Introduction to Network Concepts

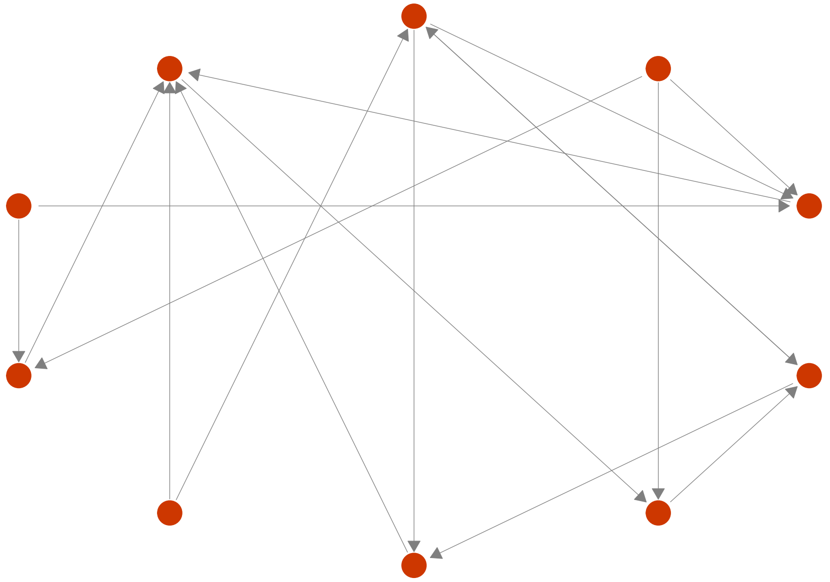
Today

- Introduction network terms
- The role of graph theory
- How matrices (ew math) help us

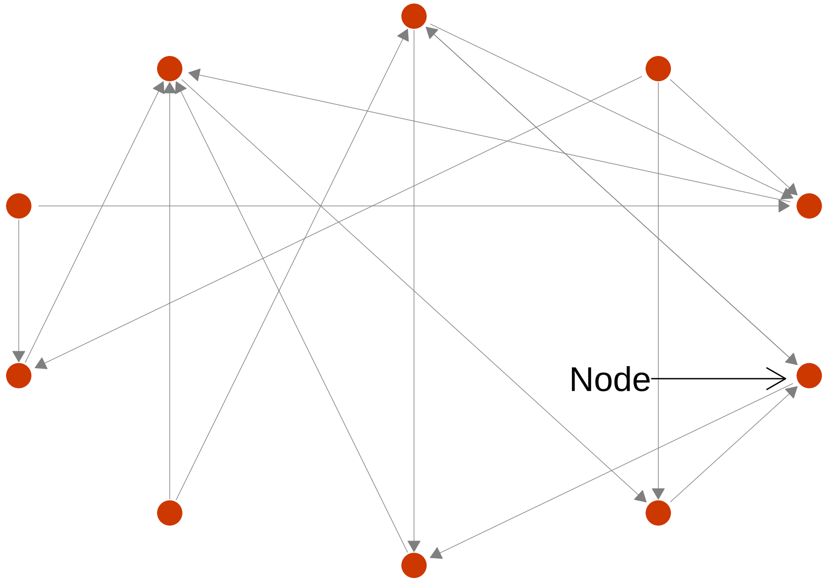
Nodes and Edges

Networks are made up of nodes that are connected with edges.

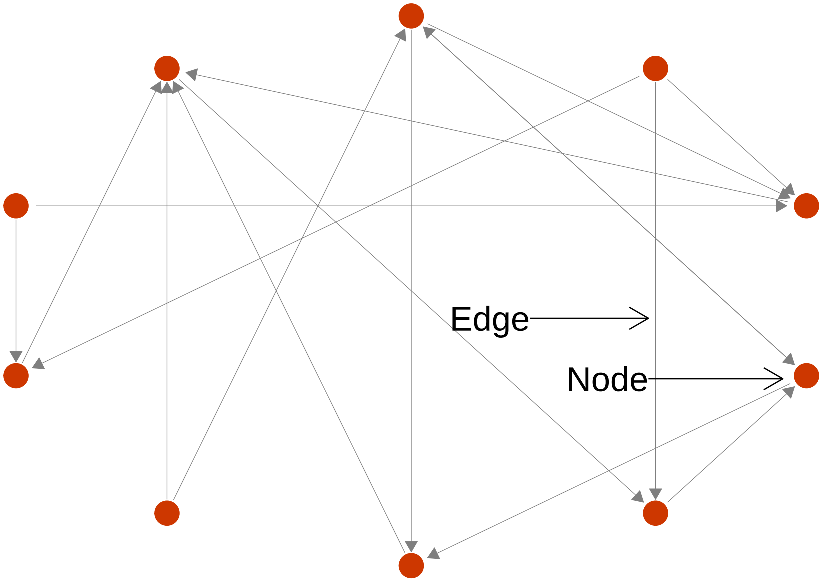
Network



Network



Network



Graph Theory

This makes sense visually, but in order to analyze it we need to use math

Graph theory

Graph:

- Is defined as a set of vertices (nodes) and edges: G(V, E)

Graph theory

Graph:

- Is defined as a set of vertices (nodes) and edges: G(V, E)
- Start with a set of vertices: V={A, B, C, D}

Graph theory

Graph:

- Is defined as a set of vertices (nodes) and edges: G(V, E)
- Start with a set of vertices: V={A, B, C, D}
- Edges connect two vertices, making them adjacent: (A, C)

Graph theory

Graph:

- Is defined as a set of vertices (nodes) and edges: G(V, E)
- Start with a set of vertices: V={A, B, C, D}
- Edges connect two vertices, making them adjacent: (A, C)
- E then the set of such connections: E={(A,C), (A,D), (B,C)}

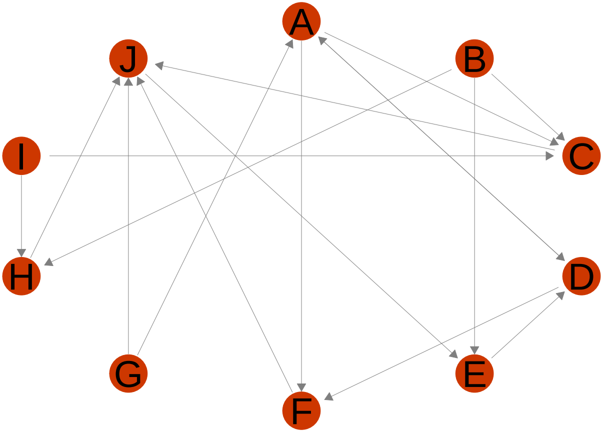
Graph theory

Graph:

- Is defined as a set of vertices (nodes) and edges: G(V, E)
- Start with a set of vertices: V={A, B, C, D}
- Edges connect two vertices, making them adjacent: (A, C)
- E then the set of such connections: E={(A,C), (A,D), (B,C)}
    - Note: these edges have a direction: (A, C), is different from (C,A)

Network as a Graph

E={(A,C), (A,D), (A,F), (B,C), (B,E), (B,H), (C,J), (D,A), (D,F), (E,D), (F,J), (G,A), (G,J), (H,J), (I,C), (I,H), (J,E)}

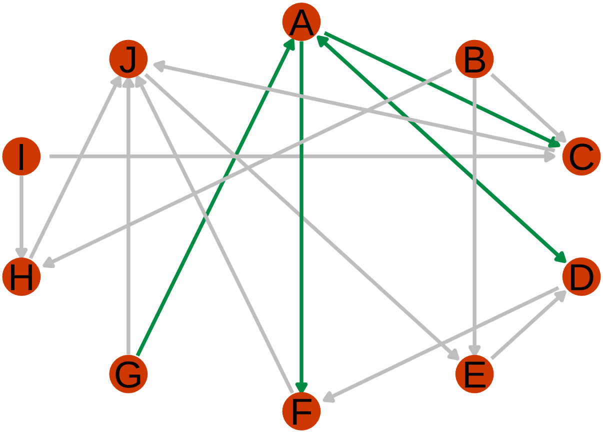


Adjacent, Incident, and Degree

- Two vertices that are connected by an edge are **adjacent**
- An edge that is connected to a vertice is said to be **incident** to it.
- The number of edges **incident** to a vertex is its **degree**

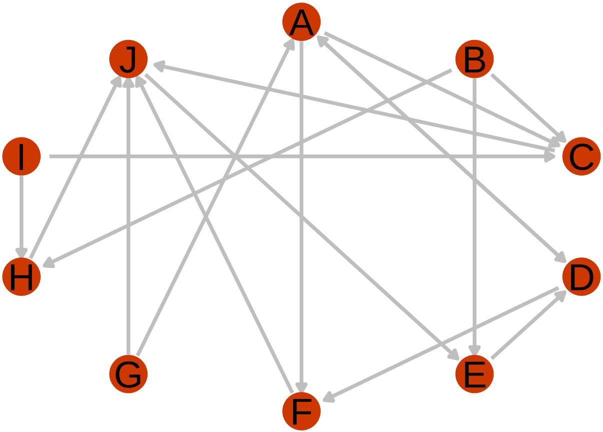
Degree

A has 4 edges incident to it, so it has degree of 4

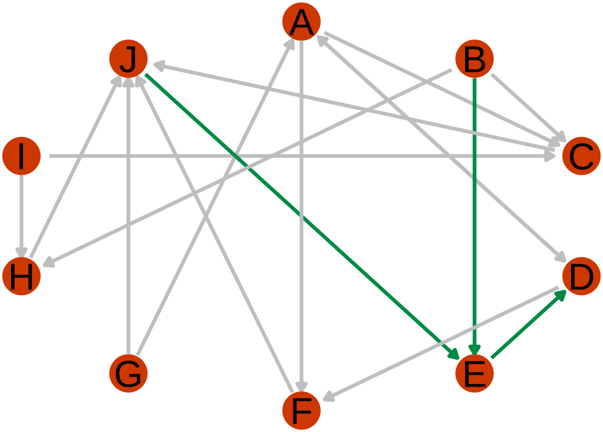


Degree for E

What is E’s degree?



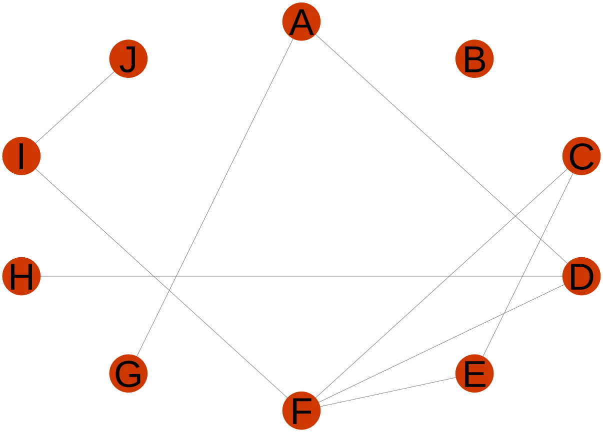
Degree for E



Undirected networks

It is also possible to have networks where edges do not have a direction:

E={(A,D), (A,G), (C,E), (C,F), (D,F), (D,H), (E,F), (F,I), (I,J)}



Directed vs Undirected Network

Whether a network is directed or undirected is dependent on the types of relationships:

Directed:

- Parent-Child
- Donor-Donee

Undirected

- Friend (hopefully)
- Classmate
- Ally

Characteristics for nodes

What are some things that might matter if you are at a node? - Where can you move to? - How fast can you get there? - Who can you reach?

Moving along a network

We can move from one vertex (node) to another through edges.

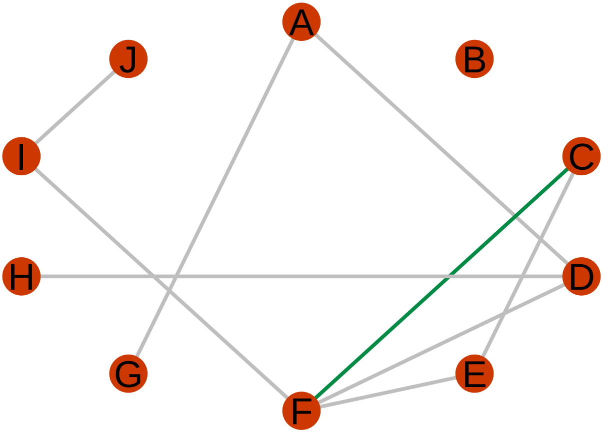
- A walk is a sequence of **adjacent** vertices:
  - G-A-D-H
  - F-C-E-F-I
  - H-D-A
- A path is a sequnce of **adjacent** vertices where none are repeated:
  - G-A-D-H
  - H-D-A
- All paths are walks, but not all walks are paths.

The length of paths and walks

- The **length** of a walk or paths is how many edges are in them.
- The shortest path is called a **geodesic**.
- The **(geodesic) distance** between two vertices is the length their **geodesic**

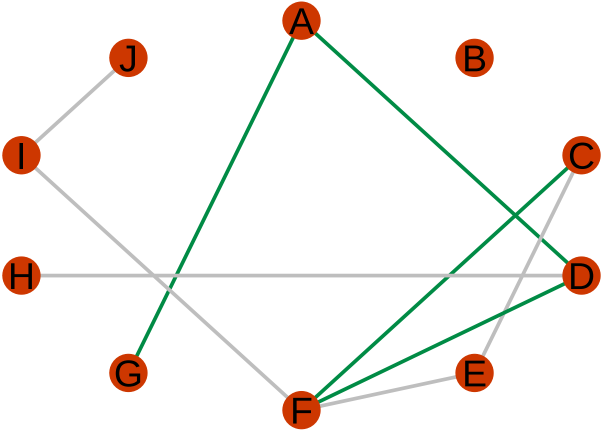
Geodesic Distance

The distance between F and C is 1



Geodesic Distance

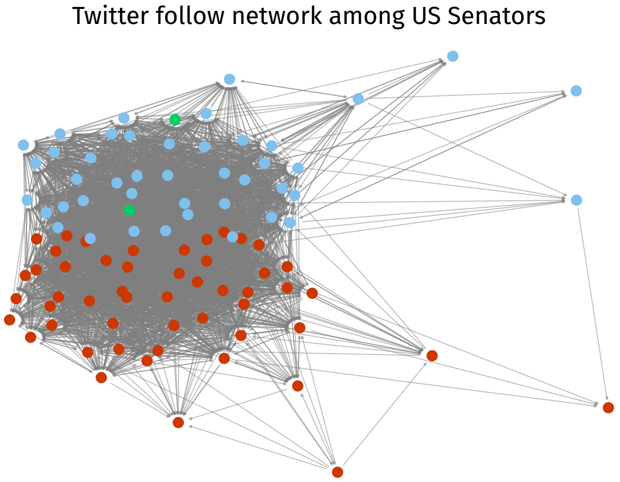
The distance between G and C is 4



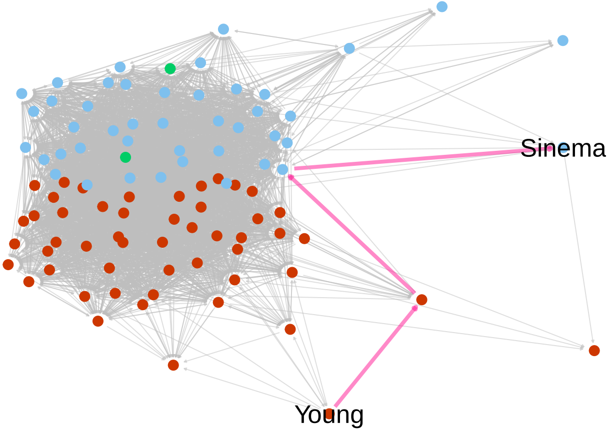
Why care about distance?

- Distance is viewed as how ‘close’ two vertices are.
- Information spreads faster between two vertices that are closer

Looking at some real data: Twitter Senate Data

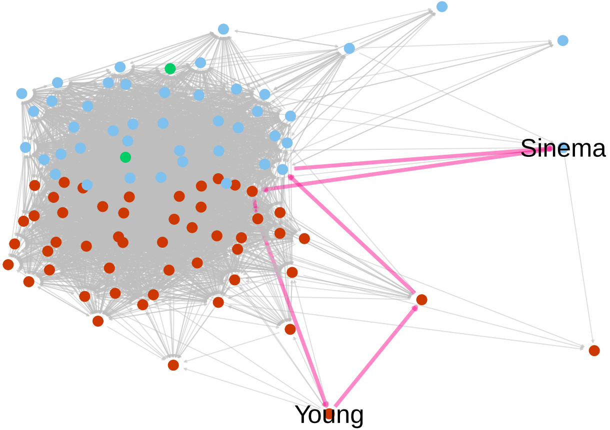


Looking at some real data



Young follows Portman, who follows Booker, who follows Sinema

Looking at some real data



Sinema follows Scott, who follows Cornyn, who follows Young

What can we learn?

- The mean distance between any two senators is 1.45
- The median distance between any two senators is 1
- The median degree is 110.5

Degree for Directed Networks

Directed networks have both **outdegree** (how many edges start at a vertex) and **indegree** (how many edges point at a vertex):

- Median outdegree (how many each Senator follows): 51
- Median indegree (how many followers each Senator has): 57

Who is the most popular?

	Following	Followers
Daniel Sullivan	59	55
Lisa Murkowski	77	74
Gordon Jones	52	42
Richard Shelby	58	65
John Boozman	36	63

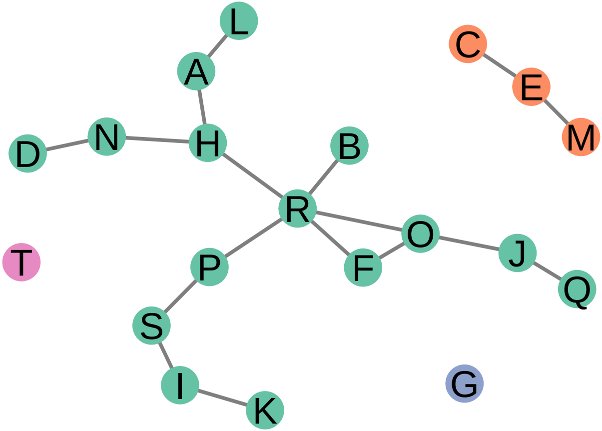
- Mark Warner follows the most (follows 87)
- Susan Collins is followed by the most (follows 78)

Who can reach who?

- A vertex with no incident edges is known as an **isolate**
- We can also group vertices by who they can reach.

A **component** is the maximal (largest) group of vertices where every vertex within it can reach every other vertex.

Some Components



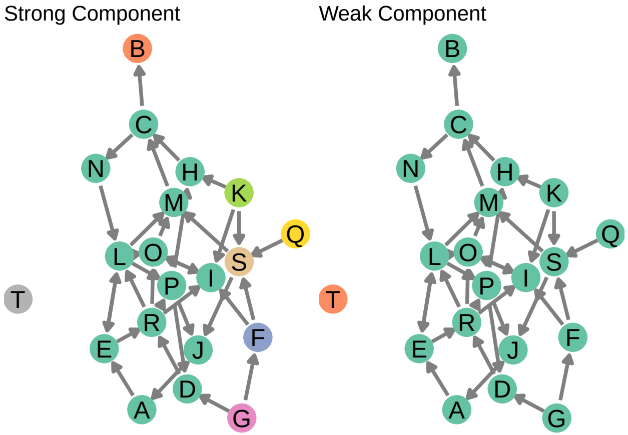
This network has 4 components, the largest has 15 vertices in it.

Components with Directed Graphs

When a network is directed, then we need to think about direction.

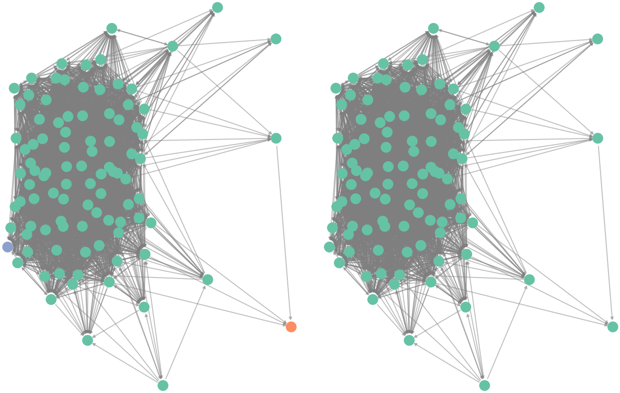
- **Weak Component:** Is a component if we disregard direction of edges.
- **Strong Component:** Is a component if we follow direction of edges

Strong/Weak Components:



Real Life Data

Strong Component　　　　Weak Component



Bring the math back in

How do we organize this data? An **adjacency matrix**

Matrix Referesher

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

- A matrix is a rectangular collection of numbers ($n \times m$ in size).
- Each number (element) is given a position by what row and column it is.
- The 1 above is in row 2, column 3.

Adjacency matrix

- For now our matrices will be square: $n \times n$ where n is the number of vertices.
- A 1 means an edge exists from that row to that column, a 0 means no edge exists.

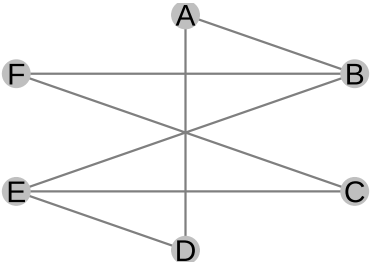
Adjacency matrix

- For now our matrices will be square: $n \times n$ where n is the number of vertices.
- A 1 means an edge exists from that row to that column, a 0 means no edge exists.

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

- In $\mathbf{A}$ there is an edge from node 3 to node 2, and from node 1 to node 3 (directed).

Adjacency Matrix vs Network Graph



	A	B	C	D	E	F
A	0	1	0	1	0	0
B	1	0	0	0	1	1
C	0	0	0	0	1	1
D	1	0	0	0	1	0
E	0	1	1	1	0	0
F	0	1	1	0	0	0

Undirected Adjacency Matrix

- For undirected graphs the adjacency matrix is mirrored across the diagonal
- Meaning that the entry for $a_{i,j}$ is the same as for $a_{j,i}$

Senator Adjacency Matrix

	Daniel Sullivan	Lisa Murkowski	Doug Jones
Daniel Sullivan	0	1	0
Lisa Murkowski	1	0	0
Doug Jones	0	1	0
Richard Shelby	0	1	1
John Boozman	1	1	0