Global Measures

CME4422 Graph Theory

Permutations

- Order doesn't matter
- How do we select r objects from a pool of n?
- Assume we don't replace objects; they can be chosen only once.
- There are $P(n,r) = n!/(n-r)!$

Example

- I have 16 numbered balls. How do I select three of them?
- P(16,3) = 16!/(13!) = 14.15.16

Clustering Measures

- Local Clustering Coefficient
- Global Clustering Coefficient
- Clustering is a measure of Transitivity:

If A and B are connected and B and C are connected, then A and C are connected.

Local Clustering Coefficient

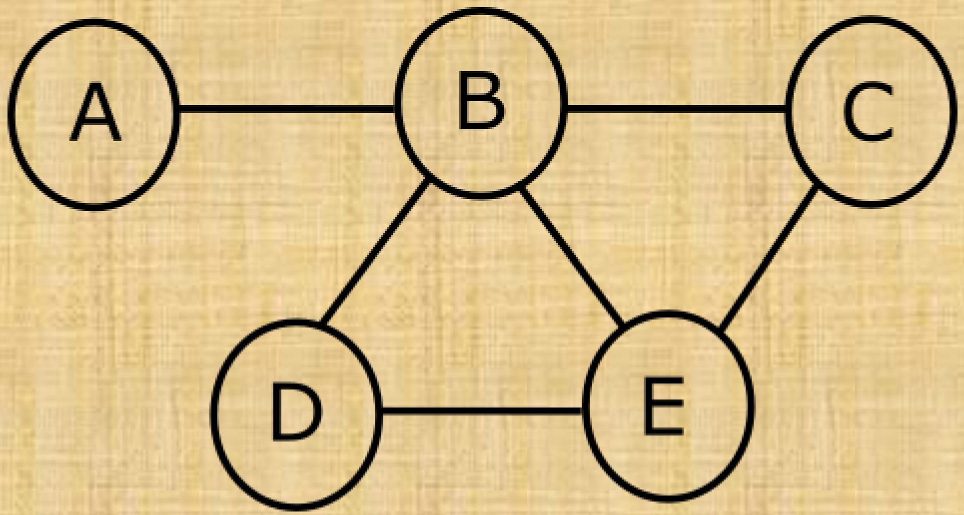
- $C_{cc}(v_i) = \frac{\#\ of\ v_i's\ connected\ neighbors}{\#\ of\ v_i's\ neighbor\ pairs}$

Example: Find the local clustering coefficient for node B.

- Neighbors of node B: A,C,E,D
- 2 connections: CE,ED.
- 4 nodes can form (4x3)/2=6 neighbor pairs
- $C_{cc}(B)$=2/6 = 0.33

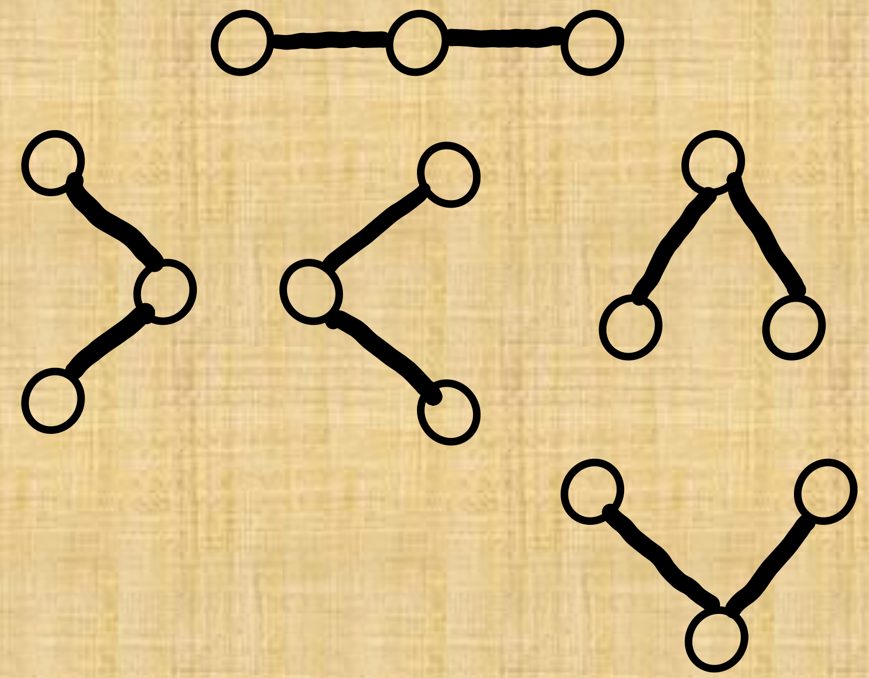
P(4,2)

Undirected Graph:
AE=EA



Triples and Triangles

- Connected Triples



- Triangles

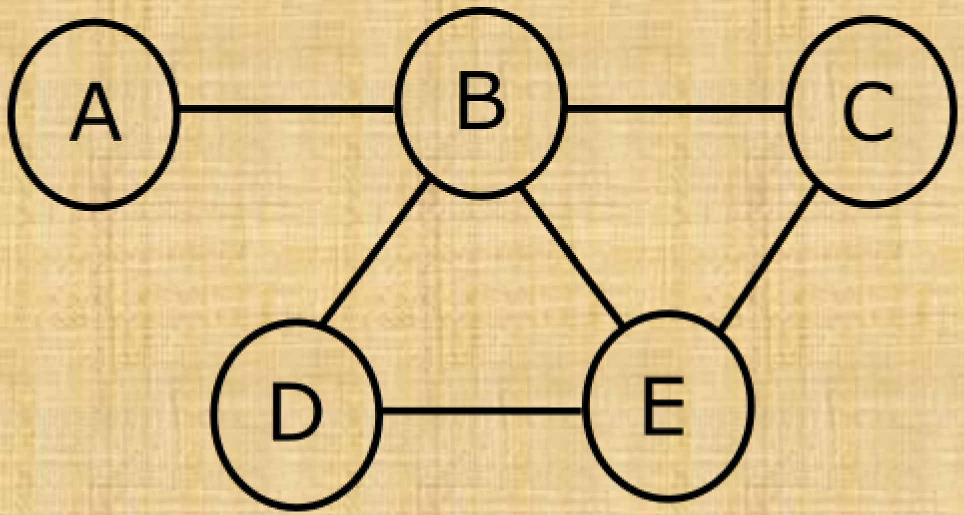
Each triangle has 3 triples.

Another Definition

- $C_{cc}(v_i) = \frac{\#\ of\ triangles\ that\ include\ v_i}{\#\ of\ triples\ with\ center\ v_i}$

Example: Find the local clustering coefficient for node B.

- Triangles which include B: BCE, BDE.
- Triples with B at the center: (4x3)/2



P(4,2)

Undirected Graph:
AE=EA

Global Clustering Coefficient

- $C_{cc}(G) = \dfrac{\#\ of\ triangles\ \times 3}{\#\ of\ triples}$

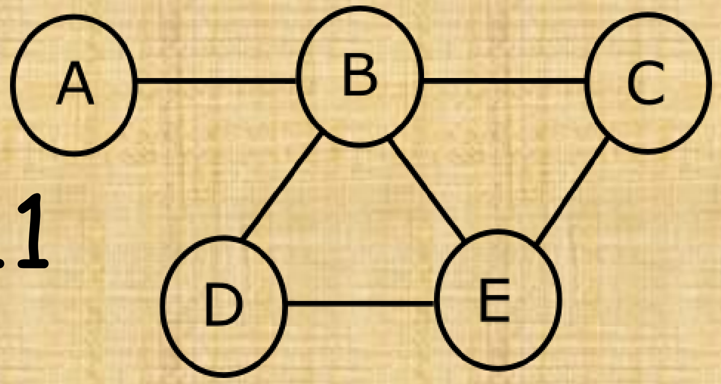
of Triples

- The # of triples in a graph is given by:

$$\frac{1}{2}\sum_{i}^{|V|}[\deg(v_i) \times (\deg(v_i) - 1)]$$

- For the graph to the right:

# of triples
= (1/2)(1×0+4×3+2×1+3×2+2×1) = 11

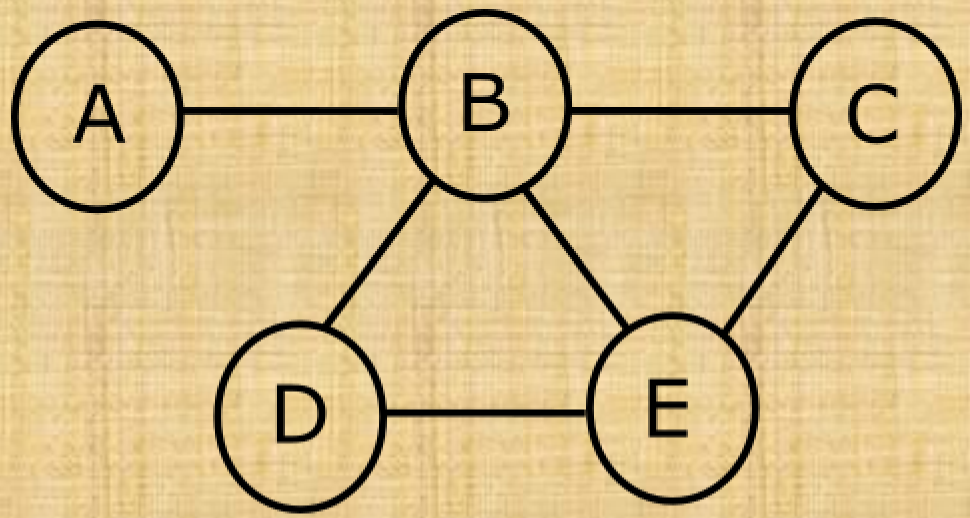


Example

- There are 2 triangles: BDE,BEC.
- There are 11 triples(see slide above).

$C_{cc}(G)=(2x3)/11=0.54$

Note: The triples outside the triangles: ABC,ABD,ABE,CBD,CED

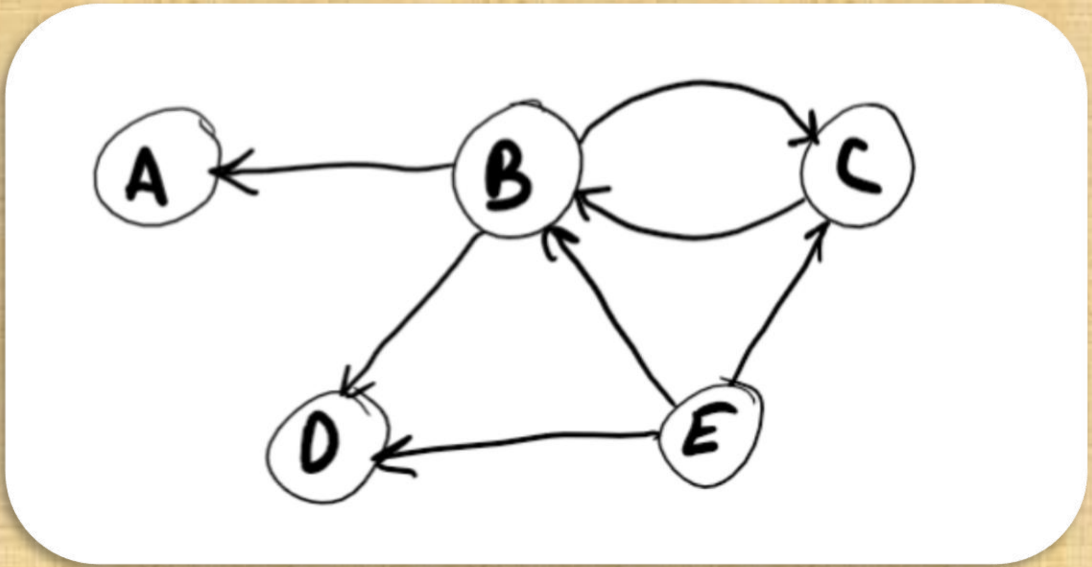


Reciprocity

- Defined only for directed graphs.
- It's a measure of reciprocity between nodes.
- $C_{rec}(G) = \dfrac{\#\ of\ existing\ reciprocal\ edge\ pairs}{\max \#\ of\ edge\ pairs\ \left(\frac{|E|}{2}\right)}$

Example

- |E|=7, so max. # of reciprocal pairs can be 3.5.
- There is 1 reciprocal edge pair.
- Crec(G) = 1/3.5 = 0.29



Average Degree

Undirected Graphs

- $\langle \deg(i) \rangle = \frac{1}{|V|} \sum_i^{|V|} \deg(i) = \frac{2|E|}{|V|}$

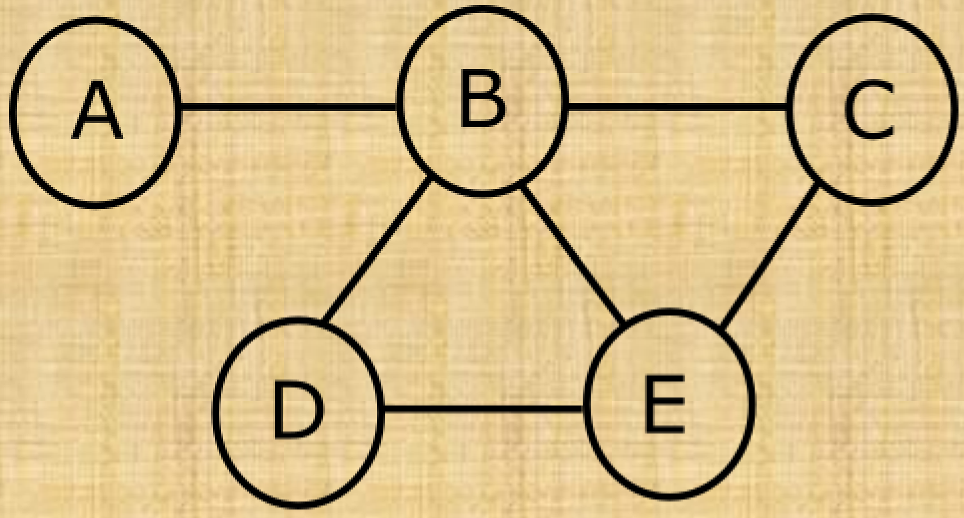
Directed Graphs

- $\langle \deg_{in}(i) \rangle = \frac{1}{|V|} \sum_i^{|V|} \deg_{in}(i) = \frac{|E|}{|V|}$
- $\langle \deg_{out}(i) \rangle = \frac{1}{|V|} \sum_i^{|V|} \deg_{out}(i) = \frac{|E|}{|V|}$

Example

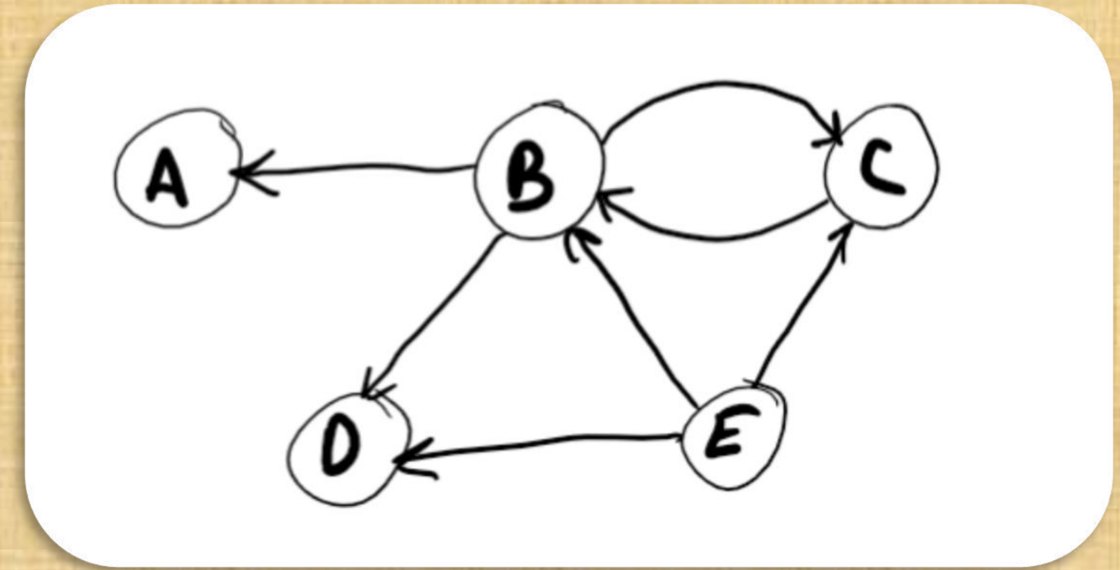
- $|\mathrm{V}| = 5$
- $\langle \deg(i) \rangle$ = (1/5)(1+4+2+2+3)

= 12/5 = 2.4

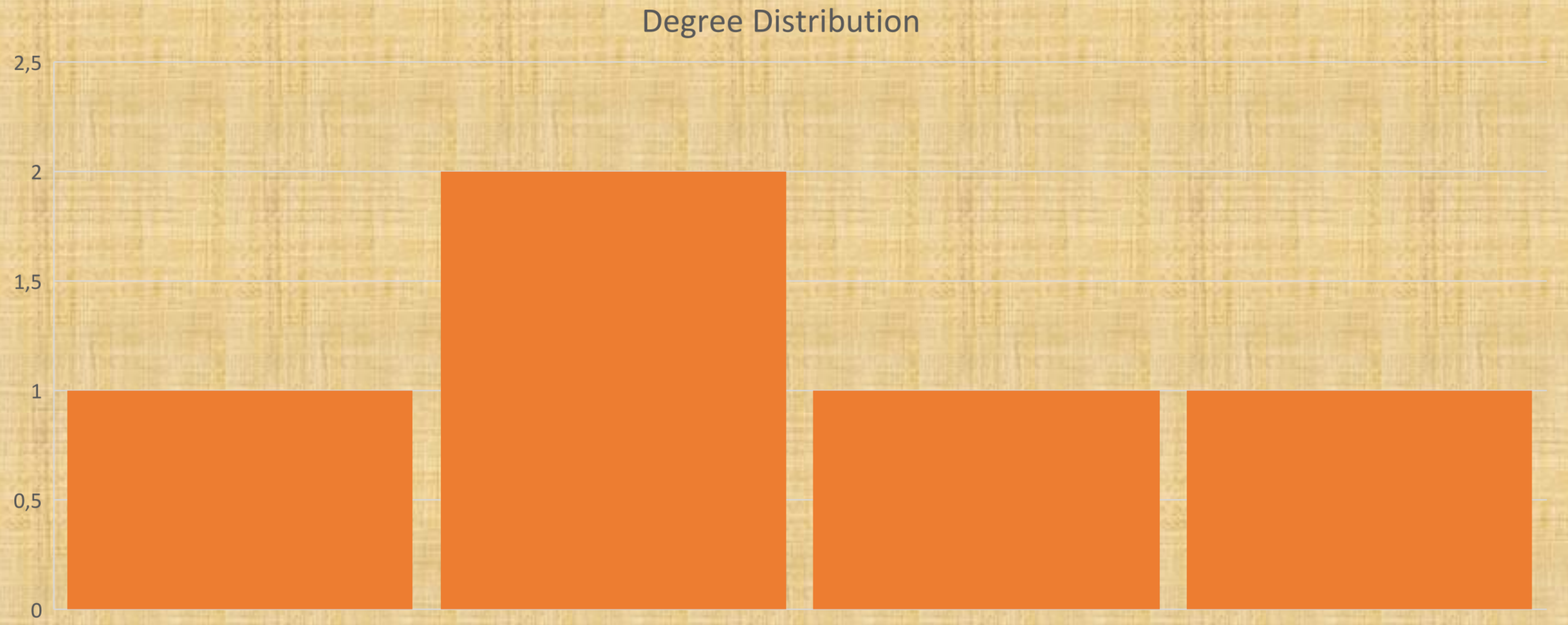


Example

- |V| = 5, |E| = 7
- $\langle \deg_{in}(i) \rangle = \frac{1}{|V|}\sum_i^{|V|} \deg_{in}(i) = \left(\frac{1}{5}\right)(1 + 2 + 2 + 2 + 0) = \frac{7}{5} = \frac{|E|}{|V|}$
- $\langle \deg_{out}(i) \rangle = \frac{1}{|V|}\sum_i^{|V|} \deg_{out}(i) = (1/5)(0 + 3 + 1 + 0 + 3) = \frac{7}{5} = \frac{|E|}{|V|}$



Degree Distribution



Social Similarity Coefficient

- $r = \frac{4\langle k_i k_j \rangle - \langle k_i + k_j \rangle^2}{2\left\langle k_i^2 + k_j^2 \right\rangle - \langle k_i + k_j \rangle^2}$
- $r \epsilon [-1,1]$.
- r>0: high assortativity
- r = 0: no correlation
- r<0: disassortativity

Density

- Density is a measure of connectedness of the graph.

$$d = \frac{\#\ of\ existing\ edges\ |E|}{\#\ of\ max.possible\ edges\ (E_{max})}$$

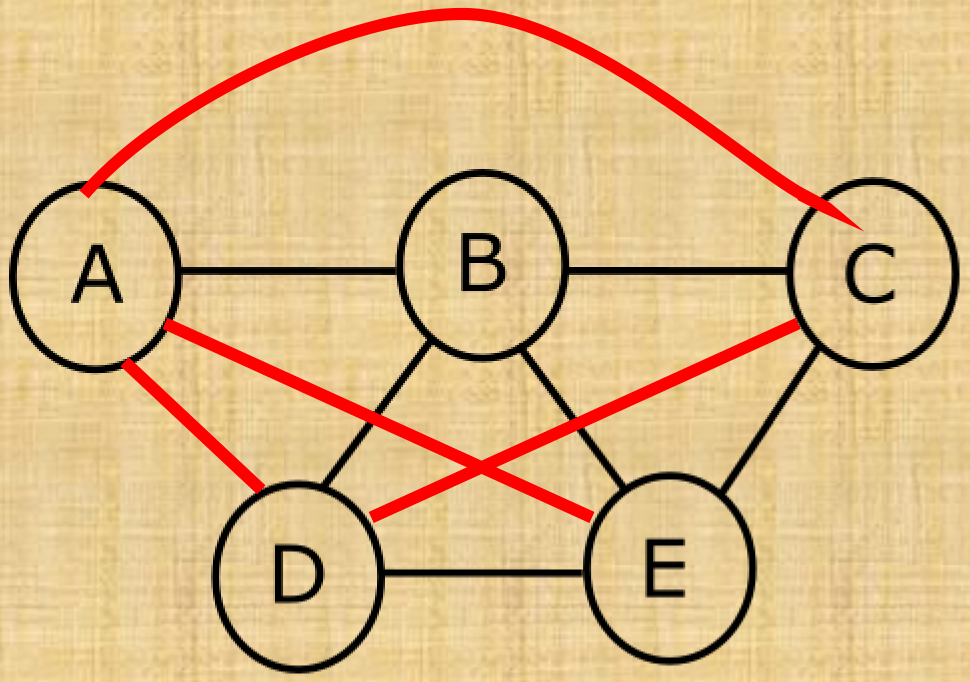
Undirected Graphs

- Max. # of edges E_{max} is $|V|(|V|-1)/2$
- # of existing edges is $|E|$

$$d = \frac{|E|}{|V|(|V|-1)/2}$$

Example

- $d = 6/[(5\times4)/2]$
  $= 0.6$



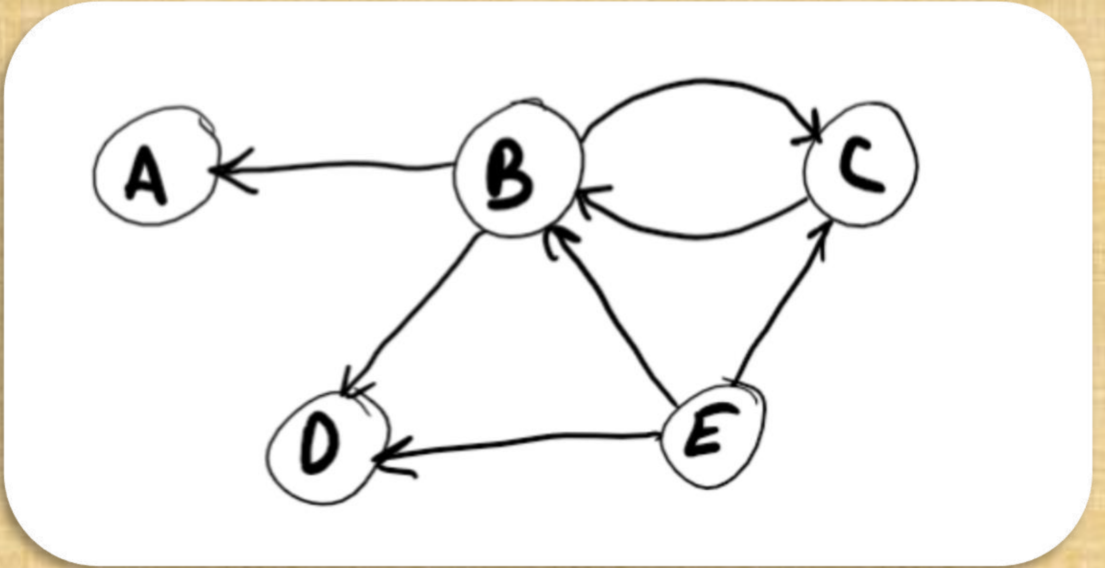
Directed Graphs

- Max. # of edges E_{max} is |V|(|V|-1)
- # of existing edges is |E|

$$d = \frac{|E|}{|V|(|V| - 1)}$$

Example

- $d = 7/(5 \times 4)$
  $= 0.35$

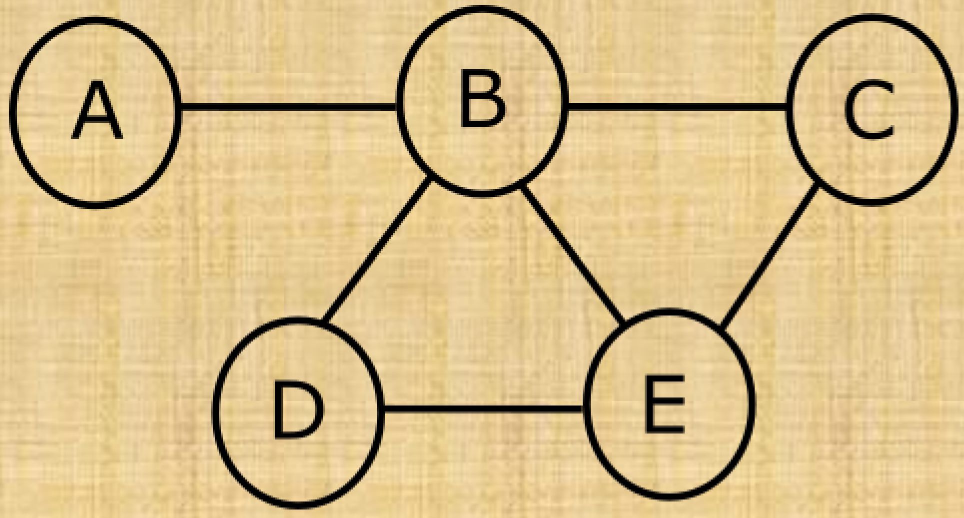


Connectivity\Cohesion

- Connectivity is the min. # of nodes that should be removed so the graph can be divided into one or more connected components.
- It gives an idea about the resilience of the graph.

Example

- con = 1, only option is removing B.



Centralization

- Centralization is a measure of Centrality distribution of nodes.
- It has a value between 0 and 1.
- 0: All nodes have the same centrality value.
- 1: A node dominates all other nodes.

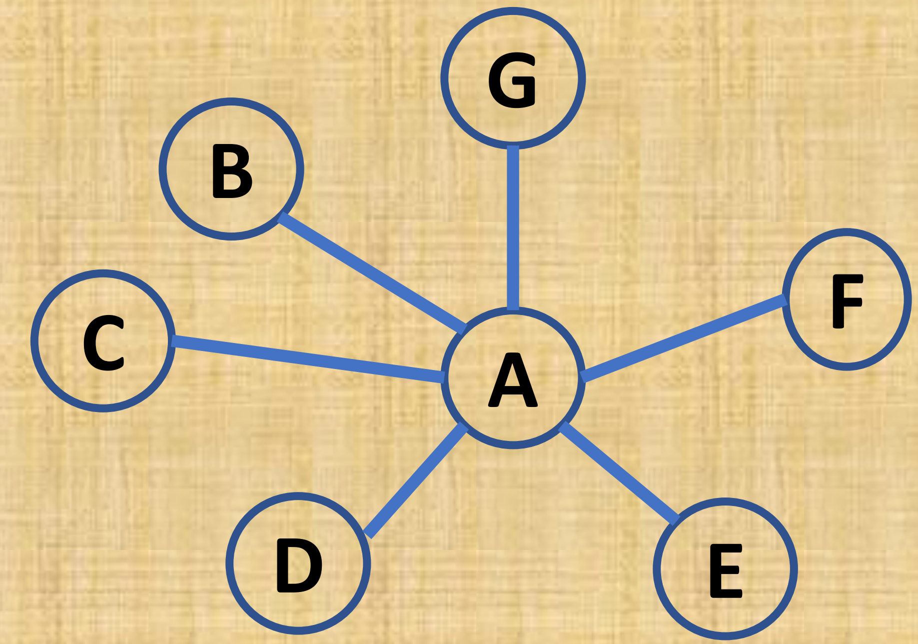
Centralization Formula

- $C(G) = \frac{\sum_{i=1}^{|V|} C(v^*) - C(v_i)}{max\left(\sum_{i=1}^{|V|} C'(v^*) - C'(v_i)\right)} \epsilon [0,1]$
- Where v^* is the node with the highest centrality score and C' is the centrality measure calculated for the

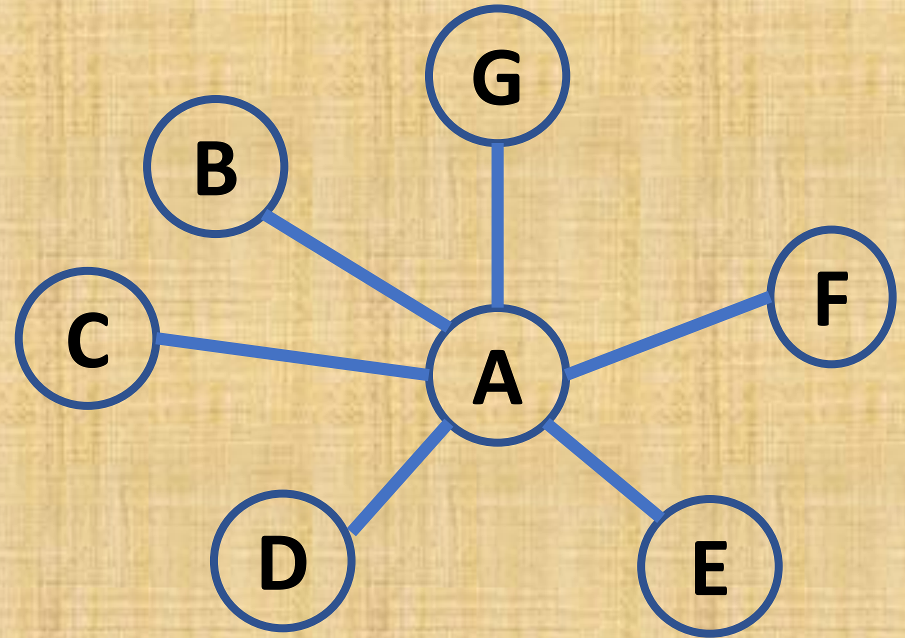
  worst case(star topology).
- Of course $C(v_i)$ can be calculated by any centrality measure.

The Denominator of Centraliation Formula

- The maximum sum of differences between the node with the highest centrality and other nodes appears in star topology.



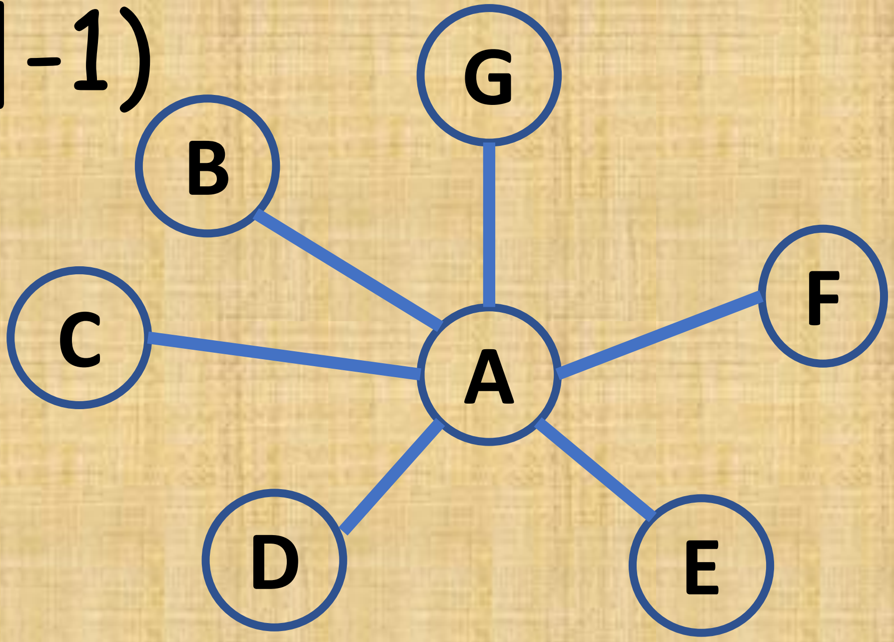
Degree Centrality



- $|V|=7$
- $deg(A)=|V|-1=6$
- $deg(B,C,D,E,F,G)=1$
- There are $|V|-1=6$ differences: $deg(A)-deg(B), deg(A)-deg(C)\ldots$
- All differences are $6-1=5=(|V|-2)$
- Max $\sum$ of differences: $(|V|-1)(|V|-2)$

Closeness Centrality

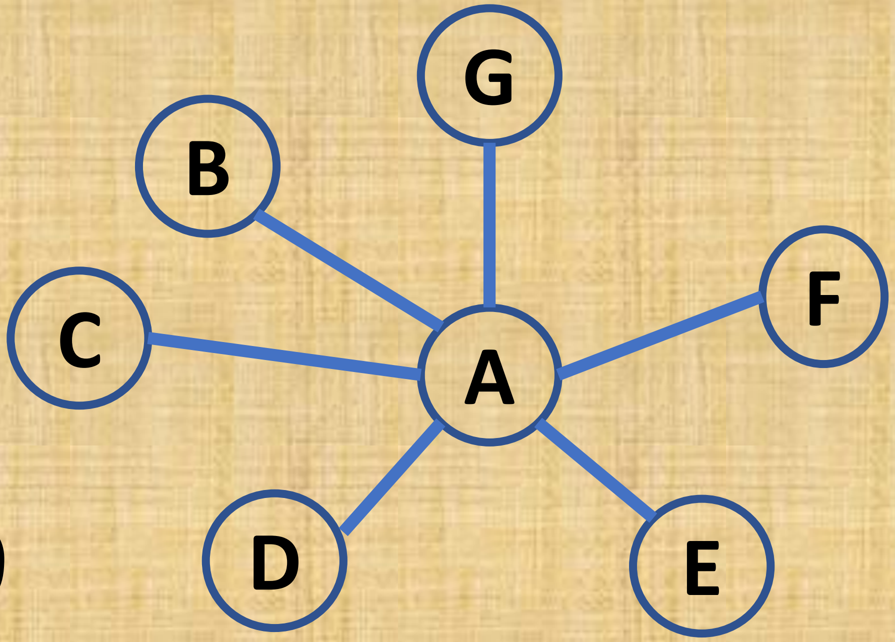
- $C_{cc}=1/(\sum$ of shortest paths$/|V|-1)$
- For A: $1/(6/6)=1$
- For B,C,D,E,F,G $=1/[(1+2\times5)/6]$
  $=6/11$
- There are $|V|-1=6$ differences.
- All differences are $1-6/11=5/11$
- Max $\sum$ of differences: $(|V|-1)(|V|-2)/(2|V|-3)$



Betweenness Centrality

- # of times node appears in the shortest paths
- P(6,2)/2=15 paths excluding A and A appears on all of them:
- $C_{bc}(A)$:15×(1/1) =15
- For B,C,D,E,F,G = 0
- There are |V|-1=6 differences.
- All differences are 15-0=15
- Max ∑ of differences: 15x6=90

$(|V|-1)^2(|V|-2)/2$



Global Efficiency

- Let $d(v_i, v_j)$ be the shortest path between nodes v_i and v_j.
- For each node, $1/d(v_i, v_j)$ is efficiency.
- If v_i and v_j are not connected then $d(v_i, v_j) = \infty$ and $\frac{1}{d(v_i,v_j)} = 0$.

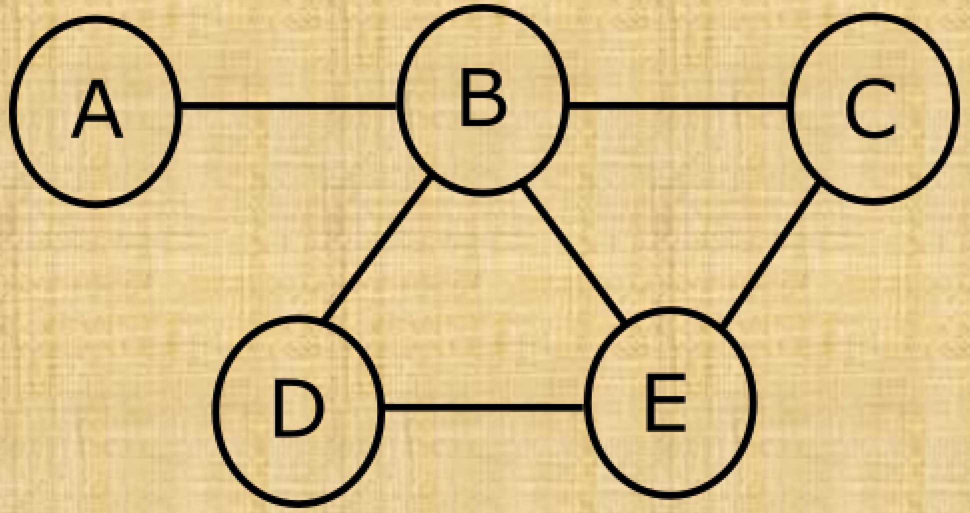
Efficiency Formula

$$Eff(G) = \frac{1}{|E|} \sum_{i<j} \frac{1}{d(v_i, v_j)} \in [0,1]$$

- If $Eff(G) = 1$ then G is a complete graph.
- Note $|E| = |V|(|V| - 1)/2$

Example

- There are 5 nodes, so $5 \times 4/2=10$ possible edge pairs.
- $d(A,B)=1; d(A,C)=2, d(A,D)=2$
  $d(A,E)=2$
- $d(B,C)=1; d(B,D)=1; d(B,E)=1$
- $d(C,D)=2; \ d(C,E)=1$
- $d(D,E)=1$



$\sum = \frac{1}{1} \times 6 + \frac{1}{2} \times 4$ so

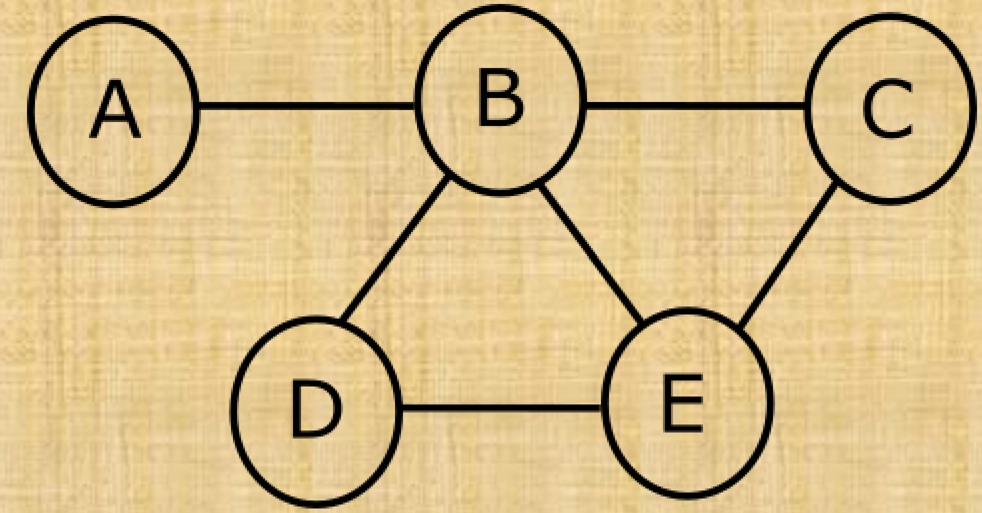
$Eff(G)=8/10=0.8$

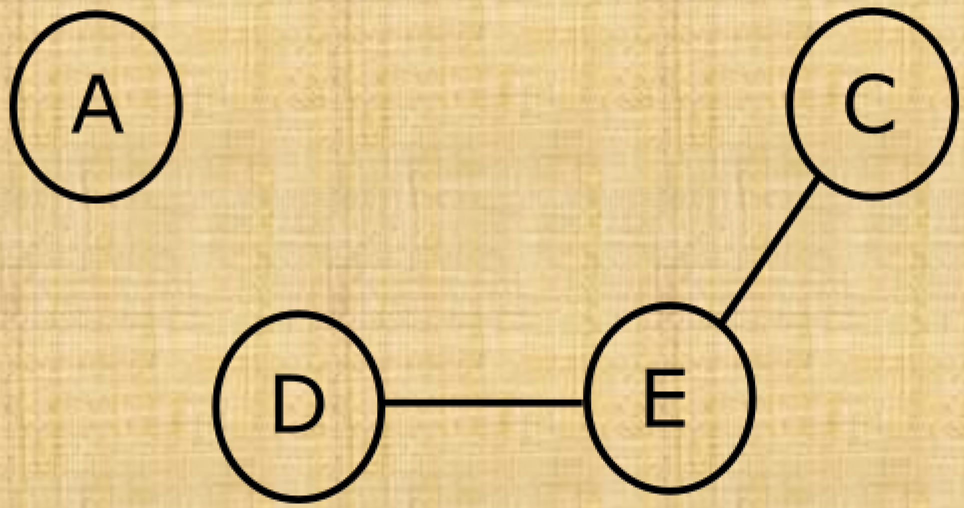
Local Efficiency

- Also called local fault tolerance, can the graph tolerate the node being removed?

Example

- The graph with B removed.
- $d(A,C)=d(A,D)=d(A,E)=\infty$
- $d(C,D)=2; d(C,E)=1$
- $d(D,E)=1$
- $\sum = (1/\infty)+(1/\infty)+(1/\infty)$
  $+(1/1)+(1/1)+(1/2)$
  $=5/2$
- $Eff(G) = (2/4\times3)\sum = 5/12$





END of Week 4