Poster J-20

Does the Yeast PPI Network Look Like a Jellyfish?

Authors:
Shaohua Zhou (*Department of Developmental and Cell Biology, University of California Irvine*)
Arthur D. Lander (*Department of Developmental and Cell Biology, University of California, Irvine*)
Natasa Przulj (*Department of Computer Science, University of California Irvine*)

Short Abstract: It has been shown that the Autonomous System Internet topology "looks like a jellyfish" (Siganos et al., 2006). Since the purpose of both the Internet and protein-protein interaction (PPI) networks is to deliver information, we ask if the PPI networks are also structured like a jellyfish. We show that the PPI network of yeast Saccharomyces cerevisiae looks like a jellyfish.

Long Abstract:
It has recently been shown that the Autonomous System Internet topology "looks like a jellyfish" (Siganos et al., 2006). Since the purpose of both the Internet and protein-protein interaction (PPI) networks is to deliver information, it is possible that PPI networks are also structured like a jellyfish. The most comprehensive data on PPIs are available for Saccharomyces cerevisiae (baker's yeast), and thus we focus our analysis on PPI networks obtained from different studies of yeast.

The Jellyfish Model is a conceptual model of network structure including the core and hierarchically organized layers (Siganos et al., 2006). The core is built by starting with the highest degree node and adding nodes to it in the non-increasing order of their degree if they form a clique with the nodes already in the core; layers are defined by their proximity to the core. The Internet has been show to have approximately 80-90% of nodes in the first 3 layers, and to contain a total of six layers despite its huge size.

We analyze several yeast PPI data sets. We start with the von Mering et al. data set containing the total of 78,390 PPIs (edges) amongst 5,321 proteins (nodes) classified into three confidence levels: 2,455 high confidence PPIs amongst 988 proteins, 9,400 medium confidence PPIs amongst 2,298 proteins, and 66,535 low confidence PPIs amongst 5,058 proteins. We find that the high confidence PPI network has 7 jellyfish layers with 50.20% of the nodes in the first 4 layers and a large percentage of the nodes not participating in the jellyfish structure. The medium confidence network has 10 layers with 68.19% of the nodes in the first 4 layers and 8.92% of the nodes not participating the jellyfish structure. The low confidence PPI network has 6 layers with 97.88% of the nodes in the first 4 layers and 1.74% of nodes not participating in the jellyfish structure of this network. Finally, the entire von Mering et al. PPI network has 5 layers with 98.50% nodes in the first 4 layers and 1.24% of the nodes not participating in the jellyfish structure.

These results suggest that the larger the network, the fewer layers it has. It could be argued

that this is expected, since the density of the edges in these PPI networks increases from high towards lower confidence networks. However, we later demonstrate that this is not necessarily expected in general.

Next we analyze two yeast PPI dataset from the Database of Interacting Proteins (DIP) (Xenarios
et al., 2002) downloaded on April 2nd, 2006. The two yeast PPI data sets that we analyze are: the complete yeast PPI data set from DIP (henceforth denoted by "DIP-yeast-full") excluding self-interactions and containing 17,511 interactions amongst 4,959 proteins; the subset of DIP-yeast-full that contains validated interactions (henceforth denoted by "DIP-yeast-core")
containing 5,952 interactions amongst 2,554 proteins. DIP-yeast-core PPI network has 8 jellyfish layers with 80.54% of nodes in the first 4 layers and 8.22% of nodes not participating in the jellyfish structure. DIP-yeast-full PPI network has 6 layers with 97.88%
of nodes in the first 4 layers and 1.17% of nodes not participating in the jellyfish structure. As in the von Mering et al. PPI data set, the larger the PPI network, the fewer layers it has.

We also analyze the high confidence filtered yeast PPI set (Han et al., 2004) containing 2,493 PPIs amongst 1,379 proteins (henceforth denoted by "FYI"). To our surprise, it contains a much larger number of layers than the other yeast PPI networks: it has 20 layers with 6.45% of nodes in the first 4 layers and 56.42% of the nodes not participating in the jellyfish structure. Thus, FYI yeast PPI data set seems to have a different network topology than the other yeast
PPI networks that we analyzed. We are doing further investigations to determine why this is the case.

For each of the above 8 PPI networks we analyze 10 instances of model networks from each of the
following 4 different types of random networks (the total of 8x10x4=320 model networks): Erdos-Renyi random networks (ER), Erdos-Renyi random networks with the same degree distributions as the PPI networks (ER-DD), Barabasi-Albert scale-free networks (SF), and
3-dimensional Euclidean space and distance geometric random graphs (GEO-3D). These model networks have the same number of nodes and edges as the PPI networks. All ER, ER-DD, and SF model networks have less than 9 of layers, while the GEO-3D networks have at least 22 layers. Model networks corresponding to the DIP-yeast-full PPI network have fewer layers than those
corresponding to the DIP-yeast-core PPI network. Similar holds for the model networks corresponding to the von Mering et al. PPI networks. This is in agreement with our previous observations that the denser the network, the fewer layers it has.

Geometric random networks are different from other model networks with respect to their "jellyfish structure" in the sense that they have significantly larger number of layers than do ER, ER-DD, or SF model networks of the same size and edge density (note that this could be expected from their larger network diameters). Von Mering et al. (2002) and Han et al. (2055) higher confidence PPI networks have more layers than other, lower-confidence PPI networks. Similarly, geometric random graphs have more layers compared to ER, ER-DD, or SF networks. This confirms the previous relative graphlet frequency distribution-based fit of geometric graphs to these data (Przulj et al., 2004). The differences in PPI data structure

could be due to the bias introduced by filtering the data sets, or the noise present in low-confidence data sets. Thus, the "jelly-fish algorithm" could possibly be used to determine the quality of the data sets.

References:

Han et al., "Evidence for dynamically organized modularity in the yeast protein-protein interaction network ", Nature 430, pp 88-93, 2004.

von Mering et al., "Comparative assessment of large-scale data sets of protein-protein interactions", Nature 417:6887, pp 399-403, 2002.

Przulj, Corneil, and Jurisica, "Modeling Interactome: Scale-Free or Geometric?", Bioinformatics, 20: 18, pp 3508-3515, 2004.

Siganos, Tauro, and Faloutsos, "Jellyfish: A Conceptual Model for the AS Internet Topology',

Journal of Communications and Networks, to appear, 2006.

Xenarios et al., "DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions", Nucleic Acids Research, 30:1, pp 303-305, 2002.