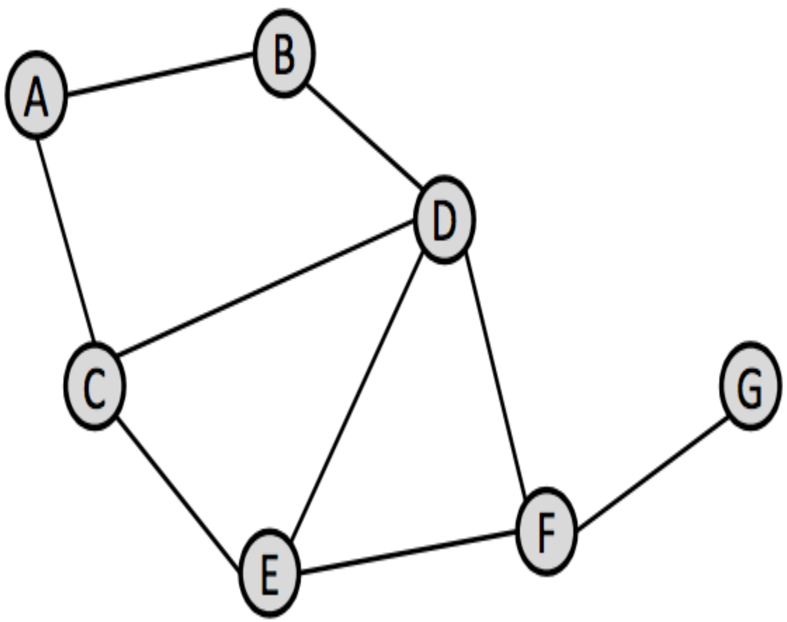
CMSC5733 Social Computing

Tutorial 4: Assignment 1 Solution

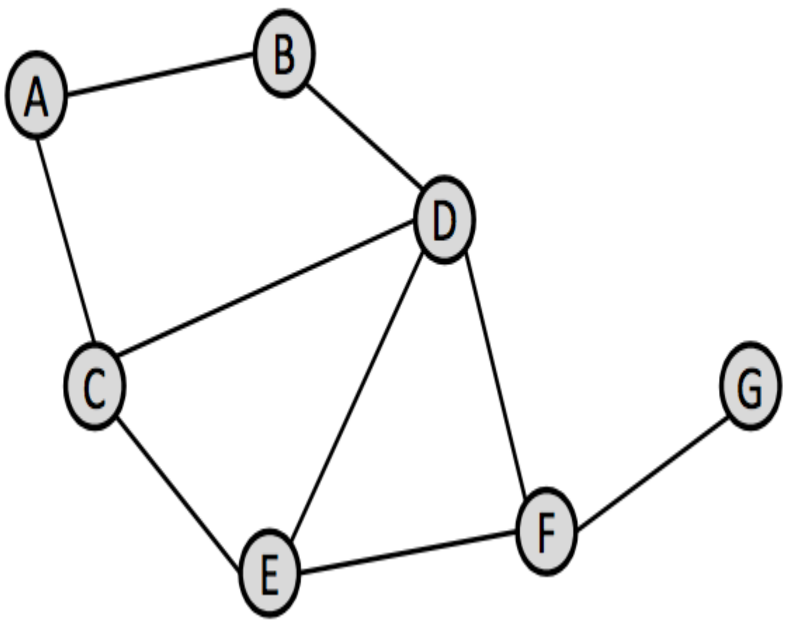
Yuanyuan, Man
The Chinese University of Hong Kong
sophiaqhsw@gmail.com

1. 1 Radius



- The longest path from a node v1 to all other nodes of a connected graph be defined as the radius of node v1
- Answer:
  - A: 4 (A-B-D-F-G)
  - D: 2 (D-B-A or D-F-G)
  - F: 3 (F-D-B-A or F-E-C-A)

1. 2 Diameter



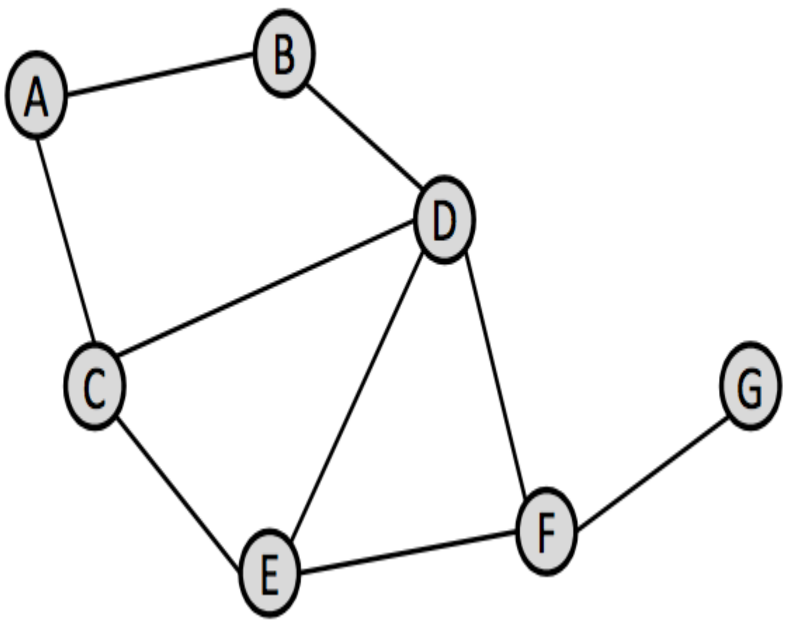
- The "longest shortest path"
- Answer: 4

  A – C – E – F – G

  Or A – B – D – F – G

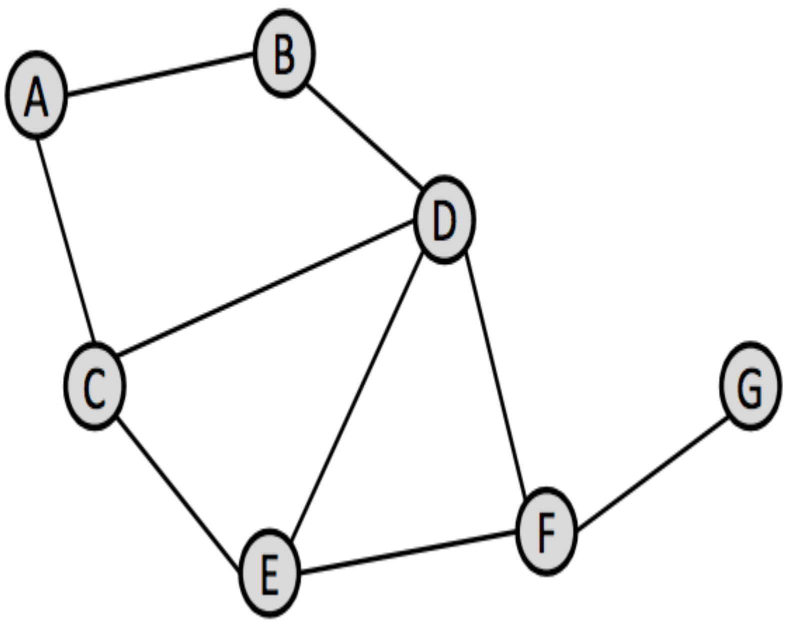
  Or A – C – D - F – G

1. 3 Center



- The center of the graph is the node with the smallest radius
- Answer:
  - A(4),B(3),C(3),D(2),E (2),F(3),G(4)
  - D or E
  - D or E is the best place to locate the supermarket

1.5 Adjacency Matrix



- The adjacency matrix ignores duplicate links between node
- Answer:

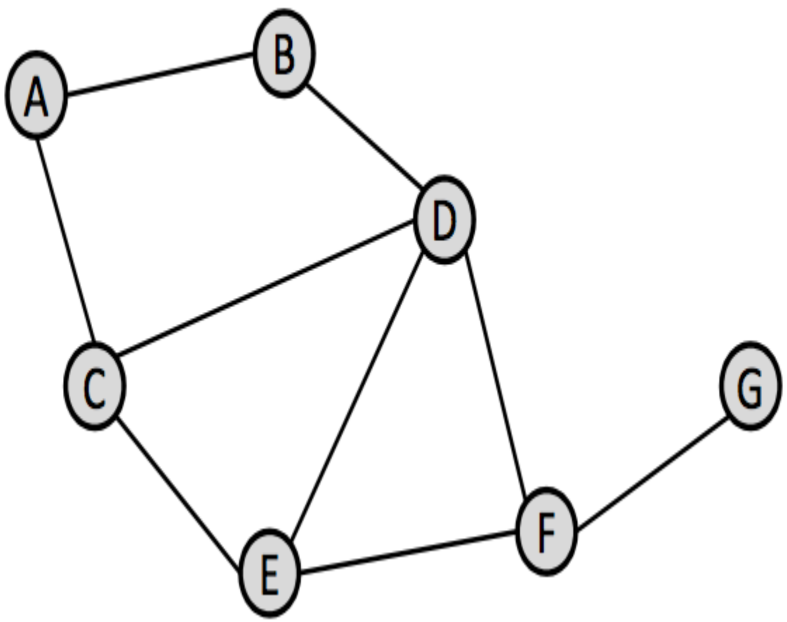
	A	B	C	D	E	F	G
A	0	1	1	0	0	0	0
B	1	0	0	1	0	0	0
C	1	0	0	1	1	0	0
D	0	1	1	0	1	1	0
E	0	0	1	1	0	1	0
F	0	0	0	1	1	0	1
G	0	0	0	0	0	1	0

1. 5 Laplacian Matrix

The *Laplacian matrix* of graph *G*, namely, *L*(*G*), is a combination of the connection matrix and (diagonal) degree matrix: *L* = *C* - *D*, where *D* is a diagonal matrix and *C* is the connection (adjacency) matrix

$$d_{i,j} = \begin{cases} \text{degree}(v_i) & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

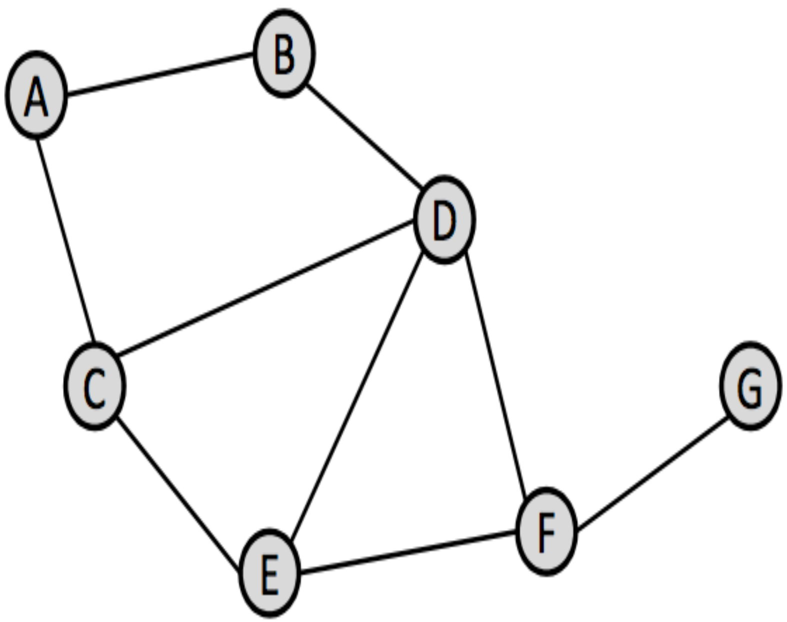
1. 5 Laplacian Matrix



- Answer:

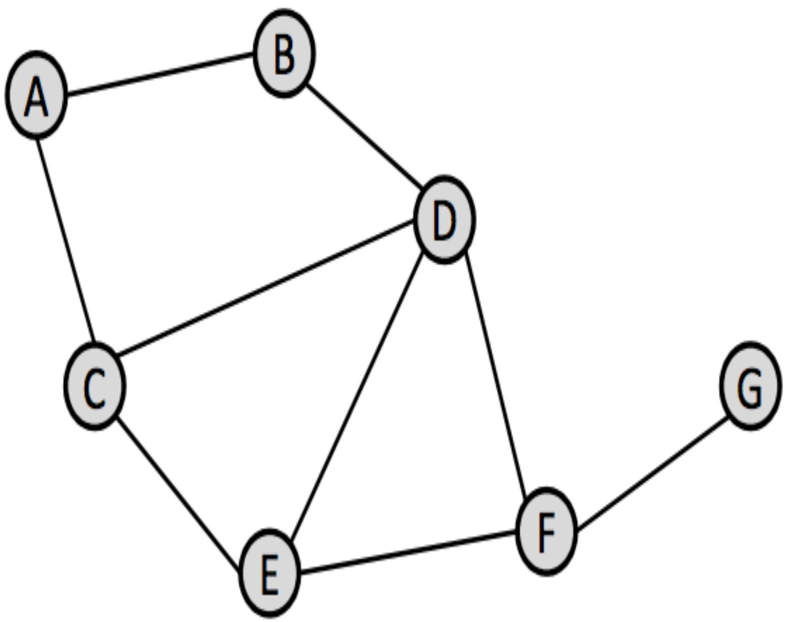
	A	B	C	D	E	F	G
A	-2	1	1	0	0	0	0
B	1	-2	0	1	0	0	0
C	1	0	-3	1	1	0	0
D	0	1	1	-4	1	1	0
E	0	0	1	1	-3	1	0
F	0	0	0	1	1	-3	1
G	0	0	0	0	0	1	-1

1.6



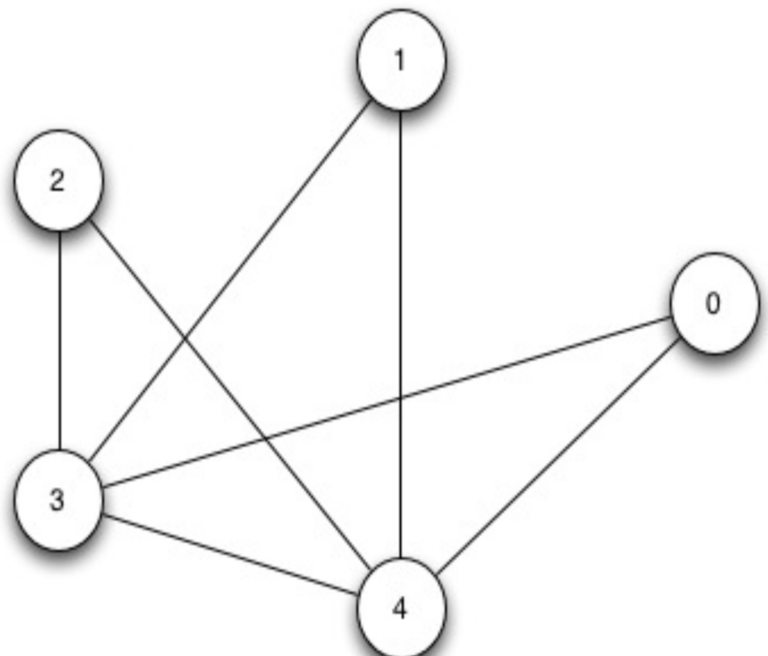
- What node(s) is (are) the farthest from the central node, and how far?
- Central node: D/E
- Answer:
  - D: A, G
  - E: A, B, G

1.7



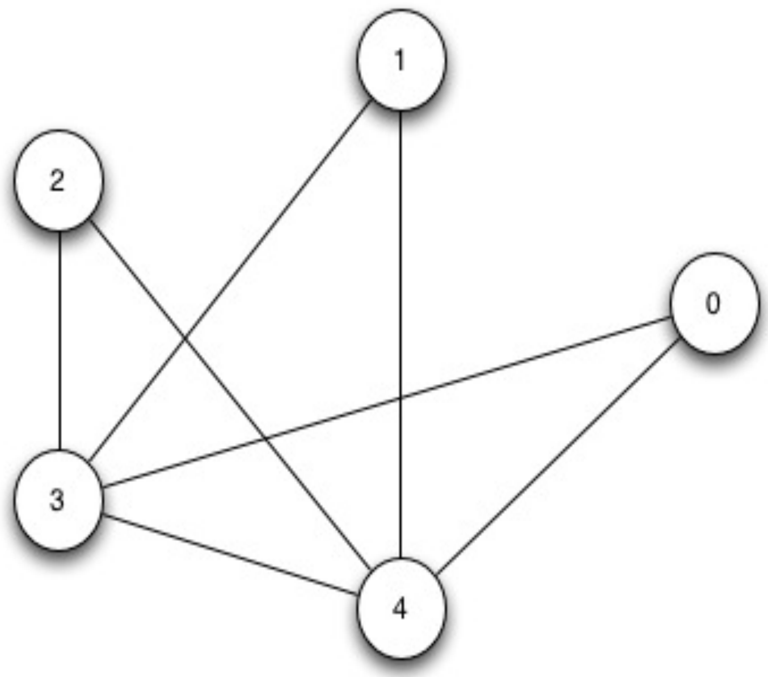
- Answer:
  - F

2. 1 Density



- Density = $\frac{2|E|}{|V|(|V|-1)}$
- Answer:
  - 2*7/ 5*4 = 0.7

2. 2 Degree Sequence



- Degree sequence: g=[d_1, d_2,…,d_n] define a degree sequence containing the degree values of all n nodes in G
- Answer:
  - g= [2,2,2,4,4]
    0 1 2 3 4

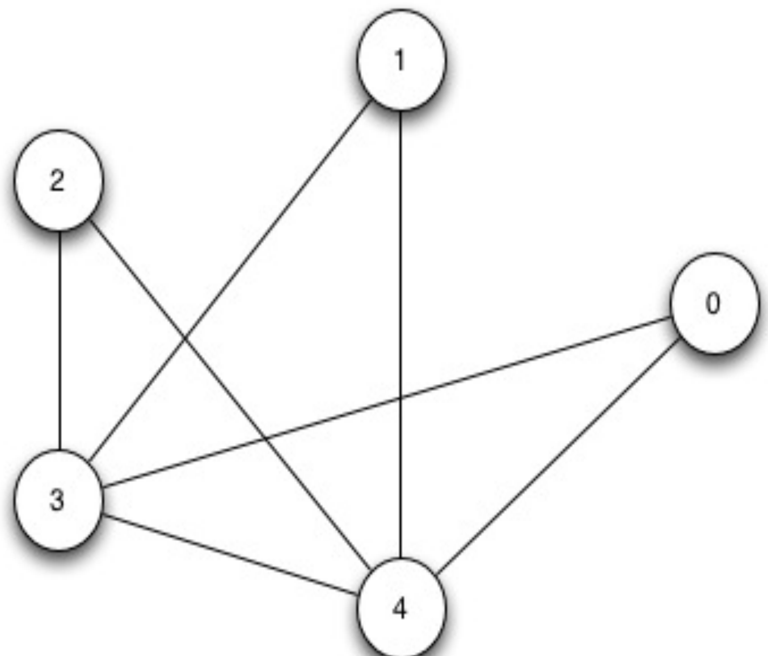
2. 1 Average Path Length

- The average path length of G is equal to the average over all shortest paths

$$l_G = \frac{1}{n \cdot (n-1)} \cdot \sum_{i,j} d(v_i, v_j)$$

- For unweighted graph G, let d(v1,v2) denote the shortest distance between v1 and v2; n is the number of node.

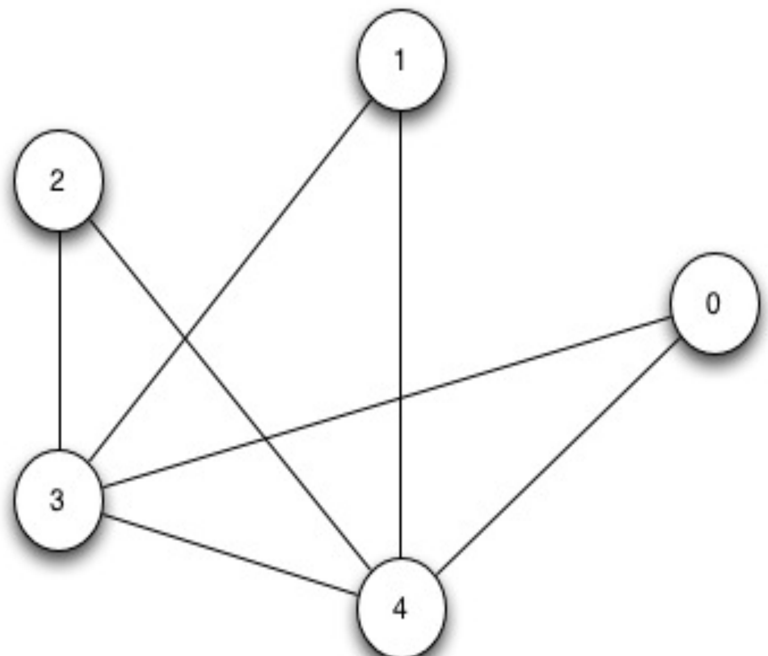
2. 1 Average Path Length



- Path matrix: path matrix P(G) stores the number of hops along the direct path between all node pairs in a graph

$$P(G) = \begin{bmatrix} 0 & 2 & 2 & 1 & 1 \\ 2 & 0 & 2 & 1 & 1 \\ 2 & 2 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

2. 1 Average Path Length



- Answer:
  - 26/5*4 = 1.3

3.1 Cluster Coefficient

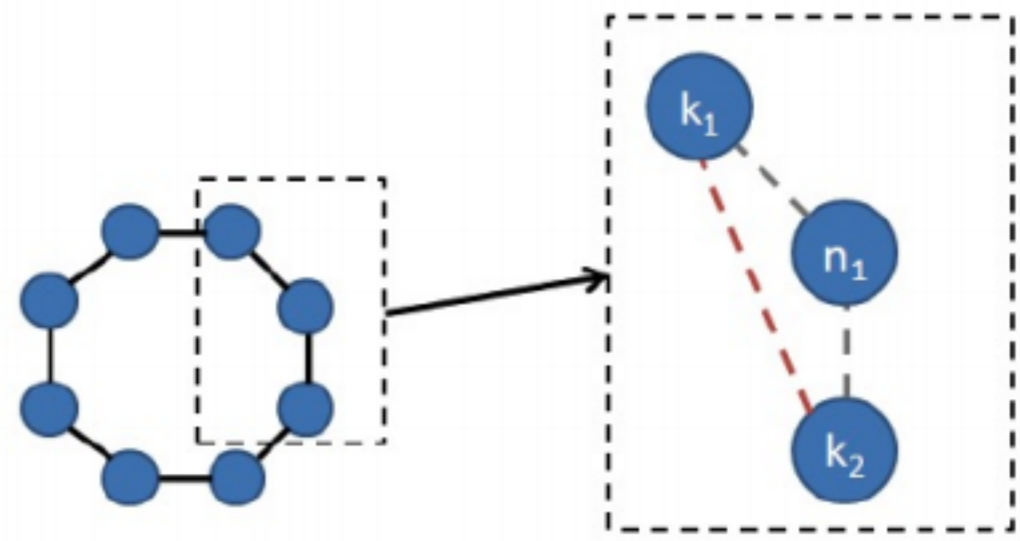
- For a node u, suppose that the neighbors share c links, then the cluster coefficient of node u, Cc(u), is

$$Cc(u) = \frac{2c}{\text{degree}(u)(\text{degree}(u) - 1)}$$

$$CC(G) = \sum_{i=1}^{n} \frac{Cc(v_i)}{n}$$

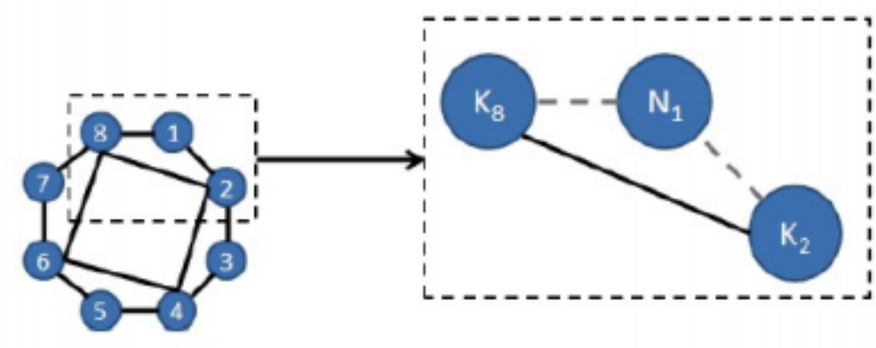
3. 1 Cluster Coefficient

- Answer: the average clustering coefficient of this network is 0. There are no triads in this network.

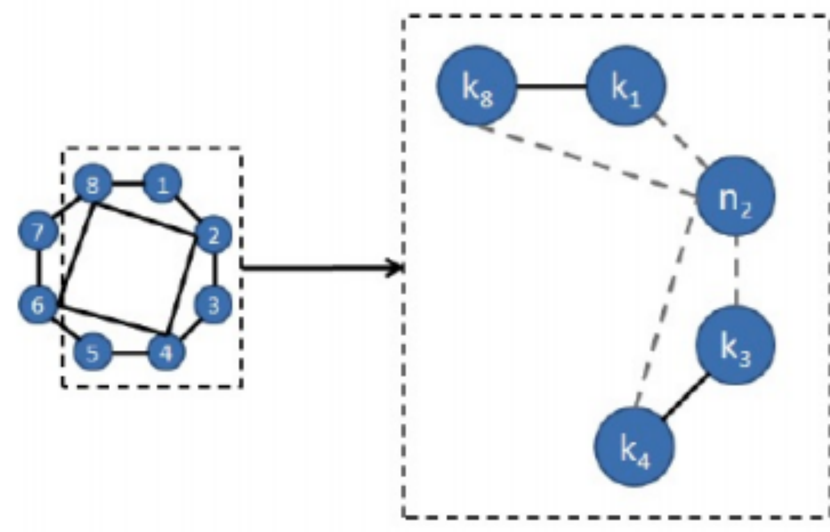


3. 2 Cluster Coefficient

- Answer: All the odd nodes will have a cluster coefficient of 1 because they only have two neighbors and those two neighbors know one another. So their cluster coefficient is 2*1/(2*1)=1



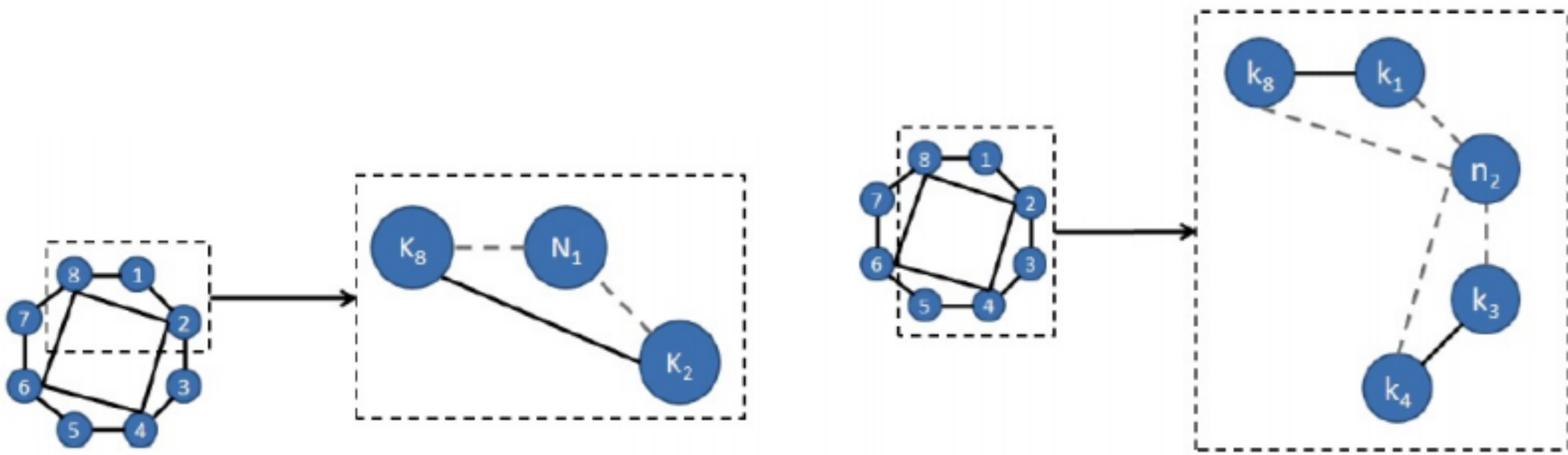
odd nodes



even nodes

3.2 Cluster Coefficient

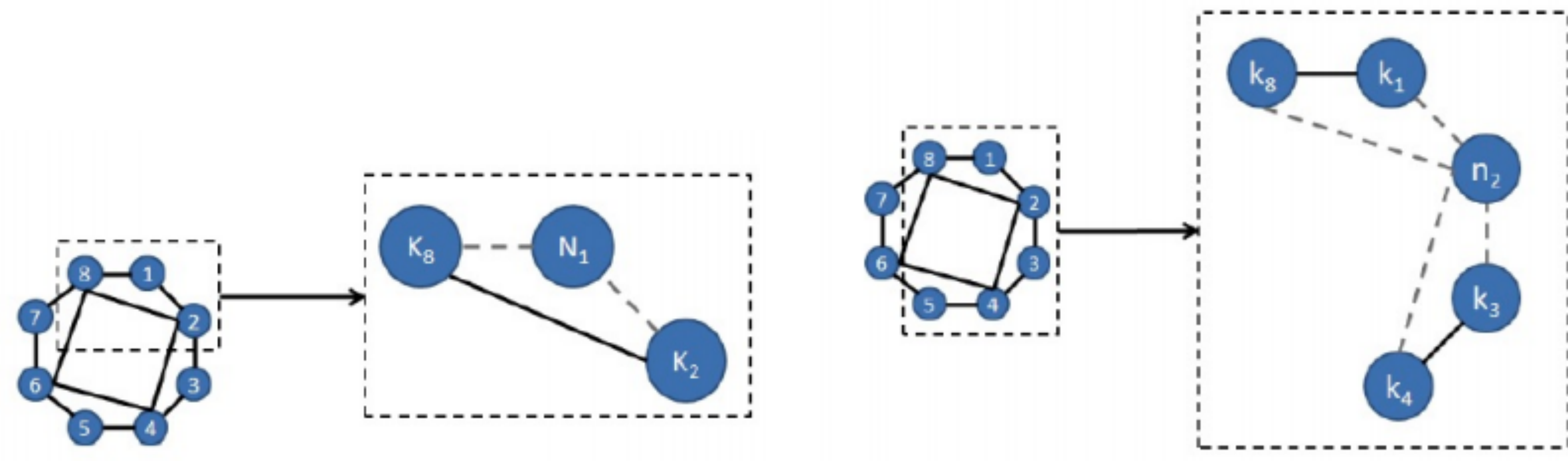
- Answer: The even nodes have four neighbors, and the two pairs of neighbors on either side know one another. So their cluster coefficient is 2*2/(4*3) = 1/3



odd nodes even nodes

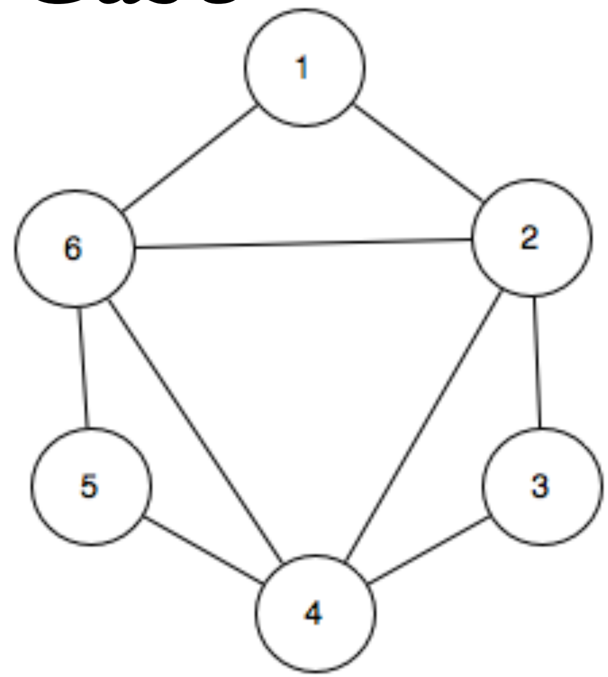
3. 2 Cluster Coefficient

- Answer: The cluster coefficient for the whole network is therefore n/2∗(1+1/3)/n= 2/3



odd nodes even nodes

3.3 Special Case



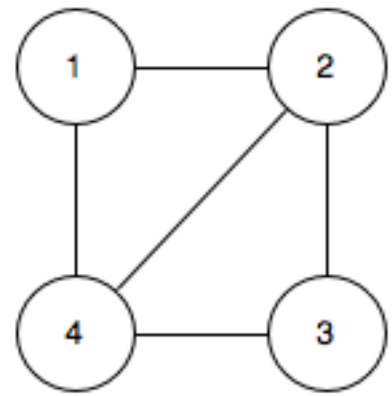
- In a special case n=6,

  cluster_coefficient(even node) = 3*2/4*3 = ½

  cluster_coefficient(odd node) = 1

  average_cc(graph) = n/2 * (1+1/2) /n = 3/4

3.3 Special Case



- In a special case n=4,

  cluster_coefficient(even node) = 2*2/3*2 = 2/3

  cluster_coefficient(odd node) = 1

  average_cc(graph) = n/2 * (1+2/3) /n = 5/6

4. 1 Closeness

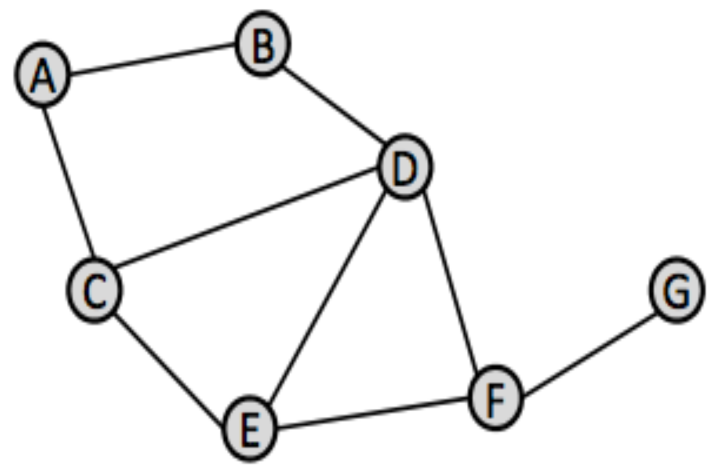
- Closeness centrality of a node u is the reciprocal the sum of the shortest path distance from u to all n-1 other nodes.

$$C(u) = \frac{1}{\sum_{v=1}^{n} d(v,u)}$$

where d(v, u) is the shortest path distance between v and u, and n is the number of nodes in the graph.

4. 1 Closeness

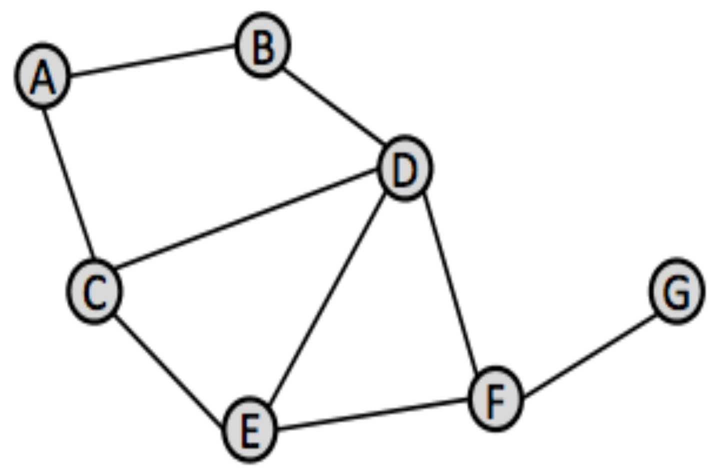
- Answer:
  - D -> (A, B, C, E, F, G)
    (2 1 1 1 1 2)
    D(D) = 8
    Closeness(D) = 1/8
  - F -> (A, B, C, D, E, G)
    (3 2 2 1 1 1)
    D(F) = 10
    Closeness(F) = 1/10



4. 1 Closeness

- Normalization
  - Closeness is normalized by the sum of minimum possible distance n-1

$$C(u) = \frac{n-1}{\sum_{v=1}^{n} d(v,u)},$$



  - Closeness(D) = 1/8
  - N_Closeness(D) = 6/8
  - Closeness(F) = 1/10
  - N_Closeness(F) = 6/10

4. 2 Betweenness

Betweenness Centrality of a node counts the number of times that a node lies along the shortest path between two others vertices in the graph. It is defined as

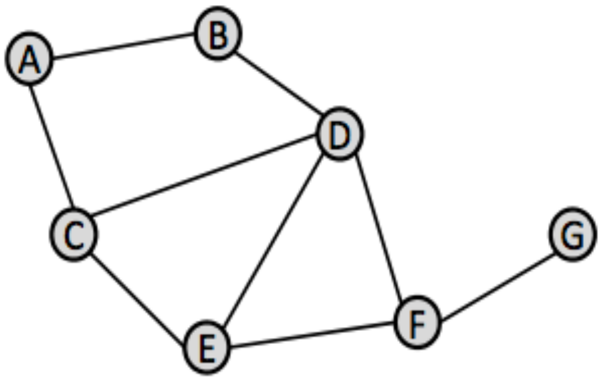
$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{g\sigma_{st}}. \quad (1)$$

where σ_{st} is the number of shortest paths from s to t and $\sigma_{st}(v)$ is the number of shortest paths from s to t that pass through a vertex v.

4. 2 Betweenness

- 1. For each pair of vertices (s, t), compute the shortest paths between them.
- 2. For each pair of vertices (s, t), determine the fraction of shortest paths that pass through the vertex in question (here, vertex v).
- 3. Sum this fraction over all pairs of vertices (s, t).

4. 2 Betweenness

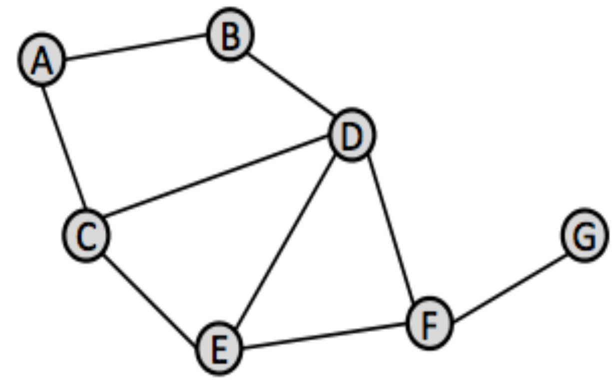


- Answer:
  - For node D:

Pairs	Shortest Paths	Total	Via D	Fraction
(A, F)	A-B-D-F, A-C-D-F, A-C-E-F	3	2	2/3
(A, G)	A-B-D-F-G, A-C-D-F-G, A-C-E-F-G	3	2	2/3
(B, C)	B-A-C, B-D-C	2	1	1/2
(B, E)	B-D-E	1	1	1/1
(B, F)	B-D-F,	1	1	1/1
(B, G)	B-D-F-G	1	1	1/1
(C, F)	C-D-F, C-E-F	2	1	1/2
(C, G)	C-D-F-G, C-E-F-G	2	1	1/2
CLOSENESS(D) = 2/3 + 2/3 + 1/2 + 1 + 1+ 1 + 1/2 + 1/2= 5.833				

4. 2 Betweenness

- Answer:
  - For node F:



Pairs	Shortest Paths	Total	Via F	Fraction
(A, G)	A-B-D-F-G, A-C-D-F-G, A-C-E-F-G	3	3	3/3
(B, G)	B-D-F-G	1	1	1/1
(C, G)	C-D-F-G, C-E-F-G	2	2	2/2
(D, G)	D-F-G	1	1	1/1
(E, G)	E-F-G	1	1	1/1
CLOSENESS(F) = 1+ 1+ 1+1+1 = 5				

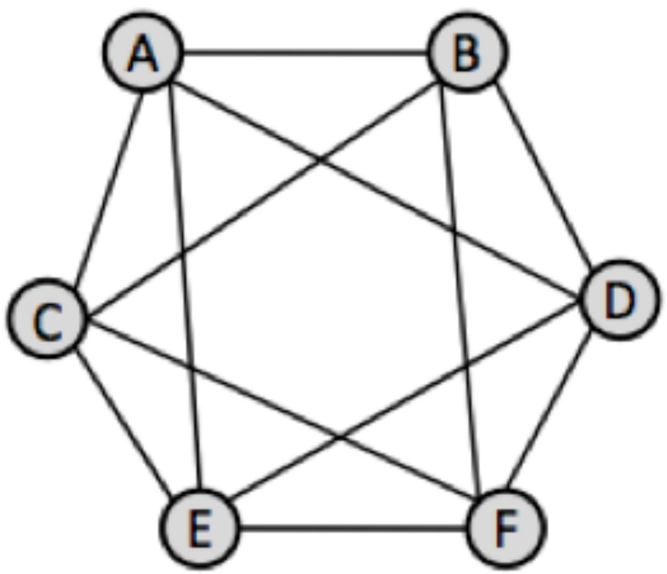
4. 2 Betweenness

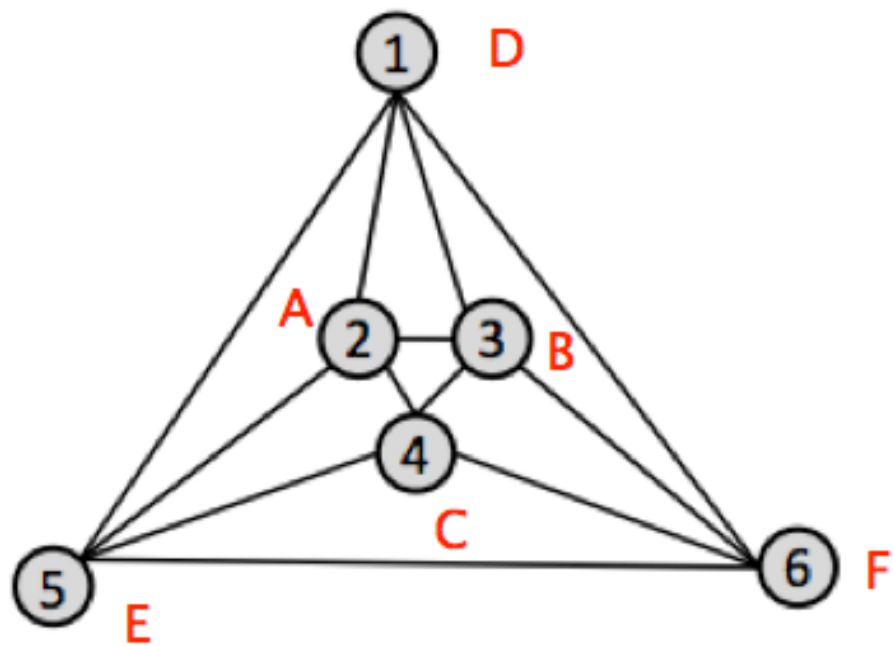
- Normalization
  - Betweenness is normalized by 2/((n-1)(n-2)) for undirected graphs, and 1/((n-1)(n-2)) for directed graphs
  - CLOSENESS(D) = 5.833 * 2/(6*5) = 0.389
  - CLOSENESS(F) = 5 * 2/(6*5) = 0.333

5. 1 Isomorphism

- Two graphs G and H are said to be isomorphic, denoted by G ~ H, if there is a one-to-one correspondence, called an isomorphism, between the vertices of the graph such that two vertices are adjacent in G if and only if their corresponding vertices are adjacent in H.
- Answer: D~1, E~5, F~6, A~2, B~3, C~4

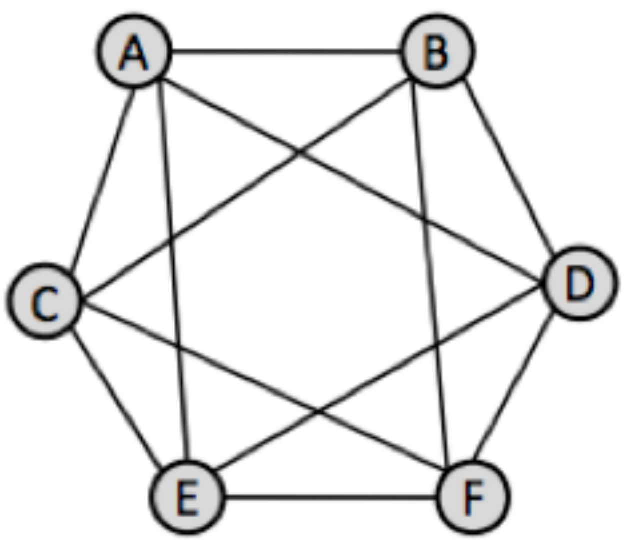
5.1 Isomorphism





5. 2 Eulerian Path

- In graph theory, an Eulerian path is a trail in a graph which visits every edge exactly once.
- Answer: A-E-D-A-C-F-B-C-E-F-D-B-A



5. 3 Cluster Coefficient

- Graph clustering coefficient
  - Global clustering coefficient

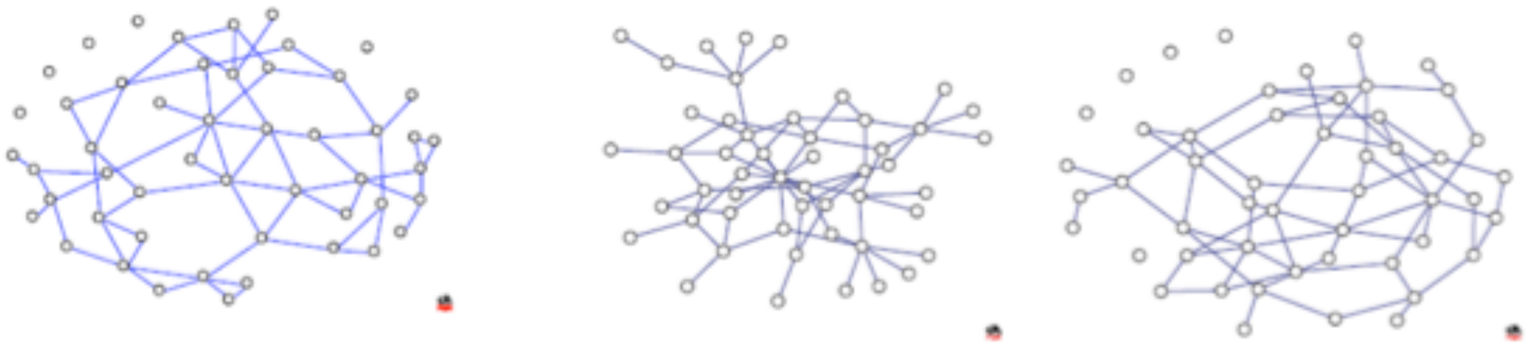
$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}}$$

  - Average clustering coefficient

5. 3 Cluster Coefficient

- The correct answer is (a). This is the only network of the three that has a significant occurrence of closed triads. Triads contribute to the clustering coefficient whether it is calculated as a the average proportion of connected neighbor pairs over all the vertices, or as (3 times the number of triangles)/(number of connected triples) in the entire graph.

5. 3 Cluster Coefficient



a b c

Q&A