

High-Betweenness Proteins in the Yeast Protein Interaction Network

Maliackal Poulo Joy, Amy Brock, Donald E. Ingber, and Sui Huang

Vascular Biology Program, Departments of Surgery and Pathology, Children's Hospital, Harvard Medical School, Boston, MA 02115, USA

Received 1 June 2004; accepted 13 August 2004

Structural features found in biomolecular networks that are absent in random networks produced by simple algorithms can provide insight into the function and evolution of cell regulatory networks. Here we analyze “betweenness” of network nodes, a graph theoretical centrality measure, in the yeast protein interaction network. Proteins that have high betweenness, but low connectivity (degree), were found to be abundant in the yeast proteome. This finding is not explained by algorithms proposed to explain the scale-free property of protein interaction networks, where low-connectivity proteins also have low betweenness. These data suggest the existence of some modular organization of the network, and that the high-betweenness, low-connectivity proteins may act as important links between these modules. We found that proteins with high betweenness are more likely to be essential and that evolutionary age of proteins is positively correlated with betweenness. By comparing different models of genome evolution that generate scale-free networks, we show that rewiring of interactions via mutation is an important factor in the production of such proteins. The evolutionary and functional significance of these observations are discussed.

INTRODUCTION

The availability of genome-scale databases of pairwise protein interactions data in yeast [1] has made it possible to analyze the structure of the entire protein interaction network (PIN) in light of concepts from graph theory and the study of complex networks [2]. In these models of cell regulatory networks, proteins are represented by the nodes and the interactions between these components by the edges of the graph. Such genome-scale analysis of the PIN revealed that these molecular components form a “genome-wide” network, that is, the largest connected network component (“giant component”) encompasses a dominant portion of the proteome. The large-scale topology (architecture) of this genome-wide PIN exhibits several interesting features that distinguish it from an Erdos-Renyi (ER) random graph [3]. For instance, the distribution of the connectivity (or degree, as used in graph theory) k which refers to the number of first neighbors

of a given node approximates a power law, or, in other words, the PIN may be a scale-free network. PIN contains a larger number of highly connected proteins (hubs) than one would expect to find in an ER random network [4]. The connectivity of a protein appears to be positively correlated with its essentiality [4] in that highly connected proteins tend to be more essential for the viability of the organism.

Barabasi and Albert [5] proposed a simple algorithm for network growth (BA model) in which incoming nodes (newly evolved proteins) attach preferentially to existing nodes with higher degree. However, the yeast PIN exhibits additional structural details not observed in these randomly generated, scale-free networks. For instance, there are correlations between the connectivities of directly interacting proteins, in that connections between hubs are almost entirely absent. This feature has been postulated to be partly responsible for the robustness of biological networks [6]. Some specific local network structures, so-called network motifs, have also been shown to occur more frequently in molecular networks than in random networks [7]. Another structural feature of biological systems is their modularity, for example, the metabolic network exhibits a hierarchical modular structure [8].

In contrast to the genome-scale perspective, characterization of the biological functions of proteins has traditionally assumed the existence of distinct signaling modules that can be associated with particular cellular functions [9]. Hence, much effort has been spent

Correspondence and reprint requests to Sui Huang, Vascular Biology Program, Departments of Surgery and Pathology, Children's Hospital, Harvard Medical School, Boston, MA 02115, USA, E-mail: sui.huang@childrens.harvard.edu

This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

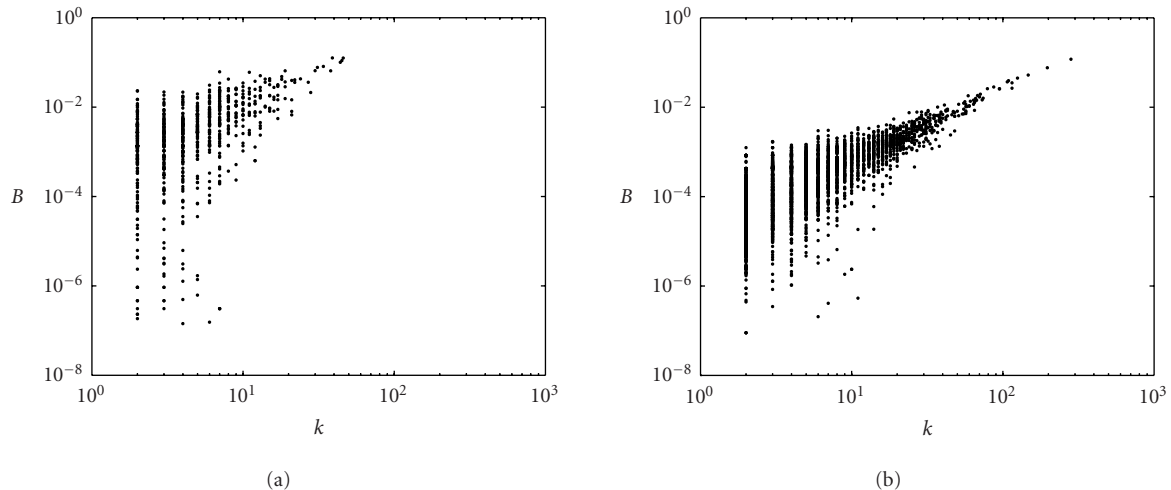


FIGURE 1. Degree (k) versus betweenness (B) plotted in logarithmic scale for the measured yeast interaction network based on DIP data [15, 16]. (a) Core data. (b) Full DIP data.

in defining and identifying discrete, functional network modules within the PIN. However, the ad hoc structural criteria used to define a module in physical networks remain somewhat arbitrary. Here we set out to examine a feature of complex networks that unites local and global topological properties of a node: the betweenness centrality. Measures, such as the connectivity of a node, k , and the clustering coefficients of networks, C [10], used previously to describe global architectural features capture only the local neighborhood of network nodes (nearest neighbors). In contrast, betweenness B_i of a given node i in a network is related to the number of times that node is a member of the set of shortest paths that connect all the pairs of nodes in the network (see “data and methods” for details). Hence, betweenness accounts for direct and indirect influences of proteins at distant network sites and hence it allows one to relate local network structure to global network topology [11]. Betweenness has also been used to characterize the “modularity” (eg, community structure) of various natural and man-made networks (see [12, 13]).

The functional relevance of the betweenness centrality B_i of a node is based on the observation that a node which is located on the shortest path between two other nodes has most influence over the “information transfer” between them. The betweenness distribution $P(B)$ of the nodes in a scale-free network also follows a power law or has a scale-free distribution, $P(B) \sim B^{-\rho}$ [14]. Although the distribution of the connectivity k across the nodes of the network has been used as a measure to characterize natural networks and the value of k has been suggested to correlate with the importance of the protein, this is truly valid only if the immediate neighbors are the only ones determining the properties of a protein in the network. In contrast, betweenness indicates how important the node is within the wider context of the entire network.

Based on analysis of the betweenness measure, we report here a new topological feature in the yeast PIN that is not found in randomly generated scale-free networks: the abundance of proteins characterized by high betweenness, yet low connectivity. The existence of such proteins points to the presence of modularity in the network, and suggests that these proteins may represent important connectors that link these putative modules. We describe here an extended network-generating algorithm that produces networks containing high betweenness nodes with low connectivity. We then discuss the evolutionary and functional significance of these findings.

RESULTS

We studied yeast protein interaction data obtained from different databases [1], including the Database of Interacting Proteins (DIP), and the Munich Information Center for Protein Sequences (MIPS) [15, 16, 17]. Although these networks differ at the level of individual protein-protein interactions, they exhibited the same global statistical properties. Here we present results for the most recent “full” [15] and “core” DIP data. In the core data only confirmed interactions were included [16]. The data set used contains 15 210 interactions between 4721 proteins for the “full” data set, and 6438 interactions among 2605 proteins for the “core” data set.

High-betweenness, low-connectivity proteins

Unlike the connectivity k which ranged from 1 to 282 in the PIN, values for betweenness B ranged over several orders of magnitude. The few highly connected nodes (hubs) in the PIN must have high-betweenness values because there are many nodes directly and exclusively connected to these hubs and the shortest path between these nodes goes through these hubs. However,

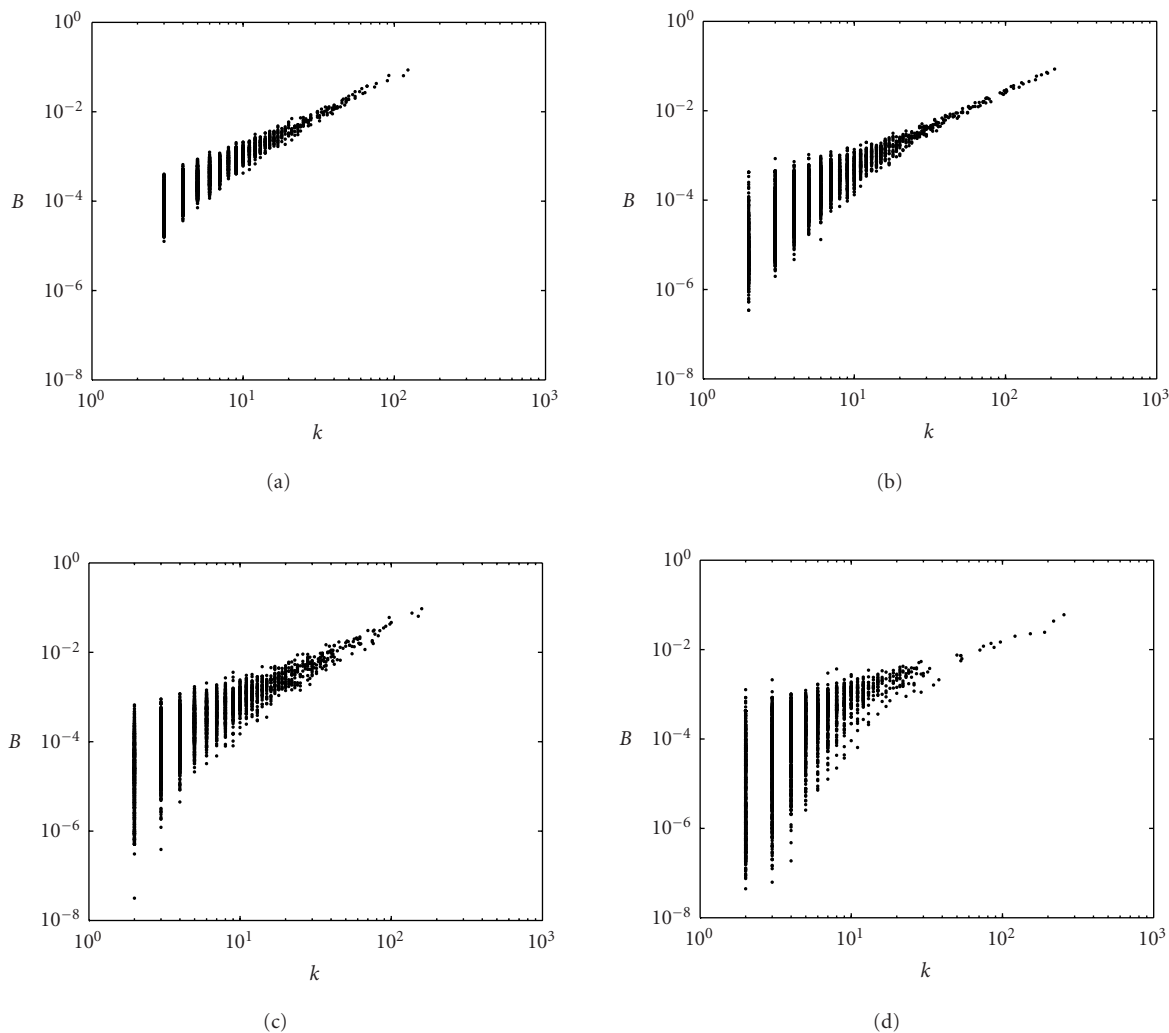


FIGURE 2. The $k - B$ plot for various model generative algorithms. (a) Barabasi-Albert (BA); (b) extended BA (EBA); (c) Sole-Vazquez (SV); (d) duplication mutation (DM).

the low-connectivity nodes also exhibited a wide range of betweenness values in the yeast PIN, as shown in Figure 1a (core data) and in Figure 1b (full data), where betweenness (B) is plotted as a function of connectivity (k). This indicates the existence of a large number of nodes with high betweenness but low connectivity (HBLC nodes). Importantly, such nodes are absent in computer-generated, random scale-free networks [5]. Although the low connectivity of these HBLC proteins would imply that they are unimportant, their high betweenness suggests that these proteins may have a global impact. From a topological point of view, HBLC proteins are positioned to connect regions of high clustering (containing hubs), even though they have low local connectivity.

Models

Can models for network evolution reproduce HBLC behavior? To address this question, we analyzed different computational models of biological network evolu-

tion that generate scale-free networks. The simplest generative algorithm, first proposed by Barabasi and Albert [5] (BA model) to explain the power-law distribution of connectivity, does not predict the existence of HBLC nodes: betweenness and connectivity were almost linearly correlated (Figure 2a). The extended Barabasi-Albert (EBA) model [18], where link addition and rewiring occur along with node addition with preferential attachment, also did not produce networks with HBLC nodes similar to that found in our analysis of the PIN, although low k nodes showed some spread of betweenness (Figure 2b). Moreover, this algorithm has no biological basis. A biologically motivated model put forward by Sole et al [19] and Vazquez et al [20] incorporated “gene duplication” as the driving mechanism for genome growth. In this model, the existing nodes (proteins) are copied with all their existing links, followed by divergence of the duplicated nodes introduced by rewiring and/or addition of connections, imitating mutations of duplicated genes. For the model

parameter range that produces power-law networks, the Sole-Vazquez (SV) model also failed to produce the same bias towards HBLC exhibited by the PIN (Figure 2c).

Berg et al (see [21]) have proposed a model that attempts to capture the actual molecular mechanism of genome growth based on evolutionary data. We asked whether that model can produce HBLC-node-containing networks. For our simulation of network growth, we used a modified version of the Berg model [21] which considered gene duplications and point mutations. “Duplications” relate to the process by which a gene is duplicated with all of its connections and which accounts for the increase in genome size, and hence network growth. “Point mutations” affect the structure of a protein such that it changes its interacting partners and hence connections within the network. The time scales involved in these two processes are different. Gene duplication is very slow compared to point mutation. The observed rate of gene duplication is less than 10^{-2} per million years per gene in *Saccharomyces cerevisiae*, while the point mutation rate is at least one order of magnitude higher [21]. Point mutations which affect a protein’s ability to engage in molecular interactions are modeled as attachment or detachment of links, while the number of nodes is fixed (“link dynamics”). Since node duplication in evolutionary time scales is slow, compared to the time scale of link dynamics, gene duplication is modeled as addition of nodes without any links, while link dynamics occurs at each time step. This has been justified by the observation that in duplicated genes complete diversification occurs almost immediately after duplication. Usually, this divergence is biased, in that one of the proteins retains most of the interactions while the other retains a few or none [22]. Thus, for link dynamics in our simulation, a new attachment is established as follows: a random node is selected and attached to another node with preferential attachment, that is, with a rate proportional to its connectivity k as in the BA model. In contrast, for detachment, a link between two nodes is selected with a detachment rate proportional to the sum of inverses of their connectivities. This is motivated by the observation of higher mutation rates for less connected proteins [22, 23]. Importantly, simulation of network growth based on this duplication-mutation (DM) model led to the evolution of a network that exhibited power-law behavior with HBLC nodes (Figure 2d) similar to that exhibited by the yeast PIN. (See “data and methods” for details of model implementation.)

To compare the extent to which the various models produced HBLC nodes consistent with our experimental PIN data, we quantified the variation of betweenness values for a particular connectivity and its change with the value of the connectivity. In the basic BA network, betweenness and connectivity were almost linearly correlated in a logarithmic plot (Figure 2a). Thus, an increase in the standard deviation of betweenness values $D_B(k)$ among the nodes of a particular connectivity k , with decreasing k , reflects the presence of HBLC nodes. The plot of D_B versus the logarithm of k falls on a straight line,

which will be flat if HBLC nodes are absent. The slope S of the best-fit straight line can thus be used as a measure for the presence of HBLC nodes (Figure 3). Our DM model had a slope very close to that of the PIN data while other models had significantly lower values of S .

Taken together, these results show that existing growth algorithms that produce scale-free networks do not predict the existence of HBLC nodes found within the yeast PIN. In contrast, a new model that is biologically more realistic, and considers mutations (random rewiring) in addition to duplication (node and link addition), produces a global network architecture with HBLC nodes that is consistent with the PIN of living cells. This finding supports the general idea that a trait, in this case, a network topology feature, may arise during evolution because of its inherent robustness due to mechanistic and historical constraints [24, 25]. However, it does not exclude contributions due to functional adaptation driven by natural selection, since the two mechanisms of genesis are not mutually exclusive.

Essentiality

Therefore, to address a possible role of selective pressure in the bias in betweenness in the PIN, we examined the relationship between a protein’s essentiality and its betweenness value. Overall, we found that essential proteins of the yeast PIN had a higher mean betweenness and the frequency of high-betweenness nodes is greater for essential proteins. Mean betweenness for all proteins was 6.6×10^{-4} but for the essential proteins it was 1.2×10^{-3} ; this represents an increase of 82%. In the case of connectivity, the increase of the connectivity value of essential proteins relative to all proteins was 77%. Thus, the betweenness of a protein reflects its essentiality to at least the same degree as its connectivity [4]. In Figure 4, the percentage of essential proteins among proteins within a particular range of betweenness values is displayed as a function of betweenness. The increase in the variance of betweenness values for low-connectivity proteins disrupts this correlation for low-connectivity values, whereas it does not disrupt the correlation between betweenness and essentiality. This is interesting, because HBLC proteins are not “protected” from mutation by the constraint imposed by a high number of interaction partners as in the case of high-connectivity nodes [23] and thus they could easily lose their betweenness property.

Evolutionary age

The association of essentiality with low connectivity embodied by the HBLC proteins raises the question about the relationship between betweenness and the evolutionary age of a protein. The BA model of preferential attachment would suggest that high-connectivity proteins, which are typically essential, evolved earlier, while low-connectivity proteins are more likely to be recent additions to the network [26]. To estimate the evolutionary age of proteins, we used the list of isotemporal categories

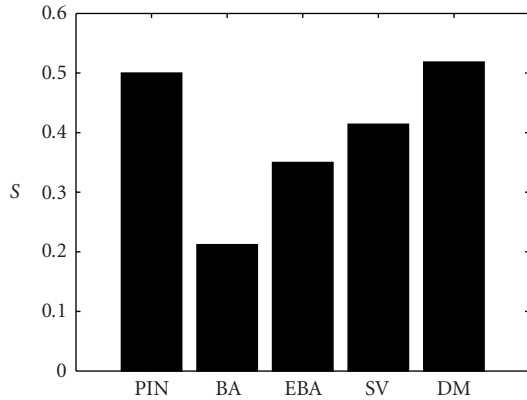


FIGURE 3. Magnitude of slope S , for the PIN data and different models (measured yeast protein interaction network (PIN) (data as in Figure 1b); the models are Barabasi-Albert (BA); extended BA (EBA); Sole-Vazquez (SV); duplication mutation (DM)). S measures the decrease of variance of the betweenness values of proteins with increasing degree, and hence indicates the relative prevalence of HBLC proteins.

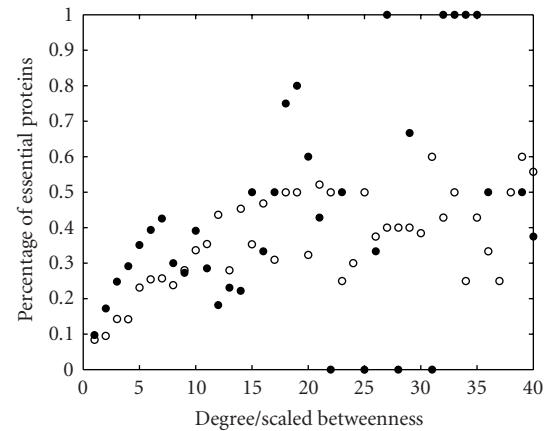


FIGURE 4. Percentage of essential genes with a particular degree (open circle) or betweenness (filled circle). Betweenness is scaled in such a way that the maximum value of betweenness is equal to the maximum degree. The plot was truncated at $k/B = 40$, since the number of essential genes beyond that is too small to have statistical significance.

of yeast protein orthologs provided by Qin et al [27], and classified them into four different age groups based on the phylogenetic tree, as in [26]. The core data set with confirmed interactions [16] showed a linear dependence of age and connectivity, while the dependence was not linear for the full data set [15], although there was a positive correlation (see Figure 5). The latter finding is consistent with the notion that some of the connections listed in the full data set are false positives [16].

Since betweenness correlates with essentiality and evolutionary age, it would be of particular interest to determine if the group of HBLC proteins has a different age or essentiality than the non-HBLC proteins of the same connectivity degree. Unfortunately, the number of proteins that falls into this class is too small to make statistically robust conclusions. This is because essentiality expressed as a continuous quantity as is done here and elsewhere [4, 26] is actually a group property (percentage of indispensable proteins in a given group) and not an attribute of individual proteins. Age is also a crude measure in that only four age groups can be defined; thus both measures require large numbers of proteins. With these caveats, our analyses found no statistically significant difference in evolutionary age or in essentiality between the HBLC proteins and their low-betweenness counterparts of the same connectivity.

DISCUSSION

Here, we report a new topology feature in the PIN not found in random networks: the prevalence of low-connectivity-degree nodes with high-betweenness values. It is also not predicted by the elementary growth model that explains the scale-free property of the PIN [5]. The existence of architectural features that deviate from that

of a random graph immediately raises the fundamental question of how such a nonrandom network structure first originated. In general, one can distinguish two main mechanisms of genesis that can contribute to a particular biological, nonrandom feature: (i) adaptive evolution toward optimization of a function and (ii) inherent robustness due to constraints imposed by the particular history and mechanism of its formation [24, 25]. The former explanation, which represents Darwinian selection of the fittest, is equivalent to the engineer's notion of functional optimization. Its validation typically rests on the demonstration of convergent evolution and of a functional advantage. Thus, it requires analysis of the specific identity of the nominal proteins, their evolutionary (historical) relationships, as well as the phenotypic consequences of that network structure [28, 29, 30]. In contrast, inherent robustness due to network constraints is more fundamental and implies that a nonrandom feature is the unavoidable consequence of some elementary physical, mechanistic, or other less obvious, self-organizing principles [25, 31]. As for networks, this second mechanism can be reduced to a simple, generic, generative algorithm that may represent a plausible mechanism for the genesis of a given system, as has been studied by researchers in the field of complexity [31, 32, 33]. Hence, network structures are particularly well suited for addressing the relative contribution of either mechanism responsible for formation of a nonrandom trait [25].

By comparing network growth models, we found that mutation (changes in network links due to addition and deletion) is central to the mechanism of network genesis that produces HBLC nodes. Thus, our simple algorithm explains this network topology feature without invoking functional adaptation. In this study on the generic architecture of the PIN, we do not discuss the molecular

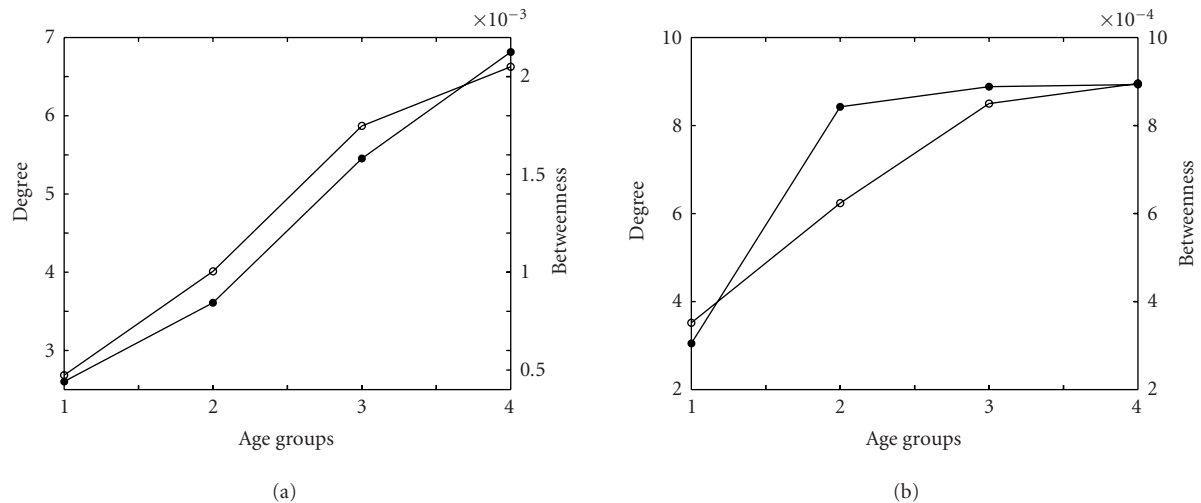


FIGURE 5. Degree and betweenness dependence of protein age. Average degree (left axis, open circle) and average betweenness (right axis, filled circle) of the four age groups of the yeast proteins. Group 1 contains proteins existing only in *S cerevisiae* and hence supposed to be the youngest while group 4 contains proteins existing in the all four branches and hence the oldest [32]. (a) Core data. (b) Full DIP data.

identity of HBLC proteins, but we show that their existence can at least be explained as an unavoidable consequence given certain assumed molecular mechanisms of network growth that involve random link rewiring due to mutations. This, together with the finding that HBLC nodes appear not to be evolutionary older proteins, favors the idea that the presence of HBLC proteins is due to intrinsic, structural, and mechanistic constraints of network growth rather than selective pressure on the growing network. However, to support a contribution of adaptive evolution to this distinct feature of network topology, it will be necessary to obtain larger data sets that can reveal an increased essentiality or higher evolutionary age of HBLC proteins compared with other proteins of the same connectivity class.

The HBLC feature also provides some insight into the modular organization of a large network. Real biological networks have a high clustering coefficient [34], indicating that the immediate neighbors of a given node are likely to be interconnected themselves. As a consequence, there are many alternate paths between two nodes. Betweenness can therefore be relatively small even if a node is highly connected, despite the overall correlation between connectivity and betweenness in the random networks. This could contribute to some variance of betweenness values of a protein with a particular (high) connectivity. On the other hand, the existence of high-betweenness nodes specifically with low connectivity suggests that there are proteins outside such clusters that connect those clusters. Thus, even without a precise definition for what constitutes a particular module, HBLC nodes point to the existence of modularity in the PIN. More specifically, HBLC proteins can be viewed as proteins that link

putative network modules within a genome-wide network.

Overall, this work illustrates that nonrandom network topology features represent one of the most simple phenotypic traits, simple enough to stimulate the formulation of generating algorithms, and therefore they provide a useful handle for addressing the fundamental dualism between adaptive evolution and intrinsic constraints in shaping the traits of living organisms.

DATA AND METHODS

Data

Yeast protein pairwise interaction information was from the yeast20040104.lst and ScereCR20040104.tab files, corresponding to the full and core data, respectively, obtained from <http://dip.doe-mbi.ucla.edu> [15, 16].

Calculation of betweenness centrality B

To calculate B of node i , one first counts the number of shortest paths between two nodes going through node i . Let b_i be the ratio of this number to the total number of shortest paths existing between those two nodes. The sum of b_i over all pairs of nodes in the network gives the betweenness B'_i of the node i . In this paper we use the quantity B_i , the scaled B'_i with respect to the maximum possible B in a network having n nodes, given by

$$B_i = \frac{2B'_i}{(n-1)(n-2)}. \quad (1)$$

B_i is positive and always less than or equal to 1 for any network. Betweenness of the whole graph is defined as the

average of the differences of all B_i from the largest value among the n nodes of the graph.

Model implementation

BA [5], EBA [18], and SV [19, 20] models were implemented as described in the corresponding references. In all these cases we investigated a range of parameters and selected the ones which gave power-law degree distributions. Among them, we searched for the best set of parameters which gave HBLC-type behavior.

Our generative model (DM) was implemented as follows. We start with a few connected nodes, as in [5]. For t number of steps, we apply the link dynamics, the preferential attachment, and the inverse-degree-dependent detachment of links, and then add a node without any links. This process is repeated until the network grows to the desired size. At each step, probability for attachment, p , and detachment, q , are set to be almost equal and adjusted to obtain the desired final mean connectivity. In our simulations we evolved the network till it reached 6000 nodes, corresponding to the approximate total number of genes in *S cerevisiae*. After this evolution process we selected the largest connected component for further network analysis. We selected parameters in such a way that the size of the largest connected component and mean connectivity are similar to that in PIN data. For many sets of parameters, this model produces a scale-free network with HBLC. Figure 2d gives the $k - B$ plot for one such parameter set.

ACKNOWLEDGMENTS

We would like to thank Luis Amaral and Gabriel Eichler for helpful discussions. This work was supported by grants from AFSOR (F49620-01-1-0564) to Sui Huang and from the NIH (CA55833) to Donald E. Ingber.

REFERENCES

- [1] Xenarios I, Eisenberg D. Protein interaction databases. *Curr Opin Biotechnol*. 2001;12(4):334–339.
- [2] Strogatz SH. Exploring complex networks. *Nature*. 2001;410(6825):268–276.
- [3] Erdos P, Renyi A. On random graphs I. *Publicationes Mathematicae*. 1959;6:290–297.
- [4] Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature*. 2001;411(6833):41–42.
- [5] Barabasi AL, Albert R. Emergence of scaling in random networks. *Science*. 1999;286(5439):509–512.
- [6] Maslov S, Sneppen K. Specificity and stability in topology of protein networks. *Science*. 2002;296(5569):910–913.
- [7] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. *Science*. 2002;298(5594):824–827.
- [8] Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. *Science*. 2002;297(5586):1551–1555.
- [9] Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature*. 1999;402(suppl 6761):C47–C52.
- [10] Watts DJ, Strogatz SH. Collective dynamics of “small-world” networks. *Nature*. 1998;393(6684):440–442.
- [11] Freeman LC. A set of measures of centrality based upon betweenness. *Sociometry*. 1977;40(1):35–41.
- [12] Girvan M, Newman MEJ. Community structure in social and biological networks. *Proc Natl Acad Sci USA*. 2002;99(12):7821–7826.
- [13] Guimera R, Amaral LAN. Modelling the world-wide airport network. *Eur Phys J B*. 2004;38:381–385.
- [14] Goh KI, Oh E, Jeong H, Kahng B, Kim D. Classification of scale-free networks. *Proc Natl Acad Sci USA*. 2002;99(20):12583–12588.
- [15] Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res*. 2002;30(1):303–305.
- [16] Deane CM, Salwinski L, Xenarios I, Eisenberg D. Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol Cell Proteomics*. 2002;1(5):349–356.
- [17] Mewes HW, Frishman D, Guldener U, et al. MIPS: a database for genomes and protein sequences. *Nucleic Acids Res*. 2002;30(1):31–34.
- [18] Albert R, Barabasi AL. Topology of evolving networks: local events and universality. *Phys Rev Lett*. 2000;85(24):5234–5237.
- [19] Sole RV, Pastor-Satorras R, Smith E, Kepler TB. A model of large-scale proteome evolution. *Adv Compl Syst*. 2002;5(1):43–54.
- [20] Vazquez A, Flammini A, Maritan A, Vespignani A. Modeling of protein interaction networks. *Complexus*. 2003;1:38–44.
- [21] Berg J, Lässig M, Wagner A. Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. *BMC Evol Biol*. 2004;4:51.
- [22] Wagner A. How the global structure of protein interaction networks evolves. *Proc R Soc Lond B Biol Sci*. 2003;270(1514):457–466.
- [23] Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. Evolutionary rate in the protein interaction network. *Science*. 2002;296(5568):750–752.
- [24] Gould, SJ. *The Structure of Evolutionary Theory*. Cambridge, Mass: Harvard University Press; 2002.
- [25] Huang S. Back to the biology in systems biology: what can we learn from biomolecular networks? *Brief Funct Genomic Proteomic*. 2004;2(4):279–297.

-
- [26] Eisenberg E, Levanon EY. Preferential attachment in the protein network evolution. *Phys Rev Lett*. 2003;91(13):138701.
- [27] Qin H, Lu HHS, Wu WB, Li WH. Evolution of the yeast protein interaction network. *Proc Natl Acad Sci USA*. 2003;100(22):12820–12824.
- [28] Conant GC, Wagner A. Convergent evolution of gene circuits. *Nat Genet*. 2003;34(3):264–266.
- [29] Alon U. Biological networks: the tinkerer as an engineer. *Science*. 2003;301(5641):1866–1867.
- [30] Wagner A. Does selection mold molecular networks? *Sci STKE*. 2003;2003(202):PE41.
- [31] Kauffman S. *The Origins of Order: Self-Organization and Selection in Evolution*. New York, NY: Oxford University Press; 1993.
- [32] Wolfram S. *A New Kind of Science*. Champaign, Ill: Wolfram Media; 2002.
- [33] Bar-Yam Y. *Dynamics of Complex Systems*. Reading, Mass: Addison-Wesley; 1997.
- [34] Wagner A, Fell D. The small world inside large metabolic networks. *Proc R Soc Lond B Biol Sci*. 2001;268(1478):1803–1810.