

REVIEW

Open Access



Latent feature models for large-scale link prediction

Jun Zhu* and Bei Chen

*Correspondence:
dcszj@tsinghua.edu.cn
Department of Computer Science & Technology, Center for Bio-Inspired Computing Research, Tsinghua National Lab for Information Science & Technology, State Key Lab of Intelligence Technology & System, Tsinghua University, 100084 Beijing, China

Abstract

Link prediction is one of the most fundamental tasks in statistical network analysis, for which latent feature models have been widely used. As large-scale networks are available in various application domains, how to develop effective models and scalable algorithms becomes a new challenge. In this paper, we provide a review of the recent progress on latent feature models for the task of link prediction in large-scale networks, including the nonparametric Bayesian models which can automatically infer the latent social dimensions and the max-margin models which can learn strongly discriminative latent features for highly accurate predictions as well as dealing with the imbalance issue in large real networks. We also review the progress on scalable algorithms for posterior inference in such models, including stochastic variational methods and MCMC methods with data augmentation.

Keywords: Latent feature model, Social network, Link prediction

Introduction

As the pervasiveness and scope of network data (e.g., social networks, biological gene networks, document networks, citation networks, etc.) increase, statistical network analysis has attracted a great deal of attention. Those networks are typically represented as a graph, whose vertices represent entities and edges represent links or relationships between these entities. Given a network, it is very useful to answer the query that: which new interactions among entities are likely to occur given some partially observed information? This problem is known as *link prediction* [1]. Link prediction is of significant importance in network analysis with extensive applications, where latent feature models have been widely used. Compared with other methods, latent feature models can learn expressive representations from network structures to achieve state-of-the-art prediction performance. However, the link prediction problem meets a lot of challenges as the networks become larger, which have motivated the development on effective models and scalable algorithms.

Large-scale networks

With the fast development of Internet and information science, there are more and more large-scale networks to be analyzed. Table 1 shows the statistics of a diverse set of real networks that are often used for estimating the performance of methods in network analysis. Each network has its own characteristics. For example, WebKB [2] is a hyperlink network,



Table 1 Statistics of different networks, where positive links mean relationships between entities in a network

Dataset	Entities	Positive links	Sparsity rate
Kinship [42]	104	415	4.1 %
NIPS [43]	234	1196	2.2 %
WebKB [2]	877	1608	0.21 %
AstroPh [44]	17903	391462	0.12 %
CondMat [44]	21363	182684	0.040 %
Gowalla [3]	196591	1900654	0.0049 %
US Patent [4]	3774768	16522438	0.00012 %

whose entities represent webpages from the computer science departments of different universities and the webpage contents provide rich information. So both the network structure and the text contents can help with analysis. Gowalla [3] is a social network, where the user locations of check-ins are collected. Combining the relationships between users and the extra geographical information, we can analyze user movements and friendships. US Patent [4] is a massive citation network, whose entities represent patents and links represent citations between patents. There is not extra information for entities, so the sparse network structure is the only thing we can use.

Several challenges exist when analyzing such large networks. First, a large number of vertices and edges will lead to the increase of computational complexity, which asks for efficient algorithms. Second, the relationships between entities will become more complicated. If we represent each entity with a feature vector, we need a space with a higher dimension. That is, both entities and relationships in large-scale networks are harder to be depicted. Finally, in real networks the positive links are often much fewer than the negative ones, which leads to serious imbalance issues in supervised learning. As shown in Table 1, positive links are much sparser with larger networks. Therefore, it is imperative for improving models to adapt in large-scale network analysis.

Link prediction

Link prediction is one of the most fundamental problems in network analysis. For static networks, it is often defined as predicting the missing links from a partially observed network topology (and maybe some attributes as well), while for dynamic networks, it is typically defined as predicting network structure at the next time $t + 1$ given the structures up to the current time t . Link prediction is of significant application value in many areas (see [5] for a comprehensive survey). In social networking websites like Facebook and LinkedIn, link prediction can be used to predict the existence of friendships between pairs of users [6]. In citation networks, link prediction can be applied in suggesting the most likely coauthorships in the near future [7]. In bioinformatics, link prediction offers a cheaper way to predict if two proteins will interact than the laboratory experiments [8, 9]. Moreover, the link prediction approaches can be applied to user-item recommendation in collaborative networks [10], and describe the link structure in document networks [11].

In the rest of the paper, we first survey some existing prominent approaches for link prediction, followed by the recent progress in latent feature models for large-scale link prediction with several experimental results. Finally, we conclude and discuss about future work.

Background

A wide range of models have been proposed for link prediction. In this section, we survey some prominent approaches.

Proximity based models

The early work on link prediction has been focused on designing good proximity (or similarity) measures between nodes, using features related to certain topological properties of the graph. For instance, graph-based models [1] compute a measure for each pair of nodes and rank the unobserved links in a descending order. Popular measures include common neighbors, Jaccard's coefficient [12], Adamic/Adar [13], Katz [14], etc. These methods are unsupervised and depend heavily on the manually designed proximity measure.

Well-conceived feature models

Supervised learning methods have also been popular for link prediction [15, 16]. These methods learn predictive models on labeled training data with a set of manually designed features that capture the statistics of the network. For example, Hasanand et al. [15] and Lichtenwalter et al. [17] identify a set of features and cast the link prediction problem as a classification task. Backstrom and Leskovec [6] use random walks to combine the information from the network structure with node and edge attributes. Although we can design effective domain-specific features from the graph topology as well as node attributes, this process can be too time demanding and only restrictive to some particular application domains.

Latent class models

Latent class models assume that there is a number of clusters (or classes) underlying the observed entities and each entity belongs to certain clusters. The observed link between two entities is determined by their cluster assignments (or social roles). The early work of stochastic block models [18] is a representative work that places a probability distribution over the clusters and reveals a soft clustering of the entities by posterior inference. Later advancements are with nonparametric techniques, such as the infinite relational model [19] and the infinite hidden relational model [20], which allow a potentially infinite number of clusters. The mixed membership stochastic block model (MMSB) [21] increases the expressiveness of latent class models by allowing entities to be members of multiple communities. But the number of latent communities is required to be externally specified. The nonparametric extension of MMSB is a hierarchical Dirichlet process relational model (HDPR) [22], which allows mixed membership in an unbounded number of latent communities.

Latent feature models for link prediction

Latent feature models (LFM) are powerful to link prediction, in which each entity is assumed to be associated with a feature vector which is unobserved (thus latent). Then the probability of a link is determined by the interactions among such latent features. Latent feature models have been popular due to three reasons: (1) they can model the social phenomena in the real world; (2) they can discover the underlying structures of the relations between entities and sometimes correlated with extra known attributes; (3) they can make accurate predictions of the link structures using

automatically learned latent features and have been shown to give state-of-the-art performance.

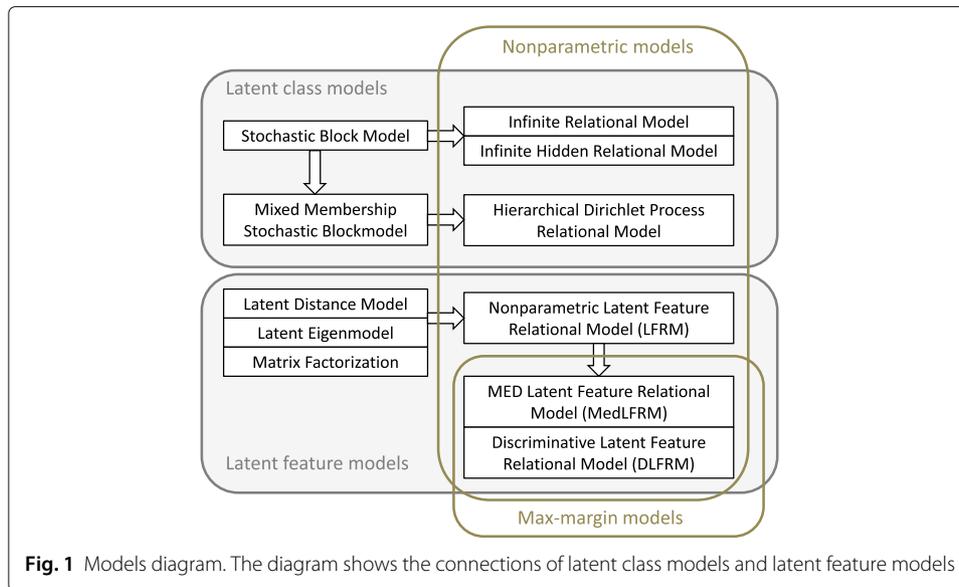
Formally, in a network with N entities, latent feature models represent each entity i by a K -dimensional feature vector $Z_i \in \mathbb{R}^K$, which is a point in a latent feature space. In some cases, we have extra known attributes $X_{ij} \in \mathbb{R}^D$ to help with prediction. Let Y denote the $N \times N$ binary link indicator matrix, while $y_{ij} = \pm 1$ indicates the positive link or negative link from entity i to entity j . Then, the model's score for a link (i, j) can be generally defined as:

$$f((i, j); Z_i, Z_j, X_{ij}) = \Phi(\psi(Z_i, Z_j) + \beta^\top X_{ij} + b), \quad (1)$$

where $\psi(Z_i, Z_j)$ is a function that measures the similarity of entity i and entity j in the latent space, $\beta^\top X_{ij}$ is the score that considers the influence of known attributes, β is the real-valued weight, and b is an offset. $\Phi(\cdot)$ is a link function and the common choices of $\Phi(\cdot)$ include the sigmoid function, the probit function and so on.

Link prediction problem in dynamic networks is typically defined to predict the evolutions of the network structure. We focus on static networks, where we can only observe a part of links, containing positive links and negative links. Some links are unobserved to be predicted. It is similarly formulated as predicting missing (unobserved) entries in a partially observed link matrix Y . The formulation in Eq. (1) covers various interesting latent feature models. The latent distance model [23] measures the similarity of two entities using a distance function $d(\cdot)$ in the latent feature space as $\psi(Z_i, Z_j) = -d(Z_i, Z_j)$. The latent eigenmodel [24], which based on the idea of the eigenvalue decomposition, defines $\psi(Z_i, Z_j) = Z_i^\top D Z_j$, where D is a diagonal matrix that is estimated from observed data. As the latent feature vectors form a latent feature matrix $Z = [Z_1^\top; \dots; Z_N^\top]$, the matrix factorization approach [25] can be seen as a kind of latent feature models, which defines $\psi(Z_i, Z_j) = Z_i^\top \Lambda Z_j$.

However, there are still some challenges in link prediction problems that the above models can not deal with. 1) **Model complexity.** The dimensionality of the latent feature space K is assumed to be known a priori. However, this assumption is often unrealistic for data analysis, especially when dealing with large networks. A typical way is using model selection (e.g., cross-validation or likelihood ratio test), which may be computationally prohibitive by comparing many candidate models. 2) **Imbalance Issue.** As we mentioned before, the real networks are often very sparse, where the positive links are much fewer than the negative ones. That leads to serious imbalance issues in supervised learning. 3) **Performance.** It is challenging to think about how to improve the models and design inference algorithm to obtain better prediction performance. 2) **Scalability.** In the real world, there are more and more large-scale networks to be analyzed, which needs scalable models. But most previous proposed latent feature models can not meet the requirements. In this section, we will review nonparametric latent feature relational models (LFRM) [7], and two extended models (i.e., MedLFRM [26] and DLFRM [27]). We can see how these models can meet the above four challenges. Because of the distributed representation, LFMs are more flexible than latent class models. Figure 1 shows the connections of latent class models and latent feature models that we introduce.

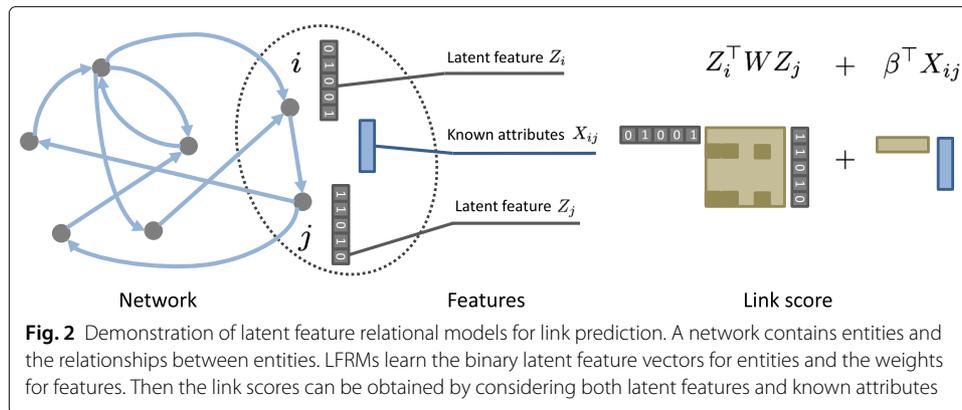


Nonparametric latent feature relational model

Bayesian nonparametric techniques have shown promise in bypassing model selection and automatically resolving the model complexity from empirical data by imposing an appropriate stochastic process prior on a rich class of models, such as the mixture models with an infinite number of components [28], or the latent factor models with an infinite number of features [29]. For link prediction, the recently developed nonparametric latent feature relational models (LFRM) [7] leverage the advance of Bayesian nonparametric methods to automatically resolve the unknown dimensionality of the feature space by applying a flexible nonparametric prior. It assumes that each entity i has an infinite number of binary features, that is $Z_i \in \{0, 1\}^\infty$, then $Z = [Z_1^T; \dots; Z_N^T]$ is a latent feature matrix with N rows and an unbounded number of columns. So each column of Z corresponds to a feature and $z_{ik} = 1$ if entity i has feature k , $z_{ik} = 0$ otherwise. Indian Buffet Process (IBP) [29] prior is used as a prior of Z to produce sparse latent feature vector for each entity, which defines a stochastic process that generates sparse binary matrices of an unbounded number of columns. According to Eq. (1), the prediction function of LFRM can be defined as:

$$f((i, j); Z_i, Z_j, X_{ij}, W, \beta) = \Phi(Z_i^T W Z_j + \beta^T X_{ij}), \tag{2}$$

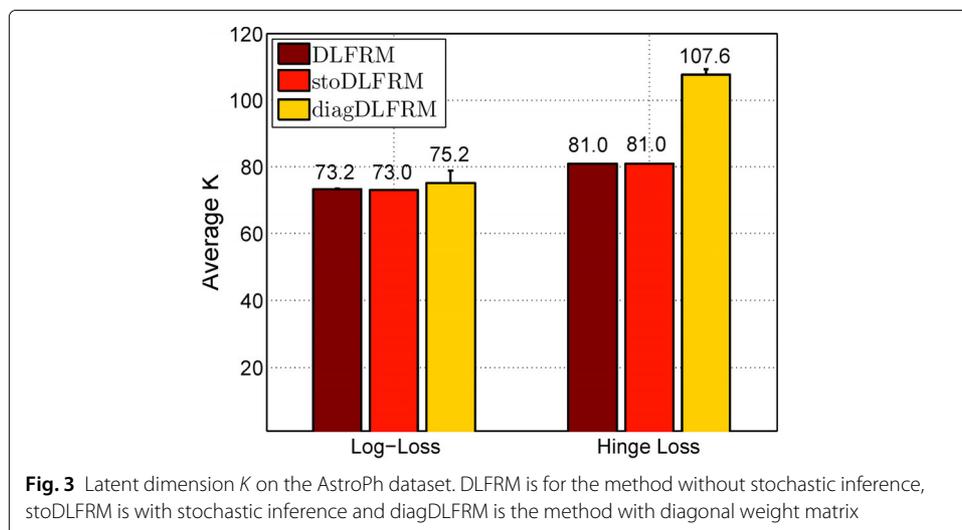
where W is a real-valued weight matrix. Each entry $w_{kk'}$ in W denotes the weight of the feature pair k and k' . That is, if entity i has feature k and entity j has feature k' , then the value $w_{kk'}$ will be added to the link score. Note that $w_{kk'}$ can be negative, which means the mutual exclusion of these two features. So $Z_i^T W Z_j$ is the score that considers the interaction patterns among the latent features. Although the dimension of Z_i is infinite, we only take the finite number of K non-zero columns to represent entities, while K is determined automatically during learning. A demonstration of LFRM for link prediction is shown in Fig. 2.



Discriminative latent feature relational model

Discriminative nonparametric latent feature relational models (DLFRM) [27], an extension of LFRM, employ regularized Bayesian inference (RegBayes) [30] to handle the imbalance issue in the real networks and learn discriminative latent features. With the prediction function in Eq. (2), the link likelihood can be defined as the product over all pairs of entities $p(Y|Z, X, W, \beta) = \prod_{i,j \in \mathcal{I}} \Phi(Z_i^T W Z_j + \beta^T X_{ij})$, where \mathcal{I} is the set of training links (observed links). For fully Bayesian inference, the feature matrix Z follows IBP, and the weight matrix W and β are also treated as random and assumed to be isotropic Gaussian distribution. Once Z , W and β are given, the predictions can be made using the sign rule, $\hat{y}_{ij} = \text{sign}(Z_i^T W Z_j + \beta^T X_{ij})$, and the error is measured on the training data. As the nonparametric prior IBP is employed, the latent dimension K can be determined automatically. Figure 3 shows the average K on AstroPh dataset, where stoDLFRM is DLFRM with stochastic algorithm and diagDLFRM is with diagonal W .

As there are not conjugate priors on W and β , exact posterior inference is intractable. DLFRM exploits the ideas of data augmentation with simpler Gibbs sampling [31, 32] under the regularized Bayesian inference (RegBayes) framework [30]. RegBayes treats inferring the posterior distribution as solving an optimization problem with a non-negative regularization parameter c . The parameter c balances the prior part and the



posterior regularization part. c can be controlled to deal with the imbalance issue in real networks. For example, we can choose a larger c value for the fewer positive links and a relatively smaller c for the larger negative links. The sensitivity of c^+/c^- on AstroPh dataset is shown in Fig. 4, which demonstrates the effectiveness of dealing with the imbalance issue. Data augmentation technique introduces auxiliary variables λ , so that the likelihood can be represented as the marginal of a higher dimensional distribution that includes λ . With the proper design of likelihood, we can directly obtain posterior distributions as a mixture of Gaussian components and develop efficient Gibbs sampling algorithms. In order to make DLFRM scalable, stochastic gradient Langevin dynamics [33] is employed to get the approximate sampler of W , so that DLFRM can handle large networks with hundreds of thousands of entities and millions of links.

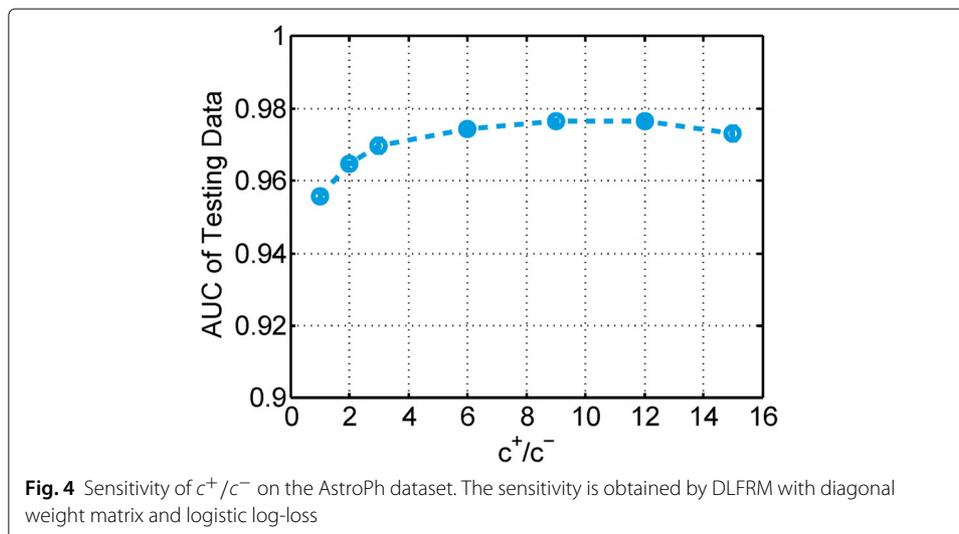
MED latent feature relational model

In [34], maximum entropy discrimination latent feature relational model (MedLFRM) unites the ideas of max-margin learning and Bayesian nonparametrics, which directly minimizes the hinge loss that measures the quality of link prediction. An averaging classifier is defined as making the predictions using the sign rule, $\hat{y}_{ij} = \text{sign}(\mathbb{E}_q[Z_i^\top W Z_j + \beta^\top X_{ij}])$. In learning, it adopts the hinge-loss as a surrogate to the training error. MedLFRM is defined as solving the problem:

$$\min_{p(Z,W,\beta) \in \mathcal{P}} \text{KL}(p(Z,W,\beta) || p_0(Z,W,\beta)) + c\mathcal{R}_\ell(p(Z,W,\beta)), \tag{3}$$

where \mathcal{P} denotes the space of normalized distribution, $\mathcal{R}_\ell(p(Z,W,\beta)) = \sum_{(i,j) \in \mathcal{I}} \max(0, \ell - y_{ij} \mathbb{E}_q[Z_i^\top W Z_j + \beta^\top X_{ij}])$ is the hinge-loss, and c is the regularization parameter that can be controlled for handling the imbalance issue as DLFRM.

For IBP prior, there is an equivalent augmented stick-breaking construction [35]. With the stick-breaking representation of IBP and the truncated mean-field assumption, a variational algorithm is presented for posterior inference. Although variational methods are approximate, they are usually more efficient and also have an objective to monitor the convergence behavior. For the AstroPh dataset in Table 1, 90 % of the positive links are



randomly selected for training and the number of negative training links is 10 times the number of positive training links. The testing set contains the remaining positive links and the same number of negative links. The AUC scores are shown in Fig. 5, where aMMSB is assortative MMSB [21], aHDPR is assortative HDP Relational model [22], stoDLFRM^l is DLFRM with stochastic algorithm and logistic log-loss, and stoDLFRM^h is with hinge-loss. For MedLFRM, the truncated K should be set beforehand. We can observe that DLFRM and MedLFRM outperform all the baselines, that demonstrates the superior performance of the two models.

Moreover, an efficient stochastic variational inference [36] algorithm was proposed to scale MedLFRM up to large-scale networks with millions of entities and tens of millions of positive links [26]. As far as we know, none of the previous Bayesian latent feature models can handle the two largest networks in Table 1. The results on US Patent dataset are shown in Table 2. In the experiments, 21,796,734 links are extracted, containing all the positive links and randomly sampled negative links. Then 17,437,387 links are uniformly chosen for training. Three proximity based methods are used as baselines. We can observe that MedLFRM achieve significantly better performance.

Overall, DLFRM and MedLFRM can meet the four challenges as mentioned before. They present discriminative models to achieve state-of-the-art performance, deal with the imbalance issue via RegBayes, handle large-scale networks using stochastic algorithms, and at the same time employ nonparametric techniques to determine the latent dimension automatically.

Discussions and conclusions

We firstly discuss the large-scale networks analysis and the challenges they meet, along with the importance and usefulness of the link prediction problem. In order to tackle the challenges in large-scale link prediction, progresses have been made in latent feature models. We review the latent feature models, especially LFRM, which impose IBP prior to solve the unknown dimension problem. Then two recently improved model DLFRM and MedLFRM are introduced under RegBayes with their efficient inference using stochastic algorithm. The experimental results demonstrate that these improved latent feature

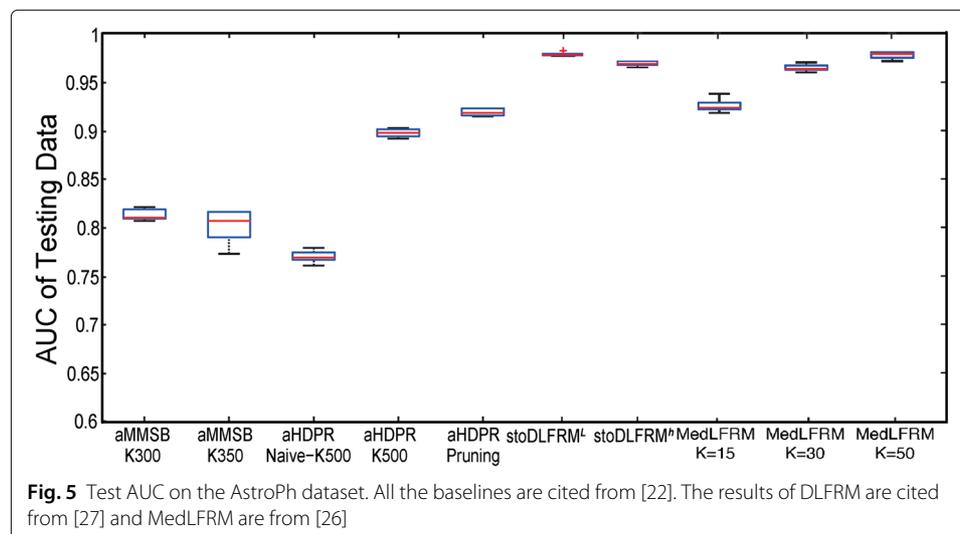


Table 2 Results on the US Patent dataset

Method	Test AUC	Running time (s)
CN	0.619 ± —	87.00 ± 2.58
Jaccard	0.618 ± —	52.15 ± 2.46
Katz	0.639 ± —	21975 ± 259
MedLFRM (K=15)	0.653 ± 0.0033	2787 ± 132
MedLFRM (K=30)	0.670 ± 0.0029	10342 ± 648
MedLFRM (K=50)	0.685 ± 0.0035	37860 ± 1224

models not only have effective and elegant model structures, but also have efficient inference algorithm that can obtain state-of-the-art performances.

There are several future directions to be discussed. The LFRM and its extended models we introduce in the paper belong to Bayesian methods, which represent one important class of statistic methods for machine learning. As Bayesian methods can get good performance in network analysis, improving them to scale up to large-scale networks is of great importance. Recently, many advances are in big learning with Bayesian methods (see [37] for a survey). Besides those techniques we mentioned in the paper, there are many other methods, such as scalable algorithms and distributed computing. Taking full advantage of big Bayesian learning, we can improve our methods effectively.

Besides Bayesian methods, deep learning is another powerful technique for learning latent features. Deep learning has been widely used in computer vision (e.g. deep convolutional neural network [38]) and natural language processing (e.g. word embedding [39]). Recently, a novel approach, DeepWalk [40], was proposed in network analysis, which incorporated random walk with deep learning to learn latent representations for entities in networks. How to take advantage of deep learning to solve the problem in network analysis, such as link prediction and community detection, is very worth studying.

Moreover, learning latent features in dynamic networks is more complicated, as both the networks and the features change over time. The dynamic relational infinite feature model (DRIFT) [41] is the dynamic extension of LFRM for link prediction, where the latent features for each entity in the network evolve according to a Markov process. It is significant but also challenging to enrich and scale up these kinds of latent feature models for dynamic network analysis.

Acknowledgments

The work was supported by the National Basic Research Program (973 Program) of China (No. 2013CB329403), National NSF of China Projects (Nos. 61620106010, 61322308, 61332007), and Tsinghua Initiative Scientific Research Program (No. 20141080934).

Funding

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in the published articles [2–4, 42–44].

Authors' contributions

JZ carried out the whole structure and the main idea, participated in drafting the manuscript. BC carried out the model development and experiments, participated in drafting the manuscript. Both of the authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Received: 14 May 2016 Accepted: 21 October 2016

Published online: 01 February 2017

References

- Liben-Nowell D, Kleinberg J. The link prediction problem for social networks. In: ACM Conference of Information and Knowledge Management (CIKM). New York: ACM; 2003. <http://dl.acm.org/citation.cfm?id=956972>.
- Craveny M, DiPasquoy D, Freitag D, McCallumzy A, Mitchelly T, Nigamy K, Slattery S. Learning to Extract Symbolic Knowledge from the World Wide Web. In: Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence. Menlo Park: American Association for Artificial Intelligence; 1998. p. 509–516. <http://dl.acm.org/citation.cfm?id=295240.295725>.
- Cho E, Myers SA, Leskovec J. Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM; 2011. p. 1082–1090. <http://dl.acm.org/citation.cfm?id=2020579>.
- Leskovec J, Kleinberg J, Faloutsos C. Graphs over time: densification laws, shrinking diameters and possible explanations. In: Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining. New York: ACM; 2005. p. 177–87. <http://dl.acm.org/citation.cfm?doi=1081870.1081893>.
- Lü L, Zhou T. Link prediction in complex networks: A survey. *Physica A: Stat Mech Appl*. 2011;390(6):1150–1170.
- Backstrom L, Leskovec J. Supervised random walks: predicting and recommending links in social networks. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. New York: ACM; 2011. p. 635–44. <http://dl.acm.org/citation.cfm?id=1935914>.
- Miller K, Jordan MI, Griffiths TL. Nonparametric latent feature models for link prediction. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2009. p. 1276–1284. <http://papers.nips.cc/paper/3846-nonparametric-latent-feature-models-for-link-prediction>.
- Clauset A, Moore C, Newman ME. Hierarchical structure and the prediction of missing links in networks. *Nature*. 2008;453(7191):98–101.
- Redner S. Networks: teasing out the missing links. *Nature*. 2008;453(7191):47–8.
- Xu M, Zhu J, Zhang B. Nonparametric max-margin matrix factorization for collaborative prediction. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2012. p. 64–72. <http://papers.nips.cc/paper/4581-nonparametric-max-margin-matrix-factorization-for-collaborative-prediction>.
- Chen N, Zhu J, Xia F, Zhang B. Discriminative relational topic models. *IEEE Trans Pattern Anal Mach Intell*. 2015;37(5):973–86.
- Salton G, McGill MJ. Introduction to modern information retrieval. New York: McGraw-Hill, Inc.; 1986.
- Adamic LA, Adar E. Friends and neighbors on the web. *Soc Networks*. 2003;25(3):211–30.
- Katz L. A new status index derived from sociometric analysis. *Psychometrika*. 1953;18(1):39–43.
- Hasanand MA, Chaoji V, Salem S, Zaki M. Link prediction using supervised learning. In: SDM: Workshop on Link Analysis, Counterterrorism and Security. Bethesda, Maryland; 2006. <http://www.siam.org/meetings/sdm06/workproceed/Link%20Analysis/12.pdf>.
- Shi X, Zhu J, Cai R, Zhang L. User grouping behavior in online forums. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM; 2009. <http://dl.acm.org/citation.cfm?id=1557105>.
- Lichtenwalter R, Lussier J, Chawla N. New perspectives and methods in link prediction. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM; 2010. <http://dl.acm.org/citation.cfm?id=1835837>.
- Nowicki K, Snijders TAB. Estimation and prediction for stochastic blockstructures. *J Am Stat Assoc*. 2001;96(455):1077–87.
- Kemp C, Tenenbaum J, Griffiths T, Yamada T, Ueda N. Learning systems of concepts with an infinite relational model. In: the American Association for Artificial Intelligence (AAAI). Boston, Massachusetts: AAAI Press; 2006. <http://dl.acm.org/citation.cfm?id=1597600>.
- Xu Z, Tresp V, Yu K, Kriegel HP. Infinite hidden relational models. In: International Conference on Uncertainty in Artificial Intelligence (UAI). Arlington, Virginia: AUAI Press; 2006. https://dmlpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=1291&proceeding_id=22.
- Airoldi E, Blei D, Fienberg S, Xing E. Mixed membership stochastic blockmodels. *J Mach Learn Res (JMLR)*. 2008;9:1981–2014. <http://dl.acm.org/citation.cfm?id=1442798>.
- Kim DI, Gopalan P, Blei DM, Sudderth EB. Efficient online inference for Bayesian nonparametric relational models. In: Advances in Neural Information Processing Systems (NIPS). Curran Associates, Inc.; 2013. <http://papers.nips.cc/paper/5072-efficient-online-inference-for-bayesian-nonparametric-relational-models>.
- Hoff P, Raftery A, Handcock M. Latent space approaches to social network analysis. *J Am Stat Assoc*. 2002;97(460):1090–8.
- Hoff P. Modeling homophily and stochastic equivalence in symmetric relational data. In: Advances in Neural Information Processing Systems (NIPS). Curran Associates, Inc.; 2007. <http://papers.nips.cc/paper/3294-modeling-homophily-and-stochastic-equivalence-in-symmetric-relational-data>.
- Menon AK, Elkan C. Link prediction via matrix factorization. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 437–52. http://link.springer.com/chapter/10.1007%2F978-3-642-23783-6_28.
- Zhu J, Song J, Chen B. Max-margin nonparametric latent feature models for link prediction. arXiv preprint arXiv:1602.07428. 2016. <https://arxiv.org/abs/1602.07428>.
- Chen B, Chen N, Zhu J, Song J, Zhang B. Discriminative nonparametric latent feature relational models with data augmentation. In: the American Association for Artificial Intelligence (AAAI). Phoenix, Arizona: AAAI Press; 2016. <http://aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12136>.

28. Antoniak CE. Mixture of Dirichlet process with applications to Bayesian nonparametric problems. *Ann Stat.* 2(6): http://www.jstor.org/stable/2958336?seq=1#page_scan_tab_contents.
29. Griffiths T, Ghahramani Z. Infinite latent feature models and the indian buffet process. In: *Advances in Neural Information Processing Systems (NIPS)*. MIT Press; 2005. <http://papers.nips.cc/paper/2882-infinite-latent-feature-models-and-the-indian-buffet-process>.
30. Zhu J, Chen N, Xing E. Bayesian inference with posterior regularization and applications to infinite latent SVMs. *JMLR.* 2014;15(1):1799–847.
31. Polson N, Scott S. Data augmentation for support vector machines. *Bayesian Anal.* 2011;6(1):1–23.
32. Polson N, Scott J, Windle J. Bayesian inference for logistic models using Polya-Gamma latent variables. *J Am Stat Assoc.* 2013;108:1339–1349. <http://www.tandfonline.com/doi/abs/10.1080/01621459.2013.829001>.
33. Welling M, Teh YW. Bayesian learning via stochastic gradient langevin dynamics. In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. New York: ACM; 2011. p. 681–8. http://machinelearning.wustl.edu/mlpapers/papers/ICML2011Welling_398.
34. Zhu J. Max-margin nonparametric latent feature models for link prediction. In: *International Conference on Machine Learning (ICML)*. New York: Omnipress; 2012. <http://www.icml.cc/2012/papers/>.
35. Teh YW, Görür D, Ghahramani Z. Stick-breaking construction for the indian buffet process. In: *International Conference on Artificial Intelligence and Statistics*. *JMLR*; 2007. p. 556–63. <http://jmlr.csail.mit.edu/proceedings/papers/v2/>. <https://www.mendeley.com/catalog/stickbreaking-construction-indian-buffet-process/>.
36. Hoffman MD, Blei DM, Wang C, Paisley J. Stochastic variational inference. *J Mach Learn Res.* 2013;14(1):1303–47.
37. Zhu J, Chen J, Hu W. Big learning with Bayesian methods. *arXiv preprint arXiv:1411.6370*. 2014. <https://arxiv.org/abs/1411.6370>.
38. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc.; 2012. p. 1097–1105. <http://papers.nips.cc/paper/4824-i>.
39. Bengio Y, Schwenk H, Senécal JS, Morin F, Gauvain JL. Neural probabilistic language models. In: *Innovations in Machine Learning*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006. p. 137–86. http://link.springer.com/chapter/10.1007%2F3-540-33486-6_6.
40. Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM; 2014. p. 701–10. <http://dl.acm.org/citation.cfm?doi=2623330.2623732>.
41. Foulds JR, DuBois C, Asuncion AU, Butts CT, Smyth P. A dynamic relational infinite feature model for longitudinal social networks. In: *International Conference on Artificial Intelligence and Statistics*. *JMLR*; 2011. p. 287–95. <http://www.jmlr.org/proceedings/papers/v15/foulds11b.html>.
42. Denham WW. The detection of patterns in Aiyawara nonverbal behavior. PhD thesis, University of Washington, Seattle. 1973.
43. Globerson A, Chechik G, Pereira F, Tishby N. Euclidean embedding of co-occurrence data. In: *Advances in Neural Information Processing Systems*. MIT Press; 2004. p. 497–504. <http://papers.nips.cc/paper/2733-euclidean-embedding-of-co-occurrence-data>.
44. Leskovec J, Kleinberg J, Faloutsos C. Graph evolution: Densification and shrinking diameters. *ACM Trans Knowl Discov Data (TKDD)*. 2007;1(1):2.

Submit your next manuscript to BioMed Central
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

