EVOLUTION VERSUS "INTELLIGENT DESIGN": COMPARING THE TOPOLOGY OF PROTEIN-PROTEIN INTERACTION NETWORKS TO THE INTERNET

Q. Yang, G. Siganos, M. Faloutsos and S. Lonardi*

Department of Computer Science and Engineering, University of California, Riverside, CA 92521, USA

**Email: stelo@cs.ucr.edu*

Recent research efforts have made available genome-wide, high-throughput protein-protein interaction (PPI) maps for several model organisms. This has enabled the systematic analysis of PPI networks, which has become one of the primary challenges for the system biology community. In this study, we attempt to understand better the topological structure of PPI networks by comparing them against man-made communication networks, and more specifically, the Internet.

Our comparative study is based on a comprehensive set of graph metrics. Our results exhibit an interesting dichotomy. On the one hand, both networks share several macroscopic properties such as scale-free and small-world properties. On the other hand, the two networks exhibit significant topological differences, such as the cliqueshness of the highest degree nodes. We attribute these differences to the distinct design principles and constraints that both networks are assumed to satisfy. We speculate that the evolutionary constraints that favor the survivability and diversification are behind the building process of PPI networks, whereas the leading force in shaping the Internet topology is a decentralized optimization process geared towards efficient node communication.

1. INTRODUCTION

From an engineering perspective, cells are complex systems that process information. The main mechanism by which cells are able to process information is through protein-protein interactions (PPI). Cellular proteins either aggregate in protein complexes or act concertedly to assemble, store and transduce biological information in an efficient and reliable way. Pathways of interactions between proteins can be found in essentially every cellular process, e.g., signal transduction cascades, metabolism, cell cycle control, apoptosis. Recently, a number of experimental, genome-wide, high-throughput studies have been conducted to determine protein-protein interactions and the consequent interaction networks in several model organisms (see, e.g., Refs. 1 and 2). They provide a unique opportunity to study the complex dynamics of "message passing" in cellular networks at the genome-scale.

The overarching goal of our study is to understand better the topological properties and structure of PPI networks. To do this, along the lines of comparative genomics, we propose to compare PPI networks against one of the largest and the most successful communication networks, the Internet. The main question we tackle in this paper is: how different or similar are the two types of networks? This comparison can constitute a valuable reference when attempting to understand the design principles that underlie PPI networks. Interestingly, PPI networks could be thought of as a type of communication networks since the protein interactions implicitly convey information on biological processes. Clearly, the building process behind the two networks is very different. PPI networks resulted as a byproduct of processes at the evolutionary scale and are constrained by the laws of physics and chemistry. The Internet was built to optimize the communication efficiency through a decentralized process and under the constraints imposed by technological, geographical, social and economical factors.

In recent years, several research groups have studied large complex systems and their topologies, from social networks to the structure of the web. In what follows, we provide a quick overview of the most related previous work on PPI and Internet topologies. The rapid developing theoretical models for complex networks, such as the ER random model [3], and the small-world [4], scale-free [5] and hierarchical network models [6], have greatly influenced the analysis of the topology of complex biological networks (see, e.g., Refs 7, 8 and, 9). PPI networks have

*Corresponding author.

been characterized as scale-free networks that follow a power-law degree distribution with a sharp cutoff for large degrees [7]. Recently, it has been shown that PPI networks show hierarchical organization [9]. The literature on the analysis of the Internet topology[a] is even richer than the one for PPI networks. A recent study provides a good overview of this body of work [10]. The study in this field was jump-started in 1999, when Faloutsos *et al.* [5] used power-laws to characterize the degree distribution of the AS-level Internet topology. It has also been argued that the Internet topology is organized with a natural semantic proximity, such as geography or business interests [11], and exhibits a hierarchical structure [12].

The contribution of this work is an extensive topological comparison of PPI and Internet networks. On the one hand, both network types exhibit some similar properties, such as skewed degree distribution. On the other hand, the networks have been built by completely different processes, over a very different time-scale, and to optimize different criteria. Our study uses the most important and diverse graph metrics that have been proposed and used in a wide range of studies in multiple disciplines. To our knowledge, this is the first such extensive study of these two types of networks.

We classified the results of our study in six categories, namely, (1) connectivity, (2) small-world, (3) modular/hierarchical organization, (4) entropy, (5) communication efficiency, and (6) robustness. Some of our findings are somewhat surprising, and are discussed in Section 4 and summarized in Section 5. We speculate that the differences found by our study can be attributed to the distinctive objectives and constraints that the two types of networks are supposed to satisfy. We conjecture that the goals are robustness and survivability for cellular networks and communication efficiency in man-made networks.

2. NOTATIONS AND METRICS

First, we briefly review all the metrics used in this study. Formally, a *graph metric* is a function $M : \mathcal{G} \to \mathbb{R}^t$, where $\mathcal{G}$ is the space of all possible graphs and t is a positive integer.

2.1. Connectivity

In the domain of connectivity metrics, we selected average degree and the degree distribution to measure the global connectivity, and the rich club connectivity [13] to measure the core connectivity.

Definition 2.1. The *average degree* of a graph $G = (V, E)$ is defined as $\bar{k} = 2m/n$, where $n = |V|$ and $m = |E|$.

Definition 2.2. The *degree distribution* of a graph $G = (V, E)$ is a function $P : [0, \ldots, k_{\max}] \to [0, 1]$, where $P(k)$ is the fraction of the vertices in G that have degree k for $0 \leq k \leq k_{\max}$, and $k_{\max}$ is the largest degree in G.

High degree vertices play an essential role in communication networks. They carry most of the communication flow and together form the *backbone* of the network, which is also referred to as the *core* of the network. We used *rich club connectivity* to measure how densely connected are high degree vertices in the network [13].

Definition 2.3. The *rich club connectivity* of a graph $G = (V, E)$ is a function $\phi : 2^V \to \mathbb{R}$ defined as follows

$$\phi(\rho) = \frac{|\{(u, v) \in E : u \in \rho, v \in \rho\}|}{|\rho|(|\rho| - 1)/2}, \qquad (1)$$

where ρ is the set containing the first $|\rho|$ highest degree vertices in the list of vertices ranked according to their degree in non-increasing order.

2.2. Small-World metrics

The small-world hypothesis states that everyone in the world can be reached through a short chain of social acquaintances. According to Watts and Strogatz [4], a small-world network is mainly characterized by two structural properties, namely, (1) a shorter

[a]Note that we study the Internet topology at the *AS-level*, which is defined as follows. The Internet consists of a large number of independently managed networks, which we call *Autonomous Systems* (AS). For example, an Internet Service Provider of a large company network constitutes usually an AS. In the AS-level graph, the vertices are the Autonomous Systems and an edge represents the fact that the two adjacent nodes are physically connected and exchange information in the form of packets. In this work, we use the term *Internet* or *Internet topology* to refer to the AS-level Internet graph.

characteristic path length and (2) a higher clustering coefficient when compared to random networks.

Definition 2.4. Given a graph $G = (V, E)$, the *characteristic path length* L of G is defined as $L = (\sum_{u,v \in V} L(u,v))/[n(n-1)/2]$, where $L(u,v)$ is the shortest path length between vertex u and v.

Definition 2.5. The *clustering coefficient* $C(v)$ of a vertex $v \in V$ is defined as

$$C(v) = \frac{|\{(s,t) \in E : (v,s) \in E, (v,t) \in E\}|}{d(v)(d(v)-1)/2}, \quad (2)$$

where $d(v) > 1$ is the degree of vertex v. The *clustering coefficient* C of a graph $G = (V, E)$ is defined as $C = (\sum_{v \in V} C(v))/n$.

2.3. Modular and Hierarchical Organization

The modular and hierarchical structures in networks can be quantified by the scaling relation between the clustering coefficient C_k and the vertex degree k, where $C_k = (\sum_{d(v)=k} C(v))/N(k)$ and $N(k)$ is the number of vertices having degree k. Several studies [6, 14] have shown that if a network has modular and hierarchical structure, the distribution of C_k is power-law-like, that is $C_k \sim k^{-\alpha}$ for some real positive α.

2.4. Entropy

We selected graph entropy [15] and target entropy [16] to evaluate the randomness of a graph. Let X and Y be two discrete random variables associated with the degree of the two vertices of a randomly chosen edge.

Definition 2.6. The *graph entropy* $E(G)$ of a graph $G = (V, E)$ is defined as follows

$$E(G) = H(X) + H(Y) - H(X,Y), \quad (3)$$

where $H(X)$ and $H(Y)$ are the entropy of random variable X and Y, and $H(X,Y)$ is the joint entropy of X and Y (as defined e.g., in Ref 17).

The graph entropy $E(G)$ corresponds to the mutual information between random variable X and Y, which measures the amount of information that one random variable contains about another random variable. The mutual information quantifies the reduction in the uncertainty of one random variable due to the knowledge of the other [17].

Our second metric of randomness is *target entropy*, which measures the predictability of the amount of traffic in the neighborhood of any given vertex [16]. More specifically, assume that every vertex in the network sends one unit flow to vertex u using shortest path. Let $c(u,v)$ denote the fraction of the flows with destination u that passes through vertex v, where v is the immediate neighbor of u.

Definition 2.7. The *target entropy* $T(u)$ of a vertex $u \in V$ is defined as follows

$$T(u) = -\sum_{v \text{ neighbor of } u} c(u,v) \log_2 c(u,v). \quad (4)$$

The *target entropy* T of a graph $G = (V, E)$ is defined as $T = (\sum_{u \in V} T(u))/n$.

2.5. Performance Measures

We selected two metrics to evaluate the performance of the network, namely eccentricity and edge congestion. The former metric is related to the notion of reachability of a graph [18], whereas the latter measures the congestion on the edges assuming a flow model.

Definition 2.8. The *eccentricity* $\varepsilon(u)$ of a vertex $u \in V$ is defined as $\varepsilon(u) = \max_{v \in V} L(u,v)$.

Edge congestion measures the amount of flows traveling through the edges of a network assuming a given traffic model and routing policy. In this study, we assume that one unit of flow between every pair of vertices is routed using the shortest-path routing policy [19, 20].

Definition 2.9. The *edge congestion* $e_c(u,v)$ of an edge $(u,v) \in E$ is defined as $e_c(u,v) = f(u,v)/[n(n-1)]$, where $f(u,v)$ denotes the total number of flows traveling through the edge (u,v).

2.6. Robustness

We selected two simple methods to measure the robustness of network topology under failures. In the first, we remove vertices at random, which corresponds to random failures. In the second, we remove vertices in the order of decreasing degree, which corresponds to "intelligent attacks" [21]. In both cases, the network eventually gets decomposed into a set of connected components. To characterize this process, we measured (a) $L_c = |S|/n$, where S is the largest connected component, and (b) N_c, the number of components in the network.

Table 1. Statistic summary: n is the number of vertices, m is the number of edges, $\overline{k}$ is the average degree, L is the characteristic path length, and C is the clustering coefficient.

	Yeast	Fly	AS990220	skitter
n	4687	6926	4686	9200
m	15138	20745	8772	28957
$\overline{k}$	6.4596	5.9905	3.7439	6.2927
L	4.18519	4.45931	3.72621	3.118
C	0.126	0.0154	0.3786	0.6212

3. DATASETS

In our work, we used four networks whose global statistics are summarized in Table 1. `Yeast` and `Fly` are two PPI networks downloaded from the DIP database [22], in which vertices represent proteins and edges represent physical interactions between pairs of proteins. `AS990220` and `Skitter` are two AS-level Internet instances obtained from two different methods[b].

We also employed two models of random graphs, namely, $G(n,p)$ and degree-based random graphs. A $G(n,p)$ random graph is a graph composed of n vertices where each pair of vertices is connected with probability[3] p . Given a degree distribution d and an integer n, a *degree-based random graph* (DBRG) is a graph with n vertices where vertices u, v are connected with probability proportional to the product of their degree[26] $d(u)d(v)$. In the following, $G(n,p)$ random graphs were generated based on the same number of vertices and edges as in the real networks. DBRG random graphs were produced based on the same degree distribution of the real networks along with the same number of vertices and edges as in real networks.

4. RESULTS AND DISCUSSION

4.1. Connectivity

Average Degree. Table 1 summarizes the average degree of the four networks. `Yeast` and `Fly` PPI networks have an average degree around 6. `AS990220` has approximately the same number of vertices as `Yeast` PPI network, but its average degree is much lower (≈ 4). The other AS-level Internet `Skitter` has an average degree of about 6, much closer to the two PPI networks.

Skewed Degree Distribution. Figure 1 shows the complementary cumulative density function (CCDF) of degree distribution of the four networks. The two Internet networks show a "perfect" power-law degree distribution with $\gamma \approx 1.1$. Observe that although the degree distribution of the two PPI networks is highly skewed, they do not follow closely a power-law distribution ($\gamma \approx 1.7$).

The degree distribution of PPI networks has been characterized as *truncated scale-free*, which has a power-law regime followed by a sharp cutoff, like an exponential or Gaussian decay of the tail [27]. In Ref 27, the authors showed that they could generate networks with such degree distribution by imposing constraints on the process that adds new links to vertices. We speculate that such constraints could potentially exist in the evolutionary process that shaped the topology of PPI networks. For example, one constraint is the physical and chemical limitations on the number of interacting partners that a protein could possibly have. Moreover, compartmentalization and inherent functional modular organization of various components inside the cell would restrict spatially and functionally the number of links added between two different compartments or two different functional modules. Although it is not clear what is the evolutionary advantage for PPI networks to have scale-free topology, we argue that the physical, chemical and thermodynamic constraints in the cell could account for the lack of a perfect scale-free topology in PPI networks.

[b] `AS990220` is an AS-level topology collected by the Oregon routeviews project [23], which extracts the information from BGP routing updates [24]. `Skitter` was collected by CAIDA (Cooperative Association for Internet Data Analysis) using traceroute and then carefully processed [10, 25].

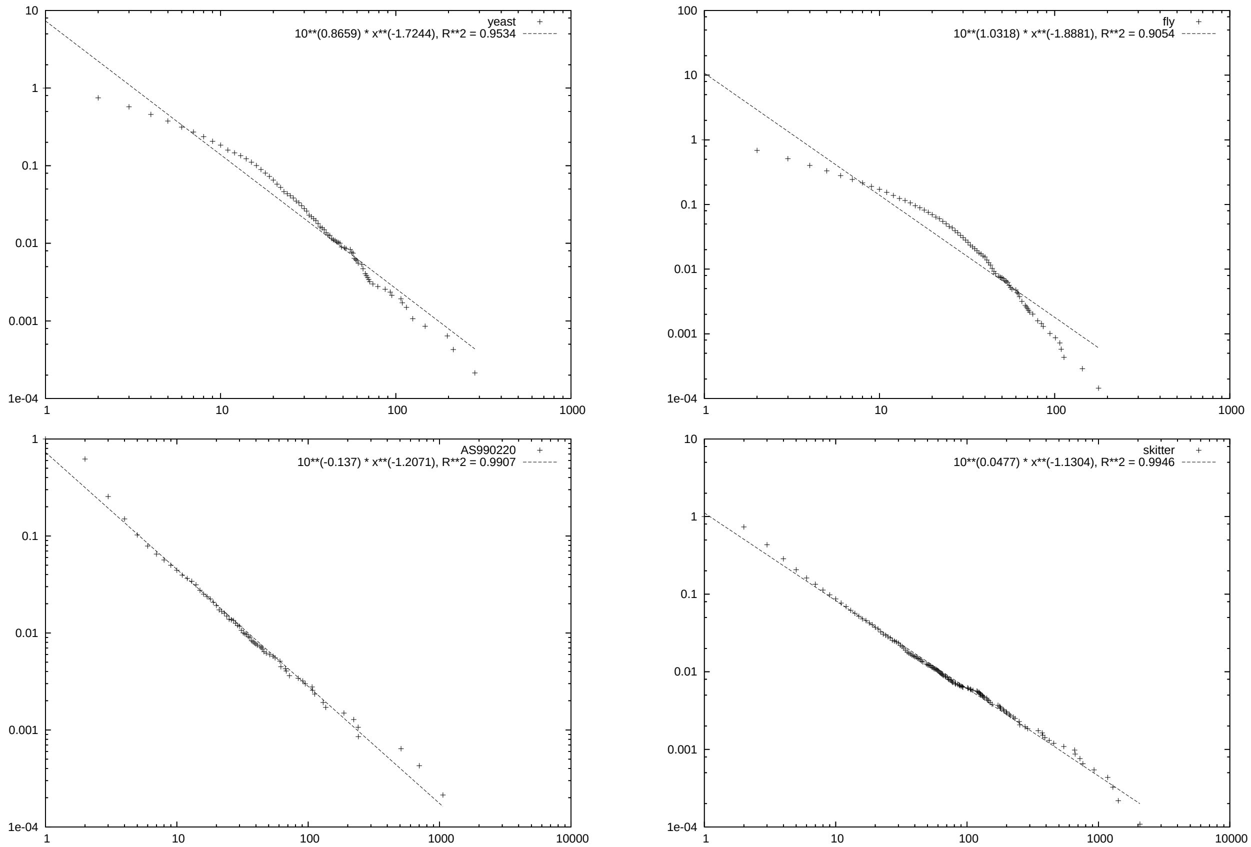

Fig. 1. Complementary cumulative density function (CCDF) of the degree distribution.

Rich club connectivity. Figure 4 shows that about 10% of the vertices with the highest degree in AS-level Internet are more densely connected with each other than those in PPI networks. In another words, links between high degree vertices in PPI networks are suppressed, which is consistent with previous observations [28]. A comparison with $G(n, p)$ and DBRG random graphs further illustrates that the number of links between high degree vertices in PPI networks is significantly lower than expected. In contrast, the number of links connecting high degree vertices in AS-level Internet matches the expected number observed in their corresponding random networks (data not shown).

The core connectivity analysis shows that high degree vertices in PPI networks do not connect with each other as much as in the Internet or when compared to random networks. This feature is consistent with the theory of functional modular organization of the cell. Functional modules can be insulated from or connected to each other. Insulation allows the cell to carry out many diverse reactions without the cross-talks that would harm the cell, whereas connectivity allows one function to influence another one. The most notable effect of suppressing the connections between high degree vertices is to prevent the deleterious perturbations from propagating rapidly over the network through densely connected high degree vertices. In contrast, such concern does not exist in Internet, in which high degree vertices (i.e., large Internet Service Providers) are expected to connect to each other to promote communication between different cities, countries and continents.

4.2. Small-world metrics

Characteristic Path Length. Table 1 summarizes the characteristic path length for the four networks. The two AS-level Internet have an average shortest path length almost one hop shorter than that of two PPI networks. This indicates that on average it takes fewer edges to reach one another using shortest path in Internet than in PPI networks.

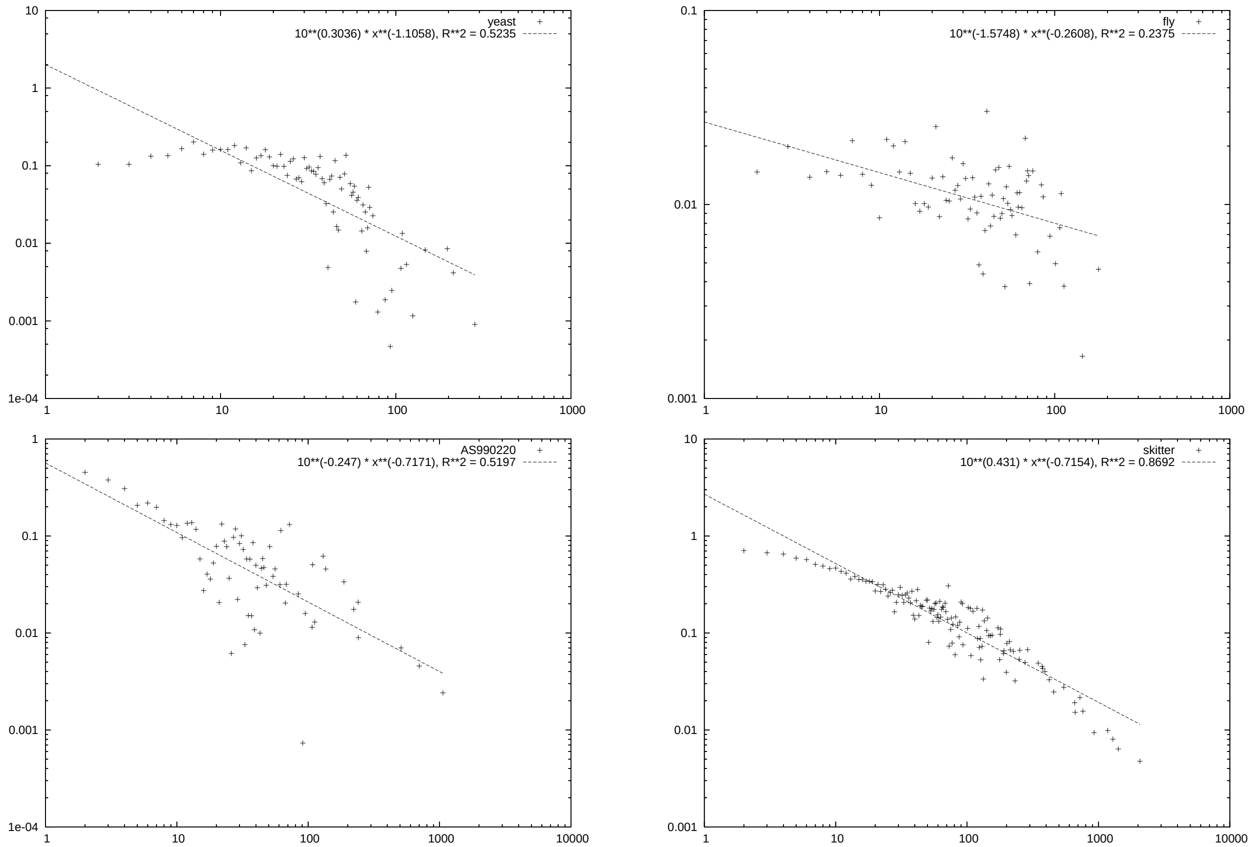

Fig. 2. Clustering coefficient C_k as a function of the degree k.

Clustering Coefficient. The clustering coefficient C is also shown in Table 1. The two AS-level Internet have a much higher clustering coefficient than that of two PPI networks, indicating that the neighboring vertices in Internet are well connected when compared to PPI networks. This may also imply that prominent cluster structures exist in the Internet [6, 9].

The fact that Internet networks have a shorter characteristic path length and a higher clustering coefficient than PPI networks indicates that the former graphs are more small-world than PPI networks. This results imply that the overall design of the Internet promotes well-connected neighborhood and can propagate messages through a short chain between long distance vertices, which in turn implies more efficient communication. This conclusion is rather expected and not surprising. The primary goal of the Internet is to deliver messages in a fast and reliable manner, thus the small-world properties are more desirable in such type of networks. In contrast, fast communication is perhaps not one of the primary concerns in PPI networks.

4.3. Modular and Hierarchical Organization

Figure 2 shows the clustering coefficient C_k versus degree k. The two AS-level Internet show a power-law-like distribution of $C_k \sim k^{-\alpha}$, which is an indication of hierarchical organization. In PPI networks, however, there is almost no indication of hierarchical structures, except perhaps among high degree vertices, where we detected weak signs of hierarchical structure.

In addition to the analysis of the scaling relation between C_k and k, we designed another experiment to explicitly demonstrate that the combination of high clustering coefficient and the presence of the relation $C_k \sim k^{-\alpha}$ results in modular and hierarchical structures in AS-level Internet. The experiment was designed as follows. (1) Iteratively remove the most *critical* edge (for example, we removed the edge with

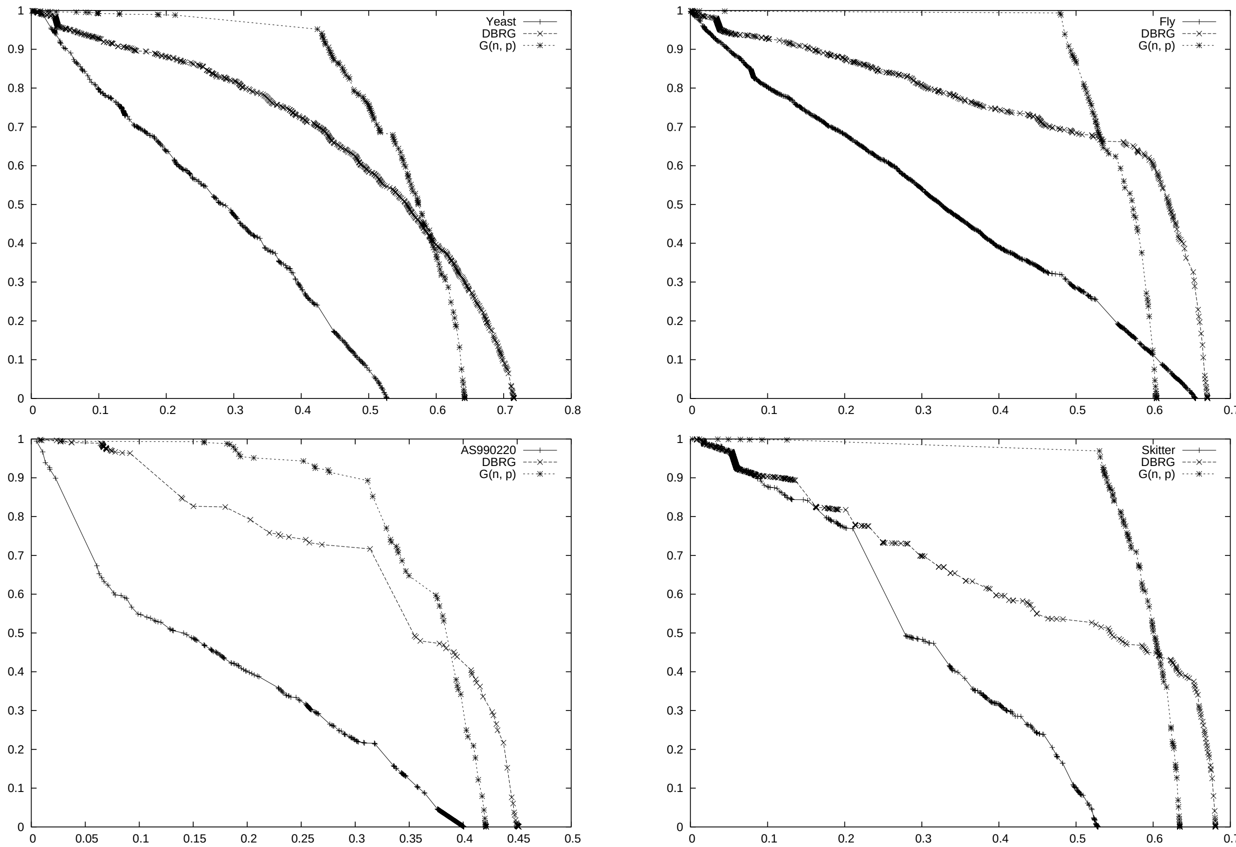

Fig. 3. The size L_c of the largest normalized component (represented on the y-axis) in real networks and the corresponding $G(n, p)$ and DBRG random networks after successive removal of critical edges. The x-axis represents the fraction of edges removed.

the highest betweenness [29]) in the network, until the network breaks into two components, (2) Measure the size L_c of the largest normalized connected component, (3) Repeat step (1) and (2) on the largest component until its size reaches one node. The rationale behind this decomposition process is that if the network is modular and hierarchically organized, then one expects the decomposition to separate large components from the network.

Figure 3 shows the comparison between the real networks and their corresponding $G(n, p)$ and DBRG random networks. The graph illustrates that L_c decays much faster in four real networks than that in their two random counterparts, which in turn indicates that the four real networks are more modular and hierarchically organized than random networks. Figure 5 shows a comparison of L_c between the four real networks. Observe that the size of the largest component in the AS-level Internet decays faster than that in the PPI networks. Although the decay rate of L_c in `Skitter` is comparable to that of two PPI networks, in `Skitter`, larger size components are separated from the network, indicating a much stronger modularity.

The measurements on the scaling relation and the decomposition experiment suggest that modular and hierarchical structures exist in all networks examined. Moreover, the topology of the Internet is significantly more modular and hierarchical than that of PPI networks. In fact, hierarchical organization is inherent in the topology of the Internet, which mirrors the hierarchical structure in business relationships. It is a well known fact that on the Internet there are a few tens of vertices that provide international world-wide connectivity and they practically form a clique. Then, within each country we have national, regional and local Internet Service Providers (ISPs). Typically, the smaller ISPs are the customers of the larger ISPs. This hierarchical structure which emerges as a reflection of business policies, is not a strict hierarchy but it definitely provides

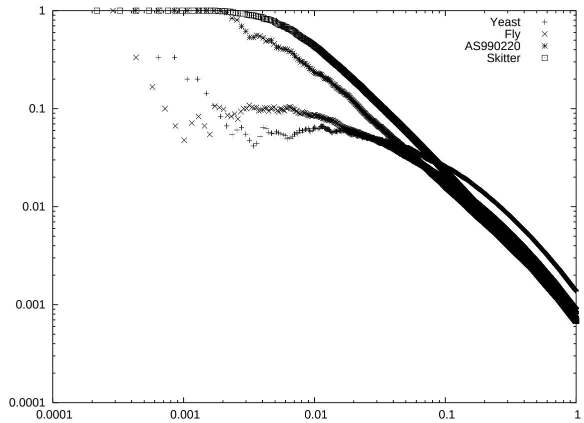

Fig. 4. Rich club connectivity $\phi(\rho)$ as a function of ρ.

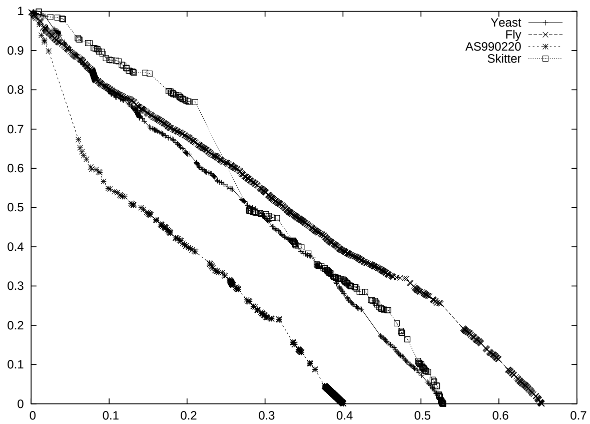

Fig. 5. L_c as a function of the fraction of edges removed.

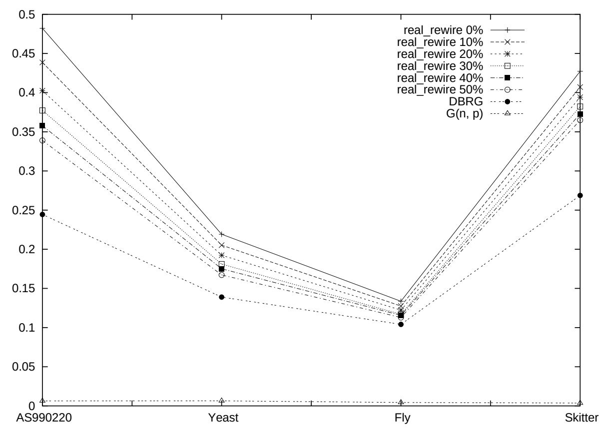

Fig. 6. Graph entropy $E(G)$ of the networks under study.

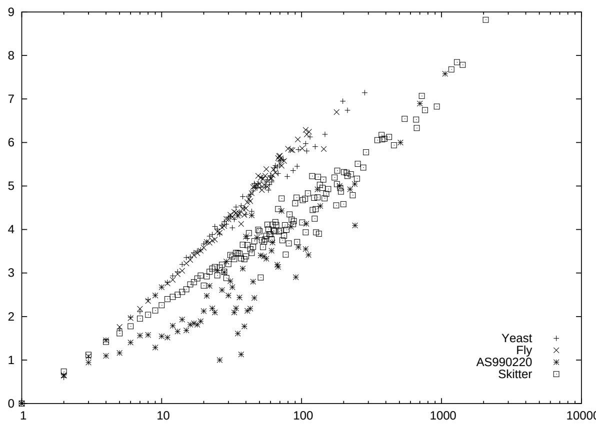

Fig. 7. Target entropy T_k as a function of the degree k.

a topological structure.

4.4. Entropy

Graph Entropy. Figure 6 illustrates the graph entropy for real networks and their corresponding $G(n,p)$ and DBRG random networks. The figure also shows the graph entropy for rewired networks, in which 10, 20, 30, 40, 50% of all edges are rewired[c] (both in real and random networks). The data points for random and rewired networks were averaged over 10 corresponding networks. Observe that the entropy of all real networks approaches the corresponding DBRG random networks with more and more rewiring. In contrast, the entropy of random networks remains almost the same regardless of how much rewiring was performed (data not shown). Since the vertices in $G(n,p)$ random networks are connected uniformly at random, their graph entropy is expected to be zero, as shown in Figure 6. These observations show that graph entropy reflects the randomness of a network in a quantitative way, since the value of graph entropy varies accordingly with the randomness of a network.

The figure also shows that the two AS-level Internet have a much higher graph entropy than that of the two PPI networks. This result implies that on the Internet we can observe a regular connection pattern between different classes of vertices, e.g., low or high degree vertices. The analysis of the *assortativity coefficient* [30] confirms this result by showing preferential connections between low and high degree vertices in Internet networks (data not shown).

The analysis of graph entropy shows that the connectivity between vertices with different degrees

[c]Rewiring is a process that randomly switches the edges in the network in such a way that the degree distribution of the nodes in the networks remains constant [4].

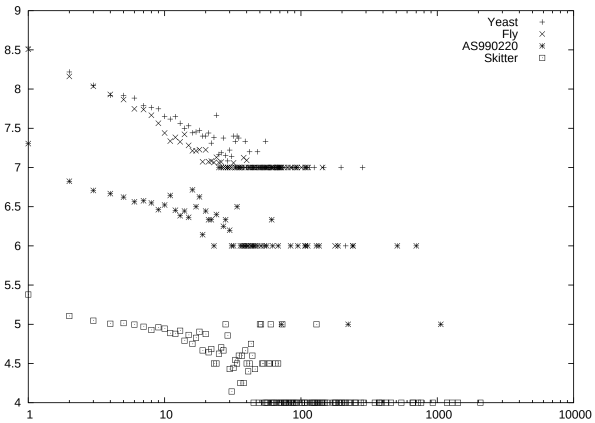

Fig. 8. Eccentricity ε_k as a function of the degree k.

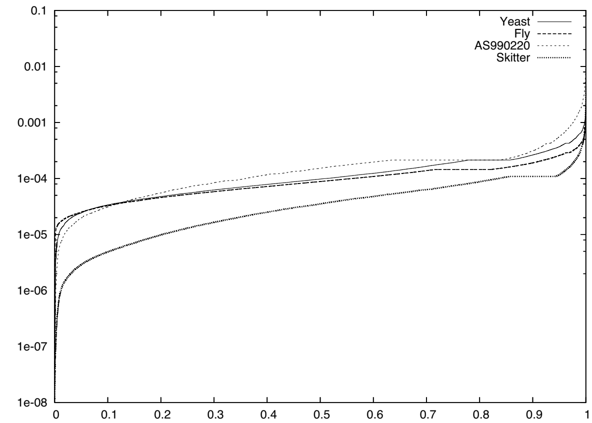

Fig. 9. Distribution of edge congestion.

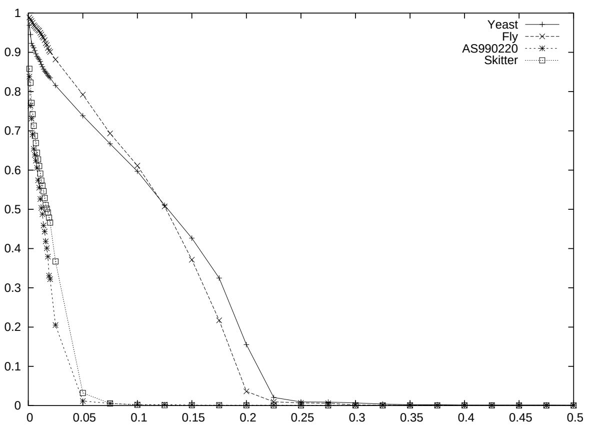

Fig. 10. L_c as a function of the fraction of vertices removed.

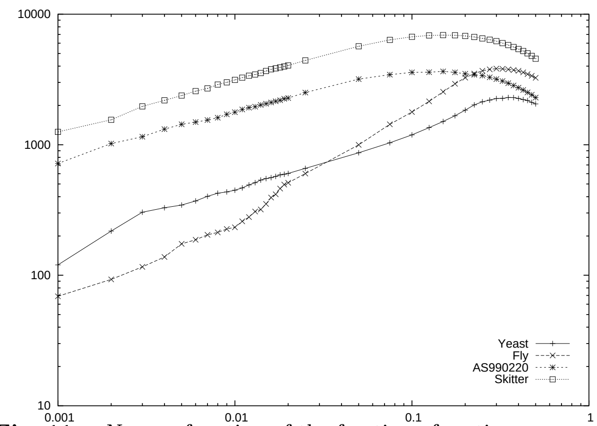

Fig. 11. N_c as a function of the fraction of vertices removed.

in PPI networks is close to uniformly random, which is analogous to the notion of *diversification* in evolution. Diversification is a process in which multiple phenotypes and genotypes are simultaneously present in a population, which increases the probability that some individuals will survive and reproduce in a heterogeneous and changing environment. The same mechanism could possibly be adopted in the building process of PPI networks too, since the heterogeneous connectivity pattern could potentially increase the robustness of the network by redundancy or/and degeneracy mechanism [31].

Target Entropy. Recall that the target entropy of a vertex u is the entropy of the distribution of the number of times a vertex in the neighborhood of u is traversed to route messages. The closer is the distribution to a uniform distribution, the higher is its entropy. Figure 7 shows T_k versus degree k, where $T_k = (\sum_{d(u)=k} T(u))/N(k)$ and $N(k)$ is the number of vertices having degree k. The figure shows that the vertices in the Internet are less uniform at choosing neighbors to route messages when compared to PPI networks. This is due to the fact that all the routing domains in the Internet have to visit some large ISPs in order to exchange information with other administrative domains. Therefore, the choice of which vertex to use for routing is highly selective in Internet. Since such constraints do not exist in PPI networks, the vertices have more freedom to choose which vertex to pass the message to.

The results presented here on the target entropy appear to somewhat contradict the ones obtained by Sneppen *et al.* in Ref 16. In their work, the authors show that the Internet has higher target entropy than the PPI network of Yeast and Fly, and that the PPI of Fly has higher target entropy than the one of Yeast. Our results show that the Internet has lower target entropy than the two PPI networks and that the PPI of Fly has lower target entropy than that of Yeast. This discrepancy is likely to be explained by the fact that they used much smaller networks than the ones we used in our study. In addition, the authors of Ref 16 do not report the average degree of the networks used in their dataset. Comparing the target entropy of networks with very different average degrees might not be very meaningful.

4.5. Performance Measure

Our main measure of performance is the communication efficiency. Many factors influence the performance of the Internet, such as routing policy, traffic flow, etc. Since for PPI networks the notions of routing and traffic might not be very meaningful, we used the simplest model for both types of networks, namely we modeled packets using one unit of flow between every pair of vertices where packets are routed using shortest path policy. Under these assumption, our goal was to determine whether PPI networks have any advantages over the Internet as communication networks. We measured eccentricity, which estimates how quickly one vertex can reach any vertex in the network, and edge congestion, which is related to the traffic flow in the network.

Eccentricity. Figure 8 shows the quantity ε_k versus degree k, where $\varepsilon_k = (\sum_{d(u)=k} \varepsilon(u))/N(k)$ is the average eccentricity for the vertices having the same degree k. The figure shows that on average the vertices in Internet reach the rest of the vertices in the network with fewer hops than those in PPI networks.

Edge Congestion. The edge congestion for all edges in the network was sorted into non-decreasing order and the distribution of the congestion of each edge was plotted in Figure 9. The figure shows that most of the edges in `Skitter` have less flow traffic than the other networks. Although the edge congestion in `AS990220` is comparable to that of the two PPI networks, we need to recall that `AS990220` has an average degree of 3.74 in contrast to an average degree of about 6 in PPI networks. In other words, `AS990220` achieves the same level of edge congestion as the PPI networks with a significantly smaller number of edges. This indicates that the topology of the Internet has inherent structural properties that tend to reduce edge congestion.

The performance analysis suggests that the Internet is highly optimized for communication efficiency (under the assumptions we made on the routing and the traffic). In contrast, it appears that PPI networks are not optimized to route messages and minimize traffic.

4.6. Robustness

When vertices were randomly removed from the network, along with all incident edges, all four networks behave similarly in terms of the size of the largest normalized component L_c and the number of components N_c (data not shown). However, when we targeted first vertices with the highest degrees, AS-level Internet collapses much faster than PPI networks as shown by a smaller L_c and a larger N_c (see Figure 10 and 11). The results indicate that high degree vertices play a critical role in Internet. In contrast, the fact that high degree vertices in PPI networks tend not to be connected with each other (as shown by the rich club connectivity analysis) prevents the deleterious effects, such as gene knockout or protein malfunction, from spreading throughout the network too fast. Thus, suppressed cross-talking between high degree vertices in PPI networks clearly contributes to the robustness of the network by localizing the effects of deleterious perturbations. The distinction between the robustness of these two types of networks supports the idea that the underlying driving force that shape the topology of these two types of networks are distinct, namely, for PPI networks, the survivability of the cell favored by evolution, and for Internet, the optimized communication requirements.

5. CONCLUSION

In this paper we showed that by comparing PPI networks to AS-level Internet, one can possibly gain some insights on the topological properties and the design principles underlying the two types of networks. Such cross-disciplinary comparison brings together tools, expertise, and ideas from different communities, and benefits both research areas.

Our results suggest that although both types of networks have been characterized as scale-free topologies, they also exhibit non-trivial topological differences.

- *Connectivity.* The Internet has a highly-connected "core", which does not appear to exist in PPI networks.
- *Small-world.* The Internet topology exhibits stronger small-world properties than PPI networks.
- *Modular/Hierarchical organization.* The Internet topology shows a more prominent modular and hierarchical organization than PPI networks.
- *Entropy.* Vertices with different degrees are more uniformly connected in PPI networks than those in the Internet.
- *Communication efficiency.* The Internet topology is more efficient in routing messages and minimizing traffic than PPI networks with respect to the metrics that capture the communication efficiency.
- *Robustness.* The Internet and PPI networks seem equally robust against random failures. However, PPI networks are more robust under "targeted" (e.g., toward high degree nodes) attacks compared to the Internet.

We speculate that the structural and functional differences between PPI networks and AS-level Internet are originated from different constraints and objectives that govern and shape the building process of these complex networks. Specifically, the building process of PPI networks is driven by the evolutionary constraints that favor the survivability and diversification. In contrast, the architecture of Internet is built under the needs of fast and reliable communication.

ACKNOWLEDGMENTS

This project was supported in part by NSF CAREER IIS-0447773 and NSF DBI-0321756. We thank Dr. M. Newman for providing us the implementation of the edge betweenness clustering algorithm and for helpful discussions. We also want to thank the anonymous reviewers that helped improve the quality of this manuscript.

References

1. Uetz P, Giot L, *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae. Nature* 2000; **403**: 623–627.
2. Giot L, Bader J, *et al.* A protein interaction map of *Drosophila melanogaster. Science* 2003; **302**: 1727–1736.
3. Erdös P, Rényi A. On random graphs. *Publicationes Mathematicae* 1959; **6**: 290–297.
4. Watts D, Strogatz S. Collective dynamics of 'small-world' networks. *Nature* 1998; **393**: 440–442.
5. Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the Internet topology. *Proc. of ACM SIGCOMM* 1999: 251–263.
6. Ravasz E, Barabási A-L. Hierarchical organization in complex networks. *Phys. Rev. E* 2003; **67**: 026112.
7. Jeong H, Mason S, *et al.* Lethality and centrality in protein networks. *Nature* 2001; **411**: 41–42.
8. Jeong H, Tombor B, *et al.* The large-scale organization of metabolic networks. *Nature* 2000; **407**: 651–654.
9. Barabási A-L, Deszo Z, *et al.* Scale-free and hierarchical structures in complex networks. *Seventh Granada Lectures on Modeling of Complex Systems* 2002.
10. Mahadevan P, Krioukov D, *et al.* The Internet AS-level topology: three data sources and one definitive metric. *ACM SIGCOMM Computer Communication Review* 2006; **36**: 17–26.
11. Gkantsidis C, Mihail M, *et al.* Spectral analysis of Internet topologies. *Proc. of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'03)* 2003: 364–374.
12. Jaiswal S, Rosenberg A, *et al.* Comparing the structure of power-law graphs and the Internet AS graph. *Proc. of the 12th IEEE International Conference on Network Protocols (ICNP'04)* 2004: 294–303.
13. Zhou S, Mondragon R. Accurately modeling the internet topology. *Phys. Rev. E* 2004; **70**: 066108.
14. Dorogovtsev S, Goltsev A, *et al.* Pseudofractal scale-free web. *Phys. Rev. E* 2002; **65**: 066122.
15. Mahadevan P, Krioukov D, *et al.* Comparative analysis of the Internet AS-level topologies extracted from different data sources. `http://www.krioukov.net/~dima/pub/`.
16. Sneppen K, Trusina A, *et al.* Hide-and-seek on complex networks. *Europhys. Lett.* 2005; **69**: 853–859.
17. Cover T, Thomas J. Elements of information theory. John Wiley & Sons, Inc. New York, NY. 1991.
18. Harary F. Graph theory. Addison-Wesley publishing company, Reading, MA. 1994.
19. Akella A, Chawla S, *et al.* Scaling properties of the Internet graph. *Proc. of the Twenty-second Annual Symposium on Principles of Distributed Computing* 2003: 337–346.
20. Gkantsidis C, Mihail M, *et al.* Conductance and congestion in power law graphs. *Proc. of the 2003 ACM SIGMETRICS international conference on Measurement and modeling of computer systems* 2003: 148–159.
21. Albert R, Jeong H, *et al.* Error and attack tolerance

of complex networks. *Nature* 2000; **406**: 378–382.

22. Xenarios L, Salwinski L, *et al.* DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Research* 2002; **30**: 303–305.
23. University of Oregon Route Views Project. Online data and reports. `http://www.routeviews.org/`.
24. Rekhter Y, Li T(Eds). A Border Gateway Protocol 4 (BGP-4). 1995.
25. Mao Z, Rexford J, *et al.* Towards an accurate AS-level traceroute tool. *Proc. of ACM SIGCOMM* 2003: 365–378.
26. Aiello W, Chung F, *et al.* A random graph model for massive graphs. *Proc. of the Thirty-second Annual ACM Symposium on Theory of Computing* 2000: 171–180.
27. Amaral L, Scala A, *et al.* Classes of small-world networks. *Proc. Natl. Acad. Sci.* 2000; **97**: 11149–11152.
28. Maslov S, Sneppen K. Specificity and stability in topology of protein networks. *Science* 2002; **296**: 910–913.
29. Girvan M, Newman M. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* 2002; **99**: 7821–7826.
30. Newman M. Assortative mixing in networks. *Phys. Rev. Lett.* 2002; **89**: 208701.
31. Tononidagger G, Sporns O, *et al.* Measures of degeneracy and redundancy in biological networks. *Proc. Natl. Acad. Sci.* 1999; **96**: 3257–3262.