



基于资源传输节点拓扑紧密性的链路预测方法

李英乐, 何赞园, 王 凯, 许明艳

(中国人民解放军战略支援部队信息工程大学 信息技术研究所, 郑州 450002)

摘要: 针对现有基于网络拓扑结构的局部相似性RA指标未考虑传输节点拓扑紧密性的问题, 提出一种节点拓扑紧密性指标及链路预测方法。根据多跳节点资源传输情况确定重要传输节点, 基于传输节点周围拓扑集聚程度对拓扑紧密性进行量化, 并根据传输节点紧密性对共同邻居传输资源量的影响刻画节点间相似性。实验结果表明, 该方法具有较高的普适性, 所提相似性指标适合于Precision标准, 与CN、AA和CAR等现有相似性指标相比, 具有较高的预测精度。

关键词: 复杂网络; 链路预测; 资源传输节点; 拓扑紧密性; 传输资源

开放科学(资源服务)标志码(OSID):



中文引用格式: 李英乐, 何赞园, 王凯, 等. 基于资源传输节点拓扑紧密性的链路预测方法[J]. 计算机工程, 2021, 47(1): 50-57.

英文引用格式: LI Yingle, HE Zanyuan, WANG Kai, et al. Link prediction method based on topological tightness of resource transmission nodes[J]. Computer Engineering, 2021, 47(1): 50-57.

Link Prediction Method Based on Topological Tightness of Resource Transmission Nodes

LI Yingle, HE Zanyuan, WANG Kai, XU Mingyan

(Institute of Information Technology, PLA Strategic Support Force Information Engineering University, Zhengzhou 450002, China)

[Abstract] To address the problem that the existing local similarity RA index based on network topology does not consider the topological closeness of transmission nodes, this paper proposes a node topology closeness index and link prediction method. The important transmission nodes are determined according to the resource transmission details of multi-hop nodes, and then the topological closeness is quantified based on the analysis of the topological clustering degree around the transmission nodes. Finally, the similarity between nodes is described by using the impact of the closeness of the transmission nodes on the transmission resources of the common neighbors. Experimental results show that this method has high universality, and the proposed similarity index is suitable for the Precision standard, and has higher prediction accuracy compared with the existing similarity indexes such as CN, AA and CAR.

[Key words] complex network; link prediction; resource transmission nodes; topological tightness; transmission resource

DOI: 10.19678/j.issn.1000-3428.0056092

0 概述

链路预测是网络科学的一个重要研究方向, 其目的是通过分析网络当前的拓扑结构和属性信息, 来预测网络中不存在链接的节点之间出现链接的可能性^[1], 对研究复杂网络的演进机制具有十分重要的意义^[2], 在生物网络^[3]和社交网络^[4]等现实网络实践中也具有很好的应用价值。

近年来, 越来越多的学者对链路预测的方法进

行了深入研究, 并在多个方面都取得了一系列成果, 特别是在基于网络拓扑结构的链路预测方法上的进展较大。基于网络拓扑结构的链路预测方法根据节点之间的链接关系计算节点之间的相似性, 得到节点之间出现链接的可能性。节点相似性指标大体上包括全局相似性指标和局部相似性指标两大类。局部相似性指标主要根据节点的共同邻居来计算相似性, 包括CN指标^[5]、Salton指标^[6]、AA指标^[7]和RA指标^[8]等。全局性指标是从整个网络拓扑结构着眼

基金项目: 国家自然科学基金(61803384)。

作者简介: 李英乐(1985—), 男, 副研究员, 主研方向为复杂网络分析; 何赞园(通信作者)、王 凯、许明艳, 副研究员。

收稿日期: 2019-09-23 **修回日期:** 2019-12-26 **E-mail:** 13513802510@139.com

计算节点相似性,主要有基于随机游走的 Cos+ 指标^[9]、ACT 指标^[10]、RWR 指标^[11]和基于路径的 LP 指标^[12]、LHZ-II 指标^[13]、Katz 指标^[14]。局部性指标仅考虑邻居节点,其计算复杂度和精度较低;全局性相似指标考虑整个网络结构,预测精度较高,但计算复杂度也较高,难以用于大规模网络中。需要注意的是,由于 RA 指标在局部相似性指标中取得相当高的预测精度,因此受到了广泛的关注。但 RA 指标仅从简单的资源传输过程描述出发对其进行量化,忽略了传输节点周围拓扑信息对资源传输节点传输概率的影响^[15-16]。在现实网络中,传输节点周围拓扑的聚集程度与资源传输的效果具有相关性^[17-18],传输节点越聚集,资源的传输效果越好。

针对上述问题,本文提出一种基于资源传输节点拓扑紧密性的链路预测方法。研究资源传输过程中多跳节点资源的传输情况,确定重要传输节点,对传输节点聚类系数进行分析,基于聚类系数改进拓扑紧密性量化方法,并利用传输节点紧密性对共同邻居传输资源进行参数调整,进而重新刻画节点间相似性,给出资源传输节点拓扑紧密性指标。此外,在 Precision 衡量标准下,利用多个实际网络对所提指标进行有效性分析,以验证该方法的预测效果。

1 相关工作

当前基于相似性的链路预测方法有很多,研究人员基于不同原理提出多种预测指标,本节对几种知名预测指标及网络资源传输相关指标简要介绍如下:

1) CN 指标^[5]。LORRAIN 等人直接利用共同邻居的数目衡量节点 x 和 y 的相似度,该方法复杂度低,在集聚性较高的网络中表现较好,但在稀疏网络中表现较差。 $\Gamma(x)$ 为节点邻居的集合,表示为:

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

2) AA 指标^[7]。根据节点对共同邻居的重要程度不同,ADAMIC 等人利用高度惩罚机制对共同邻居加权,权值为节点度对数的倒数,表示为:

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

3) CAR 指标^[17]。在 CN 的基础上考虑了共同邻居之间存在链接的情形,该方法在集聚性较高的网络中表现较好,CAR 指标可表示为:

$$s_{xy}^{CAR} = |\Gamma(x) \cap \Gamma(y)| \cdot \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{|\gamma(z)|}{2} \quad (3)$$

4) LP 指标^[12]。吕林媛等人在共同邻居(即 2 阶路径)基础上,综合考虑了 3 阶路径对节点相似性的影响,LP 指标可表示为:

$$S = A^2 + \alpha \cdot A^3 \quad (4)$$

其中, S 为相似度矩阵, A 是邻接矩阵, $(A^n)_{xy}$ 表示节点 x 和 y 之间长度为 n 的路径数目, α 为调节参数。该方法是预测效果与计算复杂度的合理折中,效果好于 CN 指标。

5) Katz 指标^[14]。KATZ 等人从网络全局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,具有很好的预测精度,但复杂度较高,表示为:

$$s_{xy}^{Katz} = \sum_{l=1}^{\infty} \alpha^l \cdot |\text{paths}_{xy}^{<l>}| = \alpha A_{xy} + \alpha^2 (A^2)_{xy} + \dots + \alpha^n (A^n)_{xy} \quad (5)$$

其中, $\text{path}_{xy}^{<l>}$ 表示节点 x 到 y 之间路径长度为 l 的数目, α 为调节各个路径比例的参数。

6) ACT 指标^[10]。KLEIN 等人将一个随机游走的粒子从一个节点 x 游走至另一个节点 y 所需要走的平均步数,作为两节点产生连接的可能性,提出平均通勤时间指标,表示为:

$$s_{xy}^{ACT} = \frac{1}{L_{xx}^+ + L_{yy}^+ - 2L_{xy}^+} \quad (6)$$

其中, L_{xy}^+ 表示该网络的拉普拉斯矩阵的伪逆 L^+ 中第 x 行第 y 列对应的元素值。

7) Cos+ 指标^[9]。FOUSS 等人在基于随机游走的基础上,提出余弦相似性指标,在矩阵 L^+ 计算 2 个向量之间的相似性:

$$s_{xy}^{Cos+} = \frac{\mathbf{v}_x^T \cdot \mathbf{v}_y}{|\mathbf{v}_x| \cdot |\mathbf{v}_y|} = \frac{L_{xy}^+}{\sqrt{L_{xx}^+ \cdot L_{yy}^+}} \quad (7)$$

8) RA 指标^[8]。周涛等人基于资源传输原理,认为共同邻居传递资源的多少和其路径上共同邻居的节点度成反比,RA 指标具体表示为:

$$s_{xy}^{RA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k_z} \quad (8)$$

9) ERA 指标^[19]。刘树新等人在 RA 指标的基础上,详细分析节点间多跳的局部路径传输的资源,提出了扩展的资源分配 ERA 指标,具体表示为:

$$s_{xy}^{ERA} = R(x \rightarrow y) + R(y \rightarrow x) + R'(x \rightarrow y) + R'(y \rightarrow x) = \sum_{z \in C_{xy}} \frac{2 + \sigma(n_{zy} + n_{zx})}{k_z} + \sum_{z \in C'_{xy}} \frac{\sigma(n_{zy} + n_{zx})}{k_z} \quad (9)$$

其中,参数 σ 表示多跳路径资源对相似性的贡献,当 $\sigma=0$ 时,该指标退化为 RA 指标。

10) EP 指标^[20]。王凯等人考虑资源传输路径有效程度对传输性能的影响,分析了潜在的资源传输路径有效性并对双向资源传输量进行刻画,EP 指标具体表示为:

$$s_{xy}^{EP} = Ep(y \rightarrow x) + Ep(x \rightarrow y) = \sum_{z \in \Gamma(x)} \frac{wl_{yz}^w}{W_z} + \sum_{z \in \Gamma(y)} \frac{wl_{xz}^w}{W_z} \quad (10)$$

其中, W_z 表示节点 z 所有连边的路径有效量化值之和, wl_{ij} 表示节点 i, j 之间的有效资源传输量。

11) QN 指标^[21]。王凯等人在研究路径传输有效性基础上,分析传输路径的资源承载能力,提出资源承载度及相应的预测方案,具体表示为:

$$s_{xy}^{QN} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} (Q^1(x, z) + Q^1(z, y)) \quad (11)$$

其中, $Q^1(i, j)$ 表示节点 i, j 之间量化的资源承载能力。

综上所述,资源传输原理是节点相似性刻画的重要参考,其不仅应从传输过程分析,而且在传输路径、局域拓扑等方面有更多待研究之处。

2 资源传输节点拓扑紧密性链路预测方法

2.1 资源传输节点拓扑紧密性分析与量化

研究发现,网络演进时连边的变化会影响到网络中的信息的传播^[23],同时网络中信息的传播也会影响到连边的变化^[24],此两者具有相关性^[25]。任意2个未连接点之间的资源传输需要经过多跳节点进行,共同邻居节点作为最短跳的资源传播节点,是资源传输过程的核心。图1所示为不同类型资源传输节点在资源传输过程中的作用(图中仅以节点 x 传输一定的资源至节点 y 的过程为例)。

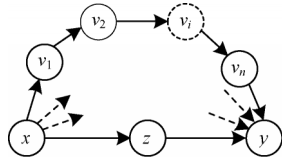


图1 不同类型资源传输节点在资源传输中的作用

Fig.1 Effect of different types of resource transmission nodes in resource transmission

从图1可以看出,在多跳路径传输中,随着路径上节点数目的增加,资源传输过程趋于复杂,资源量的传输“衰落”现象也越来越明显,导致节点 x 到节点 y 的资源传输量也会以一定的比例逐步减少。而仅经过两跳传输的共同邻居节点 z ,其传输资源的效率和确定性明显高于其他传输节点。因此,在资源传输过程中,共同邻居是资源传输的核心节点,发挥着关键作用。在资源传输的后续分析中,将重点针对核心资源传输节点共同邻居。

资源传输指标(RA)是直接利用核心传输节点共同邻居进行资源传输过程分析,然后对任意2个未连接节点进行相似性刻画。该方法效果较好,但现实网络中核心传输节点共同邻居周围拓扑结构较为复杂,仅利用节点度对其进行资源传输量的刻画难以更加有效地量化传输过程。共同邻居节点周围拓扑结构的紧密性会对资源传输有着较大影响,如图2所示。对于节点 x 和节点 y 的共同邻居节点 z ,其周围存在邻居节点 $\{v_1, v_2, v_3, v_4, v_5\}$ 和许多连边构成了相对稠密的局部结构,这对于资源传播具有很大的促进作用。为量化局部结构对节点间资源传输的作用,需要分析刻画传输节点周围拓扑结构的资源传播紧密性。

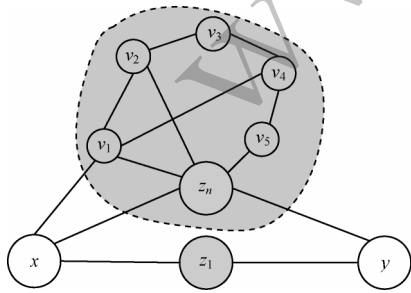


图2 资源传输节点周围拓扑结构示意图

Fig.2 Schematic diagram of topology around resource transmission nodes

网络节点的聚类系数是刻画网络局部拓扑结构紧密性的重要参数^[26]。对于网络中任意一个节点 i ,有 k_i 个邻居节点。若所有邻居节点间两两互联,则所有连边数目表示为 $k_i(k_i-1)/2$,此时网络邻居节点的紧密性最大。在正常情形下,节点 i 的聚类系数可以表示为:

$$C_i = \frac{E_i}{k_i(k_i-1)/2} = \frac{2E_i}{k_i(k_i-1)} \quad (12)$$

其中, E_i 表示节点 i 的所有邻居节点间实际存在连边的数目。节点聚类系数会随着邻居节点连边数目不断增加,如图3所示。图3(a)中邻居节点之间无任何连接,此时节点 i 的拓扑集聚性最低, $C_i=0$;图3(b)中邻居节点之间存在一条连接,此时聚类系数 $C_i=2 \times 1 / (4 \times 3) = 1/6$;图3(c)中邻居节点之间的连接大幅度增加,此时拓扑间关系更加紧密,聚类系数为 $C_i=2 \times 3 / (4 \times 3) = 1/2$;图3(d)中所有邻居节点都存在连接,局部拓扑集聚性达到最大,此时聚类系数为 $C_i=1$ 。可以看出,网络节点聚类系数一定程度上表征了独立节点周围拓扑紧密性。然而,不同于独立节点仅考虑局部拓扑结构,对于一对未连接节点的核心资源传输节点,其资源传输紧密性的刻画需要考虑资源传输节点周围拓扑结构对资源传输过程的影响,而且直接使用聚类系数难以表征不同邻居节点对传输资源紧密性的作用。

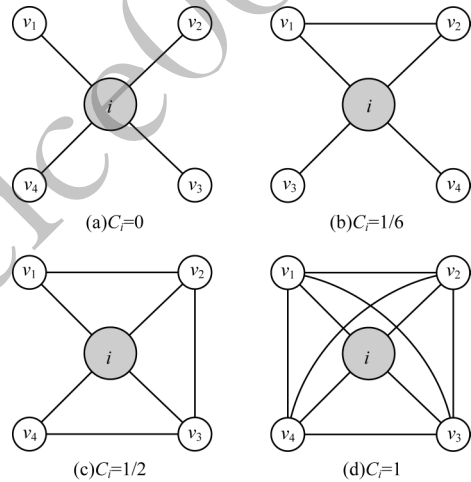


图3 独立节点不同拓扑结构下的紧密性分析

Fig.3 Tightness analysis of independent nodes under different topology

为分析在不同拓扑结构下资源传输节点周围局部拓扑对节点间资源传输的影响,进而刻画资源传输节点拓扑结构的紧密性,本文引入多个局部拓扑结构进行对比分析,如图4中的4个网络结构中均存在一对未连接的节点 x 和 y ,以及待分析的资源传输节点 z_i 。下文将对比分析 z_i 周围不同拓扑结构对于节点 x 和 y 资源传输过程的影响,即对资源传输节点紧密性的影响。对于图4(a)和图4(b),传输节点 z_i 具有相同数目的邻居节点,但后者邻居节点之间连边数目明显高于前者,因此从节点 x 和 y 资源传输的

角度,后者传输节点 z_i 的资源传输紧密性明显高于前者;对于图4(b)和图4(c),传输节点 z_i 既具有相同数目的邻居节点,也具有相同数目的邻居连边,不同的是前者是 $\{v_1, v_3\}$ 和 $\{v_2, v_3\}$,而后者是 $\{v_1, x\}$ 和 $\{v_1, z_1\}$ 。虽然从聚类系数角度,图4(b)和图4(c)的集聚性相同,但是很明显后者 z_i 的邻居节点 v_1 的连边为资源传输提供了更大的可能(对比来看,后者邻居节点 v_1 为节点 x 和 y 提供了更少跳数的稠密拓扑传输结构),因此,从节点 x 和 y 资源传输的角度,图4(c)传输节点 z_i 的资源传输紧密性高于图4(b);对于图4(c)和图4(d),传输节点 z_i 具有相同数目的邻居节点,但后者具有更多邻居连边和有效邻居节点 v_4 。因此,从节点 x 和 y 资源传输的角度,图4(d)传输节点 z_i 的资源传输紧密性高于图4(c)。

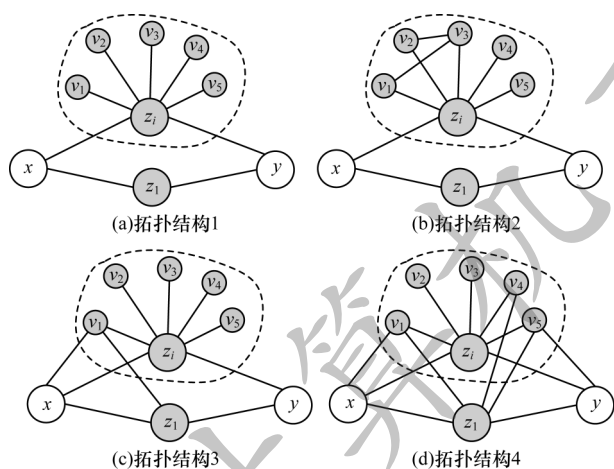


图4 不同资源传输节点周围拓扑结构

Fig.4 Topology around different resource transmission nodes

通过上述分析可以看出,传输节点的邻居在资源传输紧密性的刻画上发挥着不同的作用。对于资源传输始发、终结节点 x 和 y ,若传输节点 z_i 的邻居与其存在连接,则其对于资源传输的贡献较大,如图5中的 v_1, v_2, v_3 。与之相反,若传输节点 z_i 的邻居与始发、终结节点并无连接,其对于资源传输的贡献非常小,考虑到路径上资源的“衰落”,可以忽略其对于节点资源传输紧密性的影响,如图5中的 v_4, v_5 。因此,为从资源传输的角度更好地刻画传输节点的紧密性,对传输节点的邻居中对紧密性贡献较大的邻居进行定义。

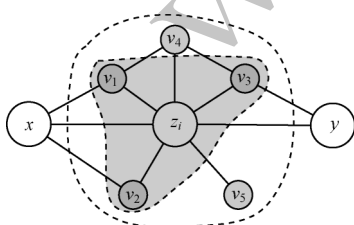


图5 资源传输节点周围拓扑结构示意图

Fig.5 Schematic diagram of topology around resource transmission nodes

定义1(资源传输节点的紧密邻居) 网络中任意一对未连接的资源传输始发、终结节点 x 和 y ,共同邻居节点 z 是其核心传输节点。对于传输节点 z 的所有邻居节点,其紧密邻居包含始发、终结节点 x 和 y ,主要囊括与其存在直接连边的邻居节点。传输节点 z 的紧密邻居集合 V_z^{xy} 可表示为:

$$V_z^{xy} = (\Gamma(z) \cap \Gamma'(x)) \cup (\Gamma(z) \cap \Gamma'(y)) = \Gamma(z) \cap (\Gamma'(x) \cup \Gamma'(y)) \quad (13)$$

其中, $\Gamma(z)$ 为传输节点 z 的所有邻居节点集合, $\Gamma'(x)$ 、 $\Gamma'(y)$ 为包括自身节点 x 、 y 及其所有邻居节点的集合。图6所示为资源传输节点的紧密邻居,对于资源传输始发、终结节点 x 和 y ,其资源传输节点 z 的紧密邻居为:

$$V_z^{xy} = \{x, y, v_1, v_2, v_3\} \quad (14)$$

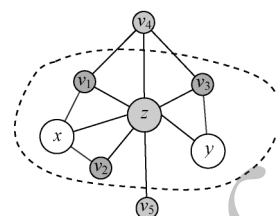


图6 资源传输节点的紧密邻居示意图

Fig.6 Schematic diagram of tight neighbors of resource transmission nodes

对于同样的网络结构,若其资源传输的始发、终结节点发生了变化,其紧密邻居也会发生变化。若图6中资源传输始发、终结节点是 v_4 和 v_5 ,其资源传输节点 z 的紧密邻居为:

$$V_z^{v_4 v_5} = \{v_1, v_3, v_4, v_5\} \quad (15)$$

在对资源传输节点 z 的紧密邻居进行定义后,便可以对其紧密性进行量化,本文基于聚类系数,从紧密邻居的角度定义传输节点的紧密性。

定义2(资源传输节点拓扑紧密性) 网络中任意一对未连接的资源传输始发、终结节点 x 和 y ,共同邻居节点 z 是其核心传输节点。对于传输节点 z ,其资源传输拓扑紧密性量化为紧密邻居间实际连边数目占比紧密邻居最大连边数目,即:

$$TT_z = \frac{\text{紧密邻居间实际连边数目}}{\text{紧密邻居间最大连边数目}} \quad (16)$$

具体表示为:

$$TT_z^{xy} = \frac{E_z^{xy}}{m_z^{xy}(m_z^{xy} - 1)/2} = \frac{2E_z^{xy}}{m_z^{xy}(m_z^{xy} - 1)} \quad (17)$$

其中, E_z^{xy} 为资源传输节点 z 的紧密邻居 V_z^{xy} 之间存在实际连接的数目,而 m_z^{xy} 则是其紧密邻居 V_z^{xy} 的数目。对于图6中的网络结构,以 x 和 y 为始发、终结节点,资源传输节点 z 的紧密性 $EC_z^{xy} = 2 \times 3 / (5 \times 4) = 0.3$ 。

2.2 资源传输节点拓扑紧密性指标

对资源传输节点拓扑紧密性进行量化后,便可以基于拓扑紧密性分析节点之间的相似性,进而预

测任意节点之间存在的可能性。在一般情况下,与资源分配指标RA相似,可以直接利用传输节点拓扑紧密性对资源传播进行加成:

$$s_{xy} = \sum_{z \in I(x) \cap I(y)} \frac{TT_z^{xy}}{k_z} = \sum_{z \in I(x) \cap I(y)} \frac{2E_z^{xy}}{k_z \cdot m_z^{xy} (m_z^{xy} - 1)} \quad (18)$$

式(18)直接利用存在相似度量上的问题,若资源传输节点 z 的紧密邻居之间并无连接,其相似度则量化为0。该量化方法会一定程度上忽视共同邻居节点本身的资源传输能力。因此,在刻画资源传输节点拓扑紧密性指标时,需要同时考虑两个方面:一方面考虑关键资源传输节点本身的资源传输能力;另一方面需要考虑该资源传输节点拓扑紧密性对资源传输能力的影响。如图7所示,若资源从节点 x 传输指定资源到端节点 y ,则该传输过程的资源量与共同邻居节点本身资源传输和其拓扑紧密性对资源传输影响相关。

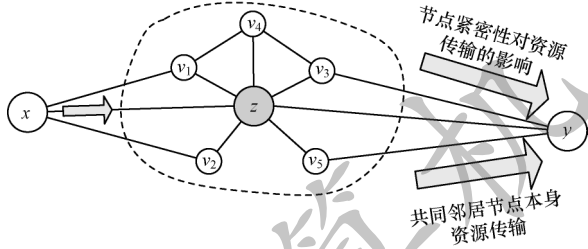


图7 资源传输节点拓扑紧密性指标示意图

Fig.7 Schematic diagram of tightness index of resource transmission node topology

定义3(资源传输节点拓扑紧密性指标) 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 ,从网络中关键资源传输节点的局部拓扑出发,即共同邻居节点拓扑紧密性的角度,节点 x 和 y 之间的相似性可量化为所有共同邻居节点本身传递资源量与其拓扑紧密性加成资源量之和,具体表示为:

$$s_{xy}^{TT} = \sum_{z \in I(x) \cap I(y)} \left(\frac{1}{k_z} + \frac{(TT_z^{xy})^\beta}{k_z} \right) = \sum_{z \in I(x) \cap I(y)} \frac{1 + (TT_z^{xy})^\beta}{k_z} \quad (19)$$

其中, β 为调节参数,用于调整不同网络中资源传输节点拓扑紧密性对资源传输量的影响程度, $1/k_z$ 表示共同邻居节点本身传递资源量,而 $(TT_z^{xy})^\beta/k_z$ 表示拓扑紧密性加成资源量。当 $\beta=0$ 时,资源传输节点拓扑紧密性指标(TT指标)与资源分配指标RA相同。而对于图7所示的拓扑结构中, $s_{xy}^{TT}=(1+0.3^\beta)/7$ 。

3 网络数据集

为了验证基于传输节点拓扑紧密性的链路预测方法的有效性,本文选取了多个不同类型的实际网络进行实验,对具体网络数据分别介绍如下:

1)爵士音乐家合作网络 Jazz^[27]:一种表示爵士音乐家之间的合作关系的网络,其中音乐家用网络节点表示,音乐家之间的合作关系用网络中的连边表示。

2)食物链网络 FWEW^[28]:一种生活在 Everglades Graminoids 湿季的 69 种生物组成的网络,不同生物用网络节点表示,生物之间的捕食关系用网络的边表示。

3)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。

4)线虫神经网络 CE^[30]: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。

5)邮箱通信网络 Email(EM)^[31]:一种规模为中型的工厂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。

6)线虫新陈代谢网络 Metabolic^[32]:秀丽隐杆线虫的新陈代谢网络,节点表示代谢物,而连边表示它们之间的相互作用关系。

7) Infectious(Infect)^[33]: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,人与人之间的沟通用连边表示。

8) mac-animal^[34]:一种 62 个成年雌性日本猕猴的支配行为有向网络,节点表示猕猴,网络连边表示左侧节点对右侧节点的支配行为。

上述网络数据集的具体网络统计特征参数如表1所示。表1主要包含网络节点数目($|V|$)、网络中边的数目($|E|$)、网络平均聚类系数(C)、网络平均节点度 $\langle k \rangle$ 和网络平均最短路径 $\langle d \rangle$ 。在后续网络数据实验中,本文设置训练集合 E^T 中连边数目占比为0.9,测试集 E^P 的连边占比为0.1,此次每个数据测试点的结果均为20次实验的平均值。

表1 实验数据集统计信息

数据集	$ V $	$ E $	$\langle k \rangle$	$\langle d \rangle$	r	C
Jazz	198	2 742	27.70	2.24	0.020	0.633
FWEW	69	880	25.51	1.64	-0.298	0.552
USAir	332	2 128	12.81	2.74	-0.208	0.749
CE	297	2 148	14.46	2.46	-0.225	0.308
Email	167	5 784	69.26	1.87	-0.295	0.541
Metabolic	453	2 025	8.94	2.66	-0.226	0.647
Infectious	410	2 765	13.49	3.63	0.226	0.456
mac-animal	62	1 166	37.61	1.38	-0.073	0.667

4 实验结果与分析

为验证资源传输节点拓扑紧密性指标的有效性,本文在多个不同类型网络中进行实验验证。在实验研究中,由于该指标主要针对 Precision 指标,因此重点对 Precision 结果进行实验。具体地,首先实

验分析了不同网络中调节参数对 TT 的 Precision 结果的影响,然后与现有指标进行对比,分析 TT 的实验效果和有效性,并给出现实网络应用时的合理使用建议值。

4.1 Precision 结果的参数影响分析

资源传输节点拓扑紧密性指标的具体相似性刻画,利用了参数 β 调节不同网络中资源传输节点拓扑紧密性的影响程度。图 8 显示了不同网络中参数 β 对资源传输节点拓扑紧密性指标预测结果的影响。

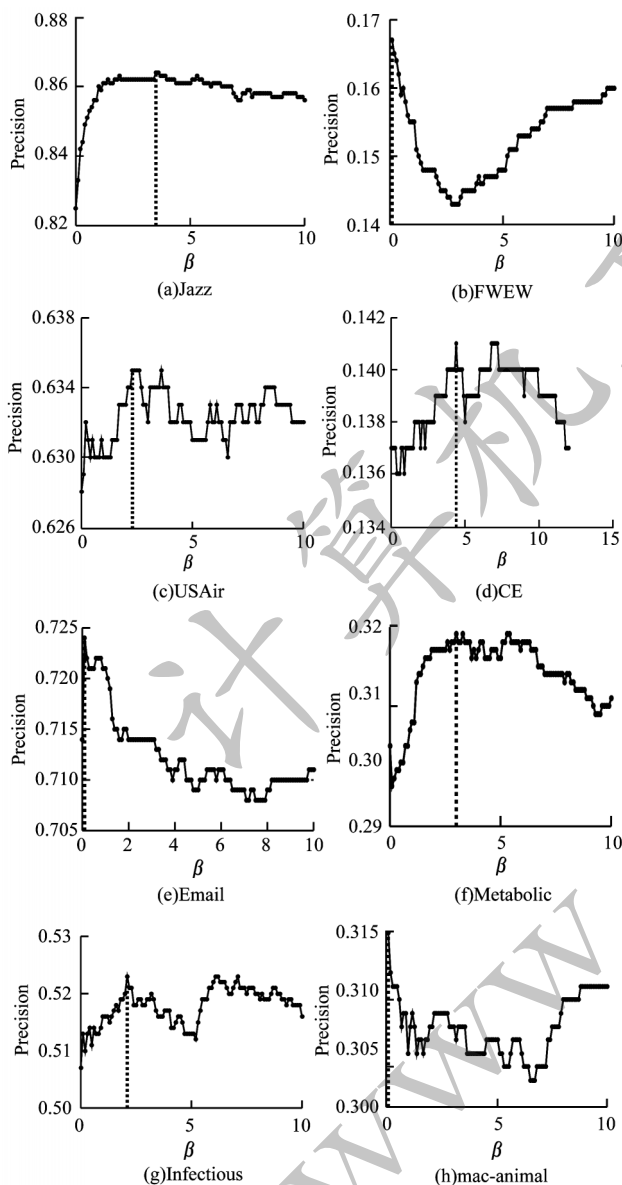


图 8 TT 指标 Precision 结果随参数变化示意图

Fig.8 Schematic diagram of TT index Precision results with parameters changes

可以看出,不同网络中参数 β 对 TT 指标的影响较大,在 Jazz 和 Metabolic 网络中随着 β 值的增大,在 $\beta=0$ 处迅速增长,说明传输节点拓扑紧密性影响非常明显。当 Precision 值达到较高值后,随参数持续稳定,说明此时传输节点拓扑紧密性强度的增大对节

点间相似性的贡献不再明显增大;在 FWEW 和 mac-animal 网络中,参数 β 对 TT 指标的影响较为复杂,随着 β 值的增大, Precision 值逐渐下降到极点后逐步上升,说明较低的拓扑紧密性强度对于相似性刻画促进作用逐渐下降,而到达较大强度后会逐渐增强其影响力; USAir 和 CE 网络相似,在参数 β 较小时其 Precision 值上升至最大值,然后随着参数 β 的增大呈现轻微下降趋势;与其他网络不同,在 Email 网络中,在参数 β 接近 0 时 Precision 值呈现急剧上升趋势,表明较低的传输节点拓扑紧密性强度便可以产生较好的效果。同样,当达到最高点后,随着参数 β 的增大,呈现快速下降趋势,也说明了此时传输节点拓扑紧密性强度对其影响正在逐渐降低;在 Infectious 网络中,当参数 β 较小时 Precision 呈现持续上升态势,到达最大值顶点后逐步下降,而后又呈现上升,说明传输节点拓扑紧密性强度在不同阶段对其影响各异。总之,在多数网络中,当参数 β 较小时,传输节点拓扑紧密性强度对 Precision 结果促进作用较强,部分网络中提升幅度较大且相对稳定,说明传输节点拓扑紧密性对相似度刻画具有一定的有效性。

4.2 Precision 结果对比分析

为进一步研究分析资源传输节点拓扑紧密性指标的有效性和特点,本文选取了 8 个相似性指标进行对比分析。对比指标方法主要包括局部相似性和全局相似性指标,其中局部相似性指标包括共同邻居指标 CN、资源分配指标 RA、AA 指标和考虑共同邻居相互关系的 CAR 指标,全局相似性指标包括全路径指标 Katz、平均通勤时间 ACT 和余弦相似性指标 Cos+,还包括介于局部和全局之间的局部路径指标 LP。

表 2 为多个不同类型网络中 TT 指标与现有指标的 Precision 结果对比情况。其中,(a)可调参数 $\alpha=0.001$, (b)可调参数 $\alpha=0.01$ 。

表 2 Precision 结果对比分析

Table 2 Comparative analysis of Precision results

指标	Jazz	FWEW	USAir	CE	Email	Metabolic	Infectious	mac-animal
CN	0.829	0.136	0.608	0.133	0.714	0.202	0.429	0.310
RA	0.825	0.167	0.628	0.137	0.714	0.305	0.507	0.315
AA	0.842	0.156	0.627	0.137	0.722	0.244	0.452	0.311
CAR	0.851	0.136	0.604	0.131	0.707	0.195	0.403	0.306
LP ^(a)	0.811	0.152	0.612	0.138	0.714	0.201	0.378	0.311
LP ^(b)	0.785	0.193	0.609	0.138	0.710	0.201	0.371	0.314
Katz ^(a)	0.811	0.152	0.612	0.138	0.714	0.201	0.378	0.311
Katz ^(b)	0.753	0.200	0.598	0.138	0.698	0.196	0.364	0.309
ACT	0.000	0.124	0.000	0.000	0.000	0.000	0.000	0.129
Cos+	0.367	0.000	0.097	0.096	0.618	0.098	0.175	0.003
TT	0.864	0.167	0.635	0.141	0.724	0.319	0.523	0.315

从表2可以看出,相比其他的相似性指标,总体上TT指标具有较高的预测精度,8个网络中有7个网络TT指标的Precision值最高(见粗体数字),只有在FWEW网络中排名第三,这在一定程度上说明了资源节点拓扑紧密性对节点间相似性刻画的有效性。作为最简单的共同邻居指标CN,由于仅考虑了共同邻居的数目,其Precision预测结果具有一定的效果,但相比其他相似性方法,大多数网络中CN的预测结果略低(除在Email、Metabolic和Infectious网络中,CN的Precision高于CAR、LP、Katz)。资源分配指标RA通过利用资源分配过程计算节点间传输的资源量进行量化节点间相似性,其复杂度较低但效果明显好于一般局部相似性指标,尤其是在FWEW、USAir、CE、Metabolic、Infectious和mac-animal等网络中,部分甚至高于全局相似性指标LP或Katz。与RA指标类似,AA指标利用共同邻居的节点度对共同邻居进行加权,其效果在多数网络中好于共同邻居指标,部分网络如Jazz、USAir、CE、Email、Metabolic和Infectious中其结果好于全局指标。CAR指标考虑了共同邻居之间存在的部分连接对相似性的影响,其效果明显好于共同邻居指标,但在多数网络中比RA和AA效果略差。LP和Katz是基于长路径的相似性指标,在多数网络中效果较好且相对稳定,尤其是在FWEW、CE、Email和mac-animal网络中效果最好。平均通勤时间ACT指标是基于随机游走的相似性指标,不过在Precision衡量标准下,多数网络中普遍较低,许多Precision值甚至接近于0,这可能是与ACT指标更多地适用于AUC衡量标准,而不适用于Precision标准。余弦相似性指标Cos+同样是基于随机游走的相似性指标,相比ACT其效果有了一定的提升,但与其他相似性指标相比,其总体效果较差。ACT和Cos+的Precision效果分析结果表明,部分基于随机游走的相似性指标更多地倾向于AUC衡量标准,难以适应于Precision衡量标准。在考虑了资源传输节点拓扑紧密性后,TT指标在8个网络中表现普遍较好。多个网络中如在Jazz、USAir、Email和Infectious中,其Precision的提高幅度较高,明显高于其他局部和全局相似性指标。在mac-animal中,TT指标与RA相似,说明了当前网络中资源传输节点拓扑紧密性的影响较低,其设置参数为0。在现实网络的链路预测应用中,建议参数设置为11.1,各个网络中表现普遍较高。此外,本文所提方法时间复杂度介于共同邻居指标CN和局部路径指标LP之间,复杂度相对较低,可以应用于大规模复杂网络链路预测。

综上所述,本文方法能够同时在2个衡量指标AUC和Precision下取得较好的效果。AUC和Precision作为链路预测2个重要但角度不同的衡量标准,在一定程度上说明了本文方法的普适性较高。

此外,本文方法在8个不同网络的实验预测效果相对较高,也验证了其在不同网络数据集上的普遍应用价值。

5 结束语

本文根据传输节点周围拓扑紧密性对资源传输节点传输概率的影响,结合节点拓扑紧密性的特点,提出一种基于资源传输节点拓扑紧密性的链路预测方法。从资源传输过程分析重要的传输节点,重点研究传输节点周围拓扑关系的量化问题,以节点聚类系数为突破点,通过分析当前聚类系数在刻画资源传输节点有效性中存在的问题,提出相应的资源传输紧密性量化方法,并基于资源传输紧密性,对任意未连边节点间的资源传输量进行刻画,给出资源传输节点拓扑紧密性指标。在多个实际网络数据中的实验结果表明,本文相似性指标适合于Precision标准,且相比其他指标具有较好的预测效果。本文方法和思路可以为资源传输过程的刻画提供一种新思路 and 借鉴。在后续研究中,将通过探讨传输路径有效性对资源传输过程的影响,进一步丰富和拓展基于资源传输过程的链路预测方法。

参考文献

- [1] LYU L, ZHOU T. Link prediction in complex networks: a survey [J]. *Physica A: Statistical Mechanics and Its Applications*, 2011, 390(6): 1150-1170.
- [2] LIU Hongkun, LÜ Linyuan, ZHOU Tao. Uncovering the network evolution mechanism by link prediction [J]. *Science in China: Physica, Mechanica & Astronomica*, 2011, 41(7): 816-823. (in Chinese)
刘宏鲲, 吕琳媛, 周涛. 利用链路预测推断网络演化机制[J]. *中国科学: 物理学 力学 天文学*, 2011, 41(7): 816-823.
- [3] YU H, BRAUN P, YILDIRIM M A, et al. High-quality binary protein interaction map of the yeast interactome network [J]. *Science*, 2008, 322(5898): 104-110.
- [4] YUAN Weiwei, HE Kangya, GUAN Danghai, et al. Graph kernel based link prediction for signed social networks [J]. *Information Fusion*, 2019, 46: 1-10.
- [5] LORRAIN F, WHITE H C. Structural equivalence of individuals in social networks [J]. *The Journal of Mathematical Sociology*, 1971, 1(1): 49-80.
- [6] SALTON G, MCGILL M J. Introduction to modern information retrieval [M]. Auckland, New Zealand: McGraw-Hill, 1983.
- [7] ADAMIC L A, ADAR E. Friends and neighbors on the Web [J]. *Social Networks*, 2003, 25(3): 211-230.
- [8] ZHOU Tao, LÜ Linyuan, ZHANG Yicheng. Predicting missing links via local information [J]. *The European Physical Journal B*, 2009, 71(4): 623-630.
- [9] FOUSS F, PIROTTE A, RENDERS J M, et al. Randomwalk 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.
- [10] KLEIN D J, RANDI M. Resistance distance[J]. *Journal of Mathematical Chemistry*, 1993, 12(1):81-95.
- [11] BRIN S, PAGE L. The anatomy of a large-scale hypertextual Web search engine[J]. *Computer Networks and ISDN Systems*, 1998, 30(1):107-117.
- [12] LÜ Linyuan, JIN Cihang, ZHOU Tao. Similarity index based on local paths for link prediction of complex networks[J]. *Physical Review E*, 2009, 80(4):46-62.
- [13] LEICHT E A, HILME P, NEWMAN M E J. Vertex similarity in networks[J]. *Physical Review E*, 2006, 73(2):61-80.
- [14] KATZ L. A new status index derived from sociometric analysis[J]. *Psychometrika*, 1953, 18(1):39-43.
- [15] TAN Fei, XIA Yongxiang, ZHU Boyao. Link prediction in complex networks; a mutual information perspective[J]. *PloS One*, 2014, 9(9):107-126.
- [16] WU Zhihao, LIN Youfang, WANG Jing, et al. Link prediction with node clustering coefficient[J]. *Physica A; Statistical Mechanics and Its Applications*, 2016, 452(15):1-8.
- [17] LIU Shuxin, JI Xinsheng, LIU Caixia, et al. Similarity index for link prediction based on the clustering coefficient of nodes[J]. *Journal of Information Engineering University*, 2017, 18(6):698-702. (in Chinese)
刘树新, 季新生, 刘彩霞, 等. 一种基于节点集聚程度的相似性链路预测方法[J]. *信息工程大学学报*, 2017, 18(6):698-702.
- [18] DING Dazhao, CHEN Yunjie, JIN Yanqing, et al. Link prediction method for complex network based on closeness between nodes[J]. *Journal of Computer Applications*, 2017, 37(8):2129-2132, 2138. (in Chinese)
丁大钊, 陈云杰, 靳彦青, 等. 基于拓扑连接紧密度的相似性链路预测算法[J]. *计算机应用*, 2017, 37(8):2129-2132, 2138.
- [19] LIU Shuxin, JI Xisheng, LIU Caixia, et al. Extended resource allocation index for link prediction of complex network[J]. *Physica A; Statistical Mechanics and Its Applications*, 2017, 479(2):174-183.
- [20] WANG Kai, LI Xing, LAN Julong, et al. A new link prediction method for complex networks based on topological effectiveness of resource transmission paths[J]. *Journal of Electronics and Information Technology* 2020, 42(3):653-660.
王凯, 李星, 兰巨龙, 等. 一种基于资源传输路径拓扑有效性的链路预测方法[J]. *电子与信息学报*, 2020, 42(3):653-660.
- [21] WANG Kai, LIU Shuxin, CHEN Hongchang, 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. (in Chinese)
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.
- [22] 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, 4(3):16-23.
- [23] IRIBARREN J L, MORO E. Impact of human activity patterns on the dynamics of information diffusion[J]. *Physical Review Letters*, 2009, 103(3):38-47.
- [24] WENG L, RATKIEWICZ J, PERRA N, et al. The role of information diffusion in the evolution of social networks[C]//*Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA: ACM Press, 2013:356-364.
- [25] GRINDROD P, HIGHAM D J. Evolving graphs: dynamical models, inverse problems and propagation[J]. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2009, 466(2115):753-770.
- [26] VAN ECK P S, JAGER W, LEEFLANG P S H. Opinion leaders' role in innovation diffusion: a simulation study[J]. *Journal of Product Innovation Management*, 2011, 28(2):187-203.
- [27] GLEISER P M, DANON L. Community structure in jazz[J]. *Advances in Complex Systems*, 2003, 6(4):565-573.
- [28] ULANOWICZ R E, DEANGELIS D L. Network analysis of trophic dynamics in south florida ecosystems [J]. *Florida Bay Ecosystem*, 1998, 55:20688-20738.
- [29] BATAGELJ V, MRVAR A. Pajek-program for large network analysis[J]. *Connections*, 1998, 21(2):47-57.
- [30] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998, 393(6684):440-442.
- [31] GUIMERA R, DANOD L, DIAZ-GUILEAR A, et al. Self-similar community structure in a network of human interactions[J]. *Physical Review E*, 2003, 68(6):65-73.
- [32] GUIMERA R, AMARAL L A N. Functional cartography of complex metabolic networks[J]. *Nature*, 2005, 433(7028):895-906.
- [33] 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.
- [34] TAKAHATA Y. Diachronic changes in the dominance relations of adult female Japanese monkeys of the Arashiyama B group[M]. Albany, USA: State University of New York Press, 1991.