Churn in Social Networks

Churn in Social Networks

Marcel Karnstedt[1], Tara Hennessy[1], Jeffrey Chan[1], Partha Basuchowdhuri[1], Conor Hayes[1], and Thorsten Strufe[2]

[1] Digital Enterprise Research Institute, NUI Galway, Ireland
{firstname}.{lastname}@deri.org
[2] TU Darmstadt, Darmstadt, Germany
strufe@cs.tu-darmstadt.de

1 Introduction

In the past, churn has been identified as an issue across most industry sectors. In its most general sense it refers to the rate of loss of customers from a company's customer base. There is a simple reason for the attention churn attracts: churning customers mean a loss of revenue. Emerging from business spaces like telecommunications (telcom) and broadcast providers, where churn is a major issue, it is also regarded as a crucial problem in many other businesses, such as online games creators, but also online social networks and discussion sites. Companies aim at identifying the risk of churn in its early stages, as it is usually much cheaper to retain a customer than to try to win him or her back. If this risk can be accurately predicted, marketing departments can target customers efficiently with tailored incentives to prevent them from leaving.

Telcom networks, online gaming communities, online communities and discussion forums all have one thing in common: they all can be represented by a network of the social links between people. The links in this social network may be based on calls between customers or explicit and implicit connections extracted from online communities, such as friendship relations, shared activity in popular forums, common contributions to discussion threads and so on. While churn in the telcom sector has been studied extensively, no research has been carried out on the general meaning and consequence of churn in online social networks and communities. This chapter aims at filling this gap. Churn is an important factor for social network providers, as it leads to a loss of social capital and ultimately effects service sustainability. Churn undermines the health and ability of communities to self-govern and self-maintain. The risks to digital social networks arise not only through members stopping their activity, but also through significantly reducing activity. This becomes particularly evident where network services rely upon member activity.

Often, the combination of increasing and decreasing activity (of different customers) is an important indicator for risk. As an example, consider a forum where the popular topic initiators are gradually replaced by spammers. Such a change in the population composition indicates the risk of "losing" the forum, its associated services and revenue model. Therefore, churn analysis should be understood in close relation to analysing activity.

Usually churn prediction is based on pure *feature-based* approaches. This means, key characteristics and features of customers are analysed in order to predict the probability of churning. However, this neglects the importance of social influence between individuals, which can be analysed by examining their social networks. We believe that churn is not only a phenomenon due to individual decisions and profiles, rather it is influenced by external events (e.g., elections or poor reviews for a company) and, more importantly, by community effects. To analyse and understand community effects, a crucial aspect is to understand social roles of single individuals and their influence on the community. Some recent work has identified this issue and argues to replace or extend the feature-based approaches by *influence-based* techniques, mostly by the use of information diffusion models [34, 13]. The idea underlying this approach is that a subscriber is more likely to churn if he is connected to other subscribers that have churned. Thus churn diffuses from subscriber to subscriber, where the degree of influence of the previous churners depends on their importance and social weight. This approach is not only a focus of recent research, it is also receiving attention in the business sector [3].

Measuring the popularity of an individual in a social network can be examined with social network analysis (SNA) measures such as connectedness, betweenness and centrality. Popularity can also be analysed by measuring the use of the network. For example, in terms of of how many profile visits a subscriber receives. For providers of decentralised online social networks, this can be used to optimise the physical structure of the underlying network. But, with the idea of information diffusion and influence, popularity has an increased importance for churn prediction. Intuitively, popular subscribers that decide to churn have influence on the probability of other customers to churn. We propose to understand popularity as a *meta feature* for churn analysis, which can be used in feature-based and diffusion-based models of prediction.

In this chapter, we summarise the field of churn analysis and prediction with focus on digital social networks. We first discuss the differences between churn in the telcom sector and the emerging sector of digital social networks. We present the reasons and motivations underlying user activity. Then, we discuss appropriate definitions of churn and identify research directions and challenges. The general applicability of a definition proposed by us is analysed on the basis of an empirical study. Finally, we present state-of-the-art approaches for churn prediction and highlight their strengths and weaknesses. We discuss the importance of the underlying network structure with reference to observations in recent literature and to the results of our own experiments. We introduce the idea using popularity as an important feature determining influence in social networks. We present a brief analysis of how to determine popularity, where again we highlight the importance of the underlying structural features. Throughout the chapter, we provide implications for further research in order to leverage novel diffusion-based methods for churn prediction by using social roles and popular-

[3] http://yro.slashdot.org/story/09/08/01/1946208/IBM-Uses-Call-Detail-Records-To-Identify-Friends

ity, and discuss how to combine these methods with traditional feature-based approaches.

2 Understanding Churn in Social Networks

In this section, we focus on what the concept of churn means in digital social networks. Churn has been analysed in a wide range of industries: most widely in the telcom sector [44, 50, 41, 66, 17, 32, 13, 65, 30], but also in the field of retail business [12], banking [69], Internet service providers [31], service industries [53], P2P networks [28] and online games [34]. In its most general sense churn refers to customer loss. In the telcom industry, a subscriber is said to have churned when he leaves one carrier to move to another [49, 33]. Churn rate is defined as the total gross number of subscribers who leave the service in the period divided by the average total customers in the period. The churn rate of a telcom company is a key measure of risk and uncertainty in the marketplace and will be quoted in the company annual report[4]. Annual churn rate may be as high as 40 percent while monthly churn rates tend to be around 2 to 3 percent[5]. Several studies suggest that retaining existing customers is considerably less expensive than winning new customers [50], and that new customers tend to be less profitable than existing customers [6]. An excerpt from the 2009 Vodafone annual report illustrates uncertainty and risks inherent in not being able to stem customer churn:

> There can be no assurance that the Group will not experience increases in churn rates, particularly as competition intensifies. An increase in churn rates could adversely affect profitability because the Group would experience lower revenue and additional selling costs to replace customers or recapture lost revenue [55].

As such, there is considerable ongoing research focused on extracting features and developing predictive models so that a telcom provider might intervene before a subscriber moves to a competitor. Lifetime value (LTV) analysis is often used to predict the future profitability of a potential churner so that only the most valuable subscribers are targeted for retention [33]. The idea of segmenting contributors in social networks into different value categories based on their predicted value to the community has only begun to be explored [9]. In Sections 6 and 7, we begin to develop this idea further.

2.1 Reasons for Churn

Factors contributing to churn have intrinsic and extrinsic features, see Figure 1. Intrinsic features relate to inherent qualities of the service/product and/or service provider. For example, Keaveney gives several factors influencing churn

[4] http://www.vodacom.com/reports/ar_2009/pdf/full.pdf, http://www.vodafone.com/annual_report09/downloads/VF_Annual_Report_2009.pdf

[5] http://insidemr.blogspot.com/2006/06/comparison-of-churn-rates.html

in the service industry such as pricing, inconvenience, core service failures, customer service failures and dissatisfaction with provider ethics [35]. In subsequent studies, factors such as these have been decomposed into feature profiles of the customer and customer experience in order to be able to predict churn [49, 33, 1]. Examples of such features include call quality, presence of loyalty points, service usage levels and handset functionality [1]. On the other hand, extrinsic features describe the service in terms of the value it accrues through its social role. This intuition is captured by Metcalfe's "law", which states that the value of a communications network is a function of the number of connected users of the system [46, 3]. Telcom providers that offer discounted calls to the subscribers' family and friends make use of this effect to tie the service to the stability and longevity of a customer's real social network. A related extrinsic function is where the value of a network service is a function of the *opinion* of the customer's social call network. Although there is not a "law" to describe this function, *word-of-mouth* algorithms that combine neighbour preferences to predict the preference of a target user are well established in the domain of on-line recommender systems [27, 18]. Such algorithms explicitly recognise that extrinsic, social features are as important as intrinsic features in determining the perceived value of an object or service. While much research has been carried out on how intrinsic features can be used to predict potential churners, there is comparatively little research on how extrinsic, social features contribute to churn. In Section 6, we explore this idea in greater detail.

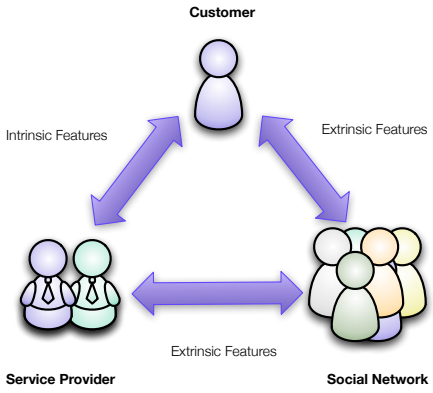


Fig. 1. Intrinsic and extrinsic features of churn

Value in Online Communities In this section, we examine the value that subscribers derive from activity in online digital networks. As we will see, such value is tied directly to the social capital of the network. The loss of capital through subscriber loss or significant drop in activity may precipitate spiralling

cycles of churn. A key observation of user behaviour in online networks is that users, with the exception of spammers, make contributions to online discourse without expecting any immediate return [39, 11]. In sociological discourse, this type of activity is described in terms of the 'gift economy' [58]. In contrast to the commodity or service economy, which is driven by the exchange of good/services for money, economic exchange in the gift economy is defined in terms of an implicit social contract. In a gift transaction, there is an unstated expectation that the benefits of a gift will be reciprocated by the recipient at some reasonable time in the future. A more risky transaction involves 'generalised exchange', whereby the giver's generosity is reciprocated, not by the recipient, but by someone else in the group. In social networks, this exchange mechanism applies to those contributors who give of their time and expertise but do not appear to receive immediate benefits. However, there is a risk that the group will not assume responsibility for the debt and the contributor will never be reimbursed in kind. In the worst case, if all members of the group never contribute (free-load), no one benefits and the exchange system breaks down. Despite this risk, several motivations for contribution to digital social networks have been proposed [39, 11]. One such is the contributor's expectation of reciprocity under the generalised exchange system. Contributors expect that there will be sufficient payback in terms of information and agreeable social relations from other contributors over time. Another motivation is the contributor's desire to enhance his reputation and prestige by being recognised as a source of valuable information or help. The contributor may also be motivated by trying to have his/her ideas adopted by others in the group. A further motivation concerns the value the contributor places in his attachment to the group or the values represented by the group. Thus, contribution is partly made to sustain the group and the contributor's continuing attachment. For each motivation, the contributor derives value in part from his contribution in a social context. We can hypothesise that churn is an outcome that tends to occur when the expected return drops below a certain threshold - perhaps the cost of contribution. In short, we hypothesise that the contributor's utility function is dependent on the social characteristics of the network.

2.2 Churn in Digital Social Networks

As we discussed earlier, the subscriber to an online community is weakly-tied by a non-binding social contract [11]. On the other hand, the telcom subscriber is generally bound by a service contract or by the purchase of advance credit. In general, subscribers have a contract with one telcom provider and tend to break that contract in response to industry-understood 'triggering events', such as the expiry of the service contract or a poor customer service experience [26]. The main prohibition to customer churn is the switching cost involved in moving to another service provider, such as the loss of loyalty points [37, 43]. Online communities do not have a switching costs. There is a low-entry barrier to join and the contributor may leave without incurring an explicit penalty. More importantly, the contributor may return at any time, and may have simultaneous in-

volvement in several communities. As such, churn is defined by telcom providers in terms of complete loss of custom - the customer has moved to a rival network, while in online communities the notion of churn is less easily defined.

In fact, the ease with which contributors can alter their behaviour online suggests that, in contrast to the telcom view, churn is a normative behaviour in social networks and that there will always be an an underlying turnover in contributor activity. In P2P networks, churn is defined as the collective effect of the independent join-participate-leave cycles of all peers observed over a given period. Churn in P2P networks effects key design parameters and the resiliency and structure of the overlay network [62]. While this introduces the notion of cyclic join-participate-leave behaviour in low entry cost dynamic systems, P2P churn differs from telcom and social network churn because there are no social or contractual ties affecting participation, i.e., peer behaviour is independent. In P2P simulation analysis, Herrera and Znati [28] introduced the concept of different classes of peers, some of whose behaviours play a crucial role in the applications that the network can support. Current research in social networks suggests that contributors can be classed according to different roles, and that the presence of roles such as 'popular initiator' or 'supporter' are important for the health and sustainability of the network [9]. Likewise, telcom churn prediction focuses on identifying the profitable subscribers that are about to churn. As such, churn in social networks refers to turnover where there is a net loss of the types of contributors that are needed to maintain the service level of the network. Therefore, churn can be analysed on different levels: per user, per thread, per forum, per community, etc. Unlike in telcom networks where churn is defined in terms of a customer's complete defection, a contributor in a social network may be deemed to be a 'partial defective' if his activity drops significantly below previously established levels of engagement. The overall effect of churn on a social network may be a function of the number of partial defectives who are also high value contributors. By moving from a binary decision (activity stopped or not) to one based on a significant drop in activity, the question of an appropriate churn definition arises. We discuss this in the next section. Moreover, as we show as well, this results in the existence of different types of churn, depending on the significance and duration of drop in activity – and the actual risks and defectives that shall be identified.

3 Definitions of Churn in Digital Social Networks

As discussed in Section 2, the low cost in contributing and in exiting means that online contributors demonstrate behaviour that cannot easily be categorised in terms of binary churn. In this section, we propose an intuitive definition for churn and several variants, while indicating the implications and suitability of different alternatives for several aims and domains. In its most general sense churn is perceived as a significant and sustainable change in the activity of single individuals and/or communities. We argue that no general definition exists that would be the best for each case. Rather, the chosen definition should depend on

the application domain and the risks that churn analysis and prediction should indicate. This section is not meant to provide an exhaustive discussion of possible definitions. Rather, its main purpose is to highlight the crucial points for choosing such a definition.

Currently, churn figures highly in research on predicting customer loyalty. Thus, an intuitive definition is a 'partial defective' – customers whose established buying or usage patterns drop significantly, suggesting that they have moved most of their custom elsewhere.

Definition 1 ***Individual Churn:*** *The* previous activity (PA) *window is a time window consisting of time steps t_1 to $t_1 + n - 1$ inclusive, $n \in \mathbb{N}, n \geq 0$. Let $\mu_{PA}(v_i)$ denote the average activity of a user v_i over the previous activity window. The* churn (C) *window is a time window $t_2 = t_1 + n$ to $t_2 + m - 1$ inclusive, $m \in \mathbb{N}, m \geq 0$. Let $\mu_C(v_i)$ denote the average activity of a user v_i over the churn window. A user v_i is considered to have churned during the churn window if:*

$$\mu_C(v_i) \leq T(\mathcal{S}) \cdot \mu_{PA}(v_i)$$

$0 \leq T(\mathcal{S}) < 1$ is a threshold factor dependent on the relevant system parameters $\mathcal{S}$.

The definition is more simply stated as:
A user has churned if his or her average activity over a window of m time steps has dropped to less than a fraction T of their average activity in the previous n time steps, where m and n are positive integers and $T < 1$ is a suitable system-dependent threshold factor.

In Section 4, we investigate the sensitivity of the above definition to varying parameters n, m and T. Note that it is also possible to define the threshold factor dependent on the user him- or herself, i.e., use v_i as an additional parameter for T. The identification of a suitable threshold factor, the system parameters it depends on and the appropriate past activity presents a particularly interesting problem. Several variants of Definition 1 can be differentiated based on these aspects. One alternative is to make churn dependent on whatever constitutes typical activity of the average user in the entire system.

Definition 2 ***Mean activity:*** *Let $a(v_i, t)$ denote the activity of a user v_i at time t, N the total number of time steps in the observations and V denote the set of all users of the network. The mean activity μ_u across all users is defined as:*

$$\mu_u = \frac{1}{N \cdot |V|} \sum_{v_i \in V} \sum_{t=1}^{N} a(v_i, t)$$

Based on this mean activity, the following variants of Definition 1 are intuitive:

- **Global Churn:** Replace $\mu_{PA}(v_i)$ in Definition 1 by μ_u, use a global threshold factor defined as some function $T(\mu_u, \mathcal{S})$, or both.

– **Role Churn:** Define separate $\mu_R(r)$ for each existing role r. The definition of $\mu_R(r)$ follows straightforwardly from Definition 2 by restricting to only users with role r. Then, replace $\mu_{PA}(v_i)$ in Definition 1 by $\mu_R(r)$, use a role threshold factor defined as some function $T(\mu_R(r), \mathcal{S})$, or both.

Roles can be found in a wide range of social networks. For instance, users of forum sites can be classified by the common behaviour roles they play [9]. Some users could play the *popular initiator* role, i.e., tend to initiate many threads and get many replies to their posts. In contrast, other users can play the role of *taciturn*, i.e., have low posting and replying behaviour. The normal and abnormal churning behaviours of the two sets of users are going to be different. Hence, it makes sense to define churn on the basis of the importance of a user, i.e., on the basis of the role they play. Similarly, this can be extended to other levels, dependent on the domain of the system. For instance, for discussion boards, it might be interesting to inspect the typical activity in each forum:

– **Forum Churn:** Define separate $\mu_F(f)$ for each existing forum f. The definition of $\mu_F(f)$ follows straightforwardly from Definition 2 by restricting to only users active in forum f. Then, replace $\mu_{PA}(v_i)$ in Definition 1 by $\mu_F(f)$, use a forum threshold factor defined as some function $T(\mu_F(f), \mathcal{S})$, or both.

All the above variants of Definition 1 are only examples of how to define churn. Some obvious possibilities to further adapt them, without changing their general meaning, are:

(i) Replace the mean by median.
(ii) Choose other relative thresholds.
(iii) Choose absolute thresholds, i.e., do not include any past activity in the churn definition.
(iv) Require the activity levels to be below the threshold for a continuous span of time, rather than on average.

For certain reasons it might be interesting to identify a user as churner as soon as his activity drops below the threshold for the first time. This can be achieved with Definition 1 by setting $m = 1$. As introduced in Section 2, in social networks different types of churn can occur on different levels. The definition is general and flexible enough to allow several different types of churn and related phenomena to be identified and analysed. Examples of such types for discussion boards are:

– Detect the emergence of a new forum population: Forum Churn for existing users and increasing activity of other users in the same forum.
– Detect movement of activity from one forum to another: Forum Churn in one forum and increasing activity of the same users in other forums.
– Detect the change of the role of an individual user: Role churn of the user while staying above the threshold for another role.
– Churn on different levels of an application can always be mapped to the churn of single individuals: a "dying" discussion thread can be identified if most of the posting users are classified as churners, a "dying" forum is composed of dying threads, etc.

We further investigate the sensitivity of the definition with respect to the included parameters and resulting types of identified churn in Section 4. Another popular type of churn definition cannot directly be mapped to Definition 1. Churn can be defined based on the degree of change in activity, i.e., a point in time when the magnitude of the decreasing rate of change of activity of a user is above some absolute threshold. There are several definitions possible based on a decreasing gradient, where the simplest among them is:

Definition 3 ***Gradient Churn:*** *A user v_i is considered to have churned at time $m, t_1 < m \leq N$ if:*

$$-\frac{a(v_i, m) - a(v_i, t_1)}{m - t_1} \geq T(\mathcal{S})$$

where $T(\mathcal{S})$ denotes the threshold associated with this definition.

Simply stated, this definition says:
A user has churned if the absolute slope of his or her decreasing activity over a window of m time steps is above a threshold T, where m is a positive integer and T is a suitable system-dependent threshold.

Again, this definition can be adapted along the dimensions discussed above, i.e., the window size m, absolute and relative thresholds, etc. Other definitions for churn may be based on the ratio between variance and mean of activity, change point detection, etc. However, the discussions presented here highlight the crucial parts in the definition of churn: the applied threshold (factor) T, the window sizes for past activity and churn window as well as the type of past activity used. The right choice is dependent on the domain and the risks that need to be identified. As an example, consider the telcom and online game sectors. To identify likely churners, it is sufficient to identify customers that will drop to an absolute threshold of $T = 0$. The earlier potential churners are required to be identified, the higher this absolute threshold should be. This is based on the observation that usually activity slightly decreases in the months before the decision to churn [34]. The gradient churn definition may be useful to detect churn in any state of the customer lifecycle, rather than only shortly before the churn event. On the other hand, as resources for customer care are usually restricted, providers might want to focus on high-value customers. In this case, applying Role Churn might be the right choice.

A different picture is drawn in social networks that do not imply churning costs for customers, such as discussion boards. Here, risks related to churn materialise in different states and should usually be classified according to the characteristics of the single user, such as his role in a forum. Obviously, relative thresholds for Individual or Role Churn are an intuitive choice in this case. Even if the activity slightly increases after a significant drop, but stays below the critical threshold, this should be understood as an alarming signal in such domains. Whereas in the telcom space it implies that a customer is still active in the network of the operator. In the following section, we present a brief empirical analysis to highlight the effects of different parameters for Definition 1 in such social networks.

4 Empirical Analysis

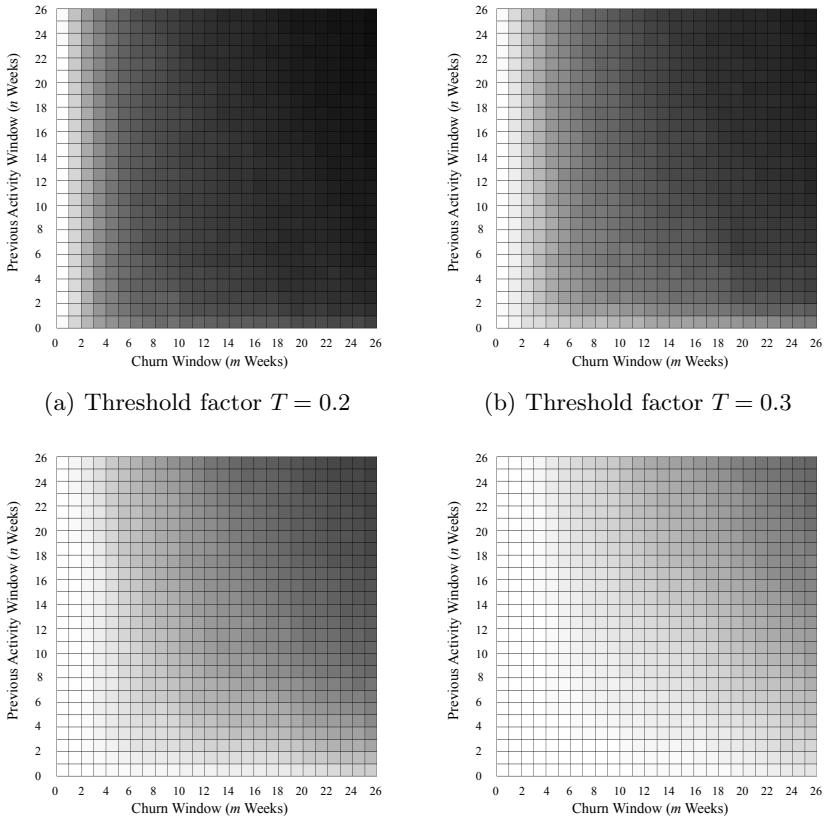
Identification of churn in a network will critically depend on whatever constitutes typical behaviour for a user in that network. In this section, we present a basic empirical analysis on two data sets gained from online networks. We indicate the sensitivity of Definition 1 from Section 3 with respect to the choice of the window sizes m and n, and the threshold factor T. We further show that with different parameter combinations it is possible to isolate significantly different types of churn. First, we briefly introduce the two data sets used.

4.1 Data Set 1: User Activity in a Discussion Board

The first data set, *Set* 1, consists of the number of posts made per week by individual users in individual forums of the popular Irish forum site *boards.ie*[6] over the course of the year 2006. In the following, we refer to the number of posts vs. time as *activity profile*. On initial inspection of the data the activity profiles for single users in *Set* 1 are very discontinuous and "jagged" (particularly in comparison to *Set* 2, see Section 4.2). *Set* 1, in its entirety, consists of almost 150,000 activity profiles. However, a large percentage of the users in *Set* 1 make 1-2 posts and never return. These users are of little interest for churn analysis. The loss of regular users is of much greater concern to a service provider. Thus, the data were filtered to isolate users who made an average of at least 10 posts per week. The remaining sample had a size of 300 users. The number of users identified as churners in any window over the course of 2006 was recorded for a range of window sizes and threshold factors. A maximal churn window size m of 26 weeks is a natural choice, as it is half of the total time spanned by the full data set investigated. We applied Definition 1 to each activity profile. Selected results of this analysis are shown in Figure 2. The individual figures correspond to different threshold factors. The colour gradient ranges from black, when nobody churns, to white, when all of the sample of 300 are identified as churners.

In Figure 2(a) it can be seen that when the threshold factor is very low (20%), the condition for churn is quite strong. Thus, for large window sizes m and n very few users are identified as churners. For smaller window sizes more churners are identified. This is due to the nature of the data, as almost all users have a short drop of activity at some point in time. This explains the white stripe at the far left hand side and shows that such a small window size is not appropriate to identify "real" churn. Along this line the churn window is only one week long, which means that for churn to be identified the user has to be away from the site or has uncharacteristically low activity for only one week. Figure 2(b) shows the same picture for a threshold factor of 30%. Here the condition for churn is weaker, i.e., the drop in activity indicating churn is not as big. Consequently, more churn is identified. In Figures 2(c) and 2(d) it can be seen that as the threshold is raised the conditions for churn become less strict. The number of users identified as churners raises accordingly.

[6] http://boards.ie



(a) Threshold factor $T = 0.2$ (b) Threshold factor $T = 0.3$

(c) Threshold factor $T = 0.5$ (d) Threshold factor $T = 0.7$

Fig. 2. The number of users identified as churners (from black=0% to white=100%) is very sensitive to the chosen parameters: previous activity window size n, churn window size m and threshold factor T.

In Sections 2 and 3, we mentioned the importance of different types of churn that can occur in social networks. In the following, we present a more detailed analysis using a threshold factor of 20%. Regarding the number of identified churners for different window sizes, it is crucial to understand the relations between parameter values and the according types of identified churn. The threshold factor of 20% represents a rather strict condition and is in-line with an intuitive understanding of churn.

An Approximate Hierarchy To analyse the effects of parameter choices in more detail, the parameter combinations at A, B, C, D and E in Figure 3 were chosen. They result in significantly contrasting variants of the definition and thus provide insights into the different types of churn that are identified.

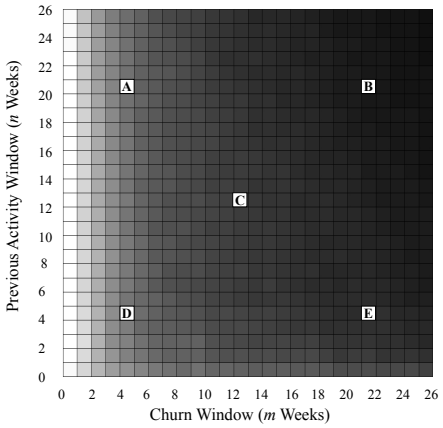


Fig. 3. The five parameter sets chosen for further investigation.

Parameter Set	Previous Activity Window size n	Churn Criteria Window size m	No. of churners (out of 300)
A	21	5	130
B	21	21	32
C	13	13	69
D	5	5	137
E	5	21	53

Table 1. Number of churners identified by the different parameter combinations.

As the parameter choices vary from the bottom left of Figure 3 in any positive direction (i.e., as either of the window sizes increases), the criteria for identifying churn becomes stricter. This holds not only along the diagonal, but also, for instance, as we move from parameter combination D to combination A or E. Table 1 supports this observation by showing the numbers of users identified as churners for the different parameter combinations. Apparently, the different combinations seem to form a hierarchy, meaning that, for example, all churners identified in A are also identified in D. In fact, this is only an approximate hierarchy, illustrated in Figure 4.

Now, the question is what different types of churn are identified by the different parameter combinations. The best way to analyse the differences is by identifying exceptions to the hierarchy. We show examples for these exceptions in Figure 5. In each of the figures, the circles represent the beginning of a churn window in which churn is identified and the stars represent the end of the window. Due to the definition, there may be multiple consecutive churn windows for which churn is detected. The circle and star corresponding to the first churn window are shown in bold.

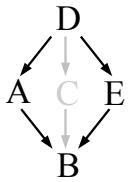


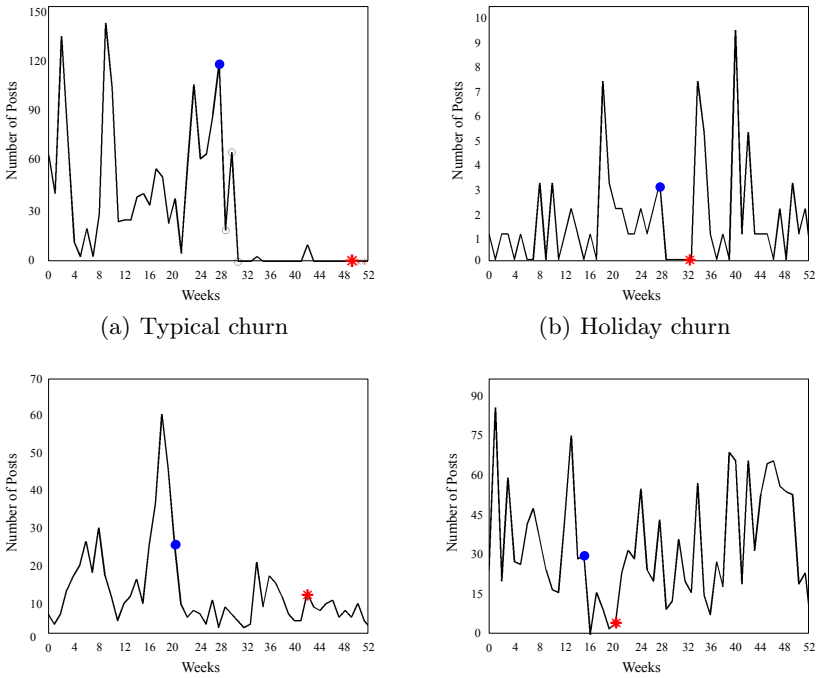
Fig. 4. The approximate hierarchy.

Typical churn: churn identified by parameter combination *B* Criteria B can be seen as the strictest criteria of all five. It requires that the average activity of the user drops significantly for the longest time, while having been high for an equally long time. The users identified by B are those that we see as *typical churners*. An example is shown in Figure 5(a). Out of the 32 churners identified by criteria B, all but 2 are also identified by D and E, all but one are also identified by C and all 32 are identified by A.

Holiday churn: churn identified by parameter combination *A*, but not by *B* or *D* We found examples of churn that are identified by parameter combination A, but not by combinations B and D. Such examples might be called *holiday churn*, where a pattern of consistently high activity is interrupted for a short period of 2-4 weeks. Such short drops in activity have to be pronounced enough to significantly affect the average activity when they fall into a short churn window. But, they do not significantly affect the average when they represent only a small part of a much longer churn window. We show an according example activity profile in Figure 5(b).

Bursty behaviour: churn identified by parameter combination *E* only E is the parameter combination that refers least to an intuitive understanding of churn. It tends to identify short, uncharacteristically high periods of activity that are preceded by long periods of low or even just average behaviour. This is due to the short previous activity window and the long churn window. We refer to this as *bursty behaviour*. An example, which is identified by E but not by any other combination, is shown in figure Figure 5(c).

Inconsistent behaviour: churn identified by parameter combination *D* only Combination D imposes the least strict criteria for churn. Most users have some uncharacteristically low period of activity for some weeks in the year. Since criteria D has two short windows, it flags a kind of *inconsistent behaviour* as churn, see Figure 5(d) for an example. By inspecting the amount of churn identified by combination D but not by any other combination, one could try to measure the consistency of users.



(a) Typical churn (b) Holiday churn

(c) Bursty behaviour (d) Inconsistent behaviour

Fig. 5. Examples for different types of churn from the approximate hierarchy.

Combination *C* Combination C, right in the centre of Figure 3, represents a balance between all other combinations. It works well for identifying typical churn, as well as detecting significant periods of churn before recovery. Due to this generality, it is the combination that we use in Section 6 to analyse network effects for churn.

4.2 Data Set 2: Activity in an Online Social Network

The second data set, *Set* 2, is taken from a popular Central-European business-oriented online social network (OSN) and has been used before in [61]. The gathered data are publicly available and delivered automatically through the Web interface of the OSN. Rather than measuring user activity by, for instance, tracing interactions with other users, we could use an explicit activity measure, provided by the OSN operator. This measure does not allow to conclude on session length or exact times of login. But, as simple tests with specially created profiles revealed, represents a reliable indication of the broad frequency of a user's OSN utilisation. It is correlated with each individual user's activity in the previous days and, as a result, never tends to drop off sharply (except in a small percentage of instances where the activity suddenly becomes 0 – this is very

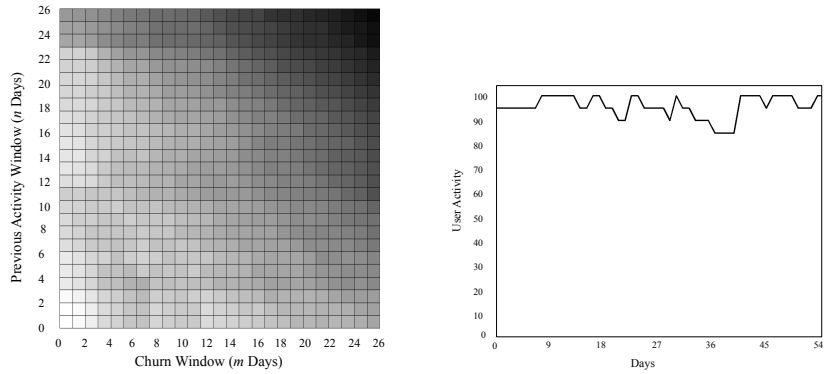
likely due to a user's decision to change privacy settings). The activity measure is quantised to multiples of 5 between 5 and 100. In any one day the measure is never observed to change by more than one step. Since the activity measure is slowly varying, the profiles in *Set* 2 are quite smooth. The activity measure was read at 240 intervals over the course of 54 days for a sample of 31,643 users (21436 male, 10207 female). The sample has been validated in [61]. The time intervals between each reading were not of consistent duration. As the data gathering was performed on the basis of different random walks, only a random portion of the sample users was measured at each of the 240 steps. On average, 66 activity readings are available for each user over the 54 days. However, since the readings were taken very frequently and the activity measure falls off rather slowly with time, we could reconstruct the activity profiles of all included users. Wherever multiple readings were present for one day, the maximum for that day was taken. If no data were available for a day, the gaps were filled by interpolating between existing values. Because of the short time domain associated with *Set* 2 and the smoothness of the pre-processed activity profiles, we omit a deep analysis and focus on showing that the amount of churn identified is similarly sensitive to the parameters as in the data of *Set* 1.

Definition 1 from Section 3 was applied to each activity profile from a random sample of 3000 users (we used a random sample for computational reasons). Again, the number of users identified as churners for any window over the course of the 54 days was recorded for a range of window and threshold sizes. Results of this analysis are shown in Figure 6(a). The colour gradient ranges from black, when no user is identified as churner, to white, when the maximum of 3000 users is identified as churners. For the combinations presented, the maximum number of users identified as churners is 380 out of the 3000. A maximal window size of 27 days is a natural choice, as it is half of the total time spanned by the full dataset investigated. We chose a threshold factor of 50%. Due to the smoothness of the data in *Set* 2, a higher threshold factor is more effective here than for *Set* 1. Figures 6(b) to 6(d) show example activity profiles for this threshold factor and previous activity window size $n = 21$ days and churn window size $m = 5$ days. Figure 6 shows that, although less data is available than in *Set* 1, the general trends in the sensitivity of the definition to the parameters are the same for both data sets.

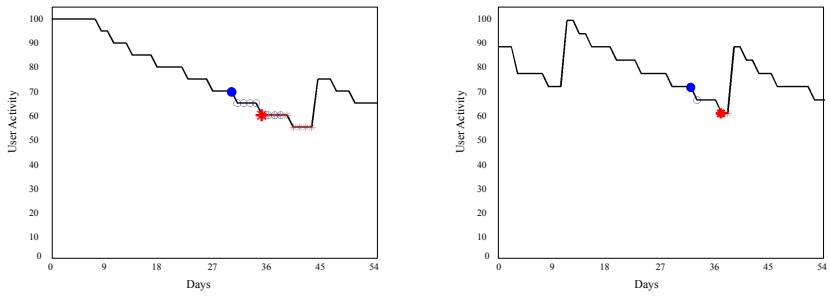
Figure 6 illustrates that again a variety of different churn types can be identified. Figure 6(b) shows the activity profile of a user who is not identified as a churner for any combination of window sizes. Users like this one, whose average activity over the entire period is greater than 85%, make up almost 40% of the entire data set. In contrast, Figure 6(c) shows the activity profile of a user who is identified as a churner for a wide range of window sizes. Finally, a potential periodic pattern of activity is illustrated in Figure 6(d).

4.3 Summary

The contrasting nature of the two data sets makes them interesting case studies for the application of the proposed churn definition. By applying Definition 1



(a) Parameter sensitivity for $T = 0.5$ (b) Non-churner



(c) Churner (d) Potential periodic behaviour

Fig. 6. Churn in *Set* 2.

to *Set* 1, we have shown that the amount and variety of churn identified is highly sensitive to parameter choices. Decreasing the size n of the past activity window increases the sensitivity of the churn detection, i.e., low activity for shorter periods is more often understood as churn. Consequently, increasing n results in detecting more significant churn in an intuitive sense. Increasing the size m of the churn window obviously increases the time until churn is detected as well as the duration of periods of low activity required for indication. Further, we have shown that some interesting types of churn can be isolated by using various combinations of parameters. The data of *Set* 1 is better suited to analyse the meaning of the definition and parameters, as it is not as smooth as *Set* 2. However, the similarities in the trends between Figure 6(a) and Figure 2 indicate the same correlations for *Set* 2 as for *Set* 1.

The brief empirical analysis presented here shows that the proposed definition for churn is suited for a wide range of different social networks. Depending on the actual motivation and needs of the analyser, the parameters can be set to detect several different types of churn. The risks and phenomena that can be identified

by that range from typical churn, over bursty and (in-)consistent behaviour, to sensitively recognised fluctuations in activity. However, other definitions for churn, as well as data from different kinds of social networks, should be analysed and compared in future work. The presented discussion highlights the importance of the correct understanding and modelling of churn, and the significance of carefully investigating this issue before any meaningful prediction or prevention mechanisms can be developed.

5 Models for Churn Prediction

In this section, we briefly summarise the existing state-of-the-art literature on predicting churn. We start with the traditional *feature-based* approaches. Afterwards, we present the few novel approaches based on *social network analysis*.

5.1 Feature-Based Approaches

As introduced in Section 2, traditionally churn prediction has been approached as a feature-based problem. In this section, we list some concrete techniques and refer to according works. The features of a customer are formed by characteristical data available, such as number of purchased products, number of calls, etc. Feature-based prediction models consist of two parts, *feature selection* and the actual *prediction model*.

During feature selection the most important features of a customer are selected in order to reduce the size of the feature vector. Further, focusing on the most significant features usually increases the accuracy and efficiency of the prediction model [70]. However, feature selection is only used if the number of available features is large enough [49, 6, 12]. Several works discuss feature selection for churn prediction in the telcom sector [49] and other service industries, such as banking [16], Pay-TV [7] and newspapers [12].

[63] suggests two general approaches for feature selection, namely sequential forward selection (SFS) and sequential backward selection (SBS). SFS starts with an empty set and subsequently adds features to it, whereas SBS starts with the complete set of features and eliminates one by one based on a performance test. Several concrete techniques for feature selection are applying clustering techniques, such as k-nearest neighbours (kNN) techniques [14] and self-organising maps (SOM) [65], on the values of feature. Features are chosen based on the quality of the resulting clusterings. Though, most approaches combine the selection of features with the training step of the actual prediction model. This is the case when building decision trees (or random forests, which basically are multiple decision trees) based on chosen split criteria [14, 52, 8]. [52] applies an induction algorithm to build decision trees. Usually, the best resulting trees, and thus features, are selected using measures like the minimum error rate. Similarly to decision trees, genetic algorithms use fitness functions [63, 14] and neural networks [47, 65] several epochs and hidden layers to select features during the training phase.

Statistical survival analysis has been used to predict churn and is the approach taken in [41]. Ordinal regression has been proposed [22] as an alternative to survival analysis techniques for churn prediction. Latent semantic analysis has also been used to predict churn among insurance policy holders [48]. In this work, the authors highlight the difficulties associated with dealing with time-stamped data. The applicability and suitability of the different approaches varies from one case to another and none of them can be clearly preferred in any of the churn-related industries.

Probably the most popular method for churn prediction (and included feature selection) is logistic regression. It has been applied very successfully in a wide range of works, e.g., [49, 6, 51, 7, 53, 12, 57, 8]. The general idea behind it is to create a system of linear equations that combine the available feature values of the training data. Solving the system provides the included coefficients. These represent weighting factors for the different features. For prediction, the gained linear equation is computed based on the features of the user to classify. Logistic regression is a probabilistic classifier, conceptually simple and produces robust results. [7] showed that simple linear regression can outperform sophisticated models like multiple-regression models in special cases. Similar to logistic regression, decision trees (and thus, random forests) are easy to use and provide high accuracy in the classification task [49, 6–8, 12, 17, 57, 69, 2]. One advantage of them is that they are easy to modify during the validation phase, e.g., by pruning nodes with high error-rates. Disadvantages of decision trees are their lack of robustness and suboptimal performance in special cases [7]. A third popular class of prediction models is based on Markov chains [56, 7]. Markov chains represent correlations between successive observations of a random variable. Thus, they are especially effective for evolving populations in rather dynamic systems. Consequently, they are popular if churn data over different time slots is available. On the negative side, the runtime complexity of Markov chains can be very high.

Several alternatives to these traditional approaches have been analysed for churn prediction. One example are neural networks [49, 6, 65, 32, 57, 69]. [2] shows that neural networks can outperform decision trees and logistic regression in the case of the analysed Malaysian wireless provider. Other proposals suggest to apply genetic algorithms [17, 30] and support vector machines (SVM) [12, 69] for the prediction task. In most cases, these alternative approaches may perform as well as the traditional ones, but do not provide a significant increase in accuracy or efficiency. Despite the various approaches for feature selection and churn prediction that can be independently combined, there are also a few systems that process all involved tasks in an automated fashion. Two of them are DMEL (Data Mining by Evolutionary Learning) [2] and CHAMP (CHurn Analysis, Modeling, and Prediction) [44, 14]. Both support to load raw data and produce churn prediction based on them, without requiring the user to care for feature selection or the actually used prediction model.

Evaluation metrics for binary classifiers (e.g., decision trees) are the well-known measures based on the number of true and false positives and negatives, such as precision, recall and accuracy. Hit ratio (or precision) and lift (also lift

ratio) are maybe the two most popular metrics used for evaluating the accuracy of churn prediction models. The hit ratio is measured as the ratio of true positives to true positives plus false positives [32, 13]. Lift measures the improvement of a prediction model compared to a classification based on randomly picking classes and is well used in marketing practice. Hung et al. [32] defined lift as the ratio of hit ratio to the monthly churn rate (i.e., to the number of all churners divided by the number of all customers). Each probabilistic classifier (e.g., logistic regression) can be turned into a binary classifier by introducing a threshold. To evaluate probabilistic classifiers without restricting to a specific threshold, ROC analysis can be used. The ROC graph is built by relating 1-*specificity* on the x-axis to specificity on the y-axis for all possible thresholds, where specificity is the number of true negatives divided by the number of all negatives. [15] showed that the ROC curve is equivalent to the precision-recall curve. The *area under the ROC curve* (AUC) [70] can be used to evaluate a probabilistic classifier and is close to the intuitive understanding of classification quality. It expresses the probability that a randomly picked churner has a higher score than a randomly picked non-churner. Thus, a random classifier has AUC of 0.5, a perfect classifier has AUC 1.0. AUC has the advantage that it is independent from the actual churn rate, in comparison to, for instance, the lift curve (plotted as *Yrate* [8] versus lift, where the Yrate is the number of false and true positives divided by the number of all customers). [8] gives a good overview of the different methods for evaluating churn prediction models and further focuses on the problem of class imbalance. This is a problem if the number of positive cases (churners) is by far below the number of negative cases in the training data. As such, it is particularly crucial for churn prediction. Intuitively, class imbalance can effect the accuracy of the prediction model significantly. [8] discusses several approaches for that problem, such as under-sampling [40] or using specific class ratios (ratio between churners and non-churners) [66]. Similar to [68], one observation is that an equal distribution between both classes is not fruitful, but that otherwise there is no optimal choice for all cases. Further, [8] concludes that advanced sampling techniques do not provide significant improvement and that advanced classification techniques, such as cost-sensitive learners [67] like weighted random forests [10] and boosting [19], should always be compared to logistic regression.

5.2 Social Network Analysis for Churn Prediction

As we discuss in Section 2, the feature-based approaches are ignoring the social relations in the underlying network. Only recently, some first works consider social network analysis as an alternative or extension to customer churn prediction models solely based on features. [13] models churn in telcom sector as a spread of influence applying the *spreading activation* [60] method. The underlying social networks are created from the call detail records of the customers. Evaluation based on the lift ratio shows that this approach is very promising, as it allows to increase the precision of prediction accuracy significantly. Similarly, [34] models churn for multi-player online games on the basis of the spreading activation

model. Churn is represented as a negative influence that propagates from one player to another. In [5], Birke describes a churn prediction model based on the underlying network structure using regression and a diffusion model similar to spreading activation. Hill et al. studied how marketing of a new telcom service was improved by applying a viral marketing process that takes the network structure into consideration [29]. This can be understood as opposite to churn as we define it in this work. All these works are still pioneering and the authors note that the applied diffusion models still require research. Moreover, the common understanding is that the feature-based approaches should probably be enriched by diffusion-based approaches, rather than being replaced. We discuss the use of diffusion models for churn prediction and the combination with feature-based approaches in more detail in the next section.

6 Network Effects and Propagation of Churn

As mentioned above, only recently [13, 34] discussed the correlation between the probability of a user to churn and the number of his neighbours that already churned. Both works proposed diffusion models to describe how churn, specifically, churn influence, is propagated between users. Using these models, they were able to increase the accuracy of traditional feature-based prediction methods. We posit that network effects are a crucial factor of churn in social networks and that they are an important component in modelling and predicting churn. To this end, we already highlighted the importance of extrinsic features in Section 2. In this section, we provide some details how basic diffusion models were used in the works mentioned above to predict churn and discuss important improvements. Afterwards, we highlight the importance of user roles and how different roles are likely to have different influences on fellow users.

6.1 Network Views

Graphs built from social networks are a natural representation of the social influence between people. Vertices represent people and edges can represent a variety of relationships. Examples include friendship links in OSNs, frequent interactions in discussion boards, or similarity in interests. In studying churn, we are interested in how churners influence other people to churn, hence, we are interested in networks that are associated with influences.

In many cases, an explicit network structure exists, e.g., a call graph in the telcom sector or friendship relations in OSNs. However, this explicit network might not be the most appropriate representation for modelling influence between users, or it might be worth analysing different network *views* of the same social network. As an example, consider OSNs, where user-defined friendship relations constitute the explicit network, but other views can be constructed, e.g., on the basis of group membership, blog following, etc. Moreover, other social networks like boards.ie do not always define an explicit network structure. Different influence network views can be built based on connections between users that refer to replies, activity in the same forum or thread, etc.

One useful network view for discussion boards like boards.ie (*Set* 1, cf. Section 4) is the user-to-user interaction network. It represents the users and the amount of pairwise interaction, in terms of number of posts, number of threads and number of forums two users are involved in. In this section, we want to measure whether there is a network effect for churning based on these interaction networks. The idea is to infer the influence of one user on another by the amount of communication between them. The more communications a user A has with another user B, the more influence user B might have on user A. If user B churns, this will more likely affect user A than some other user who has low communications with A.

In contrast, the users in the business-oriented OSN (*Set* 2, cf. Section 4) have explicit links between professional associates. Users can see the activities of associates, and in general, a professional network is only valuable to a user if he or she has links to associates of value. If these relatively important associates churn, this might cause a user to churn as well. Therefore, the explicit associate links can be used to model the spread of influence. However, other possible influence paths can be constructed via similarity in features (also known as homophily [45]). The network view here is to link individuals who are similar, such as in hobbies, employment, gender, combinations of individual features, etc.

Based on the different possible network views briefly discussed here, in the following we introduce how diffusion models can be applied on top of these network views. It is important to understand that different network views should be investigated and that different diffusion models (with different extensions and different models for influence) have to be applied.

6.2 Diffusion Models

The idea behind diffusion models for churn prediction is that some few key individuals, the churners, may influence other individuals by word-of-mouth effects to churn themselves. These, in turn, influence their neighbours (e.g., friends) in the underlying social network, and so on. Diffusion models have a long history in social sciences [4, 24] and epidemiology [36]. Since that, they have been adopted for a wide range of applications. The basic idea behind modelling churn as a diffusion process is illustrated in Figure 7. The two already influenced nodes (i.e., churners) 1 and 2 have some influence on their neighbours. Consequently, in the next time step some of their neighbours (nodes 3 and 4) are influenced and decide to churn, while some (nodes 5, 6 and 8) stay uninfluenced. This process continues over time, so that nodes 8-10 might become influenced in later time steps. Note that there are models that also allow nodes 5-7 become influenced later on, e.g., in the *spreading activation* (SPA) model the energy (i.e., influence) of an already influenced node can increase over time.

In [29], the authors explored options for improving direct-mail marketing to reach people who might be interested to subscribe to their service. Such target-based marketing can be seen as efforts to initiate activity (as opposite to churn). As resources for such direct marketing initiatives are usually limited, the idea is to identify people who are especially prone to subscribe to their service. The

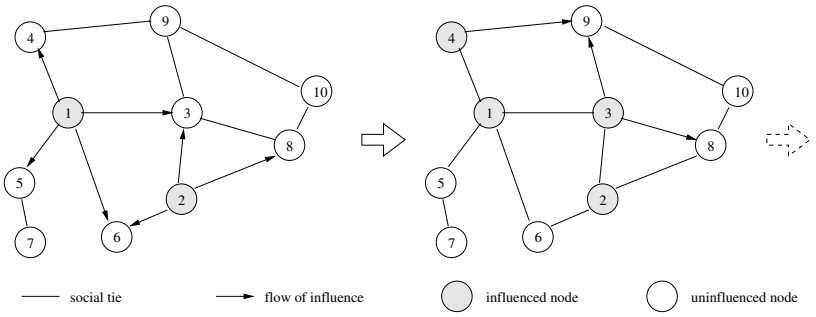


Fig. 7. Sketch of an influence diffusion process.

authors observed that this is true for people who have regular communication with the present customers. They developed a multivariate model to predict the people who would be most likely to join their customer base, including network features like degree and connected component size. Thus, this work combines the analysis of *structural* features with a first step of diffusion models – targeting those who are in direct communication with the present customers builds on the assumption that influence diffuses by word-of-mouth over direct relations. [13] applies the SPA model to predict churn based on the idea of social influence in a telcom network. The SPA model is a basic model for information diffusion. Influence is spread by an infected node to its neighbours. The spread takes place in discrete time steps and at any time step the total amount of influence (i.e., energy) over all nodes remains the same. How the influence is spread is modelled using a global spreading factor and a transfer function. The already infected nodes act as seeds from where the spread starts. If the spreading factor is low, the amount of influence transferred is less but the spread takes place quickly. If the spreading factor is high, the spread is more effective in the current neighbourhood of an infected node. Hence, it becomes slow and cannot propagate fast, as it gets "trapped" inside highly connected neighbourhoods. The transfer function depends on the relative weight of the edge between two nodes and signifies what fraction of influence should be transferred from one node to another. At each step, the amount of influence transferred by a node to its neighbour is determined by the product of the spreading factor, the output of the transfer function and the influence remaining in the node at that step. The process terminates under two conditions. Firstly, if the number of the activated nodes remains same for two consecutive steps. Secondly, if the energy transferred by a node at a step is not greater than the *accuracy threshold*. [34] also suggests churn prediction based on the SPA model, but in the area of online games. The authors extend the original model by several parameters based on the features of online gamers. They further model negative influence, include the engagement level of players and develop mechanisms for the exchange of positive and negative influences. [5] uses a model called survival analysis. This is based on the notion of how

an infected node infects its neighbours. By that, it is similar to the analysis of virus spreads and viral marketing. Similar to SPA, this is based on the idea of diffusion of information in the underlying network.

SPA can be applied easily to many applications. However, it is a rather simple model for information diffusion. Tuning spreading factor and transfer function to reflect actual influences poses a problem. Some observations that do not adhere to the intuitive understanding of spreading churn influence are:

- churners can collect more energy over time, but actually their influence should not increase
- spread of energy is mainly determined by global factors, rather than by individual relations
- the model does not differentiate between different roles and popularity of individuals, which will result in different influence effects in the network

[13] also states that there are several directions for improving their approach. This involves optimising the SPA parameters, evaluating other diffusion models, focusing on influential individuals for churn prevention, and investigating how social influences can be incorporated. Further, the authors mention the promising idea to combine their approach with decision trees on conventional and structural features. Alternatively, this could be based on link-based classification. We discuss this direction in more detail in Section 6.3. [34] comes to a similar conclusion. The authors plan to accommodate players engagement versus group engagement into influence propagation, to provide recency-frequency-money (RFM) analysis, and to apply queueing theory for prediction.

We believe that some of the well-known models are better suited to predict churn based on the influence of individuals than the introduced SPA model. In the *linear threshold model* [59] neighbours are only successfully influenced if the influence summed over all seed nodes is larger than a pre-defined threshold. The *independent cascade model* [21] is based on individual cascades between nodes and the *heat diffusion model* [42] works similar to SPA. Further, there exist several extensions of these basic models. [25] provides a broad survey and further pointers to more detailed surveys for most of the traditional models. For an improved churn prediction, these models should be combined with positive and negative influences, e.g., by combining linear thresholds with a sentiment measure. Furthermore, similarity in an user's and neighbour's features (homophily) and its effect on churn diffusion should be investigated. In addition, diffusion-based models for predicting churn should take external circumstances into account. This should include external events (e.g., a new provider enters the arena with special offers), the importance of providers' pricing politics and reputation to geographical aspects. A last aspect that up to now did not receive enough attention is to analyse the "direction" of influence, i.e., is the influence implicit or is it explicitly initiated by the churner.

As a key observation, the degree of influence and thus the process of diffusion has to be analysed with respect to the social roles and popularity of the involved individuals. This involves local popularity (influence among the individual's direct friends, this relates to the strength of social ties [24]) as well as global

popularity (with respect to the whole network, which relates to social roles in the network [9]). Intuitively, popularity and resulting influence are strongly dependent on the underlying application domain and can therefore not be defined in general. In this light, we discuss approaches for determining popularity, social roles and according influence in Section 7.

6.3 Combining Feature-Based Approaches and Diffusion Models

Although recent works successfully predicted churn on the basis of diffusion models, the long history of feature-based approaches for churn prediction shows that they are also well-suited for that task. In the literature (e.g., [1]), there is the common agreement that the key to tackle churn is to identify the point at which customers experience a change of their status, e.g., before they become dissatisfied with the service and decide to switch to an alternative provider. According data mining approaches are based on several indicators for predicting customers who are likely to churn, such as the initial activation period, the number of customer service queries, price band and the original sales channel. Such approaches were successfully used to significantly reduce churn rates in telcom industries. However, these numbers cannot be achieved for all service industries and it is difficult to apply the feature-based approach in social networks. Reasons for this are the special characteristics of churn in social networks (cf. Section 2) and the problem that many required features are often not available or not trustworthy (e.g., naming a gender on a forum site can be done by choice and with the intention to cheat).

Intuitively, one should not have to decide between either a feature-based or a diffusion-based approach. Rather, both ideas should be combined accordingly to achieve prediction accuracy and efficiency that none of the two approaches could achieve on their own. This idea is also suggested by [13]. After comparing the proposed SPA-based approach with a simple decision tree approach, the authors conclude that a combination of both is the most promising future direction. They suggest a list of features that should be considered for the telcom industry. Apparently, this is only a first list that should be modified and fine-tuned with respect to the prediction accuracy. Feature selection models, as mentioned in Section 5, can be used to automate this task. To this end, a combination of decision trees with diffusion-based techniques has been identified as most promising. [34] also indicates such a combined approach by using more features to determine player engagement, which shall be used to steer and fine-tune the underlying diffusion process.

To the best of our knowledge, up to now there is no concrete suggestion or proposal of how to exactly combine both worlds. Further research has to analyse the possibilities of using features to adapt the influence of individuals effective in the diffusion process. We propose to focus further research on an enrichment of the *conventional* features by *structural* features (connectivity, degree, triangles, path lengths, etc.). As we show in Section 7.2, such features are better suited to identify and define social roles, popularity and resulting influence. In a preliminary analysis, [13] found that decision trees combining both types of features

perform better than those restricted to conventional features. The authors come to the conclusion that SPA further helps "to learn" the important ties in a network, which cannot be achieved with feature-based approaches only. This goes along with research in the area of *link mining* as a new discipline between link analysis, Web mining, relational learning and graph mining [20]. [20] also suggests to combine the analysis of probabilistic dependencies with the analysis of link structure, resulting in *link-based classification*. Link-based classification uses conventional features of an individual, the links that the individual participates in, as well as features of individuals connected by a path in the network.

An alternative direction worth investigation is the use of an inverse approach, i.e., use information diffusion to adapt (as input for) the feature-based approach. However, this direction seems to be not as obvious. Further, we argue to include a kind of *meta features* and *external features*. Meta features are defined by the combination of several "atomic" features of individuals. That is, several features might not be well-suited as indicators for churn, but still they influence the actual process of diffusion. Classical metrics from the social network analysis, such as local clustering coefficient and betweenness measures, can be understood as meta features on top of structural features. One meta feature on top of structural as well as conventional features is popularity. As we discuss in the following section, popularity of individuals has very strong impact on their local and global influence in the network. Again, this relates to the computation of *aggregate features* as mentioned in [20]. External features relate to external events, such as a general hype due to new players or aggressive advertising, bad reviews, hacking of sites, new alternatives on the market, etc.

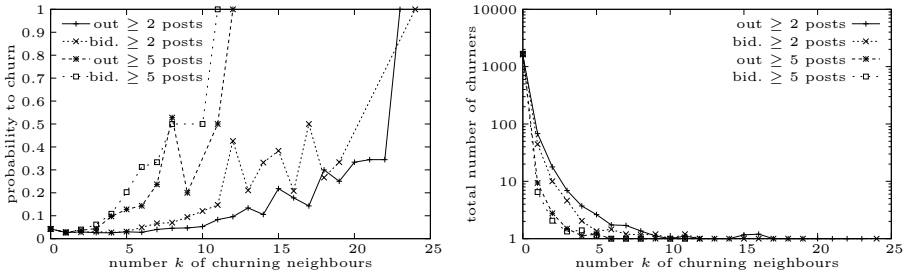
7 Popularity and Influence in Social Networks

In the previous section, we highlighted the importance of influence that individuals have on their neighbours and other individuals in the network. Intuitively, such influence can be seen in relation to the *popularity* of individuals. The more popular a user is seen by others, the more influence he or she will have on them. We have already briefly discussed the approach of using popularity to model influence in diffusion processes by the means of a meta feature. In this section, we discuss the notion of popularity, its relation to influence and information diffusion as well as approaches for analysing popularity on the basis of the two data sets introduced in Section 4.

Most service providers provide information about their users by default. Such conventional features include the registration date of the users, their usage frequency, etc. However, the actual notion of popularity and the determining features strongly depend on the kind of social network and application. We posit that in social networks popularity stronger depends on structural features (cf. number of contacts, betweenness) than on conventional features.

7.1 Social Roles and Influence in Discussion Boards

In this section, we discuss the roles users play in discussion boards and how these can affect churn. In previous work [9], the authors grouped the common features of users in boards.ie (*Set* 1 from Section 4), and proposed 8 different social roles played by users in boards.ie. These were: joining conversationalists, popular initiators, taciturns, supporters, elitists, popular participates, grunts, and ignored. The roles were determined on the basis of conventional and structural features, where the underlying network was constructed from the reply-to structure (nodes represent users, an edge represents one user replying to another in a forum thread).



(a) Probability to churn (b) Number of users

Fig. 8. Network effects on the probability to churn

We posit that different roles have different amount of influence on other roles. For example, users playing the elitist role might have strong influence on other users playing the same role, but users in all other roles will have little or no influence on the users of the elitist role. Regarding that the roles are determined based on the reply-to structure, the notion of popularity is intuitively an important factor. The more popular a user is, the more post views and replies he or she gets from other users, therefore exerting more influence on him or her fellow users. We indicate the relation between popularity and influence in Figure 8. Similar to preliminary experiments in [13, 34], Figure 8(a) shows an individual's probability to churn in relation to the number k of the neighbours that already churned. This probability is determined by dividing the number of all current churners that have k churned neighbours by the total number of users that have k churned neighbours. In other words, it is the percentage of all people with k churning neighbours who decided to churn as well. The figure shows this relation for four different reply graphs. The number of replies constitutes the weight of an edge. We averaged the values over all forums and over all analysed time slots. The labels with `out` refer to the graph based on only outgoing edges, `bid.` refers to the graph containing only bidirectional edges (thus, it is sparser). To highlight the effects of popularity on influence, we filtered edges by their weight:

$\geq$ 2 contains only edges with a weight larger than 1, $\geq$ 5 contains only edges with weight larger than 4.

As expected, the churn probability increases with k, which illustrates the existence of network effects (i.e., the more churning neighbours there are, the more likely a user is to churn also). The fact that this probability is 1.0 for some k is because only a few users have that many churning neighbours, and all these users churned as well. To illustrate this, we show the total number of churners with k churned neighbours in Figure 8(b). Note that the y axis is shown in log scale. Furthermore, comparing the outgoing edge results for $\geq$ 2 and $\geq$ 5 posts in Figure 8(a) shows that the amount of influence is apparently increasing with the number of replies between users. Intuitively, an increased outgoing edge weight can be interpreted as increased popularity. This means, the more popular the neighbours are, the more likely a user will churn. It demonstrates the validity of determining influence based on popularity, where popularity is determined based on structural features of the graphs. The results provide only first insights into the relation between popularity, influence and diffusion, but highlight the appropriateness of further research in that direction.

In the literature, there are several approaches to measure influence. The popular Pagerank [54] and HITS algorithms [38] are two examples. In these approaches, influential Web pages (or people) are those that receive many in-links from reputable Web pages (or other people) and have fewer out-links. These approaches work well for measuring absolute influence and reputation. However, there are better approaches for measuring relative influence, which should be assumed in the case of roles and churn. One value or measure of influence is hence not appropriate. We propose that we need to examine conditional influence - conditional on the users and possibly features. There has been some recent work in this direction for determining different influence measures for different topics [64] and movies [23], which we plan to explore for determining relative influence of roles.

The crucial point for an appropriate application of diffusion models for churn prediction will be the inclusion of social roles, popularity and the according individual (conditional) influence of users into the used diffusion model. This is not possible in the basic version of the SPA model introduced in Section 6.2. The main questions that have to be answered are:

1. How should popularity be defined and determined in different social networks?
2. Is the influence of an individual a function of his or her popularity?

In the next section, we briefly investigate an intuitive notion for popularity in OSN and the feasibility of different features to determine it.

7.2 Popularity in Online Social Networks

Due to the limitations for *Set* 2 from Section 4 (explained in that section), and the fact that the relation can be expected to be similar (see also [13, 34]), we

omit an analysis of the churn probability in this section. However, the gathered data allows to indicate the importance of an analysis of popularity based on structural features. In the following, we highlight that they are more important than conventional features for this task. To this end, this section indicates one direction of how to combine feature-based churn prediction with diffusion-based approaches.

Social networking services contain a wealth of information. Users voluntarily feed self-descriptive details into the OSN and their utilisation behaviour is completely observable by the OSN provider. Analysing this complex data facilitates understanding of the psychological and sociological properties of online social networks and their users. Identifying key properties of users and their profiles, which allow for the prediction of their popularity, however, is not only interesting for social scientists. Especially system designers and developers of social networking services may capitalise on the extracted knowledge. In case of centralised, server-based systems, the user experience may be enhanced by decreasing service delays of frequently requested profiles. They may prove even more valuable for the development of decentralised designs (cmp. Chapter 17: "Decentralized Online Social Networks"), i.e., for the parametrisation of this entirely different approach to provide social networking services. Being able to predict the popularity may help suggesting necessary availability levels for selected profiles, and they may indicate which profiles can more frequently tolerate temporary inavailabilities without causing a significant deterioration of the service experience. System designers additionally may be able to derive timing constraints for different profiles, and hence be able to decide which profiles need to be presented with very low delay, and be able to identify properties of profiles that may tolerate higher response times.

The same as mentioned for the activity of users (Section 4.2) holds for the popularity of their profiles. Messages sent to the wall, commenting photos or posts left in guest books indicate at least one profile impression. More detailed information is given by explicit counters of the number of requests to a profile. It consequently allows for the identification of relations between characteristics of profiles and the frequency with which they are viewed. Thus, in the following we refer to the popularity of users by the frequency of profile impressions, i.e., the number of visits to their profile by other users of the OSN.

Features for Determining Popularity Access to profiles and the full user behaviour for scientific purposes is rather difficult, since, for obvious commercial and privacy reasons, OSN providers are keen on their protection and highly unwilling to disclose them. Some information, however, can be deduced by directly accessing the publicly available profiles through the Web interface of the social networking service. The main purpose of participating in an OSN being to publish information about oneself, the users willingly share quite detailed information on themselves, usually including the list of other users they have contacted. The information that usually is contributed by the users can be grouped into a few categories:

– identifying data of the person
– contact details
– personal interests
– work experience
– curriculum vitae (affiliations, educational information)
– recommendations (or direct comments and messages from other users, left at the wall or guest book)
– the list of friends and contacts

The number of impressions of a profile is usually integrated into the presented profile automatically, without any participation of the user (cmp. Chapter 21: "Security and Privacy Threats in Current Online Social Networks" for a concise list of the data that frequently is published in OSN profiles). The gender, though usually not explicitly stated, can easily be derived from the name, even in an automated fashion.

Relations between Features and Popularity A previous analysis [61] reveals different relations between selected features and the popularity of profiles, based on data from a large set of monitored profiles. It studies in detail which influence the gender and inclusion of a profile photo have on the popularity and additionally correlates the popularity of profiles with the number of contacts, the number of joined interest groups inside the OSN, and the average activity of the users. The study is divided into two parts. The first part uses Mann-Whitney U tests to analyse non-parametric relationships between classes of users and their corresponding profile popularity. Potential correlations between parametric features of the users and their corresponding profiles are analysed in the second part, using Pearson's correlation coefficient r. In the following, we summarise the main observations of this study, which are collected in Table 2.

Similarity and difference of means of *Popularity* π_{v_i} (rank sum)			
Gender	Male $Mdn = 0.039$	Female $Mdn = 0.041$	*no sign. difference*
Profile Picture	Picture ($Mdn = 0.043$) > No picture ($Mdn = 0.016$) ($p < 0.001$)		
Correlational observations (Pearson's r)			
Accepted contacts	correlation: $r \approx 0.75$ ($p < 0.001$)		
	Male	$0.74 < r < 0.75$	($p < 0.05$)
	Female	$0.81 < r < 0.83$	($p < 0.05$)
Profile alterations	correlation: $r \approx 0.62$ ($p < 0.001$)		
Subscribed groups	correlation: $r \approx 0.37$ ($p < 0.001$)		
	Male	$0.37 < r < 0.4$	($p < 0.05$)
	Female	$0.33 < r < 0.37$	($p < 0.05$)
Average activity	correlation: $r \approx 0.17$ ($p < 0.001$)		
	Male (0.18)	Female (0.16)	*no sign. difference*

Table 2. Summary of the popularity experiment

A first hypothesis stated that the gender of users would have impact on the popularity of their profiles. Profiles of male ($Mdn = 0.039$) and female ($Mdn = 0.041$) users did not experience significantly different popularity (ns). The impact

of providing a profile photo has been analysed in the next step. The results supported the hypothesis that profiles including a photo ($Mdn = 0.043$) were significantly more often requested than profiles without a photo ($Mdn = 0.016$, $p < 0.001$, estimated distance = 0.024, or one profile impression every second day). Significance and estimated distance where almost identical when testing the subsamples of male and female profiles, including vs. excluding profile pictures.

A correlational study has been performed in the second part. First, the study examined if the membership duration has an impact on the popularity of the member's profile. With preferential attachment in mind, the correlation between membership duration and the number of contacts was estimated. Testing H_0 that the true correlation between membership duration and the number of contacts was 0 failed ($p < 0.001$), indicating that a correlation indeed exists. Pearson's r was estimated to $r \approx 0.19$, a very slight correlation, with $0.18 < p < 0.20$ to a confidence of 95%.

Since this factor does not seem to have a high influence on the profile popularity, it was further investigated if correlations between popularity and selected features of the profiles could be detected. Investigated were the hypotheses that the popularity of profiles increased in correlation with the activity of the corresponding user, with its degree, and with the number of subscribed interest groups.

The first hypothesis was that profiles of highly active users will experience a generally higher popularity than the profiles of inactive users. Testing the hypothesis that the correlation between activity and popularity is 0 fails ($p < 0.001$). Pearson's product moment correlation is determined to be 0.17, again indicating a very slight correlation between activity of the users and the popularity of their profiles. This correlation is slightly, but not significantly, higher for male users ($0.16 < r < 0.19$ with $p < 0.05$) than for female users ($0.14 < r < 0.18$ with $p < 0.05$). The activity tested in this case was the measure given by the social network provider, which was not very fine-grained and reliable (cf. Section 4.2). In contrast, using the alteration frequency of profiles (the frequency with which user-maintained details of the profiles had been changed) yielded surprising results. Again, Pearson's test was significant ($p < 0.001$). The product moment correlation between the alteration frequency of a profile and its popularity was estimated to a high 0.62 ($0.61 < r < 0.63$ with $p < 0.025$). This correlation may be caused by the fact that the last five changes to profiles of contacts are presented to the users after logging into the OSN. A constant profile alteration consequently increases the chances of a user to be visible with his contacts.

Modifying the hypothesis, it was studied if a correlation between the degree of a profile (the number of accepted contacts) and its popularity could be found. The hypothesis that the correlation between degree and popularity is 0 fails, too. Pearson's r indicates a quite high correlation between the two variables of $r = 0.75$ ($p < 0.001$). This correlation is significantly higher for profiles of female users ($0.81 < r < 0.83, p < 0.05$) compared to male users ($0.74 < r < 0.75, p < 0.05$). A final correlational study based on the features of the profiles has been

performed to determine the relation between the number of subscribed groups and the popularity of the subscribed users. The results refuse a correlation of 0 between groups and popularity and suggest a correlation of $r = 0.37$. In this case, the correlation is significantly higher for male users ($0.37 < r < 0.4, p < 0.05$) compared to female users ($0.33 < r < 0.37, p < 0.05$).

Roughly said, intuitive beliefs about profile popularity are abundant. Unverified rumours frequently state that the unfortunate possession of a last name late in the alphabet will inevitably condemn a user's profile to eternal lack of popularity. The simple rationale behind this conjecture is that profiles are usually listed in increasing alphabetical order of the last names. Users in the beginning of that list might plainly enjoy a higher visibility due to the fact that other users start browsing contact lists, which usually contain only ten contacts per presented page. If they do not pursue to the later pages, they hence do not reach users from the end of the alphabetical list. A last analysis reflects this intuitive question of the relation between a user's name and the profile popularity. The tested correlation is very slightly, but not significantly, if determined for the whole group of users ($r = -0.01, ns$). Analysing the *rich club* of profiles with the highest popularity, in contrast, leads to an impressive change of results. Considering the 5% profiles with the highest popularity already yields a correlation of $r \approx -0.09$ ($ns, -0.26 < r < 0.08$ with $p < 0.025$). This correlation gets more significant with increasing "exclusivity" of the rich club: Analysing the top 2‰ of profiles the correlation increases to $r \approx -0.22$ ($ns, -0.47 < r < 0.06$ with $p < 0.025$), for the top 1‰ of profiles to $r \approx -0.29$ ($ns, -0.62 < r < 0.11$ with $p < 0.025$), and for the top 10 users, it finally increases to $r \approx -0.9, p < 0.001$ ($-0.98 < r < -0.61$ with $p < 0.025$). Considering the very small samples size, this result of course has to be taken with a grain of salt.

Discussion Considering the results of the study, certain features can actually indicate the expected popularity of profiles in OSNs. They strongly support correlations between the activity of users, their participation in interest groups, and most importantly the number of accepted contacts of a profile with its popularity. Additionally, they suggest that profiles with pictures will be more frequently viewed than profiles without, while no difference between the profiles and male vs. female users could be determined. As expected, this highlights the importance of combining the right choice of conventional features with structural features in order to meaningfully determine popularity. Next steps should be to analyse different OSN and other types of social networks, and to combine an appropriate classification technique (such as decision trees) with the proposed diffusion-based churn prediction.

8 Summary and Conclusion

This chapter dealt with churn in social networks. Research on churn currently enjoys great popularity, since churning customers cause effective loss of revenues at service providers and similarly affected companies. But, the focus usually lies

on contractual services like in the telcom sector. The notion of churn, the factors driving it, such as the social costs compared to purely monetary reasons, and the risks in social networks are not well understood nor researched. This chapter aims at filling this gap by reflecting on the specifics of churn in social networks. First, we discussed different notions, reasons and facets of churn. In this context, and as an extension to the traditional understanding, churn was defined as an individual's act of significantly decreasing activity in a social network. The ability to estimate churn behaviour could enable stake-holders to react early, trying to change the potential churner's mind, and churn prediction consequently promises to help companies avert potentially decreasing income. Moreover, in social networks churn is also relating to the health of the underlying communities and its prediction is therefore mandatory for successful (self-)governance.

In this chapter, we have further given an overview of the current research on churn detection and interpretation, including discussions on appropriate definitions of churn. After providing our own suggestion for defining churn, we showed its effectiveness and crucial parameters on the basis of two example social networks. The first data set comprises the activity measures of users in discussion boards (over ten years of data from http://www.boards.ie, from which we took the data for 2006). The second data set contains a random set of over 30.000 user profiles from a predominant central European online social network for professional purposes, which we have gathered using a publicly available interface.

Existing approaches for churn prediction have been introduced. With a look on novel approaches and the intuition of the importance of the underlying social network structure, we discussed required extensions of the traditional approaches by techniques from social network analysis. One focus on this is the introduction of diffusion models, which has also been pointed out in other recent work. Finally, we have presented results of first studies on popularity and churn prediction. We propose that popularity should be regarded as a main factor for modelling the differences in social influence, which is mandatory for a successful application of information diffusion for churn prediction.

The results presented here show that the proposed churn definition is appropriate for identifying a wide range of different churn types that are special for social networks. This enables to detect and predict churn in social networks for a wide range of application domains and potential risks that are of interest. Further, it highlights the crucial factors of understanding and handling churn. We showed that there is a relationship between churn, network effects and influence. Based on the literature review, we proposed to combine traditional feature-based churn prediction with diffusion models. We highlighted possible directions and open issues for this novel field of research in social networks. Finally, we assessed the relationship between popularity, determining the degree of social influence, and features that determine popularity. The results emphasise the need for focusing on structural features and relations as an extension to mining conventional features. Open issues and future research directions resulting from the gained insights and discussions have been highlighted throughout the chapter.

Acknowldegements

This work was carried out in part in the CLIQUE Strategic Research Cluster, which is funded by Science Foundation Ireland (SFI) under grant number 08/SRC/I1407, and under partial funding of ETRI and DFG FOR 733 ("QuaP2P").

References

1. J.-H. Ahn, S.-P. Han, and Y.-S. Lee. Customer churn analysis: Churn determinants and mediation effects of partial defection in the korean mobile telecommunications service industry. *Telecommunications Policy*, 30(10-11):552 – 568, 2006.
2. W.-H. Au, K. C. C. Chan, and X. Yao. A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Trans. Evolutionary Computation*, 7(6):532–545, 2003.
3. A. O. B. Briscoe and B. Tilly. Metcalfe's law is wrong, July 2006.
4. F. M. Bass. A Dynamic Model of Market Share and Sales Behavior. In *Winter Conference American Marketing Association*, 1963.
5. D. Birke. Diffusion on networks: Modelling the spread of innovations and customer churn over social networks. In *GI Jahrestagung (2)*, pages 480–488, 2006.
6. W. Buckinx and D. V. den Poel. Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 164(1):252–268, 2005.
7. J. Burez and D. V. den Poel. CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Syst. Appl*, 32(2):277–288, 2007.
8. J. Burez and D. V. den Poel. Handling class imbalance in customer churn prediction. *Expert Syst. Appl*, 36(3):4626–4636, 2009.
9. J. Chan, E. M. Daly, and C. Hayes. Decomposing discussion forums and boards using user roles. In *AAAI Conference on Weblogs and Social Media*, pages 215–218, 2010.
10. C. Chen, A. Liaw, and L. Breiman. Using Random Forest to Learn Imbalanced Data. Technical report, University of California at Berkley, 2004.
11. D. Constant, L. Sproull, and S. Kiesler. The kindness of strangers: The usefulness of electronic weak ties for technical advice. *Organization Science*, 7(2):119–135, 1996.
12. K. Coussement and D. V. 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.
13. 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*, pages 668–677, 2008.
14. P. Datta, B. M. Masand, D. R. Mani, and B. Li. Automated cellular modeling and prediction on a large scale. *Artif. Intell. Rev.*, 14(6):485–502, 2000.
15. J. Davis and M. Goadrich. The relationship between Precision-Recall and ROC curves. In *ICML '06*, pages 233–240, 2006.
16. D. V. den Poel and B. Larivière. Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, 157(1):196–217, 2004.

17. J. Ferreira, M. B. R. Vellasco, M. A. C. Pacheco, and C. R. H. Barbosa. Data mining techniques on the evaluation of wireless churn. In *ESANN*, pages 483–488, 2004.
18. Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. *J. Mach. Learn. Res.*, 4:933–969, 2003.
19. J. H. Friedman. Stochastic gradient boosting. *Comput. Stat. Data Anal.*, 38(4):367–378, 2002.
20. L. Getoor. Link mining: a new data mining challenge. *SIGKDD Explor. Newsl.*, 5(1):84–89, 2003.
21. J. Goldenberg, B. Libai, and E. Muller. Using Complex Systems Analysis to Advance Marketing Theory Development: Modeling Heterogeneity Effects on New Product Growth Through Stochastic Cellular Automata. In *Academy of Marketing Science Review*, 2001.
22. S. Gopal, R.K. Meher. Customer churn time prediction in mobile telecommunication industry using ordinal regression. *LECTURE NOTES IN COMPUTER SCIENCE*, 884-889(5012):252–268, 2008.
23. A. Goyal, F. Bonchi, and L. V. Lakshmanan. Discovering leaders from community actions. In *CIKM '08*, pages 499–508, New York, NY, USA, 2008. ACM.
24. M. Granovetter. The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.
25. D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *WWW '04*, pages 491–501, 2004.
26. A. Gustafsson, M. Johnson, and I. Roos. The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention. *Journal of Marketing*, 69(4):210–218, 2005.
27. J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *SIGIR '99*, pages 230–237, New York, NY, USA, 1999. ACM.
28. O. Herrera and T. Znati. Modeling churn in P2P networks. In *Annual Simulation Symposium*, pages 33–40. IEEE Computer Society, 2007.
29. S. Hill, F. Provost, and C. Volinsky. Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science*, 22:2006, 2006.
30. B. Huang, B. Buckley, and T. M. Kechadi. Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications. *Expert Syst. Appl.*, 37(5):3638–3646, 2010.
31. B. Q. Huang, M. T. Kechadi, and B. Buckley. Customer churn prediction for broadband internet services. In *DaWaK*, volume 5691 of *Lecture Notes in Computer Science*, pages 229–243, 2009.
32. S.-Y. Hung, D. C. Yen, and H.-Y. Wang. Applying data mining to telecom churn management. *Expert Syst. Appl*, 31(3):515–524, 2006.
33. H. Hwang, T. Jung, and E. Suh. An ltv model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert Syst. Appl.*, 26(2):181–188, 2004.
34. J. Kawale, A. Pal, and J. Srivastava. Churn Prediction in MMORPGs: A Social Influence Based Approach. In *CSE '09*, pages 423–428, 2009.
35. S. M. Keaveney. Customer switching behavior in service industries: An exploratory study. *The Journal of Marketing*, 59(2):71–82, 1995.
36. W. O. Kermack and A. G. McKendrick. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, 1927.

37. M. Kim, M. Park, and D. Jeong. The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. *Telecommunications Policy*, 28(2):145–160, 2004.
38. J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, 1999.
39. P. Kollock. *The Economies of Online Cooperation: Gifts and Public Goods in Cyberspace*. Routledge, London, 1999.
40. C. X. Ling and C. Li. Data Mining for Direct Marketing: Problems and Solutions. In *KDD*, pages 73–79, 1998.
41. J. Lu. Predicting customer churn in the telecommunications industry – an application of survival analysis modeling using sas. In *SAS Proceedings, SUGI 27*, pages 114–127, 2002.
42. H. Ma, H. Yang, M. R. Lyu, and I. King. Mining social networks using heat diffusion processes for marketing candidates selection. In *CIKM '08*, pages 233–242, 2008.
43. J. P. Maicas, Y. Polo, and F. J. Sese. Reducing the level of switching costs in mobile communications: The case of mobile number portability. *Telecommunications Policy*, 33(9):544 – 554, 2009.
44. B. M. Masand, P. Datta, D. R. Mani, and B. Li. CHAMP: A prototype for automated cellular churn prediction. *Data Min. Knowl. Discov*, 3(2):219–225, 1999.
45. M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27:415–444, 2001.
46. B. Metcalfe. Metcalfe's law: A network becomes more valuable as it reaches more users, 1995.
47. A. Meyer-Bäse and R. Watzel. Transformation radial basis neural network for relevant feature selection. *Pattern Recognition Letters*, 19(14):1301–1306, 1998.
48. K. Morik and H. Köpcke. Analysing customer churn in insurance data - a case study. In *PKDD '04*, pages 325–336, 2004.
49. M. Mozer, R. Wolniewicz, D. Grimes, E. Johnson, and H. Kaushansky. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on Neural Networks*, 11(3):690–696, May 2000.
50. M. Mozer, R. H. Wolniewicz, D. B. Grimes, E. Johnson, and H. Kaushansky. Churn reduction in the wireless industry. In *NIPS*, pages 935–941, 1999.
51. T. Mutanen. Customer churn analysis - a case study. Technical report, Helsinki University of Technology, System Analysis Laboratory, 2006.
52. K. Ng and H. Liu. Customer retention via data mining. *Artif. Intell. Rev.*, 14(6):569–590, 2000.
53. G. Nie, G. Wang, P. Zhang, Y. Tian, and Y. Shi. Finding the hidden pattern of credit card holder's churn: A case of china. In *ICCS '09*, pages 561–569, 2009.
54. L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford, 1998.
55. Plc-Vodafone-Group. Vodafone annual report for the year ended 31 march 2009: Principal risk factors and uncertainties, 2009. http://www.vodafone.com/annual_report09/downloads/VF_Annual_Report_2009.pdf.
56. A. Prinzie and D. V. den Poel. Investigating purchasing-sequence patterns for financial services using markov, mtd and mtdg models. *European Journal of Operational Research*, 170(3):710–734, 2006.
57. J. Qi, Y. Zhang, and H. Shu. Churn prediction with limited information in fixed-line telecommunication. In *Symposium on Communication Systems Networks and Digital Signal Processing*, pages 423–426, 2006.
58. H. Rheingold. *The Virtual Community: Homesteading on the Electronic Frontier*. MIT Press, Cambridge, MA, USA, 2000.

59. B. Ryan and N. C. Gross. The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociology*, 8(1):15–24, 1943.
60. G. Salton and C. Buckley. On the use of spreading activation methods in automatic information. In *SIGIR '88*, pages 147–160, 1988.
61. T. Strufe. Profile Popularity in a Business-oriented Online Social Network. In *Social Network Systems, EuroSys*, 2010.
62. D. Stutzbach and R. Rejaie. Understanding churn in peer-to-peer networks. In *ACM SIGCOMM conference on Internet measurement*, page 202, 2006.
63. Z. Sun, G. Bebis, and R. Miller. Object detection using feature subset selection. *Pattern Recognition*, 37:2165–2176, 2004.
64. J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *SIGKDD '09*, pages 807–816, 2009.
65. C.-F. Tsai and Y.-H. Lu. Customer churn prediction by hybrid neural networks. *Expert Syst. Appl.*, 36(10):12547–12553, 2009.
66. C.-P. Wei and I.-T. Chiu. Turning telecommunications call details to churn prediction: a data mining approach. *Expert Syst. Appl*, 23(2):103–112, 2002.
67. G. M. Weiss. Mining with rarity: a unifying framework. *SIGKDD Explor. Newsl.*, 6(1):7–19, 2004.
68. G. M. Weiss and F. Provost. Learning when training data are costly: the effect of class distribution on tree induction. *J. Artif. Int. Res.*, 19(1):315–354, 2003.
69. Y. Xie, X. Li, E. W. T. Ngai, and W. Ying. Customer churn prediction using improved balanced random forests. *Expert Syst. Appl.*, 36(3):5445–5449, 2009.
70. L. Yan, R. H. Wolniewicz, and R. Dodier. Predicting customer behavior in telecommunications. *IEEE Intelligent Systems*, 19(2):50–58, 2004.