基于划分社区和差分共邻节点贡献的链路预测*

伍杰华[1,2]

(1. 广东工贸职业技术学院 计算机工程系, 广州 510510; 2. 华南理工大学 信息科学与技术学院, 广州 510641)

摘 要: 通过改进基于节点相似度的朴素贝叶斯模型,引入 GN 和 CMN 两种经典的划分社区算法挖掘网络社区属性对预测节点对的影响,赋予共邻节点不同的连接度和社区贡献度并计算其贡献权重,同时把模型应用于五种相似度算法,采用 ROC 和 Precision-Recall 曲线进行实验评价。人工网络和真实网络中的实验证明,该模型能够在深入挖掘社会网络结构信息的基础上提高预测的精确度,同时为该类模型的研究提供一种新的方案。

关键词: 链路预测; 社会网络; 社区划分; 相似度算法; 共邻节点

中图分类号: TP391 **文献标志码**: A **文章编号**: 1001-3695(2013)10-2954-04

doi:10.3969/j.issn.1001-3695.2013.10.018

Link prediction based on partitioning community and differentiating role of common neighbors

WU Jie-hua[1,2]

(1. *Dept. of Computer Science & Engineering, Guangdong College of Industry & Commerce, Guangzhou* 510510, *China*; 2. *College of Information Science & Technology, South China University of Technology University, Guangzhou* 510641, *China*)

Abstract: This paper examined a new measure of link prediction based on an enhance local naive Bayesian model which applying two classic community partition algorithm: GN and CMN to mine network's communities attributes and impact on the predicted node, then entrusted common-neighbors connectivity and community participation degree to calculate the weight of their contribution, finally improved five similarity based algorithm and took ROC and Precision-Recall curve as experimental evaluation. Artificial networks and real network experiments show that the model can mine the latent social network structure information and enhance accuracy of link prediction.

Key words: link prediction; social network; community partition; similarity algorithms; common neighbors

社会网络分析主要研究社会中实体及其相互活动关系[1],这种关系和活动可以用网络图的结构[2]来表示,其中节点(顶点)表示一个参与者,链路(边)表示两个参与者之间的关系。利用网络图可以挖掘其结构特征并找出感兴趣的信息。链路预测是其开展研究的重点领域[3]。

链路预测根据网络的历史结构信息预测网络的演化及链路关系发生的潜在可能,其在众多领域有着重要的应用价值[4],例如通过分析生物信息网络,预测蛋白质的相互作用;在社交网络中预测现在尚未成为朋友的未来是否"应该是朋友";在市场营销网络中判断哪些客户应该需要去开发等。

1 相关工作

链路预测作为数据挖掘和社会网络分析研究的方向已经有一些工作正在开展,基于网络结构信息挖掘的链路预测则是其中的一个思路和方向,其主要包括:

a)基于节点相似度,相似度表示两个节点之间(后称节点对)结构属性的相似性及产生链路可能性。假设相似性越大,它们之间存在链路的可能性就越大。Liben-Nowell 等人[4]总结了基于网络拓扑结构的相似性定义方法,并将这些指标分为基于节点和基于路径的两类,并分析了若干指标对相互作用,网络中链路预测的效果。

b)网络层次聚类。基于网络层级结构的链路预测模型主要通过对网络中的节点构造成一棵生成树实现。Clauset 等人[5]提出了一种通过网络层次结构的模型,该模型通过蒙地卡罗采样,估计节点间产生链路概率的极大似然函数进行预测;Donghyuk 等人[6]提出了一种多尺度的链路预测模型,该模型通过对网络构建不同尺度的矩阵,然后低秩近似分解进行预测;Valverde-Rebaza 等人[7]提出一个基于聚类信息的模型,该模型通过挖掘网络的聚类信息进行预测,获得了较好的效果。

本文提出一种新的网络层次聚类思路进行链路预测——基于社区划分的链路预测模型。社区[8]由一组具备高密度且用户找到感兴趣的信息并帮助 Internet/Intranet 服务提供者有效地组织门户;同时它可以帮助制造商准确地找到消费者。对于链路预测,划分社区有助于挖掘社会网络结构属性并应用到基于节点相似度运算的模型当中[9]。如图 1 所示,社区内部节点链接比社区之间的链接更加紧密,同时处于同一个社区的节点对的共同邻接节点(后简称共邻节点)要比处于不同社区间节点对的要多,所以预测节点对受所在社区内部邻居节点的影响显然要比处于其他社区的邻居节点要大[7];况且两个节点之间的链接可能性还受整个社区的聚类系数和节点的数目影响,也有可能受社区划分评价属性模块度的影响。因此提出一种新的挖掘社会网络社区属性进行链路预测的算法十分必要。

在本文中首先给出社会网络模型和社区结构定义,然后分

收稿日期: 2013-01-13; **修回日期**: 2013-03-07 **基金项目**: 国家自然科学基金资助项目(61003045)

作者简介:伍杰华(1982-),男,广东广州人,高级工程师,讲师,博士,主要研究方向为机器学习、数据挖掘、社会网络分析(kodakwu@126.com).

析了 GN[10] 和 CMN[11] 两种经典的社区划分算法,并结合 CN 等五种基于共邻节点的相似度算法赋予具备不同社区属性邻接节点的贡献权重构建链路预测模型,同时结合 Precision-recall 和 ROC 曲线对人工网络和真实网络进行实验分析,从而把该方法推广到所有基于层次结构和相似度的链路预测模型中。

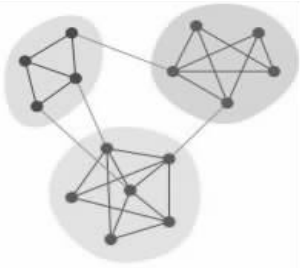

图1　社区图模型

综合来说,本文的贡献主要有以下三点:

a)采用社区划分算法能够挖掘网络中的社区属性并有效地应用于链接预测模型中。

b)引入了社区连接度和社区贡献度的概念,提出了差分化处理共邻节点对预测节点对权重贡献的模型,改进了原有链接预测算法中所有共邻节点贡献一致的假设,并将其推广到所有基于共邻节点的相似度算法中。

c)对人工和真实网络分析,了解到不同社区划分算法和差异化的社区结构属性,如模块度、聚类系数对链路预测的精确度会产生影响。

2　GN 和 CMN 社区划分算法

2.1　网络和社区模型

定义 1　$G=(V,E)$ 定义整体网络结构,其中,$N=\{n_1,n_2,n_3,\cdots,n_m\}$ 和 $E=\{(v_i,v_j)\mid v_i,v_j\text{has\_connect}\}$ 分别定义网络中节点和链路的集合,其中 (v_i,v_j) 定义为节点对,$\Gamma(v_i,v_j)$ 为节点对的共邻节点集合。

定义 2　划分社区结果为 $C=\{C_1,C_2,C_3,\cdots,C_n\}$,每一个 C_i 都有其包含的节点和链接。划分结果中不存在社区重叠的节点,同时任意的节点只属于一个社区,链接则可以连接同一个社区内的节点和不同社区之间的节点。

定义 3　r_{xy} 定义 x、y 两个节点间相似度值。

2.2　社区划分

2.2.1　GN 划分算法

GN 社区划分算法由 Newman 等人[10] 提出。该算法基本思想是:计算网络中所有节点对之间的边介数(betweenness),找到其中介数最高的边并将它从网络中移除,然后重新计算余下节点对的边介数值并进行迭代处理,直到每个节点就是一个退化社团为止,并提出可以用 shortest paths、resistor networks 和 random walks 三种算法定义边介数,采用计算网络中经过每条边的最短路径的数目 shortest paths 实现,时间复杂度较高。

2.2.2　CNM 划分算法

CNM 则由 Clauset 等人[11] 在贪婪凝聚算法的基础上提出,该算法采用最大堆结构不断更新社区划分模块度,使得在划分过程中模块度增加最多而减少最少。

2.2.3　模块度

由于社区内部节点间联系紧密而社区间节点联系较为稀疏,如果社区划分时较多的连接被分到社区内部,那么这样的划分显然是符合原理的。一般用模块度(modularity)的定义来衡量网络能否有效划分为社区度量。

3　基于社区划分差分化共邻节点贡献的链路预测

3.1　问题描述

给定一个网络 G 中不存在预测节点对 (x,y) 及其共同邻接节点集 $\Gamma(x,y)$,其中节点 x 属于社区 C_i,节点 y 属于社区 C_j,计算两者之间的相似度 r_{xy}。

3.2　基于社区角色定义

基于共邻节点的相似度算法假设所有节点对预测节点对的贡献权重视为一致,不利于区分具备不同属性共邻节点的角色及其贡献。针对这一点,Adamic-Adar(AA)和 resource allocation(RA)算法作出了改进[4],采用共邻节点的度数来差分其角色权重,但是该算法不能深入挖掘社会网络的结构属性。Lv 等人[12] 提出了一种基于朴素贝叶斯模型的节点角色定义算法,把共邻节点的角色 R_ω 定义为其在链接前提下聚类系数比率的和:

$$R_\omega=\prod_{i=1}^{n}\frac{P(e\mid\omega_i)}{P(\bar{e}\mid\omega_i)}=\frac{C_{\omega_i}}{1-C_{\omega_i}}\tag{1}$$

其中:e 是 (x,y) 存在链路;C_{ω_i} 为该节点的聚类系数。从而经典的 CN 算法改进后定义为

$$r_{xy}^{LBN-CN}=\mid\Gamma(x,y)\mid\times\log s+\sum_{\omega\in\Gamma(x,y)}\log R_\omega\tag{2}$$

其中:$\mid\Gamma(x,y)\mid\times\log s$ 是经典 CN 算法,计算共邻节点数目;$\sum_{\omega\in\Gamma(x,y)}\log R_\omega$ 相当于共邻节点贡献的权重和。

但是该算法从全局角度进行计算,而网络由各个社区所组成,由于社区具备相同的性质和兴趣,基于不同社区的不同共邻节点的角色不一致,无法差分化处理其贡献,同时文献[7]也提出,链路预测节点对受社区内共邻节点的影响要比处于社区外的共邻节点要大。本文结合文献[7,12]的相关定义,以 CN 算法为例,提出基于社区划分定义共邻节点贡献权重的算法,在对网络划分社区的基础上采用共邻节点连接度和不同社区对其的参与度来划分邻接节点的贡献,其修改后的定义是

$$Rc_\omega=\prod_{i=1}^{n}\frac{P(e\mid\omega_i,\omega_i\in C_k)}{P(\bar{e}\mid\omega_i,\omega_i\in C_k)}\times f(\omega_i)\tag{3}$$

其中:$f(\omega_i)$ 是社区对共邻节点的影响。$P(e\mid\omega_i,\omega_i\in C_i)$ 是 ω_i 存在并属于 C_k 条件下存在链路的概率。根据条件概率式:

$$P(e\mid\omega_i,\omega_i\in C_k)=\frac{P(\omega_i\in C_k\mid e,\omega_i)}{P(\omega_i\in C_k\mid\omega_i)}\times P(e\mid\omega_i)\tag{4}$$

$$P(\bar{e}\mid\omega_i,\omega_i\in C_k)=\frac{P(\omega_i\in C\mid\bar{e},\omega_i)}{P(\omega_i\in C_k\mid\omega_i)}\times P(\bar{e}\mid\omega_i)\tag{5}$$

式(4)和(5)相除:

$$\frac{P(e\mid\omega_i,\omega_i\in C_k)}{P(\bar{e}\mid\omega_i,\omega_i\in C_k)}=\frac{P(\omega_i\in C_k\mid e,\omega_i)}{P(\omega_i\in C_k\mid\bar{e},\omega_i)}\times\frac{P(e\mid\omega_i)}{P(\bar{e}\mid\omega_i)}\tag{6}$$

简化后可得

$$Rc_\omega=R_\omega\times\prod_{i=1}^{n}\frac{P(\omega_i\in C_k\mid e,\omega_i)}{P(\omega_i\in C_k\mid\bar{e},\omega_i)}\tag{7}$$

其中:Rc_ω 指共邻节点 ω 在社区划分的基础上对链路预测节点对的贡献。由于共邻节点处于某个社区中,其贡献会被社区结构属性所影响,定义 ω 的贡献为社区内的影响和社区外的影响的比率。由于 e 和 ω_i 独立,所以式(7)的连乘部分可转换为

$$\frac{P(\omega_i\in C_k\mid e,\omega_i)}{P(\omega_i\in C_k\mid\bar{e},\omega_i)}=\frac{P(\omega_i\in C_k\mid e)}{P(\omega_i\in C_k\mid\bar{e})}\times\frac{P(\omega_i\in C_k\mid\omega_i)}{P(\omega_i\in C_k\mid\omega_i)}=$$

$$\frac{P(\omega_i \in C_k \mid e)}{P(\omega_i \in C_k \mid \bar{e})} = \frac{N_{C_k}}{|N| - N_{C_k}} \tag{8}$$

其中：N_{C_k}和$|N|$分别是 C_k 社区内所有节点数目和网络总节点数目。那么，如何定义其社区影响 $f(\omega_i)$，本文采用共邻节点社区连接度和贡献度进行度量。

首先用 D_{in_ω}表示共邻节点 ω 与其所在社区 C_k 其他邻接节点之间出入度，$\bar{D}_{\text{in}_\omega}$表示社区 C_k 所有节点出入度的均值，$\sigma_{\text{in}_\omega}$表示标准差，则共邻节点 ω 在社区 C_k 内部的连接度为

$$k_{\text{in}\_\omega} = \frac{(D_{\text{in}\_\omega} - \bar{D}_{\text{in}\_\omega})}{\sigma_{\text{in}\_\omega}} \tag{9}$$

其中：k_{in_ω}代表共邻节点和其所在社区内部节点的连接度。

同时，由于不同社区的属性对共邻节点会有影响，还需要定义社区参与系数来刻画该节点与其他社区共邻节点的链接情况，并差分化处理不同社区对不同节点的贡献。那么共邻节点 ω 的参与系数 p_ω 定义为

$$p_\omega = 1 - \sum_{\omega=1}^{N_c} \left(\frac{D_{\omega c}}{D_\omega} \right)^2 \tag{10}$$

其中：D_{ic}描述 ω 与社区 C_k 中节点的连接数；D_ω 描述 ω 的总连接数；N_c 是社区的数目。所以，共邻节点的贡献定义为社区内的影响和社区外的影响的比率：

$$f(\omega_i) = \frac{k_{\text{in}\_\omega_i}}{1 - k_{\text{in}\_\omega_i}} \times \frac{p_{\omega_i}}{1 - p_{\omega_i}} \tag{11}$$

在这个比率中，分子和分母分别表示共邻节点受社区内部属性和外部属性的影响，分子越大，贡献越大，符合文献[7]提出的预测节点对的预测准确度受社区内部节点的影响相对较大这一原理。所以，改进后的 CN 算法公式如下：

$$r_{xy}^{\text{Community-}CN} = |\Gamma(x,y)| \times \log s + \sum_{\omega \in \Gamma(x,y)} \log Rc_\omega \tag{12}$$

3.3 模型推广

基于共邻节点相似度的链路预测[4]有许多经典模型。为了进一步证明本文提出算法的有效性和扩展性，把基于社区划分的定义共邻节点贡献算法应用到以下算法中，其中 k_ω 表示 ω 节点的度：

a) Adamic-Adar (AA)。采用共邻节点度的对数反比作为贡献权重。

$$r_{xy}^{AA} = \sum_{\omega \in \Gamma(x,y)} \frac{1}{\log k_\omega} \tag{13}$$

b) Resource allocation (RA)。采用共邻节点度的反比作为贡献权重。

$$r_{xy}^{RA} = \sum_{\omega \in \Gamma(x,y)} \frac{1}{k_\omega} \tag{14}$$

c) Jaccard coefficient(JC)。主要用于计算共同邻接节点和所有邻接节点的相似度。

$$r_{xy}^{\text{Jaccard}} = \frac{1}{|\Gamma(x) \cup \Gamma(y)|} \sum_{\omega \in \Gamma(x,y)} \log s \tag{15}$$

d) S∅rensen(SOR)。采用预测节点对度的和作为贡献权重。

$$r_{xy}^{\text{Sor}} = \frac{1}{k_x + k_y} \sum_{\omega \in \Gamma(x,y)} \log s \tag{16}$$

其对应改进形式如下所示：

$$r_{xy}^{\text{Community-}AA} = \sum_{\omega \in \Gamma(x,y)} \frac{1}{\log k_\omega} (|\Gamma(x,y)| \times \log s + \log Rc_\omega) \tag{17}$$

$$r_{xy}^{\text{Community-}RA} = \sum_{\omega \in \Gamma(x,y)} \frac{1}{k_\omega} (|\Gamma(x,y)| \times \log s + \log Rc_\omega) \tag{18}$$

$$r_{xy}^{\text{Community-Jaccard}} = \frac{1}{|\Gamma(x) \cup \Gamma(y)|} \sum_{\omega \in \Gamma(x,y)} (|\Gamma(x,y)| \times \log s + \log Rc_\omega) \tag{19}$$

$$r_{xy}^{\text{Community-Sor}} = \frac{1}{k_x + k_y} \sum_{\omega \in \Gamma(x,y)} (|\Gamma(x,y)| \times \log s + \log Rc_\omega) \tag{20}$$

4 实验评价

4.1 实验评价模型

本实验基于 complex networks package for MATLAB[13] 和 stanford network analysis platform[14] 平台，采用C ++ 和 MATLAB 语言进行开发。为了检测算法的正确性，在 $G = (V, E)$ 中把随机移除 E 中 10% 的边记为预测子集 E_r，其他 90% 的边作为训练数据集 $E = E_r \cup E_t$，则 $\varnothing = E_r \cap E_t$，节点总数保持不变。采用预测精确度、召回率和 ROC 曲线作为结果检验方法。

a) 预测精确度采用 10-folder 交叉验证方法进行，交叉重复验证 10 次，每次选择一个子集作为测试集，并将 10 次的平均交叉验证 precision 作为评价指标。Precision 指的是预测值 P_r 排 top-100 的预测链路的比率。

$$\text{precision} = \frac{\text{count}(\text{top100\_}P_r)}{100} \tag{21}$$

b) 召回率指的是预测准确的链路与所有存在链路 E_r 的比率。

$$\text{recall} = \frac{\text{count}(P_r)}{\text{count}(E_r)} \tag{22}$$

c) ROC(receiver operating characteristic curve) 是显示链路预测模型识别结果真正率与假正率之间折中的一种图形化方法，真阳性率 TPR 和假阳性率 FPR 分别为

$$TPR = \frac{TP}{TP + FN} \tag{23}$$

$$FPR = \frac{FP}{FP + TN} \tag{24}$$

ROC 曲线如图 2 所示。

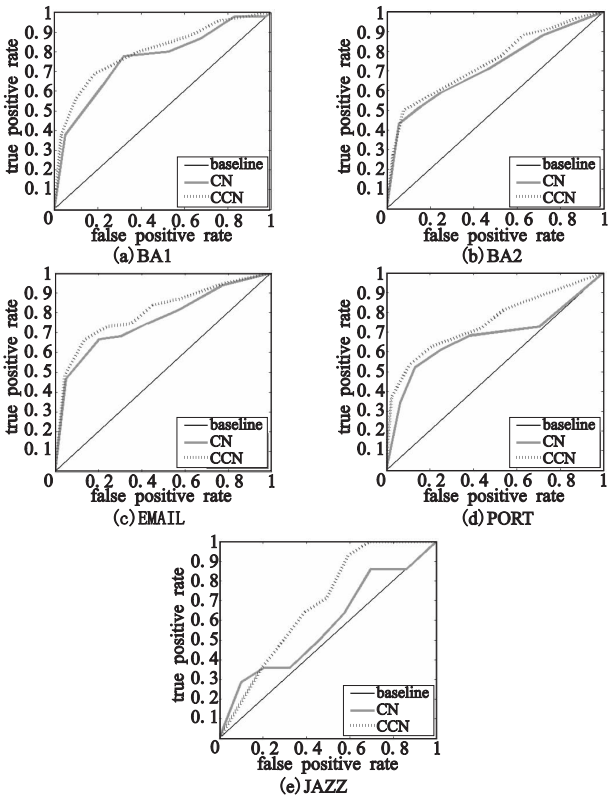

图2 ROC曲线

4.2 实验数据集

为了评价算法的有效性，分别使用了人工生成网络和真实网络开展实验。人工网络采用 Barabási 等人[15] 提出的随机网

络生成算法,生成具备不同属性的两个人工网络;真实网络采用 EMAIL[16]、PORT[16]和 JAZZ[16]三个真实的网络数据集进行实验,其度分布如图 3 所示,同时具体结构属性参照表 1。

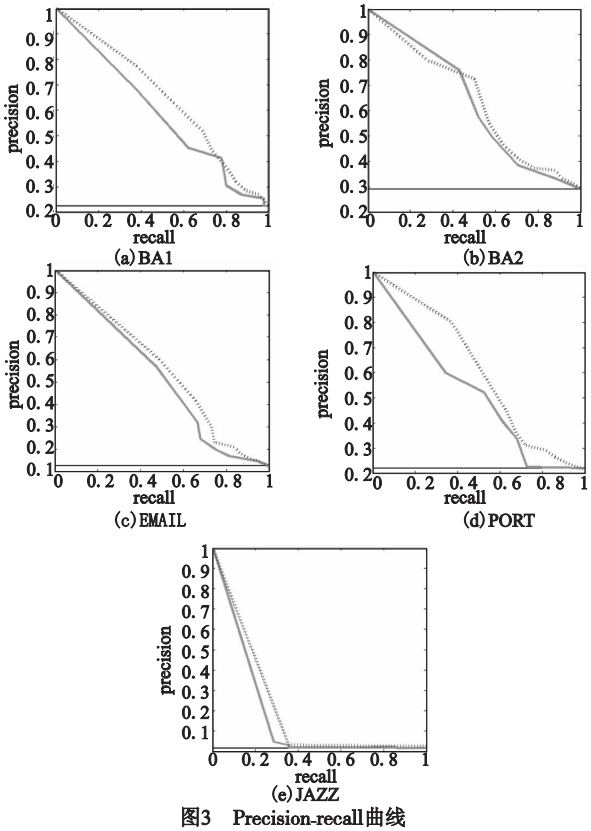

图3　Precision-recall曲线

表 1　人工网络和真实网络结构属性表

		BA1	BA2	EMAIL	PORT	JAZZ
node		200	300	1 116	327	198
edge		3 130	5 353	4 906	1 913	2 716
cluster coefficient		0. 476 3	0. 392 5	0. 235 5	0. 661 6	0. 629 3
average degree		43. 34	48. 13	8. 79	11. 7	27. 43
community	GN	171	260	68	99	40
	CNM	4	4	18	8	3
modularity	GN	0. 018 9	0. 015 3	0. 520 2	0. 112 9	0. 385 1
	CNM	0. 124 3	0. 116 2	0. 507 2	0. 324 1	0. 438 1

4.3　实验结果与分析

4.3.1　结果分析

表 2 为原始算法和改进后算法链路预测精确度。

表 2　原始算法和改进后算法链路预测精确度

model		CN	CCN	AA	CAA	RA	CRA	JC	CJC	SOR	CSOR
BA1	GN	0. 7143	0. 7449	0. 6837	0. 6939	0. 5204	0. 7347	0. 6939	0. 7347	0. 6224	0. 7245
	CNM	0. 7143	0. 7449	0. 6837	0. 6939	0. 5204	0. 7245	0. 6837	0. 6939	0. 6224	0. 7245
BA2	GN	0. 9082	0. 9388	0. 8878	0. 8980	0. 7245	0. 9184	0. 7531	0. 9592	0. 9184	0. 9286
	CNM	0. 9082	0. 9388	0. 8878	0. 8980	0. 7245	0. 9286	0. 9115	0. 9592	0. 9184	0. 9286
EMAIL	GN	0. 2245	0. 2653	0. 2653	0. 2755	0. 2245	0. 2653	0. 0816	0. 2143	0. 1735	0. 2551
	CNM	0. 2551	0. 2653	0. 2653	0. 2755	0. 2245	0. 2653	0. 0816	0. 2143	0. 1735	0. 2551
PORT	GN	0. 2143	0. 1939	0. 2143	0. 2041	0. 2041	0. 2857	0. 0816	0. 1837	0. 1429	0. 1837
	CNM	0. 2143	0. 2374	0. 2143	0. 2143	0. 2041	0. 2857	0. 0816	0. 1837	0. 1429	0. 1837
JAZZ	GN	0. 1122	0. 1224	0. 1224	0. 1224	0. 1122	0. 1224	0. 0613	0. 1020	0. 1327	0. 1531
	CNM	0. 1124	0. 1224	0. 1224	0. 1224	0. 1122	0. 1224	0. 1327	0. 1531	0. 066	0. 1020

从表 2 可以看出,无论是人工网络还是真实网络,采用两种算法进行社区划分,同时差分化共邻节点贡献后五类相似度改进算法的预测精确度都有所提高;其中,人工生成网络 CJC 算法的实验效果最佳,而 SOR 算法对真实网络的预测准确度最高。同时,对比 GN 和 CNM 两种社区划分算法,CNM 算法的效果整体更好,这是由于其采用贪婪凝聚算法和最大堆的数据结构对网络进行更加有效的社区划分,这可以从表 1 中 community 和 modularity 属性表现,CNM 算法生成模块度更高、数目相对较少的社区,同时相应算法的时间复杂度变少。

4.3.2　案例分析

为了更好地了解该算法的优越性,本文对比了邻近节点 CN 和 CCN 算法的 ROC 曲线和 Precision-recall 曲线。ROC 曲线越靠近左上角,实验的准确性就越高。从图 2 中曲线可以看出,CCN 算法基本一开始就比 CN 算法往左靠并一直延伸到整个预测集中,并且 ROC 曲线下的面积(AUC)也较大,证明改进后算法更有效。

Precision-recall 曲线由准确率和召回率这两个相互关联的统计量构成。召回率(recall)衡量预测值相比整个预测集的能力,而准确率(precision)衡量排除不相关预测结果的能力。从图 3 可以看出,CCN 曲线一直都在 CN 右上角,整体链路预测性能更佳。

5　结束语

本文研究了基于分析节点相似度的链路预测模型中共邻节点的角色问题。主要使用基于社区划分的算法挖掘网络社区属性,并把社区对共邻节点的影响定义为其贡献权重,同时把算法扩展到其他模型中,通过人工和现实的网络验证算法能够产生较好的效果。同时在本文算法的实现当中,$f(\omega_i)$的选择度量方法是非常关键的。本文采用社区连接度和贡献度,是否还有更好的贡献选择方法,是下一步需要研究的内容;同时,在实验中能否用多维网络验证算法有效性的工作也有待进一步开展。

参考文献:

[1] HOLLAND P W, LASKEY K B, LEINHARDT S. Social networks [M]. 1983.
[2] WASSERMAN S, FAUST K. Social network analysis: methods and applications[M]. [S. l.]:Cambridge University Press,1994.
[3] LV Lin-yuan, ZHOU Tao. Link prediction in complex networks: a survey[J]. Physica A,2011,390(6):1150-1170.
[4] 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.
[5] CLAUSET A, MOORE C, NEWAMAN M E J. Hierarchical structure and the prediction of missing links in networks[J]. Nature,2008, 453(7191):98-101.
[6] DONGHYUK S, SI S, DHILLON I S. Multi-scale link prediction [C]//Proc of the 21st ACM International Conference on Information and Knowledge Management. 2012:215-224.
[7] VALVERDE-REBAZA J C, De ANDRADE L A. Link prediction in complex networks-based on cluster information[C]//Proc of the 21st Brazilian Conference on Advances in Artificial Intelligence. 2012:92-101.
[8] FORTUNATO S. Community detection in graphs[J]. Physics Reports,2010,486(3):75-174.
[9] NEWMAN M E J. Communities, modules and large-scale structure in networks[J]. Nature Physics,2011,8(1):25-31.
[10] NEWMAN M E, GIRVAN M. Finding and evaluating community structure in networks [J]. Physical Review E, 2004, 69 (2): 026113.
[11] CLAUSET A, NEWMAN M E J, MOORE C. Finding community structure in very large networks[J]. Physical Review E,2004,70 (6): 066111.
[12] LV Lin-yuan, ZHOU Tao. Link prediction in weighted networks: the role of weak ties[J]. Europhysics Letters,2010,89(1):18001.
[13] Complex networks package for MATLAB[EB/OL]. http://www.levmuchnik.net/Content/Networks/ComplexNetworksPackage.html.
[14] Stanford network analysis platform (SNAP)[EB/OL]. http://snap.stanford.edu/snap/.
[15] BARABÁSI A L, ALBERT R. Emergence of scaling in random networks[J]. Science,1999,286(5439):509-512.
[16] The koblenz network collection[EB/OL]. http://konect.uni-koblenz.de/.