Applying data mining in telecommunications
Radosavljevik, D.

Author: Radosavljevik, D.
Title: Applying data mining in telecommunications
Issue Date: 2017-12-11

Chapter 3

Combining Customer Attribute and Social Network Mining for Prepaid Mobile Churn Prediction

Kusuma, P. D., Radosavljevik, D., Takes, F. W., van der Putten, P.

Published in Proceedings of the 22nd Belgian Dutch Conference on Machine Learning (Benelearn), pp. 50–58 (2013)

Customer churn, i.e. losing a customer to the competition, is a major problem in mobile telecommunications. This chapter investigates the added value of combining regular tabular data mining with social network mining, leveraging the graph formed by communications between customers. We extend classical tabular churn data sets with predictors derived from social network neighborhoods. We also extend traditional social network spreading activation models with information from classical tabular churn models, which did improve their performance. Nevertheless, the traditional tabular churn models scored best.

3.1 Introduction

Churn, which is defined as the loss of customers to another company, is a crucial problem in the telecommunication industry. As the telecom market has matured and opportunities for growth are limited, retaining existing customers has become a higher priority. In order to minimize the churn rate, mobile telecom players have to form defensive strategies to identify and present the appropriate incentive to subscribers with high churn propensity. The conventional churn models that exploit traditional predictors, such as demographic information (e.g., age, gender or location), contractual details (e.g., package plan type, contract duration or price), usage facts (e.g., voice call duration, the frequency of sending text messages) and/or other service related information (e.g., number of interactions with customer service or number of dropped calls), are typically simple and have a good predictive accuracy (Ferreira et al., 2004; Hadden, Tiwari, Roy and Ruta, 2006). However, the predictive accuracy of these models cannot be guaranteed if there is less customer data available, namely in the prepaid segment of the telecommunication industry.

This chapter investigates the extent to which social network features derived from the graph formed by communications between customers can be exploited to improve churn prediction accuracy in the prepaid segment. Examples of such features include the number of neighbors of a customer and the number of interactions that a customer has with churned neighbors. The research question for this chapter is:

Do social network mining or attributes stemming from a social network graph add value in terms of model performance to traditional prepaid churn modeling in T-Mobile Netherlands?

This research study was conducted at one of the largest telecom providers in the Netherlands, and a data set containing 700 million call records was used to assess the quality of the various techniques discussed throughout the chapter.

We propose two novel models for churn prediction. The first is a hybrid tabular model, which combines both traditional predictors and social network features to predict churn, aiming to gain significant lift. Logistic Regression and the CHAID algorithm are utilized to derive the tabular models. These churn models, however, do not take into account the influential effect of an individual's decision to his/her social network. Dasgupta et al. (2008) have been able to address this problem by constructing a churn model based on a traditional social network mining technique, i.e. spreading activation models. Their model propagated the negative churn influence from one subscriber to another in a cascade manner. Besides building hybrid tabular churn models using a combination of the traditional predictors and the social network features, we also propose a second approach, which extends the traditional propagation model to include the output by traditional churn models.

The rest of the chapter is organized as follows. Section 3.2 presents some related work within the field of churn prediction. Section 3.3 discusses the call graph and proposed algorithms. The research setup and the empirical models are introduced in Section 3.4. In Section 3.5, the experimental results and implications of all scenarios are presented. Finally, Section 3.6 summarizes the chapter and presents some suggestions for future work.

3.2 Related Work

Churn has been widely analyzed not only in the telecommunication industry (Ferreira et al., 2004; Hadden et al., 2006; Radosavljevik et al., 2010*a*), but also, among others, in the online gaming industry (Kawale, Pal and Srivastava, 2009) and banking (Prasad and Madhavi, 2012). Many machine learning techniques, such as decision trees, naive Bayes, logistic regression, neural networks and genetic algorithms, are often used to build the tabular churn prediction models.

Ferreira et al. (2004) utilized contractual and demographic information of a Brazilian mobile telecommunication provider to build several postpaid churn models using neural networks, decision trees, genetic algorithms and hierarchical neuro-fuzzy systems. Besides evaluating the predictive power, they also assessed the profitability value of those models, claiming that even the churn models with the worst performance are still able to save significant cost in the postpaid segment. Hadden et al. (2006) exploited provisions, complaints and repair interaction data to build churn models. They claimed that the regression tree model performed better than models built using neural networks or logistic regression. However, there is no further information regarding the performance comparison between the complaints-based model and the benchmark model based on demographic and contractual variables.

In chapter 2, we investigated the extent to which Customer Experience Management (CEM) data could improve prepaid churn prediction. Several Key Performance Indicators (KPI) of service quality combined with other subscriber data were used to train the decision tree models. Since the CEM data was always available, the constraint on lacking demographic information on the prepaid subscribers could be eliminated. Although some of the CEM variables were predictive, the empirical study showed that there was insufficient gain on this model's performance compared to the benchmark.

Several social network studies have been conducted by utilizing mobile call graph data to examine the structure and evolution of social networks (Backstrom, Huttenlocher, Kleinberg and Lan, 2006; Seshadri et al., 2008), the human mobility patterns

(Gyan, Hui, Zhi-Li and Jean, 2012) and their social interactions (Dasgupta et al., 2008). Dasgupta et al. (2008) analyzed the influential impact of the churned neighbors to their social circle by applying a spreading activation-based technique similar to trust metric computations (Ziegler and Lausen, 2004). Using call graph data, they were able to show that churn can be propagated through a social network. Although the study was limited to using social ties information only, reasonable predictive accuracy could still be achieved. Their analysis identified that the churn propensity of a subscriber correlates positively with the number of churned neighbors.

Kawale et al. (2009) conducted a similar study using social network data from a popular online gaming community. They proposed a new twist to the existing churn propagation model proposed by Dasgupta et al. (2008) by combining the social influence and user engagement in the game. The user engagement property, which refers to the length of the playing session during the observation period, can be classified as an intrinsic variable. This research showed that the models trained using a combination of social factors and this user engagement property performed better than traditional propagation models. Using collective classification techniques, Oentaryo, Lim, Lo, Zhu and Prasetyo (2012) were also able to demonstrate that the churn prediction accuracy could substantially be improved by utilizing the combination of traditional user profile and social features.

We applied ideas similar to the above mentioned works. A customer's decision to churn might not only depend on the social influences but also on how they perceive the products and services. Initially we found that the ratio of the immediate churned neighbors to the number of adjacent neighbors (degree) positively correlates to the churn behavior. When half of the neighbors have churned, the probability of a subscriber to churn is two times higher than the baseline churn rate. It implies to some extent that social behavior might have an impact on subscribers' churning decision. It could be that the hybrid models, which exploit both traditional predictors and social relationships, could outperform the simple social network and the tabular churn model built exclusively using traditional predictors. However, a question should be raised whether this adds actionable value over existing data. We suspected there may have been some element of publication bias: positive results get published more often, thus easier to find than non-significant or negative results, at least for trending topics. Hence, we decided to evaluate the business value experimentally. A call graph can be derived from raw data of communications between customers. This graph, further discussed in Subsection 3.2.1, is essentially a social network which can be leveraged in two ways. Classical "tabular" models are built on rectangular data sets, one row per customer with subscriber level information. This can be simply extended with attributes (columns) that contain information derived from the social network, as we will outline in Subsection 3.2.2. Likewise, a traditional approach to modeling social network dynamics is the spreading activation model, which can be used to model how customer behavior such as churn spreads over the network. Insights from traditional tabular models, more specifically churn scores, can be used

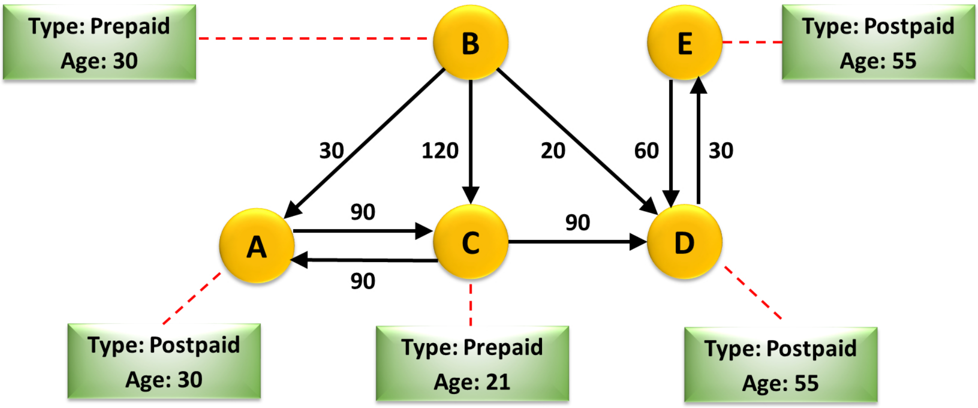

Figure 3.1: Telecom call graph.

to improve these classical social network models, a technique on which we will elaborate in Subsection 3.2.3.

3.2.1 The Call Graph

The *call graph* can be constructed from the Call Detail Records (CDRs) provided by the telecom provider. These CDRs contained detailed facts about mobile interactions, such as source phone number, destination phone number, the type of mobile communication, duration and a timestamp. This information is mapped to a directed social graph (Borgatti, 1994) $G = (V, E)$ as illustrated in the Figure 3.1.

In this call graph, *nodes* denote subscribers and an *edge* represents a mobile interaction between two subscribers. The *edge weight* can be calculated from one variable or a combination of interaction variables, e.g., voice call duration or SMS frequency. It could indicate the interaction intensity or the relationship strength between two nodes.

As several interactions could exist between the same pair of nodes, we treated duplicate edges between two nodes as a single edge, by aggregating the weight values. The aggregation method applied in this research is explained in Subsection 3.2.3.

We used Neo4j technology to store the graph structure and the content of graph elements. Neo4j differs from relational database management systems, as it is oriented to store semi-structured and network data, which makes it appropriate to store social graphs (Neo4j, 2012; Kusuma, 2013). It also provides an intuitive representation of the graph and it is easy to traverse through the graph's nodes and relationships. The scalability of this system presents a great advantage because its functionality can be easily extended to perform a large scale social network analysis.

Table 3.1: Social network features used in the extended tabular churn models.

CATEGORY	VARIABLE
CONNECTIVITY	Count of in/out-degree
	Sum & average of in-/out-weight
	Count & average of voice, SMS & voice+SMS to/from neighbors
	Total and average of edge weight*
	Total interaction frequency with neighbors*
	Total and average frequency with neighbors for voice & SMS separately*
	Degree, 2nd degree & 3rd degree count*
CHURNER CONNECTIVITY	Count of in/out-degree churners
	Sum & average of in/out-weight with churners
	Count & average of voice, SMS & voice+SMS to/from churners
	Total & average edge weight with churners*
	Total interaction frequency with churners*
	Ratio of in/out-degree churners to the total in/out-degree
	Ratio of in/out-weight churners to the total in/out-weight
	Ratio of in/out voice, SMS & voice+SMS frequency with churners to the total in/out-weight
	Ratio of churner weight to the total weight*
	Ratio of interaction frequency with churners to the total interaction frequency*
	Churner degree, 2nd & 3rd degree count*
	Ratio of churner degree to the total degree*
	Ratio of 2nd churner degree to the total 2nd degree*
	Ratio of 3rd churner degree to the total 3rd degree*

*direction is not taken into account

3.2.2 Extended Tabular Churn Models

Many tabular churn models generally exploit either subscriber profile information or social network statistics separately. The predictive power of churn models based merely on the traditional predictors might be reduced in case of many missing values. In our prepaid churn study, we only had access to limited demographic data because prepaid subscribers are not required to fill in their (accurate) personal information. On the other hand, the social network features might not be predictive enough to influence the churn decision. Neither the traditional models nor the models based exclusively on social networks can cover all aspects of churn on their own.

Therefore, we propose to combine both elements to predict churn, adding the features listed in Table 3.1.

When creating the extended tabular churn models we started with a model based on traditional predictors and added connectivity features from the social network call graph: the in-degree and out-degree, the number of second degree neighbors,

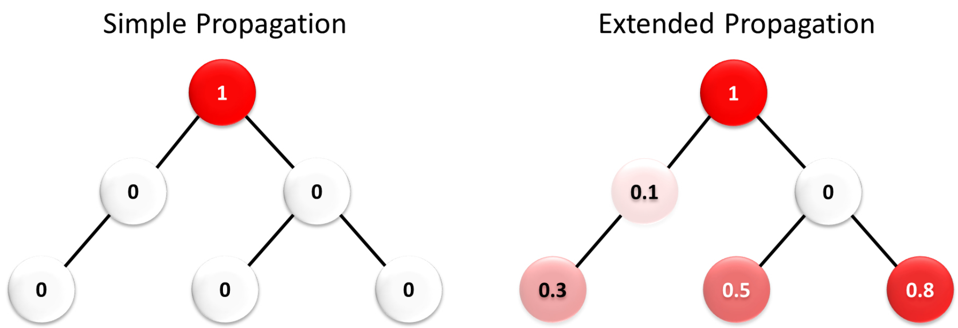

Figure 3.2: Initial energy of the simple and extended propagation technique.

sum and average of in-weight and out-weight calculated from duration of voice conversations, SMS and a combination thereof. We also added churn connectivity variables (in-degree and out-degree with churners, etc.), as well as the ratios of the total connectivity measures vs. the churners connectivity measures. A detailed overview of the added social network graph features is presented in Table 3.1. For a more detailed feature analysis, we refer the reader to Kusuma (2013).

3.2.3 Extended Social Propagation Models

In this subsection, we discuss an extension of the spreading activation model to measure how churn is diffused around telecom social network (Dasgupta et al., 2008). The churn propagation process begins by initialization of all nodes. In this study, we set the energy of *non-churners* using two different values (see Figure 3.2). For the *simple propagation approach*, the initial energy of non-churners was set to 0; for the hybrid *extended approach*, it was set to the churn score returned from the regular tabular models.

In the propagation process, for a node $x \in V$, the value of $En(x)$ represents the current amount of energy of a node, and the $En(x, i)$ represents the amount of energy or social influence transmitted to the node x via one or more of its neighbors at stage i (Dasgupta et al., 2008). After energy initialization, a set of previous churners (seeds) is activated. In stage 0, the current energy of the seeds $En(x)$ is used as initial spreading value. Therefore, the current energy value $En(x)$ becomes 0 and amount of energy in a node x at step 0 or $En(x, 0)$ becomes equal to 1.

In each consecutive stage i, the activated nodes transfer a portion of their energy to their neighbors and retain certain portion for themselves. The spreading factor $\delta \in [0, 1]$ controls the proportion of the transmitted energy, denoted by $\delta * En(x, i)$ and the amount of retained energy $(1 - \delta) * En(x, i)$. A spreading factor value of

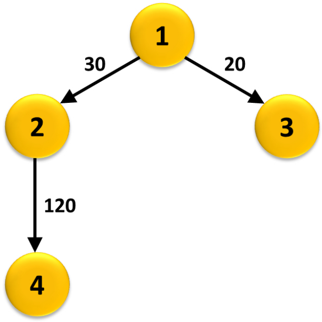

Figure 3.3: Spreading activation in a weighted graph.

$\delta = 0.8$ means that 80% of the energy is transferred to the neighboring nodes and 20% of the activated energy is retained by the node. This factor value could also be seen as a decay measure because the transferred energy will decline as it gets further away from the source. It implies that the direct neighbors will receive more influence than second degree neighbor and so on. The trust propagation study of Ziegler and Lausen (2004) has shown that people tend to trust individuals trusted by own friends more than individuals trusted only by friends of friends.

Since nodes can have multiple neighbors, the amount of the distributed energy from an active node to each neighbor depends on the tie strengths between the node pair. In Figure 3.3, for example, the amount of energy transferred from node 1 to node 2 might not be the same as the amount transferred from node 1 to node 3, because the edge weights are not equal.

Let y be a neighboring node of an active node x (with $x, y \in V$). We denote the amount of energy transferred from node x to node y in the i-th stage with $En(x, y, i)$. This amount depends on the relative edge weight of the paired nodes. This is determined by a transfer function $f(x, y)$, described in equation 3.3 below. The amount of energy transferred is then:

$$En(x, y, i) = \delta * En(x, i) * f(x, y) \tag{3.1}$$

The amount of energy of node x after the spreading computation is as follows:

$$En(x) = En(x) + (1 - \delta) * En(x, i) \tag{3.2}$$

There are multiple functions to determine the relative weight between two nodes. The simplest method is using linear edge weight normalization function (Ziegler and Lausen, 2004).

$$f(x, y) = w(x, y) / \sum_{(x,z)} w(x, z) \tag{3.3}$$

Here, $f(x, y)$ denotes the relative weight of the edge between x and y, $w(x, y)$ represents the weight of that corresponding edge, and $\sum_{(x,z)} w(x, z)$ represents the total weight of all edges connecting node x to its adjacent nodes.

We propagated the churn energy through both a directed and an undirected version of the graph. In the directed graph, energy is propagated only to outgoing edges, and in the undirected graph, both outgoing and incoming edges are used. For churn propagation, the remaining energy after termination ultimately determines the probability of a network member to churn. These churn probability scores are then distributed into score intervals. The upper interval groups contain more subscribers with high churn propensity behavior compared to the lower interval groups. Using the threshold score-based technique, the subscribers/groups with churn scores above a predefined threshold score can each be labeled as a ″churner″, and otherwise as a ″non-churners″. As an alternative, a cut-off point can also be determined by specifying the target group size.

3.3 Experimental Setup

This section describes our operational definition of churn in Subsection 3.3.1, after which the data set and weighting technique is discussed in Subsection 3.3.2. We then give an overview of the seven different scenarios that were used to construct the churn models, outlining our experimental setup in Subsection 3.3.3.

For our experiments, we used the software Predictive Analytics Director (Pegasystems, 2008) to automate variable discretization, variable selection and grouping, to train the scoring models and also to compare the models performance. The default evaluation statistic that is used to measure the performance of the predictors and models is Coefficient of Concordance (CoC) (Kendall, 1938). As explained in chapter 2, CoC measures the area under the ROC curve formed by the percentage of cases with positive behavior against the percentage of cases with negative behavior for each unique score (Harell, 2001).

3.3.1 Operational Definition of Churn

We constructed models for both prepaid and postpaid telecom segments. Although the definition of churn is different for each segment, we will only discuss the prepaid results because both studies have come to similar conclusions. Unlike postpaid subscribers, prepaid subscribers are not bound by a contract, which makes it easier for them to churn. Prepaid subscribers need to purchase a credit voucher before using any telecom service. If they do not have sufficient credit, they cannot initiate any calls, send SMS/MMS or connect to internet. They could re-enable the service by recharging or topping-up their credit.

A prepaid subscriber is disconnected from the network and he/she is marked as a churner after six consecutive months of inactivity. A prepaid activity could be translated to an outbound voice call, an inbound voice call, an outbound SMS, a data usage or a commercial voucher recharge, also known as top-up. As churn should be detected as early as possible, the disconnection date might not be an appropriate

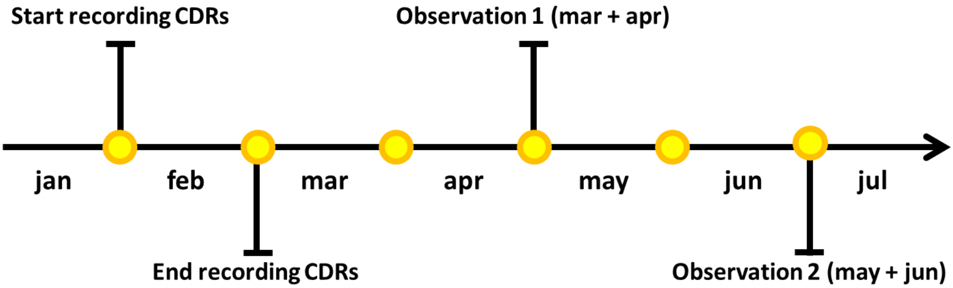

Figure 3.4: Call Graph Details.

churn date measure (Kraljevic and Gotovac, 2010). The prepaid subscribers might be long gone before they are actually disconnected from the network. Therefore, we define churn as two consecutive months of inactivity, or more. This is the same definition of churn we used in Chapter 2 (Definition 2). It was aligned with many internal studies that are conducted within the company.

3.3.2 Data Set

We used the CDRs from the whole month of February 2012, which is roughly about 700 million records, to construct the social graph. We included subscribers who have at least one call in February and we based our social network graph on the interactions that occurred in that month. The traditional predictors were also collected in this period.

We assumed that churn is also a social networking phenomenon, thus subscribers that communicated with people that have churned are more likely to churn themselves. Therefore, we labeled the nodes/subscribers that churned in the period before May 1, 2012 ('observation 1' in Figure 3.4) as seeds/churners of the propagation graph explained in Subsection 3.2.3. These are not the churners we are trying to predict.

The end goal was to use the traditional predictors as well as the social network information obtained in February 2012 to predict churn at the end of June 2012 ('observation 2' in Figure 3.4). This is a different experimental setup than the one described in chapter 2. In this research, we were trying to predict churn four months after the initial data recording. In Chapter 2, this period was shorter: it was set to two months.

In this research study, we only used the duration of voice calls in minutes and the count of text messages to construct the social graph. We could not explore mobile interactions utilizing the data connection, i.e. using over the top (OTT) services[1], due

[1] An over the top service is utilizing the telecom network to perform. However, it does not require any explicit affiliation with the network provider. Examples of over the top applications are WhatsApp, Skype or Viber.

to legal issues. Within the company, the postpaid cost of making one minute of a voice call was the same as one SMS. In the prepaid segment, one SMS was typically charged roughly the same as half of a minute of voice call. Therefore, we made the assumption that a text message is equivalent to a voice call of 30 seconds. Hence, we could generalize the edge weight $w(x, y, t)$ between a pair of nodes x and y at time t to include both types of mobile communication, voice calls and SMS and all interactions could all be measured uniformly in seconds. The identifier t represents the hourly timestamp at which the interaction starts, and is ranged from 1 until 29 February 2012.

$$w(x, y, t)' = w(x, y, t) * \begin{cases} 1, & \text{if voice call} \\ 30, & \text{if SMS} \end{cases} \tag{3.4}$$

Interactions that occurred outside working hours are assigned twice the weight to emphasize their importance. The underlying assumption here was that interactions within working hours mostly indicate communication of professional nature, whereas interactions outside working hours may involve communication of more personal nature (e.g., friends, family), which could have higher influence on the decision to churn. Motahari et al. (2012) showed that members of a family/friends social network are more likely to call each other on the weekend and the engagement ratio value within the family/friends network is at least twice as much compared to the rest of the population. Therefore, we introduce a weight scale $\rho(t)$, which is defined as follows:

$$\rho(t) = \begin{cases} 1, & \text{if t=weekdays (8-17)} \\ 2, & \text{otherwise} \end{cases} \tag{3.5}$$

$$w(x, y, t)'' = \rho(t) * w(x, y, t)' \tag{3.6}$$

We also assumed that a recent interaction should carry more weight than older ones. Therefore, the daily decay rate $\alpha = 0.2$ was manually selected. The weight value of an edge that is measured on a certain day exponentially decayed according to a predefined rate as follows:

$$w(x, y, t)''' = w(x, y, t)'' * e^{-\alpha * d} \tag{3.7}$$

Here, the symbol d corresponds to the gap measured in days between the interaction timestamp and the end of the observation period. At the end of the observation period, the weight values are aggregated. As a result, each node pair could only have maximum one edge in each direction, so two edges in total. The equation below formulates the aggregation process of the weight values.

$$w(x, y) = \sum w(x, y, t)''' \tag{3.8}$$

For an undirected graph, we could simply add up the weights for both directions together as follows:

$$w(x, y) = \sum w(x, y, t)''' + \sum w(y, x, t)''' \tag{3.9}$$

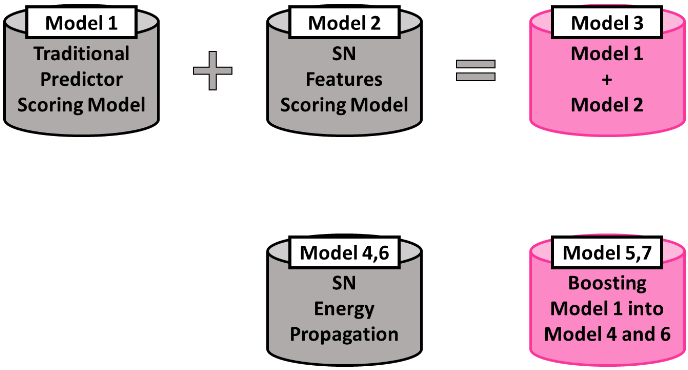

Figure 3.5: Implementation scenarios.

3.3.3 Churn Predictive Models

To investigate to which extent social network data could be used to predict churn and possibly improve churn prediction performance, we trained three tabular data mining models using scoring algorithms and four social network models using a spreading activation algorithm (see Figure 3.5).

Scoring Models

We applied logistic regression and a CHAID decision tree algorithm (Witten and Frank, 2005) to train our three scoring models:

- Model 1: simple scoring model
- Model 2: social network (SN) scoring model
- Model 3: extended scoring model

Model 1, a simple scoring model, was trained using traditional churn predictors, namely features such as prepaid credit, handset and usage information. We employed this model as the benchmark model. Model 2 was a social network scoring model, which focused solely on the social network attributes extracted from the call graph, such as the number of incoming and outgoing ties of the first and second degree neighbors. The extended scoring model, Model 3, was built by using the combined data set of the first and the second model. This hybrid model was trained using both traditional churn variables and social network features.

Propagation Models

The remaining four models were trained using energy propagation techniques based on the previously discussed spreading activation algorithm:

- Model 4: simple propagation model
- Model 5: extended propagation model
- Model 6: simple propagation model undirected
- Model 7: extended propagation model undirected

Churners in April 2012 were used as the source of the energy propagation. Each churned node was given an initial energy of 1. Model 4, which was a simple propagation model, set the initial energy of non-churners to 0. Model 5 was a hybrid model created by boosting of Model 1 into Model 4. It incorporated subscribers' intrinsic churn information into the propagation model. Instead of setting the energy of non-churners to 0, this model assigned the churn score obtained from Model 1 as the initial energy of the non-churner nodes. The intuition behind this idea is that a subscriber might already have a certain tendency to churn due to his/her experience with the provided service. Model 6 and Model 7 are similar to Model 4 and Model 5 respectively, except that those models were trained using an undirected instead of a directed graph.

The total energy value that remained after termination is assumed to be the probability of a network member to churn. To study the influential effect of churned neighbors in the social network, we then compared the propensity values of non-churners to the actual known churn class.

3.4 Results

In this section, we report the empirical result for each of the implementation scenarios (see Table 3.2 and Figure 3.6). We present and discuss only the scoring models based on decision trees, because these models had a slightly better predictive performance compared to the ones built using logistic regression. Moreover, we only include propagation models with the spreading factor that yield the best prediction results. The performance of any of these models cannot be compared with the models described in chapter 2, due to a difference in the experimental setup. As explained in Section 3.3.2, in this chapter we were trying to predict churn further into the future compared to the experimental setup in chapter 2, which makes the prediction task more difficult.

Model 3, which is the hybrid model that combined tabular churn predictors and social network variables derived from the social network graph, had the highest CoC score on the test set (64.98). Since it only slightly outperformed Model 1 (64.88), we can conclude that adding social network features on top of the traditional churn predictors did not appear to provide a substantial improvement for our scoring model. Model 2 built solely using social network predictors had the lowest predictive accuracy compared to the rest of the scoring models (56.57). By targeting the top 30% of the subscribers, Model 2 could find only 37% of the churners, while Model 1 and Model 3 were able to return about 50% of the churners. The lift chart shows that in

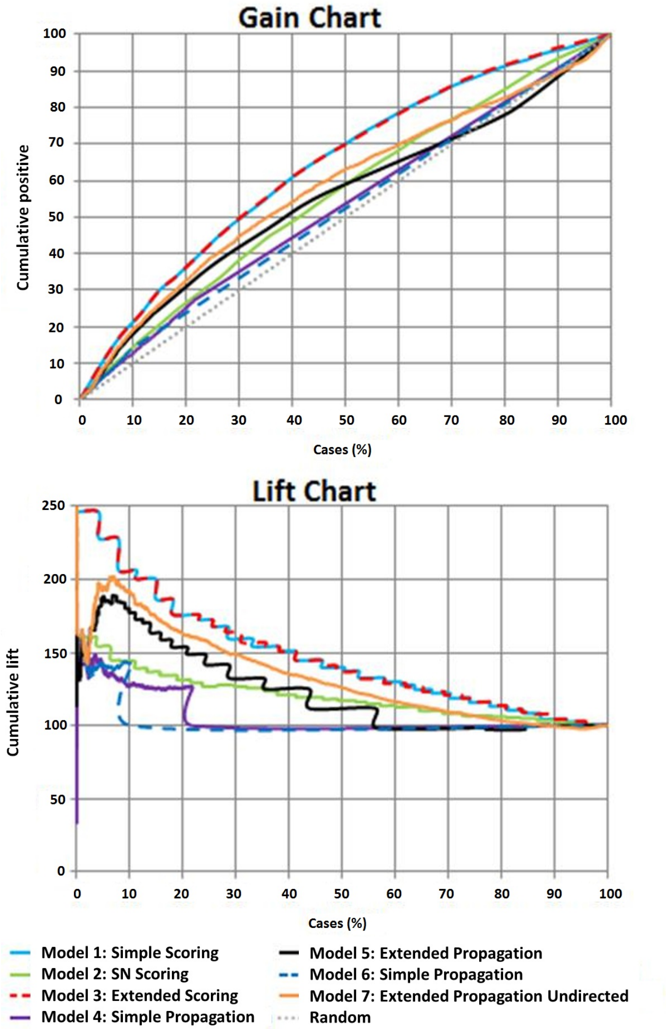

Figure 3.6: Gain and Lift chart of all models.

Table 3.2: Coefficient of Concordance of the scoring and propagation models.

Model Description	Performance on		
	Train Set	Validation Set	Test Set
Model 1- simple scoring model	65.48	64.47	64.88
Model 2- social network (SN) scoring model	57.93	56.72	56.57
Model 3- extended scoring model	65.65	64.45	64.98
Model 4- simple propagation model	53.34	53.43	53.04
Model 5- extended propagation model	55.26	54.58	55.24
Model 6- simple propagation model undirected	52.07	52.15	52.26
Model 7- extended propagation model undirected	58.39	57.66	58.30

the top 30% of the cases Model 2 had cumulative lift of 130%, whereas other scoring models had cumulative lift of 160%. In other words, the information derived from the social network was weakly predictive by itself and it failed to outperform the predictive power of the traditional predictors.

As expected, the extended propagation models (Model 5 and Model 7), which incorporated churn scores of the simple scoring model as the initial energy value in the propagation process, outperformed the traditional social network propagation models (Model 4 and Model 6). These extended or hybrid models provided better predictive accuracy than the simple propagation models for the directed and the undirected graph. By targeting 30% subscribers, Model 7 was able to correctly predict about 45% churners. It returned 5% less than the tabular churn models, Model 1 and Model 3. Although Model 7 incorporated the traditional predictor elements in the propagation process, the predictive power was still lower than that of the traditional tabular churn scoring models.

The simple propagation models that incorporated only the social neighborhood information, Model 4 and Model 6, had even lower performance compared to Model 2. Unlike Model 2, the simple propagation models used only the previous churner information within the social network without considering the individual churn propensity. This led us to believe that the churning behavior of neighbors does not have substantial influence on other members within a prepaid telecom subscriber social network. Traditional churn predictors apparently had a stronger influence on churn compared to social relationships.

3.5 Conclusions and Future work

Throughout this chapter we have investigated the extent to which social network information can be used to predict telecom churn, and how this information could potentially improve the predictive performance of conventional churn prediction methods. We have assessed the performance of models constructed using classical tabular data mining, social network mining and hybrid models combining both techniques. The first hybrid model was built by extending traditional tabular churn predictors with social network variables extracted from the social graph. The second hybrid model was obtained by incorporating the results of a traditional tabular churn model into the social propagation graph, using them as initial energies of the non-churner nodes.

The performance of our models was verified using a large data set of 700 million call data records. Our initial observation showed that the churn probability was positively aligned with the number of churned neighbors. The regular tabular churn models constructed exclusively using social network information and the traditional social network models scored the least. This indicated that social network information alone is not sufficient to predict churn. Overall, the traditional tabular churn models had the best predictive performance. The added value of the social network variables to the tabular churn models was rather minimal. Although the second hybrid models were able to outperform the regular propagation models, they still could not beat the performance of the traditional tabular churn models. The contribution of traditional predictors to churn prediction was substantially higher than that of the social network behavior. Moreover, the performance gain of both hybrid models was not substantial enough to justify the computational costs. In a nutshell, the answer to the research question posed in section 3.1 is that social network mining and attributes stemming from a social network graph did not add substantial value in terms of model performance to traditional prepaid churn modeling in T-Mobile Netherlands. This was in in contrast to most of the statements made in research literature, but it did not come as a surprise to us because we were suspecting that in the other cases the models might have not been benchmarked well enough against standard models based on rich data, and that there may have been some instances of publication bias.

The current research study only explored the negative influential effect of previous churners within the social network. Future research could potentially be focused on removing this limitation. The influences from both churners and non-churners could be taken into account, as subscribers might spread messages based on how they perceive the product/service quality. Assuming bad news can have a stronger influential effect than good news, positive influence from non-churners to stay within the network might not be as strong as negative influence from churners. Since our energy propagation model was purely derived from node and neighborhood-based relationships, the spreading activation computations were done locally and subscribers did not have knowledge beyond their direct neighbors. Other algorithms, for example from the field of community detection, are capable to identify the role of subscribers within the social network, such as influencer or adopter. Rather than targeting all future churners, we can minimize our resources by focusing only on churners with high influential power.