

## 一种基于移动用户行为的回路融合社区发现算法

肖 觅 孟祥武\* 史艳翠

(北京邮电大学智能通信软件与多媒体北京市重点实验室 北京 100876)

(北京邮电大学计算机学院 北京 100876)

**摘 要:** 针对现有基于派系的重叠社区发现算法难以对移动社会化网络实施的问题, 该文给出一种基于移动用户行为的回路融合社区发现算法。该算法首先通过分析移动用户行为构建移动社会化网络, 利用  $k$ -EC( $k$ -Elementary Circuit)简单回路发现算法寻找移动社会化网络的  $k$  阶回路作为社区核, 并按照给定的规则对社区核进行融合, 得到初步社区; 然后通过计算移动用户行为的相关度将余下的离散节点加入到相应的初步社区, 得到最终的社区; 最后通过公开数据集和仿真数据集验证了该算法在移动社会化网络社区发现方面的可行性和有效性。

**关键词:** 移动社会化网络; 社区发现; 相关度; 回路融合

**中图分类号:** TP393

**文献标识码:** A

**文章编号:** 1009-5896(2012)10-2369-06

**DOI:** 10.3724/SP.J.1146.2012.00331

## A Circuits Merging Community Discovery Algorithm Based on Mobile User Behaviors

Xiao Mi Meng Xiang-wu Shi Yan-cui

(Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia,

Beijing University of Posts and Telecommunications, Beijing 100876, China)

(School of Computer, Beijing University of Posts and Telecommunications, Beijing 100876, China)

**Abstract:** Due to the characteristics of the mobile device, it is difficult to implement the existing network discovery algorithm based on clique in mobile social network. In order to solve the problem, through analyzing the mobile user behaviors, an algorithm of community discovery algorithm based on circuit merging is proposed. Firstly, the mobile social network is constructed according to the mobile user behaviors. The method finds the  $k$ -circuit as community core by using the  $k$ -Elementary Circuits ( $k$ -EC) finding algorithm. And then the community cores that have common nodes will be merged according to some given rules. Secondly, the remainder discrete nodes are added to elementary communities according to the mobile user correlation degree. Finally, the algorithm is proved more accurate and feasible in the experiments by using the public data sets and the simulated data sets.

**Key words:** Mobile social network; Community discovery; Correlation degree; Circuits merging

### 1 引言

随着移动网络服务的日益涌现及其广泛应用, 移动网络服务类型和信息内容的增长将逐渐超出人们所能接受的范围, 加之移动设备的界面显示、终端处理、输入输出等能力有限, 将导致严重的“移动信息过载”问题<sup>[1]</sup>。如何实时、准确地为移动用户提供个性化的移动网络服务是其盈利最大化的关键。社区发现是一种解决用户需求个性化问题的可行方法, 并且广泛应用于互联网社会化网络服务中。与互联网相比, 移动网络中移动用户主要与朋友,

同事及家人等联系, 所构成的移动社会网络比较稀疏, 因此不能完全依照互联网中社会化网络服务的方法对移动社会化网络进行分析。“位置”、“实时性”、“绑定身份”和“动态交互”是移动互联网区别于传统互联网的关键特性<sup>[2]</sup>。因此将移动用户行为引入到社区发现中, 能起到很好的辅助作用, 提高移动社区发现的准确性。

移动社会化网络是一个现实网络, 每个用户可以属于多个社区, 比如朋友社区、同事社区和家人社区等, 因此本文着重研究重叠社区发现问题。文献<sup>[3]</sup>中总结了目前重叠社区发现算法, 大体上分为派系(完全子图)过滤算法和非派系过滤算法。派系过滤算法: Palla 等人<sup>[4]</sup>提出了用于发现重叠社区的派系过滤 CPM(Clique Percolation Method)算法,

2012-03-27 收到, 2012-07-02 改回

国家自然科学基金(60872051)和北京市教育委员会共建项目资助课题

\*通信作者: 孟祥武 mengmxw@163.com

如果网络中的派系非常密集,该算法将把整个网络认为是一个社区;OCBC (Overlapping Communities Based on Community Cores)算法是将派系作为初始的社区核心<sup>[5]</sup>,然后将社区核有条件地进行融合,最后将剩余节点根据给定的规则加入到相应的社区。以上两种算法都是以派系为基础,对派系较少的网络难以实施。非派系过滤算法:基于标签传播是一类主要的算法,RAK(文献[7]根据文献[6]的作者 Raghavan, Albert, Kumara 命名)算法是一个简单的标签传播算法<sup>[6]</sup>,RAK 算法受多种随机性因素影响,其结果不稳定,每次产生的社区结构存在一定差异;COPRA(Community Overlap Propagation Algorithm)在 RAK 标签传播的基础上<sup>[7]</sup>,使每个节点携带多个标签,通过扩展标签的方式进行重叠社区划分,但是 COPRA 需要指定  $v$  参数,此参数对社区发现的结果影响很大,所以该算法有一定的局限性。

近年来,实体用户行为的研究越来越广泛,文献[8]利用移动用户行为信息建立模型,得出用户之间的信任关系及其信任程度,并将其与基于项目评分的相似度结合,应用到项目评分预测模型中。文献[9]提出以实体用户行为和时间戳为条件的信任预测模型,引入多维测量指标度量实体交互满意度,使得满意度计算更加精确。这些研究结果表明实体用户行为可以很好地预测用户之间的信任关系,提高算法的精确性。

根据上述研究,本文提出一种基于移动用户行为的回路融合社区发现算法,简称 CM(Circuits Merging)算法。该算法具有以下意义:(1)相比于目前的基于派系的社区发现算法,CM 算法对派系较少的网络可以得到更好的社区划分;(2)提出的回路融合策略和节点添加策略能有效地提高模块度;(3)通过加入移动用户行为,使得社区的划分具有了移动社会化网络的特性,提高了划分的准确性,这样的社区为文献[10]中阐述的移动服务推荐提供了基础。本文第2节提出和分析了基于移动用户行为的回路融合社区发现算法;第3节利用公开数据集和仿真数据集进行实验验证,结果表明了该算法的有效性和合理性;最后给出结论。

## 2 算法设计及分析

### 2.1 移动社会化网络模型

设图  $G = \langle V, E \rangle$  是一个简单无向图,顶点集为  $V = \{v_1, v_2, \dots, v_N\}$ ,其中  $v_i$  表示移动通信网中的移动用户;边集为  $E$ ,如果移动用户  $v_i$  与  $v_j$  有通话记录,则  $v_i$  与  $v_j$  用无向边相连,记为  $e_{ij} = \langle v_i, v_j \rangle$ ;  $N$

表示移动用户的个数,本文用网络图来对移动社会化网络进行建模。现实世界中通话方式分为主叫方和被叫方,但实际研究中不能简单通过主叫或被叫来判断两人的地位关系,所以本文将这种有向关系转化为无向关系进行研究。

### 2.2 基于移动用户行为的移动社会化网络构建

与互联网及固话网相比,移动用户行为具有如下特征:(1)移动用户行为的可确定性,这种可确定性表现为移动通信设备通常是一个号码对应一个确定的用户,这是本文研究移动用户行为的基础,而固定电话和互联网没有这种对应关系,用户行为不可确定;(2)上下文信息,与互联网及固话网相比,由于移动网络的特点,移动用户受周围环境的影响更加明显,并且通过移动终端、传感器、GPS 以及基站等可以获取移动用户实时的上下文信息,例如时间、位置、情绪、周围人员以及社会关系;(3)移动用户可以通过语音、短信、飞信等方式进行通信,而且手机还提供拍摄、视频录制、音乐播放、网页浏览、下载应用软件、GPS 定位等功能,而固定电话除了语音通话,缺少以上这些丰富的服务体验。为了提高移动社区划分的准确性,本文将移动用户间的通信行为、移动用户与周围人员位置相处信息(位置、时长)引入到移动社区发现算法中,其中位置信息包括实验室、家、工作单位等。

**定义 1** 移动用户间总的语音通信时长 TTC (Total Time of Communication): 表示移动用户  $v_1$  与  $v_2$  总的语音通信时长,在不考虑主叫和被叫差异的前提下, TTC 定义如下:

$$TTC(v_1, v_2) = \sum_{i=1}^{m_1} ttc(v_1, v_2)_i \quad (1)$$

其中  $ttc(v_1, v_2)_i$  表示移动用户  $v_1$  与  $v_2$  第  $i$  次的通信时长,  $m_1$  表示通信次数。

**定义 2** 移动用户与周围人员总的相处时长 TTL (Total Time of Location): 表示移动用户  $v_1$  与  $v_2$  总的相处时长, TTL 定义如下:

$$TTL(v_1, v_2) = \sum_{l_j \in L} \delta_j \cdot \sum_{i=1}^{m_2} ttl(v_1, v_2, l_j)_i \quad (2)$$

其中  $ttl(v_1, v_2, l_j)_i$  表示在给定位置  $l_j$  下,移动用户  $v_1$  与  $v_2$  在位置  $l_j$  第  $i$  次的相处时长,  $m_2$  表示在该位置下的相处次数,  $L$  表示位置集合,  $n_l = |L|$ ,  $\delta_j$  表示对应位置  $l_j$  的权重,且  $\delta_j \geq 0$ ,  $\sum_{j=1}^{n_l} \delta_j = 1$ 。  $\delta_j$  根据 3.3 节(1)中介绍的遗传算法调优得到。

以移动用户  $v_1$  和  $v_2$  为例,将  $v_1$  与其联系人的所有通话时长求和得到  $TTC(v_1)$ ,  $v_1$  与  $v_2$  的所有通话时长为  $TTC(v_1, v_2)$ , 将  $v_1$  与周围人员的相处时长求

和得到总相处时长  $TTL(v_1)$ ,  $v_1$  与  $v_2$  的相处时长为  $TTL(v_1, v_2)$ 。

**定义 3** 移动用户  $v_1$  对  $v_2$  的相关度 DC (Degree of Correlation), DC 定义如下:

$$DC(v_1, v_2) = \alpha \times \frac{TTC(v_1, v_2)}{TTC(v_1)} + (1 - \alpha) \times \frac{TTL(v_1, v_2)}{TTL(v_1)} \quad (3)$$

其中  $\alpha \geq 0$  为权重参数, 用来调节通话时长和通信次数的权重,  $\alpha$  根据 3.3 节(1)中介绍的遗传算法调优得到。

**定义 4** 移动用户间相关度阈值 TDC (Threshold of DC): 在一次实验中固定不变的数值, 用来限定用户间相关度的最小值。利用这个数值把低于阈值的用户间相关度值所对应的联系去除, 即将由图  $G$  表示的移动社会化网络的相应边去除。当 TDC 设定太小, 所得到的移动社会化网络将引入一些不必要的边, 即移动用户之间的联系; 当 TDC 太大时, 所得到的移动社会化网络将丢失一些移动用户之间的联系。本文根据 3.1 节的评价方法进行多次重复试验选取合适的 TDC。

**定义 5** 移动用户间频繁相关度 FDC (Frequency of DC): 借鉴文献[11]中频繁模式的定义思想, FDC 表示移动用户间相关度值与相关度阈值的差值, FDC 定义如下:

$$FDC(v_1, v_2) = DC(v_1, v_2) - TDC \quad (4)$$

FDC( $v_1, v_2$ ) 表示移动用户  $v_1$  与  $v_2$  间通信时长的参照值。

按照以下规则对移动社会化网络模型图  $G$  进行筛选, 完成基于移动用户行为的移动社会化网络构建。

(1) 当  $FDC(v_1, v_2) \geq 0$  时, 保留图  $G$  中节点  $v_1$  与  $v_2$  所对应的边  $\langle v_1, v_2 \rangle$ ;

(2) 当  $FDC(v_1, v_2) < 0$  时, 删除图  $G$  中节点  $v_1$  与  $v_2$  所对应的边  $\langle v_1, v_2 \rangle$ 。

**定义 6** 孤立子群: 社会化网络中与其他群体没有边关联的群体称为孤立子群, 包括只有一个节点的孤立节点<sup>[12]</sup>。

### 2.3 $k$ -EC 算法 $k$ 值合理性分析及算法效率分析

通过对文献[13]所使用的 EC 简单回路生成算法进行修改得到  $k$ -EC 算法,  $k$ -EC 算法用限定生成回路的长度为  $k$  来得到适合本文社区发现的回路。

**2.3.1  $k$ -EC 算法  $k$  值合理性分析** 基于小世界理论及六度分割理论<sup>[14]</sup>, 将  $k$  的上限设定为 6。当  $k > 6$  时, 算法的效率会降低, 并且太长的回路有可能会使社区发现结果不精确, 即在密集的网络中只能发现一个或少数几个很大的社区, 不能够真正区分社区结构, 失去社区发现的意义。根据简单回路的

定义可知最小回路的长度为 3, 因此将  $k$  的下限设定为 3。综上所述,  $k$  的取值范围为  $3 \leq k \leq 6$ 。

**2.3.2  $k$ -EC 算法效率分析** 由文献[13]可知, EC 算法的时间复杂度为  $O((|V| + |E|)(|C| + 1))$ , 其中  $|V|$  和  $|E|$  分别为图  $G$  的节点数和边数,  $|C|$  为图  $G$  存在的简单回路个数。 $k$ -EC 算法在 EC 算法的基础上通过限定回路的长度的上限为  $k$ , 减少了找到所有长度大于  $k$  的回路的时间, 此处用  $|C_k|$  表示回路长度不大于  $k$  的简单回路个数, 则可以得出  $k$ -EC 算法的时间复杂度为  $O((|V| + |E|)(|C_k| + 1))$ 。EC 算法与  $k$ -EC 算法时间复杂度的差值为  $O((|V| + |E|)(|C| - |C_k|))$ , 当  $|C_k| < |C|$  时, 差值为  $k$ -EC 算法能有效减小的时间复杂度, 当  $|C_k| = |C|$ ,  $k$ -EC 算法和 EC 算法时间复杂度相同, 当网络规模越大, 时间复杂度优化越明显, 因此在大规模实际应用中  $k$ -EC 算法在效率上要优于 EC 算法。

### 2.4 CM 算法及正确性证明

#### 2.4.1 CM 算法描述

(1) 利用 2.3 节介绍的  $k$ -EC 算法提取每个孤立子群中满足规定长度的所有回路, 将得到的回路当作是社区核, 将这样的社区核的集合表示为  $\Gamma$ ; 同时将没有社区核的孤立子群看作一个独立社区, 加入  $\Gamma$  中。

(2) 社区核按照以下规则进行融合:

$$\text{count}(C_1 \cap C_2) \geq \frac{\min\{\text{count}(C_1), \text{count}(C_2)\}}{2}, \quad C_1, C_2 \in \Gamma \quad (5)$$

其中  $\text{count}(C)$  表示社区  $C$  包含的节点个数。如果社区核  $C_1$  和  $C_2$  满足式(5), 将这两个社区核进行融合, 形成一个新的社区核  $C'$ , 新的社区核  $C'$  添加到  $\Gamma$  中, 并在  $\Gamma$  中删除社区核  $C_1$  和  $C_2$ 。

(3) 如果  $\Gamma$  中没有新的社区核生成, 则  $\Gamma$  为生成的初步社区集合, 执行(4), 否则执行(2)。

(4)  $\Psi$  表示网络中没有加入任何一个社区并且与  $\Gamma$  中的一个或多个社区有邻边的节点的集合。对于任意节点  $\theta \in \Psi$ , 如果  $\theta$  只与一个社区相连, 则将  $\theta$  加入到这个社区中; 如果  $\theta$  与多个社区 ( $C_1, C_2, \dots, C_m$ ) 相连, 则分别计算  $\theta$  与每个社区中有连接边的节点的相关度之和, 记为  $DC_i (i = 1, \dots, n_1)$ 。将  $\theta$  加入  $DC_i$  取最大值时所对应的社区中, 如果  $DC_i$  的最大值有几个社区对应, 则将  $\theta$  加入对应的几个社区中。

(5) 如果所有节点都加入到一个或多个社区, 则得到了最后的社区集合  $\Gamma$ , 算法终止, 否则执行(4)。

**2.4.2 算法正确性证明** 本文选择利用文献[15]所使用的循环不变式证明方法对 CM 算法进行证明, 将

CM 算法归纳为如下的循环不定式: 在每一轮循环开始前, 对  $\forall \theta_1 \in C_i, C_i \in \Gamma$ , 这样的节点  $\theta_1$  最后属于哪个社区是确定的, 每一轮循环将没有社区归属的节点  $\theta$  归类到一个或多个社区, 使得  $\theta \in C_i, C_i \in \Gamma$ 。

**初始化:** 步骤(1)利用  $k$ -EC 算法将网络中的回路提取出来作为社区核,  $\Gamma$  为社区核集合, 步骤(2)和步骤(3)将社区核进行合并形成初步社区, 并得到新的集合  $\Gamma$ , 对  $\forall \theta_1 \in C_i, C_i \in \Gamma$ , 节点  $\theta_1$  在后面的循环中是不会再被处理, 所以节点  $\theta_1$  社区归属确定不变, 这样就证明了循环不变式在循环前是成立的;

**保持:** 在步骤(4)的循环中, 每次处理一个节点  $\theta \in \Psi$ , 计算  $\theta$  与有边连接的社区的相关度  $DC_i (i = 1, \dots, n_1)$ , 选择  $\max(DC(\theta)) = DC_i$ , 将  $\theta \rightarrow C_i$ , 因此一旦  $DC_i$  计算出来,  $\theta$  的社区归属通过这一规则确定了, 在重叠社区的环境下,  $\theta$  可能加入一个或多个社区, 使得  $\theta \in C_i, C_i \in \Gamma$ , 因此每一轮循环都可以保证循环的不变式成立;

**终止:** 由于网络中的节点数目的有限的, 即为移动用户个数  $N$ , 初始化中部分节点归属已经确定, 而步骤(4)中每次都会处理完一个节点, 因此算法会重复执行不多于  $N$  次, 最终会在步骤(5)停止, 此时  $\forall \theta_1 \in V$  都有  $\theta_1 \in C_i, C_i \in \Gamma$ , 即每个节点都有社区归属, 综上可以得出算法是正确的。

### 3 实验

本小节描述实验方案设计以及实验结果分析, 实验环境为: 2 GB 内存, 2.66 GHz 双核 CPU, Windows7 操作系统, Java1.6 开发语言, Myeclipse8.5 和 Matlab R2010a 集成环境, Mysql5.1 数据库。

#### 3.1 实验数据

**公开数据集:** 麻省理工学院多媒体实验室 MIT 收集的数据集<sup>[6]</sup>, 包括 94 个移动用户从 2004 年 9 月到 2005 年 6 月共 9 个月移动用户的行为信息, 并包括通话记录(通话和短信的时间、电话的主叫和被叫、短信的发送和接收)、位置、手机状态、蓝牙设备等信息。数据集中包括通过问卷调查得出的 94 个移动用户之间的好友关系, 以及用户在实验室内和实验室外与其他人员的相处时间等数据。

**仿真数据集:** 参考公开数据集的格式, 生成实验所需要的仿真数据集。该数据集包含移动用户通信行为列表、移动用户位置列表, 其中移动通信行为列表包含用户标识, 通信联系人标识及每次通信时长; 移动用户位置列表分为实验室内和实验室外两种情况, 分别包括用户标识, 相处联系人标识及

每次相处时长。移动用户规模设为 5000, 将其分成 5 组。假定组内用户间关系较为紧密, 组间用户较为稀疏, 即设定组内移动用户间的通信时长(相处时长)占本组用户总通信时长(总相处时长)的 80% 的比例, 组内用户和其他组的用户通信时长(相处时长)占 20% 的比例。参考公开数据集, 限定移动用户标识, 每次通信时长和每次相处时长的范围(如表 1)。为了适应于本文考察派系较少网络的社区发现情况, 需要对随机生成的数据集进行派系删减, 将网络中 90% 的派系随机去掉一条或多条边, 使得这些派系的结构不满足。

表 1 属性范围限定

属性名	取值范围
移动用户标识	[0,4999]
每次通信时长(s)	[5,500]
每次相处时长(h)	[1,5]

#### 3.2 评价方法

本文采用文献[5]中提出的适用于重叠社区发现的模块度  $Q_o$  评价方法, 定义如下:

$$Q_o = \frac{1}{2m_e} \sum_{C \in \Gamma} \sum_{vu} \alpha_{Cv} \alpha_{Cu} \left( A_{vu} - \frac{k_v k_u}{2m_e} \right) \quad (6)$$

其中  $\mathbf{A}$  代表社区社会化网络图  $G$  的邻接矩阵,  $m_e$  是网络中的总边数,  $\Gamma$  是社区的集合,  $k_v$  是节点  $v$  的度数,  $\alpha_{Cv}$  表示表示节点  $v$  属于社区  $C$  的程度, 定义如下:

$$\alpha_{Cv} = \frac{n_{Cv}}{\sum_{C \in \Gamma} n_{Cv}} \quad (7)$$

其中  $n_{Cv}$  表示在社区  $C$  中与节点  $v$  的连接边数, 由式(7)可知  $\forall C \in \Gamma, v \in V, 0 \leq \alpha_{Cv} \leq 1$ , 且  $\sum_{v \in C \in \Gamma} \alpha_{Cv} = 1$ 。

如果节点  $v$  只属于一个社区  $C$ , 则  $\alpha_{Cv} = 1$ ; 如果节点  $v$  不属于社区,  $\alpha_{Cv} = 0$ 。也就是说,  $Q_o$  与非重叠社区的模块度评价方法  $Q$  保持一致。模块度体现了社区内部节点紧密程度及社区间稀疏程度, 即社区结构的模块化程度, 模块度越高表明社区发现算法就越好。

#### 3.3 实验描述及结果分析

(1)CM 参数调优及可行性验证实验: 利用 3.1 节介绍的数据集对 CM 算法进行实验, 选定两个用户之间的相处行为为实验室内和实验室外的相处时长。首先, 对数据进行清洗, 删去通信行为为空或

通信行为记录为 0 的用户；然后用 CM 算法对移动通信网进行社区发现，得到生成的社区；由于目标函数与用户相关度密切相关，所以利用 matlab 数据包中的遗传算法(Genetic Algorithm, GA)函数分别对两个数据集参数进行调优，适应度函数定义如下：

$$|N(v_1, v_2)| = |DC(v_1, v_2) - 1| \quad (8)$$

通过遗传算法调优得到两个数据集的最优参数值如表 2 所示。

然后根据表 2 中的参数和 3.2 节介绍的评价方法对模块度和 TDC 进行调优。通过对公开数据集反复试验对参数 TDC 进行调优，得到最优的模块度值  $Q_{o1} = 0.3521$ ，此时  $TDC = 0.1$ ；对仿真数据集反复试验对参数 TDC 进行调优，得到最优的模块度值  $Q_{o2} = 0.5362$ ，此时  $TDC = 0.15$ 。

表 2 调优参数值

参数名	公开数据集	仿真数据集
$\delta_1$	0.70	0.64
$\delta_2$	0.30	0.36
$\alpha$	0.60	0.55

图 1 为公开数据集的 CM 社区图，区域 A 和区域 B 表示对应的社区 1 和社区 2 所包含的节点，区域 C 表示重叠的节点。由文献[5]所计算出的模块度可知本文计算出的模块度  $Q_{o1}$  和  $Q_{o2}$  在一个合理的范围内。由图 1 可知，本文算法很好地区分了整个网络图稀疏和密集的部分，因此可得本文提出的 CM 算法对派系较少的网络具有良好的可行性和有效性。

(2)CPM, OCBCC, CM 对比实验：第 1 节介绍的 CPM 是当前最流行的重叠社区发现算法之一<sup>[4]</sup>，OCBCC 是比较新的重叠社区发现算法<sup>[5]</sup>，CM 是本文所提出的算法，CM 的参数基于实验 1 的参数进行设置。本节利用 3.1 节介绍的数据集并结合 3.2 节介绍的评价方法对这 3 种算法进行对比实验，实验结果如表 3 所示。

表 3 模块度  $Q_o$  对比

对比算法	公开数据集	仿真数据集
OCBCC	0.1524	0.2760
CPM	0.1690	0.3597
CM	0.3521	0.5362

通过对比分析可以得出对 3.1 节介绍的公开数据集来构建的派系较少网络时，使用 CM 算法比 OCBCC 算法的模块度提高了 131%，比 CPM 算法模块度提高了 108%；对仿真数据集构建的派系较少的网络时，使用 CM 算法比 OCBCC 算法模块度提高了 94%，比 CPM 算法模块度提高了 48%。以上结果充分说明 CM 算法对于派系较少的网络和具有移动特性的网络的社区发现具有更好的性能。

#### 4 结束语

本文面向移动社会化网络，提出了一种基于移动用户行为的回路融合移动社区发现(简称 CM)算法，通过加入移动用户行为，采用所提出的回路融合算法，来对移动社会化网络进行 CM 算法社区发

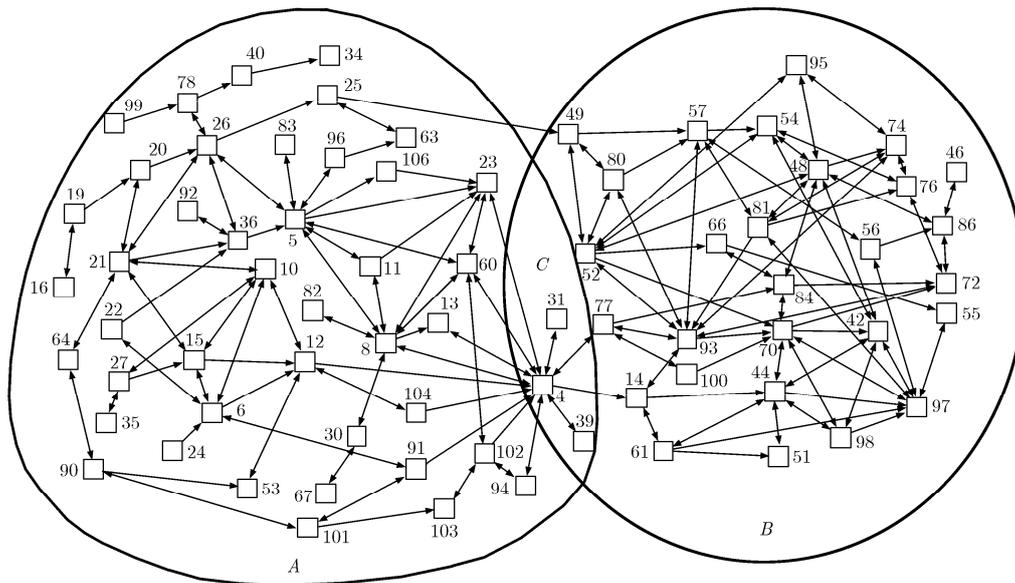


图 1 公开数据集的移动社会化网络社区图

现。实验部分,首先通过实验 1,对 CM 算法的参数进行调优并且对该算法的有效性进行了验证;然后通过实验 2,验证了 CM 算法对于派系较少网络的社区发现具有更好的模块度  $Q_o$ 。实验结果表明,CM 算法能够解决移动社会化网络的派系较少的网络社区发现的问题,并且具有较好的模块度  $Q_o$ 。由于 CM 算法基于简单回路,相比于基于派系的算法条件更加宽松,所以对于一般派系足够多的网络来说,也能够实施 CM 算法。

### 参考文献

- [1] Chiu Po-huan, Kao Yi-ming, and Lo Chi-chun. Personalized blog content recommender system for mobile phone users[J]. *International Journal of Human-Computer Studies*, 2010, 68(8): 496-507.
  - [2] 刘经南, 郭迟, 彭瑞卿. 移动互联网时代的位置服务[J]. 中国计算机学会通讯, 2011, 12(7): 40-50.  
Liu Jing-nan, Guo Chi, and Peng Rui-qing. The location-based services of mobile Internet era[J]. *Communications of China Computer Federation*, 2011, 12(7): 40-50.
  - [3] Santo F. Community detection in graphs[J]. *Physics Reports*, 2010, 486(3-5): 75-174.
  - [4] Palla G, Derenyi I, and Farkas I. Uncovering the overlapping community structure of complex networks in nature and society[J]. *Nature*, 2005, 435: 814-818.
  - [5] Shang Ming-sheng, Chen Duan-bing, and Zhou Tao. Detecting overlapping communities based on community cores in complex networks[J]. *Chinese Physics Letters*, 2010, 27(5): 058901(1-4).
  - [6] Raghavan U N, Albert R, and Kumara S. Near linear time algorithm to detect community structures in large-scale networks[J]. *Physical Review E*, 2007, 76(3): 036106(1-11).
  - [7] Gregory S. Finding overlapping communities in networks by label propagation[J]. *New Journal of Physics*, 2010, 12(10): 103018.
  - [8] 黄武汉, 孟祥武, 王立才. 移动通信网中基于用户社会化关系挖掘的协同过滤算法[J]. 电子与信息学报, 2011, 33(12): 3002-3007.  
Huang Wu-han, Meng Xiang-wu, and Wang Li-cai. A collaborative filtering algorithm based on users' social relationship mining in mobile communication network[J]. *Journal of Electronics & Information Technology*, 2011, 33(12): 3002-3007.
  - [9] 李峰, 申利民, 司亚利, 等. 一种基于实体用户行为和时间戳的信任预测模型[J]. 电子与信息学报, 2011, 33(5): 1217-1223.  
Li Feng, Shen Li-ming, and Si Ya-li, et al. A trust forecasting model based on entity context and time stamp[J]. *Journal of Electronics & Information Technology*, 2011, 33(5): 1217-1223.
  - [10] Wang Li-cai, Meng Xiang-wu, and Zhang Yu-jie. A heuristic approach to social network-based and context-aware mobile services recommendation[J]. *Journal of Convergence Information Technology*, 2011, 6(10): 339-346.
  - [11] 万里, 廖建新, 朱晓民. 一种时间序列频繁模式挖掘算法及其在WSAN行为预测中的应用[J]. 电子与信息学报, 2010, 32(3): 682-686.  
Wan Li, Liao Jian-xin, and Zhu Xiao-min. Time series frequent pattern mining algorithm and its application to WSAN behavior prediction[J]. *Journal of Electronics & Information Technology*, 2010, 32(3): 682-686.
  - [12] 吴鹏, 李思昆. 适于社会网络结构分析与可视化的布局算法[J]. 软件学报, 2011, 22(10): 2467-2475.  
Wu Peng and Li Si-kun. Layout algorithm suitable for structural analysis and visualization of social network[J]. *Journal of Software*, 2011, 22(10): 2467-2475.
  - [13] Radde N. Fixed point characterization of biological networks with complex graph topology[J]. *Bioinformatics*, 2010, 26(22): 2874-2880.
  - [14] 黄永生, 孟祥武, 张玉洁. 基于社会网络特征的P2P内容定位策略[J]. 软件学报, 2010, 21(10): 2622-2630.  
Huang Yong-sheng, Meng Xiang-wu, and Zhang Yu-jie. Strategy of content location of P2P based on the social network[J]. *Journal of Software*, 2010, 21(10): 2622-2630.
  - [15] Cormen H, Leiserson E, and Rivest L, et al. Introduction to Algorithms[M]. Second Edition, Beijing: Higher Education Press, 2002, Chapter 2.
  - [16] Eagle N, Pentland A, and Lazer D. Inferring friendship network structure using mobile phone data[J]. *Proceedings of the National Academy of Sciences*, 2009: 106(36): 15274-15278.
- 肖 冕: 男, 1987年生, 硕士生, 研究方向为社会化网络分析方法与推荐系统。  
孟祥武: 男, 1966年生, 博士, 教授, 博士生导师, 研究领域为网络服务、智能信息处理、通信软件。  
史艳翠: 女, 1982年生, 博士生, 研究方向为用户偏好自适应技术、推荐系统、个性化服务。