Network Science (CS 544) Date: 3.09.19

ASSIGNMENT 1

1. Given the Nepal earthquake dataset that contains the tweets information that is posted during the earthquake. Every tweet has a long list of attributes, including fundamental attributes such as id , created_at , and text etc.

   Considering the tweets as nodes, create two separate networks with edges between two nodes created based on the following respective conditions:

   a. Two tweets will be connected by an undirected edge if one has been posted within *t* time interval of the other. You may vary *t* from 15 to 60 minutes.

   b. There will be an edge between two tweets, if the cosine similarity between two tweet is greater than some threshold (assume 0.65), where the cosine similarity is calculated based on the similarity of the text vectors (after removing stopwords, punctuation, URLs and numbers), or one of them is a reply or retweet of the other.

   Use these networks to

   a. Find the important nodes from both the networks based on the centrality values calculated using the degree, eigen vector, katz, closeness and betweenness centrality metrics.

   b. Find out the overlapping nodes between both the networks based on the top 10 centrality values.

   c. Draw the inferences that you may draw from these overlapping tweets.