基于资源传输匹配度的复杂网络链路预测方法

刘树新[1,2]，李星[1,2]，陈鸿昶[1,2]，王凯[1,2]
（1. 信息工程大学信息技术研究所，河南 郑州 450002；2. 国家数字交换系统工程技术研究中心，河南 郑州 450002）

摘 要：为了解决基于资源传输的链路预测方法忽略节点间匹配度对资源传输过程影响的问题，提出了一种基于资源传输匹配度的复杂网络链路预测方法。首先，对资源传输路径上的 2 个端点进行详细分析，提出任意节点间匹配度的量化方法；然后，为了刻画匹配度对于节点间资源传输过程的影响程度，定义了资源传输匹配度；最后，基于资源传输匹配度，考虑节点间双向传输的资源量，提出资源传输匹配度指标。在 9 个实际网络数据集上的实验测试表明，相比其他基于相似性指标，所提方法在 AUC 和 Precision 衡量标准下能够取得更好的效果。

关键词：复杂网络；链路预测；资源传输；匹配度
中图分类号：TP391
文献标识码：A
doi: 10.11959/j.issn.1000−436x.2020124

Link prediction method based on matching degree of resource transmission for complex network

LIU Shuxin[1,2], LI Xing[1,2], CHEN Hongchang[1,2], WANG Kai[1,2]
1. Information Technology Institute, Information Engineering University, Zhengzhou 450002, China
2. National Digital Switching System & Engineering Technology Research Center, Zhengzhou 450002, China

Abstract: In order to solve the problem that many existing resource-transmission-based methods ignore the important influence of the matching degree of two endpoints on resource transmission, a link prediction method was proposed based on matching degree of resource transmission for complex networks. Firstly, by analyzing the two endpoints on the resource transmission path in detail, the method of quantifying the matching degree between two nodes was proposed. Then, in order to describe the influence of matching degree on resource transmission process between nodes, the matching degree of resource transmission was defined. Finally, based on the matching degree of resource transmission, a resource transmission matching index was proposed considering the resource amount of bidirectional transmission between nodes. The experimental results of nine datasets show that compared with other similarity indices, the proposed index can achieve higher prediction accuracy under the AUC and Precision metrics.

Key words: complex network, link prediction, resource transmission, matching degree

1 引言

近年来，随着网络科学研究的不断开展，越来越多的复杂性系统通过复杂网络进行深度挖掘和分析，包括生物网络[1]、互联网[2]、交通网络[3]、社交网络[4]等复杂性系统[5]。链路预测作为网络科学领域的研究热点[6-7]，旨在利用现有网络信息预测未连接的节点对之间的连边概率，具体可应用于网络中缺失连边的预测[8]、未来可能存在连接的预测[9]及错误连边关系的发现[10]。

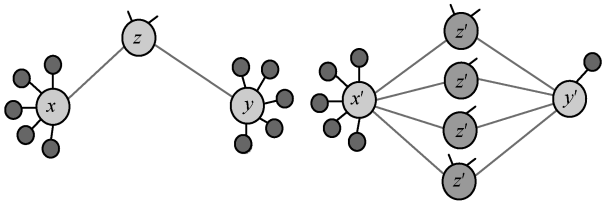
当前，链路预测相关研究方法较多，其中基于网络结构的相似性方法具有复杂度低、效果好的特点，受到各领域学者普遍关注。共同邻居（CN, common neighbor）[11]是最简单的相似性指标，其通

收稿日期：2019−12−17；**修回日期**：2020−03−11
基金项目：国家自然科学基金资助项目（No. 61803384）
Foundation Item: The National Natural Science Foundation of China (No. 61803384)

过计算任意两点之间共同邻居的数目刻画相似度。在 CN 的基础上，研究者相继提出了 AA（Adamic Adar）[12]、CAR[13]、资源分配（RA, resource allocation）[14]等局部相似性指标，且预测效果普遍好于 CN。Lyu 等[15]在 CN 基础上考虑了三阶路径，提出了局部路径（LP，local path），在牺牲一定复杂度的前提下，取得了非常好的效果。Liu 等[16]在 RA 的基础上提出了扩展的资源分配（ERA, extended resource allocation），Li 等[17]提出了资源传输容量（PIC, potential information capacity）。RA、ERA、PIC 等指标均对复杂网络中资源传输过程进行了不同角度的刻画和实际应用。此外，考虑网络全局信息，全局指标如全路径指标 Katz[18]、平均通勤时间（ACT, average commute time)[19]和余弦相似性指标 Cos+[20]被提出，其效果一般优于局部方法，但复杂度较高，难以应用于大型复杂网络。上述方法中，RA 在复杂度较低的情形下，取得了非常好的效果，在部分网络中接近甚至高于全局指标。然而，RA 仅从简单的资源传输过程描述出发对传输资源进行量化，忽略了节点间匹配度对资源传输过程的影响。

复杂网络中的关联匹配特性是多数网络中的常见现象，不同网络中或多或少具备同配和异配特性。现实网络中连边建立过程存在倾向性的现象也较常见[21]，如社交媒体网络中高影响力用户可能更倾向于和同等影响力节点建立联系，食物链网络中高度节点（食物链顶层的动物）更倾向于和低度节点（弱小动物）建立捕食关系。这种网络中的匹配度在实际网络中有着重要的偏好影响，因此导致每一个节点在资源传输过程中存在普遍的倾向性。

图 1 给出了资源传输过程中匹配度对连接建立的影响。图 1(a)中节点 x 和 y 具有非常接近的节点度，而图 1(b)中节点 x'和 y'的节点度相差较大。若在同配网络中，由于图 1(a)中节点 x 和 y 之间匹配度较高（即节点度较为接近），两节点之间进行资源传输的内在动力较大，此时仅有一个共同邻居，其对两节点之间建立联系提供较高的可能性；图 1(b)中节点 x'和 y'之间匹配度较低，两节点之间资源传输的可能性相对较小，需要大量的共同邻居方可为两节点之间建立联系提供较高的可能性。此时，若两对节点存在同样数目的共同邻居，则节点 x 和 y 之间更有可能建立连接关系。相反地，若在异配网络中，由于图 1(a)中节点 x 和 y 之间匹配度较高，两节点之间进行资源传输的内在动力较小，此时仅有一个共同邻居对两节点之间建立联系提供较低的可能性；图 1(b)中两节点 x'和 y'之间匹配度较低，两节点之间资源传输的可能会极大增强，仅需要少量的共同邻居即可为两节点之间建立联系提供较高的可能性。此时，若两对节点存在相同数目的共同邻居，则节点 x'和 y'之间更有可能建立连接关系。通过对比可以发现，网络中节点间的匹配度对于资源传输过程具有较大影响，对节点间建立连接的可能性也有明显的影响。



(a) 结构1　　(b) 结构2

图 1　资源传输中匹配度对连接建立的影响

针对上述问题，本文提出了一种基于资源传输匹配度的链路预测方法。首先，为了解决节点间匹配度的量化问题，对资源传输路径上的 2 个端点进行详细分析，提出节点间匹配度的量化方法；其次，针对匹配度对于节点间资源传输过程的影响程度量化问题，通过详细分析匹配度对资源量的影响，提出资源传输匹配度的具体量化方法；最后，为了解决基于资源传输匹配度的相似度刻画问题，从资源传输角度，考虑了节点间双向传输的资源量，对节点间相似性进行了重新刻画，提出资源传输匹配度指标。在曲线下面积（AUC, area under curve）和精确度 Precision 这 2 个衡量标准下，在多个实际网络中对所提指标进行了有效性分析，验证了该方法的预测效果。

2　基于资源传输匹配度的链路预测方法

2.1　资源传输匹配度的分析与量化

复杂网络中资源传输过程是指各类资源由起始节点出发，经过逐跳节点传播或转移至目的节点这一动力学过程[17]，其涉及不同的路径和节点，多跳路径中的每一条连边均影响着最后的资源传输总量。图 2 表示资源传输中多跳节点间存在匹配度问题。节点 x 和 y 之间经过多跳进行资源传输，其中的任意一段连边，如 v_i 和 v_j 的匹配度会对该连边的资源传输能力有影响。因此，节

点 x 和 y 之间传输的资源量与路径上各跳连边 $\{l_{xv_1}\cdots l_{v_iv_j}\cdots l_{v_{n-1}y}\}$ 的资源传输量息息相关，而任意多跳连边的端点匹配度对于资源量传输的“积极性”有很大影响。

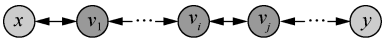


图 2 资源传输中多跳节点间存在匹配度问题

现实网络中，有的高度节点倾向于和高度节点分享热点话题（即资源传输），有的高度节点倾向于和低度节点进行热点话题分享。鉴于节点间匹配度对资源传输的重要影响，需要针对性研究如何量化节点间的匹配度。匹配系数是刻画一个复杂网络整体匹配程度的全局统计量。Newman 曾提出一种关联系数来刻画不同网络的匹配程度，该统计参数针对的是整个网络，并不适合于 2 个节点间的差异性。图 3 表示同配和异配这 2 种基本匹配模式的网络结构。

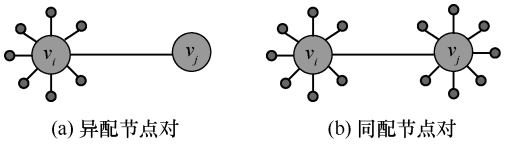


图 3 2 种基本匹配模式的网络结构

显然，图 3(a)中的节点 v_i 和 v_j 具有异配特性，图 3(b)中的节点 v_i 和 v_j 具有同配特性。可以看出节点度的差距是异配的主要表现形式，仅简单利用节点度之差进行表示为

$$\mathrm{Ad}'_{ij}=\frac{1}{\left|k_i-k_j+\tau\right|} \tag{1}$$

其中，k_i 和 k_j 分别为节点 v_i 和 v_j 的节点度，τ 防止出现分母为零的情形。式(1)能够一定程度上反映节点之间的节点度差距，但存在节点度差值较近时无法区分匹配度的问题。例如，一对节点的节点度分别为 $k_i=3$, $k_j=4$，另一对节点的节点度分别为 $k_i=13$, $k_j=14$。很明显，后者的匹配度应该更高，然而利用式(1)计算其结果是相同的。因此，需要详细地分析不同结构对节点间匹配度的影响，针对性地提出一种相对合理的量化方法。

图 4 为不同结构下节点间匹配度对比示意图。其中，图 4(a)和图 4(b)中的节点 v_j 具有相同的节点度，节点 v_i 具有不同的节点度，由于图 4(a)的 v_i 节点度更高，因此图 4(b)比图 4(a)具有更高的匹配度；对于图 4(b)和图 4(c)中的节点 v_i 和 v_j，虽然其节点度的差值相同，但图 4(c)的节点度相对较大，归一化后其相对差值更小，故图 4(c)比图 4(b)具有更高的匹配度；图 4(c)和图 4(d)中的节点 v_i 具有相同节点度，节点 v_j 具有不同的节点度，但由于图 4(d)的 2 个端点节点度相同，故图 4(d)比图 4(c)具有更高的匹配度。

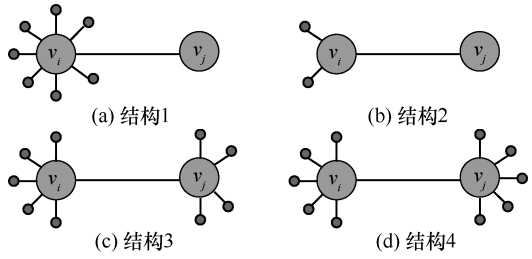


图 4 不同结构下节点间匹配度对比

考虑到上述情形，本文提出一种量化两节点间匹配程度的方法，即节点匹配度，具体表示为

$$\mathrm{Ad}_{ij}=\frac{2k_ik_j}{k_i^2+k_j^2} \tag{2}$$

显然，当 $k_i=k_j$ 时，$\mathrm{Ad}_{ij}=1$，此时匹配度最高为同配。通过式(2)对图 4 中各节点对之间匹配度进行计算，分别为

$$\mathrm{Ad}_{(a)}=\frac{7}{25}$$

$$\mathrm{Ad}_{(b)}=\frac{4}{5}$$

$$\mathrm{Ad}_{(c)}=\frac{40}{41}$$

$$\mathrm{Ad}_{(d)}=1$$

可以看出，$\mathrm{Ad}_{(a)}<\mathrm{Ad}_{(b)}<\mathrm{Ad}_{(c)}<\mathrm{Ad}_{(d)}$，较为符合上述结构匹配度的分析。

在对节点间匹配度量化后，便可以分析匹配度对资源传输过程的影响程度。通常情况下，任意两节点 v_i 和 v_j 之间的资源传输能力可以利用其多跳路径进行简单刻画，存在连接的节点间资源传输匹配度示意如图 5 所示。然而，考虑到两点之间匹配度的影响，节点间传输能力随着匹配度不同而发生不同变化（减少或增加）。

对于图 5 中直接相连的两点 v_i 和 v_j，本文将资源传输匹配度定义为两点之间匹配度对资源传输能力的影响程度，如式(3)所示。

$$\mathrm{Rad}_{ij}=(1+\left|\Gamma(i)\cap\Gamma(j)\right|)^{\mathrm{Ad}_{ij}\theta} \tag{3}$$

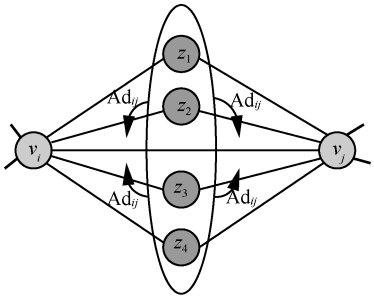


图 5　存在连接的节点间资源传输匹配度示意

其中，$\left|\Gamma(i)\cap\Gamma(j)\right|$ 为节点 v_i 和 v_j 的共同邻居数目，θ 为匹配度调节参数。当 $\theta\leqslant 1$ 时，$\mathrm{Ad}_{ij}\theta\leqslant 1$，此时会对固有的资源传输能力进行一定幅度的衰减；同样，$\mathrm{Ad}_{ij}\theta>1$ 会对固有的资源传输能力进行一定幅度的增加。

与直接连边的两点相似，对于图 6 中无直接连边的 2 个节点 v_i 和 v_j，影响节点间资源传输能力的主要因素是共同邻居和匹配度，此时其资源传输匹配度可以量化表示为

$$\mathrm{Rad}_{ij}=\left(\left|\Gamma(i)\cap\Gamma(j)\right|\right)^{\mathrm{Ad}_{ij}\theta} \tag{4}$$

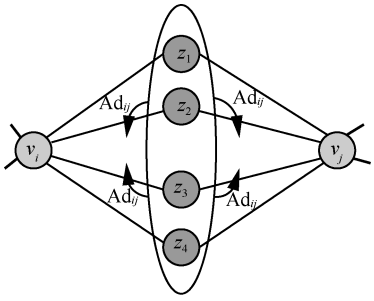


图 6　不存在连接的节点间资源传输匹配度示意

结合上述 2 种情形下的资源传输匹配度量化方法，下面给出网络中任意两点的资源传输匹配度定义。

定义 1　资源传输匹配度。对于网络中任意一对节点 v_i 和 v_j，两点之间存在资源传输的可能。节点间资源传输匹配度可用于量化两点之间匹配度对于节点间资源传输能力的影响程度，具体表示为

$$\mathrm{Rad}_{ij}=(a_{ij}+\left|\Gamma(i)\cap\Gamma(j)\right|)^{\mathrm{Ad}_{ij}\theta} \tag{5}$$

其中，a_{ij} 为邻接矩阵 $\boldsymbol{A}$ 中第 i 行第 j 列的元素值。该资源传输匹配度也一定程度上表征了节点间资源匹配度与资源传输能力的相互作用关系。

2.2　基于资源传输匹配度的相似性指标

在对任意节点间的资源传输匹配度进行量化后，便可以针对一个网络中所有节点对进行资源传输匹配度的计算。进一步地，需要在资源传输匹配度基础上，解决任意 2 个未连接节点之间相似度的定义问题。由于在资源传输匹配度计算量化中产生的 Rad 值各不相同，在传输资源过程中针对节点间的共同邻居两侧连边的 Rad 进行资源传输量的计算中便存在差异，因此在分析两点之间相似度时需要考虑资源的双向传输问题。

为了从节点间匹配度对资源传输影响的角度刻画任意两点间的相似性，本文将基于资源传输匹配度进行定义分析节点间资源传输能力。首先，对于网络中的任意 2 个未连接的节点 x 和 y，其中 z_1、z_2 是两点间的共同邻居，如图 7 所示。

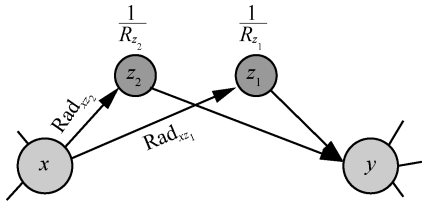


图 7　节点 x 传输资源至节点 y 的传输过程分析

假设节点 x 的连边上存在一个单元的资源经过节点间复杂拓扑结构传输到 y，由于资源传输匹配度影响了资源传输的偏好，其连边上的资源量分别变为 $1\times\mathrm{Rad}_{xz_1}$ 和 $1\times\mathrm{Rad}_{xz_2}$，则节点 y 接收到的资源量经过资源传输匹配度的影响和共同邻居节点的资源传输后，可以量化为

$$\mathrm{AR}(x\to y)=\sum_{z\in\Gamma(y)}\frac{\mathrm{Rad}_{xz}}{R_z}=\sum_{z\in\Gamma(y)}\frac{\left(a_{xz}+\left|\Gamma(x)\cap\Gamma(z)\right|\right)^{\mathrm{Ad}_{xz}\theta}}{R_z} \tag{6}$$

其中，$R_z=\sum_{i\in\Gamma(z)}\mathrm{Rad}_{iz}$ 表示了节点 z 与所有邻居节点的 Rad 值之和（与节点强度的概念类似）。上面分析了 x 和 y 资源传输过程，接下来分析 y 到 x 的资源传输过程。对于网络中的任意 2 个未连接的节点 x 和 y，其中 z_1、z_2 是两点间的共同邻居，节点 y 传输资源至节点 x 的传输过程如图 8 所示。

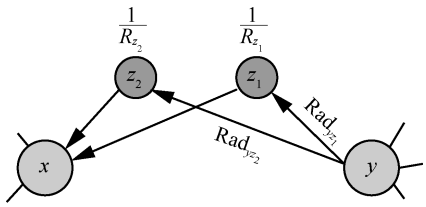


图 8　节点 y 传输资源至节点 x 的传输过程

假定节点 y 的连边上存在一个单元的资源经过节点间复杂拓扑结构传输到 x，由于资源传输匹配度影响了资源传输的偏好，其连边上的资源量分别变为 $1\times \mathrm{Rad}_{yz_1}$ 和 $1\times \mathrm{Rad}_{yz_2}$，则节点 x 接收到的资源量经过资源传输匹配度的影响和共同邻居节点的资源传输后，可以量化为

$$\mathrm{AR}(y\to x)=\sum_{z\in\Gamma(x)}\frac{\mathrm{Rad}_{yz}}{R_z}=\sum_{z\in\Gamma(x)}\frac{\left(a_{yz}+\left|\Gamma(y)\cap\Gamma(z)\right|\right)^{\mathrm{Ad}_{yz}\theta}}{W_z} \tag{7}$$

在定量分析网络中节点 x 与 y 之间互相资源传输后，接下来从资源传输匹配度角度描述整体传输过程，进而刻画节点间的相似性。

定义 2 节点间资源传输匹配度（AR，assortative resource）。对于一个无权无向网络 $G(V, E)$，x 和 y 为任意 2 个未连接的节点。2 个节点之间资源传输匹配度对于节点间相似性有着直接影响，其相似度可以定义为 x 和 y 相互之间经过资源传输匹配度的影响后，传输节点的资源双向传输量之和，如式(8)所示。

$$s_{xy}^{\mathrm{AR}}=\mathrm{AR}(x\to y)+\mathrm{AR}(y\to x)=\sum_{z\in\Gamma(y)}\frac{\mathrm{Rad}_{xz}}{R_z}+\sum_{z\in\Gamma(x)}\frac{\mathrm{Rad}_{yz}}{R_z} \tag{8}$$

式(8)中包含了两部分，一部分表示从 x 到 y 传输的资源量，另一部分表示从 y 到 x 传输的资源量。

3 衡量标准及数据集

3.1 算法衡量标准

在链路预测过程中，一般将一个网络中的连边 E 分为训练集 E^{T} 和测试集 E^{P}，$E=E^{\mathrm{T}}\cup E^{\mathrm{P}}$，且 $E^{\mathrm{T}}\cap E^{\mathrm{P}}=\varnothing$。选取的算法预测效果衡量指标为 AUC 和 Precision。其中，AUC 更多关注算法的整体预测准确性，是链路预测方法应用最广泛的评价方法之一；Precision 并不侧重整体预测效果，而更注重前 L 条预测连边的预测效果。在生物神经网络、蛋白质交互网等网络的隐含连接的研究中，排名靠前的预测连边准确度对于指导网络交互关系的发现实验具有重要实际价值，高 Precision 能够极大地节省实验时间，提高科研效率。

AUC 衡量指标可以理解为在测试集 E^{P} 中随机选择一条边的分数值比未连接边的分数值大的概率[22]。独立比较 n 次，其中测试集中边的分数大于未连接边的分数则加 1 分，测试集中边的分数等于未连接边的分数则加 0.5 分。分别用 n' 和 n'' 记录上述 2 种情况的个数，则 AUC 表示为

$$\mathrm{AUC}=\frac{n'+0.5n''}{n} \tag{9}$$

显然，如果所有的实验结果均是随机产生的，则 AUC≈0.5，因此 AUC 超过 0.5 的多少可以衡量当前算法比随机方法精确的程度。

Precision 指标可以理解为在前 L 个预测边中预测准确的比例[23]，定义为

$$\mathrm{Precision}=\frac{m}{L} \tag{10}$$

其中，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。对于具体的 L，Precision 越大，表明预测结果越准确。一般情况下，设置为 L=100。

3.2 数据集

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（acquired immuno deficiency syndrome）博客网[24]，下文简称为“AIDS”。一个 AIDS 博客相关的引用关系网络，节点代表该网络的用户，边表示用户之间存在的链接和关注关系。

2) 食物链网络（FW, food Web）[25]。在 Florida 海岸湿季的食物链网络，节点表示各个物种，边则为物种间存在捕食关系。

3) 线虫神经（CE, caenorhabditis elegans）网络[26]。线虫神经元用网络节点表示，线虫的神经元突触或其之间的连接用网络的连边表示。

4) 邮箱通信（EM, email）网络[27]。一种中型规模工厂的工人之间的邮箱通信网络，工人的邮箱地址用网络节点表示，不同工人之间的邮件往来用网络连边表示。

5) 美国政治博客（PB, politics blog）网络[28]。美国某政治论坛的博客首页之间的关系网络，节点为该论坛成员的首页，边代表他们之间存在的超链接。

6)美国航空线路网络 USAir[29]。网络的节点代表不同的机场，网络中的连边代表机场之间有直达航线。

7) HS（Hamster）[30]。在 Hamster 网站的网页上用户间朋友关系网络，节点代表用户，边表示用户间存在好友关系。

8) Infec（infectious）[31]。2009 年，在柏林科

学美术馆的表演 *Infectious: stay away* 中人们面对面行为的网络，参与的人员用网络节点表示，人们之间的沟通用连边表示。

9)线虫新陈代谢（Met, metabolic）网络[32]。秀丽隐杆线虫的新陈代谢网络，节点表示代谢物，连边表示他们之间的相互作用关系。

上述网络具体的特征参数如表 1 所示，包括基本节点数目$|V|$、边的个数$|E|$、平均度$<k>$、聚类系数（C）、平均最短路径$<d>$和匹配系数 r。

在实验测试中，设置训练集合 E^{T} 中边数占比为 0.9，测试集合 E^{P} 中边数占比为 0.1，每个实际测试结果均为 20 次结果的均值。

4　方法验证及结果分析

为了验证所提指标的预测效果，本节对比分析了 9 个网络数据中的 AUC 和 Precision，对预测效果进行对比。

4.1　AUC 指标分析

首先，在 AUC 衡量标准下，对 AR 指标进行对比分析。图 9 为 9 个网络中，参数θ对 AUC 的影响。9 个网络中 AUC 均随着参数的增加有一定程度的升高，这说明了资源传输匹配度考虑的合理性。大多数网络如 CE、EM、PB、USAir、HS、Infec、Met 中，AUC 随着参数迅速升高，然后维持在较高水平上，这说明了资源传输匹配度对节点间存在连接与否影响较大，且参数较小时便已经能够涵盖其主要影响。尤其是 HS 网络中，θ在 0～0.01 这一较小范围内便达到高值，这与网络的匹配系数为–0.085，较为接近 0 有关。FW 网络中随着θ变化，AUC 缓慢提升到达最大值，然后保持在较高值上，这说明资源匹配度对连边影响是缓慢的，当θ达到一定值时效果才最好。不同的是，AIDS 网络中随着θ增大，AUC 出现先下降后上升的趋势，这说明了当前网络中资源传输匹配度对连边影响较为复杂。总体上，多数网络中，在θ取值较小时，AUC 值均呈现较大程度的增长，即此时 AUC 值已经具有较好的预测效果。因此在实际网络的链路预测中，选取较小的θ值便具备较高的链路预测准确性。

图 10 显示了 AR 与现有指标的 AUC 对比结果。可以看出，相比其他相似性指标，在 9 个网络中 AR 均取得了较高的预测精度。CN 仅考虑了二阶邻居数目，其 AUC 结果普遍一般。AA 和 RA 考虑了共同邻居节点度，明显好于 CN，部分网络中（CE 和 EM）甚至好于全局指标。同样，考虑了共同邻居间存在的社团关系，CAR 在一些网络中效果好于 CN，但其他网络中效果并不理想。由于 LP 在 CN 基础上考虑了三阶路径，其效果明显好于局部相似性指标，且复杂度较低。全局指标 Katz 在考虑了所有路径信息后，效果普遍较好，但复杂度较高。此外，全局指标 ACT 和 Cos+的 AUC 结果也相对较高，但仍然存在复杂度较高的问题。

在考虑了资源传输匹配度后，本文所提方法 AR 预测效果明显较高，也说明了现实网络连边构建中均存在一定程度的匹配度倾向性。多个网络中如 AIDS、FW、USAir 和 HS 中，提高幅度较高，其中相比局部相似性指标提升比例为 20%~69%，相比全局指标提升比例为 1%~53%。总体上，相比其他相似性指标，AR 指标的平均提升比率为 15.4%，最高提升比率则高达 69%。实际网络的链路预测应用中，建议θ设置在 1.5 左右，此时，所有网络的 AUC 指标普遍较为理想。

4.2　Precision 结果分析

本节在 Precision 衡量标准下对 AR 指标进行结果分析。图 11 为 9 个不同网络中，Precision 结果

表 1　　**网络特征参数**

数据集	$\|V\|$	$\|E\|$	$<k>$	$<d>$	r	C
AIDS	146	180	2.47	3.42	–0.725	0.052
FW	128	2 075	32.42	1.78	–0.112	0.335
CE	297	2 148	14.46	2.46	–0.163	0.308
EM	167	5 784	69.26	1.87	–0.295	0.541
PB	1 222	16 717	27.36	2.74	–0.221	0.361
USAir	332	2 128	12.81	2.74	–0.208	0.749
HS	1 858	12 534	13.49	3.39	–0.085	0.090
Infec	410	2 765	5.76	3.98	–0.331	0.04
Met	453	2 025	8.94	2.66	–0.226	0.647

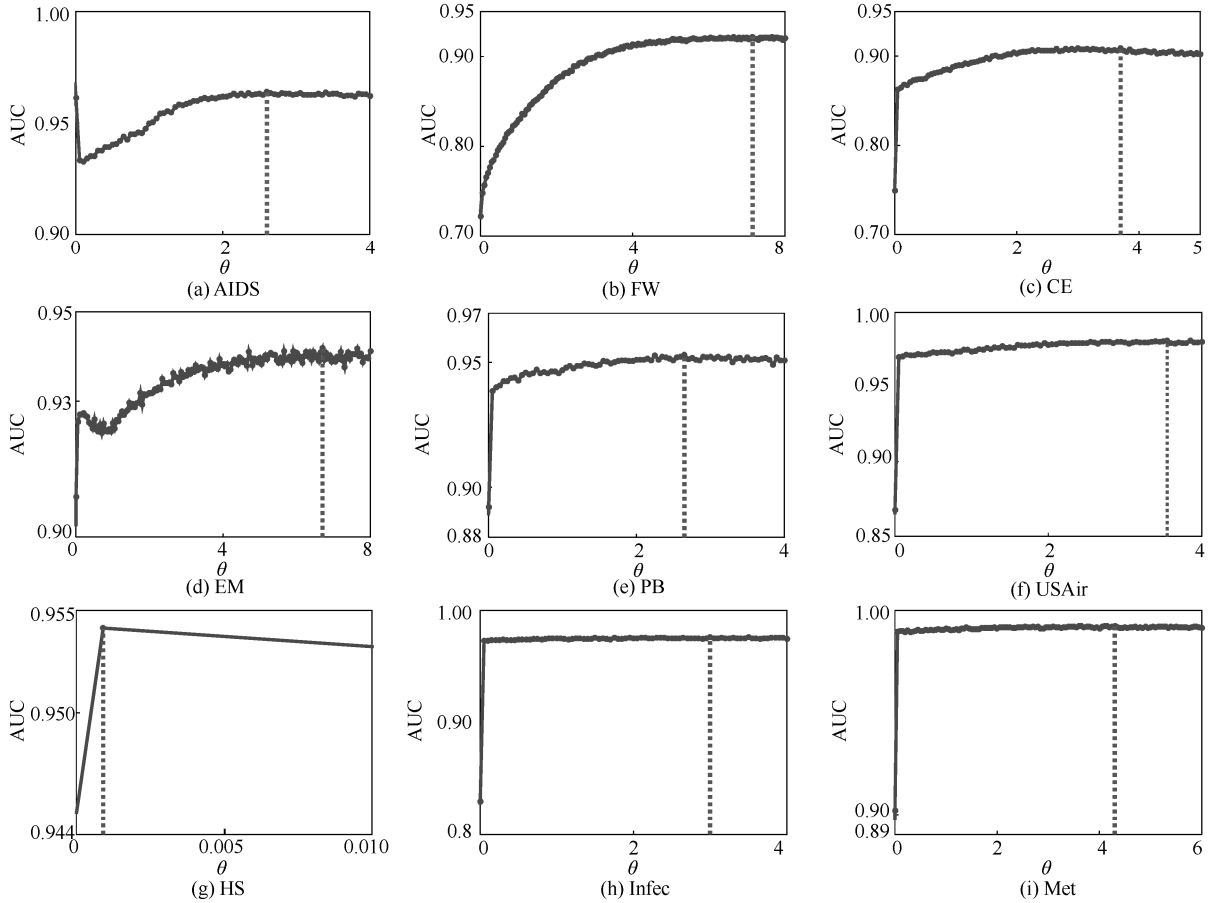


图 9　调节参数对 AUC 影响曲线

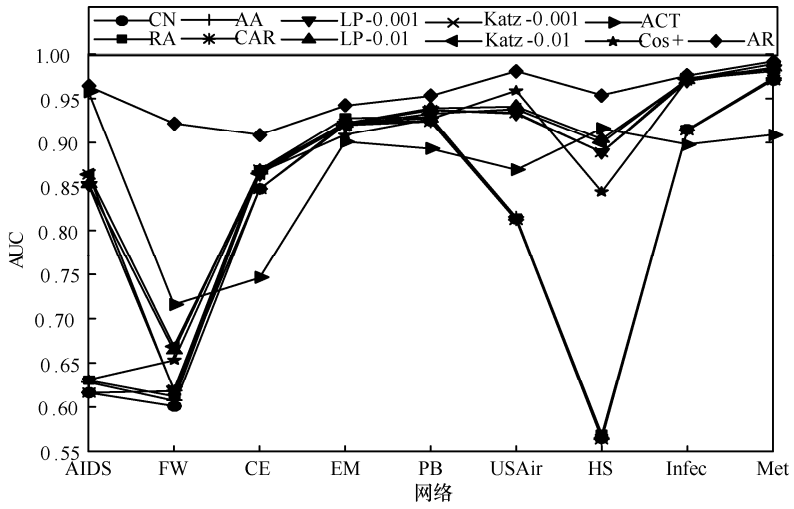


图 10　AUC 结果对比分析

随着参数的变化曲线。对于 AIDS、PB、HS 和 Infec，Precision 值随着 θ 增大迅速升高，后续则保持在较高水平上，这说明在这些网络中轻微考虑资源传输匹配度便具有较好的效果。对于 FW、CE、USAir 和 Met，Precision 曲线则呈现缓慢上升，到达最大值后略微下降，表明在这些网络中合理范围内的资源传输匹配度更加有利于连边预测。不同的是，在 EM 中，Precision 呈现先下降后上升并保持在较高水平上的趋势，说明了资源传输匹配度影响较为复杂，当其强度较高时对连边预测结果有较大效果。

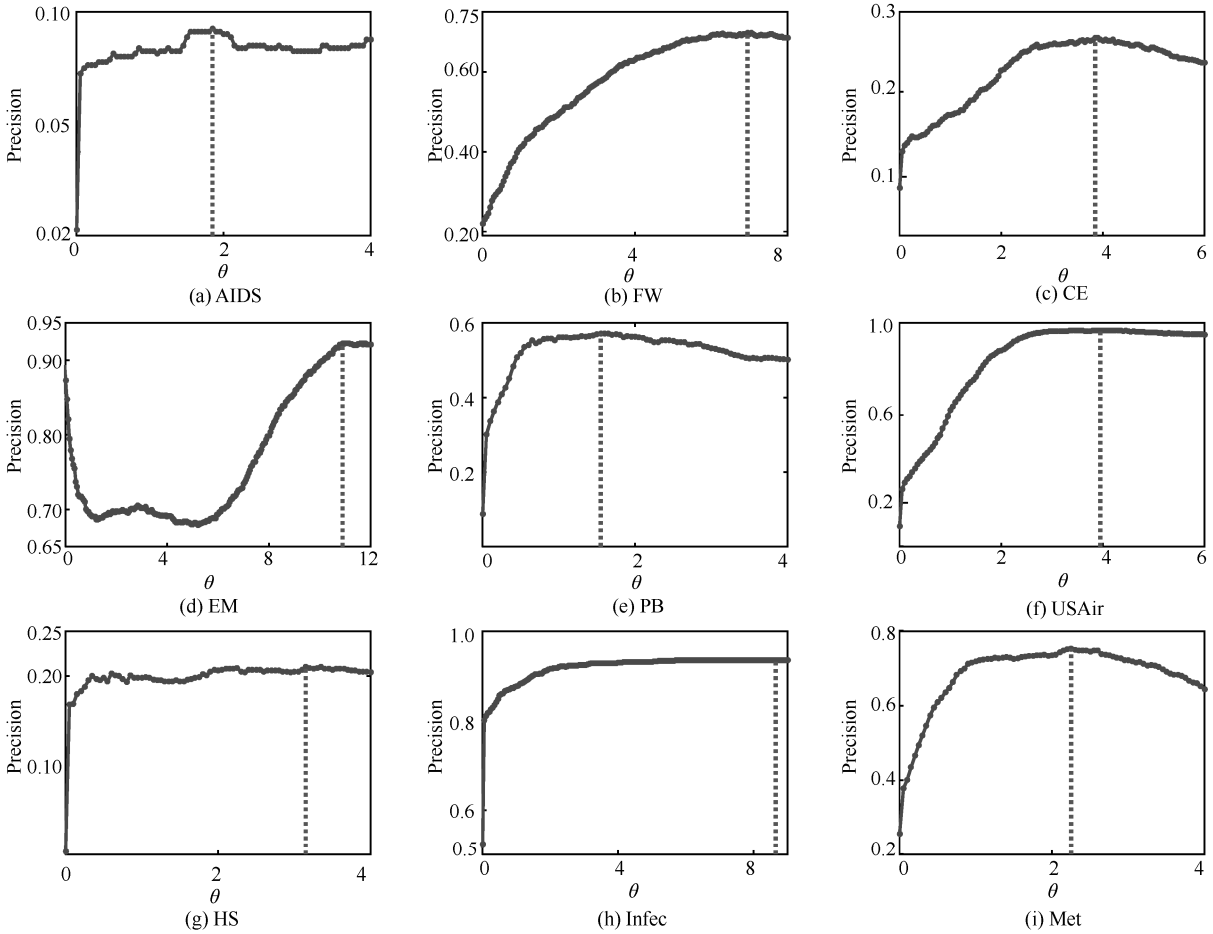


图 11　调节参数对 Precision 影响曲线

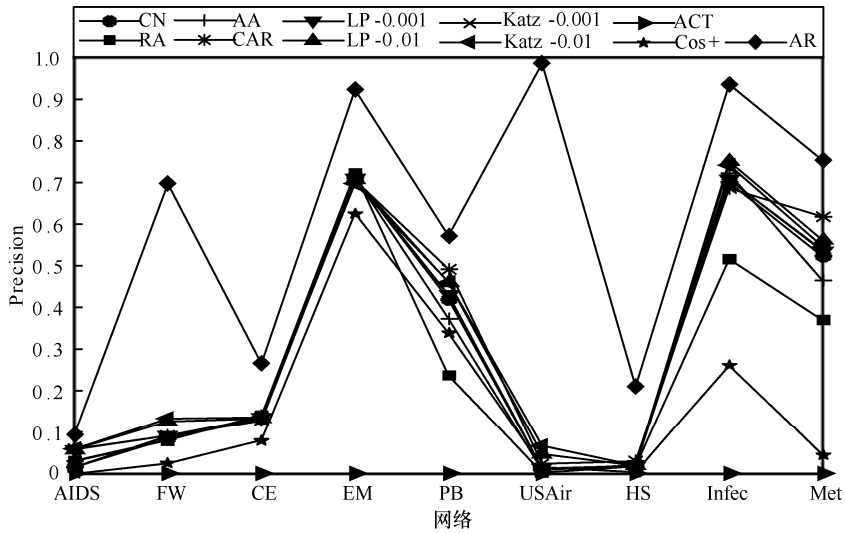


图 12　Precision 对比结果

总体上，多数网络中，在参数取值较小时，Precision 值均呈现快速有效的增长，即此时已经具有较好的 Precision 预测结果。因此，在实际网络的链路预测中，选取较小的参数值便具备较高的链路预测准确性。

图 12 为 Precision 对比结果。很明显，相比其他相似性方法，本文所提 AR 在 9 个网络中均具有较高的 Precision 值。与 AUC 结果不同，局部相似

性指标 CN、RA、AA 和 CAR 在不同网络中表现各异，某些网络中 CN 表现较好，而在另一些网络中则表现较差。这说明了在 Precision 衡量标准下，共同邻居节点本身信息的利用对于相似度的刻画贡献难以界定，额外考虑的信息并非一定起到正面促进的预测效果。相比局部相似性指标，LP 指标普遍较高，部分网络中甚至高于全局指标 Katz。Katz 指标在多个网络中表现高于其他指标，且预测结果相对稳定。此外，与 AUC 结果不同，全局指标 ACT 和 Cos+在 Precision 衡量标准下表现较差，很多网络中接近于 0，说明了这 2 个指标更倾向于 AUC。

在考虑了资源传输匹配度后，AR 指标的 Precision 结果明显较高。多个网络中如 FW、EM、USAir、HS、Infec 和 Met 中，提高幅度较高，其中相比局部相似性指标，Precision 值提升幅度为 0.19~0.98，相比全局指标提升幅度为 0.19~0.99（ACT 和 Cos+的 Precision 值较小，故提升幅度较高）。总体上，相比其他相似性指标，AR 指标的 Precision 结果平均提升幅度为 0.353，最高提升幅度为 0.99。具体链路预测应用中，针对 Precision 建议参数设置在 0.6 左右，此时对于所有网络中其 Precision 效果普遍较高。由于 AR 存在相关参数和建议值，因此实际应用中使用建议值并没有极值效果好，虽然多数网络中会明显高于其他方法，但部分网络中存在略好于其他方法的情况。

5 结束语

基于网络结构的相似性方法具有简单、复杂度低且效果好的特点，受到该领域学者普遍关注。针对现实网络中连边存在偏好的问题，从资源传输角度出发，本文提出了一种基于资源传输匹配度的链路预测方法。所提方法首先定义了匹配度用于刻画传输过程中任意两点间的匹配程度；然后，提出资源传输匹配度刻画传输过程的影响；最后提出资源传输匹配度指标。在多个实际网络中测试表明，考虑了资源传输匹配度后，其预测效果有了明显提升，这也证实了节点间匹配程度对网络中节点建立连边过程影响较大。所提方法复杂度较低，可以应用于大型复杂网络的链路预测。

参考文献：

[1] CUI Y, CAI M, DAI Y, 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.

[2] SHANMUKHAPPA T, IVAN W H, CHI K T. Spatial analysis of bus transport networks using network theory[J]. Physica A: Statistical Mechanics and Its Applications, 2018, 502: 295-314.

[3] CHENG Y, TAO F, XU L, et al. Advanced manufacturing systems: supply–demand matching of manufacturing resource based on complex networks and Internet of Things[J]. Enterprise Information Systems, 2018, 12(7): 780-797.

[4] KIM J, HASTAK M. Social network analysis[J]. International Journal of Information Management: The Journal for Information Professionals, 2018, 38(1): 86-96.

[5] 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): 1-11.
LIU S H, JI X S, LIU C X, et al. A complex network evolution model for network growth promoted by information transmission[J]. Acta Physica Sinica, 2014, 63(15): 158902.

[6] 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.
WANG K, LIU S X, CHEN H C, et al. A new link prediction method for complex networks based on resources carrying capacity between nodes[J]. Journal of Electronics and Information Technology, 2019, 41(5): 1225-1234.

[7] 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 S H, JI X S, LIU C X, et al. Information coupling of local topology promoting the network evolution[J]. Journal of Electronics and Information Technology, 2016, 38(9): 2180-2187.

[8] VON M 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): 433-437.

[9] SCELLATO S, NOULAS A, 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. New York: ACM Press, 2011: 1046-1054.

[10] HOLLAND P W, LASKEY K B, LEINHARDT S. Stochastic blockmodels: first steps[J]. Social Networks, 1983, 5(2): 109-137.

[11] LORRAIN F, WHITE H C. Structural equivalence of individuals in social networks[J]. Social Networks, 1977, 1(1): 67-98.

[12] ADAMIC L A, ADAR E. Friends and neighbors on the Web[J]. Social Networks, 2003, 25(3): 211-230.

[13] 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.

[14] ZHOU T, LÜ L Y, ZHANG Y C. Predicting missing links via local information[J]. The European Physical Journal B, 2009, 71(4): 623-630.

[15] LYU 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.

[16] LIU S H, JI X S, LIU C X, et al. Extended resource allocation index for link prediction of complex network[J]. Physica A: Statistical Mechanics and its Applications, 2017, 479: 174-183.

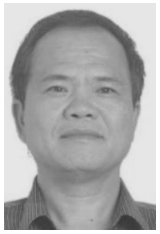
[17] LI X, LIU S X, CHEN H C, et al. A potential information capacity index for link prediction of complex networks based on the cannikin

law[J]. Entropy, 2019, 21(9): 863.
[18] KATZ L. A new status index derived from sociometric analysis[J]. Psychometrika, 1953, 18(1): 39-43.
[19] KLEIN D J, RANDIĆ M. Resistance distance[J]. Journal of Mathematical Chemistry, 1993, 12(1): 81-95.
[20] 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.
[21] MELAMED D, HARRELL A, SIMPSON B. Cooperation, clustering, and assortative mixing in dynamic networks[J]. Proceedings of the National Academy of Sciences, 2018, 115(5): 951-956.
[22] WU Y, YU H, ZHANG J, et al. USI-AUC: an evaluation criterion of community detection based on a novel link-prediction method[J]. Intelligent Data Analysis, 2018, 22(2): 439-462.
[23] CHUAN P M, ALI M, KHANG T D, et al. Link prediction in co-authorship networks based on hybrid content similarity metric[J]. Applied Intelligence, 2018, 48(8): 2470-2486.
[24] GOPAL S. The evolving social geography of blogs[M]. Berlin: Springer, 2007: 275-293.
[25] ULANOWICZ R E, 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.
[26] WATTS D J, STROGATZ S H. Collective dynamics of ‘small-world’ Networks[J]. Nature, 1998, 393(6684): 440
[27] GUIMERA R, DANON L, DIAZ-GUILERA A, et al. Self-similar Community structure in a network of human interactions[J]. Physical Review E, 2003, 68(6): 065103.
[28] ADAMIC L A, GLANCE N. The political blogosphere and the 2004 US election: divided they blog[C]//Proceedings of the 3rd International Workshop on Link Discovery. New York: ACM Press, 2005: 36-43.
[29] BATAGELJ V, MRVAR A. Pajek-program for large network analysis[J]. Connections, 1998, 21(2): 47-57.
[30] LYU L Y, PAN L M, ZHOU T, et al. Toward link predictability of complex networks[J]. Proceedings of the National Academy of Sciences, 2015, 112(8): 2325-2330.
[31] ISELLA L, STEHLÉ J, BARRAT A, et al. What's in a crowd? analysis of face-to-face behavioral networks[J]. Journal of Theoretical Biology, 2011, 271(1): 166-180.
[32] GUIMERA R, AMARAL L A N. Functional cartography of complex metabolic networks[J]. Nature, 2005, 433(7028): 895.

[作者简介]

刘树新（1987– ），男，山东临朐人，博士，信息工程大学助理研究员，主要研究方向为链路预测、通信网络安全。

李星（1987– ），男，河南新乡人，博士，信息工程大学助理研究员，主要研究方向为链路预测、社团挖掘。



陈鸿昶（1964– ），男，河南郑州人，信息工程大学教授、博士生导师，主要研究方向为通信与信息系统、数据科学与人工智能。

王凯（1980– ），男，河南许昌人，博士，信息工程大学副研究员，主要研究方向为链路预测、社会网络分析。