# Tutorial 6

CSL-471 (Probability and Computing)

November 4, 2016

1. Consider the branching process for a contagion. Given that the contact network exists in the form of a tree and every node except the leaves have 4 children each. Moreover, there are 11 levels in the tree from level 0 to level 10. Every level except the last do not have any leaf node. The probability of disease transmission across every link is 0.5. Answer the following questions.
   (a) Given a lady Elsa, standing at level 5 in this tree, what is the probability that the infection reaches level 10 by passing through Elsa?
   (b) Is the disease going to be an epidemic? If yes, how can you alter the given parameters such that the contagion can not convert into an epidemic?
2. Give the coalescent process, where there are $10^6$ individuals in every generation. Answer the following questions.
   (a) Given there are 100 lineages present at a given generation, what is the probability that the number of lineages in the previous(parent) generation is also 100, i.e. No two lineages collide?
   (b) Given there are 100 lineages present at a given generation, what is the probability that 3 or more than 3 lineages merge in the previous generation?
   (c) Find the expected number of generations one encounters in the past when number 0f lineages= (i)10, (ii)$10^2$, (iii)$10^3$, (iv)$10^4$, (v)$10^5$,(vi)$10^6$? What do your infer from these observations?
   (d) Find the expected number of generations one encounters in the past when number 0f lineages= (i)10, (ii)8, (iii)6, (iv)4, (v)2? What do your infer from these observations?
3. In hubs and authority analysis, we see that a symmetric matrix can be written as, $M = p_1 z_1 + p_2 z_2 + ... + p_n z_n$, where $z_1, z_2, ...$ are eigen vectors of the matrix. In this presentation prove that why the value of $p_1$ can not be zero.
4. In hubs and authority convergence analysis, we considered $|c_1| \geq |c_2| \geq |c_3| .... \geq |c_n|$. Whether this process will converge or not if $|c_1| = |c_2| = |c_3| ... = |c_l|$, where $l < n$.
5. What is Perron's Theorem?
6. What is the best known general attack to break the pre-image resistant property? Describe the attack and also derive it's complexity in terms of the bits used for hashing. Hence, comment on the hash size one must choose. Describe a balls and bins problem that precisely models the same attack. You can show the analysis for the attack in terms of the modelled problem as well.

7. What is the birthday attack? Describe the relationship between the probability of success in the birthday attack anf the bits used for hashing.
8. In Chao's estimator for species accumulation problem, what will be the expected value of Sample Coverage $C$ in terms of the probabilities of occurence of species $p_i$'s, the sample size $n$ and the total number of unique species $N$.
9. Explain two real world scenarios which are similar to the species accumulation problem and hence Chao's estimator may be useful in coming up with a solution.
10. Considering a Wikipedia article a collection of the edits made by various editors over time, is there any way to use Chao's estimator to predict the completeness of a given Wikipedia article? Explain in detail.