags. For example, the first bookmark is "Technology Review: 3-D Printing for the Masses" with 62 saves and tags like "prototyping", "printing", "3dprinting", "fabrication", and "3d". To the right of the bookmarks, there's a sidebar titled "Popular Tags" which lists various tags such as "design", "blog", "video", "software", etc. At the bottom of the page, there's a footer with links to "delicious", "about", "blog", "terms of service", "privacy policy", "copyright policy", "forums", "support", and "What's new?".

# 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

The screenshot shows the YouTube homepage. On the left, there's a sidebar with "What To Watch" sections for "My Channel", "My Subscriptions" (with 20 items), "History", "Watch Later" (with 7 items), "PLAYLISTS" (including "Favourite videos", "dirty dancing wedding...", "Liked videos", and "will smith"), and "SUBSCRIPTIONS" (including "wahabana", "George Payne", "clicknetwork", and "TheSevenTronic"). The main area has tabs for "Recommended" and "My Subscriptions". Under "Recommended", there are several video thumbnails with titles like "Volkswagen Golf R 2014 review - Carbuyer" (4.52M views), "God or No God? The Quest for Absolute Meaning by Rev. Zacharias International Religious Foundation" (36,380 views), and "NOTE 3 Review: All You Need to Know by Erica Griffith" (1,035,141 views). There are also sections for "Abandoned (Spontaneous)" (1,035 views), "Six Signs of a Rising Warrior" (51,524 views), and "SMOKING GUN PROOF that Malaysia Flight 370 was...". At the bottom, there's a "Show more" link.

# 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

The screenshot shows a Yelp page for a Croissant at Tartine Bakery. The product is priced at \$3.85. A review by Stephanie S. from Loma Linda, CA, dated 10/23/2012, describes the experience: "This was our first stop from the airport and we were starving! The line was long, but it went pretty fast. This was our first time here and we couldn't decide what to order. We tried the morning bun, chocolate and almond croissant, bread pudding, & the chocolate eclair. Everything was delicious, but the morning bun was sooo amazing. I loved the hints of citrus and the flakiness of the bun. I made my hubby go back & buy me another one to save for later. Oh, Tartine! I wish you were also located in So. Cal." Below the review are links to 'Was this review...?' and 'Flag this review'.

**Menu for Tartine Bakery**

Item	Reviews	Photos	Price
Croissant	1147 reviews	49 photos	\$3.85 Popular
Frangipane Croissant	63 reviews	5 photos	\$4.50
Double Pain Au Chocolat	78 reviews	12 photos	\$4.50
Morning Buns	355 reviews	13 photos	\$3.85
Buttermilk Scones	3 reviews	1 photo	\$3.25
Tea Cake	36 reviews	9 photos	\$3.75
Bread Pudding	898 reviews	85 photos	Price details Popular
Pain Au Jamon	24 reviews	8 photos	\$4.95
Gougeré	130 reviews	14 photos	\$3.50
Cake Aux Olives	9 reviews		\$4.95
Quiche	379 reviews	29 photos	Price details Popular
Muesli			Price details

## 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

The screenshot shows a Quora question titled "What percentage of visits would Omniture / Google Analytics / Coremetrics etc miss?". The question was added to the topic "Google Analytics". An answer by Ross Allen, Front End Engineer at Airbnb, discusses the limitations of tracking via Javascript pinit.js file. Another user, Anon User, provides a follow-up answer about the Pin It button. The sidebar includes sections for "Top Answerers" (Mike Sullivan, Ozberk Ocer) and "Followed by 5455 People".

# 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?

## Lecture Overview

- Throughout the lectures, we learn to answer questions such as
  - Who are the most important people in a social network?
  - How do people befriend others?
  - How can we find interesting patterns in user-generated content?
  - How community are formed and evolved?
  - How the qualities of detected communities are evaluated?