

Observations on Failure in Blogs

Vahed Qazvinian
Department of Computer
Engineering
Sharif University of
Technology
Tehran, Iran
qazvinian@ce.sharif.edu

Abtin Rassoulia
Department of Computer
Engineering
Sharif University of
Technology
Tehran, Iran
rassolian@ce.sharif.edu

Jafar Adibi
Information Science Institute
University of Southern
California
USA
adibi@isi.edu

Abstract

Blogs form a large social network, and their analysis are becoming an important research area today. Blogs are growing rapidly in the Internet, because bloggers can rapidly change the content and linking patterns of them. Visitors of blogs may comment on the postings of a blog, and this leads to a complex interaction between groups of bloggers. One of the interesting phenomenon in blog space is “blogger failure” when a blogger stops writing after a certain amount of time and will not return to blogspace for a long time, or when a blogger does not get any comment from her audience.

In this paper we illustrate our observation on bloggers failure in a unique blogspace. First, we introduce, PersianBlog blogspace and dataset, and then we will describe our observations in commenting behaviors of bloggers. Finally, we will provide our definition of failure, and give a broad future research path to bring out a model for this phenomenon.

Keywords

Failure Model, Failure in Blogspace, commitment failure, connection failure

1. Introduction

Web, by design is distributed and heterogenous. Alongside, Components of the new web, known as web 2.0, have given it a more dynamic structure. In fact, web seems to be exploding by weblogs (or shortly blogs [10]), as the bloggers talk on the internet about their daily happenings, from writing diaries to political discussions. Many of us are involved in blogs these days, as the number of blogs is more than 9 million with a new blog, created every 7.4 seconds [12].

Blogs are indeed, forming a large social network. The importance of studying this network is so clear that Secko [12] describes the opportunity of reading one as:

Blogging is a form of communication that is sweeping through business, and although it's yet to significantly break into the pharmaceutical and biotech industries, few believe it's going to stop at their gates. So, if you're not reading one, or better yet, writing one, you're missing the opportunities others are taking advantage of.

In the blog social network, bloggers have relationships with others, through links and comments. Any blogger usually have a sort of friends, and a regular number of readers, which tend to comment on her postings.

Despite all other works on blogspahres, the blogspace is so complex that it has still unidentified mysteries behind. One of these facts is the phenomenon of “Blogger Failure”. As, to our knowledge, there is no previous work on this concept, different interpretations of failure is expected. However our understanding of failure focuses on two different types of failures:

First, suppose a blogger loses her visitors gradually, until she has no visitors at all, and gets no comments on her postings. Indeed, Connection of this blog shrinks over time. Of course this blogger might continue writing posts, even when she does not get comments. This is the case, when the blogger only writes regardless of the response of the network. This type of blogging is not considered in our study, because, in our definition of blog, the identity of a blog is defined by its interactions, and its role in a larger network, say blogspere. We will call this type of failure as *Connection-failure*.

Second, Suppose a blogger that starts a blog and *quits* writing after a while. We also consider this situation as failure. There could be reasons for quitting a blog. From now on we will call this type of failure as *Commitment-failure*

In this paper our aim is to give a brief observation on the phenomenon of *Failure*. The definition and the framework for *failure* can be used successfully in future attempts to devise a model for the phenomenon.

2. Data Corpus

Persian weblogs are growing very fast, as stated in [1] by October 2005, Persian weblogs were estimated to be about 700,000 (out of an estimated total value of 100 million blogs worldwide), of which about 40,000-110,000 are active, mostly written in Persian language. We chose persian blogspaces to collect our data.

2.1 Corpus Dimension

In this section we will overview our data, on which we did the analysis. We have previously described data in [11], yet we will give a brief notification on characteristics of data here. The data is gathered by crawling a Persian blog host.¹ As stated in [11], the data, contains archives of more than 22,000 weblogs in a 15 months period. There are nearly 347,800 posts with a total number of 1,258,000 comments, which to

¹ www.persianblog.com

Table 1: Basic analysis on corpus size

Weblogs	22,306
Posts	348,700
Posts with Invalid IDs	492
Correct Comments	1,257,561
Incomplete Comments	89,349
unavailable Comments	385
Accuracy	93%
Commented Posts	339,884 (97.5%)
Uncommented Posts	8,816 (2.5%)

an average of 3.6 comments per post. This average is comparable to that of previous corpus [6, 9], for which the average value was 0.3, 0.9 respectively. Table 1 shows some other statistics of our collected dataset.

Our data consists of almost 80,000 XML files, each for monthly archive of every blog. Related information about data is available at [2].

2.2 Different Links

We observed that there are four different types of transactions in persian weblogs which will further form a link. Accordingly, we modeled the data with the concept of typed graphs [11], having four types:

- *Blog Roll Link* which is a hyperlink put in the side bar of blog page. These hyperlinks usually link to blogs or homepages of the blog maintainer’s friends, and are not updated very often.
- *Post Link* also we call it an *entry-to-entry* link, is a hyperlink put in the content (body) of a post. It might probably point to a related material or post, what so ever.
- *Comment outLink* is a hyperlink put in the content (body) of a comment.
- *Comment inLink* is a hyperlink put in the footer of a comment and points to the blog, homepage or email address of the comment leaver.

The number of *comment outlinks* are negligible in comparison with other types of links. The Venn diagram in Fig.1 shows the distribution of different links. It is clear that *comment inLinks* with nearly 70% of all links in corpus have a noticeable contribution in formation of the blogspace graph.

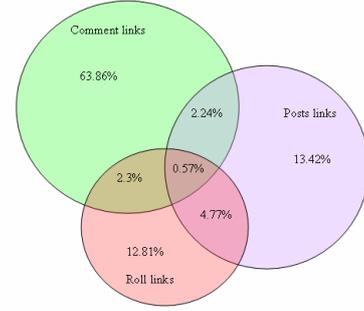
This shows that, comment graph, with a negligible shared links with other graphs, is a good test bench for further analysis. With the clearance of comment graph importance, we based our observations and definitions on *comment inLink* graph.

3. Justification of Failure and Its Background

In this section we will give our understanding and definition of *failure* in blogspaces. Before, some background information could be valuable.

3.1 Properties of The Comment Graph

As stated in previous section, comments are important part of our data. Not only in that, they form a large portion of data, but also because they show daily interactions between bloggers. Blogroll links form a quite static network,

**Fig. 1:** Venn Diagram for different links distribution

and Post Links, do not have the contribution of Comment inLinks. Trevino et al. [14] and Gumbrecht et al. [5] show the importance of comments in blog analysis, and conclude that comments play a essential role in interactive nature of blogs. So we focused on Comment inLink graph to define an Connection-failure phenomenon.

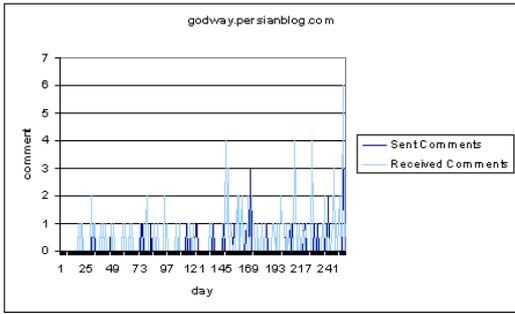
Here will be a set of properties which we believe any *commentspace* should convince, along with our reasons. The set of properties are divided into 5 categories, based upon a blogger’s interactions to collect more comments.

Category I. Two (or more) bloggers in a limited period of time, involve in discussions, Blogger *A*, comments on a posting of Blogger *B*, and subsequently *B* comments on a posting of *A*. This is even sometimes as an acknowledgment message or a reply for a comment. These discussions are usually with friends . If a blogger only gets comments of this type, she will lose his comments, if he delays writing posts, or does not tend to reply or leave comments. godway.persianblog.com is a clear example of such blogs. As you see in Fig 2 (a), godway gets comments on the time she leaves comments on his friends’ postings.

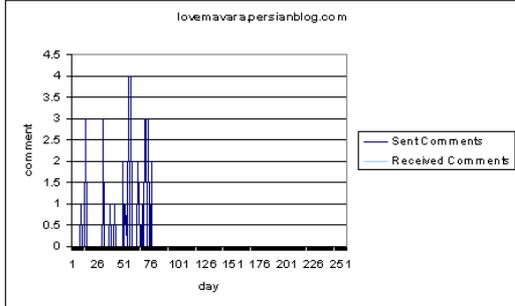
Category II. Here a blog at his origin, puts comments for other bloggers, but do not receive any comments subsequently. As a result, the blogger might stop commenting on other blogs’ postings. A sample of this category is clear in Fig. 2 (b).

Category III. This category is somehow the opposite form of previous category. A blogger receives many comments, where she rarely leaves comments herself. Bloggers of this type, can be considered powerful bloggers. They in fact, write postings that many people read and comment. kelash.persianblog.com is a good example, seen in Fig. 3 (a). The blogger receive comments nearly 8 times the comments she leaves for others in a period of time.

Category IV. Here, the number of comments left, and received are dependant and coupled. A blogger may receive comments, and this fact could be a motive for her to read and leave comments for her commenters. On the other hand, one might comment others, and subsequently put a link. This link will later be a path from that blog to hers, for visitors, and she will probably get comments in future. A sample of such blogs is given in Fig. 3 (b).



(a)



(b)

Fig. 2: Sample Observations. (a) is a blog of category I. (b) blog of category II.

Category V. This category contains blogs that neither receive comments, nor tend to comment others. They might have even banned the commenting possibility on their blogs. These bloggers are not of much importance in our study, and we ignore them in our definitions.

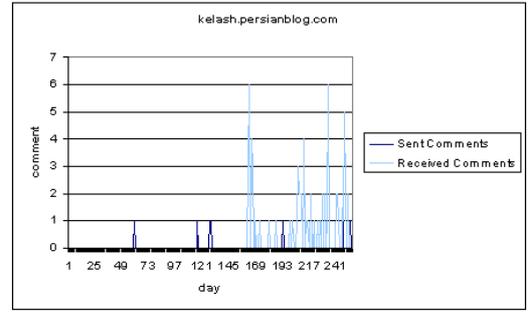
3.2 Failure Definition

3.2.1 Connection Failure

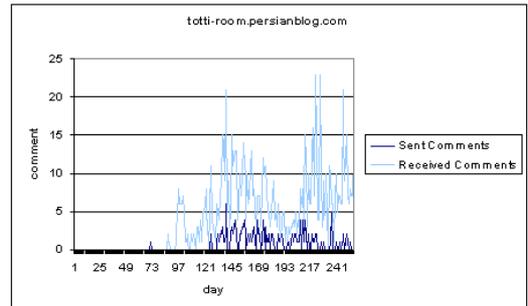
There are some assumptions in our definition, that should be considered. First, we assume that, any blogger, to visit another blog uses a blog (most probably her own) as a start point and follows a series of links to reach a desired blog. So we have no random visit for which going directly to the blog by typing the URL is needed. Lets also suppose, that a constant percentile of readers leave comments and each comment is followed by a link to commenter's blog. So the number of comments could be a representative for the number of readers and thus visitors of that blog.

Any blogger is a *potential source of visitors*. A visitor might choose an outlink of her current visiting blog to another blog and visit that blog. This is where, putting links in others blog, in order to gather visitors, is tricky. Of course it costs one to leave comment for a blog, and it is not logical to leave comment for any blog you visit. So to gather visitors, one might use a strategy of commenting others and putting her links to make her cost minimum and her visitors maximum. Fig 4 shows how visitors flow in a blogspace via links.

As we stated before, when the blogger *A* comments on a posting of *B*, she puts a link from the post comment of *B* to her blog, *A*. If a blogger desires to have more comments, she should have links from other blogs, especially those with many visitors. To put a link, as we said, she can comment



(a)



(b)

Fig. 3: Sample Observations. (a) is a blog of category III. (b) blog of category IV.

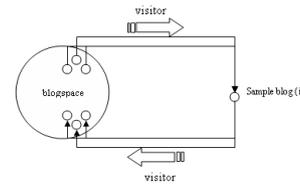


Fig. 4: Visitors flow in a blogspace

others with a definite cost and a definite benefit. A blogger, to our understanding, actually *fails* if she cannot maintain a non-decreasing number of visitors over time.

Visitors often, read the most recent post of a blog they visit, specially if they have already read the previous posts. So a link from the most recent and top post is more valuable than a link from other posts. These links may carry more visitors and will gather more readers for a blog. The strategy could be so that ones links do not get very out-dated.

3.2.2 Commitment Failure

As we stated, another form of failure is *commitment-failure*, when a blogger quits writing after a while. We examined a 9 month period of our data. It is clear in Fig. 5 that the number of bloggers that started a new blog in the period and continued for only one month are nearly 80% of all quits. A basic observation is that, the probability of quitting is much less for a blogger which has been writing for a longer time, in comparison with a new blogger. There should be some triggers to motivate this failure, and probably a model behind this distribution, for which proposing a model is a nice future work and is beyond the scope of this paper.

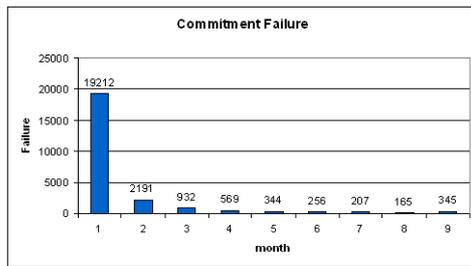


Fig. 5: Chart of Commitment Failure: Each bar shows the number of failures in which the blogger only wrote posts for the given length of months.

4. Related Work

To our knowledge there is no previous work that directly addresses the problem of failure in blogs, though other works on analyzing weblogs are growing rapidly, as the importance of such analysis is becoming more vivid. This is obvious when you see Adamic et. al [3] have studied the network structure of political blogs during the 2004 U.S. Presidential election and showed that the linking patterns between conservative and liberal blogs formed two diverse communities. [13] gives a descriptive analysis of blog postings around the London attacks, with daily trends for the top links and topics for different content templates.

Dynamics of blogospheres are studied in [4, 7]. In fact Adar et al. [4] studies information epidemics in blogs and introduces a new ranking algorithm for blog pages. Some works on blog communities have been proposed. [8] defines a mutual awareness relationship and discovers communities in blog social network based on two factors: (a)communities form due to individual blogger actions; (b) semantics of the hyperlink structure are different from traditional web analysis problems. Above researches, and many others on weblogs, focus on post data. Yet few researchers have focused on comments. Trevino et al. [14] and Gumbrecht et al. [5] show the importance of comments in blog analysis, and conclude that comments play a essential role in interactive nature of blogs. Herrig et al. [6] studies a small comment dataset of 203 weblogs (average of 0.3 comment per post). A larger scale study on comments investigates the relations of comments and posts, and extracts commenting pattern based on blog popularity [9].

5. Future Work

With the definition of failure in mind, and based on the categories, that we stated above, it is clear that a new research trend is now open for researchers for exploring different experiments on failure in various blogspaces.

We believe, there is a vast area for new work, in this field. We are currently working to give a formal definition for failure, together with a game theory-based model the failure phenomenon in blog graphs, to solve some open problems.

First, research and study on the triggers that may cause a failure. For instance if a heavily negative comments could be the main cause of the failure. Or what are the most important triggers and external events, posts etc. that cause a blog success.

Second, modeling failure classification. While many bloggers join and leave the blog space every day, we are interested to find a common grounds for various group of bloggers to clas-

sify them. For instance what are the common features among ten bloggers who failed after the first month of trial.

Furthermore, there are some ideas to predict the connection-failure in a blog based on her interactions and role in the blogspace.

References

- [1] The blog herald blog count october 2005: over 100 million blogs created. available online: www.blogherald.com/2005/10/10/the-blog-herald-blog-count-october-2005.
- [2] Blogscience. available online : www.blogscience.org/data.html.
- [3] L. Adamic and N. Glance. The political blogosphere and the 2004 u.s. election: Divided they blog. In *Proceedings of the WWW2005 Conference's 2nd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis, and Dynamics*, 2005.
- [4] E. Adar, L. Zhang, L. Adamic, and R. Lukose. Implicit structure and the dynamics of blogspace. In *Workshop on the Weblogging Ecosystem, 13th International World Wide Web Conference*, 2004.
- [5] M. Gumbrecht. Blogs as "protected space". In *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, at WWW 04: the 13th international conference on World Wide Web*, 2004.
- [6] S. C. Herring, L. A. Scheidt, S. Bonus, and E. Wright. Bridging the gap: A genre analysis of weblogs. In *The 37th Annual Hawaii International Conference on System Sciences (HICSS04)*, 2004.
- [7] R. Kumar, J. Novak, P. Raghavan, and A. Tomkins. On the bursty evolution of blogspace. In *Proc. of the 12th International World Wide Web Conference*, pages 568–576, 2003.
- [8] Y. Lin, H. Sundaram, Y. Chi, J. Tatemura, and B. Tseng. Discovery of blog communities based on mutual awareness. In *Proceedings of the WWW06 Workshop on Web Intelligence*, 2006.
- [9] G. Mishne and N. Glance. Leave a reply: An analysis of weblog comments. In *Third annual workshop on the Weblogging ecosystem*, Edinburgh, Scotland, May 2006.
- [10] T. Nanno, Y. Suzuki, T. Fujiki, and M. Okumura. Automatic collection and monitoring of japanese weblogs. In *WWW2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*, 2004.
- [11] V. Qazvinian, A. Rassoulain, and M. Shafiei. A large-scale study on persian weblogs. In *to appear in the proceedings of 12th international joint conference on Artificial Intelligence, workshop of TextLink2007*, 2007.
- [12] D. Secko. The power of the blog. *The Scientist* 2005, 19(15):37, 2005.
- [13] M. Thelwall. Bloggers during the london attacks: Top information sources and topics. In *Proceedings of the WWW06 Workshop on Web Intelligence*, 2006.
- [14] E. M. Trevino. Blogger motivations: Power, pull, and positive feedback. In *Internet Research 6.0*, 2005.