ON QUANTITATIVE NETWORK COMPLEXITY MEASURES FOR BUSINESS PROCESS MODELS

Stephan Borgert, Matthias Dehmer and Erwin Aitenbichler

Abstract. This work investigates measures to quantify structural complexity to be applied to complex business process models. In particular, we mainly focus on applying information-theoretic measures to characterize the underlying process graph structurally. Moreover, we examine the relatedness between these measures by using a set containing business processes representing networks.[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 [12]. A challenging area in this research field is the investigation of quantitative methods to establish certain metrics, e.g., to estimate the probability of errors [14]. One important example for such quantitative measures to analyze business process models relates to determining their complexity. Generally, the problem of measuring the complexity can not be uniquely defined, because complexity largely lies in the eye of the beholder.

[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.

In this paper, we want to put the emphasis on measures which are suitable to quantify the structural complexity of business process models [12]. So far, some studies have been performed to examine the just mentioned problem, e.g., [2, 1, 7, 10, 12, 17]. For example, it turned out that unnecessarily high complexity can lead to an increasing probability of errors in such models and, thus, raises additional problems when maintaining these models [14]. Furthermore, this kind of unnecessary complexity can also have a negative effect on the comprehensibility of the models [2, 1, 7, 14]. Interestingly, Mendling [14] investigated about 2000 processes and concluded that the more errors he found, the more complex was the corresponding area of the model [14].

As mentioned, structural complexity measures to analyze business processes have already been proposed in [12]. The contribution of this paper is to apply information-theoretic measures to quantify the structural complexity of business processes. In addition, 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 [16]. $d(u, v)$ 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 as a branching index is the Wiener index [19]. It is defined by

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

where

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

Starting from a graph $G = (V, E)$, the Randić connectivity index [15] is defined by

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

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 [13] and is defined as

$$C(G) = |E| - |V| + p \quad (4)$$

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 [13].

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 [18].

Now, we start by defining the set

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

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

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

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

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

We end up with a family of entropy measures

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

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

In order to evaluate how unique the measures are, we determine their degeneracy, see, e.g., [11]. The uniqueness could be used to measure if the index would change when performing graph operations that result in only small differences in the networks under consideration. To determine the degeneracy, we calculate the sensitivity measure

$$S(I) = \frac{|N| - |N_i|}{|N|}, \quad (9)$$

of a topological Index I with respect to a set of graphs denoted by N. Generally, $|N|$ stands for the cardinality of N and $|N_i|$ stands for the number of Graphs $G_i \in N$ which can not be distinguished (by using I), respectively. By definition, it holds $S(I) = 1$ iff it does not exist any pair of non-isomorphic graphs $G \in N$ whose graphs possess the same value of I.

3 Scenario

In this work, we chose the Eco Calculator service prototype from the Theseus project as evaluation scenario. This service calculates the ecological impact of production, transportation, recycling, etc. of a product and automatically issues compliance certificates accordingly. A service engineer may discover one or multiple such services on the Internet and send construction details of, e.g., a car seat to this service. The behavior of the Eco Calculator service is described by a formal business process. The process description language has a graphical notation and well-defined semantics, based on a process calculus.

Because the process involves knowledge of various disciplines, we assume that the provider of the Eco Calculator service is willing to outsource certain parts of the process.

Hence, the Eco Calculator is decomposed into three different service types: the Eco Calculator service itself, a chemical database service, and a chemical laboratory service. All services are described by formal business processes. The whole process model is now formed by a service composition that integrates all three different types of services. For each service type, we have modeled five different variants, which all differ a bit in their behavior. This resembles the practical situation, where different service providers will provide services showing an equivalent behavior on an abstract level, but differ in small details. For example, every Eco Calculator service can receive construction details and issue certificates, but there may be differences in administrative processes, such as delivery or payment.

4 Evaluation

By combining the three different service types with five different services for each type, we created 125 process models. Note that the generated process models do not differ much based on their underlying network structure. As future work, our hypothesis is that a change of behavior will have an impact on the structure of the process model and, therefore, also affects the complexity of this model. That is the reason why we particularly need such measures which are sensitive to small changes within the network structure. Besides examining the correlations between the used information-theoretic measures, we also investigate the discrimination power of the selected complexity measures by using Equation (9).

Figures 1-4 show the obtained results. On the x-axis we plotted the values of the reference entropy measure and the y-axis represents the value range of the comparison entropy measure. All values were normalized. Figure 1 depicts the correlation between the Wiener Index and the entropy due to Dehmer.

From Figures 1-3, we clearly see that underlying topological measures are highly uncorrelated. Hence, we can conclude that these measures detect structural complexity very differently. In contrast, when we visualize the correlation between Wiener- and Randić index, we observe that there is a linear correlation with increasing values of both indices (Figure 4).

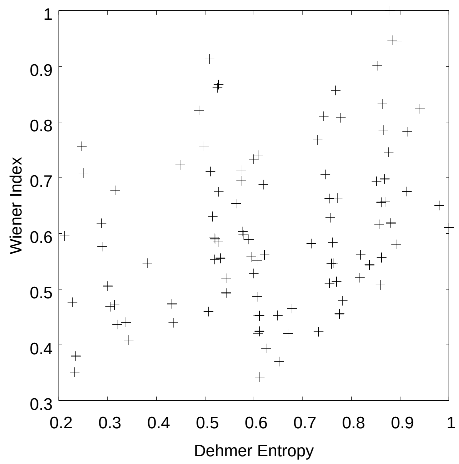

Figure 1: Correlation between Dehmer Entropy and Wiener Index.

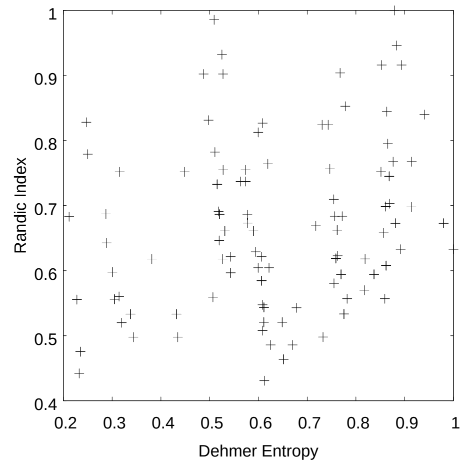

Figure 2: Correlation between Dehmer Entropy and Randić Index.

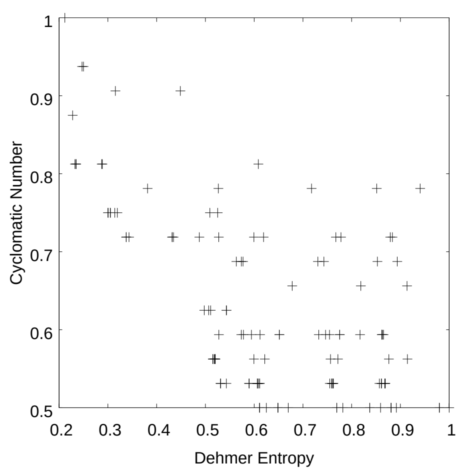

Figure 3: Correlation between Dehmer Entropy and Cyclomatic Number.

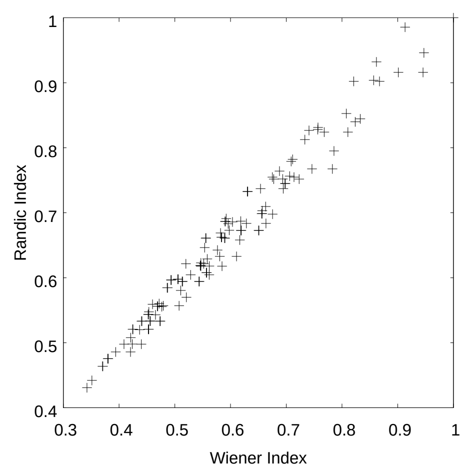

Figure 4: Correlation between Wiener Index and Randić Index.

In Figure 3, the measurement values appear to be aligned along several discrete vertical lines. This indicates that the cyclomatic number does not have a high discrimination power to characterize the graphs uniquely. Table 1 summarizes the results when evaluating the degeneracy of the used measures. All calculations were performed four times, where the number of decimal places was increased in each step.

The best result was obtained with the Dehmer Entropy, which allows to uniquely distinguish between all graphs starting with five decimal places. The Cyclomatic Number gives the worst result and virtually does not allow to distinguish between graphs. The Randić index performs not much better and shows only a very little discrimination power. Interestingly, the Wiener Index performs quite well when applied to our process graphs. In contrast to this result, it has only little discrimination power when it is applied to chemical graphs [11].

Table 1: Comparison of Complexity Measures

	Dehmer Entropy		Wiener Index		Randić Index		Cyclomatic #	
Digits	N_i	$S(I)$	N_i	$S(I)$	N_i	$S(I)$	N_i	$S(I)$
2	113	0.096	106	0.152	110	0.1200	123	0.016
3	56	0.552	31	0.752	86	0.312	123	0.016
4	8	0.936	6	0.952	80	0.360	123	0.016
5	0	1.000	6	0.952	80	0.360	123	0.016

5 Summary and Conclusion

In this paper, we applied structural complexity measures to graphs representing business processes and compared the results. As data set, we used 125 processes graphs which were quite similar to each other. Hence, we searched for a measure that is able to discriminate graphs uniquely when allowing only a low number of decimal places. As a result, we found that the entropy measure developed by Dehmer possesses a high discrimination power. Moreover, it turned out that the Wiener Index, Randić Index, and the Cyclomatic Number are not appropriate for our used data set.

In our future work, we will study the relationships between the degeneracy of such a topological measure and the behavior of the underlying processes. In addition, we plan large scale calculations to apply such complexity measures to large sets of process graphs. Another important aspect we want to investigate is to measure the complexity of subgraphs meaningfully. Then, the resulting method could be used to detect subgraphs within large process networks representing relevant sub-processes. Again, the hypothesis that needs to be analyzed more deeply is whether similar complexities correspond to processes which behave similarly.

References

[1] Jorge Cardoso. Evaluating the Process Control-Flow Complexity Measure. In *IEEE Intl. Conference on Web Services (ICWS 2005)*, pages 803–804. IEEE Press, 2005.

[2] Jorge Cardoso. Business Process Quality Metrics: Log-Based Complexity of Workflow Patterns. In *On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS*, volume 4803 of *LNCS*, pages 427–434. Springer, 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 Intl. Conference on Business Information Systems (BIS 2006)*, pages 1–12. Springer, 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 *3rd Intl. Conference on Innovative Computing Information and Control (ICICIC'08)*, pages 246–246, 2008.

[11] E. V. Konstantinova. The Discrimination Ability of Some Topological and Information Distance Indices for Graphs of Unbranched Hexagonal Systems. *Journal of Chemical Information and Computer Sciences*, 36:54–57, 1996.

[12] Antti Latva-Koivisto. Finding a Complexity Measure for Business Process Models. Helsinki University of Technology, 2001.

[13] TJ McCabe. A Complexity Measure. *IEEE Transactions on Software Engineering*, pages 308–320, 1976.

[14] Jan Mendling. *Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness*, volume 6 of *Lecture Notes in Business Information Processing*. Springer, 2008.

[15] Milan Randić and Dejan Plavšić. On the Concept of Molecular Complexity. *Croat. Chem. Acta*, 75:107–116, 2002.

[16] V. A. Skorobogatov and A. A. Dobrynin. Metrical Analysis of Graphs. *Commun. Math. Comp. Chem.*, 23:105–155, 1988.

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

[18] 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.

[19] 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*