an online video in a video network connected based on their content. However, [6, 7] also require the definition on the node and edge functions. Moreover, the HeteroSales framework also incorporates the results of external classifiers to the label propagation, which further improves its flexibility and robustness. In [18, 20], the authors use a propagation-based method to estimate product quality on a heterogeneous network. However, they construct the graph from raw data, instead of combining multiple graphs.

Finally, we also find several related articles [9,14,24] focusing on improving the effectiveness of selling process. For example, [24] studies the problem how to combine several products into one bundle and recommend it to customers through emails. Since discounts are usually also applied to these product bundles, users may more likely to purchase them, if the bundle contains different product he/she may need. However, all these research works focus on online sales process, which require a large number of training data to fully unleash the power of their proposed methods.

6. CONCLUSION

In reality, online and offline sales are the two important parts in the modern selling activities. Unlike online sales selling a small quantity of a product to each customer through methods like sending recommendation emails, offline sales normally target at enterprise customers to sell a large quantity of the product in one single deal. Since the offline sales usually require human contact between the sales agents and the representatives of the buying company, it is more costly and requires longer time and additional human labor to collect the data to be used to research on this problem. Therefore, how to find a new enterprise customer usually depends on the experience and insights of the sales personnels, and computer scientists have rarely get involved in this problem. In this paper, we have introduced a method to utilize the information in an online heterogeneous social network to improve the effectiveness of offline sales. We propose a two-step framework, HeteroSales, to achieve this goal. The HeteroSales first constructs a Company Homophily Graph (CHG) through learning from semantics based meta-paths in the social network, and then adopts a label propagation algorithm on it to find new potential enterprise customers. Based on the offline sales records of a third-party company and a large professional social network, LinkedIn, we introduce statistical findings to show that not all the users' social connections in a network can be helpful for the HeteroSales. For example, we found that those people who are more familiar with the product are more likely to have indicative social connections to help us find new enterprise customers. Finally, based on the extracted data set, we conduct extensive experiments to evaluate the proposed HeteroSales, and show it can constantly outperform other baselines in this task.

7. ACKNOWLEDGEMENT

This work is supported in part by LinkedIn Corporation, NSF through grants III-1526499, CNS-1115234, and OISE-1129076, and Google Research Award.

8. REFERENCES

[1] L. Breiman. Random forests. *Machine learning*, 2001.

[2] L. Breiman. Manual on setting up, using, and understanding random forests v3. 1. *Statistics Department University of California Berkeley, CA, USA*, 2002.

[3] L. Clark, D. Pregibon, J. Chambers, and T. Hastie. Tree-based models. *Statistical models in S*, 1992.

[4] A. J. Dobson and A. Barnett. *An introduction to generalized linear models*. Chapman and Hall, 1990.

[5] D. A. Freedman. *Statistical models: theory and practice*. Cambridge University Press, 2009.

[6] Q. Hu, G. Wang, and P. S. Yu. Assessing the longevity of online videos: A new insight of a video's quality. In *DSAA '14*, pages 1–10, 2014.

[7] Q. Hu, G. Wang, and P. S. Yu. Deriving latent social impulses to determine longevous videos. In *WWW' 14*, 2014.

[8] Q. Hu, S. Xie, S. Lin, W. Fan, and P. S. Yu. Frameworks to encode user preferences for inferring topic-sensitive information networks. In *SDM '14*, 2014.

[9] J. Karat. *Designing personalized user experiences in eCommerce*. Springer Science & Business Media, 2004.

[10] X. Kong, P. S. Yu, Y. Ding, and D. J. Wild. Meta path-based collective classification in heterogeneous information networks. In *CIKM '12*, 2012.

[11] P. McCullagh and J. A. Nelder. *Generalized linear models*. CRC press, 1989.

[12] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 2001.

[13] B. D. Ripley. *Modern applied statistics with S*. Springer, 2002.

[14] J. B. Schafer, J. Konstan, and J. Riedl. Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce*, 1999.

[15] Y. Sun, R. Barber, M. Gupta, C. C. Aggarwal, and J. Han. Co-author relationship prediction in heterogeneous bibliographic networks. In *ASONAM '11*, 2011.

[16] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *VLDB '11*, 2011.

[17] G. Wang, Q. Hu, and P. S. Yu. Influence and similarity on heterogeneous networks. In *CIKM '12*, 2012.

[18] G. Wang, S. Xie, B. Liu, and P. S. Yu. Review graph based online store review spammer detection. In *ICDM '11*, 2011.

[19] B. L. Welch. The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, 1947.

[20] S. Xie, Q. Hu, J. Zhang, J. Gao, W. Fan, and P. S. Yu. Robust crowd bias correction via dual knowledge transfer from multiple overlapping sources. In *2015 IEEE International Conference on Big Data*, 2015.

[21] X. Yu, X. Ren, Y. Sun, Q. Gu, B. Sturt, U. Khandelwal, B. Norick, and J. Han. Personalized entity recommendation: A heterogeneous information network approach. In *WSDM '14*, 2014.

[22] X. Yu, X. Ren, Y. Sun, B. Sturt, U. Khandelwal, Q. Gu, B. Norick, and J. Han. Recommendation in heterogeneous information networks with implicit user feedback. In *RecSys '13*, 2013.

[23] J. Zhang, P. S. Yu, and Z.-H. Zhou. Meta-path based multi-network collective link prediction. In *KDD '14*, 2014.

[24] T. Zhu, P. Harrington, J. Li, and L. Tang. Bundle recommendation in ecommerce. In *SIGIR ' 14*, 2014.

[25] X. Zhu and Z. Ghahramani. Learning from labeled and unlabeled data with label propagation. Technical report, Carnegie Mellon University, 2002.

[26] X. Zhu, Z. Ghahramani, J. Lafferty, et al. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML '03*, 2003.