

according to their relevance weight. In fact, there are 36 official accounts in cluster C, nearly all members of the community share the same geo location with the test organization. After ranking all users in this group on their RScore, we could easily figure out that lots of the top ranked users are real world members of the test organization from their personal description. Cluster D includes 2 official accounts, also includes some real world members, though not as many as cluster C. In cluster B, only a small number of top ranked users are relevant enough. As for cluster A, it keep expanding as k increasing, which suggests that it consists of mainly low RScore users. Obviously, cluster C and cluster D are quite relevant communities, while cluster A is not. Figure 2 show the RScore of the four representative communities. As Cluster C includes 36 official accounts(official accounts' RScore is $|U_o|$, i.e. 49), its top ranked members exhibit obvious advantage over the other sample communities.

The relevance weight emphasizes the relevance of the most relevant users of each community, it proofs that the top ranked users' relevance is highly effective in finding the most relevant communities. It can be noted from Figure 2 that the RScores of all four communities decrease sharply with the growing of K and quickly turn into a stable value afterward. In fact, in each community, there are still a large number of users that don't display any relevance from either their user information or RScore. Those users could be friends or relatives of the relevant users, but are not real world users of the organization. This means that the graph still keeps a lot of noise. A further de-noising approach is to prone nodes with low RScore in each community with certain threshold, only keep the most relevant users.

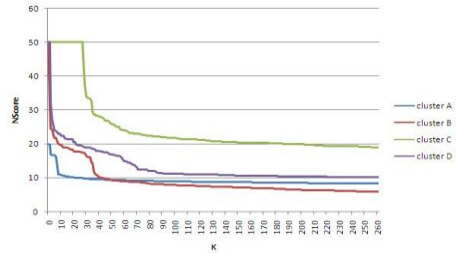


Fig. 3 top ranked users in the sample communities

5. Conclusion

In this paper, we propose a new approach based on community detection to mine real world relevant members on microblog platform. All bi-directional relationship are kept to find all true friends on the network. We also remove the noisy users according to their following information. After detect the user communities, we define communities' relevance weight to distinguish the most relevance user groups. Observation shows that the relevant groups still includes lots of noise, which could be eliminated according to their RScore.

6. References

- [1] <http://en.wikipedia.org/wiki/Twitter>.
- [2] Java, A., Song, X., Finn, T., & Tseng, B. Why we Twitter: Understanding microblogging usage and communities. WebKDD/SNA-KDD '07 Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. ACM New York, NY, USA, 2007. Pages 56-65
- [3] http://en.wikipedia.org/wiki/Community_structure
- [4] <https://dev.twitter.com/docs/api>
- [5] M. E. J. Newman (2004). Fast algorithm for detecting community structure in networks. *Phys. Rev. E* **69** (6): 066133. 2004