

# The Clustering of Hot Topics on Plurk

Zong-Hua Yang\* and Hung-Yu Kao

Department of Computer Science and Information Engineering,  
National Cheng Kung University,  
No.1, Ta-Hsueh Road, Tainan, Taiwan  
bobyang79@livemail.tw, hykao@mail.ncku.edu.tw

*Received 17 March 2011; Revised 22 May 2011; Accepted 15 June 2011*

**Abstract.** In the new era of Web 2.0, Microblogging has become a new way to let users communicate with friends promptly and efficiently. Due to the huge amount messages generated in microblogging sites, it is an important task to find users and messages that you are interesting. We observed that users in Plurk suffer from the inconvenient and insufficient interaction interface. The recommended friends on Plurk are not easy to find. Users can not also easily find the topics in that they get interested to follow. Therefore, in this paper we aim to provide an easy mechanism for users to find other users they are interesting by clustering users with their discussion topics. We link the topics that are mentioned in the same time and are from the same theme. These topics will be considered a topic event. We call this method "Timing and Correlation Clustering" (TCC). We can then build a complete event list and recommend to users that are interested in this event and let users know this event and other users who are also interested in this event.

**Keywords:** Microblogging, event detection, clustering

## References

- [1] F. Wu and B. A. Huberman, "Novelty and Collective Attention," in *Proceedings of the National Academy of Sciences of the United States of America (PNAS'07)*, Vol. 104, No. 45, pp. 17599-17601, 2007.
- [2] S. Yardi, S. A. Golder, M. J. Brzozowski, "Blogging at Work and the Corporate Attention Economy," in *Proceedings of the 27th international Conference on Human Factors in Computing Systems, (CHI'09)*, Boston, Massachusetts, USA, pp. 2071-2080, 2009.
- [3] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, M. Hurst, "Cascading Behavior in Large Blog Graphs," in *Proc. of 7th SIAM International Conference on Data Mining (SDM'07)*, Minneapolis, Minnesota, pp. 1-21, 2007.
- [4] A. L. Barabási, "The Origin of Bursts and Heavy Tails in Human Dynamics," *Nature*, Vol. 435, pp. 207-211, 2005.
- [5] R. Crane and D. Sornette, "Robust Dynamic Classes Revealed by Measuring the Response Function of A Social System," *PNAS*, Vol. 105, No. 41, pp. 15649-15653, 2008.
- [6] K.J. Chen and W.Y. Ma, "Unknown Word Extraction for Chinese Documents," in *Proceedings of the 19th International Conference on Computational linguistics (COLING'02)*, U.S.A, Vol. 1, pp. 169-175, 2002.
- [7] W.Y. Ma and K.J. Chen, "A Bottom-up Merging Algorithm for Chinese Unknown Word Extraction," in *Proceedings of ACL, Second SIGHAN Workshop on Chinese Language Processing*, Sapporo, Japan, pp. 31-38, 2003.
- [8] J. Yang and J. Leskovec, "Patterns of Temporal Variation in Online Media," in *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM'11)*, Hong Kong, China, pp. 177-186, 2011.
- [9] J. L. Michail, J. Lin, M. Vlachos, E. Keogh, D. Gunopulos, "Iterative Incremental Clustering of Time Series," in *Proceedings of 9th International Conference on Extending Database Technology (EDBT'04)*, Crete, Greece, pp. 106-122, 2004.

---

\*Correspondence author

- [10] U.Brandes, "A Faster Algorithm for Betweenness Centrality," *Journal of Mathematical Sociology*, Vol. 25, No. 2, pp. 163-177, 2001.
- [11] A.Chin and M.Chignell, "A Social Hypertext Model for Finding Community in Blogs," in *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia (HT'06)*, Odense, Denmark, pp. 11-22, 2006.