Homework of Week 6

Deadline: 9:00am, December 1(Sunday), 2019

1. Bloom filters can be used to estimate set differences. Suppose Alice has a set X and Bob has a set Y, both with n elements. For example, the sets might represent their 100 favorite songs. Alice and Bob create Bloom filters of their sets respectively, using the same number of bits m and the same k hash functions. Determine the expected number of bits where our Bloom filters differ as a function of m, n, k and $|X \bigcap Y|$. Explain how this could be used as a tool to find people with the same taste in music more easily than comparing lists of songs directly.

2. Recall that $\mathcal{G}_n$ is the uniformly distributed n-vertex random graph, and that $\mathcal{G}_{n,p}$ is the n-vertex random graph each of whose edge appears independently with probability p. Prove that $\mathcal{G}_n$ and $\mathcal{G}_{n,\frac{1}{2}}$ are identically distributed.

3. (**Bonus score 5 points**) We know that $\lim_{n\to\infty} \Pr(\mathcal{G}_{n,p}$ has an isolated vertex$) = 1 - e^{-e^{-c}}$ when $p = \frac{\ln n + c}{n}$. Based on this fact, prove that $\lim_{n\to\infty} \Pr(\mathcal{G}_{n,m}$ has an isolated vertex$) = 1 - e^{-e^{-c}}$ when $m = \frac{n \ln n + cn}{2}$. (Hint: it may be helpful to follow the basic idea in proving the similar result of coupon collector problem.)

4. Do Bernoulli experiment for 20 trials, using a new 1-Yuan coin. Record the result in a string $s_1 s_2 ... s_i ... s_{20}$, where s_i is 1 if the i^{th} trial gets Head, and otherwise is 0.