Visualizing the Liveliness of Discussions and Reply Relationships on Online Discussion Boards

Hidekazu Shiozawa

Department of Software Science, Tamagawa University, Japan
shiozawa@eng.tamagawa.ac.jp

Abstract - This paper proposes a technique to interactively visualize the liveliness of discussions and the reply relationships of messages on online discussion boards, and describes prototype software that visualizes the liveliness of any topic thread on 2channel (Ni Channeru) [1], Japan's largest online discussion site. This visualization is based on a line graph of the number of postings versus time. To show reply relationships on the same chart, this visualization utilizes the technique of Thread Arcs [14], which show reply relationships as semicircular arcs that connect message plots. Furthermore, for users to know the length of each message, the visualization can plot a message as a pair of small bar graphs on the same line graph.

Keywords: liveliness of discussion, visualization of discussion structure, online discussion boards, 2channel (2ch.net), Thread Arcs

1 INTRODUCTION

An enormous amount of information is exchanged every day in online discussion boards or Web forums on the Internet. Although such sites are mixtures of various information, some lively discussions in which many people participate would include valuable, interesting, or entertaining information. However, conventional discussion boards provide only user interfaces in which users need to read messages one by one from the beginning of each thread, even if they are looking for only the lively parts of the discussion.

Therefore, this paper proposes a technique to interactively visualize the liveliness of discussions and the reply relationships of messages on online discussion boards. It also describes prototype software that visualizes the liveliness of any topic thread on 2channel (Ni Channeru) [1], Japan's (and probably the world's) largest online discussion site. Although the "liveliness" of a discussion can relate to various factors of the discussion, this paper roughly defines liveliness as "the number of posted messages per unit time". The number of replies per message and other statistical data are used as secondary information.

The 2channel consists of many Web discussion boards on different themes, and each board allows a user to make a new "thread" for every new topic. In each thread, posted messages are displayed sequentially from 1 up to 1000 message numbers. On each board, the display order of its threads is determined by floating the thread with a new posting at the top. At the same time, the user can use the "sage" (down) command to keep the thread from floating.

In addition, 2channel features anonymity[1], that is, users are not required to provide any registration or contact information. Actually, most of the postings are contributed by "anonymous". However, to reduce the negative effects of anonymity, many boards have recently appended a string called "ID", which is apparently a random string to each message for identifying the user. The ID string is calculated with a one-way hash function from both the posting IP address and the posting date, thus it remains the same when the user posts any message from the same address on the same day.

2 RELATED WORK

2.1 Analysis of Online Discussion Boards

As for the liveliness of online discussion boards, Matsumura et al. [2], [3] collected and analyzed log data from all discussion boards in 2channel to propose the quantification of characteristics of each discussion thread according to the following eight indices:

Contents (C) The average size of a message in bytes (except for the following AA and V).
Activity (A) The average number of messages per thread.
Interactions (I) The average number of replies per message.
Speed (S) The average number of messages per day.
Vocabulary (V) The byte ratio of "2channel words".
ASCII Art (AA) The byte ratio of text arts.
Nameless (N) The ratio of messages posted by anonymous ("nanashi-san").
ABON The ratio of messages deleted[2] by administrators.

Researchers calculated these indices by the category of which discussion boards they belonged to, and showed the rough trends of discussions in each category by analyzing the relation of these indices with the method of covariance structure analysis. Then, they proposed a statistical model to show the overall dynamism of 2channel with the cause-effect relationships of the three trends: the discussion divergence trend (A and S are large), the discussion deepening trend (C and I are large), and the stylized expression trend (V and AA are large).

In addition, Matsumura et al. [5] also performed statistical analysis for the dynamism of Yahoo! Japan Message Board [4], which requires a user's registration unlike 2channel. They used a weighted, directed graph representing the relation of users to analyze reply relationships and their influences in

[1] Indeed, it is not so anonymous, because the servers maintain the log data of all connections.

[2] "Abon" is the system message indicating deletion of the message.

the discussion. Then, they focused on the role of users and classified discussions in Yahoo! Japan Message Boards into three types: leader-driven communication, leader-follower-collaborative communication, and follower-driven communication.

These researchers investigated users' collective behavior in online discussions from a macro perspective by quantifying a discussion with several index values. Therefore, the techniques can be significant in the fields of sociology, mass psychology, etc. However, they are not directly useful for general users reading online discussion boards.

2.2 Time-Series Analysis of Discussion

From the users' viewpoint, discussion boards with a greater number of postings are considered more worth reading, because such boards usually hold lively and active discussions and contain abundant fresh information. As for 2channel, the degree of discussion liveliness is often called "ikioi" (momentum) or "nobi" (growth) [3]. These words roughly signify the number of postings per unit time ("growth" also means the total number of postings).

For users to find such lively and interesting discussions, some software [6], [7] and Web sites [8], [9] for reading online discussion boards display the value called "ikioi" (momentum), or the posting speed per hour, for every discussion thread, which is the average number of postings per time. In addition, some boards show ranking lists of the values.

Furthermore, there is software called V2C [10] that can visualize the time variation of the value on a chart. V2C draws a thread in a line graph using the x-axis for time and y-axis for message number (i.e., the number of postings). Therefore, the graph slope represents the number of postings per time or the momentum.

Matsumura et al. [11] extended the research previously mentioned [2] and attempted a time series analysis of the liveliness of a discussion in 2channel with the following six new indices: the posting interval rate (the average interval between postings), the interaction rate (the average number of replies per message), the anonymous rate (the rate of anonymous postings), the 2channel word rate (the rate of jargon per message bytes), the ASCII arts rate (the ratio of characters often used in text arts), and the new information rate (the rate of new nouns per message).

2.3 Visualization of the Discussion Structure

Visualizations of discussion structures for character-based online communication have been developed and used conventionally. Most of the software for e-mail and Netnews (Usenet) enables users to gather a related discussion and view relationships of messages in a tree structure by using the messages' headers such as "Subject" and "References". This method visualizes a lively discussion in Netnews as a complicated tree structure because of its message format such as e-mail.

[3] Moreover, the state of extremely lively discussion is sometimes called "matsuri" (festival), where users join the discussion and post many messages in the aim of liveliness itself.

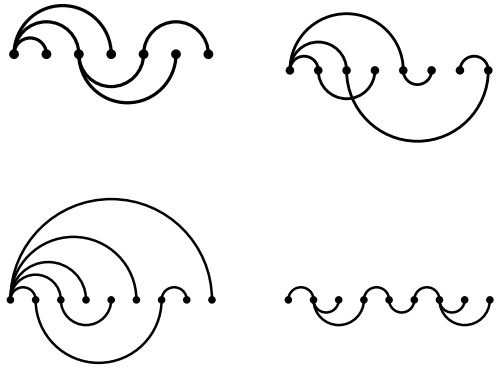

Figure 1: Thread Arcs (B. Kerr, IBM Research).

In the CSCW (computer-supported cooperative work) field, many discussion systems [13] with visualization interfaces have been proposed traditionally. Most of such systems show that the discussion flows with branching and merging as directed graph structures on a two-dimensional plane. For example, Matsumura et al. [12] proposed the analysis method and the visualization of the discussion structure on their online discussion board as a kind of directed graph.

On the other hand, this paper refers to the research of Thread Arcs [14], which is a visualization of e-mail reply relationships in a personal mailbox. As Figure 1 shows, the Thread Arcs visualization can show reference relationships among messages placed on a straight line, by drawing semi-circular arcs connecting replying and replied messages. As a result, a compact visualization is realized that does not need two-dimensional layouts.

3 INFORMATION TO BE VISUALIZED

As suggested by the words "ikioi" (momentum) and "nobi" (growth), it is considered that users' feelings about the liveliness of an online discussion are greatly affected by the passage of time. Therefore, this research uses visualization based on a statistical graph with a time axis to display the following information:

The cumulative number of postings Visualizing this is the most basic view for the liveliness of a discussion over time. A line graph such as V2C is used. It relates Matsumura's A and S indices.

Postings by the same ID (user) Messages posted by the same ID (as currently focused by the user) are highlighted in the view.

The cumulative number of unique IDs If this is small, it means that few users perform many postings.

Postings with "sage" command Posting messages with the "sage" (down) command prevents the thread from floating to the top of the board.

Size of every message The visualization can show the size of each message, excluding symbols and spaces.

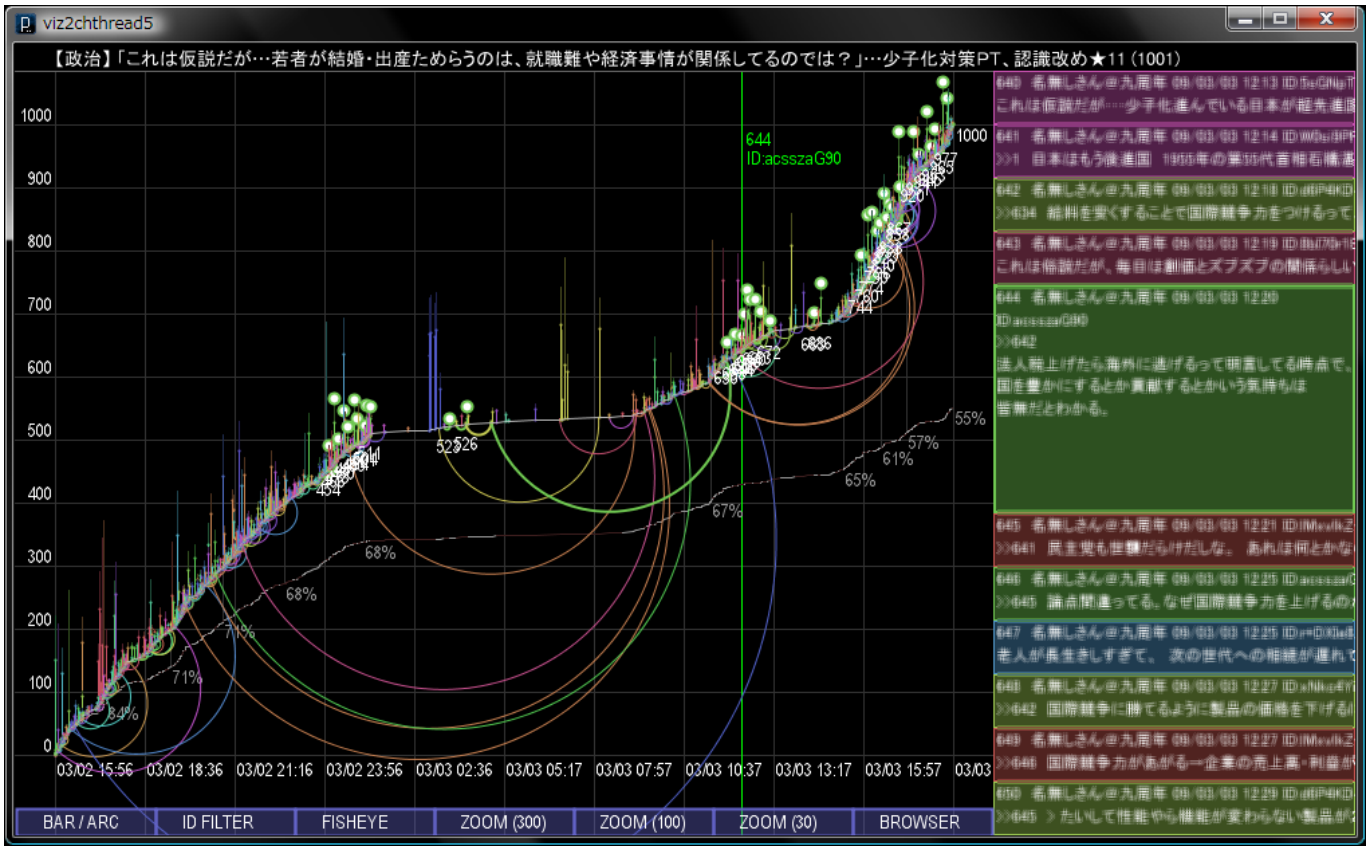

Figure 2: The visualization overview (the left area is for visualization, the right area is for messaeg texts).

Reply relationships between messages Every reply in the thread is visualized. It relates to Matsumura's I index.

The cumulative number of replies It also relates to Matsumura's I index.

For now, there are no metrics corresponding to the number of 2channel words (V), the rate of anonymous postings (N), and the number of deletions by administrators (ABON) in Matsumura's indices [2], [11]. The author considers that these are less important for the interactive visualization of a single thread, because most new jargon is used frequently only for short periods of time, and the number of signed postings and message deletions is small in a thread.

On the other hand, this research uses information about the ID and "sage" (down), although this is not included in Matsumura's indices. These are considered useful to characterize each thread, because the number of IDs can be a useful index for the number of unique posters, and "sage" postings are commonly seen in threads that are little changed by the participating users.

In addition, this research does not count replies to the first message, which presents the topic of the entire thread, as effective replies.

4 VISUALIZATION TECHNIQUE

This section describes the visualization technique in detail. The prototype software has been developed with Processing [15], a programming language that is suitable for dynamic graphics. When a user drag-and-drops the log file or the URL of the thread, which is saved by a 2channel browser like Open Jane [6] or displayed on a Web browser like Firefox, onto the window of the visualization, the screen is displayed as shown in Figure 2 [16], [17].

4.1 Line Graph for Discussion Liveliness

This visualization technique is based on a line graph of the cumulative number of postings versus time. By using this technique, users can read the total number of postings up until a certain point in time, and view by the slope, the progress of making the discussion lively. If the graph slope is extreme at some point in time, the discussion at that time is very lively. On the other hand, if the slope is small, the discussion is not so lively.

Each posted message is plotted with a different color (hue) calculated from its ID on the graph. When a user selects a message to investigate its relationships, the messages posted by the same ID are highlighted on the chart with particularly large plots (Figure 3 and 4). These features enable users to determine the posting tendencies of other participants.

4.2 Line Graph for the Number of Users

The software also visualizes the cumulative number of unique IDs as a line graph, which shares the vertical axis with the number of postings, and shows the percentage of unique IDs for all postings in digits, every hundred messages. If

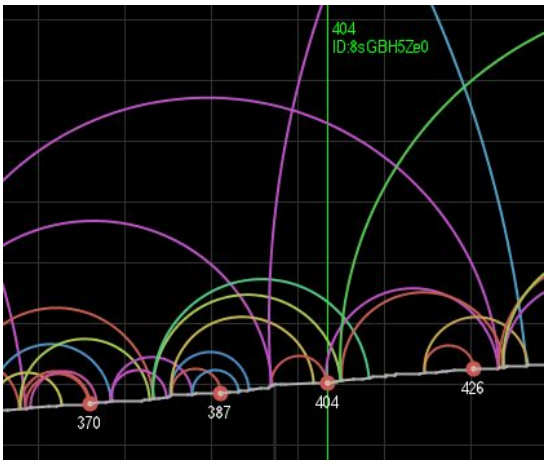

Figure 3: Arcs for reply relationships and coloring for IDs.

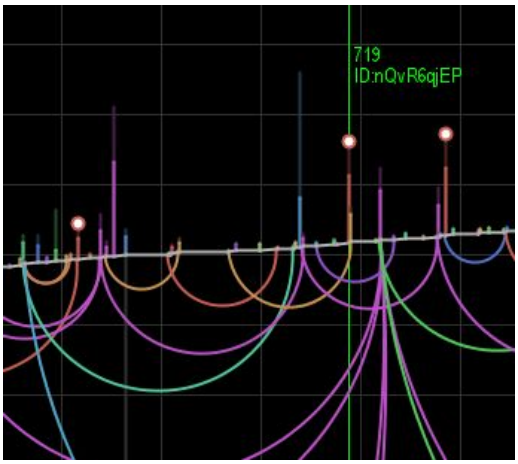

Figure 4: Small bar graphs for the amount of information.

the percentage is small, some regular users are continuously posting in the thread. On the other hand, if the percentage is large, many new posters are participating constantly.

Furthermore, the parts of the graph where messages are posted by "sage" (down) commands are drawn in a lighter color. Usually, threads that are sinking in this manner are hard for newcomers on the discussion boards to discover.

4.3 Small Bar Graphs for Message Size

The software can visualize the approximate amount of information contained in each message using the two small bar graphs at a position of each plot on the line graph of the number of postings (Figure 4). These two graphs represent the length of each message in bytes and the length excluding symbols and spaces in bytes, respectively. The latter's bar is drawn more thinly than that of the former's.

If the lengths of the pair of bars on a plot are very different, the corresponding message contains many symbol characters. Therefore, this feature enables a user to estimate the effective size of every message except for text arts, formatting spaces, etc. Although Matsumura removed text arts with the dictionary for calculating the effective size of every message, this interactive visualization enables users to determine the amount of information visually from the bar graphs.

4.4 Arcs for Reference Relationships

A feature of this interactive visualization is to display reply relationships of messages in semicircular arcs on the same line graph of the number of postings by using the technique of Thread Arcs (Figure 3). As a result, a user can investigate the tendency of replies in the entire thread, find the messages that have particularly many replies, and read the sequences where the discussion becomes much livelier with many replies.

Applying the technique of Thread Arcs to an existing monotone increasing graph enables us to visualize the relationships of data as additional information on the same graph, because the technique visualizes connections between two plots on one dimension.

The arcs are shown on the upper side of the line graph when the small bar graphs are not drawn, and on the other side when the bar graphs are drawn to avoid visual overlapping. Their colors are the same as the plots of the corresponding reply messages.

In addition, the cumulative number of replies is also visualized as a line graph like the number of postings and the number of unique IDs. This number often increases when some users in the thread have opposite views, because their lively debate elicits many replies.

4.5 Partial Zooming and Fisheye View

The time axis of the graph ranges from the thread creation to the latest post. The vertical axis ranges from 1 to 1000 message numbers. Plotting up to a thousand messages in a chart can make the visualization too dense and occluded, so the zooming feature is provided to show the range between the last ten and the next ten messages around the one that the user has selected.

However, some reply relationships connect temporally separated messages in a thread, so only clipping and zooming a part of a chart can be insufficient for a user to view relationships in detail. Therefore, the visualization provides a typical fisheye view with the graphical fisheye view technique [18] (Figure 5). It is also possible to apply simple zooming at the same time.

When you select a message in the fisheye view mode, the x-coordinate is recalculated interactively by the formula,

$$x_{feye} = \frac{(d+1)\, x_{max}\, x}{dx + |x_{max}|},$$

where x is the x-coordinate before the transformation when the x-coordinate of the focus is zero, x_{feye} is the x-coordinate after the transformation, x_{max} is the x-coordinate of the display limit in the zooming direction (depending on the sign of x), and the d is the zooming parameter ($d = 8$ in the figure).

By this technique, a user can overview the entire thread and detail around the selected message, and roughly know the destination point of arcs representing reference relationships.

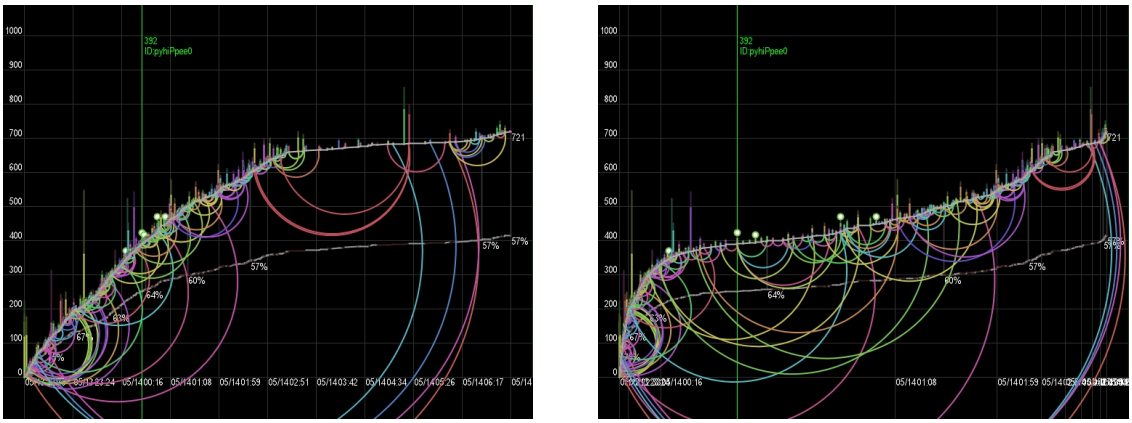

Figure 5: Nonlinear zooming with fisheye view, on the left is before applying and on the right is after applying.

5 APPLICATION AND DISCUSSION

This technique enables users to visually understand tendencies of individual threads in a temporal view, although Matsumura et al. clarified the tendencies of online discussions in a macro perspective with several quantified indices.

This section discusses the characteristic tendencies of threads found by the visualization in three perspectives: thread on event vs. thread on theme, thread with impressions vs. thread with debate, and thread for open discussion vs. thread for members. These were found by applying this visualization to various types of threads in 2channel.

5.1 Thread on Event vs. Thread on Theme

By observing the time variations of liveliness on many different threads from creation to end, some threads become most lively at their early stage from creation then gradually decay over time, and others become lively at random times regardless of their creation and end.

Such a phenomenon may relate to the purpose of each thread or the reason for its creation. The former type is often seen when a kind of event triggers a user to create a thread for discussion. The latter type is often seen when the thread is continuously maintained by some theme, such as a local area or sports team.

For example, the right-hand side of Figure 6 is the thread of breaking news about the national debt balance (over the course of a day). The left-hand side of the figure is a different thread on the topic of the Kingdom of Thailand (over the course of three years). The latter has been maintained in a variety of topics related to Thailand, but the tremendous surge of the coup d'etat is shown in comparison to before or after it.

5.2 Threads with Impressions vs. Threads with Debates

In many threads, some are lively on some level in the number of postings, even though there are very few replies in the thread. In contrast, some are not that lively and the increase of postings is slow, but discussion is actively performed with long messages and many replies.

One of the factors causing these differences can be whether or not the topic of the thread is easy to debate. For some kinds of news, there is not much room for debate among users other than reported in the first message, so the discussions do not become so interactive even if people post many messages of impression such as surprise and repulsion.

Moreover, when the first message reports shocking news, occasionally a number of messages without any replying comments are posted much more rapidly, for example, simple messages of surprise, praise, opinion, and agreement with the first message. It depends on the readers to deem such communication as lively and important.

On the other hand, social problems, historical issues, and topics involving users often expand discussions in various directions and accelerate lively communications among users. It enlarges the number of replies in the thread.

For example, Figure 7 shows two visualizations of threads, both in the newsflash board. On the left is news of Prime minister's remarks; on the right is news of the exhibition of the Dead Sea Scroll. The number of postings is not much different, but the number of replies is contrasting: the former is 15% of the number of postings, the latter is 48%.

5.3 Threads for Open Discussion vs. Threads for Regular Members

It can be said that threads with a high percentage of unique IDs are open for discussion, and various users are continuously participating and immediately leaving. On the contrary, threads with a low percentage of unique IDs are for regular members, and a relatively small number of participants continue to post messages.

The threads for regular members often continue to progress with "sage" (down) postings, because such threads are not very interesting for other people. Although there is a small number of new participants in such thread, sometimes the discussion becomes very lively, as in an explosion. It is often triggered by the news of an event related to the topic of the thread, such as result of sports match, and indicates that the number of postings per member is increasing.

For example, the left-hand side of Figure 8 shows a thread on certain local news of a city, and the right-hand side is

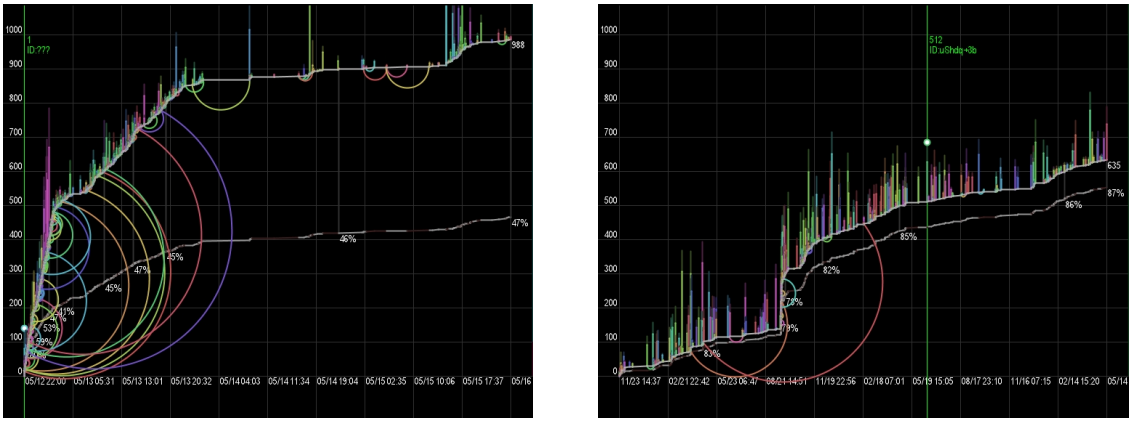

Figure 6: Visualizations of the threads on an event (left) and on a theme (right).

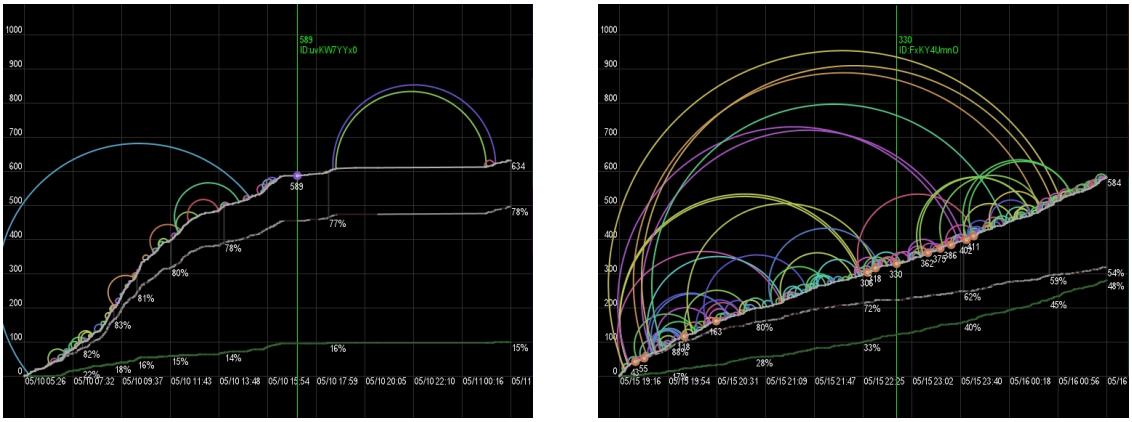

Figure 7: Visualizations of the threads with impressions (left) and with debates (right).

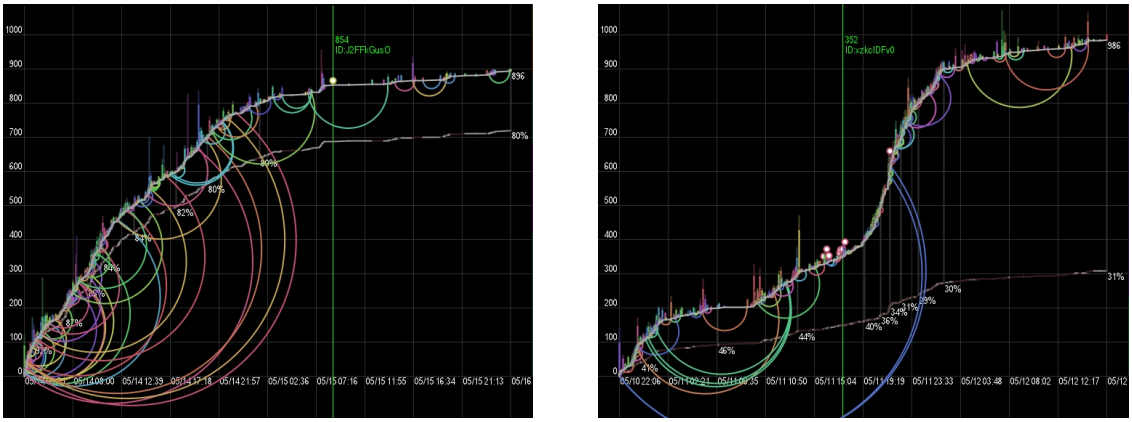

Figure 8: Visualizations of the threads for open discussion (left) and for regular members (right).

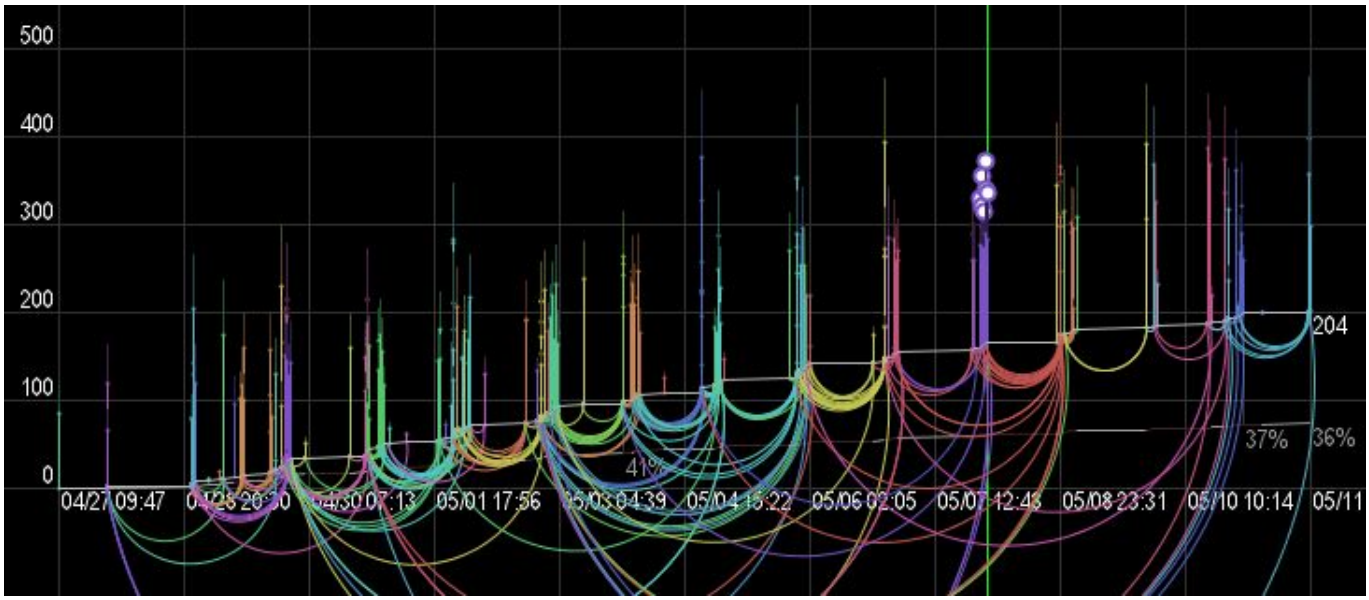

Figure 9: An example of very high correlation between discussion liveliness and time of day.

a thread for fans of a specific professional baseball team. The former indicates over 80% unique IDs, while the latter indicates 31%, in keeping with floating with "sage" progress.

6 SUMMARY

In this paper, the author has proposed a technique to interactively visualize the liveliness of a discussion and the reply relationships of messages on online discussion boards. This visualization consists of a line graph of the cumulative number of postings versus time to visualize the liveliness of discussion, and semicircular arcs connecting replying pairs of messages on the same graph.

In other words, this research extends the technique of Thread Arcs to visualize relationships of two points on a monotone increasing polygonal line or curve, although originally it can visualize only on a straight line. Therefore, this visualization is considered applicable to other existing line graphs to visualize additional information among data.

The author applied the method to the real threads in 2channel, and then the various characteristics in different threads were found. The author then discussed them in three perspectives: thread on event vs. thread on theme, thread with impressions vs. thread with debate, and thread for open discussion vs. thread for regular members.

Some topics are considered for future work, such as experimental evaluation, visualization with the effect of time of day, and feedback to the technique to statistical analysis. Three-dimensional visualization such as the Natto View [19] may also be useful to provide visual interaction on a chart.

Many threads are affected by human life rhythm regarding the effect of time of day, so it is known that the number of the postings decreases from midnight until the next morning [16]. Furthermore, there are threads that have a very high correlation between their liveliness and time; Figure 9 shows such a thread, where many messages are posted at the same time every day.

ACKNOWLEDGEMENT

The author would like to thank Hiroyuki Nakayama for his assistance. He also performed basic analysis of collected log data in the perspective of effects of time of day [16].

REFERENCES

[1] 2channel (Ni Channeru), http://www.2ch.net
[2] N. Matsumura, A. Miura, Y. Shibanai, Y Ohsawa, M. Ishizuka: The Dynamism of 2channel, Transactions of Information Processing Society of Japan, Vol.45, No.3, pp. 1053–1061 (2004). (in Japanese)
[3] N. Matsumura, A. Miura, Y. Shibanai, Y. Ohsawa, T. Nishida: The Dynamism of 2channel, AI & Society, Vol. 19, No. 1, pp. 84–92, Springer Verlag (2005).
[4] Yahoo! Japan Message Boards, http://messages.yahoo.co.jp
[5] N. Matsumura: The Dynamism of Yahoo! Japan Message Boards, Proceedings of The 19th Annual Conference of The Japanese Society for Artificial Intelligence, 2C1-02 (2005). (in Japanese)
[6] Open Jane Project Top Page, http://sourceforge.jp/projects/jane/ (2002–).
[7] noby: 2ch Hot Thread Search, http://www.forest.impress.co.jp/article/2006/03/01/2ch_hotthread.html (2006).
[8] The Thread Ranking, http://www.bbsnews.jp
[9] R. Nakajima: 2channel News Navigator (2NN), http://www.2nn.jp (2004–).
[10] n|a: V2C, The 2ch Browser using Java+Swing, http://v2c.s50.xrea.com/ (2004–).
[11] N. Matsumura, Y. Shibanai, H. Nishimura, A. Miura, T. Nishida: Time-series Analysis for the Communications on 2channel, Online Proceedings of The Fourth MYCOM, The Japanese Society for Artificial Intelligence (2003). (in Japanese)

[12] N. Matsumura, Y. Kato, Y. Ohsawa, M. Ishizuka: Visualization of Discussion Structure for Discovering and Understanding Discussion Points, Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 15, No. 5, pp. 554–564 (2003). (in Japanese)
[13] P. Kirschner, S. Buckingham-Shum, C. Carr (ed.): Visualizing Argumentation – Software Tools for Collaborative and Educational Sense-Making, Springer (2003).
[14] B. Kerr: THREAD ARCS: An Email Thread Visualization, Proceedings of The Ninth IEEE Symposium on Information Visualization, pp. 211-218 (2003).
[15] C. Reas, B. Fry: Processing: A Programming Handbook for Visual Designers and Artists, MIT Press (2007). (http://processing.org)
[16] H. Shiozazwa, H. Nakayama: Visualizing Liveliness of Discussion and Reference Relations on Online Discussion Boards, Proceedings of DICOMO 2008, Information Processing Society of Japan, pp.1308-1315 (2008). (in Japanese)
[17] H. Shiozazwa: Visualization of Liveliness of Discussion and Reference Relationships on Online Discussion Board, Proceedings of Interaction 2009, Information Processing Society of Japan (2009). (in Japanese)
[18] M. Sarkar, M. H. Brown: Graphical Fisheye Views of Graphs, Communications of the ACM, Vol. 37, No.12, pp.73–84 (1994).
[19] H. Shiozawa, H. Nishiyama, Y. Matsushita: The Natto View : An Architecture for Interactive Information Visualization, Transactions of Information Processing Society of Japan, Vol. 38, No.11, pp. 2331–2342 (1997). (in Japanese)

(Received August 29, 2009)
(Revised June 23, 2010)

Hidekazu Shiozawa received his B.E., M.E. and Ph.D. degrees in instrumentation engineering from Keio University, Japan, in 1994, 1996 and 2000, respectively. He is currently Associate Professor at the Department of Software Science at Tamagawa University, Japan. His research interests include information visualization, human-computer interaction, network application and computer-supported cooperative work. He is a member of ACM, IEEE CS, IPSJ and IEICE.