Evaluation of customer behavior with temporal centrality metrics for churn prediction of prepaid contracts.

Laura Calzada-Infante[a]*, María Óskarsdóttir[b], Bart Baesens[c,d]

[a] Faculty of Economics and Business, University of León, Campus de Vegazana 24071, León, Spain, lcali@unileon.es

[b] Department of Computer Science, Reykjavík University, Menntavegi 1, 102 Reykjavík, Iceland, mariaoskars@ru.is

[c] Faculty of Economics and Business, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium, bart.baesens@kuleuven.be

[d] Dept. of Decision Analytics and Risk, University of Southampton, United Kingdom

*Corresponding author:

Address: Facultad de Economía y Empresa

                Campus de Vegazana,

                24071, León, Spain

E-mail: lcali@unileon.es

Phone: +34 987 291000/5480

ABSTRACT

The telecommunication industry is a saturated market where a proper implementation of a retention campaign is critical to be competitive, since retaining a customer is cheaper than attracting a new one. Hence, it is crucial to detect customer behavioral patterns and define accurate approaches to predict potential churners. Multiple researchers have used binary classification methods to predict churn of customers. Some of them verify that customers' social relationships influence the decision of changing the operator.

We propose a novel method to extract the dynamic relevance of each customer using social network analysis techniques with a binary classification method called similarity forests. The dynamic importance of each customer is determined by applying various centrality metrics over temporal graphs, to represent the relationships between customers and to extract behavioral patterns of churners and non-churners. These relationships are established in a temporal graph using the call detail records (CDR) of telco's customers. In this paper, we compare the performance of different centrality metrics applied over two types of temporal graphs: Time-Order Graph and Aggregated Static Graph.

Keywords: Churn prediction; Social Network Analysis; Similarity Forests; Time Series; Centrality Metrics

1. INTRODUCTION

The mobile telecommunication industry (telco) is rapidly expanding. In 2017, two-thirds of the world population were connected by mobile devices, and it is estimated that almost 75% of the global population will be connected through mobile phones by 2020. Furthermore, the milestone of five billion mobile phone customers was achieved in 2017 (Hollander, 2017). The telco market is reaching saturation, as most people have a mobile phone. At the same time, it is relatively easy for customers to change telco operators if they become unhappy with their current services or if the competitors offer a better plan. In this context, telco operators have a high interest in applying effective customer retention campaigns. The main reasons are that retaining customers is five times less expensive than attracting a new customer, and long-term customers are more profitable due to their satisfaction, increased likelihood of buying other products and attracting new customers as well. (Kotler & Keller, 2006, Ganesh, Arnold, & Reynolds, 2000). As a result, retention campaigns are one of the main activities of customer relationship management (CRM) departments (Owczarczuk, 2010) where there is a high demand for effective and complete solutions that are capable of detecting potential churners before it is too late.

The development of new methodologies to detect potential churners has been an active research field for some years. It is typically viewed as a classification problem and various data mining techniques have been proposed. Verbeke, Dejaeger, Martens, Hur, and Baesens (2012) provide an extensive overview of these techniques. For example, decision-rule based classifier (Amin *et al.*, 2017); decision tree approaches (Wanchai, 2018, Caigny, Coussement & Bock., 2018); neural networks (Vafeiadis, Diamantaras, Sarigiannidis & Chatzisavvas, 2015), nearest neighbor, ensemble methods (Óskarsdóttir, Bravo, Verbeke, Sarraute, Baesens & Vanthienen, 2017), classic statistical methods like logistic regression (Wanchai, 2018), or support vector machine (SVM) (Vafeiadis *et al.*, 2015) are some of the main techniques that have been used. These techniques use a wide variety of features to detect churners. Recently, call detail records (CDR) have become a popular data source for churn prediction. These logs of phone calls between the telco's customers contain information about people's calling behaviour and have been used to build social networks and show that churn is a social phenomenon, in the sense that the propensity of the customers to churn can be studied based on their churned friends (Dasgupta, Singh, Viswanathan, Chakraborty, Mukherjea, & Nanavati, 2008). The information in the CDR in combination with other data and machine learning techniques, such as linear classification, logistic regressions, decision trees, naive Bayes, multilayer perceptron neural networks, evolutionary data mining algorithm and support vector machines, enables the detection of more churners (Huang, Kechadi, & Buckley, 2012). Moreover, Dasgupta *et al.* (2008) affirm that social relationships play an important role in churn prediction, and not only that, structural relationships found in CDR affect churn apart from the individual relationships of churners. Ferreira, Telang and De Matos (2019) studied the effect of relationships between customers in the context of prepaid service when consumers pay different tariff depending on whether they call inside or outside the carrier.

Modeling churn prediction for customers with prepaid contracts is more challenging than for customers with postpaid contracts since they do not pay a monthly subscription and their frequency of using the service is lower (Owczarczuk, 2010). The definition of churn could be

the expiration date of the SIM cards, however, some telcos deactivate the SIM card after one year of inactive period. In this research, we consider the typical definition of churn as the first day of 30 days of an inactive period.

In this paper, we evaluate the relevance of telco customers in a graph based on their connections established in the CDR. The main goal of this paper is to detect potential churners and consider them in the retention campaigns, as it is observed that the structural relationships play an important role in churn prediction (Dasgupta *et al.*, 2008; Ferreira *et al.*, 2019; Pushpa & Shobha, 2013; Kim, Jun & Lee 2014). The relevance of each customer is studied from a well-known sociological tool called social network analysis (SNA) (Aggarwal, 2011; Costa *et al.,* 2011, Holme, 2014). In Section 2, there is a literature review of applications of SNA in the context of churn prediction and some propagation techniques to obtain the structural relationships to find churners. However, as a far as we known, there is no comparative study about the evolution of different structural relationships over telco graphs. In this study, we represent prepaid telco relationships with two different types of temporal graphs to compare the evolution of the structural properties of the customers in the graph by extracting multivariate time series of various centrality metrics. We compare the multivariate time series with a binary classification technique called similarity forests. This tool classifies the time series of churners and non-churners and detects evolution patterns and distinctive behavior in the centrality metrics for each group of customers, having as target variable whether the customer churned at a later time.

The paper is organised as follows. In Section 2, we gather the related literature on the application of SNA in the context of churn prediction, and explain the main concepts of SNA used in this paper. In Section 3, we describe the methodology to define the influence of the churners with centrality metrics over static and temporal graphs, and the binary classification method used to evaluate the performance of each metric. Section 4 describes the used database and also how the data was prepared to be assessed by the binary classification method. In Section 5, the results of the performance of the binary classification method are shown and discussed. And last but not least, Section 6 gathers the conclusion of the paper.

2. RELATED LITERATURE REVIEW

2.1 SOCIAL NETWORK ANALYSIS FOR CHURN PREDICTION

In the context of telco, there are two different types of social networks that represent the interaction among customers (Pushpa & Shobha, 2013): Homogeneous and Heterogeneous. The former collects only one kind of relationship among customers, for example, friendship, whilst the later represents multiple kinds of relationships.

SNA provides tools to study the structure of these graphs; for instance, centrality metrics determine the relative importance of each customer in the telco graph. These centrality metrics have been applied by Pushpa and Shobha (2013) and Kim *et al.* (2014). In the first case, Pushpa and Shobha (2013) apply Regular Equivalent Algorithm after identifying the central players with different centrality metrics over heterogeneous graphs. On the other hand, Kim *et al.* (2014) have nice results examining the communication patterns among customers with a propagation process called SPA, based on the influenced of each customer. The influence of

each customer is extracted using the eigenvector centrality over a homogeneous graph based on CDR.

Although it has been shown that there is an influence among customers' relationships and their churn decision (Óskarsdóttir, Van Calster, Baesens, Lemahieu, & Vanthienen, 2018), as far as we know, centrality metrics over temporal graphs have not been applied over telco customers in the context of churn prediction.

2.2 TEMPORAL GRAPHS

Static graphs allow a representation of the relationships among people over a fixed period, but relationships are dynamic, and people's interactions change constantly. In a dynamic process, the order and the delay among their interactions are crucial to define a representative system of the real situation. This fact is a limitation of static graphs; indeed, there are many paths over the static graphs that do not exist in the real system. Temporal graphs appear to solve this situation and represent the evolution of the interactions as can be observed in Figure 1.

---Figure 1---

Different definitions of temporal graphs allow representing dynamic systems (Holme & Saramäki, 2012). For example, some models represent the system as an interval graph where the edges are active over a set of intervals. The activation of each edge is done by a presence function. In the particular case of Time-Varying Graphs (TVG), which was defined by (Nicosia, Tang, Mascolo, Musolesi, Russo, & Latora, 2013) and (Casteigts, Flocchini, Quattrociocchi, & Santoro, 2012), the presence function can be used with the latency function, which defines the time that the edge appears.

Another kind of temporal representations is through a static graph structure. For example, Kostakos (2008) represents each entity of the system as a directed chain of nodes, as Figure 1 shows. Each chain represents all stages of one node over time. As a result, the temporal graph is a directed flow of interactions, where each node is linked to itself in the next stage where it appears and to the nodes with which it interacts in the current stage. A similar approach, called Time-Ordered Graphs (TOG), is used by Kim and Anderson (2012). TOG represent the dynamic system as a directed sequence of static graphs, where each static graph gathers the information of a span window. The main difference between the definition of the temporal graphs of (Kostakos, 2008) and TOG is that the latter gathers the information of a span window and all nodes are linked to themselves in the subsequent window. This allows for discretizing the time with the desired window granularity through the size of the span window. Tang, Musolesi, Mascolo and Latora (2009) studied the influence of the span window and showed that the larger the size of the window, the lower the precision due to neglecting the order of edge appearances.

2.3 CENTRALITY METRICS

One of the main features of graphs is centrality metrics. In SNA, centrality metrics are applied to identify how the information shared among people is spread on the graph. In this situation, the relative position of each node is relevant to determine what the influence of each node is by knowing: who is going to receive the information first, who can spread the information

faster, who has a relevant position based on the importance of their connections, etc. The influence of each node over the graph is defined by different points of view through typical centrality metrics in static graphs like degree, closeness, betweenness and PageRank.

The definition of each centrality metric is as follows. Degree counts the number of connections each node has. Closeness measures how far the nodes are from the other nodes in the graph, as the sum of the inverse of the shortest path between them. Betweenness assesses the ratio of how many times a node appears on the shortest path between all pairs of nodes in the graph, and PageRank defines the importance of each node based on the number of connections that they have and the importance of their connections. These centrality metrics measure the importance of nodes from different perspectives and together provide a holistic view of the nodes' position in and influence over the graph.

There are some studies about the application of centrality metrics on temporal graphs. Holme and Saramäki (2012) define temporal centrality metrics using latency functions. In the case of TOG, multiple comparisons between static and temporal graphs exist. Pan and Saramäki (2011) define the term of average temporal distance to compare closeness over static and temporal graphs on homogenous telco graph based on phone calls. They found a high correlation between both metrics, although there is a wide range between them. In static graphs, a group of nodes can seem to be closed, while they are far in temporal graphs, or there is no temporal path at all.

Tang, Musolesi, Mascolo, Latora, and Nicosia (2010) establish a new definition of centrality over TOG and compare them with aggregated graphs to determine the significant nodes to spread the information. It is important to highlight that the definition of the temporal shortest path allows a maximum number of exchanges among nodes in the same window, and that the temporal betweenness metric considers how much time a node keeps the information until it gives the information in another window. Kim and Anderson (2012) discuss other definitions of temporal centrality metrics over TOG that are more correlated with the static metrics. In this case, the main difference is that temporal betweenness and closeness consider all temporal shortest paths evaluating all snapshots from their beginning until the end of the evaluated period. After comparing the defined temporal centrality metrics with the static ones over an aggregated graph using unweighted bidirectional edges, Kim and Anderson (2012) show that there is a significant difference in the role of the nodes when the temporal order of the edges is considered. These metrics are available in the R package TNC (Hanke, 2017).

3.   METHODOLOGY

Our proposed methodology looks at churn from a social perspective using dynamic behavior by comparing different structural relationships over temporal graphs. The methodology consists of three steps. In the first step, we build homogeneous graphs using CDR of prepaid customers. We consider both aggregated static and temporal ordered graphs. In the second step, we compute in each graph both temporal and static centrality of each node across time and obtain multivariate time series throughout various time steps. In the final step, we classify the resulting multivariate time series using a state-of-the-art classification technique called similarity forests. We evaluate the model performance using three distinct performance measures. This approach allows us to establish whether a customer´s position and importance

in the graph changes in the time leading up to churn and which centrality metric is affected the most.

3.1 GRAPHS

As shown in the previous section, there is a vast literature on the comparison between temporal and static centrality metrics. However, as far as we know, this comparison has not been done in the context of churn prediction in telco. In this paper, two different undirected unweighted graphs are used to model the dynamic influence of churners to compare the performance of centrality metrics based on Kim and Anderson (2012). In churn prediction based on SNA, the important fact is the existence of connections among customers. Óskarsdóttir *et al.* (2017) formulate that undirected graphs offer better results than the directed ones when they represent the interactions among customers. Moreover, Óskarsdóttir *et al.* (2017) conclude that there is no significant difference between using weighted graphs based on the length of the calls or the number of calls, and unweighted graphs. Therefore, we use graphs with undirected and unweighted edges to represent relationships between telco customers which we extract from CDR data.

The first type of graphs, the aggregated static graphs, are an ordered sequence of static graphs, where each graph gathers the information of a span window. The second type is a temporal ordered graph based on TOG, which similarly gathers information of a span window, but orders each span in a directed sequence of ordered graphs. In the case of TOG, as mentioned previously, all the nodes have a directed edge pointing to themselves in the subsequent window. Therefore, all the sequences are related in a directed flow. Both graphs consider a rolling window of seven days with daily steps. This fact allows enhancing the evolution of the relationships over time, considering important those that constantly appear along the week independent of the duration of the phone calls.

Both static graphs and temporal graphs are defined by a set of nodes V and edges E_t that appear in snapshot t with $0 \leq t \leq n$ where n is the total number of snapshots. The information of each snapshot is gathered in a symmetric adjacency matrix (A^t). However, in the case of the aggregated static graphs, the result is a list of n static graphs, while in the case of the temporal ordered graphs is a graph composed by a directed flow of n snapshots. In our application, nodes are the customers of the telco with prepaid mobile plans and the edges are phone calls between them.

The transformation of the temporal graph into a static graph of directed flows allows for applying static centrality metrics defined by Kim and Anderson (2012) over the temporal graphs. Both types of graphs facilitate a comparison between static and temporal centrality metrics through their time series of length n.

3.2 CENTRALITY METRICS APPLIED OVER TEMPORAL GRAPHS

Centrality metrics identify how information is spread in the graph and the power of each node over their relationships while taking into account the flow of information between nodes. In this paper, we only apply existing centrality metrics that have not been used in the context of churn prediction with temporal graphs. There are four different types of centrality metrics, namely degree, closeness, betweenness and PageRank (eigenvector centrality). Our goal is to compare static centrality metrics, which are applied on the aggregated static graphs, and

dynamic centrality metrics as defined by Kim and Anderson (2012) which are applied on the temporal ordered graphs. We distinguish these two types using the adjectives "static" and "temporal". Each centrality metric, as mentioned before, gives different information about the relative position of the nodes in the graph. The centrality metrics are formulated in Table 1.

---Table 1---

The first centrality metric is the degree, which measures the number of neighbors that the evaluated node has in the graph in the snapshot t. In the case of static graphs, as the adjacency matrix (A^t) is symmetric, the degree of the node $x \in V$, $D_t(x)$, defined in equation (1), is the sum of the rows of the adjacency matrix. In the case of TOG, the temporal degree ($D_{ij}(x)$) defined in equation (2) collects all the edges of the node x on the interval $[i, j]$ being $0 < i < j < n$. This means that the degree of the node x is the sum of all the degrees $D_{ij}(x)$ of the t snapshots which belongs to the interval $[i, j]$. However, we are working with time series of length n defined by the mentioned snapshots to allow the comparison between them, so we do not need the aggregated degree $D_{ij}(x)$, we only need the degree $D_t(x)$, whose value is the same independently of using static graphs or temporal graphs. As a consequence of this, in the rest of the paper, we use degree only.

The second centrality metric is the closeness. This metric assesses how close the customers are to each other depending on the shortest paths between them. The shortest path between each pair of nodes (u, v) is called the geodesic path $g(u, v)$. In the case of the aggregated static graphs, the closeness of the node x, see equation (3), is the inverse of the sum of the shortest paths of the node x with the rest of the graph. This metric can be normalized if it is multiplied by $|V| - 1$, being then the inverse of the average distance from the node x to the rest of the graph. The temporal closeness, defined by equation (4), is the inverse of the temporal shortest path $\Delta_{t,j}$ on the time interval $\{[t, j] : i \leq t < j\}$, so in this case we are considering m time intervals where $m = j - i$. This metric can also be normalized by multiplying with $(|V| - 1)m$.

Betweenness of a node x is the third centrality metric we apply. It is based on the sum of the ratios between the shortest path between each pair of nodes passing through the node x, and the shortest path between the pair of nodes being evaluated. The static betweenness (5) is normalized by $(|V| - 1)(|V| - 2)$. To compute the temporal betweenness (6) we must define two different subsets. The subset $S_{ij}(u, v)$ collects all the temporal shortest paths from node u to destination node v on the interval $[i, j]$, and the subset $S_{ij}(u, v, x)$ gathers all the temporal shortest paths from the subset $S_{ij}(u, v)$ that go through node x. In the formulation of temporal betweenness, $\sigma_{t,j}(s, d, x) \equiv |S_{tj}(u, v, x)|$ and $\sigma_{t,j}(s, d) \equiv |S_{tj}(u, v)|$. This formulation is normalized by $(V_s^x V_d^x m)$ being V_s^x, $V_d^x \subseteq V \backslash v$, where $\sigma_{t,j}(s, d) > 0$ for each $s \in V_s^x$ and $d \in V_d^x$.

The fourth centrality metric is Static PageRank, which is applied only over the aggregated static graphs as it has not been developed for TOG. Static PageRank, see equation (7), is the typical formula from Brin and Page (1998) which is an iterative algorithm that calculates the importance of the node x, ($PR_t(x)$) depending on the importance of its neighbors and the number of connections that they have.

All these centrality metrics allow observing the evolution of influence of each node in the graph from a different perspective and the changes in the structural relationships between them over time. After computing the temporal and static centrality metrics with the temporal graph based on TOG and the ordered sequence of static graphs respectively, the result is various multivariate time series that reflect the structural relationships of the nodes. The length of these time series corresponds to the number of days that are used when aggregating data for the graphs.

3.3 SIMILARITY FORESTS

After extracting the structural relationships of the graph, a binary classification method called similarity forests compares the different multivariate time series with respect to churn labels. This technique has already been used in the context of churn prediction with good results (Óskarsdóttir *et al.*, 2018). Similarity forests are a novel supervised machine learning technique for binary classification that can work with arbitrary data. It was developed by Sathe and Aggarwal (2017) and is an extension of random forests. During the training time, it creates multiple decision trees, which are used later to do a prediction for each object. However, the main difference between the two techniques is that similarity forests do not use random directions to split at the nodes. Instead, splits are based on the pairwise similarity of the objects in each node. This means that it computes over a small fraction of the objects, making it faster. The result is an efficient technique, which increases the performance of the traditional random forest.

Similarity forests work as follows. Each decision tree is split recursively until the leaf nodes belong only to a single class. The objects in the dataset $O_1,.., O_n$ are embedded in some multidimensional space as the points $X_1,..X_n$. The technique works as a random forest technique defining a vector direction of a sample X_i, X_j selected randomly. The dataset is projected over the defined direction, the projection of each specific data point O_k is proportional to the similarity difference of datapoints O_i and O_j, as the similarity among O_i and O_j is constant $p(\overline{X_j}) \propto S_{kj} - S_{ki} - S_{ij} + S_{ii}$. After that, the split point is defined by minimizing the weighted Gini Index based on the pairwise similarity of the sample. This split point defines a hyperplane perpendicular to the vector direction, which divides the objects into two groups. This process is repeated recursively to construct each ensemble component of the similarity forests.

The pairwise similarity can be computed in multiple ways. In this paper, the pairwise similarity is calculated using Euclidea distance as it was shown in the case of churn prediction in telco that this measure does not have significant differences with others distances and allows results without high computational requirements (Óskarsdóttir *et al.*, 2018).

3.4 PERFORMANCE MEASURES

In this paper, we use three threshold-independent measures to evaluate the predictive performance of the similarity forest classifier with each centrality metric. These measures are area under the ROC curve (AUC), area under the precision-recall curve (AUPR) and the top decile lift (TDL).

The AUC is commonly used to evaluate binary classification models. It summarizes the trade-off between the true and false positive rates using the receiver operating characteristic curve (ROC). The AUC value is calculated as the area under the ROC curve. In contrast, the area under the precision-recall curve measures the trade-off between precision and recall, being more suitable for imbalanced datasets as it does not consider true negatives. Both measures can be compared with a random model using their baseline. In the case of AUC, the baseline is 0.5, while in the case of the precision-recall curve it is the ratio of positive labels over the whole number of customers.

Top decile lift is a ratio between the ratio of churners in a fraction of customers with the highest probabilities and the ratio of churners in the actual customer base. In this paper, we use the top decile lift, also called Ten Lift, which is common in churn prediction (Óskarsdóttir *et al.*, 2017). This measure represents how much better the model is at identifying churners, compared to a random sample of customers.

4. DATASET AND EXPERIMENTAL SETUP

Before applying the described methodology to detect churners based on the structural relationships of the prepaid customers, it is critical to clean the dataset properly. This preprocessing step removes all the phone calls that last less than 4 seconds, all the duplicated observations, self-calls and all the customers who have never used the telco service during the whole selected period. The cleaned dataset has around 170 thousand customers and contains over 17 million phone calls among the telco's customers during six consecutive months.

Inevitably, the high computational requirements of the proposed temporal centrality metrics limit the size of the evaluated sample. Therefore, we extract a random sample of customers together with their most connected neighbors for 60 days in an attempt to have highly connected graphs. The desired sample must have customers, which have already made phone calls over the selected period, and all the customers must meet the definition of being churners after 30 days of the inactive period.

The process of choosing the sample of the dataset starts by selecting a random sample of 5,000 customers. After that, the closest neighbors of the previously selected customers are gathered in the sample. In this case, the threshold to define whether two customers are highly connected is 25 phone calls. This threshold defines the size of the end sample, and it is selected based on the highest limitation to compute some of the centrality metrics.

It is important to highlight that the first random sample must have enough churners to keep the churn rate of the dataset. In this case, the selected sample has 13,454 customers, the length of the analysed period, that is, the number of snapshots n, is 60 days, and the churn rate is 10.24%. Note that the churn rate is not constant as it depends on the period that is being analysed.

After preprocessing the raw data and selecting the sample, the interactions among customers are represented in undirected and unweighted static graphs and temporal graphs to define the influence of each customer over their relationships. A sliding window collects the call information of one week and moves with a step of one day. In the case of the static graphs, each undirected unweighted graph represents all the interactions of each sliding window,

while in the temporal graphs the interactions are represented as a directed flow of a sequence of the undirected static graphs. We use a seven day window in order to have enough connections in each graph and move the window one day at a time to capture daily changes in the customers' behaviour.

Once the data is cleaned and transformed into graphs, we compute the centrality metrics for each node. Degree, closeness, betweenness and PageRank are applied over the aggregated static graphs, and temporal closeness and betweenness are applied over the temporal ordered graphs. These metrics are calculated using the R package (Csardi & Nepusz, 2006) for aggregated static graphs. For the temporal ordered graphs, we use the temporal metrics implemented in the R package (Hanke, 2017) which uses the REN algorithm defined by Hanke and Foraita (2017), to reduce the computational burden. The result is a multidimensional time series for each customer that measure their importance in the graph. To estimate how long the churn process is in temporal graphs, and for how long the customer behavior should be observed to make accurate predictions, we compare two different lengths of time series 14 and 28 days.

---Figure 2--

We propose two approaches to evaluate the performance of the centrality metrics to predict churn and to overcome the limitations of using only a sample of the whole dataset. These approaches are illustrated in Figure 2, where the green boxes represent the periods where the customer is active and the time series of centrality metrics are extracted and the black circle represents the moment when the customers churned. If there is no black circle, it means that the customer did not churn.

The first approach, approach A, evaluates the individual behavior of each customer while checking if there is a different pairwise behavior among churners and non-churners and considering how many days it takes to churn . This approach selects the active period of churners some days before the start of the inactive period (w), while the active period for the non-churners is fixed (k). For non-churners, the active period is the last available time series of the dataset with length k. This means that the approach evaluates all the churners of the selected sample. The time series of churners are selected depending on which day churn is predicted. As a consequence of this, we can observe how accurate the predictions are in both short and long term. We consider various churn moments in the month along the observation period, i.e., whether it happens in the next 1, 3, 7, 10, 14, 17, 21, 24 and 28 days. This way we can see a precision in churn predictions, that is, how the approach performance varies in time.

The second approach, approach B, considers a fixed observation period for all customers and predicts churn in the following 1 2, 3, or 4 weeks. This approach represents a realistic way to evaluate the performance of the centrality metrics because all the inactive periods start in different moments, as can be observed in Figure 2. In this approach, it is important to highlight that we are training for different scenarios, that is, one prediction for each of the four weeks. In this way, we can see how well the centrality metrics predict churn immediately after the observation period, and also after some time has passed.

There are two main differences between both approaches. The first one is the churn rate. In the first approach the churn rate is around 9% and it decreases as predictions are made further in the future due to the desirable time series are not available for the customers who churn in the first weeks of the sample. The churn rate in the second approach is around 2% and reflects the ratio of churners who churn in each week. The second main difference of the

approaches is when the prediction is done: in the first approach, all churners churn "x" days after the end of the time series, while in approach B churners churn within a whole week.

After extracting the multivariate time series based on the two described approaches, similarity forests are applied using a stratified cross-validation of 10 folds to extend the sample keeping the original balance between churners and non-churners. The result of this classification is evaluated by the performance measure AUC, TDL and AUPR, as mentioned in the previous section.

5. RESULTS AND DISCUSSION

In this section, we compare the results of static and temporal centrality metrics to reveal whether the relative importance of churners is different from that of non-churners. As mentioned before, two different approaches for the experimental setup are implemented and similarity forests are used to predict potential churners. Figure 3 summarizes the results of the first approach, with time series of length two or four weeks. This figure reports the capacity of the centrality metrics to predict churners in short and long term. For both lengths of time series, all centrality metrics have a stable behavior to detect churners in middle and long term according to all three performance measures AUC, TDL and AUPR.

--Figure 3---

In the short term, most of centrality metrics predict better the churners whether they churn in the first seven days. PageRank, static closeness and degree are the centrality metrics, which perform better in the short term. All of them have an AUPR value over 75%, compared to the baseline of about 14.2% and TDL is over 6, when predicting churn in the following day. Although their performance drops sharply, they still perform better in the short term than the rest of the centrality metrics. Moreover, after the seventh or tenth day, the behavior of the centrality metrics is stable, except for temporal betweenness, which increases monotonously.

Temporal betweenness is the only centrality metric whose behavior increases monotonously in the long term. This metric has poor behavior in short-term like static betweenness and temporal closeness. However, in the long term, the behavior of temporal betweenness matches that of the degree when considering TDL for short and long series, and is furthermore similar to the static closeness when measured in terms of AUPR for long series.

In the long term, the centrality metrics that perform better are PageRank, static closeness, degree and temporal betweenness for long and short time series. In the case of short time series, PageRank performs better than for long time series in the long term, and the opposite is observed for the static closeness. The rest of the centrality metrics performs quite similar for both and short time series.

Although the behavior of the centrality metrics using short and long time series is rather similar, the performance is considerably different. In the case of static closeness, this metric needs long time series to be able to detect over 40% of churners in the first decile both in the short and the long term. However, PageRank performs better with short time series, the AUC is always over than 85%, while the AUPR is over 62%. This particular metric is able to detect over 60% of churners both in the short and the long term in approach A.

The centrality metrics with the lowest performance overall, are temporal closeness and static betweenness. Although both centrality metrics are over the baseline of AUC, both of them detect less than 10% of churners in the first decile and in the case of static betweenness for some predictions, the AUPR is under the baseline defined by the rate of churners. As a result of this performance, these metrics are not considered in approach B.

--Figure 4--

As mentioned earlier, in approach B the centrality metrics with the highest performance in approach A are compared using two different time series of length two and four weeks over a fixed period to check whether the studied churn behavior lasts short or long time as it was done in (Óskarsdóttir *et al.*, 2018). Similarity forests are used to detect churners in the next month after the analysis within one week periods. Figure 4 illustrates the performance of these centrality metrics. Here, static closeness and degree detect more churners than PageRank. Temporal betweenness finds fewer churners than any of the static centrality metrics.

In the short term, the performance of static closeness and degree is completely different inside a narrow range of values depending on whether the time series are long or short. For the first week, these centrality metrics can predict at least 30% of the churners in the first decile when the time series are short, while in long time series they are around 20%. In the second week, the performance increases for long time series, however, this behavior is not observed for the short time series.

In the long term, static closeness and degree find over 25% of churners in the first decile in the fourth week independently of using long or short series. The performance of predicting in the fourth week is better than in the third week attending to TDL and AUPR the performance of all the centrality metrics increase except for temporal betweenness whose performance is better for the third week.

--Table 2--

As mentioned previously, one of the main differences between both approaches is the difference in churn rate. To illustrate a fair comparison between both approaches, Table 2 and Table 3 report the performance of centrality metrics using approach A in short and long term with a similar churn rate observed in approach B. The new sample is achieved by randomly removing observations from the minority class. In this situation, the undersampled dataset highlights the impact of both approaches' definition to predict churn. Approach A follows, in general, the same pattern described previously that was observed in Figure 2 independently of using short or long series. PageRank centrality performs the best of all centrality metrics, followed closely by static closeness. Temporal betweenness increases its performance when the predictions are done with short time series, and degree and closeness have a similar performance. Due to the lowering values of churn rate, it is expected a drastically decreased of the AUPR curve for degree, static closeness and PageRank centrality, although their performance is more accurate in the 7th and 28th days and their values are over the baseline.

--Table 3--

As already mentioned, there are some differences between applying short and long time series. However, these differences are not significant enough to justify the need of using CDR

of one month. Hence, in this case we agree with (Óskarsdóttir *et al.*, 2018) on the use of short time series to predict churn applying SNA.

6. CONCLUSION

6.1. MAIN CONTRIBUTION

SNA is a useful tool applied in multiple fields that can be used to represent the interaction between customers. This methodology allows defining churn prediction models not only on customers' features but also on their behavior among their relationships. Most of the available literature on applying SNA to detect potential churners in telco use static graphs to map the relationships of the customers. Although some of the approaches use time series, as far as we know, there is no research about applying centrality metrics over temporal graphs.

The managerial implications of our research are twofold. On the one hand, they further establish that the behavior of customers changes in the time leading up to their decision to churn. Therefore, it is not enough to consider only static graphs or behavioral features extracted from such graphs. Features that represent changes in behavior are necessary to predict churn more accurately. On the other hand, our work focusses on centrality metrics which represent the influence a customer has in the call network. Proper analysis of customer characteristics relative to their centrality, before taking on a retention campaign has the potential to be more effective as, more influential customers can be targeted and they influence other potential churners to also not churn. Thus, the campaign could have a cascading effect.

This paper tests the influence of customers' behavior before churning based only on the evolution of the relative position of the customers among their relationships. With this purpose, customers' relationships have been modeled by SNA in two different temporal graphs and multiple centrality metrics have been applied to measure customers' behavior. This analysis evaluates the performance of SNA centrality metrics on representing dynamic customer behavior on churn. This behavior is extracted from the CDR and it is based on the interactions between customers through connections that are created by means of their phone usage within the telco. These connections are represented in their real time order of interaction due to the application of temporal graphs. The second main contribution was to compare centrality metrics over the temporal graphs in terms how well they represent and distinguish the behavior of customers before they churn.

We consider two approaches to evaluate the predictive performance of time series of centrality metrics. In the first approach, the individual behavior of each customer is observed checking the key centrality metric to represent the behavior of customers. While in the second one, the time series belong to the same period to evaluate the capability of each centrality metric to detect churn in a practical situation.

From the first approach, we can observe that most of the static centrality metrics perform better than the temporal ones. Therefore, the order of the interactions is not relevant for the evaluation of the customer relationships allowing quick results without using high computational methods. PageRank and static closeness detect churners better than any other centrality metric, both short and long term independently of using short or long time series.

This means, that there is a clear pattern to differentiate churners from non-churners when all-time series are aligned and the churners start their inactive period at the same time. In the second approach, static closeness outperforms the rest of centrality metrics both in short and long term, although the accuracy of the predictions changes depending on the length of the time series. This means that the pattern variation of PageRank and closeness in the time series allows predicting churners due to they are more sensitive to the different behavior between churners and nonchurners. It is interesting to highlight how the number of connections of the customers have not outperformed these centrality metrics, enhancing the influence of the closeness among customers and importance of their position among their relationships.

It is important to consider that there is an expected gap in the performance of the centrality metrics when comparing both approaches with a similar churn rate, due to the influence of when the churners start their inactive period. In the case of approach B, the prediction is done over a wider number of days. This gap is different depending on the centrality metric, for example PageRank is affected significantly by this fact, but it is not the case for the static closeness.

6.2. LIMITATIONS AND FUTURE WORK

As all studies, this work has some limitations, which are important to highlight in case that any researcher would like to go on this new field. In this paper, the first limitation is the high computational requirements needed for the temporal centrality metrics applied over TOG. Working with CRD that have hundreds of thousands of observations is hard, and handle structures are needed to work with them. These temporal centrality metrics have become the bottleneck of this research, although it was implemented the REM algorithm, so it has been necessary to reduce the sample and limit the use of time series longer than a month. In the case of static graphs, these high requirements do not exist, and multivariate time series are achievable in a reasonable time. Overcoming this limitation, the evaluated sample allowed us to do a comparison of the performance of this centrality metrics to predict churn.

In a future research, we propose to apply SNA techniques to simulate the influence of churners as a social contagion over a temporal aggregated static graph (Holme, 2014) to evaluate the power of churners instead of defining whether the behavior of customers allow to predict churn. Relational learners, which learn churn labels directly from the network, have been used with some success in the literature, although it has been shown that they are worse than their non-relational counterparts (Óskarsdóttir *et al.*, 2017). Thus far, they have been used with static networks only. However, developing relational classifiers on time evolving networks for churn prediction would be a very interesting and highly relevant research project and something worth working on. In a follow up study, our goal is to combine network features, centrality features and local customer features in one holistic churn prediction model, where we will apply a number of binary classification techniques in order to find an optimal approach that furthermore takes into account numerous reasons for churning.

7. ACKNOWLEDGEMENT

8. REFERENCES

Aggarwal, C. C. (Ed.). (2011). *Social Network Data Analytics*. New York, USA: Springer

Brin, S., & Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. In *Proceedings of the 7th World-Wide Web Conference*. Brisbane, Australia.

Caigny, A. D., Coussement, K., & Bock, K. W. D. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, *269*, 760–772. https://doi.org/10.1016/j.ejor.2018.02.009

Casteigts, A., Flocchini, P., Quattrociocchi, W., & Santoro, N. (2012). Time-varying graphs and dynamic networks. *International Journal of Parallel, Emergent and Distributed Systems*, *27*(5), 387–408. https://doi.org/10.1080/17445760.2012.668546

Costa, L. D. F., Oliveira Jr, O. N., Travieso, G., Rodrigues, F. A., Villas Boas, P. R., Antiqueira, L., ... & Correa Rocha, L. E. (2011). Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, *60*(3), 329–412.

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

Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., & Nanavati, A. A. (2008). Social Ties and their Relevance to Churn in Mobile Telecom Networks. In *Proceedings of the 11th international conference on Extending database technology: Advances in database technology* (pp. 668–677)

Ferreira, P., Telang, R., & De Matos, M. G. (2019). Effect of Friends' Churn on Consumer Behavior in Mobile Networks. *Journal of Management Information Systems*, *36*(2), 355–390. https://doi.org/10.1080/07421222.2019.1598683

Ganesh, J., Arnold, M. J., & Reynolds, K. E. (2000). Understanding the Customer Base of Service Providers: An Examination of the Differences between Switchers and Stayers. *Journal of Marketing*, *64*(3), 65–87. https://doi.org/10.1509/jmkg.64.3.65.18028

Hanke, M. (2017). *TNC: Temporal Network Centrality (TNC) Measures*. R package version 0.1.0. Retrieved from https://cran.r-project.org/package=TNC

Hanke, M., & Foraita, R. (2017). Clone temporal centrality measures for incomplete sequences of graph snapshots. *BMC Bioinformatics*, *18*(261), 1–18. https://doi.org/10.1186/s12859-017-1677-x

Hollander, R. (2017). Two-thirds of the world's population are now connected by mobile devices. Retrieved from https://www.businessinsider.com/world-population-mobile-devices-2017-9?r=US&IR=T

Holme, P. & Saramäki, J. (2012). Temporal networks. *Physics Reports*, *519*(3), 97–125. https://doi.org/10.1016/j.physrep.2012.03.001

Holme, P. (2014). Analyzing Temporal Networks in Social Media. *Proceedings of the IEE,* *102*(12)

Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems With Applications*, *39*(1), 1414–1425. https://doi.org/10.1016/j.eswa.2011.08.024

Kim, H., & Anderson, R. (2012). Temporal node centrality in complex networks. *Physical Review E*, *85*(026107), 1–8. https://doi.org/10.1103/PhysRevE.85.026107

Kim, K., Jun, C., & Lee, J. (2014). Improved churn prediction in telecommunication industry by analyzing a large network. *Expert Systems with Applications*, *41*, 6575–6584.

Kostakos, V. (2008). Temporal graphs. *Physica A: Statistical Mechanics and Its Applications*, *388*(6), 1007–1023.

Kotler, P., & Keller, K. (2006). *Dirección de Marketing*. Naucalpan de Juárez, Estado de México: Pearson Educación de México. https://doi.org/10.1017/CBO9781107415324.004

Nicosia, V., Tang, J., Mascolo, C., Musolesi, M., Russo, G., & Latora, V. (2013). Graph Metrics for Temporal Networks. In P. Holme & J. Saramäki (Eds.), *Temporal Networks* (pp. 15–40). Berlin: Springer Berlin Heidelberg.

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

Óskarsdóttir, M., Van Calster, T., Baesens, B., Lemahieu, W., & Vanthienen, J. (2018). Time series for early churn detection: Using similarity based classification for dynamic networks. *Expert Systems with Applications*, *106*, 55–65. https://doi.org/10.1016/j.eswa.2018.04.003

Owczarczuk, M. (2010). Churn models for prepaid customers in the cellular telecommunication industry using large data marts. *Expert Systems with Applications*, *37*, 4710–4712. https://doi.org/10.1016/j.eswa.2009.11.083

Pan, R. K., & Saramäki, J. (2011). Path lengths, correlations, and centrality in temporal networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *84*(016105). https://doi.org/10.1103/PhysRevE.84.016105

Pushpa, & Shobha, G. (2013). Social Network Classifier for Churn Prediction in Telecom Data. In *International Conference on Advanced computing and Communication Systems*. Coimbatore, INDIA.

Sathe, S., & Aggarwal, C. C. (2017). Similarity Forests. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 395–403). Halifax, NS, Canada. https://doi.org/10.1145/3097983.3098046

Tang, J., Musolesi, M., Mascolo, C., & Latora, V. (2009). Temporal Distance Metrics for Social Network Analysis. In *Proceedings of WOSN '09* (pp. 31–36). Barcelona, Spain.

Tang, J., Musolesi, M., Mascolo, C., Latora, V., & Nicosia, V. (2010). Analysing Information Flows and Key Mediators through Temporal Centrality Metrics. In *Proceeding SNS '10 Proceedings of the 3rd Workshop on Social Network Systems*.

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, *55*, 1–9. https://doi.org/10.1016/j.simpat.2015.03.003

Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector : A profit driven data mining approach. *European Journal of Operational Research*, *218*(1), 211–229. https://doi.org/10.1016/j.ejor.2011.09.031

Wanchai, P. (2017). Customer churn analysis : A case study on the telecommunication industry of Thailand. *2017 12th International Conference for Internet Technology and Secured Transactions, ICITST 2017*, 325–331. https://doi.org/10.23919/ICITST.2017.8356410

TABLES

Table 1 Definition of centrality metrics for aggregated static graphs and time-ordered graphs

Aggregated Static Graphs		Time-Order Graphs	
Static Degree: $$D_t(x) = \sum_u^{\lvert V \rvert} A_{u,x}^t$$	(1)	Temporal Degree: $$D_{ij}(x) = \sum_{t=i}^{j} D_t(x)$$	(2)
Static Closeness: $$C_t(x) = \frac{1}{\sum_{v=1; x \neq v}^{\lvert V \rvert} g(x,v)}$$	(3)	Temporal Closeness: $$C_{ij}(x) = \sum_{i \leq t < j} \sum_{u \in V \backslash v} \frac{1}{\Delta_{t,j}(v,u)}$$	(4)
Static Betweenness: $$B_t(x) = \sum_{s \neq x \neq d \in V} \frac{g_{t,j}(s,d,x)}{g_{t,j}(s,d)}$$	(5)	Temporal Betweenness: $$B_{i,j}(x) = \sum_{i \leq t < j} \sum_{\substack{s \neq x \neq d \in V \\ \sigma_{i,j}(s,d) > 0}} \frac{\sigma_{t,j}(s,d,x)}{\sigma_{t,j}(s,d)}$$	(6)
Static PageRank: $$PR_t(x) = 0.15 + 0.85 \sum_j \frac{PR_t(j)}{D_t(j)}$$	(7)		

Table 2 Reduction of Churn rate over model A using short time series

	Predicting in X days	7	14	21	28
	Churn rate	1.90%	1.70%	2.10%	2.00%
Degree	AUC	79.22%	70.35%	69.51%	74.23%
	Ten Lift	4.06	2.41	2.07	3.06
	AUPR	12.05%	3.46%	4.48%	6.74%
Static Closeness	AUC	78.87%	73.95%	72.05%	79.00%
	Ten Lift	3.85	2.77	2.50	3.91
	AUPR	9.02%	4.10%	4.64%	7.95%
Static PageRank Centrality	AUC	84.28%	78.64%	78.18%	81.88%
	Ten Lift	6.23	5.41	5.24	6.32
	AUPR	48.28%	42.01%	39.31%	47.53%
Temporal Betweenness	AUC	76.30%	76.59%	80.17%	79.31%
	Ten Lift	2.84	3.41	3.91	4.02
	AUPR	20.12%	21.85%	31.18%	31.37%

Table 3 Reduction of Churn rate over model A using long time series

	Predicting in X days	7	14	21	28
	Churn rate	2.20%	1.90%	2.10%	2.70%
Degree	AUC	76.87%	68.84%	66.09%	70.88%
	Ten Lift	3.87	2.27	1.99	2.59
	AUPR	15.86%	6.24%	4.10%	9.65%
Static Closeness	AUC	81.83%	75.05%	72.93%	79.34%
	Ten Lift	4.82	3.12	2.77	4.28
	AUPR	20.72%	8.77%	10.14%	15.80%
Static PageRank Centrality	AUC	77.37%	70.59%	68.12%	77.04%
	Ten Lift	4.56	3.27	3.23	4.55
	AUPR	27.36%	17.44%	18.43%	30.18%
Temporal Betweenness	AUC	69.76%	72.64%	77.40%	76.00%
	Ten Lift	1.41	1.96	2.89	2.83
	AUPR	6.47%	10.11%	19.91%	17.94%

FIGURES

Figure 1 Visualization of a Sequence of Static Graphs and a Time-Ordered Graph
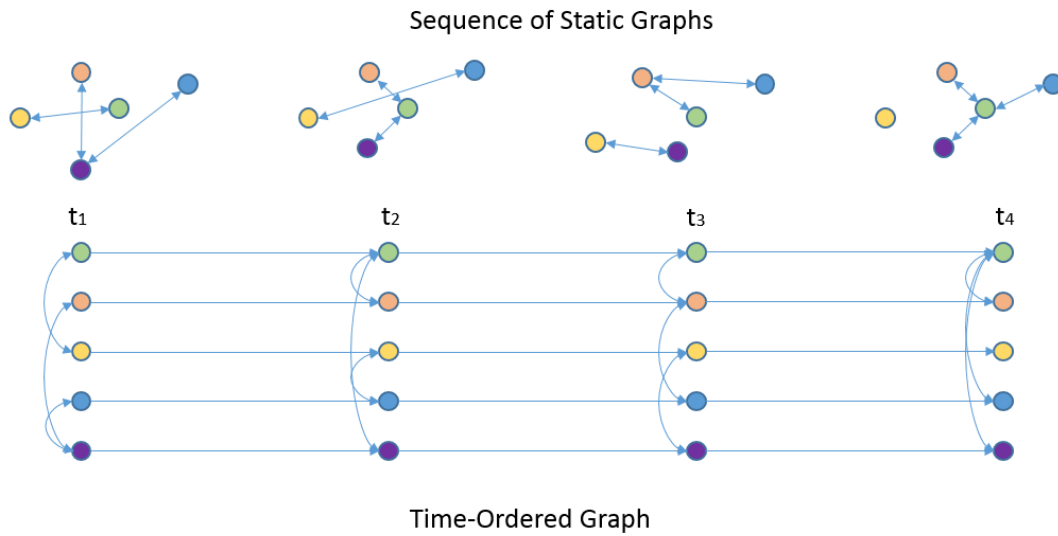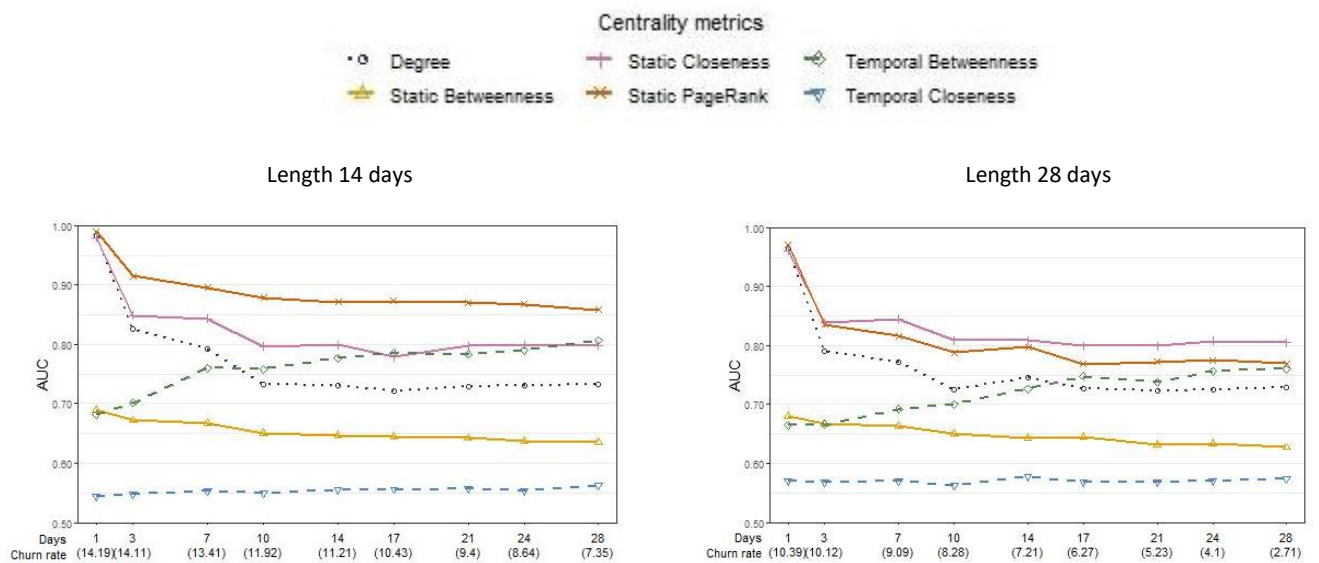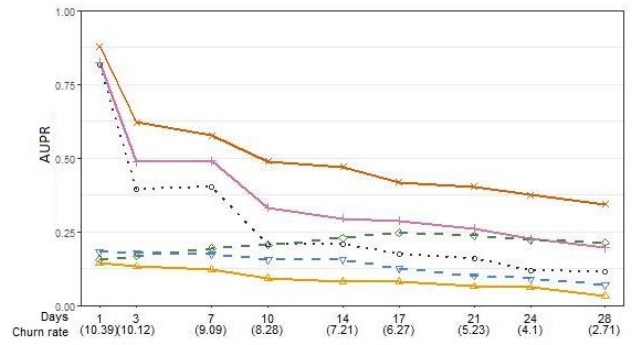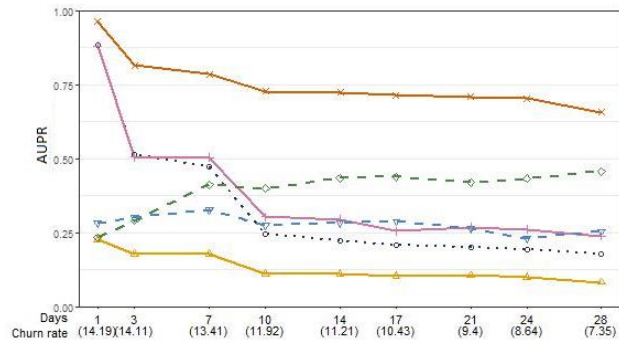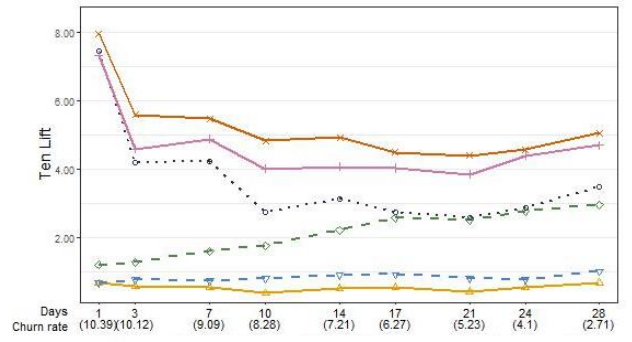

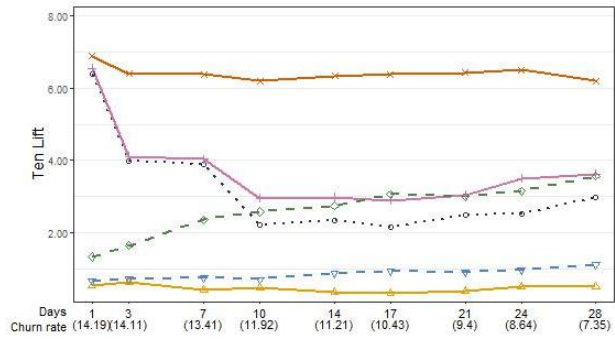Figure 2 Definition of approach A and approach B
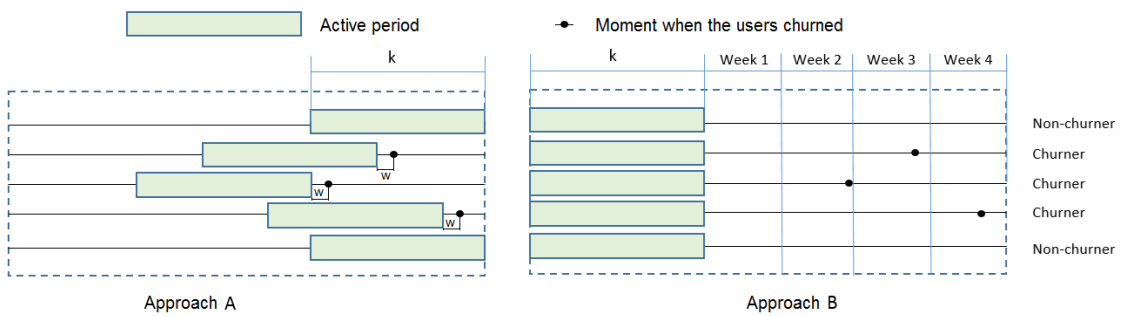
Figure 3 Churn Prediction in approach A

Figure 4 Churn Prediction in approach B