Assignment 3

COMP 596: Network Science

September 2019

Using the datasets from assignment 1, pick one of these options [one is enough for getting the full credit]:

1. Compare performance of 2 (or more) **node classification** algorithms on:

   (a) apply the algorithms to predict the labels on nodes of the graphs
   you are free to use any package or code off-the-shelf, or implement your own algorithm. For all results report average over 10 runs.
   Experiment on these three sets of data:
   real-classic: strike, karate, polblog, polbooks, football [30%]
   For these datasets, randomly drop the labels and report the accuracy as a function of the portion of the labels that are dropped, from 95% to 5%
   real-node-label: citeseer, cora, pubmed [50%]
   You can use or ignore the feature matrix based on the algorithm, use the same split for test and train as the GCN paper
   synthetic: LFR [20%]
   The common practice is to sample for varying values of μ which controls how well separated are the communities, i.e. generating synthetic graphs with $\mu = .1$ to $\mu = .9$, reporting average performance for 10 realizations at each difficulty level, and use n=1000, tau1 = 3, tau2 = 1.5, average degree=5, min community=20, for all, drop 20 percents of the labels randomly and report the accuracy of predicting them.

2. Compare performance of 2 (or more) **link prediction** algorithms on the same set of data [same scoring proportions as part1]. Here, drop 20% of edge at random, and report the AUC (average over 10 runs). For node attributed datasets, you can ignore or use labels depending on the algorithm.

bonus the best performing algorithm in each task [10%]

bonus doing both parts [30%]

If this assignment, Submit the report and code as separate attachments, by email and use COMP596Ass3 in the title.