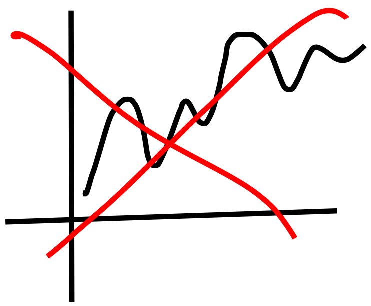
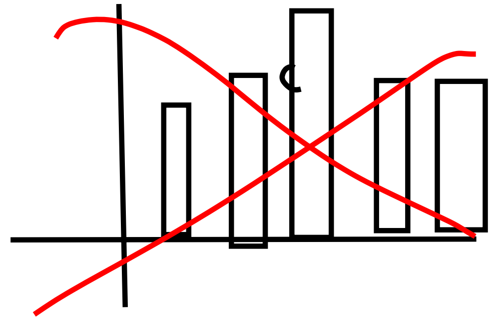
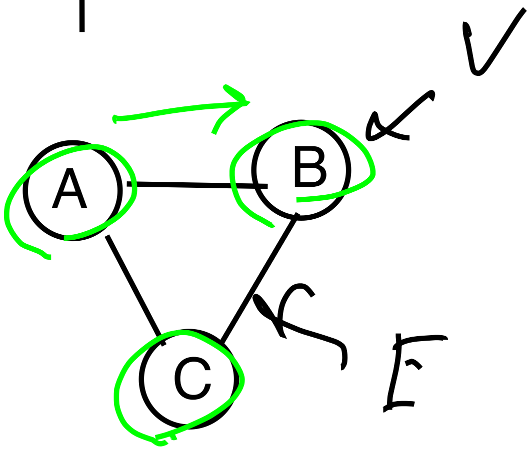
Graph Theory

Darren Reger Lecture for Galvanize DSI

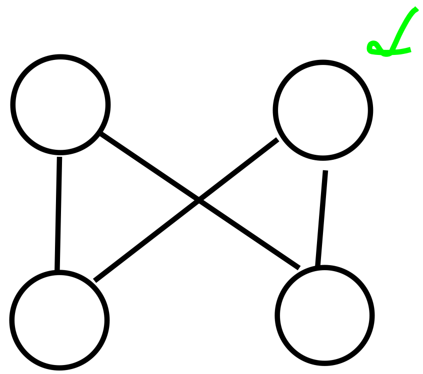
What is a graph?







Order 3
Size 3



Order 4
Size 4

$$G = (V,E)$$

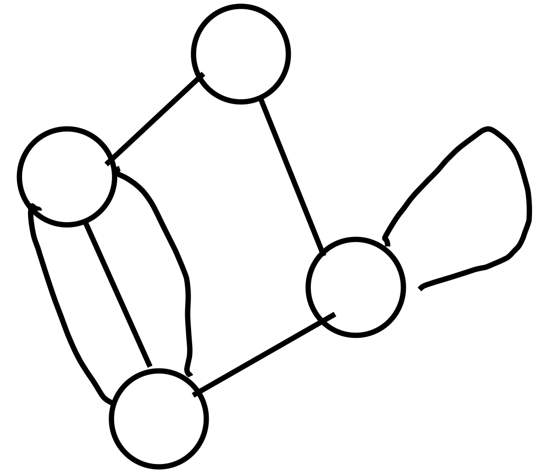
V is a (finite) set of elements
E is a sets of 2 subsets of V

$V = \{A, B, C\}$

$E = \{\{A,B\}, \{B,C\}, \{A,C\}\}$

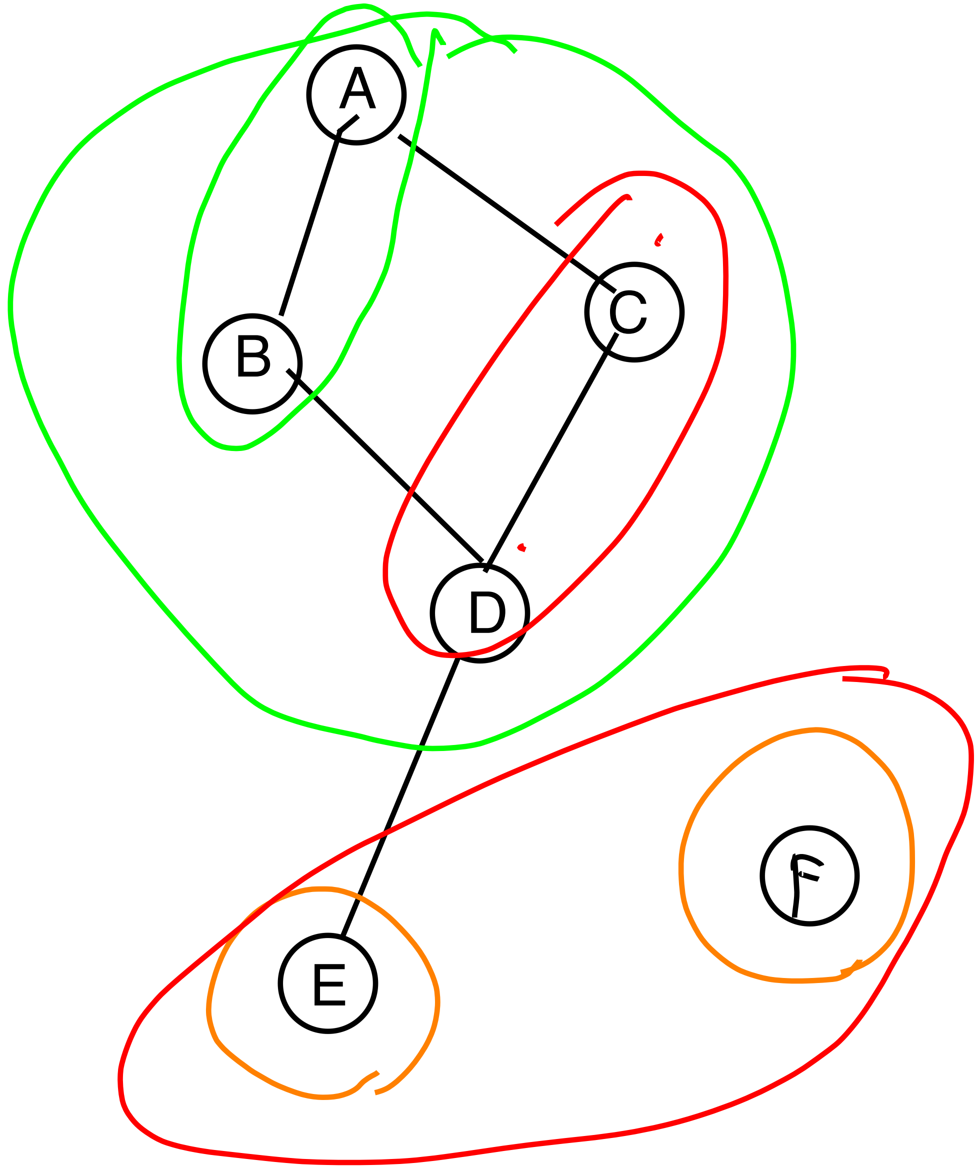
Simple graphs

~~Loops~~
~~Multiple Edges~~



Terminology

Neighbors
Degree
Path
Complete
Connected
Subgraph
Conn. Component

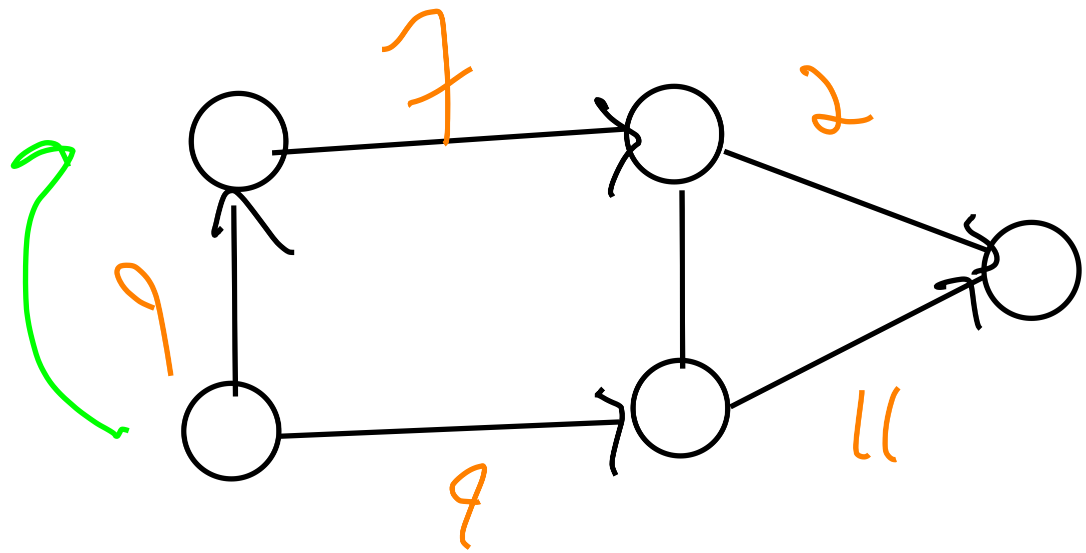


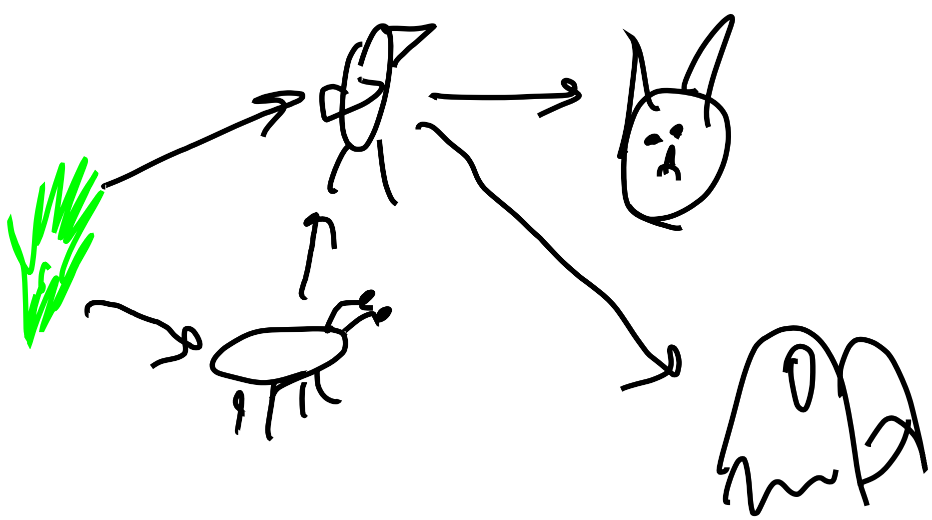
Types of Graphs

Directed

Undirected

Weighted





How do we represent graphs in our computer?

Edge List

(A, B, 3)

(A, C, 5)

(B, C, 7)

(B, D, 9)

Adjacency Matrix

	A	B	C	D
A	0	1	1	0
B	1	0	1	1
C	1	1	0	0
D	0	1	0	0

	A	B	C	D
A	0	3	5	$\boxed{\infty}$
B	3	0	7	9
C	5	7	0	$\boxed{\infty}$
D	$\boxed{\infty}$	9	$\boxed{\infty}$	0

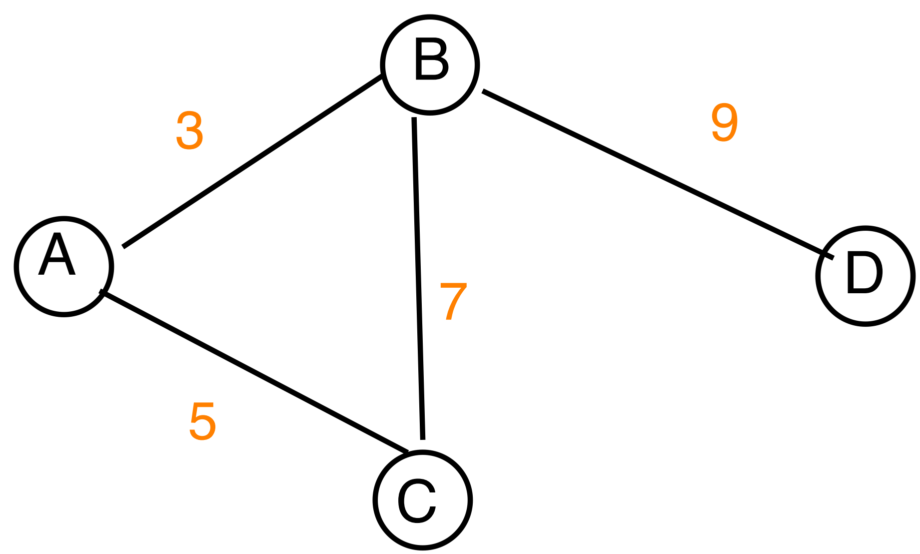
Adjacency list

A: (B, 3) (C, 5)

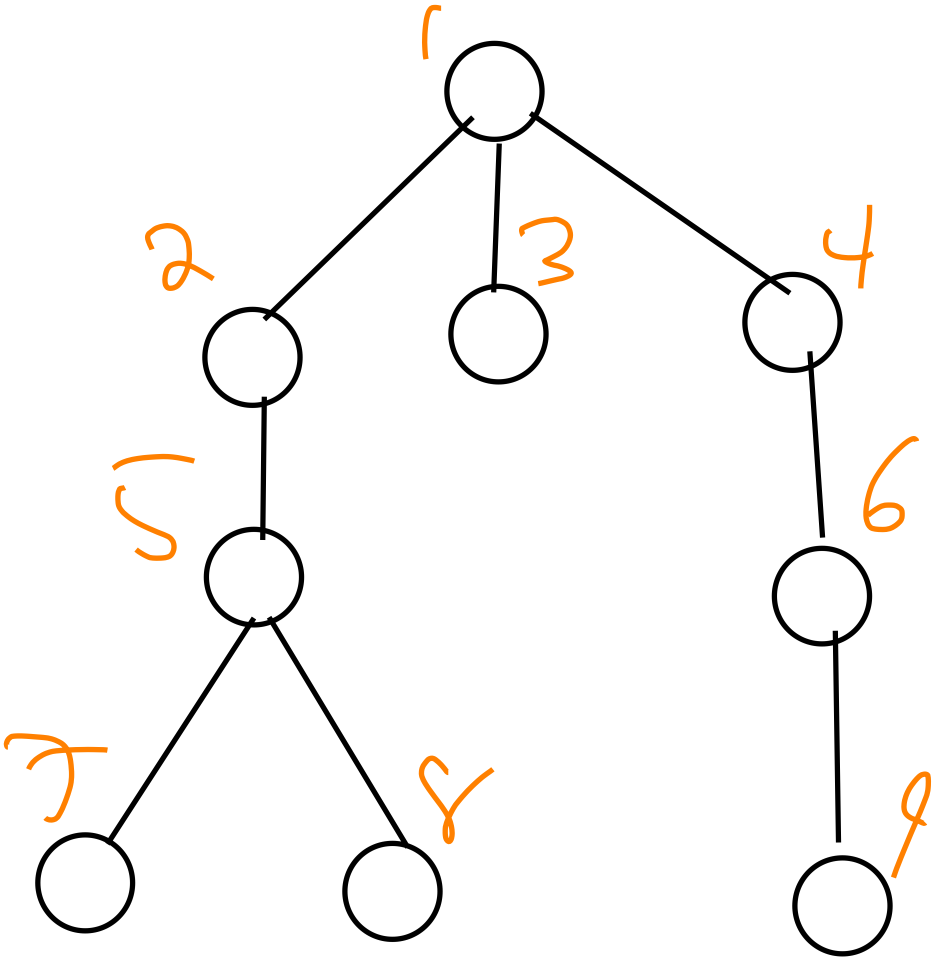
B: (A, 3) (C, 7) (D, 9)

C: (B, 7) (A, 5)

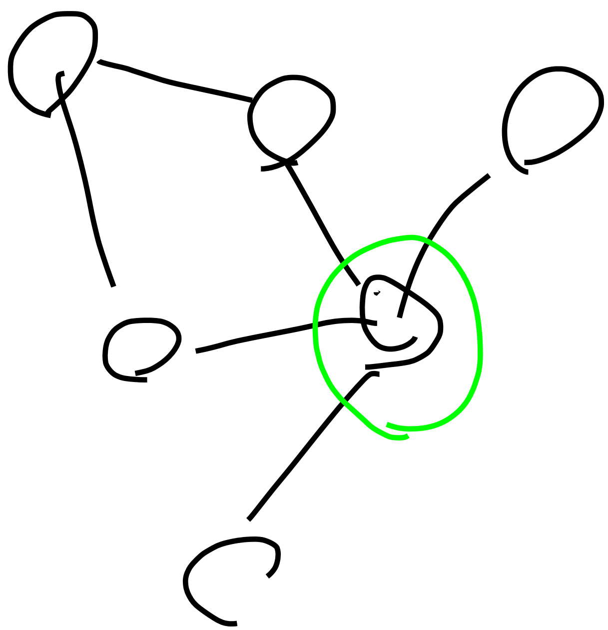
D: (B, 9)



Breadth First Search

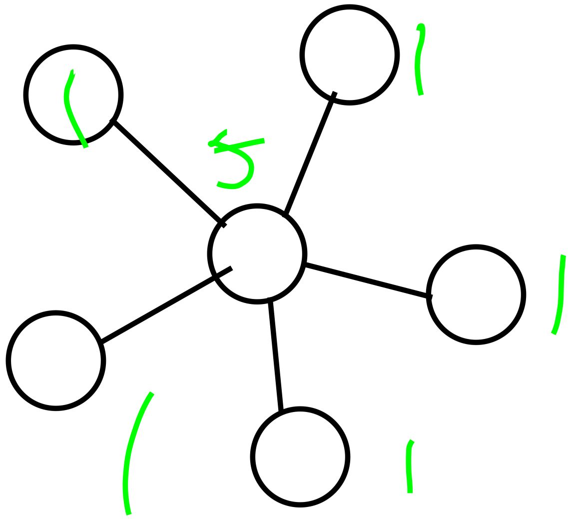


How Important Is a Given Node?

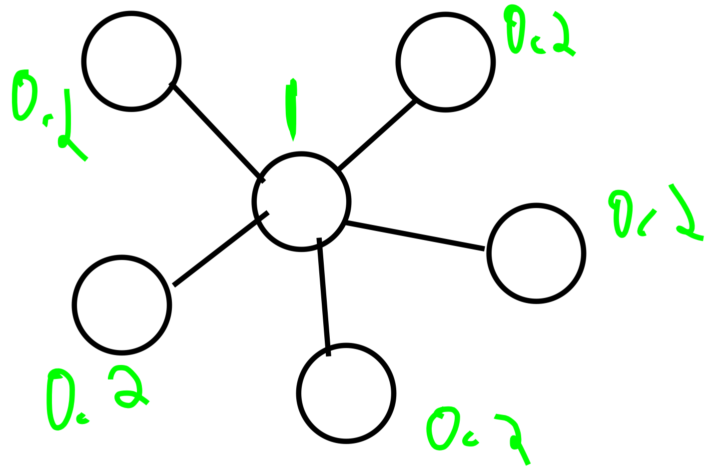


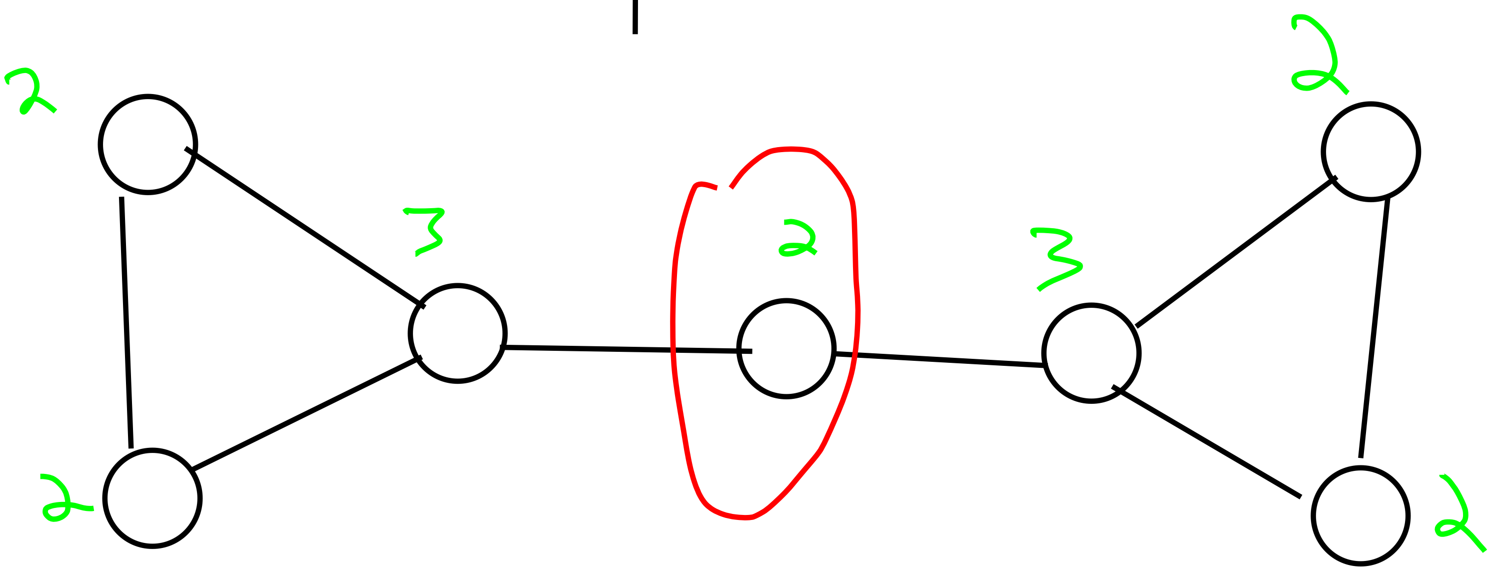
Centrality

Degree - # of Connections



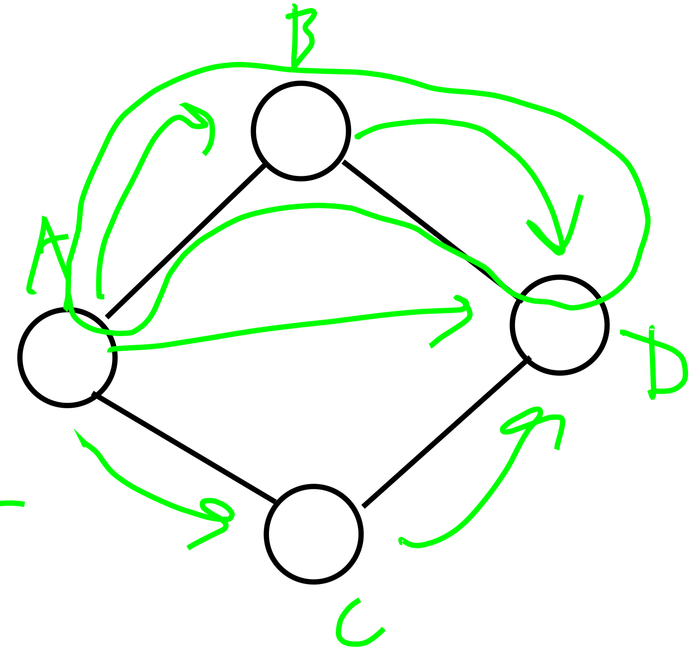
Normalize - divide by # of nodes





Betweenness Centrality

$$C_B(i) = \sum_{i \neq j \neq k} \frac{g_{jk}(i)}{g_{jk}}$$

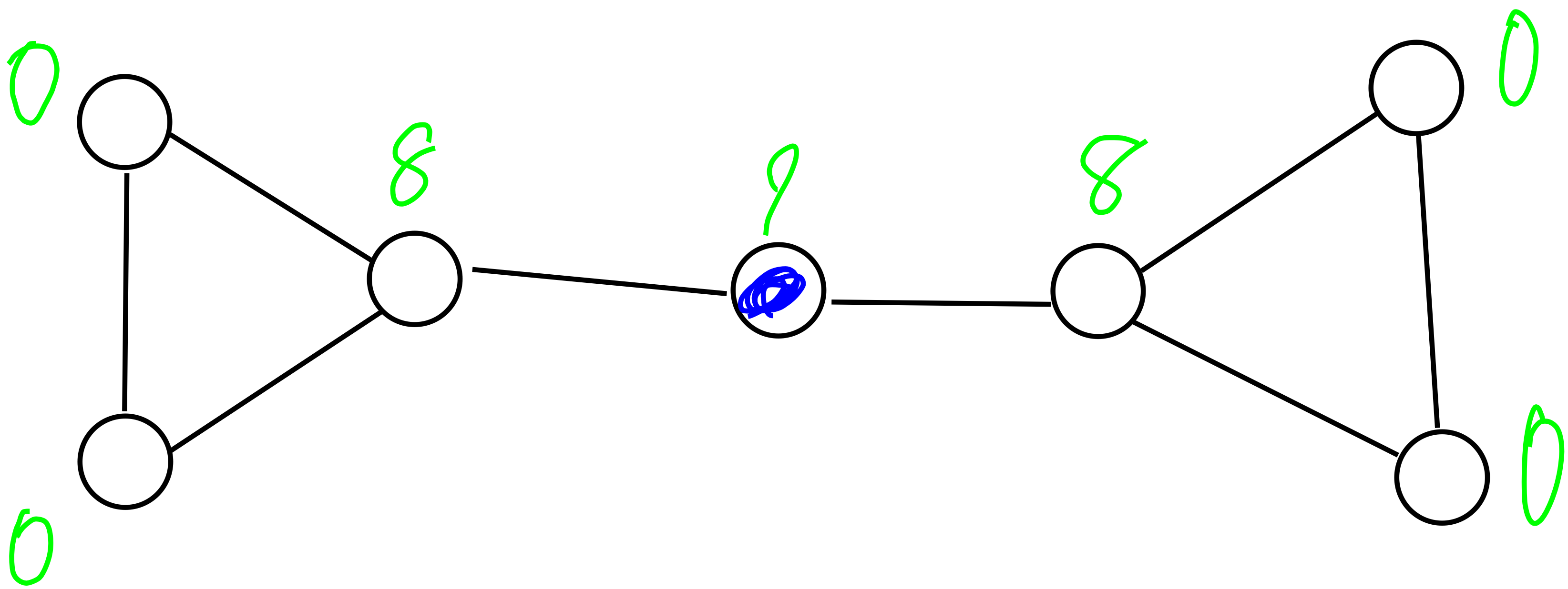


$\frac{1}{2}$

Normalized -

Divide by # of pairs of vertices excluding the vertex itself

$$C'_B(i) = \frac{C_B(i)}{[(n-1)(n-2)/2]}$$



Eigenvector Centrality

$$C_e(i) = \frac{1}{\lambda} \sum_{j=1}^{n} A_{j,i} C_e(j)$$

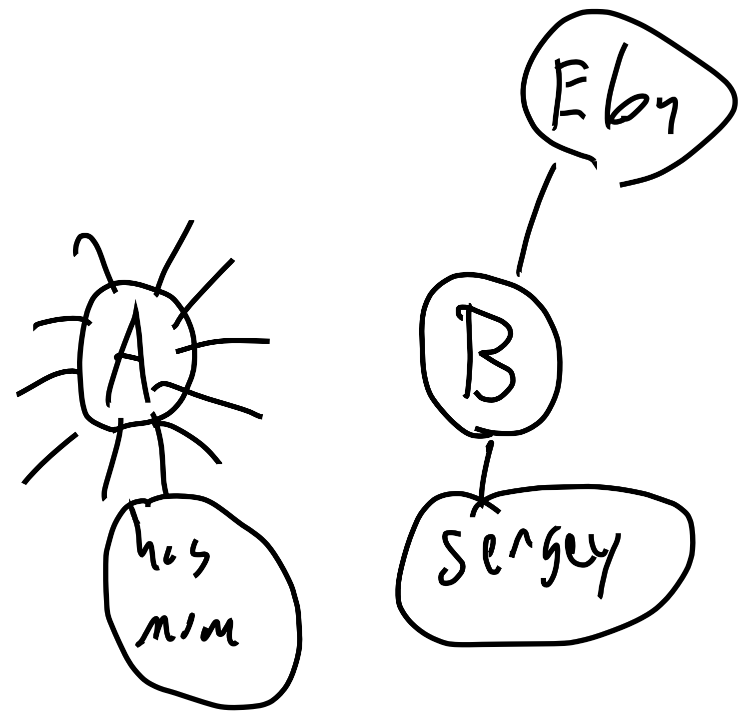
adj matrix

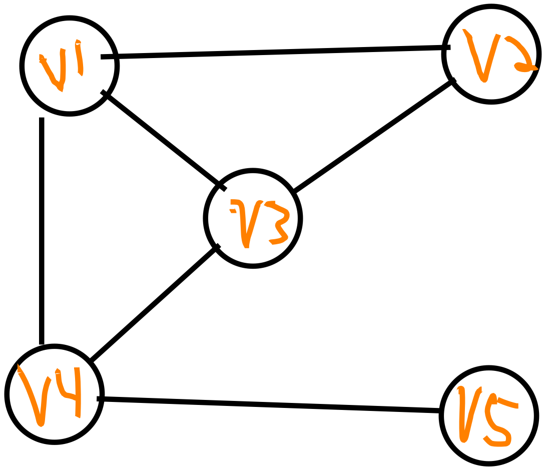
$$\mathbf{C}_e = (C_e(i), C_e(j), C_e(k), \ldots, C_e(n))^T$$

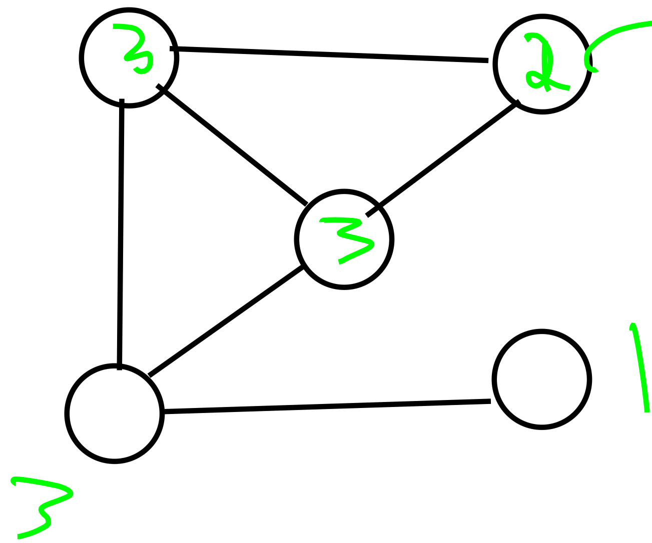
$$\lambda \mathbf{C}_e = A^T \mathbf{C}_e \implies \lambda V = AV$$

Which eigenvector?

Perron-Frobenius Theorem







$$\begin{bmatrix} 3 \\ 2 \\ 3 \\ 3 \\ 1 \end{bmatrix}$$

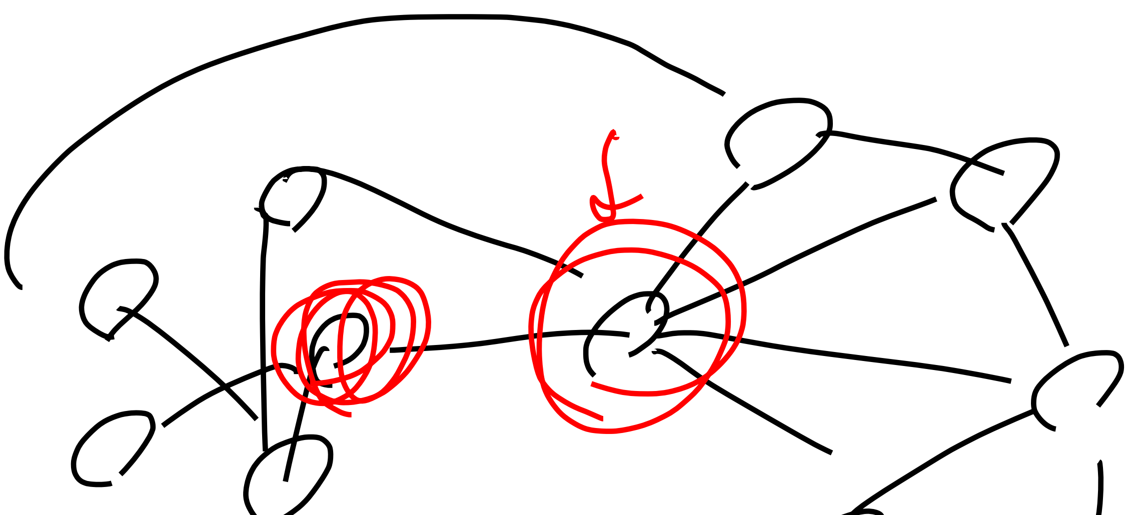
$$Ax = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 3 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \\ 8 \\ 7 \\ 3 \end{bmatrix}$$

$$3 \cdot 0 + 2 \cdot 1 + 3 \cdot 1 + 3 \cdot 1 + 1 \cdot 0$$

$$A \cdot \begin{bmatrix} 8 \\ 6 \\ 8 \\ 7 \\ 3 \end{bmatrix} = \begin{bmatrix} 21 \\ 16 \\ 21 \\ 19 \\ 7 \end{bmatrix}$$

$$\frac{[21, 16, 21, 19, 7]}{\| [21, 16, 21, 19, 7] \|}$$

$$\begin{bmatrix} 0.53 \\ 0.41 \\ 0.53 \\ 0.48 \\ 0.18 \end{bmatrix}$$



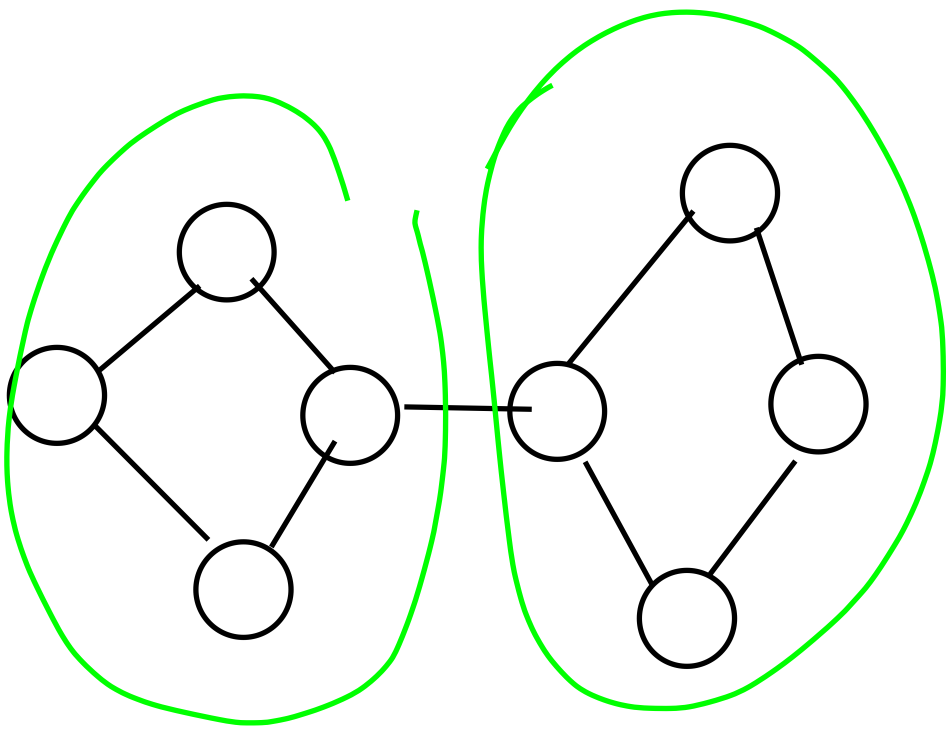
Communities

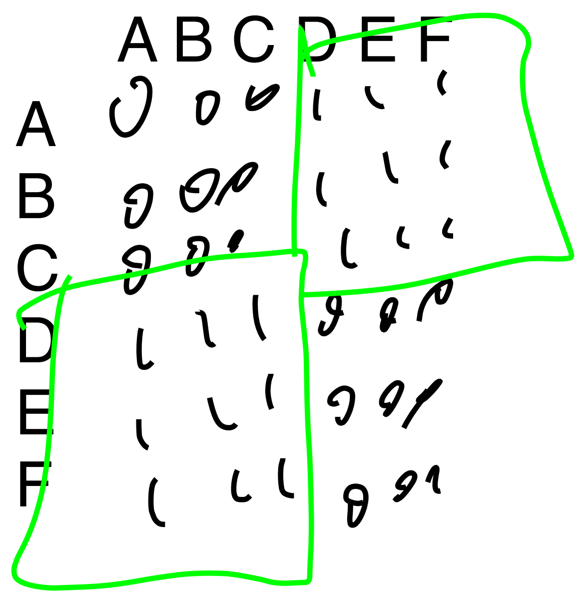
Mutual Ties - within the group know each other

Compactness - reach other members in few steps

Dense Edges - high frequency of edges within groups

Separation from other groups - frequency within groups higher than out of group

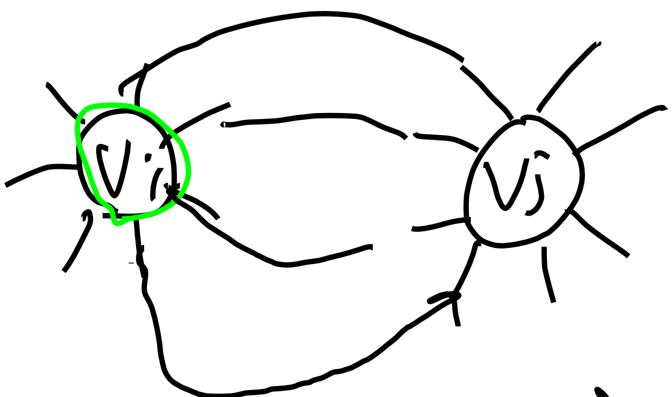




Modularity

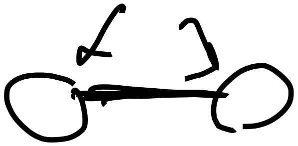
Measure that defines how likely the group structure seen was due to random chance

$G = (V, E) \qquad |E| = m$



$$\frac{d_j}{\sum_{i=1}^{n} d_i} = \frac{d_j}{2m} + \frac{d_j}{2m} + \frac{d_j}{2m}$$

$2m \propto$

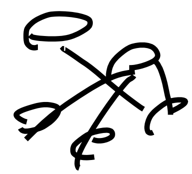


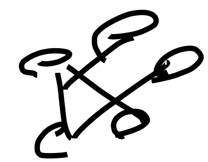
$$\text{expected \# of edges} = \frac{d_i d_j}{2m}$$

Modularity = Q = $\sum$ (Observed fraction of links within the group — Expected fraction of links in the group)

$$Q = \frac{1}{2m}\sum_{c\in C}\left[\sum_{i,j\in c} A_{i,j} - \frac{d_i d_j}{2m}\right]$$

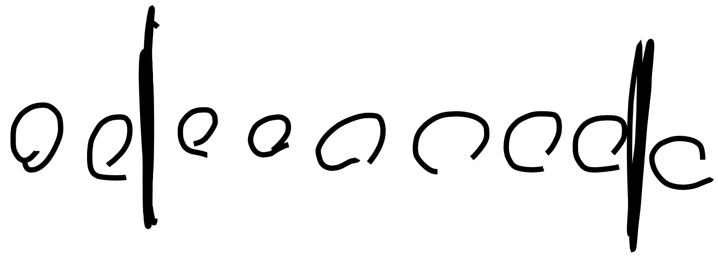
adj matrix





Graph Partitioning Problem

It's combinatorial!



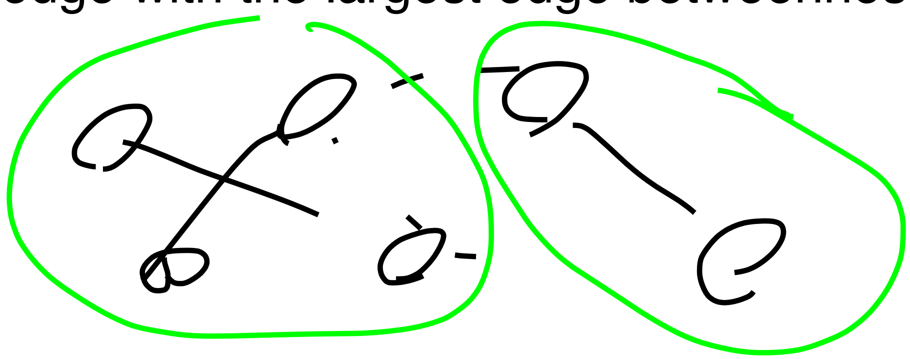
$$B_{20} \approx 5{,}000{,}000{,}000$$

Heuristic Approach

Focus on the edges that connect communities

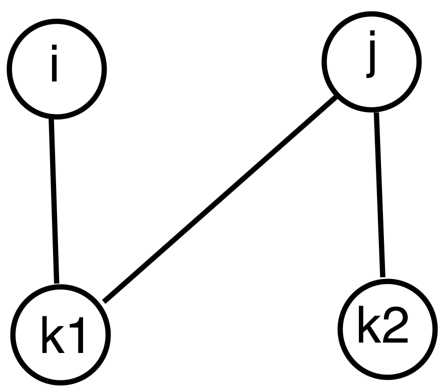
Girvan-Newman 2004

For all edges, compute the edge betweenness and remove the edge with the largest edge betweenness



Node Similarity

The # of neighbors that 2 nodes share



$$n_{ij} = \sum_{k} A_{ik} A_{jk}$$

$$n_{ij} = (A_{ik_1} A_{jk_1}) + (A_{ik_2} A_{jk_2})$$

$$1 \cdot 1 \quad + \quad 1 \cdot 0$$

$$= 1$$

$$k = \frac{n(A \cap B)}{\sqrt{n(A) \times n(B)}}$$

$$d_{ij} = \sum_{k} (A_{ik} - A_{jk})^2$$

Hierarchical Clustering

Assign each vertex to a group of it's own

Find 2 groups with the highest similarity and join them together

Calculate the similarity between groups:

1) single linkage
2) complete linkage
3) average linkage

Repeat until all in 1 group