Social Network Mining
An Introduction

Jiawei Zhang
Assistant Professor
Florida State University

A Questionnaire

Please raise your hands, if you

(1) use Facebook

(2) use Instagram

(3) use Snapchat

(4) use LinkedIn

(5) don’t use any online social networks

Background Knowledge

Online Social Networks & Web 2.0

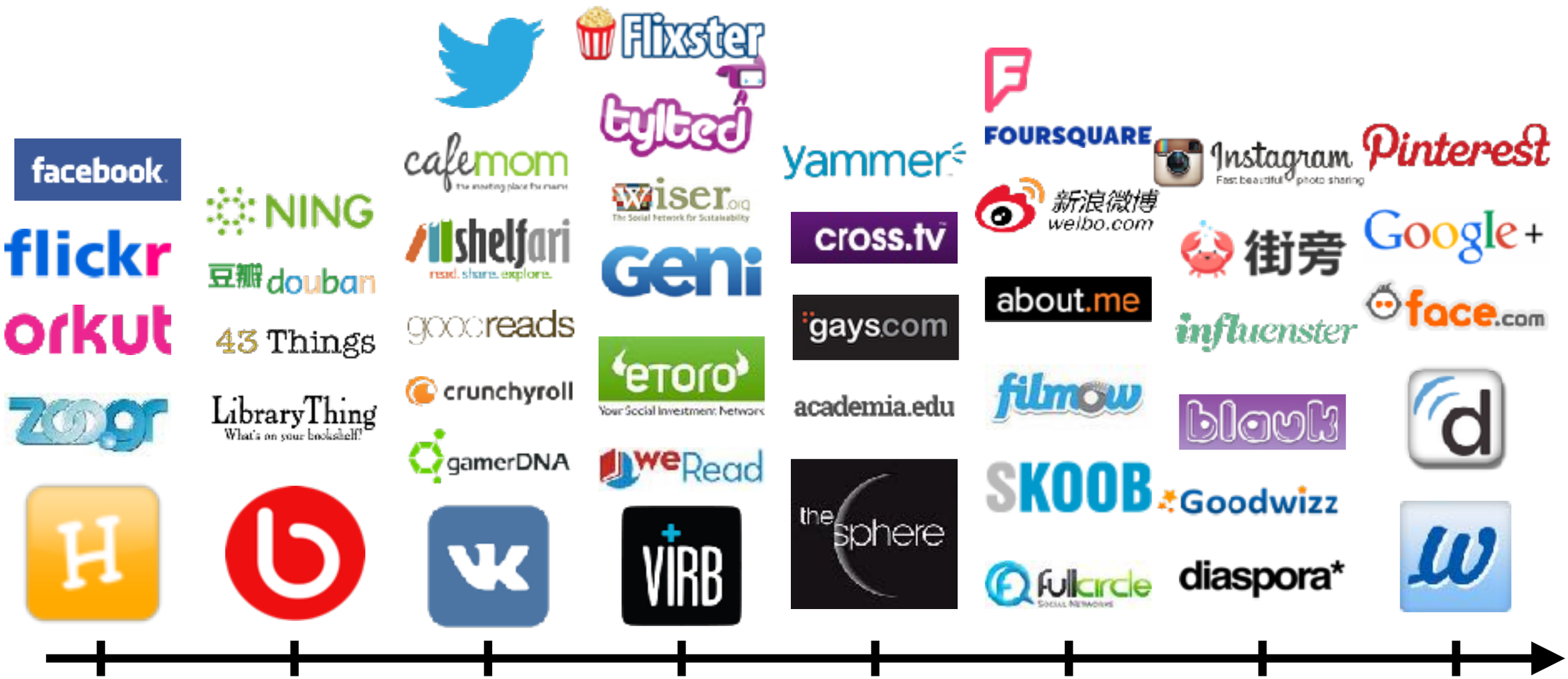
- Web 1.0 vs Web 2.0
  - Web 1.0: users take information
  - Web 2.0: users create information

- Online Social Networks & Web 2.0
  - OSNs use Web 2.0 technology to share a user-focused approach for communication
  - User actively participate in content creation and editing through open collaboration between members of communities of practice

Online Social Networks

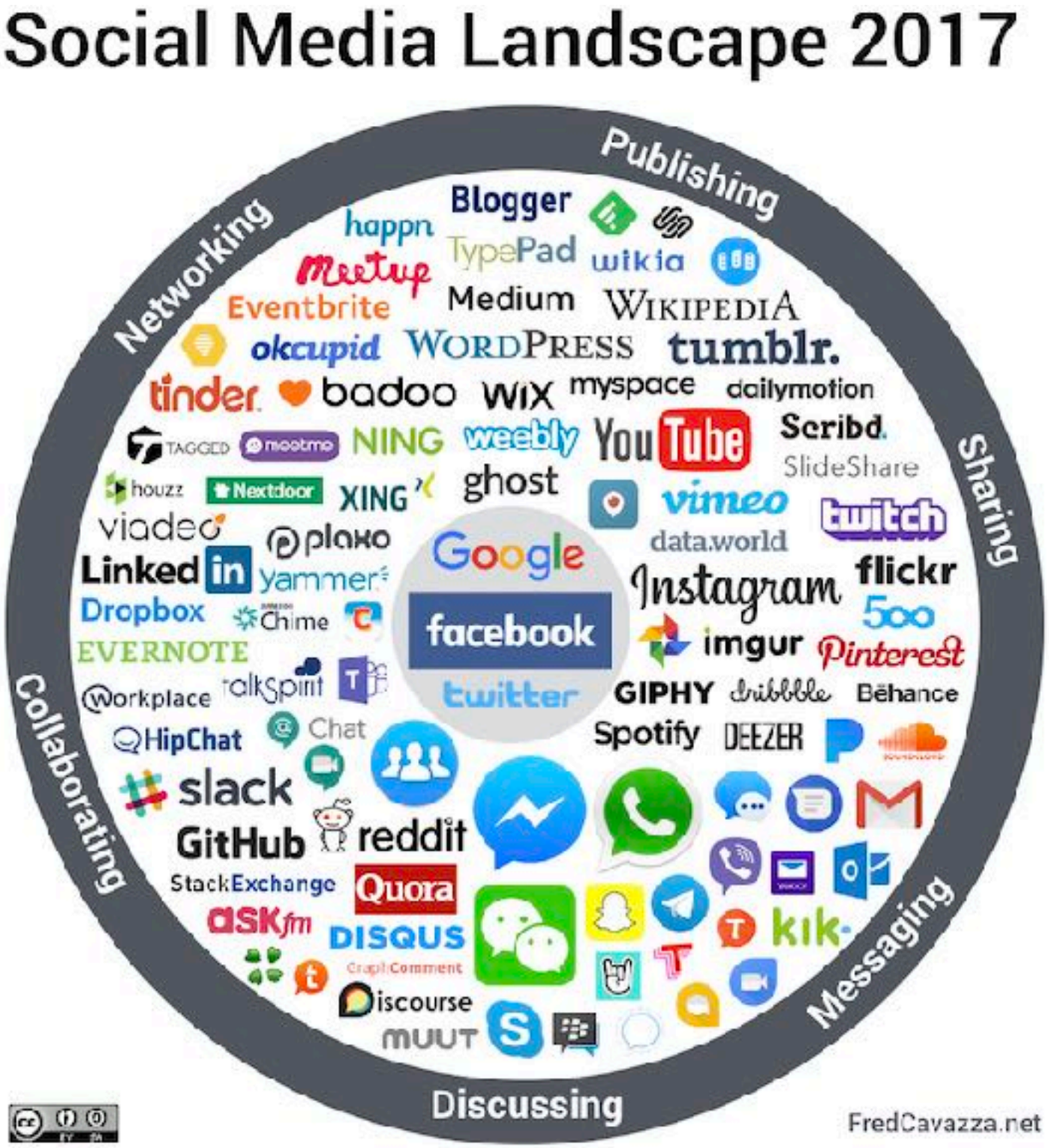
Online social media is the use of electronic and Internet technology for the purpose of sharing and discussing information and experiences with other human beings in more effective and efficient ways.

New Social Networks Emerge Every Year



http://en.wikipedia.org/wiki/List_of_social_networking_websites

Social Media Landscape In Recent Years



Social Network Statistics and Functions

PINTEREST

SOCIAL SITE THAT IS ALL ABOUT **DISCOVERY**

LARGEST OPPORTUNITIES

USERS ARE:

32% MALE

68% FEMALE

70 MILLION ACTIVE USERS

TWITTER

MICRO BLOGGING SOCIAL SITE THAT LIMITS EACH POST TO **140** CHARACTERS

LARGEST PENETRATION

BUT SPREADING SLOWLY AND STEADILY

5,700 TWEETS HAPPEN EVERY SECOND

560 MILLION ACTIVE USERS

FACEBOOK

SOCIAL SHARING SITE THAT HAS **1 BILLION** USERS WORLDWIDE

LARGEST OPPORTUNITIES

COMMUNICATING WITH **CONSUMERS** IN A NON-OBTRUSIVE WAY

USERS SHARE **2.5 BILLION** PIECES OF CONTENT EACH DAY

1 BILLION ACTIVE USERS

INSTAGRAM

SOCIAL SHARING SITE ALL AROUND **PICTURES** AND NOW 15 SECOND **VIDEOS**

MANY BRANDS ARE PARTICIPATING THROUGH THE USE OF **#HASHTAGS**

AND POSTING **PICTURES** CONSUMERS CAN RELATE TO

MOST FOLLOWED BRAND IS MTV

150 MILLION ACTIVE USERS

GOOGLE+

SOCIAL NETWORK BUILT BY GOOGLE THAT ALLOWS FOR **BRANDS** AND **USERS** TO BUILD CIRCLES

NOT AS MANY BRANDS ACTIVE, BUT THE ONES THAT ARE **TEND TO BE A GOOD FIT** WITH A GREAT FOLLOWING

GROWING RAPIDLY WITH **925,000** NEW USERS EVERY DAY

400 MILLION ACTIVE USERS

LINKEDIN

BUSINESS ORIENTED SOCIAL NETWORKING SITE

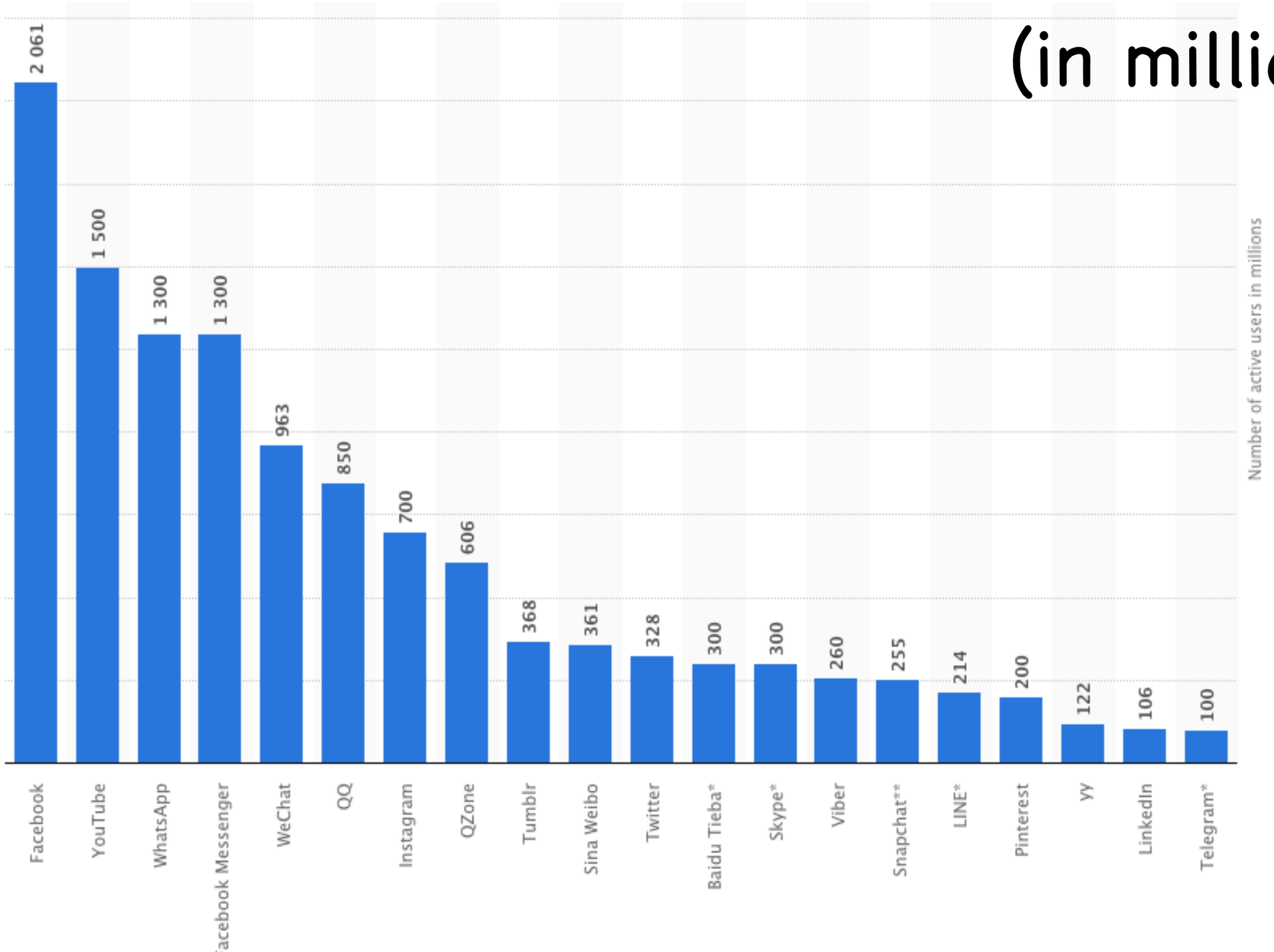
BRANDS THAT ARE PARTICIPATING ARE **CORPORATE BRANDS** GIVING POTENTIAL AND CURRENT ASSOCIATES A PLACE TO NETWORK & CONNECT

79% OF USERS ARE **35** OR OLDER

240 MILLION ACTIVE USERS

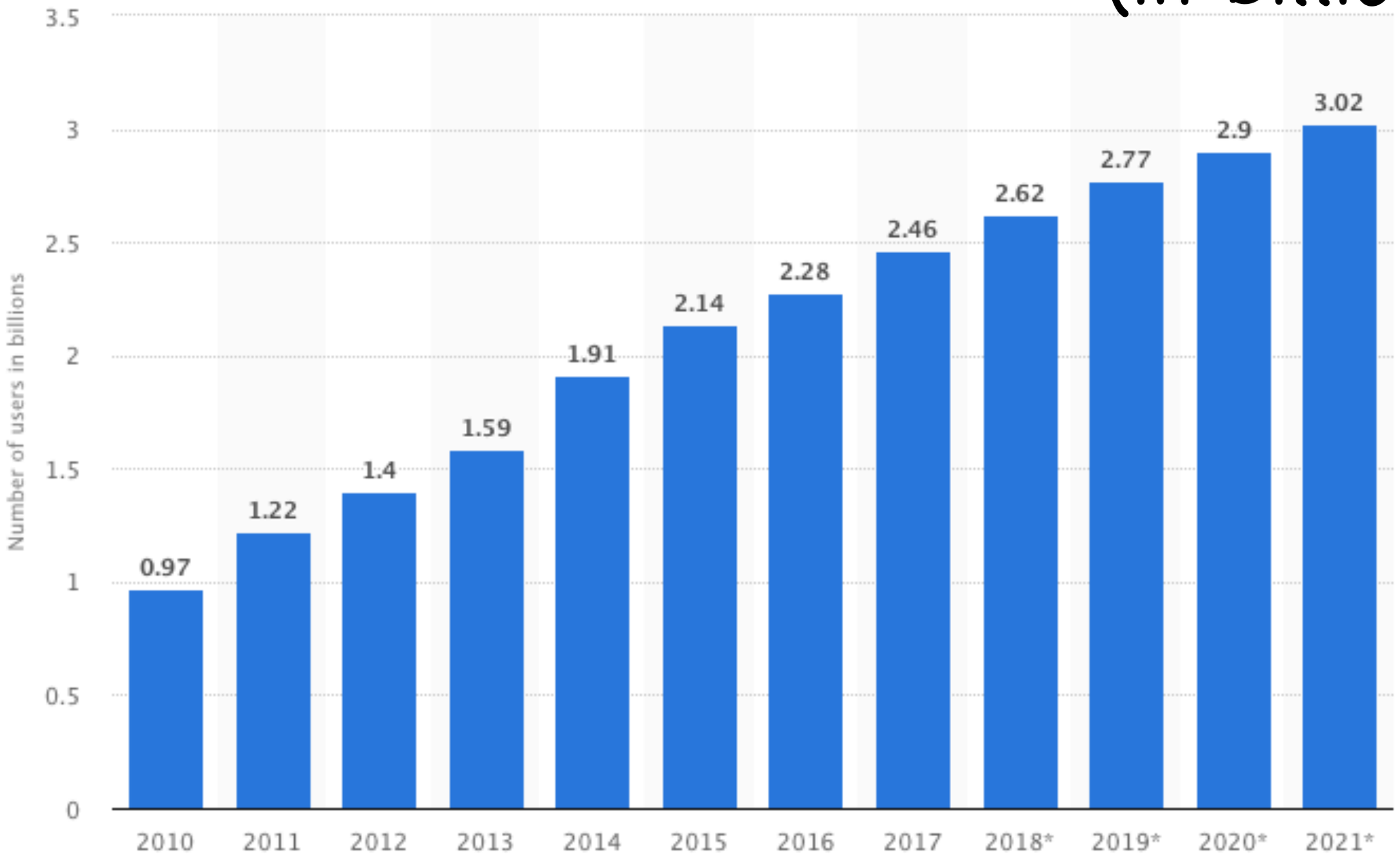
Social Networks are Very BIG (Statistical Data 2017)

(in millions)



Social Networks are Very BIG (2010-2021)

(in billions)



Social Media Types

- 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

http://dmml.asu.edu/smm/slide/SMM-Slides-ch1.pdf

Online Social Networks

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

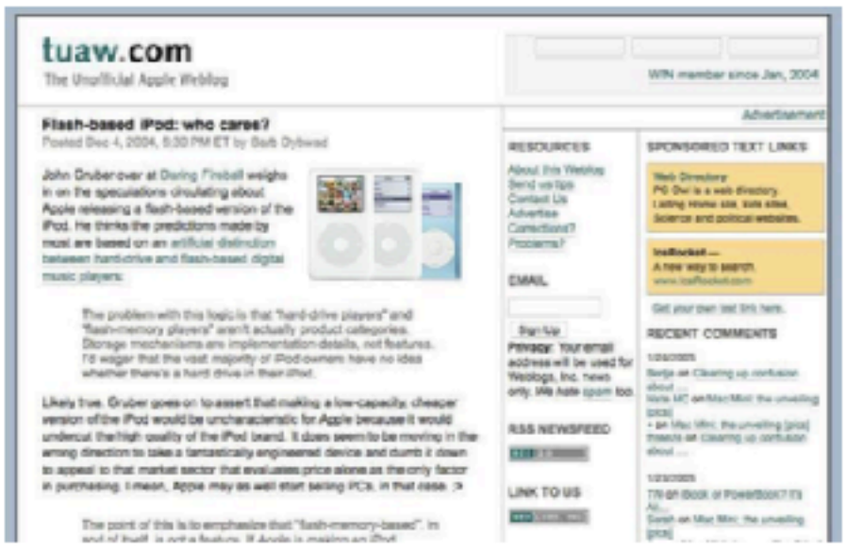
- Examples
  - Facebook.com
  - MySpace.com
  - Bebo.com
  - Orkut.com

http://dmml.asu.edu/smm/slide/SMM-Slides-ch1.pdf

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/

http://dmml.asu.edu/smm/slide/SMM-Slides-ch1.pdf

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

http://dmml.asu.edu/smm/slide/SMM-Slides-ch1.pdf

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

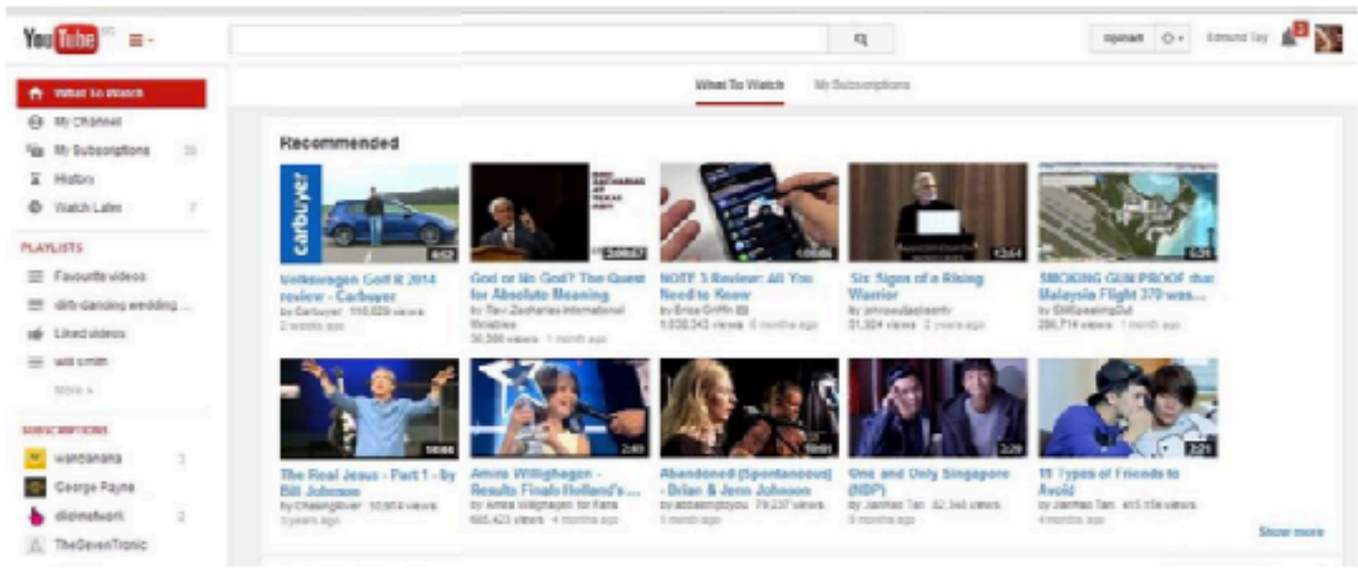
http://dmml.asu.edu/smm/slide/SMM-Slides-ch1.pdf

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

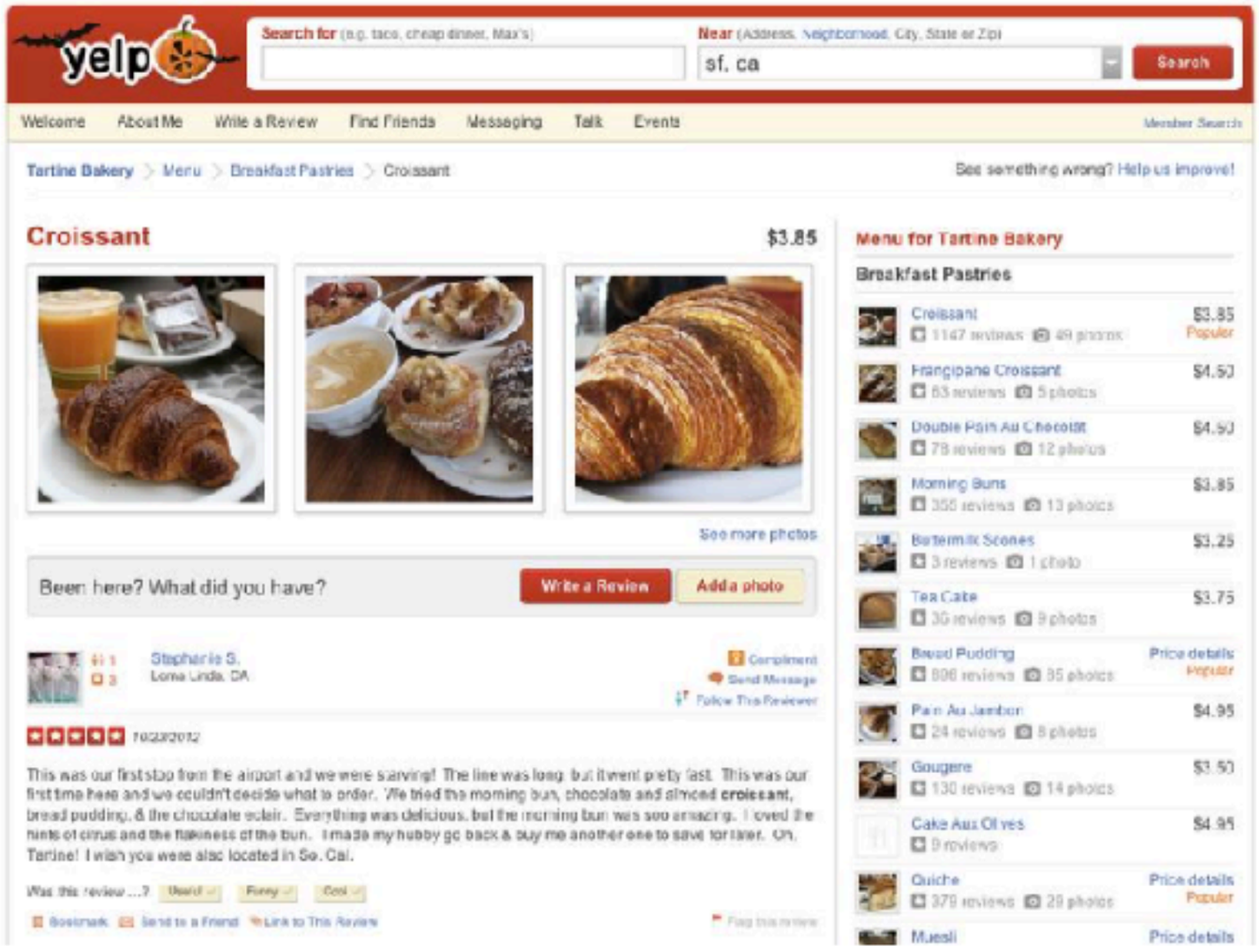


http://dmml.asu.edu/smm/slide/SMM-Slides-ch1.pdf

Opinion, Review, and Rating 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

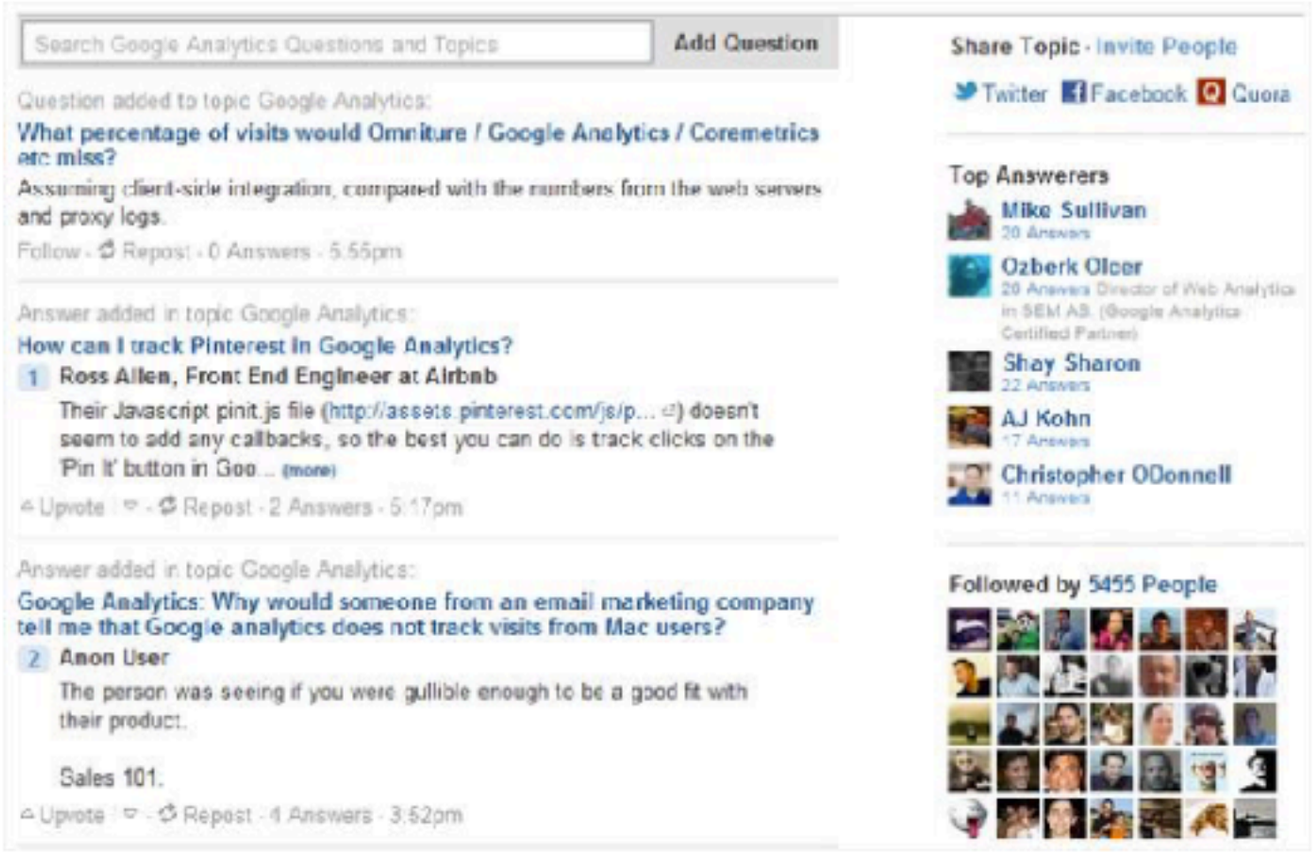


Question & Answer Sites

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.

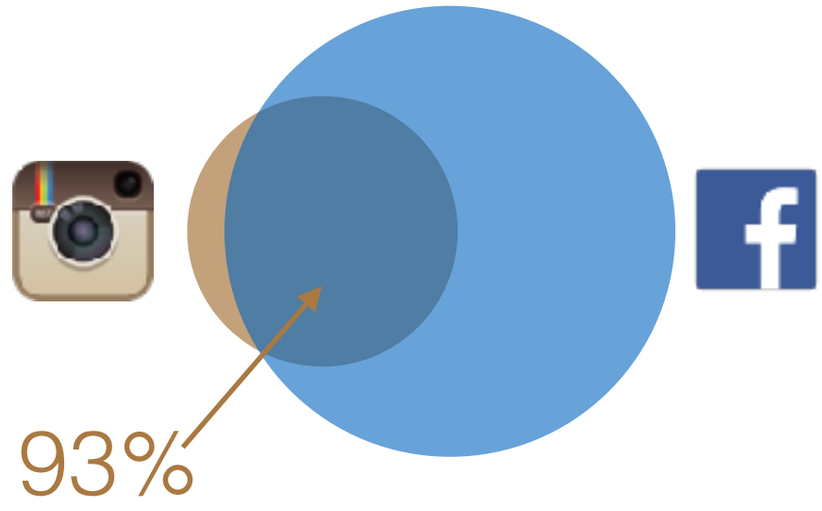
  - Unlike review and opinion sites, which contain self-motivated contribution of opinions, answer sites contain knowledge shared in response to a specific query.

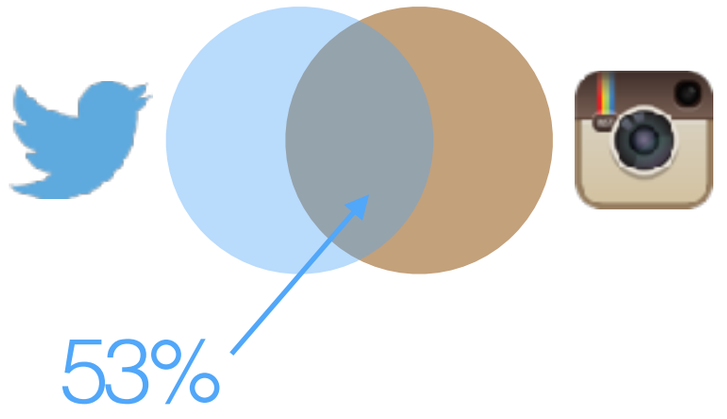
- Samples:
  - WikiAnswers, Yahoo Answers, Quora

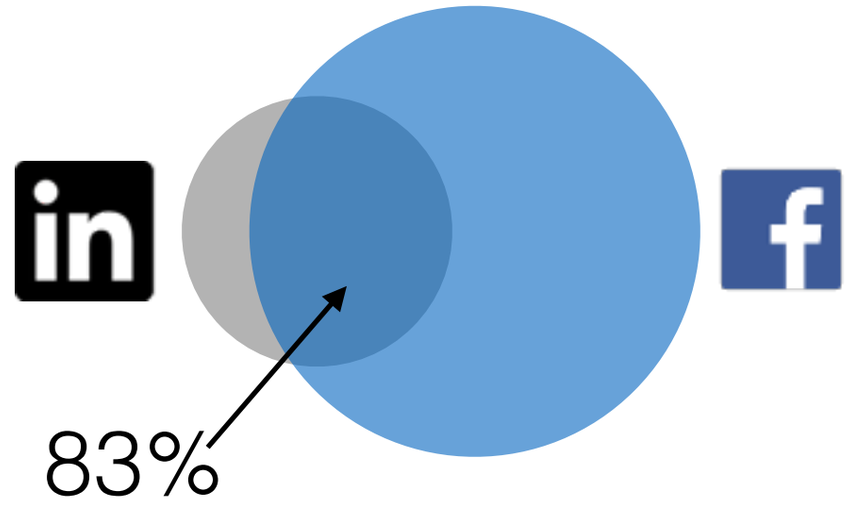


http://dmml.asu.edu/smm/slide/SMM-Slides-ch1.pdf

An Observation: People are using multiple online social networks simultaneously



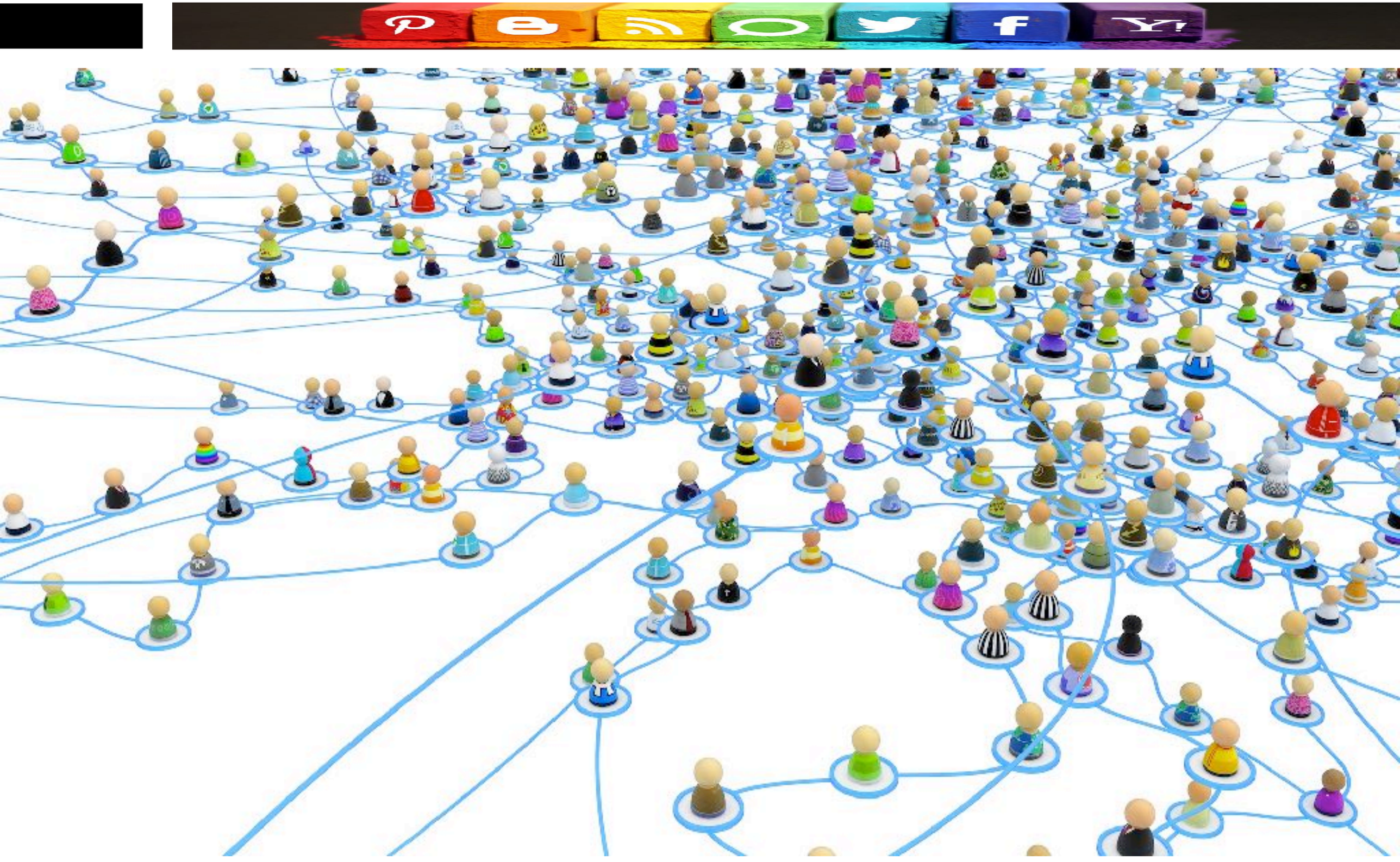




[1] Duggan et al. Social media update, 2013.

Network Structure Representation

Concept Definition: Homogeneous Social Networks



Concept Definition: Heterogeneous Social Networks

- *Representation:* Homogeneous Social Networks

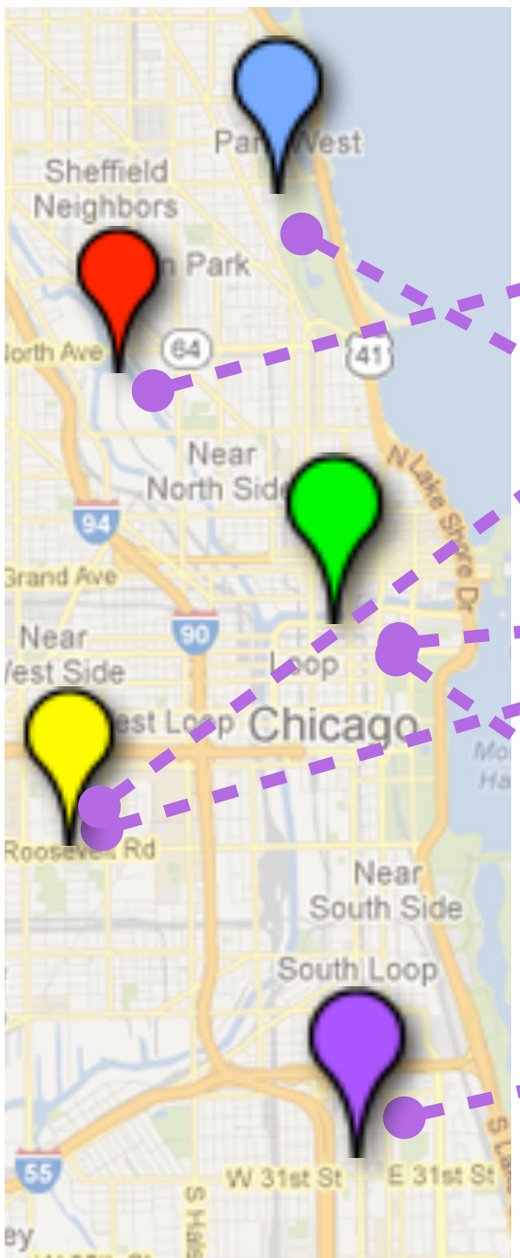
$$G = (V, E)$$

Where V is the sets of various kinds of nodes in the network and set E denotes the various types of links in the network

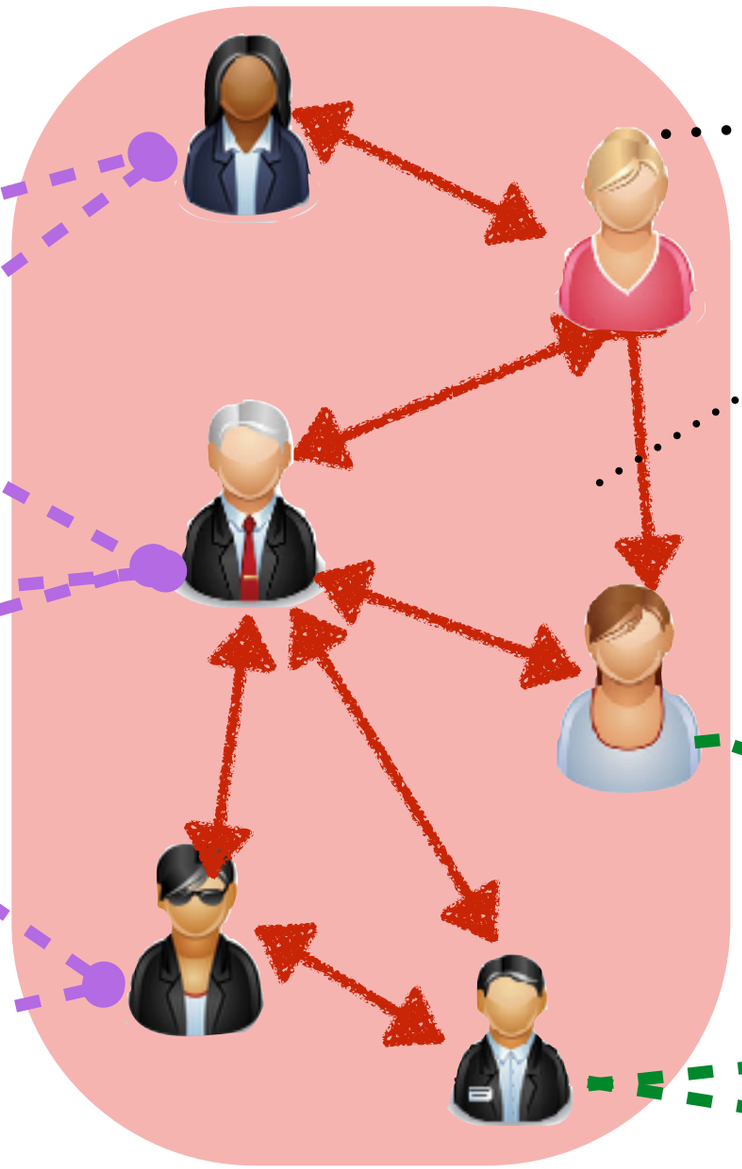
Social Network Structures are Very Complex

Who, **Where**, **When**, **What**

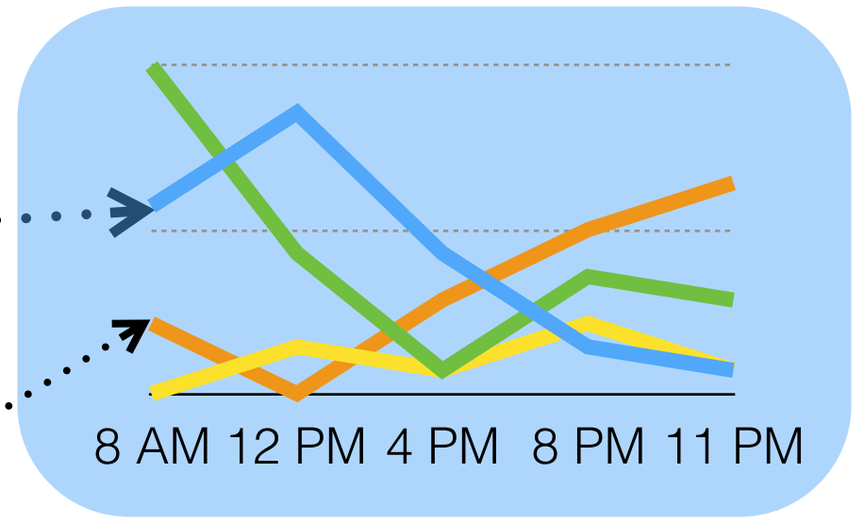
Locations



Social Links



Temporal Activities



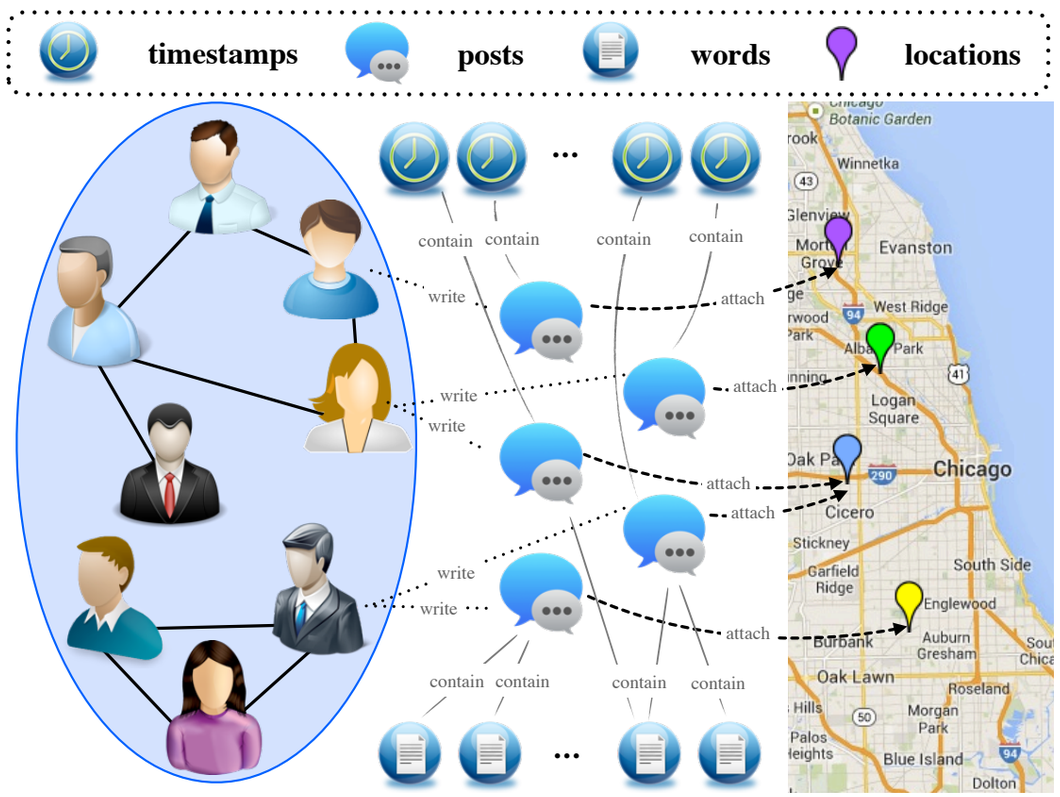
Contents: Tweets

Concept Definition: Heterogeneous Social Networks

- *Representation:* Heterogeneous Social Networks

$$G = (V, E)$$

where $V = \bigcup_i V_i$ is the sets of various kinds of nodes in the network and $E = \bigcup_j E_j$ is the set of various types of links in the network



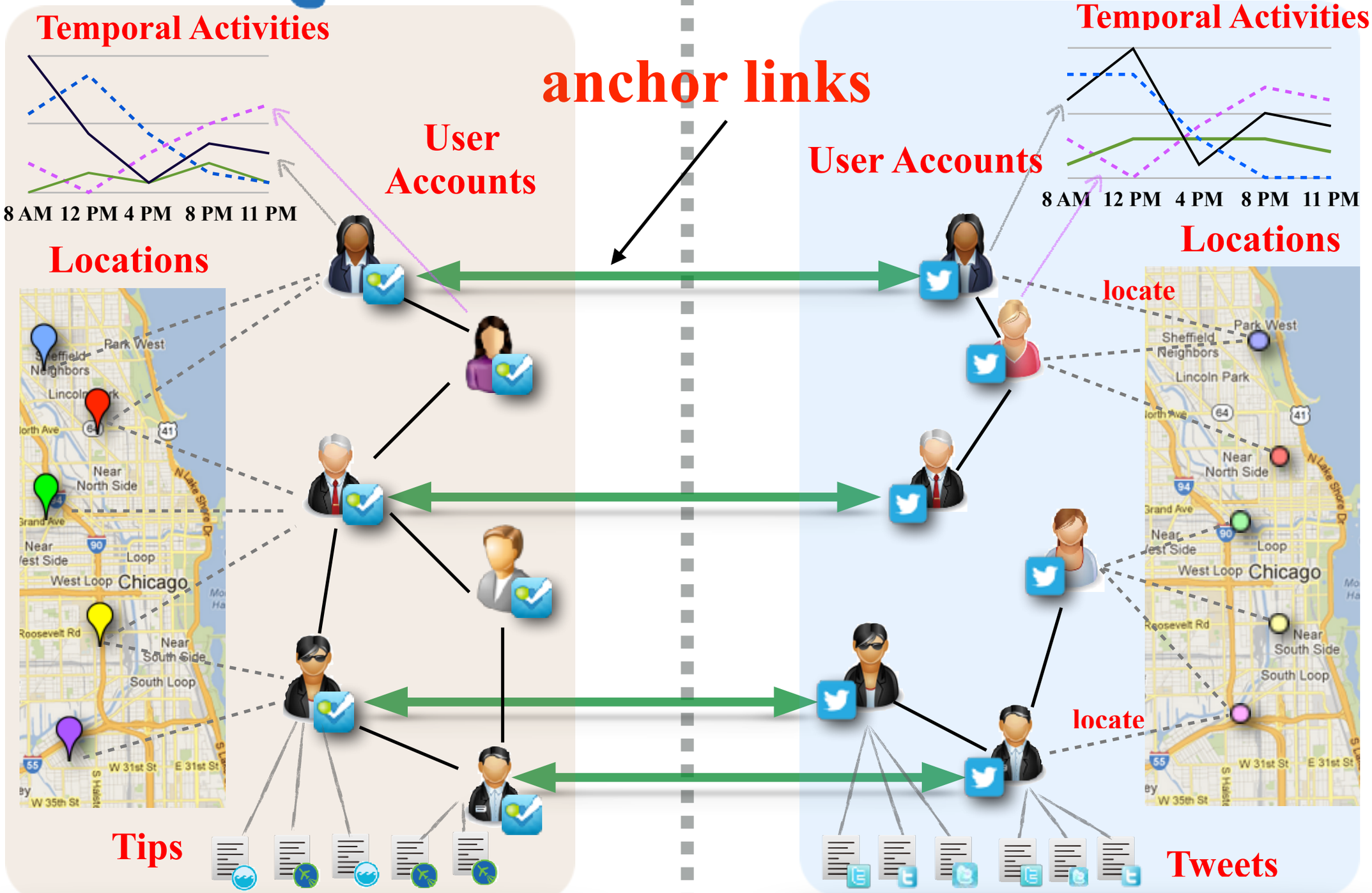
V = {**user node** set, **post node** set, **word node** set, **time node** set, **location node** set}

E = {**user-user link** set, **user-post link** set, **post-word** link set, **post-time** link set, **post-location** link set}

Social Networks Belong to Different Varieties

foursquare

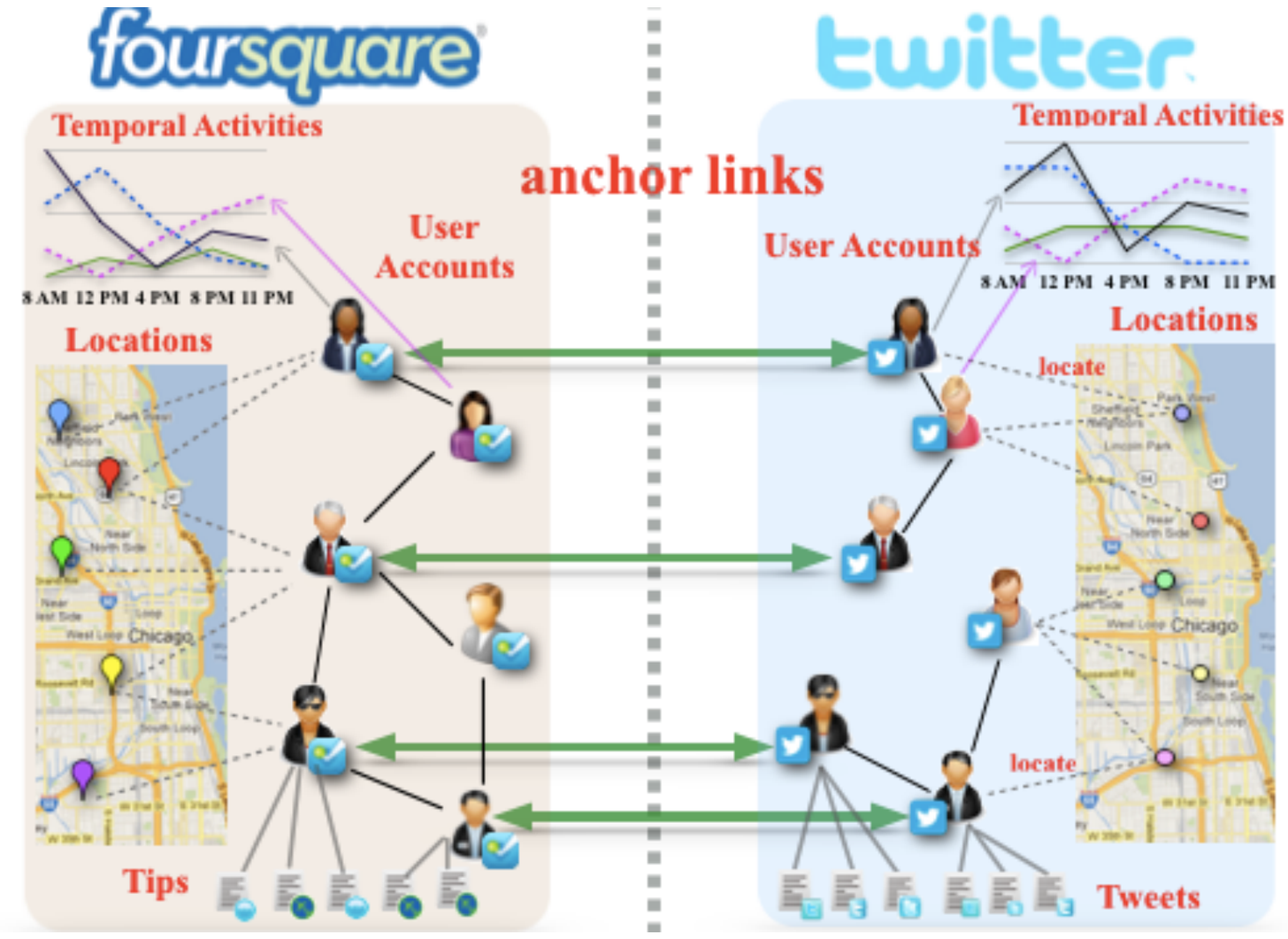
twitter



Concept Definition: Multiple Aligned Heterogeneous Social Networks

- Representation: Multiple Aligned Heterogeneous Social Networks

$$\mathcal{G} = \left((G^{(1)}, G^{(2)}), (A^{(1,2)})\right)$$



Heterogeneous Social Networks

$G^{(1)}$: Foursquare

$G^{(2)}$: Twitter

Anchor Links

$A^{(1,2)}$: Anchor links between Foursquare and Twitter

Social Network Mining Problems (An Overview)

Social Network Mining Problems: An Overview

User-Centric

Role Analysis

Social Spammer Detection

Social Ties

Negative Links

Information Diffusion

Network Alignment

Network Summarization

Network Embedding

Content-Centric

Misinformation

Event Detection

Content Quality and Popularity

Sentiment Analysis

Social Tags

Social Summarization

Social Recommendations

Social Media Q&A

Interdisciplinary

Personality Analysis

Crisis Informatics

Social Media Healthcare

Social Media Privacy and Security

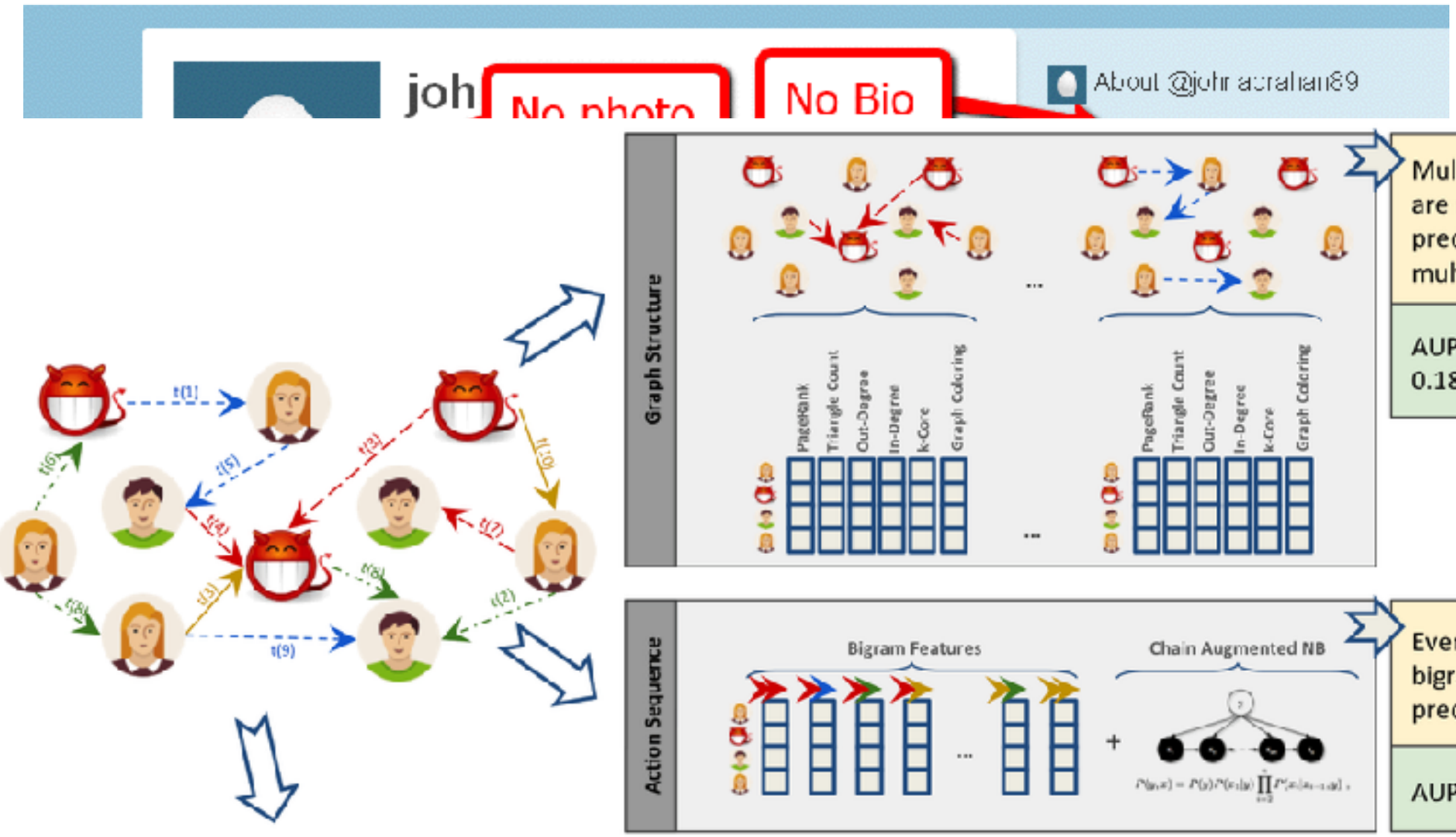
Social Media Education

Computational Social Science

Social Media Marketing

Social Media Visualization

Social Spammer Detection



Multiple relations are more predictive than multiple features

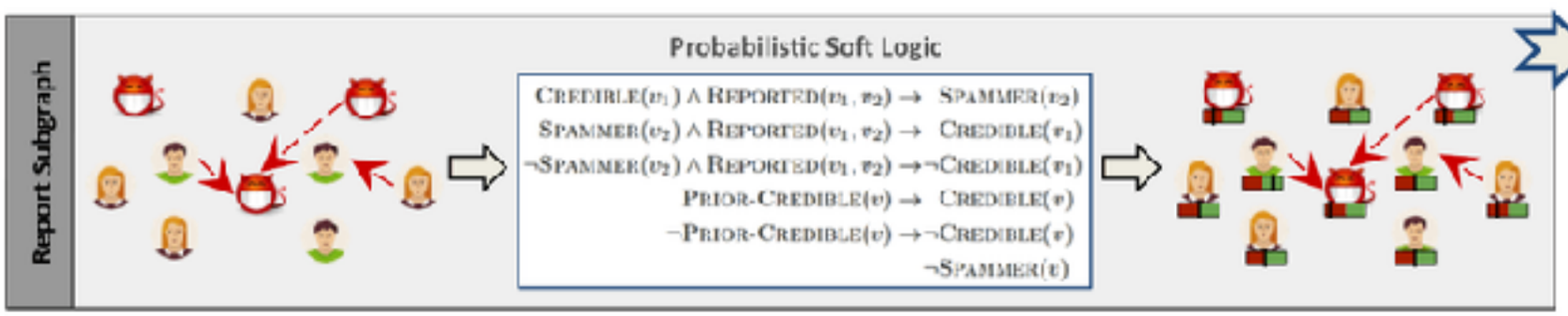
AUPR:
0.187 → 0.328

Even simple bigrams are highly predictive

AUPR: 0.471

Can classify 70% of the spammers that needed manual labeling with 90% accuracy

AUPR: 0.779

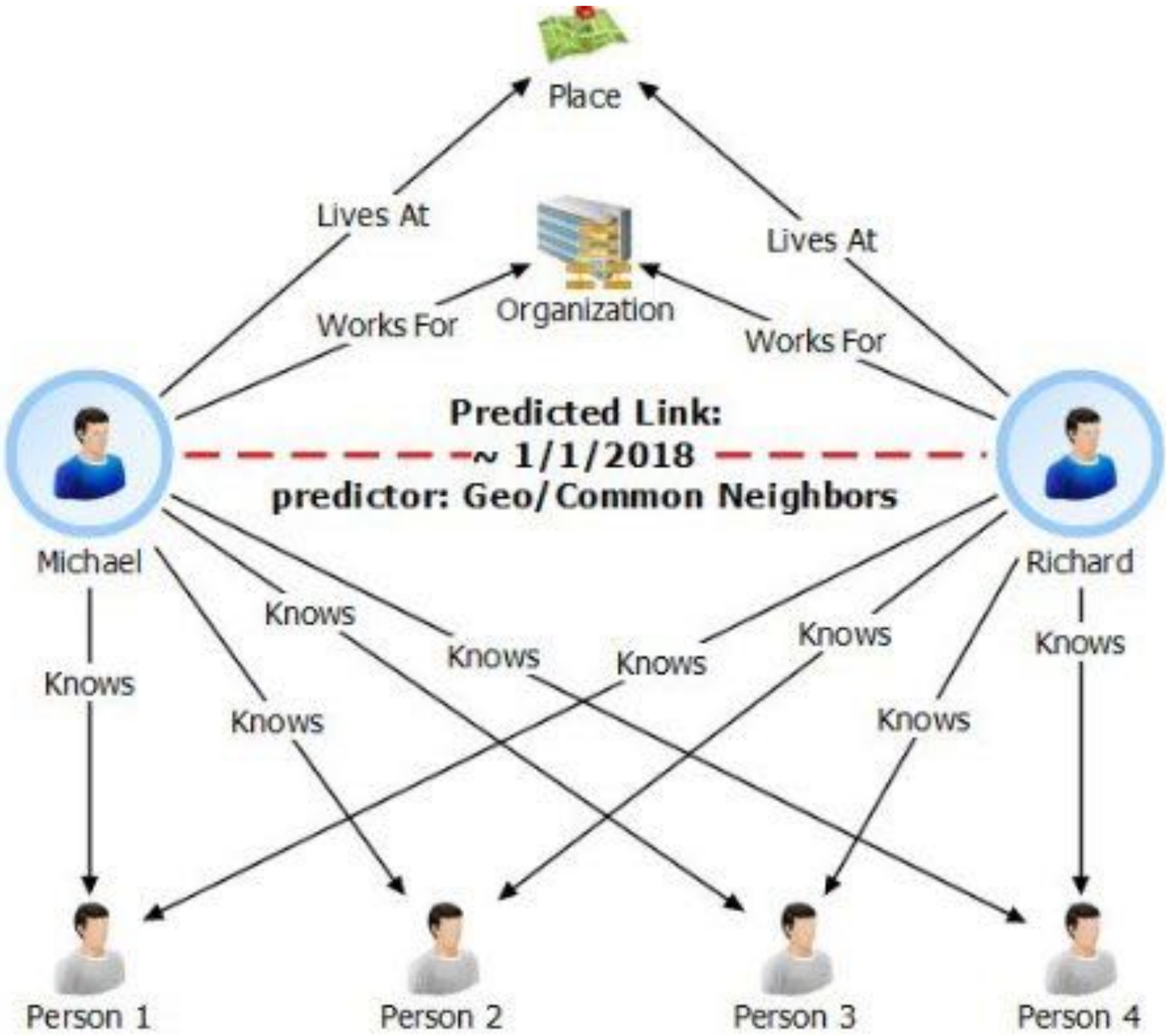


Jointly refining the credibility of the source is highly effective!

AUPR:
0.674 → 0.884

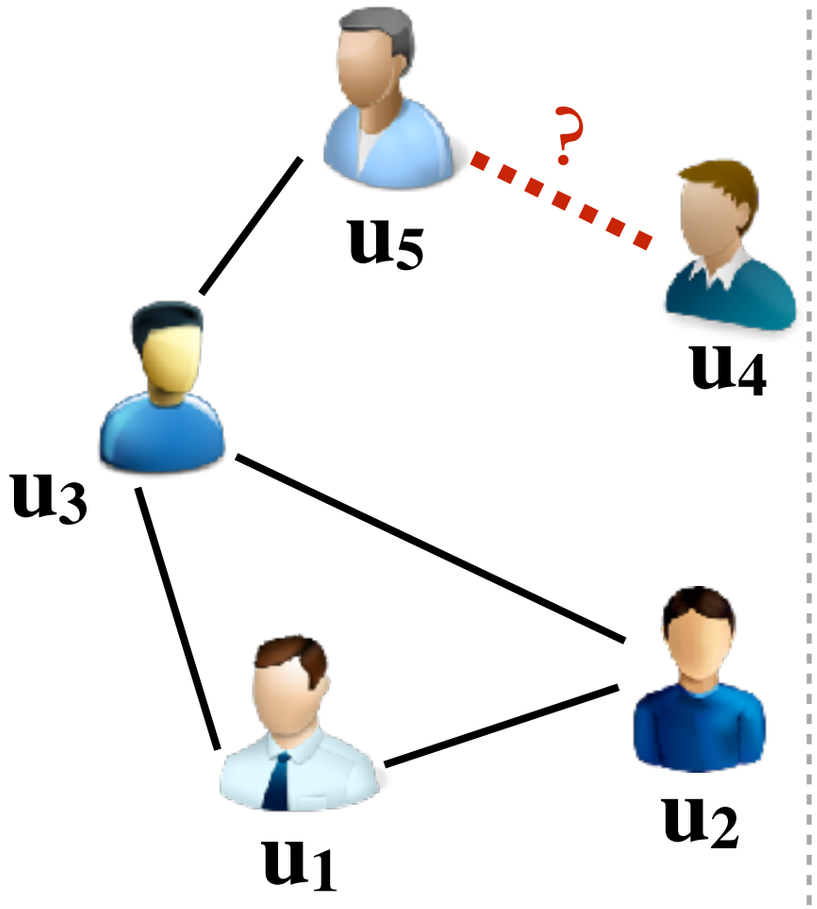
Collective Spammer Detection in Evolving Multi-Relational Social Networks

Social Ties Prediction



Supervised Link Prediction Setting

network structure



information used to extract feature vectors for these links

	link	features	label
existing links	(u_1, u_2)		+1
	(u_3, u_5)		+1
non-existing links	(u_3, u_4)		-1
	(u_4, u_2)		-1

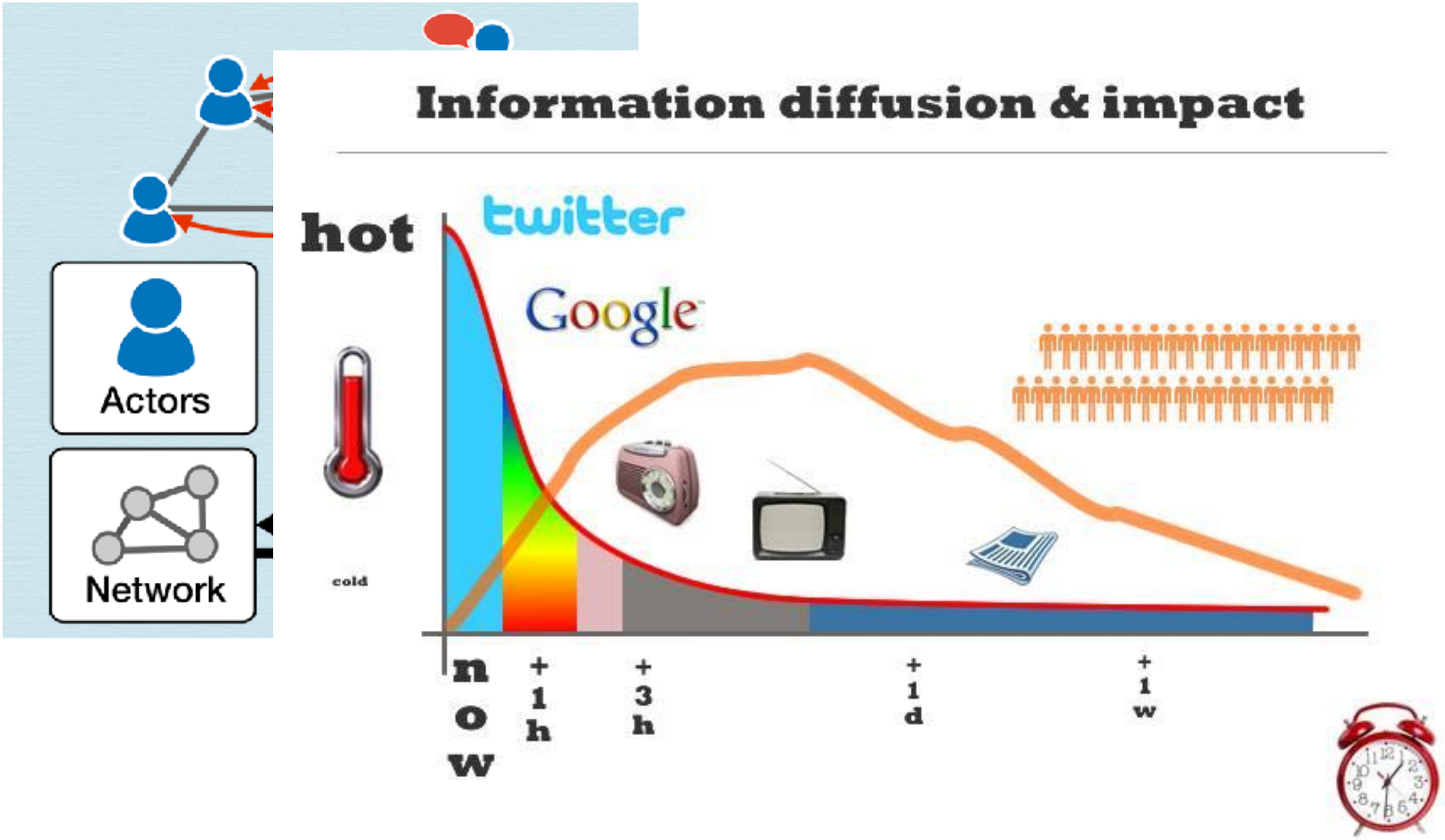
link to be predicted

(u_5, u_4)

supervised learning model

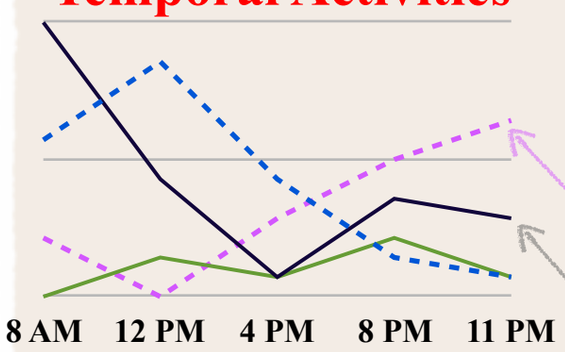
label/score

Information Diffusion

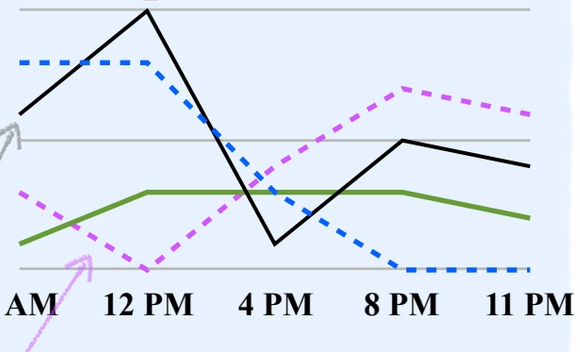


Social Network Alignment

Temporal Activities



Temporal Activities



User Accounts

User Accounts

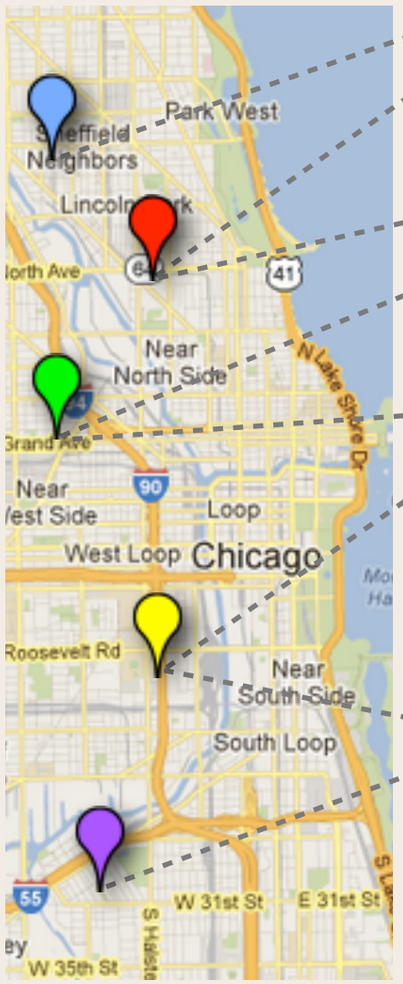
anchor links

?

?

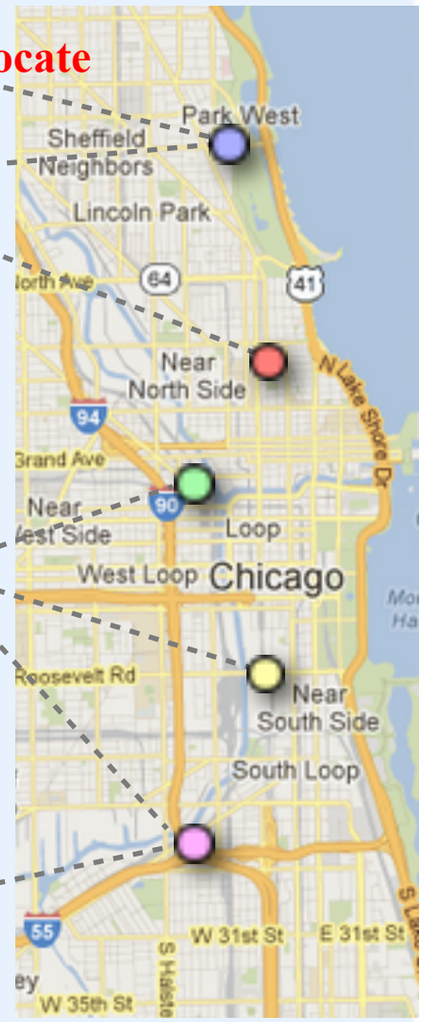
?

Locations



Locations

locate

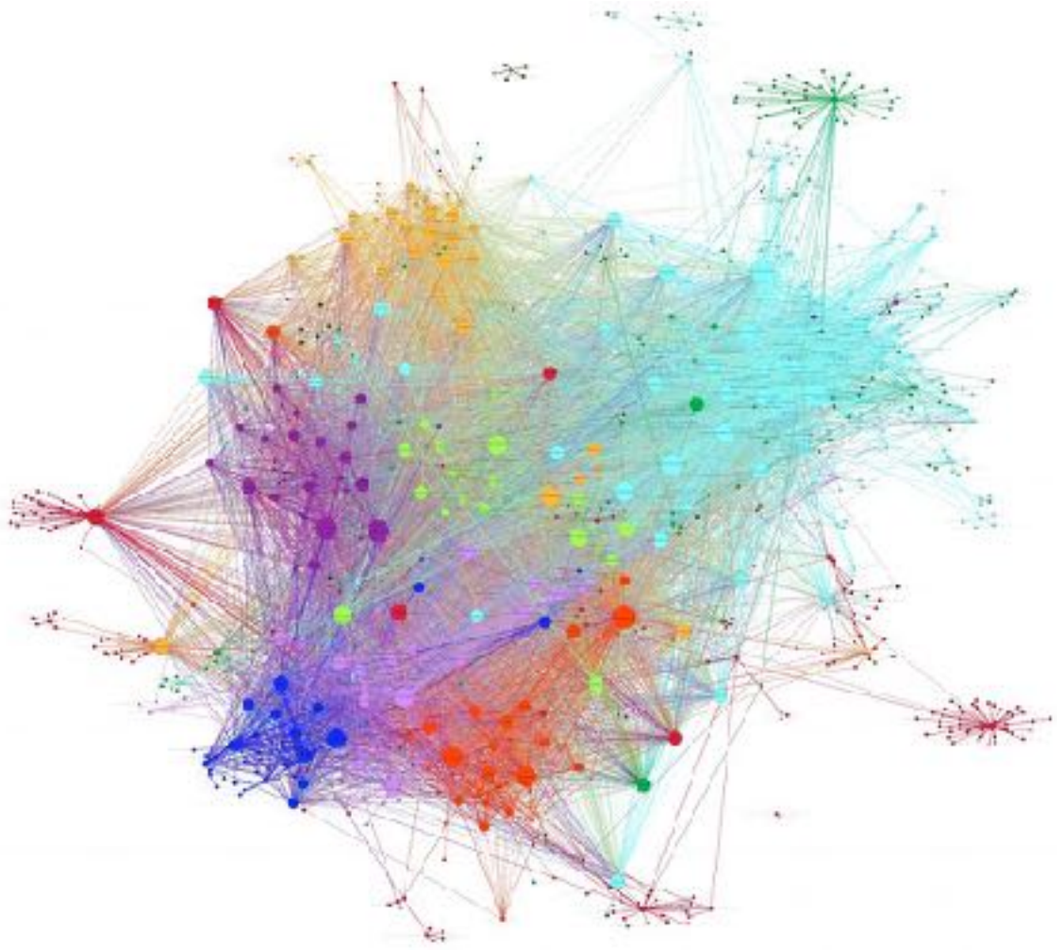


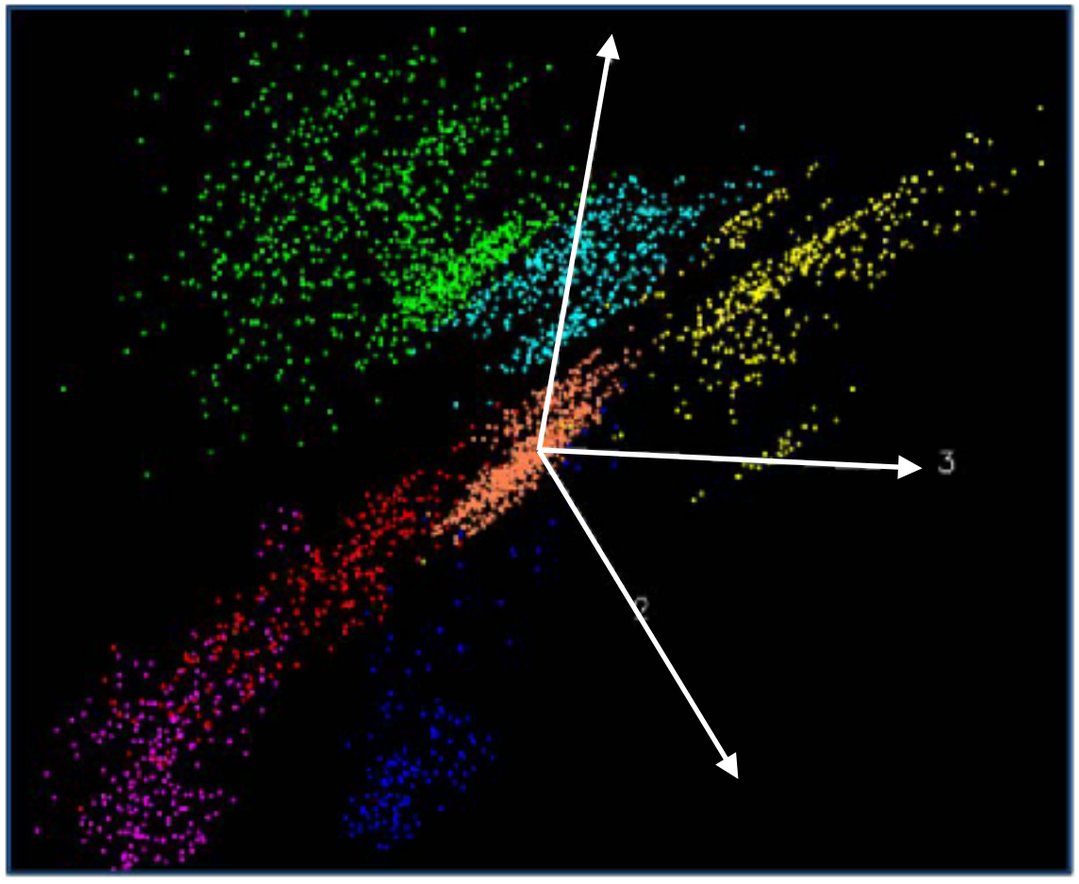
locate

Tips

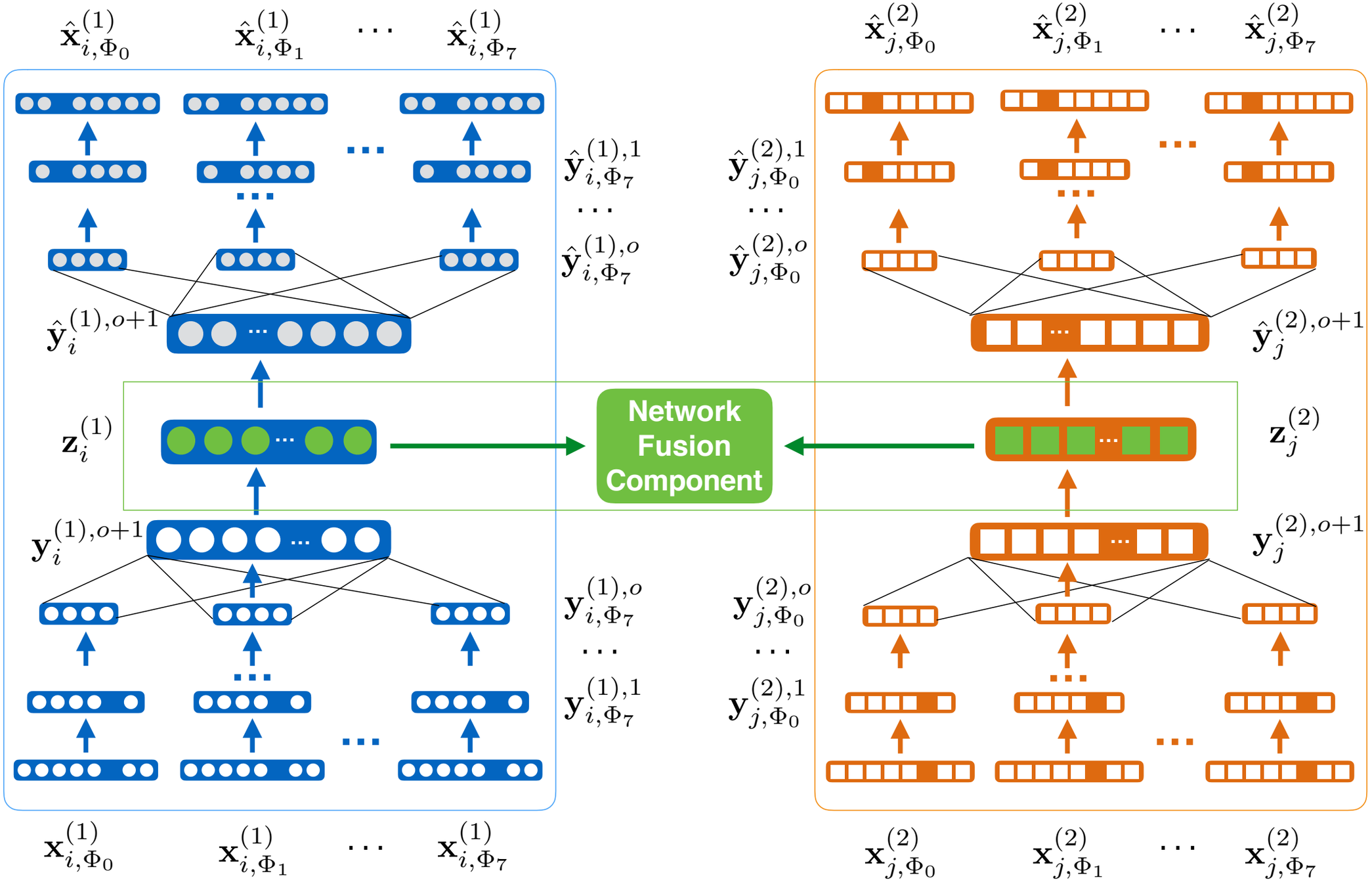
Tweets

Network Embedding





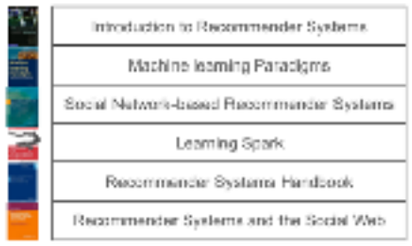
Network Embedding



Social Recommendation

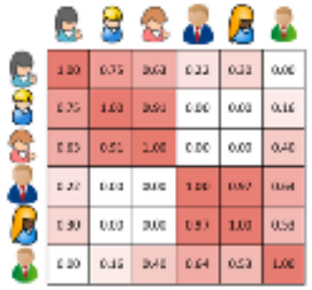
INPUT

4	3			5	
5		4		4	
4		5	5	4	
	3				5
	4				4
		2	4		5

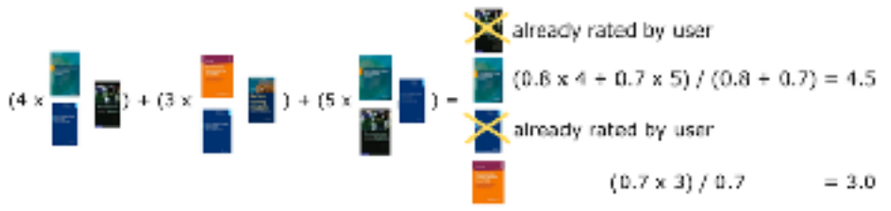


Introduction to Recommender Systems
Machine learning Paradigms
Social Network-based Recommender Systems
Learning Spark
Recommender Systems Handbook
Recommender Systems and the Social Web

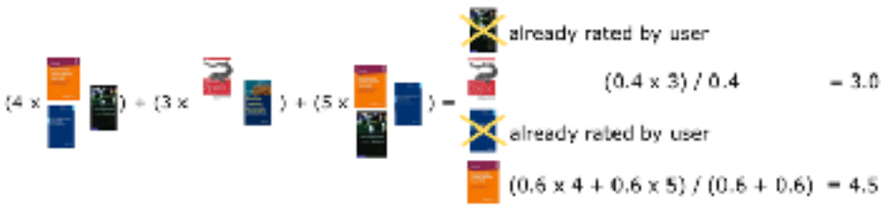
USER-BASED COLLABORATIVE FILTERING



ITEM-BASED COLLABORATIVE FILTERING



CONTENT-BASED FILTERING

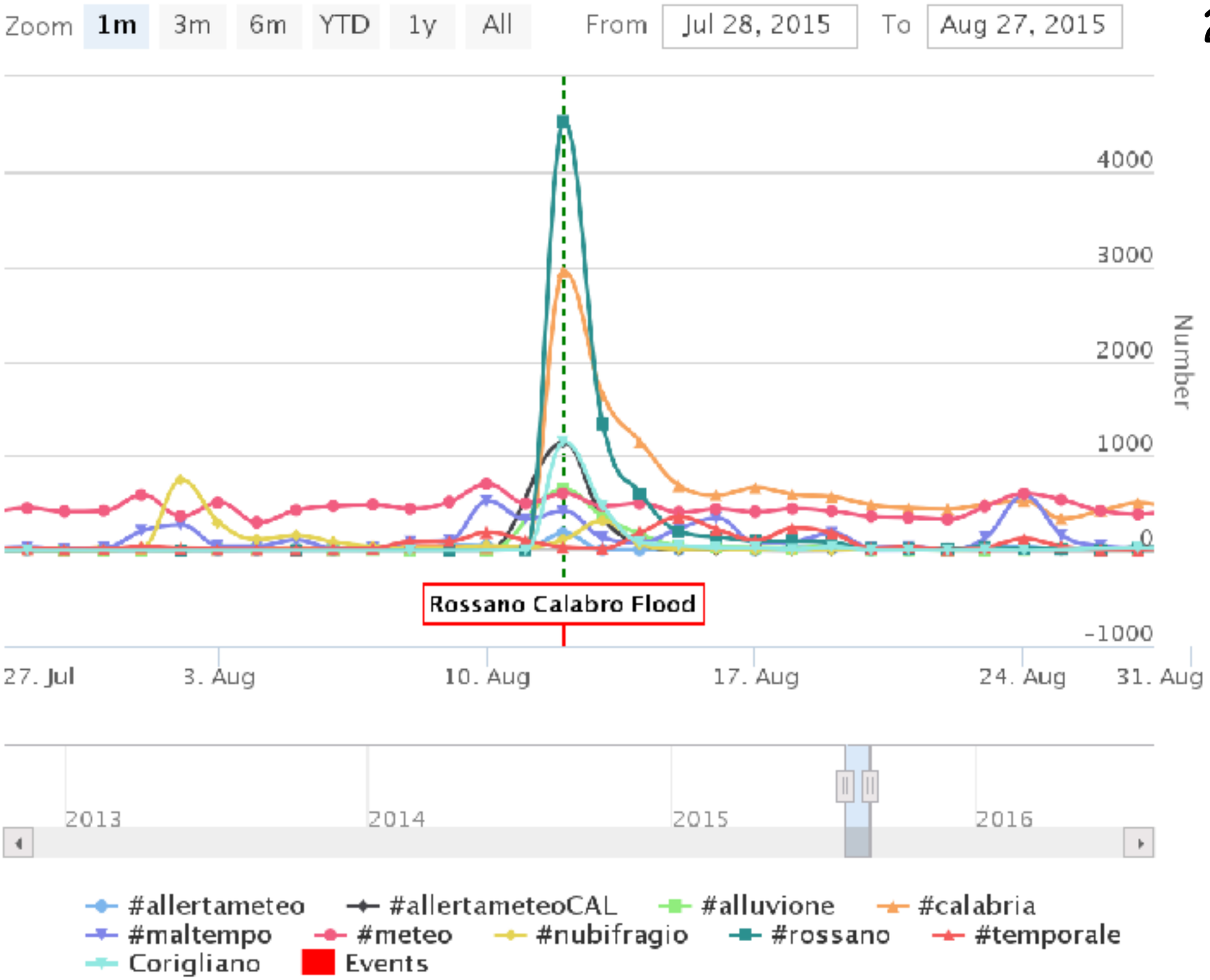


HYBRID

OUTPUTS

FOUR RECOMMENDER ALGORITHMS ARE FED THE SAME INPUT AND PRODUCE DIFFERENT OUTPUTS.
THEIR DEFINITIONS OF USER AND/OR ITEM SIMILARITY ACCOUNT FOR MOST OF THE DIFFERENCES.

Social Event Detection



2015 Twitter Channels for Rossano Flood

Social Network Mining Problems: An Overview

User-Centric

Role Analysis

Social Spammer Detection

Social Ties

Negative Links

Information Diffusion

Network Alignment

Network Summarization

Network Embedding

Content-Centric

Misinformation

Event Detection

Content Quality and Popularity

Sentiment Analysis

Social Tags

Social Summarization

Social Recommendations

Social Media Q&A

Interdisciplinary

Personality Analysis

Crisis Informatics

Social Media Healthcare

Social Media Privacy and Security

Social Media Education

Computational Social Science

Social Media Marketing

Social Media Visualization

An Advertisement:
CIS5930-04 Social Network Mining
Call for Enrollment

CIS5930-04 Call for Enrollment

http://www.ifmlab.org/recent_news.html

IFM Lab

Information Fusion and Mining Laboratory

HOME NEWS PEOPLE PUBLICATIONS LEARNING RESEARCH CLONE IFMLAB

Recent News

Timeline

OCTOBER 18, 2017

Course Call For Enrollment

A new course entitled Social Network Mining will be offered to the graduate students in 2018 Spring. This course will cover some advanced topics like 'social network mining', 'graph mining', 'deep learning', 'broad learning', 'text mining' and 'recommender system', etc. We are calling for enrollment.

October 16, 2017

IEEE BigData

IEEE BigData Notification

Two research papers Inverse Extreme Learning Machine for Learning with Label Proportions and PUDroid: Positive and Unlabeled Learning to Detect Zero-day Android Malware are accepted by IEEE BigData 2017.

CIS5930-04 Course Page: http://www.ifmlab.org/courses.html

CIS 5930-04 Social Network Mining (*syllabus*)

Course Information

CIS 5930-04 Social Network Mining
Date: Jan 8-Apr 27, 2018
Time: M/W 3:35-4:50PM
Location: Lov 103

Instructor Information

Instructor: Jiawei Zhang
Email: jzhang [AT] cs.fsu.edu
Office: 171 James Love Building
Office Hours: M/W 5:00-6:00PM

Teaching Assistant: TBD
Email: TBD
Office: TBD
Office Hours: TBD

Course Description

The course on Social Network Mining focuses on selective areas of importance in social network mining, data mining and machine learning. Social network mining is a recently emerged hot topic in machine learning and data mining research, and solutions developed in the research field substantial impacts in many important applications. Selective topics will be covered in the Social Network Mining course.

Topics Covered

- Social Network Mining & Graph Mining
- Deep Learning
- Broad Learning
- Text Mining
- Recommender System
- Misc.

Textbook

No required textbook.

Prerequisites

You are expected to have background knowledge in Data Structure, Algorithm, Discrete Mathematics. You will also need to be familiar with basic Linear Algebra, basic Statistics, and can master at least one programming language and have programming experiences.

Course Format

The objective of this course is to familiarize students with the latest research topics related to social network mining, data mining and machine learning. Course activities include 1) paper reading and paper review; 2) paper presentation and discussion; and 3) research oriented course paper writing.

- **Paper Reading and Paper Review**: Each class will discuss one academic paper. Before class, the students should read the paper to be presented in class, and write a short review (no longer than 1 page) for the paper. The review should cover: (a) a summary of the paper; (b) 3 strong points of the paper; (c) 3 weakness of the paper; (d) potential ideas of future works;

Presentation Schedule and Progress

Schedule Sheet Link

Time

Presentation

Final Paper

Grading Policy

In-class presentation: 30%. Powerpoint presentation needs to be submitted on the day of the presentation, before 11:59PM (midnight) of your presentation day. Copying existing presentation from the web is regarded as plagiarism.
Course participation and Q&A: 20%. A summary/review of each in-class discussion paper needs to be submitted before each class starts (i.e., before 3:35PM M/W). During the class, presenter and audiences can have Q&A with the pre-prepared questions in the review report.
Course paper: 50%. Single-authored original work on social network mining/data mining/machine learning. Not recycled published/submitted/on-going work with another faculties or classes.

- **Paper proposal: 10%.** Due on March 5, 2018, 11:59PM (midnight).
- **Final paper: 40%.** Due on April 25, 2018, 11:59PM (midnight).

Final Grade

- A: 100-90; A-: 90-85;
- B+: 85-80; B: 80-75; B-: 75-70;
- C: 70-60;
- F: 60-0.

Late Submission Policy

- Late paper review submission will not be accepted. During the semester, you can miss up to 2 paper reviews without penalty.
- Late presentation slides submission will not be accepted.

Social Network Mining: An Introduction

Thank You!

Jiawei Zhang

jzhang@cs.fsu.edu

Florida State University