Intro to Network Analysis

A crash course in terminology, metrics, and visualization

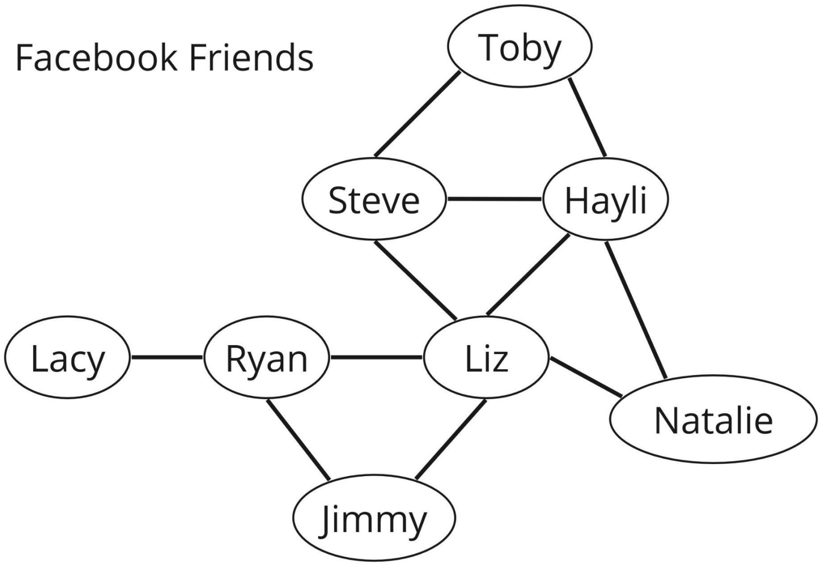
E 388M - Spring 2024

Part I: Nodes & Edges

What's a **network**?

A model of **things** and the **relationships** between them

Also called a **graph** (more common in mathematical contexts)



Key Terms: Nodes and Edges

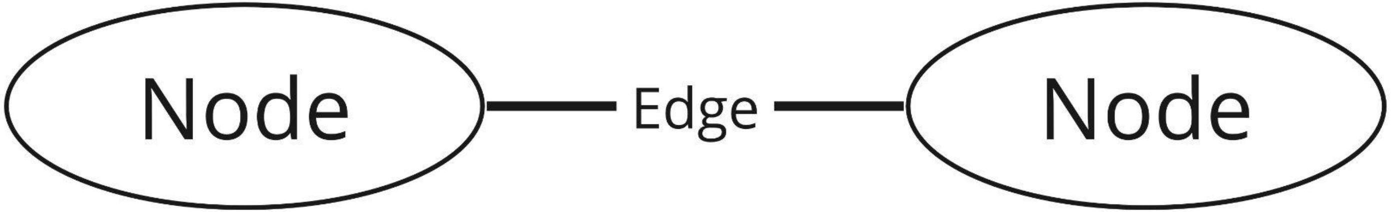
Networks are made of **nodes** and **edges**

Nodes are “things”

Edges are connections between things

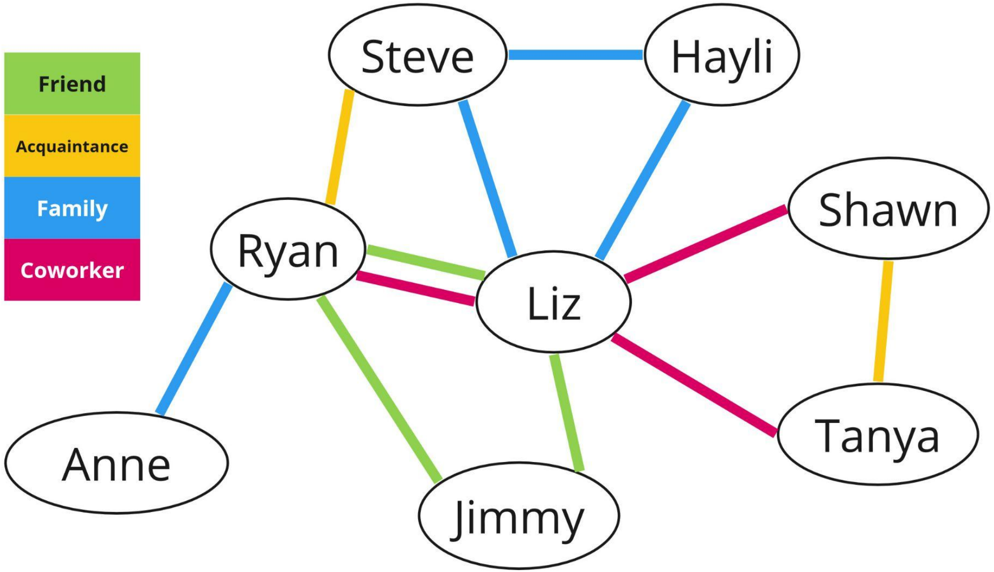
In **social networks**, nodes are **people**

But **nodes** can be **anything**, and **edges** can represent **any relationship**

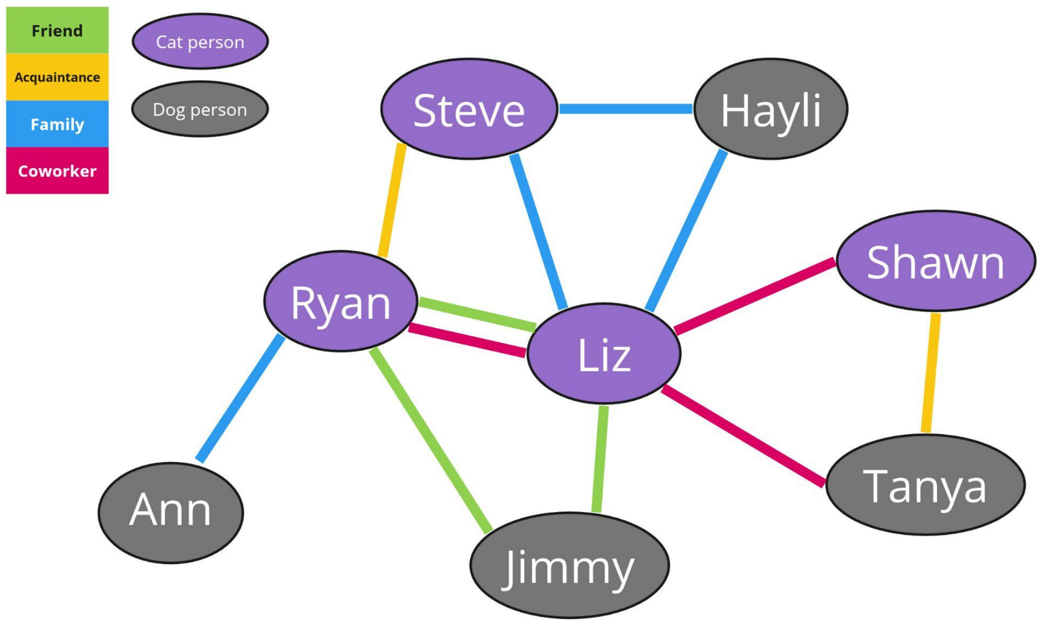


Nodes are typically drawn as circles/ovals, **Edges** are typically drawn as lines

Edges may represent **different relationships** within a single network

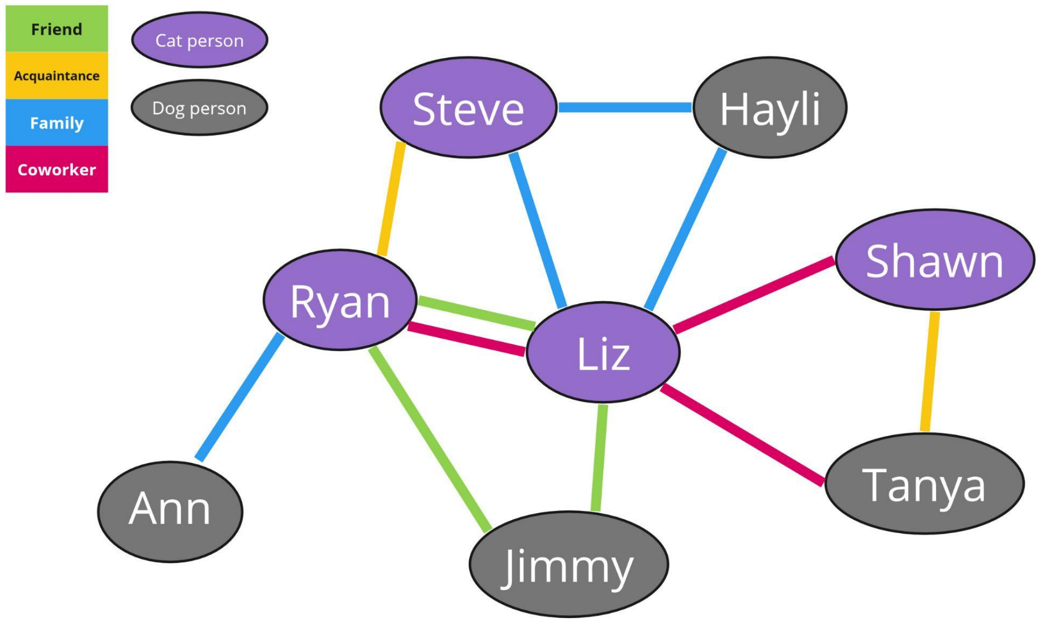


Nodes can have any number of **attributes** that may be used in analysis



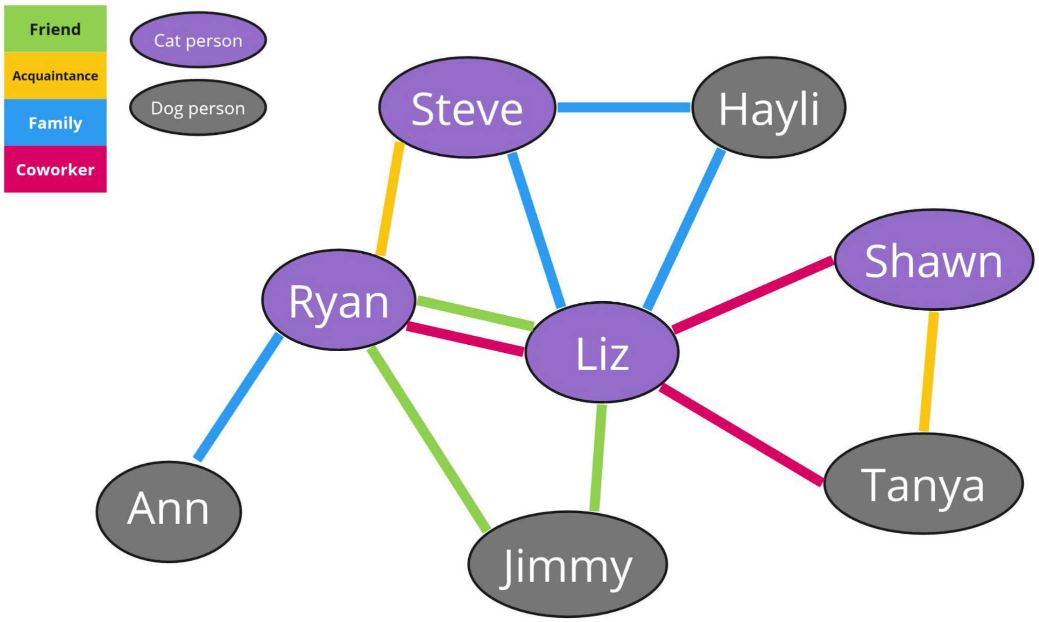
Checkpoint!

- This diagram shows a _____, also called a _____
- The ovals are called _____
- The lines are called _____
- In this diagram, ovals represent ______ and lines represent ______



Checkpoint!

- This diagram shows a **network**, also called a **graph**
- The ovals are called **nodes**
- The lines are called **edges**
- In this diagram, ovals represent **people** and lines represent **relationships between people**



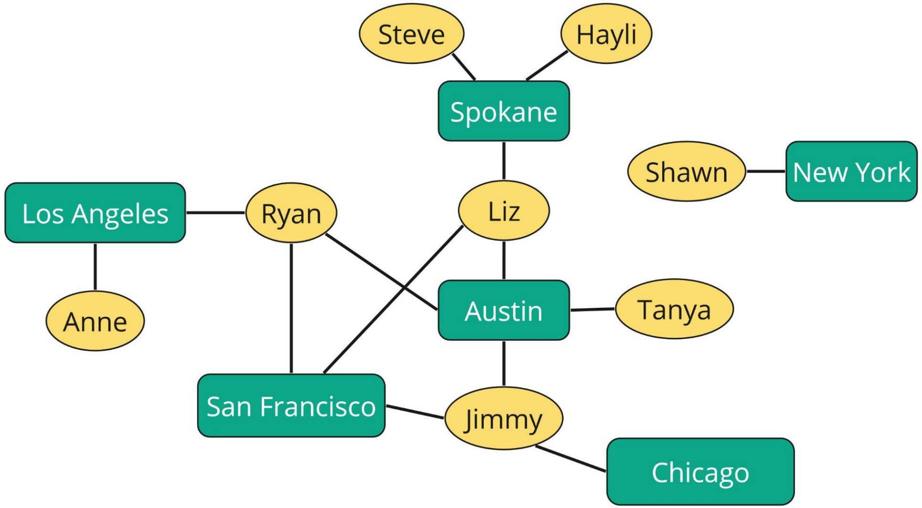
Network **Modality** / **k-partite** networks

Modality refers to the number of **node types** in a network.

Nodes are different **types** if they're representing different conceptual **things**
(unlike the last example, same type with variety in attributes)

A network with **one type** of node is called **unimodal** or **monopartite**

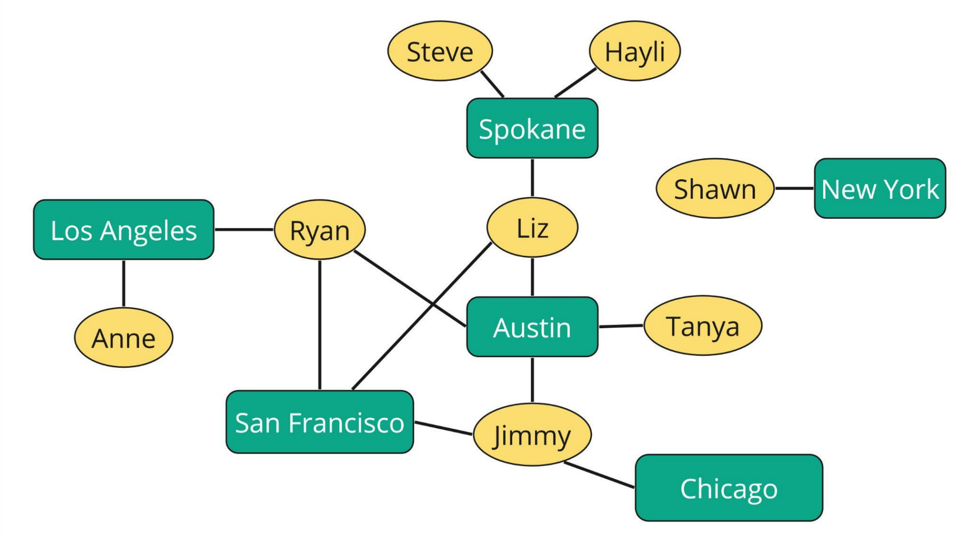
- 2 types = **bimodal** or **bipartite**
- 3 types = **trimodal** or **tripartite**
- etc.



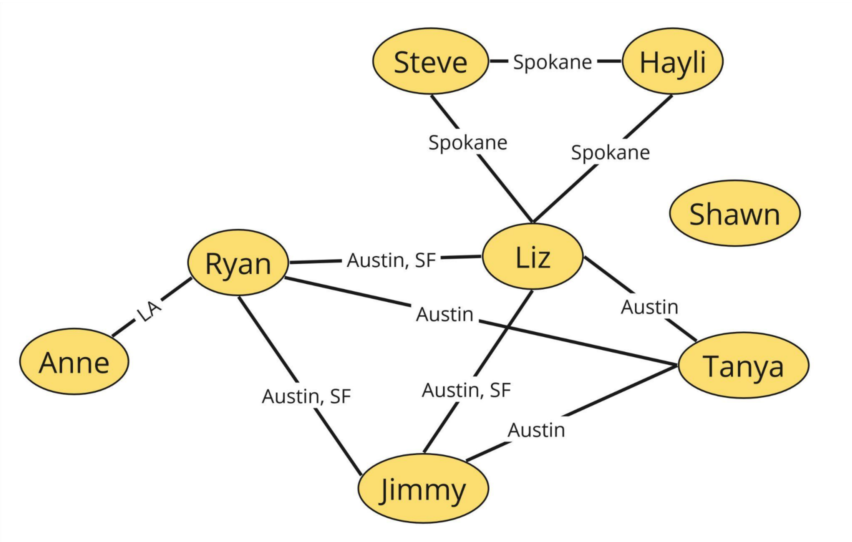
A **bipartite/bimodal** network of people and places

Collapsing to lower modalities

Bimodal graphs can be **collapsed** (or “projected”) to **unimodal** graphs by converting one **node type** into **edges**



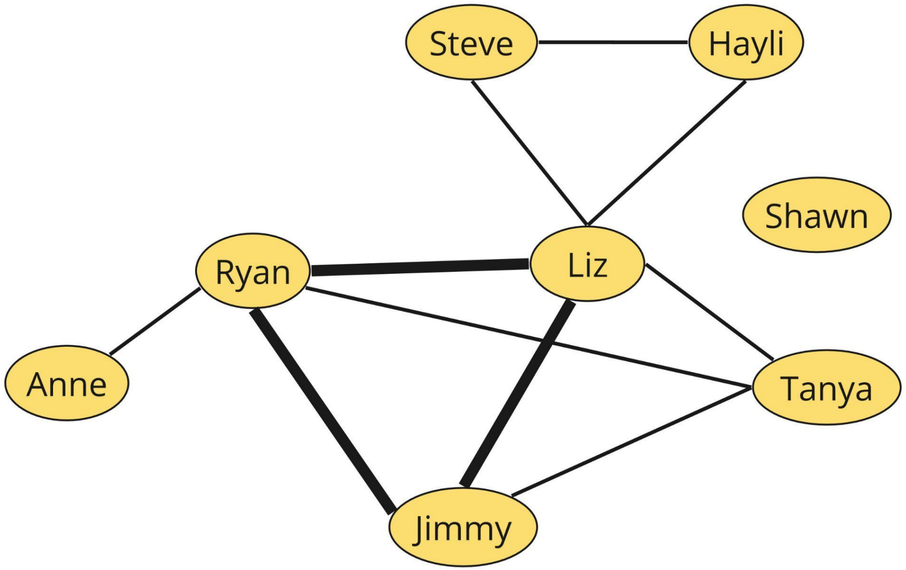
Can also do the reverse, changing some element of edges into nodes. Less common.



Edges can have

Weight

Weight indicates the "strength" of the relationship



Edge weight is commonly represented by line **thickness**

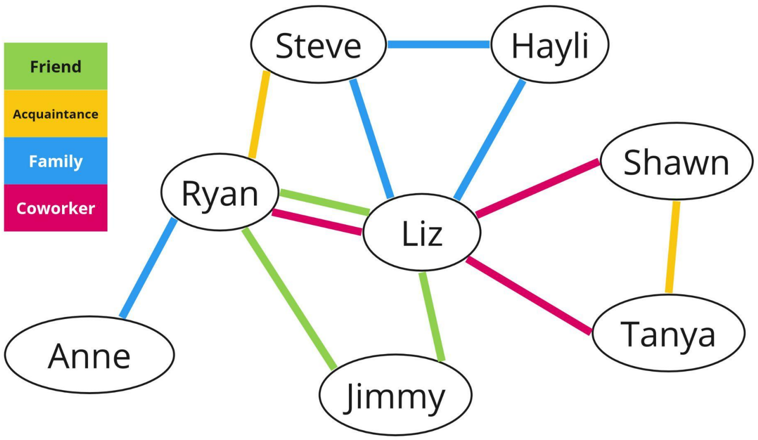
Edges can be

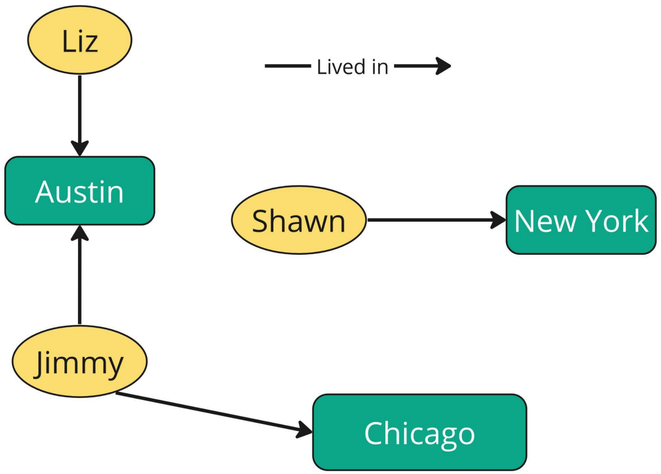
directed or undirected

Directed edges represent relationships that only go **one way**

Undirected edges represent relationships that go **both ways**

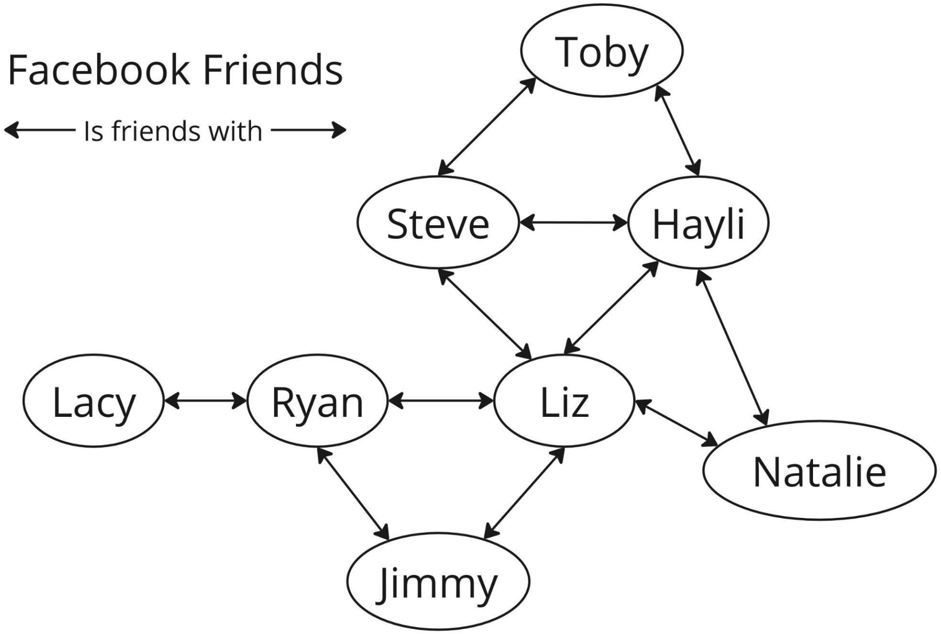
These relationships are reciprocal—**you are also your family's family**. So these edges are **undirected**



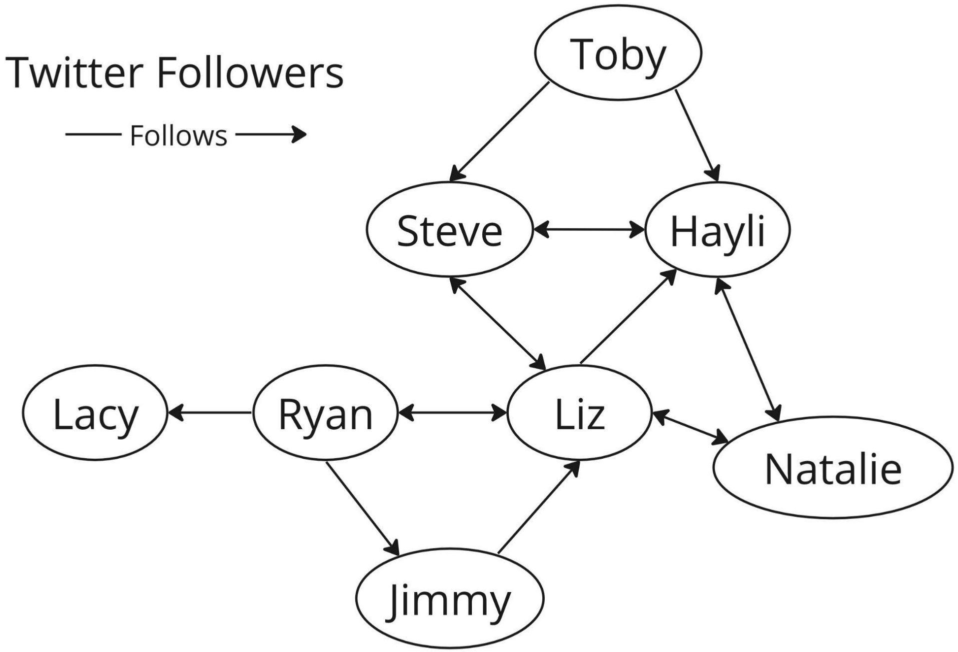


People *live in* places, but places don't *live in* people—so these edges should be **directed**

Facebook Friendship is an **undirected** relationship

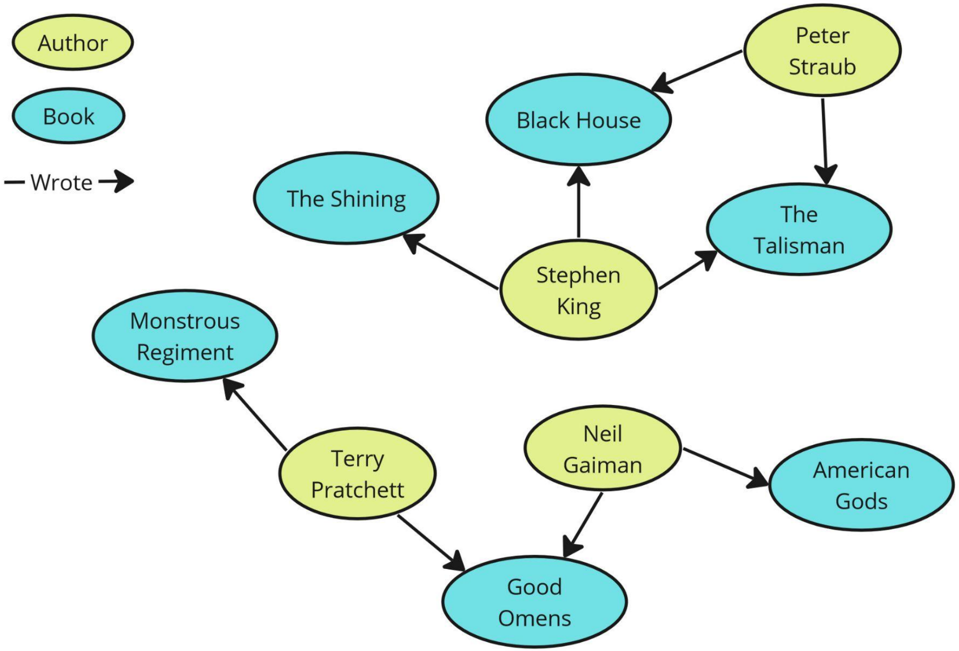


Twitter Following is a **directed** relationship



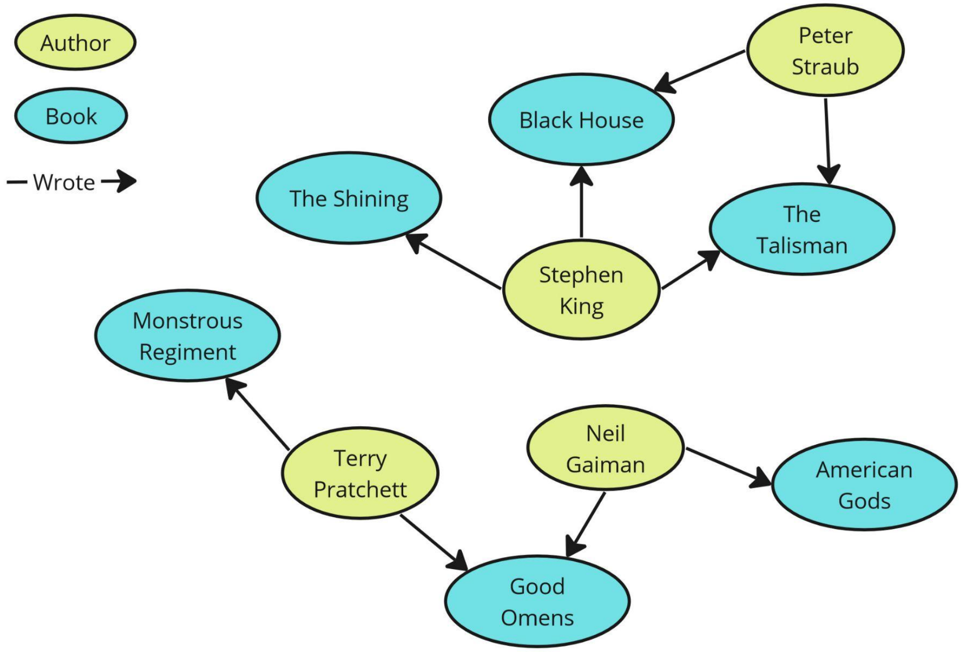
Checkpoint!

- This is a ___-modal network
- Its edges are ______
- We could convert this to a ___-modal network focused on author collaborations by converting ______ to _____
- In the converted network, the edge between ____ and ____ would have a ____ of 2



Checkpoint!

- This is a **bi**modal network
- Its edges are **directed**
- We could convert this to a **uni**modal network focused on author collaborations by converting **book nodes** to **edges**
- In the converted network, the edge between **King** and **Straub** would have a **weight** of 2



Part II: Metrics & Layouts

Network analysis can involve both **visualization** and **statistical analysis**

The statistical side of network analysis involves calculating various **metrics**, at both the node level & network-wide level.

Most common node metrics are measures of **centrality**—kind of like “importance”

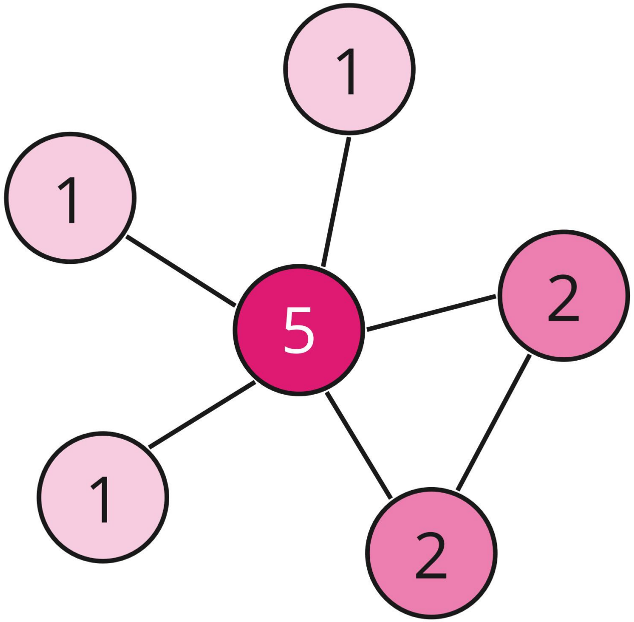
The visualization side of network analysis involves representing **metrics** by varying **color, size, and shape**.

Visualization also involves **arranging nodes** according to some **layout algorithm**

One way of measuring centrality is with

Degree,

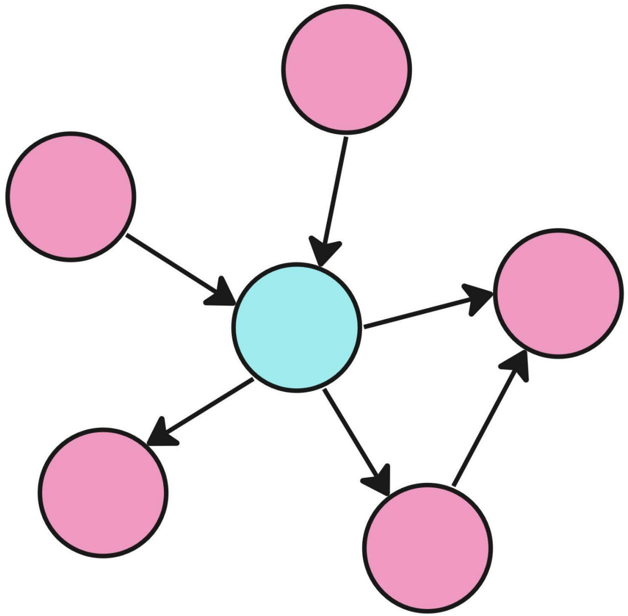
or how many connections a node has



A node's **degree** equals the number of **edges** connect to that node. Higher degree = higher centrality. Think "the popular kids," or a busy transportation hub.

In graphs with **directed edges**, there are two degree measures:

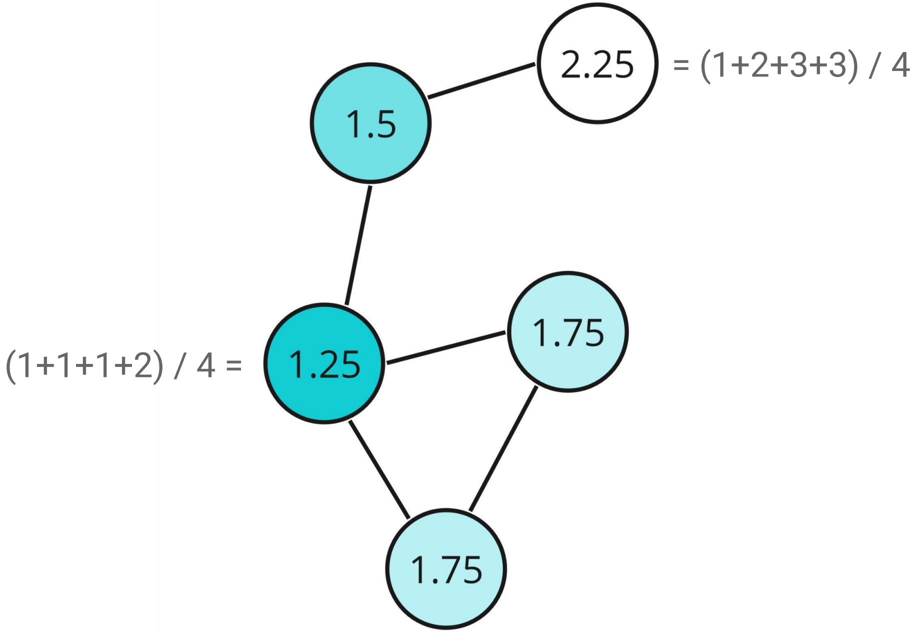
“in-degree” & **“out-degree”**



For example, the **blue node** has an **in-degree of 2** and an **out-degree** of 3.

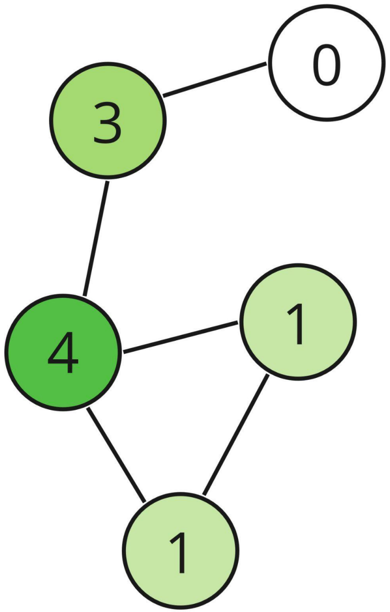
Another way of measuring centrality: **Closeness**,

or how close a node is to other nodes, on average



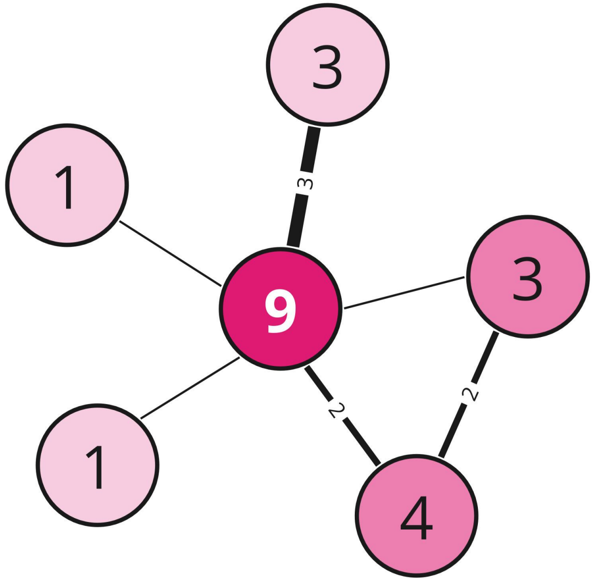
A node's **closeness** equals the sum of the number of **hops** it takes to get from a node **to each other node**, divided by the **number of other nodes**. Lower closeness = higher centrality. Imagine minimizing travel for a group get-together.

Another way of measuring centrality: **Betweenness**, or how good of a bridge the node is to others



A node's **betweenness** is the **number of times** it falls on the **shortest path** between two other nodes. Higher betweenness = higher centrality. Think of an information bottleneck, or someone who is friends with *radically* different groups of people.

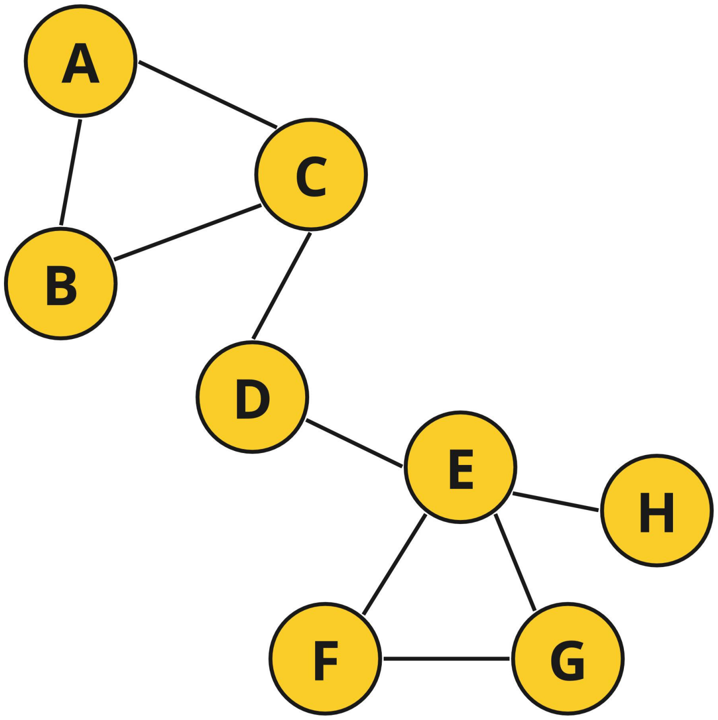
In graphs with **weighted edges,** these metrics factor in edge weight



For example, **weighted degree** counts the total weight of edges connecting to a node

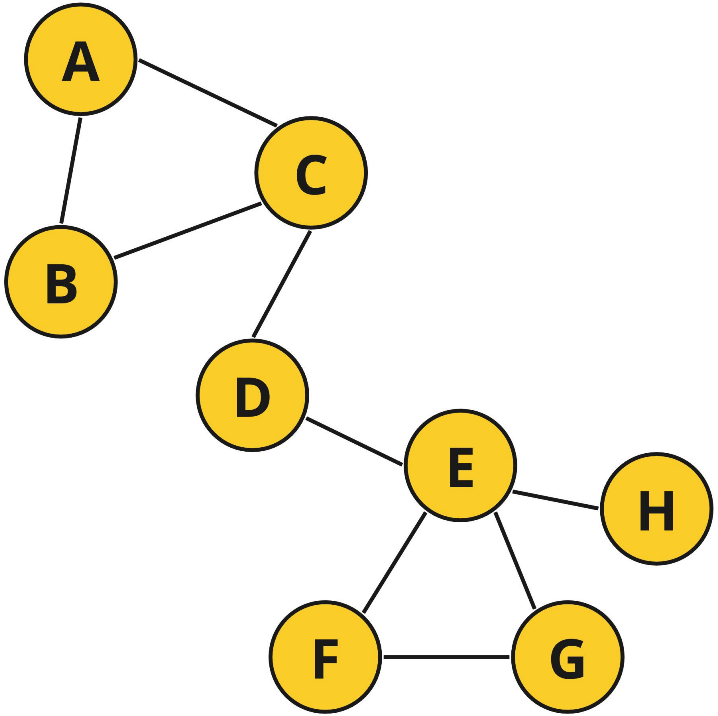
Checkpoint!

- Node **C** has a **degree** of ____
- The node with the **highest degree** is ____
- Node **D** has a **betweenness** of _____
- Node **H** has a **closeness** of __ /__



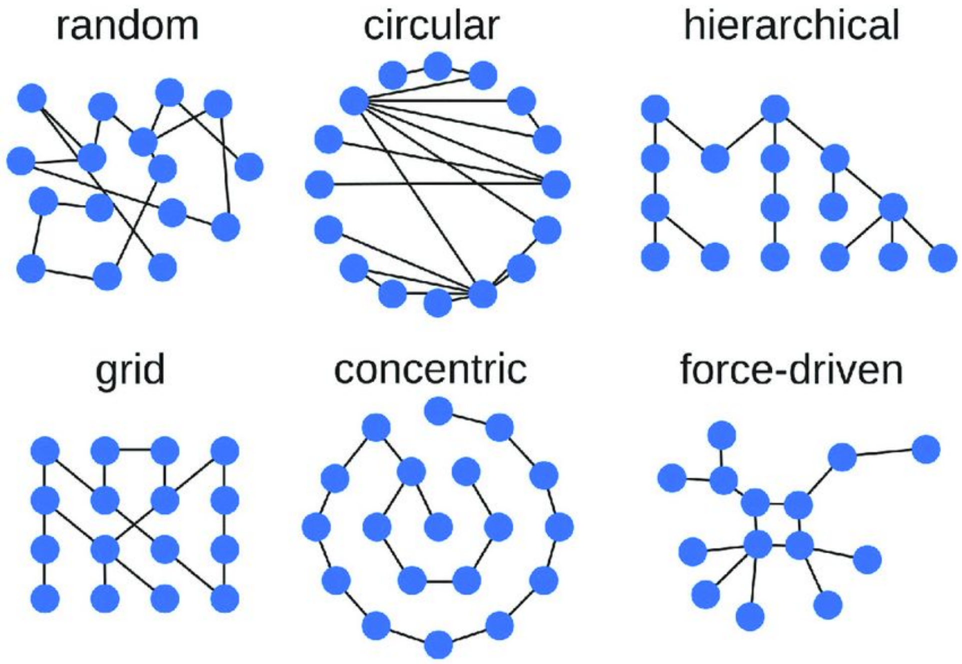
Checkpoint!

- Node **C** has a **degree** of **3**
- The node with the **highest degree** is **E** (4)
- Node **D** has a **betweenness** of **12**
- Node **H** has a **closeness** of **18 / 7** (~2.57)



Calculating these by hand is annoying—we let software do it for us

Network **visualizations** arrange nodes according to some **layout** algorithm



Force-directed

layouts use **edge weights** and **simulated gravity** to position nodes.

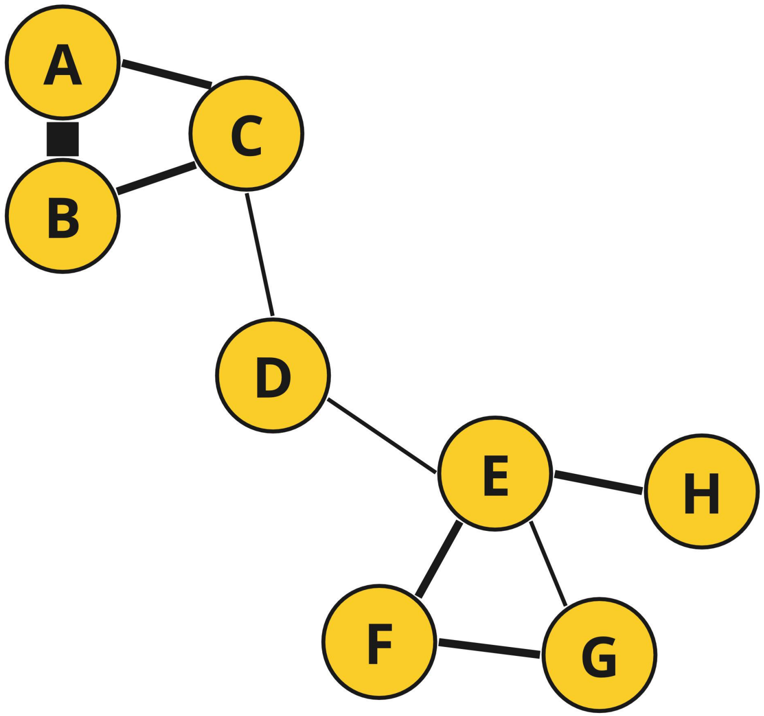
They tend to make **clusters** more obvious

Layouts can be deceptive

Physical proximity does **not** always equal **actual closeness**

In this example, H and F are **drawn** roughly the same distance from G

But H is really **farther** from G than F is (2 hops vs 1 hop)



Recap

- **Network** (or **graph**) is a collection of **nodes** (things) connected by **edges** (relationships)
- Nodes and edges can have **attributes**
- Edges can be **directed** or **undirected**, and can be **weighted** or **unweighted**
- Network analysis can involve both **visualization** and **statistical metrics**
- The three most common measures of a node's importance (**centrality**) are **degree, closeness, and betweenness**
- **Force-directed** visualization layouts emphasize **clusters** in a network