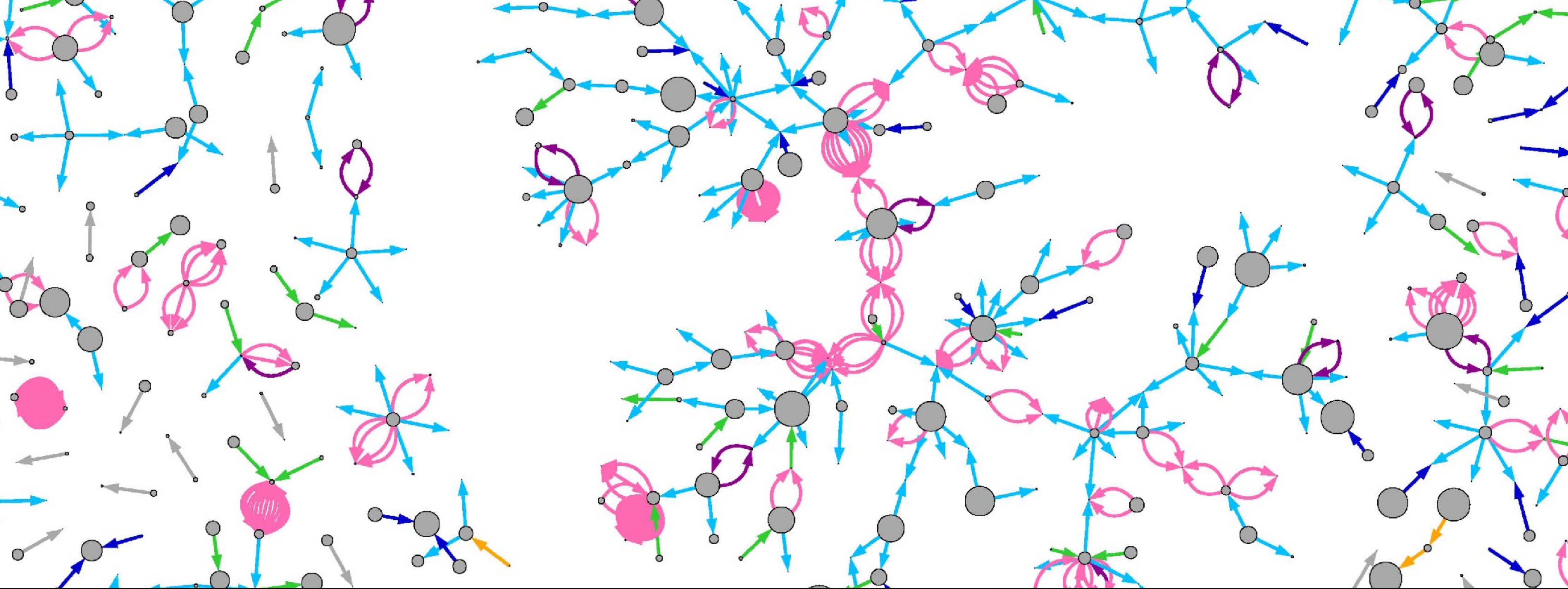


Introduction to Social Network Analysis

MY560 Workshop
Dr. Milena Tsvetkova

tsvetkova.me

- Assistant Professor of Computational Social Science, Dept. of Methodology
  - Program Director: MSc in Applied Social Data Science
  - Lecturer: MY470 Computer Programming, MY461 Social Network Analysis
- Training
  - PhD in Sociology, Cornell University
  - Postdoc in Computational Social Science, Oxford Internet Institute
- Research
  - Network analysis of online data, large-group online experiments, agent-based modeling
  - Cooperation, emergent inequality, social contagion, human-machine networks

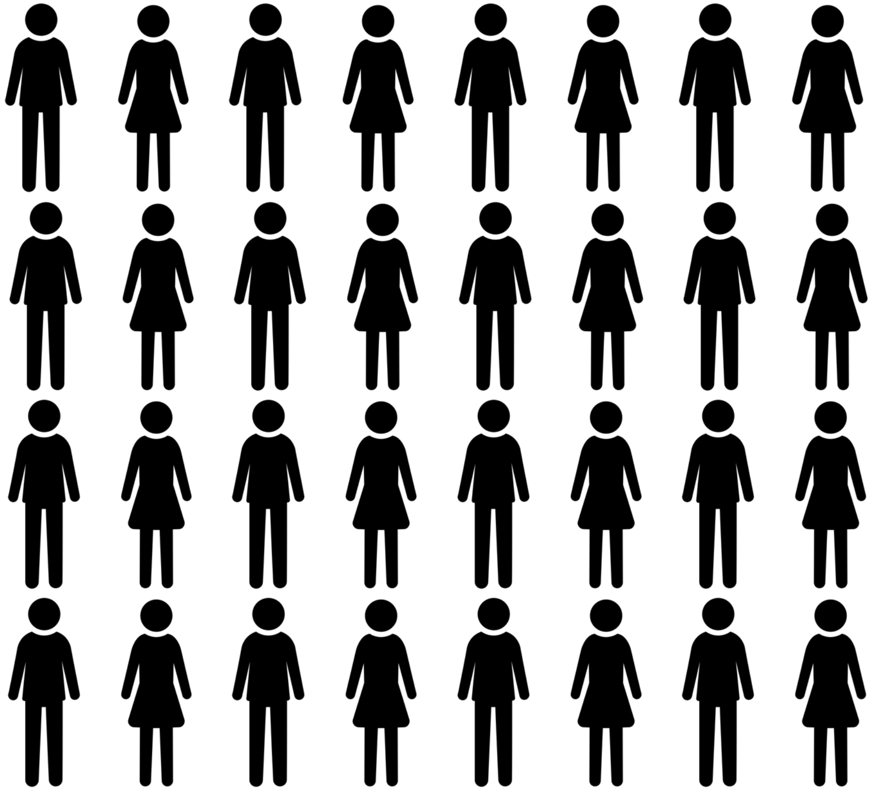
Overview

- (morning) What is social network analysis?
    1. Networks are everywhere!
    2. … but what are networks?
    3. Describing nodes
    4. Describing dyads and triads
    5. Describing networks
    6. Statistical inference on networks
- (afternoon) Social network analysis with R
    1. Predicting individual outcomes with ego networks
    2. Visualizing and describing whole networks

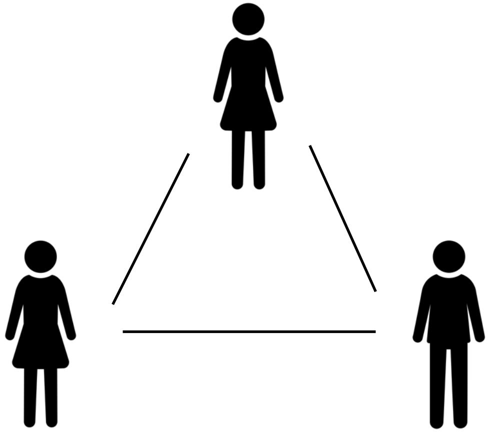
Social networks

Individuals and groups

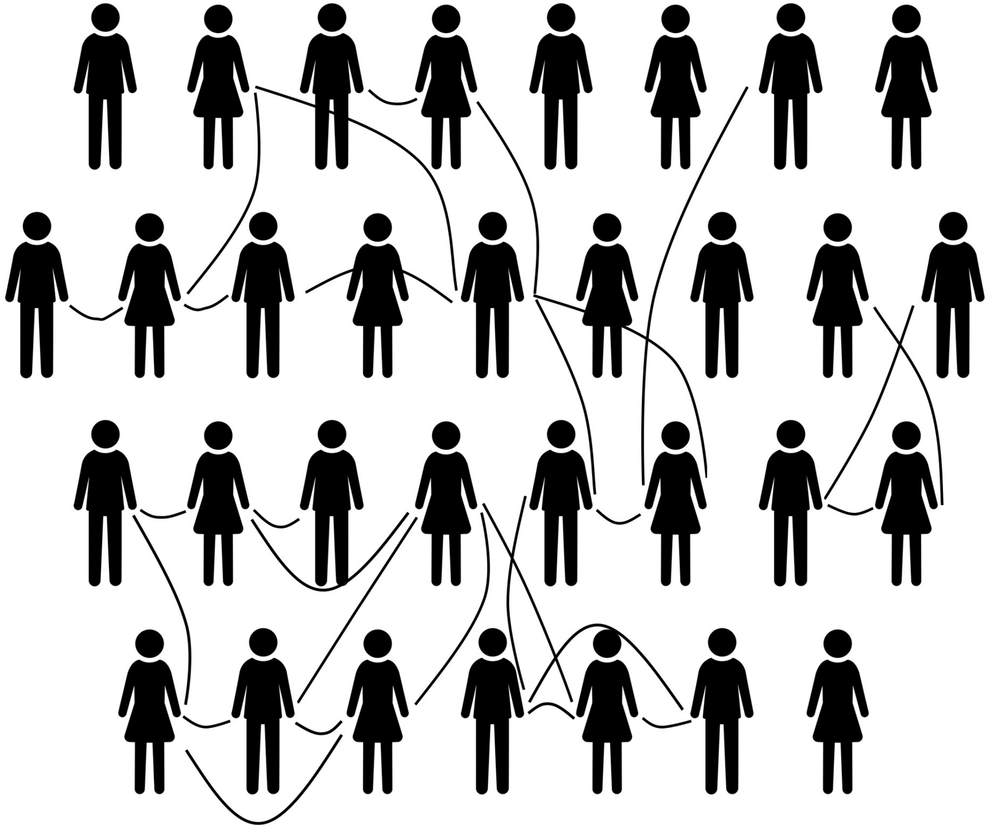
Survey analysis



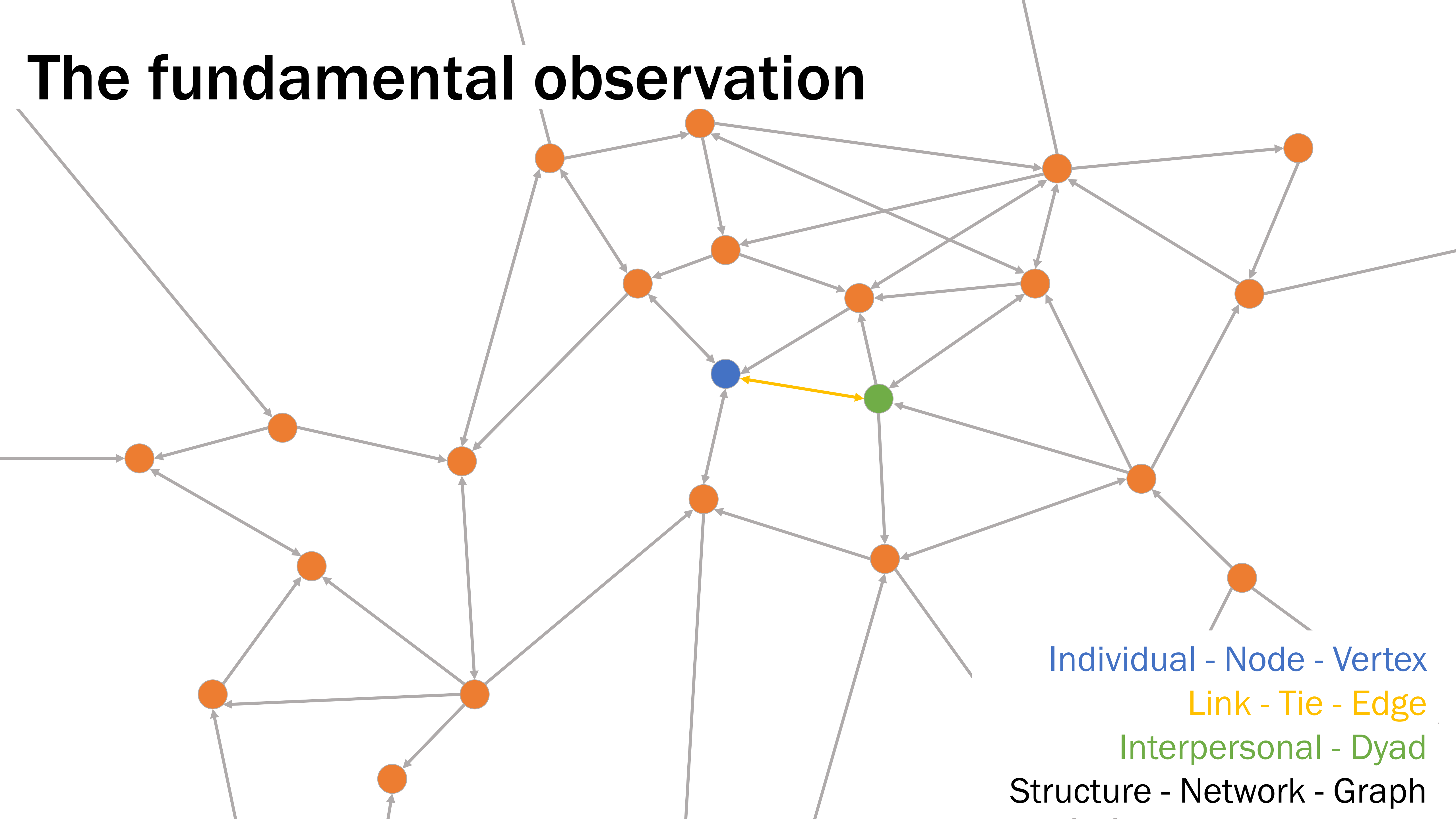
Laboratory experiments



Networks



The fundamental observation



Individual - Node - Vertex
Link - Tie - Edge
Interpersonal - Dyad
Structure - Network - Graph

What is Social Network Analysis?

- Focuses on social structure
- Studies **relationships and interactions** between social entities
- Uses methods from network and graph theory

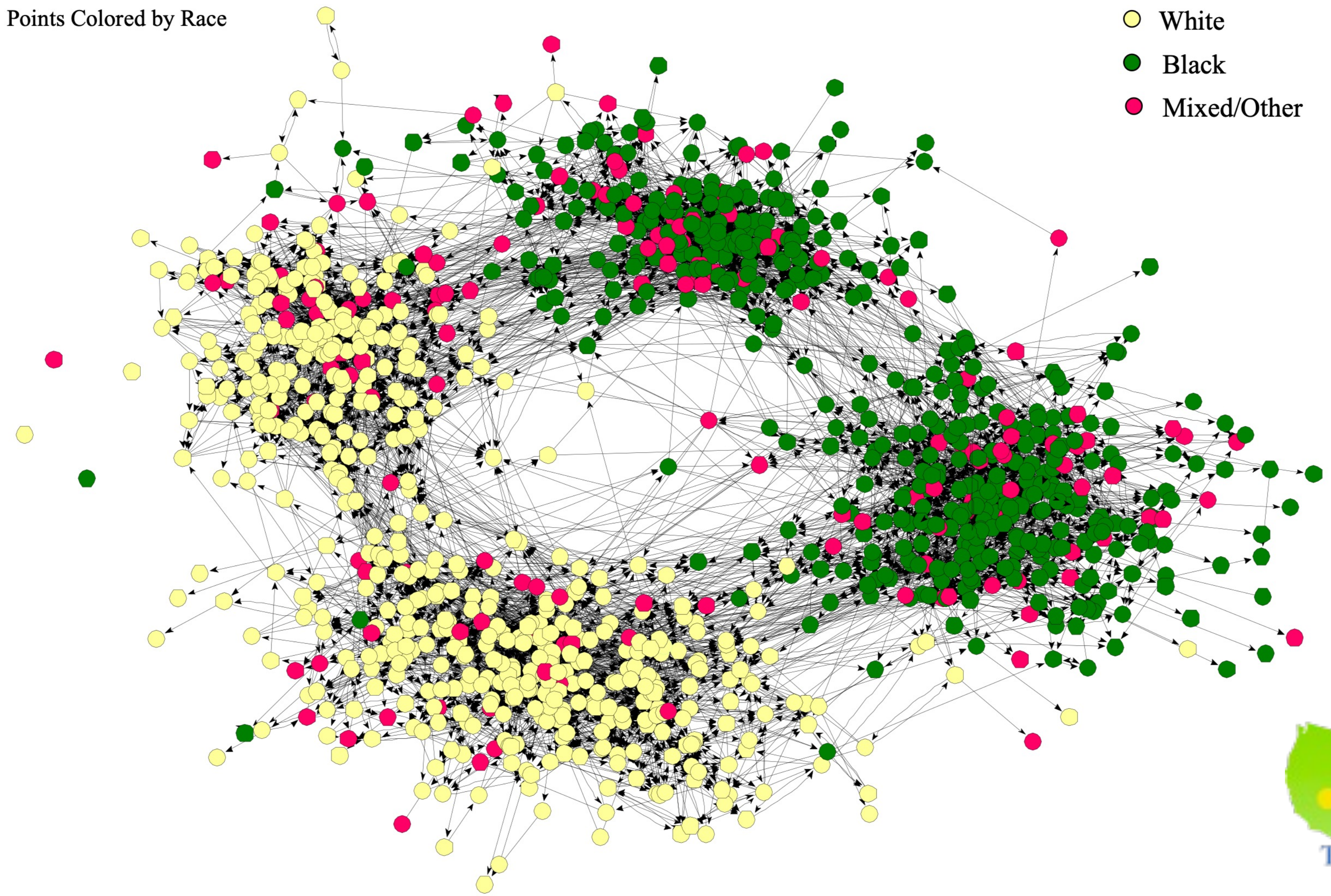
What makes up a social network?

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.,
Same spatial and temporal space	Same clubs	Same gender	Mother of	Friend of	Likes	Knows	Sex with	Information
	Same events	Same attitude	Sibling of	Boss of	Hates	Knows about	Talked to	Beliefs
	etc.	etc.		Student of	etc.	Sees as happy	Advice to	Personnel
				Competitor of		etc.	Helped	Resources
							Harmed	etc.
							etc.	

Fig. 3. A typology of ties studied in social network analysis.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, *323*(5916), 892–895.

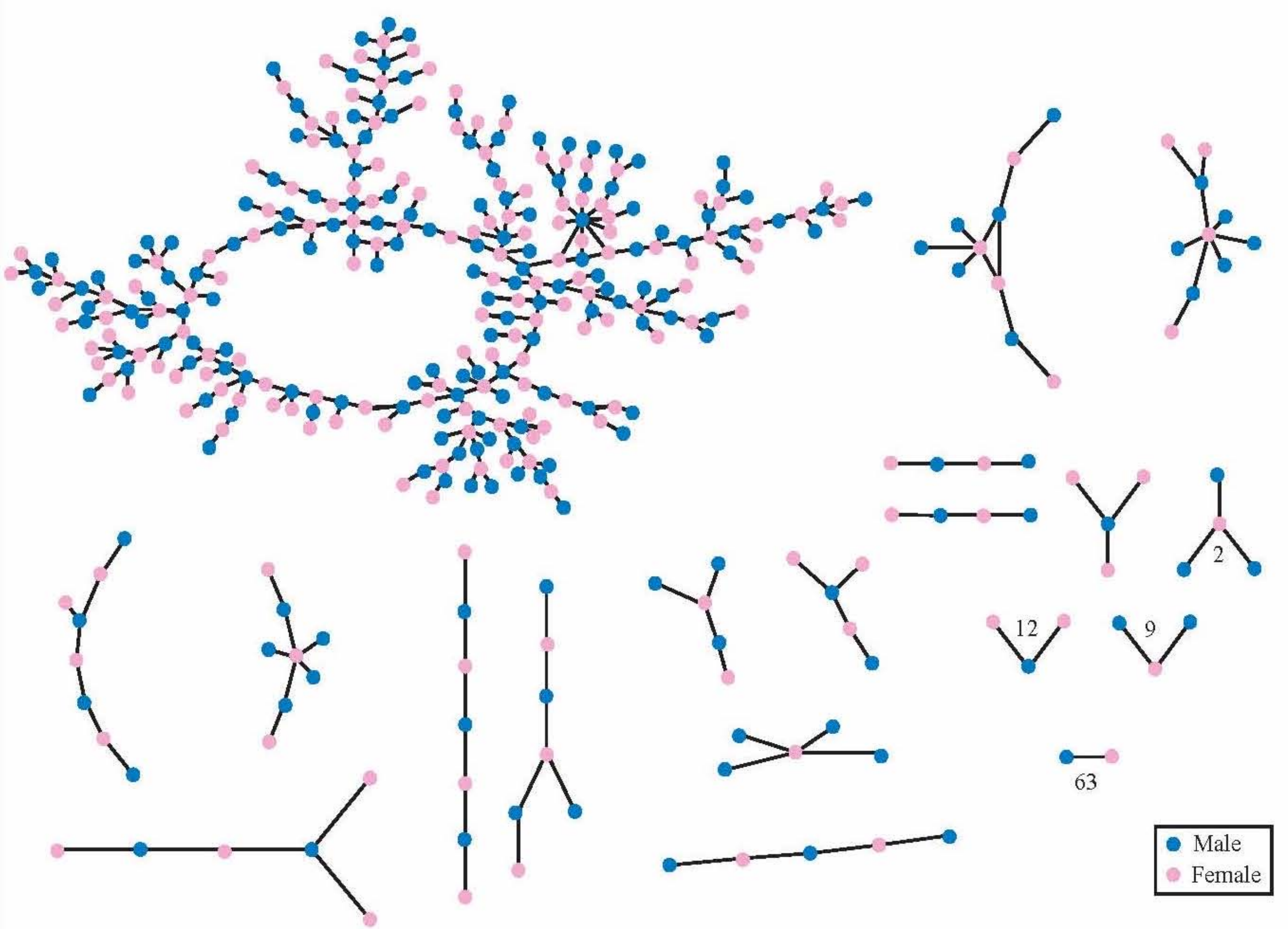
Networks are everywhere: Friendships



The National Longitudinal Study of Adolescent to Adult Health

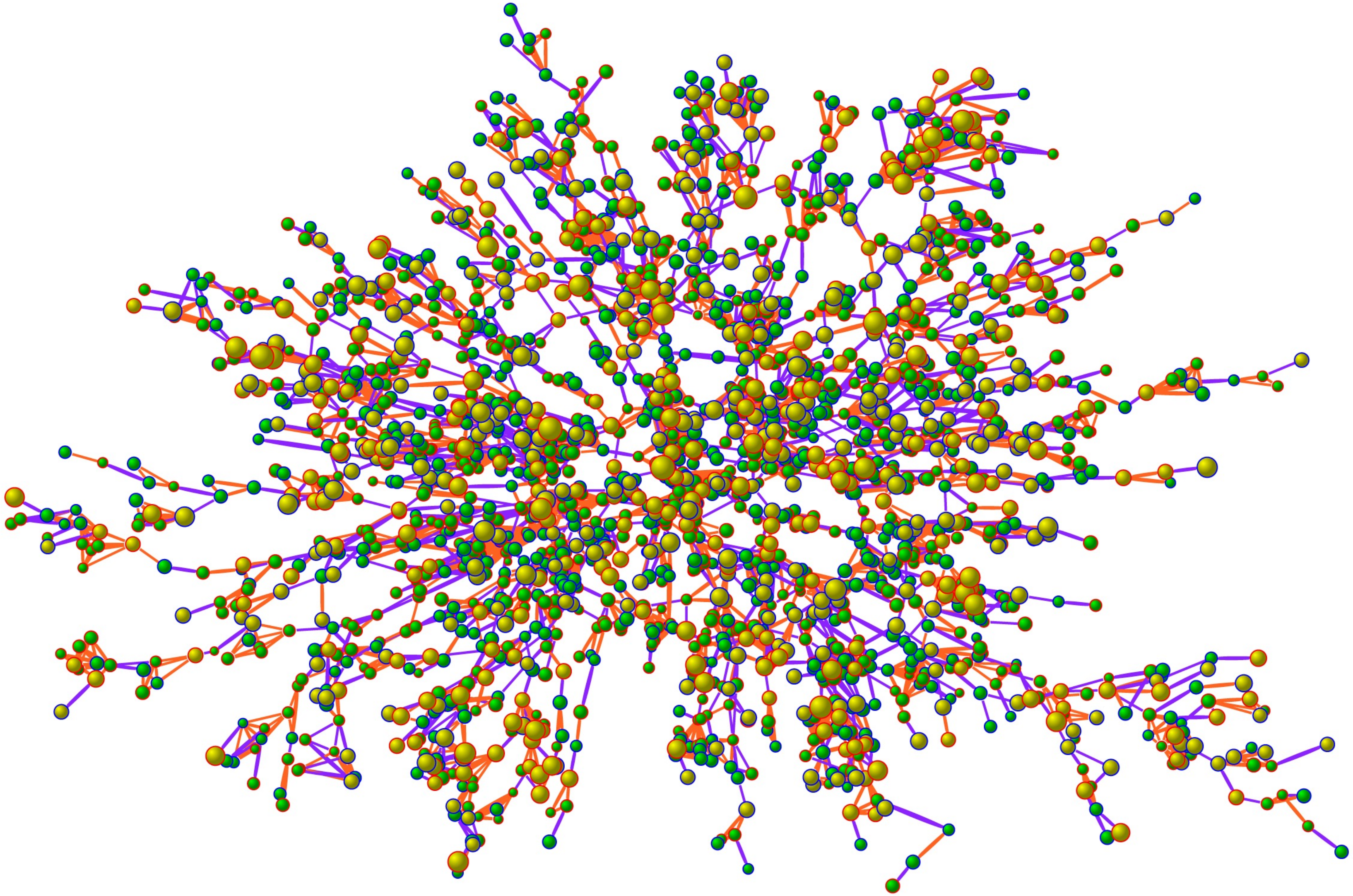
Moody, J. (2001). Race, school integration, and friendship segregation in America. *American Journal of Sociology,* 107(3), 679–716.

Networks are everywhere: Dating



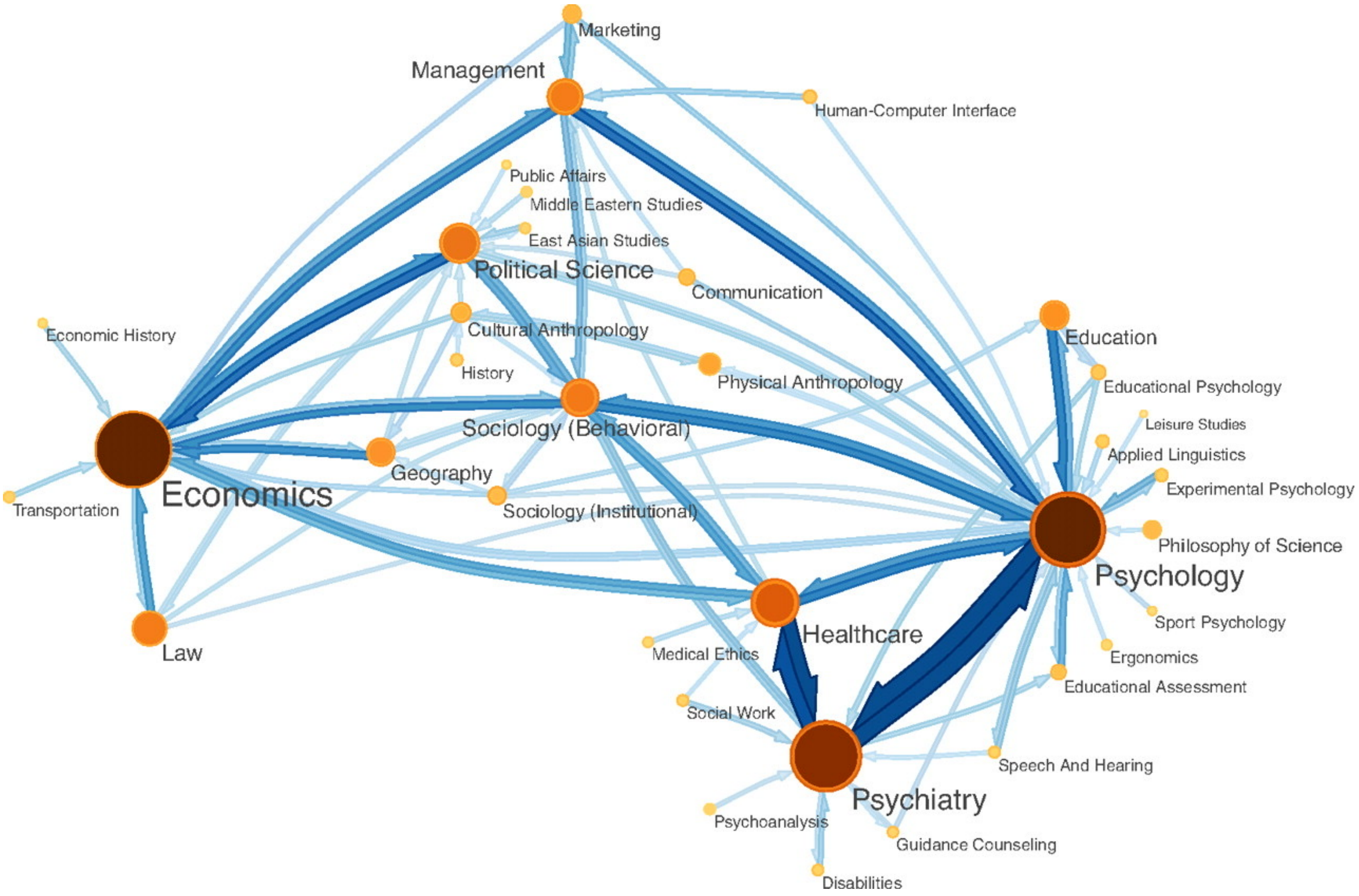
Bearman, P. S. et al. (2004). Chains of affection: the structure of adolescent romantic and sexual networks. *American Journal of Sociology*, *110*(1), 44–91.

Networks are everywhere: Emergency contacts



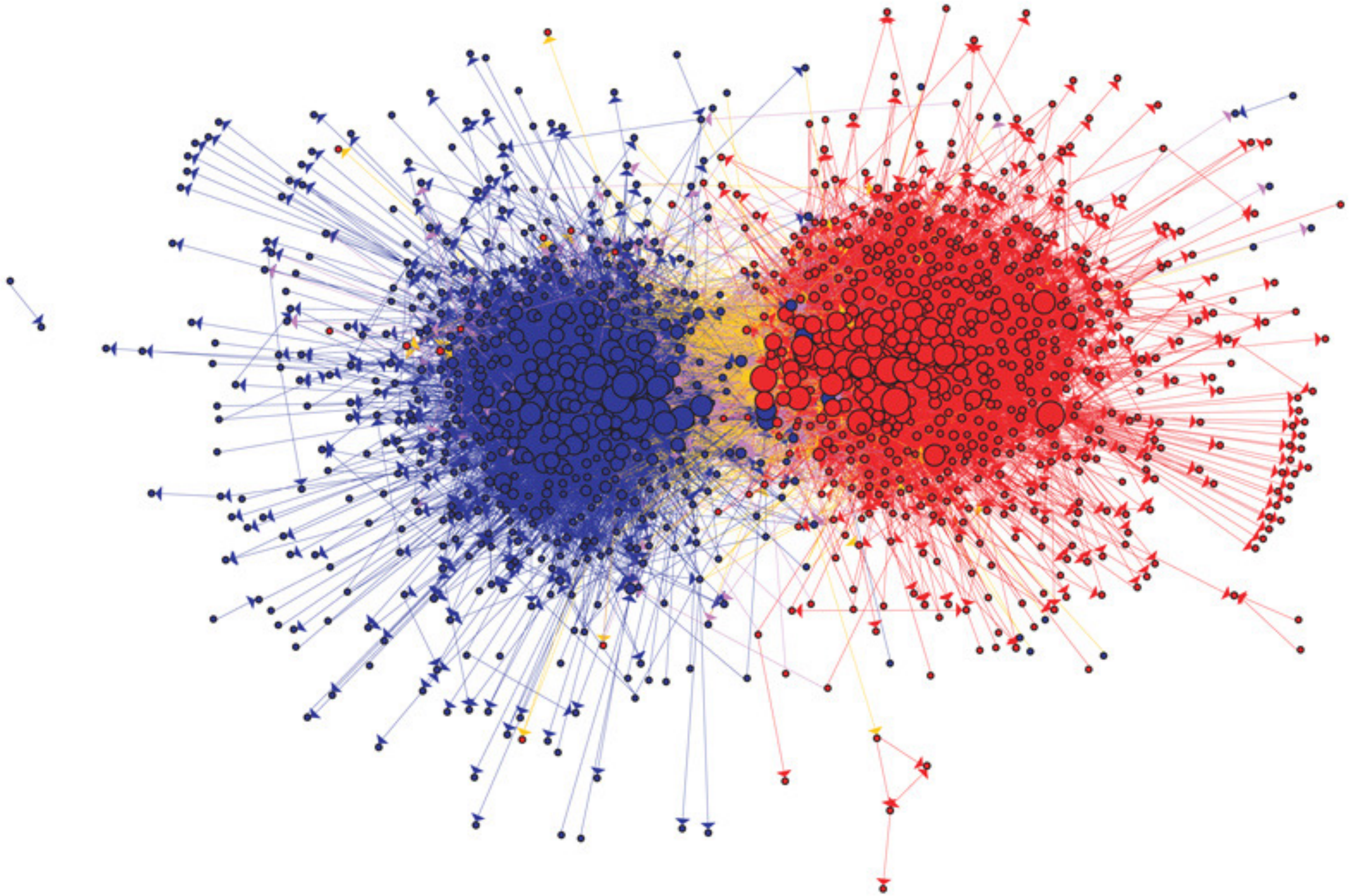
Christakis, N.A. & Fowler, J.H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357, 370–379.

Networks are everywhere: Article citations



Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *PNAS*, 105(4), 1118–1123.

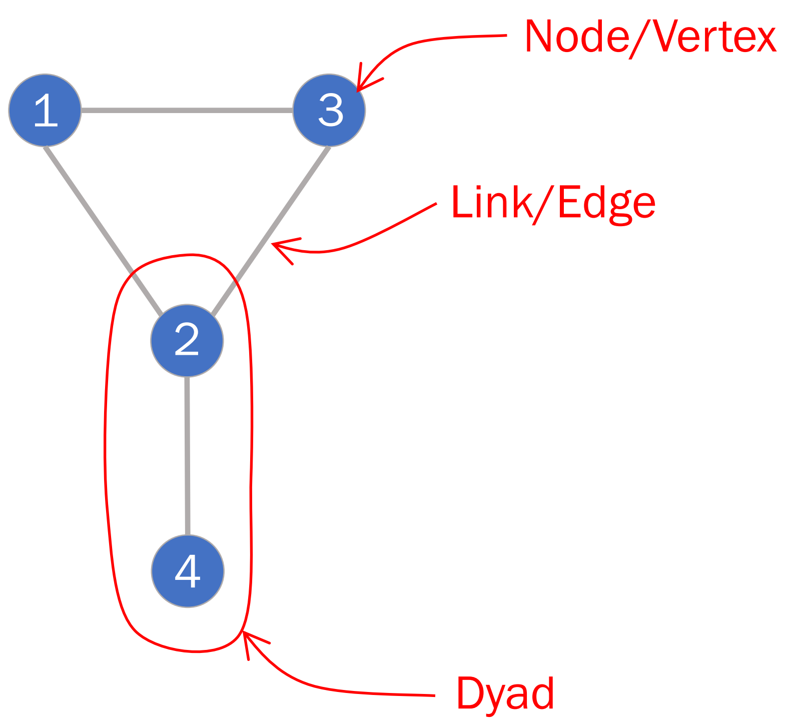
Networks are everywhere: Political blogs



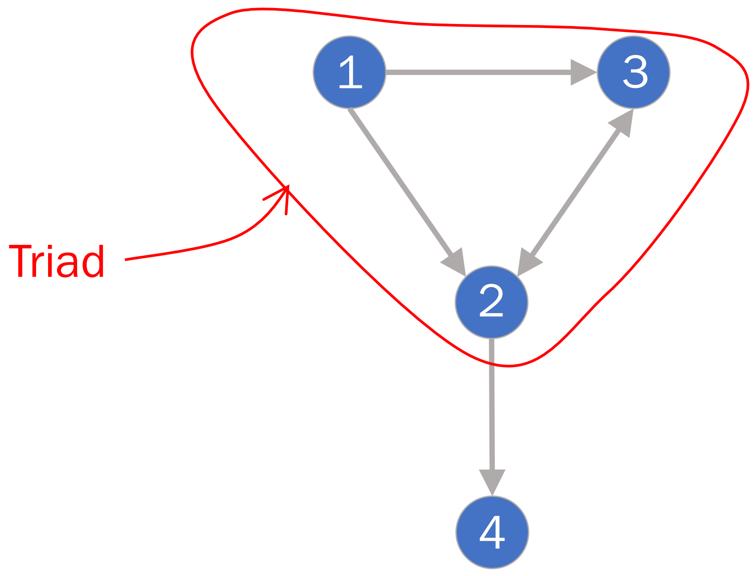
Adamic, L. & Glance, N. (2005). The political blogosphere and the 2004 US election: divided they blog. *Proc. of the 3rd Int. Workshop on Link Discovery*, 36–43.

Network terminology

Undirected Network

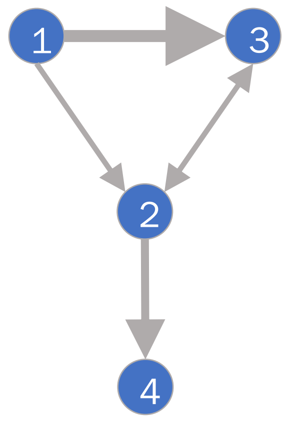


Directed Network

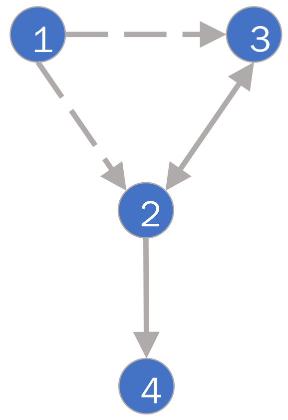


Types of networks

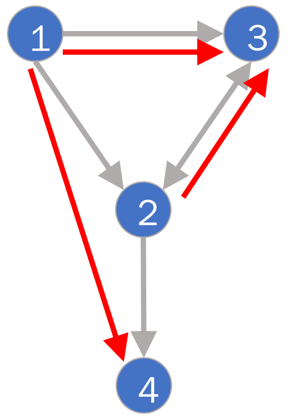
Weighted Network



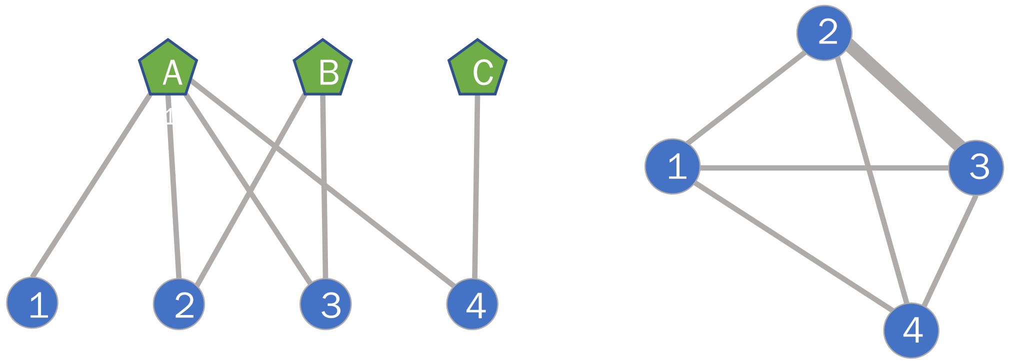
Signed Network



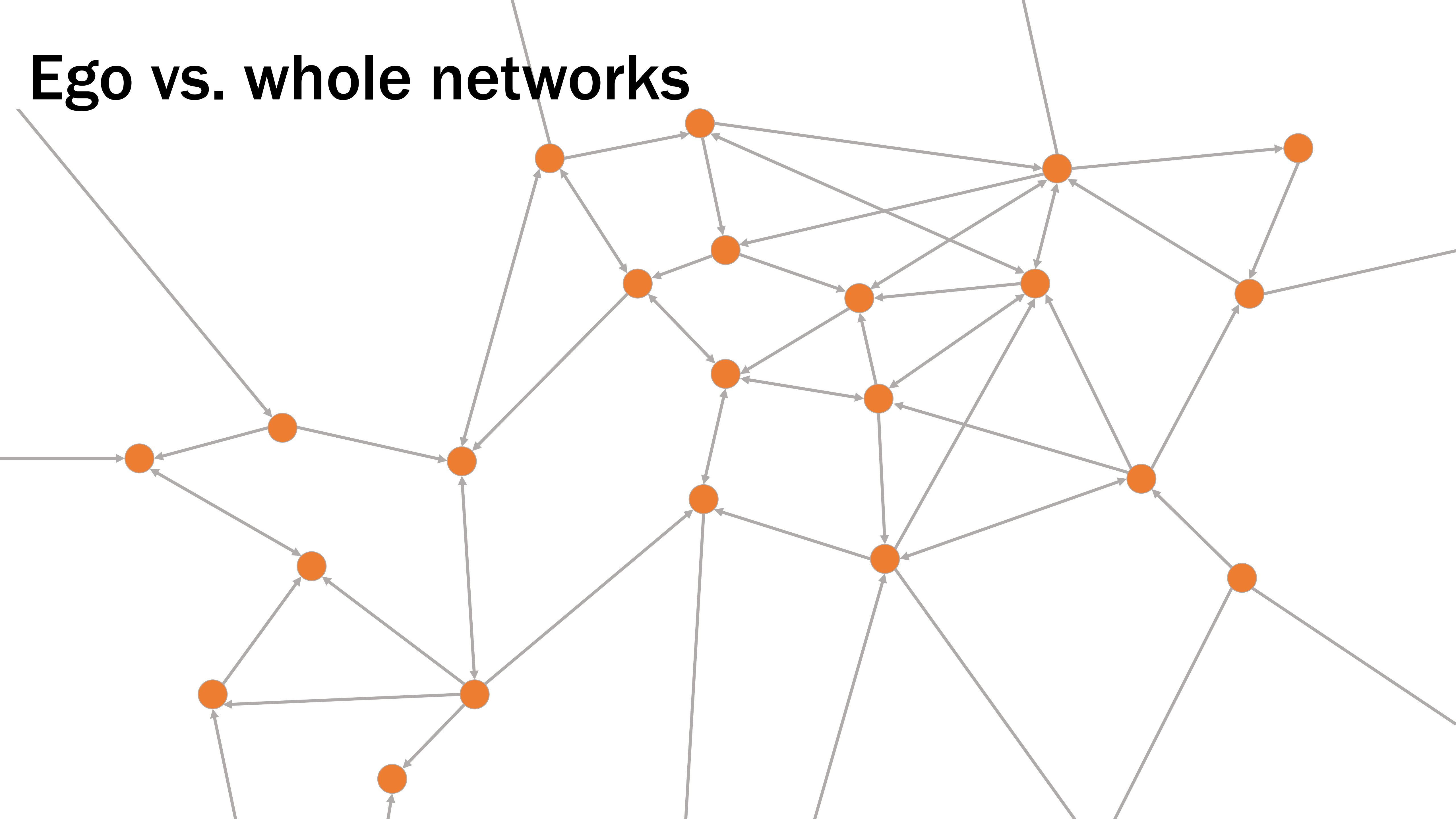
Multilayer Network



Affiliation Network



Ego vs. whole networks



Ego Networks

- "Ego networks" are the personal network of an individual
- They are usually gathered by talking only to the focal individual via a survey
- Often, this uses a "name generator." The General Social Survey asks:

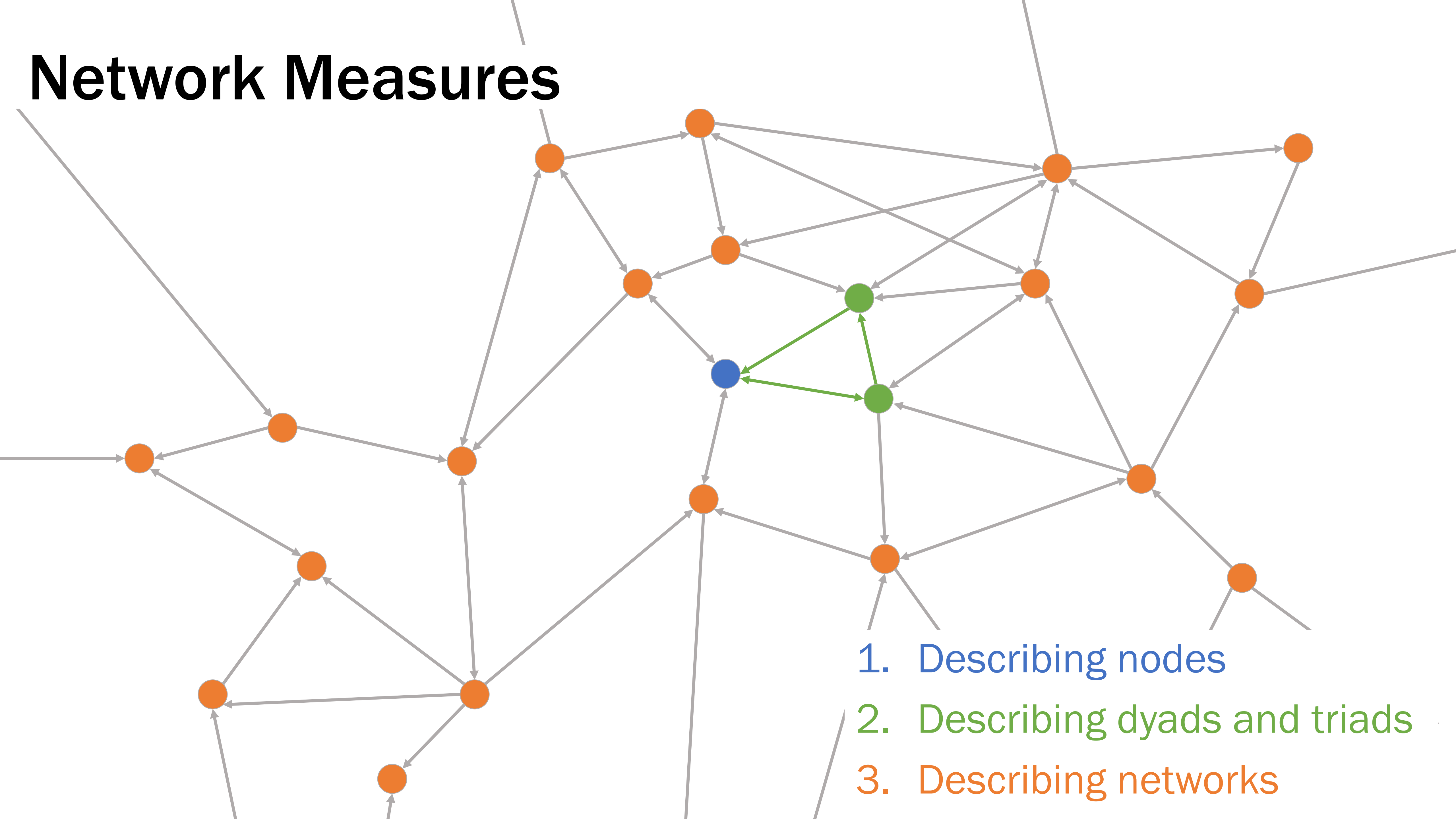
  "From time to time, most people discuss important matters with other people. Looking back over the last six months — who are the people with whom you discussed matters that are important to you?"

- Then, the focal individual reports further information on those named
  - Age, gender, type of relationship, etc.
  - Do the named know each other?

Whole Networks

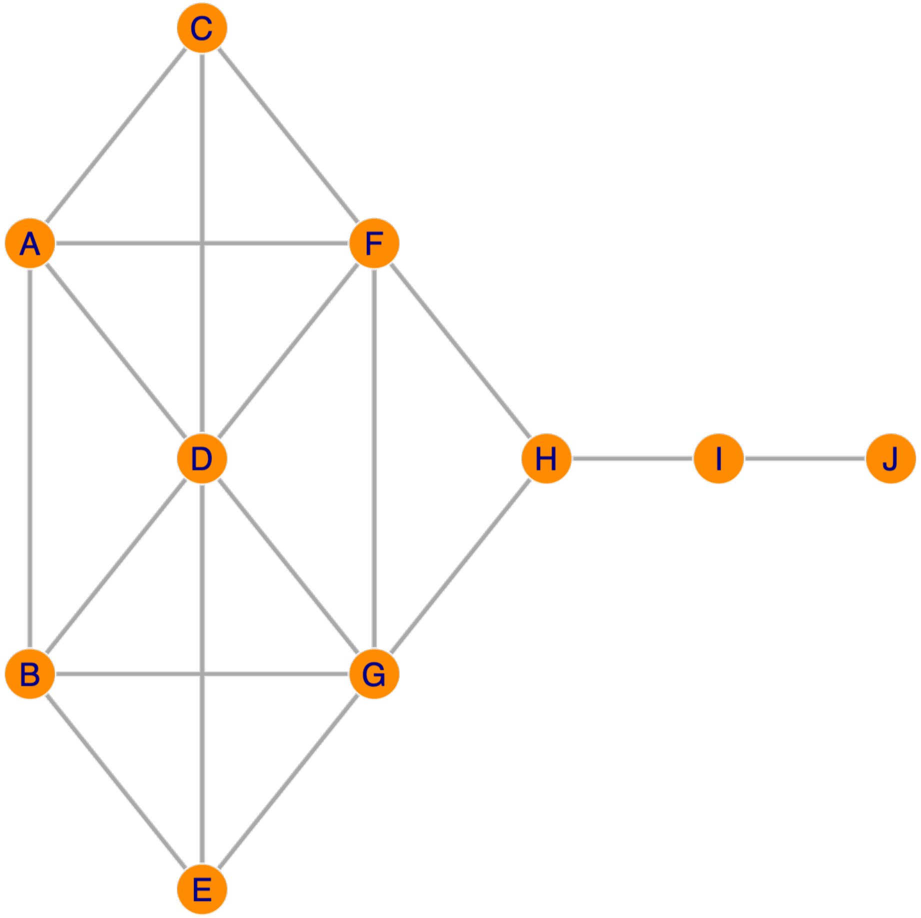
- "A whole network" is a census of everyone at a particular organization, school, community, village, or any other well defined and delineated group
- Whole networks can be gathered by name-generator surveys but also digital records of online communication on social media platforms and online communities
- In these type of data, **observations (whether individuals or links) are not independent**
- If we want to explain the network or how the network affects individuals, we cannot employ standard regression models
  - Instead, we should use Exponential Random Graph Models, Stochastic Actor Oriented Models, network permutation approaches, etc.

Network Measures



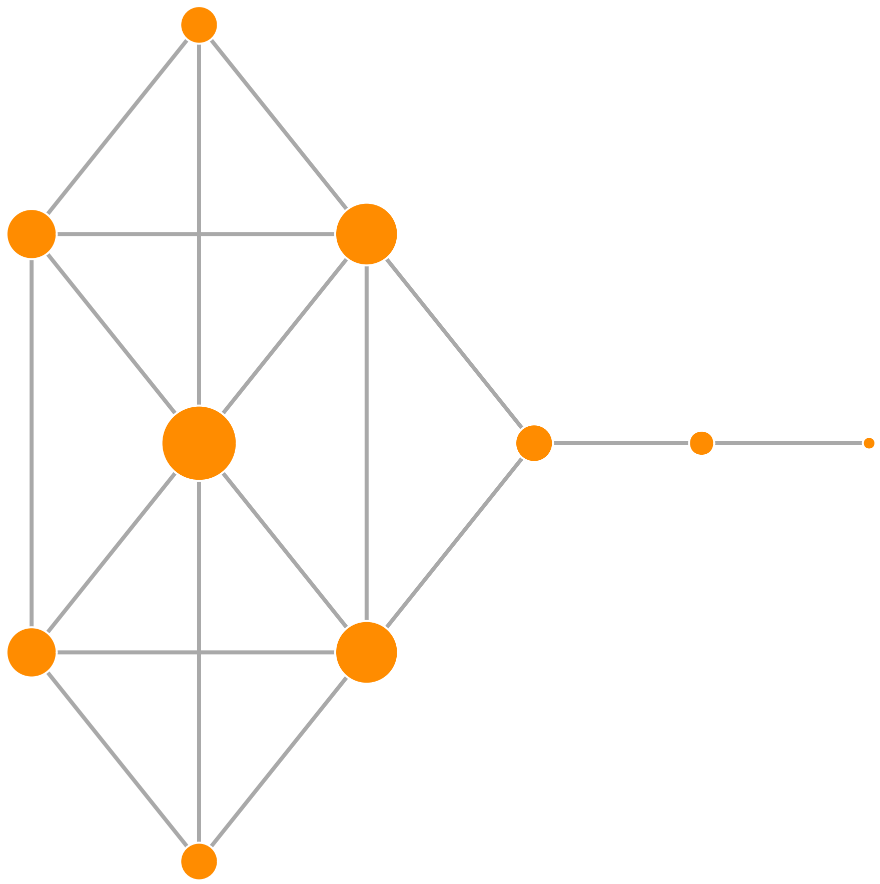
1. Describing nodes
2. Describing dyads and triads
3. Describing networks

Node centrality: Which is the most important node?



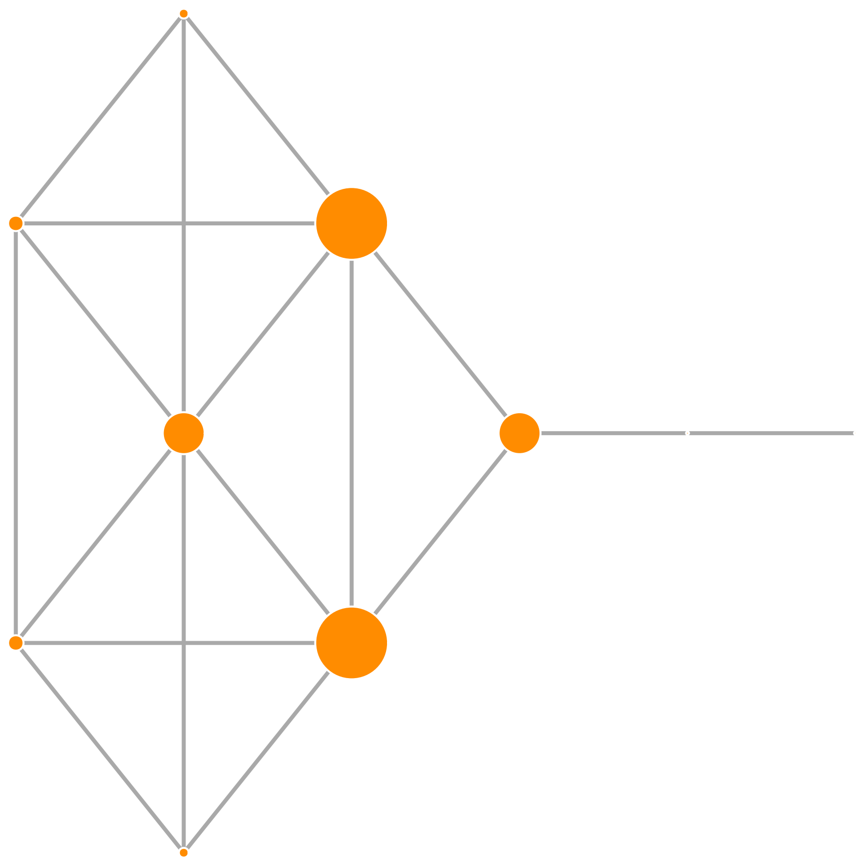
Degree centrality

- The number of adjacent edges
- For directed networks:
  - In-degree centrality
  - Out-degree centrality
- Value range depends on the network
- Meaning depends on the ties
  - In-degree: Prominence, popularity, generosity?
  - Out-degree: Support, reach?



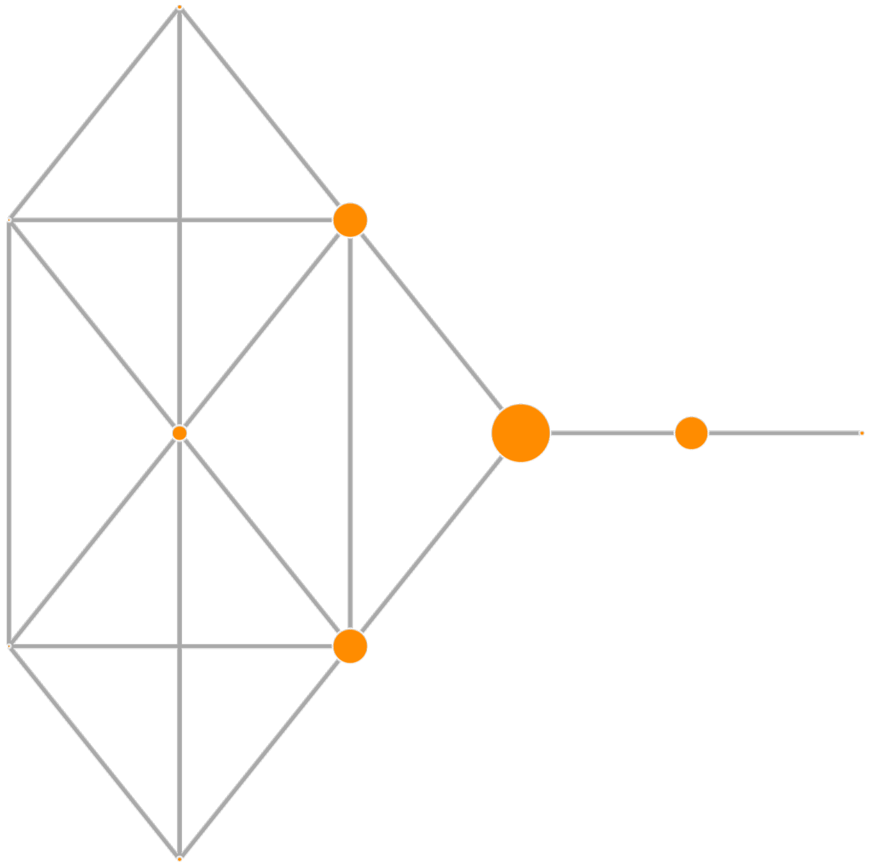
Closeness centrality

- How quickly can you reach someone else? How many intermediaries would it take?
- Closeness centrality measures how far a node is from all other nodes (really, the inverse of this, so that higher = more)



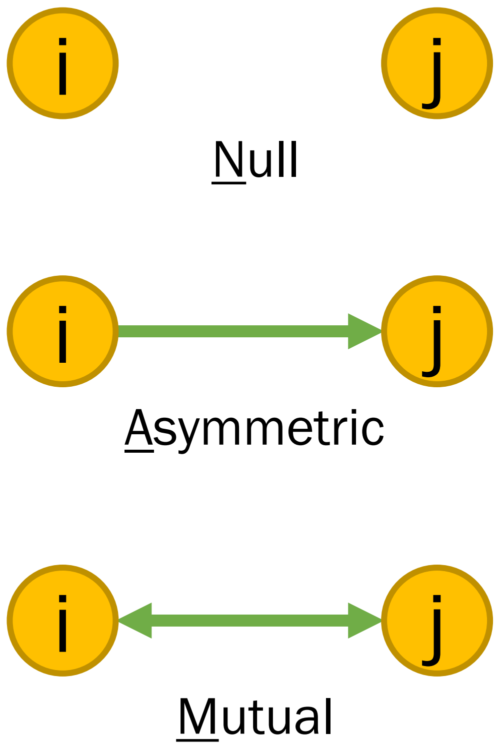
Betweenness centrality

- Betweenness centrality uses paths in a different way
- It measures how often a node lies on the path between *other* nodes



Dyads: Reciprocity

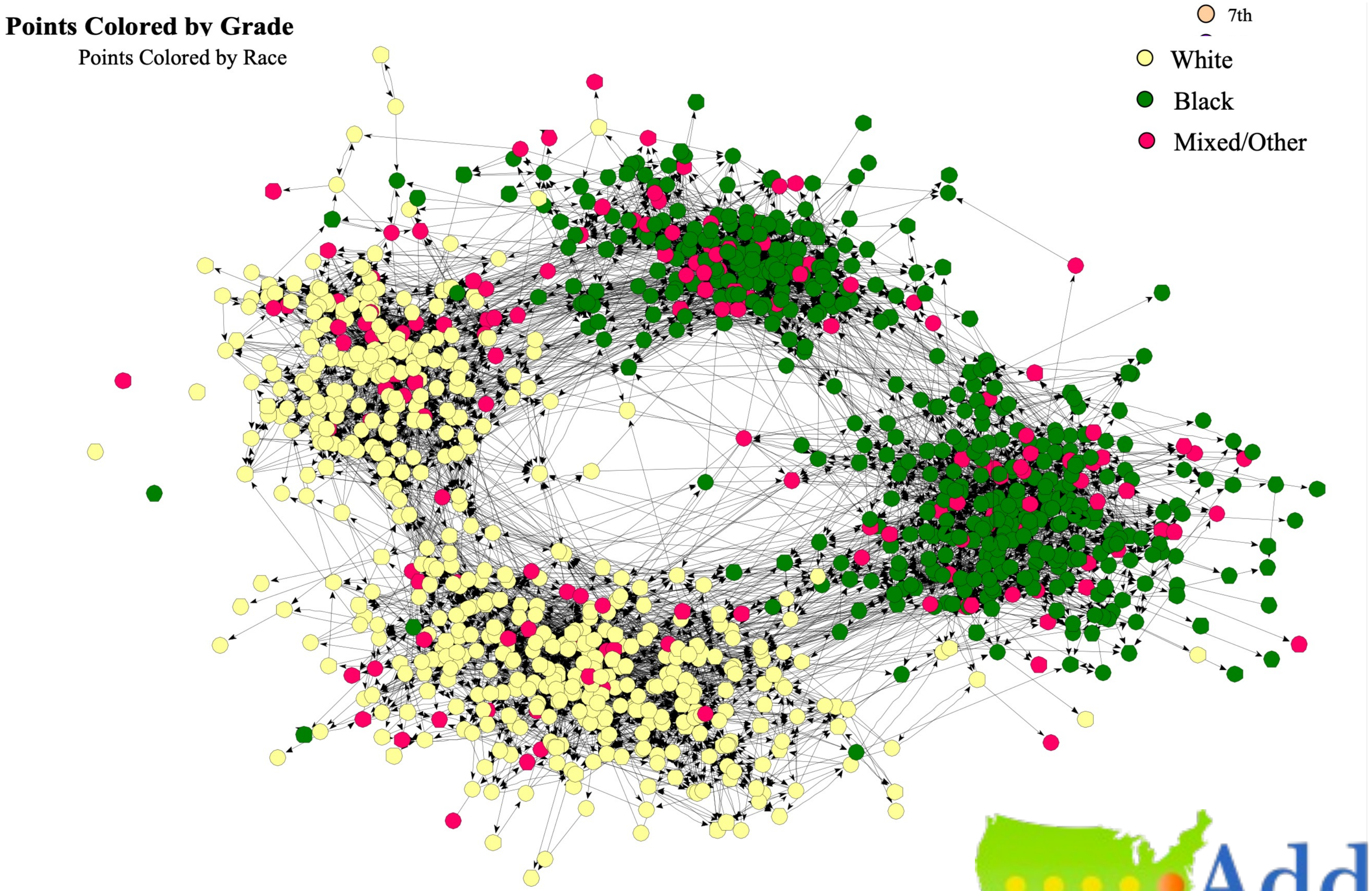
- If j is connected to i, there's a pretty good chance that i is connected to j
  - i.e., their relationship is ***mutual/reciprocal***
- <u>Reciprocity</u>:
  - The fraction of edges that are reciprocated
  - Or, the probability that the opposite counterpart of a directed edge is present, given that one edge exists



Dyads: Homophily

- Often, if i and j are alike in some way, they are also more likely to have a relationship
- This is known as <u>homophily</u> (like associates with like) or <u>assortative mixing</u>
- There can also be <u>disassortative mixing</u>, as for example with romantic & sexual relationships
- Nodes can be "alike" in many ways: age, gender, ethnicity, income, occupation, class, beliefs, behaviours, country, etc.
- To quantify homophily, compare what we would expect if there were *no* assortative mixing; if type had *no effect* (i.e., perfect random mixing)

Dyads: Homophily



Add Health
The National Longitudinal Study of Adolescent to Adult Health

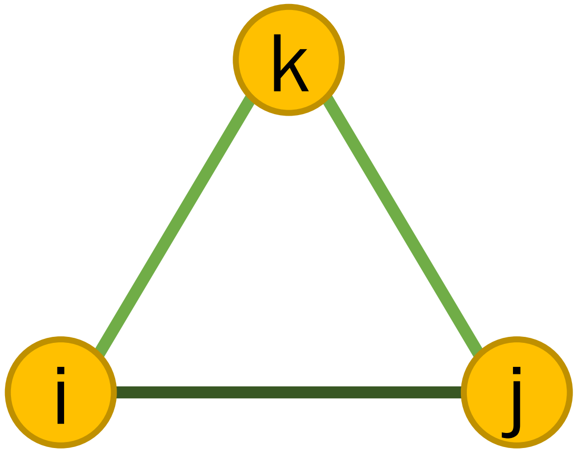
Dyads: Homophily

- This association can be because of an actual interest in associating with like, but can also result from other social structural forces
  - Geography
  - Demographics
  - Kinship
  - Organizational foci (school, office)
  - Ease (cognitive and social processes)
- Some ties may be less likely to form, or more easily dissolved
- Some homophily may be because people who are friends come to resemble each other (on mutable qualities: attitudes, drug use, obesity, etc.)

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, *27*, 415–444.

Triads: Triadic closure

- i and j might be more likely to have a relationship because of their shared association with a third node, k
- <u>Triadic closure</u> is the propensity of nodes i and j to become connected, given that they both have a connection to k
- A closed triad, is one where all 3 nodes are connected



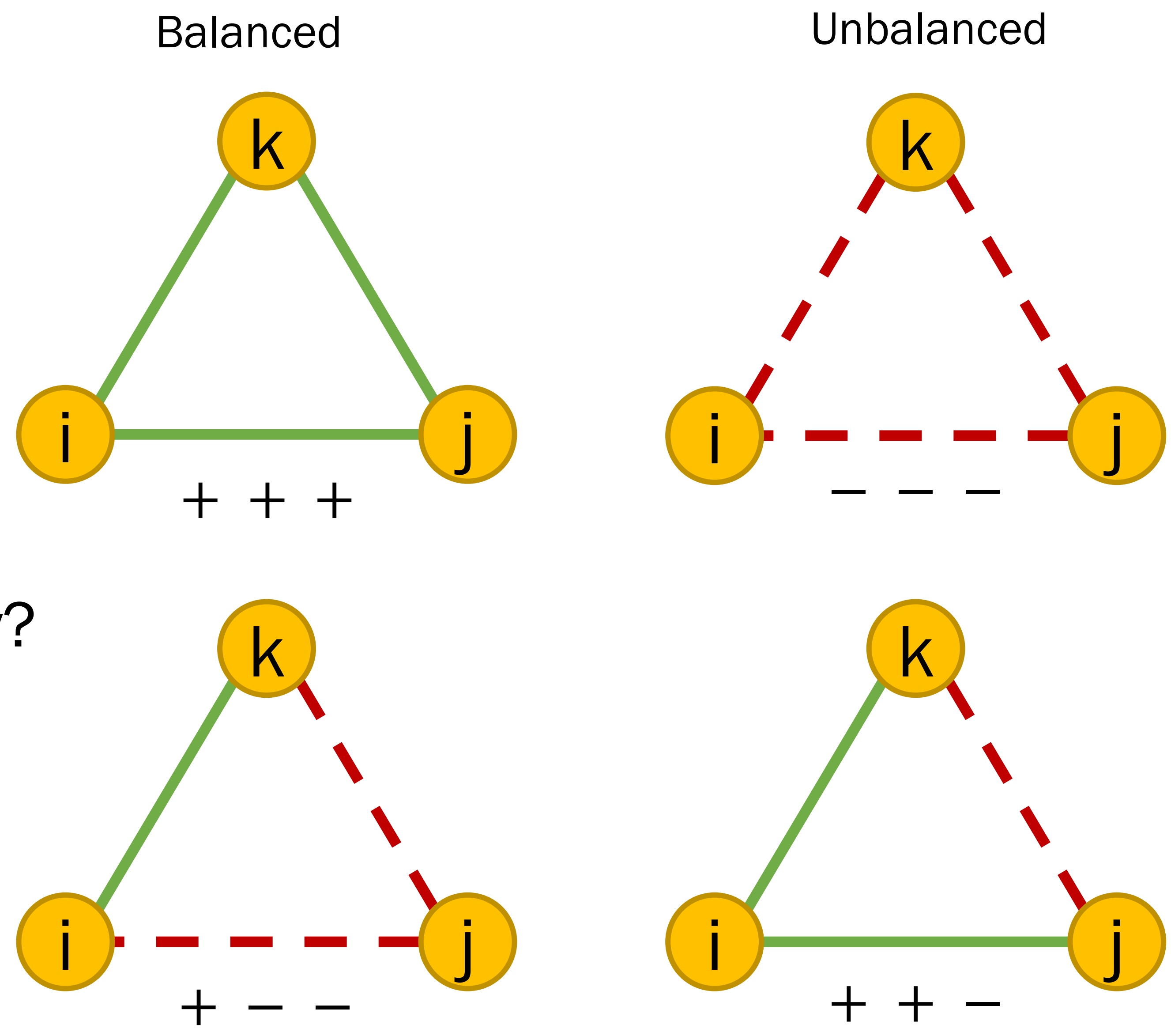
Triads: Structural balance

Balanced:

- A friend of a friend is a friend
- An enemy of my enemy is a friend

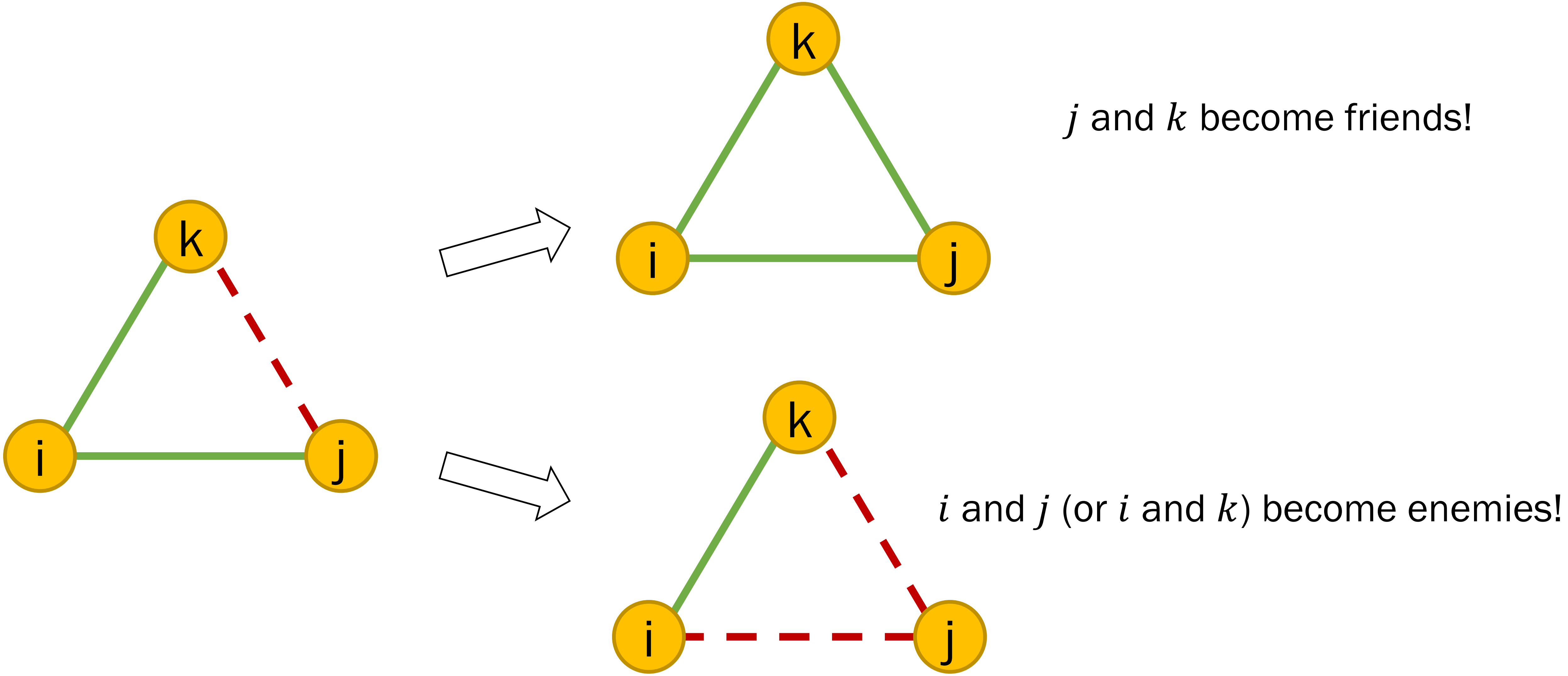
Unbalanced:

- An enemy of my enemy is an enemy?
- A friend of a friend is an enemy?



Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology,21*(1), 107–112.

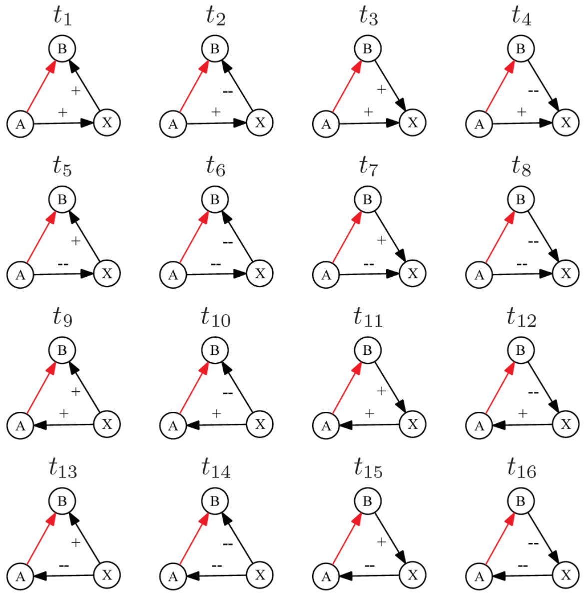
Triads: Structural balance



Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology*, *21*(1), 107–112.

Triads: Structural balance

- Leskovec et al (2010) use data from online opinion measures, which can be + or – and are directed (Epinions, Slashdot, Wikipedia)
- 16 isomorphism classes for the completion of a triad with **directed signed edges**
- Find that in some ways the patterns align with structural balance, but not always, so bring in an idea of relative status

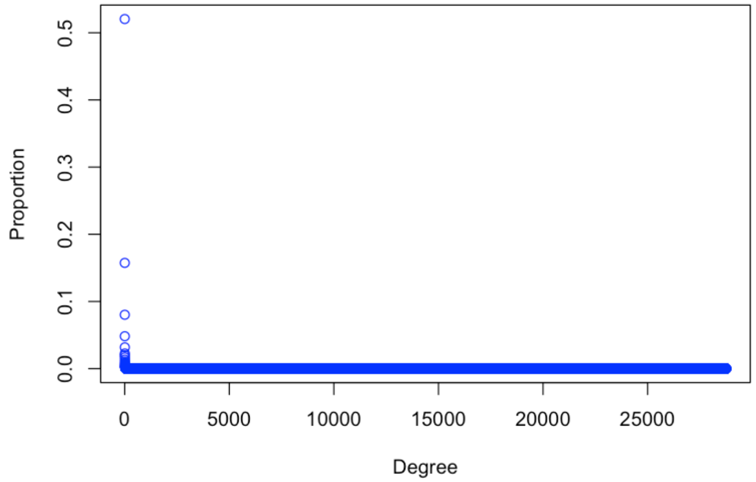


t_i	count	$P(+)$	s_g	s_r	B_g	B_r	S_g	S_r
t_1	178,051	0.97	95.9	197.8	✓	✓	✓	✓
t_2	45,797	0.54	-151.3	-229.9	✓	✓	✓	○
t_3	246,371	0.94	89.9	195.9	✓	✓	○	✓
t_4	25,384	0.89	1.8	44.9	○	○	✓	✓
t_5	45,925	0.30	18.1	-333.7	○	✓	✓	✓
t_6	11,215	0.23	-15.5	-193.6	○	○	✓	✓
t_7	36,184	0.14	-53.1	-357.3	✓	✓	✓	✓
t_8	61,519	0.63	124.1	-225.6	✓	○	✓	✓
t_9	338,238	0.82	207.0	-239.5	✓	○	✓	✓
t_{10}	27,089	0.20	-110.7	-449.6	✓	✓	✓	✓
t_{11}	35,093	0.53	-7.4	-260.1	○	○	✓	✓
t_{12}	20,933	0.71	17.2	-113.4	○	✓	✓	✓
t_{13}	14,305	0.79	23.5	24.0	○	○	✓	✓
t_{14}	30,235	0.69	-12.8	-53.6	○	○	✓	○
t_{15}	17,189	0.76	6.4	24.0	○	○	○	✓
t_{16}	4,133	0.77	11.9	-2.6	✓	○	✓	○
				Number of correct predictions	8	7	14	13

Figure 2. Top: All contexts $(A, B; X)$. Red edge is the edge that closes the triad. Bottom: Surprise values and predictions based on the competing theories of structural balance and status. t_i refers to triad contexts above; *Count*: number of contexts t_i; $P(+)$: prob. that closing red edge is positive; s_g: surprise of edge initiator giving a positive edge; s_r: surprise of edge destination receiving a positive edge; B_g: consistency of balance with generative surprise; B_r: consistency of balance with receptive surprise; S_g: consistency of status with generative surprise; S_r: consistency of status with receptive surprise.

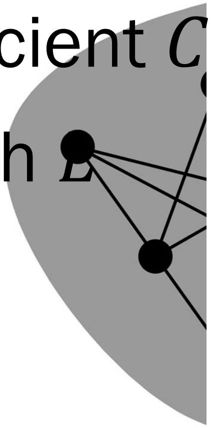
Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010). Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*(pp. 1361–1370). New York, NY, USA: ACM.

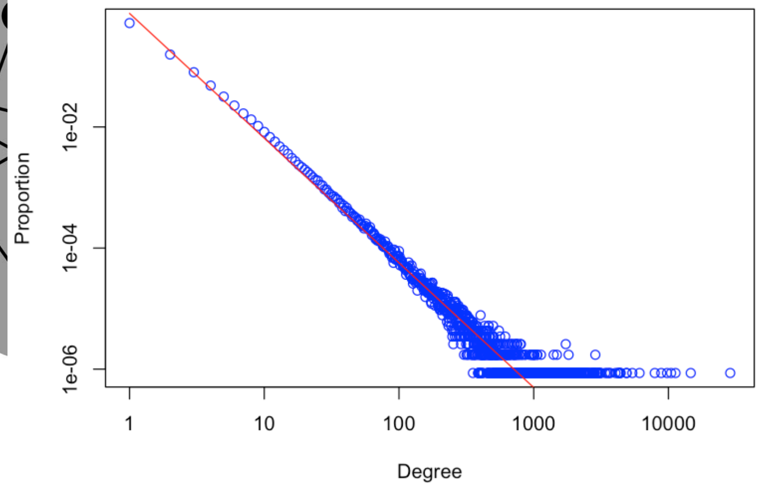
From nodes, dyads, and triads to networks



- Node degree ⟶ Network degree distribution
- Local clustering ⟶ Communities

- Local clustering ⟶ Average clustering coefficient C
- Shortest path for dyad ⟶ Average path length L



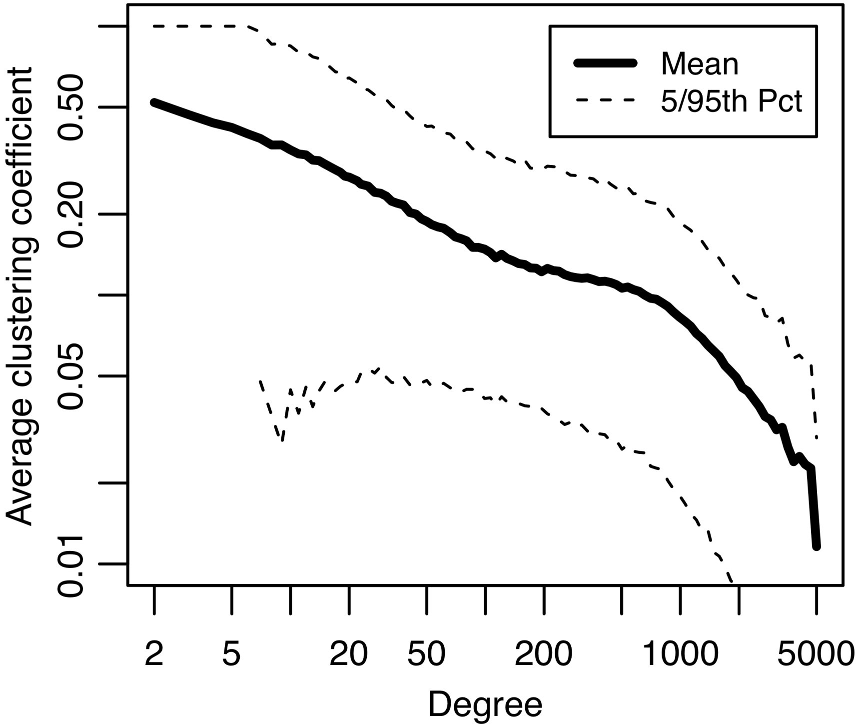


C and L in a real-world social network

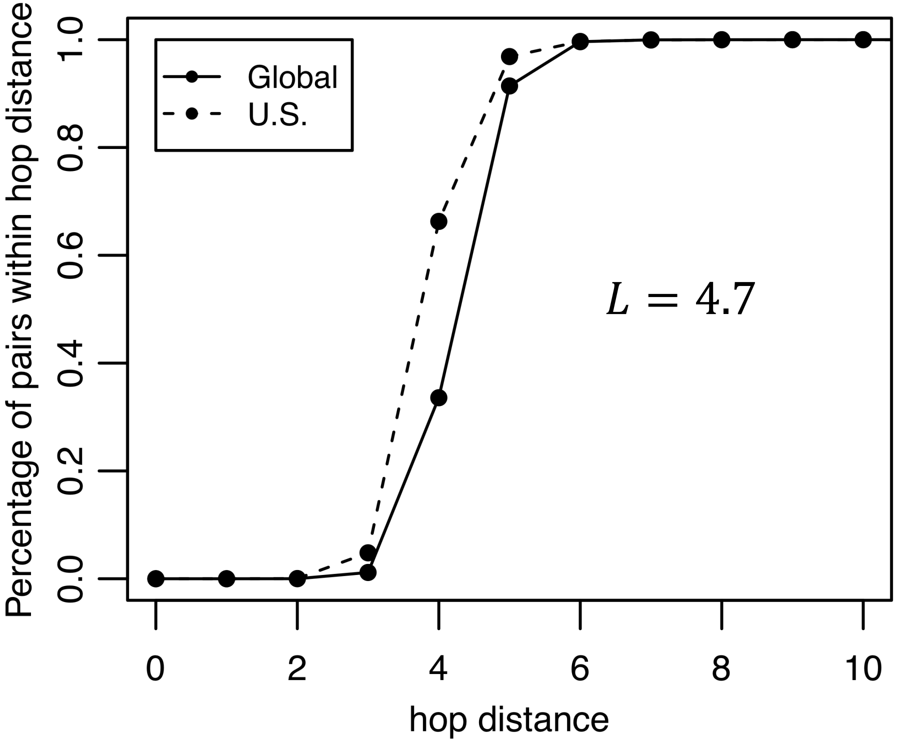
Source: https://www.wired.com/2011/11/facebook-social-graph-study/

C and L in the Facebook social network

- 721 million active Facebook users in May 2011



It's a small world!



Ugander, J., Karrer, B., Backstrom, L., & Marlow, C. (2011). The anatomy of the Facebook social graph. *arXiv preprint arXiv:1111.4503*.

Source: http://www.famouspsychologists.org/stanley-milgram/

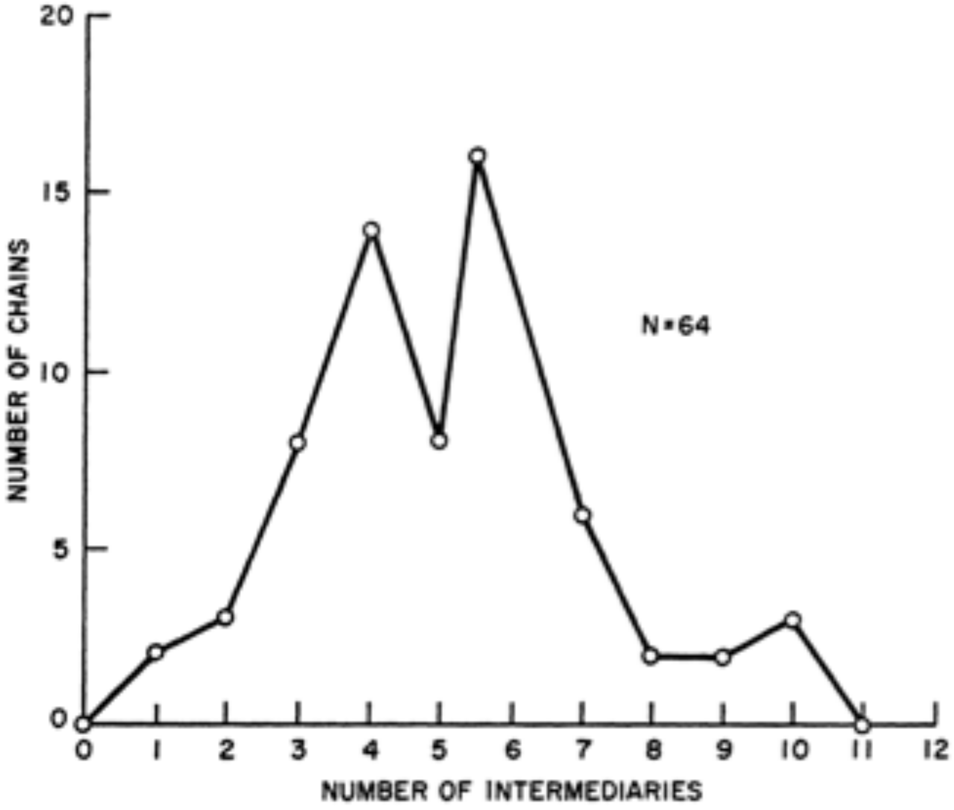
The small world experiment

- How long is the path between any two people?

- Selected 296 people in Omaha, NE and Wichita, KS
- Ask them to pass a letter to a stockbroker in Boston, MA
- Letter can be passed only to people they know by first name

Travers, J., & Milgram, S. (1969). An experimental study of the small world problem. *Sociometry*, 425–443.

Six degrees of separation!



- For the 64 completed chains, it took 6.2 steps on average

Travers, J., & Milgram, S. (1969). An experimental study of the small world problem. *Sociometry*, 425–443.

Small worlds affect social processes

Spread/diffusion/contagion:

- Contagious disease
- Innovations
- Information
- Fake news
- Collective action
- Health-related behavior
- Prosocial behavior
- ...

Statistical inference on networks

- Ego networks from large-scale surveys
  - Observations are independent and identically distributed
  - Can apply common statistical models e.g. linear regression
  - E.g., younger, more educated urbanites have larger networks (Marsden 1987)
- Whole networks
  - Network data is inherently relational
  - Observations are not independent
  - For example, a link could be explained with:
    - Ego's outgoingness (node property)
    - Alter's attractiveness (node property)
    - Homophily (dyad property)
    - Triadic closure (triad property)

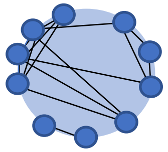
Marsden, P. V. (1987). Core discussion networks of Americans. *American Sociological Review*, *52*(1), 122–131.

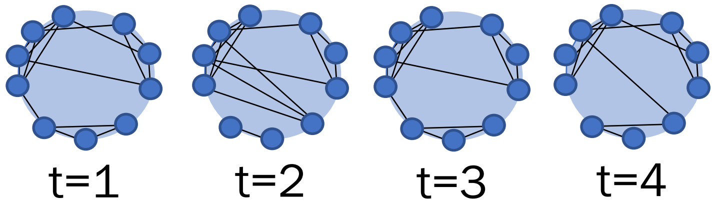
Statistical inference on networks

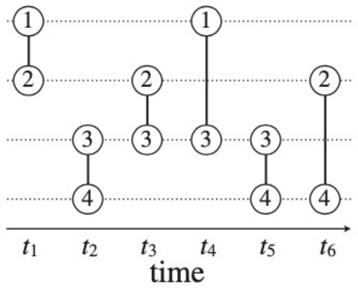
- In order to say what processes influence the formation of a network, we need a statistical model that considers the set of all possible alternative networks weighted on their similarity to the observed network
- Methods
  - Compare to a suitably randomized version of the network
    - E.g. rewire the edges but preserve the node degrees
    - Repeat multiple times
    - Estimate the observed metric to the expected distribution
  - Exponential random graph models
  - Stochastic actor-oriented models

Dynamic Networks

	Static Networks	Network Panel Data	Temporal Networks
Time	Single snapshot / Aggregate record	Multiple "snapshots"	Time-stamped observations
Ties	States / Events	States	Events
Sources	Name generators / Digital-trace data	Longitudinal surveys	Digital-trace data
Context	Everything so far*	Friendship, collaboration	Phone calls, e-mails, Bluetooth proximity
Examples	Everything so far*	[Add Health]()	[SocioPatterns]()

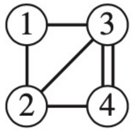






Coarse aggregation ⟷ Fine event detail

* In practice, many static networks are simply single snapshots of many events aggregated over time



Summary: What is Social Network Analysis?

- Networks present a paradigm to approach social structure and phenomena
- The focus is on **relationships and interactions between actors**
- Our relationships and interactions with others define who we are
- Social network analysis uses methods from matrix algebra, statistical inference, graph theory, physics, and computer science to describe and explain social interaction mechanisms, patterns, and dynamics

- In the afternoon: How do we analyze social networks in R?

Further SNA resources

- Collection of literature, datasets, courses, and many other resources related to network analysis by François Briatte: https://github.com/briatte/awesome-network-analysis
- Textbook on egocentric network analysis in R by Raffaele Vacca: https://raffaelevacca.github.io/egocentric-r-book/
- Tutorials on network analysis and visualization by Katya Ognyanova: https://kateto.net/tutorials/

MY461 Social Network Analysis

- Lent Term
- Lecturers: Eleanor Power, Milena Tsvetkova
- No prerequisites
- Content
  - Read key papers in the development of social network analysis
  - Cover essential measures and methods of network analysis
  - Introduce network modelling, analysis, and visualisation using R
- Auditors are welcome to attend lectures and seminars, if there’s room
  - Sign up at https://www.lse.ac.uk/Methodology/Study/Auditing