

*Statistical Analysis of Network Data*

---

# Introduction to Network Analysis

---

Jacob Young

[jacob.young.1@asu.edu](mailto:jacob.young.1@asu.edu)

<https://jacobtnyoung.github.io/SAND>



# Why Networks?



# Why Networks?

- ❖ In 1989, the German Democratic Republic (aka “East Germany”) collapsed.
- ❖ What happened?





---

# Why Networks?

---

- ❖ Understanding exit / voice theory
  - ❖ When people have grievances, they can: leave (exit) or express grievances (voice)
  - ❖ What happened in the GDR?
    - ❖ Exit was not an option after 1961 (visa application sanctions; a wall!)
    - ❖ So people used voice...not for nearly 30 years!
      - ❖ *Why?*

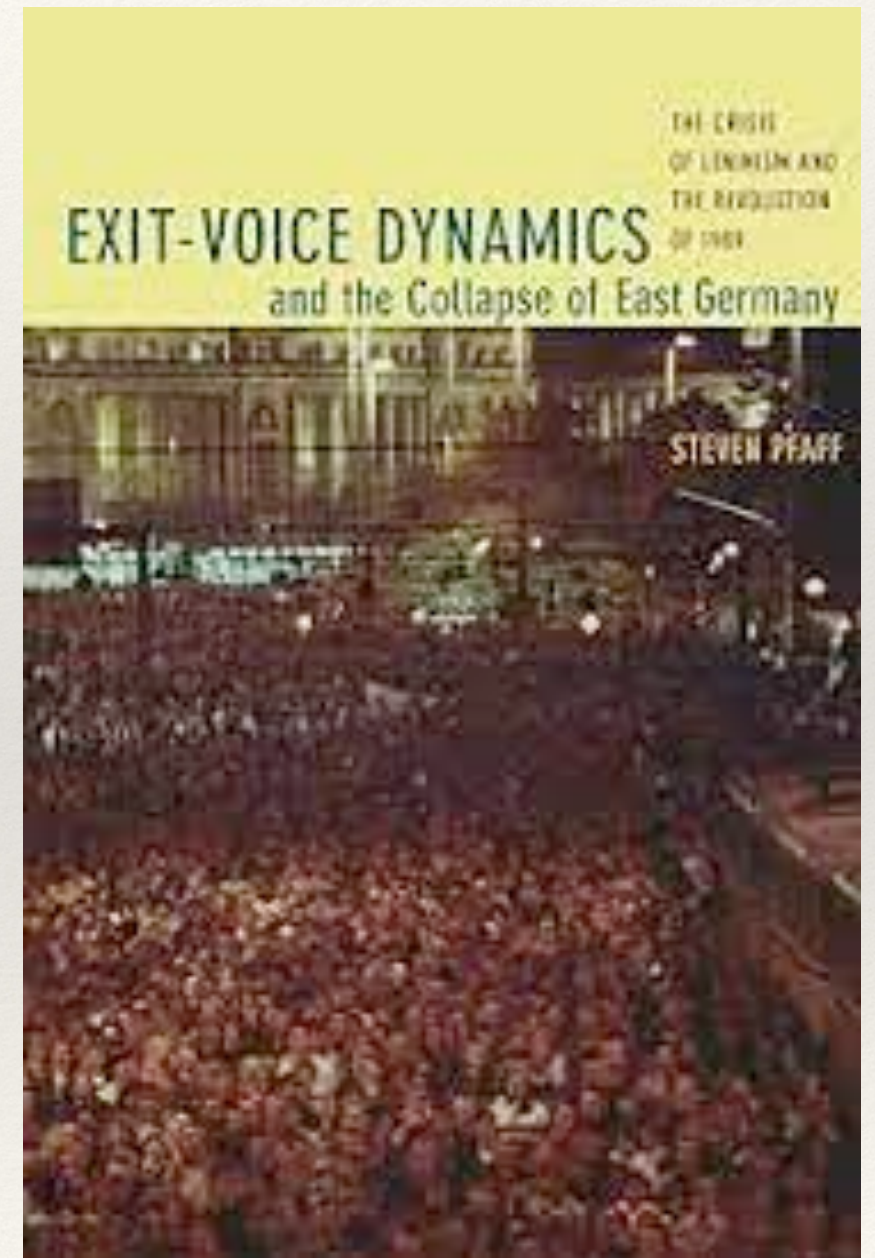


---

# Why Networks?

---

- ❖ Stasi as a surveillance tool led to a “niche society” where individuals discuss grievances, but only with close, trusted friends



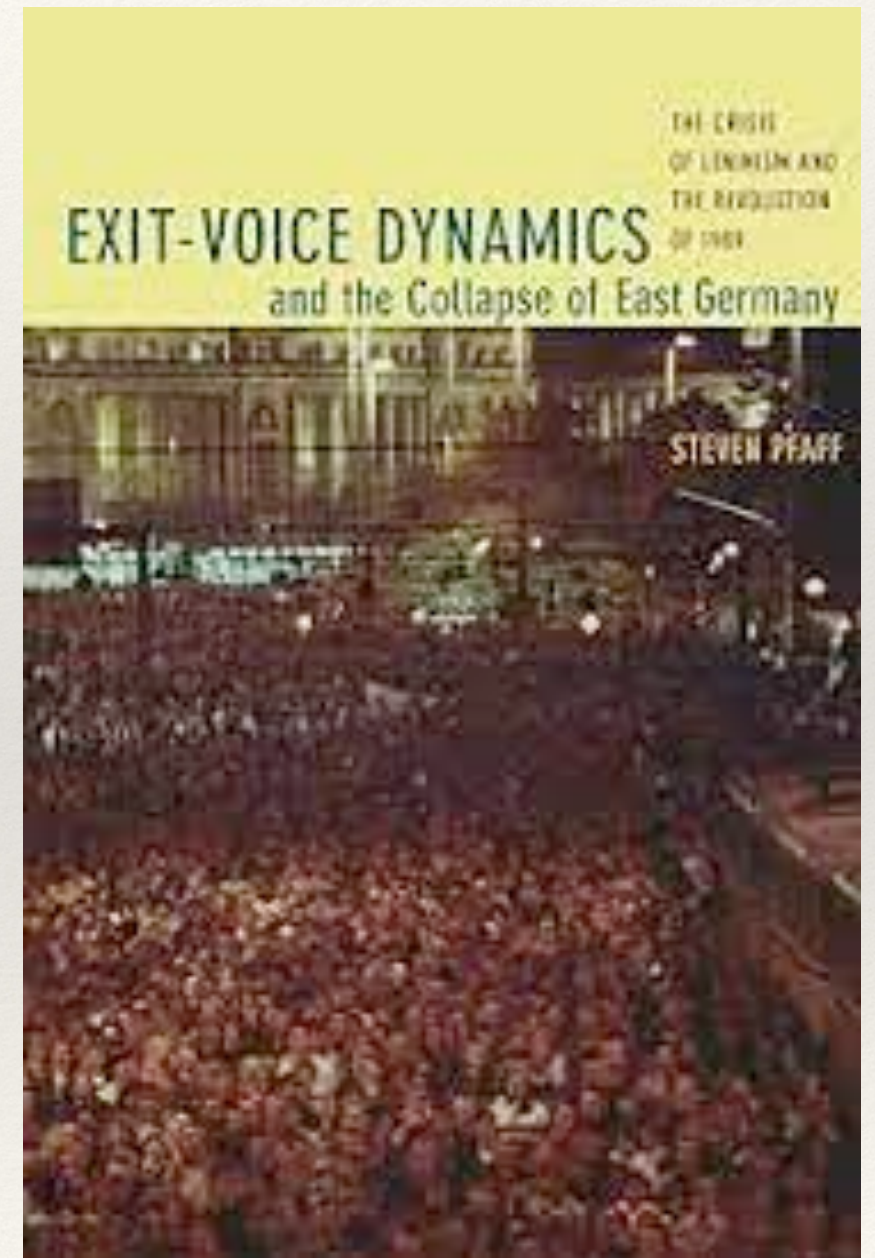


---

# Why Networks?

---

- ❖ Eventually, the Lutheran Church would serve as a place for individuals to discuss grievances.
- ❖ Solved the underlying **coordination problem**



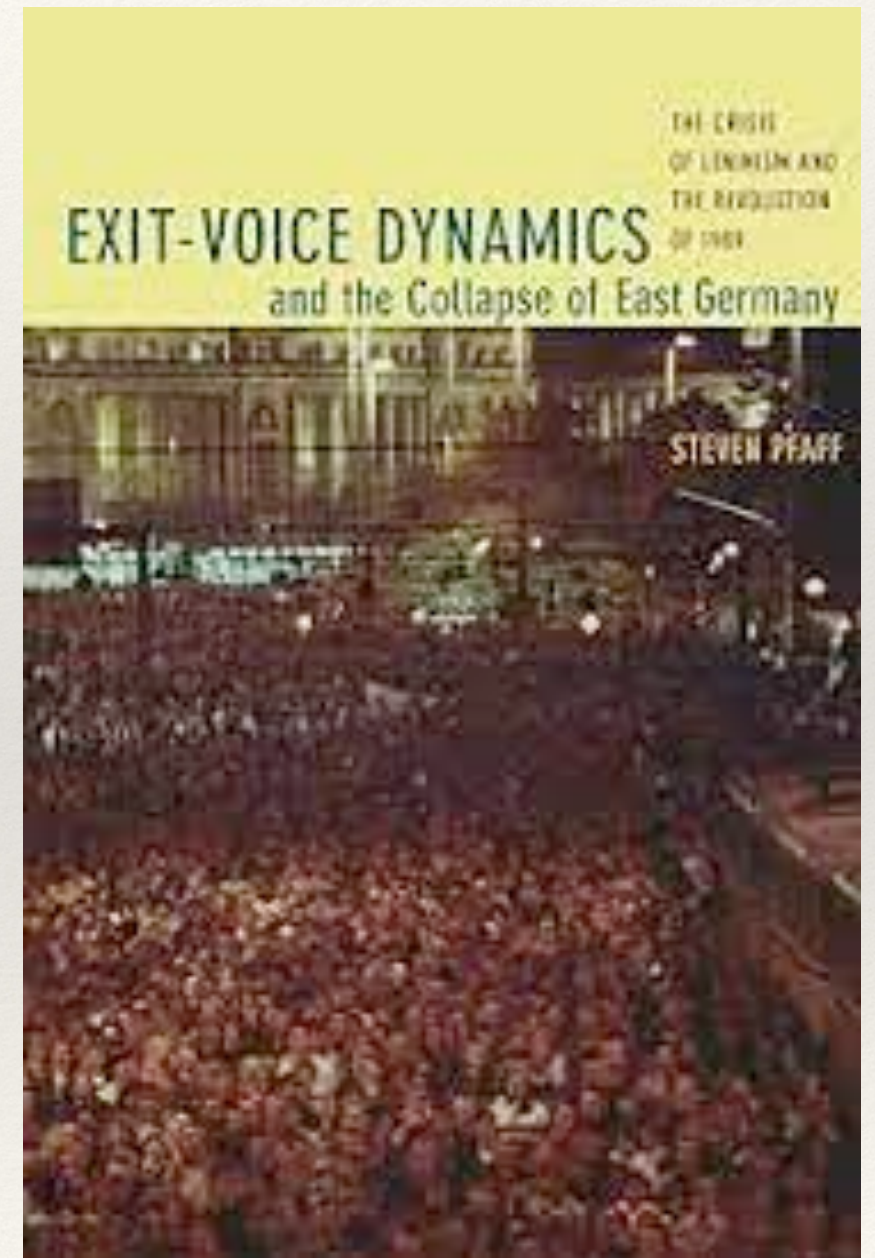


---

# Why Networks?

---

- ❖ “we usually expect revolutions to usher in a radical regime through violent means, in the GDR mass demonstrations forced the old regime from power almost entirely without violence” (p. 3)





So what?

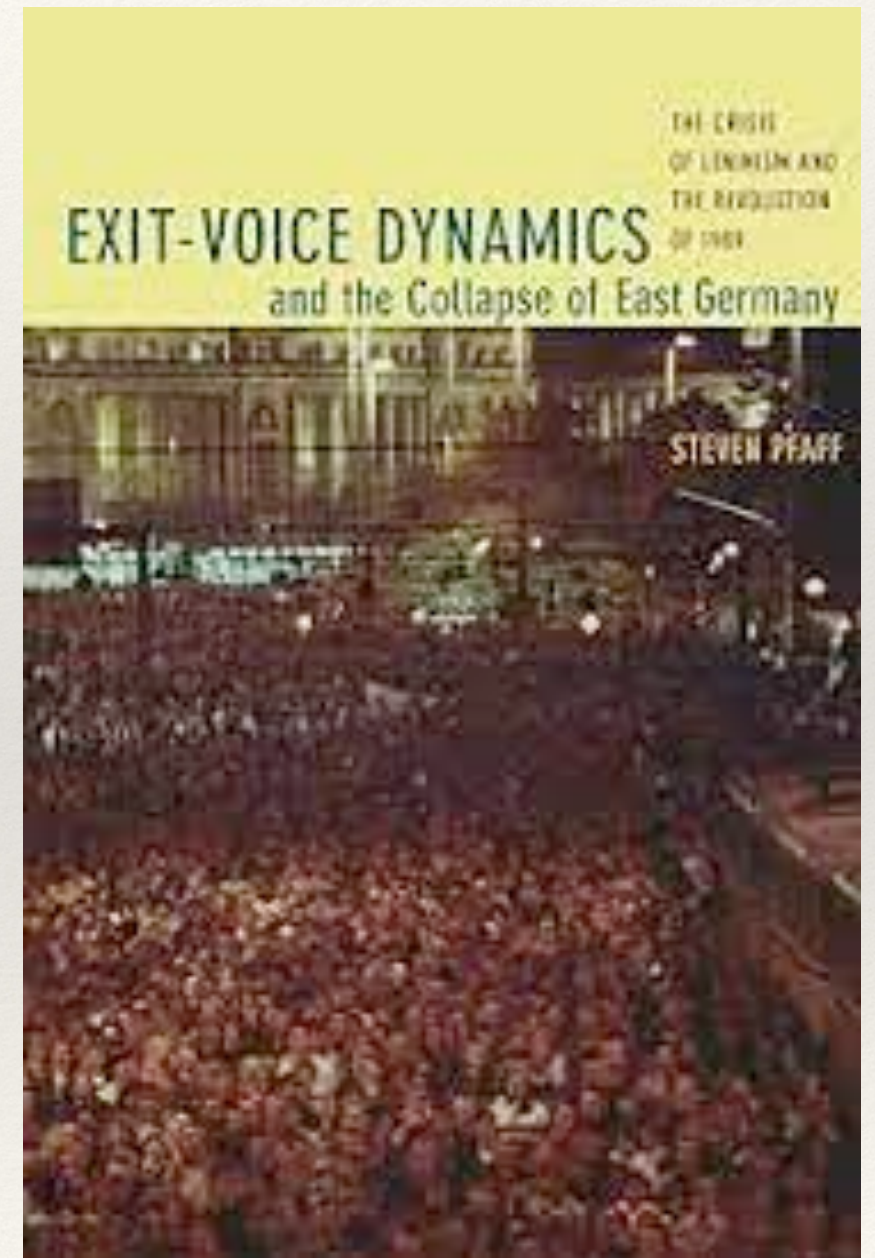


---

# Why Networks?

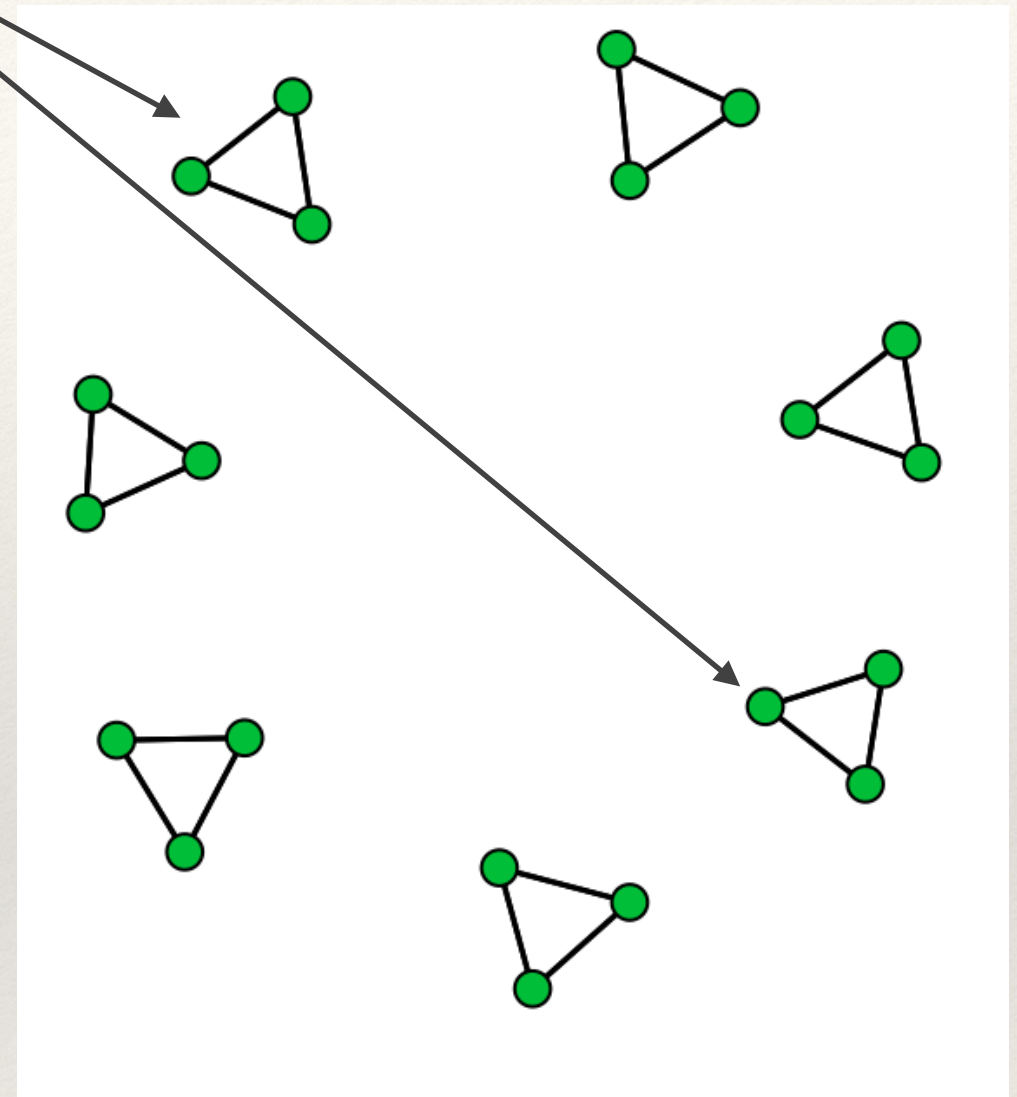
---

- ❖ This was a relational problem.
  - ❖ Not about individuals, it was about social relationships.
  - ❖ To understand this, we can reformulate the argument using pictures!





Tight-knit groups  
discussion concerns

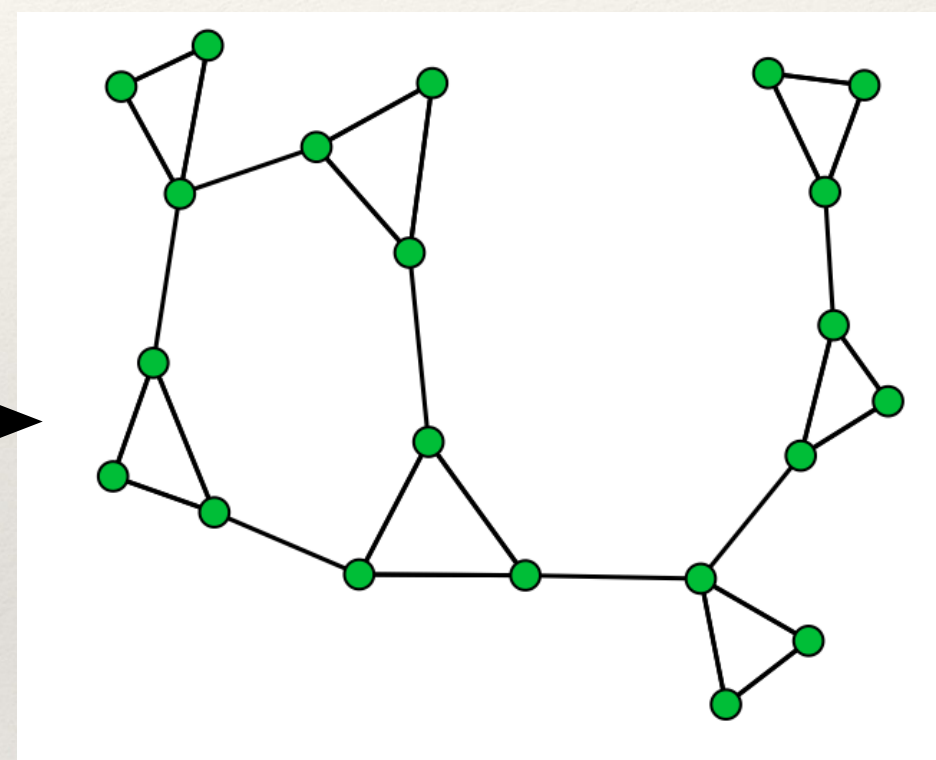
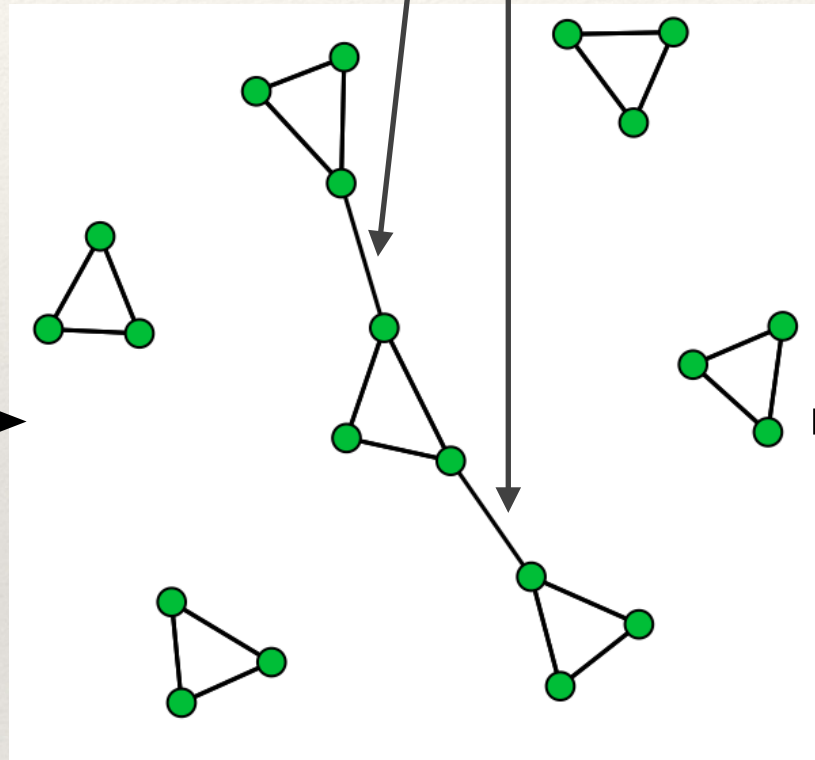
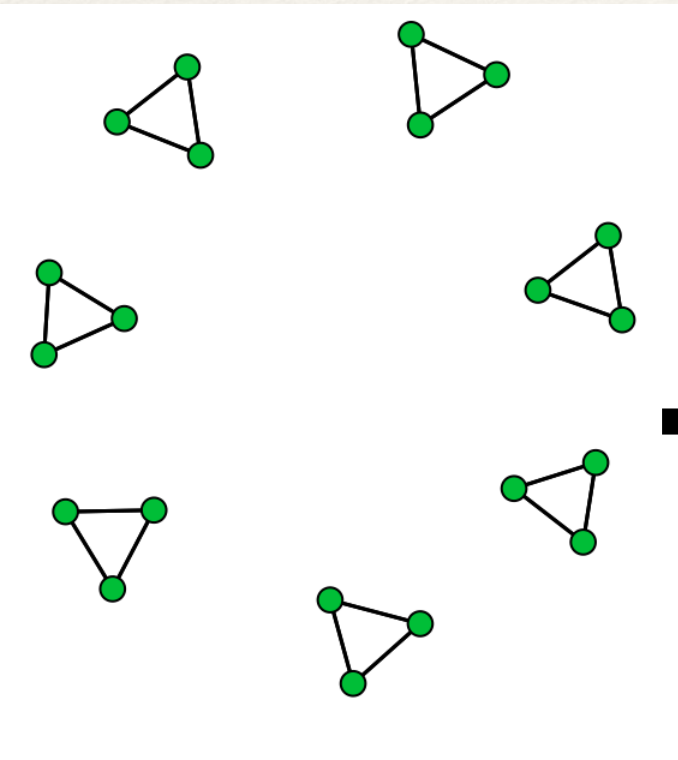


*But, what is the problem?*



Tight-knit  
groups  
discussion  
concerns

Church relations  
connect groups



*So what happens?*



---

# Why Networks?

---

- ❖ Many research problems (and solutions!) in the world are **relational** in nature.
- ❖ To answer questions of this sort, we need a specific set of tools.
- ❖ Those tools are what we call *network analysis*, broadly.



*Statistical Analysis of Network Data*

---

# Introduction to Network Analysis

---

Jacob Young

[jacob.young.1@asu.edu](mailto:jacob.young.1@asu.edu)

<https://jacobtnyoung.github.io/SAND>



---

# Learning Goals

---

- ❖ Be able to answer these questions:
  - ❖ What is “network science” and how is it different from “usual” research?
  - ❖ What do networks “look like”?
  - ❖ Where do network data “come from”?



---

# Network Science

---

- ❖ Network **science** is an approach to science that views the world as being composed of systems of actors connected through relational ties (i.e. a **network**).
- ❖ *What are some ways people can be connected?*



---

# Network Science

---

- ❖ Network science takes these *relational structures* as the primary domain of interest:
  - ❖ How does the network matter? (explanans/IV)
  - ❖ What effects (affects?) the network? (explanandum/DV)



---

# Network Analysis

---

- ❖ Network **analysis** is the set of tools used to study *relational variables*.
- ❖ A set of methods for systematically understanding and identifying connections among actors.



NEWS IN BRIEF

# Sudden Death Of Aunt Creates Rupture In Family Gossip Pipeline

11/23/15 8:45am • SEE MORE: LOCAL ▾



<https://local.theonion.com/sudden-death-of-aunt-creates-rupture-in-family-gossip-p-1819578447>



# Sudden Death Of Aunt Creates Rupture In Family Gossip Pipeline

VIRGINIA BEACH, VA—Grieving family members of local aunt Laurie Shelton confirmed Monday that the 48-year-old woman's unexpected death had caused a major breach in their gossip pipeline, suddenly disrupting access to the latest dirt on all their relatives. "Since Aunt Laurie passed, news about how Stephanie's new boyfriend can't hold down a job and updates on Uncle Jeff's gambling habit have slowed to a trickle," said Shelton's niece Arielle, mourning the loss of a woman who for years had reportedly ensured a steady stream of the juiciest tidbits about relatives' layoffs, unplanned pregnancies, personal bankruptcies, and misdemeanor shoplifting charges. "All the best gossip flowed through her, and now she's gone. For all I know, the twins in North Carolina could have been caught smoking pot, Grandma could be back together with Leon, and Uncle Mike could be considering a vasectomy. It's a devastating loss for the whole family." Several in the family expressed hope that, for the time being, a sufficient supply of idle chatter could be rerouted through Cousin Staci to meet their immediate needs.



## Sudden Death Of Aunt Creates Rupture In Family Gossip Pipeline

VIRGINIA BEACH, VA—Grieving family members of local aunt Laurie Shelton confirmed Monday that the 48-year-old woman's unexpected death had caused a major breach in their gossip pipeline, suddenly disrupting access to the latest dirt on all their relatives. "Since Aunt Laurie passed, news about how Stephanie's new boyfriend can't hold down a job and updates on Uncle Jeff's gambling habit have slowed to a trickle," said Shelton's niece Arielle, mourning the loss of a woman who for years had reportedly ensured a steady stream of the juiciest tidbits about relatives' layoffs, unplanned pregnancies, personal bankruptcies, and misdemeanor shoplifting charges. "All the best gossip flowed through her, and now she's gone. For all I know, the twins in North Carolina could have been caught smoking pot, Grandma could be back together with Leon, and Uncle Mike could be considering a vasectomy. It's a devastating loss for the whole family." Several in the family expressed hope that, for the time being, a sufficient supply of idle chatter could be rerouted through Cousin Staci to meet their immediate needs.

Conceptually, what does this story tell us about the **relational structure** of information transmission in the Shelton family?



## Sudden Death Of Aunt Creates Rupture In Family Gossip Pipeline

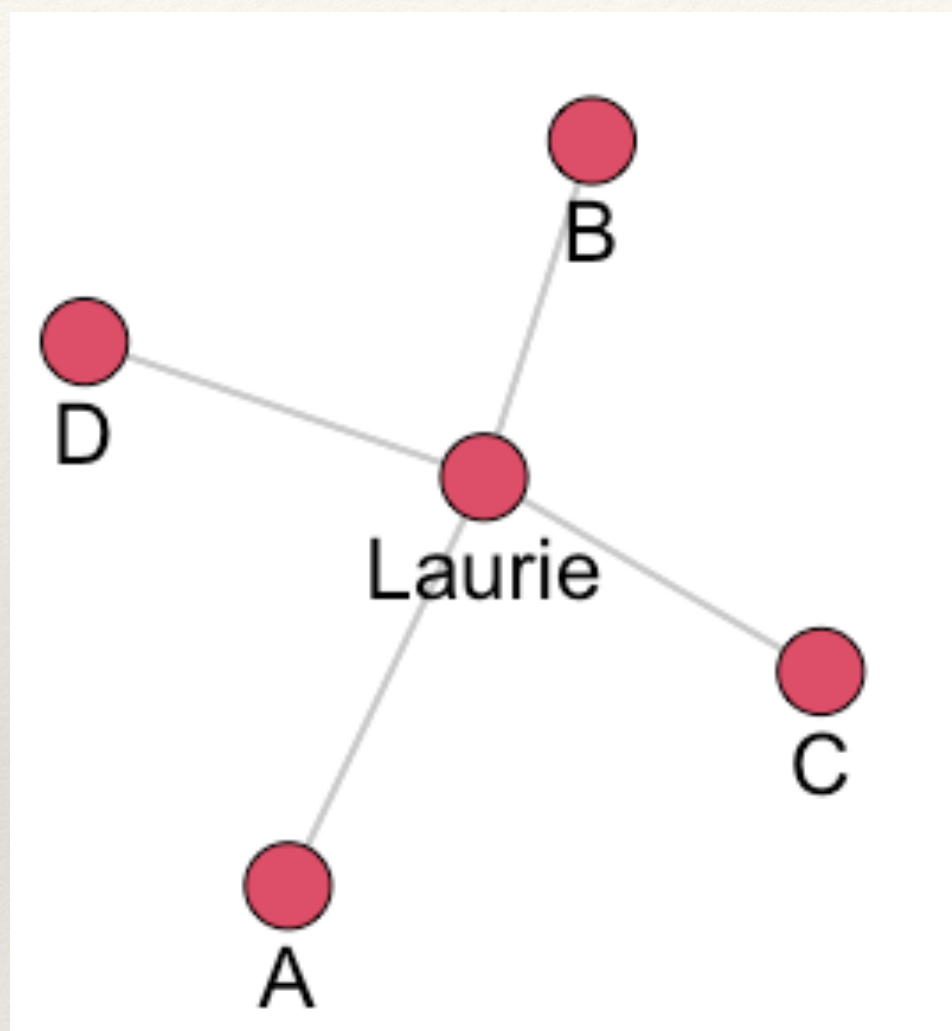
VIRGINIA BEACH, VA—Grieving family members of local aunt Laurie Shelton confirmed Monday that the 48-year-old woman's unexpected death had caused a major breach in their gossip pipeline, suddenly disrupting access to the latest dirt on all their relatives. "Since Aunt Laurie passed, news about how Stephanie's new boyfriend can't hold down a job and updates on Uncle Jeff's gambling habit have slowed to a trickle," said Shelton's niece Arielle, mourning the loss of a woman who for years had reportedly ensured a steady stream of the juiciest tidbits about relatives' layoffs, unplanned pregnancies, personal bankruptcies, and misdemeanor shoplifting charges. "All the best gossip flowed through her, and now she's gone. For all I know, the twins in North Carolina could have been caught smoking pot, Grandma could be back together with Leon, and Uncle Mike could be considering a vasectomy. It's a devastating loss for the whole family." Several in the family expressed hope that, for the time being, a sufficient supply of idle chatter could be rerouted through Cousin Staci to meet their immediate needs.

Conceptually, what does this story tell us about the **relational structure** of information transmission in the Shelton family?

*It is vulnerable...*

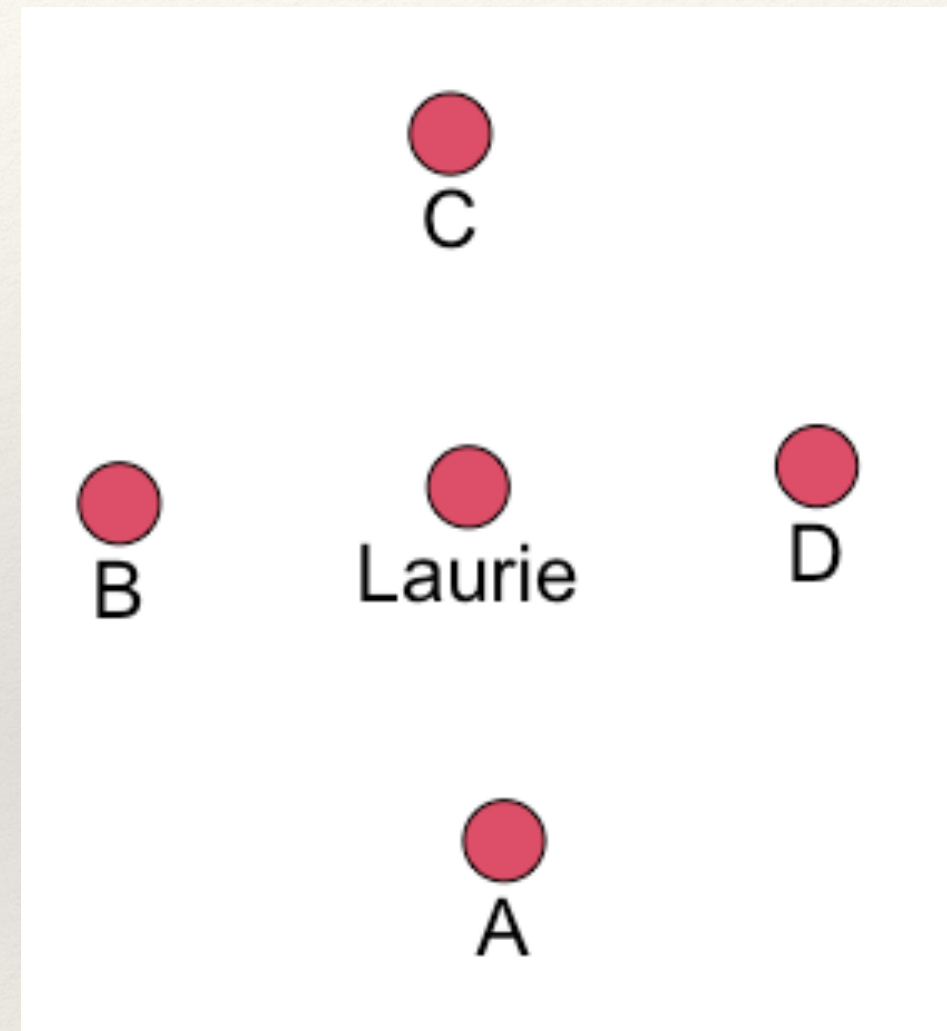
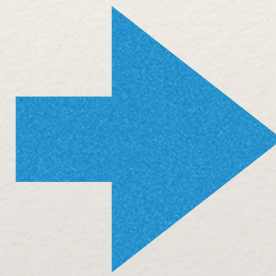
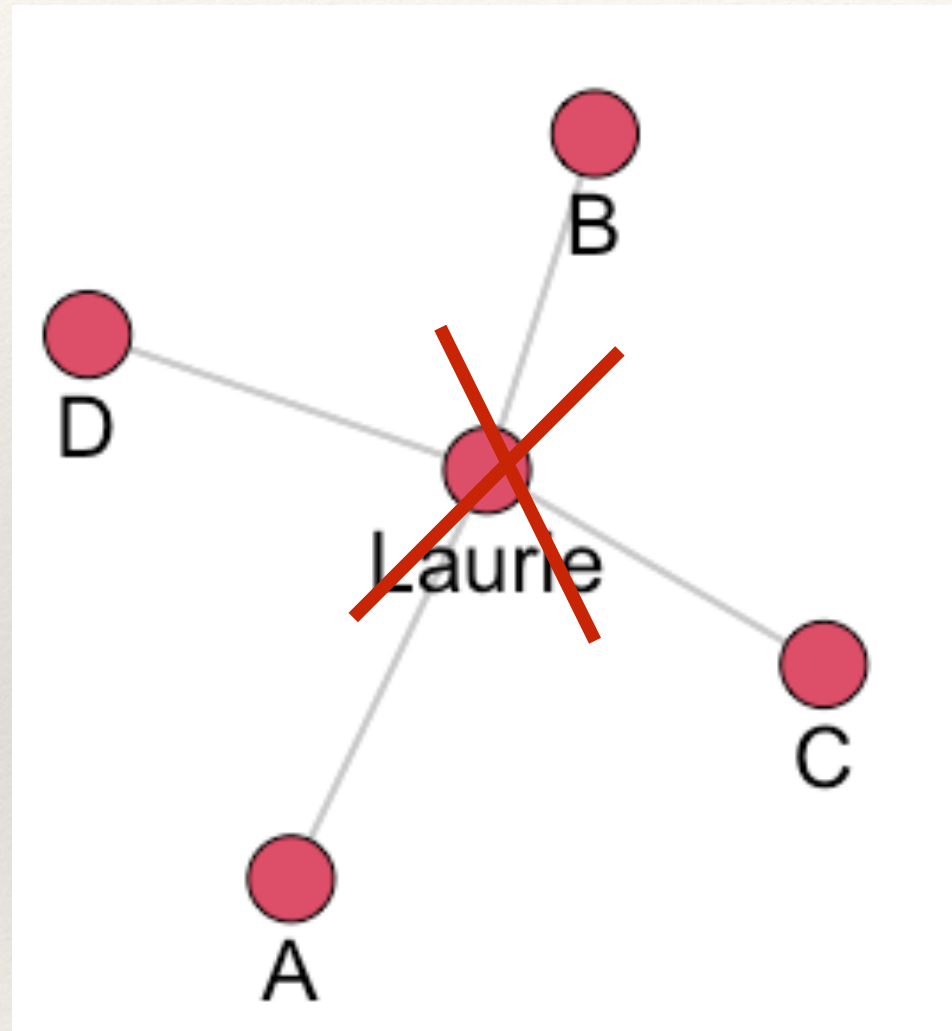


## Sudden Death Of Aunt Creates Rupture In Family Gossip Pipeline





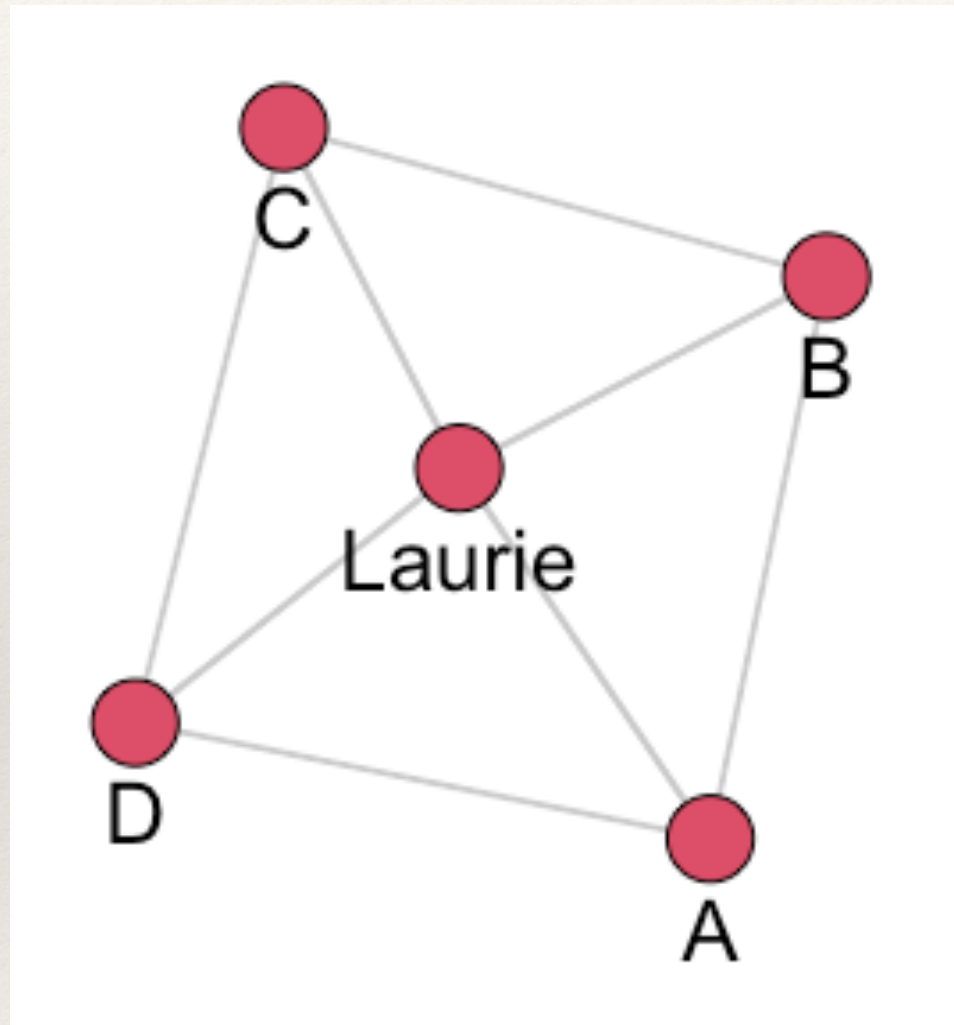
## Sudden Death Of Aunt Creates Rupture In Family Gossip Pipeline



*Why is it is vulnerable?*

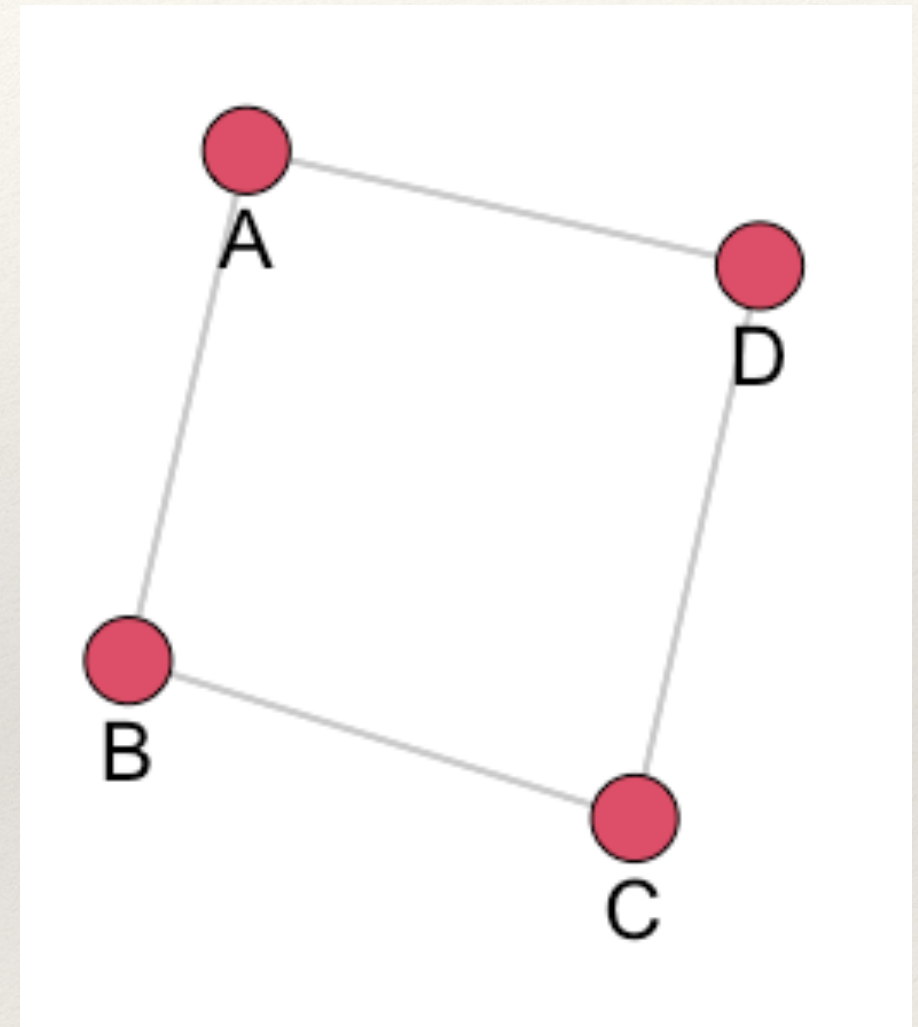
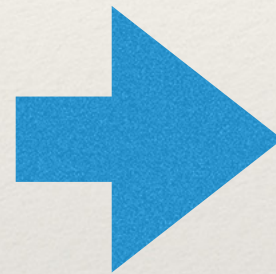
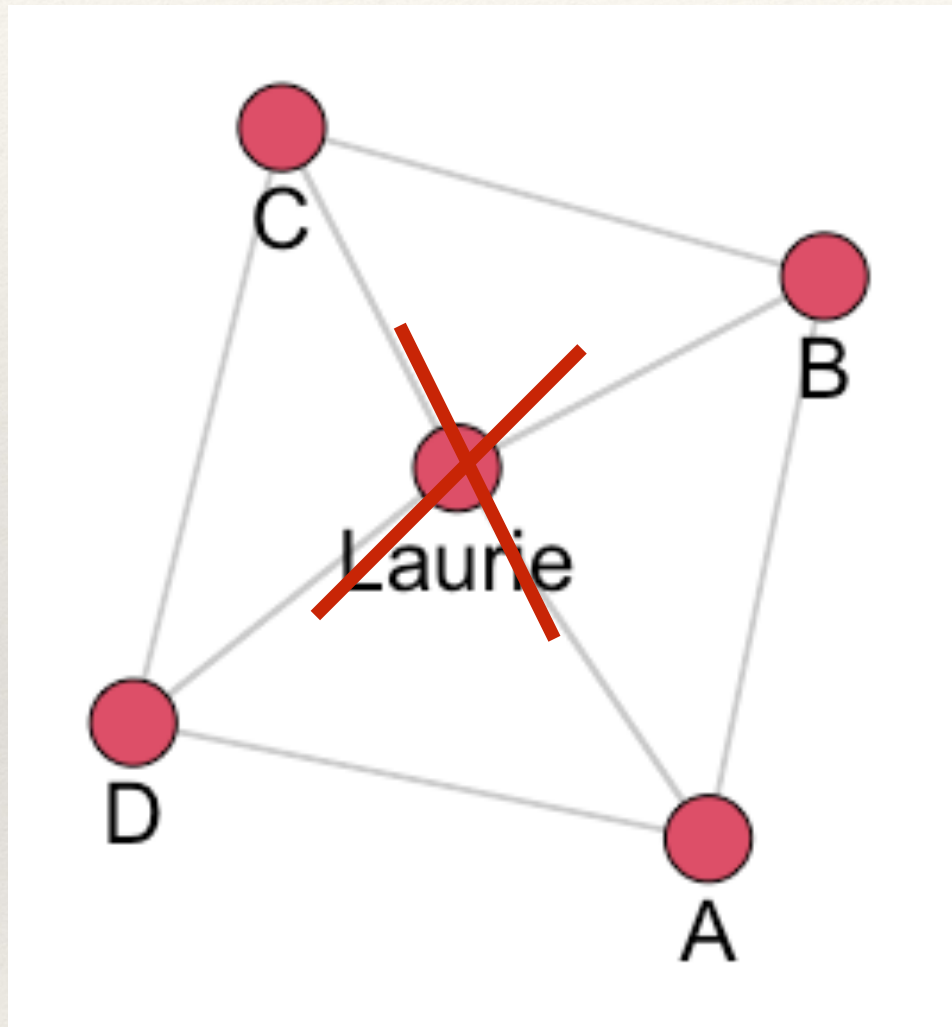


## Sudden Death Of Aunt Creates Rupture In Family Gossip Pipeline





# Sudden Death Of Aunt Creates Rupture In Family Gossip Pipeline



Why is it not vulnerable?



---

# The Point!

---

- ❖ These are topics that are inherently **relational**.



---

# Conceptualization & Operationalization

---

- ❖ Network science **conceptualizes** theoretical concepts that are inherently relational.
  - ❖ *Can you think of a relational theoretical concept?*
- ❖ Network research **operationalizes** theoretical constructs by drawing on the formal properties of graphs.
  - ❖ *Can you think of how that relational concept may be operationalized?*



---

# What "varies"?

---

- ❖ **Structural vs. Compositional variables**
  - ❖ The network itself is the variable of interest. It is a **structural** variable. When we have a matrix, that is the variable of interest. The variation in the 0s and 1s of the matrix (or whatever the values are).
  - ❖ Structural variables are measured on pairs of actors.
  - ❖ This is different from a **compositional** variable which is the nodal attributes.
    - ❖ Compositional variables are measurements of actor attributes.
- ❖ Think of the difference as "relational data" vs. "attribute data"



---

# Difference from “usual” research

---

- ❖ Network science:
  - ❖ Adopts a theoretical lens where causal processes are relational in nature
  - ❖ Rejects assumption of independence between units



# Basic Data Elements

*What do networks “look like”?*



---

# Basic Data Elements

---

- ❖ Network (relational) data represent:
  - ❖ **Connections** (aka ties, arcs, edges, lines, ties) among,
  - ❖ **Entities** (aka nodes, vertices, actors, points, dots).
- ❖ I will use *node* to mean **entities** and *edge* to mean **connections**.



---

# Basic Data Elements

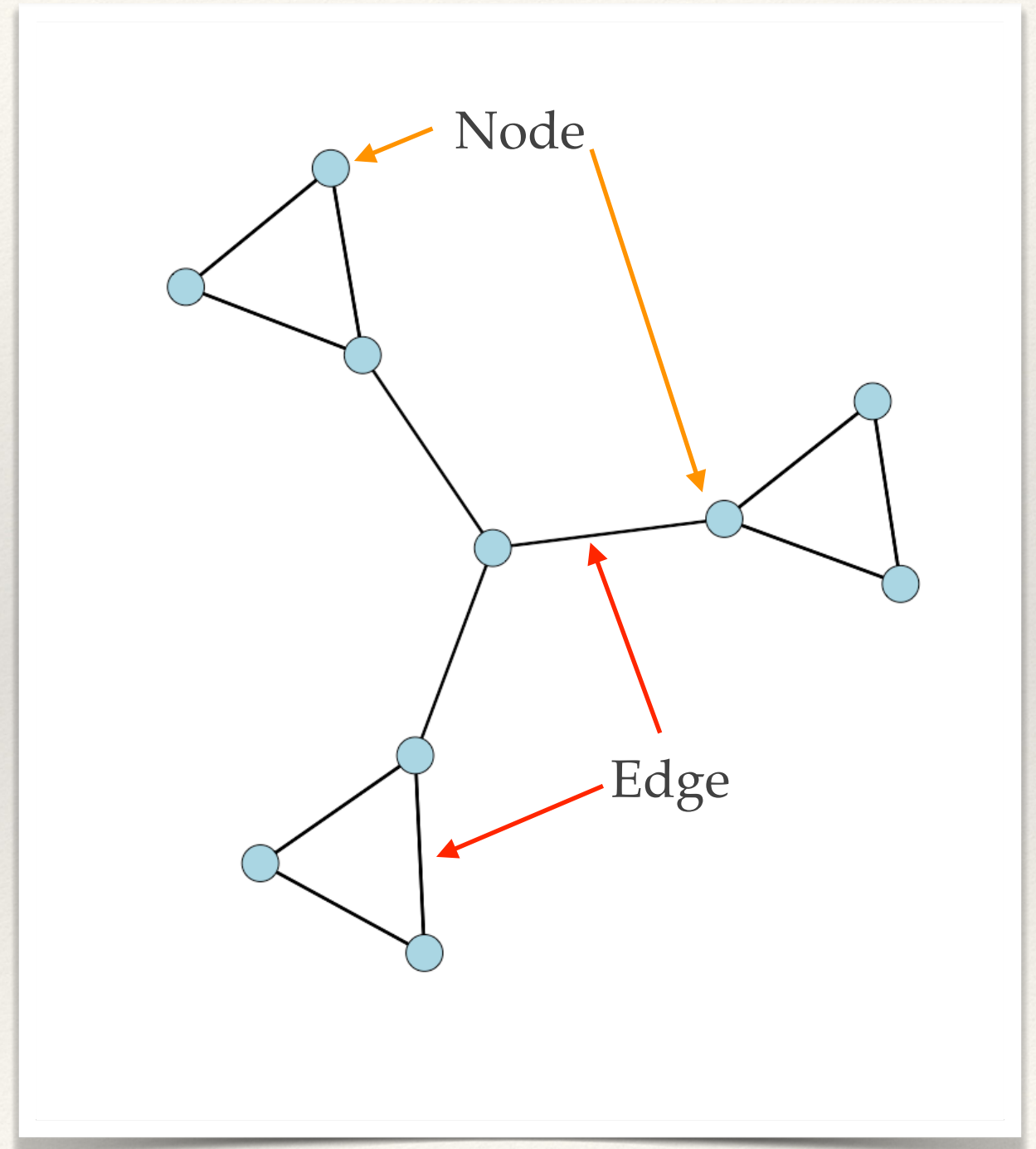
---

- ❖ *A node* can be anything that can link to something else.
- ❖ *An edge* can be anything that can record a connection between nodes.
- ❖ *What are some nodes and edges that come to mind?*



# Basic Data Elements

- ❖ On a graph, nodes are represented by *points* and edges are represented by *lines*.





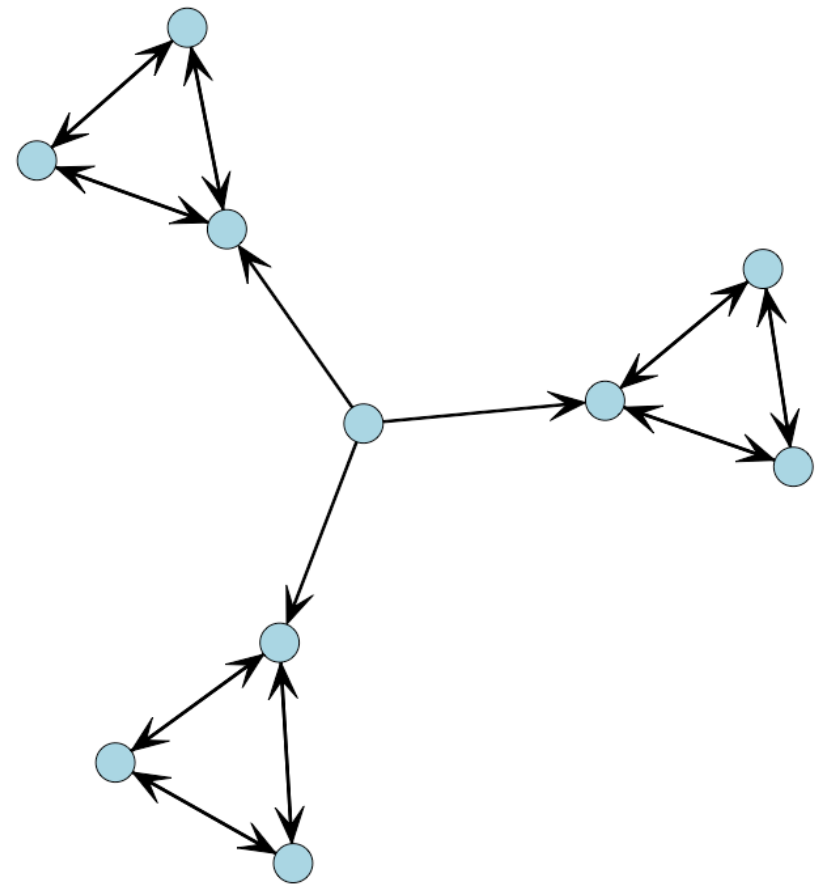
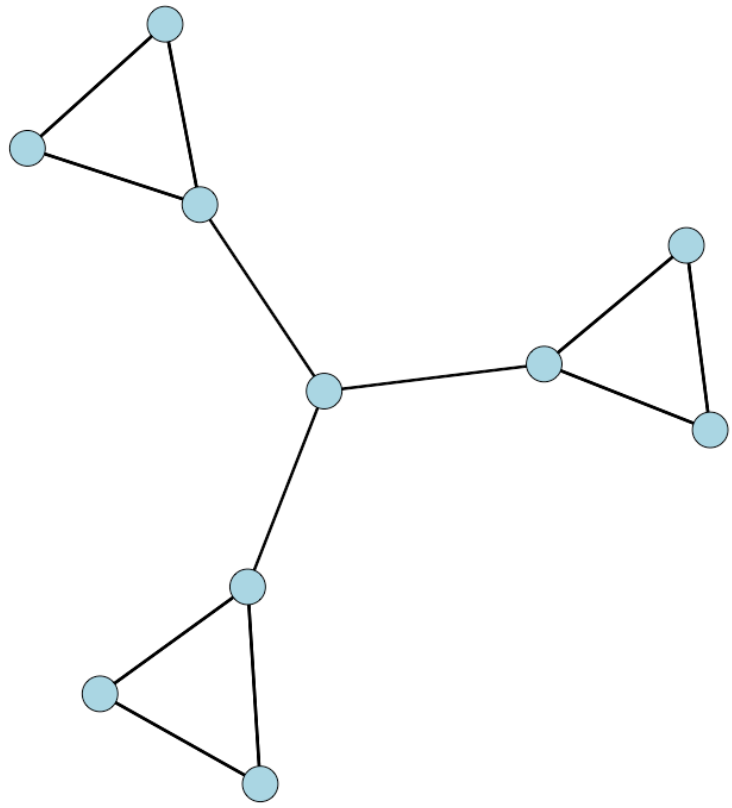
---

# Basic Data Elements

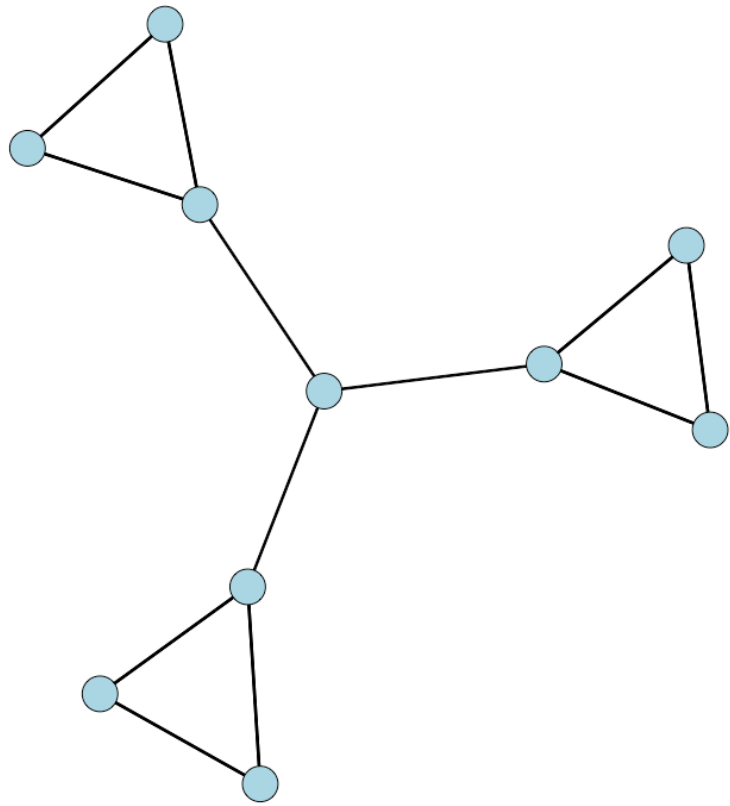
---

- ❖ Edges can be:
  - ❖ Directed or Undirected.



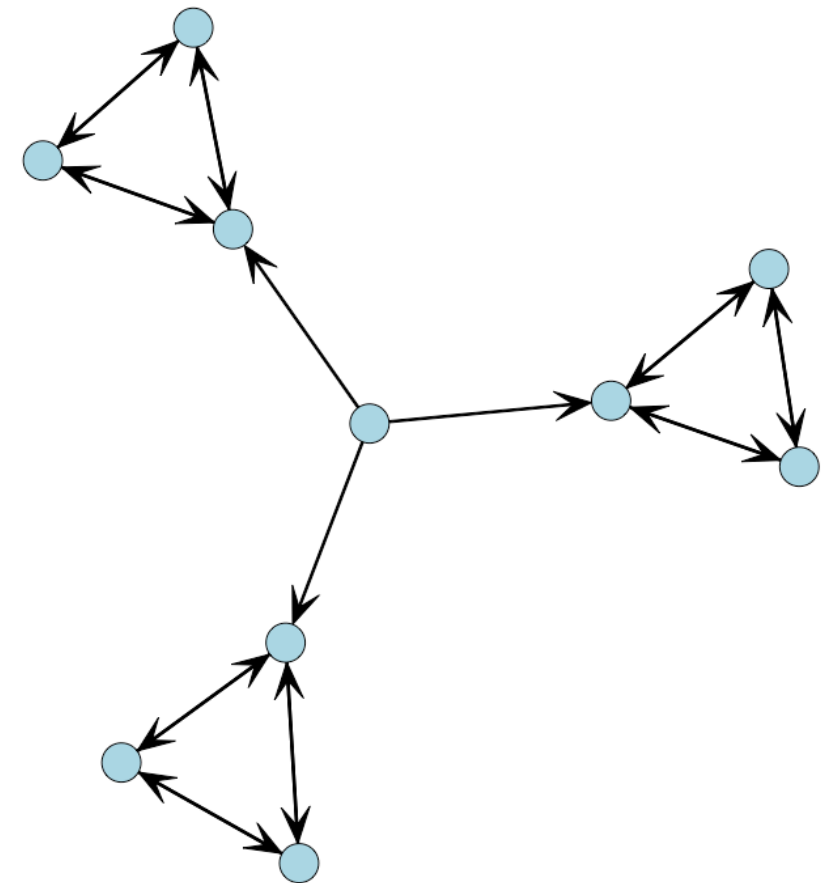






Suppose the edges  
measure  
communication...

*How are these  
structures  
different?*





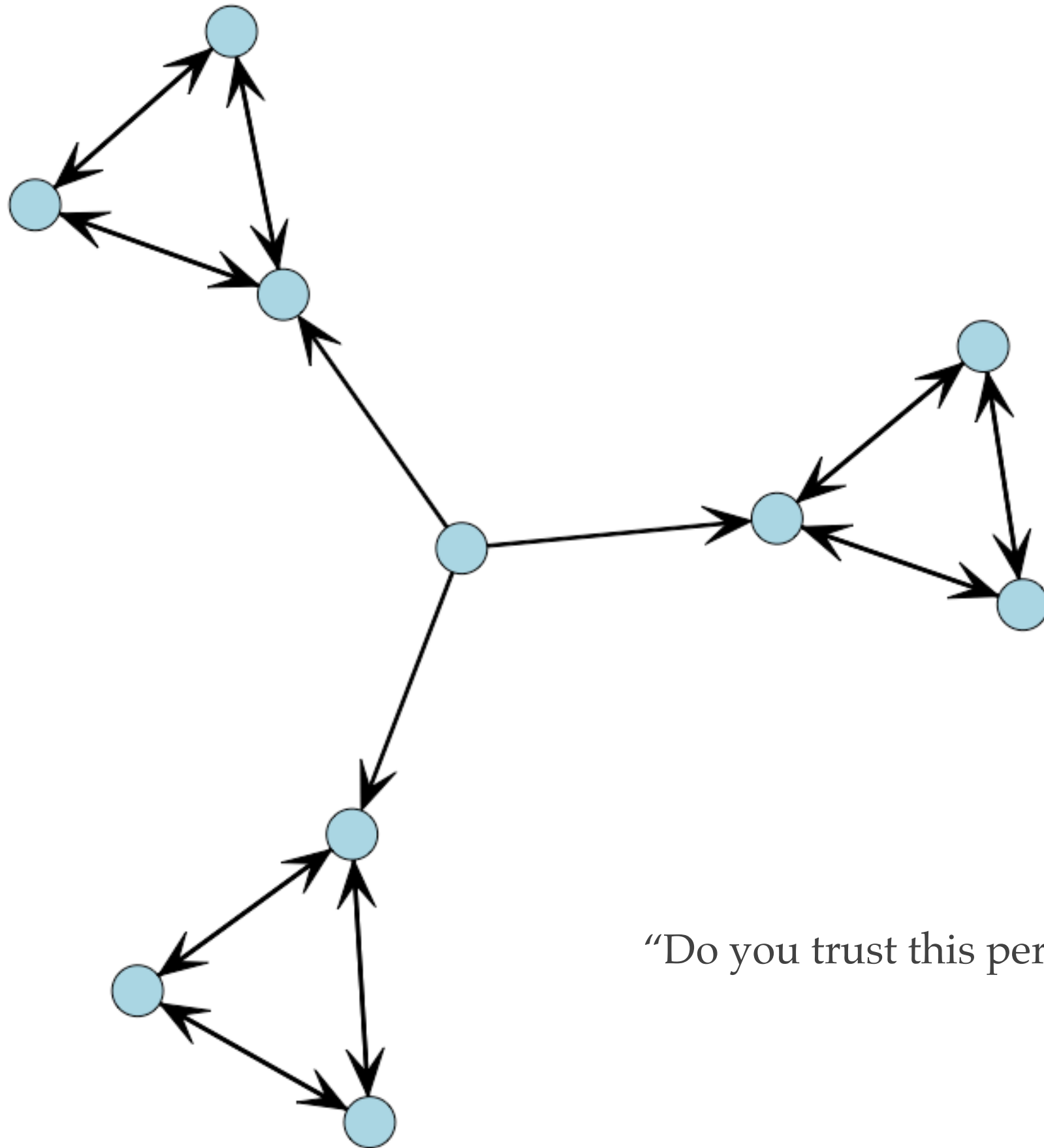
---

# Basic Data Elements

---

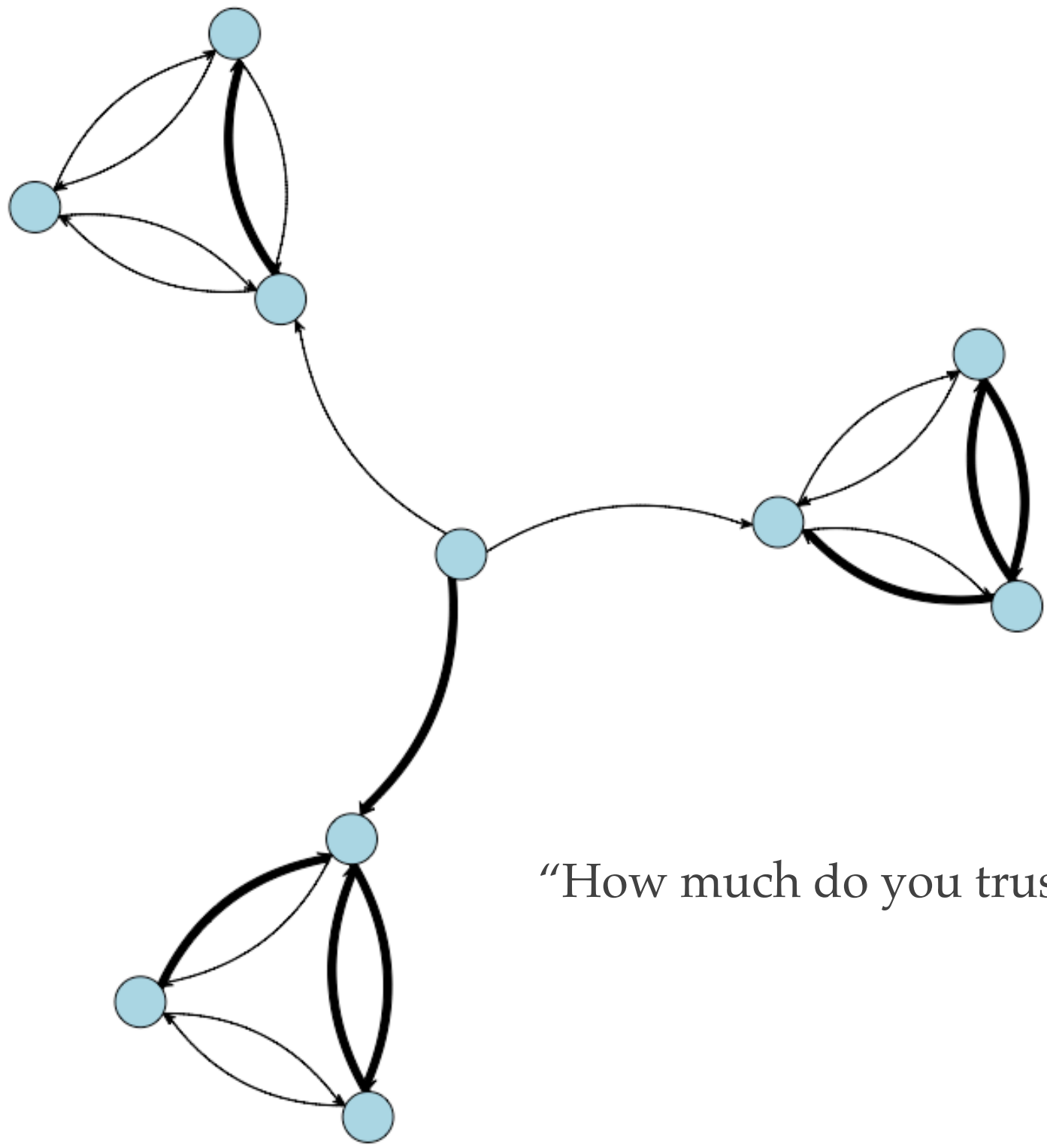
- ❖ **Edges** can be:
  - ❖ Binary (0 / 1; present / absent); Valued Integers (0 / 1 / 2...); Continuous Weights (0.24 / 1.76 / ...); Signed (+ / -).





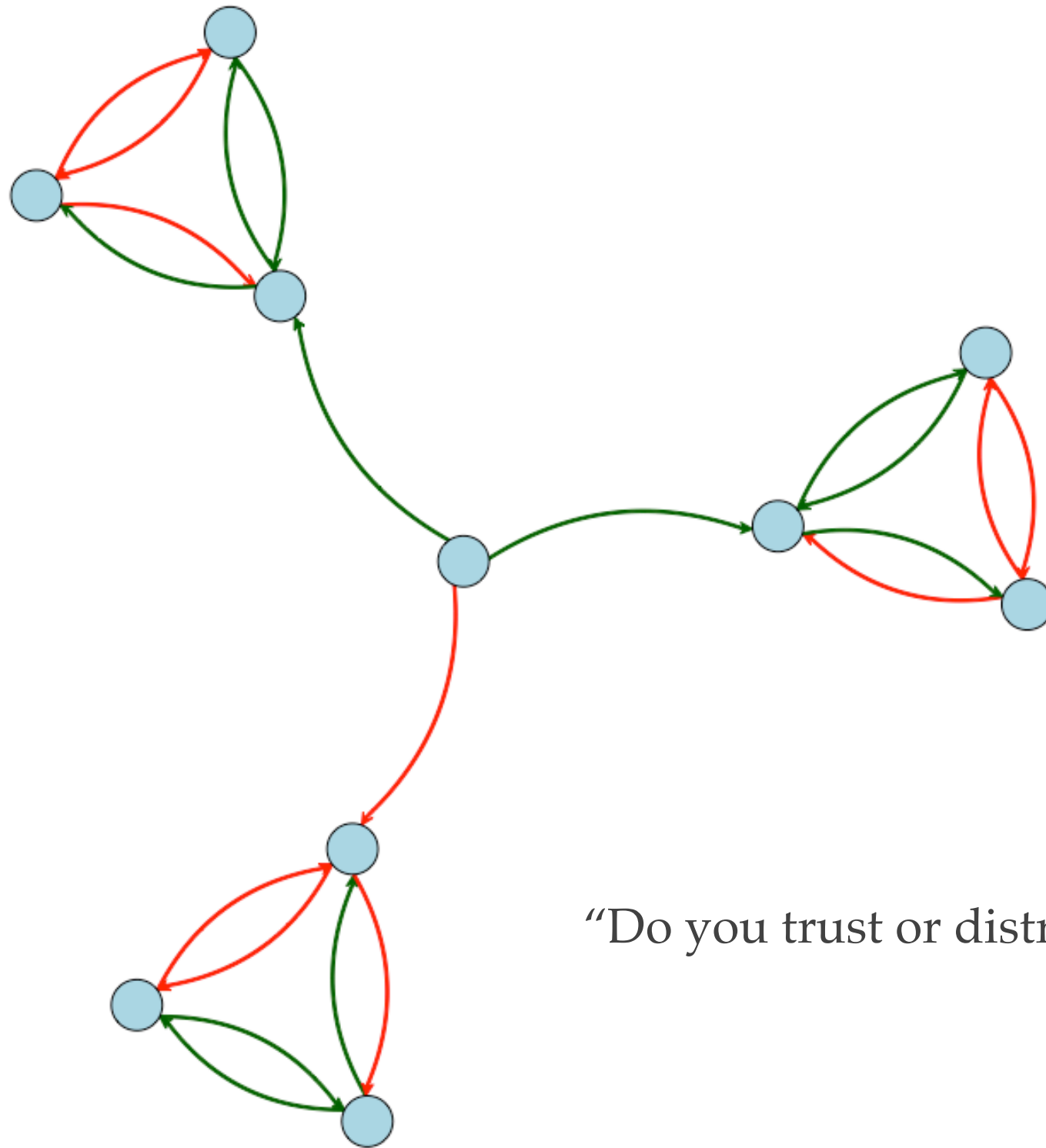
“Do you trust this person?”





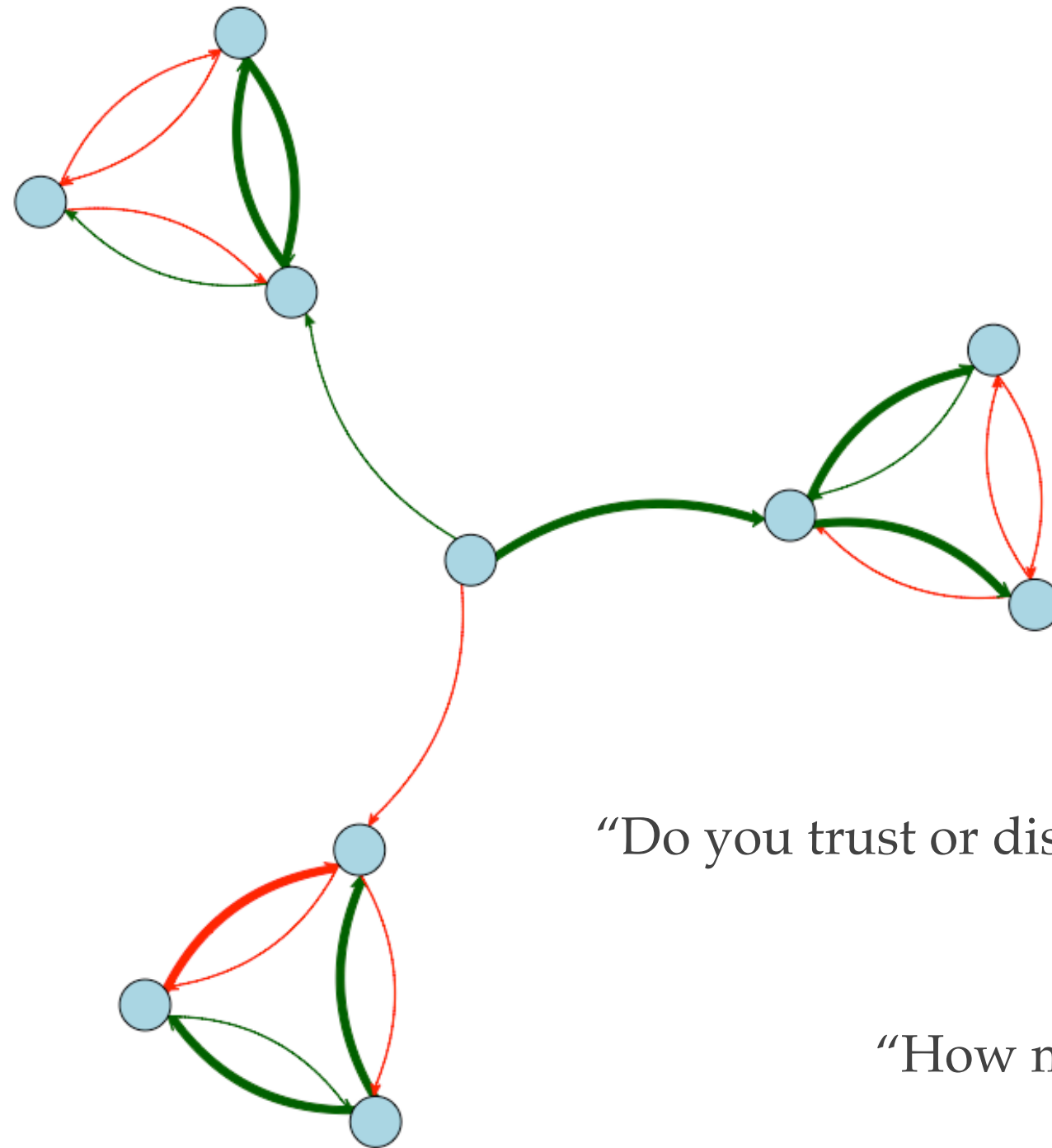
“How much do you trust this person?”





“Do you trust or distrust this person?”





“Do you trust or distrust this person?”

“How much?”



---

# Basic Data Elements

---

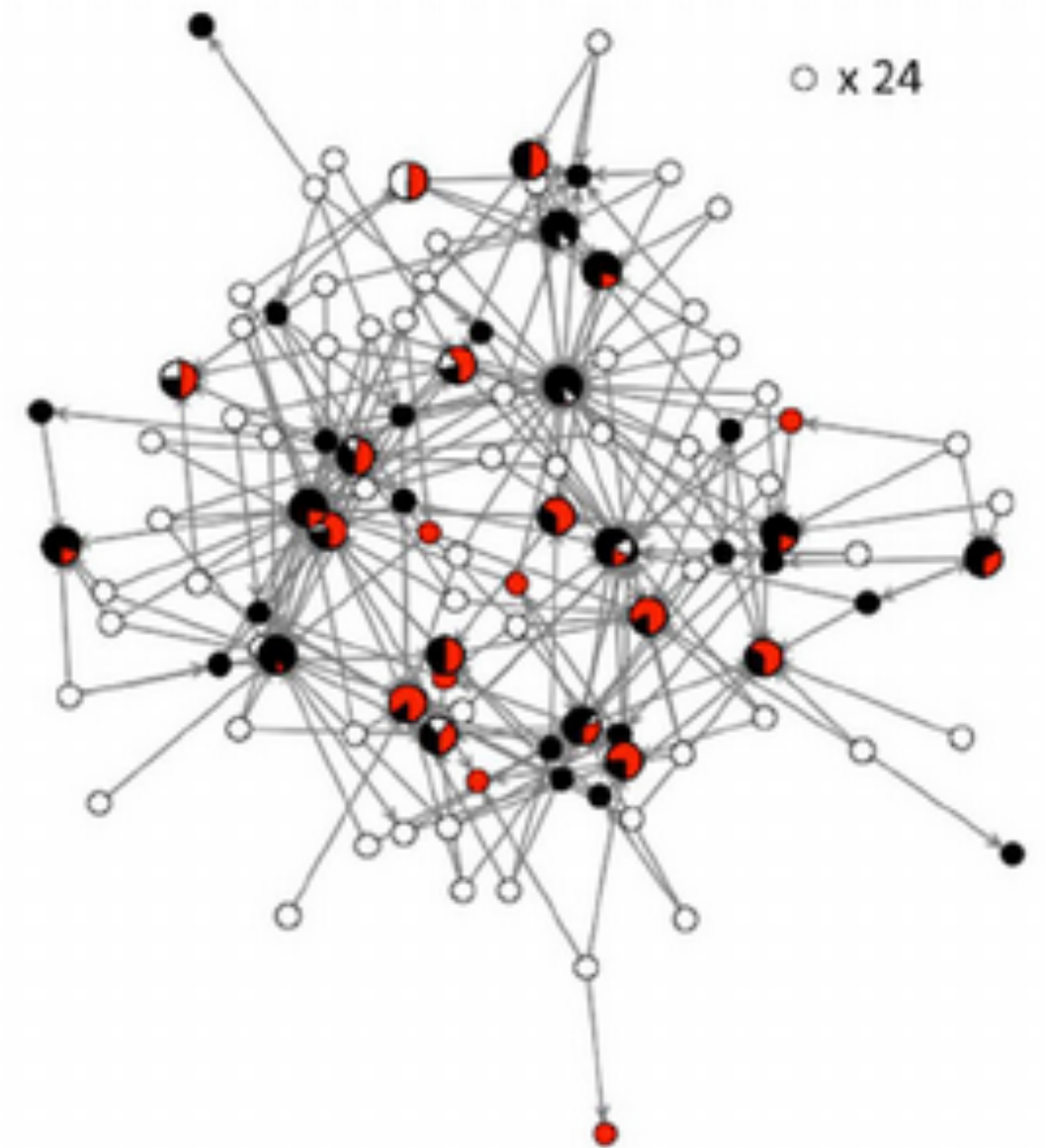
- ❖ **Edges** can have different meanings and therefore be of different *types*.
- ❖ Social relationships (sister, friend, likes, knows)
- ❖ Interactions (has sex with, talks to, seeks advice from)
- ❖ Flows (diseases, attitudes)



# In the eye of the beholder: Meaning and structure of informal status in women's and men's prisons\*

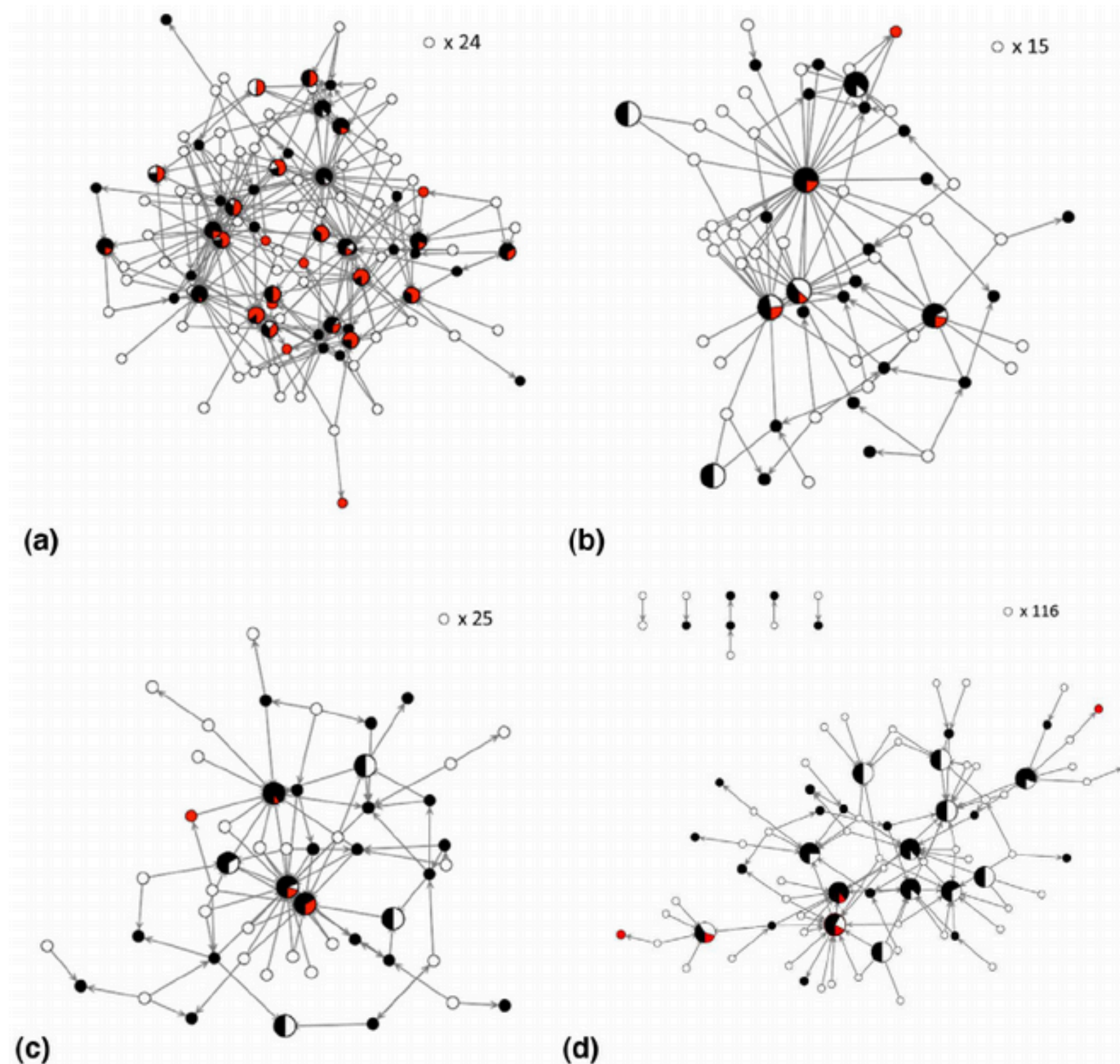
Derek A. Kreager<sup>1</sup> | Jacob T.N. Young<sup>2</sup> | Dana L. Haynie<sup>3</sup> |  
David R. Schaefer<sup>4</sup> | Martin Bouchard<sup>5</sup> | Kimberly M. Davidson<sup>1</sup>

Positive/Negative/Neutral  
Power nominations



(a)





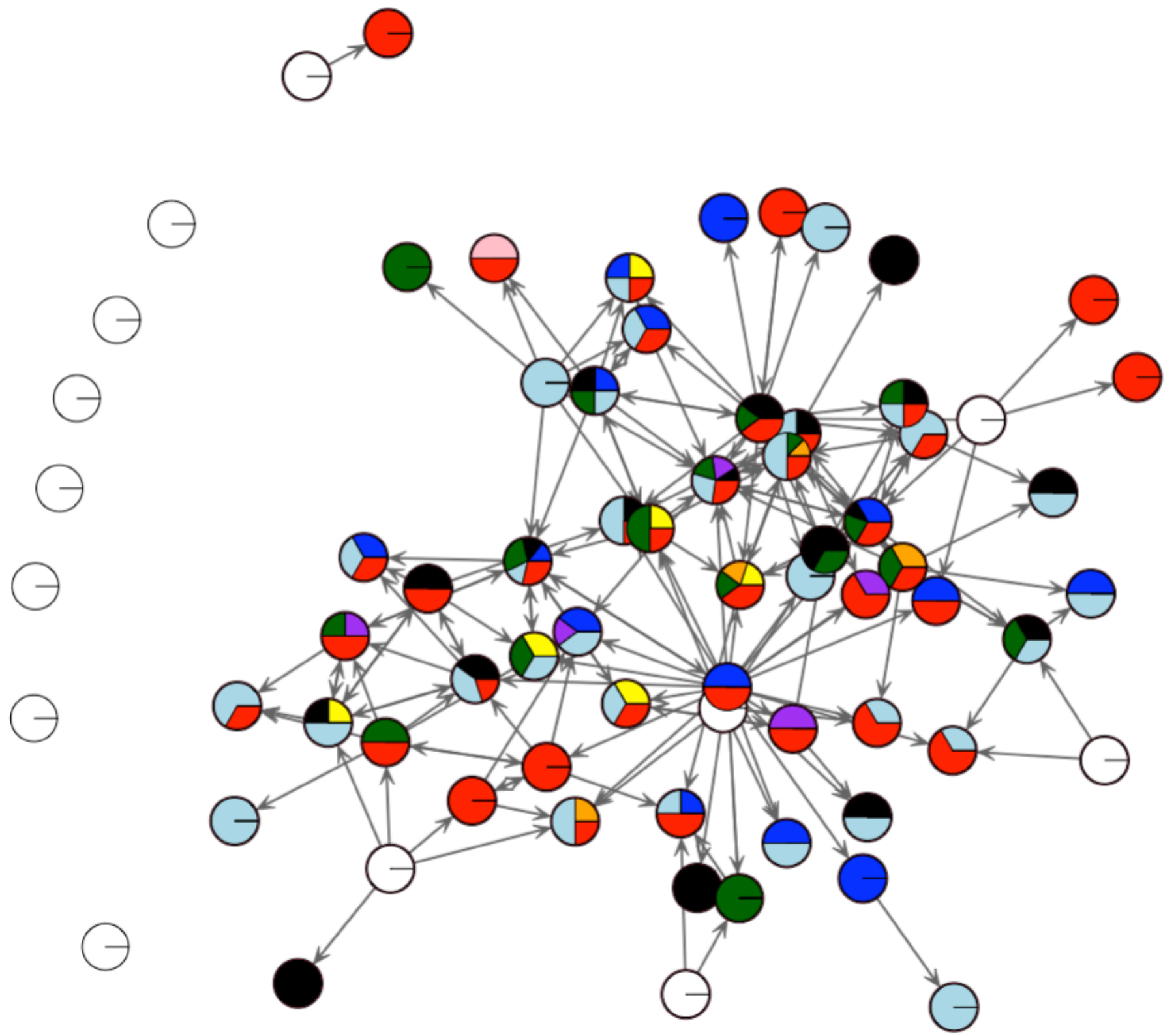
**FIGURE 1** Balance of positive, neutral, and negative ties in the status networks of (a) Unit 1, (b) Unit 2, (c) Unit 3, and (d) Men's Unit [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

*Notes:* Nodes sized by indegree, isolate frequencies listed in top-right of each graph, and pie charts reflect proportion of incoming nominations that are positive [black], neutral [white], and negative [red].



Family Role  
Nominations  
among  
Incarcerated  
Women

Good Behavior Family Network  
(w/ Non-Roles Noms)





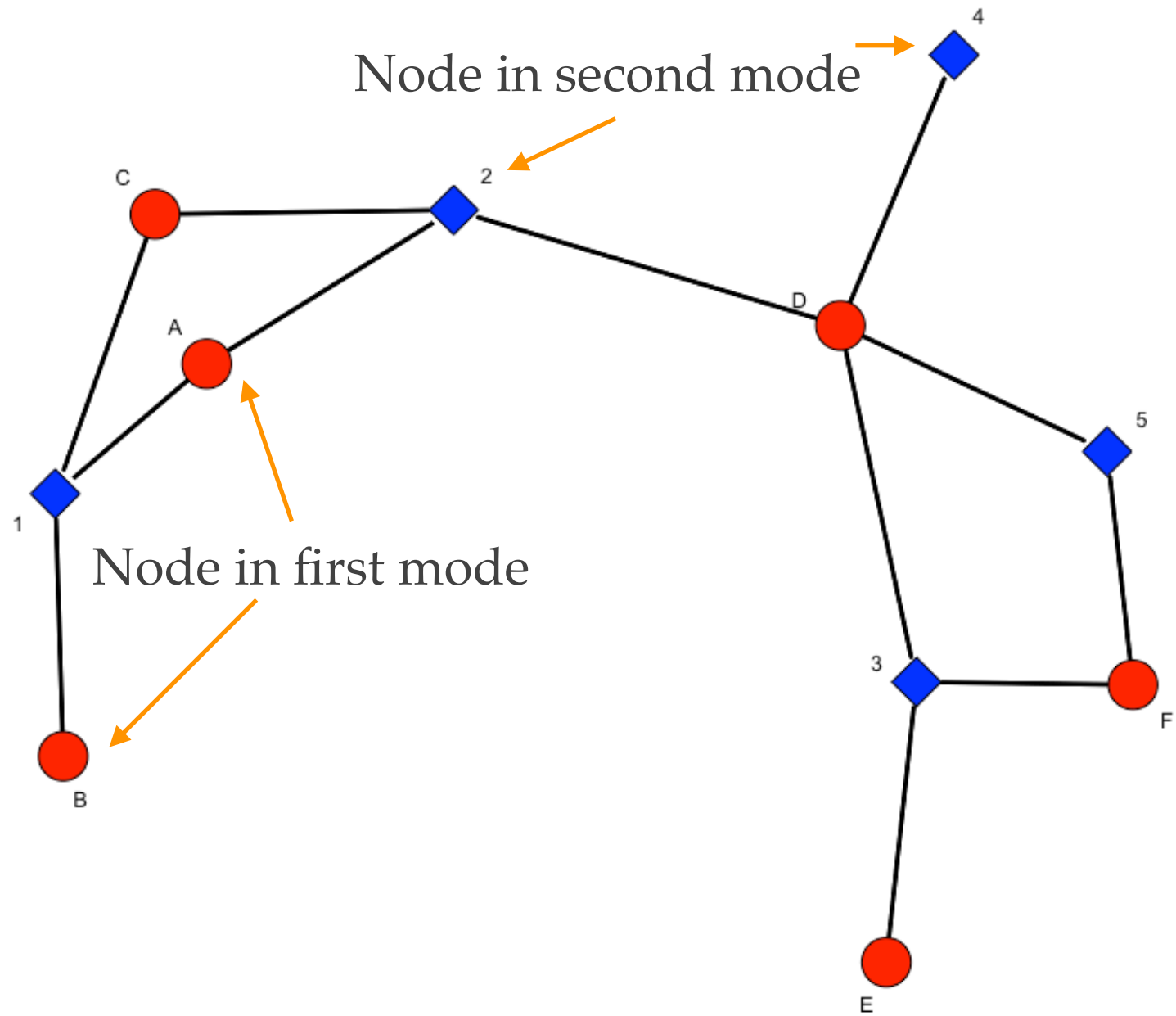
---

# Basic Data Elements

---

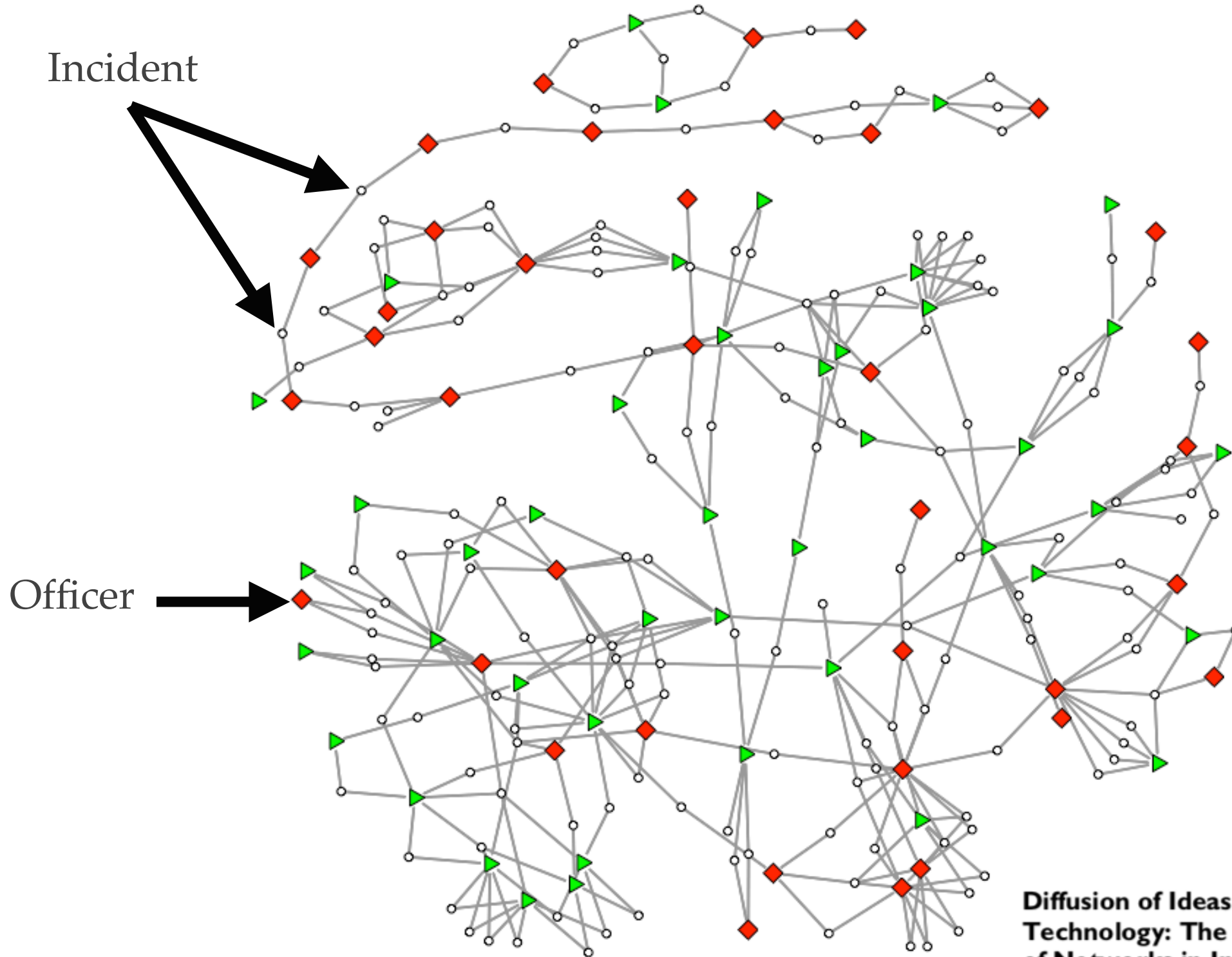
- ❖ *Networks* can differ with respect to their **nodes**:
  - ❖ One-mode / uni-partite (connections among one type of node).
  - ❖ Multi-mode / multi-partite (connections among two or more types of nodes).







# Bipartite Graph of Incidents and Officers by Treatment or Control Condition



Red/Square=Treatment Condition  
Green/Triangle=Control Condition  
White/Circle=Incidents

**Diffusion of Ideas and  
Technology: The Role  
of Networks in Influencing  
the Endorsement and Use  
of On-Officer Video Cameras**

Jacob T. N. Young<sup>1</sup> and Justin T. Ready<sup>1</sup>



**B: Plot of 291 Signers of 101 Bills in the Senate**

Am J Crim Just (2018) 43:197–221  
DOI 10.1007/s12103-017-9395-5



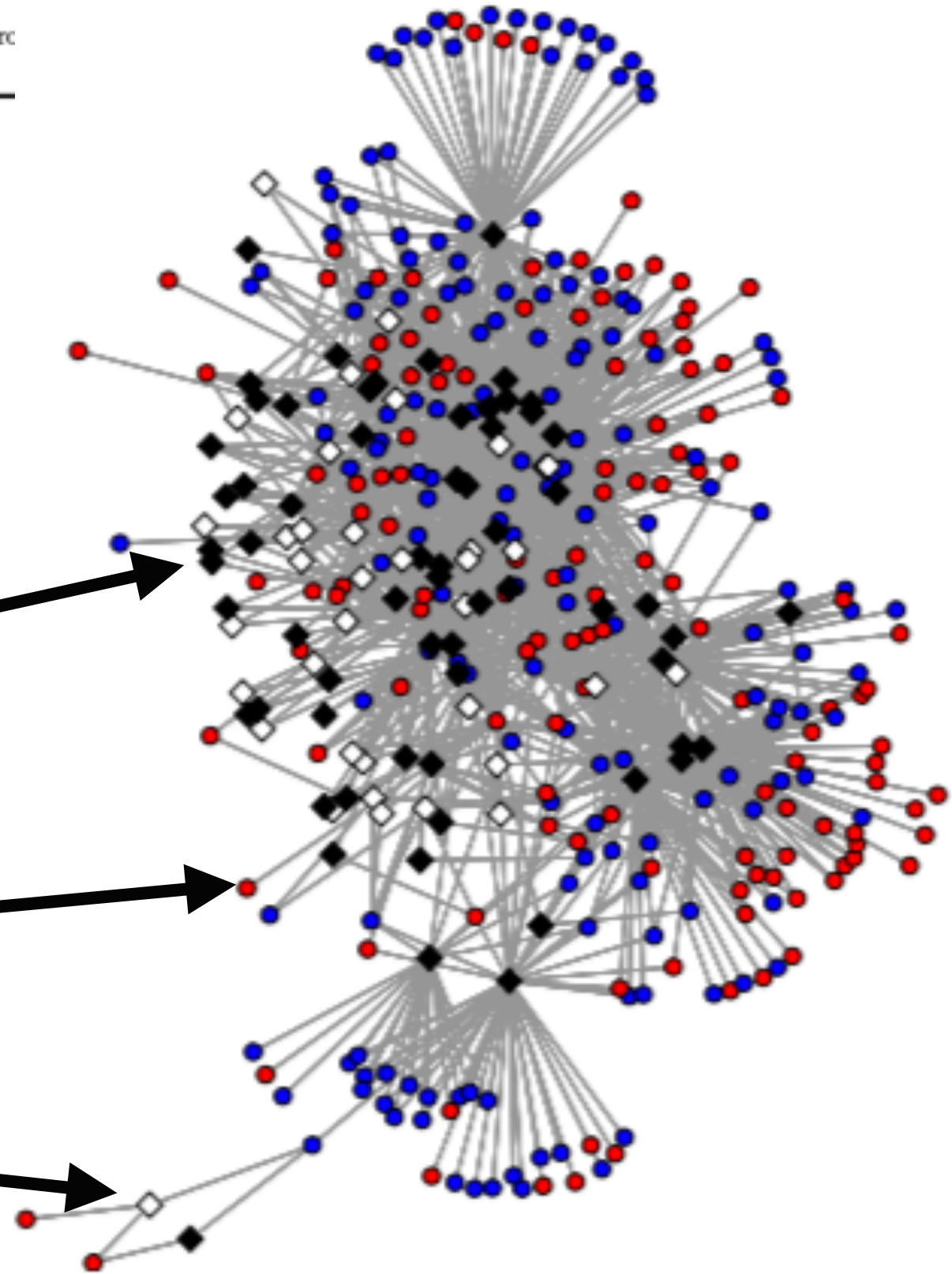
**The “Tough on Crime” Competition: a Network Approach to Understanding the Social Mechanisms Leading to Federal Crime Control Legislation in the United States from 1973–2014**

John A. Shjarback<sup>1</sup> · Jacob T. N. Young<sup>2</sup>

Crime Control Bill

Senator

Non-Crime Control Bill



Democrats = Blue; Republicans = Red; CC Bills = Black; NCC Bills = White



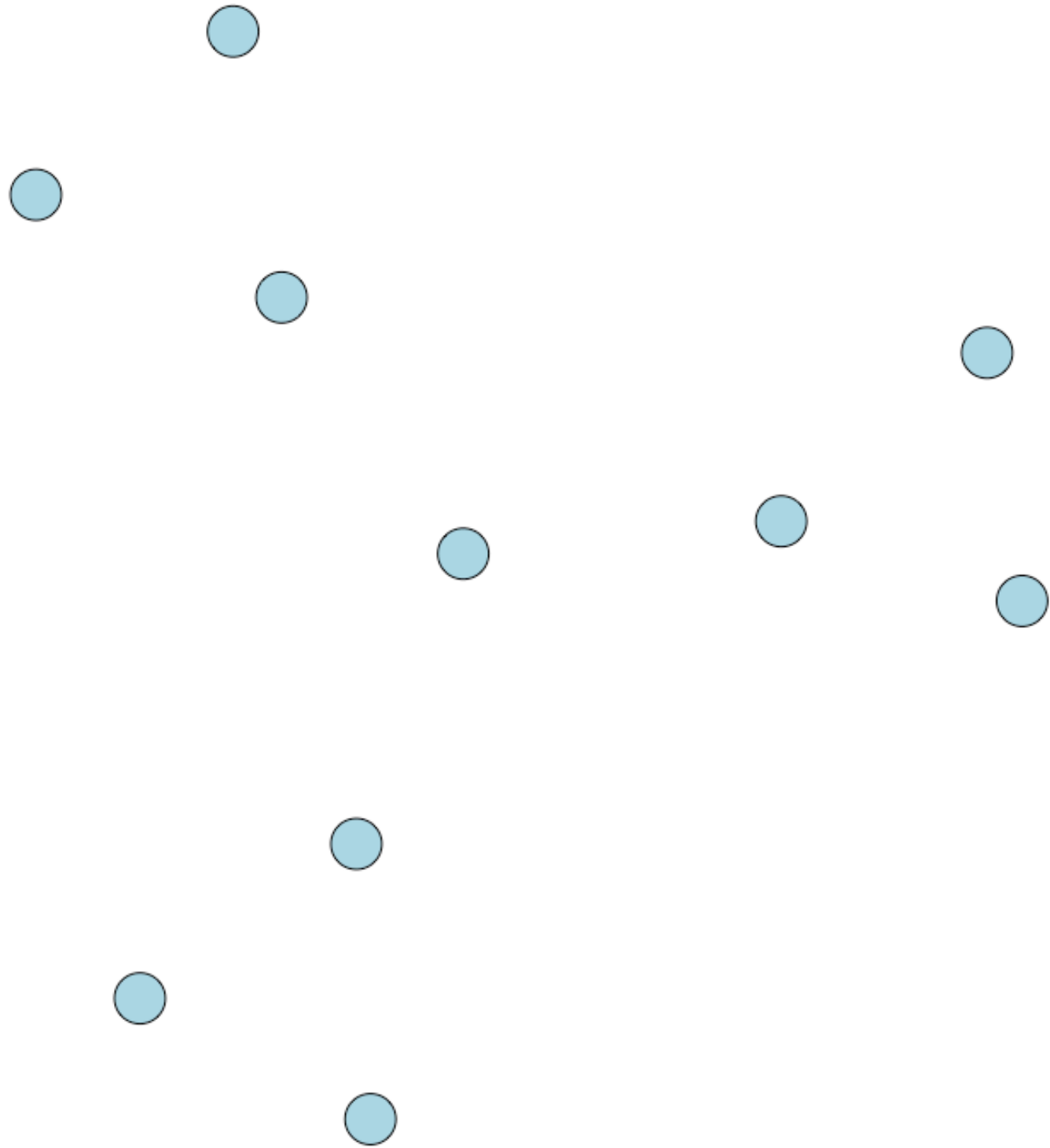
---

# Basic Data Elements

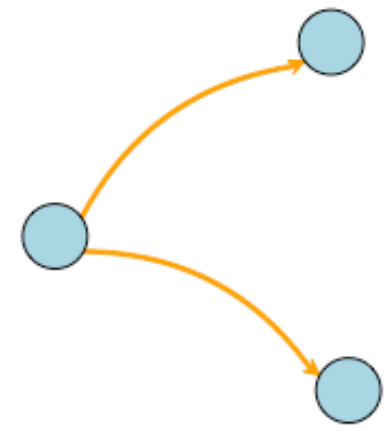
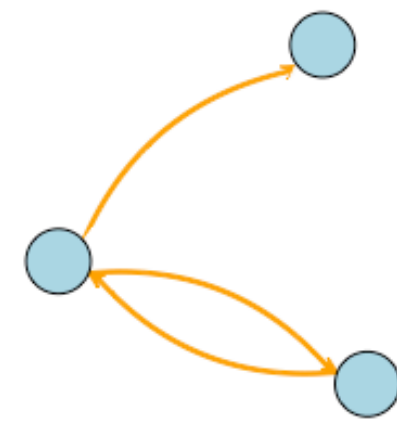
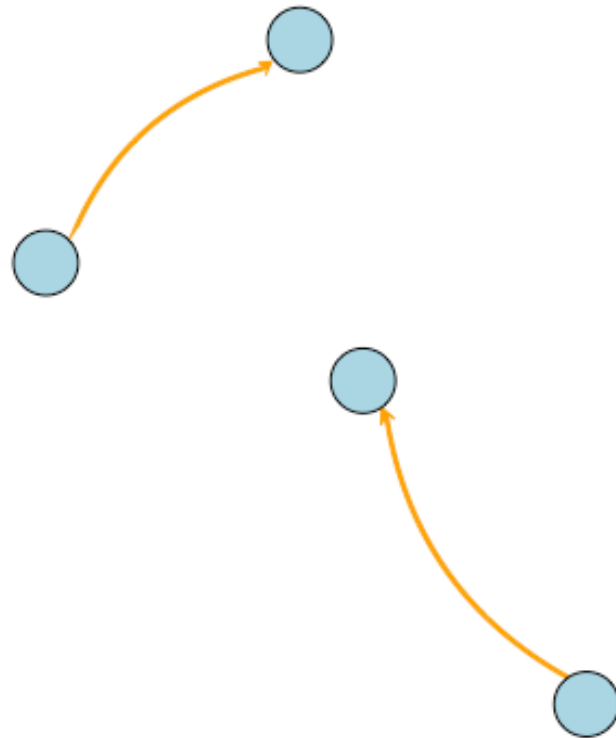
---

- ❖ *Networks* can differ with respect to their **edges**:
  - ❖ Simplex (connections among nodes are of one type).
  - ❖ Multiplex (connections among nodes are of multiple types).



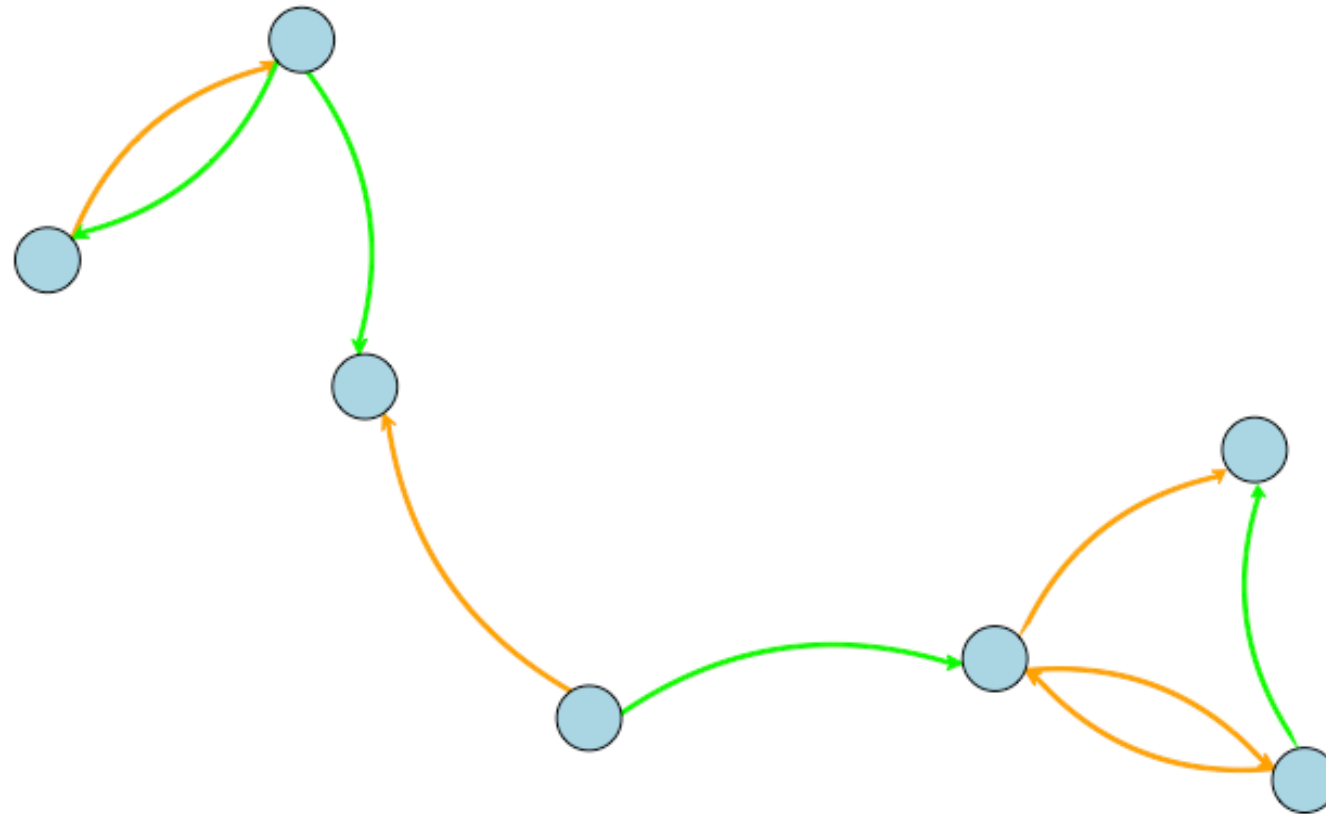






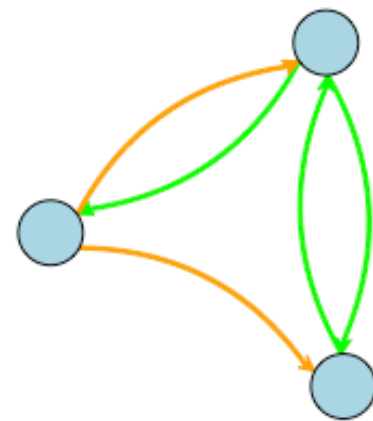
“Do you get along with this person?”



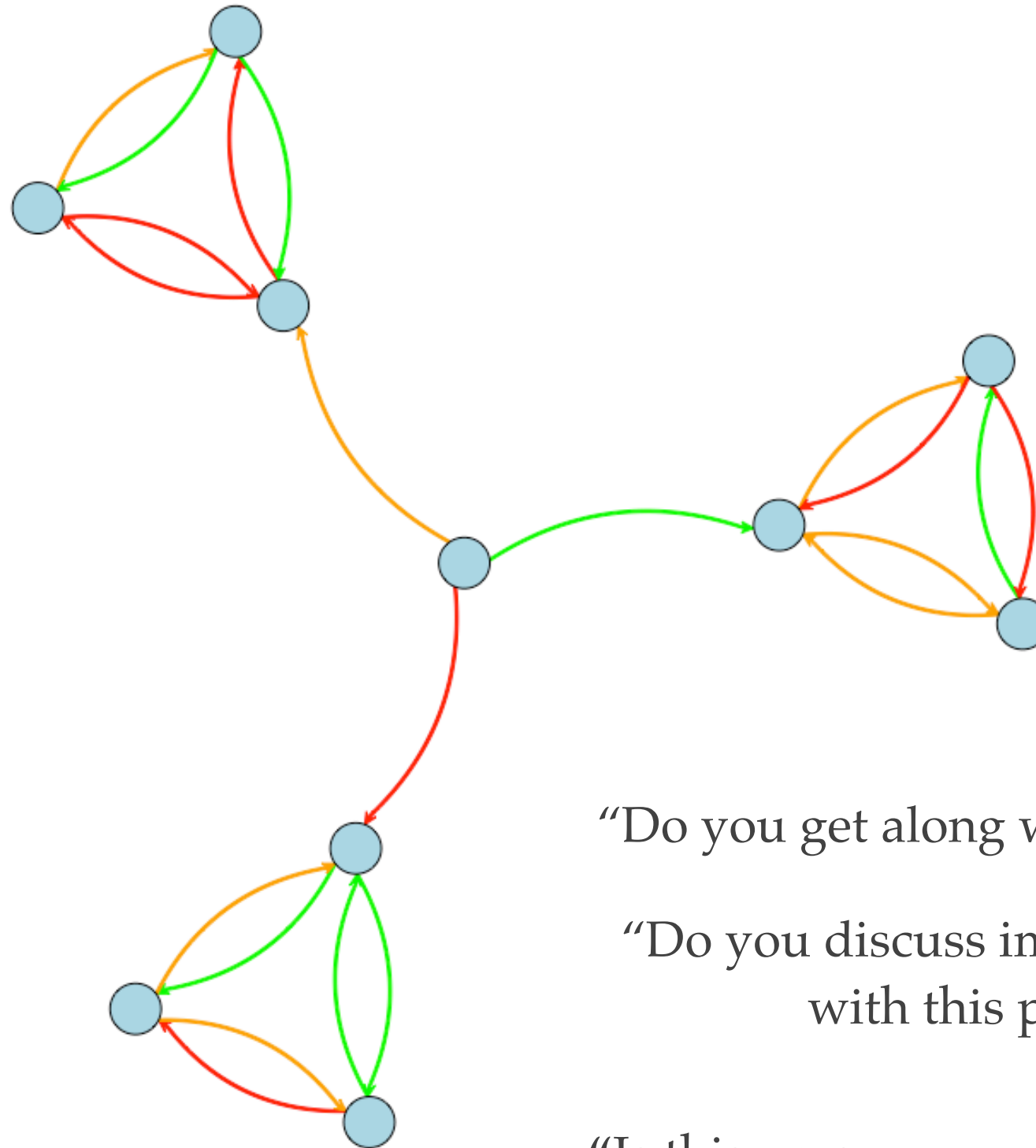


“Do you get along with this person?”

“Do you discuss important matters  
with this person?”







“Do you get along with this person?”

“Do you discuss important matters  
with this person?”

“Is this person powerful / influential?”



---

# Summary

---

- ❖ *Relational* data can take many different forms and be represented many different ways.



Questions?



# Network Data Collection

*How do we collect network data?*



---

# Network Data Collection

---

- ❖ Where do network data come from?
  - ❖ **Everywhere!**
    - ❖ Types of data collection:
      - ❖ Observational (e.g. Miller project)
      - ❖ Archival (e.g. Capone project)
      - ❖ Questionnaires (e.g. Add Health, PINS, GSS)



# ARE THESE NETWORK DATA?

O	NHD:StK	A	MM2	I	
FRI				Lo	Moe's
AG				Ac	hanging
CP	GAM <sub>g</sub>				
Q	<del>hon</del>				

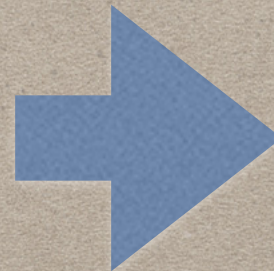
Pat Lynch and Billy Fuccillo were very incensed that Moe would not pay off \$4.00 in winning games on a pinball machine, especially since Moe sometimes paid off to certain individuals, probably boys he strongly liked, and boys who were older than Pat or Billy.

NO	act:	EV-	D	1/3/56	rt	BB
----	------	-----	---	--------	----	----



# THESE ARE NETWORK DATA

O	NHD:StK	A	MM2	I	
FRI			C3	Lo	Moe's
AG				Ac	hanging
CP	GAMg				
Q	hon				
<p>Pat Lynch and Billy Puccillo were very incensed that Moe would not pay off \$4.00 in winning games on a pinball machine, especially since Moe sometimes paid off to certain individuals, probably boys he strongly liked, and boys who were older than Pat or Billy.</p>					
NO	act: EV-	D	1/3/56	rt	BB



Event

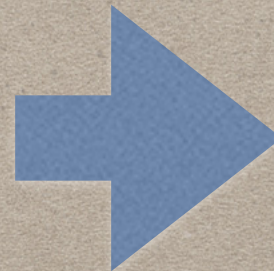


# THESE ARE NETWORK DATA

O	NHD:StK	A	MM2	I
FRI			C3	Lo Moe's
AG				Ac hanging
CP	GAMg			
Q	hon			

Pat Lynch and Billy Puccillo were very incensed that Moe would not pay off \$4.00 in winning games on a pinball machine, especially since Moe sometimes paid off to certain individuals, probably boys he strongly liked, and boys who were older than Pat or Billy.

act:	EV-	D	1/3/56	rt	BB
------	-----	---	--------	----	----



Event

Pat

Billy

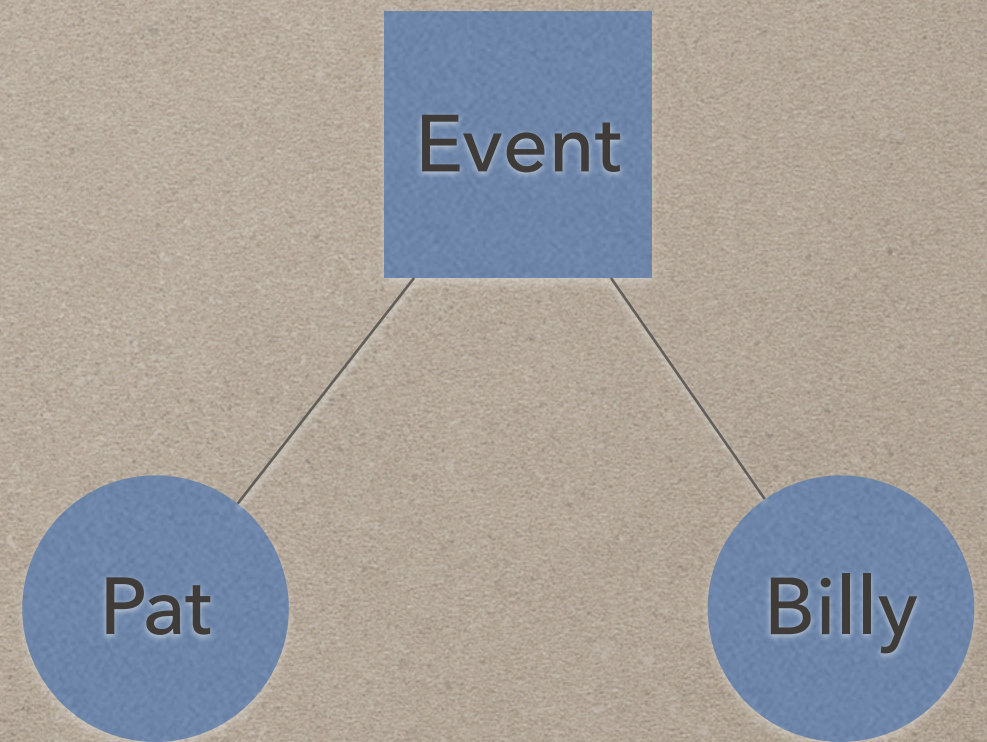
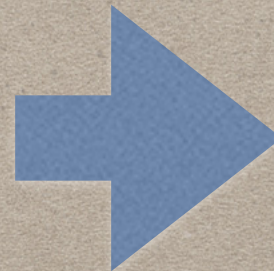


# THESE ARE NETWORK DATA

O	NHD:StK	A	MM2	I
FRI			C3	Lo Moe's
AG				Ac hanging
CP	GAMg			
Q	hon			

Pat Lynch and Billy Puccillo were very incensed that Moe would not pay off \$4.00 in winning games on a pinball machine, especially since Moe sometimes paid off to certain individuals, probably boys he strongly liked, and boys who were older than Pat or Billy.

act:	EV-	D	1/3/56	it	BB
------	-----	---	--------	----	----



A bipartite or two-mode network



---

# Boston Special Youth Project (SYP) Affiliation Data

---

- ❖ **Young, Jacob T.N.,** Scott H. Decker, and Gary Sweeten. 2018. “The Boston Special Youth Project Affiliation Dataset”. *Connections*, 37(1&2): 85-88. <https://www.exeley.com/journal/connections>.
- ❖ Data available here: <https://jacobtnyoung.github.io/SYP-Networks/>







---

# Prison Inmate Network Study (PINS)

---

- ❖ The PINS is a two-wave study of relationships among incarcerated men in a medium security prison.
- ❖ See: <https://pure.psu.edu/en/projects/the-prison-inmate-networks-study-pins>
- ❖ Each wave captures survey and network data for individuals in the same unit with a capacity of 205 inmates.



---

# PINS Instrument

---

Q8b Here is a list of residents currently in the unit. Please refer to this list as you answer the following questions. Who do you get along with most?

Other (269) \_\_\_\_\_

Other (270) \_\_\_\_\_

Other (271) \_\_\_\_\_

Do not get along with anyone (272)

.. (273)



---

# Boundary Specification

---

- ❖ The theoretical *and* methodological challenge of determining the appropriate set of actors and connections to analyze in order to identify the relevant social network within a given context.
- ❖ Is there some boundary that really exists?
- ❖ Or, is a boundary necessarily imposed to conduct the research?



---

# Instruments and Design

---

- ❖ *Instruments* are the tools used to elicit information from respondents.
- ❖ *Design* corresponds to the protocol for determining how information should be elicited, who should be sampled, etc.
  - ❖ Examples:
    - ❖ Ego-centric networks
    - ❖ Partial networks
    - ❖ Complete (global) networks



---

# Ego-Centric Networks

---

- ❖ Data on a focal actor (ego) and ties to neighbors (alters) and the ties among the alters.
- ❖ *Instrument*: name generator
  - ❖ “who are the people with whom you discuss important matters?”
  - ❖ For each person named, “which of these individuals discuss important matters”?
  - ❖ Why?-costs, generalizability, interest in local structure.



---

# Partial Network

---

- ❖ Ego networks, plus some amount of tracing to reach contacts of contacts.
- ❖ *Instrument*: tracing mechanism
  - ❖ Using tickets to trace across a network
  - ❖ Why?-difficult to reach population, hard to specify sampling frame.
  - ❖ **Does this instrument seem familiar?**



---

# Complete (Global) Network

---

- ❖ Data on all actors within a particular (defined) boundary, sampling frame is known.
  - ❖ *Instruments:*
  - ❖ roster
    - ❖ “For each of the following persons, please indicate whom you trust to share personal information with?”
  - ❖ Free response
    - ❖ “Who are the people in this prison that you trust with person information about you?”



---

# Things to consider...

---

- ❖ Domain:
  - ❖ “What type of network is this?”
- ❖ Sample:
  - ❖ “What is the population of interest and how was it sampled?”
- ❖ Temporality:
  - ❖ “Are the data cross-sectional or longitudinal?”
  - ❖ “Is it a panel or continuous measurement?”
- ❖ Tie Meaning:
  - ❖ “Are ties discrete events or enduring states?”
- ❖ Instrument:
  - ❖ “How was the information collected?”



---

# Things to keep in mind...

---

- ❖ Butts (2009: 24)

- ❖ Representational formalism:

- ❖ “Researchers begin with a finite set of identifiable entities...”

- ❖ “This representational framework is quite restrictive. To represent a system in this way, we must be able to reduce it to a well-defined set of discrete components whose interactions are strictly dyadic in nature...; although such a framework may seem so restrictive as to be useless, its typical purpose is to serve as an approximation...”

- ❖ *What does this mean?*



---

# More things to keep in mind...

---

- ❖ *What is the level of analysis?*
  - ❖ Dyad Level
    - ❖ “are individuals whose offices are close to each other more likely to be friends?”
  - ❖ Node Level
    - ❖ “are more popular youth more likely to engage in unprotected sex?”
  - ❖ Network Level
    - ❖ “do viruses spread faster in particular network structures?”



# Basic Network Analysis



---

# Description vs. Inference

---

- ❖ Description

- ❖ What proportion of possible ties are observed? (density)
- ❖ Who has the most ties? (degree centrality)
- ❖ Are there clusters in the network? (graph modularity / subgroup analysis)

- ❖ Inference

- ❖ How did this graph form? (Exponential random graph models)
- ❖ Why do nodes change their edges? (Stochastic actor based models)
- ❖ Do edges / nodes influence nodes / edges? (Co-evolution models)



---

# Network Analysis

---

- ❖ What sort of analysis you do depends on the question:
  - ❖ Do attitudes spread through simple or complex processes? (diffusion)
  - ❖ Do adolescents sort into subgroups? (modularity)
  - ❖ Are particular networks more vulnerable than others? (degree assortativity)
  - ❖ And so on...



---

# Summary

---

- ❖ Network analysis can be simple or complex, descriptive or inferential.
- ❖ But, remember the crucial steps:
  - ❖ **Conceptualizing** theoretical concepts that are inherently relational.
  - ❖ **Operationalizing** theoretical constructs by drawing on the formal properties of graphs.



---

# Twelve sites with data

---

- ❖ <https://icon.colorado.edu/#!/>
- ❖ <https://github.com/awesomedata/awesome-public-datasets#socialnetworks>
- ❖ [www-personal.umich.edu/~mejn/netdata](http://www-personal.umich.edu/~mejn/netdata)
- ❖ <http://snap.stanford.edu/data/index.html>
- ❖ <https://lincs.soe.ucsc.edu/data>
- ❖ <https://networkdata.ics.uci.edu/>
- ❖ <http://jmcauley.ucsd.edu/data/amazon/>
- ❖ <http://konect.uni-koblenz.de/networks>
- ❖ <https://www.icpsr.umich.edu/icpsrweb/NAHDAP/studies/22140>
- ❖ <https://www.dimensions.ai/widgets/access/>
- ❖ <https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/36975>



---

# Learning Goals

---

- ❖ Be able to answer these questions:
  - ❖ What is “network science” and how is it different from “usual” research?
  - ❖ What do networks “look like”?
  - ❖ Where do network data “come from”?



Questions?