research articles Journal of proteome research

# Protein Bipartivity and Essentiality in the Yeast Protein−Protein Interaction Network

**Ernesto Estrada***

*Complex System Research Group, X-rays Unit, RIAIDT, University of Santiago de Compostela, Edificio CACTUS, Santiago de Compostela 15782, Spain*

**Received March 21, 2006**

Protein−protein interaction networks (PINs) are structured by means of a few highly connected proteins linked to a large number of less-connected ones. Essential proteins have been found to be more abundant among these highly connected proteins. Here we demonstrate that the likelihood that removal of a protein in a PIN will prove lethal to yeast correlates with the lack of bipartivity of the protein. A protein is bipartite if it can be partitioned in such a way that there are two groups of proteins with intergroup, but not intragroup, interactions. The abundance of essential proteins found among the least bipartite proteins clearly exceeds that found among the most connected ones. For instance, among the top 50 proteins ranked by their lack of bipartivity 62% are essential proteins. However, this percentage is only 38% for proteins ranked according to their number of interactions. Protein bipartivity also surpasses another 5 measures of protein centrality in yeast PIN in identifying essential proteins and doubles the number of essential proteins selected at random. We propose a possible mechanism for the evolution of essential proteins in yeast PIN based on the duplication-divergence scheme. We conclude that a replica protein evolving from a nonbipartite target will also be nonbipartite with high probability. Consequently, these new replicas evolving from nonbipartite (essential) targets will with high probability be essential.

**Keywords:** essential proteins • protein interactions • protein networks • yeast • centrality measures • graph theory

## I. Introduction

Protein−protein interaction networks (PINs) are complex systems which can be represented by means of networks, in which the nodes represent proteins and links, protein−protein interactions.[1−3] The building of these networks is made possible by the availability of protein interaction data obtained from direct large-scale experiments, such as yeast two-hybrid technology and mass spectrometry.[4−6] It is known that protein−protein interactions play a prominent role in most biological processes, and in recent years, there has been an intensive study of the properties of PINs, including the study of functional modules,[7−10] protein function,[11,12] domain−domain interactions,[13] protein evolution,[14−17] and protein essentiality[18,19] as well as the investigation of global network topologies, such as "small-worldness" and "scale-freeness".[20−22] The discovery of the scale-free properties of PINs in particular has attracted much attention.[22,23] This property is intimately related to the statistical distribution of the number of interactions per protein.[1] In a scale-free PIN, there are very few proteins exhibiting a large number of interactions, the so-called hubs; most of the proteins display a low connectivity. One of the most important consequences of this topological characteristic of PINs is that a few highly connected proteins guarantee the functional and structural integrity of the network.[24] The last can be exemplified by the fact that the removal of such highly connected nodes makes the network collapse into several isolated clusters. It has been observed that a correlation exists between the connectivity and essentiality of a given protein.[18,19] A protein is considered to be essential if a knock-out results in (conditional) lethality or infertility. On the contrary, nonessential proteins are those for which knock-outs yield viable and fertile individuals.

The identification in silico of essential proteins in a PIN is important from both a theoretical and practical point of view. On one hand, the correlation observed between connectivity and essentiality has triggered a debate about the emergence of correlations between the number of interactions of a protein and its evolutionary conservation.[24−29] On the other hand, proteins essential to the viability of a pathogen would make the most obvious and experimentally feasible targets for drug design.[30−32] It is clear that if a protein is essential for a cellular process in a pathogenic microorganism then a drug that blocks its activity will cause cessation of growth or death of the cell. However, a biological function cannot easily be assigned to a single protein but is more likely to emerge from the interaction of many proteins in the PIN.[33−35] We have previously demonstrated that, in correlating with protein essentiality, the implication of a protein in a large number of

* To whom correspondence should be addressed. E-mail: estrada66@yahoo.com. Fax: 34 981 547 077.

10.1021/pr060106e CCC: $33.50 © 2006 American Chemical Society



Published on Web 07/29/2006

network substructures is significantly more important than the number of interactions.[36]

Network substructures include single cycles, such as triangles, quadrilaterals, and pentagons, as well as fully connected clusters. The protein modules formed by these substructures can be of two different topological forms based on their bipartivity.[37,38] In bipartite modules, proteins can be divided into two subgroups, in such a way that proteins in one subgroup interact only with proteins in the other subgroup. The nonbipartite modules are those in which such division is not possible, the most critical one being the fully connected module in which all proteins interact with each other. Then, we explore here to what extent an individual protein contributes to the global network bipartivity.[38] This measure of protein bipartivity, $\beta(i)$, which can take values between 1/2 and 1, i.e., $1/2 < \beta(i) \leq 1$, shows a very good capability for identifying essential proteins in yeast PIN. In addition, we provide an assessment of the causes of this high correlation, of the differences between our findings and those previously reported for essentiality-connectivity correlations, as well as a scheme of the most probable evolutionary routes pursued by essential proteins.

## II. Methods

**Protein Bipartivity.** A protein in a PIN is bipartite if it does not take part in any closed-walk (CW) of odd length.[38] A walk of length $k$ in a PIN is a sequence of (not necessarily different) proteins $v_1, v_2, \cdots, v_r, v_{k+1}$ such that for each $i = 1,2 \cdots, k$ there is an interaction (link) between $v_i$ to $v_{i+1}$. A closed walk (CW) is a walk in which $v_{k+1} = v_1$ and protein bipartivity is measured as the proportion of even to total number of CW in the PIN. Let $G = (V,E)$ be a PIN with $N$ proteins and having eigenvalues of the adjacency matrix $\sigma = (\lambda_1, \cdots, \lambda_N)$. It is known that the total number of CWs in which protein $i$ takes part is given by the following expression:[39]

$$\mathrm{SC}(i) = \sum_{j=1}^{N} [v_j(i)]^2 \mathrm{e}^{\lambda_j} \tag{1}$$

where $(v_1, v_2, \ldots, v_n)$ is an orthonormal basis of $R^N$ composed by eigenvectors of the adjacency matrix associated with the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$, and $v_j(i)$ is the $i$th component of $v_j$. The total number of even CWs in which protein $i$ participates is given by the following expression:

$$\mathrm{SC}_{\mathrm{even}}(i) = \sum_{j=1}^{N} [v_j(i)]^2 \cosh(\lambda_j) \tag{2}$$

Then the protein $i$ bipartivity, $\beta(i)$, is given by[38]

$$\beta(i) = \frac{\sum_{j=1}^{N} [v_j(i)]^2 \cosh(\lambda_j)}{\sum_{j=1}^{N} [v_j(i)]^2 \mathrm{e}^{\lambda}{}_j} \tag{3}$$

When a protein does not takes part in any odd-length CW, i.e., $\mathrm{SC}_{\mathrm{odd}}(i) = 0$, and $\beta(i) = 1$, this means that this protein is bipartite. Protein bipartivity can take values between 1/2 and 1, $1/2 < \beta(i) \leq 1$, where the lower bound is reached when all proteins interacting with $i$ interact which each other, a fully connected motif or clique. The exact value of 1/2 is reached in the limit when the number of proteins in such cluster tends to infinity, i.e., $\infty$-clique. In general, the effect of "adding" a new interaction into a bipartite cluster of interacting proteins affects protein bipartivity in different degrees.[38]

**Protein Centrality Measures.** The other protein centrality measures used in this work include degree (DC), closeness (CC), betweenness (BC), eigenvector (EC), information (IC), and odd subgraph (SC) centralities. DC($i$) is the number of proteins that are interacting with protein $i$. CC($i$) is the sum of graph-theoretic distances from all other proteins, with respect to protein $i$, in the PIN, where the distance $d(i,j)$ from one protein $i$ to another $j$ is defined as the number of links in the shortest path from one to the other. The closeness centrality of protein $i$ in a PIN is given by the following expression:[40]

$$\mathrm{CC}(i) = \frac{N-1}{\sum_j d(i,j)} \tag{4}$$

BC characterizes the degree of influence a protein has in "communicating" between protein pairs and is defined as the fraction of shortest paths going through a given node. If $\rho(i,j)$ is the number of shortest paths from protein $i$ to protein $j$, and $\rho(i,k,j)$ is the number of these shortest paths that pass through protein $k$ in the PIN, then the betweenness centrality of node $k$ is given by[41]

$$\mathrm{BC}(k) = \sum_i \sum_j \frac{\rho(i,k,j)}{\rho(i,j)}, i \neq j \neq k \tag{5}$$

EC is defined as the principal eigenvector of the adjacency matrix **A** defining the network. It simulates a mechanism in which each node affects all of its neighbors simultaneously. The defining equation of an eigenvector is $\lambda \mathbf{e} = \mathbf{A}\mathbf{e}$, where **A** is the adjacency matrix of the graph, $\lambda$ is an eigenvalue, and **e** is the eigenvector. Thus, EC of protein $i$ is defined as the $i$th component of the eigenvector $\mathbf{e}_1$, $e_1(i)$, that corresponds to the largest eigenvalue of **A**, $\lambda_1$ (principal eigenvalue or index)[42,43]

$$\mathrm{EC}(i) = e_1(i) \tag{6}$$

Information centrality (IC) is based on the information that can be transmitted between any two points in a connected network. In the case of PINs, the transmission of information can be understood as the process in which a protein interacts with another and changes it, by a conformational or chemical transformation.[44] If **A** is the adjacency matrix of a network, **D** a diagonal matrix of the degree of each node, and **J** a matrix with all its elements equal to one, then IC is defined by inverting the matrix **B** defined as $\mathbf{B} = \mathbf{D} - \mathbf{A} + \mathbf{J}$ in order to obtain the matrix $\mathbf{C} = (c_{ij}) = \mathbf{B}^{-1}$ from which the information matrix is obtained as follows:[45]

$$\mathbf{I}_{ij} = (c_{ii} + c_{jj} - c_{ij})^{-1} \tag{7}$$

The information centrality of the protein $i$ is then defined by using the harmonic average

$$\mathrm{IC}(i) = \left[\frac{1}{N}\sum_j \frac{1}{\mathbf{I}_{ij}}\right]^{-1} \tag{8}$$

Stephenson and Zelen proposed defining $\mathbf{I}_{ii}$ as infinite for computational purposes, which makes $1/\mathbf{I}_{ii} = 0$.[45] The odd subgraph centrality, $\mathrm{SC}_{\mathrm{odd}}(i)$, is defined by counting the number

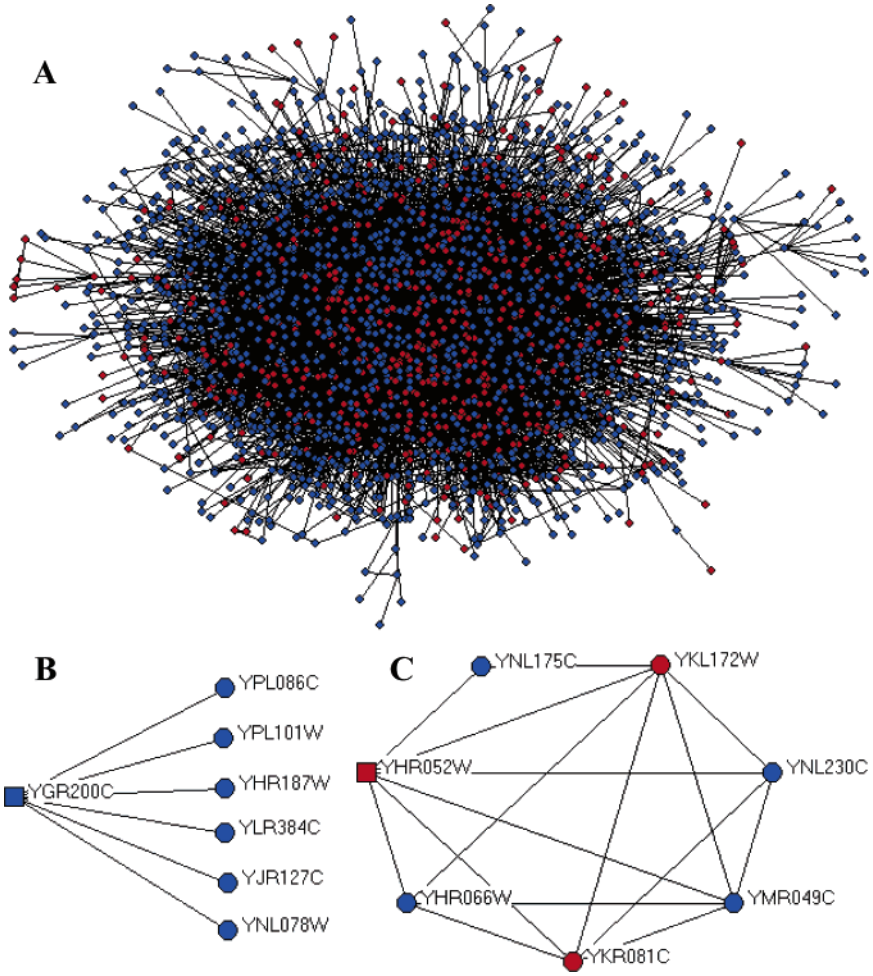


**Figure 1.** (A) The PIN of yeast showing essential proteins (red) and nonessential proteins as well as proteins with unknown essentiality (blue). (B) Subnetwork of the yeast PIN illustrating a bipartite protein (YGR200C). All proteins in this cluster are nonessential. (C) Subnetwork of the yeast PIN illustrating a nonbipartite protein (YHR052W), which takes part in a large number of odd cycles.

of odd-length CW in the PIN as was previously carried out for the CWs of even-length[39]

$$SC_{odd}(i) = \sum_{j=1}^{N} [v_j(i)]^2 \sinh(\lambda_j) \quad (9)$$

**PIN Dataset.** The protein−protein interaction network (PIN) of yeast *Saccharomyces cerevisiae* was compiled by Bu et al.[46] The protein interaction network is viewed as a set of nodes representing proteins and a set of links, which join together pairs of proteins that physically interact as determined experimentally by any of the technologies used with this end, mainly the yeast two-hybrid system and mass spectrometry identification of proteins that co-affinity purify with a bait protein.[3] The original data was obtained by von Mering et al. by assessing a total of 80 000 previously reported interactions among 5400 proteins and assigning each interaction a confidence level.[47] Bu et al. focused on 11 855 interactions with high and medium confidence between 2617 proteins in order to reduce the interference of false positives.[45] They reported a network of 2361 nodes and 6646 links (http://vlado.fmf.uni-lj.si/pub/networks/data/bio/Yeast/Yeast.htm). This interaction map may be considered to be a network in which proteins are represented as the nodes and two nodes are linked by an edge if the two proteins that are connected can be expected with high or medium confidence to be interacting.

The indispensability of a protein defines the functional significance of a gene at its most basic level. Essential genes are those upon which the cell depends for its viability. Using the GENECENSUS database (http://bioinfo.mbb.yale.edu/genome/), we checked for all proteins in the main cluster of the yeast PIN for essentiality. The main cluster, i.e., the larger cluster containing no isolated protein, of 2224 proteins sharing 6608 interactions which is used here for the analysis is illustrated in Figure 1. Protein bipartivity was calculated by an in-house program using Matlab.

## III. Results and Discussion

According to our calculations, most of the proteins in the yeast PIN show very low bipartivity. As illustrated in Figure 2A, 86% of all proteins in the yeast PIN display bipartivity values lower than 0.6, i.e., $0.5 < \beta(i) \leq 0.6$ and only 4% of proteins can be considered to be highly bipartite, i.e., $0.9 \leq \beta(i) \leq 1$. These results indicate that most yeast proteins are implicated in a large number of odd-length cycles in the PIN. In Figure 2B, we illustrate the percentage of essential proteins which are found among the least bipartite proteins in the PIN, e.g., those proteins for which $0.5 < \beta(i) \leq 0.6$. As can be seen, there is a

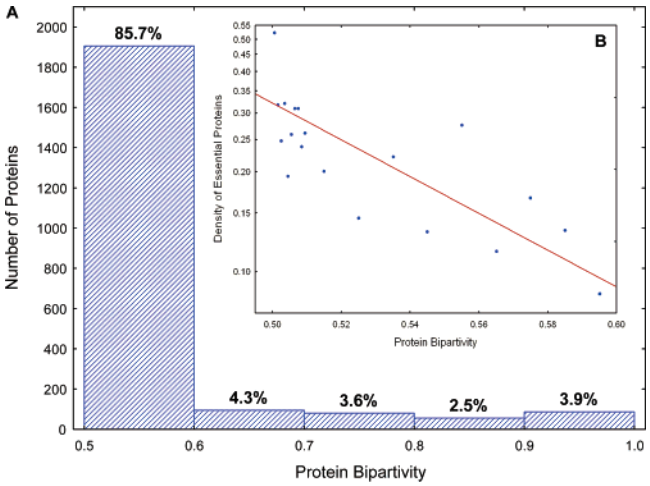


**Figure 2.** (A) Likelihood of protein essentiality for proteins with different bipartivity. The least bipartite proteins in the yeast PIN tend to be essential and only a small portion of bipartite proteins are essential. (B) Scatterplot of the percentage of essential proteins with lowest bipartivity in the yeast PIN versus their bipartivity. The largest abundance of essential proteins is found among the least bipartite proteins.

linear trend indicating that the larger number of essential proteins is found among the least bipartite proteins and that this abundance of essential proteins decreases as the protein bipartivity increases.

An immediate consequence of the distribution of essential proteins with respect to the protein bipartivity, observed in Figure 2 is that we can use this criterion for selecting essential proteins in the yeast PIN. Our selection strategy consists of ranking proteins according to their values of bipartivity; we select the top 10, 50, 100, etc. proteins, and determine how many in each selected group are essential in the yeast PIN. For purposes of comparison we also select proteins according to 8 other criteria which rank proteins according to their centrality in the PIN. These centrality measures include odd subgraph (SC), degree (DC), closeness (CC), betweenness (BC), eigenvector (EC), and information (IC) centralities (see Methods). We also use a random selection method, taking the average of the number of essential proteins after 20 random selections (realizations). Our results are illustrated in Figure 3 where we plot the number of essential proteins identified by each centrality measure for the first top 10−400 ranked proteins. Protein bipartivity is clearly superior to all other centrality measures in identifying essential proteins, including our previous results based on the use of the subgraph centrality. For instance, among the top 10 proteins bipartivity identifies 60% of essential proteins, whereas random selection identifies less than 25%. A very important result is that essential proteins are more abundant among the least bipartite proteins than among the most connected ones, the so-called network hubs. It was previously concluded that "the likelihood that removal of a protein will prove lethal correlates with the number of interactions the protein has".[18] This very intuitive idea has enjoyed great resonance because the most "popular" proteins in the PIN, i.e., those having the largest number of interactions, are indeed the most important ones, i.e., they are essential. However, if we take, for instance, the top 50 proteins in Figure 3, we can see that the number of interactions a protein has (DC) only identifies 38% of essential proteins in the PIN versus 62% identified by protein bipartivity. In general, for the first top 300 proteins, bipartivity identifies 20 essential proteins more effectively than the protein degree. Protein bipartivity doubles the number of essential proteins identified at random in all cases. Thus the general hypothesis about protein essentiality should be reformulated as "the likelihood that removal of a protein will prove lethal correlates with its lack of bipartivity". Unfortunately, the combination of several centrality criteria in a linear discriminant analysis does not improve the number of essential proteins identified. Thus, we think that the best possible choice for combining the information obtained from different centrality measures would be the use of partial order ranking techniques in order to improve the predictability of this method of essential proteins identification.

Our previous results raise several questions concerning the importance of protein bipartivity and its role in the structure, function, and evolution of the yeast PIN. The first question is related to the differences found between bipartivity and connectivity to identify essential proteins. Simple reasoning suggests that the most connected proteins in the PIN are the least bipartite ones. This is based on a simple transitivity principle, if protein A interacts with proteins B and C there is a high probability that proteins B and C interact. Despite the fact that this principle is in general obeyed, in the yeast PIN there are many cases in which proteins with the same number of interactions have different bipartivities. One simple example is provided in Figure 1, panels B and C, where two proteins, YGR200C and YHR052W, having interactions with 6 other proteins show very different bipartivities, very high for YGR200C and very low for YHR052W. This situation is even more critical for proteins with a low number of interactions which can display bipartivities ranging from 0.5 to 1.0 (data not shown). A simple example is provided by proteins forming a triangle or a square in the PIN. Both proteins have interactions with only two other proteins, but although the one forming the square is bipartite, the other is not. A similar lack of correlation exists between the bipartivity and the transitivity measured through the so-called "clustering coefficient",[20] which is the number of triangles in which a protein takes part divided by the number of triples centered on such a protein (data not shown). In this case, we can find pairs of proteins having zero clustering and different bipartivities. This is the case of proteins forming a square or a pentagon; both have zero clustering but the one forming the square is bipartite and that forming the pentagon is not. There are also proteins with the same bipartivity and different clustering, such as proteins in a triangle and a pentagon; both are bipartite but have clustering of one and zero, respectively. Thus, protein bipartivity captures topological information which is not represented by other centrality measures of a protein in a PIN and more importantly, this information is relevant for protein essentiality.

It is known that biological function generally emerges from the interaction of many separate components, such as genes, proteins, or metabolites, which form distinct functional modules.[33−35] These modules are organized in the form of structural motifs, topologically distinct interaction patterns within a network.[48,49] In contrast, a biological function can only rarely be assigned to a single component and some of these motifs, in particular the fully connected ones, tend to identify protein complexes responsible for distinct biological functions.[50] From the point of view of our current study, this means that those proteins forming part of highly connected motifs representing functional modules have the highest probability of being essential in a PIN. However, not all proteins in a large

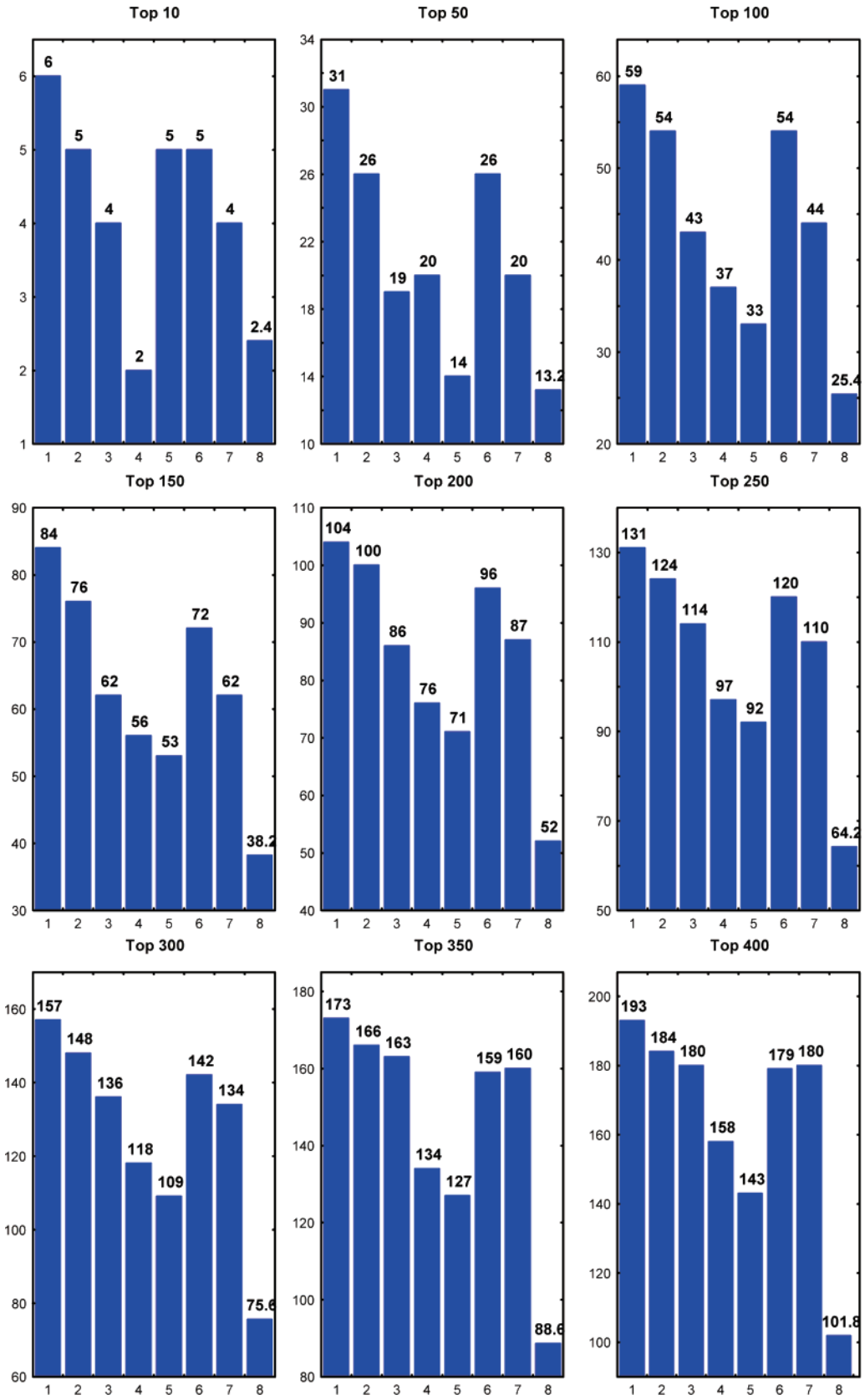


**Figure 3.** Number of essential proteins selected by ranking proteins according to their values of centrality and at random (after 20 realizations). The measures used are as follows: bipartivity (1); subgraph centrality (2); degree (3); closeness (4); betweenness (5); eigenvector (6); information (7); random (8).

protein complex interact with each other, and in many cases, highly interconnected motifs are formed among proteins of different complexes. An illustrative example is provided by the subnetwork formed by 11 proteins of the 26S proteasome

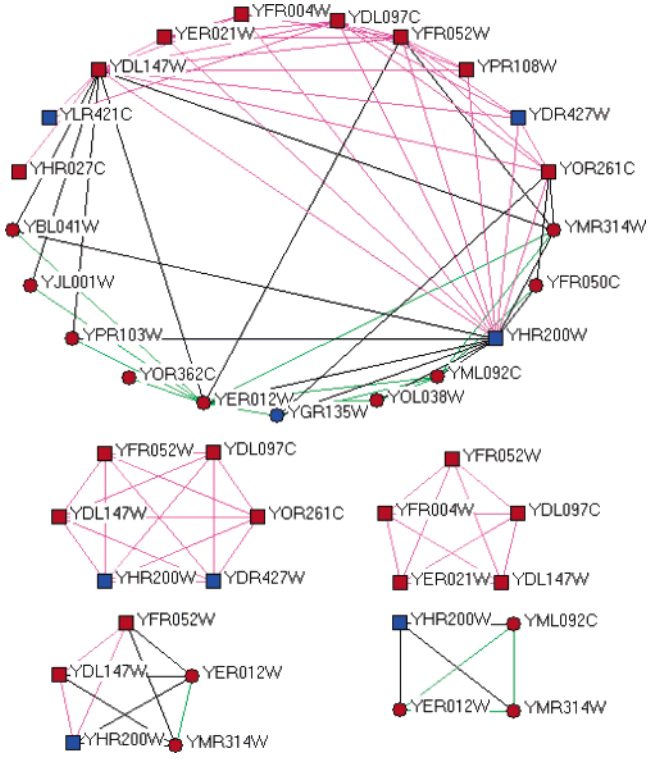


**Figure 4.** Subnetwork of the yeast PIN formed by proteins of the 26S proteasome regulatory subunit (squares) and the proteasome component (circles). Four fully connected motifs formed by proteins in this subnetwork are also illustrated. Interactions between proteins of the 26S proteasome regulatory subunit are in magenta and those between proteins of the proteasome component are in green. Interactions between proteins of both complexes are in black. Essential proteins are in red; nonessential, in blue.

regulatory subunit and 10 proteins of the proteasome component (Figure 4). The first acts as a regulatory subunit of the 26S proteasome which is involved in the ATP-dependent degradation of ubiquinated proteins and the second is a multicatalytic proteinase complex. Most proteins in this subnetwork show very low bipartivity, and consequently, 81% of them are essential. Close examination of this subnetwork permits identification of several fully connected motifs, such as those illustrated in Figure 4. The largest of these motifs is formed by 6 proteins of the 26S proteasome complex, which includes the protein YFR052W necessary for the activation of the CDC28 kinase. There are also fully connected motifs formed by combinations of the proteins of both complexes as illustrated at the bottom of this figure.

Finally, we would like to note that our current findings shed some light on the nature of the evolution of essential proteins in yeast. We have shown here that essential proteins are the least bipartite with respect to their interactions in the yeast PIN. Assuming a duplication-divergence model of protein evolution[14,51,52] we arrive at the hypothesis that essential proteins evolve from others which have very low bipartivity and, with high probability, are also essential. Our reasoning is as follows. Consider a bipartite protein $i$ such as the one illustrated in black in Figure 5A. After the duplication of $i$, a new protein $i'$, a replica, (in gray) with an interaction to all the neighbors of $i$, is created. At this point, the new protein forms cycles of even but not of odd length, i.e., it is also bipartite. If we allow an interaction between $i$ and $i'$ (self-interacting proteins), the bipartivity of both proteins decreases but their values of $\beta(i)$ are only slightly different from one, as they are taking part only in a pair of triangles. For the sake of simplicity, we will not consider here interaction between $i$ and $i'$. The next step is the divergence, where some common interactions will be lost, and new interactions may be gained by either protein. During this process, to generate a protein displaying very low bipartivity, it is necessary that all previous interactions are retained and new ones created to complete links between all pairs of proteins (see the top option in Figure 5A). However, according to a simple parsimony principle this process is not very probable and the most probable one evolves in proteins with bipartivity close to one as illustrated at the bottom of Figure 5A. Consequently, a replica protein evolving from a bipartite target will be also bipartite with high probability. In Figure 5B, we illustrate a duplication−divergence process starting from a target with very low bipartivity, e.g., $\beta(i) \approx 0.5$. After the duplication of this protein, the replica protein will display a very low bipartivity, similar to its target. During the divergence stage, it is necessary to remove all interactions between pairs of proteins which are connected to $i$ and $i'$ in order to produce a bipartite replica. This process appears to have very low probability as it will remove almost all interactions existing previously in the cluster. It appears to be more probable that only some of these interactions are removed, retaining some others and creating new ones, which result in a replica protein with low bipartivity, similar to its target. Our conclusion is that a replica protein evolving from a nonbipartite target will be also nonbipartite with high probability. As we shown here that nonbipartite proteins display high probability of being essential then these new replicas evolving from nonbipartite (essential) targets will also display high probability of being essential.

Our hypothesis is supported by several observations. First, there is a large number of fully connected motifs (cliques) in yeast PIN. Here we have calculated the existence of 3530 triangles (3 cliques), 2576 fully connected squares (4 cliques), and 1711 fully connected pentagons (5 cliques). This abundance of very low bipartivity clusters has been also observed by others[46,50,53] and has been explained by evolutionary models based on duplication-divergence network growth.[52] Even more interesting is the fact that nonbipartite motifs display significantly larger evolutionary conservation rates than are seen in bipartite motifs.[50] For instance, the 5 clique shows the highest conservation rate among all motifs (47.24%), followed by the 4 cliques (32.53%) and the 3 clique (20.51%).[50] This trend is also observed in nonfully connected motifs, such as squares and pentagons. Although the bipartite motif (square) shows a conservation rate of only 6.71%, the nonbipartite motif (pentagon) doubles this rate to 14.77%. Bipartite motifs never exceed 15% of the evolutionary conservation rate.[50]

## IV. Conclusions

Our current study provides a thorough assessment of the relevance of protein bipartivity for understanding essentiality in yeast PIN. Protein bipartivity not only clearly surpasses other protein topological measures in a PIN but also gives insights about the possible mechanisms associated with the evolution of essential genes in yeast. It is important to remark that the abundance of essential proteins found among the least bipartite proteins clearly exceeds that found among the most connected ones. Based on our findings, we have proposed a

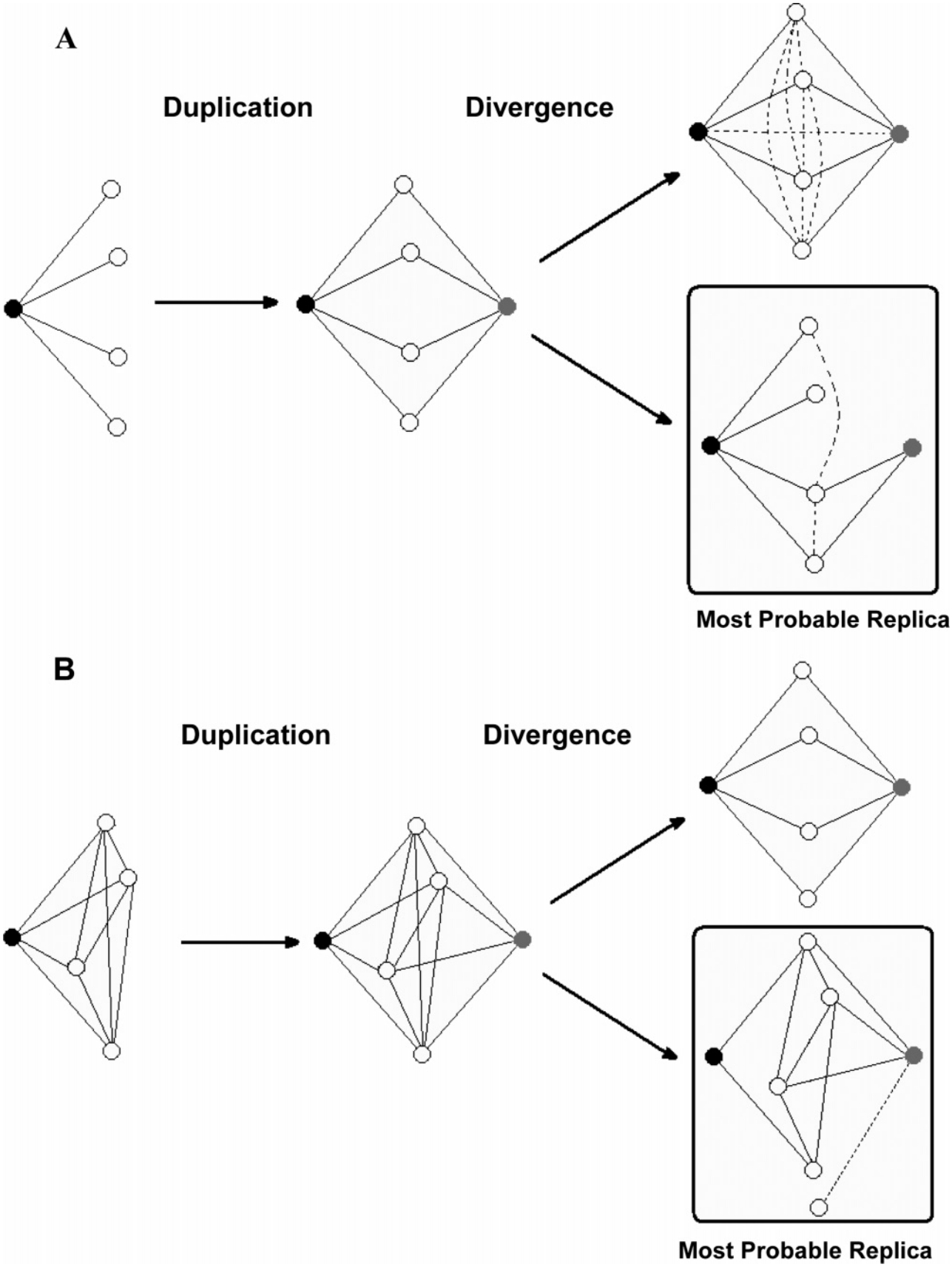


**Figure 5.** Possible mechanisms of protein evolution starting from a bipartite protein (A) and from a nonbipartite one (B). Target is in black, replica, in gray, and the rest of proteins, in white (see text for explanation).

possible mechanism for the evolution of essential proteins in yeast PIN based on the duplication-divergence scheme. Using such a model, we can conclude that a replica protein evolving from a nonbipartite target will also be nonbipartite with high

probability. Consequently, these new replicas evolving from nonbipartite (essential) targets will with high probability be essential.

**Acknowledgment.** The author thanks the "Ramón y Cajal" program, Spain for partial financial support.

## References


(1) Barabási, A.-L.; Oltvai, Z. N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **2004**, *5*, 101−113.
(2) Bork, P.; Jensen, L. J.; von Mering, C.; Ramani, A. K.; Lee, I.; Marcotte, E. M. *Curr. Opin. Struct. Biol.* **2004**, *14*, 292−299.
(3) Uetz, P.; Finley, R. L., Jr. From protein networks to biological systems. *FEBS Lett.* **2005**, *579*, 1821−1827.
(4) Ito, T.; Chiba, T.; Ozawa, R.; Yoshida, M.; Hattori, M.; Sasaki, Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4569−4574.
(5) Uetz, P.; Giot, L.; Cagney, G.; Mansfield, T. A.; Judson, R. S.; Knight, J. R.; Lockshon, D.; Narayan, V.; Srinivasan, M.; Pochart, P.; Qureshi-Emili, A.; Li, Y.; Godwin, B.; Conover, D.; Kalbfleisch, T.; Vijayadamadar, G.; Yang, M.; Johnston, M.; Fields, S.; Pothberg, J. M. A comprehensive analysis of protein−protein interactions in *Saccharomyces cereviciae*. *Nature* **2000**, *403*, 623−627.
(6) Pandey, A.; Mann, M. Proteomics to study genes and genomes. *Nature* **2000**, *405*, 837−846.
(7) Ravasz, E.; Somera, A. L.; Mongru, D. A.; Oltvai, Z. N.; Barabási, A.-L. Hierarchical organization of modularity in metabolic networks. *Science* **2002**, *297*, 1551−1555.
(8) Rives, A. W.; Galitski, T. Modular organization of cellular networks. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1128−1133.
(9) Pereira-Leal, J. B.; Enright, A. J.; Ouzounis, C. A. Detection of functional modules from protein interaction networks. *Proteins. Struct. Funct. Bioinf.* **2004**, *54*, 49−57.
(10) Campillos, M.; von Mering, C.; Jensen, L. J.; Bork, P. Identification and analysis of evolutionarily cohesive functional modules in protein networks. *Genome Res.* **2006**, *16*, 374−382.
(11) Vázquez, A.; Flammini, A.; Maritan, A.; Vespignani, A. Global protein function prediction from protein−protein interaction networks. *Nat. Biotechnol.* **2003**, *21*, 697−700.
(12) Schlitt, T.; Palin, K.; Rung, J.; Dietmann, S.; Lappe, M.; Ukkonen, E.; Brazma, A. From gene networks to gene function. *Genome Res.* **2003**, *13*, 2568−2576.
(13) Deng, M.; Mehta, S.; Sun, F.; Chen, T. Inferring domain-domain interactions from protein−protein interactions. *Genome Res.* **2002**, *12*, 1540−1548.
(14) Wagner, A. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol. Biol. Evol.* **2001**, *18*, 1283−1292.
(15) Pastor-SaTorras, R.; Smith, E.; Solé, R. V. Evolving protein interaction networks through gene duplication. *J. Theor. Biol.* **2003**, *222*, 199−210.
(16) Amoutzias, G. D.; Robertson, D. L.; Oliver, S. G.; Bornberg-Bauer, E. Convergent evolution of gene networks by single-gene duplications in higher eukariotes. *EMBO Rep.* **2004**, *5*, 1−6.
(17) Wutchy, S. Evolution and topology in the yeast protein interaction network. *Genome Res.* **2004**, *14*, 1310−1314.
(18) Jeong, H.; Mason, S. P.; Barabási, A.-L.; Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **2001**, *411*, 41−42.
(19) Yu, H.; Greenbaum, D.; Lu, H. X.; Zhu, Z.; Gerstein, M. Genomic analysis of essentiality within protein networks. *Trends Genet.* **2004**, *20*, 227−231.
(20) Watts, D. J.; Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **1998**, *393*, 440−442.
(21) Barabási, A.-L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509−512.
(22) van Noort, V.; Snel, B.; Huynen, M. A. The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple mode. *EMBO Rep.* **2004**, *5*, 1−5
(23) Vázquez, A.; Flammini, A.; Maritan, A.; Vespignani, A. Modeling of protein interaction networks. *ComPlexUs* **2003**, *1*, 38−44.
(24) Hurst, L. D.; Smith, N. G. C. Do essential genes evolve slowly? *Curr. Biol.* **1999**, *9*, 747−750.
(25) Hirsh, A. E.; Fraser, H. B. Protein dispensability and rate of evolution. *Nature* **2001**, *411*, 1046−1049.
(26) Hirsh, A. E.; Fraser, H. B. Genomic function (communication arising): rate of evolution and gene dispensability. *Nature* **2003**, *421*, 497−498.
(27) Pál, C.; Papp, B.; Hurst, L. D. Genomic function (communication arising): rate of evolution and gene dispensability. *Nature* **2003**, *421*, 496−497.
(28) Castillo-Davis, C. I.; Hartl, D. L. Conservation, relocation and duplication in genome evolution. *Trends Genet.* **2003**, *19*, 593−597.
(29) Bloom, J. D.; Adami, C. Evolution rate depends on number of protein−protein interactions independently of gene expression level: Response. *BMC Evol. Biol.* **2004**, *4*, 14.
(30) Haselbeck, R.; Wall, D.; Jiang, B.; Ketela, T.; Zyskind, J.; Bussey, H.; Foulkes, J. G.; Roemer, T. Comprehensive essential gene identification as a platform for novel anti-infective drug discover. *Curr. Pharm. Des.* **2002**, *8*, 1155−1172.
(31) Nikolsky, Y.; Nikolskaya, T.; Bugrim, A. Biological networks and analysis of experimental data in drug discovery. *Drug Discov. Today* **2005**, *10*, 653−662.
(32) Odds, F. C. Genomics, molecular targets and the discovery of antifungal drugs. *Rev. Iberom. Micol.* **2005**, 22, 229−237.
(33) Hartwell, L. H.; Hopfield, J. J.; Leibler, S.; Murray, A. W. From molecular to modular cell biology. *Nature* **1999**, *402*, C47−C52.
(34) Snel, N.; Bork, P.; Huynen, M. A. The identification of functional modules from the genomic association of genes. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5890−5895.
(35) Oltvai, Z. N.; Barabási, A.-L. Life's complexity pyramid. *Science* **2002**, *298*, 763−764.
(36) Estrada, E. Virtual identification of essential proteins within the protein interaction network of yeast. *Proteomics* **2006**, *6*, 35−40.
(37) Holme, P.; Liljeros, F.; Edling, C. R.; Kim, B. J. Network bipartivity. *Phys. Rev. E* **2003**, *68*, 056107.
(38) Estrada, E.; Rodríguez-Velázquez, J. A. Spectral measures of bipartivity in complex networks. *Phys. Rev. E.* **2005**, *72*, 046105.
(39) Estrada, E.; Rodríguez-Velázquez, J. A. Subgraph centrality in complex networks. *Phys. Rev. E.* **2005**, *71*, 056103.
(40) Freeman, L. C. Centrality in social networks, I. Conceptual clarification. *Social Networks* **1979**, *1*, 215−239.
(41) Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* **1977**, *40*, 35−41.
(42) Bonacich, P. Factoring and weighting approach to clique identification. *J. Math. Sociol.* **1972**, *2*, 113−120.
(43) Bonacich, P. Power and centrality: A family of measures. *Am. J. Sociol.* **1987**, *92*, 1170−1182.
(44) Carmi, S.; Levanon, E. Y.; Havlin, S.; Eisenberg, E. Connectivity and expression in protein networks: Proteins in a complex are uniformly expressed. *Phys. Rev. E* **2006**, 031909.
(45) Stephenson, K.; Zelen, M. Rethinking centrality: methods and examples. *Social Networks* **1989**, *11*, 1−37.
(46) Bu, D.; Zhao, Y.; Cai, L.; Xue, H.; Zhu, X.; Lu, H.; Zhang, J.; Sun, S.; Ling, L.; Zhang, N.; Li, G.; Chen, R. Topological structure analysis of the yeast protein−protein interaction network of budding yeast. *Nucl. Acid Res.* **2003**, *31*, 2443−2450.
(47) von Mering, C.; Krause, R.; Snel, B.; Cornell, M.; Oliver, S. G.; Field, S.; Bork. P. Comparative assessment of large-scale data sets of protein−protein interactions. *Nature* **2002**, *417*, 399−403.
(48) Milo, R.; Shen-Orr, S.; Itzkovitz, S.; Kashtan, N.; Chklovskii, D.; Alon, U. Network motifs: Simple building blocks of complex networks. *Science* **2002**, *298*, 824−827.
(49) Yeger-Lotem, E.; Sattath, S.; Kashtan, N.; Itzkovitz, S.; Milo, R.; Pinter, R. Y.; Alon, U.; Margalit, H. Network motifs in integrated cellular networks of transcription-regulation and protein−protein interaction. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5934−5939.
(50) Wutchy, S.; Oltvai, Z. N.; Barabási, A.-L. Evolutionary conservation of motifs constituents in the yeast protein interaction network. *Nat. Genet.* **2003**, *35*, 176−179.
(51) Ohno, S. *Evolution by Gene Duplication*; Springer: New York, 1970.
(52) Ispolatov, I.; Krapivsky, P. L.; Mazo, I.; Yuryev, A. Cliques and duplication-divergence network growth. *New J. Phys.* **2005**, *7*, 145.
(53) Zhang, L. V.; King, O. D.; Wong, S. L.; Goldberg, D. S.; Tong, A. H. Y.; Lesage, G.; Andrews, B.; Bussey, H.; Boone, C.; Roth, F. P. Motifs, themes and thematic maps of an integrated *Saccharomyces cereviciae* interaction network. *J. Biol.* **2005**, *4*, 6.


PR060106E