Social Network Analysis

Introduction & Preliminaries

Introduction

- Social Network Analysis (SNA) is about analyzing networks arising in various contexts, especially those arising in social contexts: as a result of people connecting with each other on online social networks such as twitter and facebook, as well as "who-calls-whom" graph arising out of Telecom networks.

- However, the techniques for analysing social networks can be extended to other non-social networks as well.

Wikipedia Definitions

- A **social network** is a social structure made up of a set of social actors (such as individuals or organizations), sets of dyadic ties, and other social interactions between actors.

- **Social network analysis** (**SNA**) is the process of investigating social structures through the use of networks and graph theory.

- **Network science** is an academic field which studies complex networks such as telecommunication networks, computer networks, biological networks, cognitive and semantic networks, and social networks, considering distinct elements or actors represented by *nodes* (or *vertices*) and the connections between the elements or actors as *links* (or *edges*). The field draws on theories and methods including graph theory from mathematics, statistical mechanics from physics, data mining and information visualization from computer science, inferential modeling from statistics, and social structure from sociology.

Use of Network Science

- Search of the World-Wide Web
- How to understand the spread of diseases?
- How to understand the formation of links and communities?
- Recommendation and Viral Marketing

Course Outline (Tentative)

- Course introduction and structure of Graphs
- Measuring networks and random graph models
- Network analysis and visualization tools
- Link analysis
- Community detection in graphs
- Link prediction
- Network effects and Cascade behavior
- Influence maximization and Outbreak detection
- Graph Representation Learning
- Graph Convolutional Networks
- Other applications of networks

Grading (Tentative)

- 1 quiz – 10%
- 1 assignment – 10%
- Minor –15%
- Major – 25%
- Course project – 40%

Readings

- Tanmoy Chakraborty. Social Network Analysis, Wiley India Pvt. Ltd., 2021, ISBN 978-81-265-2007-7.
  - Amazon link: https://www.amazon.in/Social-Network-Analysis-Tanmoy-Chakraborty/dp/9354247830/ref=sr_1_2?dchild=1&keywords=9789354247835&qid=1634206218&sr=8-2)

- "Networks, Crowds, and Markets: Reasoning About a Highly Connected World" by David Easley and Jon Kleinberg Cambridge University Press 2010.
  - Online: https://www.cs.cornell.edu/home/kleinber/networks-book/

- Research Papers discussed in class.

Instructors

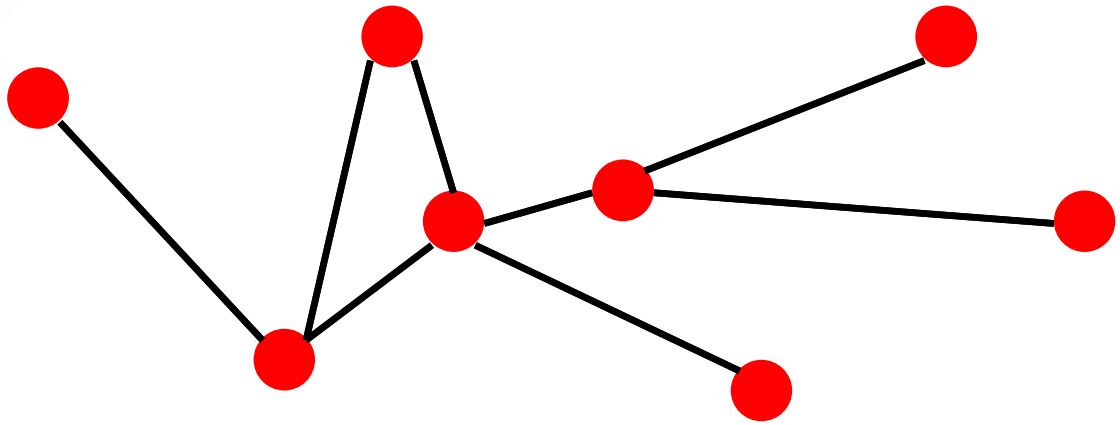
- Sougata Mukherjea
  - Email: sougatam@iitd.ac.in

- Tanmoy Chakraborty
  - Email: tanchak@iitd.ac.in

Preliminaries

• • •

Graph (Network)



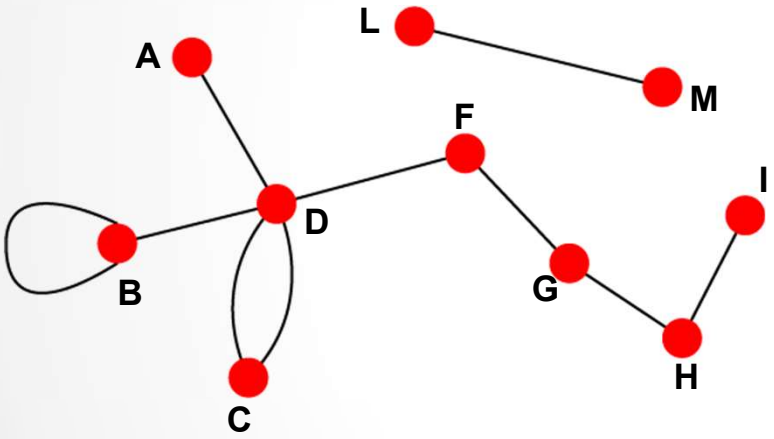
- **components**: nodes, vertices N
- **interactions**: links, edges L
- **system**: network, graph (N,L)

Directed & Undirected Graphs

Undirected

Links: undirected (*symmetrical*)

Graph:

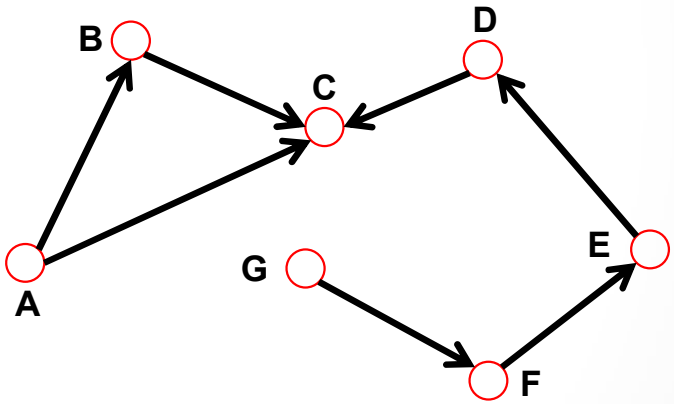


Undirected links :
coauthorship links
Actor network
protein interactions

Directed

Links: directed (*arcs*).

Digraph = directed graph:

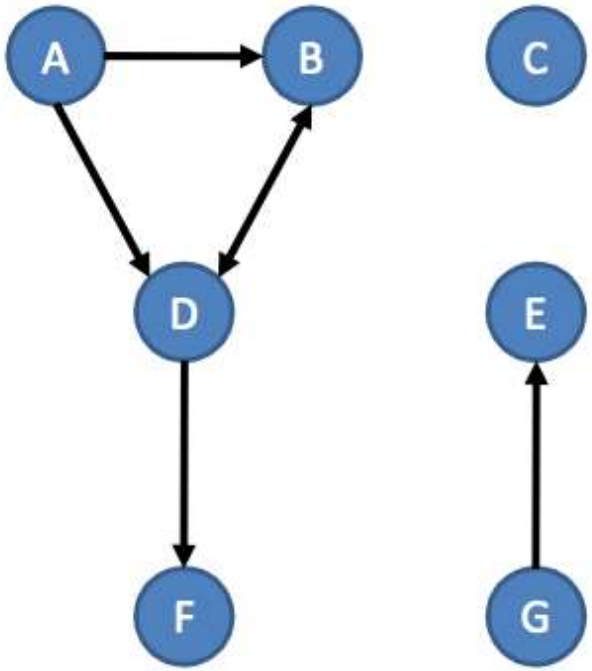


An undirected link is the superposition of two opposite directed links.

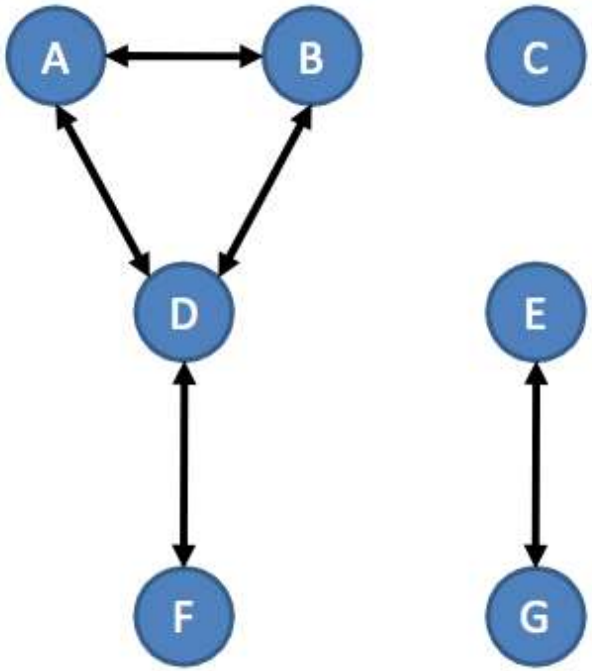
Directed & undirected

- Communication vs. friendship networks

twitter



facebook

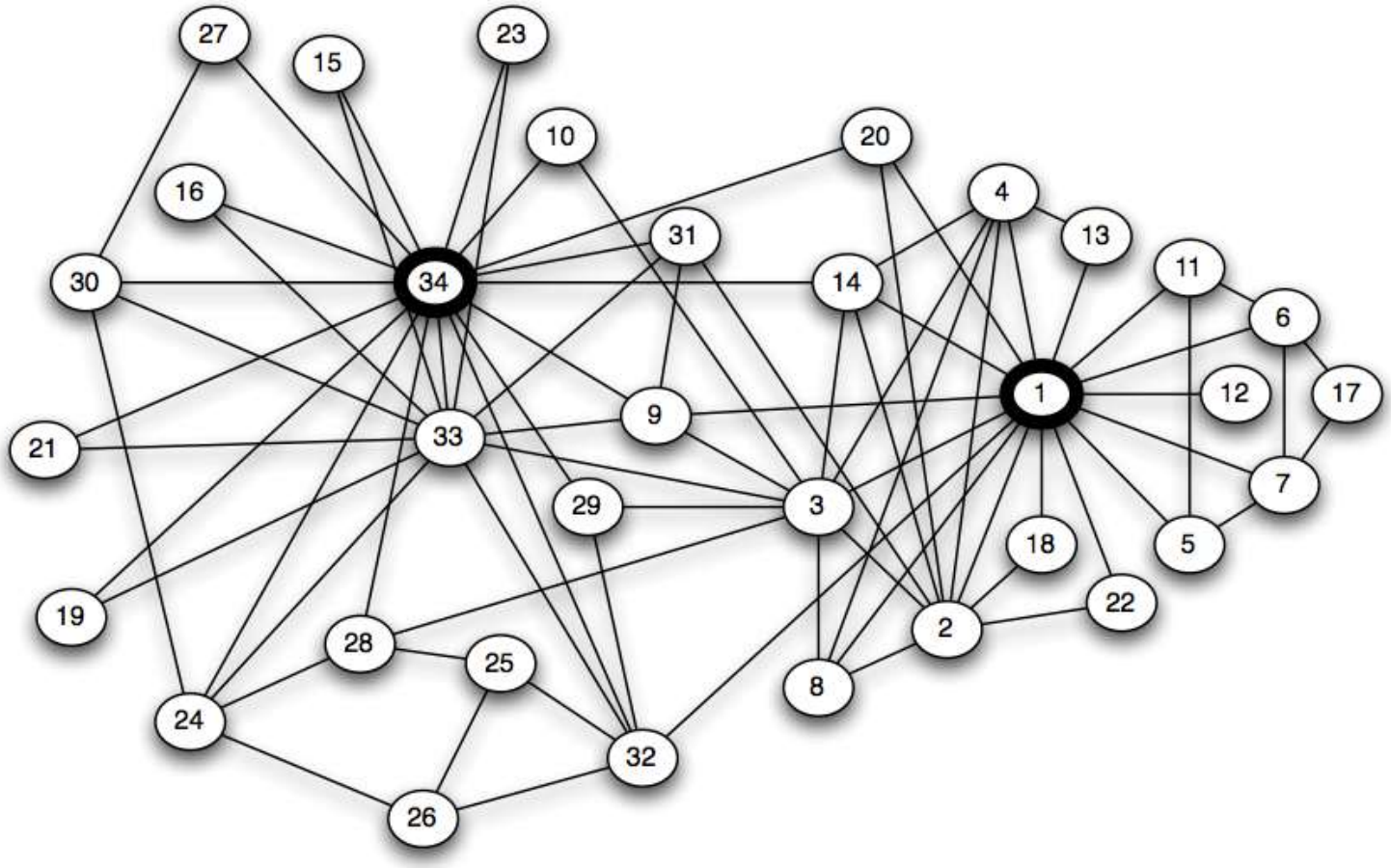


Example Networks

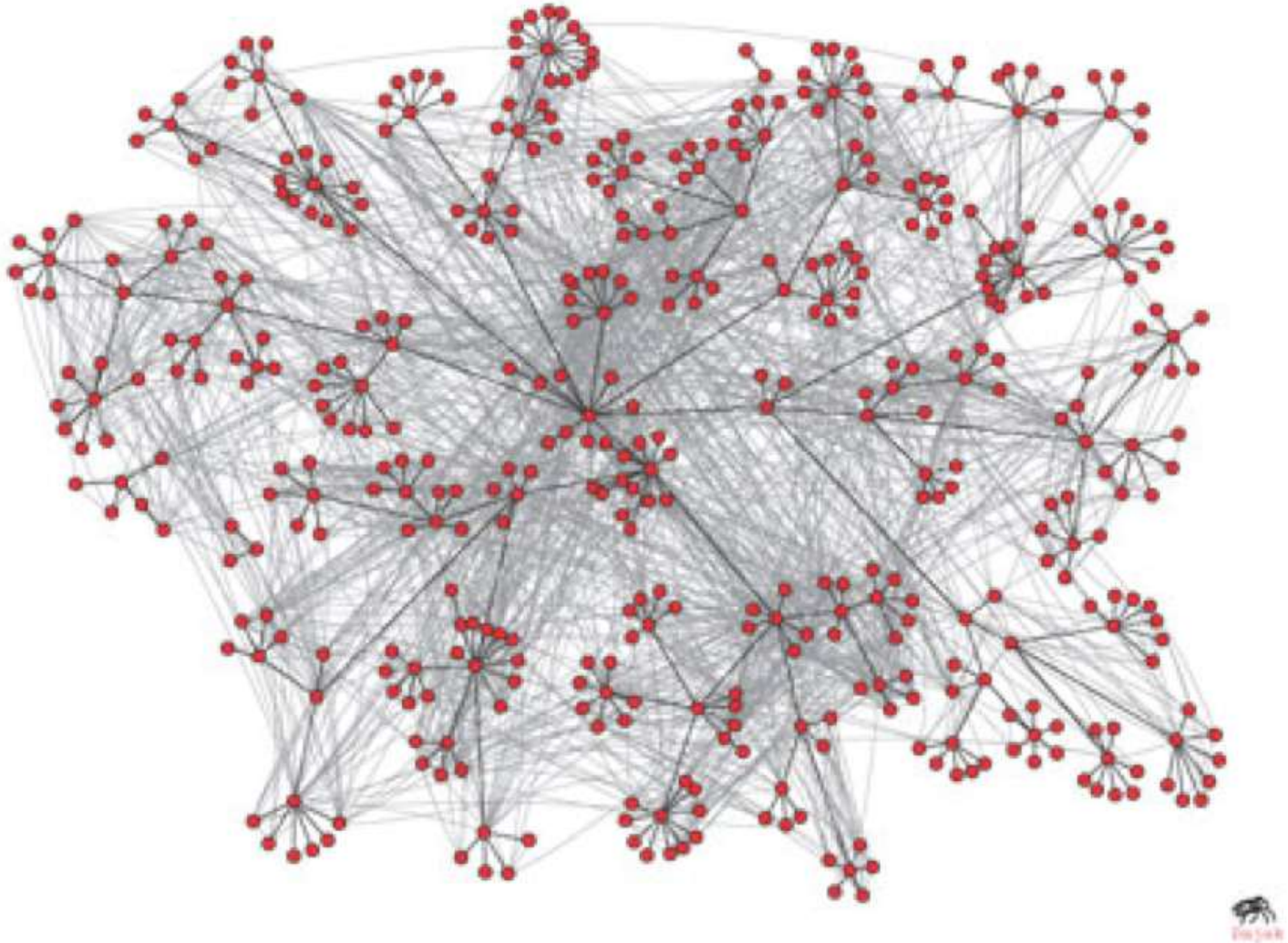
NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L
Internet	Routers	Internet connections	Undirected	192,244	609,066
WWW	Webpages	Links	Directed	325,729	1,497,134
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826
Email	Email addresses	Emails	Directed	57,194	103,731
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908
Citation Network	Paper	Citations	Directed	449,673	4,689,479
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930

Actors & relations

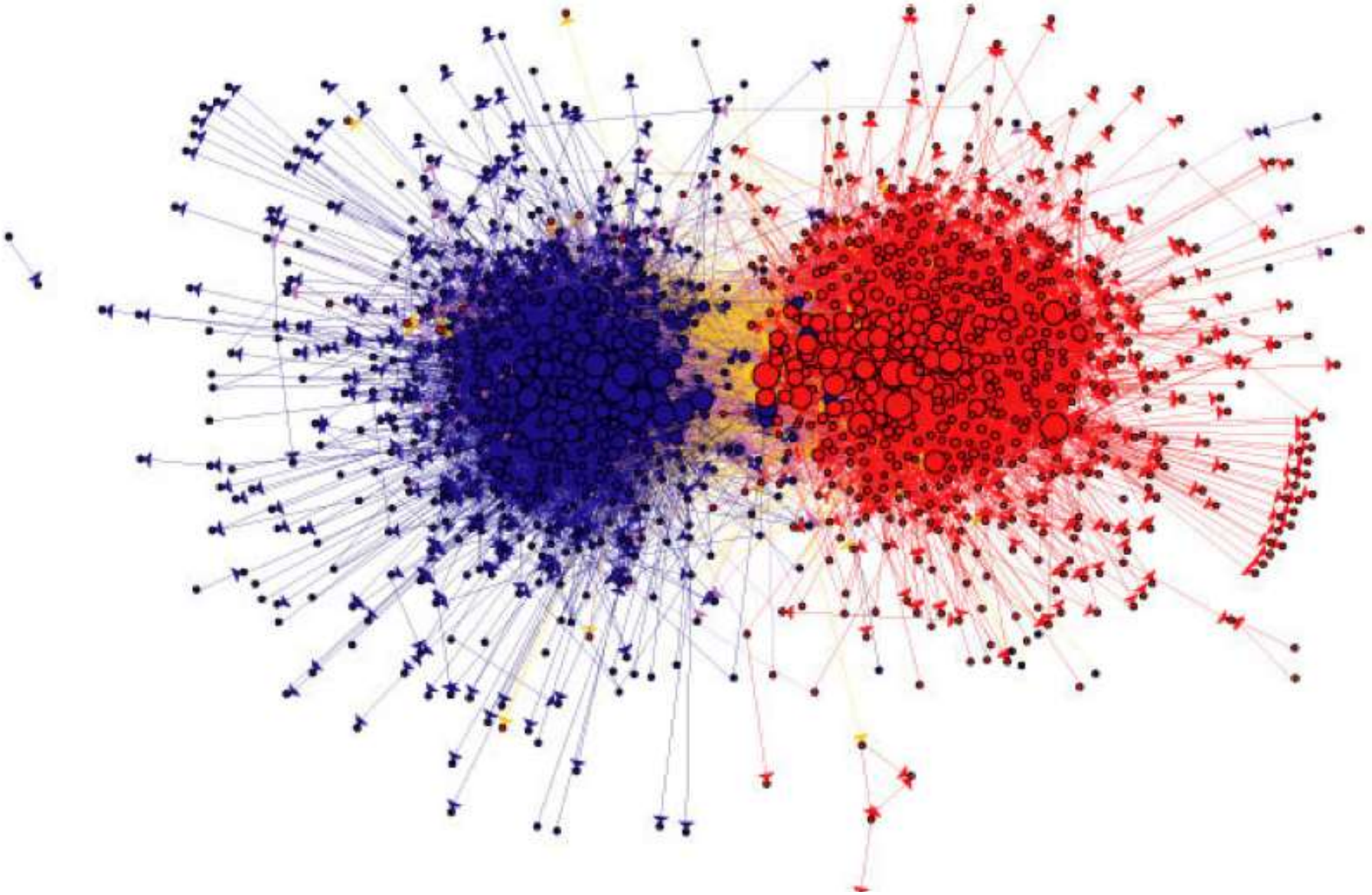
- Actors (nodes, vertices)
  - People, groups, organizations, communities, nation-states
- Relations (links, ties, edges)
  - Evaluations of one person by another (friendship, liking, ...)
  - Transfers of material resources (lending, donations, ...)
  - Association or affiliation (membership, attendance, ...)
  - Behavioral interaction (communication, intercourse, ...)
  - Movement between places or statuses (migration, mobility, ...)
  - Physical connections (roads, routers, ...)
  - Formal relations (authority, supply chain, ...)
  - Biological relations (kinship, descent, ...)



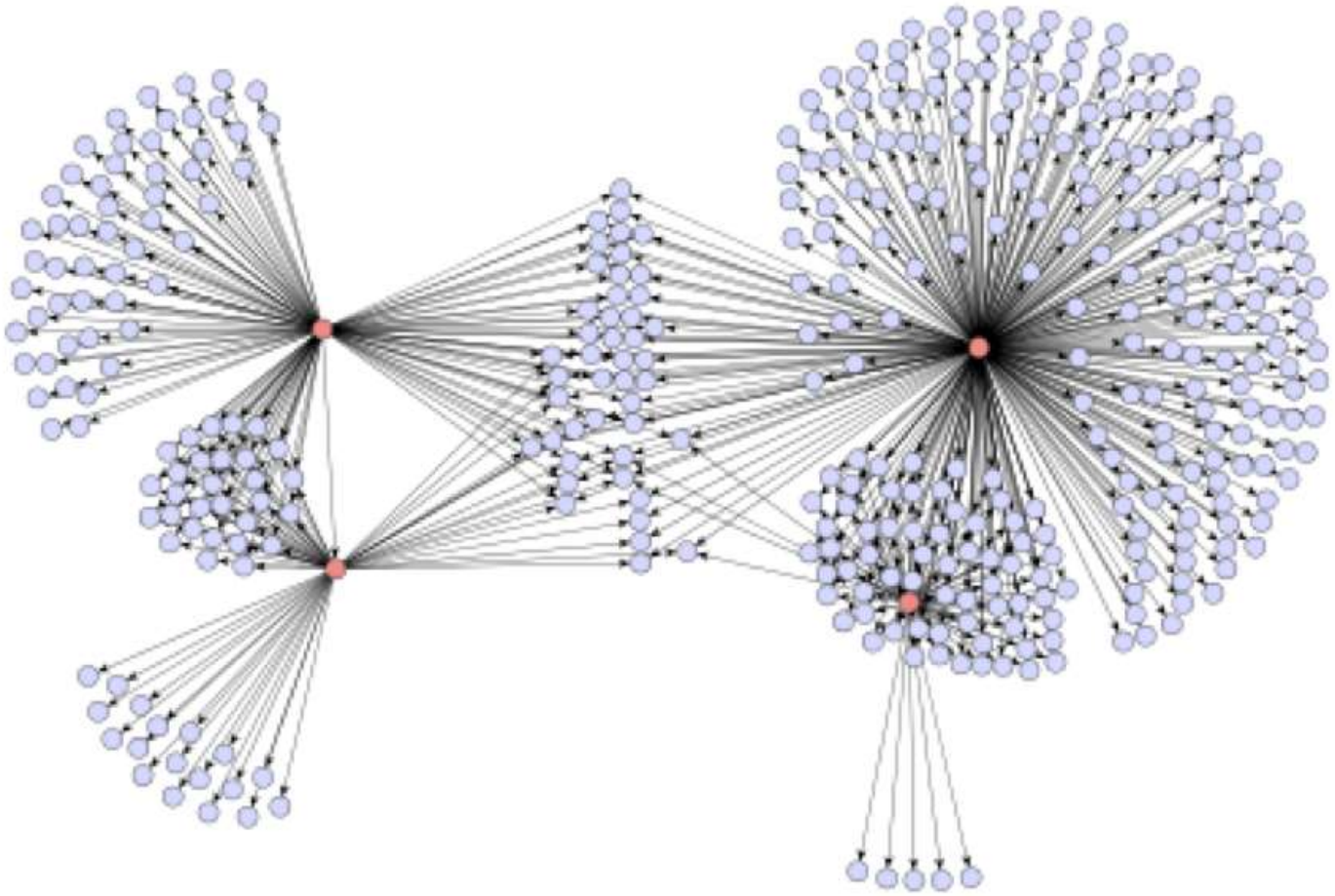
The social network of friendships within a 34-person karate club [1]



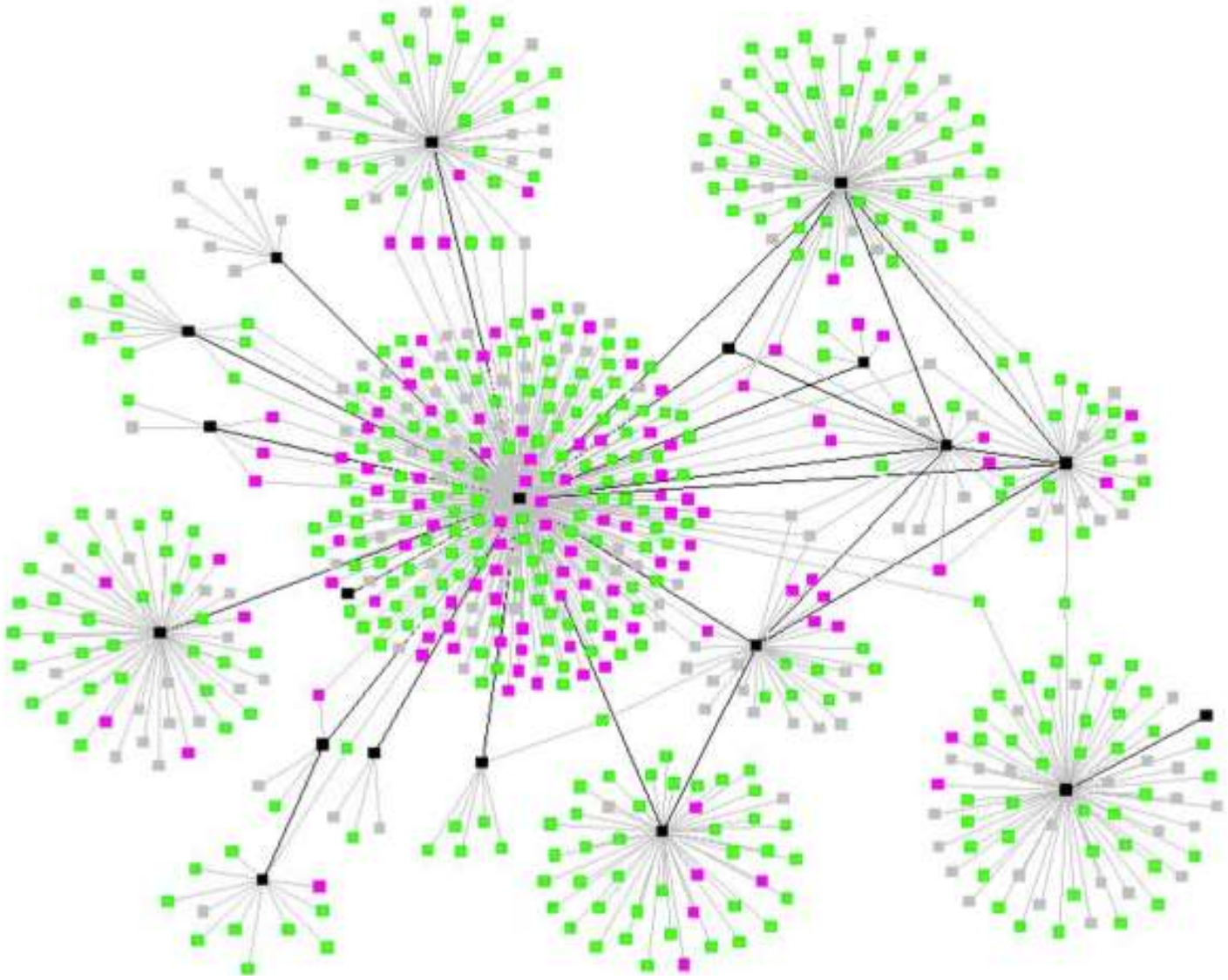
The pattern of e- mail communication among 436 employees of Hewlett Packard Research Lab is superimposed on the official organizational hierarchy [2].



The network structure of political blogs prior to the 2004 U.S. Presidential election reveals two natural and well-separated clusters [3].

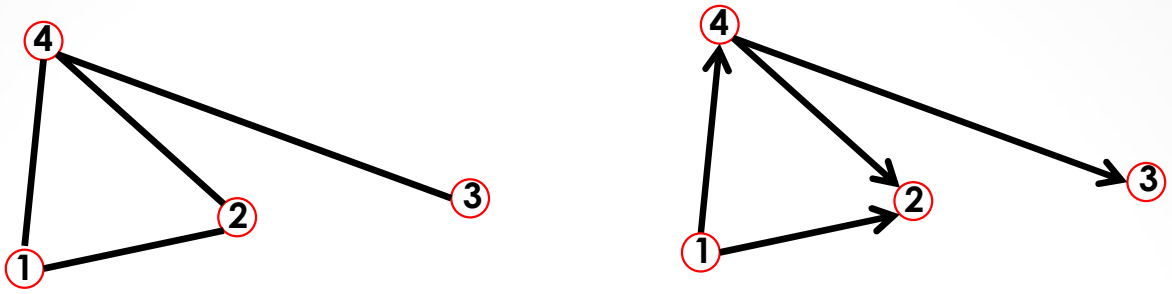


When people are influenced by the behaviors their neighbors in the network, the adoption of a new product or innovation can cascade through the network structure. Here, e-mail recommendations for a Japanese graphic novel spread in a kind of informational or social contagion. [4].



The spread of an epidemic disease (such as the tuberculosis outbreak shown here) [5].

Adjacency Matrix



A_{ij}=1 if there is a link between node *i* and *j*

A_{ij}=0 if nodes *i* and *j* are not connected to each other.

$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

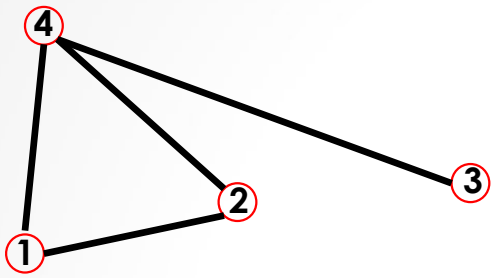
Note that for a directed graph (right) the matrix is not symmetric.

$A_{ij} = 1$ if there is a link pointing from node *j* and *i*

$A_{ij} = 0$ if there is no link pointing from *j* to *i*.

Adjacency Matrix & Node Degrees

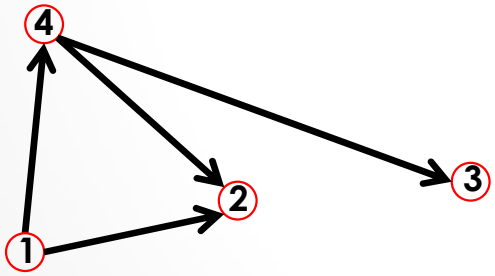
Undirected



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$A_{ij} = A_{ji}$$
$$A_{ii} = 0$$

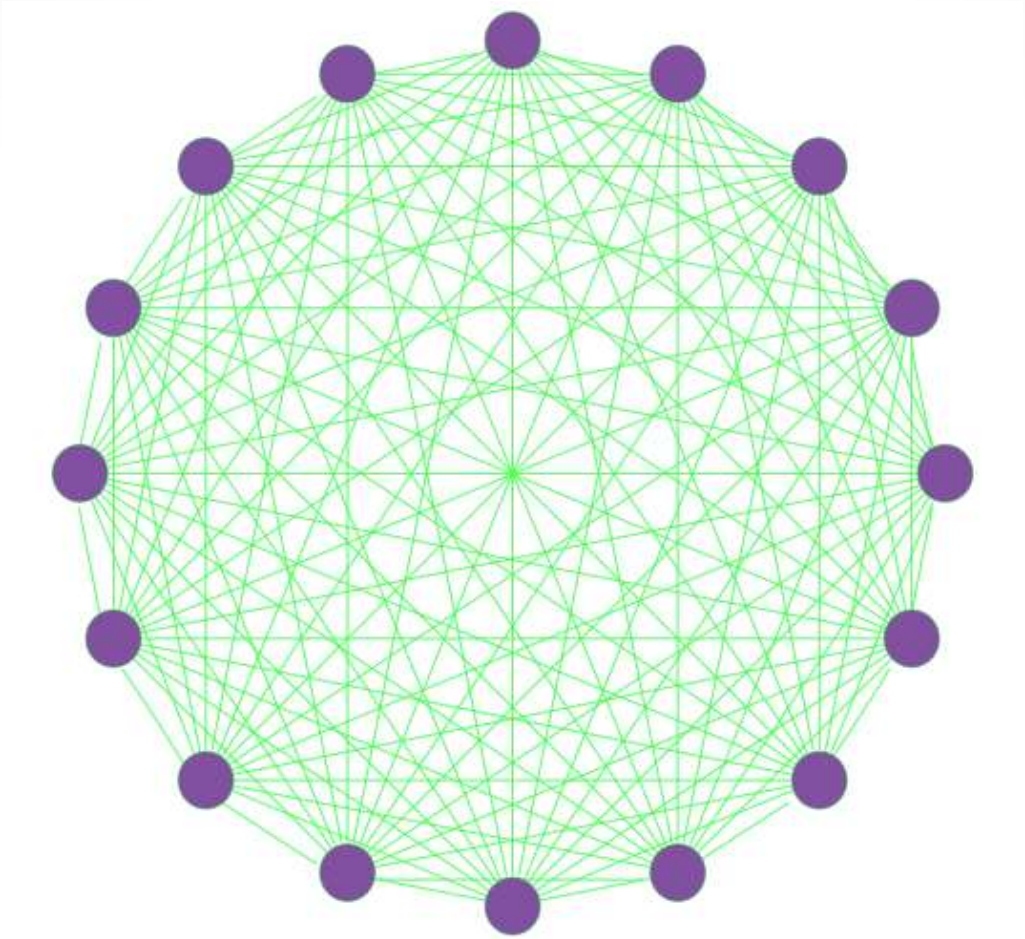
Directed



$$A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{ij} \neq A_{ji}$$
$$A_{ii} = 0$$

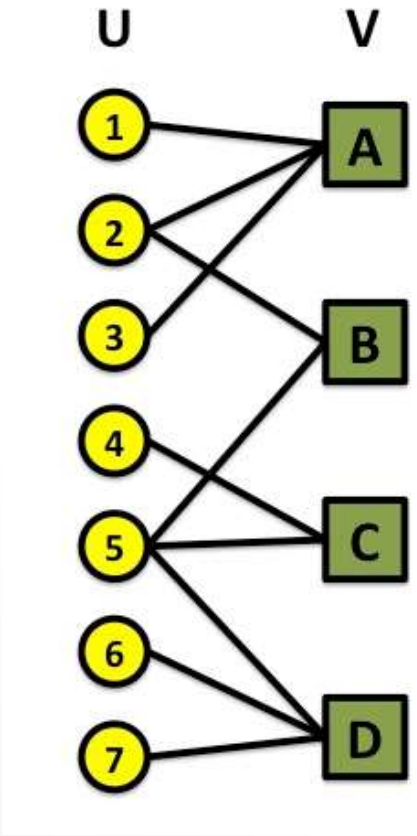
Complete Graph



has all pairs of vertices connected with each other:
$|E| = N(N-1)/2$

Bipartite Graph

Bipartite graph (or **Bigraph**) is a graph whose nodes can be divided into two disjoint sets *U* and *V* such that every link connects a node in *U* to one in *V*; that is, *U* and *V* are independent sets.

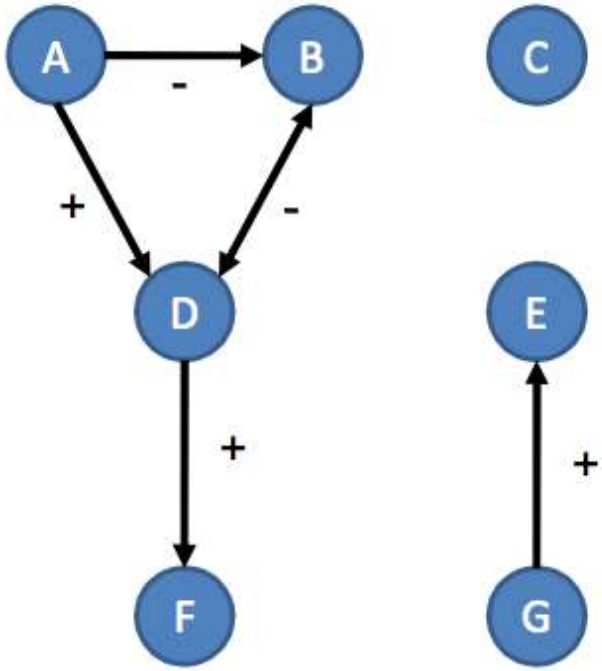


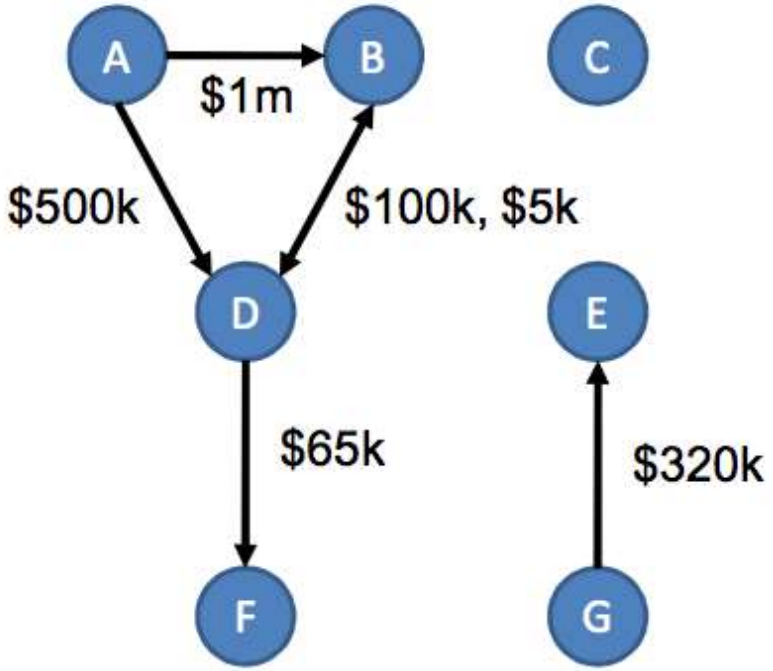
Examples:

U – People, V – Hobbies
U – Recipies, V – Ingredients
U – Documents, V – Keywords

Signed & valued

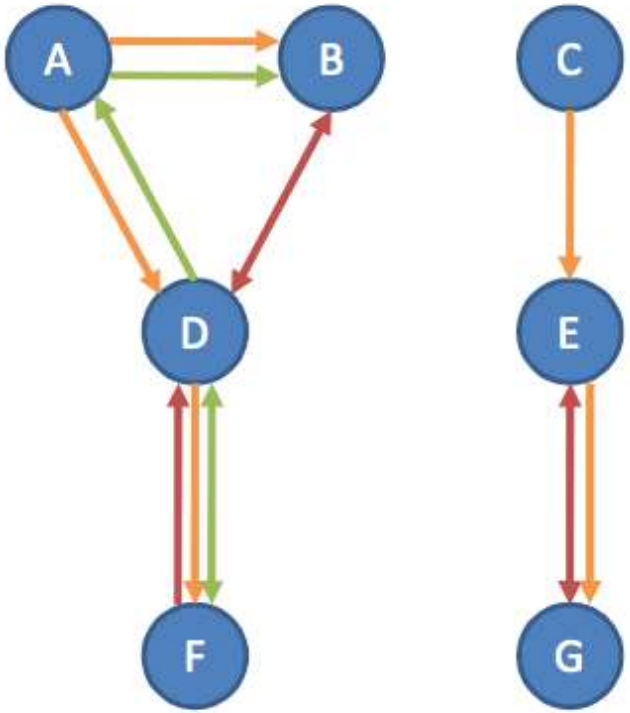
- Affect in a sorority vs. campaign financing





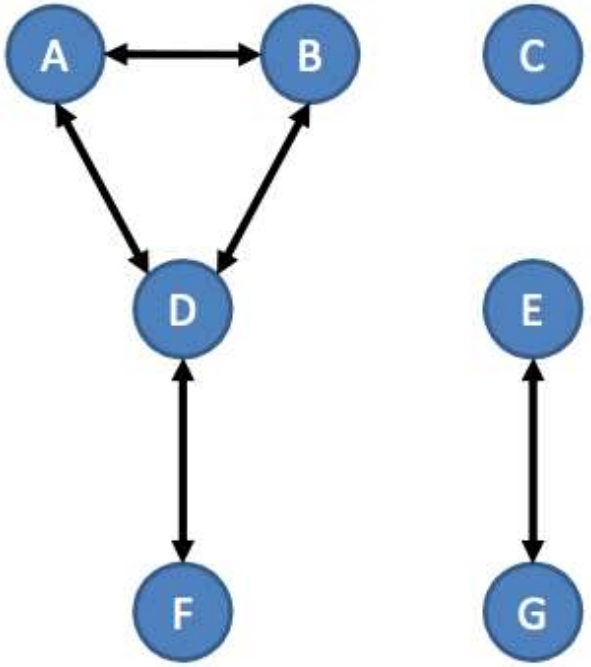
Multi-relational

- Organizations: **authority**, **trust**, & **friendship**



Egos & alters

If **A** is "*ego*"

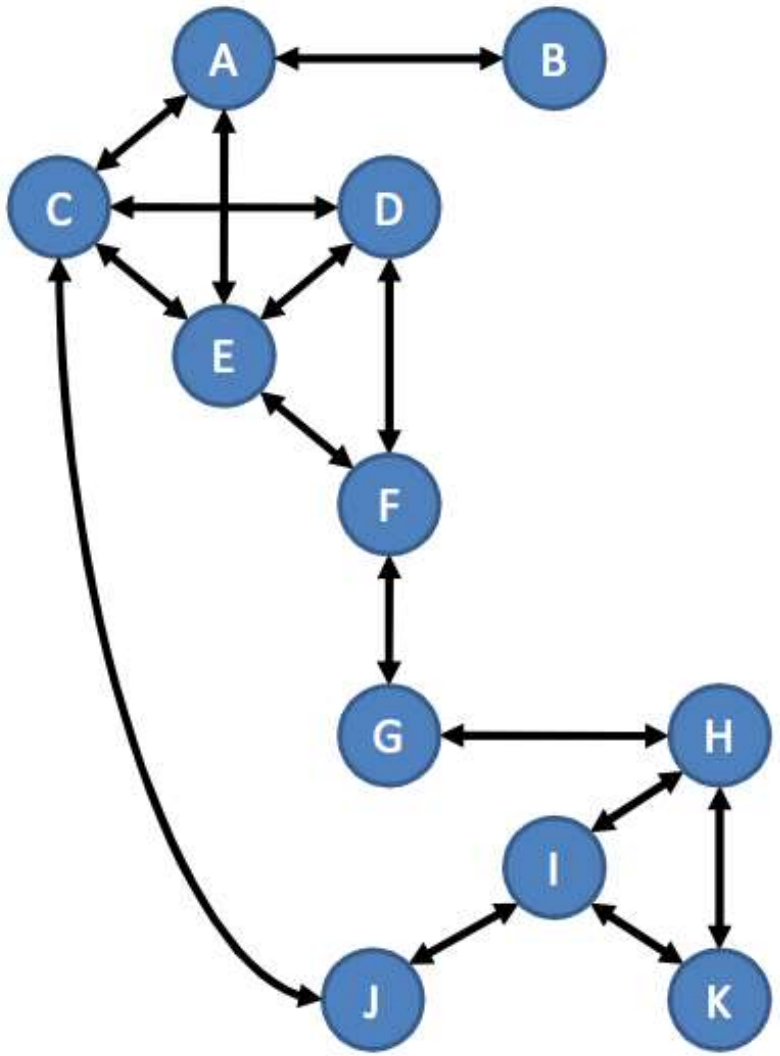


B and **D** are his "*alters*"

- Ego-centric networks (or shortened to "ego" networks) are a particular type of network which specifically maps the connections of and from the perspective of a single person (an "ego").

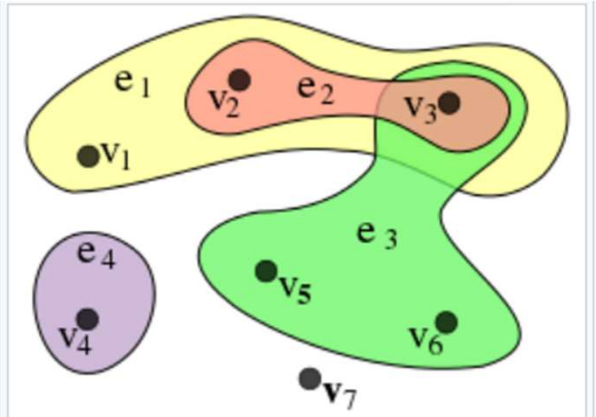
Ego network

- **N-step ego network**: network of all actors and their shared ties, N steps away from ego
  - E's 1-step ego network
  - E's 2-step ego network
  - E's 3-step ego network
  - E's 4-step ego network



Hypergraph

A **hypergraph** is a generalization of a graph in which an edge can join any number of vertices. In contrast, in an ordinary graph, an edge connects exactly two vertices.



An example of an undirected hypergraph, with $X = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ and $E = \{e_1, e_2, e_3, e_4\} = \{\{v_1, v_2, v_3\}, \{v_2, v_3\}, \{v_3, v_5, v_6\}, \{v_4\}\}$. This hypergraph has order 7 and size 4. Here, edges do not just connect two vertices but several, and are represented by colors.

References

1. Wayne Zachary. An information flow model for conflict and fission in small groups. Journal of Anthropological Research, 33(4):452–473, 1977.
2. Lada A. Adamic and Eytan Adar. How to search a social network. Social Networks, 27(3):187–203, 2005.
3. Lada Adamic and Natalie Glance. The political blogosphere and the 2004 U.S. election: Divided they blog. In Proceedings of the 3rd International Workshop on Link Discovery, pages 36–43, 2005.
4. Jure Leskovec, Lada Adamic, and Bernardo Huberman. The dynamics of viral marketing. ACM Transactions on the Web, 1(1), May 2007.
5. McKenzie Andre, Kashef Ijaz, Jon D. Tillinghast, Valdis E. Krebs, Lois A. Diem, Bev- erly Metchock, Theresa Crisp, and Peter D. McElroy. Transmission network analysis to complement routine tuberculosis contact investigations. American Journal of Public Health, 97(3):470–477, 2007.