Longevity in Second Life

Chun-Yuen Teng and Lada A. Adamic
School of Information
1085 S. University Ave.
Ann Arbor, MI 48103

Abstract

User retention is important to the success of online social media, particularly in virtual world settings where users shape one another's online experience. We study a rich set of variables, including social network and group membership, chatting, and transactions, in order to predict which users will stay and which ones will leave. We find that simple variables directly measuring the intensity and diversity of a user's interaction with others are most predictive.

Introduction

The past few years have seen an rise in number and popularity of online spaces where individuals can socialize, play, and learn. All of these spaces face the challenge of retaining the interest of users over time. We study this problem in the context of Second Life (SL). Launched by Linden Lab in 2003, SL is a well-established online virtual world. Unlike most other online virtual worlds, whose focus is on gaming, and whose content is created by game developers, SL relies on its users to build and create their own spaces and interactions. Users, as individuals, corporations, and nonprofits have created spaces whose purpose ranges from gaming, to socializing, to learning, to product promotion, to training. This open-endness creates both very diverse opportunities for interaction, but also a hazard that users may not find purpose, and might leave.

Much of social media relies on "social networking" features that allow users to maintain and accrue social ties. Even the presence of one's friends on such sites is no guarantee that a user will remain active. According to a Nielson report (Flint 2009), 60% of Twitter users don't user Twitter after the first month. Even wildly successful sites such as Facebook are not complacent about user retention (Milian 2009).

What makes SL a particularly interesting substrate to study retention in is the freedom and richness of user experience. While SL provides users the ability to add "buddies" with whom they can then easily converse and share within the virtual world, it also allows users to create groups that enable everything from expressing different interests and affiliations to coordinating in-world activities, such as land ownership, clubbing, and education. Furthermore, users can be consumers as well as producers of content, which is sometimes given away and other times comes with a fee.

We therefore divide our user retention analysis according to different facets of the user experience:

1. Usage: How much time a user has spent in SL.
2. Networking: How many contacts and groups users are associated with, how tightly knit their social networks are and how diverse their groups are.
3. Interaction: Social and group ties can grow stale if not maintained through regular chat. SL also makes it easy to meet and interact with new people.
4. Transaction: Creating content or providing services in SL can be profitable, with 150M USD in user-to-user transactions taking place in the third quarter of 2009 (Linden 2009).

Besides the obvious relevance of our analysis to user retention in SL in particular and social media sites in general, the study provides insights into how social structures contribute to user engagement and retention.

Related work

Several previous studies have addressed factors affecting users' participation in online communities. Butler (2001) studied the effect of community size and user contribution on online listserv participation. Brandtzaeg et al. (2007) linked lack of interesting participants and low quality content to user disengagement. Backstrom et al. (Backstrom et al. 2006) showed that having contacts, particularly contact who know one another, is predictive of who will join a community. Not all participation is motivated by social factors. Two studies, (Hertel, Niedner, and Herrmann 2003) (Rossi 2004) found that enjoyment of problem solving, and the ability to earn reputation and money motivate users to participate. Motivation for contributing content differs by experience level. e.g. (Bryant, Forte, and Bruckman 2005), and users are motivated both intrinsically and extrinsically (Kim, Na, and Ryu 2007).

In a massively multiplayer online game setting, Wolf (Wolf 2007) for that individual roles and goals, community and practice attributes are correlated with user activity. Other studies have found that even just the initial user interaction with others can have an impact on whether the user continues contributing to online social media such as Wikipedia (Panciera, Halfaker, and Terveen 2009), or online forums (Joyce and Kraut 2006; Arguello et al. 2006).

In this study we expand on previous work by looking at a rich set of variables in the context of SL. Rather than looking at initial interactions, we aim to predict which users are likely to leave, and base our predictions on the structure, intensity, and profitability of a user's activity.

Description of Data and Methods

The dataset was provided to us by Linden Lab, and included various facets of users' activity, including weekly snapshots of the social network, group affiliations, summary data on users including their first and most recent login, transaction data between users, and pairwise user chat frequencies. We focused our analysis on a two month period spanning May and June of 2009 for which the chat data was available.

In order to quantify user retention, we defined users as "active" if they logged in within 30 days prior to the date of interest. Specifically, we take the set of users active as of May 22, 2009, and check whether they are still active by June 26, 2009. We further divide users into two groups: positive-revenue users and zero revenue users, depending on whether they swapped outside-world currency for the in-world currency, Linden (as of Dec. 27, 2009, a dollar traded for approximately 260 Linden). They might have done so to obtain a paid membership, which allows them privileges such as land ownership, or to purchase goods and services in-world.

As of May 22, 2009, there were 537,610 active users whose first login was 30 or more days prior. 30% of those users had contributed to revenue. Already here we see an expected difference in retention, with users who had invested in their Second Life being more likely to remain. In the set of zero-revenue users, 79.4% remain active in the next 30 days. In the set of positive-revenue users, 95.4% remain active. We perform the following analysis for positive revenue users, and omit similar results for zero revenue users for brevity.

Our primary method is logistic regression, with a binary variable indicating whether the user remained active or not. We generate a balanced dataset from the full dataset by taking all users who became inactive and sampling an equal number of users who remained active. With the balanced dataset, a random guess would be correct 50% of the time, and this becomes our baseline. We divide user parameters according to the four groups above: usage, networking, interaction, and transaction, first testing their predictive accuracy individually, and then combining them.

Table 1: Usage and survival. Accuracy is reported for logistic regressions using each variable individually.

parameter	Estimate	s	p	acc.
log(usage minutes + 1)				
	0.641	0.013	$< 2e^{-16}$	0.697
age in days since first login				
	$-1.6e^{-04}$	$4.3e^{-05}$	$1.34e^{-4}$	0.515

Analysis

Usage attributes

The most straightforward variables relate to how long the individual has been a user of SL, and how much time they spent in SL. The length of time one had been an SL user was not a significant predictor, but the intensity, measured as the number of minutes spent in-world, was.

Networking attributes

For the users' networks, we consider their size in terms of the number of friends (N_F), the number of groups (N_G), the number of friends active in the previous month (N_{aF}) and the ratio of N_{aF}/N_F. But we also look at how constrained those social networks and groups are. The clustering coefficient (CC_F) is the proportion of one's friends who know one another. Having a high clustering coefficient means that one moves primarily in one or very few social circles. Similarly, belonging to different groups may expose an individual to many different kinds of people, or the same people over and over again. (GO_u) measures the pairwise group overlap for all groups the user belongs to using the Jaccard coefficient: $GO_u = \sum \frac{|N_{Gi} \bigcap N_{Gj}|}{|N_{Gi} \bigcup N_{Gj}|}$, where N_{G_i} and N_{G_j} are membership sizes of groups i and j. For example, if a user belongs to two groups, A and B, and group A has 5 members, group B has 10 members, and there are 3 members in common, then $GO_u = \frac{|A \cap B|}{|A \cup B|} = 3/12 = 1/4$. We similarly define ($GO_F$) to be the pairwise overlap in groups for a user's friends.

Table 2 shows that while all parameters are highly correlated with users' survival in SL, the raw number of contacts and groups, rather than their diversity, is key in identifying users who will stay vs. leave. Although the correlation is weak, diversity does play a positive role in user retention. The proportion of one's friends who are active also correlates positively with staying on.

Interaction

Social network and group membership data can become stale, as a user grows less interested in certain friends and groups without necessarily removing them

Table 2: Social attributes and user retention

parameter	Estimate	S	p	accuracy
log(# friends + 1)				
	0.831	0.017	$< 2e^{-16}$	0.697
log(# active friends + 1)				
	0.848	0.017	$< 2e^{-16}$	0.700
proportion of friends who are active				
	0.061	0.002	$< 2e^{-16}$	0.614
clustering ceff. of buddy graph				
	-0.259	0.062	$3.12e^{-05}$	0.511
# groups				
	0.111	0.002	$< 2e^{-16}$	0.681
group overlap by user				
	-1.031	0.187	$3.32e^{-04}$	0.502
group overlap by friend				
	6.344	0.373	$< 2e^{-16}$	0.604

Table 3: Chat and retention

parameter	Estimate	S	p	accuracy
log(# chat partners + 1)				
	1.116	0.018	$< 2e^{-16}$	0.766
log(# friends chatted with + 1)				
	1.844	0.035	$< 2e^{-16}$	0.756
proportion of chat partners who are friends				
	3.289	0.078	$< 2e^{-16}$	0.717
clustering coeff. of chat network				
	2.384	0.080	$< 2e^{-16}$	0.679
entropy in chat partners				
	0.771	0.013	$< 2e^{-16}$	0.763
# days on which one chatted				
	1.33051	0.02101	$< 2e^{-16}$	0.782
entropy of chat times				
	2.653	0.0453	$< 2e^{-16}$	0.777

from their profile. In addition, SL interactions can occur outside of explicitly defined social ties and groups. Since text chat is a frequent form of communication in SL, we derive several chat-related variables, listed in Table 3, and test their ability to predict a user's continued participation. We include two entropy measures, $-\sum_i p_i \log p_i$. The first captures how dispersed a user's chatting activity is among different partners, and has p_i as the proportion of chats with user i. For example, if user A chatted with user B once and user C 3 times, then the entropy in chat partners is given by $-\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4}$. The second captures how dispersed a user's chat activity is in time, with p_i being the proportion of chats on day i. For example, if a user chats twice on day 1 and once on day 8, the temporal chat entropy is $-\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}$.

We observe that almost all chat parameters are more predictive than the static network measures above. Furthermore, one need not resort to complex metrics because the best predictions are also the simplest, e.g, the number of chat partners (not necessarily friends), or the number of days on which the user chatted.

Transaction-related metrics

Users in SL not only socialize and seek entertainment, but are the ones creating the whole experience, from places and objects to games and music. Therefore, the production, distribution, and consumption of these goods can have an impact of users' retention. In addition to data on purchases, sales, and transfer of free goods, we include variables relating to the proximity of one's customers to oneself (average shortest path), and proximity of one's customers to one another (clustering coefficient and "clumpy distance"). Here, average path length between customers in the social graph is defined as $D_a = \frac{1}{N_c*(N_c-1)}\sum_{i,j} 1/d(i,j)$, where $d(i,j)$ is the distance between customer i and customer j. The clumpy distance $D_c = \frac{2}{N_c(N_c-1)}\sum_{C_A,C_B}\frac{1}{d_(C_A,C_B)} \cdot \frac{N_{C_A}N_{C_B}}{2}$, where C_i and C_j are connected components of customers, $d_(C_A, C_A)$ is the distance between the components, and N_{C_A} and N_{C_B} are the component sizes.

Table 4 shows that economic activity is correlates with survival, albeit more weakly than chat. Making, as opposed to spending, money is not essential. Profit, the difference between money obtained and does not improve predictions of whether a user will stay. The amount of money the user paid to Linden (as opposed to other users) is only weakly positive for survival. Interestingly, having a high proportion of free transactions is highly predictive. Free transactions are more likely to occur between friends, and could be, similar to chat, an indication of social interaction. We again observe that complex parameters relating to the customer network, such as clumpy distance, clustering coefficient, and average shortest path length are less predictive than simple parameters relating to the intensity of the user's activity, such as transaction counts.

Analysis of overall parameters

Finally, we combine all categories of user variables in a single regression, and obtain an overall accuracy of 0.802. Figure 1 is a heatmap summarizing the correlations between all parameters, with transaction-related parameters clustering near the top and social and interactive attributes clustering at the bottom.

Discussion and Future work

In this paper we studied factors associated with user retention in the online virtual world Second Life. After observing that a high percentage of users who had once invested in SL are likely to remain, we looked at additional specific social, interaction, transaction variables associated with retention. For those interested in identifying users likely to leave, we found that complex variables are no better than simple predictors. And among those simple predictors, by far it is interaction with oth-

Table 4: Transactional parameters and user retention

parameter	Estimate	S	p	acc.
log(Amount made + 1)				
	0.327	0.008	$< 2e^{-16}$	0.662
log(Amount spent + 1)				
	0.261	0.006	$< 2e^{-16}$	0.696
log(Profit - min(Profit) + 1)				
	1.0846	0.7252	0.135	0.5
log(lifetime revenue + 1)				
	0.166	0.008	$< 2e^{-16}$	0.57
log(# customers + 1)				
	0.796	0.023	$< 2e - 16$	0.669
log(# sellers this user bought from + 1)				
	2.155	0.044	$< 2e - 16$	0.705
log(# selling transactions + 1)				
	1.200	0.035	$< 2e^{-16}$	0.669
log(# buying transactions + 1)				
	0.971	0.022	$< 2e^{-16}$	0.705
proportion of free transactions				
	2.659	0.053	$< 2e^{-16}$	0.706
clustering coeff. of customers				
	4.5129	0.3366	$< 2e^{-16}$	0.556
average path length				
	2.34037	0.10785	$< 2e^{-16}$	0.567
clumpy distance				
	17.47341	0.63659	$< 2e^{-16}$	0.597

ers, whether friends or strangers, that correlated most with long user life. This implies that successful online services should encourage and facilitate users to interact with others more regularly and more diversely. In future work we would like to further explore the mechanisms by which users influence each others' continued participation.

Acknowledgements

We would like to thank Linden Lab for sharing SL data and for valuable discussions.

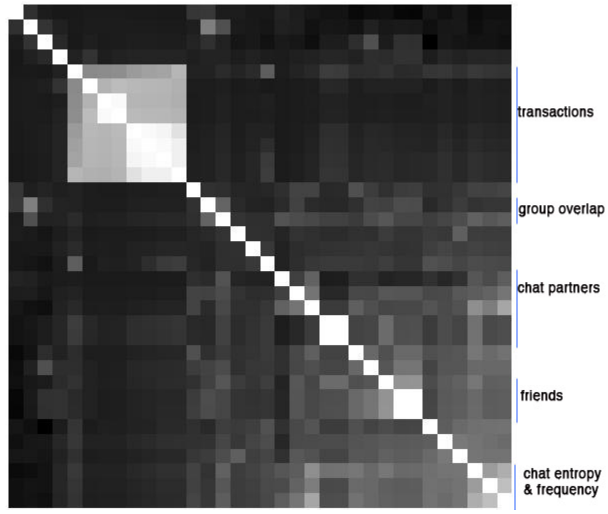

Figure 1: Heatmap of all parameters

References

Arguello, J.; Butler, B.; Joyce, E.; Kraut, R.; Ling, K.; Rosé, C.; and Wang, X. 2006. Talk to me: Foundations for successful individual-group interactions in online communities. In *SIGCHI '06*, 968. ACM.

Backstrom, L.; Huttenlocher, D.; Kleinberg, J.; and Lan, X. 2006. Group formation in large social networks: membership, growth, and evolution. In *KDD '06*, 44–54. ACM.

Brandtzaeg, P. B., and Heim, J. 2007. User loyalty and online communities: why members of online communities are not faithful. In *INTETAIN '08*, 1–10.

Bryant, S. L.; Forte, A.; and Bruckman, A. 2005. Becoming wikipedian: transformation of participation in a collaborative online encyclopedia. In *GROUP '05*, 1–10. ACM.

Butler, B. 2001. Membership size, communication activity, and sustainability: A resource-based model of online social structures. *ISR*- 12(4):346–362.

Flint, J. 2009. Nielsen on twitter: It's so five seconds ago. `http://latimesblogs.latimes.com/entertainmentnewsbuzz/2009/04/nielsen-o%n-twitterits-so-five-seconds-ago.html`.

Hertel, G.; Niedner, S.; and Herrmann, S. 2003. Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel. *Research policy* 32(7):1159–1177.

Joyce, E., and Kraut, R. 2006. Predicting continued participation in newsgroups. *JCMC*- 11(3):723.

Kim, S.; Na, E.; and Ryu, M. 2007. Factors Affecting User Participation in Video UCC (User-Created Contents) Services. In *CCT '07*, 209.

Linden, T. 2009. The second life economy - third quarter 2009 in detail. `https://blogs.secondlife.com/community/features/blog/2009/11/02/`.

Milian, M. 2009. Twitter creator jack dorsey on user retention, facebook comparisons. `http://latimesblogs.latimes.com/technology/2009/05/twitter-jack-dorsey-%nielsen.html`.

Panciera, K.; Halfaker, A.; and Terveen, L. 2009. Wikipedians are born, not made: a study of power editors on wikipedia. In *GROUP '09*, 51–60. ACM.

Rossi, M. 2004. Decoding the Free/Open Source (F/OSS) Software Puzzle a survey of theoretical and empirical contributions. *DISCE* 424:1–40.

Wolf, K. 2007. Communities of Practice in MMORPGs: an entry point into addiction? In *CCT '07*, 191.