Association for Information Systems
AIS Electronic Library (AISeL)

ECIS 2013 Completed Research

ECIS 2013 Proceedings

7-1-2013

Using Social Network Analysis To Measure Information Management Performance Introduced By Business Process Optimization

Olga Levina
Berlin Institute of Technology, Berlin, Germany, olga.levina.1@tu-berlin.de

Annette Bobrik
Berlin Institute of Technology, Berlin, Germany, annette.bobrik.1@tu-berlin.de

USING SOCIAL NETWORK ANALYSIS TO MEASURE INFORMATION MANAGEMENT PERFORMANCE INTRODUCED BY BUSINESS PROCESS OPTIMIZATION

Levina, Olga, Berlin Institute of Technology, Straße des 17. Juni 135, 10587 Berlin, Germany, olga.levina.1@tu-berlin.de

Bobrik, Annette, Berlin Institute of Technology, Straße des 17. Juni 135, 10587 Berlin, Germany, annette.bobrik.1@tu-berlin.de

Abstract

This paper demonstrates how social network analysis (SNA) can be used in the context of knowledge management and process optimization including the accordant proceeding, metrics and their interpretation theory. These measures offer quantitative information on process and knowledge structure for process designers and managers who are faced with making change and accordant investment tradeoffs, especially as the process optimization task is being undertaken. Business processes that strongly rely on interaction and communication – so called knowledge-intensive business processes (KIBP) – have a high demand of interrelating process change aspects with knowledge management aspects. A methodology is proposed and demonstrated with the help of a case study showcasing the robustness and flexibility of the proposed SNA-related measures. It is demonstrated how content-based clustering for knowledge identification on process tasks can access the hidden knowledge layer to improve information management performance in course of process optimization activities.

Keywords: social network analysis, content-based clustering, knowledge management, business process optimization.

1 Introduction

Business processes, their analysis and management have been on the agenda of researchers and practitioners for over a decade now. Nevertheless, analysis of the changes induced by the managerial decisions aimed at the process optimization is often left to the post-implementation experience. In this paper a quantitative approach of (social) network analysis (SNA) is used to describe and thus to assess the induced changes to the process and information structure that appeared due to the process optimization activities before their implementation using process as-is and to-be models. Focusing on the information and process flow structure in the context of process optimization enables a holistic view on the process as a workflow and knowledge structure. Thus, the design and implementation of process improvement activities need to consider knowledge management aspects as well as process performance.

SNA has been already used to address process performance resulted from introduced IT-process support by Hassan (2009). Furthermore, research has been done to assess changes in the process (information) structure after process optimization by Levina (2012a). Thus, the goal of this paper is to combine actor performance metrics as interpreted by Hassan (2009) and process assessment metrics as presented by Levina (2012a). Furthermore, changes in expertise and role process structure as induced by process optimization activities are explored to derive a holistic set of quantitative process metrics for process designers and managers allowing a more objective picture on process performance and knowledge change. The knowledge aspects are elevated using content-based clustering for knowledge identification on the process network derived from the process model. This approach has been introduced by Bobrik (2013). It is applicable on any kind of data of employees' interaction and communication in knowledge-intensive business processes (Abecker et al., 2002, Heisig, 2002, Davenport et al., 1996). In addition to the network perspective of conventional SNA it allows accessing the underlying knowledge by analyzing the content layer of the data and combining the results with the network perspective. Therefore, the analysis of the process structure can be extended by the investigation of the process context to retrieve a detailed picture of the requirements of knowledge demand and supply within the process.

This paper shows how SNA can be used in the context of knowledge management and process optimization including the accordant proceeding, metrics and their interpretation. The audience for the paper includes business and IT managers as well as process designers and researchers interested in studying and designing the most effective optimized business processes. Method used here is network-based analysis of the process structure and content that is validated by research activities of Hassan (2009) and Levina (2012a) as well as a real-life case study to demonstrate the approach and its results. Thus, the paper is structured as follows: First the domains of business process optimization and social network analysis including content-based clustering of network structures as well as accordant research activities in these areas are introduced. Metrics used to assess process performance and knowledge structure are presented and applied in a case study in section four. Results of the case study as well as the outlook are discussed in the fifth section.

2 Related Work and Definitions in Business Process Optimization and SNA

In their work Vergidis et al. (2008) argue that a holistic approach toward business processes should capture a business process e.g., via business process modelling and provide the necessary means for bottleneck identification and performance analysis. They further claim that using (predefined) quantitative process performance measures is the appropriate systematic approach to business process optimization. Following this approach, business process optimization (BPO) aims at enhancing measurable process performance. Research by Zhou and Chen (2003) suggests that BPO should aim at reducing lead time and cost, improving quality of product, and enhancing the satisfaction of customer and personnel so that the competitive advantage of an organization can be retained. Furthermore, Reijers (2002) suggests that the goals of BPO are often the reduction of costs and flow time. Additionally, Hofacker and Vetchera (2001) underline that the concept of "optimality" of process designs is not trivial, and the quality of processes is defined by many, often conflicting criteria. This latter view on BPO is also supported in this paper. A process is a complex construct that embeds process logic, rules as well as actors including their know-how on the process execution and the accordant information structure. These aspects are often not captured in the time/cost based view on the process performance and thus will be in the centre of the research presented here.

A case study in the context of BPO has been taken here as an example to elevate and observe the information and knowledge based changes induced by the optimization activities using the quantitative metrics of SNA. Analysis of (social) networks focuses on the exploration of the links between system elements (actors), thus on their relations rather than on the elements themselves. A social network is a set of social entities such as people or organizations connected by a set of social relationships such as friendship or information exchange (Jamali and Abolhassami, 2006). Networks are represented by mathematical structures, the graphs. Thus, SNA aims at measuring relationships and information flows between social entities (Jamali and Abolhassami, 2006) using the graph theory. Networks are characterized by their quantitative characteristics such as e.g., diameter, density, average degree, path length, connectivity, clustering coefficient and reach (Wasserman and Faust, 2009). The average path length indicates the average number of steps along the shortest paths for all possible pairs of network nodes, while reach denotes the degree of any member of the network to which it can reach other network members. The average path length can be interpreted as a measure for the efficiency of the information transport in the network. The connectivity of a network indicates how many nodes need to be removed to separate the network in several groups (Wasserman and Faust, 2009). Clustering coefficient measures the interaction of nodes within an ego-network including transitive connections, and can therefore be useful for the identification of sub-grouping in the network. Density is defined as the ratio of links present in the network and the maximum number of possible links (Wasserman and Faust, 2009). Thus, it can also be used to refer to the stability of the network with respect to structural changes. Diameter of a network is the largest geodesic distance in the connected network (Wasserman and Faust, 2009). It can also be used as a metric for network size or complexity. Complexity is also measured here as the quotient between the number of links (L) and nodes (n) within the network as suggested by Cardoso (2005).

An important construct in the analysis of social structures is the social power or capital of an actor within a network. The measure used in this context is the centrality of a node. There are three commonly used approaches to the centrality of an actor in the network and therefore of the metric (Freeman, 1978): degree, closeness and betweenness centrality. Nodal degree or degree centrality measures the number of ties, i.e., connections, of an actor/node. Nodes with a high value of this metric are interpreted as well connected actors. Closeness centrality measures the path length from a node to other actors. Nodes with high values of the closeness centrality are interpreted as being involved in close exchange with other actors. A small closeness centrality value indicates an autonomous and independent node. The average closeness centrality of a network can provide insight on the collaboration and information distribution productivity within the network (Okamoto and Chen, 2008). Betweenness centrality indicates a node that is situated in between of other pair of nodes (Jamali and Abolhassami, 2006) and thus can be regarded as controlling towards information transportation and dissemination. The metrics described above measure static properties of a network. Dynamic network analysis provides a possibility to analyze the network structure in its temporal development. Thus, the network context is enriched with the notion of the linkevent allowing the weighting of given relationships. Each linkevent can have one sender as well as no, one or multiple recipient(s) (Bender-Demoll, 2006). A network measure that builds on the notion of the linkevent is the average link strength of a network. Link strength identifies the weight of a link in terms of multiple interactions performed on a given relation. Brokering activity is another metric from dynamic network analysis and defines important actors in the network according to their activities that resulted in shortening the network's paths (Trier and Bobrik, 2007). High brokering activity values indicate nodes that strongly influence the network structure by shortening the paths and thus providing more efficient information structures and are regarded here as core activities. In the following these definitions will be operationalized for further analysis.

3 Applying SNA for Business Process Analysis

In his research Hassan (2009) uses SNA to assess the changes induced by the managerial activities aiming at process enhancement by an increased IT-process support and thus to evaluate alternative process designs. A similar goal is pursuit by Levina (2012a) on the process structure and activity levels. In this paper these approaches are enriched by the exploration of the process content using a content-based clustering algorithm to derive knowledge and information structures of the process and the induced changes. This section briefly presents the network-based business process analysis (NBPA) and the content-based clustering approaches that are applied in the following case study.

3.1 Network- based business process analysis- a brief introduction

To address the question on how to measure process changes due to process optimization using a business process model, network-based business process analysis as presented in (Levina and Hillmann, 2012b) was chosen. This approach allows analysis of process structure and its characteristics using network metrics. Thus, business processes under analysis were transformed into accordant business process networks with business process activities as nodes and control as well as information flows as links. Network metrics used for the analysis are described above and were calculated using Commetrix® network analysis software and inserted into the discriminant functions and metrics in Table 1 to define process types and activity roles (see (Levina, 2012b) for details on process definitions and approach). Following process types are distinguished here and can be identified using discriminant functions and metrics from Table 1: core, automatable, distributed, information and decision intensive. Core processes are defined here as value-adding processes requiring direct customer interaction. A core process model includes an internal or external process customer. Automatable processes are defined here as repetitive, predictable (Georgakopoulos et al., 1995) processes with a low level of variance, i.e., decisions or exceptions, in the control flow. Information intensive processes are defined here as processes, which activities use information as a main resource, implying frequent information exchange within the process flow. A decision in this context can be identified as a choice between several process variants rather than a choice between two alternatives (Bromberg, 2007). Thus, decision intensive processes were identified as processes containing an increased number of operative decisions (Levina, 2011). Distributed processes were defined here as processes, which sub-processes are executed by different actors situated in different geographical locations. Discriminant functions in Table 1 include following network metrics as variables defining behavioural characteristics of a process: average path length (PL), average clustering coefficient (CluCo), average connectivity (conn) and reach. Business process activity roles were also analyzed here using the network-based business process analysis approach. These roles are: information sink and source, process sink and control activity. An information sink is a process activity that requires an increased amount of information for its execution. This role is defined using the number of received linkevents per node. Activities with an amount of received linkevents that is found in the third quartile of received linkevents is considered as an information sink. An activity that delivers information to other activities is considered as information source. These are activities from the third quartile according to number of sent linkevents. A process sink is an activity that is considered as the process goal, i.e., the activity that captures the control and information flows of the process. This is the activity with the highest number of linkevents received in the process. Core activities are defined using the brokering activity metric and its interpretation. Control activities are identified using the metric of betweenness centrality (Wasserman and Faust, 2009).

3.2 Content- based clustering of network structures- a brief introduction

Some business aspects cannot be designed and organized by predefined process models that – once established – never or rarely ever change. There only exist more or less detailed parameters how and when certain activities have to be performed. A collection of these activities can be subsumed as a knowledge-intensive business process which is characterized by the individuals that are involved, their knowledge, and their interactions that establish a network among them. They typically have a high degree of complexity, of communication, of exceptions from predefined business rules, of work autonomy and of information need as well as a low degree of structuredness and predictability as being hard to schedule (Davenport et al., 1996, Abecker et al., 2002, Heisig, 2002). Analyzing these processes can help to identify sources and demands of knowledge in process-oriented knowledge management (Trier and Bobrik, 2007). While the use of structural network metrics on business processes and the incorporation of the network perspective in process analysis is a popular field of research most approaches do not take the content layer of the data into account. Therefore, the method of content-based clustering for knowledge identification is introduced by Bobrik (2013) to access the content layer and to identify the requirements of knowledge demand and supply as part of an enhanced business process analysis. It combines elements from text mining to access the content, cluster analysis to establish groups of shared context and SNA to identify and evaluate their embeddedness in network structures. Cluster analysis is a method for classifying objects into meaningful sets with similar characteristics (Aldenderfer and Blashfield, 1984). In contrast to discriminant analysis which assigns the data to predefined categories cluster analysis tries to establish previously unknown groups that are inherent in the data (Jain and Dubes, 1988, Everitt et al., 2001). The basic elements for cluster analysis are any kind of data objects that can be characterized by a set of attributes or features. Using suitable clustering algorithms these objects can be assigned to groups with similar attributes. There is a large number of clustering approaches that can be categorizes into hierarchical and partitioning clustering algorithms which can be used with different measures of similarity (e.g., cosine similarity, Euclidean distance). Depending on the type of data these algorithms are more or less suitable to reveal the inherent clustering structure of the data. Popular clustering algorithms to detect subgroups in network structure is the hierarchical divisive edge-betweenness clustering algorithm by (Girvan and Newman, 2002). It is designed to identify areas of strong interaction with in the graph structure. In contrast to this approach that only exploits the network structure the method for content-based clustering for knowledge identification is based on the comparison of the content of interaction and communication among the network's participants. The participants in the communication and interaction process (actors) share content objects with each other. For example, a company's employees work together on project reports and presentations. On content level a collection of content objects can be extracted from the data. Working on or with these content objects can be interpreted on network level as linkevent and aggregated to edges between nodes (actors) establishing the network structure. The content objects can then be assigned to groups of similar context depending on similarity metrics using cluster analysis. An overview of suitable clustering algorithms and similarity measures for content-based clustering for knowledge identification is provided in (Bobrik, 2013). Based on the research guideline for content-based clustering on node level the approach can be structured into the following steps (Bobrik, 2013): (1) collect the data, (2) transform the data into a network representation with the content objects of communication and interaction being interpreted as linkevents connecting the affiliated actors, (3) extract keywords from the content objects using text mining techniques, (4) assign keyword collections to actors, (5) group together actors in content-based clusters based on the similarity of their keyword collections, (6) map the clustering results to the graph structure, (7) analyze the results on content level and network level.

4 Case study

The case study is situated in a leading European logistics company and was conducted as a project with the scope on the current project planning approach in the area of global and regional EDI-

(Enterprise Data Interface) design focusing on Quality Management. Thus, organizational and technical processes were elevated with the purpose of their homogenization and identification of possible potentials and weaknesses. The as- is process includes process participants such as: project manager (PM), customer, logical mapper[1] (LM), developer, coordinator, and quality assurance (QA). The project goal was to find a process structure to optimize the planning of the interface design towards the go-live phase. The project outcomes included a general as-is process model as well as a re-designed (to-be) process model in BPMN v1.2. Independent from the concrete project goal, a network- based analysis of the process structure and content of the as-is as well as of the to-be process was conducted. The goal was to identify and compare process performance metrics of the two process variants to provide quantitative statements on the results of process optimization and enhancement for the process manager and owner. The method of network-based analysis of the process structure and content is described in the following sections.

4.1 Method for network-based Process Performance Analysis

Here two aspects of the process are analyzed using network-based analysis: the process structure and process knowledge and information flow as captured by the BPMN model. Both aspects are quantified using metrics form the SNA domain that are derived using the (extended version) of the Commetrix®[2] network analysis tool as provided by the IKM research group and the prototype for content-based clustering by Bobrik (2013). The proceeding of the analysis can be summarized as follows and is applied after the process was elevated and documented in a BPMN v1.2 diagram that was transformed and visualized as a process network: Network-based business process analysis as described in Levina (2012a) interprets process diagrams as networks with activities as nodes and control as well as message flow as links. Then, network metrics to quantitatively define a process type and activity type are derived for the network (see Table 1 and (Levina and Hillmann, 2012b)). Based on the results, managerial decisions for the process optimization or assessment can be taken. In a second step structural clustering is employed to identify densely connected subgroups based on interaction patterns using the hierarchical divisive edge betweenness clustering algorithm by (Girvan and Newman, 2002). The optimal solution is detected by calculating overall modularity value that expresses the amount of non-randomness of the clustering solution. In practice, this value ranges from 0.3 to 0.7 for comparatively well structured data (Newman and Girvan, 2004). For the case study, the structural clusters relate to functional units and an optimal solution of nine clusters with a modularity value of 0.61 is obtained. Comparing process roles and phases with structural clusters and analyzing within and between cluster interrelations provides support for managerial decisions on technical and organizational process support (i.e., information systems/data exchange, coordinative activities, quality management) based on deviations between originally designed and actual functional units. In contrast, content-based clustering for knowledge identification on nodes as described in (Bobrik, 2013) identifies subgroups with similar context by assessing the content layer of the process tasks. Therefore, tasks labels and descriptions are interpreted as keywords to describe the information context of sender and recipient nodes. The resulting clusters relate to units of shared context (e.g., process activity, expertise, etc.). Selecting suitable cluster parameters strongly depends on the data at hand. For the case study, the similarity between two nodes is calculated by the cosine similarity (Salton and McGill, 1984) of their keyword vectors. The optimal number of clusters is identified using agglomerative hierarchical clustering with UPGMA linkage rule (Sokal and Michener, 1958) and the Calinski & Harabsz stopping rule (Calinski and Harabasz, 1974). This initial solution is refined using a partitioning clustering algorithm with K-means clustering (McQueen, 1967) and medoid

[1] The logical mapper relates the specific data of the customer to the data structure of the application without regard of technical aspects of the actual development process.

[2] http://www.commetrix.de/

representatives (Kaufman and Rousseeuw, 1990). The clustering is established on the content layer itself without regard to the network structure. Each node, i.e., process activity is assigned to a content cluster based on the similarity of the shared context with the other cluster members. Afterwards, the clustering solution is mapped on the graph structure. In contrast to the structural clustering approach this type of grouping is not based on interaction patterns. Each subgroup may be densely connected but can also consist of several unconnected components. Similarly, subgroups may be loosely or densely connected among each other. Comparing process roles and phases with content clusters and analyzing within and between cluster interrelations provides support for managerial decisions on technical and organizational process support (i.e., information systems/data exchange, coordinative activities, quality management) on deviations between originally designed functional units and actual functional units.

4.2 Analysis of the process structure metrics

Here the metrics for the process performance in terms of process goal definition, information structure as well as information distribution roles are described. The interpretation of network metrics as process performance and assessment indicators was introduced by Hassan (2009) and extended by Levina (2012a). Here the proof of concept as well as enrichment of the approach with further and more information and knowledge oriented metrics with further semantic interpretation are presented. Therefore, first, the structure of the as-is and the to-be processes are analyzed. In the next subsection the content layer of the process is considered. Both of the processes are identified as not automatable in their entire workflow as well as not distributed (see Table 1) indicating a dense interaction between the enterprise and the customer. Also, their indication as information and decision intensive supports the assumption that information exchange and know-how plays an important role in the process execution and thus has to be supported by the optimization or re-design measures. The as-is process network contains 49 nodes vs. 50 in the to-be process, 76 vs. 73 linkevents and 79 vs. 69 links with the diameter of 10 vs. 11 and density 6.46 vs. 5.55. Average degree centrality in the as-is process is 6.46 vs. 5.59 in the to-be process and avg. Betweeness centrality is 7.29 vs. 7.76. Avg. Path length as-is is 4.57 vs. 4.88 and network complaexity in the as-is process as the quotient of the number of links to nodes is 1.61 vs. 1.38.

Analysis of the network metrics of the two processes shows that the density of the to-be process decreased indicating that the group of actors involved in the process is less cohesive (Hassan, 2009), while the complexity of the process decreases as indicated by the rise of the diameter and complexity metric for the to-be process. So does the overall sum of linkevents and links within the process, indicating the overall decrease in communication within the process, while the degree of the involvement of the actors into the process communication increases (see average degree centrality) as well as the communication intensity for most of the actors but the QA, see Table 2. The average path length of a network provides an indicator of how easily an actor interacts with another actor, and thus can be used to assess the efficiency of the information transportation within the process. Increased average path length in the network can indicate the increased need for consultation between the actors and thus a slightly less efficient information transport within the process.

Process Type	Discriminant Function	Result and value as-is	Result and value to-be
Automatable	D_a= 0,514- 5,923LS + 0,146conn+ 0,04reach- 0,08PL+ 0,006CluCo	No (-1.86)	No (-2.0)
Core	D_c= 8,894-6,447LS+ 0,144conn- 0,034reach- 0,08PL- 0,016CluCo	Yes (-1.754)	Yes (-1.93)
Information intensive	D_i= -10,421+4,473LS-0,106conn+ 0,06reach+ 0,204PL+ 0,03CluCo	Yes (1.56)	Yes (1.76)

Decision intensive	D_e= -20,554+ 16,805LS- 0,081conn+ 0,045reach-0,104PL+ 0,062CluCo	Yes (1.83)	Yes (2.23)
Distributed	D_v= -21,016+ 18,839LS+ 0,022conn+ 0,028reach-0,338PL+ 0,050CluCo	No (0.58)	No (0.93)

Table 1. Discriminant functions and results

After the processes have been analyzed on the network level, more detailed information can be obtained from the analysis of the egocentric measures such as average betweenness (BC) and degree centralities (DC) of the actors involved, see Table 2. In terms of the average betweenness of the actors involved into the process, customer's as well as LM's BC increased implicating the rise of importance of these two actors in the new process design, while QA, PM and developer roles decreased implying more flexible process paths for these roles (Hassan, 2009). The degree centrality of the actors has also changed through the suggested process enhancement measures. The degree centrality measure suggests for the to-be process that all of the actors are now less involved in communication with other actors than during the course of the as-is process.

Analysis of the process networks on the activity level shows that the main activities in the to-be process have not substantially changed comparing to the as-is process. The core, control and communication activities stayed the same after the process enhancement measures; they also stayed assigned to the same actor. Nevertheless, the activity identified as the process sink is different in the to-be process. Furthermore, this activity is more specific to the actual, i.e., intended process goal, that is to develop an EDI for the customer.

	AsIs_BC	AsIs_DC	ToBe_BC	ToBe_DC	As-is_ avg. % LE	To-be avg. % LE
Developer	9.13	6.95	7.72	6.12	7.59	12.31
LM	3.06	6.25	8.08	6.12	17.72	19.86
PM	7.46	7.39	6.34	5.1	25.95	26.03
QA	11.23	6.25	5.42	5.44	10.13	6.16
Customer	11.14	4.95	13.19	4.85	12.03	13.01
Coordinator	5.77	6.73	6.03	6.12	26.58	22.60

Table 2. Actor metrics and results

Table 2 lists the values of the centrality metrics for each actor. The values were calculated using the network analysis software. In addition to these rather classic metrics, the linkevent metric is used to assess the communication intensity of an actor and provides an insight on how effective the relations, i.e., links, are used. The percentage in Table 2 is calculated as a relation between the linkevents exchanged by the actor to the overall sum of linkevents exchanged in the process. The metric shows that PM had and still has the highest communication intensity in the to-be process comparing to the other actors. The communication load is even slightly increased for the actor in the to-be process. The increase in communication is supported by the increased average path length in the network. According to the values of actors' BC this fact did not lead to a rise of importance of the role of the developer or project manager. In contrary, their roles even decreased although the communication load increased. The average network metrics of the two processes demonstrate that the to-be process has an overall decreased consultancy effort between the actors as indicated by DC, has raised the information potential of the process activities but has also increased the interaction effort between the process activities as measured by the average path length. Also, the decreased average closeness centrality of the to-be process indicates a decreased potential for effective information spreading between the process activities. The overall complexity of the process has been nevertheless reduced. These performance changes are represented in the metrics on the actor level in Table 2.

4.3 Results of the content analysis of the process

This section presents the results from the content analysis of the process in order to add new insights about the process context and its requirements. First, a structural clustering is employed to evaluate the deviation between originally designed process phases and actual functional units depending on the density of interaction among process tasks. Afterwards, the content-based clustering approach for knowledge identification by Bobrik (2013) is applied on node level to group those tasks together with share a common context – regardless of their strength of interaction or affiliation to process phases and roles.

Structural Cluster	Cluster Content Description	Corresponding Process Phases
SCluster 0	Administration	Testing
SCluster 1	Data Procurement	Process Initialisation
SCluster 2	Go-Live/Error Handling	Go-Live
SCluster 3	Development Preparation	Development
SCluster 4	Feasibility Check	Planning
SCluster 5	Administration/Project Initiation (Internal)	Process Initialisation
SCluster 6	Project Initiation (by Customer)	Process Initialisation
SCluster 7	Coordination/Development/Testing	Planning, Development, Testing
SCluster 8	Testing (Integration Test)	Testing

Table 3. Overview of structural clustering

To evaluate the quality of the originally designed functional units structural clustering is employed. The structural clusters relate to the actual functional units comprising several business tasks as established by the process flow. An optimal solution of nine clusters with a modularity value of 0.61 is obtained. In practice, the modularity value ranges from 0.3 to 0.7 for comparatively well structured data (Newman and Girvan, 2004). Therefore, the process tends to structure itself into several functional units. Comparing these functional units with the originally designed process phases reveals that there is a strong match between the originally designed process phases and the clustering solution based on interaction and interrelation of tasks and events. Table 3 provides an overview of the clustering solution (see also Figure 1a) by providing a general content description of a process activity groups, i.e. cluster, and the correspondent phases. The process flow is expressed by the following order of the structural clusters: 6, 5, 1, 4, 3, 7, 8, 2. Obviously, the structure of the original process as illustrated in Figure 1 is still present when it is reduced to the units of strong interaction. The main functional loop is expressed by clusters 1, 4, 3, 8 with more peripheral clusters for project initialisation (5, 6), development preparation (3) and testing and go-live (0, 2). In general, interaction within these process phases is high, whereas interaction between these process phases is low. As a result, the process is already well-structured into functional units. As an exception, the development phase (cluster 7) is strongly connected with planning (i.e. resource allocation) and testing. As it comprises three different but interrelated functional units this indicates a potential to reduce coordination effort or to introduce a policy (organizational redesign).

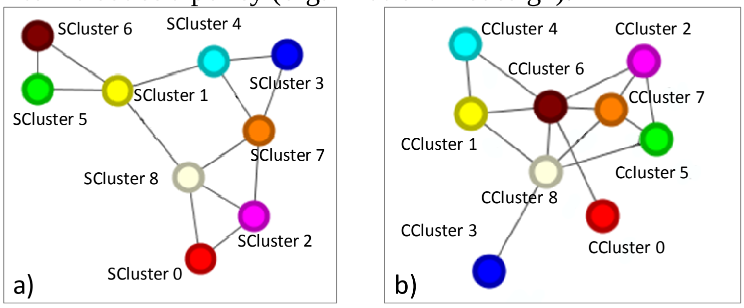

Figure 1. Graph representation of clustering solutions from a) structural clustering (“scluster”) and b) content-based clustering (“ccluster”). Each node relates to a single cluster.

These results do not take the hidden information layer of the business tasks into account. Using content-based clustering for knowledge identification task labels and descriptions can be used to identify subgroups with similar context. As a result, nine clusters with an overall content dissimilarity of 2.31 are obtained. Table 4 provides an overview of the clustering solution (see also Figure 1b). With the exception of cluster 0 and 3 all clusters are strongly interrelated. Therefore, the actual process flow cannot be as easily identified as with the strucutral clustering solution. Most clusters comprise tasks from different process phases. When inspecting the business tasks assigned to each cluster and the interrelation between the clusters new insights can be gained on process context, customer orientation and QA that have been missed by the previous analysis of the process structure. The shared context of the business activities integrates tasks from different process phases and roles. Development activities are assigned to different clusters depending on their scope (e.g., development preparation, implementation) together with surrounding and supporting tasks. This indicates, that a strict orientation on process phases and process roles may miss the specific interoperational requirements of these tasks. There is one cluster that comprises the main functional tasks as well as the customer (cluster 8). These tasks have the same context and are well-supported by QA and coordination (data transfer). The strong contextual interrelation of these tasks indicates a potential for technical and organizational support.

Content Cluster	Description	Corresponding Process Activities
CCluster 0	Process Initiation (Internal)	-
CCluster 1	Customer Feedback	Process Initialisation
CCluster 2	Feasibility Check	Planning
CCluster 3	Error Handling (after Go-Live)	Go-Live
CCluster 4	Process Initiation (Customer)	Process Initialisation
CCluster 5	Coordination (Resource Planning)	Planning
CCluster 6	Process Initiation (Internal) & Development Preparation	Process Initialisation, Development
CCluster 7	Testing (Internal Integration Test) & Problem Solving	Planning, Development, Testing
CCluster 8	Data Procurement, Development, Testing, Go-Live	Process Initialisation, Development, Testing, Go-Live

Table 4. Overview of content-based clustering results

However, customer-orientation takes only place in pre-/post-development phases (i.e., provide data and integration test). In contrast to the functional units indicated by the process analysis and the structural clustering, the context of the main implementation process is already customer-oriented. Here, further contact to the customer should be considered as inter-process tasks (e.g., feedback, revision, etc.). Early and late process activities like feasibility check (cluster 2), coordination for resource allocation (cluster 5) and error handling (cluster 3) are not supported by QA measures but form separated functional units. Thus, QA measures are only involved in the main functional tasks.

5 Summary of the Findings and Outlook

On the network level, the comparison of the as-is and to-be process characteristics shows that the process network in the to-be process is less dense, suggesting that the activities are now less connected by the logical or communication flows. This reduction in the coupling of the process activities also resulted in a less complex network as indicated by the increased diameter and decreased complexity metric and might be regarded as a higher flexibility of the process structure. On the level of node analysis the results show that the suggested optimization measures in the to-be process did not substantially change the activity roles. Nevertheless, the process optimization resulted in a process sink of the to-be process that is aligned with the process design goal. The control and information flows of the process are now in accordance with the purpose of the process. However, the decrease in average path length and closeness centrality values in the to-be process points to a less effective and less efficient information exchange structure within the process. Thus, the process might need a stronger support via e.g. information management system that could be implemented to support information and knowledge exchange between the actors. This is also an implication that can be drawn from the identification of the process type as being information and decision intensive (Levina, 2012a). On the actor and thus knowledge exchange level in the process some insights on the role structure in the to-be process could be won. The to-be process design reduced the consulting effort between the process actors as shown by the decreased degree centrality metric but it did not fully reduce the communication intensity as measured by the number of linkevents in the process. Process redesign also assigned a bigger weight to the customer and LM within the process, making them more powerful in terms of information distribution within the process. The process tends to structure itself into several functional units. Comparing these functional units with the originally designed process phases reveals that there is a strong match between the originally designed process phases and the clustering solution based on interaction and interrelation of tasks and events. Thus, the design of the process phases and its technical and organizational support seems to meet the demand of interrelating process tasks. However, when inspecting the shared context of process tasks it becomes obvious that the functional perspective on the graph structure misses the requirements of information and knowledge demand and supply between business tasks with shared context. Although the main functional tasks have the same context and are well-supported by quality QA and coordination, QA takes only place in the actual development phase. Furthermore, customer-orientation is only well-integrated in pre- or post-development phases. Future work in network-based business process analysis will include further case studies to provide robust interpretations and values for the presented metrics and method as well as workshops with process owners, managers and workers to collect and apply the insights they gained from the quantitative analysis results.

References

Abecker, A., K. Hinkelmann, H. Maus and H.-J. Müller (2002) Integrationspotenziale für Geschäftsprozesse und Wissensmanagement (in German). In Abecker, A., Hinkelmann, K., Maus, H. & Müller, H.-J. (Eds.) Geschäftsprozessorientiertes Wissensmanagement. Berlin Heidelberg, Springer.

Aldenderfer, M. S. and R.K. Blashfield (1984) Cluster Analysis, Newbury Park, Sage Publications.

Bender-Demoll, S., Mcfarland, D. (2006) The art and science of dynamic network visualization. Journal of Social Structure 7, 1-38.

Bobrik, A. (2013) Content-based Clustering for Knowledge Identification - A New Method for Knowledge Identification based on Text Mining and Cluster Analysis. (to be published). Department of Systems Analysis and IT. Berlin, Berlin Institute of Technology.

Bromberg, D. (2007) BPM for Knowledge Workers Inside Decision Intensive Processes (DIPs): Knowledge, Practice, Context, and Characteristics. BPTrends.

Calinski, R. B. and J. Harabasz (1974) A Dendrite Method or Cluster Analysis. Communications in Statistics - Theory and Methods, 3, 1-27.

Cardoso, J. (2005) How to measure the control-flow complexity of web processes and workflows. In Fischer, L. (Ed.) Workflow Handbook. Future Strategies Inc.

Davenport, T., S. Jarvenpaa, and M. Beers (1996) Improving Knowledge Work Processes. Sloan Management Review, 37, 53-65.

Everitt, B. S., S. Landau and M. Leese (2001) Cluster Analysis, New York, Oxford University Press.

Freeman, L. C. (1978) Centrality in social networks conceptual clarification. Social Networks, 1, 215-239.

Georgakopoulos, D., M. Hornick and A. Sheth (1995) An overview of workflow management: From process modeling to workflow automation infrastructure. Distributed and Parallel Databases, 3, 119-153.

Girvan, M. & M.E.J. Newman (2002) Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America, 99, 7821 -7826.

Hassan, N. R. (2009) Using Social Network Analysis to Measure IT-Enabled Business Process Performance. Information Systems Management, 26, 61-76.

Heisig, P. (2002) GPO-WM - Methoden undWerkzeuge zum geschäftsprozessorientierten Wissensmanagement. (in German) In Abecker, A., Hinkelmann, K., Maus, H. & Müller, H.-J. (Eds.) Geschäftsprozessorientiertes Wissensmanagement. Berlin Heidelberg, Springer.

Hofacker, I., Vetschera, R. (2001) Algorithmical approaches to business process design. Computers and Operations Research 28, 1253–1275.

Jain, A. K. and R.C. Dubes (1988) Algorithms for Clustering Data, Endelwood Cliffs, New Jersey, Pretence Hall.

Jamali, M. and H. Abolhassani (2006) Different Aspects of Social Network Analysis. 2006 IEEWICIACM International Conference on Web Intelligence. IEEE.

Kaufman, L. and P.J. Rousseeuw (1990) Finding Groups in Data - An Introduction to Cluster Analysis, Hoboken, New Jersey, Wiley-Interscience.

Levina, O. (2011) Design of Decision Service Using Cause-and-Effect Business Process Analysis. Computer Science- Scientific Journal of Riga Technical University, 46, 19-27.

Levina, O. (2012a) Using Network-based Business Process Analysis as a Tool for Delta Analysis of Process Models. . BUSTECH 2012.

Levina, O. and R. Hillmann (2012b) Network-based Business Process Analysis 45th Hawaii International Conference on System Sciences.

McQueen, J. B. (1967) Some Methods of Classification and Analysis of Multivariate Observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 281-297.

Newman, M. E. J. and M. Girvan (2004) Finding and Evaluating Community Structure in Networks. Physical Review E, 69, 026113.

Okamoto, K., W. Chen and X.-Y. Li (2008) Frontiers in Algorithmics Ranking of Closeness Centrality for Large-Scale Social Networks. Springer.

Reijers, H. A. (2002) Product-based design of business processes applied within the financial services. Journal of Research and Practice in Information Technology, 34, 110–122.

Salton, G. and M. J. McGill (1984) Introduction to Modern Information Retrieval, New York, McGraw-Hill.

Sokal, R. R. and C.D. Michener (1958) A Statistical Method for Evaluating Systematic Relationships. University of Kansas Scientific Bulletin, 38, 1409-1438.

Trier, M. and A. Bobrik (2007) IT-gestützte Visualisierung und Analyse von virtuellen Kontaktnetzwerken - Anwendungsfelder, Methodik und Vorteile. In Mueller, C. & Gronau, N. (Eds.) Analyse sozialer Netzwerke und Social Software - Grundlagen und Anwendungsbeispiele. Berlin, GITO-Verlag.

Trier, M., and A. Bobrik (2007) Analyzing the Dynamics of Community Formation Using Brokering Activities. Communities and Technologies. Berlin, Springer.

Vergidis, K., A. Tiwari, B. Majeed (2008) Business Process Analysis and Optimization: Beyond Reengineering. Transactions on IEEE Computer, Systems, Man, and Cybernetics, Part C: Applications and Reviews, 38, 69-82.

Wasserman, S. and K. Faust (2009) Social Network Analysis - Methods and Applications, Cambridge University Press.

Zhou, Y., and Y. Chen (2003) Project-oriented business process performance optimization. IEEE International Conference on Systems, Man and Cybernetics. IEEE.