

Estimating the Effect of Word of Mouth on Churn and Cross-Buying in the Mobile Phone Market

with Markov Logic Networks

Torsten Dierkes and Martin Bichler*
Department of Informatics, TU München, Germany

Ramayya Krishnan
School of Information Systems and Management and iLab, The Heinz College
Carnegie Mellon University, USA

Abstract: Much has been written about word of mouth and customer behavior. Telephone call detail records provide a novel way to understand the strength of the relationship between individuals. In this paper, we predict using call detail records the impact that the behavior of one customer has on another customer's decisions. We study this in the context of churn (a decision to leave a communication service provider) and cross-buying decisions based on an anonymized data set from a telecommunications provider. Call detail records are represented as a weighted graph and a novel statistical learning technique, Markov Logic Networks, is used in conjunction with logit models based on lagged neighborhood variables to develop the predictive model. In addition, we propose an approach to propositionalization tailored to predictive modeling with social network data. The results show that information on the churn of network neighbors has a significant positive impact on the predictive accuracy and in particular the sensitivity of churn models. The results provide evidence that word of mouth has a considerable impact on customers' churn decisions and also on the purchase decisions, leading to a 19.5% and 8.4% increase in sensitivity of predictive models.

Keywords: Data Mining, Marketing, Telecommunications, Social Network Analysis

*Contact author: Martin Bichler, Department of Informatics, TU München, Boltzmannstr. 3, 85748 Garching, Tel.: +49-89-289-17500, Fax: +49-89-289-17535, bichler@in.tum.de

1 Introduction

The churn rate refers to the proportion of contractual customers or subscribers who leave a service provider during a given time period. It is a possible indicator of customer dissatisfaction, cheaper and/or better offers from the competition, more successful sales and/or marketing by the competition, or reasons related to the customer life cycle [4, 19, 56]. Customer lifetime value is affected by acquisition cost, customer retention, and margin, and several studies show that customer retention is the most important factor, with a significant impact on the financial performance of a company [19, 22]. Neslin et al. [50] show that the accuracy of churn prediction models matters and that just using one method rather than another can easily amount to changes in profit in the hundreds of thousands of dollars. Accurate churn prediction is a prerequisite for effective customer retention activities. It is estimated that in the year 2000, the industry average of monthly cellular churn rate was 2.2% in the U.S. [3]. According to a Yankee Group report, the weighted average churn rate of all carriers in North America was 2.6% at the end of 2002. In Western Europe, the situation was similar with 2–3% monthly churn rates [23]. According to Mattison [47], wireless providers are experiencing annual churn ranging from 25% in Europe to over 30% in the U.S. and 48% in Asia.

Churn prediction aims to determine the background stimulators and characteristics of potential churners, and predict whether a customer has a high or low risk of switching supplier. This is often modeled as a classification problem [50]. Given a training and a test data set, analysts can compare different models based on overall accuracy (or error rate), or their lift curves on the test data set [46]. Among those customers who exhibit a high churn likelihood, marketers want to intervene to retain those who have a high "customer lifetime value" (CLTV) [64] as well as those who could influence the churn decision making behavior of other customers [40]. This influence is exercised via word of mouth (WOM), and might also be important for buying decisions as the literature indicates [5]. WOM refers to the informal communication between customers about a product or service. In this paper, we estimate the effects of

WOM on churn. We use customers' anonymized call detail records as a way of modeling WOM. Although we cannot observe the content of the communication among customers, we use the intensity of the communication as an observable variable for latent WOM

For comparison, we also analyze the impact of WOM on a specific purchase decision based on anonymized customer calling records of a mobile phone provider. The mobile phone carrier sells games to be played on handsets as a service and it appears that customers who purchase these games are highly connected. The analysis of game downloads and the impact of WOM on this buying decision provides another context to study the effect of WOM. The comparison between these contexts illustrate that the impact of WOM can be quite different for purchase and for churn decisions.

1.1 Statistical relational learning

Traditionally, classification has focused on attribute-value learning where each example or instance can be characterized by a fixed set of attributes [57]. Econometric discrete choice models also fall into this category. In machine learning terms, the hypothesis language is propositional logic and the learning algorithms are referred to as propositional learners [38]. In industries such as the telecommunications sector or for many online social networks, data about the customer network is available. A research study that was conducted by Keaveney [26] showed that 75% of defecting customers tell their negative experiences to at least one other person. Information about the communication partners of a customer and their decisions to churn might, therefore, improve the prediction of a customer's churn likelihood.

In this paper, we aim to improve models for predicting churn and buying decisions by leveraging network information. In contrast to traditional classification methods, we take into account the information about who a customer calls – i.e., a customer's neighbors in the communication graph derived from the call detail record data. We interpret these graphs as social networks. These networks can be stored in a relational data model. While propositional learners find patterns in a given single relation, statistical relational learning algorithms (also called multi-relational data mining) aim at finding patterns

in multiple relations such as in relational databases [11]. Statistical relational learning is a relatively young field and there is still only limited empirical evidence on the performance of respective learners.

There are two fundamentally different approaches to analyzing multi-relational data. Propositionalization describes methods that transform a relational representation of a learning problem into a propositional (feature-based or attribute-value) representation. In the context of social networks, this would mean using summary statistics of lagged neighbor covariates as predictors. For example, this could be the mean churn rate of neighbors, or their average call volume in a previous time period. This requires the construction of features (attributes) that capture relational properties. This representation can then be analyzed with traditional propositional methods, such as decision tree learners or logit models. Alternatively, statistical relational learning algorithms estimate models on multi-relational data (also known as multi-relational data mining, MRDM). Most statistical relational learning algorithms come from the field of inductive logic programming (ILP) [41] and derivatives. ILP systems dealing with classification tasks typically adopt the covering approach of rule induction systems [63]. ILP suffers from the high computational complexity of the task because the algorithm has to search all the relations and all the relationships between the relations [30]. Other MRDM techniques that are not based on logic formalisms [15, 35] have been proposed, but they suffer from similar problems and are not yet suitable for large data sets that typically need to be analyzed in marketing and Customer Relationship Management (CRM) settings.

Markov logic networks (MLNs) have recently been suggested as a significant step forward in this field [9]. The method draws on Markov Random Fields and ILP and is able to handle larger data sets compared to earlier ILP implementations such as FOIL [53]. The analysis of social network data is one of the potential application fields of MLN [55], although we do not know of any application of MLNs to large social networks to date. The need to consider network effects in customer modeling as well as the use of new machine learning methods has been discussed in the literature (see Gupta et al. [22] for an example).

1.2 Contribution of this paper

The prediction of customers' churn or buying decisions is important to marketers, in particular in service industries. In this paper, we analyze whether WOM has an impact on customer behavior or not. Based on the data set, we analyze if customer churn or game purchasing decisions of individuals in previous periods have an impact on the churn or purchase decisions of their neighbors, i.e., other customers whom the target customer interacted with either via a voice call, short message service (SMS), or multimedia message service (MMS). We compare the predictive accuracy of MLNs to that of a standard discrete choice model ignoring social network information, and a propositionalization approach, i.e., a logit model with aggregate information about the social network of a customer. Propositionalization applied to social network data describes a way to include lagged neighbor covariates as predictors of a discrete choice model. It has been discussed in general, but we do not know of related work on social network data. The analysis is based on anonymized calling data from a telecom provider.

Our contributions are as follows: First, we develop a Markov logic network for the churn and purchase prediction in a large-scale customer network. This is the first paper where MLNs are applied to data from a large-scale social network. Second, we propose and evaluate an approach to propositionalization addressing the specific needs of social network data. Third, we provide results comparing MLNs and propositionalization with a traditional logit model ignoring information about communication neighbors as a benchmark. The logit model resembles the type of models that are often used in CRM [50]. We found the churn behavior of a customer's neighbor has a significant positive impact on predictive accuracy (+8%) and sensitivity (+20%) of churn models, which provides evidence that the intensity of communication can be used as an observable variable for WOM. We are not aware of any other work that estimates the effect of WOM on churn from call data. Interestingly, the effect on sensitivity in our analysis of purchase of games was lower than that of churn (a 8.46% vs. a 19.57% increase in sensitivity), although such customers are much more connected. These results do not necessarily carry over to other products or industries, but they provide evidence for the effect that WOM can have in specific industries.

In the following section we discuss churn and the related WOM literature. Section 3 provides an overview of statistical relational learning and propositionalization. Section 4 introduces the research design and the data, and Section 5 summarizes the results. Finally, Section 6 provides a summary and conclusions.

2 Related Literature on Churn and WOM

There are different strands of literature that are relevant to this paper. In this section we discuss related literature on churn and WOM. One of the central steps of customer churn management is to determine the reasons for churn and to predict the potential churners. One approach is customer satisfaction surveys [47]. Call quality, pricing options, coverage area, customer service, and image are important factors of customer satisfaction that impact the duration of a customer relationship [4]. However, such surveys may fail to find the real reasons of churn and they may even be misleading in some cases. Kon [36] found that in one specific survey, 80% of the churners had described their satisfaction level as either “satisfied” or “very satisfied” within the 12 months before their switching. Reichheld emphasizes the danger of this “satisfaction trap” and states that “*What matters is not what customers say about their level of satisfaction but whether the value they felt they’ve received will keep them loyal*” [54].

Several authors emphasize the role that *word of mouth (WOM)* plays for a customer’s churn or buying decision. WOM has extensively been studied in the marketing literature [2, 5, 20, 21, 40, 62, 65, 66]. Studies have shown positive WOM to be an outcome of high customer satisfaction [59]. There is also a relation between customer tenure and the tendency of customers to engage in word of mouth. For example, East et al. [12] found that the recommendation (positive word of mouth) is the prevailing reason for switching services and that the recently acquired customers recommend more frequently than the existing (long-term) ones. An empirical study which targets the German energy sector shows that the switchers give more positive WOM about their new supplier in comparison to the ones who stay. Another finding of

the same study is that the referral switchers (i.e., the ones who are affected by WOM in their switching decision) tend to give even more WOM than any other switchers [65].

In addition, a huge body of literature has emerged on the analysis of social networks. Recent research on marketing applications has focused on models for simulating the spread of information in a social network [8, 24, 34, 61] or suggests techniques to maximize influence in social networks [27]. Leskovec et al. [45] analyze the data of an online incentivized viral marketing program, in which the retailer uses a recommendation referral program. Here, customers could generate recommendation e-mails upon purchase of an item and both the sender and the receiver of the email receive a 10% discount or credit upon the receiver purchasing the same item through a referral link. This allowed the measurement of explicit WOM recommendations. They propose a simple model for the propagation of recommendations in the network.

Dasgupta et al. [6] have recently looked at churn prediction based on networked customer data. The authors study the evolution of churners in an operator's network of pre-paid customers and the propensity of a subscriber to churn out of a service provider's network depending on the number of ties (friends) that have already churned. In pre-paid networks there is typically little information available about the customers. Therefore, they focus only on the network topology and use a spreading activation-based technique to predict potential churners based on assumptions on how influence propagates in the network. In their discussion they encourage the use of the so-called collective classification for respective churn prediction tasks, which uses node-level attributes and link information, an approach that we pursued in this paper. We focus on post-paid customers, which allows us to leverage information about customers as is typically done in churn prediction [1, 13, 44, 49, 58]. For example, Neslin et al. [50] found that logit models and decision tree learners performed best in a tournament on churn prediction with different models from 33 participants. Those methods have already performed well in earlier comparisons with a wider variety of data sets [48] and is also used in our paper as a benchmark. In summary, much of the literature in social network analysis focuses on the topology of networks, but does not leverage additional attributes of nodes and edges available in most applications.

Apart from this, there has been an active community focusing on machine learning techniques and predictive modeling for networked data, in particular research in statistical relational learning [7, 9, 10, 14-17, 55]. This literature does not explicitly discuss predictive modeling based on social network data, but networks can be analyzed in a multi-relational data model so that the techniques are potential candidates for churn prediction based on customer networks. We discuss this approach in more detail in the next section.

3 A Brief Review of Statistical Relational Learning and Propositionalization

Conventional discrete choice models and data mining algorithms are defined for analyzing data in a single relation only [31, 32]. Statistical relational learning (SRL) has been an emerging research topic in the data mining community in recent years. “SRL attempts to represent, reason, and learn in domains with complex relational and rich probabilistic structure” [18, p. 4]. There are two main strands in the literature. One approach is to develop relational learning extensions of existing machine learning algorithms so that they can handle the instances directly in a multi-relational representation without any transformation. The second approach is to convert the multi-relational representation of data into a single table [38]. Following Kramer [37], we understand *propositionalization* as a transformation of multi-relational learning problems into attribute-value representations, i.e., into a single table amenable for conventional data mining methods, also referred to as *propositional learners*.

3.1 Statistical Relational Learning and Markov Logic Networks

Most propositional learners are based on statistical learning methods such as decision trees, neural networks, and generalized linear models. In contrast, approaches that learn from multiple interrelated tables are referred to as multi-relational approaches, as the patterns they find are expressed in the relational formalism of first-order logic. Most SRL algorithms come from the field of inductive logic programming (ILP). ILP aims at inductively learning relational descriptions (in the form of logic programs as a restrict-

ed first-order logic) from examples and background knowledge [28, p. 6]. In contrast to propositional logic, first-order logic allows for predicates (i.e., properties of objects or relations) and quantification. Based on the known background knowledge and a set of examples represented as a logical database of facts, an ILP system derives a hypothesized logic program. Unlike many other machine learning approaches, ILP has traditionally dealt with multi-relational data. ILP tools can be applied directly to multi-relational data to find first-order rules from relational data. However, much of the art of ILP lies in the appropriate selection and formulation of background knowledge to be used by the selected ILP system. Therefore a considerable amount of expert knowledge is required. In addition, ILP algorithms suffer from the high computational complexity of the task because the algorithm has to search over all the relations and all the relationships between the relations [30].

SRL is an approach to combine the power of both ILP and statistical learning [17]. It attempts to learn and reason from complex relational and probabilistic structures. There have been a number of advances in SRL including conditional random fields [60], relational dependency networks [51], Bayesian logic programs [29], relational association rules, regression trees in first-order logic, and relational decision trees [7].

Markov logic networks (MLN) have become very popular in statistical relational learning recently [55]. MLNs are a collection of formulas from first-order logic, to each of which a weight is assigned. In other words, it describes a probabilistic logic. Ideas from estimating Markov networks are then applied to learn the weights of the formulas. The vertices of the MLN graph are atomic formulas, and the edges are the logical connectives used to construct the logical formula. A Markov network is a model for the joint distribution of the properties of underlying objects and relations among them. A detailed description of MLNs can be found in Richardson and Domingos [55]. The authors explicitly mention collective classification (i.e., classification on multiple relations) and social network analysis as potential applications of MLNs. In our analysis, we use Alchemy (<http://alchemy.cs.washington.edu/>), an open source software tool for learning MLNs from data.

3.2 Propositionalization

There are two main approaches to propositionalization in the literature, logic-oriented and database-oriented literature. *Logic-oriented propositionalization* constructs features from relational background knowledge and structural properties in terms of first-order rules. Some systems are based mainly on logic programming in order to construct the attributes of the final table and so represent a *logic-oriented approach*. SINUS and its previous version LINUS by Lavrac can be mentioned as examples of logic-oriented approaches [42], as well as RSD systems [67]. These systems construct clauses to derive binary features for propositional learners. This approach performs well on structurally complex but small problems [42, 43]. Business databases present different challenges than those found in the classical showcase areas of ILP and logic-based propositionalization, such as molecular biology or language learning [39]. Whereas the latter often involves highly complex structural elements, perhaps requiring deep nesting and recursion, business databases are usually structurally simpler.

RELAGGS is a *database-oriented* approach to propositionalization, which was developed by Krogel and Wrobel [39]. A wide range of aggregation functions is used when joining relevant tables. For numerical values, *minimum*, *maximum*, *average*, and *sum* operations are used. For nominal attributes, each possible value of an attribute may be counted. Another example of database-oriented propositionalization is the two-step transformation system (POLKA) of Knobbe et al. [35]. Like RELAGGS, POLKA constructs a single final table for propositional learners using joins and aggregate functions. Both methods differ in the way that a large number of relevant tables are joined. In our application, we have a simple relational schema with customers referencing their neighbors via call detail records, so that the join is straightforward. Krogel et al. [39] showed experimentally that database-oriented propositionalization performed better than ILP-based systems both in speed and accuracy. Similar results were found in [38, 39].

4 Data and Research Design

4.1 Data for the Churn Prediction

For our analysis, we had available anonymized historical data for about 120,000 customers and their call detail records for the time period from January to October 2008. The sample consists of all target and non-target customers from a selected geographic region and all their neighboring customers in the call graph. Information about whether and when a customer churned was available. Here, a *churner* (or positive) is defined as a customer who gives notice about their intent to cancel the contract and does not revoke his decision by extending his contract at some point afterwards. A *non-churner* (or negative) is a customer who does not give notice at any time. Note that there are customers who give notice and conduct a contract extension afterwards, i.e., they do not churn in the end. In our data, 6,800 customers told the phone provider that they wanted to cancel the contract (notification). Roughly 1,000 revoked their decision afterwards by extending their contracts. We knew for 645 customers that their contracts were deactivated and they definitely left the provider. For the remaining 5,135 customers the contracts were neither deactivated nor extended within the time frame of the available data. In this paper, we have decided to focus on predicting those customers who churned, i.e., left the service provider. We selected a data sample consisting of 2,645 customers containing 645 positives and 2,000 negatives. This was also about the size of data sets which could be handled by Alchemy. We were not able to run Alchemy effectively with much larger data sets. The negatives were selected such that they had positives as neighbors in the call graph. From those, the 2,000 with the highest communication levels were chosen.

Each customer is described by about 70 attributes, from which we select the 32 best attributes based on a ranking by information gain. Attribute selection is a standard procedure in data mining [33]. These attributes included attributes on the usage of voice and different data services, number of callees, handset equipment, contract type, contract duration, and trends in usage; company interaction data, such as the number of contract extensions, tariff migrations, and interactions with the service center; and cus-

customer demographics, including age and gender. Some customer attributes are subject to change over time due to customer behavior (e.g., usage patterns), while some stay constant (e.g., gender).

In addition to the attributes describing a customer, we had monthly aggregates of the call detail records available for the same time window of January to October 2008. The dataset contained monthly sums of call minutes, calls, short messaging service (SMS) messages, and multimedia messaging service (MMS) messages. Among the 2,645 customers in the subsample there are about 18,000 communication edges from January to October 2008. 600 edges could be found among the 645 positives, 9,500 among the 2,000 negatives, and 8,000 edges with a churner and a non-churner involved in all ten months. For our analysis, we consider the average connection strength of the edges across all months. This results in 4,250 edges in total, thereof 60 among the 645 positives, 2,050 among the 2,000 negatives, and 2,140 edges connecting a churner with a non-churner. Figure 1 (a) and (b) shows a visualization of connections between positives and customers with a game download in the test data set. The graphs could suggest that game downloading is contagious because there are many connections between persons downloading games, while this is less so in the case of churn, where there are typically only connections between a few pairs of churn customers. In our results section, we comment on the impact of neighbors downloading games on other neighbors' purchase decision.

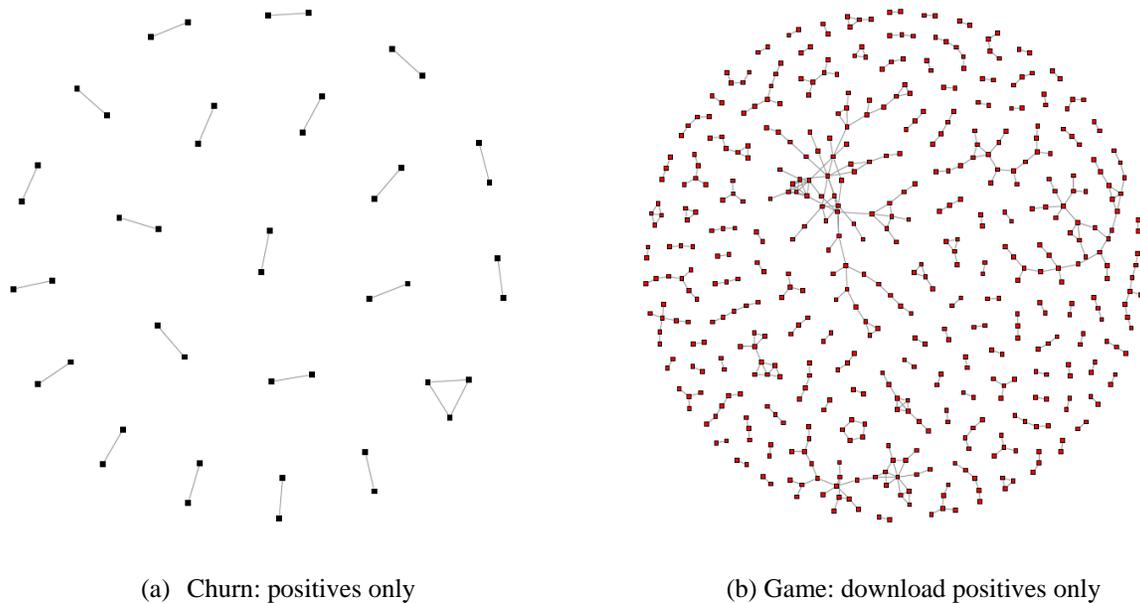


Figure 1: Visualization of selected connections (a) between positives, and (b) between game download customers

We split the set of 2,645 customers in training and test datasets such that it was stratified with respect to the number of positives and edge counts. Training and test data contained 1325 and 1320 customers, respectively. Positives were assigned according to their notification date; customers who notified before July 1, 2008 were assigned to the training data, while those who notified afterwards were assigned to the test data. Negatives were allocated to training and test data randomly, subject to a stratified distribution of communication edges. In other words, training and test data contained about the same numbers of edges strictly among positives, strictly among negatives, and from positives to negatives. For all positives we used the attribute values (e.g., voice usage or number of callees) during the month in which they cancelled. For negatives we used the attribute values in the month of July as a representative month.

4.2 Data for the Prediction of Game Download

In this set of analyses, we focus on customers who download games either via HTTP or WAP. In the dataset there were 8,020 customers with a game download (i.e., positives). There were connections from these to 24,130 customers who did not download a game (i.e., negatives). In order to yield a manageable

dataset for Alchemy, we selected 1,500 positives and 1,500 negatives. For our analysis, we again considered the average connection strength of the edges across all months. Among the selected 3,000 customers there were 7,950 edges – 1,865 among positives, 2,127 among negatives, and 3,950 between both groups across all months.

The 3,000 customers were split into training and test data randomly subject to a stratified distribution of positives and negatives. Training and test data contained a similar amount of edges among positives, among negatives, and from positives to negatives. There were 1,950 and 2,020 communication edges in the training and test data, respectively, thereof 429 exclusively among download customers. In addition, there were 3,950 communication edges from customers in the training dataset to customers in the test dataset. From the 70 available customer attributes we selected again the 32 best ones based on information gain with respect to the target variable.

4.3 Research Design

To overcome the effect of unobserved factors such as churn due to unsatisfactory service, we compared MLN and propositionalization and benchmarked it against logistic regression on the same data set. More specifically, we compared the following settings (T):

<i>Setting</i>	<i>Methods</i>	<i>Relational attributes</i>
T1	Logistic regression	None
T2	Logistic regression with propositionalization	Call minutes and call counts
T3	MLN	Call minutes and call counts

4.3.1 MLN Models

In MLN, attributes can be represented as predicates of the form $A(x,v)$, where A is an attribute, x is an object, and v is x 's value of A . The class attribute is described as $C(x,v)$, where v is x 's class. In terms of

MLN, classification is the problem of inferring the truth value of $C(x, v)$ for all x and v of interest (i.e., whether a customer churns) given all known attributes $A(x, v)$, such as age or voice usage. In propositional classification, $C(x_i, v)$ and $C(x_j, v)$ are assumed to be independent for all x_i and x_j given the known $A(x, v)$. In our example, the churn of customer x_i would be considered independent of the churn of other customers. In relational learning problems, more specifically in collective classification, dependencies between objects can be represented by relational predicates of the form $R(x_i, x_j)$ (Richardson et al. 2009).

We model the collective classification task in MLN using two types of rules. The first models the traditional relationship between an instance's class and its attributes: $A(x, v) \Rightarrow C(x, v)$, where \Rightarrow refers to a logical implication. In addition, we model an influence relationship between connected customers' churn behavior saying that customer x_j is likely to churn if customer x_i did already and x_i and x_j have called each other s times or s minutes, respectively, and vice versa. This is described in a rule $C(x_i, v) \wedge Connection(x_i, x_j, s) \Rightarrow C(x_j, v)$, where $Connection(x_i, x_j, s)$ represents a relation $R(x_i, x_j)$ between x_i and x_j of strength s (either call minutes or call counts).

4.3.2 Propositionalization of Social Network Data

Propositionalization typically deals with multiple interrelated relations. On the one hand, the relational structure of social networks is easier as we are only interested in customers and their neighbors. On the other hand, there is typically more information on how people interact with each other, and consequently there are more possibilities of how connections among customers are modeled. An analyst can look at the number of voice minutes, calls, or SMS messages sent in a period of observation in order to determine a metric for closeness. If there are multiple periods, one might just look at average communication strength, assuming that links among customers are fairly stable, or weight more recent communication higher than that of previous periods. We have used different metrics for connection strength in our research design, as is described in the next section.

After one decides how connection strength is measured, it is still unclear which neighbors are taken into account. One might look at the average churn rate of the closest k neighbors, the closest $k\%$ of the

neighbors, or all neighbors. A fundamental problem with database-oriented propositionalization has been referred to as degree disparity [25]. It describes the systematic variation in the distribution of the degree with respect to the target variable. Due to the degree disparity, any aggregated attribute may correlate highly with the target variable, although the target attribute and the individual, non-aggregated attribute are statistically independent from each other. For example, when taking the number of churn neighbors into account for churn prediction, a customer with a large number of neighbors would also have more churn neighbors than a customer with just a few neighbors. Finally, one can model different relational attributes such as average churn rate, churn rate weighted by connection strength, and/or aggregates of a number of additional relational attributes, such as the number of contract extensions of neighbors over time.

In our results in the next section, we report on (i) the average churn rate of all neighbors, (ii) the average churn rate of the five “closest” neighbors according to connection strength, and (iii) the weighted churn calculated as the sum of positives weighted by connection strengths.

5 Results

In this section, we present the results of our study, comparing propositionalization approaches and MLN models for predicting the target variables churn and game download behavior. Starting with the former, we discuss the results of nine different settings and compare them with the baseline model, i.e., the results of the logistic regression. We have analyzed six settings (T) with propositionalization (T2.x) and three settings for MLNs (T3.x). For propositionalization, we analyze, first, the average churn rate of all neighbors, based on the number of calls (T2.1) and the number of voice minutes (T2.2), second, the average churn rate of the five “closest” neighbors according to connection strength, based on the number of calls (T2.3) and the number of voice minutes (T2.4), and third, the weighted churn calculated as the sum of churning neighbors weighted by connection strength, based on the number of calls (T2.5) and on the number of voice minutes (T2.6). The difference among MLN settings is again the way in which connec-

tion strength is measured. T3.1 and T3.2 include relational attributes representing the average connection strength between two customers, measured in number of calls and call minutes, respectively, and T3.3 includes both.

For each setting, we present and discuss (i) the model results, (ii) some standard metrics, and (iii) the Receiver Operating Characteristic (ROC). Due to confidentiality and privacy restrictions we only provide a summary of the model results. The resulting models for the logistic regression (benchmark model) as well as for all propositionalization settings showed that three groups of customer attributes were especially important: First, three attributes describing the duration of the actual and past contract periods; second, two attributes about the products and services a customer utilizes; and third, two revenue and usage-related attributes. In the logit model, all these seven customer attributes were significant, i.e., of those all but one variable about a customer's utilized products were highly significant ($\alpha < 0.001$).

In contrast, for propositionalization settings T2.1 to T2.4 we found only five significant customer attributes in the respective logit model, of which several had lower statistical significance levels than in the benchmark model. Especially attributes about the past and present contract periods, which were all highly significant in the benchmark logit model, were here either less significant or not significant at all; one attribute remains as significant ($\alpha < 0.001$), one was less significant than before ($\alpha = 0.001$), and one turned out to be not significant at all. The propositionalized relational attribute describing the churn of neighbors in those settings showed high significance ($\alpha < 0.001$). Results for propositionalization settings T2.5 and T2.6 are similar to those of the benchmark model with regard to customer attributes. The relational attributes, i.e. the weighted churn aggregates in those settings were significant, however, with lower significance levels than in settings T2.1 to T2.4. The MLN models are weighted Horn clauses. The higher a weight, the more important the rule is. However, at the current state of MLN research, the weights of a learned rule cannot be translated into a significance level or a probability.

The standard metrics we adopt are defined as follows. Let TP be the true positives, TN the true negatives, FP the false positives, and FN the false negatives. *Accuracy* measures the proportion of predictions,

both for true and false positives that are correct. *Precision* measures the proportion of the claimed true positives that are indeed true positives. *Sensitivity* measures the proportion of true positives that are correctly recognized as true positives. *Specificity* measures the proportion of false positives that are correctly recognized as false positives. Then the above measures are defined as:

- $accuracy = (TP+TN)/(TP+FN+TN+FP)$,
- $sensitivity = TP/(TP+FN)$,
- $specificity = TN/(TN+FP)$, and
- $precision = TP/(TP+FP)$.

%	Logistic (T1)	Propositionalization						MLN		
		Full Cnt (T2.1)	Full Min (T2.2)	Top 5 Cnt (T2.3)	Top 5 Min (T2.4)	Weighted Cnt (T2.5)	Weighted Min (T2.6)	Cnt (T3.1)	Min (T3.2)	Cnt & Min (T3.3)
Accuracy	77.03	85.44	85.44	84.89	84.91	77.86	78.17	81.52	81.52	81.52
Precision	54.15	79.09	79.09	76.58	77.63	57.37	58.42	75.33	75.33	75.33
Sensitivity	34.7	54.37	54.37	53.12	53.12	34.06	34.69	35.31	35.31	35.31
Specificity	90.6	95.04	95.04	94.79	95.1	91.89	92.09	96.3	96.3	96.3
TP	111	174	174	170	170	109	111	113	113	113
TN	905	953	953	947	950	918	920	963	963	963
FP	94	46	46	52	49	81	79	37	37	37
FN	209	146	146	150	150	211	209	207	207	207

Table 1: Results for churn prediction

Table 1 presents the results for churn prediction. The overall accuracy and precision were highest for the propositionalization settings T2.1 to T2.4 followed by the MLN settings T3.1 to T3.3. In terms of sensitivity, propositionalization settings with churn aggregates (T2.1 to T2.4) dominate all other settings, whereas MLN settings did not perform better than the benchmark in this respect. For churn prediction, sensitivity is typically more important than specificity, since it is more costly to lose a valuable customer than to act on a non-churner. Note that the sensitivity of the benchmark logistic regression was rather low. In contrast to propositionalization with churn rates, propositionalization with weighted churn aggregates (T2.5 and T2.6) showed no improvement compared to the benchmark. There was no difference in performance among the three MLN settings. There was no difference between the two alternative relational attributes call counts and call minutes in the performance of MLNs and propositionalization. In addition

to the logistic regression, we have used a C4.5 decision tree learner for the benchmark model, as it is often used as an alternative to the logistic regression [52]. The results of both were similar with respect to predictive accuracy and sensitivity. To summarize the results in terms of standard metrics, the propositionalization approach performs best, followed by the MLN models. Especially in terms of overall sensitivity, the best propositionalization approach outperforms the best MLN model.

While the overall hit rate is expressed by the sensitivity of a model, the Receiver Operating Characteristic illustrates the trade-off between hit rate and error rate, i.e., sensitivity and specificity. The ROC can be represented by plotting the fraction of true positives (TPR = true positive rate) vs. the fraction of false positives (FPR = false positive rate) for every possible cutoff. Figure 2 shows the ROC curve of the three MLN settings and the logistic regression. MLNs dominate the logistic regression, and there is no difference among the three MLN settings T3.1 (Voice_Cnt), T3.2 (Voice_Min), and T3.3 (Voice_Cnt_Voice_Min) so that the respective ROC curves overlap.

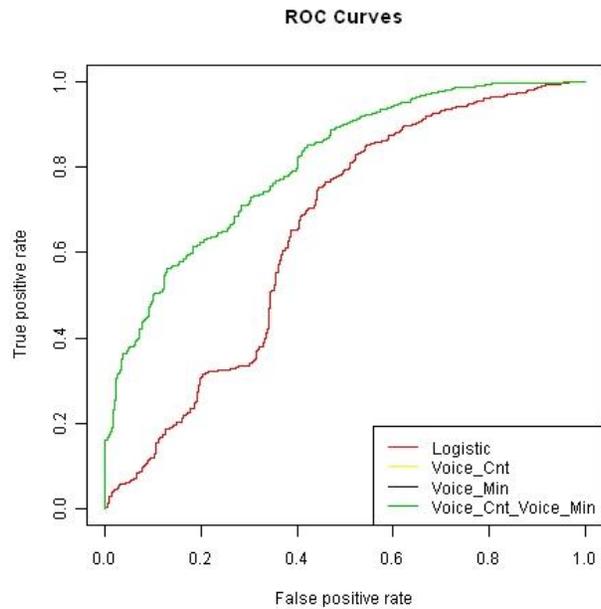


Figure 2: ROC curves of three MNL settings and the logistic regression

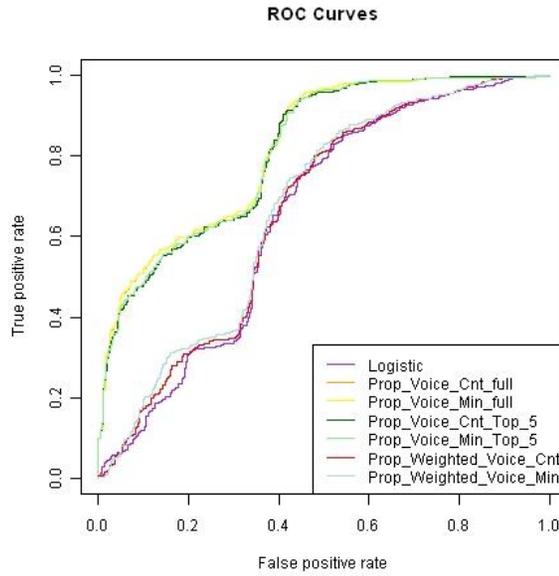


Figure 3: ROC curves of propositionalization settings and the logistic regression

Figure 3 shows that propositionalization with aggregate churn rates also dominate the logistic regression. Settings T2.5 (Prop_Weighted_Voice_Cnt) and T2.6 (Prop_Weighted_Voice_Mnt) with the number of calls and voice minutes weighted, however, are similar to the benchmark logistic regression.

Finally, Figure 4 compares the ROC curves of the best MLN setting (T3.3, Best_MLN_Voice_Cnt_Min) and the best propositionalization setting (T2.1, Prop_Voice_Cnt_full) to that of the benchmark logistic regression (T1, Logistic). It illustrates that although propositionalization with churn rate aggregates (T2.1 to T2.4) has the highest overall sensitivity and accuracy, MLNs yield comparable results for smaller samples.

Overall, despite the superiority of MLN’s sophisticated theoretical framework, propositionalization performed surprisingly well, and MLN could not beat the performance of those models (T2.1). While both are on the same level in terms of ROC, propositionalization performs significantly better in terms of overall sensitivity. However, both propositionalization and MLN clearly outperform the baseline model.

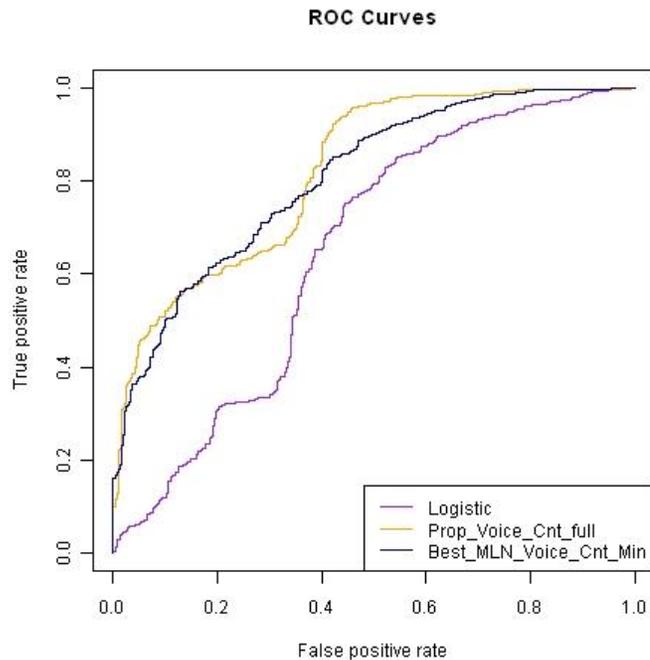


Figure 4: ROC curves of MLNs, propositionalization, and the logistic regression

5.1 Game Download

We were interested in ways how information about neighbors can help predict alternative target variables and looked at game download, since people with game downloads appear to be well connected (see Figure 1 (b)). Game download showed exceptionally many connections among the positives, suggesting that there might also be an influence of customers on each other. Therefore, we also analyzed the nine different settings for customers' game downloading behavior and report the results as a comparison. Data preparation for game download was done as described in Section 4.1. We present and discuss (i) the model results, (ii) some standard metrics, and (iii) the Receiver Operating Characteristic (ROC).

The resulting indicated that three groupings of customer attributes were important: 1) two attributes indicating whether or not a customer signed up for specific online services; 2) three attributes providing information on the duration of the actual and past contract periods; and 3) one general usage-related attribute. For the propositionalization settings T2.x, customer attributes showed the same signifi-

cance level as the benchmark model, except for one attribute about the duration of the present contract period, which was not significant for the benchmark logistic regression but significant for all propositionalization settings ($\alpha=0.01$). Attributes about the utilized services as well as the general usage-related attribute showed high significance ($\alpha<0.001$) in the settings with propositionalization, whereas attributes about past and present contract durations were less significant ($\alpha=0.01$). While the relational attributes, i.e. average churn rate aggregates in settings T2.1 to T2.4, were highly significant ($\alpha<0.001$), the weighted churn aggregates in T2.5 and T2.6 were not significant.

Regarding the standard metrics, the overall accuracy and precision were again highest for the propositionalization settings T2.1 to T2.4 and the three MLN settings T3.1 to T3.3, which all yield comparable results. In terms of sensitivity, propositionalization settings with average churn aggregates of all neighbors (T2.1 and T2.2) were best, closely followed by the MLN settings. All nine settings performed better than the benchmark logistic regression. Table 2 shows the results in detail.

	Logistic (T1)	Propositionalization						MLN		
		Full Cnt (T2.1)	Full Min (T2.2)	Top 5 Cnt (T2.3)	Top 5 Min (T2.4)	Weighted Cnt (T2.5)	Weighted Min (T2.6)	Cnt (T3.1)	Min (T3.2)	Cnt & Min (T3.3)
%										
Accuracy	61.06	69.72	69.72	67.38	67.38	65.04	65.17	68.0	68.0	68.0
Precision	61.16	69.6	69.6	67.49	67.49	64.99	65.2	68.25	68.25	68.25
Sensitivity	62.7	71.16	71.16	68.39	68.39	66.8	66.67	68.43	68.43	68.43
Specificity	59.38	68.24	68.24	66.35	66.35	63.24	63.65	67.56	67.56	67.56
TP	474	538	538	517	517	505	504	518	518	518
TN	440	505	505	491	491	468	471	502	502	502
FP	301	235	235	249	249	272	269	241	241	241
FN	281	218	218	239	239	251	252	239	239	239

Table 2: Results for game download prediction

Figure 7, Figure 8, and Figure 9 in the Appendix show the ROC curves of the MLN settings, the propositionalization settings, and a comparison of the best MLN and propositionalization settings with the benchmark model, respectively. MLNs perform slightly better than the logistic regression, and there is also no difference among the three MLN settings so that the respective ROC curves overlap (Figure 7). Propositionalization settings with aggregate churn rates (T2.1 to T2.4) dominate the logistic regression (

Figure 8). Settings T2.5 and T2.6 with the number of calls and voice minutes weighted did not, however, perform better than the benchmark. So, while one could think of weighting as an appropriate approach to address the level of interaction between two customers, this did not have a strong effect. The comparison of the best MLN and propositionalization settings with that of the logistic regression shown in Figure 9 illustrates that propositionalization settings with churn rate aggregates for all neighbors (T2.1 and T2.2) perform best and MLN settings yield only slightly better results than the benchmark for smaller samples.

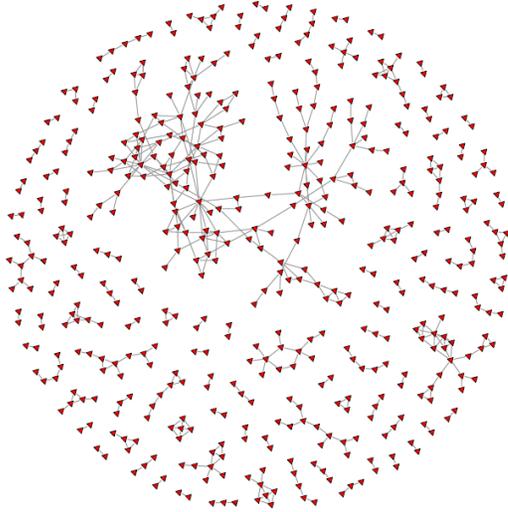
Overall, both propositionalization and MLN outperform the baseline model. Again, propositionalization performed surprisingly well, and MLN could not beat the performance of those models (T2.1).

5.2 Comparison of Results for Churn and Game Download

Interestingly, the impact of relational attributes was highly significant for predicting game download as well, but had a lower effect on sensitivity and the ROC curve compared to churn prediction. Just by adding information about the average churn rate of neighbors, the sensitivity of the churn prediction model increased by almost 20 percentage points. In contrast, the sensitivity in the game download model only increased by roughly eight percentage points. The impact on overall accuracy was almost equal (8.4 vs. 8.66 percentage points difference in accuracy in churn and game download, respectively). Looking at Figures 1 a) and b) illustrating the connections among the positives in both samples, this is somewhat surprising as the connections suggest that game download is more contagious than churn. Three observations from Figures 5 and 6 might provide an explanation. In Figure 5, we have plotted the communication edges among negatives ((a), (c)) and between positives and negatives ((b), (d)) for both churn and game download. In addition, the degree distributions are shown in Figure 6.

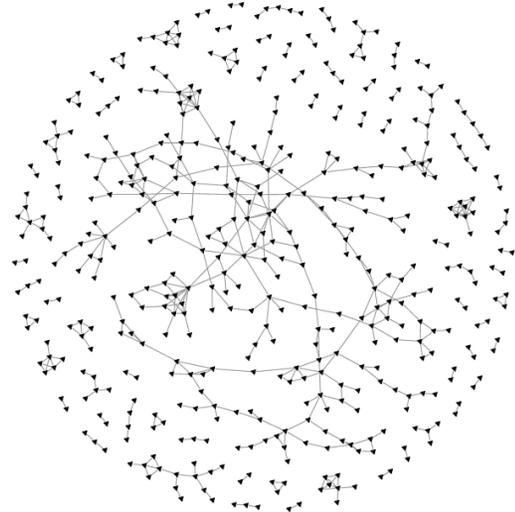
First, while positives are less connected than negatives for the churn scenario, it is the opposite way for game download. Second, there are more connections from positives to negatives than connections among negatives in the game download sample. In contrast, there are fewer connections from positives to negatives than from negatives to negatives in the churn sample.

Churn

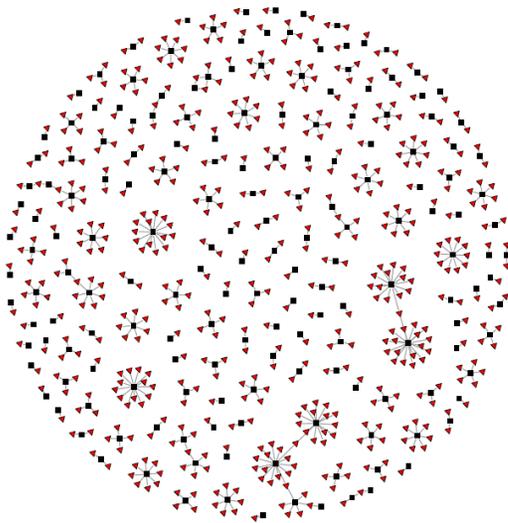


(a) Churn: negatives only

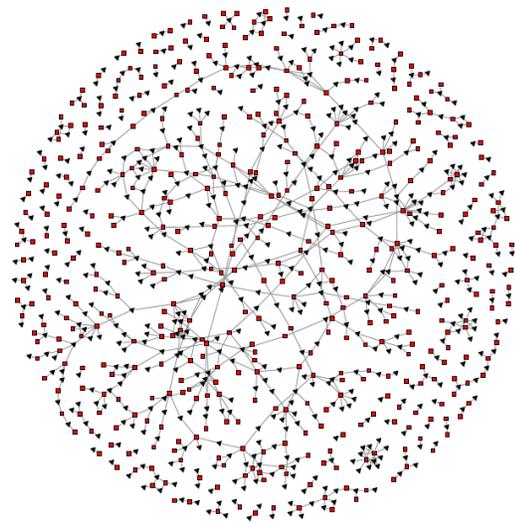
Game Download



(c) Game download: negatives only



(b) Churn: positives to negatives



(d) Game download: positives to negatives

Figure 5: Visualization of connections a) between negatives (without churn), b) from positives with game download (squares) to negatives (triangles), and c) between negatives (without game download), and d) from positives with game download (squares) to negatives (triangles)

Third, churn positives are less common than customers with a game download in the sample. Overall, positives are much less connected in the churn sample than in the game download sample. However, relatively speaking, there are more connections strictly among positives or negatives than connections between positives and negatives for churn than for game download, i.e. the set of positives and the set of negatives are relatively speaking more separated in the churn sample than in the game download sample.

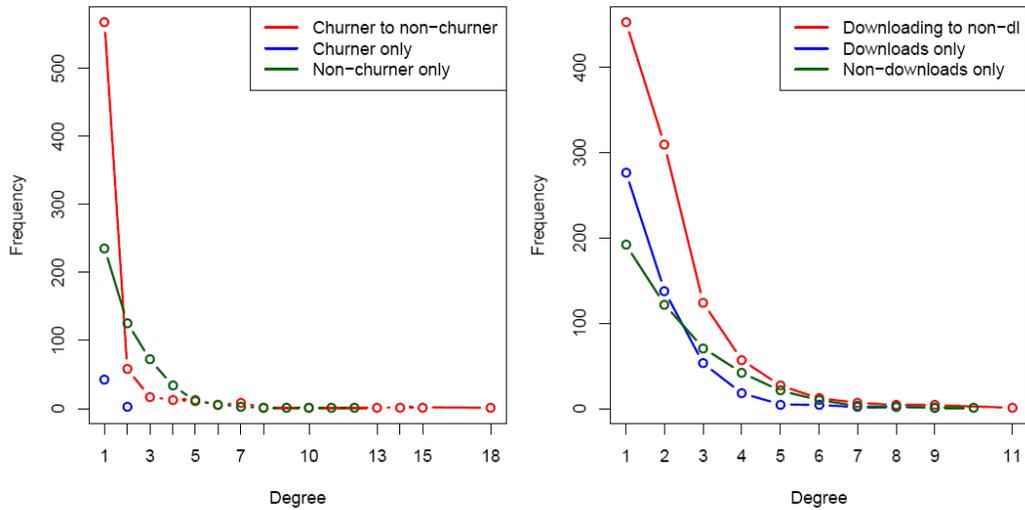


Figure 6: Degree distribution of churners (left) and game download customers (right)

In summary, churn was rare in our sample, and if there was churn in the neighborhood of a customer, this event was a powerful predictor for churn of this customer, as compared to game download, which is relatively widespread in particular communities. This provides evidence for the impact that word of mouth has on churn decisions.

6 Conclusions

Churn prediction is among the most important tasks of marketing departments in today's services industries, and improving the prediction of churn can have a considerable impact on a firm's profit, as effective customer retention measures can prevent customers from churning [50]. Word of mouth (WOM) has long

been recognized as a determinant for churn. Traditional discrete choice models, however, do not adequately allow the influence of peers through a social network to be modeled. Research in statistical relational learning has developed a number of new and powerful techniques. While original approaches such as Inductive Logic Programming have suffered from their high computational complexity, Markov Logic Networks allow for the analysis of larger data sets and can be considered a significant advancement in statistical relational learning. Social network modeling has been proposed as an application domain, although we are not aware of any publication where MLNs have been applied to large-scale social networks.

In this paper, we developed an MLN for churn prediction based on the anonymized data set of a mobile phone provider and found the MLN to have significantly higher predictive accuracy (+8%) and sensitivity (+19.7%) than the benchmark logistic regression. Similar results on accuracy were achieved when predicting game download, but the increase in sensitivity was lower. In addition, we have proposed a straightforward approach to the propositionalization of social network data. We found this approach to provide even better results than MLNs in terms of increasing sensitivity of the benchmark logit model. Note that the size of our dataset was restricted due to the computational effort of learning MLNs. Given that propositionalization is simpler, it is a viable alternative to marketers, in particular as it can be applied to much larger data sets than MLNs, which can lead to even better results.

From a marketing point of view, our results suggest that customer WOM significantly affects churn and cross-buying decisions of their neighbors in the mobile phone industry. To our knowledge, this is the first paper estimating the impact of WOM on churn. Empirical work measuring the effect of WOM is challenging. There is an issue of homophily and endogeneity due to other sources influencing the churn decision of a customer. For example, there might be other, unobserved friends or factors influencing a churn decision. Also, we only focus on dyadic relationships between a customer and his neighbors in the network and did not take into account communication among neighbors and the influence of communities of connected customers. For the churn sample, neighbors who churned were rarely connected, so that we do not expect a significant effect of this on our churn analysis. This might, however, matter in other situations and it is a worthwhile research question that we leave for future research.

Acknowledgements

We would like to thank Florian Wangenheim, the anonymous associate editor and the reviewers for their valuable comments.

7 Bibliography

- [1] W.-H. Au, K.C.C. Chan, X. Yao, A Novel Evolutionary Data Mining Algorithm with Applications to Churn Prediction, *IEEE Transactions on Evolutionary Computation*, 7(6) (2003) 532-545.
- [2] H.S. Bansal, P.A. Voyer, Word-of-mouth processes within a services purchase decision context, *Journal of Service Research*, 3(2) (2000) 166-177.
- [3] A. Berson, *Building Data Mining Applications for CRM*, (McGraw Hill, New York, 2000).
- [4] R.N. Bolton, A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction, *Marketing Science*, 17(1998) 45-65.
- [5] J. Chevalier, D. Mayzlin, The Effect of Word of Mouth on Sales: Online Book Reviews, *Journal of Marketing Research*, XLIII(2006) 345-354.
- [6] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. Nanavti, A. Joshi, Social Ties and their Relevance to Churn in Mobile Telecom Networks, in: *EDBT*, (Nantes, France, 2008).
- [7] L. De Raedt, H. Blockeel, L. Dehaspe, W. Van Laer, Three Companions for data mining in first order logic, in: S. Dzeroski, N. Lavrac Eds. *Relational Data Mining*, (Springer, 2001).
- [8] P. Domingos, M. Richardson, Mining the Network Value of Customer, *Seventh International Conference on Knowledge Discovery and Data Mining*, (2001) pages 57--66.
- [9] P. Domingos, M. Richardson, Markov Logic Networks, *Machine Learning*, 62(2006) 107-136.
- [10] S. Dzeroski, Data Mining in a Nutshell, in: S. Dzeroski, N. Lavrac Eds. *Relational Data Mining*, (Springer, 2001), pp. 3-27.
- [11] S. Dzeroski, N. Lavrac, *Relational Data Mining*, (Springer, Berlin, 2001).

- [12] R. East, W. Lomax, R. Narain, Customer Tenure, Recommendation and Switching, *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 14(2001) 46-54.
- [13] J.B. Ferreira, M. Vellasco, M.A. Pacheco, C.H. Barbosa, Data Mining Techniques on the Evaluation of Wireless Churn, in: *European Symposium on Artificial Neural Networks*, (Bruges, Belgium, 2004), pp. 483-488.
- [14] N. Friedman, L. Getoor, D. Koller, A. Pfeffer, Learning Probabilistic Relational Models, *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, (1999).
- [15] L. Getoor, Multi-relational data mining using probabilistic relational models: research summary, in: A.J. Knobbe, D.M.G. Van der Wallen Eds. *Proceedings of the First Workshop on Multi-relational Data Mining*, (2001).
- [16] L. Getoor, Link Mining: A New Data Mining Challenge, *ACM SIGKDD*, (2003) 84 - 89.
- [17] L. Getoor, B. Taskar, Introduction to statistical relational learning, in, (MIT Press, Cambridge, Mass., 2007), pp. 586.
- [18] L. Getoor, B. Taskar, *Introduction to Statistical Relational Learning*, (MIT Press, Boston, MA, 2007).
- [19] N. Glady, B. Baesens, C. Croux, Modeling churn using customer lifetime value, *European Journal on Operational Research*, 197(1) (2009) 402-411.
- [20] D. Godes, D. Mayzlin, Using Online Conversations to Measure Word-of-Mouth Communication, *Marketing Science*, 23(4) (2004) 545-560.
- [21] D. Godes, D. Mayzlin, Firm-Created Word-of-Mouth Communication: Evidence from a Field Test, *Marketing Science*, 28(4) (2009) 721-739.
- [22] S. Gupta, V. Zeithaml, Customer Metrics and Their Impact on Financial Performance, *Marketing Science*, 25(6) (2006) 718-739.
- [23] D. Hawley, International Wireless Churn Management Research and Recommendations in: Y. Group (Ed.), (2003).

- [24] O. Hinz, M. Spann, Managing information diffusion in Name-Your-Own-Price auctions, *Decision Support Systems*, 49(4) (2010) 474-485.
- [25] D. Jensen, J. Neville, M. Hay, Avoiding Bias when Aggregating Relational Data with Degree Disparity, in: *Twentieth International Conference on Machine Learning (ICML-2003)*, (Washington DC, 2003).
- [26] S.M. Keaveney, Customer Switching Behavior in Service Industries: An Exploratory Study, *Journal of Marketing*, 59(1995) 71-82.
- [27] D. Kempe, J. Kleinberg, E. Tardos, Maximizing the Spread of Influence through a Social Network, in: *ACM CIKM*, (Washington, DC, USA, 2003).
- [28] K. Kersting, *An Inductive Logic Programming Approach to Statistical Relational Learning*, (IOS Press, Amsterdam, The Netherlands, 2006).
- [29] K. Kersting, L. DeRaedt, Bayesian logic programs, in: J. Cussens, A.M. Frisch (Eds.) *10th International Conference on Inductive Logic Programming*, (London, 2000).
- [30] J.-U. Kietz, Some lower bounds for the Computational Complexity of Inductive Logic Programming, in: *European Conference on Machine Learning*, (1993).
- [31] E. Kim, W. Kim, Y. Lee, Combination of multiple classifiers for the customer's purchase behavior prediction, *Decision Support Systems*, 34(2) (2003) 167-175.
- [32] Y.S. Kim, An intelligent system for customer trageting: a data mining approach, *Decision Support Systems*, 37(2) (2004) 215-228.
- [33] Y.S. Kim, Toward a successful CRM: variable selection, sampling, and ensemble, *Decision Support Systems*, 41(2) (2006) 542-553.
- [34] C. Kiss, M. Bichler, Identification of Influencers - Measuring Influence in Customer Networks, *Decision Support Systems*, 46(1) (2008).
- [35] A.J. Knobbe, A. Siebes, D.M.G. Van der Wallen, Multi-relational decision tree induction, in: *3rd European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, (1999).

- [36] M. Kon, Customer Churn: Stop it Before it Starts, *Mercer Management Journal*, 17(2004).
- [37] S. Kramer, N. Lavrac, P. Flach, Propositionalization Approaches to Relational Data Mining, in: S. Dzeroski, N. Lavrac Eds. *Relational Data Mining*, (Springer, 2001), pp. 262-291.
- [38] M.-A. Krogel, S. Rawles, F. Zelezny, P. Flach, N. Lavrac, S. Wrobel, Comparative Evaluation on Approaches to Propositionalization, in: *I.C.i.I.L. Programming (Ed.)*, (Springer, Berlin, 2003), pp. 142-155.
- [39] M.-A. Krogel, S. Wrobel, Transformation-Based Learning Using Multirelational Aggregation, in: C. Rouveirol, M. Sebag (Eds.) *Proceedings of the Eleventh International Conference on Inductive Logic Programming (ILP)*, (Springer, 2001).
- [40] V. Kumar, A. Petersen, R.P. Leone, How Valuable is the Word of Mouth?, *Harvard Business Review*, (October) (2007) 139-146.
- [41] N. Lavrac, S. Dzeroski, *Inductive Logic Programming: Techniques and Applications*, in, (Ellis Horwood, Chichester, 1994).
- [42] N. Lavrac, S. Dzeroski, M. Grobelnik, Learning nonrecursive definitions of relations with LINUS, in: *Fifth European Working Session on Learning*, (Springer, Berlin, 1991), pp. 265-281.
- [43] N. Lavrac, F. Zelezny, P. Flach, Relational subgroup discovery through first-order feature construction, in: S. Matwin, C. Sammut (Eds.) *Twelfth International Conference on Inductive Logic Programming (ILP)*, (Springer, 2002).
- [44] A. Lemmens, C. Croux, Bagging and Boosting Classification Trees to Predict Churn, in: *DTEW Research Report 361*, (2003), pp. 40.
- [45] J. Leskovec, L. Adamic, B. Huberman, The dynamics of viral marketing, in: *7th ACM Conference on Electronic Commerce*, (2006), pp. 228-237.
- [46] C.X. Ling, C. Li, *Data Mining for Direct Marketing: Problems and Solutions*, in: *4th International Conference on Knowledge Discovery and Data Mining*, (American Association for Artificial Intelligence, New York, NY, 1998).
- [47] R. Mattison, *The Telco Churn Management Handbook*, (2005).

- [48] D. Michie, D.J. Spiegelhalter, C.C. Taylor, Machine Learning, Neural and Statistical Classification, (Ellis Horwood, New York, 1994).
- [49] M.C. Mozer, R. Wolniewicz, D.B. Grimes, E. Johnson, H. Kaushansky, Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry, IEEE Transactions on Neural Networks, (2000).
- [50] S.A. Neslin, S. Gupta, W. Kamakura, J. Lu, C. Mason, Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models, Journal of Marketing Research, 43(2) (2006) 204-211.
- [51] J. Neville, D. Jensen, Relational Dependency Networks, in: L. Getoor, B. Taskar Eds. Introduction to statistical relational learning, (MIT Press, Cambridge, Mass., 2007), pp. 239-268.
- [52] C. Perlich, F. Provost, J.S. Simonoff, Tree Induction vs. Logistic Regression: A Learning-Curve Analysis, Journal of Machine Learning Research, 4(2003) 211-255.
- [53] J.R. Quinlan, Learning logical definitions from relations, Machine Learning, 5(3)(1990) 239-266.
- [54] F.F. Reichheld, Learning from Customer Defections, Harvard Business Review, 74(1996) 56-69.
- [55] M. Richardson, P. Domingos, Markov Logic Networks, Machine Learning, (2009).
- [56] D.A. Schweidel, P.S. Fader, E.T. Bradlow, A Bivariate Timing Model of Customer Acquisition and Retention, Marketing Science, 27(2008) 829-843.
- [57] M.J. Shaw, C. Subramaniam, G.W. Tan, M.E. Welge, Knowledge management and data mining for marketing, Decision Support Systems, 31(1) (2001) 127-137.
- [58] H. Shin-Yuan, W. Hsiu-Yu, Applying Data Mining to Telecom Churn Management, in: Asia Pacific Conference on Information Systems, (2004).
- [59] D.S. Sudaraman, K. Mitra, C. Webster, Word-of-mouth communications: a motivational analysis, in: E.J. Arnould, L.M. Scott Eds. Advances in consumer research, (Association for Consumer Research, 1998), pp. 527-531.

- [60] C. Sutton, A. McCallum, An Introduction to Conditional Random Fields for Relational Learning, in: L. Getoor, B. Taskar Eds. Introduction to statistical relational learning, (MIT Press, Cambridge, Mass., 2007), pp. 93-129.
- [61] M. Trusov, A. Bodapati, R.E. Bucklin, Determining Influential Users in Internet Social Networks, *Journal of Marketing Research*, 47(4) (2010) 643-658.
- [62] R. van der Lans, G. van Bruggen, J. Eliashberg, B. Wierenga, A Viral Branching Model for Predicting the Spread of Electronic Word of Mouth, *Marketing Science*, to appear(2009).
- [63] W. Van Laer, L. De Raedt, How To Upgrade Propositional Learners to First Order Logic: A Case Study, in: S. Dzeroski, N. Lavrac Eds. *Relational Data Mining*, (Springer Verlag, Berlin, 2001).
- [64] R. Venkatesan, V. Kumar, A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy, *Journal of Marketing*, 68(4) (2004) 106-125.
- [65] F. Wangenheim, Post-Switching Negative Word-of-Mouth, *Journal of Service Research*, 8(1) (2005) 67-78.
- [66] F. Wangenheim, T. Bayon, The Effect of Word of Mouth on Services Switching: Measurement and Moderating Variables, *European Journal of Marketing*, 38(9/10) (2004) 1173-1185.
- [67] F. Zeleny, N. Lavrac, Propositionalization-based relational subgroup discovery with RSD, *Machine Learning*, 62(1-2) (2006) 33-63.

8 Appendix

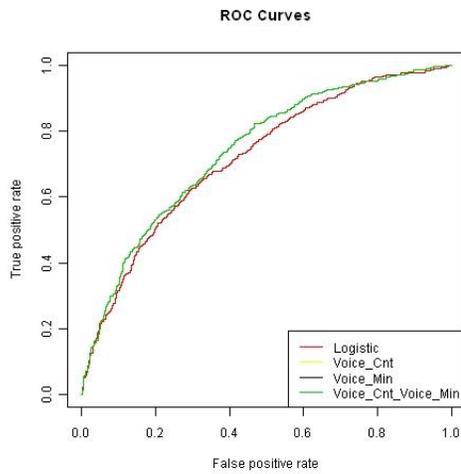


Figure 7: ROC curves of three MNL settings and the logistic regression for game download

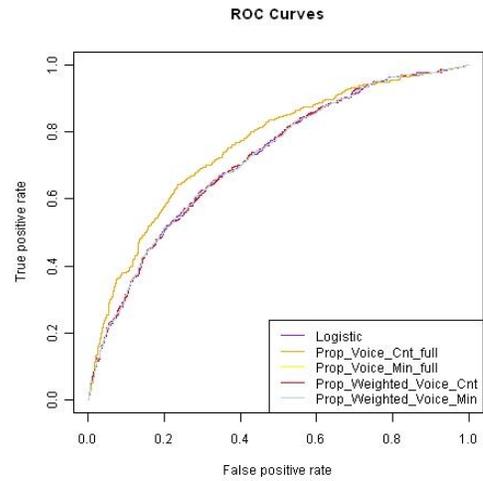


Figure 8: ROC curves of propositionalization settings and the logistic regression for game download

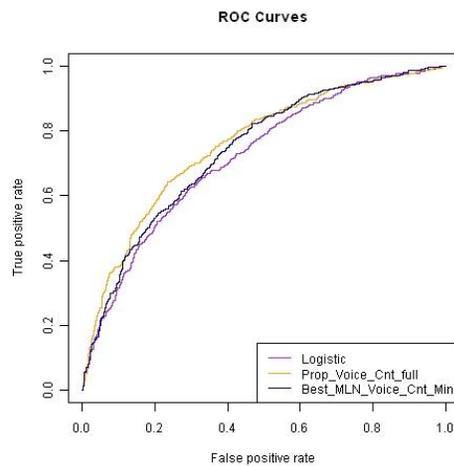


Figure 9: ROC curves of MLNs, propositionalization, and the logistic regression for game download