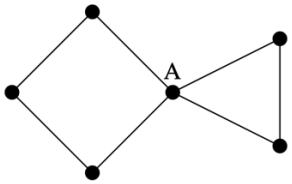
BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

I SEMESTER 2022-2023 CS G519 – Social Media Analytics, Comprehensive Exam (Part-A closed book)

Weightage: 22% **Time: 3 Hours** **Date: 20th Dec 2022**

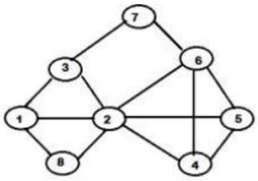
1. What are the limitations of BB's and Leung's concept-based approaches for query clustering? Explain through an example. [2]

2. Calculate the betweenness centrality of A in the following network [1]



3. (a) Draw centralized and non-centralized graphs of 4-nodes each.
   (b) suppose a network refers to traffic flow network of a city. If you are interested in finding the section that can most frequently control traffic in the network, which centrality measure is the most appropriate and why?
   [4]

4. Find communities for the given graph using clique percolation method. [4]



5. What is the randomization test for influence and homophily. How do you measure them using it? [3]

6. There is assortativity by degree in a network where connections are created based on the degree i.e., high degree nodes will be preferentially connected to other high-degree nodes, and the low to low. Another term is disassortativity. Define it and give one-one real life examples of each one of these. [3]

7. MNIST dataset has thousands of images of handwritten digits. Time series are generated from MNSIT images from top left pixel to right bottom pixel. You need to prepare the generated TS dataset for classification. What is/are the invariant would you like to use and why? [3]

8. What are models of diffusion of innovations? Pointwise describe their differences and observations. [3]

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

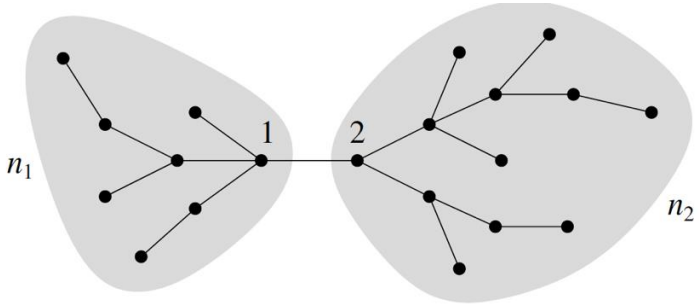
I SEMESTER 2022-2023 CS G519 – Social Media Analytics, Comprehensive Exam (Part-B Open book)

Weightage: 18% **Time: 3 Hours** **Date: 20th Dec 2022**

1. **Pinterest** is a web and mobile application that operates a software system designed to discover information on the World Wide Web. Registration is required to use it. The social networking features of Pinterest are built around the activity of collecting digital images and videos, and, in Pinterest terms, *pinning* them to a pinboard (collection). Each pin consists of the item, a user-generated brief description and a link back to the source of the item. Once a pin is created, other community members can add comments, *like* it, or *repin* it. Liking a pin will add the image to the Likes section of a user's profile but will not add it to their boards; repining an image allows a user to copy and categorize the image onto one of their own boards while maintaining the link back to the original web source. Comments are displayed beneath the item in a comment stream, like other social media sites.
   Pinterest is a social and transparent site; usernames, profiles, boards, and pins are viewable by other users and the general web public. Registered Pinterest users can opt to "follow" either the activity of other users or pinboards. Activity and certain statistics are public, which allows users to see the number of repins and likes a pin has received and to read and contribute comments. This metadata can act both as a form of social validation for users looking for information and resources (e.g., *Hundreds of people have repined this recipe. It must be good*") and as a social reward for users who pin content (e.g., *"Hundreds of people have repined this recipe. I must be cool*").
   (a) You are required to build an appropriate recommender system which recommends items, pinboards and other members of interest to a user.
   (b) Design an example inspired from the above network to show herd behavior.
   (c) If you want to measure influence in the network, what are the features would you use and why?
   (d) Which model of information diffusion from epidemics would you apply on Pinterest network? Justify the suitability of the model.

   [4+2+3+2]

2. Given closeness centrality of a node *i* is ($C_i = n / \sum_j d_{ij}$). Consider the undirected graph of *n* nodes given below:



   A particular edge in the tree joins nodes 1 and 2 and divides the tree into two disjoint regions of n_1 and n_2 nodes as shown above. Show that the closeness centralities of the C_1 and C_2 satisfy the following:

$$\frac{1}{c_1} + \frac{n_1}{n} = \frac{1}{c_2} + \frac{n_2}{n}$$

   [3]

3. Apply Girvan-Newman algorithm to find communities in the following network. Show all the steps. How do you choose the number of communities in a network? [4]