Discussion Paper

Predicting Users' Future Level of Communication Activity in Online Social Networks: A First Step towards More Advertising Effectiveness

by

Florian Probst

Europäische Union
„Investition in Ihre Zukunft"
Europäischer Fonds für
regionale Entwicklung

in: Proceedings of the 17th Americas Conference on Information Systems, AMCIS.
Paper 39

Predicting Users' Future Level of Communication Activity in Online Social Networks: A First Step towards More Advertising Effectiveness

Florian Probst
FIM Research Center,
Department of Information Systems
Engineering & Financial Management,
University of Augsburg
Universitaetsstr. 12, Augsburg, Germany
florian.probst@wiwi.uni-augsburg.de

ABSTRACT

Many Online Social Networks do not generate sustainable revenues through advertising, even though active usage has reached enormous scales. To enable more effective advertising strategies in Online Social Networks, it is essential to identify users who can affect a large number of friends, acquaintances, or other users in the network. In this context, especially users' future level of communication activity in the Online Social Network plays an important role. A highly active past, however, does not guarantee high levels of future communication activity. Thus, approaches for the prediction of users' future level of communication activity are needed. Therefore, we transfer a probability-based method that has been primarily developed to forecast purchasing behavior of customers to the context of users' communication activity in Online Social Networks. In addition, we demonstrate the method's applicability and suitability by using a publicly available dataset of Facebook.com.

Keywords

Online Social Networks, advertising effectiveness, communication activity, probability-based method, forecasting.

INTRODUCTION

In the last couple of years, Online Social Networks (OSN) have become popular Internet platforms that connect people around the globe. Thereby, the active usage of OSN has reached enormous scales: In March 2010, the OSN Facebook.com surpassed Google.com to become the most visited website of the week in the US (Dougherty, 2010), while a few months later, the number of active users of Facebook.com exceeded 500 million (Facebook, 2010). According to a recent study, two-thirds of the US Internet users already visit OSN each month, with 43% of them using OSN more than once a day (Alison, 2010). Thus, the phenomenon OSN has evolved into a global mainstream medium that generates an increasing social and economic impact.

However, many OSN face the question of how to leverage on their fast growing popularity to achieve sustainable revenues (Heidemann, Klier and Probst, 2010). Nowadays, the majority of OSN relies on an advertisement-based business model (Gnyawali, Fan and Penner, 2010). Nevertheless, many OSN do not generate sustainable revenues through advertising: Even though the worldwide advertisement spending on OSN is expected to grow from US$ 2.0 billion in 2008 to US$ 3.5 billion in 2013 (Williamson, 2009), OSN often do not know how to unleash this potential (Clemons, 2009; Lu and Hsiao, 2010). Therefore, more effective advertising strategies in OSN are needed in order to remain financially viable (Wen, Tan and Chang, 2009; Xu, Zhang, Xue and Yeo, 2008).

To tap the enormous potential originated by the dramatic increase in the popularity of OSN, particularly the identification of users who can affect a large number of friends, acquaintances, or other users in an OSN is essential (Heidemann et al., 2010; Hill, Provost and Volinsky, 2006). Such users can for example be addressed in marketing campaigns to achieve a high awareness of a product or service. This strategy is often referred to as network-based marketing, word-of-mouth marketing, or viral marketing (cf. e.g. Brown, Broderick and Lee, 2007; Hill et al., 2006; Lee, Lee and Lee, 2009). The underlying key assumption is that users propagate "positive" information about a product or service after they have either been made aware by traditional marketing techniques or experienced it by themselves (Hill et al., 2006). In this context, Ray, Bernoff and Wise

(2010) found "that people in the US generate more than 500 billion online impressions on each other regarding products and services". However, as only "16% of online consumers generate 80% of these impressions" (Ray et al., 2010), only a subset of users is particularly valuable for marketers (Trusov, Bodapati and Bucklin, 2010). For the identification of these users, especially users' future level of communication activity, i.e. each user's number of future communication activities, plays an important role (Cheung and Lee, 2010; Hoffman and Fodor, 2010; Willinger, Rejaie, Torkjazi, Valafar and Maggioni, 2010; Xu et al., 2008). But even though some users might have been highly active in the past, high levels of future communication activity cannot be taken for granted (Cummings, Butler and Kraut, 2002; Viswanath, Mislove, Cha and Gummadi, 2009). Hence, approaches for the prediction of users' future level of communication activity are needed, which might serve as a first step towards more effective advertising strategies in OSN. However, to the best of our knowledge, approaches for predicting users' future level of communication activity in OSN are missing. Therefore, we transfer a probability-based method, which has been primarily developed by Fader, Hardie and Lee (2005), to forecast purchasing behavior of customers, to the context of users' communication activity in OSN. In addition, we demonstrate the practical applicability of the method and evaluate its suitability for predicting users' future level of communication activity in OSN by using a dataset of Facebook.com.

After the discussion of the general relevance of the problem and its motivation within this introduction, we specify the problem context and review prior research on users' communication activity in OSN. Thereby, we identify our research gap. Afterwards, we propose our artifact as a probability-based method. In the penultimate section, we illustrate the method's practical applicability and suitability to predict users' future level of communication activity in OSN. Finally, we conclude with a summary of the results and an outlook on future steps.

PROBLEM CONTEXT AND RELATED WORK

OSN are a particular type of virtual communities[1] (Dwyer, Hiltz and Passerini, 2007). According to Boyd and Ellison (2007, p. 211) we define OSN as "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system". Current OSN are primarily used for maintaining existing relationships from an offline context, although they also allow for creating pure online relationships (Ellison, Steinfield and Lampe, 2007; Lampe, Ellison and Steinfield, 2006). Especially the visibility and searchability of the users' relationships is a distinctive feature of OSN. Thus, OSN can "create substantial value for the individuals who participate in them, the organizations that sponsor them, and the larger society in multiple ways" (Agarwal, Gupta and Kraut, 2008, p. 243). The majority of OSN that rely on the advertisement-based business model, however, face the challenge to tap the enormous potential originated by the dramatic increase in the popularity of OSN in order to generate sustainable revenues (Clemons, 2009; Lu and Hsiao, 2010). Therefore, more effective advertising strategies are needed (Wen et al., 2009; Xu et al., 2008). Word-of-mouth marketing, for example, "is known to be the most effective form of advertising, but, until recently, was the most expensive" (Jacks and Salam, 2009, p. 2). In contrast, word-of-mouth and targeted marketing can be much more efficient and cost-effective in OSN (Trusov, Bucklin and Pauwels, 2009; Zhang, Wang and Xia, 2010). However, only a small subset of users has actually significant influence on other users (Trusov et al., 2010). Thus, the identification of the most influential users is necessary to enable more effective advertising strategies in OSN (Heidemann et al., 2010; Trusov et al., 2010; Zhang et al., 2010). In this context, literature indicates that particularly the identification of users with high levels of future communication activity is essential (Cheung and Lee, 2010; Hoffman and Fodor, 2010; Willinger et al., 2009; Xu et al., 2008).

Communication activity in OSN can be any sort of interaction among users facilitated by methods provided by OSN, such as messages or wallposts (cf. Wilson, Boe, Sala, Puttaswamy and Zhao, 2009). Prior work emphasizes the importance of users' communication activity: "No matter what resources are available within a structure, without communication activity those resources will remain dormant, and no benefits will be provided for individuals" (Butler, 2001, p. 350). Recent work supports that the value of OSN lies in the communication activity between users (Krasnova, Hildebrand and Günther, 2009; Willinger et al., 2009). Latest studies further show that user's communication activity is highly relevant for advertising effectiveness in OSN (e.g. Cheung and Lee, 2010; Ganley and Lampe, 2009). The record of communication activities between users in OSN can be used to identify users' with high levels of communication activity in the past (Xiang, Neville and Rogati, 2010). However, prior research found that users who have been highly active in the past are not necessarily highly active in the future (Cummings et al., 2002; Viswanath et al., 2009). Hence, approaches for the prediction of each user's future level of communication activity need to "abandon the traditional treatment of OSNs as static networks" (Willinger et al., 2010, p. 49) and incorporate the dynamic of communication activity in OSN.

[1] A definition of virtual communities can be found in Leimeister, Sidiras and Krcmar (2004). While Dwyer et al. (2007) and Boyd and Ellison (2007) use the term social networking site, we are using the term OSN throughout the paper synonymously.

Plenty of research addressing the dynamic nature of OSN can be found with respect to network structures in OSN. Thereby, previous studies focus on the evolution of network structures in general (for an overview cf. Dorogovtsev and Mendes, 2003) and the establishment of static social links, i.e. friendship relationships, between users in particular (e.g. Liben-Nowell and Kleinberg, 2007). However, Xu et al. (2008, p. 14) emphasize "that interaction information is invaluable to marketers, more important than the static links". Consequently, some studies take into account the dynamic nature of OSN and users' communication activity: De Choudhury, Sundaram, John and Seligmann (2007), for example, determine the intent to communicate and the communication delay between users based on several contextual factors in OSN, such as the relevance of a topic. Therefore, it is first assumed that a user receives a message. Second, the likelihood that the receiver will communicate with the sender on a particular topic and the delay in communication are predicted. However, to the best of our knowledge existing approaches do not allow for determining conditional expectations about users' future level of communication activity in OSN on an individual-level, i.e. making predictions about each user's future level of communication activity given information about his or her past communication activity. Thus, approaches for predicting users' future level of communication activity in OSN are missing.

METHOD

For predicting users' future level of communication activity, we draw on a probability-based method that has been primarily developed by Fader et al. (2005) to address a very similar problem, i.e. to forecast purchasing behavior of customers. This beta-geometric/negative binomial distribution (BG/NBD)-based method goes back to the highly regarded "counting your customers" framework introduced by Schmittlein, Morrison and Colombo (1987). In the following, we discuss the BG/NBD-based method as possibility to predict users' future level of communication activity in OSN.

First, it is assumed that a user $i \in \{1,2,...,N\}$ is active at $t_{0_i}=0$. This will generally be satisfied if we take t_{0_i} as the point of time at which the user's initial communication activity occurred. Furthermore, the BG/NBD-based method requires three pieces of information about each user, represented by $(X_i=x_i, t_{x_i}, T_i)$. Thereby, $x_i \in IN$ denotes the "frequency", i.e. the number of communication activities after the initial communication activity within the observation period $(0,T_i]$, $t_{x_i} \in IR^+$ (with $0 \leq t_{x_i} \leq T_i$) is the "recency", i.e. the point of time of the last communication activity (if $x_i=0$ then $t_{x_i}=0$), and $T_i \in IR^+$ represents the length of the observation period, i.e. the time between the initial communication activity and the end of the observation period. Notice, that the relation between the period $(0,T_i]$ and calendar time will vary from user to user depending on when the user's initial communication activity occurred. Based on this information, we aim to predict each user's future level of communication activity during the forecasting period of length t_i (cf. Figure 1).

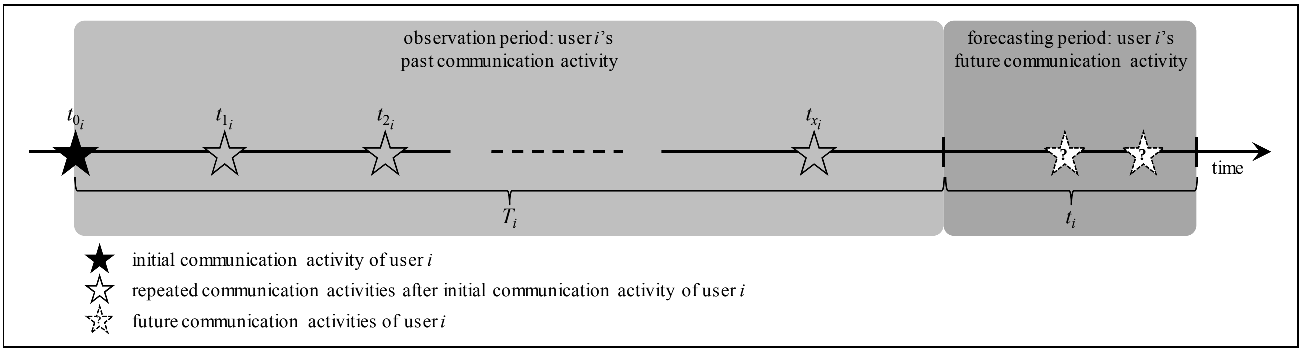

Figure 1. A User's Communication Activity over Time

To predict the future level of a user's communication activity, assumptions about the communication activity process and the time that users stay active are needed (cf. Fader et al., 2005; Schmittlein et al., 1987):

Assumption 1 (Users' communication activity):

> *While being active, each user i's communication activity during a forecasting period of length t_i follows a homogeneous Poisson process with rate $\lambda_i \in IR^+$. This is equivalent to assuming that the time between communication activities is distributed exponentially with rate λ_i and mean $\lambda_i t_i$ (Fader et al., 2005).*

Even though the Poisson assumption has a long validated history for frequently purchased consumer goods (Ehrenberg,

1972) and is often used to describe aspects of human activity (Malmgren, Stouffer, Motter and Amaral, 2008), the transferability to the context of online communication behavior is also critically discussed (Malmgren, Hofman, Amaral and Watts, 2009). Particularly circadian cycles, i.e. for example some people's habit to answer e-mails only in the morning and the evening, are reported to lead to heavy-tailed power-law distributions of inter-communication times (cf. Malmgren et al., 2008). However, many studies confirm the applicability of a homogeneous Poisson process for modeling repeated communication behavior within these cycles, i.e., referring to our example stated above, while being in one session of answering e-mails (e.g. in the morning) (Malmgren et al., 2009). In a first approximation, we assume that users in OSN are in only one cycle while being active and thus a homogeneous Poisson process qualifies for modeling users' communication activity in OSN. After each communication activity, there is a certain probability that a user becomes inactive and does not continue to communicate within the OSN:

Assumption 2 (Users' probability of becoming inactive):

After any communication activity, user i becomes inactive with probability $p_i \in [0;1]$. The point at which the user "drops out" is distributed across communication activities according to a (shifted) geometric distribution with probability mass function:

$$P(\text{inactive immediately after } j\text{th communication activity}) = p_i(1-p_i)^{j-1}, \text{ with } j=1,2,3,....$$

This assumption is supported by prior work on customer retention in general (cf. Fader and Hardie, 2007) and the dropout process of users in OSN in particular (cf. Ahmed, Berchmans, Neville and Kompella, 2010). As there are users with high and users with low levels of communication activity as well as users with high probability to drop out and vice versa, some heterogeneity assumptions are mandated. The gamma distribution is a flexible distribution and can capture the spirit of most of the reasonable distributions on λ_i and p_i (cf. Fader et al., 2005; Schmittlein et al., 1987):

Assumption 3 (Heterogeneity in users' communication activity):

Heterogeneity in λ_i follows a gamma distribution with shape parameter $r \in IR^+$, scale parameter $\alpha \in IR^+$, and probability density function:

$$f(\lambda_i|r,\alpha) = (\alpha^r \lambda_i^{r-1} e^{-\lambda_i \alpha})/\Gamma(r), \text{ with } \lambda_i > 0. \tag{1}$$

Assumption 4 (Heterogeneity in users' probability of becoming inactive):

Heterogeneity in p_i follows a beta distribution with probability density function:

$$f(p_i|a,b) = p_i^{a-1}(1-p_i)^{b-1}/B(a,b), \text{ with } 0 \le p_i \le 1, \tag{2}$$

where $B(a,b)$ with $a \in IR^+$ and $b \in IR^+$ is the beta function, which can be expressed in terms of gamma functions, i.e.:

$$B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b).$$

Finally, there is no a priori reason to favor a positive correlation between users' future communication activity and their probability to drop out over a negative correlation. On the one hand, users with high levels of communication activity may have more frequent opportunities to be disenchanted by the OSN (e.g. due to privacy concerns, system malfunctions) and drop out. On the other hand, these users are probably more strongly attached to the OSN and hence less easily disenchanted. Hence, neglecting interdependencies seems to be a reasonable first approximation:

Assumption 5 (Independence of user's communication activity and probability of becoming inactive):

The transaction rate λ_i and the dropout probability p_i vary independently across users.

With the random variable $Y(t_i)$ denoting the number of a user i's future communication activities initiated in the forecasting period of length t_i, we finally aim to derive the expected number of a user's future communication activities by computing the conditional expectation $E(Y(t_i)|X_i=x_i,t_{x_i},T_i)$ for a user i with observed behavior $(X_i=x_i,t_{x_i},T_i)$. However, λ_i and p_i (cf. assumptions 1 and 2) are unobserved. While there is usually not enough observed user-specific behavior to reliably estimate these parameters for each user, there is generally enough information to estimate the distribution of (λ_i,p_i) over all users. Hence, we can derive the desired probabilities for a randomly chosen user by taking the expectation of the individual-level results over the mixing distributions for λ_i and p_i as given in formulas (1) and (2) (cf. Fader et al., 2005). Thus, the four BG/NBD parameters (r, α, a, b) (cf. assumptions 3 and 4) can be estimated via the method of maximum likelihood. For N users, the sample log-likelihood function is given by (cf. Fader et al., 2005):

$$LL(r,\alpha,a,b) = \sum_{i=1}^{N} \ln\left[L\left(r,\alpha,a,b \middle| X_i = x_i, t_{x_i}, T_i\right)\right], \tag{3}$$

which can be maximized by using standard numerical optimization routines. Thereby, $L(r,\alpha,a,b|X_i=x_i,t_{x_i},T_i)$ represents the likelihood function for a single user i, which can be derived according to formula (6) in Fader et al. (2005). Afterwards, we can calculate the expected number of each user's communication activities in the forecasting period of length t_i by (for a detailed derivation of the formula cf. Fader et al., 2005):

$$E\left(Y(t_i)\middle|X_i=x_i,t_{x_i},T_i,r,\alpha,a,b\right)=\frac{\dfrac{a+b+x_i-1}{a-1}\left[1-\left(\dfrac{\alpha+T_i}{\alpha+T_i+t}\right)^{r+x_i}{}_2F_1\left(r+x_i,b+x_i;a+b+x_i-1;\dfrac{t_i}{\alpha+T_i+t_i}\right)\right]}{1+\delta_{x_i>0}\dfrac{a}{b+x_i-1}\left(\dfrac{\alpha+T_i}{\alpha+t_{x_i}}\right)^{r+x_i}}, \quad (4)$$

where ${}_2F_1(\cdot)$ is the Gaussian hypergeometric function, which can be closely approximated with a polynomial series, and $\delta_{x_i>0}=1$ if $x_i>0$, 0 otherwise (cf. Fader et al., 2005). By calculating the expected number of each user i's future communication activities, users' future level of communication activity in OSN can be predicted. In the following, we demonstrate the method's practical applicability and evaluate its suitability for predicting users' future level of communication activity.

DEMONSTRATION AND EVALUATION

As many other OSN, Facebook.com allows users to set up personal profiles and to establish undirected social links by entering virtual friendship relationships. One of the most popular mechanisms for communication activity within OSN in general and Facebook.com in particular is a message board called "wall" that is included in each profile (Wilson et al., 2009).

For the demonstration and evaluation of the BG/NBD-based method for predicting users' future level of communication activity, we use a publicly available dataset provided by Viswanath et al. (2009) that has also been used and described in detail in Heidemann et al. (2010). It contains 63,731 users of the Facebook.com New Orleans Network connected by 817,090 undirected social links and exhibits the OSN specific characteristics (cf. Heidemann et al., 2010). The dataset also contains information on users' communication activity in terms of 876,687 wallposts initiated and received by the users covered by the dataset. Each wallpost includes information about the initiator, the receiver, and the time at which the wallpost was made. Overall, the wallposts span 227 weeks from September 14, 2004 to January 22, 2009. In the following, we use these wallposts to represent users' communication activity. To account for the potential bias induced by the strong growth of the number of active users after week 175, we chose two scenarios, i.e. a low volatility scenario spanning nine weeks from week 150 to 158 (scenario 1) and a high volatility scenario spanning from week 216 to 224 (scenario 2). Figure 2 displays the development of the number of wallposts and the number of distinct active users covered by the dataset over time.

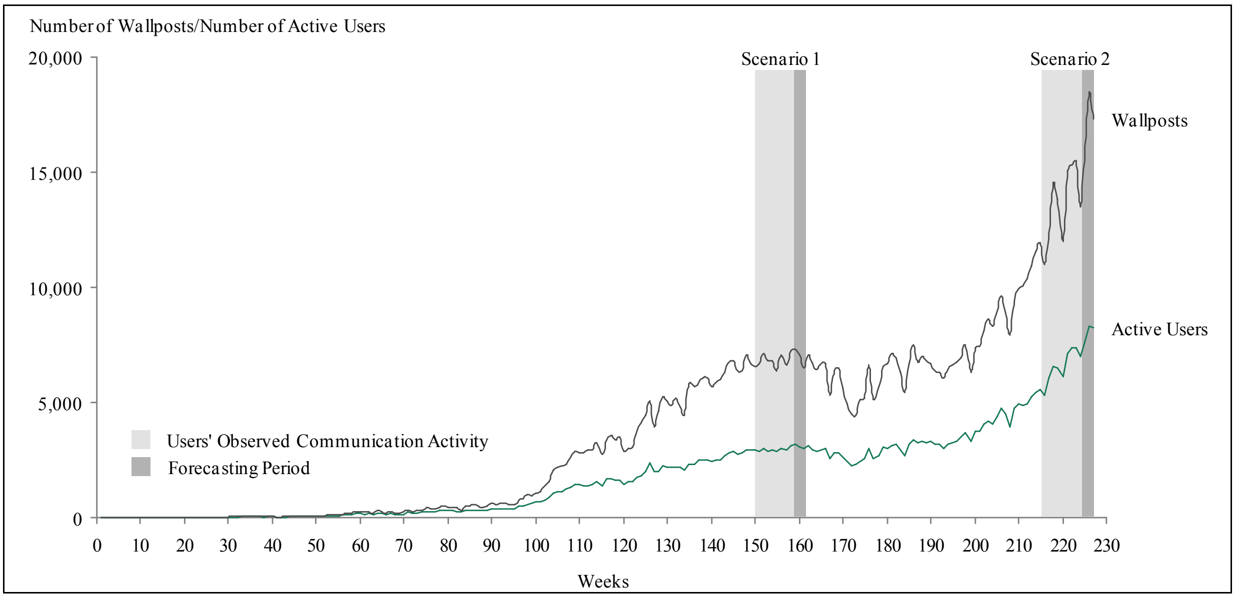

Figure 2. Number of Wallposts and Number of Active Users over Time

Table 1 summarizes both scenarios' characteristics.

Characteristic	Scenario 1	Scenario 2
Total Number of Active Users	9,815	25,182
Total Number of Wallposts	60,138	130,824
Average Number of Active Users per Week	2,994	6,965
Standard Deviation of Active Users per Week	100	894
Average Number of Wallposts per Week	6,682	14,536
Standard Deviation of Wallposts per Week	296	2,182

Table 1. Overview of Scenarios' Characteristics

For each scenario, we aim to predict the users' number of wallposts during a forecasting period of three weeks, i.e. from week 159 to 161 (scenario 1) and week 225 to 227 (scenario 2), respectively. Therefore, we first derive each user i's observed behavior (X_i=x_i,t_{x_i},T_i) and estimate the parameters (r, α, a, b) by applying formula (3). Table 2 summarizes the results.

Parameter	Scenario 1	Scenario 2
r	0.485	0.420
α	0.332	0.293
a	0.433	0.742
b	4.346	5.555

Table 2. Estimation Results

Second, we calculate for both scenarios the expected number of each user's wallposts in the forecasting period according to formula (4). To evaluate the suitability of the BG/NBG-based method for predicting users' future level of communication activity in OSN, we apply the evaluation approach suggested by Fader et al. (2005). In Figure 3, we report the average of the predicted along with the average of the actual number of wallposts that took place in the forecasting period broken down by the number of wallposts in the users' observation periods for scenario 1.

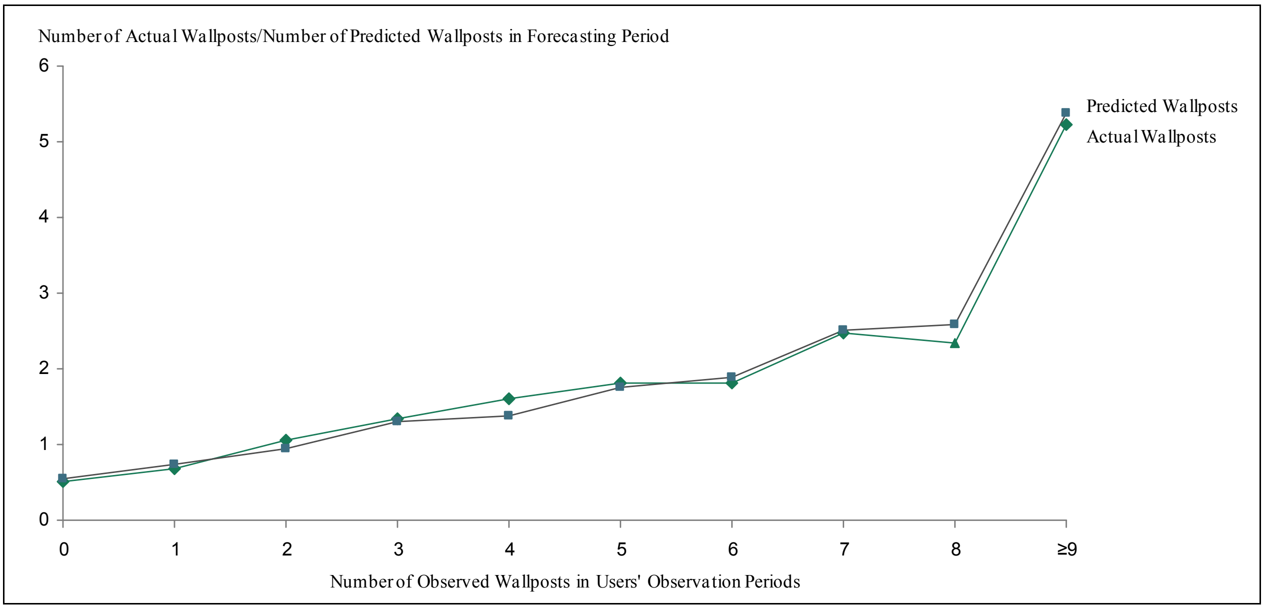

Figure 3. Evaluation Scenario 1

The virtually absent deviation between the users' predicted and actual wallposts highlights that in scenario 1 the BG/NBG-based method provides excellent predictions of the users' number of wallposts in the forecasting period. Figure 4 displays the evaluation's results for scenario 2.

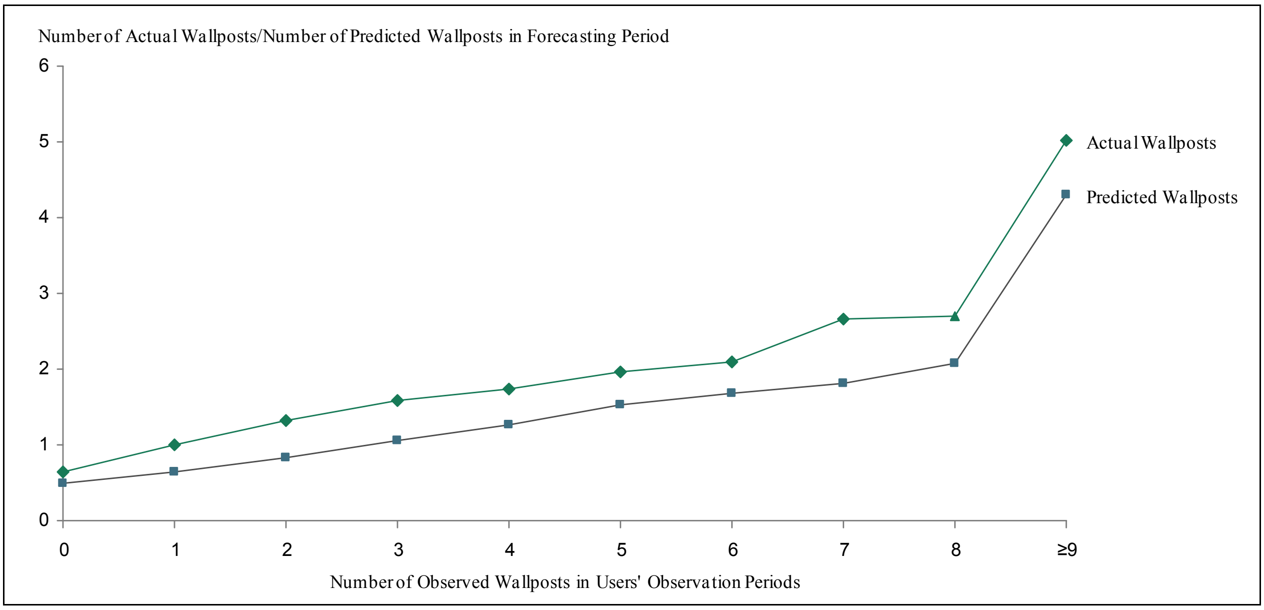

Figure 4. Evaluation Scenario 2

Here, the evaluation reveals that the users' actual level of communication activity is slightly underestimated. However, the deviation of 0.4 on average is quite small and constant. Thus, we think future work will be able to account for high volatility in the data by slightly adapting the method. Taken together, the BG/NBG-based method seems to be suitable for the prediction of users' future level of communication activity in OSN.

CONCLUSION

Even though active usage has reached enormous scales, the majority of OSN relying on the advertisement-based business model face the challenge of generating sustainable revenues. Particularly the identification of users with high levels of future communication activity plays an important role when developing more effective advertising strategies by addressing users deliberately. For the identification of these users, we transferred a probability-based method, which has been primarily developed to forecast purchasing behavior of customers, to the context of users' communication activity in OSN. The application and evaluation illustrated that the BG/NBG-based method seems to be suitable for predicting users' future level of communication activity in OSN. Even though the method seems to be rather complex at first sight, the necessary parameters can be derived easily and the calculation can be automated and even be done in Excel, as pointed out by Fader et al. (2005). Nevertheless, future work is needed to further evaluate the approach (e.g. regarding further sample groups), for instance by using other datasets and by considering also economic aspects. The method's formal denotation implicates assumptions and limitations, for instance regarding the underlying probability distributions. Future work needs to address these issues either by confirming or by relaxing these assumptions. Moreover, upcoming challenges, for instance due to changing privacy practices, need to be carefully observed and considered when developing approaches for the identification of users with high levels of future communication activity. Finally, in line with prior research on the identification of influential users in OSN (e.g. Trusov et al., 2010) we did not address the question how responsive highly active users are to certain marketing strategies (e.g. word-of-mouth marketing). Even though this question is subject to future research, the BG/NBG-based method serves as a first step towards more effective advertising strategies in OSN.

REFERENCES

1. Agarwal, R., Gupta, A. K. and Kraut, R. (2008) The Interplay Between Digital and Social Networks (Editorial Overview), *Information Systems Research*, 19, 3, 243-252.
2. Ahmed, N. K., Berchmans, F., Neville, J. and Kompella, R. (2010) Time-Based Sampling of Social Network Activity Graphs, in *Proceedings of the 8th Workshop on Mining and Learning with Graphs (MLG 10)*, July 25 - 28, Washington, DC, USA, 1-9.
3. Alison, D. (2010) Social Media Up 230% Since 2007, http://www.informationweek.com/news/software/web_services/ showArticle.jhtml?articleID=225701600&subSection=News, 10-20-2010.
4. Boyd, D. M. and Ellison, N. B. (2007) Social Network Sites: Definition, History, and Scholarship, *Journal of Computer-Mediated Communication*, 13, 1, 210-230.
5. Brown, J., Broderick, A. J. and Lee, N. (2007) Word of Mouth Communication within Online Communities: Conceptualizing the Online Social Network, *Journal of Interactive Marketing*, 21, 3, 2-20.
6. Butler, B. S. (2001) Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures, *Information Systems Research*, 12, 4, 346-362.
7. Cheung, C. M. K. and Lee, M. K. O. (2010) A Theoretical Model of Intentional Social Action in Online Social Networks, *Decision Support Systems*, 49, 1, 24-30.
8. Clemons, E. K. (2009) The Complex Problem of Monetizing Virtual Electronic Social Networks, *Decision Support Systems*, 48, 1, 46-56.
9. Cummings, J. N., Butler, B. and Kraut, R. (2002) The Quality of Online Social Relationships, *Communications of the ACM,* 45, 7, 103-108.
10. De Choudhury, M., Sundaram, H., John, A. and Seligmann, D. D. (2007) Contextual Prediction of Communication Flow in Social Networks, in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI 07)*, November 02 - 05, Silicon Valley, CA, USA, 57-65.
11. Dorogovtsev, S. N. and Mendes, J. F. F. (2003) Evolution of Networks: From Biological Networks to the Internet and WWW, Oxford University Press, Oxford.
12. Dougherty, H. (2010) Facebook Reaches Top Ranking in US, http://weblogs.hitwise.com/heather dougherty/2010/03/ facebook_reaches_top_ranking_i.html, 10-20-2010.
13. Dwyer, C., Hiltz, S. and Passerini, K. (2007) Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and MySpace, in: *Proceedings of the 13th Americas Conference on Information Systems (AMCIS 07)*, August 09 - 12, Keystone, Colorado, USA, Paper 339.
14. Ehrenberg, A. S. C. (1972) Repeat Buying, North-Holland, Amsterdam.
15. Ellison, N., Steinfield, C. and Lampe, C. (2007) The Benefits of Facebook Friends: Social Capital and College Students Use of Online Social Network Sites, *Journal of Computer-Mediated Communication*, 12, 4, 1143-1168.
16. Facebook (2010) 500 Million Stories, http://blog.facebook.com/blog.php?post=409753352130, 10-20-2010.
17. Fader, P. S. and Hardie, B. G. S. (2007) How to Project Customer Retention, *Journal of Interactive Marketing*, 21, 1, 76-90.
18. Fader, P. S., Hardie, B. G. S. and Lee, K. L. (2005) "Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model, *Marketing Science*, 24, 2, 275-284.
19. Ganley, D. and Lampe, C. (2009) The Ties That Bind: Social Network Principles in Online Communities, *Decision Support Systems*, 47, 3, 266-274.
20. Gnyawali, D. R., Fan, W. and Penner, J. (2010) Competitive Actions and Dynamics in the Digital Age: An Empirical Investigation of Social Networking Firms, *Information Systems Research*, 21, 3, 594-613.

21. Heidemann, J., Klier, M. and Probst, F. (2010) Identifying Key Users in Online Social Networks: A PageRank Based Approach, in *Proceedings of the 31st International Conference on Information Systems (ICIS 10)*, December 12 - 15, St. Louis, MO, USA, Paper 79.
22. Hill, S., Provost, F. and Volinsky, C. (2006) Network-Based Marketing: Identifying Likely Adopters via Consumer Networks, *Statistical Science*, 21, 2, 256-276.
23. Hoffman, D. L. and Fodor, M. (2010) Can You Measure the ROI of Your Social Media Marketing? *MIT Sloan Management Review*, 52, 1, 41-49.
24. Jacks, T. and Salam, A. F. (2009) Computer-Mediated Friendship Networks, in *Proceedings of the 30th International Conference on Information Systems (ICIS 09)*, December 15 - 18, Phoenix, AZ, USA, Paper 115.
25. Krasnova, H., Hildebrand, T. and Günther, O. (2009) Investigating the Value of Privacy in Online Social Networks: Conjoint Analysis, in *Proceedings of the 30th International Conference on Information Systems (ICIS 09)*, December 15 - 18, Phoenix, AZ, USA, Paper 173.
26. Lampe, C., Ellison, N. and Steinfield, C. (2006) A Face(book) in the Crowd: Social Searching vs. Social Browsing, in *Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW 06)*, November 04 - 08, Banff, AB, Canada, 167-170.
27. Lee, J., Lee J.-H. and Lee, D. (2009) Impacts of Tie Characteristics on Online Viral Diffusion, *Communications of the AIS*, 24, 1, Article 32.
28. Leimeister, J. M., Sidiras, P. and Krcmar, H. (2004) Success factors of virtual communities from the perspective of members and operators: An empirical study, in: *Proceedings of the Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS 04)*, January 05 - 08, Big Island, HI, USA.
29. Liben-Nowell, D. and Kleinberg, J. (2007) The Link-Prediction Problem for Social Networks, *Journal of the American Society for Information Science and Technology*, 58, 7, 1019-1031.
30. Lu, H.-P. and Hsiao, K.-L. (2010) The Influence of Extro/Introversion on the Intention to Pay for Social Networking Sites, *Information & Management*, 47, 3, 150-157.
31. Malmgren, R. D., Hofman, J. M., Amaral, L. A. N. and Watts, D. J. (2009) Characterizing Individual Communication Patterns, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 09)*, June 28 - July 1, Paris, France, 607-616.
32. Malmgren, R. D., Stouffer, D. B., Motter, A. E. and Amaral L. A. N. (2008) A Poissonian Explanation for Heavy Tails in E-Mail Communication, *Proceedings of the National Academy of Sciences of the United States of America*, 105, 47, 18153-18158.
33. Ray, A., Bernoff, J. and Wise, J. (2010) Peer Influence Analysis - Mass Influencers are the Key to Achieving Scale in Social Media Marketing, http://www.forrester.com/rb/Research/peer_influence_analysis/q/id/56766/t/2, 10-20-2010.
34. Schmittlein, D. C., Morrison, D. G. and Colombo, R. (1987) Counting Your Customers: Who Are They and What Will They Do Next?, *Management Science*, 33, 1, 1-24.
35. Trusov, M., Bodapati, A. V. and Bucklin, R. E. (2010) Determining Influential Users in Internet Social Networks, *Journal of Marketing Research*, 47, 4, 643-658.
36. Trusov, M., Bucklin R. E. and Pauwels, K. (2009) Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site, *Journal of Marketing*, 73, 5, 90-102.
37. Viswanath, B., Mislove, A., Cha, M. and Gummadi, K. P. (2009) On the Evolution of User Interaction in Facebook, in *Proceedings of the 2nd ACM Workshop on Online Social Networks (WOSN 09)*, August 17, Barcelona, Spain, 37-42.
38. Wen, C., Tan, B. C. Y. and Chang, K. T.-T. (2009) Advertising Effectiveness on Social Network Sites: An Investigation of Tie Strength, Endorser Expertise and Product Type on Consumer Purchase Intention, in *Proceedings of the 30th International Conference on Information Systems (ICIS 09)*, December 15 - 18, Phoenix, AZ, USA, Paper 151.
39. Williamson, D. A. (2009) Social Networking Worldwide: Ad Spending and Usage, http://www.emarketer.com/Reports/All/Emarketer_2000567.aspx, 10-20-2010.

40. Willinger, W., Rejaie, R., Torkjazi, M., Valafar, M. and Maggioni, M. (2010) Research on Online Social Networks: Time to Face the Real Challenges, *ACM SIGMETRICS Performance Evaluation Review,* 37, 3, 49-54.

41. Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P. N. and Zhao B. Y. (2009) User Interactions in Social Networks and Their Implications, in *Proceedings of the 4th ACM European Conference on Computer Systems*, April 01 - 03, Nuremberg, Germany, 205-218.

42. Xiang, R., Neville, J. and Rogati, M. (2010) Modeling Relationship Strength in Online Social Networks, in *Proceedings of the 19th International Conference on World Wide Web (WWW 10)*, April 26 - 30, Raleigh, NC, USA, 981-990.

43. Xu, Y., Zhang, C., Xue, L. and Yeo, L. L. (2008) Product Adoption in Online Social Network, in *Proceedings of the 29th International Conference on Information Systems (ICIS 08)*, December 14 - 17, Paris, France, Paper 200.

44. Zhang, Y., Wang, Z. and Xia, C. (2010) Identifying Key Users for Targeted Marketing by Mining Online Social Network, in *Proceedings of the 24th International Conference on Advanced Information Networking and Applications Workshops (WAINA 10)*, April 20 - 23, Perth, WA, Australia, 644-649.