A COMPARATIVE STUDY OF COMPLEXITY MEASURES TO ANALYZE BUSINESS PROCESS MODELS

Stephan Borgert, Matthias Dehmer and Erwin Aitenbichler

Abstract. This work reports on structural complexity measures to be applied to complex business process models. Here, we mainly focus on applying information-theoretic measures to characterize the underlying process graph structurally. Moreover, we compare the considered measures by using a set of certain process graphs.[1]

Keywords: *complexity, entropy, business processes*

2000 *Mathematics Subject Classification*: 68R10, 94C15

1. Introduction

Business processes can be understood as coordinated tasks and activities either instantiated by human beings or an equipment which lead to fulfilling a specific organizational goal. Hence, the task of analyzing business processes is an important issue for companies and, therefore, has been a central research area since years [11]. A challenging area in this research field research deals with investigating quantitative methods to establish certain metrics, see, e.g., [13]. One important example for such quantitative measures to analyze business process models relates to analyze their complexity. Generally, the problem of measuring the complexity can not be uniquely defined because complexity is in the eye of the beholder.

In this paper, we want to put the emphasis on measures which are suitable to quantify the structural complexity of business process models [11]. So far,

[1]The project was funded by means of the German Federal Ministry of Economy and Technology under the promotional reference 01MQ07012. The authors take the responsibility for the contents.

some studies have been performed to examine the just mentioned problem, e.g., [2, 1, 7, 10, 11, 16]. For example, it turned out that unnecessarily high complexity can lead to an increasing probability for errors in such models and, thus, raises additional problems when maintaining these models [13]. Further, this kind of unnecessary complexity can also have a negative effect on the comprehensibility of the models [2, 1, 7, 13]. Interestingly, Mendling [13] investigated about 2000 processes and concluded that the more errors he found, the more complex was the corresponding area of the model [13].

As mentioned, structural complexity measures to analyze business processes have already been proposed in [11]. The contribution of this paper is to apply information-theoretic measures to quantify the structural complexity of business processes. Further, we compare the results with some other existing (non-information-theoretic) measures. By saying complexity, we here always refer to the problem of measuring the structural complexity of the underlying *process graph*. We think that applying a statistical framework to these process models, e.g., information-theoretic complexity measures can have a substantial impact when dealing with erroneous graphs [6]. This issue is almost unexplored so far and, therefore, the paper can be seen as a first attempt to analyze process graphs statistically (i.e., by using information-theoretic methods).

2. Complexity Measures

First, we briefly introduce some mathematical preliminaries [3, 9, 8] to formulate our approach. For this, we first start with the definition of finite, undirected and connected graphs. We call $G = (V, E), |V| < \infty$, $E \subseteq \binom{V}{2}$ a finite, undirected and connected graph. whereas $\mathcal{G}_{UC}$ denotes the set of such graphs. Moreover, we also repeat the definitions of some metrical properties of graphs [15]. $d(u, v)$ denotes the shortest distance between $u \in V$ and $v \in V$. For $G \in \mathcal{G}_{UC}$, $\sigma(v) = \max_{u \in V} d(u, v)$ is called the eccentricity of $v \in V$ and $\rho(G) = \max_{v \in V} \sigma(v)$ the diameter of G, respectively.

In the following, we state the definitions of some complexity measures we want to apply for performing our study. A well-known complexity measure that has been intensely used a branching index is the Wiener index [18]. It is defined by

$$W(G) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}(v_i, v_j) = \frac{1}{2} \sum_{j=1}^{n} d(v_i), \tag{1}$$

where

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

Starting from a graph $G = (V, E)$, the Randić connectivity index [14] is defined by

$$R(G) = \sum_{(v_i,v_j)\in E} [\delta(v_i)\delta(v_j)]^{-\frac{1}{2}}, \tag{3}$$

where $\delta(v_i)$ is called the degree of the vertex $v_i \in V$. $\delta(v_i)$ equals the number $e \in E$ which are incident with v_i.

The cyclomatic number was introduced by McCabe [12] and is defined as

$$C(G) = |E| - |V| + p \tag{4}$$

where p is the number of components. A utilized process graph possesses a unique entry node and usually several exit nodes. Here, the number of components is given by the number of exit nodes. As a result, the cyclomatic number was used for the determination of complexity in the program code of software [12].

Now, we define the entropy of the underlying topology of a process graph where we here utilize the graph entropy measure due to Dehmer [4, 5]. We want to point out that this measure has been used in several contexts, e.g., to characterize chemical structures [5] and to examine the uniqueness of such measures when being applied to large chemical databases [17].

Now, we start by defining the set

$$S_j(v_i, G) := \{v \in V \,|\, d(v_i, v) = j,\ j \geq 1\}. \tag{5}$$

$S_j(v_i, G)$ denotes the j-sphere of v_i regarding $G \in \mathcal{G}_{UC}$. To define a probability distribution, we use the entities

$$p^V(v_i) := \frac{f^V(v_i)}{\sum_{j=1}^{|V|} f^V(v_j)}. \tag{6}$$

By now using the definition of the j-spheres, see Equation (5), we further derive the information functional

$$f^V(v_i) := c_1|S_1(v_i, G)| + c_2|S_2(v_i, G)| + \cdots + c_{\rho(G)}|S_{\rho(G)}(v_i, G)|, \quad c_k > 0, 1 \leq k \leq \rho(G). \tag{7}$$

Finally, we end up with a family of entropy measures

$$I_{f^V}(G) := -\sum_{i=1}^{|V|} \frac{f^V(v_i)}{\sum_{j=1}^{|V|} f^V(v_j)} \log\left(\frac{f^V(v_i)}{\sum_{j=1}^{|V|} f^V(v_j)}\right), \tag{8}$$

to quantify the structural complexity of process graphs. To infer a concrete information measure, we choose $c_1 := \rho(G)$, $c_2 := \rho(G) - 1, \ldots, c_{\rho(G)} := 1$.

3. Results

In order to evaluate the presented complexity measures we created process graphs from a set of six different reference processes. The names of the processes, their number of nodes and the diameter of the corresponding process graphs are shown in Table 1.

Table 1: Overview of the used set of process graphs

Business Process	$\lvert V\rvert$	$\rho(G)$
Material Procurement	11	5
Purchaser – Vendor	13	5
Invoicing	13	5
Application for permit	13	6
Vacation Request	13	7
Credit Management	13	10

Figures 1–4 show the obtained results. On the x-axis (Figures 1–3) we plotted the values of the entropy measure and the y-axis represents the value range of the remaining complexity measures. All values were normalized. Figures 1 and 2 indicate that our entropy-based measure is almost linearly correlated with Wiener Index and Randic Index. In contrast to that, the correlation between the cyclomatic number and our entropy measure shows a different behavior.

Figure 4 summarizes the results of our short numerical section. One can clearly see that the used complexity measures capture structural information differently. In addition, we observe that the entropy values are always higher than the values of the other measures. An overlap is only given at the normalization point. We will investigate this property in more detail in our future work.

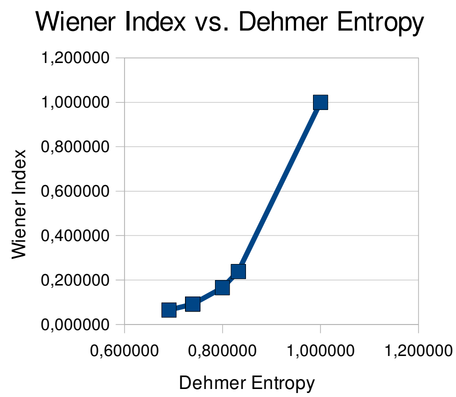

Figure 1: Correlation between Entropy and Wiener Index.

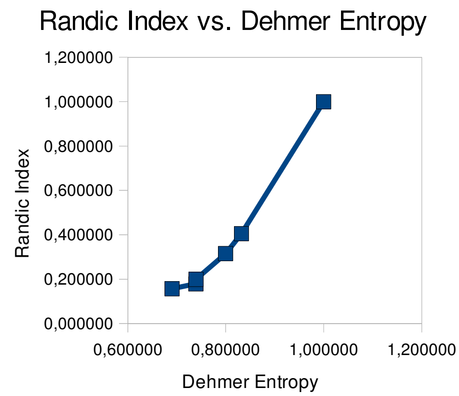

Figure 2: Correlation between Entropy and Randic Index.

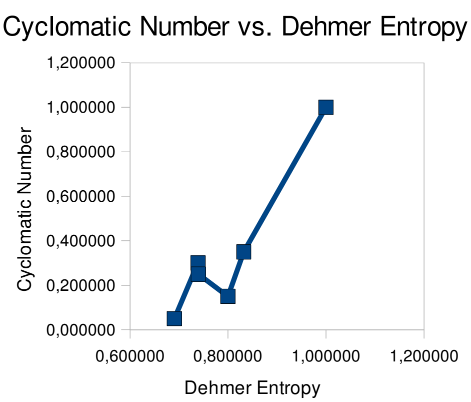

Figure 3: Correlation between Entropy and Cyclomatic Number.

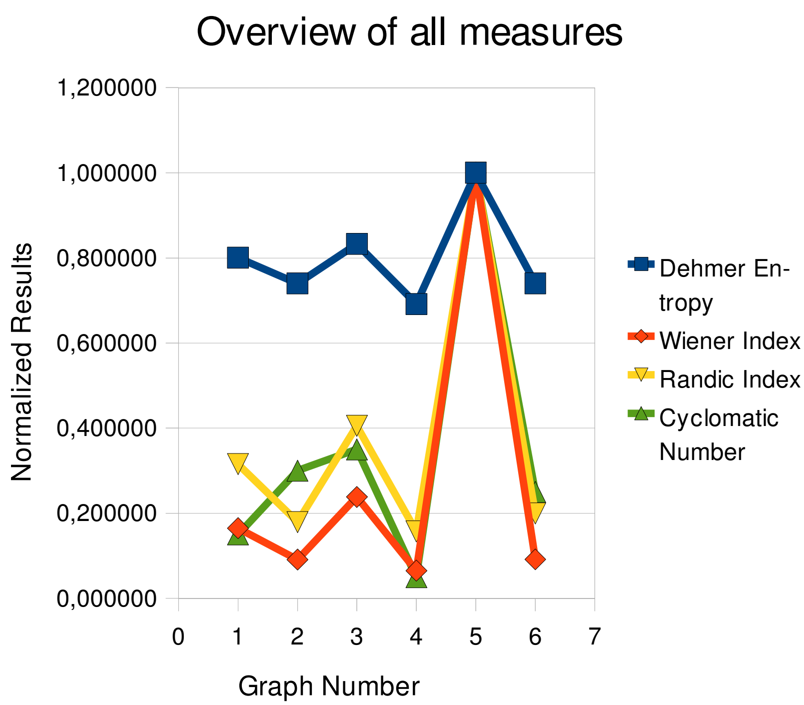

Figure 4: Normalized Entropies vs. Graphs with increasing order.

References

[1] Jorge Cardoso. Evaluating the process control-flow complexity measure. In *2005 IEEE International Conference on Web Services, 2005. ICWS 2005. Proceedings*, page 804, 2005.

[2] Jorge Cardoso. Business Process Quality Metrics: Log-Based Complexity of Workflow Patterns. *Lecture Notes in Computer Science*, 4803:427, 2007.

[3] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing. Wiley & Sons, 2006.

[4] Matthias Dehmer. A Novel Method for Measuring the Structural Information Content of Networks. *Cybernetics and Systems*, 39:825–843, 2008.

[5] Matthias Dehmer and Frank Emmert-Streib. Structural Information Content of Chemical Networks. *Zeitschrift für Naturforschung, Part A*, 63a:155–159, 2008.

[6] F. Emmert-Streib and M. Dehmer. Information processing in the transcriptional regulatory network of yeast: Functional robustness. *BMC Systems Biology*, 3, 2009. doi:10.1186/1752-0509-3-35.

[7] Volker Gruhn and Ralf Laue. Complexity metrics for business process models. In *9th international conference on business information systems (BIS 2006)*, volume 85, pages 1–12. Citeseer, 2006.

[8] Rudolf Halin. *Graphentheorie*. Akademie Verlag, 1989. Berlin, Germany.

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

[10] Jae-Yoon Jung. Measuring Entropy in Business Process Models. In *Innovative Computing Information and Control, 2008. ICICIC'08. 3rd International Conference on*, pages 246–246, 2008.

[11] Antti M. Latva-Koivisto. Finding a complexity measure for business process models. *Helsinki University of Technology, Systems Analysis Laboratory*, 2001.

[12] TJ McCabe. A complexity measure. *IEEE Transactions on software Engineering*, pages 308–320, 1976.

[13] Jan Mendling. *Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness (Lecture Notes in Business Information Processing)*. Springer, 1 edition, 11 2008.

[14] Milan Randić and Dejan Plavšić. On the concept of molecular complexity. *Croat. Chem. Acta*, 75:107–116, 2002.

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

[16] Jos J.M. Trienekens, Rob Kusters, Dirk Kriek, and Paul Siemons. Entropy based software processes improvement. *Software Quality Journal*, 17(3):231–243, 2009.

[17] Kurt Varmuza, Matthias Dehmer, and Stephan Borgert. Molecular Descriptors Based on Entropy and the Full Topological Neighborhood of All Atoms. In *3rd German Conference on Chemoinformatics - 21. CIC-Workshop*, 2008. accepted poster presentation.

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

Stephan Borgert, Darmstadt University of Technology, Hochschulstr. 10, 64289 Darmstadt, Germany,
email: *stephan@tk.informatik.tu-darmstadt.de*

Matthias Dehmer, Discrete Mathematics and Geometry, Vienna University of Technology, Wiedner Hauptstrasse 8-10, A-1040 Vienna, Austria,
email: *mdehmer@dmg.tuwien.ac.at*

Erwin Aitenbichler, Darmstadt University of Technology, Hochschulstr. 10, 64289 Darmstadt, Germany,
email: *erwin@tk.informatik.tu-darmstadt.de*