



# 链路预测的若干基础问题探讨

毕祎琳, 焦鑫善, 万书言, 周涛\*

(电子科技大学 计算机科学与工程学院, 成都 611731)

**摘要** 链路预测是网络科学最具活力的分支之一, 其目标是基于已知的网络拓扑结构估计未观察到的链接的存在可能性。该文对链路预测中仍需重点关注的 4 个基础性问题——网络选取、链路抽样、模型训练和算法评价进行了研究, 报告了这 4 个方面目前的研究进展, 并指出尚未解决的关键问题。最后, 对亟待解决的一些关键研究问题进行了总结。

**关键词** 链路预测; 网络选取; 链路抽样; 模型训练; 算法评价

**中图分类号** TP181 **文献标志码** A **DOI** 10.12178/1001-0548.2024205

## On Fundamentals of Link Prediction

BI Yilin, JIAO Xinshan, WAN Shuyan, and ZHOU Tao\*

(School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

**Abstract** Link prediction is one of the most productive branches in network science, aiming to estimate the likelihoods of unobserved links based on known network topology. This paper critically examines four fundamental issues in link prediction, say network selection, link sampling, model training and algorithm evaluation. It reviews the current research progresses and highlights some significant yet unresolved issues that urgently require scientific answers.

**Key words** link prediction; network selecting; link sampling; model training; algorithm evaluating

现实世界中的很多复杂系统, 都是由大量同质或者异质的个体组成, 而这些系统所表现出来的功能与特性, 包括一些涌现出来的复杂行为和现象, 很大程度上都根植于上述个体之间的相互作用。网络是用来表示这类由大量个体与个体间相互作用所构成的复杂系统的最优数学工具, 这些网络包括但不限于社交网络<sup>[1]</sup>、生物网络<sup>[2]</sup>、信息网络<sup>[3]</sup>、交通网络<sup>[4]</sup>、技术网络<sup>[5]</sup>、经济网络<sup>[6]</sup>等。网络科学作为新兴交叉学科, 专注于研究网络的结构、演化和功能<sup>[7-10]</sup>, 其中链路预测是一个基础性的任务, 旨在基于已知的拓扑结构信息预测缺失的连边以及未来可能出现的连边<sup>[11-13]</sup>。因为链路预测问题同时具备数学上的简洁性、理论上的基础性和应用上的广泛性, 并且好的链路预测方法往往与对网络结构和演化的深度理解紧密结合, 所以链路预测的研究受到了越来越多的关注, 目前已经成为网络科学领域最具活力的分支之一<sup>[14]</sup>。

在过去的二十年中, 研究者们提出了成百上千种链路预测算法(一些代表性的研究工作参见文献[15-25]), 极大地拓展了链路预测研究的前沿阵地。一般而言, 在一个新兴领域发展的初期, 大部分的研究人员会投身于前沿性的研究。但是, 待该领域具备一定成熟度后, 批判性的反思研究就变得重要, 因为如果没有坚实的理论基础、科学的方法、达成共识的研究框架和广泛共享的标准数据库, 许多研究结果可能是不可信的, 或者不同研究人员使用截然不同的方法选取网络、处理样本和评价算法, 使得后续的研究者既很难重现以前的结果, 也没有办法开展横向的比较分析。类似的问题已在人类心理-行为研究中显现出来。如开放科学合作组织重复进行了 100 项心理学实验, 发现仅有不到 40% 的实验结果能够重现<sup>[26]</sup>。此外, 他们还复现了 2011—2014 年发表在《美国经济评论》和《经济学季刊》上的 18 项经济学实验室实验, 以

收稿日期: 2024-08-05; 修回日期: 2024-08-28

基金项目: 国家自然科学基金(T2293771, 42361144718)

作者简介: 毕祎琳, 博士生, 主要从事复杂网络、链路预测等方面的研究。

\*通信作者 E-mail: zhutou@ustc.edu

及2010年至2015年间发表在《自然》和《科学》上的21项社会科学实验研究。在前者的情况下,他们在11次重复实验(61%)中报告了与原始研究相同方向上的显著效应,而复现实验的效应量平均为原始效应量的66%<sup>[27]</sup>;在后者的情况下,这两个数字分别是13次(62%)和50%<sup>[28]</sup>。网络科学的研究人员或许会乐观地认为,基于数学、物理学和计算机科学的网络科学,其研究方法是完全定量的,因此不会出现社会科学研究中遇到的问题。但实际上,如果我们在网络选取、链路抽样、模型训练方法、算法评价指标等问题上无法达成基本共识,并在实施过程中随意选择且描述的时候语焉不详,那么在社会科学实验研究中出现的问题同样会在链路预测研究中出现,从而导致学术资源的巨大浪费。

一个合理的链路预测研究框架,其最基本的功能就是在给定任何一个链路预测算法后,能够公平且准确地对算法性能进行评价,使得算法研究者可以通过此评价横向比较不同算法,应用研究者可以基于此评价提前判断该算法在具体领域应用的效果。在给定链路预测算法后,至少需要以下4个步骤来获得对算法的评价:1)选择一批用于评价算法性能的网络,既包括人工生成的网络,也包括真实的网络,但后者往往更具说服力;2)选择合适的数据抽样方法,将网络中的链路分成两个部分,一部分用来训练模型和确定参数,一部分用来评价算法性能;3)确定一种公平的方法来训练模型和学习参数,这种方法要保证用于测试的链路信息是严格未知的;4)选择合适的评价指标对算法表现给出量化的打分。

在全面回顾链路预测已有的相关研究后,同时发现了可忧和可喜的两面:一方面,当前研究尚未就相关问题达成广泛的学术共识,导致一些显著的缺陷频繁出现在近期的学术出版物中,很多研究人员在上述重要问题上随意选择或片面选择,使得其结论可信度较低;另一方面,目前已经有一些网络科学领域的领军科学家注意到这些问题,并且完成了一些虽然初步但具有重大借鉴意义的研究工作。本文所讨论的问题,是链路预测中重要的问题,但还没有形成系统、成熟的研究成果。讨论这些问题的目的是:1)明确识别并指出当前链路预测研究中存在的缺陷与谬论;2)报告近几年围绕相关问题取得的最新结果,包括若干尚未发表的重要探索;3)指出为了夯实链路预测研究的基础,目前

亟须回答的问题和应该开展的研究。本文旨在推动形成一个具有广泛共识的链路预测研究框架,使得后续的研究成果具有更好的可信度、可比性和可复现性。

## 1 网络选取

为了让叙述变得简洁和紧凑,本文聚焦于简单无向图上的链路预测问题,其中连边的方向和权重都不用考虑,也不允许出现节点自环的情况。通常用 $G(V,E)$ 表示一个简单图,其中 $V$ 是节点的集合, $E$ 是连边的集合。在链路预测的研究中,假设因为实验分析、调查问卷和观察记录等手段的局限性,已知的连边 $E$ 只是真实存在的连边的一部分——这在生物网络和社会网络中特别常见<sup>[29-32]</sup>,而链路预测的主要任务就是基于已知的网络拓扑结构,去推断实际存在但未被观察到的连边——称其为缺失边。对于动态演化的网络,如Internet、WWW等技术网络和Facebook、Twitter等在线社交网络,一个自然推广的链路预测任务就是基于当前的网络拓扑结构和历史演化轨迹,预测未来将会出现但现在尚不存在的连边——称其为未来潜在边。

对于任何链路预测方面的研究,如果要说明算法的有效性,一个自然且必要的步骤是在真实网络中测试算法的表现。有少量关于算法的研究是针对特定类型的网络开展的,如社交网络<sup>[33-34]</sup>和生物网络<sup>[35]</sup>,而更多的研究是适用于各领域的网络——至少从研究者的描述来看是这样的。对于后者而言,一个自然的问题就是针对来自不同领域的网络,算法的表现是否存在差异。如会不会某算法A在社交网络上预测的精度显著好于算法B,但针对交通网络,算法A的精度却显著差于算法B。当然,为了能够针对一个领域的网络,比较若干算法的预测性能,需要选择足够多的网络——从数学上讲,应该多到无论去掉哪一个网络,都不会对算法性能的排名产生任何影响。如果不管多少网络都无法达到这个条件,就说明针对该领域的网络,有一些算法的性能是非常接近的,分不出胜负。因此,在还不完全清楚领域对于算法性能的影响时,如果研究者想要说明他们所提出的算法性能优异,一种让人信服的做法是选择多个领域的网络,并且保证在每个领域都选择足够多的网络。

再次回到上面提到的关键问题:针对不同领域的网络,算法的表现有显著差异吗?尽管没有专门针对这个问题的研究,但是从最近一些涉及大量网

网络的链路预测研究中可以发现端倪。我们对比分析了基于二阶路径和基于三阶路径的局部相似性指标在 16 个领域的 128 个真实网络上的链路预测精度<sup>[36-37]</sup>，发现在软件网络和食物链网络中，三阶指标的表现优于二阶指标。与此同时，在合作网络，人类接触网络和动物接触网络中，二阶指标的表现远胜于三阶指标。同样是与蛋白质相关的网络，在蛋白质相互作用网络中，三阶指标的表现远胜于二阶指标，但在以二级结构单元为节点的蛋白质结构网络中，三阶指标却远不如二阶指标。文献 [38] 对比分析了 CHA<sup>[39]</sup>、SPM<sup>[20]</sup>、SBM<sup>[17]</sup>、Stacking<sup>[25]</sup> 等算法在 6 个领域 550 个网络中的表现。从文献 [38] 的实验结果可以看出，针对不同领域的网络，算法的相对表现有显著差异。根据预测精确度<sup>[40]</sup> 的胜率来看，CHA 在生物、信息、技术、经济、社会这 5 类网络中胜率都排名第一，但在交通网络中胜率却是上述 4 种算法中最低的；SPM 在交通网络中胜率第一，在其他 5 类网络中胜率均排第二；SBM 在生物、信息、经济 3 类网络中胜率超过了 Stacking，但在技术、社会和交通 3 类网络中胜率又低于 Stacking。文献 [41] 针对上述 6 个领域 550 个网络，对比分析了基于低阶点轨道度和低阶边轨道度的链路预测器的预测精度，发现对于不同领域的网络，占优的预测器差异巨大。如在社会网络中，与共同邻居等价的边轨道度预测器胜率高达 91%，但在生物、技术和经济网络中，同一个预测器的胜率还没有超过 5%。我们测试了 9 种著名的链路预测算法在来自 6 个领域中的 131 个网络上的表现，如果按照平均 AUC<sup>[42]</sup> 分值给 9 种算法排名，发现两个排名之间的平均关联，如果用肯德尔等级相关系数<sup>[43]</sup> 和斯皮尔曼等级相关系数<sup>[44]</sup> 的方法来计算，其关联值分别为 0.518 5 和 0.606 7。可见，排名之间虽然存在正向关联，但也存在显著的不一致，说明算法在不同领域网络中的表现是不一致的。

基于目前的文献和计算，可以确认算法在不同领域网络上的表现是不一致的——算法 A 在某领域网络中表现好于算法 B，并不意味着算法 A 在另外一个领域网络中的表现也会好于算法 B。由于本文所讨论的链路预测算法都只依赖于网络的拓扑特征，所以原则上只要对网络拓扑特征有充分的理解，就能够很好估计给定算法的表现。算法在不同领域的网络上表现不一致可以被归因为：1) 不同领域的网络拓扑性质不同；2) 拓扑特性会影响链

路预测算法的表现。

对于第一个原因，网络科学家早已达成共识。如大部分真实网络都表现出小世界效应，也就是网络平均距离的增长速度不超过网络节点数的对数<sup>[45]</sup>，但是对于高度依赖于基础设施的陆上交通网络，如地铁网络、公路网络和街道网络等<sup>[46]</sup>，因为直接修建“长程连接”是不经济的，所以网络的平均距离会显著高于同等规模和密度的典型的小世界网络<sup>[47]</sup>。类似地，许多真实网络都展现出无标度特性，即节点的度分布近似符合幂律分布<sup>[48]</sup>，与此同时，也有很多真实网络不具备无标度特性。根据文献 [49] 针对 928 个真实网络的实证研究，不同领域的网络中无标度网络的丰富程度大不相同，如即便按照最宽松的定义，也有 63% 的生物网络不是无标度网络，但有 92% 的技术网络可以归为不同严格程度的无标度网络。如果继续深入一个个大领域中，会进一步发现，不同子领域网络的度分布也大相径庭。如在社交网络大类中，Facebook、Youtube、Flickr 等在线社交网络的度分布是幂律的<sup>[50]</sup>，科学家和演员合作网络的度分布是介于幂律和指数之间的<sup>[51-52]</sup>，而某些特定的熟人关系网络的度分布却是高斯的<sup>[53]</sup>。又如在交通网络中，航空网络的节点度分布是幂律的<sup>[54]</sup>，但铁路网络的节点度分布是指数的<sup>[55]</sup>，而地铁网络<sup>[56]</sup> 和公交网络<sup>[57]</sup> 节点的度分布比指数还要狭窄<sup>[58]</sup>。从目前已知的研究结果来看，来自不同领域的网络，在各个拓扑特征方面都存在显著的不同。如社交网络的度相关性往往大于零，而生物与技术网络的度相关性却几乎都小于零<sup>[59]</sup>。又如三角结构在大部分真实网络中随处可见，然而在蛋白质相互作用网络<sup>[24]</sup> 和人类浪漫关系网络<sup>[60-61]</sup> 中却非常少见。事实上，来自不同领域的网络，在社团结构<sup>[62]</sup>、模体结构<sup>[63-64]</sup>、富人俱乐部效应<sup>[65-66]</sup>、自相似性<sup>[67]</sup> 等方面，也呈现出显著的差别。

对于第二个原因，相关的研究非常少。基于 18 个真实网络，文献 [68] 分析了常见的网络结构特征和若干知名链路预测算法精确度之间的关系，研究发现网络度分布的异质性越强或者度序列的 Gini 系数越大，基于优先连接的指数表现越好，而网络的簇系数越大，基于共同邻居的指数表现越好。

综上，本文可以断定链路预测算法在不同领域网络中的表现是不一致的，因此，在对一个普适性的算法进行评价的时候，有必要选取来自不同领域的网络，并且在每个领域都要选择足够多数量的网

络, 否则评价的结果极可能是有偏差的。我们认为, 要想真正理解算法在不同领域网络中表现不同的原因, 就必须挖掘网络不同结构特征对于算法表现的影响, 从而未来有望通过结构分析快速筛选合适的算法。文献 [68] 的工作开了一个好头, 但是在这方面还需要开展深入的分析 and 全面的实证。

## 2 链路抽样

链路预测是一个特殊的二分类问题<sup>[69]</sup>——已存在的边被视为正例, 而未出现的边则被视为负例。为了评价算法的性能, 通常要从已经观察到的边集  $E$  中选取一部分边作为测试集  $E^T$ ——测试集是用来评价算法的, 因此涉及测试集的所有信息在训练算法模型的时候都是严格不可知的。如果待训练的模型还有参数, 通常要把剩余的边  $E - E^T$  进一步划分成训练集和验证集, 前者用于特征提取, 后者用于确定模型的参数。为了保持数据分布的一致性, 从  $E - E^T$  中抽取验证集的方法应该与从  $E$  中抽取测试集的方法保持一致, 因此关键的问题是如何从  $E$  中抽取测试集。这个过程叫作链路抽样, 以便于和一个紧密相关的图抽样过程<sup>[70-71]</sup>有所区分, 后者往往关注如何抽取一部分节点, 其首要目的是判断抽样后的较小网络和原始网络结构及功能的差异, 而不是为了预测缺失链路。

应该怎么进行链路抽样呢? 如果网络中每条边出现的时间是已知的<sup>[72]</sup>, 一种自然的方法就是把最后出现的一批连边作为测试集, 因为用过去的信息预测未来的信息是一个基本的原则。如果没有时间信息, 所面对的问题就是如何预测缺失边。即, 假设在那些不存在的边中 ( $E$  的补集), 有一些是实际存在但没有被观察到的边, 而链路预测的任务就是要把这些边找出来。假设未被观察到的缺失边的集合是  $E^M$ , 那么, 如同从  $E - E^T$  中抽取验证集的方法应该与从  $E$  中抽取测试集的方法保持一致, 从  $E$  中抽取测试集的方法也应该与从  $E \cup E^M$  中抽取  $E^M$  的方法保持一致, 这样训练得到的算法模型才能够很好地基于  $E$  预测  $E^M$ 。如果能够知道为什么有一些实际存在的连边在真实网络中无法被观察到 (也就是  $E^M$  出现的原因, 称其为缺失模式), 那么就可以通过模拟这种缺失模式而从  $E$  中抽样出  $E^T$ 。遗憾的是, 这种模式通常是未知的。

对于“找到真实网络的缺失模式以便更好设计链路抽样方法”这个问题, 研究人员要么没有意识到这个问题的重要性, 要么没有找到解决这个问题

的切入口, 因此目前已知的几乎所有不涉及时间信息的链路预测研究, 都是采用完全随机的方式从  $E$  中抽样出  $E^T$ , 但事实上边的缺失模式受到网络本身特性或数据收集过程的复杂影响, 展现出明显的非均匀性和非随机性<sup>[73]</sup>。一些零星的研究考虑了连边的缺失模式。文献 [74] 认为低度节点之间的连边可能更易遗漏, 因此采用基于节点度的抽样策略, 并让连接低度节点的连边有更高的概率被抽到测试集中。在此基础上对比研究不同算法的表现, 得到了一个让人惊讶的结果: Leicht-Holme-Newman 相似性指标<sup>[75]</sup>表现最优, 而这个指标在完全随机抽样的前提下表现很差<sup>[74]</sup>。由此可见, 不同的链路抽样方法对结果影响很大, 因此深入开展这方面的研究很有必要。最近特别值得关注的是文献 [76] 对链路缺失模式的系统性研究, 该研究再次指出, 基于均匀随机分布假设的抽样可能会因忽视了网络数据的非随机缺失特性而引入系统性偏差, 进而误导对模型效能的正确评估。文献 [77] 的研究系统地分析了常见的 20 种不同的抽样方法, 发现不同缺失模式下算法的表现不尽相同, 再次验证了缺失模式的重要性。该研究所涉及的抽样方法主要来自图抽样的研究, 并不一定与真实网络的链路缺失模式一致, 但因为覆盖的方法很多, 所以具有较大的参考价值。

要找到恰如其分的链路抽样方法, 一个首要的任务是要搞清楚真实网络的缺失模式到底是什么样子的。对于时序网络, 要进一步观察和分析最后一段时间出现的连边具有什么特性, 这些连边的特征就代表了链路缺失模式。对于本身没有时序信息, 但是逐步被构造出来的网络, 特别是通过大量实验逐渐累积构建的生物网络, 可以认为理论上具有共识的或实验容易验证的连边先出现, 因此相关论文发表顺序也可以作为类似的参考, 从而可以把后发表论文所报道的连边的特征看作链路的缺失模式。当对目标网络的链路缺失模式了解清楚后, 就可以设计对应的链路抽样方法, 或者从常见的抽样方法<sup>[77]</sup>中选择最接近的方法。另外, 如何确定测试样本的数量也是一个需要关心的问题, 因为过大的测试集可能会削弱模型训练 (训练数据过少), 而过小的测试集则可能导致对算法性能评估不准确。本文集中讨论的是无向简单图上的链路预测问题, 如果要考虑更复杂的网络上的链路预测, 如针对多层网络<sup>[78]</sup>或者高阶网络<sup>[79]</sup>, 也需要考虑如何进行链路抽样, 以符合真实目标网络的链路缺失模

式。这个时候进行抽样还需要考虑层间交互或高阶交互，问题变得更具挑战性。

### 3 模型训练

如上节所述，在训练模型参数时，绝不当使用测试集中的信息。如果使用了部分或者全部测试集的信息，然后再用算法在测试集上的表现来评价算法的性能，就相当于提前知道了考试的答案，那么考试的价值就失去了。这样得到的算法可能在测试集上表现优异，但是在真实场景中表现不佳，用机器学习的术语来说，就是泛化能力差——泛化能力是指从已有的学习经验中提取和应用知识到新的、未经处理的情境中的能力。

遗憾的是，在链路预测实际发表的研究工作中，研究人员经常比较预测结果与测试集来获得所谓的最优参数，相当于利用了测试集的信息。这是一个严重的错误，在链路预测研究的早期就存在，并且一直延续至今。这在很大程度上对无参数算法是不公平的，因为这些算法无法从测试集的信息中得到任何附加的好处，与之相对，泛化能力差的算法则“不当得利”，因为它们既使用测试集的信息来拟合参数，又使用测试集来评价算法，所以它们泛化能力差的缺陷被错误地掩盖了。文献 [80] 的研究发现了“选择偏差”的现象，即选择的模型对于特定数据集过于拟合，无法泛化到其他数据集。为了避免这种现象，该研究提出一种修正方法，即在进行交叉验证时，每次模型选择后应在独立的验证集上进行评估，而不是在同一个数据集上评估，以更准确地评估模型的泛化能力。文献 [81] 也讨论了模型选择中的过拟合问题以及由此导致的选择偏差，建议在进行模型选择时应使用一个独立的验证集进行性能评估，并且在整个模型选择过程中不再使用该验证集。

因此，我们迫切需要分析曾经可能因为这种错误被高估或者低估的算法。可以想象，那些对参数非常敏感且本质上容易过拟合的算法性能会急剧下降，而无参算法或具有强大泛化能力的算法的相对性能倾向于提升。2024 年，文献 [82] 对图神经网络的泛化性能进行了研究，展示了泛化误差如何被图噪声等因素的相互作用所塑造，其中也阐述了模型算法的泛化性能受到训练集相对于所有节点比例的影响。事实上，更早一点的时候，文献 [83] 的研究表明存在一种违反直觉的双下降机制，即更多的训练数据和更大的模型反而会损害泛化能力。如

果是先按照一个固定的比例进行链路抽样并得到测试集，再按照一个固定的比例（这两个比例通常相同，但也可以不同）从剩下的链路中选出验证集，那么按照文献 [82] 和文献 [83] 的建议，选择什么样的比例才能提高算法的泛化能力是需要研究的。除了用一次抽样后便固定的测试集外，文献 [84] 使用  $k$  折交叉验证的方法可以提高分类算法的泛化性，并提倡通过考虑数据集的大小和计算资源来选取最优的  $k$  值。类似的结论是否也适用于链路预测，针对链路预测又怎么选择合适的  $k$  值呢？这些问题目前还没有答案。总体而言，本文认为要厘清这些问题，除了重新使用正确的方法确定参数外，还需要进一步研究已知链路预测算法的泛化能力。

### 4 算法评价

最近十几年有数以千计的链路预测算法被提出，因此，如何选择合适的评价指标来科学、公正地评估这些算法的性能，进而正确引导算法设计和优化的方向，成为了链路预测领域亟待解决的核心问题。不同的评价指标通常具备独特的属性，且各自适用于特定的场景。如精确度关注的是预测为正样本的准确性，特别适用于假阳性（预测为正样本但实际为负样本）成本较高时的场景，如医学诊断、金融风险评估等<sup>[85]</sup>；马太关联参数（Matthews Correlation Coefficient, MCC）特别适用于数据类别不平衡或样本数量不均匀的情况，如药物筛选、信用评分等<sup>[86]</sup>；归一化折损累计增益（Normalized Discounted Cumulative Gain, NDCG）考量了预测结果的排序及每个位置的重要性，主要用于推荐系统、搜索引擎等排序任务中<sup>[87]</sup>；ROC（Receiver Operating Characteristic）曲线下的面积（Area Under Curve, AUC）反映的是模型区分正负样本的能力，在数据不平衡的情况下，能较好地评估模型的整体性能<sup>[42]</sup>；Precision-Recall 曲线下的面积（Area Under the Precision-Recall, AUPR）适用于类别不平衡且关注高精度的任务<sup>[88]</sup>，如异常检测等；Precision 曲线下的面积（AUC-Precision）关注模型在不同阈值下召回率和精确度之间的关系，其更加重视模型在少数类上的表现<sup>[89]</sup>。

到目前为止，尚没有选择评价指标的标准：一些研究人员习惯于遵循经典的指标，而另一些则有自己的偏好。如果不同的指标给出的算法性能排序是相同或者几乎相同的，那么如何选择评价指标就不是一个值得关注的问题。然而，近期的实证研究

显示, 不同评价指标呈现出显著的不一致性<sup>[89-90]</sup>。即, 在一个评价指标下 A 算法性能好于 B 算法, 如果换了另一种评价指标, 则很有可能 B 算法的性能会超过 A 算法。在这种情况下, 如何选择合适的指标来评价算法性能, 就成了一个亟待解决的问题。

由于在所有链路预测算法性能的评价指标中, AUC 是最为流行的, 因此在仅存不多的关于评价指标性能方面的研究中, 往往都是讨论 AUC 的缺点和提出改进方案<sup>[91]</sup>。文献 [92] 的研究认为在正负样本严重不均衡时, 使用 AUPR 会更优于 AUC。文献 [93] 提出 AUC 会给许多负样本排序在底部的算法过高得分, 而这种能力对于找到需要的正样本而言, 没有太大意义。文献 [94] 则强调了 AUC 在评估只有少量头部预测才有价值的场景 (常见于搜索和推荐) 时存在局限性。针对 AUC 可能存在的不足, 一些科学家提出了 AUC 的若干改良版本。如文献 [95] 认为 AUC 存在一个重大的缺陷, 就是在衡量不同分类器的性能时使用了不同的错误分类代价矩阵, 基于此, 提出了名为 H-measure 的指标, 它采用了 AUC 指标的设计思路, 又可以保证不同分类器下错误分类代价局部是一致的。文献 [89] 认为靠前的位置重要性要大于靠后的位置, 因此提出了一种基于放缩 ROC 曲线横纵坐标的新指标 AUC-mROC。

在没有明确选择标准情况下, 有些研究人员同时选择多个指标, 以期达到一种更全面的评价。如文献 [96] 使用了 6 个评价指标来评估算法性能, 文献 [38] 从预测的精确度、排序性能、类别不平衡的角度选取了 7 个评价指标来量化预测性能。尽管这种做法可以有效避免依赖单一指标可能带来的偏见或限制, 但占用了更多的计算资源, 并且在评价指标出现不一致时, 并没有办法判断哪一个评价指标更加可信。我们认为衡量评价指标好坏的关键是看一个评价指标能否正确分辨性能相近的算法孰优孰劣。基于此, 我们针对一个链路预测算法性能可控的人工网络模型, 提出了一套刻画评价指标分辨能力的方法<sup>[97-98]</sup>, 这是较少见的定量刻画任意指标特定能力并在不同指标之间对比分析的研究。总体而言, 探索链路预测中评价指标的有效性, 形成具备广泛共识的一套评价算法优劣的标准, 是目前推动链路预测领域坚实发展的关键性问题。相关研究刚刚起步, 值得特别关注。

## 5 结束语

链路预测是网络科学研究中最具活力的分支之一, 在最近的十多年, 有数以千计的算法被提出。与此同时, 在快速发展的过程中, 很多基础性的问题没有得到足够重视, 特别地, 在网络选取、链路抽样、模型训练和算法评价等方面要么没有广泛达成共识的规范和标准, 要么虽然形成了一定的共识 (如完全随机的链路抽样和利用 AUC 指标进行算法评估, 都非常流行), 但这些共识背后的理论是不充分的。

本文回归链路预测问题的研究本质与核心, 挖掘并梳理出在此研究领域中仍需重点关注的 4 个基础性研究方向。总结起来, 我们认为在链路预测领域中, 未来应首先在以下 8 个方面重点开展研究: 1) 网络的领域性如何影响链路预测算法性能, 针对不同领域网络应选择何种算法? 2) 同领域的网络具备什么样的有差异性的拓扑特性? 3) 不同类型网络连边的内在缺失机制是什么, 针对这些机制应采用何种链路抽样方法? 4) 如何度量链路预测算法的泛化能力? 5) 经典链路预测算法的泛化能力如何, 如何有效提高算法的泛化能力? 6) 网络特征与算法特性是如何影响泛化能力的? 7) 如何量化评价指标对算法优劣的分辨能力? 8) 选取哪一个或一组评价指标能更高效且全面地评估链路预测算法的性能?

## 参考文献

- [1] VEGA-REDONDO F. Complex social networks[M]. Cambridge: Cambridge University Press, 2007.
- [2] PAVLOPOULOS G A, SECRIER M, MOSCHOPOULOS C N, et al. Using graph theory to analyze biological networks[J]. *BioData Mining*, 2011, 4: 10.
- [3] SHI C, LI Y T, ZHANG J W, et al. A survey of heterogeneous information network analysis[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(1): 17-37.
- [4] GANIN A A, KITSACK M, MARCHESE D, et al. Resilience and efficiency in transportation networks[J]. *Science Advances*, 2017, 3(12): e1701079.
- [5] ALBERT R, ALBERT I, NAKARADO G L. Structural vulnerability of the North American power grid[J]. *Physical Review E*, 2004, 69(2): 025103.
- [6] SCHWEITZER F, FAGIOLO G, SORNETTE D, et al. Economic networks: The new challenges[J]. *Science*, 2009, 325(5939): 422-425.
- [7] 汪小帆, 李翔, 陈关荣. 网络科学导论[M]. 北京: 高等教育出版社, 2012.  
WANG X F, LI X, CHEN G R. Network science: An

- introduction[M]. Beijing: Higher Education Press, 2012.
- [8] PÓSFAL M, BARABÁSI A L. Network science[M]. Cambridge, Cambridge University Press, 2016.
- [9] NEWMAN M. Networks[M]. 2nd ed. Oxford: Oxford University Press, 2018.
- [10] 周涛, 张子柯, 陈关荣, 等. 复杂网络研究的机遇与挑战[J]. *电子科技大学学报*, 2014, 43(1): 1-5.  
ZHOU T, ZHANG Z K, CHEN G R, et al. The opportunities and challenges of complex networks research[J]. *Journal of University of Electronic Science and Technology of China*, 2014, 43(1): 1-5.
- [11] 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.  
LYU L Y. Link prediction on complex networks[J]. *Journal of University of Electronic Science and Technology of China*, 2010, 39(5): 651-661.
- [12] LYU L Y, ZHOU T. Link prediction in complex networks: A survey[J]. *Physica A: Statistical Mechanics and Its Applications*, 2011, 390(6): 1150-1170.
- [13] 吕琳媛, 周涛. 链路预测[M]. 北京: 高等教育出版社, 2013.  
LYU L Y, ZHOU T. Link prediction[M]. Beijing: Higher Education Press, 2013.
- [14] ZHOU T. Progresses and challenges in link prediction[J]. *iScience*, 2021, 24(11): 103217.
- [15] 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.
- [16] 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.
- [17] GUIMERÀ R, SALES-PARDO M. Missing and spurious interactions and the reconstruction of complex networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106(52): 22073-22078.
- [18] ZHOU T, LYU L Y, ZHANG Y C. Predicting missing links via local information[J]. *The European Physical Journal B*, 2009, 71(4): 623-630.
- [19] CANNISTRACI C V, ALANIS-LOBATO G, RAVASI T. Erratum: From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks[J]. *Scientific Reports*, 2015, 5: 9794.
- [20] LYU L Y, PAN L M, ZHOU T, et al. Toward link predictability of complex networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(8): 2325-2330.
- [21] PAN L M, ZHOU T, LYU L Y, et al. Predicting missing links and identifying spurious links via likelihood analysis[J]. *Scientific Reports*, 2016, 6: 22955.
- [22] PECH R, HAO D, PAN L M, et al. Link prediction via matrix completion[J]. *EPL (Europhysics Letters)*, 2017, 117(3): 38002.
- [23] XU T, ZOU L. Link prediction with simple path-aware graph neural networks[C]//International Conference on Neural Information Processing. Singapore: Springer, 2024: 570-582.
- [24] KOVÁCS I A, LUCK K, SPIROHN K, et al. Network-based prediction of protein interactions[J]. *Nature Communications*, 2019, 10(1): 1240.
- [25] 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 the United States of America*, 2020, 117(38): 23393-23400.
- [26] COLLABORATION O S. Estimating the reproducibility of psychological science[J]. *Science*, 2015, 349(6251): eaac4716.
- [27] CAMERER C F, DREBER A, FORSELL E, et al. Evaluating replicability of laboratory experiments in economics[J]. *Science*, 2016, 351(6280): 1433-1436.
- [28] CAMERER C F, DREBER A, HOLZMEISTER F, et al. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015[J]. *Nature Human Behaviour*, 2018, 2(9): 637-644.
- [29] VON MERING C, KRAUSE R, SNEL B, et al. Comparative assessment of large-scale data sets of protein-protein interactions[J]. *Nature*, 2002, 417(6887): 399-403.
- [30] KOSSINETIS G. Effects of missing data in social networks[J]. *Social Networks*, 2006, 28(3): 247-268.
- [31] STUMPF M P H, THORNE T, DE SILVA E, et al. Estimating the size of the human interactome[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(19): 6959-6964.
- [32] CAMACHO D M, COLLINS K M, POWERS R K, et al. Next-generation machine learning for biological networks[J]. *Cell*, 2018, 173(7): 1581-1592.
- [33] HASAN M A L, ZAKI M J. A survey of link prediction in social networks[M]//AGGARWAL C. *Social Network Data Analytics*. Boston, MA: Springer, 2011: 243-275.
- [34] DAUD N N, HAMID S H A, SAADOON M, et al. Applications of link prediction in social networks: A review[J]. *Journal of Network and Computer Applications*, 2020, 166: 102716.
- [35] SULAIMANY S, KHANSARI M, NEJAD A M. Link prediction potentials for biological networks[J]. *International Journal of Data Mining and Bioinformatics*, 2018, 20(2): 161.
- [36] ZHOU T, LEE Y L, WANG G N. Experimental analyses on 2-hop-based and 3-hop-based link prediction algorithms[J]. *Physica A: Statistical Mechanics and Its Applications*, 2021, 564: 125532.
- [37] 李艳丽, 周涛. 链路预测中的局部相似性指标[J]. *电子科技大学学报*, 2021, 50(3): 422-427.  
LI Y L, ZHOU T. Local similarity indices in link prediction[J]. *Journal of University of Electronic Science and Technology of China*, 2021, 50(3): 422-427.
- [38] MUSCOLONI A, CANNISTRACI C V. "Stealing fire or stacking knowledge" by machine intelligence to model link prediction in complex networks[J]. *iScience*, 2023, 26(1): 105697.
- [39] MUSCOLONI A, MICHELI U, ZHANG Y, et al.

- Adaptive network automata modelling of complex networks[EB/OL]. [2022-05-10]. <https://doi.org/10.20944/preprints202102.0808.v3>.
- [40] 朱郁蓂, 吕琳媛. 推荐系统评价指标综述[J]. *电子科技大学学报*, 2012, 41(2): 163-175.  
ZHU Y X, LYU L Y. Evaluation metrics for recommender systems[J]. *Journal of University of Electronic Science and Technology of China*, 2012, 41(2): 163-175.
- [41] HE Y J, RAN Y, DI Z, et al. Uncovering multi-order popularity and similarity mechanisms in link prediction by graphlet predictors[EB/OL]. [2024-08-18]. <https://arxiv.org/pdf/2408.09406>.
- [42] HANLEY J A, MCNEIL B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve[J]. *Radiology*, 1982, 143(1): 29-36.
- [43] KENDALL M G. A new measure of rank correlation[J]. *Biometrika*, 1938, 30(1/2): 81-93.
- [44] SPEARMAN C. The proof and measurement of association between two things[J]. *International Journal of Epidemiology*, 2010, 39(5): 1137-1150.
- [45] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998, 393: 440-442.
- [46] BARTHÉLEMY M. Spatial networks[J]. *Physics Reports*, 2011, 499(1-3): 1-101.
- [47] TELESFORD Q K, JOYCE K E, HAYASAKA S, et al. The ubiquity of small-world networks[J]. *Brain Connectivity*, 2011, 1(5): 367-375.
- [48] BARABASI A L, ALBERT R. Emergence of scaling in random networks[J]. *Science*, 1999, 286(5439): 509-512.
- [49] BROIDO A D, CLAUSET A. Scale-free networks are rare[J]. *Nature Communications*, 2019, 10(1): 1017.
- [50] MISLOVE A, MARCON M, GUMMADI K P, et al. Measurement and analysis of online social networks [C]//Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement. New York: ACM, 2007.
- [51] NEWMAN M E J. The structure of scientific collaboration networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2001, 98(2): 404-409.
- [52] ZHANG P P, DI Z R, FAN Y, et al. Model and empirical study on some collaboration networks[J]. *Physica A*, 2006, 360(2): 599-616.
- [53] AMARAL L A, SCALA A, BARTHELEMY M, et al. Classes of small-world networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2000, 97(21): 11149-11152.
- [54] GUIMERÀ R, MOSSA S, TURTSCHI A, et al. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(22): 7794-7799.
- [55] SEN P, DASGUPTA S, CHATTERJEE A, et al. Small-world properties of the Indian railway network[J]. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 2003, 67(3): 036106.
- [56] LATORA V, MARCHIORI M. Is the Boston subway a small-world network?[J]. *Physica A: Statistical Mechanics and Its Applications*, 2002, 314(1-4): 109-113.
- [57] SIENKIEWICZ J, HOLYST J A. Statistical analysis of 22 public transport networks in Poland[J]. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 2005, 72(4): 046127.
- [58] XIE Y B, ZHOU T, BAI W J, et al. Geographical networks evolving with an optimal policy[J]. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 2007, 75(3): 036106.
- [59] NEWMAN M E J, PARK J. Why social networks are different from other types of networks[J]. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 2003, 68(3): 036122.
- [60] HOLME P, LILJEROS F, EDLING C R, et al. Network bipartivity[J]. *Physical Review E*, 2003, 68(5): 056107.
- [61] HOLME P, EDLING C R, LILJEROS F. Structure and time evolution of an Internet dating community[J]. *Social Networks*, 2004, 26(2): 155-174.
- [62] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(12): 7821-7826.
- [63] MILO R, SHEN-ORR S, ITZKOVITZ S, et al. Network motifs: Simple building blocks of complex networks[J]. *Science*, 2002, 298(5594): 824-827.
- [64] MILO R, ITZKOVITZ S, KASHTAN N, et al. Superfamilies of evolved and designed networks[J]. *Science*, 2004, 303(5663): 1538-1542.
- [65] ZHOU S, MONDRAGON R J. The rich-club phenomenon in the Internet topology[J]. *IEEE Communications Letters*, 2004, 8(3): 180-182.
- [66] COLIZZA V, FLAMMINI A, SERRANO M A, et al. Detecting rich-club ordering in complex networks[J]. *Nature Physics*, 2006, 2: 110-115.
- [67] SONG C M, HAVLIN S, MAKSE H A. Self-similarity of complex networks[J]. *Nature*, 2005, 433(7024): 392-395.
- [68] YANG J X, ZHANG X D. Revealing how network structure affects accuracy of link prediction[J]. *The European Physical Journal B*, 2017, 90(8): 157.
- [69] HASTIE T, TIBSHIRANI R, FRIEDMAN J. *The elements of statistical learning: data mining, inference, and prediction*[M]. 2nd ed. New York: Springer, 2009.
- [70] LESKOVEC J, FALOUTSOS C. Sampling from large graphs[C]//Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2006.
- [71] AHMED N K, NEVILLE J, KOMPPELLA R. Network sampling[J]. *ACM Transactions on Knowledge Discovery from Data*, 2014, 8(2): 1-56.
- [72] HOLME P, SARAMÄKI J. Temporal networks[J]. *Physics Reports*, 2012, 519(3): 97-125.
- [73] AMARAL L A. A truer measure of our ignorance[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(19): 6795-6796.
- [74] ZHU Y X, LYU L Y, ZHANG Q M, et al. Uncovering missing links with cold ends[J]. *Physica A: Statistical Mechanics and Its Applications*, 2012, 391(22): 5769-

- 5778.
- [75] LEICHT E A, HOLME P, NEWMAN M E. Vertex similarity in networks[J]. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 2006, 73(2): 026120.
- [76] HE X, GHASEMIAN A, LEE E, et al. Link prediction accuracy on real-world networks under non-uniform missing-edge patterns[J]. *PLoS One*, 2024, 19(7): e0306883.
- [77] ROZEMBERCZKI B, KISS O, SARKAR R. Little ball of fur: A python library for graph sampling[C]//Proceedings of the Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM, 2020: 3133-3140.
- [78] KIVELA M, ARENAS A, BARTHELEMY M, et al. Multilayer networks[J]. *Journal of Complex Networks*, 2014, 2(3): 203-271.
- [79] BICK C, GROSS E, HARRINGTON H A, et al. What are higher-order networks? [J]. *SIAM Review*, 2023, 65(3): 686-731.
- [80] VARMA S, SIMON R. Bias in error estimation when using cross-validation for model selection[J]. *BMC Bioinformatics*, 2006, 7: 91.
- [81] CAWLEY G C, TALBOT N L C. On over-fitting in model selection and subsequent selection bias in performance evaluation[J]. *Journal of Machine Learning Research*, 2010, 11: 2079-2107.
- [82] SHI C, PAN L M, HU H, et al. Homophily modulates double descent generalization in graph convolution networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2024, 121(8): e2309504121.
- [83] NAKKIRAN P, KAPLUN G, BANSAL Y, et al. Deep double descent: Where bigger models and more data hurt[J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2021, 2021(12): 124003.
- [84] KOHAVI R. A study of cross-validation and bootstrap for accuracy estimation and model selection[J]. *IJCAI International Joint Conference on Artificial Intelligence*, 1995, 2: 1137-1143.
- [85] BUCKLAND M, GEY F. The relationship between recall and precision[J]. *Journal of the American Society for Information Science*, 1994, 45(1): 12-19.
- [86] MATTHEWS B W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme[J]. *Biochim Biophys Acta*, 1975, 405(2): 442-451.
- [87] JÄRVELIN K, KEKÄLÄINEN J. Cumulated gain-based evaluation of IR techniques[J]. *ACM Transactions on Information Systems*, 2002, 20(4): 422-446.
- [88] DAVIS J, GOADRICH M. The relationship between Precision-Recall and ROC curves[C]//Proceedings of the 23rd international conference on Machine learning. New York: ACM, 2006: 233-240.
- [89] MUSCOLONI A, CANNISTRACI C V. Early retrieval problem and link prediction evaluation via the area under the magnified ROC[EB/OL]. [2024-05-10]. <https://doi.org/10.20944/preprints202209.0277.v2>.
- [90] BI Y, JIAO X, LEE Y L, et al. Inconsistency of evaluation metrics in link prediction[EB/OL]. [2024-02-24]. <https://arxiv.org/pdf/2402.08893>.
- [91] LOBO J M, JIMÉNEZ-VALVERDE A, REAL R. AUC: A misleading measure of the performance of predictive distribution models[J]. *Global Ecology and Biogeography*, 2008, 17(2): 145-151.
- [92] YANG Y, LICHTENWALTER R N, CHAWLA N V. Evaluating link prediction methods[J]. *Knowledge and Information Systems*, 2015, 45(3): 751-782.
- [93] LICHTENWALTER R, CHAWLA N V. Link prediction: Fair and effective evaluation[C]//Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. New York: IEEE, 2012: 376-383.
- [94] SAITO T, REHMSMEIER M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets[J]. *PLoS One*, 2015, 10(3): e0118432.
- [95] HAND D J. Measuring classifier performance: Acoherent alternative to the area under the ROC curve[J]. *Machine Learning*, 2009, 77(1): 103-123.
- [96] KOU H Z, LIU H W, DUAN Y C, et al. Building trust/distrust relationships on signed social service network through privacy-aware link prediction process[J]. *Applied Soft Computing*, 2021, 100: 106942.
- [97] ZHOU T. Discriminating abilities of threshold-free evaluation metrics in link prediction[J]. *Physica A: Statistical Mechanics and Its Applications*, 2023, 615: 128529.
- [98] JIAO X, WAN S, LIU Q, et al. Comparing discriminating abilities of evaluation metrics in link prediction[J]. *Journal of Physics: Complexity*, 2024, 5(2): 025014.