

Visualizing a Personal Timeline by Adding Multiple Social Contexts

Haewoon Kwak* Yoonsung Hong† Jinyoung You‡ Sue Moon†

*Telefonica Research, Barcelona, Spain

†Division of Web Science and Technology, KAIST, Korea

‡Advanced Research Institute, LG Electronics, Korea

Abstract

We improve the readability of a personal timeline by weaving multiple social contexts of tweets into a visualization. Our social contexts consist of three dimensions: community membership, key persons, and interesting tweets within a personal timeline. A person is often a member of several communities, such as a family, a class, or a team, simultaneously. We identify all communities that a user participates in. Labeling a tweet with a visual representation to indicate what community it belongs to can help readers to understand why the tweet is written, since different communities are likely to carry tweets in different contexts. We then discover key persons and interesting tweets within a personal timeline. Our prototype design demonstrates how three social contexts work together for visualizing a personal timeline.

Introduction

A Twitter user typically has a few hundreds of followers. As a mixture of their tweets fills a personal timeline, a user has a difficult time to read all of tweets. Several methods are proposed to improve the readability of a personal timeline. Most of them aim to summarize the timeline or extract little number of key tweets from a stream of tweets. Through the interaction history between a user and followers, few tweets that are written by close friends involving tangible interaction or sufficiently retweeted are highlighted, but most of tweets are falling into neither of those categories. The majority of tweets in a personal timeline require more contexts for rich visualization.

In this work, we visualize a personal timeline by adding multiple social contexts to tweets. Our social contexts consist of three dimensions: community membership, key persons, and interesting tweets within a personal timeline. A person is often a member of several communities, such as a family, a class, a sports team, simultaneously. People in different communities have different interests and are under different circumstances. That is, different communities are likely to carry tweets in different contexts. Labeling a tweet with a visual representation to indicate what community it belongs to, thus, can help readers to understand why the

tweet is written. As a result, most of tweets are presented with social contexts through community membership. We then identify key persons and interesting tweets by several factors based on theoretical and practical knowledge, such as homophily, link exchange, tie strength, interaction history, social hierarchy, informativeness, and collective intelligence. Researchers have proved that those factors fundamentally explain how and why people pay attention to others in various social systems not limited in Twitter, and it strongly supports our design. We note that our social contexts are quite independent; we highlight key persons or interesting tweets while showing community structures. Our prototype design shows how our multiple social contexts work together well when visualizing a personal timeline.

The rest of the paper is organized as follows. We first explain our community mining algorithms and its technical details. Next, we introduce how we can identify key persons and interesting tweets within a personal timeline by taking account of several factors. Then, we demonstrate our prototype design with an image and explain an evaluation plan. Finally, we conclude with future work.

Adding Multiple Social Contexts

We augment a personal timeline by allotting community membership, identifying key persons, and recognizing interesting tweets. As the multiple contexts are independent of each other, a user can choose the level of details by activating or deactivating each context in our prototype design.

Our design of social contexts is deeply rooted in previous literature from social science. We explain each social context with seminal studies so as to understand how they help users who struggle with unstructured stream of tweets.

Mining community membership

There is no perfect agreement on the rigorous definition of social community. Despite such a disagreement, in light of hypothesis that the number of links (relationships) within a community is larger than that of links bridging communities, various community mining algorithms has been developed in many disciplines, such as sociology, biology, computer science, engineering, economics, and politics (Fortunato 2010).

Most of approaches assigning a node (a user) to only a single community are called *partitioning* algorithm. They

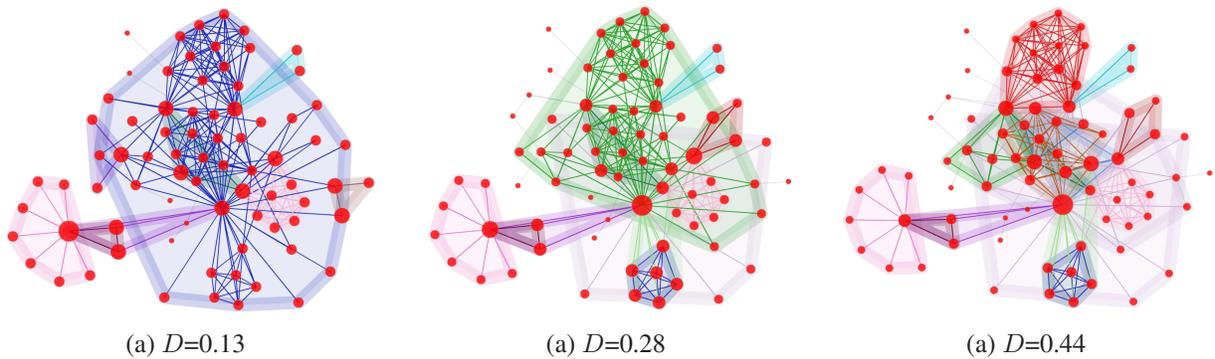


Figure 1: [Best in color] Identified communities according to varying cutting point

are inappropriate for our application, since people have multiple community memberships and those communities give users multiple contexts. To address the limitation of partitioning algorithm, we use Hierarchical link clustering (HLC) proposed by Ahn *et al.* in order to find overlapping communities (Ahn, Bagrow, and Lehmann 2010). HLC assigns a link, instead of a node, to a single community. A node having multiple links thus can belong to multiple communities. Ahn *et al.* demonstrate the promising results of HLC with empirical networks. HLC libraries implemented by Python, C++, and R are publicly available via their web page¹.

One of the strengths of HLC is that we can find communities under different resolutions if we change the cutting point instead of using the optimal cutting point to maximize the partition density, D . Figure 1, depicted by HLC library, shows how outcomes can be changed from big, sparse communities to small, dense ones by varying D . This enables people to find the exact communities that fit for their mental models of communities they feel sense of belonging.

In our prototype design, people can easily edit community membership, such as adding or removing members by drag-and-drop. When people have new followees over time, people deal with it by manual editing. People can assign an appropriate name to each community, such as ‘writers’, ‘collaborators’, or ‘family’. The result of identified communities can be automatically synchronized with the *list* features in Twitter for enlarging the adaptability of our application. Although Twitter offers list features for managing multiple timelines for a subset of followees, it is not prevalently used due to burden of creating and managing lists. Our automatic syncing schemes address such problem and can be accessed from any other Twitter clients including the genuine Twitter web page.

Finding key persons within a personal timeline

Then we find key persons among hundreds of followees. The theoretical and empirical backgrounds of social science guide us to extract several factors from Twitter as below.

Homophily Homophily is the tendency of people hanging out with similar people rather than with dissimilar

ones (McPherson, Lovin, and Cook 2001). Homophily works in Twitter; Weng *et al.* find that Twitter users are likely to follow those who write tweets of similar topics (Weng et al. 2010), and Kwak *et al.* report that Twitter users break the follow relationship when followees’ tweets are about uninteresting topics (Kwak, Chun, and Moon 2011). In our design, we compute the topical similarity between a user and each followee by the common hashtags, which is a good indicator of predicting unfollow (Kwak, Moon, and Lee 2012), and highlight the most similar followees.

Relationship exchange Recent studies reveal that bidirectional follow relationships decrease the likelihood of being unfollowed in Twitter (Kivran-Swaine, Govindan, and Naaman 2011; Kwak, Chun, and Moon 2011; Kwak, Moon, and Lee 2012). We offer an option to distinguish those who exchange follow relationships from those who do not.

Strong and weak ties Kwak *et al.* report that a personal network in Twitter is a mixture of strong and weak ties (Kwak, Chun, and Moon 2011), and interviewees answer that tweets from strong ties are more important than those from weak ties.

Tie strength can be approximated by the overlap of friends between two people (Gilbert and Karahalios 2009; Onnela et al. 2007). Since follow relationships in Twitter have directionality, we borrow the concept of outlink and inlink equivalence, which are the overlaps of outlinks and inlinks, respectively, defined in a directed network (Wasserman and Faust 1994). Due to high correlation between outlink and inlink equivalence (Kwak, Moon, and Lee 2012), we compute the number of common followees only and intuitively present tie strength by the width of links.

Actual interaction partners Interactions between two people bring emotional closeness and are one of the important dimensions in tie strength (Granovetter 1973). Several studies report that the number of those who actually interact with is much smaller than that of followees (Huberman, Romero, and Wu 2009; Kwak, Chun, and Moon 2011). Thus we emphasize a set of followees who actually interact with the user in the past among many followees.

¹<http://barabasilab.neu.edu/projects/linkcommunities/>

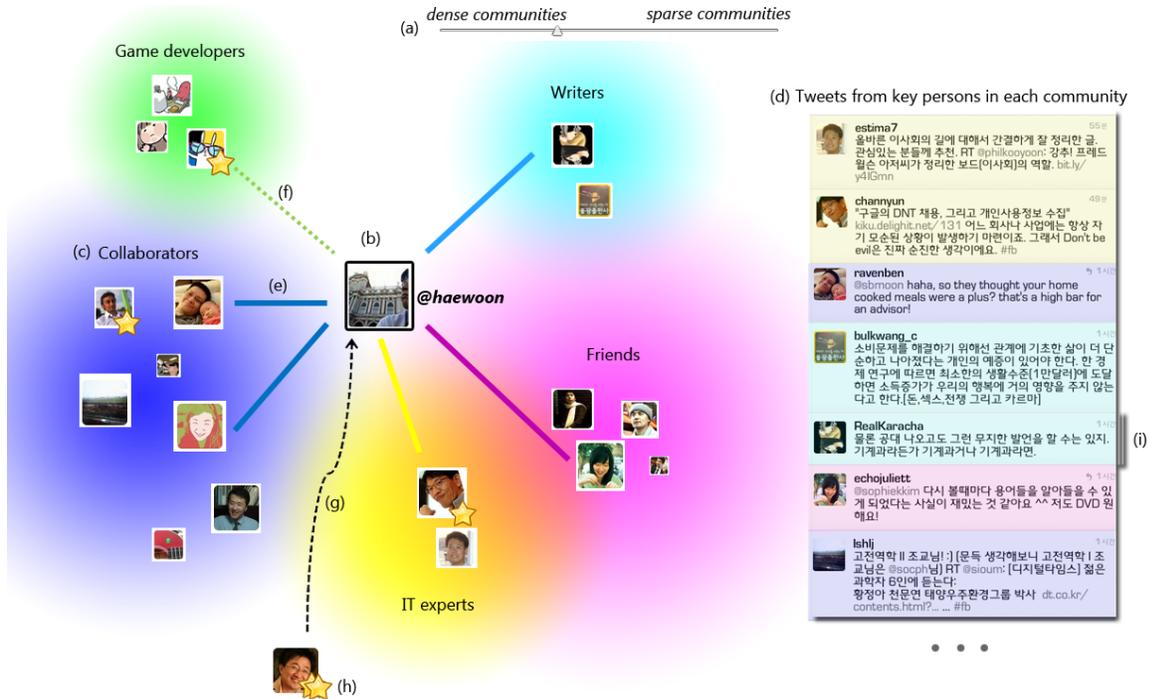


Figure 2: [Best in color] Our prototype design showing a personal timeline. @haewoon is a member of game developers, collaborators, IT experts, friends, and writers, simultaneously.

Social hierarchy The number of followers represents how many people pay attention to. However, due to people’s habits of reciprocal follows in Twitter, having many followers is not so difficult if one has many followees (Kwak, Chun, and Moon 2011). Instead, we compute the ratio of the number of followers to that of followees, which is commonly interpreted as a more accurate indicator of higher social position. Then, we offer a filter that scale photos of followees according to such ratio.

Informativeness From the perspective of an information network, Kwak *et al.* show followees who write tweets a reader favored are less likely to be unfollowed than by chance (Kwak, Moon, and Lee 2012). We highlight followees who are favored by a user in the past.

Marking interesting tweets in a personal timeline

As the third social context we add, we mark interesting tweets within a personal timeline. The number of retweeted counts (or favored counts) is a good signal of finding such tweets. In addition to the total number of retweeted counts of a tweet, we visualize who (or which community) retweeted the tweet within a personal network in our design. It enables a user to understand who (or which community) pay attention to such tweet. Twitter offers API to list all users who retweet a particular tweet, and thus it does not need expensive computations. Recent studies reveal that showing the characteristics of the first source of a tweet is good for

assessing the trustworthiness of such tweet (Diakopoulos, Choudhury, and Naaman 2012). However, a few API calls and computations are additionally required for that, so we leave the choice to a user about a trade-off between performance and a view that is more comprehensive.

We offer features to not only mark important tweets but also filter annoying tweets. Although followees are those who a user is interested in, sometimes their tweets are not interesting to a user. For example, people do not like followees’ burst tweets in short time interval, automatically generated tweets, such as Foursquare logs, or very short pointless tweets such as lol (Kwak, Chun, and Moon 2011). We offer a filtering option for such tweets. Fortunately, those tweets are easily recognizable.

Prototype Design

Figure 2 shows our prototype design visualizing a personal timeline by allotting above three social contexts to tweets. As these multiple contexts are independent of each other, a user can choose the level of details by activating or deactivating each context.

We explain each element from (a) to (i) in Figure 2 and demonstrate how they improve the readability of a personal timeline. (a) is a partition density slider that adjust the boundary of communities as mentioned in Figure 1. (b) is a currently logged user. All profile photos are scaled to the relative importance computed by several factors including the

ratio of the number of followers to that of followees. (c) is a community name. People add a name to each community while referring to key persons in each group. (d) is a personal timeline, a stream of tweets colored by community membership. The colors in the timeline are matched to those in the left diagram. (e) is a strong tie. In our prototype, thick and solid lines show that the relationship is strong. It is computed by the overlap of relationships and interaction history, as we mentioned earlier. (f) and (g) show weak ties. Twitter is closer to information network rather than other social networks (Kwak et al. 2010), so we should be able to effectively show the weak ties that play an important role to disseminate information. (h) shows favorited users by counting the number of favorites in the past. (i) indicates the existence of filtered tweets, such as burst tweets or automatically generated tweets. We also provide several interactive views. First, we offer a timeline of a specific community. People can see tweets from a specific community just by clicking that community. Second, we offer a view of those who retweeted a tweet when a user moves a mouse over the interesting tweet. Next, we offer a view of previous conversations by mentions or replies when clicking strong ties.

In our prototype design, through three social contexts that are community membership, key persons, and interesting tweets, we add rich contexts to most of tweets within a personal timeline rather than to extract few popular tweets.

Evaluation Plan

We plan to evaluate our design in both quantitative and qualitative manner with a prototype implementation. For the implementation, we are inspired by Processing.js², which is a javascript version of an open source programming language especially for visualization, Processing. It creates beautiful visualizations that are enough for the user study eliciting valuable feedback. Moreover, it is running on most modern web browsers even on the mobile devices without additional plugin. We thus implement our prototype as a web-based application with Processing.js. D3.js³ can be a good alternative.

For the quantitative evaluation, we collect user behavior recorded in application logs as an implicit feedback. We figure out how people interact with a prototype implementation. For instance, we inspect what size of communities people create by means of the value of partition density slider. We verify whether people pay more attention to key persons or interesting tweets by the number of clicks and reading time. In the long term, we examine our design helps not only to peruse but also to skim a timeline.

A survey with respondents complements the quantitative log analysis. The survey consists of Likert-scale questions and semi-structured interviews with a part of respondents. Here we aim to capture user behavior hardly recorded in application logs. In particular, the subjective satisfaction is our focal point. Moreover, we are interested in what factors are additionally needed to help people to read a personal timeline. We expect to get hints from semi-structured interviews.

²<http://processingjs.org/>

³<http://mbostock.github.com/d3/>

Finally, the level of sense of belongings people feel toward each community can be different. Understanding the mechanism behind how people feel a sense of belonging is a key to enhance our future design with improved community identification methods.

Summary and Conclusion

We improve the readability of a personal timeline by adding multiple social contexts to most of tweets. Our social contexts are community membership, key persons, and interesting tweets within a personal timeline. We plan to evaluate our design in both quantitative and qualitative manner with a web-based prototype implementation.

Acknowledgments

Yoonsung Hong and Sue Moon are supported by WCU (World Class University) program under the National Research Foundation of Korea and funded by the Ministry of Education, Science and Technology of Korea (Project No: R31-30007).

References

- Ahn, Y.-Y.; Bagrow, J.; and Lehmann, S. 2010. Communities and hierarchical organization of links in complex networks. *Nature* 466:761–764.
- Diakopoulos, N.; Choudhury, M. D.; and Naaman, M. 2012. Finding and assessing social media information sources in the context of journalism. In *Proc. CHI*.
- Fortunato, S. 2010. Community detection in graphs. *Physics Reports* 486:75–174.
- Gilbert, E., and Karahalios, K. 2009. Predicting tie strength with social media. In *Proc. CHI*.
- Granovetter, M. 1973. The strength of weak ties. *American Journal of Sociology* 78(6):1360–1380.
- Huberman, B. A.; Romero, D. M.; and Wu, F. 2009. Social networks that matter: Twitter under the microscope. *First Monday* 14(1).
- Kivran-Swaine, F.; Govindan, P.; and Naaman, M. 2011. The impact of network structure on breaking ties in online social networks: Unfollowing on Twitter. In *Proc. CHI*.
- Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is Twitter, a social network or a news media? In *Proc. WWW*.
- Kwak, H.; Chun, H.; and Moon, S. 2011. Fragile online relationship: A first look at unfollow dynamics in Twitter. In *Proc. CHI*.
- Kwak, H.; Moon, S.; and Lee, W. 2012. More of a receiver than a giver: Why do people unfollow in twitter? In *Proc. ICWSM*.
- McPherson, M.; Lovin, L. S.; and Cook, J. M. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27(1):415–444.
- Onnela, J.-P.; Saramäki, J.; Hyvönen, J.; Szabó, G.; Lazer, D.; Kaski, K.; Kertész, J.; and Barabási, A.-L. 2007. Structure and tie strengths in mobile communication networks. *PNAS* 104(18):7332–7336.
- Wasserman, S., and Faust, K. 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- Weng, J.; Lim, E.-P.; Jiang, J.; and He, Q. 2010. TwitterRank: finding topic-sensitive influential Twitterers. In *Proc. WSDM*.