11-28-2018

Detecting Structure In Chaos: A Customer Process Analysis Method

Elisabeth Z. Palvölgyi
Frankfurt School of Finance & Management, e.palvoelgyi@fs.de

DETECTING STRUCTURE IN CHAOS: A CUSTOMER PROCESS ANALYSIS METHOD

Research paper

Palvölgyi, Elisabeth Z., Frankfurt School of Finance & Management, Frankfurt, Germany, e.palvoelgyi@fs.de

Abstract

Detecting typical patterns in customer processes is the precondition for gaining an understanding about customer issues and needs in the course of performing their processes. Such insights can be translated into customer-centric service offerings that provide added value by enabling customers to reach their process objectives more effectively and rapidly, and with less effort. However, customer processes performed in less restrictive environments are extremely heterogeneous, which makes them difficult to analyse. Current approaches deal with this issue by considering customer processes in large scope and low detail, or vice versa. However, both views are required to understand customer processes comprehensively. Therefore, we present a novel customer process analysis method capable of detecting the hidden activity-cluster structure of customer processes. Consequently, both the detailed level of process activities and the aggregated cluster level are available for customer process analysis, which increases the chances of detecting patterns in these heterogeneous processes. We apply the method to two datasets and evaluate the results' validity and utility. Moreover, we demonstrate that the method outperforms alternative solution technologies. Finally, we provide new insights into customer process theory.

Keywords: Customer Process, Process Mining, Clustering, Sequence Mining, Service Design.

1 Introduction

The value of a service emerges when it is used by customers in the course of their own processes of value creation – this is one of the key statements of customer-dominant logic of service (Grönroos and Ravald, 2011; Heinonen et al., 2010). Aligning service processes with customer processes can accelerate customers' success in value creation by enabling them to reach process goals more effectively and rapidly, and with less effort (Bettencourt et al., 2013; Grönroos and Ravald, 2011; Payne et al., 2008; Heinonen et al., 2010). Knowledge about customers' activities and requirements in the course of their entire processes can enable companies to proactively and comprehensively provide customers with what they need in their next steps throughout the process (Behara et al., 2002; Vandermerwe, 2000). Moreover, the analysis of customer processes can lead to identification of issues and pitfalls that customers are facing in their processes. Such insights can be used to design services that support customers in reaching their process objectives more efficiently, for instance by taking over process tasks from customers or even managing the whole process for them (Bettencourt et al., 2013; Vandermerwe, 2000).

Customer processes can be defined as sequences of activities customers perform to reach a certain objective, such as buying a car, going on vacation or arranging retirement provisions (Payne et al., 2008; Behara et al., 2002). The precondition for taking customer processes as a basis for service design is to identify patterns in specific customer processes relevant to a certain service company. Several analysis methods consider customer processes in a great level of detail, but only look at a small section of the overall process (e.g. Le et al., 2017; Liu et al., 2017; Duan and Zhang, 2014; Chen et al., 2013; Fournier-Viger et al., 2012). Other approaches regard the full scope of customer processes but do this on a very high level only (e.g. Sands et al., 2016; Voorveld et al., 2016; Prinzie and Van Den Poel, 2011; Salazar et al., 2007). Naturally, such limited insights on customer processes restrict the chances of identifying opportunities for designing customer-process-centric services.

Following the design science research (DSR) paradigm, we design a method that enables the identification of activity clusters making up customer process networks (CPNs) (Hevner and Chatterjee, 2010; Hevner et al. 2004). Using this inner structure, customer processes can be considered in both large scope *and* detail when searching for patterns. Hence, in this study we strive to answer the following research question: How can the activity-cluster structure of customer processes be identified?

The answer to this question provides valuable and directly applicable implications for companies that wish to understand how different customer groups move through their customer processes, and what issues they might face – the precondition for designing services that meet these customers' needs in their processes. Moreover, we contribute to customer process research by uncovering the hidden substructure of these processes and designing a method to analyse customer processes in spite of their great heterogeneity. By providing such a method, and hence the chance to analyse and understand customer processes, we also go one step further towards aligning service and customer processes, as demanded by the paradigm of customer-dominant logic of service (Grönroos and Ravald, 2011; Heinonen et al., 2010).

First, we present the related literature and explain which business issue our suggested artefact addresses. We then explain the conceptual background and the customer process analysis method developed. We test the new method by applying it to two datasets and evaluating the validity and utility of the resulting activity clusters. Subsequently, we compare the performance of the developed method with that of alternative solution technologies. Finally, we summarize the study's results and implications.

2 Related Literature and Business Problem Addressed

Customer processes are discussed in the literature at different levels of detail and scope. Online clickstream sequences reflect customers' individual paths in the online world at the operations level. According to activity theory, this is the most detailed level for considering an activity (Hacker, 2001; Kuutti, 1996). Plenty of literature considers how clickstream sequences can be analysed to detect patterns (e.g. Le et al., 2017; Liu et al., 2017; Duan and Zhang, 2014; Chen et al., 2013; Salehi, 2013; Fournier-Viger et al., 2012; Wei et al., 2012; Wang and Wang, 2007; Jenamani et al., 2003; Mobasher et al., 2002; Spiliopoulou and Pohle, 2001). Knowledge about such sequential patterns can be used to optimize website content and structure according to customers' preferences (Jenamani et al., 2003; Spiliopoulou and Pohle, 2001), or can be fed into recommendation systems, proactively providing customers with information they will probably need for current or next activities (Li et al., 2017; Salehi, 2013; Wang and Wang, 2007; Mobasher et al., 2002).

Another stream of literature regards customer processes on a more comprehensive but less detailed level when searching for patterns in channel choice or product/service purchase sequences (e.g. Sands et al., 2016; Voorveld et al., 2016; Van Rijnsoever et al., 2012; Prinzie and Van Den Poel, 2011; Frambach et al., 2007; Paas et al., 2007; Salazar et al., 2007; Kamakura et al., 1991). Such information can be used for optimizing multi-channel service offerings or realizing cross-selling opportunities by offering the right service at the right time to the right customer.

Online clickstream sequences reflect customer activities at a detailed level, but such data naturally reflects only customer activities performed online. Furthermore, by considering a certain number of website clicks, usually just a few minutes of customer activities are captured (e.g. Dunne et al., 2011; Pan and Fesenmaier, 2006). However, especially customer processes that include the purchase of complex services can take weeks, months or even over a year, and usually also include offline activities (e.g. Gough and Nurullah, 2009; Savolainen 2009; Vroomen et al., 2005). In the channel choice and purchase sequences literature, customer processes are regarded over a longer period, which may even comprise several years (e.g. Paas et al., 2007; Kamakura et al., 1991). However, besides information on which channels, or products or services, were chosen in which chronological order, no further insights about customers' activities are given, e.g. around searching for and using the service. Hence, customer processes are considered in the literature either in a great level of detail but with limited scope, or vice versa.

The more of a customer process is known, the more opportunities for customer-centric service improvements or innovations can be identified. Consequently, customer processes should be analysed in large scope *and* detail. The current literature does not offer methods for doing this, which represents a major research gap. To address this business problem, we suggest a method for identifying the inner structure of CPNs that can be used for detecting patterns in customer processes by considering them in full scope and at a high level of detail. Thus, we focus on processes of individual customers entailing the purchase of a complex service.

3 Conceptual Background

Several papers present theoretical models of customer processes and state that they can be split into different phases, for instance pre-purchase, purchase and post-purchase (Behara et al., 2002; Ives and Learmonth, 1984; Lunn, 1974; Howard and Sheth, 1969; Nicosia, 1966). Choi et al. (2011) state that these phases occur not once, but again and again in a customer process. Pan and Fesenmaier (2006) discover that the information search phase of customer processes comprises several different sub-problems which the customer has to resolve one after another in order to finalize that phase. What all these approaches have in common is the idea of the existence of a substructure within customer processes, comprising different sub-problems or phases.

Customer processes are defined as sequences of activities (Payne et al., 2008; Behara et al., 2002). Hence, customer processes contain activities *and* the elements of the above-described substructure, which – corresponding to activity theory – describe the same phenomenon at different levels of detail (Hacker, 2001; Kuutti, 1996). Accordingly, each of these substructure elements contains several process activities (Pan and Fesenmaier, 2006). Thereby, each process activity can be assigned to only one of these elements (e.g. Pan and Fesenmaier, 2006; Behara et al., 2002; Ives and Learmonth, 1984). As each substructure element represents a certain sub-problem or phase (e.g. Martin and Woodside, 2012; Pan and Fesenmaier, 2006; Behara et al., 2002; Lunn, 1974; Howard and Sheth, 1969; Nicosia, 1966), we can assume that activities of the same substructure element are related to each other in terms of their contents.

In less restrictive environments, customers are free to choose which activities they wish to perform in what order for completing their processes (Günther and Van Der Aalst, 2007). This "will lead to more diverse and less-structured behaviour (and hence to) 'spaghetti'-models'" (Günther and Van Der Aalst, 2007, p. 330), when merging the individual customer processes into a process network for further analysis. However, given that 1) for each phase/sub-problem of a customer process a certain group of activities is available which is suitable for solving that particular phase (Choi et al., 2011; Savolainen, 2009) and 2) customers move in processes from phase to phase (Martin and Woodside, 2012; Choi et al., 2011; Pan and Fesenmaier, 2006) and 3) choose in each phase an individual combination of activities suitable for completing it (Van Rijnsoever et al., 2012; Choi et al., 2011; Pan and Fesenmaier, 2006), possibly jumping back and forth between activities before moving to the next phase (Savolainen, 2009), we can conclude that activities in the same phase succeed each other more often than they succeed activities of other phases. This differs from the generic heterogeneity and chaos in "spaghetti" models (Günther and Van Der Aalst, 2007). Hence, we expect that CPNs comprise clusters of highly interlinked process activities, each representing a different customer process phase/sub-problem.

Moreover, given that each customer comes up with their own idea of what activities to execute for solving a certain process phase/sub-problem (Günther and Van Der Aalst, 2007), the diversity and hence the number of different activities are expected to be high, while the frequency of each individual activity is expected to be low. Furthermore, given that customers choose different activity combinations for solving a certain process phase/sub-problem and perform these activities in individual chronological order (Van Rijnsoever et al., 2012; Choi et al., 2011; Pan and Fesenmaier, 2006; Günther and Van Der Aalst, 2007), it can be expected that not only these activities, but also the edges between them, are infrequent. We conclude that CPNs, which represent processes performed in non-restrictive environments, comprise clusters containing infrequent activities that are connected via infrequent edges. According to the customer process literature, the number of these clusters is limited (e.g. Martin and Woodside, 2012; Van Rijnsoever et al., 2012; Choi et al., 2011; Nussbaumer et al., 2011; Pan and Fesenmaier, 2006); hence, the number of activities per cluster is expected to be high.

Besides their large number of infrequent activities, CPNs contain several activities that are performed frequently. This phenomenon can be due, for example, to regulatory (e.g. in Germany one must visit a notary to buy a house) or rational reasons (e.g. for most people it makes sense to inspect a house before buying it; Savolainen, 2009). As each customer is free to determine how to approach such a frequent activity and what to do after completing it (Günther and Van Der Aalst, 2007), we expect that these frequent activities are connected with many different predecessor and successor activities. Consequently, frequent activities come with high numbers of edges. There is no common order of consecutive activities, as it would be the case if frequent activities had few edges. Overall, we can conclude the following: CPNs comprise clusters of highly-interlinked infrequent activities. This cluster structure is overlapped by the numerous edges of frequent activities interlinking them with numerous infrequent activities all across the network. Due to their many edges, we call these frequent activities “hyper-connectors”.

Based on the available customer process knowledge discussed above, the following requirements (R) for a customer process analysis method can be formulated. The method should be capable of detecting clusters of infrequent activities in CPNs in spite of the overlapping impact of hyper-connectors (R1). It should generate preferably large clusters (R2) comprising activities with a maximized number of interconnections (R3). The method should be able to balance R2 and R3 (R4). Activities assigned to the same cluster should be related to each other from a content perspective, representing the solution of a certain sub-problem (R5). The resulting clusters’ positions in the overall CPN should reflect a reasonable order of dealing with these sub-problems (R6). In the next section, we present the components of the novel customer process analysis method and explain how they are meant to realize these requirements.

4 The Customer Process Analysis Method

The customer process analysis method is based on an existing clustering algorithm that is enhanced in a way that allows the generation of activity clusters by taking into account the specifics of CPNs. Algorithms suitable for performing clustering tasks in networks are available in the fields of process mining and social network analysis (SNA).

Process mining is an approach to “unveil previously hidden knowledge” (Günther and Van Der Aalst, 2007, p. 328) in process data to understand the structure of real-world processes (Van Der Aalst, 2011). To deal with unstructured spaghetti models, which are a major challenge in process mining (Bose et al., 2013), several methods based on trace, hierarchical and activity clustering are suggested. Trace and hierarchical clustering aim at splitting up the overall process model into more easily understandable sub-models, the former by clustering event logs and the latter by aggregating process parts to a more abstract level. The precondition for clustering is the identification of similarities in the process data in the form of frequent patterns (Koschmider, 2017; Diamantini et al., 2013, 2016; Cho et al., 2014; Bose et al., 2011; Song et al., 2008) or semantic similarities (Nguyen et al., 2016; Richetti et al., 2014). Additionally process attributes (Hompes et al., 2014; Cho et al., 2014; Seret et al., 2014) and existing domain knowledge (Yahya et al., 2016) are utilized for clustering. However, due to the heterogeneity of customer processes, it is questionable whether sufficient frequent patterns can be identified at activity level. Moreover, besides the names and chronological order of the activities, no further information is available for analysing customer processes. Furthermore, it cannot be assumed that customer processes contain semantically related elements.

Clustering activities is another approach suggested for mining unstructured processes. Again, existing process knowledge (Baier and Mendling, 2013; Rozsnyai et al., 2011; Li et al., 2010; Günther and Van Der Aalst, 2006) or semantic similarities (Smirnov et al., 2012) are taken as a basis for clustering. Also, causal relations or frequent successions between activities are considered for clustering (Tax et al., 2016; Verbeek et al., 2016; Verbeek and Van Der Aalst, 2014; Van Dongen and Adriansyah, 2009). However, as the majority of customer process activities are infrequent, it is difficult to identify such relations in CPNs.

These approaches are either applied to business processes (e.g. Nguyen et al., 2016; Verbeek et al., 2016) or to such parts of customer processes that are influenced by interaction with the given supplier (Diamantini et al., 2016; Song et al., 2008). Although these are unstructured real-life processes, they do not reach the degree of heterogeneity of customer processes made up of infrequent activity clusters. Two papers deal with mining of "activities of daily life", recorded, for example, by a smart home application (Koschmider, 2017; Tax et al., 2016). Due to repetition and the existence of uniform patterns in many people's daily routines (e.g. switching on lights at night, or going to work in the morning), these processes again are less heterogeneous than the customer processes discussed in this study.

To analyse customer processes, we require a clustering approach that works without the need of having frequent sequences in the process data. Furthermore, the approach should utilize the only customer process activity information available to us: how activities are connected to each other. The following two process mining approaches meet these criteria: "fuzzy mining" (Günther and Van Der Aalst, 2007) and "activity mining" (Günther et al., 2009).

Finding clusters of highly interlinked items within a network is also a common application in SNA (Borgatti et al., 2013). In this field, networks of persons interlinked by social ties, such as friendship, are investigated to find "cohesive sub-groups" – groups of people "who interact with each other to such an extent that they could be considered to be a separate entity" (Borgatti et al., 2013, p. 181). SNA has already been applied in process mining contexts to detect who is interacting with whom throughout a process (Mans et al., 2008; Van der Aalst and Song, 2004). However, to analyse CPNs, the SNA algorithm would be applied to the process activities themselves and not the individuals performing them.

The SNA algorithms "Factions" and "Girvan-Newman" do not allow network items to be assigned to more than one cluster (Borgatti et al., 2013; Girvan and Newman, 2002; De Amorim et al., 1992; Glover, 1989, 1990), which is in line with customer process literature. The "Girvan-Newman" algorithm strives to "find the structurally important edges whose removal fragments the network" (Borgatti et al., 2013, p. 195) into cohesive sub-groups. However, in "spaghetti-like" CPNs, the existence of such "structurally important edges" is questionable. The SNA algorithm "Factions" uses combinatorial optimization techniques to identify such sub-groups by assigning all network items randomly to one of the groups, calculating the overall fit, moving some items to other groups, recalculating the overall fit and repeating this procedure until no further improvement is possible (Borgatti et al., 2013; De Amorim et al., 1992; Glover, 1989, 1990). The factor determining how well an activity fits a group is the number of edges connecting the given activity with other group members. Hence, in addition to the two process mining approaches mentioned, the "Factions" algorithm also seems to be appropriate for identifying activity clusters in CPNs.

Process mining tools focus on those items within a network which are frequent and hide all non-frequent items in order to build process models reflecting the most important aspects of the underlying processes (Günther and Van Der Aalst, 2007). The reason for this is that process mining tools were designed for detecting and analysing business processes which are represented best by those elements that occur repeatedly (Van Der Aalst, 2011; Günther et al., 2009; Günther and Van Der Aalst, 2007). Customer processes, by contrast, are performed in less restrictive environments and are extremely heterogeneous.

Thus, the structure of CPNs is characterized by those activities and edges which are non-frequent. In SNA the frequency of activities and edges is ignored and solely the structure of the network is analysed (Borgatti et al., 2013). Therefore, SNA algorithms seem to be more suitable for identifying infrequent activity clusters in CPNs than process mining approaches. Consequently, the new customer process analysis method is based on the SNA algorithm “Factions”. In Section 5.6, we validate this rationale by comparing the new method’s performance with that of the two identified process mining approaches.

The extent to which sub-group members are tied together is measured by group density. The density of a group in a directed graph is $D = e / n\ (n-1)$, with $0 \leq D \leq 1$, where e = number of directed edges, and n = number of items belonging to the group (Borgatti et al., 2013). For small groups it is easier to reach high densities due to their smaller maximal number of potential edges. Hence, generating activity clusters by considering their densities favours the generation of many small clusters, which is in contrast to the customer process literature. To mitigate this issue and meet the requirements defined earlier in this section, we require a measure that rewards both an increasing density (R3) and a growing number of activities per cluster (R2) to the same extent (R4). Hence, this measure should be a strictly monotonically increasing and symmetric function of these two values. Calculating the simple factor product of a cluster’s density and the number of activities within that cluster fulfils these criteria. We call this new measure “weighted density”. It is meant to “reward” large activity groups with a higher value and “punish” small groups accordingly.

The customer process analysis method comprises a round-based approach to detect infrequent activity clusters in CPNs, in spite of the existence of hyper-connectors that overlap the cluster structure with their many edges. First, the process activities are sorted by number of edges in ascending order. Then, the activities are added individually to the network in the same order. After each round of adding one further activity, the “Factions” algorithm is applied to the network and the sum of the generated clusters' weighted densities calculated. Following the customer process literature, we do not regard pairs as groups and therefore calculate the weighted density sum by considering activity groups containing three or more activities only. The round with the maximum weighted density sum is identified. This procedure allows infrequent activity clusters to be generated first and identifies the round in which the addition of activities with many edges starts to spoil the clusters’ generation, which addresses R1.

The results of the round-based procedure are (1) those clusters that were generated in the round with the maximum weighted density sum and (2) several activities (in fact, the hyper-connectors) that would have been added in later rounds and therefore are not yet assigned to any cluster. Again, we do not regard pairs as groups; therefore, clusters containing two activities are split up and their activities added to those currently unassigned. These leftover activities are assigned to the identified clusters by checking for each of the activities in which group they would generate maximum density gain. The assigned activities must not be added to their new clusters before the optimal assignment for each activity is identified, otherwise the first hyper-connector added to one of the clusters would, with its many edges, pull the remaining activities into the same cluster. This would inflate the respective cluster and significantly worsen its density. The final result of the described procedures is a set of clusters. Each cluster contains a number of highly interlinked activities that are related to each other from a content perspective (R5). Moreover, the order of these clusters in the CPN is expected be reasonable in terms of their content (R6).

Following the DSR paradigm, we performed three design cycles for designing the customer process analysis method (Hevner and Chatterjee, 2010; Hevner et al., 2004). The method can be summarized as follows. Build up the process network P by adding its process activities according to the number of their edges in ascending order individually to the network. After each round, apply the SNA algorithm “Factions” to the network data and calculate the sum of weighted densities of clusters containing at least three activities. Identify the round with the maximum weighted density sum. Split up those of the resulting clusters that contain only two activities and put their activities aside for reassignment.

This procedure generates the following outputs: (1) the set of clusters B identified in the round with the maximum weighted density sum, excluding clusters containing two activities and (2) the set of free activities F not yet assigned to any cluster. The subsequent procedure aims to assign each activity $A_i \in F$ to that cluster $C_j \in B$ where maximum density gain is reached by adding A_i. The density gain of cluster C_j by adding A_i is $G_{ij} = D_2 – D_1$, where D_1 is the initial density of C_j, and D_2 is its density after adding A_i. We use the following running indices: $0 < i \leq$ number of activities $A_i \in F$; and $0 < j \leq$ number of clusters $C_j \in B$.

algorithm ActivityAssignment (P, B, F)

repeat

  for each activity $A_i \in F$ **do**

    calculate density gain $G_{ij}(A_i, C_j)$;

  for each activity $A_i \in F$ **do**

    if $G_{ik} > 0$, with $G_{ik} = \max \{G_{ij}\}$, where k refers to the cluster with maximal positive density gain, **do**

      put A_i into cluster C_k (make random selection if more such clusters exist);

    else

      leave A_i in F;

until F = { } **or** $G_{ij} \leq 0$ for each i and each j;

return B, F;

This method is suitable for analysing spaghetti-like CPNs, made up of customer processes that differ considerably regarding activities and their order, reflecting unstructured and individual behaviour in a less restrictive environment. An indicator for such a diversity of activity sequences in customer processes is a strong positive correlation of process activities' frequency with their number of edges. The method is unsuitable for analysing customer processes in restrictive environments, where customers are forced to perform steps in a certain order. The method performs well for customer processes including several observable activities, like when purchasing complex services. It is less suitable for very short and rather emotionally triggered processes.

5 Empirical Study

This section evaluates the designed method for detecting high-density activity clusters within CPNs. First, we test whether the developed customer process analysis method is indeed capable of identifying high-density activity clusters in customer processes. We do this by applying the method on a dataset representing the customer process "Buying a house/apartment".

Then we assess the validity and utility of the results produced by the method (as suggested by Gregor and Hevner, 2013). For evaluating the validity of the resulting activity clusters, we consider their densities and examine from a content perspective whether the activities within each cluster are similar. Additionally, we investigate whether the positioning of the clusters in the network makes sense. To perform this, we require the process network on a cluster level, which we generate by replacing the activities in the individual processes by their respective clusters and merging the resulting high-level processes into a process network by applying process mining techniques (Günther and Van Der Aalst, 2007). In the succeeding section we demonstrate the utility of the generated clusters by presenting how they can be utilized to identify different customer groups, their specific processes and the issues they face when moving through their processes. For this purpose, we apply the sequence mining algorithm CM-SPAM (Fournier-Viger et al., 2014) on the high-level processes to identify frequent sequences on a cluster level. Then we cluster customers according to the type of frequent sequence they have performed and their previous knowledge and experience of the products and processes involved, as these are seen as relevant for differing between purchasing behaviours in the literature (Jaillet, 2003; Flynn and Goldsmith, 1999; Brucks, 1985).

Subsequently we apply the method as a robustness check on a second dataset, reflecting the customer process "Making an investment", in order to demonstrate that it is valuable for analysing customer processes in general. Finally, we compare the performance of our designed method with those of the identified alternative solution technologies (as suggested by Venable, 2006): the process mining approaches "fuzzy mining" (Günther and Van Der Aalst, 2007) and "activity mining" (Günther et al., 2009).

5.1 Data

A first set of customer process data was gathered by conducting a customer survey at a German direct bank. In an online survey, individual customers who had a mortgage from this bank were asked about their "buying a house/apartment" process. First, customers were asked to write each of their process activities in free-text fields and/or choose (further) appropriate activities from an activity list entered by previously questioned customer groups. Next, customers were asked to order the entered/chosen activities chronologically to obtain their individual processes. In total, 321 individual customer processes were collected. By applying the constant comparison procedure, 145 activities were derived from the text data entered (Glaser, 1965). Two researchers independently mapped the free-text field entries with the 145 activities. Conflicting cases were resolved by discussing them and agreeing the most appropriate match. The most frequent 50% of the total 145 standard activities represented 95% of activity volume. Hence, we deleted 73 less frequent activities with the aim of increasing efficiency by analysing a reduced dataset (Han et al., 2011).

Additionally, we measured customers' subjective knowledge and experience concerning the products/services involved in the customer process "buying a house/apartment": the house and the mortgage (following Park and Moon, 2003; Flynn and Goldsmith, 1999; Johnson and Russo, 1984; Park and Lessig, 1981; Bettman and Park, 1980). We also measured participants' knowledge and experience concerning the *processes* of purchasing those products/services (Flynn and Goldsmith, 1999). Furthermore, we asked customers to indicate on a seven-point Likert scale the extent to which they are satisfied with their "buying a house/apartment" process (with 7 = very satisfied and 1 = very dissatisfied).

Via the same procedure, we collected the "Making an investment" processes of 49 persons. We chose this process as this study focuses on analysing customer processes entailing the purchase of a complex service. Services allowing the customer to make an investment are regarded as such (Nussbaumer et al., 2011; Gough and Nurullah, 2009; Vroomen et al., 2005).

5.2 Primary study results

When arranging the 72 available customer process activities in the dataset according to the number of edges in ascending order (see left of Figure 1), about two-thirds of activities have fewer than the average of 27 edges per activity, while at the other end of the exponentially shaped curve there are activities with edge numbers far beyond average – the "hyper-connectors". A Spearman's correlation test revealed a strong positive correlation between frequency of activities and number of edges, which is statistically significant, $rs = 0.893$, $p = 0.000$. This means that in individual processes the same activity is indeed approached from different predecessor activities and followed by different successor activities.

We applied process mining techniques to integrate all individual customer processes into a process network containing the 72 activities and the edges between them (Günther and Van Der Aalst, 2007). As expected, the resulting CPN was a chaotic "spaghetti-like real-life process" (Swinnen et al., 2012, p. 89). To detect cohesive sub-groups of activities within the CPN, we applied the developed customer process analysis method and utilized the SNA software UCINET with its plugin NetDraw (Borgatti et al., 2002).

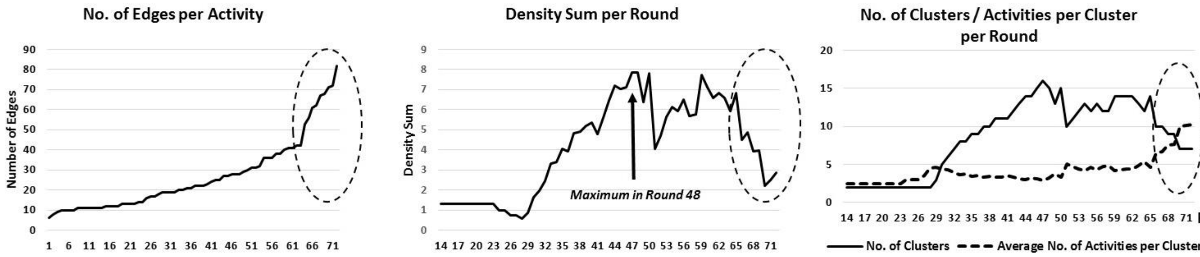

Figure 1. Hyper-connectors spoil the generation of high-density activity clusters.

As shown in Figure 2, the sum of weighted densities across all generated groups first increases, while adding one activity after another to the network, but decreases again when hyper-connectors are included. Their numerous edges blur the network structure built in the previous rounds. The reason can be seen on the right of Figure 1: initially the number of clusters generated per additional round increases, with each group contributing to the increasing density sum. As of Round 49 the number of clusters remains quite stable, while the number of activities added to the clusters keeps growing, leading to increasing cluster sizes. When adding the hyper-connectors in the last rounds, the number of clusters rapidly decreases. The available process activities are allocated to fewer groups, leading to even more activities per group. For larger groups it is harder to reach high densities due to their higher maximal number of potential edges. Hence both the smaller number of clusters and the larger number of activities per cluster yield lower densities for groups generated in the last rounds. Due to their interconnections to numerous activities all across the network, the hyper-connectors seemingly have the effect of merging the initially more but smaller groups into fewer but larger ones, bringing down the densities of these groups. Considering the clusters' density sum per round makes the effect of adding hyper-connectors to the network even more visible. In contrast to the weighted density sum, this measure does not "reward" the occurrence of large groups in the last rounds. Therefore, as can be seen in the middle of Figure 1, the density sum rapidly decreases after having reached a maximum in Round 48.

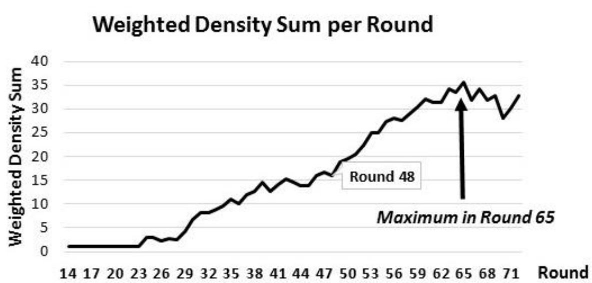

Figure 2. Identification of best round by weighted density sum measure.

As displayed in Figure 2, the maximum value of the weighted density sum of groups with more than two activities was 35.57 and was generated in Round 65. As it can be seen on the right of Figure 1, Round 65 is the last spot before the number of clusters and their density values start decreasing. The 14 activity clusters generated in Round 65 had an overall density of 6.85 (as shown in Table 1). Seven activities had not been assigned to any group yet. Applying the above procedure on the results of Round 65 for assigning the remaining activities led to 14 final clusters with a density sum of 10.31.

We performed the same assignment procedure for the groups identified in Round 48 – the round with the maximum density sum. Table 1 shows that the final result of Round 48 (after activity assignment) is outperformed by that of Round 65 regarding density sum. Hence, measuring the sum of weighted densities of groups with more than two activities is indeed a better indicator for identifying the round with the best activity groups than considering the groups' density sum.

	Clusters Produced By SNA / Process Mining Algorithm					Final Clusters Produced By Activity Assignment Algorithm			
	No. of Clusters	Avg. No. of Activities per Cluster	Avg. Density	Density Sum	Weighted Density Sum	No. of Clus-ters	Avg. No. of Activities per Cluster	Avg. Density	Density Sum
SNA, Round 48	15	3.20	0.52	**7.85**	**15.94**	12	5.25	0.68	**8.15**
SNA, Round 65	14	4.64	0.49	**6.85**	**35.57**	14	4.36	0.74	**10.31**
Fuzzy Mining, Round 70	14	3.93	0.36	**5.03**	**14.71**	14	4.21	0.57	**7.95**
Activity Mining, Round 71	12	5.92	0.36	**4.35**	**34.85**	9	7.00	0.68	**6.10**

Table 1. Results of clustering approaches.

All in all, the novel customer process analysis method succeeded in identifying 14 activity clusters in the CPN by preventing hyper-connectors from spoiling cluster generation (R1). The method identified the round where the clusters profited from increasing activity counts (as displayed on the right of Figure 1), but still showed high densities (R2, R3, R4). These clusters served as a starting point for assigning additional activities to the clusters and hence further increasing the clusters' sizes and average densities (R2, R3, R4).

5.3 Validity of results

The density values of the detected activity clusters indicate to which extent the members of each group are interlinked with each other. The average density per cluster is 0.74, which can be regarded as a good result (R3), especially when compared with the overall CPN's density of 0.19 (Borgatti et al., 2013).

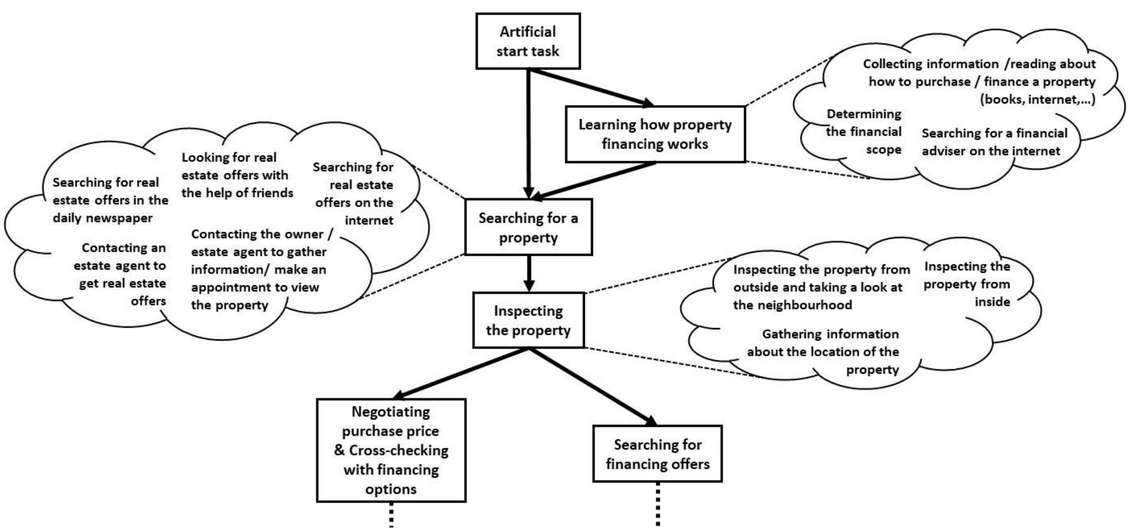

Figure 3. Excerpt of the CPN "Buying a house/apartment" on a cluster level, including the contents of three selected clusters.

Each identified cluster contains activities that are related to each other from a content perspective and represent a certain sub-problem that has to be handled in the course of the customer process (R5). This has been confirmed by subject-matter experts of the direct bank whose customers had been interrogated for this study, and is also in line with previous observations reported in the customer process literature (e.g. Pan and

Fesenmaier, 2006). The contents of three clusters are displayed as examples in Figure 3. Also, the arrangement of the clusters in the network does make sense concerning their contents, which also has been confirmed by subject-matter experts (R6). For example, clusters containing search activities are placed near the starting point of the network, while clusters around signing the mortgage or the purchase contract are found at a later stage. All in all, we can state that the designed method detected actually existing sub-structures in customer processes.

5.4 Utility of results

Sequence mining is an approach to detect patterns – that is, frequently occurring sequences – in sequential data (Fournier-Viger et al., 2014). Uncovering the cluster structure of CPNs allows processes to be mined not only on an activity but also on a cluster level. This offers additional and previously unavailable chances to gain further insights into customer processes. In the following we present a case example to demonstrate how sequential patterns detected in customer processes on a cluster level can be used for identifying issues different customer groups face in the course of their processes.

We applied the sequence mining algorithm CM-SPAM (Fournier-Viger et al., 2014) on the processes at cluster level and identified the following frequent sequences. Some customers start dealing with the topic of mortgages before searching for and/or inspecting eligible properties (Sequence A). Other customers perform these activities after having searched for and/or visited properties (Sequence B). When performing a two-step cluster analysis, an attribute indicating whether the respective customer performed Sequence A or B and two further experience-related attributes served as the most relevant predictors for revealing three clusters of customers: 34.8% of customers did not deal with the mortgage topic before searching for and/or visiting properties (Sequence B) and had high levels of experience and knowledge concerning the relevant products/services and processes when buying a house. Therefore we call them the Experienced. Seemingly the Experienced can assess their financial abilities without gathering further information about mortgages. Consequently they reach their process objective faster and are more satisfied with the overall process than other groups. The second cluster is made up of 29.4% of customers. According to a Kruskal-Wallis test, they are significantly less experienced and have significantly less knowledge than the Experienced (Kruskal and Wallis, 1952). They have difficulties in assessing financial options and therefore inform themselves about mortgages before searching for/visiting properties (Sequence A). The third group has significantly less experience and knowledge than the second cluster, therefore we call them the Greenhorns. They immediately start searching for and/or visiting properties, without gathering information about mortgages in advance (Sequence B). Furthermore, they are less satisfied with the process than other customers.

Seemingly the Greenhorns have so little knowledge and experience available that they fail to come up with the idea of checking information about mortgages first, even though this might have been helpful for them. Hence, offering such information via the bank’s channels (e.g. the website) will not serve the Greenhorns. Instead, they have to be provided with support in such a way that they receive it automatically when moving through the process. For example, the bank could cooperate with parties the customer is in contact with at an early stage, such as property developers or real estate agents, and empower them to provide the customer with relevant information. Another option is the implementation of an integrated financial services and real estate online portal, providing Greenhorns with both at once, real estate offerings and matching mortgages.

These results show that identification of activity clusters in CPNs is useful for detecting patterns in customer processes on a cluster level, which can reveal further insights about different customer types and their issues and needs. Such insights can be utilized for identifying opportunities for customer-centric service design.

5.5 Secondary study results

We applied the customer process analysis method on a second dataset, reflecting the customer process “Making an investment”, and succeeded in identifying 8 clusters, containing an average of 3.13 activities each, with a density sum of 4.42. As the average density sum of 0.44 per cluster is higher than the overall network density of 0.11, and as the grouping of activities into different clusters again makes sense from a content perspective, we can state that the developed method also succeeded in identifying cohesive activity groups in this network, and hence can be considered generically applicable for the analysis of customer processes performed in less restrictive environments.

5.6 Comparison with other solution technologies

Two process mining approaches, "fuzzy mining" (Günther and Van Der Aalst, 2007) and "activity mining" (Günther et al., 2009), have been identified as alternative solution technologies in Section 4 of this study. The former is implemented as the Fuzzy Miner plugin and the latter as the Stream Scope plugin in the ProM framework (Van Dongen et al., 2005). In the following we report the results of comparing the performance of the developed method with those of the alternative solution technologies.

We applied the Fuzzy Miner on the data reflecting the customer process "Buying a house/apartment" in the round-based manner reported. The highest weighted density sum achieved by this approach was 14.71 in Round 70. When applying the Fuzzy Miner there was no decline of density sum in the last third of the 72 rounds. This is because the Fuzzy Miner does not force each activity into a group, as is done by the SNA Factions algorithm. Instead the most frequent activities are left outside. As the frequency of activities goes hand in hand with the number of edges, those activities are in fact the hyper-connectors. Consequently they do not "spoil" the generated activity clusters. But the round-based approach can still be regarded as beneficial, as the maximum density occurs before the last round. The activity-assignment procedure led to the generation of 14 clusters with an overall density sum of 7.95 (displayed in Table 1).

As a next step we tested the ability of the Stream Scope ProM plugin to generate high-density activity groups (Günther et al., 2009; Van Dongen et al., 2005). The activity mining algorithm in that tool generates clusters by identifying highly correlated activities. Activities are seen as highly correlated if they "frequently occur closely together" (Günther et al., 2009, p. 130). By considering not only one but a certain number of previous activities when searching for correlations in processes, the algorithm can identify correlations in highly unstructured and heterogeneous process networks (Günther et al., 2009). We applied this tool in the same round-based approach on our network data. With a weighted density sum of 34.85, the best results were found in Round 71. Again the hyper-connectors added in the last rounds did not have such a big impact on the quality of the resulting activity groups. The reason is that the activity mining algorithm primarily searches for pairs of highly correlated activities, before merging the identified pairs into larger clusters. Therefore the hyper-connectors have little chance to spoil activity-cluster detection. The round-based approach can still be considered beneficial, as the maximum density does not occur in the last but the penultimate round. Then we applied the activity-assignment procedure on the clusters identified in Round 71, which led to the generation of 9 groups with a density sum of 6.1. Thus none of the alternative solution technologies succeeded in outperforming the results generated by our suggested round-based SNA approach for identifying high-density activity clusters in CPNs.

6 Summary and Implications

Following the DSR paradigm, we designed a method for detecting clusters of highly interlinked activities in CPNs by extending an existing SNA algorithm (Hevner and Chatterjee, 2010; Hevner et al., 2004). We applied the new method on two datasets and have shown that the resulting clusters make sense from both a measurement and a content perspective. This confirmed that the method is capable of analysing customer processes performed in less restrictive environments. Finally we demonstrated that the performance of the novel method outperforms other solution technologies from the field of process mining (Venable, 2006).

By designing this method, we contributed to closing the research gap explained in the literature review in this paper. In contrast to approaches in the channel/product choice sequence literature (e.g. Van Rijnsoever et al., 2012; Prinzie and Van Den Poel, 2011; Frambach et al., 2007), process activities are not cut off but integrated into a more generic view on a data-driven and hence customer-behaviour-driven basis when building activity clusters. Hence, contrary to approaches reported in the clickstream literature (e.g. Duan and Zhang, 2014; Chen et al., 2013; Salehi, 2013; Fournier-Viger et al., 2012), the designed method allows the analysis of customer processes at the detailed level of process activities *and* comprehensively on a cluster level. This allows to understand the activity-cluster structure of customer processes and increases the chances of revealing patterns in them.

In a case example, we showed that these patterns are helpful for identifying issues different customer groups face in the course of their processes. These insights can be translated into opportunities for customer-centric service design. But the process structure itself can also be considered for this purpose. For example, websites or mobile apps meant to support customers in performing a specific process type can be structured according to activity clusters and hence offer customers an intuitively usable service (Schmidt-Rauch and Schwabe, 2014; Dohmen et al., 2009). Furthermore, analogous to business process optimization, the CPN structure can be improved: services could make certain customer activities or activity clusters obsolete, for instance by automatizing them or sourcing them from customers, or by changing the cluster order (Dumas et al., 2013). All in all, this method for customer process analysis offers the possibility to gain further insights about what customers do and what they might need in the course of their processes. Such knowledge can be utilized to better align service processes with customer processes to enable customers to reach their process goals better, faster and with less effort (Grönroos and Ravald, 2011; Heinonen et al., 2010). Thus, by introducing the novel customer process analysis method, we contribute to realizing an alignment of customer and business processes as demanded by the paradigm of customer-dominant logic of service (Grönroos and Ravald, 2011; Heinonen et al., 2010).

This paper provides new insights into customer process theory. First, it confirms that customer processes, which comprise the purchase of a complex service and are performed in non-restrictive environments, are extremely heterogeneous, making pattern detection difficult. This non-homogeneity is reflected by the broad range of activity combinations performed by customers in the course of their processes. However, there are also activities which occur in almost every process. Each of these high-frequency activities, which we call hyper-connectors, is approached from and succeeded by many different activities, which again shows that customers move through processes in a unique way. With their many edges, hyper-connectors overlap the underlying cluster structure and prevent the identification of cohesive activity groups when applying SNA algorithms. The suggested round-based approach mitigates this issue by identifying hyper-connectors that must be excluded from the network to be able to detect activity clusters.

Furthermore, we discovered that not only information search (Pan and Fesenmaier, 2006), but the complete customer process comprises a series of sub-problems that must be resolved to reach the process objective. This is also supported by previous observations that the search phase is carried out not once but repeatedly in the course of a customer process (Choi et al., 2011). This makes sense, as there might be several sub-problems requiring repeated search activities to resolve; for instance, in the house-buying process the sub-problems “Finding an appropriate property”, “Finding an appropriate mortgage offer” and “Familiarizing oneself with the topic ‘mortgages’”. Each sub-problem is represented by a cluster of highly interlinked activities. This means that a certain number of activities is available for resolving each sub-problem and each customer chooses one or more of these activities to handle the respective sub-problem. Interestingly, there are also sub-problems comprising activities for solving two issues, which means customers have to work in parallel on two interdependent topics, making the process even more cumbersome. By applying the designed method on a second dataset representing another customer process, we demonstrated that the insights gained on customer processes are in general valid for these types of processes.

Designing an IT artefact that is capable of performing the procedures of the customer process analysis method developed will be the next step of this research. Furthermore, the design of an IT artefact which additionally allows the user to visualize and further analyse CPNs would be helpful.

References

Baier, T. and J. Mendling (2013). "Bridging abstraction layers in process mining by automated matching of events and activities." In: *Business Process Management. Lecture Notes in Computer Science*. Ed. by F. Daniel, J. Wang and B. Weber. Berlin: Springer, p. 17.

Behara, R. S., G. F. Fontenot and A. B. Gresham (2002). "Customer process approach building loyalty." *Total Quality Management* 13 (5), 603–611.

Bettencourt, L. A., S. W. Brown and N. J. Sirianni (2013). "The secret to true service innovation." *Business Horizons* 56 (1), 13–22.

Bettman, J. R. and C. W. Park (1980). "Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis." *Journal of Consumer Research* 7 (3), 234–248.

Borgatti, S. P., M. G. Everett and J. C. Johnson (2013). *Analyzing Social Networks*. Los Angeles CA: SAGE.

Borgatti, S. P., M. G. Everett and L. C. Freeman (2002). *Ucinet for Windows: Software for Social Network Analysis*. Harvard MA: Analytic Technologies.

Bose, R. J. C., R. S. Mans and W. M. P. Van Der Aalst (2013). "Wanna improve process mining results?" In: *IEEE Symposium on Computational Intelligence and Data Mining*. Ed. by B. Hammer, Z. H. Zhou, L. Wang and N. Chawla. Singapore: IEEE, p. 127.

Bose, R. J. C., H. M. W. Verbeek and W. M. P. Van Der Aalst (2011). "Discovering hierarchical process models using ProM." In: *CAiSE Forum 2011. Volume 107 of Lecture Notes in Business Information Processing*. Ed. by S. Nurcan. Berlin: Springer, p. 33.

Brucks, M. (1985). "The effects of product class knowledge on information search behaviour." *Journal of Consumer Research* 16 (1), 1–16.

Chen, J., X. Zhou and Q. Jin (2013). "Recommendation of optimized information seeking process based on the similarity of user access behavior patterns." *Personal and Ubiquitous Computing* 17 (8), 1671–1681.

Cho, M., M. Song and S. Yoo (2014). "A systematic methodology for outpatient process analysis based on process mining." In: *Asia-Pacific Conference on Business Process Management Proceedings*. Ed. by C. Ouyang and J.-Y. Jung. Cham: Springer, p. 31.

Choi, S., X. Y., Lehto, A. M. Morrison and S. S. Jang (2011). "Structure of travel planning processes and information use patterns." *Journal of Travel Research* 51 (1), 26–40.

De Amorim, S. G., J. P. Barthélemy and C. C. Ribeiro (1992). "Clustering and clique partitioning: simulated annealing and tabu search approaches." *Journal of Classification* 9 (1), 17–41.

Diamantini, C., L. Genga and D. Potena (2016). "Behavioral process mining for unstructured processes." *Journal of Intelligent Information Systems* 47 (1), 5–32.

Diamantini, C., L. Genga, D. Potena and E. Storti (2013). "Pattern discovery from innovation processes." In: *2013 International Conference on Collaboration Technologies and Systems (CTS)*. Ed. by G. C. Fox and W. W. Smari. San Diego CA: IEEE, p. 457.

Dohmen, A., J. Moormann and M. Rosemann (2009). "Using a smartphone application for customer-centric banking." In: *ACIS 2009 Proceedings*. Ed. by H. Scheepers and M. Davern. Atlanta GA: AIS, p. 385.

Duan, W. and J. Zhang (2014). "Customer journeys on online purchase: Search engine, social media, and third party advertising." In: *PACIS 2014 Proceedings*. Ed. by K. Siau, Q. Li and X. Guo. Atlanta GA: AIS, p. 96.

Dumas, M., M. La Rosa, J. Mendling and H. A. Reijers (2013). *Fundamentals of Business Process Management*. Heidelberg: Springer.

Dunne, G., S. Flanagan and J. Buckley (2011). "Towards a decision making model for city break travel." *International Journal of Culture, Tourism and Hospitality Research* 5 (2), 158–172.

Flynn, L. R. and R. E. Goldsmith (1999). "A short, reliable measure of subjective knowledge." *Journal of Business Research* 46 (1), 57–66.

Fournier-Viger, P., U. Faghihi, R. Nkambou and E. M. Nguifo (2012). "CMRules: Mining sequential rules common to several sequences." *Knowledge-Based Systems* 25 (1), 63–76.

Fournier-Viger, P., A. Gomariz, M. Campos and R. Thomas (2014). "Fast vertical mining of sequential patterns using co-occurrence information." In: *Advances in Knowledge Discovery and Data Mining*. Ed. by V. Tseng, T. Bao Ho, Z. Zhou, A. Chen and H. Kao. Cham: Springer, p. 40.

Frambach, R. T., H. C. Roest and T. V. Krishnan (2007). "The impact of consumer internet experience on channel preference and usage intentions across the different stages of the buying process." *Journal of Interactive Marketing* 21 (2), 26–41.

Girvan, M. and M. E. Newman (2002). "Community structure in social and biological networks." *Proceedings of the National Academy Of Sciences* 99 (12), 7821–7826.

Glaser, B. G. (1965). "The constant comparative method of qualitative analysis." *Social Problems* 12 (4), 436–445.

Glover, F. (1989). "Tabu search-part I." *ORSA Journal on Computing* 1 (3), 190–206.

Glover, F. (1990). "Tabu search-part II." *ORSA Journal on Computing* 2 (1), 4–32.

Gough, O. and M. Nurullah (2009). "Understanding what drives the purchase decision in pension and investment products." *Journal of Financial Services Marketing* 14 (2), 152–172.

Gregor, S. and A. R. Hevner (2013). "Positioning and presenting design science research for maximum impact." *MIS Quarterly* 37 (2), 337–355.

Grönroos, C. and A. Ravald (2011). "Service as business logic: Implications for value creation and marketing." *Journal of Service Management* 22 (1), 5–22.

Günther, C. W., A. Rozinat and W. M. P. Van Der Aalst (2009). "Activity mining by global trace segmentation." In: *Business Process Management Workshops. BPM 2009. Lecture Notes in Business Information Processing*. Ed. by S. Rinderle-Ma, S. Sadiq and F. Leymann. Berlin: Springer, p. 128.

Günther, C. W. and W. M. P. Van Der Aalst (2006). "Mining activity clusters from low-level event logs." In: *BETA Working Paper Series*. Eindhoven: Eindhoven University of Technology, p. 165.

Günther, C. W. and W. M. P. Van Der Aalst (2007). "Fuzzy mining–adaptive process simplification based on multi-perspective metrics." In: *Business Process Management. BPM 2007. Lecture Notes in Computer Science*. Ed. by G. Alonso, P. Dadam and M. Rosemann. Berlin: Springer, p. 328.

Hacker, W. (2001). "Psychological Activity Theory." In: *International Encyclopedia of Social & Behavioral Sciences*. Ed. by N. J. Smelser and P. B. Baltes. Oxford: Elsevier Science, p. 58.

Han, J., M. Kamber and J. Pei (2011). *Data Mining: Concepts and Techniques*. 3rd edition. Amsterdam: Elsevier.

Heinonen, K., T. Strandvik, K. J. Mickelsson, B. Edvardsson, E. Sundström and P. Andersson (2010). "A customer-dominant logic of service." *Journal of Service Management* 21 (4), 531–548.

Hevner, A. and S. Chatterjee (2010). "Design science research in information systems." In: *Design Research in Information Systems. Integrated Series in Information Systems*. Ed. by A. Hevner and S. Chatterjee. Boston MA: Springer, p. 9.

Hevner, A., S. T. March, J. Park and S. Ram (2004). "Design science in information systems research." *MIS Quarterly* 28 (1), 75–105.

Hompes, B. F. A., H. M. W. Verbeek and W. M. P. Van Der Aalst (2014). "Finding suitable activity clusters for decomposed process discovery." In: *International Symposium on Data-Driven Process Discovery and Analysis*. Ed. by P. Ceravolo, B. Russo and R. Accorsi. Cham: Springer, p. 32.

Howard, J. A. and J. N. Sheth (1969). *The Theory of Buyer Behavior*. New York: Wiley.

Ives, B. and G. P. Learmonth (1984). "The information system as a competitive weapon." *Communications of the ACM* 27 (12), 1193–1201.

Jaillet, H. F. (2003). "Web metrics: Measuring patterns in online shopping." *Journal of Consumer Behaviour* 2 (4), 369–381.

Jenamani, M., P. K. Mohapatra and S. Ghose (2003). "A stochastic model of e-customer behaviour." *Electronic Commerce Research and Applications* 2 (1), 81–94.

Johnson, E. J. and J. E. Russo (1984). "Product familiarity and learning new information." *Journal of Consumer Research* 11 (1), 542–550.

Kamakura, W. A., S. N. Ramaswami and R. K. Srivastava (1991). "Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services." *International Journal of Research in Marketing* 8 (4), 329–349.

Koschmider, A. (2017). "Clustering event traces by behavioral similarity." In: *Advances in Conceptual Modeling*. Ed. by S. De Cesare and U. Frank. Cham: Springer, p. 36.

Kruskal, W. H. and W. A. Wallis, (1952). "Use of ranks in one-criterion variance analysis." *Journal of the American Statistical Association* 47 (260), 583–621.

Kuutti, K. (1996). "Activity theory as a potential framework for human–computer interaction research." In: *Context and Consciousness: Activity Theory and Human-Computer Interaction*. Ed. by B. A. Nardi. Cambridge MA: MIT Press, p. 17.

Le, B., H. Duong, T. Truong and P. Fournier-Viger (2017). "FCloSM, FGenSM: Two efficient algorithms for mining frequent closed and generator sequences using the local pruning strategy." *Knowledge and Information Systems* 53 (1), 1–37.

Li, J., R. J. C. Bose and W. M. P. Van Der Aalst (2010). "Mining context-dependent and interactive business process maps using execution patterns." In: *Business Process Management Workshops. BPM 2010. Lecture Notes in Business Information Processing*. Ed. by M. Zur Muehlen and J. Su. Berlin: Springer, p. 109.

Li, J., P. Ren, Z. Chen, Z. Ren, T. Lian and J. Ma (2017). "Neural attentive session-based recommendation." In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. Ed. by E.-P. Lim and M. Winslett. New York NY: ACM, p. 1419.

Liu, Q., S. Wu and L. Wang (2017). "Multi-behavioral sequential prediction with recurrent log-bilinear model." *IEEE Transactions on Knowledge and Data Engineering* 29 (6), 1254–1267.

Lunn, J. A. (1974). "Consumer decision-process models." In: *Models of Buyer Behaviour*. Ed. by J. Sheth. New York NY: Harper & Row, p. 34.

Mans, R. S., M. H. Schonenberg, M. Song, W. M. P. Van Der Aalst and P. J. Bakker (2008). "Application of process mining in healthcare–a case study in a dutch hospital." In: *Biomedical Engineering Systems and Technologies*. Ed. by A. Fred, J. Filipe and H. Gamboa. Berlin: Springer, p. 425.

Martin, D. and A. G. Woodside (2012). "Structure and process modeling of seemingly unstructured leisure-travel decisions and behaviour." *International Journal of Contemporary Hospitality Management* 24 (6), 855–872.

Mobasher, B., H. Dai, T. Luo and M. Nakagawa (2002). "Using sequential and non-sequential patterns in predictive web usage mining tasks." In: *Proceedings 2002 IEEE International Conference on Data Mining*. Ed. by V. Kumar, S. Tsumoto, P. Yu and N. Zhong. Washington DC: IEEE, p. 669.

Nguyen, P., A. Slominski, V. Muthusamy, V. Ishakian and K. Nahrstedt (2016). "Process trace clustering: A heterogeneous information network approach." In: *Proceedings of the 2016 SIAM International Conference on Data Mining*. Ed. by S. Chawla and W. Meira. Philadelphia PA: Society for Industrial and Applied Mathematics, p. 279.

Nicosia, F. M. (1966). *Consumer Decision Processes: Marketing and Advertising Implications*. Englewood Cliffs NJ: Prentice-Hall.

Nussbaumer, P., I. Matter, I. Slembek and G. Schwabe (2011). "Information search behavior of investors and the role of advisory services." In: *ECIS 2011 Proceedings*. Ed. by V. Tuunainen, M. Rossi and J. Nandhakumar. Atlanta GA: AIS, p. 46.

Paas, L. J., T. H. Bijmolt and J. K. Vermunt (2007). "Acquisition patterns of financial products: A longitudinal investigation." *Journal of Economic Psychology* 28 (2), 229–241.

Pan, B. and D. R. Fesenmaier (2006). "Online information search: vacation planning process." *Annals of Tourism Research* 33 (3), 809–832.

Park, C. W. and V. P. Lessig (1981). “Familiarity and its impact on consumer decision biases and heuristics.” *Journal of Consumer Research* 8 (2), 223–231.

Park, C. W. and B. J. Moon (2003). “The relationship between product involvement and product knowledge: Moderating roles of product type and product knowledge type.” *Psychology & Marketing* 20 (11), 977–997.

Payne, A. F., K. Storbacka and P. Frow (2008). “Managing the co-creation of value.” *Journal of the Academy of Marketing Science* 36 (1), 83–96.

Prinzie, A. and D. Van Den Poel (2011). “Modeling complex longitudinal consumer behavior with Dynamic Bayesian networks: an Acquisition Pattern Analysis application.” *Journal of Intelligent Information Systems* 36 (3), 283–304.

Richetti, P. H. P., F. A. Baião and F. M. Santoro (2014). “Declarative process mining: reducing discovered models complexity by pre-processing event logs.” In: *Business Process Management. BPM 2014. Lecture Notes in Computer Science*. Ed. by S. Sadiq, P. Soffer and H. Völzer. Cham: Springer, p. 400.

Rozsnyai, S., A. Slominski and G. T. Lakshmanan (2011). “Discovering event correlation rules for semi-structured business processes.” In: *Proceedings of the 5th ACM International Conference on Distributed Event-Based System*. Ed. by D. Eyers. New York NY: ACM, p. 75.

Salazar, M. T., T. Harrison and J. Ansell (2007). “An approach for the identification of cross-sell and up-sell opportunities using a financial services customer database.” *Journal of Financial Services Marketing* 12 (2), 115–131.

Salehi, M. (2013). “An effective recommendation based on user behaviour: a hybrid of sequential pattern of user and attributes of product.” *International Journal of Business Information Systems* 14 (4), 480–496.

Sands, S., C. Ferraro, C. Campbell and J. Pallant (2016). “Segmenting multichannel consumers across search, purchase and after-sales.” *Journal of Retailing and Consumer Services* 33, 62–71.

Savolainen, R. (2009). “Interpreting informational cues: An explorative study on information use among prospective homebuyers.” *Journal of the American Society for Information Science and Technology* 60 (11), 2244–2254.

Schmidt-Rauch, S. and G. Schwabe (2014). “Designing for mobile value co-creation – the case of travel counselling.” *Electronic Markets* 24 (1), 5–17.

Seret, A., S. K. L. M. Vanden Broucke, B. Baesens and J. Vanthienen (2014). “A dynamic understanding of customer behavior processes based on clustering and sequence mining.” *Expert Systems with Applications* 41 (10), 4648–4657.

Smirnov, S., H. A. Reijers and M. Weske (2012). “From fine-grained to abstract process models: A semantic approach.” *Information Systems* 37 (8), 784–797.

Song, M., C. W. Günther and W. M. P. Van der Aalst (2008). “Trace clustering in process mining.” In: *Business Process Management Workshops. BPM 2008. Lecture Notes in Business Information Processing.* Ed. by M. Dumas, M. Reichert and M.-C. Shan. Berlin: Springer, p. 109.

Spiliopoulou, M. and C. Pohle (2001). “Data mining for measuring and improving the success of web sites.” *Data Mining and Knowledge Discovery* 5 (1–2), 85–114.

Swinnen, J., B. Depaire, M. J. Jans and K. Vanhoof (2012). “A process deviation analysis – A case study.” In: *Business Process Management Workshops. BPM 2011. Lecture Notes in Business Information Processing.* Ed. by F. Daniel, K. Barkaoui and S. Dustdar. Berlin: Springer, p. 87.

Tax, N., N. Sidorova, R. Haakma and W. M. P. Van Der Aalst (2016). “Mining local process models.” *Journal of Innovation in Digital Ecosystems* 3 (2), 183–196.

Van der Aalst, W. M. P. and M. Song (2004). “Mining social networks: Uncovering interaction patterns in business processes.” In: *Business Process Management. BPM 2004. Lecture Notes in Computer Science*. Ed. by J. Desel, B. Pernici and M. Weske. Berlin: Springer, p. 244.

Van Der Aalst, W. M. P. (2011). *Process Mining – Discovery, Conformance and Enhancement of Business Processes*. Berlin: Springer.

Van Dongen, B. F. and A. Adriansyah (2009). "Process mining: fuzzy clustering and performance visualization." In: *Business Process Management Workshops. BPM 2009. Lecture Notes in Business Information Processing.* Ed. by S. Rinderle-Ma, S. Sadiq, F. Leymann. Berlin: Springer, p. 158.

Van Dongen, B. F., A. K. A. de Medeiros, H. M. W. Verbeek, A. J. M. M. Weijters and W. M. P. Van Der Aalst (2005). "The ProM framework: A new era in process mining tool support." In: *Applications and Theory of Petri Nets*. Ed. by G. Ciardo and P. Darondeau. Berlin: Springer, p. 444.

Van Rijnsoever, F. J., C. Castaldi and M. J. Dijst (2012). "In what sequence are information sources consulted by involved consumers? The case of automobile pre-purchase search." *Journal of Retailing and Consumer Services* 19 (3), 343–352.

Vandermerwe, S. (2000). "How increasing value to customers improves business results." *Sloan Management Review* 42 (1), 27–38.

Venable, J. (2006). "A framework for design science research activities." In: *Emerging Trends And Challenges in Information Technology Management*. Ed. by M. Khosrow-Pour. Washington DC: Idea Group Publishing, p. 21.

Verbeek, H. M. W. and W. M. P. Van Der Aalst (2014). "Decomposed process mining: The ILP case." In: *Business Process Management Workshops. BPM 2014. Lecture Notes in Business Information Processing.* Ed. by F. Fournier and J. Mendling. Cham: Springer, p. 264.

Verbeek, H. M. W., W. M. P. Van Der Aalst and J. Munoz-Gama (2016). "Divide and conquer." In: *Technical report, BPM Center Report Series.* BPM Center, p.1.

Voorveld, H. A., E. G. Smit, P. C. Neijens and A. F. Bronner (2016). "Consumers' cross-channel use in online and offline purchases". *Journal of Advertising Research* 56 (4), 385–400.

Vroomen, B., B., Donkers, P. C. Verhoef and P. H. Franses (2005). "Selecting profitable customers for complex services on the internet." *Journal of Service Research* 8 (1), 37–47.

Wang, H. and S. Wang (2007). "Mining purchasing sequence data for online customer segmentation." *International Journal of Services Operations and Informatics* 2 (4), 382–390.

Wei, J., Z. Shen, N. Sundaresan and K. L. Ma (2012). "Visual cluster exploration of web clickstream data." In: *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*. Ed. by G. Santucci and M. Ward. Seattle WA: IEEE, p. 3.

Yahya, B. N., M. Song, H. Bae, S. O. Sul and J. Z. Wu (2016). "Domain-driven actionable process model discovery." *Computers & Industrial Engineering* 99, 382–400.