

The alphabet distribution reflects the user preference for specific letter. Some letter can also reflect a user's country or region to a certain extent. Therefore, the closer the letter distribution of two display names is, the more likely the two names belong to the same user.

(4) The evolutionary analysis results show that the above attributes remain unchanged over time.

(5) The similarity of two display names from Facebook and Foursquare is generally more striking. This is mainly due to the user tend to choose his display name closer to his real name on these two social networks.

7. CONCLUSION

A display name is a name that an individual chooses shown to other avatars on an OSN site. By comparing the display names from the same users and the different users, we know that the character similarity and the letter distribution of the positive instances are very high. The results of our measurements demonstrate that the same individual on different OSNs tends to use the same display names or similar display names. The presented attributes are very helpful for identifying whether accounts belong to the same individual or not based on their display names.

8. ACKNOWLEDGEMENTS

This research was supported in part by Shaanxi Provincial Natural Science Foundation Research, China under grant No. 2014JM2-6104.

9. REFERENCES

- [1] F. Buccafurri, G. Lax, A. Nocera, and D. Ursino. Discovering links among social networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 467–482. Springer, 2012.
- [2] Y. Chen, C. Zhuang, Q. Cao, and P. Hui. Understanding cross-site linking in online social networks. In *Proceedings of the 8th Workshop on Social Network Mining and Analysis*, page 6. ACM, 2014.
- [3] W. Cohen, P. Ravikumar, and S. Fienberg. A comparison of string metrics for matching names and records. In *Kdd workshop on data cleaning and object consolidation*, volume 3, pages 73–78, 2003.
- [4] D. Liu, Q. Wu, W. Han, and B. Zhou. User identification across multiple websites based on username features. *Chinese Journal of Computers*, 38(10):2028–2040, 2015.
- [5] O. Goga, H. Lei, S. H. K. Parthasarathi, G. Friedland, R. Sommer, and R. Teixeira. Exploiting innocuous activity for correlating users across sites. In *Proceedings of the 22nd international conference on World Wide Web*, pages 447–458. ACM, 2013.
- [6] F. Hussain and U. Qamar. Identification and correction of misspelled drugs' names in electronic medical records (emr). *Science and Technology Publications*, 2:333–338, 2016.
- [7] T. Iofciu, P. Fankhauser, F. Abel, and K. Bischoff. Identifying users across social tagging systems. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 522–525, 2011.
- [8] P. Jain, P. Kumaraguru, and A. Joshi. @ i seek'fb.me': identifying users across multiple online social networks. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1259–1268. ACM, 2013.
- [9] K. Kim, M. Khabsa, and C. L. Giles. Random forest dbscan for uspto inventor name disambiguation. *arXiv preprint arXiv:1602.01792*, 2016.
- [10] J. Liu, F. Zhang, X. Song, Y.-I. Song, C.-Y. Lin, and H.-W. Hon. What's in a name?: an unsupervised approach to link users across communities. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 495–504. ACM, 2013.
- [11] A. Malhotra, L. Totti, W. Meira Jr, P. Kumaraguru, and V. Almeida. Studying user footprints in different online social networks. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, pages 1065–1070. IEEE Computer Society, 2012.
- [12] M. Motoyama and G. Varghese. I seek you: searching and matching individuals in social networks. In *Proceedings of the eleventh international workshop on Web information and data management*, pages 67–75. ACM, 2009.
- [13] R. Ottoni, D. B. Las Casas, J. P. Pesce, W. Meira Jr, C. Wilson, A. Mislove, and V. Almeida. Of pins and tweets: Investigating how users behave across image-and text-based social networks. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 2014.
- [14] D. Perito, C. Castelluccia, M. A. Kaafar, and P. Manils. How unique and traceable are usernames? In *International Symposium on Privacy Enhancing Technologies Symposium*, pages 1–17. Springer, 2011.
- [15] E. Raad, R. Chbeir, and A. Dipanda. User profile matching in social networks. In *Network-Based Information Systems (NBIS), 2010 13th International Conference on*, pages 297–304. IEEE, 2010.
- [16] J. Vosecky, D. Hong, and V. Y. Shen. User identification across multiple social networks. In *2009 First International Conference on Networked Digital Technologies*, pages 360–365. IEEE, 2009.
- [17] P. Wang, W. He, and J. Zhao. A tale of three social networks: User activity comparisons across facebook, twitter, and foursquare. *IEEE Internet Computing*, 18(2):10–15, 2014.
- [18] R. Zafarani and H. Liu. Connecting users across social media sites: a behavioral-modeling approach. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 41–49. ACM, 2013.