Network Analysis and Modeling
CSCI 5352, Fall 2014
Prof. Aaron Clauset
Problem Set 6, due 11/19

1. (100 pts total) Consider Price's model of a citation network (Chapter 14.1 of *Networks*), applied to publications in a single field.

   (a) (35 pts) Implement the simulation algorithm described in Chapter 14.1.1. Choosing $c = 3$, $a = 1.5$ and $n = 10^6$, produce a figure showing the complementary cumulative distribution function $\Pr(K \geq k_{\text{in}})$ (the ccdf) for in-degree k_{in}. Figures whose axes are unlabeled will receive no credit. Briefly discuss the impact of the uniform attachment mechanism on the distribution's shape and comment about the fraction of vertices with $k_{\text{in}} = 0$.

   (b) (35 pts) Now use your numerical simulation to recreate Figure 14.4 from *Networks*, which shows the average in-degree of vertices as a function of their rescaled time of creation. Give a brief discussion of what this means in terms of the way individual vertices accumulate edges, and how this behavior varies with a.

   (c) (30 pts) Reasonable values of the model parameters for real citation networks are $c = 20$ and $a = 5$. For these choices, use your numerical simulation to calculate (i) the average number of citations to a paper (in-degree) in the first 10% of published papers (vertices) and (ii) the average number for a paper in the last 10%. Briefly discuss the implications of your results with respect to the "first-mover advantage," and the corresponding bias in citation counts for the first papers published in a field.

   Hint: Choose a reasonably large value of n.

   (d) (20 pts extra credit) Now consider a variation of Price's model of a citation network in which we remove the preferential attachment part. That is, each time a new vertex joins the network, each of its c edges attaches to an existing vertex with equal probability. Using the same parameter choices as in question (1a), produce a figure showing the ccdf for both this model and Price's model. Briefly discuss the differences in terms of how citations (edges) are distributed across papers (vertices).