NETWORK COMPLEXITY MEASURES. AN INFORMATION-THEORETIC APPROACH

Matthias Dehmer[1] *and Stefan Pickl*[1]

[1] Department of Computer Science
Universität der Bundeswehr München
Werner-Heisenberg-Weg 39
85577 Neubiberg, Germany

1. INTRODUCTION

Quantitative graph analysis by using structural indices has been intricate in a sense that it often remains unclear which structural graph measures is the most suitable one, see [1, 12, 13]. In general, quantitative graph analysis deals with quantifying structural information of networks by using a measurement approach [5]. As special problem thereof is to characterize a graph quantitatively, that means to determine a measure that captures structural features of a network meaningfully. Various classical structural graph measures have been used to tackle this problem [13]. A fruitful approach by using information-theoretic [21] and statistical methods is to quantify the structural information content of a graph [1, 8, 18].

In this note, we sketch some classical information measures. Also, we briefly address the problem what kind of measures capture structural information uniquely. This relates to determine the discrimination power (or also called uniqueness) of a graph measure, that is, how is the ability of the measures to discriminate non-isomorphic graphs structurally.

2. GRAPH ENTROPY MEASURES FOR NETWORKS

2.1. Classical Measures

Many classical information indices for characterizing graphs are based on grouping the elements given by an arbitrary graph invariant according to a certain equivalence criterion [1, 18]. Typical graph invariants are vertices, edges, vertex degrees and distances in a graph. As a result, we derive probability values for each group (partition) and finally the structural information content of a network. The structural information content of a graph is the entropy of the underlying graph topology.

Let $G = (V, E)$ be a graph, let X be a graph invariant and α is assumed to be an equivalence criterion. By applying α, distributions of X may be obtained and thus partitions X_i with cardinality $|X_i|$. According to the scientific literature, the following general graph entropy measures have been developed [1]:

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

and

$$\bar{I}(G,\alpha) := -\sum_{i=1}^{k} P_i \log(P_i) = -\sum_{i=1}^{k} \frac{|X_i|}{|X|}\log\left(\frac{|X_i|}{|X|}\right). \quad (2)$$

k is the number of partitions. Concrete information measures for graphs have been also developed [1, 18]:

$$\bar{I}_{orb}(G) := -\sum_{i=1}^{k} \frac{|N_i|}{|V|}\log\left(\frac{|N_i|}{|V|}\right), \quad (3)$$

is called the topological information content of G, see [18, 19]. Here, $|N_i|$ denotes the number of topologically equivalent vertices in the i-th vertex orbit of G. k stands for the number of different orbits. A similarly defined measure that is based on determining the edge orbits of G is due to Trucco [26]:

$$^{E}\bar{I}_{orb}(G) := -\sum_{i=1}^{k} \frac{|N_i^E|}{|E|}\log\left(\frac{|N_i^E|}{|E|}\right). \quad (4)$$

$|N_i^E|$ denotes the number of edges belonging to the i-th edge orbit [1] of G.

Mowshowitz [18] further developed an important information measure using chromatic decompositions of graphs:

$$I_{cr}(G) := \min_{\hat{V}} \left\{ -\sum_{i=1}^{h} \frac{n_i(\hat{V})}{|V|} \log\left(\frac{n_i(\hat{V})}{|V|}\right) \right\}, \quad (5)$$

where $\hat{V} = \{V_i | 1 \leq i \leq h\}$. $|V_i| = n_i(\hat{V})$ denotes an arbitrary chromatic decomposition of a graph G, $h = \chi(G)$ is the chromatic number of G. Finally, Bonchev [1] generalized these classical partition-based measures by introducing weighted probability distributions. For example, he derived the so-called magnitude-based information index [1],

$$\begin{aligned} \bar{I}_D(G) := &-\frac{1}{|V|}\log\left(\frac{1}{|V|}\right) \\ &-\sum_{i=1}^{\rho(G)} \frac{2k_i}{|V|^2} \log\left(\frac{2k_i}{|V|^2}\right). \quad (6) \end{aligned}$$

Here, we assume that the distance of a value i in the distance matrix appears $2k_i$ times. $\rho(G)$ is the diameter of a graph G.

2.2. Complexity Measures based on Information Functionals

Dehmer developed an approach to derive information measures for graphs which are based on using so-called information functionals, see [4, 8]. An information functional is a mapping that captures structural information of a graph, see [4, 8]. Instead of determining probability values for each obtained partition (see previous section), one defines a probability value to each vertex in the graph. We derive

$$p_f(v_i) := \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)}, \quad \forall\, v_i \in V. \quad (7)$$

f represents an arbitrary information functional. In fact, these values are vertex probabilities as

$$p_f(v_1) + p_f(v_2) + \ldots + p_f(v_{|V|}) = 1. \quad (8)$$

By employing the obtained vertex probabilities, the entropy of the underlying graph topology of G has been defined as [4, 8]:

$$\begin{aligned} I_{f^\star}(G) &:= -\sum_{i=1}^{|V|} p_f(v_i) \log\left(p_f(v_i)\right), \quad (9) \\ &= -\sum_{i=1}^{|V|} \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)} \log\left(\frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)}\right). \quad (10) \end{aligned}$$

An example of such information functional is $f(v_i) := \sigma(v_i)$. $\sigma(v_i)$ is the eccentricity of $v_i \in V$. More recent work relates to examine the degree-powers, see [3]. If $f^\star(v_i) := d_i^k$, $k > 0$, we yield

$$I_f(G,k) = -\sum_{i=1}^{|V|} \frac{d_i^k}{\sum_{j=1}^{|V|} d_j^k} \log\left(\frac{d_i^k}{\sum_{j=1}^{|V|} d_j^k}\right). \quad (11)$$

We see $I_{f^\star}(G, k = 1)$ is a special case that yields to

$$\begin{aligned} I_f(G) &:= I_{f^\star}(G, k = 1) \quad (12) \\ &= -\sum_{i=1}^{|V|} \frac{d_i}{\sum_{j=1}^{|V|} d_j} \log\left(\frac{d_i}{\sum_{j=1}^{|V|} d_j}\right). \end{aligned}$$

Here, $f(v_i) := d_i$. Note that Cao et al. [3] proved extremal properties of this graph entropy measure. But most of the results obtained in [3] have been obtained for $k = 1$. The case $k > 1$ has been more intricate, see [3].

Other information functionals have been developed too, see [4, 8]. We emphasize that the resulting measures have been used in chemical and biological network analysis [10, 11].

3. DISCRIMINATION POWER OF GRAPH MEASURES

The discrimination power or uniqueness of graph measures relates to the ability to discriminate the structure of non-isomorphic graphs. Following Todeschini et al. [24], the degeneracy of a graph measure is an undesired aspect as from a theoretical point of view, non-isomorphic graphs should be distinguished by the measure. With other words it does not make sense that a particular graph measure maps non-isomorphic graphs to the same measured value. The situation may be different when it comes to special classification in structural chemistry or data mining problems, see [23, 25].

Now we briefly survey the main results in this area. Classical work in this area is due to Bonchev et al. [2] and Konstantinova [16, 17] who determined the degeneracy of information-theoretic graph measures by using sets of special chemical structures. The indices were simple information-theoretic indices based on distances in graphs [1, 17].

However, these sets are quite small and, therefore, they are not suitable to perform any statistical analysis on a large scale. Also, it is difficult to generalize these results as (small) special graph classes were used only. Dehmer et al. [8, 7] tackled the problem differently by performing a compelling analysis on exhaustively generated networks. For instance, the generated all connected and non-isomorphic graphs from 5 to 10 vertices. Note that by considering the graphs having 10 vertices each we deal already with almost 12 million graphs. In order to perform this analysis, they calculated various known structural indices on exhaustively generated trees and graphs and found that most of the measures are not unique at all [8, 7]. That means their discrimination power is very low and they are not suitable to discriminate graphs practically [8, 7]. As a further result, a special information-theoretic quantity was found to be highly discriminating on exhaustively generated graphs [8, 7]. Thus it represents a highly discriminating graph invariant which can be useful for structure searching and graph isomorphism testing.

4. MEASURES FOR COMPARATIVE GRAPH ANALYSIS

Besides graph characterization by using information-theoretic measures (see preceding sections), comparative graph analysis has been an important problem. It relates to compare graphs by using graph distance or similarity measures [9, 14, 15, 22, 27].

Here, we want to sketch an approach for defining comparative graph measures by using real similarity and distance measure for real numbers namely $s : \mathcal{G} \times \mathcal{G} \longrightarrow$ R and $d : \mathcal{G} \times \mathcal{G} \longrightarrow$ R. $\mathcal{G}$ is a class of graphs. By employing graph measures like $I : \mathcal{G} \longrightarrow$ R, we get comparative graph measures such as

$$d(G_1, G_2) := |M(G_1) - M(G_2)|. \tag{13}$$

Also, by using the real distance measure [20]

$$d(x, y) = 1 - e^{-\left(\frac{x-y}{\sigma}\right)^2}, \tag{14}$$

we obtain [6]

$$d_I(G, H) := d(I(G), I(H)) = 1 - e^{-\left(\frac{I(G)-I(H)}{\sigma}\right)^2}. \tag{15}$$

It can be easily verified that $d(x, x) = 0$, $d(x, y) \geq 0$, and $d(x, y) = d(y, x)$ holds. Properties of this measure have been investigated in [6]. Moreover it has been demonstrated that this measure also has useful properties when applying it to structural data sets [6].

5. SUMMARY AND CONCLUSION

We sketched classical and recent information indices for performing quantitative network analysis. Also, we discussed the problem of measuring the discrimination power of structural graph measures. Further we sketched an approach to come up with comparative graph measures by employing graph measures. This class of measures will be investigated in depth as future work.

We hope that the problems we have tackled here are stimulating for those who deal with structural network analysis and who want to apply this apparatus interdisciplinarily.

6. ACKNOWLEDGEMENT

Matthias Dehmer and Stefan Pickl gratefully acknowledges financial support from the German Federal Ministry of Education and Research (BMBF) (project RiKoV, Grant No. 13N12304).

7. REFERENCES

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

[2] D. Bonchev, O. Mekenyan, and N. Trinajstić. Isomer discrimination by topological information approach. *J. Comp. Chem.*, 2(2):127–148, 1981.

[3] S. Cao, M. Dehmer, and Y. Shi. Extremality of degree-based graph entropies. *Information Sciences*, 278:22–33, 2014.

[4] M. Dehmer. Information processing in complex networks: Graph entropy and information functionals. *Appl. Math. Comput.*, 201:82–94, 2008.

[5] M. Dehmer and F. Emmert-Streib. *Quantitative Graph Theory. Theory and Applications*. CRC Press, 2014.

[6] M. Dehmer, F. Emmert-Streib, and Y. Shi. Interrelations of graph distance measures based on topological indices. *PLoS ONE*, 9:e94985, 2014.

[7] M. Dehmer and M. Grabner. The discrimination power of molecular identification numbers revisited. *MATCH Commun. Math. Comput. Chem.*, 69(3):785–794, 2013.

[8] M. Dehmer, M. Grabner, and K. Varmuza. Information indices with high discriminative power for graphs. *PLoS ONE*, 7:e31214, 2012.

[9] M. Dehmer and A. Mehler. A new method of measuring similarity for a special class of directed graphs. *Tatra Mountains Mathematical Publications*, 36:39–59, 2007.

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

[11] F. Emmert-Streib and M. Dehmer. Networks for systems biology: Conceptual connection of data and function. *IET Systems Biology*, 5:185–207, 2011.

[12] F. Emmert-Streib and M. Dehmer. Exploring statistical and population aspects of network complexity. *PLoS ONE*, 7:e34523, 2012.

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

[14] F. Kaden. Graphmetriken und Distanzgraphen. *ZKI-Informationen, Akad. Wiss. DDR*, 2(82):1–63, 1982.

[15] F. Kaden. Halbgeordnete Graphmengen und Graphmetriken. In H. Sachs, editor, *Graphs, Hypergraphs, and Applications*, volume 73, pages 92–95. Teubner, 1985. Leipzig, DDR.

[16] E. V. Konstantinova. The discrimination ability of some topological and information distance indices for graphs of unbranched hexagonal systems. *J. Chem. Inf. Comput. Sci.*, 36:54–57, 1996.

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

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

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

[20] C. Schädler. *Die Ermittlung struktureller Ähnlichkeit und struktureller Merkmale bei komplexen Objekten: Ein konnektionistischer Ansatz und seine Anwendungen*. PhD thesis, Technische Universität Berlin, 1999.

[21] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.

[22] F. Sobik. Modellierung von Vergleichsprozessen auf der Grundlage von Ähnlichkeitsmaßen für Graphen. *ZKI-Informationen, Akad. Wiss. DDR*, 4:104–144, 1986.

[23] R. Todeschini, R. Cazar, and E. Collina. The chemical meaning of topological indices. *Chemometrics and Intelligent Laboratory Systems*, 15:51–59, 1992.

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

[25] R. Todeschini, V. Consonni, and A. Maiocchi. The k correlation index: theory development and its applications in chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 46:13–29, 1998.

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

[27] B. Zelinka. On a certain distance between isomorphism classes of graphs. *Časopis pro pěst. Mathematiky*, 100:371–373, 1975.