Seminar of Computer Networks

Homework 1

You can do the homeworks alone or in groups of two students. Write a single solution with both names.

Make sure that the solutions are typewritten or clear to read.

Hand in your solutions and keep a copy for yourself. After the due date we will post the solutions and in the final exam we will ask you to explain us what were your mistakes.

Due date: 18/5/2011, before the class.

Problem 1. Lex X be a discrete random variable that takes values different than 0 with expectation $\mu = \mathbf{E}[X]$. Define the random variable $Y = 1/X$. Does it hold in general that $\mathbf{E}[Y] = 1/\mu$? You should provide a simple proof if you believe that it is true, or a simple counterexample if you believe that it isn't.

Problem 2. Consider a random graph $G = (V, E)$ with $|V| = n$ nodes created according to the Erdős-Rényi $G_{n,p}$ random-graph model. Let us define a *3-star* to be a subgraph $V' = \{v_0, v_1, v_2, v_3\} \subseteq V$ of V such that all the edges (v_0, v_i) for $i = 1, 2, 3$ exist in G, and none other of the edges (v_1, v_2), (v_1, v_3), and (v_2, v_3) exist in the graph (but other edges may exist, including edges between nodes in V' and other nodes in V).

1. What is the expected number of 3-stars in G?
2. Define similarly a *k-star* (a subset $V' = \{v_0, v_1, \ldots, v_k\} \subseteq V$ such that for $i = 1, \ldots, k$ we have that $(v_0, v_i) \in E$, and for $i, j = 1, \ldots, k$ we have that $(v_i, v_j) \notin E$). What is the expected number of k-stars in G?

Problem 3. Consider the following modification of the Barabassi-Albert preferential attachement model that we did in class: When a new node arrives at time t again it comes with ℓ edges. However, this time each edge selects a node v with probability proportional to the degree d_v plus a constant c, that is, the probability equals

$$\frac{d_v + c}{(t-1)(2\ell + c)},$$

where $c \geq -\ell$, as we describe at the end of Chapter 4 in the notes (so for $c = 0$ this is the Barabassi-Albert model). Show that the degree distribution that we obtain as $t \to \infty$ is approximately a power law with exponent $3 + c/\ell$.

Problem 4. Let $G = (V, E)$ be the circle graph, that is, the graph with node set $V = \{v_1, v_2, \ldots, v_n\}$ and edge set the $E = \{(v_i, v_{i+1});\ i = 1, \ldots, n\}$ (where we assume that the edge (v_n, v_{n+1}) stands for edge (v_n, v_1)). Find the best partitioning for G that maximizes the modularity (and show that it is the best possible).

Note: Use the following fact without having to prove it: Among all the possible sets of n numbers $a_1, a_2, \ldots, a_n$ for which we know that the sum $\sum_{i=1}^{n} a_i = L$ is some fixed value L, the sum of the squares $\sum_{i=1}^{n} a_i^2$ is minimized when we have that $a_1 = a_2 = \cdots = a_n = L/n$.

Problem 5. Write a simple program in the programming language that you prefer to calculate the degree distributions of the graphs in

`http://www.dis.uniroma1.it/~socialnet/graphs`

and plot them in regular and in log-log scale. For plotting you can also use any program that you wish (Excel, OpenOffice, Matlab, R, gnuplot, etc.).

Next find the best power-law function that fits the data. To do that there are various difinitions of "best." A simple one (not necessarily the best) is the following. Recall that a power law distribution function is a distribution function of the form

$$y(x) = b \cdot x^a.$$

If we take logarithms in both sides we have

$$\ln y(x) = \ln b + a \ln x,$$

that is, we have a linear relationship between $\ln x$ and $\ln y$. We can find the parameters a and b that create the line that best fits a set of data (x_i, y_i) by minimizing the mean-square error of the logarithms. That is, we find the values of a and b that minimize the expression

$$\text{RSS} = \sum_{i=1}^{N} \left(\ln y_i - (a \ln x_i + \ln b)\right)^2,$$

where (x_i, y_i) is an input pair, and N the total number of points. In our case a pair (x_i, y_i) corresponds to a pair (degree, number of nodes with that degree), and N is the number of different degrees. This would give the line that best fits the points in the log-log plot as measured by the *residual sum of squares* (RSS) error.

Since the relationship between the $\ln y$ and $\ln x$ is linear, it turns out that the values of a and $\ln b$ that minimize this expression are given by

$$a = \frac{S_{xy}}{S_{xx}} \qquad \text{and} \qquad \ln b = \widehat{\ln y} - a\,\widehat{\ln x},$$

where

$$\widehat{\ln x} = \frac{1}{N} \sum_{i=1}^{N} \ln x_i$$

$$\widehat{\ln y} = \frac{1}{N} \sum_{i=1}^{N} \ln y_i$$

$$S_{xx} = \sum_{i=1}^{N} (\ln x_i - \widehat{\ln x})^2$$

$$S_{xy} = \sum_{i=1}^{N} (\ln x_i - \widehat{\ln x})(\ln y_i - \widehat{\ln y}).$$

After you compute the parameters a and b, plot in the same figures with the points (x_i, y_i) that you ploted earlier the function $y = b \cdot x^a$.

Mail the code that you write to compute the degree distributions `aris@cs.brown.edu`, with subject: "Seminar Computer Networks - Homework 1."