

一种基于节点间资源承载度的链路预测方法

王凯 刘树新* 陈鸿昶 李星

(国家数字交换系统工程技术研究中心 郑州 450002)

摘要: 链路预测旨在发现网络的未知、缺失连接,具有重要的实际应用价值。基于网络结构相似性的链路预测方法具有简单且有效的特点,受到各领域学者的普遍关注。然而,许多现有方法在计算节点间存在连接可能性时,忽视了节点间资源承载能力的影响。鉴于此,该文提出一种基于节点间资源承载度的链路预测方法。该方法首先通过分析节点间资源传输过程,进而对节点间资源承载能力进行量化,提出资源承载度。然后,基于资源承载度对节点间连接可能性的影响进行分析,并提出相应的链路预测方法。9个真实网络的实验结果表明,相比其他链路预测方法,该方法在3个衡量标准下均具有较高的预测精度。

关键词: 复杂网络; 链路预测; 资源承载度; 相似性

中图分类号: N94; TP393

文献标识码: A

文章编号: 1009-5896(2019)05-1225-10

DOI: 10.11999/JEIT180553

A New Link Prediction Method for Complex Networks Based on Resources Carrying Capacity Between Nodes

WANG Kai LIU Shuxin CHEN Hongchang LI Xing

(National Digital Switching System Engineering and Technological R&D Center, Zhengzhou 450002, China)

Abstract: Link prediction aims to discover the unknown or missing links of complex networks, which plays an important role in practical application. The similarity-based link prediction methods attract a lot of attention due to their briefness and effectiveness. However, most of similarity indices ignore the influence of resource carrying capacity between nodes when calculating the likelihood that a link exists between two endpoints. Because of the problem, a new link prediction method based on resources carrying capacity between nodes is proposed. Firstly, the resource carrying capacity is proposed to quantify the capability of resource carrying between nodes. Then, based on the resource carrying capacity, a new link prediction method is proposed by analyzing the impact of node connectivity. The experimental results of nine real networks show that compared with other link prediction methods, the proposed method can achieve higher prediction accuracy under three standard metrics.

Key words: Complex network; Link prediction; Resources carrying capacity; Similarity

1 引言

近年来,网络科学相关领域的进展不断加快,许多不同类型的实际网络包括航空交通运输网^[1]、生物网络^[2]、引文网络^[3]、电力网络^[4]、社交网络^[5]等均已成为复杂网络的重要研究对象。链路预测作为网络科学领域的重要研究方向,旨在发现现有网络中未知连接^[6]、缺失连接^[7]或错误连接^[8],具有重要的实际应用价值。

当前,针对复杂网络链路预测的研究已经取得较大进展,尤其是基于网络结构的链路预测方法^[9]更是受到广泛关注。近年来,基于网络结构相似性提出了大量链路预测方法,其大致分为局部相似性指标和全局相似性指标。局部相似性指标主要基于共同邻居这一基本结构,有共同邻居指标(Common Neighbor, CN)^[10]、资源分配指标(Resource Allocation, RA)^[11]、Adamic-Adar(AA)指标^[12]和CAR指标^[13],其普遍具有较低的时间复杂度。此外,刘树新等人也在资源分配指标RA的基础上,提出了扩展的资源分配指标(ERA)^[14],其在两个不同的衡量标准下均表现较好;与局部相似性指标不同,全局指标则利用了网络的全局拓扑信息,包括考虑所有路径的全局指标Katz^[15]、基于随机游走的

收稿日期: 2018-06-05; 改回日期: 2019-01-16; 网络出版: 2019-01-30

*通信作者: 刘树新 liushuxin11@126.com; liushuxin11@gmail.com

基金项目: 国家自然科学基金(61521003, 61803384)

Foundation Items: The National Natural Science Foundation of China (61521003, 61803384)

ACT^[16]和Cos+^[17],其效果一般好于局部指标,但复杂度较高不适合于大型网络;当然,为了在算法预测性能和复杂度之间找一个折中,周涛等人^[18]在共同邻居指标基础上考虑了3跳路径提出了局部路径指标(Local Path, LP),在付出一定时间复杂度后,取得一定预测效果,但预测精度仍然需要进一步提高。当前,虽然提出了大量的链路预测方法,也取得了较好的效果,但仍然缺少复杂度相对较低但具有较高预测精度的方法。此外,许多现有方法在计算节点间存在连接可能性时,也忽视了节点间资源承载能力的影响。而现实网络中,通信行为随时随地发生,节点间的资源交互是建立连接的重要动力来源^[19],节点间资源承载能力越大,则建立连接可能越大。

基于上述分析,本文在将从资源交互促进网络演化的角度^[20],通过分析节点间多路径资源传输机理,对任意两点间的资源承载能力进行量化,定义了节点间资源承载度。然后,基于资源承载度提出链路预测方法。通过多个实际网络中试验对比,验证了所提方法的合理性和有效性。

2 相关工作介绍

基于相似性的链路预测方法相对较多,现在对相关指标简要介绍如下:

(1)CN^[10]:通过共同邻居的数目衡量节点 x 和 y 的相似度, $\Gamma(x)$ 为节点 x 邻居的集合,表示为

$$s_{xy}^{CN} = |\Gamma(x) \cap \Gamma(y)| \quad (1)$$

(2)AA^[11]:在CN的基础上对其中共同邻居节点加权,权值为节点度对数的倒数,表示为

$$s_{xy}^{AA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\lg k_z} \quad (2)$$

(3)RA^[12]:基于资源传输原理,认为共同邻居传递资源的多少和其节点度成反比,具体表示为

$$s_{xy}^{RA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k_z} \quad (3)$$

(4)CAR^[13]:在CN的基础上,考虑了共同邻居之间的存在连接的情形,表示为

$$s_{xy}^{CAR} = |\Gamma(x) \cap \Gamma(y)| \cdot \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{|\gamma(z)|}{2} \quad (4)$$

其中, $|\gamma(z)|$ 表示共同邻居 z 中相互存在连接关系的集合。

(5)LP^[18](Local Path):在共同邻居(即2阶路径)基础上,考虑了3阶路径,可表示为

$$\mathbf{S} = \mathbf{A}^2 + \alpha \cdot \mathbf{A}^3 \quad (5)$$

其中, \mathbf{S} 为相似度矩阵, \mathbf{A} 是邻接矩阵, $(\mathbf{A}^n)_{xy}$ 表

示节点 x 和 y 之间长度为 n 的路径数目, α 为调节参数。

(6)Katz^[15]:从全局路径角度,充分考虑节点间所有的路径数目量化节点间的相似度,表示为

$$s_{xy}^{Katz} = \sum_{l=1}^{\infty} \alpha^l \cdot |\text{path}_{xy}^l| \\ = \alpha \mathbf{A}_{xy} + \alpha^2 (\mathbf{A}^2)_{xy} + \cdots + \alpha^n (\mathbf{A}^n)_{xy} \quad (6)$$

其中, path_{xy}^l 为节点间路径长度为 l 的数目, α 是路径调节参数。

(7)ACT^[16]:即平均通勤时间,粒子从 x 游走至 y 所需的平均步数,表示为

$$s_{xy}^{ACT} = \frac{1}{l_{xx}^+ + l_{yy}^+ - 2l_{xy}^+} \quad (7)$$

其中, l_{xy}^+ 为对应拉普拉斯矩阵的伪逆 \mathbf{L}^+ 中对应的元素值。

(8)Cos+^[17]:即余弦相似性指标,基于随机游走,在矩阵 \mathbf{L}^+ 的基础上,计算两个向量之间的相似性

$$s_{xy}^{Cos+} = \frac{v_x^T v_y}{|v_x| \cdot |v_y|} = \frac{l_{xy}^+}{\sqrt{l_{xx}^+ \cdot l_{yy}^+}} \quad (8)$$

3 基于节点间资源承载度的链路预测方法

网络内部不同资源传输过程是复杂网络的重要动力学行为,其相关研究也涉及多个领域。Wang等人^[21]便在研究故障传播过程的基础上,提出了邻居节点优先连接的策略提升相互依存网络的鲁棒性。此外,该团队也基于级联故障传播过程提出了一种优先连接机制^[22]。当前,基于资源传输过程的链路预测缺乏对节点间资源承载能力的刻画和利用。本文将从节点间资源承载度的角度分析网络节点间连边的可能性。

3.1 节点间资源承载度分析与量化

复杂网络的节点之间存在大量的传输路径,所有路径均可作为节点间资源传输的承载通道。图1示意了节点 v_i 和 v_j 之间资源传输过程。可以看出,邻居节点尤其是共同邻居为 v_i 和 v_j 之间的资源传输

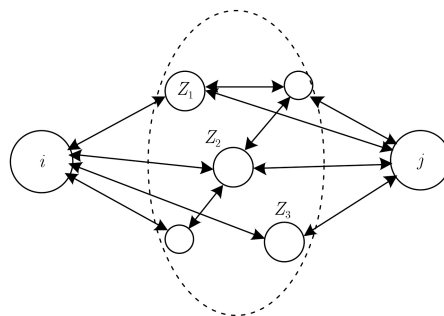


图1 网络中任意两点之间资源承载示意图

提供了大量可能的承载。图2示意了资源从 v_i 经过多跳路径传输到 v_j 的过程，假定 v_i 有资源量为1的资源需要传输。由于 v_1 存在多个连边， v_2 接收的资源量可表示为 $1/k_{v_1}$ 。多跳传输后， v_j 接收的资源量 R_j 表示为

$$R(i \rightarrow j) = \frac{1}{k_{v_1} k_{v_2} \cdots k_{v_n}} \quad (9)$$

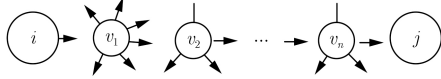


图2 网络节点间资源传输示意图

为了更加定量的分析节点通过多跳路径传输后所接收的资源量，本文将以前述网络平均节点度对上式进行计算分析。以平均度 \hat{k} 为3~10分析，1个单位的资源经过1跳传输后资源接收量 R_j 大约为 $1/3 \sim 1/10$ ，3跳传输后约为 $1/27 \sim 1/1000$ 。在量化节点间资源承载中，若考虑3跳以上路径，则存在复杂度的付出和资源承载量化效果之间的矛盾。因此，本文仅考虑3跳以下的资源承载过程。图3示意了不同拓扑结构下节点间资源承载度对比。对比分析来看，图3(b)比图3(a)多1个共同邻居，前者比后者多1条资源传输路径，故资源承载能力较高；而图3(b)和图3(c)虽然共同邻居数目相同，但后者共同邻居节点度均小于前者，以式(9)的传输模型分析，后者的资源传输承载能力更强；此外，图3(d)和图3(c)的共同邻居数目和节点度均不同，前者拥有更多的传输路径且每条路径资源传输能力更高，故其资源传输承载能力也最高。总结来看，各种网络结构的资源承载度依次为：图3(d) > 图3(c) > 图3(b) > 图3(a)。

在分析了不同网络拓扑结构下，共同邻居的数目和节点度对资源承载度的影响后，仍然需要分析在1对节点间共同邻居节点度不同时，如何量化资源承载能力的问题。图4便显示了不同共同邻居节点度对资源承载度的影响。若每个邻居节点接收相同的资源量输入，则节点度越大，其传输的资源量

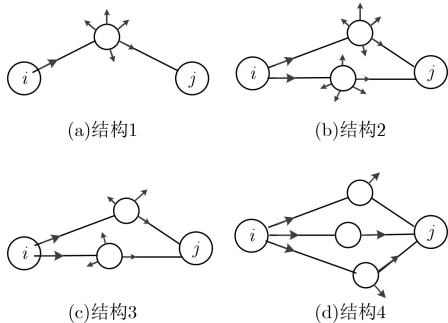


图3 不同拓扑结构下节点间资源承载度对比

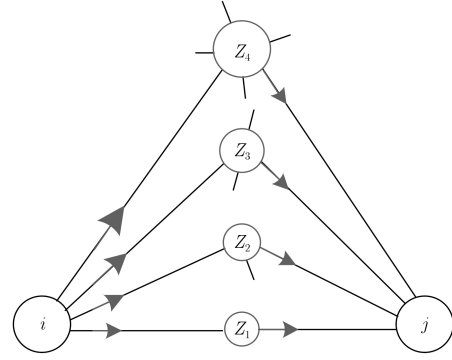


图4 不同节点对节点间资源承载度的影响

则越小。然而，实际网络中，网络资源更多地偏向于大度节点 z_4 。文献[23]在电力网络等相互依存网络鲁棒性研究中，也证实了节点度越大，其承受的电力负载越高；同样，现实网络中普遍存在的“富者愈富”现象[24]也揭示了大度节点在网络资源占有量上的优势。综上分析，从资源传输角度上看，大度节点 z_4 所在路径是重点路径。虽其资源传输量较低，但作为核心路径节点，将以一种传输短板的形式决定着整体的资源承载能力。因此，本文在对节点间资源承载度刻画时，重点考虑了共同邻居的大度节点和数目，具体的定义如下：

定义1 节点间资源承载度：对于一个无向网络 $G(V, E)$ ，其中 V 代表网络中节点的集合， E 代表所有连边的集合。网络中任意存在两个节点 i 和 j ，两个节点间资源承载度量化为

$$Q(i, j) = \varepsilon (|\Gamma(i) \cap \Gamma(j)|) \cdot \left(1 + \frac{|\Gamma(i) \cap \Gamma(j)|}{\text{Max}\{k_{z_1}, k_{z_2}, \dots, k_{z_n}\}} \right), \quad z_x \in \Gamma(i) \cap \Gamma(j) \quad (10)$$

其中， $\text{Max}\{k_{z_1}, k_{z_2}, \dots, k_{z_n}\}$ 为共同邻居中节点度的最大值。

3.2 基于节点间资源承载度的链路预测方法

在对节点间资源承载度进行量化后，便可探讨基于节点间资源承载度的链路预测方法。一般情形下，可通过计算两端节点与共同邻居的资源承载度量相似度。然而，若仅考虑共同邻居，节点周围潜在资源承载能力对于连接可能性的影响难以刻画。为了更好的刻画节点间相似性，对其定义如下：

定义2 基于节点间资源承载度的链路预测方法(QN)：对于一个无向网络 $G(V, E)$ ，其中网络中任意存在两个节点 x 和 y ，从网络中节点间资源承载度的角度，节点 x 和 y 存在连接的可能性表示为

$$s_{xy}^{QN} = \sum_{z \in \Gamma(x) \cup \Gamma(y)} (Q^\lambda(x, z) + Q^\lambda(z, y)) \quad (11)$$

其中，参数 λ 作为调节资源承载度强度，调节不同

网络中其对相似度影响的强度。节点间资源承载度和相似性具有明显的正相关,故 λ 取值范围设定为 $[0, +\infty)$,当 $\lambda = 0$ 时, $s_{xy}^{QN} = 2(k_x + k_y)$,此时其复杂度与CN相近。QN方法的复杂度相对较低,具体在CN和LP之间。

4 衡量指标及数据集介绍

4.1 算法衡量指标

对于一个指定的链路预测算法,给所有未连接的边赋予一个分数值,其分数值越高,则其连接的可能性则越高。具体上,网络连边 E 分为训练集 E^T 和测试集 E^P , $E = E^T \cup E^P$,且 $E^T \cap E^P = \emptyset$ 。

AUC衡量指标可以简单理解为在测试集 E^P 中随机选择一条边的分数值大于未连接边的分数值高的概率^[25]。若测试集中边大于未连接边的分数(n'),则加1分,若两者相等(n''),则加0.5分,表示为

$$AUC = \frac{n' + 0.5n''}{n} \quad (12)$$

其中, n 为比较的次数,随机预测下 $AUC \approx 0.5$,因此AUC超过0.5的多少可以衡量算法精确程度。

Precision衡量指标(以下简称Pre)则为在前 L 个预测边中预测准确的比例^[26],定义为

$$Pre = m/L \quad (13)$$

式(13)中, m 表示前 L 个预测结果排序中出现在测试集 E^P 中的个数。一般情况下,设置 $L = 100$ 。

ROC曲线^[27],主要用于判断检测结果的好坏。ROC曲线以真正例率(TPR)为纵坐标,假正例率(FPR)为横坐标的曲线(TP是真正例, N_+ 是真正例和假负例之和;FP是假正例, N_- 是真正例和真负例之和)

$$TPR = TP/N_+, FPR = FP/N_- \quad (14)$$

4.2 网络数据

为了验证所提算法的有效性,选择了多个实际网络数据进行测试,分别介绍如下:(1)AIDS-Blog^[28]:一个AIDS博客相关的引用关系网络;(2)FWFB^[29]:在Florida海岸湿季的食物链网络;(3)FWEW^[29]:

生活在Everglades Graminoids湿季的捕食关系网络;(4)Caenorhabditis Elegans(CE)^[30]:线虫的神经元网络;(5)Email^[31]:一个小型企业的公司邮件收发网络;(6)PB(PoliticalBlogs)^[32]:美国某政治论坛的博客首页之间的关系网络;(7)Hamster(HS)^[33]:hamsterster.com网页用户间的朋友关系网络;(8)Figeys^[34]:人类蛋白质的相互作用网络;(9)UCsocial^[35]:美国加州大学UC在线讯息交互网络。上述网络具体的特征参数如表1所示,包括节点数目 $|V|$ 、边数 $|E|$ 、平均度 $\langle k \rangle$ 、集聚系数 C 、平均最短路径 $\langle d \rangle$ 和匹配系数 r 。实验中,设置训练集中连边占比为0.9,测试集为0.1,每个实验结果均为20次结果的均值。

5 试验结果及分析

5.1 AUC结果及分析

针对文中所提节点间资源承载度指标(QN),首先需要分析强度参数 λ 对不同网络中预测效果的影响。图5中示意了9个实际网络中AUC结果随着强度参数 λ 的变化情况。可以看出,在多数网络中随着强度参数 λ 的增大,预测结果AUC会逐渐增大,然后趋于平稳,说明在这些网络中资源承载度在节点间是否存在连接上起着较大的权重。对于少部分网络中(AIDS和Figeys),强度参数 λ 接近于0时,其AUC结果较好,随着 λ 的增大其会轻微下降,然后一直保持平稳状态,这说明在一些网络中资源承载度对节点间是否存在连接的影响上权重较小。总体来看,随着强度参数 λ 的变化,AUC可以在不同网络中达到较高的预测精度,而且在很大范围内,预测精度会趋于平稳。鉴于大多数网络中AUC曲线在 λ 值较大时,AUC值会相对平稳,后续对比分析主要对比AUC最大值和 $\lambda = 18.5$ 时AUC值。

表2示意了9个实际网络中QN指标和其他方法的对比结果,其中QN-18.5为 $\lambda = 18.5$ 时的结果,而QN-max则为最大值结果。可以看出,CN作为最简单的相似性指标,其表现一般,而考虑了部分节点信息后,AA,RA和CAR总体预测结果均有一定幅度的提升,尤其是RA指标,在部分网络中已

表1 网络数据特征参数

Network	AIDS	FWFB	FWEW	CE	Email	PB	Hamster	Figeys	UC
$ V $	146	128	69	297	167	1222	1858	2239	1899
$ E $	180	2075	880	2148	5784	16717	12534	6432	13838
C	0.052	0.335	0.552	0.308	0.541	0.361	0.0904	0.04	0.109
$\langle k \rangle$	2.47	32.42	25.51	14.46	69.26	27.36	13.49	5.76	14.57
$\langle d \rangle$	3.42	1.78	1.64	2.46	1.87	2.74	3.39	3.98	3.06
r	-0.725	-0.112	-0.298	-0.225	-0.295	-0.221	-0.085	-0.331	-0.188

经接近甚至超过全局指标(如CE, Email, PB网络)。LP考虑了3跳路径, 在预测结果上明显高于

基于共同邻居的局部相似性方法, 而仅多耗费了一跳的计算复杂度。而考虑全局所有路径的Katz方

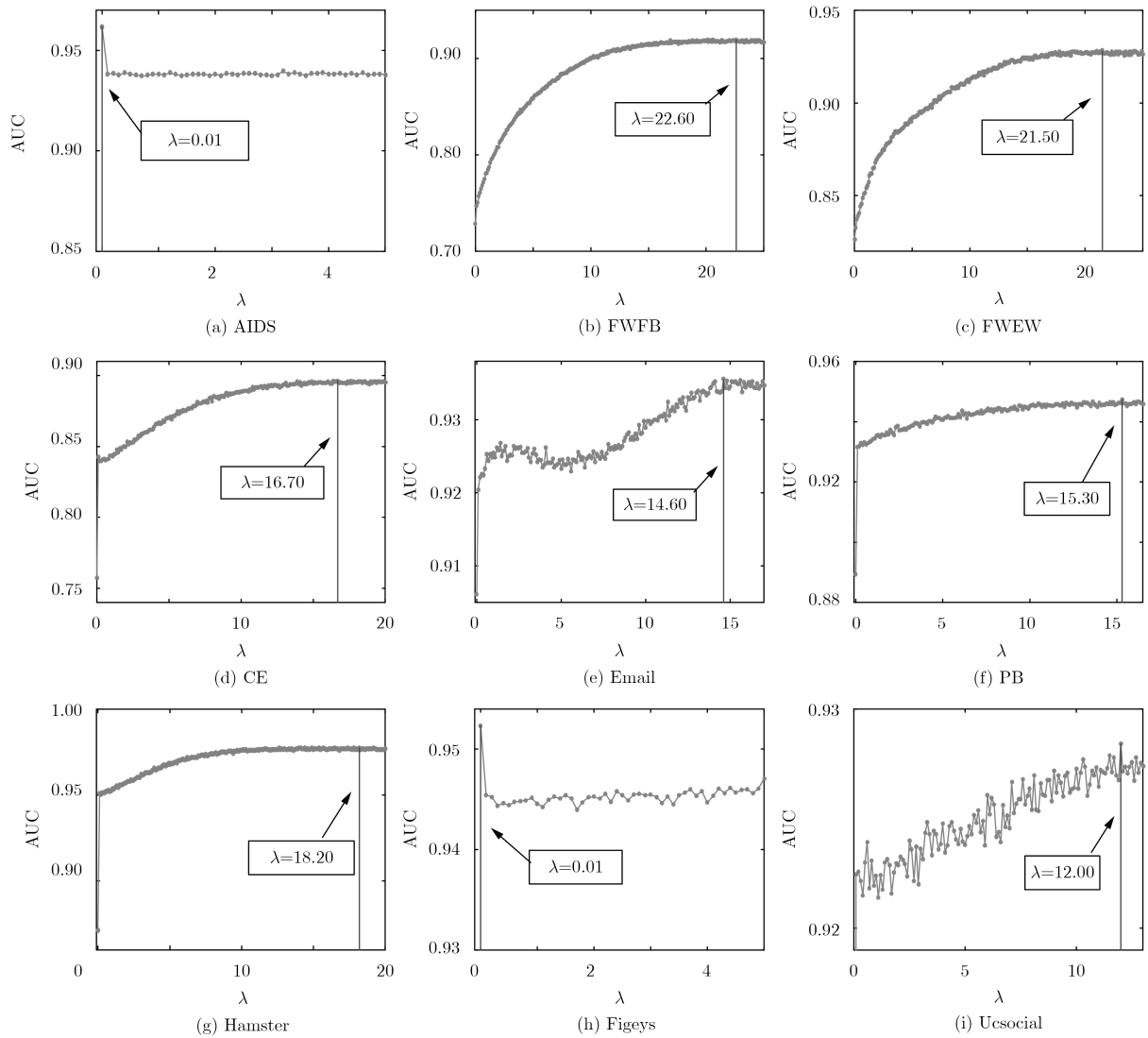


图 5 强度参数对AUC结果影响曲线图

表 2 AUC结果对比

Network	AIDS	FWFB	FWEW	CE	Email	PB	Hamster	Figeys	UC
CN	0.588	0.605	0.686	0.853	0.920	0.923	0.817	0.563	0.782
RA	0.601	0.609	0.701	0.873	0.928	0.927	0.822	0.568	0.786
AA	0.602	0.608	0.695	0.870	0.922	0.926	0.821	0.567	0.786
CAR	0.589	0.620	0.689	0.853	0.919	0.921	0.817	0.564	0.780
LP-0.001	0.831	0.622	0.707	0.871	0.922	0.935	0.936	0.889	0.891
LP-0.01	0.831	0.670	0.730	0.871	0.921	0.937	0.942	0.903	0.902
Katz-0.001	0.847	0.620	0.706	0.870	0.921	0.934	0.935	0.886	0.891
Katz-0.01	0.848	0.675	0.737	0.869	0.919	0.932	0.939	0.900	0.901
ACT	0.951	0.722	0.784	0.755	0.900	0.891	0.871	0.918	0.895
Cos+	0.584	0.650	0.514	0.862	0.906	0.925	0.961	0.843	0.871
QN-18.5	0.936	0.918	0.927	0.896	0.935	0.946	0.977	0.945	0.928
QN-max	0.962	0.919	0.928	0.896	0.936	0.947	0.977	0.952	0.928

法, 虽然效果较好, 多数预测结果和LP非常接近, 但复杂度远高于局部相似性方法。同样, 全局相似性指标ACT和Cos+在考虑了全局网络拓扑结构后, 预测结果也非常显著, 但其复杂度也相对较高。而在考虑了节点间资源承载度后, 文中所提QN链路预测方法在9个实际网络中AUC结果均为最好, 在固定参数强度下其表现也较好。

横向比较来看, 少数网络中如Email和PB, 现有指标的AUC均达到了0.9以上的预测精度。而对于大多数网络Hamster, Figeys, UC, AIDS和CE来说, 局部指标的AUC都较小, 仅在全局指标中的预测精度达到了0.9。其他网络中, 多数现有指标的AUC均在0.6左右。相比而言, 本文所提QN指标能够把9个网络的AUC提升到0.9左右(仅CE中为0.89, 略低于0.9), 一定程度上说明了方法的有效性。从QN的预测结果上, 参数 λ 较大时取得最大

AUC结果, 说明AUC结果更倾向于加大资源承载度的影响权值或强度(即: 若节点间资源承载度较大, 其建立连接的可能性则是指数级增长的)。在实际网络预测中, 强度参数 λ 建议选取一个较大的值, 如18.5左右, 其在不同网络中整体预测结果也非常稳定且较好。

5.2 Precision结果及分析

图6示意了参数 λ 对Pre结果影响曲线图, 与AUC相似, 多数网络中Pre结果随参数 λ 的增大逐渐增大, 然后趋于稳定。部分网络中(AIDS和Figeys), 则会瞬间增大到最大值, 然后缓慢下降至平稳。Email网络中, 曲线变化呈现先下降后上升, 接着再次下降的趋势, 这说明在Email网络中, 资源承载度的强度对Pre结果影响较为复杂。总体来看, 随着参数 λ 的变化, Pre可以在不同网络中达到较高的预测精度, 且在很大范围内, 多数网络中的预

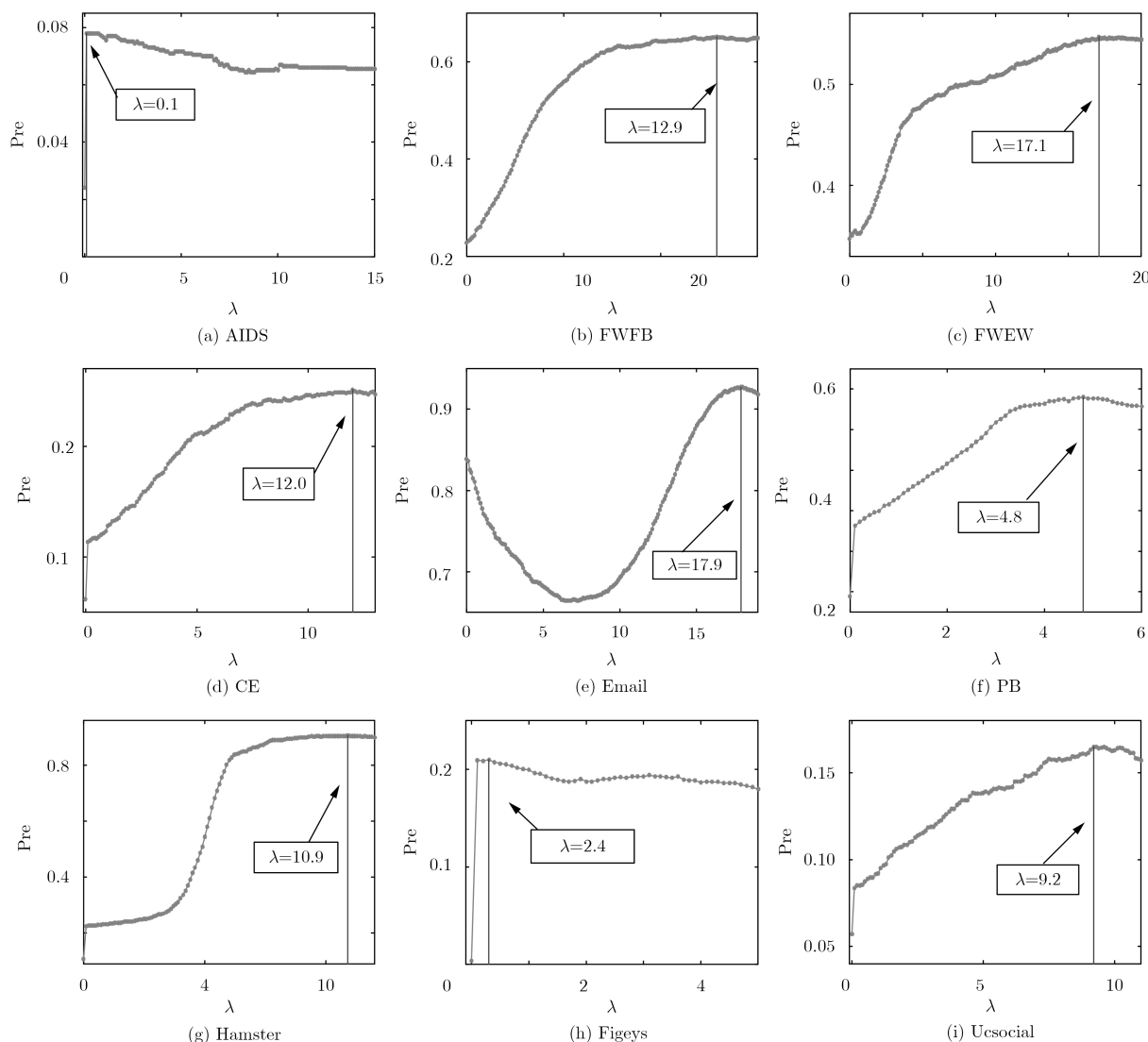


图6 强度参数对Pre结果影响曲线图

测精度会趋于平稳。不同于AUC,许多网络中Pre曲线变化趋势复杂,但 λ 取较小值时,其相对其它指标已经效果较好,故后续对比中选取了最大值和 $\lambda = 2.5$ 。

表3示意了9个实际网络中QN指标和其他方法的对比结果。与AUC结果相似,CN指标的Pre预测精度较低,而AA,RA和CAR在考虑了部分局部

信息后预测效果均有所的提升。尤其是CAR,在部分网络中甚至超过了全局相似性指标(PB, Figeys和UC)。LP和Katz的Pre效果整体好于局部相似性指标,但令人意外的是ACT和Cos+的结果较差,可能基于随机游走的指标更适合AUC。考虑了节点间资源承载度后,QN在9个实际网络中Pre结果均为最高,而在固定参数强度下其表现也较好。

表3 Pre结果对比

Network	AIDS	FWFB	FWEW	CE	Email	PB	Hamster	Figeys	UC
CN	0.014	0.086	0.161	0.131	0.708	0.417	0.015	0.011	0.022
RA	0.026	0.088	0.170	0.129	0.727	0.247	0.007	0.014	0.020
AA	0.026	0.090	0.164	0.138	0.720	0.380	0.010	0.012	0.022
CAR	0.014	0.088	0.150	0.131	0.703	0.478	0.030	0.026	0.052
LP-0.001	0.051	0.094	0.171	0.137	0.709	0.421	0.017	0.011	0.025
LP-0.01	0.051	0.123	0.198	0.136	0.701	0.455	0.052	0.012	0.034
Katz-0.001	0.053	0.093	0.171	0.137	0.709	0.422	0.017	0.011	0.025
Katz-0.01	0.053	0.134	0.202	0.136	0.696	0.454	0.071	0.012	0.037
ACT	0.000	0.000	0.126	0.000	0.000	0.000	0.000	0.000	0.000
Cos+	0.000	0.039	0.000	0.081	0.620	0.333	0.017	0.008	0.011
QN-2.5	0.075	0.397	0.415	0.156	0.734	0.460	0.251	0.192	0.114
QN-max	0.078	0.651	0.547	0.251	0.927	0.580	0.906	0.210	0.165

横向比较来看,与AUC相似,少数网络(如Email和PB)的Pre预测精度相对较高。而其他网络中,现有指标的结果都普遍较低,多数甚至小于0.1。而本文所提QN指标能够较大幅度的提升Pre结果,许多结果能够从0.1提升至0.2~0.4。在 $\lambda = 2.5$ 时,8个网络中QN效果最好,且提升幅度非常可观,仅在PB中略低于CAR。在实际网络预测中,参数 λ 建议取2.5左右,其整体预测结果较为稳定。

5.3 ROC曲线结果及分析

为了更深入的验证文中所提方法QN的预测效果,本文对9个网络中的ROC曲线做了实验对比分析。鉴于现有指标曲线重合较多,故选取了4个典型的指标曲线作对比。图7示意了不同网络中相关指标的ROC曲线变化图,其中QN为最大值曲线(由于参数在18.5和最大值曲线基本重合故只显示最大值ROC),LP和Katz均为参数取0.01时曲线。总体上,在9个实际网络中,QN的表现均好于其他相关指标,即其ROC曲线更接近“左上角”。在多数网络如AIDS,FWFB,FWEW,Hamster,Figeys和Ucsocial中,QN的表现明显好于相比其他指标,能够清楚的看到其ROC曲线的优势。而对于其他网络中,QN的ROC曲线和其他指标非常接

近。让人意外的是,FWFB和FWEW网络中,全局指标Cos+的表现最差,可能该指标并不适合于食物链等相关网络的链路预测。从ROC曲线上可以看出,相比局部和全局指标,QN指标在预测效果上具有一定优势,这也一定程度上证明了所提方法的有效性。

6 结束语

近年来,基于网络结构的链路预测方法受到学者广泛关注,相关方法也相继提出和应用。许多全局指标效果较好但复杂度较高^[36,37]。而局部方法,虽然复杂度较低,但预测结果难以达到理想效果。针对上述问题,本文从节点间资源承载度的角度重新研究链路预测问题,通过分析节点间的资源传输过程,定义资源承载度量节点间的资源承载能力,并基于资源承载度提出一种链路预测方法。多个实际网络试验结果表明,在3个衡量标准AUC,Pre和ROC曲线下,基于资源承载度的链路预测方法能够提高链路预测的精度,印证了资源承载度与节点间存在连接有明显的相关性,也为复杂网络演化的深层次内在动力学机理的相关研究提供了一种思路。此外,本方法时间复杂度较低,可以应用于大型的复杂网络链路预测。

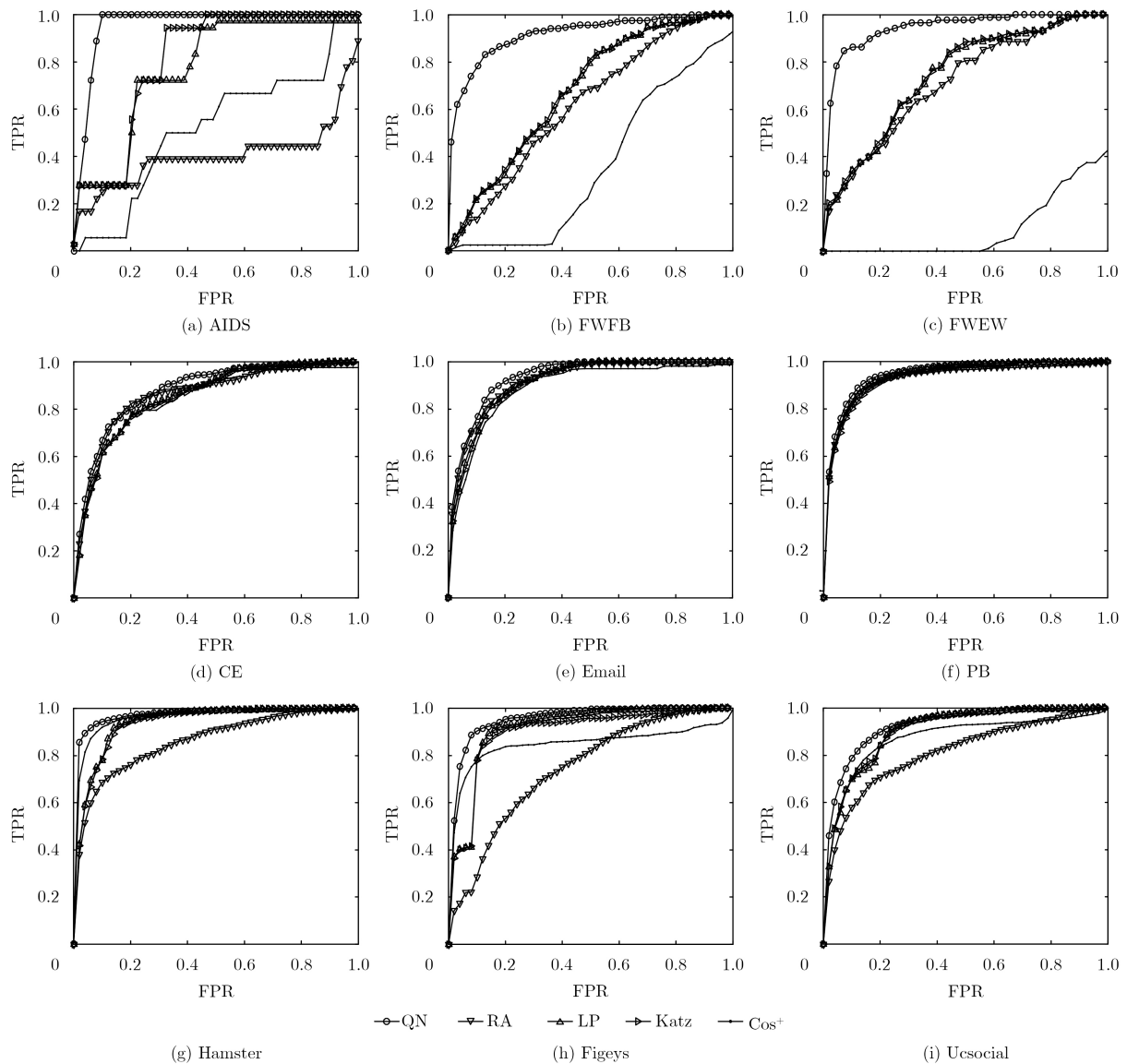


图7 9个网络中ROC曲线对比结果

参考文献

- [1] SHANMUKHAPPA T, HO I W H, and TSE C K. Spatial analysis of bus transport networks using network theory[J]. *Physica A: Statistical Mechanics and Its Applications*, 2018, 502: 295–314. doi: [10.1016/j.physa.2018.02.111](https://doi.org/10.1016/j.physa.2018.02.111).
- [2] 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. doi: [10.1016/j.physa.2017.10.026](https://doi.org/10.1016/j.physa.2017.10.026).
- [3] VINCENOT C E. How new concepts become universal scientific approaches: insights from citation network analysis of agent-based complex systems science[J]. *Proceedings of the Royal Society B: Biological Sciences*, 2018, 285(1874): 20172360. doi: [10.1098/rspb.2017.2360](https://doi.org/10.1098/rspb.2017.2360).
- [4] CHEN Zhenhao, WU Jiajing, XIA Yongxiang, *et al.* Robustness of interdependent power grids and communication networks: A complex network perspective[J]. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2018, 65(1): 115–119. doi: [10.1109/TCSII.2017.2705758](https://doi.org/10.1109/TCSII.2017.2705758).
- [5] KIM J and HASTAK M. Social network analysis: characteristics of online social networks after a disaster[J]. *International Journal of Information Management*, 2018, 38(1): 86–96. doi: [10.1016/j.ijinfomgt.2017.08.003](https://doi.org/10.1016/j.ijinfomgt.2017.08.003).
- [6] VON MERING C, JENSEN L J, SNEL B, *et al.* STRING: known and predicted protein-protein associations, integrated and transferred across organisms[J]. *Nucleic Acids Research*, 2005, 33(1): D433–D437. doi: [10.1093/nar/gki005](https://doi.org/10.1093/nar/gki005).
- [7] SCCELLATO S, NOULAS A, and MASCOLO C. Exploiting place features in link prediction on location-based social networks[C]. *Proceedings of the 17th ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*, San Diego, California, USA, 2011: 1046–1054. doi: [10.1145/2020401.2020488](https://doi.org/10.1145/2020401.2020488).

- 10.1145/2020408.2020575.
- [8] HOLLAND P W, LASKEY K B, and LEINHARDT S. Stochastic blockmodels: first steps[J]. *Social Networks*, 1983, 5(2): 109–137. doi: [10.1016/0378-8733\(83\)90021-7](https://doi.org/10.1016/0378-8733(83)90021-7).
- [9] SANZ-CRUZADO J, PEPA S M, and CASTELLS P. Structural novelty and diversity in link prediction[C]. Companion of the the Web Conference, 2018, Lyon, France, 2018: 1347–1351. doi: [10.1145/3184558.3191576](https://doi.org/10.1145/3184558.3191576).
- [10] LORRAIN F and WHITE H C. Structural equivalence of individuals in social networks[J]. *The Journal of Mathematical Sociology*, 1971, 1(1): 49–80. doi: [10.1080/0022250X.1971.9989788](https://doi.org/10.1080/0022250X.1971.9989788).
- [11] ZHOU Tao, LÜ Linyuan, and 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](https://doi.org/10.1140/epjb/e2009-00335-8).
- [12] ADAMIC L A and ADAR E. Friends and neighbors on the web[J]. *Social Networks*, 2003, 25(3): 211–230. doi: [10.1016/S0378-8733\(03\)00009-1](https://doi.org/10.1016/S0378-8733(03)00009-1).
- [13] CANNISTRACI C V, ALANIS-LOBATO G, and 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](https://doi.org/10.1038/srep01613).
- [14] 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](https://doi.org/10.1016/j.physa.2017.02.078).
- [15] KATZ L. A new status index derived from sociometric analysis[J]. *Psychometrika*, 1953, 18(1): 39–43. doi: [10.1007/BF02289026](https://doi.org/10.1007/BF02289026).
- [16] KLEIN D J and RANDIĆ M. Resistance distance[J]. *Journal of Mathematical Chemistry*, 1993, 12(1): 81–95. doi: [10.1007/BF01164627](https://doi.org/10.1007/BF01164627).
- [17] 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. doi: [10.1109/tkde.2007.46](https://doi.org/10.1109/tkde.2007.46).
- [18] LÜ Linyuan, JIN Cihang, and 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](https://doi.org/10.1103/PhysRevE.80.046122).
- [19] YANG Yujie, ZHANG Jianhua, ZHU Xuzhen, et al. Link prediction via significant influence[J]. *Physica A: Statistical Mechanics and Its Applications*, 2018, 492: 1523–1530. doi: [10.1016/j.physa.2017.11.078](https://doi.org/10.1016/j.physa.2017.11.078).
- [20] 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. doi: [10.7498/aps.63.158902](https://doi.org/10.7498/aps.63.158902).
- 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): 158902. doi: [10.7498/aps.63.158902](https://doi.org/10.7498/aps.63.158902).
- [21] WANG Xingyuan, ZHOU Wenjie, LI Rui, et al. Improving robustness of interdependent networks by a new coupling strategy[J]. *Physica A: Statistical Mechanics and Its Applications*, 2018, 492: 1075–1080. doi: [10.1016/j.physa.2017.11.037](https://doi.org/10.1016/j.physa.2017.11.037).
- [22] WANG Xingyuan, CAO Jianye, LI Rui, et al. A preferential attachment strategy for connectivity link addition strategy in improving the robustness of interdependent networks[J]. *Physica A: Statistical Mechanics and Its Applications*, 2017, 483: 412–422. doi: [10.1016/j.physa.2017.04.128](https://doi.org/10.1016/j.physa.2017.04.128).
- [23] WANG Xingyuan, CAO Jianye, and QIN Xiaomeng. Study of robustness in functionally identical coupled networks against cascading failures[J]. *PLoS One*, 2016, 11(8): e0160545. doi: [10.1371/journal.pone.0160545](https://doi.org/10.1371/journal.pone.0160545).
- [24] DEWHURST D R, DANFORTH C M, and DODDS P S. Continuum rich-get-richer processes: mean field analysis with an application to firm size[J]. *Physical Review E*, 2018, 97(6): 062317. doi: [10.1103/PhysRevE.97.062317](https://doi.org/10.1103/PhysRevE.97.062317).
- [25] ZENG Guoping and ZENG E. On the three-way equivalence of AUC in credit scoring with tied scores[J]. *Communications in Statistics-Theory and Methods*, 2017, 46(17): 1–16. doi: [10.1080/03610926.2018.1435814](https://doi.org/10.1080/03610926.2018.1435814).
- [26] 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. doi: [10.1016/j.physa.2017.11.103](https://doi.org/10.1016/j.physa.2017.11.103).
- [27] ZENG Xiangxiang, LIU Li, LÜ Linyuan, et al. Prediction of potential disease-associated microRNAs using structural perturbation method[J]. *Bioinformatics*, 2018, 34(14): 2425–2432. doi: [10.1093/bioinformatics/bty112](https://doi.org/10.1093/bioinformatics/bty112).
- [28] GOPAL S. The evolving social geography of blogs[M]. MILLER H J. *Societies and Cities in the Age of Instant Access*. Dordrecht, Springer, 2007: 275–293. doi: [10.1007/1-4020-5427-0_18](https://doi.org/10.1007/1-4020-5427-0_18).
- [29] MICHALSKI R, PALUS S, and KAZIENKO P. Matching organizational structure and social network extracted from email communication[C]. Proceedings of the 14th International Conference on Business Information Systems, Poznań, Poland, 2011.
- [30] ULANOWICZ R E and DEANGELIS D L. Network analysis of trophic dynamics in south Florida ecosystems[J]. *US Geological Survey Program on the South Florida*

- Ecosystem*, 2005, 114: 45–47. (未找到本条文献信息, 请核对)
- [31] WATTS D J and STROGATZ S H. Collective dynamics of ‘small-world’ networks[J]. *Nature*, 1998, 393(6684): 440–442. doi: [10.1038/30918](https://doi.org/10.1038/30918).
- [32] MICHALSKI R, PALUS S, and KAZIENKO P. Matching organizational structure and social network extracted from email communication[C]. Proceedings of the 14th International Conference on Business Information Systems, Poznań, Poland, 2011: 197–206. doi: [10.1007/978-3-642-21863-7_17](https://doi.org/10.1007/978-3-642-21863-7_17).
- [33] ADAMIC L A and GLANCE N. The political blogosphere and the 2004 U.S. election: divided they blog[C]. Proceedings of the 3rd International Workshop on Link Discovery, Chicago, USA, 2005: 36–43. doi: [10.1145/1134271.1134277](https://doi.org/10.1145/1134271.1134277).
- [34] LÜ Linyuan, PAN Liming, ZHOU Tao, *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. doi: [10.1073/pnas.1424644112](https://doi.org/10.1073/pnas.1424644112).
- [35] EWING R M, CHU P, ELISMA F, *et al.* Large-scale mapping of human protein-protein interactions by mass spectrometry[J]. *Molecular Systems Biology*, 2007, 3: 89. doi: [10.1038/msb4100134](https://doi.org/10.1038/msb4100134).
- [36] OPSAHL T and PANZARASA P. Clustering in weighted networks[J]. *Social Networks*, 2009, 31(2): 155–163. doi: [10.1016/j.socnet.2009.02.002](https://doi.org/10.1016/j.socnet.2009.02.002).
- [37] 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. doi: [10.11999/JEIT151338](https://doi.org/10.11999/JEIT151338).
- LIU Shuxin, JI Xincheng, LIU Caixia, *et al.* Information coupling of local topology promoting the network evolution[J]. *Journal of Electronics & Information Technology*, 2016, 38(9): 2180–2187. doi: [10.11999/JEIT151338](https://doi.org/10.11999/JEIT151338).
- 王 凯: 男, 1980年生, 副研究员, 博士生, 研究方向为链路预测、社会网络分析.
- 刘树新: 男, 1987年生, 助理研究员, 博士, 研究方向为复杂网络演化、链路预测、通信网络安全.
- 陈鸿昶: 男, 1964年生, 教授, 博士生导师, 研究方向为电信网安全、社团发现.
- 李 星: 男, 1987年生, 助理研究员, 博士生, 研究方向为社会网络分析.