

• 复杂性科学 •



## 链路预测中的局部相似性指标

李艳丽, 周涛\*

(电子科技大学复杂性科学实验室 成都 611731)

**【摘要】**链路预测是网络科学中一个重要且充满挑战的研究方向,其在社交网络中的朋友推荐、生物实验中的关系发现、搜索引擎中的链接导航以及电商平台中的商品推荐等领域发挥着不可忽视的作用。链路预测研究兴起的 20 余年里,各类链路预测算法层出不穷,其中局部相似性指标以其简洁性、可解释性、较低的运算时间、灵活的可扩展性以及有竞争力的预测准确度等优势迅速在多个相关研究领域和实际应用场景中被广泛应用。这些指标大多基于同质性、聚集性、三角闭包等理论在 2 阶邻居分析框架中提出。但最近 10 年,局部社团范式理论的提出、赫布律的应用以及针对 2 阶框架合理性的争议等一系列重要工作的出现极大推动了局部相似性指标的深入研究和讨论。该文旨在针对这些新的理论和争议进行梳理和讨论。

**关键词** 2 阶路径; 3 阶路径; 复杂网络; 链路预测; 局部相似性指标  
**中图分类号** TP391 **文献标志码** A **doi**:10.12178/1001-0548.2021062

## Local Similarity Indices in Link Prediction

LI Yan-li and ZHOU Tao\*

(ComplexX Lab, University of Electronic Science and Technology of China Chengdu 611731)

**Abstract** Link prediction is a significant and challenging task in network science, which plays an important role in friend recommendations, the discovery of biological interactions, link navigation, and product recommendations. Since the rise of link prediction, many methods have been proposed. Due to the simplicity, interpretability, high efficiency, scalability, and satisfactory performance, local similarity indices are widely used in various research fields and applications. Under the 2-hop-based neighborhood analytical framework, most of the indices are proposed based on the network organization mechanisms including homophily, clustering and triadic closure. In the last decade, the emergence of local community paradigm, Hebb theory and new arguments about the rationality of the 2-hop-based framework has greatly promoted the development of local similarity indices. This paper aims at sorting out and discussing these new findings.

**Key words** 2-hop; 3-hop; complex network; link prediction; similarity indices

链路预测的核心任务是预测两个没有连接关系的节点产生链接的可能性<sup>[1-4]</sup>。该研究方向最初的兴起是为了辅助万维网用户从海量的网页中寻找感兴趣的网页<sup>[5-6]</sup>。后来逐渐在朋友关系预测<sup>[7]</sup>、恐怖分子的发现<sup>[8]</sup>、潜在的学术合作关系预测<sup>[9]</sup>、生物学相互作用关系的揭示<sup>[10-11]</sup>、商品推荐<sup>[12]</sup>、网络生成机制探索<sup>[13-14]</sup>和网络可预测性<sup>[15-17]</sup>等问题上发挥了巨大价值。其预测类型囊括了根据观测到的网络结构预测可能存在的缺失连边 (missing link), 未来可能产生的连边 (future links) 以及甄别虚假连边 (spurious link) 等<sup>[1,18]</sup>。其研究对象涵盖了简单网

络、有向网络、二分网络、异构网络和时序网络等<sup>[19]</sup>。其方法论涉及了基于相似性的链路预测算法 (similarity-based algorithms)<sup>[9,20]</sup>、概率模型 (probabilistic models)<sup>[21-23]</sup>、最大似然模型 (maximum likelihood methods)<sup>[8,18,24]</sup>、网络嵌入模型 (network embedding)<sup>[25-28]</sup> 和其他方法<sup>[29-31]</sup>。

基于相似性的链路预测算法尤其是局部相似性指标可应用领域最为广泛。因为它设计简洁、可解释性强、运算时间低、灵活可扩展,同时其预测准确度有时甚至优于相对复杂的概率模型、最大似然模型和网络嵌入模型<sup>[31-33]</sup>。该类算法的基本思路是

收稿日期: 2021-02-05; 修回日期: 2021-03-08

基金项目: 国家自然科学基金(11975071, 61673086, 61673085)

作者简介: 李艳丽(1990-), 女, 博士生, 主要从事链路预测、关键节点挖掘、推荐系统等方面的研究。

通信作者: 周涛, E-mail: zhutou@ustc.edu

利用节点的局部拓扑结构信息为每一条候选连边分配一个相似性得分, 得分越高的连边被认为有更大可能是缺失边, 因此在预测列表中排序更靠前。这些算法中最具代表性的几个工作分别是文献 [9] 总结的纯粹基于网络拓扑结构的一系列局部相似性指标的工作、文献 [34] 提出 AA 指标的工作和文献 [20] 提出 RA 指标 (resource allocation) 的工作。文献 [9] 首次明确提出了基于网络拓扑结构的相似性定义框架, 同时一个非常简单且符合直觉的相似性指标——共同邻居指标 (common neighbors, CN) 也是在该文中被明确强调, 即两个节点如果拥有越多的共同邻居, 则越相似。定义  $\Gamma_x$  为节点  $x$  的邻居集合, 则网络中节点  $x$  和节点  $y$  的共同邻居指标可定义为:

$$S_{xy}^{CN} = |\Gamma_x \cap \Gamma_y| \quad (1)$$

不同于 CN 指标, AA 指标不仅考虑两个节点拥有的共同邻居数目, 而且考虑共同邻居的角色差异。具体定义如下:

$$S_{xy}^{AA} = \sum_{z \in \Gamma_x \cap \Gamma_y} \frac{1}{\log(k_z)} \quad (2)$$

式中,  $k_z$  表示节点  $z$  的度。该指标大大削弱了共同邻居中大度节点的贡献。RA 是众多局部相似性指标中少有的包含物理过程的指标 [35]。由资源分配的物理过程出发, 两个节点的相似性由从其中一个节点传递到另一个节点的资源量决定。节点  $x$  的每一个邻居都拥有一个单位的资源, 并将各自拥有的资源平均分配给它们的邻居, 节点  $y$  从中接收到的资源量即为两者的相似性值, 即:

$$S_{xy}^{RA} = \sum_{z \in \Gamma_x \cap \Gamma_y} \frac{1}{k_z} \quad (3)$$

从数学形式上看, CN、AA 和 RA 指标的区别在于是否考虑共同邻居的角色差异。在 CN 指标中, 所有共同邻居的贡献相同, AA 和 RA 则都是削弱了大度共同邻居的贡献, 只是一个从物理过程出发设计, 一个从信息论角度出发设计。从预测效果上看, 三者节点度相差不大的网络, 即有较小的度异质性的网络中, 预测效果不会有显著差异。但在度异质性大, 且平均度较大的网络, CN、AA 和 RA 指标在预测效果上差异显著。在大多数网络中, RA 预测效果最好, AA 次之, CN 最差。原因有三: 一是大多数真实网络的度分布呈长尾特性, 度异质性较强 [36-37]; 二是大度节点成为共同邻

居比较常见, 难以体现节点间的相似性, 因此惩罚大度节点可以更好刻画大度共同邻居与小度共同邻居不同的角色; 最后, CN 简并度高, 很多候选边的 CN 值相同, 而 AA 和 RA 的区分度远胜于 CN。

这些工作可用相同且坚实的理论和实证依据进行解释, 包括社会学中的同质性原理 [38], 即两个相似的节点更大概率产生连边。三角闭包理论 [39], 即“朋友的朋友就是朋友”和物理学中的聚集性原理 [40], 即两个拥有更多共同邻居的节点更大概率产生连边。这一系列暗含同一内涵的理论早在 2001 年的合作演化网络 [40] 和 2006 年的邮件通信演化网络 [41] 中得到了验证。同时也被证实是一些非社交网络中同样成立 [42]。

至此, 大多数局部相似性指标的研究工作都在上述理论框架中开展。通过更精细的刻画共同邻居节点以及候选节点对的自身拓扑结构差异, 引入并更好地利用节点的属性信息。研究人员发展出了一系列局部相似性指标, 并一次又一次地刷新了预测效果 [1,19,43-45]。一般而言, 在如此简单的理论框架下很难产生突破, 但最近的若干工作跳出了该框架的约束, 提出了新的连边理论并对原始基于共同邻居的局部相似性指标设计框架发起了挑战。这些工作的出现将使我们节点间局部连接模式的理解进入一个新的阶段。

## 1 LCP 和 Cannistraci-Hebb 理论

在社团结构明显的网络中, 同一社团内部产生连边的概率要大于社团之间产生连边的概率。基于该思想, 只要增加同一社团内部连边产生的概率, 预测性能便可以提高。最直接的做法是使用一些现成的社团识别算法先识别社团, 再区分对待社团内部和社团之间的节点连边概率 [46-47], 但这样做并不能为我们理解和利用上述思想提供更深刻的洞见和启发。针对没有显式社团标签的网络, 有没有可能设计出一套考虑节点间社团连接模式的高效且性能良好的链路预测指标呢?

2013 年, 文献 [48] 做出了非常有价值的尝试和推动, 如果把最近十年局部相似性指标发展史上有推动意义的工作收录成榜, 这一贡献绝对名列前茅。受到以边为研究对象对社团进行划分的启发 [48], 文献 [49] 提出了两个节点之间由边构成的局部社团内部连接越紧密, 两个节点越容易产生连边的局部社团范式 (local-community-paradigm, LCP)。LCP 指由共同邻居及其之间的连边形成的局部结构, 展

示的是一种不依赖于节点标签和边标签的局部社团描述方法。尽管也有少量工作有类似的出发点<sup>[47,50]</sup>，但并没有形成一种坚实清晰且干净利落的方法论。LCP作为一种简单独立的结构，很容易作为一种性能增强插件融入到过去已有的局部相似性指标中。如针对CN指标，LCP融入后形成的CAR指标表示如下：

$$S_{xy}^{\text{CAR}} = S_{xy}^{\text{CN}} \sum_{z \in \Gamma_x \cap \Gamma_y} \frac{|\gamma_z|}{2} \quad (4)$$

式中， $\gamma_z$ 表示在节点 $x$ 和节点 $y$ 形成的LCP结构中，节点 $z$ 的节点度，亦即节点 $z$ 与 $x$ 和 $y$ 的其他共同邻居的连边数。类似地，针对RA指标，LCP融入后形成的CRA指标可表示为：

$$S_{xy}^{\text{CRA}} = \sum_{z \in \Gamma_x \cap \Gamma_y} \frac{|\gamma_z|}{k_z} \quad (5)$$

从数学表达形式上来看，相对于CN和RA等指标，基于LCP的局部相似性指标对连边相似度有更强的区分度。但这些LCP增强指标存在一个明显的缺陷，在共同邻居之间连边稀疏甚至没有连边的网络中，它们的预测性能将差于原始的RA指标，甚至CN指标。

LCP的核心思想与著名的“赫布律”(Hebb Theory)一致<sup>[51]</sup>。赫布律的核心思想是知识、学习过程或记忆痕迹存储于由神经元(可视作网络中的节点)和突触(可视作网络中的边)构成的突触集合中，而非神经元个体。文献[52]在2018年指出网络的拓扑结构所起的关键作用便是将不同功能的“突触集合”隔离开，每个“突触集合”中的神经元通过在集合内部加强或者产生新的突触来完成信息的局部处理过程。这也就意味着连边更大概率产生于由边构成的隔离度良好的局部社团内部，这恰与前文提到的LCP范式一致。基于CRA进行微调后的Cannistraci-Hebb指标可表示为：

$$S_{xy}^{\text{CH}} = \sum_{z \in \Gamma_x \cap \Gamma_y} \frac{1 + k_z^{(i)}}{1 + k_z^{(e)}} \quad (6)$$

式中， $k_z^{(i)}$ 为式(4)和式(5)中的 $\gamma_z$ ； $k_z^{(e)}$ 则表示为节点 $z$ 与除 $x$ 、 $y$ 本身及其共同邻居以外节点的连边数。该指标在沿袭了LCP范式的同时，更直观地刻画了两个节点共同邻居构成的局部社团与外界的隔离度，既奖励了局部社团内的连边，又惩罚了跨出局部社团因而降低隔离度的连边。此外，Cannistraci-Hebb指标进行了拉普拉斯平滑操作，

即在分子、分母中添加常数项1。该常规意义上规避零概率的操作把具有共同邻居但邻居之间不存在连边的情况纳入进来，从而弥补了前文中所述的LCP增强指标存在的缺陷。该缺陷的弥补大幅度提升了Cannistraci-Hebb指标相对于CRA的预测效果。

LCP和Cannistraci-Hebb理论是局部相似性指标研究历程重要且令人印象深刻的系列研究成果，该工作仅仅是对局部社团连接范式的初步尝试，就已获得卓越的预测效果，它必将为我们理解节点之间的局部连接模式、刻画无标签依赖的局部社团结构以及设计出更优美有效的相似性指标带来灵感。

## 2 路径越短价值越高？

长久以来研究人员一个不证自明的认识是节点间较短连边的数目要比较长连边的数目更能刻画节点之间产生连边的似然。因为节点之间的互动关系和相互影响会随着连边距离的增大而衰减<sup>[53-54]</sup>。该现象在众多的基于2阶路径设计的相似性指标上可见一斑。即便考虑了较长路径连边数目的相似性指标(如Katz<sup>[55]</sup>和LP<sup>[56]</sup>)，短路径也永远扮演主要角色，路径长度越长，贡献越小。这些长路径的加入主要起到增加连边似然分辨率的作用。然而2018年之后连续出现了一系列工作表明3阶路径比2阶路径能更有效地预测未知链路<sup>[52,57-58]</sup>。这非常反直觉甚至令人生疑，如果这一发现存在于大多数网络中，将会是对过去认知的巨大挑战。

文献[57]假定节点 $x$ 和节点 $y$ 的连边似然可以由 $x$ 的所有邻居贡献值的线性求和表示，即：

$$s_{xy} = (\mathbf{AZ})_{xy} = \sum_{j \in \Gamma_x} a_{xj} z_{jy} \quad (7)$$

式中， $\mathbf{Z}$ 为一个待求解的矩阵。通过最小化 $\mathbf{Z}$ 与观测到的邻接矩阵 $\mathbf{A}$ 的差异得到以下优化函数：

$$\min_{\mathbf{Z}} \alpha \|\mathbf{A} - \mathbf{AZ}\|_F^2 + \|\mathbf{Z}\|_F^2 \quad (8)$$

该优化问题有一个闭合解，即

$$\mathbf{Z}^* = \alpha(\alpha \mathbf{A}^T \mathbf{A} + \mathbf{I})^{-1} \mathbf{A}^T \mathbf{A} \quad (9)$$

式中， $\mathbf{I}$ 为单位矩阵。从而，节点 $x$ 和节点 $y$ 的LO连边似然可表示为：

$$S_{xy}^{\text{LO}} = (\mathbf{AZ}^*)_{xy} = [\alpha \mathbf{A}(\alpha \mathbf{A}^T \mathbf{A} + \mathbf{I})^{-1} \mathbf{A}^T \mathbf{A}]_{xy} = [\alpha \mathbf{A}^3 - \alpha^2 \mathbf{A}^5 + \alpha^3 \mathbf{A}^7 - \alpha^4 \mathbf{A}^9 + \dots]_{xy} \quad (10)$$

LO算法的预测效果显著优于当前一些利用全局信息的代表性算法，包括SPM<sup>[15]</sup>和Katz<sup>[55]</sup>等。令人惊奇的是LO的退化指标——直接计算节点3阶路



径数目的 $A^3$ 指标也优于 CN 指标。

同时期, 在专门针对蛋白质相互作用网络 (protein-protein interaction, PPI) 进行关系预测的研究中, 文献 [58] 从蛋白质相互作用网络的结构和演化证据出发, 指出相互作用的两个蛋白质大多数情况下需要互补的界面表征 (interface)。共享大量相同蛋白质的两个蛋白质界面表征大概率相似或者相同, 而非互补。他们进一步通过实证分析表明传统的基于共同邻居的相似性指标与连边概率呈负相关性。为了迎合 PPI 网络的结构特性, 他们启发式地设计了一个名为 L3 的 3 阶路径指标, 其数学形式如下:

$$S_{xy}^{L3} = \sum_{u,v} \frac{a_{xu}a_{uv}a_{vy}}{\sqrt{k_u k_v}} \quad (11)$$

该指标显著提升了 PPI 网络中潜在在相互作用关系的预测准确度。

文献 [57] 是针对一般性网络, 通过抽象理论分析自动得到衍生指标 $A^3$ ; 文献 [58] 则从一个具体的网络出发, 基于对网络本身的连边结构特征的深刻理解启发式地提出 L3 指标。两种完全不同的方法论, 在几乎相同的时间获得几乎一致的结果! 文献 [52] 迅速关注了该项研究, 并敏锐地认为有必要提出一个一般性的理论框架将该重要发现推广到已有的局部相似性指标和  $n$  阶路径上。基于几何平均数思想, 即  $n$  个变量的几何平均为这  $n$  个变量连乘的  $n$  次方根, RA 指标在  $n$  阶路径上的推广可自然而然的表示为:

$$S_{xy}^{RA(n)} = \sum_{z_1, z_2, \dots, z_{n-1} \in L(n)} \frac{1}{\sqrt[n-1]{k_{z_1} k_{z_2} \dots k_{z_{n-1}}}} \quad (12)$$

式中,  $L(n)$  为连接节点  $x$  和节点  $y$  的  $n$  阶路径上的  $n-1$  个中间节点集合。不难发现, 前文提到的 L3 指标即为 RA 指标在 3 阶路径上的推广。同时他们也进一步通过实证表明基于 3 阶路径的局部相似性指标不仅在蛋白质相互作用网络, 而且在国际贸易网络和食物链网络上都优于基于 2 阶路径的局部相似性指标。

如果说上述发现所考虑的网络太少并不足以挑战我们对局部连边模式的传统认知, 接下来两个大规模实验的结果促使我们不得不回看最开始的局部相似性指标的设计原理, 并斟酌它是否绝对适合所有网络。首先文献 [59] 基于 137 个来自 16 个领域的真实网络进行的一次大规模实证<sup>[59]</sup>。该实证研究对比了 CN、AA、RA、Cannistraci-Hebb 指标和它

们在 3 阶路径上的推广指标的预测性能。结果发现, 3 阶路径指标在相当大一部分网络中表现占优, 但并不存在一类指标可以在所有领域绝对占优, 即有些领域是基于 2 阶路径的相似性指标表现更好, 而有些领域是基于 3 阶路径的相似性指标表现更好。如果把不同领域的网络放到一起综合评价, 基于 3 阶路径的相似性指标在 AUC 指标 (即 ROC 曲线下面积) 上优于基于 2 阶路径的相似性指标, 即在超过 50% 的网络中, 基于 3 阶路径的相似性指标预测效果更好。而在 Precision 指标上, 两者表现相当。该实证研究一个意外的发现就是验证了 LCP 理论和赫布律的价值: 在涉及的 4 个系列的算法中, Cannistraci-Hebb 系列指标成为表现最好的指标。文献 [32] 进一步基于 1371 个来自 14 个领域的真实网络进行了又一次大规模实证分析, 结果再一次表明, 在以 AUPR 为评价指标的实验中, 网络所属领域不同, 表现占优的算法类别也不同。

针对基于 2 阶路径的相似性指标和基于 3 阶路径相似性指标预测性能的争议的意义并不在于选出一个表现更好的相似性指标, 而是重塑我们对于节点间局部连接模式的认识, 鼓励并启发我们设计出更有效的局部相似性指标。

### 3 结束语

在链路预测兴起的最初几年, 得益于网络科学关于节点连边模式和网络拓扑结构特点的丰硕成果, 产生了一系列在同一框架下的经典的局部相似性指标。然而想要在该框架下进一步大幅度提升预测性能更多地需要借助拓扑以外的节点或边的属性信息。本文总结了近十年有别于传统框架的新理论和新发现, 其中必将孕育新的预测算法。在以应用为导向追求无限刷新预测效果的同时, 链路预测研究的另一个重要使命是探索网络的形成和演化机制<sup>[13-14, 60]</sup>, 理解网络的组织结构以及节点连边范式。如果能够对网络背后的组织逻辑精确理解, 对应的链路预测效果自然水到渠成。从这个角度而言, 局部相似性指标便是以此目标为驱动发展形成的一个重要分支。该分支对应的算法简洁、易用、时间复杂度和空间消耗低、预测性能也具有竞争力, 且每一个算法的提出都对应着对一般网络或特定网络节点连边模式的一个新的理解, 在丰富了具体网络先验知识的同时, 也可以进一步推广到利用全局信息的概率模型和网络嵌入等模型中, 从而激

发更多的后续研究。因此,即便在机器学习技术蓬勃发展的今天,局部相似性指标在链路预测研究的方法论中仍然有很大的不可替代性,且作为众多研究领域的技术基础,其科研价值和应用价值都不容小觑。

### 参 考 文 献

- [1] LÜ L, ZHOU T. Link prediction in complex networks: A survey[J]. *Physica A*, 2011, 390(6): 1150-1170.
- [2] ZHOU T. Progresses and challenges in link prediction[EB/OL]. (2021-02-23). <https://arxiv.org/abs/2102.11472>.
- [3] 吕琳媛,周涛. 链路预测[M]. 北京: 高等教育出版社, 2013.  
LÜ Lin-yuan, ZHOU Tao. Link prediction[M]. Beijing: Higher Education Press, 2013.
- [4] 吕琳媛. 复杂网络链路预测[J]. 电子科技大学学报, 2010, 39(5): 651-661.  
LÜ Lin-yuan. Link prediction on complex network[J]. Journal of University of Electronic Science and Technology of China, 2010, 39(5): 651-661.
- [5] SARUKKAI R R. Link prediction and path analysis using markov chains[J]. *Computer Networks*, 2000, 33(1-6): 377-386.
- [6] CRAVEN M, DIPASQUO D, FREITAG D, et al. Learning to construct knowledge bases from the World Wide Web[J]. *Artificial Intelligence*, 2000, 118(1-2): 69-113.
- [7] AIELLO L M, BARRAT A, SCHIFANELLA R, et al. Friendship prediction and homophily in social media[J]. *ACM Transactions on the Web*, 2012, 6(2): 9.
- [8] CLAUSET A, MOORE C, NEWMAN M E J. Hierarchical structure and the prediction of missing links in networks[J]. *Nature*, 2008, 453(7191): 98-101.
- [9] 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.
- [10] BARZEL B, BARABÁSI A L. Network link prediction by global silencing of indirect correlations[J]. *Nature Biotechnology*, 2013, 31(8): 720-725.
- [11] DING H, TAKIGAWA I, MAMITSUKA H, et al. Similarity-based machine learning methods for predicting drug-target interactions: A brief review[J]. *Briefings in Bioinformatics*, 2014, 15(5): 734-747.
- [12] LÜ L, MEDO M, YEUNG C H, et al. Recommender systems[J]. *Physics Reports*, 2012, 519(1): 1-49.
- [13] WANG W Q, ZHANG Q M, ZHOU T. Evaluating network models: A likelihood analysis[J]. *Europhysics Letters*, 2012, 98(2): 28004.
- [14] ZHANG Q M, XU X K, ZHU Y X, et al. Measuring multiple evolution mechanisms of complex networks[J]. *Scientific Reports*, 2015, 5(1): 10350.
- [15] LÜ L, PAN L, ZHOU T, et al. Toward link predictability of complex networks[J]. *Proceedings of the National Academy of Sciences, USA*, 2015, 112(8): 2325-2330.
- [16] SUN J, FENG L, XIE J, et al. Revealing the predictability of intrinsic structure in complex networks[J]. *Nature Communications*, 2020, 11(1): 574.
- [17] XIAN X, WU T, QIAO S, et al. NetSRE: Link predictability measuring and regulating[J]. *Knowledge-Based Systems*, 2020, 196: 105800.
- [18] GUIMERÀ R, SALES-PARDO M. Missing and spurious interactions and the reconstruction of complex networks[J]. *Proceedings of the National Academy of Sciences of USA*, 2009, 106(52): 22073-22078.
- [19] KUMAR A, SINGH S S, SINGH K, et al. Link prediction techniques, applications, and performance: A survey[J]. *Physica A*, 2020, 553: 124289.
- [20] ZHOU T, LÜ L, ZHANG Y C. Predicting missing links via local information[J]. *European Physical Journal B*, 2009, 71(4): 623-630.
- [21] NEVILLE J, JENSEN D. Relational dependency networks[J]. *Journal of Machine Learning Research*, 2007, 8(3): 653-692.
- [22] YU K, CHU W, YU S, et al. Stochastic relational models for discriminative link prediction[C]//Proceedings of the 19th International Conference on Neural Information Processing Systems. Berlin: Springer, 2006: 1553-1560.
- [23] WANG C, SATULURI V, PARTHASARATHY S. Local probabilistic models for link prediction[C]//Proceedings of the 7th IEEE International Conference on Data Mining. Washington DC: IEEE Computer Society, 2007: 322-331.
- [24] PAN L, ZHOU T, LÜ L, et al. Predicting missing links and identifying spurious links via likelihood analysis[J]. *Scientific Reports*, 2016, 6(1): 22955.
- [25] GROVER A, LESKOVEC J. Node2vec: Scalable feature learning for networks[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 855-864.
- [26] TANG J, QU M, WANG M, et al. Line: Large-scale information network embedding[C]//Proceedings of the 24th International Conference on World Wide Web. Florence: ACM, 2015: 1067-1077.
- [27] CAO S, LU W, XU Q. Deep neural networks for learning graph representations[C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence. Arizona: AAAI, 2016: 1145-1152.
- [28] WANG D, CUI P, ZHU W. Structural deep network embedding[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 1225-1234.
- [29] PECH R, HAO D, PAN L, et al. Link prediction via matrix completion[J]. *Europhysics Letters*, 2017, 117(3): 38002.
- [30] BENSON A R, ABEBE R, SCHAUB M T, et al. Simplicial closure and higher-order link prediction[J]. *Proceedings of the National Academy of Sciences of USA*, 2018, 115(48): E11221-E11230.
- [31] GHASEMIAN A, HOSSEINMARDI H, GALSTYAN A, et al. Stacking models for nearly optimal link prediction in complex networks[J]. *Proceedings of the National Academy of Sciences of USA*, 2020, 117(38): 23393-23400.

- [32] MUSCOLONI A, MICIELI U, CANNISTRACI C V. Adaptive network automata modelling of complex networks[EB/OL]. (2020-12-31). <https://www.preprints.org/manuscript/202012.0808/v1>.
- [33] MARA A C, LIJFFIJT J, DE BIE T. Benchmarking network embedding models for link prediction: Are we making progress?[C]//Proceedings of the 7th International Conference on Data Science and Advanced Analytics. Setubal: IEEE, 2020: 138-147.
- [34] ADAMIC L A, ADAR E. Friends and neighbors on the web[J]. *Social Networks*, 2003, 25(3): 211-230.
- [35] OU Q, JIN Y D, ZHOU T, ET AL. Power-law strength-degree correlation from resource-allocation dynamics on weighted networks[J]. *Physical Review E*, 2007, 75(2): 021102.
- [36] ALBERT R, JEONG H, BARABÁSI A-L. Diameter of the world wide web[J]. *Nature*, 1999, 401(6749): 130-131.
- [37] 汪小帆. 无标度网络研究纷争: 回顾与评述[J]. 电子科技大学学报, 2020, 49(4): 499-510.  
WANG Xiao-fan. Controversial issues in researches on scale-free networks: An overview with a network perspective[J]. *Journal of University of Electronic Science and Technology of China*, 2020, 49(4): 499-510.
- [38] MCPHERSON M, SMITH-LOVIN L, COOK J M. Birds of a feather: Homophily in social networks[J]. *Annual Review of Sociology*, 2001, 27(1): 415-444.
- [39] OPSAHL T. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients[J]. *Social Networks*, 2013, 35(2): 159-167.
- [40] NEWMAN M E J. Clustering and preferential attachment in growing networks[J]. *Physical Review E*, 2001, 64(2): 025102.
- [41] KOSSINETIS G, WATTS D J. Empirical analysis of an evolving social network[J]. *Science*, 2006, 311(5757): 88-90.
- [42] OPSAHL T, PANZARASA P. Clustering in weighted networks[J]. *Social Networks*, 2009, 31(2): 155-163.
- [43] MARTÍNEZ V, BERZAL F, CUBERO J C. A survey of link prediction in complex networks[J]. *ACM Computing Surveys*, 2016, 49(4): 69.
- [44] 王凯, 刘树新, 于洪涛, 等. 基于共同邻居有效性的复杂网络链路预测算法[J]. 电子科技大学学报, 2019, 48(3): 432-439.  
WANG Kai, LIU Shu-xin, YU Hong-tao, et al. Predicting missing links of complex network via effective common neighbors[J]. *Journal of University of Electronic Science and Technology of China*, 2019, 48(3): 432-439.
- [45] 李治成, 吉立新, 刘树新, 等. 基于拓扑有效连通路径的有向网络链路预测方法[J]. 电子科技大学学报, 2021, 50(1): 127-137.  
LI Zhi-cheng, JI Li-xin, LIU Shu-xin, et al. A method of link prediction in directed network based on effective connectivity path[J]. *Journal of University of Electronic Science and Technology of China*, 2021, 50(1): 127-137.
- [46] SOUNDARAJAN S, HOPCROFT J. Using community information to improve the precision of link prediction methods[C]//Proceedings of the 21st International Conference on World Wide Web. Lyon: ACM, 2012: 607-608.
- [47] FIRE M, TENENBOIM-CHEKINA L, PUZIS R, et al. Computationally efficient link prediction in a variety of social networks[J]. *ACM Transactions on Intelligent Systems and Technology*, 2014, 5(1): 10.
- [48] AHN Y Y, BAGROW J P, LEHMANN S. Link communities reveal multiscale complexity in networks[J]. *Nature*, 2010, 466(7307): 761-764.
- [49] 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(1): 1613.
- [50] DONG Y, KE Q, WANG B, et al. Link prediction based on local information[C]//Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining. Kaohsiung: IEEE Computer Society, 2011: 382-386.
- [51] HEBB D O. The organization of behavior: A neuropsychological theory[M]. London: Psychology Press, 2005.
- [52] MUSCOLONI A, ABDELHAMID I, CANNISTRACI C V. Local-community network automata modelling based on length-three-paths for prediction of complex network structures in protein interactomes, food webs and more[EB/OL]. (2018-06-18). <https://doi.org/10.1101/346916>.
- [53] CHRISTAKIS N A, FOWLER J H. The spread of obesity in a large social network over 32 years[J]. *New England Journal of Medicine*, 2007, 357(4): 370-379.
- [54] CHRISTAKIS N A, FOWLER J H. The collective dynamics of smoking in a large social network[J]. *New England Journal of Medicine*, 2008, 358(21): 2249-2258.
- [55] KATZ L. A new status index derived from sociometric analysis[J]. *Psychometrika*, 1953, 18(1): 39-43.
- [56] LÜ L, JIN C H, ZHOU T. Similarity index based on local paths for link prediction of complex networks[J]. *Physical Review E*, 2009, 80(4): 046122.
- [57] PECH R, HAO D, LEE Y L, et al. Link prediction via linear optimization[J]. *Physica A*, 2019, 528: 121319.
- [58] KOVÁCS I A, LUCK K, SPIROHN K, et al. Network-based prediction of protein interactions[J]. *Nature Communications*, 2019, 10(1): 1240.
- [59] ZHOU T, LEE Y L, WANG G. Experimental analyses on 2-hop-based and 3-hop-based link prediction algorithms[J]. *Physica A*, 2021, 564: 125532.
- [60] 王文强, 张千明. 链路预测的网络演化模型评价方法[J]. 电子科技大学学报, 2011, 40(2): 174-179.  
WANG Wen-qiang, ZHANG Qian-ming. New method of assessing network evolving models based on link prediction[J]. *Journal of University of Electronic Science and Technology of China*, 2011, 40(2): 174-179.