Assignment 1

You can check Lada Adamic's social network analysis course video lectures Week 2: Random Graph Models for problem 1

1. Given the definition of following models:

- Static geographic model: Given N nodes randomly dispersed in the space, each node has a fixed position and each node is connected to its m nearest neighours, m is the same for all the nodes
- Random growth model: Similar with preferential attachment growth model (Barabasi-Albert model), except that for each new added node, the probability p_i that the new node is connected to node i is the same for all the existed nodes $p_i = \frac{1}{\sum_j \mathbf{1}} = \frac{1}{n}$, n is the total number of existed nodes when the new node is added

In this question we will compare these two models with Erdos-Renyi model and preferential attachment model.

(a) For an Erdos-Renyi model with N nodes and link probability p, what is the expected degree $\bar{k}$?

(b) Consider a static geographic model with N nodes and $m \approx \bar{k}$, then compare it with the previous Erdos-Renyi model, whether the following statements are true or false, and give your explainations?

1. the static geographic model has stronger locality
2. the static geographic model has shorter average shortest path

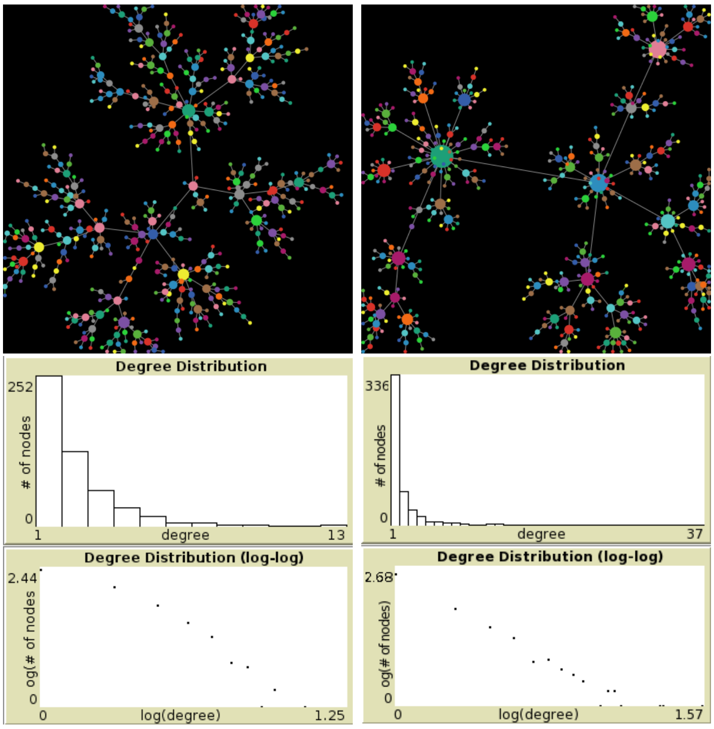
(c) Compare random growth model with Erdos-Renyi model, explain why links are unevenly distributed in random growth model? Which model has more nodes with degree 1?

(d) Compare random growth model with preferential attachment model, for the following two pictures, each has 500 nodes, which model fits best to each picture and explain why?

(e) Suppose for random growth model, when each new node is added, it's connected to m other nodes, and at each time only one new node is added, then at time t, consider the degree $K^{t_0}(t)$ of the node which was added at time t_0, it satisfies

$$\frac{dK^{t_0}(t)}{dt} = \frac{m}{t}$$

with $K^{t_0}(t_0) = m$, give the solution for $K^{t_0}(t)$?



(a) Model1 (b) Model2

(f) For preferential attachment model,

$$K^{t_0}(t) = m(\frac{t}{t_0})^{\frac{1}{2}}$$

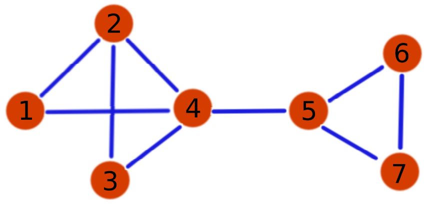
when t_0 is small, $K^{t_0}(t)$ is large, which means old nodes have large degrees at time t. Now consider when t is large, the percentage of nodes with degree smaller than $K^{t_0}(t)$ is

$$P(k \leq K^{t_0}(t)) = \frac{t - t_0}{t}$$

explain why and give the degree distribution $P(k)$?

You may check this notebook as a reference for problems 2 and 3

2. Given the following network,



(a) What is the normalized degree centrality for node 2? What are the normalized betweenness centrality, normalized closeness centrality and clustering coefficient for node 4?

(b) What is the value of network constraint for node 5?

(c) What is the modularity matrix for this network? You can check with the notebook's results and also see how a brutal force search is used to find the optimum community strture with maximum modularity.

3. Plot your facebook ego-network,

- Go to http://snacourse.com/getnet
- Choose which user data (e.g. "wall posts count") you would like to include, check all the checkbox
- Be patient. If you have lots of Facebook friends, the app will fetch the data gradually, and report on its progress (e.g. (30 out of 100, 50 out 100...). Once its finished, at the bottom of the page it will provide a download link. Right click on 'gml file' and save it to your computer
- Plot your facebook networks, using Gephi or the sample NetworkX codes provided in the notebook. If you checked all the checkboxes when you download the data, you will only need to change the gml file path in the sample code and change the node sizes and colors according to your taste.

Bring your answers with your printed facebook networks to the course on 25th/Mar. If the questions are not clear for you, or you have any other questions, contact TAs as soon as possible.