

基于共同邻居惩罚的复杂网络链路预测方法

邹剑升¹, 李玉珩²

(1. 浙江中烟工业有限责任公司, 浙江 宁波 315504;

2. 秦皇岛烟草机械有限责任公司, 河北 秦皇岛 066004)

摘要: 链路预测是确定用户间关系的基本工具; 通过相似性度量进行链路预测是一种常见的方法, 提出一种基于相似度的链路预测算法, 根据网络结构及拓扑特性来确定相似度, 引入优化链路预测度量方法, 将聚类系数作为网络结构性质; 此外, 并考虑共享邻域, 得到较其他同类链路预测方法更好的性能; 实验结果表明, 提出的算法性能优于经典算法; 结合在社交网络环境中的实验结果可知, SLP-CNP 法较其他算法具有更优精度与效率; 在未来的工作中, 还可尝试在所提方法的基础上, 提升在加权网络、有向网络和二部网络中的适用性。

关键词: 共同领域; 复杂网络; 链路预测; 相似度计算

Complex Network Link Prediction Method Based on Common Neighbor Punishment

WU Jiansheng¹, LI Yuheng²

(1. China Tobacco Zhejiang Industrial Co., Ltd., Ningbo 315504, China;

2. Qinhuangdao Tobacco machinery Co., Ltd., Qinhuangdao 066004, China)

Abstract: Link prediction is a fundamental tool for determining relationships between users. Link prediction by similarity measure is a common method. This paper proposes a link prediction algorithm based on similarity, which determines the similarity according to the network structure and topological characteristics, introduces the optimized link prediction measure, and takes the clustering coefficient as the network structure property. In addition, considering the shared neighborhood, the performance is better than other similar link prediction methods. Experimental results show that the proposed algorithm outperforms the classical algorithm. Combined with the experimental results in the social network environment, it can be seen that SLP-CNP method has better accuracy and efficiency than other algorithms. In the future work, we can also try to improve the applicability of the proposed method in weighted networks, directed networks and bipartite networks.

Keywords: common domain; complex network; link prediction; similarity calculation

0 引言

随着社会网络中信息量的迅速扩大, 链路预测已成为推荐系统、决策和刑事侦查等领域的重要问题^[1]。链路预测涉及计算网络中节点间丢失或未来链路的可能性^[2-3]。为精确定义链路预测问题, 假设复杂网络为无向图 $G = (V, E)$, 其中 V 为一组节点, E 表示节点对间的边。考虑到在时间 t 处的网络 G 的快照, 链路预测问题涉及在时间 $t + \Delta$ 处形成的当前快照中定义丢失的子集^[4]。

现有的复杂网络链路预测问题面临两大挑战: 第一类是海量数据, 需要低复杂度的预测方法; 第二个挑战是预测方法涉及高预测精度。而传统的数据挖掘方法忽略了实体间关系, 无法有效地解决链路预测问题。现有研究采用不同方法来处理链路预测问题, 其中多数基于节点间的相似性实现计算^[4]。在相似性计算技术中, 相似节点间更容易形成链接。此外, 有一些链路预测方法考虑了共享更多

邻居的节点^[7]。

在上述方法中, 为给网络中的每对节点分配相似性分数, 首先定义函数 $s(x, y)$, 基于不同特征 (如拓扑特征和在相似度评分) 将网络中的所有节点按其得分的降序排列, 并在丢失的链路列表中选择具有最高等级的链路作为可预见链路。基于相似度的链路预测方法根据计算相似度函数时所考虑的信息量可分为局部、全局和准局部三类^[8]。在局部技术中, 更多关注直接的邻居节点信息, 常适用于大型复杂网络, 与线性时间复杂度相比具有较高精度。利用整个网络拓扑结构的全局技术能够计算每对节点间的相似度, 而不局限于节点共邻。而与局部方法相比, 全局方法由于对噪声的敏感性和较高的计算复杂度具有较低精度。而准局部技术寻求利用局部和全局技术点, 通过考虑邻节点的邻节点而不仅仅是直接邻节点, 并限制每对节点间的距离。

上述方法均可在考虑信息量和计算复杂度间找到平衡。

收稿日期: 2022-08-03; 修回日期: 2022-09-07。

作者简介: 邹剑升(1981-), 男, 浙江台州人, 大学本科, 工程师, 主要从事信息系统与自动化方向的研究。

引用格式: 邹剑升, 李玉珩. 基于共同邻居惩罚的复杂网络链路预测方法[J]. 计算机测量与控制, 2023, 31(3): 71-75, 139.

链路预测范围有两个主要问题。第一个是引入一种低计算复杂度的链路预测方法，特别是在面对大规模数据集时；第二个是预测精度，故需高精度的链路预测方法。而链路预测的两个主要挑战是时间复杂度和准确性。由于传统的链路预测方法不能有效地解决这一问题，还没有一种既能获得低复杂度又能获得高准确度的方法被提出。本文提出了基于公共邻域惩罚的相似度链路预测方法 (similarity link prediction method for common neighborhood punishment, SLP-CNP)，根据网络拓扑特征 (包括每两个节点的公共邻域) 和平均聚类系数确定相似度，与其它同类算法的主要区别在于区分节点的公共邻域，是一种同时兼顾局部和全局特征的准局部相似方法。实验结果表明，该方法在精度和计算复杂度方面优于同类方法。

1 相关工作

链路预测在链路分析、信息检索和网络演化等领域变得越来越重要。在社交网络中，链路预测常被用于预测潜在的社交关系，并基于此为用户推荐好友或信息。最经典的相似度链路预测方法为公共邻居 (CN)^[9]、Adamaic Adar (AA)^[10] 和资源分配 (RA)^[11]。公共邻居对每对节点给出的相似度得分涉及到这些节点间共享邻居的数量，且假设如两个节点有多个共邻，则其间形成边的概率将增加。Adamaic Adar 法根据其程度对每个共享邻居进行分组，并通过研究每个节点间的公共邻居为节点分配相似性分数。资源分配法考虑了两个非连接节点间通过其邻居的资源分配，使得每个邻居节点接收到一些资源并在其邻居之间平均分配。两个节点间的相似性准则可通过共享邻居从一个节点从另一节点接收到的资源量表示。而 Jaccard 指数^[12]、S0rensen 指数^[13] 和 Leicht-Holme-Newman 指数^[8] 是链接预测中采用的其他基于相似性的度量。在链路预测范围内，有几种基于相似度的方法，其中两个节点间的链路概率是根据其共享邻居来确定。刘留等^[14] 提出了基于公共邻居的动态社会网络链路预测算法，使用 3 个特定度量为两个节点间的所有边分配权重，然后将其总和确定为所述节点间链接的概率。而 Wu 等^[15] 提出了节点耦合聚类系数，将节点间的公共邻域部分与聚类信息相结合，采用每个节点相同的公共邻域聚类系数。Dong 等^[16] 建立了结合邻居和群体信息的复杂网络链路预测模型，其考虑了称为基序的网络结构单元，为确定每个公共邻域的两个节点间的相似度得分，将所有公共邻域的结果汇总，并将结果划分为这些公共邻域的个数，最终实现对节点进行相似度评分。

而翟东升等^[17] 提出了合著网络中链接预测概念和基于主题建模的链接预测算法，证明了经典转移相似度区间的范围，用模糊系统理论来表示相似度。复杂网络图中顶点的两阶段选择链接预测方法^[18]，旨在预测图流中最有可能连接到目标顶点的 top-k 顶点。黄璐等^[19] 引入了时间链路预测方法，利用复杂网络局部和全局拓扑结构，将科学家互引预测问题描述为引文网络链接预测问题，其中链接预

测方法通过使用时间链接预测度量来预测链接和链接权重。另外，有学者使用影响最大化算法从目标的当前影响用户集合中确定一组可能的影响用户^[20]。而 Bastami 等^[21] 提出了基于无监督链路预测方法，利用节点特征、社区信息和图特征组合来提高局部和全局预测精度。Grover 等^[22] 以半监督方式使用这些信息进行节点聚类，并使用不同的统计抽样方法来生成网络中的节点上下文，但有时统计抽样方法也无法有效保持节点的高阶拓扑关系。而 Cao 等^[23] 提出的 GraRep 算法依赖于奇异值分解 (singular value decomposition, SVD) 和矩阵乘法法进行链路预测，但具有较高的时间复杂度。另外，Wang 等^[24] 所提 SEMAC 法也是基于矩阵乘法进行链路预测优化，并应用于每个节点对应的子图预测，Dharavath 等^[25] 将链路预测定义为二分类问题，根据节点相似性选择一组特征，而 Aghabozorgi 等^[26] 利用网络图作为监督学习下的结构特征来进行链路预测。所以综上所述，在预测社交网络中的链接时，不完整是另一个严峻的挑战，这是由于几乎所有的社交网络数据都包含缺失值，这是出于匿名和隐私保护，通常只能收集部分数据，且当网络规模较小或缺少严重时，冷启动问题尤为严重，而在处理耦合网络时也会遇到这种情况。所以在网络中完善和实施链路预测方法是现在复杂网络研究的核心问题之一。

2 算法设计

2.1 预备知识

链路预测方法旨在根据不同域上不同结构特征对网络进行准确预测。自适应度惩罚算法根据复杂网络聚类系数对公共邻域的度进行惩罚，一般的相似度度量方法可定义为式 (1) 所示。

$$\text{Similarity}(x, y) = \sum_{z \in \Gamma_x \cap \Gamma_y} |\Gamma_z|^{-\alpha} \quad (1)$$

其中： α 为常量， z 为 x 和 y 间的公共邻居，且 Γ_z 为 z 的度数。不同方法间的区别为 α 值。在自适应度惩罚算法中， α 值通过考虑节点间最短路径和平均聚类系数得出节点间是否存在较强相关性及聚类系数，作为两个节点 x 和 y 间的链接概率可被表示如式 (2) 所示。

$$\text{ADP}(x, y) = \sum_{z \in \Gamma_x \cap \Gamma_y} |\Gamma_z|^{-\beta} \quad (2)$$

其中： C 为平均聚类系数， β 为常量值。将聚类系数作为复杂网络结构属性，对每对节点间共享邻居的个数进行分组。自适应度惩罚算法可在多种网络上进行，并取得良好性能。网络中有几个结构属性，特征包括节点间最短路径、节点间路径的信息熵、网络中最长最短路径的网络直径及节点聚类系数。平均聚类系数是整个网络的常数，可通过计算网络中每个节点的聚类系数得到。故可根据式 (3) 计算节点 x 的聚类系数，且可得平均聚类系数如式 (4) 所示。

$$CE = \frac{\sum_{y, z \in \Gamma_x, E_{y,z} \in E} E_{y,z}}{|\Gamma_x| (|\Gamma_x| - 1)} \quad (3)$$

$$\overline{CE} = \frac{\sum_{x \in V} CE}{|V|} \quad (4)$$

2.2 所提 SLP-CNP 法

基于相似度的链路预测方法具有相同的框架, 节点间相似度是不同方法间的唯一区别。其主要目的是提供更准确的指标来估计网络中节点间链路存在的概率, 是每对节点间的相似度得分。而每两个节点间的链接概率取决于其间的公共邻居数量。现有方法多没有将处罚程度与网络特征和结构进行联系, 而自适应度惩罚算法适当利用相似性指数中的平均聚类系数来关注复杂网络特征和网络结构, 但其缺乏对公共邻域形式的关注。为克服这一挑战, 本文提出了 SLP-CNP 法, 从一个新的角度看待邻域。通过区分公共邻域对自适应度惩罚算法进行改进。如需计算节点 x 和 y 间的相似性得分, 为提高链路预测的效率, 在度量中考虑了这种差异。如节点 x 和 y 有已经是互为邻节点的节点对, 则节点 x 和 y 将来成为好友的概率将比节点 x 和 y 不属于朋友的概率要大。需要注意, 当公共邻节点的数量增加时, 链路预测方法的精度和效率提高。为此, 所提方法以不同方式考虑共享邻居, 还可根据网络结构进行调整, 具体如式 (5) 所示。

$$Similarity(x, y) = \sum_{z \in \Gamma_z} |C_z| * \{ADP = (|\Gamma_z|)^{-\alpha}\} \quad (5)$$

其中: z 为两个节点 x 和 y 的公共邻居, $|C_z|$ 为邻居数量, 除节点 x 和 y 外还包括其公共邻居。 Γ_z 为 z 的邻域数量, C 为平均聚类系数。本文所提 SLP-CNP 法的具体步骤如算法 1 所示。首先计算每个节点的平均聚类系数, 将网络分成训练集和测试集 (10% 和 90%), 采用 5 次交叉验证将原网络的总边划分为 5 个等边。计算每个边的相似度得分按降序排列, 然后将排列列表中的边添加到序列列表中。在列车网络图的边上添加与测试集完全相同的边数, 并从主网络中任意减去。这些添加的边是预测边。最后, 确定真阳性 (正确预测) 和假阳性 (错误预测) 数量, 并基于此计算精度。

算法 1: 本文所提 SLP-CNP 算法

输入: 复杂网络图 G

输出: 平均精度与 AUC

01: **for** 每个节点 i **do**

02: 对于节点 i 计算簇系数

03: **end for**

04: 基于簇系数之和与簇数量的商计算平均簇系数

05: 将图 G 按 5-fold 切分为训练网络 G_{train} 与测试网

络 G_{test}

06: **for** 训练网络 G_{train} 中的每条边 (x, y) **do**

07: 对每条边 (x, y) 计算相似性得分 S_{xy}

08: **end For**

09: 按降序排列所有相似性得分 S_{xy}

10: 基于有序列表插入边至训练网络 G_{train}

11: 基于式 (6) 和式 (7) 计算精度和 AUC 值

12: 基于上述精度和 AUC 值结果, 并使用式 (8) 和式 (9) 计算平均精度和 AUC 值

$$AUC = \frac{n' + 0.5n''}{n} \quad (6)$$

$$Precision = \frac{l}{L} \quad (7)$$

其中: n' 是错误链接分数大于不存在链接分数的次数, n'' 为两个分数相等的次数, n 为比较总次数。如得分来自独立分布, 则 AUC 值预计为 0.5, 故如 AUC 值高于 0.5 表示性能优于纯随机情况。为获得预测精度, 对不存在的链接分数计算精度, 并将得到值可按降序排序。然后, 选择得分最高的 L 条链接, 得到 l 作为正确预测的链接数量。

$$AVERAGE AUC = \sum_i \frac{n'_i + 0.5n''_i}{n_i} / i \quad (8)$$

$$AVERAGE APrecision = \sum_i \frac{l_i}{L_i} / i \quad (9)$$

3 实验设计

为证明本研究所提 SLP-CNP 算法的有效性, 本研究使用 3 个真实社交网络数据集进行数值模拟, 以便观察算法对现实情况的适应度 (实验中使用的真实社交网络数据集特征如表 1 所示)。其中, 激活率根据网络图的稀疏性进行设置, 并基于社交网络中节点度和二阶度的平均值进行计算。本文使用 Python 软件以近期的热点事件 “HUAWEI event” 和 “华为事件” 的 30 个热点评论用户节点作为初始节点, 爬取了知乎、CSDN 和新浪微博的社交网络用户数据集作为实验仿真的基础数据 (爬取时间为 2020 年 4 月 23 日 - 2020 年 9 月 6 日)。本研究将每个用户作为一个节点, 使用节点间的边界表示用户间关系。本研究选择了 10 个具有较强影响力的用户及其好友列表作为社交网络初始节点, 以此生成了简单社交网络。实验在 MATLAB 2017b 环境下实施, 并都是在 Windows10 操作系统的服务器 (Intel Xeon 处理器 (34 GHz) 和 32 GB 内存) 上进行。首先随机选择 10% 的链接并从网络中删除。为了获得更精确的结果并避免算法的随机行为, 此选择将执行五次。接下来, 应用 5 倍交叉验证方法, 将网络划分为 5 个相等的部分, 每次将一个部分视为测试集或信用集。

表 1 社交网络数据集说明

网络序号	社交网络名称	节点数量	节点边界数量	平均度	节点平均路径	聚类系数
1	知乎	25 231	232 311	10.98	3.24	0.558
2	CSDN	38 231	532 411	34.08	3.65	0.546
3	新浪微博	68 790	452 738	24.49	5.35	0.509

为评估本文所提 SLP-CNP 链路预测算法的性能, 将其结果与现有较为成熟的重要节点识别算法进行比较。其中包括 4 种中心度算法: 即度中心性法^[15] (degree centrality, DC)、k-shell 法^[27]、PageRank 法^[18] 和介数中心性法^[16] (intermediate centrality, IC)。两种启发式算法: 双折扣法^[16] (double discount, DD) 和启发式聚类法^[22] (heuristic

clustering, HC)。两种元启发式算法：度递减搜索策略 (degree descending search strategy, DDSE)^[23] 和自适应度惩罚算法 (ADP)^[19]。由于实验中 SLP-CNP 链路预测算法预测列表在每次运行时的结果都有可能不同, 故设置评估结果为迭代 100 次运行后的平均值, 运行的平均标准差为 1.524。如前所述, β 为一个常数参数, 该参数值影响两个节点间存在链路的概率, 而该参数在一定程度上决定了该方法的性能。故考虑到所提用于计算两个节点间存在链路的可能性的指标, 为每个节点设置不同的值。本文采用试错法, 首先评估标准的应用基于参数 γ 为每个网络, 然后最好的表现 γ 对于不同的网络可获得的多个 $\gamma \in [-1.0, 1.5]$ 范围内, 每个网络的性能最好 γ 实现价值。然后在网络的聚类系数和最佳性能 γ 值间进行线性回归以确定 β 数值, 故对于精密测量而言 $\beta=1.74$ (合成网络图如图 1 所示)。

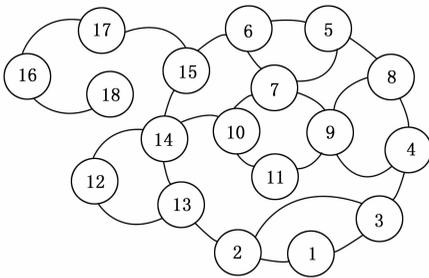


图 1 合成网络图结果

4 实验结果与分析

4.1 基准实验结果

首先随机选择 10% 的链接从网络中删除, 为获得更精确的结果并避免算法的随机行为, 并应用 5-fold 交叉验证法将网络划分为 5 个相等部分, 为保证算法的正确性在图 1 所示的合成网络结构中进行。在网络中提出的链路预测方法后, 该算法将最相似的分值赋予所提边, 实际上可被视为预

测边。在上述情况下精度值是 1, 作为正确的预测边相对于预测边数量的比率, 这与边的正确预测完全相同。此外 AUC 值为 0.939, 且根据精度和 AUC 值, 并以该方法提供的边作为结果, 可知该方法能正确地预测丢失链路。

将本文所提 SLP-CNP 算法与现有算法进行性能比较。用于比较的链路预测方法都是基于相似度的, 由于本文所采用的数据集是无向的, 故通过考虑已经与边相连的每对节点之间的两个方向来施加测度, 并将其结果与所提方法进行比较。实验结果表明, 本研究所提度量方法较其他方法更为有效。表 2 通过选择 3 个真实网络中 20% 的边来说明不同算法的性能。结果表明, 本研究所提 SLP-CNP 算在所有状态下的性能都最优。表 3 报告了不同真实网络的 5-fold 交叉验证, 展示了不同算法的性能, 可知本文所提 SLP-CNP 算法较其他算法更优。

4.2 时间复杂度

如前所述, 计算复杂度和执行时间是链路预测方法的关键挑战, 故当一种方法能够在基于评估标准的执行时间和效率间取得良好平衡时更为优越。总的来说, 所有基于邻域的相似性度量都有相同过程, 所有这些方法间的唯一区别是计算相似性的过程。在进行 SLP-CNP 法计算时, 对于节点 x 首先搜索 x 的所有邻节点。遍历节点邻域的时间复杂度仅为 k , 而本文所提 SLP-CNP 法的时间复杂度为 $O(nk^2)$, n 表示节点数量, k 表示平均度。除时间复杂度外, 内存空间是算法实现面临的另一个限制。在进行 SLP-CNP 法计算时, 所需的内存约为 $O(nk)$, 顾所需内存和 CPU 时间相对较少。表 4 报告了本文所提 SLP-CNP 法与其他方法的运行时间 (以秒为单位) 比较结果。由结果可知, 本文为所提方法的时间复杂度较其他方法较优。总的来说, 本文所提方法的主要优点是获得了高性能, 比基于相似性的方法更好, 计算复杂度非常低, 这大大减少了基准数据集的运行时间。

表 2 不同真实复杂网络中选择 10% 边的算法比较结果

测度指标	数据集名称	DC	k-shell	PageRank	IC	DD	HC	DDSE	ADP	SLP-CNP
精度	知乎	0.124	0.523	0.131	0.131	0.538	0.319	0.648	0.213	0.837
	CSDN	0.367	0.314	0.238	0.223	0.421	0.472	0.341	0.427	0.701
	新浪微博	0.313	0.398	0.219	0.218	0.429	0.321	0.319	0.481	0.889
AUC	知乎	0.412	0.237	0.574	0.483	0.437	0.527	0.692	0.769	0.850
	CSDN	0.397	0.311	0.528	0.322	0.541	0.698	0.646	0.538	0.841
	新浪微博	0.231	0.427	0.519	0.349	0.427	0.592	0.528	0.534	0.738

表 3 不同真实复杂网络中通过 5-fold 的算法比较结果

测度指标	数据集名称	DC	k-shell	PageRank	IC	DD	HC	DDSE	ADP	SLP-CNP
精度	知乎	0.218	0.438	0.218	0.128	0.427	0.438	0.538	0.328	0.953
	CSDN	0.327	0.534	0.242	0.327	0.312	0.421	0.483	0.307	0.849
	新浪微博	0.320	0.555	0.285	0.437	0.489	0.329	0.491	0.390	0.910
AUC	知乎	0.328	0.423	0.548	0.452	0.438	0.438	0.348	0.231	0.850
	CSDN	0.538	0.437	0.516	0.644	0.547	0.502	0.231	0.472	0.810
	新浪微博	0.399	0.548	0.349	0.317	0.280	0.479	0.538	0.408	0.948

表 4 不同真实复杂网络中算法时间复杂度比较结果

数据集名称	DC	k-shell	PageRank	IC	DD	HC	DDSE	ADP	SLP-CNP
知乎	131	241	149	563	327	249	454	131	32
CSDN	42	271	462	234	314	137	231	103	34
新浪微博	62	182	535	371	128	231	230	78	55
平均值	78	231	382	389	256	206	305	104	40

表 5 不同算法的 Friedman 检验比较结果

测度指标	数据集名称	DC	k-shell	PageRank	IC	DD	HC	DDSE	ADP	SLP-CNP	卡方检验值	df	P 值
精度	知乎	2.112	2.319	2.133	1.837	1.237	2.840	1.039	1.433	4.911	24.192	3	0.000
	CSDN	2.123	2.212	1.893	1.730	1.419	2.739	1.210	1.538	4.482	23.227	3	0.000
	新浪微博	2.283	2.137	1.739	1.833	1.321	2.019	1.279	1.839	5.943	27.810	3	0.000
AUC	知乎	2.319	2.192	2.123	1.938	1.732	1.303	1.238	3.293	4.123	25.293	3	0.000
	CSDN	2.839	2.910	2.394	1.839	1.928	1.343	1.847	2.839	4.028	21.903	3	0.000
	新浪微博	2.738	1.930	2.185	1.933	1.783	1.238	1.293	2.817	4.128	24.239	3	0.000

4.3 统计性检验

本研究使用 Friedman 检验^[27], 分析了从不同相似性链路预测方法获得的结果。Friedman 检验是一种非参数统计检验, 用于发现多种方法的行为差异。非参数 (无分布) 意味着测试不假设数据来自特定的分布。Friedman 检验可用于评价 N 种不同方法对 K 个数据集的结果。在这个测试中, 这些方法是根据它们的性能标准来排序的。本文以精密度和 AUC 为评价标准, 以高阶方法的评价效果最好。表 4 报告了上述链路预测方法的排名。可知本文所提 SLP-CNP 法的分类精度和 AUC 分别为 4.00 和 3.77, 高于其他基于相似度的分类方法, 且 P 值小于 0.05。故可知上述结果通过统计性检验是显著的。

5 结束语

本文提出了新的链路预测度量方法, 将聚类系数作为网络的结构属性加以考虑, 该方法除考虑每对节点的共享邻居, 还考虑了共享邻节点的邻节点, 故比其他类似链路预测方法具有更好性能。为验证该方法的有效性, 在多个真实网络上进行对比实验。结合在知乎、CSDN 与新浪微博等社交网络环境中的实验结果可知, 本研究所提 SLP-CNP 法较其他算法具有更优精度与效率。在未来的工作中, 将尝试提出新的系统化方法, 提出并行算法显著提高效率的方法来改进所提出的方法。其次, 还可尝试本文所提方法在加权网络、有向网络和二部网络中的适用性。再次, 可尝试使用不同的操作符, 以有效提升链路预测算法的时间复杂度, 以提升算法的运行效率。另外, 可尝试使用深度学习等先进技术以提升复杂网络链路方法的效率与精度。最后, 可提出链路预测方法以确定适当参数值, 以优化相似度方法。在应用场景方面, 可尝试在如蛋白质网络、恐怖分子网络、科研合作网络、多层网络等其他复杂网络结构中对本文所提算法的适用性进行验证。

参考文献:

[1] 郭丽媛, 王智强, 梁吉业. 基于边重要度的矩阵分解链路预测

算法 [J]. 模式识别与人工智能, 2018, 31 (2): 150 - 157.

- [2] PIETER B, KOLLER A D. Link prediction in relational data [J]. Advances in Neural Information Processing Systems Nips, 2003, 16: 659 - 666.
- [3] MARTINEZ V, BERZAL F, CUBERO J C. Adaptive degree penalization for link prediction [J]. Journal of Computational Science, 2016, 13: 1 - 9.
- [4] BU D, ZHAO Y, LUN C, et al. Topological structure analysis of the protein-protein interaction network in budding yeast [J]. Nucleic Acids Research, 2003 (9): 2443 - 2450.
- [5] LIBEN-NOWELL D, KLEINBERG J. The link-prediction problem for social networks [J]. Journal of the American Society for Information Science and Technology, 2007, 58 (7): 1019 - 1031.
- [6] ISELLA L, STEHLÉ J, BARRAT A, et al. What's in a crowd? Analysis of face-to-face behavioral networks [J]. Journal of Theoretical Biology, 2010, 271 (1): 166 - 180.
- [7] LEICHT E A, HOLME P, NEWMAN M. Vertex similarity in networks [J]. Physical Review E, 2006, 73 (2): 26 - 120.
- [8] 李冬, 申德荣, 寇月, 等. 基于层次化混合特征图的链路预测方法 [J]. 中国科学: 信息科学, 2020, 50 (2): 221 - 238.
- [9] NEWMAN M E. Clustering and preferential attachment in growing networks [J]. Physical Review E, 2001, 64 (2): 25 - 102.
- [10] ADAMIC L A, ADAR E. Friends and neighbors on the Web [J]. Social Networks, 2003, 25 (3): 211 - 230.
- [11] ZHOU T, Lü L, ZHANG Y C. Predicting missing links via local information [J]. European Physical Journal B, 2009, 71 (4): 623 - 630.
- [12] JACCARD P. Etude comparative de la distribution florale dans une portion des Alpes et des Jura [J]. Bulletin del la Societe Vaudoise des Sciences Naturelles, 1901, 37 (142): 547 - 579.
- [13] SØRENSEN T A. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons [J]. Biol Skr, 1948, 5 (1): 1 - 34.

(下转第 139 页)