Table 3: Comparison of different methods: In all MF based methods, the number of latent factors K is set to 10.

Method	*Precision*	*Recall*	*F1-score*
SVM	0.781 ± 0.027	0.952 ± 0.030	0.857 ± 0.020
SF + SVM	0.810 ± 0.029	0.949 ± 0.021	0.874 ± 0.020
MF + SVM	0.807 ± 0.036	0.966 ± 0.038	0.879 ± 0.029
MFSR + SVM	0.839 ± 0.026	0.978 ± 0.027	0.903 ± 0.013
SMFSR	0.851 ± 0.032	0.980 ± 0.037	0.911 ± 0.031

We plan to extend our work in the following directions. Firstly, we wish to use other matrix factorization methods to explore additional hidden factors to represent users. In particular, recent advances in tensor factorization could be used to factorize the performer-activity-receiver interaction tensor, which contains richer information than the user-activity matrix used in this work. However, the great sparseness in such representation poses challenges. Secondly, we will integrate our method as a new component into the existing immune system in Renren. Thirdly, we are particularly interested in identifying normal users whose accounts are hijacked by spammers to send spam to his social friends.

Acknowledgment

We thank the support of Hong Kong RGC GRF projects 621010 and 621211.

References

Benevenuto, F.; Magno, G.; Rodrigues, T.; and Almeida, V. 2010. Detecting spammers on twitter. In *Anti-Abuse and Spam Conference (CEAS)*.

Blanzieri, E., and Bryl, A. 2008. A survey of learning-based techniques of email spam filtering. *Artif. Intell. Rev.* 29(1):63–92.

Brown, G.; Howe, T.; Ihbe, M.; Prakash, A.; and Borders, K. 2008. Social networks and context-aware spam. In *ACM Conference on Computer Supported Cooperative Work (CSCW)*, 403–412.

Chang, C.-C., and Lin, C.-J. 2011. Libsvm: A library for support vector machines. *ACM TIST* 2(3):27.

Duch, J., and Arenas, A. 2005. Community detection in complex networks using extremal optimization. *Physical Review E* 72(2):027104.

Gyöngyi, Z., and Garcia-Molina, H. 2005. Web spam taxonomy. In *Adversarial Information Retrieval on the Web (AIRWeb)* , 39–47.

Koren, Y. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *KDD*, 426–434.

Kwak, H.; Lee, C.; Park, H.; and Moon, S. B. 2010. What is twitter, a social network or a news media? In *WWW*, 591–600.

Landauer, T.; McNamara, D.; Dennis, S.; and Kintsch, W. 2007. *Handbook of latent semantic analysis.* Lawrence Erlbaum Associates Publishers.

Lee, K.; Caverlee, J.; and Webb, S. 2010. Uncovering social spammers: social honeypots + machine learning. In *SIGIR*, 435–442.

Lee, K.; Eoff, B. D.; and Caverlee, J. 2011. Seven months with the devils: A long-term study of content polluters on twitter. In *The International AAAI Conference on Weblogs and Social Media (ICWSM)*, 185–192.

Li, F., and Hsieh, M.-H. 2006. An empirical study of clustering behavior of spammers and group-based anti-spam strategies. In *Anti-Abuse and Spam Conference (CEAS)*.

Liu, Z.; Shu, G.; Li, N.; and Lee, D. 2006. Defending against instant messaging worms. In *In Proceedings of IEEE GLOBECOM*, 1–6.

Ma, H.; Liu, C.; King, I.; and Lyu, M. R. 2011. Probabilistic factor models for web site recommendation. In *SIGIR*, 265–274.

Rendle, S. 2012. Factorization machines with libfm. *ACM Transactions on Intelligent Systems and Technology (TIST 2012)* 3(3).

Rennie, J. D. M. 2004. Smooth hinge classification. http://people.csail.mit.edu/jrennie/writing.

Singh, A. P., and Gordon, G. J. 2008. Relational learning via collective matrix factorization. In *KDD*, 650–658.

Stein, T.; Chen, E.; and Mangla, K. 2011. Facebook immune system. In *Proceedings of the EuroSys Social Network Systems (SNS)*, 8:1–8:8.

Von Ahn, L.; Maurer, B.; McMillen, C.; Abraham, D.; and Blum, M. 2008. recaptcha: Human-based character recognition via web security measures. *Science* 321(5895):1465.

Wang, A. H. 2010. Don't follow me - spam detection in twitter. In *Intl. Conf. on Sec. and Crypto.*, 142–151.

Xu, Q.; Xiang, E.; Du, J.; Zhong, J.; and Yang, Q. 2012. Sms spam detection using content-less features. *IEEE Intelligent Systems* http://doi.ieeecomputersociety.org/10.1109/MIS.2012.3.

Yang, Z.; Wilson, C.; Wang, X.; Gao, T.; Zhao, B. Y.; and Dai, Y. 2011. Uncovering social network sybils in the wild. In *Internet Measurement Conference*, 259–268.

Zhu, S.; Yu, K.; Chi, Y.; and Gong, Y. 2007. Combining content and link for classification using matrix factorization. In *SIGIR*, 487–494.

Zinkevich, M.; Smola, A. J.; and Langford, J. 2009. Slow learners are fast. In *NIPS*, 2331–2339.