Social Computing: Introduction

Mainack Mondal

CS 60017
Autumn 2021

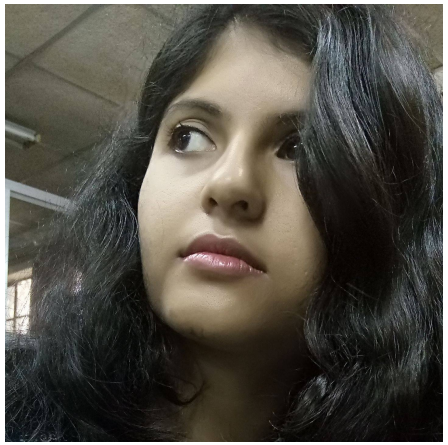
- Course logistics
- Social networks in the offline and online worlds
- Research challenges on social networks / social media

Instructors

- **Mainack Mondal:** social systems, usable security and privacy, system security and privacy
  - Office: CSE 316
  - Also teaching Usable Security and Privacy

TAs (1)

Paramita Das



Rima Hazra

TAs (2)

Sayantan Adak

Jagdish Singh Samant

TAs (3)

Ankan Mullick

Website

- https://cse.iitkgp.ac.in/~mainack/courses/2021-autumn/socialcomp/
  - Read the instructions carefully
  - Check the announcements
  - Regularly visit

Course timings

- Credit : 3 – 0 – 0

- Wednesday 11:00 am - 11:55 am
- Thursday 12:00 noon - 12:55 pm
- Friday 8:00 am – 9 am

Mode of teaching

- Live lectures via zoom
  - Zoom link will be announced in MS classroom
  - Might not be recorded, please attend classes

- Pre-recoded lectures
  - We will upload the recorded lectures via MS teams

CSE Moodle

- All submissions will be via CSE moodle unless otherwise stated
  - CSE Moodle
    - https://moodlecse.iitkgp.ac.in/moodle/login/index.php
    - Course name: Social Computing (Autumn 2021-22)
    - Enrollment key: STU60017 (checking with the admins)

Course evaluation: Test / viva

- Three tests / Viva (40%)
  - Syllabus : Everything until that point
  - Dates are in the webpage

Course evaluation: Assignment

- 40% of the evaluation
  - 3/4 in number
  - Programming based
  - Will test how well can you apply the knowledge acquired in the class

Course evaluation: Research papers

- 20% of the evaluation
- Requirements:
  - Present research papers (in class)
  - Present in groups
  - The covered papers will be part of the test
  - More details will be announced shortly

Course logistics

- Questions?

- Course logistics

- Social networks in the offline and online worlds

- Research challenges on social networks / social media

Social networks in off-line world

- Social networks studied for several decades
  - Friendship networks among students of a school, members of a club, ...
  - Collaboration networks among scientists, movie actors, ...
  - Citation networks: scientists / papers referring to other scientists / papers

What are possible rules? Ideas

- Proximity – how close the person is sitting or living
- Common interest
- Organization
- Common friends – mutual friends
- Common experiences

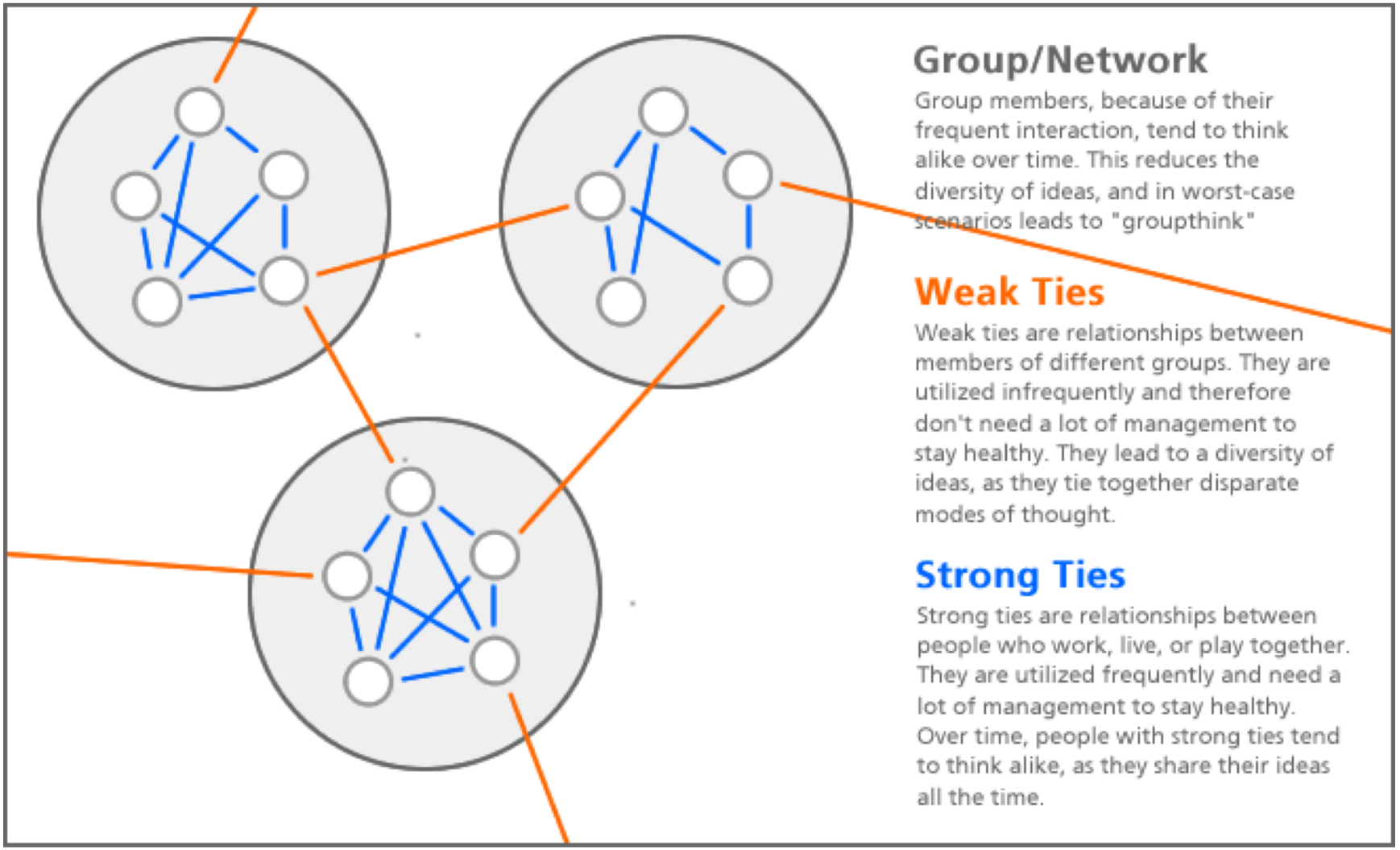
Sociological theories

- Several sociological theories developed
  - Homophily – birds of a feather flock together
  - Six degrees of separation - Milgram's experiments (1967)
  - Strength of weak ties (1973)
  - Spread of epidemics / conventions / news / rumors

Milgram's experiment in 1967

- Sent packets to people in Omaha, Nebraska and Wichita, Kansas
  - You need to get the packets to a specific person in Boston
  - If you know the recipient, send the packet directly to him
  - If not, think of a friend you know, who is likely to be closer to the recipient in Boston; sign your name to a roster, and send the packet to your friend
- Boston recipient examined the roster and saw how many steps it took for the letter to arrive
- 64 letters reached recipient, average number of links: between 5 and 6

Strength of ties



Group/Network

Group members, because of their frequent interaction, tend to think alike over time. This reduces the diversity of ideas, and in worst-case scenarios leads to "groupthink"

Weak Ties

Weak ties are relationships between members of different groups. They are utilized infrequently and therefore don't need a lot of management to stay healthy. They lead to a diversity of ideas, as they tie together disparate modes of thought.

Strong Ties

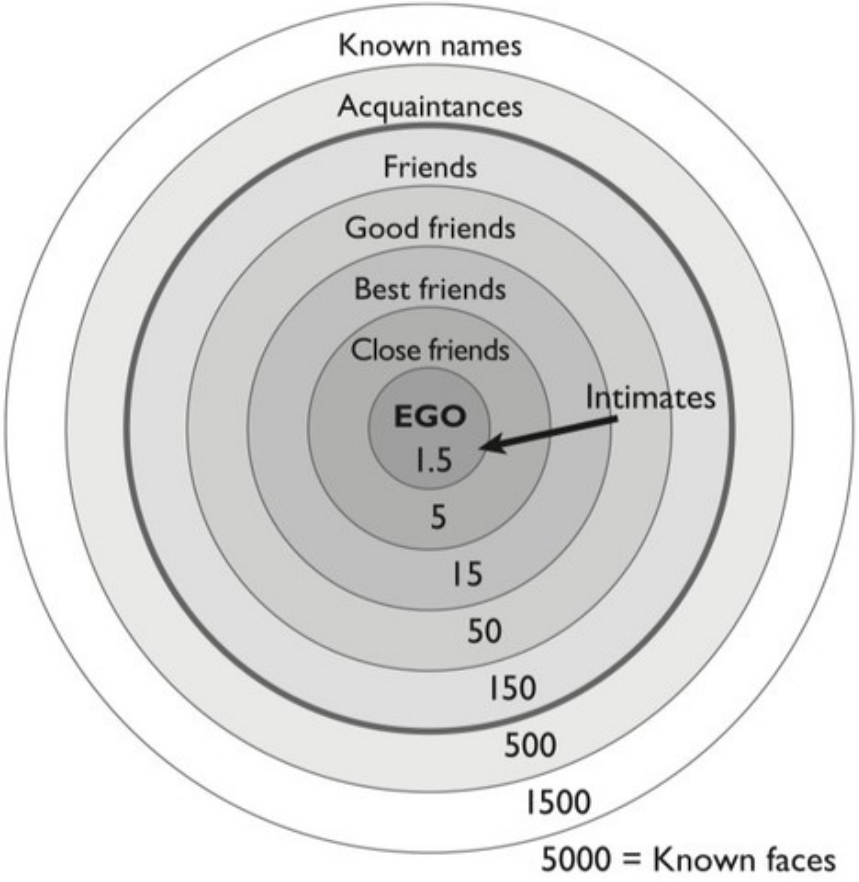
Strong ties are relationships between people who work, live, or play together. They are utilized frequently and need a lot of management to stay healthy. Over time, people with strong ties tend to think alike, as they share their ideas all the time.

Dunbar number and circles

Dunbar number and circles

- Robin Dunbar checked the frequency of sending Christmas cards over the years for a number of people
  - The more frequently sent, the closer they are
  - Then via statistical analysis he found that these frequencies are discretized
  - There are “levels” of friendship

Dunbar number and circles



https://www.theatlantic.com/family/archive/2021/05/robin-dunbar-explains-circles-friendship-dunbars-number/618931/

Today's online social networks

Online Social Networks (also called social media and OSN)

- Among the most popular sites on today's Web
- Billions of users world-wide
  - Celebrities, media houses, politicians, commoners, …
  - Spammers, cyber-bullies, hatemongers, …
- Huge impact
  - Advertisers reach large population at minimal cost

OSN in research

- Huge data readily available
  - Volume – networks of billions of users, petabytes of user generated content every day
  - Variety – text, image, speech, video, …
  - Velocity – thousands of posts / minute during major events

- Automated data collection rather than surveys

Multi-disciplinary research on OSNs

- Computer networks & distributed systems
- Sociology, social psychology, linguistics, …
- Network science, complex network theory
- Data mining, machine learning, information retrieval, natural language processing, privacy…

Multi-disciplinary research on OSNs

- Computer networks & distributed systems
- Sociology, social psychology, linguistics, …
- Network science, complex network theory

THIS COURSE

- Data mining, machine learning, information retrieval, natural language processing, privacy …

Two important aspects in a social network / social media

- The network structure
- The content

- This course
  - First part – network structure – we will consider only simple networks
  - Second part – content – we will focus on only textual content

Social networks vs. Social media

- “What is Twitter, a Social Network or a News Media?”, Kwak et al., WWW 2010

- What is the difference between social network and social media?
  - Social network – interpersonal links, interactions in focus
  - Social media – information exchange in focus
- Sometimes these two influence each other
  - Case study: Why do FB sometimes show things that you were texting about on whatsapp (which is encrypted)?

- Course logistics
- Social networks in the offline and online worlds
- Research challenges on online social networks (OSN) / social media

1. Modelling OSN

How to model / represent OSNs?

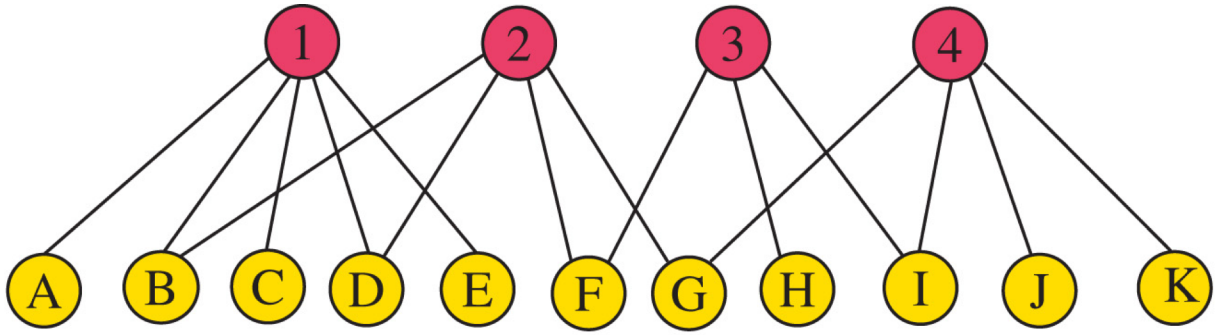
- Most common representation: a graph
  - Nodes: users, edges: social links

- Undirected networks: Facebook
- Directed networks: Twitter
- Weighted networks
  - Edge-weights usually measure “strength” of social link, e.g., number of interactions

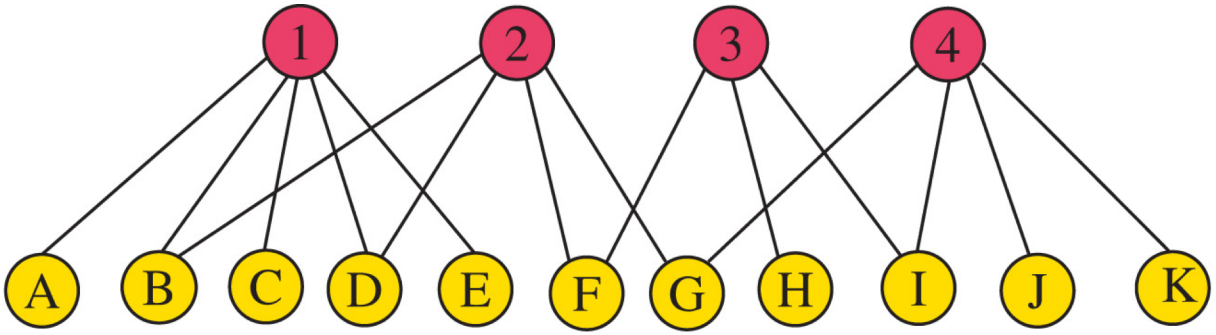
Graph models of OSNs

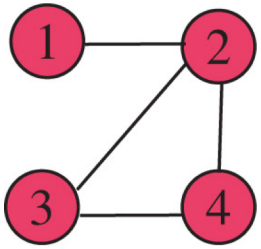
- Other varieties of networks
  - Networks among blogs, videos, …
  - Bipartite networks, e.g., viewer-video model of Youtube
  - Folksonomy: Users annotate resources with tags,modeled as tri-partite hypergraphs [Cattuto, AI Communications 2007]

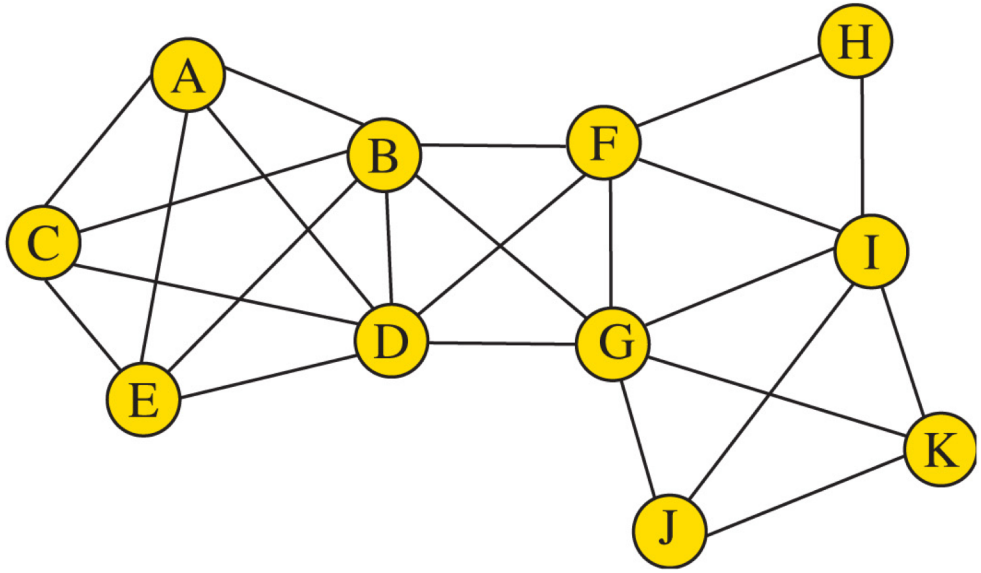
Bipartite networks and projections



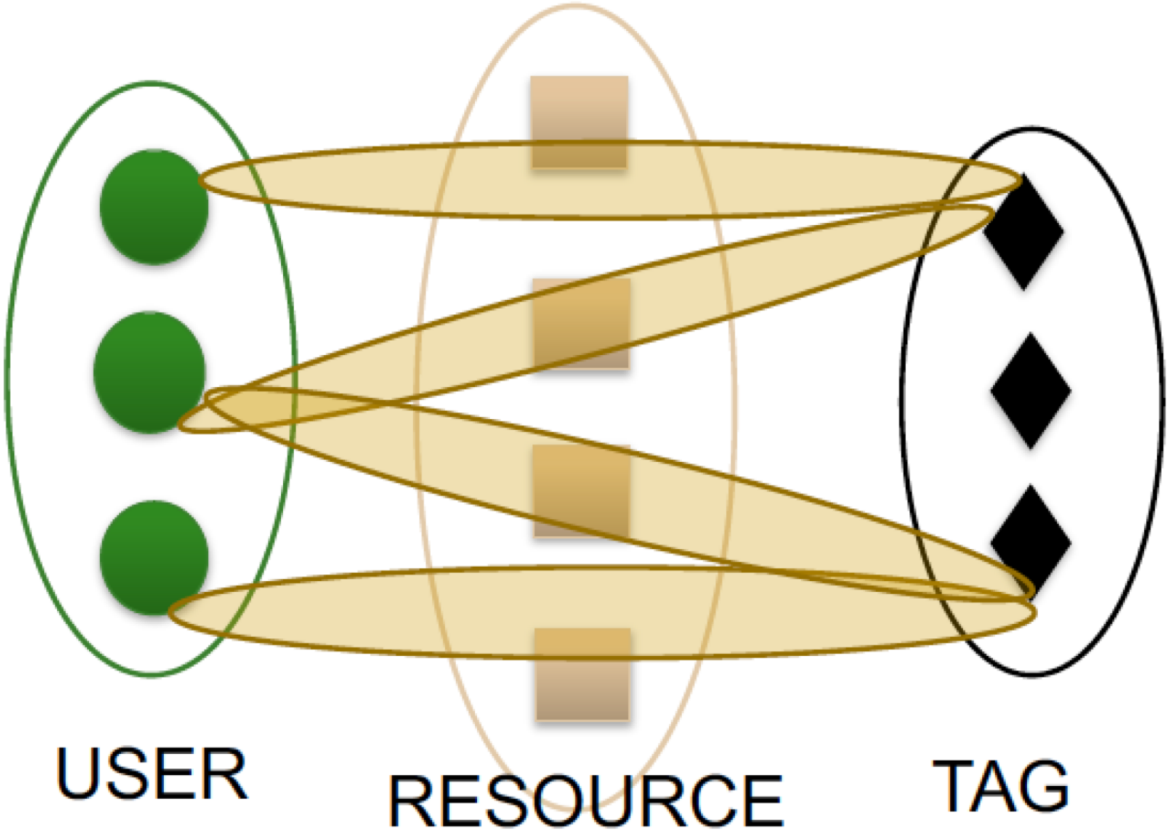
Bipartite networks and projections







Tri-partite model for folksonomies



Sociological differences between offline and online graph

- Sociological theories investigated on OSNs
  - Homophily, strength of weak ties [Grabowicz, Plos ONE, 2012]
  - Emergence and spread of conventions [Kooti, ICWSM 2012]

- OSNs different from offline SNs in some aspects
  - Offline social networks – an individual can maintain only a certain number of meaningful social links – Dunbar number (~150); OSNs – almost zero cost of maintaining
  - social links - many more links can be maintained
  - Important users readily connect to many ordinary ones
  - Geographical distance does not matter

Locality of friendship in Facebook

- http://www.techprone.com/facebook-displays-visually-its-impact-and-spread

2. Properties of online social graphs

Network properties of OSNs [Mislove, IMC 2007]

- Most users have few links, few have many links
  - Degree distributions: power-law, exponential, …
- Presence of numerous triangles (transitivity)
- Small-world, e.g., 6 degrees of separation
- Assortativity, homophily

Explaining the network properties

- What nature of link-creation dynamics explain the empirically observed properties of OSNs?
- Several evolution models proposed
  - Global rules, e.g., Preferential Attachment [Barabasi, Science 1999]
  - Local rules, e.g., triangle closure [Kleinberg, ICWSM 2010], random walk starting from a node [Vasquez, PRE 2003]
  - Biased Preferential attachment, based on different types of users: inactive, linkers, inviters [Kumar, KDD06]
  - Co-evolution of social and content networks [Singer, Making Sense of Microposts, 2012]

Dynamic network properties

- Dynamic nature: how do properties of OSNs change with time?
  - Network density varies non-monotonically [Kumar, KDD06]
  - Assortativity varies non-monotonically [Hu, Physics Letters A, 2009]

- Models to explain temporal variation of properties

Link analysis

- Classification of social links
  - Strong and weak links (e.g. based on level of interaction) [Gilbert, CHI'09][Wilson, EuroSys09][Valafar, WOSN09] [Xiang, WWW10]
  - Some OSNs allow positive and negative links (friends and enemies)

- Variation of strength of links with time [Viswanath, WOSN09]

Centrality (importance) of nodes

- How important is a node in a network?
  - How influential is a person in a social network?
  - How important is a website on the Web?

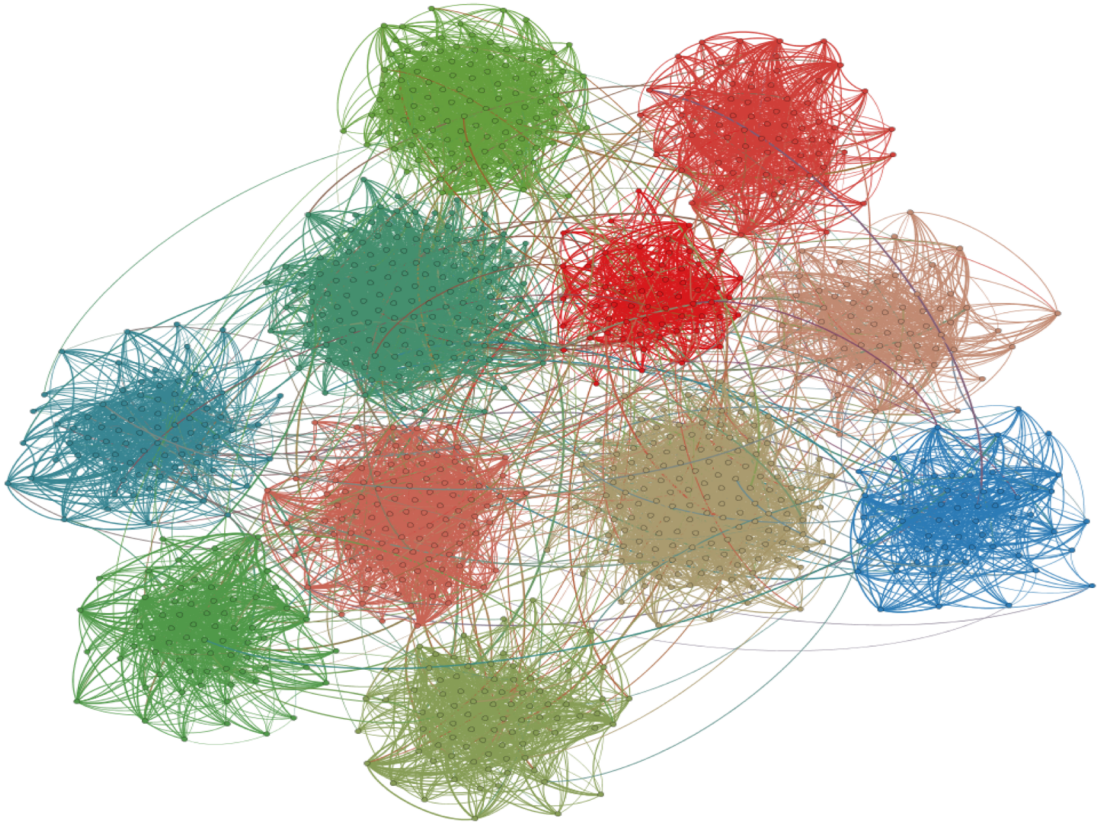
- Many proposed centrality metrics
  - Degree centrality
  - Closeness centrality
  - Betweenness centrality
  - Eigenvector centrality, PageRank

Community detection / clustering

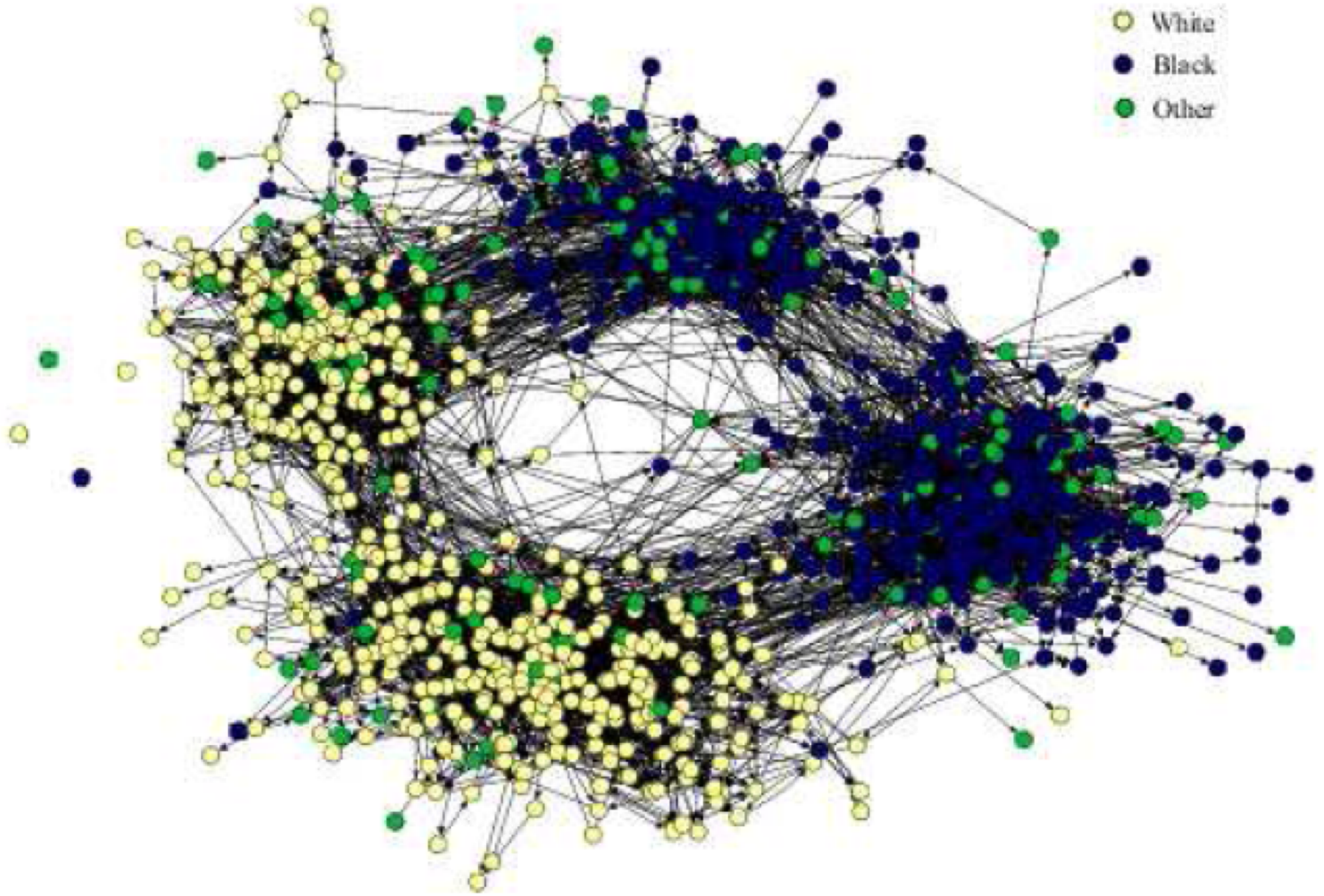
- Identifying communities of ‘similar’ users
  - Traditionally, only rely on network structure: several algorithms [Fortunato, Physics Reports 2010] [Leskovec, WWW10]
  - Content can also be leveraged in case of OSNs

- Dynamic communities: how do communities change with time? [Mitra, Computer Networks, 2012]

Community example



Friendship network among students in a US school

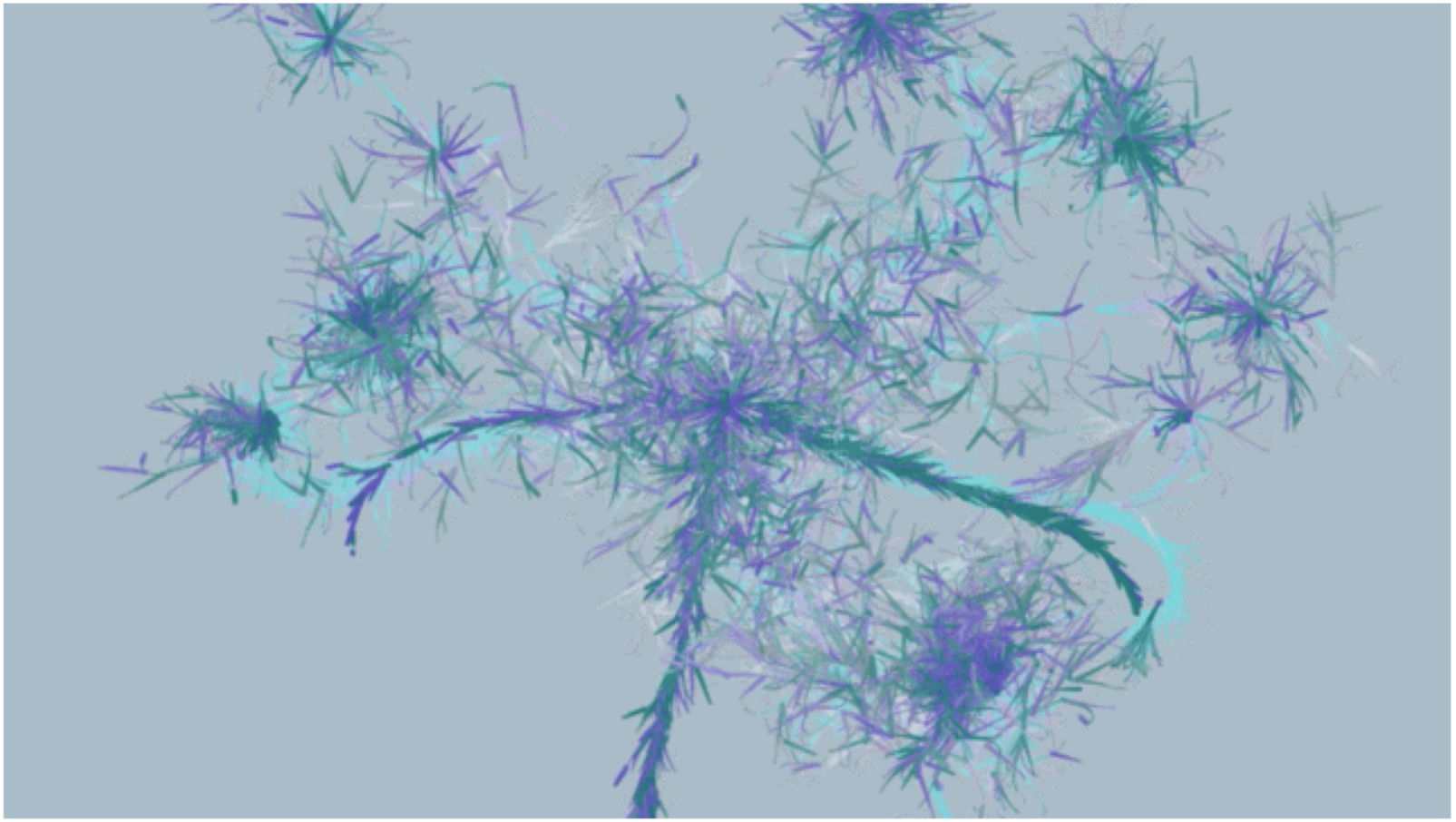


3. Modeling spread of information in OSN

Information spread / diffusion

- Understanding information spread / diffusion in OSNs
  [Cha, WWW 2009] [Lerman, ICWSM 2010] [Bakshy, WWW 2012]
  - To what extent does information (news) spread?
  - How fast? Along which links?
  - Who are the most influential in spreading information?
  - How does a topic / video become viral?

Spread of viral images in Facebook



http://www.gizmodo.com.au/2012/10/how-viral-images-spread-on-facebook-visualised/

4. Utilizing content uploaded in OSN

Utilizing information content in OSNs

- Recommendation and search
- Identifying influential users / experts
- Misinformation detection
- Authority identification
- Enabling private communication

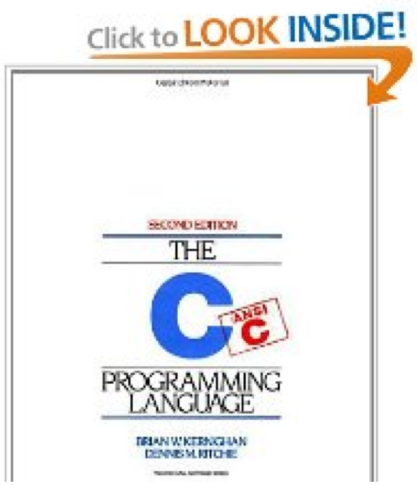
Recommendation and search

- Help users discover interesting content, friends, groups
- Motivation: The amount of information has become so large that it is impossible for an individual user to find out on her own interesting content / friends / groups
- Recommend friends, groups to join [Chen, WWW09], resources [Konstas, SIGIR09], tags [Sen, WWW09][Song, SIGIR08]
- Personalized answers to queries [Xu, SIGIR08] [Bao, WWW07] [Mislove, HotNets06]

Recommendation

- Two broad ways
  - Content-based, e.g., based on your profile information (e.g., you study in IITKGP), or some keywords given by you while creating account
  - Collaborative filtering – identify “similar” users or items – how to find “similar” users or items?

Recommendation of books in Amazon



C Programming Language (2nd Edi

Brian W. Kernighan (Author), Dennis M. Ritchie (Author)

★★★★½ (367 customer reviews)

Buy New

$52.49 & **FREE Shipping**. Details

In Stock.

Ships from and sold by **Amazon.com**. Gift-wrap available.

Want it Tuesday, June 4? Order within **55 hrs 50 mins** and choose **One-Day Shipping** at checkout. Details

Customers Who Bought This Item Also Bought

The C Answer Book: Solutions to the …

› Clovis L. Tondo

★★★★☆ (15)

Computer Systems: A Programmer's …

› Randal E. Bryant

★★★★½ (23)

Programming in C (3rd Edition)

› Stephen G. Kochan

★★★★½ (70)

Social recommendations

Similar friends might have similar tastes

Identify influential users / experts

- Several metric of influence: #followers or #friends, PageRank, number of times retweeted [Cha, ICWSM10]
- Identifying topical experts [Weng, WSDM10] [Pal, WSDM11] [Ghosh, SIGIR12]
- How to measure topic-specific expertise / interests of users?

Emotion / opinion mining

- Identify user's emotion / opinion from posts
  - Identify opinion on movies / political issues [Fang, WSDM12]
  - Comparison among different methods[Goncalves, COSN13]
  - Summarization of opinions [Ganesan, WWW12]
  - Twitter used to predict election results [Tumasjan, ICWSM10]

Various types misinformation on social media

- Spam, phishing, …
- Hate speech (against a particular social / religious /ethnic group), cyberbullying, …
- Fake news, rumors, …

Spam detection

- Identify spam / users with malicious intentions [Heymann, IEEE Internet Computing 2007]
  - Identify spam in Facebook [Gao, IMC10] ,Twitter [Lee, SIGIR10], Youtube [Benevenuto, SIGIR09], blogs [Shin, Infocom11], …
  - Sybil detection [Yu, SIGCOMM 2006][Viswanath, SIGCOMM10]

- Identifying trustworthy entities, e.g., reviews, ratings [Chandra, Trustcom 2012]

Fake news – some ideas

- First step – identify claims / factual statements that need to be verified – NLP features used
- Second step – verify the claims
- How to verify?
  - From trustworthy information sources, e.g., claims about COVID19 can be verified from medical research papers
  - Crowdsourcing from trusted people, e.g., relief workers present in the region of a disaster

Privacy of social content

- What information to share / what not to share?
- What privacy settings to choose while uploading?
  - Online social media does not have the geographical separation of online world – context collapse
- Privacy of old content
  - Privacy preferences get changed over time – users almost never go back and reevaluate privacy settings

Thank you

http://cse.iitkgp.ac.in/~mainack