H-0317 (HAI1209-008) September 23, 2012
Computer Science

IBM Research Report

The Effect of Social Affinity and Predictive Horizon on Churn Prediction Using Diffusion Modeling

Dorit Baras, Amir Ronen, Elad Yom-Tov*
IBM Research Division
Haifa Research Laboratory
Mt. Carmel 31905
Haifa, Israel

*Presently with Yahoo Research, New York

The Effect of Social Affinity and Predictive Horizon on Churn Prediction using Diffusion Modeling

Dorit Baras* Amir Ronen† Elad Yom-Tov‡

Abstract

The social influence of people on their peers in the selection of products and services is frequently modeled as a diffusion process. Recently, such processes have been successfully applied as a tool for the prediction of customer turnover, or churn, in mobile communication carriers. Such prediction is most accurate when the appropriate social ties are used, and is primarily useful when it provides a long forecast horizon, so as to enable the service provider to take mitigating actions. In this work we investigate several measures of social affinity and compare their performance for churn prediction. We demonstrate that these measures capture different calling patterns and show that combining these measures can significantly improve the accuracy of the prediction. We study the predictive horizon of diffusion processes and show that it deteriorates significantly as the horizon increases. Our results from two large mobile phone carriers show how the usefulness of diffusion processes can be enhanced for churn prediction and provide insights to their limitations.

Keywords: Churn Prediction, Diffusion Processes, Telco, Social Networks

1 Introduction

In the past few decades, telecommunication, and especially mobile telecommunication, has become a dominant communication medium among people. Many countries report over 100% saturation of the telecom market, indicating the importance customers assign to this form of communication. Public regulators and standartization bodies have made the telecom market a highly competitive one, by enabling customers to easily move from one telecommunication provider to another. Such transitions are known as **churn**. From this point, we refer to churn as transition from the current carrier to a different one.

Churn is one of the most costly items affecting a telecom carriers' bottom line as it decreases revenue. Additionally, the carrier investment of winning a new customer is far greater than the cost of preserving an existing one [9]. In many cases, churn can be lessened by offering customers incentives to remain customers of the existing carrier. Therefore, there are salient reasons for predicting which customers are likely to churn in the near future, so as to try and prevent their churn.

In addition to identifying likely churners, a good churn prediction system should provide a sufficiently long horizon forecast in its predictions: First, a long forecast horizon is required to enable the customer care department to approach the customer and make him a retention offer. Second, a long forecast horizon is advantageous in that the further away the customer is from actually making the churn decision, the easier it is to prevent that decision at a significantly lower cost.

Most churn prediction systems (see for example, [3, 5, 8, 14, 18]) consider each customer individually. The goal of these systems is to predict which customers are likely to churn in the immediate future, which is usually set to between one month and three months. Such systems rely on hundreds of complex Key Performance Indicators (KPIs) which are generated for each customer from their call attributes, financial attributes, and service information, all evaluated over long (many months and years) periods. These KPIs then serve as input to a statistical regression model (usually a logistic regression variant) that outputs a churn score. This approach has one major drawback in that it relies on the assumption that a churning customer changes some of his attributes (calling patterns or otherwise) prior to switching carriers. While this may be true in some cases, there are certainly many scenarios in which these assumptions are violated. For example, this may occur when customers come to believe that they have found a better deal with a competitor and churn immediately.

The dominant approach to churn prediction focuses exclusively on the individual customer without taking into account social influence. Clearly, there are many social aspects to churn, as witnessed in other consumer areas [7, 21], where a dominant example is when a churning customer influences other customers to churn as well. Thus, developing churn prediction systems

*IBM Research - Haifa

†IBM Research - Haifa

‡IBM Research - Haifa. Current affiliation: Yahoo Research, New York, NY

that take social aspects into account poses an emerging theoretical challenge with potentially great practical implications.

Several works (e.g. [10, 2, 4, 15]) have started exploring the possibility of using social ties to predict churn. The dominant approach in this direction is known as **diffusion**. The underlying assumption of diffusion is that recent churners are known and they are likely to affect the churning decisions of their social neighborhood. In order to predict churn, the network of subscribers is modeled as a weighted directed graph where nodes represent the customers, and weights on the edges correspond to the strength of the social connections between them. Next, a diffusion process is used to model the flow of information from recent churners to their social environment. Specifically, each node in the network that corresponds to a recent churner is assigned an initial numerical value termed energy. A decaying diffusion process propagates this energy across the global network until convergence. At that point, each subscriber in the network has some associated energy corresponding to the amount of "churning" information or influence that has been assigned to him. The individual churn scores are then derived directly from these energies [4]. We discuss the technical details of this approach below.

In addition to churn prediction, diffusion processes play a central role in the study of various fields in marketing and social network analysis. These include the study of the spread of innovation, viral marketing and more. A recent survey of this literature can be found at [15] and pioneering study of graph theoretic problems that stem from some of these applications can be found in [11]. We believe that understanding the properties of diffusion-based models can enhance the research on these topics as well.

As noted above, the basic assumption underlying diffusion is that friends who churn affect ones' churn propensity. Recently, Nitzan and Libai [13] showed that exposure to a churning customer increases the likelihood of ones' churn by 80% (after controlling for homophile, i.e., user similarity). Interestingly, this study also demonstrated that two thirds of customers who churn do not have an immediate churning acquaintance (where the latter is defined as a person called by the churner). This means that diffusion can explain the churn of at least 33% of the churning population, which is a significant social effect.

Two technical factors influence the performance of diffusion. First, the way in which the strength of connection between customers is estimated, and second, the specific algorithm used to diffuse energy across the graph.

Most papers to date modeled the connection strength between customers using the number of minutes of calls between subscribers, normalized by the total number of calls made by a customer [4, 10, 15, 2]. This has an obvious advantages in that it is easy to measure and is easy to justify as a measure of peoples' relationship strength. Adamic and Adar [1] suggested that a more informative measure is to count the number of shared friends every pair of customers have between them. Richter et al. [16] modified this measure by normalizing it to the total number of subscribers called, through the use of point-wise mutual information. However, to the best of our knowledge, no study has shown a rigorous comparison of the effect of these measures on the accuracy of the diffusion process.

Our goal in this paper is to investigate three important aspects of diffusion-based algorithms for churn prediction. First, we study the effective predictive horizon of diffusion algorithms, which is of major importance to the applicability of these algorithms to realistic business problems. We demonstrate that diffusion algorithms can provide impressive churn prediction capabilities, though only for a limited horizon. For longer prediction horizons we show that there is a significant deterioration in performance. Second, we compare different methods of estimating the connection strength between users, and show that some methods have significantly better performance than others. Finally, we show that combining results from diffusion processes based on different types of connection strength measures improves the accuracy of the prediction. Specifically, in one of the carriers, this method yielded an average of 50% improvement compared to the standard baseline measure. These findings can guide telecommunication carriers who wish to implement diffusion algorithms, and have important insights into human behavior.

2 Diffusion Processes for Churn Prediction

2.1 Description of the Diffusion Process The diffusion process that we studied is based on [4]. We now describe it in detail (see also Algorithm 1). The diffusion process starts with the construction of a call graph and a seed. The call graph is a directed graph in which each node corresponds to a subscriber in the network and the weight on each directed edge reflects the strength of connection between the caller (head of edge) and callee (tail of edge). The weight associated with each edge is based on calls data (for example, the total number of calls or the total duration of calls over a period of time). The seed is a list of subscribers that are known to have churned during a predefined period of time, typically a subset of the time period that was used to construct the graph, e.g. the same

Algorithm 1 The diffusion algorithm of [4]

Calculate W_{ij}, $\forall i, j$ from calls data:
$W_{ij} = \#$ calls where i is the caller and j is the callee
Set initial energy for all subscriber as follows:
$$E_0(i) = \begin{cases} c, & \text{if } i \in \text{seed (churner);} \\ 0, & \text{else;} \end{cases}$$
repeat
  Update $E(j) \forall j$ by having each subscriber i in the graph transfers $d \cdot E(i)$ energy to its neighbors according to the formula: $\frac{d \cdot W_{ij} \cdot E(i)}{\sum_k W_{ik}}$
until $\max_i |E_{new}(i) - E_{old}(i)| < \epsilon$.

period, the last two weeks, etc. Each such churner is assigned with an initial positive energy and all other subscribers are assigned with zero energy. Finally, a diffusion-like process is initiated in the graph, where at each iteration nodes transfer a fraction of their energy to their outgoing neighbors in the graph. The exact value depends linearly on the weight associated with the edge and on a spreading coefficient $d \in (0, 1)$ that determines the fraction of energy that can be given away. After the stopping condition is met, each subscriber is associated with a certain amount of energy, where higher values are considered higher probability candidates for churning. In the remainder of the paper we use the implementation of the diffusion algorithm as described in [4], with additional details given in [12]. It should be noted that many variants of the above process are possible (e.g., a non-uniform initialization of the churn energies).

2.2 Additional Aspects of the Diffusion Process

The diffusion process described in Section 2.1 gives rise to several natural questions that we attempt to investigate in the current work. These include the following:

2.2.1 Predictive horizon The predictive horizon is the period of time in the future for which an algorithm can provide predictions. Churn prediction is typically used as a means for customer preservation. Since resources are limited, only a small fraction of the population may be contacted by the customer care service, and the main goal is reaching the customers that are most likely to churn as early as possible. Therefore, the time horizon plays a role in the efficiency of the obtained results. An optimal predictor would provide long-term prediction so that the it can be used more effectively to prevent churn by customers in risk, a task which naturally requires time and effort the telecommunication company. Previous work [13] demonstrated degradation of prediction accuracy over time when using a "word-of-mouth" scenario, for a specific similarity measure, and a simple diffusion model. Our goal is to ascertain whether the degradation is a general characterizer of diffusion processes for churn prediction, and whether it is affected by the choice of relationship measure (see below a discussion of measures).

2.2.2 Relationship measure Most previous works used a social affinity measure between subscribers that is based on direct correspondence between two subscribers. Several recent works demonstrated the usefulness of using alternative measures for various applications (for example [1, 17, 16]). Therefore, our second goal is to determine the effect of the relationship measure (i.e., the corresponding edge weight) on the accuracy of the algorithm with respect to churn prediction. In our work we compare four types of measures:

- *Calls measure*: The weight is proportional to the count of calls between the caller and the callee. This is the most common measure used in the context of diffusion.

- *Weighted number of shared neighbors measure*: The weight is proportional to the number of shared outgoing neighbors [1] in the call graph, weighted by the number of calls made to those neighbors. The weighting was added to emphasize the effect of frequently called subscribers, with which one has a stronger connection.

- *Social measure*: This measure, defined in [16], measures the point-wise mutual information between caller and callee, according to the number of shared versus unshared outgoing neighbors, if and only if the caller made at least one call to the callee. We refer to this measure as social because it takes into account the social environment of the caller and callee and not just the fact that two subscribers talked to each other.

- *Cosine of the angle between call vectors measure*: The weight is proportional to the cosine between the two vectors representing the calls that the caller and the callee performed to all other nodes in the calls graph.

Naturally, each such measure will result in a different network structure. For example the measures are asymmetric. Section 2.3 describes these measures in more detail and discusses some of the differences in their resulting network structure.

The four similarity measures require different computational complexities. Whereas the calls measure can be built using a simple sparse matrix, computing the other three measures requires second-order computations, and are thus more expensive computationally.

2.2.3 Methods for performance enhancement

A natural question is whether the accuracy of diffusion processes can be enhanced by generic techniques. This question may also have implications on other applications of marketing and social networks. In this paper we focus on the possibility of combining different diffusion processes and show that indeed such combination can significantly increase the performance of churn prediction methods.

2.3 The Interplay between the Relationship Measure and the Coverage

It is important to note that the graphs obtained by each of the measures described above are different in structure (i.e., which nodes are connected through non-zero weighted edges) as well as in the edge weights.

The call graph contains all the users who made or received at least one call during the corresponding period. Hence, it has the largest coverage among the studied relationship measures. Other measures invoke smaller graphs, as we demonstrate using the following simple example: Consider a network of 5 nodes, numbered $1, 2, 3, 4, 5$. Let $v_i[j]$ denote the number of times caller i called j and consider the following five vectors: $v_1 = [0, 100, 0, 0, 1]$, $v_2 = [0, 0, 100, 0, 1]$, $v_3 = [0, 0, 0, 100, 0]$ and both v_4 and v_5 are zero vectors. In this example callers 4 and 5 made no outgoing calls, caller 1 called caller 2 100 times and caller 2 called caller 5 once. The call graph that corresponds to this data is presented at the top of Figure 1. By definition, in the calls measure, the weights are $w_{ij} = v_i[j]$. In this figure, large weights are denoted by solid lines and small weights by dashed lines.

Let us now consider each of the measures that we study on this call graph.

When constructing the social graph, we take into account the social environment as well as the immediate connections. The mutual information between two nodes (A and B) takes into account the number of nodes that both A and B called, the number of nodes that A called and B did not, the number of nodes that B called and A did not, and the number of nodes that both didn't call. The resulting network appears in the middle of Figure 1. Again solid lines represent large weights and dashed lines represent small weights. The strength of connection between nodes 1 and 2 is high because their social environment is non empty (node 5 is an outgoing neighbor of both). The other weights are smaller, and although we keep a large fraction of the edges, low valued edges are pruned in our implementation. Therefore depending on actual settings of the algorithms, the resulting network may not retain connections to nodes 3,4,5 or any subset of those.

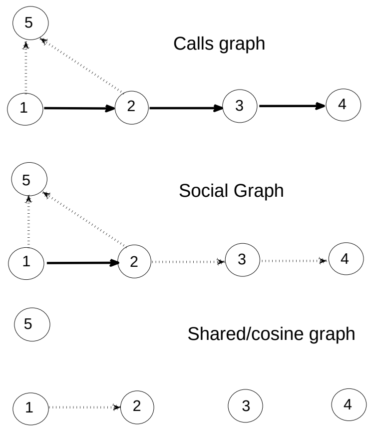

Figure 1: The networks resulting from various relationship measures. From top to bottom: calls, social, shared neighbors and cosine.

Consider the network that is constructed from this data using the weighted number of shared neighbors measure. Formally, this measure is defined by the dot product between of calls vectors above: $w_{ij} = v_i \cdot v_j$, if $v_i[j] > 0$ and zero otherwise. The resulting network is shown at the bottom of Figure 1. As can be seen it contains only one edge - the edge from node 1 to node 2, because these are the only two nodes that have shared outgoing neighbors and are connected.

Using similar consideration we can show that the cosine network has the same structure as the shared one, with different weights: $w_{ij} = v_i \cdot v_j / (\|v_i\| \, \|v_j\|)$

This small example demonstrates that even before running the diffusion algorithm, the choice of relationship measure has a major effect on the network structure, and in particular on the node coverage. While the calls measure guarantees that all callers and callees are part of the networks, the social measure prunes a fraction of the smallest edges resulting in a smaller network. The shared and cosine measures networks might be significantly smaller in terms of nodes and edges. As will be demonstrated this indeed happens and should be taken into account when choosing a measure. Various modification of these two measures may be attempted (e.g. adding some constant to the actual number of mutual neighbors). We leave the study of such modifications and their effect on the diffusion algorithm to future work.

3 Experiments

3.1 Data Sets We used calls data from two telecommunication providers located in different countries. We divided the data into two week periods, with an overlap of one week between each pair of consecutive periods. Unless otherwise noted, the reported results are averaged over these time periods.

Following the discussion of Section 2.3, Table 1 summarizes the average number of nodes in each graph (averaged over time periods). Figure 2 shows a logarithmic histogram of the undirected node degrees. The social graph contains about 55% of the nodes of one carrier and 79% of the second, and the shared and cosine graphs contain about 20% of the nodes of one carrier and 54% of the second. As explained earlier, there is an option of disregarding edges below a certain threshold for the social graph (as in [16]), which implies a tradeoff between the size and accuracy that can result from different choices of this threshold. A higher threshold (i.e., retaining only stronger edges) decreases the size of the graph and leaves only edges which represent stronger connections. We did not explore the effect of the threshold parameter. The data summarized in the table demonstrates that indeed different measures construct different networks in terms of nodes and edges.

In all relationship measures, we see that there are more nodes with a small number of connection compared to nodes with large number of connections. The decrease in all relationship measures is rapid. However, whereas in the calls measure it follows closely a power-law decay, the social, shared and cosine measures exhibit a phase transition at a value of around 5. This is partially because the graph contains fewer nodes, and partially because only subscribers who have shared connections are kept.

Each carrier defines a subset of the subscribers that is considered relevant for churn prediction. These can be different segments of the population (private, business, etc) or on customer value. Table 1 also summarizes the average number of nodes that are targeted by the carrier in each graph. Interestingly, the number of targeted nodes is less affected by the choice of measure. Finally, churn rates are 1% per month for the first carrier, and 0.4% for the second carrier, which are considered relatively low rates.

3.2 Evaluation of the Results Lift [6] is one of the most commonly used performance measures for this type of application. For a given fraction P, where $0 < P \leq 1$, the lift is defined as the ratio between the number of churners among the fraction of P subscribers that are ranked highest by the proposed system, and the expected number of churners in a random sample from the general subscribers pool of equal size. For example, a lift of k at a fraction $P = 0.05$ means that among the 5% of the subscribers that are the highest ranked by the algorithm, there exist k times more churners than in a random sample of the population.

The *lift curve* characterizes the performance of a given churn prediction system. This curve plots as a function of fraction of the population ($0 < P \leq 1$) to the lift value obtained for this fraction. In general it is a monotonically decreasing function because the larger the fraction, the more difficult it is to provide meaningful lift. By definition for $P = 1$, the lift is 1.

Typically, since carriers can only invest targeted efforts in a small fraction of the population, the lift that corresponds to the small fractions is more important to the carriers. In practice, common working points range from 0.1% to 10%.

Measure	Average number of nodes		Average number of targeted nodes	
	Carrier 1	Carrier 2	Carrier 1	Carrier 2
Calls	7731K	8572K	270K	306K
Social	4325K	6794K	246K	303K
Shared	1581K	4620K	211K	299K
Cosine	1581K	4620K	211K	299K

Table 1: Nodes statistics for the two dataset, for the four similarity measures.

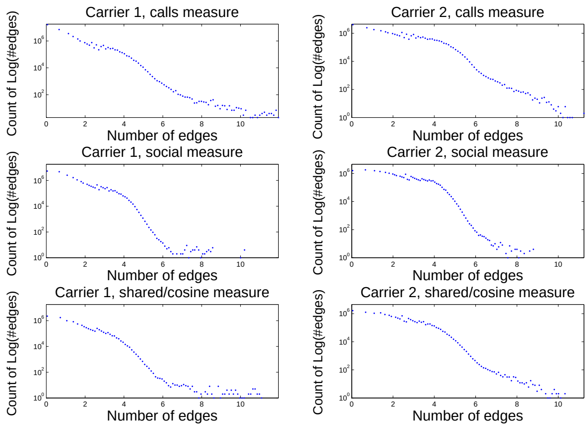

Figure 2: Histogram of the number of edges for each carrier and each relationship measure. Isolated nodes (for example churners that made no calls in the relevant time period) are not taken into account. Note that the vertical axis is logarithmic

4 Experimental Results

We implemented the diffusion algorithm of [4] using the IBM Parallel Machine Learning toolbox [20]. The output of each run of the diffusion algorithm is a list of nodes, along with their corresponding level of energy (which we refer to as scores). Our experiments compare the performance of the algorithm over the entire population. In cases where nodes did not receive a score by a measure, we set this score to zero.

4.1 Effects of the Predictive Horizon In this set of experiments we studied the effect of the predictive horizon on the quality of the results. We calculated the lift for every period, for prediction horizons varying between 5 and 90 days (at intervals of five days). In all the experiments we conducted, on both carriers and on all four measures, we see a clear deterioration of the predictive accuracy as the prediction horizon is increased. The effect is more evident in smaller fractions of the population. After approximately 30 days the decrease is significant (30% - 40% is some cases) and beyond 60 days, the lift values are almost constant.

Figure 3 plots the lift values of the first carrier (denoted by asterisks) along with the best exponential fit (solid line) of the form $f(x) = Ae^{-\alpha \cdot x} + B$ where $A, B, \alpha \in \Re$. These curves represent the lift on the entire population. Similar phenomena can be observed in both carriers regardless of the fraction of the population. In the exponential fit, the decay constants are in the range [0.025,0.047] and the fit is extremely high (average R^2 is 0.98). In the limit of infinite time horizon ($x \rightarrow \infty$), the theoretical value of the lift approaches 1. Therefore we expect the fit to also approach 1 for large horizons. Probably due to the limited maximal horizon of 90 days we did not reach that value, but a close one (the mean value of B is 1.2). Results on the second carrier are similar.

In [13], Nitzan end Libai showed that the hazardous effect that churners have on their neighbors decreases exponentially over time. Our results are consistent with their findings and indicate that diffusion is especially suited for short to medium prediction horizons.

4.2 The effect of different relationship measures As mentioned above, we tested several relationship measures on the same populations, time frames, and predictive horizons. This allows us to estimate the contribution of the measures directly, given that all other parameters are equal.

Figure 4 shows the lift at the 1% fraction of the population as a function of the prediction horizon, for the first carrier. As this figure demonstrates, the social measure outperformed the calls, shared, and cosine measures, up to an horizon of 40 days. Table 3 quantifies these improvements, averaged over several prediction horizons, and shows that the differences between the individual measures are often significant. Significantly, only the social measure outperforms the calls measure in both carriers. Therefore, we deduce that the social measure is a good candidate for being the default measure of choice in diffusion, while taking into account its larger computational complexity. However, individual carriers may find that other measures offer (sometimes greatly) superior performance for their specific population.

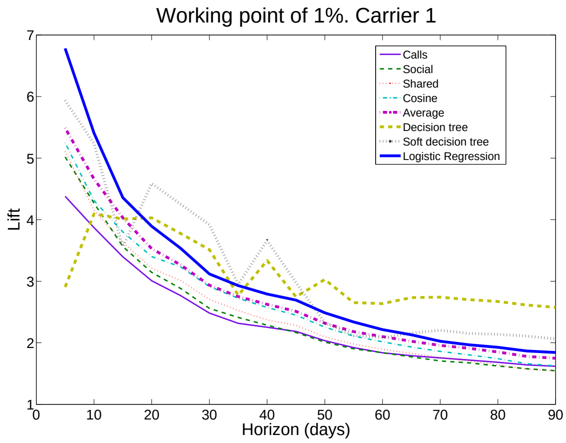

Figure 4: Lift as function of time horizon, on all basic measures and combinations in working point of 1%. This figure is best viewed in color.

4.3 Combining Multiple Diffusion Scores The previous section demonstrated that differences between the performance of similarity measures exist. Therefore, we hypothesize that each measure may be better at identifying different churners. Hence, in this section we demonstrate that combining multiple diffusion measures can improve the quality of the results for churn prediction.

First, the Spearman correlation [19] between the scores resulting from the different relationship measures is shown in Table 2, for different fractions of the population. The top 1% refers to 1% of the population which received the highest averaged score across all measures. Interestingly, for most pairs of measures, the correlation coefficients that correspond to these top 1% of scores are small in their absolute value and differ in their sign, while the correlation for the entire population is somewhat higher and always positive. This suggests that there is a large disagreement between measures on the ranking of the most likely churning subscribers, and

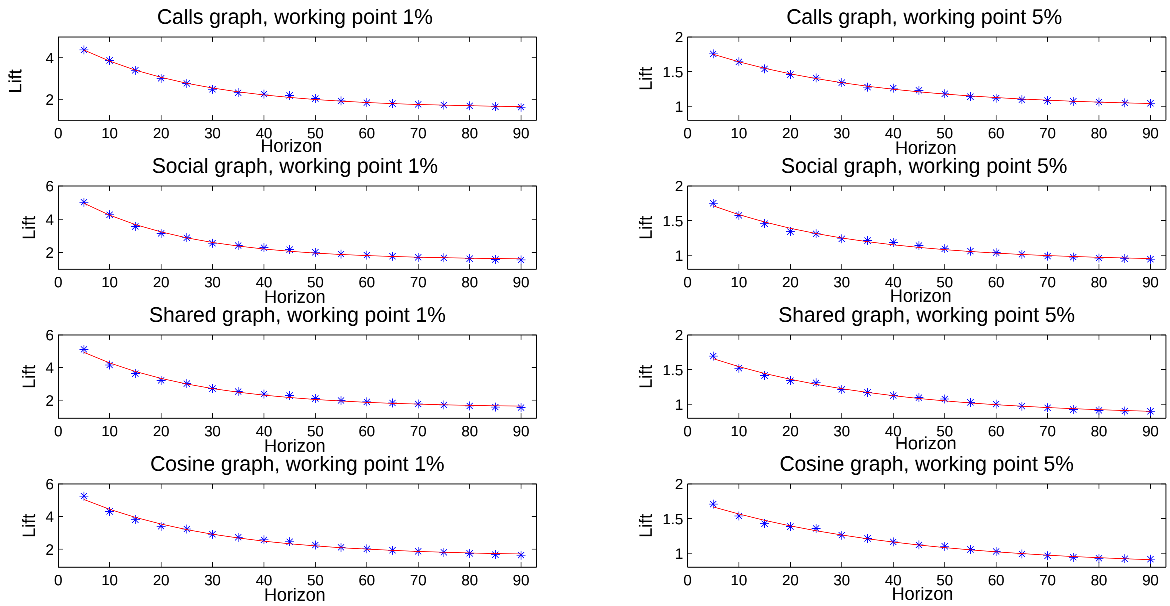

Figure 3: Lift as function of time horizon (asterisks), with best exponential fit (solid line). Each plot corresponds to one of the similarity measures at a given lift working point.

Carrier	Carrier 1		Carrier 2	
	Top 1%	All scores	Top 1%	All scores
Calls vs. Social	-0.054	0.572	-0.018	0.286
Calls vs. Shared	0.283	0.586	0.077	0.494
Calls vs. Cosine	0.302	0.576	0.075	0.452
Social vs. Shared	0.151	0.666	0.061	0.412
Social vs. Cosine	0.153	0.658	0.139	0.458
Shared vs. Cosine	0.928	0.982	0.878	0.907

Table 2: Spearman correlation coefficient values

a general agreement on the scoring of less likely churn candidates. The only pairs that have a high correlation are shared and the cosine. This is not surprising because these measures are highly related to each other by definition, and they form the same network structure (for a more detailed discussion see Section 2.3).

We tested several classifiers for combining the four similarity measures. The inputs to the classifiers were the diffusion scores, and the classifiers tested were logistic regression and regression tree. Since the regression tree maps large fractions of the populations to the same score, we also tried to smoothen the scores by adding to each tree score the average over all measures. We term this soft decision tree. The classifiers were constructed for each horizon using the first (training) period, and applied to the test periods. For comparison, we also provide the lift obtained by a simple averaging of the four diffusion scores.

The results of this analysis are shown in Figure 4, which depicts the lift values over time, and in Table 3, which shows the average lift values, using the calls measure as the baseline. As can be seen, the decision and regression trees offer the best improvements in lift, obtaining results which are far superior than the ones obtained for each measure separately, with an improvement over the baseline of over 50% for the first carrier and around 18% for the second. This lends additional evidence to the finding that different measures identify different churners, and that these can be exploited through a learning algorithm. Our finding may also hint that different measures identify subscribers who churn for different social influences. This, however, requires additional study.

Figure 5 shows the structure of the regression tree for the working point of 20 days for the first carrier. The tree was pruned to a depth of five levels to avoid over fitting. As can be seen, all measures we tested appear in the tree. Interestingly, the most indicative feature is the social measure, where a low social score indicates a lower propensity to churn (which is further partitioned using the calls measure). Conversely, a high social score, combined with a high score in cosine and shared is indicative of an extremely high churn likelihood.

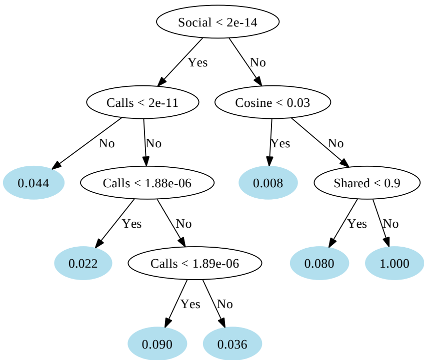

Figure 5: Decision tree structure for horizon of 20 days (Carrier 1). Values in leaf nodes indicate the likelihood of churn for subscribers mapped to these nodes.

Measure	Carrier 1	Carrier 2
Social measure	1.70	5.73
Shared measure	4.86	-31.88
Cosine measure	11.62	-19.00
Average measure	15.34	-2.16
Decision tree	38.79	18.46
Logistic regression	26.60	-18.22
Soft decision tree	54.5	6.19

Table 3: Average percentage improvement in diffusion performance using different similarity measures and combination classifiers, compared to the calls measure. Predictive horizon is averaged between 15 and 25 days, and the lift fraction is 1%.

5 Conclusions and Future Work

Diffusion processes play a central role in applications of viral marketing and social networks. A deeper understanding of the properties of such processes can lead into theoretical and practical insights on these applications. In this context, churn is a useful ground truth because it is relatively easy to define and measure.

In this paper we studied a number of phenomena related to the usage of diffusion based algorithms for churn prediction in Telco networks. We demonstrated three main phenomena: the dependence of the accuracy of the prediction on the prediction horizon, the effect of the social affinity measure used, and the usefulness of combining social affinity measures for enhancing the performance of churn prediction algorithms.

Several research directions stem from this work: Our paper showed that significant gains can be gained from both the usage of different social affinity measures and their combination via ensemble methods. We believe that these gains could be further improved by both introducing additional measures and by additional ways of combining them. Additionally, as is the case in many such application, each carrier may find that different measures are the best for their specific population, though the social measure is the best single measure in the measures we tested.

The fact that different measures identify different churners is interesting from a social perspective, as it may be due to different modes of churn or to differences in the churning subscribers. Future work will investigate this phenomena in more detail.

The deterioration of prediction quality as a function of the time horizon has implications to marketing applications. A theoretical explanation of this phenomenon may lead into novel algorithms and deeper insights. One possible conjecture is that this phenomenon stems from the expander like structure of such networks. Yet, a generic theoretical explanation of this deterioration seems challenging.

References

[1] L. A. Adamic and E. Adar. Friends and neighbors on the web. *SOCIAL NETWORKS*, 25:211–230, 2001.

[2] D. Birke. Diffusion on networks: Modelling the spread of innovations and customer churn over social networks. *Workshop on Formation of Social Networks in Social Software Applications*, 2006.

[3] K. Coussement and D. Van den Poel. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Syst. Appl.*, 34(1):313–327, 2008.

[4] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi. Social ties and their relevance to churn in mobile telecom networks. In *EDBT '08: Proceedings of the 11th international conference on Extending database technology*, pages 668–677. ACM, 2008.

[5] P. Datta, B. Masand, D. R. Mani, and B. Li. Automated cellular modeling and prediction on a large scale. *Artif. Intell. Rev.*, 14(6):485–502, 2000.

[6] E. de Oliveira Lima. *Domain knowledge integration in data mining for churn and customer lifetime value modelling: new approaches and applications*. PhD thesis, University of Southampton, School of Management, 2009.

[7] S. Doyle and S. Barn. The role of social networks in marketing. *The Journal of Database Marketing & Customer Strategy Management*, 15(1):60–64, October 2007.

[8] R. K. Gopal and S. K. Meher. *Customer Churn Time Prediction in Mobile Telecommunication Industry Using Ordinal Regression*, volume 5012 of *Lecture Notes in Computer Science*, pages 884–889. Springer, 2008.

[9] J. Hadden, A. Tiwari, R. Roy, and D. Ruta. Computer assisted customer churn management: State-of-the-

art and future trends. *Computers and Operations Research*, 34(10):2902–2917, 2007.

[10] J. Kawale, A. Pal, and J. Srivastava. Churn prediction in mmorpgs: A social influence based approach. *Computational Science and Engineering, IEEE International Conference on*, 4:423–428, 2009.

[11] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '03, pages 137–146, New York, NY, USA, 2003. ACM.

[12] A. A. Nanavati. Personal communication.

[13] I. Nitzan and B. Libai. Social effects on customer retention. *Marketing Science Institute (MSI) Working Paper*, 2010.

[14] P. C. Pendharkar. Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services. *Expert Syst. Appl.*, 36(3):6714–6720, 2009.

[15] R. Peres, V. Mahajan, and M. Eitan. Innovation diffusion and new product growth: A critical review and research directions. *International Journal of Research in Marketing*, (1):91–106, 2010.

[16] Y. Richter, E. Yom-Tov, and N. Slonim. Predicting customer churn in mobile networks through analysis of social groups. In *SDM*, pages 732–741, 2010.

[17] P. Sarkar, D. Chakrabarti, and A. Moore. Theoretical justification of popular link prediction heuristics. In *23st Annual Conference on Learning Theory - COLT 2010*, 2010.

[18] G. Song, D. Yang, L. Wu, T. Wang, and S. Tang. A mixed process neural network and its application to churn prediction in mobile communications. In *Proceedings of the Sixth IEEE International Conference on Data Mining Workshops (ICDM Workshops)*, pages 798–802, 2006.

[19] C. Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(2):72–101, 1904.

[20] H. Toledano, E. Yom-Tov, D. Pelleg, E. Pednault, and R. Natarajan. Support vector machine solvers: Large-scale, accurate, and fast (pick any two). Technical Report H-0260, IBM Research, 2008.

[21] J. Yang, X. He, and H. Lee. Social reference group influence on mobile phone purchasing behaviour: a cross-nation comparative study. *International Journal of Mobile Communications*, 5(3):319–338, 2007.