Statistics of the network of organic chemistry – Supplementary Electronic Information

Philipp-Maximilian Jacob, Alexei Lapkin*

Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK

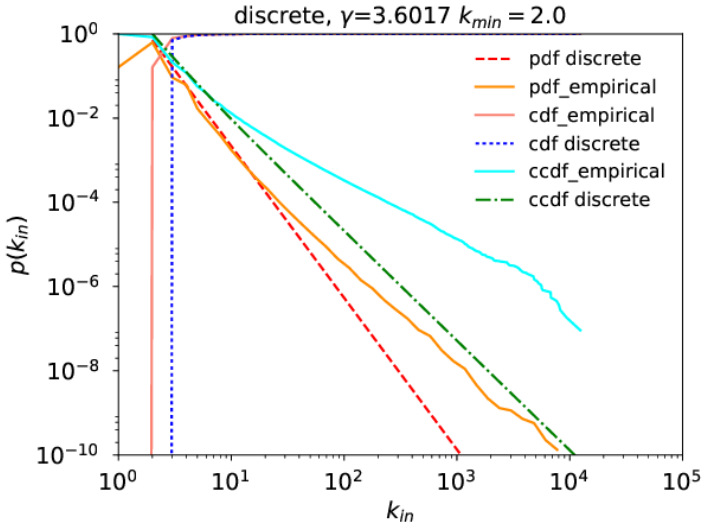

Figure S-1. Plot of $p(k_{in})$ for $k_{min} = 2.0$.

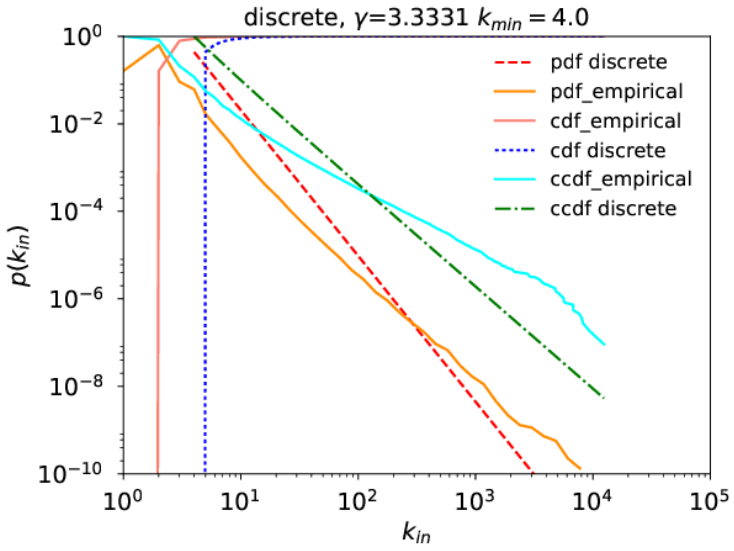

Figure S-2. Plot of $p(k_{in})$ for $k_{min} = 4.0$.

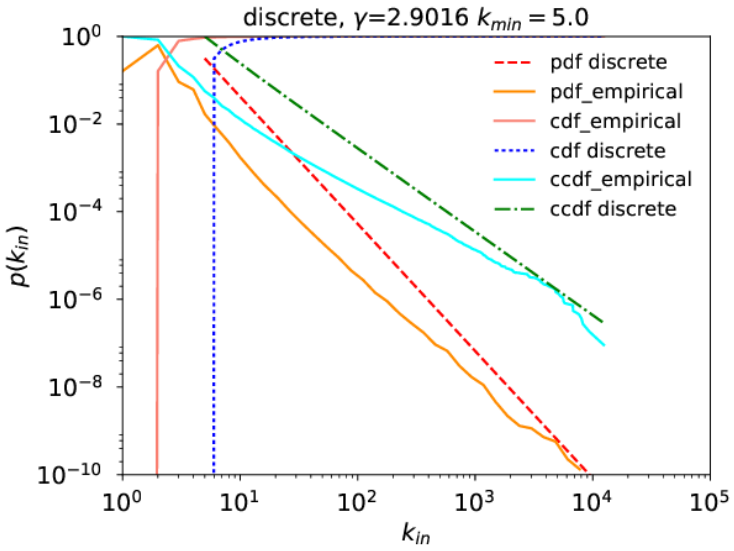

Figure S-3. Plot of $p(k_{in})$ for $k_{min} = 5.0$

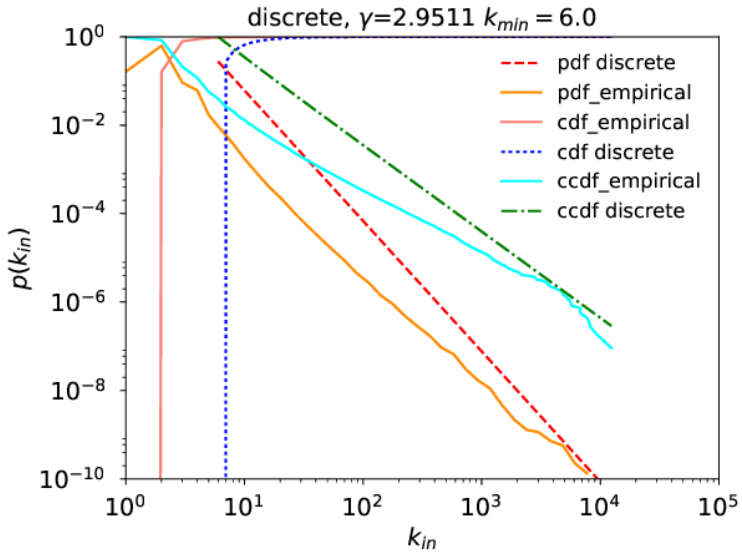

Figure S-4. Plot of $p(k_{in})$ for $k_{min} = 6.0$.

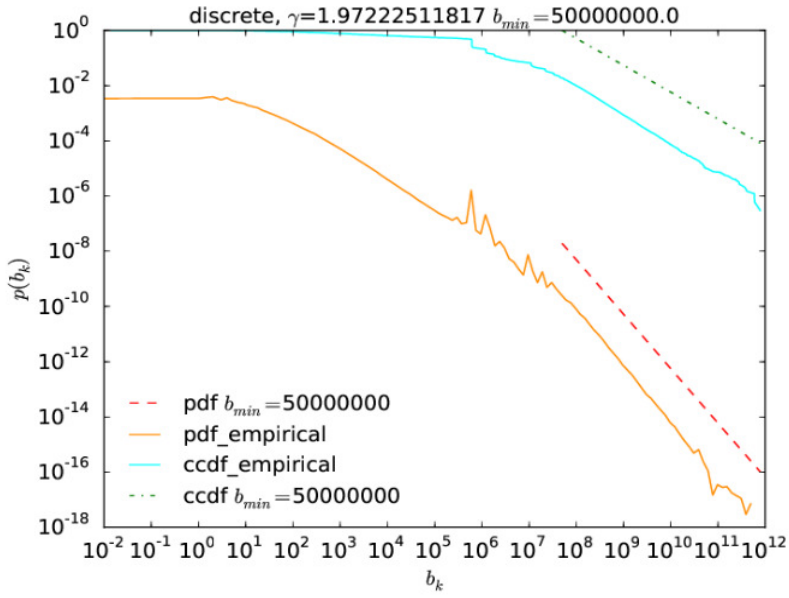

Figure S-5. Distribution functions giving the probability of observing a given betweenness centrality value for $1.0 \leq b_k \leq 10^5$. This is compared to the power-law exponent calculated using powerlaw for the different regimes observed. Curves labelled as "pdf" are probability density functions giving the probability of observing a given value of b_k. "cdf" denotes the cumulative distribution function, giving the probability that the betweenness centrality will be less than or equal to b_k. The "ccdf" is the complementary cumulative distribution function giving the probability of the betweenness centrality being greater than b_k. If a curve additionally carries the label "empirical" this denotes that this is the actual observed data while a curve not carrying this label shows the model's values.

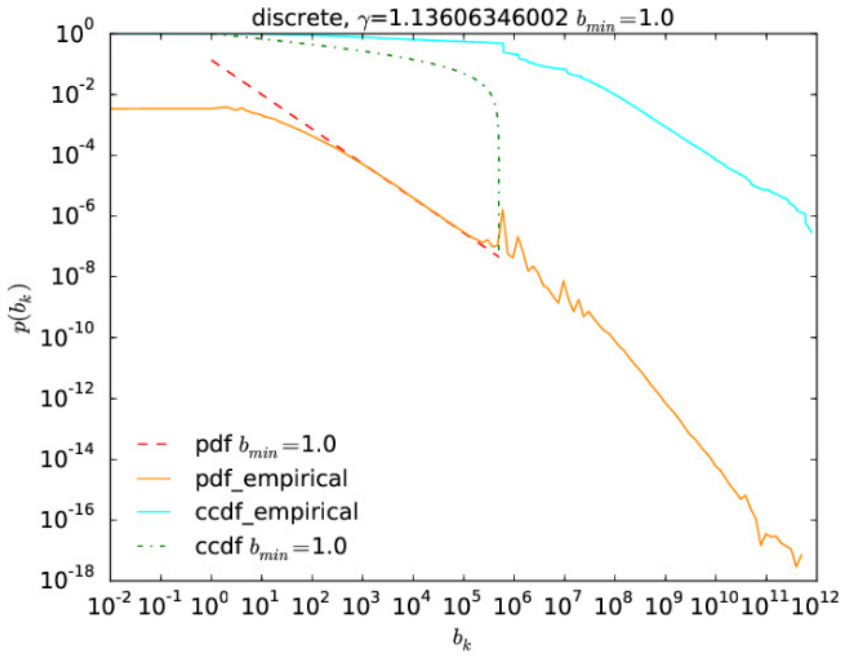

Figure S-6. Distribution functions giving the probability of observing a given betweenness centrality value for $10^5 \leq b_k$. This is compared to the power-law exponent calculated using powerlaw for the different regimes observed. Curves labelled as "pdf" are probability density functions giving the probability of observing a given value of b_k. "cdf" denotes the cumulative distribution function, giving the probability that the betweenness centrality will be less than or equal to b_k. The "ccdf" is the complementary cumulative distribution function giving the probability of the betweenness centrality being greater than b_k. If a curve additionally carries the label "empirical" this denotes that this is the actual observed data while a curve not carrying this label shows the model's values.

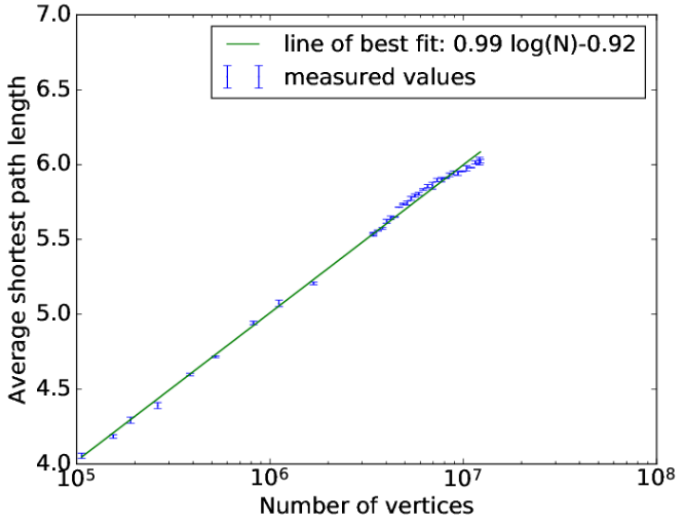

Figure S-7. The average shortest path length versus the number of nodes in the network. The number of nodes are plotted on a logarithmic axis showing lines of best fit to illustrate growth in the average shortest path length compared to $log\ N$.

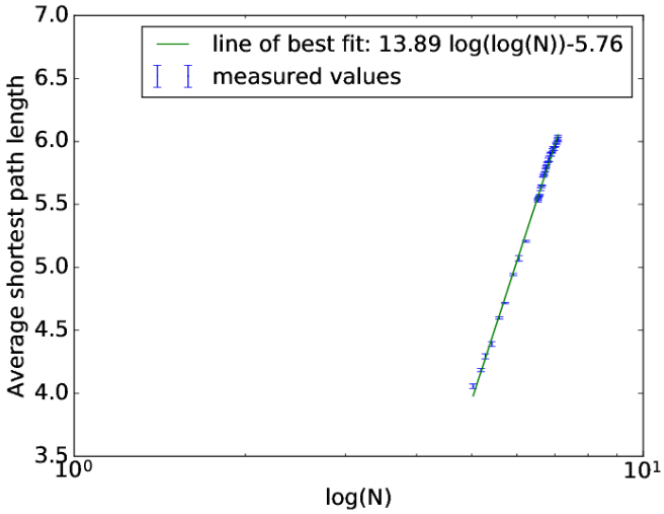

Figure S-8. The average shortest path length versus the logarithm of the number of nodes in the network. The logarithm of the number of nodes are plotted on a logarithmic axis showing lines of best fit to illustrate growth in the average shortest path length compared to $log\ log\ N$.