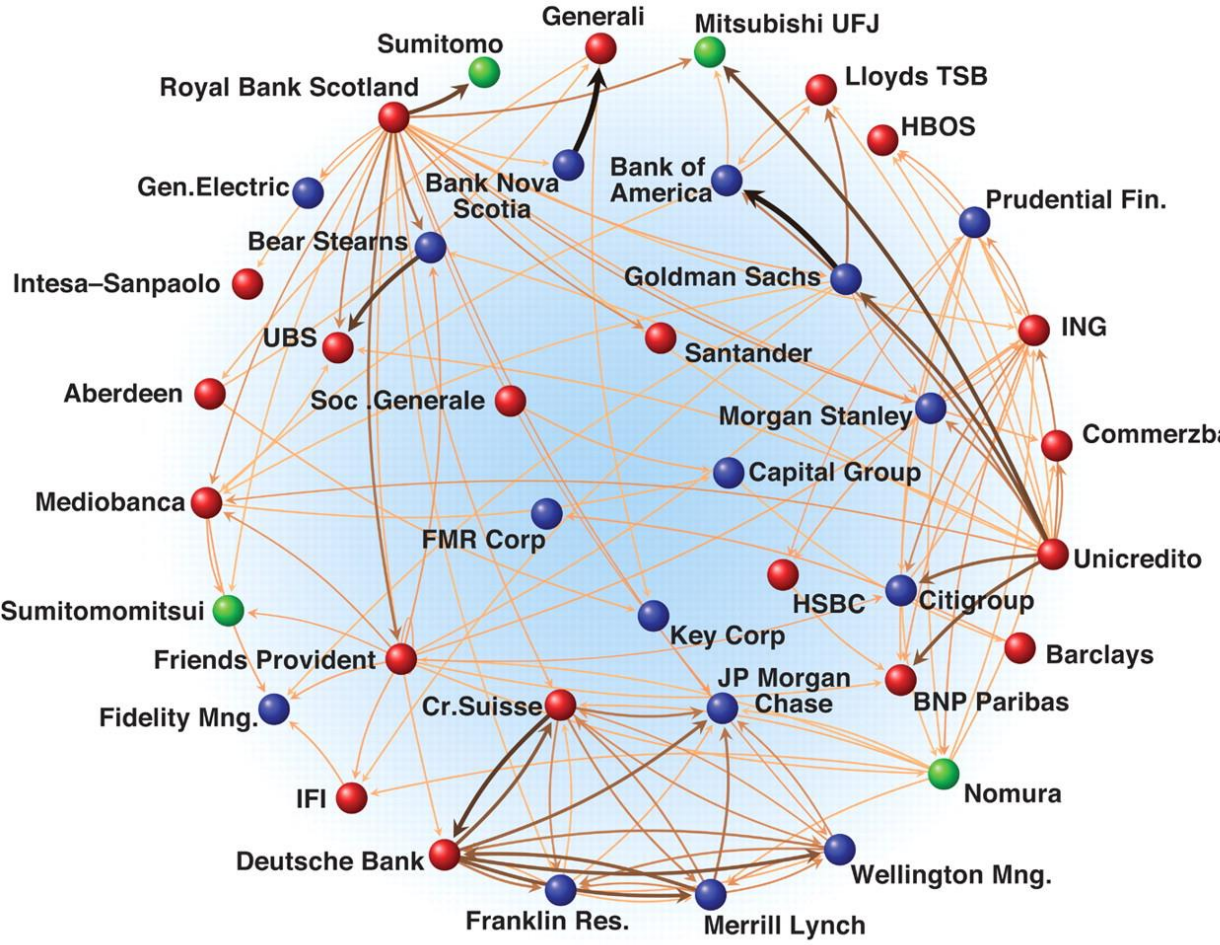
Business Network Analytics

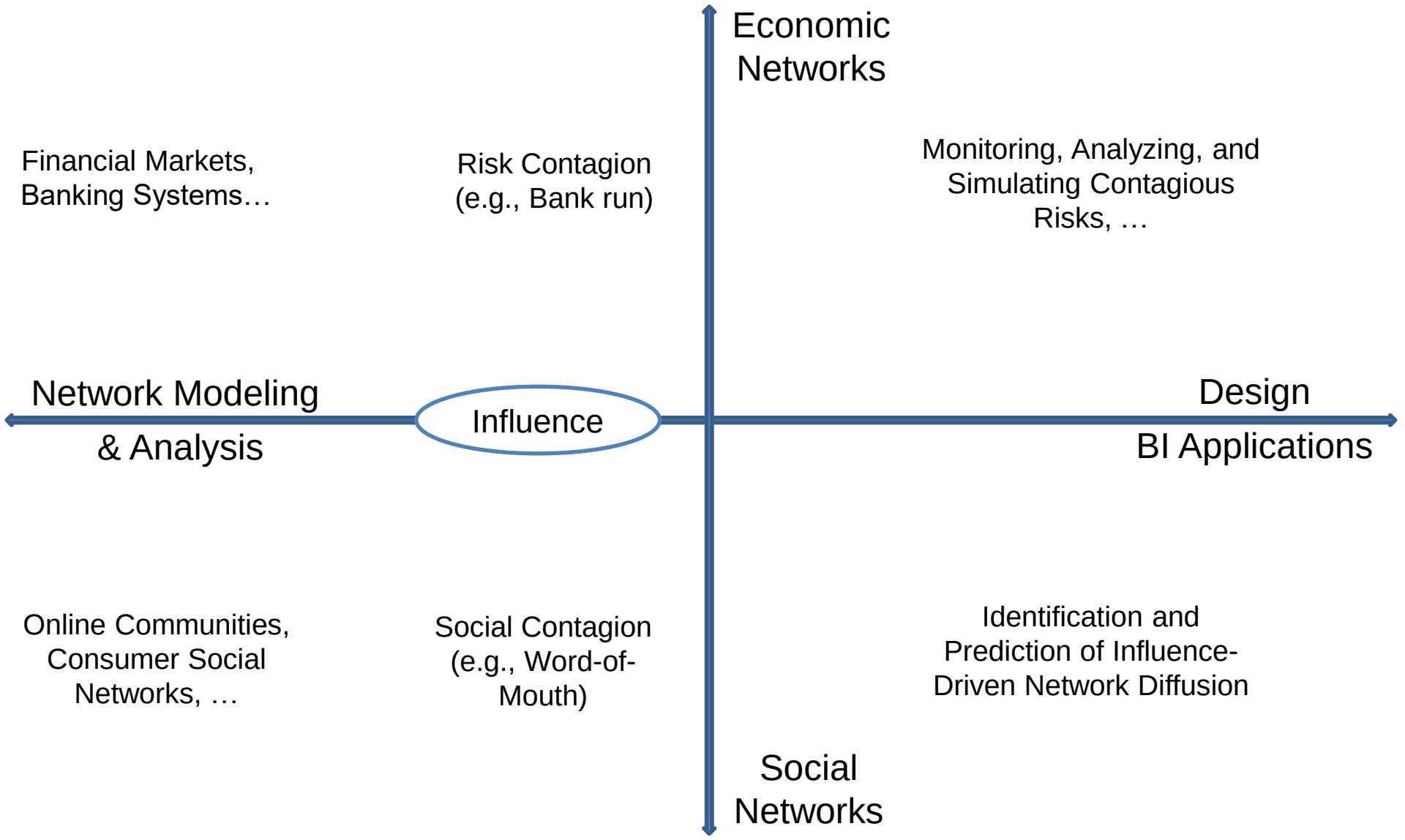


F Schweitzer et al. Science 2009

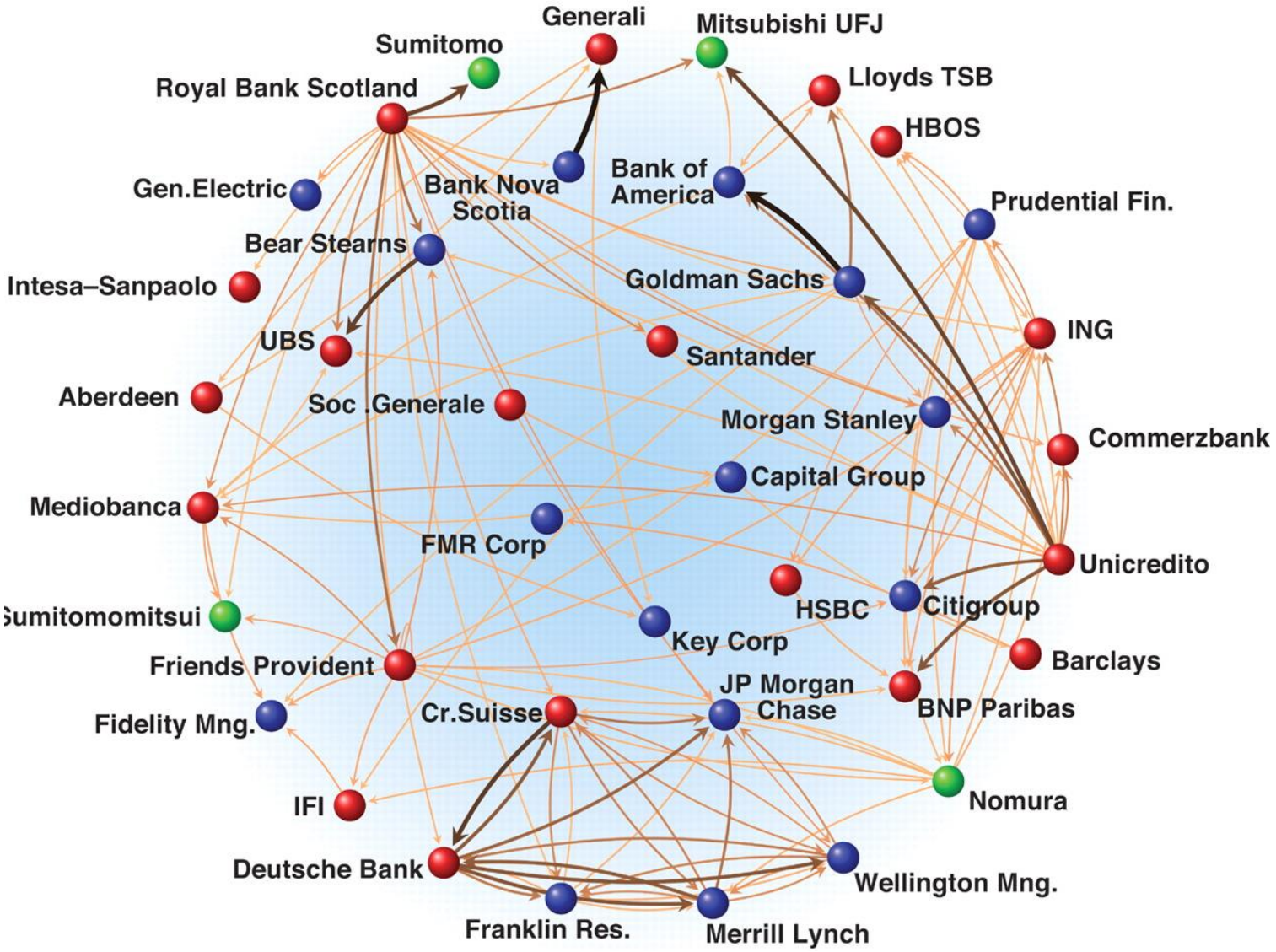
Sep 12th, 2017

Daning Hu
Department of Informatics
University of Zurich
Business Intelligence Research Group

Business Network Analytics and Business Intelligence

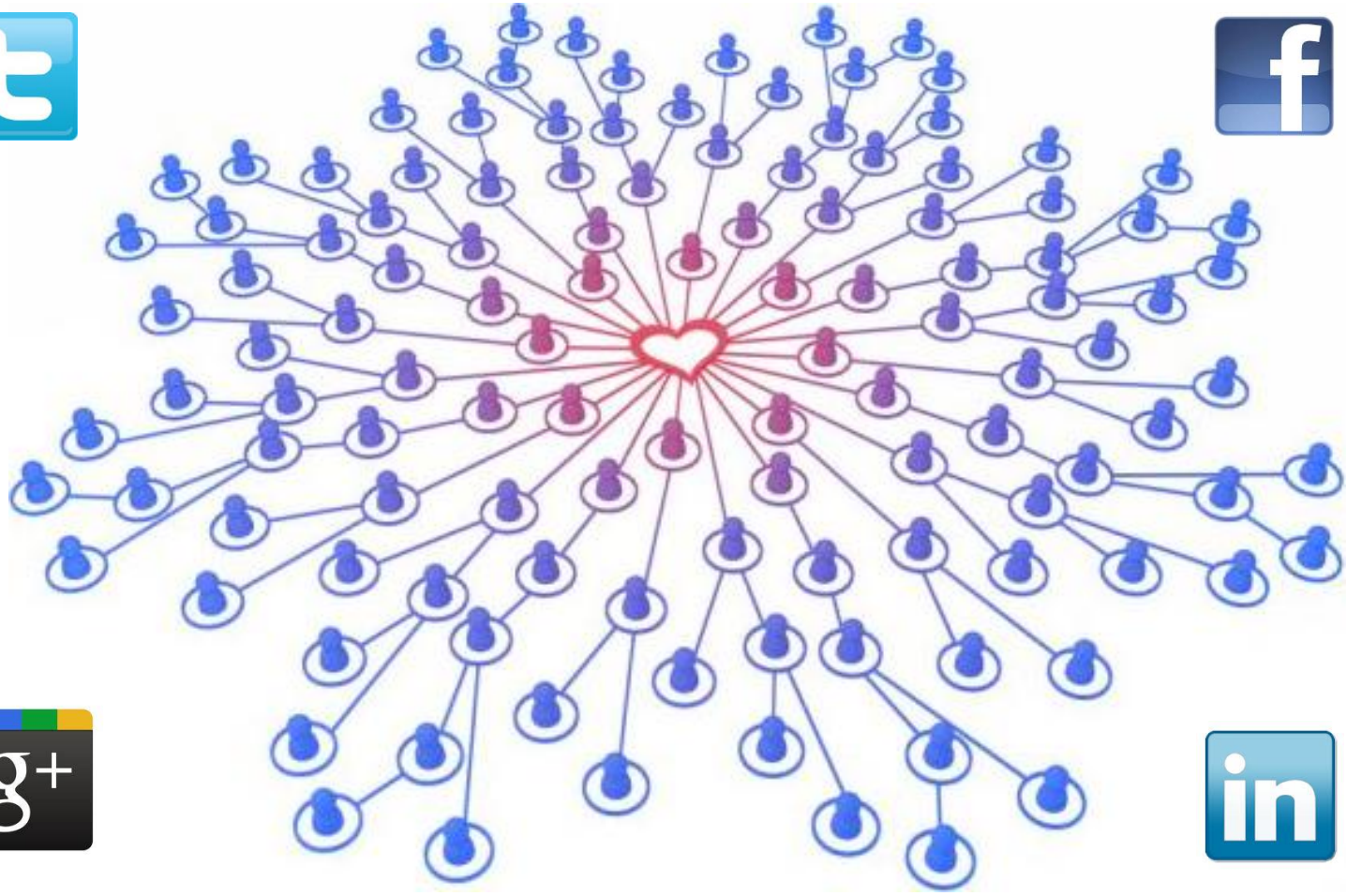


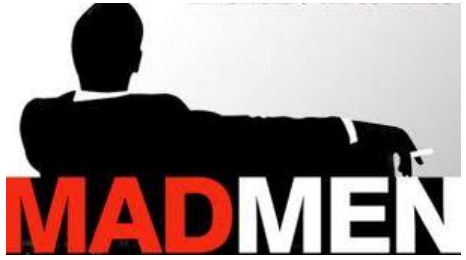
Stop Contagious Failures in Banking Systems



- During 2008 financial tsunami, which bank(s) we should inject capital ***first*** to stop contagious failures in bank networks?

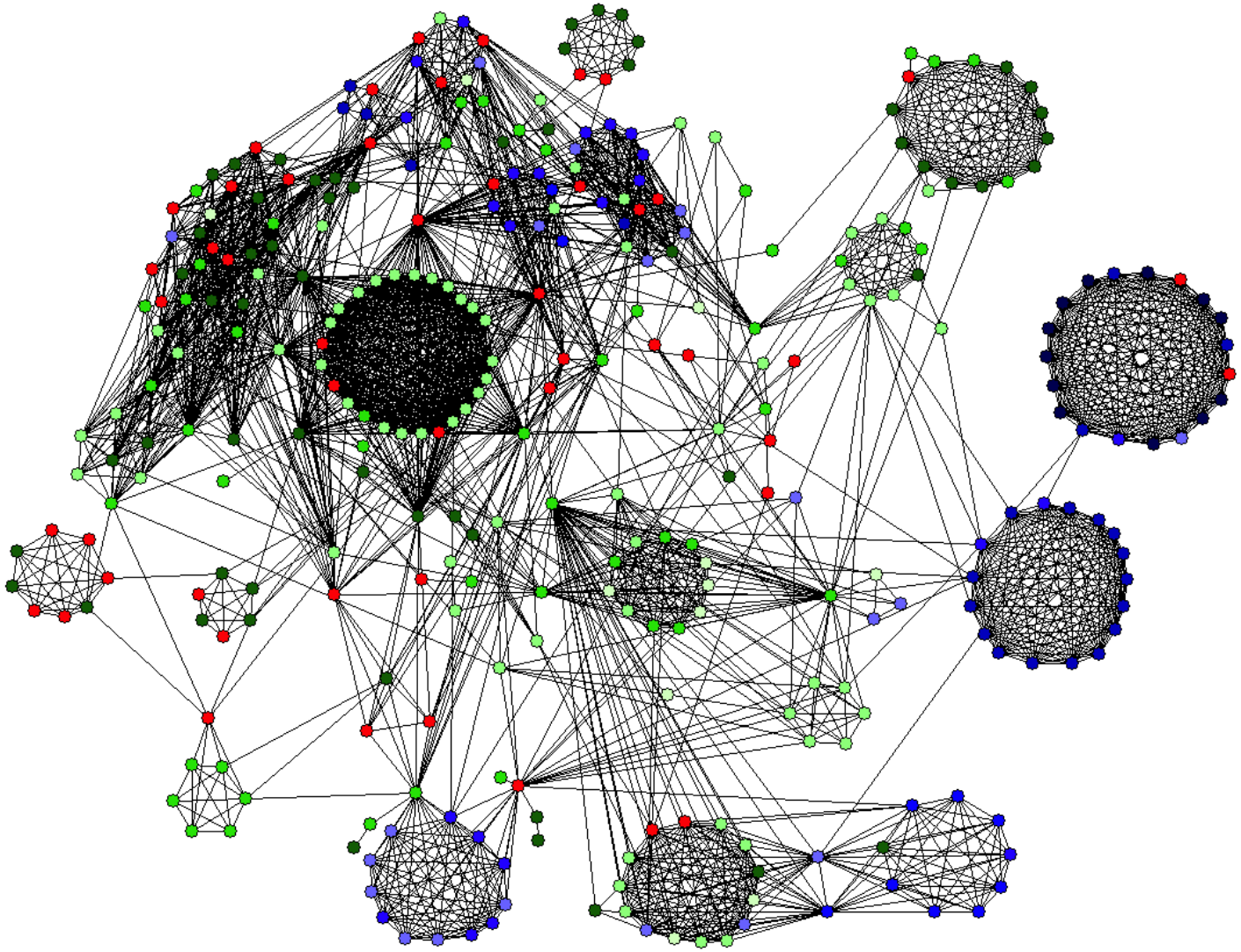
Utilize Peer Influence in Online Social Networks





- Influencer Marketing, Product Recommendation
  - Who are the most influential people?
  - What are the patterns of information diffusion?

Develop Strategies to Attack Terrorist Networks



A Global Salafi Jihad Terrorist Network
Hu et al. JHSEM 2009

- How to effectively break down a terrorist network?

Business Networks Analytics and Applications

- Instructor: Prof. Dr. Daning Hu, BIN. 2.A.12
- TA: Xiao Li, BIN.2.A.24
- Email: hdaning@ifi.uzh.ch
- Credits: 5 ECTS credits
- Class Schedule: http://www.ifi.uzh.ch/en/bi/teaching/Fall2017/NA.html
- Language: English
- Audience: Undergraduate and Master students
- Office Hours: Email for appointment, Room 2.A.12.

Grading and Course Goals

- 1. One course project (90%)

- 2. Active participation and interaction during the lectures and tutorials (10%)

- The project report should include the following four major components:
  - **Network/Relational Data Collection** (15%)
  - **Network Data Processing and Modeling** (20%)
  - **Network Visualization** (15%)
  - **Network Analysis** (30%)

Example 1: Network Data Collection

- Social Networks: Online communities, Social networking websites, Personal blogs and micro-bloggings, online video sharing websites. (e.g., **Programmable Web**)

- E-Business: Amazon Web Service, Ebay Data API, Taobao.

- Others: Financial, Education data sources: Stanford SNAP Portal

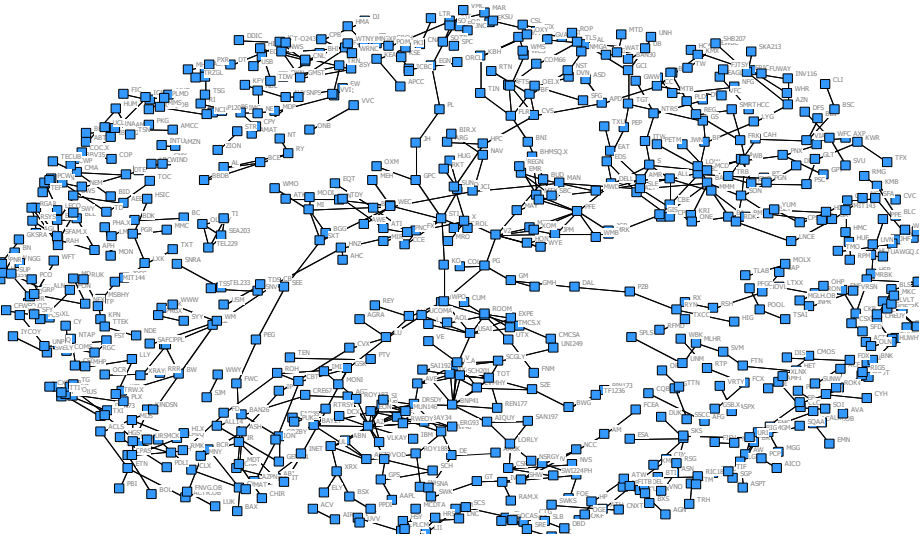
Example 2: Network Data Processing Modeling

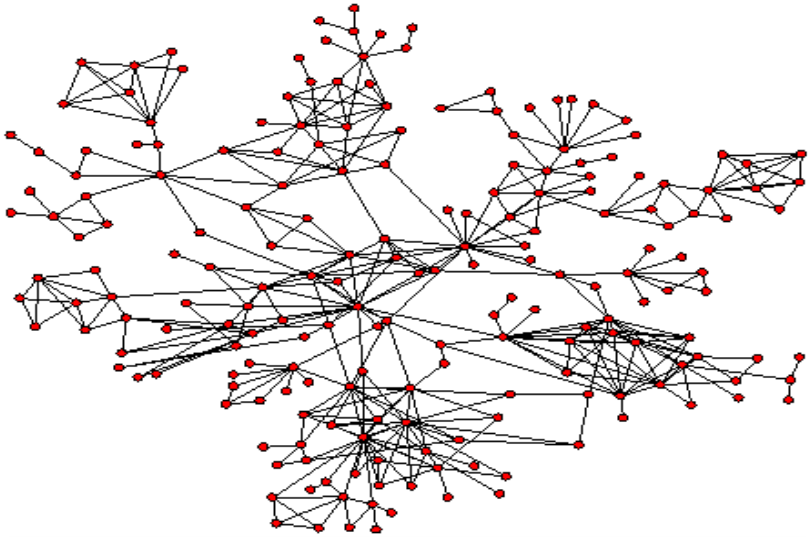
- Extract relations/links from raw data in database tables

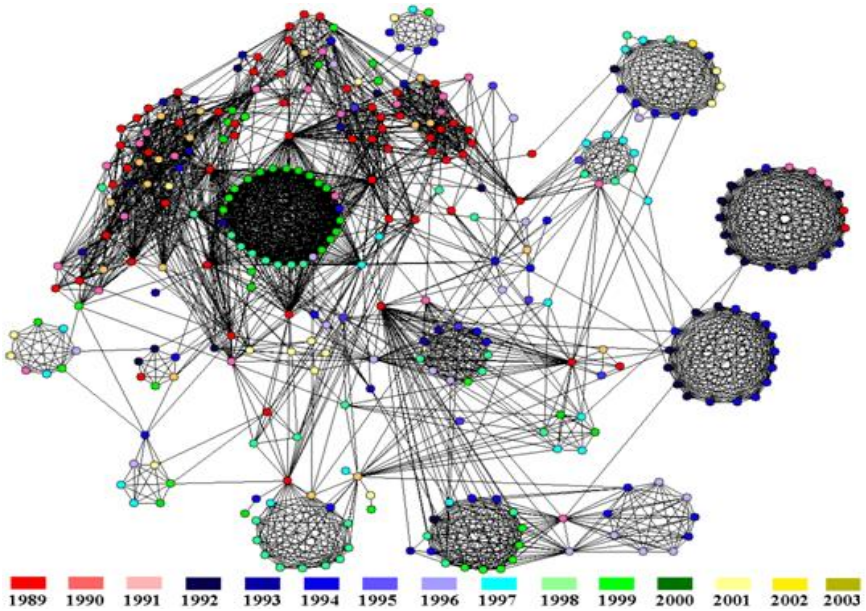
Name	Thread pitch (mm)	Minor diameter tolerance	Nominal diameter (mm)	Head shape	Price for 50 screws	Available at factory outlet?	Number in stock	Flat or Phillips head?
M4	0.7	4g	4	Pan	$10.08	Yes	276	Flat
M5	0.8	4g	5	Round	$13.89	Yes	183	Both
M6	1	5g	6	Button	$10.42	Yes	1043	Flat
M8	1.25	5g	8	Pan	$11.98	No	298	Phillips
M10	1.5	6g	10	Round	$16.74	Yes	488	Phillips
M12	1.75	7g	12	Pan	$18.26	No	998	Flat
M14	2	7g	14	Round	$21.19	No	235	Phillips
M16	2	8g	16	Button	$23.57	Yes	292	Both
M18	2.1	8g	18	Button	$25.87	No	664	Both
M20	2.4	8g	20	Pan	$29.09	Yes	486	Both
M24	2.55	9g	24	Round	$33.01	Yes	982	Phillips
M28	2.7	10g	28	Button	$35.66	No	1067	Phillips
M36	3.2	12g	36	Pan	$41.32	No	434	Both
M50	4.5	15g	50	Pan	$44.72	No	740	Flat

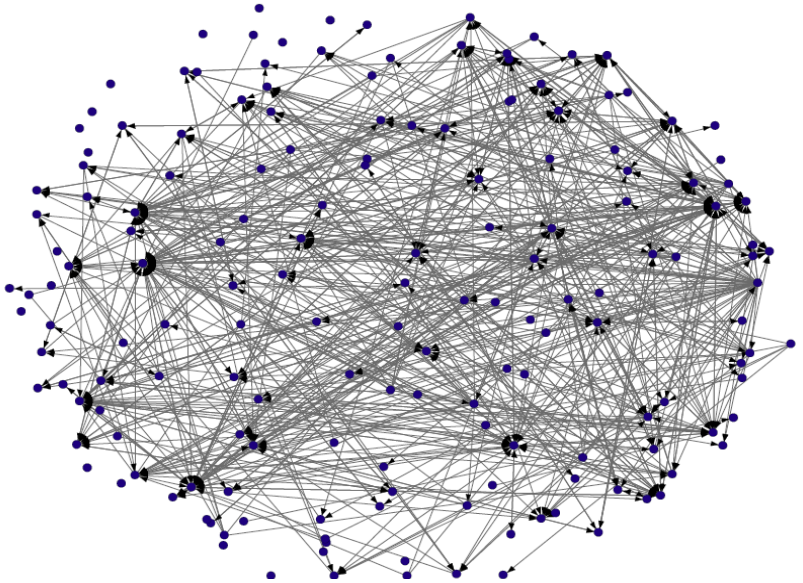
- Model such relations/links into network data.
  - Node data
  - Link data

Example 3: Network Visualization

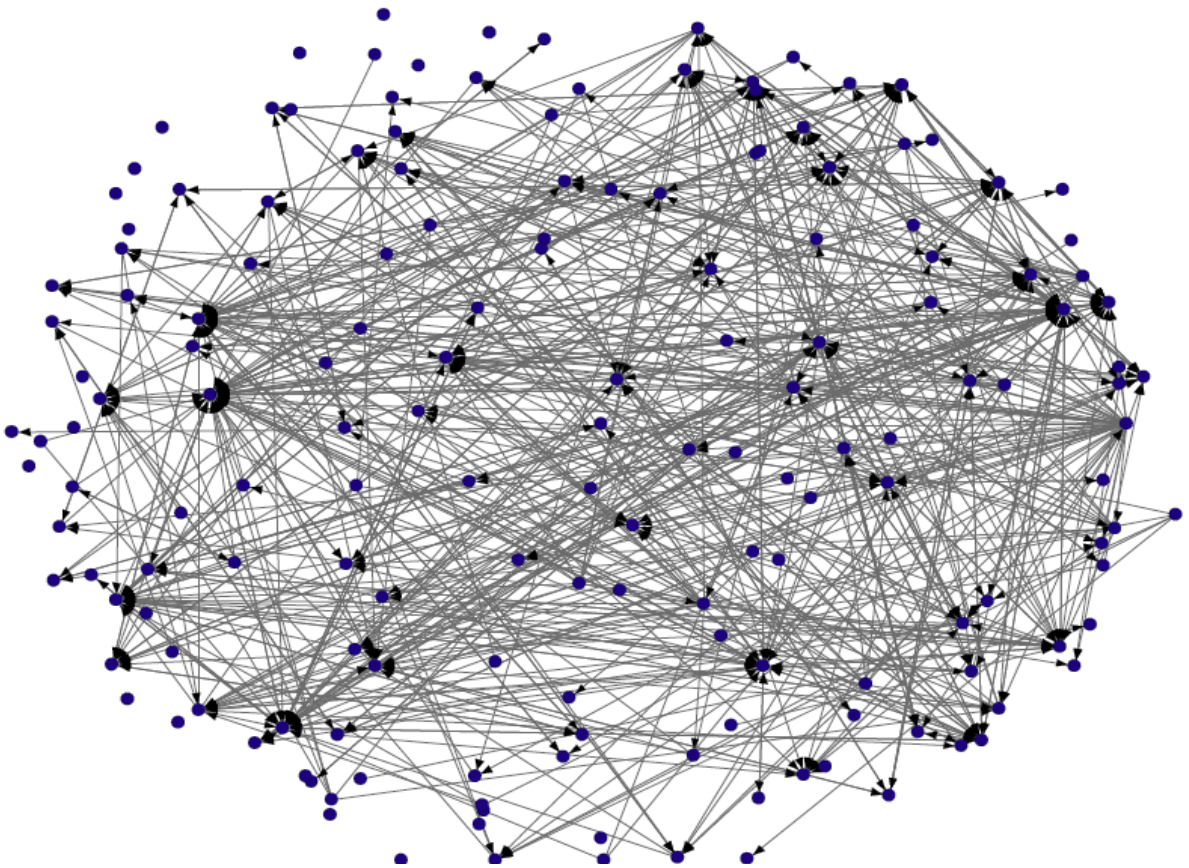






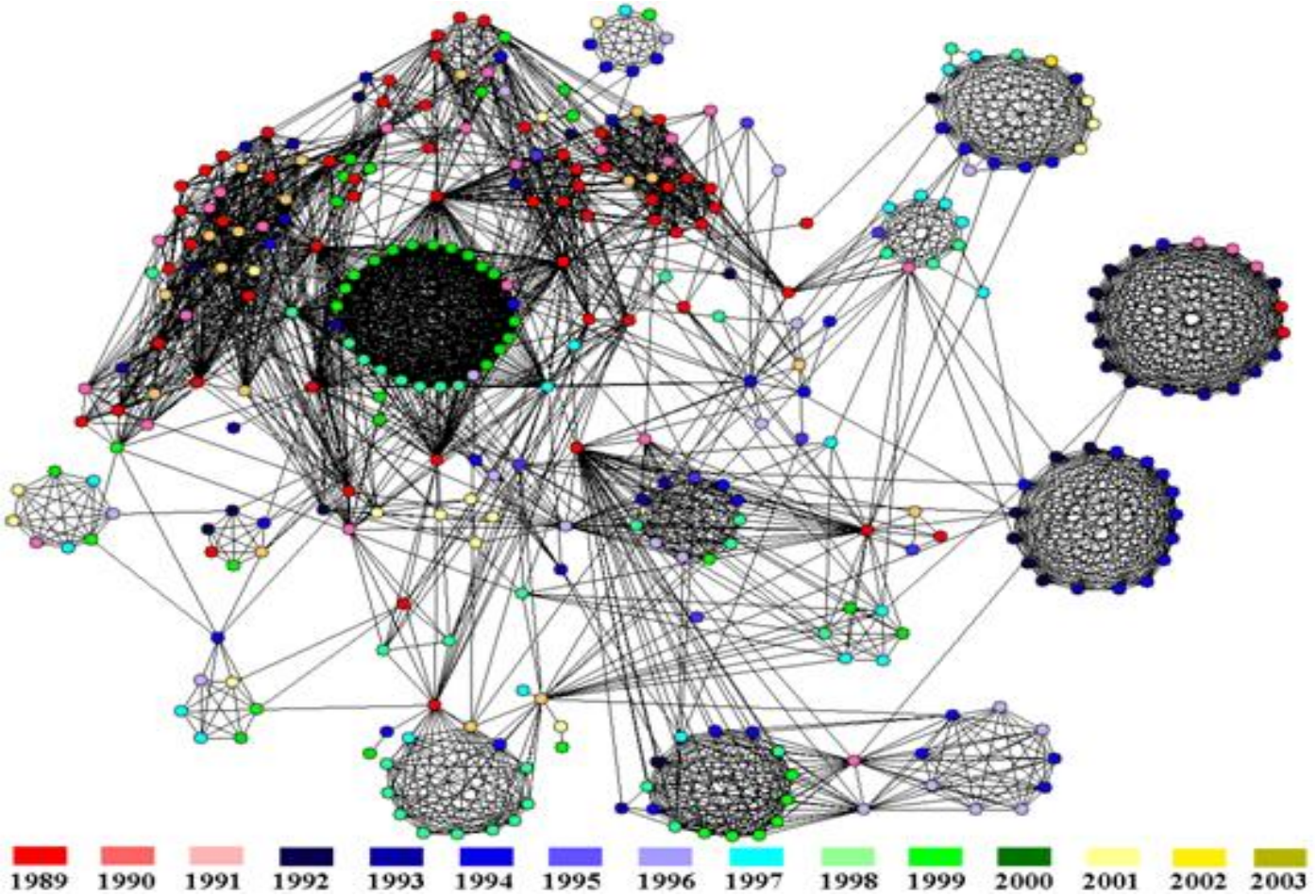


Example 4: Network Analytics



- A Microblogging Network: Who possesses the most advantageous position in brokering information and knowledge in this network?

Example 5: A Global Terrorist Network



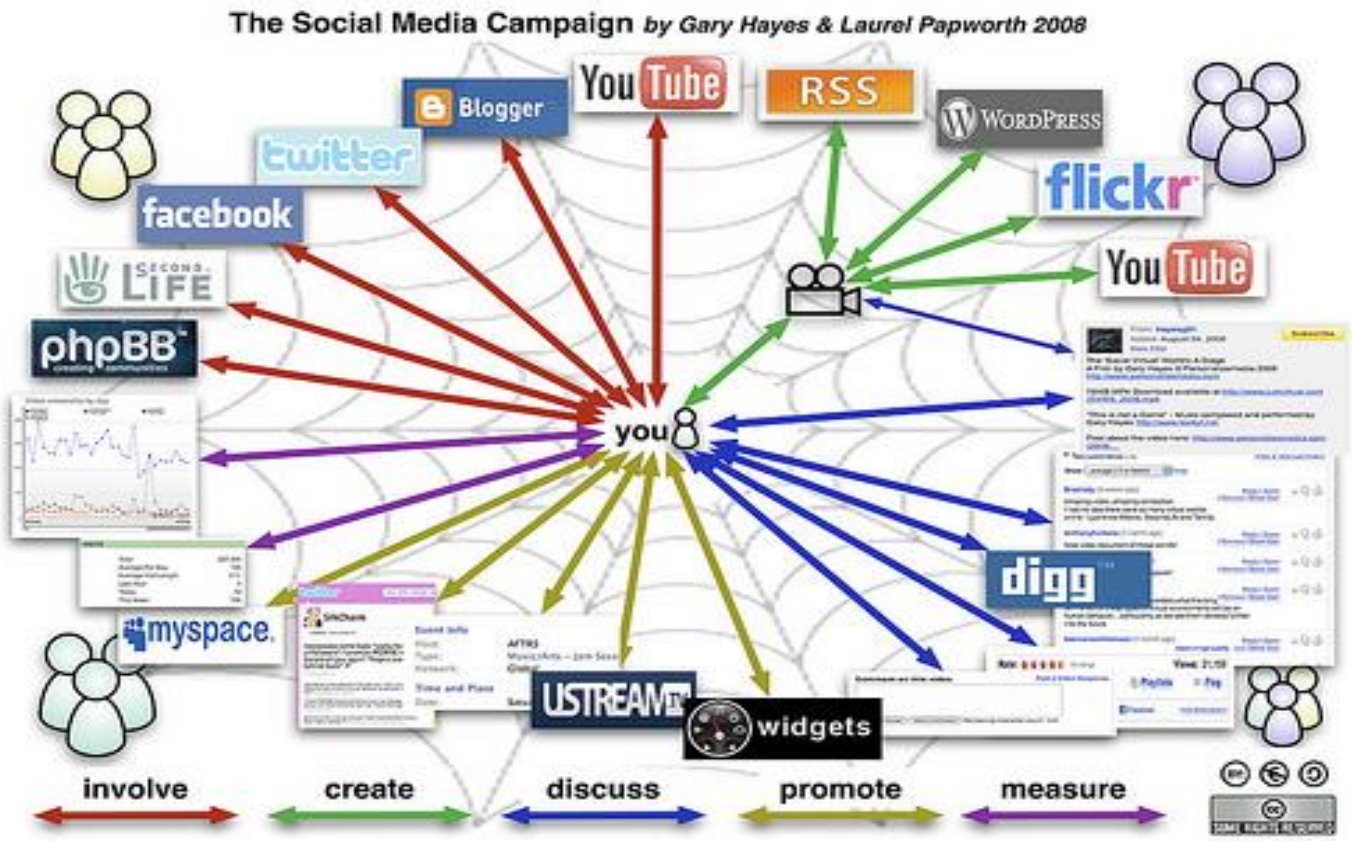
- How to effectively break down terrorist networks?

Business Network Analytics or Applications

- Recommender Systems:

Business Network Analytics or Applications

- Social Media based Marketing, Word-of-Mouth Effect



Computing Tools Required In Tutorials

- Database Management Software
  - MySQL or other common DBMS such as MS SQL Server, Oracle, etc.
- Network Visualization Tool: NetDraw, R (Statnet or iGraph)
- Network Analysis Tool: R (Statnet or iGraph), or UCINet

Outline

- Syllabus
- Examples
- Introduction
- Social Network Analysis
- Social Network Data Modeling and Analysis

* Some of the contents are adapted from Prof. James Moddy's slides at Duke University, and Prof Jure Leskovec and Lada Adamic from Standford University

Introduction: Why Study Networks?

- One of the most profound changes in today's world is - **Decentralization**
  - Economical: BitCoin, Blockchain, P2P lending, etc.
  - Social: Social Media News, Online Communities, Terrorist Cells, etc.
  - Technological: Open Source Software, Virtual Teams, etc.

- The power, information, resources in real world networks are becoming increasingly decentralized ->
  - Nodes are distributed more equally, Less hierarchical;
  - Good representations of entities and relationships in decentralized systems.

The Focus: The Influence Mechanisms in Networks

- An average individual (node) can affect system outcome by **influencing** its linked peers.

- This course will focus on the "**influence**" mechanisms that network actors affect each other in terms of opinions, behaviors, or risks through various ties, thereby changing the network outcomes
  - Economic influence: Risk contagion, etc.
  - Social influence: Word-of-Mouth, observation learning, herding, etc.
  - Network outcomes: Bank run, stock market crash, product diffusions.

A "Random" History of Network Science

1736

1930

1990

2000

2017

- **Mathematical** foundation – Graph Theory
- **Social** Network Analysis and Theories
  - Sociogram: Network visualization
  - Six degree of separation
  - Structural hole: Source of innovation
- **(Physicists)** Complex Network Topologies
  - Small-world model (e.g., WWW)
  - Scale-free model ("Rich get richer")
- **Network Science**
  - Economic networks (Agent modeling & simulation)
  - Dynamic network analysis
  - BI applications: product diffusion in social media, recommendation systems
- **?**

Network Science

- Network science is an **interdisciplinary** academic field which studies complex networks such as information networks, biological networks, cognitive and semantic networks, and **social networks**. It draws on theories and methods including:
  - **Graph theory** from mathematics, e.g., Small-world
  - **Statistical mechanics** from physics, e.g., Rich get richer,
  - **Data mining** and **information visualization** from computer science,
  - **Inferential modeling** from statistics, e.g., Collaborative filtering
  - **Social structure** from sociology, e.g., weak tie, structural holes

- Network science can be defined as "*the study of network representations of physical, biological, and social phenomena leading to predictive models of these phenomena.*"

A "Random" History: Math, Psychology, Sociology…

- The study of networks has emerged in diverse disciplines as a means of analyzing complex relational data.

- Network science has its root in ***Graph Theory***.
  - Seven Bridges of Königsberg written by Leonhard Euler in 1736.
  - Focusing on the properties of pairwise relations in a network structure.

- Social Network Analysis
  - Jacob Moreno, a psychologist, developed the Sociogram and to "precisely describe the interpersonal structure of a group".
  - Stanley Milgram (Small World Experiment: Six Degrees of Separation, 1960s). Facebook: 5.28 steps in 2008, 4.74 in 2011.

"For the last thirty years, empirical social research has been dominated by the sample survey. But as usually practiced, …, the survey is a sociological meat grinder, tearing the individual from his social context and guaranteeing that nobody in the study interacts with anyone else in it."

Allen Barton, 1968 (Quoted in Freeman 2004)

Moreover, the complexity of the relational world makes it impossible to identify social connectivity using only our intuition.

Social Network Analysis (SNA) provides a set of tools to empirically extend our theoretical intuition of the patterns that compose social structure.

The Origin of Modern Network Science: Social Network Analysis

- Social network analysis (SNA) is a set of *relational* methods for systematically understanding and identifying connections /ties /relationships among actors.

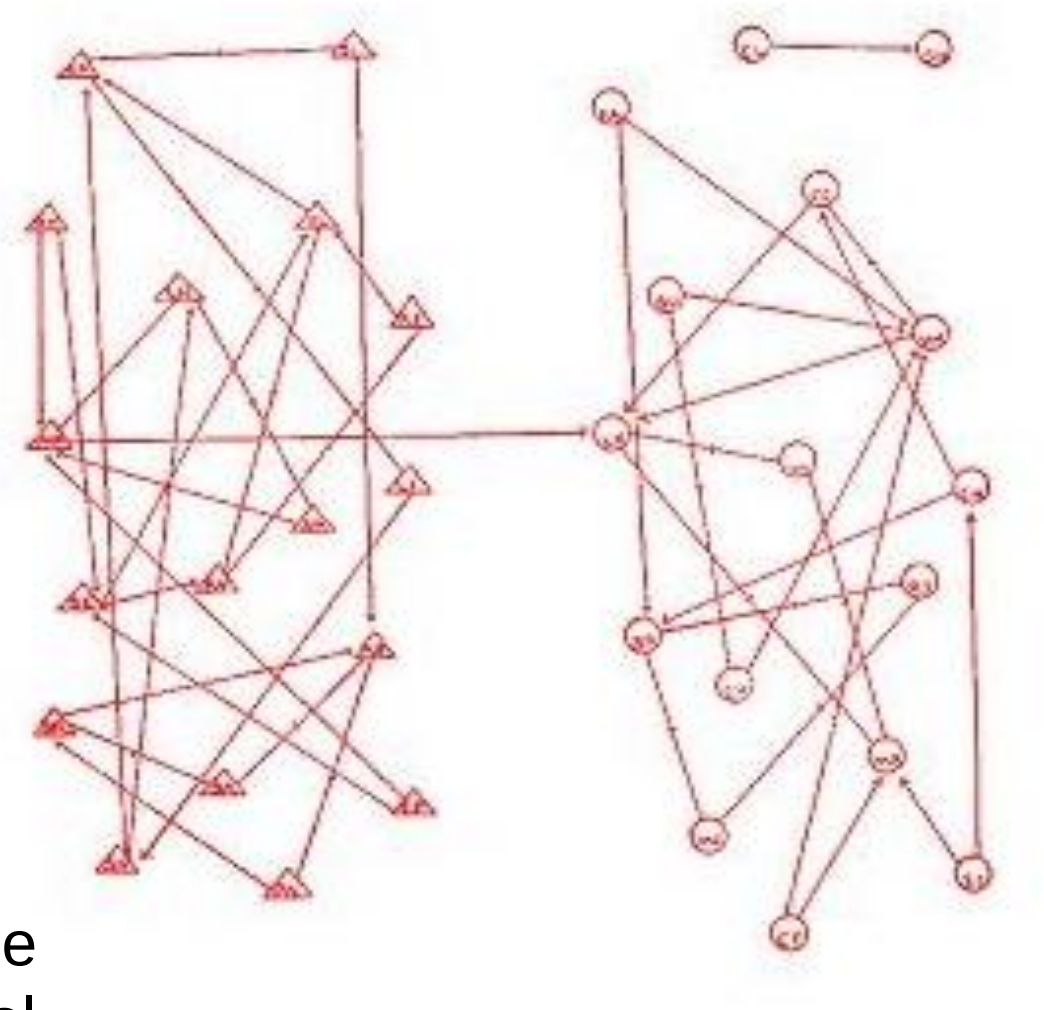
- Social network analysis (SNA)
  - is motivated by a structural intuition based on ties linking social actors
  - is grounded in systematic empirical data
  - draws heavily on graphic theory and imagery
  - relies on the use of mathematical and computational models.

Jacob Moreno's experiment on Friendship Network

EMOTIONS MAPPED BY NEW GEOGRAPHY

Charts Seek to Portray the Psychological Currents of Human Relationships.

New York Times
April 3, 1933



Jacob's experiment is the first to use Social Network Analysis

What Does Social Network Analysis Study?

Social Network analysis lets us answer questions about social interdependence. These include:

"Networks as Variables" approaches

- Are kids with smoking peers more likely to smoke themselves?
- Do unpopular kids get in more trouble than popular kids?
- Do central actors control resources?

"Networks as Structures" approaches

- What generates hierarchy in social relations?
- What network patterns spread diseases most quickly?
- How do role sets evolve out of consistent relational activity?

We don't want to draw this line too sharply: emergent role positions can affect individual outcomes in a 'variable' way, and variable approaches constrain relational activity.

Now…

Complex Networks in the Real World

	Nodes	**Links**
Social network	People	Friendship, kinship, collaboration
Inter-organizational network	Companies	Strategic alliance, buyer-seller relation, joint venture
Citation network	Documents/authors	Citations
Internet	Routers/computers	Wire, cable
WWW	Web pages	hyperlink
Biochemical network	Genes/proteins	Regulatory effect
…	…	…

Now…

- Universal modeling and analysis methods for complex network data
- Shared vocabulary between fields: Computer Science, Physics, Sociology, Economics, Statistics, Biology
- “Big” Data availability: Internet, mobile, bio, health, security…
- Impact/usage: social networking, social media, marketing, etc.

Our Approach

What

- Social network analysis (Metrics)
- **Describe** the changes in network evolution
  - Temporal changes in network topological measures
- Dynamic network recovery
- (Relational) data mining

Why

- Econometric **identification** of casual Social and Economic influence
  - Distinguishing homophily
  - Confounding factors
  - PSM, DID, RD, etc.
  - Explanations

How

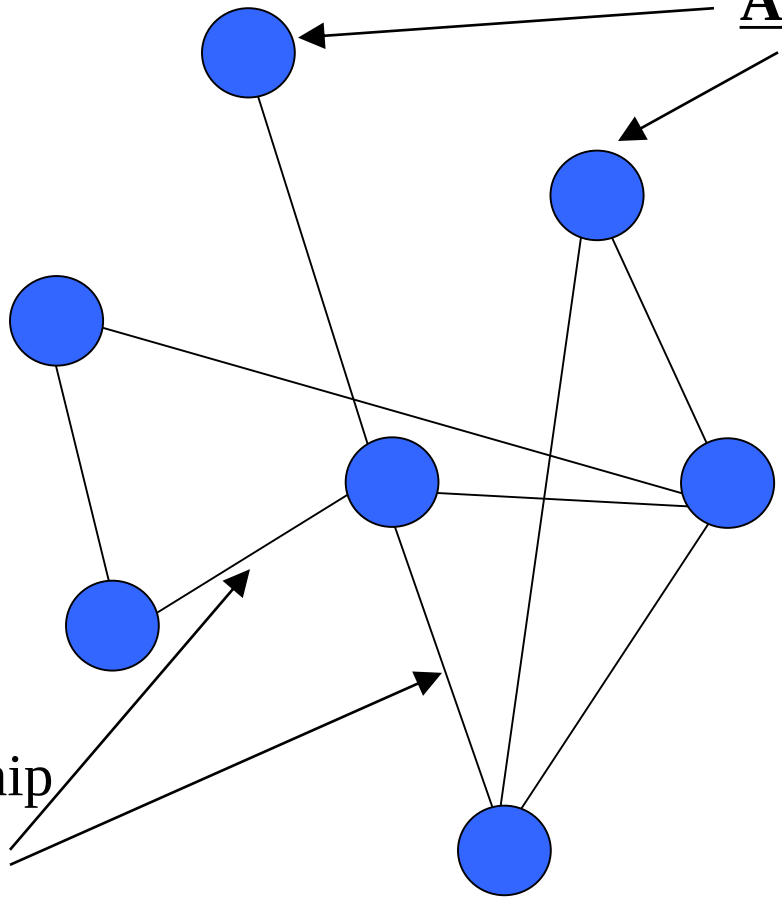
- **Combine** social science methods, data mining, machine learning with econometric analysis
- **Predict** link formation
- **Simulate** the evolution of networks

What: Social Network Analysis

- Social network analysis (SNA) is a set of **metrics** and methods for systematically **describing,** modelling, and analyzing relationships among actors.

- Social network analysis (SNA)
  - is motivated by a structural intuition based on ties linking social actors
  - is grounded in systematic empirical data
  - draws heavily on graphic imagery
  - relies on the use of mathematical and/or computational models.

What is a Network?



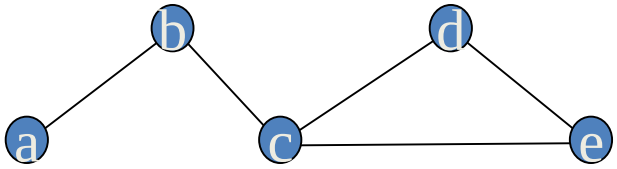
Actor/Node: Any entity in a network (person, system, group, organization)

Tie/Link: Relationship or interaction between two nodes.

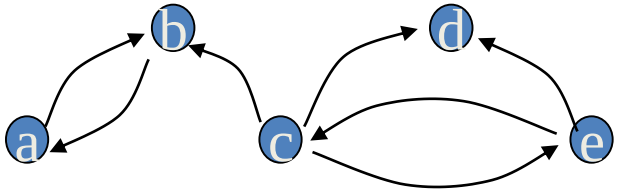
Basic Concepts in (Social) Network Analysis

- Node, Actor, Vertex ***V***
- Tie, Link, Edges ***E***
- Network, System ***G (V, E)***

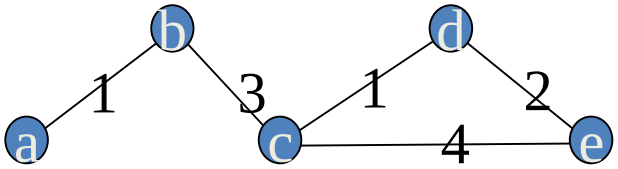
- A link can be (1) Binary or Valued, (2) Directed or Undirected.



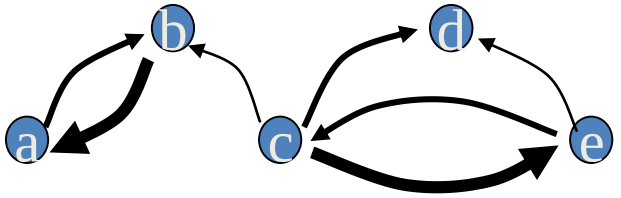
Undirected, binary



Directed, binary



Undirected, Valued



Directed, Valued

Nodes or Social Actors

- Social Network data consists of two linked classes of data: Nodes and Links.
- Node Example: Products in a purchase newtork
- Actor Examples: people in a group, departments within in a corporation, public service agency in a city, nation-states in the world system. “Node” does not imply that they have intention or the ability to “act”.
- Network nodes are most often people, but can be any other unit capable of being linked to another (schools, countries, organizations, personalities, etc.)

Links or Ties

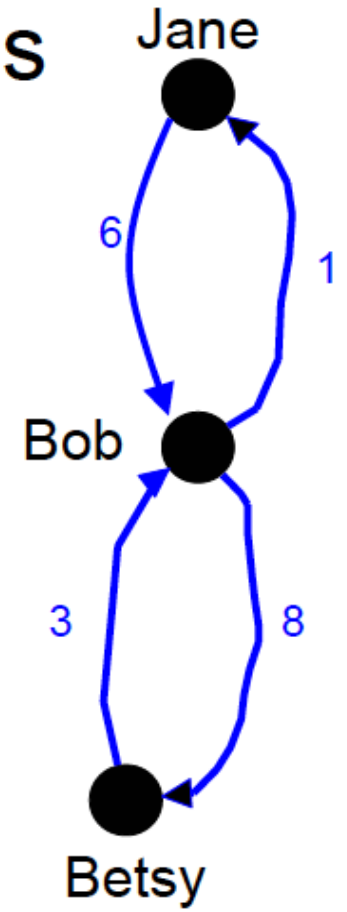
- Actors (nodes) are linked to one another by social ties (links)
  - Kinship, role-based, cognitive, affective, interactions, affliations

- Example of **direct** links in SNA (Wasserman/Faust 2008:17):
  - Evaluation of one person by another (friendship, liking, or respect)
  - Transfers of material resources (business transactions, lending or borrowing things)
  - Behavioral interaction (talking together, sending messages)
  - Physical connection (a road, river, or bridge connecting two points)

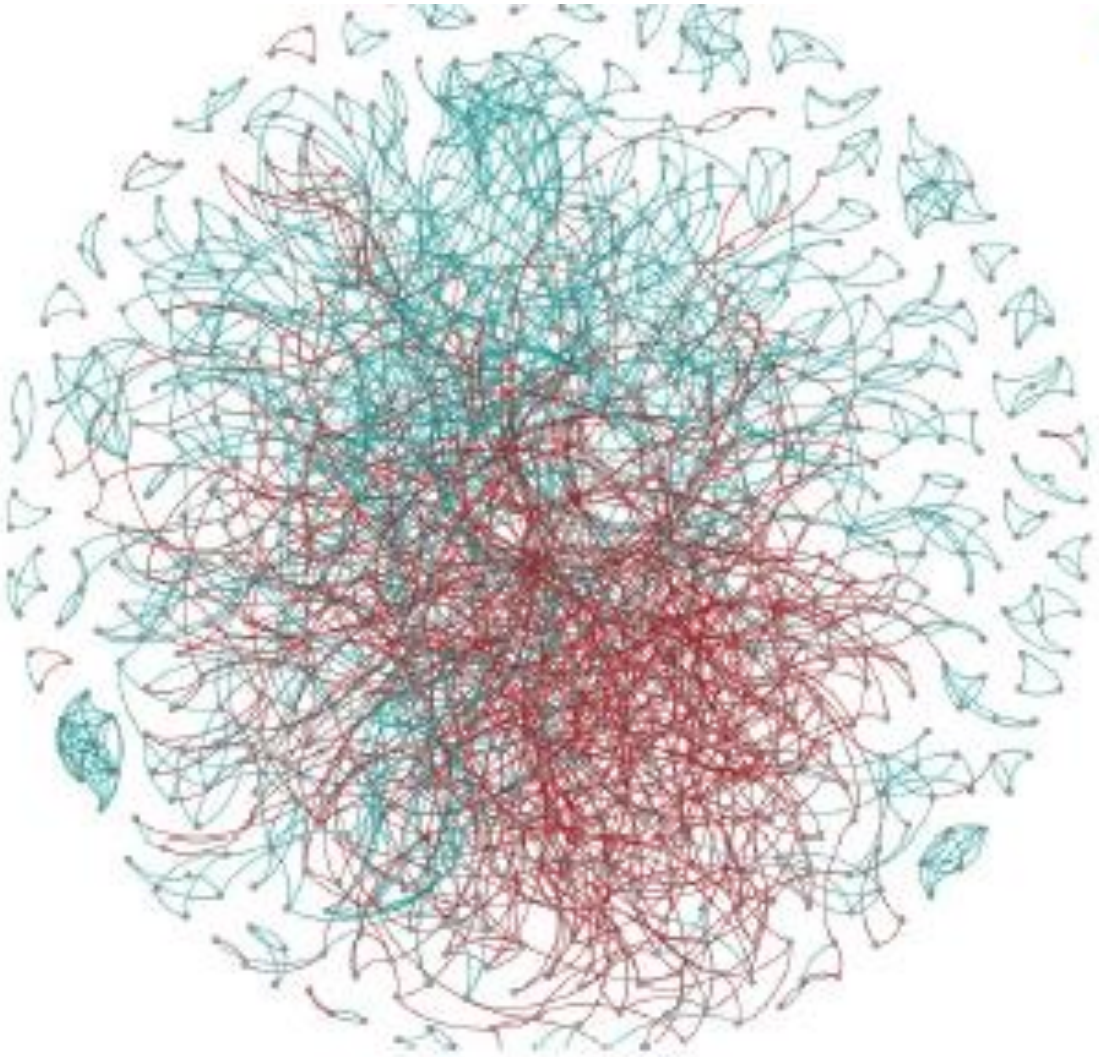
Similarities			Social Relations				Interactions	Flows
Location	Membership	Attribute	Kinship	Other role	Affective	Cognitive	e.g.,	e.g.,
e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	Sex with	Information
Same spatial and temporal space	Same clubs	Same gender	Mother of	Friend of	Likes	Knows	Talked to	Beliefs
	Same events	Same attitude	Sibling of	Boss of	Hates	Knows about	Advice to	Personnel
	etc.	etc.		Student of	etc.		Helped	Resources
				Competitor of		Sees as happy	Harmed	etc.
						etc.	etc.	

Weight/Strength of Ties

- We can attach values to ties, representing quantitative attributes
  - Strength of relationship
  - Information capacity of tie
  - Rates of flow or traffic across tie
  - Distances between nodes
  - Probabilities of passing on information
  - Frequency of interaction
- Valued graphs or vigraphs



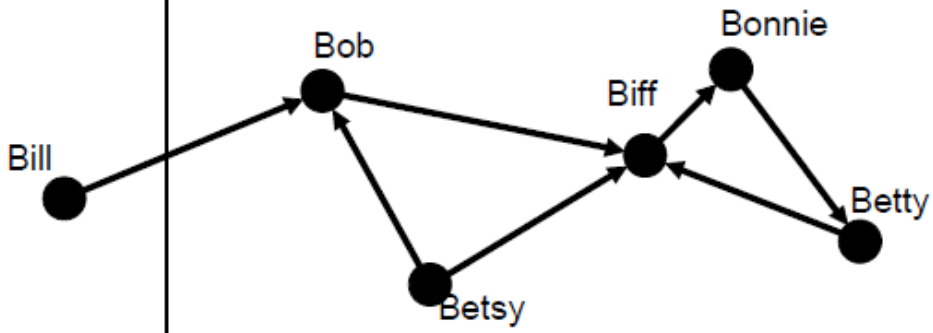
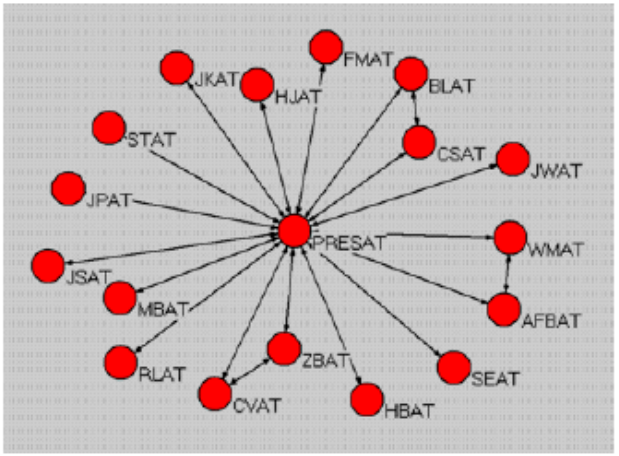
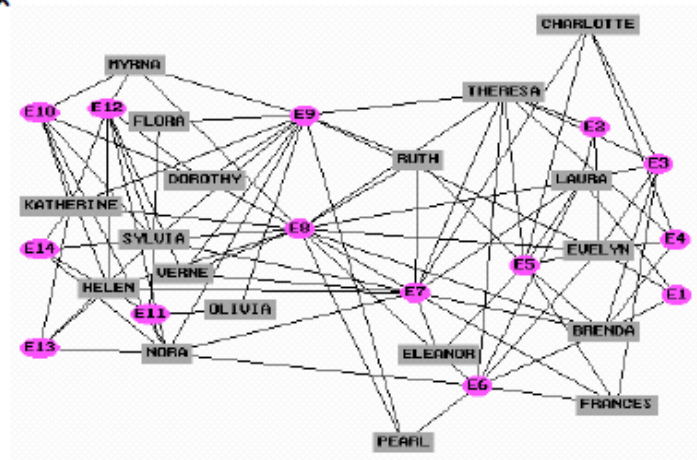
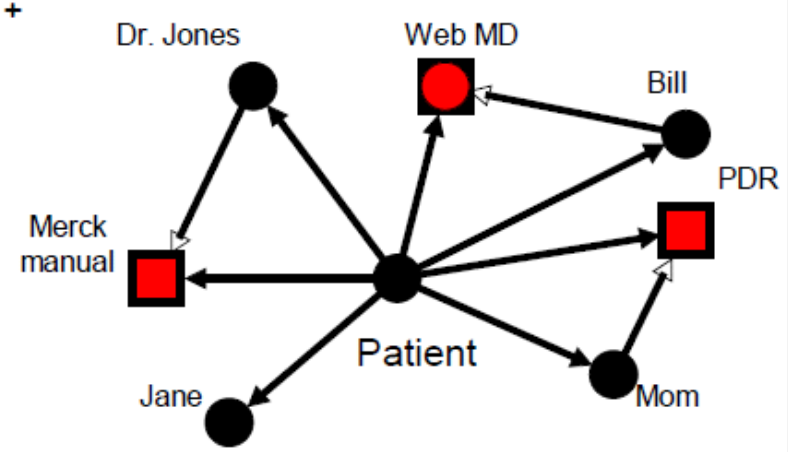
Positive and Negative Weights



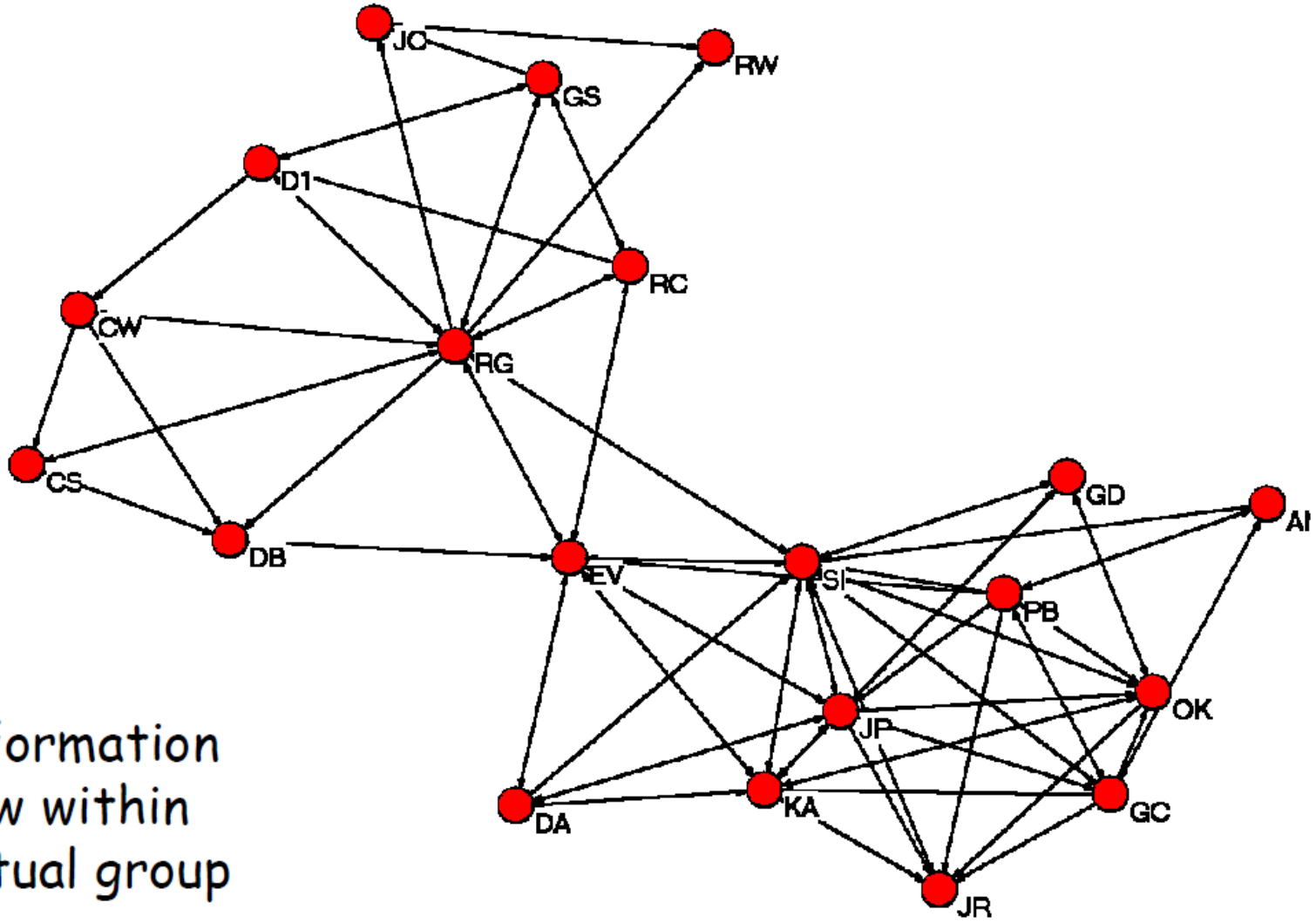
- e.g. one person trusting/distrusting another
  - Research challenge: How does one 'propagate' negative feelings in a social network? Is my enemy's enemy my friend?

sample of positive & negative ratings from Epinions network

Two Modes of Social Network Analysis

	Complete	Ego
1-mode	++++ 	++++ 
2-mode	** 	+ 

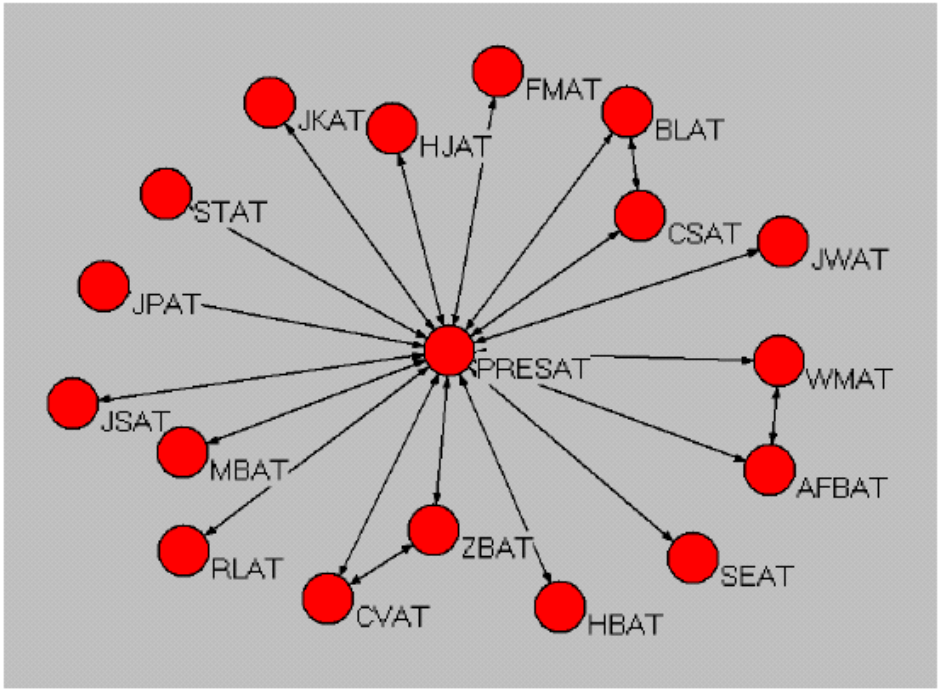
One-mode Complete network



Information flow within virtual group

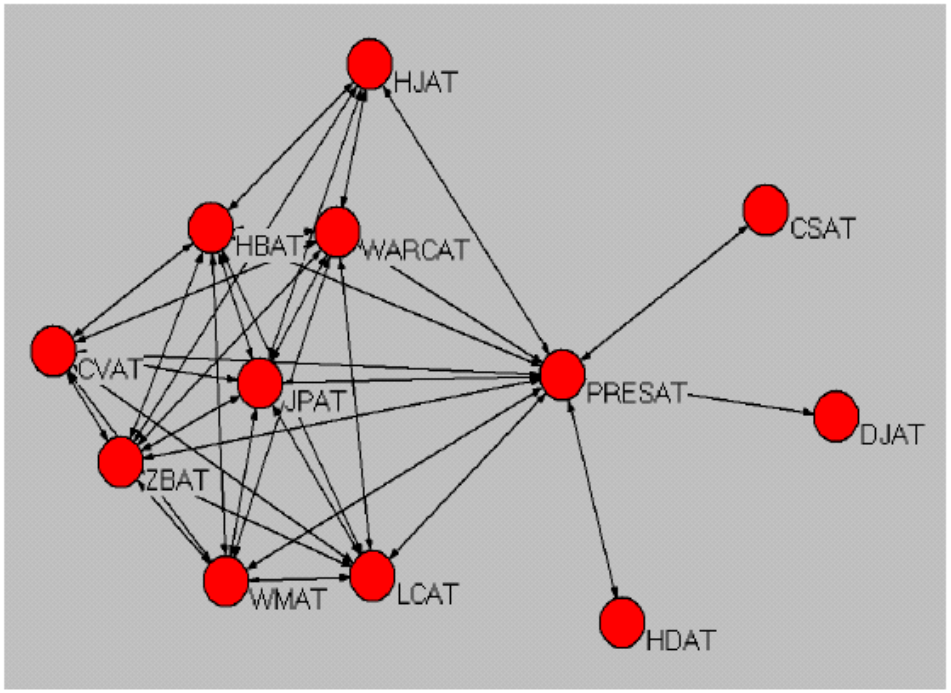
Data collected by Cross

One-mode Ego network



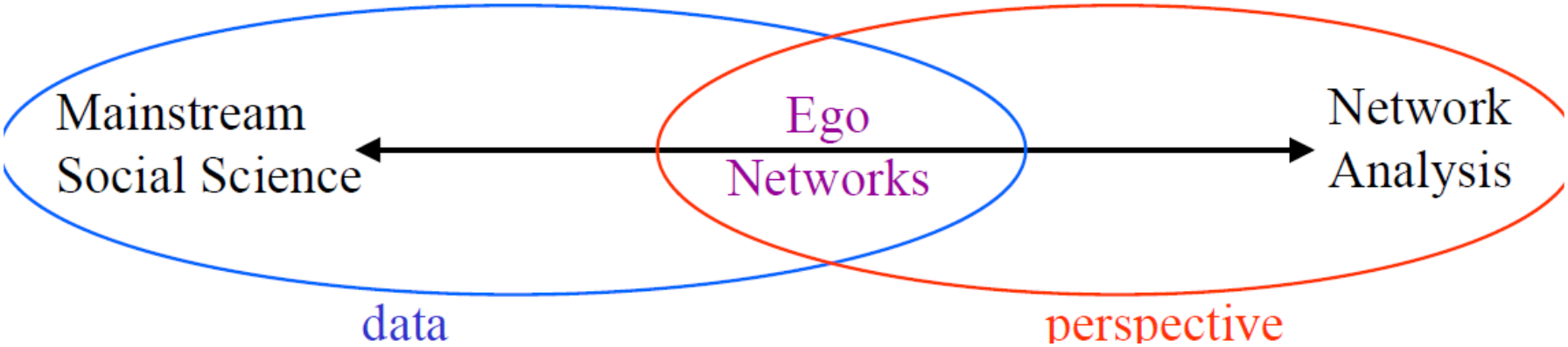
Year 1

Data courtesy of Michael Link



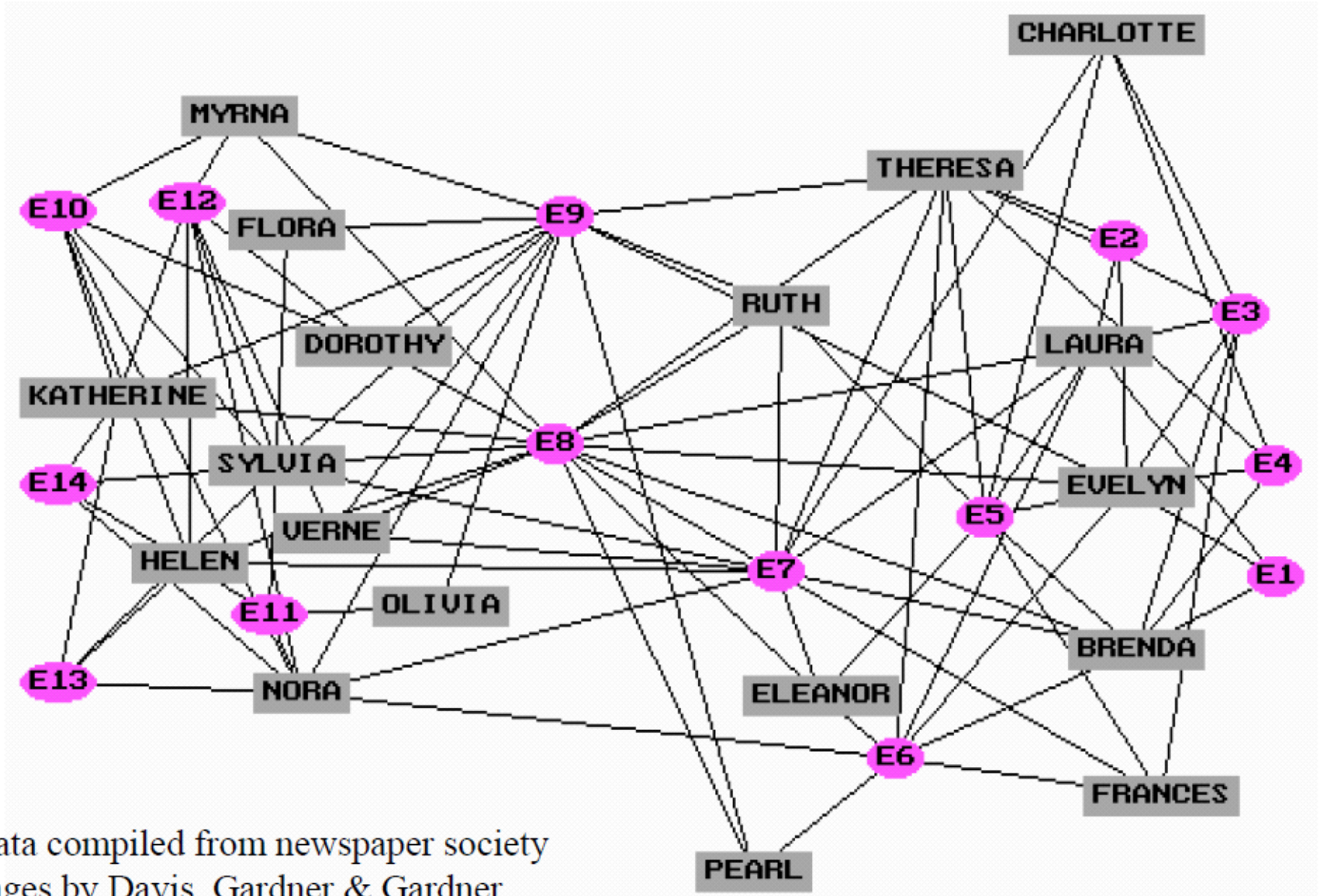
Year 4

Ego Network Analysis



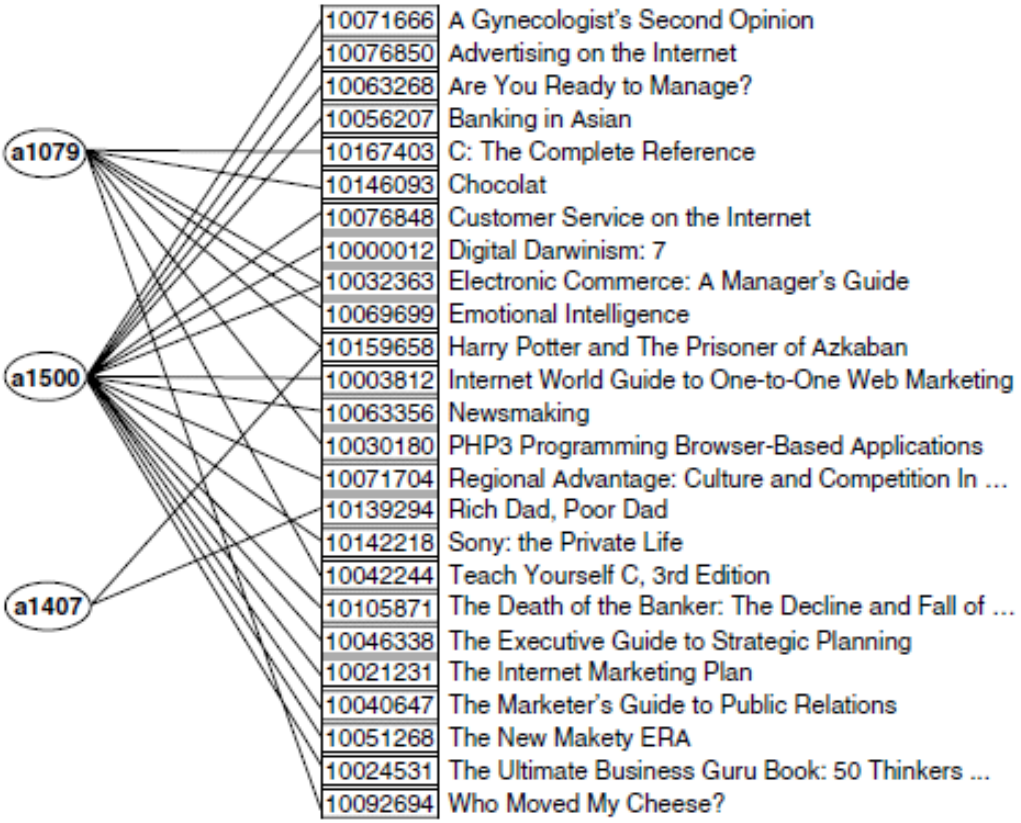
- Ego Network Analysis combine the perspective of network analysis with the data of mainstream social science
- No computer assisted analysis needed

Two-mode Complete Network (Bipartite Graph)



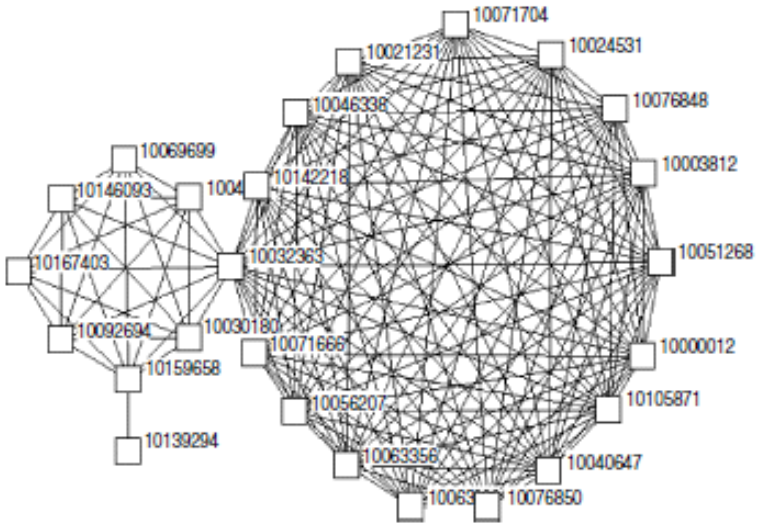
Data compiled from newspaper society pages by Davis, Gardner & Gardner

Two-mode (Bipartite) Network Transformation



(a) A bipartite consumer-product graph consisting of three consumers and 25 books

(b) The consumer graph projected from the bipartite graph depicted in (a)



(c) The product graph projected from the bipartite graph depicted in (a)

- From Zan Huang et al., 2009, Management Science

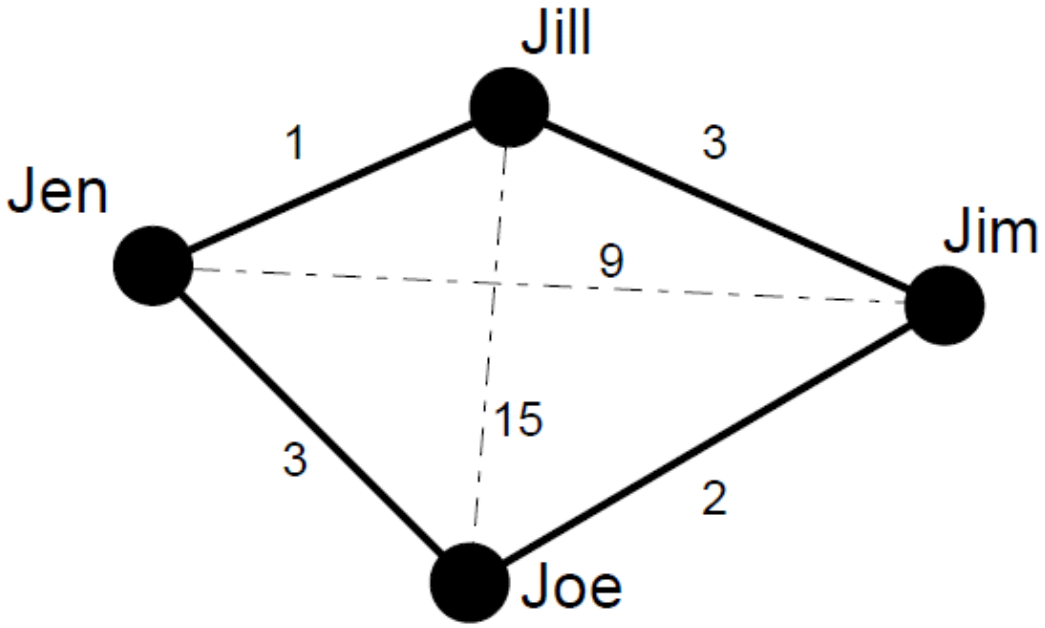
Network Data Modeling: Adjacency Matrix

Friendship

	Jim	Jill	Jen	Joe
Jim	-	1	0	1
Jill	1	-	1	0
Jen	0	1	-	1
Joe	1	0	1	-

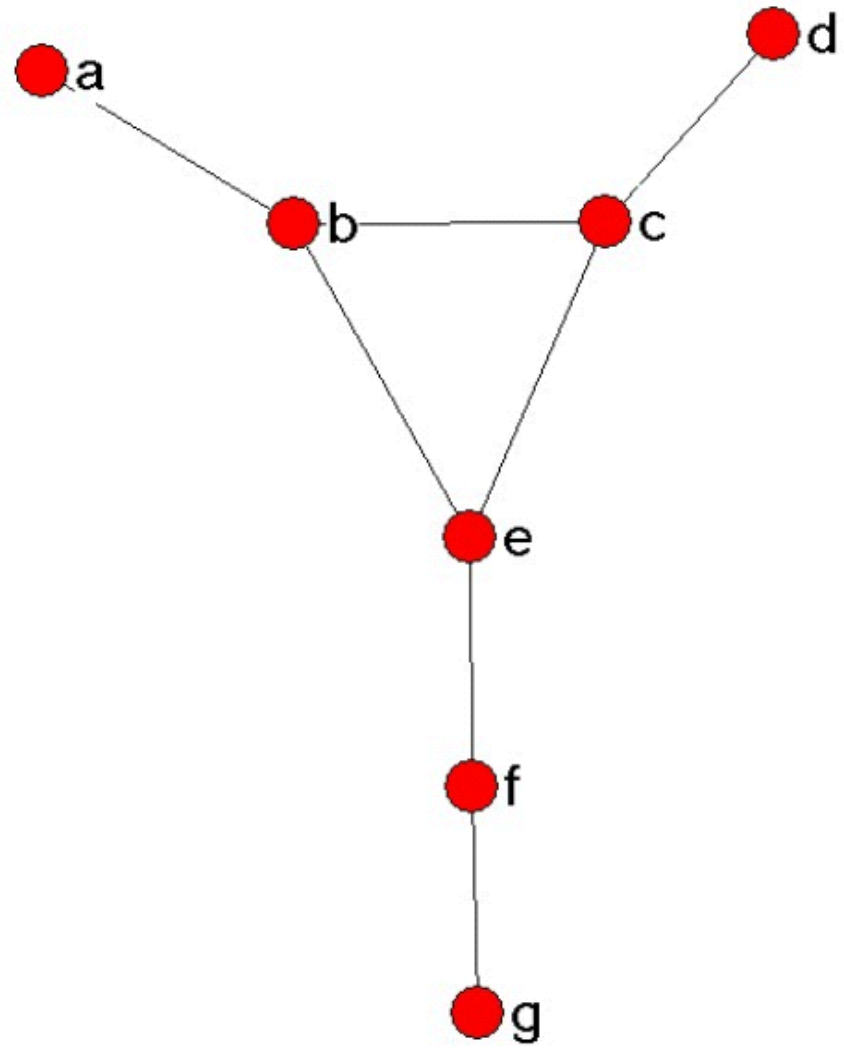
Proximity

	Jim	Jill	Jen	Joe
Jim	-	3	9	2
Jill	3	-	1	15
Jen	9	1	-	3
Joe	2	15	3	-



Network Distance (Weighted) Adjacency Matrix

	a	b	c	d	e	f	g
a	0	1	2	3	2	3	4
b	1	0	1	2	1	2	3
c	2	1	0	1	1	2	3
d	3	2	1	0	2	3	4
e	2	1	1	2	0	1	2
f	3	2	2	3	1	0	1
g	4	3	3	4	2	1	0



Major Network Data Formats (in UCINet)

Dl n = 5
Format = fullmatrix
Labels embedded
Data:

	billy	john	jill	mary
billy	0	1	0	0
john	1	0	1	0
jill	0	0	0	1
mary	1	0	1	0

(Values optional)

Dl n = 5
Format = nodelist
Labels embedded
Data:
Billy jill john jim jane
Jill billy bob bertha
Dick jane
Jim bob billy brenda

(No values possible)

Dl n = 5
Format = edgelist
Labels embedded
Data:
Billy jill
Billy john 6.3
Dick jane
Jim bob 2.5

(Values optional – assigned 1 if omitted)

Real World Networks are Sparse Graphs

Most real-world networks are sparse

$$E \ll E_{max} \quad (\text{or } \bar{k} \ll N-1)$$

WWW (Stanford-Berkeley):	N=319,717	⟨k⟩=9.65
Social networks (LinkedIn):	N=6,946,668	⟨k⟩=8.87
Communication (MSN IM):	N=242,720,596	⟨k⟩=11.1
Coauthorships (DBLP):	N=317,080	⟨k⟩=6.62
Internet (AS-Skitter):	N=1,719,037	⟨k⟩=14.91
Roads (California):	N=1,957,027	⟨k⟩=2.82
Proteins (S. Cerevisiae):	N=1,870	⟨k⟩=2.39

(Source: *Leskovec et al., Internet Mathematics, 2009*)

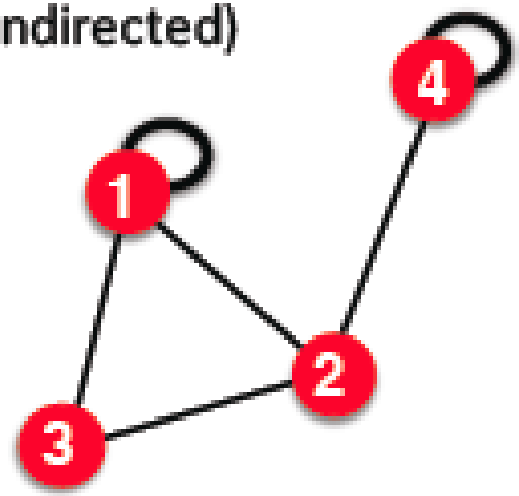
Consequence: Adjacency matrix is filled with zeros!

(**Density of the matrix (E/N^2):** WWW=1.51×10^{-5}, MSN IM = 2.27×10^{-8})

More Types of Networks

- **Self-edges (self-loops)**

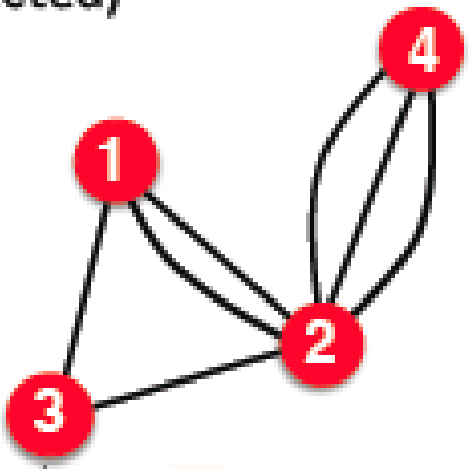
(undirected)



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

- **Multigraph**

(undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$