

CSI4H3

Analisis Media Sosial

Lecture 1 – Introduction



Outline

- Introduction
- Definitions
- Basic Concepts



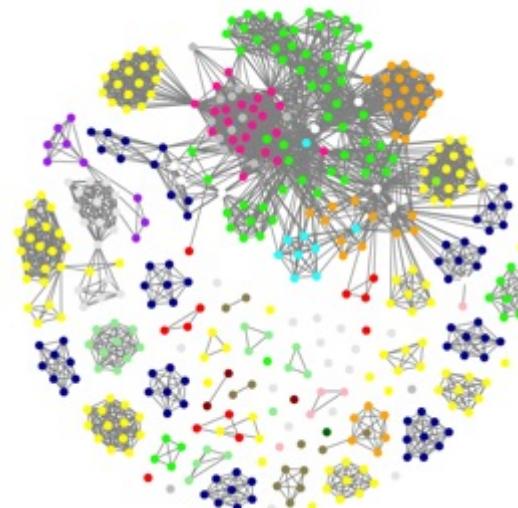


Course Learning Outcomes (CLO)

- Explain the definitions, basic concepts, and theories of social network analysis
- Understand how these concepts and theories can help explain different factors in network behaviors

Introduction

Social Network Analysis



Social Media: Many-to-Many

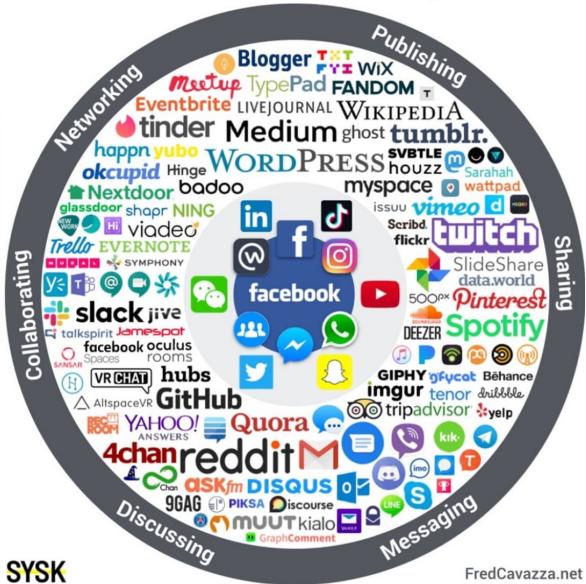
Social Media Landscape



Various forms of Social Media

- **Blog**: Wordpress, blogspot, LiveJournal
- **Forum**: Yahoo! Answers, Epinions
- **Media Sharing**: Flickr, YouTube, Scribd
- **Microblogging**: Twitter, FourSquare
- **Social Networking**: Facebook, LinkedIn, Orkut
- **Social Bookmarking**: Del.icio.us, Diigo
- **Wikis**: Wikipedia, scholarpedia, AskDrWiki
- **Publishing**: Blogging, Wiki

Social Media Landscape 2019



Social media landscape 2021



Social media landscape 2020



Social media landscape 2022



Characteristics of Social Media

- “Consumers” become “Producers”
- Rich User Interaction
- User-Generated Contents
- Collaborative environment
- Collective Wisdom
- Long Tail

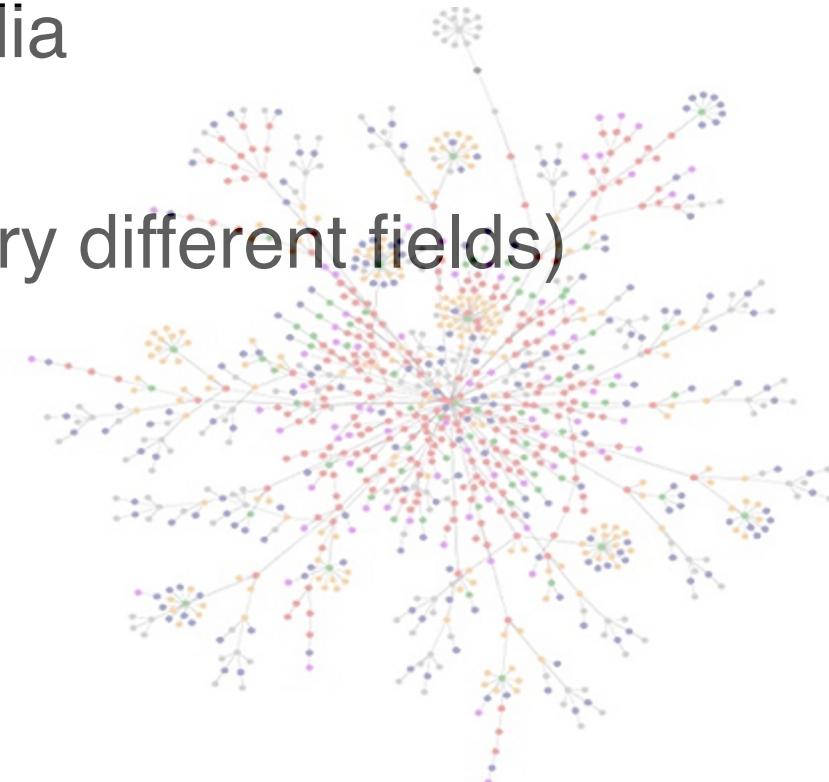


Characteristics of Social Media

- **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.

Why Networks?

- Rise of the Web and Social Media
- More data
- Shared vocabulary between (very different fields)



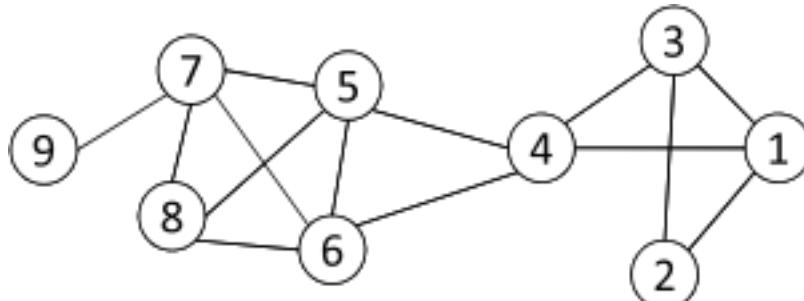
Why Networks?

- Network data is increasingly available:
 - Large on-line computing applications where data can naturally be represented as a network:
 - On-line communities: Facebook
 - Communication: Instand Messenger
 - News and Social Media: Twitter, Blogging
- Network is a set of weakly interacting entities
- Links give added value:
 - Google realized web-pages are connected
 - Collective classification

Networks and Representation

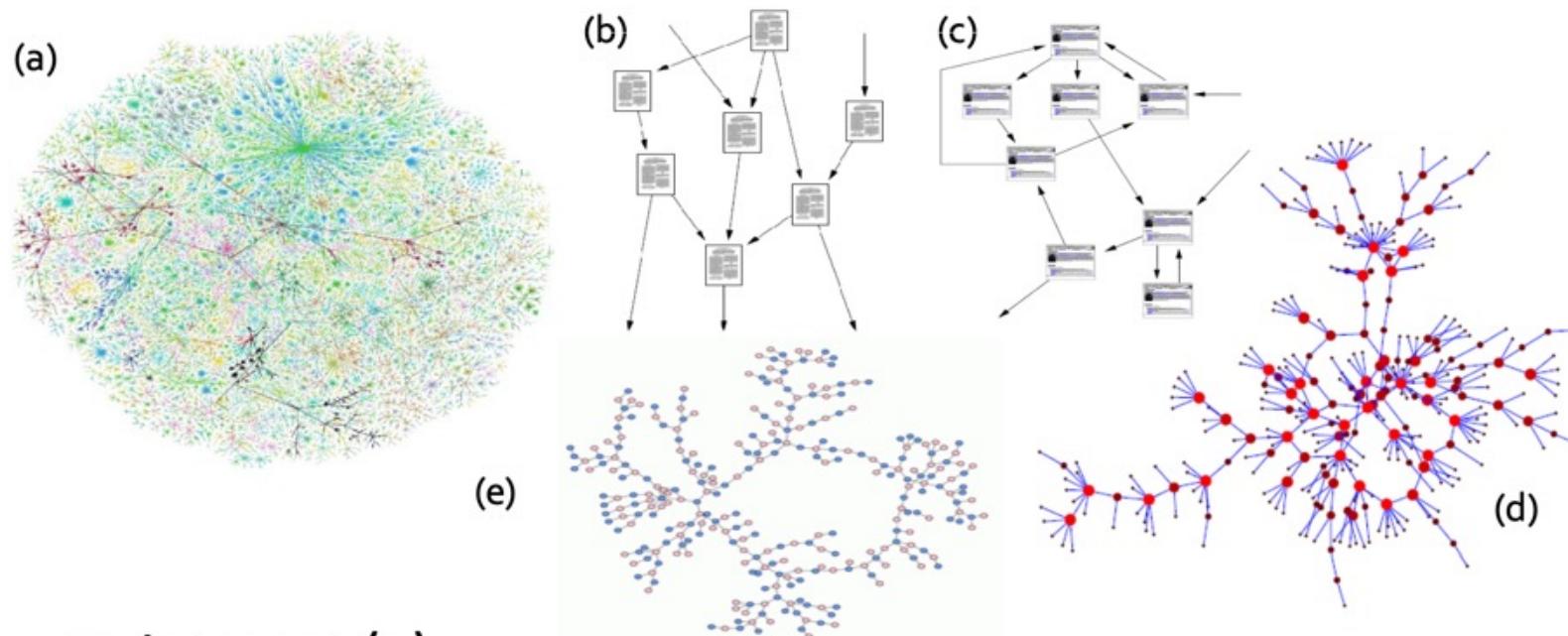
Social Network: A social structure made of nodes (individuals or organizations) and edges that connect nodes in various relationships like friendship, kinship etc.

- Graph Representation
- Matrix Representation



Node	1	2	3	4	5	6	7	8	9
1	-	1	1	1	0	0	0	0	0
2	1	-	1	0	0	0	0	0	0
3	1	1	-	1	0	0	0	0	0
4	1	0	1	-	1	1	0	0	0
5	0	0	0	1	-	1	1	1	0
6	0	0	0	1	1	-	1	1	0
7	0	0	0	0	1	1	-	1	1
8	0	0	0	0	1	1	1	-	0
9	0	0	0	0	0	0	1	0	-

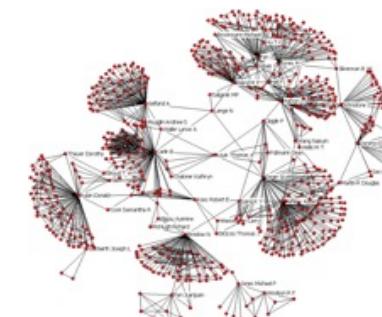
Networks: Rich data



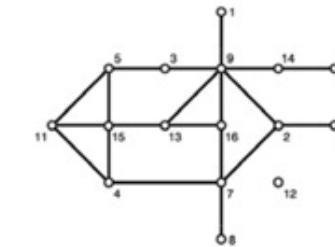
- Internet (a)
- Citation network (b)
- World Wide Web (c)
- Sexual network (d)
- Dating network (e)

Networks of the Real-world [1]

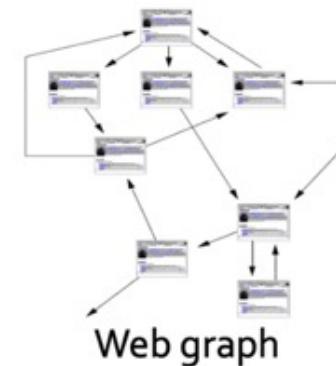
- **Information networks:**
 - World Wide Web
 - Citation networks
 - Blog networks
- **Social Networks:**
 - Organizational networks
 - Communication networks
 - Collaboration networks
- **Technological Network:**
 - Power grid
 - Airline, road, river networks
 - Telephone networks
 - Internet
 - Autonomous systems



Citation networks



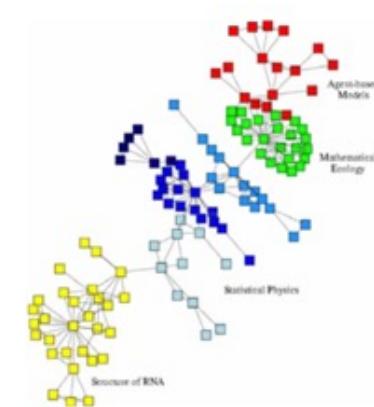
Florentine families



Web graph



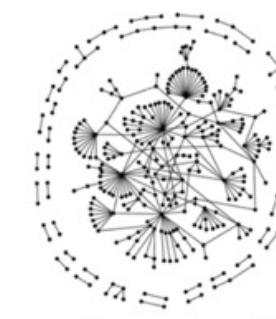
Friendship network



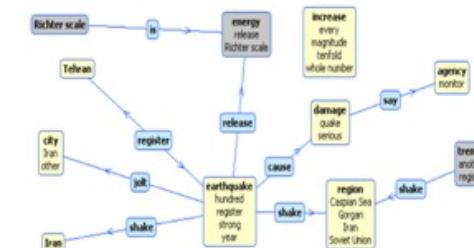
Collaboration network

Networks of the Real-world [2]

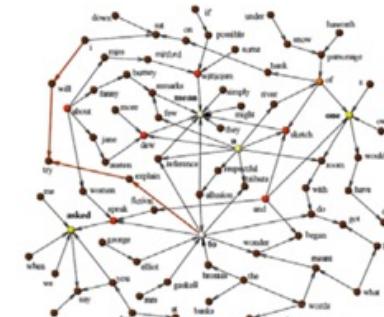
- **Biological networks:**
 - Metabolic networks
 - Food webs
 - Neural networks
 - Gene regulatory networks
 - **Language networks:**
 - Semantic networks
 - **Software networks:**
 - Call graphs



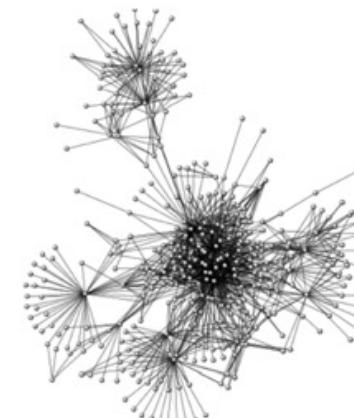
Yeast protein interactions



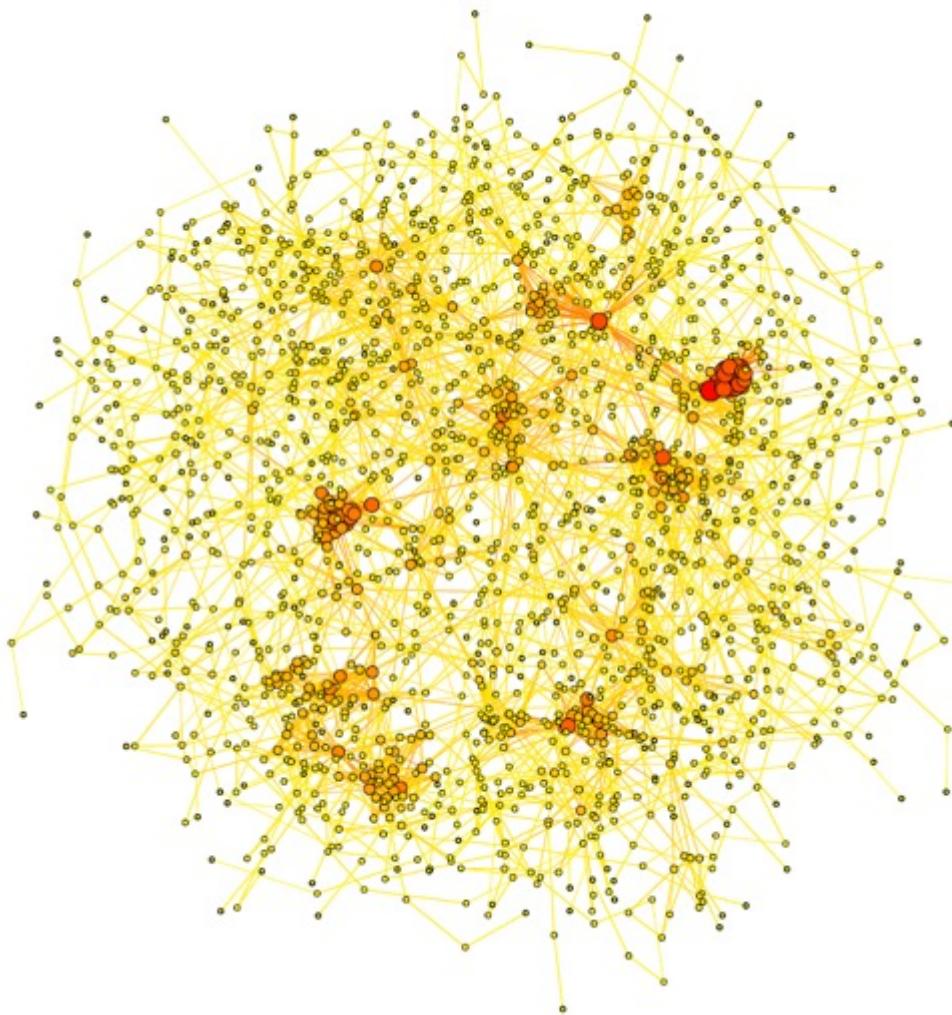
Semantic network



Language network



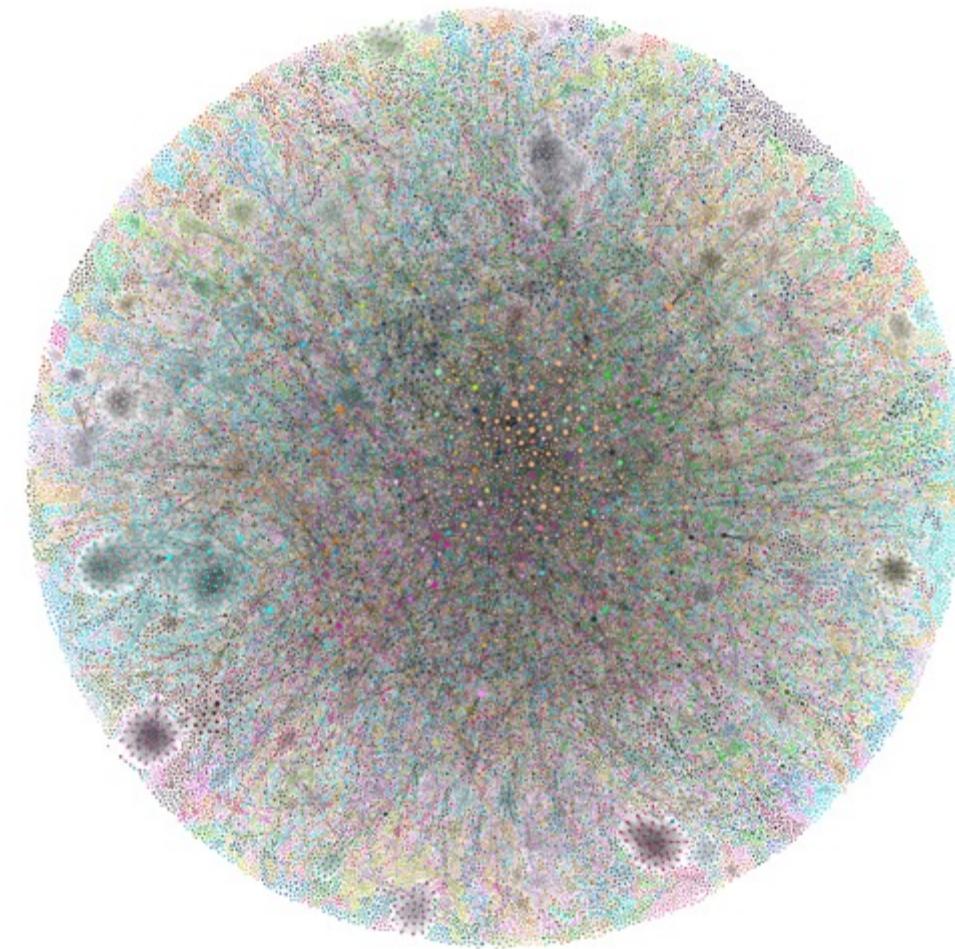
Software network



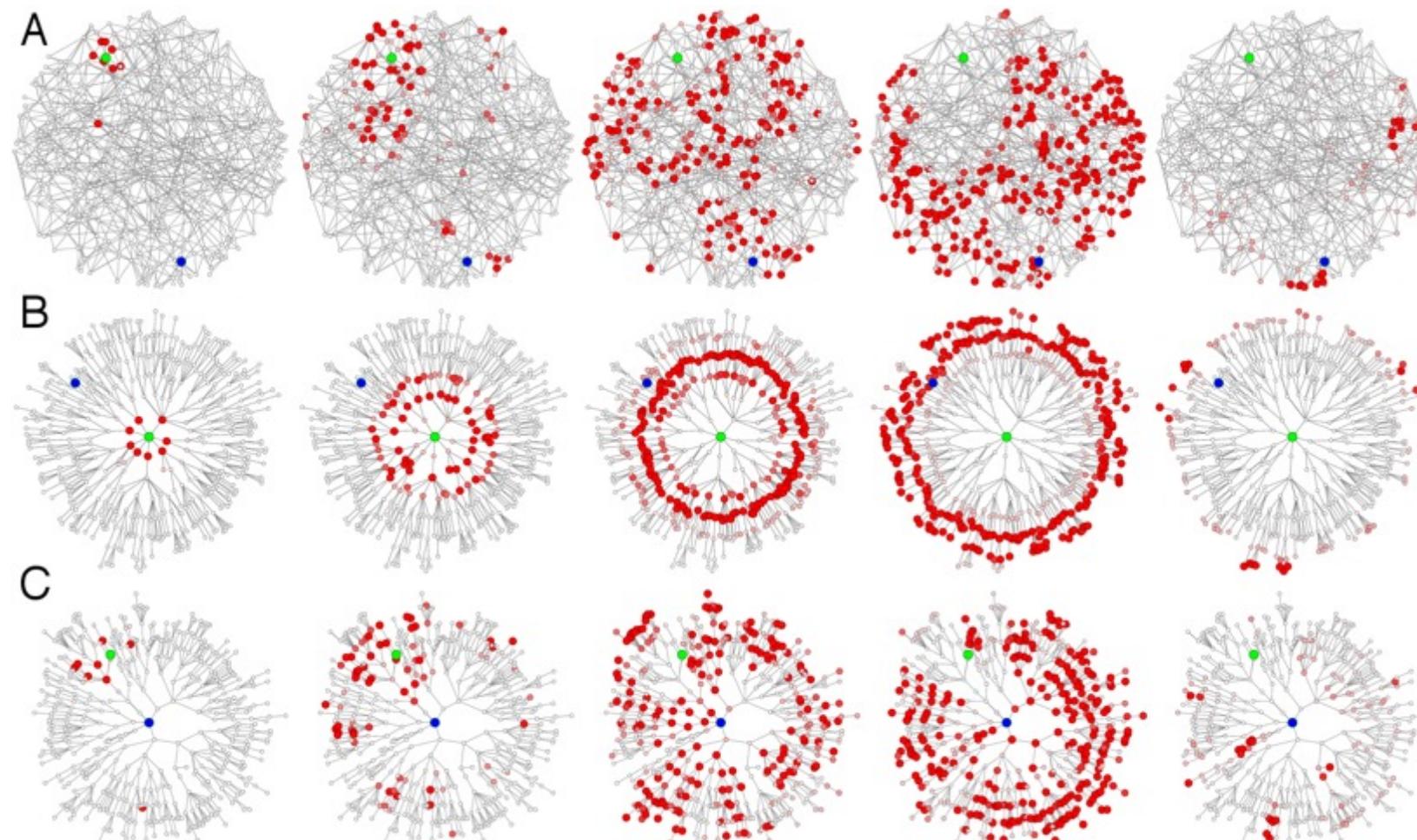
Sample of an online social network



Protein interaction network



Sweden's economic network of interlocked corporations



Three models of epidemic spread in human contact networks.

Ice Breaking

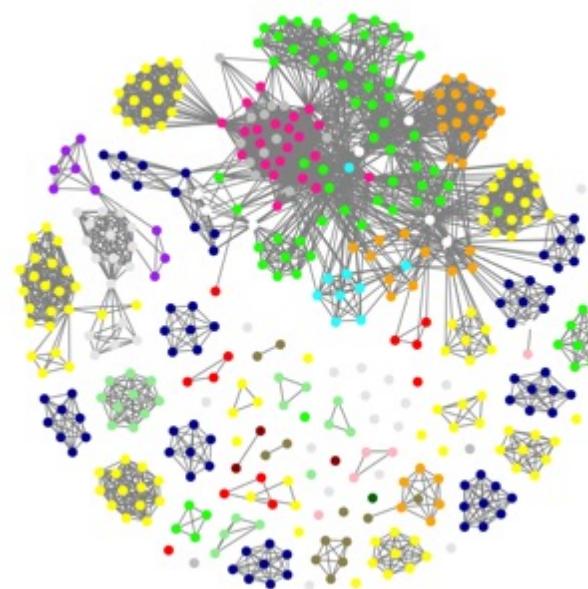


Share your opinions about social media and social network analysis:

<https://padlet.com/wmaharani/my-opinions-anrgiopu9lhv341j>

Definitions

Social Network Analysis

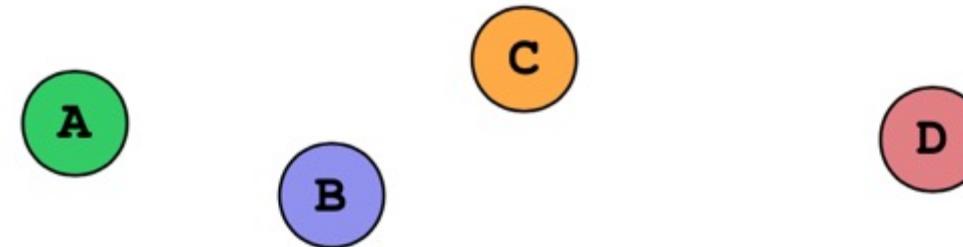


Definitions

- Actor/Node/Vertex
- Edge/Relation/Tie
- Path
- Network/Graph

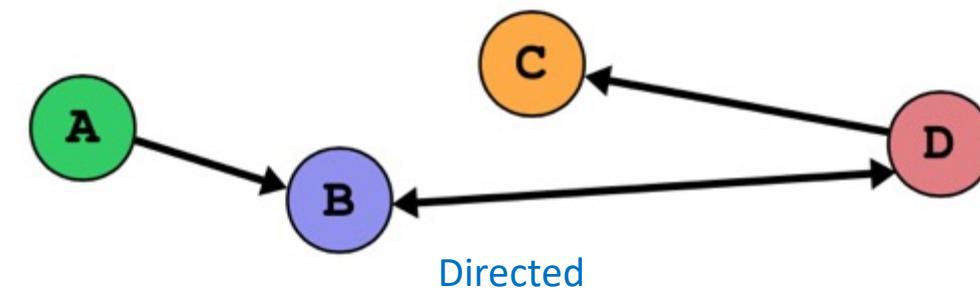
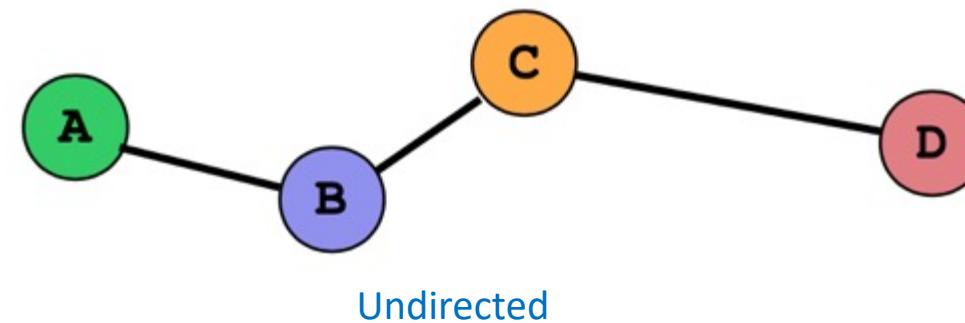
Actor/Node/Vertex

- An individual can have relationships with other individuals.



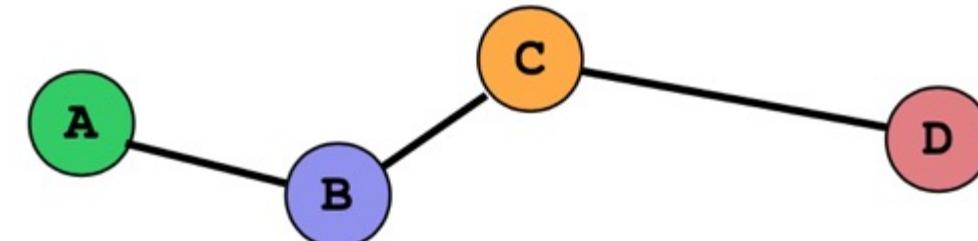
Edge/Relation/Tie

- Describes a particular, well specified, relationship between two **Actors/Nodes/Vertex**.



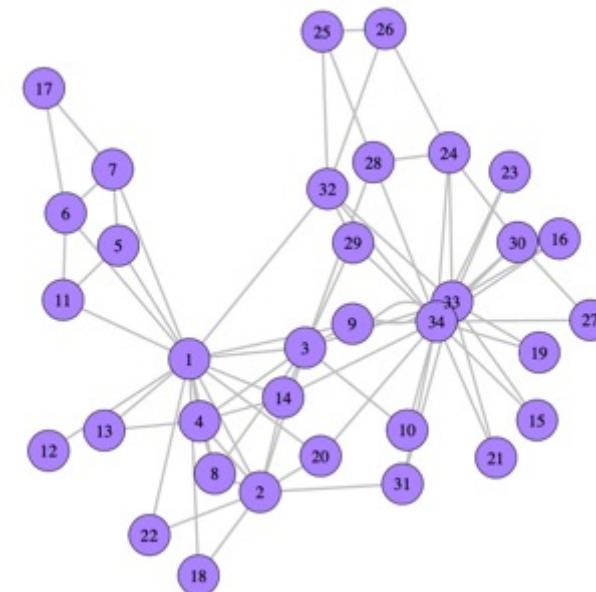
Graph

- Graph $G(V, E)$:
 - A set V of vertices or nodes
 - Connected by a set E of edges or links
 - Elements of E are unordered pairs (u, v) , where $u, v \in V$
- Vertices $V = \{A, B, C, D\}$
- Edges $E = \{(A,B), (B,C), (C,D)\}$



From Networks to Graphs

- A network is a set of *nodes* connected by a set of *edges*
- Graphs are mathematical representations of networks
 - Networks are also called graphs



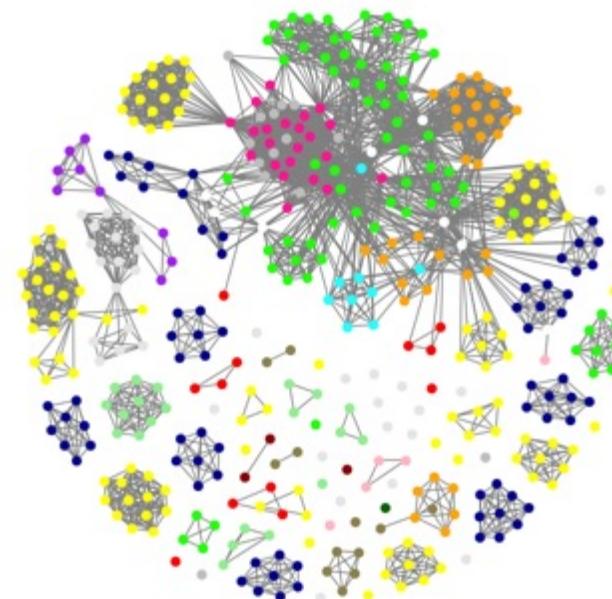
Vertices and edges in Networks

Network	Vertex	Edge
Internet	Computer/router	Cable or wireless link
Metabolic network	Metabolite	Metabolic reaction
WWW	Web page	Hyperlink
Food web	Species	Predation
Gene-regulatory network	Gene	Regulation of expression
Friendship network	Person	Friendship or acquaintance
Power grid	Substation	Transmission line
Affiliation network	Person and club	Membership
Protein interaction	Protein	Physical interaction
Citation network	Article/patent	Citation
Neural network	Neuron	Synapse
:	:	:

Go to page 1

Basic Concept

Social Network Analysis



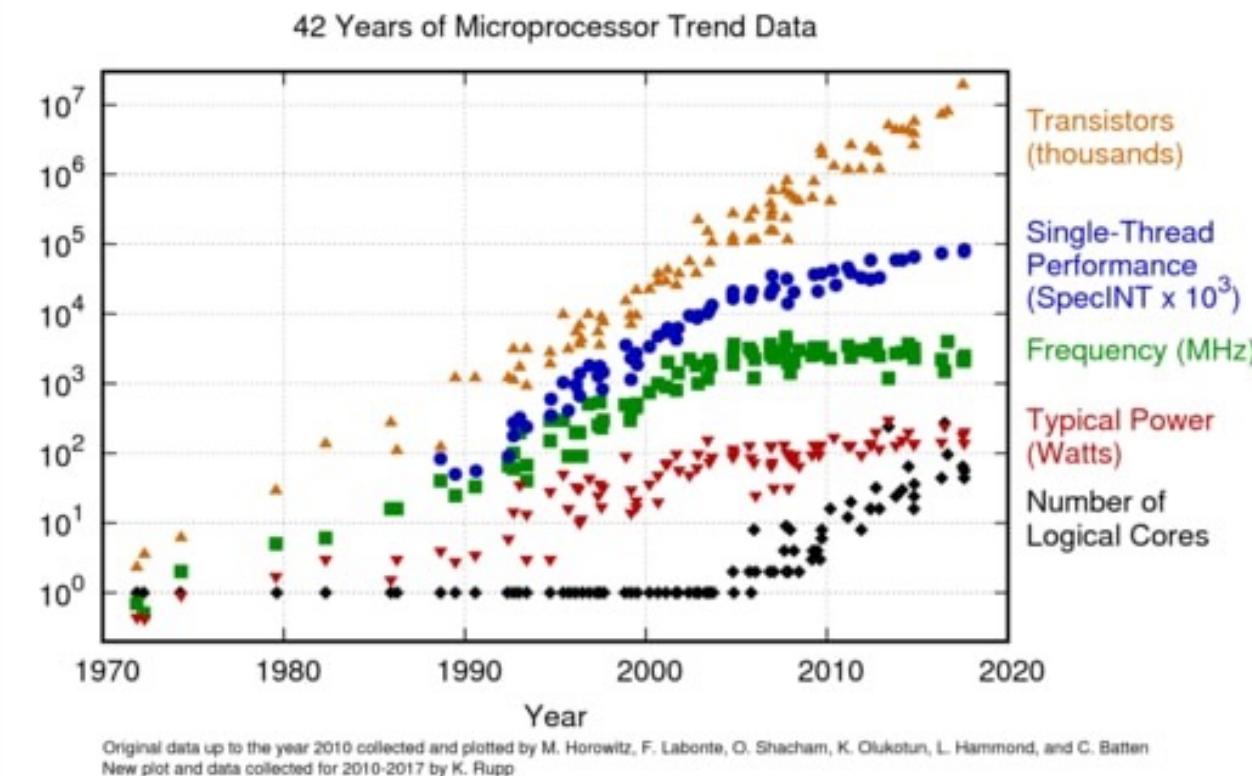
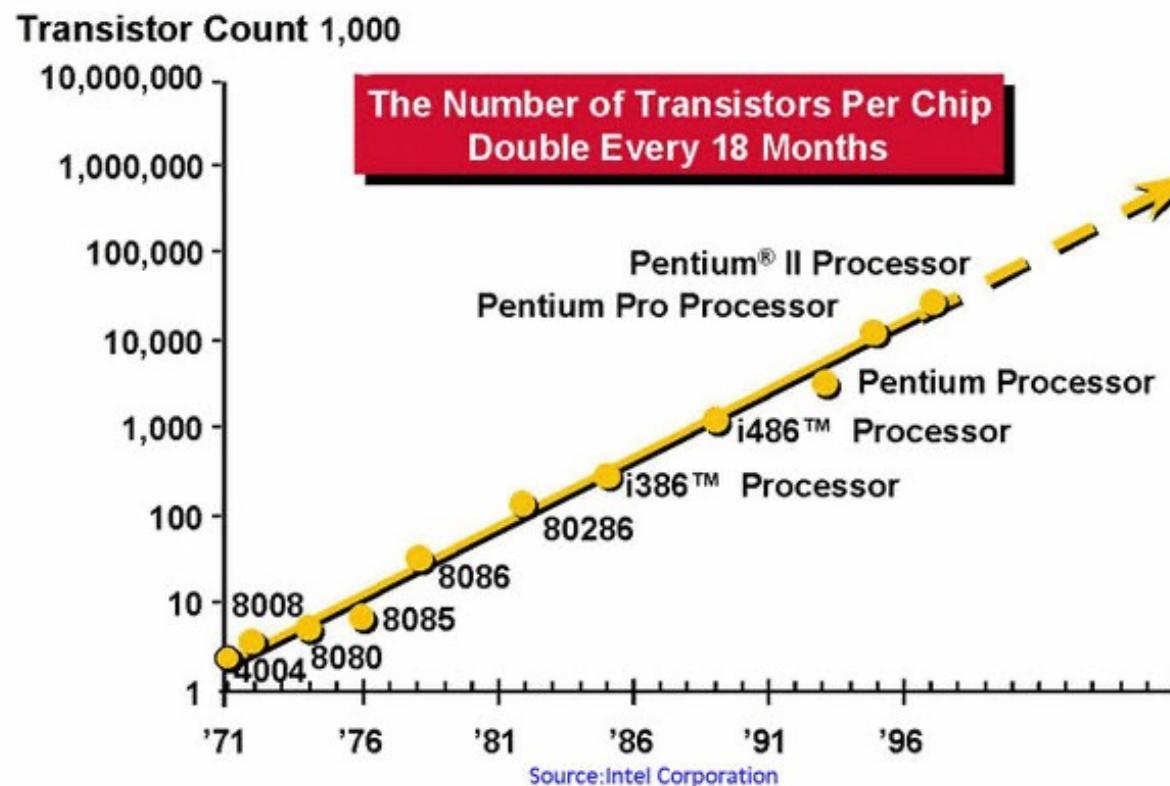
Data

- Data: facts, measurements or text collected for reference or analysis (Oxford dictionary)
 - Unstructured data: data that does not fit a certain data structure (text, images, audio, video, a list of numeric measurements)
 - Structured data: data that fits a certain data structure (table, graph/network, tree, etc.)

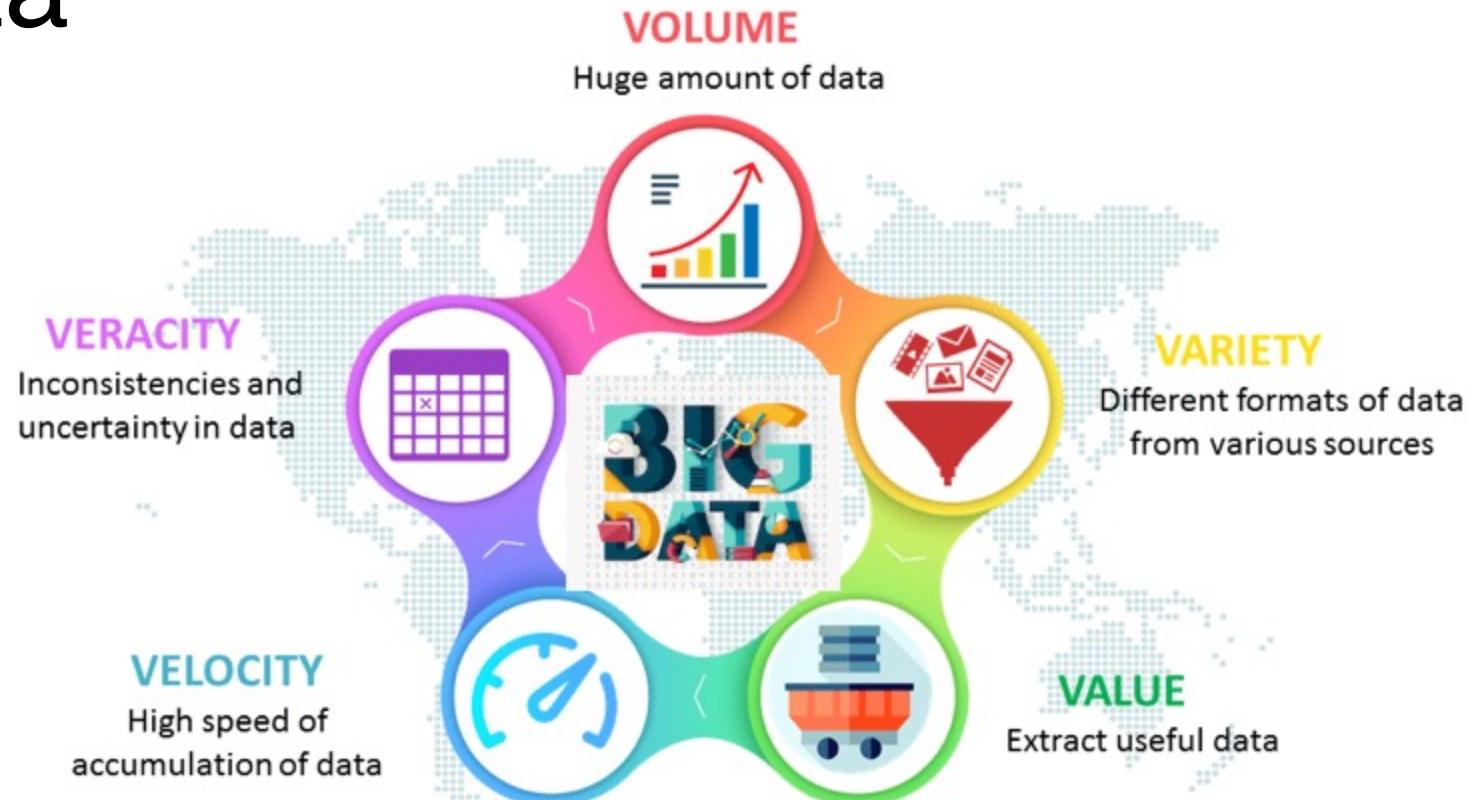
Data Evolution

- Sensus data (60's)
- Transaction data (80's)
- Micro event data (00's)
- Social data (2010)
- 2020 ??
- 2022 ??

Moore's law & Transistors

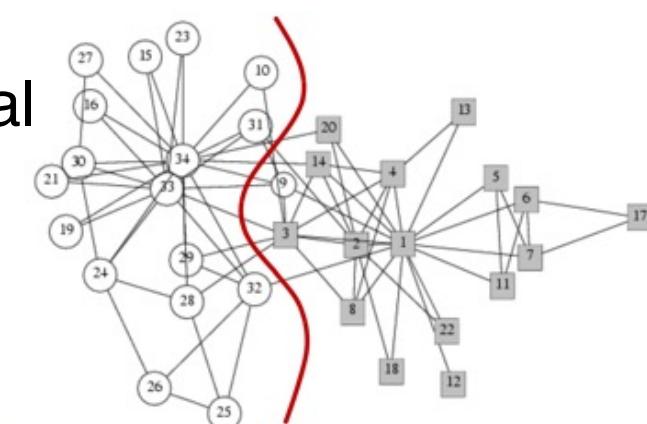


Big Data



Mining Social Network Data

- Mining social network has a long history in social science
 - Wayne Zachary's PhD work (1970) observe **social ties and rivalries in a university karate club**
 - During his observation, conflicts led the group to **split**
 - Split could be explained by a **minimum cut** in the social network



Social Media Mining

- Social media **platforms**: Facebook, Twitter, LinkedIn, Reddit, YouTube, Blogger, . . .
- Platforms generate enormous amounts of (un)structured data
- **Social media mining & analytics**: analyzing this data in order to get insight in user(s), trends, usage patterns, the platform itself, . . .
 - Text mining
 - Trend analysis
 - Sentiment mining
 - Topic modelling
 - **Social network analysis**

Networks: Rich Social Data

- Traditional obstacle:
 - Large-scale
 - Realistic
 - Completely mapped
- **Now:** large on-line systems leave detailed records of social activity
 - On-line communities: MySpace, Facebook, LiveJournal
 - Email, blogging, ecommerce, instant messaging
 - On-line publications repositories, arXiv, MedLine

Networks: A Matter of Scale

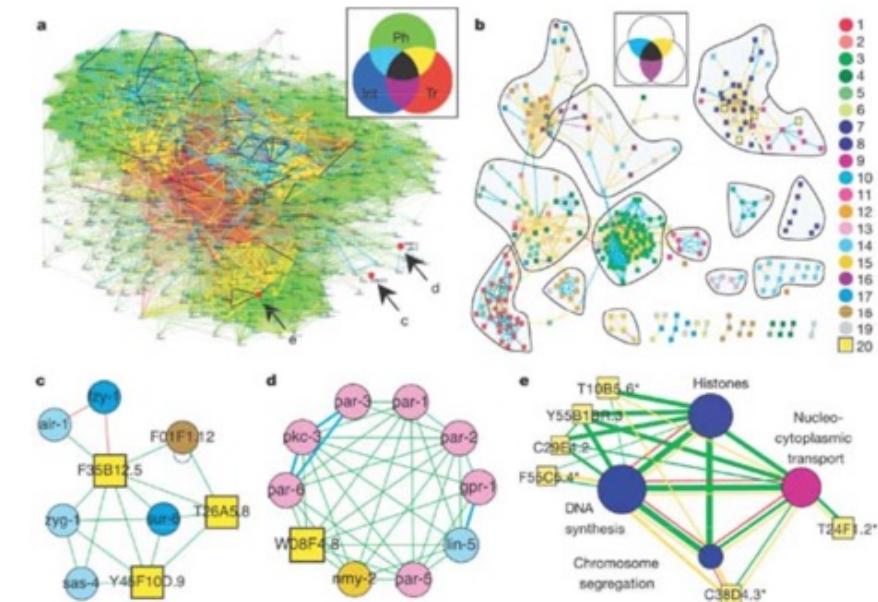
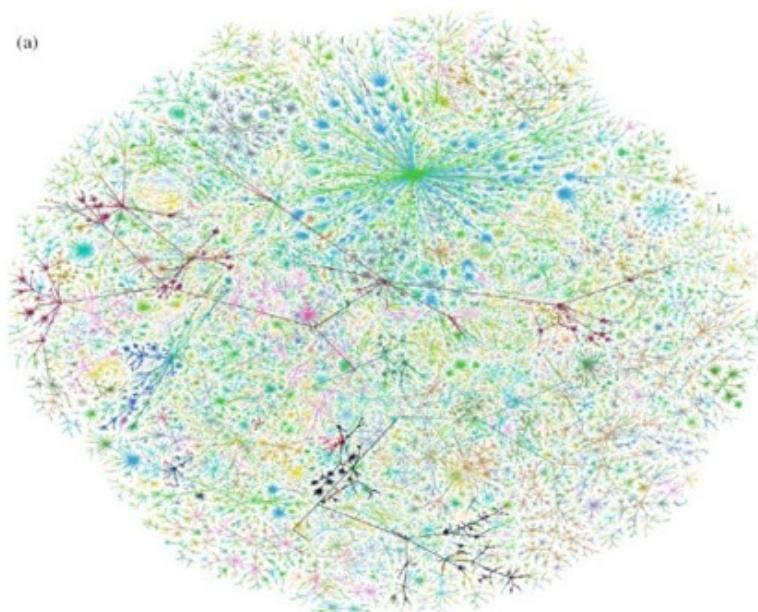
- Network data spans many orders of magnitude:
 - **436-node** network of email exchange over 3-months at corporate research lab [Adamic-Adar, SocNets '03]
 - **43,553-node** network of email exchange over 2 years at a large university [Kossinets-Watts, Science '06]
 - **4.4-million-node** network of declared friendships on a blogging community [Liben-Nowell et al., PNAS '05, Backstrom et al., KDD '06]
 - **240-million-node** network of all IM communication over a month on Microsoft Instant Messenger [Leskovec-Horvitz, WWW '08]

Networks: Structure & Process

- What have we learned about large networks?
- Structure: Many recurring patterns
 - Scale-free, small-world, locally clustered, bow-tie, hubs and authorities, communities, bipartite cores, network motifs, highly optimized tolerance
- Processes and dynamics:
 - Information propagation, cascades, epidemic thresholds, viral marketing, virus propagation, diffusion of innovation

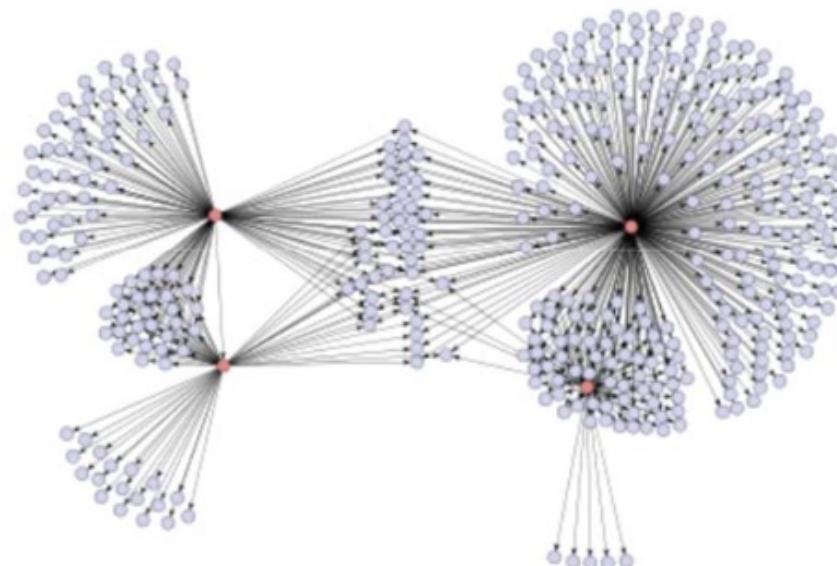
Structure of Networks

- What is the structure of a large network?
- Why and how did it became to have such structure?

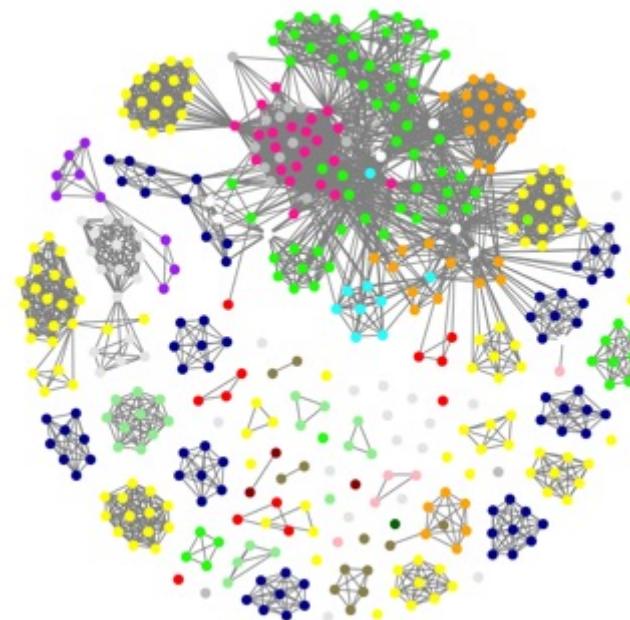


Diffusion in Networks

- One of the networks is a spread of a disease, the other one is product recommendations
- Which is which?



Applications and Challenges in Social Network Analysis

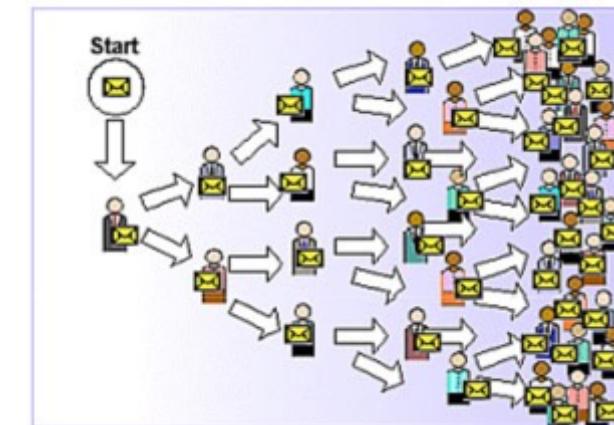


Applications [1]

- Web as a graph
 - Google PageRank
 - How to estimate webpage importance from the structure of the web-graph?
- Routing in peer-to-peer networks:
 - BitTorrent, ML-donkey, Kazaa, Gnutella
 - Can we find a file in a network without a central server?

Applications [2]

- Marketing and advertising:
 - How to define influence?
 - How to find influencers?
 - Who to give free products to
 - so that we create a network effect?
- Diffusion of information and epidemics:
 - How to trace information as it spreads?
 - How to efficiently detect epidemics and information outbreaks?



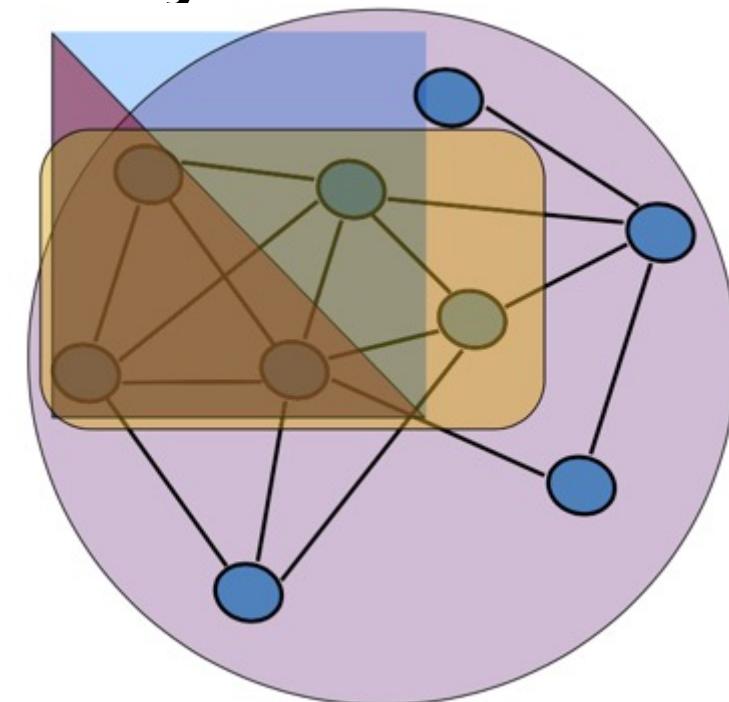
Applications [3]

- Friend/link prediction:
 - How to predict/suggest friends in networks?
- Trust and distrust:
 - How to predict who are your friends/foes? Who to trust?
- Community detection:
 - How to find clusters and small communities in social networks



Level of Social Network Analysis

- Nodal/Actor level
 - focuses on nodal level attributes and phenomena
- Dyadic level
 - focuses on the pairs of nodes
- Triadic level
 - focuses on triplets of nodes
- N-adic/Subset level
 - focuses on sub-graphs of N nodes
- Network/Group level
 - focuses on the whole graph and network level phenomena



Typically a cross-level analysis, combining all of these levels

Challenges

- Scale
 - This work considers “extreme-scale” graphs – billion+ vertices and up to trillion+ edges
 - Processing these graphs requires at least hundreds to thousands of compute nodes or tens of thousands of cores
 - Graph analytic algorithms are generally memory-bound instead of compute-bound; in the distributed space, this results in a ratio of communication versus computation that increases with core/node count

Complexity

- Real-world extreme-scale graphs have similar characteristics: small-world nature with skewed degree distributions
- Small-world graphs are difficult to partition for distributed computation or to optimize in terms of cache due to “too much locality”
- Skewed degree distributions make efficient parallelization and load balance difficult to achieve
- Multiple levels of cache/memory and increasing reliance on wide parallelism for modern HPC systems compounds the above challenges

Challenges

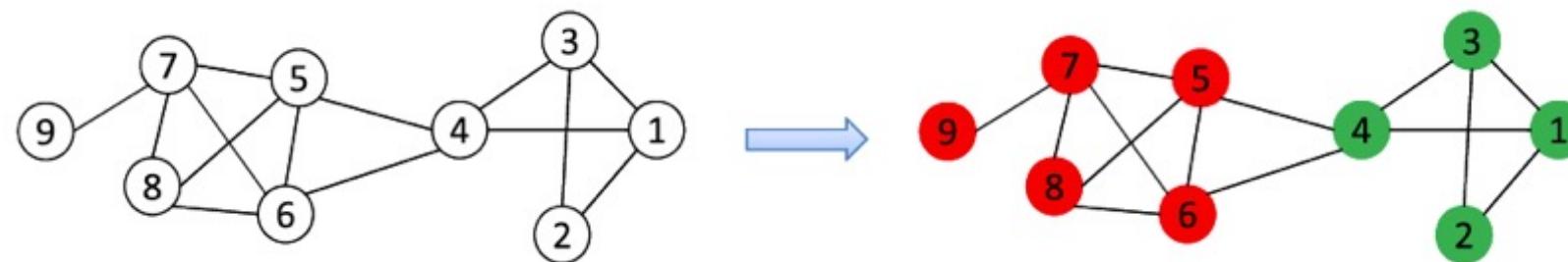
- **Heterogeneity**
 - Various types of entities and interactions are involved
- **Evolution**
 - Timeliness is emphasized in social media
- **Collective Intelligence**
 - How to utilize wisdom of crowds in forms of tags, wikis, reviews
- **Evaluation**
 - Lack of ground truth, and complete information due to privacy

Centrality Analysis and Influence Modeling

- Centrality Analysis:
 - Identify the **most important** actors or edges
 - E.g. PageRank in Google
 - Various other criteria
- Influence modeling:
 - How is information diffused?
 - How does one influence each other?
- Related Problems
 - Viral marketing: word-of-mouth effect
 - Influence maximization

Community Detection

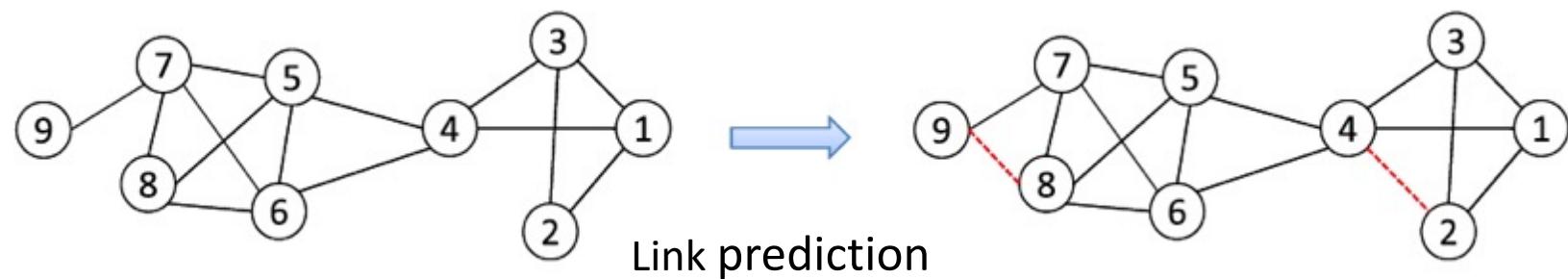
- A **community** is a set of nodes between which the interactions are (relatively) frequent
 - A.k.a., *group, cluster, cohesive subgroups, modules*



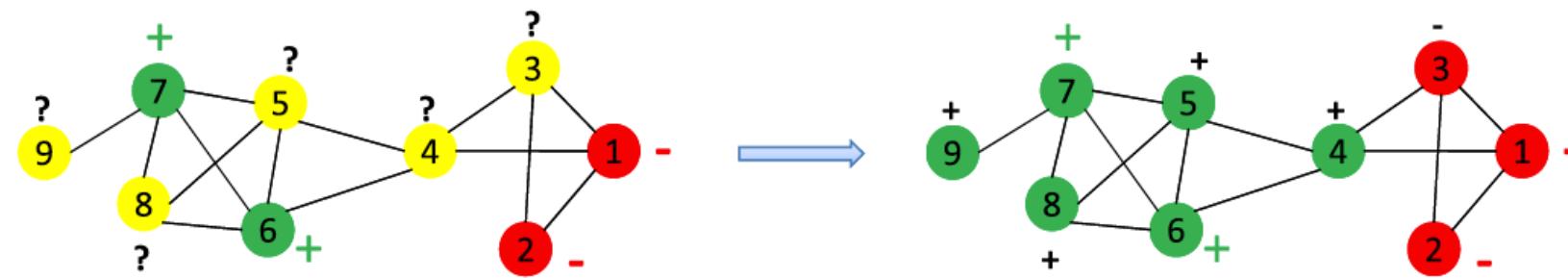
- Applications: *Recommendation based communities, Network Compression, Visualization of a huge network*
- New lines of research in social media
 - Community Detection in Heterogeneous Networks
 - Community Evolution in Dynamic Networks
 - Scalable Community Detection in Large-Scale Networks

Classification and Recommendation

- Common in social media applications
 - Tag suggestion, Product/Friend/Group Recommendation



Link prediction



Network-Based Classification

Privacy, Spam and Security

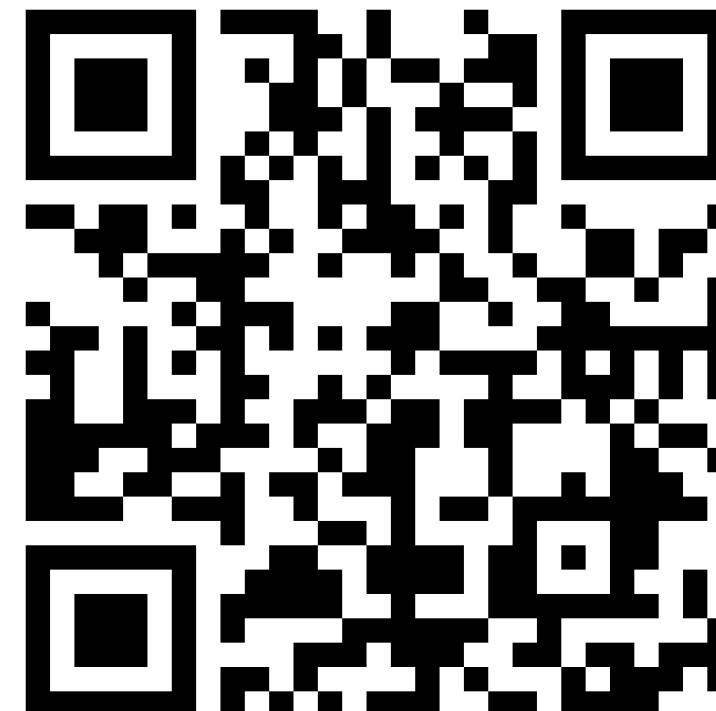
- Privacy is a big concern in social media
 - Facebook, Google buzz often appear in debates about privacy
 - NetFlix Prize Sequel cancelled due to privacy concern
 - Simple annoymization does not necessarily protect privacy
- Spam blog (splog), spam comments, Fake identity, etc., all requires new techniques
- As private information is involved, a secure and trustable system is critical
- Need to achieve a balance between sharing and privacy

Conclusions

- Network:
 - Scale
 - Structure
 - Information Diffusion
- Challenges
 - Heterogeneity
 - Evolution
 - Evaluation

Define your target & efforts

- <https://padlet.com/wmaharani/my-target-efforts-5npss5d1pjdljr9a>



References

- Community Detection and Mining in Social Media. Lei Tang and Huan Liu, Morgan & Claypool, September, 2010.
- Newman, ME, The Structure and Function of Complex Networks, SIAM 2003
- Watts, DJ; Strogatz, S H. 1998.
Collective dynamics of 'small-world' networks, *NATURE* 393(668).
- Barabasi, AL. Network Science Network Science

CSI4H3

Analisis Media Sosial

Thank You

