BIOINFORMATICS

Vol. 20 no. 18 2004, pages 3508–3515
doi:10.1093/bioinformatics/bth436

Modeling interactome: scale-free or geometric?

N. Pržulj[1], D. G. Corneil[1] and I. Jurisica[1,2,*]

[1]Department of Computer Science, University of Toronto, Toronto, M5S 3G4, Canada and [2]Ontario Cancer Institute, Division of Cancer Informatics, Princess Margaret Hospital, University Health Network, 610 University Avenue, Toronto, M5G 2M9, Canada

Received on June 2, 2004; revised on July 19, 2004; accepted on July 24, 2004
Advance Access publication July 29, 2004

ABSTRACT

Motivation: Networks have been used to model many real-world phenomena to better understand the phenomena and to guide experiments in order to predict their behavior. Since incorrect models lead to incorrect predictions, it is vital to have as accurate a model as possible. As a result, new techniques and models for analyzing and modeling real-world networks have recently been introduced.

Results: One example of large and complex networks involves protein–protein interaction (PPI) networks. We analyze PPI networks of yeast *Saccharomyces cerevisiae* and fruitfly *Drosophila melanogaster* using a newly introduced measure of local network structure as well as the standardly used measures of global network structure. We examine the fit of four different network models, including Erdös-Rényi, scale-free and geometric random network models, to these PPI networks with respect to the measures of local and global network structure. We demonstrate that the currently accepted scale-free model of PPI networks fails to fit the data in several respects and show that a random geometric model provides a much more accurate model of the PPI data. We hypothesize that only the noise in these networks is scale-free.

Conclusions: We systematically evaluate how well-different network models fit the PPI networks. We show that the structure of PPI networks is better modeled by a geometric random graph than by a scale-free model.

Contact: juris@cs.utoronto.ca

Supplementary information: Supplementary information is available at http://www.cs.utoronto.ca/∼juris/data/ppiGRG04/

1 INTRODUCTION

Many real-world phenomena have been modeled by large networks including the World Wide Web, electronic circuits, collaborations between scientists, metabolic pathways and protein–protein interactions (PPIs). A common property of these phenomena is that they all consist of components (modeled by network nodes) and pairwise interactions between the components (modeled by links between the nodes, i.e. by network edges). Studying statistical and theoretical properties of large networks (also called graphs) has gained considerable attention in the past few years. Various network models have been proposed to describe properties of large real-world networks, starting with the earliest models of Erdös–Rényi random graphs (Erdös and Rényi, 1959, 1960, 1961) and including more recent small-world (Watts and Strogatz, 1998), scale-free (Barabási and Albert, 1999) and hierarchical (Ravasz *et al*., 2002) models. Excellent review papers have recently appeared describing this emerging, large research area (Newman, 2003; Barabási and Oltvai, 2004; Albert and Barabási, 2002; Strogatz, 2001).

This paper uses a method for detecting local structural properties of large networks and proposes a new model of PPI networks. Our new measure of local network structure consists of 29 network measurements. Using this new measure of network structure, we find that the PPI networks of *Saccharomyces cerevisiae* and *Drosophila melanogaster* are more accurately modeled by geometric random graphs (defined below) than by the scale-free model. The extent of this improvement is such that even perturbing the network by random additions, deletions and rewiring of 30% of the edges introduces much smaller error when compared to the error from modeling the network by scale-free, or other currently available network models (details are provided below). In addition, we show that three out of four standard network parameters measuring a global network structure also show an improved fit between the experimentally determined PPI networks and the geometric random graph model than between the PPI networks and the scale-free model.

2 SYSTEMS AND METHODS

2.1 Definitions

To our knowledge, this study is the first one to use geometric random graphs to model PPI networks. Thus, we give a brief description of geometric random graphs. The descriptions of more popular Erdös–Rényi and scale-free network models are presented in the Supplementary information.

*To whom correspondence should be addressed.

2.1.1 Geometric random graphs A geometric graph $G(V,r)$ with radius r is a graph with node set V of points in a metric space and edge set $E = \{\{u,v\}|(u,v \in V) \wedge (0 < \|u-v\| \leq r)\}$, where $\|\cdot\|$ is an arbitrary distance norm in this space. That is, points in a metric space correspond to nodes, and two nodes are adjacent if the distance between them is at most r. Often, two-dimensional space is considered, containing points in the unit square $[0,1]^2$ or unit disc, and $0 < r < 1$ (Diaz *et al*., 1999, 2000), with the distance norms being l_1 (Manhattan distance), l_2 (Euclidean distance) or l_∞ (Chessboard distance). The distance between two points (x_1,y_1) and (x_2,y_2) is $|x_1-x_2|+|y_1-y_2|$ in the l_1 norm, $\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$ in the l_2 norm, and $\max(|x_1-x_2|,|y_1-y_2|)$ in the l_∞ norm. A random geometric graph $G(n,r)$ is a geometric graph with n nodes which correspond to n independently and uniformly randomly distributed points in a metric space. Many properties of these graphs have been explored when $n \to \infty$ (Penrose, 2003). Similar to Erdös–Rényi random graphs, certain properties of these graphs also appear suddenly when a specific threshold is reached.

2.1.2 Global network properties The most commonly studied statistical properties to measure the global structure of large networks are the degree distribution, network diameter and clustering coefficients, defined as follows. The degree of a node is the number of edges (connections) incident to the node. The degree distribution, $P(k)$, describes the probability that a node has degree k. This network property has been used to distinguish amongst different network models; in particular, Erdös–Rényi random networks have a Poisson degree distribution, while scale-free networks have a power-law degree distribution $P(k) \sim k^{-\gamma}$, where γ is a positive number. The smallest number of links that have to be traversed to get from node x to node y in a network is called the distance between nodes x and y and a path through the network that achieves this distance is called shortest path between x and y. The average of shortest path lengths over all pairs of nodes in a network is called the network diameter. [Note that in classical graph theory, the diameter is the maximum of shortest path lengths over all pairs of nodes in the network (West, 2001).] This network property also distinguishes different network models, e.g. the diameter of Erdös–Rényi random networks on n nodes is proportional to $\log n$, the network property often referred to as the small-world property; the diameters of scale-free random networks with degree exponent $2 < \gamma < 3$, which have been observed for most real-world networks, are ultra-small (Chung and Lu, 2002; Cohen and Havlin, 2003), i.e. proportional to $\log\log n$. The clustering coefficient of node v in a network is defined as $C_v = 2e_1/[n_1(n_1-1)]$, where v is linked to n_1 neighboring nodes and e_1 is the number of edges amongst the n_1 neighbors of v. The average of C_v over all nodes v of a network is the clustering coefficient C of the whole network and it measures the tendency of the network to form highly interconnected regions called clusters. The average clustering coefficient of all nodes of degree k in a network, $C(k)$, has been shown to follow $C(k) \sim k^{-1}$ for many real-world networks indicating a network's hierarchical structure (Ravasz and Barabási, 2003; Ravasz *et al*., 2002). Many real-world networks have been shown to have high clustering coefficients and to exhibit small-world and scale-free properties.

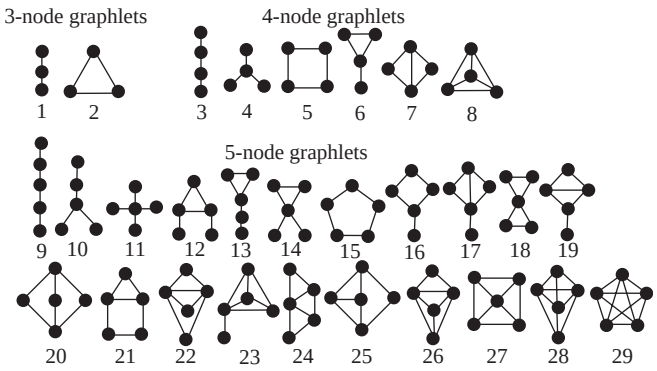

Fig. 1. All 3-node, 4-node and 5-node connected networks (graphlets), ordered within groups from the least to the most dense with respect to the number of edges when compared to the maximum possible number of edges in the graphlet; they are numbered from 1 to 29.

2.1.3 Local network properties In addition to the above global properties of network structure, a new bottom-up approach focusing on finding small, over-represented patterns in a network has recently been introduced (Milo *et al*., 2002, 2004; Shen-Orr *et al*., 2002; Itzkovitz *et al*., 2003). In this approach, all small subgraphs (subnetworks whose nodes and edges belong to the large network) of a large network are identified and the ones that appear in the network significantly more frequently than in the randomized network are called network motifs. Different types of real-world networks have been shown to have different motifs (Milo *et al*., 2002). The *S.cerevisiae* PPI network constructed on combined, mostly two-hybrid analysis data (Uetz *et al*., 2000; Xenarios *et al*., 2000), has been shown to have two network motifs (Milo *et al*., 2002), those corresponding to graphs 2 and 4 presented in Figure 1. Furthermore, different real-world evolved and designed networks have been grouped into superfamilies according to their local structural properties (Milo *et al*., 2004). In addition, the shortest path distribution and the frequencies of 3–15-node cycles in the high-confidence fruitfly PPI network have been shown to differ from those of randomly rewired networks which preserve the same degree distribution as the original PPI network (Giot *et al*., 2003).

2.2 Graphlet analysis of PPI networks

Our approach to analyzing large networks belongs to the bottom-up type. Similar to the approach of Milo *et al*. (2004), we identify all 3–5-node subgraphs of PPI networks for

S.cerevisiae and *D.melanogaster*. We compare the frequencies of the appearance of these subgraphs in PPI networks with the frequencies of their appearance in four different types of random networks: (1) Erdös–Rényi random networks with the same number of nodes and edges as the corresponding PPI networks (ER); (2) Erdös–Rényi random networks with the same number of nodes, edges and the same degree distribution as corresponding PPI networks (ER-DD); (3) scale-free random networks with the same number of nodes and the number of edges within 1% of those of the corresponding PPI networks (SF); and (4) several types of geometric random graphs with the number of nodes and the number of edges within 1% of those of the corresponding PPI networks (GEO) (see Supplementary information). We used three different geometric random graph models, defining points uniformly at random in two-dimensional Euclidean space (GEO-2D), three-dimensional Euclidean space (GEO-3D) and four-dimensional Euclidean space (GEO-4D); the Euclidean distance measure between the points was used to determine if two points are close enough to be linked by an edge in the corresponding geometric random graph (see Supplementary information).

The number of different connected networks on n nodes increases exponentially with n. For $n = 3, 4$ and 5, there are 2, 6 and 21 different connected networks on n nodes, respectively. To avoid terminology confusing network motifs with network subgraphs (motifs are special types of subgraphs), we use the term graphlet to denote a connected network with a small number of nodes. All 3–5-node graphlets are presented in Figure 1. [Note that in their analysis of undirected networks, Milo *et al*. (2004) examined the presence of the eight graphlets of size 3 or 4.] We use the graphlet frequency, i.e. the number of occurrences of a graphlet in a network, as a new network parameter and show that PPI networks are closest to geometric random graphs with respect to this new network parameter (details are given below). In addition, despite the difference in degree distributions of PPI networks and geometric random graphs and the similarity between degree distributions of PPI networks and scale-free networks, we show that the diameter and clustering coefficient parameters also indicate that PPI networks are closer to the geometric random graph model than to the ER, ER-DD and SF models. We hypothesize that the discrepancy between the degree distributions of PPI and GEO networks is caused by a high percentage of false negatives in the PPI networks and that when PPI datasets become denser and more complete, the degree distributions of PPI networks will be closer to Poisson distributions, characteristic of geometric random graphs.

We analyzed graphlet frequencies of four PPI networks: the high-confidence yeast *S.cerevisiae* PPI network involving 2455 interactions amongst 988 proteins (von Mering *et al*., 2002); the yeast *S.cerevisiae* PPI network involving 11 000 interactions amongst 2401 proteins (von Mering *et al*., 2002) [these are the top 11 000 interactions in von Mering *et al*. (2002) classification]; the high-confidence fruitfly *D.melanogaster* PPI network involving 4637 interactions amongst 4602 proteins (Giot *et al*., 2003); and the entire fruitfly *D.melanogaster* PPI network as published in Giot *et al*. (2003) involving 20 007 interactions amongst 6985 proteins, which includes low confidence interactions. We computed graphlet frequencies in the PPI and the corresponding random networks of the previously described four different types.

Graphlet counts quantify the local structural properties of a network. Currently, our knowledge of the connections in PPI networks is incomplete (i.e. we do not know all the edges, and for many organisms, we do not even know all the nodes). The edges we do know are dominated by experiments focused around proteins that are currently considered 'important'. However, we hypothesize that the local structural properties of the full PPI network, once all connections are made, are similar to the local structural properties of the currently known, highly studied parts of the network. Thus, we would expect that the relative frequency of graphlets among the currently known connections is similar to the relative frequency of graphlets in the full PPI network, which is as yet unknown. Thus, we use the relative frequency of graphlets $N_i(G)/T(G)$ to characterize PPI networks and the networks we use to model them, where $N_i(G)$ is the number of graphlets of type i ($i \in \{1, \ldots, 29\}$) in a network G, and $T(G) = \sum_{i=1}^{29} N_i(G)$ is the total number of graphlets of G. In this model, then, the 'similarity' between two graphs should be independent of the total number of nodes or edges, and should depend only upon the differences between relative frequencies of graphlets. Thus, we define the relative graphlet frequency distance $D(G, H)$, or distance for brevity, between two graphs G and H as:

$$D(G, H) = \sum_{i=1}^{29} |F_i(G) - F_i(H)|,$$

where $F_i(G) = -\log(N_i(G)/T(G))$. We use the logarithm of the graphlet frequency because frequencies of different graphlets can differ by several orders of magnitude and we do not want the distance measure to be entirely dominated by the most frequent graphlets.

3 RESULTS AND DISCUSSION

3.1 Graphlet frequency counts

Using this method, we computed the distances between several real-world PPI networks and the corresponding ER, ER-DD, SF and GEO random networks. We found that the GEO random networks fit the data an order of magnitude better in the higher-confidence PPI networks, and less so (but still better) in the more noisy PPI networks (see Supplementary Table 3 of the Supplementary information). The only exception is the larger fruitfly PPI network, with ∼77% of its edges corresponding to lower confidence interactions (Giot

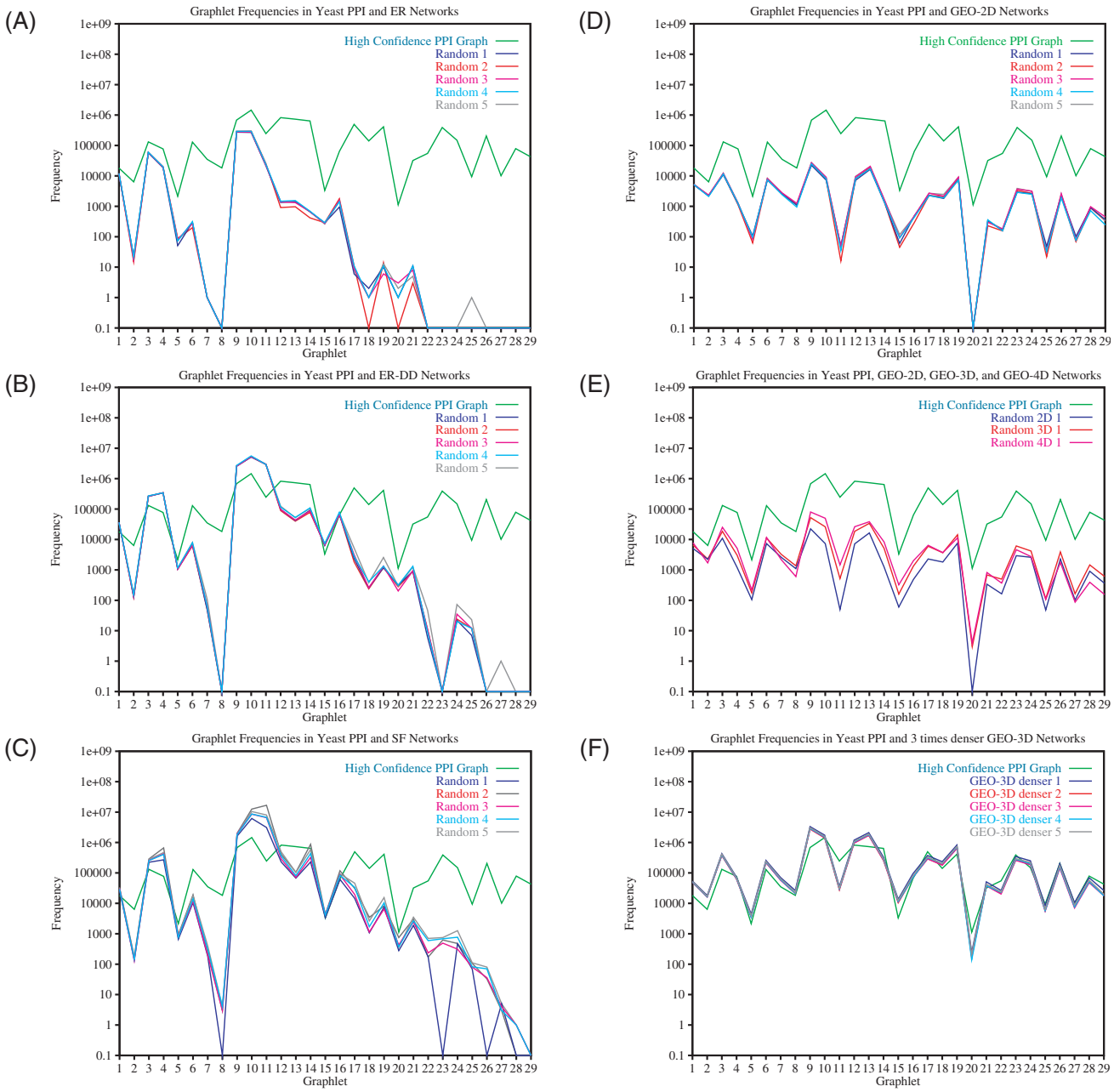

Fig. 2. Comparison of graphlet frequencies in the high-confidence *S.cerevisiae* PPI network (von Mering *et al.*, 2002) (green line) with corresponding ER, ER-DD, SF and GEO random networks. Zero frequencies were replaced by 0.1 for plotting on log-scale. (**A**) PPI network versus five corresponding ER random networks. (**B**) PPI network versus five corresponding ER-DD random networks. (**C**) PPI network versus five corresponding SF random networks. (**D**) PPI network versus five corresponding GEO-2D random networks. (**E**) PPI network versus a corresponding GEO-2D, GEO-3D and GEO-4D random network. (**F**) PPI network versus five GEO-3D random networks with the same number of nodes and approximately three times as many edges as the PPI network.

et al., 2003); this PPI network is ∼2.7 times closer to the scale-free than to the geometric network model with respect to this parameter (see Supplementary information). We hypothesize that this behavior of the graphlet frequency parameter is the consequence of a large amount of noise present in this fruit-fly PPI network; our analysis of the diameters and clustering coefficients of these networks further support this hypothesis (see below).

An illustration showing graphlet frequencies in the high-confidence yeast PPI network and the corresponding random model networks is presented in Figure 2. As mentioned above, the current yeast high-confidence PPI network is missing

many edges, so we expect that the complete PPI network would be much denser. Also, we believe that the maximum degree of this PPI network is not likely to change significantly due to the extent of research having been done on the highly connected regions of the network. Thus, we constructed two sets of three-dimensional geometric random networks with the same number of nodes, but about three and six times as many edges as the PPI network, respectively. By making the GEO-3D networks corresponding to this PPI network about six times as dense as the PPI network, we matched the maximum degree of the PPI network to those of these geometric random networks. The resulting geometric random network models provide the closest fit with respect to the graphlet frequency parameter to the PPI network (see Fig. 2 F and Supplementary information).

3.2 Robustness of Graphlet frequency counts

When studying PPI networks, it should be noted that all of the current publicly available PPI datasets contain a percentage of false positives and are also largely incomplete, i.e. the number of false negatives is arguably much larger than the number of false positives. Since the genomes of many species have already been sequenced, it is expected that the predicted number of proteins in PPI datasets will not change significantly, but the number of known interactions will grow dramatically.

Since PPI networks contain noise, we examined the robustness of the graphlet frequency parameter by adding noise to the yeast high-confidence PPI network and comparing graphlet frequencies of the perturbed networks and the PPI network. In particular, we perturbed this PPI network by randomly adding, deleting and rewiring 10, 20 and 30% of its edges. We computed distances between the perturbed networks and the PPI network by using the distance function defined above. We found the exceptional robustness of the graphlet frequency parameter to random additions of edges encouraging, especially since the currently available PPI networks are missing many edges. In particular, additions of 30% of edges resulted in networks which were $\sim$21 times closer to the PPI network than the corresponding SF random networks. We also found that graphlet frequencies were fairly robust to random edge deletions and rewirings (deletions and rewirings of 30% of edges resulted in networks which were $\sim$6 times closer to the PPI network than the corresponding SF random networks), which further increases our confidence in PPI networks having geometric properties despite the presence of false positives in the data (see Supplementary information).

3.3 Global network properties of PPI and model networks

Recently, there has been a lot of interest in the global properties of PPI networks. PPI networks for the yeast *S.cerevisiae* resulting from different high-throughput studies (Uetz *et al*., 2000; Xenarios *et al*., 2000; Ito *et al*., 2001) have been shown to have scale-free degree distributions (Jeong *et al*., 2001; Maslov and Sneppen, 2002). They have hierarchical structures with $C(k)$ scaling as k^{-1} (Barabási *et al.*, 2003). The degree distributions of this yeast PPI network, as well as the PPI network of the bacterium *Helicobacter pylori*, have been shown to decay according to a power law (Jeong *et al*., 2001; Rain *et al*., 2001). However, the high confidence *D.melanogaster* PPI network and a larger *D.melanogaster* PPI network have been shown to decay close to, but faster than a power law (Giot *et al*., 2003).

We compared the commonly studied statistical properties of large networks, namely the degree distribution, network diameter and clustering coefficients C and $C(k)$, of the PPI and various model networks. Despite the degree distributions of the PPI networks being closest to the degree distributions of the corresponding scale-free random networks (Supplementary Figures 9 and 10), the remaining three parameters of the two yeast PPI networks differ from the scale-free model with most of them being closest to the corresponding geometric random networks (Supplementary Tables 4 and 5, and Supplementary Figures 11 and 12). An illustration of the behavior of $C(k)$ in the yeast high-confidence PPI network and the corresponding model networks is presented in Figure 3. Also, many of these properties of the two fruitfly PPI networks were close to ER, ER-DD and SF models possibly indicating the presence of noise in these PPI networks (Supplementary Tables 4 and 5, and Supplementary Figures 13 and 14). Nevertheless, the high-confidence fruitfly PPI network exhibits some geometric network properties; e.g. the clustering coefficient of this PPI network is only an order of magnitude smaller than the clustering coefficients of the corresponding geometric random networks, but it is at least four orders of magnitude larger than the clustering coefficients of the corresponding scale-free networks (Supplementary Table 5). We expect that ongoing improvements in the fruitfly PPI dataset will make the structural properties of its PPI network closer to those of the geometric random graphs.

4 CONCLUSIONS

Despite recent significant advances in understanding large real-world networks, this area of research is still in its infancy (Barabási and Oltvai, 2004; Newman, 2003). Novel techniques for analyzing, characterizing and modeling structures of these networks need to be developed. As new data becomes available, we must ensure that the theoretical models continue to accurately represent the data. The scale-free model has been assumed to provide such a model for PPI networks (Jeong *et al*., 2001; Rain *et al*., 2001; Maslov and Sneppen, 2002). The current scale-free model of human PPI network has been used for planning experiments in order to optimize time and cost required for their completion (Lappe and Holm, 2004). In particular, the model was used to form the basis of an algorithmic strategy for guiding experiments which would detect up to 90% of the human interactome with less

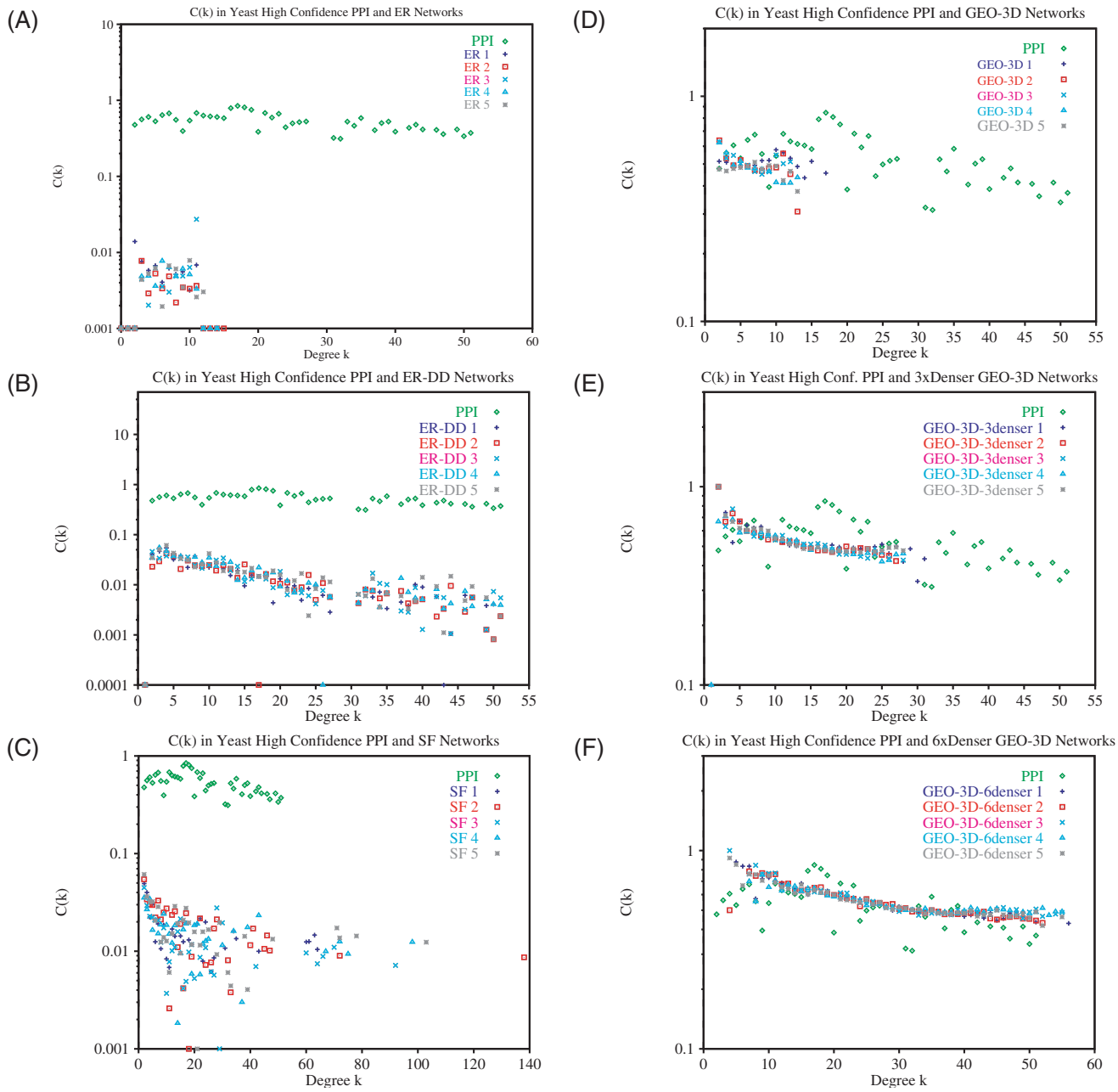

Fig. 3. Comparison of average clustering coefficients $C(k)$ of degree k nodes in the high-confidence *S.cerevisiae* PPI network (von Mering *et al.*, 2002) (green dots) with corresponding ER, ER-DD, SF and GEO random networks. Since we use a log-scale, zero values were placed on the abscissa. (**A**) PPI network versus five corresponding ER random networks. (**B**) PPI network versus five corresponding ER-DD random networks. (**C**) PPI network versus five corresponding SF random networks. (**D**) PPI network versus five corresponding GEO-3D random networks. (**E**) PPI network versus five corresponding GEO-3D random networks with the same number of nodes and approximately three times as many edges as the PPI network. (**F**) PPI network versus five corresponding GEO-3D random networks with the same number of nodes and approximately six times as many edges as the PPI network.

than one-third of the proteome used as bait in high-throughput pull-down experiments (Lappe and Holm, 2004). However, if an incorrect model is used to plan experiments then clearly the experiments will be at best inefficient at gaining the desired information. At worst, they could even be misleading by failing to direct experimenters to find actual PPIs that exist but will remain hidden because the experiments will be looking in the wrong place. Therefore, having an improved model for the PPI networks is crucial for effective experimental planning.

We have shown compelling evidence that the structure of yeast PPI networks is closer to the geometric random graph model than to the currently accepted scale-free model. For yeast PPI networks, three out of four of the commonly studied statistical properties of global network structure, as well as the newly introduced graphlet frequency parameter describing local structural properties of large networks, were closer to geometric random graphs than to scale-free or Erdös–Rényi random graphs. In addition, despite the noise present in their PPI detection techniques and the lack of independent verification of its PPIs by various laboratories, fruitfly PPI networks do show properties of geometric random graphs. Other designed and optimized communication networks, such as wireless multihop networks (Bettstetter, 2002), electrical power-grid and protein structure networks (Milo *et al*., 2004), have been modeled by geometric random graphs as well. Thus, it is plausible that PPI networks, which possibly emerged, similar to the World Wide Web, through stochastic growth processes, but unlike the World Wide Web have gone through millions of years of evolutionary optimization, are better modeled by the geometric random graph model than by the scale-free model (the scale-free model seems to be appropriate for networks that have emerged through stochastic growth processes and have not been optimized, such as the World Wide Web). Also, similar to the limited coverage that wireless networks have, currently available PPI data cover only a portion of the interactome. Once a more complete interactome data becomes available, we will be able to validate the correctness of the current model and possibly design better models for PPI networks.

ACKNOWLEDGEMENTS

We thank Rudi Mathon, Gil Prive, Wayne Hayes, Jeff Wrana and Isidore Rigoutsos for helpful comments and discussions, and Andrew King for implementing some of the random graph generators. We thank the anonymous referee for useful comments and suggestions which helped us improve the paper. Financial support from the Natural Sciences and Engineering Research Council of Canada, the Ontario Graduate Scholarship Program and IBM Canada was gratefully received.

REFERENCES

Albert,R. and Barabási,A.-L. (2002) Statistical mechanics of complex networks. *Rev. Modern Phys.*, **74**, 47–97.

Barabási,A.L. and Albert,R. (1999) Emergence of scaling in random networks. *Science*, **286**, 509–512.

Barabási,A.-L. and Oltvai,Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev.*, **5**, 101–113.

Barabási,A.-L., Dezso,Z., Ravasz,E., Yook,S.H. and Oltvai,Z.N. (2003) Scale-free and hierarchical structures in complex networks, AIP Conf. Proc. No. 661, Modeling of Complex Systems: Seventh Granada Lectures Granada (Spain), 2–7 September 2002, Springer-Verlag, Heidelberg, pp.1–16.

Bettstetter,C. (2002) On the minimum node degree and connectivity of a wireless multihop network. *Proceedings of the 3rd ACM international symposium on mobile ad hoc networking and computing*, ACM Press, NY, pp. 80–91.

Chung,F. and Lu,L. (2002) The average distances in random graphs with given expected degrees. *Proc. Natl Acad. Sci., USA*, **99**, 15879–15882.

Cohen,R. and Havlin,S. (2003) Scale-free networks are ultra small. *Phys. Rev. Lett.*, **90**, 058701.

Diaz,J., Penrose,M.D., Petit,J. and Serna (2001) Linear orderings of random geometric graphs. *J. Algorithm.*, **38**, 78–116.

Diaz,J., Penrose,M.D., Petit,J. and Serna,M. (2000) Convergence theorems for some layout measures on random lattice and random geometric graphs. *Combinatorics, Prob. Comput.*, **10**, 489–511.

Erdös,P. and Rényi,A. (1959) On random graphs. *Publ. Math.*, **6**, 290–297.

Erdös,P. and Rényi,A. (1960) On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, **5**, 17–61.

Erdös,P. and Rényi,A. (1961) On the strength of connectedness of a random graph. *Acta Math. Sci. Hung.*, **12**, 261–267.

Giot,L., Bader,J., Brouwer,C., Chaudhuri,A., Kuang,B., Li,Y., Hao,Y., Ooi,C., Godwin,B., Vitols,E. *et al*. (2003) A protein interaction map of *Drosophila melanogaster. Science*, **302**, 1727–1736.

Ito,T., Chiba,T., Ozawa,R., Yoshida,M., Hattori,M. and Sakaki,Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci., USA*, **98**, 4569–4574.

Itzkovitz,S., Milo,R., Kashtan,N., Ziv,G. and Alon,U. (2003) Subgraphs in random networks. *Phys. Rev. E*, **68**, 026127.

Jeong,H., Mason,S.P., Barabási,A.L. and Oltvai,Z.N. (2001) Lethality and centrality in protein networks. *Nature*, **411**, 41–42.

Lappe,M. and Holm,L. (2004) Unraveling protein interaction networks with near-optimal efficiency. *Nat. Biotechnol.*, **22** (1), 98–103.

Maslov,S. and Sneppen,K. (2002) Specificity and stability in topology of protein networks. *Science*, **296**, 910–913.

Milo,R., Shen-Orr,S.S., Itzkovitz,S., Kashtan,N., Chklovskii,D. and Alon,U. (2002) Network motifs: simple building blocks of complex networks. *Science*, **298**, 824–827.

Milo,R., Itzkovitz,S., Kashtan,N., Levitt,R., Shen-Orr,S., Ayzenshtat,I., Sheffer,M. and Alon,U. (2004) Superfamilies of evolved and designed networks. *Science*, **303**, 1538–1542.

Newman,M.E.J. (2003) The structure and function of complex networks. *SIAM Rev.*, **45**, 167–256.

Penrose,M. (2003) *Geometric Random Graphs*. Oxford Univeristy Press, Oxford.

Rain,J.-D., Selig,L., DeReuse,H., Battaglia,V., Reverdy,C., Simon,S., Lenzen,G., Petel,F., Wojcik,J., Schachter,V., Chemama,Y., Labigne,A. and Legrain,P. (2001) The protein–protein interaction map of *Helicobacter pylori. Nature*, **409**, 211–215.

Ravasz,E. and Barabási,A.-L. (2003) Hierarchical organization in complex networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **67**, 026112.

Ravasz,E., Somera,A.L., Mongru,D.A., Oltvai,Z.N. and Barabási,A.-L. (2002) Hierarchical organization of modularity in metabolic networks. *Science*, **297**, 1551–1555.

Shen-Orr,S.S., Milo,R., Mangan,S. and Alon,U. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.*, **31**, 64–68.

Strogatz,S.H. (2001) Exploring complex networks. *Nature*, **410**, 268–276.

Uetz,P., Giot,L., Cagney,G., Mansfield,T.A., Judson,R.S., Knight,J.R., Lockshon,D., Narayan,V., Srinivasan,M., Pochart,P. *et al*. (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature*, **403**, 623–627.

von Mering,C., Krause,R., Snel,B., Cornell,M., Oliver,S.G., Fields,S. and Bork,P. (2002) Comparative assessment of large-scale datasets of protein–protein interactions. *Nature*, **417**, 399–403.

Watts,D.J. and Strogatz,S.H. (1998) Collective dynamics of 'small-world' networks. *Nature*, **393**, 440–442.

West,D.B. (2001) *Introduction to Graph Theory*, 2nd edn. Prentice-Hall, Upper Saddle River, NJ.

Xenarios,I., Salwinski,L., Duan,X.J., Higney,P., Kim,S.M. and Eisenberg,D. (2000) Dip: the database of interacting proteins. *Nucleic Acids Res.*, **28**, 289–291.