End Semester Examination

Winter 2025, Network Science

26th April 2025; Time: 50 Minutes [**60+2 Points Maximum**]

Roll Number: _____ Name: _____

1. Choose the correct alternative. *Negative points for wrong answers.* [$20 \times \pm 1 = +20/-20$]

(1) The 'Coefficient of Assortativity' for a Erdos-Renye graph is expected to be:

(a) 0 (b) +1

(c) -1 (d) -2

(2) For a 'divisive strategy' used for network clustering, which of the following metric/property can be used as the basis for partitioning the network?

(a) clustering coefficient (b) edge weight

(c) node degree (d) edge betweenness

(3) Which of the following are the characteristic features of the 'Barabasi-Albert Algorithm' for generating scale-free networks'?

(a) Preferential Attachment & Sparseness (b) Growth & Preferential Attachment

(c) Sparseness & Growth (d) Sparseness & Rich getting poorer

(4) Most biological networks are known to be:

(a) Scale-free and Disassortative (b) Scale-free and Assortative

(c) Sparse and Assortative (d) Disassortative and Dense

(5) For which of the following conditions on their degree exponent, the Scale-Free networks have been shown to be non-existent?

(a) $\gamma < 0$ (b) $\gamma < 1$

(c) $\gamma < 2$ (d) $\gamma < 3$

(6) Which of the following metric computes the 'nature of degree correlations' in a network?

(a) Coefficient of assortativity (b) Coefficient of correlativity

(c) Degree exponent (γ) (d) Correlativity index

(7) Which of the following was the factor observed to be the best explanation for the 'clustering of obese people' from the study of 'spread of obesity on social networks'?

(a) Clustering (b) Induction

(c) Homophily (d) Confounding

(8) Which of the following was found to be the most critical factor in the 'spread of obesity'?

(a) Geographic Distance (b) Ego perceived friendship

(c) Mutual Friendship (d) Alter perceived friendship

(9) Which of motifs are commonly found in transcriptional gene regulation networks as well as neuronal networks?

(a) bifan and three chain (b) bifan and feedbackward loop

(c) feedforward and feedbackward loops (d) feedforward loop and bifan

(10) Which of the following have been **traditionally** used for assessing the power of a language?
(a) Number of Wikipedia Pages (b) Demography and Economic Power
(c) Number of Twitter Accounts (d) Number of books published

(11) Which of the following data is **not** the correct 'space' discussed in the research article 'Drug-target and disease networks: polypharmacology in the post-genomic era'?
(a) Target Space (b) Chemical Space
(c) Disease Space (d) Genomic Space

(12) In a network, if a 4-node pattern of connectivity (subgraph) is found with numbers significantly **lesser** compared to that in its random graph counterparts, then such a pattern would be labelled—
(a) a motif (b) a community
(c) an anti-motif (d) a clique

(13) In the 'Flavor Network' in which nodes are ingredients, the edges represent—
(a) any shared characteristics between ingredients.
(b) shared intensity of flavor.
(c) number of shared flavor atoms.
(d) number of shared flavor molecules.

(14) In the 'network of recipes' where nodes are recipes, the edges represent—
(a) number of shared ingredients. (b) number of shared flavor molecules.
(c) number of shared flavor atoms. (d) shared color intensity.

(15) Which of the following situations demands the use of graph partitioning?
(a) Graph Coloring (b) Integrated Circuitry (IC) Design
(c) Shorted Path Analysis (d) Finding Giant Component

(16) The last step of Hierarchical Clustering involves building a—
(a) Dendrogram (b) Anagram
(c) Dendrotree (d) Clustogram

(17) In a random graph, the probability that a node with degree k links to a node with degree k' is—
(a) $p_{kk'} = \frac{kk'}{L}$ (b) $p_{kk'} = \frac{kk'}{2L}$
(c) $p_{kk'} = \frac{L}{kk'}$ (d) $p_{kk'} = \frac{2L}{kk'}$

(18) The mean of a binomial distribution is expressed as—
(a) $< x >= Np$ (b) $< x >= Np^2$
(c) $< x >= N^2p$ (d) $< x >= N^2p^2$

(19) In a simple, undirected, random graph with N nodes and p as the probability that any two nodes are connected, average degree would be—
(a) $< k >= pN$ (b) $< p >= pC_2^N$
(c) $< k >= p(N-1)$ (d) $< p >= pN^2$

(20) The adjacency matrix data structure for storing a graph takes more space compared to the edge list due to its—

(a) redundant data storage.
(b) unstructured data storage.
(c) inability to represent edges.
(d) efficient data organization.

2. In a scale-free network with degree exponent $\gamma = 3$ (i.e. $P(k) = k^{-3}$), prove that nodes with degree (interactions) $k = z$ are approximately eight times more populated than nodes with $k = 2z$. [2]

$\gamma = 3$

$K_{max} = K_{min} N^{\frac{1}{\gamma-1}} \Rightarrow K_{min} N^{\frac{1}{2}}$

$\Rightarrow K_{min} \sqrt{2z} \Rightarrow 8z^3$

Hence eight times more

3. Briefly explain the Kerninghan-Lin Algorithm for partitioning the graph into two parts. [2]

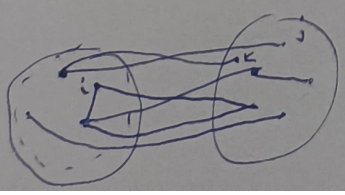
⇒ Kerninghan - Lin Algorithm is used to segregate the clusters in graph. The algorithm works as:

→ Seperate the nodes as per your intituition in the first step

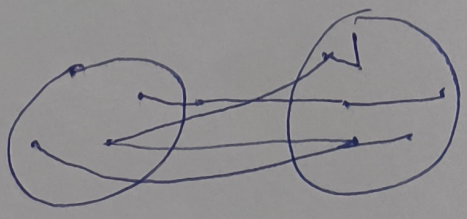
→ Choose node i from one cluster & j from another cluster & check if swapping these nodes will minimize the edge cut b/w these cluster (what if cut size is increased)

⇒ If the edge cut is minimize after swapping then swap them

⇒ Repeat the step until the No. of edges to between these two clusters is ~~minimized~~



→



4. For an undirected and unweighted graph, show that $< k > = 2L/N$. [1]

⇒ for an undirected graph, the average degree is defined as:

$$\langle K \rangle = \frac{\text{Sum of the degree of nodes}}{\text{total No. of nodes}}$$

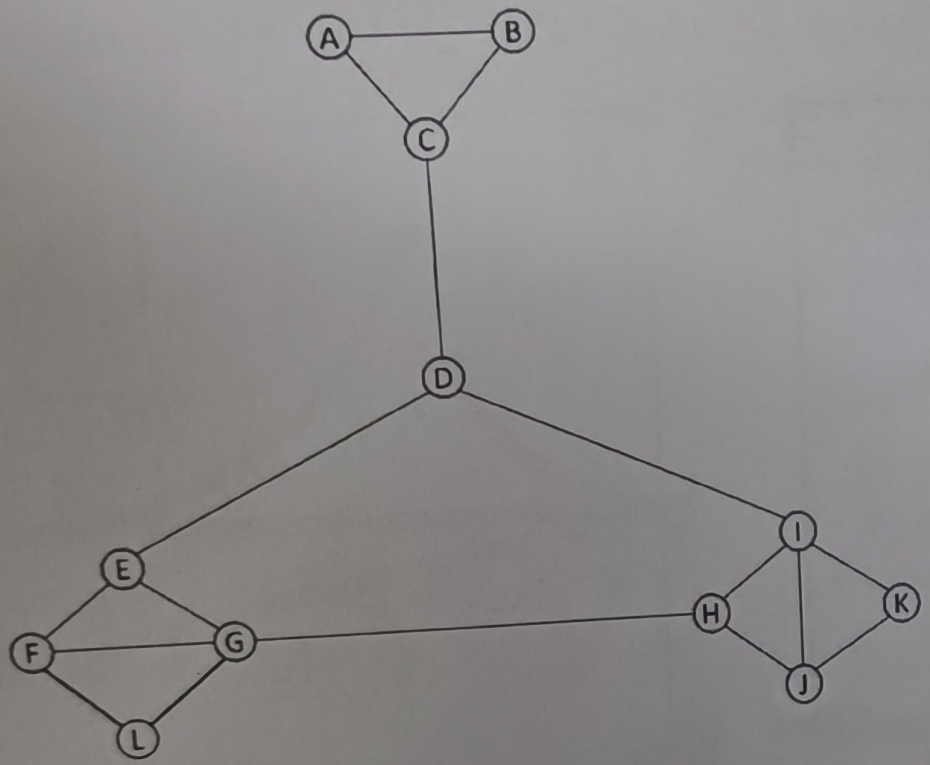
By handshaking lemma, we know the sum of degree of Nodes is twice the edge length.

Hence we got,

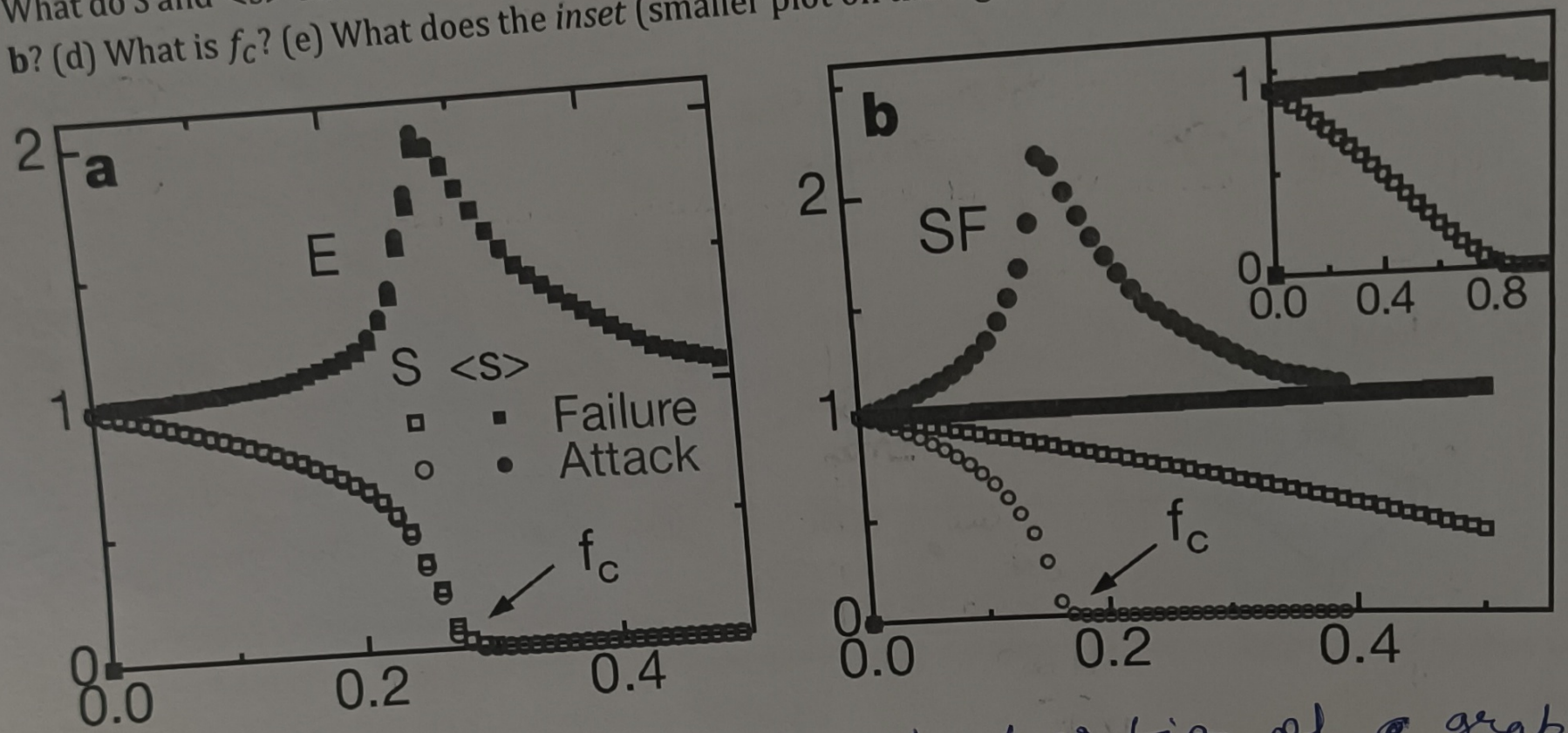
$$\langle K \rangle = 2L/N \quad (1)$$

5. Briefly explain the 'Configuration Model for generating networks of desired degree sequence'. (a) How are the number of stubs and the total number of edges related? (b) Prove the probability that an arbitrary node i connecting to node j is given by, $p_{ij} = \frac{k_i k_j}{2L-1}$. [3]

6. Define the 'Topological Overlap Index' used for measuring the overlap between different modules. Using the same, compute the index for every edge in the following network. [1+3=4]

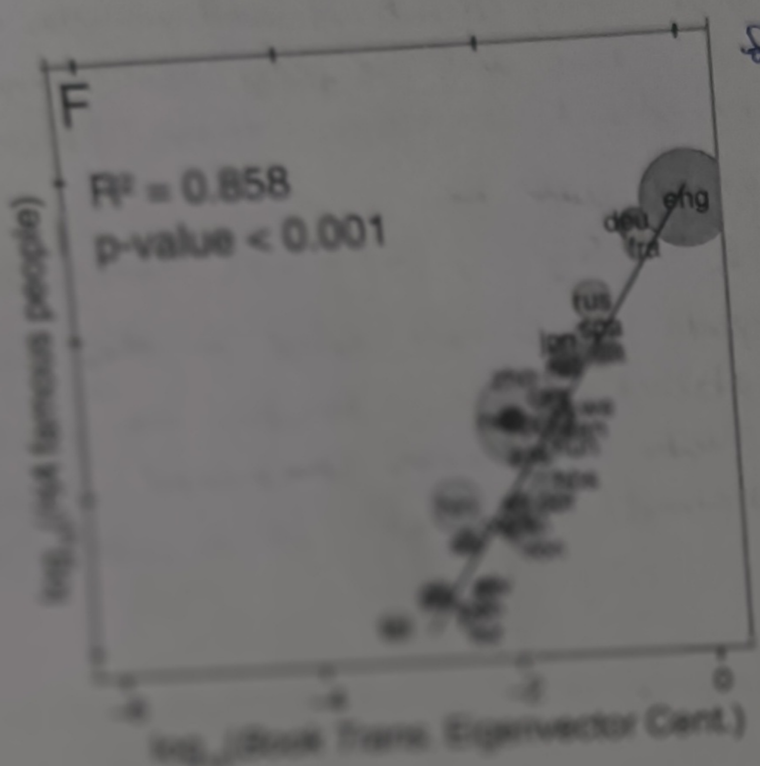


7. Briefly explain results, depicted in the following figures, obtained by Albert *et al.* in their studies o 'error and attack tolerance of complex networks'. (a) What is represented on the x- and y-axis? (b) What do S and <s> stand for? (c) What are the conclusions from each data curve in figure **a** and figure **b**? (d) What is f_c? (e) What does the *inset* (smaller plot on the right-top) in figure **b** imply? [1×5=5]



graph

8. The following figure presents a result from the PNAS article 'Links that speak: The global language network and its association with global fame.' It depicts 'the position of a language in the GLN (x-axis) and the global impact of its speakers (y-axis)' for the Global Language Network created using the 'book translations data.' The size of the bubble represents the number of speakers for each language and the color intensity represents Gross Domestic Product (GDP) per capita for the language/country. What are the inferences that you would draw regarding: (a) correlation between the x- and y-axis parameters, (b) correlation between the GDP of the country and the centrality of its official language in GLN? (c) Which aspect best explains the global impact of the speakers? Why? [½ ×2 + 1 = 2]



a) The graph represent the quantity of famous people speaking a particular language, wrt the Book translation that happens in that language.

The graph suggest that the more the No. of famous people of a particular language is more is ~~their~~

from other language to

9. Rooted in the tripartite network of Recipes, Ingredients, and their constituent Flavor Compound state and describe all the bipartite and their monopartite network projections that can be created. For each of these networks, clearly state the notion of in-degree, out-degree and/or degree. [4]

Recipes(S_1) Ingredient(S_2) Flavor(S_3)

R_1 I_1 F_1

R_2 I_2 F_2

The above tripartite graph will have edge b/w Recipes & Ingredient, if that Ingredient is used in the making of the recipe, similarly there will be edge b/w Ingredient & Flavor if the Ingredient contains those flavor compounds

Those the bipartite graphs will be:

$S_1 - S_2$, $S_2 - S_3$, $S_1 - S_3$ (Theoretically possible)

The above bipartite graph will have edge in the same senses as i have mentioned above. For Recipe to Flavor, there will be edge b/w Recipe & Flavor compound, if the recipe has that Flavor compound (Practically it is not implemented)

Monopartite, weighted graph for flavors compounds where edge weights will represent its prevalence The graph will have edge if

10. Define the following terms and illustrate by drawing an example each: (a) Cycle, (b) Eulerian Path, (c) Self-avoiding Path, and (d) Shortest Path. [1+1+1+1 = 4]

11. The analysis of a real-world network ($n = 10000, e = 50000$) led to following observations about frequency of 5 subgraphs in the network, and their frequency and standard deviation (SD) from 1000 instances of random controls. Define 'motif' and state the formula for computational of Z_{score}, compute Z_{score} for each subgraph. Conclude which of the subgraphs could be called a 'motif'. [3]

Subgraph	N_{real}	$N_{rand} \pm SD$	Z_{score}
Subgraph-1	9800	9000 ± 400	
Subgraph-2	355	300 ± 50	
Subgraph-3	12300	200 ± 50	
Subgraph-4	100	10 ± 5	
Subgraph-5	655	700 ± 15	

$= 800/400 = 2$
$= 55/50 = 1.1$
$= 12100/50 = 242$
$= 90/5 = 18$
$= -45/15 = -3$

Motif - is a small subgraph of a large Graph which is found a relatively higher No. of times than the others or in its random counterpart.

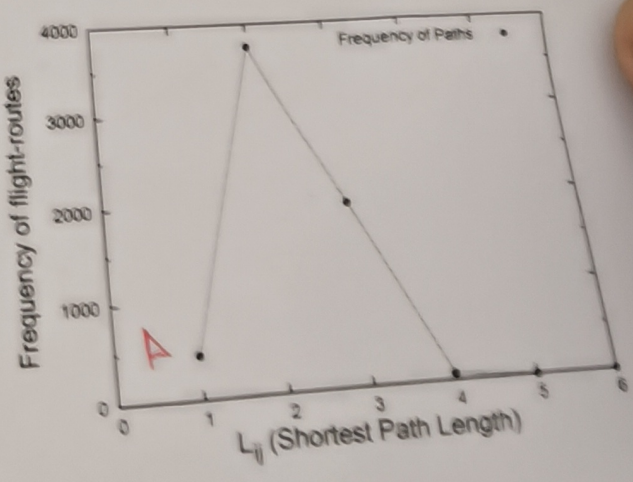
$$Z_{score} = \frac{N_{real} - N_{rand}}{\sigma}$$

3

from the above calculation we found out that the most suitable candidate for subgraphs are:
Subgraph 3, Subgraph 4

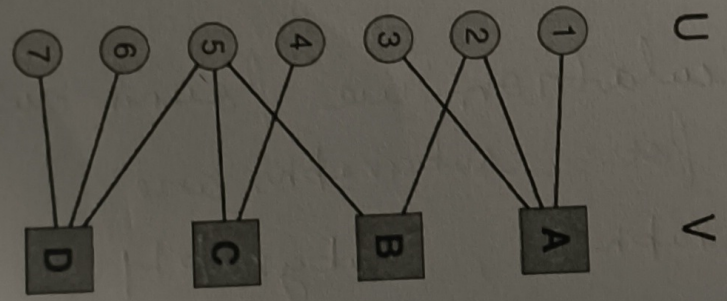
12. For the research article characterizing the 'Airport Network of India,' India's civil aviation infrastructure, answer the following: [2+1+1+1 = 4]

(a) What is the definition of a node and an edge?
(b) What does the edge weight represent?
(c) Provide a succinct interpretation of the graph.
(d) Interpret each of the data points.



a) The node is the airports
the edges is No. the flights
travelling per week from
airport

13. For the following bipartite graph, draw the monopartite 'Projection U' and 'Projection V'. [1+1 = 2]



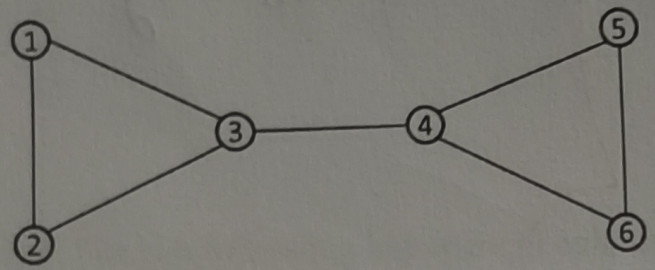
Projection U	Projection V

14. Prove that in the limits of large N and small $< k >$, the binomial degree distribution leads to the Poisson degree distribution. State all the approximations clearly and mention every step of the derivation. **[4]**

Initially the probability of random graph

BONUS QUESTION

15. Compute the value of Betweenness for '**node 4**' in the following graph. Clearly mark all the steps. **[2]**



Node 2 1 Node 3 Node 4 (5)