

# The effect of social affinity and predictive horizon on churn prediction using diffusion modeling

Dorit Baras · Amir Ronen · Elad Yom-Tov

Received: 2 October 2013 / Revised: 8 September 2014 / Accepted: 16 September 2014 / Published online: 26 September 2014  
© Springer-Verlag Wien 2014

**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 predicting customer turnover, or churn, in mobile communication carriers. These predictions are most accurate when specific social ties are used in the diffusion process, and are primarily useful when they provide a long forecast horizon, so as to enable a service provider to take mitigating actions. Here, we investigate several measures of social affinity and compare their performances for churn prediction, using data from two large mobile phone carriers. Our analysis demonstrates that the various measures of social ties capture different calling and texting patterns, and that a significant improvement in the accuracy of prediction is reached by combining them. We study the predictive horizon of diffusion processes and show that it deteriorates significantly as the horizon increases. Our findings underline the usefulness of diffusion processes for enhancing churn prediction while providing insights to their limitations.

**Keywords** Churn prediction · Diffusion processes · Telco · Social networks

---

D. Baras (✉) · A. Ronen  
IBM Research-Haifa, Haifa University Campus, Mount Carmel,  
3498825 Haifa, Israel  
e-mail: doritb@il.ibm.com

A. Ronen  
e-mail: amirro@il.ibm.com

E. Yom-Tov  
Microsoft Research, 13 Shenkar st., 46875 Herzeliya, Israel  
e-mail: eladyt@microsoft.com

## 1 Introduction

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

Churn is one of the most costly items affecting a telecom carrier's bottom line as it decreases revenue. Additionally, a carrier's investment in winning a new customer is far greater than the cost of preserving an existing one (Hadden et al. 2007). In many cases, churn can be lessened by offering customers incentives to remain customers of the existing carrier. Therefore, salient reasons exist for predicting which customers are likely to churn in the near future as a means to try and prevent their churn.

In addition to identify 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 a customer and make a retention offer. Second, a long forecast horizon is advantageous in that the further away a 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, Coussement and Van den Poel 2008; Datta et al. 2000; Gopal and Meher 2008; Pendharkar 2009; Song et al. 2006; Neslin et al. 2006; Idris et al. 2012; Huang et al. 2012; Lu

et al. 2012; Jadhav and Pawar 2011) consider each customer individually. The goal of these systems is to predict which customers are likely to churn in the immediate future, usually defined as between 1 and 3 months. Such systems rely on hundreds of complex key performance indicators (KPIs) that are generated for each customer from their call attributes, financial attributes, and service information—all evaluated over long periods (many months and years). 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 attributes (calling patterns or otherwise) prior to switching carriers. While this may be true in some cases, certainly many scenarios occur in which these assumptions are violated. For example, customers who have come to believe that they have found a better deal with a competitor may churn immediately.

The dominant approach to churn prediction focuses exclusively on the individual customer without taking social influence into account. Clearly, churn entails many social aspects, as witnessed in other consumer areas (Doyle and Barn 2007; Yang et al. 2007). A dominant example is when a churning customer influences other customers to churn as well. Thus, developing churn prediction systems that take social aspects into account poses an emerging theoretical challenge with potentially great practical implications.

Several works (e.g., Kawale et al. 2009; Birke 2006; Dasgupta et al. 2008; Peres et al. 2010; Dierkes et al. 2011) have started exploring the possibility of using social ties to predict churn. The dominant approach in this direction is known as *diffusion*. The underlying assumptions of diffusion are that recent churners are known and that they are likely to affect the churning decisions of their social neighborhoods. To predict churn, the network of subscribers is modeled as a weighted directed graph on which 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 (Dasgupta et al. 2008).

In addition to churn prediction, diffusion processes play central roles 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 (Ma et al. 2008; Wang et al. 2010; Karnstedt et al. 2010; Iyengar et al. 2011; Gruhl et al. 2004). A recent survey of this literature can be found at Peres et al. (2010) and a pioneering study of graph theoretic problems that stem from some of these applications can be found in Kempe et al. (2003). 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 (2010) showed that exposure to a churning customer increases the likelihood of churn by 80 % (after controlling for homophily, i.e., user similarity). Interestingly, this study also demonstrated that two-thirds of customers who churn do not have an immediate churning acquaintance, 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: the way in which the strength of a connection between customers is estimated, and the specific algorithm used to diffuse energy across the graph.

Most papers to date, which we are aware of, 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 (Dasgupta et al. 2008; Kawale et al. 2009; Peres et al. 2010; Birke 2006). This has an obvious advantage in that it is easy to measure and is easy to justify as a measure of the strength of relationships. Adamic and Adar (2001) suggested that a more informative measure is to count the number of shared friends every pair of customers have between them. Richter et al. (2010) 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 a significant deterioration in performance occurs. 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 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.

the remainder of this paper, we use the implementation of the diffusion algorithm as described in Dasgupta et al. (2008) and according to additional details obtained from personal communication with A. Nanavati. Note that many variants of the above process are possible, e.g., a non-uniform initialization of the churn energies.

---

**Algorithm 1** The diffusion algorithm of Dasgupta et al (2008)

---

1: Calculate  $W_{ij}, \forall i, j$  from calls data:

$$W_{ij} = \# \text{ calls where } i \text{ is the caller and } j \text{ is the callee}$$

2: Set initial energy for all subscribers as follows:

$$E_0(i) = \begin{cases} c, & \text{if } i \in \text{seed (churner);} \\ 0, & \text{else;} \end{cases}$$

3: **repeat**

4: 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:  $\frac{d \cdot W_{ij} \cdot E(i)}{\sum_k W_{ik}}$

5: **until**  $\max_i |E_{new}(i) - E_{old}(i)| < \epsilon$ .

---

## 2 Diffusion processes for churn prediction

### 2.1 Description of the diffusion process

The diffusion process we studied is based on Dasgupta et al. (2008). We describe it in detail in this section (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 the callee (tail of edge). The weight associated with each edge is based on call data, such as the total number of calls or the total duration of calls over a period of time. The seed is a list of subscribers who 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 last 2 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

### 2.2 Additional aspects of the diffusion process

The diffusion process described in Sect. 2.1 gives rise to several natural questions that we attempt to investigate in this work. These include the effect of the prediction horizon, the effect of the used performance measure, and the usage of enhancement techniques. We describe them in detail in the next sections.

#### 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 customer care service, with the main goal being to reach 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 as to be most effectively used to prevent churn by customers at risk, a task which naturally requires time and effort on the part of the telecommunication company. Previous work (Nitzan and Libai 2010) demonstrated a degradation in prediction accuracy over time when using a “word-of-mouth” scenario, for a specific similarity measure and a simple diffusion. Our goal is to ascertain whether the sharp degradation is a general characterizer of diffusion processes for churn prediction and whether it is affected by the choice of relationship measure, addressed in the next section.

### 2.2.2 Relationship measure

Most previous works used a social affinity measure that is based on direct correspondence between two subscribers. Several recent works demonstrated the usefulness of alternative measures for various applications (for example, Adamic and Adar 2001; Sarkar et al. 2010; Richter et al. 2010). 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 this 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, and it provides similar results to measures that are based on the number of minutes (normalized). Let  $\mathbf{v}_i$  denote the calls vector of the  $i$ th caller, where  $\mathbf{v}_i[j]$  is the number of times caller  $i$  called  $j$ . In the calls measure,  $w_{ij} = \mathbf{v}_i[j]$  (where  $w_{ij}$  denotes the weight of the edge from  $i$  to  $j$ ).
- **Weighted number of shared neighbors measure:** The weight is proportional to the number of shared outgoing neighbors (Adamic and Adar 2001) 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 whom a caller is assumed to have stronger connections. Let  $\mathbf{v}_i$  denote the calls vector of the  $i$ 'th caller, where  $\mathbf{v}_i[k]$  is the number of times caller  $i$  called  $k$ . In the shared measure,  $w_{ij} = \sum_k (\mathbf{v}_i[k] \cdot \mathbf{v}_j[k])$ . Note that contribution to the sum is not zero only for callers  $k$  to whom both  $i$  and  $j$  called.
- **Social measure:** This measure, defined in Richter et al. (2010), 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 for two reasons. First, it takes into account the fact that two subscribers talked to each other. Second, when estimating connection strength among linked subscribers, the social environment of the caller and callee is weighted such that a much higher weight is given to cases where two subscribers share a large proportion of their friends. None of the other three measures shares both attributes. Let  $\mathbf{v}_i$  denote the calls vector of the  $i$ 'th caller, where  $\mathbf{v}_i[k]$  is the number of times caller  $i$  called  $k$ . In the social measure,  $w_{ij} = 1[\mathbf{v}_i[j] > 0] \cdot MI(\mathbf{v}_i, \mathbf{v}_j)$  where  $1[\cdot]$  equals 1 when the condition in the brackets holds, and  $MI(\mathbf{v}_i, \mathbf{v}_j) = -P_{i,j} \log_2 P_{i,j} - P_{i,\wedge j} \log_2 P_{i,\wedge j} - P_{\wedge i,j} \log_2 P_{\wedge i,j} - P_{\wedge i,\wedge j} \log_2 P_{\wedge i,\wedge j}$ . Here,  $P_{i,j} = \text{sum}_k(1[\mathbf{v}_i[k] > 0, \mathbf{v}_j[k] > 0])$ ,  $P_{i,\wedge j} = \text{sum}_k(1[\mathbf{v}_i[k] > 0, \mathbf{v}_j[k] = 0])$ ,  $P_{\wedge i,j} = \text{sum}_k(1[\mathbf{v}_i[k] = 0, \mathbf{v}_j[k] > 0])$  and  $P_{\wedge i,\wedge j} = \text{sum}_k(1[\mathbf{v}_i[k] = 0, \mathbf{v}_j[k] = 0])$ .

- **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. Cosine similarity is frequently used as a measure of similarity between sparse entities, for example, in information retrieval, and more recently in social networks Crandall et al. (2008); Kossinets and Watts (2009). Let  $\mathbf{v}_i$  denote the calls vector of the  $i$ th caller, where  $\mathbf{v}_i[k]$  is the number of times caller  $i$  called  $k$ . In the cosine measure,  $w_{ij} = 1[\mathbf{v}_i[j] > 0] \frac{\mathbf{v}_i \cdot \mathbf{v}_j}{\|\mathbf{v}_i\| \cdot \|\mathbf{v}_j\|}$  where  $1[\cdot]$  equals 1 when the condition in the brackets holds and 0 otherwise,  $\mathbf{v}_i \cdot \mathbf{v}_j$  is the dot product of  $\mathbf{v}_i, \mathbf{v}_j$ , and  $\|\mathbf{v}_k\|$  denotes the Euclidean norm of  $\mathbf{v}_k$ .

Naturally, each measure will result in a different network structure, such as in cases in which the measures are asymmetric. Section 2.3 describes these measures in more detail and addresses 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

In many applications, using more features or sources of information can increase model accuracy. In the current study, a natural question is whether the accuracy of diffusion processes can be enhanced by combining scores from several sources. 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 a combination can significantly increase the performance of churn prediction methods.

### 2.3 The interplay between the relationship measure and the coverage

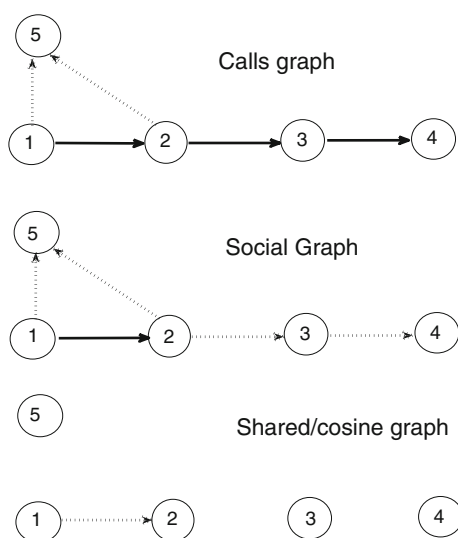
It is important to note that the graphs obtained by each of the measures described above are different in structure (i.e., which nodes are connected through non-zero weighted edges) as well as in the edge weights.

The call graph represents 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 5 once. Caller 2 called caller 3 100 times and caller 5 once, and caller 3 called caller 4 100 times. The call graph that corresponds to this data is shown at the top of Fig. 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 took 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 neither called. The resulting network appears in the middle of Fig. 1. Again, solid lines represent large weights and dashed lines represent small weights. The strength of the 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 the actual settings of the algorithms, the resulting network may not retain connections to nodes 3, 4, 5, or any subset of those.



**Fig. 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 call 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 Fig. 1. As shown, 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 a similar consideration, we can show that the cosine network has the same structure as the shared one, with different weights:  $w_{ij} = v_i \cdot v_j / (\|v_i\| \|v_j\|)$ . Specifically, with the shared number of connections, the link between node 1 and node 2 would be rated as having a strength of 1, which is the minimal strength value between friends who share a connection. This is in compared to cosine, where the strength is 0.0001 (on a range of [0,1]). As will be demonstrated, even though the cosine and shared networks have the same structure, the results in terms of converged energy are different.

Table 1 summarizes the weights that corresponds to each of the measures.

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 effects on the diffusion algorithm to future work.

### 3 Experiments

This section describes the experiments that we conducted to check the hypotheses raised in the paper. We describe in

**Table 1** Values of weights obtained by the various measures for the networks of Fig. 1  $W_{ij}$  corresponds to the weights of the edge from  $i$  to  $j$ .

Weight	Calls	Social	Shared	Cosine
$W_{12}$	100	1.9219	1	0.0001
$W_{23}$	100	1.5219	No edge	No edge
$W_{34}$	100	0.7219	No edge	No edge
$W_{15}$	1	0.7219	No edge	No edge
$W_{25}$	1	0.7219	No edge	No edge

**Table 2** Nodes statistics for the two datasets, for the four similarity measures

Measure	Average number of nodes		Average number of targeted nodes	
	Carrier 1	Carrier 2	Carrier 1	Carrier 2
Calls	7,731K	8,572K	270K	306K
Social	4,325K	6,794K	246K	303K
Shared and cosine	1,581K	4,620K	211K	299K

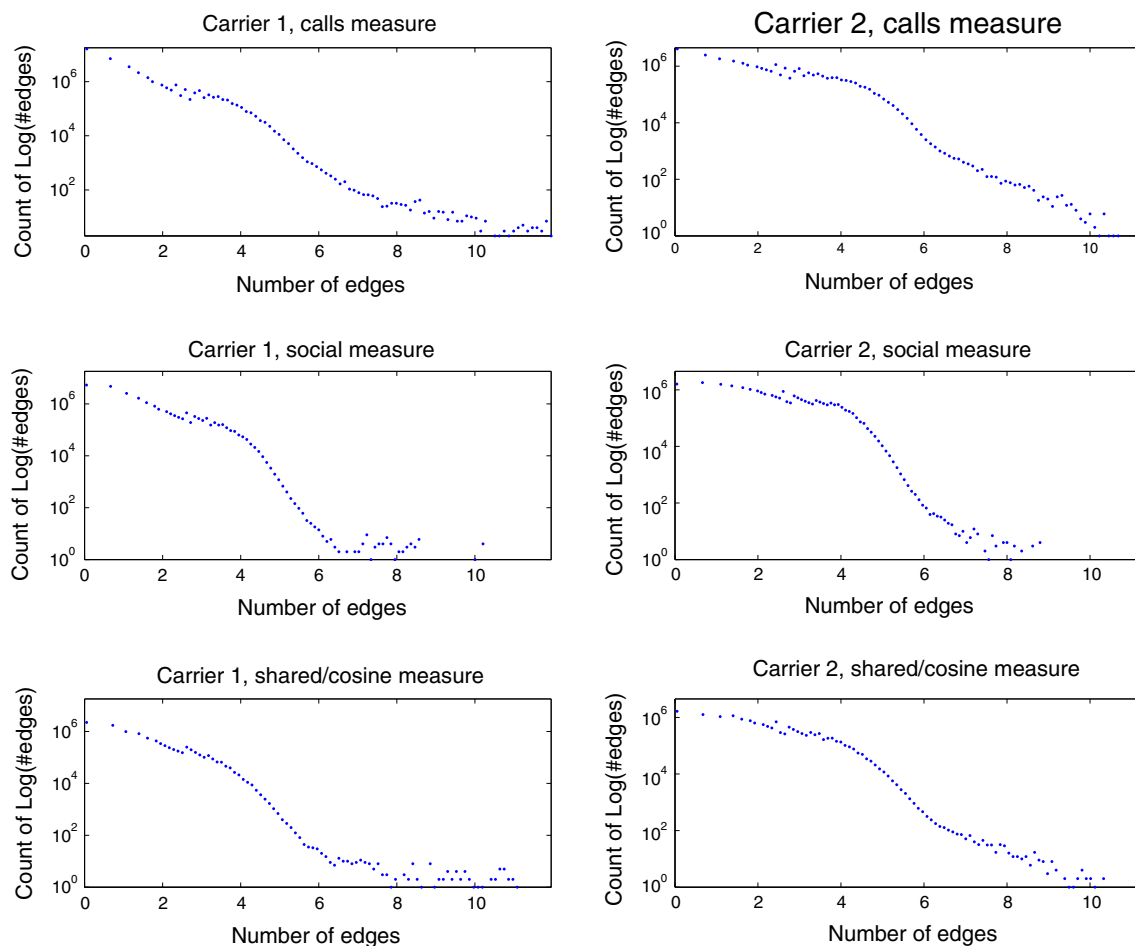
detail the data sets that were used as well as the methods for results evaluation. We present the results of these experiments in Sect. 4.

### 3.1 Data sets

We used call data from two telecommunication providers located in different countries. Each data set contains the call data records (CDRs) of the provider, which is a list of the calls made by subscribers of the provider.

Each record contains the number of the caller, number of the callee, a time stamp of the call and its duration and sometimes additional information. We only used the tuple caller, callee, time stamp. For each carrier, we had a list of churners and their corresponding churn dates. We divided the data into 2-week periods, with an overlap of 1 week between each pair of consecutive periods. The seed (list of churners – nodes with positive energy) is the set of churners who churned during the 2-week period according to the churn dates provided by the operator. When predicting, day 1 refers to the first day after the 2-week data. Unless otherwise noted, the reported results are averaged over these time periods. Note that each experiment utilizes 2 weeks of data to predict from day 1 onwards, and is performed independently of other experiments.

Following the discussion of Sect. 2.3, Table 2 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



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

second, and the shared and cosine graphs contain about 20 % of the nodes of one carrier and 54 % of the second. As explained earlier, an option exists to disregard edges below a certain threshold for the social graph (as in Richter et al. 2010), 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 that represent stronger connections. We did not explore the effects of the threshold parameter. The data summarized in the table demonstrates that indeed different measures yield different networks in terms of nodes and edges.

Table 3 describes the fraction of the edges  $a \rightarrow b$  in which  $b$  churned in a small time period after  $a$ . The fraction is taken over all the edges of the graph. We present the fraction for each metric and for two time periods—up to 1 week and between 8–14 days.

In all relationship measures, we see that more nodes with a small number of connections exist as compared to nodes with a large number of connections. The decrease in all relationship measures is rapid, and closely follow an exponential decay fit ( $R^2 \geq 0.9$  for both carriers on all).

Each carrier defines a subset of the subscribers that are considered relevant for churn prediction. In the induced network we refer to them as the set of targeted nodes. These can be different segments of the population (private, business, etc.) or based on customer value. Table 2 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 (Lima et al. 2009) 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,  $k$  times as many churners exist 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

**Table 3** Fraction of edges that connect churners who churned within certain time differences

Measure	Carrier 1		Carrier 2	
	1–7 days ( $\mu$ )	8–14 days ( $\mu$ )	1–7 days ( $\mu$ )	8–14 days ( $\mu$ )
Calls	3.9	0.55	8.42	2.7
Social	3.03	0.77	9.24	3.15
Shared & Cosine	10.79	1.66	18.21	5.53

$\mu = 10^{-6}$

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 %.

## 4 Experimental results

We implemented the diffusion algorithm of Dasgupta et al. (2008) using the IBM Parallel Machine Learning toolbox (Toledano et al. 2008). 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 in which 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 5 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 % in some cases) and beyond 60 days, the lift values are almost constant.

Figures 3 and 4 plot the lift values of the first carrier for working points of 1 and 5 % correspondingly. Figures 5 and 6 plot the lift values along with the best exponential fit (solid line) of the form  $f(x) = Ae^{-\alpha \cdot x} + B$ , in which  $A, B, \alpha \in \mathfrak{R}$ . These curves represent the lift on the entire

population. Similar phenomena can be observed in both carriers regardless of the fraction of the population. Table 4 presents the values obtained in the exponential fit for the various measures. In this fit, the resulted decay parameters 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 we did reach a close one (the mean value of  $B$  is 1.23). Results on the second carrier are similar. We tested the difference in performance among the four similarity measures for statistical significance using the Fisher paired test (Demšar 2006). The social measure was found to be statistically

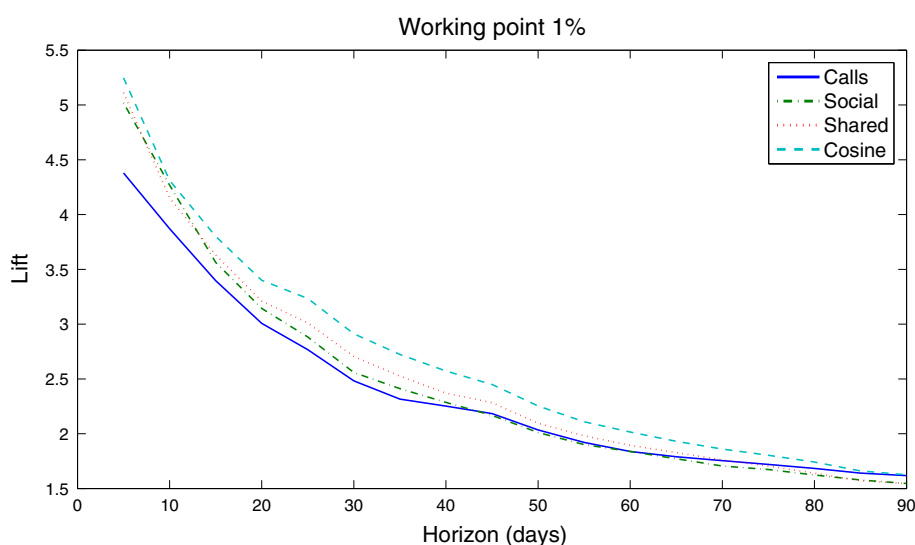
significantly better than the other three measures. The null hypothesis, namely, that the other 3 measures (calls, cosine, and shared) have indistinguishable performance could not be rejected at  $P = 0.05$ .

In Nitzan and Libai (2010), the authors 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.

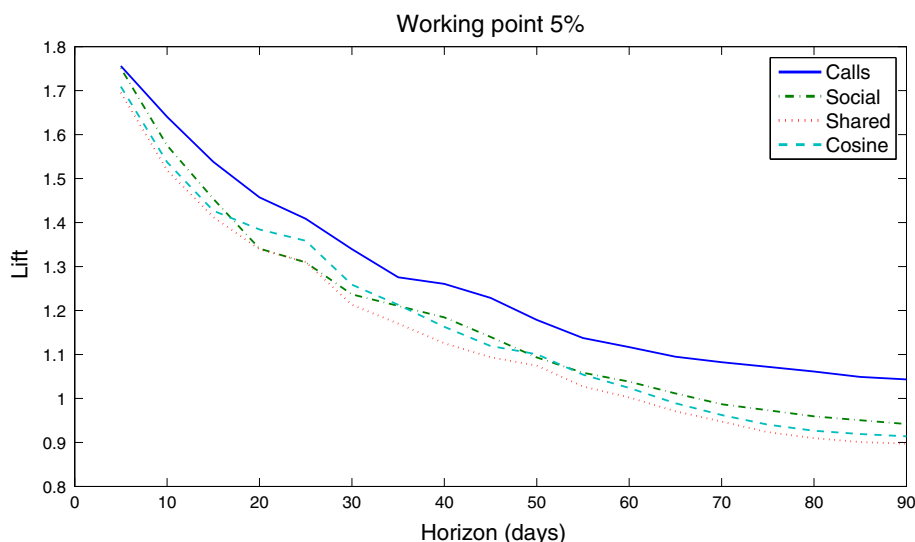
#### 4.2 The effect of different relationship measures

As mentioned above, we tested several relationship measures on the same populations, time frames, and predictive

**Fig. 3** Lift as function of time horizon for working point of 1 %

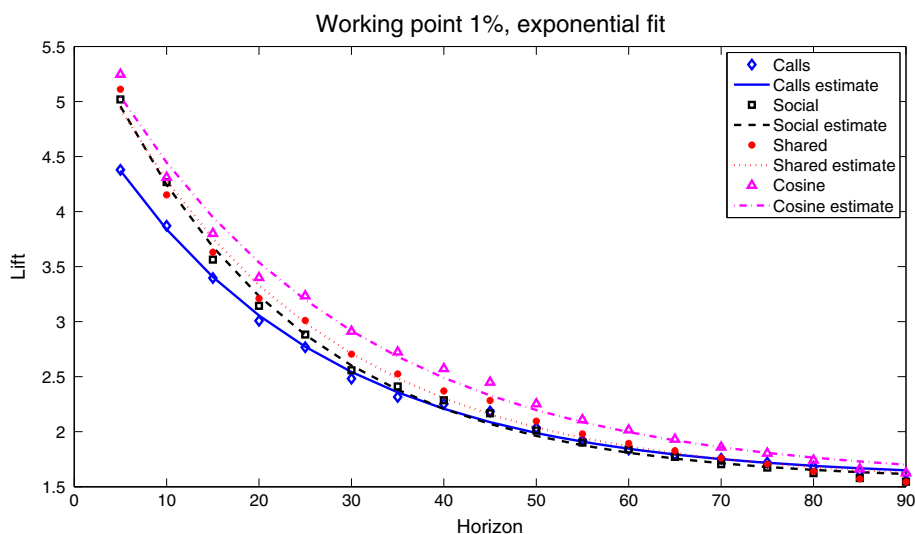


**Fig. 4** Lift as function of time horizon for working point of 5 %

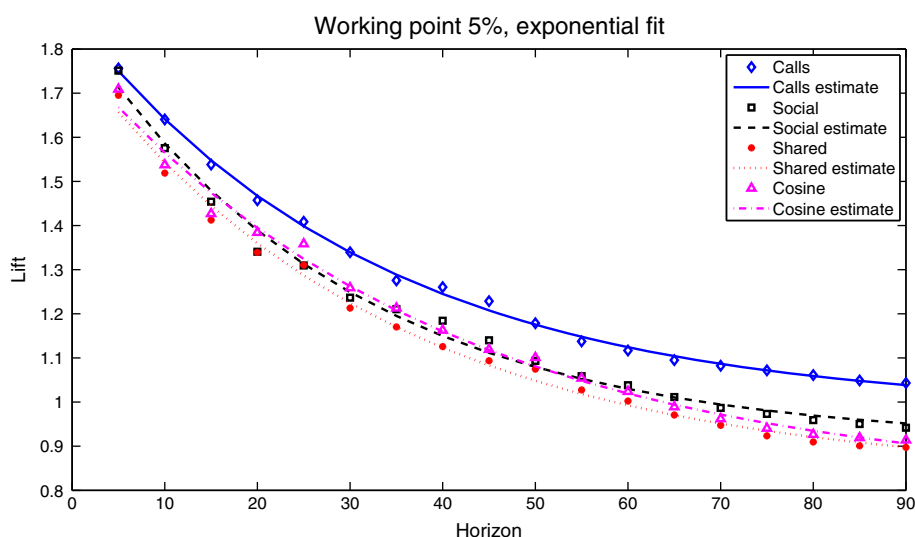




**Fig. 5** Lift as function of time horizon for working point of 1 %, with best exponential fit (solid line). This figure is best viewed in color



**Fig. 6** Lift as function of time horizon for working point of 5 %, with best exponential fit (solid line). This figure is best viewed in color



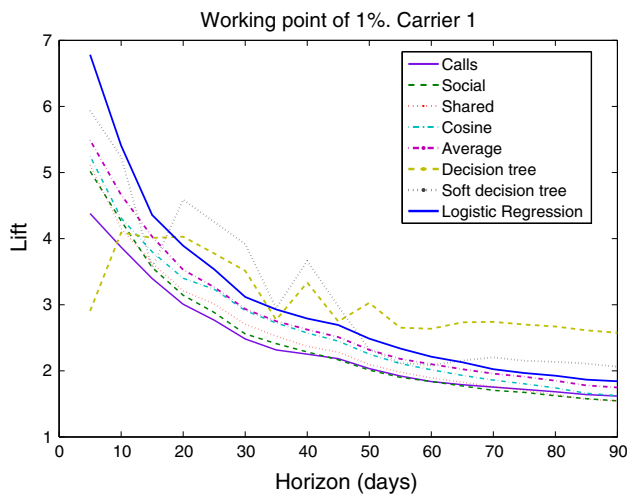
horizons. This allowed us to estimate the contribution of the measures directly, given that all other parameters were equal.

Figure 7 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, up to horizon of 40 days, the calls measure is outperformed by the social, shared, and cosine measures, as well as some combined measures to be discussed at the next section. Table 6 quantifies these improvements, averaged over several prediction horizons, and shows that the differences between the individual measures are often significant. Additionally significant, the social measure outperforms the calls measure in both carriers. Therefore, we deduce that the social measure is a good can-

**Table 4** Fitted parameters of the exponential fit

Measure	Working point of 1 %			Working point of 5 %		
	A	B	$\alpha$	A	B	$\alpha$
Calls	3.4601	1.5739	0.0424	0.8958	0.9812	0.0306
Social	4.306	1.5542	0.0471	0.9559	0.9085	0.0344
Shared	4.1966	1.5400	0.0426	0.9571	0.8319	0.0297
Cosine	4.2088	1.5595	0.0378	0.9771	0.8977	0.0255

didate 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



**Fig. 7** Lift as functions of time horizon, on all basic measures and combinations in working point of 1 %. This figure is best viewed in color

**Table 5** Spearman correlation coefficient values

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

### 4.3 Combining multiple diffusion scores

The previous section demonstrated that similarity measures differ in performance. Therefore, we hypothesize that each measure may be better at identifying a different set of 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 (Spearman 1904) between the scores resulting from the different relationship measures is shown in Table 5, 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 values and differ in their signs, while the correlation for the entire population is somewhat higher and always positive. This suggests a large disagreement between measures on the ranking of the most likely churning subscribers, and a general agreement on the

**Table 6** Average percentage improvement in diffusion performance using different similarity measures and combination classifiers, compared to the calls measure

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

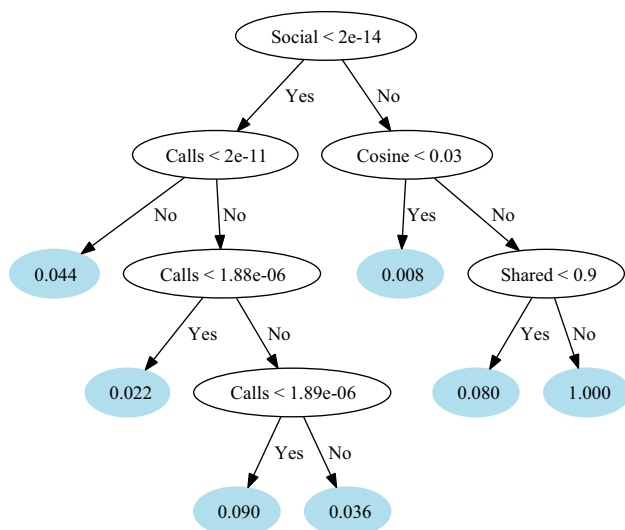
Predictive horizon is averaged between 15 and 25 days, and the lift fraction is 1 %

scoring of less likely churn candidates. The only pairs that have a high correlation are shared and the cosine. This is not surprising, given that these measures are highly related by definition, and they form the same network structure (for a more detailed discussion, see Sect. 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 the 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 Fig. 7, which depicts the lift values over time, and in Table 6, which shows the average lift values, using the calls measure as the baseline. As shown, the decision and regression trees offer the best improvements in lift, obtaining results that 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 8 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. All measures we tested appear in the tree. Interestingly, the most indicative feature is the social measure, in which a low social score indicates a lower propensity to churn (which is further partitioned using the calls measure).



**Fig. 8** 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

Conversely, a high social score, combined with a high score in cosine and shared is indicative of an extremely high churn likelihood.

## 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 to 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 addressed a number of phenomena related to the use of diffusion-based algorithms for churn prediction in Telco networks. We demonstrated three main phenomena: the fact that accuracy deteriorates sharply with 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.

During this study, we fixed several parameters of the diffusion algorithm such as the initial energy and the spreading coefficient  $d$ . The values were set as suggested in Dasgupta et al. (2008), which also provide a discussion on the effect of changing the value of  $d$ . We leave for future work to check whether the optimal value of  $d$  should depend on the chosen social affinity measure, and whether the initial energy has a major effect on the results.

Several research directions stem from this work: our paper showed that significant gains can be achieved from both the use 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 applications, each carrier may find that a different set of measures is best for its specific population, though the social measure is the best single measure among those we tested.

The deterioration of prediction quality as a function of the time horizon has implications for 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, deducing a generic theoretical explanation of this deterioration would be challenging.

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 should investigate this phenomena in more detail.

**Acknowledgments** We thank Amit A. Nanavati for clarifying some implementation details of Dasgupta et al. (2008).

## References

- Adamic LA, Adar E (2001) Friends and neighbors on the web. *Soc Netw* 25:211–230
- Birke D (2006) Diffusion on networks: modelling the spread of innovations and customer churn over social networks. *Workshop on Formation of Social Networks in Social Software Applications*
- Coussement K, Van den Poel D (2008) Churn prediction in subscription services: an application of support vector machines while comparing two parameter-selection techniques. *Expert Syst Appl* 34(1):313–327
- Crandall D, Cosley D, Huttenlocher D, Kleinberg J, Suri S (2008) Feedback effects between similarity and social influence in online communities. In: *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp 160–168
- Dasgupta K, Singh R, Viswanathan B, Chakraborty D, Mukherjea S, Nanavati AA, Joshi A (2008) Social ties and their relevance to churn in mobile telecom networks. In: *EDBT '08: Proceedings of the 11th international conference on extending database technology*
- Datta P, Masand B, Mani DR, Li B (2000) Automated cellular modeling and prediction on a large scale. *Artif Intell Rev* 14(6):485–502
- de Oliveira Lima E (2009) Domain knowledge integration in data mining for churn and customer lifetime value modelling: new approaches and applications. PhD thesis, School of Management, University of Southampton
- Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. *J Mach Learn Res* 7:1–30
- Dierkes T, Bichler M, Krishnan R (2011) Estimating the effect of word of mouth on churn and cross-buying in the mobile phone market with markov logic networks. *Decis Support Syst* 51(3):361–371

- Doyle S, Barn S (2007) The role of social networks in marketing. *J Database Mark Customer Strategy Manag* 15(1):60–64
- Gopal RK, Meher SK (2008) Customer churn time prediction in mobile telecommunication industry using ordinal regression. In: *Lecture Notes in Computer Science*, vol 5012, Springer, pp 884–889
- Gruhl D, Guha R, Liben-Nowell D, Tomkins A (2004) Information diffusion through blogspace. In: *Proceedings of the 13th international conference on World Wide Web*, ACM, WWW '04, pp 491–501
- Hadden J, Tiwari A, Roy R, Ruta D (2007) Computer assisted customer churn management: State-of-the-art and future trends. *Comput Oper Res* 34(10):2902–2917
- Huang B, Kechadi MT, Buckley B (2012) Customer churn prediction in telecommunications. *Expert Syst Appl* 39(1):1414–1425
- Idris A, Rizwan M, Khan A (2012) Churn prediction in telecom using random forest and pso based data balancing in combination with various feature selection strategies. *Comput Electr Eng* 38(6):1808–1819
- Iyengar R, Van den Bulte C, Valente TW (2011) Opinion leadership and social contagion in new product diffusion. *Mark Sci* 30(2):195–212
- Jadhav RJ, Pawar UT (2011) Churn prediction in telecommunication using data mining technology. *Int J Adv Comput Sci Appl* 2(2):17–19
- Karnstedt M, Hennessy T, Chan J, Hayes C (2010) Churn in social networks: a discussion boards case study. In: *IEEE Second international conference on social computing (SocialCom)*, 2010, pp 233–240
- Kawale J, Pal A, Srivastava J (2009) Churn prediction in mmorpgs: a social influence based approach. In: *IEEE international conference on computational science and engineering*, vol 4, pp 423–428
- Kempe D, Kleinberg J, Tardos E (2003) 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
- Kossinets G, Watts DJ (2009) Origins of homophily in an evolving social network. *Am J Sociol* 115(2):405–450
- Lu N, Lin H, Lu J, Zhang G (2012) A customer churn prediction model in telecom industry using boosting. *Industrial Informatics, IEEE Transactions on PP(99)*:1–1.
- Ma H, Yang H, Lyu MR, King I (2008) Mining social networks using heat diffusion processes for marketing candidates selection. In: *Proceedings of the 17th ACM conference on information and knowledge management*, ACM, New York, NY, USA, CIKM 2008, pp 233–242
- Neslin SA, Gupta S, Kamakura W, Lu J, Mason CH (2006) Defection detection: measuring and understanding the predictive accuracy of customer churn models. *J Mark Res* 43(2):204–211
- Nitzan I, Libai B (2010) Social effects on customer retention. *Marketing Science Institute (MSI) Working Paper*
- Pendharkar PC (2009) Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services. *Expert Syst Appl* 36(3):6714–6720
- Peres R, Mahajan V, Eitan M (2010) Innovation diffusion and new product growth: a critical review and research directions. *Int J Res Mark* 1:91–106
- Richter Y, Yom-Tov E, Slonim N (2010) Predicting customer churn in mobile networks through analysis of social groups. In: *SDM*, pp 732–741
- Sarkar P, Chakrabarti D, Moore A (2010) Theoretical justification of popular link prediction heuristics. In: *23st annual conference on learning theory*, COLT 2010
- Song G, Yang D, Wu L, Wang T, Tang S (2006) 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)*, pp 798–802
- Spearman C (1904) The proof and measurement of association between two things. *Am J Psychol* 15(2):72–101
- Toledano H, Yom-Tov E, Pelleg D, Pednault E, Natarajan R (2008) Support vector machine solvers: large-scale, accurate, and fast (pick any two). *Tech Rep H-0260*, IBM Research
- Wang Y, Cong G, Song G, Xie K (2010) Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In: *Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining*, ACM, New York, NY, USA, KDD 2010, pp 1039–1048
- Yang J, He X, Lee H (2007) Social reference group influence on mobile phone purchasing behaviour: a cross-nation comparative study. *Int J Mob Commun* 5(3):319–338