Predicting Interpurchase Time in a Retail Environment using Customer-Product Networks: An Empirical Study and Evaluation

Jasmien Lismont[a], Sudha Ram[b], Jan Vanthienen[a], Wilfried Lemahieu[a], Bart Baesens[a,c,*]

[a] *KU Leuven, Dept. of Decision Sciences and Information Management, Naamsestraat 69, B-3000 Leuven, Belgium*
[b] *University of Arizona, Eller College of Management, Department of MIS, 430J McClelland Hall, Tucson, AZ 85721, USA*
[c] *University of Southampton, Southampton Business School, Highfield, Southampton SO17 1BJ, United Kingdom*

Abstract

In predictive analytics and statistics, entities are frequently treated as individual actors. However, in reality this assumption is not valid. In the context of retail, similar customers will behave and thus also purchase similarly to each other. By combining their behavior in an intelligent way, based on transaction history, we can leverage these connections and improve our ability to predict purchase outcomes. As such, we can create customer-product networks from which we can deduce information on customers expressing similar purchasing behavior. This allows us to exploit their preferences and predict which products are going to be sold significantly less often. We want to use this information mainly for gaining novel marketing insights on products. For example, if customers refrain from buying products this might be due to contextual reasons such as new complements or supplements, or new nearby shops. By using these networks on data from an offline European retail corporation, we are able to boost performance of the predictive models by 6% and the identification of these specific products by

*Corresponding author
Email addresses: Jasmien.Lismont@kuleuven.be (Jasmien Lismont), ram@eller.arizona.edu (Sudha Ram), Jan.Vanthienen@kuleuven.be (Jan Vanthienen), Wilfried.Lemahieu@kuleuven.be (Wilfried Lemahieu), Bart.Baesens@kuleuven.be (Bart Baesens)

20%. This indicates that the development of customer-product graphs in retail can lead to improved marketing intelligence. To our knowledge, this is one of the first studies to use customer-product networks for prediction modeling in an offline retail setting. Furthermore, we suggest an extensive set of product and network features which can guide future researchers and practitioners in their model development.

Keywords: customer-product graph, interpurchase time, offline retail, purchase behavior, social network analytics, transactional data

---

1. Introduction

Increasingly, marketing applications, such as product recommendations and online advertising are based on leveraging large amounts of data from various sources. Retailers often exploit links between customers to improve their targeting actions, i.e. (social) network-based marketing (Hill et al., 2006). For example, if a person clicks on an advertisement on social media, her friends may also be more inclined to do so. This is based on the concept of homophily (McPherson et al., 2001), where similar people tend to connect with each other. There are also studies where retailers, such as Amazon, target customers based on their similarities in purchase behavior, i.e. instead of using a friendship or social network they construct networks, also called graphs, of customers based on product purchases. These 'pseudo-social' networks (Martens et al., 2016) are used to recommend other products to specific individuals.

In this research, our goal is to construct such a network and use it to predict significant increase in interpurchase time (IPT), i.e. the frequency with which specific products will be purchased. We construct a customer-product network based on transaction data, in which products are connected with the customers who have purchased them. Knowing the interpurchase time for products can help us understand which products are likely to be sold in lower quantities in a specific time period and is of particular interest to grocery chains. In the grocery industry, there is often considerable variation between products in

terms of the repetitiveness of their purchases and, as such, repetition of consumer choices cannot be assumed (Adamowicz & Swait, 2013). This variance in repetitiveness may be attributed to customer habits, variety-seeking behavior of customers, novelty of products or utility maximization. For more details on consumer decision-making strategies, the reader is referred to Adamowicz & Swait (2013). Marketing decision makers can use the predicted IPT to boost marketing actions of certain products, make store-layout decisions or for planning purposes. Additionally, this information can be used for supply chain and logistics decisions.

Using customer-product networks may help overcome some limitations of friendship or social networks. For instance, Aral et al. (2009) have shown that a large portion of observed correlation in product adoption in social networks is actually due to similarity of people rather than social influence. They claim that actual social influence should not be overestimated.

We make several contributions to network-based analytics in this study. Firstly, we introduce the application of customer-product networks in predictive modeling. Given that prior work in this domain was only descriptive, we produce novel results in the context of network analytics for generating insights in retail. We introduce a set of network features which, when combined with product characteristics, are helpful for predicting IPT. These features contain more simple unipartite characteristics, as well more complex and detailed bipartite characteristics, and centrality measures. Furthermore, we discuss which of these product and network characteristics contribute the most to the predictions. Second, we are able to improve performance of our prediction significantly in terms of area under the receiver operating characteristic curve by almost 6%. In addition, our models are better capable in identifying products of interest with increased sensitivity scores of almost 20%. This means that our model can be used in practice by retailers to identify the top-n most relevant products for marketing campaigns. These results are based on real-life data from a European retailer, multiple analytics techniques and cross-validation of the results, leading to more robust and reliable conclusions. Our model and final insights

can then be used by retailers as an expert tool in their decision-making. Finally, our results also demonstrate the usefulness of transaction data in addition to product and customer data in a practical offline retail setting.

In the next section, we first discuss related research. Then, in Section 3, we discuss our methodology. Section 4 presents the results and discusses the findings. We also cover the limitations of this study and propose future opportunities. Finally, in Section 5, we conclude this paper.

2. Related research

Although we do not work with friendship or social networks, i.e. the products do not 'know' each other and the customers do not necessarily have a relationship in real life, we employ (social) network analytics (SNA) techniques. In this context, Hill et al. (2006) built consumer networks using direct interactions. Although they do work with actual social networks, their work proves that being connected in such a network can directly affect product adoption. 'Pseudo-social' networks have been proposed and used before in other studies. Martens et al. (2016) introduce the concept of a pseudo-social network in a financial context. They also use transaction data to create a customer network in which customers are linked based on their purchase history to target specific customers for different products. For the purpose of brand advertising, Zhang et al. (2016) build a network of interactions between people on the social media platform Facebook, even though they are not friends.

Also for products, network-based analytics has been applied in previous research. In general, there are two types of networks which can be constructed, namely market basket graphs and customer-product graphs, although no consensus about terminology currently exists. Firstly, market basket graphs connect products which are sold together at the same time, by the same customer (Kim et al., 2012). For example, Dhar et al. (2014) aim to predict future demand of an online retailer by developing a co-purchase network of products which are frequently purchased together. They claim that co-purchase networks, to which

they also refer as 'product networks' or 'economic networks', reflect an aggregation of preferences of people external to the network. As such, they are able to catch smoothed trends. Similarly, Raeder & Chawla (2009) and Videla-Cavieres & Ríos (2014) create a market basket network for the purpose of frequent item set mining. They are able to perform a market basket analysis by means of community detection techniques. Ríos & Videla-Cavieres (2014) extend this work by introducing time dynamics in order to study the stability of the communities. Secondly, there has been some prior work on customer-product networks, the topic of this paper. These networks connect products which are purchased together by the same customer, regardless the timing (Kim et al., 2012). Huang et al. (2007) study this type of networks in order to gain more insights into consumer purchase behavior in an e-commerce setting. In comparison to standard random graphs, they find that average path lengths are larger than expected and the tendency to clustering is higher. This suggests that customers' product choices are not random, which justifies further research. Kim et al. (2012) also study customer-product graphs in a more descriptive manner. They claim that this type of network is more sensitive to customer preferences compared to market basket networks. The choice for one type of network should thus be mainly driven by the application. Market basket networks could be more interesting for frequent item set mining or recommender systems. Customer-product graphs, on the other hand, are interesting for capturing customer preferences, for example for classification models and studying purchase behavior dynamics.

Furthermore, some work has been done on network-based, predictive classification problems. Specifically, we can link our work to the concept of (partial) churn. In the context of customer churn, e.g. in the telecommunications sector, marketing aims to identify those customers who are (gradually) ceasing their purchases of products or services at a company. In our work, we categorize products into two groups and attempt to identify those products which experience a significant increase in IPT. These products are thus facing partial and involuntary passive churn. Therefore, previous work on customer churn can be of particular interest to our study. Verbeke et al. (2014) developed a

network of customers based on call details and study how the spread of churn through this network can be used in a predictive manner. They, furthermore, discuss how network characteristics can be incorporated in classification models, i.e. by building a relational classifier or by featurizing the network and adding these variables to a non-relational learner. Óskarsdóttir et al. (2017) compare several SNA methods for churn prediction in telecommunications. They emphasize the importance of edge and weight definitions in this context. Moreover, their best-performing model was a non-relational classifier enriched with network variables. Benoit & Van den Poel (2012) use kinship network data to improve a churn prediction model in the financial services industry. They show that network variables increase model performance and often prove to be more impactful than local variables not containing network effects.

Furthermore, our work is linked to the domains of market basket analysis, recommender systems and forecasting, although it is different in a number of ways. Firstly, market basket analysis zooms in on products which are frequently bought together, by the same customer at the same time, i.e. market basket networks. Our network, on the other hand, is a customer-product network which only focuses on products bought together by the same customer and not necessarily at the same time (Kim et al., 2012). Secondly, our work differs from recommender systems in two ways. The goal is inherently different since recommender systems aim to offer several items to customers while we focus on particular products and their dynamics. Moreover, collaborative filtering is focused on finding similar customers and uses the purchase preferences of these customers to make recommendations (Adomavicius & Tuzhilin, 2005). We, on the other hand, focus on products and finding similar products based on customer purchase behavior. Lastly, the concept of product attrition is linked to sales or demand forecasting. However, our methods are focused on classification which differs from the techniques in the forecasting domain. Moreover, the goals are also different since our paper aims mainly to provide new insights regarding the interpurchase time for specific products.

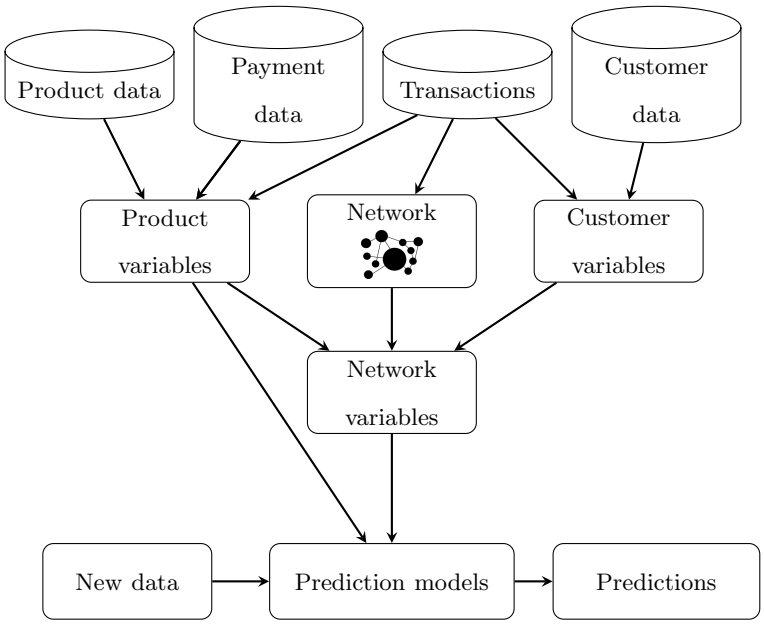

Figure 1: Methodology.

3. Methodology and data

This section describes the data and explains how we apply analytics techniques in order to predict which products will be sold significantly less often. Both product characteristics and network features will be taken into account. The former are also referred to as local features, since they do not take the network into account. Figure 1 illustrates an overview of the methodology.

3.1. Data from an offline food retailer

Our methodology is applied on a dataset from a low-cost European retailer specialized in grocery shopping. They provide both food (e.g. butchery, fruits and vegetables, dairy, beverages, snacks) and non-food (e.g. health and beauty, household, pets supplies) products. They are mainly active in Belgium where they have a 31.5% market share (2015/2016) and 234 retail stores. We collect data from 30 stores of which two are pick-up points for online ordered goods. After the removal of outliers and seasonal products, we are left with 6,355 prod-

ucts and 406,678 customers with a total of more than 100 million transactions over a period of 12 months. Here, a customer is a household and may be a single person, couple or family all living together at one address. It is to be noted that the data is anonymized so the identity of the customer is not revealed.

3.2. Variable extraction

In this section, we explain how we define the outcome variable, namely product attrition. Next, we identify and justify features that are used for the prediction including product characteristics and features extracted from the network of products. These features are summarized in Table A.1, in Appendix A.

3.2.1. Product attrition

We define product attrition based on IPT. A product is undergoing attrition if its purchase frequency decreases significantly over time. Product attrition can be regarded as partial passive, involuntary churn since customers are ceasing their purchases for this particular product. In order to determine IPT, we calculate the number of days in between product purchases by a specific customer during a specific time period of five months. Consecutively, we take the average per month. For example, a product that is bought every day in September, has an average IPT of 0. Next, we calculate the slope of the IPT throughout the five consecutive months. If this slope is positive, thus the IPT is increasing, and this slope is significant with a confidence level of 90%, we identify this as product attrition. This means that we subdivide products into two categories: those products which experience a significant increase in the IPT slope and those which do not. This allows us to apply predictive, binary classification techniques. Note that we do not take products into account which are not sold on a regular basis, since these are probably seasonal or temporal products, and thus not of interest to the retailer.

Our study is not the first to use the change in a particular value as a dependent variable. Baesens et al. (2004), for example, use the slope of customer

spending to define passive customer churn. Miguéis et al. (2012), on the other hand, also focus on partial churn, namely of customers, and recognize that churn in retail occurs less abrupt. Moreover, in prior work, IPT has proven to be a valuable feature. Benoit & Van den Poel (2012) illustrate the importance of interpurchase time for predicting customer churn in the financial services industry. In the context of customer lifetime value (CLV) measurement, Borle et al. (2008) found that longer IPTs are associated with a greater risk of leaving the firm, while Kumar et al. (2004) claim that more stable IPTs are linked to more durable and profitable customer relationships. Moreover, IPT has been included in multiple CLV models (Borle et al., 2008; Kumar et al., 2004; Venkatesan & Kumar, 2004) and in predictive models, e.g. for the purpose of customer loyalty prediction (Buckinx et al., 2007).

3.2.2. The collection of product characteristics

Product characteristics are extracted from product, transaction and payment datasets. Each transaction line is linked to a specific product. With regards to payment information, we can link one or more payment details to a transaction. For example, a customer may have paid the largest part of a particular transaction with foodstamps, and the remainder with debit card.

Firstly, we extract recency, frequency and monetary (RFM) features, which are popular measures in marketing literature and practice (Buckinx & Van den Poel, 2005) and a well-known customer value analysis method ((Kaymak, 2001) in (Cheng & Chen, 2009)). We chose not to aggregate these values but rather treat each value as a separate feature. We follow the definition of Cheng & Chen (2009), and discretize each feature into five quintiles. The higher the value, the more recent, frequent or profitable the product is. Next, we take the average and standard deviation of the interpurchase time (in days) into account. Additionally, we compose a feature 'regularity', similar to Buckinx & Van den Poel (2005), which expresses the stability of IPT and is defined as the standard deviation divided by the average. The number of unique stores in which the product is sold, the number of unique customers of the product, how often it

was in promotion, its mean price and its mean discount are also calculated based upon Baesens et al. (2004) and Buckinx & Van den Poel (2005). Each of these characteristics give information about the likelihood that customers buy the product. For example, discounts can significantly influence the shopping behavior of customers (Walters, 1991). Finally, we include payment data, similar to Buckinx & Van den Poel (2005), and calculate how frequently a product is bought by means of cash, card, foodstamps, the retailer's debit card, or by mobile phone. How customers paid for a specific shopping basket can reflect customer, timing, store, purpose and basket characteristics, and is therefore valuable to include. For example, we note that young adults (18 up to 34 years) use on average significantly more their electronic card compared to middle group (35 up to 54 years) customers and 55-plus customers. On the other hand we note that 55-plus customers use significantly more often the retailer's card or cash to pay. Furthermore, we observe significant positive correlations of 0.67 and 0.12 (p-value < 0.0001) between the percentage a customer paid with the retailer's debit card and their total number of visits and total spending during the observed year respectively. Additionally, a positive correlation between the average number of distinct products a customer purchases during a shopping trip and how they paid, can be discovered. Paying by means of an electronic card has a significant positive correlation of 0.11 (p-value < 0.0001). The same can be said about the retailer's card and foodstamps, although these correlations are smaller, i.e. 0.023 and 0.015 respectively. Paying in cash, on the other hand, has a significant negative correlation of 0.15 (p-value < 0.0001).

Since we measure these variables across a time span of five months, some data might be lost due to aggregation. For this reason, we would like to incorporate the dynamics of the previously introduced product variables during this time span. As such, we create an additional variable for each characteristic which represents the slope of the change in the variable. For example, if a product is bought 100 times each month, its frequency slope will be 0. Moreover, we cap outliers at the 5^{th} and 95^{th} percentile and standardize each variable.

Finally, we also include the time period as a categorical variable in our

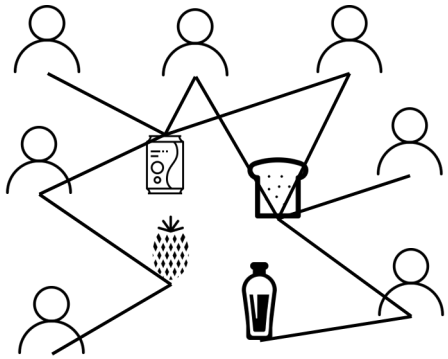

Figure 2: A small example of how a bipartite customer-product graph can be visualized. Products are connected if they are bought by the same customer at the same store.

models.

3.2.3. Building a customer-product network

We create a network of products which are connected if they are bought by the same customer at the same store during the specified time period. These connections are also called edges and the objects they connect are called nodes. In this setting, we can distinguish between two types of graphs. Firstly, we have bipartite graphs which exist out of two types of nodes. A small example is visualized in Figure 2. The nodes are products and the customers who buy them. However, we can also map this bipartite graph onto a unipartite graph existing only out of product nodes. Then, products are directly linked to other products by means of edges if they have customers in common.

3.2.4. Extracting network characteristics

There are, in general, two methods to incorporate network effects in our models. This is illustrated nicely in the work of Verbeke et al. (2014, p.432). We can either extract features from the network, this process is called featurization or propositionalization (Kramer et al., 2001), or we can implement a relational learner (RL). In addition, a RL can be complemented by a collective inferencing component (Macskassy & Provost, 2007). The advantage of a RL is that it entails all network information and details. However, the featurization

technique allows us to still apply well-known, proven and robust non-relational classifiers. Moreover, it allows for analyzing the effect of both local and network features alongside each other. In addition, good results were obtained using this technique in previous studies in churn prediction (Óskarsdóttir et al., 2017; Verbeke et al., 2014). Therefore, we opt for featurization. We distinguish between four types of network features.

(1) Firstly, we apply an adaption (Van Vlasselaer et al., 2017) of Google's PageRank algorithm (Page et al., 1998) as can be observed in Equation 1. The PageRank algorithm has been applied in a similar context by Dhar et al. (2014). However, this algorithm focuses on unipartite graphs, whereby we would lose the information of the customers. Therefore, we implement an adapted version which allows us to work with both the product and customer characteristics of the bipartite graph. We set the number of iterations k to 100 as suggested by Van Vlasselaer et al. (2017) and the damping factor α to 0.85 based on Page et al. (1998). In Equation 1, $\overrightarrow{z}_{norm}$ represents the normalized degree-adapted restart vector and ξ_k the exposure scores after k iterations. Note that the final exposure scores are independent of the initial ξ_0 (Page, 2001; Van Vlasselaer et al., 2017). However, we make slight changes to the restart vector in order to be able to calculate the exposure to each of the RFM values, a popular metric in retail. We propose three starting vectors. Each vector contains the respective product and customer values for each of the three RFM variables. This leads to three PageRank exposure scores for recency, frequency and monetary.

$$\xi_{k+1} = \alpha \cdot \xi_k + (1 - \alpha) \cdot \overrightarrow{z}_{norm} \tag{1}$$

(2) Secondly, we calculate edge attributes in the bipartite graph. We take into account the connections a product has. Then, we adhere three types of weights to the edges. We determine the recency, frequency and monetary value for each specific product-customer connection, which we again discretize into five quintiles. Finally, we take the average edge weights for each of the RFM values for each product. By including edge attributes, we are able to focus on specific customer-product behavior in a particular store. For example, if there

would a new bakery in town, the purchase behavior of a local inhabitant for bread can change and impact her IPT.

(3) Thirdly, we calculate network features based on the direct neighborhood of products in the unipartite graph. In general, the direct neighborhood already contains a lot of information (Macskassy & Provost, 2007; Neville & Jensen, 2007; Óskarsdóttir et al., 2017; Verbeke et al., 2014) which will allow us to deduce relevant features. As such, we measure the number of first-order neighbors, namely products who share customers with the product of interest. Furthermore, we take into account how many of these neighbors belong (percentage-wise) to the same category as the product of interest. This is considered for three hierarchical categorical levels. For example, cereals and muesli belong to the same group breakfast. It is common that marketing applies product hierarchies to describe their goods, for store-lay-outs and for visualizing advertisements. For example, Kim et al. (2012) explicitly take advantage of a product taxonomy in their customer-product graphs. In addition, we compute the general number of passively churning neighbors and repeat this to detect attrition among neighboring products belonging to the same category.

(4) Lastly, we calculate centrality measures on the unipartite graphs. This type of measures allows us to estimate how a product is situated in the network. We measure the degree, closeness, betweenness and finally apply the personalized PageRank algorithm of Google (Page et al., 1998) and a local clustering coefficient. Degree, closeness, and betweenness are common centrality measures (Sun & Tang, 2011). Degree is similar to the number of neighboring customers with the only difference that we normalize the degree by the number of nodes. Closeness takes into account how many steps are necessary to reach every other product in the network, which is also normalized by the number of nodes (Freeman, 1978). Betweenness expresses the number of shortest paths going through a specific product node and is again normalized (Brandes, 2001; Freeman, 1978). All centrality measures are determined using the implementation of Csardi & Nepusz (2006). The final two measures are recommended by Dhar et al. (2014) in a similar context. For calculating the PageRank, we apply the "prpack"

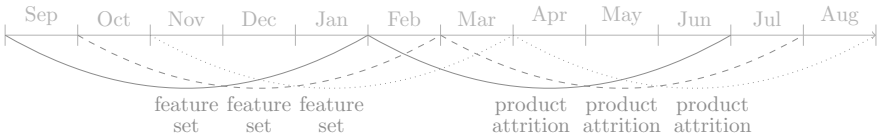

Figure 3: Sampling of data for the extraction of the feature set and outcome variable.

algorithm (Csardi & Nepusz, 2006), we set the damping factor again to 0.85, and we personalize the algorithm using product attrition of the current time period. Next, the local clustering coefficient is defined as the ratio of triangles connected to the node and the triples centered on the node (Csardi & Nepusz, 2006). Dhar et al. (2014) provide us with two good reasons to include this clustering measure. Firstly, it gives an indication of customer behavior, in the sense that it influences the path potential customers follow going from one product to another. Secondly, it suggests groups of more closely linked products.

3.3. Model building

We split our data in three samples of each five months in order to extract the features. Consecutively, we use the next five months, as depicted in Figure 3, to determine product attrition. The duration of the time period, i.e. five months, was chosen to be significantly long in order to discover a product trend and, at the same time, allow for comparison of different time periods bounded by the availability of the sample dataset. This is in line with previous literature (Baesens et al., 2004; Buckinx & Van den Poel, 2005; Miguéis et al., 2012) applying slopes in customer analysis which use three, five or six months of data.

We create three models, one with only local, product features; one with only network features; and one with a combination of local and network features. We will refer to these model as the local, network and hybrid model. Using repeated five-fold cross-validation, we train and validate our model on 80% of the dataset, and consecutively test our model on 20% of the remaining observations. We measure performance as the area under the receiver operating characteristic (ROC) curve (AUC), and sensitivity and specificity with a cut-off rate of 50% and a cut-off rate following the actual attrition rate of 14%. ROC curves display

the sensitivity versus the specificity. The closer this curve is to the top left, and thus the higher the AUC, the better the model is able to distinguish between the products which experience attrition and those which do not. Thus, AUC will vary between 0 and 1 with a value of 0.50 representing random classification performance and a value of 1 representing a perfect model. Sensitivity or the true positive rate, measures the percentage of actual product attrition predicted as such and ranges from 0 to 1. Specificity or the true negative rate, on the other hand, measures the percentage of actual non-attrition identified as such and ranges from 1 to 0. Note that the maximum and best value for both sensitivity and specificity is 1. This cross-validation process is repeated ten times after which we calculate the average in order to report a more stable result.

We construct each model using four popular data analytics techniques, namely logistic regression, decision trees, random forests, and neural networks.

(1) Logistic regression is perhaps one of the most well-known techniques. This technique allows for easy deduction of variable importance. In the context of this paper, this adds the advantage of distinguishing between the importance of product and network characteristics. Before applying the technique, we performed variable selection based on their importance in a random forest model, using a hold-out validation set containing 20% of the dataset. Importance is calculated as the mean decrease in node impurity, measured by the Gini index, if that particular variable would be removed from the variable set. Accordingly, we perform forward feature selection by removing highly correlated ($> 50\%$) variables.

(2) Next, we apply decision trees using the C5.0 algorithm implemented by Kuhn et al. (2014). This algorithm is an extension of the C4.5 algorithm (Kuhn et al., 2014; Quinlan, 1993). We tune a binary Boolean parameter which determines whether feature selection should be used by means of four-fold cross-validation on the training and validation set.

(3) Additionally, we apply random forests (Breiman, 2001), an ensemble technique which constructs multiple decision trees and combines them into one model. It has been shown that this technique can achieve superior performance

compared to other techniques (Lessmann et al., 2015). For this purpose, we follow the implementation of Apache's H2O math engine (The H2O.ai team, 2017). We optimize the number of trees in the random forest model by means of grid search on a hold-out validation set consisting of 20% of the original dataset, and set it to an odd number in order to improve tie-breaking.

(4) For the reason of providing a more exhaustive overview, we also apply a classic neural network. In this technique, the number of neurons in the hidden layers (for simplicity set to one hidden layer) is important. Therefore we tune this parameter similar to how we tune the number of trees in the random forests. The algorithm itself is an implementation of Fritsch et al. (2016) using resilient backpropagation with weight backtracking (Riedmiller, 1994).

4. Results and discussion

4.1. Comparison in performance of prediction models

We build the models using the four techniques mentioned in the previous section, namely logistic regression, decision trees, random forests and neural networks. Firstly, we want to compare the different techniques. As illustrated in Figure 4, random forests outperform all other techniques while decision trees clearly underperform. This might be due to the fact that random forests are able to catch interaction effects. Moreover, random forests is an ensemble technique which frequently offers increased performance (Lessmann et al., 2015). Therefore, our main focus will be on random forests.

Table 1 provides the average AUCs of all models in their local, network and hybrid variant. By means of the test of DeLong et al. (1988), we compare per iteration whether the values differ significantly and provide the average. Additionally, we perform a paired non-parametric statistical test, namely Wilcoxon signed-rank test. The results are shown in Table 2. Note that by applying four analytics techniques, a repeated five-fold cross-validation, and providing both a test for comparing individual ROCs and a non-parametric statistical hypothesis test, the reliability and validity of our findings are demonstrated in a robust

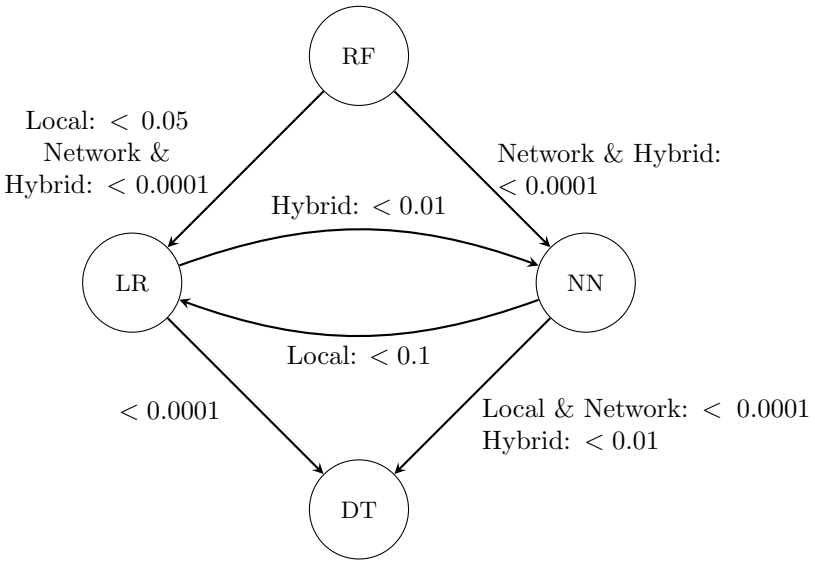

Figure 4: Domination graph of logistic regression (LR), decision tree (DT), neural network (NN) and random forest (RF) models. AUCs are compared using the Mann-Whitney U test.

manner. We observe that for all techniques except the neural networks, the network and hybrid model achieve better performance than the local model. When we zoom in on the random forests, we even observe a significant average difference of 4 percentage points (i.e. a 6% performance increase) by including network variables. This suggests that customer-product networks indeed capture customer preferences and behavior, and that we can transfer this in order to gain new insights into product attrition.

In addition, Figure 5 illustrates the median ROC curves of the local, network,

Table 1: Performance of the different techniques in terms of AUC. Standard deviations of the models are relatively small and indicated in parentheses.

	Local (σ)	**Network (σ)**	**Hybrid (σ)**
Logistic regression	0.6503 (0.0248)	0.6521 (0.0183)	0.6623 (0.0217)
Decision tree	0.5587 (0.0648)	0.6055 (0.0600)	0.6130 (0.0536)
Random forest	0.6616 (0.0204)	0.6902 (0.0229)	0.6989 (0.0213)
Neural network	0.6595 (0.0239)	0.6512 (0.0246)	0.6462 (0.0271)

Table 2: Difference in AUC performance between the local, network and hybrid model. (1) Average p-values determined by the test of (DeLong et al., 1988) are shown, as well as (2) the results of a Wilcoxon signed-rank test.

	Logistic regression		**Decision tree**	
	(1)	**(2)**	**(1)**	**(2)**
Local vs. Network	0.5182	0.6922	0.2141	0.002577
Local vs. Hybrid	0.3611	< 0.0001	0.1365	< 0.0001
Network vs. Hybrid	0.4057	< 0.0001	0.2243	0.4286
	Random forest		**Neural network**	
	(1)	**(2)**	**(1)**	**(2)**
Local vs. Network	0.3154	< 0.0001	0.4657	0.004019
Local vs. Hybrid	0.05284	< 0.0001	0.4111	0.0005685
Network vs. Hybrid	0.4618	0.0002536	0.4742	0.3956

and hybrid random forest model. We note that the advantage of including network variables is higher for sensitivity rates up till 80%. Therefore, we take a closer look at the sensitivity and specificity of the random forest models.

Table 3: The sensitivity (sens) and specificity (spec) of the random forest models. Both metrics are calculated for a 50% cut-off rate (Sens/Spec 50) and a cut-off rate (Sens/Spec 14) similar to the actual attrition ratio of 14%.

	Sens 50	**Spec 50**	**Sens 14**	**Spec 14**
Local model	0.02983	0.9963	0.3038	0.8891
Network model	0.09389	0.9892	0.3482	0.8952
Hybrid model	0.05160	0.9953	0.3634	0.8969

Table 3 presents the measures for the local, network and hybrid random forest model. In a marketing context, one might say that sensitivity is important since it informs us how well our model is able to identify actual product attrition. In general, it is believed to be more costly to miss the products of interests than the other way around. We can immediately observe the advantage of including network features. If we focus on the sensitivity when we follow the current

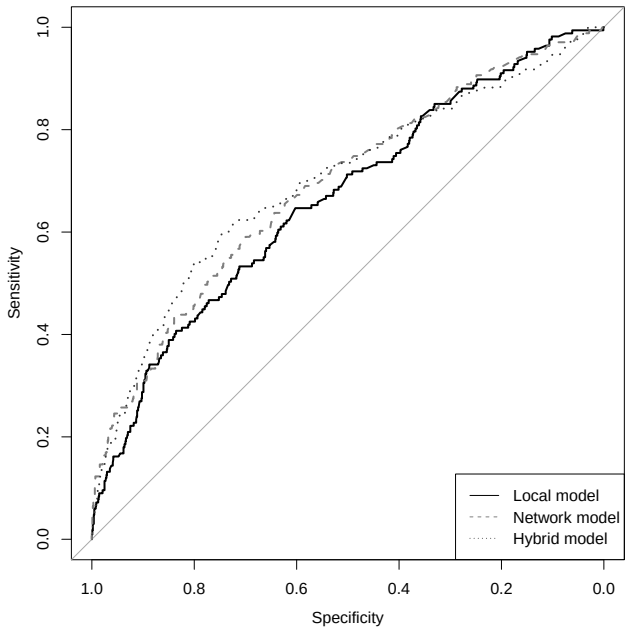

Figure 5: Median ROC curves of the local, network and hybrid random forest model.

attrition rate of 14%, we notice an increase of almost 6 percentage points (i.e. a 20% performance increase) in sensitivity for the hybrid model compared to the local model. Moreover, the network model already offers an increase of more than 4 percentage points (i.e. a 15% performance increase).

4.2. Importance of product and network characteristics in the hybrid model

Logistic regression does allow for easy evaluation of its model variables. Therefore, we take a better look at the hybrid model. The significance of each variable included in the model after feature selection can be observed in Table 4. Note that no causal relation may be inferred and that interaction effects remain unnoticed in this model. However, we find some significant positive and negative effects. With regards to local variables, average IPT has a negative correlation with attrition probability. Products with an already high IPT have a lower probability of increasing their IPT. Furthermore, if the number of customers is

increasing, the probability of attrition increases. The same can be said about a positive change in the number of customers over the last five months. These last findings seem unusual and might be due to interaction effects. When we take a look at the feature 'Promo', we observe that the more frequently a product was in promotion, the lower its chance of attrition, which is in line with general understanding and marketing research (Walters, 1991). Next, the higher the percentage of payments by card, the lower its chance of attrition. This might also relate to the fact that most customers pay by card (more than 70%). Contrarily, an increase in the percentage of payments by means of foodstamps leads to a rise of the attrition probability. However, foodstamps only account for 0.13% of the payments on average. As we discussed in Section 3.2.2, there are various reasons why payment methods can influence product attrition. The size and direction of the means of payment impact might additionally be country-specific (Schmiedel et al., 2012). Secondly, with regards to network variables, we can also deduce findings. We notice that the number of product neighbors in the same group with medium aggregation has a positive influence while neighbors in the same group with high aggregation has a negative influence on attrition probability. On the one hand, interaction effects may be due here. On the other hand, we might be observing substitution and complementary effects. Products in the middle group are perhaps rather substitutes, while products in the higher-level group can be rather complements. We also find that the number of passively churning neighbors in the same group (most detailed categorization), has a negative impact on attrition probability. Similarly, this can be due to hidden interaction effects as well as substitution effects. Finally, the edge frequency weight has a negative effect on attrition probability. The higher the frequency of the purchase, the lower the chance of attrition. This is in line with previous research which positively correlates a higher frequency to more loyal customers (Buckinx & Van den Poel, 2005). Moreover, it is worthy to note that the time period also has a significant impact. Therefore, it is important to take into account this period when making new predictions.

Table 4: The estimate and significance of each variable included in the best performing hybrid logistic regression model with an AUC of 0.70 on the holdout test set.

Variable	**Estimate**	**p-value**
Intercept	0.09602	0.9475
Time period 2	−0.2149	0.1534
Time period 3	−1.0277	< 0.0001
Local variables		
RecencyChange	0.08869	0.1492
FrequencyChange	−0.0002564	0.4884
AvIPT	−0.2699	0.003385
RegularityIPT	0.006392	0.7435
NoCustomers	0.00001704	0.04533
Promo	−1.3846	0.03102
AvPrice	−0.005853	0.5134
AvDiscount	−0.001684	0.9301
Card	−2.0820	0.01114
MobilePay	−27.4102	0.1276
SlopeStores	0.06362	0.2668
SlopeNoCustomers	0.1363	0.01255
SlopePromo	0.01068	0.8540
SlopeAvDiscount	−0.04973	0.3812
SlopeCard	0.01289	0.8047
SlopeFoodstamps	0.1293	0.02022
SlopeMobilePay	0.04530	0.3733
Network variables		
NeighborsGroupL2	20.5352	0.008195
NeighborsGroupL1	−13.5442	< 0.0001
NeighborsChurn	20.0299	0.4737
ChurnGroupL3	−0.2323	0.006564
ChurnGroupL2	−0.1390	0.8264

ChurnGroupL1	0.1770	0.8645
EdgeR	0.2067	0.4826
EdgeF	−0.3211	0.000352
PageRankChurn	−2557.2767	0.6019

p-value < 0.1; **p-value** < 0.05; ***p-value*** < 0.01; ***p-value*** < 0.001

The logistic regression model indicates that customer behavior can indeed be captured in product and network variables and, as such, impacts the IPT of products. However, the presented model does not show any interaction effects hidden underneath. Therefore, it would be valuable to look at the importance of the variables in the final hybrid random forest model. We can analyze the importance of each variable by its relative influence in the model. This metric is determined based on whether the variable was selected for splitting a tree and how much the squared error over all trees improved as a result (The H2O.ai team, 2017). The relative importance of each variable is visualized in Figure 6. For numeric details, the reader is referred to Table B.1 in Appendix B. Firstly, we observe that network and local variables alternate each other continuously in importance. The local variable, change in number of customers, and the network variable, number of neighboring products in the same high-level group, prove to be very important. Furthermore, the means of payment also turns up in the random forest as important. While only the edge frequency deemed important in the logistic regression, all edge variables turn out to be valuable in the final random forest model. Next, there are still some network variables which are in the upper half of importance and which are worth mentioning. The local clustering coefficient seems to be useful. This metric indicates how clustered a product is within customer purchases. The motivation behind this might be that it is important to know which products are frequently purchased by the same customers, or rather bought by one-time shoppers. Furthermore, the PageRank personalized with product attrition suggests that attrition spreads through the network. The finding that these last two features are important is in line with the findings of Dhar et al. (2014) in a similar setting. Finally, RFM

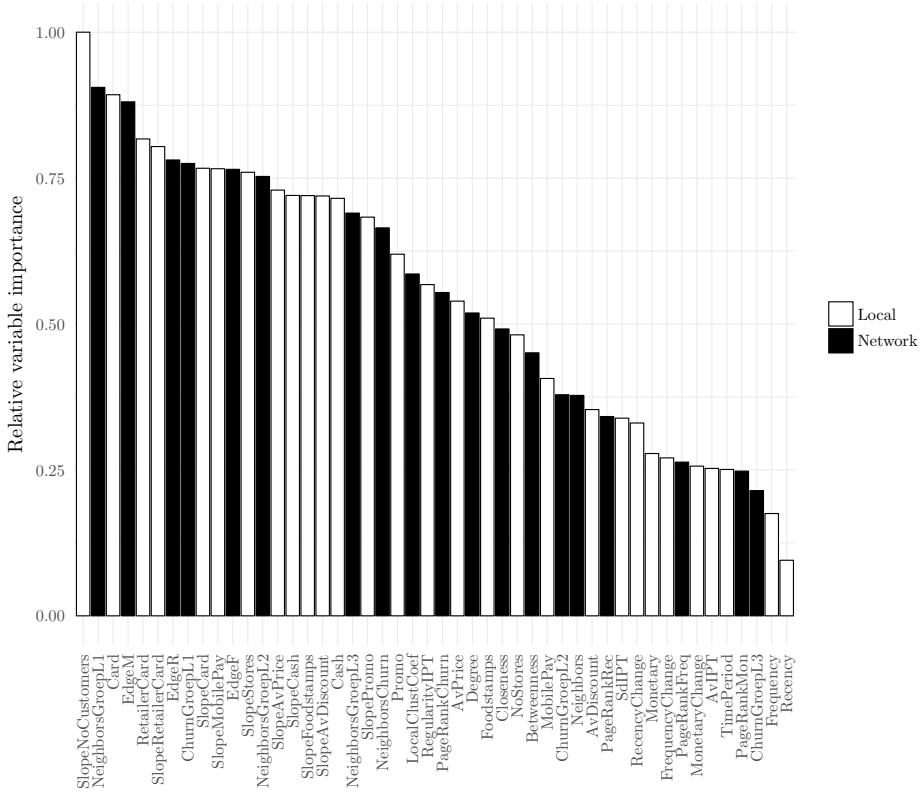

Figure 6: Variable importance of the best performing (on the hold-out validation set) hybrid random forest model in terms of mean decrease in node impurity measured by the Gini index if that particular variable would be removed from the variable set.

variables do not prove to be very important as well as the adapted PageRank using RFM values as a starting point. Thus, although RFM values can be important for the identification of customer loyalty (Cheng & Chen, 2009), this does not necessarily lead to product loyalty. Rather, the RFM values of specific customer-product relations are deemed to be more important, emphasizing the importance of applying more intelligent bipartite network structural measures. Furthermore, we would like to add that in general adapting the PageRank with local variables (e.g. attrition, RFM) can be valuable, but results depend naturally on the choice of the local variable as shown in this paper.

4.3. Discussion

We built a local model using only product characteristics, a network model using only unipartite and bipartite network characteristics of a customer-product graph, and a hybrid model using both feature sets. We found that, by including network characteristics, we can improve the predictive value of product attrition models. This indicates that customer behavior and preferences can indeed be leveraged by means of a customer-product graph. This was already suggested by previous literature (Huang et al., 2007; Kim et al., 2012) in a descriptive setting, but now also confirmed in a practical setting in offline retail.

These results contribute both to practitioners and researchers. Companies can use our models as expert tools to focus their marketing attention on certain products and to better map customer behavior. Furthermore, we included several product characteristics as well as network characteristics and illustrated their importance in product attrition prediction. This can deliver novel insights to the retailer and guide them in the development of new models. We found that, although RFM values seemed to be less valuable, the payment method and promotions do matter. With regards to the customer-product graph, purchase behavior of similar products is relevant. We demonstrated that personalized PageRank can be a valuable metric, depending on the starting vectors employed. Moreover, we found that edge attributes also improve performance. This means that both unipartite as well as more complex bipartite networks contribute to these results. With regards to the existing knowledge base, we contribute by, to the best of our knowledge, being the first to apply customer-product graphs in a predictive setting. Moreover, we incorporated an extensive set of product and network features, including network and bipartite features, and illustrated their importance in the final model. Moreover, our findings are robust across four techniques (logistic regression, decision tree, random forest, neural network) and are validated by means of repeated five-fold cross-validation. Finally, we provide an empirical study with a focus on offline retail of everyday products, and illustrate as such how transactional data can be leveraged by using new ways of structuring data, i.e. by means of networks.

We also researched homophily in our customer-product networks. Homophily assumes that more similar entities are more likely to form a connection (McPherson et al., 2001). In the training sets, we found significant (p-value < 0.0001) indications of homophily in each time period. This indicates that products decreasing in purchase frequency are more closely clustered together than expected randomly. However, these results do not hold if we apply the attrition rates of the next five months. This might be due to the unusual characteristics of the customer-product graphs. They are a lot denser than graphs studied in other works (Kim et al., 2012) with density rates of more than 77% in the unipartite product graphs. Nevertheless, additional tests show that these networks do have dyadic properties, meaning that products decreasing in purchase frequency are more densely connected than expected if they would have been randomly connected.

4.4. Limitations and further research

Research on customer-product graphs, especially in offline retail, has been limited up till now. Nevertheless, it contains real value thanks to its ability to incorporate customer spending behavior. In this paragraph, we shortly present some interesting opportunities for further research.

Firstly, our study used network analytics and evaluated a prediction model using a large offline retail dataset. The study was constrained by the time duration of the sample dataset. As such, it could be repeated over a longer time period to include time dependencies. We observed a significant impact of the third time period in our research. This network, in addition, appeared to be a lot denser than the other two graphs. A longer time period could thus provide more insights into seasonal effects. Moreover, in future work, seasonal effects could even be explicitly included by adding features indicating holidays, weather, season, etc. Note, however, that we excluded products which are heavily subjected to seasonal effects by setting a maximum IPT. Seasonal effects do not necessarily need to be included by means of explicit features; they are reflected in the customer-purchase graphs since these implicitly express customer

preferences. Additionally, it would be valuable to repeat this study for other retailers, potentially including online retail. Furthermore, it would be useful to investigate different adaptions of the PageRank algorithm. In this study, we investigated the spread of each of the RFM metrics. However, the recency, frequency and monetary local values did not prove to add value relative to the other metrics. Therefore, it could be interesting to research the spread of other local features such as price, promotion and payment information. Additionally, instead of focusing on classification, we could focus on predicting a product's repeat purchase by means of survival analysis. Prinzie & Van den Poel (2007), for example, study survival analysis in the context of a cross-sell model for home appliances. Finally, the dynamics of the customer-product graph itself could be researched. We already found that our customer-product graphs are much denser and that our nodes are more clustered than typically expected in social networks or market basket networks. Similarly, Kim et al. (2012) found that their customer-product network is 20 times more dense than their market basket network and has a higher clustering coefficient. The characteristics as well as how customer behavior evolves over time could be more thoroughly studied in future work. As such, fluctuations in customer behavior could be reflected in descriptive and predictive analytics models. A potential approach for this could be, for example, community mining or co-clustering.

5. Conclusion

Customers do not behave and purchase in a vacuum, but rather their choices are influenced by social interactions, context and environment. We focus on leveraging purchasing similarities by building a customer-product network for the purpose of product attrition prediction. In these networks, products are connected to the customers who have purchased them, regardless of the timing and basket properties. We applied four techniques in order to create three models, one with local features, one with network features and one with both. For our network features, we relied on a technique called featurization which proved

to be interesting in the past and allowed us to apply well-known and proven classification techniques. By comparing the three models via four different analytics techniques and validating them by means of repeated cross-validation and two types of statistical tests, we improve the robustness of our findings. The hybrid random forest model proved to deliver improved predictive performance and allows us to better recognize those products which are going to be sold significantly less. This information can be used by the marketing department to tailor their marketing actions and promotions. In addition, this information can be used for supply chain and logistics decisions as well as strategic planning and store lay-out decisions. Moreover, our study indicates the importance of using similarities in customer behavior. The fact that customers who are more alike, behave similarly can be used in numerous analytics applications, both descriptive and predictive. Lastly, we proposed new network features such as an adaption of the PageRank algorithm based on RFM scores and edge attributes. Specifically the local neighborhood of a product as well as its edge attributes proved to be valuable and can be of interest for future studies. In addition, we also confirmed the importance of product characteristics concerning price and promotion details as well as payment details.

Given the value of customer-product networks for gaining novel retail insights, we listed several opportunities for future research. First, the study itself could be extended, e.g. to a longer time period or to different markets such as e-commerce. Additionally, instead of focusing on classification, one could apply survival analysis to predict time to attrition. Finally, it would be particularly interesting to research the dynamics of customer-product graphs. These networks show interesting properties which deviate from typical social graphs, e.g. they are denser. Studying how customer preferences change over time and identifying valuable customer groups, presents itself as an interesting follow-up research.

References

Adamowicz, W. L., & Swait, J. D. (2013). Are food choices really habitual? Integrating habits, variety-seeking, and compensatory choice in a utility-maximizing framework. *Am. j. of agric. econ.*, *95*, 17–41. doi:10.1093/ajae/aas078.

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE trans. on knowl. and data eng.*, *17*, 734–749. doi:10.1109/TKDE.2005.99.

Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, *106*, 21544–21549. doi:10.1073/pnas.0908800106.

Baesens, B., Verstraeten, G., Van den Poel, D., Egmont-Petersen, M., Kenhove, P. V., & Vanthienen, J. (2004). Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers. *Eur. j. of oper. res.*, *156*, 508–523. doi:10.1016/S0377-2217(03)00043-2.

Benoit, D. F., & Van den Poel, D. (2012). Improving customer retention in financial services using kinship network information. *Expert syst. with appl.*, *39*, 11435–11442. doi:10.1016/j.eswa.2012.04.016.

Borle, S., Singh, S. S., & Jain, D. C. (2008). Customer lifetime value measurement. *Manag. sci.*, *54*, 100–112. doi:10.1287/mnsc.1070.0746.

Brandes, U. (2001). A faster algorithm for betweenness centrality. *J. of math. sociol.*, *25*, 163–177. doi:10.1080/0022250X.2001.9990249.

Breiman, L. (2001). Random forests. *Mach. learn.*, *45*, 5–32. doi:10.1023/A:1010933404324.

Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *Eur. j. of oper. res.*, *164*, 252–268. doi:10.1016/j.ejor.2003.12.010.

Buckinx, W., Verstraeten, G., & Van den Poel, D. (2007). Predicting customer loyalty using the internal transactional database. *Expert syst. with appl.*, *32*, 125–134. doi:10.1016/j.eswa.2005.11.004.

Cheng, C.-H., & Chen, Y.-S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert syst. with appl.*, *36*, 4176–4184. doi:10.1016/j.eswa.2008.04.003.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Systems*, 1695. URL: http://igraph.org.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biom.*, *44*, 837–845. doi:10.2307/2531595.

Dhar, V., Geva, T., Oestreicher-Singer, G., & Sundararajan, A. (2014). Prediction in economic networks. *Inf. syst. res.*, *25*, 264–284. doi:10.1287/isre.2013.0510.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Soc. netw.*, *1*, 215–239. doi:10.1016/0378-8733(78)90021-7.

Fritsch, S., Guenther, F., Suling, M., & Mueller, S. M. (2016). *neuralnet: Training of neural networks*. R package version 1.33. https://CRAN.R-project.org/package=neuralnet.

Hill, S., Provost, F., & Volinsky, C. (2006). Network-based marketing: Identifying likely adopters via consumer networks. *Stat. sci.*, *21*, 256–276. doi:10.1214/088342306000000222.

Huang, Z., Zeng, D. D., & Chen, H. (2007). Analyzing consumer-product graphs: Empirical findings and applications in recommender systems. *Manag. sci.*, *53*, 1146–1164. doi:10.1287/mnsc.1060.0619.

Kaymak, U. (2001). Fuzzy target selection using RFM variables. In *IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th* (pp. 1038–1043). USA: IEEE volume 2. doi:10.1109/NAFIPS.2001.944748.

Kim, H. K., Kim, J. K., & Chen, Q. Y. (2012). A product network analysis for extending the market basket analysis. *Expert syst. with appl.*, *39*, 7403–7410. doi:10.1016/j.eswa.2012.01.066.

Kramer, S., Lavrač, N., & Flach, P. (2001). Propositionalization approaches to relational data mining. In S. Džeroski, & N. Lavrač (Eds.), *Relational Data Mining* (pp. 262–291). Berlin, Heidelberg: Springer. doi:10.1007/978-3-662-04599-2_11.

Kuhn, M., Weston, S., Coulter, N., & Quinlan, R. (2014). *C50: C5. 0 decision trees and rule-based models*. R package version 0.1.0-24. https://CRAN.R-project.org/package=C50.

Kumar, V., Ramani, G., & Bohling, T. (2004). Customer lifetime value approaches and best practice applications. *J. of interact. mark.*, *18*, 60–72. doi:10.1002/dir.20014.

Lessmann, S., Baesens, B., Seow, H., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *Eur. j. of oper. res.*, *247*, 124–136. doi:10.1016/j.ejor.2015.05.030.

Macskassy, S. A., & Provost, F. J. (2007). Classification in networked data: A toolkit and a univariate case study. *J. of mach. learn. res.*, *8*, 935–983. URL: http://dl.acm.org/citation.cfm?id=1314532.

Martens, D., Provost, F. J., Clark, J., & de Fortuny, E. J. (2016). Mining massive fine-grained behavior data to improve predictive analytics. *MIS quart.*, *40*, 869–888. URL: http://misq.org/mining-massive-fine-grained-behavior-data-to-improve-predictive-analytics.html.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annu. rev. of sociol.*, *27*, 415–444. http://www.jstor.org/stable/2678628.

Miguéis, V. L., Van den Poel, D., Camanho, A. S., & e Cunha, J. F. (2012). Modeling partial customer churn: On the value of first product-category purchase sequences. *Expert syst. with appl.*, *39*, 11250–11256. doi:10.1016/j.eswa.2012.03.073.

Neville, J., & Jensen, D. D. (2007). Relational dependency networks. *J. of mach. learn. res.*, *8*, 653–692. URL: http://dl.acm.org/citation.cfm?id=1314522.

Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B., & Vanthienen, J. (2017). Social network analytics for churn prediction in telco: Model building, evaluation and network architecture. *Expert syst. with appl.*, *85*, 204–220. doi:10.1016/j.eswa.2017.05.028.

Page, L. (2001). Method for node ranking in a linked database. US Patent 6,285,999.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). *The PageRank citation ranking: Bringing order to the web*. Technical Report 1999-66 Stanford InfoLab.

Prinzie, A., & Van den Poel, D. (2007). Predicting home-appliance acquisition sequences: Markov/markov for discrimination and survival analysis for modeling sequential information in NPTB models. *Decis. support syst.*, *44*, 28–45. doi:10.1016/j.dss.2007.02.008.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

Raeder, T., & Chawla, N. V. (2009). Modeling a store's product space as a social network. In *2009 International Conference on Advances in Social Network Analysis and Mining, ASONAM 2009, 20-22 July 2009, Athens, Greece* (pp. 164–169). doi:10.1109/ASONAM.2009.53.

Riedmiller, M. (1994). *Rprop - Description and implementation details*. Technical Report University of Karlsruhe.

Ríos, S. A., & Videla-Cavieres, I. F. (2014). Generating groups of products using graph mining techniques. In *18th International Conference in Knowledge Based and Intelligent Information and Engineering Systems, KES 2014, Gdynia, Poland, 15-17 September 2014* (pp. 730–738). doi:10.1016/j.procs.2014.08.155.

Schmiedel, H., Kostova, G., & Ruttenberg, W. (2012). *The social and private costs of retail payment instruments: a European perspective*. Technical Report 1698269709 Federal Reserve Bank of St Louis. URL: http://ideas.repec.org/p/ecb/ecbops/20120137.html (Last revised: 10/03/2015).

Sun, J., & Tang, J. (2011). A survey of models and algorithms for social influence analysis. In *Social Network Data Analytics* (pp. 177–214). Springer. doi:10.1007/978-1-4419-8462-3_7.

The H2O.ai team (2017). *h2o: R Interface for H2O*. URL: https://CRAN.R-project.org/package=h2o r package version 3.10.5.3.

Van Vlasselaer, V., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2017). GOTCHA! Network-based fraud detection for social security fraud. *Manag. sci.*, *63*, 3090–3110. doi:10.1287/mnsc.2016.2489.

Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *J. of mark.*, *68*, 106–125. doi:10.1509/jmkg.68.4.106.42728.

Verbeke, W., Martens, D., & Baesens, B. (2014). Social network analysis for customer churn prediction. *Appl. soft comput.*, *14*, 431–446. doi:10.1016/j.asoc.2013.09.017.

Videla-Cavieres, I. F., & Ríos, S. A. (2014). Extending market basket analysis with graph mining techniques: A real case. *Expert syst. with appl.*, *41*, 1928–1936. doi:10.1016/j.eswa.2013.08.088.

Walters, R. G. (1991). Assessing the impact of retail price promotions on product substitution, complementary purchase, and interstore sales displacement. *J. of mark.*, *55*, 17–28. doi:10.2307/1252234.

Zhang, K., Bhattacharyya, S., & Ram, S. (2016). Large-scale network analysis for online social brand advertising. *MIS quart.*, *40*, 849–868. URL: http://misq.org/large-scale-network-analysis-for-online-social-brand-advertising.html.

Appendix A. Description of model variables

Table A.1 presents an overview of local (product) and network characteristics extracted from a customer-product network.

Table A.1: Product and network variables and their description. All variables are measured during the specified time period.

Variable	**Description**
Local variables	
Recency	Number of days in between last purchase and end of time period. Discretized into 5 quintiles.
Frequency	Number of purchases. Discretized into 5 quintiles.
Monetary	Total revenue of purchases. Discretized into 5 quintiles.
RecencyChange	Slope of change in recency per month; capped at the 5^{th} and 95^{th} percentile and standardized.
FrequencyChange	Slope of change in frequency per month; capped at the 5^{th} and 95^{th} percentile and standardized.
MonetaryChange	Slope of change in monetary value per month; capped at the 5^{th} and 95^{th} percentile and standardized.
AvIPT	Average of IPT (in days).
SdIPT	Standard deviation of IPT (in days).
RegularityIPT	Ratio of SdIPT and AvIPT; if AvIPT equals zero, this value is set to SdIPT.

NoStores	Number of unique stores in which the product is sold.
NoCustomers	Number of unique customers.
Promo	Percentage of purchases for which the product was in promotion.
AvPrice	The average price of the product.
AvDiscount	The average discount of the product in absolute value.
Cash	Percentage of purchases paid in cash.
Card	Percentage of purchases paid electronically.
Foodstamps	Percentage of purchases paid with paper foodstamps.
RetailerCard	Percentage of purchases paid with the retailer's debit card.
MobilePay	Percentage of purchases paid with a mobile phone.
SlopeStores	Slope of the change in NoStores per month; capped at the 5^{th} and 95^{th} percentile and standardized.
SlopeNoCustomers	Slope of the change in NoCustomers per month; capped at the 5^{th} and 95^{th} percentile and standardized.
SlopePromo	Slope of the change in Promo per month; capped at the 5^{th} and 95^{th} percentile and standardized.
SlopeAvPrice	Slope of the change in AvPrice per month; capped at the 5^{th} and 95^{th} percentile and standardized.
SlopeAvDiscount	Slope of the change in AvDiscount per month; capped at the 5^{th} and 95^{th} percentile and standardized.
SlopeCash	Slope of the change in Cash per month; capped at the 5^{th} and 95^{th} percentile and standardized.
SlopeCard	Slope of the change in Card per month; capped at the 5^{th} and 95^{th} percentile and standardized.
SlopeFoodstamps	Slope of the change in Foodstamps per month; capped at the 5^{th} and 95^{th} percentile and standardized.
SlopeRetailerCard	Slope of the change in RetailerCard per month; capped at the 5^{th} and 95^{th} percentile and standardized.
SlopeMobilePay	Slope of the change in MobilePay per month; capped at the 5^{th} and 95^{th} percentile and standardized.
Network variables	

PageRankRec	Adapted GOTCHA PageRank algorithm applied to Recency.
PageRankFreq	Adapted GOTCHA PageRank algorithm applied to Frequency.
PageRankMon	Adapted GOTCHA PageRank algorithm applied to Monetary.
Neighbors	Number of first-degree neighbors in unipartite network (number of connected products).
NeighborsGroepL3	Number of first-degree neighbors in unipartite network belonging to the same L3 product category (most detailed).
NeighborsGroepL2	Number of first-degree neighbors in unipartite network belonging to the same L2 product category (medium detailed).
NeighborsGroepL1	Number of first-degree neighbors in unipartite network belonging to the same L1 product category (least detailed).
NeighborsChurn	Number of first-degree neighbors in unipartite network that passively churned.
ChurnGroepL3	Number of first-degree neighbors in unipartite network belonging to the same L3 product category (most detailed), that passively churned.
ChurnGroepL2	Number of first-degree neighbors in unipartite network belonging to the same L2 product category (medium detailed), that passively churned.
ChurnGroepL1	Number of first-degree neighbors in unipartite network belonging to the same L1 product category (least detailed), that passively churned.
NoConnections	Number of direct edges to customers.
EdgeR	Average of the Recency weights of direct edges with the Recency weights discretized into 5 quintiles.
EdgeF	Average of the Frequency weights of direct edges with the Frequency weights discretized into 5 quintiles.

EdgeM	Average of the Monetary weights of direct edges with the Monetary weights discretized into 5 quintiles.
Degree	Degree centrality, normalized.
Closeness	Closeness centrality, normalized.
Betweenness	Betweenness centrality, normalized.
LocalClustCoef	Local clustering coefficient.
PageRankChurn	Personalized PageRank using product attrition, with damping factor set to 0.85.

Appendix B. Variable importance

Table B.1 presents the variable importance in the hybrid random forest model.

Table B.1: The relative importance of each variable in the hybrid random forest model. In addition, its importance on a scale of 0 to 1 is given. Network variables are accentuated in bold.

Variable	**Relative importance**	**Scaled importance**
SlopeNoCustomers	3865.2598	1.00000000
NeighborsGroepL1	3499.9412	0.90548666
Card	3451.0300	0.89283263
EdgeM	3404.0449	0.88067688
RetailerCard	3158.2720	0.81709178
SlopeRetailerCard	3107.8982	0.80405933
EdgeR	3019.0454	0.78107180
ChurnGroepL1	2995.9548	0.77509793
SlopeCard	2964.1155	0.76686061
SlopeMobilePay	2960.9929	0.76605276
EdgeF	2957.0359	0.76502902
SlopeStores	2937.6428	0.76001175
NeighborsGroepL2	2910.0046	0.75286134
SlopeAvPrice	2819.5332	0.72945504
SlopeCash	2784.1602	0.72030350
SlopeFoodstamps	2782.9639	0.71999401

SlopeAvDiscount	2780.7424	0.71941929
Cash	2764.9990	0.71534624
NeighborsGroepL3	2667.1936	0.69004252
SlopePromo	2640.4956	0.68313536
NeighborsChurn	2569.2712	0.66470856
Promo	2395.0117	0.61962504
LocalClustCoef	2263.8835	0.58570023
RegularityIPT	2193.8025	0.56756923
PageRankChurn	2141.2820	0.55398139
AvPrice	2084.0913	0.53918532
Degree	2005.7013	0.51890466
Foodstamps	1971.3920	0.51002833
Closeness	1900.0089	0.49156047
NoStores	1860.8704	0.48143475
Betweenness	1741.8317	0.45063767
MobilePay	1571.9622	0.40668991
ChurnGroepL2	1464.1974	0.37880957
Neighbors	1459.8702	0.37769007
AvDiscount	1366.0131	0.35340783
PageRankRec	1319.9918	0.34150145
SdIPT	1309.1505	0.33869664
RecencyChange	1277.1089	0.33040700
Monetary	1074.9142	0.27809623
FrequencyChange	1045.8759	0.27058359
PageRankFreq	1017.9736	0.26336485
MonetaryChange	991.2912	0.25646173
AvIPT	976.2747	0.25257674
TimePeriod	969.4575	0.25081302
PageRankMon	958.2082	0.24790266
ChurnGroepL3	829.4179	0.21458271
Frequency	677.5262	0.17528607
Recency	367.9931	0.09520527