BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 1(62), 2010, Pages 3–22
ISSN 1024–7696

Towards Structural Network Analysis

Matthias Dehmer, Marina Popovscaia

Abstract. Structural network analysis is an intricate problem. In fact, the majority of techniques that have been developed so far are only applicable to investigate deterministic network models. This gives rise to develop novel graph-theoretical methods for applying them to more complex graphs and especially to statistically inferred networks. In this regard, we review methods for analyzing complex networks structurally putting the special emphasis on network partitioning and quantifying network complexity. Both areas are of general importance in structural graph theory as well as useful for exploring biological networks.

Mathematics subject classification: 68R01, 68R10, 90B10, 92E10.
Keywords and phrases: Complex networks, graph classes, graph clustering, graph partitioning, biological networks, structural complexity, information measures, entropy, quantitative graph theory.

1 Introduction

Prominent areas in which graph-theoretical methods have been intensely used are, e.g., social network analysis [71, 110], biological network analysis [59], chemical graph theory [103] and investigating technological networks [82]. In terms of developing methods for exploring complex networks, random graph models have been frequently investigated [37, 41]. But besides merely exploring random graphs, it turned out that many real world phenomena can be modeled by using non-random network topologies and, hence, meaningful methods for their structural analysis are crucial [30]. From a mathematical point of view, either descriptive or quantitative methods could be used to explore graphs structurally. To name some well-known examples, we mention metrical properties of graphs [97], general graph measures [54], graph polynomials [50], graph decompositions [20], graph colorings [54], graph complexity [72] and the partitioning of graphs [22]. Importantly, we want to remark that most of the just mentioned approaches are only suitable to analyze deterministic graphs. But the observation that complex networks are often the result of a dynamical processes led to the insight that their analysis can not be adequately performed in a deterministic framework [40]. Thus, there is a strong need to design novel techniques to meet this challenge.

In this paper we provide a review about the structural analysis of complex networks. Here, we focus on such techniques which have been preferably used in computational and systems biology. Concretely, we will put the emphasis on approaches to partition complex networks and to quantify network complexity. Both problems

are challenging and there is a future need to find novel approaches when considering networks which were inferred statistically. Altogether, the main goal of this review paper is to demonstrate the usefulness and potential of Structural Graph Analysis and to stimulate the interest of other researchers to observe graph theory as a tool for solving interdisciplinary problems.

2 Graph-Theoretical Applications in Bioinformatics and Computational Biology

Various examples in the scientific literature have been demonstrated that biological phenomena and processes can be tackled by applying graph theory, see, e.g., [38, 80, 84].

In this section, we provide a general overview about important areas dealing with graph-based approaches in computational biology:

- Phylogenetics: During the last thirty years, various graph-based techniques have been successfully applied for solving problems in phylogenetics, see, e.g., [42, 89, 99]. A prominent example, for instance, is phylogenetic tree reconstruction that has been a major research goal for biologists because it often serves as indispensable interpretive framework for the analysis of evolutionary processes by representing the interrelationships among biological entities as graphs [42, 59]. Further, distance-, character-, and likelihood-based methods are three important approaches which have been used for phylogenetic tree analysis [42, 99, 100]. Besides the problem of inferring phylogenetic trees from biological data sets, the structural analysis of such graphs has been found as crucial. In this context, various tree distance measures and metrics [89, 92, 93] were used to determine the structural similarity between phylogenetic trees.

- RNA-Structure Analysis: Graphs play an important role when analyzing secondary structures inferred from biological sequences [109, 111]. For example, Nussinov [79] did the first attempt to calculate secondary structures for simplified energy models based on base-pairing rules. After the model was elaborated, it turned out that there is a further need for considering loops in the RNA secondary structure and, consequently, Zuker and Stiegler [116] proposed a recursive algorithm to take the loop types [116] into account. Moreover, an important contribution when analyzing secondary structures comparatively was proposed by McCaskill [70]. In order to compare secondary structures structurally, it turned out to be useful encoding them as trees and to use existing tree distance measures [92, 101] for determining their similarity. Note that a recent survey on graph-based techniques to model and process A-structures has been contributed by Washietl et al. [109].

- Molecular Biology: For instance, regulatory, signal transduction, or metabolic networks are often represented by networks to analyze molecular biological processes [57]. For this, special graph classes like bipartite graphs, hypergraphs

and directed acyclic graphs [58, 59, 67] were particularly used. Apart from applying existing graph classes [21] to represent networks, graph-theoretical techniques have been intensely used to analyze molecular biological pathways. Exemplarily, we here mention a contribution due to Rosselló et al.[90] who describe development pathways by using graph grammars.

- High-Throughput Analysis: A hype dealing with employing graphs in computational biology started after the development of high-throughput techniques [24, 38] because they allow a large-scale identification of genes, RNAs, proteins. In this context, a key problem is to identify and study functional components of a biological system meaningfully, based on their molecular interactions involving, e.g., genes, proteins or metabolites, instead of exploring these components in isolation. For example, a challenging problem in the above mentioned area is to investigate complex diseases by investigating underlying network representations [38]. To tackle these problems, methods from statistical data analysis and machine learning have been used [26, 38, 39].

- Drug Design and Bio-chemical Graph Analysis: A still challenging and ongoing problem is to predict physico-chemical or toxic properties of bio-chemical molecules using structural graph descriptors [33, 102]. Particularly entropy-based measures to perform such studies within QSPR (quantitative structure-property relationship) and QSAR (quantitative structure-activity relationship) have been found to be powerful, see, e.g., [8, 14, 33]. But because a large number of measures to quantify molecular complexity have been developed so far, there is a strong need to examine which kind of structural information the measures do detect. Contributions to shed light on this problem were recently made in [7, 32]. Similarly, Pathway Analysis [88] using graph-based techniques became a crucial field when analyzing bio-chemical processes and complex diseases [38]. In particular, it allows, e.g., the identification of gene networks and to study how genes are regulated [31].

3 Applied Graph Partitioning

The investigation of general graph partitioning methods for finding community structures is currently of considerable interest when analyzing complex networks quantitatively as well as descriptively [49, 78, 115]. In this section, we briefly describe such methods by using graph partitioning. Before we start outlining concrete techniques, we sketch some seminal work concerning classical graph partitioning [61].

3.1 Classical Methods

To understand the underlying idea of graph partitioning properly, we firstly state the following definition that describes the problem intuitively, see, e.g., [60].

Definition 1. *Let $G = (V, E)$ be a graph. Then, we define the k-way graph partitioning problem as follows: Partition the vertex set V into k subsets $V_1, V_2, \ldots, V_k$*

such that $V_i \bigcap V_j = \emptyset$, $i \neq j$, $|V_i| = \frac{n}{k}$, $\bigcup_i V_i = V$, *and the number of edges of* E *whose incident vertices belong to different subsets is minimized.*

An important contribution in this area is the algorithm due to Kernighan and Lin [61]. Apart from this work, other approaches to partition graphs based on spectral clustering and multilevel partitioning have been explored, see, e.g., [76]. For instance, spectral methods produce a partition based on the eigendecomposition [51] of the graph. Also, spectral approximations for a variety of partitioning criteria have been formulated including the minimum cut [83], ratio cut [25] and normalized cut [95]. Interestingly, most of the multilevel algorithms are based on the work we already mentioned above, see [61]. A strong point of this heuristic algorithm is the fact that its time complexity is $O(|V|^3)$ on sparse graphs [61]. Improvements possessing lower computational complexity can be found in [106]. Let's now describe the original method presented in [61] in more detail: Let $G = (V, E)$ be a graph with weighted edges (costs) and $|V| = 2n$. Let S be a set of $2n$ points with an associated cost matrix, $C = (c_{ij}), i, j = 1, \ldots, 2n$. Further, without loss of generality, it is assumed [61] that C is a symmetric matrix and $c_{ii} = 0, \forall\, i$. Then, the aim of the algorithm is to partition S into two sets A and B, $|A| = |B| = n$, such that the so-called external costs $T = \sum_{A \times B} c_{ab}$ will be minimized.

Note that the work initiated by Kernighan and Lin [61] has already been successfully improved, see, e.g., [43, 60, 61]. A well-known example of such a recently developed multilevel approach is the METIS algorithm [60] that aims to partition graphs from different application domains efficiently. In addition, we mention another fast multi-level algorithm developed by Dhillon et al. [34] that directly optimizes various weighted graph clustering objectives. In particular, Dhillon et al. [34] show that a general weighted k-means objective is mathematically equivalent to a weighted graph clustering objective. The main advantage of this method is that it approximates graph clustering objectives without requiring an eigendecomposition, which can be computationally intensive for large graphs [34]. Another advantage of this algorithm, compared to other multilevel approaches, is that it does not require the partitions to be of equal sizes [34].

3.2 Community Structure Detection

In this section, we sketch known approaches to detect community structures within biological networks. Generally, to find community structures in complex networks, classical and recent graph partitioning methods have often been applied [48, 59]. Until now, the concept of graph partitioning has been used for detecting community structures in social networks, WWW-graphs, and biological or biochemical networks [44, 48, 56, 59, 77]. Informally speaking, the community structure property of a network can be understood by considering a graph in which the vertices are joined together in tightly-knit groups and there are only looser connections between them, see [48]. It is important to mention that the traditional method for detecting community structures in networks is hierarchical clustering [48]. If we start with a weighted graph $G = (V, E)$, $|V| = n$, (i) we first have to calculate a

weight w_{ij} for every pair i, j of vertices in the network, (ii) select all vertices in the network with no edges between them and (iii) add edges between pairs one by one in the order of their weights, starting with the pair with the strongest weight and progressing to the weakest, see [48]. If edges are added, the resulting graph shows a nested set of connected subsets of vertices, which are expected to be the communities [48]. Note that algorithms of this kind are called agglomerative, see, e.g., [12].

To overcome existing shortcomings of agglomerative methods, see, e.g., [35], Girvan and Newman [48] proposed an alternative approach for detecting communities that represents a so-called divisive algorithm [48]. The main procedure works as follows: Start with the entire graph and iteratively cut the edges, thus dividing the network progressively into smaller and smaller disconnected sub-networks finally identified as the communities. The crucial point of this algorithm is the selection of the edges to be cut, which has to be those connecting communities and not those within them. The main steps of algorithm proposed in [48] can be stated as follows:

1. Calculate the betweenness for all edges in the network.
2. Remove the edge with the highest betweenness.
3. Recalculate betweenness for all edges affected by the removal.
4. Repeat from the step 2 until no edges remain.

Until now, several improvements and extensions using shortest path versions of this algorithm have already been proposed [48]. For example, Holme et al. [56] modified this method and then applied this modification, based on global centrality measure (betweenness), to a number of metabolic networks from different organisms for finding communities that correspond to functional units within these networks. Also, Wilkinson and Huberman [113] have applied the approach to a network representing relationships between genes, as established by the co-occurrence of gene names found in published research articles. For finding communities in network they used a nonlocal process exploiting the concept of betweenness centrality.

For finalizing, we state two more contributions in this area. The CONGA (Cluster-Overlap Newman Girvan Algorithm) [52] is an extension of [48]. It can be also used with undirected, unweighted graphs and performs hierarchical clustering but it allows overlapping clusters. Finally, the CFinder algorithm [81], that is a bottom-up approach, provides a method to interpret communities as union of cliques. For more details refer to [48, 81].

4 Topological Complexity Measures for Graphs

The problem of determining the structural complexity of a network can be understood as characterizing the graphs taking structural features into account [11,17,62]. Clearly, this task is not uniquely defined because no complexity index can measure

all structural features which contribute to the complexity of a graph. Before starting with describing concrete non-information-theoretic and information-theoretic complexity measures, we outline existing applications in computational biology and bioinformatics:

- To investigate the evolution of PPI domains and the impact on organismal complexity and the complexity of protein-protein interaction networks [114].
- General studies to examine how, e.g., biological and technological networks differ by calculating their structural complexity [108].
- To find interrelations between the structure and complexity of the pathways and the phylogeny of species by using non-information-theoretic and information-theoretic complexity measures [17, 69].
- To use entropy-based measures for problems in QSPR (quantitative structure-property relationship) and QSAR (quantitative structure-activity relationship), see, e.g., [8, 14, 33].
- To employ non-information-theoretic and information-theoretic measures in the field of chemoinformatics [47], e.g., to perform correlation analyses [7] and develop similarity/diversity measures [107].

4.1 Distance-Based Measures

A large number of complexity measures that have been developed so far are based on distances in a graph [102]. As a strong point, such distances are simple to calculate by any shortest path algorithm to be applied to the underlying adjacency matrix. Often, a weak point of such measures is that they do not capture structural information uniquely, that means, the measures are highly degenerated [64]. Let $G = (V, E)$ be a graph. We now start by expressing the well-known Wiener-index [112],

$$W(G) := \frac{1}{2} \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} d(v_i, v_j). \tag{1}$$

Originally, it was developed to detect branching of chemical graphs [53]. $d(v_i, v_j)$ denotes the shortest distance between v_i and v_j. Similarly, the Harary-index [4,36],

$$H(G) := \frac{1}{2} \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} (d(v_i, v_j))^{-1}, \quad i \neq j, \tag{2}$$

is based on reciprocal distances. A more complex example of such a measure is the Balaban J-index [2],

$$J(G) := \frac{|E|}{\mu + 1} \sum_{(v_i, v_j) \in E} [DS_i DS_j]^{-\frac{1}{2}}. \tag{3}$$

Note that DS_i denotes the distance sum (row sum) of $v_i \in V$ and $\mu := |E|+1-|V|$ is the cyclomatic number. Other important distance-based measures are, for instance, the mean distance deviation [97],

$$\Delta\mu(G) := \frac{1}{|V|} \sum_{i=1}^{|V|} |\mu(v_i) - \bar{\mu}|, \tag{4}$$

where

$$\mu(v_i) := \sum_{j=1}^{|V|} d(v_i, v_j) \quad \text{and} \quad \bar{\mu} := \frac{2W}{|V|}, \tag{5}$$

the product of row sums-index [91] given by

$$\log\big(\mathrm{PRS}(G)\big) := \log\left(\prod_{i=1}^{|V|} \mu(v_i)\right), \tag{6}$$

and, finally, the hyper-distance-path index [102],

$$D_P(G) := \frac{1}{2} \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} d(v_i, v_j) + \frac{1}{2} \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} \binom{d(v_i, v_j)}{2}. \tag{7}$$

For further distance-based measures, refer to [102].

4.2 Other and Related Complexity Indices

Besides distance-based measures, various other complexity indices for networks based on other graph invariants have been developed. To pursue outlining known graph complexity measures, we now state some important examples which have been used to measure molecular complexity [86]. Note that the purpose for deriving such indices was either to find measures with low computational complexity or with high discrimination power [64]. For example, the index of total adjacency [17] can be easily derived from the underlying adjacency matrix,

$$A(G) := \frac{1}{2} \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} a_{ij}. \tag{8}$$

a_{ij} denotes the entry in the i-th row and j-th column of A. From this, it straightforwardly follows the normalized edge complexity [17],

$$E_{|V|}(G) := \frac{A}{|V|^2}. \tag{9}$$

Let k_{v_i} be the degree of the vertex $v_i \in V$. Interestingly, the vertex degrees seen as a graph invariant have also been used to define measures to quantify the structural complexity of graphs. In order to list some well-known examples, we now state the Zagreb group-indices [36],

$$Z_1(G) := \sum_{i=1}^{|V|} k_{v_i} \quad \text{or} \quad Z_2(G) := \sum_{(v_i,v_j)\in E} k_{v_i} k_{v_j}, \tag{10}$$

and the Randić connectivity-index [85],

$$R(G) := \sum_{(v_i,v_j)\in E} [k_{v_i} k_{v_j}]^{-\frac{1}{2}}. \tag{11}$$

An interesting generalization of such measures was developed by Bonchev [14] who developed the so-called Overall (OX) indices given by

$$OX(G) := \sum_{k=0}^{|E|} {}^{k}X; \quad \{OX(G)\} = \{{}^{0}X, {}^{1}X, \ldots, {}^{|E|}X\}. \tag{12}$$

OX is called the overall value of a certain graph invariant X by summing up its values in all subgraphs, and partitioning them into terms of increasing orders (increasing number of subgraph edges k). For instance, $OX = SC$ is equal to the subgraph count [14, 17].

More recent complexity measures were developed by Kim et al. [62]. To name an example, we here express the so-called Efficiency complexity C^e of a graph G that is based on calculating path lengths. Starting from

$$E'(G) := \frac{2}{|V|(|V|-1)} \sum_i \sum_{j>i} \frac{1}{d(v_i, v_j)}, \tag{13}$$

expressing the arithmetic mean of all inverse path lengths. Further, by defining

$$E_{\text{path}}(G) := \frac{2}{|V|(|V|-1)} \sum_{i=1}^{|V|-1} \frac{(|V|-i)}{i}, \tag{14}$$

the Efficiency complexity C^e yields to

$$Ce(G) := \left(\frac{E' - E_{\text{path}}}{1 - E_{\text{path}}}\right)\left(1 - \frac{E' - E_{\text{path}}}{1 - E_{\text{path}}}\right) \in [0, 1]. \tag{15}$$

Moreover, a measure that crucially relies on the largest eigenvalue of an undirected graph G was defined in [62]. If r stands for the largest eigenvalue calculated from the adjacency matrix of G, then, the graph index Cr is defined as

$$Cr := 4c_r(1 - c_r) \in [0, 1], \tag{16}$$

where

$$c_r := \frac{r - 2\cos\left(\frac{\pi}{N+1}\right)}{|V| - 1 - 2\cos\left(\frac{\pi}{|V|+1}\right)}. \tag{17}$$

Numerical examples to calculate these measures and details regarding their interpretation can be found in [62].

4.3 Information-theoretic Complexity Measures

The key concept for obtaining a further class of important graph complexity measures relies on Shannons's information theory [94]. Starting from inferred structural features of a network, the crucial step for quantifying its structural information is to infer a probability distribution and, then, to apply Shannons's entropy formula. As a result, one obtains topological entropies for characterizing networks [13,32,98]. Prior to start explaining concrete information measures, we emphasize that the main application domains of general information-theoretic methods to analyze networks have been biology [27,68,73,87], ecology [55,105], mathematical chemistry [6,8,13], software technology [1], and operations research [23,45].

4.3.1 Classical Information Measures for Graphs

The development of information measures represented by entropies to characterize the underlying topology of a given network was the starting point of applying information theory to investigate biological and chemical systems structurally [87,104]. These measures are based on the principle that by assuming a graph $G = (V, E)$, a graph invariant X and an equivalence criterion, distributions of X can be obtained. Particularly, this process can be understood by considering the following scheme [13]:

$$\begin{pmatrix} 1 & 2 & \cdots & k \\ |X_1| & |X_2| & \cdots & |X_k| \\ p_1 & p_2 & \cdots & p_k \end{pmatrix}. \tag{18}$$

The first row of this matrix contains the equivalency classes and the second row the cardinalities of the obtained partitions, respectively. The probability values, calculated by $p_i = \frac{|X_i|}{|X|}$, for each partition form the third row. Hence, $\mathcal{P}_G = (p_1, \ldots, p_k)$ represents a probability distribution. Then, the application of the well known Shannon-entropy [94]

$$H(X) := H\left(p(x_1), \ldots, p(x_k)\right) = -\sum_{i}^{k} p(x_i)\log(p(x_i)), \tag{19}$$

of a discrete random variable $X = (x_1, x_2, \ldots, x_k)$ leads to the following graph entropies [13],

$$I_t(G) := |X|\log(|X|) - \sum_{i=1}^{k} |X_i|\log(|X_i|), \tag{20}$$

$$I_m(G) := -\sum_{i=1}^{k} p_i \log(p_i) = -\sum_{i=1}^{k} \frac{|X_i|}{|X|} \log\left(\frac{|X_i|}{|X|}\right). \tag{21}$$

$I_t(G)$ is called the total structural information content of G and can be recalculated by using $I_m(G)$. The latter is called the mean structural information content of a graph.

Now, by using certain graph invariants X, special entropy measures can be obtained which serve as graph complexity measures. The starting point by developing concrete measures was done by Rashevsky [87] and Trucco [104]. Rashevsky's information measures to characterize G are concretely given by [87]

$$I_t^V(G) := |V| \log(|V|) - \sum_{i=1}^{k} N_i \log\left(N_i\right), \tag{22}$$

$$I_m^V(G) := -\sum_{i=1}^{k} \frac{|N_i|}{|V|} \log\left(\frac{|N_i|}{|V|}\right), \tag{23}$$

where $|N_i|$ denotes the number of topologically equivalent vertices in the i-th vertex orbit of G. k stands for the number of different vertex orbits. Trucco's measure [104] can be analogously obtained by using the edge automorphism group. After this seminal work, Mowshowitz [74] also investigated the measure I_m^V (see Equation (23)) in depth and additionally explored the chromatic information content of a graph [75]:

$$I_c(G) := \min_{\hat{V}} \left\{ -\sum_{i=1}^{h} \frac{n_i(\hat{V})}{|V|} \log\left(\frac{n_i(\hat{V})}{|V|}\right) \right\}. \tag{24}$$

$\hat{V} = \{V_i | 1 \leq i \leq h\}$ is an arbitrary chromatic decomposition of G, $|V_i| = n_i(\hat{V})$, and $h = \chi(G)$ is the chromatic number of G. Note that the computation of the chromatic number is a costly procedure for arbitrary graphs [54].

Apart from defining and calculating information measures for networks, there is also a strong need to understand the meaning of these measures in depth. This could be done by establishing their mathematical properties under certain theoretical assumptions (e.g., bounds and the behavior under certain graph operations etc.). Such a concrete result has been proven by Mowshowitz [74].

Theorem 1. *For graphs G and H*

$$I_m^V(G \times H) \leq I_m^V(G) + I_m^V(H), \tag{25}$$

and

$$I_m^V(G \circ H) \leq I_m^V(G) + I_m^V(H), \tag{26}$$

where $\times$ and $\circ$ represent the cartesian product and composition, respectively.

The assertion of this theorem is that the information measure is semi-additive on the cartesian product and on the composition of two graphs. Interestingly, we emphasize that formal properties like the just shown one or bounds for the entropies (by using important graph classes) are unknown for the majority of network information measures.

After the just outlined work, Bonchev [13,18] introduced the so-called magnitude-based information indices by defining a weighted probability scheme. These indices can be considered as generalization of the measures due to Rashevsky and Mowshowitz. It follows easily that such a scheme can be analogously applied to a system with $|N|$ elements to group these elements into k partitions according to the magnitude. Then, the modified scheme is [13]:

$$\begin{pmatrix} 1 & 2 & \cdots & k \\ |N_1| & |N_2| & \cdots & |N_k| \\ p_1 & p_2 & \cdots & p_k \\ w_1 & w_2 & \cdots & w_k \\ p_1^M & p_2^M & \cdots & p_k^M \end{pmatrix}. \tag{27}$$

In addition to the already existing rows of the introduced probability scheme (see Matrix (18)), the magnitudes representing weights $(w_1, w_2, \ldots, w_k)$ and the weighted probability values $(p_1^M, p_2^M, \ldots, p_k^M)$ were introduced [13,18]. Because it holds $M = \sum_{i=1}^{k} w_i N_i$ and $p_i = \frac{w_i}{M}$, the graph entropies represented by Equation (20) and Equation (21) can be rewritten as

$$I_t^M(G) := M \log(M) - \sum_{i=1}^{k} N_i w_i \log(w_i), \tag{28}$$

$$I_m^M(G) := -\sum_{i=1}^{k} \frac{N_i w_i}{M} \log\left(\frac{w_i}{M}\right). \tag{29}$$

By using this approach, concrete magnitude-based information were defined, for instance [13],

$$I_D(G) := -\frac{1}{N} \log\left(\frac{1}{N}\right) - \sum_{i=1}^{\rho(G)} \frac{2k_i}{N^2} \log\left(\frac{2k_i}{N^2}\right), \tag{30}$$

$$I_D^W(G) := W(G) \log(W(G)) - \sum_{i=1}^{\rho(G)} i k_i \log(i). \tag{31}$$

k_i is the occurrence of a distance possessing value i in the distance matrix of G. A strong point of these measures is their low degeneracy [63] compared to the classical measures mentioned in the beginning of this section. In general, one calls such a measure degenerated if for more than one graph the measure possesses the same value [3]. By using chemical graphs, numerical results are reported in [13, 65].

Again Bonchev [16,17] developed a substructure-based approach to detect molecular complexity. Let $G = (V, E)$ be a graph and X be a graph invariant. Then, the following entropy-based complexity measure

$$I(G, OX) := OX \log(OX) - \sum_{k=0}^{|E|} {}^{k}X \log\left({}^{k}X\right), \tag{32}$$

relies on the overall value OX (see Equation (12)) by summing up its values in all subgraphs [16]. The values will be partitioned into terms of increasing orders (increasing number of subgraph edges k) [16]. As an example, one can set $OX = SC$, i.e., OX equals the subgraph count [16]. Starting from this construction, Bonchev [16, 17] obtained several overall information indices such as overall connectivity (the sum of total adjacency of all subgraphs) [14], overall Wiener-index (the sum of total distances of all subgraphs) [15], overall Zagreb-indices [19], and the overall Hosoya-index [16]. Known earlier and also substructure-based contributions to detect molecular complexity were developed by, e.g., Bertz et al. [9, 10]. As a further remark, note that many further information measures for graphs which are similar to the outlined ones or which are based on the same construction principle (e.g., simple finite probability scheme, weighted probability scheme, etc.) can be found in [8, 13, 17, 33, 102].

To finalize this section as well as to show a different paradigm to derive graph entropies, we state the well-known Körner entropy [66, 96] that has been applied in information theory. The measure is defined by

$$H(G, P) := \lim_{t \longrightarrow \infty} \min_{U \subseteq V^t, P^t(U) > 1-\epsilon} \frac{1}{t} \log(\chi(G^t(U))). \tag{33}$$

For $V' \subseteq V(G)$, the induced subgraph on V' is denoted by $G(V')$ and $\chi(G)$ is the chromatic number [5] of G, G^t the t-th co-normal power [66] of G and

$$P^t(U) := \sum_{x \in U} P^t(x). \tag{34}$$

Examples and an interpretation of this measure can be found in [66, 96].

4.3.2 Parametric Information Measures

To compute the structural information content of arbitrary large networks, one needs a method whose underlying algorithm is efficient, i.e., its time complexity is polynomial. From Section (4.3.1), it follows that classical network information measures are often rely on algebraic principles, e.g., determining automorphism groups of graphs or chromatic decompositions. However it is known that for arbitrary networks, the computational complexity of the corresponding algorithms is often very high [46].

In order to overcome this problem, we now present parametric entropy measures whose time complexity has been proven to be polynomial [28]. The key principle to

construct such information measures is as follows: Let $G = (V, E)$ be an arbitrary graph and let S be a given set, e.g., a set of vertices or paths etc. The function $f : S \longrightarrow \mathbb{R}_+$ is called an abstract information functional of G. Instead of inducing partitions using an equivalence criterion (see Section (4.1)), we start from an abstract information functional f and define the quantity [29],

$$p^f(v_i) := \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)}, \quad \forall\, v_i \in V. \tag{35}$$

Because the following equation

$$p^f(v_1) + p^f(v_2) + \ldots + p^f(v_{|V|}) = 1, \tag{36}$$

holds by definition, these entities can be interpreted as vertex probabilities. Hence, $(p^f(v_1), \ldots, p^f(v_{|V|}))$ forms a probability distribution. From this, it is straightforward to obtain families of graph entropy measures like

$$I_f(G) := -\sum_{i=1}^{|V|} \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)} \log\left(\frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)}\right), \tag{37}$$

or

$$I_f^{\lambda}(G) := \lambda\left(\log(|V|) + \sum_{i=1}^{|V|} \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)} \log\left(\frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)}\right)\right), \tag{38}$$

$\lambda > 0$. By incorporating special information functionals, one clearly obtains special entropies. To give an example for a special information functional that is based on metrical properties, we express [29]

$$f^{V_1}(v_i) := \alpha^{c_1|S_1(v_i,G)|+c_2|S_2(v_i,G)|+\cdots+c_{\rho(G)}|S_{\rho(G)}(v_i,G)|}, \quad c_k > 0, 1 \leq k \leq \rho(G), \alpha > 0. \tag{39}$$

c_k are arbitrary real positive coefficients. $\rho(G)$ denotes the diameter of G and

$$S_j(v_i, G) := \{v \in V |\, d(v_i, v) = j,\, j \geq 1\}, \tag{40}$$

the j-sphere of a vertex v_i of G, respectively. f^{V_1} is a parametric information functional that depends on both the parameter α and the vector $(c_1, c_2, \ldots, c_{\rho(G)})$. The meaning of these parameters has been explained in [29]. Then, the resulting (parametric) information measure representing the entropy of the underlying graph topology is

$$I_{f^{V_i}}(G) := -\sum_{i=1}^{|V|} \frac{f^{V_i}(v_i)}{\sum_{j=1}^{|V|} f^{V_i}(v_j)} \log\left(\frac{f^{V_i}(v_i)}{\sum_{j=1}^{|V|} f^{V_i}(v_j)}\right), \quad i = 1, 2. \tag{41}$$

Of course, it is also possible to define

$$f^{V_2}(v_i) := c_1|S_1(v_i, G)| + c_2|S_2(v_i, G)| + \cdots + c_{\rho(G)}|S_{\rho(G)}(v_i, G)|, \tag{42}$$

that does not depend on α. Importantly, the process to design an information functional and, thus, the resulting information measures strongly depends on the specific problem when characterizing a graph using an information measure.

5 Summary and Conclusion

In this paper, we reviewed some concepts known in structural graph analysis. We emphasized that we particularly put the underscore on such methods which have been used in bioinformatics and systems biology. After outlining graph-theoretical approaches in these areas, we firstly began to survey graph partitioning methods to find clusters or communities within complex networks. Due to the steadily increasing complexity of real-world networks, we believe that it will be fruitful to further develop this field to process statistically inferred networks.

As future work, we want to focus on approaches combining graph-theoretical and information-theoretic techniques. Secondly, we studied the challenging problem to determine the structural complexity of graphs and reviewed classical and recent methods. We want to emphasize that finding a meaningful complexity measure to quantify structural information of a graph is far from trivial and usually not unique. These facts give an idea about the complexity of such measures. Also, in consideration of the fact that a vast number of graph complexity measures have been developed so far, the problem to examine which kind of structural information the measures do detect is not solved properly. Therefore, we would like to shed light on this important aspect in the future by examining correlations and interrelations between graph complexity measures.

6 Acknowledgments

We are grateful to Danail Bonchev, Frank Emmert-Streib, Armin Graber, and Alexander Mehler for fruitful discussions.

This work was supported by the COMET Center ONCOTYROL and funded by the Federal Ministry for Transport Innovation and Technology (BMVIT) and the Federal Ministry of Economics and Labour/the Federal Ministry of Economy, Family and Youth (BMWA/BMWFJ), the Tiroler Zukunftsstiftung (TZS) and the State of Styria represented by the Styrian Business Promotion Agency (SFG) [and supported by the University for Health Sciences, Medical Informatics and Technology and BIOCRATES Life Sciences AG].

References

[1] Allen E.B. *Measuring graph abstractions of software: An information-theory approach.* Proceedings of the 8-th International Symposium on Software Metrics, page 182. IEEE Computer Society, 2002.

[2] Balaban A.T. *Highly discriminating distance-based topological index.* Chem. Phys. Lett., 1982, **89**, 399–404.

[3] Balaban A.T., Balaban T.S. *New vertex invariants and topological indices of chemical graphs based on information on distances.* J. Math. Chem., 1991, **8**, 383–397.

[4] Balaban A.T., Ivanciuc O. *Historical development of topological indices.* In: Balaban A.T., Devillers J., editors, *Topological Indices and Related Descriptors in QSAR and QSPAR*, pages 21–57. Gordon and Breach Science Publishers, Amsterdam, The Netherlands, 1999.

[5] Bang-Jensen J., Gutin G. *Digraphs. Theory, Algorithms and Applications.* Springer, London, Berlin, Heidelberg, 2002.

[6] Basak S.C. *Information-theoretic indices of neighborhood complexity and their applications.* In: Balaban A.T., Devillers J., editors, *Topological Indices and Related Descriptors in QSAR and QSPAR*, pages 563–595. Gordon and Breach Science Publishers, Amsterdam, The Netherlands, 1999.

[7] Basak S.C., Balaban A.T., Grunwald G.D., Gute B.D. *Topological indices: Their nature and mutual relatedness.* J. Chem. Inf. Comput. Sci., 2000, **40**, 891–898.

[8] Basak S.C., Magnuson V.R. *Molecular topology and narcosis.* Arzeim.-Forsch./Drug Design, 1983, **33(I)**, 501–503.

[9] Bertz S.H., Herndon W.C. *The similarity of graphs and molecules.* In: Artificial Intelligence Applications in Chemistry, Chapter 15, 169–175. American Chemical Society, 1986.

[10] Bertz S.H., Wright W.F. *The graph theory approach to synthetic analysis: Definition and application of molecular complexity and synthetic complexity.* Graph Theory Notes of New York, 1998, **35**, 32–48.

[11] Boccaletti S., Latora V., Moreno Y., Chavez M., Hwang D.-U. *Complex networks: Structure and dynamics.* Physics Reports, 2006, **424(4-5)**, 175–308.

[12] Bock H.H. *Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten.* Studia Mathematica. Vandenhoeck & Ruprecht, Göttingen, 1974.

[13] Bonchev D. *Information Theoretic Indices for Characterization of Chemical Structures.* Research Studies Press, Chichester, 1983.

[14] Bonchev D. *Overall connectivities and topological complexities: A new powerful tool for QSPR/QSAR.* J. Chem. Inf. Comput. Sci., 2000, **40(4)**, 934–941.

[15] Bonchev D. *The overall Wiener index – a new tool for characterization of molecular topology.* J. Chem. Inf. Comput. Sci., 2001, **41**, 582–592.

[16] Bonchev D. *The overall topological complexity indices.* In: Simos T., Maroulis G., editors, *Advances in Computational Methods in Science and Engineering*, Vol. 4B, pages 1554–1557. VSP Publications, 2005.

[17] Bonchev D., Rouvray D.H. *Complexity in Chemistry, Biology, and Ecology.* Mathematical and Computational Chemistry. Springer, New York, NY, USA, 2005.

[18] Bonchev D., Trinajstić N. *Information theory, distance matrix and molecular branching.* J. Chem. Phys., 1977, **67**, 4517–4533.

[19] Bonchev D., Trinajstić N. *Overall molecular descriptors. 3. Overall Zagreb indices.* SAR QSAR Environ. Res., 2001, **12**, 213–236.

[20] Bosák J. *Decompositions of Graphs.* Mathematics and its Applications. Springer, 1990.

[21] Brandstädt A., Le V.B., Sprinrad J.P. *Graph Classes. A Survey.* SIAM Monographs on Discrete Mathematics and Applications, 1999.

[22] Broere I., Hajnal P., Mihók P. *Partition problems and kernels of graphs.* Discussiones Mathematicae Graph Theory, 1997, **17**, 311–313.

[23] Calinescu A., Efstathiou J., Sivadasan S., Schirn J., Huaccho Huatuco L. *Complexity in manufacturing: an information theoretic approach.* Proceedings of the International Conference on Complex Systems and Complexity in Manufacturing, pages 30–44, 2000.

[24] Causton H.C., Brazma A.,Quackenbush J. *Microarray Gene Expression Data Analysis.* Blackwell Publishers, 2003.

[25] Chan P., Schlag M., Zien J. *Spectral k-way ratio cut partitioning.* IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 1994, **13**, 1088–1096.

[26] Cristianini N., Shawe-Taylor J. *An Introduction to Support Vector Machines.* Cambridge University Press, Cambridge, UK, 2000.

[27] Dancoff S.M., Quastler H. *Information content and error rate of living things.* In: Quastler H., editor, *Essays on the Use of Information Theory in Biology*, pages 263–274. University of Illinois Press, 1953.

[28] Dehmer M. *Information-theoretic concepts for the analysis of complex networks.* Appl. Artif. Intell., 2008, **22(7 & 8)**, 684–706.

[29] Dehmer M. *A novel method for measuring the structural information content of networks.* Cybernetics and Systems, 2008, **39**, 825–843.

[30] Dehmer M. *Structural Analysis of Complex Networks.* Birkhäuser/Springer, Boston/Basel, 2010.

[31] Dehmer M., Emmert-Streib F. *Analysis of Complex Networks: From Biology to Linguistics.* Wiley VCH, Weinheim, Germany, 2009.

[32] Dehmer M., Varmuza K., Borgert S., Emmert-Streib F. *On entropy-based molecular descriptors: Statistical analysis of real and synthetic chemical structures.* J. Chem. Inf. Model., 2009, **49**, 1655–1663.

[33] Devillers J., Balaban A.T. *Topological Indices and Related Descriptors in QSAR and QSPR.* Gordon and Breach Science Publishers, Amsterdam, The Netherlands, 1999.

[34] Dhillon I.S., Guan Y., Kulis B. *Weighted graph cuts without eigenvectors: a multilevel approach.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, **29**, 1944–1957.

[35] Diao Y., Lia M., Fenga Z., Yina J., Pana Y. *The community structure of human cellular signaling network.* Journal of Theoretical Biology, 2007, **247**, 608–615.

[36] Diudea M.V., Gutman I., Jäntschi L. *Molecular Topology.* Nova Publishing, Ney York, NY, USA, 2001.

[37] Dorogovtsev S.N., Mendes J.F.F. *Evolution of Networks. From Biological Networks to the Internet and WWW.* Oxford University Press, 2003.

[38] Emmert-Streib F., Dehmer M. *Analysis of Microarray Data: A Network-Based Approach.* Wiley-VCH, 2008.

[39] Emmert-Streib F., Dehmer M. *Information Theory and Statistical Learning.* Springer, New York, USA, 2008.

[40] Emmert-Streib F., Dehmer M. *Information processing in the transcriptional regulatory network of yeast: Functional robustness.* BMC Syst. Biol., 2009, **3**:35.

[41] Erdös P., Rényi P. *On the evolution of random graphs.* Magyar Tud. Akad. Mat. Kutató Int. Közl, 1960, **5**, 17–61.

[42] Felsenstein J. *Coalescents, phylogenies, and likelihoods.* Biological Bulletin, 1999, **196**, 343–344.

[43] Fiduccia C.M., Mattheyses R.M. *A linear-time heuristic for improving network partitions.* Proceedings of the 19-th Design Automation Conference, pages 241–247, 1982.

[44] Flake G.W., Lawrence S.R., Giles C.L., Coetzee F.M. *Self-organization and identification of web communities.* IEEE Computer, 2002, **35**, 66–71.

[45] Frizelle G., Woodcock E. *Measuring complexity as an aid to developing operational strategy.* International Journal of Operations & Production Management, 1995, **15**, 26–39.

[46] Garey M.R., Johnson D.S. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* Series of Books in the Mathematical Sciences. W. H. Freeman, 1979.

[47] Gasteiger J., Engel T. *Chemoinformatics - A Textbook.* Wiley VCH, Weinheim, Germany, 2003.

[48] Girvan M., Newman M.E.J. *Community structure in social and biological networks.* PNAS, 2002, **99**, 7821–7826.

[49] Gleiser P.M., Danon L. *Community structure in jazz.* Advances in complex systems, 2003, **6(4)**, 565–574.

[50] Godsil C., Royle G. *Algebraic Graph Theory.* Graduate Texts in Mathematics. Academic Press, 2001.

[51] Golub G.H., Van Loan C.F. *Matrix Computations.* John Hopkins University Press, Baltimore, USA, 1996.

[52] Gregory S. *An algorithm to find overlapping community structure in networks.* Proceedings of 11th European conference on principles and practice of knowledge discovery in databases (PKDD'07), pages 91–102, 2007.

[53] Gutman I., Randić M. *Algebraic characterization of skeletal branching.* Chemical Physics Letters, 1977, **47(1)**, 15–17.

[54] Harary F. *Graph Theory.* Addison Wesley Publishing Company, Reading, MA, USA, 1969.

[55] Hirata H., Ulanowicz R.E. *Information theoretical analysis of ecological networks.* Int. J. Syst. Sci., 1984, **15**, 261–270.

[56] Holme P., Huss M., Jeong H. *Subnetwork hierarchies of biochemical pathways.* Bioinformatics, 2003, **19**, 532–538.

[57] Huber W., Carey V., Long L., Falcon S., Gentleman R. *Graphs in molecular biology.* BMC Bioinformatics, 2007, **8(Suppl 6):S8**.

[58] Jungnickel D. *Graphs, networks and algorithms.* Springer/Verlag, 2005.

[59] Junker B.H., Schreiber F. *Analysis of biological networks.* Wiley-Interscience, 2008.

[60] Karypis G., Kumar V. *A fast and high quality multilevel scheme for partitioning irregular graphs.* SIAM Journal on Scientific Computing, 1999, **20**, 359–392.

[61] Kernighan B.W., Lin S. *An efficient heuristic procedure for partitioning graphs.* Bell Systems Technical Journal, 1970, **49**, 291–307.

[62] Kim J., Wilhelm T. *What is a complex graph?* Physica A, 2008, **387**, 2637–2652.

[63] Konstantinova E.V. *On some applications of information indices in chemical graph theory.* In: Ahlswede R., Bäumer L., Cai N., Aydinian H., Blinovsky V., Deppe C., Mashurian H., editors, *General Theory of Information Transfer and Combinatorics*, Lecture Notes of Computer Science, pages 831–852. Springer, 2006.

[64] Konstantinova E.V., Paleev A.A. *Sensitivity of topological indices of polycyclic graphs.* Vychisl. Sistemy, 1990, **136**, 38–48 (in Russian).

[65] Konstantinova E.V., Skorobogatov V.A., Vidyuk M.V. *Applications of information theory in chemical graph theory.* Indian Journal of Chemistry, 2002, **42**, 1227–1240.

[66] Körner J. *Coding of an information source having ambiguous alphabet and the entropy of graphs.* Transactions of the 6-th Prague Conference on Information Theory, pages 411–425, 1973.

[67] Lauritzen S.L. *Graphical Models.* Oxford Statistical Science Series. Oxford University Press, 1996.

[68] Linshitz H. *The information content of a battery cell.* In: Quastler H., editor, *Essays on the Use of Information Theory in Biology*, University of Illinois Press, Urbana, IL, USA, 1953.

[69] Mazurie A., Bonchev D., Schwikowski B., Buck G.A. *Phylogenetic distances are encoded in networks of interacting pathways.* Bioinformatics, 2008, **24(22)**, 2579–2585.

[70] McCaskill J.S. *The equilibrium partition function and base pair binding probabilities for RNA secondary structure.* Biopolymers, 1990, **29**, 1105–1119.

[71] Mehler A. *A Quantitative Graph Model of Social Ontologies by Example of Wikipedia.* In: Dehmer H., Emmert-Streib F., editors, *Towards an Information Theory of Complex Networks: Statistical Methods and Applications*, Birkhäuser, Boston/Basel, 2010, to appear.

[72] Minoli D. *Combinatorial graph complexity.* Atti. Accad. Naz. Lincei, VIII. Ser., Rend., Cl. Sci. Fis. Mat. Nat., 1975, **59**, 651–661.

[73] Morowitz H. *Some order-disorder considerations in living systems.* Bull. Math. Biophys., 1953, **17**, 81–86.

[74] Mowshowitz A. *Entropy and the complexity of the graphs I: An index of the relative complexity of a graph.* Bull. Math. Biophys., 1968, **30**, 175–204.

[75] Mowshowitz A. *Entropy and the complexity of graphs IV: Entropy measures and graphical structure.* Bull. Math. Biophys., 1968, **30**, 533–546.

[76] Narasimhamurthy A., Greene D., Hurley N., Cunningham P. *Partitioning large networks without breaking communities.* Springer-Verlag, London, 2009.

[77] Newman M.E.J. *Detecting community structure in networks.* The European Physical Journal B, 2004, **38**, 321–330.

[78] Newman M.E.J., Girvan M. *Finding and evaluating community structure in networks.* Physical Review E, 2003, **69**, 026113(15).

[79] Nussinov R., Jacobson A.B. *Fast algorithm for predicting the secondary structure of single-stranded RNA.* PNAS, 1980, **77(11)**, 6309–6313.

[80] Ogata H., Fujibuchi W., Goto S., Kanehisa M. *A heuristic graph comparison algorithm and its application to detect functionally.* Nucleic Acids Res., 2000, **28(20)**, 4021–4028.

[81] Palla G., Derenyi I., Farkas I., Vicsek T. *Uncovering the overlapping community structure of complex networks in nature and society.* Nature, 2005, **435**, 814–818.

[82] Patil A., Esfahanian A.-H., Xiao L., Liu Y. *Resource allocation using multiple edge-sharing multicast trees.* IEEE Transactions on Vehicular Technology, 2007, **58**, 3178–3186.

[83] Pothen A., Simon H.D., Liou K.P. *Partitioning sparse matrices with eigenvectors of graphs.* SIAM Journal on Mathematical Analysis, 1990, **11**, 430–452.

[84] Pržulj N. *Network comparison using graphlet degree distribution.* Bioinformatics, 2007, **23**, e177–e183.

[85] Randić M. *On characterization of molecular branching.* J. Amer. Chem. Soc., 1975, **97**, 6609–6615.

[86] Randić M., Plavšić D. *Characterization of molecular complexity.* International Journal of Quantum Chemistry, 2002, **91**, 20–31.

[87] Rashevsky N. *Life, information theory, and topology.* Bull. Math. Biophys., 1955, **17**, 229–235.

[88] Rison S.C.G., Thornton J.M. *Pathway evolution, structurally speaking.* Current Opinion in Structural Biology, 2002, **12**, 374–382.

[89] Robinson D.F., Foulds L.R. *Comparison of phylogenetic trees.* Mathematical Biosciences, 1981, **53**, 131–147.

[90] Rosselló F., Valiente G. *Graph transformation in molecular biology.* In: Kreowski H.J., Montanari U., Orejas F., Rozenberg G., Taentzer G., editors, *Formal Methods in Software and Systems Modeling*, pages 116–133, 2005.

[91] Schultz H.P., Schultz E.B., Schultz T.P. *Topological organic chemistry. 4. Graph Theory, Matrix Permanents, and Topological Indices of Alkanes.* Journal of Chemical Information and Computer Sciences, 1992, **32(1)**, 69–72.

[92] Selkow S.M. *The tree-to-tree editing problem.* Information Processing Letters, 1977, **6**, 184–186.

[93] Semple C., Steel M. *Phylogenetics.* Graduate Series Mathematics and its Applications. Oxford University Press, 2003.

[94] Shannon C.E., Weaver W. *The Mathematical Theory of Communication.* University of Illinois Press, Urbana, IL, USA, 1997.

[95] Shi J., Malik J. *Normalized cuts and image segmentation.* Proceedings of IEEE conference on computer vision and pattern recognition (CVPR '97), pages 731–737, 1997.

[96] Simonyi G. *Graph entropy: A survey.* In: Cook W., Lovász L., Seymour P., editors, *Combinatorial Optimization*, Vol. 20, pages 399–441. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 1995.

[97] Skorobogatov V.A., Dobrynin A.A. *Metrical analysis of graphs.* Commun. Math. Comp. Chem., 1988, **23**, 105–155.

[98] Solé R.V., Valverde S. *Information theory of complex networks: On evolution and architectural constraints.* Lecture Notes in Physics, Vol. 650, pages 189–207, 2004.

[99] Steel M., Penny D. *Parsimony, likelihood, and the role of models in molecular phylogenetics.* Molecular Biology and evolution, 2000, **17**, 839–850.

[100] Steel M., Székely L., Mossel E. *Phylogenetic information complexity: Is testing a tree easier than finding it?* Journal of Theoretical Biology, 2009, **258**, 95–102.

[101] Tai K.C. *The tree-to-tree correction problem.* Journal of the Association for Computing Machinery, 1979, **26**, 422–433.

[102] Todeschini R., Consonni V., Mannhold R. *Handbook of Molecular Descriptors.* Wiley-VCH, Weinheim, Germany, 2002.

[103] Trinajstić N. *Chemical Graph Theory.* CRC Press, Boca Raton, FL, USA, 1992.

[104] Trucco E. *A note on the information content of graphs.* Bull. Math. Biol., 1956, **18(2)**, 129–135.

[105] Ulanowicz R.E. *Information theory in ecology.* Computers and Chemistry, 2001, **25**, 393–399.

[106] Vahid F., Dm Le T. *Extending the kernighan/lin heuristic for hardware and software functional partitioning.* Design Automation for Embedded Systems, 1997, **2(2)**, 237–261.

[107] Varmuza K., Demuth W., Karlovits M., Scsibrany H. *Binary substructure descriptors for organic compounds.* Croat. Chem. Acta, 2005, **78**, 141–149.

[108] Wang J., Provan G. *Characterizing the structural complexity of real-world complex networks.* In: Zhou J., editor, *Complex Sciences*, Vol. 4 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 1178–1189, Springer, Berlin/Heidelberg, Germany, 2009.

[109] Washietl S., Gesell T. *Graph representations and algorithms in computational biology of RNA secondary structure.* Structural Analysis of Complex Networks. Birkhäuser/Springer, 2010. To appear.

[110] Wasserman S., Faust K. *Social Network Analysis: Methods and Applications.* Structural Analysis in the Social Sciences. Cambridge University Press, 1994.

[111] Waterman M.S. *Secondary structure of single-stranded nucleic acids.* Studies on foundations and combinatorics, Advances in mathematics supplementary studies, Academic Press N.Y., 1978, **1**, 167–212.

[112] Wiener H. *Structural determination of paraffin boiling points.* Journal of the American Chemical Society, 1947, **69(17)**, 17–20.

[113] Wilkinson D., Huberman B.A. *A method for finding communities of related genes.* PNAS, 2004, **101(1)**, 5241–5248.

[114] Xia K., Fu Z., Hou L. *Impacts of protein-protein interaction domains on organism and network complexity.* Genome Res., 2008, **18**, 1500–1508.

[115] Zachary W. *An information flow model for conflict and fission in small groups.* Journal of Anthropological Research, 1977, **33**, 452–473.

[116] Zuker M., Stiegler P. *Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information.* Nucleic Acids Research, 1981, **9(1)**, 133–148.

Matthias Dehmer, Marina Popovscaia
Institute for Bioinformatics and Translational Research
UMIT, Eduard Wallnoefer Zentrum 1
A-6060, Hall in Tyrol
Austria

E-mail: *matthias.dehmer@umit.at;*
marina.popovscaia@umit.at

Received January 5, 2010