ECS 253 / MAE 253, Network Theory
Spring 2016
Problem Set # 2, Solutions

Problem 1: Clustering coefficient of the Erdős-Rényi random graph

(a) We can denote the average degree of an Erdős-Rényi random graph $G(n,p)$ as m; so, $p = \frac{m}{n}$. Picking from all possible combinations of 3 nodes for which a triangle could exist is $\binom{n}{3}$. Since each edge is added independently, the probability that a triangle is created between a set of 3 nodes is $p^3 = \left(\frac{m}{n}\right)^3$. Thus the number of expected triangles can be derived as follows:

$$\begin{aligned}
&\binom{n}{3}\left(\frac{m}{n}\right)^3 \\
&= \frac{n!}{3!(n-3)!}\left(\frac{m}{n}\right)^3 \\
&\approx \frac{(n-3)!}{6(n-3)!}m^3 \\
&= \frac{1}{6}m^3
\end{aligned}$$

(b) Connecting a set of 3 vertices by 2 edges has 3 different possibilities, each with probability $\left(\frac{m}{n}\right)^2$ (since each edge is added independently). Since there are still $\binom{n}{3}$ ways to pick a set of 3 nodes, we can derive the expected number of connected triples as follows:

$$\begin{aligned}
&3 \cdot \binom{n}{3}\left(\frac{m}{n}\right)^2 \\
&= 3 \cdot \frac{n!}{3!(n-3)!}\left(\frac{m}{n}\right)^2 \\
&\approx 3 \cdot \frac{(n-2)!}{3!(n-3)!}m^2 \\
&\approx 3 \cdot \frac{n}{6}m^2 \\
&= \frac{1}{2}nm^2
\end{aligned}$$

(c) The clustering coefficient, $\mathcal{C}$, is the equation found in (a) divided by the equation found in (b). Simplifying:

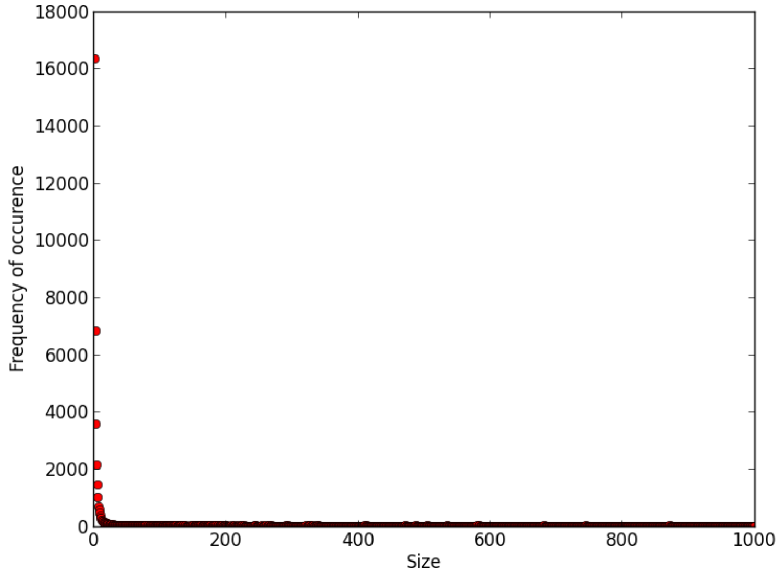
$$\mathcal{C} = \frac{\frac{1}{6}m^3}{\frac{1}{2}nm^2}$$

$$= \frac{1}{3} \cdot \frac{m^3}{nm^2}$$
$$= \frac{m}{3n}$$

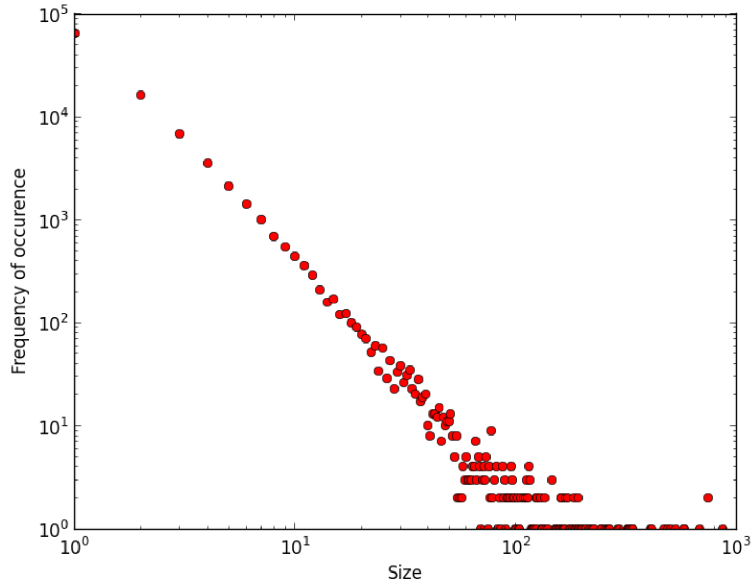
(d) If m is held constant, and n is increased, the clustering coefficient $\mathcal{C}$ approaches 0. This makes sense, due to the fact that the number of triangles (the numerator) is independent of the size of the graph, while the expected number of connected triples is dependent on n.

Problem 2: Fitting Power Law Distributions (Mostly thanks to Sam Johnson)

(a) Linear scale plot of "transformation mode" power law distribution:



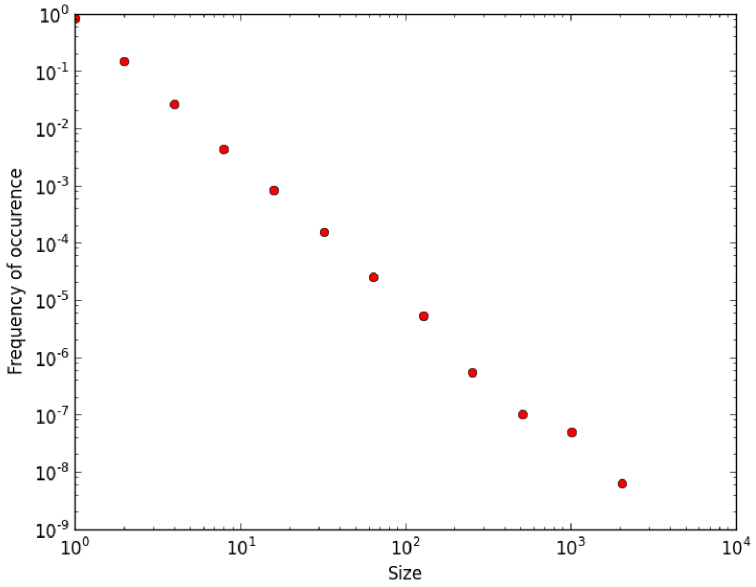
Log-log scale plot of "transformation mode" power law distribution:



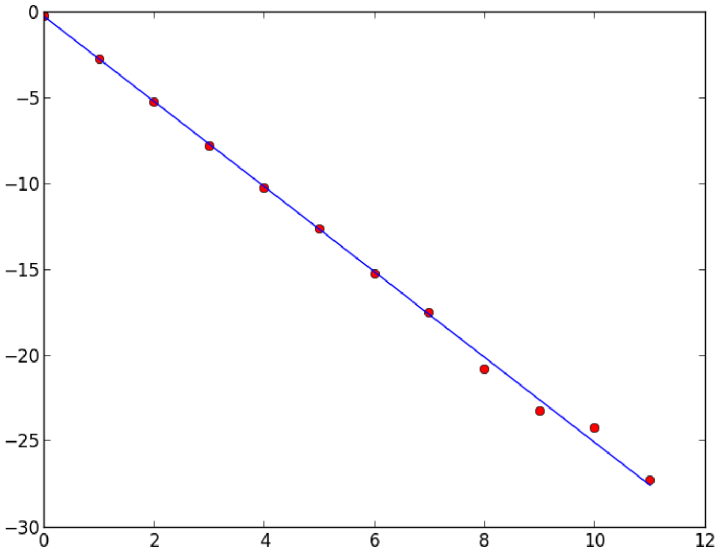
(b) For brevity, I will say that "size i occurs" if a point x is in the window $[i, i+1)$, as plotted in a). As seen in a), **the slope will be overestimated/less steep (i.e., γ will be underestimated), since the slope appears to be significantly greater than -2.5 due to the noise in the tail**. Note that the frequency of points whose size is greater than say, 100, will be nearly the same as in an exact power law, because the probability of choosing a point in this tail is not too miniscule compared to the number of samples. However, the number of times a specific size occurs, say 789, is extremely small and has a good chance of never occurring. Since many sizes do not occur at all, the ones which do

occur will tend to skew high, resulting in an overestimation of the slope. More formally, let $N_i(t)$ be the number of times that size i occurs after t samples. In order for the histogram to be an exact power law, it would be necessary for $N_i(t) = \langle N_i(t) \rangle$. But for large i, the probability of size i occurring is very small, so $N_i(t)$ is likely to differ greatly from its expectation as we observe clearly in a). It is possible to go deeper into the probability of this by noting that the distribution on $N_i(t)$ is Binomial and hence approximately Poisson. That is, $\Pr(N_i(t) = k) = \binom{t}{k} p^k (1-p)^{t-k} \approx \frac{e^{-p} p^k}{k!}$ where $p = A \int_i^{i+1} i^{-\gamma} di$.

(c) Histogram of data from (a) plotted with bins of exponentially increasing size:



(d) Linear least-squares fit of (c) = -2.48:



(e)

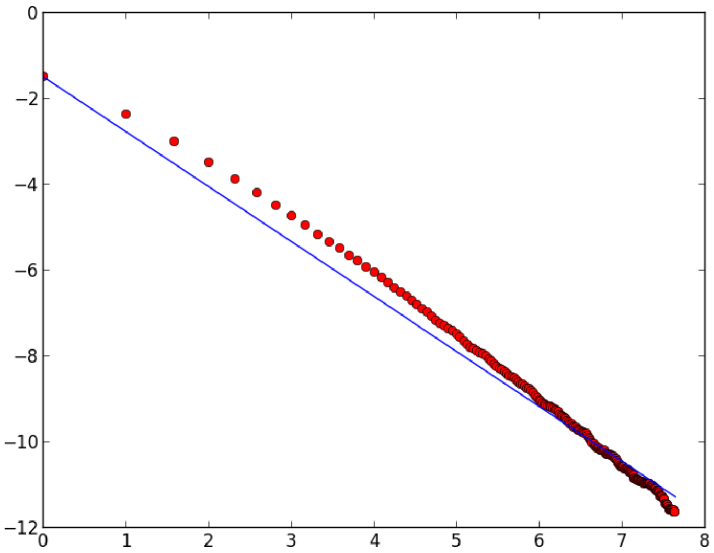
$$C(x) = A \int_x^{\infty} x^{-\gamma} dx$$

$$= \frac{A}{-\gamma+1}\left(\lim_{t\to\infty} t^{-\gamma+1} - x^{-\gamma+1}\right)$$

$$= \frac{A}{\gamma-1}x^{-\gamma+1}$$

$$= \frac{A}{\gamma-1}x^{-(\gamma-1)}$$

(f) The estimated slope of $C(x)$ found using linear least squares is approximately -1.28, which gives an estimate of $\gamma = -2.28$.



(g) In this example, exponential binning gave a much better approximation of the slope $\gamma = 2.5$.