A Random Graph Algorithm For Modeling Social Networks

Ryan Parker, Andrew Wilson

Introduction

Social networks represent an interesting reservoir of potential information on the workings of society and many researchers seek to model these networks in a random fashion. Three of the trademark features of social networks include a low average path length, a relatively high clustering coefficient, and a degree distribution following a power law [1]. Several attempts have been made to model these networks, such as those by Erdős and Rényi [2] and Barabási and Albert [3].

However, these attempts fail to display at least one of the three aforementioned properties. Here we attempt a novel algorithm to create multiple clusters of small, highly connected subgraphs of normal nodes using the ER method, which are connected to each via a cloud of high-degree 'influencer' nodes, reminiscent of the BA method. In doing so, this method seeks to simulate the power law, while still maintaining low average path length and high clustering coefficients.

Methods

To construct the random graphs, the following algorithm was performed with parameters (n,p,k,m,q):

1. k ER(n,p) graphs were created [2].
2. m additional influencer nodes were initialized.
3. For each influencer node, and for each of the other nk+m-1 nodes, an edge was placed, independently and with probability q between them.

Due to computational limitations, a scaled-down version of the Instagram network was simulated using various parameter sets intuited from the full-size network. Expectations of various metrics were derived analytically, as well as calculated from the simulations.

Figure 1: Social Network graph properties. **(a)** Low path length and high clustering. **(b)** Power law degree distribution..

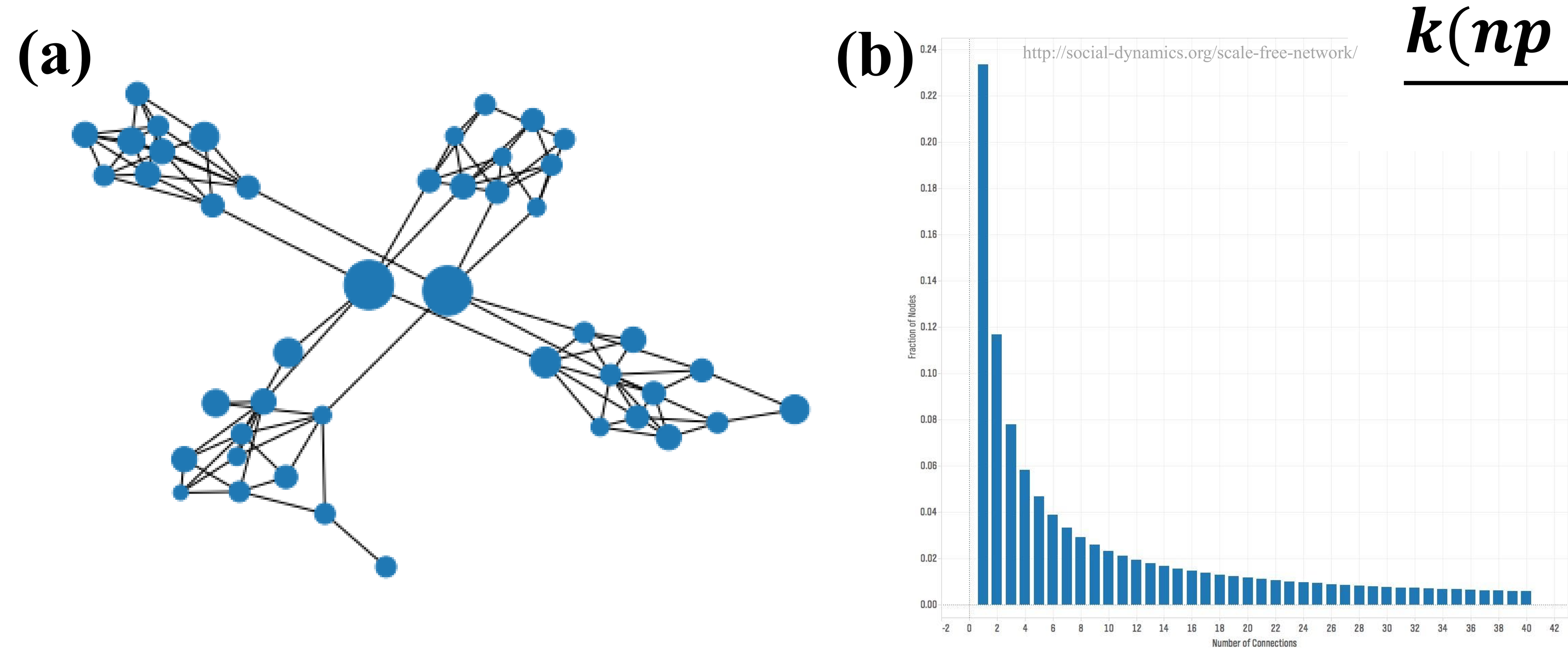
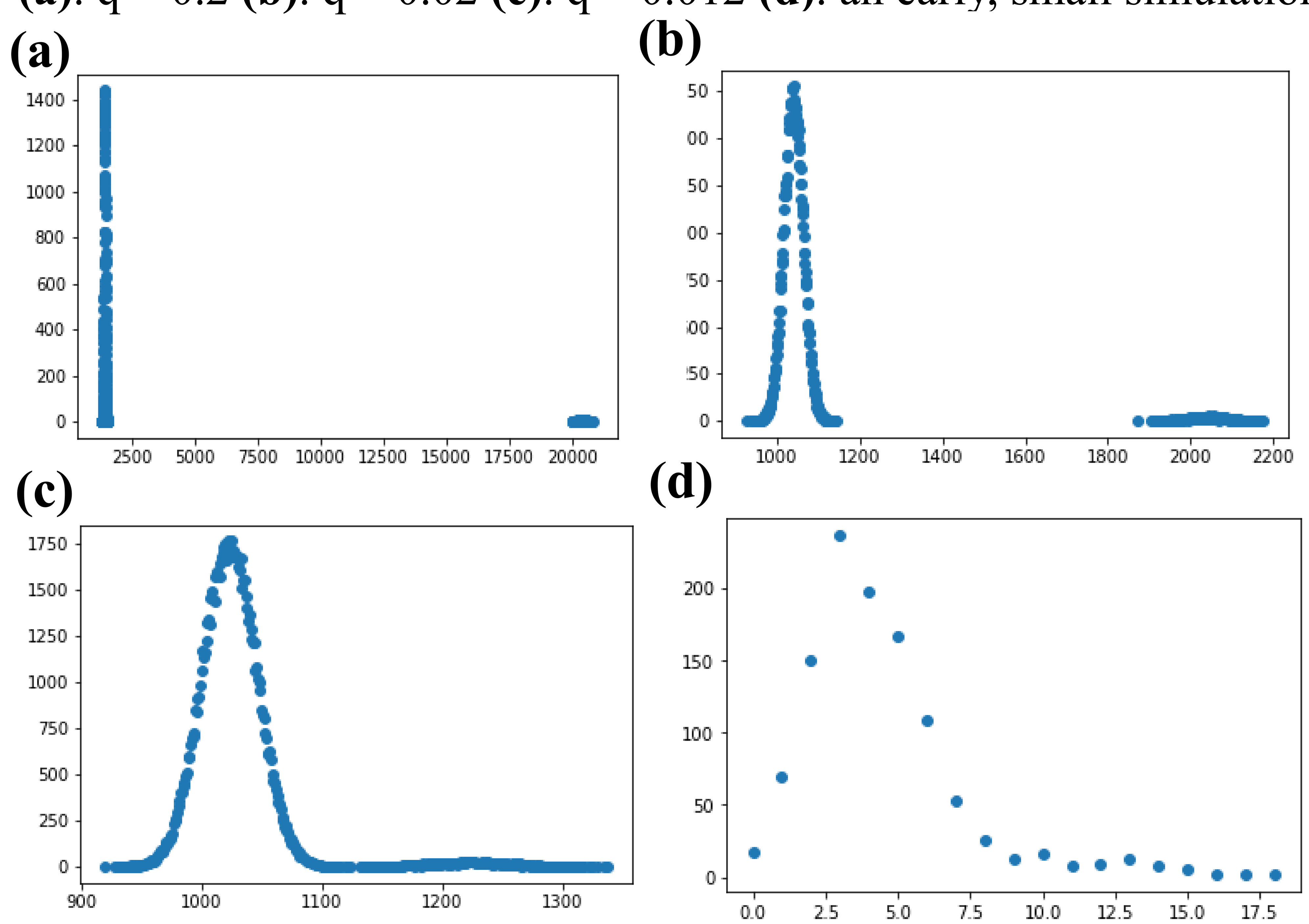


Table 1: Comparison of expected and calculated metrics from the random graph model using different values of q.

	q = 0.2		q = 0.02		q = 0.012	
	Expected	Actual	Expected	Actual	Expected	Actual
Normal Node Connections	1400	1399	1040	1039	1024	1023
Influencer Node Connections	20400	20398	2040	2039	1224	1223
Total Connections	9.04x10^{7}	9.04x10^{7}	5.40x10^{7}	5.40x10^{7}	5.24x10^{7}	5.24x10^{7}
Average Path Length	<4.614	1.986	<5.330	2.420	<5.630	2.694
Clustering Coefficient	0.3463	0.3517	0.4547	0.4634	0.4682	0.4768

Figure 2: Degree distributions of the random graph simulations. **(a)**. q = 0.2 **(b)**. q = 0.02 **(c)**. q = 0.012 **(d)**. an early, small simulation



Equation 1: Probability Mass Function of Degree Distribution

$$P(deg(V) = y) = \frac{k(np+mq)^y e^{-(np+mq)} + (knq+mq)^y e^{-(knq+mq)}}{y!\,(k+1)}$$

Results

Most experimental and calculated metrics correlate very well (Figure 1). An obvious difference exists in average path length, where the experimental number is far lower than the expected value, which was calculated as an upper bound. The intended degree distribution is a large Poisson curve, with an extended right tail, which is observed with lowest q-value, but not with the higher values, for reasons discussed below (Figure 2).

Discussion

The approximate power law is enabled by the clusters forming stacked Poisson distribution with the influencers forming the right tail, assuming a well-tuned q. The clusters further serve to shield the clustering coefficient, allowing it to approach p when nk>>m. The path length is heavily dependent on q and m, as most paths must pass through the influencer cloud. These results demonstrate a very low path length, and one far smaller than the expected due to the scaling down of the network, which led to an artificially inflated q. During experimentation, the path length was increased to nearly 4.0 before the network disconnected, by decreasing q.

Next steps involve determining an intuitive guideline to select q and testing the algorithm on an upscaled network. This latter will, however, require significant optimization of the code and considerably more computational resources.

References

1. Ugander, Johan et al. "The Anatomy of the Facebook Social Graph." ArXiv abs/1111.4503 (2011): n. pag.
2. Erdos, Paul L. and Alfréd Rényi. "On the evolution of random graphs." Transactions of the American Mathematical Society 286 (1984): 257-257.
3. Albert, Réka and A L Barabasi. "Statistical mechanics of complex networks." ArXiv cond-mat/0106096 (2001): n. pag.