Protein Network Topology Metric Conservation: From Yeast to Human

Gil Alterovitz[1], Michael Xiang[2], Isaac S. Kohane[2,3], and Marco F. Ramoni[2,3]

Keywords: proteomics, protein-protein interactions, networks

1 Introduction

One of the most surprising results of the human genome project and the current sequencing efforts of other organisms is the remarkable similarity of the number of genes between species: the number of genes in *C. elegans* is of the same order of magnitude as that of *Homo sapiens*. It has been conjectured that these surprising results can be explained by a layer of protein-protein interactions, responsible for the expected difference in functional richness between worms and humans. For example, as the number of proteins **n** increases, the number of potential interactions increases proportional to $\mathbf{n^2}$. The quantity of interactions does appear to vary between organisms of different complexity. But, are there some constant parameters in these networks of interactions? As with the (Fibonacci) golden ratio ~1.618 (which appears in nature everywhere from flower pedals to shell patterns), perhaps there is a constant of nature with regard to biological network topology statistics. While it is not expected that such an absolutely invariant metric emerges, evolutionary pressures may have led to certain constraints on network parameters. By comparing yeast and human interaction data, we find that while the number of proteins and interactions are very different, topologic properties such as the average minimum distance between nodes, the power-law degree exponent, and the power-law proportionality constant appear to be relatively conserved.

2 Background

Barabasi [2] has shown that a number of metabolic and protein networks are scale-free. A scale free network is a graph where the probability distribution of the number of edges **k** can be described as: $\mathbf{P(k) = a k^{-\gamma}}$, where **a** is the proportionality constant and $\boldsymbol{\gamma}$ is the degree exponent. This construct results in a small number network hubs (nodes have many interactions) relative to the more common nodes that have few links. Other studies have shown that network hubs share certain properties such as evolutionary stability [5] and centrality/lethality [4]. All of these have been done with yeast (and other lower complexity organisms).

3 Methods and Results

In this work, network topology metrics of a human protein-protein interaction network are analyzed and compared with those of yeast. For the human protein interaction network, a superset of the previous human protein meta-interaction database work [1] is used- containing over 162,000 interactions and around 20,000 proteins. For yeast, the data/metrics are derived from databases/analysis done via TopNet as described by the Gerstein lab [6]. This includes 69,592 interactions and 4,957 proteins.

[1] Health Science and Technology/Electrical Engineering & Computer Science, Massachusetts Institute of Technology, Cambridge, 02139, USA, Email: ga@alum.mit.edu
[2] Children's Hospital Informatics Program, Harvard Medical School, Boston, 02115, USA.
[3] Harvard Partners Center for Genetics and Genomics, Boston, 02115, USA.

The following topology metrics were compared between yeast and human interaction networks (see Table 1): average degree (average number of interactions per protein), clustering coefficient (normalized number of interactions between neighbors of each protein), average shortest pairwise distance, subgraph diameter (longest path between any two nodes), power-law distribution exponent (γ as discussed in introduction), and power-law distribution proportionality constant (**a** as discussed in introduction). For the analysis of the topology of the protein-protein interaction network in humans, the Dijkstra algorithm was used to calculate the shortest distance between all nodes (for further investigation beyond the statistics presented here). This resulted in over 207 million inter-node distances that were calculated with Matlab via a Sun Grid-based cluster with 21 nodes, each with 2 gigabytes of main memory. The other metrics were all calculated using a single PC workstation.

The results of the human protein topology metric analysis (including previous work's average shortest pairwise distance calculation of 4.3 by Monte Carlo simulation [1]) can be compared with yeast results in Table 1. In short, certain broad topology metrics such as power-law degree exponent (1.4±-0.1), power-law degree exponent (0.4), and average shortest distance (4.1±-0.6) appear to be invariant across human and yeast. One of the first applications of scale-free networks, namely Internet routers, was found to have very different parameters with a power-law degree exponent of 2.4-2.5 and average shortest distance of 11-12 (even though it also had approximately 150,000 nodes, as in the human case) [3]. The invariance between certain yeast and human protein network metrics may be due to evolutionary pressure that keeps signaling cascades and other pathways optimized as specific scale-free network instantiations.

4 Figures and Tables

	Yeast	**Human**
Average Degree	28	15
Clustering Coefficient	0.19	0.11
Average Shortest Pairwise Distance	3.5	4.3, 4.7
Subgraph Diameter	10	15
Power-Law Distribution Exponent	1.3	1.5
Power-Law Distribution Proportionality Constant	0.42	0.41

Table 1: Comparing human and yeast protein-protein interaction topologic network metrics.

5 References

[1] G. Alterovitz, D. Patek, I. S. Kohane, and M. Ramoni, "Human Protein Meta-Interaction Database (HPMD) Potentiates Integration for Meta-Analysis," IEEE GENSIPS, 2005 (accepted).

[2] A. L. Barabasi and Z. N. Oltvai, "Network biology: understanding the cell's functional organization," *Nat Rev Genet*, vol. 5, pp. 101-13, 2004.

[3] R. Govindan and H. Tangmunarunkit, Proceedings of IEEE INFOCOM 2000, Tel Aviv, Israel, 2000.

[4] H. Jeong, S. P. Mason, A. L. Barabasi, and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, pp. 41-2, 2001.

[5] I. K. Jordan, L. Marino-Ramirez, Y. I. Wolf, and E. V. Koonin, "Conservation and coevolution in the scale-free human gene coexpression network," *Mol Biol Evol*, vol. 21, pp. 2058-70, 2004.

[6] H. Yu, X. Zhu, D. Greenbaum, J. Karro, et al., "TopNet: a tool for comparing biological sub-networks, correlating protein properties with topological statistics," *Nucleic Acids Res*, vol. 32, pp. 328-37, 2004.