2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

The Influence of Feedback with Different Opinions on Continued User Participation in Online Newsgroups

Teng Wang, Keith C. Wang, Fredrik Erlandsson, S. Felix Wu
Department of Computer Science
University of California, Davis
Davis, CA 95616, U.S.A
Email: {tewang, kcwang, fer, sfwu}@ucdavis.edu

Robert Faris
Department of Sociology
University of California, Davis
Davis, CA 95616, U.S.A
Email: rwfaris@ucdavis.edu

***Abstract*—With the popularity of social media in recent years, it has been a critical topic for social network designer to understand the factors that influence continued user participation in online newsgroups. Our study examines how feedback with different opinions is associated with participants' lifetime in online newsgroups. Firstly, we propose a new method of classifying different opinions among user interaction contents. Generally, we leverage user behavior information in online newsgroups to estimate their opinions and evaluate our classification results based on linguistic features. In addition, we also implement this opinion classification method into our SINCERE system as a real-time service. Based on this opinion classification tool, we use survival analysis to examine how others' feedback with different opinions influence continued participation. In our experiment, we analyze more than 88,770 interactions on the official Occupy LA Facebook page. Our final result shows that not only the feedback with the same opinions as the user, but also the feedback with different opinions can motivate continued user participation in online newsgroup. Furthermore, an interaction of feedback with both the same and different opinions can boost user continued participation to the greatest extent. This finding forms the basis of understanding how to improve online service in social media.**

Keywords*—*Online Newsgroups, Opinion Classification, Social Influence, Continued User Participation

I. Introduction

Social media has influenced many aspects of people's lifestyle. It not only changed the way people collect news and information, but even reformed the way people communicate with each other. One of the most popular uses of social media is to support online newsgroups [1]. These groups allow people to seek the latest news and exchange opinions on a wide variety of topics, from entertainment and education to religions and politics. In order to improve the effectiveness of online newgroups, it is an important task for social media designers to motivate user continued participation on their websites. Actually, social media services are highly dependent on user participation to provide value to their products [2]. However, it is not an easy task. Member participation in online newsgroups is often sparse and uneven [3]. In this paper, we examine the factors that influence user continued participation in online newsgroups. To be specific, our work focuses on the effect of feedback with different opinions.

In 2008, S.L. Johnson [4] gave a detailed definition of online groups from the views of group membership and interaction. Besides the criteria mentioned by S.L. Johnson, online newsgroups, as a special case of online groups, also have their own characteristics: members in online newsgroups are less socially connected in real life; most of them are strangers with little off-line communication. These characteristics make online newsgroups an ideal resource for researchers to examine user influence. Firstly, because most of the members in newsgroups are strangers in real life, they are more likely to share their opinions online, while people in private friendship groups may have more concerns. Additionally, because most of people's interaction happens in the online environment, offline influence will have very little effect on the analysis result. In all, user interaction data in online newsgroups can give us more comprehensive information about their mutual influence pattern.

In general, there are mainly two challenges to examine the effect of feedback with different opinions on continued user participation. First of all, we need to find out an effective and efficient method to classify user comments into different opinions. In this paper, instead of only focusing on the content, we leverage user behavior information and build user-like graph to do opinion classification. Besides that, we also use linguistic analysis tools to evaluate the classification results and develop this method into a real-time service on our SINCERE system[1]. The second challenge of this topic is to distinguish the influence of different feedback from other factors on continued user participation. In this paper, we perform a large scale study of user interaction on the official Occupy LA Facebook page which has a total of 20,569 unique users, 56,937 comments, 31,833 posts and 66,758 likes. Based on this dataset, we use the Cox proportional hazards model [5] in survival analysis to explore the relationship between a user's lifetime in an online newsgroup and several explanatory variables. Our final result shows that the content of feedback in an online newsgroup is significantly related to continued user participation: Not only the feedback with the same opinions as the user, but also the feedback with different opinions can motivate continued user participation. Furthermore, an interaction of feedback with both the same and different opinions can boost continued user participation to the greatest extent.

The rest of the paper is organized as follows. In section

[1] http://sincere.se/

ASONAM'13, August 25-29, 2013, Niagara, Ontario, CAN

II, we discuss related work on influences on continued user participation. Then we describe our opinion classification method and result evaluation in section III. Based on this method, we design an experiment to analyze the influence of different feedback on user continued participation in section IV. Discussion and conclusions are in section V and we also mention the limitations of our method and future work in section VI.

II. Related Work

Moira et al. [2] grouped the theories of user online participation into three high level categories: social learning, distribution and feedback. In this paper, we focus our work on examining the influence of feedback on user continued participation. In previous work, theories of reciprocity [6] and reinforcement both showed that feedback from other users should predict long-term participation. More specifically, feedback can be broken into the following three parts: the role of people who give the feedback, the amount of feedback and the content of feedback.

Regarding the role of people who give feedback, Steven Johnson [7] showed that interaction with online group leadership is associated with higher participation continuance and participation intensity. During the interaction with leadership in online group, the participant can feel psychological safety. Johnson indicated that members of online groups with higher levels of psychological safety report higher levels of continued participation intentions.

Furthermore, people also found positive relationship between the amount of received feedback and user continued participation. Previous work [3], [2] on newcomers in online newsgroups showed that newcomers will be more likely to post again if anyone responds to their initial post. Besides the work on newcomers in online groups, Y. Wang et al. [8] examined the factors that influence the continued participation of any member in online health support groups. Their results showed that the count of all comments in the threads in a week in which the user had posted is highly correlated (r=.67) with user's continued participation in online health support group.

When it comes to the content of feedback, there are many different ideas. Some previous work focused on newcomers in online newsgroups [3], [9]. E.Joyce [3] indicated that length, tone, content and personal affirmation are not found to be significant predictors of long-term engagement for newcomers in online newsgroup. However, Y. Wang et al. [8] examined the influence of feedback on all the group members and found a significant relationship between the content of feedback comments and user commitment in online health support groups. In this paper, our work also considers all the group members, but unlike Y. Wang's work, which analyzed different types of social supports among user interaction content, we focus on different feedback opinions in online newsgroups.

The influence of feedback with different opinions is widely discussed in democratic deliberation research in political communication [10], [11] and this issue is still controversial now. There are two different perspectives in this area. Some people believe that expressions of disagreement may violate expected norms of politeness in social interactions [12]. Mutz [13] proposed that the negative effects of disagreement may make people avoid political discussion and deliberation. However, J. Stromer Galley et al. [10] indicated that expressions of disagreement do not generally harm participants' future participation and an interaction of agreement and disagreement can even boost expected future participation in democratic deliberations. However, their work only focused on discussion of political topics and most of these studies used questionnaires or phone surveys as their data collection method, which severely limits their experiment sample. This paper, to the best of our knowledge, is the first work to give a large scale analysis of the influence of feedback with different opinions on continued user participation in online newsgroup. We offer a new opinion classification algorithm to partition feedback comments into different opinions automatically, which allows us to do analysis on a much larger dataset. Additionally, instead of only focusing on political topics, where people's stances are always sharp and in opposition to each other, our experiment broadens user interaction data to general discussion topics in online newsgroups.

III. Opinion Classification

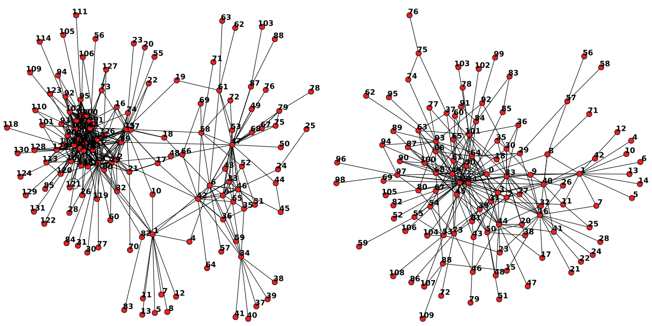
In the area of Natural Language Processing, there are many works discussing opinion classification in online threaded discussions [14], [15]. In the latest research work, Rob Abbott et al. [16] identified disagreement in political blogs using only lexical features. However, their machine learning method need a complex training process and every training model is only valuable for its source corpus. Instead of only focusing on the corpus itself, we can also leverage user behavior information in online newsgroups to assist us on opinion classification. Some recent work showed that user behavior features can be used to capture contextual information present in textual features very accurately [17]. Taking the public newsgroup on Facebook as an example, in addition to the text of user interaction under each post, the information of like behavior is also a valuable tool for us to recognize user's opinion. In this paper, we leverage the user-like-graph under each post to classify user interaction content into different opinions based on its link structure.

A. Data Collection

Our dataset is crawled from public pages on Facebook. These pages are open to the public and anyone with a Facebook account can post or comment on existing posts. To be specific, each public page is a tree-shaped structure in which multiple posts are organized in temporal ordering. Facebook users can share the latest news and discuss their opinions on different topics on those pages, which makes them ideal examples of online newsgroups. Under each post, we can get all of the user interaction information, such as the content of comments, likes information and their time-stamps. In general, our method leverages user like information on each comment to do opinion classification, which can be applied to analyze user interaction information on any Facebook public page in real time.

B. User-like Graph

As to each post in a Facebook public page, we define a *user-like* graph $G = (V, E)$, where V is a set of nodes and E is a set of edges among V. For simplicity, in this paper, we consider the *user-like* graph to be undirected. A node stands for a user who has liked a comment or whose comments were liked by others in this post. An edge e stands for the connection between two users u and v in this *user-like* graph and its weight w_{uv} equals to the number of likes they have clicked on each other's comments in this post. Figure 1 shows two examples of *user-like* graph of the posts on the official Occupy LA Facebook group.



(a) Post1 (b) Post2

Fig. 1: Examples of User-Like Graph

C. Opinion Classification method

The general idea of our opinion classification method is as follows: we first classify people into different groups based on its *user-like* graph of each post. Then we collect the comments made by different groups of people as different opinions contents. Therefore, as the first step, we need to find a partition of the *user-like* graph such that edges between different groups have a very low weight (which means people in different clusters are holding different opinions from each other) and the edges within a group have high weight (which means that people within the same cluster are holding similar opinions with each other). Obviously, this is a classic graph partitioning problem.

In order to make our output clusters reasonably large groups of nodes, we use the concept of Ratiocut [18], [19] to formulate our objective function. In the *user-like* graph $G = (V, E)$, we denote a subset of vertices $A \subset V$ and $\overline{A}$ for the complement of A. In addition, for two set $A, B \subset V$, we define

$$W(A, B) = \sum_{i \in A, j \in B} w_{ij} \tag{1}$$

For a given number k of subsets, we choose a partition $A_1, ..., A_k$ which minimizes

$$Partition(A_1, A_2, ..., A_k) = \frac{1}{2} \sum_{i=1}^{k} \frac{W(A_i, \overline{A_i})}{|A_i|} \tag{2}$$

In the objective function 2, the size of a subset A of a *user-like* graph is measured by its number of vertices, i.e. the amount of people in this subset. It will get a small value if the clusters A_i are not too small. Therefore, this objective function can make our output clusters balanced, measured by both the connections between each cluster and the number of their vertices. Unfortunately, introducing balancing conditions makes our partition problem become NP hard. In order to solve this objective function, we choose to use the spectral clustering algorithm [19], which is the most popular algorithm to solve relaxed versions of Ratiocut problem.

Based on our discussion above, we can now get a near-optimal partition of any *user-like* graph using the spectral clustering algorithm when the number of clusters k is fixed. Now we need to decide on the optimal number of cluster k for a *user-like* graph. In this paper, for simplicity, we only choose the cluster number from 1 to 2, i.e. we only want to decide if the comments under the same post can be clustered as two different opinions or people are just holding the same opinion in this post. In 2004, Newman and Girvan [20] proposed a modularity function which can directly measure the quality of a particular clustering of nodes in a graph. Their function Q measures the fraction of the edges in the graph that connect nodes in the same group minus the expected value of the same quantity in a graph with the same partitioning result but random connections between the nodes. If the sum weight of edges in the same groups is the same with that got by random connection, we will get $Q = 0$. And if the partitioning result has a strong community structure, the value of Q will be very close to 1 [20].

In our method, after getting the partitioning result using the spectral clustering algorithm with $k = 2$, we will use modularity Q to check the clustering quality of our partitioning result and decide if the number of clusters in this *user-like* graph is 1 or 2. Newman's work showed that real-world unweighted networks with high community structure generally have Q values within a range from 0.3 to 0.7. Figure 2 shows four examples of our clustering results and their corresponding Q values. Each of these graphs represents the user-like structure of one of the posts in Occupy LA Facebook group. The two different colors (pink and blue) stands for the spectral clustering result when we fix $k = 2$. We find that the Q value can be a good measurement for deciding the number of clusters in the user-like graph: The clustering results with $Q \geq 0.2$ show strong community structure, where k remains to be 2, while results with $Q < 0.2$ converge at some special nodes, where k is determined as 1. In this paper, we set $Q = 0.2$ as a threshold. Therefore, if the value of modularity measurement Q is less than 0.2 when $k = 2$, we will consider the user-like graph to be one cluster, or we accept its partitioning result and cluster the nodes into two different groups. As the last step, for each post, we consider the comments made by different clusters of people to be different opinions contents.

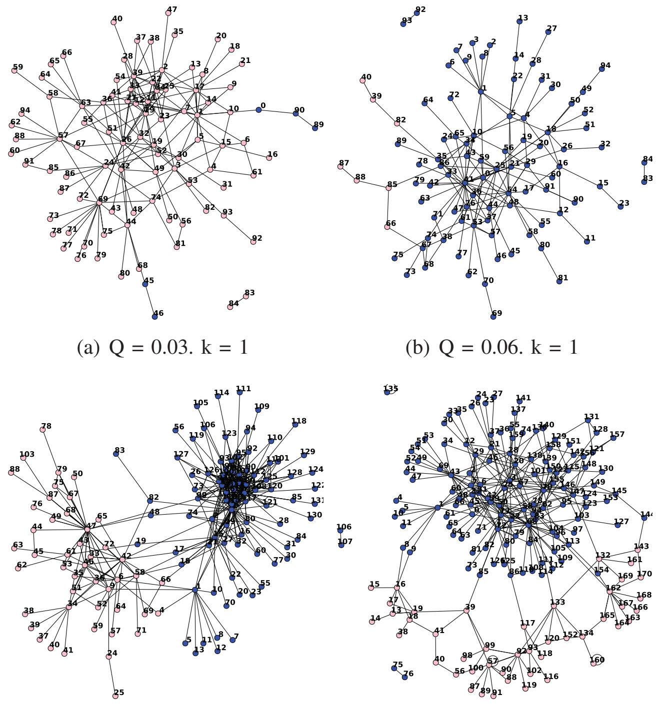
D. Evaluation from Linguistic Features

In order to evaluate the effectiveness of our opinion classification method, we select three Facebook public pages as our dataset: OccupyLA[2], Occupy Wall Street[3] and Occupy Together[4]. In Sept. 2011, the Occupy movement called for people to protest against social and economic inequality, which attracted a wide range of people all round the world. Online social networks, during this time, played an important role by offering people an ideal platform to get latest news and share opinions. Among all the public newsgroups about the Occupy movement on Facebook, the public pages of Occupy Wall Street and Occupy LA are the largest two groups. and OccupyTogether is a comprehensive public page where people

[2]http://www.facebook.com/OccupyLA

[3]http://www.facebook.com/OccupyWallSt

[4]http://www.facebook.com/OccupyTogether



(a) Q = 0.03. k = 1

(b) Q = 0.06. k = 1

(c) Q = 0.27. k = 2

(d) Q = 0.31. k = 2

Fig. 2: Examples of Clustering Results with Different Q Values

share information and opinions on any occupy movement. In order to examine user interaction content in each post, the number of comments in the post that we analyze should reach a certain amount. In this experiment, we select all the posts with more than 35 comments inside. At last, we collect a total of 205,198 comments on 1929 posts in these three public pages from Sept. 2011 to Apr. 2012. Because our dataset is very large, it is impossible for us to get the ground truth of the opinion classification result for each post by employing workers to rate each post manually. In this paper, we evaluate our opinion classification method from the linguistic features of user comments, which is a totally different view from the structure of *user-like* graph.

We use the Linguistic Inquiry and Word Count (LIWC) tool [21] to study the linguistic characteristic of user comments. LIWC is a popular tool which calculates the frequency with which words in a text match each of 68 categories representing linguistic dimensions, psychological constructs and personal concerns [8]. Many previous work [8], [22] have shown that the categories in LIWC are effective in determining linguistic differences on attentional focus and emotionality of the relationship. In this paper, we consider 39 categories in 13 areas, which is shown in Table I. Among these 39 categories, 32 of them belong to psychological processes and 7 of them belong to the personal concerns area. The analysis results in [8] and [16] reveal that the LIWC scores on psychological process are very helpful for identifying people's emotional attitudes, i.e. agreement and disagreement, in their online comments. Besides that, [17] also showed the LIWC score of categories on personal concerns is also effective at identifying people's different references in online groups.

Areas	**Categories**	**Selected Examples**
Social [1]		mate, talk, they, child
	Family	husband, aunt
	Friends	friend, neighbor
	Humans	adult, baby, boy
Affective [1]		happy, cried, abandon
	Pos. Emotion	love, nice, sweet
	Neg. Emotion	hurt, ugly, nasty
	Anxiety	fearful, nervous
	Anger	hate, kill, annoyed
	Sadness	crying, grief, sad
Cognitive [1]		cause, know, ought
	Insight	think, know, consider
	Causation	because, effect, hence
	Discrepancy	should, would, could
	Tentative	maybe, perhaps, guess
	Certainty	always, never
	Inhibition	block, constrain, stop
	Inclusive	and, with, include
	Exclusive	but, without, exclude
Perceptual[1]		heard, feeling
	See	View, saw, seen
	Hear	Listen, hearing
	Feel	Feels, touch
Biological[1]		eat, blood, pain
	Body	cheek, hands, spit
	Health	clinic, flu, pill
	Sexual	horny, love, incest
	Ingestion	dish, eat, pizza
Relativity[1]		area, bend, exit, stop
	Motion	arrive, car, go
	Space	down, in, thin
	Time	end, until, season
Work[2]		job, majors, xerox
Achievement[2]		earn, hero, win
Leisure[2]		cook, chat, movie
Home[2]		kitchen, family
Money[2]		audit, cash, owe
Religion[2]		altar, church, mosque
Death[2]		bury, coffin, kill

TABLE I: The 39 textual Categories in 13 areas used in our linguistic evaluation. Areas marked with [1] are psychological processes, and areas marked with [2] are personal concerns

During the evaluation process, our theoretical basis is: Comments on different opinions have different characteristics in their linguistic features [16]. After using our opinion classification method to analyze those 1929 posts, 1216 posts were classified as discussions with two different opinions. For each of these 1216 posts, we denote G as the set of comments of one post and $P, Q \subset G$ as the two groups of comments on the different opinions. Suppose there are m comments in set P and n comments in set Q. The LIWC analysis result of one comment is denoted as S_{ij}, where i denotes the ID of the comment in set P or Q and j denotes the serial number of the 39 categories. In set P, we define the LIWC scores in the ith category as $\mathbf{X}_{pj} = [S_{1j}, S_{2j}, ..., S_{nj}]^T$ $(j \in [1, 39])$. Similarly, in set Q, the LIWC scores in the ith category is defined as

$\mathbf{X}_{qj} = [S_{1j}, S_{2j}, ..., S_{mj}]^T$ $(j \in [1, 39])$. Therefore, the LIWC result of these two groups of comments can be denoted as follows:

$$L(P) = [\mathbf{X}_{p1}, \mathbf{X}_{p2}, ..., \mathbf{X}_{p39}] \tag{3}$$

$$L(Q) = [\mathbf{X}_{q1}, \mathbf{X}_{q2}, ..., \mathbf{X}_{q39}] \tag{4}$$

Then as to each of the 39 LIWC categories, we use Welch's t-test [23] to test whether the means of the two population $\mathbf{X}_{pj}$ and $\mathbf{X}_{qj}$ are different from each other. For LIWC category j, we denote the p-value of their t test as p_j $(j \in [1, 39])$. So we get the result of linguistic comparison as

$$Compare(P, Q) = [p_1, p_2, ..., p_{39}] \tag{5}$$

Because the posts in our dataset cover diversified topics, we cannot limit their different linguistic features show in only one particular item of those 39 categories. Therefore, as the final step, if any of these 39 p-values is less than the predetermined significance level $\alpha(= 0.05)$, we will conclude the two groups of comments obtained by our opinion classification method reveal different characteristics in their linguistic features, which indicates that our opinion classification result is acceptable for this post. Table II shows our evaluation results of all the 1216 posts in the three Facebook public pages. The item **Total** denotes the amount of posts with more than 35 comments in each online newsgroup. The item **Graph** denotes the amount of posts which are recognized as two opinion groups inside by our opinion classification method. The item **Linguistic** denotes the amount of posts within **Graph** which reveal different linguistic features. From the result we can see that 959 of those 1216 posts reveal different linguistic features, which achieves an accuracy of 78.9%.

Group	Total	Graph	Linguistic	Ratio
Occupy LA	380	264	200	75.8%
Occupy Together	451	285	230	80.7%
Occupy Wall Str	1098	667	529	79.3%

TABLE II: Evaluation Results from Linguistic Features

E. System Development

Besides theoretical evaluation, we also implemented this opinion classification method into our SINCERE system (Social Interactive Networking and Conversation Entropy Ranking Engine) as a real-time service. The SINCERE system is a diversified search engine based on user social informatics. Its database offers all the interactions (such as likes, shares, comments, timestamps) of 1391 Facebook public pages.

Figure 3 shows a screen shot of our opinion classification service on SINCERE. This system can automatically classify the comments of a post into different opinions (one or two opinions) and show the result in a pull-down menu. If there is only one opinion group recognized, the system will list all of the comments in one list. Otherwise, it will show the different groups of comments in two parallel columns, as shown in Figure 3. From the content of those classified comments in our example, we find that the opinion classification result is very effective: Under the post where people are discussing the recent actions of Los Angeles Police, the comments on the left (marked as green color) express support and kindness, while most of the comments on the right (marked as pink color) show a skeptical attitude.

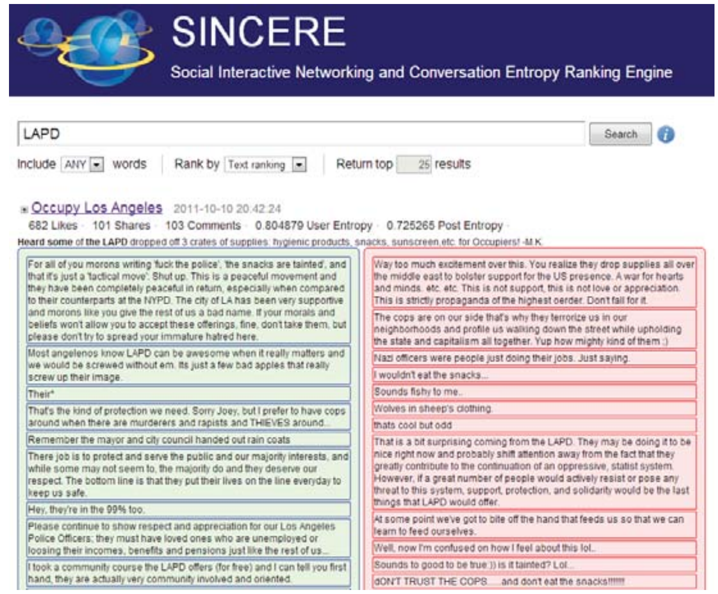


Fig. 3: Screen-Shot of Our Opinion Classification Service on SINCERE System

IV. The Influence of Feedback with Different Opinions on User Continued Participation

In section III, we introduced a new opinion classification method for online newsgroups and evaluated its effectiveness from linguistic features. In this section, we use this method as a tool to analyze the influence of feedback with different opinions on continued user participation in online newsgroups. The dataset we use in this experiment is crawled from the official Occupy LA Facebook public page. We collected all the posts, comments and like information on this page from Sep. 2011 to Apr. 2012. During this period of time, there were a total of 20,569 users who posted 56,937 comments on 31,833 posts. Additionally, there are also 66,758 likes among all of these comments. In order to analyze the influence of feedback on online newsgroup participation, the users we examine should have enough amount of activity records in this group. Therefore, in this experiment, we are only interested in those users who have made more than 20 comments on the Occupy LA public page, which includes 622 users in total.

We use the Cox proportional hazard (PH) model [24] in survival analysis [5] to explore the relationship between user lifetime in this newsgroup and several explanatory variables on the influence of feedback comments. Survival analysis is the main method to examine and model the time it takes for some special events to occur [24]. In our experiment, the specified event is defined as the end of the user's active lifetime on this page. In previous work, this technique has been widely used in medical science, sociology and engineering [9]. As the most widely used method of survival analysis, Cox regression can provide estimates of coefficients for each covariate and allow the assessment of the impact of multiple covariates in the same model.

A. Experiment Design

1) Dependent Variable:

- **Lifetime**: Because we are interested people's active lifetime on online newsgroups, user's first and last comment time may not be ideal indicators for the actual user lifetime in online newsgroup. [25] gave a definition on the lifetime of IRC channels based on their level of activity. In our experiment, we include user comment frequency into the definition of user lifetime. In our dataset, the total time duration is 220 days. First, we divide this period of time into 22 time-blocks with the same 10-day interval. Then we consider the start time of each participant when he/she makes more than 3 comments in two consecutive time-blocks. And we consider him/her to have left this newsgroup the day after when he/she gives no comments in two more consecutive time-blocks. Figure 4 shows the distribution of user lifetime of those selected 622 participants in Occupy Los Angeles public page. Their online lifetime ranges from 202 days to 0 days. In addition, because people whose last comment is found within the last time-block may still be participating in this Occupy LA group, we treat them as right censored in the survival analysis.

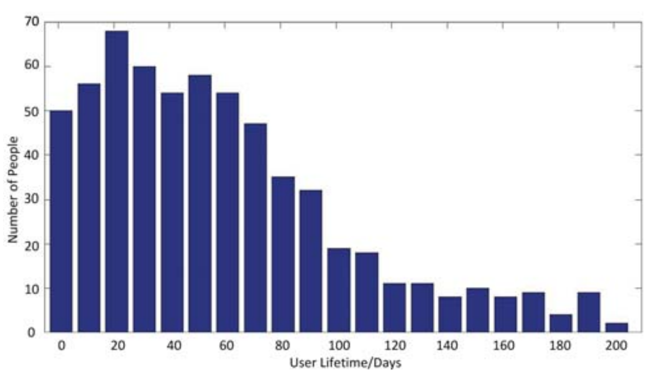


Fig. 4: Distribution of User Lifetime in Occupy Los Angeles Group

2) Control Variables:

- **OriginalPostWriter**: Among all the 31,833 posts in Occupy Los Angeles public page, most of them are written by the official maintainers of this public homepage named *OccupyLA*. However, there are also many posts written by normal users. Considering people who write original posts on this page may have different participating enthusiasm than people who just take part in discussions started by others, for each of those 622 members we are interested in, we define a control variable OriginalPostWriter as the percentage of his/her original posts among all of his comments in Occupy Los Angeles public page.
- **NonDiscussionPostsInvolved(NonDiscussion)**: Among all the posts written by the official maintainer of Occupy Los Angeles group, we also split them into two parts: the posts with more than 35 comments and those with less than 35 comments. The posts in the first part are considered discussion posts. Meanwhile, we use the posts in the second part to define a control variable *NonDiscussionPostsInvolve* as the number of an individual's comments in non-discussion posts divided by the amount of his/her comments in all posts.
- **ReceivedCommentsPerActivity (ComPerAct)**: In each discussion post in Occupy LA group, there is no exactly *reply* during user interaction: people just make comments one after another along the timeline of each post. So we define comments made within three hours after the individual makes a comment as his/her replies. We assume that a participant reads all comments within three hours after his/her comment and regard them as feedback for his/her activity. Based on this definition, we calculated the average number of received replies during the three hours after the individual makes a comment.

3) Independent Variables:

- **SameOpinionsPercentage(SamePercent)**: This variable measures the percentage of the replies in the same opinion with this person among all of the replies this person received. To be specific, during the three hours after this person makes a comment, all the received comments can be classified into three groups: comments with the same opinion with him/her in this post (these comments are made by people who are in the same opinion group in this post); comment with a different opinion with him/her in this post (these comments are made by people who are in a different opinion group in this post); comments with unclear opinion (these comment are made by people who neither clicked like on any other comments nor are liked by others).
- **DifferentOpinionsPercentage(DifferentPercent)**: This variable measures the percentage of the replies in a different opinion group from this person among all of the replies this person received.

Table III shows the descriptive statistics of all these variables. Based on the definitions above, we standardize all the control and independent variables with a mean of zero and standard deviation of one and use Cox regression model to analyze the relationship between the user's lifetime in Occupy LA group and these variables.

Variable	Min	Max	Mean	Std.Dev.
OriginalPostWriter	0	1	0.320	0.318
NonDiscussion	0	1	0.415	0.237
ComPerAct	0	57.625	5.679	6.900
SamePercent	0	1	0.375	0.283
DifferentPercent	0	0.8	0.110	0.133

TABLE III: Descriptive Statistics for Variables

B. Experiment Result

The results of the Cox regression model are shown in Table IV. The exponential coefficient indicates the direction of the effect of variables: when exp(coef) is smaller than 1, it represents a positive relationship between the variable and the lifetime. For example, because the exp(coef) of OriginalPost-Writer is 0.641 which is less than 1, we can say that when all other variables have average values, the more original posts one user writes, the longer the lifetime he/she will have in this newsgroup. Std. Err. indicates its standard error.

Control/Indepedent Variable	exp(coef)	Std. Err.
OriginalPostWriter	0.652***	0.123
NonDiscussion	0.723**	0.108
ComPerAct	0.770*	0.120
SamePercent	0.813**	0.074
DifferentPercent	0.875**	0.055
SamePercent **X** DifferentPercent	0.761**	0.087
ComPerAct **X** SamePercent	1.324***	0.071
ComPerAct **X** DifferentPercent	1.140*	0.062

*: p<0.05, **: p<0.01, ***: p<0.001

TABLE IV: Results of Cox Regression Model

Among the results of two control variables, the exp(coef) value (0.652) of OriginalPostWriter tells us that when all other variables are at their average values, the members who posted an average of one standard deviation (0.320) more original posts were 35%[5] more likely to remain in the group. Similarly, the exp(coef) value of NonDiscussionPostsInvolved indicates that members who are involved one standard deviation more in Non-Discussion Posts were 28% more likely to remain in the group. And those who received a standard deviation more feedback on his/her comment were revealed to be 23% more likely to remain in the group.

In addition, both of the independent variables, SamePercent and DifferentPercent show significant influence on user survival rate in social newsgroup. A member who received a standard deviation more feedback with the same opinion as him/her was 19% more likely to remain in the group and members who received a standard deviation more feedback with different opinions also showed a higher preference to remain in the group, with the rate of 13%. In other words, both positive and negative feedbacks from others can motivate users to participate in social newsgroups, and positive feedback has only slightly more driving effect.

Furthermore, we also consider the interaction between independent variables and those with the control variable ComPerAct. When controlling all of the control variables, the two independent variables SamePercent and DifferentPercent revealed super-linear positive influence on the user's lifetime: not only did each of them reveal positive influence on the lifetime, their interaction (SamePercent **X** DifferentPercent) showed a significant positive relationship with user longtime participation as well. In other words, compared to people who received an average number of positive feedback and negative feedback, members who received a standard deviation more feedback with both positive and negative feedback were 46%[6] more likely to remain in the group, which is much higher than their linear combination of 29%[7].

V. Discussion and Conclusions

In this paper, we built *user-like* graph to classify different opinions within user interaction content in online newsgroup. Then we used the Cox regression model to evaluate the influence of feedback with different opinions on user lifetime in official Occupy LA Facebook group. From the results shown above on different variables, we obtain many important conclusions which can help designers of social network systems to get a deeper understanding of user behavior and improve their online service.

Among the three control variables, firstly, the results show that members who started more original posts had a longer lifetime in this newsgroup. Therefore, designers of online social networks can motivate user participation by offering more opportunities for normal users to post their own news and become a discussion starter. Secondly, we also found that those who have more comments in Non-discussion posts preferred to stay in this newsgroup longer. Our explanation for this result is that although Non-discussion posts do not have enough comments to host a discussion environment, the information they offered is also very important. D. Fisher et al. [26] indicates that the topic of the forum is one of the factors that we can use to predict user engagement. Therefore, this result tells website designers that despite the importance of discussion posts with many comments and people involved, they should not neglect the information offered by Non-discussion posts. Last but not least, the result of control variable ComPerAct shows that the more feedback one individual receives after his/her comment, the more likely he/she will remain in this newsgroup.

When controlling all the control variables, our final result indicates that not only the number of replies, but also their content has a significant influence on a user's commitment to an online newsgroup. To be specific, our conclusion is that not only the feedback with the same opinions as the user, but also the feedback with different opinions can motivate a user to continue participating in an online newsgroup. Furthermore, an interaction of feedback with both the same and different opinions can boost continued user participation to the greatest extent. Based on our conclusion, we think that although feedback with different opinions may result in a unpleasant interaction, they help to form a comprehensive and healthy discussion environment. De Dreu et. al. [11] indicate that conflicts in group discussion can increase creativity and divergent thinking. Therefore, we believe that when people are involved in a discussion with several different perspectives, their understanding of certain topics may be improved, which will increase people's evaluation of this online newsgroup and motivate their future participation. This result tells website designers that they can also motivate user continued participation by digging into user interaction content. For example, website designers can highlight or send notices to a user when feedback from different opinions show up in his/her involved discussion posts.

VI. Limitations and Future Work

One of the limitations in this paper comes from our opinion classification method. Firstly, we modeled the *user-like* graph as an undirected graph which makes no discrimination between the writer of one comment and the likers of that comments. However, the degree of their preference on certain opinions may different from each other. Future work can model it as a directed graph so as to get a more accurate model of user opinions. Secondly, we also only choose limited number of opinion groups in one post to 1 or 2 and classified the person in *user-like* graph into one of the two opinion groups.

[5] $35\% = (1 - 0.652) * 100\%$

[6] 46% = (1-0.813*0.875*0.761)*100%

[7] 29% = (1-0.813*0.875)*100%

However, user interaction in online newsgroup is a much more complex scenario: Different from content in a debate [16] where people's opinions form two parties, the discussion in online newsgroup may consist of many different opinions, each of which starts from a different view and is not necessarily opposed to others. Therefore, in future work, we can consider more than two opinion-groups among user interaction content and assign various degrees to feedback comments instead of simply agreement and disagreement.

Furthermore, when we analyze the influence of feedback on continued user participation, we average the effect of the feedback comments during the whole user's lifetime in an online newsgroup. However, the influence of feedback may change during user participation in one group and different factors of feedback, such as the role of speaker, the amount of feedback and the content of feedback, may have a varied, dynamic influence pattern during user participation. Future work can analyze how those different factors change their influence during a user's lifetime in online newsgroup.

Acknowledgment

We want to thank Mohammad Rezaur Rahman and Roozbeh Nia for their help on data collection. We are also very grateful to Ran Cao in Statistics Department of Michigan State University for her insightful comments on the survival analysis part. This paper is based upon work supported in parts by the National Science Foundation under Grant No. CNS-1152320, IIP-1161015, and CNS-0832202, Army Research Office under the Multi-University Research Initiative (MURI) grant W911NF-07-1-0318.

References

[1] N. O'Neill, "Google now indexes 620 million facebook groups," http://allfacebook.com/google-now-indexes-620-million-facebook-groups_b10520, Feb. 2010.

[2] M. Burke, C. Marlow, and T. Lento, "Feed me: motivating newcomer contribution in social network sites," in *Proceedings of the 27th international conference on Human factors in computing systems*. ACM, 2009, pp. 945–954.

[3] E. Joyce and R. Kraut, "Predicting continued participation in newsgroups," *Journal of Computer-Mediated Communication*, vol. 11, no. 3, pp. 723–747, 2006.

[4] S. Johnson, *Impact of Leadership on continued participation in online groups*. ProQuest, 2008.

[5] D. R. Cox and D. Oakes, *Analysis of survival data*. Chapman & Hall/CRC, 1984, vol. 21.

[6] A. W. Gouldner, "The norm of reciprocity: A preliminary statement," *American sociological review*, pp. 161–178, 1960.

[7] S. Johnson, "Should i stay or should i go? continued participation intentions in online communities," in *Continued Participation Intentions in Online Communities (September 1, 2010). PROCEEDINGS OF ACADEMY OF MANAGEMENT ANNUAL CONFERENCE, Leslie A. Toombs, ed*, 2010.

[8] Y. Wang, R. Kraut, and J. Levine, "To stay or leave? the relationship of emotional and informational support to commitment in online health support groups," in *Proceedings of the ACM conference on computer-supported cooperative work*, 2011.

[9] J. Yang, X. Wei, M. Ackerman, and L. Adamic, "Activity lifespan: An analysis of user survival patterns in online knowledge sharing communities," *Proceeding of ICWSM*, 2010.

[10] J. Stromer-Galley and P. Muhlberger, "Agreement and disagreement in group deliberation: Effects on deliberation satisfaction, future engagement, and decision legitimacy," *Political Communication*, vol. 26, no. 2, pp. 173–192, 2009.

[11] C. K. De Dreu and M. A. West, "Minority dissent and team innovation: the importance of participation in decision making." *Journal of applied Psychology*, vol. 86, no. 6, p. 1191, 2001.

[12] N. Eliasoph, *Avoiding politics: How Americans produce apathy in everyday life*. Cambridge University Press, 1998.

[13] D. C. Mutz, *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press, 2006.

[14] J. Andreas, S. Rosenthal, and K. McKeown, "Annotating agreement and disagreement in threaded discussion," in *Proceedings of the 8th International Conference on Language Resources and Computation (LREC), Istanbul, Turkey, May*, 2012.

[15] S. Germesin and T. Wilson, "Agreement detection in multiparty conversation," in *Proceedings of the 2009 international conference on Multimodal interfaces*. ACM, 2009, pp. 7–14.

[16] R. Abbott, M. Walker, P. Anand, J. E. Fox Tree, R. Bowmani, and J. King, "How can you say such things?!?: Recognizing disagreement in informal political argument," in *Proceedings of the Workshop on Languages in Social Media*. Association for Computational Linguistics, 2011, pp. 2–11.

[17] S. Adali, F. Sisenda, and M. Magdon-Ismail, "Actions speak as loud as words: predicting relationships from social behavior data," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 689–698.

[18] L. Hagen and A. B. Kahng, "New spectral methods for ratio cut partitioning and clustering," *Computer-aided design of integrated circuits and systems, ieee transactions on*, vol. 11, no. 9, pp. 1074–1085, 1992.

[19] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[20] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 2, p. 026113, 2004.

[21] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods," *Journal of Language and Social Psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[22] M. A. Walker, P. Anand, R. Abbott, J. E. F. Tree, C. Martelly, and J. King, "That's your evidence?: Classifying stance in online political debate," *Decision Support Systems*, 2012.

[23] B. L. Welch, "The generalization ofstudent's' problem when several different population variances are involved," *Biometrika*, vol. 34, no. 1/2, pp. 28–35, 1947.

[24] J. Fox, "Cox proportional-hazards regression for survival data," *An R and S-PLUS companion to applied regression*, pp. 1–18, 2002.

[25] D. R. Raban, M. Moldovan, and Q. Jones, "An empirical study of critical mass and online community survival," in *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. ACM, 2010, pp. 71–80.

[26] D. Fisher, M. Smith, and H. T. Welser, "You are who you talk to: Detecting roles in usenet newsgroups," in *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on*, vol. 3. IEEE, 2006, pp. 59b–59b.