Outline

Day 1: Social Network analysis

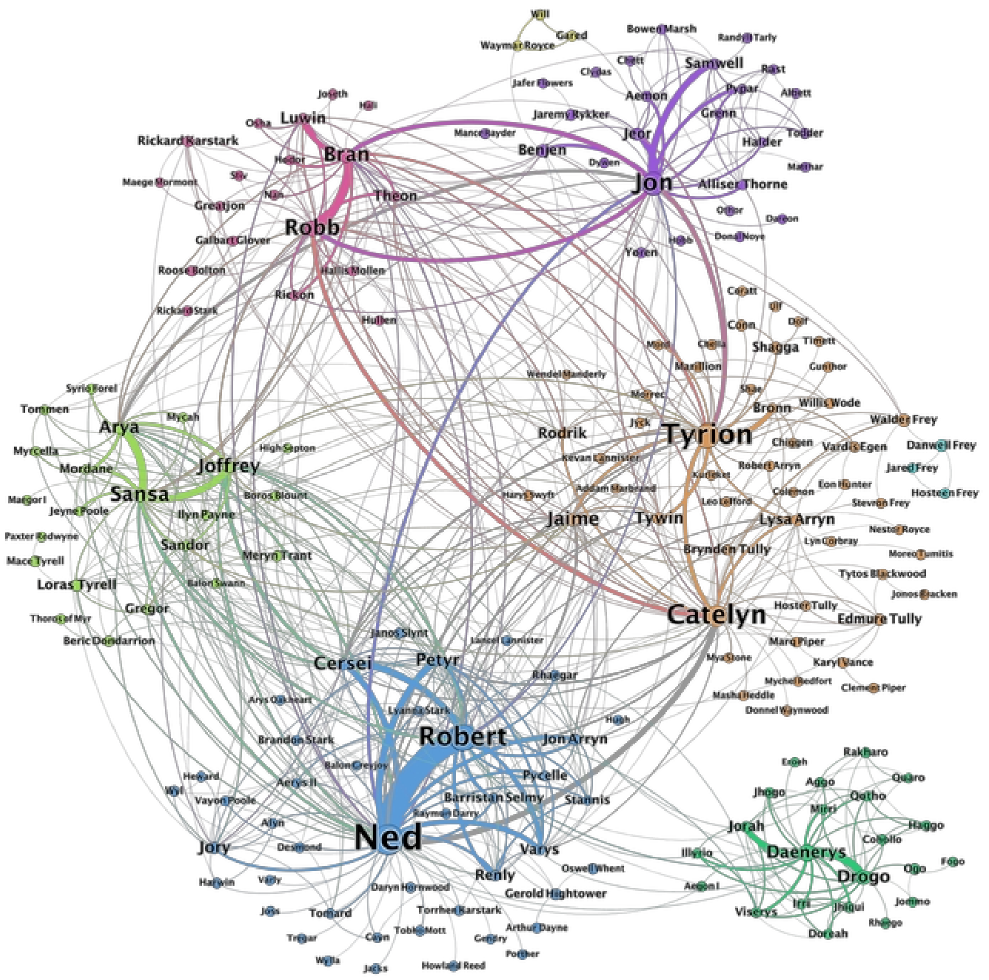
Part 1:
- **Fundamental concepts**
- **Visualization**
- **Connectivity and Cohesion**
- **Centrality**
- **Roles**

Part 2:
- **Introduction to python library networkx**
- **apply concepts from part 1**

Day 2: Social media analysis

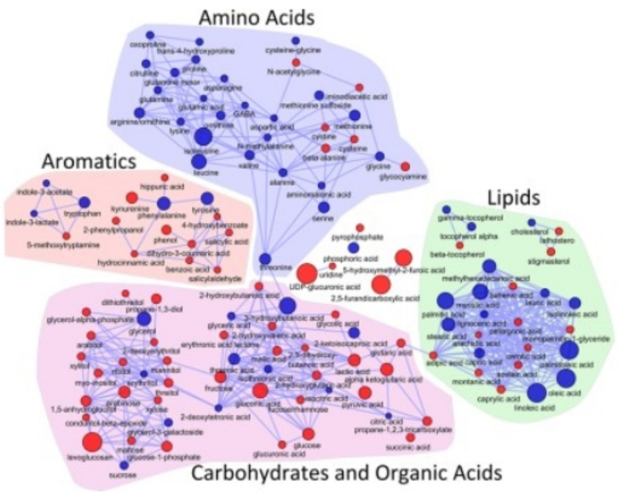
- **Twitter APIs**
- **How to get data and what does the data look like**
- **Create a small network from twitter data and analyse it**
- **simple text analysis**

Network Visualization and Theory: A ”Very” Condensed Guide to Exploratory Social Network Analysis



Nikolas Zöller and Jonathan H. Morgan

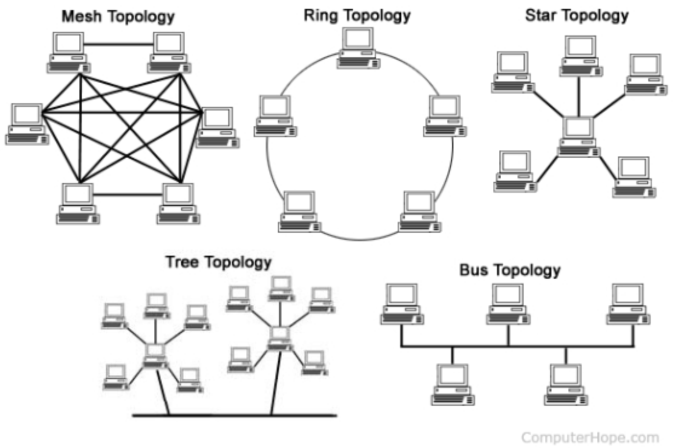
I. Main types of networks found in nature



Example of **biological networks**: Nodes represent metabolites and edges can be many things.

Source: https://www.r-bloggers.com/tutorial-building-biological-networks/

Telecommunication networks:
Collection of computers, servers, mainframes, network devices.

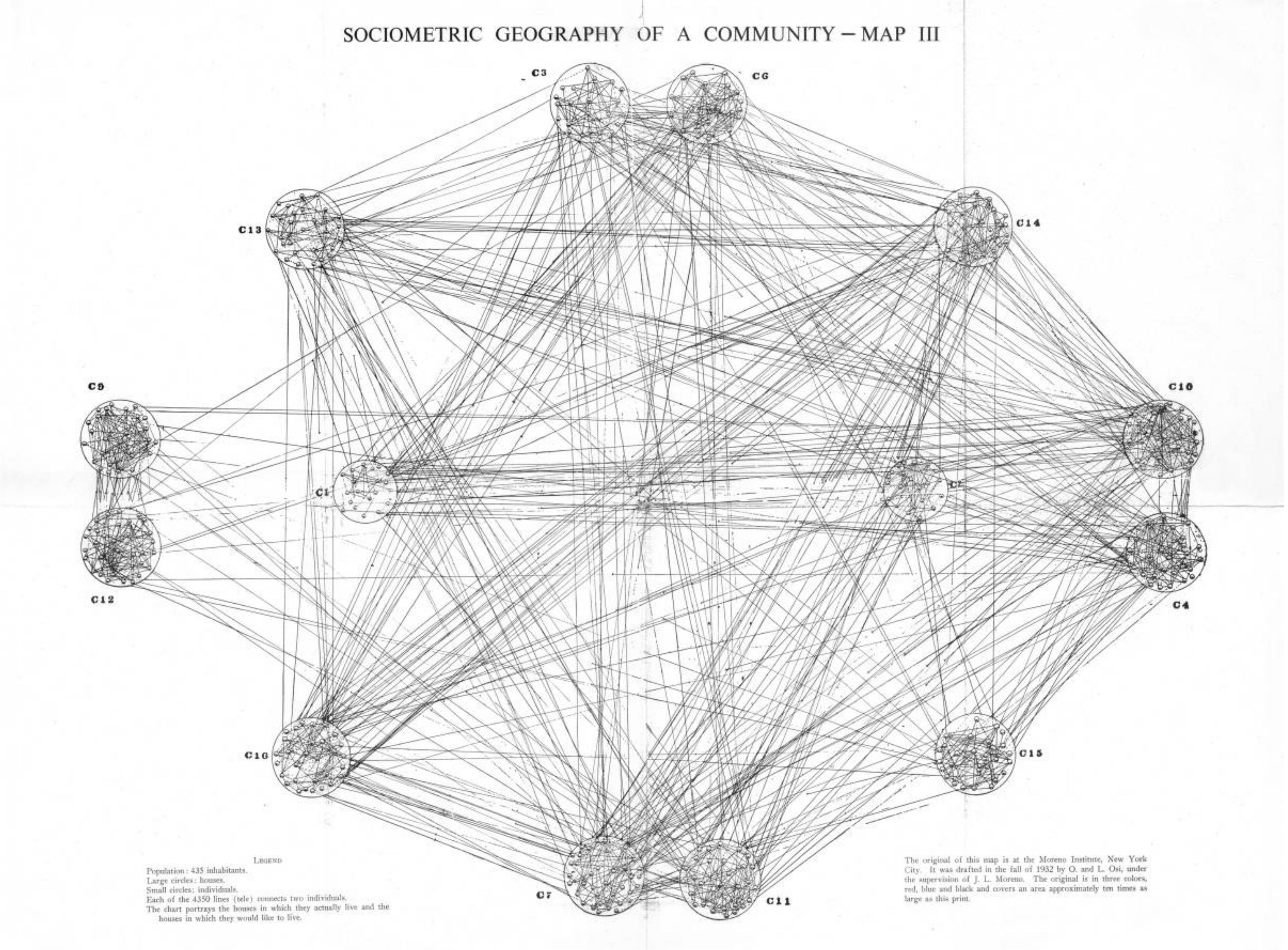


Source: https://www.computerhope.com/jargon/n/network.htm

Social networks:
Facebook, Twitter, Wikipedia, etc.

Source: https://blog.stateofthedapps.com/blockchain-based-social-networks-1b7b729beb4d

Sociometry and the Cultural Order: 1936



EMOTIONS MAPPED BY NEW GEOGRAPHY

Charts Seek to Portray the Psychological Currents of Human Relationships.

FIRST STUDIES EXHIBITED

Colored Lines Show Likes and Dislikes of Individuals and of Groups.

MANY MISFITS REVEALED

Dr. J. L. Moreno Calculates There Are 10 to 15 Million Isolated Individuals in Nation.

A new science, named psychological geography, which aims to chart the emotional currents, cross-currents and under-currents of human relationships in a community, was introduced here yesterday at the scientific exhibit of the Medical Society of the State of New York, which opens its 127th annual meeting here today at the Waldorf-Astoria.

The first series of maps of the new human geography were shown by Dr. Jacob L. Moreno of New York, consulting psychiatrist of the National Committee of Prisons and Prison Labor and director of research, New York State Training School for Girls, Hudson, N. Y. The maps represent studies of the forces of attraction and repulsion of individuals within a group toward one another and toward the group, as well as the attitude of the group as a whole toward its individual members, and of one group toward another group.

Emotions are represented on these psychological maps by various colored lines. Red stands for liking, black for disliking. If individual A likes B a red line with an arrow points from A to B. If B reciprocates a similar red line points from him. If he dislikes A this is indicated by a black line with an arrow pointing toward A. If B is merely indifferent the feeling is shown by a blue line.

Group of 500 Girls Studied.

Fundamentals

Fundamental Concepts

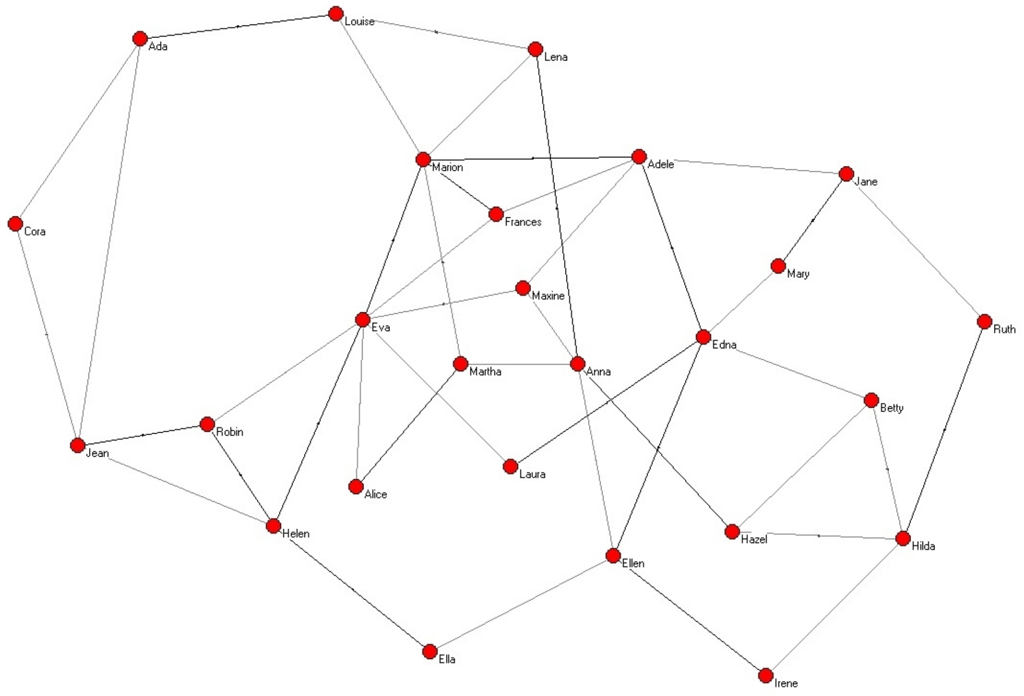
Graph: A set of ***vertices*** and a set of ***lines*** between pairs of vertices.

Key Elements of a Graph

Vertices, Nodes, or Dots

Lines, Edges, Links

Layout

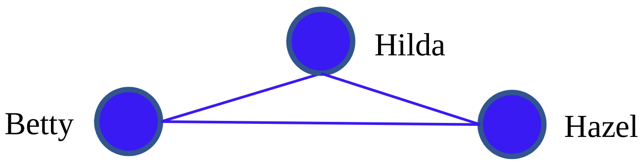
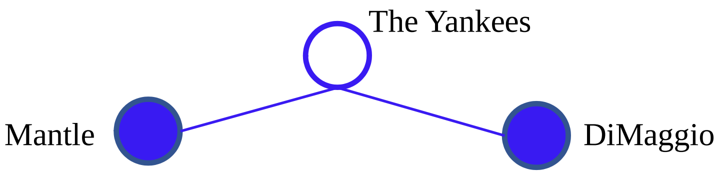


A **network** is a graph where the vertices are assigned attributes such as names (e.g., Laura or Intel), categories (e.g., messenger ribonucleic acid or sensory neurons), or continuous features such as size or weight).

Fundamental Concepts

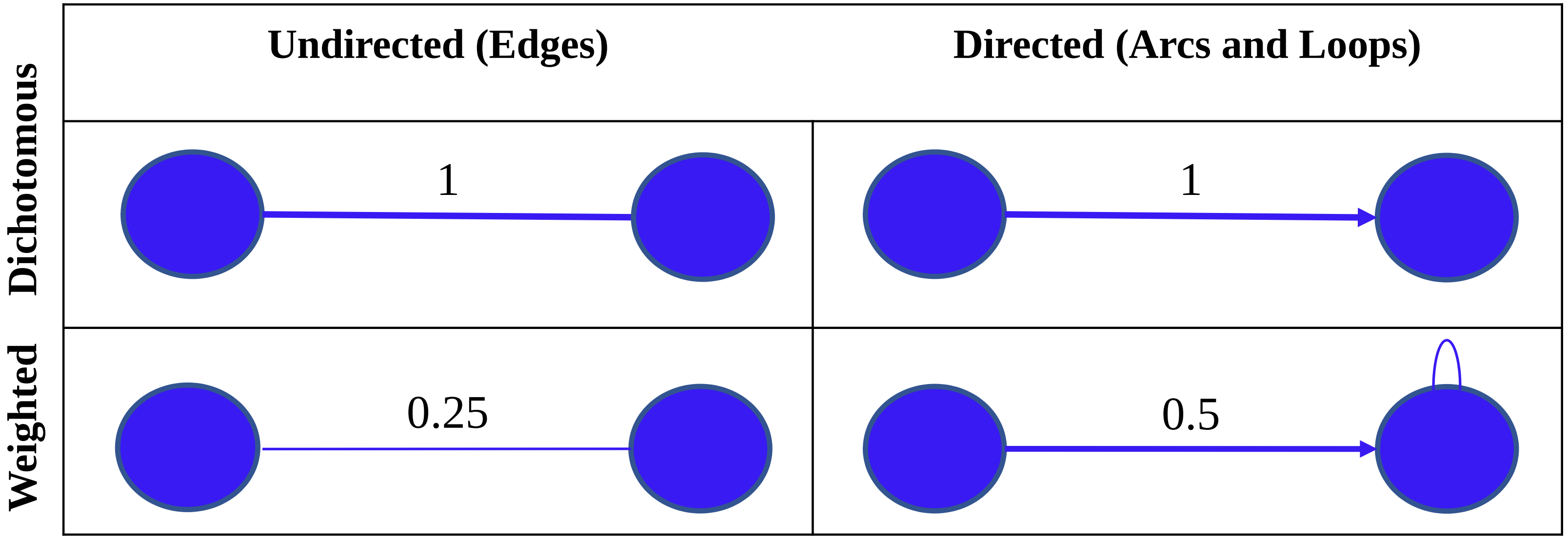
Networks can **differ** with respect to their nodes, edges, and layout

Nodes

Unimodal	Multimodal: Bi-partite
	

Fundamental Concepts

Edges: **Direction** and **Weight** (e.g., frequency or distance) provide more information



Loops refer to arcs directed to oneself. Dichotomous networks are also called binary networks.

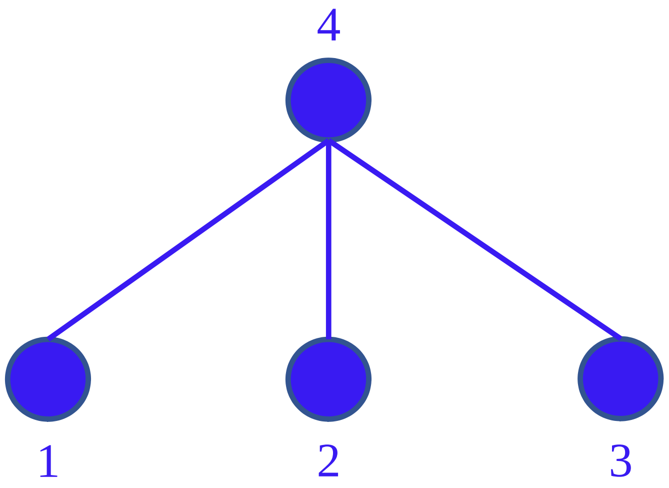
Multiplex networks refer to networks with more than one tie type (e.g., friends you ask advice from on important matters vs. friends you work with on projects).

Although some people may be connected through both types of ties; the ties are, nevertheless, different because they represent different kinds of relations.

Fundamental Concepts

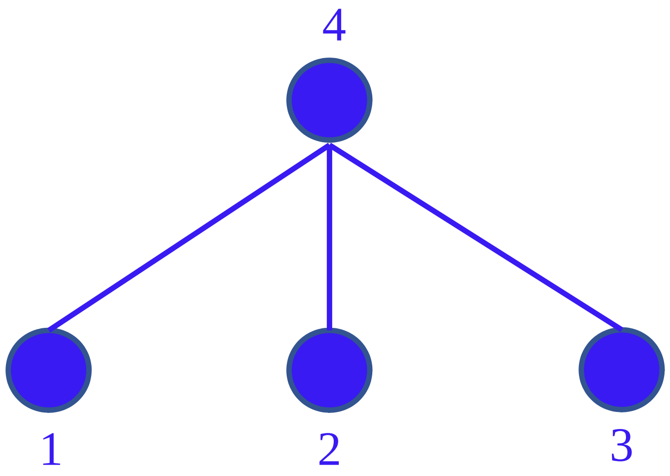
Data Structures

Adjacency Matrix



	1	**2**	**3**	**4**
1	**0**	0	0	1
2	0	**0**	0	1
3	0	0	**0**	1
4	1	1	1	**0**

Edgelist

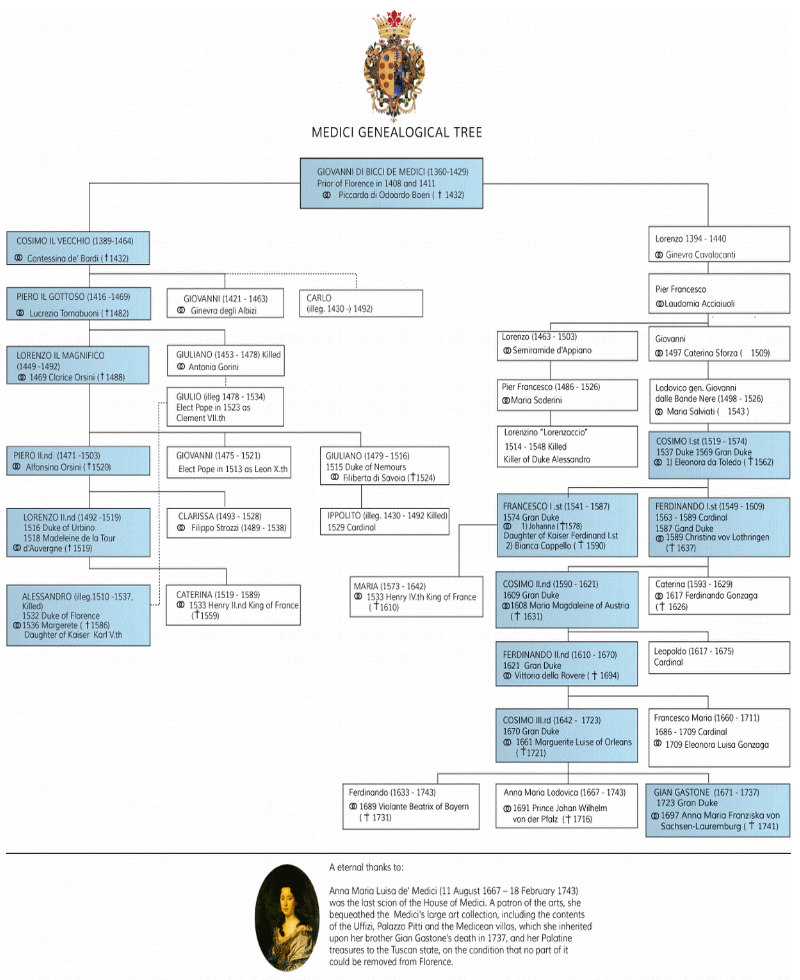
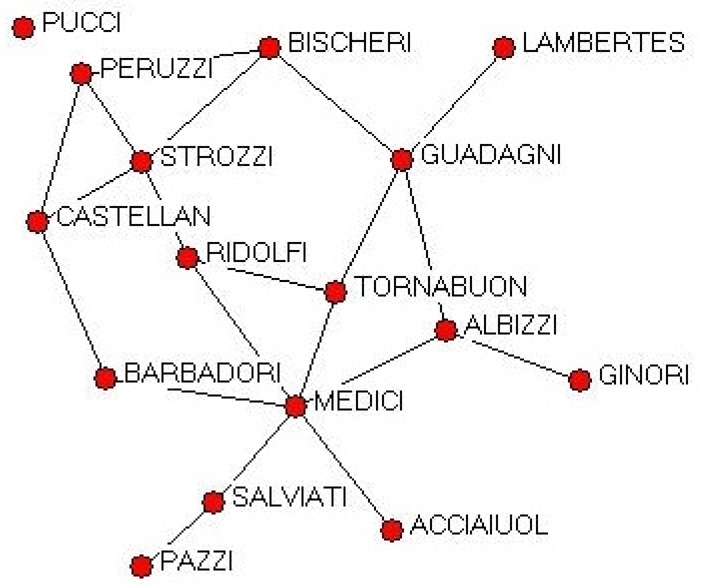


Sender	Target	Weight
1	4	1
2	4	1
3	4	1

Visualization and a Little History

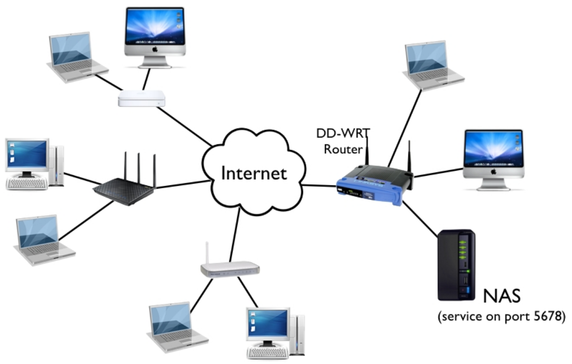
Fundamental Concepts

Layouts

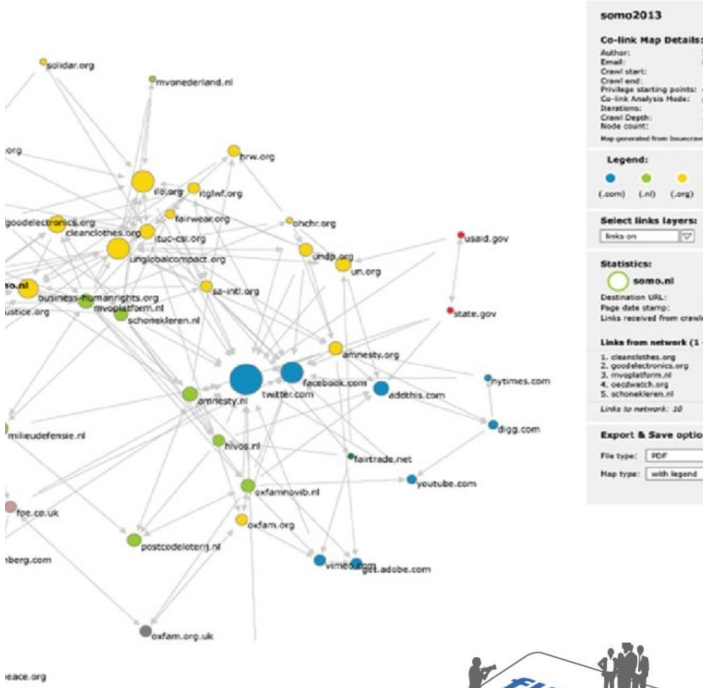
Hierarchical	Spring or Energy
	 Padgett's Florentine Marriage Alliances

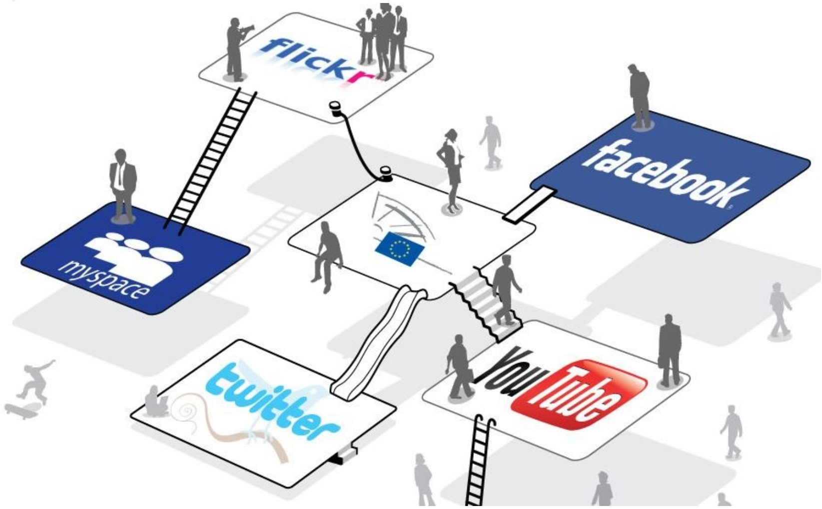
Medici Genealogical Tree: http://www.museumsinflorence.com/musei/genealogic_tree.htm

The internet infrastructure
→ geographical meaning

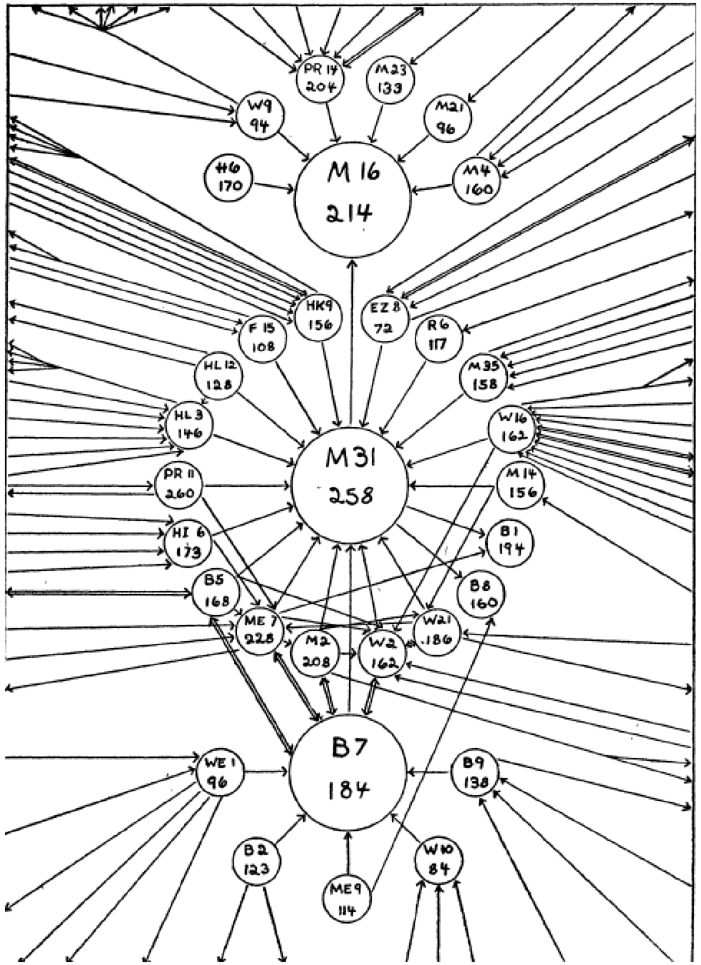


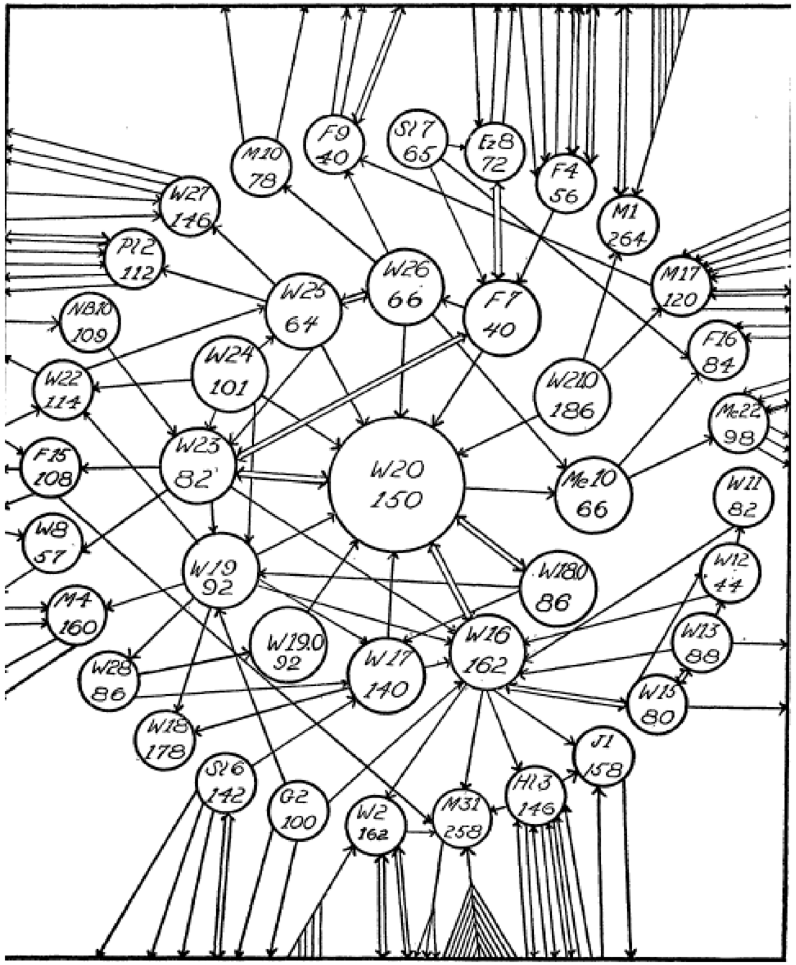
www → no geographical meaning





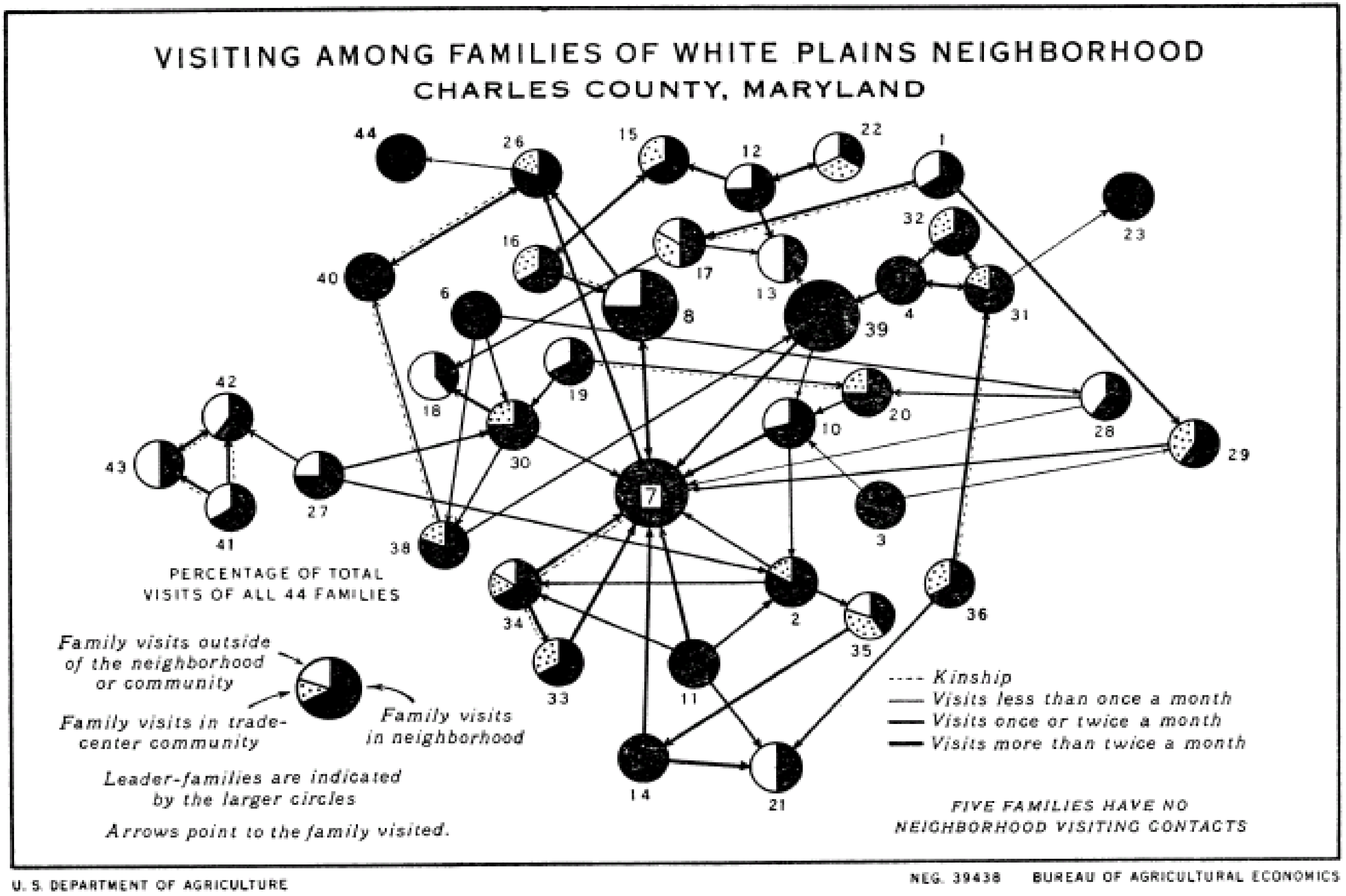
Lundberg & Steel 1938: Visualizing Directed Ties





Moreno’s idea was quickly taken up in community studies, as the social analog to “atoms” in physics….

Charles Loomis 1948: Early Example of Layered Information



Northaway’s (1940) Target Sociograms

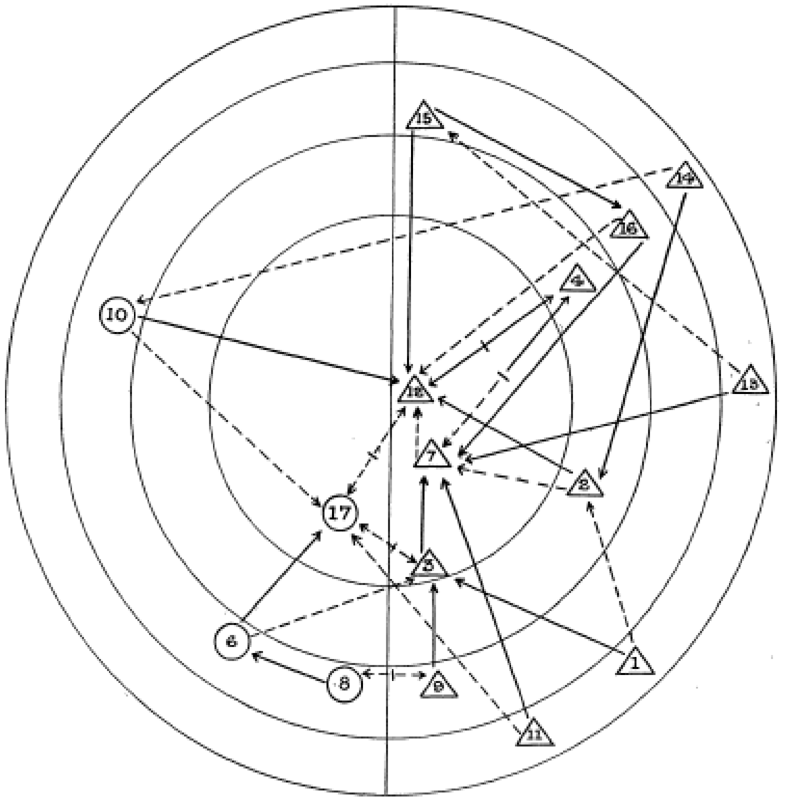


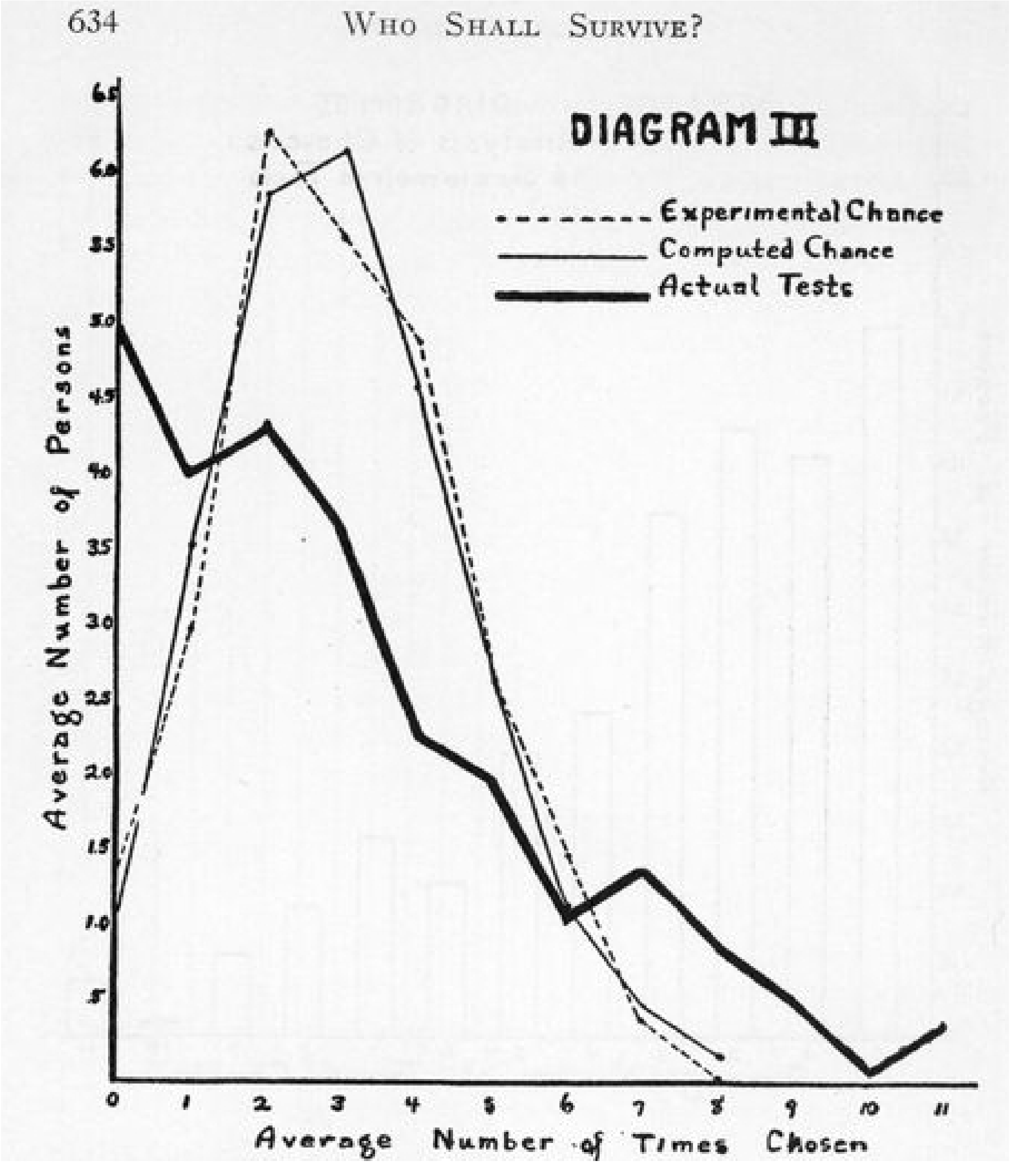
FIG. B.—SOCIOGRAM

For a Third Grade Group

Girls 4
Boys 12

Criterion:
Luncheon preference

More than Sociograms:
The First Long-Tailed Distribution Observed in a Social Network?



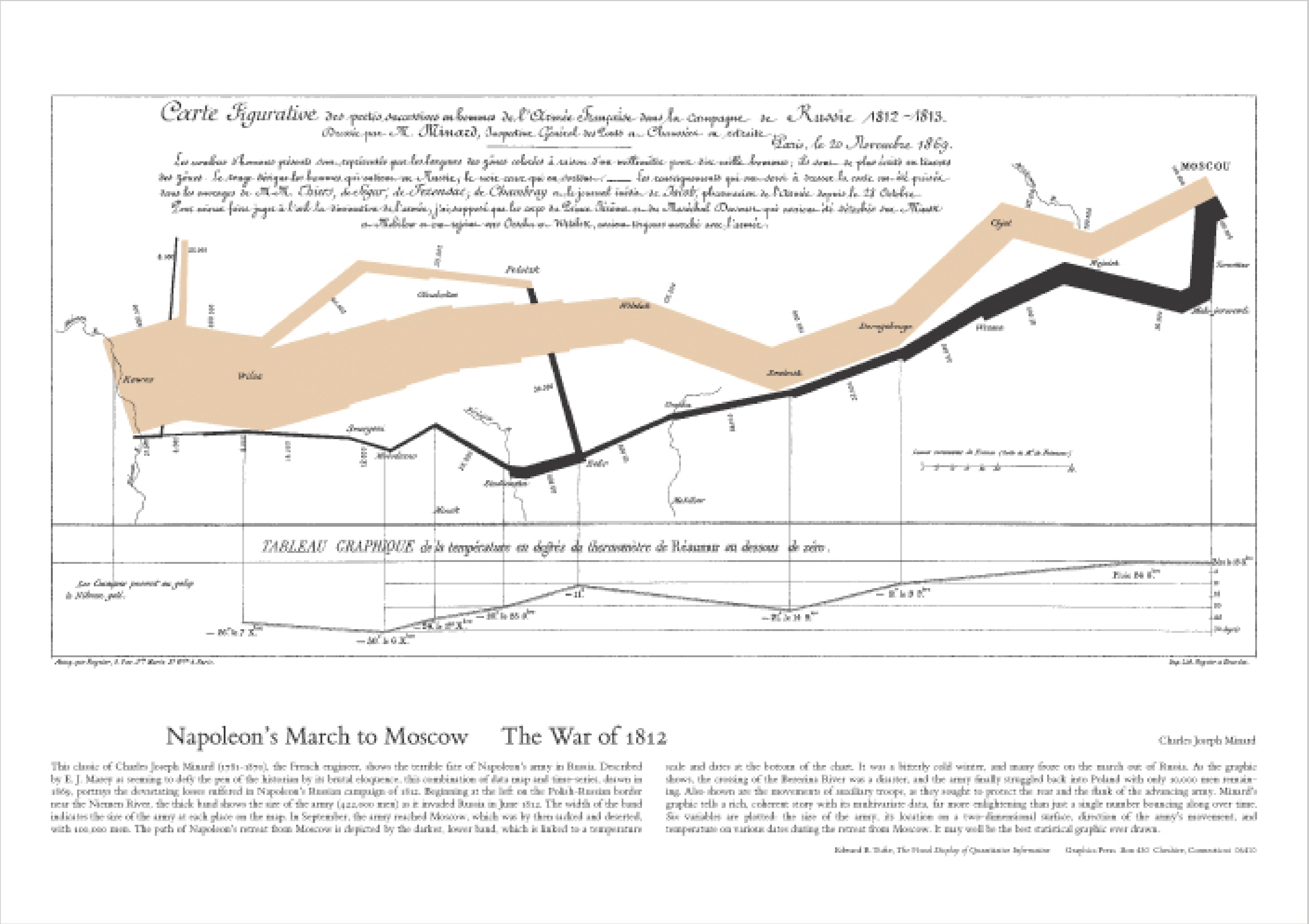
Goals to Aspire to When Visualizing a Network

Clarity

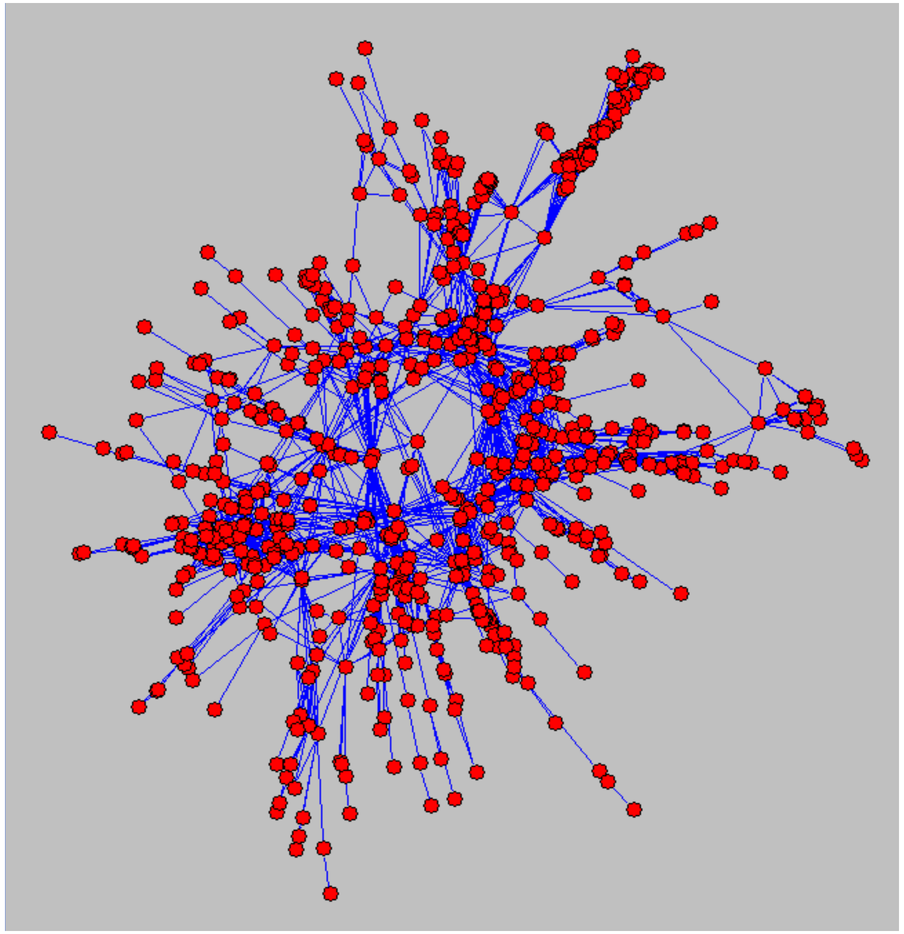
Scaling

Effective Layering information

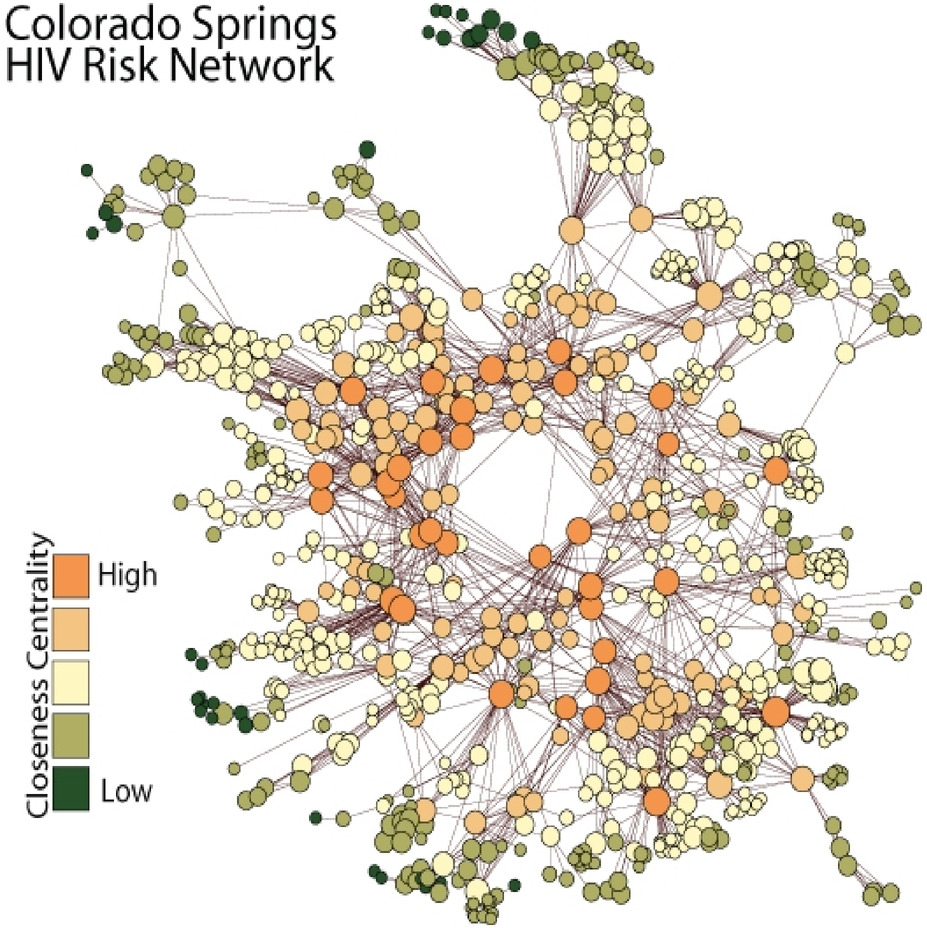
Effective use of color



Example of an Effective Sociogram



Before: Out-of-the-Box

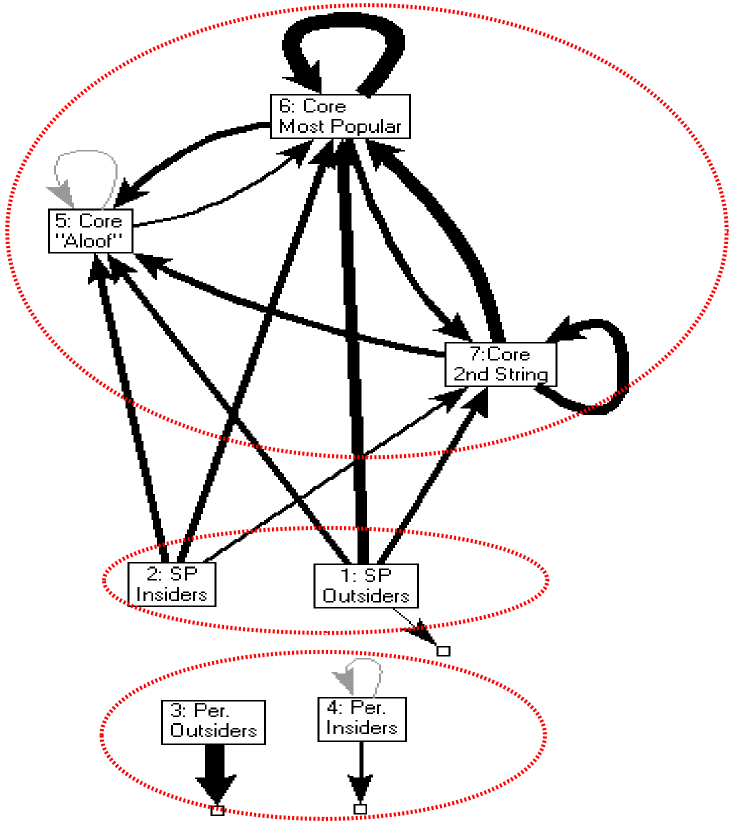


After: Moody 2007

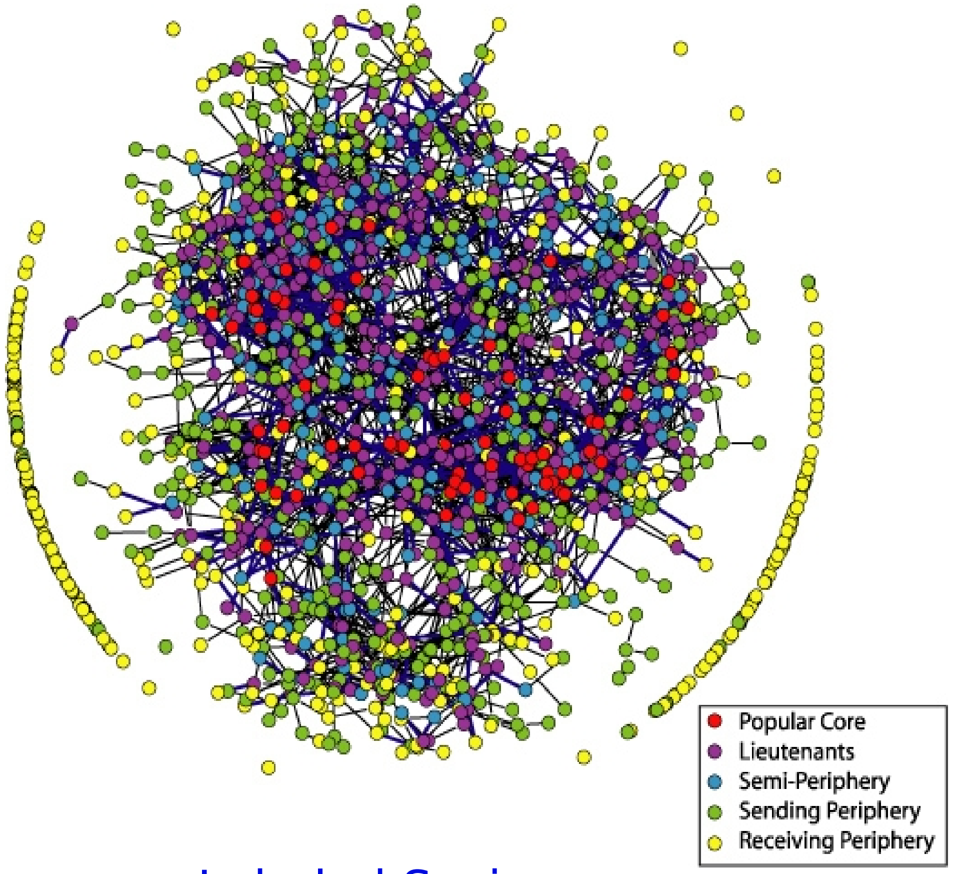
Visualization Challenges

Network Specific graphic problem: Which social space?

For example, are we trying to show the distinctiveness of the nodes or the general topology of the network?



Reduced Block Model



Labeled Sociogram

Basic Tradeoffs & Considerations

Network Visualization Challenges/Considerations

Model of Social Space

Resolution (Local or Global)

Multiple Relations

Dynamics

Density

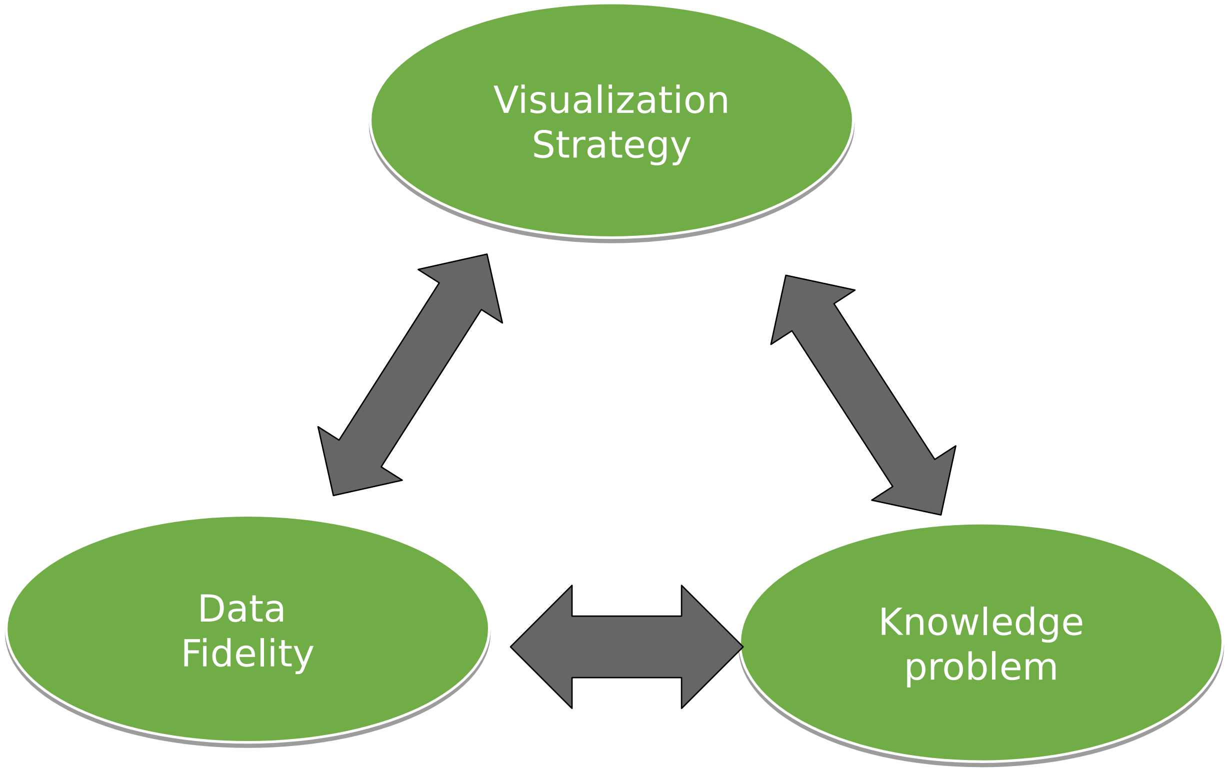
Node-Level Attributes

Edge-Level Attributes

Hierarchical Ordering

Diffusion Potential

Exploration or Communication

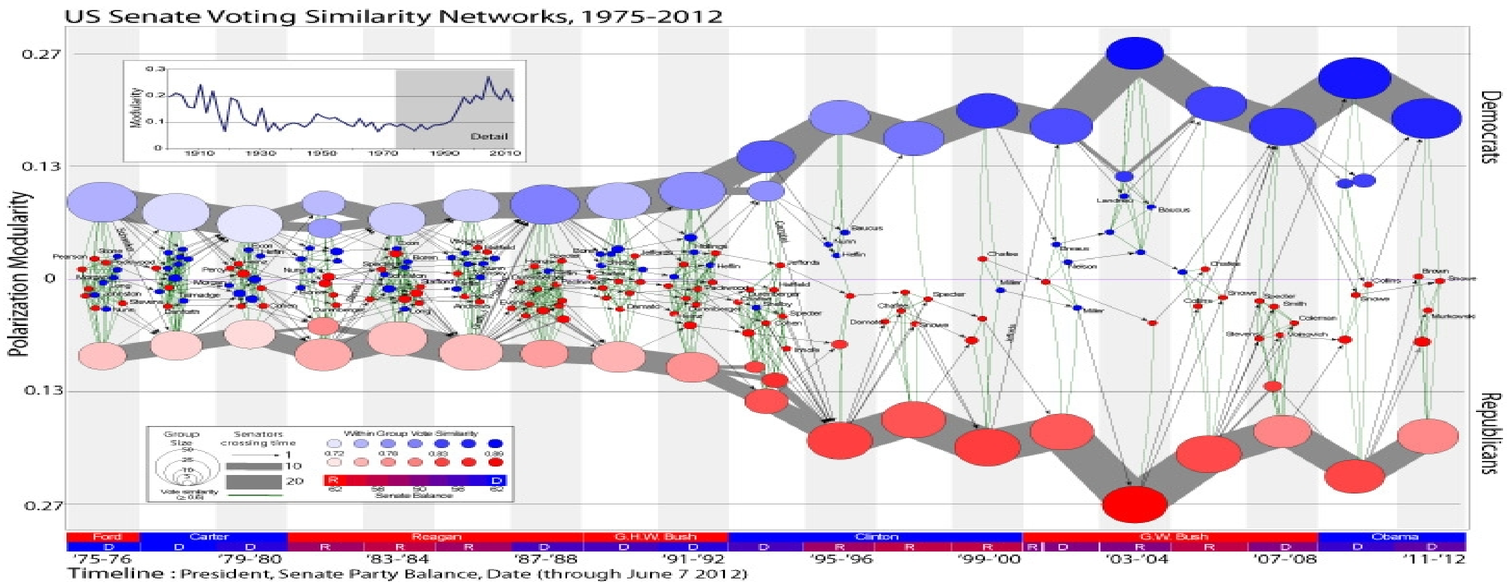


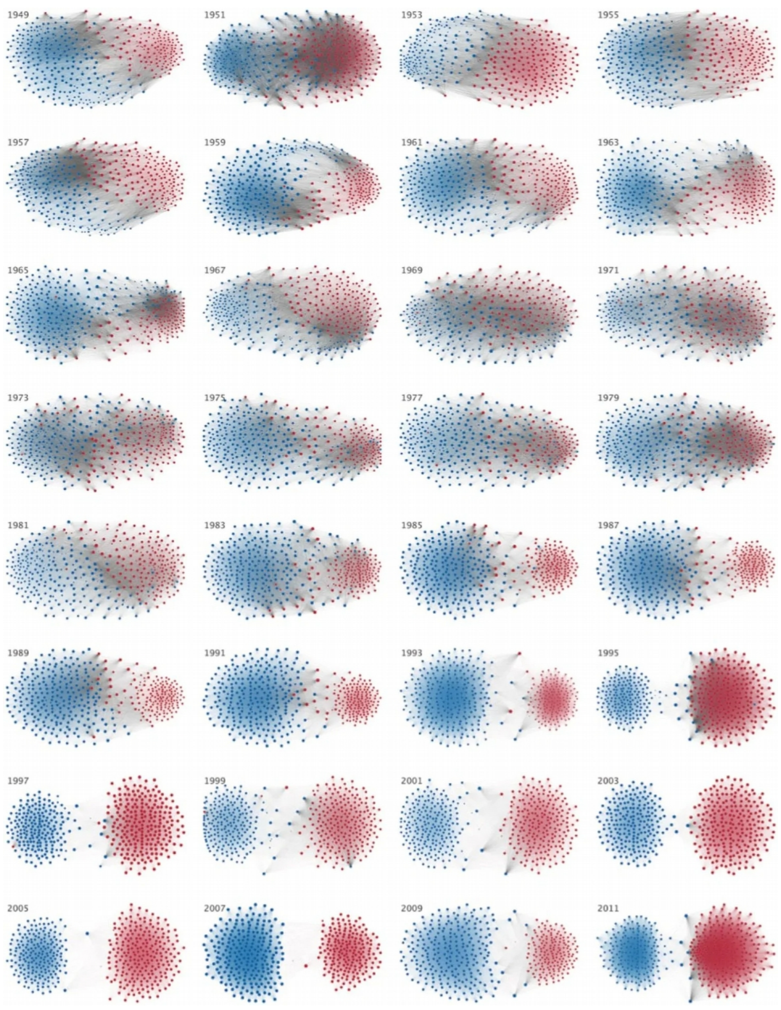
Final Goals When Visualizing a Network

Repeatable: Same data should produce the same picture

Quantification: Most features in the graph should be translatable to a measurement in the graph

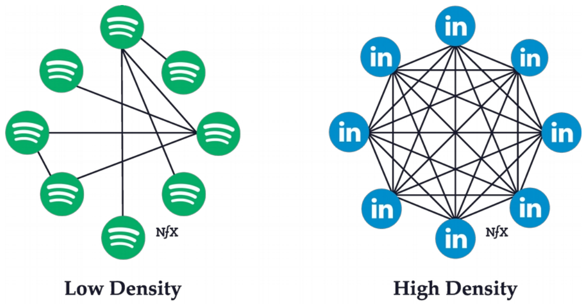
Relevant: Either give us something new to theorize about, or use the visualization a theory building tool.



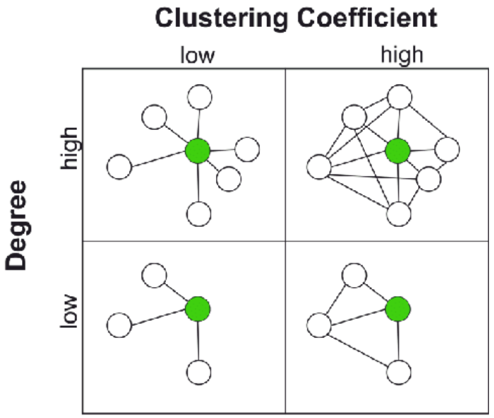


Connectivity and Cohesion

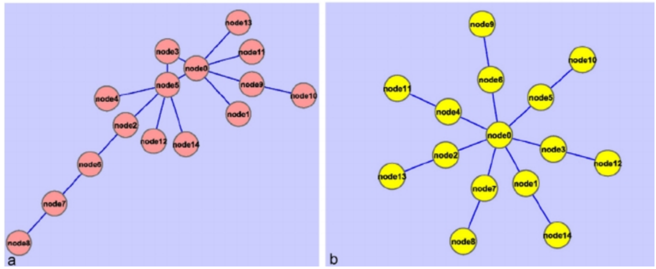
System-Level Measures



Density: The proportion of actually connected ties over all possible connections. Dense networks are highly connected networks where information, diseases, or resources can quickly move across the network.

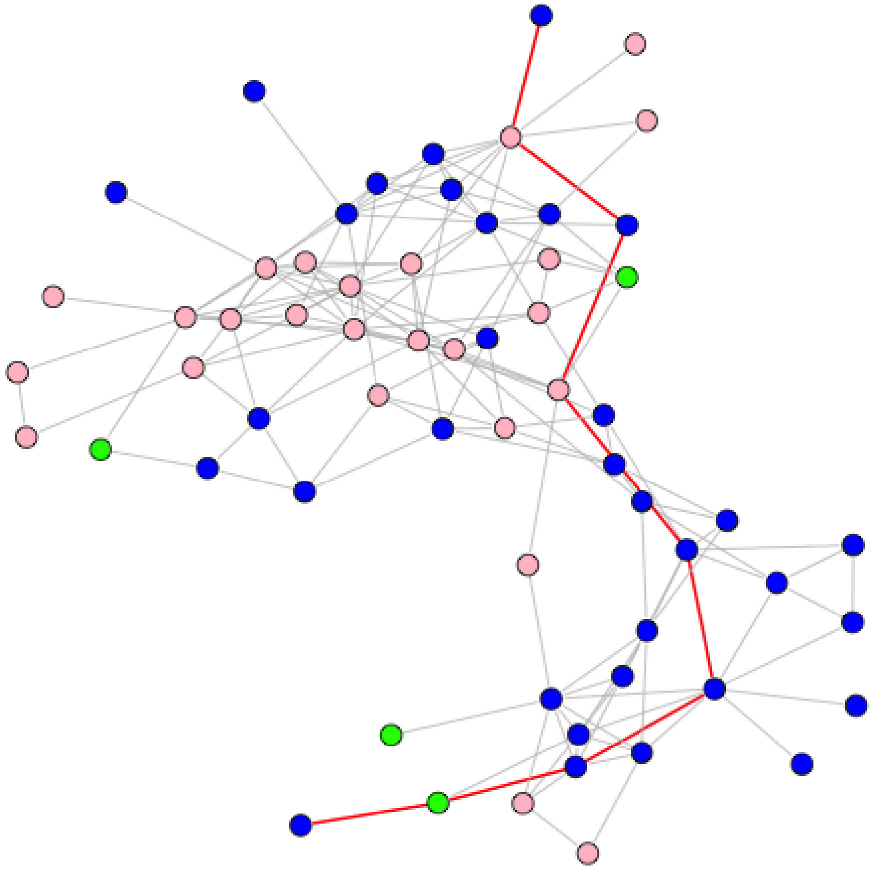


Clustering: The proportion of closed triads over the total number of triads both closed and open

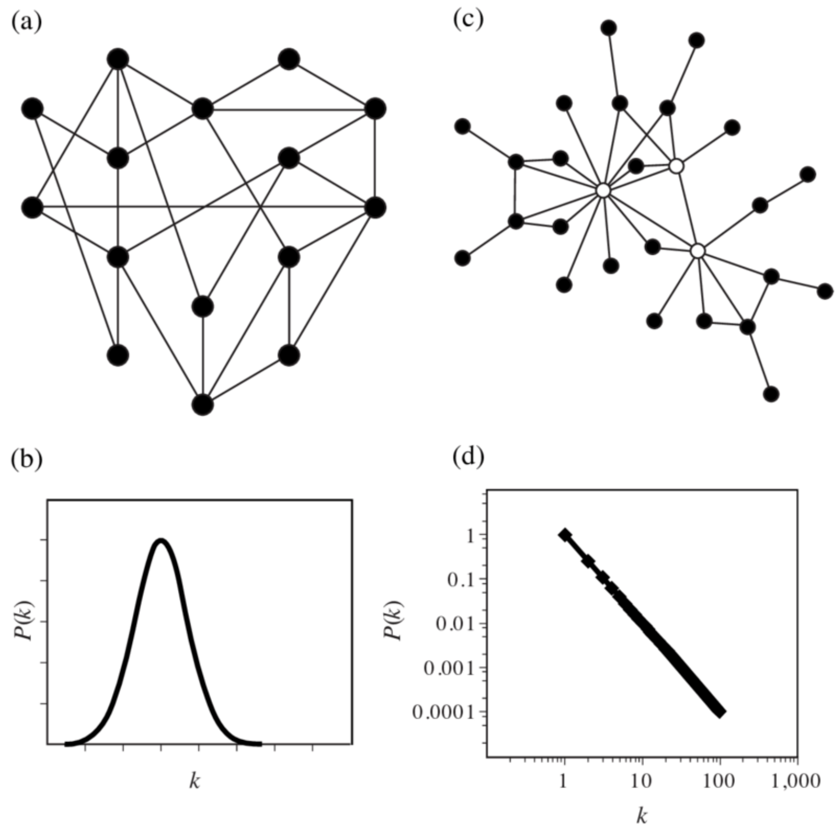


Diameter: The longest geodesic distance in the network.

Particularly Important System-Level Measures



Geodesic (Shortest-Path) Distance and Average Geodesic Distances

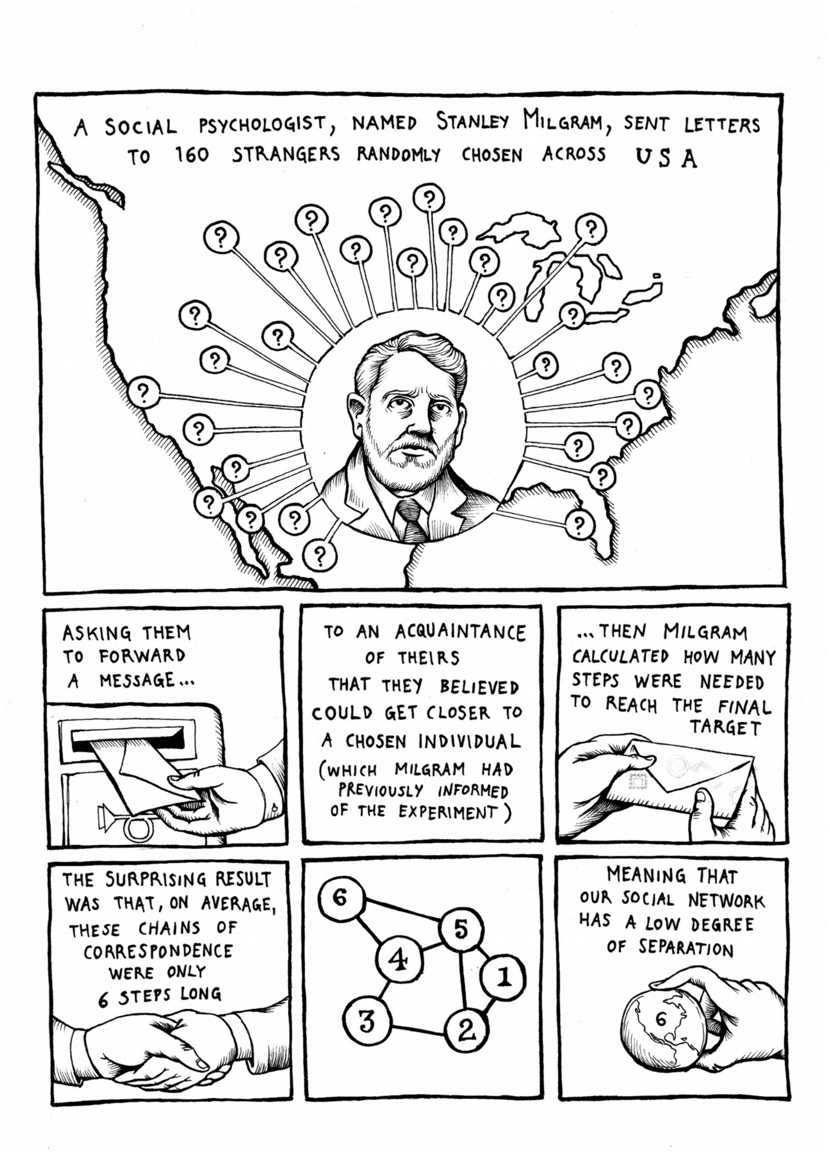


Random / Erdös-Rényi (ER) network

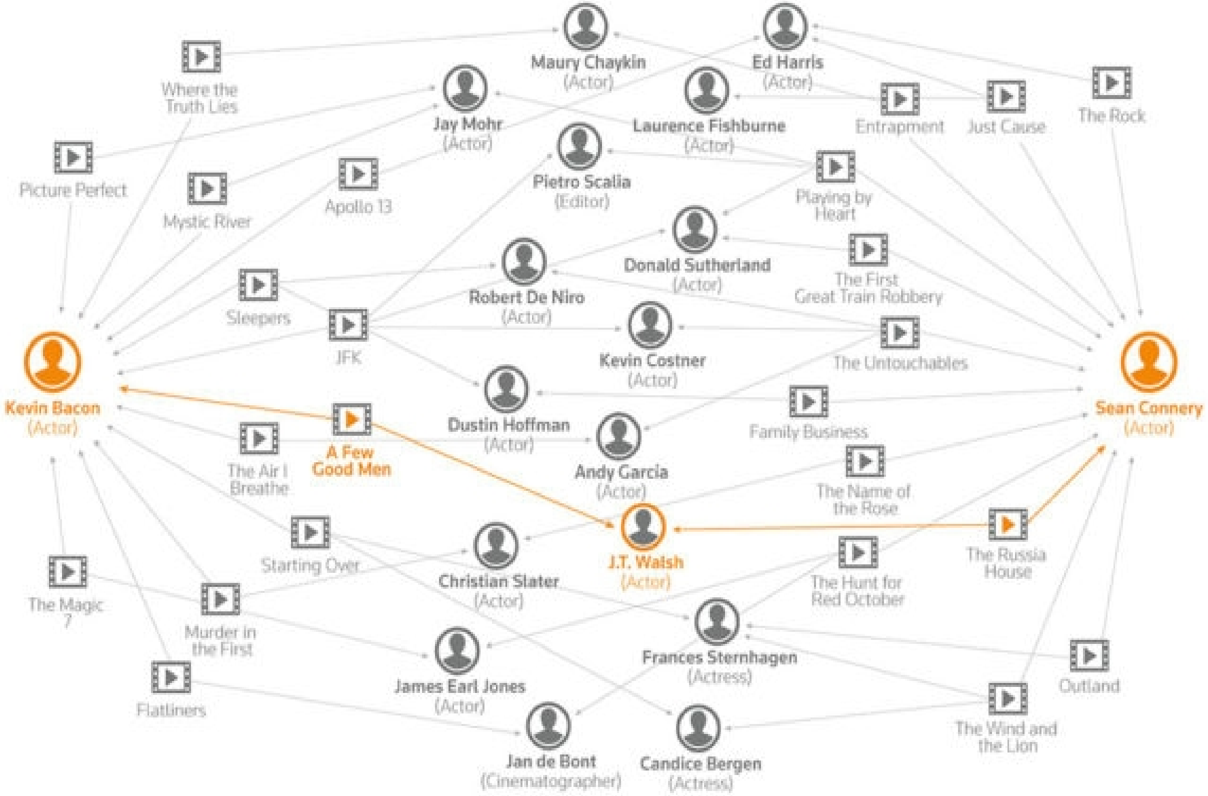
Power-law / Scale-free network

Degree Distribution

A Small-Worlds: Stanley Milgram and Kevin Bacon

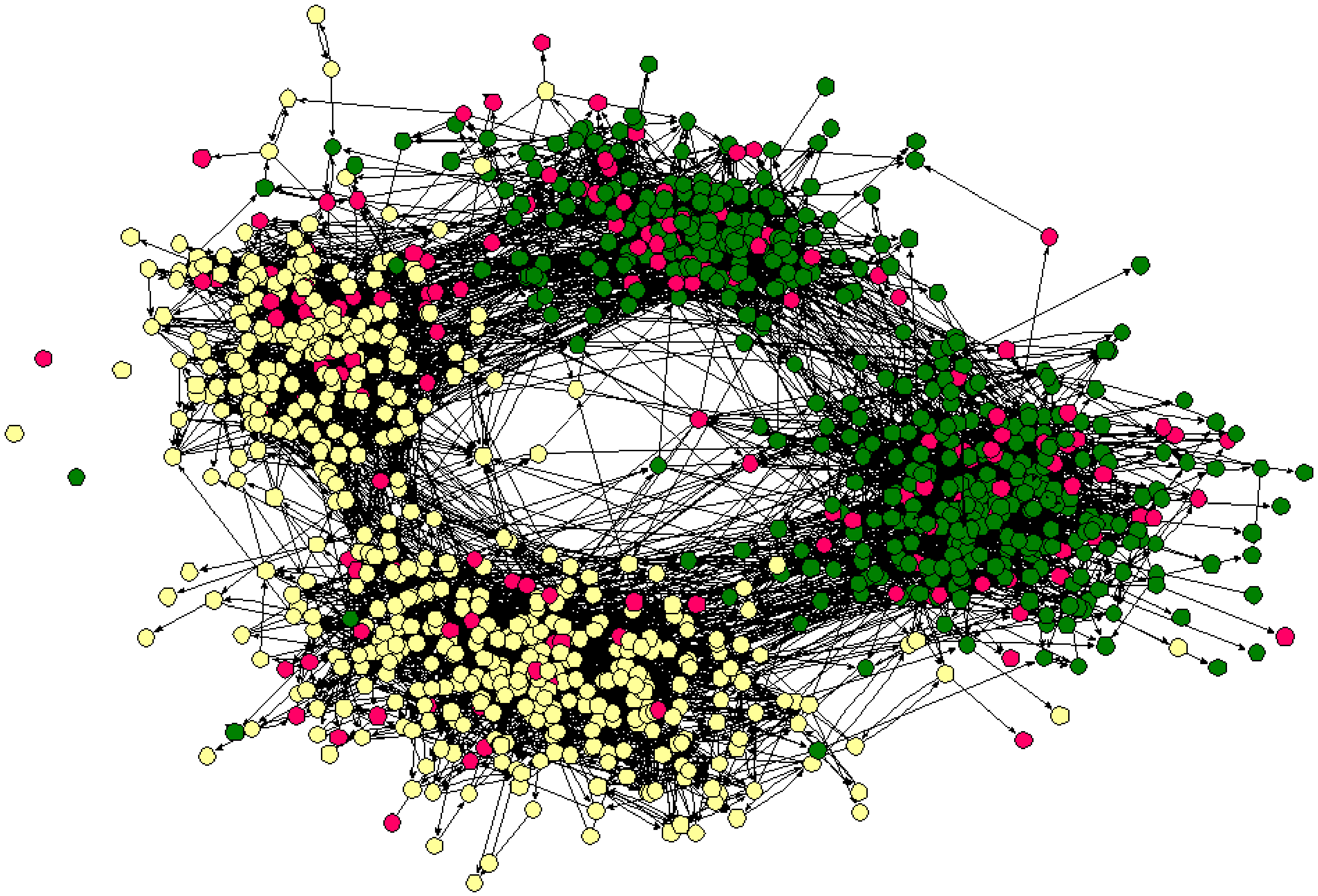


Small World: Stanley Milgram experiment



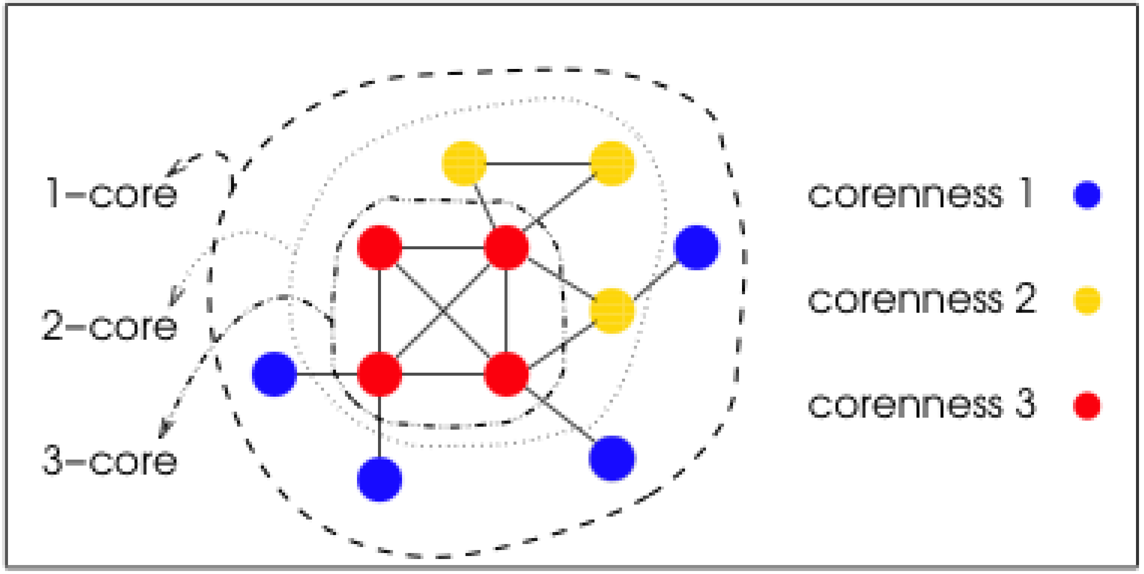
The Shortest Path from Kevin to Sean through Movies

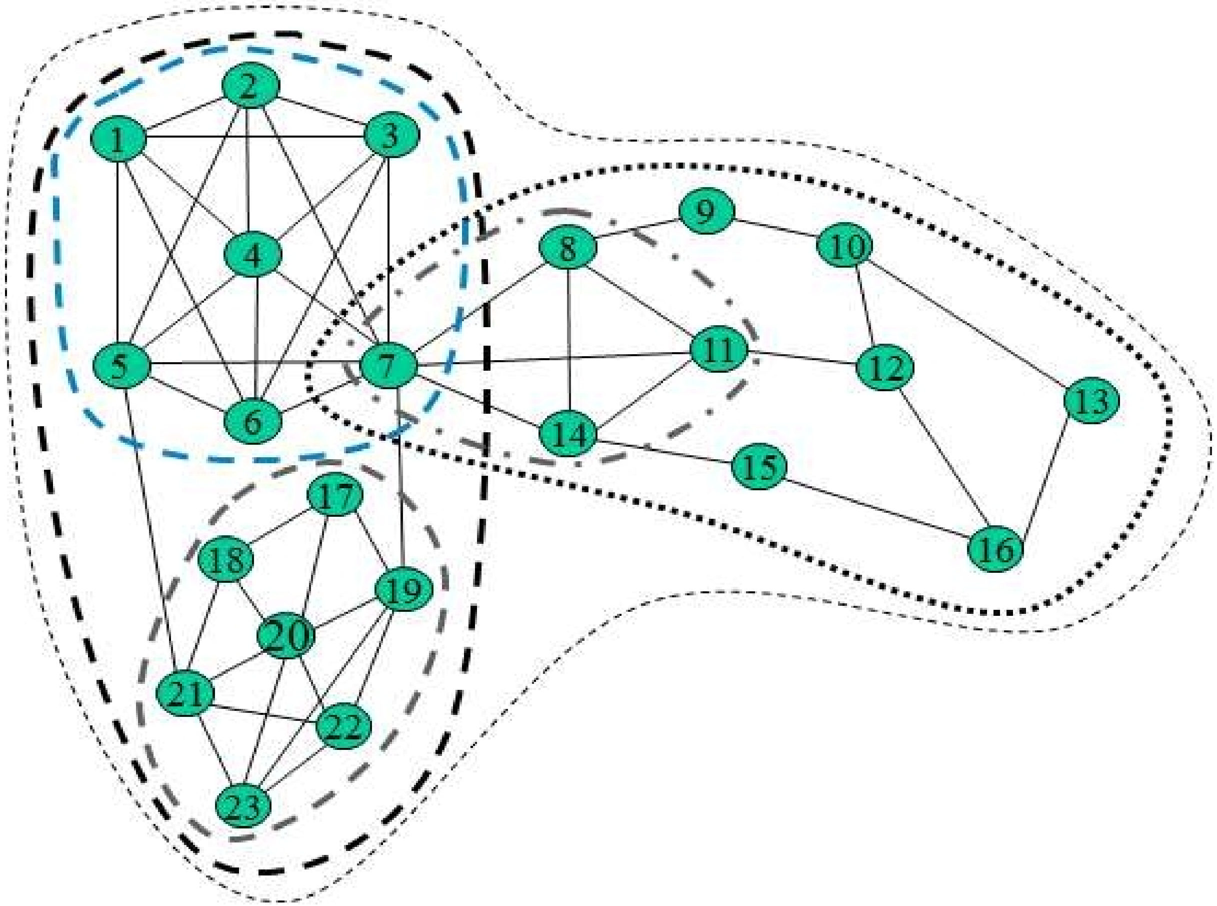
Meso-Level View: Homophily



Moody 2001: Friendship Segregation in a U.S. High School

Meso-Level View: Cohesiveness and Embeddedness

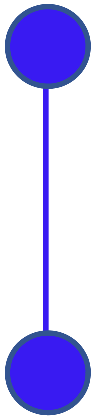




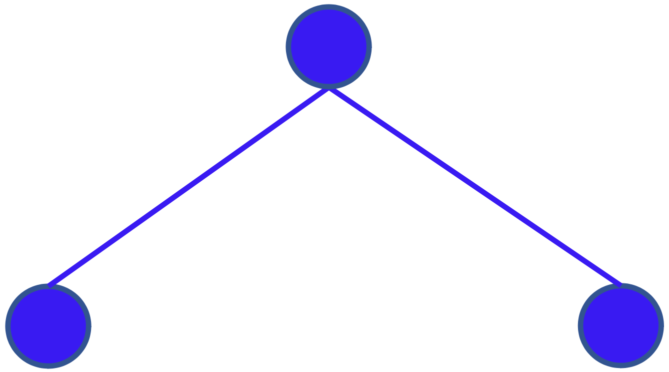
Embeddedness: Identify cohesive groups in a network, then remove k-cut sets to identify successively deeper embedded groups.

Triads: The Building Blocks of Networks

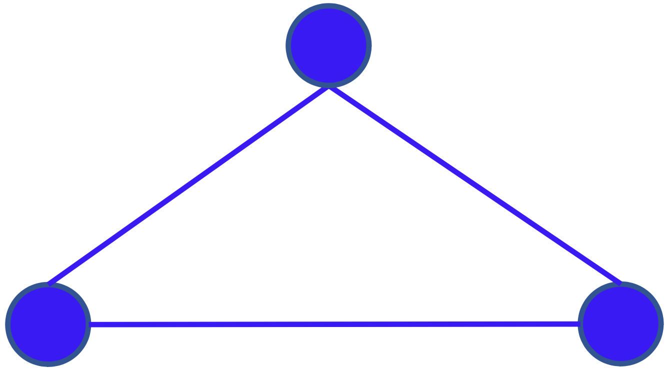
Dyad



Tryads



Open Triad

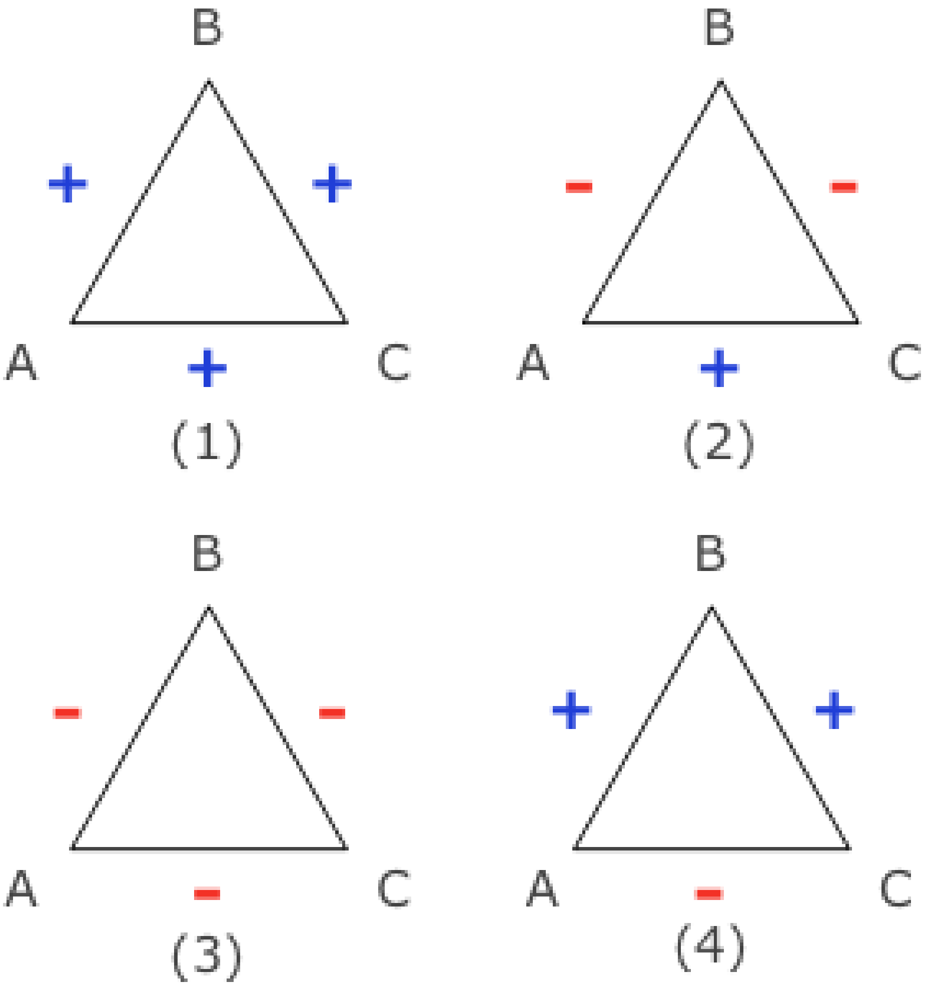


Closed Triad

Transitivity refers to the extent to which nodes making up a dyad that are also connected to a third node will both be connected to that third node: if a $\rightarrow$ b and b $\rightarrow$ c then a $\rightarrow$ c

In social networks, high transitivity can indicate social closure mechanisms that might be important such as mutualism, shared costs, a common enemy or threat, or social sanctioning.

Heider: Structural Balance Theory

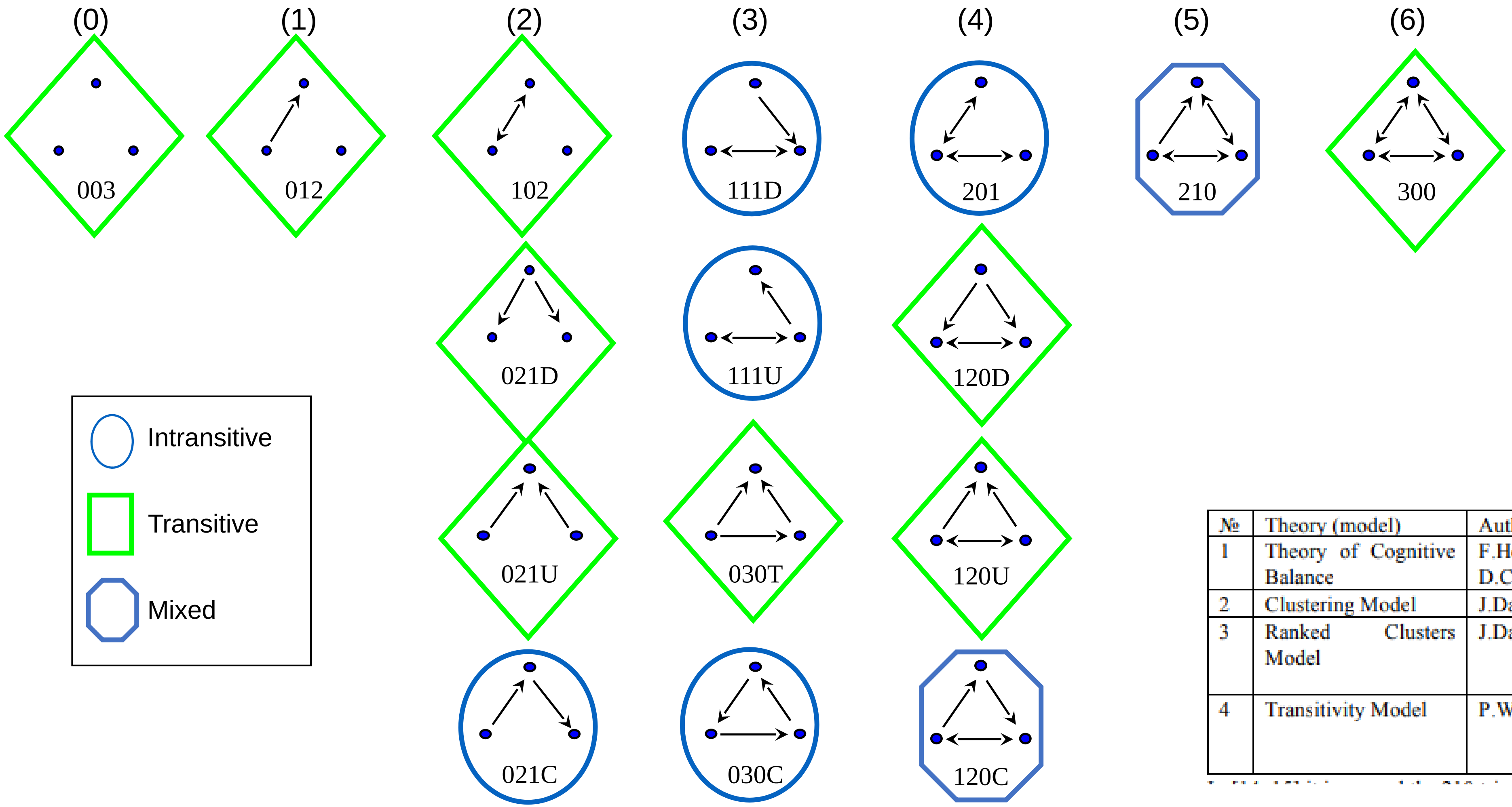


Positive ties are transitive:
The friend of a friend is a friend

Negative ties are intransitive:
The friend of a friend is a friend

→ Leads to Segregation, Polarization, Homophily

The Triad Census



№	Theory (model)	Authors	Permitted Triad Types
1	Theory of Cognitive Balance	F.Heider, D.Cartwright, F.Harary	300, 102
2	Clustering Model	J.Davis	300, 102, 003
3	Ranked Clusters Model	J.Davis,S. Leinhardt	300, 102, 003, 120D, 120U, 030T, 021D, 021U
4	Transitivity Model	P.W.Holland, S.Leinhardt	300, 102, 003, 120D, 120U, 030T, 021D, 021U, 012, 210.

Centrality

Periodic Table of Network Centrality

	1 IA	2 IIA	3 IIIA	4 IVB	5 VB	6 VIB	7 VIIB	8 VIIIB	9 VIIIB
1	8000 1979 **DC** Degree								
2	224 1971 **BC** Betweenness	239 2008 **EBC** Endpoint BC							
3	942 1966 **CC** Closeness	239 2008 **PBC** Proxy BC							
4	1279 1972 **EC** Eigenvector	239 2008 **LSBC** LscaledBC	224 1971 **EBC** Edge BC	53 2009 **CBC** Commun. BC	236 2007 **ΔC** Delta Cent.	5 2010 **MDC** MD Cent.	0 2015 **EYC** Entropy C.	2 2013 **CAC** Comm. Ability	56 2007 **EPTC** Entropy PC
5	1306 1953 **KS** Katz Status	239 2008 **DBBC** DBounded BC	979 2005 **RWBC** RWalk BC	477 1991 **TEC** Total Effects	42 2009 **LI** Lobby Index	11 2008 **MC** Mod Cent.	0 2014 **COMCC** Community C.	45 2012 **ECCoef** ECCoef	0 2015 **SMD** Super Mediat.
6	8053 1999 **PR** Page Rank	239 2008 **DSBC** DScaled BC	291 1953 **σ** Stress	477 1991 **IEC** Immediate Eff.	1 2014 **DM** Degree Mass	10 2012 **LAPC** Laplacian C.	0 2012 **ABC** Attentive BC	1699 2001 **STRC** Straightness C	0 2015 **SNR** Silent Node R.
7	484 2005 **SC** Subgraph	613 1991 **FBC** Flow BC	14 2012 **RLBC** RLimited BC	477 1991 **MEC** Mediative Eff.	69 2010 **LEVC** Leverage Cent.	35 2010 **TC** Topological C.	X X **SDC** Sphere Degree	15 2010 **ZC** Zonal Cent.	14 2013 **CI** Collab. Index

	10 VIIIB	11 IB	12 IIB	13 IIIA	14 IVA	15 VA	16 VIA	17 VIIA	18 VIIIA
1									518 1989 **IC** Information C
2				26 1989 **kPC** kPath C.	275 2002 **EGO** Ego	51 2004 **HYPER** Hypergraphs	279 1997 **AFF** Affiliation C.	399 2 001 **α-C** α-Cent.	178 1995 **ECC** Eccentricity
3				9068 1999 **HITS** Hubs/Authority	573 2006 **g-kPC** geodesic kPath	296 1999 **GROUP** Groups/Classes	80 2006 **HYPSC** Hyperg. SC	34 2010 **t-SC** t-Subgraph	116 1998 **RAD** Radiality
4	281 1971 **CCoef** Clust. Coef.	42 2012 **PeC** PeC	427 2007 **BN** Bottleneck	43 2009 **EI** Essentiality I.	573 2006 **e-kPC** e-disjoint kPC	573 2006 **v-kPC** v-disjoint kPC	505 2010 **WEIGHT** Weighted C.	17 2013 **TCom** Total Comm.	116 1998 **INT** Integration
5	1 2014 **UCC** United Comp.	4 2012 **WDC** WDC	119 2008 **MNC** MNC	43 2009 **KL** Clique Level	179 2005 **BIP** Bipartivity	426 1988 **GPI** GPI Power	116 1991 **kRPC** Reachability	58 2007 **SCodd** odd Subgraph	586 2004 **RWCC** RWalk CC
6	15 2011 **HPC** Harm. Prot.	26 2011 **LAC** Local Average	119 2008 **DMNC** DMNC	3 2013 **LR** Lurker Rank	2457 1987 **β-C** β Cent.	X X **HYP** Hyperbolic C.	27 2012 **kEPC** k-edge PC	13 2007 **FC** Functional C.	0 2014 **HCC** Hierar. CC
7	11 2013 **CoEWC** CoEWC	45 2012 **NC** NC	108 2010 **MLC** Moduland C.	X X **RSC** Resolvent SC	1 2014 **SWIPD** SWIPD	36 2009 **XXXX** LinComb	0 2014 **BCPR** BCPR	0 2014 **TPC** Tunable PC	0 2015 **EDCC** Effective Dist.

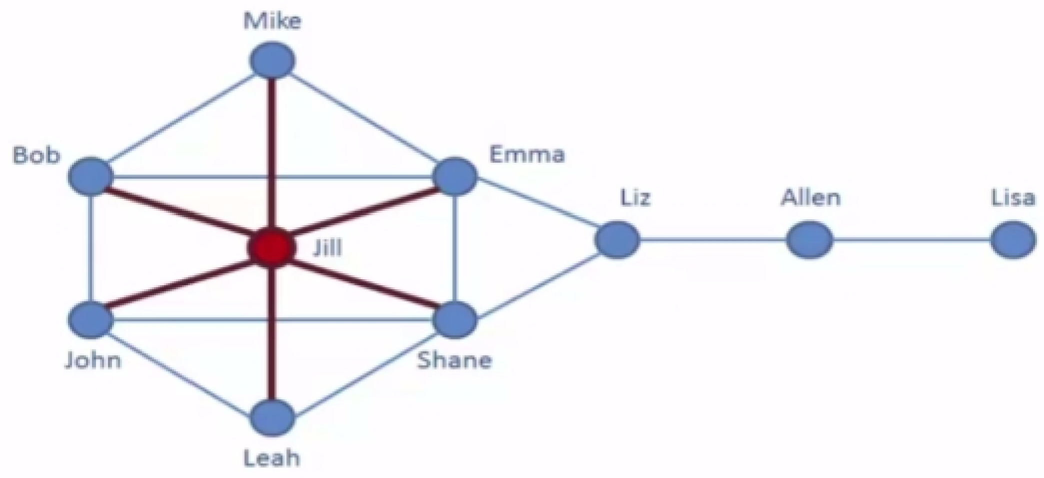
citations	year
C	
Name	

8000 1979 **Freeman** Conceptual	942 1966 **Sabidussi** Axiomatic	573 2006 **Borgatti/Everett** Conceptual	1130 2005 **Borgatti** Conceptual	24 2014 **Boldi/Vigna** Axiomatic	252 1974 **Nieminen** Axiomatic	6 1981 **Kishi** Axiomatic	3 2012 **Kitti** Axiomatic	3 2009 **Garg** Axiomatic
2065 1934 **Moreno** Historic	1546 1950 **Bavelas** Historic	780 1948 **Bavelas** Historic	1475 1951 **Leavitt** Historic	297 1992 **Borgatti/Everett** Conceptual	3649 2001 **Jeong et al.** Empirical	4167 1998 **Tsai/Ghoshal** Empirical	961 1993 **Ibarra** Empirical	71 2008 **Valente** Empirical

- "Traditional"
- Betweenness-like
- Friedkin Measures
- Miscellaneous
- Path-based
- Specific Network Type
- Spectral-based
- Closeness-like

Common Measures of Centrality

Degree Centrality: A node's degree centrality refers to the number of ties it has.

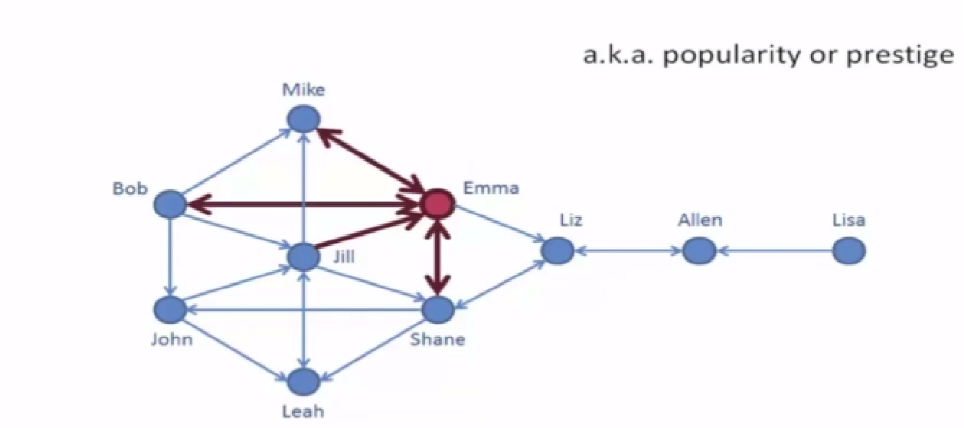


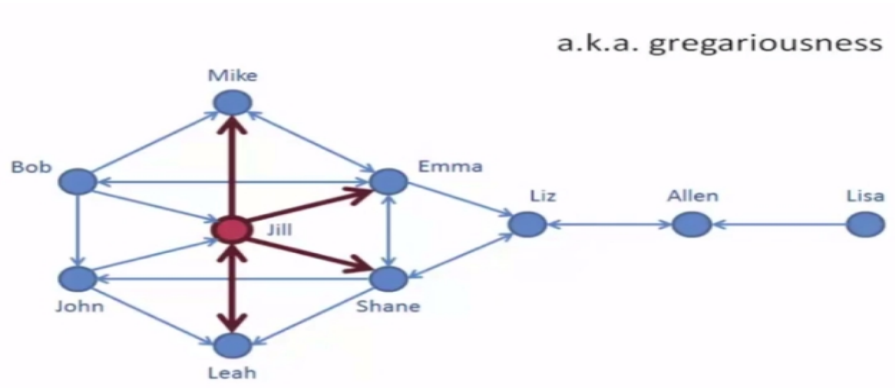
Figures From: https://friendsarena.wordpress.com/category/uncategorized/

Common Measures of Centrality

In-Degree Centrality: A node's in-degree centrality refers to the number of ties it receives.

Out-Degree Centrality: A node's out-degree centrality refers to the number of ties it sends.



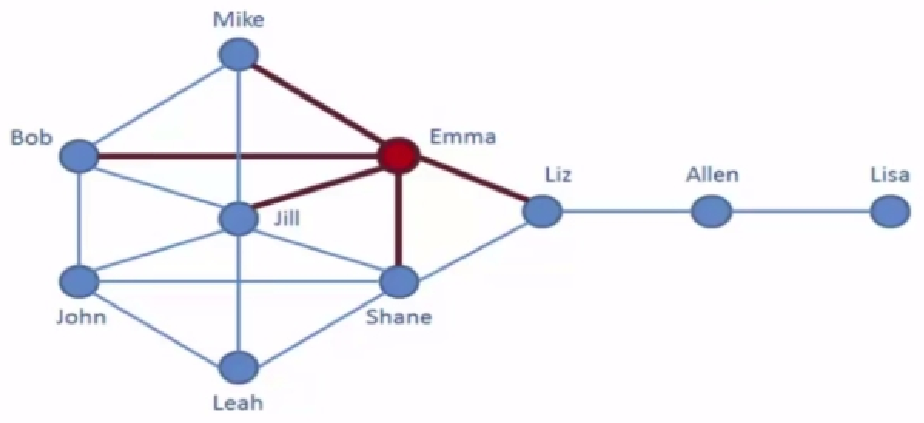


Note: To calculate these measures, you need to have a directed network.

Figures From: https://friendsarena.wordpress.com/category/uncategorized/

Common Measures of Centrality

Closeness Centrality: The closeness centrality of a node is the average length of the shortest path between the node and all other nodes. The most central the node the closer it is to all the other nodes.



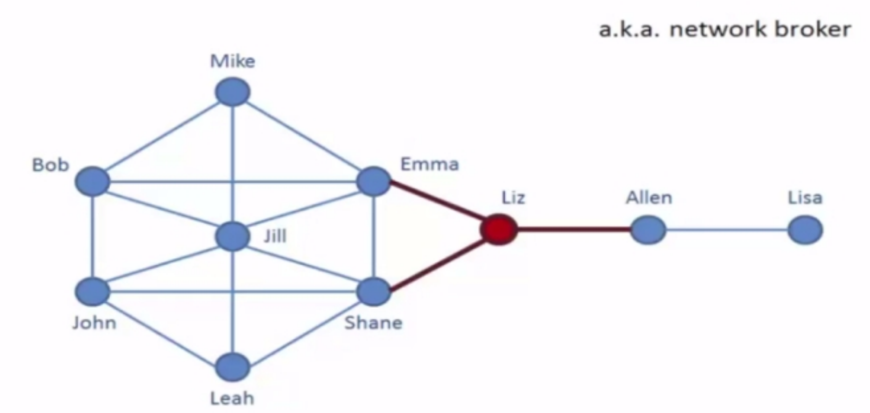
Note: Shane's closeness centrality is equivalent to Emma's.

Closeness centrality like degree centrality can be calculated in terms of closest ***to*** (incoming ties) and ***from*** (outgoing ties) in directed networks.

Figures From: https://friendsarena.wordpress.com/category/uncategorized/

Common Measures of Centrality

Betweeness Centrality: A node's betweeness centrality is the number of times a node acts as a bridge along the shortest path between two other nodes.



Betweeness Centrality is calculated by:

1. For each pair of nodes, compute the shortest paths between them.

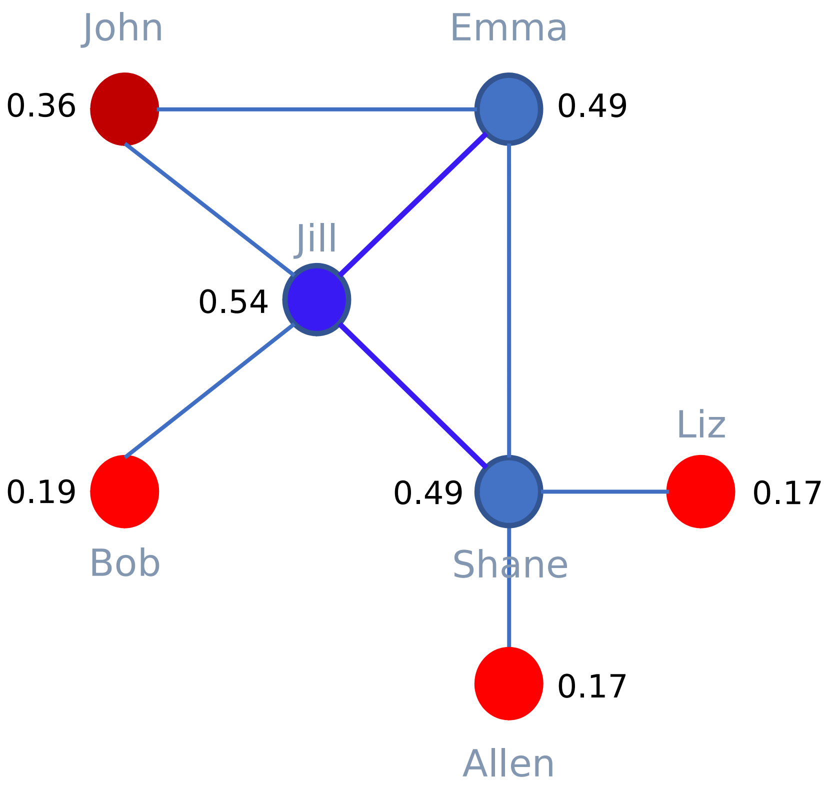
2. For each pair of nodes, determine the fraction of shortest paths through the node in question.

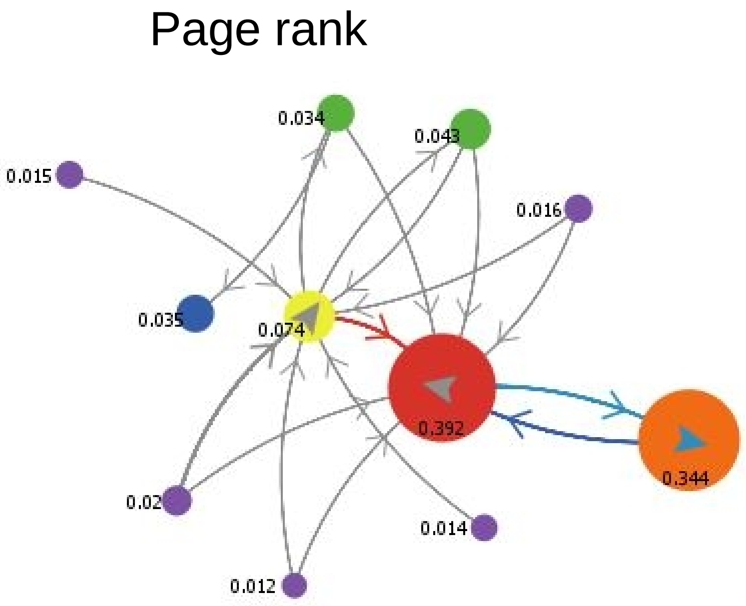
3. Sum this fraction over all pairs of nodes.

Figures From: https://friendsarena.wordpress.com/category/uncategorized/

Common Measures of Centrality

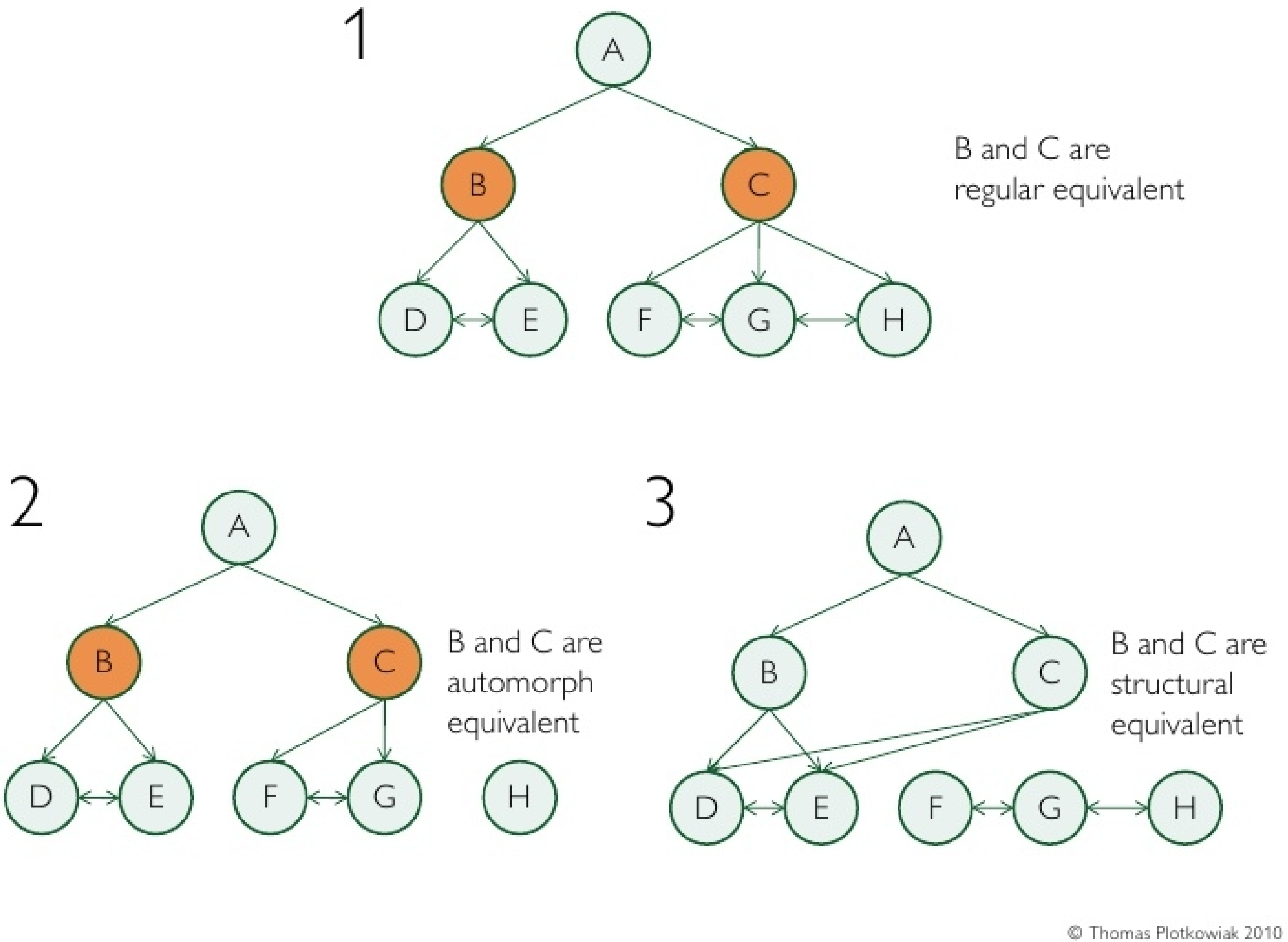
Eigenvector Centrality: A node's eigenvector centrality measures its influence in terms of how connected the node is to other highly connected nodes.





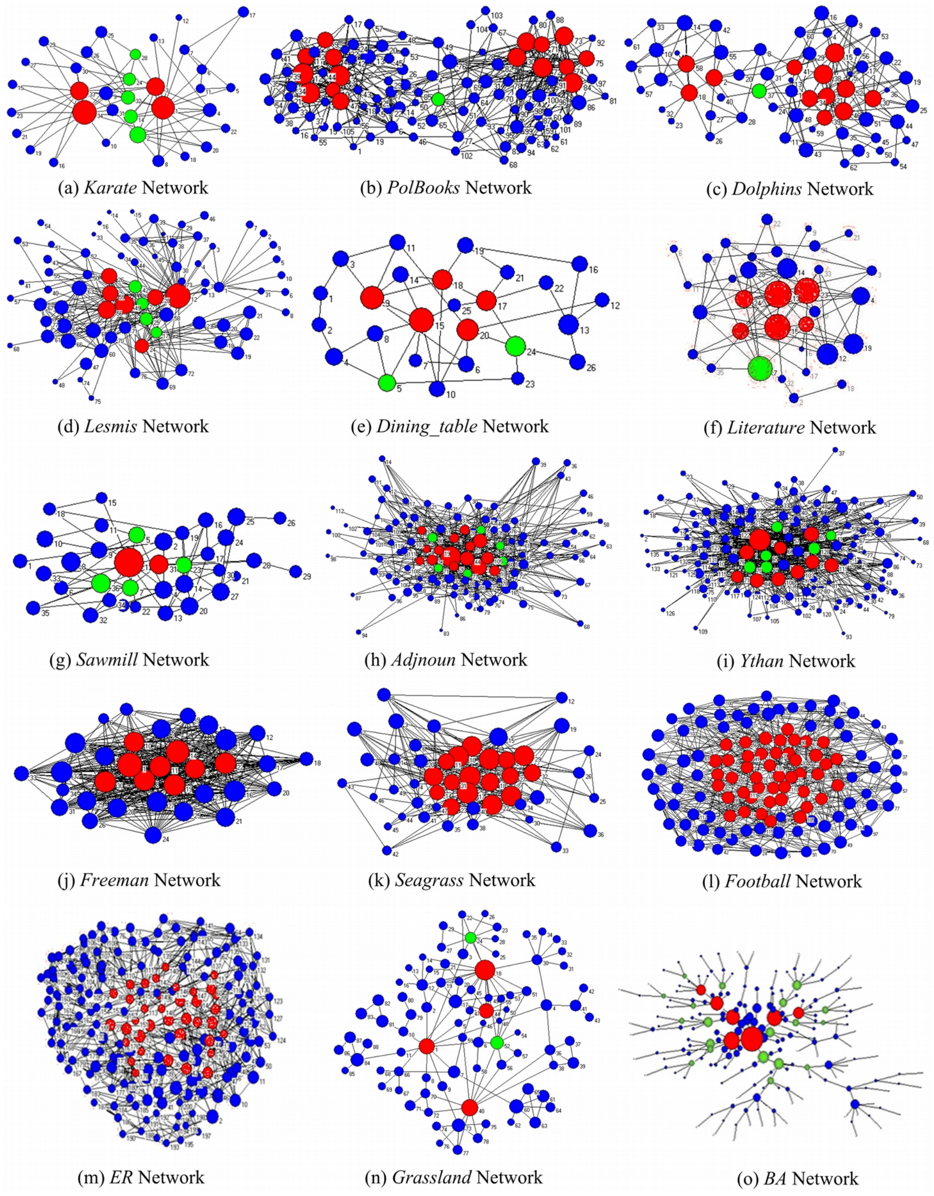
Roles

Identifying Roles in Networks: Structural Equivalence



Identifying Roles in Networks

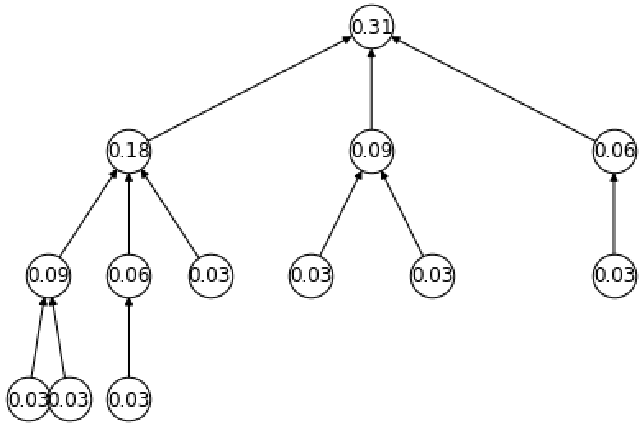
They use several centrality measures to find roles:
degree,
betweenness centrality
and eigenvector centrality

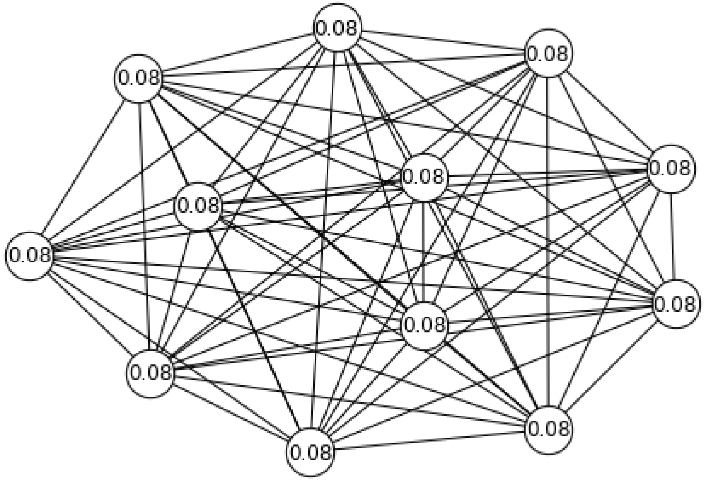


Identifying Node Role in Social Network Based on Multiple Indicators, 2014
Shaobin Huang, Tianyang Lv , Xizhe Zhang, Yange Yang, Weimin Zheng, Chao Wen

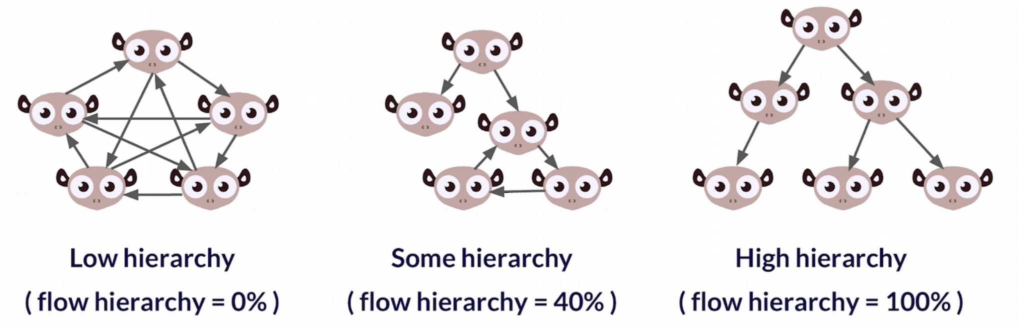
Hierarchy in networks

Pagerank





Flow hierarchy



$$h = \frac{\sum_{i=1}^{L} e_i}{L}, \ \ e_i = 0 \text{ if edge in a cycle (1 otherwise)}$$

Calculate hierarchy value for the whole network
→ useful to compare networks

Structural Holes: Why Some Roles Are More Advantageous than Others

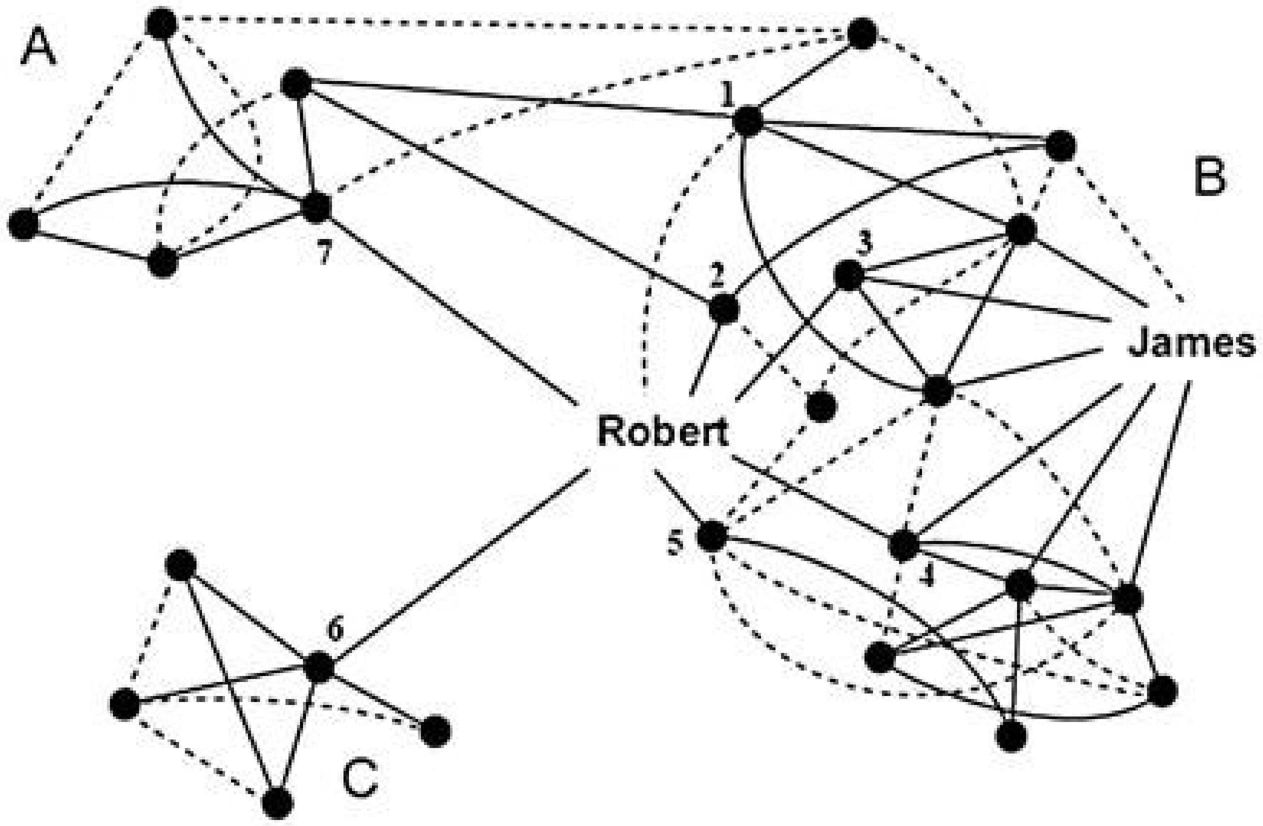
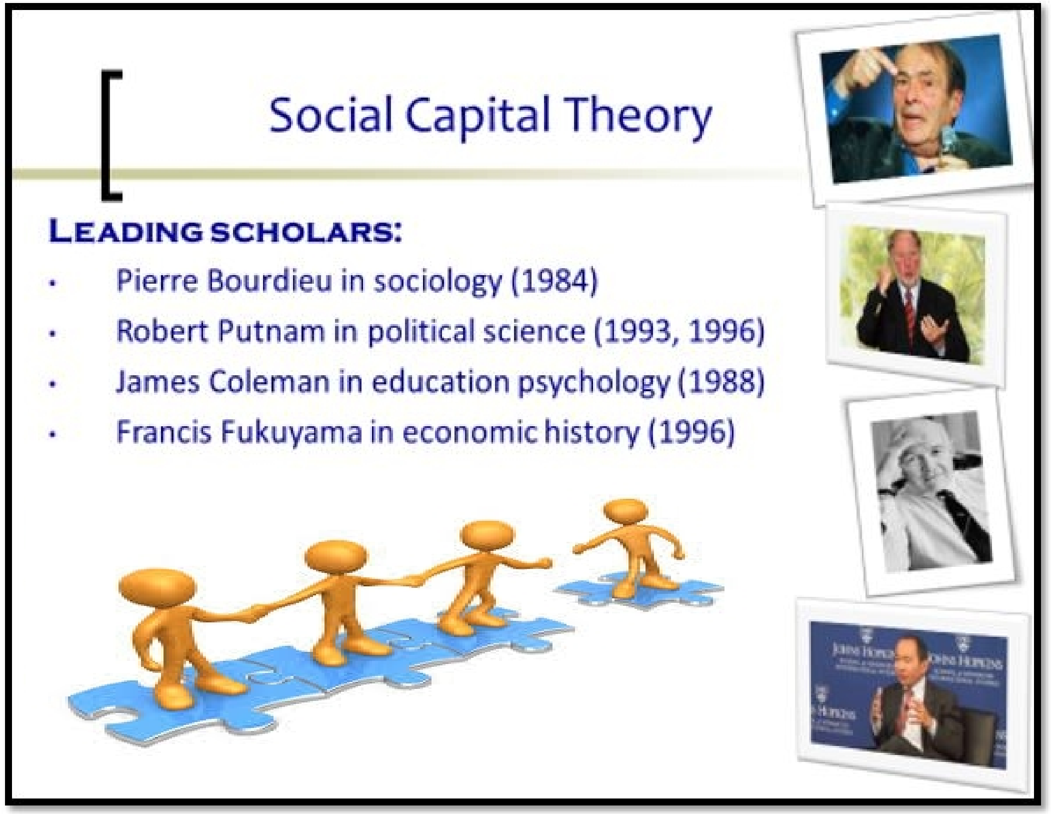


Figure 1. Robert Fills a Structural Hole in the Network.
From: Mauro F. Guillen, Randall Collins, Paula England, and Marshall Meyer, eds. "The Social Capital of Structural Holes," in *New Directions in Economic Sociology* (New York: Russell Sage Foundation, 2002).


Social Capital Theory
Leading scholars:
Pierre Bourdieu in sociology (1984)
Robert Putnam in political science (1993, 1996)
James Coleman in education psychology (1988)
Francis Fukuyama in economic history (1996)


SMALL WORLD.
SMALL SOCIAL CLASS.
COAT CHECK

Part 2: Network analysis with networkx

NetworkX