Fall Semester 2020

KAIST AI607

Graph Mining and Social Network Analysis

Homework 3: Take-Home Midterm Exam

Release: October 9, 2020

Due: October 23, 2020, 11:59 PM

Name: Fanchen Bu

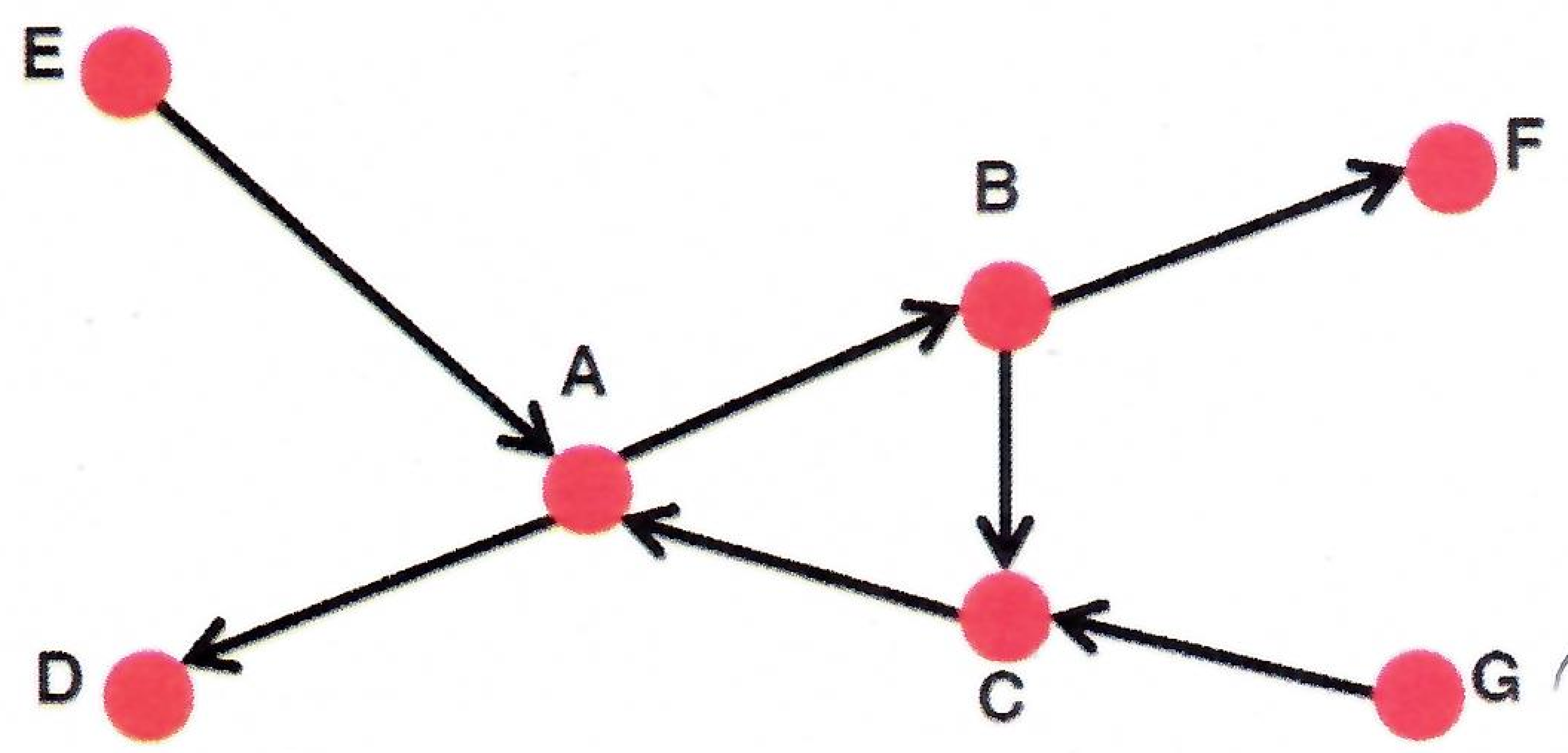
Student ID: 20194185

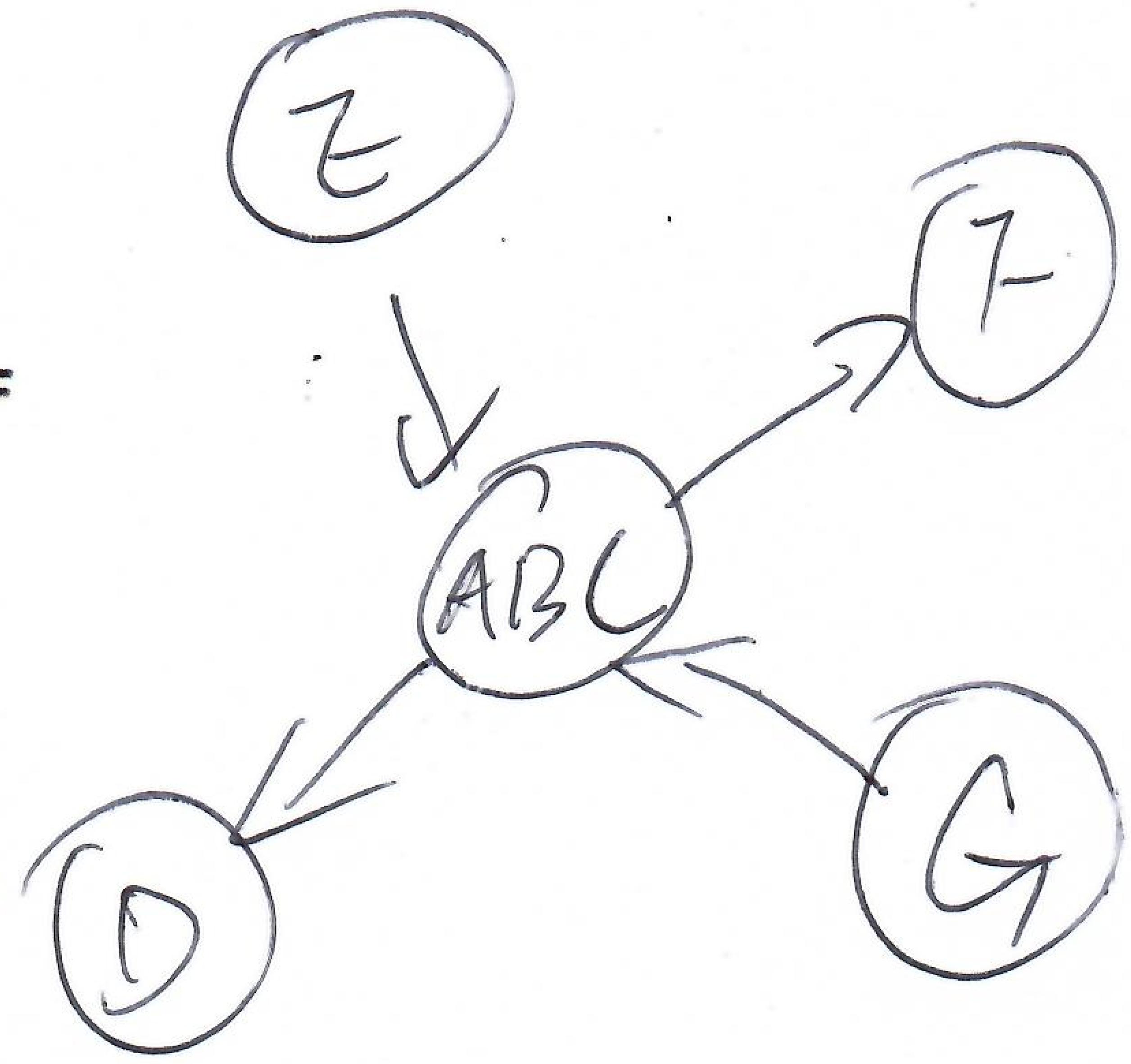
This exam is open book and notes. Read the questions carefully and focus your answers on what has been asked. You are allowed to ask the instructor/TAs for help only in understanding the questions, in case you find them not completely clear. Be concise and precise in your answers and state clearly any assumption you may have made. Good Luck

Question 1	/10
Question 2	/ 10
Question 3	/ 10
Question 4	/ 10
Question 5	/ 10
Question 6	/ 15
Question 7	/ 10
Question 8	/ 15
Question 9	/ 10
Total	/ 100

1. **(10 points) Basic Graph Theory**

Consider the following directed graph G:





*(a) **(5 points)** How many weakly connected components (WCCs) does G have?

Answer: There are ___1___ WCCs in G.

as the whole graph is "weakly connected", as defined in the lecture.

Hint: Recall that WCCs should be maximal. Adding any node to a WCC should make it no longer a WCC.

(b) **(5 points)** How many strongly connected components (SCCs) does G have?

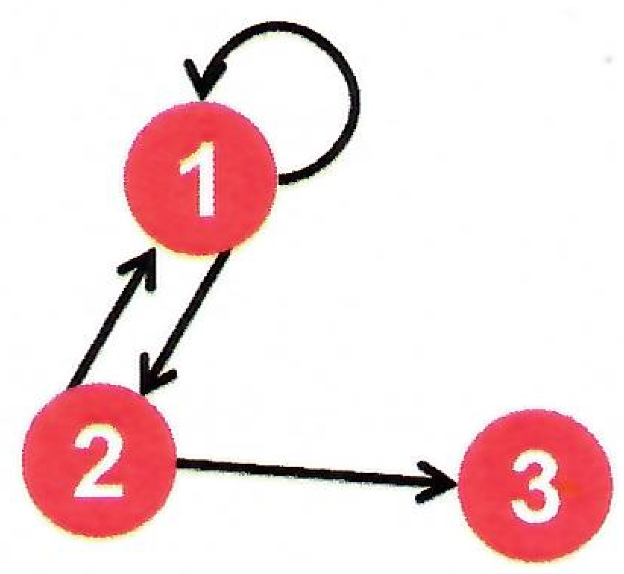
Answer: There are ___5___ SCCs in G.

Hint: Recall that SCCs should be maximal. Adding any node to a SCC should make it no longer a SCC.

(*) The definition of WCC I learned is the maximal subgraph induced by a subset of vertex set where $\forall i \neq j$, $\exists$ path from $i \to j$ or from j to i, which gives the answer 4.

2. **(10 points) PageRank**

Consider the following unweighted directed graph G:



$$r_j = \sum_{i \to j} \beta \frac{r_i}{d_i} + (1-\beta)\frac{1}{N}$$

(a) (5 points) What is the adjacency matrix A of G? Fill the following matrix:

A =

	1	2	3
1	1	1	0
2	1	0	1
3	0	0	0

(b) (5 points) Computing $\vec{x} \leftarrow \mathbf{B} \times \vec{x}$ repeatedly makes $\vec{x}$ converge to the PageRank score vector (with damping factor 0.85) of G.

β

What is the matrix **B**? Fill the following matrix:

$$M = \begin{bmatrix} 1/2 & 1/2 & D \\ 1/2 & 0 & \\ 0 & 1/2 & E \end{bmatrix}$$

$$\downarrow$$

$$\begin{bmatrix} 1/2 & 1/2 & 1/3 \\ 1/2 & 0 & 1/3 \\ 0 & 1/2 & 1/3 \end{bmatrix} \times 0.85$$

$$+ \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \times 0.15$$

B =

	1	2	3
1	0.475	0.475	1/3
2	0.475	0.05	1/3
3	0.05	0.475	1/3

Hint:

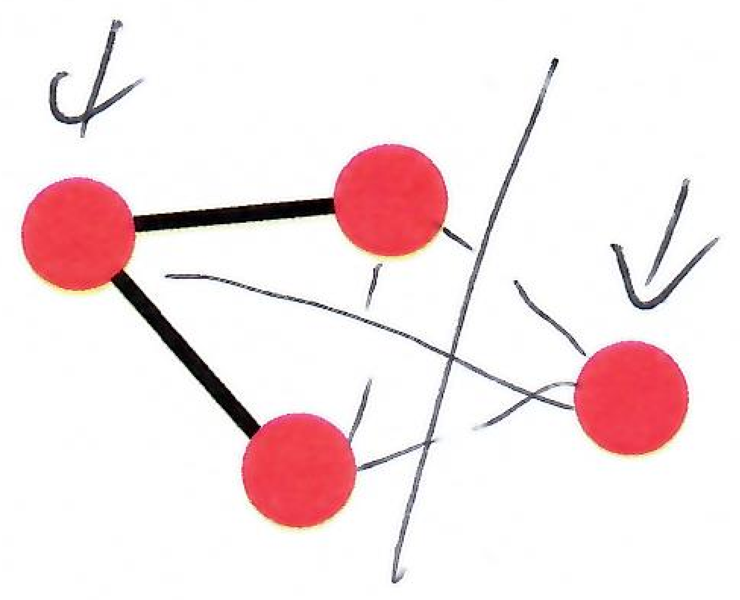
- Consider the version of PageRank where the dead-end problem and the spider-trap problem are solved by teleports.
- The probability that the random surfer teleports is (1 - damping factor) $= 0.15$
- When the random surfer teleports, it jumps to a uniformly random node.

3. **(10 points) Random Graphs**

Assume we generate an undirected graph using the random graph model G(N, p) with N = 4 and p = 0.3

(a) **(5 points)** What is the probability that the generated graph has the following structure?

$4 \times 3 \times (0.3)^2 \times (0.7)^4$



Answer: the probability is $\frac{64827}{250000} = 0.259308$

(Feel free to write an equation as your answer)

Hint:
- N is the number of nodes
- p is the probability that each pair of two distinct nodes are connected
- There can be multiple graphs with the same structure

(b) **(5 points)** What is the expected value of the average degree of the generated graph?

Answer: the expected average degree is 0.9

(Please provide an exact value)

$E[\bar{d}] = 2E[|E|]/|V| = 2\binom{n}{2}p/n = (n-1)p =$

4. **(10 points) Community Guided Attachment**

Assume we generate an undirected graph with 9 nodes using the community guided attachment model with b = 3, c = 0.5

(a) **(5 points)** What is the probability that the generated graph is complete (i.e., all nodes in the generated graph are connected to each other)?

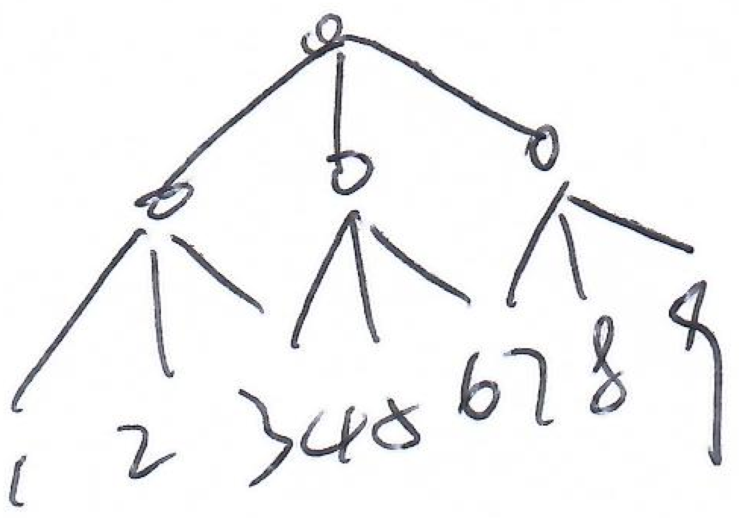
Answer: the probability is 2^{-63}

(Feel free to write an equation as your answer)

Hint:
- b is the community tree branching factor
- two nodes with tree distance h is connected with probability c^h

(Do not ask the definitions of 'community tree branching factor' and 'tree distance'.).

$f(h) = c^{-h} = 2^{-h}$



Total $\binom{9}{2} = 36$ possible edges

Level 1 : $\binom{3}{2} \times 3 = 9$ w/ $Pr = 1/2$

Level 2 : 27 w/ $Pr = 1/4$

$(1/2)^9 \times (1/4)^{27} =$

(b) **(5 points)** What is the expected value of the average degree of the generated graph?

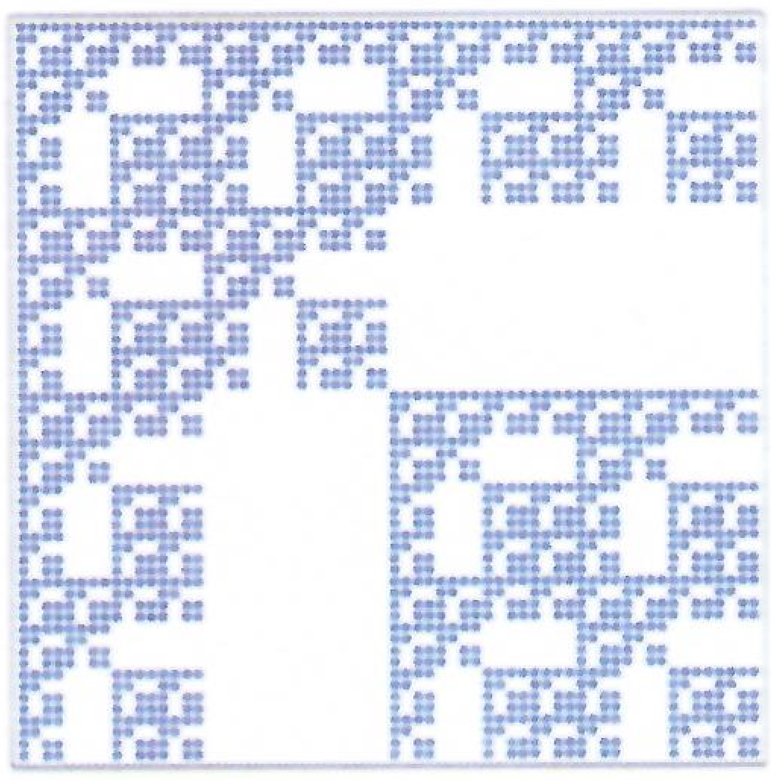
Answer: the expected average degree is $\frac{5}{2}$

(Feel free to write an equation as your answer)

$E[\bar{d}] = 2E[|E|]/|V| = \frac{2}{9} \times (9 \times 1/2 + 27/4)$

5. **(10 points) Kronecker Graph**

Assume that an undirected graph G_3 with the following adjacency matrix is generated by the (deterministic) Kronecker graph model.



<Sparsity pattern of the adjacency matrix of G_3>

(a) **(5 points)** What is the adjacency matrix of the seed graph? Fill the following matrix:

G_1 =

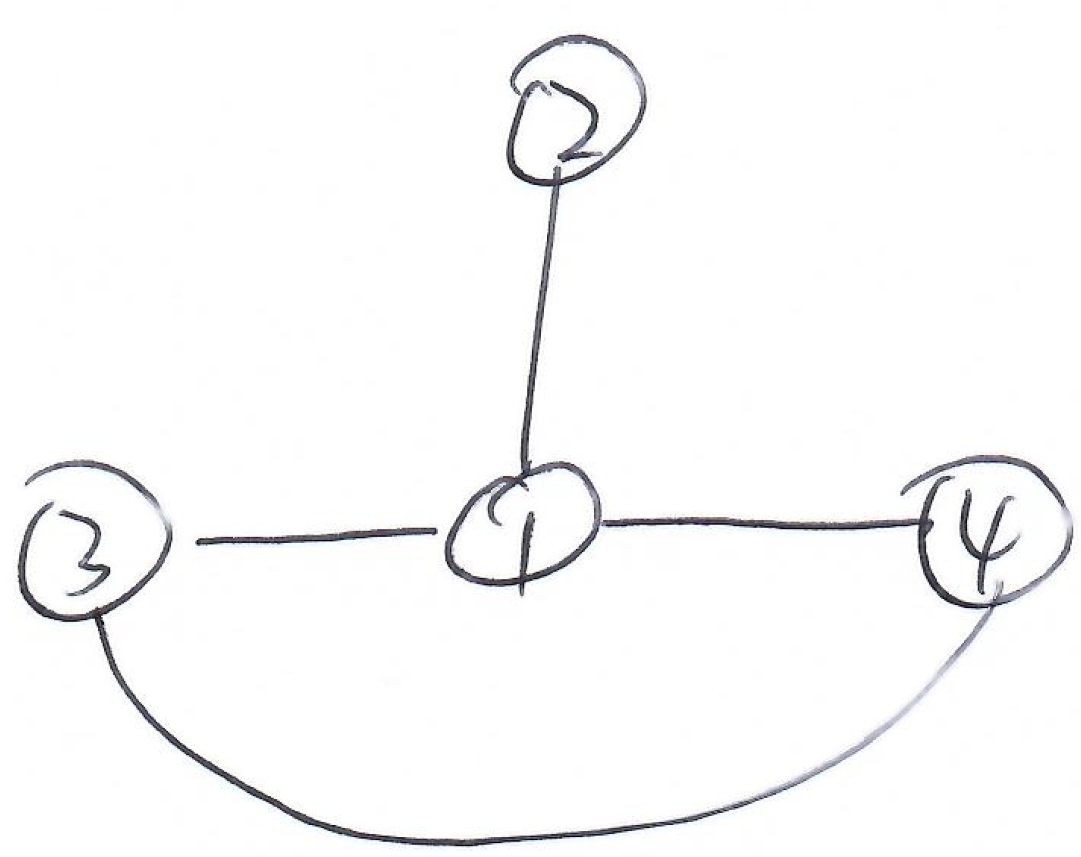
	1	2	3	4
1	1	1	1	1
2	1	1	0	0
3	1	0	1	1
4	1	0	1	1

(b) **(5 points)** what is the diameter of G_3?

Answer: the diameter is 2

$D(G_3) = D(G_1)$

D	1	2	3	4
1	ϕ	1	1	1
2	1	ϕ	2	2
3	1	2	ϕ	1
4	1	2	1	ϕ



6. **(15 points) Coreness**

Consider the following undirected graph G:

(a) **(5 points)** What is the maximum coreness in G?

Answer: the maximum coreness is 1

(b) **(5 points)** To increase the maximum coreness, how many edges should we add to G?

Answer: at least 1 edge(s) should be added

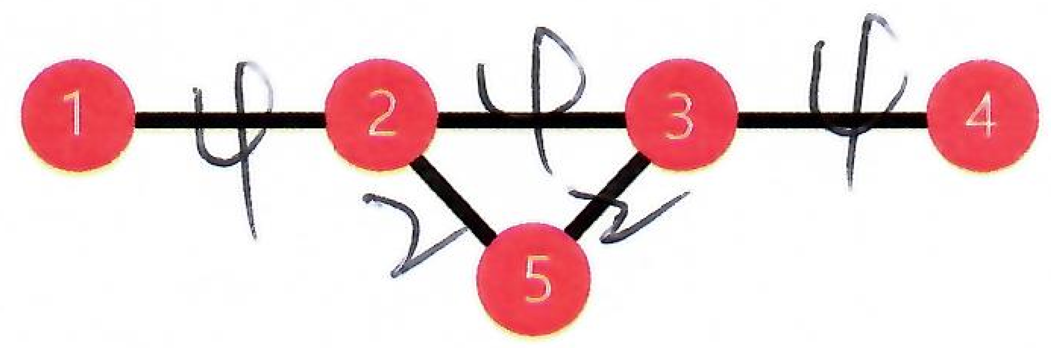
2-core $\Leftarrow K_3$

(c) **(5 points)** To decrease the maximum coreness, how many edges should we remove from G?

Answer: at least 3 edge(s) should be removed any edge $\rightarrow$ 1-core

7. **(10 points) Edge Betweenness Centrality**

Consider the following undirected graph G:



(a) **(5 points)** Choose <u>one</u> edge with the highest betweenness centrality?

Answer: edge between 2 and 3

(b) **(5 points)** Choose <u>one</u> edge with the lowest betweenness centrality?

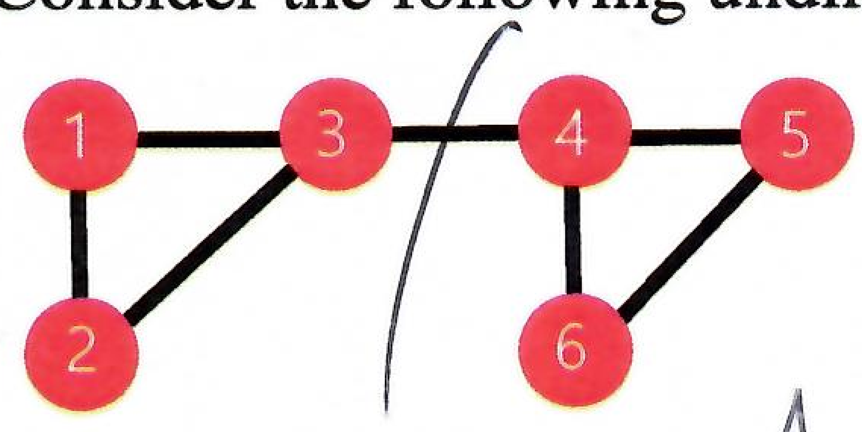
Answer: edge between 2 and 5

$$C_B(e)=\sum_{s,t\in V}\frac{\sigma(s,t|e)}{\sigma(s,t)}$$

	1	2	3	4	5
1	φ	12	123	1234	125
2		φ	23	234	25
3			φ	34	35
4				φ	435
5					φ

8. **(15 points) Laplacian Matrix and Eigenvectors**

Consider the following undirected graph G and its adjacency matrix A:



$A =$

0	1	1	0	0	0
1	0	1	0	0	0
1	1	0	1	0	0
0	0	1	0	1	1
0	0	0	1	0	1
0	0	0	1	1	0

$D = \mathrm{diag}(2, 2, 3, 3, 2, 2)$

Let $\vec{x}$ be an eigenvector corresponding to the second smallest eigenvalue of the Laplacian matrix of G.

Hint: you do not have to calculate $\vec{x}$ to answer the following questions.

(a) **(5 points)** What are the **two** entries of $\vec{x}$ closest to 0?

① $\vec{x}_1$ ② $\vec{x}_2$ ③ $\vec{x}_3$ ④ $\vec{x}_4$ ⑤ $\vec{x}_5$ ⑥ $\vec{x}_6$

Answer: ___3___ **and** ___4___

(b) **(5 points)** Compare the two numbers below using >, = or <.

Answer: $\vec{x}_1 \times \vec{x}_3$ ___>___ 0

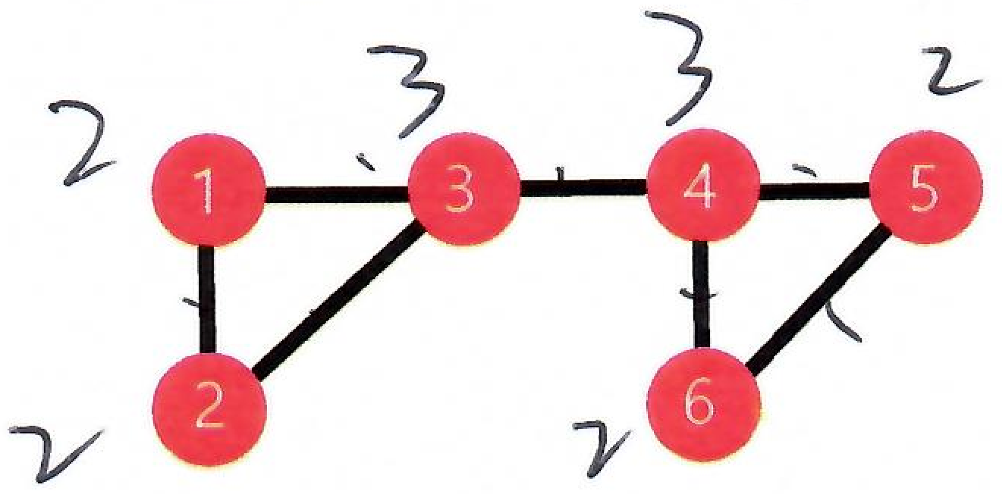
(c) **(5 points)** Compare the two numbers below using >, = or <.

Answer: $\vec{x}_3 \times \vec{x}_4$ ___<___ 0

$$Q = \frac{1}{2m} \sum_{\substack{i,j \\ \text{Same Com}}} \left(A_{ij} - \frac{k_i k_j}{2m} \right)$$

9. **(10 points) Modularity**

Consider the following undirected graph G.



$$= \sum_{c \in Com} \left[\frac{L_c}{m} - \left(\frac{K_c}{2m} \right)^2 \right]$$

L_c: # of intra-Com. links in C

K_c: Σ of degrees of nodes in C

$m = 7$

Hint: use the definition of modularity in the lecture note.

(a) **(5 points)** For three communities {1,2}, {3,4}, and {5,6}, compute the modularity.

Answer: the modularity is $4/49$.

$$2 \times \left(\frac{1}{7} - \left(\frac{4}{14} \right)^2 \right) + \left(\frac{1}{7} - \left(\frac{6}{14} \right)^2 \right)$$

(b) **(5 points)** For two communities {1,2,3} and {4,5,6}, compute the modularity.

Answer: the modularity is $5/14$.

$$2 \times \left(\frac{3}{7} - \left(\frac{7}{14} \right)^2 \right)$$