MCS 590 – Foundations of Data Science
Fall 2017
Problem Set 3

Lev Reyzin

Due: 12/8/17 at the beginning of class

Instructions: Atop your problem set, please write your name and list your collaborators.

Problems

1. Find the sharp threshold for p for the existence of 4-cliques in $G(n, p)$. Prove your answer correct.

2. In class we showed that if the degrees in $G(n, \frac{1}{n})$ were independent, there would be a vertex of degree

$$d = \frac{\log n}{\log \log n}$$

with constant positive probability. However, the degrees are not independent. Show how to overcome this difficulty.

3. Give an example of a set H of hash functions such that $h(x)$ is equally likely to be any element of $\{0, ..., M - 1\}$ but H is not 2-universal.

4. For the k-median problem, give an upper bound on the ratio between the optimal value when we either require all cluster centers to be data points or allow arbitrary points to be centers.