

Uncorrected Proof
Copy

31

Complexity of Protein–Protein Interaction Networks, Complexes, and Pathways

Danail Bonchev

1. Introduction

1.1. *The Network Approach to Proteomics*

The focus of proteomic research in developing experimental techniques for protein identification and interaction studies is shifting from individual proteins to their organization in reaction pathways, complexes, and networks, i.e., to the proteome—the large-scale network comprising all protein–protein interactions in a cell, tissue, or organism. The number of complete proteomes in accessible databases exceeds 100 (1), thus making possible proteome-wide and across-proteomes analyses. Such a systemic approach offers a view of the biological machine as a whole, revealing important new details of its work. Thus, one could regard aging and diseases as specific patterns of protein network degradation and, vice versa, evolutionary beneficial factors as creating patterns of larger proteome complexity. Medicines' side effects could be analyzed in terms of the extremely high network connectivity, thus orienting the search for new medicines toward protein complexes, rather than individual compounds (2). Potential drug and marker candidates could be identified proceeding from protein connectivity and centrality patterns.

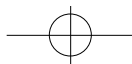
1.2. *Quantifying Networks*

Network analysis and applications are necessarily related to numbers that uniquely characterize each network, making possible comparison, classification, structure-activity/toxicity relationships, and prediction. Regarding networks as graphs (3), one can use graph theory (4,5) to generate such numbers, usually termed *topological indices* (6,7), proceeding from different graph invariants. A variety of topological indices have been proposed in an area of theoretical chemistry called *mathematical chemistry*. From this rich arsenal of descriptors we select for characterizing proteomic networks only a few representatives that mirror patterns of *network topological complexity*. Information theory (8,9) is also used as a source of descriptors of *network diversity and composition*.

1.3. *The Concept of Network Complexity*

Two concepts of complexity as a general property of systems have been developed during the last half century. The concept of new phenomena emerging in a highly com-

Uncorrected Proof



Uncorrected Proof
Copy

plex system focuses on the nonlinearity of processes in complex dynamic systems (10). In this chapter, we present the lesser known concept of *topological (or structural) and compositional complexity*. The roots of this concept can be traced back to the ideas of the information content of a system (a molecule, a cell, an organism) advanced in the 1950s (11), to the first estimates of graph complexity (12), and to the later hierarchical concepts of complexity (13–15).

Definition 1. The larger the number and connectivity of the subnetworks, the larger the network topological complexity.

Definition 2. The more diverse the distribution of the network elements, the larger the network compositional complexity.

Definition 1 proceeds from counting all subgraphs and the links within each of them. It quantifies the idea that the whole is more than its parts and provides nonlinear quantitative measures of network structural complexity. Definition 2 accounts for the different aspects of network complexity, based on the element's distribution, properties, weights, interactions, and so on. When applied to dynamic evolutionary networks, the definitions presented might be viewed as a step toward unifying the two alternative approaches to systems complexity.

2. Some Basic Notions from Graph Theory

Protein–protein interaction networks are presented as graphs (4,5). A *graph* is a structure composed of points (*vertices* or *nodes*), connected by lines (*edges* or *links*). A *sub-graph* is a graph obtained from the parent graph by deleting at least one edge or vertex with its incident edges. A *loop* is an edge that begins and ends in the same vertex. A *multigraph* is a graph in which some pairs of vertices are linked by more than one edge. *Simple graphs* are graphs having no multiple edges and loops. In a *complete graph* any two vertices are connected by an edge. A *directed graph* is a graph having at least one directed edge. Directed edges are termed *arcs*. The graph is *connected* when there is a path between any pair of vertices in it; otherwise the graph is *disconnected*. A *path* in the graph is a sequence of adjacent edges without traversing any vertex twice. A *walk* is an alternating sequence of vertices and edges, each of which could be traversed more than once. The *walk length* is the number of edges in it. A *cycle* is a path that starts from and ends in the same vertex. *Trees* are graphs containing no cycles. Two vertices j and i are called *adjacent* when they are connected by an edge $\{i, j\}$. The adjacency relation is quantified by the term $a_{ij} = 1$. *Graph components* are connected subgraphs or vertices that are not connected to each other. Euler's theorem relates the number of vertices V , edges E , cycles C , and components K :

$$C = E - V + K \quad (1)$$

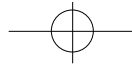
Illustrations of the notions introduced above are given in **Fig. 1**.

Definition 3. *Vertex degree* a_i is the number of edges $\{i, j\}$ connecting vertex i with its adjacent vertices j (denoted as $j \leftrightarrow i$):

$$a_i = \sum_{j \leftrightarrow i} a_{ij} \quad (2)$$

In a multigraph one may distinguish between *minimal* vertex degree, for which $a_{ij} = 1$, and *multiple* vertex degree, which accounts for all links to the nearest neighbors ($a_{ij} \geq 1$). In directed graphs, *in-degree* and *out-degree* are defined as the number of arcs entering into and emanating from the vertex, respectively.

Uncorrected Proof



Uncorrected Proof
Copy

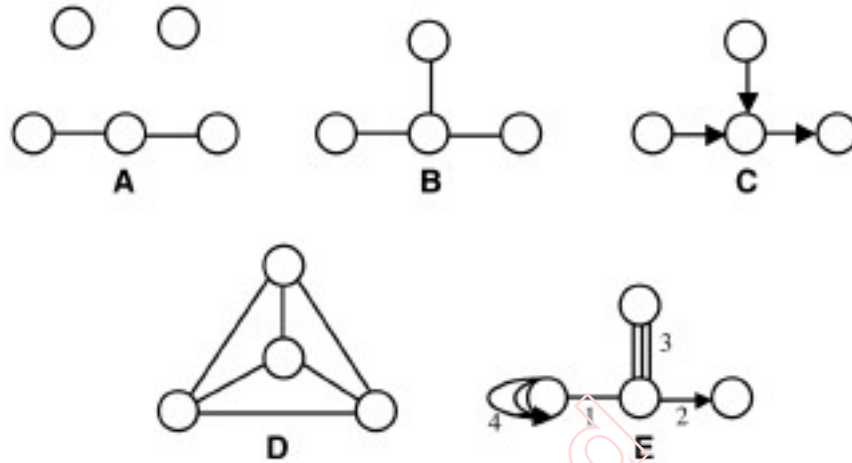


Fig. 1. (A) A disconnected graph with three components. (B) A simple connected graph. (C) A directed graph. (D) A complete graph with three cycles. (E) a multigraph with loops. 1, edge; 2, arc; 3, triple edge; 4, two loops.

Definition 4. Graph adjacency $A(G)$ is the sum of all vertex degrees a_i or all vertex adjacencies a_{ij} . In directed graphs, the total adjacency $A(DG)$ is the sum of the respective in- and out-degrees:

$$A(G) = \sum_{i=1}^V a_i = \sum_{i=1}^V \sum_{j \leftarrow i} a_{ij} \quad A(DG, out) = \sum_{i=1}^V a_i(out) = \sum_{i=1}^V \sum_{j \leftarrow i} a_{ij}(out) \quad (3)$$

Graph invariants are numbers uniquely derived from the graph; they do not depend on the way the graph is drawn or on the manner in which the vertices are labeled. Graph adjacency is a graph invariant, whereas vertex degree is a vertex (or local) invariant.

Definition 5. (a) The distance d_{ij} between the vertices i and j is the number of edges along the shortest path connecting i and j . (b) The vertex distance degree d_i is the sum of distances between vertex i and all other vertices j in the graph:

$$d_i = \sum_{j \neq i} d_{ij} \quad (4)$$

Definition 6. The graph distance $d(G)$ is the sum of distances between all pairs of vertices i and j ; it is also a half sum of the distance degrees of all V vertices in the graph:

$$d(G) = \sum_{i=1}^V \sum_{j \neq i} d_{ij} = \frac{1}{2} \sum_{i=1}^V d_i \quad (5)$$

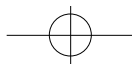
In nondirected graphs G , $d(G) = 2W(G)$, $W(G)$ being the Wiener number of G (16). In directed graphs DG , $d(DG) = W(DG)$.

Definition 7. (a) Vertex eccentricity e_i is the maximum distance between vertex i and the remaining graph vertices. (b) Graph center is the vertex(es) having minimum eccentricity and minimum distance degree (see ref. 17 for a more detailed definition):

$$e_i = \min; \quad d_i = \min \quad (6)$$

Illustrations of Definitions 3–7 are given in Fig. 2.

Uncorrected Proof



Uncorrected Proof
Copy

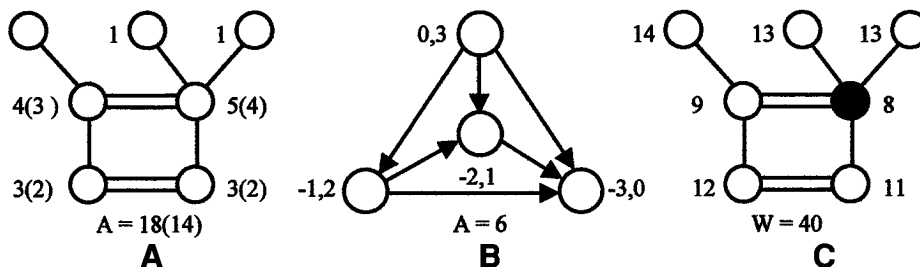


Fig. 2. (A) Vertex degrees and adjacency of a multigraph; the minimal degrees are given in parentheses. (B) In- and out-degrees and adjacency of a directed graph; in-degrees are denoted by a minus sign. (C) Vertex distance degrees and the Wiener number (16) of a graph. The filled vertex having eccentricity 2 and distance degree 8 is the graph center.

3. Topological Complexity Measures

For a review of the entire area of complexity descriptors, the reader is addressed to ref. 15. Here, we focus on several complexity measures that satisfy the requirements to increase with the increase in network size and connectivity, to vary regularly with the topological patterns of branching (18), cyclicity (19), centrality (17), and clustering (18,19), and to be in compliance with the idea of biological complexity regarding the whole as more than the sum of its parts.

3.1. Connectedness (Connectance)

Definition 8. Connectedness *Conn* is the ratio between the number of edges *E* in the network and the number of edges in the complete graph having the same number of vertices *V*:

$$Conn(\%) = \frac{2E}{V(V-1)} \times 100 \tag{7}$$

For simple graphs (graphs having no loops or multiple edges), *Conn* varies from zero for a totally disconnected graph having *E* = 0, to 100 for a complete graph. An example is given in Fig. 3A. For multigraphs, *Conn* could exceed 100. In such cases one may regard separately the multiple connectedness *Multiconn* from the simple one, which measures only whether the pairs of vertices are connected but not how many edges connect them. Similarly, for characterizing the local (or vertex) complexity one may use the vertex degrees and multiple vertex degrees introduced in Subheading 2. The factor 2 in Eq. 7 is omitted for directed graphs. Connectedness can be used as a preliminary estimate of topological complexity; it is not sensitive to variations in topology of networks having the same number of vertices and edges.

3.2. The Substructure Count SC

See refs. 20–22.

Definition 9. (a) The complexity index *SC* is the total number of subgraphs in the graph (Fig. 3). (b) The *e*th order index ^{*e*}*SC* is the count of all subgraphs having *e* edges:

$$SC = {}^0SC + {}^1SC + {}^2SC + \dots + {}^E SC = \sum_{e=0}^E {}^e SC \tag{8}$$

Uncorrected Proof

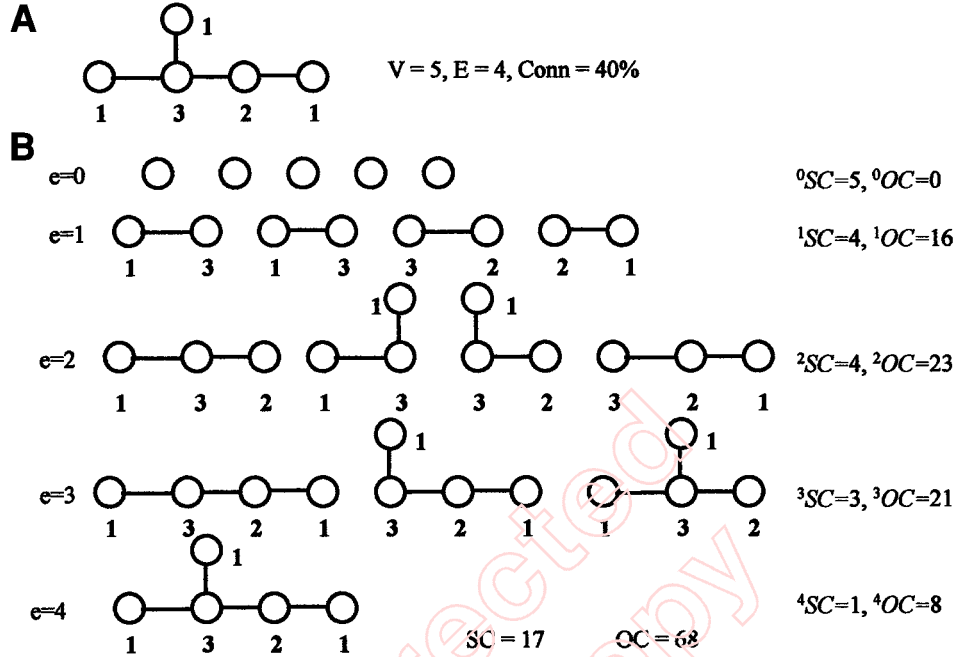
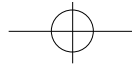


Fig. 3. (A) Calculation of connectedness according to Eq. 7. (B) Illustration of subgraph count (20–22) and overall connectivity indices (22,23) calculation (Eqs. 8 and 9).

where 0SC , 1SC , and 2SC are the number of vertices V , edges E , and two-edge subgraphs, respectively, and ${}^E SC = 1$ stands for the graph itself (Fig. 3B).

3.3. Overall Connectivity

See refs. 22 and 23.

The overall connectivity concept (Definition 1) combines subgraph count with subgraph connectivity.

Definition 10. (a) The overall connectivity index OC is the sum of total adjacencies A_i of all subgraphs G_i belonging to the graph G . (b) The e th order overall connectivity term ${}^e OC$ is the sum of the total adjacencies ${}^e A_i$ of all subgraphs ${}^e G_i$ having e edges (Fig. 3B).

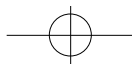
$$\begin{aligned}
 OC &= {}^1OC + {}^2OC + \dots + {}^E OC = \sum_{e=1}^E {}^e OC \\
 {}^e OC &= \sum_{i=1}^{e SC} {}^e A_i ({}^e G_i) = \sum_{i=1}^{e SC} \sum_{j=1}^{V_i} a_j (j \in {}^e G_i)
 \end{aligned} \tag{9}$$

3.4. The Graph Walk Count

Another realization of the idea to characterize complexity by regarding the structure as a whole is to account for all walks within it (24–26; see also Chap. 2 in ref. 15).

Definition 11. (a) The graph walk count TWC is the total number of walks ${}^l WC$ of length $l = 1$ to $V - 1$. (b) The walk count of order l , ${}^l TWC$ is the number of walks of length l . (c) The vertex walk count vwc_i is the total number of walks of length l that start in the vertex i .

$$TWC = {}^1 WC + {}^2 WC + {}^3 WC + \dots + {}^{V-1} WC = \sum_{l=1}^{V-1} {}^l WC = \sum_{i=1}^V vwc_i = \sum_{l=1}^{V-1} \sum_{i=1}^V {}^l vwc_i \tag{10}$$



Uncorrected Proof
Copy

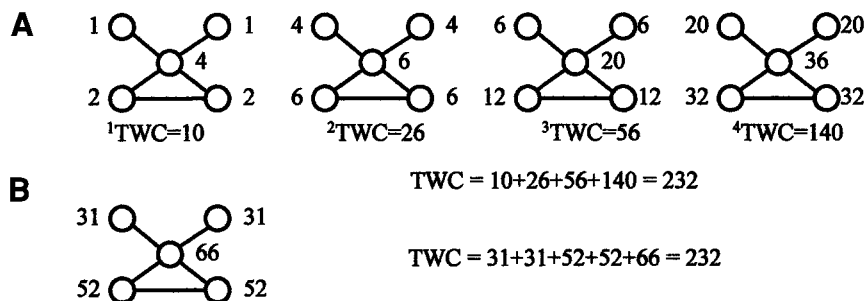


Fig. 4. (A) Calculation of the total walk count TWC (24–26) as a sum of walk counts of lengths 1 to $V - 1$. (B) Calculation of TWC from the vertex walk count sums (Eq. 10).

An example is shown in Fig. 4. A simple method for calculating vertex walk counts and total walk count is based on the Morgan algorithm. Starting with vertex degrees, one calculates the count of walks of length 1 as the sum of vertex degrees, and then an *extended degree* of l th order is calculated for each vertex by summation of the $(l - 1)$ th order degrees of the vertex neighbors.

4. Compositional Complexity Measures

Diversity of network composition is another aspect of complexity (see Definition 2). Graph representation can incorporate such an aspect by assigning weights to graph vertices; however, additional insight can be gained by Shannon’s information theory (8).

4.1. Shannon’s Basic Formulas as Applied to Finite Networks

See refs. 8, 9, and 27.

Consider a network composed of N elements, distributed into k classes, according to a certain equivalence criterion α . The elements could be network vertices, edges, or any other type of subgraphs, as well as distance-based invariants. Denote the number of elements in classes 1, 2, . . . , k as N_1, N_2, \dots, N_k , respectively. The probability for a single randomly chosen network element to belong to the class i is $p_i = N_i/N$, where $N = \sum N_i$, and $\sum p_i = 1$.

Definition 12. Shannon’s entropy $H(\alpha)$ of the network distribution $\{N_1, N_2, \dots, N_k\}$ is

$$H(\alpha) = N \log_2 N - \sum_{i=1}^k N_i \log_2 N_i \text{ bits} \tag{11}$$

$$\bar{H}(\alpha) = - \sum_{i=1}^k p_i \log_2 p_i \text{ bits/element} \tag{12}$$

Here, base 2 logarithms are used for calculating entropy in binary digits (bits). Entropy is maximum when $N_i = 1$, and $p_i = 1/N$:

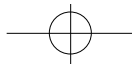
$$H_{\max}(\alpha) = N \log_2 N; \quad \bar{H}_{\max}(\alpha) = \log_2 N \tag{13}$$

Definition 13. The information content $I(\alpha)$ of network probability distribution, the average and the normalized information content, $\bar{I}(\alpha)$ and $I_n(\alpha)$, are respectively

$$I(\alpha) = H_{\max}(\alpha) - H(\alpha) = \sum_{i=1}^k N_i \log_2 N_i; \quad \bar{I}(\alpha) = \frac{1}{N} \sum_{i=1}^k N_i \log_2 N_i \tag{14}$$

$$I_n(\alpha) = \frac{1}{N \log_2 N} \sum_{i=1}^k N_i \log_2 N_i; \quad 0 \leq I_n(\alpha) \leq 1$$

Uncorrected Proof



Uncorrected Proof Copy

Different information-theoretic indices can be introduced depending on the specific equivalence relation α . Such a relation for proteins in a proteome network means they belong to a certain protein complex or to a certain network component. The corresponding information indices of compositional complexity may be termed *information on proteome complexes*, $I(\text{complexes})$, and *information on proteome components*, $I(\text{component})$, respectively.

The Shannon entropy/information measures based on equivalence of system elements were previously shown (27) to fail in evaluating structural or topological complexity. Thus, proceeding from vertex equivalence, one obtains the same null information content for three graphs with very different complexity: a totally disconnected, a single-cycle, and a complete graph. A modification of the theory captured many of the complexity features (9,28). According to this modification, each system element is assigned a weight, uniquely derived from its topology. Then, the distribution of the total graph weight M into weights M_i of individual elements i (vertices, edges, two-edge sub-graphs, and so on) is also a probability distribution with the probability of a randomly chosen element to have weight M_i equal to $p_i = M_i/M$.

Definition 14. Weighted information content ${}^M I$, of the network elements distribution, average and normalized weight information content, ${}^M \bar{I}$ and ${}^M I_n$, are respectively

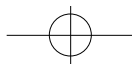
$${}^M I = \sum_{i=1}^m M_i \log_2 M_i; \quad {}^M \bar{I} = \frac{1}{M} \sum_{i=1}^m M_i \log_2 M_i; \quad {}^M I_n = \frac{1}{M \log_2 M} \sum_{i=1}^m M_i \log_2 M_i \quad (15)$$

Typical network “weight” distributed among its elements is the total adjacency A as partitioned into vertex degrees a_i (Eqs. 2 and 3). Similarly, the total network distance $d(G)$ (Eq. 5) is considered as composed of vertex distance degrees d_i or, alternatively, of the distance values $d(i) = 1, 2, \dots, d_{max}$. The information indices thus constructed are termed *vertex degree information index*, I_{vd} , *distance degree information index*, I_{dd} , and *distance information index*, I_d , respectively.

5. Examples of Complexity Assessments of Protein–Protein Interaction Networks, Complexes, and Pathways

5.1. Networks Presented by Nondirected Graphs

Nondirected graphs and used to represent protein–protein interaction network, the direction of interaction is not known or disregarded. Also included here are networks the nodes of which are protein complexes, and an edge between two nodes indicates the presence of a protein common to the two complexes. An example is the functional net of membrane biogenesis and traffic in the *Saccharomyces cerevisiae* proteome extracted from the data of Gavin et al. (29), (Fig. 5). The network contains 147 proteins organized into 20 complexes. The corresponding compositional distribution is $147\{20, 15, 13, 13, 10, 10, 5 \times 7, 5, 5, 3 \times 4, 3, 3 \times 2\}$, and Eq. 14 yields $I(\text{complexes}) = 463.9$ bits, $\bar{I} = 3.16$, and $I_n = 0.438$. The 20 *protein complexes* are distributed into 10 components, 1 large including 11 complexes, and 9 single-complex ones: $20\{11, 9 \times 1\}$, from whence $I(\text{component}) = 38.05$, $\bar{I} = 1.903$, and $I_n = 0.440$. The *protein* distribution into components $147\{94, 10, 10, 3 \times 7, 4, 4, 2, 2\}$ differs considerably from the even distribution, which produces high values of information indices: $I(\text{component}) = 761.5$, $\bar{I} = 5.18$, and $I_n = 0.720$. Thus, the presence of a very large component, typical for biological networks, can be related to a trend toward higher information content.



Uncorrected Proof
Copy

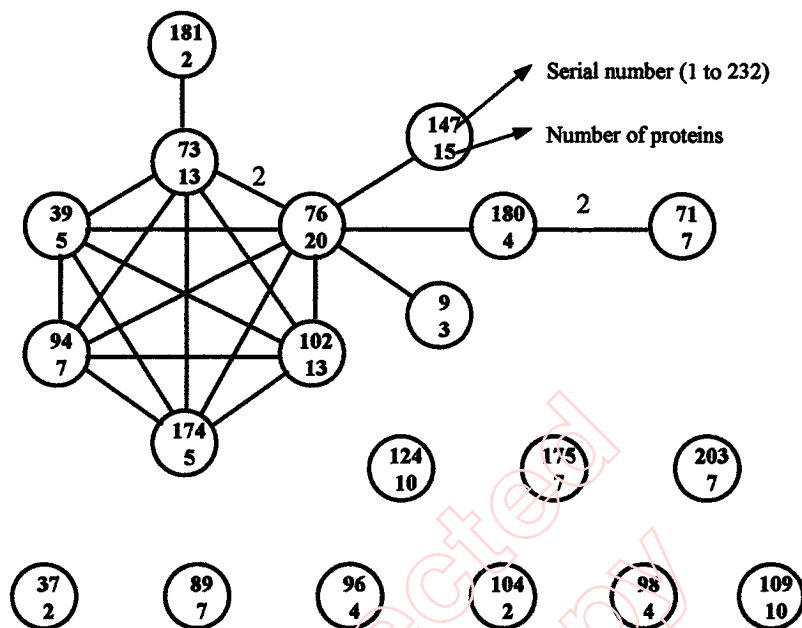
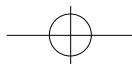


Fig. 5. The membrane biogenesis and traffic network in yeast proteome. Each node represents a protein-protein interaction complex, each edge stands for a protein shared between two complexes. Edge weight denotes the number of proteins shared. (Data from ref. 29.)

The large component of the network contains 20 edges, 2 of which are double edges. Connectedness is quite high: $(2 \times 22 \times 100)/(11 \times 10) = 40\%$, owing to the presence of a complete six-vertex subgraph. The vertex degree distribution $\{9, 7, 4 \times 5, 3, 2, 3 \times 1\}$ obeys a power law to a good approximation (30,31). The central point (17) of the network is complex 76, which shares nine proteins with eight other complexes and reaches to all other complexes via only one or two steps. Such central points are regarded as potential marker/drug candidates. They are of crucial importance for the stability of the network.

The Wiener number of the large component is $W = 102$, and the average distance in it is only $\langle d \rangle = (102 \times 2)/(11 \times 10) = 1.85$. Thus, each pair of complexes in the functional set is connected on average by less than two links, a typical manifestation of a “small-world” network (32), which is another feature of biological networks. The distance distribution $\{20 \times 1, 24 \times 2, 10 \times 3, 1 \times 4\}$ is not very different from the least complex even distribution, as demonstrated by the calculated low value of the information index $I_{d,n} = 0.152$. The distance degree distribution $\{28, 24, 21, 21, 19, 4 \times 16, 15, 12\}$ is considerably more diverse, placing the corresponding information index $I_{dd,n} = 0.554$ in the middle of the 0 to 1 complexity scale.

The three topological complexity measures, when applied to the connected component regarded as a simple graph, yield very high values ($SC = 729,449$; $OC = 26,594,270$; $TWC = 122,082,804$). It would suffice for characterizing the complexity of large-scale biological networks to use the first several terms of these indices, eSC , eOC , lTWC , with e or $l = 1$ to 3 or 4. Such terms are, for example, the number of two-edge subgraphs, known as the Platt index (33), ${}^2SC = Pt$, and that of the three-edge subgraphs, 3SC . For



Uncorrected Proof
Copy

Complexity of Protein Networks

459

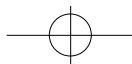
the large component of the membrane biogenesis and traffic network regarded as a multigraph, one obtains $Pt = 99$ and ${}^3SC = 518$, respectively. Other examples are ${}^1OC = 246$, ${}^2OC = 1714$; ${}^2TWC = 248$, and ${}^3TWC = 1338$. All these descriptors can be used for quantitative comparative studies in proteomics, as well as for network structure-activity/toxicity relationships (NSAR). Another application is to quantify the degradation of the network as a result of a sickness or environmental effects. As an example, consider the degradation of the membrane biogenesis and traffic network resulted from the elimination of the central complex 76. The connected component of the network decomposes into a subgraph of six vertices, another subgraph of two vertices, and two isolated vertices. The dramatic loss of network complexity can be illustrated with the decrease in the corresponding complexity descriptors: *Conn*—from 40 to 28.9%; the Wiener index—from 102 to 20; the Platt index—from 99 to 35; 1OC —from 246 to 98; and 2OC —from 1724 to 422. The information index on network components also reduces strongly to half its initial value ($38.05 \rightarrow 17.51$).

5.2. Networks Presented by Directed Graphs

The direction of arcs in the graphs representing protein-protein interaction networks is assumed from the bait protein to the interacting partner. As an example, consider the DNA damage response network in yeast proteome (data taken from Ho et al. (34); **Fig. 6**). The network includes 76 proteins organized in five components, a large one with 59 proteins, and four small ones having 7, 4, 4, and 2 proteins, respectively. This highly uneven distribution produces a normalized compositional complexity index of 0.810, close to the upper limit of 1.0. The directed interactions prohibit many paths between proteins thus diminishing network topological complexity. The network descriptors are calculated as the sum of the respective values for all components, their values remaining close to the values of the large dominant component. Thus, for the entire DNA damage response network and its large component, one obtains a connectedness of 3% versus 4.2%, a Platt index of 72 versus 57, and a Wiener number of 273 versus 230. The in and out first-order overall connectivities of the complex are 316 versus 274 and 316 versus 266, respectively; the second-order overall in-connectivities are 1503 versus 1381, and the out ones are 1302 versus 1152. The total walk count of lengths two and three is 72 versus 57, and 52 versus 35, respectively.

The vertex degrees in directed networks also could not reach very high values, 10 being the highest value of in-degree in the large component (however, this indicates that protein Dun1 could be a good potential marker/drug target) versus only 6 for the out-degree. This makes vertex degree distribution less complex, and closer to the even distribution, as witnessed by the low values (0.210 for out-degrees versus 0.177 for the in-degrees) of the corresponding information indices. In contrast, the distance degree distribution is considerably more uneven, owing to the hindered protein-protein communication. Thus, the in-degree distribution in the large component is $\{16 \times 1, 3 \times 2, 8 \times 3, 3 \times 4, 2 \times 5, 3 \times 6, 2 \times 7, 10, 11, 17, 23, 31, 38\}$. However, the presence of 16 distance degrees equal to unity reduces the complexity of the distribution, and the corresponding normalized information index $I_{dd}(in)$ has the medium range value of 0.429. For comparison, if the DNA damage response network were nondirected, the average distance in the large component would increase from 3.90 to 4.53 (still remaining a small-world network), the maximum distance would go up from 4 to 11, and distance

Uncorrected Proof



Uncorrected Proof
Copy

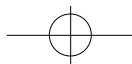
Acknowledgments

The comments and suggestions made by Dr. Christoph Rucker (Bayreuth, Germany) are gratefully acknowledged.

References

1. www.ebi.ac.uk/proteome/.
2. Vogelstein, B., Lane, D., and Levine, A. J. (2000) Surfing the p53 Network. *Nature* **408**, 307–310.
3. Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. (2002) The KEGG Databases at GenomeNet. *Nucleic Acids Res.* **30**, 42–46; see also KEGG Encyclopedia, <http://fire2.scl.genome.ad.jp/kegg/kegg2.html>.
4. Harary, F. (1972) *Graph Theory*. Addison-Wesley, Reading, MA.
5. Trinajstić, N. (1992) *Chemical Graph Theory*. CRC Press, Boca Raton, FL.
6. Todeschini, R. and Consonni, (2000) *V. Handbook of Molecular Descriptors*. Wiley Europe, Weinheim.
7. Devillers, J. and Balaban, A. T. (eds.) (1999) *Topological Indices and Related Descriptors in QSAR and QSPR*. Gordon & Breach, Reading, UK.
8. Shannon, C. and Weaver, W. (1949) *Mathematical Theory of Communications*. University of Illinois Press, Urbana, IL.
9. Bonchev, D. (1983) *Information-Theoretic Indices for Characterization of Chemical Structures*. Research Studies Press, Chichester, U. K.
10. Nicolis, G. and Prigogine, I. (1989) *Exploring Complexity*. Piper, Munich.
11. Rashewsky, N. (1955) Life, information theory, and topology. *Bull. Math. Biophys.* **17**, 229–235.
12. Mowshowitz, A. (1968) Entropy and the complexity of graphs. I. An index of the relative complexity of a graph. *Bull. Math. Biophys.* **30**, 175–204.
13. Bertz, S. H. (1983) A mathematical model of molecular complexity, in *Chemical Applications of Topology and Graph Theory* (King, R. B., ed.), Elsevier, Amsterdam, pp. 206–221.
14. Bonchev, D. and Polansky, O. E. (1987) On the topological complexity of chemical systems, in *Graph Theory and Topology in Chemistry* (King, R. B. and Rouvray, D.H., eds.), Elsevier, Amsterdam, pp. 126–158.
15. Bonchev, D. and Rouvray, D. H. (eds.) (2003) *Mathematical Chemistry*, vol. VII, *Complexity in Chemistry*. Taylor and Francis, London, UK.
16. Wiener, H. (1947) Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **69**, 17–20.
17. Bonchev, D. (1989) The concept for the center of a chemical structure and its applications. *Theochemistry* **185**, 155–168.
18. Bonchev, D. (1995) Topological order in molecules. 1. Molecular branching revisited. *Theochemistry* **336**, 137–156.
19. Bonchev, D., Balaban, A. T., Liu, X., and Klein, D. J. (1994) Molecular cyclicity and centrality of polycyclic graphs. I. Cyclicity based on resistance distances or reciprocal distances. *Int. J. Quantum Chem.* **50**, 1–20.
20. Bertz, S. H. and Herndon, W. C. (1986) The similarity of graphs and molecules, in *Artificial Intelligence Applications to Chemistry* (Pierce, T. H. and Hohne, B. A., eds.). ACS, Washington, DC, pp. 169–175.
21. Bertz, S. H. and Wright, W. F. (1998) The graph theory approach to synthetic analysis: definition and application of molecular complexity and synthetic complexity. *Graph Theory Notes NY Acad. Sci.* **35**, 32–48.

Uncorrected Proof



Uncorrected Proof
Copy

22. Bonchev, D. (1997) Novel indices for the topological complexity of molecules. *SAR QSAR Environ. Res.* **7**, 23–43.
23. Bonchev, D. (2000) Overall connectivities/topological complexities: a new powerful tool for QSPR/QSAR. *J. Chem. Inf. Comput. Sci.* **40**, 934–941.
24. Rücker, G. and Rücker, C. (2000) Walk count, labyrinthicity and complexity of acyclic and cyclic graphs and molecules. *J. Chem. Inf. Comput. Sci.* **40**, 99–106.
25. Gutman, I., Rücker, C., and Rücker G. (2001) On walks in molecular graphs. *J. Chem. Inf. Comput. Sci.* **41**, 739–745.
26. Rücker, G. and Rücker, C. (2001) Substructure, subgraph and walk counts as measures of the complexity of graphs and molecules. *J. Chem. Inf. Comput. Sci.* **41**, 1457–1462.
27. Bonchev, D. and Seitz, W. A. (1996) The concept of complexity in chemistry, in *Concepts in Chemistry: A Contemporary Challenge* (Rouvray, D. H., ed.). Research Studies Press, Taunton, UK, pp. 348–376.
28. Bonchev, D. and Trinajstić, N. (1977) Information theory, distance matrix, and molecular branching. *J. Chem. Phys.* **67**, 4517–4533.
29. Gavin, A.-C., Böschke, M., Krause, R., et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147.
30. Barabasi, A.-L. and Albert, R. (1999) Emergence of scaling in random networks. *Science* **286**, 509–512.
31. Jeong, H., Tombor, B., Albert R., Oltvai Z. N., and Barabasi, A.-L. (2000) The large-scale organization of metabolic networks. *Nature* **407**, 651–654.
32. Watts, D. J. and Strogatz, S. H. (1998) Collective dynamics of “small-world” networks. *Nature* **393**, 440–442.
33. Platt, J. R. (1947) Influence of neighbor bonds on additive bond properties in paraffins. *J. Chem. Phys.* **15**, 419–420.
34. Ho, Y., Gruhler, A., Heilbut, A., et al. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183.

Uncorrected Proof