Social Media Mining

Introduction
• How does Facebook use your data?
• Where do you think Facebook can use your data?
Twitter

Social Media Mining

Introduction
Objectives of Our Course

• Understand social aspects of the Web
  – Social Theories + Social media + Mining
  – Learn how 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
  – Understanding friendships in social media, perform recommendations, and analyze behavior

• Study or ask interesting research issues
  – e.g., start-up ideas

• Learn representative algorithms and tools
Overview – Dependency Graph

1. Graph Essentials
2. Data Mining Essentials
3. Network Measures
4. Network Models
5. Influence and Homophily
6. Community Analysis
7. Information Diffusion in Social Media
8. Recommendation in Social Media
9. Behavior Analytics

Social Media Mining

Introduction
Social Media
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

Social Media

Publish

Share

Discuss

Social Networks

Virtual Worlds

Social Games

Livecast

Lifestream

Microblog

Pownce

PLURK

addub

tweetpawk

Socialize

justin.tv

socializr

friendfeed

socialthing!

Hi5

Ning

meebo

comms

PHORUM

forumposting

phpbb

Crowdstorm

lost-fm

slideshare

YouTube

flickr

social networks

Lifestyle

MMO

mystruggle

The Sims

Wii

World of Warcraft
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?