

# 基于互信息的复杂网络链路预测

齐方鹏,王童,傅忠谦

(中国科学技术大学电子科学与技术系,安徽合肥 230027)

**摘要:** 互信息在复杂网络中的应用为解决链路预测问题提供了一个新的思路. 传统的互信息方法(MI)不仅考虑了节点的邻居信息,还加入了共同邻居之间的结构信息,这种方法比传统的基于共同邻居的方法预测精度更高;但是该方法没有对共同邻居进行有效的区分,即没有考虑到共同邻居之间的差异性. 为此进行了相应的改进,提出了改进的互信息方法(MMI),实验结果表明,MMI方法可以在一定程度上提高链路预测的精度.

**关键词:** 复杂网络;链路预测;互信息

**中图分类号:** TP391      **文献标识码:** A      doi: 10.3969/j.issn.0253-2778.2020.01.007

**引用格式:** 齐方鹏,王童,傅忠谦. 基于互信息的复杂网络链路预测[J]. 中国科学技术大学学报,2020,50(1):57-63.

QI Fangpeng, WANG Tong, FU Zhongqian. Link prediction in complex networks based on mutual information[J]. Journal of University of Science and Technology of China, 2020,50(1):57-63.

## Link prediction in complex networks based on mutual information

QI Fangpeng, WANG Tong, FU Zhongqian

(Department of Electronic Science and Technology, University of Science and Technology of China, Hefei 230027, China)

**Abstract:** A new perspective of dealing with link prediction problem was derived due to the application of mutual information in complex networks. Traditional mutual information algorithm (MI) not only considers the neighbor information of nodes, but also the structural information of common neighbors. Although MI has better performance compared with traditional methods which are based on common neighbors, it doesn't effectively differentiate between different common neighbors. A new algorithm (MMI) was proposed by considering the influence of different common neighbors, which performs better than MI in precision.

**Key words:** complex network; link prediction; mutual information

## 0 引言

复杂系统普遍存在于自然界和人类社会中,包括生态系统、道路交通系统和电力系统等,一般可以采用复杂网络对复杂系统进行研究<sup>[1-2]</sup>. 在复杂网络中,节点表示网络中的个体,连边表示个体之间的某

种联系. 例如,在蛋白质相互作用网络<sup>[3]</sup>中,节点可以表示某种蛋白质,而连边则可以表示蛋白质之间的相互作用. 链路预测作为复杂网络研究中重要的领域之一,受到了各领域研究人员的青睐,不仅仅因为它在各领域应用广泛,还因为链路预测具有理论研究价值. 比如在微博用户网络中,通过链路预测挖

收稿日期: 2018-03-31; 修回日期: 2018-12-21

作者简介: 齐方鹏,男,1992年生,硕士生. 研究方向: 链路预测. E-mail: fpqi@mail.ustc.edu.cn

通讯作者: 傅忠谦,副教授. E-mail: zqfu@ustc.edu.cn

掘用户之间的隐藏关系,从而进行精准的推荐,提高用户的满意度,使公司的收益得到增加.在电力网络中,通过预测两个电力站点之间建立直接联系的可能性,从而使负载更加均衡,避免过载和空载的现象,使网络的抗干扰能力更强.在网络理论研究中,链路预测可以通过网络的演化特征来揭示网络的演化机制,从而可以对网络进行更好的控制和传播,进一步推动复杂网络基础理论的建立和完善<sup>[4]</sup>.

网络中的链路预测是指通过已知的网络信息(包括节点信息和结构信息)来预测网络中尚未产生连边的两个节点之间产生连接的可能性<sup>[5]</sup>.这里既包括了对未知链接的预测,也包括了对未来链接的预测.未知链接指的是网络中尚且存在的但是未被发现的链接,而未来链接是指网络中尚且不存在但是以后可能存在的链接.目前研究链路预测的方法多种多样,其中应用最广泛的是基于结构相似性<sup>[6]</sup>的方法,该方法利用网络的结构信息和节点属性等信息为网络中的节点赋予一个相似性分数,分数越高则表明该节点对产生连接的可能性越大.根据所使用拓扑结构信息不同范围可以进一步分为基于局部信息的相似性方法,如共同邻居(CN)方法,资源分配(RA)方法等;基于全局信息的相似性方法,如Katz方法,随机游走的余弦相似性(Cos+)方法和带重启的随机游走(RWR)方法;基于半局部信息的相似性方法,如局部路径(LP)方法和局部随机游走(LRW)方法等.

近年来,信息论在复杂网络中得到了广泛的应用. Anand等<sup>[7]</sup>对网络的香农熵给出了明确的定义,并用来描述网络的异质性; Halu等<sup>[8]</sup>研究了空间多层和交互网络的最大集成熵,还研究了网络传播过程和随机行走等网络的动态熵.同时,信息论还被应用于社区识别、年龄识别和癌症的趋势以及链路预测中. 2014年, Tan等<sup>[9]</sup>利用互信息的概念刻画了共同邻居对连边的贡献,并提出了基于互信息的相似性算法,该算法比传统的算法预测效果更好; Zhu等<sup>[10]</sup>将该算法进一步应用到加权网络中,提出了加权的互信息算法,并结合弱连接理论,讨论了权重惩罚参数的选取问题; Xu等<sup>[11]</sup>在互信息方法的基础上,与Katz算法相结合,首次提出了路径熵的概念.

链路预测本质上是计算两个节点之间产生链接的可能性,而这种可能性在信息论中可以用信息量来表示,即一个事件发生的可能性越大,那么该事件的信息量就越小,因此利用信息量进行链路预测是

一种行之有效的方式.近几年,互信息方法在链路预测中得到了广泛的应用,传统的互信息方法不仅考虑了节点的邻居信息,还考虑了共同邻居之间的结构信息,在链路预测中取得了较好的预测效果<sup>[9]</sup>.传统的互信息方法没有对不同的共同邻居进行有效区分,由于在同一个网络中,不同的节点因其位置和自身属性的不同,其作用也不一样,因此区分不同的共同邻居有重要的意义.本文通过共同邻居的度信息对不同的共同邻居进行区分,通过取指数控制共同邻居对端节点的影响,从而提出了改进的互信息方法.实验表明,该方法可以在一定程度上提高链路预测的精度.

## 1 互信息

在信息论中,信息量是用来对信息多少的度量.对单个事件而言,可以用自信息量衡量一个事件发生的概率;而对于多个事件而言,则可以用互信息量表示事件之间的相互作用.互信息在链路预测中的形式多种多样,如Tan等<sup>[9]</sup>通过计算共同邻居与端节点产生连边之间的互信息,从而进行链路预测,并取得了较好的预测效果.

### 1.1 互信息的定义

一个事件或变量包含的信息量称为自信息量,简称自信息,它取决于事件发生的不确定性程度,即事件发生的概率.对于一个随机变量 $X$ ,如果随机事件 $x_k$ 发生的概率为 $p(x_k)$ ,那么该事件的自信息量定义为

$$I(x_k) = \log\left(\frac{1}{p(x_k)}\right) = -\log(p(x_k)) \quad (1)$$

信息量的单位与 $\log$ 所用的底有关,通常取底数为2,此时信息量的单位是比特;取自然常数 $e$ 时,信息量的单位是奈特;取10时,信息量的单位是哈特莱.事件发生的概率越小,其信息量越大.考虑在一个事件发生的条件下,另一个事件发生的概率时,可以使用条件自信息,其公式为

$$I(x_i | y_j) = -\log(p(x_i | y_j)) \quad (2)$$

式中, $I(x_i | y_j)$ 表示在事件 $y_j$ 发生的条件下,事件 $x_i$ 发生的信息量.

对于两个随机变量 $X$ 和 $Y$ ,如果其联合概率密度为 $p(x, y)$ ,边缘概率密度分别为 $p(x)$ 和 $p(y)$ ,那么互信息 $I(X; Y)$ 的计算公式为

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} =$$

$$\sum_{x,y} p(x,y) \log \frac{p(x|y)}{p(x)} \quad (3)$$

当  $X = x_k, Y = y_j$  时,互信息  $I(x_k; y_j)$  计算公式为

$$I(x_k; y_j) = I(x_k) - I(x_k | y_j) \quad (4)$$

### 1.2 基于互信息的链路预测方法

定义  $\Gamma(x)$  为节点  $x$  邻居的集合,  $O_{xy}$  表示节点  $x$  和节点  $y$  共同邻居的集合, 则  $O_{xy} = \Gamma(x) \cap \Gamma(y)$ , 对于一个未连接的节点对  $(x, y)$ , 其相似性分数定义为

$$S_{xy}^{MI} = -I(L_{xy}^1 | O_{xy}) \quad (5)$$

式中,  $I(L_{xy}^1 | O_{xy})$  表示在共同邻居存在的条件下, 节点对  $(x, y)$  之间存在连接的条件自信息. 根据自信息的性质可以知道,  $I(L_{xy}^1 | O_{xy})$  越小, 它们之间存在连接的可能性越大, 因此这里需要加负号. 根据公式(4),  $I(L_{xy}^1 | O_{xy})$  的计算公式为

$$I(L_{xy}^1 | O_{xy}) = I(L_{xy}^1) - I(L_{xy}^1; O_{xy}) \quad (6)$$

式中,  $I(L_{xy}^1)$  是节点对  $(x, y)$  产生连接的自信息,  $I(L_{xy}^1; O_{xy})$  表示  $x$  和  $y$  之间存在共同邻居  $O_{xy}$  的同时, 节点对  $(x, y)$  产生连接的互信息.  $I(L_{xy}^1 | O_{xy})$  表示在共同邻居存在的条件下, 节点  $x$  和  $y$  之间产生连接的条件自信息.

如果  $O_{xy}$  里面的元素都是相互独立的, 那么

$$I(L_{xy}^1; O_{xy}) = \sum_{z \in O_{xy}} I(L_{xy}^1; z) \quad (7)$$

式中,  $I(L_{xy}^1; z)$  用  $I(L^1; z)$  来代替,  $I(L^1; z)$  表示在  $z$  存在的同时,  $z$  的邻居之间产生连接的互信息的平均值, 计算公式为

$$I(L^1; z) = \frac{1}{|\Gamma(z)|(|\Gamma(z)|-1)} \sum_{\substack{m \neq n \\ m, n \in \Gamma(z)}} I(L_{mn}^1; z) \quad (8)$$

根据公式(4),  $I(L_{xy}^1; z)$  可以表示为

$$I(L_{mn}^1; z) = I(L_{mn}^1) - I(L_{mn}^1 | z) \quad (9)$$

式中,  $I(L_{mn}^1)$  表示节点对  $(m, n)$  产生连接的自信息,  $I(L_{mn}^1 | z)$  表示当  $z$  是节点  $m$  和  $n$  的共同邻居时, 该节点对连接的条件自信息, 根据条件自信息的定义, 可以通过如公式(2)所示条件概率来计算.

条件概率  $p(L_{mn}^1 | z)$  可以用节点  $z$  的簇系数来近似表示, 定义为

$$p(L_{mn}^1 | z) = \frac{N_{\Delta z}}{N_{\Delta z} + N_{\wedge z}} \quad (10)$$

式中,  $N_{\Delta z}$  和  $N_{\wedge z}$  分别表示当  $z$  为两个节点的共同邻居时, 节点  $z$  的所有邻居之间节点对连接和未连接的数目. 得到  $p(L_{mn}^1 | z)$  后, 进而可以得到

$$I(L_{mn}^1 | z).$$

利用网络中各节点的度, 可以计算出节点对  $(m, n)$  不连接的概率为

$$p(L_{mn}^0) = \frac{C_{M-k_m}^{k_n}}{C_M^{k_n}} \quad (11)$$

式中,  $k_m$  和  $k_n$  分别表示节点  $m$  和节点  $n$  的度,  $M$  是训练集中总的边数. 这个公式显然是对称的, 故  $I(L_{mn}^1)$  的计算公式为

$$I(L_{mn}^1) = -\log(p(L_{mn}^1)) = -\log(1 - \frac{C_{M-k_m}^{k_n}}{C_M^{k_n}}) \quad (12)$$

通过上面的推理, 可以得到每个节点对的相似性分数为

$$S_{xy}^{MI} = \sum_{z \in O_{xy}} I(L_{xy}^1; z) - I(L_{xy}^1) \quad (13)$$

式中,  $I(L_{xy}^1; z)$  和  $I(L_{xy}^1)$  可以通过公式(8)和(12)得到.

## 2 改进的互信息方法

传统的互信息方法既考虑了节点的邻居信息, 又考虑了共同邻居之间的结构信息, 但是没有对不同的共同邻居进行区分. 网络中的节点因其位置和自身属性的不同, 导致不同的节点在网络中的重要性不同, 因此有必要对不同的共同邻居进行区分. 经典的基于共同邻居的方法有很多, 如 AA 和 RA, 这两种方法都是利用共同邻居的度信息进行链路预测, 并且可以取得较好的预测效果. 如图 1 所示, 节点  $a$  和节点  $b$  之间有 3 个共同邻居  $c, d$  和  $e$ , 这 3 个节点的度分别为 6, 4 和 3. 其中节点  $d$  不仅与节点  $a$  和节点  $b$  相连, 也是节点  $c$  和节点  $e$  的共同邻居, 因此节点  $a$  和节点  $b$  更有可能通过节点  $d$  产生连接. 虽然节点  $c$  的度最大, 但是它有 3 个与节点  $a, b$  无关的邻居节点, 因此节点  $c$  对节点  $a, b$  之间产生的影响应该是最小的.

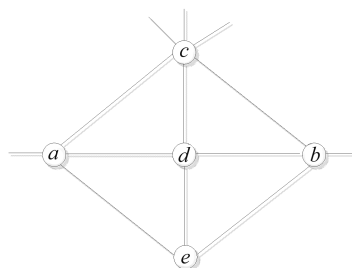


图 1 网络示例图

Fig. 1 Example of network

基于以上分析,本文对原始互信息方法进行了改进.改进的互信息(MMI)方法通过共同邻居节点的度信息对节点进行区分,因此 MMI 可以表示为

$$S_{xy}^{\text{MMI}} = -I'(L_{xy}^1 | O_{xy}) = I'(L_{xy}^1; O_{xy}) - I(L_{xy}^1) = \sum_{z \in O_{xy}} I(L_{xy}^1; z) * |\Gamma(z)|^\alpha - I(L_{xy}^1) \quad (14)$$

对于每个共同邻居  $z$ ,  $(|\Gamma(z)|)^\alpha$  表示共同邻居节点  $z$  对两个端节点产生连接的影响,  $\alpha$  是影响因子,表示共同邻居的影响程度.当  $\alpha < 0$  时,表明共同邻居节点的度对边的连接起抑制作用,即度与边的连接成反比;当  $\alpha = 0$  时,MMI 退化为 MI;当  $\alpha > 0$  时,表明共同邻居节点的度对边的连接起促进作用,即度与边的连接成正比.图 2 给出了 MMI 算法的流程图.

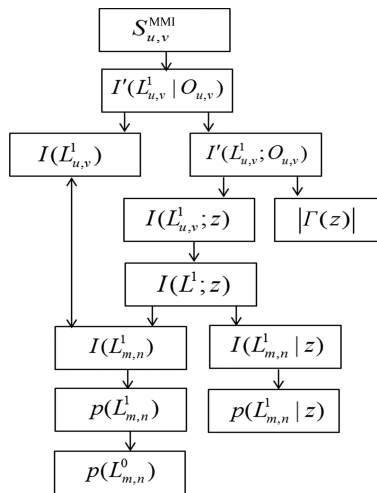


图 2 MMI 算法流程图

Fig. 2 Procedure of MMI algorithm

从图 2 可以看到,实验中用到了  $I(L_{u,v}^1)$  和  $I(L_{m,n}^1)$ ,这两个自信息量的计算公式仅字母不同.如果每一次都重新计算,将出现计算冗余,当网络较大时,冗余现象尤为严重.同时,节点  $z$  可能是多个节点对共同邻居,因此  $I(L_{u,v}^1 | z)$  也会多次重复计算,节点的度越大,重复计算次数越多.为了减少重复计算的时间消耗,首先对数据进行预处理,计算所有的  $I(L_{u,v}^1)$  和  $I(L_{u,v}^1 | z)$ ,并作为参数传入主函数中.

基于互信息的链路预测方法与传统的链路预测方法有如下几点区别:

(I) 将信息熵的概念应用到链路预测方法中,从信息论的角度对已知问题进行研究;

(II) 与其他半局部链路预测方法如 LP 相比,

互信息方法在利用节点间距离信息的基础上,还考虑了邻居节点之间的相互影响,并且考虑了邻居节点的差异性.

### 3 实验准备

#### 3.1 问题描述

对于一个简单的无向无权网络  $G(V, E)$ , 定义  $V$  为网络中节点的集合,  $E$  为网络中边的集合. 令  $N = |V|$  表示网络  $G$  中节点的个数, 用  $U$  表示网络中所有可能的边, 则  $|U| = N(N - 1)/2$ . 链路预测的问题可以描述为对一个给定的网络, 通过一种链路预测方法, 对所有未连接的节点对  $(x, y)$  分配一个分数  $S(x, y)$ , 称之为相似性分数, 然后按照相似性分数从大到小排序, 排名越靠前, 表示未来出现连边的可能性就越大. 为了对不同的链路预测方法进行评价, 首先将数据集划分为训练集  $E^T$  和测试集  $E^P$ , 这里  $E^T \cap E^P = \Delta$ , 且  $E^T \cup E^P = E$ , 其中  $E^T$  是网络中已知的信息, 而  $E^P$  则是需要预测的网络. 本次试验中随机选取 20% 的边作为测试集, 剩余 80% 的边作为训练集.

#### 3.2 评价标准和对比算法

本文使用的评价标准是精确度<sup>[12]</sup> (precision), 指在前  $L$  条边中预测准确的比例. 如果有  $m$  个预测准确, 即根据出现连接的可能性将边从大到小排列, 若排在前  $L$  的边中有  $m$  个在测试集中, 则精确度定义为

$$\text{precision} = \frac{m}{L} \quad (17)$$

由公式(17)可知, 精确度的大小和  $L$  有关, 对于给定的  $L$ , 精确度越大, 预测结果越准确.

本文使用的对比方法如下:

(I) 共同邻居(CN)方法<sup>[13]</sup>

该方法认为, 如果两个节点之间拥有更多的共同邻居, 则它们更倾向于产生连接, 其公式为

$$S_{xy}^{\text{CN}} = |\Gamma(x) \cap \Gamma(y)| \quad (18)$$

式中,  $\Gamma(x)$  和  $\Gamma(y)$  分别为节点  $x$  和节点  $y$  邻居的集合.

(II) Adamic-Adar(AA)方法<sup>[14]</sup>

该方法类似于 RA, 但是对节点的度进行对数惩罚, 公式表示为

$$S_{xy}^{\text{AA}} = \sum_{z \in O_{xy}} \log \frac{1}{|\Gamma(z)|} \quad (19)$$

(III) 资源分配(RA)方法<sup>[15]</sup>

该方法对度大的节点进行相应的惩罚,其公式为

$$S_{xy}^{RA} = \sum_{z \in O_{xy}} \frac{1}{|\Gamma(z)|} \quad (20)$$

除了以上几个对比方法之外,还有前面提到的互信息方法,这里不再赘述.

## 4 实验结果与分析

### 4.1 数据集描述

本实验共使用了 4 个数据集,均来自公开的数据集网站<sup>[16]</sup>,分别如下:

(I) Friendship. 该数据集是 hamster. com 网站上的好友关系数据集,节点代表用户,边代表用户之间的好友关系. 该数据集属于社交网络数据集.

(II) Political Blogs (PolBlog). 该数据集是由 Adamic 和 Glance 于 2005 年收集的关于美国政治的网页之间的链接,节点代表一个网页,边代表两个网页之间至少存在一个超链接. 该数据集属于网页超链接数据集.

(III) US Power Grid (PowerGrid). 该数据集是美国西部电网数据集,一个节点代表一个发动机、变压器或者变电站,边代表的是电力供应. 该数据集属于电力网络数据集.

(IV) Yeast. 该数据集是和蛋白质相关的数据集,节点表示一个蛋白质,边代表两个蛋白质之间存在相互作用. 该数据集属于蛋白质网络数据集.

本次实验在 4 个数据集上对不同的相似性方法进行了对比,表 1 列出了这几个数据集的各项参数. 对于有向连接,忽略它们的方向,看作无向连接,并且删除网络中的自环和多重边.

表 1 实验中使用数据集的相关参数

Tab. 1 The parameters of datasets in the experiment

数据集	节点数	边数	平均度	簇系数	稀疏性
PolBlog	1 490	16 715	11. 218 1	0. 312 2	0. 015 1
Yeast	2 361	6 646	2. 814 9	0. 289 5	0. 002 4
Friendship	1 858	12 534	6. 746 0	0. 197 5	0. 007 3
PowerGrid	4 941	6 594	1. 334 5	0. 416 2	5. 40E-04

从表 1 可以看到,PolBlog 数据集的平均度最大,PowerGrid 数据集的平均度最小,仅有 PolBlog 数据集的十分之一;从簇系数来看,PowerGrid 数据集的簇系数最大,Friendship 的簇系数最小,前者是

后者的 2 倍.

### 4.2 实验结果

本文以精确度作为算法的衡量指标,以 CN、AA、RA 和 MI 方法作为对比算法,将改进的互信息方法分别应用于 PolBlog、Yeast、Friendship 和 PowerGrid 数据集. 表 2 给出了  $L=100$  时,各算法的预测精度,其中加粗部分为最优值.

表 2  $L=100$  时,不同相似性方法在不同数据集集中的精确度

Tab. 2 Precision of different algorithms in different datasets when  $L=100$

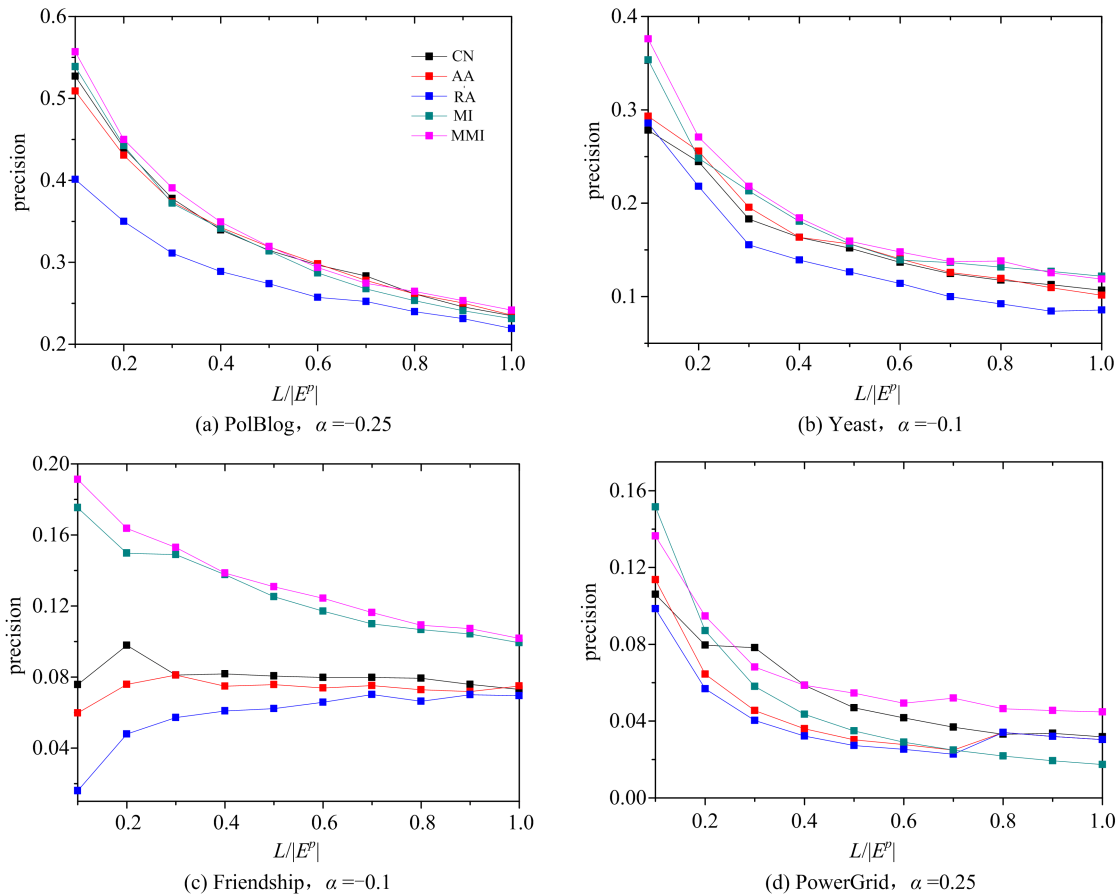
数据集	CN	AA	RA	MI	MMI
PolBlog	0. 67	0. 59	0. 47	0. 68	<b>0. 69</b>
Yeast	0. 29	0. 34	0. 34	0. 38	<b>0. 4</b>
Friendship	0. 03	0. 02	0	<b>0. 2</b>	<b>0. 2</b>
PowerGrid	0. 15	0. 14	0. 13	<b>0. 19</b>	<b>0. 19</b>

从表 2 可以看到,与基于共同邻居的方法相比,基于互信息的方法(MI 和 MMI)都能取得较好的预测效果,特别是在 Friendship 数据集中,基于共同邻居的方法预测精度很低,而基于互信息的方法则可以取得较高的预测精度. MMI 方法与 MI 方法相比,在 PolBlog 和 Yeast 数据集中有一定的提升. 为了消除  $L$  对实验结果的影响,进一步比较 MMI 对 MI 的改进效果,将  $L$  从  $0.1 * |E^p|$  到  $1 * |E^p|$  不断变化,其实验结果图 3 所示.

图 3 中,纵坐标表示精确度,横坐标表示计算精确度时所选取的  $L$  与该数据集的测试集边数  $|E^p|$  的比值. 图 3 从整体上来看,相比于对比方法,MMI 方法可以取得较好的预测精度,其中在 Friendship 和 PowerGrid 数据集中,改进效果相对明显. 在 Friendship 数据集中,基于互信息的方法表现十分突出,其预测精度远高于基于共同邻居的方法;在 PowerGrid 数据集中,当  $L/|E^p| > 0.5$  时,MMI 算法的预测精度几乎是 MI 算法的 2 倍.

根据参数  $\alpha$  的定义可以知道, $\alpha$  取正数和取负数具有完全不同的意义. 从图 3 中可以看到,MMI 方法中  $\alpha$  的最优值在不同数据集中不一样,其中在 PowerGrid 数据集中  $\alpha$  的最优值为  $0.25 > 0$ ,而在其他数据集中均为负数. 由此可以得出结论:在 PowerGrid 数据集中,共同邻居的度越大,其端节点产生链接的可能性越大,在本文所使用的其他数据集中则恰恰相反.

本文以精确度作为评价标准(精确度越高,说明

图 3 预测精度随  $L/|E^p|$  的变化Fig. 3 trend of precision with  $L/|E^p|$ 

算法越准确,效果越好),在 4 个数据集中进行测试.从表 2 中可以看到,当  $L$  固定时,在 PolBlog 和 Yeast 数据集中分别提升了 1.47% 和 5.2%. 由于每个数据集的大小不同,我们探究了不同  $L/|E^p|$  的算法表现.通过图 3 可知,随着  $L/|E^p|$  的变化,基于互信息的方法精确度整体高于传统方法.在 PolBlog 和 Yeast 数据集中,MMI 的精度优势体现在  $L/|E^p|$  较小的情况;在 Friendship 数据集中,MMI 方法和 MI 方法一直遥遥领先,充分展现了互信息方法的优点;在 PowerGrid 数据集中,除了当  $L = 0.1 * |E^p|$  之外,MMI 方法优于 MI 方法,并且其预测精度的差值一直在变大.

## 5 结论

互信息方法在复杂网络中的应用为链路预测问题的研究提供了新的思路,本文在原始互信息方法的基础上,通过理论分析发现其中的不足,即没有对不同的共同邻居进行有效区分,提出了改进的互信息(MMI)方法.实验结果表明,MMI 方法可以在一

定程度上提高链路预测的精度,并且通过对参数  $\alpha$  分析后发现,在不同的网络中,共同邻居对端节点的影响不同.在大部分网络中,共同邻居的度越大,越抑制端节点的连接,在小部分网络中则恰恰相反.为了消除分母  $L$  对实验结果的影响,本文还研究了当  $L$  变化时,预测精度随之变化的情况.实验表明,不论  $L$  取多大,MMI 方法大部分时候都能取得较好的效果,充分体现了该方法的鲁棒性.

本文是基于简单的无向无权网络进行的研究,没有考虑边的方向性和权重信息,在这些有权有向网络中,如何利用互信息进行链路预测值得进一步的思考和研究.实际中的网络都是动态演化的,静态网络中的互信息方法不一定适用,因此如何在互信息的基础上考虑网络的时间信息也是亟待解决的问题.

## 参考文献(References)

- [1] ALBERT R, BARABÁSI AL. Statistical mechanics of complex networks[J]. Reviews of Modern Physics,

- 2002, 74(1): 47-51.
- [ 2 ] NEWMAN M E J. The structure and function of complex networks[J]. SIAM Review, 2003, 45(2): 167-256.
- [ 3 ] LEI C, RUAN J. A novel link prediction algorithm for reconstructing protein - protein interaction networks by topological similarity[J]. Bioinformatics, 2013, 29(3): 355-364.
- [ 4 ] DOROGOVTSEV S N, MENDES J F. Evolution of networks [J]. Advances in Physics, 2002, 51(4): 1079-1187.
- [ 5 ] LÜ L Y, ZHOU T. Link prediction in complex networks: A survey [J]. Physics A: Statistical Mechanics and its Applications, 2011, 390(6): 1150-1170.
- [ 6 ] LIBEN-NOWELL D, KLEINBERG J. The link-prediction problem for social networks[J]. Journal of the American society for Information Science and Technology, 2007, 58(7): 1019-1031.
- [ 7 ] ANAND K, BIANCONI G. Entropy measures for networks: Toward an information theory of complex topologies [J]. Physical Review E: Statistical Nonlinear and Soft Matter Physic, 2009, 80(4): 045102.
- [ 8 ] HALU A, MUKHERJEE S, BIANCONI G. Emergence of overlap in ensembles of spatial multiplexes and statistical mechanics of spatial interacting network ensembles[J]. Physical Review E: Statistical Nonlinear and Soft Matter Physic, 2014, 89(1): 012806.
- [ 9 ] TAN F, XIA Y, ZHU B. Link prediction in complex networks: A mutual information perspective[J]. PloS One, 2014, 9(9): e107056.
- [10] ZHU B, XIA Y. Link prediction in weighted networks: A weighted mutual information model[J]. PloS One, 2016, 11(2): e0148265.
- [11] XU Z, PU C, YANG J. Link prediction based on path entropy[J]. Physica A: Statistical Mechanics and its Applications, 2016, 456: 294-301.
- [12] HERLOCKER J L, KONSTAN J A, TERVEEN L G, et al. Evaluating collaborative filtering recommender systems [J]. ACM Transactions on Information Systems, 2004, 22(1): 5-53.
- [13] MARTÍNEZ V, BERZAL F, CUBERO J C. A survey of link prediction in complex networks [J]. ACM Computing Surveys, 2017, 49(4): 69-113.
- [14] BLISS C A, FRANK M R, DANFORTH C M, et al. An evolutionary algorithm approach to link prediction in dynamic social networks [J]. Journal of Computational Science, 2014, 5(5): 750-764.
- [15] LI L, QIAN L, CHENG J, et al. Accurate similarity index based on the contributions of paths and end nodes for link prediction[J]. Journal of Information Science, 2015, 41(2): 167-177.
- [16] JÉRÔME K. KONECT: The Koblenz network collection [EB/OL]. [2018-02-27] <http://konect.uni-koblenz.de>.