

# 基于 Motif 聚集系数与时序划分的高阶链接预测方法\*

康驻关, 金福生, 王国仁

(北京理工大学 计算机学院, 北京 100081)

通讯作者: 金福生, E-mail: jfs21cn@bit.edu.cn



**摘要:** 高阶链接预测是当前网络分析研究的热点和难点, 一个优秀的高阶链接预测算法不仅可以挖掘出复杂网络中节点间存在的潜在联系, 还有助于认识网络结构随时间演化的规律, 对于探索未知的网络关系有着重要的作用。大多数传统的链接预测算法仅考虑节点间的结构相似性特征, 而忽略高阶结构的特性以及网络变化的信息。提出一种基于 Motif 聚集系数与时序划分的高阶链接预测模型(MTLP 模型), 该模型通过提取网络中高阶结构的 Motif 聚集系数特征和网络结构演变等特征, 将其构建成可表示性特征向量, 并使用多层感知器网络模型进行训练完成链接预测任务。该模型能够同时结合网络中高阶结构的聚集特征与网络结构演变信息, 从而改善预测效果。通过在不同的数据集上进行实验, 其结果表明, 所提出的 MTLP 模型具有更好的高阶链接预测性能。

**关键词:** 动态网络; 链接预测; 高阶网络结构; 图机器学习

**中图法分类号:** TP311

中文引用格式: 康驻关, 金福生, 王国仁. 基于 Motif 聚集系数与时序划分的高阶链接预测方法. 软件学报, 2021, 32(3): 712-725. <http://www.jos.org.cn/1000-9825/6172.htm>

英文引用格式: Kang ZG, Jin FS, Wang GR. High-order link prediction method based on Motif aggregation coefficient and time series division. Ruan Jian Xue Bao/Journal of Software, 2021, 32(3): 712-725 (in Chinese). <http://www.jos.org.cn/1000-9825/6172.htm>

## High-order Link Prediction Method Based on Motif Aggregation Coefficient and Time Series Division

KANG Zhu-Guan, JIN Fu-Sheng, WANG Guo-Ren

(School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China)

**Abstract:** High-level link prediction is a hot and difficult problem in network analysis research. An excellent high-level link prediction algorithm can not only mine the potential relationship between nodes in a complex network but also help to understand the law of network structure evolves over time. Exploring unknown network relationships has important applications. Most traditional link prediction algorithms only consider the structural similarity between nodes, while ignoring the characteristics of higher-order structures and information about network changes. This study proposes a high-order link prediction model based on Motif clustering coefficients and time series partitioning (MTLP). This model constructs a representational feature vector by extracting the features of Motif clustering coefficients and network structure evolution of high-order structures in the network, and uses multilayer perceptron (MLP) network model

\* 基金项目: 国家自然科学基金(61732003, 61025007, 60933001); 国家重点研发计划(2020AAA0108500); 广东省重点研发计划(2020B010164002); 北京市科技重大专项(Z171100005117002)

Foundation item: National Natural Science Foundation of China (61732003, 61025007, 60933001); Key Research and Development Program of China (2020AAA0108500); Key R&D Project of Guangdong Province (2020B010164002); Beijing Major Science and Technology Projects (Z171100005117002)

本文由“支撑人工智能的数据管理与分析技术”专利特约编辑陈雷教授、王宏志教授、童咏昕教授、高宏教授推荐。

收稿时间: 2020-07-09; 修改时间: 2020-09-03; 采用时间: 2020-11-06; jos 在线出版时间: 2021-01-21

to complete the link prediction task. By conducting experiments on different real-life data sets, the results show that the proposed MTLP model has better high-order link prediction performance than the state-of-the-art methods.

**Key words:** dynamic networks; link prediction; high-order network structure; graph based machine learning

网络中的链接预测<sup>[1]</sup>指通过已知的网络结构信息,预测网络中尚未产生链接的节点间产生新链接的可能性.如果给定网络的当前状态,则链接预测将重点放在评估未来状态中存在边的可能性.可以理解为预测社交网络中两个人未来相互认识,计算机网络中的通信或生物网络中的生物分子相互作用的可能性.一个好的链接预测算法不仅可以用来挖掘网络中节点与结构间的关系,而且可以进一步帮助研究者探索网络结构的演化规律.目前,链接预测不仅在理论研究上取得了较好的成果,而且在一些特定的场景下也具有广泛的应用价值,例如预测社交网络中的友谊发现、推断基因与疾病之间的新关系、发现学术界的合作关系等<sup>[2,3,4]</sup>具体应用.常见的链接预测算法一次只需要预测一条边,然而在分析复杂网络结构演化时往往需要考虑所有的边,这些算法则不能进行很好的扩展.本文提出的高阶链接预测任务将探测更大的特征集,这也要求算法能够捕获数据集中更多潜在的可表示性特征.原则上,好的链接预测模型可以适用于网络中任意结构的预测,具体表示为预测网络中任意高阶的交互.毋庸置疑,链接预测是复杂的网络分析中一个有重要实用价值的研究问题,在交互事件发生之前对其进行预测,是深入探索网络结构演化的一项重要能力.

目前,绝大多数的链接预测算法都是基于网络低阶结构的研究<sup>[1,5]</sup>,这些方法通常只分析了网络中成对节点间的关系,无法有效捕获高阶结构中的潜在联系,从而在高阶结构中的预测效果表现较差.本文使用网络高阶结构来进行链接预测,充分挖掘网络中高阶结构的演变信息,从而获得对复杂网络结构的新见解.为了简化研究和便捷可表示性,本文重点以高阶结构中的基础结构(无向三角形结构)为主要的研究对象.

为了解决上述提出的高阶特征表示问题,本文从高阶网络结构特性和网络结构演化两个维度进行深入分析,提出了计算高阶结构聚集性特征和网络结构演变特征的两个方法:首先,通过对高阶结构的聚集行为研究,将其进行量化用来表示高阶结构的聚集程度,从而有效捕获高阶结构的聚集性特征;其次,通过将数据集进行切割分片,分析随着时间演变网络结构的变化规律,提出节点间链接频次和链接趋势等具有网络结构演变特性的特征,有效利用了动态网络演化的历史信息.基于上述两个方法,提出了基于 Motif 聚集系数与时序划分的高阶链接预测模型用于完成预测任务.

综上所述,本文主要贡献总结如下:

- (1) 通过分析网络高阶结构,提出 Motif 聚集系数特征,将传统对单节点的聚集系数定义扩展到高阶结构的聚集系数定义.经过实验验证,Motif 聚集系数可以有效捕获网络中高阶结构的聚类信息,并且验证了使用 Motif 聚集系数表示高阶结构聚集程度的有效性;
- (2) 通过分析网络结构演变的特性,提出节点间链接频率和链接趋势等表示网络结构演变的重要特征,可以挖掘网络结构演变过程中节点之间的链接变化规律;
- (3) 提出了基于 Motif 聚集系数与时序划分的高阶链接预测模型(high-order link prediction model based on Motif aggregation coefficient and time series division,简称 MTLP),有效融合传统链接预测提出的相似性度量特征与本文提出的 Motif 聚集系数特征和表示网络结构演变的特征,通过使用多层感知器网络(multi-layer perceptron,简称 MLP)算法完成高阶链接预测任务.

本文在第 1 节介绍相关研究工作.第 2 节给出本文的问题定义.第 3 节重点介绍本文提出的高阶链接预测模型.第 4 节为实验分析.第 5 节总结全文.

## 1 相关研究

近年来,链接预测方法主要应用在社交网络和其他复杂网络等领域.针对现有的链接预测算法的研究,本文将预测模型大致分为基于局部结构相似性的链接预测方法、基于概率模型的链接预测方法和基于机器学习的链接预测方法这三大类.

### (1) 基于局部结构相似性的链接预测方法

常见的相似性指标有基于调和平均数、几何平均数和算术平均数相似性指标<sup>[3]</sup>、基于共同邻居(common neighbors,简称 CN)相似性指标<sup>[1,6]</sup>、基于 Jaccard 相似性指标<sup>[1]</sup>、基于 Adamic/Adar(Adamic-Adar,简称 AA)相似性指标<sup>[1,7]</sup>、基于优先链接(preferential attachment,简称 PA)相似性指标<sup>[1,6,8]</sup>、基于 Katz 相似性指标<sup>[1,9,10]</sup>、基于 PageRank 相似性指标<sup>[1,11,12]</sup>和基于层次化混合特征指标<sup>[13]</sup>等.基于局部结构相似性的方法主要是用于捕获局部网络的特征,从而根据相似性的特征进行链接预测.但是这些方法无法有效地捕获高阶结构的特征,仅适用于低阶的链接预测.如果将其扩展到高阶链接预测,其预测效果并不理想.

### (2) 基于概率模型的链接预测方法

Clauset 等人<sup>[14]</sup>最初提出基于最大似然估计的链接预测方法,将其应用到有明显层次的网络结构中,发现有一定的精确度;Tian 等人<sup>[15]</sup>提出了一种基于最大似然估计的链接预测模型,将脑网络的数据建立了层次随机图,再结合马尔可夫算法计算脑网络边的链接概率,显示出了良好的预测性能.Das 等人<sup>[16]</sup>提出一种基于随机马尔科夫链预测模型,其以时序图作为社交网络链接预测的基础.Liu 等人<sup>[17]</sup>提出一种基于动态博弈的双向选择机制来预测未来的网络拓扑结构.该方法适合于在静态网络下进行预测,但是动态网络的结构是频繁变化的网络,因此该方法往往难以满足要求.

### (3) 基于机器学习的链接预测方法

Chiu 等人<sup>[18]</sup>提出了弱估计器的特征指标,该特征重点考虑了随着时间推移而发展的网络中链接预测的相关趋势,并将该指标作为构建深度神经网络的有效特征向量进行学习.Yaghi 等人<sup>[19]</sup>通过提取网络中节点度中心性(DC)和紧密度中心性(CC)两个特征来增强给定边的重要性,并使用机器学习模型进行链接预测.Abuoda 等人<sup>[20]</sup>进一步对网络拓扑结构进行分析,并提出网络模式(Motif)结构应用于链接预测.Chen 等人<sup>[21]</sup>提出了一种基于图卷积网络(GCN)的嵌入式长短期记忆网络(LSTM),用于端到端动态链接预测.该方法在结构相似性链接预测方法的基础之上进一步获取网络特征,使用机器学习模型进行训练学习完成链接预测任务.但是多数的研究仅仅是更改已有的机器学习模型,缺乏对高阶网络结构中存在的特性和网络结构演变信息进一步的分析,以至于达不到预期的预测效果.

虽然当前的链接预测算法有很多,但面对上述各种链接预测的研究仍存在不足.一方面,现有的大多数链接预测算法只考虑了网络中很少一部分的局部结构特征,这些算法基本上都是从节点相似性角度进行度量.例如计算共同邻居、余弦相似性等局部结构特征,这些特征往往仅适用于节点对之间的相似性度量,但是在一些复杂的网络中,高阶结构间的特性远不止如此.另一方面,现有的链接预测算法很少考虑网络中的时序信息,以及随着时间推移网络不断变化的信息.尽管一些学者尝试使用长短期记忆网络(LSTM)解决网络结构演化过程中的时间依赖性问题,但是无法具体地分析网络演化规律,缺乏对表示时序信息的特征做出进一步解释.

本文提出的链接预测算法与上述算法不同之处如下.

- (1) 传统的链接预测算法主要提取网络低阶结构之间的相似性特征,而忽略高阶结构中存在的特征;本文通过研究高阶结构的特征来获得对网络高阶结构的新见解.本文提出了 Motif 聚集系数的方法,通过计算网络高阶结构的 Motif 聚集系数来捕获高阶结构中存在的聚类信息,从而丰富网络结构的特征;
- (2) 当前,一些算法使用 LSTM 模型解决网络结构演化过程中的时间依赖性问题,但是无法区分网络中哪些表示网络结构演变的特征在预测任务中起到决定性作用,缺乏对网络结构演化规律的有效分析;本文重点在于分析网络中存在的可表示性特征,因此本文通过对动态网络进行分析,发现节点间的链接频率和链接趋势等表示网络结构演变的重要特征,可以更加细致地解释网络结构变化信息.

## 2 问题定义

本文研究的重点是预测多个节点间产生链接形成高阶结构,因此我们通过研究三角形结构来确定高阶链接预测的模式.为了使读者更容易理解本文的研究内容,本节先介绍几个相关概念,针对高阶链接预测任务进行形式化定义.

**定义 1(单形).** 单形表示一个包含  $k$  个节点的集合,每个单形对应一个时间戳,具体表示为  $\{(S_i, t_i)\}_{i=1}^N, |S_i|=k$ , 我们称  $S_i$  为单形,  $t_i$  为该单形对应的时间戳。

**定义 2(闭合三角形).** 如果有 3 个节点存在于同一个单形,例如  $\{u, v, w\} \subset S_i$ , 则这 3 个节点形成一个闭合三角形,具体可表示为组成闭合三角形的 3 个节点至少同时出现在一个单形中的情况。

**定义 3(开放三角形).** 如果 3 个节点中的任意一对节点至少同时出现在一个单形中,但不存在某个单形同时包含这 3 个节点,我们称这 3 个节点形成一个开放三角形。

**定义 4(简单闭合事件).** 在  $t$  时刻之前,节点间可以任意的产生链接,但是没有任何一个单形同时包含  $\{u, v, w\}$  这 3 个节点,如果在  $t$  时刻之后存在一个单形同时包含了这 3 个节点,则该过程称为简单闭合事件。

**定义 5(高阶链接预测).** 为了简化表示,本文将高阶链接预测定义为预测节点三元组上的简单闭合事件.简单来说,我们研究的高阶链接预测问题是预测哪些尚未同时出现在同一个单形中的 3 个节点将来会同时出现在某一个单形中,具体表示为预测哪些开放三角形在未来会经历简单闭合事件。

### 3 基于 Motif 聚集系数与时序划分的高阶链接预测模型

为了充分利用并整合动态网络中高阶结构特征与结构演变特征,以提高链接预测模型的性能,本文提出一种基于 Motif 聚集系数与时序划分的高阶链接预测模型(MTLP).本节首先介绍该模型的基本思想,然后分别介绍 Motif 聚集系数与网络结构演变特征的计算方法,最后给出基于多层感知器网络的链接预测算法实现。

#### 3.1 MTLP模型的基本思想

本文通过构建 MTLP 模型来完成预测任务,该模型主要由 3 部分组成,分别是提取网络高阶结构 Motif 聚集系数特征、提取网络结构演化特征和构建基于 MLP 的链接预测算法,模型的整体框架如图 1 所示。

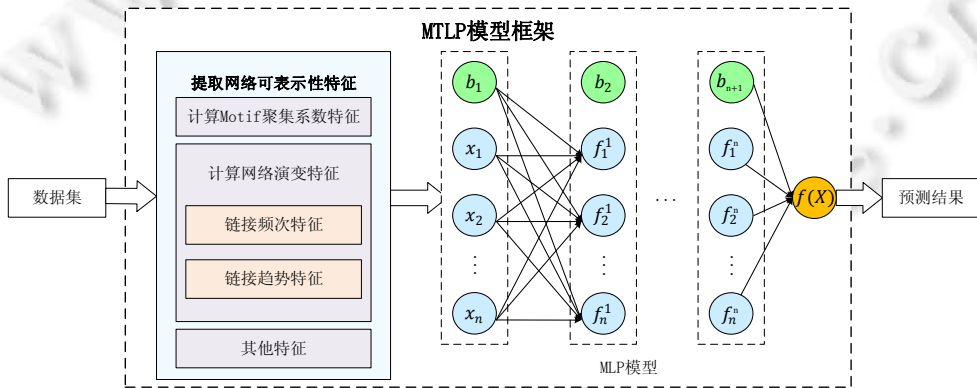


Fig.1 MTLP model framework

图 1 MTLP 模型框架图

- (1) 聚集系数是网络科学中数据建模的重要统计数据<sup>[22,23,24]</sup>,也是机器学习中非常有用的特征,例如在角色发现<sup>[25]</sup>和异常检测<sup>[26]</sup>中的应用.但是传统聚集系数本身具有限制条件,因为它仅限于测量一个简单楔形结构的闭合概率,本质上表示网络中单个节点的聚集程度,无法表示网络中高阶结构的聚集特性.本文重点研究基于高阶结构的链接预测任务,而传统的局部聚集系数并不适用于表示高阶结构的聚集行为.因此,将测量单个节点的局部聚集系数扩展至测量高阶结构的聚集系数,是提取 Motif 聚集系数特征模块主要解决的问题;
- (2) 链接预测的相关方法,本质上都是利用网络历史信息并将其压缩为一个网络,以预测下一时刻的网络.可以说,目前大多数的链接预测都是基于网络快照的方法,例如在给定的时间内分析给定的网络快照,用于下次链接预测.但是这些方法仅考虑了前一时刻的网络拓扑结构信息,而与前一时刻网络的动态发展无关.因此,通过对动态网络进行分析并捕获网络结构演变信息,是提取网络结构演化特

征模块主要解决的问题;

- (3) 本文研究的问题是预测一个三角形是否会经历简单闭合事件,该问题实际上是二分类问题.因此,本文使用 MLP 算法来完成链接预测任务.

### 3.2 提取Motif聚集系数特征

聚集系数<sup>[22,23,24]</sup>是用来量化网络中节点聚集程度的度量标准.本文扩展了传统的局部聚集系数,将单个节点的聚集系数扩展为高阶结构的聚集系数,用来衡量网络中高阶结构的聚集程度,即 Motif 聚集系数.

本文主要研究的是无向图网络,定义  $G=(V,E)$ 表示无向图, $n=|V|$ 表示无向图中的总节点数, $m=|E|$ 表示无向图中的总边数,任意节点  $u$  的邻居节点集合为  $N(u)$ .定义  $C_{motif}(g)$ 为 Motif 聚集系数, $g$  表示任意高阶结构, $G_{total}(g)$ 表示  $g$  的全局聚集系数, $N(g)$ 表示  $g$  中所有节点的共同邻居节点, $|W_{i,j}|$ 表示节点  $i$  与节点  $j$  之间是否存在链接:如果存在,则 $|W_{i,j}|=1$ ;否则 $|W_{i,j}|=0$ .Motif 聚集系数的公式定义如下:

$$C_{motif}(g) = G_{total}(g) \times \frac{2 \sum_{i,j}^{N(g)} |W_{i,j}|}{|N(g)|(|N(g)|-1)} \quad (1)$$

通过上述的定义已经了解到如何计算 Motif 聚集系数,计算 Motif 聚集系数首先需要计算每个子图的全局聚集系数,再计算子图中每个高阶结构的共同邻居,从而得到最终结果.从公式上来看,直接进行计算的代价非常昂贵,因此,我们对算法进行了优化,通过维护一个集合用来存储每个节点的邻居节点,使时间复杂度降为  $O(n)$ ,极大地缩短了计算 Motif 聚集系数的时间,计算高阶结构的 Motif 聚集系数算法详见算法 1.

**算法 1.** 计算 Motif 聚集系数.

Input:高阶结构  $g$ ,邻居节点集合  $N$ ;

Output: $C_{motif}(g)$ (高阶结构  $g$  的 Motif 聚集系数).

1 算法 *CalculateMotifCluf*( $g,N$ )

2 定义 *triNum* 记录  $g$  中闭合三角形的个数, $G_{total}(g)$ 记录全局聚集系数

3 定义 *tempNum* 用于临时计数,定义 *conns* 用于记录共同邻居中相互链接的边数

4 **for**  $g$  中的任意 3 个节点( $i,j,k$ ) **do**

5 **if** ( $i,j,k$ )任意两个节点间存在链接 **then**

6 *triNum*++

7 **end if**

8 **end for**

9 **for**  $g$  中的任意一个节点  $i$  **do**

10 *tempNum* +=  $C_2^{N(i)}$

11 **end for**

12  $G_{total}(g) = \frac{3 \times triNum}{tempNum}$

13 **for**  $N(g)$ 中的任意两个节点( $i,j$ ) **do**

14 **if** ( $i,j$ )两个节点间存在链接 **do**

15 *conns*++

16 **end if**

17 **end for**

18  $C_{motif}(g) = G_{total}(g) \times \frac{2 \times conns}{|N(g)|(|N(g)|-1)}$

19 **return**  $C_{motif}(g)$

### 3.3 提取网络结构演变特征

网络演化的历史为理论上的预测提供了更多信息,同时也为网络分析增加了一个全新的维度.我们采用图序列来描述网络的动态变化,并分析随着时间的推移,节点结构之间关联性的演变过程.

首先在数据预处理部分,我们将网络切分成表示一系列图序列的时间片集合  $G=(G_1, G_2, \dots, G_{t-1}, G_t)$ , 其中,  $G_t=(V_t, E_t)$  表示第  $t$  个时间序列的网络图, 每个网络图对应一个邻接矩阵  $W_t$ . 动态网络的拓扑结构会随着时间  $t$  的变化而不同, 主要是随着时间的变化, 节点间的链接产生了不同的变化, 有的节点间添加新的链接, 有的节点间删除原有的链接, 在不同的时间序列中, 节点对之间的链接关系也会不同. 每个时间片中节点间的链接变化如图 2 所示, 其中, 虚线表示图中节点从  $G_1$  时刻演变到  $G_2$  时刻的映射.

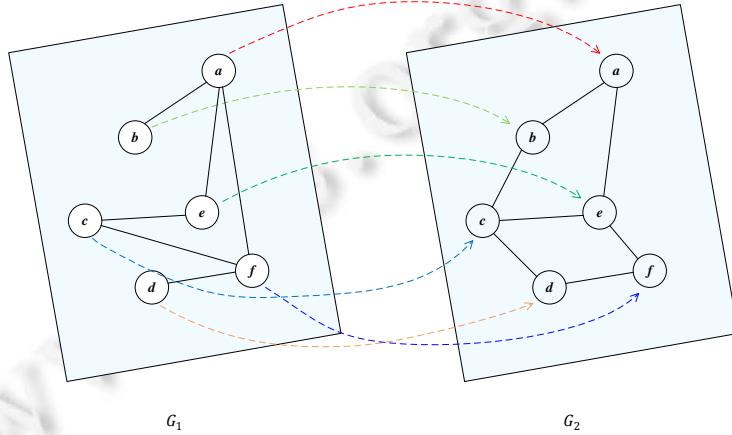


Fig.2 Representation of node link changes in different time slices  
图 2 不同时间片中节点链接变化表示

由图 2 可知:网络中的结构会随着时间的推移产生不同的变化,这些变化会在每个时间片中留下一系列的痕迹.本文主要从简单闭合事件的生命周期转换进行分析,通过考虑一系列时间序列的转换来分析节点间的结构转换关系是否能够对未来形成闭合三角形有重要的影响.通过分析发现:在大多数的数据集中,许多新形成的单形由  $k$  个节点组成,这些节点出现在同一个单形中之前也分布在不同的单形中,与其他节点产生交互.因此,本文认为,网络中产生开放三角形的一个可能假设是链接创建中的时间异步性.例如,开放三角形中的节点不能同时出现在同一个单形中,这些节点只能在不同时间段进行交互形成链接.不管开放三角形是如何创建的,随着网络的发展,这 3 个相关的节点在未来都可能在同一个单形中同时出现.通过对三角形经历简单闭合事件过程分析,每一种结构的变化对未来是否能够形成闭合三角形有着重要影响.因此,本文将整个结构演变过程进行时间上的划分,分别统计每个时间片中的局部结构关系,从而在时序维度提取重要的特征信息,主要提取了节点间链接频次和链接趋势两项特征用来表示网络结构变化规律.

在网络演化的进程中,不同节点间的链接频次各有不同.统计链接频次可以表示节点间的动态联系,因此,本文将链接频次作为动态网络中的一个重要特征.定义  $fs_n(i, j, k)$  为  $n$  个时间片中三角形的  $e_{ij}, e_{ik}, e_{jk}$  这 3 条边的链接频次算数和,其中,  $|W'_{i,j}|$  表示第  $t$  个时间片中节点  $i$  与节点  $j$  之间是否存在链接:如果存在,  $|W'_{i,j}|=1$ ; 否则,  $|W'_{i,j}|=0$ . 节点间的链接频次公式定义如下:

$$fs_n(i, j, k) = \sum_{(u,v)}^{(i,j,k)} \sum_t^n |W'_{u,v}| \tag{2}$$

网络结构在演化时呈现不同的链接规律,这些规律既具有确定性又具有不确定性,因此,可以将每条边的链接趋势作为一项重要的链接指示特征进行归纳.定义  $f'_i(i, j, k) = \left| \sum_{(u,v)} W'_{i,j} - \sum_{(u,v)}^{(i,j,k)} W'_{i,j} \right|$  表示第  $t$  个时间片与相邻时

间片中三角形边数的链接趋势,链接趋势减少为-1,链接趋势不变为 0,链接趋势增长为 1.本文将数据切割成  $n$  个时间片,通过计算  $n$  个时间片中三角形 3 条边的链接趋势的算数平均值,作为节点间链接变化的最终链接趋势,节点间的链接趋势公式定义如下:

$$ft_n(i, j, k) = \frac{1}{n} \sum_{t=1}^n f_t'(i, j, k) \quad (3)$$

节点间的链接频次和链接趋势可以有效地捕捉到动态网络演化规律,同时,这些特征也可以通过简单的计算将其表示出来.因此,本文将在后续的链接预测模型中使用这些可表示网络结构演变的特征,从而有效捕获网络演变信息.

### 3.4 构建基于MLP的链接预测算法

本节主要介绍构建基于 MLP 的链接预测算法,用于完成高阶链接预测任务,算法流程图如图 3 所示.

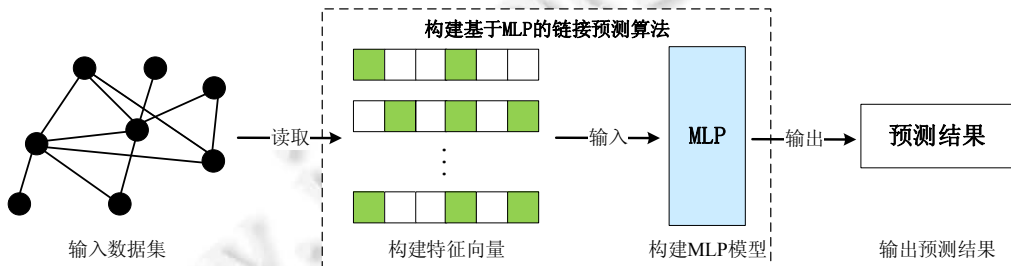


Fig.3 Flow chart of constructing MLP-based link prediction algorithm

图 3 构建基于 MLP 的链接预测算法流程图

#### 3.4.1 构建特征向量

在机器学习领域,只有数据和特征才能决定机器学习的上限,一个优秀的模型和算法只能不断地逼近这个上限.因此,选择具有代表性的特征也是一项非常重要的工作.通过前文的分析,本文已经提取出动态网络中 Motif 聚集系数特征和网络结构演变相关特征.除此之外,我们也加入了一些传统链接预测经常使用的特征,例如节点的邻居和节点的度等相似性特征,通过融合这些特征构建可供机器学习模型训练的特征向量.特征向量构建算法详见算法 2.

**算法 2.** 构建特征向量.

Input:单形  $S_i$  列表,时间列表  $t_i$ ,切割次数  $n$ ;

Output:特征向量  $LS$ .

- 1 算法  $FeatureMatrix(S_i, t_i, n)$
- 2 构建加权邻接矩阵  $W$ 、单形集合  $R$ 、节点度集合  $N$ 、加权邻接矩阵  $W_i$ 、三元组列表  $Triangles$
- 3 **for**  $Triangles$  中的每个三元组  $(i, j, k)$  **do**
- 4 定义  $vt$  用来存储三元组的特征向量
- 5 **for**  $(i, j, k)$  中的任意两条边  $(u, v)$  **do**
- 6  $vt \leftarrow \text{append } |w_{uv}|$  //同时包含节点  $u$  和  $v$  的单形
- 7  $vt \leftarrow \text{append } |N(u) \cap N(v)|$  //节点  $u$  和  $v$  的共同邻居
- 8 **end for**
- 9 **for**  $(i, j, k)$  中的任意一个顶点  $u$  **do**
- 10  $vt \leftarrow \text{append } |N(u)|$  //节点  $u$  的度
- 11  $vt \leftarrow \text{append } |R(u)|$  //包含节点  $u$  的单形
- 12  $vt \leftarrow \text{append } \sum_v^{N(u)} |N(v)|$  //节点  $u$  的二阶邻居度

```

13  end for
14  vt←append |N(i)∩N(j)∩N(k)|      //节点 i,j,k 的共同邻居
15  vt←append fsn(i,j,k)           //三角形边的链接频次
16  vt←append ftn(i,j,k)           //三角形边的链接趋势
17  vt←append Cmotif(Δ)            //三阶 Motif 聚集系数
18  LS←append vt
19  end for
20  return LS
    
```

这些特征主要分为 3 类:第 1 类为大多数链接预测算法所考虑的网络拓扑结构相似性特征,例如节点的度以及节点的邻居等;第 2 类为高阶结构特征,例如高阶 Motif 聚集系数,该特征有效捕获了高阶结构的聚集特性;第 3 类为网络演变过程中可表示结构变化的特征,例如节点间的链接频次和链接趋势等.这些特征都存在于现实网络中,可以有效地表示网络结构间的关系.

### 3.4.2 构建链接预测模型

针对链接预测问题,本文选用多层感知器(MLP)网络完成预测任务,MLP 算法自从被提出以来,主要被应用于节点分类等有监督学习任务.该算法的核心是训练学习函数  $f(X):R^m \rightarrow R^n$ ,其中,  $m$  是输入的维数,  $n$  是输出的维数.具体表示为:给定一组特征向量特征  $x$  和标签  $y$ ,它可以学习用于回归或分类的非线性函数.

图 4 列举了一个具有标量输出的单隐藏层 MLP 模型:模型最左边一组代表输入特征的神经元构成输入层;模型中间多个有向层中分布的神经元构成整个隐藏层,每个隐藏层中的神经元将前一层的值进行加权线性求和转换,再使用非线性激活函数进行计算得到该神经元的输入数据;模型最右边为输出层,主要用于接收由最后一个隐藏层经过变换计算输出的值.

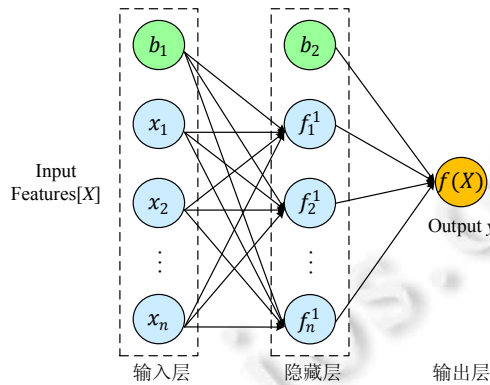


Fig.4 Single hidden layer MLP

图 4 单隐藏层 MLP

MLP 模型具体用法是:给定一组训练样本  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ,其中,  $x_i \in R^n, y_i \in \{0, 1\}$ ,则单隐藏层 MLP 模型学习到的函数为  $f(x)$ :

$$f(x) = W_2 g(W_1^T x + b_1) + b_2 \tag{4}$$

其中,  $W_1 \in R^m$  和  $W_2, b_1, b_2 \in R$  是模型参数.  $W_1, W_2$  分别是输入层与隐藏层之间和隐藏层与输出层之间的权重,  $b_1, b_2$  分别是隐藏层和输出层的偏置值,  $g(x)$  是激活函数.本文使用的激活函数为双曲正切函数,计算公式如下:

$$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{5}$$

对于二分类问题,  $f(x)$  通过 logistic 函数  $g(z) = \frac{1}{1 + e^{-z}}$  能够得到 0~1 之间的输出值.若将阈值设置为 0.5, 则输出大于等于 0.5 的样本分到正类(positive class),其他情况分为负类(negative class).同时,一般情况下,MLP



算法会根据不同问题使用不同的损失函数.本文的高阶链接预测为二分类问题,因此采用交叉熵损失函数,计算公式如下:

$$Loss(\hat{y}, y, W) = -y \ln \hat{y} - (1 - y) \ln(1 - \hat{y}) + \alpha \|W\|_2^2 \quad (6)$$

其中,  $\hat{y}$  是经过 logistic 函数计算后的预测值,  $y$  是数据的实际值;  $\alpha \|W\|_2^2$  为  $L2$  正则化的模型复杂度惩罚项,  $\alpha > 0$  为非负的超参数,主要用来调节惩罚程度的大小;  $W$  为权重矩阵.从随机初始化权重开始,MLP 算法通过不断更新这些权重值来最小化损失函数,得到最终的预测模型.

本文使用 MLP 模型的输入参数为算法 2 构建的特征向量,同时采用 3 个全连接的隐藏层进行计算,模型的输出为当前开放三角形是否会经历简单闭合事件的状态.MLP 模型最终输出结果为 0 和 1,其中,1 表示该开放三角形会经历简单闭合事件,而 0 表示该开放三角形不会经历简单闭合事件.

## 4 实验结果与分析

本节详细介绍了数据集、实验环境、评价指标等内容.首先,对 Motif 聚集系数特征和网络结构演变等特征进行了有效性验证;接着,通过与不同链接预测算法进行对比,对本文提出的 MTLP 模型进行综合评价.

### 4.1 实验数据集

本文采用来自 5 个领域的 14 个数据集<sup>[27]</sup>,分别是合著网络(coauth)、在线标记网络(tags)、在线参与网络(threads)、电子邮件网络(email)和近距离接触网络(contact).

每个数据集是一组带有时间戳的节点集.这些数据集分布在不同的领域,因此包含了丰富多样的结构,非常适合多种实验测试,每个数据集的详细信息见表 1.

Table 1 Data set information

表 1 数据集信息

数据集	节点数	边数	单形数	数据集	节点数	边数	单形数
coauth-DBLP	1 924 991	7 904 336	3 700 067	NDC-substances	5 311	88 268	112 405
coauth-MAG-Geology	1 256 385	5 120 762	1 590 335	NDC-classes	1 161	6 222	49 724
coauth-MAG-History	1 014 734	1 156 914	1812 511	DAWN	2 558	122 963	2 272 433
tags-math-sx	1 629	91 685	822 059	email-Enron	998	29 299	234 760
tags-ask-ubuntu	3 029	132 703	271 233	email-Eu	143	1 800	10 883
threads-math-sx	176 445	1 089 307	719 792	contact-high-school	327	5 818	172 035
threads-ask-ubuntu	125 602	187 157	192 947	contact-primary-school	242	8 317	106 879

### 4.2 评价指标

本文使用 ROC 曲线和 AUC-PR 值<sup>[28]</sup>作为评价指标对实验结果进行度量.

ROC 是由 True Positive Rate(TPR)和 False Positive Rate(FPR)组成的曲线,其中,TPR 代表分类器预测的正类中实际经历简单闭合事件的正实例占有所有正实例的比例,也可表示为正确预测的覆盖程度;FPR 代表分类器预测的正类中实际未经简单闭合事件的负实例占有所有负实例的比例,也可表示为预测虚报的程度.两项指标的计算公式如下:

$$TPR = \frac{TP}{TP + FN} \quad (7)$$

$$FPR = \frac{FP}{FP + TN} \quad (8)$$

对于链接预测模型而言,根据预测时设定的阈值,在测试样本集上会得到 TPR 和 FPR 的点对.如果遍历全部阈值,就可以得到多个点对,将点对连接便得到一条经过(0,0),(1,1)的曲线,这就是此链接预测模型的 ROC 曲线.同时,我们当然希望虚报的越少越好,覆盖的越多越好.因此,FPR 越低,TPR 越高,ROC 曲线越接近左上角,意味着预测模型的性能越好.

由于很多时候 ROC 曲线并不能明确地指出哪个分类器的效果更好,而 AUC-PR 指标具有明确的数值,对应 AUC-PR 值越大的分类器效果越好,因此本文也使用 AUC-PR 值作为评价模型性能的标准.同时,AUC-PR 适用

于类不平衡的预测问题,我们的数据集就是这种情况,因此该指标可以充分地对比不同模型的性能。

同时,本文使用随机分数作为基准,该分数为测试集中经历简单闭合事件的三角形个数与训练集中开放三角形个数的比例。随机分数以绝对值列出,通过比较每个模型的 AUC-PR 与基准的相对值来衡量不同模型的性能,分值越大,则表示模型的性能越好。

### 4.3 验证特征有效性

本节主要验证 Motif 聚集系数特征和网络结构演变相关特征的有效性,主要是从 Motif 聚集系数特征和网络结构演变特征两个维度对数据进行系统域划分,通过训练一个多项式逻辑回归模型来确定不同数据集的系统域。

本文采用 scikit-learn 库中的多项式逻辑回归模型进行训练学习,训练的特征向量包含开放三角形的比例、开放三角形的平均 Motif 聚集系数和开放三角形的网络结构演变这 3 项特征和一个截距项。为了防止模型过拟合,采用  $l_2$  正则化进行约束,最终训练出一个多项式逻辑回归分类器来预测划分的系统域。为了直观地表示出系统域的划分,在这里给出一个使用开放三角形比例和开放三角形的平均 Motif 聚集系数作为协变量训练出的系统域划分图。从图 5 上可以观察到:本文提出的 Motif 聚集系数可以很好地将不同的数据集划分到各自的领域,从而体现出 Motif 聚集系数可以作为网络中高阶结构的一项重要的特征。

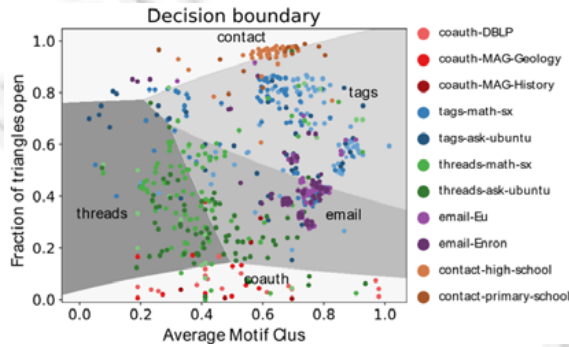


Fig.5 System domain division diagram

图 5 系统域划分图

为了更精准地验证本文提出的特征有效性,我们采用消融实验,分别控制平均 Motif 聚集系数特征和网络结构演变特征这两个变量进行实验,用于验证系统域划分的准确性,实验结果见表 2。

Table 2 Comparison of accuracy of different features

表 2 不同特征准确率对比

模型特征			截距	准确率	
开放三角形比例	平均 Motif 聚集系数	平均链接趋势	截距	随机	多项式 LR
√	√	√	√	0.21	0.70±0.02
√	√	-	√	0.21	0.67±0.02
√	-	√	√	0.21	0.62±0.02
√	-	-	√	0.21	0.51±0.02

从表 2 可以看出,任意添加一项平均 Motif 聚集系数特征或者网络结构演变特征都可以提高数据域划分的准确性。尽管最终划分数数据域的结果并不完美,但是已经可以说明本文提出的 Motif 聚集系数与网络结构演变相关特征的有效性。

### 4.4 不同链接预测算法性能对比

#### 4.4.1 对比方法

本文选取了 10 个链接预测模型用于对比实验,其中前 7 个为经典的链接预测模型,后 3 个为近两年最新的

基于机器学习的链接模型,分别如下.

- Harm.Mean<sup>[27]</sup>是基于启发式的结合边强度指示的相似性模型,主要是通过计算三角形中 3 条边对应的权值相加和的倒数,作为一项相似性指标用于链接预测;
- Geom.mean<sup>[27]</sup>是基于启发式的结合边强度指示的相似性模型,主要是通过计算三角形中 3 条边对应的权值连乘积的三次方根,作为一项相似性指标用于链接预测;
- Arith.mean<sup>[27]</sup>是基于启发式的结合边强度指示的相似性模型,主要是通过计算三角形中 3 条边对应的权值相加的均值,作为一项相似性指标用于链接预测;
- Adamic-Adar<sup>[1]</sup>是基于邻居节点的相似性模型,主要通过计算三角形中 3 条边邻居的对数倒数和,作为一项相似性指标用于链接预测;
- Pref.Attach<sup>[1]</sup>是基于邻居节点的相似性模型,主要通过计算三角形中 3 条边邻居的乘积,作为一项相似性指标用于链接预测;
- Katz similarity<sup>[1]</sup>是基于路径的相似性模型,主要通过计算三角形中 3 条边的路径和,作为一项相似性指标用于链接预测;
- PPR<sup>[1]</sup>是基于路径的相似性模型,主要通过计算三角形中 3 条边对应的随机游走的概率,作为一项相似性指标用于链接预测;
- Benson 等人<sup>[27]</sup>发现,连接强度和边缘密度是高阶组织链接的积极指标,并使用这些指标进行链接预测;
- Yaghi 等人<sup>[19]</sup>通过构造节点的中心性度量指标来提高预测的准确性;
- Chiu 等人<sup>[18]</sup>提出了网络链接中弱估计指标用于链接预测,从而提高预测性能.

#### 4.4.2 对比实验

本文首先选取对比方法中前 7 个经典的链接预测模型与本文提出的 MTL P 模型进行对比实验,并且使用 ROC 曲线来评价每个模型的性能.这些模型都是已知的模型,可在链接预测任务中提供良好的预测效果.为了验证不同模型的性能,我们选取几个典型的数据集进行对比实验,ROC 对比结果如图 6 所示.

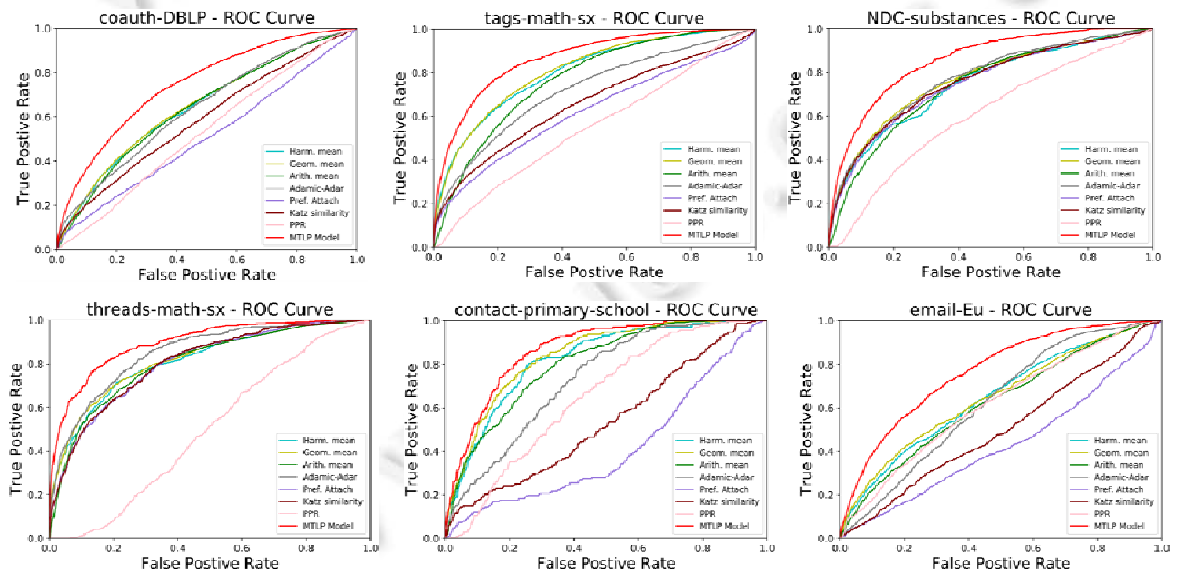


Fig.6 ROC comparison results

图 6 ROC 对比结果

从实验结果可以看出:本文提出的 MTL P 模型在不同数据集上都表现出了较优的性能,主要表现为 ROC 曲线下的面积越大,性能越优.

除了与经典的链接预测模型进行对比,我们也与近两年最新的基于深度学习的链接预测算法进行对比.为了更加详细地对比每一个模型的性能,本文同时考虑以上 10 个链接预测模型,并使用随机模型分数作为基准,使用 *AUC-PR* 指标来评价每个模型的预测性能.实验结果见表 3,图表中黑色加粗的数字表示在该数据集上链接预测性能最优的模型得分.

Table 3 Experimental results AUC-PR value

表 3 实验结果 AUC-PR 值

Dataset	Random	Harm.mean	Geom.mean	Arith.mean	Adamic-Adar	Pref.Attach
coauth-DBLP	1.68E-03	1.49	1.59	1.50	1.60	0.74
coauth-MAG-History	7.16E-04	1.69	2.72	3.20	5.82	2.49
coauth-MAG-Geology	3.35E-03	2.01	1.97	1.69	2.71	0.97
tags-math-sx	1.08E-03	9.08	8.67	2.88	6.34	2.81
tags-ask-ubuntu	1.08E-03	12.29	12.64	4.24	7.51	5.63
threads-math-sx	5.63E-05	20.86	16.01	5.03	23.32	7.46
threads-ask-ubuntu	1.31E-04	78.12	<b>80.94</b>	29.00	30.82	6.62
NDC-substances	1.17E-03	4.90	5.27	2.90	5.97	4.46
NDC-classes	6.72E-03	<b>4.43</b>	3.38	1.82	0.99	2.14
DAWN	8.47E-03	4.43	3.86	2.13	<b>4.77</b>	1.45
email-Enron	1.40E-02	1.78	1.62	1.33	0.87	0.83
email-Eu	5.34E-03	1.98	2.15	1.78	1.37	1.55
contact-high-school	2.47E-03	3.86	<b>4.16</b>	2.54	2.00	1.13
contact-primary-school	2.59E-03	5.63	6.40	3.96	3.21	0.94

Table 3 Experimental results AUC-PR value (Continued)

表 3 实验结果 AUC-PR 值(续)

Dataset	Katz similarity	PPR	Benson	Yaghi	Chiu	本文
coauth-DBLP	1.51	1.83	<b>3.37</b>	1.21	1.11	3.34
coauth-MAG-History	3.40	1.88	6.75	0.98	0.77	<b>7.44</b>
coauth-MAG-Geology	1.74	1.26	4.74	0.92	1.55	<b>4.81</b>
tags-math-sx	2.71	1.55	13.99	0.88	12.07	<b>17.67</b>
tags-ask-ubuntu	4.15	2.54	7.48	0.61	9.85	<b>32.09</b>
threads-math-sx	4.86	1.18	<b>47.18</b>	0.61	32.75	42.09
threads-ask-ubuntu	32.31	1.51	9.82	0.64	13.54	4.75
NDC-substances	2.93	1.83	8.17	0.65	0.78	<b>8.36</b>
NDC-classes	1.34	0.91	0.62	0.96	0.9	0.66
DAWN	2.04	1.37	2.86	0.58	0.67	2.20
email-Enron	1.28	3.16	0.72	2.54	1.46	<b>3.22</b>
email-Eu	1.79	1.75	3.47	1.40	0.76	<b>3.61</b>
contact-high-school	2.53	2.41	2.86	0.79	0.72	2.96
contact-primary-school	4.02	4.31	6.91	0.98	0.92	<b>7.64</b>

从上面的实验结果可以清楚地看到:虽然没有任何一个模型在所有数据集上都能表现最佳,但本文提出的算法可以在大多数的数据集上表现出更好的预测性能,例如 coauth-MAG-History,coauth-MAG-Geology,tags-math-sx 和 tags-ask-ubuntu 数据集.这说明本文提出的算法有效捕获了网络的聚集特性和结构演变特性,从而得到较好的实验结果.但是在一些数据集上表现的结果较差,例如 threads-ask-ubuntu 数据集,我们通过对该数据集分析发现:该数据集聚集程度较低,同时开放三角形的比例也较低,如果开放三角形比例较低,会出现数据分布不均匀的现象,这是导致预测结果较差的主要原因.值得注意的是:从不同的实验结果可以看出,使用不同相似性指标的模型也可以在部分数据集上表现出了较好的结果.例如基于启发式的结合边强度指示的相似性模型在 threads-ask-ubuntu,NDC-classes 和 contact-high-school 数据集上表现出了较好的预测性能,这说明网络中节点之间的链接强度也可以精准地捕获网络结构信息.但是在一般情况下,单个特征指标仅仅在特定的网络中起到作用,并不能完全捕获整个网络结构的特性.而且这些特征无法有效表示高阶结构的特征,因此在大多数的数据集上表现不出较为优异的结果.

为了进一步分析本文提出的算法在部分数据集上表现较差的原因,我们从数据集本身进行分析,通过分析每个网络的聚集分布发现,一些预测结果较好的数据集的网络聚集程度较低,例如 coauth-MAG-History,coauth-MAG-Geology 和 NDC-substances 数据集.同时,一些预测结果较差的数据集的网络聚集程度较高,例如 NDC-

classes 和 DAWN 数据集.这表明:如果节点间的聚集程度太高,整个网络趋向于完全图,则网络结构的聚集系数仅有较弱的参考价值.但是也存在例外,例如 `contact-primary-school` 数据集聚集程度较高,但是实验结果也表现出了较好的性能.通过对该数据集分析,我们发现该数据集有较长的时间跨度,从而产生了丰富多样的结构.本文提出的时序特征主要通过将数据集进行切片划分从而捕获网络结构演变的关系,因此较长的时间跨度可以使得数据切分更加充分,进而促进划分后的网络结构演变特征充分发挥了作用.因此,数据集自身存在的特性也会直接导致本文提出的模型在部分的数据集上没有表现出较好的预测性能,这也是我们今后需要继续关注的问题.

虽然本文提出的算法无论与经典的链接预测算法比较还是与最新的基于机器学习的链接预测算法比较,性能上都表现的不错,但是这并不能说明本文提出的算法性能已经非常完美,因为高阶链接预测任务还具有挑战性.但是我们的目标是确定高阶链接预测中一些重要结构特征,而不是精确地进行预测.因此,我们将探索网络中更多潜在的可表示性特征作为未来的工作进行继续研究.

## 5 总结与展望

本文重点分析了链接预测的特点以及目前面临的难点问题,分析了网络中高阶结构的基本结构和特性,说明了在动态网络中进行链接预测的目的和意义.首先,从网络结构方面进行分析,提出了 Motif 聚集系数,用于衡量高阶结构的聚集程度;其次,从时序维度分析,提出了可表示网络演变规律的节点间的链接频率和链接趋势等特征;最后,得益于神经网络对数据的建模能力,本文将高阶链接预测问题转换为二分类问题,从而采用多层感知器网络完成预测任务.为了进一步探索网络中更多潜在的可表示性特征,本文未来工作将会重点分析网络中的其他高阶结构特征,例如高阶结构中的结构中心性和结构相似性等特性,从而进一步挖掘网络中潜在的信息.

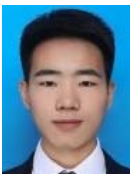
## References:

- [1] Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 2007,58(7):1019–1031.
- [2] Backstrom L, Leskovec J. Supervised random walks: Predicting and recommending links in social networks. In: *Proc. of the 4th ACM Int'l Conf. on Web Search and Data Mining*. 2011. 635–644.
- [3] Wang X, Gulbahce N, Yu H. Network-Based methods for human disease gene prediction. *Briefings in Functional Genomics*, 2011, 10(5):280–293.
- [4] Tang J, Wu S, Sun J, *et al.* Cross-domain collaboration recommendation. In: *Proc. of the 18th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. 2012. 1285–1293.
- [5] Benson AR, Gleich DF, Leskovec J. Higher-order organization of complex networks. *Science*, 2016,353(6295):163–166.
- [6] Newman MEJ. Clustering and preferential attachment in growing networks. *Physical Review E*, 2001,64(2):025102.
- [7] Adamic LA, Adar E. Friends and neighbors on the Web. *Social Networks*, 2003,25(3):211–230.
- [8] Barabási AL, Jeong H, Neda Z, *et al.* Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and Its Applications*, 2002,311(3-4):590–614.
- [9] Katz L. A new status index derived from sociometric analysis. *Psychometrika*, 1953,18(1):39–43.
- [10] Wang C, Satuluri V, Parthasarathy S. Local probabilistic models for link prediction. In: *Proc. of the 7th IEEE Int'l Conf. on Data Mining (ICDM 2007)*. IEEE, 2007. 322–331.
- [11] Page L, Brin S, Motwani R, *et al.* The Pagerank citation ranking: Bringing order to the Web. *Stanford InfoLab*, 1999.
- [12] Bahmani B, Chowdhury A, Goel A. Fast incremental and personalized pagerank. *Proc. of the VLDB Endowment*, 2010,4(3): 173–184.
- [13] Li D, Shen DR, Kou Y, *et al.* Research on a link-prediction method based on a hierarchical hybrid-feature graph. *SCIENTIA SINICA Informationis*, 2020,50(2):221–238 (in Chinese with English abstract).
- [14] Clauset A, Moore C, Newman MEJ. Hierarchical structure and the prediction of missing links in networks. *Nature*, 2008,453(7191): 98–101.

- [15] Tian T, Yang YL, Guo J, *et al.* Link prediction of brain networks based on hierarchical random graph model. *Application Research of Computers*, 2016,33:1066–1069 (in Chinese with English abstract).
- [16] Das S, Das SK. A probabilistic link prediction model in time-varying social networks. In: *Proc. of the 2017 IEEE Int'l Conf. on Communications (ICC)*. IEEE, 2017. 1–6.
- [17] Liu L, Wang YY, Ni QX, Cao J, Bu Z. A link prediction approach in temporal networks based on game theory. *Journal of Computer Research and Development*, 2019,56(9):1953–1964 (in Chinese with English abstract).
- [18] Chiu C, Zhan J. Deep learning for link prediction in dynamic networks using weak estimators. *IEEE Access*, 2018,6:35937–35945.
- [19] Yaghi RI, Faris H, Aljarah I, *et al.* Link prediction using evolutionary neural network models. In: *Proc. of the Evolutionary Machine Learning Techniques*. Singapore: Springer-Verlag, 2020. 85–111.
- [20] Abuoda G, Morales GDF, Abounlaga A. Link prediction via higher-order motif features. *arXiv preprint arXiv:1902.06679*, 2019.
- [21] Chen J, Xu X, Wu Y, *et al.* GC-LSTM: Graph convolution embedded lstm for dynamic link prediction. *arXiv preprint arXiv:1812.04206*, 2018.
- [22] Schaeffer SE. Graph clustering. *Computer Science Review*, 2007,1(1):27–64.
- [23] Newman MEJ. Random graphs with clustering. *Physical Review Letters*, 2009,103(5):058701.
- [24] Seshadhri C, Kolda TG, Pinar A. Community structure and scale-free collections of Erdős-Rényi graphs. *Physical Review E*, 2008,34(5):700–720.
- [25] Henderson K, Gallagher B, Eliassi-Rad T, *et al.* Rolx: Structural role extraction & mining in large graphs. In: *Proc. of the 18th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. ACM, 2012. 1231–1239.
- [26] LaFond T, Neville J, Gallagher B. Anomaly detection in networks with changing trends. In: *Proc. of the Outlier Detection and Description under Data Diversity at the Int'l Conf. on Knowledge Discovery and Data Mining*. 2014.
- [27] Benson AR, Abebe R, Schaub MT, *et al.* Simplicial closure and higher-order link prediction. *Proc. of the National Academy of Sciences*, 2018,115(48):E11221–E11230.
- [28] Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: *Proc. of the 23rd Int'l Conf. on Machine Learning*. 2006. 233–240.

#### 附中文参考文献:

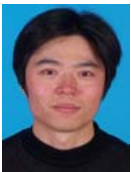
- [13] 李冬,申德荣,寇月,等.基于层次化混合特征图的链路预测方法. *中国科学:信息科学*,2020,50(2):4.
- [15] 田甜,杨艳丽,郭浩,等.基于层次随机图模型的脑网络链路预测. *计算机应用研究*,2016,33(4):1066–1069.
- [17] 刘留,王煜尧,倪琦瑄,等.一种基于博弈论的时序网络链路预测方法. *计算机研究与发展*,2019,56(9):1953–1964.



康驻关(1996—),男,硕士,主要研究领域为图计算,链接预测.



王国仁(1966—),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为数据管理,大数据计算,知识图谱,区块链,生物信息学.



金福生(1977—),男,博士,副教授,CCF 高级会员,主要研究领域为大数据,区块链,人工智能.