Journal of Computer Applications
计算机应用,2013,33(8):2095 - 2099
ISSN 1001-9081
CODEN JYIIDU
2013-08-01
http://www.joca.cn

文章编号:1001-9081(2013)08-2095-05
doi:10.11772/j.issn.1001-9081.2013.08.2095

基于协同过滤的 Web 服务动态社区发现算法

吴　钟[1,2*],聂规划[1],陈冬林[1],章佩璐[1]
(1.武汉理工大学 经济学院,武汉 430070;　2.武汉理工大学华夏学院 经济与管理系,武汉 430223)
(*通信作者电子邮箱 7849800@qq.com)

摘　要:针对现有社区发现算法挖掘结果精确度不高以及 Web 服务资源智能推荐质量较低的问题,在传统协同过滤算法的基础上,提出了基于节点相似性的动态社区发现算法。首先以连接节点最多的中心节点为起始网络社区,以社区贡献度为衡量指标不断形成多个全局贡献度饱和的社区;再使用重叠度计算将相似度高的社区进行合并,最后通过计算目标用户与社区中其他用户之间的动态相似度,将计算结果降序排列后构成邻近用户集,获得社区化推荐对象。实验结果表明,提出的社区发现算法对用户社会网络的社区分类与实际社区分类结果吻合,提高了社区挖掘的精确度,有助于实现高质量的社区化推荐。

关键词:Web 服务资源;协同过滤;社会网络;重叠社区;社区挖掘;节点相似性
中图分类号:TP393.094　　**文献标志码**:A

Dynamic community discovery algorithm of Web services based on collaborative filtering

WU Zhong[1,2*], NIE Guihua[1], CHEN Donglin[1], ZHANG Peilu[1]
(1. *School of Economics, Wuhan University of Technology, Wuhan Hubei* 430070, *China*;
2. *Department of Economic Management, Wuhan University of Technology Huaxia College, Wuhan Hubei* 430223, *China*)

Abstract: To cope with the low accuracy of the mining results in the existing community discovery algorithms and the low quality of intelligent recommendation in the Web services resource, on the basis of the conventional collaborative filtering algorithms, a dynamic community discovery algorithm was proposed based on the nodes' similarity. Firstly, the central node that had the most connected nodes was regarded as the initial network community, and the community contribution degree was taken as the metric to continuously form a plurality of global saturated contribution degree communities. Then, an overlapping calculation was used to merge the communities of high similarity. Finally, the calculated results were arranged in descending order to form neighboring user sets for obtaining community recommendation object by calculating the dynamic similarity between target user and other users in the community. The experimental results show that the user social network community classification by the proposed community discovery algorithms is consistent with the real community classification results. The proposed algorithm can improve the accuracy of the community mining and helps to achieve high-quality community recommendation.

Key words: Web service resource; collaborative filtering; social network; overlapping community; community mining; node similarity

0　引言

社会网络是人与人之间为达到某种特定目的而实现信息沟通的复杂网络。Web 服务社区发现是为了将社会网络划分为若干个互相分离的社区,通过挖掘用户社会网络结构、分析用户之间的连接关系来发现用户社区,寻求用户社区中与目标用户相近的用户集合。它是一种解决用户需求个性化问题的可行方法,近年来成为了研究者所关注的焦点。

目前,网络社区发现的算法主要有谱平分法、Kernighan Lin 算法、层次聚类算法和 GN(Givern-Newman)算法。但谱平分法在每次实施网络分割时只能对其进行平分,使得复杂网络在进行社区分割时的效率大大降低;Kernighan Lin 算法只能在知晓所分割社区大小的前提下才能进行网络分割;层次聚类算法中的单连接法难以控制和掌握算法的起始,无法确定最终划分得到的网络社区数量,完全连接法由于时间复杂度高,操作十分困难;GN 算法因为其没有有效定义网络拓扑结构,且需要进行重复计算,不适用于大规模的社会网络。总的来说,上述社区发现算法都只把社区发现问题简单地描述为普通聚类,且大都通过构建静态模型进行相似度度量,没有考虑其动态性。

针对静态社区算法中把社区发现简单描述为普通聚类的问题,国内外众多学者在考虑网络节点多样性的基础上将研究重心放在重叠社区动态发现算法上,纷纷在上述经典算法的基础上提出了新的社区发现算法。例如团渗算法(Clique Percolation Method,CPM)[1]、基于局部扩展的重叠社区挖掘算法(LFM)[2]、UEOC(Unfold and Extract Overlapping Communities)算法[3]、基于连边相似度的重叠社区发现算法(EGN)[4]、基于信息熵的社区发现(Community Detection Based on Entropy,CDBE)算法[5]、贪婪的团扩张(Greedy Clique Expansion,GCE)算法[6]等。然而,这些算法都没有从

收稿日期:2013-02-25;**修回日期**:2013-03-25。　**基金项目**:国家自然科学基金资助项目(71072077,71172043);国家科技支撑计划项目(2011BAH16B02);教育部留学回国人员科研启动基金资助项目(20101561);中央高校基本科研业务费专项资金资助项目(2012YB20)。

作者简介:吴钟(1982 -),男,湖北武汉人,讲师,博士研究生,主要研究方向:Web 服务、信息经济、服务管理;　聂规划(1958 -),男,河南周口人,教授,博士,主要研究方向:商务智能、知识管理、知识工程;　陈冬林(1970 -),男,湖北孝感人,教授,博士,主要研究方向:云计算、服务管理;　章佩璐(1986 -),女,浙江宁波人,硕士,主要研究方向:语义网、服务管理、商务智能。

用户兴趣相似性的角度实现 Web 服务社区发现，且对于社区规模不同的网络呈现出的有效性各不相同；另外，随着混合参数数值的增加，部分算法的挖掘精确度迅速下降。因此，这些算法在挖掘的社区质量上仍存在不完善之处。

协同过滤算法是通过分析用户之间的兴趣相似性来进行项目推荐，其基本思想是通过评分来反映用户对项目的兴趣，利用与用户兴趣相似的若干最近邻用户的评分来推导该用户对未知项目的评分[7]。它主要是基于其他用户的偏好完成商品推荐，因此，只需要计算用户之间的相似性[8]。目前协同过滤算法在 Web 服务上的应用非常广泛，例如预测 Web 服务的服务质量(Quality of Service，QoS)值[9-10]、建立基于协同过滤的个性化推荐系统[11-12]和识别与优先大软件项目中的需求[13]等。

本文提出了基于节点相似性的动态社区发现算法，并使用对社会网络中用户节点的动态描述来计算目标用户和其他用户间的相似度。该算法结合了 Web 服务的特点以及用户社会网络中节点的多样性，将协同过滤和社会网络应用于 Web 服务选择领域以解决目前 Web 服务选择算法的不足。

1 用户社会网络构建

对于 Web 服务推荐系统来说，挖掘用户社区能够帮助目标用户找到兴趣相投、习性相近的消费用户，同时通过适当的推荐方法对社区中的成员进行有类别的有效推荐。本文考虑将协同推荐算法中的“用户—项目评级矩阵”投射到用户社会网络中，利用“用户—项目评级矩阵”计算出各用户之间的相关性，并以此为基础得到用户的关系矩阵，如图 1 所示。

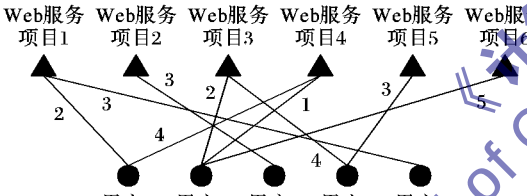


图 1 用户—项目关系图

图 1 中三角形节点表示 Web 服务的项目组；圆形节点表示用户；用户和项目间的连线表示用户使用过该 Web 服务项目且对其进行了评级，连线上的数值表示用户对项目的评级值，评级值越大，表示用户对 Web 服务项目的满意度越高，若用户与某一项目没有任何连接线，则表示该用户没使用过该项目或是使用后没给出评级值。本文假设评级值的取值范围为[1,5]。设 Web 服务项目有 s 个，用户数量为 t，根据用户—项目关系图构建得到的“用户—项目评级矩阵” $\boldsymbol{M}$：

$$\boldsymbol{M}=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1s} \\ r_{21} & r_{22} & \cdots & r_{2s} \\ \vdots & \vdots & & \vdots \\ r_{t1} & r_{t2} & \cdots & r_{ts} \end{bmatrix} \tag{1}$$

其中：r_{ij} 为用户 i 对 Web 服务项目 j 的评级值，当用户没有对项目作出评级时，$r_{ij}=0$。以此为基础，利用文献[14]中 Pearson 相关性算法计算得到用户 u 和用户 v 之间的相关系数：

$$w_{u,v}=\frac{\sum_{i\in I}(R_{u,i}-\bar{R}_u)(R_{v,i}-\bar{R}_v)}{\sqrt{\sum_{i\in I}(R_{u,i}-\bar{R}_u)^2}\sqrt{\sum_{i\in I}(R_{v,i}-\bar{R}_v)^2}} \tag{2}$$

其中：I 指用户 u 和用户 v 共同评级的 Web 服务项目集合；$R_{u,i}$ 指的是用户 u 对于项目 i 的评级值，$\bar{R}_u$ 则表示用户 u 评级的平均值；类似地，$R_{v,i}$ 表示用户 v 对项目 i 的评级值，$\bar{R}_v$ 表示用户 v 评级的平均值。该 Pearson 相关系数体现了用户和用户之间对于 Web 服务项目评级的相似度，因此可以将其转换为“用户—用户关系矩阵” $\boldsymbol{R}$：

$$\boldsymbol{R}=\begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1t} \\ s_{21} & s_{22} & \cdots & s_{2t} \\ \vdots & \vdots & & \vdots \\ s_{t1} & s_{t2} & \cdots & s_{tt} \end{bmatrix} \tag{3}$$

其中：s_{ij} 表示用户 i 和用户 j 的 Pearson 相关性，即用户 i 和用户 j 的相似度，$s_{ij}\in[0,1]$。

基于以上研究，以用户为节点、以用户之间的相关性为边、以用户之间的具体相似度为边上权重，便可构建 Web 服务的用户社会网络。

2 基于中心节点的重叠社区动态挖掘

本章以用户社会网络中连接节点最多的中心节点为起始网络社区，以社区贡献度为衡量指标考察中心节点的连接节点，通过不断地比较、增添、扩大和更新形成多个全局贡献度饱和的社区，使用重叠度计算将相似度高的社区进行合并，以此来减少社区覆盖率，精确用户分组情况，从而实现对用户社会网络的社区分类和发现不同用户社区之间的不同特点。该重叠社区挖掘主要经历社区挖掘阶段和重叠社区调整阶段。

定义 1 节点度指的是与该节点相连节点的边的数量。对于一个用户社会网络来说，节点度最大的节点一般都是网络的中心节点。

定义 2 社区贡献度用来描述社会网络中节点对某个社区的贡献度。参考文献[15]中的定义，本文的社区贡献度 c 可以表示为

$$c=\frac{L_{\text{in}}}{L_{\text{in}}+L_{\text{out}}} \tag{4}$$

其中：L_{in}、L_{out} 分别为社区内所有连接边上权重总和、社区与外部所有连接边上权重总和。

社区挖掘阶段的步骤如下：

1)构建 Web 服务用户社会网络，计算该网络中各用户节点的节点度，取其中节点度的值最大的节点 m 作为中心节点，标记 m 并以其为初始节点构建初始社区 C_m。由于假定 C_m 没有任何的内部联系和外部联系，故其社区贡献度 $c=0$。全局贡献度 C 描述的是网络社区挖掘过程中社区贡献度的最大值，故其初始值也为 0。

2)以 C_m 为核心，在用户社会网络中找到所有与该社区有连接关系的节点，并将这些节点作为候选邻近节点计算社区贡献度 c。若计算得到节点 n 的社区贡献度最大，并且该社区贡献度 $c_n\geqslant C$，那么将 n 进行标记后添加到 C_m 中，并实时更新 C，使得 $C=c_n$。由于 n 的加入，C_m 结构发生了变化，再重复执行步骤 2)，直到 C 不再变化。

3)当 C 已经达到最大值且不再发生变化时，表明社区结构已经达到一种最佳的稳定状态，即挖掘得到了网络社区 C_m。

4)查看社会网络中节点的状态，如果所有用户节点都被标记，则说明网络中所有社区都已被挖掘出来，结束；如果还有部分节点未被标记，则重新对这些未标记的节点进行检测，找到节点度最大的节点作为新的中心节点来创建一个新的初始社区，返回步骤 2)重复计算。

按照以上步骤可将整个 Web 服务用户社会网络分成若干个具有共同特点的社区，但由于挖掘得到的部分社区之间具有重叠性，因此还要对挖掘得到的初始社区群进行调整。

重叠社区的调整使用重叠度进行衡量,为了便于计算,本文简化用户社区的重叠度计算,通过计算两社区的共有节点和综合节点的数量比值得到,即社区 a 和社区 b 的重叠度 O 表示为

$$O_{a,b} = \frac{C_a \cap C_b}{C_a \cup C_b} \tag{5}$$

利用式(5)计算挖掘阶段中挖掘得到的社会网络中任意两个用户初始社区之间的重叠度 $O_{a,b}$。设定重叠度阈值为 T,当 $O_{a,b} \geqslant T$ 时将 a 和 b 两个初始社区合并成一个社区,当 $O_{a,b} < T$ 时不进行任何操作;然后重复计算用户社区之间的重叠度,直到所有重叠度的值均小于设定阈值时结束调整,最后得到最终的用户社区分类情况。

3　基于节点相似性的动态社区发现算法

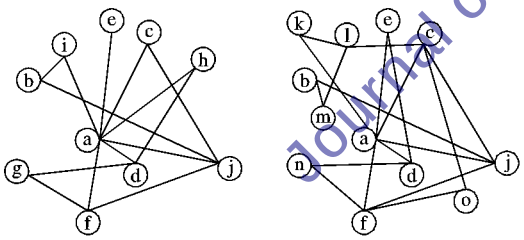
由于网络中的节点会随着时间的变化而变化,这就导致了包括社区结构在内的网络结构的动态演化。本章在前面挖掘得到的若干个用户社区的基础上,提出基于节点相似性的动态社区发现算法,从而获得与目标用户位于同一个社区且相关性高的用户组。

3.1　基于节点活跃度的动态社区描述方法

在动态的用户社会网络中,用户社区是指在一个时间片段中随着时间变化产生的共同社区特征的用户组。假定第二章挖掘得到了若干个用户社区且目标用户 U 所在的社区为 C_U,Δt_i 表示一个时间片段,G_i 表示在该时间片段中 C_U 中存在的用户状态,那么 C_U 从 Δt_1 到 Δt_s 的动态变化状态用以下二元组来表示:

$$C_U = \{(\Delta t_1, G_1), (\Delta t_2, G_2), \cdots, (\Delta t_s, G_s)\} \tag{6}$$

该二元组通过 G_i 描述社区中用户节点在不同时间片段的构成情况和相互间的关系变迁。G_i 为一个无向图,图中节点表示用户,连接边表示用户间的关系。图 2 描述了时间片段 Δt_1 和 Δt_2 的社区用户结构状态,当时间片段从 Δt_1 变迁到 Δt_2 以后,用户节点 g、h 和 i 离开了社区,而新用户 k、m、n 和 o 加入了该社区,且与社区中的原成员具有一定程度的联系。



(a) 状态一:$(\Delta t_1, G_1)$　　(b) 状态二:$(\Delta t_2, G_2)$

图 2　两个不同时间片段的社区用户变迁情况

本节提出“用户个体活跃度”的概念来衡量不同用户在社区中的活动频繁程度。由于不同的时间片段对整个用户社会网络的影响不同,且在一般情况下,时间顺序较靠后的时间片段所呈现的网络动态情况与当前的网络状态更为接近,因此对用户社会网络中的节点活跃度进行描述时必须对每一个时间片段设置一个影响权重,并通过加权平均来进行计算。

令 $A_{C_U}^{u_i}$ 代表目标用户 U 所在社区 C_U 中用户 u_i 的个体活跃度,$w_{\Delta t}$ 表示时间片段 Δt 在整个动态社区的权重影响度,$T_{C_U}^{u_i}$ 表示用户社区 C 在动态变化中包含用户节点 u_i 的时间片段集合,T_{C_U} 表示用户社区 C 在动态变化中的所有时间片段集合。本文假定目标用户 U 并不会随着时间的变化而离开社区 C_U,则用户 u_i 的个体活跃度可以表示为:

$$A_{C_U}^{u_i} = \sum_{\Delta t \in T_{C_U}^{u_i}} w_{\Delta t} \Big/ \sum_{\Delta t \in T_{C_U}} w_{\Delta t} \tag{7}$$

由于一个用户节点的活跃度并不能有效反映整个网络的连接关系和用户间的行为交往程度,所以还要对节点的邻近节点进行相应描述。利用式(7)计算 $A_{C_U}^{u_i}$ 之后,将 C_U 中的各用户节点按活跃度大小降序排列,并依次划分成 n 个活跃等级,得到社区 C_U 中节点 u_i 的动态描述方式:

$$(u_i, C_U) = \{Ulink_{u_i}, N_1^{u_i}, N_2^{u_i}, \cdots, N_n^{u_i}\} \tag{8}$$

其中:$Ulink_{u_i}$ 描述的是在 C_U 中与 u_i 相连接的节点集合;$N_1^{u_i}$,$N_2^{u_i}$,…,$N_n^{u_i}$ 描述的是与 u_i 相连接的节点按照个体活跃度大小归类于不同的活跃等级的个数。

3.2　基于用户节点相似度的动态相似度算法

根据文献[16]提出的个体之间动态相似度算法,将目标用户 U 所在的社区 C_U 中用户节点 u_i 和 u_j 的动态相似度用式(9)计算:

$$\begin{aligned} sim(u_i, u_j) = {} & w_0 \times |Ulink_{u_i} \cap Ulink_{u_j}| + \\ & w_1 \times \min(N_1^{u_i}, N_1^{u_j}) + w_2 \times \min(N_2^{u_i}, N_2^{u_j}) + \cdots + \\ & w_n \times \min(N_n^{u_i}, N_n^{u_j}) \end{aligned} \tag{9}$$

其中:$|Ulink_{u_i} \cap Ulink_{u_j}|$ 指的是分别与用户节点 u_i 和 u_j 相连接的邻近节点交集的节点个数;$\min(N_n^{u_i}, N_n^{u_j})$ 指的是在 $N_n^{u_i}$ 和 $N_n^{u_j}$ 两个数值中的较小值;$(w_0, w_1, \cdots, w_n)$ 指的是不同活跃等级节点集合对动态社区的影响权值集,一般情况下 $w_0 > w_1 > \cdots > w_n$。利用式(9)计算目标用户与其他用户间的动态相似度,然后将计算结果降序排列,取前三分之一的用户构成邻近用户集作为对目标用户推荐项目的社区化推荐对象。

4　实证分析

选取豆瓣网上 18 位电影爱好者为用户、17 部电影为 Web 视频服务项目,以这些用户对这些电影的评分为项目评级值(评级值在 1 到 5 之间,空白表示无评分),简化构建一个小型的 Web 视频服务网站的用户社会网络。豆瓣网“电影爱好者—电影评分”信息如表 1 所示,为了表述方便,电影 1~17分别代表电影七宗罪(1),画皮(2),三傻大闹宝莱坞(3),人在囧途(4),大话西游之月光宝盒(5),午夜凶铃(6),暮光之城之暮色(7),加勒比海盗 3(8),盗墓迷城(9),开心鬼上身(10),倩女幽魂(11),功夫熊猫(12),卑鄙的我(13),海扁王(14),小姐好白(15),孤岛惊魂(16),喜剧之王(17);字母 a~r 分别代表用户哥(a),三教废人陆不压(b),呦(c),Seon(d),Я|R(e),胆汁 lelele(f),Legend(g),少年 knife(h),阿夏(i),优优大头菜(j),Melonking(k),Majere(l),Copen(m),四麻言(n),姜城(o),親切的一萬.紮堆小天后(p),怪人(q),Josh(r)。

根据表 1 中的基本信息以及式(1),构建得到“用户—项目关系矩阵” $\boldsymbol{M}$。

$$\boldsymbol{M} = \begin{bmatrix} 4 & 0 & 5 & \cdots & 5 \\ 0 & 0 & 5 & \cdots & 3 \\ 0 & 3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 2 & 0 & \cdots & 0 \end{bmatrix}$$

按照表 1 中对用户和电影的项目编号,利用式(2)计算用户间的相关系数。以用户 a(“哥”)和用户 c(“呦”)为例,两者的共同评分电影项目有“人在囧途”、“大话西游之月光宝盒”、“卑鄙的我”和“小姐好白”。从中找到两者对这些项目的评分,并根据两个用户的平均评分值求得 Pearson 相关系数:

$$\begin{aligned} w_{a,c} = {} & [(3-2.06)(3-1.47) + (4-2.06)(5-1.47) + \\ & (1-2.06)(4-1.47) + (5-2.06)(4-1.47)] \times \end{aligned}$$

$$[(3-2.06)^2+(4-2.06)^2+(1-2.06)^2+(5-2.06)^2]^{-1/2}\times[(3-1.47)^2+(5-1.47)^2+(4-1.47)^2+(4-1.47)^2]^{-1/2}\approx\frac{13.0428}{19.8975}\approx 0.66$$

同理可得$w_{a,a}=1, w_{a,b}=0.86, \cdots, w_{a,r}=0.85$。由于用户之间的相关系数具有逆反性，所以$w_{u,v}=w_{v,u}(u=a,b,\cdots,r; v=a,b,\cdots,r)$，最终便可建立“用户 — 用户关系矩阵”$\boldsymbol{R}$。

$$\boldsymbol{R}=\begin{bmatrix}1 & 0.86 & \cdots & 0.85\\ 0.86 & 1 & \cdots & 0.65\\ \vdots & \vdots & & \vdots\\ 0.85 & 0.65 & \cdots & 1\end{bmatrix}$$

将用户作为网络节点，以用户间的连接线为其相关性并赋予具体权值，构建Web视频服务的模拟用户社会网络，如图 3 所示。为降低网络复杂度，图 3 中所有连接线边上的权重均省略显示。利用该模拟社会网络实现的社区挖掘结果如图 4 所示。

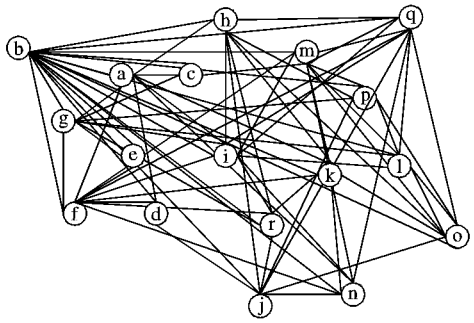


图 3 模拟的用户社会网络简图

表 1 豆瓣网“电影爱好者—影片评分”表示例

电影序号	用户																	
	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r
1	4				5			4			4		5		4	5		
2			3	2							3	4		3	3		3	2
3	5	5		5			4	5		4			5					
4	3		3		2	3		3				1	4	2			3	
5	4	5	5		5				4	5			5		1			5
6			2					1		1	5	4				5		4
7		3		3			2				4	5		4	2		4	
8						5		4	5		4		5	5		5	4	5
9								2	3	2	4		5	5		3		5
10		2	4			5	4	3			1		2					
11					3					3		4	4	3			5	2
12	4	1		5	3		4		4		3					2		4
13	1		4	2		3				5		4		1	3			
14	4	4			4	3		3			1		1			5		3
15	5		4	5		4	5			4								3
16						1					4	3		4	5	3	5	
17	5	3		5		4			5	5							3	
评分总值	35	22	25	27	22	25	22	25	21	29	33	27	36	27	18	28	27	33
评分平均值(近似)	2.06	1.29	1.47	1.59	1.29	1.47	1.29	1.47	1.24	1.71	1.94	1.59	2.12	1.59	1.06	1.65	1.59	1.94

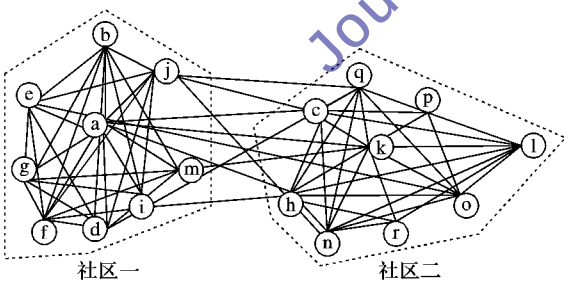


图 4 模拟社区挖掘结果

结果显示：该模拟用户社会网络被分成了两个社区。而在豆瓣网上，这些用户也归属于两个不同类别的电影社区，图 5 为实验用户在豆瓣网上的实际分组情况简图。

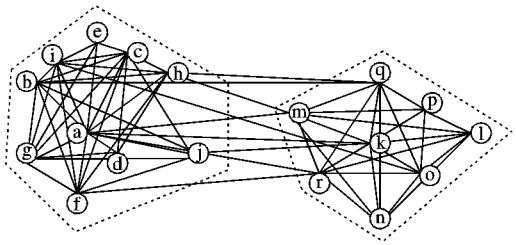
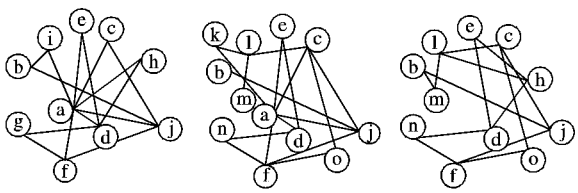


图 5 实验用户在豆瓣网上的实际分组情况简图

比较图 4 和图 5 发现，除了用户 c、h 和 m 之外，算法挖掘结果与社区实际分类一致，这主要是由于 c、h 和 m 对不少另外一小组成员感兴趣的电影进行了评分，且这些电影所占其评分电影的比例较高。因此，从理论上说 c、h 和 m 可属于任意一个社区，表明社区挖掘的精确度较高。

为了验证基于节点相似性的动态社区发现算法，再引入一个实例来演示说明。图 6 为目标用户 U 所在社区 C_U 在三个不同时间片段时的社区动态结构。



(a) 状态一:$(\Delta t_1, G_1)$ (b) 状态二:$(\Delta t_2, G_2)$ (c) 状态三:$(\Delta t_3, G_3)$

图 6 实例计算的三个时间片段中社区的动态结构

假设 Δt_1、Δt_2 和 Δt_3 对整个社区动态结构的影响权重分别为 1/4、1/2 和 1，利用式(7)计算出各用户节点在社区中的个体活跃度：用户 a 为 3/7；用户 b，c，d，e，f，j 均为 1；用户 g，i 均为 1/7；用户 h 为 5/7；用户 k 为 2/7；用户 l，m，n，o 均

为 6/7。将 C_U 中的用户节点分成六大类:

$U_{A=1}=\{b,c,d,e,f,j\}$, $U_{A=\frac{6}{7}}=\{l,m,n,o\}$

$U_{A=\frac{5}{7}}=\{h\}$, $U_{A=\frac{3}{7}}=\{a\}$

$U_{A=\frac{2}{7}}=\{k\}$, $U_{A=\frac{1}{7}}=\{g,i\}$

以用户节点 a 为例,在这三个时间片段中,与 a 有着直接联系的用户节点集合 $Ulink_a=\{c,d,e,f,h,i,j,k\}$,这些用户节点中个体活跃度属于第一类的有 5 个,属于第三类的有 1 个,属于第五类的有 1 个,属于第六类的有 1 个。根据式(8)得到用户 a 的动态描述为:

$(a,C_U)=\{\{c,d,e,f,h,i,j,k\},5,0,1,0,1,1\}$

同样地,对该社区中其他用户节点进行类似分析得到它们的动态描述分别为:

$(b,C_U)=\{\{i,j,m\},1,1,0,0,0,1\}$

$(c,C_U)=\{\{a,j,l,o\},1,2,0,1,0,0\}$

$(d,C_U)=\{\{a,e,g,h,n\},1,1,1,1,0,1\}$

$(e,C_U)=\{\{a,d,h\},1,0,1,1,0,0\}$

$(f,C_U)=\{\{a,g,j,n,o\},1,2,0,1,0,1\}$

$(g,C_U)=\{\{d,f\},2,0,0,0,0,0\}$

$(h,C_U)=\{\{a,d,e,l\},2,1,0,1,0,0\}$

$(i,C_U)=\{\{a,b\},1,0,0,1,0,0\}$

$(j,C_U)=\{\{a,b,c,f\},3,0,0,1,0,0\}$

$(k,C_U)=\{\{a,l\},0,1,0,1,0,0\}$

$(l,C_U)=\{\{c,h,k,m\},1,1,1,0,1,0\}$

$(m,C_U)=\{\{b,l\},1,1,0,0,0,0\}$

$(n,C_U)=\{\{d,f\},2,0,0,0,0,0\}$

$(o,C_U)=\{\{c,f\},2,0,0,0,0,0\}$

假设动态社区用户节点类别所占的权重 $w_0=\frac{1}{2},w_1=\frac{1}{4},w_2=\frac{1}{6},w_3=\frac{1}{8},w_4=\frac{1}{10},w_5=\frac{1}{12},w_6=\frac{1}{14}$,根据式(9),便可得到该社区中各用户节点之间的动态相似性。

由于该社区中用户节点 d 在三个时间片段中始终存在,而本文刚好假设目标用户不会离开目标社区,故设定用户 d 为本案例中的目标用户 U_o。由此计算出 d 与其他用户间的动态相似性:

$$sim(d,a)=\frac{1}{2}\times 2+\frac{1}{4}\times\min(1,5)+\frac{1}{6}\times\min(1,0)+\frac{1}{8}\times\min(1,1)+\frac{1}{10}\times\min(1,0)+\frac{1}{12}\times\min(0,1)+\frac{1}{14}\times\min(1,1)=\frac{81}{56}$$

同理可得:

$sim(d,b)=\frac{41}{84}$, $sim(d,c)=\frac{61}{60}$, $sim(d,e)=\frac{59}{40}$

$sim(d,f)=\frac{877}{420}$, $sim(d,g)=\frac{1}{4}$, $sim(d,h)=\frac{91}{60}$

$sim(d,i)=\frac{17}{20}$, $sim(d,j)=\frac{17}{20}$, $sim(d,k)=\frac{23}{30}$

$sim(d,l)=\frac{25}{24}$, $sim(d,m)=\frac{5}{12}$, $sim(d,n)=\frac{1}{4}$

$sim(d,o)=\frac{1}{4}$

将计算结果降序排列得到:

$$\frac{877}{420}>\frac{91}{460}>\frac{59}{40}>\frac{81}{56}>\frac{61}{60}>\cdots$$

取前三分之一个数值所对应的用户分别为:f、h、e 和 a,因此这四个用户便集合起来形成了候选的推荐用户小组。然后通过对目标用户 d 需求的 Web 服务资源和供应商提供的 Web 服务资源的功能属性之间进行概念匹配、结构匹配和文本匹配,以及对两者 QoS 文本型和数值型属性的匹配,以用户 d 的风险态度、印象值以及 Web 服务供应商的声誉情况为三大影响用户兴趣度的推荐因素进行匹配计算,得到的 Web 服务推荐结果,将被社区化推荐给这些候选用户小组。

5　结语

本文考虑到现实社会网络中重叠社区的存在情况,在传统社区聚类算法上利用社区中节点的动态性结构对社区挖掘结果进行精确化处理,提出了基于节点相似性的动态社区发现算法。与其他社区发现算法相比,本文提出的社区发现算法从用户兴趣相似性和用户社会网络的角度出发,提高了社区挖掘的精确度。

参考文献:

[1] PALLA G, DERENYI I, FARKAS I. Uncovering the overlapping community structures of complex networks in nature and society [J]. Nature, 2005, 435(7043): 814 - 818.

[2] LANCICHINETTI A, FORTUNATO S, KERTESZ J. Detecting the overlapping and hierarchical community structure in complex networks [J]. New Journal of Physics, 2009, 11(3): 033015.

[3] JIN D, YANG B, BAQUERO C. A Markov random walk under constraint for discovering overlapping communities in complex networks [J]. Journal of Statistical Mechanics: Theory and Experiment, 2011: P05031.

[4] 施伟, 傅鹤岗, 张程. 基于连边相似度的重叠社区发现算法研究[J]. 计算机应用研究, 2013, 30(1): 221 - 223.

[5] 王刚, 钟国祥. 基于信息熵的社区发现算法研究[J]. 计算机科学, 2011, 38(2): 238 - 240.

[6] LEE C, REID F, MCDAID A. Detecting highly overlapping community structure by greedy clique expansion [C]// Proceedings of the 4th International Workshop on Social Network Mining and Analysis. New York: ACM, 2010: 33 - 42.

[7] 王松, 徐德华. 基于产品分类的协同过滤算法应用研究[J]. 计算机应用与软件, 2012, 29(4): 183 - 185, 191.

[8] 余肖生, 孙珊. 基于网络用户信息行为的个性化推荐模型[J]. 重庆理工大学学报: 自然科学版, 2013, 27(1): 47 - 50.

[9] ZHENG Z B, MA H, LYU M R, *et al*. QoS-aware Web service recommendation by collaborative filtering [J]. IEEE Transactions on Services Computing, 2011, 4(2): 140 - 152.

[10] WU J, CHEN L, FENG Y P, *et al*. Predicting quality of service for selection by neighborhood-based collaborative filtering [J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2013, 43(2): 428 - 439.

[11] LIU Q, CHEN E H, XIONG H, *et al*. Enhancing collaborative filtering by user interest expansion via personalized ranking[J]. IEEE Transactions on Systems, Man, and Cybernetics — Part B: Cybernetics, 2012, 42(1): 218 - 232.

[12] KWON H-J, HONG K-S. Personalized smart TV program recommender based on collaborative filtering and a novel similarity method [J]. IEEE Transactions on Consumer Electronics, 2011, 57(3): 1416 - 1423.

[13] LIM S L, FINKELSTEIN A. StakeRare: using social networks and collaborative filtering for large-scale requirements elicitation [J]. IEEE Transactions on Software Engineering, 2012, 38(3): 707 - 735.

[14] PHAM M C, CAO Y W, KLAMMA R. A clustering approach for collaborative filtering recommendation using social network analysis[J]. Journal of Universal Computer Science, 2011, 17(4): 583 - 604.

[15] 马兴福, 王红. 一种新的重叠社区发现算法[J]. 计算机应用研究, 2012, 29(3): 844 - 846.

[16] 陈琼, 李辉辉, 肖南峰. 基于节点动态属性相似性的社会网络社区推荐算法[J]. 计算机应用, 2010, 30(5): 1268 - 1272.