Abstract

This paper presents a quantitative approach to business process analysis using metrics from (social) network analysis. It is claimed that using this approach not only allows quantification of business process properties, but also provides insights on business process behavior that is not obvious from business process descriptions or models. The goal is to show that process characteristics can be explored and analyzed using network theory and methods of statistical analysis. The focus of this work is put on process type definition. To validate the approach business process model networks are evaluated against artificial networks with selected random graph distributions. The identified metrics are evaluated using a sample of real-life business processes.

1. Introduction

The methods of network analysis are gaining momentum in different scientific disciplines. This approach is especially applicable when relations among entities and not the entities themselves are in the focus of analysis. Research disciplines that rely on observation and thus deal with rather “soft” features of the studied subject often struggle defining the robustness of their results and are often considered subjective. Hence, due to a distinctive behavioral and organizational component, network analysis has already found the way into Information Systems Research (ISR) both as method and tool [1-4], although its use has been so far limited to derivation or evaluation of given structures.

Business process analysis is part of business process management and thus a frequently addressed topic in ISR. It is an important component of process reengineering and can lead to essential insights into the process structure and deployment. Organizational as well as IT-related topics such as integration of new application systems are often the trigger for such analysis. As a result there is no distinct and exhaustive method for conducting the process analysis. This fact can be partly referred to the multiple purposes and thus foci a process analysis can pursue. Process analysis is also a very domain specific, knowledge intensive process based on the expert knowledge that can be hard to formalize in full detail. The analysis is being conducted with numerous purposes concerning the process activity chain itself, but several aspects are still being missed or neglected, see e.g. business process logic [5]. Another aspect is the analysis of structural relations between process activities. This aspect is based on information exchange within the process itself.

The goal of this paper is to present an approach to business process analysis based on a business process model with a focus on the behavior of business process activities that is determined by information exchange among them. The focus of this analysis is put on quantitative definition of process types. To capture the interaction and behavioral structure of the process activities (social) network analysis is chosen as an approach and its methods for analysis are applied. The use of this method is rooted in the origins of business process modeling, as many of them originate in Petri Nets [6]. To apply the network-oriented approach, business process models are converted into networks, network metrics are calculated according to (social) network analysis followed by the derivation of process related metrics. To capture dynamic communication structures in the process dynamics [7, 8] a dynamic network model is chosen to capture communication activities of the process. The derived metrics for business processes use standard network metrics as foundation. Their interpretation and definition in business process context is based on statistical evaluation of the metrics in process networks.

This approach provides a possibility to determine business process analysis results and their comparability. Having explicit definitions can support decisions and provide clearer and robust measures for process optimization, reengineering and governance. Thus, the contributions of this paper are twofold: first of all the method of network analysis using a dynamic network model is applied to business processes. This approach leads to quantification of process characteristics concerning communicational and informational behavior by interpreting the information...
exchange within process activities. To verify this approach, random network theory was applied on business process networks. The algorithms for the random networks were further adapted to the dynamic network model and thus to the concept of linkevents, providing an addition to the body of knowledge of studies of random and real-life networks. The second contribution presented in this paper is the provision of a process analysis technique that allows an objective, rigorous process classification by providing quantitative criteria. The identified process types are: core, automatable, information intensive, distributed and flexible processes.

Research methods applied here are (social) network analysis to calculate and interpret the process metrics, random network theory used for verification of small-world properties of the network, as well as statistical analysis that was used to derive and evaluate the significance of the found metrics. These methods were applied in the domain of business process management. Finally, a set of independent real-life business processes was analyzed regarding the classification of their types to provide an evaluation of the approach. Visualization and calculation of the process network metrics is realized using parts of the Commetrix\textsuperscript{1} toolset.

The research question addressed here is whether business process analysis can be quantified using the network view on business processes. Resulting artifact is the quantitative analysis approach including metrics for process type definition. This work addresses explanation and prediction theory in ISR as described by [9], as application of the derived metrics allows to analyze business processes concerning their type definition. Therefore, accordant management measures for further process enhancement can be applied and their effects considered. Being rooted in ISR the research method principles of relevance and rigor as proclaimed by [10] are complied with as the research question is chosen from an active research area and presented research artifact is derived using a robust research framework including its evaluation. Thus, the paper is structured as follows: to provide robust results, the research method is presented in chapter two. Application results of random network theory on business process networks are presented and shortly discussed. Derived process metrics are described in chapter three; their verification using supplementary real-life business processes is presented in chapter four. Discussion and future work finish the paper.

2. Research Method

\textsuperscript{1} www.commetrix.org

As already stated above, the goal of this research is to provide quantitative metrics for business process analysis. Hereby the process model is considered as a set of nodes (process activities) connected by links (control and message flows). The logical dependencies expressed as control flow and data or message flows are defined here as informational behavior of a process. The overall process behavior is thus a combination of two factors: the amount of information exchanged and the network structure emerged by information exchange. The first factor is expressed using network analysis terms as nodal degree. The second factor is included in the emerging network topology of the process.

A set of 58 business process models from various sources like case studies conducted during industry projects, research documents such as dissertations and master thesis generated within industrial projects and books related to business process management and analysis were analyzed. Thus, processes from various domains such as manufacturing, government, service as well as administration were analyzed.

To derive the network metrics, these business process models had to be converted into networks. The approach to interpret business process models as graphs is facilitated by the fact that popular business process modeling notations such as EPCs (event-driven process chains) [11], flow charts and BPMN (business process modeling notation) [12] models are based on Petri Nets notation, which is rooted in graph theory. Both notations, EPC as well as BPMN, include the following basic elements: activities- describing the process steps, events and control flow- connecting the model elements. While BPMN considers multiple types of process events and is able to describe communication between actors using the concept of message flow, EPC describes information and data objects with extra symbols that are modeled within the process activity as in- and output elements.

Thus, EPC models were converted in BPMN diagrams. Those are taken as the basis for process networks construction. The process activities are interpreted as network nodes, the information or message flows among process activities are interpreted as linkevents between nodes. The underlying data model used in this analysis consists of nodes and linkevents enriched with various properties [13]. The main difference in comparison to other network data models is the aggregation of linkevents to links. A linkevent is a relation between one or more nodes. Each linkevent can have one or more recipients as well as no, one or more than one recipients. In addition to that, each linkevent has a timestamp determining the timing of the event. The notion of linkevents stems
from the analysis of dynamic network characteristics [14]. Thus, the network is not only analyzed considering its final structure but also the interdependencies that led to the final structure are considered. The co-alignment of linkevents can lead to “weighted” links representing the communication intensity between the connected nodes. The visualization and analysis toolset Commetrix is based on this event-based data model and the usage is therefore consistent with the introduced network process approach.

As first evaluation step random network theory was used to create artificial networks with some quantitative and structural similarities according to the analyzed process networks. Quantitative metrics for process definition were identified using statistical analysis techniques such as factor analysis, discriminant analysis and statistical tests on significance. The statistical analysis was performed using SPSS 9.0 and the random network creation was performed using purpose-built software tools. The discriminant analysis was based on a precedent manual process analysis that resulted in specified process groups. Principal component analysis was used to derive the discriminant functions. The emerging quantitative metrics were finally applied on new real-life processes. Thus, event-based network analysis was applied here as a method for business process analysis. The informative value of the derived metrics describes the process specific impact on the network structure and was verified using structurally similar random networks as baseline.

3.1 Converting Business Process Models into Networks

A set of 58 business processes serve as the basis for the analysis of process type metrics. The business process models were given as EPC and BPMN diagrams. To provide a common notation, EPC models were converted into BPMN 1.2 models. The BPMN models were then reviewed to comply with defined modeling standards based on [15]. The models were designed by one single modeler so that the modeling style and model design were controlled for. In the next step, the process models were converted into the abovementioned data format to visualize the networks and calculate the network metrics.

The following network analysis metrics were computed for each network: number of nodes/links/linkevents, diameter, density, average degree/betweenness/closeness centralities, average reach, average path length, average clustering coefficient, number of core group members, number of isolated nodes. Additionally process specific data such as activity names, lanes and pools were included in the analysis. During the analysis some of the metrics that are used in strictly social network analysis context were not further included, instead some additional metrics that are able to characterize the processes, were calculated.

Network metrics such as the centralities (see eq. 2 for normalized degree centrality, eq. 3 for normalized betweenness centrality, eq. 4 for normalized closeness centrality), density, path length (eq. 1), clustering coefficient, diameter as well as reach were calculated according to [16]. The variable n represents the number of nodes in the network and d(n₀, n_j) the geodesic path between the two nodes and d_i the degree of node i. The nodal degree provides information on how many other nodes are connected with the node in focus. The centralities were normalized to make them comparable among the analyzed networks due to their different characteristics. gₖₐ(n_j) in equation 3 is the set of paths that go through n_j; gₖₐ is the number of shortest paths between nodes n₀ and n_j.

\[ l = \frac{1}{n(n-1)} \sum_{i \neq j} d(n_i, n_j) \]  
\[ C'_D(n_i) = \frac{d_i}{N-1} \]  
\[ C'_B(n_i) = \frac{2 \sum_{j=1}^{n-1} \sum_{k=1}^{n-2} b_{jk}(n_i)}{(n^2-3n+2)}, \] with \[ b_{jk}(n_i) = \frac{g_{jk}(n_i)}{g_{jk}} \]  
\[ C'_C(n_i) = \frac{n-1}{\sum_{j=1}^{n-1} a(n_i, n_j)} \] (4)

The network metrics and their interpretation used in social network analysis were reviewed and their possible impact on business process analysis was evaluated. An additional metric, the entropy of the centralities (eq. 5) was calculated and used in the analysis according to [17]. The entropy of the centrality distributions in equation 5 provides a way to examine the information behavior in a process network.

\[ H(G, P) = \sum_{i=1}^{n} p_i \log_2(p_i) \] (5)

Entropy is a measurement that can be used to explain probability distributions and describes the state of order in a system [17]. H(G, P) is the entropy of the graph G with a probability distribution P of the node set. To compute the entropy of centralities, the histogram of the nodal centrality values was first computed and converted into the probability distribution. Then, the entropy for each centrality in each process network was calculated.

3.2 Analyzing Real-World Properties of Process Networks

Using business process models as basis for process analysis poses the question of whether these are valid
representations of the real-world processes. Thus, random graphs are created to explore whether business process networks derived from process models are real-world networks as defined by [19]. The study of random graphs and thus the random network theory is generally applied to prove the deterministic properties of graphs. This analysis method is called probabilistic, as deterministic statements are provided using probabilistic arguments. The random graph theory has the goal to model a real-world network and to determine at what probability a particular property of a graph will most likely arise. This research field was initiated by Erdős and Renyi [17] in 1960ies after they discovered that probabilistic methods are often useful in solving problems in graph theory [18].

For each of the 58 process models, hundred artificial but quantitatively similar networks were generated to provide robust results. Process network creation was achieved using three different algorithms rooting in the random-network theory. For each of the random networks, the before mentioned network metrics were computed and analyzed regarding statistical significant difference compared to the original networks. This step was included to establish the baseline for the results of the metric calculation.

The purpose of this analysis was to explore whether business processes show the same properties, i.e. network metrics, as other real-world networks. According to [19-21] real-world networks are neither completely ordered nor completely random. They often show a larger path length and a larger clustering coefficient than random networks. Nevertheless, random networks allow simulation of real-world networks, so that a baseline for studying their characteristics is given.

The random networks were created using three different approaches to share certain structural similarities with the original networks. The here developed random network software created artificial networks based on selected input parameters from the original data or with certain graph distributions. Nevertheless, the applied algorithms needed to be adjusted to fit the linkevent-based data model. As null baseline for random graph creation the Bernoulli random graph distribution based on random network model from Erdős and Renyi [17] was used. The random networks were created based on a given number of nodes and linkevents from the original process network. The algorithm selects sender and recipient for each linkevent at random until the number of linkevents matches the original network. The only constraint is the avoidance of self links (sender must not be identical to the recipient).

The second baseline creates a random graph distribution with a fixed node out-degree: U|{X i+},|X +j} [22]. The algorithm takes pairs of nodes and their respective selection of linkevents and re-wires the network links [19].

The third baseline creates conditional random graph distribution according to the given dyad census in the original network. This algorithms is also known as the U|MAN algorithm [23]. The dyad census represents the number of mutual asynchronous or not existing links between pairs of nodes. The random network does have a different topology but the dyad census is the same as in the original network.

For each analyzed network, 3 times 100 random networks were created according to the mentioned algorithms. The random networks were analyzed according to their mean and standard deviation values of the network metrics as described by [19].

Table 1 provides an overview of the metrics for real-life business process networks and the random networks generated using the modified algorithms for random network generation by [22]. The table also uses results from [17] to provide an overview of the number of nodes and links of other real-life networks. For business process networks these are average numbers, as 58 processes were considered. Thus, the path length (PL) and clustering coefficient (CluCo) results are averages for analyzed business process networks. Here two examples of real-world network metrics are shown, [18] and [17] provide a larger overview on the analyzed real-work network characteristics. The results of the analysis allow the same conclusion as described in [18]: process networks show the same characteristics as other real-world networks studied by e.g. [20]. They have a bigger path length as well as a bigger clustering coefficient comparing to the random networks (PL_{random} and CluCo_{random} respectively). Furthermore, real-life and random networks were analyzed concerning the differences in their metrics. A t-test was used to identify significant differences.

<table>
<thead>
<tr>
<th>Network</th>
<th>Nodes</th>
<th>links</th>
<th>PL</th>
<th>PL_{random}</th>
<th>CluCo</th>
<th>CluCo_{random}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business</td>
<td>33,74</td>
<td>29,9</td>
<td>4</td>
<td>3,35</td>
<td>3,25</td>
<td>5,96</td>
</tr>
<tr>
<td>process</td>
<td>4</td>
<td></td>
<td></td>
<td>0,07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>4941</td>
<td>2,67</td>
<td>18,7</td>
<td>12,4</td>
<td>0,08</td>
<td>0,005</td>
</tr>
<tr>
<td>grid[18]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Movie</td>
<td>225226</td>
<td>61</td>
<td>3,65</td>
<td>2,99</td>
<td>0,79</td>
<td>2,7*10^{-7}</td>
</tr>
<tr>
<td>actor[18]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Comparing random and real-life networks
The random network analysis resulted in the conclusion that process characteristics can neither be considered ordered nor completely random. Significant differences of some of process metrics were detected in some of the processes. Nevertheless, no systematic deviation could be derived, supporting the findings of [18, 19] and highlighting the small-world characteristics of business process networks. Based on this conclusion, network analysis techniques can be applied for business process analysis as described in the following section.

4. Derived Process Metrics

The analysis of random networks indicates that standard process behavior cannot be explained using combination of a single or only a few metrics used in network analysis. An explorative factor analysis was conducted using the principal component analysis to identify possible metric sets that can be seen as process characteristic and thus reduce the amount of potential variables. The factor analysis was conducted using all of the network analysis metrics described above. Four main factors that characterized process behavior and are responsible for 85.17% of the variance were identified:
- factor1: structural properties
- factor2: communication behavior
- factor3: sub-grouping
- factor4: connectivity

The behavioral properties of business processes depend on their structure, connectivity, the communication patterns and intensiveness as well as group-building within the process. The sub-groups can be based on functional, organizational or communicational characteristics. For additional insights on sub-group identification clustering analysis can be used, see e.g. [20].

To support the explorative analysis using process data, a semantic analysis of the process models was conducted. This analysis aimed to identify process characteristics that can be implied for the process model but that are rather difficult to quantify or to analyze in an objective way. A first classification was done using manual process model analysis resulting in process groups or classes based on pre-defined criteria (see section 3.1). First of all, common process types existing in ISR literature were reviewed. Their characteristics were operationalized to derive them from the process model. According to the values of these characteristics, processes were grouped into classes. These classes or types were then analyzed using statistic methods such as discriminant analysis to provide a proof of the separability of the derived groups. Subsequently, a test on significant differences between the groups was taken. The definition of an interval characterizing a process type has been made possible.

4.1 Definition of Process Types

Previous to the process type definitions, a general process definition was provided. The classic definition of a business process is taken as a baseline for any process type [21]. The authors in [21] define the business process as “a collection of activities that takes one or more kinds of input and creates an output that is of value to the customer.” The considered process types in this analysis were: core processes, automatable processes, information intensive processes, distributed processes and flexible processes. The core processes are often defined as processes that contribute directly to the value generation of a business or an enterprise. Distributed processes are defined here as processes which sub-processes are executed by different actors that are situated in different geographical locations. Information intensive processes are defined here as processes that contain process activities with information as the major input resource for process output generation. Flexibility of a business process is an important area of research. [22] defined process flexibility referring to manufacturing processes. They considered multiple aspects of flexibility, although the operational flexibility has already found access into ISR [23]. It is defined as: “Operation flexibility of a part refers to its ability to be produced in different ways”[24]. This definition refers to a property of the part, i.e. the result of a sub-process and thus is adapted here and operationalized for the process classification. The different ways to produce the process result are defined here as branches of the process flow that are modeled using (X)OR- splits and joints. Automatable processes are defined as processes with high repeatability, and low level of variances in the control flow. Thus, these were identified using the network visualization of the process (see fig.1).

Following the process definition, process type characteristics that can be derived from the process model or description were used for further manual analysis: core processes were defined as processes including customer presence and distributed processes were identified as processes with more than one actor. To define flexible and information intensive processes the procedure was twofold: first for the definition of flexible processes the number of (X)OR-splits in the model, for information intensive processes the number of message and data flows were counted. The results were normalized using the number of nodes in the process to make them comparable. These were ordered.
and the arithmetic mean was calculated. Processes with resulting value for the number of splits or message flow respectively higher than the calculated mean were defined as flexible or information intensive processes respectively. Automatable processes were identified as processes with a linear process graph and with a low number of deviations in the control flow (see fig. 1 for comparison).

![Figure 1. Non-linear process network (up); linear process network (down)](image)

After the process types were identified, processes had to be categorized into two groups respectively: the ones complying with the defined characteristic, e.g. information intensity, being group1 and the others that are lacking these characteristics, e.g. non-information intensive, being classified as group0 (see the description above on information intensive processes). This first manual examination was conducted as an explorative study. Thus, the considerable temporal effort to complete this classification was taken into account.

### 4.2 Quantifying Business Process Types

#### Definitions

The manually grouped processes were analyzed using discriminant analysis. The discriminant function for core processes has been derived referring to the canonical discriminant function coefficients as:

\[
D_v = 8.894 - 6.44LS + 0.144conn + 0.034reach - 0.08PL + 0.016CluCo
\]

(6)

Using that function, process networks can be classified into groups of core processes \(D_v \text{ values } < 0\) and non-core processes \(D_v \text{ values }> 0\).

The discriminant function for defining automatable processes \(D_a\) based on uncorrelated network metrics was derived as:

\[
D_a = 0.514 - 5.923LS + 0.146conn + 0.04reach - 0.08PL + 0.006CluCo
\]

(7)

Processes with \(D_a > 0\) are considered automatable and processes with \(D_a < 0\) value are considered not automatable.

The analysis of information intensive processes resulted in the discriminant function \(D_t\):

\[
D_t = -10.421 + 4.473LS - 0.106conn + 0.065reach + 0.204PL + 0.038CluCo
\]

(8)

Significance value of Wilks' Lambda is here 0.046 at the level of significance \(\alpha = 5\%\). \(D_t\) values < 0 indicate small information intensiveness, while \(D_t\) values > 0 indicate information intensive processes. As the next step, the function for distributed processes was computed. The level of separability of the discriminant function can be measured as the value of significance of Wilks' lambda. \(D_c\) has the level of separability of 0.018:

\[
D_c = 21.016 + 18.839LS + 0.022conn + 0.028reach - 0.338PL + 0.050CluCo
\]

(9)

\(D_c\) values < 0 indicate distributed processes, while \(D_c\) values > 0 indicate non-distributed processes.

The discriminant coefficients exhibit different impact on the discriminant functions and thus represent different process characteristics. Link strength shows a negative separability effect for core and automatable processes; avg. connectivity has a negative effect for information intensive processes; avg. reach shows negative effects for automatable and core processes; avg. path length is positive in information intensive processes. The characteristics of the discriminant coefficients allow additional insights into the process types, e.g. information intensive processes can be expected to have a longer path length between the process activities. A high connectivity would result in less changeable processes that could not be considered as characteristically information intensive. The discriminant functions for process types reflect whether the processes comply with its purpose, i.e. the intended process type and goal.
In contrast to the process types described above, flexible processes were identified using solely the entropy metrics of the centralities. This approach was chosen, as the entropy of centralities reflects the information transportation within the process. Conclusions concerning the flexibility are based on the flexibility definition that implies that different ways lead to the process goal. For each centrality metric an entropy function was calculated and the processes were analyzed using discriminant analysis based on these metrics as input parameters. The resulting functions, i.e. entropy values for the three centralities, provided similar conclusions, i.e. processes with values < 0 of each of the discriminant function can be classified as flexible.

In addition to the analysis of the impact of discriminant coefficients on the separability of process types, differences of the discriminant coefficients between the types were examined concerning their significance using a t-test. Core processes resulted to have smaller avg. connectivity as non-core processes; automatable processes have smaller link strength than non-automatable processes. This fact might be due to the smaller level of communication embedded in such processes. Furthermore, automatable processes have a bigger average connectivity than non-automatable ones, which can indicate a higher level of robustness, i.e. a lesser reaction to changes. Distributed processes show greater values of average path length, implying less efficient information transportation; information intensive processes exhibit higher level of average connectivity which may imply their characteristics as being more sensitive to structural changes, as structure is important for content distribution and transmission in these processes.

The discriminant functions presented above together with the thresholds defined for analyzed process groups are used for analysis and categorization of new business process, i.e. of business processes that were not included in the original explorative set.

5. Application of the Network-oriented Business Process Analysis

The defined functions and criteria were applied on four additional real-world business processes derived to provide an additional verification of the found metrics. The process models were converted into BPMN-models and then transformed into suitable format for network analysis. No prior classic business process analysis, such as defining their types or goals, was conducted on these processes.

Using the defined discriminant functions, the four processes were analyzed concerning their type, i.e. being core, information intensive, flexible or automatable process. Their structural characteristics are presented in table 2.

<table>
<thead>
<tr>
<th>process</th>
<th>nodes</th>
<th>links</th>
<th>linkevents</th>
<th>diameter</th>
<th>avg. path length</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>18</td>
<td>17</td>
<td>13</td>
<td>9</td>
<td>4,37</td>
</tr>
<tr>
<td>P2</td>
<td>25</td>
<td>27</td>
<td>21</td>
<td>10</td>
<td>4,27</td>
</tr>
<tr>
<td>P3</td>
<td>20</td>
<td>33</td>
<td>19</td>
<td>6</td>
<td>2,9</td>
</tr>
<tr>
<td>P4</td>
<td>27</td>
<td>35</td>
<td>23</td>
<td>8</td>
<td>3,53</td>
</tr>
</tbody>
</table>

Figure 2 shows P1 as a BPMN- and as a network model. The different shades of nodes in the network model indicate different pools and lanes the activity is executed in. Node labels represent the activity names equal to those described in the BPMN-model. The size of the node indicates the number of linkevents sent.

Figure 2. P1 as BPMN model [24] (upper figure) and process network

P1 is a process that is situated in a car rental domain and it has been already implemented into a workflow environment. The process part in focus is car...
return, where the customer is not involved but the car is checked for possible damages or pollution. P2 is a manufacturing process called “pre-cut of the material” from the textile industry. The process is modeled but is executed mainly manually due to the (experienced) high level of coordination effort and human labor involved. The goal of the analysis here is to explore, whether the process is potentially automatable using the same workflow patterns. Also, it is interesting to know, as it is hardly noticeable from the model, whether the process indeed includes a high level of communication. P3 is a marketing process, that is, a product launch process that is performed by the product manager and the marketing together. Hereby the marketing expert is a marketing agency, that is assigned to define marketing measures and a marketing plan needed to support the product implementation by the product management. P4 is a process from the insurance domain. It describes the collection of information by the customer and the initiation of the contract with the insurance agency. Results of the process analysis including the resulting values of discriminant functions that indicate the process types are summarized in table 3.

The analysis shows that processes P2-P4 are identified as core processes, though this feature was not clearly visible from the process model and its identification would require the knowledge of process context or a process description. P1 is not identified as a core process, as it is a sub-process that is accomplished by the car rental company without providing additional economic value or value to the customer. Nevertheless, P1 was identified as automatable and not information intensive. All of the processes were identified as distributed and flexible referring to the operational flexibility as defined by [22], implying that their structure potentially allows changes in the control flow.

### Table 3. Summary of the evaluation results

<table>
<thead>
<tr>
<th>Process type/name</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
</tr>
</thead>
<tbody>
<tr>
<td>automatable</td>
<td>x (1,49)</td>
<td>0 (-1,08)</td>
<td>0 (-1,18)</td>
<td>0 (-1,71)</td>
</tr>
<tr>
<td>information</td>
<td>0 (-1,3)</td>
<td>x (0,33)</td>
<td>x (0,85)</td>
<td>x (0,56)</td>
</tr>
<tr>
<td>core</td>
<td>0 (-1,54)</td>
<td>x (0,46)</td>
<td>x (0,85)</td>
<td>x (0,88)</td>
</tr>
<tr>
<td>distributed</td>
<td>x (-0,99)</td>
<td>x (-1,02)</td>
<td>x (-0,72)</td>
<td>x (-0,80)</td>
</tr>
<tr>
<td>flexible</td>
<td>x (-0,99)</td>
<td>x (-1,41)</td>
<td>x (-1,22)</td>
<td>x (-1,47)</td>
</tr>
</tbody>
</table>

These findings can be used by process owner or manager to define measures for process improvement or optimization. Distributed as well as information intensive processes are characterized by increased path length between the activity nodes that implies a decreased efficiency of the information transportation. The identification of a process as information intensive requires information management related measures by process owner or analyst. Thus, the consequences of the analysis results can include process support using an information management system and changes in process structure to enable efficient information exchange.

### 6. Related Work

The overview on related work includes the areas of quantitative or data-oriented business process analysis, including analysis of business process models, as well as other research areas of ISR such as knowledge management, enterprise architecture design and governance, as well as direct implementation of (social) network analysis in the abovementioned areas.

Network analysis is gaining momentum as a research method in Information Systems Research (ISR), as it provides a strongly developed theoretical basis rooting in mathematical graph analysis. Thus, the approach potentially provides a quantification approach for business related behavior. Originating in organizational studies the method of network analysis is used in knowledge management and related disciplines such as the analysis of virtual communities [8] and library studies [25]. Network analysis approach is currently increasingly applied for analysis of social networks or internet communities, e. g. by [26].

Quantitative approaches, often combined with development of process related metrics, in business process management, especially of business process models is gaining attention in the business process management research community. [27] explored complexity and coupling of business process models considering the aspects of understandability and modification; [3, 28] explored the topic of quantitative process model characterization inspired by social network analysis. Constitutive of this work [29] explore the adaptation of existing quality metrics from object orientation for BPMN and [30] for error density in process models; [31] adapt several coupling metrics from software engineering to business process domain. Data-based business process analysis is an established and increasingly explored terrain by [32]. The focus of the research lies on the verification of process workflows and development of quantitative metrics for process models being the base for the workflow.

Network analysis and visualization has already been applied to the area of enterprise application integration and governance. Integration and structuring of IT-architecture are topics that have been analyzed by [1, 2] while the implementation of an IT-architecture governance approach was presented and realized in [33]. [34] adapts a business process representation as a graph. He uses this view to define,
model and evaluate business process interactions using
the π-calculus. He defines a business process graph as
a four-tuple consisting of nodes, that represent process
activities, directed edges, which define dependencies
between activities, node types and attributes. Later he
also considers message flow between the processes and
defines it as an interaction edge of a process graph.

As already stated above, [35] suggested modeling
random networks to identify specific characteristics of
social networks. The random network theory is often
used and successfully applied in complex networks
research. Increasingly, this approach is also
implemented in the area of social network analysis
[36].

The abovementioned approaches related to business
process management use the network analysis
approach to verify or evaluate business process
models. In the presented paper, the method is applied
to identify process specific characteristics that
influence or are represented by the structural properties
of the network.

7. Discussion and Future Work

In this paper, a network-oriented approach for
process analysis is presented. The analysis is focused
on quantitative definition of process types. This
approach has been chosen, as network analysis roots in
strong mathematical theory and allows quantitative
analysis of the relations between system elements.
Business processes are seen as networks, the process
activities being interpreted as nodes and the control
and information flows being interpreted as links. The
behavior and thus the core properties of the process are
assumed to be externalized in the information
exchange between the activities. The concept of
linkevents was adapted to be able to capture the
communication structure and behavior of the business
processes. To provide a baseline, random network
theory was applied. We have compared the metrics of
random networks generated in accordance to certain
structural properties of the real-world business
processes. The results indicate that business process
networks comply with the definition of real-world or
small-world networks. Thus, the (social) network
analysis approach can be applied for their exploration.

Business process models were used as
representations of real-life business processes. To
identify process types, statistical analysis methods like
factor and discriminant analysis were applied. This
analysis was supported by the definition of process
types by operationalizing the process type specific
criteria, so that their identification in the process model
is robust and repeatable. The process groups were
analyzed using statistical methods of discriminant
analysis. Accordingly, discriminant functions were
derived, to also enable the future analysis of business
processes. Finally, sample processes were used to
define their type using the identified discriminant
functions. Evaluation results were consistent with the
expectations put on the processes and provided
additional insights into the process structure and
characteristics. As the first explorative study for
process type definition was conducted manually, it is
certainly possible to use the described process criteria
for future process type analysis. However, the
significant temporal effort required for the completion
of this task needs to be taken into account.

We believe that the quantification of process
features can provide a more reliable and rigorous
process management as well as offers a structured
orientation in business process analysis and realization,
e.g. decisions on whether and how IT-based business
process execution can be supported using the presented
methods.

Being a starting point for further quantitative
process analysis, several aspects occurred during this
research. The use of discriminant analysis for process
type definition results in a black-and-white approach to
process types. Thus, process models with
heterogeneous internal granularity are assigned to a
certain type regardless the fact that its sub-processes
can show different characteristics and thus would
require different managerial support. Therefore, the
analysis results need to be discussed with the accordant
process owner or manager.

Future work will include analysis of further process
types based on network analysis to provide an
extensive pool for the approach evaluation but also
analysis of roles and types of process activities within
the process networks. Furthermore, experiments on the
perception of business process visualization as a
network using the cognitive fit theory are already being
performed by the authors. Deeper analysis on process
specific features as well as a broader analysis of the
differences between social and business process
networks are aimed at as further research questions.

10. References

[1] Aier, S. How Clustering Enterprise Architectures helps to
Design Service Oriented Architectures. in SCC ’06. IEEE
International Conference on Services Computing. 2006:
IEEE.
Analysis in ICSC 2006 Workshop. 2007: Springer.
Understanding the occurrence of errors in process models
based on metrics. in OTM Confederated international


[12] OMG, Business Process Modeling Notation (BPMN)


[26] Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., Bhattacharjee, B. Measurement and analysis of online social networks. in 7th ACM SIGCOMM conference on Internet measurement. 2007, San Diego, California, USA: ACM.


