Cross View Link Prediction by Learning Noise-resilient Representation Consensus

Xiaokai Wei*, Linchuan Xu†, Bokai Cao* and Philip S. Yu*
* Department of Computer Science, University of Illinois at Chicago,
{xwei2, caobokai, psyu}@uic.edu
† Department of Computing, The Hong Kong Polytechnic University,
cslcxu@comp.polyu.edu.hk

ABSTRACT

Link Prediction has been an important task for social and information networks. Existing approaches usually assume the completeness of network structure. However, in many real-world networks, the links and node attributes can usually be partially observable. In this paper, we study the problem of **Cross View Link Prediction (CVLP)** on partially observable networks, where the focus is to recommend nodes with only links to nodes with only attributes (or vice versa). We aim to bridge the information gap by learning a robust consensus for link-based and attribute-based representations so that nodes become comparable in the latent space. Also, the link-based and attribute-based representations can lend strength to each other via this consensus learning. Moreover, attribute selection is performed jointly with the representation learning to alleviate the effect of noisy high-dimensional attributes. We present two instantiations of this framework with different loss functions and develop an alternating optimization framework to solve the problem. Experimental results on four real-world datasets show the proposed algorithm outperforms the baseline methods significantly for cross-view link prediction.

1. INTRODUCTION

In the past decade, there have been an increasing number of information networks from a wide range of domains. Study on computer networks, biological and social networks has attracted great attention from the research community [7] [4] [26]. Link prediction [1, 2], which aims at recommending potential links between network nodes, is an important step to understand and study the characteristics of these networks. For instance, in bioinformatics, by predicting protein interaction links, one does not need to conduct expensive experiments on all possible pairs and can spend the resource wisely on the most likely interaction. For social media websites, such as Facebook and Twitter, it is fundamental to grow the user base and enhance user engagement with link prediction techniques. For security analysts/agencies, predicting (currently unobserved) links can reveal hidden but important relationship among terrorists and provides additional insights for understanding organizational structures of terrorist-attack activities.

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3038912.3052575

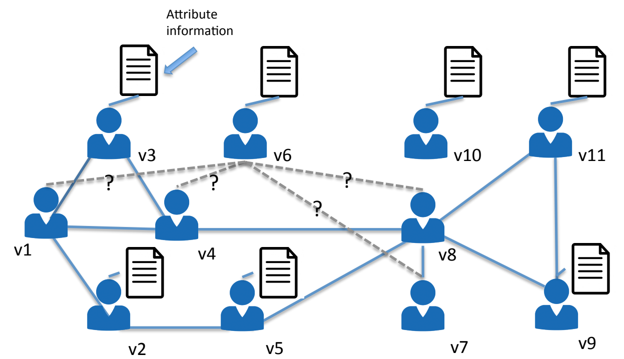


Figure 1: An example of networks with partially observable links and attributes

Many methods have been proposed for the task of link prediction [11, 1, 9, 2]. However, in various social and information networks, it is common that certain nodes do not have any link information revealed [25] and make these methods not applicable:

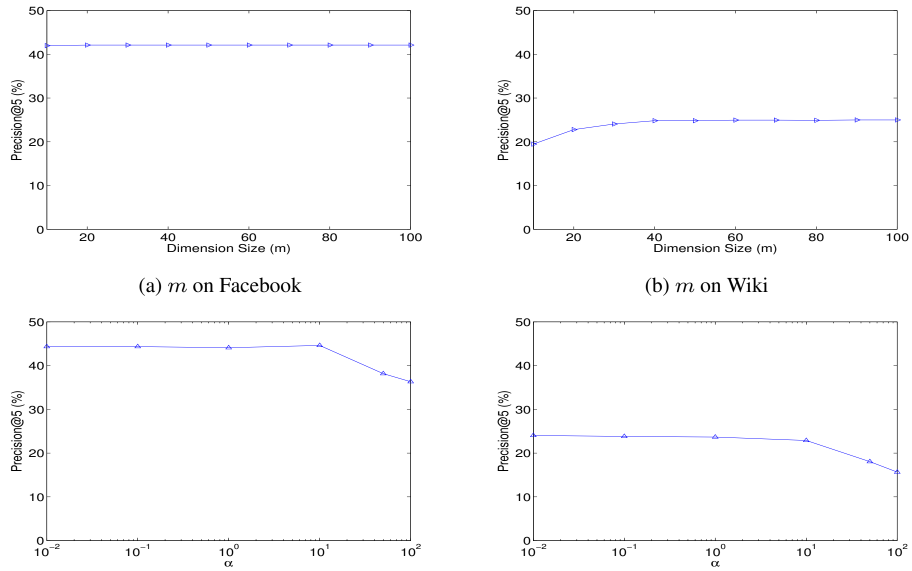
- In real-world social networks (e.g., Twitter, Facebook and LinkedIn), link prediction for new users usually has the challenge of cold start problem, since these users do not have any connection. Besides, some users may choose a strict privacy setting that restricts the visibility of their connections, personal information or posts[1][2]. Recommending links in such a partially observable setting could enhance user experience.

- In bioinformatics information networks, for example, studying protein interaction could help researchers better understand many biological processes. However, it is infeasible to collect all the experimental data for all the possible pairs of protein.

- In terrorist-attack networks, nodes represent terrorist activities and links represent terrorist attacks in which the same terrorist group is involved. Detecting hidden links in these networks is useful for understanding the underlying structure of terrorist-attack activities. However, the complete linkages between attacks are highly difficult to resolve [13].

Nonetheless, nodes in many social/information networks are often equipped with features/attributes, such as user attributes in so-

[1] https://www.facebook.com/help/325807937506242/
[2] https://help.linkedin.com/app/answers/detail/a_id/52



(a) m on Facebook

(b) m on Wiki

(c) α on Facebook

(d) α on Wiki

Figure 3: Parameter sensitivity for MM-NRCL

[13] W. Lin, X. Kong, P. S. Yu, Q. Wu, Y. Jia, and C. Li. Community detection in incomplete information networks. In *Proceedings of the 21st International Conference on World Wide Web*, pages 341–350. ACM, 2012.

[14] L. Lu and T. Zhou. Link prediction in complex networks: A survey. *Physica A*, 390(6):1150–1170, 2011.

[15] A. K. Menon and C. Elkan. Link prediction via matrix factorization. In *Proceedings of the ECML/PKDD 2011*, 2011.

[16] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[17] K. T. Miller, T. L. Griffiths, and M. I. Jordan. Nonparametric latent feature models for link prediction. In *NIPS*, pages 1276–1284, 2009.

[18] F. Nie, H. Huang, X. Cai, and C. H. Q. Ding. Efficient and robust feature selection via joint l2, 1-norms minimization. In *NIPS*, pages 1813–1821, 2010.

[19] B. Perozzi, R. Al-Rfou', and S. Skiena. Deepwalk: online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 701–710. ACM, 2014.

[20] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *UAI*, 2009.

[21] P. Sen, G. M. Namata, M. Bilgic, L. Getoor, B. Gallagher, and T. Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–106, 2008.

[22] Y. Sun, J. Han, C. C. Aggarwal, and N. V. Chawla. When will it happen?: relationship prediction in heterogeneous information networks. In *WSDM*, pages 663–672. ACM, 2012.

[23] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1067–1077, 2015.

[24] F. Wang, T. Li, X. Wang, S. Zhu, and C. H. Q. Ding. Community discovery using nonnegative matrix factorization. *Data Min. Knowl. Discov.*, 22:493–521, 2011.

[25] X. Wei, B. Cao, W. Shao, C.-T. Lu, and P. S. Yu. Community detection with partially observable links and node attributes. In *IEEE International Conference on Big Data*, 2016.

[26] X. Wei, B. Cao, and P. S. Yu. Unsupervised feature selection on networks: A generative view. In *AAAI*, 2016.

[27] X. Wei, S. Xie, and P. S. Yu. Efficient partial order preserving unsupervised feature selection on networks. In *SDM*, pages 82–90. SIAM, 2015.

[28] L. Xu, X. Wei, J. Cao, and P. S. Yu. Embedding of embedding (eoe) : Embedding for coupled heterogeneous networks. In *WSDM*, 2017.

[29] K. Yu, W. Chu, S. Yu, V. Tresp, and Z. Xu. Stochastic relational models for discriminative link prediction. In *NIPS*, pages 1553–1560, 2006.

[30] J. Zhang, P. S. Yu, and Z.-H. Zhou. Meta-path based multi-network collective link prediction. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1286–1295, 2014.