Did You Blog Yesterday? Retention in Community Blogs

Imrul Kayes[1], Xiang Zuo[1], Da Wang[2], Jacob Chakareski[3]

[1] University of South Florida, [2] HuBei University of Technology, [3]University of Alabama

ABSTRACT

- Blogging has a high impact on marketing, shaping public opinions, and informing the world about major events from a grassroots point of view [1].
- Turnover in online blogging is very high, with most people who initially join and start contributing to the community, failing to contribute in the long run [2].
- We ask what factors cause a blogger to continue participating in the community by contributing content (e.g., posts, comments).

DATASET AND METHODOLOGY

- Used Metropolis-Hastings Random Walk (MHRW) algorithm[3] to crawl a popular blogging network, ''Blogster''.
- Scraped HTML pages of 17,436 users (contributed about 91% posts), who form a social graph with 72,907 edges.
- Five categories of predictor variables:
  >Network metrics (clustering coefficient, degree, betweenness, closeness, pagerank, and communicability centrality)
  >Activity (posts, comments, photos, network age)
  >Physiology (age, gender)
  >Interactional (blog traffic, other users' comments)
  >Relational (social tie strength, friends retention)
- Performed multiple linear regression considering points as retention and response variable.

RESULTS

$$Points_i = \alpha + \beta_1.DegreeRank_i + \beta_2.CC_i + \beta_3.CommunicabilityRank_i + \beta_4.BlogTraffic_i + \beta_5.UserComments_i + \beta_6.Age_i + \beta_7 Gender_i + \beta_8 AvgFriendsRet_i + \epsilon_i$$

	Estimate (β)	Std. Error	t value	Pr($>|t|$)
(Intercept)	3.66e+03	1.86e+02	19.62	$< 2e-16$ ***
DegreeRank	-1.92e+01	1.07e+00	-17.98	$< 2e-16$ ***
Comm.Rank	-1.57e-02	4.25e-03	-3.68	0.000233 ***
CC	-8.71e+01	4.06e+01	-2.14	0.032081 *
BlogTraffic	3.62e-02	8.21e-04	44.06	$< 2e-16$ ***
UserComments	1.02e+00	1.59e-02	64.61	$< 2e-16$ ***
Age	5.03e+00	7.29e-01	6.90	$5.45e-12$ ***
Gender	-9.38e+00	1.86e+01	-4.50	0.001257 **
AvgFriendsRet	1.53e-04	2.04e-03	3.08	0.021513*

Table 1: The results of a multiple linear regression with points as the dependent variable. Significance codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.**

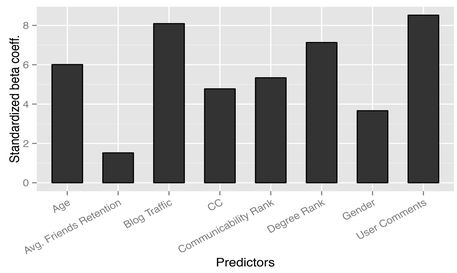

Figure 1: Standardized beta coefficients (β) show the relative effect sizes that each input variable has on points.

- Male and aged (senior) bloggers, who face fewer constraints and have more opportunities in the community are more retained than others.
- Bloggers pay a high degree of attention to retained bloggers through implicit (reading posts) and explicit (writing comments) interactions.
- A blogger has higher retention if her friends have also higher retention
- A strong social tie reduces retention imbalance between two blogger friends.
- A blogger's network age (e.g., how long ago she joined) has no effect on her retention.

CONCLUSIONS

- Our work has multiple practical and theoretical implications.
- Future research on social behavior of bloggers will benefit from the understanding of the variables that predict continued activity.
- Retention variables will allow the developers of a new community blog to make more informed design decisions.
- One could imagine a "retention score" from retention variables that can complement the incentive oriented scores (e.g., points) in a blogging platform.

FUTURE WORK

In future, we want to investigate why bloggers show sporadic retention (e.g., active and inactive in a time interval).

REFERENCES

1. D. Gillmor. We the Media: Grassroots Journalism by the People, for the People. O'Reilly, 2006.
2. E. Joyce and R. E. Kraut. Predicting continued participation in newsgroups. Journal of Computer-Mediated Communication, 11(3):723–747, 2006.
3. M. Gjoka, M. Kurant, C. T. Butts, and A. Markopoulou. Practical recommendations on crawling online social networks. Selected Areas in Communications, IEEE Journal on, 29(9):1872–1892, 2011.