Statistical Analysis of Networks

Network Data Structures

Learning Goals

- Understand various approaches to collecting network data.
- Understand the representation of networks using graph and graph notation.
- Understand the representation of undirected and directed networks using matrices.

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.
  - <u>Examples</u>:
    - Ego-centric networks
    - Partial networks
    - Complete (global) networks

Ego-Centric Networks

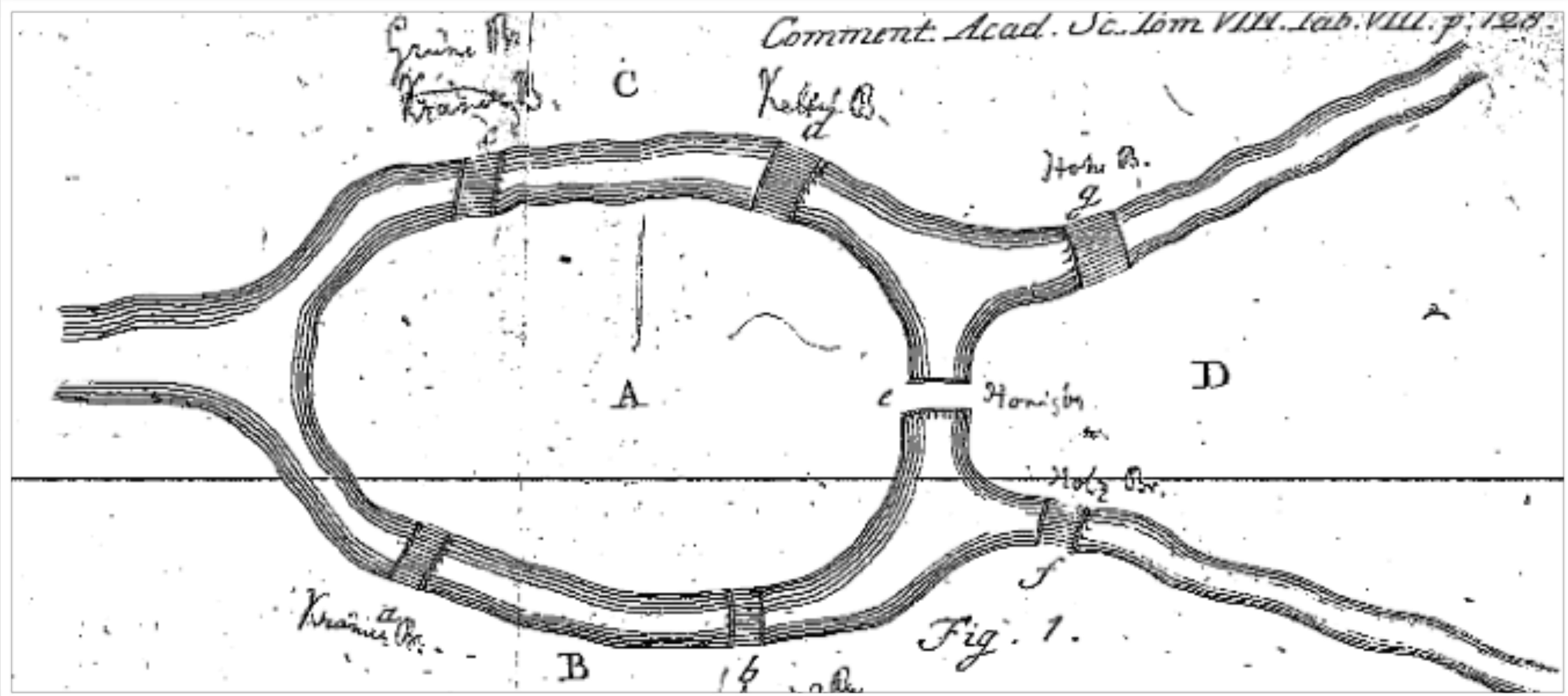
- Data on a focal actor (ego) and ties to neighbors (alters) and the ties among the alters.
  - *Instrument*: name generator
    - "who are the people with whom you discuss important matters?"
    - For each person named, "which of these individuals discuss important matters"?
    - <u>Why</u>?-costs,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

Partial Network

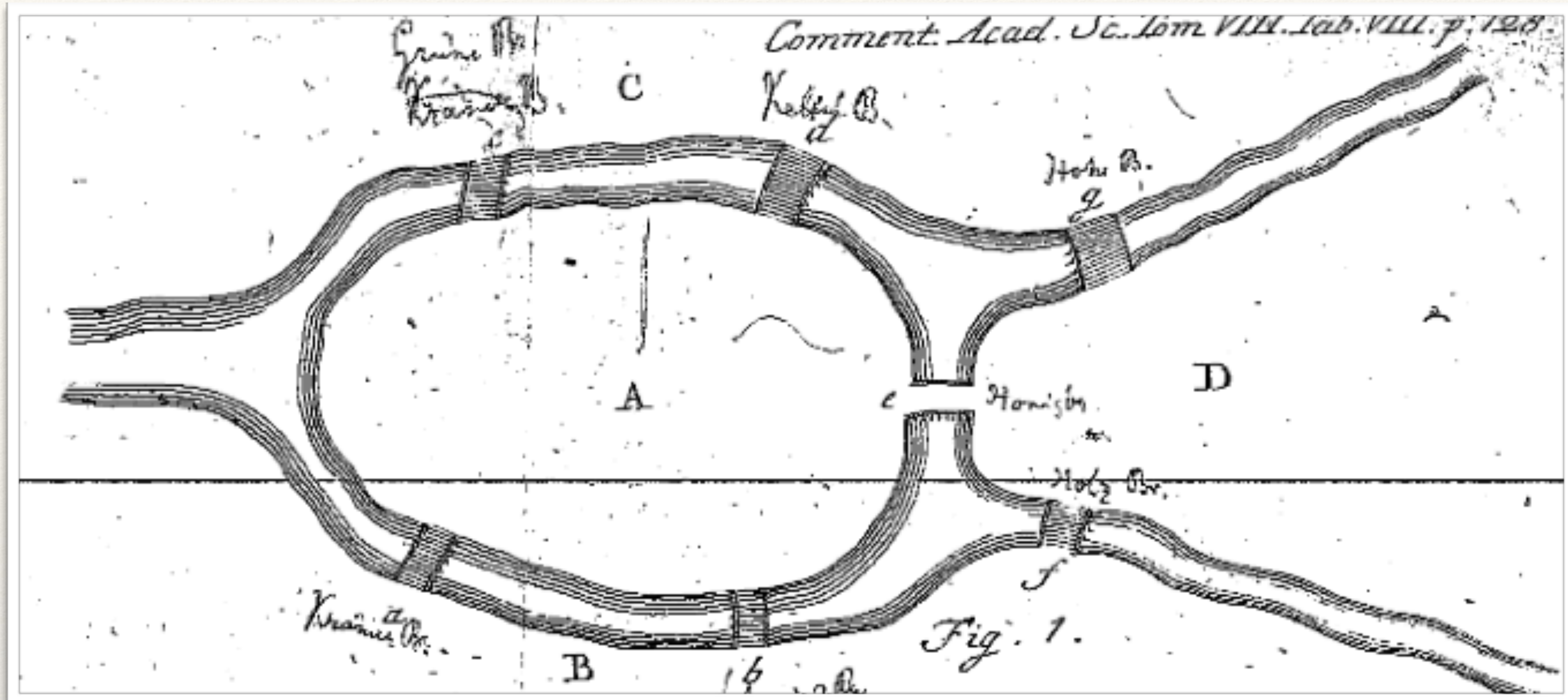
- Ego networks, plus some among of tracing to reach contacts of contacts.
  - *Instrument*: tracing mechanism
    - Using tickets to trace across a network
    - Why?-difficult to reach population, hard to specify sampling frame.

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?"

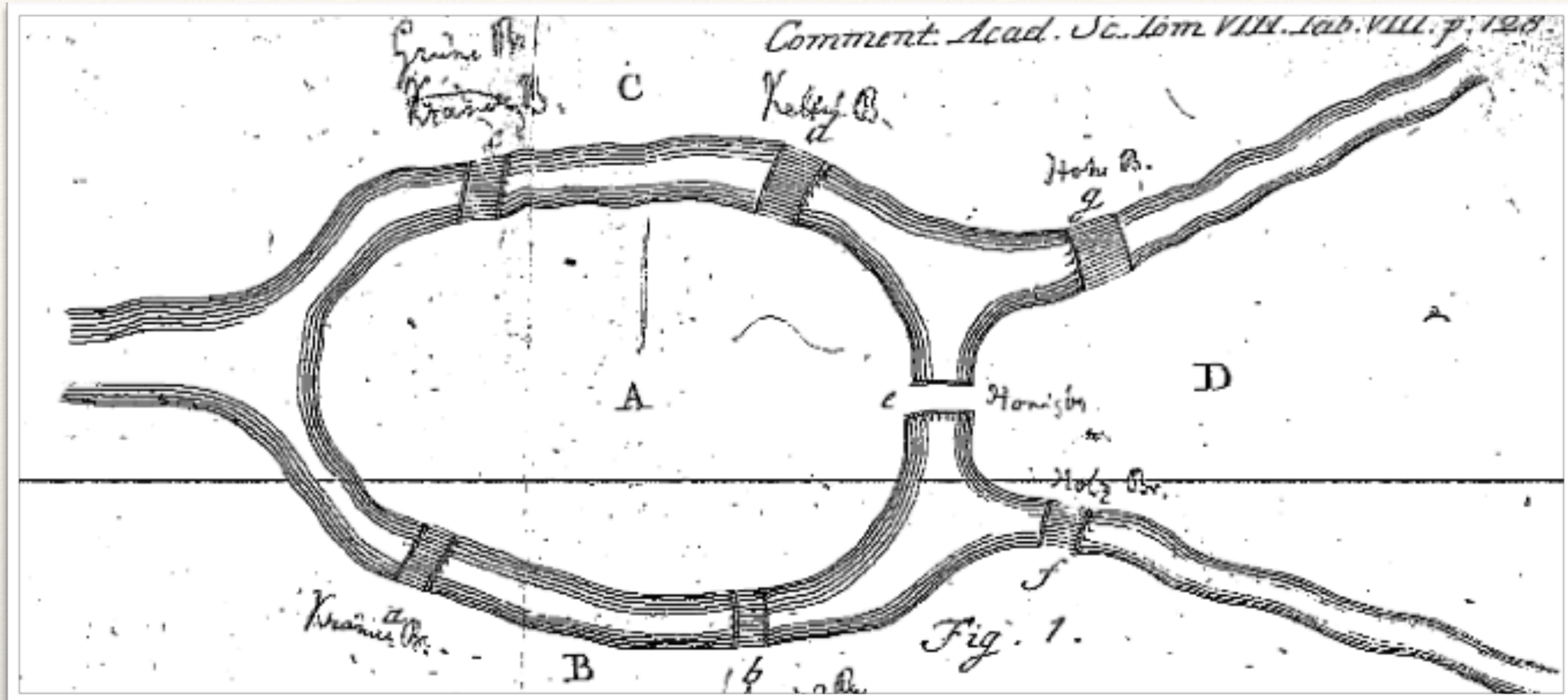


Here is a problem for you…



Konigsberg Bridge Problem

Devise a route in which you could cross all seven bridges.



Konigsberg Bridge Problem

Devise a route in which you could cross all seven bridges, **but** crossing each of the seven bridges only once.

Konigsberg Bridge Problem

- Leonard Euler worked on this problem in the mid 18th century, eventually representing the solution with a set of points and lines.
  - See: **https://www.youtube.com/watch?v=nZwSo4vfw6c**
    - Graph theory provides a foundation for operationalizing concepts of interest among network scientists.

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

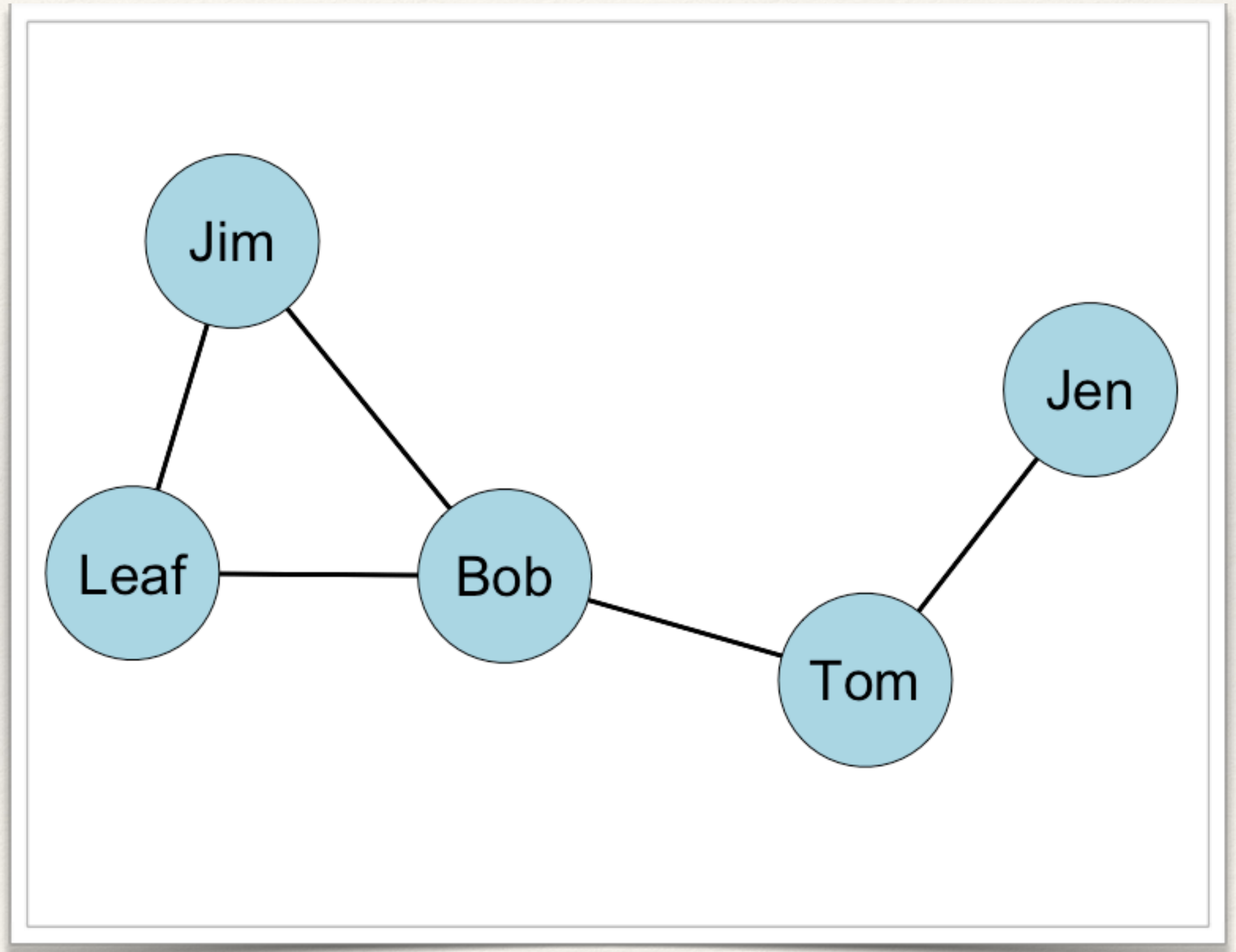
- Two nodes, n_i and n_j are **adjacent** if the line $l_k = (n_i, n_j)$
  - What this means is that in the graph, there exists a line between nodes i and j.

Example: Undirected, Binary Network

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

In other words,

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



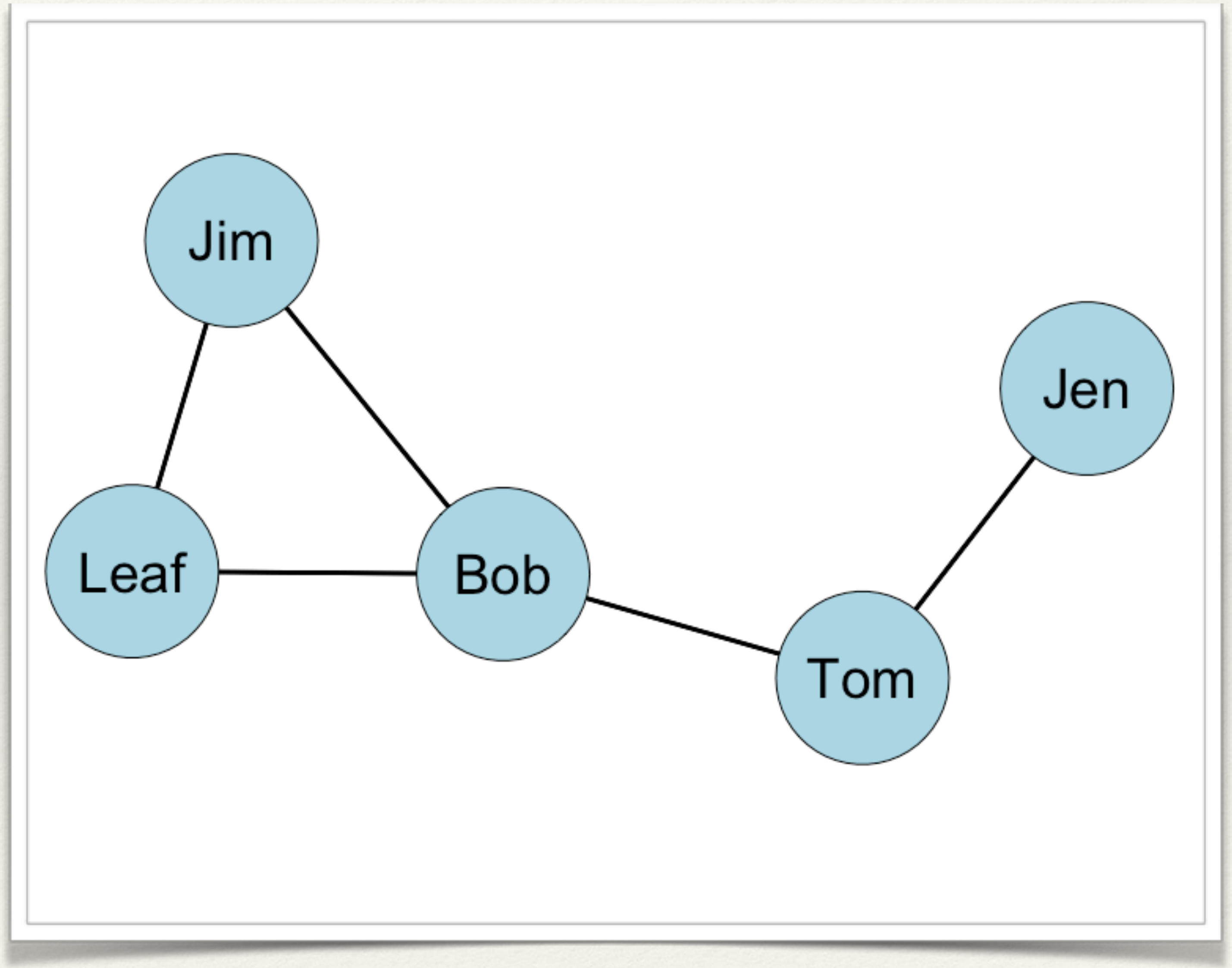
Example: Undirected, Binary Network

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

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

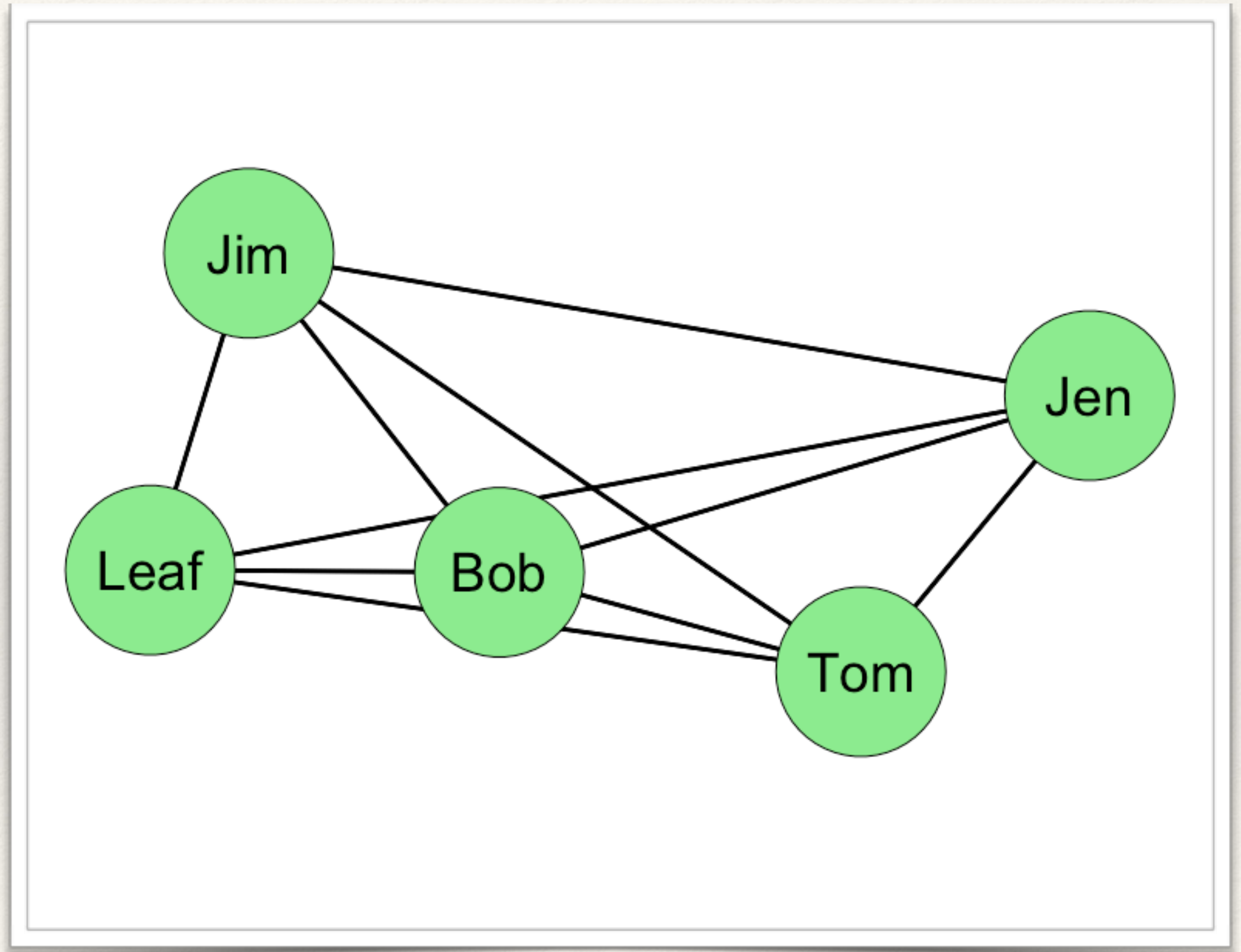
$g(g\text{-}1)/2$ possible ordered pairs.

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



Example: Undirected, Binary Network

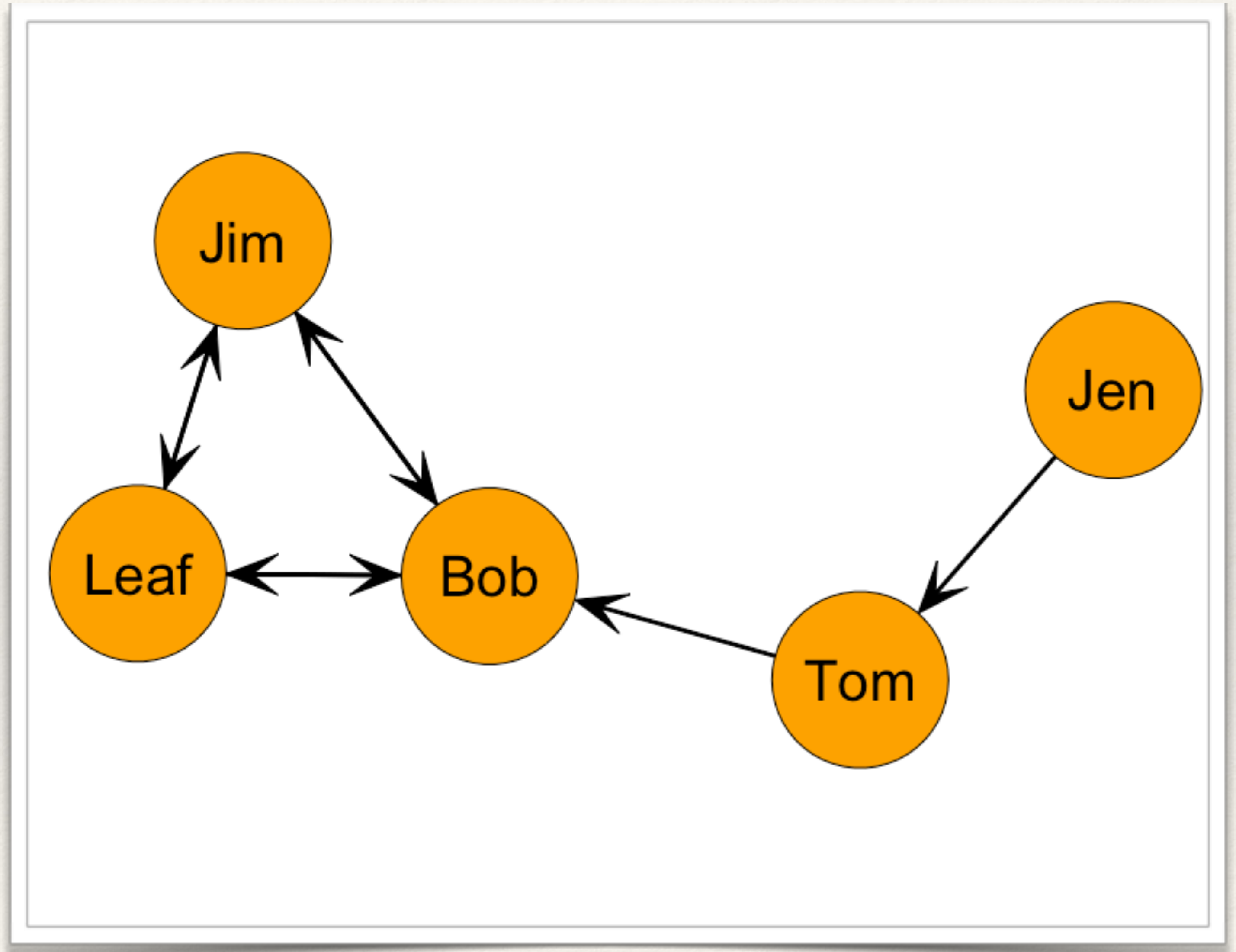
$g(g\text{-}1)/2 = 5(5\text{-}1)/2 = 10$



Example: Directed, Binary Network

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

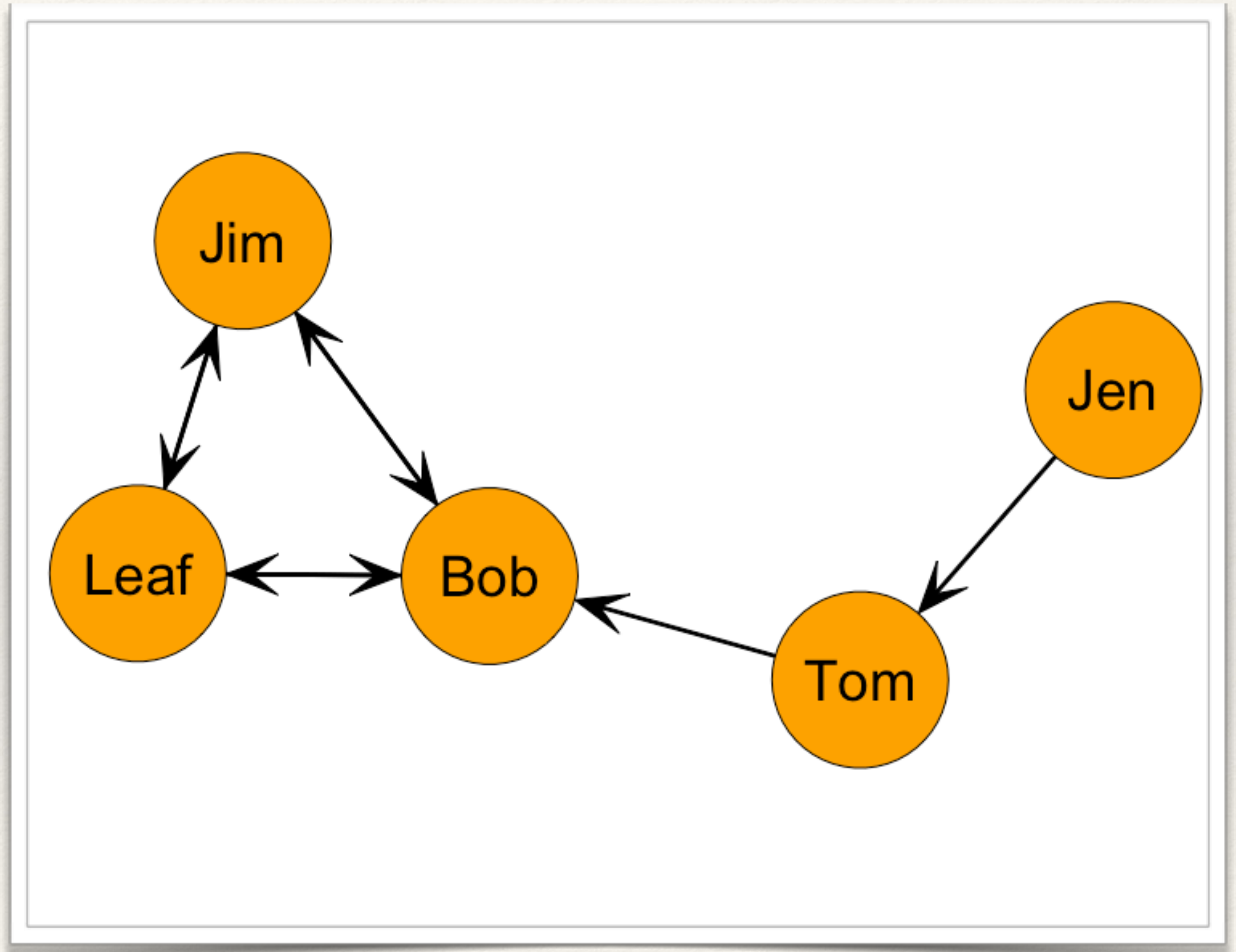
$$l_{k1} = (n_i, n_j) \neq (n_j, n_i) = l_{k2}$$



Example: Directed, Binary Network

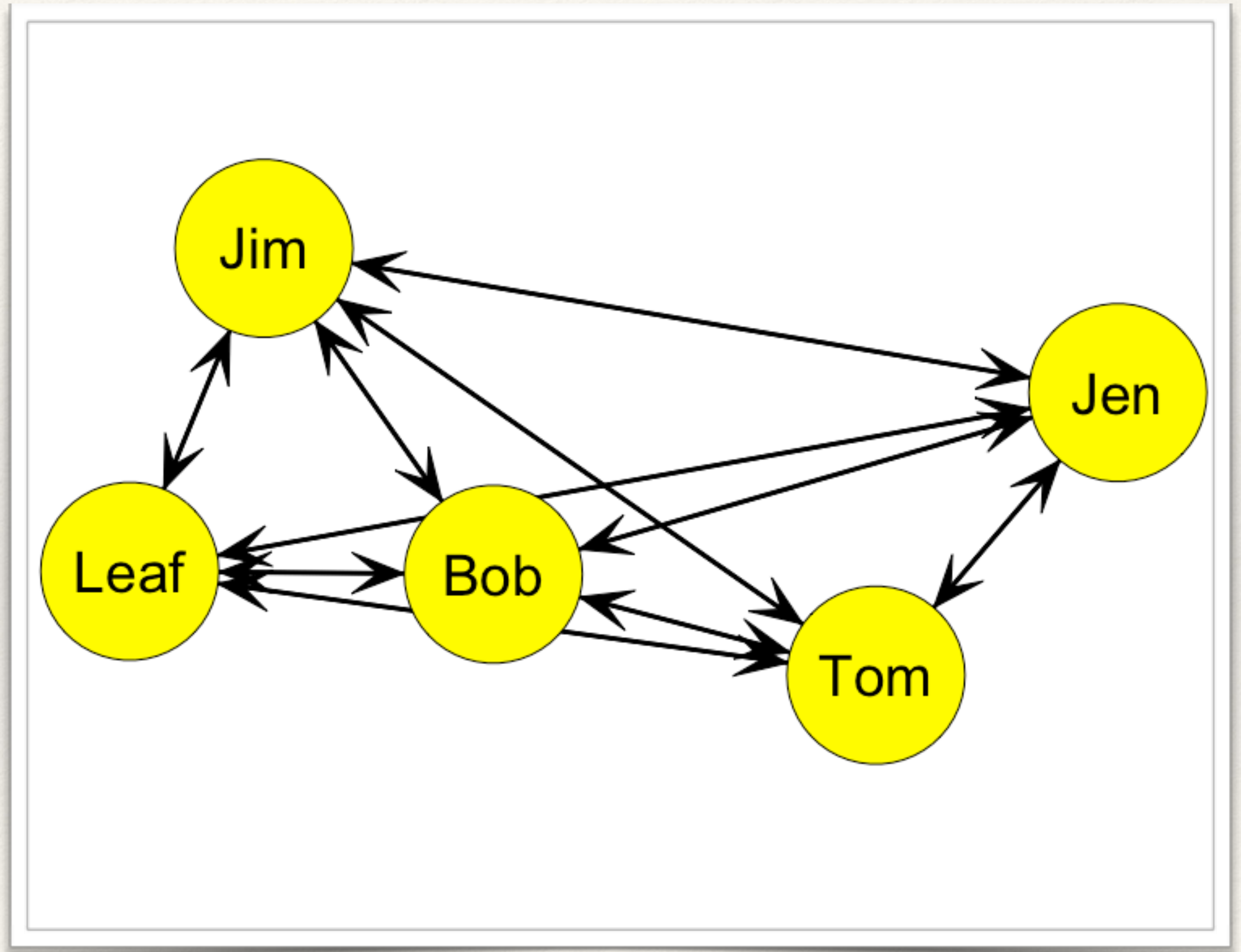
As a result, there are *g(g-1)* possible ordered pairs.

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



Example: Directed, Binary Network

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



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.

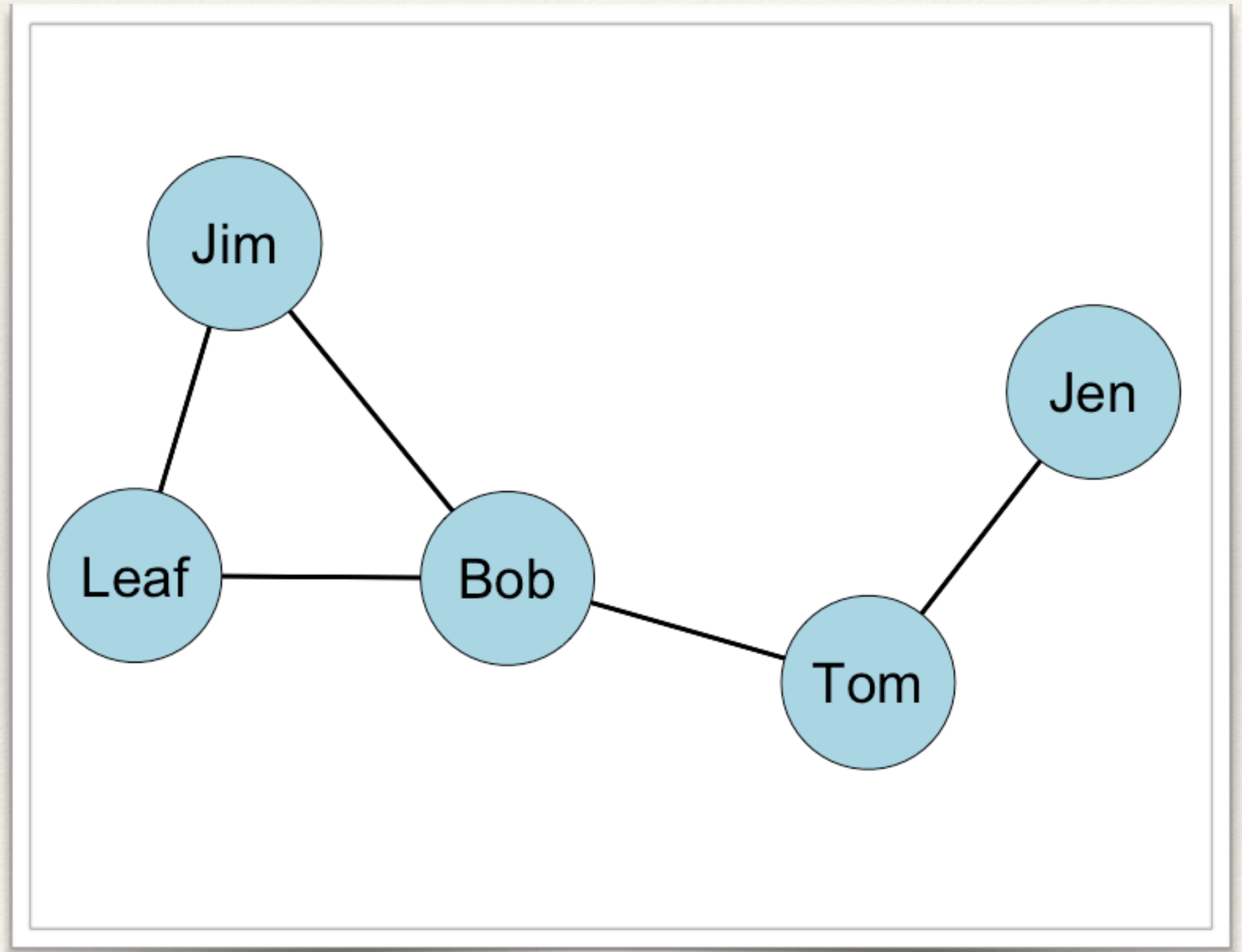
Sociometric Notation

- 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

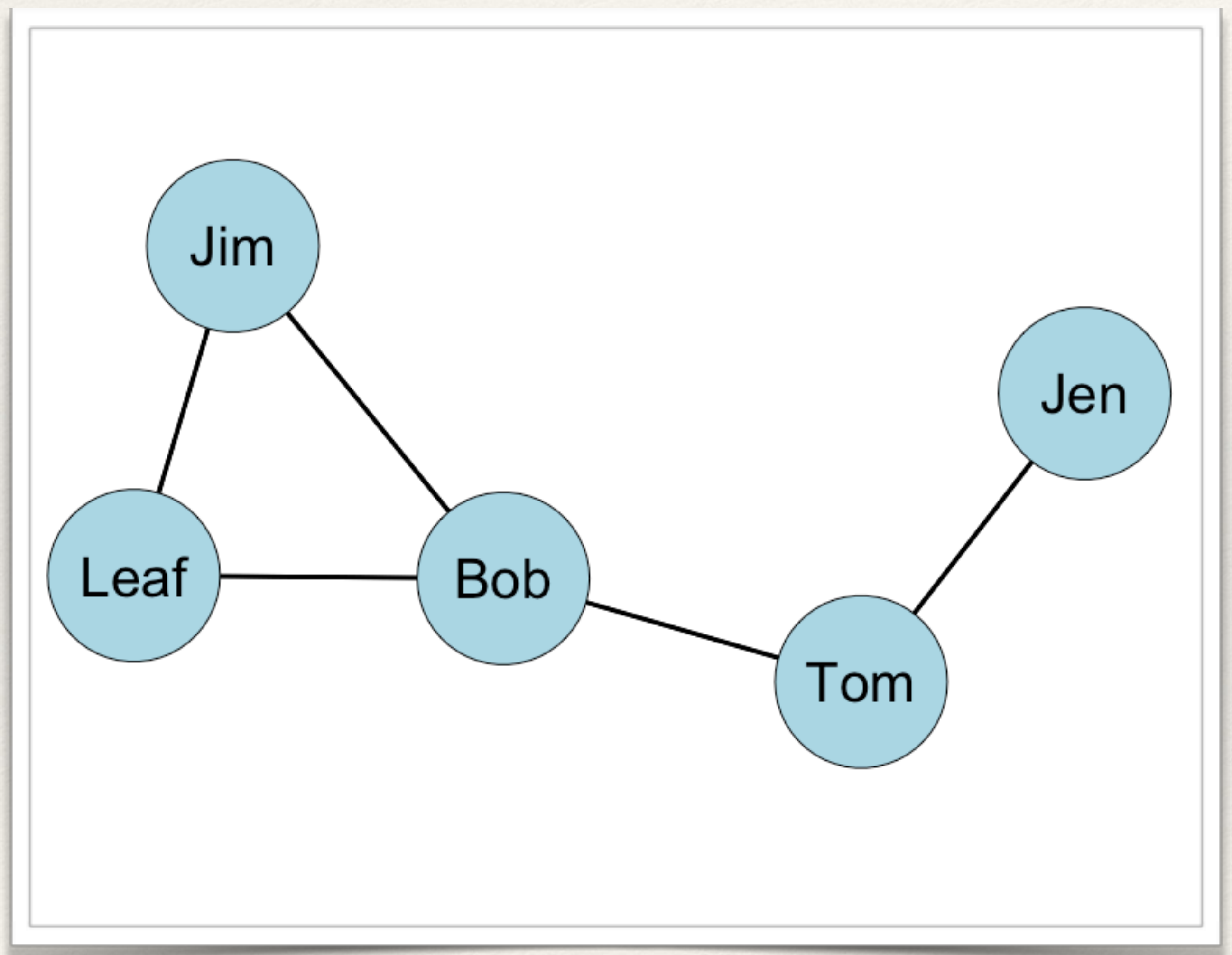
- **<u>Definitions</u>**
  - Scalar: a single number
  - Column vector: a column of numbers
  - Row vector: a row of numbers
  - Matrix: a rectangular arrow of numbers
  - Order: number of rows and columns defining the matrix
  - Square matrix: number of rows and columns of matrix are equal

Example: Undirected, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen					
Tom					
Bob					
Leaf					
Jim					

Example: Undirected, Binary Network

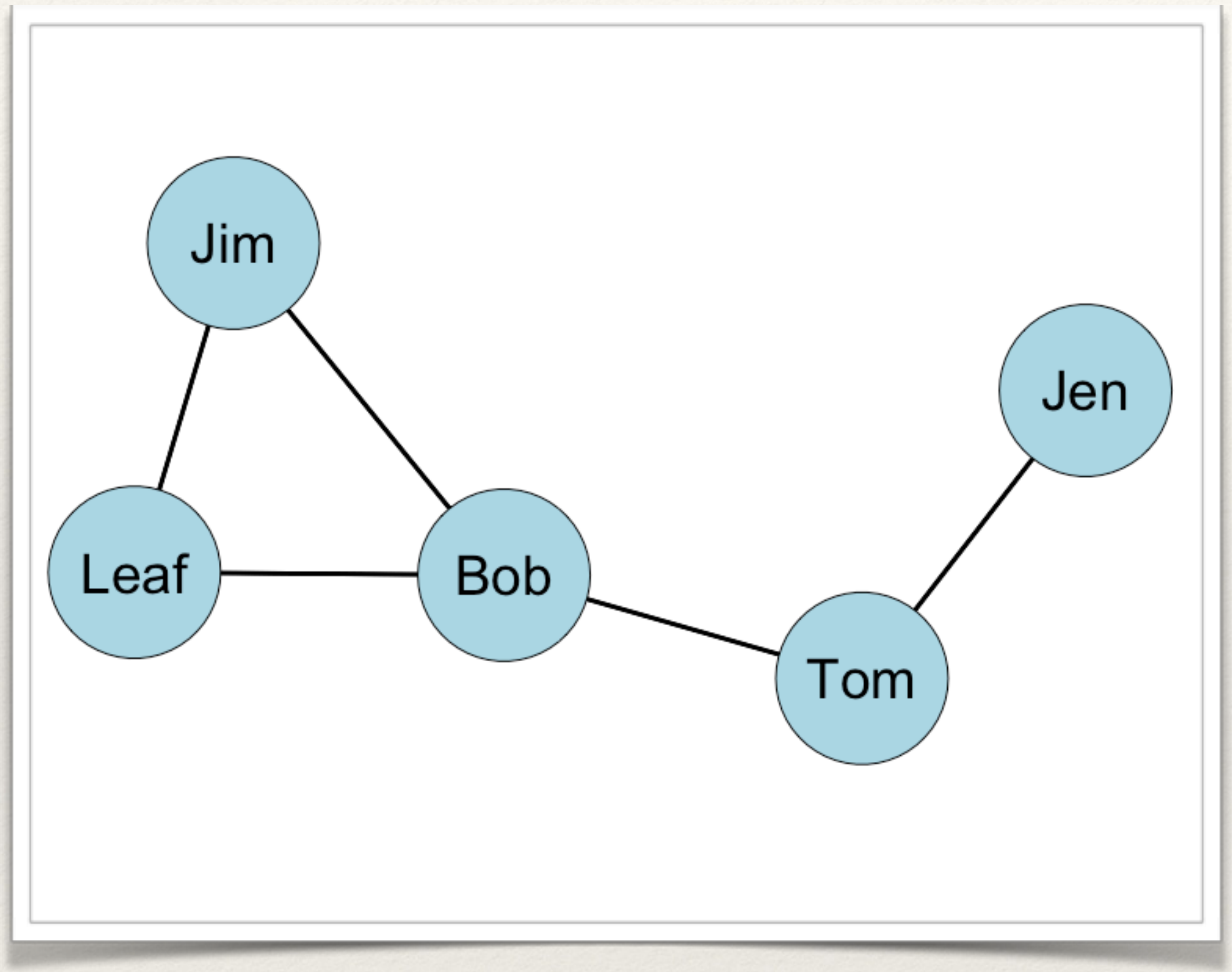


	Jen	Tom	Bob	Leaf	Jim
Jen					
Tom					
Bob					
Leaf					
Jim					

Graph or Sociogram

Adjacency Matrix or Sociomatrix

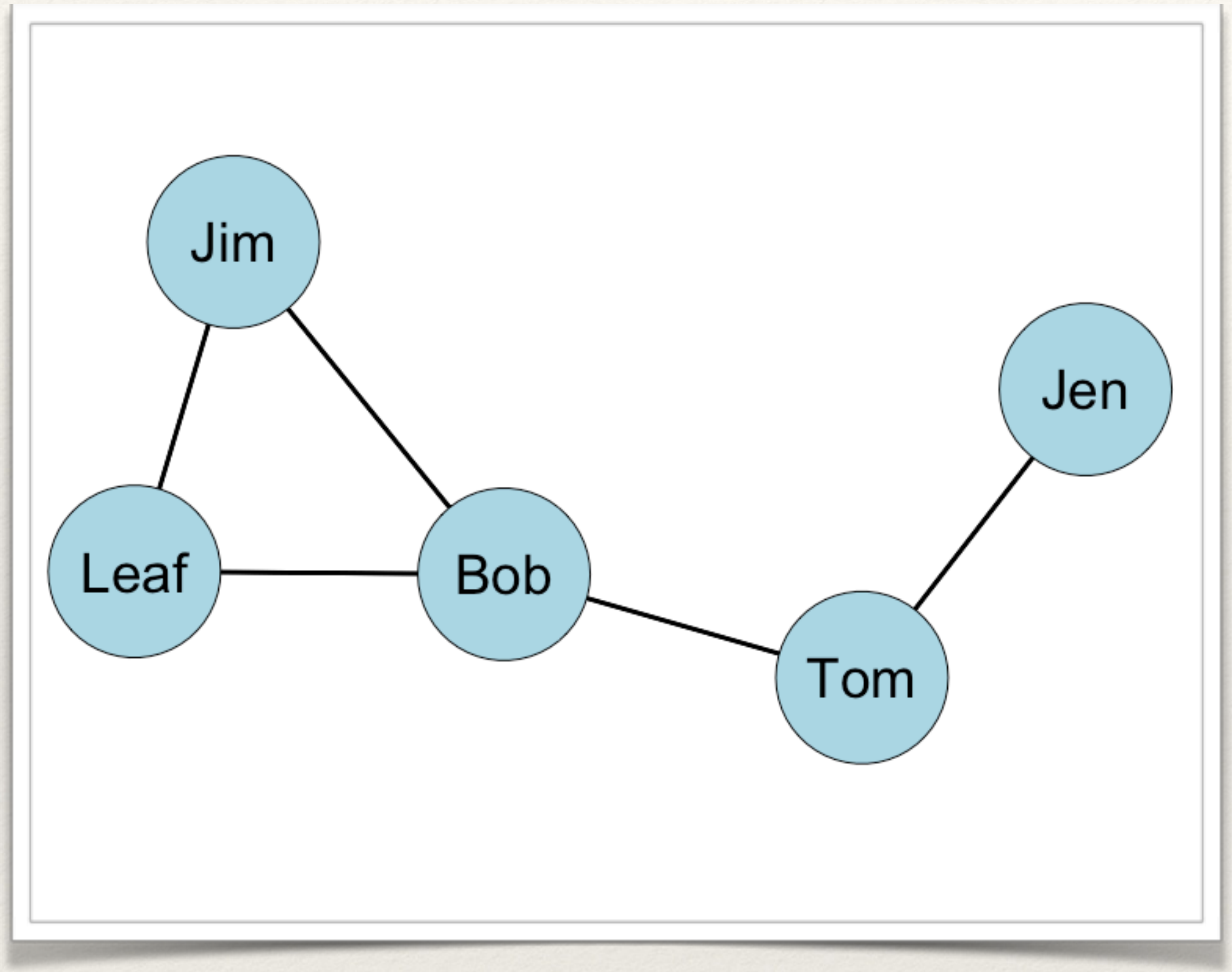
Example: Undirected, Binary Network



	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.

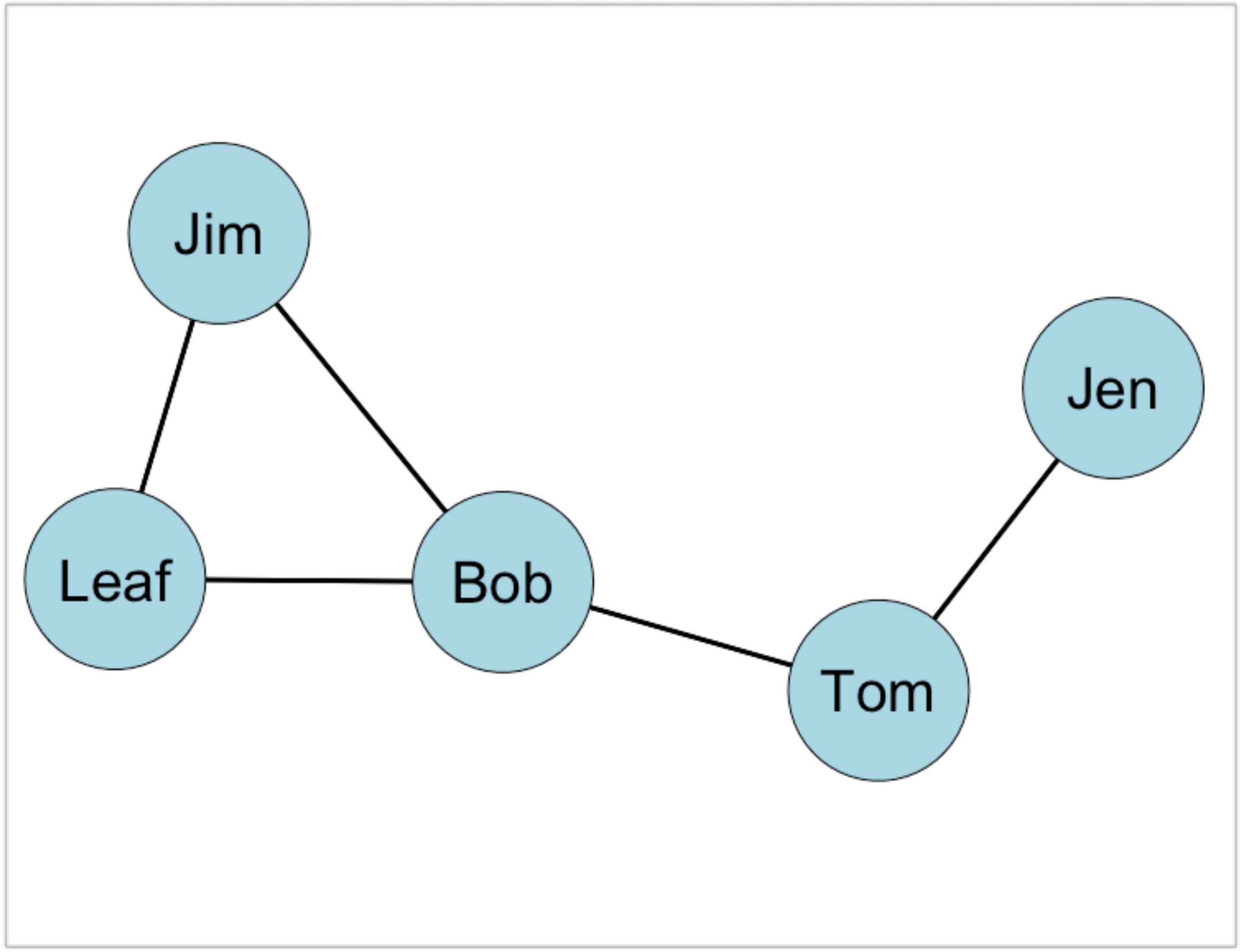
Example: Undirected, Binary Network



	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: $X_{12} = 1$; $X_{15} = 0$

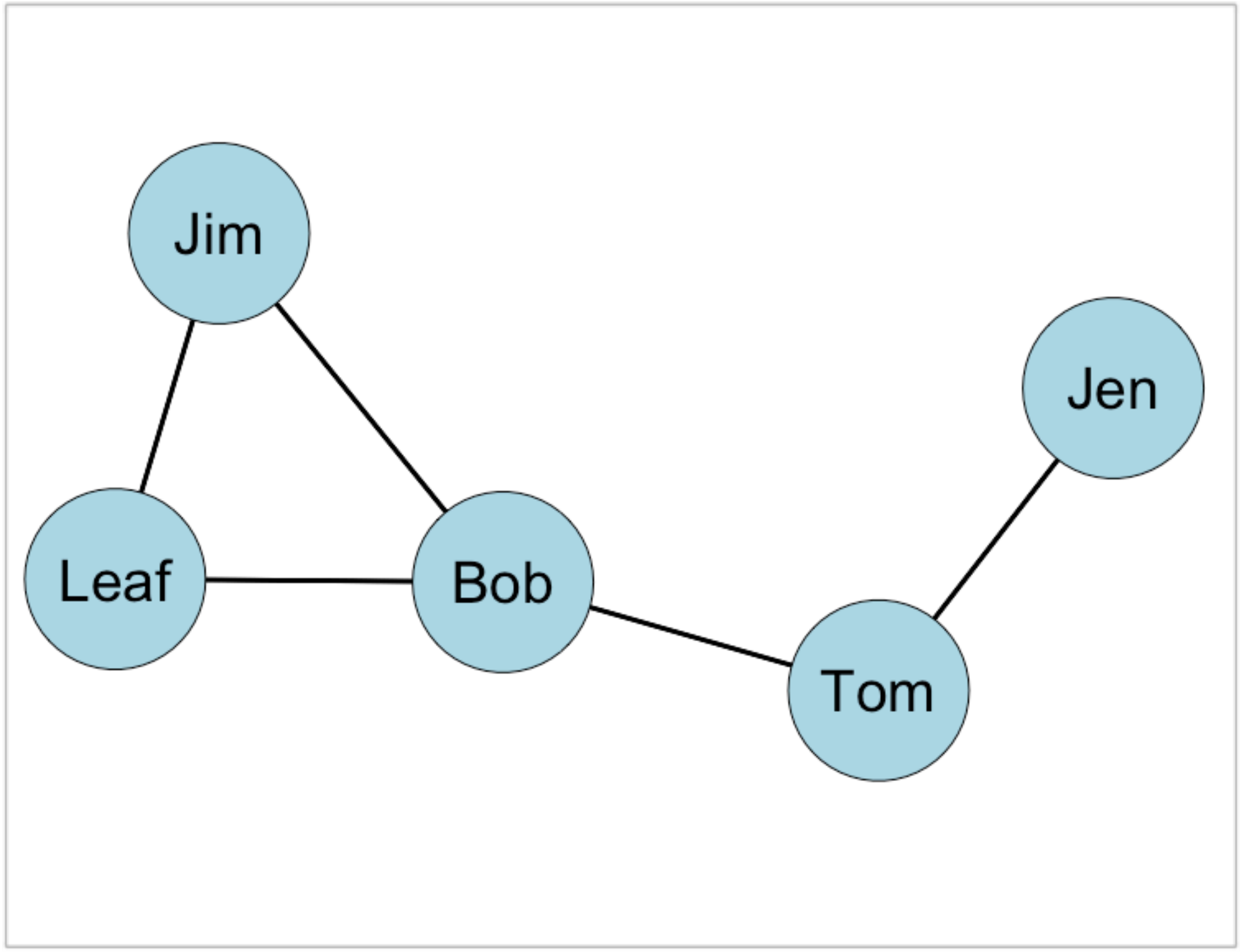
Example: Undirected, Binary Network



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

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

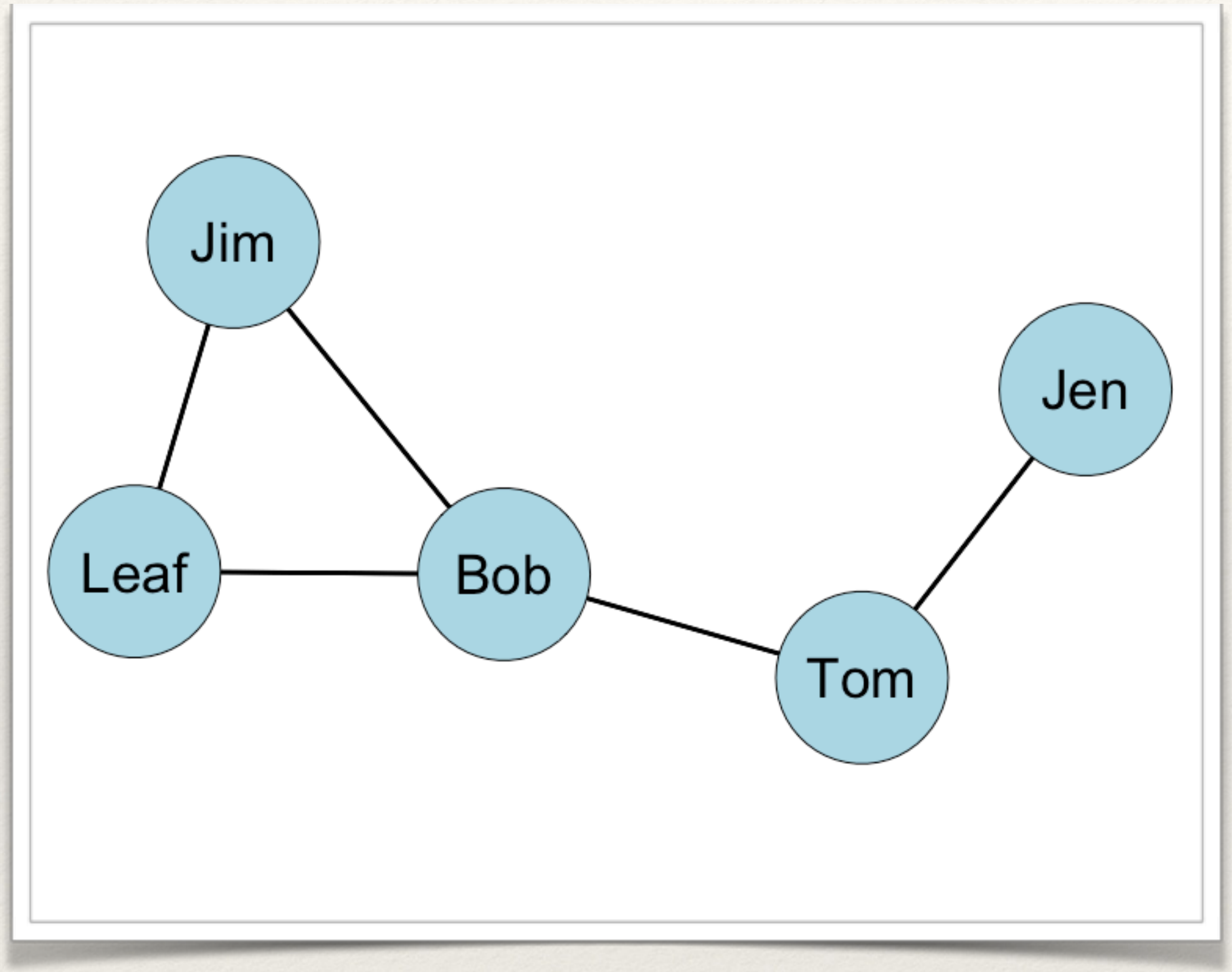
Example: Undirected, Binary Network



	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?

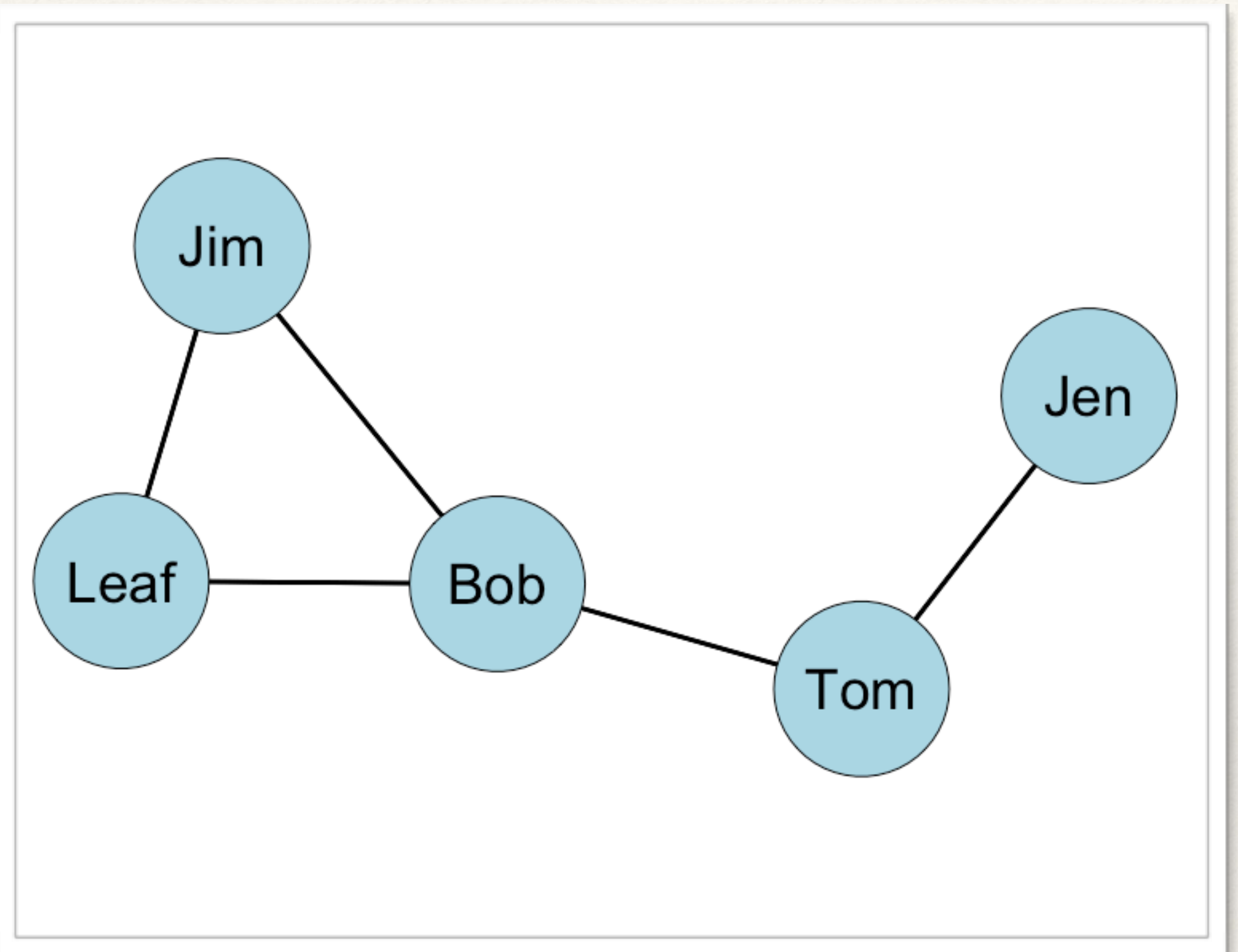
Example: Undirected, Binary Network



It looks like this.

	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	

Example: Undirected, Binary Network

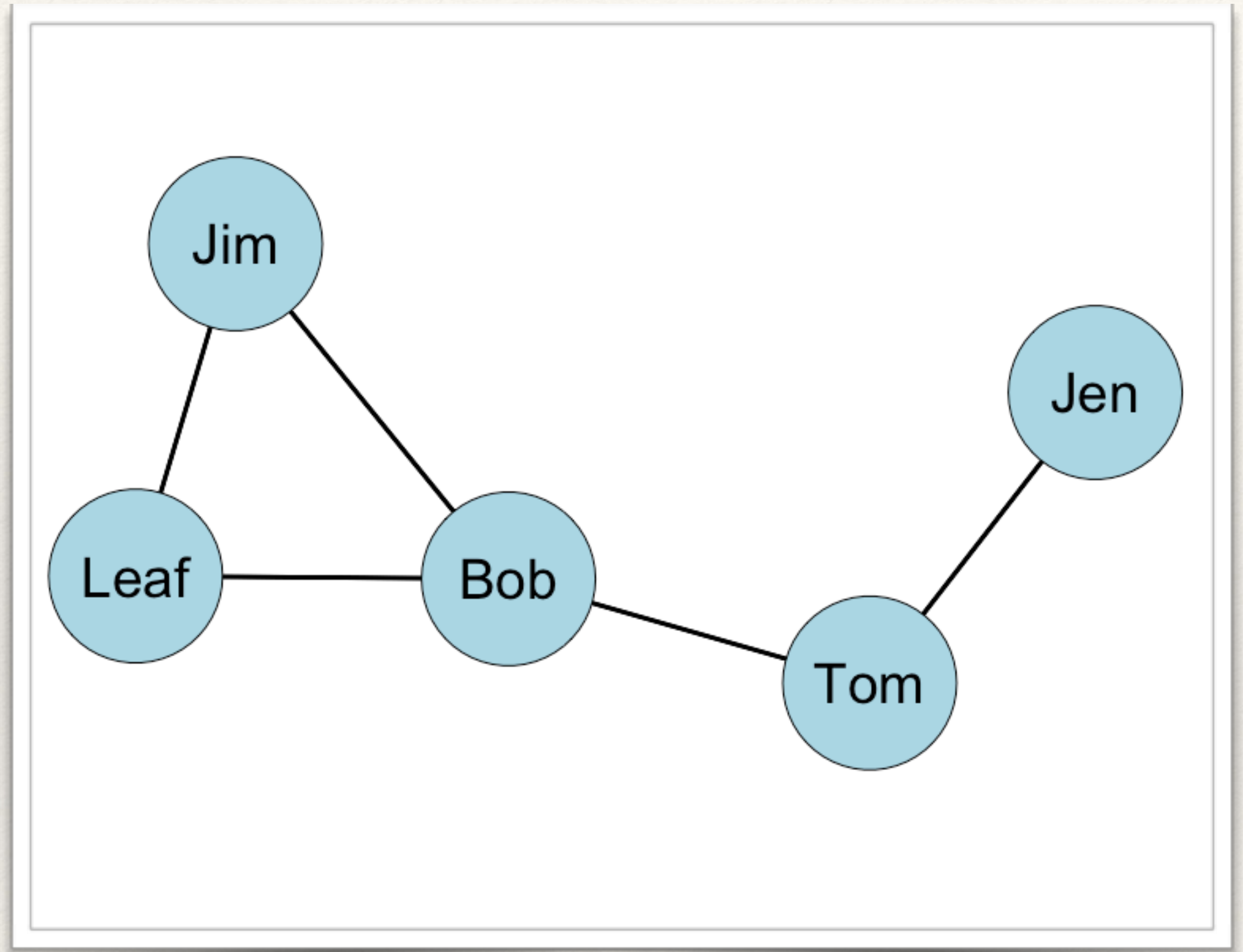


It looks like this.

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

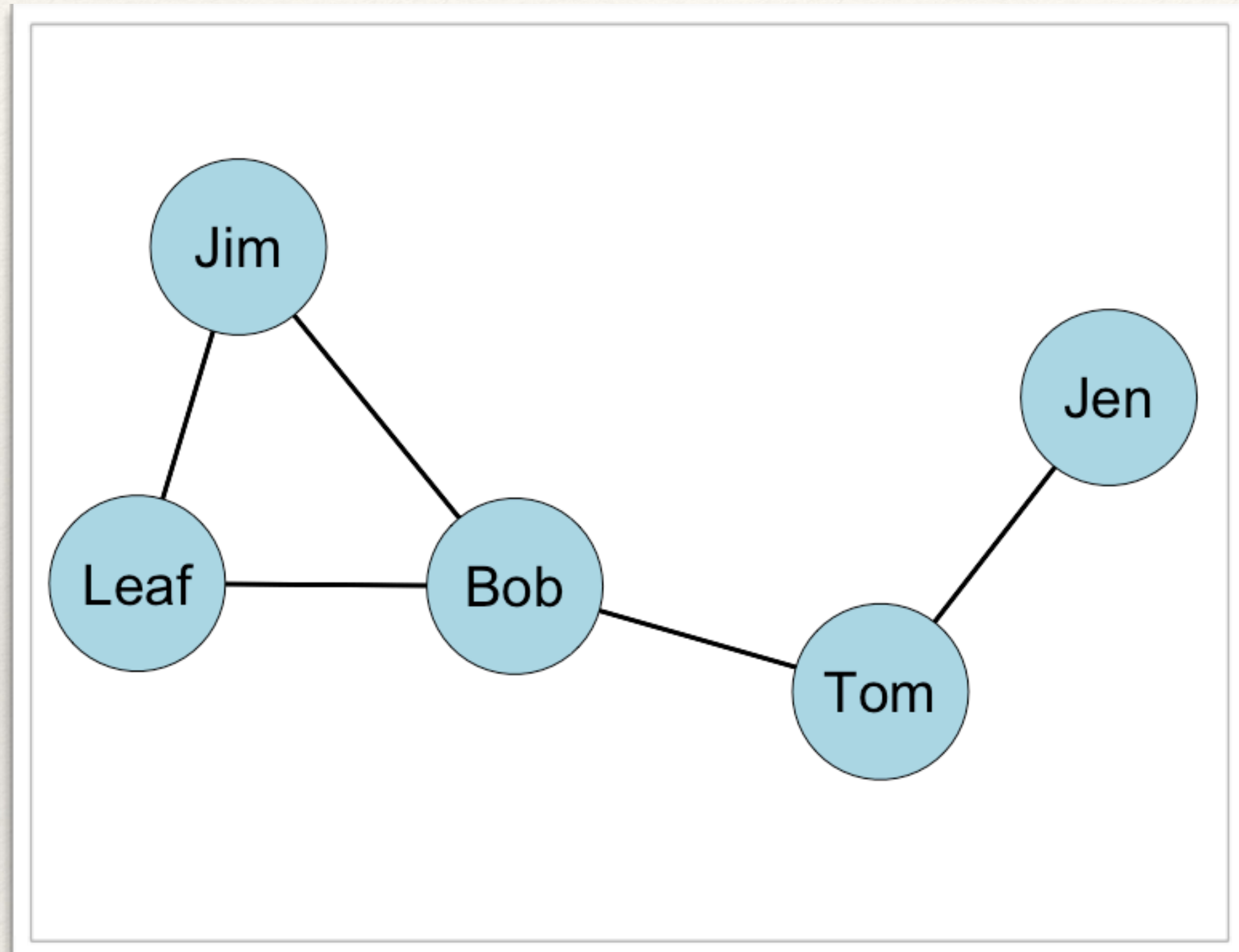


The highlighted section here is called the **lower triangle** of the matrix.

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

	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

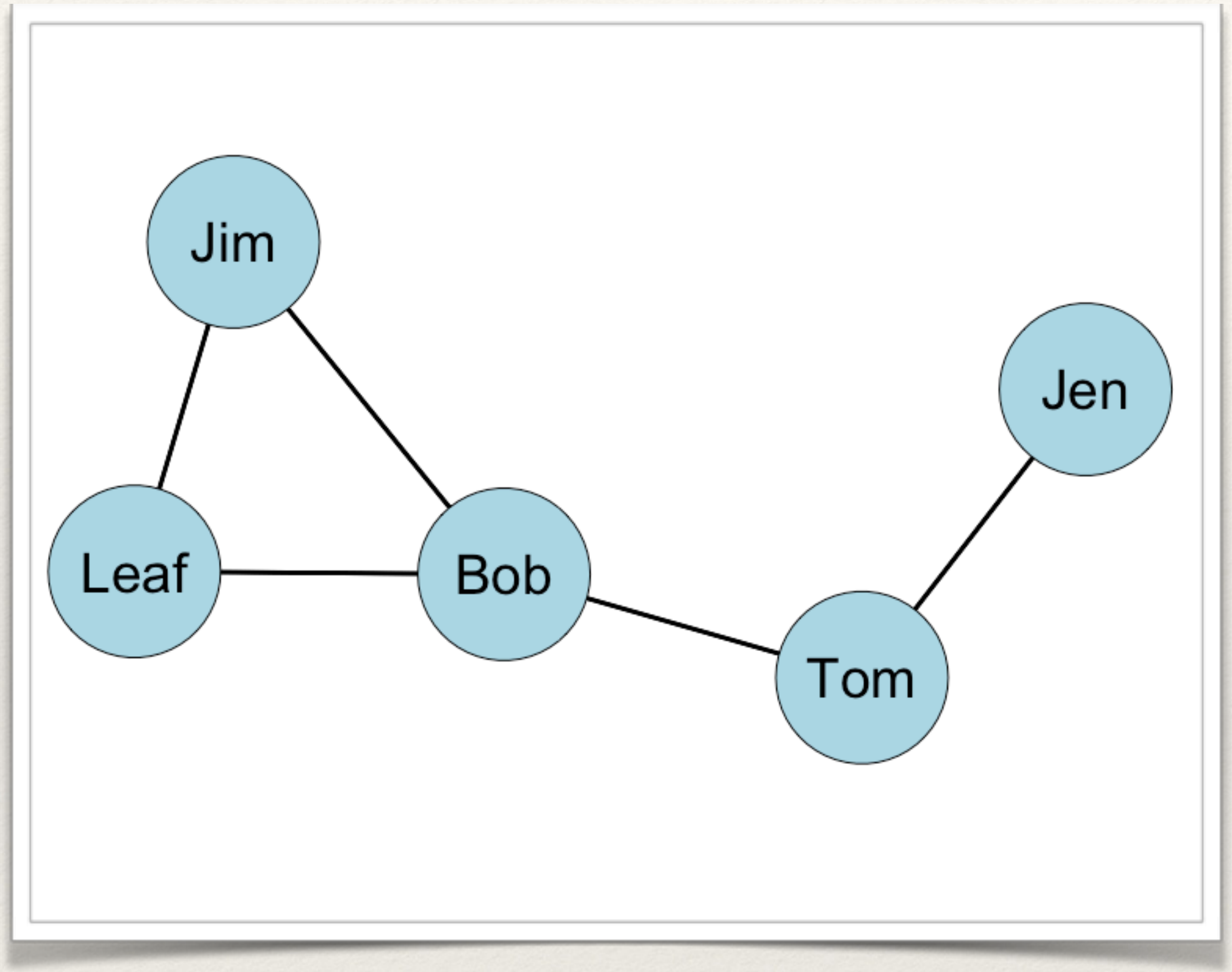


	Jen	Tom	Bob	Leaf	Jim
Jen	0	1	0	0	0
Tom	1	0	1	0	0
Bob	0	1	0	1	1
Leaf	0	0	1	0	1
Jim	0	0	1	1	0

The 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.

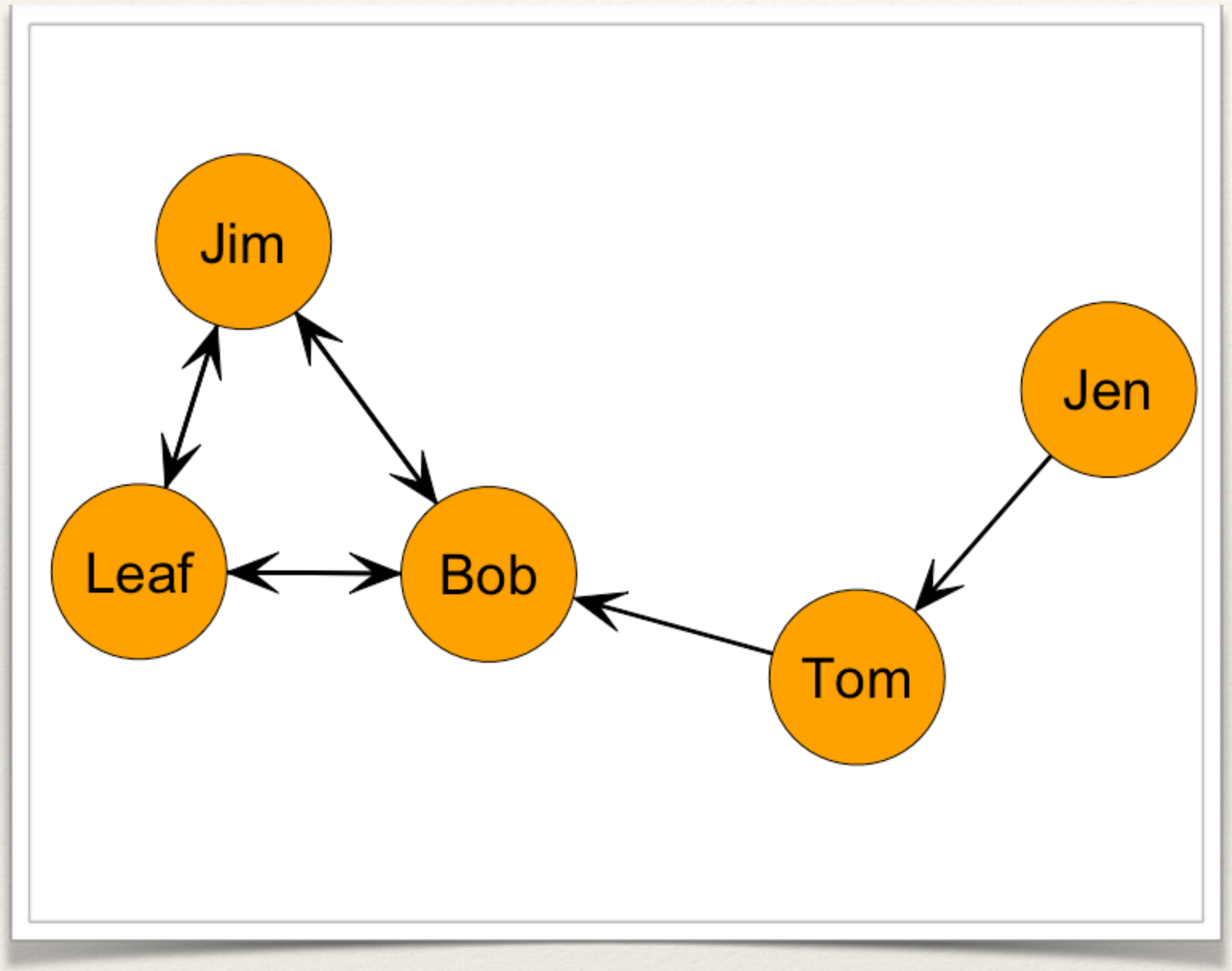
Example: Undirected, Binary Network



	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.

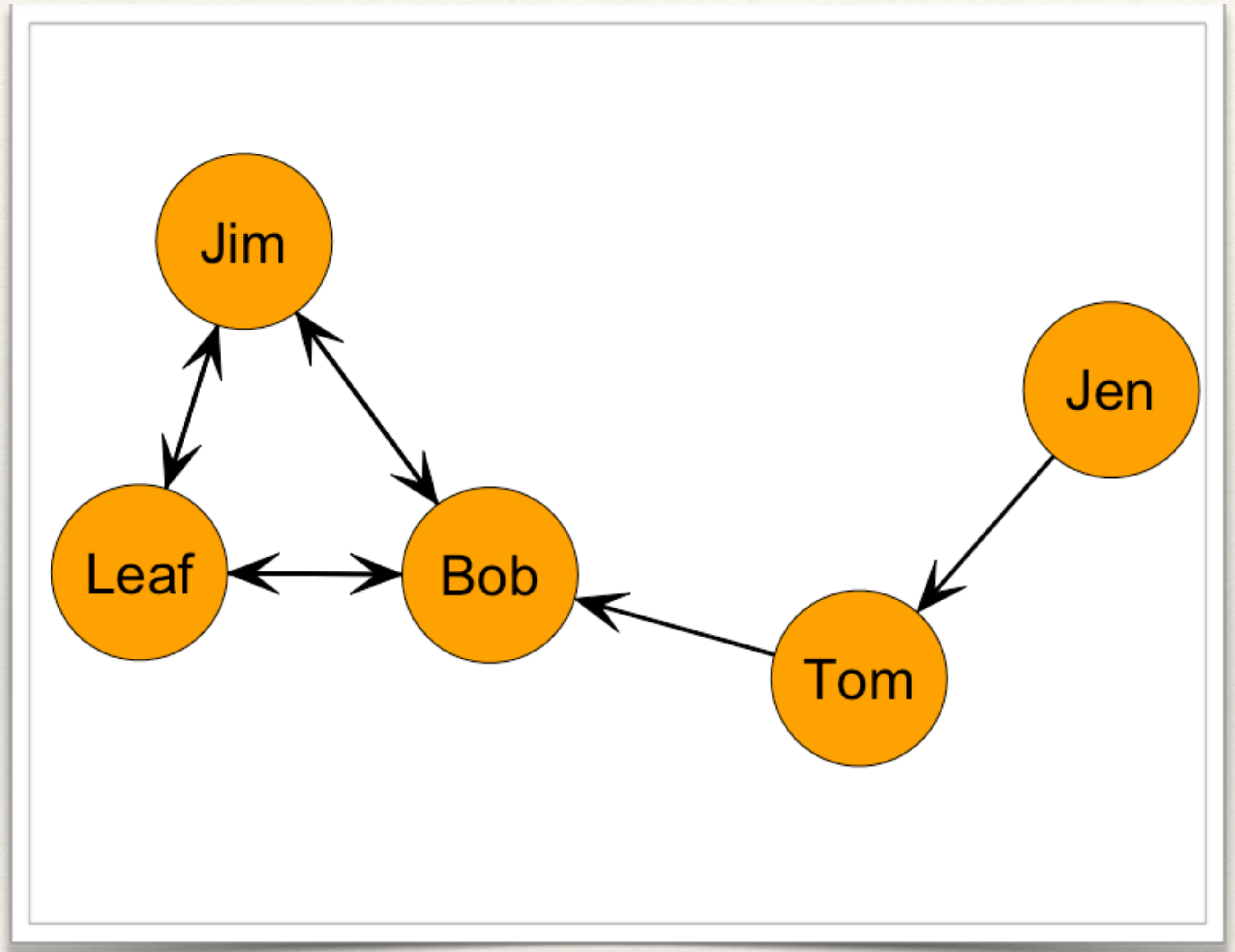
Example: Directed, Binary Network



	Jen	Tom	Bob	Leaf	Jim
Jen					
Tom					
Bob					
Leaf					
Jim					

What's different about a directed network?

Example: Directed, Binary Network

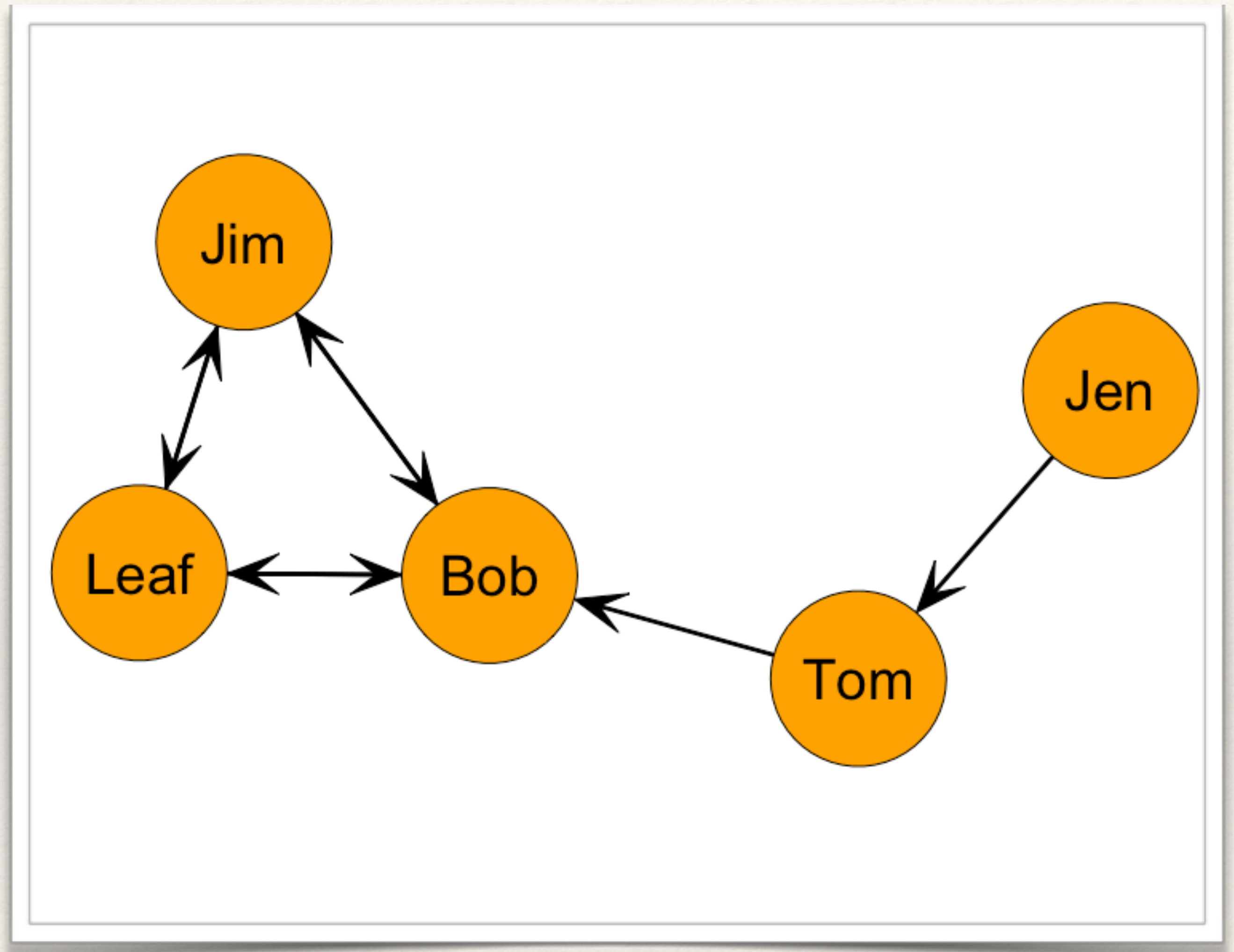


	Jen	Tom	Bob	Leaf	Jim
Jen		1			
Tom					
Bob					
Leaf					
Jim					

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

$X_{12} = 1$

Example: Directed, Binary Network

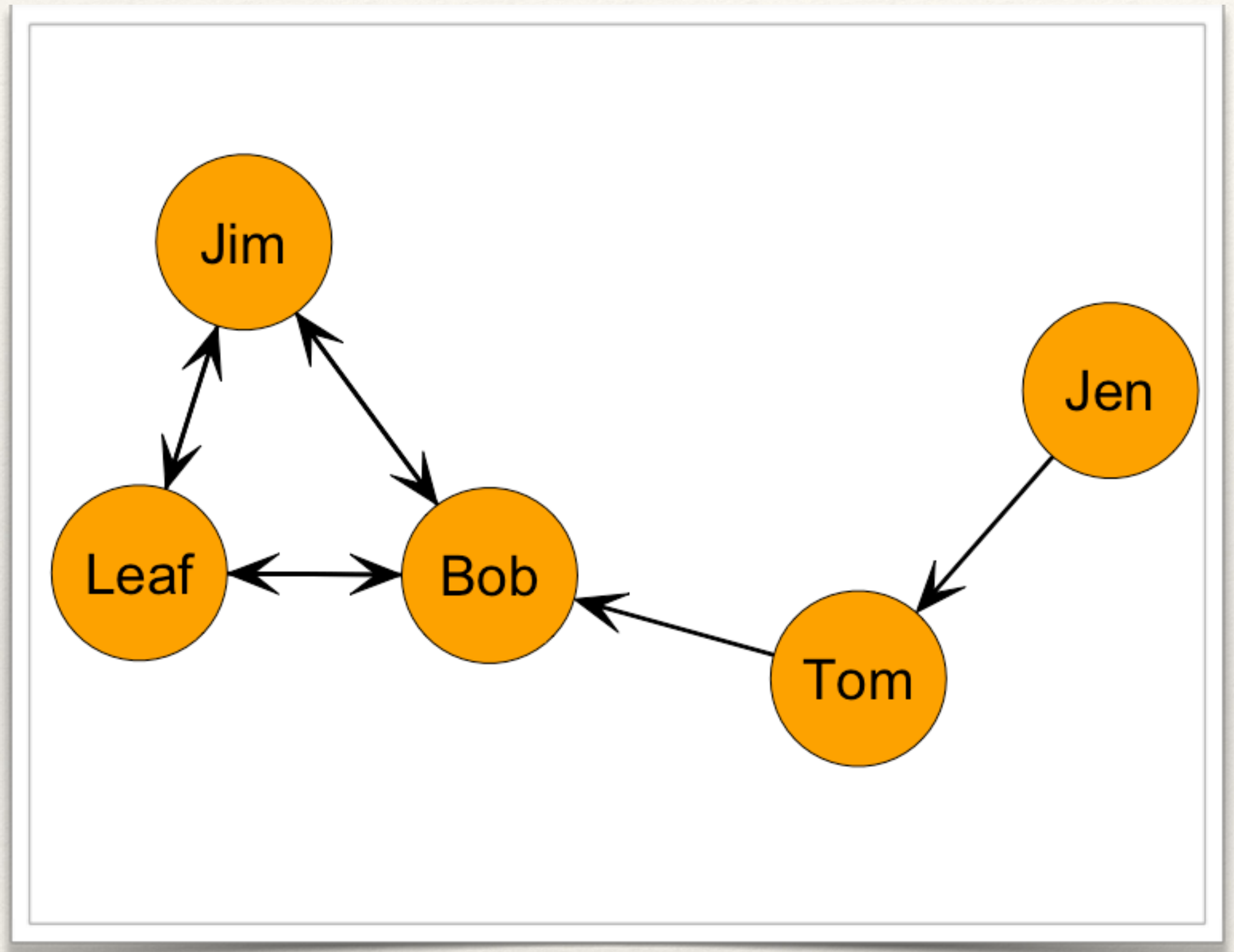


	Jen	Tom	Bob	Leaf	Jim
Jen		1			
Tom	0				
Bob					
Leaf					
Jim					

But in the second row, *j* does not send:

$X_{21} = 0$

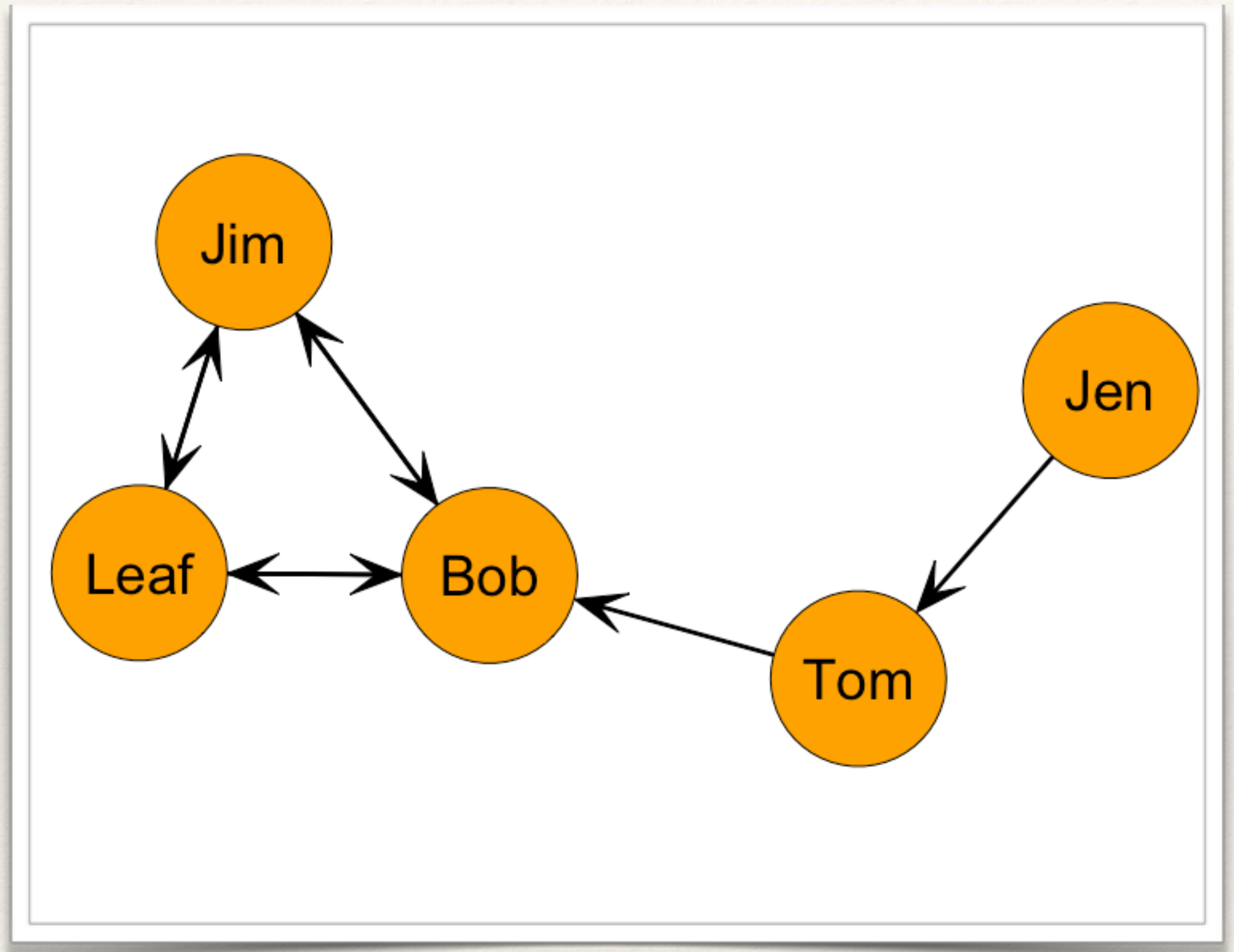
Example: Directed, Binary Network



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

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

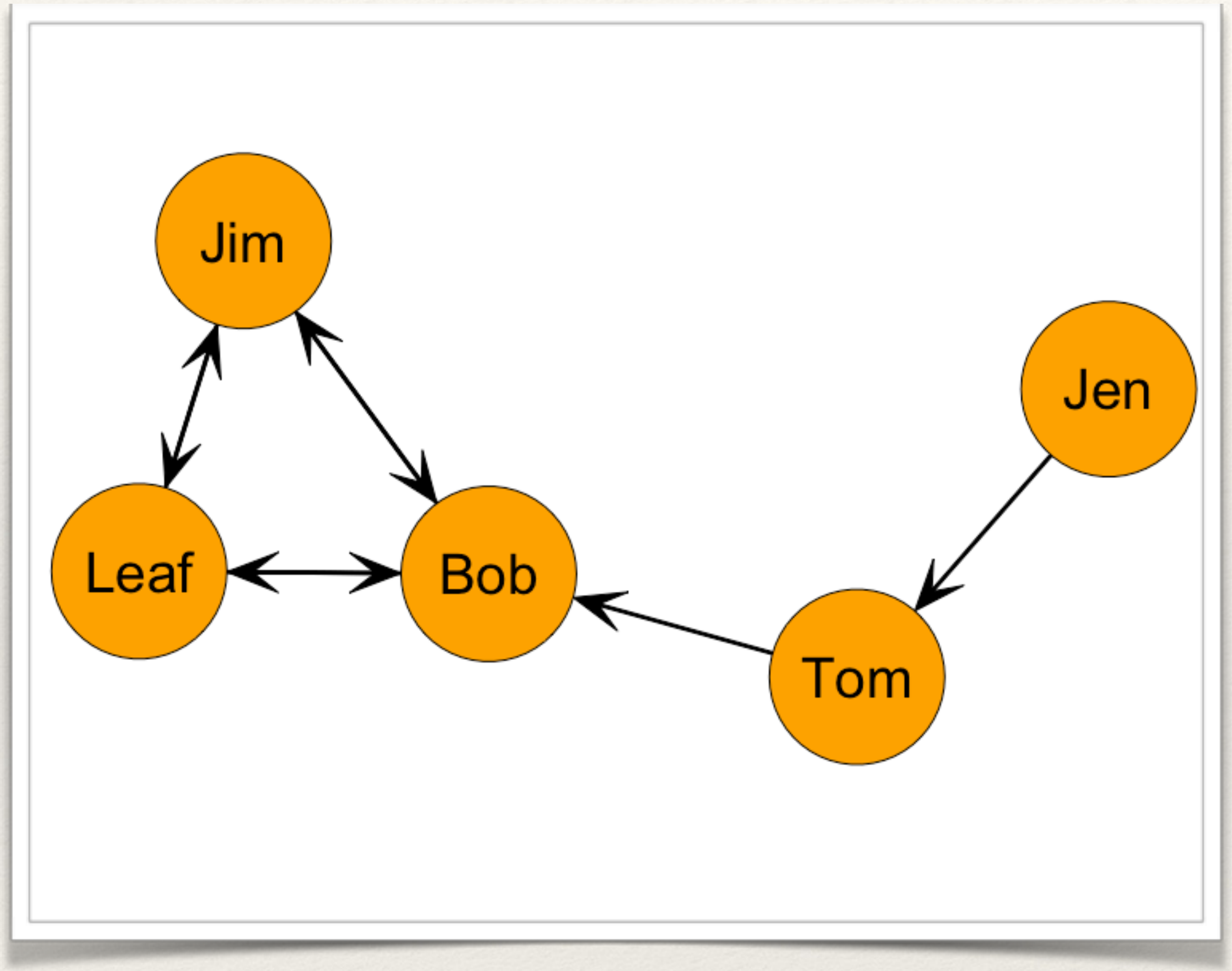
Example: Directed, Binary Network



	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**?

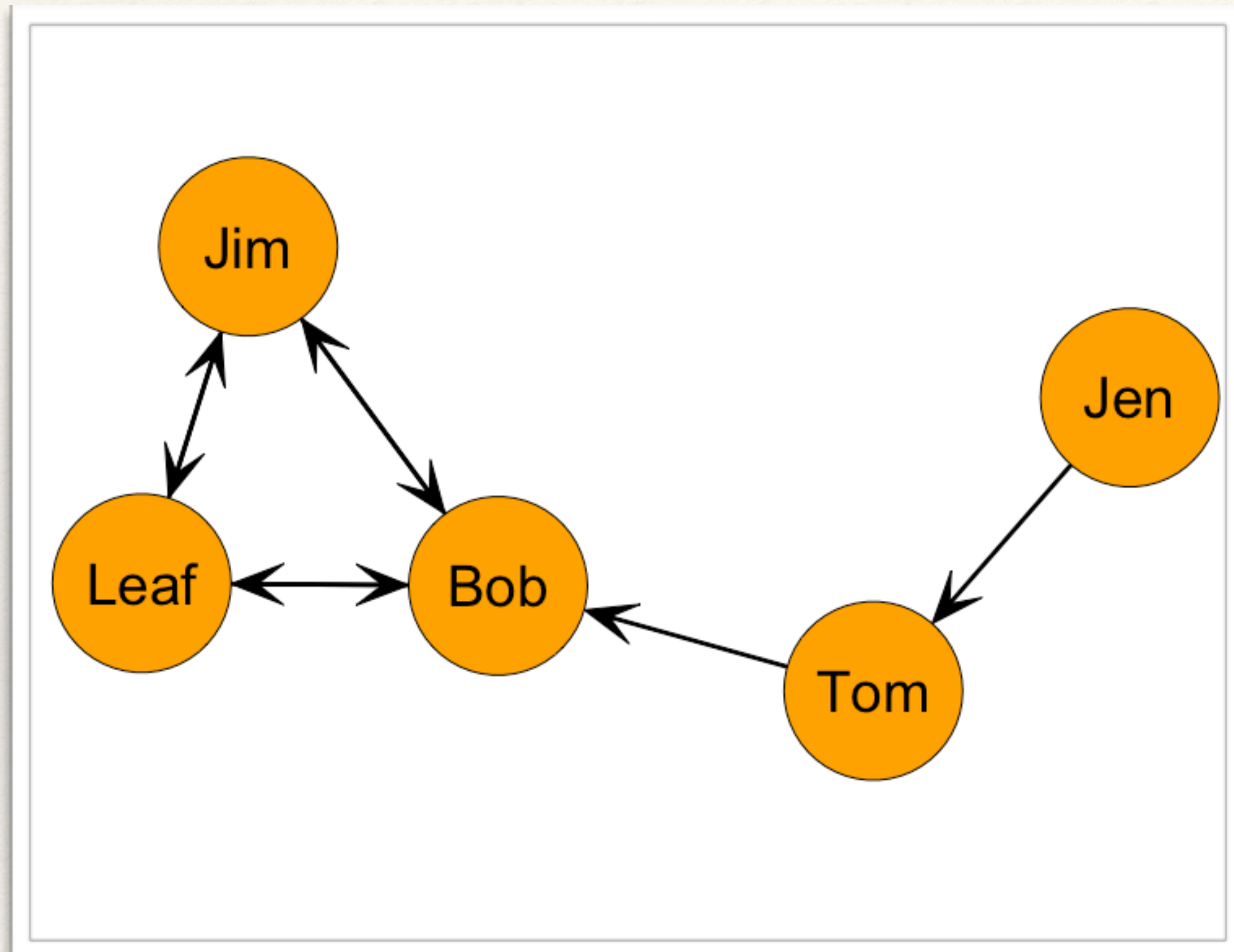
Example: Directed, Binary Network



	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

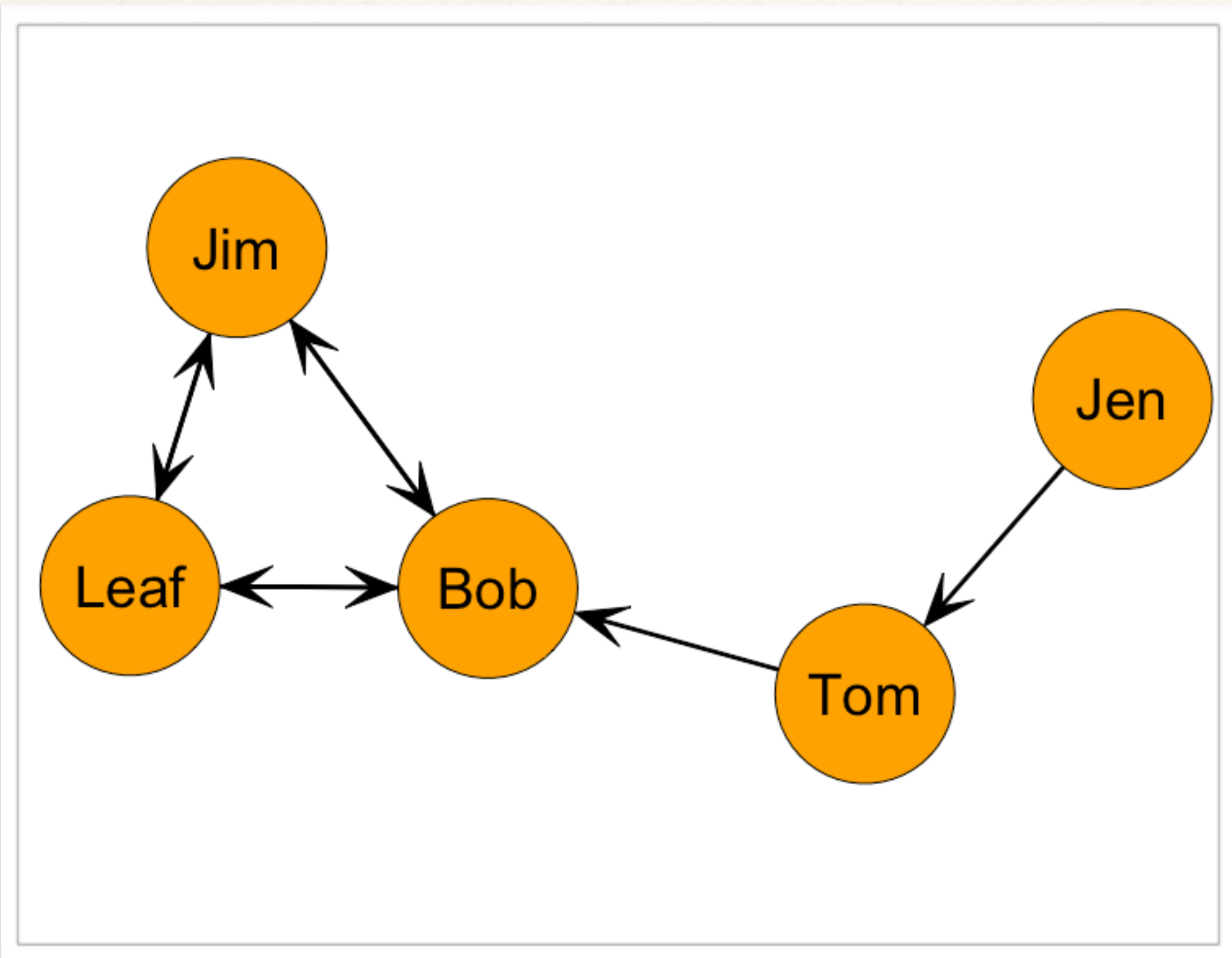


It looks like this.

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

Not that, because we are allowing directionality to matter, the total number of edges in the network is just the **sum** of the entire matrix.

Edgelists

- Very large networks are sometimes represented with an *edgelist*.
    - An edgelist lists the edges in a graph with the head of the edge in the first column and the tail of the edge in the second column.
    - Note: *isolates* (nodes without incident edges) are excluded from edgelists.

Learning Goals

- Understand various approaches to collecting network data.
- Understand the representation of networks using graph and graph notation.
- Understand the representation of undirected and directed networks using matrices.

Questions?