L3S Research Center
Large Scale Data Mining, SS 2016
Dr. Avishek Anand
Assignment 4, due: **14 June 2016**

You must return your assignment sheet and have a correct solution in order to present in the exercise groups. Please write legibily! Do no forget to put your name and matriculation number on your solution!

Problem 1.

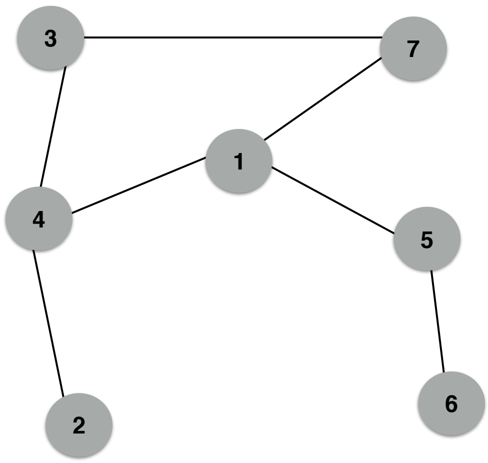

Figure 1: Sample Graph

1. For the graph in Figure 1 calculate the degree, closeness and betweenness centrality for node 5.
2. For the graph in Figure 1 calculate the edge betweenness centrality for edge (1,5).
3. Give an example graph where the nodes with the highest degree centrality, closeness centrality and betweenness centrality are different from each other.

Problem 2.

1. Give an algorithm to compute the degree centrality of a graph in map-reduce.
2. In the lecture we discussed how to compute edge betweenness centrality by the flow approach. Can you modify the algorithm to also compute node betweenness centrality ?
3. If your input graph G is a tree can you come up with a simplified procedure for closeness centrality ? If so how, and if not why not ?

Problem 3.
In the lecture we discussed how to use 2-hop querying on indexes of labels for unweighted undirected graphs. What extensions would you make to the discussed indexing technique to support distance queries on *directed unweighted* graphs.

1. What changes would you make to the index and labels ?

2. How would the query processing change ?

Problem 4.

1. Compute the labels for each vertex based on pruned landmark labelling for the given graph in Figure 1.
2. If we swap the node identifiers 1 and 6 will the index size (overall number of labels) increase or decrease ? Can you find the optimal node idenfier allocation which gives the smallest index size ?
3. Compare the query processing cost for both the above scenarios when the query is d(1,3).

Problem 5.

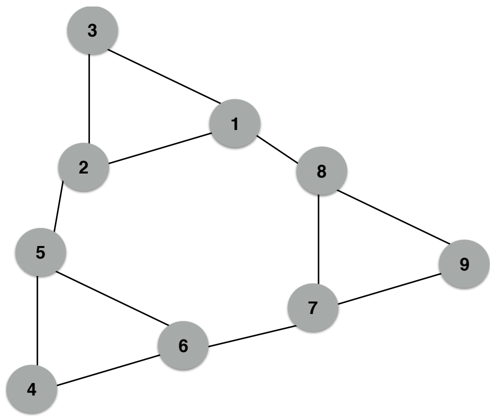

Figure 2: Clustering Graph

In Fig. 2 :

1. Use the Girvan-Newman approach to find the number of shortest paths from each of the following nodes that pass through each of the edges. (a) 3 (b) 2.
2. Using symmetry, the calculations of part 1 are all you need to compute the betweenness of each edge. Do the calculation.
3. Using the betweenness values from part 2, determine reasonable candidates for the communities in Fig. 2 by removing all edges with a betweenness above some threshold.