Statistical Analysis of Networks

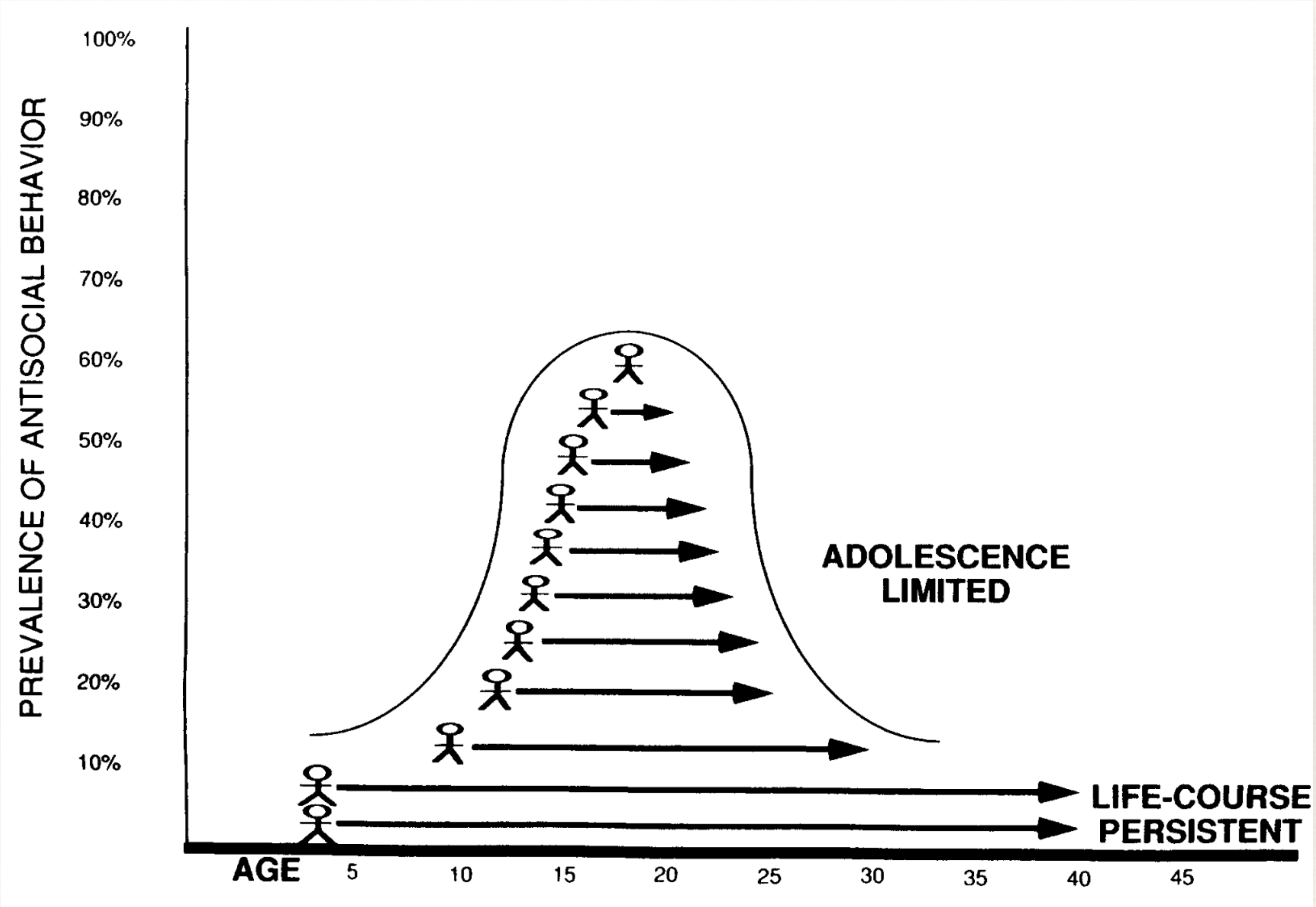
Degree Centrality

Motivating Example

- Question: Why is there a dramatic increase in delinquency during adolescence?

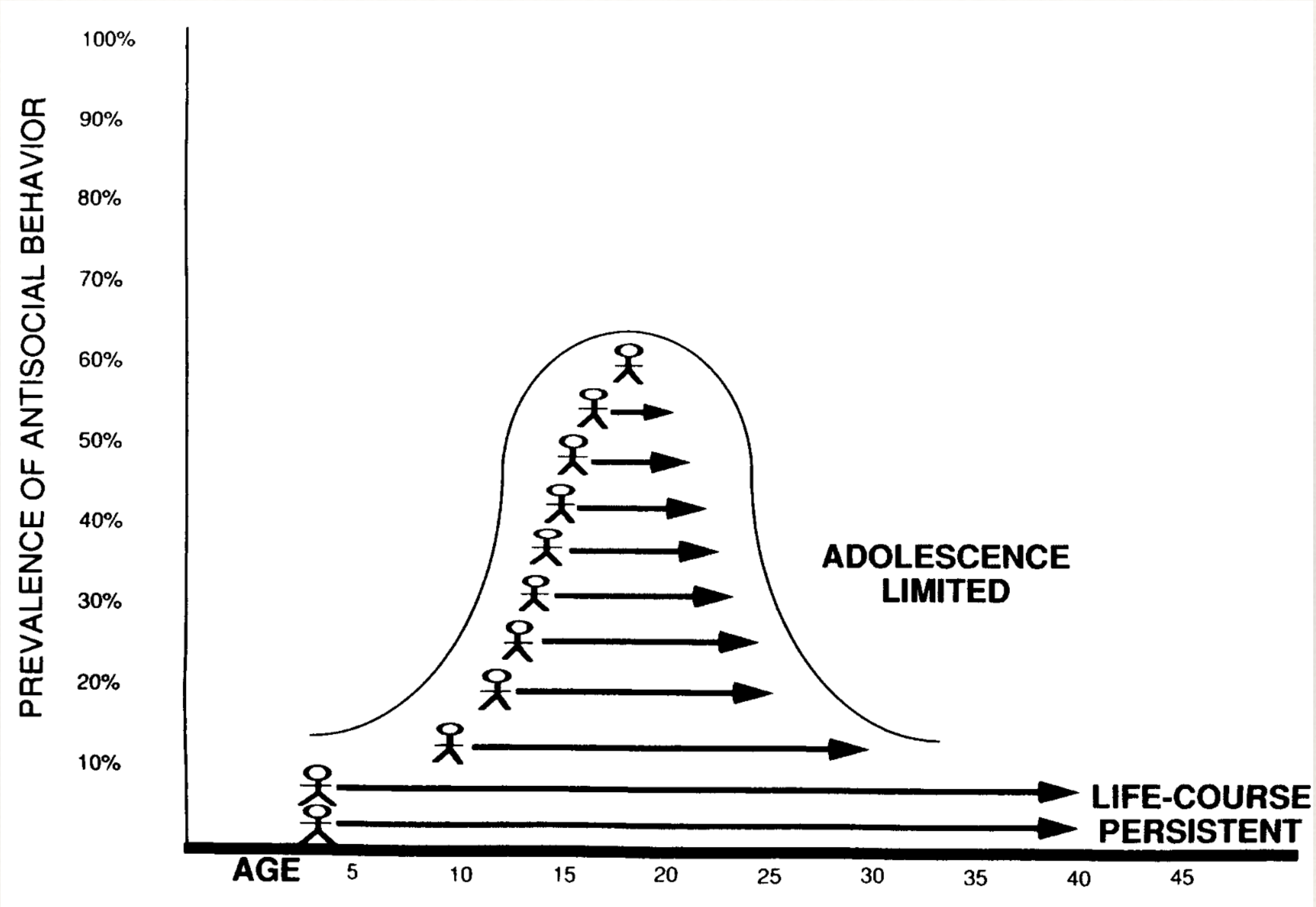
Motivating Example

- Question: Why is there a dramatic increase in delinquency during adolescence?
    - Argument: Moffitt's (1993) Dual-taxonomy theory.



Motivating Example

- Question: Why is there a dramatic increase in delinquency during adolescence?
  - Argument: Moffitt's (1993) Dual-taxonomy theory.
    - *So what?*



Motivating 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).
  - This is a hypothesis about the trajectory of centrality for a set of individuals

Motivating Example

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

EMPIRICAL RESEARCH

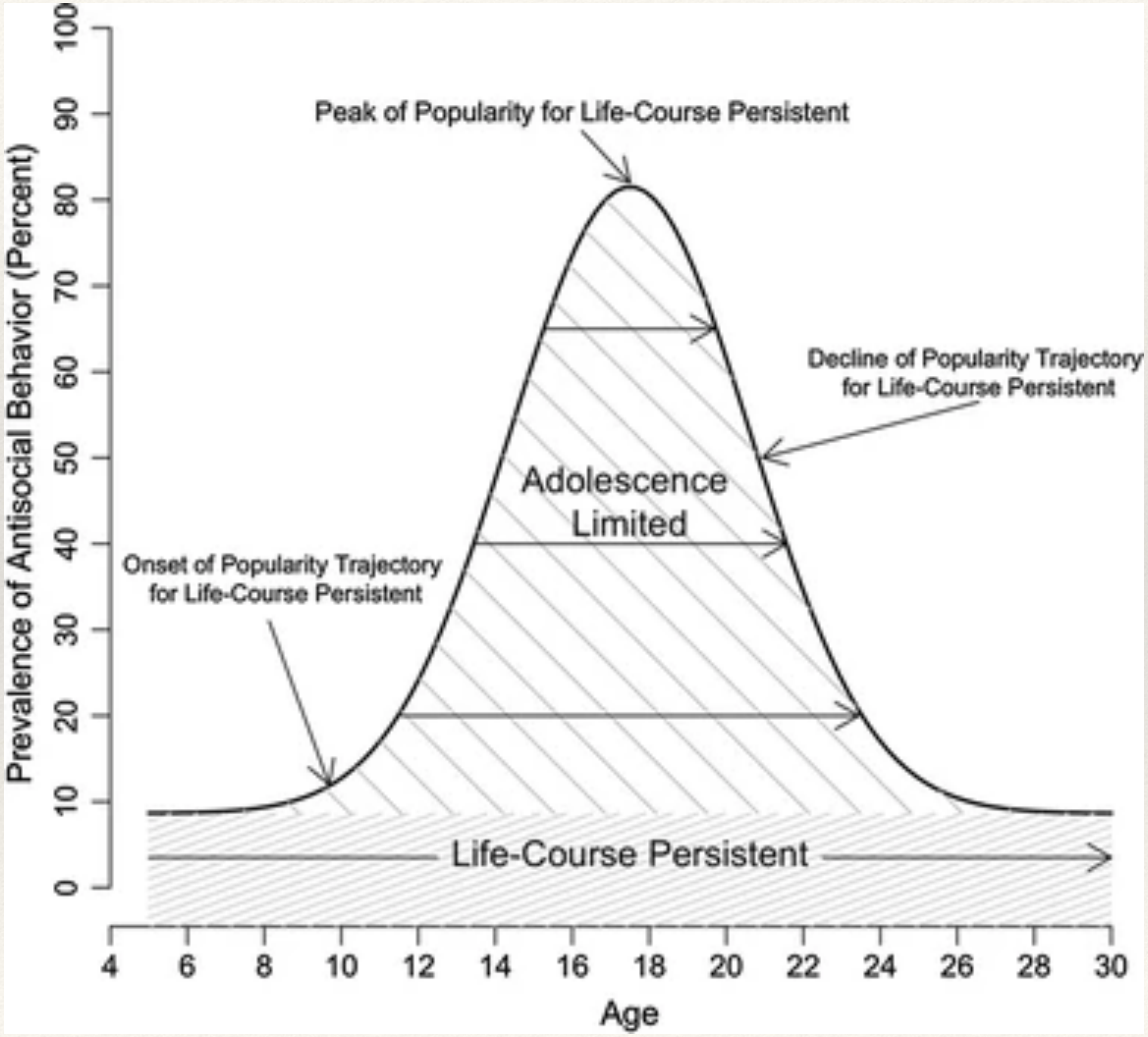
"Role Magnets"? An Empirical Investigation of Popularity Trajectories for Life-Course Persistent Individuals During Adolescence

Jacob T. N. Young

- https://link.springer.com/article/10.1007/s10964-013-9946-0
- Goal: Examine the developmental trajectory of popularity during adolescence for individuals showing persistent violence into young adulthood.

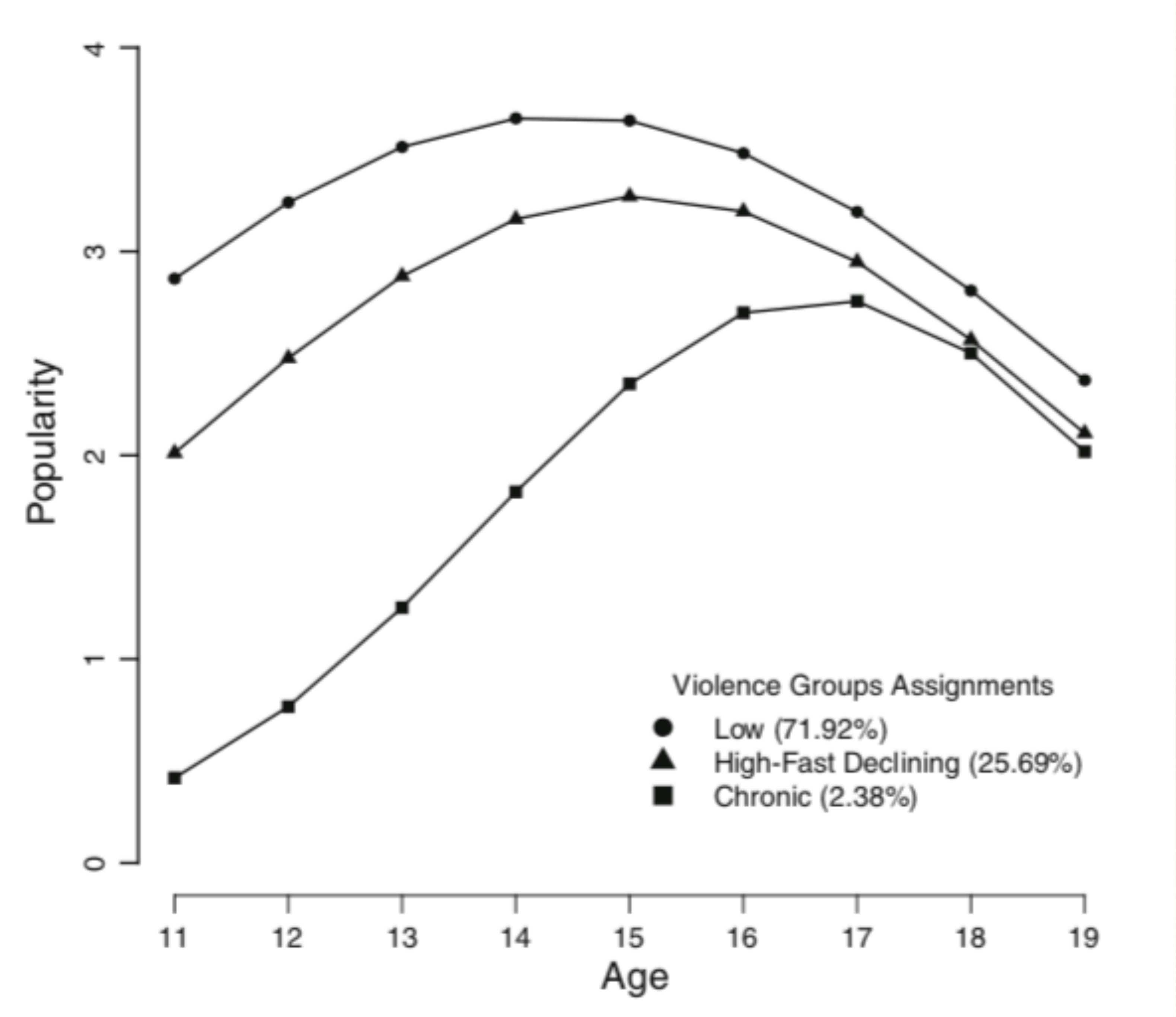
Motivating Example

- This is a hypothesis about the trajectory of centrality for a set of individuals



Motivating Example

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



Another Motivating Example

- How do you reduce hate speech in online forums?

Another Motivating Example

- "Deplatforming" is a type of **strategic network disruption**
- Why should it work? According to this intervention, what does the network look like in online hate organizations? (Draw it)

Another Motivating Example

- "Deplatforming" is a type of **strategic network disruption**

- Does it work?

PNAS RESEARCH ARTICLE | POLITICAL SCIENCES OPEN ACCESS

Disrupting hate: The effect of deplatforming hate organizations on their online audience

Daniel Robert Thomas[a,1] and Laila A. Wahedi[a]

Edited by Timothy Wilson, University of Virginia, Charlottesville, VA; received August 17, 2022; accepted January 20, 2023

Another Motivating Example

- Yes!
- Hate speech declined after the disruption events.

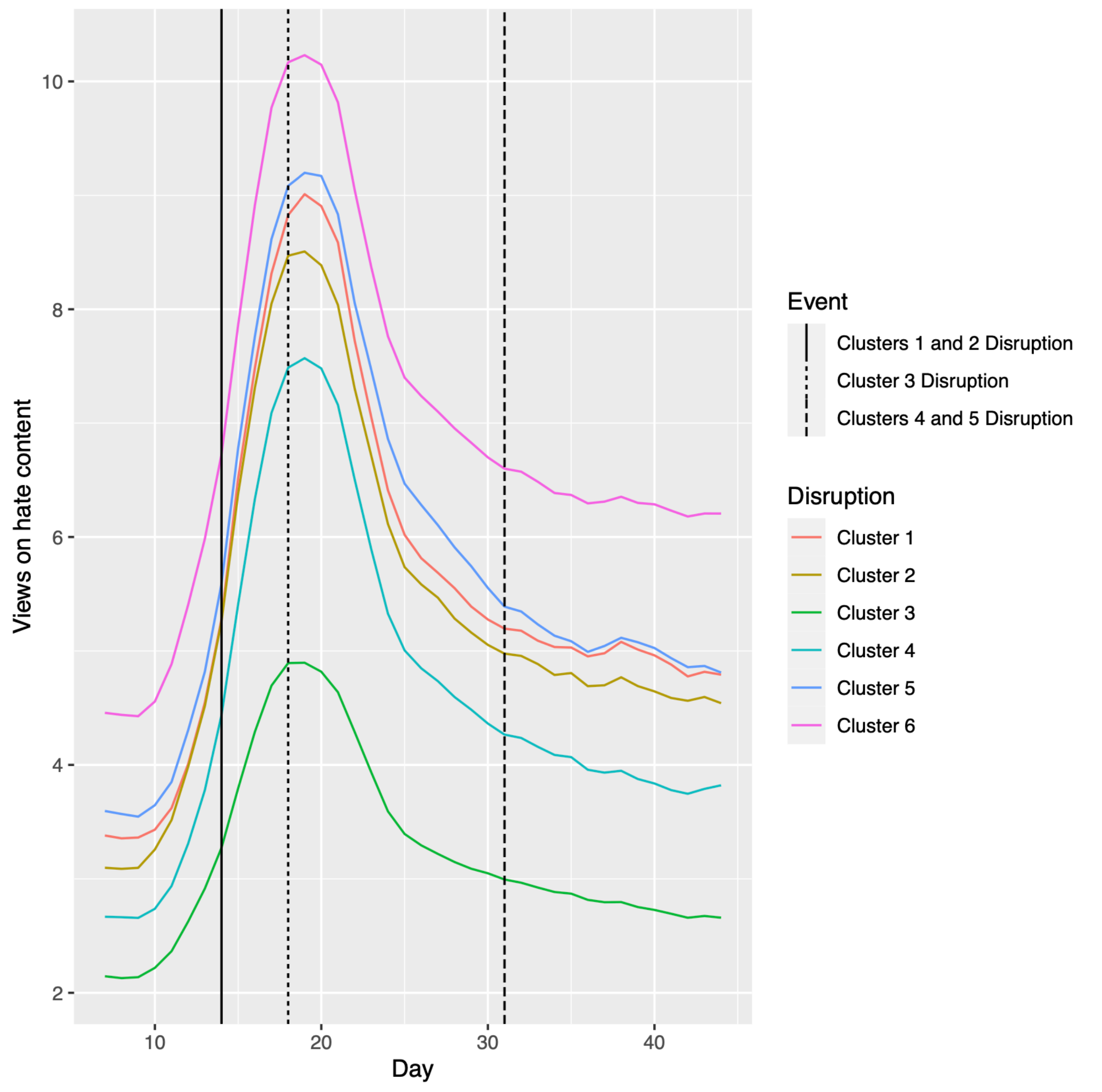


Fig. 1. Views on hateful content by organization over time: All six organizations have similar time trends prior to the initial disruptions. The spike in hateful content corresponds with the beginning of the George Floyd protests. Note that it is difficult to discern treatment effects from this descriptive plot because treatment effects are a combination of effects over the postdisruption study period. Figs. 2–5 for magnitude of the treatment effects.

Degree Centrality

Learning Goals

- At the end of the lecture, you should be able to answer these questions:
    - How can we conceptualize “centrality”.
    - How can we operationalization centrality as “degree”.
    - How do you calculate degree centrality for undirected and directed graphs?
    - What are the descriptive properties 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 description of the position of a **node/vertex** as a:
    - dependent variable (e.g. are taller individuals more likely to be trusted?)
    - independent variable (e.g. are more popular adolescents more likely to succeed in school?)
  - *And,* as a description of an entire **network**:
    - Is this needle-sharing network hierarchical or decentralized?

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.

Conceptualization

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

Undirected Networks

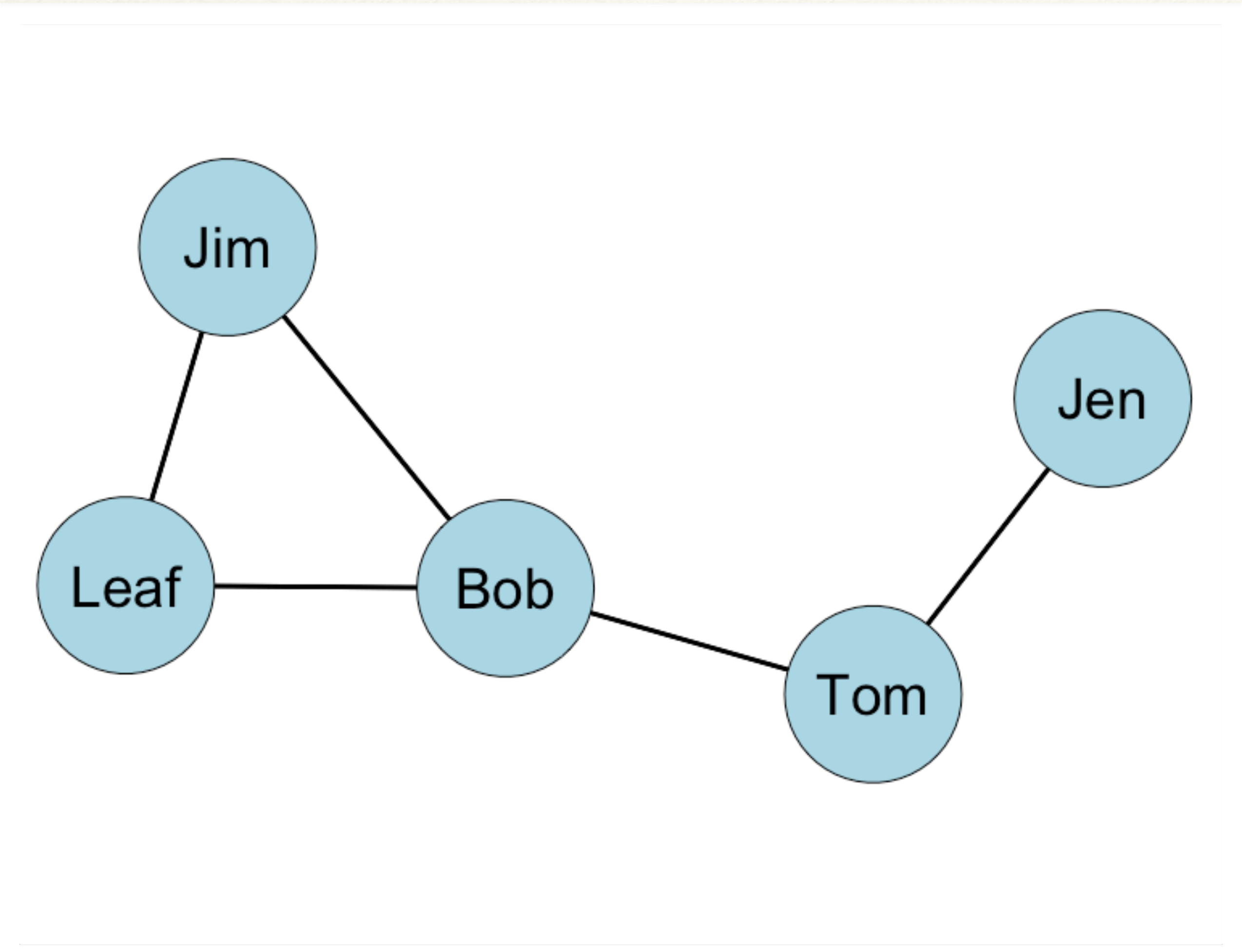
Degree Centrality: Undirected Binary Graphs

- 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

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

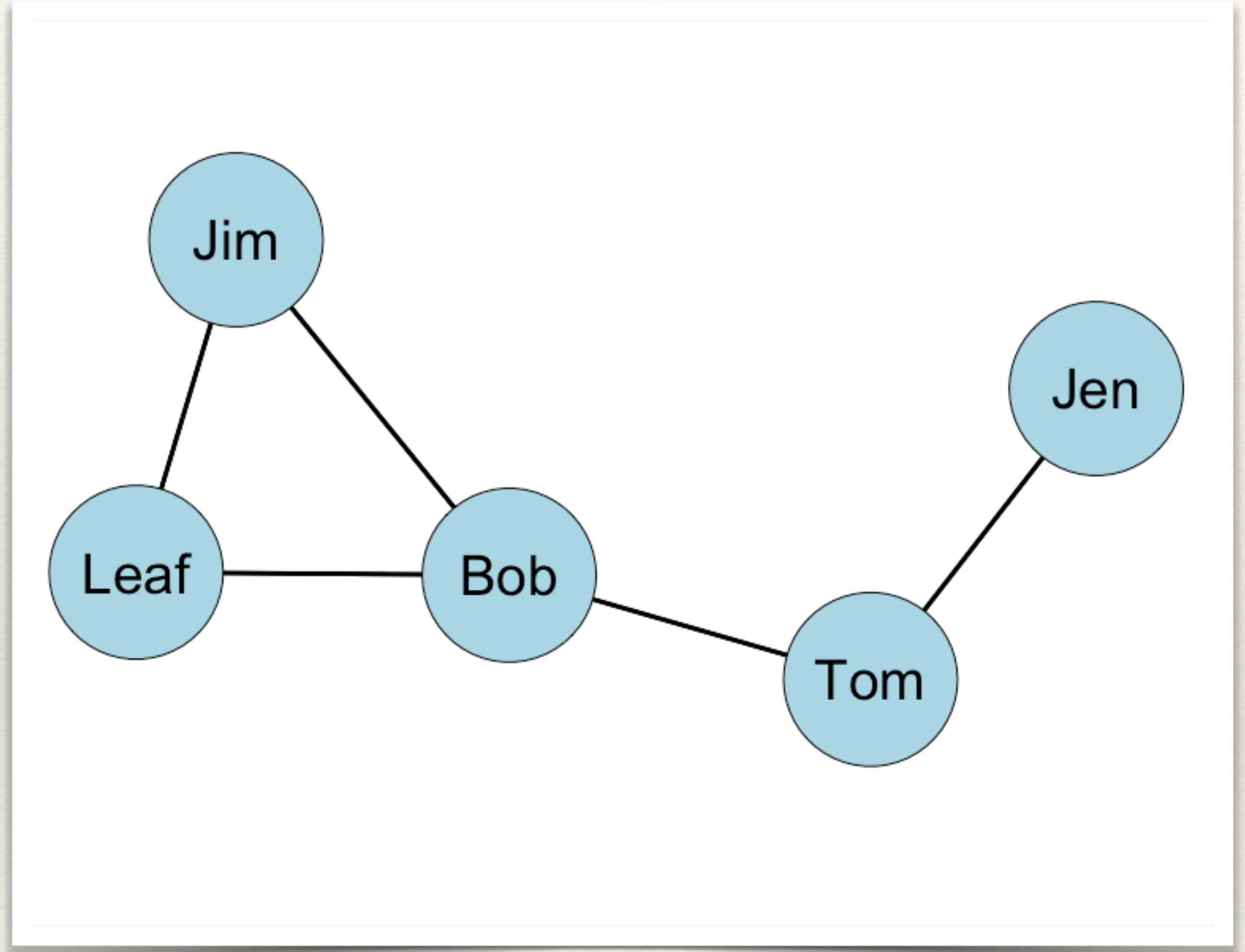
Example: Undirected, Binary Network



	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

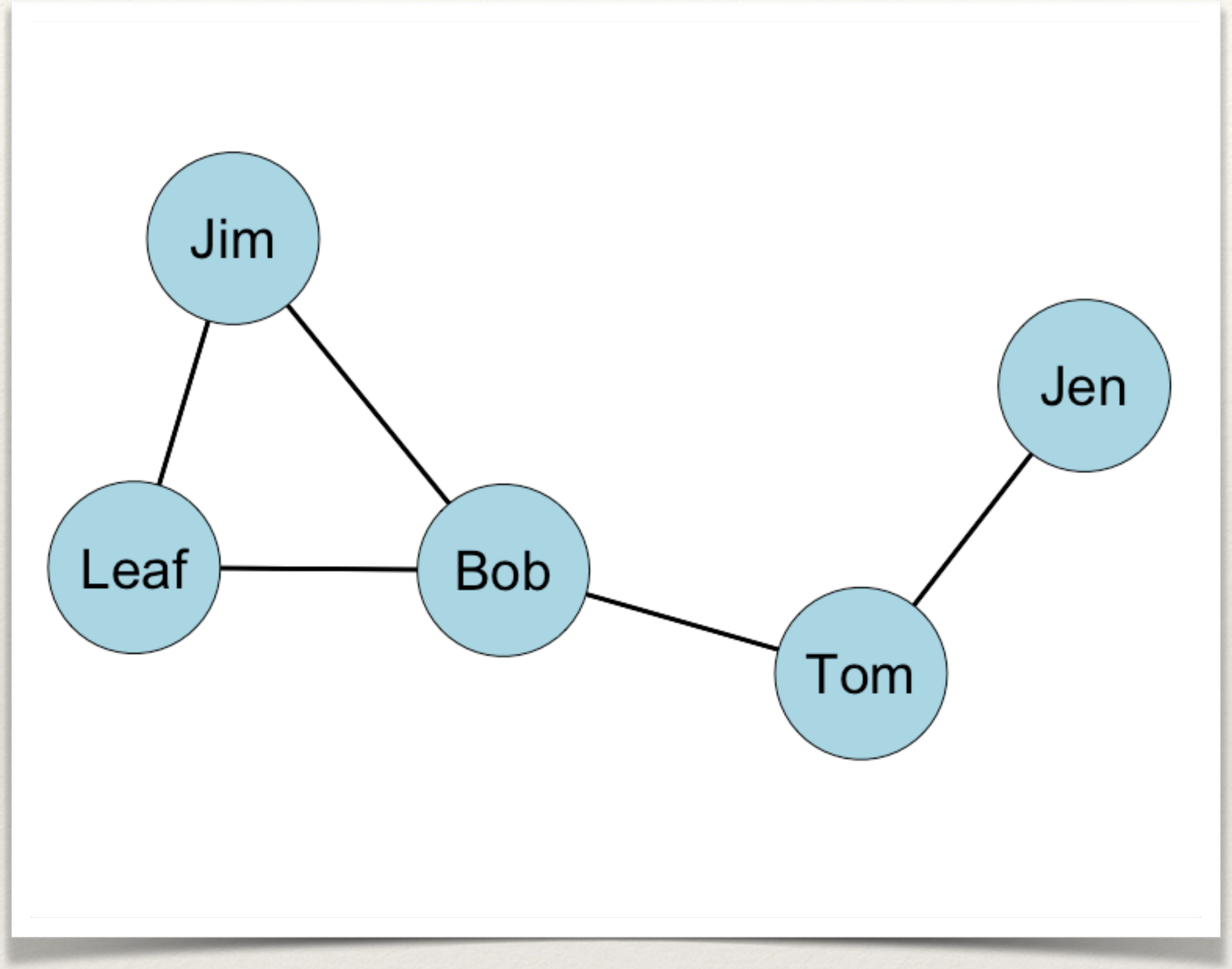
$$C'_D(n_i) = \frac{d(n_i)}{g-1} = \frac{\sum_j x_{ij}}{g-1}$$

Why do we subtract 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



<u>Raw Degree Centrality</u>

Jen = 1

Tom = 2

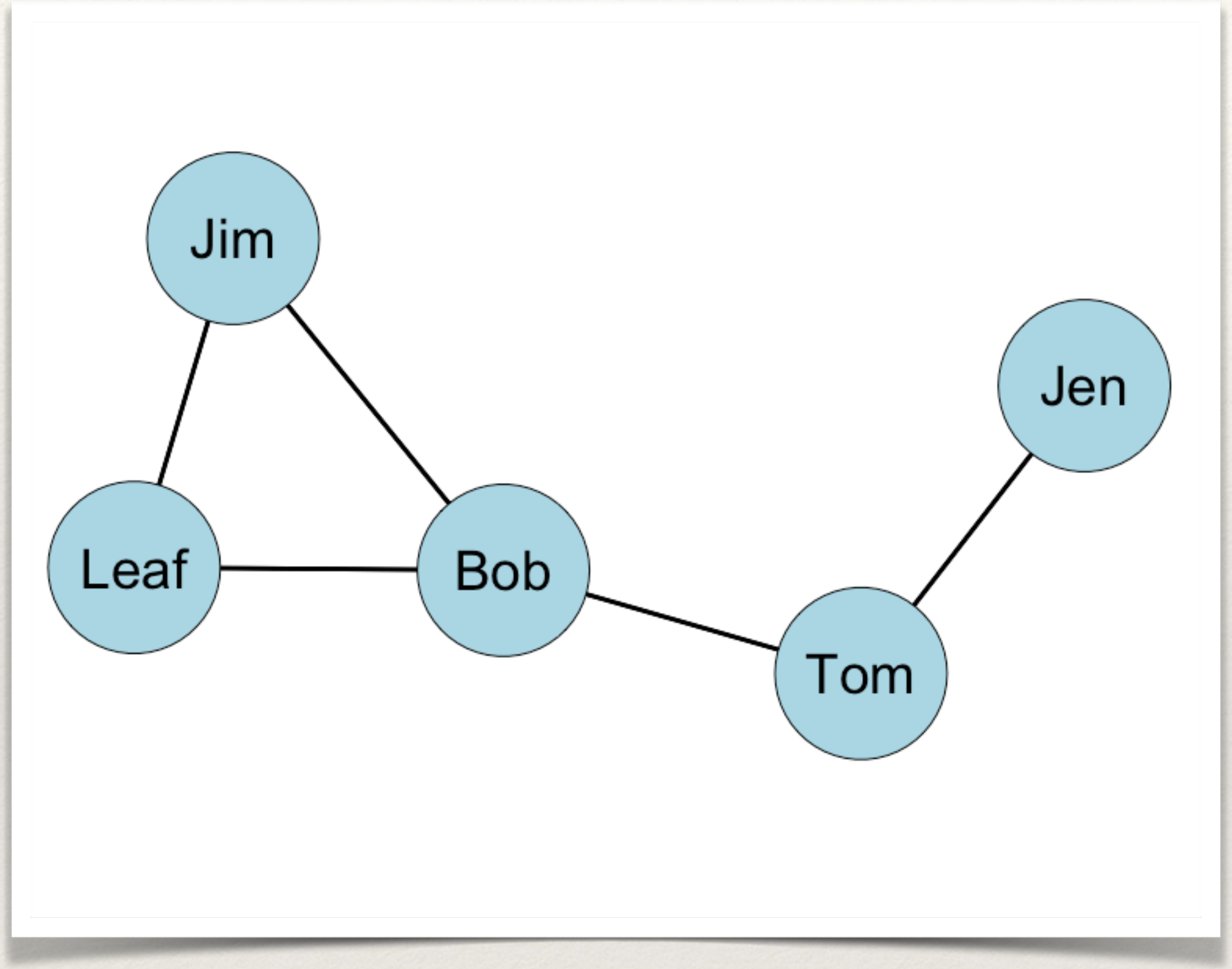
Bob = 3

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**.

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

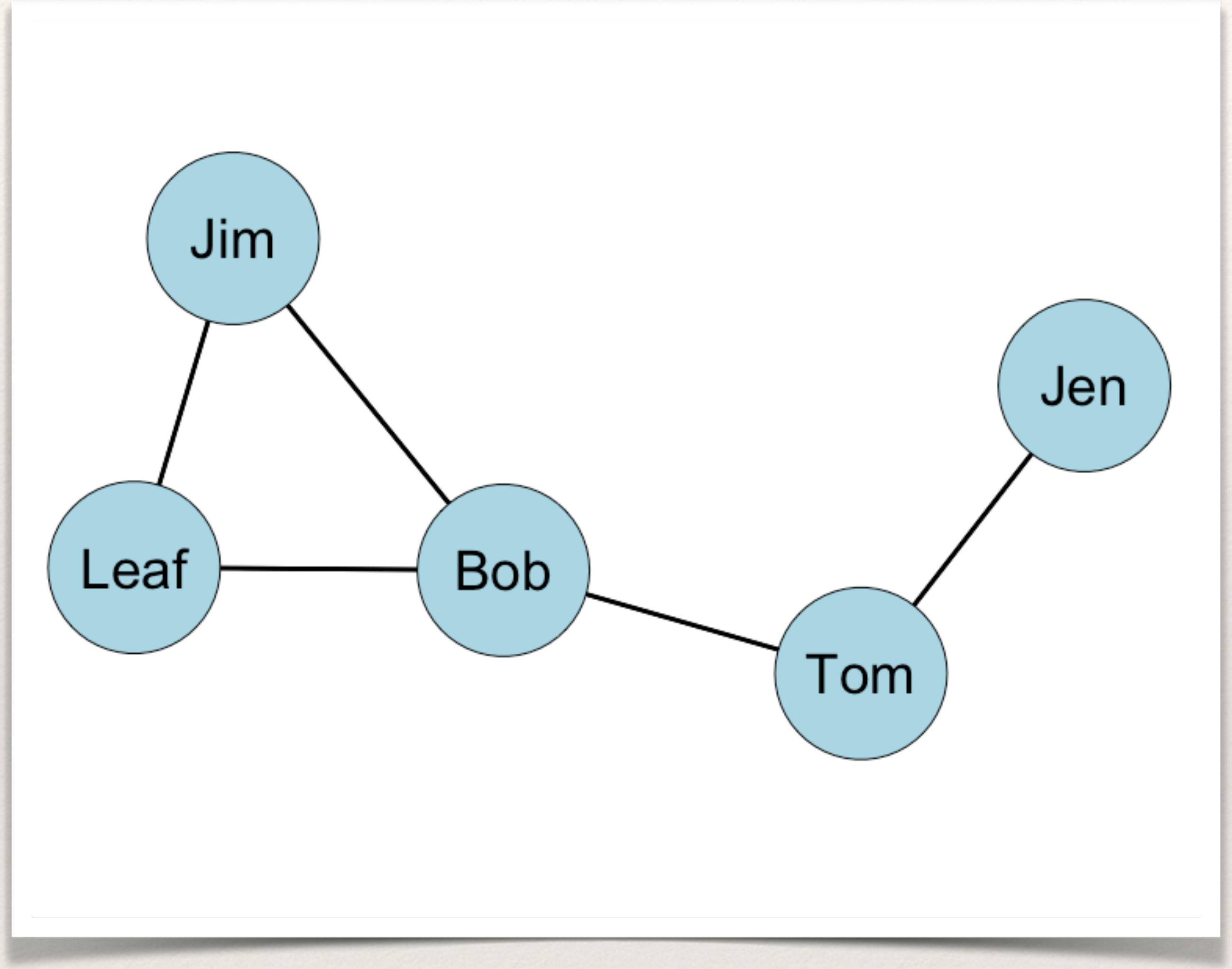
Mean Degree (undirected)

Or, multiply the number of edges by 2.

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

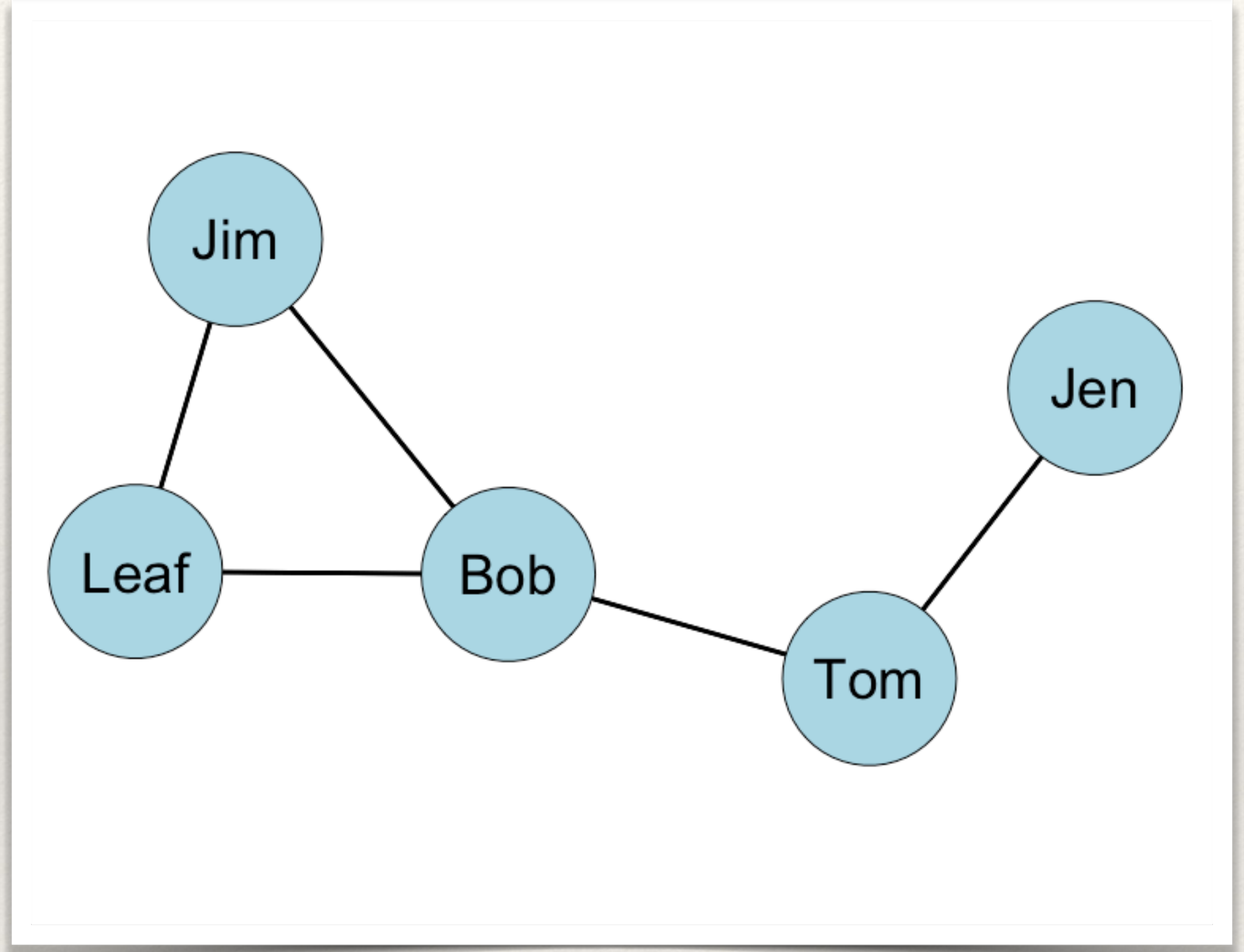
Divide by number of actors

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

- 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?
            - **Draw a picture of each…(I will wait)**

Summarizing Degree Centrality

- Note that degrees form a distribution.
  - The average degree tells us the central tendency of that distribution.
- What is another way we can describe a distribution?
  - (Hint: think back to your stats courses)

Summarizing Degree Centrality

- Dispersion!
- We can describe the dispersion in the centrality scores.
  - In sna, this is referred to as *centralization*.
  - *Group degree centralization* measures the extent to which the actors in a social network differ from one another in their individual degree centralities.
    - This should sound familiar, what does the standard deviation tell you?

Standard Deviation

$$SD_y = \sqrt{\frac{\sum (y_i - \mu)^2}{N}}$$

?

Group Degree Centralization

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

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!)

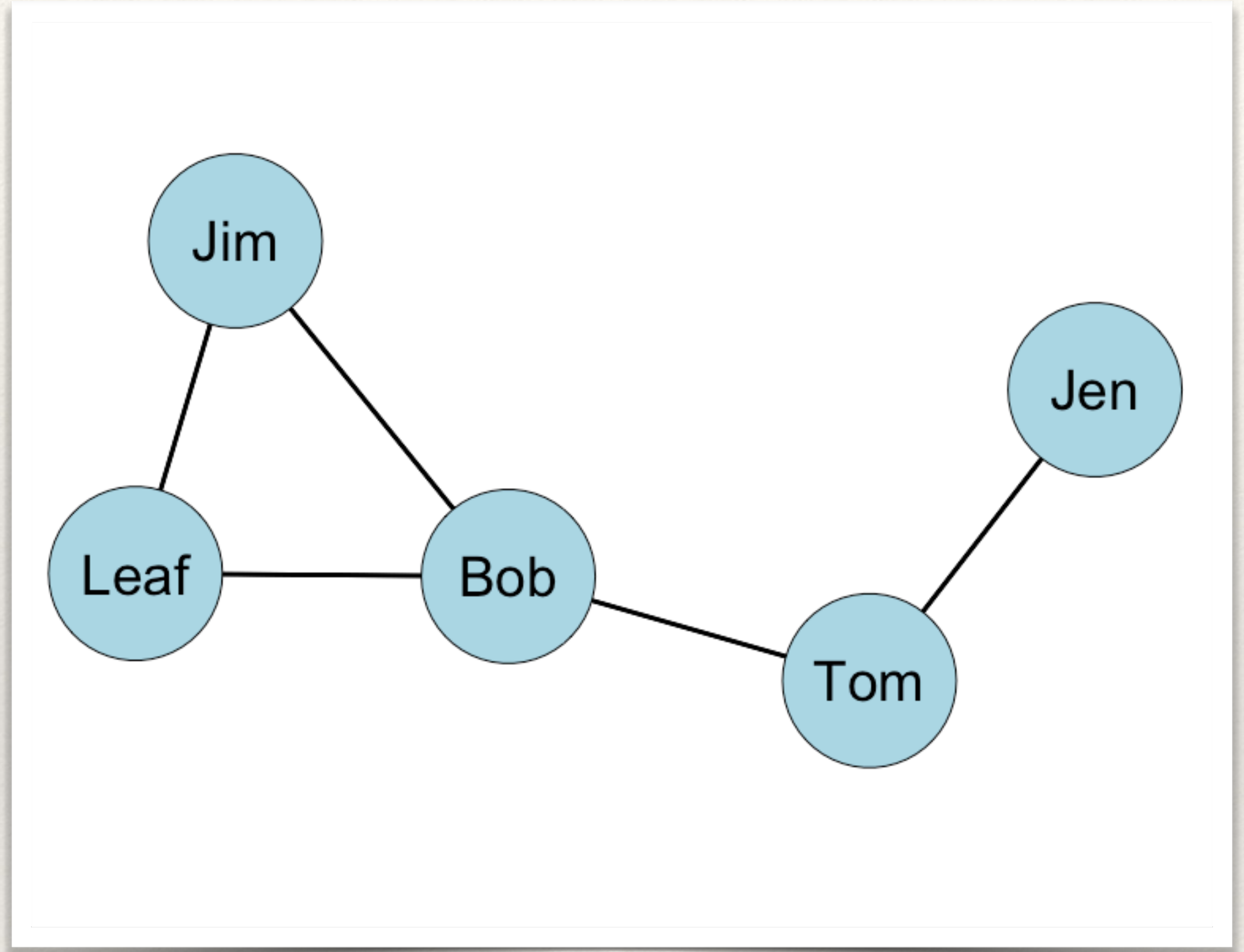
Index of Group Degree Centralization

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

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

The maximum possible sum of differences

Example: Undirected, Binary Network



<u>Raw Degree Centrality</u>

Jen = 1

Tom = 2

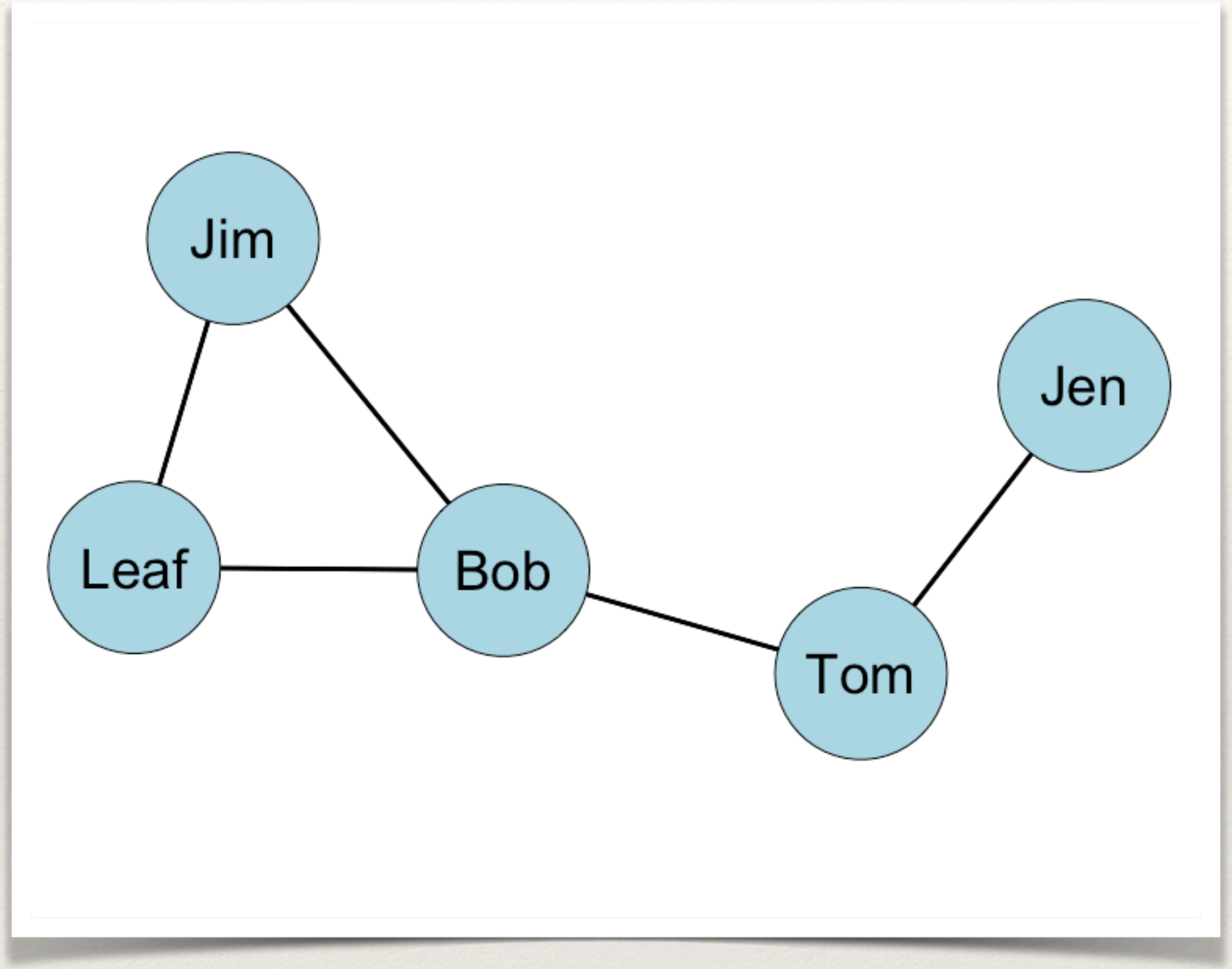
Bob = 3

Leaf = 2

Jim = 2

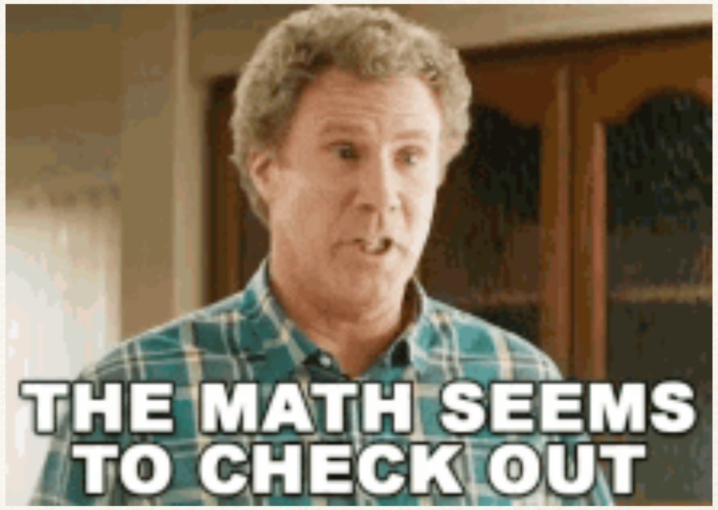
What is the index of degree centralization for this graph?

Example: Undirected, Binary Network



0.4167

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

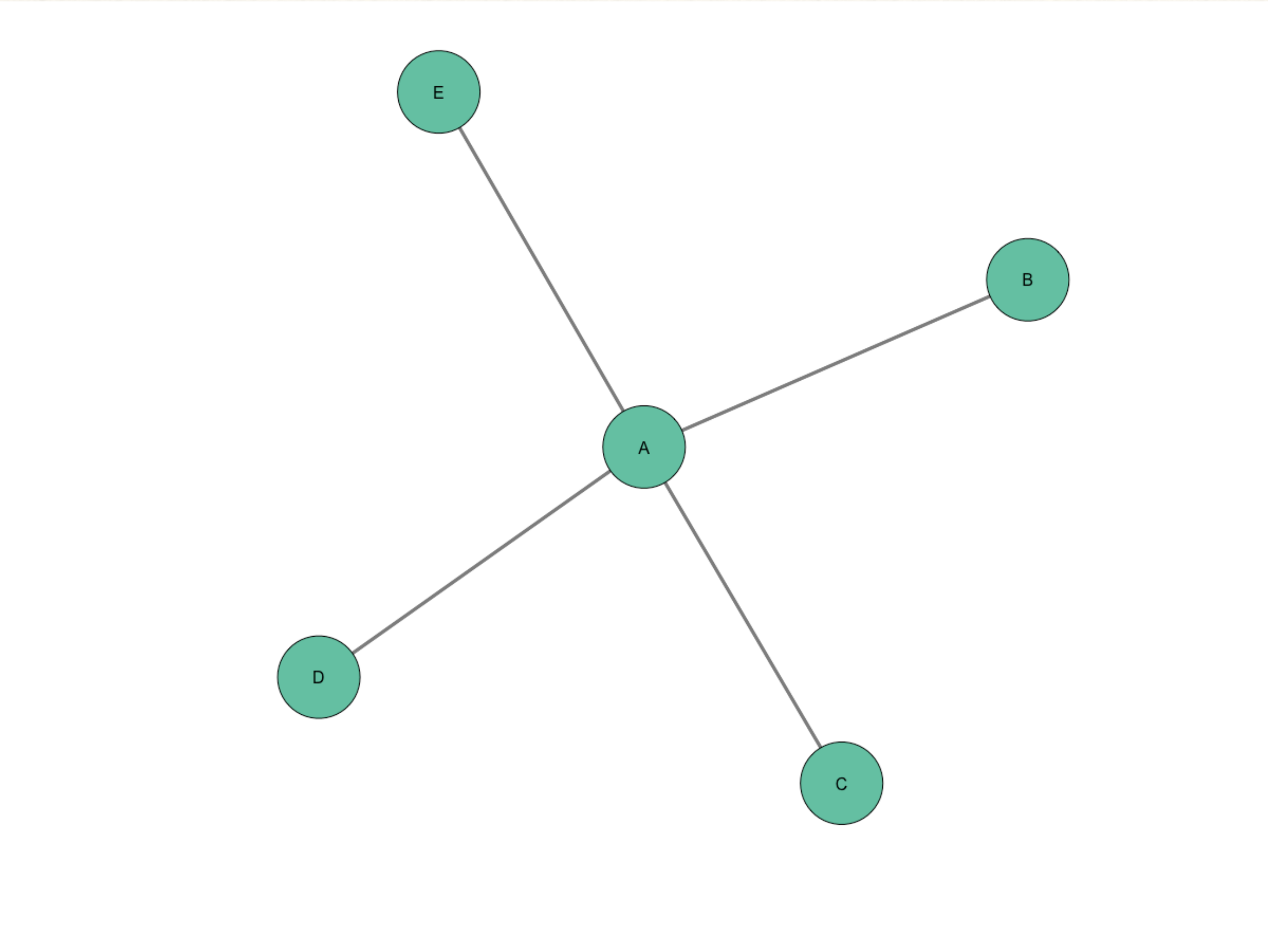
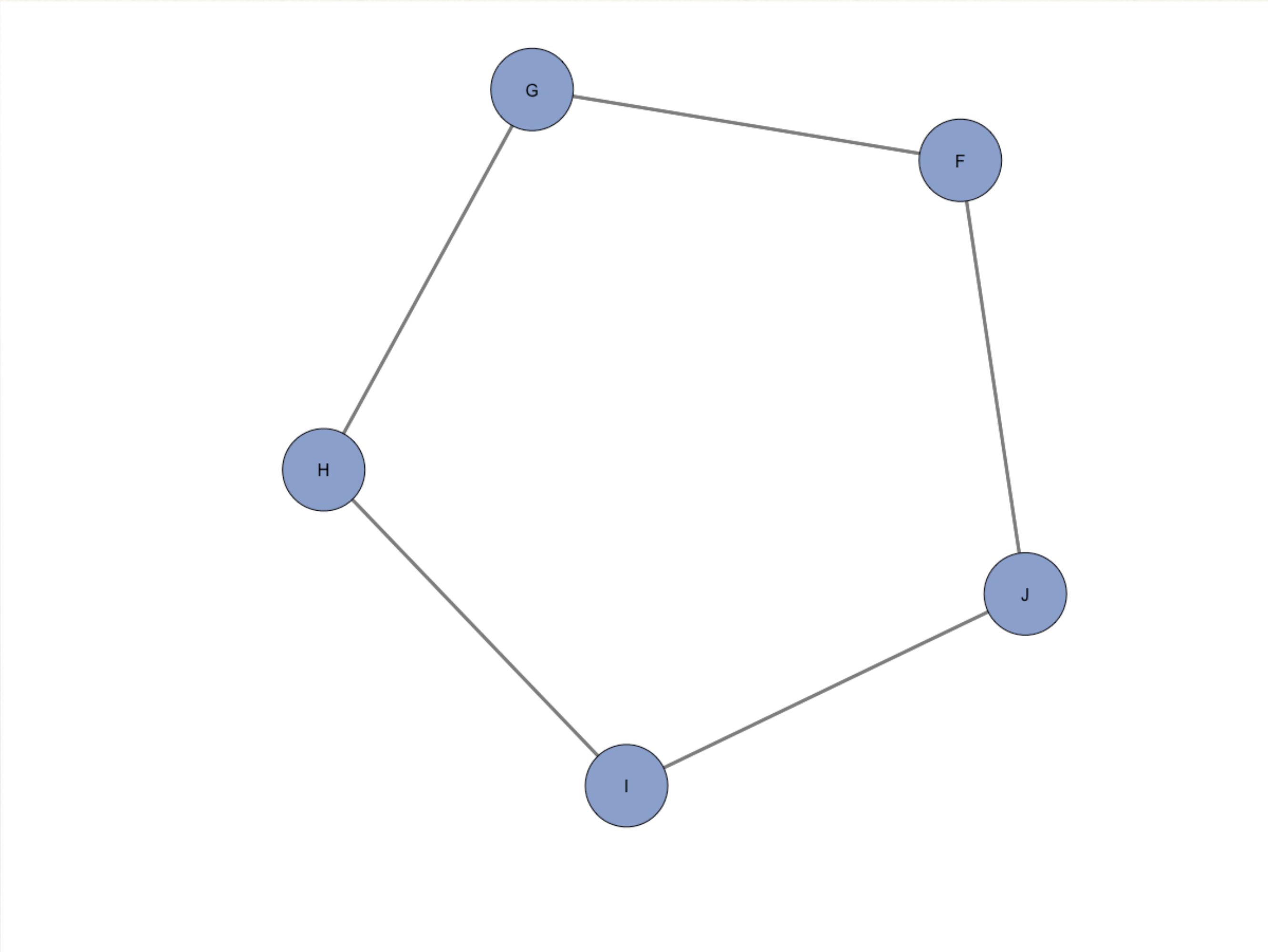


$$= \frac{(3-1)+(3-2)+(3-3)+(3-2)+(3-2)}{(5-1)(5-2)}$$

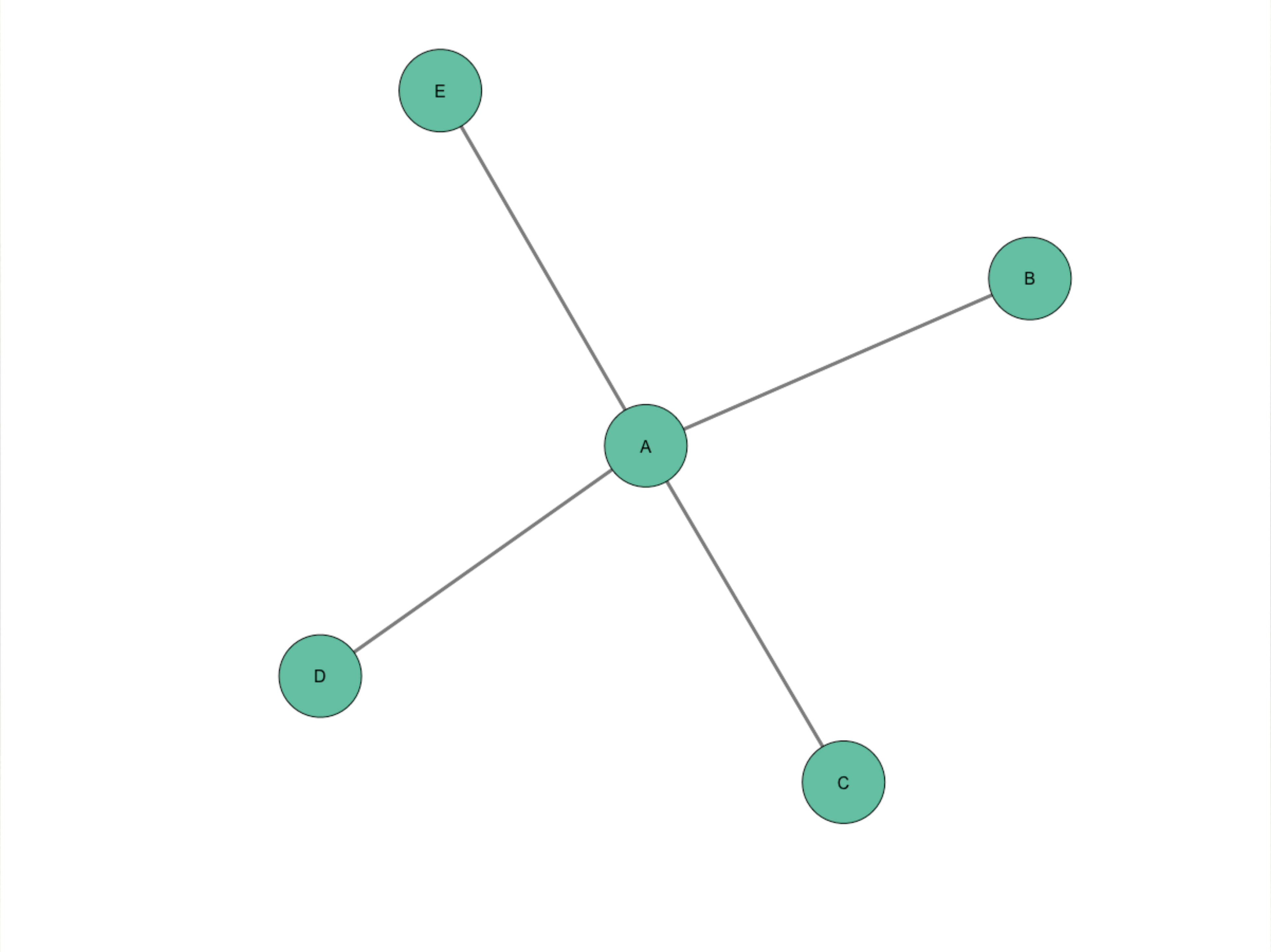
$$= \frac{2+1+0+1+1}{4*3} = \frac{5}{12} = 0.4167$$

Summarizing Degree Centrality

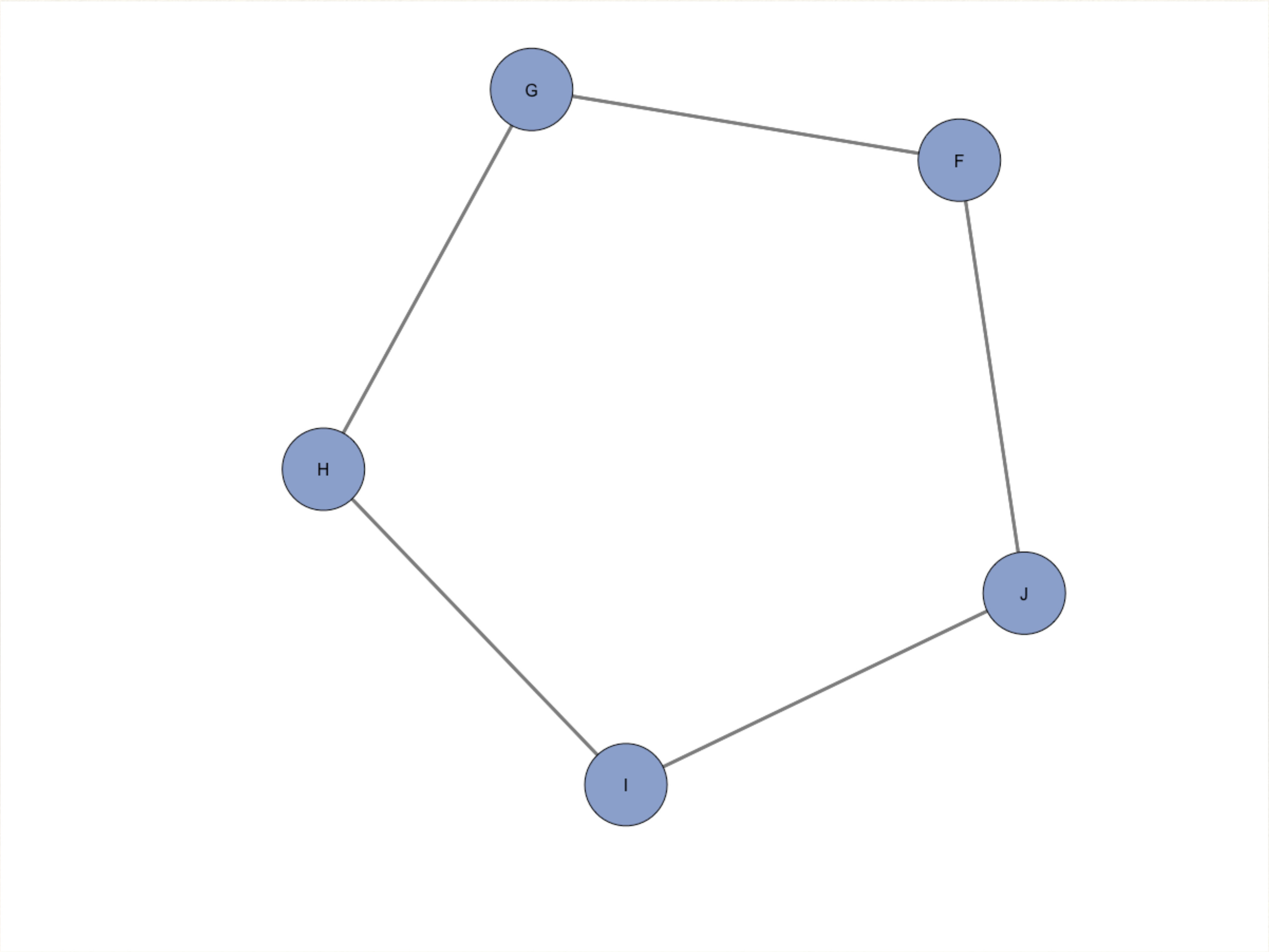
- When degree centrality is evenly dispersed, the numerator will be zero, and the quotient will be close to 0.
  - When there is considerable inequality in the actor degrees, the quotient will be closer to 1.
    - Thus, closer to 1 indicates that the graph is hierarchically structured.


E
B
A
D
C

G
F
H
J
I

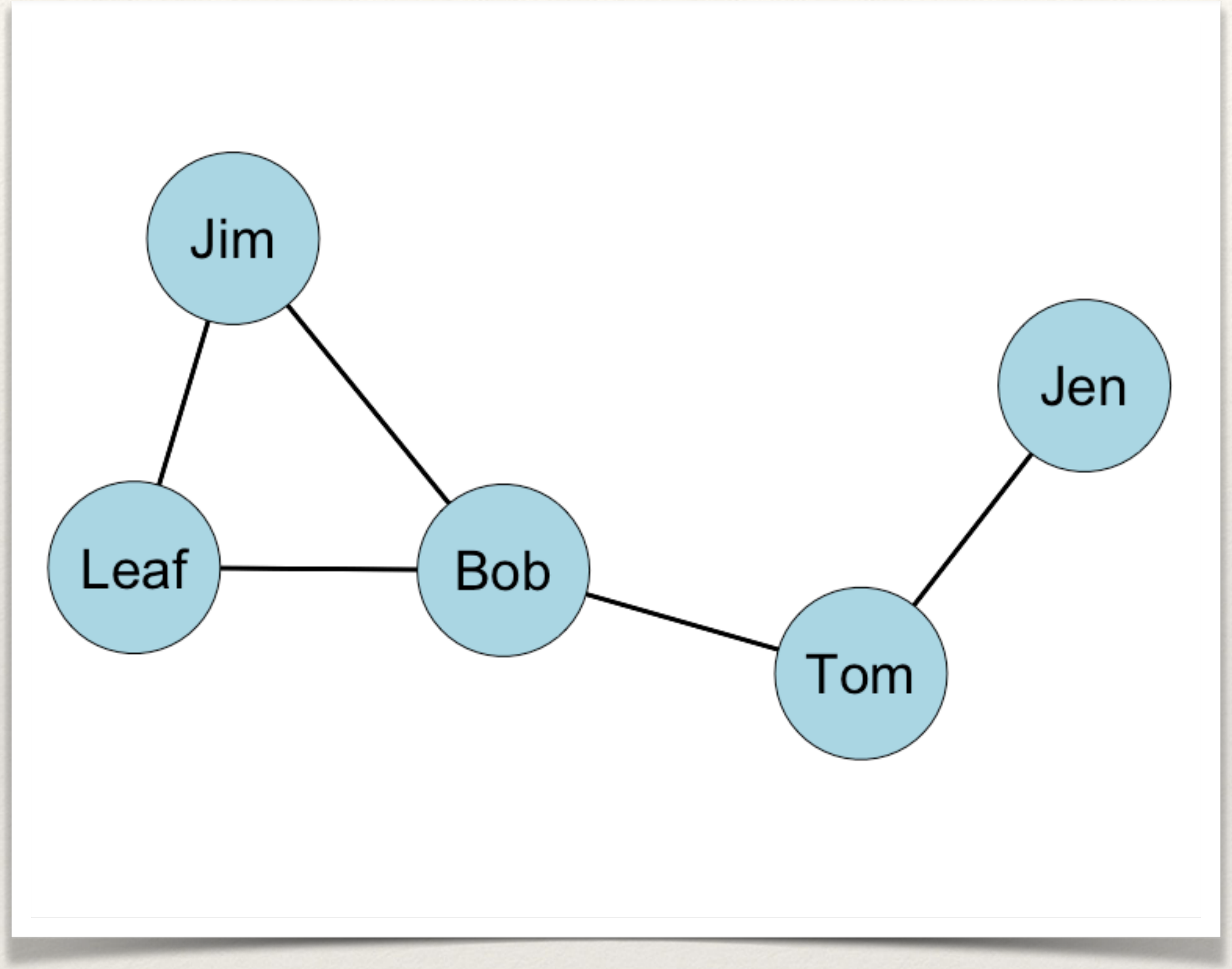
$$C_D = \frac{\sum_{i=1}^{g}[C_D(n^*) - C_D(n_i)]}{[(g-1)(g-2)]} = \frac{(4-4)+(4-1)+(4-1)+(4-1)+(4-1)}{(5-1)(5-2)} = \frac{0+3+3+3+3}{4*3} = \frac{12}{12} = 1.0$$



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



Example: Undirected, Binary Network



How should we interpret this value?

0.4167

Directed Networks

Degree Centrality: Directed Binary Graphs

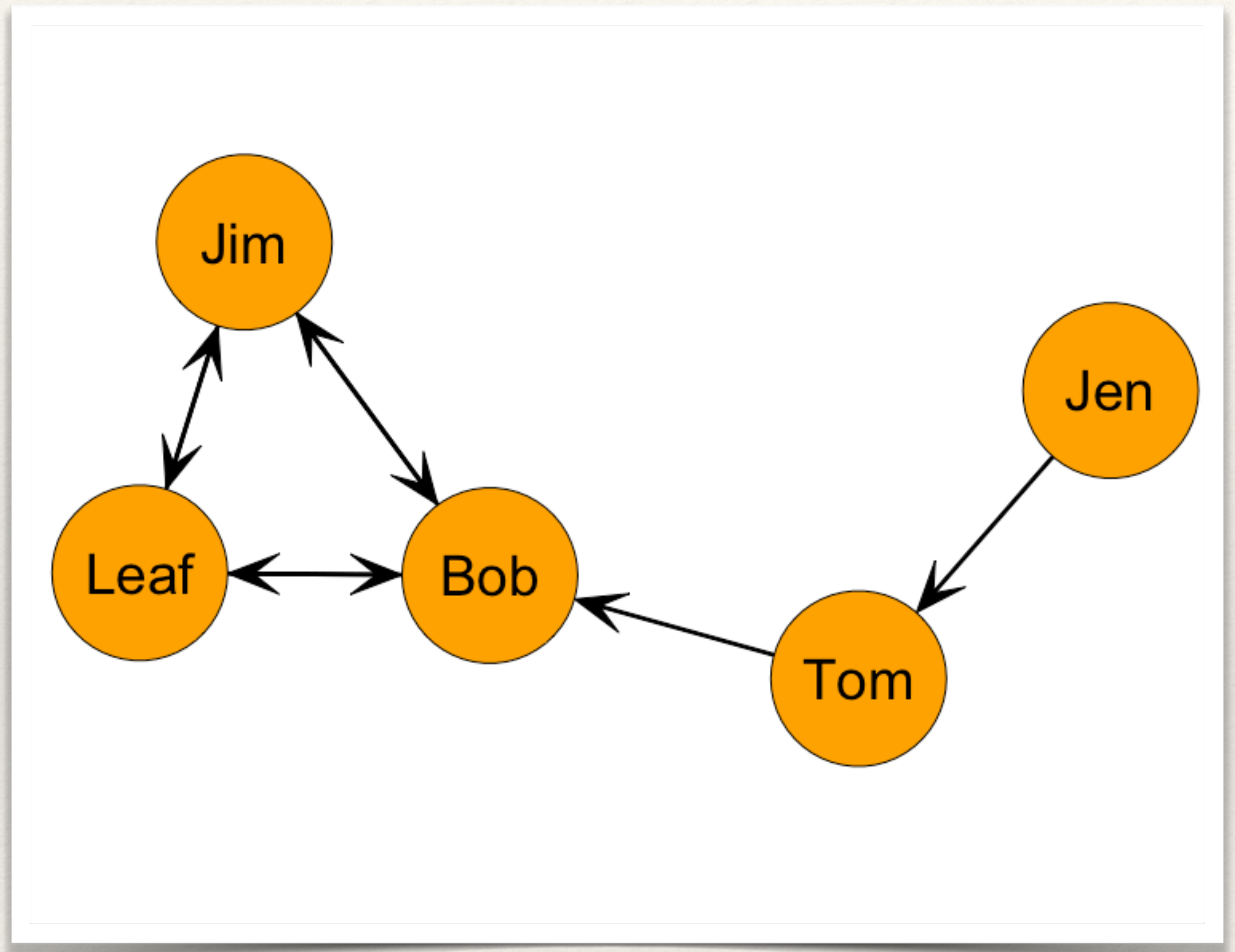
- In a directed binary graph, *actor degree centrality* can be broken down into indegree and outdegree centrality.
  - **Indegree**, $C_I(n_i)$, measures the number of ties that i receives.
    - For the sociomatrix $\boldsymbol{X}_{ij}$, the indegree for i is the column sum.
  - **Outdegree**, $C_O(n_i)$,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

NOTE: These both sum to the same value

Raw Indegree Centrality

Jen = 0
Tom = 1
Bob = 3
Leaf = 2
Jim = 2
TOTAL: 8

Raw Outdegree Centrality

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

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

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

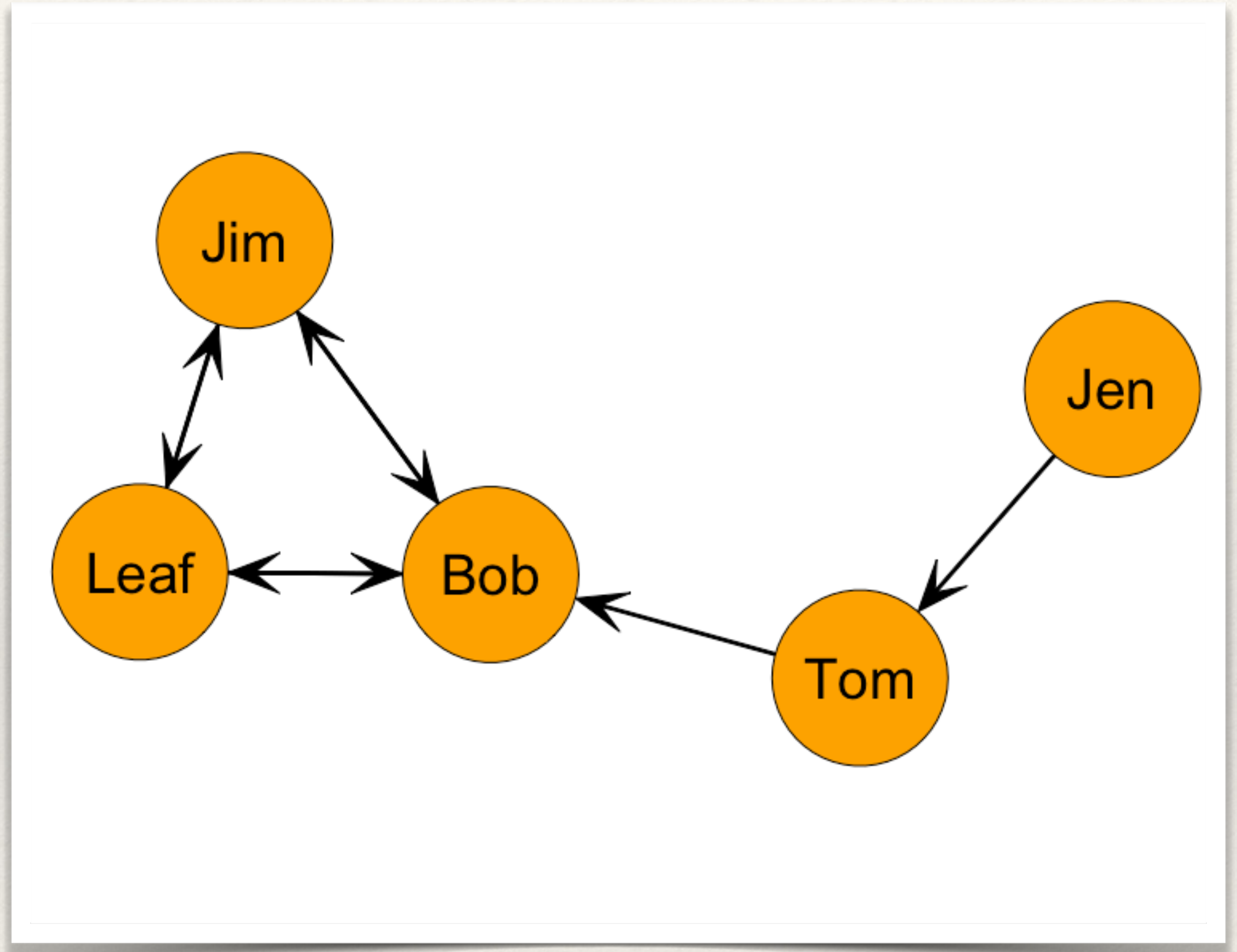
Standardized Degree Centrality: Directed Binary Graphs

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

Why do we subtract 1?

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

Example: Directed, Binary Network



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

Raw Indegree Centrality

Jen = 0

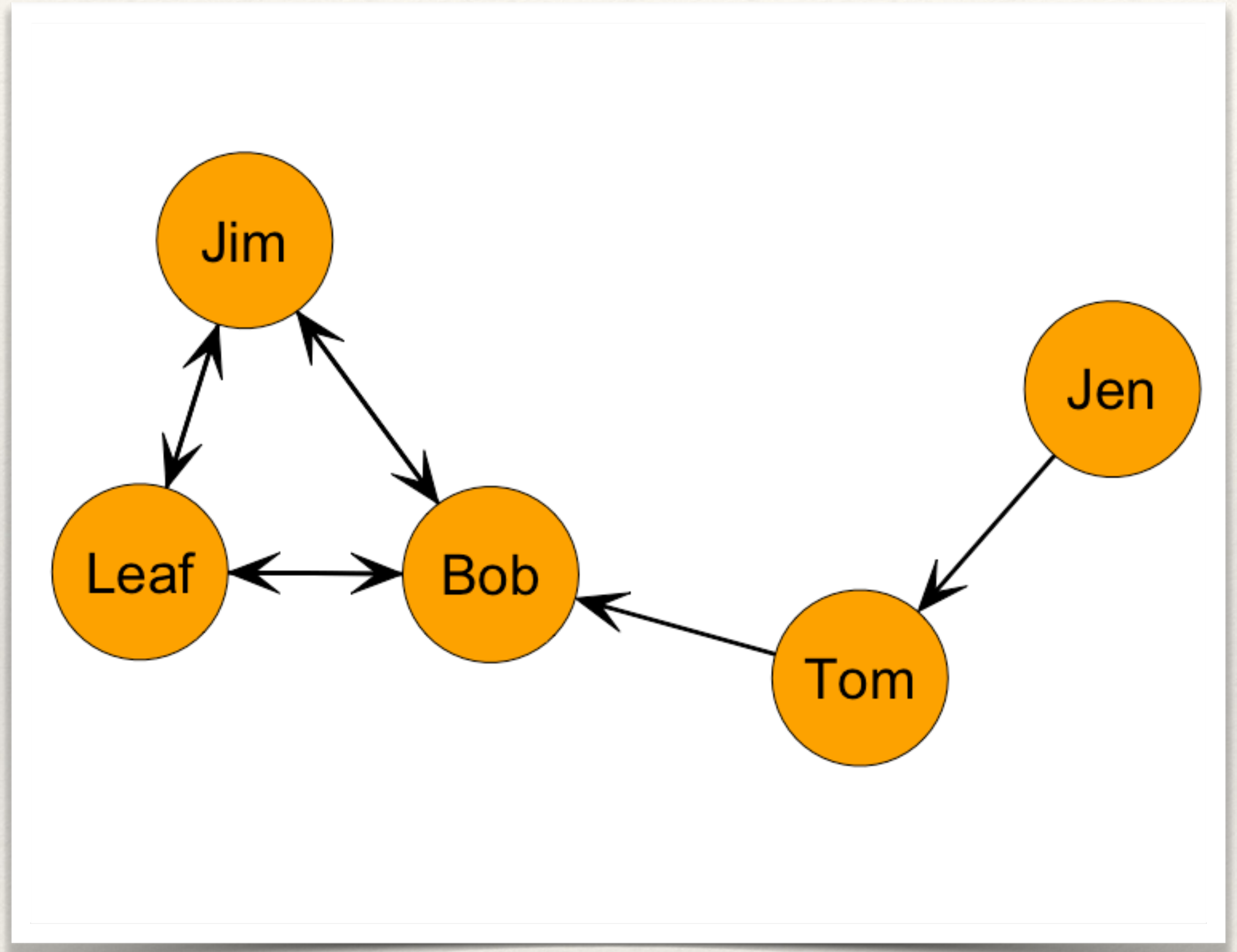
Tom = 1

Bob = 3

Leaf = 2

Jim = 2

Example: Directed, Binary Network



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

<u>Standardized Indegree Centrality</u>

Jen = 0/4 = 0

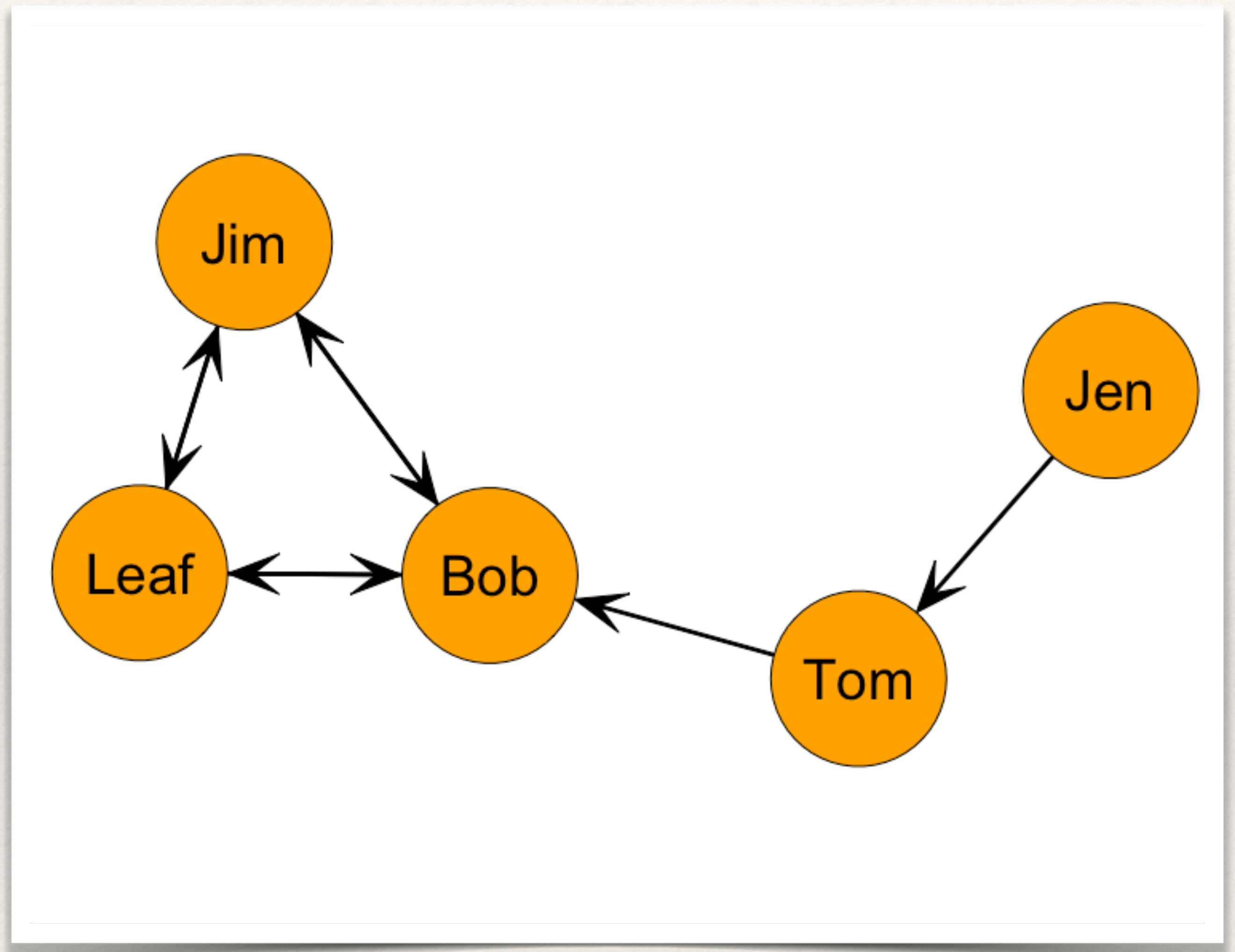
Tom = 1/4 =0.25

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?

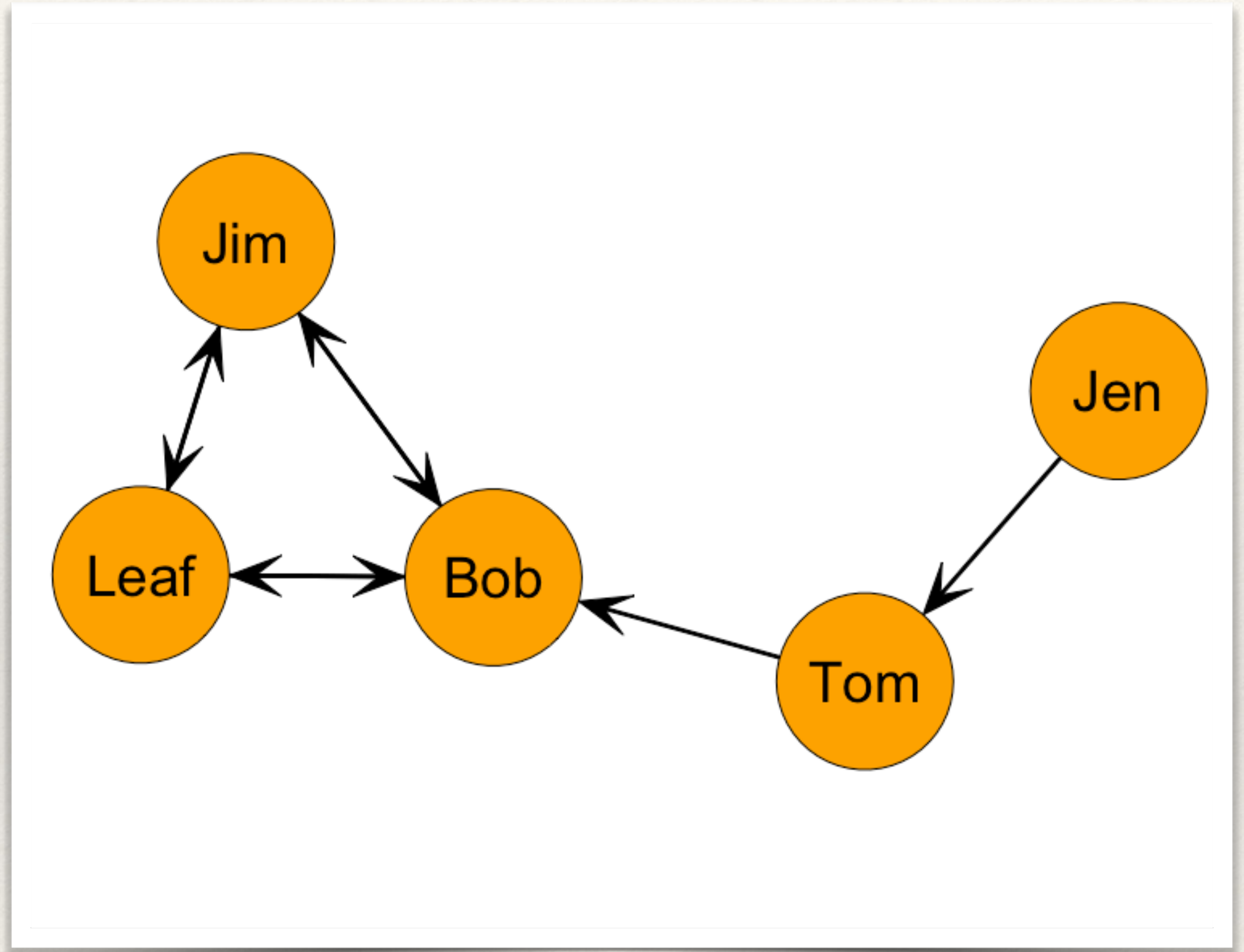
Raw Indegree Centrality

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

Raw Outdegree Centrality

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?

<u>Standardized Indegree Centrality</u>

Jen = 0/4 = 0

Tom = 1/4 =0.25

Bob = 3/4 = 0.75

Leaf = 2/4 = 0.50

Jim = 2/4 = 0.50

<u>Standardized Outdegree Centrality</u>

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

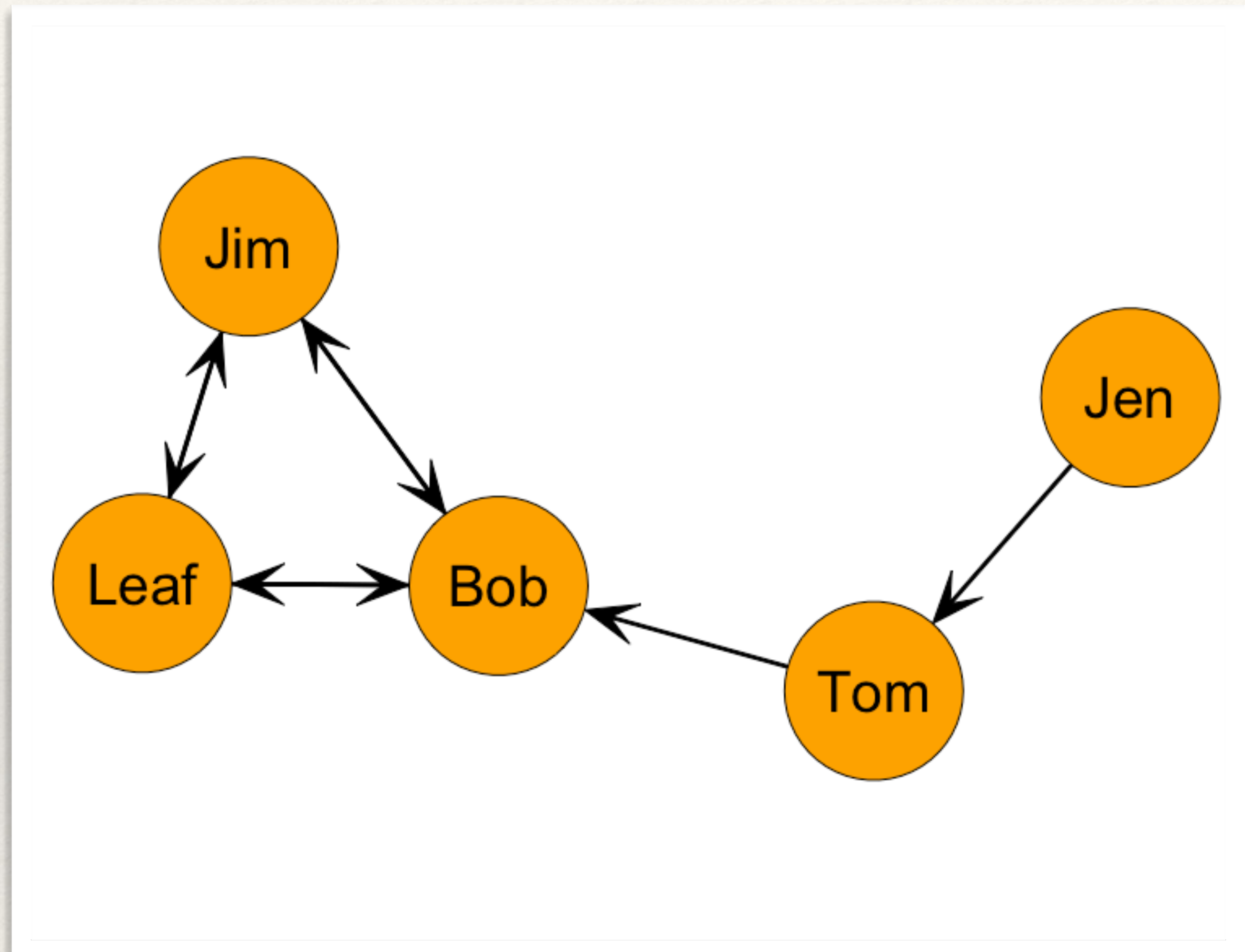
$$\bar{d} = \frac{\sum_{i=1}^{g} C_I(n_i)}{g} = \frac{\sum_{i=1}^{g} C_O(n_i)}{g} = \frac{L}{g}$$

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 for **indegree** and **outdegree**.

Index of Group Degree Centralization

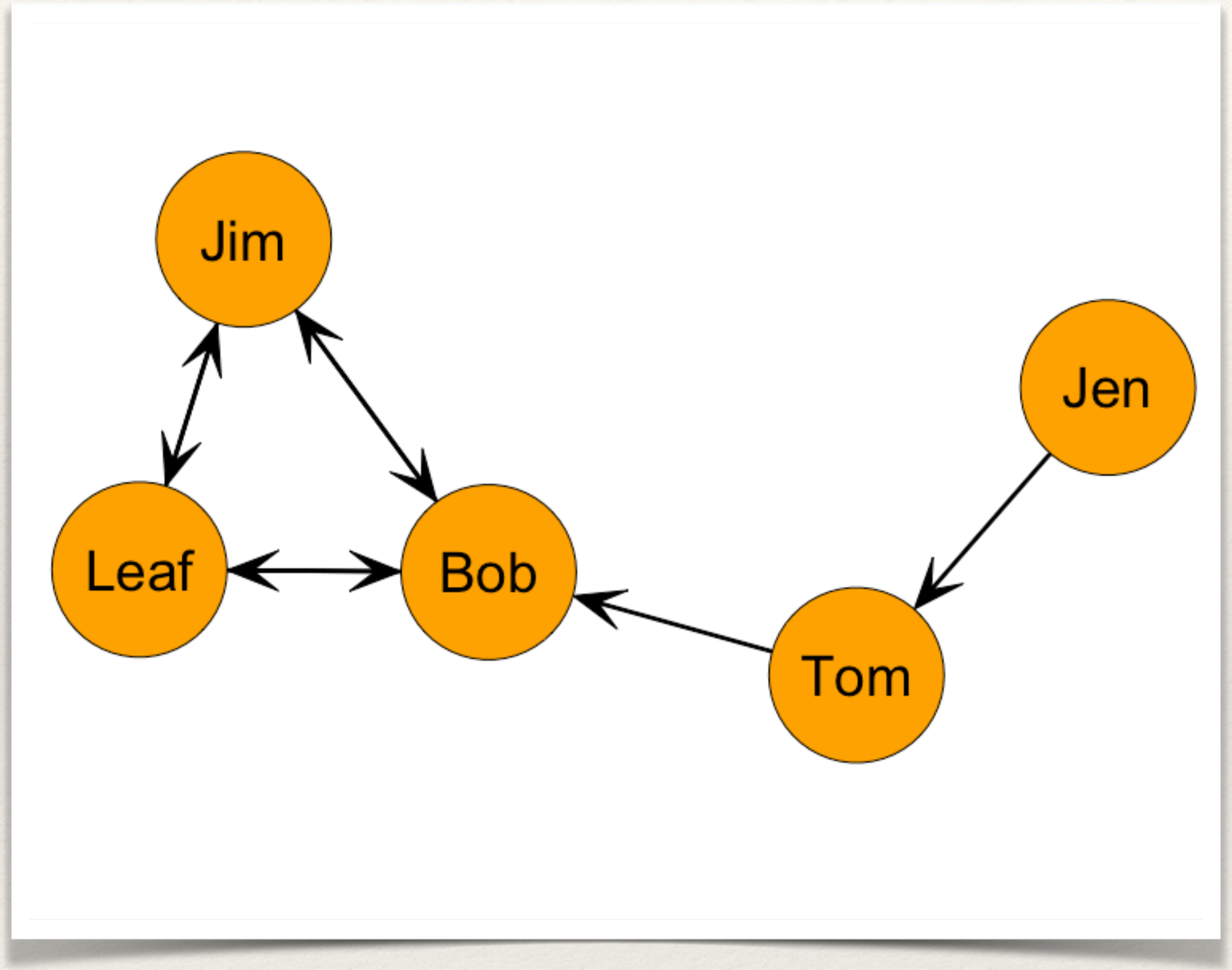
$$C_D = \frac{\sum_{i=1}^{g} [C_D(n^*) - C_D(n_i)]}{[(g-1)^2]}$$

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



Raw Indegree Centrality

Jen = 0

Tom = 1

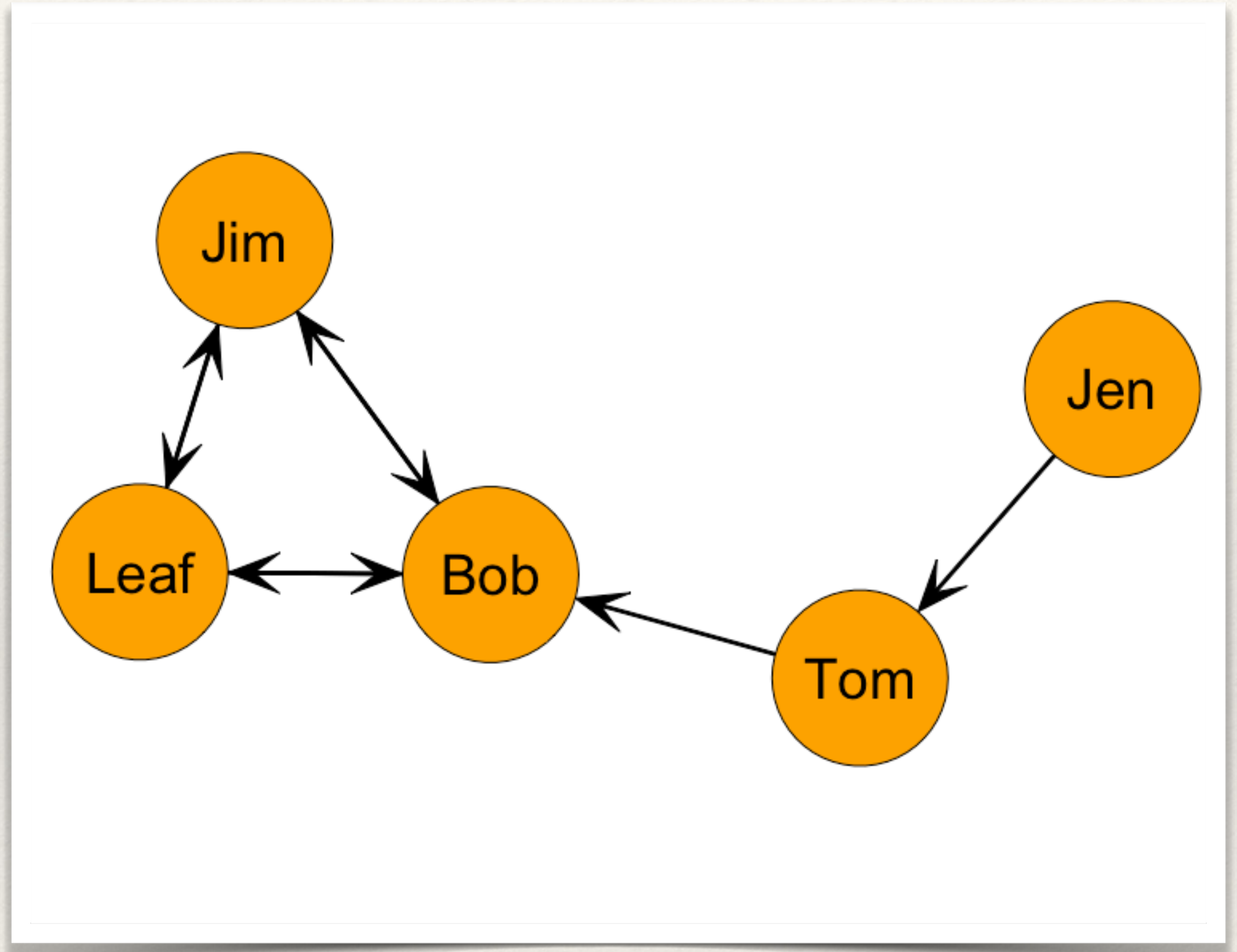
Bob = 3

Leaf = 2

Jim = 2

What is the index of indegree centralization for this graph?

Example: Directed, Binary Network



<u>Raw Indegree Centrality</u>

Jen = 0

Tom = 1

Bob = 3

Leaf = 2

Jim = 2

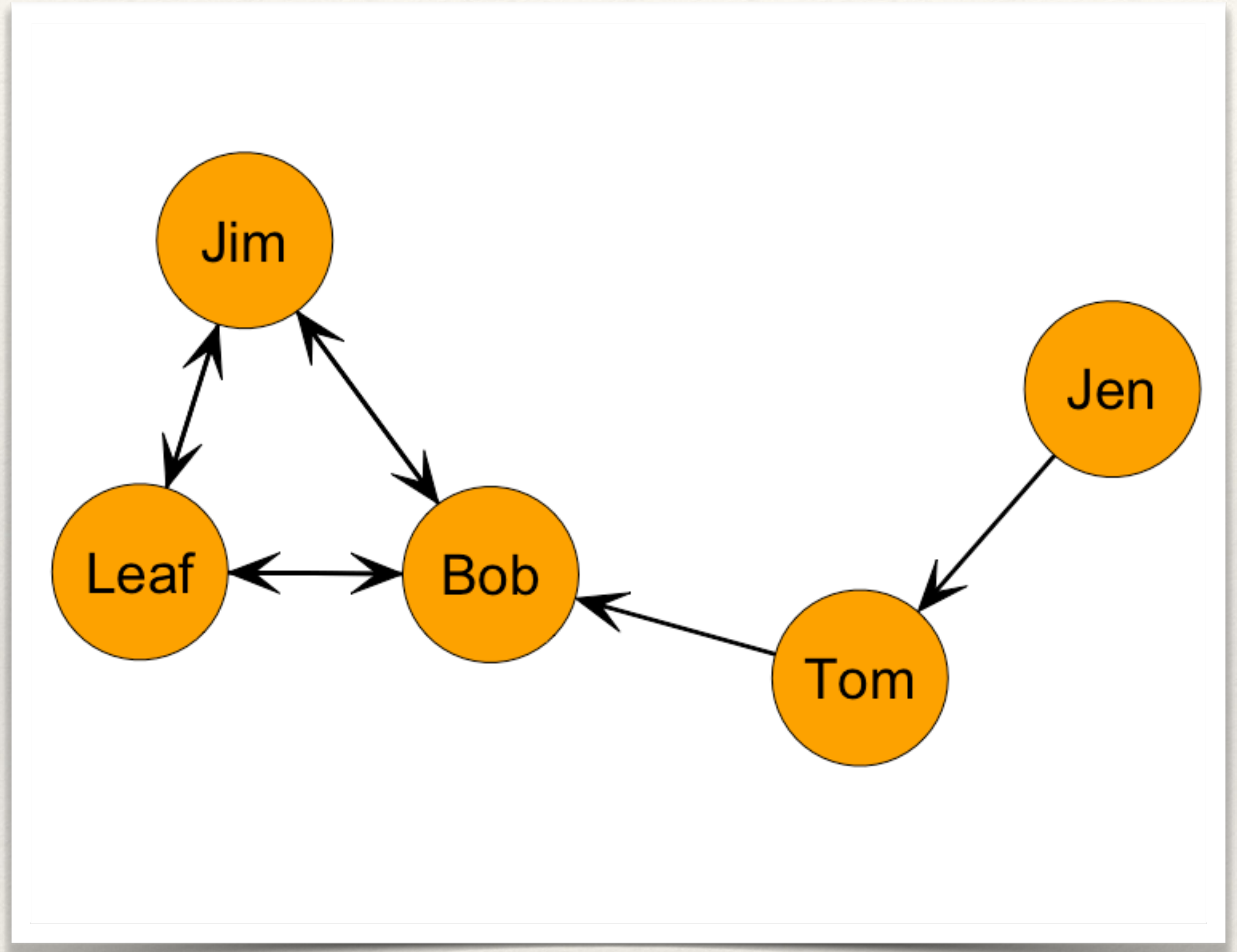
What is the index of indegree centralization for this graph?

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

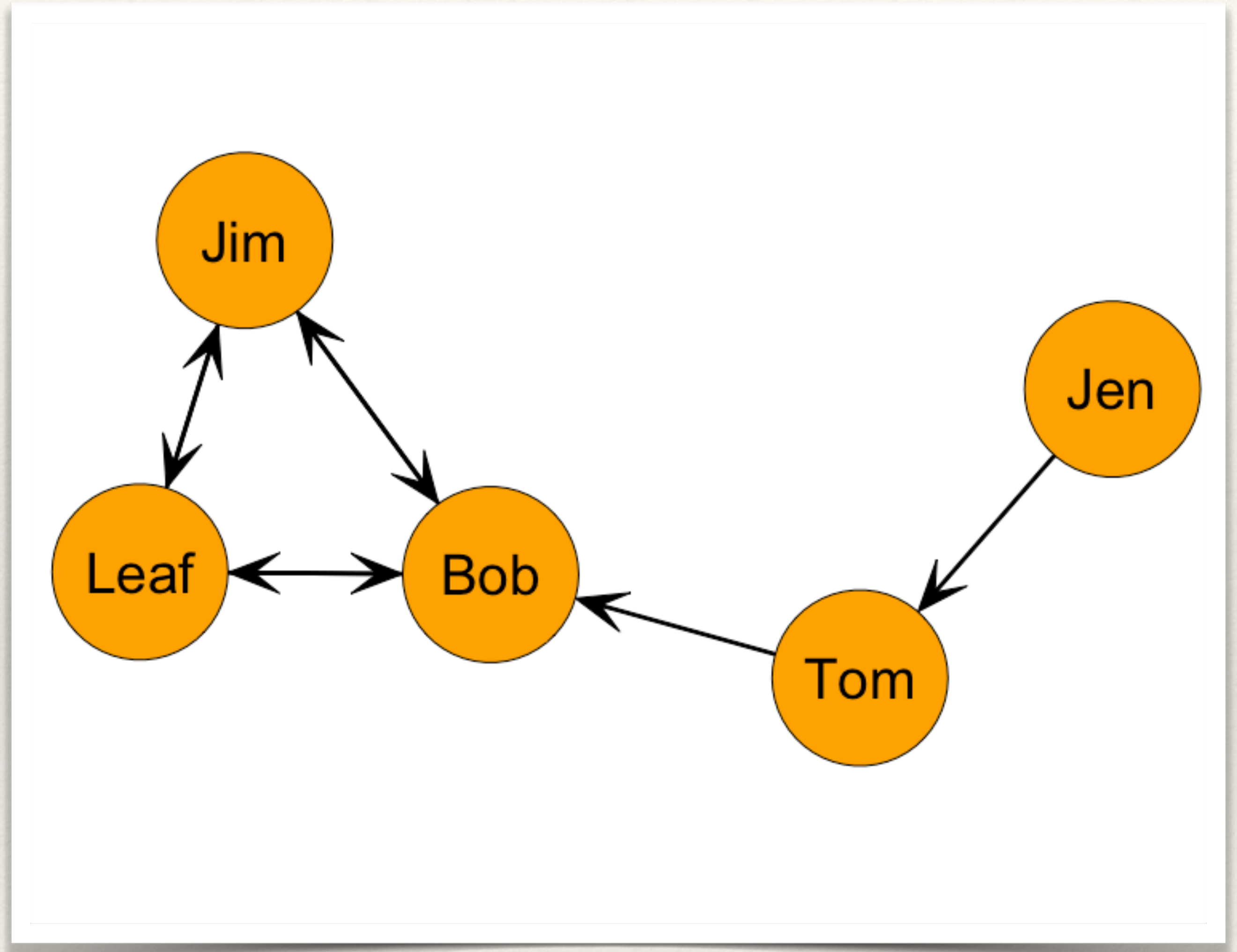
Bob = 2

Leaf = 2

Jim = 2

What is the index of outdegree centralization for this graph?

Example: Directed, Binary Network



<u>Raw Outdegree Centrality</u>

Jen = 1

Tom = 1

Bob = 2

Leaf = 2

Jim = 2

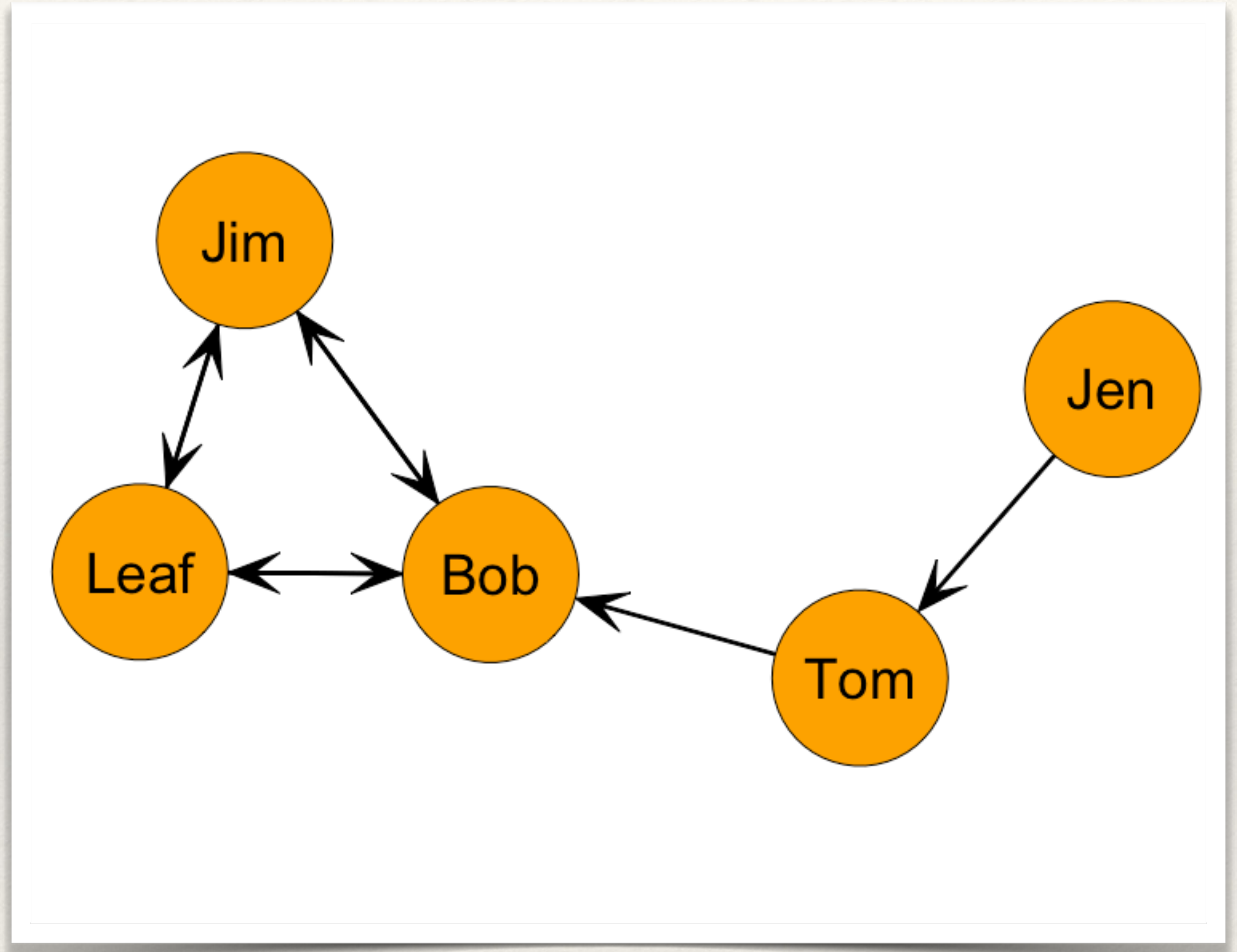
What is the index of outdegree centralization for this graph?

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

- At the end of the lecture, you should be able to answer these questions:
  - How can we conceptualize “centrality”.
  - How can we operationalization centrality as “degree”.
  - How do you calculate degree centrality for undirected and directed graphs?
  - What are the descriptive properties of degree centrality?

Questions?