

# 地点网络中的社区发现

郑香平 於志勇 温广槟

(福州大学数学与计算机科学学院 福州 350116)

(福州大学福建省网络计算与智能信息处理重点实验室 福州 350116)

**摘要** 地点网络可从一些独特的视角来刻画城市的空间结构。通过研究城市地点网络的特点及其与传统社交网络的区别,提出了基于地点网络的社区发现算法。该算法综合考虑地点临近性、地点间的连接和用户出行行为的相似性,先进行初始社区的划分,再反复迭代计算各地点隶属于本社区的程度,对隶属度较低的地点进行调整直到收敛,从而发现有意义的城市社区。通过分析社区内部地点的属性和关联,验证了算法的有效性。

**关键词** 群智感知,社区发现,地点网络

**中图分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.06.008

## Community Discovery in Location Network

ZHENG Xiang-ping YU Zhi-yong WEN Guang-bin

(College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China)

(Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing, Fuzhou University, Fuzhou 350116, China)

**Abstract** The location network can portray the spatial structure of city from some unique perspectives. By studying the characteristics of urban location network and its difference with traditional social network, a community discovery algorithm based on location network was proposed. The algorithm takes into account the proximity of location, the connection between the locations and the similarity of user's travel behavior. Firstly, the initial community is divided. Then, the extent of each site belonging to this community is iteratively calculated the places with lower membership degree are adjusted until convergence, so as to find significant urban communities. The validity of the algorithm was verified by analyzing the attributes and correlations of the internal sites.

**Keywords** Crowd sensing, Community discovery, Location network

## 1 引言

城市是一个复杂、庞大的系统,其中存在着多种多样的结构,如经济结构、产业结构、社会结构和空间结构等。城市空间结构是多种结构图谱的基础,是城市功能组织在地域空间上的投影。在城市空间配置的过程中,优良的城市空间结构可以使城市资源配置的效益最大化,使社会资源得到有效的利用,从而产生良好的经济效益、社会效益和环境效益,促进城市的可持续发展。但是,随着城市规模的不断扩大,城市发展的决策者越来越难详细地了解整个城市的总体布局,投资者很难掌握一个城市的真实需求,这些信息的缺失对城市的健康发展产生了不小的阻碍。

随着带有 GPS 等定位功能的智能手机的普及,带有位置信息的社交网络已经得到了长足的发展,吸引了数以百万计的用户。这些用户每天都产生大量的位置数据,其中包含很多个人行为、城市活动规律等信息,创造了大尺度上理解集体

用户行为的机会。

收集到大量用户的位置数据后,将其按时间线连接起来,就可以得到用户的行动轨迹和相应的地点网络。社区(community)是网络最普遍和最重要的拓扑结构之一,对于理解节点功能,发现隐含模式,预测其行为有着极其重要的作用。传统的社区发现算法在社交网络、生物网络和互联网中均有着广泛的应用,而城市地点网络必然也存在社区结构,因此基于地点网络的社区发现可以加深对城市空间结构的理解。

前期研究发现,地点网络具有高度数节点聚类系数高、低度数节点聚类系数低的特点,即传统的社区发现算法很难在地点网络中发现节点间真正的潜在联系或具有相似属性的节点。因此,本文给出一个城市社区的新定义,综合考虑地点临近性、地点间的连接、用户出行行为的相似性,提出相应的城市社区发现算法,通过反复调整和迭代收敛的计算过程,完成地点到社区的最佳匹配。最后,针对多个具体城市,对社区的划分结果进行分析和刻画,发现该种城市社区具有一定的现

到稿日期:2017-03-11 返修日期:2017-05-30 本文受国家自然科学基金(61300103)资助。

郑香平(1992-),男,硕士生,主要研究方向为移动社交网络;於志勇(1982-),男,博士,副教授,CCF 会员,主要研究方向为普适计算、移动社交网络,E-mail:yuzhiyong@fzu.edu.cn(通信作者);温广槟(1991-),男,硕士生,主要研究方向为移动社交网络等。

实意义,如居民区与附近的超市、医院、小学隶属于同一社区,同一大学的多个校区虽然距离较远但隶属于同一社区。

## 2 相关工作

社区结构是网络中具有极大研究意义的结构。对网络中社区结构的研究有助于加强人们对网络结构的了解,具有重要的理论意义和很高的应用价值。同时,进行社区发现的新方法不断被提出,应用领域不断被拓展,网络的社区发现方法已经成为了图论、复杂网络、数据挖掘等领域的重要组成部分。

社区发现方法受到广泛的关注<sup>[1-8]</sup>。传统意义上的社区发现方法仅从网络拓扑结构出发来挖掘连接紧密的簇结构;随着复杂网络研究的不断扩展,特别是在线社交网络研究的逐步深入,相关学者试图利用节点和边的内容来发现在线社交网络的社区。燕飞等人<sup>[9]</sup>提出了一种综合兴趣和网络拓扑结构的社区发现方法;Yang 等人<sup>[10]</sup>提出了一种将内容与链接相结合的概率模型;针对 Twitter 社交网络中的社区发现,Kewalramani 利用 Twitter 多个属性的相似性,通过传统的聚类算法来发现社区<sup>[11]</sup>。

早期对网络的研究大多聚焦于许多网络共有的属性:小世界属性、幂律度分布和网络传递性。Girvan 侧重于研究社区结构<sup>[12]</sup>,提出了一种通过中心性指数来寻找社区边界的探测方法。随后,他们在此基础上又提出了模块度<sup>[13]</sup>的概念,模块度可以很好地评价社区结构的强度以及社区划分的效果。Wilkinson 等人提出了一种从生物医学文献中自动识别功能相关基因社区的方法<sup>[14]</sup>,使用基于中心性的算法来识别巨型组件中的社区结构。他们提出的方法效果显著,提取的信息是详细的,适用于一个特定的问题。Palla 等人认为社会中个人之间的丰富的互动结果将导致复杂的社区结构<sup>[15]</sup>。由于个人的活动和沟通模式频繁变化,相关的社交和通信网络不断演变。这些研究结果可被用于洞察小团体和大型机构之间动态的基本差异。

当前已有的识别大规模真实世界网络中的社区结构的算法均需要先验信息(如社区的数量和大小),且算法复杂度高,无法分析大规模网络。Raghavan 等人提出了一种简单的标签传播算法,该算法仅以网络结构作为划分网络的基础,既不需要优化预定义的目标函数,也不需要关于社区的先验信息<sup>[16]</sup>。通过将算法应用于社区结构已知的网络来验证其有效性,证明该算法只需要线性的时间。

## 3 相关定义与城市社区发现算法

城市地点网络模型通过收集用户的签到数据来生成用户的行为轨迹,基于大量的用户数据来模拟城市居民的日常出行行为,这些数据蕴含了城市人口流动的信息。我们希望对城市地点网络进行社区结构的研究,而城市网络模型内部高聚集、边缘呈星状的特性,导致了一般的社区发现算法在该网络中无法获得有效的结果,其原因与该网络的生成原理以及城市居民出行的特性有关。

### 3.1 城市 POI 节点及特性分析

在地理信息系统中,一个 POI(Point of Interest)可以是一家商铺、一个游乐场、一个公交亭、一栋房子等。本文中,一个 POI 节点是一个蕴含信息的城市节点,包括节点的名称、地址、城市、类型以及经纬度,在现实中是实际存在的,具有现实意义,也是用户签到所依赖的节点。

分析出城市地点网络的模型结构不适合于一般的社区发现算法后,我们对城市地点网络进行了进一步的理解和分析。通过截取部分网络进行研究以及结合对用户出行行为的思考,发现具有相同性质的地点之间的直接联系是很少的,例如,一个小区中的用户不会在多个小区内的不同地点签到,即小区内部的连接是极少的,这使得一般的社区发现算法很难将一个小区的 POI 点聚集在一起,这与我们对城市社区结构的理解是相违背的。图 1 给出了一个用户行为分析的范例。

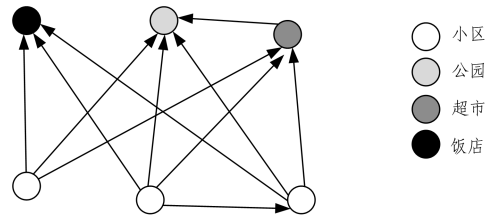


图 1 用户出行示例图

Fig. 1 Example of user travel

通过观察该范例可以发现,用户会在小区内移动,但在不同地点进行签到的情况很少,用户更多地是去往周边的公共场所,这揭示了城市网络生成的一个规律。该网络记录的是城市居民的移动规律,因而不同于社交网络或生物网络,两个 POI 节点间即使存在直接连线,也不代表两个节点的属性相似,因此进行社区分类时不应该通过 POI 节点间的联系来判断其联系的紧密程度。

为了找到性质相近的 POI 节点,我们再次对节点的行为进行观察,结果发现,虽然几个小区节点之间的连线很少,但是他们的行动却是相似的,即几个小区节点都有去往公园、超市、饭店等节点的行为。结合我们对现实世界的认知可以发现,虽然很多 POI 节点之间缺少直接联系,但是相同性质的节点间往往会存在相同的行为。因此,提出基于城市地点网络的社区发现算法,该算法结合了地点网络的特点,通过判断 POI 节点出入行为的相似性来进行城市社区结构的发现。

### 3.2 社区定义

从 3.1 节给出的简单例子可以看出,城市地点网络中节点的关联并不仅仅由节点间的连线决定;连线在现实意义上表示了人的流动,因此节点的行为即节点的人口流出分布和人口流入分布更大程度地体现了节点的特征。因此,传统的社区定义并不完全适用于城市地点网络。鉴于此,下文提出了一个社区的新定义。

城市地点网络  $G = \{V, E\}$ , 其中  $V$  为顶点集合  $\{v_1, v_2, \dots, v_n\}$ ,  $E$  为边集合  $\{e_1, e_2, \dots, e_r\}$ , 我们希望得到一个点集合的划分  $V_1, V_2, \dots, V_m$ , 相同集合中的点具有相似的出行向量和进入向量。将具有这种特征的点的集合称为地点网络中的社区结构。

### 3.3 算法流程

根据前文的分析,可通过计算 POI 节点的出入行为相似性来划分城市地点网络的社区结构。算法的流程如下:1)首先通过寻找城市中 POI 节点较为密集的区域,生成初始的社区分布;2)提取出城市地点网络的 POI-社区出行矩阵  $M$ 、社区-社区出行矩阵  $N$ 、POI-社区进入矩阵  $M'$  和社区-社区进入矩阵  $N'$ ;3)计算 POI 向量和社区向量的相似性,为每个 POI 节点选择最相似的社区,将其划分到该社区;4)重复步骤 2)和步骤 3),直至每次迭代改变的节点数小于阈值。

该算法通过比较每个节点行为与社区总体行为的相似性来决定每个节点的归属,因此算法运行之初需要一个初始的社区分布。本文通过发现城市中 POI 节点较为密集的区域来圈定初始社区的分布,并将其作为算法的初值输入。如图 2 所示,首先判断与节点相距 1000 m 范围内的 POI 点数量,选择 POI 数量最多的点为第一步的聚集点;然后在圈定的范围中再次以其中 200 m 范围内 POI 数量最多的节点作为聚集点;最后选择该聚集点周围一定距离内的 POI 节点作为一个社区,在剩下的 POI 中重复以上步骤,直到所有 POI 点找到所属社区。可以通过控制圈定社区的距离来控制所得到的初始社区的数量,同时随着 POI 节点都找到归属社区,最终会产生许多零散的 POI 点。为了不产生仅有少数 POI 的初始社区,可以设定一个阈值,若聚集到的 POI 数量太少,则将其归属到最近的聚集点所在的社区中。

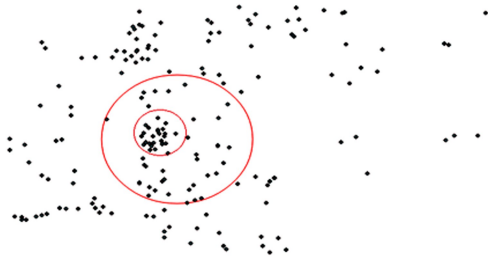


图 2 初始社区生成图

Fig. 2 Initial community generation map

生成最初的社区划分  $C_{begin}$  后,即可根据该划分提取每个 POI 节点的 POI-社区出行矩阵  $M = \{u_1, u_2, \dots, u_n\}$  (其中,  $u$  为 POI 出行分布向量,表示用户去了哪些社区,  $n$  为 POI 数量)和社区-社区出行矩阵  $N = \{c_1, c_2, \dots, c_m\}$  (其中,  $c$  为社区出行分布向量,表示该社区节点的出行分布情况,  $m$  为社区的数量)。再提取每个 POI 节点的 POI-社区进入矩阵  $M' = \{u'_1, u'_2, \dots, u'_n\}$  (其中,  $u'$  为 POI 进入分布向量,表示有哪些社区用户到达该节点)和社区-社区进入矩阵  $N' = \{c'_1, c'_2, \dots, c'_m\}$  (其中,  $c'$  为社区进入分布向量,表示有哪些社区的用户进入该社区)。然后综合这些数据,计算出每个 POI 节点分别与各个社区的相似性  $S$ 。最后将该 POI 节点分配到相似性最高的社区中。重复这个过程,直到社区结构稳定,即每次迭代发生变动的 POI 节点小于一定的阈值(因为少数节点会产生震荡,所以无法完全停止迭代)。相似性  $S$  通过余弦进行计算。本文将综合考虑以下 2 个指标:

$$S_{out_{i,j}} = \frac{\sum_{k=1}^m (M_{i,k} N_{j,k})}{\sum_{k=1}^m M_{i,k}^2 + \sum_{k=1}^m N_{j,k}^2} \quad (1)$$

$$S_{in_{i,j}} = \frac{\sum_{k=1}^m (M'_{i,k} N'_{j,k})}{\sum_{k=1}^m M'_{i,k}^2 + \sum_{k=1}^m N'_{j,k}^2} \quad (2)$$

余弦值越接近 1,表明夹角越接近 0 度,即两个向量越相似;夹角等于 0,表明两个向量相等。以上即为余弦相似性。

算法流程的代码描述如算法 1 所示。

#### 算法 1 社区划分算法

输入:POI 数据集  $\{P\}$ , OD 数据集  $\{T\}$

输出:社区划分结果  $C$

生成初始社区划分结果  $C_{begin}$

while  $t < 100 // t$  为迭代结束的阈值

计算当前社区划分下的  $M_{n \times m}, N_{m \times m}, M'_{n \times m}, N'_{m \times m}$ ;

$\alpha_1 = \alpha_2 = 1/2$ ;

$S_{max} = 0$ ;

$max = 0$ ;

$count = 0$ ;

For  $i = 1 : n$

For  $j = 1 : m$

$S_{out_{i,j}} = \text{calCosineSout}(u_i, c_j)$ ;

$S_{in_{i,j}} = \text{calCosineSin}(u'_i, c'_j)$ ;

//计算出行进入相似性

$S_j = \alpha_1 * S_{out_{i,j}} + \alpha_2 * S_{in_{i,j}}$ ;

IF  $S_j > S_{max}$

$S_{max} = S_j$  //保存相似性最高的社区和相似性值

$max = j$ ;

End for

Update  $C$  //更新节点所属社区

IF ( $C$  is changed) //判断所属社区是否变化

$count++$ ;

IF ( $count < k$ ) //如果该次迭代发生变化的节点数小于阈值

$k$

Break;

End if

End if

End for

End while

本文针对构建的城市地点网络的特点提出了该社区的发现算法。与传统的社区发现算法不同,该算法不考虑节点的直接连接,而是考虑了节点出入特征的相似性,挖掘了节点间的潜在联系。该算法的提出为社区发现算法提供了一种新思路。针对网络的特点,特别是对城市地点网络这种连接无法完全表达节点之间联系的情况,我们可以挖掘其潜在的相似特征,从而更好地发现城市的社区结构以及城市中具有潜在联系的节点。

## 4 社区划分结果的分析

根据路网数据,对天津、福州、泉州、龙岩等城市进行地块

划分。收集大众点评网和位置微博的签到数据,其中大众点评网包括 53332 个用户,105984 条签到数据;位置微博包括 589636 个用户,3186116 条签到数据。将收集到的数据以地块为基本单位进行 POI 匹配,然后进行计算;利用求得的数据再分别进行城市各地区的社区节点数量、出行向量、进入向量等的计算,从而求出各城市的空间流动情况。

本节将针对天津、福州、泉州、龙岩等城市的社区发现结果进行分析,主要基于本算法的特点提出了一个判断最终划分结果优劣的评价指标——社区内节点的整体相似度,并对划分好的社区内的节点的使用类型进行了归纳和总结,通过这些结果来评价所提算法。

4.1 社区内节点的整体相似度

由于基于地点网络的社区发现算法考量的是节点的行为相似度,因此传统的社区发现算法使用的模块度的指标对于衡量本文算法的结果缺乏实际意义。鉴于此,我们结合节点行为相似度的算法理论基础,提出一种评价社区划分优劣的指标——社区内节点的整体相似度。通过分别比较每个节点的相似度,然后判断社区整体的相似度水平,来衡量整个社区的相似度,最后对每个社区划分结果的该指标值进行分析。

首先,计算每个社区的节点相似矩阵  $K = \{s_{i,j}\}$ ,  $s_{i,j}$  为社区内第  $i$  个节点和第  $j$  个节点的相似度。社区的整体相似度的计算式如下:

$$s_{i,j} = \alpha_1 \cos(u_i, u_j) + \alpha_2 \cos(u_i', u_j') \quad (3)$$

$$S_k = \frac{1}{(n-1)(n-2)} \sum_{i,j \in c_k, 1 \leq i < j \leq n} s_{i,j} \quad (4)$$

其中,  $s_{i,j}$  为节点  $i$  和节点  $j$  的相似度,  $u$  为节点的出行向量,  $u'$  为节点的进入向量,  $\alpha_1$  和  $\alpha_2$  为各个指标的权重,本文中两个权重均设为  $1/2$ ,  $S_k$  为第  $k$  个社区的整体相似度,  $c_k$  表示第  $k$  个社区,  $n$  为该社区的节点数量。

整个网络的相似度为所有社区的整体相似度的均值:

$$S = \frac{1}{m} \sum_{k=1}^m S_k \quad (5)$$

表 1 列出了 4 个城市的初始分布和社区划分后的整体相似度,以及社区规模排名前三的社区整体相似度。可以发现,相较于初始分布,最终社区划分的整体相似度有明显的提高,都约为 0.5,划分出的社区内节点的行为在总体上有相同的趋势,包括它们的出行节点分布和进入节点分布等。上述分析说明,所提算法在发现行为相似节点方面是有效的,可以将城市地点网络中具有相似行为的节点划分到一个社区。

表 1 整体相似度

Table 1 Overall similarity

	$S_{begin}$	$S$	$S_1$	$S_2$	$S_3$
天津	0.23	0.56	0.61	0.48	0.53
福州	0.26	0.53	0.53	0.59	0.43
泉州	0.27	0.49	0.58	0.46	0.43
龙岩	0.31	0.61	0.59	0.62	0.54

表 2 列出了 4 个城市的出行相似度、进入相似度。后续工作需要在这两个指标的权重进行研究和调整,选择出最能体现节点相似度的参数进行运算。

表 2 出行相似度和进入相似度指标

Table 2 Index for out similarity and into similarity

	$S_{out}$	$S_{in}$
天津	0.53	0.52
福州	0.46	0.45
泉州	0.43	0.47
龙岩	0.59	0.56

4.2 社区划分结果的节点使用类型分析

4.1 节中提出了一个评价社区发现算法的指标,该指标主要测试的是划分出的社区的节点行为相似度,该相似度仅仅表示了节点的行为是否相近,而不具备现实意义。下面对划分的社区内节点的功能进行分析,以判断所划分的社区内节点的功能是否相近,借此来验证所提社区发现算法是否能发现城市中真实存在的社区结构。

由于最终划分的社区数量不一,因此首先选择了每个城市中节点数量最多的社区进行分析。表 3 列出了各个城市中社区节点数量最多的社区节点的功能分布情况。表 4 列出了随机抽取的各个城市中相同规模(如天津节点数最多的社区节点数为 927,则在天津市所有 POI 点中随机抽取 927 个 POI 点,并统计其土地使用类型作为比较)的 POI 节点作为对照。

表 3 社区节点的功能分布

Table 3 Function distribution of community node

(单位:%)

	餐饮	娱乐	宾馆	医疗	教育	办公	住宅	其他
天津	39	15	6	6	8	10	9	7
福州	43	17	9	4	9	5	4	9
泉州	34	16	11	5	10	6	5	13
龙岩	32	24	2	4	12	12	8	10

表 4 随机抽取相应规模节点的功能分布

Table 4 Function distribution of randomly selected nodes with corresponding size

(单位:%)

	餐饮	娱乐	宾馆	医疗	教育	办公	住宅	其他
天津	26	13	7	3	7	6	16	22
福州	22	18	3	3	9	8	13	24
泉州	21	17	6	2	6	7	18	23
龙岩	25	20	2	4	8	5	19	18

通过对照表 3 和表 4 可以发现,随机抽取的节点功能以餐饮、娱乐、住宅和其他为主,宾馆、医疗、教育和办公的比例较低,这大致体现了当前城市节点功能的分布;而经过算法迭代后的社区则拥有较突出的土地使用类型,这说明了本文提出的算法可以找到功能较为接近的节点组成的社区,而 4 个城市中节点数目最多的社区的功能以餐饮为主。

我们还统计了 4 个城市社区划分结果中各个主要土地使用类型的社区数量,如表 5 所列。划分后的社区的主要土地使用类型大部分为餐饮和娱乐,这一方面说明了当前城市中餐饮和娱乐占有很大比重;另一方面也是由新浪微博用户的签到习惯决定的,即大多数用户会选择在吃饭或游玩时进行签到,这使得餐饮和娱乐类型的节点的数据量更大,其蕴含的

信息也更为丰富。因此可以发现,餐饮和娱乐在城市网络中占据了主体地位。

表5 社区划分土地使用类型的情况

Table 5 Classification of land types in community

	总数	餐饮	娱乐	宾馆	医疗	教育	办公	住宅	其他
天津	75	33	16	3	5	2	3	8	5
福州	42	18	7	3	2	3	2	4	3
泉州	15	6	2	1	1	1	0	2	2
龙岩	8	3	1	1	0	1	0	1	1

**结束语** 本文针对城市地点网络的特点提出了不同于一般社区发现思路的基于城市地点网络的社区发现算法。该算法脱离传统社区发现算法依靠连线寻找社区结构的思路,提出了基于节点的行为相似度进行社区划分的方法,这种思路源于对城市居民日常行为的研究,具有一定的现实意义。本文还提出了评价社区划分结果的社区整体相似度的指标,该指标通过计算和整合社区内节点间的行为相似度来表征该社区的整体相似度水平,并判断算法应用在多个城市的效果。最后,对多个城市划分的社区所属节点的土地使用类型进行了分析,验证了算法可以发现节点间的潜在联系。

对城市地点网络的分析和研究,一定程度上是对城市居民日常出行的研究,具有很强的现实意义和社会价值。后续工作将继续对城市地点网络的社区结构及其评价标准进行研究。

### 参考文献

- [1] QI J S, LIANG X, WANG Y. Overlapping community detection algorithm based on selection of seed nodes[J]. *Application Research of Computers*, 2017, 34(12): 3534-3537. (in Chinese)  
齐金山, 梁循, 王怡. 基于种子节点选择的重叠社区发现算法[J]. *计算机应用研究*, 2017, 34(12): 3534-3537.
- [2] WU Z G, LU Z. A community discovery algorithm based on local similarity [J]. *Computer Engineering*, 2016, 42(12): 196-203. (in Chinese)  
吴钟刚, 吕钊. 一种基于局部相似性的社区发现算法[J]. *计算机工程*, 2016, 42(12): 196-203.
- [3] LIU Y, JI X S, LIU C X. Network community discovery optimization: based on random walk edge weight pretreatment method [J]. *Journal of Electronics & Information Technology*, 2013, 35(10): 2335-2340. (in Chinese)  
刘阳, 季新生, 刘彩霞. 网络社区发现优化: 基于随机游走的边权预处理方法[J]. *电子与信息学报*, 2013, 35(10): 2335-2340.
- [4] SONG C, ZHANG X K, FEI S, et al. Improved tag propagation algorithm based on random walk similarity matrix [J]. *Journal of Computer Applications and Software*, 2016, 33(8): 269-272. (in Chinese)  
宋琛, 张贤坤, 费松, 等. 基于随机游走相似度矩阵的改进标签传播算法[J]. *计算机应用与软件*, 2016, 33(8): 269-272.
- [5] HUANG F L, ZHANG S C, ZHU X F. Discovering Network Community Based on Multi-objective Optimization [J]. *Journal of Software*, 2013, 24(9): 2062-2077. (in Chinese)  
黄发良, 张师超, 朱晓峰. 基于多目标优化的网络社区发现方法[J]. *软件学报*, 2013, 24(9): 2062-2077.
- [6] QIAO S J, GUO J, HAN N, et al. Large-scale complex network community parallel discovery algorithm [J]. *Journal of Computer Science*, 2017, 40(3): 687-700. (in Chinese)  
乔少杰, 郭俊, 韩楠, 等. 大规模复杂网络社区并行发现算法[J]. *计算机学报*, 2017, 40(3): 687-700.
- [7] LIU H X, PENG S L. A method of community discovery based on neighboring nodes influencing intensity tag propagation [J]. *Modern Library and Information Technology*, 2015, 31(4): 58-64. (in Chinese)  
刘郝霞, 彭商谦. 一种基于邻近节点影响强度标签传播社区发现方法[J]. *现代图书情报技术*, 2015, 31(4): 58-64.
- [8] HAN L, ZHANG H. Community Discovery Method Based on Neighborhood Information [J]. *Pure Mathematics and Applied Mathematics*, 2015, 31(1): 85-92. (in Chinese)  
韩路, 张海. 基于邻域信息的社区发现方法[J]. *纯粹数学与应用数学*, 2015, 31(1): 85-92.
- [9] YAN F, ZHANG M, TAN Y W, et al. Community Discovery Method of Comprehensive Social Action Interest and Network Topology [J]. *Journal of Computer Research and Development*, 2010, 47(z1): 357-362. (in Chinese)  
燕飞, 张铭, 谭裕韦, 等. 综合社会行动者兴趣和网络拓扑的社区发现方法[J]. *计算机研究与发展*, 2010, 47(z1): 357-362.
- [10] YANG T B, JIN R, CHI Y, et al. Combining Link and content for community detection [C] // *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2009: 927-936.
- [11] KEWALRAMANI M N. Community detection in Twitter [D]. University of Maryland, 2011.
- [12] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(12): 7821-7826.
- [13] NEWMAN M E, GIRVAN M. Finding and evaluating community structure in networks [J]. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2004, 69(2): 026113.
- [14] WILKINSON D M, HUBERMAN B A. A Method for Finding Communities of Related Genes [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(Suppl. 1): 5241-5248.
- [15] PALLA G, VICSEK T. Quantifying social group evolution [J]. *Nature*, 2007, 446(7136): 664-667.
- [16] RAGHAVAN U N, ALBERT R, KUMARA S. Near linear time algorithm to detect community structures in large-scale networks [J]. *Physical Review E*, 2007, 76(2): 036106;