SCALING LAWS IN EMPIRICAL NETWORKS

Alexander W. Ray and Aaron Clauset

SIAM Workshop on Network Science 2020
July 9–10 · Toronto

Scaling laws provide a means by which to estimate the asymptotic growth behavior of a quantity from finite data. Well-known scaling relationships in other scientific domains–such as the scaling of resting metabolic rate with organismal mass [3] as well as that of city populations [1]–have been used productively to identify general patterns around which to build models and characterize deviations. Because sufficient empirical network data has been lacking, the manner in which real-world network structure scales has remained unexplored. The discovery of empirical scaling laws for statistical measures of network structure would shed new light on how networks change as they scale up and whether they exhibit "universal" behavior, providing new tests and targets for mathematical models of network structure.

Previously, network scaling has largely been studied using random graphs and generative models–often using the tools of statistical physics–whose accuracy in relation to real-world scaling behaviors often remains untested on a large scale. Despite their popularity, random graphs with fixed edge density or degree distribution are known to be unrealistic models in many contexts. They appear to broadly reproduce the low mean geodesic path lengths found in most networks, but generally fail to reproduce high values of global clustering coefficient, a hallmark of social networks [4]. A detailed, quantitative study of network scaling behavior would provide a clearer understanding of how random graphs differ from real-world networks across scales.

Here, we investigate the empirical scaling behavior of mean geodesic distance, global clustering coefficient, and degree assortativity as a function of network size, using a structurally diverse corpus of 254 networks from the Index of Complex Networks (ICON) [2], spanning four scientific domains and six orders of magnitude in size (Figure 1). First, we estimate an empirical scaling law for each structural measure. Second, we characterize the degree to which random graphs parameterized by edge density or by a degree sequence can explain the observed scaling behavior. Finally, we introduce a new random graph model which generalizes the configuration model to allow parameterized "edge localization" and show that this model produces more realistic scaling than the standard configuration model.

We find that empirical networks exhibit average shortest path lengths that robustly scale like $O(\log n)$. However, we also find that triangle densities generally scale like $O(1/n)$, even in social networks, a pattern that is broadly consistent with random graph theory, indicating that degrees and randomness alone appear to play a larger role in shaping large-scale network patterns than previously recognized. As networks scale up, we find that the residual scaling, i.e., the scaling left unexplained by degree structure alone, increases steadily, with notable differences between domains. The residual scaling of clustering coefficient, in particular, indicates that real-world networks exhibit more edge "localization" than expected given degree sequence alone, even in non-social networks. By parameterizing a form of edge localization, we present one explanation for the observed deviations, demonstrating improvement in comparison to the other tested models. These insights can help motivate the development of new classes of network models which may more accurately represent the structure of real-world systems for use as substrates for modeling dynamical processes or as null models against which to identify interesting structural patterns. Our findings can provide context for comparisons of random graphs to real-world networks and serve as quantitative targets for network models–like the stochastic block model and preferential attachment–across empirical domains.

References

[1] L. M. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104(17):7301–7306, 2007.

[2] A. Clauset, E. Tucker, and M. Sainz. The Colorado Index of Complex Networks. `https://icon.colorado.edu/`, 2016.

[3] M. Kleiber. Body size and metabolism. *Hilgardia*, 6(11), 1932.

[4] M. E. J. Newman and J. Park. Why social networks are different from other types of networks. *Physical Review E*, 68:036122, Sep 2003.

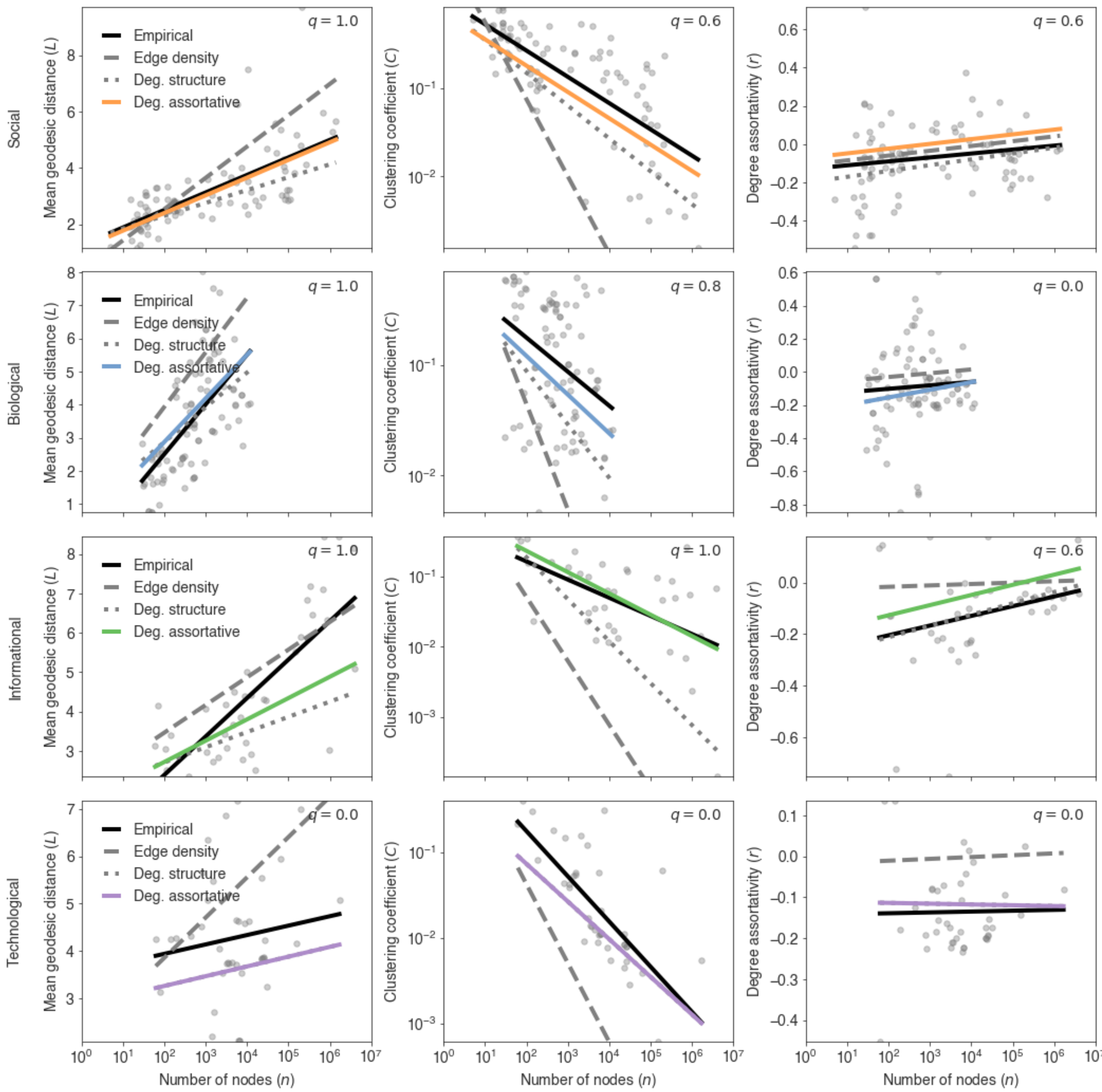

Figure 1: Scaling behaviors of mean geodesic distance L, degree assortativity r, and global clustering coefficient C as a function of the number of nodes n for 254 empirical networks. Scaling behaviors are listed for the Social, Biological, Technological, and Informational network domains. For each summary statistic and network domain, relationships are presented for empirical data (solid black line), random graphs with fixed edge density (dashed gray line), fixed degree structure (dotted gray line), and degree-assortative random networks (solid colored line). Relationships for null models are determined taking the average value of the summary statistic over 20 samples for each empirical network and calculating a least-squares regression over the null values. The value of q chosen for the degree-assortative networks used in a given statistic-domain combination is the value that minimizes the slope of the residual scaling, where residual scaling is defined as the scaling of the ratios of the statistic value of an empirical network and the average value for the null model; higher q corresponds to greater localization of edges.