

结合邻居影响和资源分配的链路预测算法

刘英杰¹, 刘士虎¹, 高海燕¹, 徐伟华²

(1. 云南民族大学 数学与计算机科学学院 云南 昆明 650504; 2. 西南大学 人工智能学院 重庆 400715)

摘要:传统的资源分配算法在进行链路预测时只考虑了共同邻居对资源分配的影响,却忽略了两个节点之间的二级节点对资源分配的影响。为了探究共同邻居和二级节点对预测两个节点之间连边存在可能性的影响,提出了一种结合邻居影响和资源分配的链路预测算法。实验结果表明,在多数网络中若两个节点接收更多由共同邻居分配给它们的资源,则可以提升预测它们之间连边存在可能性的准确率。在四个生态网络和一些平均聚类系数较低的网络中,若两个节点接收更多通过二级节点分配给它们的资源,则可以提升预测它们之间连边存在可能性的准确率。

关键词:共同邻居; 复杂网络; 资源分配; 资源传输; 链路预测; 二级节点

中图分类号: TP183

文献标志码: A

文章编号: 1671-6841(2024)01-0053-07

DOI: 10.13705/j.issn.1671-6841.2022220

Link Prediction Algorithm Combining Neighbor Influence and Resource Allocation

LIU Yingjie¹, LIU Shihu¹, GAO Haiyan¹, XU Weihua²

(1. School of Mathematics and Computer Science, Yunnan Minzu University, Kunming 650504, China;

2. School of Artificial Intelligence, Southwest University, Chongqing 400715, China)

Abstract: The traditional resource allocation algorithm only focused on common neighbors allocation of the resources in link prediction. Unfortunately, it ignored the application of second-level nodes between two nodes during the resource allocation. To investigate the impact of common neighbors and second-level nodes on predicting the potential existence of edges between two nodes, a link prediction algorithm combining neighbor influence and resource allocation was proposed, to explore the influence of common neighbors and second-level nodes on the existence probability of an edge between two nodes. The experimental results demonstrated that if two nodes received more resources allocated by common neighbors, the accuracy of predicting the possibility of an edge between them could be enhanced in most networks. In four ecological networks and some networks with a low average clustering coefficient, if two nodes received more resources allocated by second-level nodes, the accuracy of predicting the existence possibility of an edge between them could be enhanced.

Key words: common neighbor; complex network; resource allocation; resource transmission; link prediction; second-level node

0 引言

现实中的社交网络^[1]、生物网络^[2]和机会网

络^[3]等复杂网络不仅可以用来描述节点的属性信息,还能用于刻画节点之间的拓扑信息。因此,复杂网络因其潜在的数据信息而引起各领域学者的广泛关注。复杂网络中的研究分支主要有链路预测^[4]、

收稿日期:2022-07-30

基金项目:国家自然科学基金项目(61966039);云南省“兴滇英才支持计划”青年人才专项项目。

第一作者:刘英杰(1997—),男,硕士研究生,主要从事链路预测研究,E-mail:lp_lyj2021@163.com。

通信作者:刘士虎(1984—),男,副教授,主要从事网络数据挖掘研究,E-mail:gdmsh@163.com。

社团检测^[5]和聚类分析^[6]。其中,链路预测旨在探测节点之间存在连边的可能性。因此,链路预测具有巨大的实际应用价值。例如,在电子商务网络中使用链路预测技术为顾客推荐其喜爱的产品^[7]。在犯罪网络中使用链路预测技术揭示罪犯之间隐而未现的关联关系,协助警方及时侦破案件^[8]。

针对现有的链路预测问题,学者们主要采用基于节点相似性的算法来进行研究,这类算法主要分为基于局部结构的算法、基于路径信息的算法和基于随机游走的算法。

在基于局部结构的算法中,资源分配(resource allocation, *RA*)指标^[9]和聚类系数(clustering coefficient for link prediction, *CCLP*)指标^[10]分别在共同邻居的基础上考虑了共同邻居的度和聚类系数的影响;局部社区范式(cannistraci-resource-allocation, *CRA*)指标^[11]在共同邻居的基础上考虑了共同邻居之间的连边数的影响。在基于路径信息的算法中,*Katz*指标^[12]是基于两个节点之间的全路径信息的,局部路径(local path, *LP*)指标^[13]是基于两个节点之间所有二阶路径与部分三阶路径信息的。在基于随机游走的算法中,平均通勤时间(average commute time, *ACT*)指标^[14]基于网络中的全局随机游走而被提出。

除了上述三类算法,研究者们也提出了一系列具有代表性的算法和指标。例如,偏好连接(preferential attachment, *PA*)指标^[15]基于节点之间连边生长的概率与节点度之积成正比的关系而被提出。矩阵森林(matrix-forest index, *MFI*)指标^[16]利用矩阵森林理论来刻画节点相似性而被提出。*CCLP2*指标^[17]基于二级共同邻居的聚类系数而被提出。此外,二级节点的度也被相关学者考虑在链路预测算法的设计中^[18],但他们没有在拓扑性质不同的网络中对共同邻居和二级节点的影响深入分析。

值得一提的是,现有的链路预测研究还源自于网络演化中的资源传输过程^[19-20]。其基本假设是两个节点之间存在资源传输,则它们之间的关系是亲密的,从而它们就更有可能相似。李英乐等^[21]在*RA*指标的基础上,利用共同邻居周围的三角结构来量化其分配给两个节点的资源量,提出了拓扑紧密性(topological tightness, *TT*)指标。然而,*TT*指标只关注共同邻居对资源的分配,却忽略了二级节点在资源分配中的作用。

基于上述分析,为进一步探究共同邻居和二级节点在链路预测过程中的影响,本文提出了一种结合邻居影响和资源分配的链路预测算法。借助无向

网络中节点之间的双向资源传输,该算法利用平衡系数来调节共同邻居和二级节点分配给两个节点的资源总量所占的比例,以探究二者在拓扑性质不同的网络中对链路预测准确率的影响。实验结果表明,在平均聚类系数高的网络中,若两个节点接收更多由共同邻居分配的资源,则可以提升链路预测的准确率;在生态网络和平均聚类系数低的网络中,若两个节点接收更多由二级节点分配的资源,则可以提升链路预测的准确率;此外,结合共同邻居和二级节点的共同影响对链路预测算法准确率的提升将更有效。

1 预备知识

1.1 复杂网络

形式上,无向复杂网络可用一个二元组 $G=(V, E)$ 来表示, V 表示节点集合; E 表示连接任意两个节点的连边集合,即 E 中包含了 $G=(V, E)$ 中的实际存在的边。在网络 $G=(V, E)$ 中,若其内部的任意一个节点都与其余的 $n-1$ 个节点相连,则 $G=(V, E)$ 中共有 $n \cdot (n-1)/2$ 条可能存在的边,使用集合 U 来表示。于是,在 $G=(V, E)$ 中的所有未知连边可用集合 $U-E$ 来表示。其中,未知连边是指 $G=(V, E)$ 中可能存在但尚未被探测的边。

1.2 链路预测

在进行链路预测时,按照一定的比例 $b(0 < b < 1)$ 随机划分 E 中的边为训练集 E^t 和测试集 E^p ,则 E^t 和 E^p 中分别有 $b \cdot m$ 条和 $(1-b) \cdot m$ 条边,且 $E^t \cup E^p = E, E^t \cap E^p = \emptyset$ 。 E^t 中的边被用于链路预测算法中计算其相似性分数, E^p 中的边被用于和 $U-E$ 中的边进行分数比较以测试算法的性能。

基于以上叙述,链路预测问题可被描述如下:对于 $G=(V, E)$ 中任意一条边 e ,根据给定的链路预测算法为边 e 赋予一个相似性分数,然后将 $G=(V, E)$ 中的所有可能的边都按照为它们赋予的分数进行降序排列,那么排名越靠前的连边存在的可能性就越大。

2 资源分配驱动的链路预测

资源分配驱动的链路预测的基本原理是若两个节点之间存在资源交互,则它们的关系较为亲密,故它们更加相似,从而它们之间更有可能存在连边。

在无向网络中,节点之间的资源交互是指始发节点和终结节点可以相互向对方传输一定的资源。

在这个过程中,始发节点与终结节点之间的逐阶路径上的中间节点便扮演着资源承载者的角色^[21]。一般地,两个节点之间的共同邻居就是它们之间的二阶路径上的中间节点,两个节点之间的二级节点就是它们之间的三阶路径上的中间节点。下面将分别介绍两个节点之间传输的资源由共同邻居和二级节点分配时的情况。

2.1 基于共同邻居的资源分配

图1分别给出了节点 v_x 和 v_y 之间传输的资源经由共同邻居 v_z 进行分配时的三种情况。图1(a)表示节点 v_x 发送的资源经由共同邻居 v_z 平均分配给节点 v_y 的情况;图1(b)表示节点 v_y 发送的资源经由共同邻居 v_z 平均分配给节点 v_x 的情况;图1(c)表示共同邻居 v_z 自身具有的资源被平均分配给节点 v_x 和 v_y 的情况。在最简单的情况下,假设节点 v_x 和 v_y 均向对方发送一个单位的资源被共同邻居 v_z 分配,而共同邻居 v_z 的聚类系数 cc_z 被假设为它自身影响两个节点的相似性时所具有的资源量^[19-20]。因此,图1中节点 v_x 和 v_y 在经过了双向的资源传输后能够接收 $1/k_z + 1/k_z + cc_z/k_z = 0.2 + 0.2 + 0.04 = 0.44$ 单位的资源总量,其中 k_z 为节点 v_z 的度。

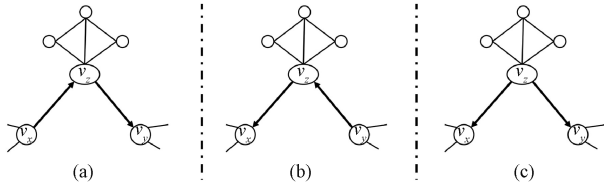


图1 资源由共同邻居分配

Figure 1 Resource allocated by common neighbor

2.2 基于二级节点的资源分配

图2给出了节点 v_x 和 v_y 进行资源传输的过程中,资源在经由二级节点 v_z 被分配到节点 v_x 和 v_y 时的三种情况。图2(a)表示节点 v_x 发送的资源经由二级节点 v_z 平均分配给节点 v_y 的情况;图2(b)表示节点 v_y 发送的资源经由二级节点 v_z 平均分配给节点 v_x 的情况;图2(c)表示二级节点 v_z 自身所具有的资源被分配给节点 v_x 和 v_y 的情况。同样地,假设节点 v_x 和 v_y 向对方发送一个单位的资源被二级节点 v_z 分配,而二级节点 v_z 的聚类系数 cc_z 被假设为它自身影响节点 v_x 和 v_y 的相似性时所具有的资源量。

不难发现,节点之间的资源由二级节点进行分配后也会经过它们的共同邻居,但探究共同邻居和二级节点在预测过程中的影响时需要控制单一变量,故对由二级节点分配给两个节点的资源会不会

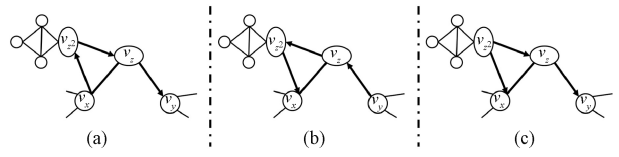


图2 资源由二级节点分配

Figure 2 Resource allocated by second-level node

因为共同邻居的影响而被稀释掉的问题,将在后续的工作中进一步讨论。本次研究在考虑资源由二级节点进行分配时,不再考虑共同邻居对资源分配的影响。以图2为例,节点 v_x 和 v_y 经过二级节点进行资源的双向传输后能接收 $1/k_z + 1/k_z + cc_z/k_z \approx 0.25 + 0.25 + 0.042 = 0.542$ 单位的资源总量。

2.3 邻居影响下的资源分配链路预测指标

本研究利用共同邻居和二级节点分配给两个节点的资源总量来探究它们在链路预测中的影响,并提出了一种结合邻居影响和资源分配(neighbor influence and resource allocation, NIRA)的链路预测指标。因此,考虑无向网络 $G = (V, E)$ 中的任意两个节点 v_x 和 v_y , v_z 和 v_z 分别是它们的共同邻居和二级节点,则NIRA的计算表达式可定义为

$$S_{xy}^{NIRA} = \lambda \cdot \sum_{v_z \in N(v_x) \cap N(v_y)} \frac{2 + cc_z}{k_z} + (1 - \lambda) \cdot \sum_{v_z \in SN(v_x, v_y)} \frac{2 + cc_z}{k_z^2}$$

其中: $N(v_x) \cap N(v_y)$ 和 $SN(v_x, v_y)$ 分别表示 v_x 和 v_y 的共同邻居集合和二级节点集合,且 $SN(v_x, v_y) = [N(N(v_x)) \cap N(v_y)] \cup [N(v_x) \cap N(N(v_y))]$,不包括 v_x 和 v_y 的共同邻居; $\lambda \in [0, 1]$ 是一个平衡系数,用于调节由共同邻居和二级节点分配给两个节点的资源比例。当 $\lambda = 0$ 时,资源只由二级节点分配给两个节点;当 $\lambda = 1$ 时,资源只由共同邻居分配给两个节点;当 $\lambda \in (0, 1)$ 时,资源由共同邻居和二级节点一起分配给两个节点。

3 实验设置

本文的实验环境为16 GB内存,1.8 GHz速度,Windows 10操作系统和AMD Ryzen 7 4800U处理器,编程语言为MATLAB 2020a。

本文使用CCLP指标、CCLP2指标、RA指标、CRA指标、PA指标、LP指标、Katz指标、MFI指标和ACT指标作为实验的基准指标,为适用本文中的二级节点集合,CCLP2指标的计算公式也被进行了相应的扩展。下面介绍本文使用的网络数据和评价指标。

3.1 网络数据

本文从公开的网络数据的学术网站上(<http://konect.cc/networks/>)收集了16个网络数据,这些网络的简称分别为LESM网络、USAI网络、NSSM网络、KING网络、STMA网络、FWEW网络、FWMW网络、FWFW网络、ADJN网络、FORU网络、EMAI网络、UCSO网络、KARA网络、POLB网络、FOOT网络和COMM网络,它们的拓补性质被列在了表1中,其中: n 为节点数; m 为边数; d 为网络直径; ρ 为网络密度; $\langle k \rangle$ 为平均度; $\langle c \rangle$ 为平均聚类系数。

表1 网络的拓补性质

Table 1 Topological properties of networks

网络名	n	m	d	ρ	$\langle k \rangle$	$\langle c \rangle$
LESM	77	254	5	0.087	6.597	0.736
USAI	332	2126	6	0.039	12.807	0.749
NSSM	379	914	17	0.013	4.823	0.798
KING	1773	9131	4	0.006	10.300	0.721
STMA	54	350	3	0.245	12.963	0.413
FWEW	69	880	3	0.375	25.507	0.552
FWMW	97	1446	3	0.311	29.814	0.468
FWFW	128	2075	3	0.254	32.413	0.336
ADJN	112	425	5	0.068	7.589	0.190
FORU	899	7036	6	0.017	15.653	0.071
EMAI	1133	5451	8	0.009	9.622	0.254
UCSO	1899	13838	8	0.008	14.574	0.138
KARA	34	78	5	0.139	4.588	0.588
POLB	105	441	7	0.081	8.400	0.488
FOOT	115	613	5	0.094	10.661	0.403
COMM	167	3250	5	0.235	38.922	0.686

3.2 评价指标

本文使用受试者工作特征曲线下的面积(area under curve, AUC)值和精确度($Precision$)进行评价,它们的定义分别如下。

AUC 值为随机地从测试集中选择一条边的相似性分数比从不存在边集中选择一条边的相似性分数更高的概率。若在 t 次的分数比较中,有 t_1 次是测试边的分数大于不存在边的分数,有 t_2 次是二者分数相等,则 AUC 的值被定义为

$$AUC = \frac{2t_1 + t_2}{2t},$$

其中: t 被设置为10000,衡量算法的性能。

$Precision$ 为排名靠前的边被准确预测的比例,即在排名前 L 的边中,若有 l 条边是测试集中的边,则 $Precision$ 计算为

$$Precision = l/L,$$

其中: L 被设置为100,衡量算法的性能。

4 算法验证及结果分析

为探究 $NIRA$ 指标的预测性能以及适用范围,本文分析了 $NIRA$ 指标与基准指标在多个网络中的 AUC 值和 $Precision$ 值。下面先分析 $NIRA$ 指标的可行性。

4.1 可行性分析

图3为 $NIRA$ 指标在16个网络中随着平衡系数 λ 的变化而生成的 AUC 曲线和 $Precision$ 曲线。由图3(a)可知 $NIRA$ 指标的 AUC 曲线具有以下两个特点。

当 $\lambda \in (0,1)$ 时, $NIRA$ 指标的 AUC 值在所有网络中都达到了一个峰值点。这表明当共同邻居和二级节点一起分配资源给两端节点时, $NIRA$ 指标可以取得最优 AUC 值。在共同邻居和二级节点的共同影响下, $NIRA$ 指标的 AUC 值达到了最优。

$NIRA$ 指标的 AUC 曲线在 $\lambda=0$ 和 $\lambda=1$ 处存在着较大的差异,在拓补性质不同的网络中表现得更为明显。如在LESM、USAI、NSSM和KING中, $NIRA$ 指标在 $\lambda=1$ 时的 AUC 值远远大于其在 $\lambda=0$ 时的 AUC 值,而在ADJN、FORU、EMAI、UCSO、STMA、FWEW和FWFW中,又呈现出相反的情况。这表明,在平均聚类系数较高的网络中,共同邻居对 $NIRA$ 指标的 AUC 影响大于二级节点;然而二级节点在平均聚类系数较低的网络和部分生态网络中对 $NIRA$ 指标的 AUC 影响大于共同邻居。此外,在FWMW、POLB、FOOT和COMM等网络中,共同邻居的影响在提升 $NIRA$ 指标的 AUC 上比二级节点更加具有优势。

由图3(b)可看出 $NIRA$ 指标的 $Precision$ 曲线与其 AUC 曲线具有相同的变化趋势。即在LESM、USAI、NSSM、KING这4个平均聚类系数较高的网络和POLB、FOOT、COMM等网络中,共同邻居对 $NIRA$ 指标的 $Precision$ 的影响大于二级节点;在STMA、FWEW、FWMW、FWFW这4个生态网络和ADJN、FORU、UCSO等三个平均聚类系数较低的网络中,二级节点对 $NIRA$ 指标的 $Precision$ 的影响更大。然而,在EMAI网络中,共同邻居对 $NIRA$ 指标的 $Precision$ 的提升更有优势。此外, $NIRA$ 指标在NSSM、KING和FOOT网络中的最优 $Precision$ 值均在 $\lambda=1$ 处。因此,在这三个网络中,共同邻居对相应算法的 $Precision$ 的提升更有优势。这表明,在多数网络中考虑共同邻居的影响来提升算法的 $Precision$ 是可行的。上述的实验结果表明,从资源分配的角度描述

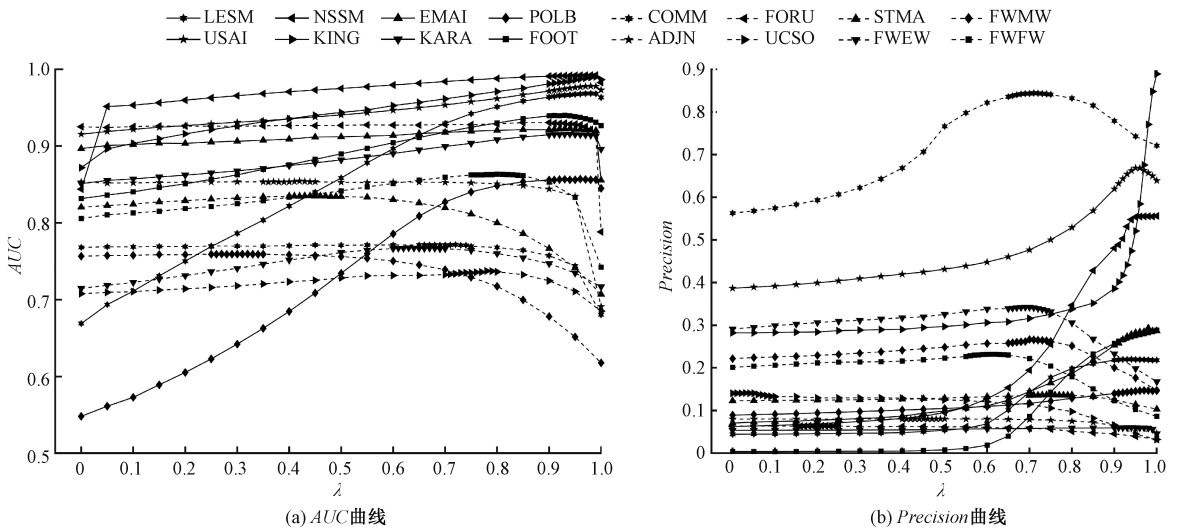


图 3 NIRA 指标在平衡系数 λ 影响下的 AUC 和 Precision 曲线

Figure 3 The AUC and Precision curves of NIRA index with the influence of balance parameter λ

共同邻居和二级节点在链路预测中的影响是可行的。

AUC 值和 Precision 值,表中每一行的加粗数字为在该网络中最高值。

4.2 有效性分析

表 2 和表 3 为 NIRA 指标与基准指标之间的

表 2 NIRA 指标与 9 个基准指标之间的 AUC 值比较

Table 2 Comparison of the AUC values between the NIRA index and nine benchmark indices

网络名	CCLP	CCLP2	RA	CRA	PA	LP	Katz	MFI	ACT	NIRA
LESM	0.951	0.669	0.962	0.915	0.832	0.952	0.953	0.928	0.883	0.968
USAI	0.957	0.886	0.972	0.934	0.911	0.953	0.957	0.941	0.901	0.978
NSSM	0.949	0.841	0.982	0.824	0.660	0.986	0.986	0.987	0.933	0.992
KING	0.971	0.860	0.985	0.903	0.824	0.977	0.977	0.972	0.805	0.990
STMA	0.690	0.710	0.669	0.663	0.733	0.683	0.683	0.676	0.710	0.737
FWEW	0.702	0.822	0.706	0.693	0.814	0.706	0.705	0.702	0.780	0.835
FWMW	0.722	0.730	0.712	0.725	0.779	0.720	0.720	0.744	0.765	0.768
FWWF	0.636	0.762	0.616	0.625	0.732	0.624	0.625	0.703	0.725	0.760
ADJN	0.689	0.756	0.678	0.575	0.763	0.732	0.732	0.688	0.744	0.772
FORU	0.699	0.842	0.693	0.565	0.838	0.803	0.802	0.776	0.824	0.855
EMAI	0.856	0.893	0.867	0.707	0.808	0.921	0.926	0.924	0.817	0.923
UCSO	0.790	0.921	0.779	0.641	0.917	0.891	0.891	0.869	0.897	0.931
KARA	0.700	0.781	0.738	0.596	0.717	0.756	0.752	0.745	0.655	0.863
POLB	0.889	0.834	0.896	0.789	0.669	0.900	0.900	0.899	0.746	0.917
FOOT	0.843	0.539	0.847	0.814	0.269	0.855	0.855	0.877	0.588	0.857
COMM	0.924	0.801	0.925	0.922	0.915	0.921	0.921	0.887	0.900	0.940

由表 2 中的结果可知,CCLP 指标和 CCLP2 指标在拓扑性质不同的网络中的预测性能表现出较大的差异。在 LESM、USAI、NSSM 和 KING 网络中,CCLP 指标比 CCLP2 指标的 AUC 值提升了 7.47%~29.6%;然而在 STMA、FWEW、FWFW、ADJN、FORU、EMAI 和 UCSO 等网络中,CCLP2 指标比 CCLP 指标的 AUC 值提升了 2.84%~17.44%。

从 CCLP 指标和 CCLP2 指标的 AUC 值可以看

出,仅考虑了共同邻居或二级节点的单一影响难以使得相应的算法在拓扑性质不同的网络中展现出较好的性能。结合了两类邻居共同影响的 NIRA 指标不仅实现了普适性,在准确率方面也有大幅度提升。此外,相比 RA 指标和 CRA 指标,NIRA 指标在较多的网络中取得了最高 AUC 值,这表明考虑资源的双向传输是更有效的。PA 指标在四个生态网络中表现良好,但在其余网络中均不如 NIRA 指标。LP、

表3 NIRA 指标与9个基准指标之间的 Precision 值比较

Table 3 Comparison of the Precision values between the NIRA index and nine benchmark indices

网络名	CCLP	CCLP2	RA	CRA	PA	LP	Katz	MFI	ACT	NIRA
LESM	0.209	0.023	0.217	0.215	0.077	0.203	0.204	0.168	0.128	0.220
USAI	0.621	0.223	0.637	0.625	0.475	0.602	0.602	0.051	0.485	0.669
NSSM	0.417	0.055	0.564	0.449	0.013	0.314	0.302	0.272	0.193	0.556
KING	0.795	0.264	0.866	0.577	0.252	0.430	0.435	0.347	0.196	0.889
STMA	0.106	0.123	0.101	0.092	0.136	0.107	0.107	0.068	0.123	0.137
FWEW	0.164	0.273	0.168	0.153	0.297	0.161	0.159	0.045	0.268	0.342
FWMW	0.144	0.218	0.143	0.145	0.263	0.144	0.142	0.051	0.254	0.267
FWFW	0.098	0.204	0.088	0.093	0.208	0.095	0.095	0.031	0.165	0.232
ADJN	0.049	0.078	0.044	0.044	0.076	0.051	0.051	0.004	0.065	0.081
FORU	0.048	0.072	0.026	0.062	0.062	0.049	0.049	0.000	0.060	0.064
EMAI	0.305	0.054	0.255	0.366	0.028	0.273	0.273	0.056	0.037	0.293
UCSO	0.055	0.114	0.038	0.065	0.070	0.035	0.040	0.001	0.079	0.141
KARA	0.041	0.051	0.046	0.021	0.040	0.047	0.045	0.035	0.039	0.060
POLB	0.143	0.084	0.148	0.135	0.046	0.129	0.128	0.097	0.113	0.147
FOOT	0.256	0.004	0.286	0.315	0.001	0.288	0.283	0.287	0.052	0.289
COMM	0.716	0.465	0.713	0.715	0.692	0.705	0.705	0.370	0.614	0.844

Katz、MFI、ACT 等指标的预测效果相对较好,但除了 FOOT 网络中的 MFI 指标,其余三个指标在剩下的 15 个网络中的 AUC 值均不如 NIRA 指标。

由表 3 可知 CCLP 指标和 CCLP2 指标的 Precision 值与它们的 AUC 值具有相同现象。由此可见,仅考虑共同邻居或二级节点的单一影响的确难以保证算法的普适性。值得一提的是,考虑共同邻居影响的算法似乎更适合用 Precision 来评价,如 RA 指标在 NSSM 和 POLB 网络中取得了最优 Precision 值,而 CRA 指标在 EMAI 和 FOOT 网络中也取得了最优的 Precision 值。在大多数网络中,CCLP、RA 和 CRA 指标相比 PA、LP、Katz、MFI 和 ACT 等也都取得了更高的 Precision 值。

4.3 复杂度分析

在本文中,利用算法在各个网络中运行时消耗的时间来分析它们的时间复杂度。为此,本文将算法 a 在网络 b 中运行时所消耗的时间归一化到区间 $[0, 1]$ 来进行更为直观的分析。此外,考虑基于局部结构的链路预测算法具有较低的复杂度,本文只在节点数大于 800 的网络上分析 Katz 指标、MFI 指标、ACT 指标、NIRA 指标和 CCLP2 指标的时间复杂度。

表 4 给出了 Katz、MFI、ACT、NIRA 和 CCLP2 等五个指标在 KING、FORU、EMAI 和 UCSO 网络中的归一化时间消耗。从表 4 中的结果可以看出,NIRA 指标相比于 CCLP2 指标在运行时间上也具有一定的优势,但相比其余三个基准指标,NIRA 指标却是

在牺牲了一定的计算时间后换来性能上的提升。这是因为在构造 NIRA 指标和 CCLP2 指标的过程中,需要先找出两个节点之间的二级节点集合,而这一过程需要在 MATLAB 编程软件中利用两个 for 循环来完成,故 NIRA 指标和 CCLP2 指标的运行时间增加了。

表4 五个指标的归一化时间消耗

Table 4 Normalized time consumption of five indices

网络名	Katz	MFI	ACT	NIRA	CCLP2
KING	0.045 0	0.038 7	0.149 3	0.988 4	1.000 0
FORU	0.073 7	0.067 2	0.099 3	0.994 8	1.000 0
EMAI	0.045 5	0.043 1	0.100 6	0.985 8	1.000 0
UCSO	0.061 9	0.060 0	0.165 7	0.994 8	1.000 0

5 结论

为了探究共同邻居和二级节点在链路预测中的影响,本文借助节点之间的双向资源传输,通过利用两个节点接收由它们的共同邻居和二级节点分配给它们的资源总量来度量它们的相似性。实验结果表明,在平均聚类系数较高的网络中,两个节点接收更多由共同邻居分配给它们的资源时,链路预测效果更佳;在平均聚类系数较低的网络和生态网络中,两个节点接收更多由二级节点分配给它们的资源,链路预测效果更佳;在拓扑性质不同的多个网络中,结合两类邻居的共同影响而提出的算法不仅具有较高

的预测准确率,还实现了良好的普适性。

本文在后续的研究中将探究共同邻居和二级节点之间的其余子节点对于提升链路预测算法性能的影响,从而进一步优化本文所提出的预测模型。

参考文献:

- [1] 刘振鹏,董妹慧,李泽园,等. 面向社交网络数据的等差数列聚类匿名算法[J]. 郑州大学学报(理学版), 2022, 54(1): 41-47.
LIU Z P, DONG S H, LI Z Y, et al. Anonymity algorithm of arithmetic sequence clustering for social network data[J]. Journal of Zhengzhou university (natural science edition), 2022, 54(1): 41-47.
- [2] MUZIO G, O'BRAY L, BORGWARDT K. Biological network analysis with deep learning[J]. Briefings in bioinformatics, 2020, 22(2): 1515-1530.
- [3] 袁培燕,黄笑妍. 机会网络中基于节点效用和能量的路由算法[J]. 郑州大学学报(理学版), 2022, 54(2): 47-55.
YUAN P Y, HUANG X Y. Routing algorithm based on node utility and energy in opportunistic networks[J]. Journal of Zhengzhou university (natural science edition), 2022, 54(2): 47-55.
- [4] 刘英杰,刘士虎,徐伟华. 基于有效路径拓扑稳定性的链路预测方法[J]. 计算机应用研究, 2022, 39(1): 90-95.
LIU Y J, LIU S H, XU W H. Link prediction method based on topology stability of effective path[J]. Application research of computers, 2022, 39(1): 90-95.
- [5] 毛北行,孟金涛. 离散复杂网络系统的混沌同步[J]. 郑州大学学报(理学版), 2013, 45(3): 9-12.
MAO B X, MENG J T. Chaos synchronization of discrete complex network systems[J]. Journal of Zhengzhou university (natural science edition), 2013, 45(3): 9-12.
- [6] 解滨,董新玉,梁皓伟. 基于三支动态阈值 K -means 聚类的入侵检测算法[J]. 郑州大学学报(理学版), 2020, 52(2): 64-70.
XIE B, DONG X Y, LIANG H W. An algorithm of intrusion detection based on three-way dynamic threshold K -means clustering[J]. Journal of Zhengzhou university (natural science edition), 2020, 52(2): 64-70.
- [7] ZHOU T. Experimental analyses on 2-hop-based and 3-hop-based link prediction algorithms[J]. Physica A: statistical mechanics and its applications, 2021, 564: 125532.
- [8] CALDERONI F. Robust link prediction in criminal networks: a case study of the Sicilian Mafia[J]. Expert systems with applications, 2020, 161: 113666.
- [9] 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.
- [10] ZHOU T, LÜ L, ZHANG Y C. Predicting missing links via local information[J]. The European physical journal B, 2009, 71(4): 623-630.
- [11] WU Z H. Link prediction with node clustering coefficient[J]. Physica A: statistical mechanics and its applications, 2016, 452: 1-8.
- [12] KATZ L. A new status index derived from sociometric analysis[J]. Psychometrika, 1953, 18(1): 39-43.
- [13] LÜ L, JIN C H, ZHOU T. Similarity index based on local paths for link prediction of complex networks[J]. Physical review E: statistical, nonlinear, and soft matter physics, 2009, 80(4 Pt 2): 046122.
- [14] ABBAS K, ABBASI A, DONG S, et al. Application of network link prediction in drug discovery[J]. BMC bioinformatics, 2021, 22(1): 187.
- [15] XIE Y B. Scale-free networks without growth[J]. Physica A: statistical mechanics and its applications, 2008, 387(7): 1683-1688.
- [16] CHEBOTAREV P, SHAMIS E. The matrix-forest theorem and measuring relations in small social groups[EB/OL]. [2022-05-23]. <https://arxiv.org/abs/math/0602070>.
- [17] KUMAR A, SINGH S S, SINGH K, et al. Level-2 node clustering coefficient-based link prediction[J]. Applied intelligence, 2019, 49(7): 2762-2779.
- [18] 陈紫扬,张月霞. 结合二层节点度和聚类系数的链路预测算法[J]. 计算机工程与应用, 2019, 55(23): 40-44.
CHEN Z Y, ZHANG Y X. Link prediction algorithm combining two-layer node degree and clustering coefficient[J]. Computer engineering and applications, 2019, 55(23): 40-44.
- [19] LIU S X. Extended resource allocation index for link prediction of complex network[J]. Physica A: statistical mechanics and its applications, 2017, 479: 174-183.
- [20] LIU Y J, LIU S H, YU F S, et al. Link prediction algorithm based on the initial information contribution of nodes[J]. Information sciences, 2022, 608: 1591-1616.
- [21] 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.
LI Y L, HE Z Y, WANG K, et al. Link prediction method based on topological tightness of resource transmission nodes[J]. Computer engineering, 2021, 47(1): 50-57.