Towards Network Complexity

Matthias Dehmer[1] and Frank Emmert-Streib[2]

[1] Institute for Bioinformatics and Translational Research, UMIT, Eduard Wallnoefer Zentrum 1, 6060, Hall in Tyrol, Austria
Matthias.Dehmer@umit.at

[2] Queen's University Belfast, Computational Biology and Machine Learning, Center for Cancer Research and Cell Biology, School of Medicine, Dentistry and Biomedical Sciences, 97 Lisburn Road, Belfast BT9 7BL, UK
v@bio-complexity.com

Abstract. In this paper, we briefly present a classification scheme of information-based network complexity measures. We will see that existing as well as novel measures can be divided into four major categories: (i) partition-based measures, (ii) non partition-based measures, (iii) non-parametric local measures and (iv) parametric local measures. In particular, it turns out that (ii)-(iv) can be obtained in polynomial time complexity because we use simple graph invariants, e.g., metrical properties of graphs. Finally, we present a generalization of existing local graph complexity measures to obtain parametric complexity measures.

Keywords: networks, network complexity, information measures.

1 Introduction

To find quantitative measures for detecting the complexity of graph-based systems is a research topic with ongoing interest. Here, we are interested in such systems which can be described as complex networks. For example, quantitative approaches to measure network complexity have been developed by [5,13]. In [5], the complexity of a network was defined to be the number of its containing spanning trees. Minoli [13] defined the so-called combinatorial complexity of a network. The key feature of this complexity measure is that it increases with the number of each factor which contributes to the complexity of a network structure. Further approaches to measure complexity of networks can be found in [15]. In this paper, we deal with information-based complexity measures [2,3,14,16]. Classical measures are based on inducing vertex partitions of a network to infer a finite probability distribution. Then, by using Shannon's entropy [17], information-based complexity measures for networks are obtained [1,2,14,16].

The contribution of the paper is twofold: First, we present a classification scheme for information-based network complexity measures. By briefly reviewing existing measures and defining a class of parameterized local complexity measures, we divide the measures into four major categories (see Figure (1)). In

J. Zhou (Ed.): Complex 2009, Part I, LNICST 4, pp. 707–714, 2009.

particular, we present general definitions to obtain so-called information functionals which are based on using efficiently computable graph invariants. We will see that those information functionals [6] can be used for defining both global and local network complexity measures. Further, by applying this principle presented in [6], we obtain straightforward a generalization of existing local graph complexity measures. Local complexity measures are here understood as measures for assigning a complexity score to each vertex of a network. As a final remark, we emphasize that for each category (see Figure (1)) we only give examples for such measures, especially in terms of the partition-based measures [2,3,16,14].

The paper is organized as follows: In Section (2), we begin with introducing basic mathematical preliminaries. Then, we present the classification scheme by starting with Section (3) that deals with classical partition-based network complexity measures. In contrast, Section (4) outlines complexity measures for networks which are not based on inducing vertex partitions. A generalization of existing local complexity measures is introduced in Section (5). The paper finishes with a summary and conclusion in Section (6).

2 Mathematical Preliminaries

We first state some mathematical preliminaries. Especially, we repeat the definitions of some known metrical properties of graphs [9,10,18]. We define an undirected, finite and connected network represented by $G = (V, E), |V| < \infty$, $E \subseteq \binom{V}{2}$. G is called connected if for arbitrary vertices v_i and v_j there exists an undirected path from v_i to v_j. Otherwise, we call G unconnected. $\mathcal{G}_{UC}$ denotes the set of finite, undirected and connected graphs. The degree of a vertex $v \in V$ is denoted by $\delta(v)$ and equals the number of edges $e \in E$ which are incident with v. In order to measure distances between vertices in a graph, we denote $d(u, v)$ as distance between $u \in V$ and $v \in V$ expressed as the minimum length of a path between u, v. $d(u, v)$ is a metric. We call the quantity $\sigma(v) = \max_{u \in V} d(u, v)$ the eccentricity of $v \in V$. Further, $\rho(G) = \max_{v \in V} \sigma(v)$ is called the diameter of G. The j-sphere of a vertex v_i regarding $G \in \mathcal{G}_{UC}$ is defined as $S_j(v_i, G) := \{v \in V \,|\, d(v_i, v) = j,\, j \geq 1\}$. Further metrical properties of graphs can be found in [18].

3 Partition-Based Complexity Measures

In this section, we give a short overview on classical partition-based information measures for determining complexity of networks [2,3,16,14]. These measures which are based on Shannon's entropy formulas are graph entropy measures which can be interpreted as so-called structural information contents. To understand how to apply Shannon's entropy to networks, we start with a network $G = (V, E)$, an equivalence criterion α, and X denotes an arbitrary graph invariant. By applying α, we get the following scheme:

$$\begin{pmatrix} 1 & 2 & \cdots & k \\ |X_1| & |X_1| & \cdots & |X_k| \\ p_1 & p_2 & \cdots & p_k \end{pmatrix}. \quad (1)$$

The first row represents the obtained equivalence classes whereas the second row represents the cardinalities of the induced vertex partitions. Now, by defining the quantities $p_i = \frac{|X_i|}{|X|}$, one obtains directly a finite probability distribution $\mathcal{P}_G = (p_1, \ldots, p_k)$ indicated by the third row. If we now apply the well know formulas for expressing the total and mean information content [3,4], we yield [3]

$$I(G, \alpha) = |X| \log(|X|) - \sum_{i=1}^{k} |X_i| \log(|X_i|), \quad (2)$$

$$\bar{I}(G, \alpha) = -\sum_{i=1}^{k} p_i \log(p_i) = -\sum_{i=1}^{k} \frac{|X_i|}{|X|} \log\left(\frac{|X_i|}{|X|}\right). \quad (3)$$

According to α, Equation (2) denotes the total structural information content of $G = (V, E)$ and Equation (3) denotes the mean structural information content of G, respectively. As examples for such measures, we express, e.g.,

$$I_V(G) = |V| \log(|V|) - \sum_{i=1}^{k} \frac{|N_i|}{|V|} \log\left(\frac{|N_i|}{|V|}\right), \quad (4)$$

$$\bar{I}_V(G) = -\sum_{i=1}^{k} \frac{|N_i|}{|V|} \log\left(\frac{|N_i|}{|V|}\right), \quad (5)$$

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

$$\hat{V} = \{V_i | 1 \leq i \leq h\}; |V_i| = n_i(\hat{V}); h = \chi(G). \quad (6)$$

The first equation (see Equation (4)) is often called the total topological information content [16] of G. Here, the equivalence criterion α corresponds to determine the automorphism group (vertex orbits) of G. The second equation (see Equation (5)) represents the mean topological information content [16] of G by using the same equivalence criterion. $|N_i|$ stands for the cardinality of the i-th vertex orbit. The third equation (see Equation (6)) representing the so-called chromatic information content was originally developed by Mowshowitz [14]. $h = \chi(G)$ denotes the chromatic number whereas $n_i(\hat{V})$ are the cardinalities of the vertex partitions induced by the underlying chromatic decomposition. As an important remark, we note that the time complexity of an algorithm to compute those vertex partitions can be very costly. For example, the problem to calculate the automorphism group of a graph is equivalent to decide wether to graphs are isomorphic. But for arbitrary graphs, there is no efficient algorithm to check if there exists an isomorphism for two given graphs [12]. Similarly, the problem of determining the chromatic number of an undirected graph is NP-complete [10].

4 Non Partition-Based Complexity Measures

To avoid applying algorithms for inducing vertex partitions which are based on algebraic principles (see Section (3)), we outline a method that results in non partition-based complexity measures for networks [6]. We start with arbitrary $G = (V, E)$. If S represents a certain set, e.g., a set of vertices or paths etc., the monotonous and positive mapping $f : S \longrightarrow \mathbb{R}_+$ is called an information functional of G. f captures structural information of a network. If we now define the quantities $p^f(v_i) := \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)}$, we infer a probability distribution $\mathcal{P}_G^f = (p^f(v_1), \dots, p^f(v_{|V|}))$. Instead of determining a probability value for induced vertex partitions, we now assign a probability to each vertex $v \in V$. As a result, we obtain the complexity measure

$$\bar{I}_f(G) := -\sum_{i=1}^{|V|} \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)} \log\left(\frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)}\right). \tag{7}$$

Equation (7) represents a family of information-based network complexity measures. We now state a definition for expressing two possible types of novel information functionals.

Definition 1. *We define two types of information functionals as*

$$f = f(v_i, X_1, \dots, X_\mu, c_1, \dots, c_\mu), \tag{8}$$
$$f = f(v_i, X_1, \dots, X_\mu). \tag{9}$$

X_i represents a graph-theoretical quantity (generally speaking a graph invariant) and c_k positive coefficients, respectively.

Example 1. We exemplary consider the information functionals

$$f_1(v_i) := c_i\delta(v_i),\ c_i > 0, 1 \le i \le |V|, \tag{10}$$
$$f_2(v_i) := \alpha^{c_i\delta(v_i)},\ c_i > 0, 1 \le i \le |V|,\ \alpha > 0, \tag{11}$$
$$f_3(v_i) := \alpha^{c_1|S_1(v_i,G)|+c_2|S_2(v_i,G)|+\cdots+c_{\rho(G)}|S_{\rho(G)}(v_i,G)|},$$
$$c_k > 0,\ 1 \le i \le |V|,\ \alpha > 0, \tag{12}$$
$$f_4(v_i) := d(v_i, v_1) + d(v_i, v_2) + \cdots + d(v_i, v_{|V|}), \tag{13}$$
$$f_5(v_i) := \sigma(v_i). \tag{14}$$

The shown information functionals are defined for $G \in \mathcal{G}_{UC}$. To simplify the notation, we write, e.g., $f_1(v_i)$ instead of $f_1(v_i, \delta(v_i), c_i)$. Especially, it holds $\mu = 1$. Equation (10), Equation (11) and Equation (12) are parametric information functionals which lead to parametric graph complexity measures (using Equation (7)). $f_1(v_i)$ is a linear information functional whereas $f_2(v_i), f_3(v_i)$ are exponential functionals. $f_3(v_i)$ was originally defined in [8] where here the functional can be derived as special case (by setting certain parameters in Definition (1)). Equation (13) and Equation (14) are non-parametric information

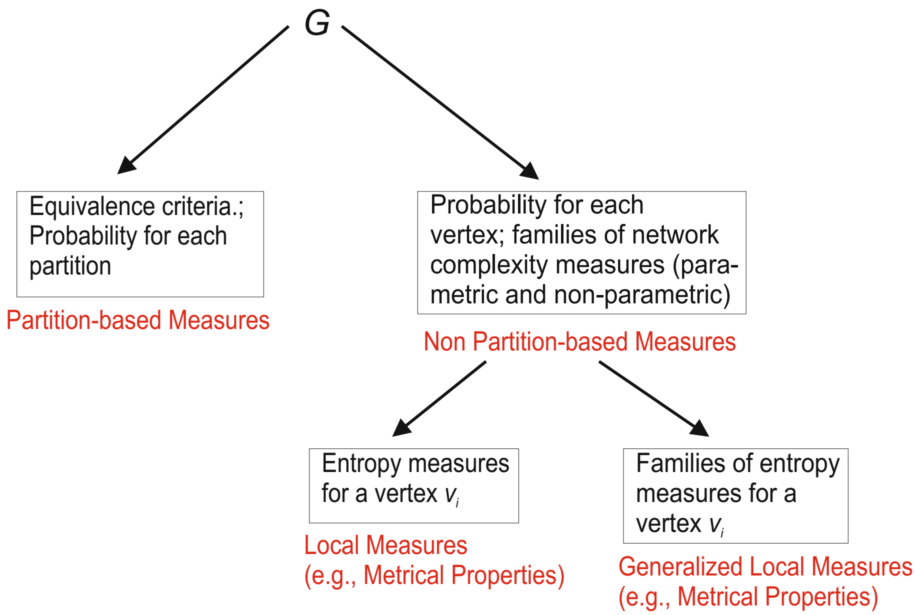

Fig. 1. Classification scheme for information-based network complexity measures

functionals based on metrical properties of graphs. As an important result, we note that especially by using Equation (12) as information functional, the calculation of the resulting graph complexity measure requires polynomial time complexity [7]. For the remaining information functionals, this result can be similarly proven.

5 Local Complexity Measures

In Section (3) and Section (4), we discussed (global) information-based complexity measures for networks. These measures can be used for determining the structural complexity of networks globally. In the following, we sketch a method to obtain local complexity measures. This relates to measure the complexity on local features or substructures of a graph. In particular, we now state a definition to determine information-based complexity for the vertices in a network.

Definition 2. *We define two types of information functionals as*

$$g = g^j(v_i, X_j, c_j), \tag{15}$$

$$g = g^j(v_i, X_j). \tag{16}$$

X_j represents a graph-theoretical quantity (generally speaking a graph invariant) and c_j positive coefficients, respectively.

Definition 3. *Let G be an arbitrary network. In general, we define the mean local complexity of vertex v_i by*

$$\bar{I}_g(v_i) := -\sum_{j=1}^{|V|} \frac{g^j(v_i)}{\sum_{j=1}^{|V|} g^j(v_i)} \log\left(\frac{g^j(v_i)}{\sum_{j=1}^{|V|} g^j(v_i)}\right). \tag{17}$$

Here, we also use the simplified notation for the local information functionals. The index j $(1 \leq j \leq |V|)$ indicates that starting from v_i, a local information functional $g^j(v_i)$ is calculated.

Example 2. We exemplarily consider the information functionals

$$g_1^j(v_i) := d(v_i, v_j),\ 1 \leq i \leq |V|, \tag{18}$$

$$g_2^j(v_i) := c_j d(v_i, v_j),\ 1 \leq i \leq |V|, c_i > 0, \tag{19}$$

$$g_3^j(v_i) := \alpha^{c_j d(v_i, v_j)},\ 1 \leq i \leq |V|, c_i > 0, \alpha > 0. \tag{20}$$

The shown information functionals are defined for $G \in \mathcal{G}_{UC}$. Again, we use the simplified notation. Here, we see that the g^j are based on metrical properties of graphs. Applying Equation (17) and Equation (18) leads to an existing local information complexity measure [11]. Equation (19) and Equation (20) are (possible) generalized versions which finally lead to parametric local complexity measures. The proof to show that the final local graph complexity measure can be computed in polynomial time is very similar to the proof presented in [7].

The overall classification scheme for the presented approaches is depicted in Figure (1).

6 Summary and Conclusion

In this paper, we presented a classification scheme of information-based network complexity measures. We began with classical partition-based complexity measures which are mostly based on inducing vertex partitions of the graph in question. By using pure algebraic principles, this can be a difficult problem. Then, for obtaining measures with better time complexity, we outlined a recently contributed method that leads to families of information-based network complexity measures. Based on Definition (1), we saw that novel information functionals can be easily inferred (see, e.g., Equation (10), (11), (14)). Finally, we expressed Definition (2) for obtaining local complexity measures (for vertices). By stating this definition, we generalized existing local complexity measures to measure the information distance for vertices of a graph. As result, local parametric measures can be easily obtained. In general, these information measures can be used to detect structural complexity of graph-based systems, e.g., in biology and chemistry. Especially in QSPR (quantitative structure property relationship) that is branch of mathematical chemistry, a main problem is to characterize molecules by using information-theoretic complexity measures. The complexity measures presented in Section (3) and Section (4) are suitable to characterize graphs (e.g., chemical structures) globally. That means starting from a graph, inferred structural features and using SHANNON's entropy, we obtain a value for its information content. In contrast, the measures given in Section (5) address the problem of calculating entropies of local graph elements, e.g., vertices. In principle, this gives us the possibility to study the importance of such graph elements among each other or between different graphs.

As future work, we will apply these measures (local and global) for analyzing biological networks and combine them with statistical techniques. Moreover, we are interested in comparing the measures presented in this paper numerically (measures of the four major categories). From this, one can gain novel insights regarding the problem of studying the interplay between the measures.

Acknowledgments

We would like to thank Danail Bonchev, Alexandru T. Balaban and Abbe Mowshowitz for fruitful discussions. Support from the 2006 Ciência fund (Portugal) is gratefully acknowledged.

References

1. Bonchev, D.: Information indices for atoms and molecules. MATCH 7, 65–113 (1979)
2. Bonchev, D.: Information Theoretic Indices for Characterization of Chemical Structures. Research Studies Press, Chichester (1983)
3. Bonchev, D., Rouvray, D.H.: Complexity in Chemistry, Biology, and Ecology. In: Mathematical and Computational Chemistry. Springer, Heidelberg (2005)
4. Brillouin, L.: Science and Information Theory. Academic Press, New York (1956)
5. Constantine, G.: Graph complexity and the laplacian matrix in blocked experiments. Linear and Multilinear Algebra 28, 49–56 (1990)
6. Dehmer, M.: Information processing in complex networks: Graph entropy and information functionals. Applied Mathematics and Computation 201, 82–94 (2008)
7. Dehmer, M.: Information-theoretic concepts for the analysis of complex networks. Applied Artificial Intelligence (in press) (2008)
8. Dehmer, M.: A novel method for measuring the structural information content of networks. In: Cybernetics and Systems (in press) (2008)
9. Halin, R.: Graphentheorie. Akademie Verlag (1989)
10. Harary, F.: Graph Theory. Addison-Wesley, Reading (1969)
11. Konstantinova, E.V., Skorobogatov, V.A., Vidyuk, M.V.: Applications of information theory in chemical graph theory. Indian Journal of Chemistry 42, 1227–1240 (2002)
12. McKay, B.D.: Graph isomorphisms. Congressus Numerantium 730, 45–87 (1981)
13. Minoli, D.: Combinatorial graph complexity. Atti. Accad. Naz. Lincei, VIII. Ser., Rend., Cl. Sci. Fis. Mat. Nat. 59, 651–661 (1975)

14. Mowshowitz, A.: Entropy and the complexity of the graphs I: An index of the relative complexity of a graph. Bull. Math. Biophys. 30, 175–204 (1968)
15. Neel, D.L., Orrison, M.E.: The linear complexity of a graph. The Electronic Journal of Combinatorics 13 (2006)
16. Rashevsky, N.: Life, information theory, and topology. Bull. Math. Biophys. 17, 229–235 (1955)
17. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press (1997)
18. Skorobogatov, V.A., Dobrynin, A.A.: Metrical analysis of graphs. MATCH 23, 105–155 (1988)