

Analysis of the Influence of Interlayer Structure Information of Multi-Layer Networks on Link Prediction

Fengqin Tang

School of Mathematics Sciences, Huaibei Normal University, Huaibei Anhui
Email: tfq05@163.com

Received: Jul. 9th, 2020; accepted: Jul. 23rd, 2020; published: Jul. 30th, 2020

Abstract

This paper discusses the link prediction problem of multi-layer networks and proposes a link prediction method that combines the information of the topological structure of the multi-layer network. We further analyze the influence of the information of the inter-layer structure on the prediction accuracy of the multilayer network link.

Keywords

Multilayer Network, Link Prediction, Similarity Matrix

浅议多层网络层间结构信息对链路预测的影响

唐风琴

淮北师范大学, 数学科学学院, 安徽 淮北
Email: tfq05@163.com

收稿日期: 2020年7月9日; 录用日期: 2020年7月23日; 发布日期: 2020年7月30日

摘要

文章讨论了多层网络的链路预测问题, 提出了结合多层网络层间拓扑结构信息的链路预测方法, 分析了层间结构信息对多层网络链路预测精度的影响。

关键词

多层网络, 链路预测, 相似矩阵



1. 引言

近年来,网络在很多领域应用广泛,包括社交网络、计算机网络、人类疾病基因网等。链路预测是网络统计分析的一个基本问题[1]。链路预测旨在基于观测到的网络的拓扑结构或者节点的特征信息预测网络中丢失的或者将会发生的链接。链路预测结果在许多学科中具有非常重要的作用,比如精准的预测结果可以指导生物学的实验—因为大部分观察到的生物网络,只是真实网络中很小的一部分,需要通过大量实验来确定其他的部分,又如社交网络的预测结果还可以用来进行好友预测—如果能够猜到用户未来交友的趋势,这种网络推荐就会变得非常精准。好的预测算法本身,还给出了很多网络演化可能机制的暗示。

鉴于现实网络的复杂多样性,单层网络已经不能满足人们对复杂系统的精确描述,近年来,人们对网络的研究逐渐从单层网络过渡到多层网络[2]。在多层网络中,利用网络层间信息对目标层网络进行链路预测是热点问题之一,文献[3]提出了一类基于 Adamic 相似度的多层网络的链路预测方法,该方法有效地解决稀疏网络的预测问题;文献[4]提出了基于概率模型的多层网络的链路预测方法;文献[5]提出了基于节点相似度的多层网络链路预测方法,该相似度均融合了层间和层内的结构信息。从上述文献的工作可见,合理地利用多层网络中的层间结构信息可以有效地提高链路预测的精度。基于此,本文进一步探讨多层网络的层间信息对目标层网络的链路预测的影响,提出了一类新的融合节点间相似度的多层网络的链路预测方法,给出基于节点对的相似度,即两对节点对间越相似,那么它们潜在的链接概率越相近。

2. 方法

首先给出本文中用到的记号。假设多层网络为 $G = (G^{(1)}, \dots, G^{(M)})$, 其中 M 代表层数, N 代表节点个数, $G^{(m)} = (V^{(m)}, E^{(m)})$ 代表第 m 层网络,该层网络的节点集合和边集合分别为 $V^{(m)}$ 和 $E^{(m)}$ 。本文假设所有网络层都是无向网络且它们的节点集合相同,即 $|V^{(1)}| = \dots = |V^{(M)}| = N$ 。令 $A^{(m)}$ 表示 $G^{(m)}$ 层网络对应的邻接矩阵,令 $S_{ij}^{(k)}$ 表示第 k 层网络节点 i, j 间的相似度,本文拟用 Jaccard 指标计算 $S_{ij}^{(k)}$, 即

$$S_{ij}^{(k)} = \frac{|N^{(k)}(i) \cap N^{(k)}(j)|}{|N^{(k)}(i) \cup N^{(k)}(j)|}, \text{ 其中 } N^{(k)}(i) \text{ 是第 } k \text{ 层网络节点 } i \text{ 的邻居节点的集合。}$$

下面给出第 k 层网络的链路预测准则(LPIMN):

$$Q^{(k)} = \arg \min_{p^{(k)}} \frac{1}{N^2} \sum_{ij} (A_{ij}^{(k)} - P_{ij}^{(k)})^2 + \frac{1}{N^4} \sum_{ii'jj'} S_{ii'}^{(k)} S_{jj'}^{(k)} (P_{ij}^{(k)} - P_{i'j'}^{(k)})^2 + \frac{1}{N^2} \sum_{kl} \sum_{ii'jj'} \theta_{kl} (A_{ij}^{(k)} - A_{ij}^{(l)})^2 \quad (1)$$

其中 $P_{ij}^{(k)}$ 是待估的网络层 k 中节点 i, j 间的潜在的链接概率, θ_{kl} 表示层间信息的权重,

$$\text{本文中令 } \theta_{kl} = \frac{1}{N} \sum_i \frac{\sum_j A_{ij}^{(l)} A_{ij}^{(k)}}{\sum_j A_{ij}^{(l)} + \sum_j A_{ij}^{(k)} - \sum_j A_{ij}^{(l)} A_{ij}^{(k)}}。$$

本文利用分块梯度下降法求解式(1)，首先对式(1)关于 $P^{(k)}$ 求导得：

$$\begin{aligned} \frac{\partial Q^{(k)}}{\partial P_i^{(k)}} = & \frac{2}{N^2} (P_i^{(k)} - A_i^{(k)}) + \frac{1}{N^4} (S_{ii}^{(k)} (W^{(k)} P_i^{(k)} - S^{(k)} P_i^{(k)})) \\ & + \sum_{i' \neq i} S_{ii'}^{(k)} (W^{(k)} P_i^{(k)} - S^{(k)} P_i^{(k)}) + \frac{1}{N^2} \left(\sum_{l \neq k} \theta_{kl} (P_i^{(k)} - A_i^{(l)}) \right) \end{aligned}$$

其中 $W^{(k)}$ 是一个 n 阶方阵， $W_{ii}^{(k)} = \sum_j S_{ij}^{(k)}$ ， $P_i^{(k)}$ 是行向量。那么令 $\frac{\partial Q^{(k)}}{\partial P_i^{(k)}} = 0$ ，可得第 $(k+1)$ 次迭代中 $P_i^{(k)}$ 的估计值为

$$\begin{aligned} P_i^{(k)(t+1)} \leftarrow & \left[N^2 I + 2\lambda_1 S_{ii}^{(k)} \left(W^{(k)} - S^{(k)} + \sum_{i' \neq i} S_{ii'}^{(k)} W^{(k)} \right) + \lambda_1 N^2 \sum_{l \neq k} \theta_{kl} \right]^{-1} \\ & \left[N^2 A_i^{(k)} + 2\lambda_1 \sum_{i' \neq i} S_{ii'}^{(k)} S P_i^{(k)(t)} + \lambda_1 N^2 \sum_{l \neq k} \theta_{kl} A_i^{(l)} \right]^{-1} \end{aligned}$$

3. 实证

本节通过两个实际多层网络数据分析所提方法 LPIMN 和现有的部分链路预测方法在预测多层网络链接时的有效性。用于对比的算法有 Najari 等[5]提出的 LPIS 算法，Yao 等[6]提出的 NSILR 算法，Zhao 等[7]提出的 LPPN 算法和 Zhang 等[8]提出的 ENNS 算法，其中 LPPN 和 ENNS 针对单层网络的链路预测，而 NSILR 和 LPIS 针对多层网络链路预测。用于实证分析的两个实际数据为：1) 律师多层网络数据[6]，包含 71 个节点，3 个网络层(Advice, Work, Friend)；2) AU-CS 数据[5]，包含 61 个节点，5 个网络层(Work, Lunch, Facebook, Leisure, Coauthor)。实验结果评价指标为 AUC [5]，AUC 是公认的评价链路预测算法精度的指标， $AUC \in [0,1]$ ，其值越大，说明算法的精度越高。具体的实验结果如表 1 和表 2 所示，AUC 数值显示 LPIMN 算法在处理多层网络链接预测问题时更为有效，虽然 NSILR 和 LPIS 两个算法均综合考虑了多层网络的层间信息，但是它们在处理有向网络图和连边较少即稀疏网络图时不及 LPIMN 算法的优势明显，LPIMN 算法很好地结合了网络的层间信息并根据其他层与目标层的相关性有针对性地获取信息，实质性地改进了算法的精度。

Table 1. The AUC values obtained by five different methods on lawyers network

表 1. 五种不同方法作用于律师三层网络数据得到的 AUC 值

	LPIMN	NSILR	LPIS	LPPN	ENNS
Advice	0.87	0.81	0.83	0.78	0.73
Work	0.85	0.78	0.82	0.71	0.67
Friend	0.85	0.83	0.81	0.70	0.78

Table 2. The AUC values obtained by five different methods on AU-CS data

表 2. 五种不同方法作用 AU-CS 数据得到的 AUC 值

	LPIMN	NSILR	LPIS	LPPN	ENNS
Work	0.85	0.78	0.82	0.71	0.67
Lunch	0.92	0.87	0.87	0.89	0.87
Facebook	0.92	0.83	0.78	0.76	0.75
Leisure	0.92	0.88	0.84	0.75	0.83
Coauthor	0.93	0.86	0.74	0.74	0.73

4. 结论

多层网络在实际中应用广泛, 在诸多领域中均涉及到网络, 比如计算机通讯系统, 金融管理系统, 新冠病毒传播网络等, 但是多层网络的统计研究尚处于成长阶段, 这就要求我们平时在课程学习中要在理论推导方面多下功夫, 逐步丰富多层网络的统计研究, 其中多层网络的链路预测又属热门课题之一, 合理地利用多层网络层间信息和层内信息有效地预测节点间潜在的链接概率将是我们继续的研究方向。

基金项目

安徽省高校自然科学研究重点项目(KJ2020A0023), 淮北师范大学重点教研项目(JY19009)。

参考文献

- [1] 吕林媛. 复杂网络的链路预测[J]. 电子科技大学学报, 2010(5): 651-661.
- [2] 张欣. 多层复杂网络理论研究进展: 概念、理论和数据[J]. 复杂系统与复杂性科学, 2015, 12(2): 103-107.
- [3] Davis, D., Lichtenwalter, R. and Chawla, N.V. (2011) Multi-Relational Link Prediction in Heterogeneous Information Networks. 2011 *International Conference on Advances in Social Networks Analysis and Mining*, Kaohsiung, 25-27 July 2011, 281-288. <https://doi.org/10.1109/ASONAM.2011.107>
- [4] Yang, Y., Chawla, N., Sun, Y., et al. (2012) Predicting Links in Multi-Relational and Heterogeneous Networks. 2012 *IEEE 12th International Conference on Data Mining*, Brussels, 10-13 December 2012, 755-764. <https://doi.org/10.1109/ICDM.2012.144>
- [5] Najari, S., Salehi, M., Ranjbar, V., et al. (2019) Link Prediction in Multiplex Networks Based on Interlayer Similarity. *Physica A: Statistical Mechanics and its Applications*, **536**, Article ID: 120978. <https://doi.org/10.1016/j.physa.2019.04.214>
- [6] Yao, Y., Zhang, R., Yang, F., et al. (2017) Link Prediction via Layer Relevance of Multiplex Networks. *International Journal of Modern Physics C*, **28**, Article ID: 1750101. <https://doi.org/10.1142/S0129183117501017>
- [7] Zhao, Y., Wu, Y.J., Levina, E., et al. (2017) Link Prediction for Partially Observed Networks. *Journal of Computational and Graphical Statistics*, **26**, 725-733. <https://doi.org/10.1080/10618600.2017.1286243>
- [8] Zhang, Y., Levina, E. and Zhu, J. (2015) Estimating Network Edge Probabilities by Neighborhood Smoothing. *Biometrika*, **104**, 771-783. <https://doi.org/10.1093/biomet/asx042>