基于共同邻居邻域拓扑稠密性加权的链路预测方法 *

李 星，朱宇航，柏 溢，李劲松

(中国人民解放军战略支援部队信息工程大学，郑州 450002)

摘 要：链路预测旨在利用已知的网络节点和拓扑结构信息，预测网络中未连接的两个节点之间存在连边的可能性。基于网络拓扑相似性的链路预测方法计算复杂度低且预测效果好，但现有的相似性指标对共同邻居的邻域拓扑信息考虑较少。针对此问题，提出一种基于共同邻居邻域拓扑稠密性加权的链路预测方法。首先，基于邻域拓扑相对稠密指数量化节点的邻域拓扑结构；然后，利用共同邻居的节点度和邻域拓扑相对稠密指数刻画共同邻居及其邻域拓扑的相似性贡献；最后，提出基于共同邻居邻域拓扑稠密性加权的节点相似性指标。在多个实际网络数据上的实验结果表明，与现有相似性指标相比，该方法能够取得更高的预测精度。

关键词：复杂网络；链路预测；邻域拓扑稠密性；拓扑加权

中图分类号：TP399 doi: 10.19734/j.issn.1001-3695.2020.06.0190

Link prediction method based on topological density weighting of common neighbor neighborhood

Li Xing, Zhu Yuhang, Bai Yi, Li Jinsong

(*PLA Strategic Support Force Information Engineering University*, *Zhengzhou* 450002, *China*)

Abstract: Link prediction aims to use known network nodes and topology information to predict the possibility of edges between two unconnected nodes in the network. The link prediction method based on network topological similarity has low computational complexity and good prediction effect, but the existing similarity indices take less consideration of the neighborhood topological information of common neighbors. To solve this problem, this paper proposed a link prediction method based on the weighted neighborhood topological denseness of the common neighbor. First, the method quantified the neighborhood topology of nodes based on the relative density index of neighborhood topology. Then, the node degree of the common neighbor and the relative density index of the neighborhood topology were used to describe the similarity contributions of the common neighbor and its surrounding topology. Finally, a node similarity index based on the weighting of the topological denseness of common neighbors was proposed. The experimental results on multiple actual networks show that the proposed method can achieve higher prediction accuracy compared with existing similarity indices.

Key words: complex network; link prediction; neighborhood topological denseness; topological weighting

0 引言

现实世界中的许多复杂系统都可以抽象为由节点和连边组成的复杂网络，节点表示网络中的实体，连边表示网络中实体间的关系。链路预测作为复杂网络中的重要研究方向之一[1~5]，旨在利用已知的网络节点和网络结构信息，预测网络中未连接的两个节点之间存在连边的可能性。其中，待预测的连边可能是网络中的缺失连边、错误连边，也可能是未知连边。链路预测不仅在社交网络、交通网络、蛋白质网络、食物链网络等现实网络中具有广泛的应用价值[6~9]，而且在网络演化、信息传播等研究中具有十分重要的理论价值[10~13]。

由于网络拓扑结构信息易于获取且数据可靠，因此基于网络结构相似性的链路预测方法一直是学者研究的热点。这类方法的基本思想是未连接节点之间的相似性越高，二者之间存在连边的可能性越大，其预测精度的高低关键是节点相似性的定义能否很好地反映网络的拓扑结构特征。根据节点相似性计算中所利用的网络结构信息，这类方法可分为基于局部信息、基于准局部信息和基于全局信息的相似性指标。

基于局部信息的相似性指标主要利用共同邻居及其周围拓扑信息进行链路预测，代表性方法有共同邻居指标 CN[14]、AA 指标[15]、偏好连接指标 PA[16]、资源分配指标 RA[17]、局部朴素贝叶斯指标 LNB[18]、CAR 指标[19]、聚类系数指标 CC[20]等。基于准局部信息的相似性指标通过多跳路径或局部随机游走引入了更多的网络结构信息，如局部路径指标 LP[21]不仅考虑了二阶路径(即共同邻居)，也考虑了三阶路径信息；刘等人[22]将 RA 指标扩展到三阶路径提出一种扩展的资源分配指标；局部随机游走指标 LRW[23]和有叠加效应的随机游走指标 SRW[23]基于有限步数的随机游走定义节点相似性。基于全局信息的相似性指标由于考虑了全部网络结构信息，其计算复杂度一般很高，难以处理大规模网络。这类指标主要有全局路径指标 Katz[24]、平均通勤时间指标 ACT[25]、余弦相似性指标 Cos+[26]、MFI 指标[27]和 SimRank 指标[28]等。

上述方法中，基于局部信息的相似性指标在保持较低计算复杂度的同时，能够取得较高的预测精度，因此非常适合解决大规模复杂网络中的链路预测问题。许多研究表明[14~20]，共同邻居在基于局部信息的链路预测方法中发挥了重要作用，并且不同的共同邻居对节点相似性的贡献也不尽相同。因此，基于局部信息的链路预测算法性能的好坏，主要取决于能否精确刻画不同共同邻居对节点相似性的贡献程度。Martínez 等人[29]认为，CN、AA 和 RA 指标是同一技术的变体。即，三个指标都假设两个节点之间存在连边的概率与共同邻居的数量成正比，并且都利用了共同邻居的节点度对相似性进行

收稿日期：2020-06-15；**修回日期：**2020-08-05 **基金项目：**国家自然科学基金资助项目

作者简介：李星(1987-)，男，河南新乡人，助理研究员，博士研究生，主要研究方向为网络科学、网络空间安全(lixing_ndsc@163.com)；朱宇航(1982-)，男，江苏徐州人，副教授，博士研究生，主要研究方向为网络科学、网络空间安全；柏溢(1975-)，男，江苏盐城人，副研究员，硕士，主要研究方向为移动网安全；李劲松(1992-)，男，山东泰安人，博士研究生，主要研究方向为网络科学、链路预测.

加权。其中，CN、AA 和 RA 指标的加权函数分别为空、$1/\log k_z$ 和 $1/k_z$。受此启发，Martínez 等人[29]通过研究共同邻居的最佳惩罚度与网络结构特征之间的关系，提出一种自适应度惩罚的链路预测方法，该方法利用网络的平均集聚系数对共同邻居节点的度进行惩罚。Wu 等人[30]利用共同邻居节点的重要性排序得分对相似性指标进行加权，提出一种基于重要节点发现指标的广义链路预测方法。该方法认为共同邻居节点的影响力越大，其对连边的贡献越小。李[31]等从资源传输角度出发，提出一种基于资源传输节点拓扑紧密性的链路预测方法。该方法利用重要传输节点紧密性对共同邻居传输资源量的影响刻画节点间的相似性。

受上述方法的启发，本文提出一种基于共同邻居邻域拓扑稠密性加权的链路预测方法。首先，通过分析节点邻域拓扑结构的稠密性，提出节点邻域拓扑结构稠密性的量化方法；其次，在分析共同邻居节点自身及其邻域拓扑对相似性计算的影响，提出基于邻域拓扑稠密性加权的相似性指标；最后，通过多个实际网络数据验证了所提方法的有效性。

1　基于共同邻居邻域拓扑稠密性加权的链路预测方法

1.1　节点邻域拓扑稠密性量化方法

网络拓扑结构对于基于相似性的链路预测方法至关重要。许多基于局部信息的相似性指标表明[14~20]，两个未连接的节点之间存在的共同邻居越多，其相似性越大。同时，由于每一个共同邻居在网络中所处的地位不同、周围拓扑结构也不同，因而其对潜在连接的贡献度也不同。例如，在社交网络中，对于素不相识的两个人来说，如果二人的共同朋友越多，那么他们成为好友的可能性就越大。但是，这些共同朋友中，活跃用户和普通用户对于促进二人相识所发挥的作用是不同的。一般来说，活跃用户的兴趣较多、社交广泛，可能经常组织或参加社交活动，因此引荐二人成为好友的概率较高；而普通用户的交际圈较小，促使二人相识的可能性较低。

如图 1 所示，节点 x 和 y 之间没有直接连边，z_1 和 z_2 为 x 和 y 的共同邻居。其中，节点 z_2 只有 x 和 y 两个邻居；节点 z_1 拥有包括 x 和 y 在内的 5 个邻居 $\{x, y, v_1, v_2, v_3\}$，且 $\{x, v_1\}$、$\{v_1, v_2\}$、$\{v_2, y\}$ 之间存在连接。可见，与节点 z_2 相比，节点 z_1 的邻域拓扑结构较为稠密，因而能够为 $\{x, y\}$ 之间连边的产生提供更大的可能性。

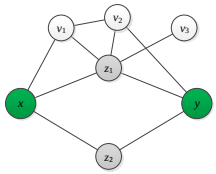


图 1　共同邻居的邻域拓扑结构对链路预测的影响

Fig. 1　Influence of the neighborhood topology of the common neighbor on the link prediction

为了刻画共同邻居节点的邻域网络拓扑结构对相似性的影响，首先给出节点的邻域拓扑相对稠密指数的定义。

定义 1　节点邻域拓扑相对稠密指数。对于网络中的任一节点 z，与该节点存在直接连边的所有邻居节点构成邻域 V_z，将邻域中两两之间存在连边的节点对数目与两两之间没有连边的节点对数目之比，定义为节点 z 的邻域拓扑相对稠密指数，即：

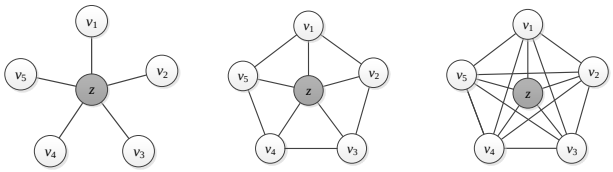
$$\varphi_z = \frac{\text{一阶邻域连边节点对数目}}{\text{一阶邻域无边节点对数目}} \tag{1}$$

可用公式表示为

$$\varphi_z = \frac{E^{V_z}}{k_z(k_z-1)/2 - E^{V_z}} \tag{2}$$

其中，V_z 为节点 z 的一阶邻域，E^{V_z} 表示一阶邻域 V_z 中实际连边数目，k_z 为节点 z 的度。

图 2 给出了几种典型的节点邻域拓扑结构示意图。其中，节点 z 有 5 个一阶邻居节点，即节点 z 的度 k_z=5。图 2(a)中节点 z 的邻居节点之间没有连边，图 2(b)中节点 z 的邻居节点之间有 5 条连边，图 2(c)中节点 z 的邻居节点两两之间都存在连边。根据式(2)，可以计算出图 2 中节点 z 的邻域拓扑相对稠密指数分别是 0，1，$+\infty$。可见，邻域拓扑相对稠密指数 φ_z 的取值范围为 $[0,+\infty)$，φ_z 越大表示节点 z 的邻域拓扑结构越稠密。



(a) 拓扑 1　　(b) 拓扑 2　　(c) 拓扑 3

图 2　节点的邻域拓扑稠密性示意图

Fig. 2　Schematic diagram of the node's neighborhood topological density

1.2　节点相似性指标

从图 2 可以看出，随着节点 z 的邻域拓扑相对稠密指数的增大，其邻域结构 V_z 的网络拓扑结构愈加复杂，从而为网络中与之有关系的节点对之间新连边的产生提供了更多的可能性。

本文综合考虑共同邻居自身及其邻域拓扑结构对节点相似性的影响，提出一种基于共同邻居邻域拓扑稠密性加权的链路预测方法。该方法利用共同邻居的节点度对共同邻居自身的相似度贡献进行加权，利用邻域拓扑相对稠密指数对共同邻居周围拓扑结构的相似度贡献进行加权。

定义 2　基于邻域拓扑稠密性加权的相似性指标(Local Density Weighting, LDW)。给定一个无向无权网络 $G(V, E)$，其中 V 表示网络节点集合，$E \subseteq V \times V$ 表示网络中连边的集合。对于网络中的任意两个节点 x 和 y，其相似性 $s^{LDW}(x,y)$ 可通过所有共同邻居节点的度及其邻域拓扑相对稠密指数进行量化，用公式表示为

$$s^{LDW}(x,y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \left(\frac{1+\varphi_z}{k_z} \right)^{\alpha} \tag{3}$$

其中，α 为调节参数，表示邻域拓扑结构稠密性强度。$\Gamma(x)$ 和 $\Gamma(y)$ 分别为节点 x 和 y 的邻居节点集合，k_z 为共同邻居 z 的节点度，φ_z 为共同邻居 z 的邻域拓扑相对稠密指数。可以看出，该方法认为共同邻居的节点度越大，节点之间的相似性越小；共同邻居的邻域拓扑结构越稠密，节点之间的相似性越大。值得注意的是，当调节参数 α=0 时，该方法等价于 CN 指标；当共同邻居 z 的邻居节点之间不存在连边(此时，$\varphi_z = 0$)且 α=1 时，该方法等价于 RA 指标。

将式(1)代入式(2)，可得

$$s^{LDW}(x,y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \left(\frac{k_z - 1}{k_z(k_z-1) - 2E^{V_z}} \right)^{\alpha} \tag{4}$$

根据式(4)，可以计算得到图 1 中节点 x 和 y 的相似性指标为 $s^{LDW}(x,y) = \left(\frac{5\text{-}1}{5\times(5\text{-}1)\text{-}2\times 3} \right)^{\alpha} + \left(\frac{2\text{-}1}{2\times(2\text{-}1)\text{-}2\times 0} \right)^{\alpha} = \left(\frac{2}{7} \right)^{\alpha} + \left(\frac{1}{2} \right)^{\alpha}$。

2　实验设置

2.1　网络数据集

为了充分验证所提 LDW 指标的有效性，本文从 Konect 开放数据库[32]中选取多个不同类型的网络数据集进行实验，分别介绍如下：

a) Email：罗维拉-威尔吉利大学的电子邮件通信网络；

b)SmaGri：一个关于“Small & Griffith and Descendants”的论文引用网络；c)SciMet：一个关于“科学计量学”的论文引用网络；d)Metabolic：一种表示秀丽隐杆线虫新陈代谢的网络；e)Figeys：一个人类蛋白质之间的交互网络；f)Jazz：一个爵士乐音乐家之间的合作网络，其中节点代表音乐家，连边代表合作关系；g)Kohonen：一个有关“自组织映射”主题的论文引用网络；h)Yeast：一个发芽酵母的蛋白质相互作用网络；i)KingJames：一个包含詹姆斯国王版《圣经》的名词以及其共现信息的词汇网络。

上述网络数据集的拓扑属性如表 1 所示，主要包括网络节点数$|V|$，连边数量$|E|$，平均节点度$\langle k\rangle$，平均集聚系数$\langle c\rangle$，平均最短路径$\langle d\rangle$。在实验过程中，一般将网络数据的边集合 E 随机划分为训练集 E^T 和测试集 E^P，其中 $E=E^T\cup E^P$ 且 $E^T\cap E^P=\varnothing$ 。

表 1　网络数据集的拓扑属性

Tab. 1　Topological properties of network datasets

网络数据集	$\|V\|$	$\|E\|$	$\langle k\rangle$	$\langle c\rangle$	$\langle d\rangle$
Email	2,029	39,264	38.70	0.890	3.38
SmaGri	1,024	4,918	4.80	0.197	4.95
SciMet	2,729	10,412	3.82	0.151	6.17
Metabolic	453	2,025	4.47	0.646	3.93
Figeys	2,239	6,452	5.76	0.040	3.98
Jazz	198	2,742	27.70	0.520	2.21
Kohonen	4,469	25,463	11.40	0.271	3.53
Yeast	2,375	11,693	9.85	0.378	5.10
KingJames	1,773	18,262	18.501	0.689	3.38

2.2　对比方法

为了验证本文提出的 LDW 指标的有效性，分别与基于局部信息的相似性指标 CN、AA、PA、RA、CC，基于准局部信息的相似性指标 LP，基于全局信息的相似性指标 Katz、ACT、MFI，以及基于节点重要性排序的相似性指标 INI-PR 进行对比实验。各对比方法简要介绍如下：

a) CN 指标[14]：将两个未连边节点之间共同邻居节点的数量作为相似性指标，其计算公式如下：

$$s^{CN}(x,y)=|\Gamma(x)\cap\Gamma(y)| \tag{5}$$

即，两个节点之间的共同邻居越多，其相似性越大。

b)　AA 指标[15]：将共同邻居节点的度数取对数，然后对共同邻居的相似性贡献进行加权。计算公式如下：

$$s^{AA}(x,y)=\sum_{z\in\Gamma(x)\cap\Gamma(y)}\frac{1}{\log(k_z)} \tag{6}$$

c) PA 指标[16]：偏好连接相似性指标，认为网络中的节点倾向于与大度节点产生连接。计算公式如下：

$$s^{PA}(x,y)=k_x\cdot k_y \tag{7}$$

d) RA 指标[17]：直接将共同邻居节点度的倒数作为共同邻居相似性贡献的权重。计算公式如下：

$$s^{RA}(x,y)=\sum_{z\in\Gamma(x)\cap\Gamma(y)}\frac{1}{k_z} \tag{8}$$

e) CC 指标[20]：将所有共同邻居节点的集聚系数之和作为相似性指标。计算公式如下：

$$s^{CC}(x,y)=\sum_{z\in\Gamma(x)\cap\Gamma(y)}c_z \tag{9}$$

其中，c_z 表示节点 x 和 y 的共同邻居节点 z 的集聚系数。

f) LP 指标[21]：不仅考虑二阶路径(共同邻居)信息，也考虑了三阶路径的影响力，计算公式如下：

$$\boldsymbol{S}^{LP}=\boldsymbol{A}^2+\alpha\boldsymbol{A}^3 \tag{10}$$

其中，$\boldsymbol{A}^2$ 和 $\boldsymbol{A}^3$ 分别表示路径长度为 2 和 3 的邻接矩阵，α 为权重调节参数。

g) Katz 指标[24]：在 LP 指标基础上，考虑全部路径信息，计算公式如下:

$$\boldsymbol{S}^{Katz}=\alpha\boldsymbol{A}+\alpha^2\boldsymbol{A}^2+\alpha^3\boldsymbol{A}^3+\cdots+\alpha^n\boldsymbol{A}^n \tag{11}$$

h) ACT 指标[25]：平均通勤时间，可以理解为网络中一个粒子从一个节点随机游走到另一个节点所需的平均步数。其计算公式如下：

$$s^{ACT}(x,y)=\frac{1}{l^+_{xx}+l^+_{yy}-2l^+_{xy}} \tag{12}$$

其中，l^+_{xy} 表示网络的拉普拉斯矩阵的伪逆中第 x 行 y 列的元素值。

i) MFI 指标[27]：基于矩阵森林理论提出的矩阵森林指标，其定义为

$$\boldsymbol{S}^{MFI}=(\boldsymbol{I}+\boldsymbol{L})^{-1} \tag{13}$$

其中，$\boldsymbol{L}$ 为网络的拉普拉斯矩阵。

j) INI-PR 指标[30]：基于节点重要性排序的相似性指标。其基本思想是利用重要节点发现方法刻画节点的全局影响力，然后将节点重要性得分作为相似性计算的惩罚因子。当以 PageRank 算法计算节点重要性得分时，该指标的定义为

$$S^{INI-PR}(x,y)=\sum_{z\in\Gamma(x)\cap\Gamma(y)}\frac{1}{PR_z}=\sum_{z\in\Gamma(x)\cap\Gamma(y)}\frac{1}{(1-d)+d\sum_{v\in\Gamma(z)}\frac{R_v}{|\Gamma(v)|}} \tag{14}$$

其中，PR_z 表示节点 x 和 y 的共同邻居节点 z 的 PageRank 得分。d 为抑制因子，一般设置为 0.85。

2.3　评价指标

本文所关注的链路预测算法评价指标为 Precision[33]，其定义为将预测得到的连边可能性由高到低排序后，前 L 个预测边中预测正确的比例。计算公式为

$$\text{Precision}=\frac{m}{L} \tag{15}$$

其中，m 表示前 L 个预测结果中出现在测试集中的连边数量。显然，该评价指标与 L 的取值有关。一般来说，L 取值为 100[34]。

实验中，设置训练集 E^T 中连边数目占比为 90%，测试集 E^P 中连边数目占比为 10%。衡量指标 Precision 的结果都是 30 次独立实验结果的平均值，并且每次实验都重新划分训练集和测试集，以保证实验结果的可靠性。同时，去掉训练集和测试集中的孤立节点，保证数据集的网络连通性。

3　实验结果及分析

首先，在实际网络数据集上测试共同邻域拓扑稠密性加权强度对 Precision 结果的影响；然后，通过与其他相似性指标的对比结果分析所提指标的有效性。

3.1　邻域拓扑稠密性强度对预测结果的影响

图 3 所示为所提方法在 9 个不同网络中的 Precision 指标随加权参数 α 的变化曲线。其中，垂直虚线为预测精度最高时加权参数的取值。

可以看出，与 $\alpha=0$ 相比(此时 LDW 指标等价为 CN 指标)，随着参数 α 的增大，所有网络的 Precision 指标均出现明显上升，说明本文所定义的邻域拓扑稠密性加权能够显著提高链路预测的精度。在 Email、Metabolic、Kohonen 和 Yeast 网络中，Precision 指标随着参数 α 的增大而迅速增大，并在 $\alpha=1$ 附近达到峰值，说明邻域拓扑稠密性强度的作用非常明显。随着参数 α 的继续增大，Precision 指标逐渐下降，并趋于平稳，说明较大的邻域拓扑稠密性强度加权对相似性的影响较弱。在 SmaGri、SciMet 和 Jazz 网络中，随着参数 α 的增大，Precision 指标出现一定程度的上升，而后下降并趋于平稳，说明邻域拓扑稠密性强度的影响先增强后减弱。与上述网络不同，在 Figeys 和 KingJames 网络中，Precision 指标随着参数 α 的增大逐渐增大，并能够始终保持在较高的取值范围内。这说明邻域拓扑稠密性随着加权强度的增大，对相

似性刻画的作用逐渐增大，且当达到一定程度后对相似性的贡献不再明显增大。值得注意的是，Jazz 网络中，在 $\alpha=0$ 时已经取得较高预测精度的情况下，通过拓扑稠密性加权后仍然能够明显提高预测精度；KingJames 网络中，通过邻域拓扑稠密性加权，预测精度提高了约 100%。这说明本文定义的邻域拓扑稠密指数加权方法能够较好的刻画共同邻居对节点相似性的贡献，对提高链路预测精度具有明显的作用。

此外，图 3 的实验数据表明，LDW 指标在不同网络中取得最高预测精度时对应的参数取值不尽相同。因此，在实际应用中，可以将已知的网络连边划分为训练集和测试集，通过仿真得到该网络的最优参数取值；然后，再利用该最优参数进行链路预测。

3.2　与其他相似性指标的对比分析

表 2 给出了所提 LDW 指标和其他 10 种对比指标在 9 个网络数据集上的 Precision 结果。其中，加粗字体表示所有方法中的最高预测精度。可以看出，除了在 Kohonen 网络中，LDW 指标的预测精度仅次于 CC 指标以外，在其余所有网络中都能够取得最好的 Precision 结果。与其他局部、准局部和全局相似性指标相比，本文方法的 Precision 指标均实现了不同幅度的提高，说明所提 LDW 指标的有效性。

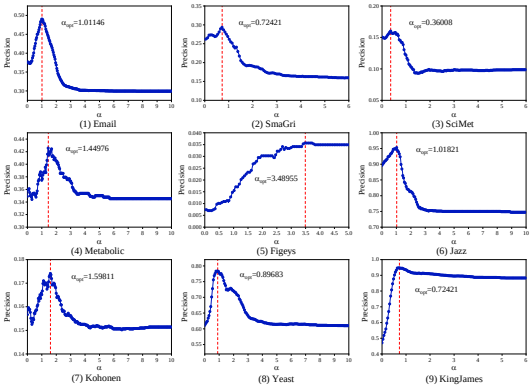


图 3　所提方法在不同网络中的 Precision 指标随加权参数的变化曲线

Fig. 3　Variation curve of Precision values of the proposed method with different weighting parameters in different networks

表 2　Precision 指标的实验结果分析

Tab. 2　Analysis of experimental results of Precision values

Precision	CN	AA	PA	RA	CC	LP*	Katz*	ACT	MFI	INI-PR	LDW
Email	0.3719	0.3724	0.2238	0.4025	0.3888	0.3651	0.3752	0.0000	0.0404	0.4060	**0.4901**
SmaGri	0.2343	0.2374	0.0909	0.2298	0.2919	0.2348	0.2737	0.0000	0.0303	0.2365	**0.2934**
SciMet	0.1425	0.1341	0.0327	0.1175	0.1548	0.1464	0.1500	0.0000	0.0115	0.1337	**0.1601**
Metabolic	0.3354	0.3280	0.2659	0.3280	0.3805	0.3415	0.3878	0.2317	0.2098	0.3046	**0.4256**
Figeys	0.0050	0.0062	0.0031	0.0070	0.0241	0.0054	0.0054	0.0015	0.0000	0.0112	**0.0361**
Jazz	0.9018	0.9109	0.3045	0.9255	0.9273	0.8973	0.8836	0.4109	0.1982	0.9155	**0.9527**
Kohonen	0.1398	0.1369	0.0357	0.1296	**0.1926**	0.1447	0.1549	0.0000	0.0439	0.1404	0.1741
Yeast	0.6004	0.6239	0.4385	0.4803	0.5957	0.6085	0.6085	0.4957	0.0598	0.6147	**0.7850**
KingJames	0.4495	0.5776	0.2675	0.8697	0.8230	0.4407	0.4224	0.2077	0.3650	0.8274	**0.9475**

注：*可调参数取值为 0.01。

在基于局部信息的相似性指标中，PA 指标只利用了待预测节点的度，其预测精度 Precision 在所有网络中表现最差；其余 CN、AA、RA 和 CC 指标在不同网络中的表现相差不大、各有千秋，CC 指标由于考虑了共同邻居的集聚性对节点相似性的影响，在 6 个网络中取得最高预测精度，RA 指标在 2 个网络中取得最高预测精度，AA 指标在 1 个网络中取得最高预测精度。在 Figeys 和 KingJames 网络中，CC 指标的 Precision 指标提高幅度较为明显。部分网络，如 Figeys、Jazz、和 KingJames 等，局部相似性指标的预测精度甚至优于准局部和全局的相似性指标。基于准局部信息的 LP 指标和基于全局信息的 Katz 指标在不同网络中预测精度相当，且与 CN 指标相差不大。但是，与 CC 指标相比，LP 指标和 Katz 指标的预测效果略差，尤其是在 Figeys、Jazz、Kohonen 和 KingJames 网络中，LP 指标和 Katz 指标的 Precision 取值明显低于 CC 指标。对于全局相似性指标 ACT 和 MFI 来说，尽管它们利用了网络的全局结构信息，但在多数网络中，其 Precision 指标普遍较低，部分网络甚至趋近于 0，说明 ACT 和 MFI 指标可能不适用于 Precision 衡量标准。新近提出的 INI-PR 指标通过计算共同邻居的重要性排序得分不仅利用了局部结构信息，而且保留了全局结构信息。与局部相似性指标相比，INI-PR 指标在大多数网络中取得比 CN、AA、PA 更高或者相当的预测精度，在 6 个网络中取得比 RA 指标更高的预测精度。而与 CC 指标相比，INI-PR 指标仅在 Email、Yeast 和 KingJames 网络中预测精度略高。与准局部和全局相似性指标相比，INI-PR 指标在部分网络，如 Email、Figeys、Jazz、Yeast 和 KingJames 中，取得了更高的预测精度。

由于在计算节点相似性时利用邻域拓扑稠密性进行加权，本文所提的 LDW 指标在所有网络中表现最好(除在 Kohonen 网络中略低于 CC 指标)。在 Precision 衡量标准下，与局部相似性指标相比，LDW 指标的平均提高幅度为 13%，其中在 Figeys、Yeast 和 Email 网络中最为明显，提高幅度分别为 50%、26%和 22%；与准局部相似性指标相比，LDW 指标的平均提高幅度为 92%，尤其是 Figeys 和 KingJames 网络中最为明显。与全局相似性指标相比，LDW 的平均提高幅度为 88%。与基于节点重要性的相似性指标 INI-PR 相比，LDW 指标的预测精度平均提高了约 44%。

综上所述，本文的 LDW 指标能够较好的刻画共同邻居在节点相似性计算中的贡献率，在 Precision 衡量标准下，其链路预测效果优于现有指标。并且，本文方法的计算复杂度较低(介于 CN 指标和 LP 指标之间)，能够适用于大规模复杂网络的链路预测。

4　结束语

通过分析共同邻居自身及其邻域拓扑结构稠密性对节点相似性的影响，提出一种基于邻域拓扑稠密性加权的链路预测方法。该方法首先根据节点的邻居之间连边稠密度，给出节点的邻域拓扑相对稠密指数定义；然后，分别将共同邻居的节点度和邻域拓扑相对稠密指数作为相似性指标的权重，提出一种基于邻域拓扑稠密性加权的链路预测方法。多个实际网络数据的实验表明，所提方法取得了比其他方法更高的预测精度。在后续工作中，将研究共同邻居更大范围内的邻域拓扑结构对节点相似性的影响。

参考文献：

[1] Lyu Linyuan, Zhou Tao. Link prediction in complex networks: a survey [J]. Physica A: Statistical Mechanics and its Applications, 2011, 390 (6): 1150-1170.

[2] Martínez V, Berzal F, Cubero J C. A survey of link prediction in complex networks [J]. ACM Computing Surveys (CSUR) , 2017, 49 (4): 69.

[3] Li Xing, Liu Shuxin, Chen Hongchang, *et al.* A potential information capacity index for link prediction of complex networks based on the cannikin law [J]. Entropy. 2019, 21 (9): 863.

[4] Benson A R, Abebe R, Schaub M T, *et al.* Simplicial closure and higher-order link prediction [J]. The National Academy of Sciences of the United States of America, 2018, 115 (48): E11221-E11230. doi: 10. 1073/pnas. 1800683115.

[5] 王凯, 刘树新, 陈鸿昶, 等. 一种基于节点间资源承载度的链路预测方法 [J]. 电子与信息学报, 2019, 41 (5): 1225-1234. doi: 10. 11999/JEIT180553. (Wang Kai, Liu Shuxin, Chen Hongchang, *et al.* A new link prediction method for complex networks based on resources carrying capacity between nodes [J]. Journal of Electronics & Information Technology, 2019, 41 (5): 1225-1234. doi: 10. 11999/JEIT180553.)

[6] Dong Yuxiao, Tang Jie, Wu Sen, *et al.* Link prediction and recommendation across heterogeneous social networks [C]// IEEE International Conference on Data Mining, 2013: 181-190.

[7] Chen Zhenhao, Wu Jiajing, Xia Yongxiang, *et al.* Robustness of interdependent power grids and communication networks: a complex network perspective [J]. IEEE Trans on Circuits and Systems II: Express Briefs, 2018, 65 (1): 115-119. doi: 10. 1109/TCSII. 2017. 2705758.

[8] Cui Ying, Cai Meng, Dai Yang, *et al.* A hybrid network-based method for the detection of disease-related genes [J]. Physica A: Statistical Mechanics and its Applications, 2018, 492: 389-394.

[9] Ai Jun, Liu Yayun, Su Zhan, *et al.* Link prediction in recommender systems based on multi-factor network modeling and community detection [J]. Europhysics Letters (EPL) , 2019, 126 (3): 38003.

[10] Du Wenbo, Zhang Mingyuan, Ying Wen, *et al.* The networked evolutionary algorithm: a network science perspective [J]. Applied Mathematics and Computation, 2018, 338: 33-43. doi: 10. 1016/j. amc. 2018. 06. 002.

[11] 刘树新, 季新生, 刘彩霞, 等. 局部拓扑信息耦合促进网络演化 [J]. 电子与信息学报, 2016, 38 (9): 2180-2187. (Liu Shuxin, Ji Xinsheng, Liu Caixia, *et al.* Information coupling of local topology promoting the network evolution [J]. Journal of Electronics & Information Technology, 2016, 38 (9): 2180-2187.)

[12] 刘树新, 季新生, 刘彩霞, 等. 一种信息传播促进网络增长的网络演化模型 [J]. 物理学报, 2014, 63 (15): 158902-158902. (Liu Shuxin, Ji Xinsheng, Liu Caixia, *et al.* A complex network evolution model for network growth promoted by information transmission [J]. Acta Physica Sinica, 2014, 63 (15) .)

[13] Li Dong, Zhang Yongchao, Xu Zhiming, *et al.* Exploiting information diffusion feature for link prediction in sina weibo [J]. Scientific Reports, 2016, 6: 20058.

[14] Lorrain F, White H C. Structural equivalence of individuals in social networks [M]. LEINHARDT S. Social Networks: A Developing Paradigm. Lausanne: Academic Press, 1977: 67-98. doi: 10. 1080/0022250X. 1971. 9989788.

[15] Adamic L A, Adar E. Friends and neighbors on the web [J]. Social Networks, 2003, 25 (3): 211-230.

[16] Xie Yanbo, Zhou Tao, Wang Binhong. Scale-free networks without growth [J]. Physica A: Statistical Mechanics and its Applications, 2008, 387 (7): 1683-1688. doi: 10. 1016/j. physa. 2007. 11. 005.

[17] Zhou Tao, Lyu Linyuan, Zhang Yicheng. Predicting missing links via local information [J]. The European Physical Journal B, 2009, 71 (4): 623-630. doi: 10. 1140/epjb/e2009-00335-8.

[18] Liu Zhen, Zhang Qianming, Lyu Linyuan, *et al.* Link prediction in complex networks: a local naive bayes model [J]. Europhysics Letters (EPL) , 2011, 96 (4): 48007.

[19] Cannistraci C V, Alanis-lobato G, Ravasi T. From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks [J]. Scientific Reports, 2013, 3: 1613. doi: 10. 1038/srep01613.

[20] Wu Zhihao, Lin Youfang, Wang Jing, *et al.* Link prediction with node clustering coefficient [J]. Physica A, 2016, 452: 1-8.

[21] Lyu Linyuan, Jin Cihang, Zhou Tao. Similarity index based on local paths for link prediction of complex networks [J]. Physical Review E, 2009, 80 (4): 046122. doi: 10. 1103/PhysRevE. 80. 046122.

[22] Liu Shuxin, Ji Xinsheng, Liu Caixia, *et al.* Extended resource allocation index for link prediction of complex network [J]. Physica A: Statistical Mechanics and its Applications, 2017, 479: 174-183. doi: 10. 1016/j. physa. 2017. 02. 078.

[23] Liu Weiping, Lyu Linyuan. Link prediction based on local random walk [J]. Europhysics Letters (EPL) , 2010, 89 (5): 58007.

[24] Katz L. A new status index derived from sociometric analysis [J]. Psychometrika, 1953, 18 (1): 39-43. doi: 10. 1007/BF02289026.

[25] Klein D J, Randić M. Resistance distance [J]. Journal of Mathematical Chemistry, 1993, 12 (1): 81-95. doi: 10. 1007/BF011 64627.

[26] Fouss F, Pirotte A, Renders J M, *et al.* Random walk computation of similarities between nodes of a graph with application to collaborative recommendation [J]. IEEE Trans on Knowledge and Data Engineering, 2007, 19 (3): 355-369. doi: 10. 1109/tkde. 2007. 46.

[27] Chebotarev P, Shamis E. The matrix-forest theorem and measuring relations in small social groups [J]. Automation and Remote Control, 2006, 58 (9): 1505-1514.

[28] Jeh G, Widom J. SimRank: a measure of structural-context similarity [C]// Proc of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'02) , Edmonton, Canada, 2002: 538-543.

[29] Martínez V, Berzal F, Cubero J C. Adaptive degree penalization for link prediction [J]. Journal of Computational Science, 2016, 13: 1-9.

[30] Wu Jiehua, Shen Jing, Zhou Bei, *et al.* General link prediction with influential node identification [J]. Physica A: Statistical Mechanics and its Applications, 2019, 523: 996-1007.

[31] 李英乐, 何赞园, 王凯, 等. 基于资源传输节点拓扑紧密性的链路预测方法 [J/OL]. 计算机工程, 2020. (2020-01-04) [2020-06-15]. https://doi. org/10. 19678/j. issn. 1000-3428. 0056092. (Li Yingle, He Zanyuan, Wang Kai, *et al.* Link prediction method based on topological tightness of resource transmission nodes [J/OL]. Computer Engineering, 2020. (2020-01-04) [2020-06-15]. https://doi. org/10. 19678/j. issn. 1000-3428. 0056092.)

[32] Kunegis J. Konect network dataset [EB/OL]. (2017) [2020-06-15]. http://konect. uni-koblenz. de/networks/.

[33] Wu Zhihao, Lin Youfang, Zhao Yiji, *et al.* Improving local clustering based top-L link prediction methods via asymmetric link clustering information [J]. Physica A: Statistical Mechanics and its Applications, 2018, 492: 1859-1874.

[34] Lyu Linyuan, Zhou Tao. Link prediction in weighted networks: the role of weak ties [J]. Europhysics Letters (EPL) , 2010, 89 (1): 18001.