

一种融合节点与链接属性的社交网络社区划分算法*

李孝伟, 陈福才, 刘力雄

(国家数字交换系统工程技术研究中心, 郑州 450002)

摘要: 针对传统社交网络社区划分算法普遍缺乏对节点属性、链接属性的综合考虑和充分表达利用节点与链接属性信息的模型和机制等问题,提出了一种融合节点与链接属性的社交网络社区划分算法。该算法融合节点属性的相似度、节点间链接权值等链接属性信息,定义了相似权值,并以此为基础,结合凝聚算法实现了对社交网络的社区划分。实验表明,该算法对社交网络中属性比较明显的社区划分效果显著。

关键词: 社交网络; 社区划分; 模块度; 相似权值

中图分类号: TP181; TP301.6 **文献标志码:** A **文章编号:** 1001-3695(2013)05-1477-04

doi:10.3969/j.issn.1001-3695.2013.05.050

Combined node and link attributes of social network community detection algorithm

LI Xiao-wei, CHEN Fu-cai, LIU Li-xiong

(National Digital Switching System Engineering & Technological R&D Center, Zhengzhou 450002, China)

Abstract: The traditional social network community detection algorithms generally lack of consideration of node and link attributes, and full expression using node and link attribute information model and mechanism. Aiming at this issue, this paper put forward a kind of combined nodes and links attribute social network community detection algorithm. This algorithm combined with node attributes similarity between nodes, link weights and link information, defined the similar weights, and, on this basis, combining condensation algorithm to realize the social network of community division. Experiments show that effect of this algorithm about social network attribute is remarkable, obviously in attribute-distinct community.

Key words: social network; community detection; modularity; similar weight

0 引言

随着社会经济的快速发展,人们越来越频繁地开展各种交往和与工作、生活相关的社会活动,信息交流也在飞速发展。在信息交流过程中,每个人的社会属性,包括社会关系、社会行为等,得以显性或隐性地展现,逐渐形成了如即时通信聊天网、电信网、微博网络等社交网络。经研究发现,社交网络普遍存在社区结构^[1]。如何划分出这种社区结构对于研究社交网络而言,具有重要意义。

传统的社区划分算法主要包括:

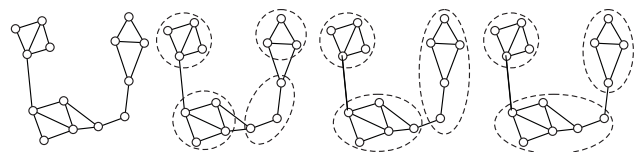
a) 基于图论的算法。其基本思想是给定一个网络,将其分解成一些子网络,各个子网络内的节点数基本相等,并且处于不同子网内的节点之间的链接非常少。著名算法有 Kernighan-Lin 算法^[2] (简称为 K-L 算法)、基于拉普拉斯图特征值的谱平分法^[3]、派系过滤算法和 W-H 快速谱分割法^[4]等。

b) 层次聚类算法。该算法属于社会学的方法,其主要是分析社会网络之间的相似性或边之间链接的强度。层次聚类算法可分为两大类算法,即凝聚算法和分裂算法,划分的依据是在网络中加边还是去边,加边的是凝聚算法,而去边的是分裂算法。凝聚算法的典型代表是 Newman 快速算法^[5]和 CNM 算法^[6],分裂算法的典型代表是 GN 算法^[7]。

后来出现的许多社区划分算法也都是在传统算法基础上优

化和改进的。由于社交网络中节点和链接关系都蕴涵有大量的属性信息,而这些属性信息对于完全呈现网络的社区结构具有重要意义。所以,近几年出现了一个新的研究方向——融合属性和链接关系的社区划分。对于社交网络而言,具有相似属性的个体间发生链接关系的可能性更大,进而有可能形成社区或团体。社交网络中,同一个组织内的成员往往不仅存在着好友关系,同时也具有共同的兴趣爱好或是其他共同的个体属性。

图1是对某一小型社交网络划分结果的比较。图1(a)为原始的链接图;(b)为基于属性划分的结果;(c)为基于链接信息划分的结果;(d)为期待的划分结果。由图1(b)和(c)可以看出,单纯地考虑属性或是链接关系,划分出的结果都不是想要的。因此,若同时考虑链接信息和属性信息能够更真实地呈现出社交网络的社区结构^[8]。



(a) 原始链接图 (b) 属性划分图 (c) 链接划分图 (d) 期待划分图

图1 社区划分实例图

加拿大西蒙菲莎大学的 Ester 等人^[9]提出了 CkC (connected k -center) 问题,将个体的属性看做与该个体对应的坐标向量,并以此建立了信息图 (informative graph)。Moser 等人^[10]在

收稿日期: 2012-09-17; 修回日期: 2012-10-26 基金项目: 国家“863”计划资助项目 (2011AA010603)

作者简介: 李孝伟 (1987-), 男, 安徽阜阳人, 硕士研究生, 主要研究方向为通信与信息系统 (lxwei_1987@163.com); 陈福才 (1974-), 男, 江西高安人, 硕士, 主要研究方向为通信与信息系统; 刘力雄 (1974-), 男, 湖南隆回人, 副教授, 主要研究方向为通信与信息系统。

CkC 问题的基础上,改进了 CkC 算法需要事先给出聚类数量 k 的限制,提出了无须事先指定聚类数量的 CXC 算法。CkC 和 CXC 并未解决边属性利用和无链接关系个体间的相似性利用问题。林有芳等人^[11]将属性信息融合到提出的紧密度中,提出了一种边稳定系数模型和一种能表达个体间关系紧密度的完全信息图模型,在此基础上设计和实现了一种有效的社区划分算法,但该算法提出的边稳定系数仅仅依赖于局部有限的网络环境,也没考虑属性对节点划分影响力的差异性,易导致划分出的社区真实度不高。Dang 等人^[12]定义了属性模块度,在社区划分中将该模块度与 Newman 提出的模块度^[6]结合,此外,提出了一种 k NN 图划分方法,但该方法受限于 k 值的大小。

社交网络中,节点和链接具有丰富的属性信息,诸如个体属性、结构属性、社会属性、网络行为属性等。这些属性对于节点所处社区的定位起着很重要的作用。本文考虑了节点属性和链接属性对社区发现的影响作用,以及无边节点间属性的相似性对社区发现的影响。算法设计上以凝聚算法为基础,将节点属性和链接关系进行函数拟合,设计了一种以此拟合函数为基础的融合属性与链接关系的社区划分算法。实验仿真验证表明,该算法对于具有属性的社会网络社区划分效果比较显著。

1 模型与相关定义

1.1 社交网络模型定义

社交网络一般用信息图来表示。简单的社交网络如图 2 所示。

定义 1 信息图。设图 $G = \{V, E\}$, V 为网络中的个体(或称节点), $V = \{v_1, v_2, \dots, v_n\}$ 共包含 n 个节点; E 为网络中节点链接(或称边)的集合, $E = \{e_1, e_2, \dots, e_m\}$, 共 m 条边。个体共有 p 个属性,节点 V 的属性集合为 $V_{\text{att}} = \{a_1, a_2, \dots, a_p\}$ 。链接共有 h 个属性,链接 E 的属性集合为 $E_{\text{att}} = \{b_1, b_2, \dots, b_h\}$ 。由以上信息可以构建图的节点和链接属性矩阵。信息图的链接矩阵可以用如下矩阵表示:

$$\begin{pmatrix} A_{11} & \cdots & \cdots & \cdots & A_{1n} \\ \vdots & \ddots & A_{ij} & & \vdots \\ \vdots & A_{ji} & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ A_{m1} & \cdots & \cdots & \cdots & A_{mn} \end{pmatrix}$$

其中: A_{ij} 为节点 i 到节点 j 的链接权值, A_{ji} 为节点 j 到节点 i 的链接权值。对于无权无向图,矩阵中元素取值为 0 或 1,并且 $A_{ij} = A_{ji}$ 。

1.2 信息增益

信息增益定义为原来的信息需求与新的需求之间的差值^[13]。原来的信息需求是指原有样本分类的期望信息。新的需求是指在已知属性 A 的划分基础上对样本分类的信息需求。

假定样本 D 中共有 m 个样本类别,第 i ($i = 1, 2, \dots, m$) 个类别占样本总数的比例为 p_i , 则

$$\text{info}(D) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

假设按照特征 A 来划分 D 中的样本,用 A 可以将 D 划分为 v 个子集 $\{D_1, D_2, \dots, D_v\}$, 其中 D_j 内的样本在 A 上具有值 a_j 。根据 A 的这种划分对 D 的样本划分所需的期望信息为

$$\text{info}_A(D) = \sum_{j=1}^v \left(\frac{|D_j|}{|D|} \cdot \text{info}(D_j) \right) \quad (2)$$

则属性 A 的信息增益为

$$\text{gain}(A) = \text{info}(D) - \text{info}_A(D) \quad (3)$$

其中: $\text{gain}(A)$ 是属性 A 的信息增益; $\text{info}(D)$ 是原来的信息需求; $\text{info}_A(D)$ 是新的信息需求。

1.3 模块度

设图 $G = \{V, E\}$, 共包含 n 个节点和 m 条边,在无权无向图中,模块度 Q 在节点上定义^[6]为

$$Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - \frac{k_i k_j}{2m}) \delta(C_i, C_j) \quad (4)$$

其中: $k_i = \sum_{j=1}^n A_{ij}$ 是 V_i 拥有的链接数(即 V_i 的度); C_i 为节点 i 所属的社区,当节点 i 和节点 j 处于同一社区时 $\delta(C_i, C_j) = 1$,反之则为 0。模块度的作用主要是评价社区划分的好坏和网络社区结构的强弱。 Q 值范围在 $(0, 1)$ 之间,其值越大则说明划分结果越好,社区结构越明显。通常模块度在 $0.3 \sim 0.7$ 之间的网络,其社区结构比较明显。

1.4 归一化互信息

归一化互信息^[14] (normalized mutual information, NMI) 是用来衡量划分后的社区与真实社区的差异。设真实的社区划分为 C_o , 算法得出的社区划分为 C_e , NMI 的定义为

$$\text{NMI}(C_o, C_e) = \frac{H(C_o) + H(C_e) - H(C_o, C_e)}{\sqrt{H(C_o)H(C_e)}} \quad (5)$$

其中: $H(C)$ 表示划分 C 的香农信息熵。当 C_e 与 C_o 完全一致时, $\text{NMI}(C_o, C_e) = 1$; 当 C_e 与 C_o 完全不同时, $\text{NMI}(C_o, C_e) = 0$; 当 $\text{NMI}(C_o, C_e)$ 在 $(0, 1)$ 之间时,其值越大则说明划分的结果越接近于真实社区。

2 基于相似权值的社区划分算法

2.1 算法总体思路

以节点的属性矩阵和链接矩阵等为基础,计算节点间的相似权值,然后结合凝聚算法进行社区划分。初始时每个节点为一个社区,首先将相似权值最大的两个社区合并成一个社区,计算合并之后的模块度。之后依次将余下具有最大相似权值的两个社区合并成一个社区,相应地计算模块度。在计算模块度的过程中存在一个极大值点,该极大值点时划分的社区即为最优社区。

2.2 算法涉及的几个定义

1) 节点属性相似性度量

在进行属性之间的相似性度量时,首先要考虑节点属性对该节点社区划分的影响力,对于影响力较大的属性要赋予较大的权值,而对于影响力较小的则赋予较小的权值。对于节点属性权值的确定,本文采用信息增益的方法。本文采用的相似性度量是 Dice 系数。节点 V_i 和 V_j 属性之间相似性度量(attribute similarity)函数 $(V_i, V_j)_{AS}$ 的计算式如下:

$$(V_i, V_j)_{AS} = 2 \times \sum_{p=1}^k (a_{ip} \cdot a_{jp}) / (\sum_{p=1}^k a_{ip}^2 + \sum_{p=1}^k a_{jp}^2) \quad (6)$$

其中:各节点共有 k 个属性信息; a_{ip} 和 a_{jp} 分别为节点 V_i 和 V_j 的第 p 个属性量化后的属性信息。

2) 链接属性度量

节点间的链接属性(link attributes)主要有链接的权值、链接发生的时间、链接发生的频率等。由于链接属性本身包含了节点间的联系,所以这里不能用相似性度量来表征链接属性。链接的各个属性对链接划分所起的作用并不相同,故要先采用

一种衡量机制来确定各个属性对链接划分的贡献。记每个链接共有 h 个属性,将节点共有邻居节点贡献也作为链接的一个属性,此时链接共有 $(h+1)$ 个属性。设两个节点的共有邻居节点的度数为 dg ,而对于这两个节点的相似性而言,其共有邻居节点所起的作用只占到 $2/dg$ 。设两个节点的共有邻居节点集合为 $(V_i, V_j)_{\text{sum}}$ 。定义共有邻居节点贡献属性 L 为: $L = \sum_{i \in (V_i, V_j)_{\text{sum}}} (2/dg_i)$ 。链接 E_{ij} 的属性量化后的属性信息为 $(b_{ij1}, b_{ij2}, \dots, b_{ijh}, L)$,各个属性对应的权值为 $(w_1, w_2, \dots, w_h, w_{h+1})$,其中 $w_1 + w_2 + \dots + w_h + w_{h+1} = 1$ 。本文拟采用的链接属性度量函数 $(E_{ij})_{LA}$ 的计算式如下:

$$(E_{ij})_{LA} = \sum_{q=1}^h (b_{ijq} \cdot w_q) + w_{h+1} \cdot L \quad (7)$$

其中: w_q 可以根据实际的社交网络来设定。

3) 相似权值

本文定义的相似权值 (similar weights) 是在紧密度^[7]基础上得来的。节点 V_i 和 V_j 的相似权值函数 $(V_i, V_j)_{SW}$ 的计算式如下:

$$(V_i, V_j)_{SW} = \alpha \times (V_i, V_j)_{AS} + (1 - \alpha) (E_{ij})_{LA} \quad (8)$$

相似权值计算是建立在属性矩阵和链接矩阵基础上的,但是对于每个节点而言,属性和链接对划分所起的作用不尽相同,这里引入一个调整参数 α 来调整节点属性和链接属性在对相似权值计算时的作用。式(5)中 $(V_i, V_j)_{AS}$ 为节点 V_i 和 V_j 属性的相似度, $(E_{ij})_{LA}$ 为节点 V_i 和 V_j 的链接属性度量。

4) 改进后的模块度

由于本文研究的社会网络均为有权有向网络,所以这里要对模块度进行相应的改进。改进后的模块度 Q_{im} 计算式如下:

$$Q_{im} = \frac{1}{2m} \sum_{i,j} (\lambda_{ij} - \frac{k_i k_j}{2m}) \delta(C_i, C_j) \quad (9)$$

其中: λ_{ij} 为参数,当节点 V_i 和 V_j 之间有链接时, $\lambda_{ij} = 1$, 否则 $\lambda_{ij} = 0$; 对于有权有向图而言, $k_i = \sum_{j=1}^n A_{ij}$ 是节点 V_i 所有链接权值的累加, A_{ij} 为节点 i 和 j 的链接权值; m 为网络中所有链接权值的和; C_i 为节点 i 所属的社区,当节点 i 和节点 j 处于同一社区时 $\delta(C_i, C_j) = 1$, 反之为 0; Q_{im} 的作用在有权有向网络中等于无权无向网络中的 Q 。

2.3 基于相似权值的社区划分算法实现

基于相似权值的社区划分算法 (community detection algorithm based on similar weights, CDASW) 采用基于加边的凝聚算法框架,以前述的相似权值和模块度为基础设计,核心是在有向有权的信息图模型下,融合相似权值和模块度得到加边凝聚算法。算法处理框图如图3所示。

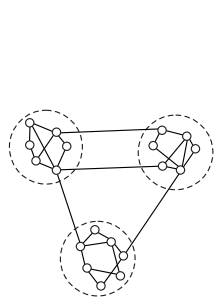


图2 简单社交网络模型

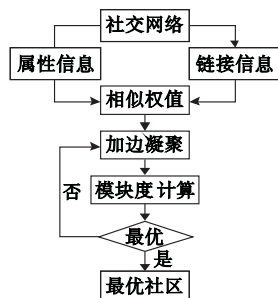


图3 算法处理框图

算法的伪代码描述如下:

算法1 CDASW

初始化 $G = \{V(n), E(m)\}$, $V_{att}(k)$, $E_{att}(h+1)$, 初始化社区 $C = \{C_1, C_2, \dots, C_n\}$ 。按照式(6)~(8)初始化 SW 矩阵 (V_{SW})。

```

1 do
2 求 ( $V_{SW}$ ) 的最大元素, ( $V_{SW}$ )ij = max ( $V_{SW}$ );
3 for  $l = 1: |C|$ 
4  max( $i$  行第  $l$  元素,  $j$  行第  $l$  元素)  $\rightarrow i$  行第  $l$  元素,
   max( $i$  列第  $l$  元素,  $j$  列第  $l$  元素)  $\rightarrow i$  列第  $l$  元素;
5 end
6 令 ( $V_{SW}$ )ij = 0;
7 delete ( $V_{SW}$ ) 的  $j$  行和  $j$  列;
8  $C \leftarrow C \setminus \{C_i, C_j\} \cup \{C_i \cup C_j\}$ ;
9  $C_{(1C1)} \leftarrow C$ ;
10 更新 ( $V_{SW}$ ), 返回 2;
11 until  $|C| = 1$ ;
12  $C_{(tree)} \leftarrow \{C_{(n)}, \dots, C_{(1)}\}$ ;
13  $C_m \leftarrow \arg \max_{C \in C_{(tree)}} Q(C)$ ;
14 返回  $C_m$  作为社区划分结果。

```

由上述算法设计循环中可以看出,本文算法的时间复杂度为 $O(n^2)$, 高于 CNM 算法的 $O(n \log^2 n)$ 。由于网络中边数目一般都远远大于节点数目,故本文方法时间复杂度低于 GN 算法。

3 实验及分析

3.1 带属性的 DBLP 数据集实验

由于社会网络本身的庞大性、多变性和隐私性,社区划分算法普遍存在验证上的困难。为了验证本文提出的基于相似权值的社区发现算法的效果,首先对公开数据集 DBLP (digital bibliography & library project) 数据进行实验。通过对实验数据的分析得到作者间的合著网络数据。

1) 数据预处理

根据对数据的分析得知,该数据为 509 篇论文数据,解析得到该数据的共有 213 个节点和 1 146 条边的合著网络。分析得到节点属性有兴趣、研究领域、所在研究小组等;链接属性有链接发生时间、链接发生次数、链接发生频度等特征。本文主要对以上节点属性和链接属性进行分析。

2) 参数选择

参数选择主要是式(8)中的 α , 其主要表示为在社区划分中节点属性和链接属性的权重比例。在社交网络中,由于社区划分时节点属性和链接属性偏重不同, α 取值也不尽相同。当偏重于节点属性时 α 应大于 0.5; 偏重于链接属性时,其值应小于 0.5; 当两者相当时,其值应在 0.5 附近。本文选取 6 个数值 (分别为 0.05, 0.25, 0.45, 0.55, 0.75, 0.95) 进行实验,从而来确定参数的大致选择范围。

对社区划分结果的质量衡量问题,采用改进后的模块度来计算划分后的社区模块性,其值越趋近于 1 说明划分结果的社区模块性越好,质量越高。由于本文算法是以凝聚算法为基础设计的,故实验中将本文算法与经典凝聚算法——CNM 算法进行了比较。具体实验结果如表1所示。

表1 CDASW 与 CNM 实验结果比较

算法	参数 α	Q_{im}	时间消耗/ms
CDASW	0.05	0.623 5	783
	0.25	0.650 7	775
	0.45	0.655 3	780
	0.55	0.656 7	782
	0.75	0.672 1	772
	0.95	0.668 4	785
CNM	/	0.408 7	347

3) 实验结果比较分析

从实验结果来看,本文算法明显好于没有考虑节点属性和链接属性的 CNM 算法。在选取的参数中,当 $\alpha=0.05$ 时可以近似地认为此时是以链接属性来进行划分的,而当 $\alpha=0.95$ 时可以近似地认为是以节点属性信息来进行划分的。由对数据的分析得知本数据集社区划分时链接属性相对于节点属性而言更重要,相应的链接属性权重应大于 0.5。从实验结果看,当 $\alpha=0.75$ 时,CDASW 算法的划分质量最好,可以得出真实的 α 取值应该在 0.75 附近。实验结果验证了该数据集划分时链接属性的重要性。

由实验结果可以看出,社交网络中节点属性、链接属性对社区发现也起着很重要的作用,单一地依靠链接信息进行社区发现存在很大缺陷。在算法时间复杂度方面,由于 CDASW 是以凝聚算法为基础进行设计的,算法复杂度也比较低,在所选的参数中,时间消耗均在 780 ms 左右。

3.2 带属性的电话网用户数据实验

为了验证本文算法对真实数据的效果,本文依托所在课题组采集的某单位工作电话网数据进行实验。数据内容为该电话网的 2012 年 5 月呼叫详细记录(CDR)。

1) 数据预处理

根据课题组预先的调查结果,该实验网络用户可划分为 13 个社区,本文也熟悉了几乎全部用户的年龄、兴趣、工作所属小单位等个体属性。从 5 月份的 CDR 数据中可分析得知,共有 263 个节点(电话用户)和 1 349 条边(通话关系,两个用户间多次通话也视为 1 个通话关系),并得到了每个通话的平均通话长度、每天通话的平均次数、通话时段等特征。本文主要对以上用户属性(节点属性)和通话属性(链接属性)进行分析。该数据集划分时节点属性和链接属性的作用相当。

2) 参数选择

参数 α 同样选取 6 个数值(分别为 0.05, 0.25, 0.45, 0.55, 0.75, 0.95)进行实验。各参数数值下划分出的社区情况如表 2 所示。

表 2 不同参数 α 时划分社区数目比较

划分社区数	参数 α					
	0.05	0.25	0.45	0.55	0.75	0.95
	14	14	12	11	11	10

由表 2 可以看出当参数取值在 0.25 和 0.45 时,划分社区数目与真实社区数目最为接近。参数选择下实验结果如图 4 所示。

对社区划分结果的质量衡量问题,采用改进后的模块度和 NMI 值来进行衡量。本文选取参数 $\alpha=0.45$ 的 CDASW 算法与 CNM 算法进行比较。具体实验结果如图 5 所示。

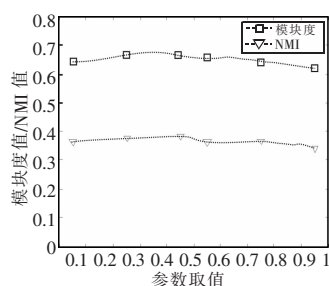


图 4 参数 α 选择下实验结果比较

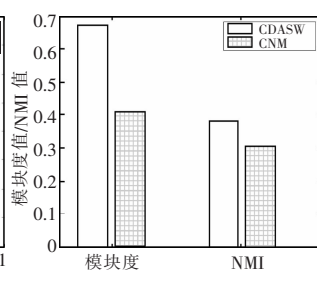


图 5 CDASW 算法与 CNM 算法结果比较

3) 实验结果比较分析

由于该数据集的节点属性和链接属性重要性相当,所以相应的 α 取值在 0.5 左右。从图 4 可以看出 $\alpha=0.45$ 时,CDASW 算法的划分质量比较好,与真实网络的社区结构也最接近,该实验结果同样也验证了该数据集的节点属性和链接属性的重要性。从图 5 可以看出,本文算法由于加入了对节点属性和链接属性的考虑,在对真实社交网络的社区划分中划分效果明显好于没有考虑节点和链接属性的 CNM 算法。

4 结束语

本文针对社交网络的社区发现问题,提出了一种融合属性的相似权值社区划分算法,结合有权有向网络对模块度进行了适当的改进,并将此算法应用于现实社交网络的社区划分问题中。以相似权值作为划分社区的依据,划分出的社区具有较高的模块度。该算法具有很好的适用性和有效性,特别是针对节点属性信息对社区划分影响比较大的网络,效果更明显。目前该算法对于节点属性过多时如何考虑以及链接属性的系数设定没有深入的研究;此外,对于链接属性的权值设定考虑也过于简单。社交网络的社区划分中如何划分出更真实有效的结果,提高 NMI 值,这些问题都需要进一步深入的研究。

参考文献:

- [1] NEWMAN M E J. The structure and function of complex networks [J]. *SIAM Review*, 2003, 45(2): 167-256.
- [2] KERNIGHAN B W, LIN S. An efficient heuristic procedure for partitioning graphs [J]. *Bell System Technical Journal*, 1970, 49(2): 291-307.
- [3] POTHEN A, SIMON H D, LIOU K P. Partitioning sparse matrices with eigenvectors of graphs [J]. *SIAM J Matrix Anal Applic*, 1990, 11(3): 430-452.
- [4] PALLA G, FARKAS I J, PLLNER P, et al. Directed network modules [J]. *New Journal of Physics*, 2007, 9: 186.
- [5] NEWMAN M E J. Fast algorithm for detecting community structure in networks [J]. *Physical Review E*, 2004, 69(6): 66-133.
- [6] CLAUSET A, NEWMAN E J, MOORE C. Finding community structure in very large networks [J]. *Physical Review E*, 2004, 70(6): 66-111.
- [7] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks [J]. *Physical Review E*, 2004, 69(2): 26-113.
- [8] 王天宇. 社交网络服务的用户群体特征分析与组织探测研究 [D]. 北京: 北京交通大学, 2009.
- [9] ESTER M, GE Rong, GAO B J, et al. Joint cluster analysis of attribute data and relationship data: the connected k -center problem [C] // Proc of SDM. 2006.
- [10] MOSER F, GE Rong, ESTER M. Joint cluster analysis of attribute and relationship data without apriori specification of the number of clusters [C] // Proc of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2007: 510-519.
- [11] 林有芳, 王天宇, 唐锐, 等. 一种有效的社会网络社区发现模型和算法 [J]. *计算机研究与发展*, 2012, 49(2): 337-345.
- [12] DANG T A, VIENNET E. Community detection based on structural and attribute similarities [C] // Proc of the 6th International Conference on Digital Society. 2012.
- [13] 汪廷华, 田盛丰, 黄厚宽. 特征加权支持向量机 [J]. *电子与信息学报*, 2009, 31(3): 514-518.
- [14] DANON L, DIAZ-GUILERA A, DUCH J, et al. Comparing community structure identification [J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2005, 2005(9): 09008.