How to Become a More Popular Thread on Tianya Club

Jindong Chen
Academy of Mathematics and
Systems Science, Chinese Academy
of Sciences, Beijing, 100190 P.R.
China
86-10-62651427
j.chen@amss.ac.cn

Lina Cao
Academy of Mathematics and
Systems Science, Chinese Academy
of Sciences, Beijing, 100190 P.R.
China
86-10-62651427
caolina@amss.ac.cn

Xijin Tang
Academy of Mathematics and
Systems Science, Chinese Academy
of Sciences, Beijing, 100190 P.R.
China
86-10-82541801
xjtang@iss.ac.cn

ABSTRACT

In Bulletin Board System, when an author submits a post to start a thread, what kinds of elements affect the popularity of this thread? To address this issue, Tianya Zatan board of Tianya Club is selected as data source for this study. Through database and online data, the daily new threads with high degree of online interactions in 2012 are collected for analysis. To explore the key elements relating to the popularity of thread, several elements are analyzed using different methods: the topics of original posts are based on Latent Dirichlet Allocation, the high frequency words of titles, the basic features (the length of titles, the titles with/without the information of picture contained in the original posts, and the length of the content of original posts) and reply modes are based on statistics. Finally, several practical tips for an author to promote the popularity of the thread are given.

Categories and Subject Descriptors

H.4.3 [**Information Systems**]: Information Systems Applications Communications Applications; H.2.8 [**Information Systems**]: Database Management Database Applications; J.4 [**Computer Applications**]: Social and behavioral sciences

General Terms

Human Factors

Keywords

Latent Dirichlet Allocation, Human Dynamics, Hot Thread, Tianya Club

1. INTRODUCTION

Bulletin Board System (BBS) plays an important role in Chinese Internet life, different cultures and Internet policies make the Chinese BBS sites more popular and important [1]. The Chinese BBS sites are an open online forum, which can be visited by any netizens and replied by registered users. To attract more users online, BBS improves the services and develops more applications for users. Hence, BBS can be treated as an electronic information center and emerging media, and fulfills the netizens' requirements of being informed and exchanging of opinions [2]. The appearance of Micro-blog, WeChat, etc., the influence of BBS is

declining, but it is still an important social media in China. Tianya Club is a representative Chinese BBS site, which provides BBS, blog, microblog and photo album services etc. [3]. Specifically, Tianya Zatan board is one of the most active boards within Tianya Club. Everyday more than 1000 new threads are published, the clicks and replies are more than millions, and the content covers many fields of the current society [4].

Previous BBS studies aim at management or monitoring the evolution of BBS [3-4]. Up to date, few researches aim at helping netizens to construct a more popular thread, which is useful for netizens to express their opinions to broader areas. To study this issue, online questionnaire survey is a normal method in previous [5]. However, due to the limitation of data size and the questionnaire survey method, it only reveals part of BBS, and is unsuited for common netizen to start a more popular thread.

Through automatic analyzing user generated content (UGC) techniques, this study aims to find several practical tips in a more general scope on Tianya Zatan broad for common netizens to create a more popular thread. Based on Tianya Club database and online data, the daily new threads with more online interactions of Tianya Zatan broad of 2012 are collected. With the assistance of Latent Dirichlet Allocation method [4], statistical method, the main elements of the original posts and the reply modes of the hot threads are analyzed in this study. Through this study, we want to answer three questions: I) what kinds of topics trigger netizens to click or reply to a thread; II) the influences of the basic features of original post to the popularity of thread; III) what kinds of reply modes improve the popularity of thread.

The rest of this paper is organized as follows. Section 2 presents the procedure of data collection and processing. The analysis, results and discussions are shown in Section 3. Finally, conclusion is given in Section 4.

2. DATA COLLECTION AND PROCESSING

To explore what kinds of elements of a thread invoking more online interaction, the 2012 daily new threads of Tianya Zatan board are selected, 409,717 daily new threads are collected. Through the method of reference [6], it can be found that the identification of a thread whether with more influence mainly relies on: clicks, replies, the ratio of replies and clicks. The hot degree of a thread is evaluated by Eq. (1) in this study.

$$HD_i = w_1 h_i/(avg(h)) + w_2 r_i/(avg(r)) + w_3 (r_i/h_i)/(max(r/h)) \qquad (1)$$

where i represents the i-th thread; h_i is the clicks of the i-th thread; r_i is the replies of the i-th thread; w_j (j=1,2,3) is the weight and $\sum_{j=1}^{3} w_j = 1$; avg represents the average value of all threads; *max*

represents the maximum value in all threads are chosen. The daily top 20 hottest threads are selected, 7320 hot threads of 2012 are collected for this study.

Aiming to extract the reply modes of these hot threads, using R package "RCurl", the repliers and corresponding reply time of all the threads in the first day are crawled online. Due to the deleting and other factors of Tianya Club, only 6528 threads can still be found in Tianya Zatan broad. Therefore, in reply mode analysis, this data set is used.

3. ANALYSIS, RESULTS AND DISCUSSIONS

To extract some practical tips for creating a more popular thread, several elements are analyzed in this section, including: the main topics of all original posts, the high frequency words of titles, the basic features in the original posts, and the reply modes.

3.1 The Main Topics Extraction of Original Posts

To exploring what kinds of topics mainly discussed by netizens, the main topics of original posts are extracted based on Latent Dirichlet Allocation (LDA) method. The processing steps of topics extraction using LDA are:

1) Original posts with urls but no contents are removed.
2) The title of each post is repeated and then added into the content of each original post. The contents are segmented to words using ICTCLAS, Baidu hot search news words are adopted as reserved words [4].
3) After segmentation, nouns, gerunds, adnoun and personal names are selected into total corpus to analysis. The number of original posts is 7247 and total number of corpus is 1,122,341.
4) Terms that occur fewer than 30 times in corpus and in fewer than 10 original posts are removed. The number of terms in the data set is 4493.

After the data processing, LDA model is applied to extract the main topics of all threads; the number of topics is set to 50. According to the posterior inference, the distribution of topics in each original post is obtained. Let θ_{ij} denotes the mixed distribution of original post j in topic i, hot_j is the hot degree of thread j. We define Hot_{topici} as the hot degree of each topic, which can be calculated through Eq. (2),

$$Hot_{topici} = \sum_{j=1}^{m} \theta_{ij} \cdot hot_j \qquad (2)$$

According to the value of Hot_{topici}, some hottest topics are presented in Table 1.

Due to the space limitation of this paper, only 10 topics are listed. At the topic level, some top words within each topic are selected.

Table 1. The top 10 hottest topics of all original posts

The topics	Top words in the topics
The issues of company	公司(Company),有限公司(Co. Ltd.), 集团(Group), 企业(Enterprise), 合同(Contract), 工程(Project), 投资(Investment),资产(Asset),资金(Funds),员工(Employee), 经营(Management), 股东(Stockholder), 国有(Nationalized),合作(Cooperation),股权(Shareholding)
Violent demolition	村民(Villager), 土地(Land), 强拆(Violent demolition), 政府(Government), 房屋(House), 房子(House), 补偿

	(Compensation), 暴力(Violence), 人员(Crew), 百姓(Civilian),情况(Situation),开发商(Firm),领导(Leader),村里(Village),违法(Illegal),法律(Law),耕地(Farmland)
Criminal cases	家属(Family), 派出所(Local police station), 死者(Decedent),民警(Policeman),公安(Police),犯罪(Offense), 嫌疑人(Suspect), 凶手(Criminal), 人员(Crew), 尸体(Corpse), 案件(Case), 受害人(Victim), 调查(Investigation), 情况(Situation), 医院(Hospital), 机关(Office),公安局(Police station), 所长(Director)
Wu Ying case lawsuit	法院(Court),证据(Proof),判决(Judgment),律师(Lawyer), 法官(Judge),法律(law),案件(case),吴英(WuYing),被告人(Accused man), 法庭(Court), 司法(Justice), 检察院(Procuratorate), 被害人(Victim), 被告(Accused), 犯罪(Crime),死刑(Capital punishment),行为(Behavior),诉讼(Lawsuit),一审(First instance)
Governmental officials	书记(Clerk),干部(Cadre),工作(Job),主任(Director),副市长(Deputy mayor),常委(Member of standing committee), 副书记(Deputy clerk),领导(Leader),党员(Party member),
Economic development	政府(Government),建设(Construction),业主(Owner),工程(Project), 建筑(Building), 部门(Department), 投资(Investment),城市(City),小区(Sub-district),规划(Plan),中心(Center),地方(local),居民(Resident),经济(Economy)
Tourism disputes	朋友(Friend),时候(When),手机(Phone),电话(Call),事情(Affair), 东西(Thing), 小时(Hour), 酒店(Hotel), 地方(Local),时间(Time),工作(Work),同事(Colleague),保安(Security),游客(Tourist),老板(Boss),结果(Result)
Politics	人民(Civilian),政治(Politics),毛泽东(Mao Zedong),主席(Chairman), 国家(Nation), 社会主义(Socialism), 发展(Development), 革命(Revolution), 制度(Regime), 改革(Reform),领导(Leader),权力(Power),共产党(Communist party),社会(Society)
Media investigation	记者(journalist),事件(Event),微博(Weibo),媒体(Media), 视频(Video), 调查(Investigation), 真相(Truth), 新闻(News),官方(Official),网友(Netizen),消息(Message),公众(Public),部门(Department),结果(Result),
Several famous lawsuits	法律(Law),社会(Society),行为(Behavior),犯罪(Crime), 司法(Justice),案件(Case),机关(Official),王立军(Wang Lijun),法治(Rule by law),唐慧(Tang Hui),违法(Illegal), 公民(Civilian), 公安(Police), 规定(Regulation), 权利(Right),受害人(Victim),幼女(Young Girl)

From Table 1, it can be found that these topics are relevant to the issues or events mainly concerned by the current society: such as famous lawsuits, government-related issues and public ethics and morals etc. The reasons of this phenomenon can be drawn as: I) these topics usually contain controversial points, such as "Wu Ying case" (A famous illegal fund-raising case in Zhejiang, China), the controversial point is whether to sentence Wu Ying to death; II) netizens are informed of the current hot events through official reports, but the vagueness and distrust of official reports lead netizens to find another way to get more information; III) netizens click the related threads on BBS and know more information about these events, which are more likely to drive netizens to join the discussion. Therefore, the threads related to the current hot events become more popular. If an author wants to attract more netizens to click or reply his thread, the topic selection in the issues or events mainly concerned by the current society will be a better choice, like the journals of Tianya Zatan broad, which focus on criticizing the hot events timely.

3.2 The High Frequency Words Analysis of the Original Posts' Titles

On Tianya Zantan board, the title of original post is the main information directly present to netizens. As previous described,

the main motivation of netizens to visit BBS is to look for more details about the current controversial and hot events, so the relative words of these events in titles should attract more online interactions. To verify this inference, the high frequency words of the original posts' titles are analyzed.

To find out what kinds of words can attract more clicks, the high frequency words of all titles are analyzed. All the titles are segmented into words using ICTCLAS; the word frequency is bigger than or equal to 20 are selected. Furthermore, words means amount are deleted, such as "一次(once)", "一场 (one scene)", etc. In the end, 292 words are selected from 7320 titles.
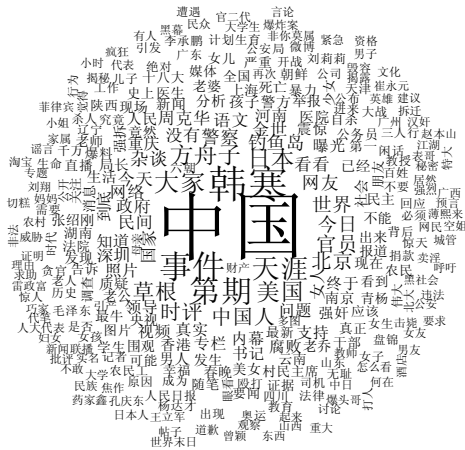

Figure 1. The word cloud of high frequency words in titles

As described in Figure 1, the most of words are mainly relating to the hot events which happened in 2012, such as "方韩大战(The online quarrel between Fang Zhouzi and Han Han)", "钓鱼岛事件 (Diaoyu islands)", "周克华事件(A homicide and robbery criminal was shot down by Chongqing polices)", "杨达才事件 (A governmental official in Shaanxi smiled at an accident spot, which infuriated netizens, then netizens found the corruption evidence of Yang through his watch)", and several journal titles of Tianya Zatan board, etc.

3.3 The Basic Features of Original Post
Several basic features of original post are analyzed in this section, including: the length distribution of titles, titles with/without the information of picture contained in the original post, and the length distribution of the contents of the original posts.

● The length distribution of titles
Title is the main information to trigger netizens to click, so only the click is considered for analyzing the features of titles. Aiming to find relationship between the length of title and click, the length distribution of titles and the average clicks of different length of titles are analyzed. The results are presented in Figure 2.

As shown in Figure 2, the lengths of titles are mainly between 11 and 28 chars, among a total of 5742 threads. It also can be found that the length of title between 8 chars to 19 chars received fewer clicks. Hence it can be said that a long title may attract more clicks; and the short title also can receive more clicks if less than or equal to 7 chars. But now, due to the modification of Tianya Club, the length of title on Tianya Zatan board is limited as 34 chars.

● Title with/without the information of picture contained in the original post

Only two key words are used to identify the title whether shows the information of picture contained in the original post: "图 (picture)", "照片(photo)", 491 threads are indexed in the hot thread data set. The average clicks of the original posts with picture is 27081.08, variance is 0.5*E10; the average clicks of the original posts with no picture is 24209.04, variance is 1.7*E10. Hence, the title with the information of picture contained in the original post will attract more clicks.
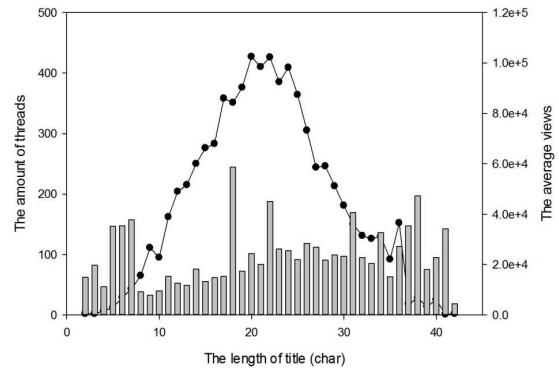

Figure 2. The length distribution of original posts' titles

● The length analysis of the contents of all original posts
The content of original post mainly affects netizens to reply, so only the reply is considered for analyzing the features of contents. The length distribution of the contents of all original posts and the corresponding average replies are counted. As can be found, although the shortest length of the content of the original post of Tianya Zatan board is 200 chars, the length less than 200 chars still can be found, the main factors includes: I) the original post contain pictures; II) the problem of Tianya Club.
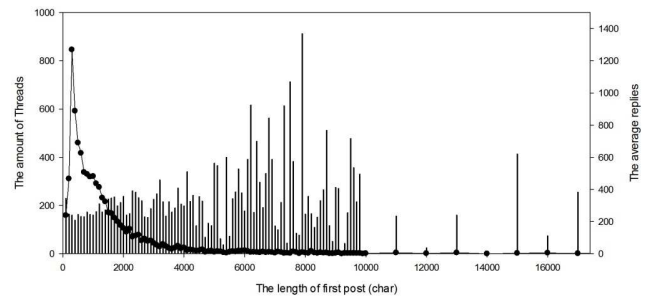

Figure 3. The length distribution of the content of all original posts and corresponding average replies

From Figure 3, it also can be found that the amount of threads which the length of the contents of original post is less than 4000 chars is 6916 (94%). From the average replies analysis, no obvious tendency is found. The correlative coefficient between replies and length is 0.014, which is not shown much relationship between replies and the length of the original post. From our previous analysis, the length distribution of the contents of the original posts of daily new threads is also similar as the profile of hot threads [1]. In general, the length of original post will affect the popularity of the thread, but it is not a main factor.

3.4 The Reply Mode Analysis
To evaluate the impact of reply mode, the reply modes of the available hot threads are analyzed in this section.

● The participating rate distribution of author

To measure the influence of author in the reply, the replying rate of author after submitting the original post is calculated. The rates are calculated in two ways: I) in the first three hour, the rates are calculated for each hour; II) the left are measured by average. The results are present in Figure 4.

As can be found in Figure 4, more than half of threads are with no replies of author after submitting the original post, and the amount of threads most replied by authors (rate >0.6) is very small (<400). From this point, it can be said that a thread hardly evoke more online interactions only relying on the replies of author. Two points can be drawn to this issue: I) a thread is mostly replied by author, which means the topic of this thread is not popular, which limits the influence of the thread; II) when this kind of thread is found by the BBS moderate, the thread will be forbidden to reply. As this reason, the author can not reply to his thread unlimitedly to keep it on the first page of BBS. Therefore, the thread most replied by author is difficult to become the hot thread.
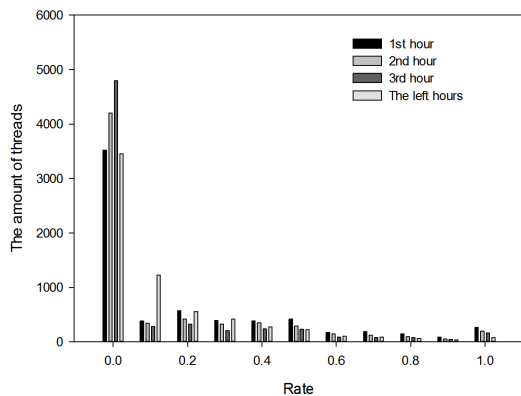

Figure 4. The participating rate distribution of author

●   The distribution of the rate of participant and his replies
The distribution of the rate of participant and his replies of all threads are presented in Figure 5. The statistical method is same as the participating rate distribution of author .
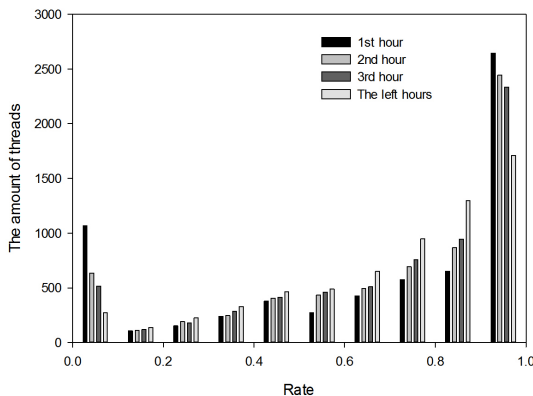

Figure 5. The distribution of the rate of participants and his replies of all threads

As can be found in Figure 5, in the first hour, most hot threads concentrate at (rate<0.1) and (rate>0.9), the amount of threads are 1070 and 2646 respectively. The rate<0.1 means one netizen reply more than ten times in one hour. Through the statistical results, the average replies in first hour are 13. From this point, the rate<0.1 means the replies are almost operated by one person. Meanwhile, the rate>0.9 means one netizen only reply once in one

hour, and almost no discussion, it is also abnormal. Through these two phenomena, it can be found more than half of hot threads are operated manually.

After the first hour, the amount of threads in (rate<0.1) and (rate>0.9) is declined, the amount of threads in other intervals are increased. As can be found, through the abnormal reply mode, the thread have more time to stay on the first page of BBS, which can be clicked and replied by more netizens, and evoke more online interactions. Therefore, the manual operative reply mode is important for a thread to evoke more online interaction.

4. CONCLUSION

In this paper, we mainly analyzed the four elements of the daily new hot threads of 2012 on Tianya Zatan board, several practical tips are drawn to help netizen to start a more popular thread:

1)   The main factor trigger netizens to click and reply to the thread is that netizens want to find another channel to know the more details of current hot events; hence the proper topics for creating a more popular thread should be related to the issues mainly concerned by current society. The high frequency words of titles also confirm this conclusion.
2)   The long title with the information of picture contained in the original post show positive effect on the spread of the thread, and the length of the content of the original post shows unobvious impact on the popularity of thread.
3)   For the reply mode, a hot thread hardly relies on the replies of author, and manual operative replies will bring more online interaction to the thread.

REFERENCES
[1]   Chen, J.D., and Tang, X.J. 2013.Exploration on Posts of Tianya Club and Preliminary Results. *The 14th International Symposium on Knowledge and Systems sciences*, (Ningbo, P.R.China., Oct. 25-27, 2013). JAIST Press, 167-174.

[2]   Jin, X.L., Cheung, C.M., Lee, M.K., and Chen, H.P. 2009. How to keep members using the information in a computer-supported social network. *Comput. Hum. Behav.* 25, 5 (Sep. 2009), 1172-1181. DOI= http://dx.doi.org/10.1016/j.chb.2009.04.008.

[3]   Tang, X.J. 2013. Exploring On-line Societal Risk Perception for Harmonious Society Measurement. *J. Syst. Sci. Syst. Eng.*, 22, 4 (Dec. 2013), 469-486. DOI=10.1007/s11518-013-5238-1

[4]   Cao, L.N., Tang, X. J. 2014. Topics and Trends of the On-line Public Concerns Based on Tianya Forum. *J. Sys. Sci. Sys. Eng.*, 23, 2 (Jun. 2014), 212-230. DOI=10.1007/s11518-014-5243-z

[5]   Tang, Y.Y., Kong, K.Q., Song, Y., and You, X.Q., 2008. Interpersonal perception in bulletin board systems among Chinese Internet users. *Comput. Hum. Behav. 24*, 5 (Sep. 2008), 2043-2054. DOI= http://dx.doi.org/10.1016/j.chb.2007.09.006.

[6]   Cui, L.J., He, H., and Liu W. 2013. Research on Hot Issues and Evolutionary Trends in Network Forums. *Int. J. U-& E-Serv., Sci. Tech.*, 6, 2 (Apr. 2013), 89-98.