

Breaking the Power Law: Improved Model Selection Reveals Increased Network Complexity

Debra S. Goldberg¹, Giovanni Franklin², and Frederick P. Roth³

Keywords: systems biology, genomic networks, complexity, model selection, maximum likelihood

1 Abstract.

Network analysis has become an indispensable tool for understanding complex systems. One global network property—the degree distribution—provides fundamental insights into processes as diverse as genome evolution and disease epidemics. In particular, a power law (“scale-free”) degree distribution often has distinct practical implications. We systematically fit various degree models to 55 biological networks of six different data types. Different models were compared using maximum likelihood fitting and the Bayesian Information Criterion (to account for model complexity). We considered four basic models from which the degree distribution can be drawn—power-law (scale-free), exponential, Poisson, and truncated power-law—and complex models that allow two underlying link types, e.g., true and false positive edges. We find that many networks in diverse fields previously reported to be scale-free or follow another simple model have more complex structural properties than has been previously understood. Previous conclusions based on simpler network models may need to be re-evaluated.

2 Results.

We find that the method used to bin the data and assumptions about variance can have a dramatic effect on maximum likelihood parameters for a model (Figure 1). The functional forms yielding the best (lowest) BIC often differed and had greater complexity than the best model reported in previous classifications of network topology. As expected, we find a power law fits many of the networks well, but a truncated power law yields a better fit for most networks previously classified as power law or exponential. Several combined functional forms frequently fit much better even than a truncated power law. The regulatory network of yeast was reported to have an exponential in-degree. We find it is better fit by a truncated power law (TPL), and an exponential convolved with a Poisson distribution (E*P) is an even better fit (Figure 2).

For protein interaction networks we find that a combination of two exponential distributions often fits quite well (Figure 3). This is a surprising result because a single exponential distribution does not fit these networks well, and suggests that networks containing multiple edge types with similar connectivity properties (same functional form with different parameters) may obscure the shape of the degree distribution to the point where a rigorous model selection approach is required to identify the correct functional form. The worm neuronal network (Figure 4) is dramatically better fit by the combination of a truncated power law and a Poisson distribution (TPL*P).

¹ Dept. of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 250 Longwood Avenue, Boston, MA 02115, USA. E-mail: debg@hms.harvard.edu

² Dept. of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 250 Longwood Avenue, Boston, MA 02115, USA. E-mail: giovanni.franklin@gmail.com

³ Dept. of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 250 Longwood Avenue, Boston, MA 02115, USA. E-mail: froth@hms.harvard.edu

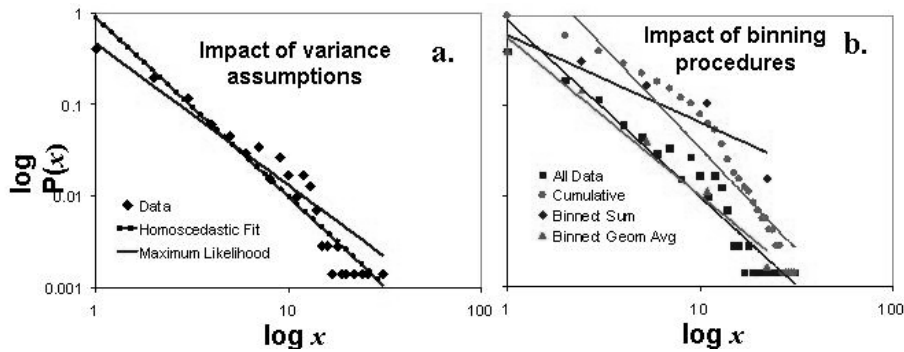


Figure 1: Effect of different methods to compute the best-fitting power law equation. a. Each point in the degree distribution given equal weight (homoscedastic fit), and our maximum likelihood method, in which each node is given equal weight. b. The effect of three binning procedures on the observed degree distributions and resulting least-squares regression power law curves. “All data”: observed degrees without binning. “Cumulative”: total frequency of all degrees \geq each degree. “Sum”: total frequency of degrees in each logarithmic bin; “Geom Avg”: geometric mean of degree frequencies in each logarithmic bin.

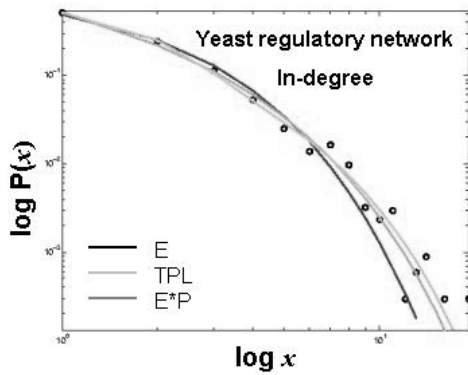


Figure 2: Yeast regulatory network.

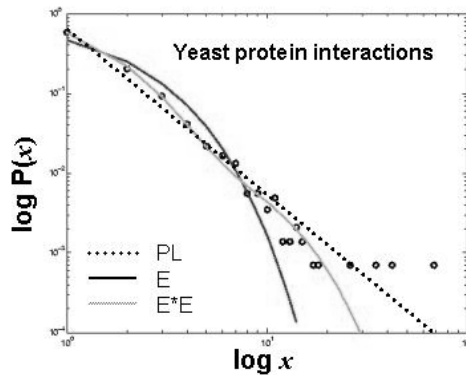


Figure 3: Yeast protein interaction network.

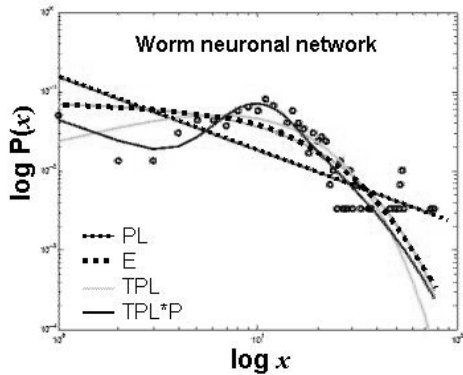


Figure 4: Worm neuronal network.