

Supplementary Information: Modeling Interactome

Nataša Pržulj,¹ Derek G. Corneil,¹ Igor Jurisica^{2,1,*}

¹Department of Computer Science, University of Toronto,
10 King's College Road, Toronto, ON, M5S 3G4, Canada

²Ontario Cancer Institute, University Health Network, Division of Cancer Informatics
610 University Avenue, Toronto, ON, M5G 2M9, Canada

*To whom correspondence should be addressed; E-mail: juris@cs.utoronto.ca.

Contents

1	Supplementary Methods and Discussion	1
1.1	Models of Large Networks	1
1.2	Graphlet Analysis	2
1.2.1	Construction of Model Networks	3
1.2.2	Graphlet Frequency Results	4
1.3	Standard Global Network Parameters	5
1.3.1	Definitions	5
1.3.2	Results on PPI and Model Networks	6
2	Supplementary Figures	8
3	Supplementary Tables	24

1 Supplementary Methods and Discussion

1.1 Models of Large Networks

The earliest model of large networks was introduced by Erdős and Rényi in late 50s and early 60s (Erdős & Rényi, 1959; Erdős & Rényi, 1960; Erdős & Rényi, 1961). They introduced the *random graph* model and initiated a large area of research, a good survey of which can be found in Bollobas's book (Bollobas, 1985). There are several versions of this model out of which the most commonly studied is the one based on the principle that in a graph on

n nodes, every edges is present with probability p and absent with probability $1 - p$; this graph model is commonly denoted by $G_{n,p}$. Another commonly used random graph model, denoted by $G_{n,m}$, is based on the principle that a graph on n nodes and m edges is chosen uniformly at random amongst all graphs on n nodes and m edges, or equivalently, a set of m edges of the graph is chosen uniformly at random among all possible $\binom{n}{m}$ sets of edges. These two models behave similarly when $p \approx \frac{m}{\binom{n}{2}}$. We used the $G_{n,m}$ model to construct ER graphs corresponding to the PPI networks (details are below). Many of the properties of random graphs can be calculated exactly in the limit of large n (Bollobas, 1985).

Since properties of the random graph model deviate from those of most real-world networks, several new network models have recently been introduced. To capture the scale-free character of real-world networks, the random graph model was modified to allow for arbitrary degree distributions while keeping all other aspects of the random graph model. This model is called the *generalized random graph* model and finding properties of these graphs has been an active area of research (Bender & Canfield, 1978; Newman, 2002; Luczak, 1990; Molloy & Reed, 1995; Molloy & Reed, 1998; Aiello *et al.*, 2001; Newman *et al.*, 2001; Wilf, 1990). The ER-DD random networks which we constructed to have the same number of nodes, edges, and the degree distributions as the PPI networks belong to this network model. We surveyed some of the main results on these graphs in (Przulj, 2004).

The *scale-free* network model, characterized by a small number of very highly connected nodes, has received a lot of attention in recent years. Barabási, Albert, and Jeong (Barabási & Albert, 1999; Barabási *et al.*, 1999) showed that a heavy-tailed degree distribution emerges automatically from a stochastic growth model in which new nodes are added continuously and they preferentially attach to existing nodes with probability proportional to the degree of the target node. That is, high-degree nodes become of even higher degree with time and the resulting degree distribution is $P(k) \approx k^{-\gamma}$. We constructed the SF networks corresponding to PPI networks using this network model (more details are given below). Many real-world networks have power-law degree distributions, such as for example, the Internet backbone (Faloutsos *et al.*, 1999), metabolic reaction networks (Jeong *et al.*, 2000), the telephone call graph (Abello *et al.*, 1998), and the World Wide Web (Broder *et al.*, 2000); degrees of all of these networks decay as a power law $P(k) \approx k^{-\gamma}$, with the exponent $\gamma \approx 2.1 - 2.4$. Thus, a large body of research on theoretical and experimental results on this network model has appeared recently (Bollobas & Riordan, 2001; Aiello *et al.*, 2001; Newman *et al.*, 2001; Krapivsky & Redner, 2001; Albert & Barabási, 2002; Albert & Barabási, 2000; Dorogovtsev & Mendes, 2000; Krapivsky *et al.*, 2000; Albert *et al.*, 2000; Broder *et al.*, 2000; Albert *et al.*, 2000; Cohen *et al.*, 2000; Callaway *et al.*, 2000; Bornholdt & Ebel, 2001; Jeong *et al.*, 2000; Wagner & Fell, 2001).

In the *geometric random graph* model, nodes correspond to independently and uniformly randomly distributed points in a metric space, and two nodes are linked by an edge if the distance between them is smaller than or equal to some radius r , where distance is an arbitrary distance norm in the metric space (see the definition in the paper; more details about geometric random graphs can be found in (Penrose, 2003)). We used 2-, 3-, and 4-dimensional “squares” and the Euclidean distance measure to construct GEO-2D, GEO-3D, and GEO-4D graphs corresponding to PPI networks (more details are given below).

Note that other network models exist and new ones will certainly be designed in the future to model the real-world phenomena better. It is not possible to predict how many and which new models will appear in the future; this will largely depend on the new data that need to be modeled.

We limited our current study to the above four network models for two reasons. First, these models are very well established and are most extensively used for modeling various real-world phenomena, so we wanted to evaluate how well each of them models PPI networks. Second, we believe that our results show that geometric random graphs provide a major modeling improvement for PPI networks over the currently well accepted scale-free model. We believe that further major improvements will only be possible when larger and less noisy PPI data sets become available.

1.2 Graphlet Analysis

We analyzed graphlet frequencies of four PPI networks: (1) the high-confidence yeast *S. cerevisiae* PPI network involving 2455 interactions amongst 988 proteins (von Mering *et al.*, 2002); (2) the yeast *S. cerevisiae* PPI network involving 11000 interactions amongst 2401 proteins (von Mering *et al.*, 2002) (these are the top 11000 interactions in von Mering *et al.* classification (von Mering *et al.*, 2002)); (3) the high-confidence fruitfly *D. melanogaster* PPI network involving 4637 interactions amongst 4602 proteins (Giot *et al.*, 2003); (4) and the entire fruitfly *D. melanogaster* PPI network as published in (Giot *et al.*, 2003) involving 20007 interactions amongst 6985 proteins which includes low confidence interactions. Since we are interested in modeling entire PPI networks of various organisms, analyzing only a subset of a PPI network obtained by a single experimental technique or by a single lab is inappropriate because of an experimental bias that a single high-throughput experimental technique inevitably carries. Thus, the yeast PPI data set compiled by von Mering *et al.* (von Mering *et al.*, 2002) containing PPI interactions detected by various techniques and verified by various labs is best suited for PPI network modeling studies. Note that the two yeast PPI networks that we analyzed contain only about 3% and 14% of the currently publicly available yeast PPI interactions respectively, contain interactions of highest confidence levels only, and are obtained by a variety of different experimental techniques and verified by various labs. Thus, we believe that these PPI data sets represent the best currently available PPI networks to try to model. Consequently, we were not surprised to obtain the best network modeling results for these two networks. We believe that the fruitfly PPI

networks, which are obtained by a single lab and a single experimental technique (Giot *et al.*, 2003), are much more noisy than the two yeast PPI networks. Therefore, the noise present in the larger fruitfly PPI network largely influences its topology and makes it correspond to a scale-free model.

All graphlet counts were obtained by the exhaustive graphlet search algorithm going through node adjacency lists rather than the adjacency matrix for increased efficiency. LEDA library for combinatorial and geometric computing was used in its implementation (Mehlhorn & Naher, 1999). In graphlet frequency figures in the paper and in the supplementary graphlet frequency figures, graphlet numbers on the abscissae are ordered as in Figure 1 in the paper, so that the node and edge numbers of the graphlets they represent are monotonically increasing; all zero graphlet frequencies are approximated by 0.1 for plotting them on log-scale. The same replacement of 0 graphlet counts by 0.1 was done for calculating the distances between networks (for taking logarithm in the distance formula described in the paper).

To evaluate how well the PPI networks fit different network models with respect to their graphlet frequencies, we compared frequencies of the 29 3-5-node graphlets in PPI networks against their frequencies in different types of random networks. For each of the PPI networks we constructed five different random graphs belonging to each of the following four random graph models: (1) Erdős-Rényi $G_{n,m}$ random graphs having the same number of nodes and edges as the corresponding PPI networks (denoted by ER); (2) Erdős-Rényi $G_{n,m}$ random graphs having the same number of nodes, edges, and the degree distribution as the corresponding PPI networks (denoted by ER-DD); (3) scale-free random graphs having the same number of nodes and the number of edges within 1% of those of the corresponding PPI networks (denoted by SF); (4) 2-, 3-, and 4-dimensional geometric random graphs with the number of nodes and the number of edges within 1% of those of the corresponding PPI networks (denoted by GEO-2D, GEO-3D, and GEO-4D, respectively). Note that it is enough to compare graphlet frequencies of the PPI networks against a very small number of random graphs of each of these types because graphs belonging to each network model have almost identical graphlet frequency distribution (this can be experimentally observed in Supplementary Figures 2 - 3 and Sup-

plementary Table 3 and also theoretically proved). In addition, searching for graphlets in large networks exhaustively is very computing intensive; for example, on Asus A7M-266D (AthlonMP1900+) and Dell PowerEdge 2650 (P4Xeon-2.80GHz) machines, it took up to one week to exhaustively find all graphlets in SF random networks corresponding to the high confidence fruitfly PPI network, while it took about two months to do this in an SF random network corresponding to the entire currently publicly available fruitfly PPI network. A heuristic sampling algorithm for estimating the number of subgraphs of a network has recently been proposed (Kashtan *et al.*, 2004). However, to evaluate how well the algorithm estimates the number of subgraphs in a specific network, results of the exhaustive search algorithms are still necessary (Kashtan *et al.*, 2004). This is why we chose to analyze the results of exhaustive graphlet searches in this study and why it is computationally infeasible to analyze a larger number of large random networks.

1.2.1 Construction of Model Networks

We tested graphlet frequencies of five different $G_{n,m}$ graphs for $n = 988$ nodes and $m = 2455$ edges with those of five different $G_{n,p}$ graphs with the same n and $p = \frac{m}{\binom{n}{2}}$ and obtained identical frequency distributions. Thus, we used the $G_{n,m}$ model in our analysis. We used the random graph generation function from LEDA library to construct ER graphs of $G_{n,m}$ type corresponding to PPI networks. The ER-DD random graphs were constructed by generating n nodes, where n is equal to the number of nodes in a PPI network, assigning the degree sequence of a PPI network to these generated nodes, sorting the nodes by degree in decreasing order, and generating edges from the sorted nodes towards randomly selected nodes while preserving the assigned degree distribution. Since this process yields some impossible edge assignments, we repeated this process several times until it yielded the desired graph; it took between 7 and 15 experiments to generate each of the desired ER-DD graphs.

Scale-free random graphs $G(V, E)$ were generated in the following way. Let $k = \frac{|E|}{|V|}$, and let an integer i be such that $i \leq k \leq i + 1$. We start the graph construction with an independent set of size i . We add a node to this graph and connect it with the i nodes of the initial

independent set. We add subsequent nodes and connect them with either i or $i + 1$ other nodes in the graph with probabilities $i + 1 - k$ and $k - i$ respectively; attachment is preferential (a new node is more likely to be attached to a high-degree than to a low-degree node) and directly proportional to the degree of the node that a new node is being attached to.

We constructed 2-, 3-, and 4-dimensional geometric random graphs with l_2 (Euclidean) norm corresponding to each of the above mentioned PPI networks. The parameters used for their construction are presented in Table 1.

1.2.2 Graphlet Frequency Results

We first established that graphlet frequencies of the two yeast PPI networks are highly correlated (see Supplementary Figure 1 A and B). To quantify the correlation between the graphlet frequencies for these two and other networks, we used the following simple method (also described in the paper). We first normalized graphlet frequencies of each graph in the following way: let $N_i(G)$ be the number of graphlets of type i ($i \in \{1, \dots, 29\}$) of graph G ; let the total number of graphlets of G be denoted by $T(G) = \sum_{i=1}^{29} N_i(G)$; we computed normalized values of frequencies, $F_i(G) = -\log\left(\frac{N_i(G)}{T(G)}\right)$. Note that we use \log because we are interested in fractional (or percentage) differences between graphlet frequencies, and we change its sign to make its value positive and plotting them more intuitive. Then we measured differences between normalized graphlet frequencies of graphs G and H using the *distance* function $D(G, H) = \sum_{i=1}^{29} |F_i(G) - F_i(H)|$; the smaller the value of the distance function $D(G, H)$ is, the more correlated the graphlet frequencies of graphs G and H are. The plots of normalized graphlet frequencies for the two yeast PPI networks are presented in Supplementary Figure 1 B; the distance between these two PPI networks is 21.33. Note that the distance between the two fruitfly PPI networks is about three times as high, 66.29, indicating presence of noise in the larger network, as expected; this can also be seen in Supplementary Figure 1 C and D, which shows non-normalized and normalized graphlet frequencies in the two fruitfly PPI networks, respectively.

Non-normalized plots of graphlet frequencies for yeast PPI networks and their corresponding ER, ER-DD, and

SF random networks are presented in Supplementary Figure 2. Clearly, these plots show that graphlet frequencies of the two yeast PPI networks are far from the graphlet frequencies of the corresponding ER, ER-DD, and SF random networks. This can also be seen from Supplementary Table 3, which shows distances between PPI networks and their corresponding random networks computed by the formula given above. In contrast, the graphlet frequencies of yeast PPI networks and their corresponding 2-dimensional geometric random graphs are very close, as presented in Supplementary Figure 3 A and B, and in Supplementary Table 3. For example, distances between graphlet frequencies of the high confidence yeast PPI network and five corresponding SF random networks are between 125.46 and 142.91, while the distances between this PPI network and five corresponding GEO-2D graphs are about 3.5 times smaller, i.e., between 35.46 and 38.96. The correlation is even stronger between yeast PPI networks and their corresponding 3- and 4- dimensional geometric random graphs, as presented in Supplementary Figure 3 C and D and Supplementary Table 3. Furthermore, the correlation between the yeast high confidence PPI network and 3-dimensional geometric random graphs with the same number of nodes, but about three times as many edges as the high confidence PPI network is particularly striking (Supplementary Figure 3 E and Supplementary Table 3) and so is the correlation between this PPI network and GEO-3D random graphs with the same number of nodes but about 6 times as many edges as the PPI network (Supplementary Figure 3 F and Supplementary Table 3); the distance between the PPI and these random graphs is even lower than the distance between the two yeast PPI networks that we analyzed. Note that we constructed the GEO-3D random graphs which are 6 times denser than the yeast high-confidence PPI network (in terms of the number of edges) in order to obtain similar maximum degrees of the PPI and these model networks (more explanation is given in the paper). We expect that once the complete PPI network for yeast becomes available, it will be much denser than the one we are working with today and it will likely have properties of 3- or 4-dimensional geometric random graphs. This hypothesis seems plausible since it is based on examining local structural properties of PPI networks (more details are given in the paper). Also, note that ER, ER-DD, and SF random networks have graphlet frequency pattern which is com-

pletely different from the graphlet frequency pattern of PPI and GEO random networks. For example, in ER network model, the probability of an edge existing between any two pair of nodes is the same, and thus, the likelihood of edges appearing “close together” and forming a dense graphlet in an ER network is very low; this is why the number of dense graphlets in these networks is orders of magnitude smaller than the number of sparse graphlets with the same number of nodes (see Supplementary Figure 2). Therefore, increasing the density of ER, ER-DD and SF random networks corresponding to the PPI networks would not result in networks with graphlet frequency pattern similar to the corresponding PPI networks and this is why we do not examine such model networks.

The results of the graphlet frequency analysis of the fruitfly PPI networks are presented in Supplementary Figures 4 and 5, and in Supplementary Table 3. Even though the fruitfly high confidence PPI network fits the 4-dimensional geometric random graph model about 1.3 times better than the next closest model, which is the ER model, and about 3.2 times better than the SF model with respect to graphlet frequencies (Supplementary Table 3), the difference in fits is not as striking as for the two yeast PPI networks. One explanation for this may be that, since this is the first publicly available fruitfly PPI data set obtained from cDNAs representing each predicted transcript of the fruitfly genome (Giot *et al.*, 2003), what we are observing is a random sample of the full fruitfly PPI network; by “full” PPI network we mean the PPI network containing all proteins and all PPIs from all cell types in an organism. An alternative explanation may be that, since *D. melanogaster* is a multicellular organism while yeast *S. cerevisiae* consists of a single cell, a different model may be required for the complete fruitfly PPI network, and yet different models may be needed for PPI networks of different fruitfly cells that belong to different tissues. If this is the case, it is reasonable to expect to observe a common PPI network model for all multicellular organisms. Also, the larger fruitfly PPI network, with about 77% of its edges corresponding to lower confidence interactions (Giot *et al.*, 2003), is closest to the scale-free model (Supplementary Figure 4 F and Supplementary Table 3). This is one of the reasons why we believe that the scale-free properties that have been observed in PPI networks are due to the presence of large amount of noise in these networks; we believe that the true structure of PPI

networks is closer to the geometric graph model than to the scale-free model.

The graphlet frequency parameter is robust to random perturbations (Supplementary Table 2 and Supplementary Figures 6, 7, and 8). We perturbed the high-confidence yeast PPI network by randomly adding, deleting, and rewiring 10, 20, and 30 percent of its edges and computed distances between the perturbed networks and the PPI network by using the distance function defined above (and in the paper). We constructed five perturbed networks in each of these nine categories (45 perturbed networks in total). We found the exceptional robustness of the graphlet frequency parameter to random additions of edges very encouraging, especially in light of the currently available PPI networks containing many false negatives (missing edges). In particular, additions of 30% of edges resulted in networks which were about 21 times closer to the PPI network than the corresponding SF random networks (the distances between the PPI and the perturbed PPI networks with additions of 30% of edges were between 5.78 and 7.14, while the distance between the PPI and the corresponding SF random networks were between 125.46 and 142.91, as shown in Supplementary Tables 2 and 3). 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 about 5.8 and 5.4 times closer to the PPI network than the corresponding SF random networks, respectively), which further increases our confidence in PPI networks having geometric properties despite the presence of false positives in the data.

1.3 Standard Global Network Parameters

1.3.1 Definitions

The most commonly studied statistical properties of large networks measuring their global structure are the degree distribution, network diameter, and clustering coefficients, defined as follows (also defined in the paper). 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 between different network models; in particular, Erdős-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 a *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, 1996); we do not use this definition.) This network property also distinguishes different network models: for example, the diameter of Erdős-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 & Lu, 2002; Cohen & Havlin, 2003), i.e., proportional to $\log \log n$. The *clustering coefficient of node v* in a network is defined as $C_v = \frac{2e_1}{n_1(n_1-1)}$, 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 & 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.

1.3.2 Results on PPI and Model Networks

We obtained the degree distributions, diameters, and clustering coefficients for the above four PPI networks, as well as for the recently published *C. elegans* PPI network (Li *et al.*, 2004) involving 5363 interactions between 3115 proteins. We did not obtain graphlet frequencies for this data set because it contains a large number of highly connected nodes which are in close proximity of each other. This makes it infeasible to find graphlets in this PPI network using the standard, brute force exhaustive search technique, which we have used to find graphlets in the

yeast and fruitfly PPI networks, as well as in all of their corresponding model networks.

We confirmed that degree distributions of all of these PPI networks approximately follow power law. In Supplementary Figure 9 we present power law functions fitted to the degree distributions of these five PPI networks. The degree distributions of the fruitfly PPI networks (Supplementary Figure 9 C and D) deviate the most from power-law functions. Note that the exponents γ of the fitted power law functions $P(x) = x^{-\gamma}$ are between 1.2 and 1.7 for most of these PPI networks (the only exception is the high confidence fruitfly PPI network, but the approximation of its degree distribution by $x^{-2.4}$ is very poor). This deviates from what was observed in many other real-world networks, including metabolic pathway networks of 43 different organisms (Jeong *et al.*, 2000; Jeong *et al.*, 2001), where these exponents were between 2 and 3. An illustration of the degree distributions of the PPI networks against the degree distributions of the corresponding model networks is presented in Supplementary Figure 10.

When calculating a network diameter, we used the standard method of considering only the lengths of shortest paths between reachable pairs of nodes, i.e., pairs of nodes which are in the same connected component of the network. Diameters of the PPI networks and the corresponding random networks are presented in Supplementary Table 4. We observed that the diameters of yeast PPI networks are closest to the diameters of GEO-3D networks which are 3 or 6 times denser than the corresponding PPI networks, while diameters of fruitfly and worm PPI networks are closest to the diameters of the corresponding SF networks (the only exception is the diameter of the larger fruitfly PPI network which is slightly closer to the diameters of ER-DD than to the diameters of SF networks; Supplementary Table 4). However, once a network is dense enough to have most of its nodes in the same connected component, with increasing density of GEO-3D (or any other type of) networks on the same number of nodes, their diameters decrease (this is theoretically expected and can also be experimentally observed in Table Supplementary 4), so it would not be hard to construct GEO-3D networks with the same number of nodes but more edges than the fruitfly and the worm PPI networks have to achieve a closer fit of the diameters of the PPI and the GEO-3D model networks. Thus, GEO graphs *do* model PPI net-

works with respect to this network parameter.

Clustering coefficients of the PPI and the corresponding model networks are presented in Supplementary Table 5. Clustering coefficients of the two yeast PPI networks are orders of magnitude larger than the clustering coefficients of the corresponding ER, ER-DD, and SF random networks while they are in agreement with the clustering coefficients of the corresponding GEO graphs (Supplementary Table 5). This can not be said for the other three PPI networks. In particular, the clustering coefficient of the high-confidence fruitfly PPI network differs by an order of magnitude from all model networks, while the clustering coefficients of the larger fruitfly and the worm PPI networks are closest to the clustering coefficients of the corresponding SF random networks; we believe that this may be due to the large amount of noise being present in these PPI networks.

We measured the average clustering coefficients of all nodes of degree k in a network, $C(k)$, for the above mentioned PPI and their corresponding model networks (Supplementary Figures 11 – 15); zero values are replaced with values close to zero and plotted along the abscissa in these supplementary figures. High correlations between $C(k)$ of the two yeast PPI and their corresponding GEO-3D networks (especially the corresponding GEO-3D networks which are about 6 times denser than the PPI networks, Supplementary Figures 11 F and 12 F) and the lack of such correlation with other model networks is blindingly obvious (Supplementary Figures 11 and 12). However, the values of $C(k)$ are much lower for the other three PPI networks and do not seem to correlate with $C(k)$ of any of the corresponding model networks (Supplementary Figures 13 – 15); we find these values of $C(k)$ of PPI networks to be unrealistically small for real-world networks and believe they are an artifact of the lack of PPI data for fruitfly and worm (false negatives). Also, we observed a lack of scaling of $C(k) \sim k^{-1}$ in all of these PPI networks (Supplementary Figures 11 – 15).

2 Supplementary Figures

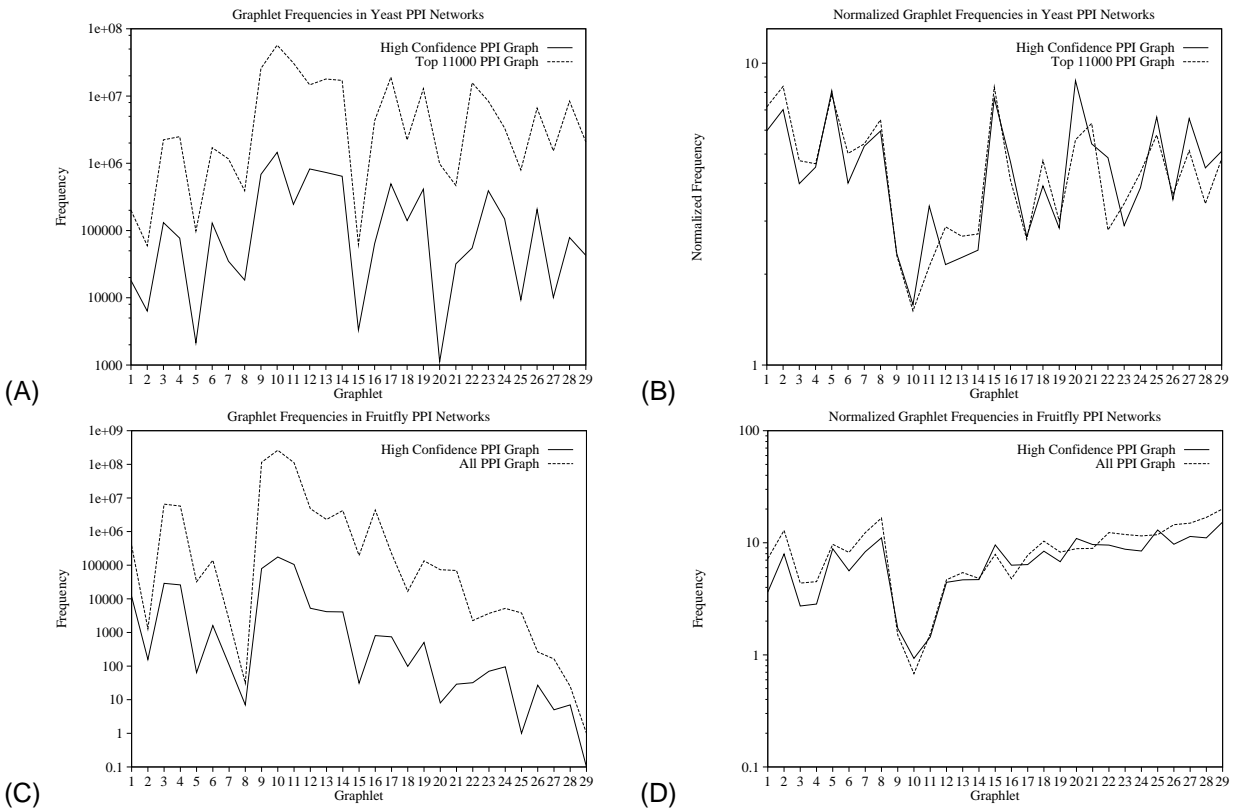


Figure 1: Graphlet frequencies in the two yeast PPI networks (von Mering *et al.*, 2002) and in the two fruitfly PPI networks (Giot *et al.*, 2003): **A.** Non-normalized frequencies for the two yeast PPI networks. **B.** Normalized frequencies for the two yeast PPI networks. **C.** Non-normalized frequencies for the two fruitfly PPI networks. **D.** Normalized frequencies for the two fruitfly PPI networks.

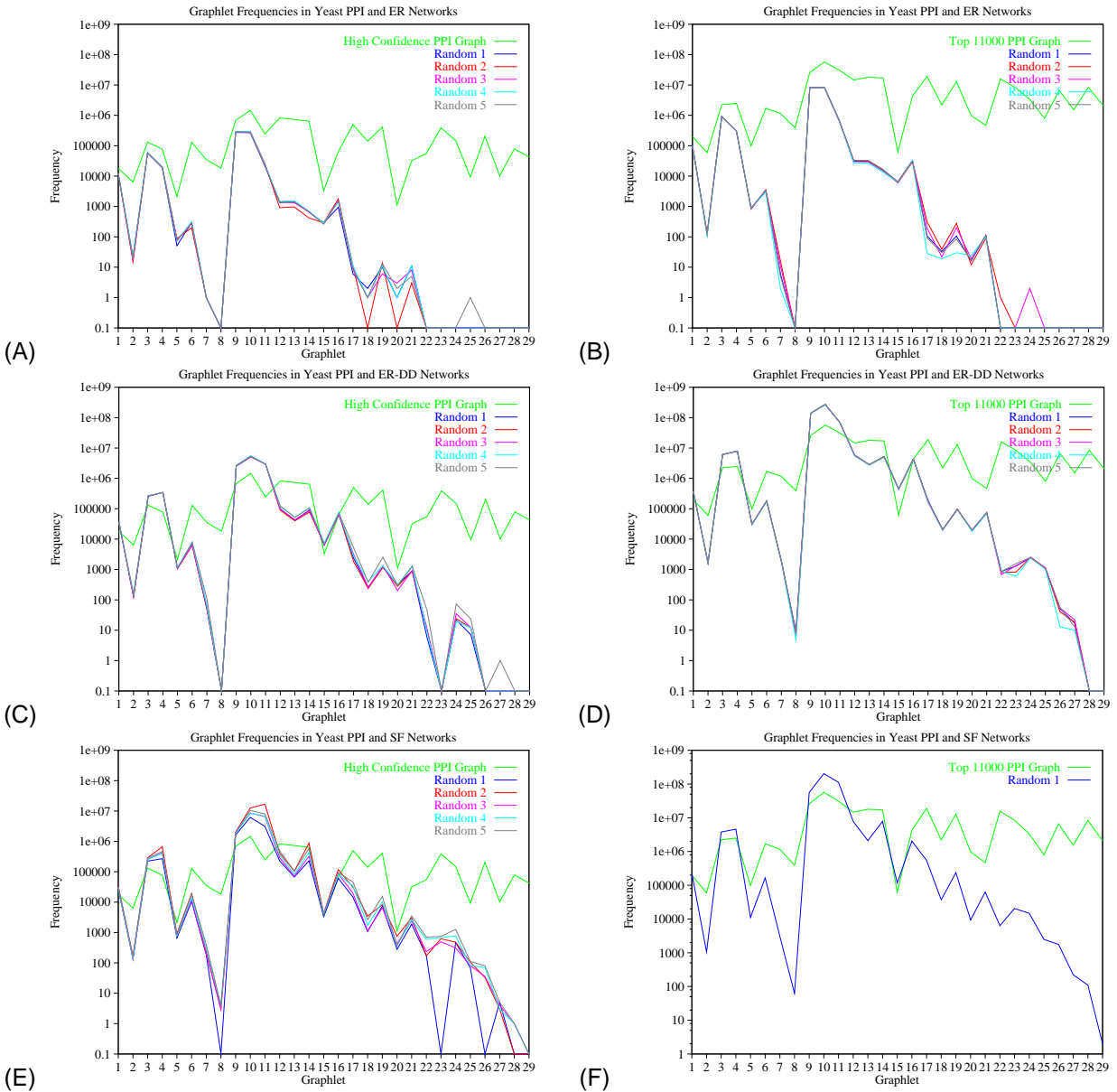


Figure 2: Comparison of frequencies of graphlets in the *S. cerevisiae* PPI networks (von Mering *et al.*, 2002) with ER, ER-DD, and SF random graphs: **A.** High confidence PPI network *versus* the corresponding ER random graphs. **B.** Top 11000 PPI network *versus* the corresponding ER random graphs. **C.** High confidence PPI network *versus* the corresponding ER-DD random graphs. **D.** Top 11000 PPI network *versus* the corresponding ER-DD random graphs. **E.** High confidence PPI network *versus* the corresponding SF random graphs. **F.** Top 11000 PPI network *versus* the corresponding SF random graphs.

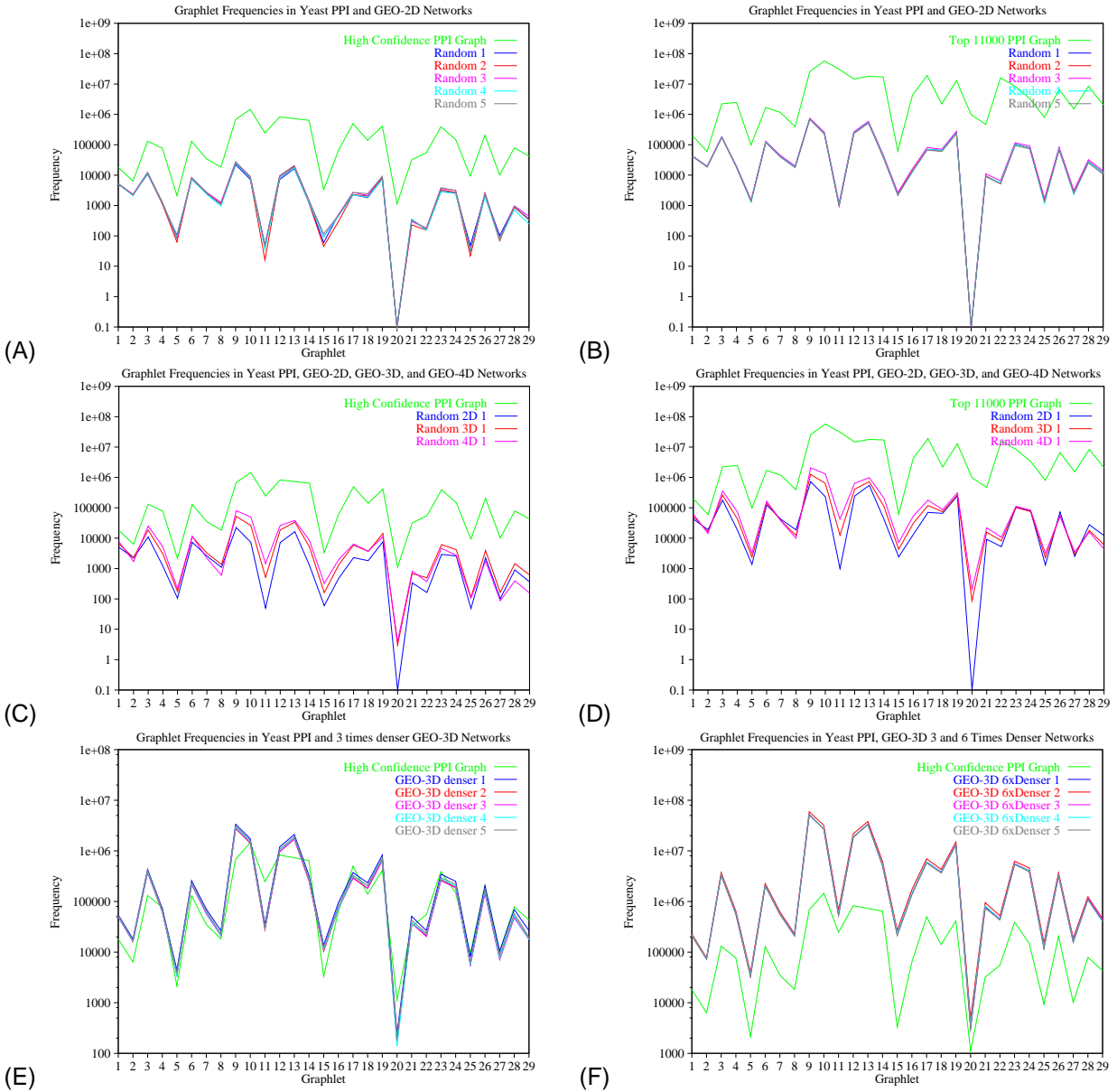


Figure 3: Comparison of frequencies of graphlets in the *S. cerevisiae* PPI networks (von Mering *et al.*, 2002) and geometric random graphs: **A.** High confidence PPI network *versus* the five corresponding GEO-2D graphs. **B.** Top 11000 PPI network *versus* the five corresponding GEO-2D graphs. **C.** High confidence PPI network *versus* a 2-, 3-, and 4-dimensional geometric random graph. **D.** Top 11000 PPI network *versus* a 2-, 3-, and 4-dimensional geometric random graph. **E.** High confidence PPI network *versus* five GEO-3D graphs with the same number of nodes, but approximately three times as many edges as the PPI network. **F.** High-confidence *S. cerevisiae* PPI networks (von Mering *et al.*, 2002) *versus* GEO-3D networks which are about 6 times as dense as the PPI network.

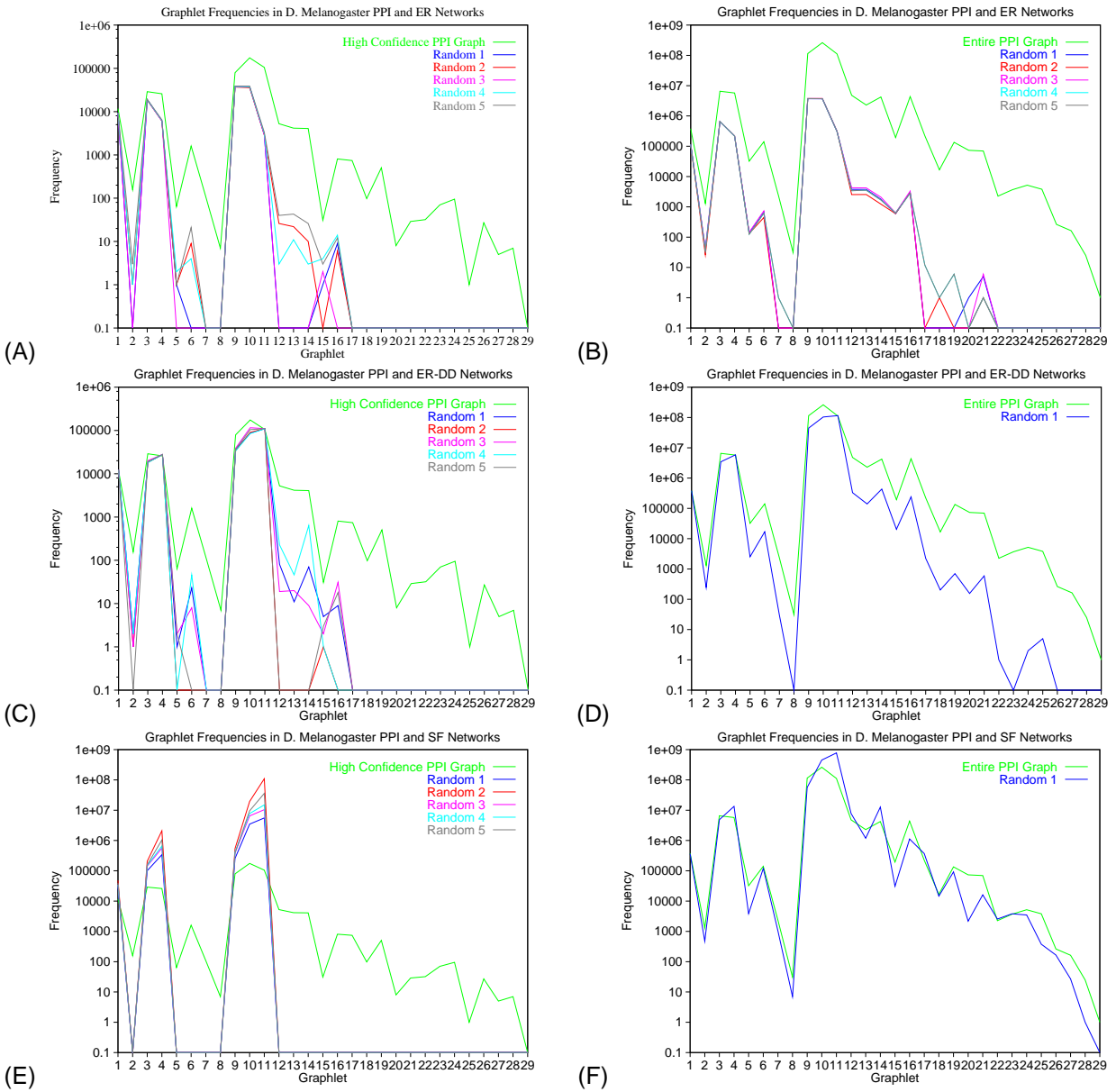


Figure 4: Comparison of frequencies of graphlets in the *D. melanogaster* PPI networks (Giot *et al.*, 2003) with ER, ER-DD, and SF random graphs: **A.** High confidence fruitfly network *versus* the corresponding ER random graphs. **B.** Entire currently available fruitfly PPI network *versus* the corresponding ER random graphs. **C.** High confidence fruitfly PPI network *versus* the corresponding ER-DD random graphs. **D.** Entire currently available fruitfly PPI network *versus* the corresponding ER-DD random graphs. **E.** High confidence fruitfly PPI network *versus* the corresponding SF random graphs. **F.** Entire currently available fruitfly PPI network *versus* the corresponding SF random graphs.

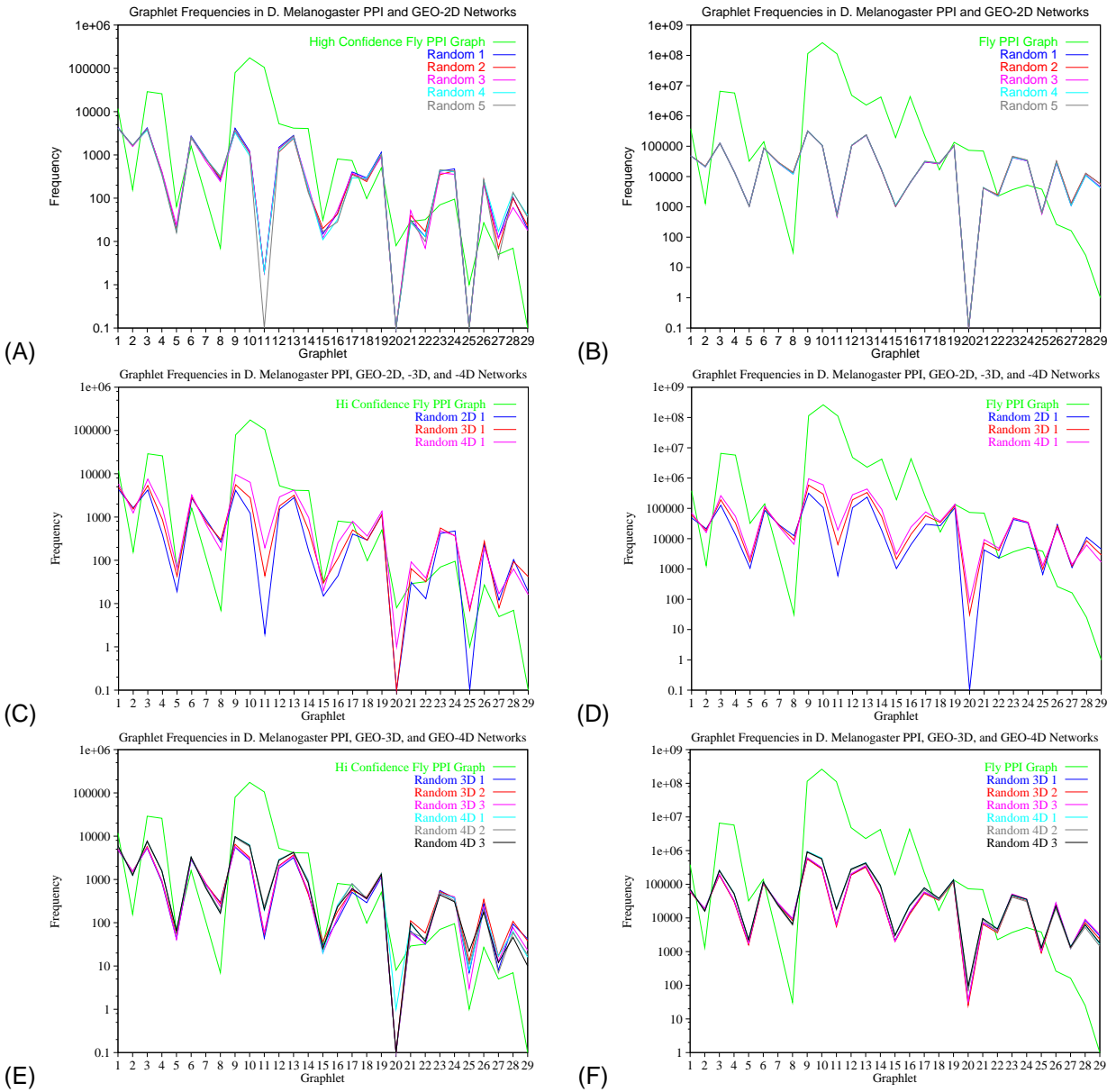


Figure 5: Comparison of frequencies of graphlets in *D. melanogaster* PPI networks (Giot *et al.*, 2003) with geometric random graphs: **A.** High confidence fruitfly PPI network *versus* GEO-2D random graphs. **B.** Entire currently available fruitfly PPI network *versus* GEO-2D random graphs. **C.** High confidence fruitfly PPI network *versus* GEO-2D, GEO-3D, and GEO-4D random graphs. **D.** Entire currently available fruitfly PPI network *versus* GEO-2D, GEO-3D, and GEO-4D random graphs. **E.** High confidence fruitfly PPI network *versus* three GEO-3D random graphs and three GEO-4D random graphs. **F.** Entire currently available fruitfly PPI network *versus* three GEO-3D random graphs and three GEO-4D random graphs.

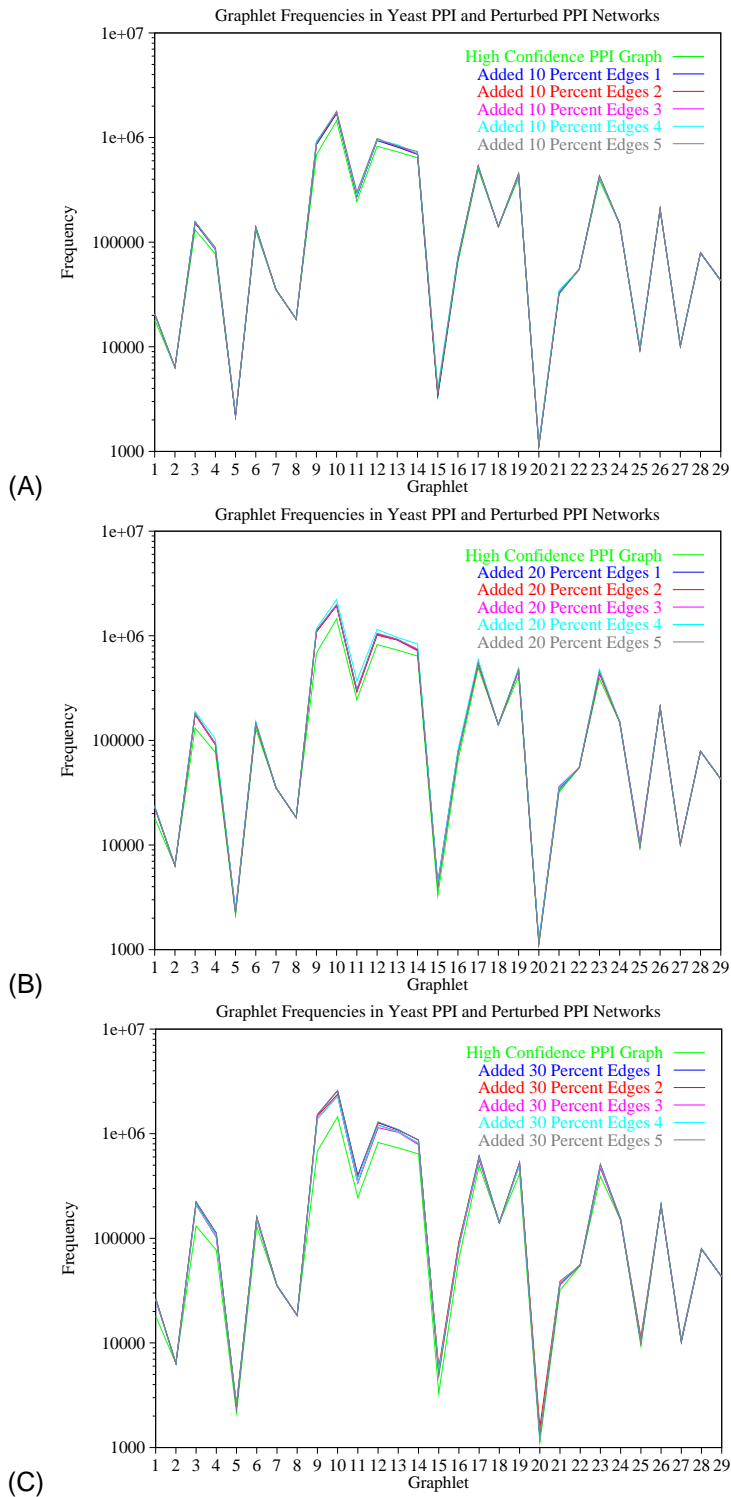


Figure 6: Comparison of graphlet frequencies in the high confidence yeast PPI network with networks obtained by adding edges at random to the PPI network: **A.** five different networks with 10% of edges added. **B.** five different networks with 20% of edges added. **C.** five different networks with 30% of edges added.

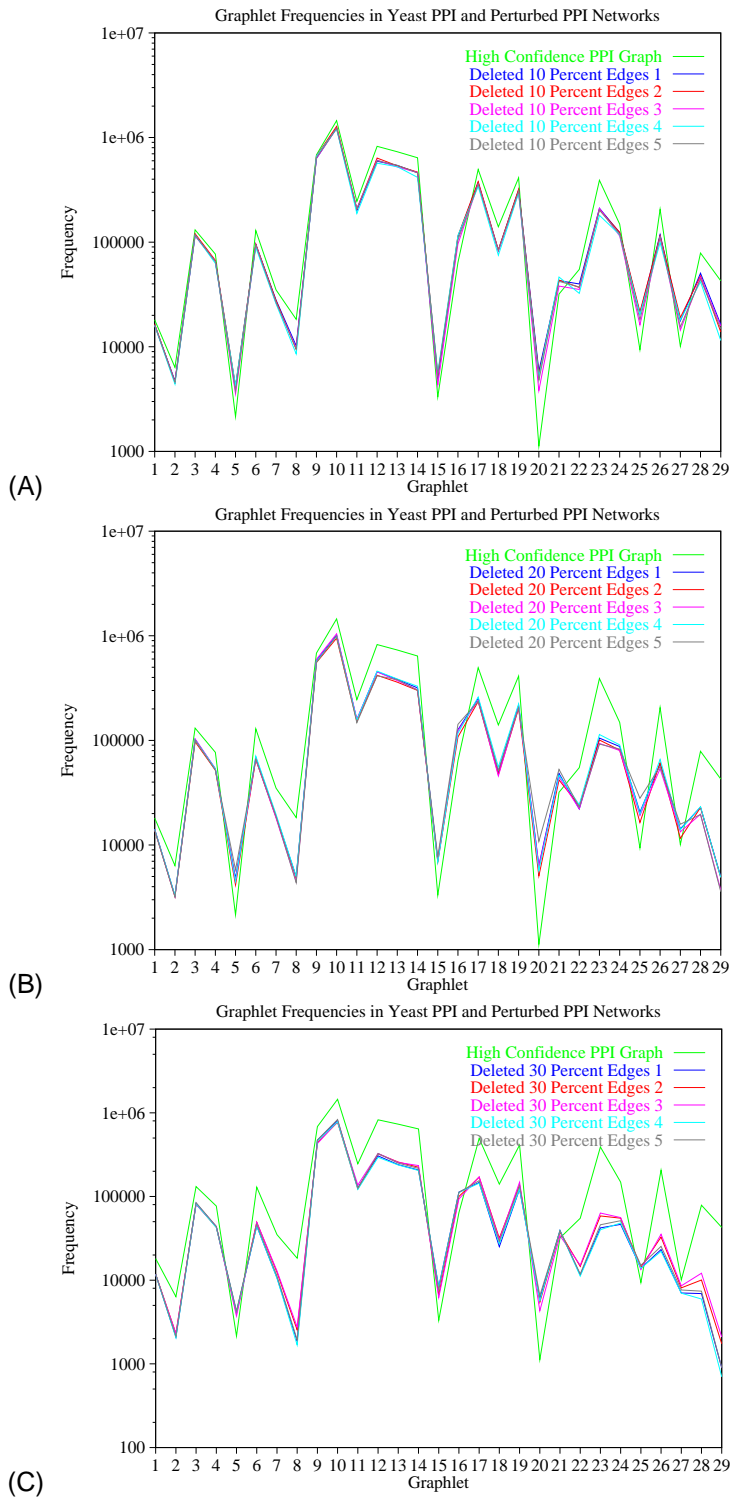


Figure 7: Comparison of graphlet frequencies in the high confidence yeast PPI network with networks obtained by deleting edges at random from the PPI network: **A.** five different networks with 10% of edges deleted. **B.** five different networks with 20% of edges deleted. **C.** five different networks with 30% of edges deleted.

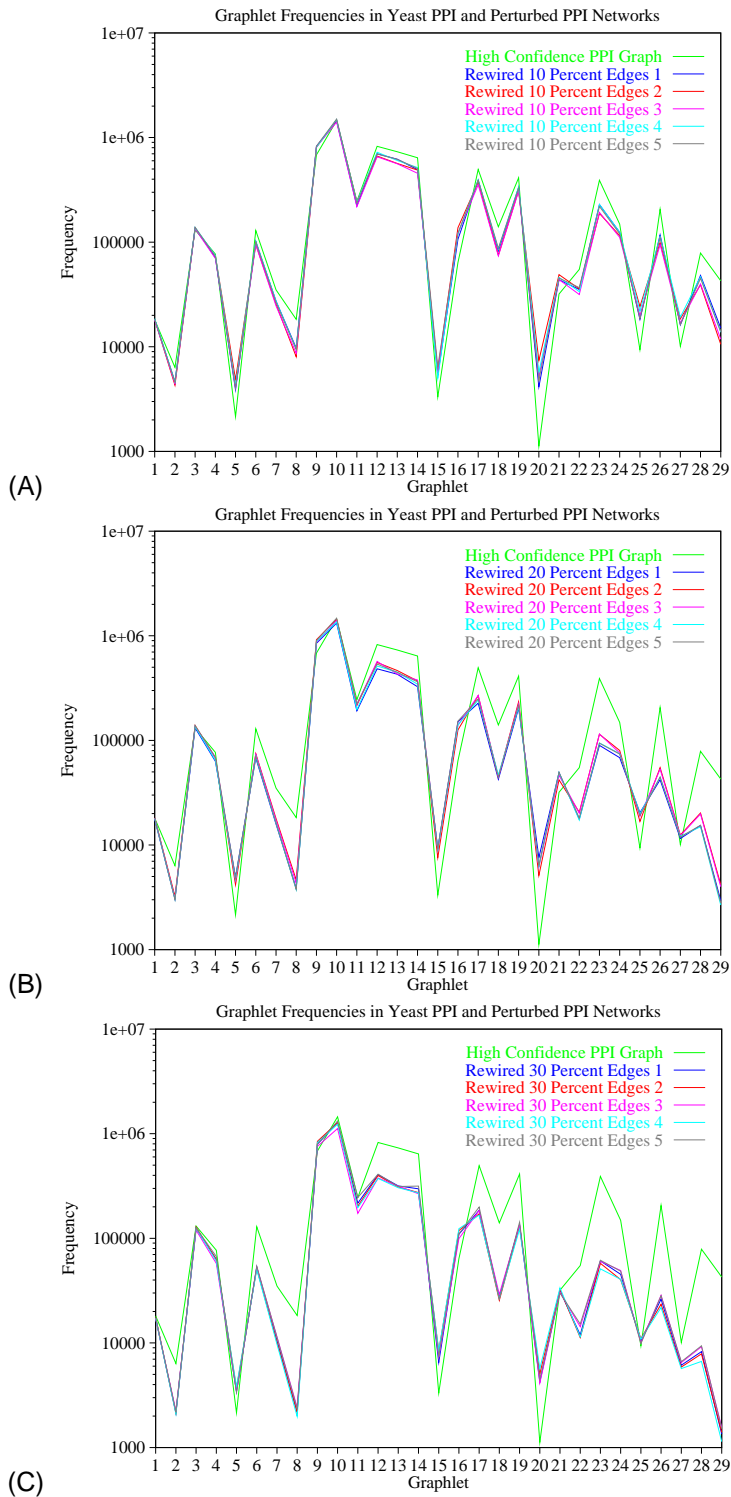


Figure 8: Comparison of graphlet frequencies in the high confidence yeast PPI network with networks obtained by rewiring edges at random in the PPI network: **A.** five different networks with 10% of edges rewired. **B.** five different networks with 20% of edges rewired. **C.** five different networks with 30% of edges rewired.

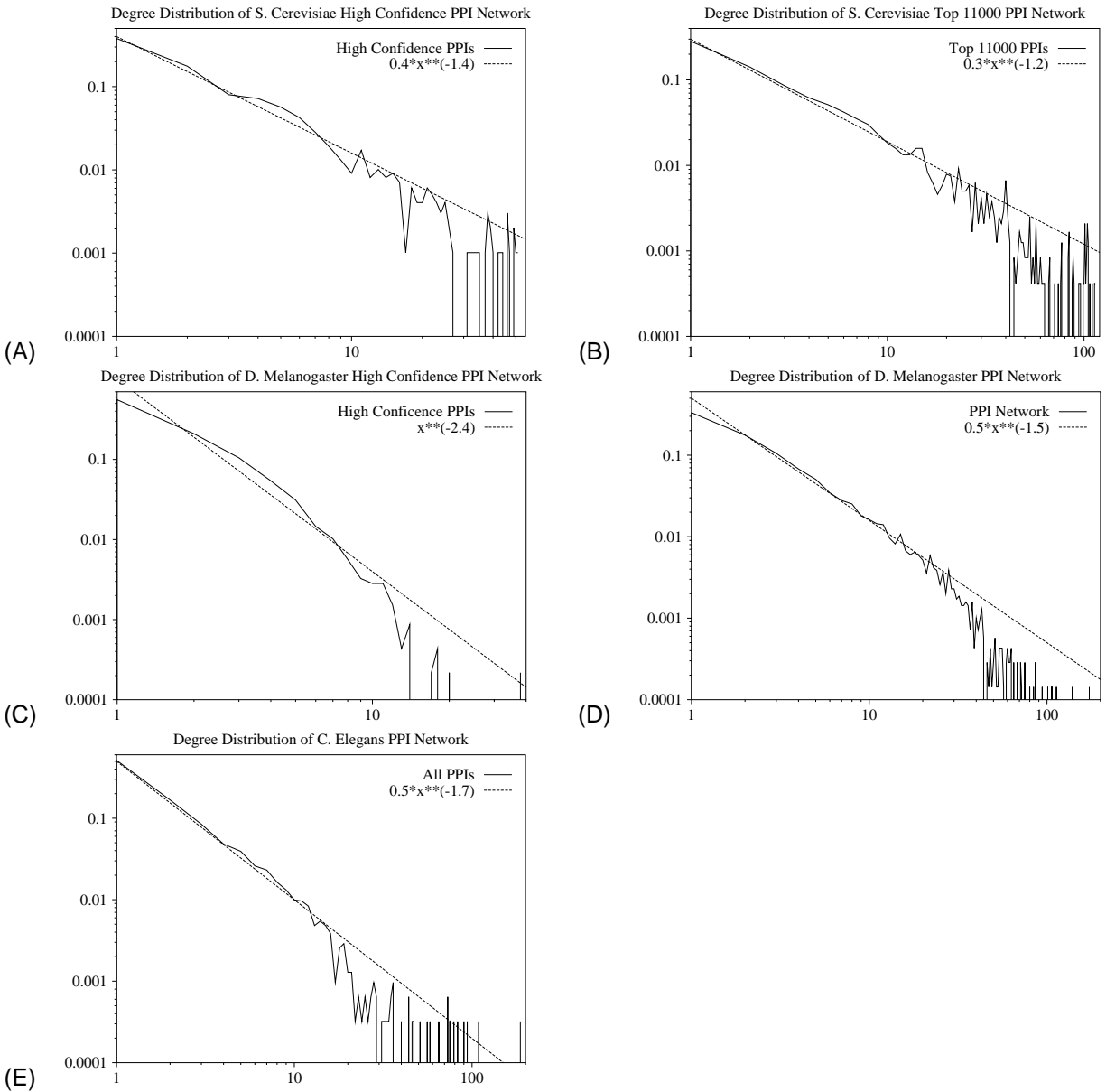


Figure 9: Degree distributions for PPI networks. The abscissa for each panel is node degree x and the ordinate is the probability distribution of degrees, i.e., the fraction of nodes having degree x . **A.** High confidence *S. cerevisiae* PPI network (von Mering *et al.*, 2002). **B.** *S. cerevisiae* “top 11000” PPI network (von Mering *et al.*, 2002). **C.** *D. melanogaster* high confidence PPI network (Giot *et al.*, 2003). **D.** Entire currently available *D. melanogaster* PPI network (Giot *et al.*, 2003). **E.** *C. elegans* PPI network (Li *et al.*, 2004).

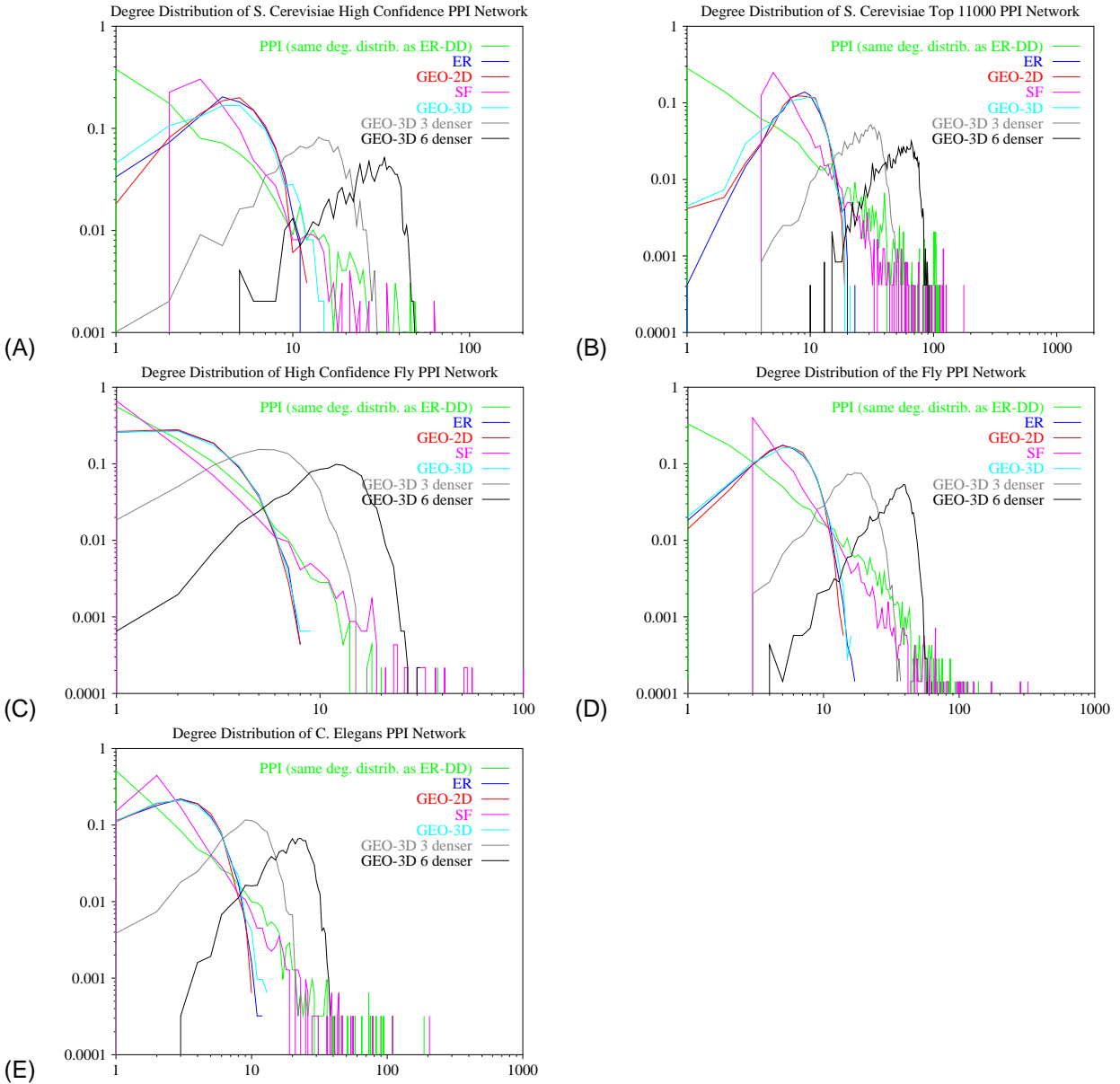


Figure 10: Degree distributions for PPI networks *versus* the degree distributions of the corresponding random graphs. Degree distribution of only one random graph belonging to each of the network models is drawn in each panel, since this is just an illustration. The abscissa for each panel is node degree x and the ordinate is the probability distribution of degrees, i.e., the fraction of nodes having degree x . **A.** High confidence *S. cerevisiae* PPI network (von Mering *et al.*, 2002). **B.** *S. cerevisiae* "top 11000" PPI network (von Mering *et al.*, 2002). **C.** *D. melanogaster* high confidence PPI network(Giot *et al.*, 2003) . **D.** Entire currently available *D. melanogaster* PPI network (Giot *et al.*, 2003). **E.** *C. elegans* PPI network (Li *et al.*, 2004).

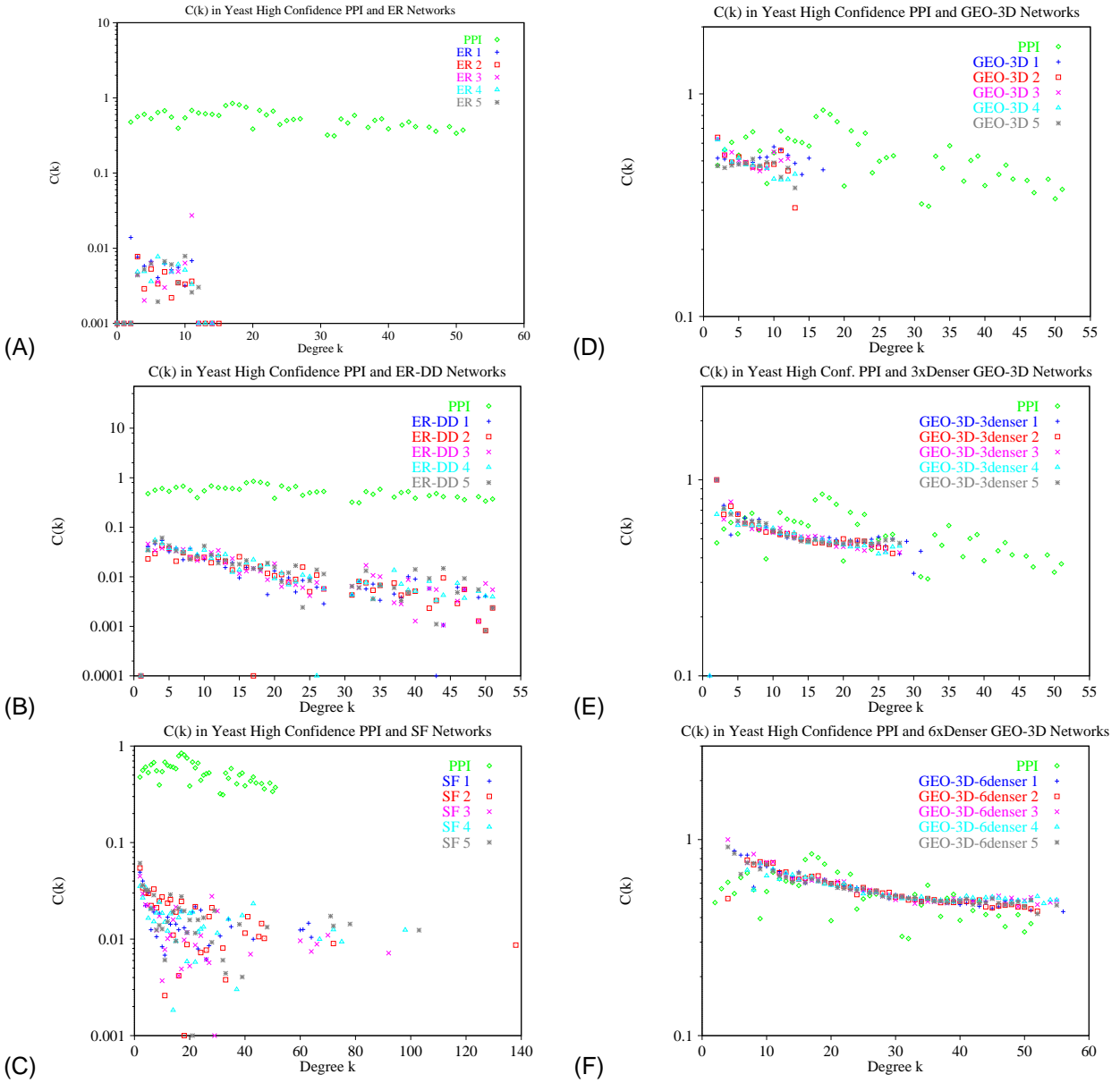


Figure 11: Comparison of degree k node average clustering coefficients $C(k)$ as a function of degree k in high confidence *S. cerevisiae* PPI network (von Mering *et al.*, 2002) and corresponding random graphs: **A.** PPI network *versus* five corresponding ER networks. **B.** PPI network *versus* five corresponding ER-DD networks. **C.** PPI network *versus* five corresponding SF networks. **D.** PPI network *versus* five corresponding GEO-3D networks. **E.** PPI network *versus* five corresponding GEO-3D networks which are 3 times denser than the PPI network. **F.** PPI network *versus* five corresponding GEO-3D networks which are 6 times denser than the PPI network.

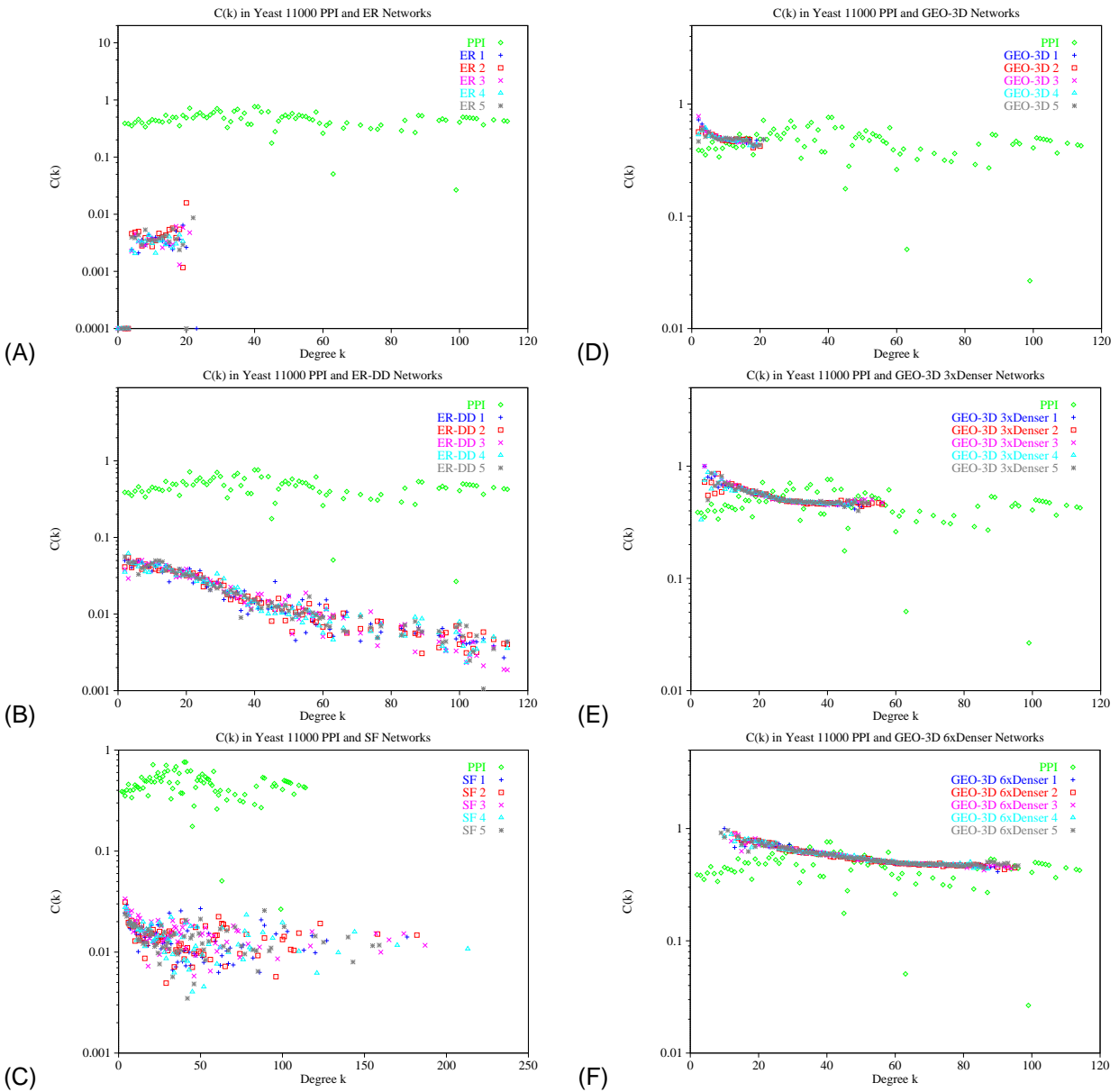


Figure 12: Comparison of degree k node average clustering coefficients $C(k)$ as a function of degree k in the “top 11000” *S. cerevisiae* PPI network (von Mering *et al.*, 2002) and corresponding random graphs: **A.** PPI network *versus* five corresponding ER networks. **B.** PPI network *versus* five corresponding ER-DD networks. **C.** PPI network *versus* five corresponding SF networks. **D.** PPI network *versus* five corresponding GEO-3D networks. **E.** PPI network *versus* five corresponding GEO-3D networks which are 3 times denser than the PPI network. **F.** PPI network *versus* five corresponding GEO-3D networks which are 6 times denser than the PPI network.

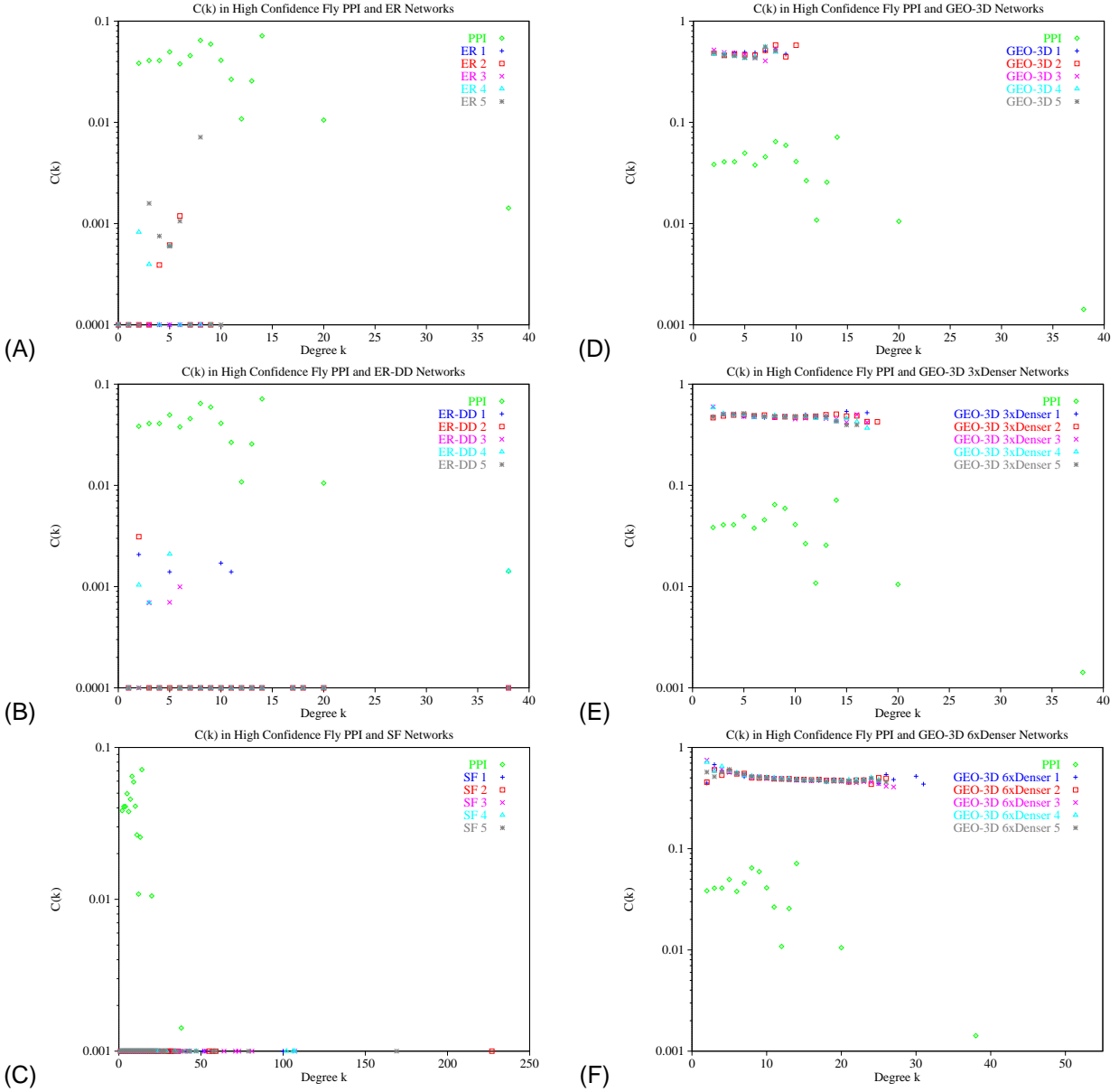


Figure 13: Comparison of degree k node average clustering coefficients $C(k)$ as a function of degree k in the high-confidence *D. melanogaster* PPI network (Giot *et al.*, 2003) and corresponding random graphs: **A.** PPI network *versus* five corresponding ER networks. **B.** PPI network *versus* five corresponding ER-DD networks. **C.** PPI network *versus* five corresponding SF networks. **D.** PPI network *versus* five corresponding GEO-3D networks. **E.** PPI network *versus* five corresponding GEO-3D networks which are 3 times denser than the PPI network. **F.** PPI network *versus* five corresponding GEO-3D networks which are 6 times denser than the PPI network.

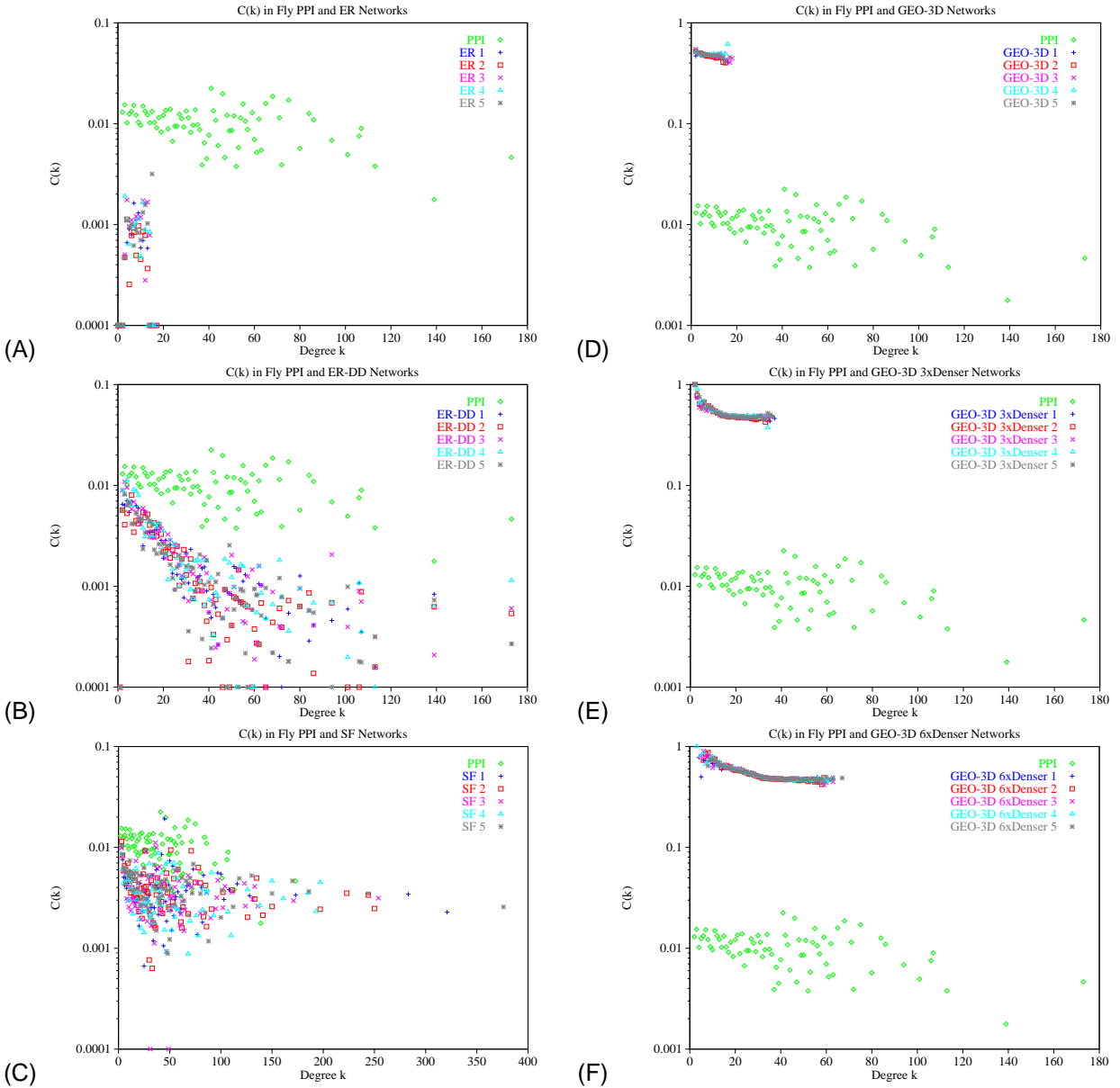


Figure 14: Comparison of degree k node average clustering coefficients $C(k)$ as a function of degree k in the entire currently available *D. melanogaster* PPI network (Giot *et al.*, 2003) and corresponding random graphs: **A.** PPI network versus five corresponding ER networks. **B.** PPI network versus five corresponding ER-DD networks. **C.** PPI network versus five corresponding SF networks. **D.** PPI network versus five corresponding GEO-3D networks. **E.** PPI network versus five corresponding GEO-3D networks which are 3 times denser than the PPI network. **F.** PPI network versus five corresponding GEO-3D networks which are 6 times denser than the PPI network.

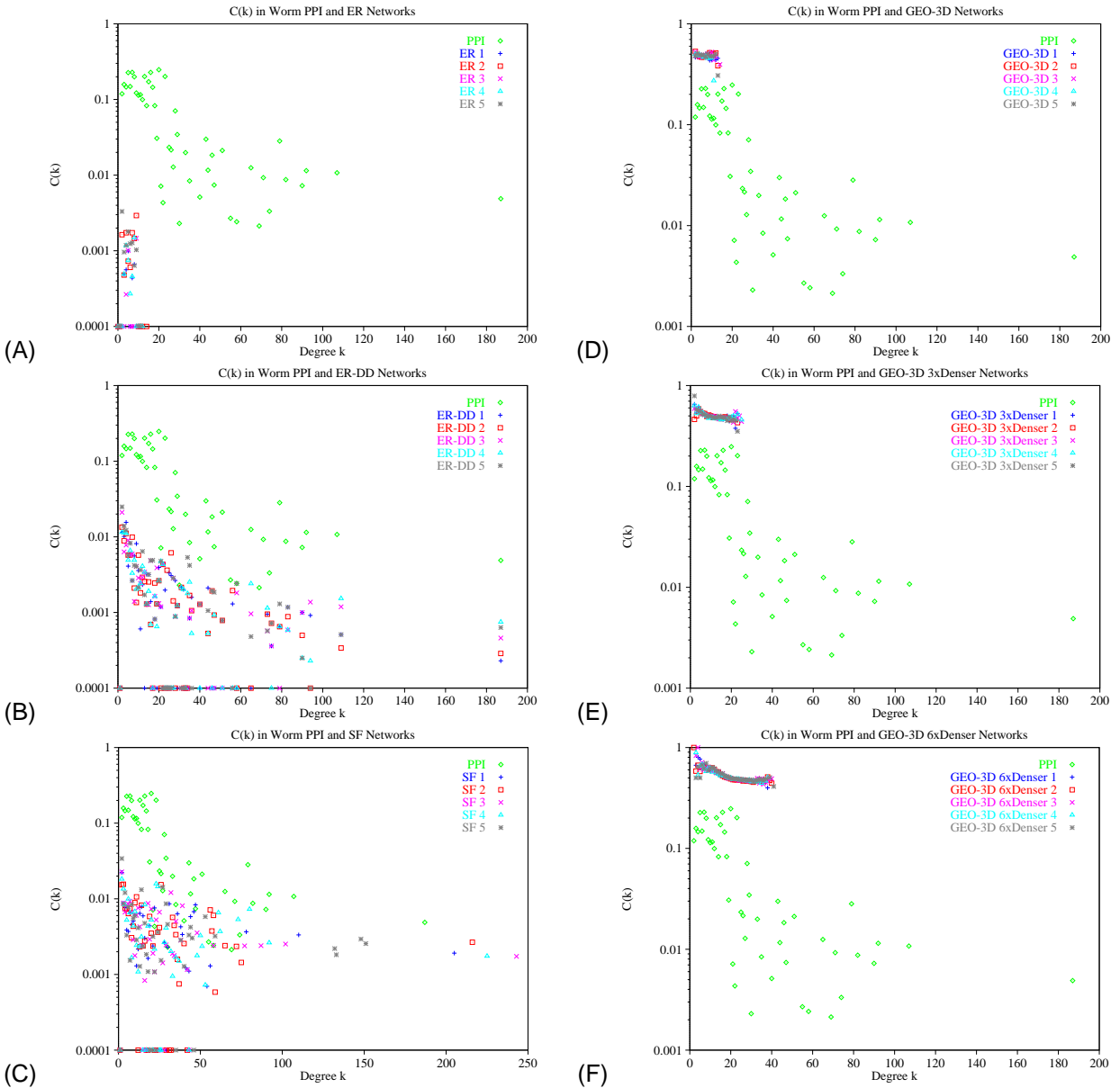


Figure 15: Comparison of degree k node average clustering coefficients $C(k)$ as a function of degree k in the *C. elegans* PPI network (Li *et al.*, 2004) and corresponding random graphs: **A.** PPI network *versus* five corresponding ER networks. **B.** PPI network *versus* five corresponding ER-DD networks. **C.** PPI network *versus* five corresponding SF networks. **D.** PPI network *versus* five corresponding GEO-3D networks. **E.** PPI network *versus* five corresponding GEO-3D networks which are 3 times denser than the PPI network. **F.** PPI network *versus* five corresponding GEO-3D networks which are 6 times denser than the PPI network.

3 Supplementary Tables

	GEO-2D	GEO-3D	GEO-4D	GEO-2D 3 denser	GEO-3D 3 denser	GEO-3D 6 denser
Yeast H.C.	95 x 95, r=4	50 x 50 x 50, r=5.55	30 x 30 x 30 x 30, r=5.74	95 x 95, r=7	50 x 50 x 50, r=8	50 x 50 x 50, r=10.3
Yeast 11K	150 x 150, r=5.3	50 x 50 x 50, r=5	30 x 30 x 30 x 30, r=5.25	150 x 150, r=9.22	50 x 50 x 50, r=7.4	50 x 50 x 50, r=9.5
Fly H.C.	250 x 250, r=3	84 x 84 x 84, r=4	30 x 30 x 30 x 30, r=3	250 x 250, r=5	43 x 43 x 43, r=3	45 x 45 x 45, r=4
Fly Larger	310 x 310, r=5	67 x 67 x 67, r=4	35 x 35 x 35 x 35, r=4.13	310 x 310, r=9	51 x 51 x 51, r=4.4	58 x 58 x 58, r=6.4
Worm	160 x 160, r=3	50 x 50 x 50, r=3.3	30 x 30 x 30 x 30, r=3.85	160 x 160, r=5	40 x 40 x 40, r=3.8	40 x 40 x 40, r=4.9

Table 1: Parameters used for construction of geometric random networks corresponding to yeast, fruitfly, and worm PPI networks. Rows represent the PPI networks and columns represent the corresponding geometric random networks. The values in the boxes represent sizes of 2-, 3-, and 4-dimensional “squares” and radii used to construct these model networks. “3 denser” and six “6 denser” denote geometric random graphs with the same number of nodes, but approximately three and six times as many edges as the corresponding PPI network, respectively.

	Add	Delete	Rewire
10% graph 1	2.01	10.02	10.28
10% graph 2	2.19	10.72	13.76
10% graph 3	2.28	9.03	12.53
10% graph 4	2.36	11.88	11.47
10% graph 5	2.61	10.05	11.26
20% graph 1	3.80	17.02	21.91
20% graph 2	3.77	16.02	18.81
20% graph 3	3.74	17.99	19.85
20% graph 4	5.22	16.02	21.45
20% graph 5	4.35	19.31	21.81
30% graph 1	7.14	24.84	24.35
30% graph 2	5.92	20.91	25.77
30% graph 3	5.96	19.72	22.89
30% graph 4	5.78	25.18	26.60
30% graph 5	7.03	24.20	23.88

Table 2: Graphlet frequency distances between high-confidence yeast PPI network and the perturbed networks with 10, 20, and 30 percent of edges added, deleted, or rewired at random. For example, 2.013974 in the top left-most field indicates that the first perturbed network with 10% of edges added at random to the yeast high-confidence PPI network is at distance 2.013974 from the yeast high-confidence PPI network.

	Yeast High Conf.	Yeast top 11000	Fruitfly High Conf.	Fruitfly Entire
ER 1	171.03	213.49	118.70	109.82
ER 2	179.22	207.73	98.23	112.95
ER 3	171.10	208.92	124.54	111.69
ER 4	171.59	217.69	98.47	100.81
ER 5	169.16	213.81	91.24	100.81
ER-DD 1	164.33	157.61	107.40	98.36
ER-DD 2	164.23	159.11	139.98	
ER-DD 3	162.80	158.04	112.84	
ER-DD 4	162.18	161.41	110.19	
ER-DD 5	154.19	158.00	133.96	
SF 1	142.91	125.50	229.71	50.55
SF 2	142.56		297.43	
SF 3	130.97		244.78	
SF 4	125.97		252.43	
SF 5	125.46		270.79	
GEO-2D 1	35.46	51.38	89.89	156.01
GEO-2D 2	38.96	51.21	90.87	156.79
GEO-2D 3	36.78	50.55	89.04	156.15
GEO-2D 4	37.54	51.15	92.28	156.17
GEO-2D 5	37.01	51.09	93.67	156.96
GEO-3D 1	27.25	40.79	90.39	142.89
GEO-3D 2	33.47	41.07	92.64	142.66
GEO-3D 3	34.26	41.52	87.95	142.71
GEO-3D 4	31.70	41.29	86.76	
GEO-3D 5	31.39	41.29	90.06	
GEO-4D 1	32.33	38.78	80.76	134.65
GEO-4D 2	30.96	40.18	82.51	134.44
GEO-4D 3	30.73	39.16	82.17	135.15
GEO-4D 4	31.91	39.02	75.95	
GEO-4D 5	31.42	38.64	82.78	
GEO-2D 3 denser 1	30.55			
GEO-3D 3 denser 1	18.38			
GEO-3D 3 denser 2	19.63			
GEO-3D 3 denser 3	20.16			
GEO-3D 3 denser 4	20.11			
GEO-3D 3 denser 5	19.46			
GEO-3D 6 denser 1	20.50			
GEO-3D 6 denser 2	20.13			
GEO-3D 6 denser 3	19.86			
GEO-3D 6 denser 4	20.10			
GEO-3D 6 denser 5	20.56			

Table 3: Graphlet frequency distances between PPI networks and the corresponding random networks. Rows represent network types (for example, PPI, ER, ER-DD etc.) and columns represent the corresponding data set. Some values are missing due to the lack of computing power needed for exhaustive graphlet searches in large networks (the values of some ER-DD and SF model networks corresponding to the larger fruitfly and yeast PPI networks), or because they are not relevant for our analysis (the remaining missing values). The last 11 rows corresponds to 2- and 3-dimensional geometric random graphs with the same number of nodes, but approximately three (“3 denser”) and six (“6 denser”) times as many edges as the corresponding PPI graph.

	Yeast High Conf.	Yeast top 11000	Fruitfly High Conf.	Fruitfly Entire	Worm Entire
PPI network	5.19	4.93	9.43	4.47	4.96
ER 1	4.48	3.75	10.94	5.27	6.55
ER 2	4.45	3.75	11.23	5.26	6.56
ER 3	4.47	3.75	11.19	5.26	6.57
ER 4	4.47	3.75	11.09	5.26	6.53
ER 5	4.46	3.75	10.88	5.26	6.56
ER-DD 1	3.65	3.33	12.92	4.23	4.66
ER-DD 2	3.66	3.32	13.26	4.23	4.67
ER-DD 3	3.65	3.33	12.50	4.23	4.65
ER-DD 4	3.65	3.33	13.23	4.23	4.65
ER-DD 5	3.65	3.33	12.82	4.22	4.67
SF 1	3.75	3.32	9.33	4.16	4.84
SF 2	3.62	3.34	7.77	4.12	4.85
SF 3	3.68	3.31	8.35	4.20	4.75
SF 4	3.70	3.30	8.05	4.19	4.81
SF 5	3.65	3.32	8.86	4.12	4.78
GEO-2D 1	26.74	20.36	3.61	49.95	12.71
GEO-2D 2	23.55	20.15	3.51	51.30	18.53
GEO-2D 3	23.32	20.37	3.46	51.26	11.43
GEO-2D 4	22.56	19.95	4.01	51.68	17.95
GEO-2D 5	24.21	19.94	3.81	52.26	14.53
GEO-3D 1	11.89	10.05	6.12	19.45	27.52
GEO-3D 2	11.75	10.21	5.96	19.37	28.87
GEO-3D 3	10.98	10.16	4.83	19.34	30.85
GEO-3D 4	11.55	10.04	8.97	19.06	25.82
GEO-3D 5	12.42	10.21	5.59	18.93	32.47
GEO-4D 1	8.29	7.37	21.74	12.49	16.39
GEO-4D 2	8.03	7.41	17.32	12.39	16.16
GEO-4D 3	8.33	7.35	22.34	12.45	16.54
GEO-4D 4	8.26	7.45	9.68	12.38	16.26
GEO-4D 5	8.59	7.45	9.78	12.65	15.66
GEO-2D 3 denser 1	9.21	9.91	43.17	21.73	22.05
GEO-3D 3 denser 1	5.84	5.78	16.37	10.41	10.44
GEO-3D 3 denser 2	5.95	5.84	16.49	10.33	10.44
GEO-3D 3 denser 3	5.92	5.83	15.99	10.34	10.33
GEO-3D 3 denser 4	5.96	5.82	16.30	10.39	10.46
GEO-3D 3 denser 5	5.90	5.88	16.63	10.29	10.37
GEO-3D 6 denser 1	4.24	4.44	10.43	7.54	7.16
GEO-3D 6 denser 2	4.21	4.41	10.73	7.53	7.15
GEO-3D 6 denser 3	4.29	4.41	10.57	7.57	7.26
GEO-3D 6 denser 4	4.25	4.41	10.67	7.54	7.25
GEO-3D 6 denser 5	4.26	4.39	10.62	7.57	7.29

Table 4: Diameters of PPI networks and their corresponding ER, ER-DD, SF, GEO-2D, GEO-3D, and GEO-4D random networks. Only distances between nodes in the same connected component are reported. Rows represent network types (for example, PPI, ER, ER-DD etc.) and columns represent the corresponding data set. “3 denser” and “6 denser” have the same meaning as in Supplementary Table 3.

	Yeast High Conf.	Yeast top 11000	Fruitfly High Conf.	Fruitfly Entire	Worm Entire
PPI network	0.343733	0.304902	0.017923	0.008434	0.069764
ER 1	0.006276	0.003471	0.000000	0.000820	0.000369
ER 2	0.003828	0.003694	0.000072	0.000632	0.000990
ER 3	0.003721	0.003283	0.000000	0.001034	0.000191
ER 4	0.004622	0.003177	0.000311	0.000972	0.000477
ER 5	0.004372	0.003795	0.000406	0.000807	0.001447
ER-DD 1	0.021663	0.029159	0.000487	0.003655	0.003763
ER-DD 2	0.016042	0.028207	0.000652	0.003210	0.005372
ER-DD 3	0.020910	0.026906	0.000109	0.004831	0.003904
ER-DD 4	0.021649	0.027990	0.000355	0.004874	0.004419
ER-DD 5	0.022328	0.029896	0.000000	0.003544	0.004275
SF 1	0.032902	0.020945	0.000000	0.007449	0.012701
SF 2	0.035412	0.021071	0.000000	0.008452	0.011027
SF 3	0.028359	0.024110	0.000000	0.007440	0.013002
SF 4	0.026749	0.021944	0.000000	0.006447	0.012038
SF 5	0.037393	0.019636	0.000000	0.007994	0.018236
GEO-2D 1	0.561738	0.579556	0.323860	0.562504	0.469442
GEO-2D 2	0.551851	0.578821	0.336352	0.560252	0.477439
GEO-2D 3	0.542434	0.585639	0.329381	0.552721	0.474385
GEO-2D 4	0.534709	0.581867	0.337864	0.555136	0.477508
GEO-2D 5	0.529600	0.577737	0.332533	0.565060	0.474463
GEO-3D 1	0.470745	0.498465	0.284394	0.469083	0.406792
GEO-3D 2	0.482708	0.497322	0.280968	0.470272	0.416314
GEO-3D 3	0.489211	0.500503	0.300616	0.473269	0.414913
GEO-3D 4	0.487208	0.496027	0.258490	0.472467	0.416880
GEO-3D 5	0.456253	0.496865	0.266211	0.470765	0.408034
GEO-4D 1	0.394535	0.430832	0.233926	0.395276	0.352444
GEO-4D 2	0.389667	0.428035	0.243421	0.400708	0.345016
GEO-4D 3	0.406101	0.428948	0.230490	0.395749	0.360883
GEO-4D 4	0.414717	0.442019	0.227551	0.393788	0.333224
GEO-4D 5	0.406548	0.433822	0.217219	0.399770	0.354642
GEO-2D 3 denser 1	0.608068	0.602326	0.563866	0.589434	0.584729
GEO-3D 3 denser 1	0.523472	0.518105	0.470257	0.498982	0.500098
GEO-3D 3 denser 2	0.520203	0.519829	0.472369	0.498337	0.499070
GEO-3D 3 denser 3	0.522614	0.517125	0.473605	0.496640	0.502030
GEO-3D 3 denser 4	0.520670	0.516553	0.478754	0.504280	0.493827
GEO-3D 3 denser 5	0.517115	0.520450	0.479796	0.497585	0.500288
GEO-3D 6 denser 1	0.538432	0.535339	0.492736	0.504445	0.506230
GEO-3D 6 denser 2	0.540154	0.531674	0.496888	0.503071	0.505796
GEO-3D 6 denser 3	0.543724	0.534059	0.495386	0.505574	0.508243
GEO-3D 6 denser 4	0.537005	0.532969	0.500067	0.502805	0.506092
GEO-3D 6 denser 5	0.539893	0.534842	0.493908	0.505338	0.514109

Table 5: Clustering coefficients of PPI networks and their corresponding ER, ER-DD, SF, GEO-2D, GEO-3D, and GEO-4D random networks. Rows represent network types (for example, PPI, ER, etc.) and columns represent the corresponding data set.

References

- Abello, J., Buchsbaum, A. & Westbrook, J. (1998) A functional approach to external graph algorithms. *Lecture Notes in Computer Science*, **1461**, 332–343.
- Aiello, W., Chung, F. & Lu, L. (2001) A random graph model for power law graphs. *Experimental Mathematics*, **10**, 53–66.
- Albert, R. & Barabási, A. L. (2000) Topology of evolving networks: local events and universality. *Phys Rev Lett*, **85** (24), 5234–7.
- Albert, R. & Barabási, A.-L. (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics*, **74**, 47–97.
- Albert, R., Jeong, H. & Barabási, A.-L. (2000) Error and attack tolerance of complex networks. *Nature*, **406**, 378–382.
- Barabási, A. L. & Albert, R. (1999) Emergence of scaling in random networks. *Science*, **286** (5439), 509–12.
- Barabási, A.-L., Albert, R. & Jeong, H. (1999) Mean-field theory for scale-free random networks. *Physica A*, **272**, 173–197.
- Bender, E. A. & Canfield, E. R. (1978) The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory A*, **24**, 296–307.
- Bollobas, B. (1985) *Random Graphs*. Academic, London.
- Bollobas, B. & Riordan, O. (2001) The diameter of a scale-free random graph. *Combinatorica*, . to appear.
- Bornholdt, S. & Ebel, H. (2001) World-wide web scaling exponent from simon’s 1955 model. *Physical Review E*, **64**, 046401.
- Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. & Wiener, J. (2000) Graph structure of the web. *Computer Networks*, **33**, 309–320.
- Callaway, D. S., Newman, M. E. J., Strogatz, S. H. & Watts, D. J. (2000) Network robustness and fragility: percolation on random graphs. *Physical Review Letters*, **85**, 5468–5471.
- Chung, F. & Lu, L. (2002) The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA*, **99**, 15879–15882.
- Cohen, R., Erez, K., ben Avraham, D. & Havlin, S. (2000) Resilience of the internet to random breakdowns. *Physical Review Letters*, **85**, 4626–4628.
- Cohen, R. & Havlin, S. (2003) Scale-free networks are ultra small. *Physical Review Letters*, **90**, 058701.
- Dorogovtsev, S. N. & Mendes, J. F. F. (2000) Evolution of networks with aging of sites. *Physical Review E*, **62**, 1842–1845.
- Erdős, P. & Rényi, A. (1959) On random graphs. *Publicationes Mathematicae*, **6**, 290–297.
- Erdős, P. & Rényi, A. (1960) On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, **5**, 17–61.
- Erdős, P. & Rényi, A. (1961) On the strength of connectedness of a random graph. *Acta Mathematica Scientia Hungary*, **12**, 261–267.
- Faloutsos, M., Faloutsos, P. & Faloutsos, C. (1999) On power-law relationships of the internet topology. *Computer Communications Review*, **29**, 251–262.
- Giot, L., Bader, J., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y., Ooi, C., Godwin, B., Vitols, E., Vijayadamodar, G., Pochart, P., Machineni, H., Welsh, M., Kong, Y., Zerhusen, B., Malcolm, R., Varrone, Z., Collis, A., Minto, M., Burgess, S., McDaniel, L., Stimpson, E., Spriggs, F., Williams, J., Neurath, K., Ioime, N., Agee, M., Voss, E., Furtak, K., Renzulli, R., Aanensen, N., Carrola, S., Bickelhaupt, E., Lazovatsky, Y., DaSilva, A., Zhong, J., Stanyon, C., Finley, R. J., White, K., Braverman, M., Jarvie, T., Gold, S., Leach, M., Knight, J., Shimkets, R., McKenna, M., Chant, J. & Rothberg, J. (2003) A protein interaction map of drosophila melanogaster. *Science*, **302** (5651), 1727–1736.

- Jeong, H., Mason, S. P., Barabási, A. L. & Oltvai, Z. N. (2001) Lethality and centrality in protein networks. *Nature*, **411** (6833), 41–2.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A. L. (2000) The large-scale organization of metabolic networks. *Nature*, **407** (6804), 651–4.
- Kashtan, N., Milo, R., Itzkovitz, S. & Alon, U. (2004) Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. In *Bioinformatics*. to appear.
- Krapivsky, P. L. & Redner, S. (2001) Organization of growing random networks. *Physical Review E*, **63**, 066123–1.
- Krapivsky, P. L., Redner, S. & Leyvraz, F. (2000) Connectivity of growing random networks. *Physical Review Letters*, **85**, 4629–4632.
- Li, S., Armstrong, C., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.-O., Han, J.-D., Chesneau, A., Hao, T., Goldberg, D. S., Li, N., Martinez, M., Rual, J.-F., Lamesch, P., Xu, L., Tewari, M., Wong, S., Zhang, L., Berriz, G., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H., Elewa, A., Baumgartner, B., Rose, D., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S., Saxton, W., Strome, S., van den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K., Harper, J., Cusick, M., Roth, F., Hill, D. & Vidal, M. (2004) A map of the interactome network of the metazoan *C. elegans*. *Science*, **303**, 540–543.
- Luczak, T. (1990) Component behavior near the critical point of the random graph process. *Random Structures and Algorithms*, **1**, 287.
- Mehlhorn, K. & Naher, S. (1999) *Leda: A platform for combinatorial and geometric computing*. Cambridge University Press.
- Molloy, M. & Reed, B. (1995) A critical point of random graphs with a given degree sequence. *Random Structures and Algorithms*, **6**, 161–180.
- Molloy, M. & Reed, B. (1998) The size of the largest component of a random graph on a fixed degree sequence. *Combinatorics, Probability and Computing*, **7**, 295–306.
- Newman, M. E. J. (2002) Random graphs as models of networks. In *Handbook of Graphs and Networks*, (Bornholdt, S. & Schuster, H. G., eds), Wiley-VHC, Berlin.
- Newman, M. E. J., Strogatz, S. H. & Watts, D. J. (2001) Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, **64**, 026118–1.
- Penrose, M. (2003) *Geometric Random Graphs*. Oxford University Press.
- Przulj, N. (2004) Graph theory approaches to protein interaction data analysis. In *Knowledge Discovery in High-Throughput Biological Domains*, (Jurisica, I. & Wigle, D., eds), to appear.
- Ravasz, E. & 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. & Barabási, A.-L. (2002) Hierarchical organization of modularity in metabolic networks. *Science*, **297**, 1551–5.
- von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S. & Bork, P. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, **417** (6887), 399–403.
- Wagner, A. & Fell, D. (2001) The small world inside large metabolic networks. *Proc. Roy. Soc. London Series B*, **268**, 1803–1810.
- West, D. B. (1996) *Introduction to Graph Theory*. Prentice Hall, Upper Saddle River, NJ.
- Wilf, H. S. (1990) *Generating Functionology*. Academic, Boston.