

Topology of resultant networks shaped by evolutionary pressure

Avi Ma'ayan, Azi Lipshtat, and Ravi Iyengar

Department of Pharmacology and Biological Chemistry, Mount Sinai School of Medicine, New York, New York 10029, USA

(Received 28 December 2005; revised manuscript received 28 March 2006; published 19 June 2006)

Understanding the topology of complex systems abstracted to networks is important for unraveling their functional capabilities. Many such networks follow the small-world and scale-free regimes. Several models of artificially growing networks lead to this observed network topology. Most previously proposed models for growing networks, such as rich-get-richer and duplication-divergence, produce realistic network topologies but do not consider the effects of exogenous forces such as optimization for adaptation in shaping network topology. It is likely that such forces have shaped complex systems throughout their evolution. To develop further insights into possible mechanisms that shape networks, a model that uses several previously proposed network growth algorithms was developed to grow networks that adapt under exogenous stress. A decision tree problem was used to generate a complex Boolean function. Growing networks were required to adapt to correctly decode this function using an evolutionary selection process. Under this growth regimen all growing network models are similarly adaptable. The newly added nodes tend to cluster into pathways emanating from few inputs, regardless of the growth algorithm. Distribution of redundant pathways from inputs to the output follow a power-law function with a scaling exponent (~ 1.3). Similar distribution of redundant pathways was observed from inputs in a cell signaling network and an air traffic control network. A flat distribution of redundant pathways from inputs was observed in growing networks that do not attempt to adapt. This analysis provides initial insights into distribution of pathways in naturally evolving complex systems that have defined input-output relationships.

DOI: [10.1103/PhysRevE.73.061912](https://doi.org/10.1103/PhysRevE.73.061912)

PACS number(s): 87.80.Vt, 87.23.Kg, 87.16.Yc, 89.75.-k

I. INTRODUCTION

Many complex systems can be abstracted to networks. For example, biological cellular components such as proteins are represented as nodes, and links describe their interactions. Other real-world examples include coauthorship of scientific papers [1], the world wide web linked web pages [2] and the Internet, connecting computers through Internet Protocol (IP) [3]. Most biological and nonbiological real-world networks are found to be “small world” [4] and “scale free” [5]. These include the protein-protein interaction networks of yeast [6,7] and fruit fly [8], metabolic networks of prokaryotes [9], and gene regulation networks in *Escherichia coli* [10]. Small-world networks have relatively high clustering coefficients and short characteristic path lengths [4] compared to randomized networks. The scale-free feature indicates that the nodal connectivity follows a power-law distribution [5] wherein there are a few but substantial number of highly connected hubs while most nodes have few links. Different algorithms have been proposed to explain the growth process that gives rise to observed topologies [11–16]. These algorithms use simple rules for stochastic growth of networks. Starting with a seed network made up of a few nodes and links, the networks grow in nodes and links and eventually reach a size where the network displays the small-world/scale-free characteristics. Most algorithms for network growth [11–16] do not consider exogenous forces that influence the development of the topological structure of complex systems abstracted to networks. Thus, in these models, all nodes have only intrinsic roles. However, it is clear that information processing networks, such as biological cellular regulatory signaling networks have defined inputs and out-

puts. These networks respond to changes in the extracellular environment and evoke phenotypic changes. Hence, it is reasonable to assume that their evolving topology may be shaped by signals from the extracellular environment. Other real-world networks, such as air traffic control networks are expected to become optimized to geographical parameters like population density, geometry, and terrain as they evolve. Developing models of network growth that include adaptation may be useful in understanding how realistic network topologies emerge. Further, since such networks have defined input-output relationships it is important to know how paths from inputs to outputs are distributed. To address this question previously proposed network growth models were extended to include inputs and outputs, as well as decision making capabilities. These network models were compared for emergent topologies resulting from growth under stress, and the ability of the resultant networks to process information.

II. METHODS

The Restaurant Problem [17] is a classical hypothetical problem used to illustrate concepts of machine learning. The problem describes a decision process by which restaurant guests who have just arrived at a restaurant need to decide whether to wait for a table or go elsewhere. The decision is made based on answers to several questions such as the approximate waiting time, and the weather outside. The restaurant problem decision tree was used to create a complex Boolean function from the problem description [Figs. 1(a) and 1(b)]. This function was chosen because its output ex-

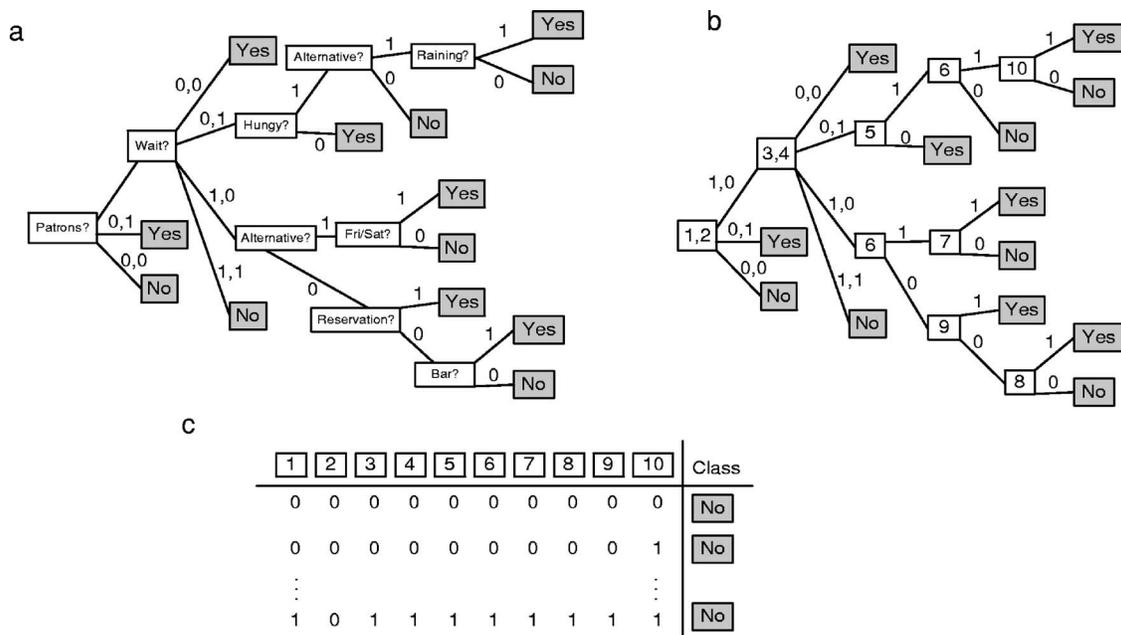


FIG. 1. A schematic representation (adapted from Russell and Norvig [17]) (a) of the decision making process in the “restaurant problem,” and the conversion of the decision tree to a Boolean function (b) and (c). The problem describes a scenario where patrons enter a restaurant and then must make a decision whether to wait for a table (Yes) or leave (No) through a series of ten questions.

hibits a nontrivial structure. For example, some inputs are more influential in determining the output than others. Each of the ten binary inputs corresponds and decodes either one question or part of a question from the decision tree. Questions with more than two answers are decoded with two inputs. For example, the “Patrons?” [Fig. 1(a)] question is decoded by inputs 1 and 2, such that 00 means none, 01 means some, 10 means full, and 11 is not a valid option. The function has ten binary inputs (attributes) and one corresponding binary output (class)

$$IO_{rel} = \{Attr_1, \dots, Attr_{10}, Class\}. \quad (1)$$

This equation produces 2^{10} input/output relationships. A number of inputs/output relationships (i.e., 100) were chosen randomly to create training sets. This complex Boolean function is used to train growing and adapting networks.

Growing network models attempt to decode correctly a set of examples randomly picked from the Boolean function solution space. The networks adapt to decode correctly examples through an evolutionary strategy growth process [18]. The growing networks are connected undirected graphs

$$G = (V^s, V^c, V^t, E), \quad (2)$$

with a fixed number of source nodes (vertices): $V^s = \{v_1^s, \dots, v_n^s\}$, a variable number of core nodes: $V^c = \{v_1^c, \dots, v_m^c\}$, and one target node (sink): $V^t = \{v\}$. The nodes are connected by a set E of edges (links). This setup is similar to an array of cell surface receptors (inputs) connected through multiple core nodes (transducers) and a downstream transcription factor (effector/output). The seed network is a network (undirected graph) where each source node is linked, with one link (edge), to a unique core node

$$e_i = (v_i^s, v_i^c), \quad i = 1, \dots, n. \quad (3)$$

Each core node is directly linked, with one link, to the single sink target node

$$e_i = (v_i^c, v), \quad i = n + 1, \dots, n + m. \quad (4)$$

In this implementation, source nodes represent $n=10$ inputs directly connected to $m=10$ core nodes connected to one output sink node [Fig. 2(a)].

These seed networks grow using three previously described algorithms: duplication/divergence (DD) [12,13], scale free (SF) [11], and exponential (EXP) [11]. Duplication/divergence networks grow by duplicating a randomly selected core node with its links, and then removing some of the links from the newly generated node

$$V_{new}^c = V_{old}^c \cup \{v_{new}^c\}, \quad (5)$$

such that

$$\{\text{Neighbors}(v_{new}^c)\} \subseteq \{\text{Neighbors}(v_{rand(1, \dots, m)}^c)\}. \quad (6)$$

The set of Neighbors (nodes that are directly connected to the specific node) of the newly added nodes is a subset of the Neighbors of a randomly chosen node that is copied. When a node is duplicated there is also a chance that the newly created node will also have a link to the node it was copied from. The algorithm for this growth process uses the probabilities $p=0.1$ for connecting the new node with the copied node and the probability $p=0.7$ for removing new links (neighbors) from the newly generated node. These p values were chosen because they were used to generate a network similar to the protein-protein interaction network of *Saccharomyces cerevisiae* [12]. Additionally, a rule was added to

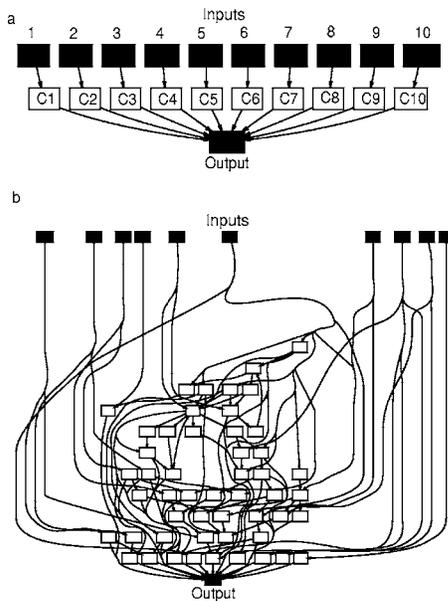


FIG. 2. Initial network and a network after growth and adaptation. (a) The initial network with ten input nodes and ten core nodes. (b) An example of a SF model network after a growth period of 41 generations. For each generation a training set of 100 inputs/output combinations were used to test 50 optional networks.

ensure that newly generated copied nodes are not isolated. New nodes must have at least one path to one source node, and at least one path to the target node. This rule was used for all three growing networks methods.

The scale-free networks (SF) grow by randomly adding one new core node [Eq. (5)] and connecting it to any other two or three existing nodes (with probability $p=0.5$)

$$E_{\text{new}} = E_{\text{old}} \cup \{(v_{\text{new}}^f, v_i^f), (v_{\text{new}}^f, v_j^f)\}, \quad \text{or}$$

$$E_{\text{new}} = E_{\text{old}} \cup \{(v_{\text{new}}^f, v_i^f), (v_{\text{new}}^f, v_j^f), (v_{\text{new}}^f, v_k^f)\}. \quad (7)$$

The preferential attachment probability selects nodes for making a connection to the new node based on their current connectivity

$$p_i = \frac{k_i}{\sum_{j=0} k_j}, \quad (8)$$

where k_i is the number of Neighbors directly connected (number of edges) to the existing core node v_i^f . Barabasi and Albert showed that this model produces scale-free networks [11]. The exponential (EXP) networks, also introduced by Barabasi and Albert, follow the same growth rules as the SF networks, except that here the probability of connecting new node to any existing nodes is even

$$p_i = \frac{1}{m}. \quad (9)$$

Barabasi and Albert showed that networks that grow with this rule display connectivity distributions that follow an exponential distribution [11].

Adding one node to a growing network is considered one generation. In each new generation, several alternative networks (N_{alt}) are generated with one additional core node. The fitness of each new network is evaluated by the network's ability to decode correctly a training set which consists of several (i.e., 100) inputs/output known relationships [Eq. (1)] from the "restaurant problem" Boolean function dataset. The fittest network is then selected, and the process is repeated until networks grow to a certain size (i.e., 65 nodes) [Fig. 2(b)]. The size of 65 nodes was chosen because networks stop improving their adaptation ability at this size.

The fitness evaluation uses the count of alternative pathways from each input node to the output node to determine the weight of each input. To determine the fitness two strategies were used: The first strategy considers the links to have a neutral sign. The number of all pathways starting from input nodes with the value 0 is compared to the number of all pathways from input nodes with input value 1. If there are more "1" pathways to the output node, the predicted output will be 1, and if there are more "0" pathways, the predicted output will be 0. Thus, the initial network, before adaptation and growth, is a simple voting function. The second strategy randomly assigns a sign (+ or -) to each link when the link is created. The sum of all positive pathways starting from input nodes with value 1 is added to the sum of all negative pathways from input nodes with input value 0; this sum is compared to the sum of all positive pathways starting from input nodes with value 0 added to the sum of all negative pathways from input nodes with input value 1. If there are more positive pathways from "1" inputs and negative pathways from "0" inputs the predicted output will be 1, and 0 if the opposite is true. Positive pathways in addition to containing positive links may contain an even number of negative links or no negative links. Negative pathways have an odd number of negative links. Thus, the algorithms for determining the output are as follows: The input parameters are a network G [Eq. (2)] and a set of attributes: I .

$$I = \{\text{Attr}_1, \dots, \text{Attr}_n\}. \quad (10)$$

The algorithm computes the class (either 0 or 1) by use of the following procedure:

(1) Count number of unique pathways (walks) W_i from each source nodes to the sink target node

$$W_i = \text{FindUniquePathsFrom}(v_i^f \rightarrow v), \quad i = 1, \dots, n. \quad (11)$$

(2) Count number of positive pathways and negative pathways from each input (this step is only needed for the second implementation)

$$W_i^+ = \text{FindAllPositivePathsFrom}(v_i^f \rightarrow v), \quad i = 1, \dots, n.$$

$$W_i^- = \text{FindAllNegativePathsFrom}(v_i^s \rightarrow v), \quad i = 1, \dots, n. \quad (12)$$

(3) Assign to each source node its matching input attribute

$$I_i^s = \text{Attr}_i. \quad (13)$$

(4) For the first implementation, the total number of pathways with input 0 is compared to the total number of pathways with input 1. If there are more pathways with the input 0 then the Class would be 0, or 1 otherwise; if

$$\left(\sum_{i=0}^n I_i^s \cdot W_i \right) \geq \left(\sum_{i=0}^n \bar{I}_i \cdot W_i \right) \quad (14)$$

then the Class is 1; else it is 0, where \bar{I} is the complement of I .

(5) For the second implementation, the total number of positive pathways with the input 1 and the total number of negative pathways with the input 0 are compared to total negative pathways with input 1 and positive pathways with input 0. If there are more positive pathways with the input 1 and negative pathways with the input 0 then the Class would be 1, or 0 otherwise; if

$$\left(\sum_{i=0}^n I_i^s \cdot W_i^+ \right) + \left(\sum_{i=0}^n \bar{I}_i \cdot W_i^- \right) \geq \left(\sum_{i=0}^n \bar{I}_i \cdot W_i^+ \right) + \left(\sum_{i=0}^n I_i^s \cdot W_i^- \right) \quad (15)$$

then the Class is 1; else it is 0.

As a control, networks that grow under selective pressure were compared to networks that grow randomly. For networks to grow randomly, in each generation one core node is added according to the same rules described above without the fitness and selection process, namely $N_{\text{alt}}=1$. The randomly growing networks are created to demonstrate that the evolutionary selection process produces networks that adapt to the Boolean function better than by chance. Thus, the randomly growing networks are not expected to improve in their adaptation as they grow. Randomly growing networks are growing according to the previously proposed methods without adaptation such that their topology can be compared with networks that adapt.

The topological structural features of the adaptation-driven artificial networks were compared to that of real-world systems abstracted to networks with many inputs and one output. For this comparison a biological mammalian regulatory cellular signaling network and an air traffic control network were used. Signaling pathways thought to exist in mammalian hippocampal CA1 neurons were assembled *in silico* from known binary interactions extracted from biomedical literature to form a connections map [19]. This abstract representation considers biochemical components, such as proteins, as nodes and binding interactions or enzymatic reactions as links. Links represent direct molecular interactions and are associated with an effect. Thus, the network is a partially directed graph with signs on the edges (+ positive, - negative, or 0 neutral). Neutral links are undirected links where the source/target relationship is not direc-

tional. Alternative pathways from different extracellular ligands (“inputs”) to the transcription factor cAMP response element binding protein (CREB), were counted if the ligand can reach CREB in a maximum of seven steps through direct intermediate cellular components (seven nodes and six links) or less. The transcription factor CREB was chosen because it is a major driver of activity dependent plasticity in these neurons. See Ref. [20] for the dataset.

An air traffic control network was constructed from the FFA (Federal Aviation Administration) NFDC (National Flight Data Center) Preferred Routes Database [21]. This network contains 1226 nodes and 2615 links. A list of airport codes was downloaded from Ref. [22]. A subset list of airport acronyms was created by choosing airport codes representing airports from U.S. states on the Atlantic coast from Maine (north) to Florida (south). The subset list contains 148 airports. These airports were used as source nodes to count potential routes to the destination LAX (Los Angeles International Airport), a major U.S. airport hub on the Pacific (west) coast. The route strings columns in the database were broken into pairs, representing links in the network. Nodes in this network are either airports or flight service stations that are used to direct small airplanes to follow preferred routes specified by the NFDC. Potential routes to LAX from east coast airports were limited to maximum seven hops to reach LAX.

III. RESULTS

All three models adapt to decode correctly about 80% of the examples after approximately 30 growth generations with the first implementation [Fig. 3(a)] and about 85–90 % with the second implementation [Fig. 3(b)], whereas the randomly growing networks fail to adapt (decode answers with about 50% correctness, as expected) [Fig. 3(c)]. The adaptation capability of all three model networks does not improve after approximately 30 generations while the networks continue to grow.

Several statistical measurements were obtained in order to quantitatively characterize the growing networks: Characteristic path length (CPL) [4] is the average length of the shortest path between any two pairs of nodes, and it measures the global cohesiveness of networks. The second measure, clustering coefficient (CC) [4], is the average fraction of links connecting node’s neighbors, out of all possible “intra-neighbors” links. It measures triangles in the network structure. For all three models, and with the two different implementations, the CPL are not significantly affected by growth under selective pressure and they are the same as in the randomly growing networks [Figs. 4(a)–4(c)]. This implies that the intercluster connectivity, or “global sparseness” of networks is not affected by adaptation.

The CC are similar for the adapted networks and for the randomly growing SF and EXP networks [Figs. 5(a) and 5(b)], but a higher CC are observed for the DD model adapting networks [Fig. 5(c)]. This implies that a high CC in DD networks may help to achieve better decoding of the Boolean function. The increase in the CC due to adaptation is only observed for the DD model implementation. This observa-

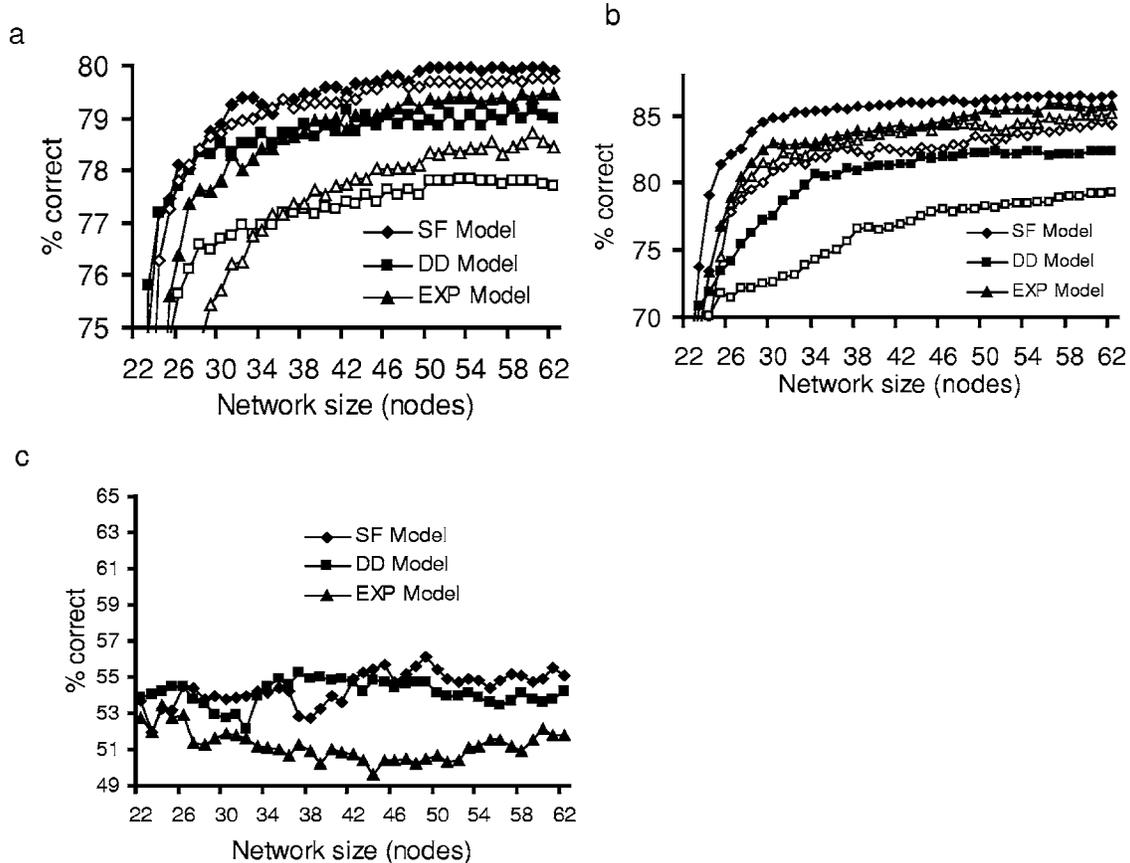


FIG. 3. Fitness of the network to the restaurant problem decision tree as function of network size. (a) First implementation (unsigned links) given $N_{alt}=50$ alternative steps (empty symbols) or $N_{alt}=500$ (filled). (b) Second implementation. (c) Fitness measurements in networks growing without fitness test ($N_{alt}=1$).

tion can be explained by the advantage of having one or two highly connected nodes (hubs). Hubs can grow in the DD model by repeated selection of the same node to be duplicated for many generations. Since the selected node is duplicated with its links, and a link to the new node is added, many triangles are formed, and the CC increases. The new link has to be added between the new copy and the old one in order to make a highly connected hub. Since the probability to add this link is $p=0.1$, and the probability to select this specific node for duplication is $1/m$, growing networks without adaptation have low probability to form as many triangles and maximally enlarge the hubs. The required number of alternative networks (N_{alt}) which guarantees hub formation is of order m/p . This explains also the need of many alternative examples in the DD model for getting high adaptation [Fig 3(b)]. On the other hand, in the EXP model, alternative networks that extend the mostly connected nodes are created randomly in $3/m$ of the alternative configurations. Thus, the required number of examples is of order $m/3$, and $N_{alt}=50$ gives similar results to those of $N_{alt}=500$. In the SF model, the chance of choosing a hub for being linked to the newly created one is even higher. The hubs in the SF and EXP models do not necessarily form triangles and thus do not increase the CC.

The number of pathways from each input node to the output node was counted. Inputs were sorted based on the

number of redundant paths from each input to the output. The most significant difference between the adapted/stressed networks and the nonadapted stressed networks was observed using this measure. For evolving and adapting networks, using the first implementation, the most influential upstream inputs have the most number of pathways to the output. Whereas networks growing without adaptation show flat distributions [Fig. 6(a)]. Interestingly, the DD and EXP model networks grown without adaptation have many more pathways from their inputs to the output [Fig. 6(a)]. This again can be explained by the selection for hubs and a shift toward the SF model topology seen in the adapting the DD and EXP models. Sorting inputs of adapted networks according to the number of routes to output yields a power-law distribution [Fig. 6(b)]. This appears to be true for the artificial networks as well as for the mammalian cellular neuronal hippocampal regulatory network [19] [Fig. 6(c)], and for the air traffic network [Fig. 6(d)]. The inputs (either ligands or airports) were sorted by the number of routes emanating from them to the output node. In both cases few “important” inputs were observed, which are the root of a high number of pathways, and many other inputs of lower importance, forming a power-law distribution. Interestingly, all distributions of pathways best fit a power-law function with a scaling exponent of approximately 1.3. Thus, pathway distribution, which is an unique measure for networks with input/output

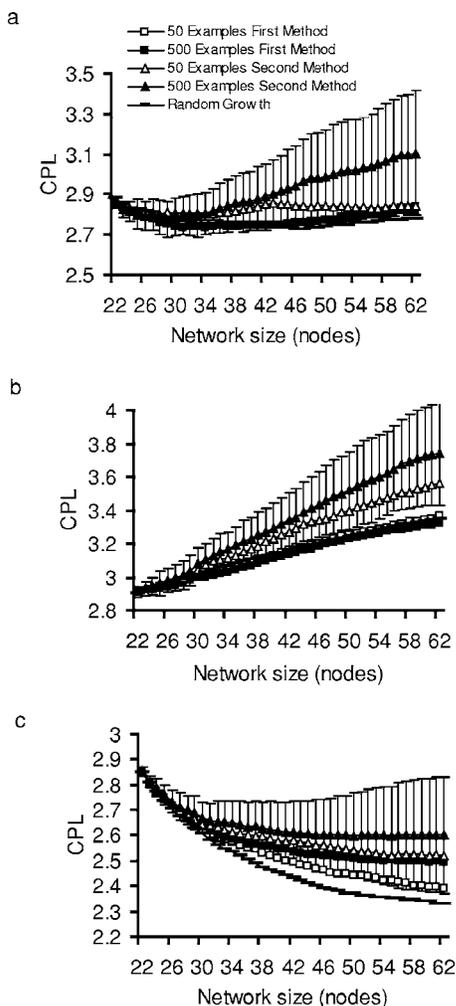


FIG. 4. Characteristic path length (CPL) as function of network size. (a) SF model, (b) EXP model, and (c) DD model. In all cases squares represent first implementation (unsigned links) and triangles represent second implementation (signed links). Empty symbols stand for $N_{alt}=50$ and filled for $N_{alt}=500$. Random growth is shown as well. Error bars relate to filled triangles.

nodes, may indicate the presence of exogenous pressure toward a certain functionality of the network throughout its growth. Additionally, the ratio between positive and negative pathways, in the artificial networks growing with the second implementation [Fig. 6(e)] is similar to the distribution of pathways observed for the signaling network [Fig. 6(f)]. The most influential inputs in the artificial networks (inputs 1 and 2) have more positive pathways than negative pathways to the output. Similarly, pathways from glutamate have the most positive compared to negative signaling pathways to CREB.

IV. DISCUSSION

The analysis presented here suggests that external pressures may affect the topology of real-world networks. Standard network measures such as CC, CPL, and connectivity scaling largely do not capture the effects of external pres-

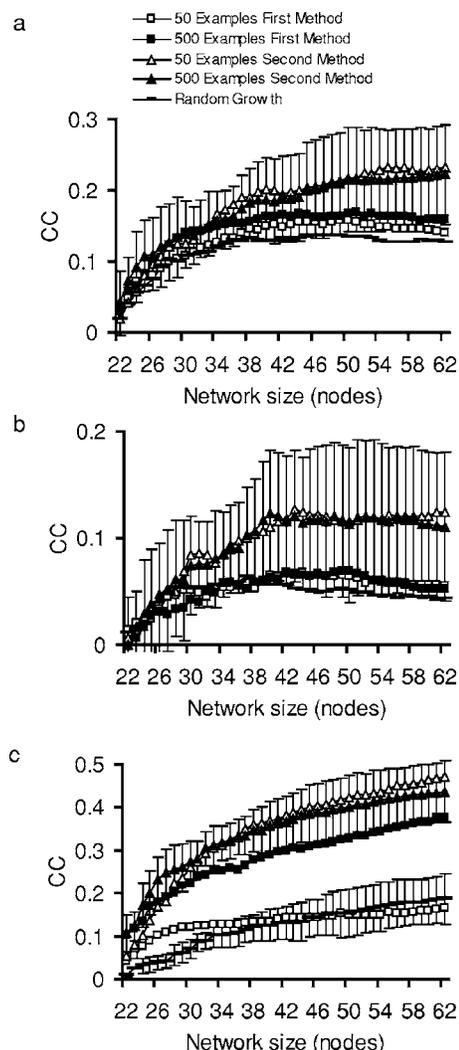


FIG. 5. Clustering coefficient (CC) as function of network size. (a) SF model, (b) EXP model, and (c) DD model. In all cases squares represent first implementation (unsigned links) and triangles represent second implementation (signed links). Empty symbols stand for $N_{alt}=50$ and filled for $N_{alt}=500$. Random growth is shown as well. Error bars relate to filled triangles at (a)–(c) and also to random growth at (c).

ures. Measuring the counts of redundant pathways from inputs to outputs is helpful in identifying the differences between networks that have grown under exogenous pressure versus networks that have grown without external pressure. The results show that the type of growth process used for adaptation is not critical for the adaptation capability. Networks that grow by the rules of all three growth models adapt similarly. This result is surprising and implies that a particular growth mechanism may be less important for shaping the network topology than previously thought. Intuitively, one might have expected that the duplication-divergence model would be the most likely mechanism by which isoforms are generated in biological networks undergoing evolutionary growth. However, SF growth of biological networks has been previously observed in real systems. The yeast protein-protein interactions network has been

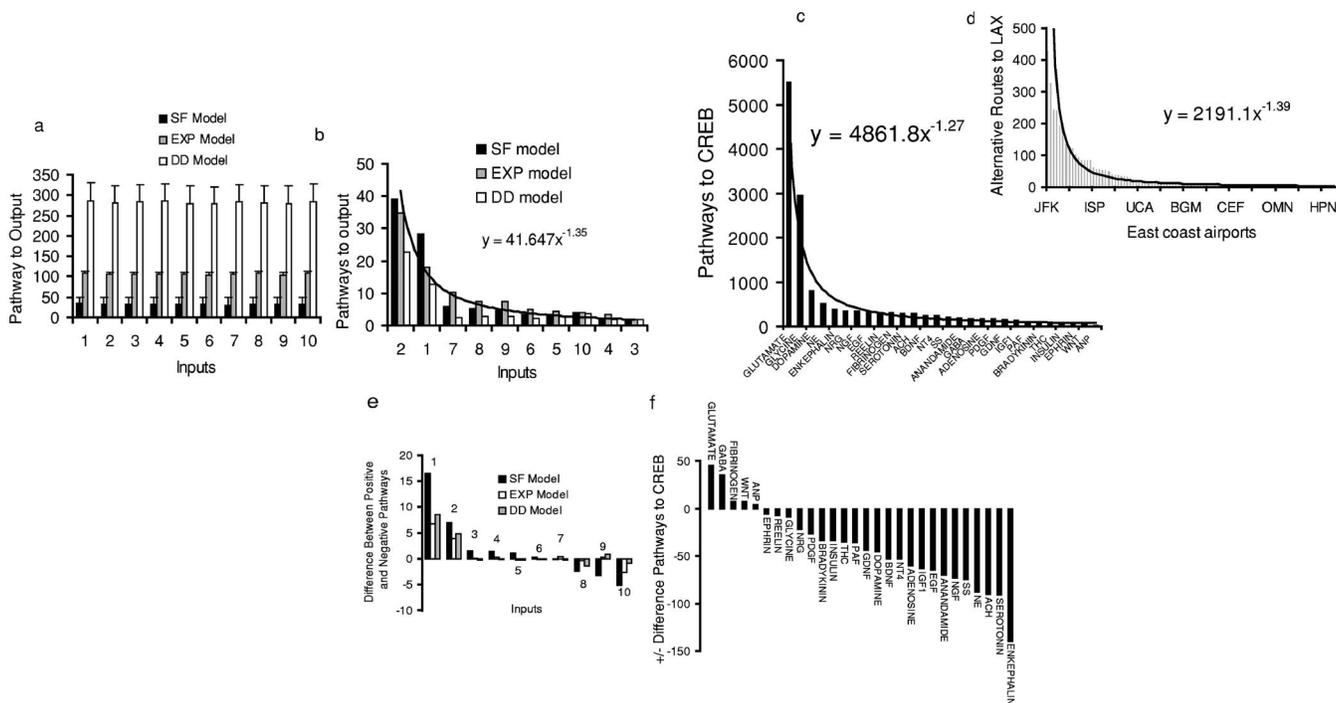


FIG. 6. Counts of pathways to the output. (a) Artificial networks without selection (SF model). Average over 100 simulations. Flat distribution was observed in individual trials. (b) Trained artificial networks (first implementation). Inputs 1 and 2 decode the “Patrons?” question (Fig. 1) and show the most pathways to the output. (c) Pathways from ligands to the transcription factor CREB in hippocampal CA1 neurons. Most pathways emanate from glutamate and glycine. (d) NFDC Air traffic control network. (e) Number of unique positive and negative pathways in the adapted artificial networks using the second implementation. (f) Positive and negative pathways from ligands to CREB in the neuronal hippocampal CA1 neuron.

shown to follow scale-free growth as assessed by the age of components following an evolutionary tree [23]. The results from the evolutionary tree of yeast protein-protein interactions and our study suggest that although biological networks may grow through duplication divergence, the external environmental pressures combined with network growth through evolution induces the formation of hubs, which results in a SF topology.

The results show that pathways in networks with defined inputs and outputs may tend to cluster underneath the most influential inputs. Pathways from these influential inputs reaching target nodes are likely to be overall more positive than negative. Why would such pathway clustering and selection for positive pathways from few inputs occur during adaptation? Perhaps, this is because of the hierarchical ordering of input nodes resulting in some inputs being canalizing. Canalizing inputs determine the output regardless of the other inputs [24]. Concentrating routes from influential inputs may be an outcome of naturally evolving complex systems with canalizing inputs. For example, the first question in the Restaurant Problem is a canalizing locus because if it is answered a certain way, the values of all other inputs do not affect the outcome. In the hippocampal neuronal regulatory network most pathways were found to emanate from glutamate and glycine. These two ligands act as inputs by binding to (NMDA) receptors which allow the entry of calcium into the cell [25]. Glutamate also binds to metabotropic

receptors which initiate G-protein signaling. The combinations of calcium and G-protein signaling with the evidence that the NMDA receptor complex is a highly connected node, containing over 100 potential binding partners [26], makes glutamate and glycine potential canalizing inputs. Analysis of air traffic routes from the east coast of the United States to Los Angeles showed that the most alternative routes originate from the New York, New Jersey, and Connecticut areas. This tristate area is the most densely populated region on the east coast. In the artificially grown networks as well as the two real networks the concentration of pathways under the most influential inputs leads to a power-law distribution of pathways. Thus, it appears that for directed networks with defined inputs and outputs “scale-freeness” exist at multiple levels of organization: at the nodal connectivity level and at the pathway distribution level.

The approach of building simple artificial network models that self-organize in response to external stress may be useful in understanding the origins of the topology of real-world networks [27]. Such computational efforts may be useful in integrating molecular biology with system-level phenotypes to predict the relationship between network topology and phenotypic behavior even when all the nodes and links are not known [28]. Thus, such analysis may lead to predictions of new proteins and interactions and complement other computational strategies used to predict biomolecular interactions [29].

Characterizing networks that have evolved under external stress through a selection process is helpful in identifying preferred network topologies. Previous growth models produce uniform symmetric network topologies, making a new link a function of local connectivity. In this study the rule of adaptation was added to the rules of growth. This addition produces nonsymmetric and more realistic artificial network topologies. Hierarchical influences of inputs lead to scale-free growth and topology. Our analysis of artificial networks suggest that external pressures may have shaped both signaling network in hippocampal neurons, and air traffic patterns in the United States. It would be interesting to see if other complex real-world systems with defined inputs and outputs

show similar redundancy of pathways property of network reorganization.

ACKNOWLEDGMENTS

This research is supported by NIH Grant No. GM-072853. A.M. was supported by NIH training Grant No. GM-62754 and an advanced center grant from NYSTAR. Special thanks to Sherry Jenkins, Roberto Sanchez, and Susan Wearne, from Mount Sinai School of Medicine, New York, NY, and Reka Albert, from Penn State University, PA for critically reading and commenting the manuscript.

-
- [1] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001).
- [2] A. Broder *et al.*, Comput. Netw. **33**, 309 (2000).
- [3] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Commun. Rev. **29**, 251 (1999).
- [4] D. J. Watts, *Small Worlds* (Princeton University, Princeton, NJ, 1999).
- [5] A. L. Barabasi, *Linked* (Plume, Penguin Group, New York, 2003).
- [6] H. Jeong, S. P. Mason, A. L. Barabasi, and Z. N. Oltvai, Nature (London) **411**, 41 (2001).
- [7] B. Schwikowski, P. Uetz, and S. Fields, Nat. Biotechnol. **18**, 1257 (2000).
- [8] L. Giot *et al.*, Science **302**, 1727 (2003).
- [9] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi, Nature (London) **407**, 651 (2001).
- [10] A. Bhan, D. J. Galas, and T. G. Dewey, Bioinformatics **18**, 1486 (2002).
- [11] A. L. Barabasi and R. Albert, Science **286**, 509 (1999).
- [12] A. Vazquez *et al.*, ComplexUs **1**, 38 (2002).
- [13] R. Pastor-Satorras, E. Smith, and R. V. Sole, J. Theor. Biol. **222**, 199 (2002).
- [14] A. R. Puniyani and R. M. Lukose, e-print cond-mat/0107391.
- [15] M. Kaiser and C. C. Hilgetag, Phys. Rev. E **69**, 036103 (2004).
- [16] B. J. Kim *et al.*, Eur. Phys. J. B **43**, 369 (2005).
- [17] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach* (Prentice-Hall, Englewood Cliffs, NJ, 1995).
- [18] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing* (Springer, Berlin, 2003).
- [19] A. Ma'ayan *et al.*, Science **309**, 1078 (2005).
- [20] <http://www.mssm.edu/labs/iyengar/resources>
- [21] The analysis was done on data which was downloaded on February 14th, 2006. The database was downloaded from <http://www.fly.faa.gov>
- [22] <http://www.airnav.com/airports>
- [23] E. Eisenberg and E. Y. Levanon, Phys. Rev. Lett. **91**, 138701 (2003).
- [24] S. A. Kauffman, *The Origins of Order* (Oxford University, New York, 1993).
- [25] J. W. Johnson and P. Ascher, Nature (London) **325**, 742 (1987).
- [26] H. Husi *et al.*, Nat. Neurosci. **3**, 661 (2000).
- [27] N. Kashtan and U. Alon, Proc. Natl. Acad. Sci. U.S.A. **102**, 13773 (2005).
- [28] T. Ideker, T. Galitski, and L. Hood, Annu. Rev. Genomics Hum. Genet. **2**, 343 (2001).
- [29] H. Ge *et al.*, Nat. Genet. **29**, 482 (2001).