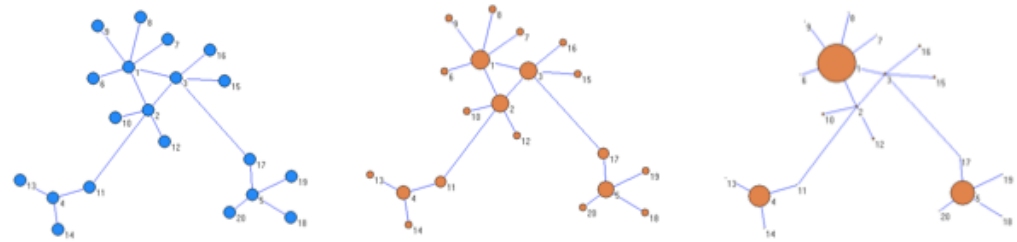
Social Computing – CS60017
Class Test
Nov 17th, 2016

Time: 2 Hours **Maximum Marks: 40**

Question 1:

Consider the figure given below and propose a design solution to the problem that follows.



(a) An illustrative network. (b) Weighting with PageRank. (c) A diverse weighting.

Figure 1: Network and node samples

Consider an example network illustrated in Figure 1(a). A simple random walk based sampling would rank the nodes as shown in Figure 1(b) with ranks proportional to size of the nodes. However, say the application demands that the nodes should be sampled as per the rank order in Figure 1(c), i.e., it should promote more diversity. In this context,

(i) Design a strategy (may be in the form of a score metric per node) that would enable a sampling as per the rank order in Figure 1(c) [i.e. larger size nodes should be at the top of the rank list and sampled first].

(ii) Propose a scheme that would take into consideration both the advantages of Figure 1(b) and (c) together while preparing the rank list for sampling.

[5+5]

Question 2:

Consider a community question-answering (CQA) platform like Quora. Routinely many questions on such CQA platforms remain unanswered. It would be very helpful for the moderators of they could automatically identify if a question would remain unanswered for long times (say till 3 months from the post). (i) Pose this problem as a supervised machine-learning task. (ii) List a set of five features that you would select for the task of supervision. State in one to two sentences reason behind choosing each feature. **[2.5+7.5]**

Question 3:

Consider the task of citation recommendation as follows: You are an author writing a new paper, and you need to cite previous 10 papers, relevant to your work (reference section). Assume that you have access to full dataset including your own paper, without references. Explain how would you design this as a supervised learning task using

(i) Topic Models: You are free to run any variation of topic models on your dataset.
(ii) Supervised Random Walks.

For each of these, you need to clearly mention how you would provide your training data, how the learning would take place and how would you validate your final system? **[6+6]**

Question 4:

In Twitter, suppose you want to build a recommendation system for "whom-to-follow." Also, assume that you have selected a list of top N followed accounts, and for each user, you know which of these N accounts, they are following. You need to now recommend to them some other accounts from these N (not yet followed) that they are interested in.

(i) Can you use a matrix factorization approach for this problem? Explain in detail how would you design your recommendation system, including the training phase
(ii) Now, suppose you also want to use the Network of the user. You may assume that there is an edge between two users if one of them has mentioned the others in atleast 3 tweets. Explain how would you modify the approach above to incorporate the network as well? **[5+3]**