

基于节点亲密度的链路预测算法

吕亚楠 韩 华 贾承丰 完颜娟

(武汉理工大学理学院 武汉 430070)

摘 要 链路预测作为复杂网络分析的一个重要分支,在不同领域中有着广泛的应用。现有的链路预测算法通常根据共同邻居节点的结构信息来度量节点对之间的相似性,忽略了节点对与其共同邻居节点之间的连接紧密程度。针对此问题,提出了一种基于节点亲密度的链路预测算法。该算法利用边聚集系数来测量节点对与其共同邻居节点之间的紧密程度,以 AUC 值作为链路预测的精确度评价指标。在 4 个真实网络上的实验结果表明,相比于其他相似性算法,所提出的算法提高了链路预测的预测精度。

关键词 复杂网络,链路预测,相似性,亲密度

中图分类号 TP393 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.11.013

Link Prediction Algorithm Based on Node Intimate Degree

LV Ya-nan HAN Hua JIA Cheng-feng WAN Yan-juan

(School of Science, Wuhan University of Technology, Wuhan 430070, China)

Abstract As an important branch of complex network analysis, link prediction has extensive application in different fields. The existing link prediction algorithms usually measure the similarity between two nodes only through the structure information of their common neighborhoods, ignoring the tightness between nodes and their common neighbor nodes. To solve this problem, the paper proposed a link prediction algorithm based on node intimate degree. This algorithm employs the clustering coefficients of edge to measure the intimacy between nodes and their common neighbor nodes, and adopts the receiver operation characteristic Area Under Curve (AUC) as the standard index of link prediction accuracy. The experimental results on four real networks show that the proposed algorithm can improve the prediction precision compared with other link prediction algorithms.

Keywords Complex networks, Link prediction, Similarity, Intimate degree

1 引言

随着复杂网络理论不断发展,可以利用复杂网络来对许多复杂系统(如生物系统、社会系统、信息系统等)进行研究。其中,节点表示生物元素(如蛋白质、基因等)、生物种类、一个人、Web 用户、邮箱等,链接表示节点之间的关系或者相互作用。链路预测作为复杂网络研究中的一个热点问题,主要利用现有的网络信息预测网络中没有连边的两节点之间产生连边的可能性,既可能是对网络中缺失连边的预测,也可能是对网络中未来产生连边的预测^[1]。

在现实生活中,复杂网络中的链路预测有着广泛的应用。如生物系统中的蛋白质相互作用网络,利用已有的蛋白质与蛋白质之间相互作用的影响,预测出蛋白质之间隐而未知的连边^[2-3];线上购物网络中,根据人们的购物习惯和购物倾向,将链路预测用于挖掘潜在的客户和推荐客户感兴趣的商

品^[4];在线上社交网络中,通过朋友之间的关系,预测和挖掘出可能相识的朋友^[5]以及网络中用户的关系^[6];此外,链路预测方法在识别犯罪网络结构、检测和控制网络攻击等应用中也发挥着重要作用^[7]。

近年来,链路预测方法的研究受到广泛关注。这些方法大致分为 3 类:基于最大似然估计的方法^[8,10]、概率模型方法^[9-10]和基于结构相似性的方法^[10]。其中,基于结构相似性的方法由于方法简单、计算复杂度相对较低、使用范围广等特点受到广泛关注。基于结构相似性的方法通常利用共同邻居节点的自身信息及其结构信息计算节点对之间的相似性得分,相似性得分越高,它们之间存在连接的可能性就越大。Yang 等^[11]结合节点的共同邻居和节点之间的距离度量节点的相似性,提出了一种基于共同邻居和距离的相似性指标,并取得了较好的预测效果。陈嘉颖等^[12]考虑了共同邻居节点在网络中的影响力,提出考虑共同邻居节点重要性的 CN,

到稿日期:2017-10-09 返修日期:2018-01-27 本文受国家自然科学基金项目(71372135,71140015),中央高校基本科研业务费专项基金项目(2015-zy-115)资助。

吕亚楠(1992—),女,硕士生,主要研究方向为链路预测、复杂网络分析,E-mail: Nan_Nan_Wuli2016@163.com;韩 华(1975—),女,博士,教授,主要研究方向为复杂性分析与评价、经济控制与决策等,E-mail: hanhua@whut.edu.cn(通信作者);贾承丰(1994—),男,硕士生,主要研究方向为链路预测、机器学习;完颜娟(1994—),女,硕士生,主要研究方向为网络级联失效、相依网络。

AA, RA 链路预测相似性指标。高扬等^[13]考虑到共同邻居节点的聚集系数对链路的产生会有影响,提出了一种结合节点度和聚集系数的链路预测算法,该算法结合了共同邻居节点自身的信息和共同邻居节点之间的结构信息进行预测,预测精度得到了提高。

2 相关工作

基于结构相似性的链路预测方法的关键是其是否能够很好地抓住目标网络的结构特征。根据所用结构信息的不同,基于结构的相似性预测算法可以分为以下 3 类:

1) 基于局部信息的相似性算法。如考虑共同邻居数目的 CN 指标^[14];利用共同邻居节点度的 Adamic-Adar 指标(AA 指标)^[15];受网络资源分配过程的启发,Zhou 等^[16]提出了资源分配指标(RA 指标);Wu 等^[17]认为局部连边和三角形结构信息可直接利用共同邻居节点的聚集系数来表达,提出了基于节点聚集系数的 CCLP 指标等。基于局部信息的相似性算法大多是以节点的共同邻居或共同邻居之间的结构信息为基础进行的相似性度量,对于聚集系数较高的网络,预测精度较高^[18],且计算复杂度较低,适用于大型复杂网络。

2) 基于路径的相似性算法。如在共同邻居的基础上考虑三阶路径因素的局部路径指标(LP 指标)^[19];考虑网络中所有路径的 Katz 指标^[20]和 LHN-II 指标^[21];考虑到端点的有效影响力和网络强的连通性对相似性有一定影响,Zhu 等^[22]提出了一种有效路径指标(EP 指标)等。基于路径的相似性算法在共同邻居的基础上考虑了多阶邻居节点的贡献,计算复杂度相对较高,因此不适用于大规模稀疏网络。

3) 基于随机游走的相似性算法。该算法是基于随机游走过程定义的,如通过随机游走粒子在节点对之间游走的时间来衡量相似性的平均通勤时间^[23];拓展网页排序算法的有重启的随机游走指标(RWR 指标)^[24];只考虑有限步数的局部随机游走相似性指标(LRW 指标)^[25];考虑在网络结构上不同邻居节点间的相似性对转移概率的作用,刘思等^[26]提出了一种基于网络表示学习与随机游走的链路预测算法等。基于随机游走的预测方法从概率的角度考虑了网络结构对连边形成的作用,刻画出了粒子随机游走的过程,在一定程度上优于前述方法。但是由于涉及全网的矩阵运算,计算复杂度较高,因此对于大规模网络并不适用。

然而,现有的基于局部信息的相似性算法主要是考虑共同邻居节点信息,忽略了共同邻居节点与节点对之间的结构信息。为了充分利用共同邻居的节点信息和共同邻居与节点对之间的信息,本文在共同邻居节点相似性指标的基础上,考虑节点对与其共同邻居节点之间结构的信息,利用边的聚集系数^[27]来评估两节点之间的结构信息,并将此结构信息看作两节点之间的亲密度,提出一种基于节点亲密度的链路预测算法。

3 问题描述与评价方法

$G=(V, E)$ 表示一个无向无权的连通网络。其中, V 是网络 G 的所有节点的集合,节点总数量 $|V|=n$, E 是网络 G 中所有边的集合,边的数量 $|E|=m$ 。网络中自连边和重连边是

不允许的。连通的网络 G 可以用邻接矩阵 A 来表示,矩阵 A 中的元素 $a_{ij}=1$ 表示节点 v_i 和节点 v_j 之间有连边, $a_{ij}=0$ 表示节点 v_i 和节点 v_j 之间没有连边。对于网络 G 中每一对未连边的节点 v_i 和节点 v_j ,通过给定的预测方法计算出相似分数值 S_{ij} ,将所有未连边的节点对按照相似分数值从高到低排列,分数越高表示两节点越有可能连边。

本文使用 6 个经典的基于局部结构的相似性指标和 1 个基于节点聚集系数的 CCLP 指标进行对比,各指标的相关定义如表 1 所列。

表 1 相似性指标
Table 1 Similarity indices

指标	定义	公式含义
CN 指标	$S_{ij} = \Gamma(i) \cap \Gamma(j) $	节点对共同邻居节点的个数
Jaccard 指标	$S_{ij} = \frac{ \Gamma(i) \cap \Gamma(j) }{ \Gamma(i) \cup \Gamma(j) }$	节点对邻居节点集合的交集与邻居节点集合的并集的比值
AA 指标	$S_{ij} = \sum_{x \in \Gamma(i) \cap \Gamma(j)} \frac{1}{\log k_x}$	节点对共同邻居节点度的对数的倒数之和
RA 指标	$S_{ij} = \sum_{x \in \Gamma(i) \cap \Gamma(j)} \frac{1}{k_x}$	节点对共同邻居节点度的倒数之和
LP 指标	$S_{ij} = A^2 + aA^3$	节点对二阶路径与三阶路径之和
LRW 指标	$S_{ij}^{LRW}(t) = \frac{k_i}{2 E } \pi_{ij}(t) + \frac{k_j}{2 E } \pi_{ji}(t)$	节点对随机游走的概率之和
CCLP 指标	$CCLP(i, j) = \sum_{x \in \Gamma(i) \cap \Gamma(j)} \frac{t_x}{k_x(k_x - 1)/2}$	节点对共同邻居节点的点聚集系数之和

表 1 中, $\Gamma(i)$ 表示节点 v_i 的邻居节点的点集, k_x 为节点 v_x 的度值, t_x 是网络中包含节点 v_x 的三角环个数。

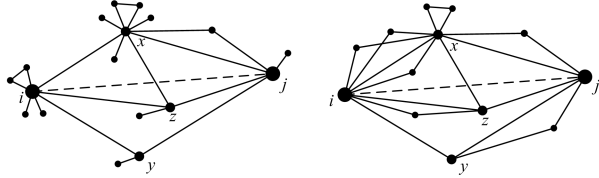
为了测试指标的准确性,一般将已知的连边 E 分为两部分:训练集 E^T 和测试集 E^P 。将训练集 E^T 作为实验时可观察到的网络信息用于计算节点的相似性得分。将测试集 E^P 作为实验时要预测的网络信息用于对比实验预测的结果。文中使用链路预测中评价准确性认可度高的指标——AUC 指标。AUC 指标是从整体上衡量指标的精确度,它是指在测试集中随机选择一条边的分数值比随机选择的一条不存在的边的分数值高的概率^[28]。实验时,每次随机从测试集中选取一条边,再从不存在的边中随机选择一条,如果测试集中边的分数值大于不存在的边的分数值,那么就加 1 分;如果两个分数值相等就加 0.5 分。这样独立比较 n 次,如果有 n' 次测试集中的边分数值大于不存在的边分数,有 n'' 次两分数值相等,那么 AUC 指标的定义为:

$$AUC = \frac{n' + 0.5n''}{n} \quad (1)$$

4 一种基于节点亲密度的链路预测算法

基于节点相似性的预测算法基本上是从被预测的节点对共同邻居节点出发,分析共同邻居节点的局部特征,却忽略了预测的两节点与其共同邻居节点之间的紧密程度,即节点之间的亲密度。节点之间的亲密度越高,两节点的关系越密切。如图 1 所示,预测节点 v_i 和节点 v_j 之间存在连接的可能性。其中,节点 v_x, v_y, v_z 是未连接的节点对 v_i, v_j 的 3 个共同邻居节点,可以看出,图 1(a)和图 1(b)中节点 v_i 和节点 v_j 的度值相等,节点 v_x, v_y, v_z 的度值也相等,但图 1(b)中的节

点 v_i 和节点 v_x 两者之间的亲密度比图 1(a) 中两者之间的亲密度更高, 因此图 1(b) 中经过节点 v_x 的二阶路径对节点对的相似性贡献比图 1(a) 中经过节点 v_x 的二阶路径更大。类似地, 图 1(b) 中分别经过节点 v_y 和节点 v_z 的二阶路径对节点对的相似性贡献比图 1(a) 中分别经过节点 v_y 和节点 v_z 的二阶路径更大。可以看出, 图 1(b) 中的节点 v_i 和节点 v_j 要比图 1(a) 中的节点 v_i 和节点 v_j 有更大的连接可能。



(a) 共同邻居节点与两节点的亲密度低 (b) 共同邻居节点与两节点的亲密度高

图 1 两种不同的节点连接方式

Fig. 1 Two different connection modes of nodes

由图 1 可以看出, 此时基于局部相似性的基础指标都无法准确地预测出节点 v_i 和节点 v_j 之间存在连接的可能性。因此, 考虑到节点对与其共同邻居之间的亲密度对相似性有一定的影响, 本文提出一种基于节点亲密度的链路预测算法。该算法在考虑节点对共同邻居节点数目时, 还考虑了共同邻居节点与两节点之间的局部网络拓扑信息。利用边聚集系数刻画两节点之间关系的亲密程度, 两节点之间的亲密程度越高, 表明两节点越相似。节点对与其共同邻居节点之间的相似性越大, 节点对之间就更有可能会产生连边。根据边聚集系数的定义, 即一条边的两端点与其共同邻居节点之间的另外两边所组成的三角环的数目与可能包含该边最大的三角环数之间的比值^[28], 可以看出这里利用边聚集系数来评估两节点之间的相似性更准确。

基于上述观点, 定义节点 v_i 和节点 v_x 之间的亲密度(Intimate Degree)为:

$$ID(i, x) = \frac{\Gamma(i) \cap \Gamma(x) + 1}{\min\{k_i - 1, k_x - 1\}} \quad (2)$$

其中, k_i 为节点 v_i 的度值, 分子上的“1”表示节点 v_i 和节点 v_x 之间有连边。对于两个没有连边的节点 v_i 和节点 v_j , 结合节点亲密度的相似性指标定义为:

$$S_{ij} = |\Gamma(i) \cap \Gamma(j)| + \sum_{x \in \Gamma(i) \cap \Gamma(j)} (ID(i, x) + ID(j, x)) \quad (3)$$

简称其为 CNID 指标。CNID 指标算法的流程如算法 1 所示。

算法 1 CNID 算法

输入: 网络 $G(V, E)$, 节点 v_i, v_j, v_x (其中 $\Gamma(i)$ 表示节点 v_i 的邻居集合, k_i 表示节点 v_i 的度, $\Gamma(i) \cap \Gamma(j)$ 表示节点 v_i 和节点 v_j 的共同邻居节点集合, 且 $v_x \in \Gamma(i) \cap \Gamma(j)$, n 表示网络节点的总数量)

输出: 节点 v_i 和节点 v_j 的相似性 S_{ij}

1. for $i=1$ to n
2. 计算 $CN(i, x) = \Gamma(i) \cap \Gamma(x)$, $CN(j, x) = \Gamma(j) \cap \Gamma(x)$
3. 计算 $ID(i, x) = \frac{CN(i, x) + 1}{\min\{k_i - 1, k_x - 1\}}$, $ID(j, x) = \frac{CN(j, x) + 1}{\min\{k_j - 1, k_x - 1\}}$
4. end
5. for $i=1$ to n
6. 计算 $CN(i, j) = \Gamma(i) \cap \Gamma(j)$
7. 计算 $S_{ij} = CN(i, j) + \sum_{x \in \Gamma(i) \cap \Gamma(j)} (ID(i, x) + ID(j, x))$
8. end

分别针对图 1(a)、图 1(b) 两种情况, 计算节点对 v_i 和 v_j 的相似性值。

对于图 1(a), 有:

$$\begin{aligned} S_{ij} &= 3 + \left(\frac{0+1}{\min\{6,8\}} + \frac{1+1}{\min\{8,4\}} + \frac{0+1}{\min\{6,3\}} + \frac{0+1}{\min\{3,4\}} + \right. \\ &\quad \left. \frac{0+1}{\min\{6,2\}} + \frac{0+1}{\min\{2,4\}} \right) \\ &= \frac{16}{3} \end{aligned}$$

对于图 1(b), 有:

$$\begin{aligned} S_{ij} &= 3 + \left(\frac{3+1}{\min\{6,8\}} + \frac{1+1}{\min\{8,4\}} + \frac{1+1}{\min\{6,3\}} + \frac{0+1}{\min\{3,4\}} + \right. \\ &\quad \left. \frac{0+1}{\min\{6,2\}} + \frac{1+1}{\min\{2,4\}} \right) \\ &= \frac{20}{3} \end{aligned}$$

可以看出, 按照本文提出的算法, 图 1(b) 的节点对 v_i 和 v_j 的相似性分数比图 1(a) 的相似性分数高, 因此图 1(b) 中的节点对 v_i 和 v_j 更有可能产生连边, 与预想吻合。而且在计算相似性分数时, 也刻画出了两节点与其每个共同邻居节点之间的亲密度大小。

5 实验

为了评价算法的有效性, 本文在 4 个典型的真实网络数据上展开实验, 用 Matlab 作为实验工具, 平均进行 100 次独立实验。

5.1 实验数据

采用的 4 个典型的真实网络数据的特征如下:

1) Dolphins 网络^[29]。该网络是新西兰神奇湾(Doubtful Sound)海豚关系网络, 该网络中的节点表示海豚, 海豚之间的连边表示海豚成对出现的次数高于预期次数。网络中共有 62 个节点, 159 条链路关系。

2) Football 网络^[30]。该网络是大学生橄榄球队在冠军争夺赛上的比赛关系网络, 网络中的节点表示球队, 边代表两支球队有过比赛关系。网络中共有 115 个节点, 615 条链路关系。

3) FFWF 网络^[31]。该网络是佛罗里达海湾雨季的食物链网络, 网络中的节点表示生物种类, 边表示各个生物之间的捕食关系。网络中共有 128 个节点, 2075 条链路关系。

4) Political blogs 网络^[32]。该网络是美国政治博客网络, 网络中的节点表示个人博客, 边表示博客之间的联系。网络中共有 1222 个节点, 19021 条链路关系。

5.2 结果与分析

本文将改进的链接预测方法应用于 Dolphins 网络、Football 网络、FFWF 网络及 Political blogs 网络 4 个真实网络, 分别以基于局部相似性链路预测算法的 CN 指标、Jaccard 指标、AA 指标、RA 指标和 CCLP 指标以及基于局部路径的 LP 指标和局部随机游走相似性的 LRW 指标作为基础性指标与改进的指标进行对比, 用 AUC 指标衡量实验结果的精确度。实验时随机地划分训练集与测试集, 第一次实验划分测试集所占比例为 10%, 第二次实验划分测试集所占比例为 20%。

实验结果如图 2 所示。

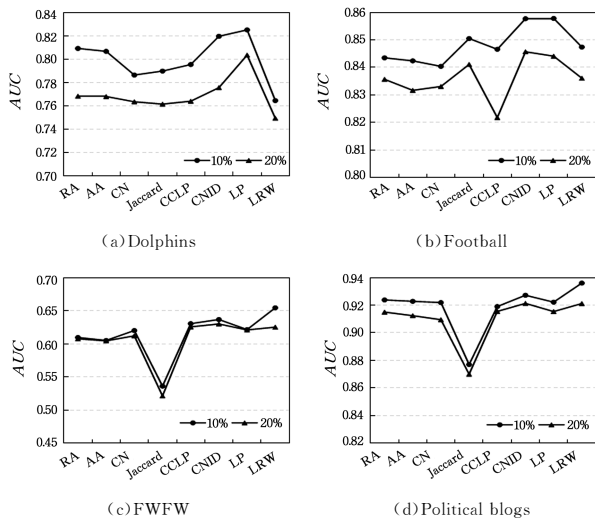


图 2 测试集由 10% 增加到 20% 时不同指标下 AUC 的变化

Fig. 2 Changes of AUC under different indices when test set increases from 10% to 20%

可以看出,当测试集由 10% 增加到 20% 时,各个指标的 AUC 值都有所下降,这是因为训练集所占比例减小,网络中节点之间的共同邻居数目也随之减小,导致预测精度有所不同。其中,Dolpins 网络的网络聚集系数比其他 3 个网络低,因此它的 AUC 值的变化幅度较大。且无论测试集是 10% 还

是 20%,CNID 指标的预测精度在基于局部相似性链路预测算法中表现较好。由于 RA 指标和 AA 指标具有相似的表达形式,因此两者的预测精度几乎一样。在 Dolphins 网络中,由于网络中的连边较少,导致 LRW 指标预测精度较低。在 FFWW 网络和 Political blogs 网络中,相比于其他指标,Jaccard 指标预测精度偏低,这说明这两个网络中节点之间的共同特征与其他特征的差异性较大,Jaccard 指标不适合预测这两个网络。

表 2 列出了网络测试集所占比例为 10% 时,4 个真实网络在不同算法下的预测精度 AUC 值。可以看出,算法横向比较时,所提出的 CNID 算法与基于局部相似性的链路预测算法相比,各个网络的预测精度 AUC 值都有一定的提高。若以 RA 算法为基准,网络的预测精度 AUC 值在 Dolphins 网络中提高了 1.04%,在 Football 网络中提高了 1.42%,在 FFWW 网络中提高了 2.71%,在 Political blogs 网络中提高了 0.34%,表明节点之间的亲密度对节点对之间的相似性是有一定影响的。算法纵向比较时,从 Dolphins 网络和 Football 网络中可以看出,LRW 指标在最优步数下的预测精度并没有 CNID 指标高,且 FFWW 网络和 Political blogs 网络的 LP 指标预测精度与 CNID 指标相差不大。进而可以看出,由于 CNID 指标抓住了节点对与共同邻居节点之间的网络结构特征,因此在网络中获取到的网络结构信息相对全面,预测精度得到了一定的提高。

表 2 测试集为 10% 时 4 个真实网络在不同算法下的预测精度 AUC 值

Table 2 AUC of different methods in four real networks when test set is 10%

网络	RA	AA	CN	Jaccard	CCLP	LP	LRW	CNID
Dolphins	0.8091	0.8066	0.7863	0.7897	0.7953	0.8250	0.7644(2)	0.8195
Football	0.8435	0.8424	0.8404	0.8505	0.8466	0.8578	0.8474(3)	0.8577
FFWW	0.6096	0.6051	0.6201	0.5362	0.6305	0.6215	0.6544(4)	0.6367
Political blogs	0.9238	0.9228	0.9219	0.8770	0.9190	0.9222	0.9360(3)	0.9272

表 3 列出了网络测试集所占比例为 20% 时,4 个真实网络在不同算法下的预测精度 AUC 值。类似地,可以看出当测试集的比例增加到 20% 时,所提出的 CNID 算法比基于局部相似性的链路预测算法的效果好,各个网络的预测精度 AUC 值也有一定的提高。若以 RA 算法为基准,各个网络的预测精度 AUC 值基本上都提高了 1% 以上。将算法纵向比较时,可以看出 Dolphins 网络和 Football 网络的 LRW 指标和

CNID 指标相差 1% 以上;FFWW 网络和 Political blogs 网络的 LP 指标预测精度与 CNID 指标相差不明显,进而说明了本文提出的算法的有效性和可行性。另外,当测试集为 20% 时,除了 FFWW 网络和 Political blogs 网络的 Jaccard 指标的预测精度 AUC 值偏低外,其他指标之间的变化较小,说明共同邻居数目的减少对各个指标的预测结果的差异不是很明显。

表 3 测试集为 20% 时 4 个真实网络在不同算法下的预测精度 AUC 值

Table 3 AUC of different methods in four real networks when test set is 20%

网络	RA	AA	CN	Jaccard	CCLP	LP	LRW	CNID
Dolphins	0.7682	0.7679	0.7634	0.7614	0.7639	0.8034	0.7493(2)	0.7755
Football	0.8357	0.8317	0.8331	0.8411	0.8218	0.8441	0.8361(3)	0.8455
FFFW	0.6075	0.6045	0.6121	0.5214	0.6256	0.6210	0.6252(4)	0.6300
Political blogs	0.9150	0.9124	0.9095	0.8700	0.9155	0.9154	0.9212(3)	0.9213

5.3 算法时间复杂度分析

在设计预测算法时,算法的效率也是考虑的因素之一。CN 算法较为简单,其时间复杂度为 $O(n\langle k \rangle^2)$ ^[33]。明显地,AA 算法、RA 算法和 Jaccard 算法与 CN 算法具有相同的时间复杂度。LP 算法考虑了三阶路径,其时间复杂度为 $O(n\langle k \rangle^3)$;LRW 算法的时间复杂度为 $O(n\langle k \rangle^a)$,其中 a 为随

机行走的步数^[34]。CNID 算法在计算所有节点对之间的相似度之前,首先计算网络中每条边的边聚集系数,其时间复杂度为 $O(n\langle k \rangle^3)$,再由 CN 算法的时间复杂度可知,其时间复杂度为 $O(n\langle k \rangle^2 + n\langle k \rangle^3)$ 。从而可以看出,CNID 算法的时间复杂度相对较低。

结束语 本文考虑到网络中节点对的相似性不仅与其共

同邻居节点的拓扑结构有关,还与节点对和其共同邻居节点之间的局部结构信息有关,利用边的聚集系数来测量节点对与其共同邻居节点之间的亲密度值,在基于共同邻居的 CN 算法的基础上,结合节点对与其共同邻居节点之间的亲密度值,提出一种基于节点亲密度的 CNID 算法。将此算法在真实网络中与其他链路预测算法进行比较,结果表明改进的算法能够提升链路的预测精度。在本研究的基础上,接下来可考虑如何将此方法拓展到加权网络和有向网络中以及在结合节点亲密度的基础上如何更好地利用共同邻居节点的属性,使预测精度有所提升。

参 考 文 献

- [1] LV L Y, LU J A, ZHANG Z K, et al. Looking into Complex Networks[J]. *Complex Systems & Complexity Science*, 2010, 7(2):173-186.
- [2] CANNISTRA 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(4):1613.
- [3] CANNISTRA C V, ALANIS-LOBATO G, RAVASI T. Minimum Curvilinearity to Enhance Topological Prediction of Protein Interactions by Network Embedding[J]. *Bioinformatics*, 2013, 29(13):199-209.
- [4] CRONE S F, SOOPRAMANIEN D. Predicting Customer Online Shopping Adoption—an Evaluation of Data Mining and Market Modelling Approaches[C]// *International Conference on Data Mining, Dmin 2005*. Las Vegas, Nevada, USA, DBLP, 2005:215-221.
- [5] MA C, ZHOU T, ZHANG H F. Playing the Role of Weak Clique Property in Link Prediction; A Friend Recommendation Model[J]. *Scientific Reports*, 2016, 6:30098.
- [6] FU Y B, CHEN Y Z. Relationship Analysis of Microblogging User with Link Prediction[J]. *Computer Science*, 2014, 41(2):201-205. (in Chinese)
傅颖斌, 陈羽中. 基于链路预测的微博用户关系分析[J]. *计算机科学*, 2014, 41(2):201-205.
- [7] CHEN B, CHEN L. A Link Prediction Algorithm Based on Ant Colony Optimization[J]. *Applied Intelligence*, 2014, 41(3):694-708.
- [8] CLAUSET A, MOORE C, NEWMAN M E J. Hierarchical Structure and the Prediction of Missing Links in Networks[J]. *Nature*, 2008, 453(7191):98-101.
- [9] SARUKKAI R R. Link Prediction and Path Analysis Using Markov Chains[J]. *Computer Networks*, 2000, 33(1):377-386.
- [10] LV L Y, ZHOU T. Link Prediction in Complex Networks: A Survey[J]. *Physica A: Statistical Mechanics and its Applications*, 2011, 390(6):1150-1170.
- [11] YANG J X, ZHANG X D. Prediction Missing Links in Complex Networks Based on Common Neighbors and Distance[J]. *Scientific Reports*, 2016, 6:38208.
- [12] CHEN J Y, YU J, YANG X Y, et al. Link Prediction Algorithm Based on Node Importance in Complex Networks[J]. *Journal of Computer Applications*, 2016, 36(12):3251-3255. (in Chinese)
- [13] GAO Y, ZHANG P, QIAN F L, et al. Combined with Node Degree and Node Clustering Coefficient of Link Prediction Algorithm[J]. *Journal of Chinese Computer Systems*, 2017, 38(7):1436-1441. (in Chinese)
高扬, 张平, 钱付兰, 等. 结合节点度和节点聚集系数的链路预测算法[J]. *小型微型计算机系统*, 2017, 38(7):1436-1441.
- [14] LORRAIN F, WHITE H C. Structural Equivalence of Individuals in Social Networks[J]. *The Journal of Mathematical Sociology*, 1971, 1(1):49-80.
- [15] ADAMIC L A, ADAR E. Friends and Neighbors on the Web[J]. *Social Networks*, 2003, 25(3):211-230.
- [16] ZHOU T, LV L Y, ZHANG Y C. Predicting Missing Links via Local Information[J]. *European Physical Journal B*, 2009, 71(4):623-630.
- [17] WU Z H, LIN Y F, WANG J, et al. Link Prediction with Node Clustering Coefficient[J]. *Physica A: Statistical Mechanics and its Applications*, 2016, 45(2):1-8.
- [18] JIA J, HU X F, HE X Y. On the Relationship Between Network Structure Features and Link Prediction Algorithms[J]. *Complex Systems and Complexity Science*, 2017, 14(1):28-37. (in Chinese)
贾珺, 胡晓峰, 贺筱媛. 网络结构特征与链路预测算法关系研究[J]. *复杂系统与复杂性科学*, 2017, 14(1):28-37.
- [19] ZHAO T, LV L, ZHANG Y C. Predicting Missing Links via Local Information[J]. *The European Physical Journal B-Condensed Matter and Complex Systems*, 2009, 71(4):623-630.
- [20] KATZ L. A New Status Index Derived from Sociometric Index[J]. *Psychometrika*, 1953, 18(1):39-43.
- [21] LEICHT E A, HOLME P, NEWMAN M E J. Verter Similarity in Networks[J]. *Physical Review E*, 2006, 73(2):026120-026130.
- [22] ZHU X, TAIN H, CAI S. Predicting Missing Links via Effective Paths[J]. *Physica A: Statistical Mechanics and its Applications*, 2014, 413(11):515-522.
- [23] KLEIN D J, RANDIC M. Resistance Distance[J]. *Journal of Mathematical Chemistry*, 1993, 12(1):81-95.
- [24] BRIN S, PAGE L. The Anatomy of a Large-Scale Hypertextual Web Search Engine[J]. *Computer Network and ISDN Systems*, 1998, 30(1):107-117.
- [25] LIU W P, LV L Y. Link Predicting Based on Local Random Walk[J]. *Europhysics Letters*, 2010, 89(5):58007-58012.
- [26] LIU S, LIU H, CHEN Q M, et al. Link Prediction Algorithm Based on Network Representation Learning and Random Walk[J]. *Journal of Computer Applications*, 2017, 37(8):2234-2239. (in Chinese)
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.
- [27] HU J, YANG B R. Community Structure Discovery Algorithm Based on Edge Clustering Coefficient[J]. *Application Research of Computers*, 2009, 26(3):858-859. (in Chinese)
胡健, 杨炳儒. 基于边聚集系数的社区结构发现算法[J]. *计算机应用研究*, 2009, 26(3):858-859.

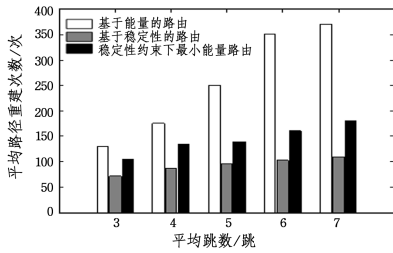


图 10 不同平均跳数下的路由重建次数的比较

Fig. 10 Times of routing reconstruction under different average hop counts

结束语 本文提出一种基于稳定性的能效路由的协议,综合考虑了时延、稳定性、能量三大重要因素。能量消耗主要考虑了节点在传输过程中的能量消耗以及簇首节点在接收和融合其他节点发来的数据所消耗的能量。在稳定性方面,通过采用 ON-OFF 模型对授权用户进行建模,计算链路的稳定性 $P_s(n_i, n_j, c_k)$,通过实例分析得出了稳定性对路由总能量消耗的重要影响。最后,实验结果表明,提出的完整算法模型在能量、稳定性方面明显优于其他两种算法。

参考文献

[1] LEE H R, CHUNG K Y, JHANG K S. A Study of Wireless Sensor Network Routing Protocols for Maintenance Access Hatch Condition Surveillance[J]. *Journal of Information Processing Systems*, 2013, 9(2): 69-78.

[2] TRAGOS E Z, FRAGKIADAKIS A, ASKOXYLAKIS I, et al. The impact of interference on the performance of a multi-path metropolitan wireless mesh network[C]// *IEEE Symposium on Computers and Communications*. IEEE Computer Society, 2011: 199-204.

[3] ARAUJO A, BLESIA J, ROMERO E, et al. Security in cognitive wireless sensor networks. Challenges and open problems [J]. *Eurasip Journal on Wireless Communications & Networking*, 2012, 2012(1): 1-8.

[4] HEINZELMAN W, CHANDRAKASAN A, BALAKRISHNAN H. Energy-Efficient Protocol for Wireless Microsensor Networks[C]// *Hawaii International Conference on System Sciences*. 2000: 3005-3014.

[5] YOUNIS O, FAHMY S. HEED: a hybrid, energy-efficient, dis-

tributed clustering approach for ad hoc sensor networks[J]. *IEEE Transactions on Mobile Computing*, 2004, 3(4): 366-379.

[6] MAO L, ZHANG Y. An energy-efficient LEACH algorithm for wireless sensor networks[C]// *Chinese Control Conference*. 2017: 9005-9009.

[7] ALNAWAFI E, MARGHESCU I. DMHT-LEACH: Dynamic multi-hop technique for wireless sensor networks[C]// *International Symposium on Signals, Circuits and Systems*. IEEE, 2017: 1-4.

[8] LIU J, REN P Y, XUE S L, et al. Primary-user behavior based joint routing and channel allocation algorithm in cognitive radio networks [J]. *Journal on Communications*, 2011, 32(11): 183-190. (in Chinese)

刘婧, 任品毅, 薛少丽, 等. 认知无线网络中基于授权用户行为的联合路由和信道分配算法[J]. *通信学报*, 2011, 32(11): 183-190.

[9] JAEWOO S, SRIKANT R. Improving Channel Utilization via Cooperative Spectrum Sensing With Opportunistic Feedback in Cognitive Radio Networks [J]. *Communications Letters IEEE*, 2015, 6(19): 1065-1068

[10] HUA N, CAO Z G. Routing of cognitive radio networks: A survey [J]. *Acta Electronica Sinica*, 2010, 38(4): 910-918. (in Chinese)

滑楠, 曹志刚. 认知无线网络路由研究综述[J]. *电子学报*, 2010, 38(4): 910-918.

[11] CALEFFI M, AKYILDIZ I, PAURA L. OPERA: Optimal Routing Metric for Cognitive Radio Ad Hoc Networks [J]. *IEEE Transaction on Wireless Communications*, 2012, 8(11): 2884-2894

[12] GROSS D. *Fundamentals of Queuing Theory*[M]. J. Wiley & Sons, 2008.

[13] HARSHA P M, KANAKARAJU R. Network lifetime enhancement of clustering approach using handoff mechanism in WSN [C]// *IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology*. IEEE, 2017.

[14] SEDGEWICK R. *Algorithms in C++: Graph Algorithms* (3rd ed.) [M]. Pearson Education, 2002.

[15] DE E, MARTINS V, PASCOAL M M B. Deviation Algorithms for Ranking Shortest Paths[J]. *International Journal of Foundations of Computer Science*, 1999, 10(3): 247-261.

(上接第 96 页)

[28] FAWCETT T. An Introduction to ROC Analysis[J]. *Pattern Recognition Letters*, 2006, 27(8): 861-874.

[29] NEWMAN M E J. Finding Community Structure in Networks Using the Eigen Vectors of Matrices[J]. *Physical Review E*, 2006, 74(3): 036104-036113.

[30] 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.

[31] ULANOWICZ R E, BONDAVALLI C, EGNOTOVICH M S. Network Analysis of Trophic Dynamics in South Florida Ecosystem[R]. FY 97: The Florida Bay Ecosystem, 1997.

[32] ACKLAND R. Mapping the US Political Blogosphere: Are Conservative Bloggers more Prominent? [C] // *Blog Talk Downunder 2005 Conference*. Sydney, 2005.

[33] DONG Y X, KE Q, WU B. Link Prediction Based on Node Similarity[J]. *Computer Science*, 2011, 18(7): 162-164. (in Chinese)

东昱晓, 柯庆, 吴斌. 基于节点相似性的链接预测[J]. *计算机科学*, 2011, 18(7): 162-164.

[34] LV L Y. Link Prediction on Complex Networks[J]. *Journal of University of Electronic Science and Technology of China*, 2010, 39(5): 651-661. (in Chinese)

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.