Social Networks “Crash Course”

Jacob Young
jacob.young.1@asu.edu
GITHUB PAGE

Goal: Answer some questions

- What is a “network perspective” and why think about the world this way?
- What do networks “look like”?
- How do we collect network data?
- What does network **data** “look like”?
- How do we analyze network data?

Schedule

- 9:00-9:30 Introductions and Overview
- 9:30-10:15 Basics of Network Science
- 10:15-10:30 Break
- 10:30-11:30 Network Data Structures and Analysis
- 11:30-12:00 Wrap-Up and Questions

Introductions

Who are you and why are you here?

Why do people get the flu?

What is the contrast?

- The old familiar…
  - Variable-based approaches to research focusing on **individuals**.
  - The *explanan* is variation between individual units.
    - For example: variation in susceptibility or resistance to the flu.
      - Premised on static “thing-concepts” as their primary unit of analysis.

Another view of research

- What is network *science*?
  - An approach to science that views the world as being composed of systems of actors connected through relational ties (i.e. a **network**).
  - The *explanan* and *explanadum* is network **structure**.
    - For example: the properties of flu virus transmission.
      - Is it dense? Sparse? Modular?

Now, *more* questions

- For example, consider two different questions:
  - Are kids who are risk-seeking more likely to drink alcohol?
  - Are kids who have friends that drink alcohol more likely to drink alcohol?
    - *Is the causal logic the same?*
    - *Are the policy implications the same?*

Network Science

- These are all questions that require a different way of thinking about the world.
- Network science takes network **structure** as the primary domain of interest.
    - **Structural variables** are quantities that measure structure.

Network Science

- As with research from an individualistic perspective, network research identifies concepts and relationships among concepts.
  - For example: Power, embeddedness, integration, …
- A key difference, though, is that network research **operationalizes** theoretical concepts by drawing on the formal properties of graphs.

Network Analysis

- Network **analysis** is the set of tools used to study *structural variables*.
  - A set of *relational* 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.

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

Can you think of a **structural variable** that could measure Laurie's importance?

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.

Questions?

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).
- I will use *node* to mean **entities** and *edge* to mean **connections.**

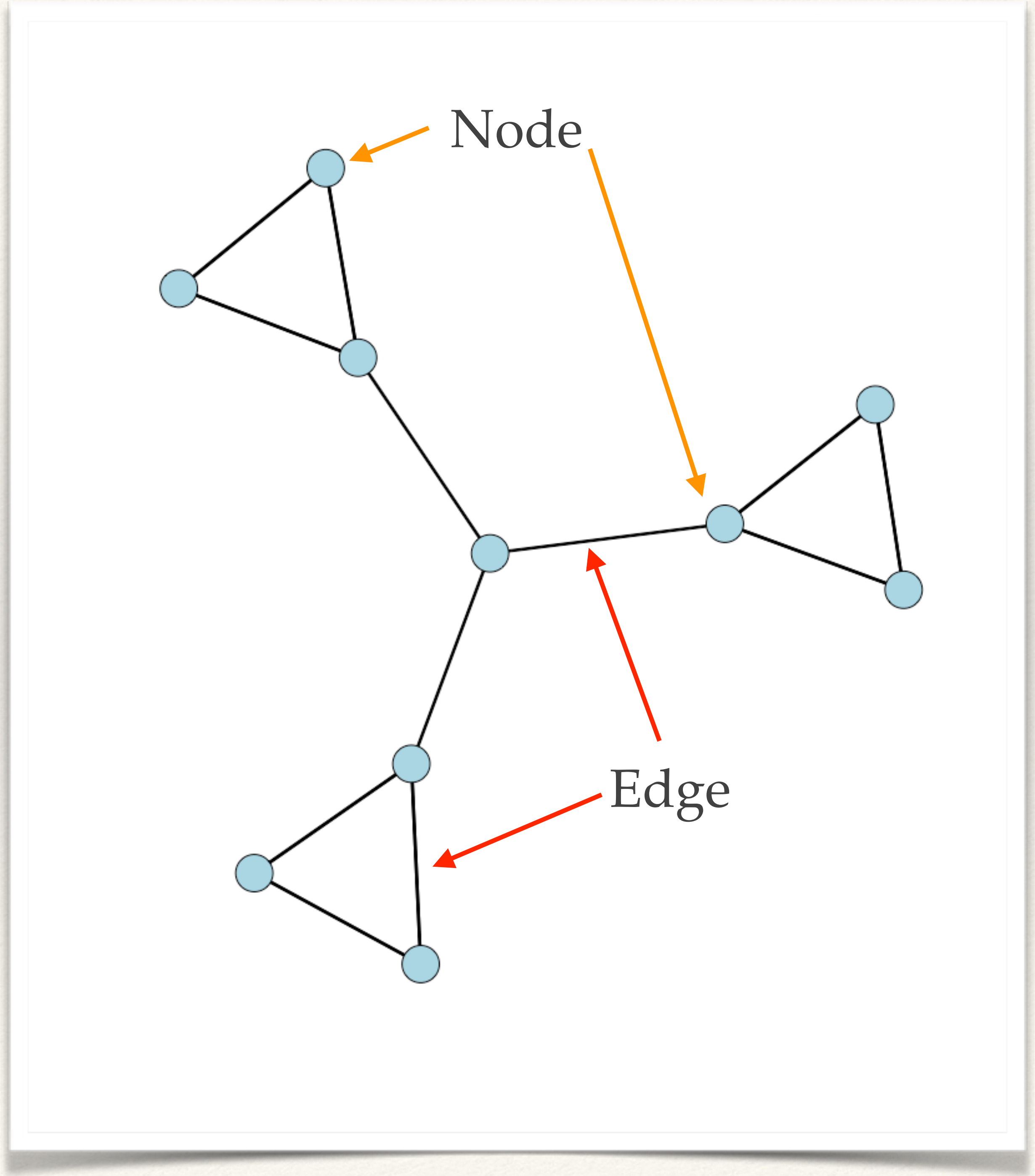
Basic Data Elements

- A *node* can be anything that can link to something else and an *edge* can be anything that can record a connection between nodes.

  - *What are some **nodes** and **edges** that come to mind?*

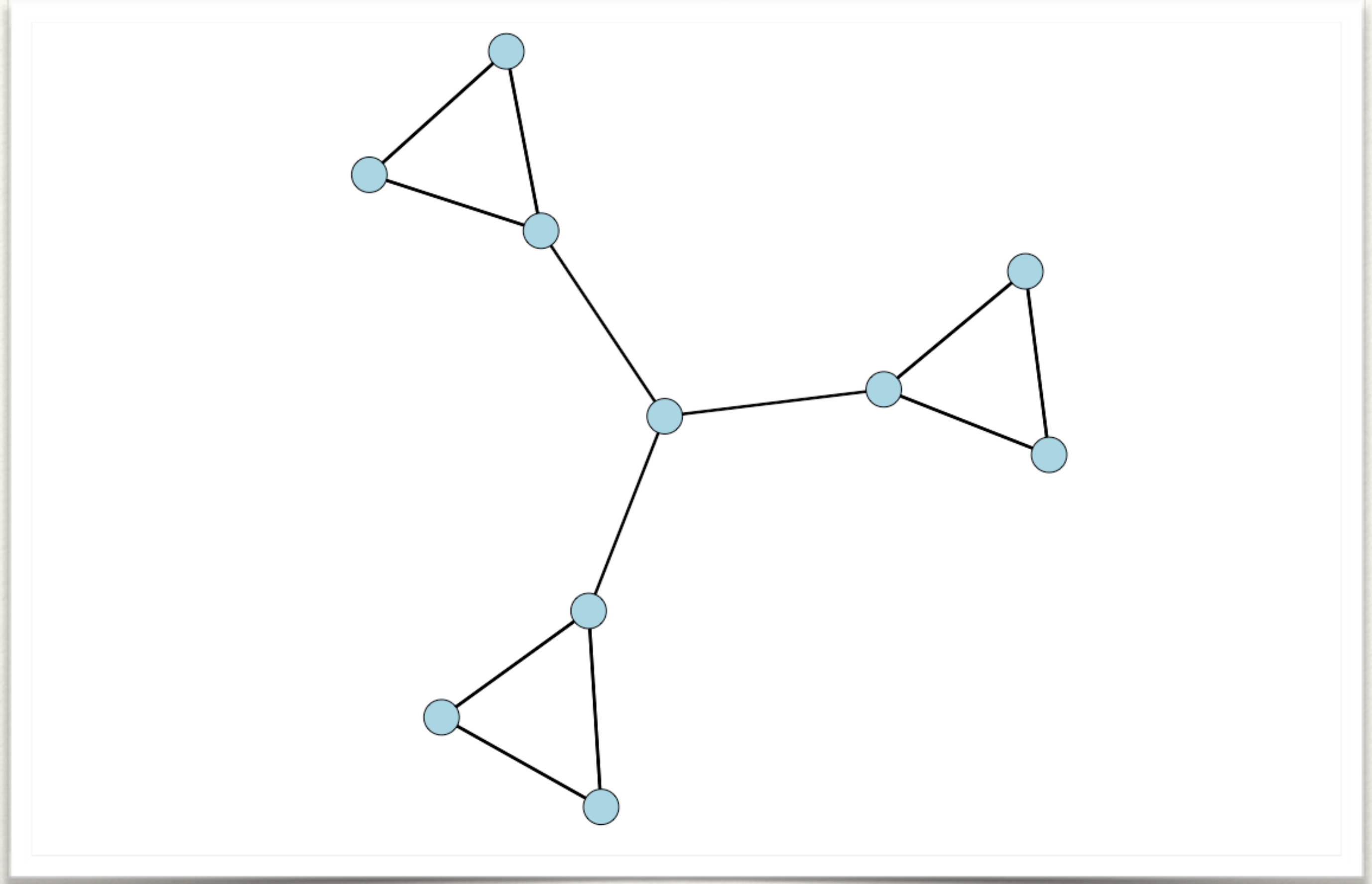
Basic Data Elements

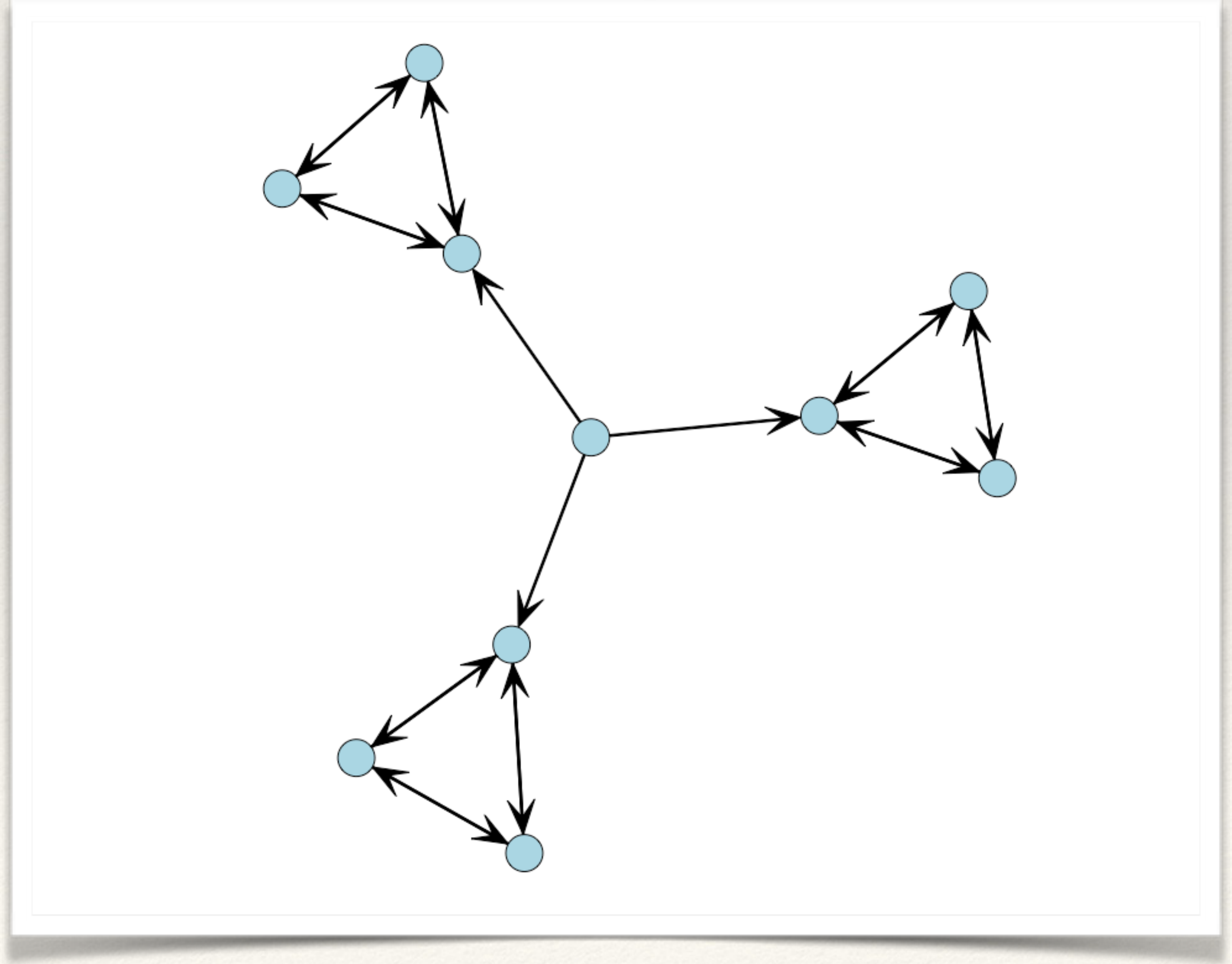
- On a graph, nodes are represented by *points* and edges are represented by *lines*.

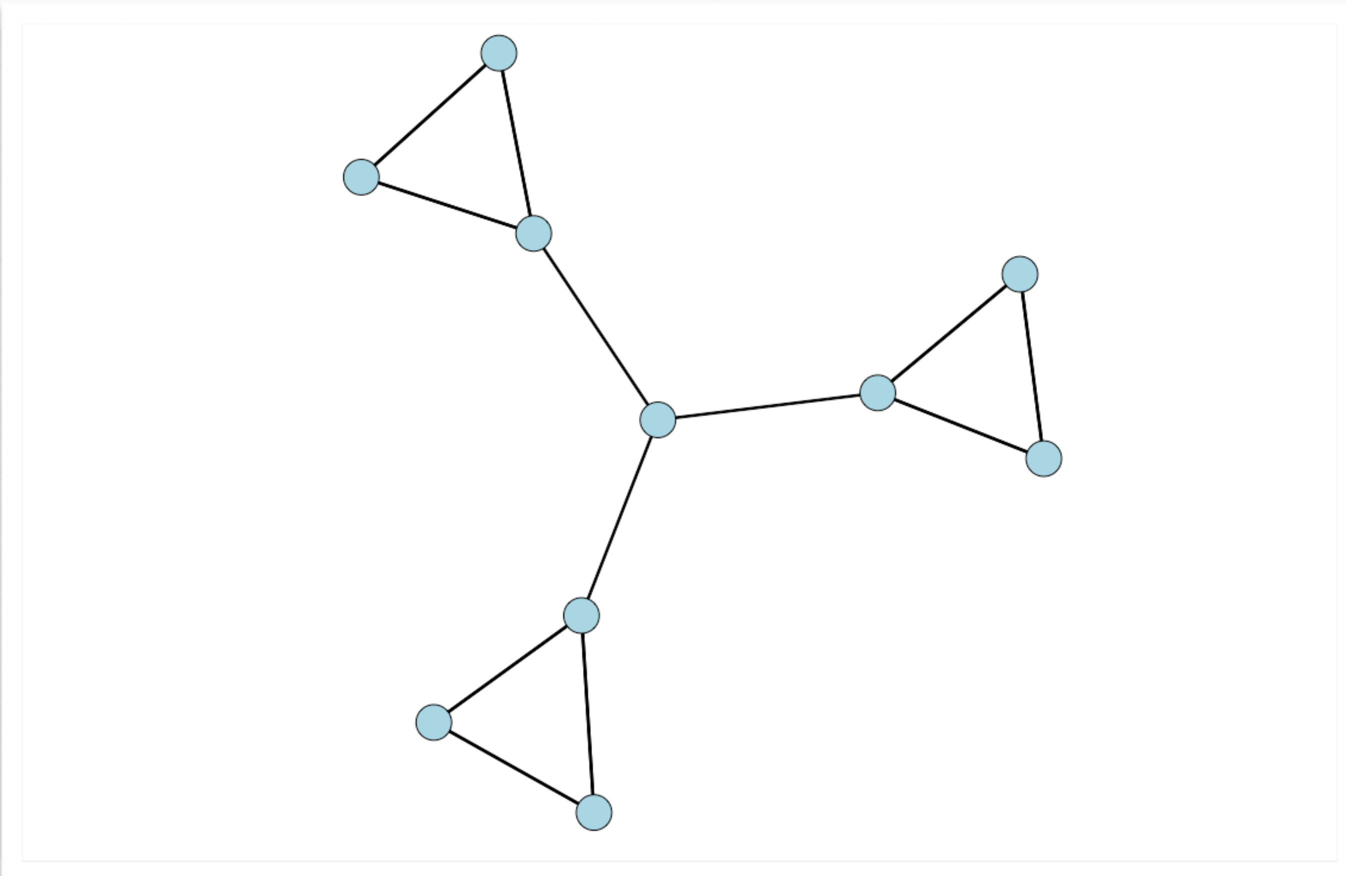


Basic Data Elements

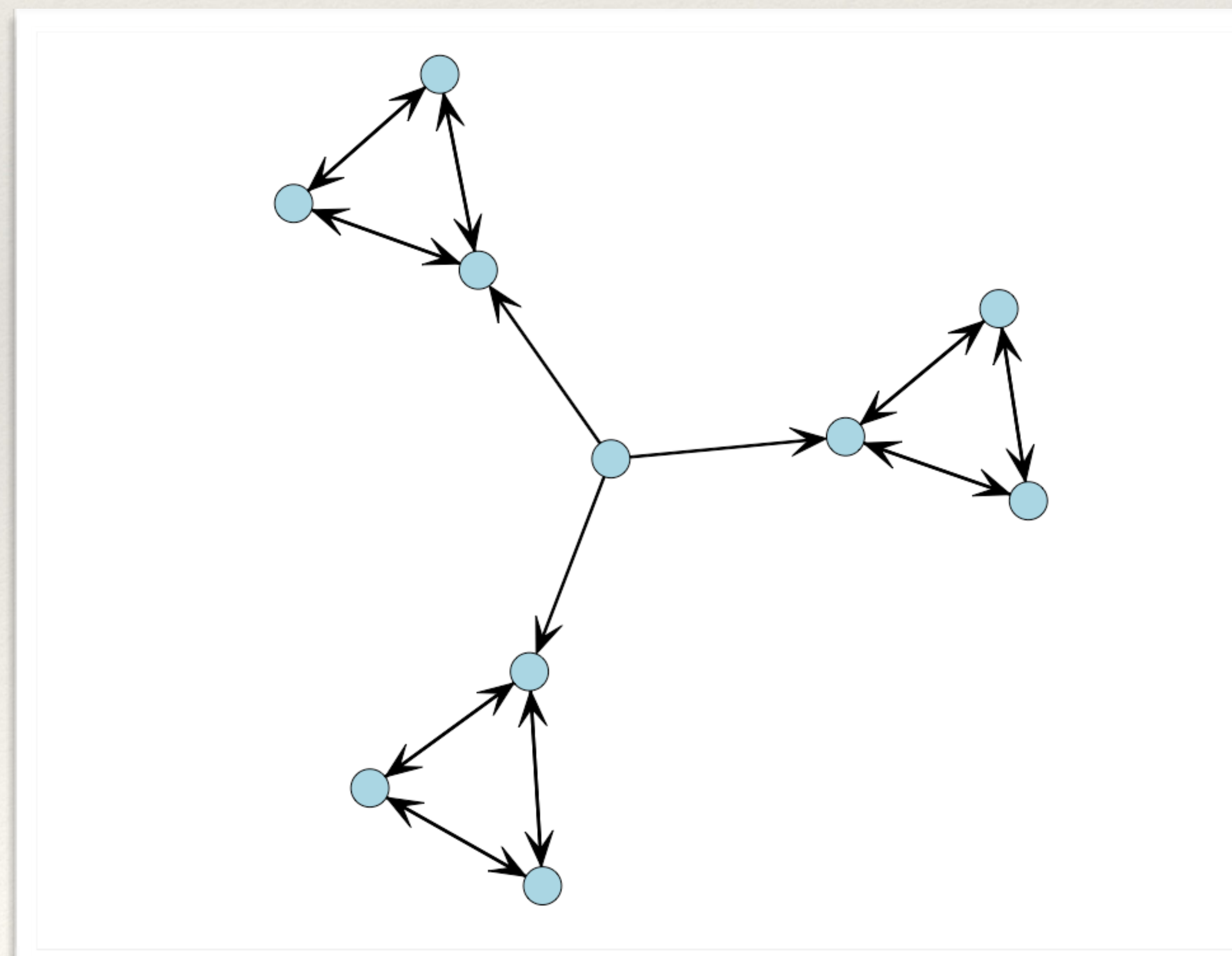
- Edges can be:
  - Directed or Undirected.





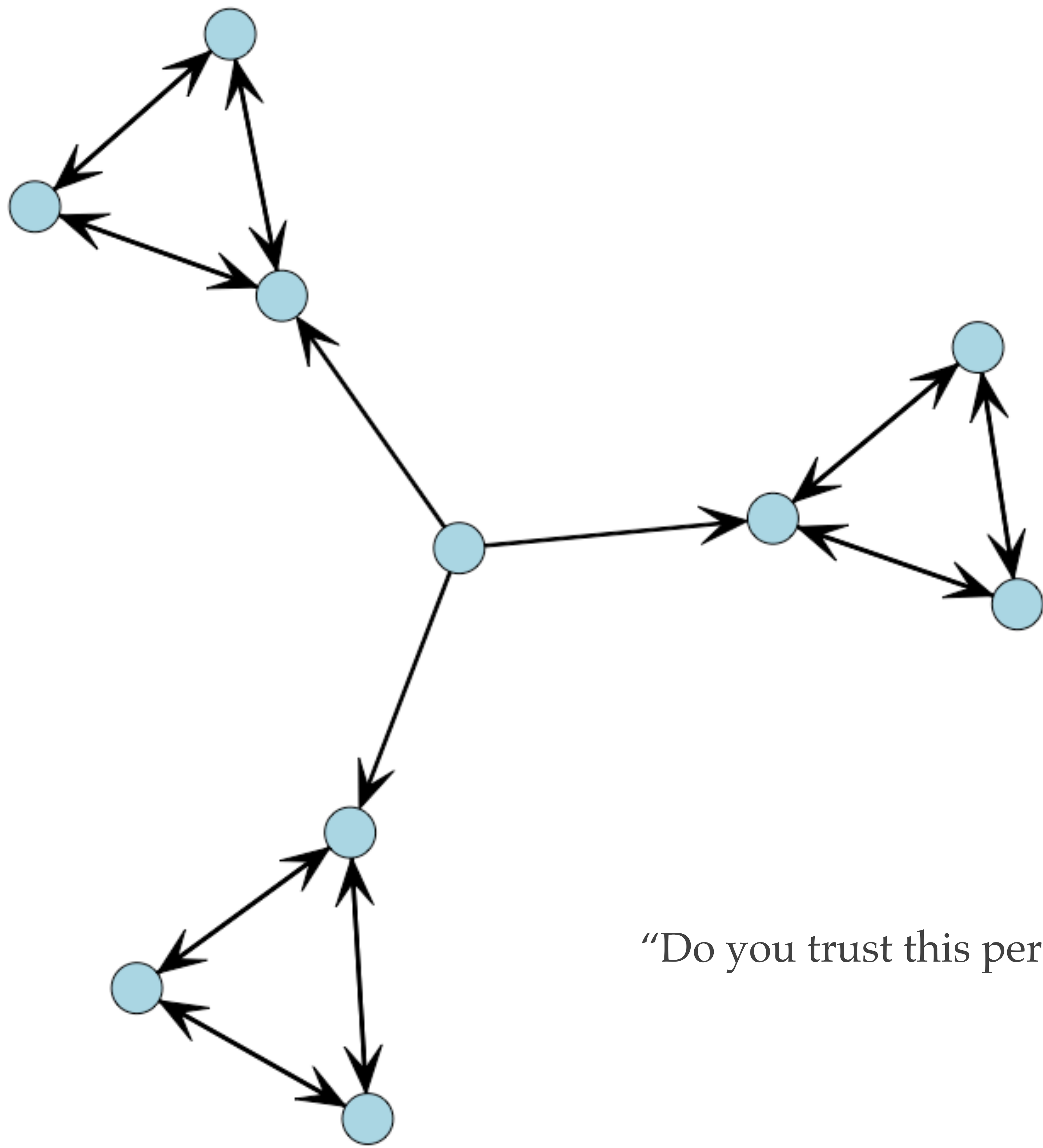


How are these structures different?

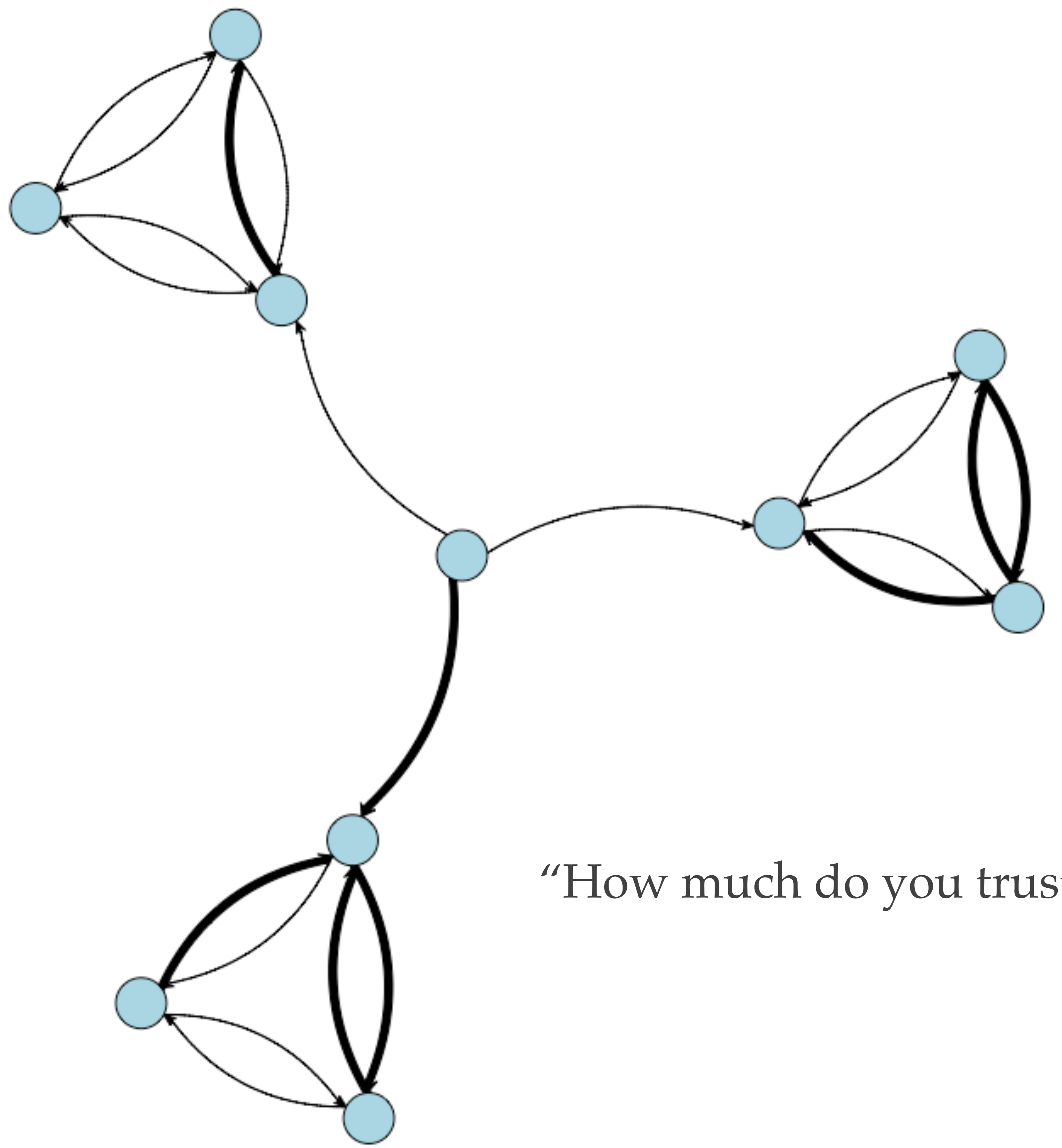


Basic Data Elements

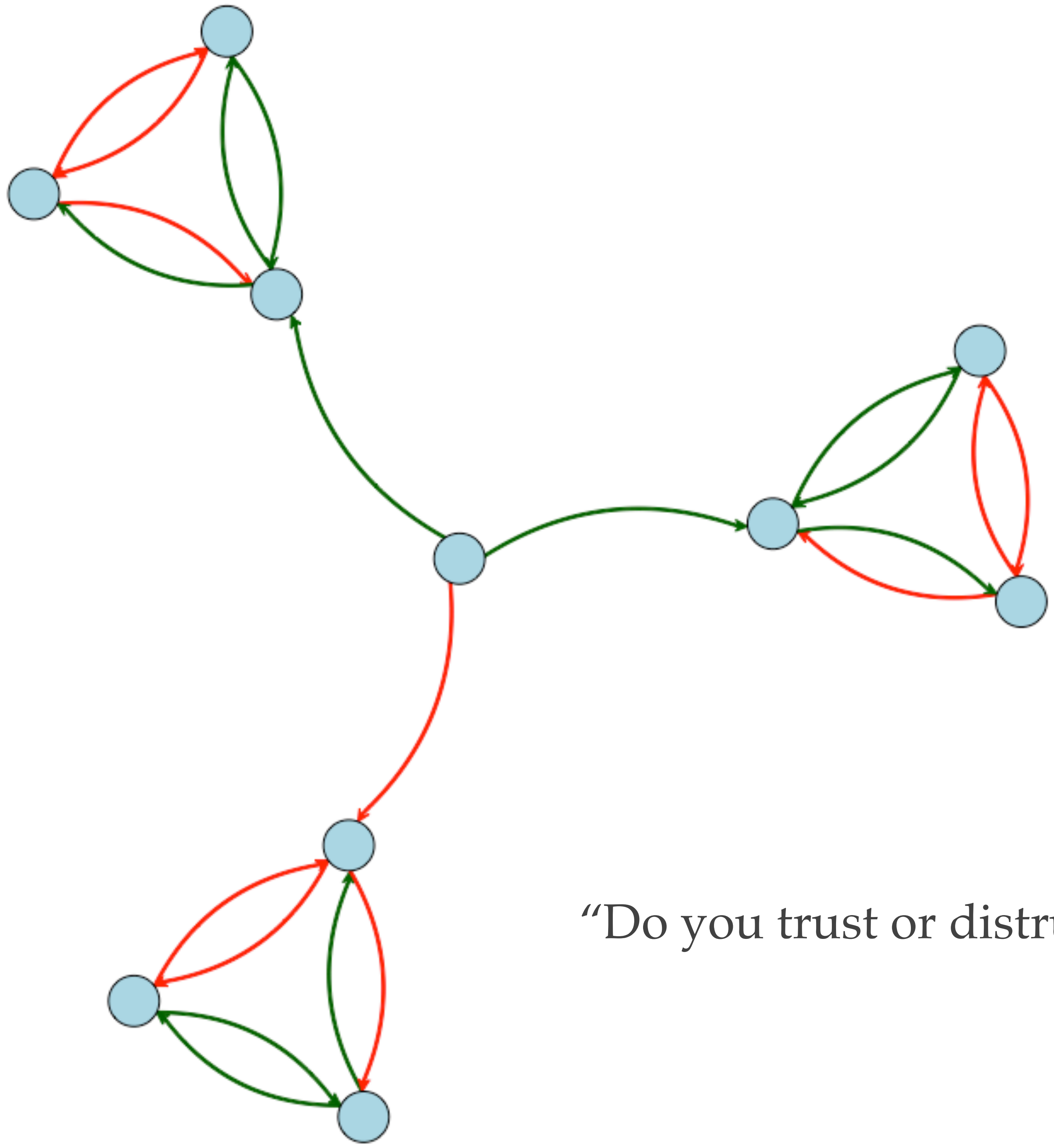
- **Edges** can be:
  - Binary (0/1; present/absent); Valued (0/1/2…); Signed (+/-).



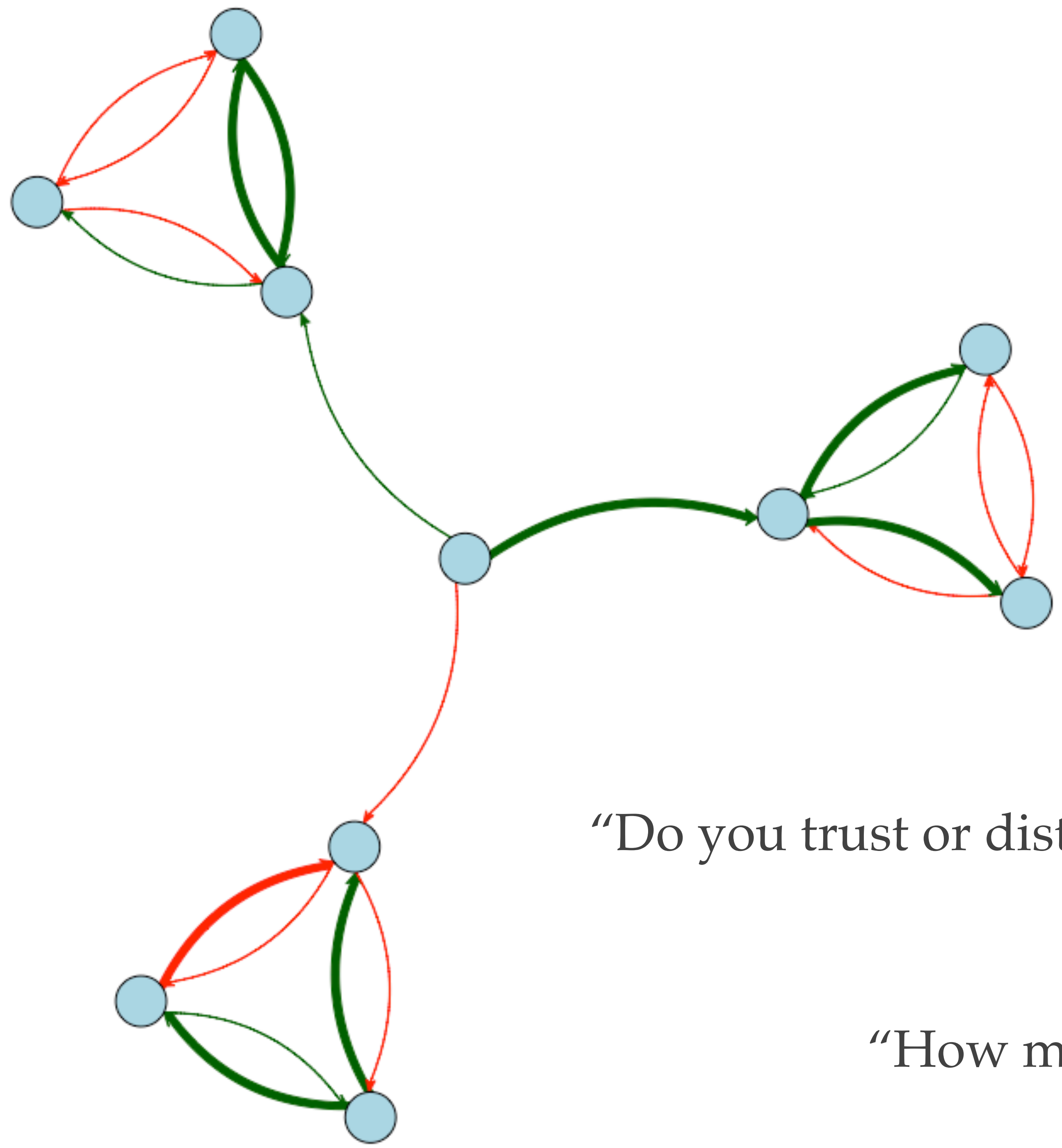
"Do you trust this person?"



“How much do you trust this person?”



“Do you trust or distrust this person?”

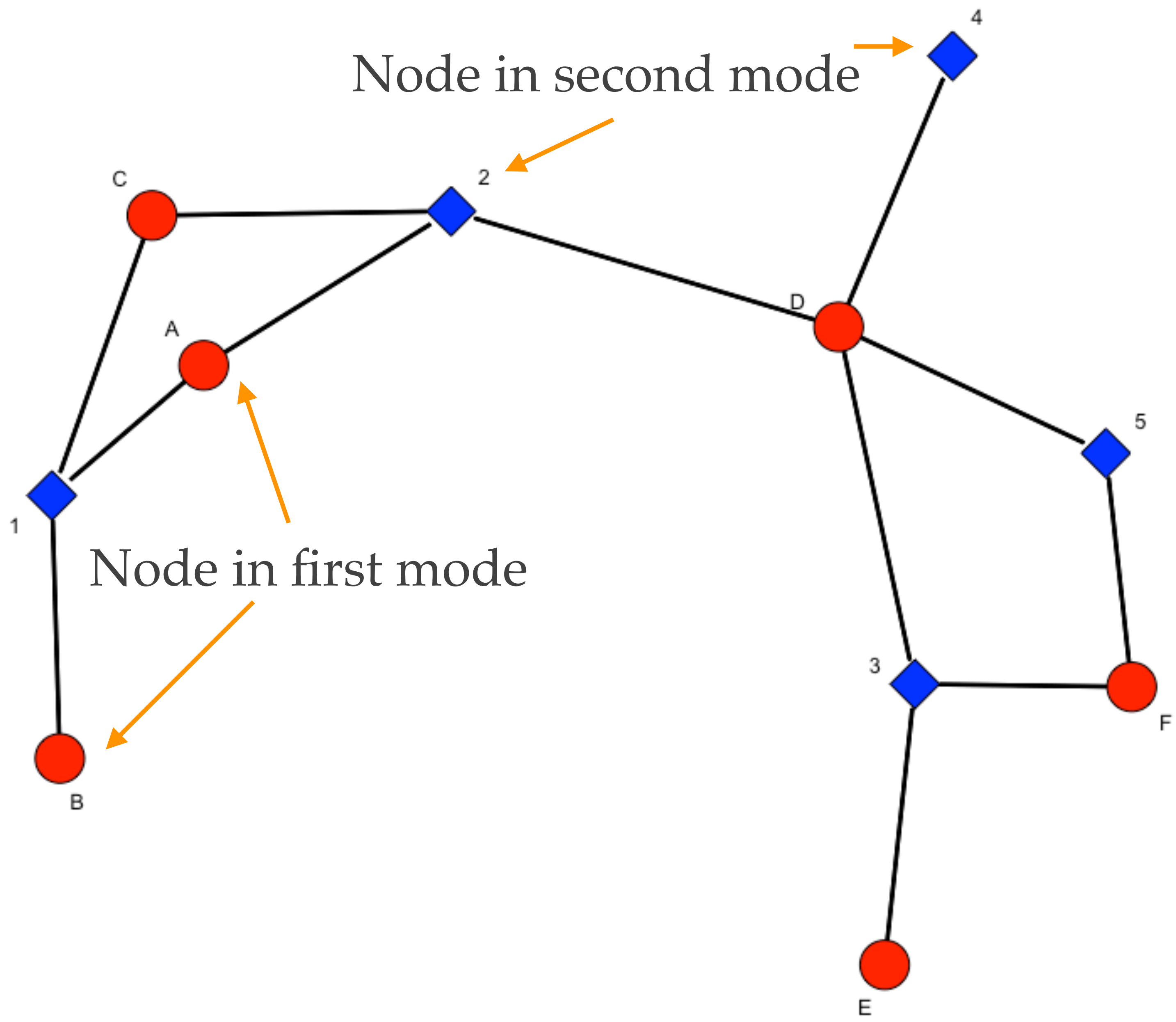


"Do you trust or distrust this person?"

"How much?"

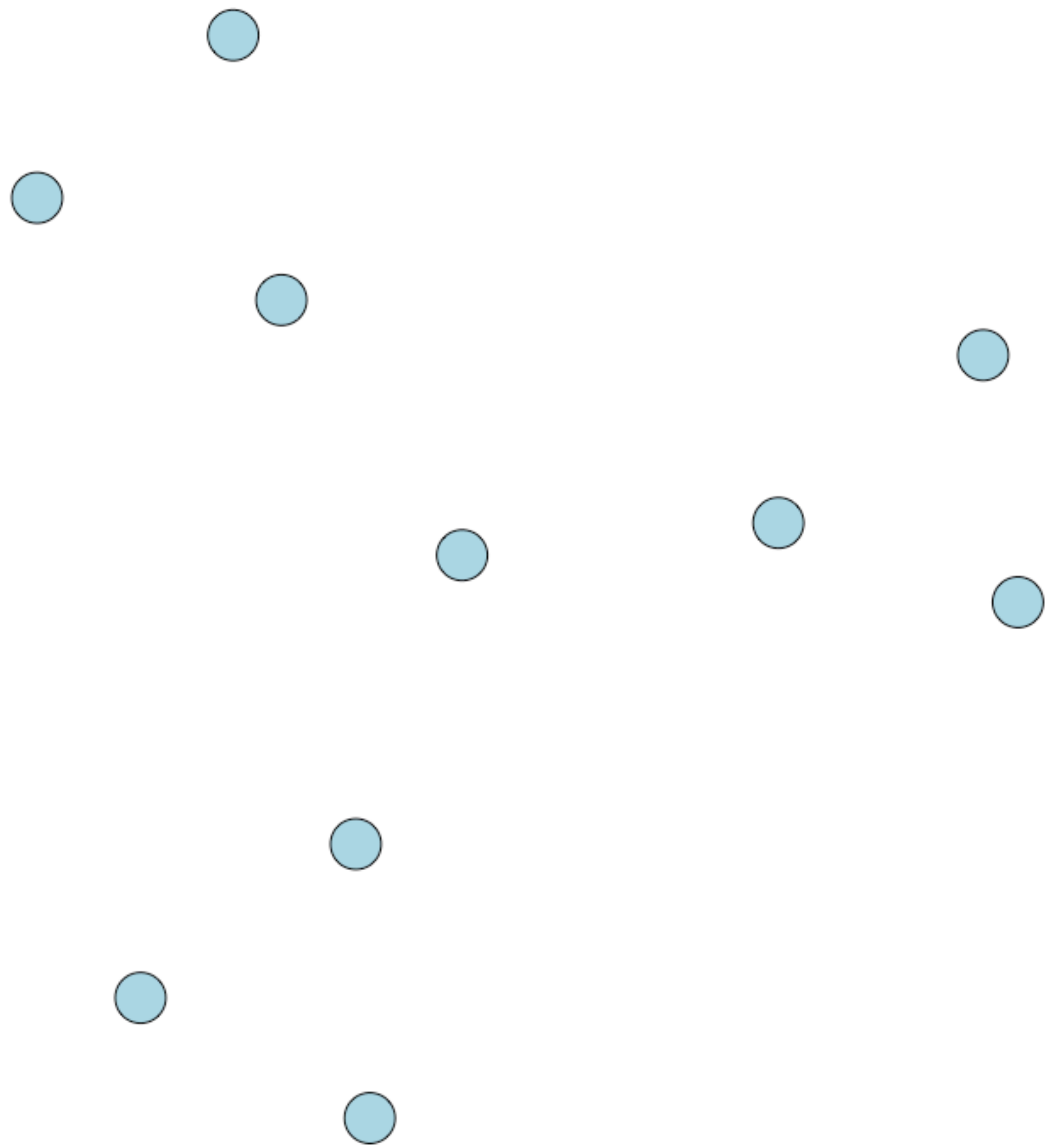
Basic Data Elements

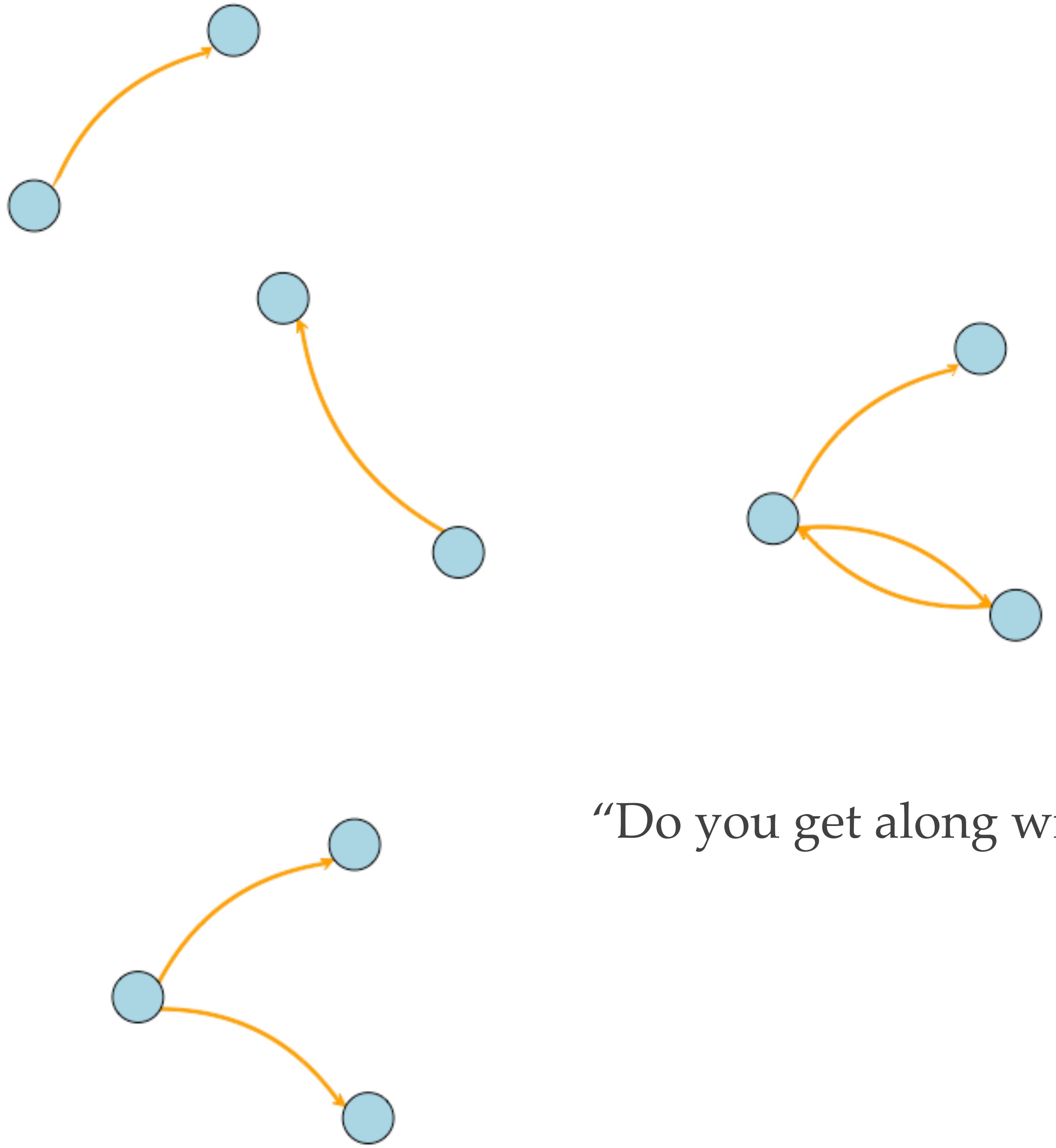
- *Networks* can differ with respect to their **nodes**:
    - One-mode/uni-partite (connections among one type of node).
    - Multi-mode/multi-partite (connections among two or more types of nodes).


Node in second mode
Node in first mode
4
2
C
D
A
5
1
3
F
B
E

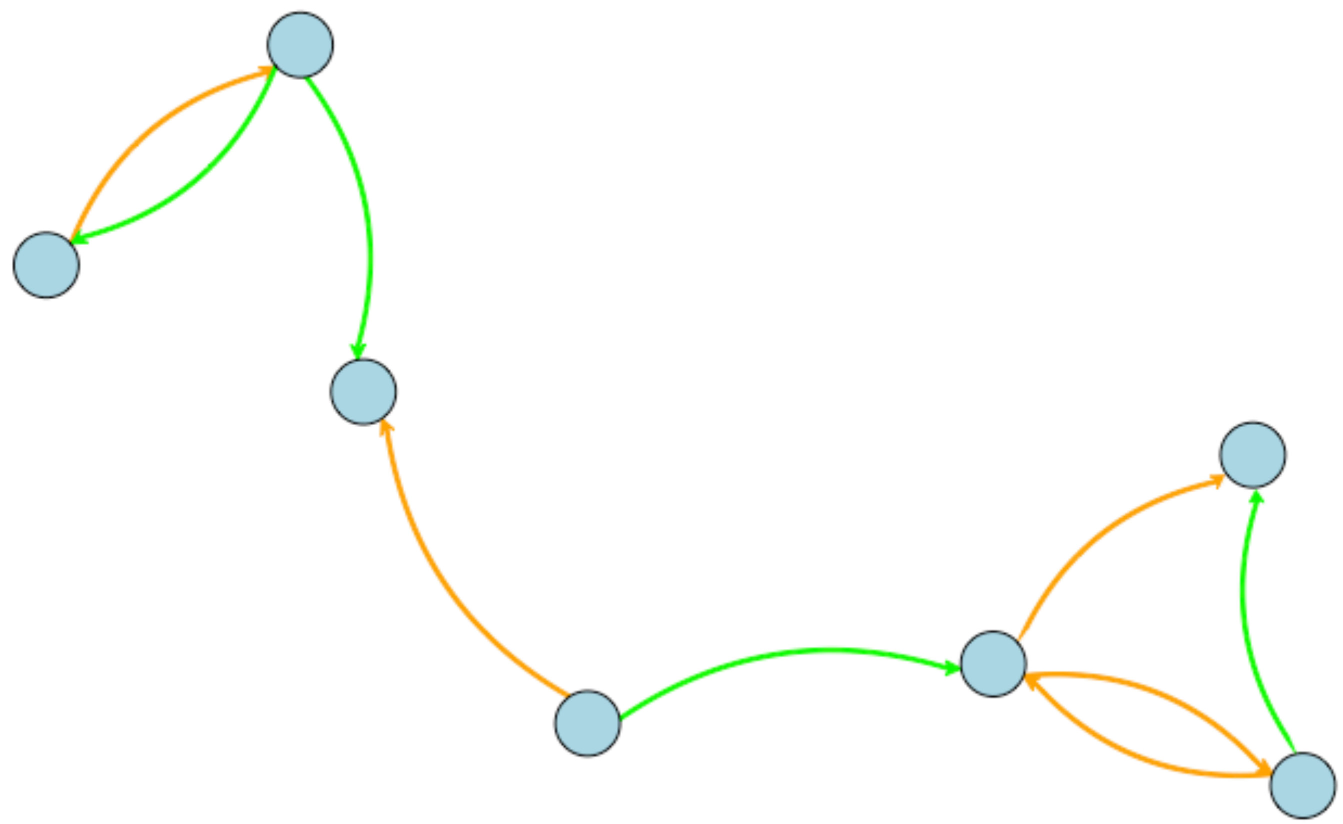
Basic Data Elements

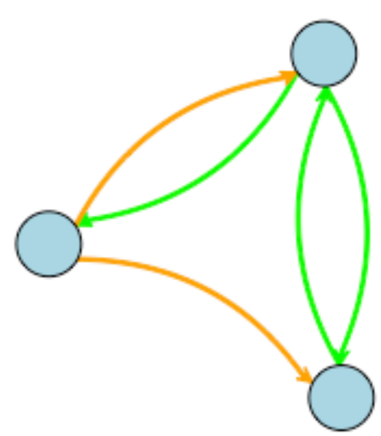
- *Networks* can differ with respect to their **edges**:
  - Simplex (connections among nodes are of one type).
  - Multiplex (connections among nodes are of multiple types).





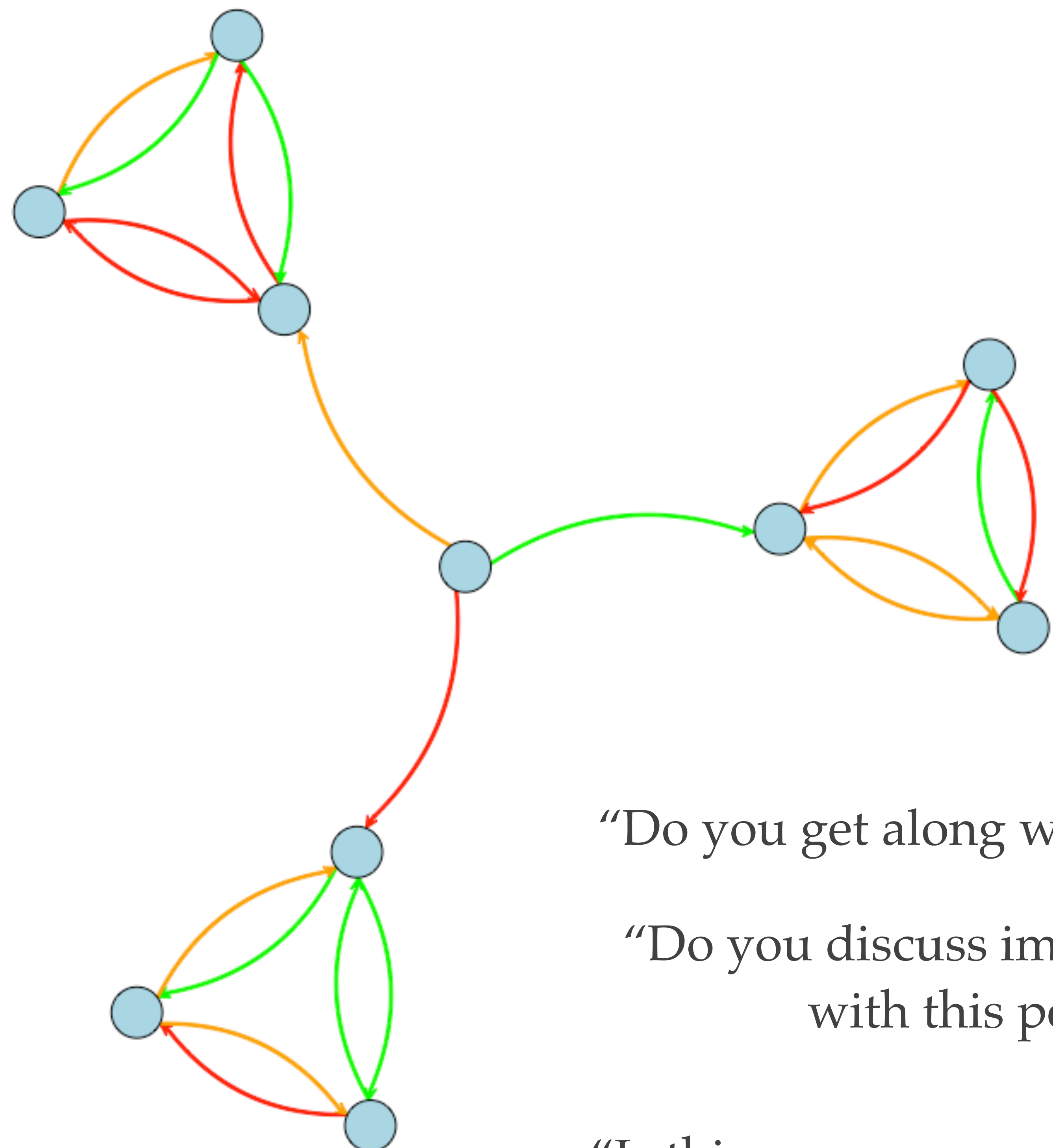
"Do you get along with this person?"





"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?"

Questions?

Network Data Collection

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

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 among 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.

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?"
  - Free response
    - "Who are the people in this school that you trust?"

Break!

Representing Networks with Graphs and Matrices

*What do network **data** "look like"?*

Graph Notation

- Definition of a **graph**: $G = (N, L)$
  - Node/Vertex set: $N = \{n_1, n_2\ldots,n_g\}$
  - Line/Edge set: $L = \{l_1, l_2\ldots,l_L\}$
    - There are N nodes/vertices and L lines/edges in a graph.

Graph Notation

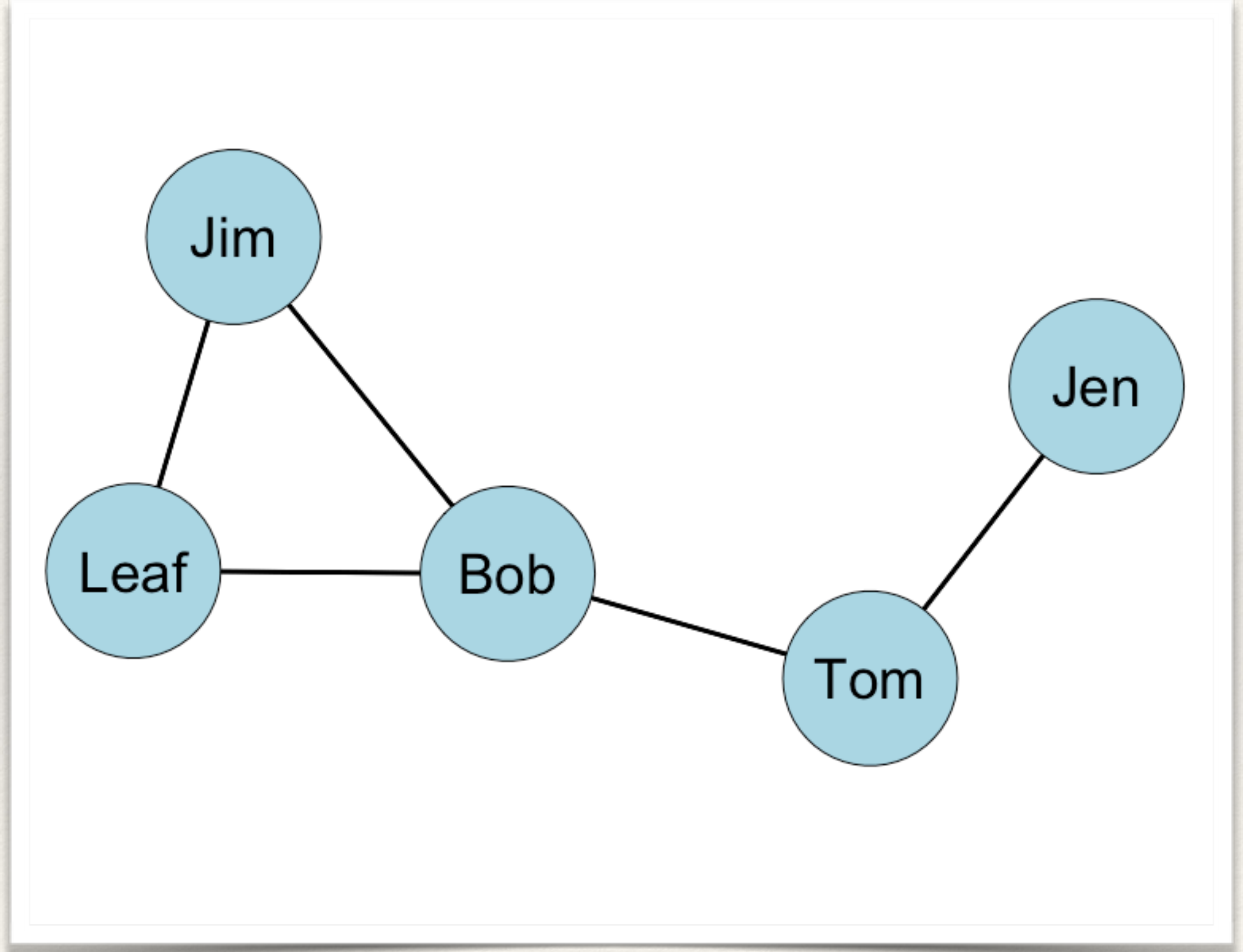
- Two nodes, n_i and n_j are **adjacent** if the line $l_k = (n_i, n_j)$
  - What this means is that in the graph, there exists a line between nodes i and j.

Example: Undirected, Binary Network

In an **undirected** graph, the order of the nodes does not matter.

In other words,

$l_k = (n_i, n_j) = (n_j, n_i)$



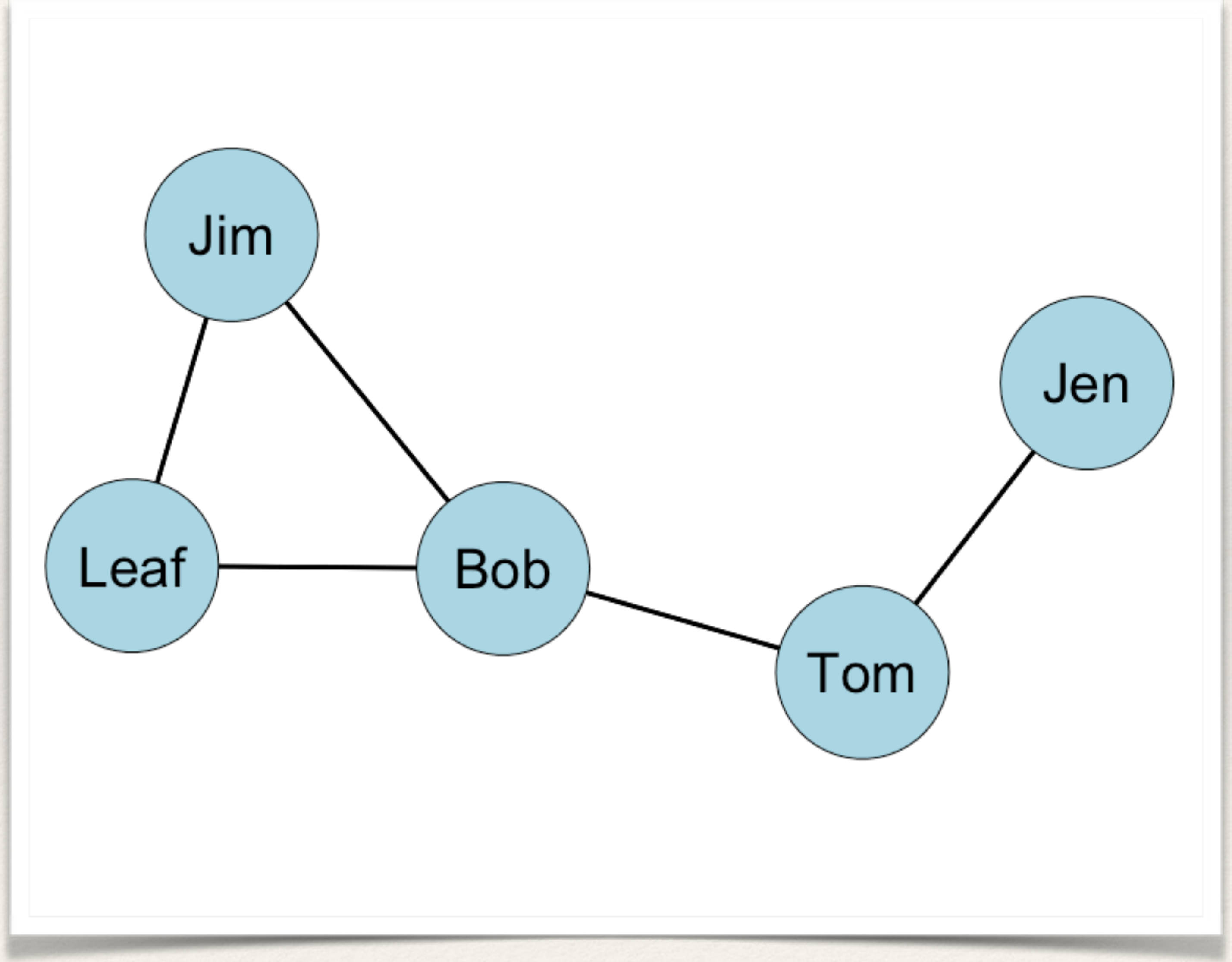
Example: Undirected, Binary Network

Let g represent the number of nodes in the graph (i.e. $g = N$).

In an **undirected** graph, there are:

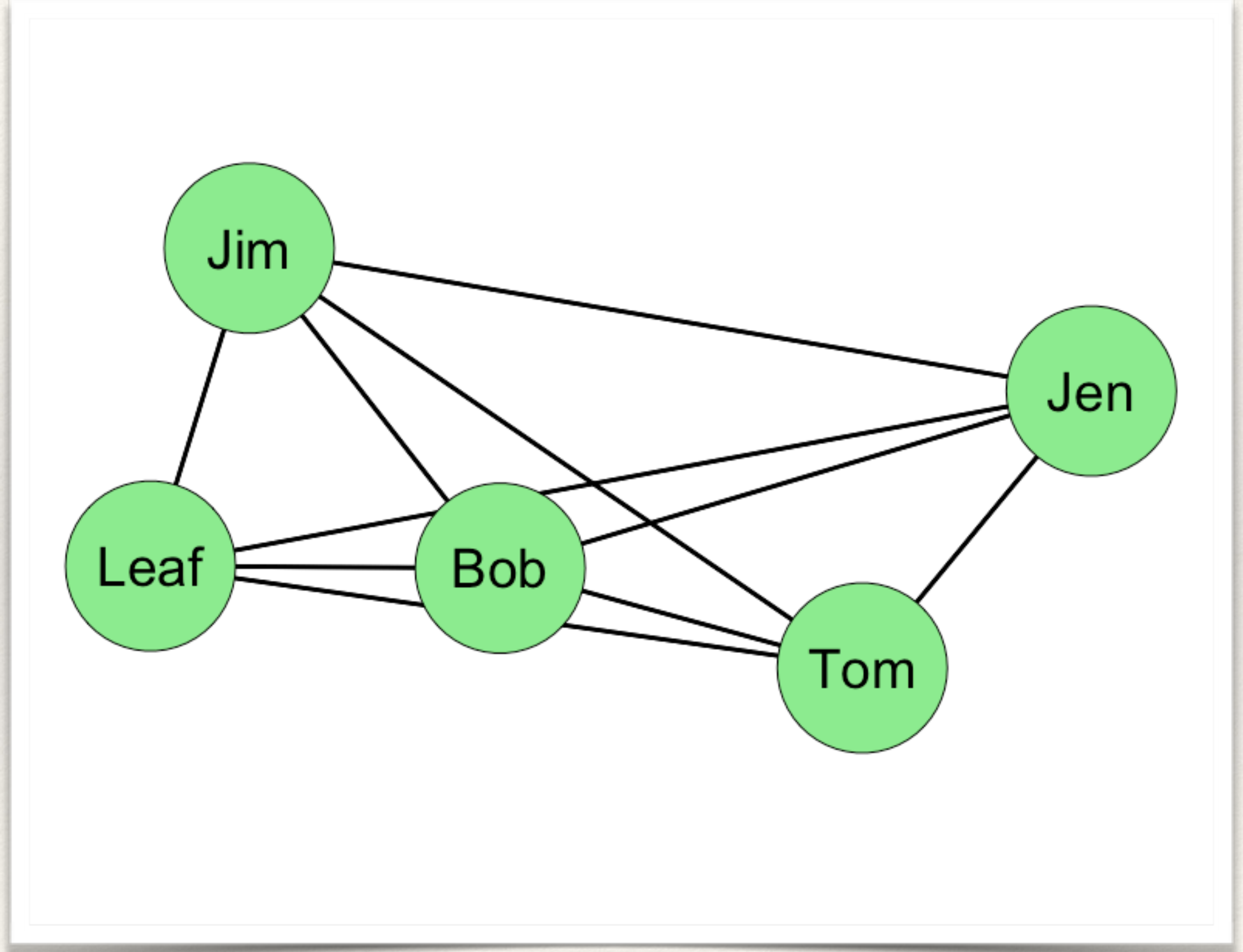
$g(g\text{-}1)/2$ possible ordered pairs.

<u>How many ordered pairs or ties could exist in this graph?</u>



Example: Undirected, Binary Network

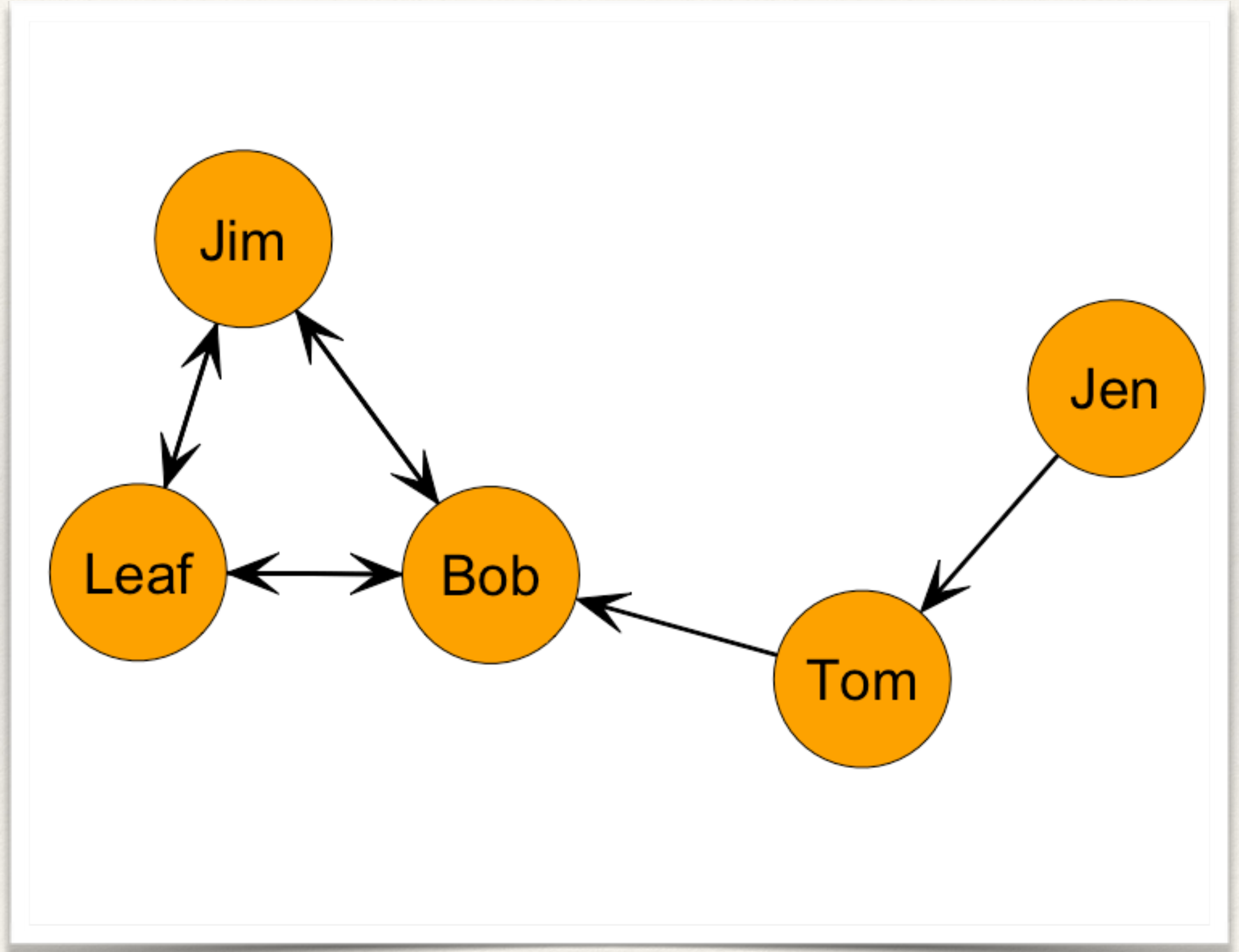
$g(g\text{-}1)/2 = 5(5\text{-}1)/2 = 10$



Example: Directed, Binary Network

In a **directed** graph, the order of the nodes does matter.

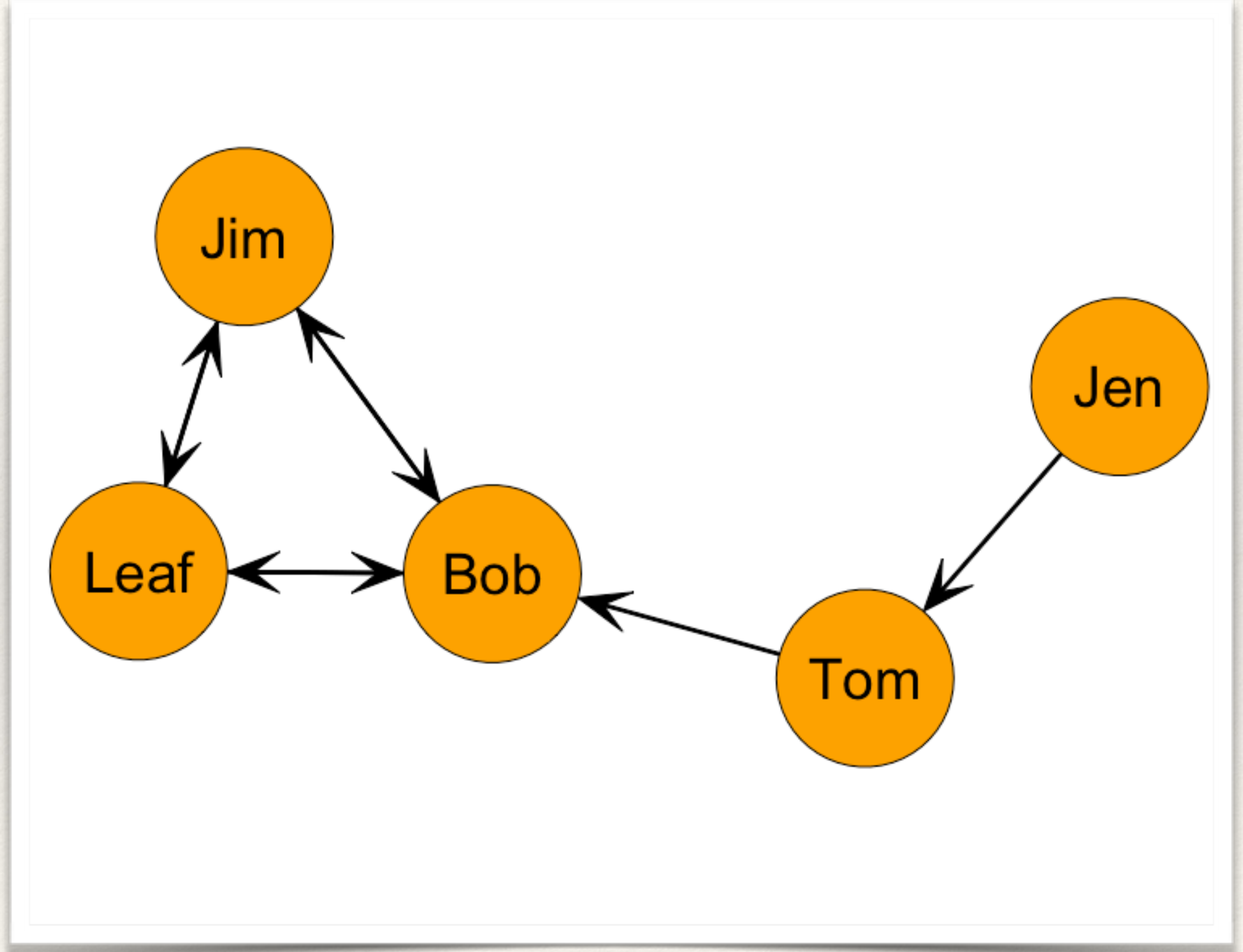
$$l_{k1} = (n_i, n_j) \neq (n_j, n_i) = l_{k2}$$



Example: Directed, Binary Network

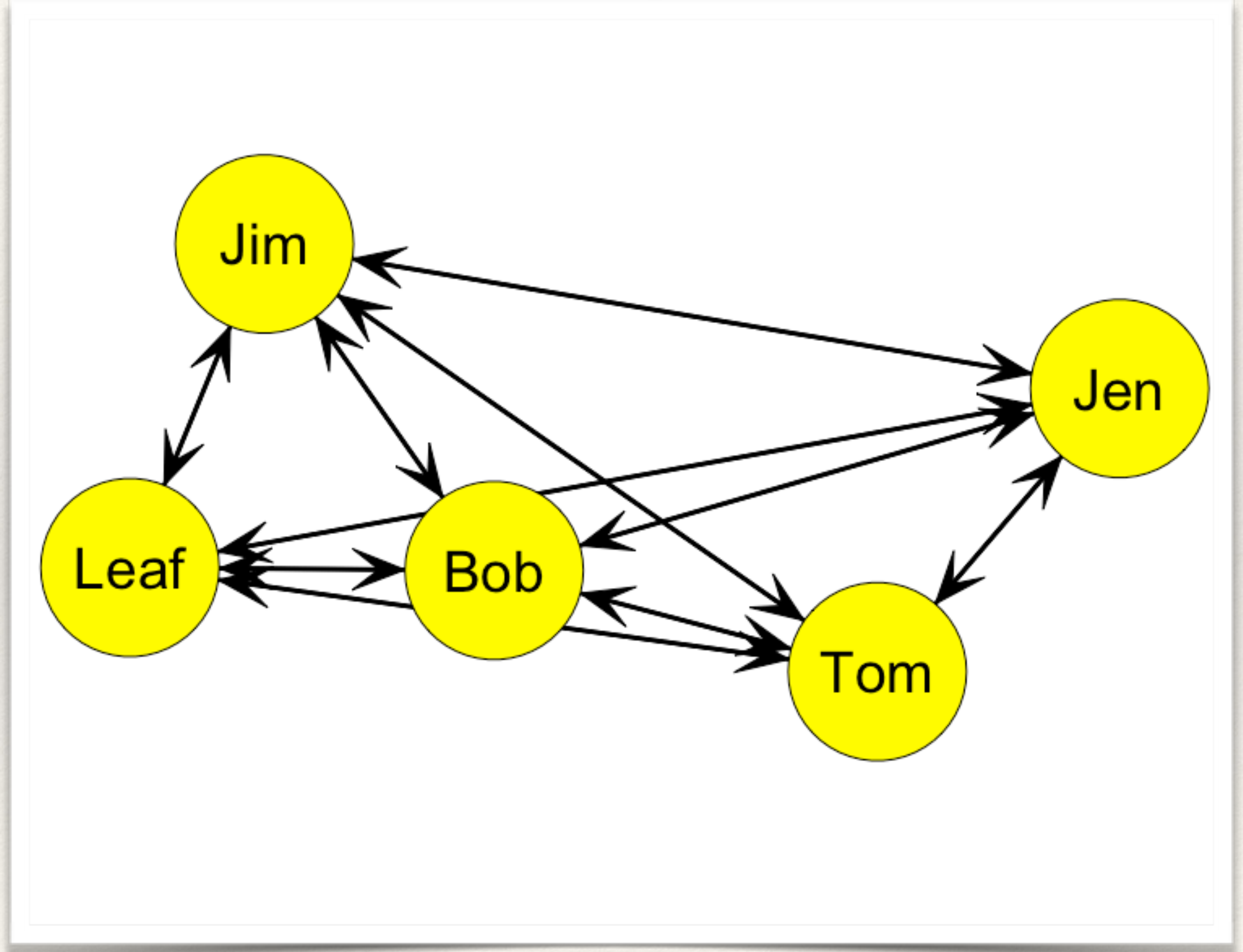
As a result, there are *g(g-1)* possible ordered pairs.

<u>How many ordered pairs or ties could exist in this graph?</u>



Example: Directed, Binary Network

$g(g\text{-}1) = 5(5\text{-}1) = 5(4) = 20$



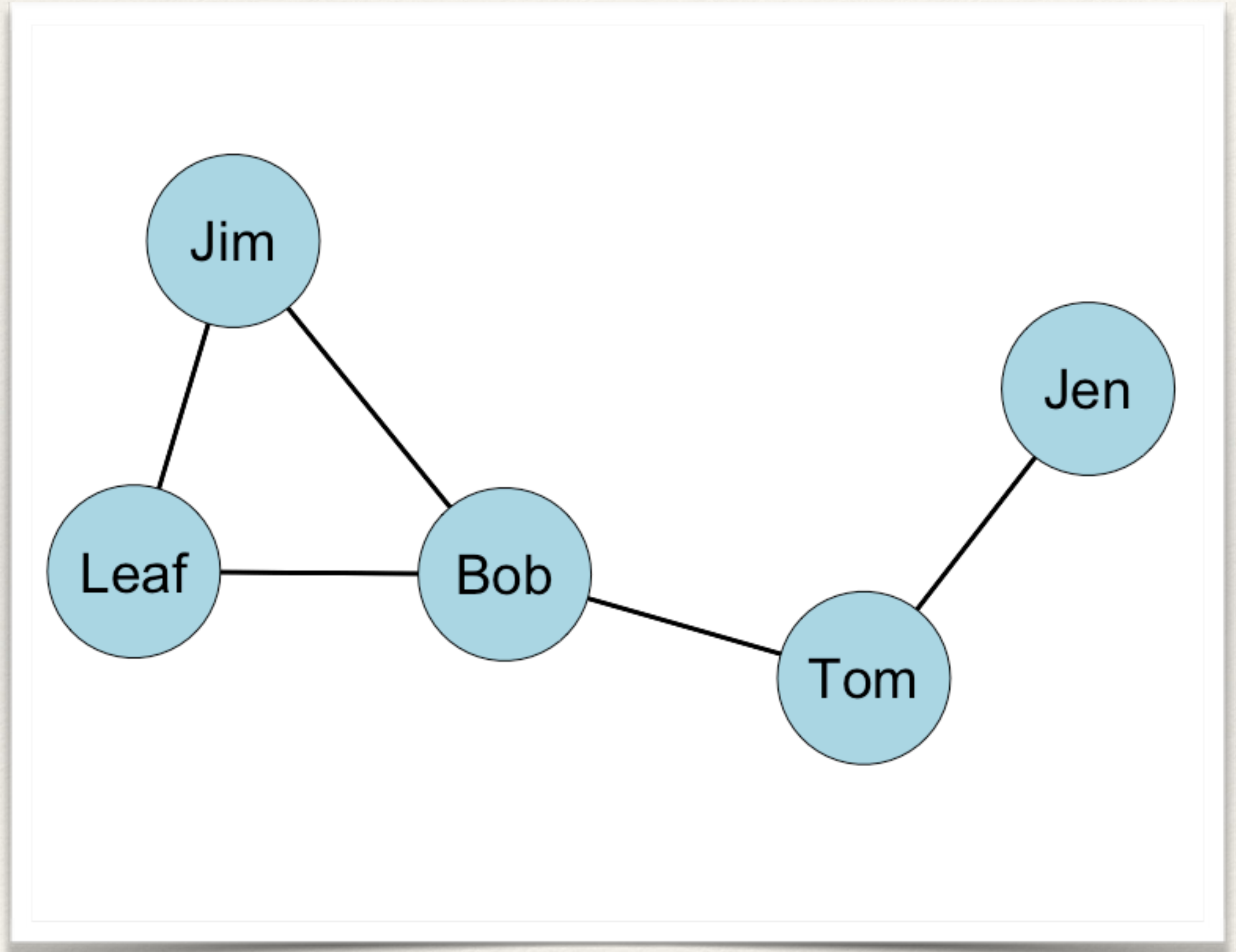
Sociometric Notation

- For a set of relations, ***X***, we can define a matrix which represents these relations.
- We commonly use an *adjacency matrix*, where each node / vertex is listed on the row and the column.
- The i_{th} row and the j_{th} column X_{ij} records the value of a tie from i to j.
- In this approach, ***X***, can be thought of as a variable.
  - The presence or absence of values in the cells represent variation.

Sociometric Notation

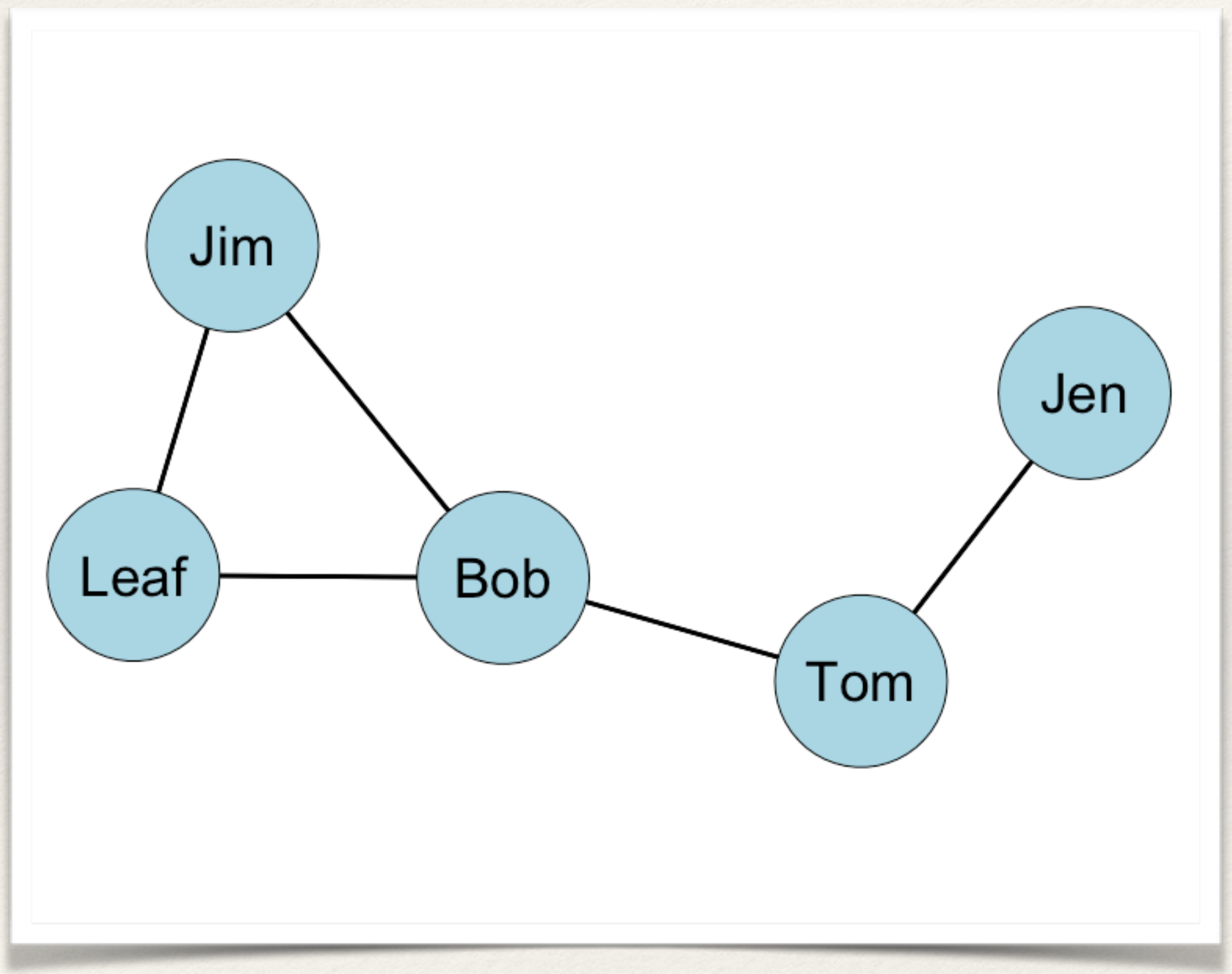
- **Definitions**
  - Scalar: a single number
  - Column vector: a column of numbers
  - Row vector: a row of numbers
  - Matrix: a rectangular array of numbers
  - Order: number of rows and columns defining the matrix
  - Square matrix: number of rows and columns of matrix are equal

Example: Undirected, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen					
Tom					
Bob					
Leaf					
Jim					

Example: Undirected, Binary Network

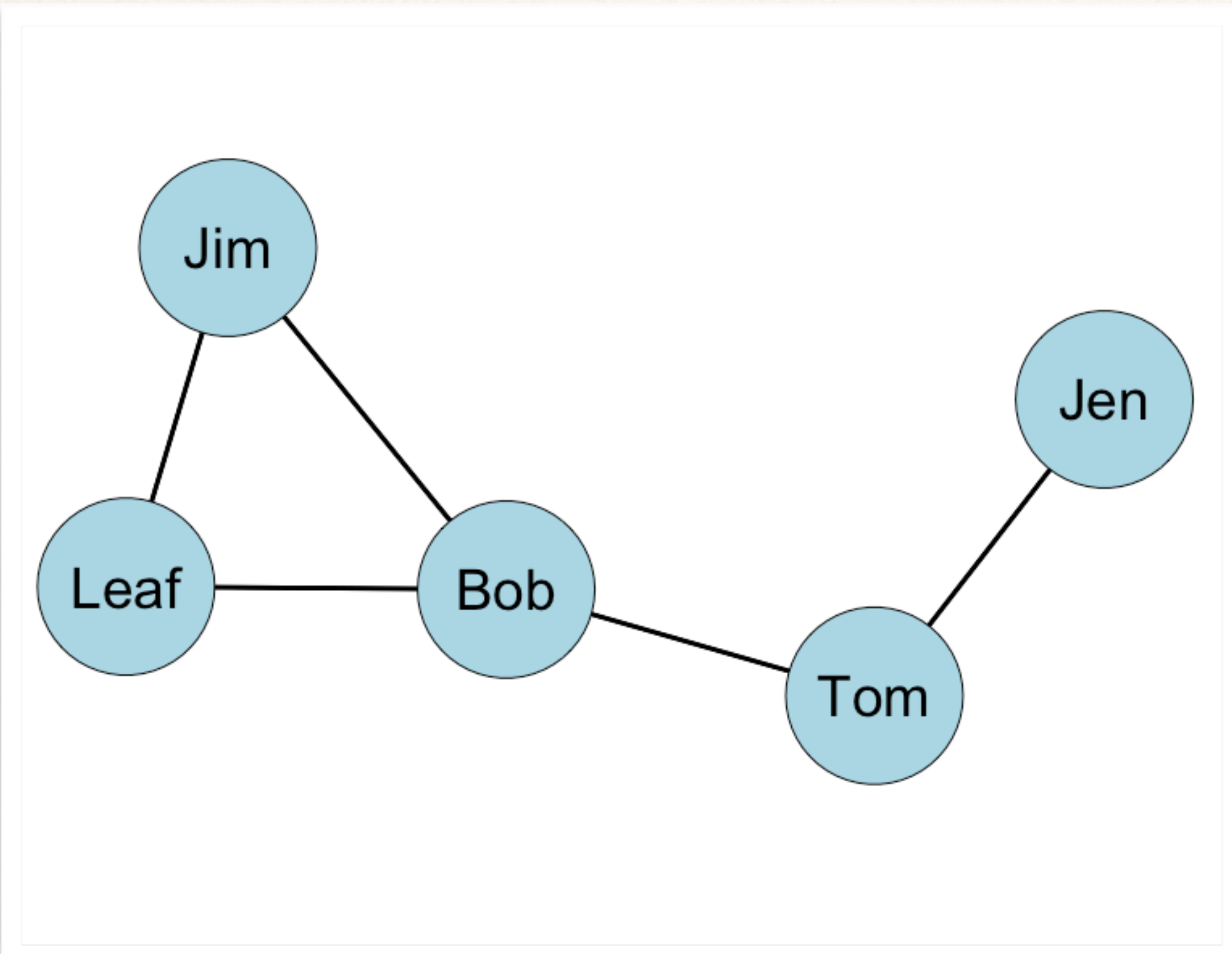


Graph or Sociogram

	Jen	Tom	Bob	Leaf	Jim
Jen					
Tom					
Bob					
Leaf					
Jim					

Adjacency Matrix or Sociomatrix

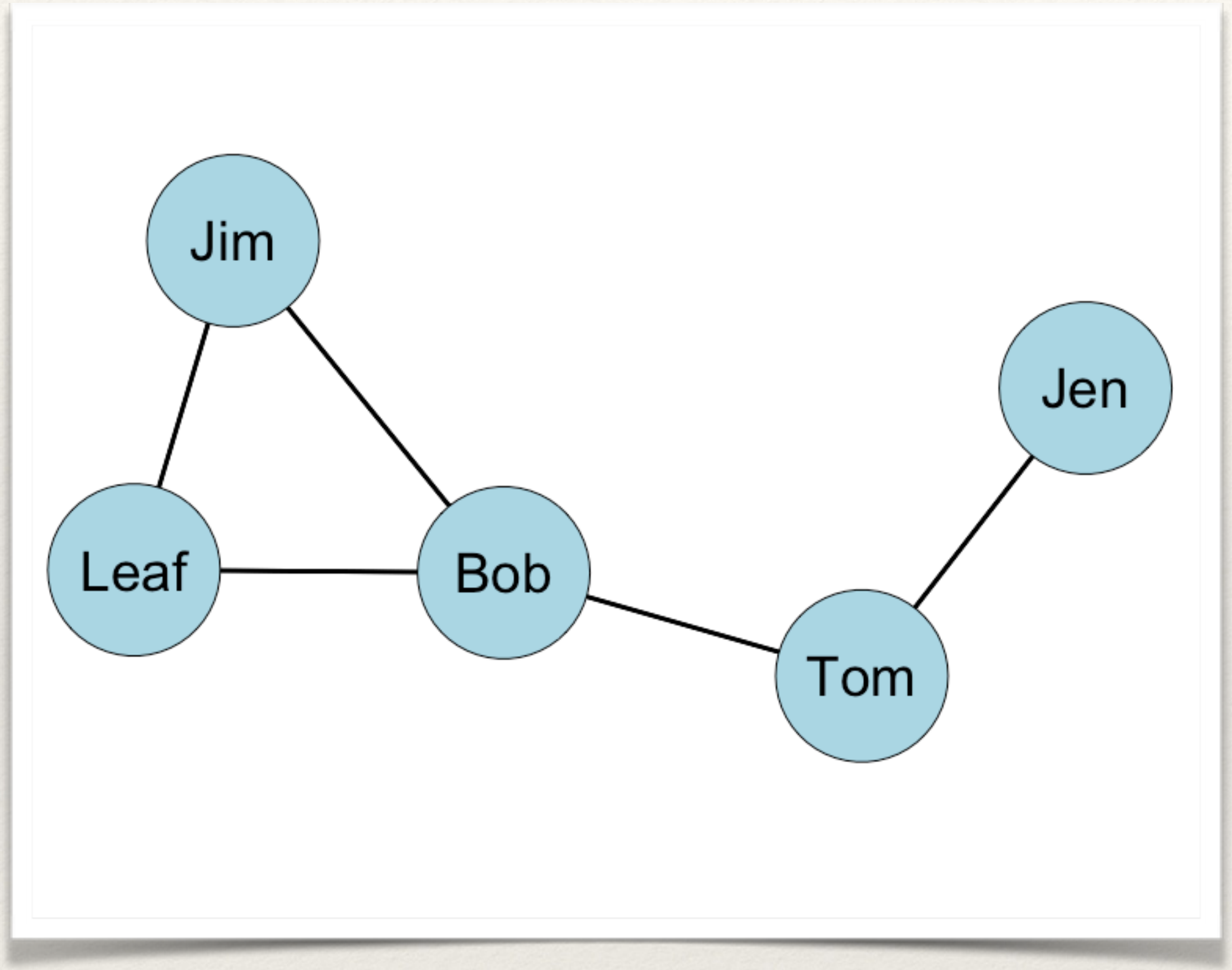
Example: Undirected, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen					
Tom					
Bob					
Leaf					
Jim					

We don't allow (in the simple case) self-nominations, so the diagonal is undefined.

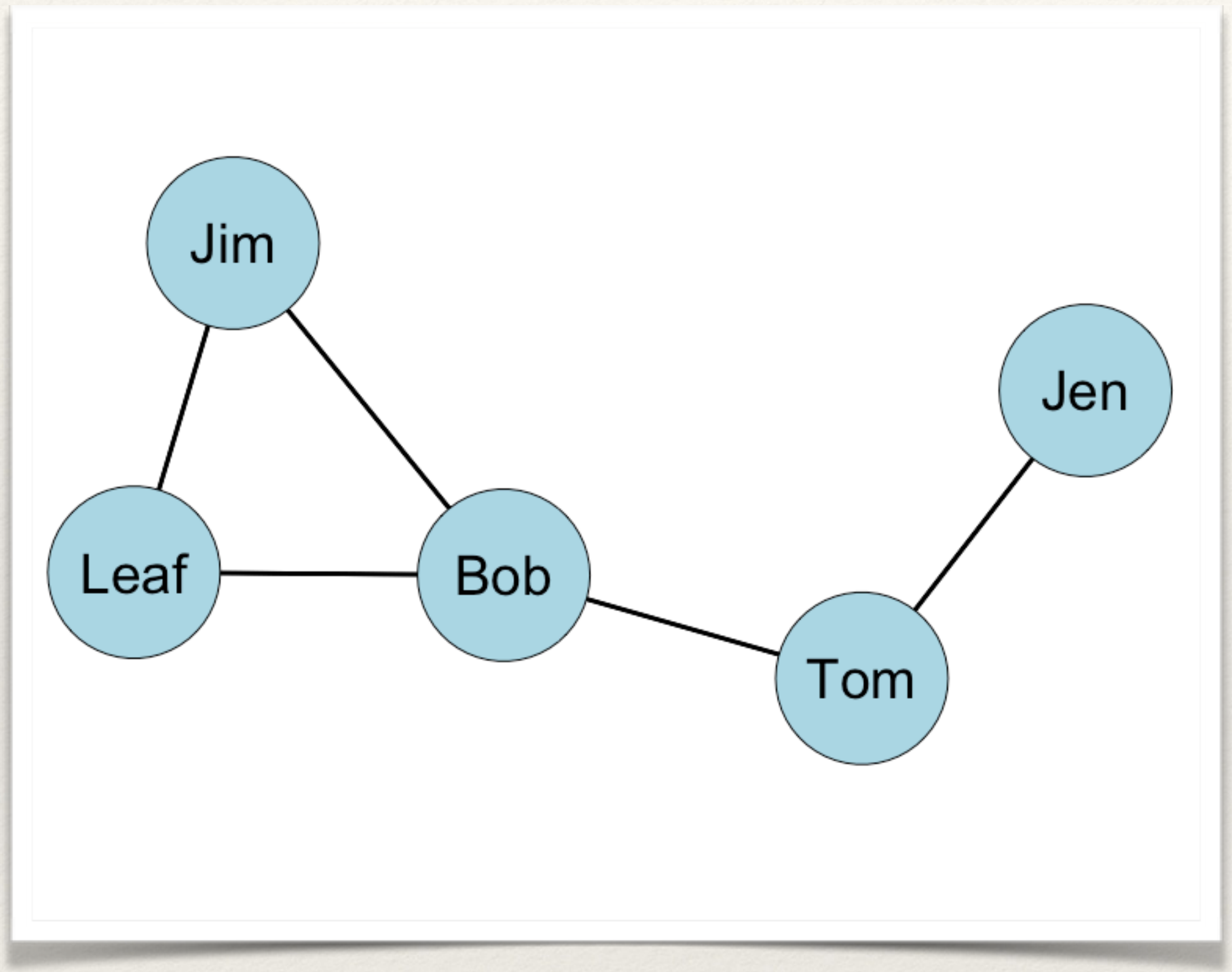
Example: Undirected, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen		1	0	0	0
Tom					
Bob					
Leaf					
Jim					

In the first row, *i* sends to the second row only: $X_{12} = 1; X_{15} = 0$

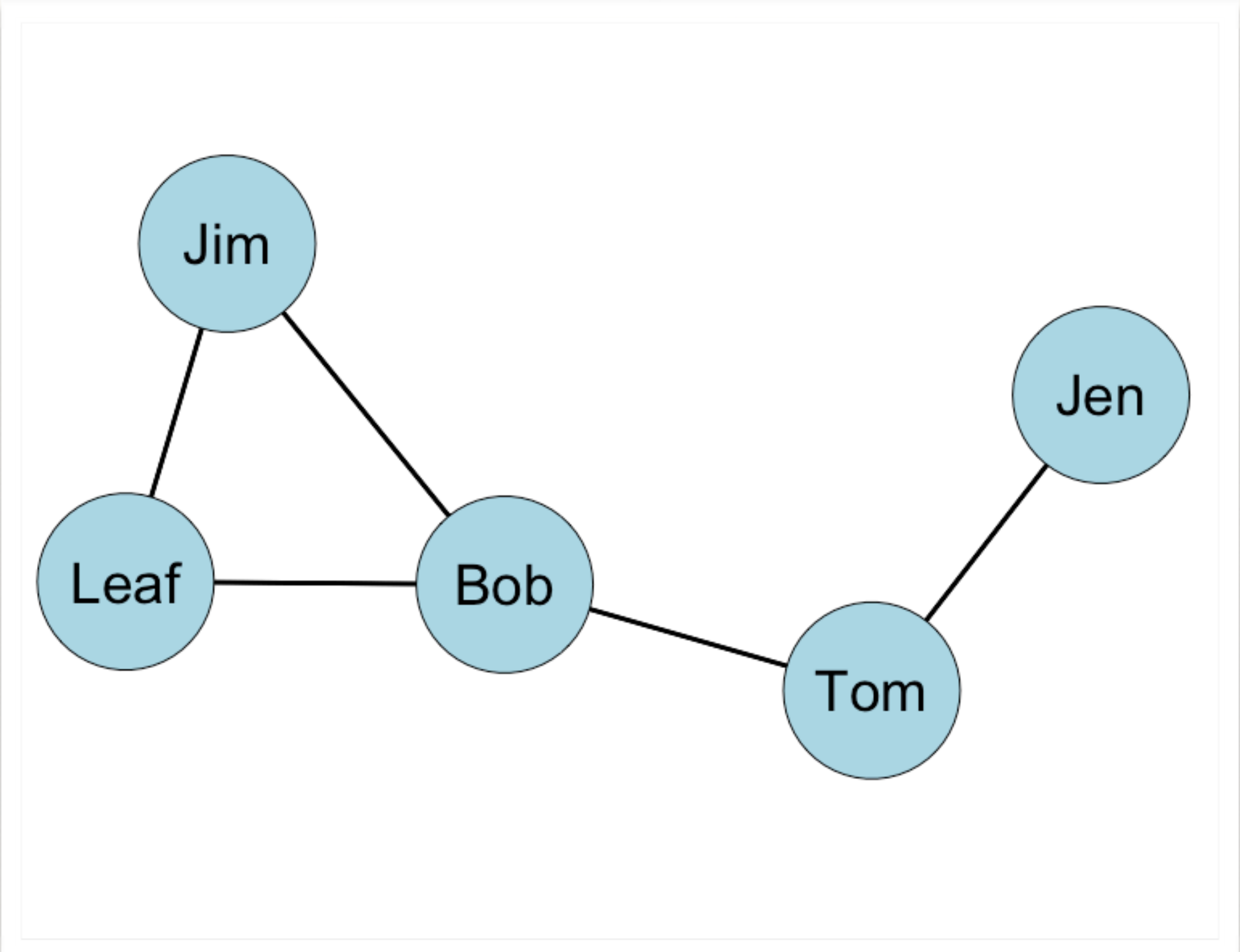
Example: Undirected, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen		1	0	0	0
Tom	1				
Bob	0				
Leaf	0				
Jim	0				

Since this is *undirected*, it is **symmetric** about the diagonal.
This means that the *ith* column is the transposition of the *ith* row.

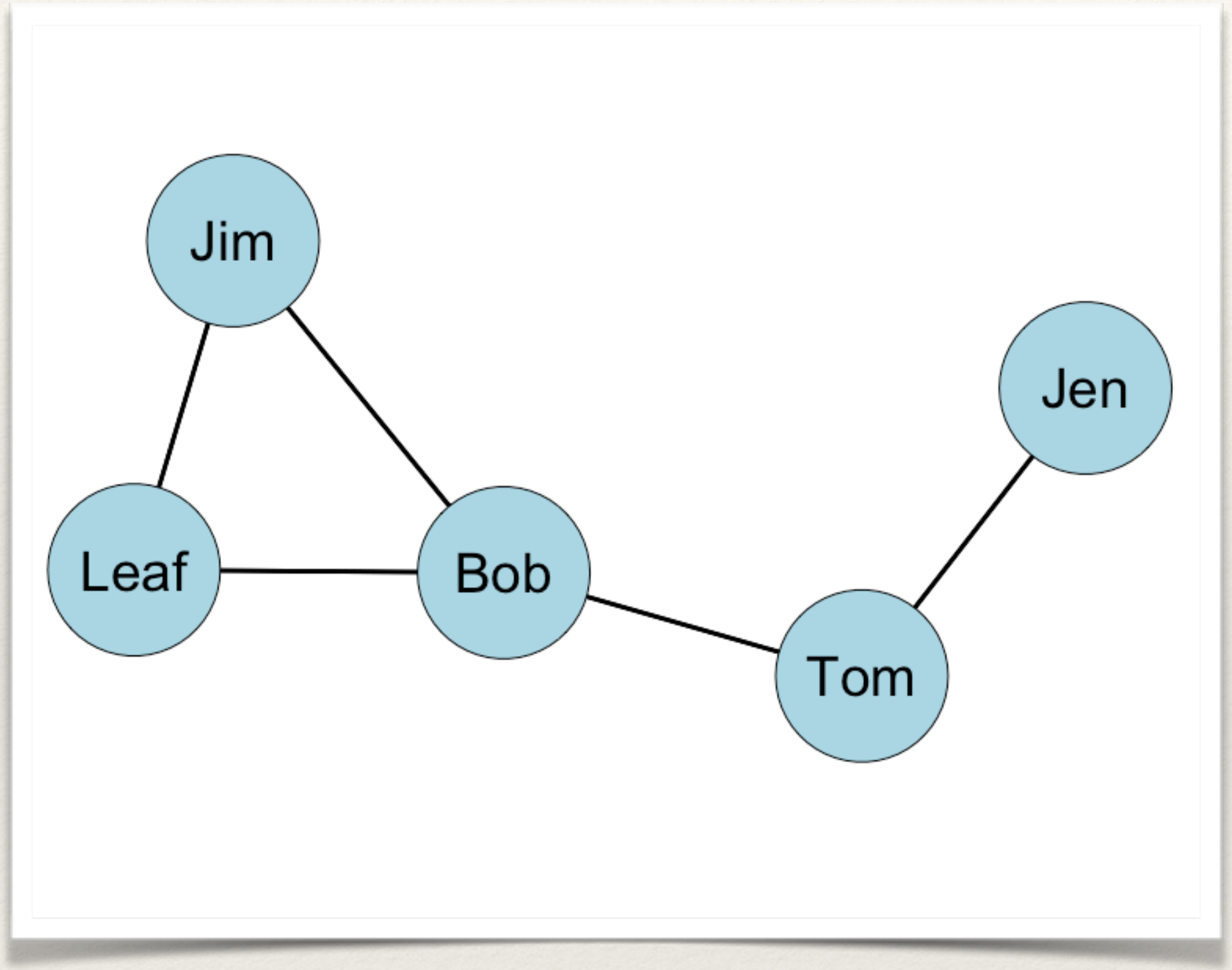
Example: Undirected, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen		1	0	0	0
Tom	1				
Bob	0				
Leaf	0				
Jim	0				

What does the rest of the matrix look like?

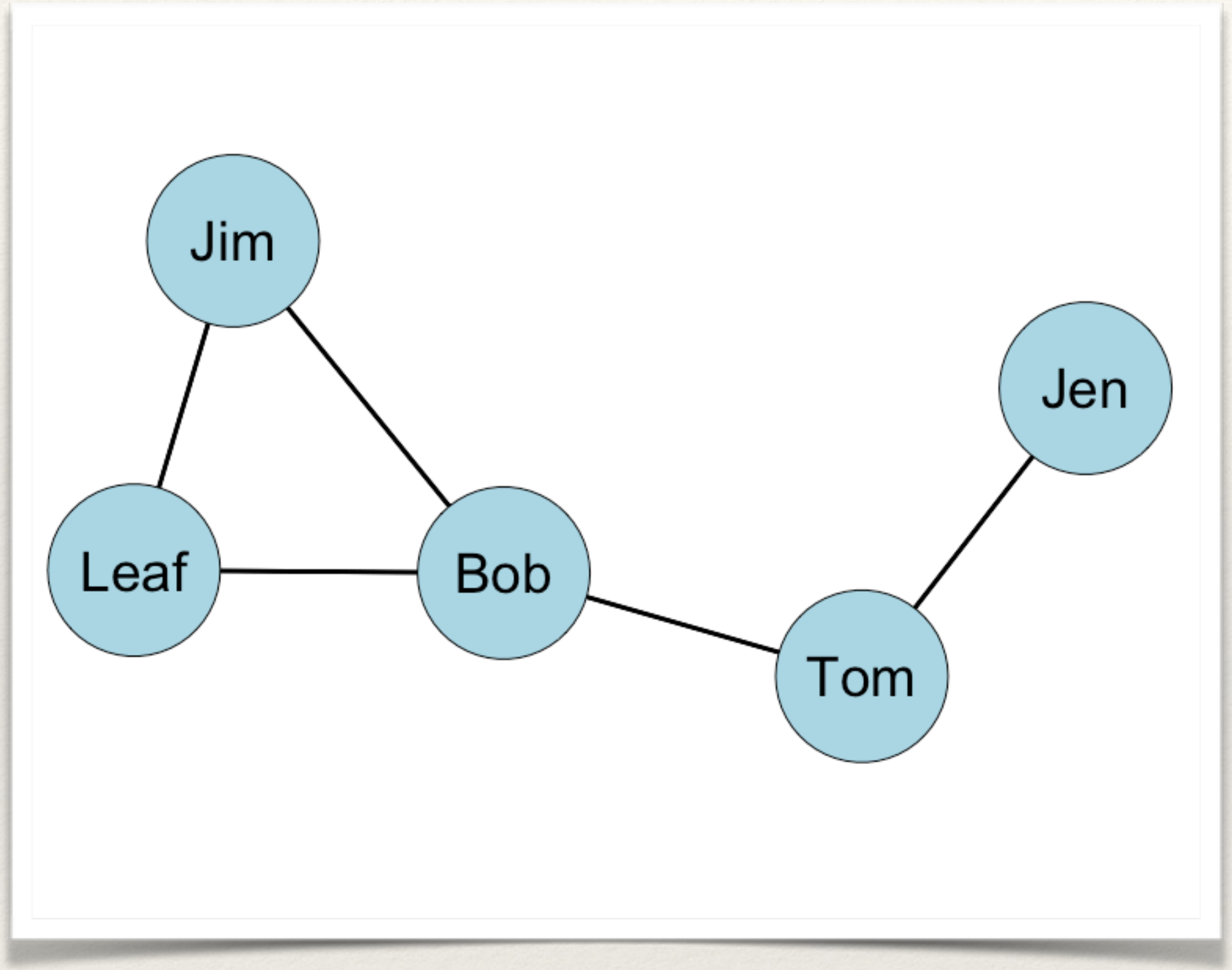
Example: Undirected, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen		1	0	0	0
Tom	1		1	0	0
Bob	0	1		1	1
Leaf	0	0	1		1
Jim	0	0	1	1	

It looks like this.

Example: Undirected, Binary Network

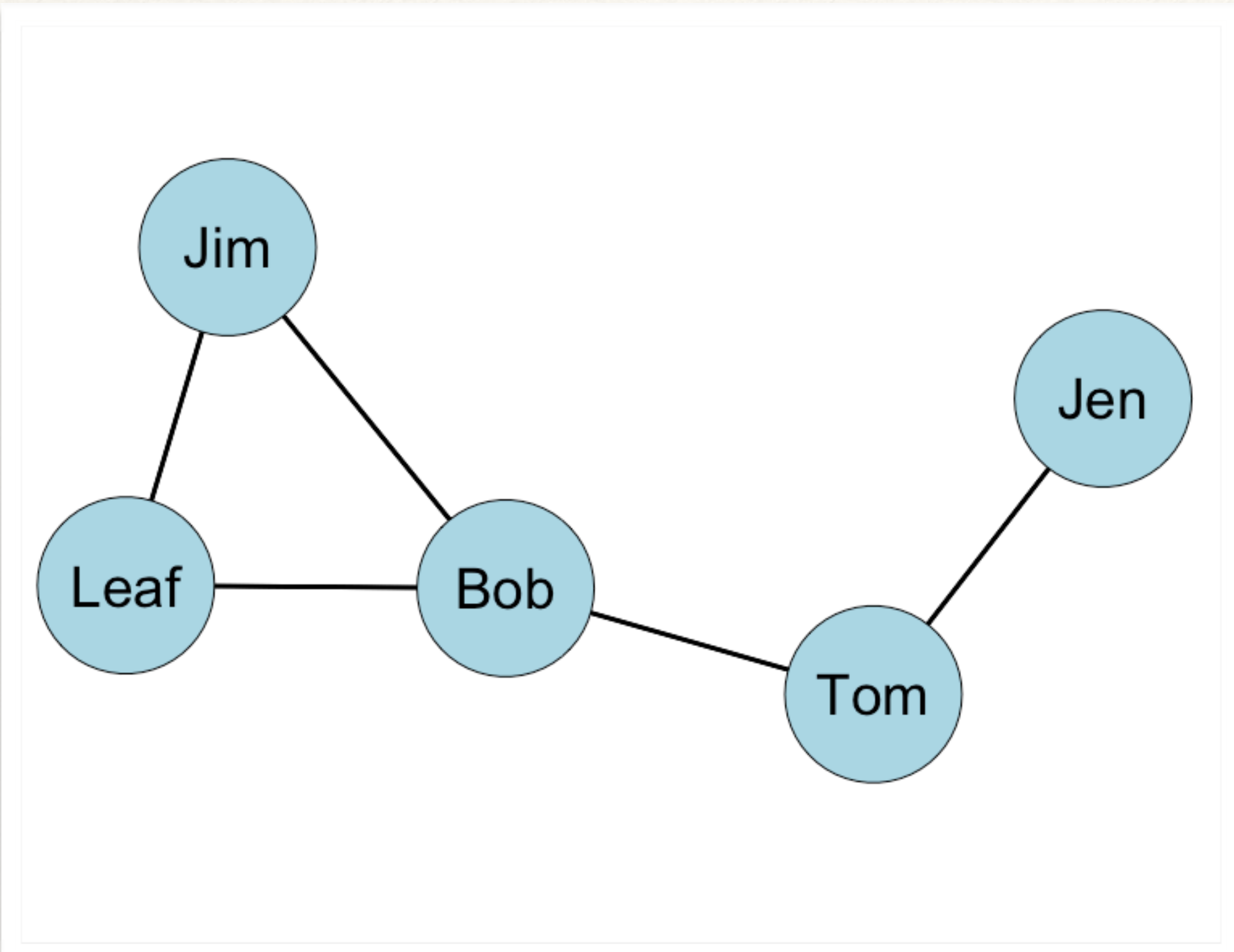


It looks like this.

Let's add zeros to the diagonals. (will explain this later…)

	Jen	Tom	Bob	Leaf	Jim
Jen	0	1	0	0	0
Tom	1	0	1	0	0
Bob	0	1	0	1	1
Leaf	0	0	1	0	1
Jim	0	0	1	1	0

Example: Undirected, Binary Network

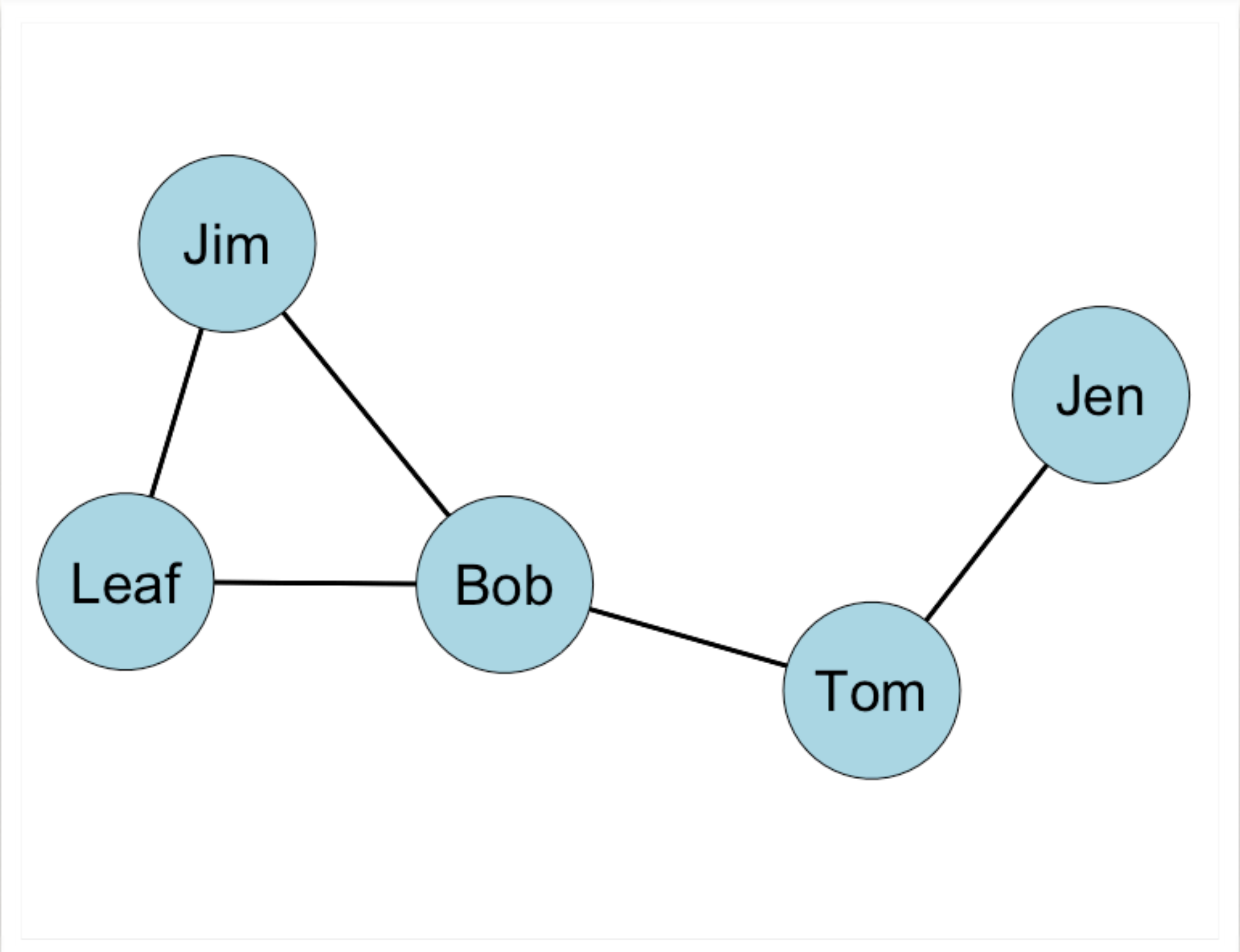


	Jen	Tom	Bob	Leaf	Jim
Jen	0	1	0	0	0
Tom	1	0	1	0	0
Bob	0	1	0	1	1
Leaf	0	0	1	0	1
Jim	0	0	1	1	0

The highlighted section here is called the **lower triangle** of the matrix.

The **sum** of the lower triangle should equal the number of edges in the graph.

Example: Undirected, Binary Network

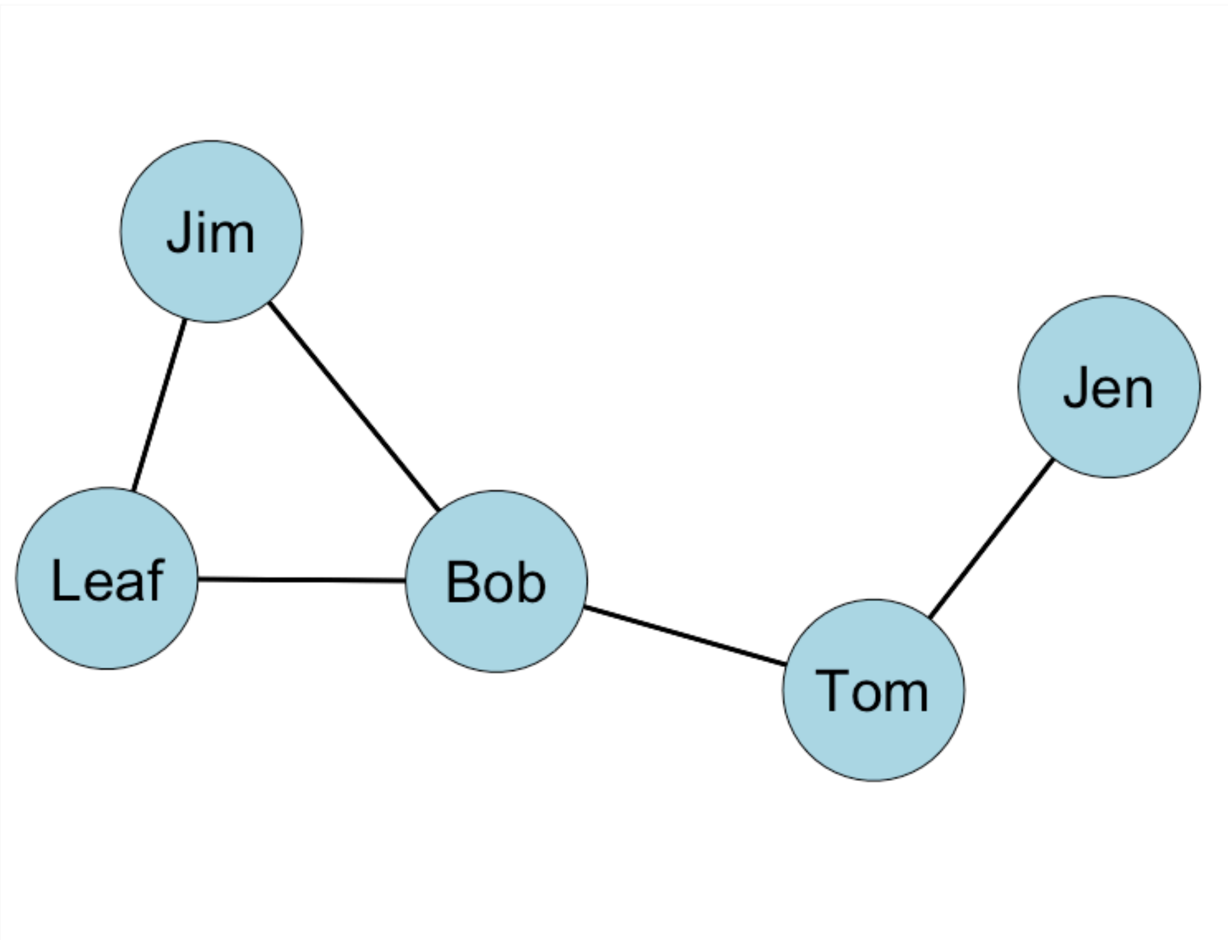


The other highlighted section here is called the **upper triangle** of the matrix.

The **sum** of the upper triangle should also equal the number of edges in the graph.

	Jen	Tom	Bob	Leaf	Jim
Jen	0	1	0	0	0
Tom	1	0	1	0	0
Bob	0	1	0	1	1
Leaf	0	0	1	0	1
Jim	0	0	1	1	0

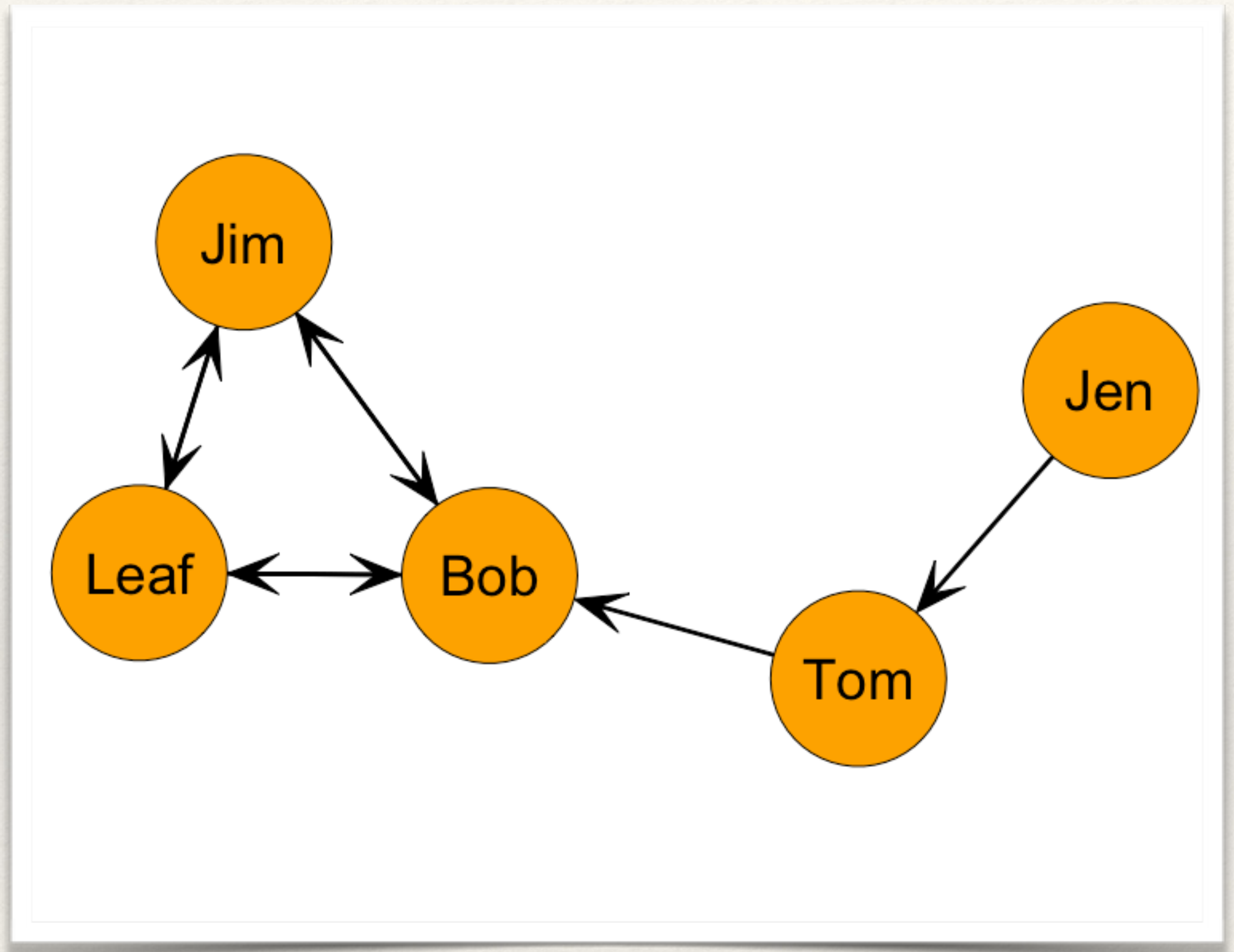
Example: Undirected, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen	0	1	0	0	0
Tom	1	0	1	0	0
Bob	0	1	0	1	1
Leaf	0	0	1	0	1
Jim	0	0	1	1	0

Alternatively, we could sum all the elements and divide by 2.

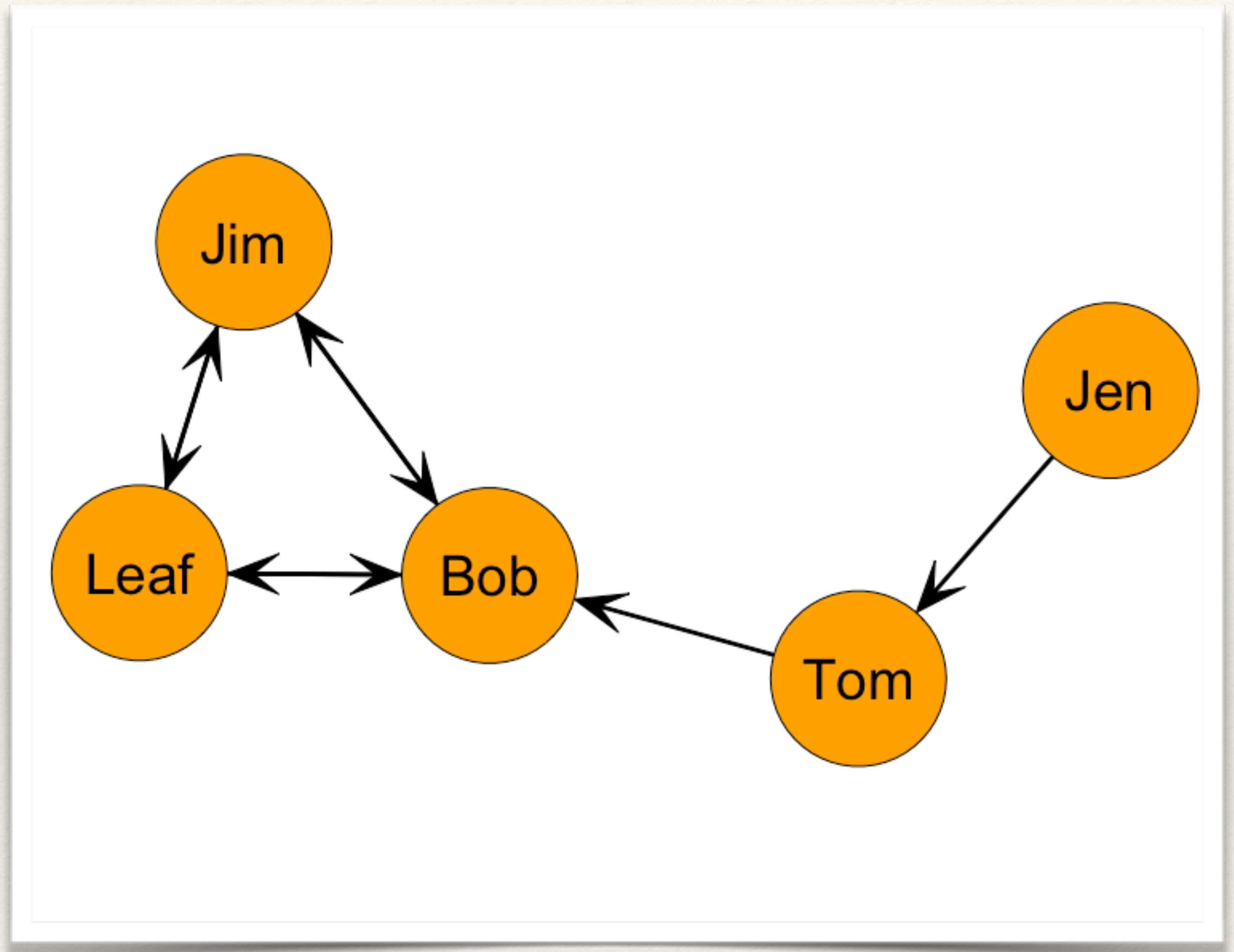
Example: Directed, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen					
Tom					
Bob					
Leaf					
Jim					

What's different about a directed network?

Example: Directed, Binary Network

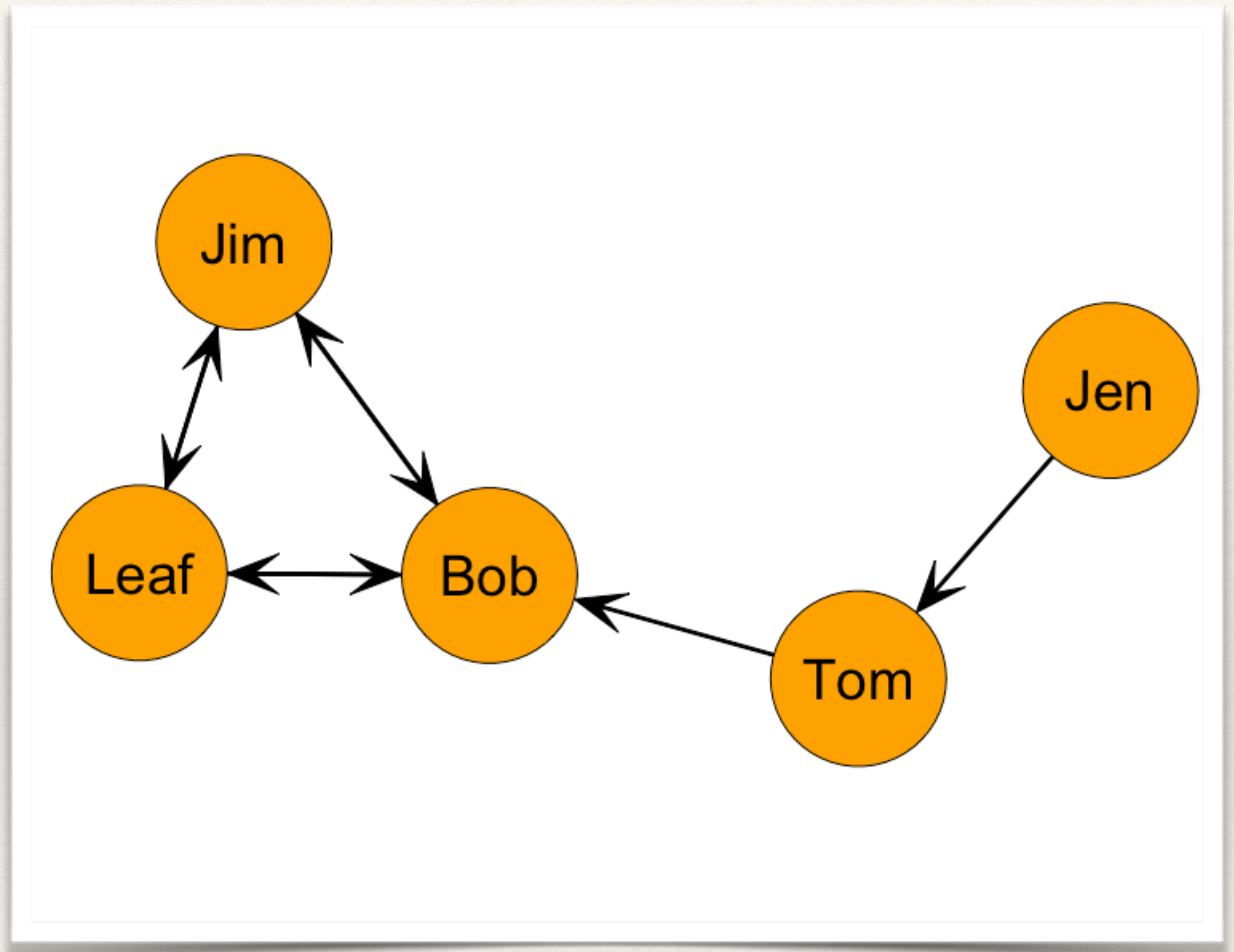


	Jen	Tom	Bob	Leaf	Jim
Jen		1			
Tom					
Bob					
Leaf					
Jim					

In the first row, *i* sends to the second row:

$$X_{12} = 1$$

Example: Directed, Binary Network

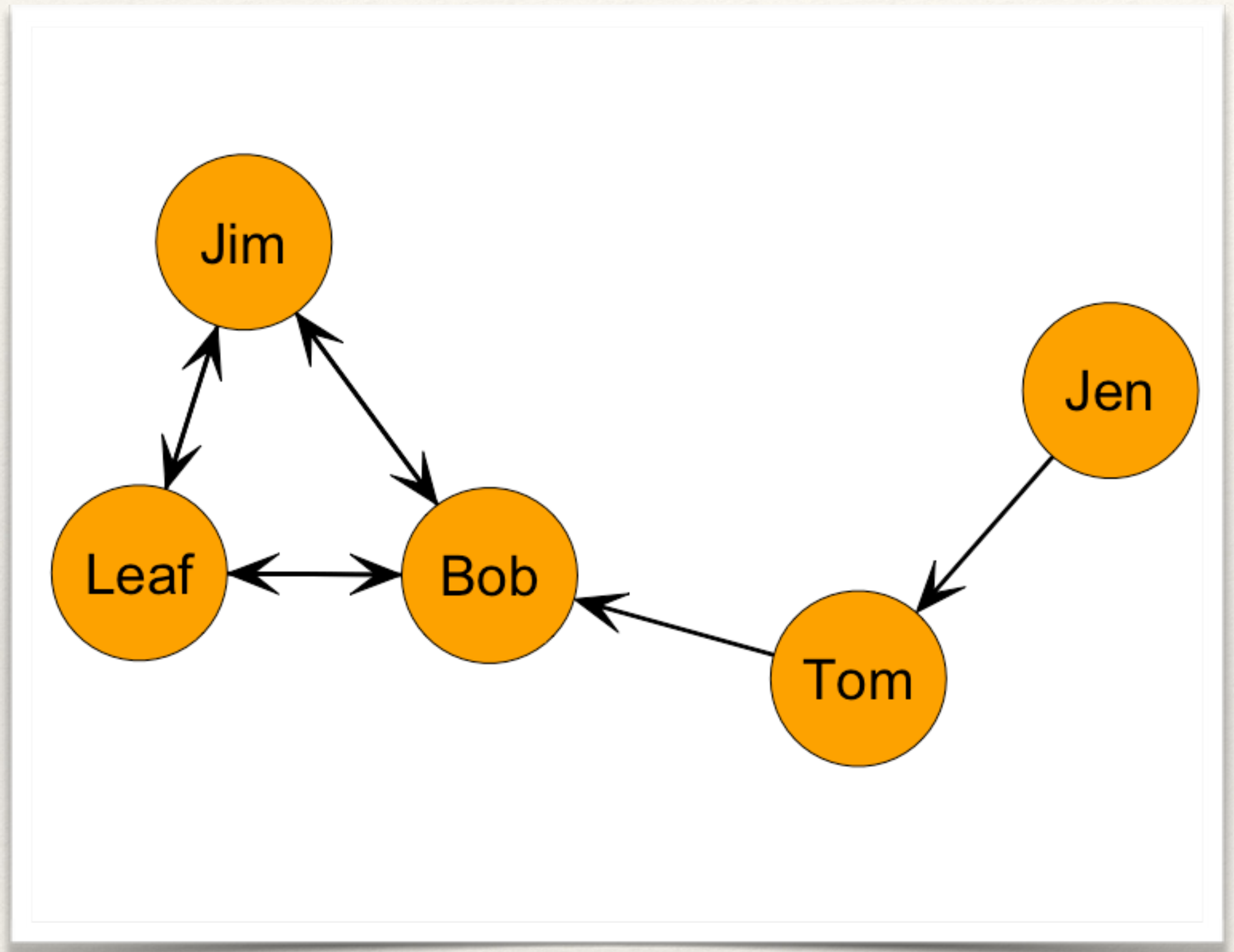


	Jen	Tom	Bob	Leaf	Jim
Jen		1			
Tom	0				
Bob					
Leaf					
Jim					

But in the second row, j does not send:

$X_{21} = 0$

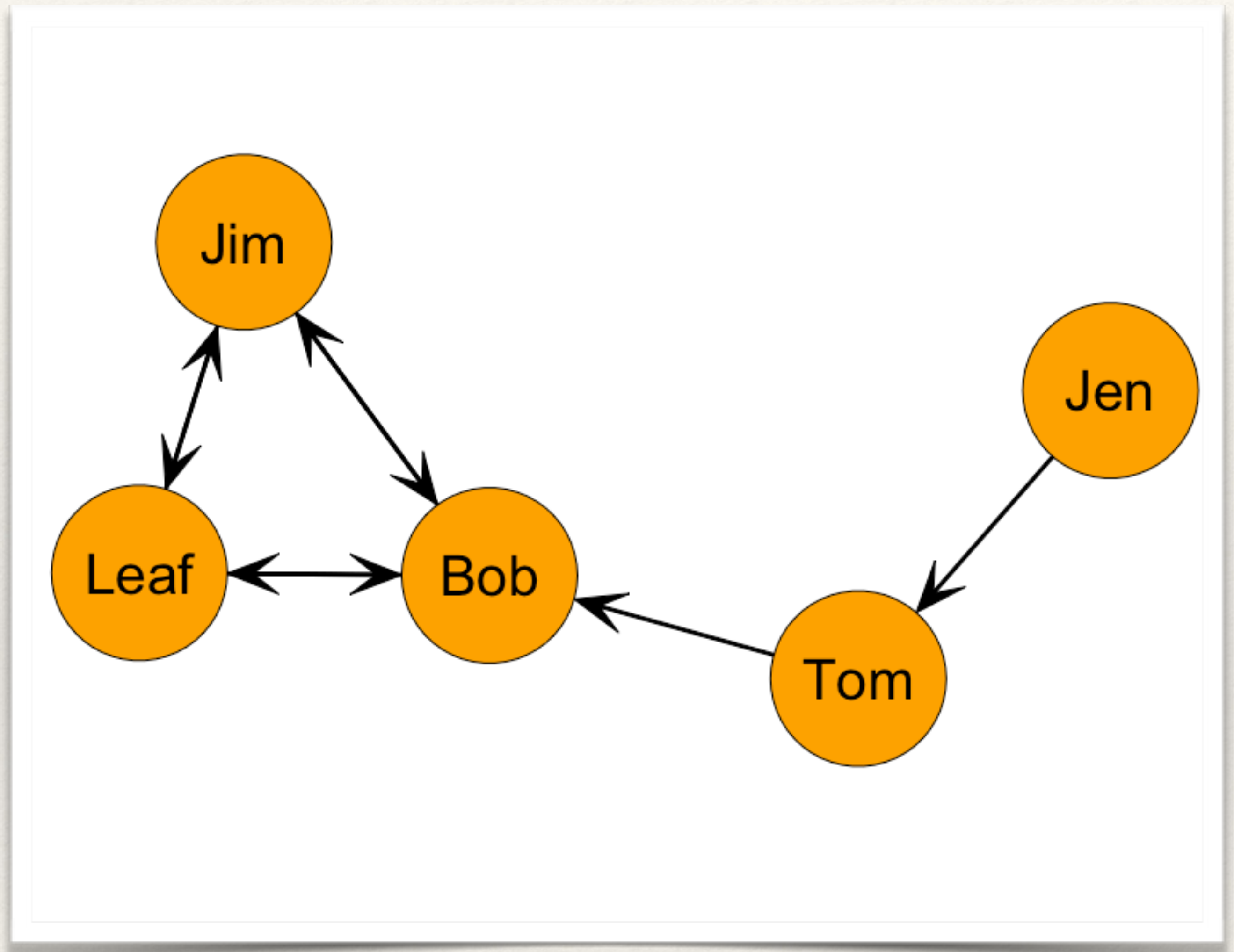
Example: Directed, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen		1			
Tom	0				
Bob					
Leaf					
Jim					

The Jen/Tom dyad is **asymmetric**. So, directed graphs permit asymmetry.

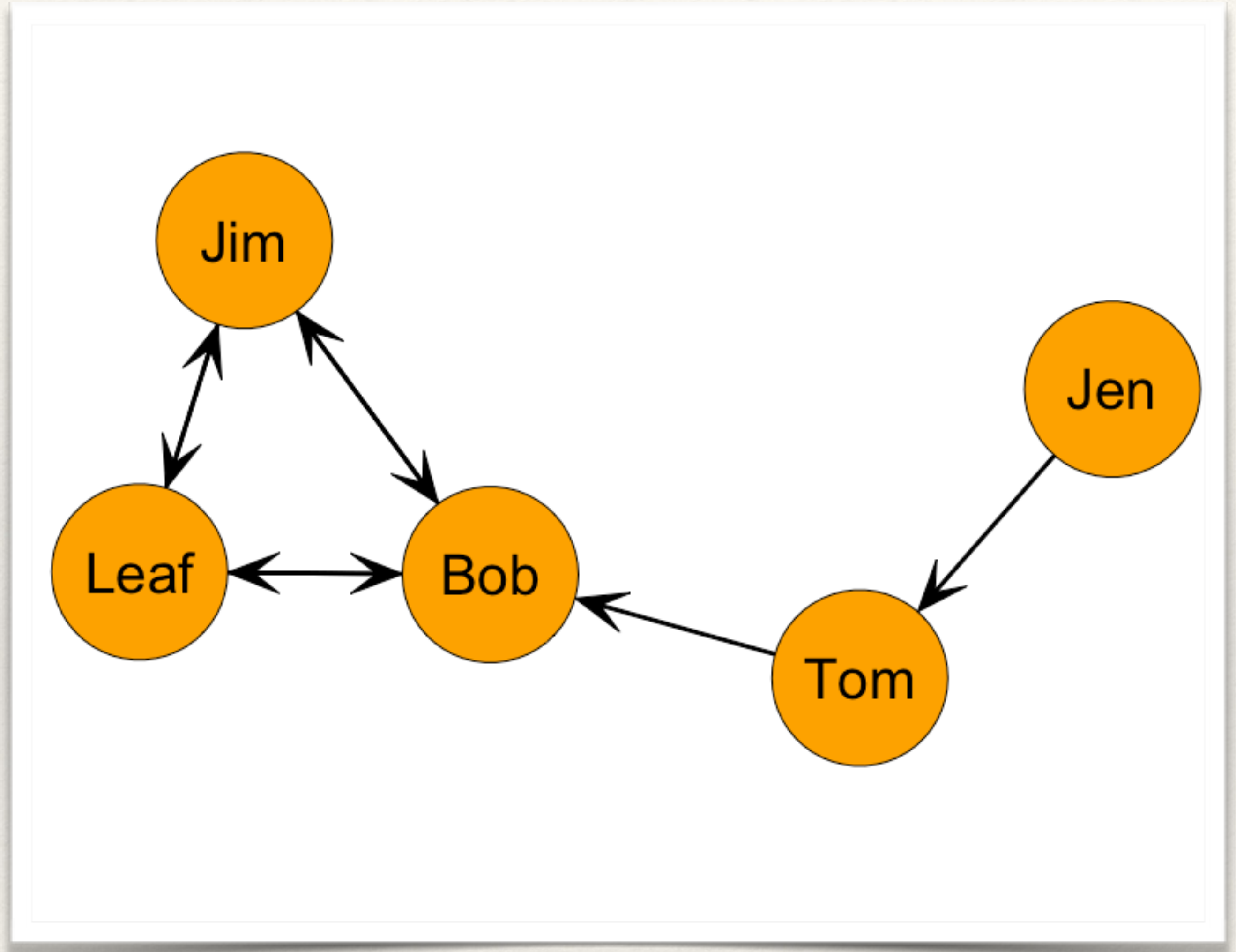
Example: Directed, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen		1			
Tom	0				
Bob				1	
Leaf			1		
Jim					

What about the Leaf/Bob dyad? Is it **asymmetric** or is it **symmetric**?

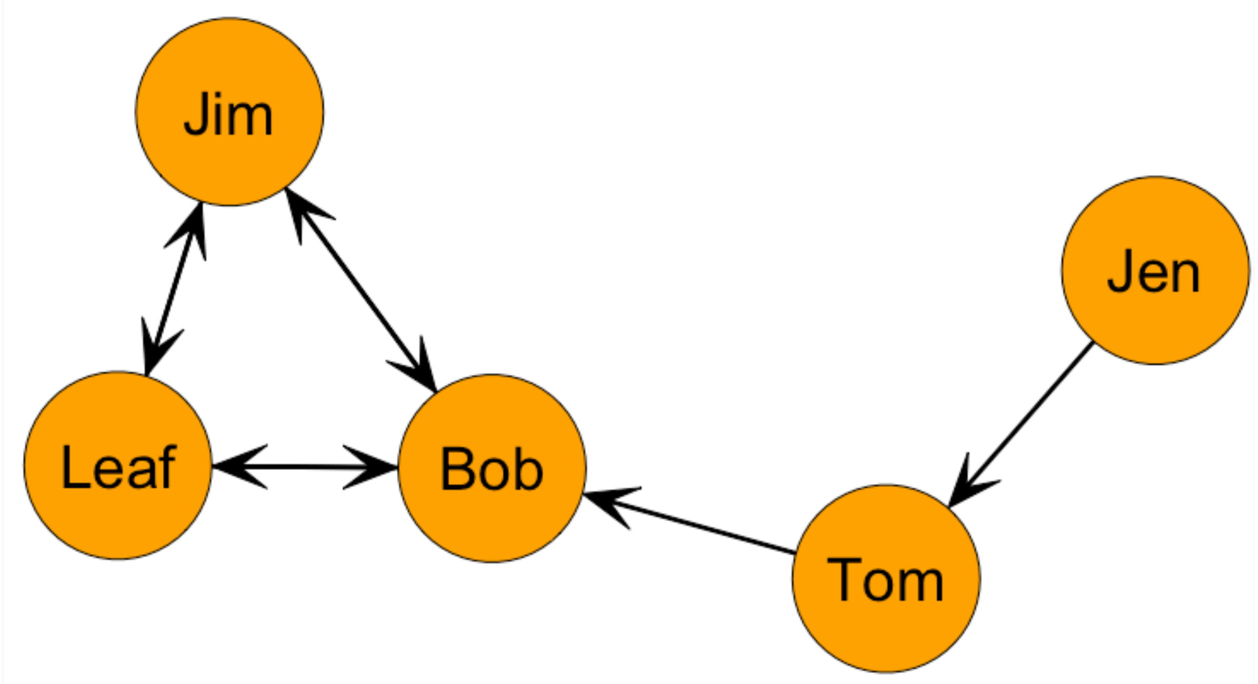
Example: Directed, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen		1			
Tom	0				
Bob				1	
Leaf			1		
Jim					

What does the rest of the matrix look like?

Example: Directed, Binary Network

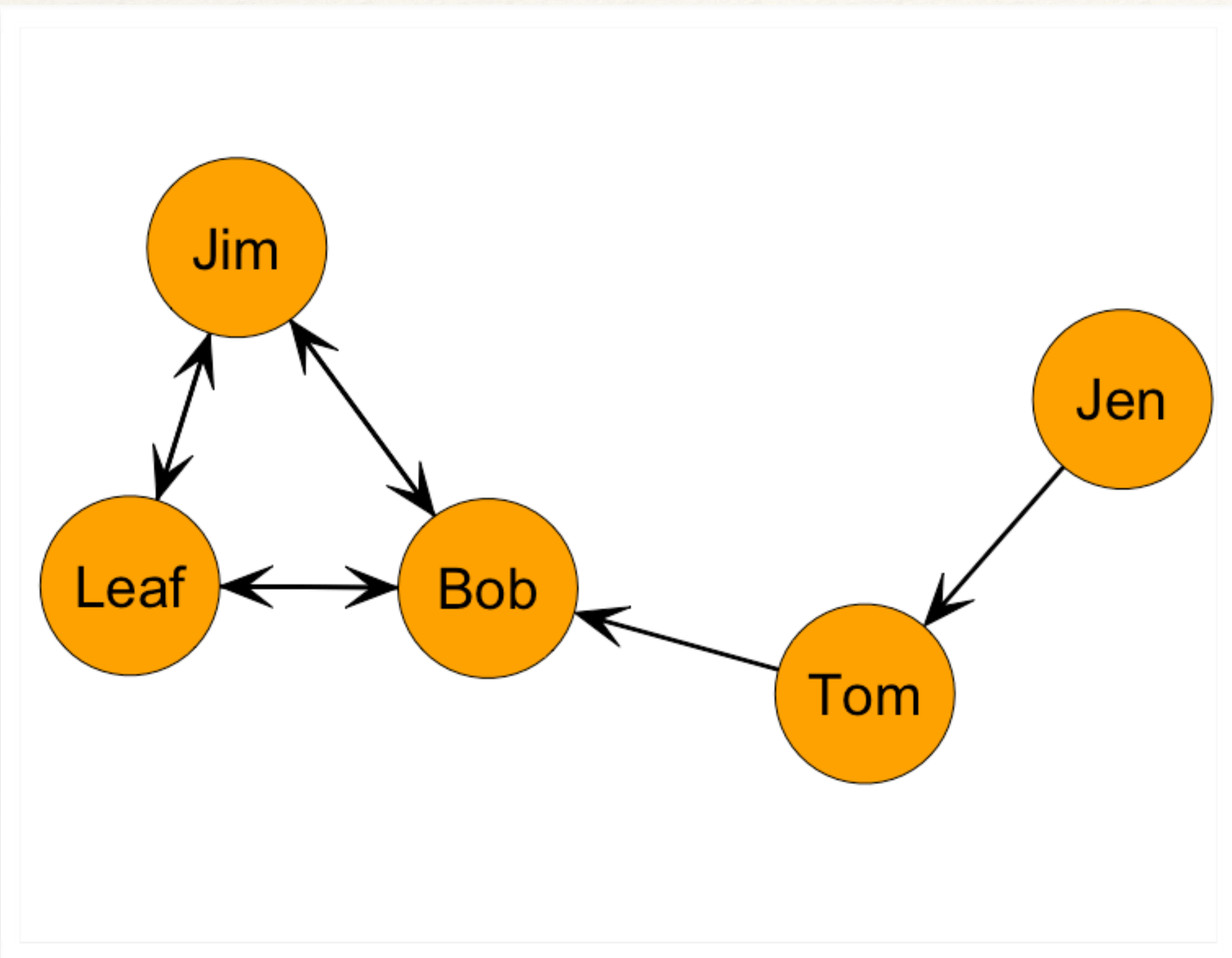


It looks like this.

Let's add zeros to the diagonals. (will explain this later…)

	Jen	Tom	Bob	Leaf	Jim
Jen	0	1	0	0	0
Tom	0	0	1	0	0
Bob	0	0	0	1	1
Leaf	0	0	1	0	1
Jim	0	0	1	1	0

Example: Directed, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen	0	1	0	0	0
Tom	0	0	1	0	0
Bob	0	0	0	1	1
Leaf	0	0	1	0	1
Jim	0	0	1	1	0

Note that, because we are allowing directionality to matter, the total number of edges in the network is just the **sum** of the entire matrix.

So what?

- The matrix is our “data” to analyze.

Empirical Example

Journal of Contemporary Criminal Justice
2015, Vol. 31(3) 243–261

Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1043986214553380
ccj.sagepub.com
SAGE

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

Jacob T. N. Young[1] and Justin T. Ready[1]

- https://journals.sagepub.com/doi/10.1177/1043986214553380

Empirical Example

Journal of Contemporary Criminal Justice
2015, Vol. 31(3) 243–261

Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1043986214553380
ccj.sagepub.com
SAGE

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

Jacob T. N. Young[1] and Justin T. Ready[1]

- Questions:
  - How do police officers "frame" body-worn cameras?
  - Is the meaning officers attribute to cameras created and transmitted in groups?

Empirical Example

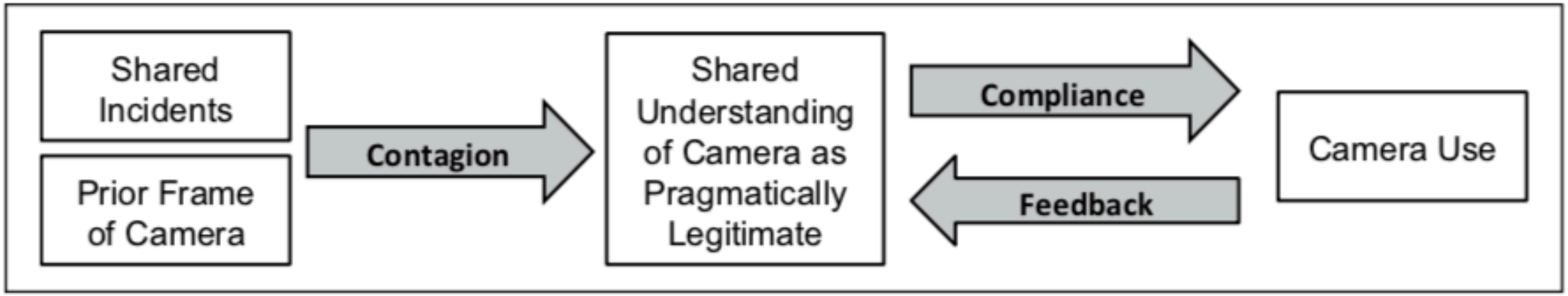
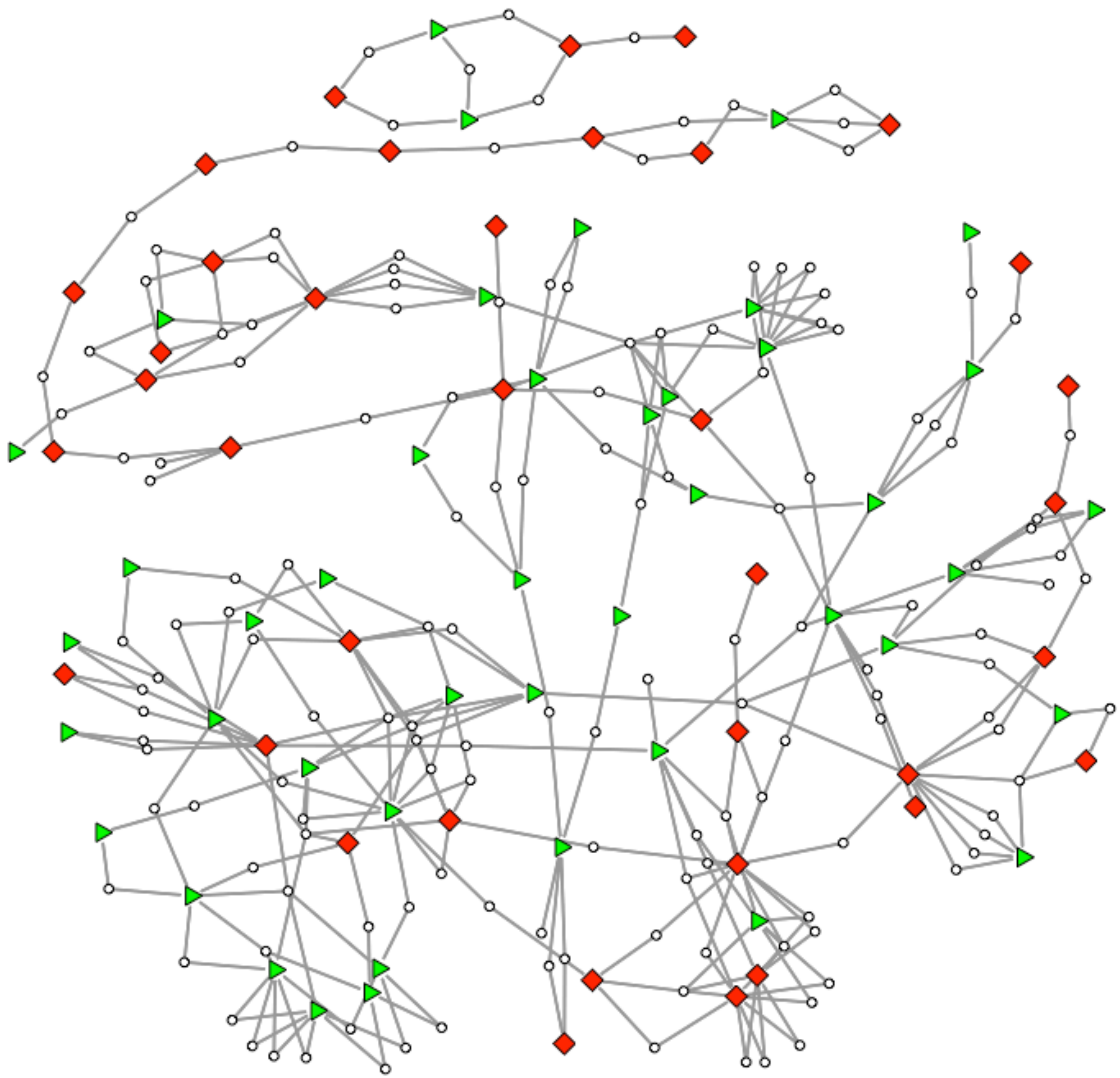


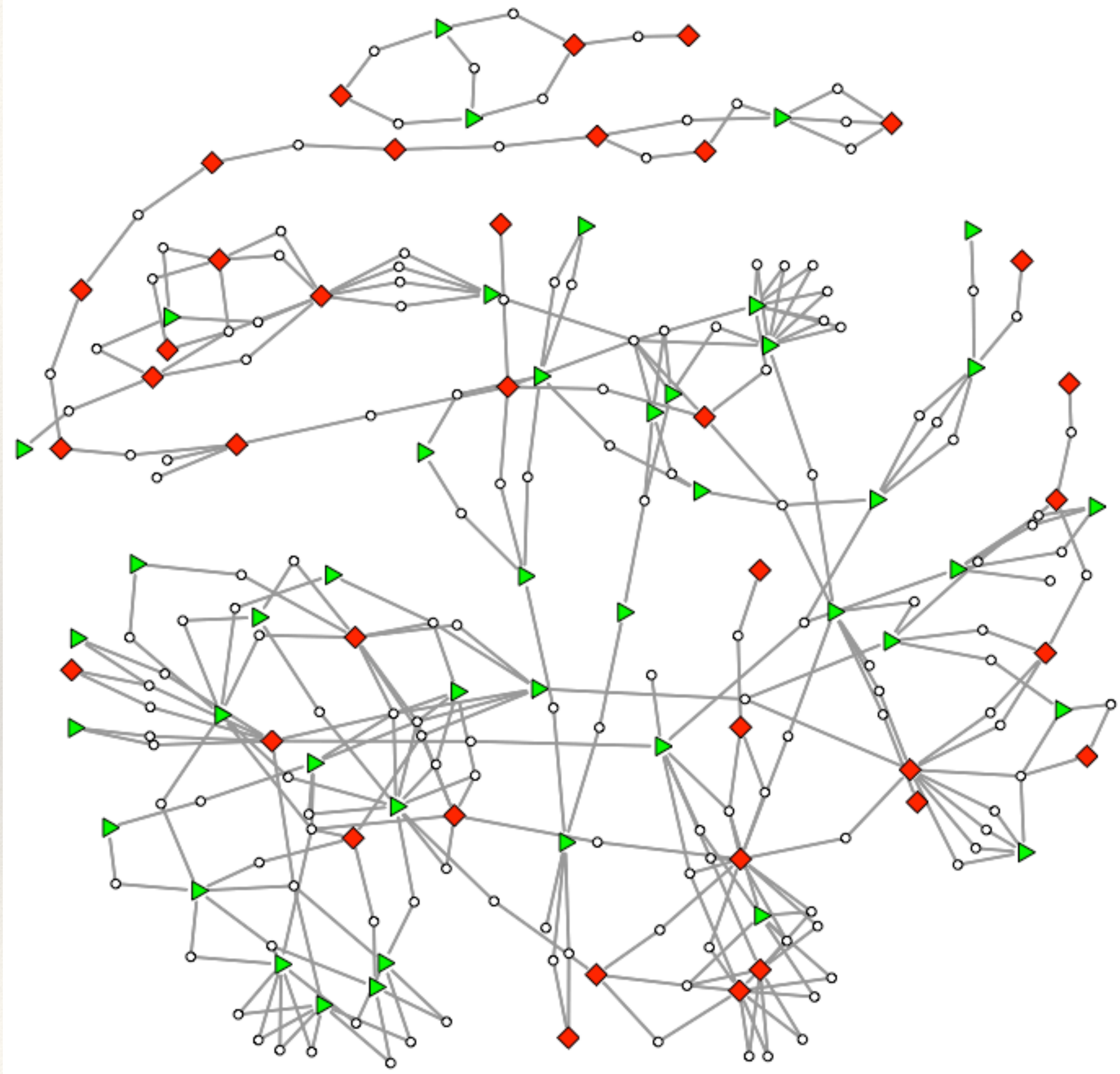
Figure 1. Diffusion of pragmatic legitimacy frame and compliance.

Bipartite Graph of Incidents and Officers
by Treatment or Control Condition



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

What do you see in this network?



What do the connections represent in this network?

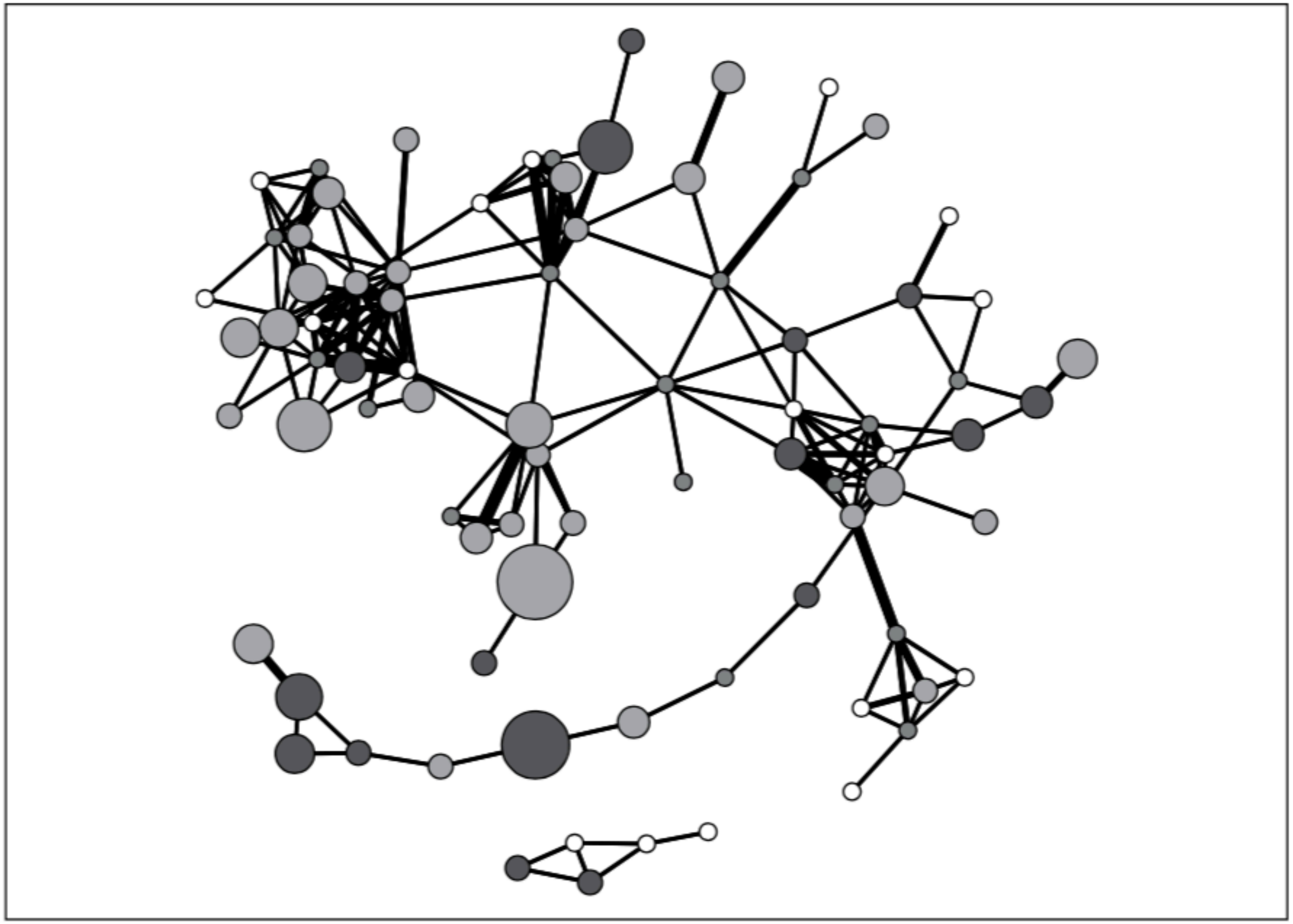


Figure 5. One-mode network of officers.
Note. Node size is proportional to change in legitimacy: darker = more negative; white = no change. Line size is proportional to number of shared incidents.

Are officers' views of body-worn cams influenced by the views of those whom they share events with?

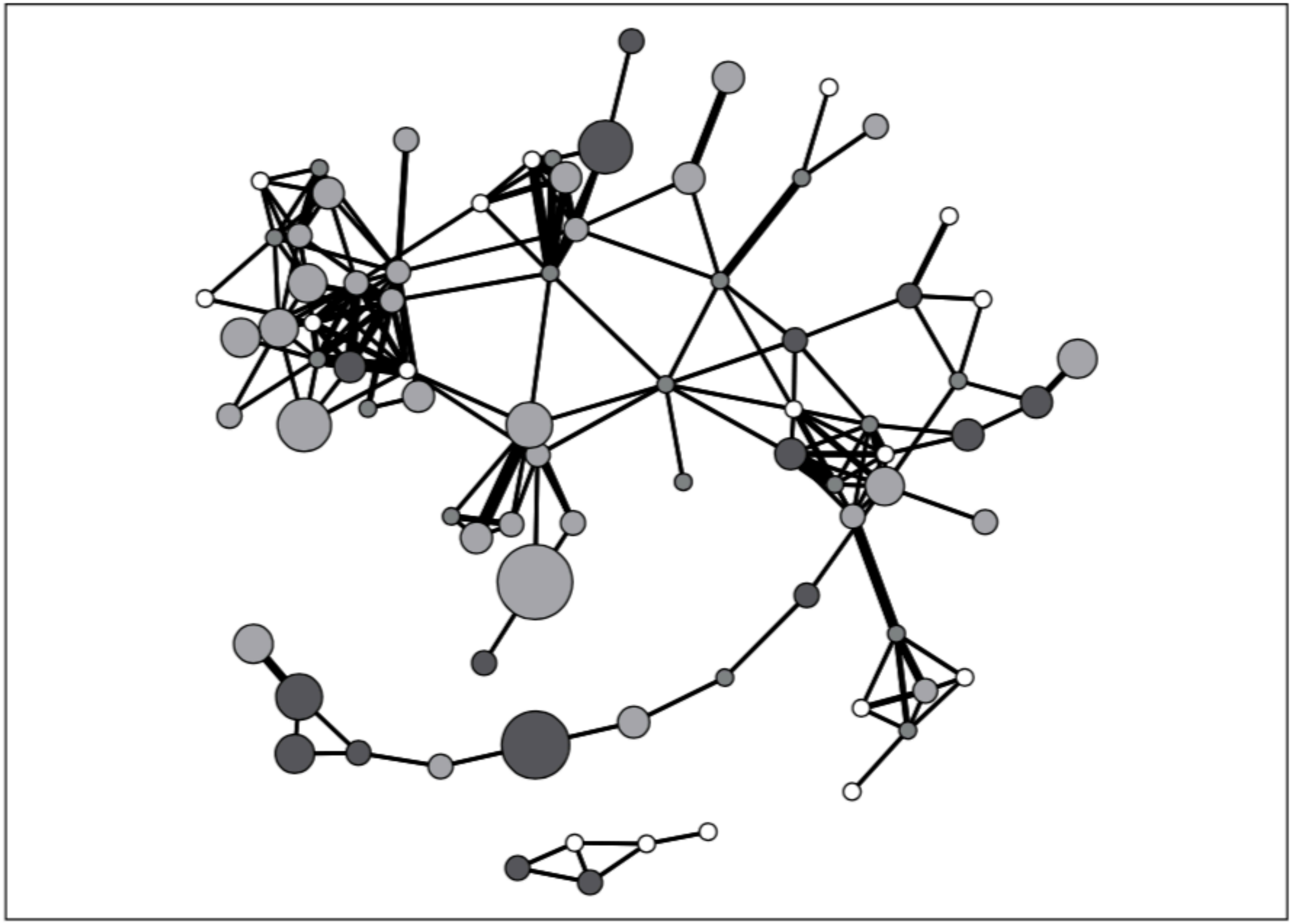


Figure 5. One-mode network of officers.

Note. Node size is proportional to change in legitimacy: darker = more negative; white = no change. Line size is proportional to number of shared incidents.

Questions?

Basic Network Analysis

How do we analyze networks?

Description vs. Inference

- Description
  - 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 influence attributes of actors? (Co-evolution models)

Degree Centrality

Learning Goals

- Understand the conceptualization of “centrality”.
- Understand calculation of degree centrality.
- Analyze descriptive features of degree centrality.

When we say a *node* is "central," what do we mean conceptually?

Concepts and Operationalization

- Speaking generally, network *position* should be interesting and important:
  - As a <u>dependent</u> variable (e.g. are taller individuals more likely to be trusted?)
  - As an <u>independent</u> variable (e.g. are more popular adolescents more likely to succeed in school?)
    - As a description of the position of a **node/vertex**
    - *And* as a description of an entire **network**

Conceptualization

- "Everyone agrees, it seems, that centrality is an important structural attribute of networks. All concede that it is related to a high degree to other important group properties and processes. But there consensus ends." (Freeman, 1978/1979: 217)
    - The type of measure we use depends on the substantive question of interest.
        - Various measures of centrality are correlated, but they <u>operationalize</u> different concepts.

Undirected Networks

Conceptualization

- Concepts and unit of analysis:
  - Point centrality (degree, betweenness, closeness)
  - Graph centrality (compactness)

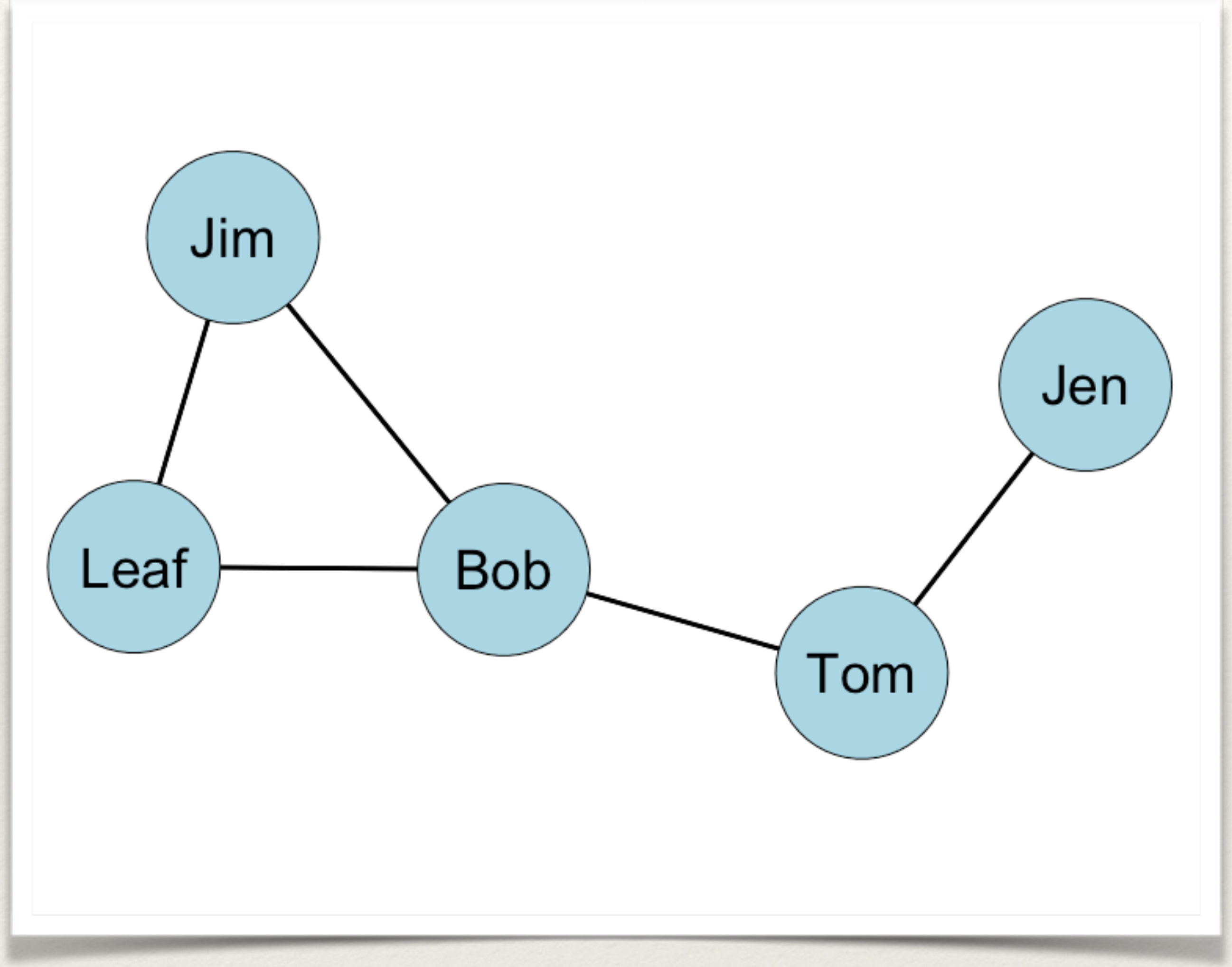
Degree Centrality: Undirected Binary Graphs

- In an undirected binary graph, *actor degree centrality* measures the extent to which a node connects to all other nodes in a social network.
  - In other words, the number of edges incident with a node.
    - This is symbolized as: $d(n_i)$
      - For an undirected binary graph, the degree $d(n_i)$ is the row or column sum.

Degree Centrality: Undirected Binary Graphs

$$C_D(n_i) = d(n_i) = \sum_j x_{ij} = \sum_j x_{ji}$$

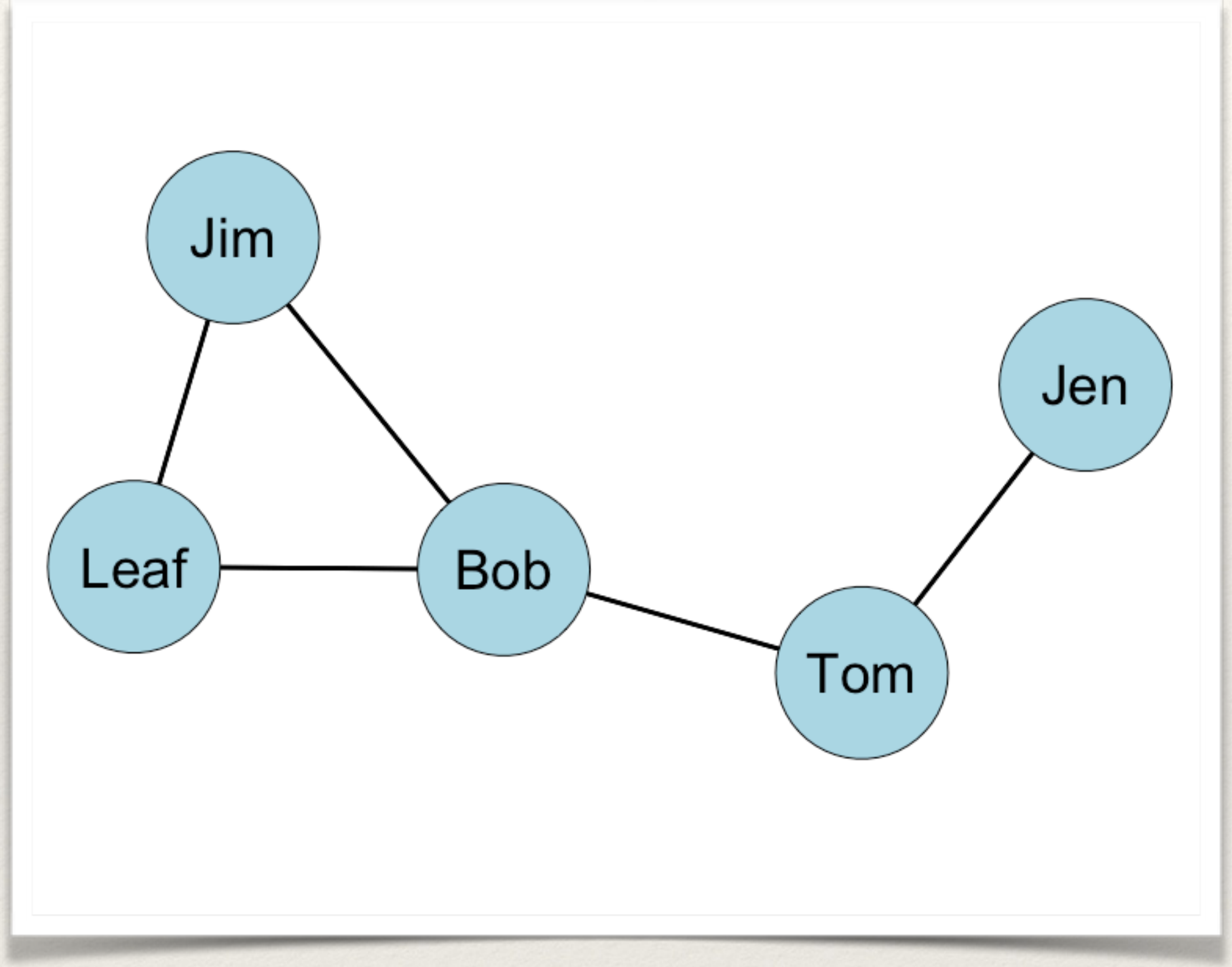
Example: Undirected, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen	0	1	0	0	0
Tom	1	0	1	0	0
Bob	0	1	0	1	1
Leaf	0	0	1	0	1
Jim	0	0	1	1	0

What is the degree for each node in this graph?

Example: Undirected, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen	0	1	0	0	0
Tom	1	0	1	0	0
Bob	0	1	0	1	1
Leaf	0	0	1	0	1
Jim	0	0	1	1	0

Note that the column sum and row sum are the same.

Degree Centrality: Undirected Binary Graphs

- Actor degree centrality not only reflects each node's connectivity to other nodes but also depends on the size of the network, g.
  - Larger networks will have a higher maximum possible degree centrality value.
    - *Solution?*

Standardized Degree Centrality: Undirected Binary Graphs

- Standardize!
  - Take into account the number of nodes and the maximum possible nodes to which *i* could be connected.
    - That is, $g-1$.

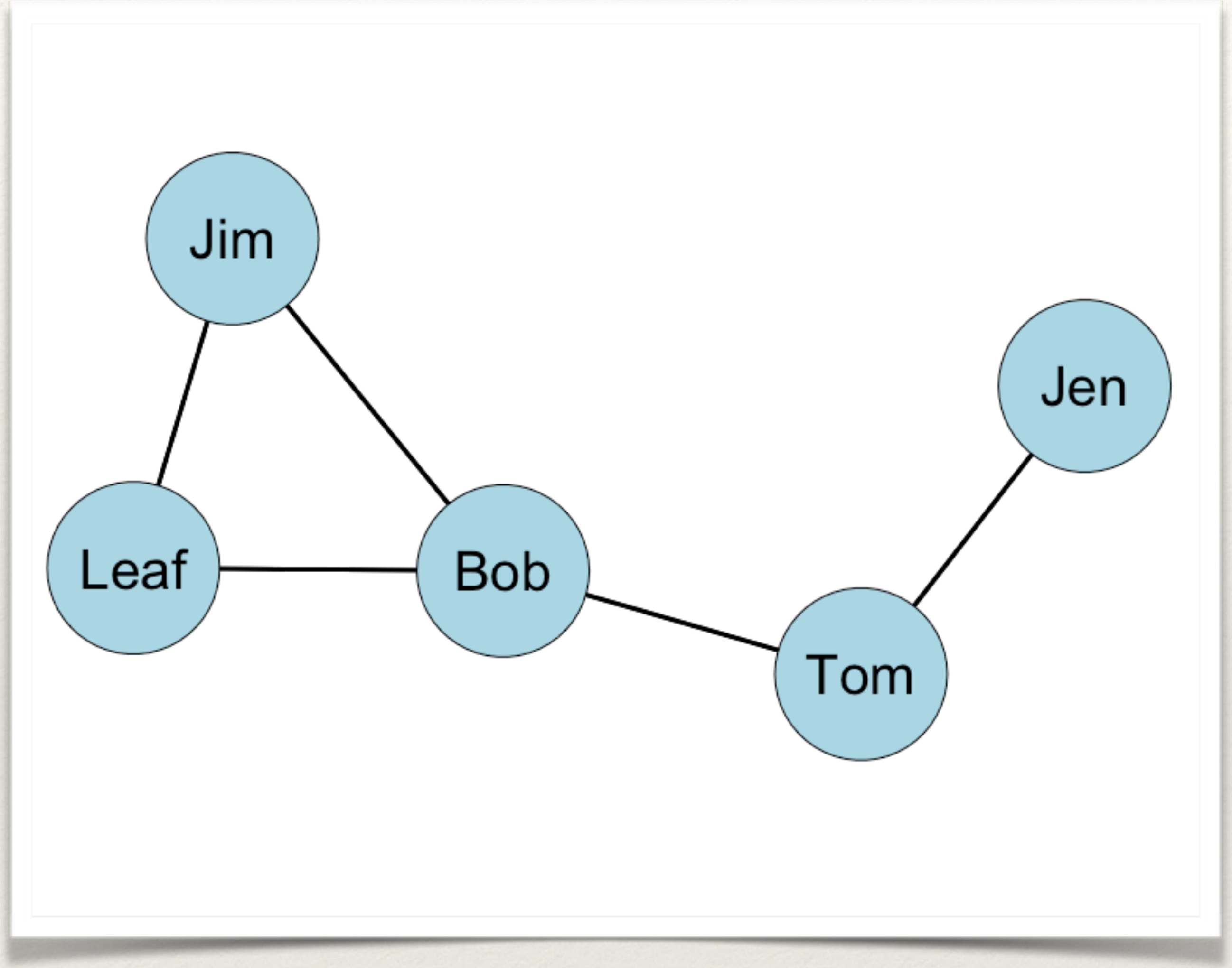
Standardized Degree Centrality: Undirected Binary Graphs

$$C'_D(n_i) = \frac{d(n_i)}{g-1} = \frac{\sum_j x_{ij}}{g-1}$$

Standardized Degree Centrality: Undirected Binary Graphs

- This yields the proportion of the network members with ties to actor *i*.
  - This varies between 0 (no connections; isolate) to 1 (ties to every actor).

Example: Undirected, Binary Network



Raw Degree Centrality

Jen = 1

Tom = 2

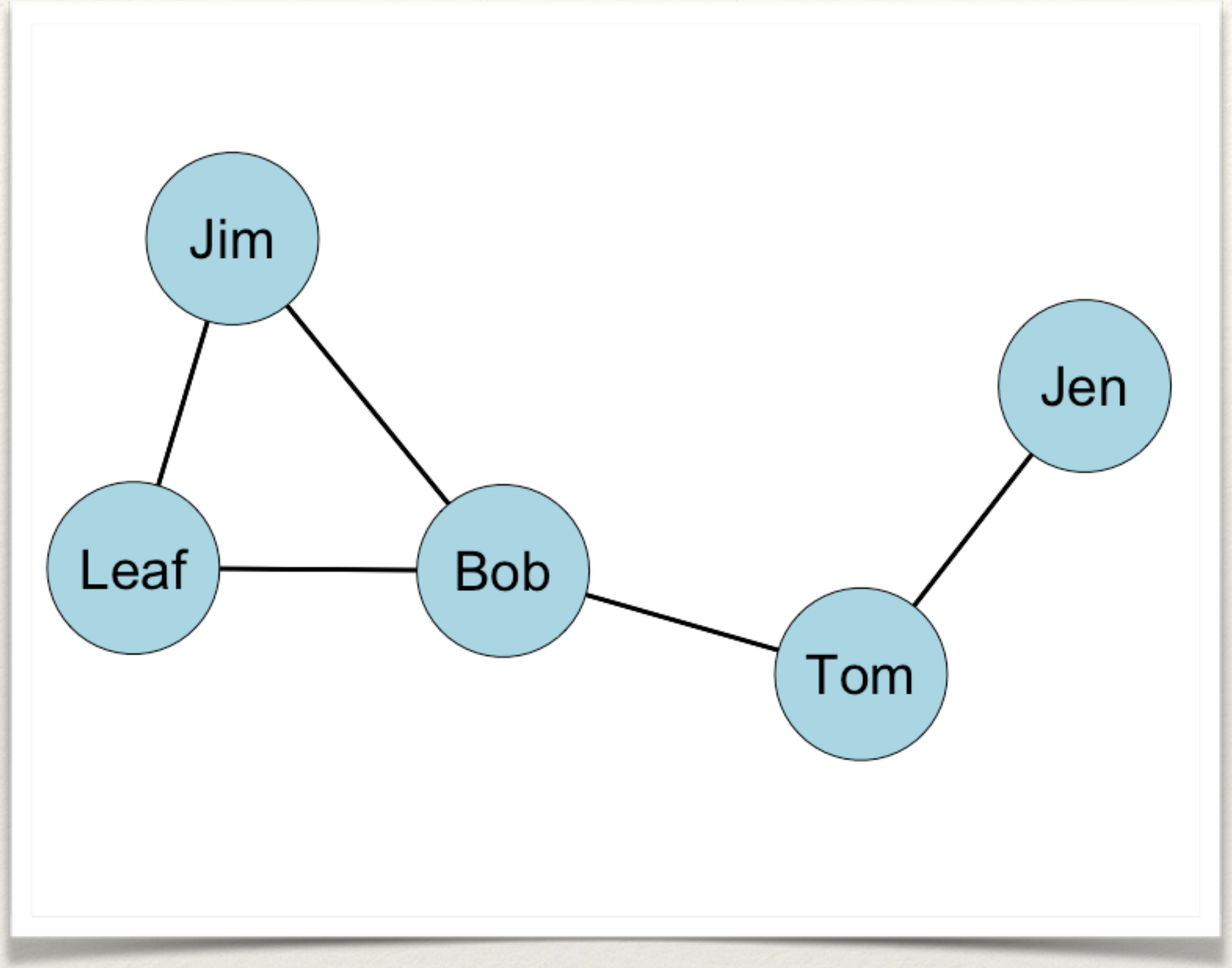
Bob = 3

Leaf = 2

Jim = 2

What is the standardized degree centrality score for each node?

Example: Undirected, Binary Network



Standardized Degree Centrality

Jen = 1/4 = 0.25

Tom = 2/4 = 0.50

Bob = 3/4 = 0.75

Leaf = 2/4 = 0.50

Jim = 2/4 = 0.50

Summarizing Degree Centrality

- We can examine the summary statistics for degree centrality by inspecting the **mean**.
  - The average degree is an important property of a network.
    - *Why? What does a network with a high average degree look like? A low average degree?*

Mean Degree (undirected)

Sum up the degrees for each actor

$$\bar{d} = \frac{\sum_{i=1}^{g} d(n_i)}{g} = \frac{2L}{g}$$

Divide by number of actors

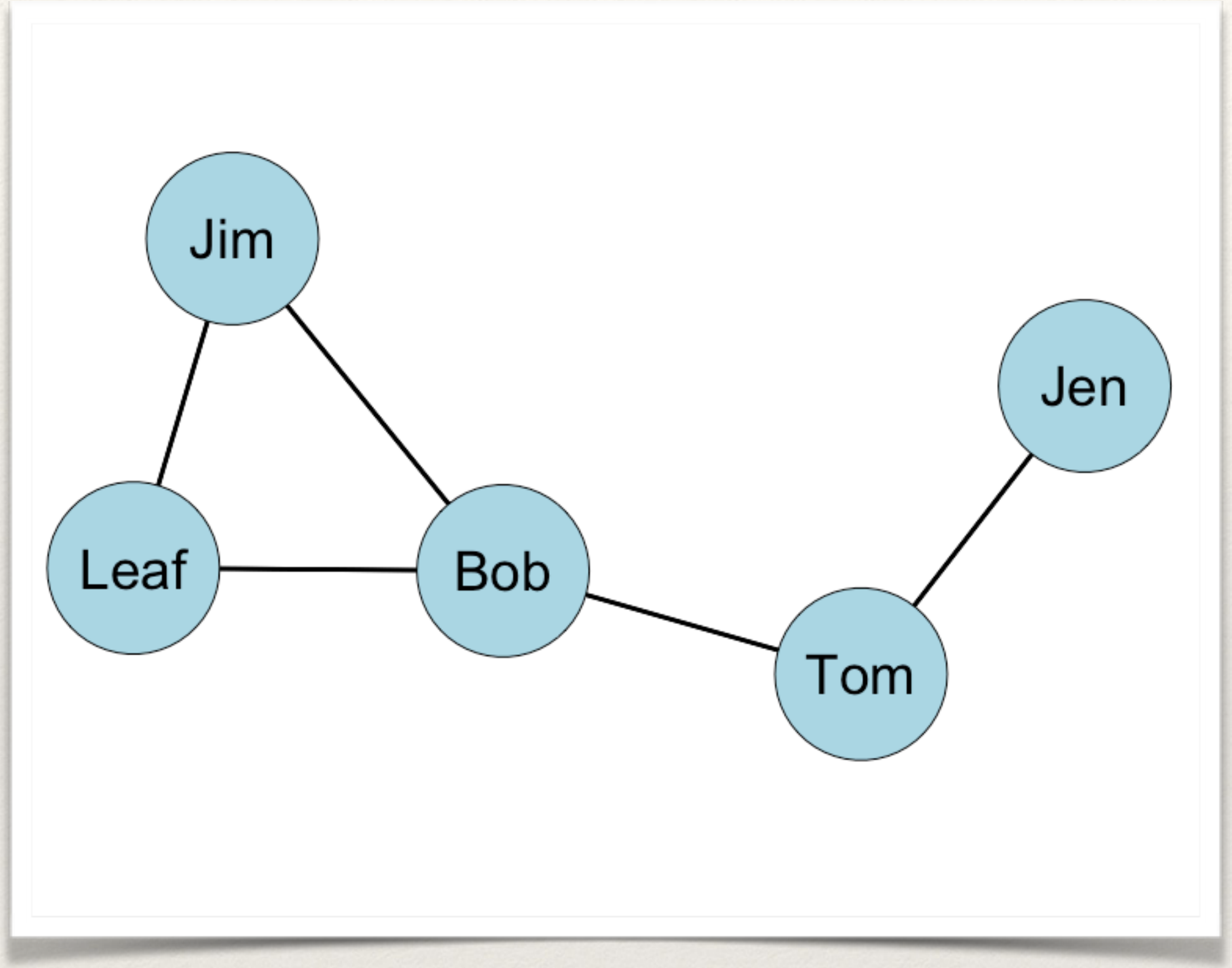
Mean Degree (undirected)

Or, multiply the number of edges by 2.

$$\bar{d} = \frac{\sum_{i=1}^{g} d(n_i)}{g} = \frac{2L}{g}$$

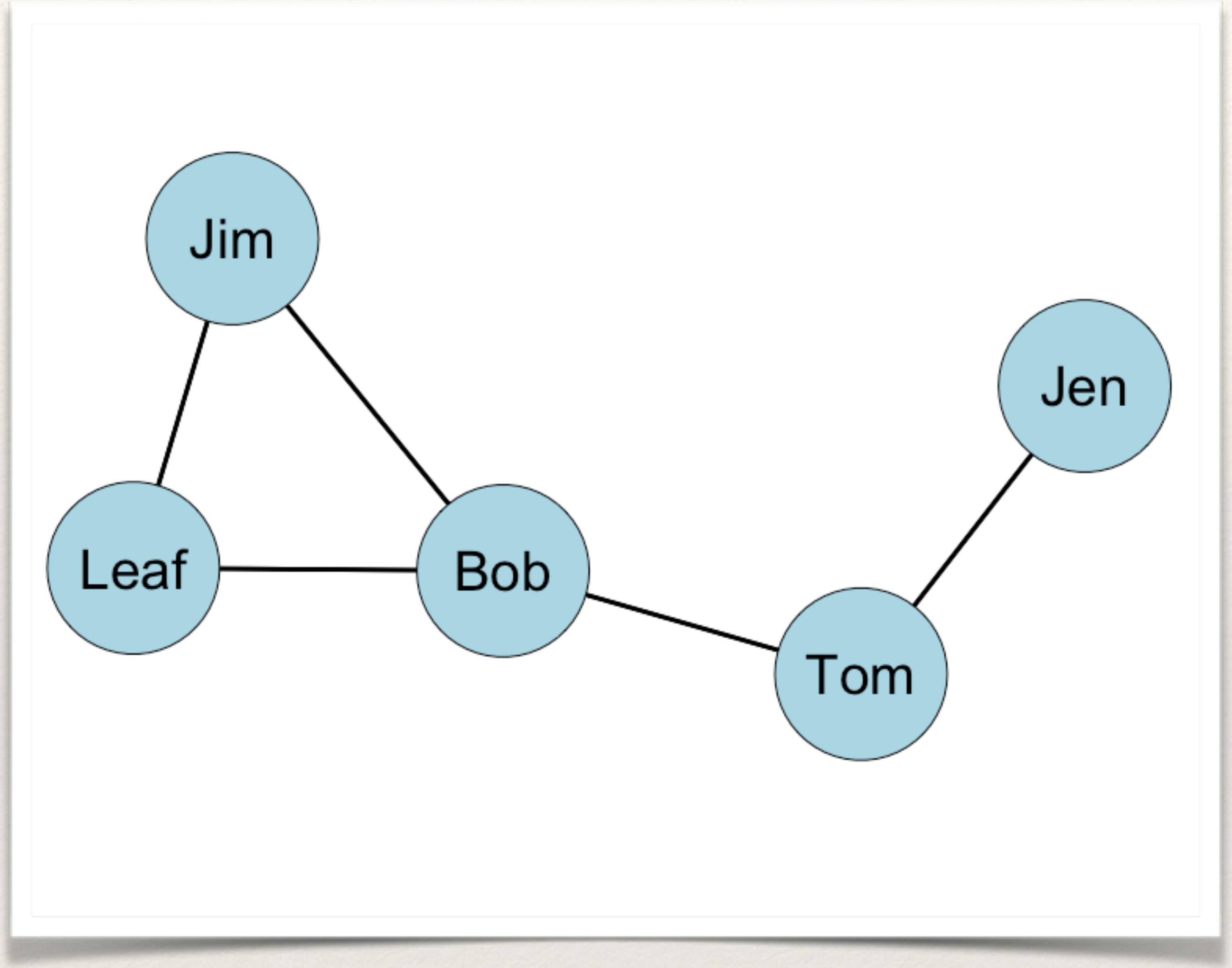
Divide by number of actors

Example: Undirected, Binary Network



What is the mean degree for this graph?

Example: Undirected, Binary Network



$$\bar{d} = \frac{\sum_{i=1}^{g} d(n_i)}{g} = \frac{2L}{g} = \frac{2 * 5}{5} = \frac{10}{5} = 2$$

What is the mean degree for this graph?

Summarizing Degree Centrality

- We can also calculate how centralized the graph itself is.
  - *Group degree centralization* measures the extent to which the actors in a social network differ from one another in their individual degree centralities.

Group Degree Centralization

Largest actor degree centrality scored observed

Degree centrality for actor *i*

$$C_A = \frac{\sum_{i=1}^{g} [C_A(n^*) - C_A(n_i)]}{max \sum_{i=1}^{g} [C_A(n^*) - C_A(n_i)]}$$

Group Degree Centralization

$$C_A = \frac{\sum_{i=1}^{g}[C_A(n^*) - C_A(n_i)]}{max \sum_{i=1}^{g}[C_A(n^*) - C_A(n_i)]}$$

Sum of observed differences between the largest actor centrality and all others

Theoretical maximum possible sum of those differences

Summarizing Degree Centrality

- Note that this is a generic measure (thanks Freeman, 1979!)
    - We can calculate the denominator as $(g-1)(g-2)$ (Thanks Wasserman & Faust, 1994!)

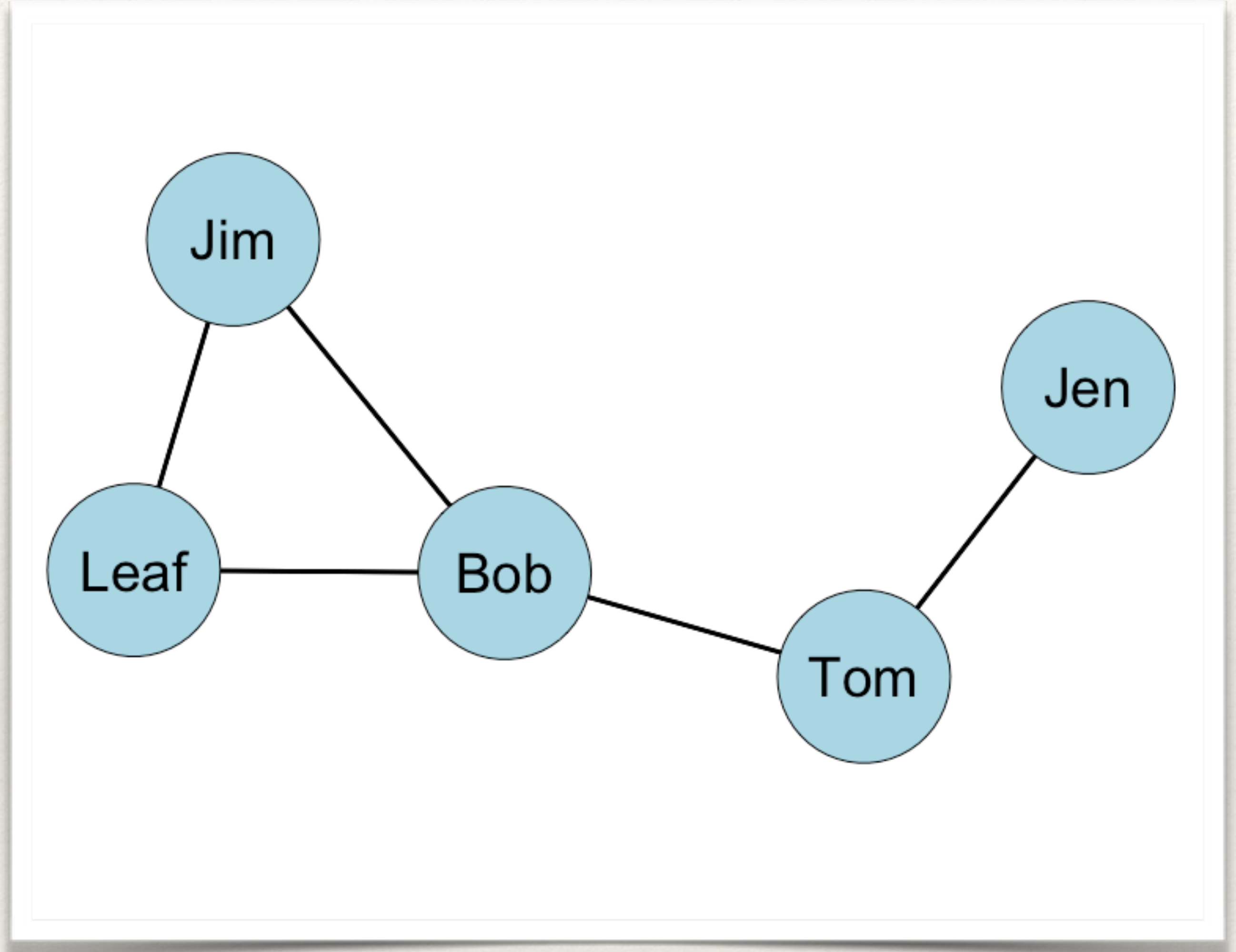
Index of Group Degree Centralization

$$C_D = \frac{\sum_{i=1}^{g} [C_D(n^*) - C_D(n_i)]}{[(g-1)(g-2)]}$$

Sum of observed differences between the largest actor centrality and all others

The maximum possible sum of differences

Example: Undirected, Binary Network



Raw Degree Centrality

Jen = 1

Tom = 2

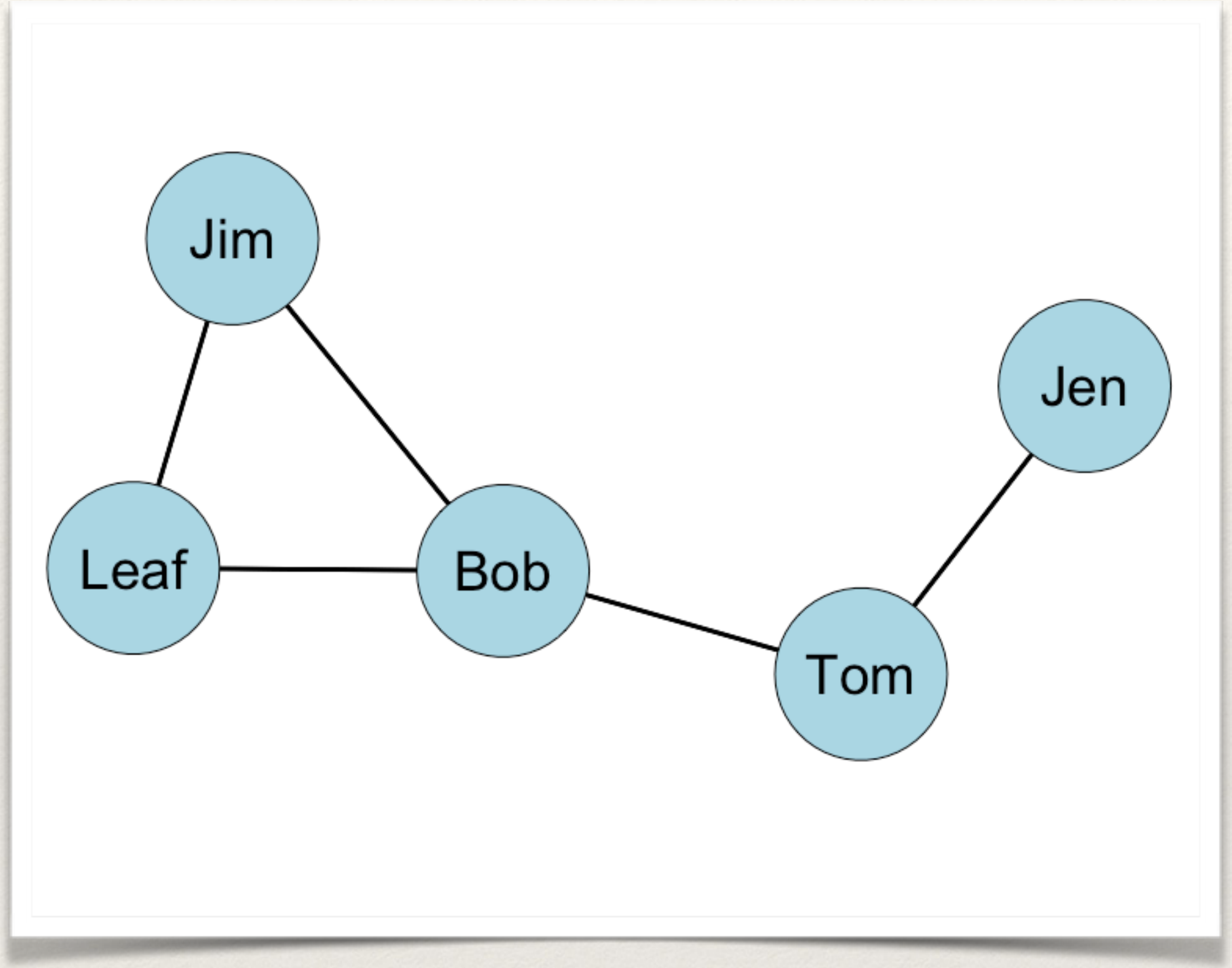
Bob = 3

Leaf = 2

Jim = 2

What is the index of degree centralization for this graph?

Example: Undirected, Binary Network



0.4167

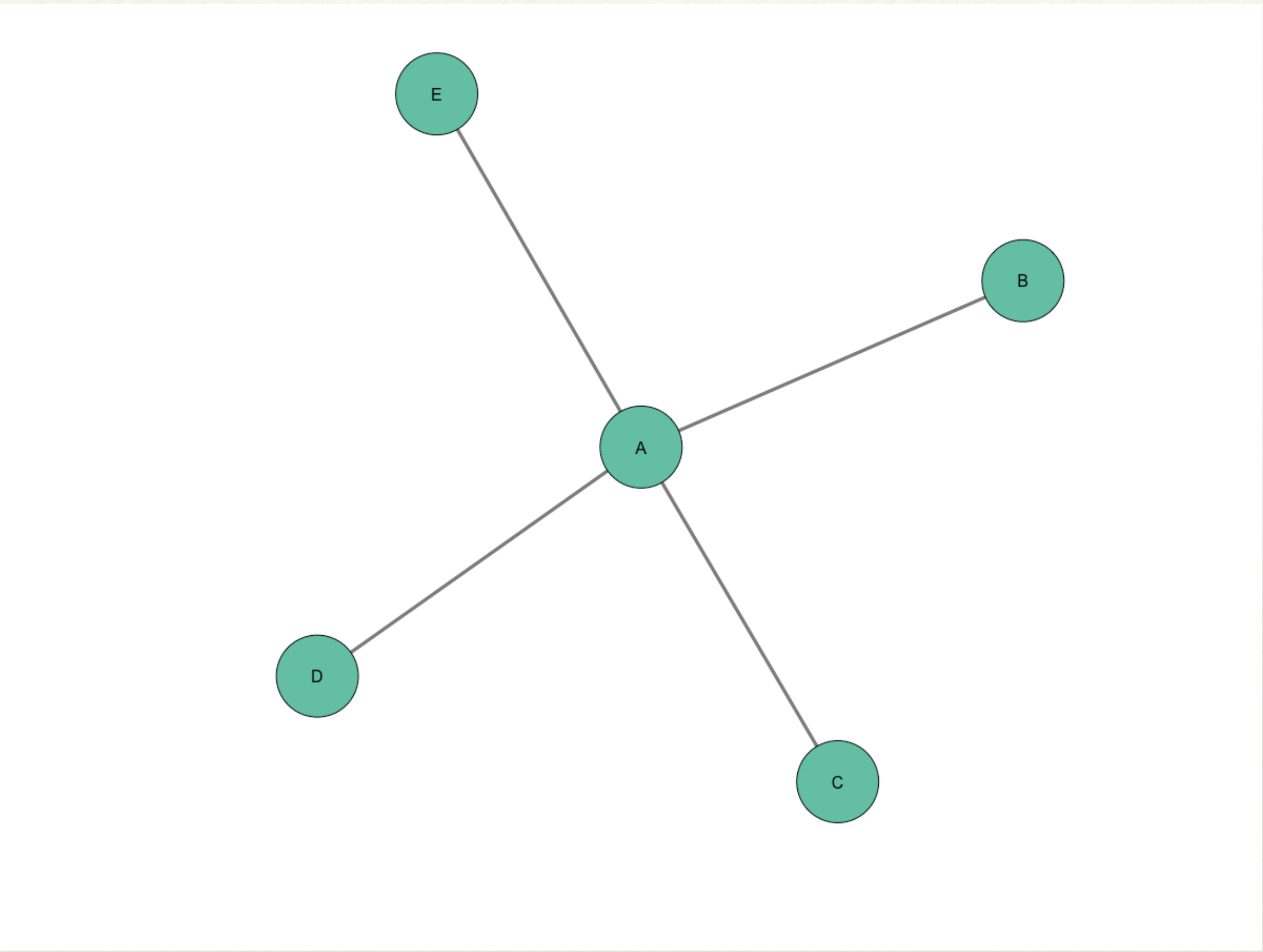
$$C_D = \frac{\sum_{i=1}^{g} [C_D(n^*) - C_D(n_i)]}{[(g-1)(g-2)]}$$

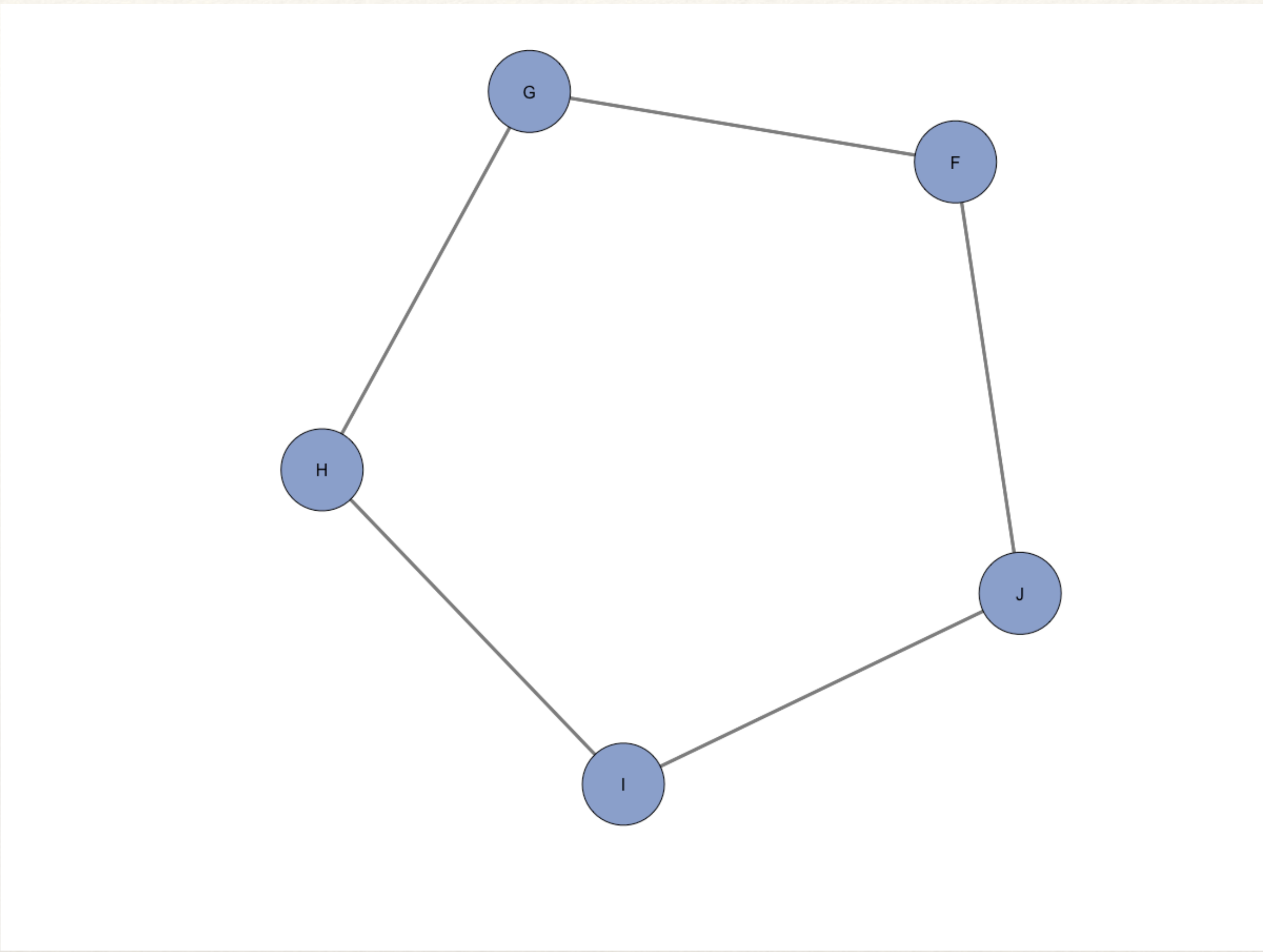
$$= \frac{(3-1)+(3-2)+(3-3)+(3-2)+(3-2)}{(5-1)(5-2)}$$

$$= \frac{2+1+0+1+1}{4*3} = \frac{5}{12} = 0.4167$$

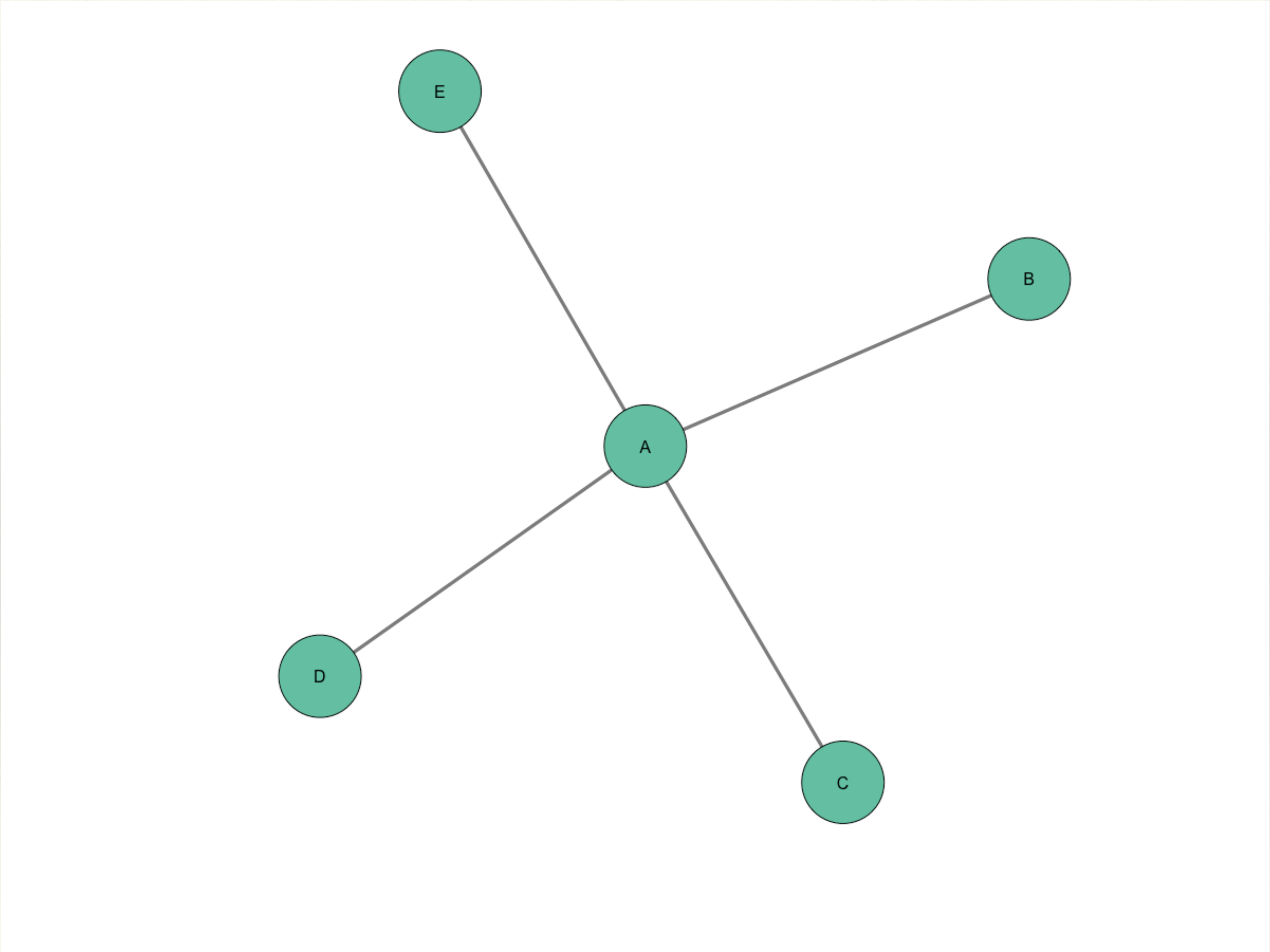
Summarizing Degree Centrality

- When degree centrality is evenly dispersed, the numerator will be zero, and the quotient will be close to 0.
    - When there is considerable inequality in the actor degrees, the quotient will be closer to 1.
        - Thus, closer to 1 indicates that the graph is hierarchically structured.

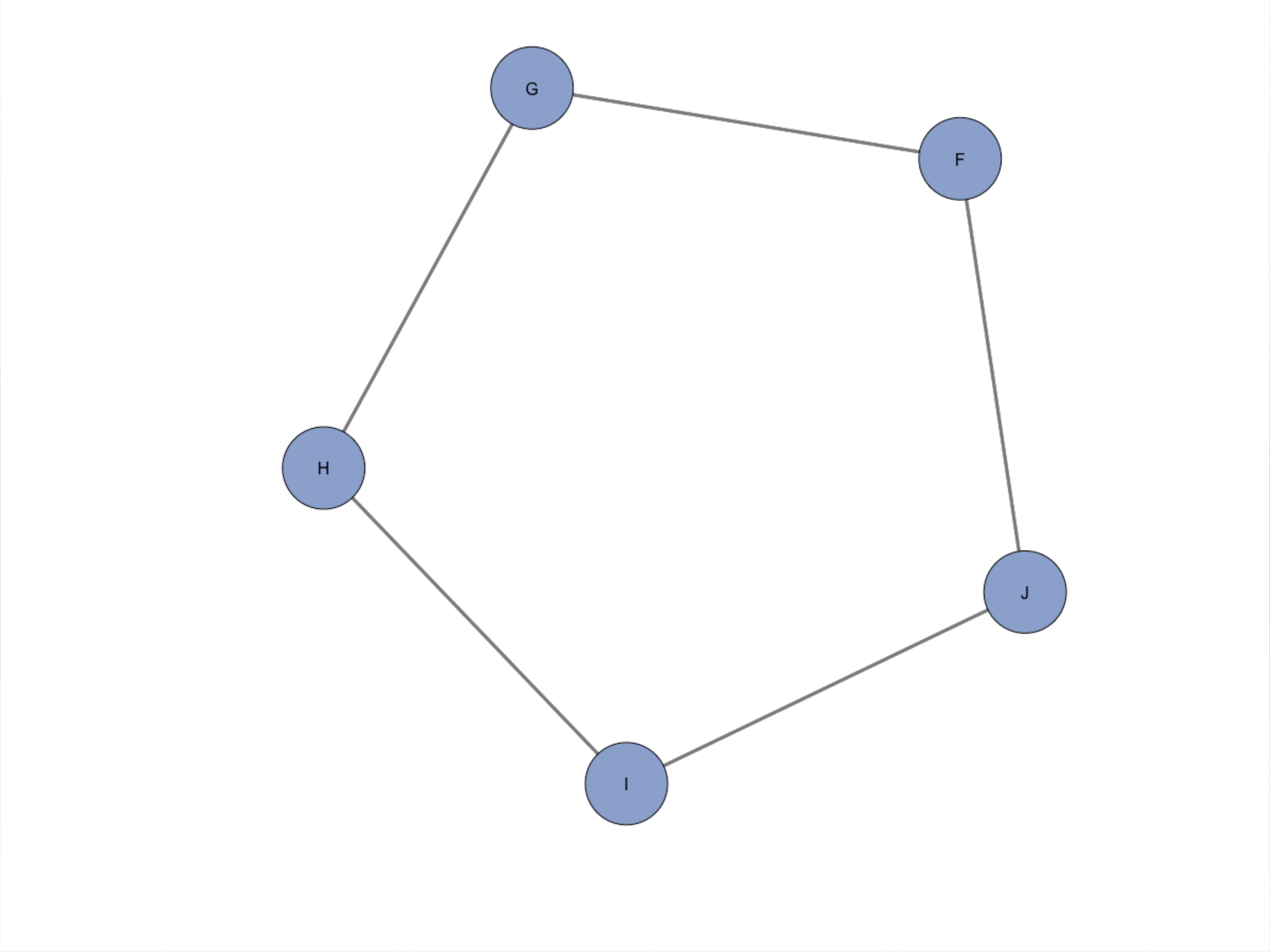

E
B
A
D
C


G
F
H
J
I

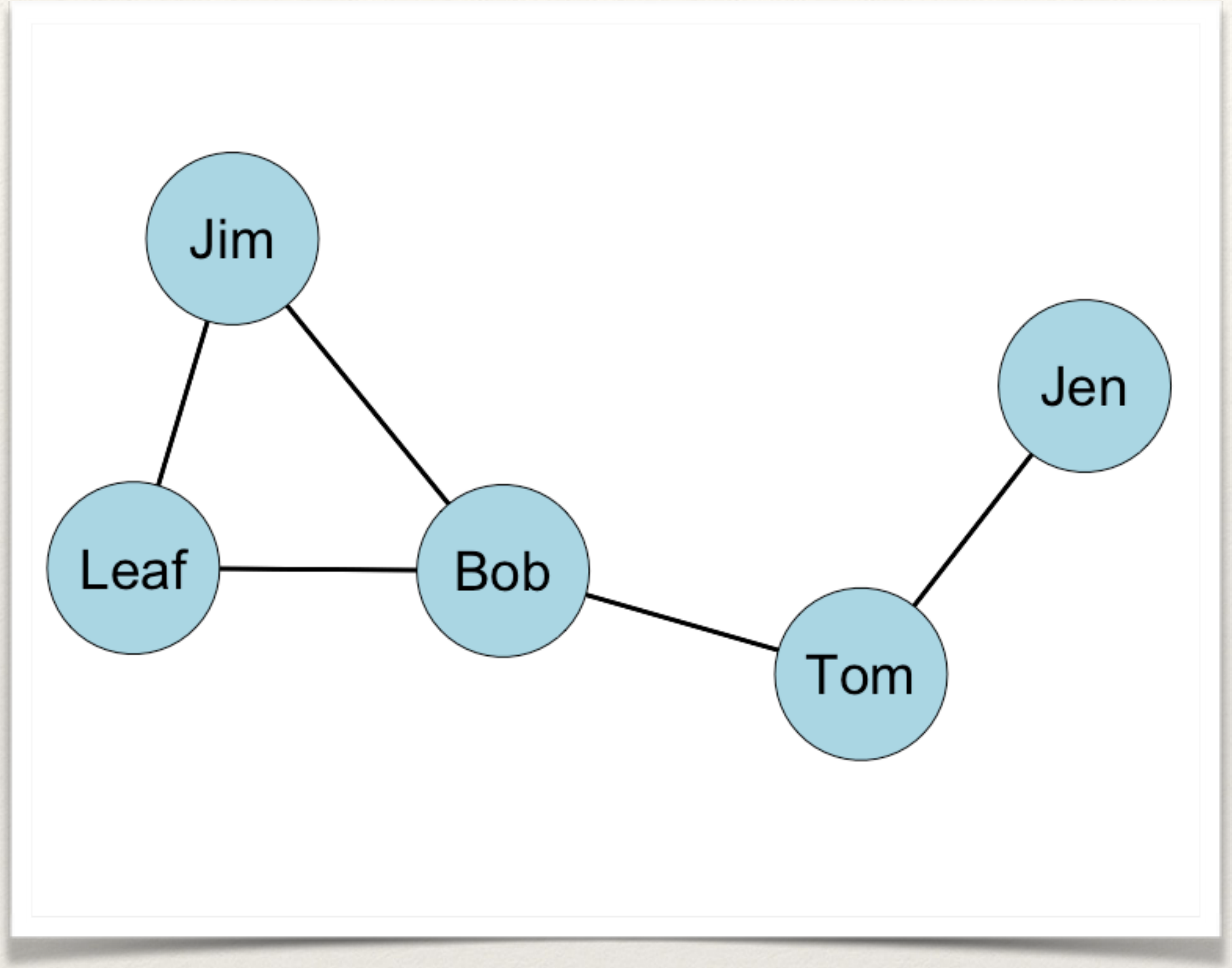
$$C_D = \frac{\sum_{i=1}^{g}[C_D(n^*) - C_D(n_i)]}{[(g-1)(g-2)]} = \frac{(4-4)+(4-1)+(4-1)+(4-1)+(4-1)}{(5-1)(5-2)} = \frac{0+3+3+3+3}{4*3} = \frac{12}{12} = 1.0$$



$$C_D = \frac{\sum_{i=1}^{g}[C_D(n^*) - C_D(n_i)]}{[(g-1)(g-2)]} = \frac{(2-2)+(2-2)+(2-2)+(2-2)+(2-2)}{(5-1)(5-2)} = \frac{0+0+0+0+0}{4*3} = \frac{0}{12} = 0.0$$



Example: Undirected, Binary Network



How should we interpret this value?

0.4167

Directed Networks

Degree Centrality: Directed Binary Graphs

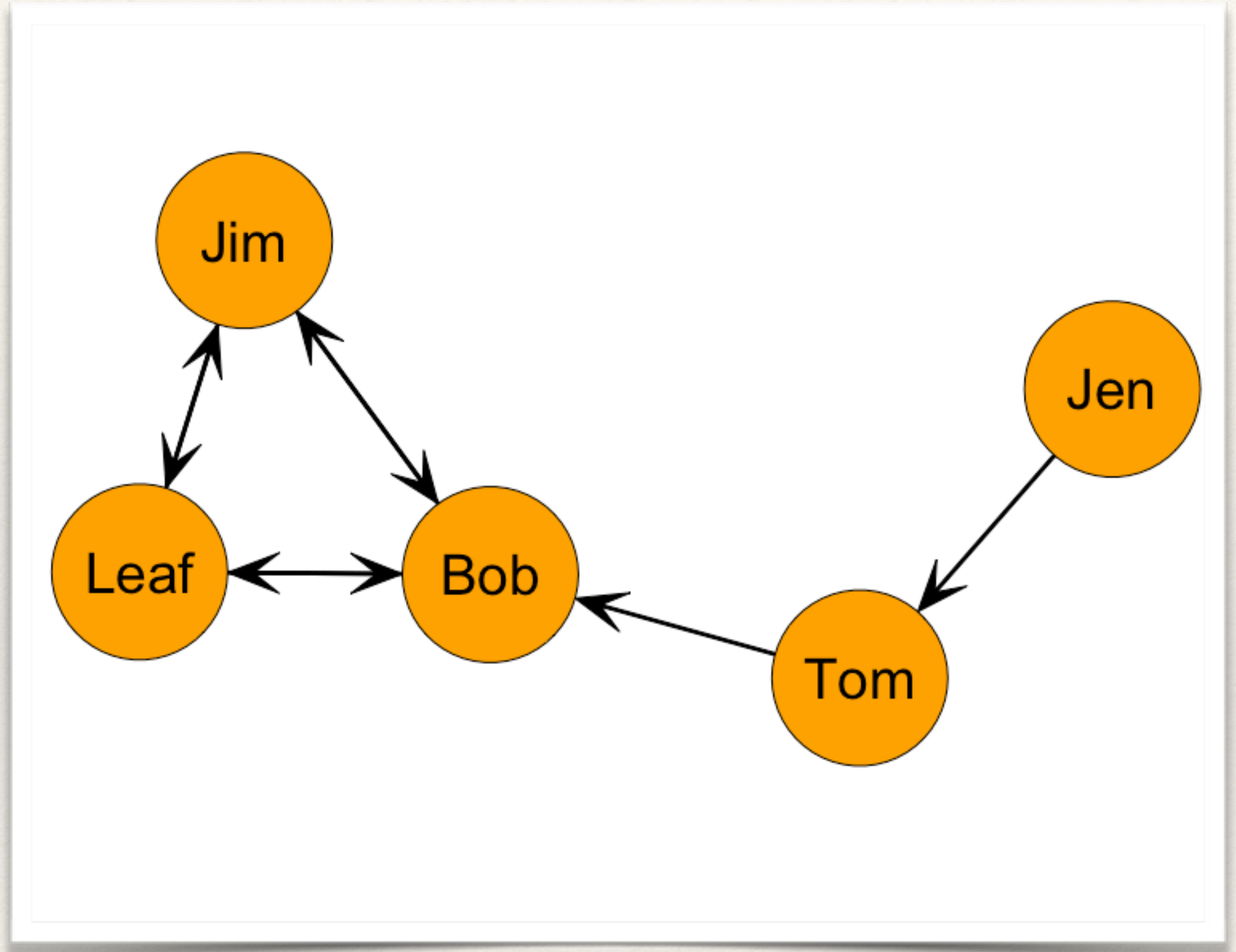
- In a directed binary graph, *actor degree centrality* can be broken down into <u>indegree</u> and <u>oudegree</u> centrality.
  - **Indegree**, $C_I(n_i)$, measures the number of ties that i receives.
    - For the sociomatrix $\boldsymbol{X}_{ij}$, the indegree for i is the column sum.
  - **Outdegree**, $C_O(n_i)$, measures the number of ties that i sends.
    - For the sociomatrix $\boldsymbol{X}_{ij}$, the outdegree for i is the row sum.

Degree Centrality: Directed Binary Graphs

$$C_I(n_i) = \sum_j x_{ji}$$

$$C_O(n_i) = \sum_j x_{ij}$$

Example: Directed, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen	0	1	0	0	0
Tom	0	0	1	0	0
Bob	0	0	0	1	1
Leaf	0	0	1	0	1
Jim	0	0	1	1	0

What is the indegree and outdegree for each node in the graph?

Example: Directed, Binary Network

<u>Raw Indegree Centrality</u>

Jen = 0

Tom = 1

Bob = 3

Leaf = 2

Jim = 2

TOTAL: 8

NOTE: These both sum to the same value

<u>Raw Outdegree Centrality</u>

Jen = 1

Tom = 1

Bob = 2

Leaf = 2

Jim = 2

TOTAL: 8

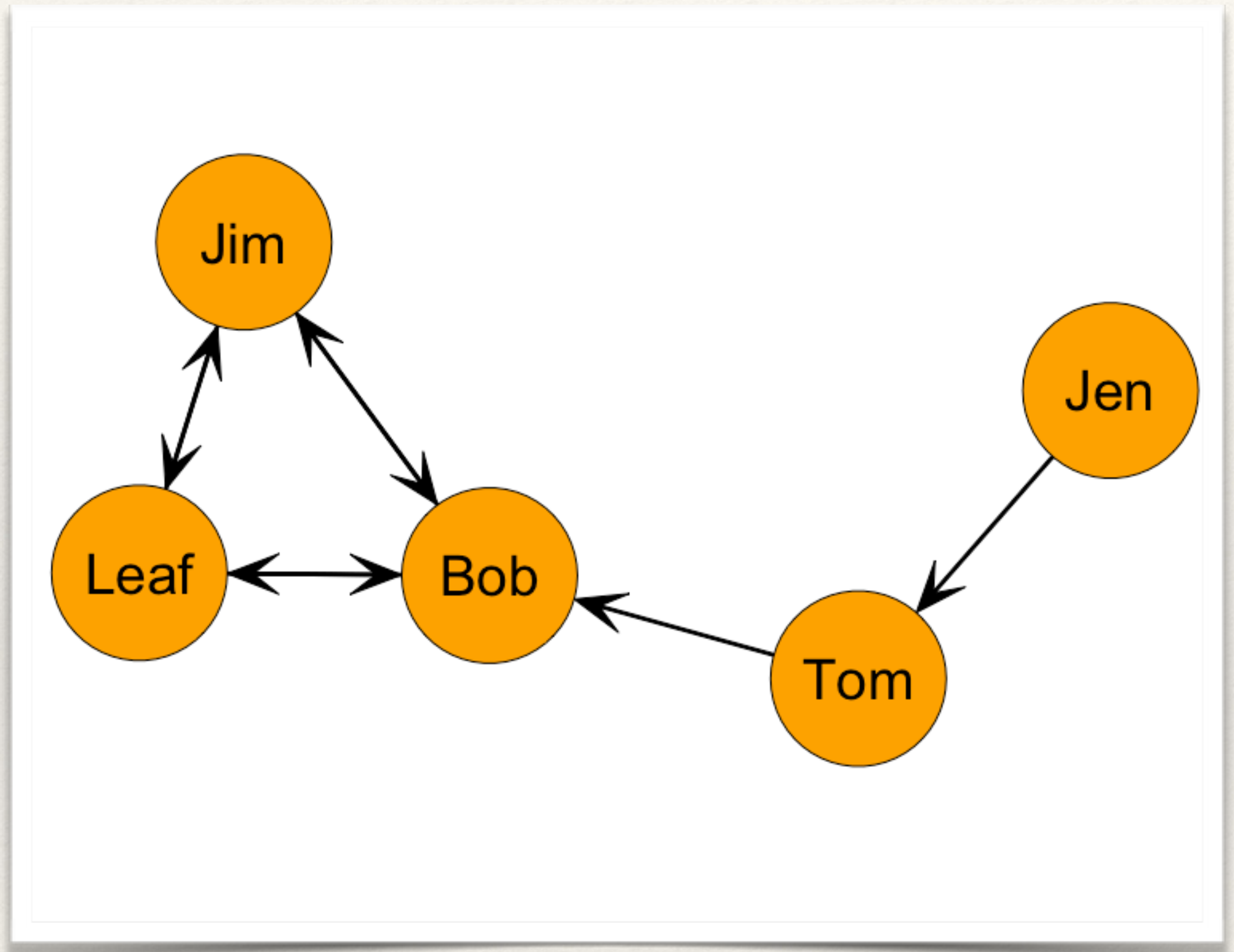
Degree Centrality: Directed Binary Graphs

- Recall that actor degree centrality not only reflects each node's connectivity to other nodes but also depends on the size of the network, *g*.
  - Larger networks will have a higher maximum possible degree centrality value.
    - We can standardize, or normalize, the same way by dividing by *g-1*.

Standardized Degree Centrality: Directed Binary Graphs

$$C'_D(n_i) = \frac{C_I(n_i)}{g-1} = \frac{C_O(n_i)}{g-1}$$

Example: Directed, Binary Network



What is the standardized indegree and outdegree centrality score for each node?

Raw Indegree Centrality

Jen = 0

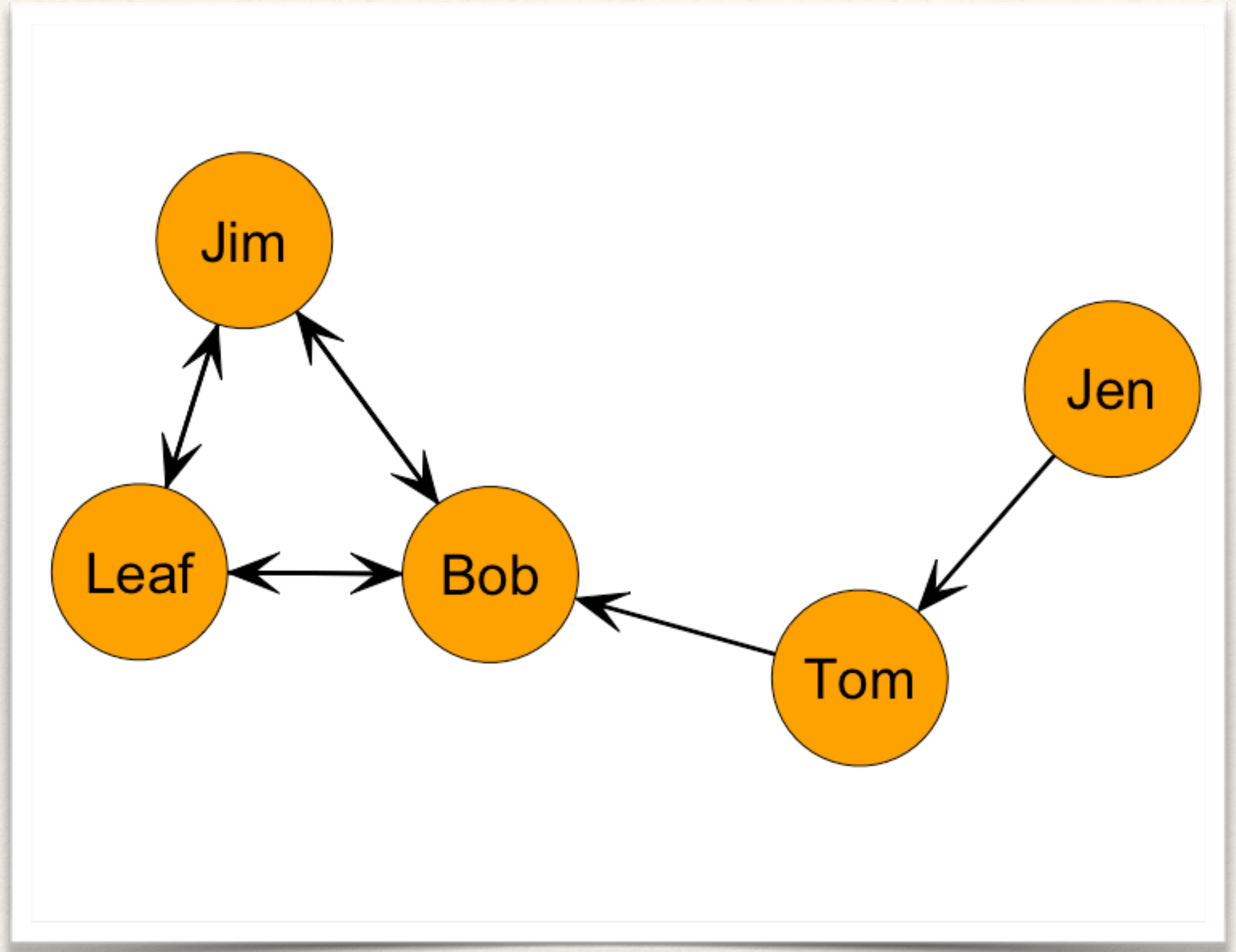
Tom = 1

Bob = 3

Leaf = 2

Jim = 2

Example: Directed, Binary Network



What is the standardized indegree and outdegree centrality score for each node?

Standardized Indegree Centrality

Jen = 0/4 = 0

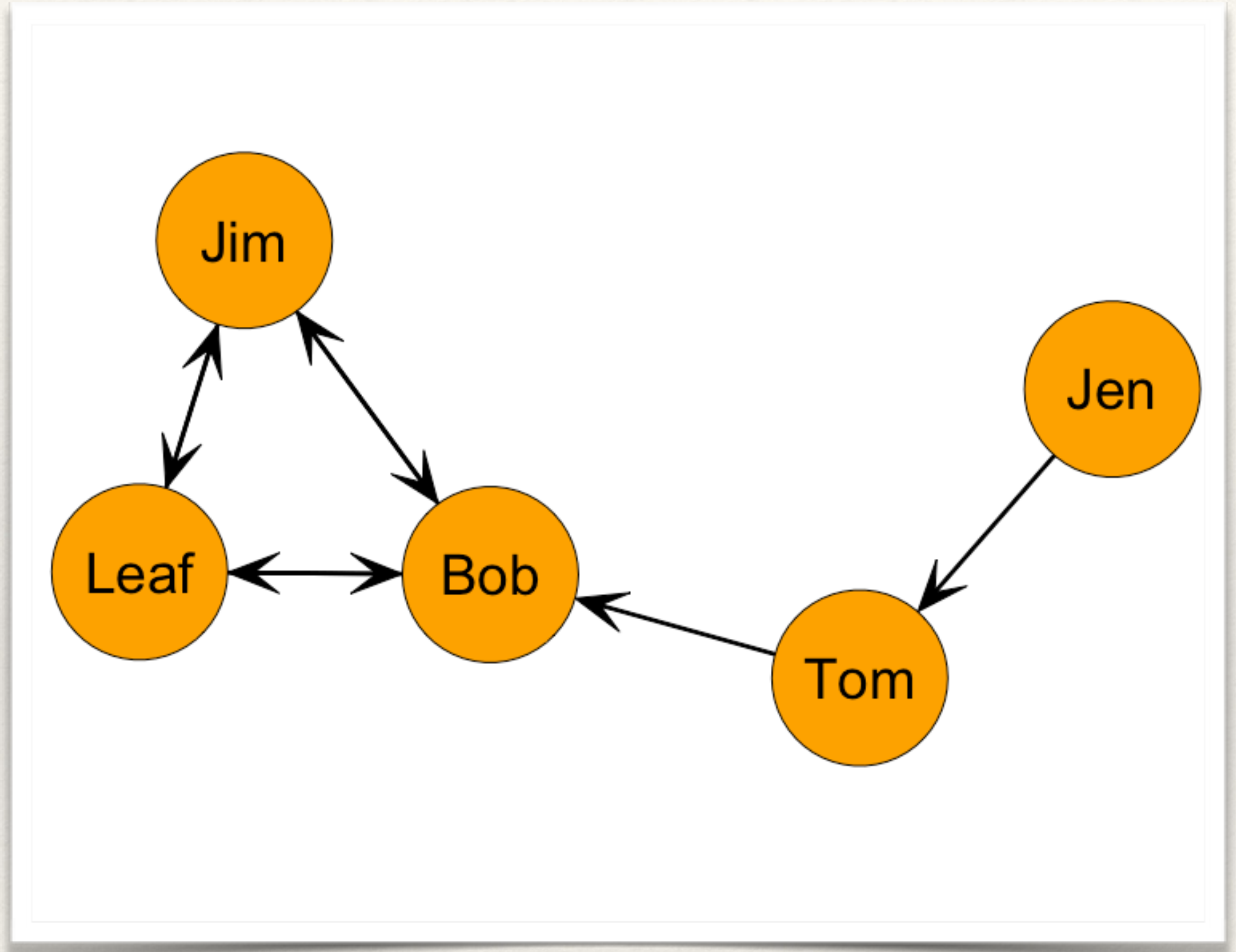
Tom = 1/4 =0.25

Bob = 3/4 = 0.75

Leaf = 2/4 = 0.50

Jim = 2/4 = 0.50

Example: Directed, Binary Network



What is the standardized indegree and outdegree centrality score for each node?

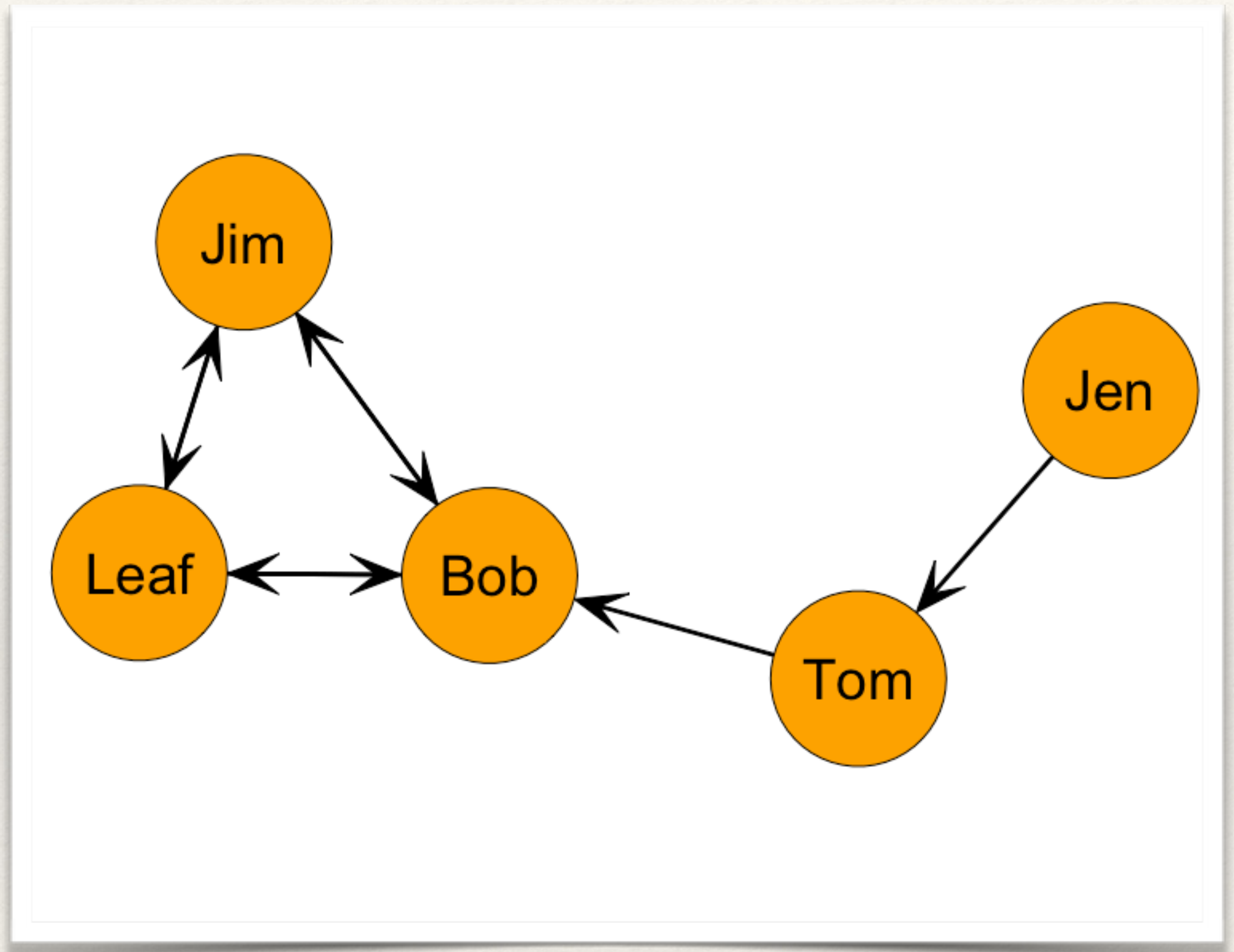
<u>Raw Indegree Centrality</u>

Jen = 0
Tom = 1
Bob = 3
Leaf = 2
Jim = 2

<u>Raw Outdegree Centrality</u>

Jen = 1
Tom = 1
Bob = 2
Leaf = 2
Jim = 2

Example: Directed, Binary Network



What is the standardized indegree and outdegree centrality score for each node?

Standardized Indegree Centrality

Jen = 0/4 = 0
Tom = 1/4 =0.25
Bob = 3/4 = 0.75
Leaf = 2/4 = 0.50
Jim = 2/4 = 0.50

Standardized Outdegree Centrality

Jen = 1/4 = 0.25
Tom = 1/4 = 0.25
Bob = 2/4 = 0.50
Leaf = 2/4 = 0.50
Jim = 2/4 = 0.50

Summarizing Degree Centrality

- As before, we can examine the summary statistics for degree centrality by inspecting the **mean**.

Mean Degree (directed)

Or, number of edges

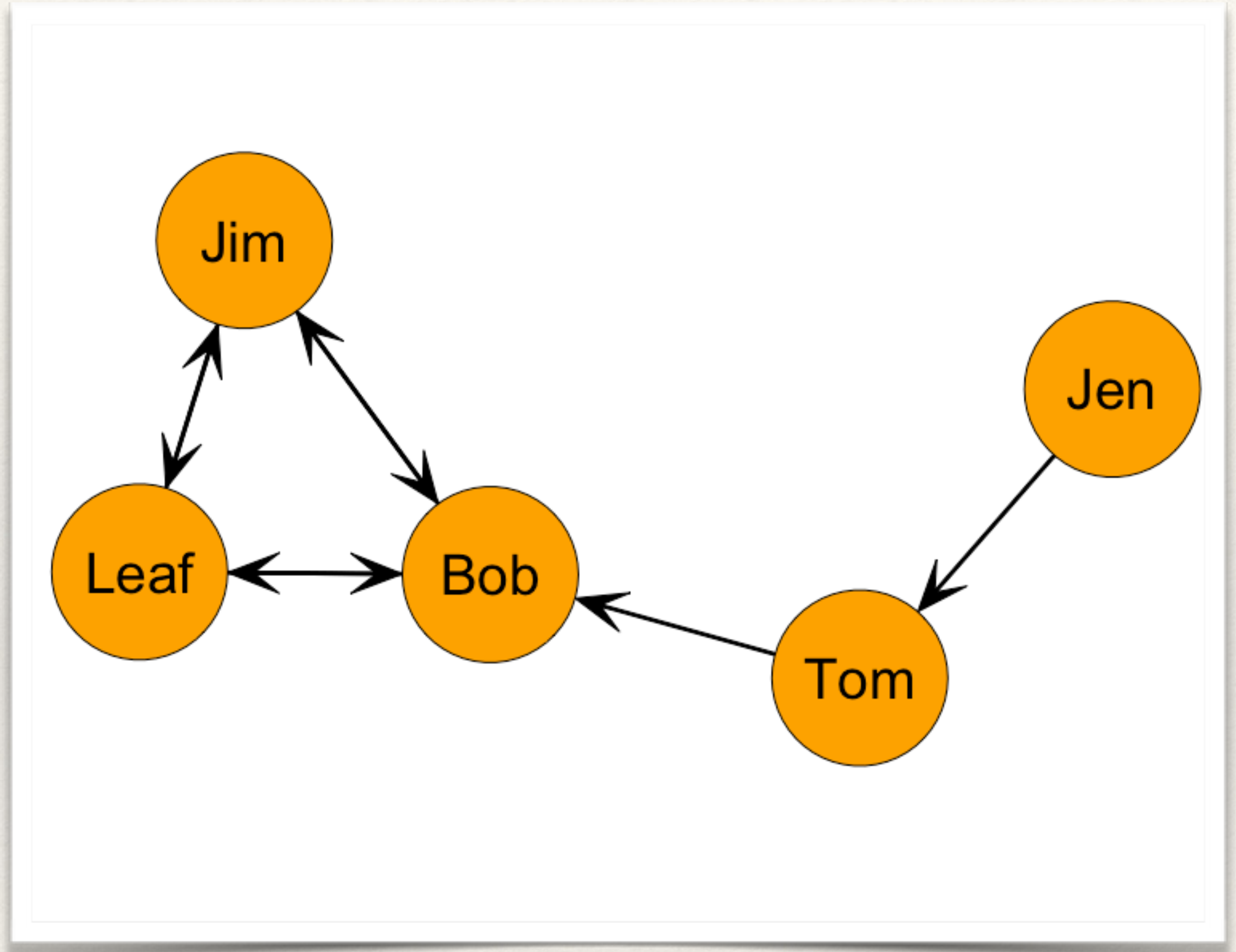
$$\bar{d} = \frac{\sum_{i=1}^{g} C_I(n_i)}{g} = \frac{\sum_{i=1}^{g} C_O(n_i)}{g} = \frac{L}{g}$$

Divide by number of actors

Summarizing Degree Centrality

- The mean indegree is equal to the mean outdegree.
  - *Why?*

Example: Directed, Binary Network



$$\bar{d} = \frac{C_I(n_i)}{g} = \frac{C_O(n_i)}{g} = \frac{L}{g} = \frac{8}{5} = 1.6$$

What is the mean indegree/ outdegree for this graph?

Summarizing Degree Centrality

- We can also calculate how centralized the graph itself is.
  - *Group degree centralization* measures the extent to which the actors in a social network differ from one another in their individual degree centralities.
    - The difference here is that the denominator is $(g-1)^2$ or $(g-1)(g-1)$.
      - Note that the numerator may differ though.

Index of Group Degree Centralization

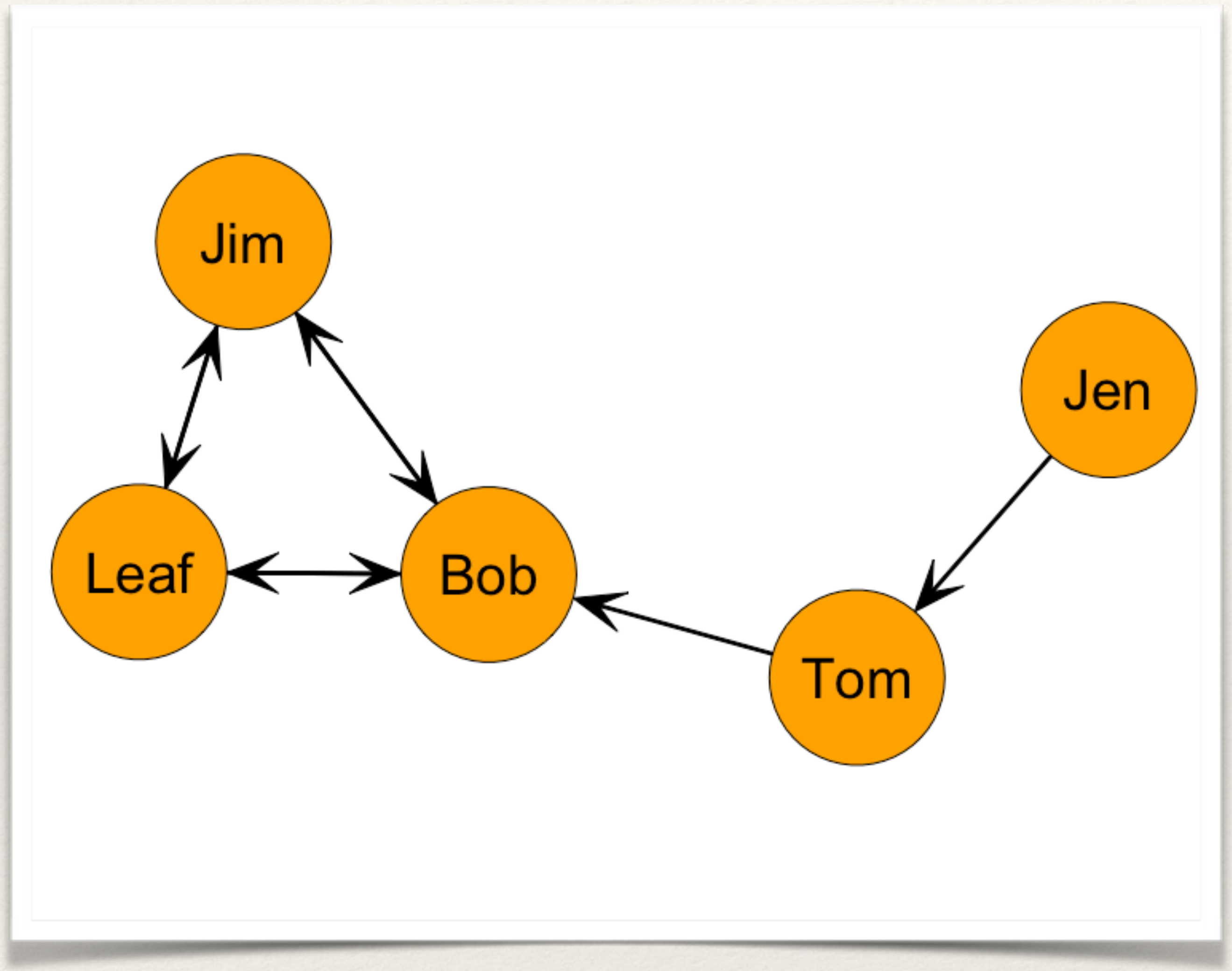
$$C_D = \frac{\sum_{i=1}^{g}[C_D(n^*) - C_D(n_i)]}{[(g-1)^2]}$$

Sum of observed differences between the largest actor centrality and all others

The maximum possible sum of differences

Note the difference (see W&F p. 199)

Example: Directed, Binary Network



<u>Raw Indegree Centrality</u>

Jen = 0

Tom = 1

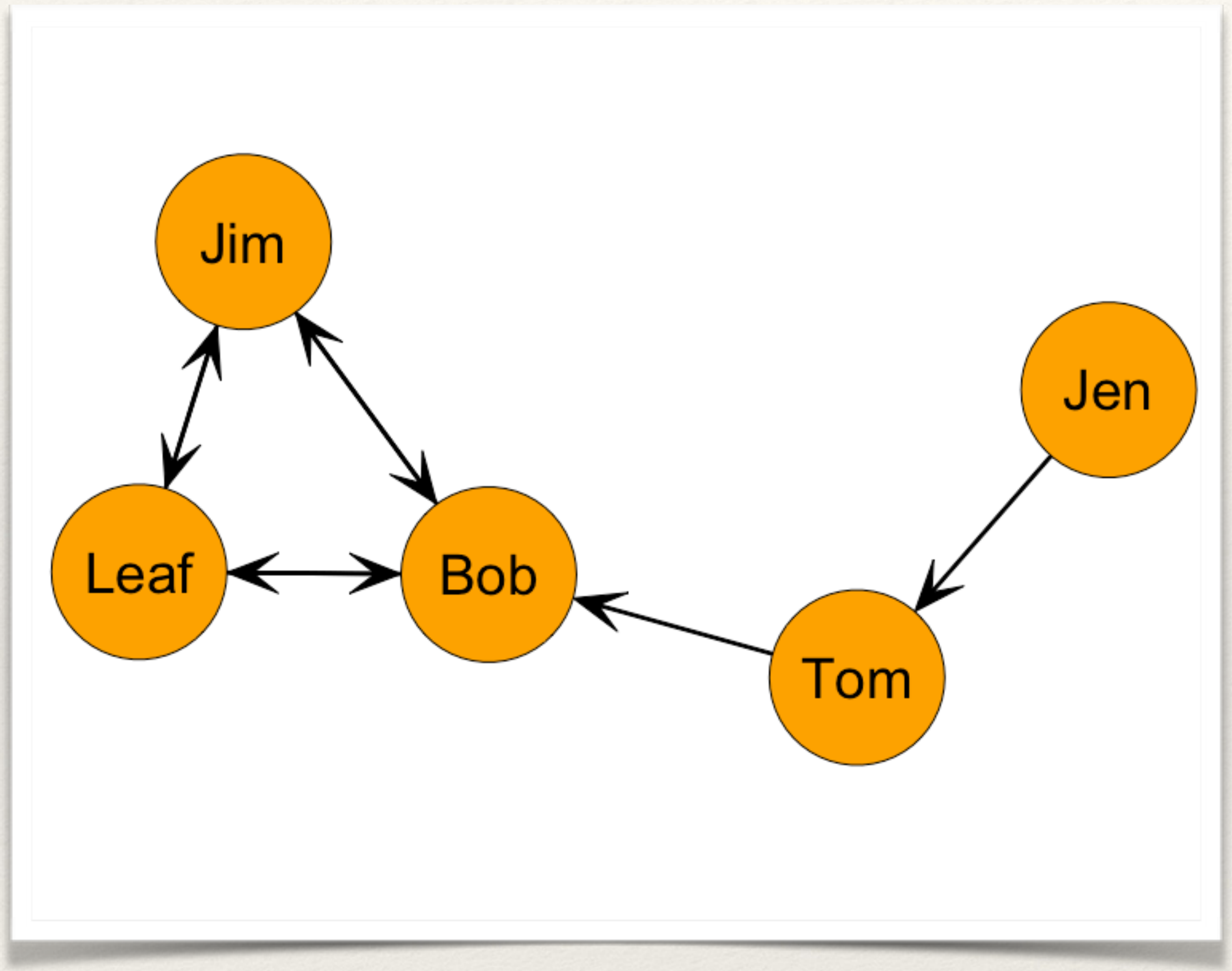
Bob = 3

Leaf = 2

Jim = 2

What is the index of indegree centralization for this graph?

Example: Directed, Binary Network



Raw Indegree Centrality

Jen = 0

Tom = 1

Bob = 3

Leaf = 2

Jim = 2

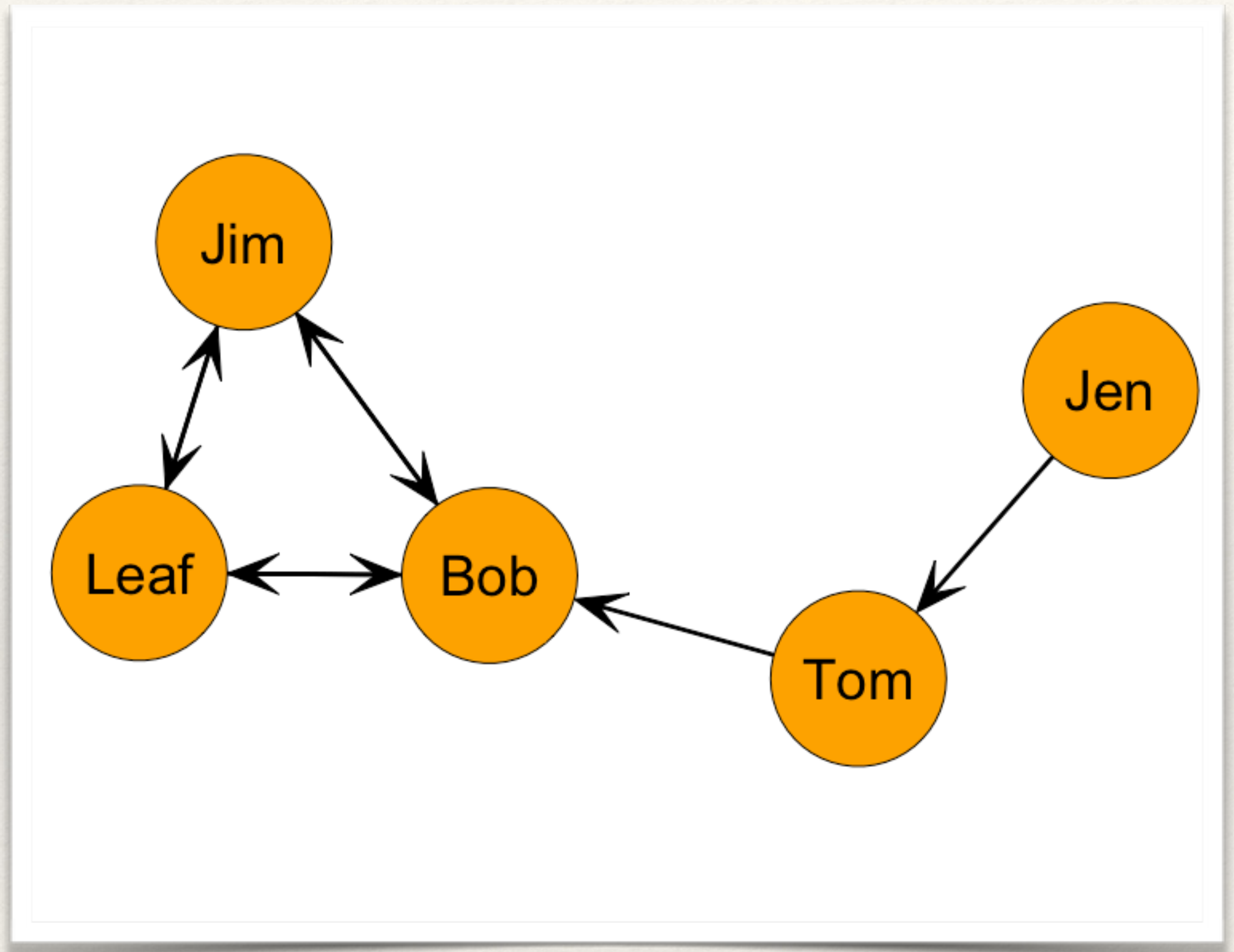
What is the index of indegree centralization for this graph?

0.4375

$$C_I = \frac{\sum_{i=1}^{g}[C_I(n^*) - C_I(n_i)]}{[(g-1)(g-1)]} =$$

$$= \frac{(3-0)+(3-1)+(3-3)+(3-2)+(3-2)}{(5-1)(5-1)} = \frac{3+2+0+1+1}{4*4} = \frac{7}{16} = 0.4375$$

Example: Directed, Binary Network



Raw Outdegree Centrality

Jen = 1

Tom = 1

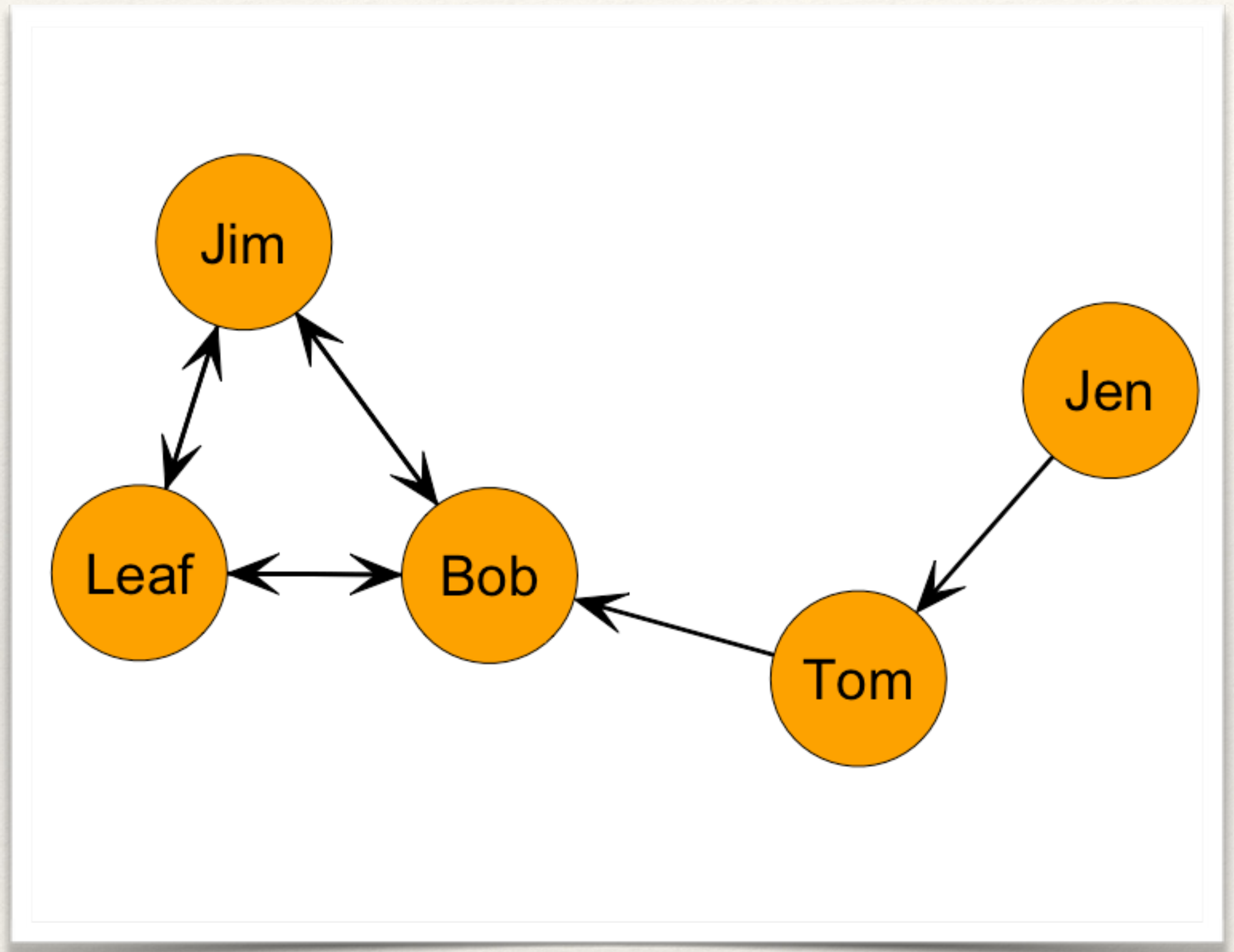
Bob = 2

Leaf = 2

Jim = 2

What is the index of outdegree centralization for this graph?

Example: Directed, Binary Network



Raw Outdegree Centrality

Jen = 1

Tom = 1

Bob = 2

Leaf = 2

Jim = 2

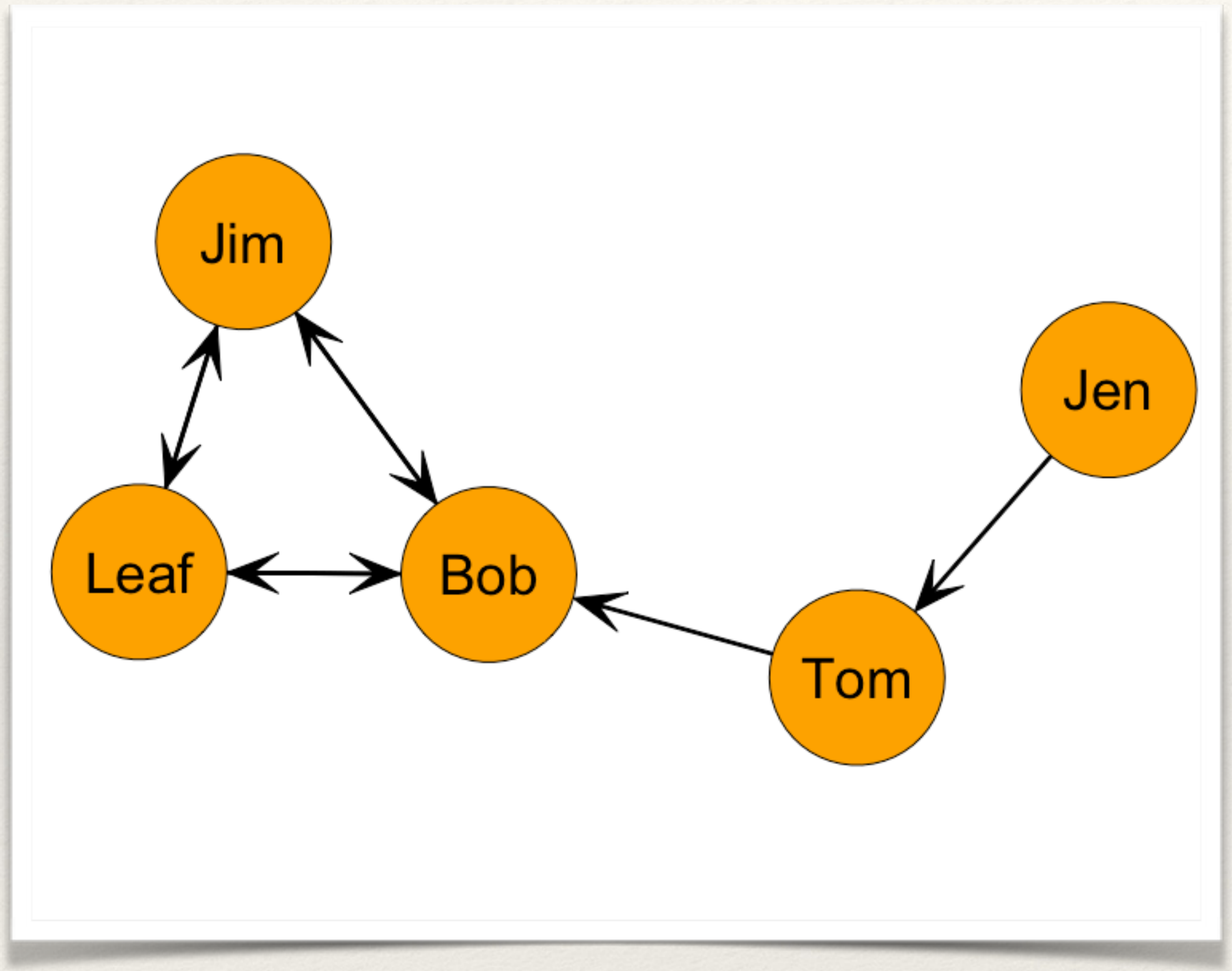
What is the index of outdegree centralization for this graph?

0.125

$$C_O = \frac{\sum_{i=1}^{g}[C_O(n^*) - C_O(n_i)]}{[(g-1)(g-1)]}$$

$$= \frac{(2-1)+(2-1)+(2-2)+(2-2)+(2-2)}{(5-1)(5-1)} = \frac{1+1+0+0+0}{4*4} = \frac{2}{16} = 0.125$$

Example: Directed, Binary Network



$$C_I = 0.4375$$

$$C_O = 0.125$$

What do the differences in the centralization scores tell us about the graph?

Learning Goals

- Understand the conceptualization of “centrality”.
- Understand calculation of degree centrality.
- Analyze descriptive features of degree centrality.

Empirical Example

J Youth Adolescence (2014) 43:104–115
DOI 10.1007/s10964-013-9946-0

EMPIRICAL RESEARCH

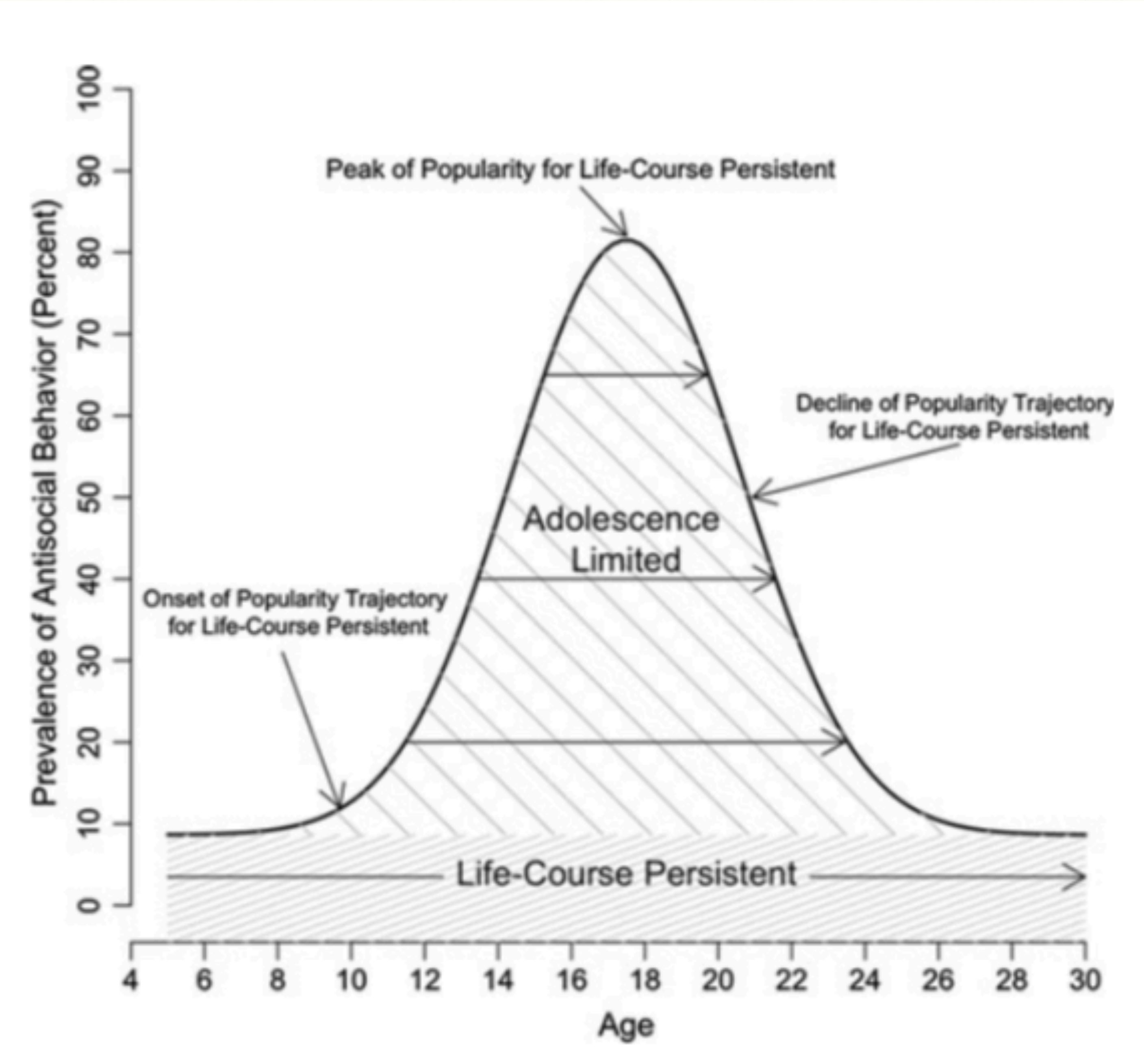
"Role Magnets"? An Empirical Investigation of Popularity Trajectories for Life-Course Persistent Individuals During Adolescence

Jacob T. N. Young

- https://link.springer.com/article/10.1007/s10964-013-9946-0

Empirical Example

- Question: Why is there a dramatic increase in delinquency during adolescence?
    - Argument: Dual-taxonomy theory.

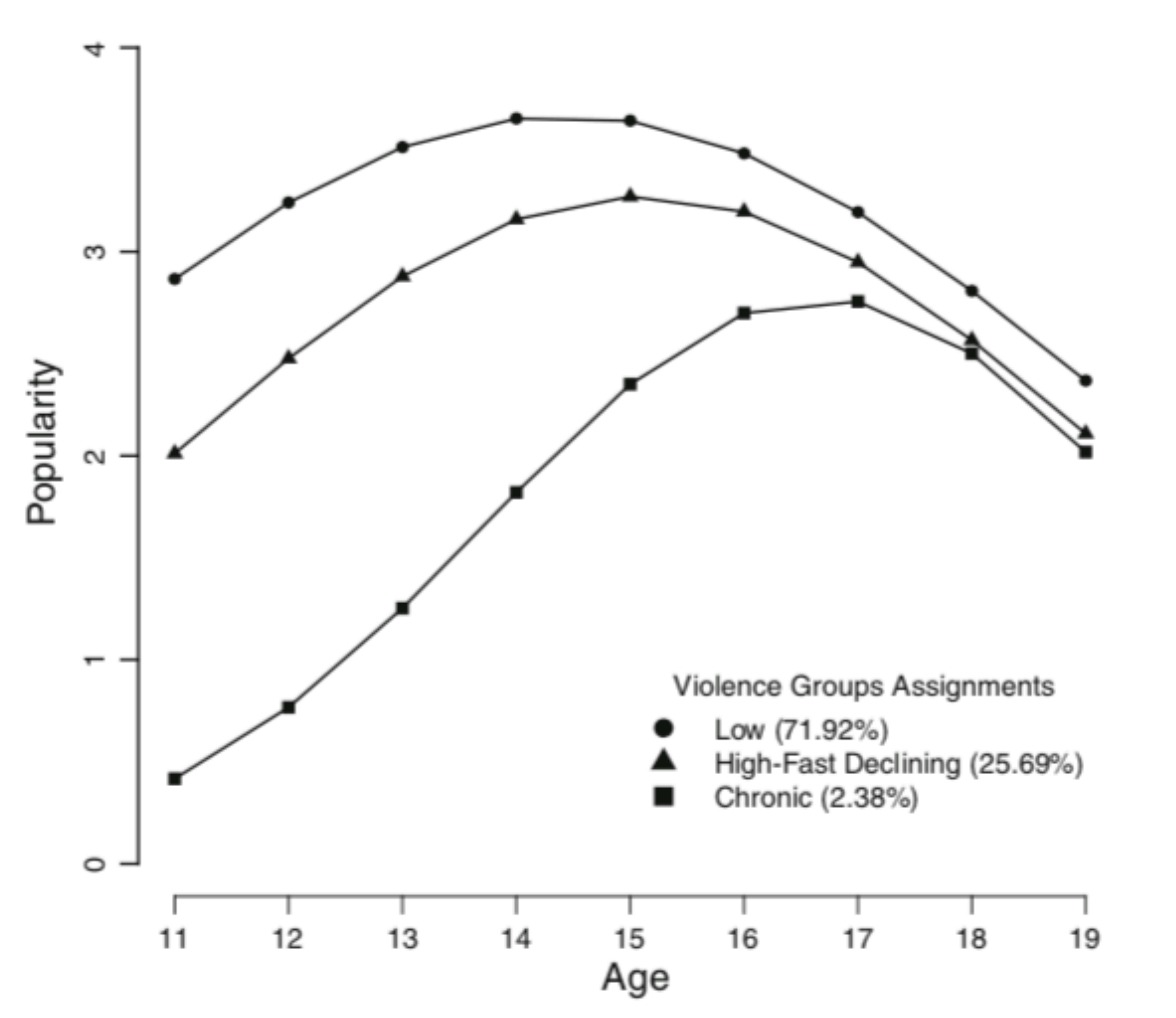


Empirical Example

- Dual-taxonomy theory argues that the causal mechanism generating the dramatic increase in delinquency during adolescence is social mimicry. As a consequence, life-course persistent individuals should occupy “more influential positions in the peer social structure” (Moffitt 1993: 687) and should be “moving toward central positions, during early adolescence” (Moffitt 1997: 28).
  - Concept: Popularity
    - Operationalization: Indegree Centrality
      - Goal: Examine the developmental trajectory of popularity during adolescence for individuals showing persistent violence into young adulthood.

Empirical Example

- Findings: Chronically violent individuals showed a more precipitous increase in indegree centrality during adolescence.



Questions?

Where to go from here…

Resources

- The internet
- Check out professional associations
  - INSNA (insna.org)
- Journals
  - *Social Networks*
  - *Network Science*

Computation

- R (cran.r-project.org)
  - https://jacobtnyoung.github.io/RWorkshop/
- UCINET
  - https://sites.google.com/site/ucinetsoftware/home

WARNING! Shameless Plug!!!

Future Classes

- Got the network fever?
  - Network Criminology (Spring 22)
  - Statistical Analysis of Network Data (Spring 23)

Thank you!!!!

Jacob Young
jacob.young.1@asu.edu
GITHUB PAGE