文章编号：1007-9831（2023）04-0019-08

基于无标度性和资源分配的网络链路预测算法

刘凯[1]，李晨璞[2]，邹龙[1]，王海龙[1]，王浩森[2]

（河北建筑工程学院 1. 信息工程学院，2. 数理系，河北 张家口 075000）

摘要：链路预测是指根据网络结构中的已知信息来推测网络产生新链接的可能性. 资源分配算法（RA）是一种简单高效的基于网络局部信息的链路预测算法，但 RA 算法中并未考虑被预测两节点自身的信息. 在 RA 算法基础上考虑被预测节点间的链接偏好（PA），提出基于网络结构的资源分配 RSF 和 RSW 算法. 在 14 个实际网络中进行链路预测实验，验证了所给算法的有效性和鲁棒性，并分析了网络结构的无标度性对算法预测性能的影响.

关键词：复杂网络；链路预测；节点度；无标度性

中图分类号：TP393　　**文献标识码**：A　　**doi**：10.3969/j.issn.1007-9831.2023.04.005

Link prediction algorithm of network based on the scale-free and the resource allocation

LIU Kai[1]，LI Chenpu[2]，ZOU Long[1]，WANG Hailong[1]，WANG Haosen[2]

（1. School of Information Engineering，2. Department of Mathematics and Physics，Hebei University of Architecture，Zhangjiakou 075000，China）

Abstract：Link prediction refers to inferring the possibility of a network generating new links based on known information in the network structure. Resource allocation algorithm（RA） is a simple and efficient link prediction algorithm based on network local information，however，the information of the two nodes that to be predicted is not be considered in the RA algorithm. Based on the RA algorithm，the link preference（PA）between the predicted nodes is considered，and resource allocation RSF and RSW algorithms are proposed based on network structure. Link prediction experiments were conducted on 14 actual networks to verify the effectiveness and robustness of the algorithm，and the impact of scale-free network structure on the algorithm′s prediction performance was analyzed.

Key words：complex network；link prediction；node degree；scale-free

近年来，复杂网络理论及其应用受到各领域学者的广泛关注，复杂网络研究是涉及到数学、物理学、信息学和社会学等多种学科的交叉学科研究[1]. 尽管复杂网络的理论还需要进一步完善，但复杂网络及相关结构模型广泛存在于自然界和人类社会中[2]，复杂网络的应用研究也已遍布多个领域，如生命科学网络、社会网络和技术网络等[3].

链路预测是复杂网络中的一个重要研究方向，链路预测的目的是在网络结构链接状况未知的情形下分析产生连边的可能性[4-6]. 链路预测的相关理论已应用在许多领域，如应用于线上社交网络的推荐系统、蛋白质网络中未知相互作用的预测、社会网络的分析[7-9]. 近年来涌现的链路预测算法可大致分为 2 类，即以节点相似性为基础的方法和以图嵌入模型为代表的机器学习方法. 节点相似性的相关算法按研究内容可分

收稿日期：2022-09-15

基金项目：河北省自然科学基金项目（D2022404003）；河北省高等学校科学研究计划拔尖项目（BJ2021054）；河北建筑工程学院功能材料与结构力学团队项目（TD202011）；河北省高等教育教学改革研究与实践项目（2018GJJG328）

作者简介：刘凯（1998-），男，湖北宜昌人，在读硕士研究生，从事复杂网络链路预测研究. E-mail：1901576600@qq.com

通信作者：李晨璞（1979-），男，河北康保人，副教授，博士，从事复杂网络理论及应用研究. E-mail：lichenpu2005@126.com

为局部信息、路径以及随机游走 3 个方向[10-13]，其中局部信息方法复杂度低且有很好的预测准确率[14].

局部信息的大多数算法基于共同邻居和待预测节点对的拓扑信息展开，最近几年，研究人员提出了大量高效的局部相似性算法. 其中常见研究方法是在共同邻居研究框架的基础上，考虑节点间的拓扑路径和节点的聚集系数等因素对预测性能的影响[15]. 除此之外，CAR 指标考虑了局部社团范式结构[16]，这对后来的研究有较大的启发；基于共同邻居或二阶路径存在的局限性，有学者对同一方法的二阶指标和三阶指标的预测性能差异开展了研究[17].

资源分配算法（RA）形式简单，预测效率也很好，但是并没有考虑相连两节点的特征对链路预测的影响. 本文对共同邻居框架下的资源分配算法做了进一步分析，在分析网络拓扑结构的基础上将 RA 算法与偏好连接算法（PA）相结合，建立二者相融合的无标度影响下资源分配链路预测算法.

1 链路预测方法及评价指标

1.1 问题描述

对任意的无向网络，可用边集 E 和节点集 V 组成的二元组表示网络结构，网络中的一条连边对应一个由两端节点组成的节点对. 为区分不同连边出现的概率，用一种链路预测的算法赋数值给网络中的未知节点对 $\left(v_x,\ v_y\right)$ ，该数值越大表明节点对之间越容易产生链路或连边[18].

1.2 数据划分

为了验证链路预测方法的有效性和改进算法的性能，从边集 E 中划分出测试集 E^{p}，剩余部分则为训练集 E^{t}. 划分采取随机抽样的方式，并且在划分的过程中，需要保证在测试集 E^{p} 从原有边集划出后，不影响原有边集的连通性. 训练集 E^{t} 用于计算预测指标的分数值，测试集 E^{p} 用于判断该方法的有效性.

1.3 相似性指标

在基于局部信息的指标中，可针对局部网络结构不同方面的特征来选择指标所利用的信息. 大部分研究在基于共同邻居特征的基础上展开，除此之外，被预测连边两端节点的度、聚类系数和节点中心性等局部信息对相似性的影响也常被视为影响指标预测性能的重要因素[19-20].

1.3.1 CN 指标 2 个节点之间共同邻居数目越多，相应节点对之间的相似性越大. CN 指标[21]的相似性 S 的计算公式为

$$S_{xy}^{\mathrm{CN}}=\left|\Gamma\left(x\right)\cap\Gamma\left(y\right)\right| \tag{1}$$

式中：$x,\ y$ 为需要预测的一组节点对；$\Gamma\left(x\right)$ 为 x 的邻居节点集；$\Gamma\left(y\right)$ 为 y 的邻居节点集.

1.3.2 Salton 指标 在共同邻居的基础上，考虑待测节点对自身度的影响，针对不同领域的研究，产生了多种相似性指标，其中一种具有代表性的指标是 Salton 指标[22]，又称余弦相似性. 它将两端节点的度 k_x，k_y 的几何平均作为分母，惩罚了待测节点对两端节点中的大度节点，即

$$S_{xy}^{\mathrm{Salton}}=\frac{\left|\Gamma\left(x\right)\cap\Gamma\left(y\right)\right|}{\sqrt{k_xk_y}} \tag{2}$$

1.3.3 RA 指标 基于度小的共同邻居节点对相似性的影响大于度大的共同节点这一假设，以及网络资源分配的特点，文献[23]提出了资源分配指标（resource allocation，RA）. 用 k_z 表示共同邻居节点 z 的度，则 RA 指标定义为

$$S_{xy}^{\mathrm{RA}}=\sum_{z\in\Gamma\left(x\right)\cap\Gamma\left(y\right)}\frac{1}{k_z} \tag{3}$$

1.3.4 PA 指标 不同于基于邻居节点相似性的指标，PA 指标基于偏好连接（preferential attachment，PA）的相似性仅考虑待预测节点对自身度的大小[24]. 偏好连接认为节点的度越大越容易产生连边，这与无标度网络增长的规则相一致. 在无标度网络模型中，新链接的节点对互相链接的概率正比于这 2 个节点度的乘积，因此 PA 指标定义为

$$S_{xy}^{\mathrm{PA}}=k_xk_y \tag{4}$$

1.4 评价指标

采用 AUC 作为评价指标[25]. 链路预测算法为每组节点对赋予一个预测分数值，其中包括不存在边的节

点对分数值，AUC 表示在随机选取的情况下，测试集边的预测分数值大于不存在边的预测分数值的概率，其定义为

$$AUC = \frac{n' + 0.5n''}{n} \tag{5}$$

式中：n' 表示在抽样中测试集中边分数值大于不存在边分数值的次数；n'' 表示两者分数值相等的次数；n 为抽样的次数.

理论上，抽样的次数 n 越大得到的 AUC 值越准确，本文实验中 n 的取值为 10 000，表示每次实验计算 AUC 时抽样了 10 000 次.

2　资源分配与偏好连接相结合的链路预测方法

2.1　无标度性对相似性的影响

在基于相似性的链路预测算法中，资源分配 RA 算法是具有代表性的算法，在链路预测中也有比较高的精度，在链路预测的一些相关应用中也表现不俗. 两节点间的偏好连接相似性 PA 指标在一部分类型的网络中，尤其是在网络结构中有显著的无标度特性的网络模型中，预测效果与其他指标相比表现突出. 在链路预测中，PA 指标和 RA 指标 2 种方法的侧重点不同. RA 指标仅考虑共同邻居集合对相似性的影响，没有考虑待预测连边两端的 2 个节点本身的影响. 而 PA 指标的衡量方法启发于无标度网络结构的产生过程，与资源分配相反，仅考虑了节点自身度对相似性的影响. 考虑这 2 种链路预测方法的特点，理论上能够提出一个同时考虑共同邻居集合和节点本身度的链路预测算法.

在无标度网络结构中，PA 指标的预测精度较高，无标度网络具有很大的异质性，网络中占总数很小比例的节点拥有网络的大部分连边，而大多数节点只有很少量的连边[26]. 而从数值上看，无标度特性对 PA 指标预测结果的影响是，大多数小度节点在相似性指标计算后有很大的概率会得到相同的数值. 如 Grid，INTRouter，Pages-food 网络中存在大量小度节点，在 INTRouter 中小度节点占比甚至接近 90%，如果用偏好连接 PA 指标计算节点对的相似性，这些网络中会有大量节点对的相似性相同（见表 1）. 因此，在 PA 指标上引入与共同邻居相关的特征值（如 RA 指标），能在预测过程中更充分地应用网络信息的同时，增加节点对相似性的区分度.

表 1　Grid，INTRouter，Pages-food 网络中的小度节点数目

网络数据	N	n_1	n_2	n_3	p（%）
Grid	4 941	1 226	1 656	1 060	79.78
INTRouter	5 022	3 259	986	214	88.79
Pages-food	620	121	93	74	46.45

注：N 为网络节点的总数；n_k 为节点度为 k（k=1，2，3）的数目；p 为网络中节点度小于等于 3 的数目占总节点数的比例.

针对某网络中 4 个节点对（见图 1），计算它们 RA 指标和 PA 指标的相似性，结果见表 2. 网络中节点对（1，4）和节点对（5，4）的 PA 指标相似性都为 2，而它们的 RA 指标相似性分别为 0.333，0.25.

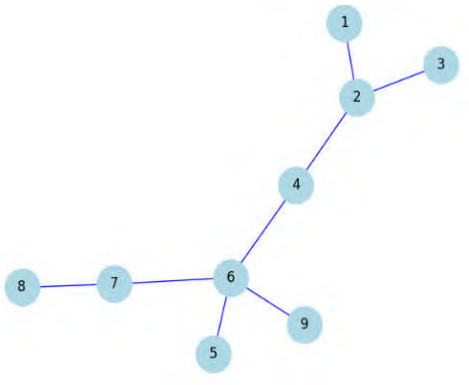

图 1　示例网络

同样，网络的无标度性对 RA 指标的预测结果也有影响. 在无标度性显著的网络中有大量的小度节点，当小度节点之间有共同邻居时，在 RA 指标中它们之间的相似性可以用共同邻居度的倒数之和来衡量. 而在无标度网络中大量的小度节点之间没有共同邻居，RA 指标就会认为他们之间没有相似性，将相似性值置为零. 假设把节点间的一条边看作一步的距离，当节点对之间不存在共同邻居时，节点对之间可能会出现 3 步（相隔 2 个邻居节点）或更多步距离可到达的情况. 在这种情形下，节点对处在相同的局部网络结构下，节点间的距离也较近，因此节点对之间存在连边的可能，即存在相似性. RA 指标将这些节点对相似性置为零，缺乏对这些没有共同邻居的节点对的考虑，而 PA 指标考虑了节点对自身度，它认为任意两个节点之间都存在连边的可能性，可以弥补 RA 指标的不足之处. 示例网络中节点对（1，6）和节点对（2，5）的 RA 指标相似性都为 0，而它们的 PA 指标相似性分别为 4，3（见表 2 ）.

表 2 示例网络中节点对经过 RA 和 PA 指标运算后的相似性值

节点对	（1，4）	（5，4）	（1，6）	（2，5）
S^{RA}	0.33	0.25	0	0
S^{PA}	2	2	4	3

2.2 RSF 指标

由分析得出，RA 指标与 PA 指标相结合可以更好地利用包括节点对自身度在内的局部信息，减少相似性相同的情况出现，使计算后的节点对相似性更有区分度. 因为 RA 指标的预测性能优异，在大部分类型的网络中表现良好，所以本文在研究 RA 指标与 PA 指标的结合方法时，以 RA 指标的相似性为主，用 PA 指标的方法增加 RA 指标相似性的区分度. 然而，一般情况下 RA 指标相似性与 PA 指标相似性相差几个数量级（见表 3），虽然 RA 指标与 PA 指标的相似性平均值只相差 2~3 个数量级左右，但是 RA 指标与 PA 指标相似性的最大值与最小值之间能相差 4~5 个数量级，故 2 种指标不能简单地相加起来.

表 3 某次数据集划分下 RA 和 PA 方法链路预测相似性数值统计

网络数据	$S^{\mathrm{RA}}_{\mathrm{avg}}$	$S^{\mathrm{RA}}_{\mathrm{max}}$	$S^{\mathrm{RA}}_{\mathrm{min}}$	$S^{\mathrm{PA}}_{\mathrm{avg}}$	$S^{\mathrm{PA}}_{\mathrm{max}}$	$S^{\mathrm{PA}}_{\mathrm{min}}$
Grid	0.450	10.533	0.052 6	5.76	256	1
INTRouter	0.135	75.756	0.010 0	5.006	10 000	1
Pages-food	0.109	30.127	0.008 2	34.245 8	14 884	1

为了便于运算处理，本文把理论上节点度的最大值（网络的节点总数减 1，即 $N-1\approx N$ ）作为实际度最大值代入，将偏好连接 PA 指标做归一化处理. 因为 PA 指标与无标度网络密切相关，故本文将 RA 指标与 PA 指标相结合后的指标称为无标度影响下的资源分配（Resource allocation under the influence of scale-free）指标，简称 RSF 指标，其具体公式为

$$S_{xy}^{\mathrm{RSF}}=\sum_{z\in\Gamma(x)\cap\Gamma(y)}\frac{1}{k_z}+\frac{k_xk_y}{N^2} \tag{6}$$

式中：k_x，k_y 分别为预测节点对自身的度；k_z 为节点对共同邻居的度；N 为网络中的节点数.

2.3 RSW 指标

无标度网络模型是复杂网络中网络拓扑结构模型的一种，当网络结构不是无标度网络时，或是网络无标度性较弱时，PA 指标不再适用这些网络的链路预测. 假设给定一个网络，网络中新添加的连边不再根据大度优先法则而是根据小度优先法则，最终会形成度分布较均匀的弱无标度性网络. 例如：对于 BA 模型，在网络初始条件和最终条件相同的情况下，小度优先规则形成的网络与大度优先相比，网络的度分布更均匀，度异质性更小[27]. 因此，本文将节点对的度乘积的倒数与 RA 指标相结合，提出弱无标度性影响下的资源分配（Resource allocation under the influence of weak scale-free）指标，简称 RSW 指标. 节点对的度乘积的倒数一般小于 RA 指标，可直接用于对 RA 指标改进，与 RSF 指标不同，此处不作归一化处理，其具体公式为

$$S_{xy}^{\mathrm{RSW}}=\sum_{z\in\Gamma(x)\cap\Gamma(y)}\frac{1}{k_z}+\frac{1}{k_xk_y} \tag{7}$$

3　实验结果与分析

3.1　网络数据集

本文在网络可视化网站（https://networkrepository.com）和 Stanford 数据集网站（http://snap.stanford.edu/data/）上从不同领域选取了 14 网络数据，包括：Usair，美国航空网络，由机场和飞机航运线路组成的网络；NS–L，学术合作网络 NS 中的最大连通集团；Grid，美国西部电力网络；Router，Internet 路由器层次网络；PB，经过网络无向处理后的美国政治博客网络；CE–LC，秀丽隐杆线虫的部分基因功能关联网络；Erdos992，J.Grossman 等收集的引文网络；FB–PF，关于食物 facebook 页面社交网络；Dol，由宽吻海豚构成的社交网络；Football，2000 年某赛季，IA 分区的橄榄球比赛网络；Polbook，2004 美国总统大选前后政治书籍网络；ENZYMES_g103，酶的化学信息网络；Jazz，爵士乐表演者合奏关系的社交网络；Road–MN，美国明尼苏达州的公路网络.

各数据集的拓扑特征见表 4.

表 4　14 个网络的网络拓扑特征统计

网络	$\|V\|$	$\|E\|$	$\langle k\rangle$	r	$\langle C\rangle$	H	D
Usair	332	2 126	12.807	–0.208	0.625	3.464	0.039
NS–L	379	914	4.82	–0.082	0.741	1.663	0.013
Grid	4 941	6 594	2.669	0.003	0.107	1.450	0.000 54
Router	5 022	6 258	2.49	–0.138	0.012	5.503	0.000 49
PB	1 222	16 714	27.355	–0.221	0.320	2.971	0.022
CE–LC	1 387	1 648	2.376	–0.17	0.076	5.309	0.001 7
Erdos992	5 094	7 515	2.951	–0.444	0.082	5.533	0.000 58
FB–PF	620	2 091	6.745	–0.032	0.331	2.952	0.011
Dol	62	159	5.129	–0.04	0.259	1.327	0.084
Football	115	613	10.661	0.162	0.403	1.007	0.094
Polbook	105	441	8.40	–0.128	0.488	1.421	0.081
ENZYMESg103	59	115	3.898	–0.174	0.235	1.084	0.067
Jazz	198	2 742	27.607	0.020	0.618	1.395	0.140
Road–MN	2 642	3 303	2.50	–0.185	0.016	1.090	0.000 95

注：$|V|$，$|E|$ 分别为网络的节点数和边数；$\langle k\rangle$，r，$\langle C\rangle$ 分别为网络平均度、同配系数和网络平均聚类系数；$H=\langle k^2\rangle/\langle k\rangle^2$，$D=2|E|/\left(|V|(|V|-1)\right)$ 分别为网络度异质性和网络密度.

3.2　实验结果分析

实验中，从原始网络随机划分出 10%作为测试集，剩余部分作为训练集，每个网络作 100 次独立重复实验. RSF 和 RSW 两相似性计算指标与 CN，Salton，RA，PA 等 4 个指标在 14 个不同的网络中进行链路预测的 *AUC* 结果见表 5.

表 5　各指标对不同网络进行链路预测的 *AUC*

网络	指标					
	CN	Salton	RA	PA	RSF	RSW
Usair	0.953	0.925	0.971	0.910	0.972	0.828
NS–L	0.977	0.976	0.981	0.657	0.981	0.916
Grid	0.626	0.623	0.626	0.579	0.641	0.536
Router	0.653	0.654	0.654	0.955	0.959	0.082
PB	0.924	0.878	0.928	0.910	0.935	0.792
CE–LC	0.687	0.683	0.690	0.868	0.914	0.251
Erdos992	0.749	0.738	0.740	0.964	0.973	0.109
FB–PF	0.907	0.895	0.911	0.837	0.933	0.659
Dol	0.803	0.796	0.804	0.678	0.810	0.712
Football	0.850	0.860	0.849	0.267	0.818	0.882
Polbook	0.889	0.882	0.899	0.670	0.893	0.899
ENZYMESg103	0.696	0.713	0.700	0.285	0.582	0.828
Jazz	0.957	0.966	0.972	0.772	0.970	0.971
Road–MN	0.522	0.521	0.522	0.200	0.220	0.820

由表 5 可以看出，与其他指标相比，RSF 指标在 9 个网络中都取得了 *AUC* 的最大值，RSW 指标在 4

个网络中取得了 *AUC* 的最大值. RSF 指标在大部分网络的预测效果都比结合前的单一的 PA 指标或 RA 指标的预测效果更好，RSF 指标改善了 PA 指标和 RA 指标在部分网络表现不佳的缺陷. RSF 指标综合了 RA 指标和 PA 指标在不同网络上的优异预测性能，并且在此基础上，RSF 指标在部分网络中超越了 RA 指标和 PA 指标的预测性能. 与 RA 指标相比，RSF 指标在 PB，Usair，Dol 3 个网络中的 *AUC* 分别提升了 0.7，0.1 和 0.6 个百分点，在 Grid，FB-PF 网络中分别提升了 1.5 和 2.2 个百分点. 与 PA 指标相比，RSF 指标在 Router，Erdos992 网络中的 *AUC* 分别提升了 0.4 和 0.9 个百分点，在 CE-LC 网络中提升了 4.6 个百分点.

与 RA 指标相比，RSW 指标在 Football，ENZYMESg103，Road-MN 这 3 个网络中的 *AUC* 分别提升了 3.3，12.8，29.8 个百分点. 尽管与 RSF 指标相比，RSW 指标适用的网络类型不多，但是 RSW 指标在 Football，ENZYMESg103，Road-MN 这 3 个网络中表现非常突出，预测效果明显优于其他预测指标.

将预测结果与表 4 中的度异质性指标 H 对比发现，当 H 值特别低时，也就是度分布较均匀时，RSW 指标性能较好，而当 H 值较大时，RSF 指标的预测性能较好. 度异质性 H 越大，网络越接近无标度网络模型. 实验结果表明，基于 RA 指标、PA 指标和网络无标度性提出的 RSF 指标，它的预测性能优异. 与 RA 指标、PA 指标或者是同样结合了共同邻居与两端节点度的 Salton 指标相比，RSF 指标在更多的网络中预测性能表现突出，这表明 RSF 指标所适用的网络类型更广，预测不同类型的网络时，RSF 指标有更强的健壮性. 除此之外，RSW 指标在少部分网络中表现突出.

3.3 网络无标度性分析

RSF 指标和 RSW 指标都是在无标度网络模型的分析上提出，为进一步分析网络无标度性对预测性能的影响，本文统计了实验所选取的 14 个网络中的度分布情况，结果见图 2，其中 n 表示度为 k 的节点数量.

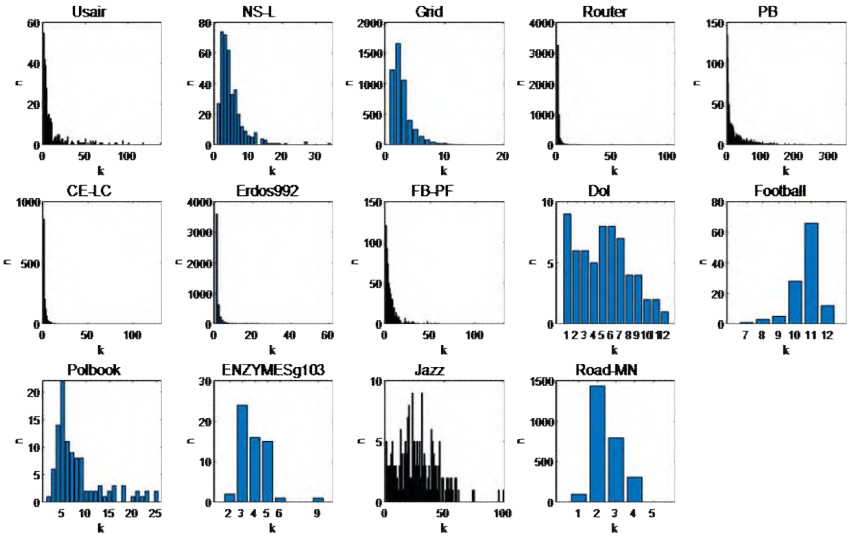

图 2 14 种网络的度分布

由柱状图 2 可以看出，在前 8 个网络中随着度 k 的增加对应的节点数 n 急剧减少，符合无标度网络中度服从幂律分布的特点，说明前 8 个网络符合或近似符合无标度网络模型. 由表 5 可以看出，RSF 指标在前 8 个网络中的预测性能优于其他预测指标，这进一步验证了 RSF 指标适用于无标度网络的链路预测. 后 6 个网络不适合用无标度网络模型来描述，这些网络的网络结构介于随机网络和无标度网络之间. 另外，由图 2 可以看出，Football，ENZYMESg103，Road-MN 这 3 个网络的度分布接近于泊松分布，它们的网络结构应更接近于随机网络而不是无标度网络. 由表 5 也可以看出，RSW 指标对 Football，ENZYMESg103，Road-MN 网络的预测效果与其他指标相比表现突出，这一结果验证了 RSW 指标适用于度分布较均匀的网络.

3.4 鲁棒性分析

为了进一步对 RSF 指标和 RSW 指标的性能进行分析，本文将在不同比例的测试集划分下，比较 RSF 指标、RSW 指标与其他指标在不同网络中的鲁棒性. 测试集占整个网络的比例（Probe Size，也称为划分比例）一般设定为从 0.05 增长到 0.5，每次增长 0.05. 部分网络会因为划出的测试集占比太多，导致剩余

网络的连通性改变时无法继续划分出测试集，因此这些网络中有划分比例的最大值 $p_{\max}$，此时划分比例仅增长到与 $p_{\max}$ 最接近的值. 在每个网络的不同划分比例下，*AUC* 取 100 次独立重复实验的平均值.

在不同的测试集划分比例条件下，网络预测的 *AUC* 变化情况见图 3. 由图 3 可以看出，在 Grid 这一网络中，随着测试集比例的增加（即训练集比例的减少），RSF 指标的 *AUC* 不降反升，而且一直优于其他指标的 *AUC*；在 PB，CE-LC，FB-PF，Dol 这 4 个网络中，随着测试集比例的增加，RSF 指标的 *AUC* 下降缓慢,而且一直优于其他指标的 *AUC*. 在 Usair,NS-L 网络中,RSF 指标的 *AUC* 接近于 RA 指标的 *AUC*，同时 RSF 指标的鲁棒性与 RA 指标保持一致. 在 Router，Erdos992 网络中，RSF 指标的 *AUC* 接近于 PA 指标的 *AUC*，同时 RSF 指标的鲁棒性与 PA 指标保持一致. 在 Football，ENZYMESg103，Road-MN 网络中，RSW 指标的 *AUC* 始终高于其他指标的 *AUC*，并且下降趋势缓慢、鲁棒性高. 在 Polbook 网络中，当划分比例增长至 0.35 之前，RSW 指标的 *AUC* 与 RA 指标的接近，而高于 RSF 指标的 *AUC*，当划分比例到达 0.35 后，RSW 指标的性能下降. 在 Jazz 网络中，RSW 与 RA 和 RSF 2 个指标的 *AUC* 保持一致.

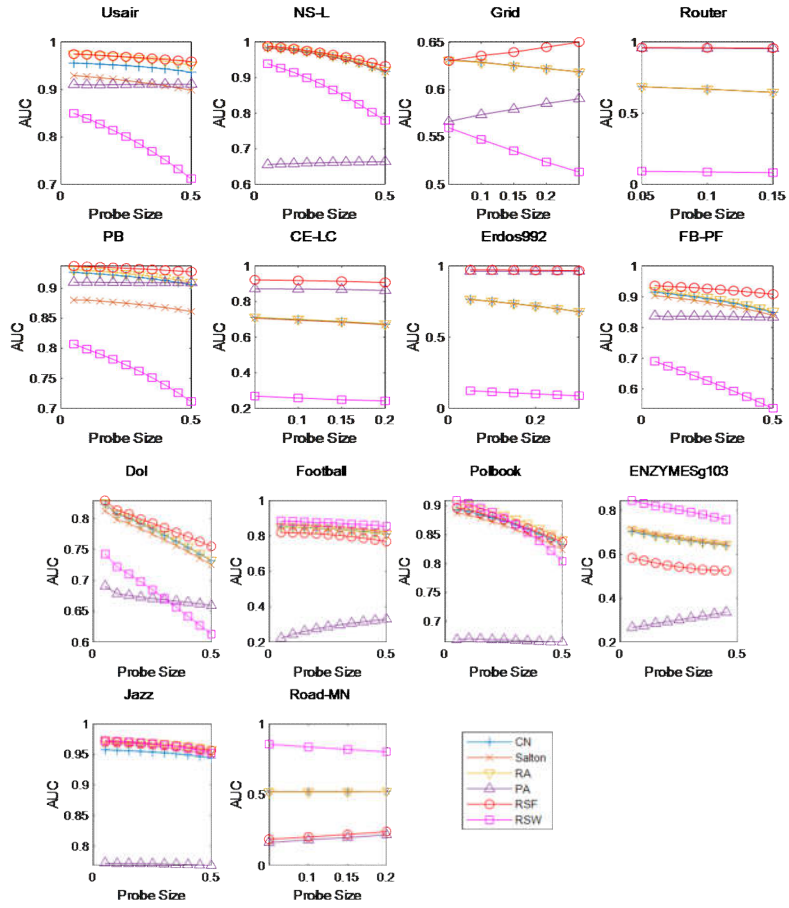

图 3　不同比例测试集划分下网络链路预测的 *AUC* 变化情况

对预测指标的鲁棒性分析表明，在不同的测试集划分比例下，RSF 指标在大部分网络中有良好的鲁棒性，并且其 *AUC* 高于其他指标，进一步验证了 RSF 指标的性能优于 RA 指标和 PA 指标. 而 RSW 指标在度分布相对较均匀的网络中，预测性能好于其他指标且有较好的鲁棒性.

4　结论

本文分析了 RA 指标和 PA 指标的特点以及网络结构对 RA 指标和 PA 指标性能的影响，并在此基础上提出了考虑网络无标度性影响的 RSF 指标和 RSW 指标. RSF 指标弥补了 RA 指标和 PA 指标仅考虑单一影响因素的不足，在算法中既考虑了共同邻居对相似性的影响又考虑了待预测连边的两端节点度对相似性的影响. 本文实验中采用 14 个实际网络，选择包括 RSF 和 RSW 在内的 6 种预测指标进行链路预测实验，实验结果表明，在大部分网络中，RSF 指标的性能优于其他预测指标. 而 RSF 指标基于 PA 指标提出，因此

与 PA 指标一样，RSF 指标的预测性能与网络的无标度性有关. 实验分析可知，RSF 指标适用于无标度性强、度异质性高的网络，且 RSF 指标具有良好的鲁棒性. RSW 指标与 RSF 指标相反，适用于弱无标度性、度分布相对较均匀的网络，并且在这类网络中有优于其他指标的预测性能和鲁棒性. 为保证 RSF 指标和 RSW 指标的简洁高效，本文没有在指标公式中增加调节参数来对公式的前后两项进行数值上的拟合. 下一步的研究中，将分析参数调节对 RSF 指标和 RSW 指标预测性能的影响.

参考文献：

[1] 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.

[2] 朱丽，葛爽，张庆红. 复杂网络结构下科技政策的创新驱动：基于网络权力和创新扩散视角[J]. 中国科技论坛，2019（1）： 29-36.

[3] 董靖巍. 基于复杂网络的网络舆情动态演进影响机制研究[D]. 哈尔滨：哈尔滨工业大学，2016.

[4] 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.

[5] 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.

[6] 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.

[7] 吴铭. 基于链接预测的关系推荐系统研究[D]. 北京：北京邮电大学，2012.

[8] 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.

[9] 赵卫绩，张凤斌，刘井莲. 复杂网络社区发现研究进展[J]. 计算机科学，2020，47（2）：10-20.

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

[11] 顾秋阳，吴宝，池仁勇. 基于高阶路径相似度的复杂网络链路预测方法[J]. 通信学报，2021，42（7）：61-69.

[12] 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 Transactions on Knowledge and Data Engineering，2007，19（3）：355-369.

[13] 刘思，刘海， 陈启买，等. 基于网络表示学习与随机游走的链路预测算法[J]. 计算机应用，2017，37（8）：2234-2239.

[14] 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.

[15] 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.

[16] Cannistraci C V，Alanis L 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）：01613.

[17] Zhou T，Lee Y L，Wang G. Experimental analyses on 2-hop-based and 3-hop-based link prediction algorithms[J]. Physica A：Statistical Mechanics and its Applications，2021，564：125532.

[18] Mcpherson M，Smith L L，Cook J M. Birds of a Feather：Homophily in social networks[J]. Annual Review of Sociology，2001，27：415-444.

[19] 郁湧，王莹港，罗正国，等. 基于聚类系数和节点中心性的链路预测算法[J]. 清华大学学报（自然科学版），2022，62（1）：98-104.

[20] 白萌，胡柯，唐翌. Link prediction based on a semi-local similarity index[J]. 中国物理 B，2011，20（12）：128902-128903.

[21] Lorrain F，White H C. Structural equivalence of individuals in social networks[J]. The Journal of Mathematical Sociology，1971，1（1）：49-80.

[22] Salton G，McGill M J. Introduction to Modern Information Retrieval[M]. Auckland：McGraw-Hill，1983.

[23] Zhou T，Lü L，Zhang Y C. Predicting missing links via local information[J]. The European Physical Journal B-Condensed Matter and Cpmplex Systems，2009，71（4）：623-630.

[24] Xie Y B，Zhou T，Wang B H. Scale-free networks without growth[J]. Physica A：Statistical Mechanics and its Applications，2008，387（7）：1683.

[25] Hanley J A，McNeil B J. The meaning and use of the area under a receiver operating characteristic（ROC）curve[J]. Radiology，1982，143：29-36.

[26] 辜芳琴. 无标度网络演化模型的研究[D]. 广州：暨南大学，2014.

[27] Barabasi A L，Albert R. Emergence of scaling in random networks[J]. Science，1999，286（5439）：509-512.