

# 基于兴趣和行为预测的移动社交网络动态资源发现机制

李致远<sup>1,2</sup>, 陈汝龙<sup>1</sup>, 王汝传<sup>2</sup>

(1. 江苏大学计算机科学与通信工程学院, 江苏 镇江 212013; 2. 南京邮电大学江苏省无线传感网高技术研究重点实验室, 江苏 南京 210003)

**摘要:** 针对时延容忍的移动社交网络中的资源发现问题, 在三维环境下提出一种基于兴趣和行为预测的动态资源发现机制(IBRD)。IBRD 首先从用户的文件资源和信息表中提取兴趣向量, 然后通过节点间的余弦相似度计算构造初始的虚拟兴趣社区。之后, 通过对移动社交数据的分析, 建立半马尔可夫链模型以预测节点的行为和运动轨迹。依据模型的预测结果, 实现虚拟兴趣社区的动态维护。最后, 基于动态的虚拟兴趣社区构建高效的资源发现策略。IBRD 机制在随机网络仿真器(ONE)上得以实现, 实验结果表明, IBRD 与同类模型相比, 具有较高的资源发现成功率、较低的平均时延和通信开销。

**关键词:** 移动社交网络; 动态资源发现; 余弦相似性; 行为预测; 马尔可夫链

中图分类号: TP393

文献标识码: A

## Exploiting interests and behavior prediction for dynamic resource discovery in mobile social networking

LI Zhi-yuan<sup>1,2</sup>, CHEN Ru-long<sup>1</sup>, WANG Ru-chuan<sup>2</sup>

(1. School of Computer Science and Telecommunications Engineering, Jiangsu University, Zhenjiang 212013, China;

2. Jiangsu High Technology Research Key Lab for Wireless Sensor Networks, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

**Abstract:** Resource discovery in delay-tolerant mobile social networks (MSN) continues to be a challenging issue. An interests and behavior prediction-based dynamic resource discovery mechanism (IBRD) in 3-dimensional cartesian coordinate system was proposed. Firstly, IBRD extracted the interest vectors from the user's file resources and the profile table, and then the initial virtual interest communities through the cosine similarity computation between the nodes were constructed. After mobile social networking data was analyzed, the semi-Markov chain model was used to predict the behavior and movement trajectory of users. According to the prediction results of the Markov model, the dynamic maintenance of the virtual interest communities were realized. Next, an efficient resource discovery strategy based on the dynamic virtual interest communities was designed. Finally, proposed method was simulated on the platform of the opportunistic network environment simulator. Simulation results show that the proposed scheme consistently outperforms the state-of-the-art resource discovery schemes in terms of the searching efficiency, the average delay and the communication cost.

**Key words:** mobile social networking, dynamic resource discovery, cosine similarity, behavior prediction, Markov chain

### 1 引言

近年来, 随着便携式设备(智能手机、iPad、笔

记本电脑及车载单元 OBU) 的普及与大众化, 短距离无线通信技术的成熟应用(Bluetooth、WiFi、IEEE 802.11p 等) 以及用户对移动社交提出新的需求, 催

收稿日期: 2015-05-31; 修回日期: 2015-07-21

基金项目: 国家自然科学基金资助项目(No.61202474, No.61373017, No.61572260, No.61572261); 中国博士后科学基金资助项目(No.2015M570469); 江苏省自然科学基金资助项目(No.BK20130528); 镇江市重点研发计划(社会发展)基金资助项目(No.SH2015020); 江苏大学高级专业人才科研启动基金资助项目(No.12JDG049)

**Foundation Items:** The National Natural Science Foundation of China(No.61202474, No.61373017, No.61572260, No.61572261), The Project Funded by China Postdoctoral Science Foundation(No.2015M570469), The Natural Science Foundation of Jiangsu Province(No.BK20130528), The Key Research and Development Program Foundation (Social Development) of Zhejiang(No.SH2015020), The Senior Professional Scientific Research Foundation of Jiangsu University(No.12JDG049)

生了移动社交网络 (MSN, mobile social networking) 及其相关应用的迅速发展。这种采用短距离通信的移动社交应用的优势在于,它不仅使用户可以随时随地进行信息共享,而且在资源共享时,无需考虑自身在流量使用过程中所招致的额外费用开销<sup>[1]</sup>。从物理网络的角度来看,移动社交网络是一种特殊的时延容忍网络 (DTN, delay tolerant network),因此,它也有机会网络中的连接频繁中断、数据转发采用“存储—携带—转发”模式等特点<sup>[2]</sup>。从覆盖网络的角度来看,移动社交网络是一种特殊的移动对等网络 (MP2P, mobile peer-to-peer networks),因此,它同时也具有 MP2P 网络中的自组织、拓扑频繁变化、节点自身资源受限及网络环境更加开放等特点<sup>[3]</sup>。上述特点使传统的在线社交网络 (OSN, online social network) 的资源发现和资源共享方法不再适用于 MSN,而资源发现是移动社交环境下进行资源共享的前提,是实现大规模移动社交应用首要解决的关键技术问题。将已有 DTN 和 MP2P 网络中的资源发现技术直接应用于 MSN,会导致资源发现的效率、消息的投递成功率降低以及消息转发的时延增大等问题。

鉴于此,本文在三维环境下提出一种基于兴趣和行为预测的动态资源发现机制 (IBRD, interests and behavior prediction-based dynamic resource discovery mechanism)。该模型首先从用户的文件资源和信息表中提取兴趣向量,然后通过节点间的余弦相似度计算构造初始的虚拟兴趣社区。之后,通过对 MIT Reality<sup>[4]</sup>和 Huggle Infocom 06<sup>[5]</sup>2 个国际上重要的移动社交数据分析发现节点的逗留时间服从幂律分布,因此,采用半马尔可夫链构建预测模型以预测节点的行为和运动轨迹。依据模型的预测结果,实现虚拟兴趣社区中节点间连接关系的动态维护。最后,基于动态的虚拟兴趣社区构建高效的资源发现策略。仿真实验表明 IBRD 与同类方法相比,可以有效地提高资源发现成功率、降低资源查找的平均时延及所带来的通信开销。

## 2 相关工作

为了满足移动社交环境下资源发现的需要,文献[6~14]提出了众多解决方案,归纳起来可以分为 3 类:基于洪泛的资源发现方法<sup>[6,7]</sup>、基于 Push(广告)-Pull(发现)模式的资源发现方法<sup>[8,9]</sup>和基于虚拟社区的资源发现方法<sup>[10~14]</sup>。

7DS<sup>[6]</sup>是最早将对等技术引入移动环境的洪泛

式资源发现方法,它利用广播模式在邻居节点间传播查询消息。文献[7]利用平均场方程在 MSN 环境下建立了具有社会特性的洪泛式查询消息转发机制。尽管基于洪泛的资源发现方法的命中率非常高,但地毯式的检索方法随着系统容量的增加会带来巨大的开销,这是资源受限的移动网络环境所不能承受的。

文献[8]中提出基于地理位置的 Push(广告)-Pull(发现)模式的资源发现方法。为了确保资源发现的成功率,它在交叉方向上广播广告信息(节点拥有的资源状况)。P2PSI<sup>[9]</sup>采用了 Push/Pull 混合模式的资源发现系统,不同的是它采用了群体智能思想,将共享的资源比作食物源,将路由表比作信息素。每一个资源节点周期性地广播广告信息给邻居节点,告知它们自身拥有的资源,然后,邻居节点再根据收集到的广告信息有选择地请求资源。当请求节点定位到资源节点后,会留下信息素以辅助后续的资源发现。尽管这类方法的网络开销比洪泛式资源发现方法要低,但相对移动网络而言,它产生的开销仍比较大。更重要的是,由于节点的移动性,使广告信息经常失效,致使资源发现的成功率无法得到保障。

基于虚拟社区的资源发现方法<sup>[10~14]</sup>是当前移动社交网络环境下主流的资源发现策略。它们利用了社交网络中节点间的社会特性,从而高效地将移动用户组织成虚拟社区,然后,基于所构建的虚拟社区完成移动社交环境下的资源发现。这种资源发现策略避免了前 2 种资源发现方法的盲目性和被动性,有效地提高了动态资源发现的效率。其中,文献[10,11]利用了节点的行为模式找出节点间长期的邻居关系,即将频繁接触的节点划分到同一个社区中,并指定一个 Agent 节点进行有效的管理。而文献[12~14]不仅考虑到了节点的移动性,还考虑到了节点的兴趣。不仅如此,它还将在多个社区间频繁移动的节点定义为大使节点,通过这些大使节点帮助提高社区之间的资源发现能力。但上述方案存在以下不足亟待解决。1) 只关注移动社交覆盖层面上的资源发现,导致获得的优质资源节点不在自己的通信范围之内,要经过远距离通信才能获得此资源,时常还会出现传输失败,即所谓的拓扑失配。2) 没有从时空角度出发考虑资源发现,使资源发现的效率大大降低。比如,在江苏大学计算机学院大楼(由 5 层组成)中,在上班时间内每一层楼都会有大量的师生从事各类移动社交应用。现有的短距离无线通信技术对障碍物的穿透能力弱,因此,在资源发现时,会优先选择同楼层

的节点进行资源查找。而现有的基于虚拟社区的资源发现方法都是按照二维平面进行资源发现方案的设计，此时相当于将三维立体的计算机学院大楼从顶部进行投影，并同时将用户映射到该投影面。那么，此时用户在投影面内找到的邻居节点在现实情况下可能分别是 1 楼和 5 楼的 2 名师生，这势必会影响到移动社交环境下资源发现的效率。3) 社区的构造方案丰富，但是均没有考虑社区的维护方案，这使无线网络的物理拓扑与社区的虚拟拓扑的拓扑失配矛盾更加突出，直接导致了资源搜索效率的下降。

### 3 基于兴趣和行为预测的资源发现机制

#### 3.1 虚拟兴趣社区的初始化构建

虚拟兴趣社区的构建是移动社交网络资源发现问题首要解决的关键技术，它关系到资源发现的效率。在本文中，资源和运动模式 2 个决定虚拟兴趣社区构建的关键因子采用向量的形式进行表示，如式(1)和式(6)所示。对于向量而言，衡量 2 个个体间差异大小的度量主要是采用余弦相似度准则计算。因此，本节采用对资源和运动模式向量化以及余弦相似度准则的方法实现对虚拟兴趣社区的初始化构建。

##### 3.1.1 节点兴趣的提取

有着共同兴趣的用户更倾向于加入相同的兴趣社区，他们会相互共享彼此感兴趣的资源。在本文中的资源主要指文件资源，包括文档、语音和短视频。对于文档的向量化表示可以通过文本聚类获得，对于语音及短视频的向量化表示可以通过对标题、简介等信息的关键词提取获得。用户在注册信息时可以定义自己的多个兴趣，兴趣以<关键词，权值>形式出现，从而获得节点的兴趣表示。下面通过对节点共享的资源向量进行聚类的方法来提取节点兴趣。具体描述如下。

对于一个资源  $R_i$ ，它可以被表示为

$$R_i = \{(k_1, w_1); (k_2, w_2); \dots; (k_n, w_n)\} \quad (1)$$

其中， $k_i$  表示资源  $R_i$  中的第  $i$  个关键词， $w_i$  表示第  $i$  个关键词的权重，这些序偶对依照权重进行降序排序。

利用式(2)来计算节点  $i$  上存储的共享资源  $R_u$  和  $R_v$  之间的兴趣相似度。

$$sim(u, v) = \frac{\sum_{i=1}^m w_{ui} w_{vi}}{\sqrt{\sum_{i=1}^m w_{ui}^2} \sqrt{\sum_{i=1}^m w_{vi}^2}} \quad (2)$$

其中， $m$  表示共享资源  $R_u$  和  $R_v$  中所包含的公共关键词的个数， $w_{ui}$  和  $w_{vi}$  分别表示这 2 个资源中第  $i$  个公共关键词的权重。对于关键词的权重主要通过 2 种方式获取：1) 对于文档资源，统计关键词在文档中出现的频率，将其作为该关键词的权值；2) 对于语音和短视频类的资源，主要通过用户自行选择关键词并定义其权重的方法进行设置。

由此可以得到用户节点中所有共享资源的相似度，根据相似度值进行降序排列，可以获得资源相似度列表。为了获得有效的资源表示，需要截取前  $r$  个高相似度的资源。目前，有 2 种截取方法：1) 设置相似度阈值；2) 设置选取相似度的个数。

然而，上述 2 种方法在相似度分布不稳定的情况下均会产生较差的选取效果，不能准确地表示节点兴趣。图 1 为资源的相似度分布。若采用设置相似度阈值的方法，并假设相似度阈值为 0.48，那么，图 1(a)状态下的相似度将会被正常截取，而图 1(b)状态下的情况将无法截取有效的相似度。如果相似度设置为 0.1，那么图 1(b)状态可以进行正常的截取，而图 1(a)状态则会截取过多的资源向量，导致无法有效表示节点兴趣。同理，采用设置选取相似度个数的方法，若选取相似度个数过多，会引起图 1(b)状态截取过多的低相似度资源，反之，会引起图 1(a)状态丢失过多的高相似度资源。鉴于此，利用资源相似度的影响大小，给出式(3)以有效地截取前  $r$  个高相似度资源。



图 1 资源相似度分布示意

$$r = \arg \min \left| \frac{\sum_{i=1}^r sim_{(u,v)}^i}{\sum_{i=1}^k sim_{(u,v)}^i} - r \right|, 0 < r < 1, r < k \quad (3)$$

其中， $k$  表示资源相似度列表的长度， $sim_{(u,v)}^i$  表示资源相似度列表中的第  $i$  项， $r$  是预先定义的参数。 $r$  的取值由参数  $r$  和变量  $sim_{(u,v)}^i$  的数字特征共同决定。当参数  $r$  确定之后， $r$  的取值仅由变量  $sim_{(u,v)}^i$  的数字特征决定。为了数值计算的方便，将  $r$  的默认值定义为 0.5。

紧接着，使用式(4)对所获得的前  $r$  个资源重新计算权重，并用  $\langle k_j, w_{k_j} \rangle$  序偶对的形式来表示节点

$n$  的兴趣向量  $R_n^p$ ，如式(5)所示。

$$w_{k_j} = \frac{\sum_{i=1}^r w_{R_i}^{k_j}}{r} \quad (4)$$

$$R_n^p = (k_j, w_{k_j}), j > 0 \quad (5)$$

其中， $w_{R_i}^{k_j}$  表示资源  $R_i$  中关键词  $k_j$  的权重。

### 3.1.2 节点运动模式的提取

**定义 1** 节点  $i$  的当前位置由向量  $I(x_i, y_i, z_i, t_i)$  表示，其中， $(x_i, y_i, z_i)$  表示节点  $i$  的三维坐标，该数据可从 GPS 获取， $t_i$  表示节点  $i$  滞留在位置  $(x_i, y_i, z_i)$  上的时间戳。

#### 定义 2 滞留区域和滞留时间

节点  $i$  在  $t_i$  时刻所处的位置是  $I_i(x_i, y_i, z_i, t_i)$ ，在  $t_j$  时刻所处的位置是  $I_j(x_j, y_j, z_j, t_j)$ ，若位置  $I_i$  与位置  $I_j$  之间的距离小于无线接入设备 AP 的无线覆盖范围，则节点  $i$  仍处在同一个滞留区域内。在滞留区域内停留的时间  $T_{ij}$  便是节点在该区域内的滞留时间。

为了提取节点的运动模式，需要获取节点的移动轨迹  $T_i$ ，包括节点  $i$  所处的区域  $V_i$ ，进入该区域的时间  $t_{start}$  和离开该区域的时间  $t_{end}$ ，见式(6)。

$$T_i = (V_i(x_i, y_i, z_i, list(t_{start}, t_{end}), d_1), L, V_m(x_m, y_m, z_m, list(t_{start}, t_{end}), d_n)) \quad (6)$$

其中， $d_i$  是节点  $i$  在区域  $V_i$  上滞留时间的权重，其计算方法见式(7)。

$$d_i = \frac{\sum_{i=0}^{m-1} (list(i) t_{end} - list(i) t_{start})}{t_{total}} \quad (7)$$

其中， $m$  表示运动轨迹列表的长度， $t_{total}$  表示总的时间。

### 3.1.3 虚拟兴趣社区的构建策略

#### 1) 局部兴趣社区的构建

##### 兴趣相似度计算

计算节点  $i$  和  $j$  之间的兴趣相似度  $S_{ij}^p$ ，如式(8)所示。

$$S_{ij}^p = \frac{\sum_{k=1}^m w_{ik} w_{jk}}{\sqrt{\sum_{k=1}^m w_{ik}^2} \sqrt{\sum_{k=1}^m w_{jk}^2}} \quad (8)$$

运动模式相似度计算

图 2 为节点  $i$  和  $j$  的运动轨迹。它们从时刻  $t_a$  到时刻  $t_b$  滞留在相同的区域  $V$ ，这种现象称之为运动模式的相似性。

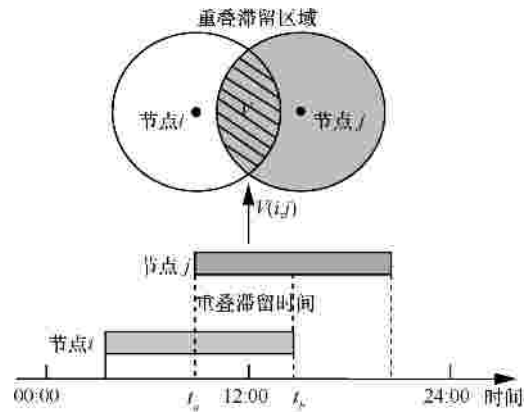


图 2 节点运动轨迹的示意

通过式(9)可以得到节点  $i$  和  $j$  的运动模式相似度。

$$S_{ij}^{mp} = \frac{3 \sum_{i=1}^n \sum_{j=1}^n V(i, j) d_i d_j}{4 n p r^3 \sqrt{\sum_{i=1}^n d_i^2} \sqrt{\sum_{j=1}^n d_j^2}} \quad (9)$$

其中， $V(i, j)$  是重叠区域  $V$  的体积， $d_i$  和  $d_j$  分别表示节点  $i$  和节点  $j$  在重叠滞留区域  $V$  的滞留时间的权重。

#### 节点间总的相似度计算

通过式(10)可以得到节点  $i$  和  $j$  之间最终的相似度。

$$S_{ij}' = a S_{ij}^p + b S_{ij}^{mp}, a + b = 1 \quad (10)$$

如果  $S_{ij}'$  大于事先设定的阈值  $T_{sim}^1$ ，则节点  $i$  和  $j$  进入相同的局部兴趣社区。

在局部兴趣社区构建完成之后，需将局部兴趣社区列表中的节点分为 3 类以提高资源发现和资源管理的效率。这 3 类节点分别是簇首管理节点、移动大使节点和普通移动节点。

簇首管理节点位于区域  $R$  的质心，节点的计算、存储和转发能力均强于其他节点，且簇首管理节点在区域  $R$  的滞留时间较长。它的主要任务包括：帮助普通移动节点在局部兴趣社区内和社区之外进行资源查找以及周期性地局部兴趣社区内去收集普通移动节点的信息并提取节点的兴趣向量。

移动大使节点是局部兴趣社区内一类特殊的

移动节点，它总是规律性地移入和移出不同的局部兴趣社区，类似城市中的公交车，总是能够按时按点进入车站。因此，充当着在局部兴趣社区之间进行信息交互桥梁的作用。

2) 局部兴趣社区的合并与全局兴趣社区的形成

局部兴趣社区的合并需要提取局部社区的兴趣向量，提取的方法与提取单个节点兴趣向量的方法和步骤类似，所不同的是首先由簇首管理节点收集局部兴趣社区内所有移动节点的兴趣向量信息，然后使用式(2)计算任意 2 个节点  $i$  和  $j$  之间的相似度  $sim(i, j)$ ，并最终得到局部社区的兴趣向量  $R_c^p$ 。

在获得局部社区的兴趣向量  $R_c^p$  之后，由移动大使节点携带兴趣向量  $R_c^p$  到它频繁到达的另一个局部兴趣社区，并交由该局部社区的簇首管理节点计算 2 个局部兴趣社区之间的相似度值，若大于阈值  $T_{sim}^2$ ，就将 2 个局部兴趣社区进行合并，直到不能再合并为止，便得到了全局兴趣社区。为了实现下一步的高效资源发现，此处需要通过从所有参与合并的局部兴趣社区中提取全局社区的兴趣向量。

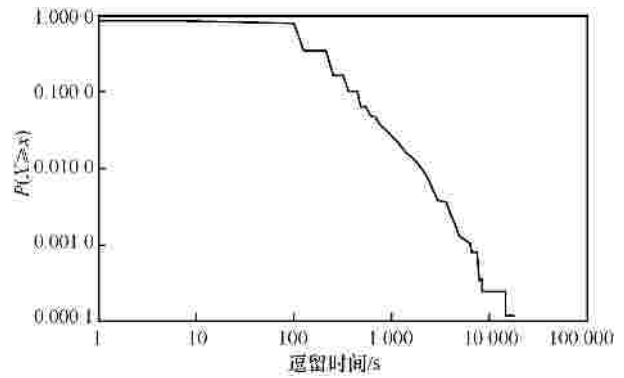
3.2 基于行为预测的虚拟兴趣社区动态维护

根据社会学规律，人与人之间的社会关系是长期且稳定的，这种关系并不会随着节点的移动而发生改变。因此，在前一部分所构造的全局兴趣社区中，社区的用户成员基本不会发生改变。但是，社区中节点之间连接关系会随着用户节点的移动而发生改变。现有的移动社交网络资源发现方案均没有考虑这种情况，从而使资源发现的效率在起初时刻较高，随着时间的推移不断降低。为了重新提高资源发现的效率，不得不反复重构虚拟兴趣社区，使方案的代价大大增加。

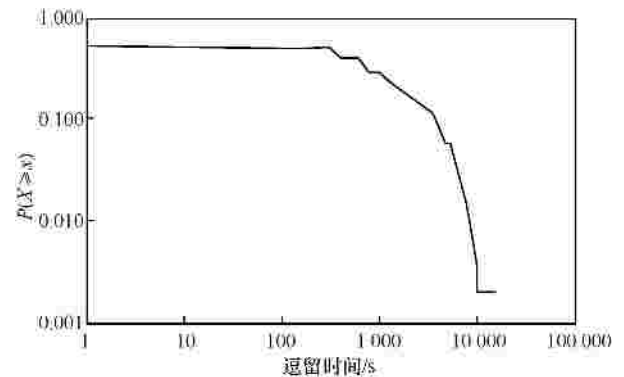
为了解决上述问题，采用马尔可夫模型为节点建立行为预测模型。然后，根据该预测模型动态实时地调整节点之间的连接关系，很好地实现了虚拟社区的动态维护。为下面进行动态的资源发现提供了强有力地保障。建模过程如下。

3.2.1 移动社交数据分析

对 MIT<sup>[4]</sup>和 Huggle<sup>[5]</sup>2 个移动社交数据集进行分析。分析重点是节点在社区中逗留时间的分布，它关系到如何构建马尔可夫预测模型。图 3 为节点在社区中的逗留时间分布，从中不难发现节点在社区中的逗留时间分布服从幂律分布，而不是指数分布，因此，它不满足连接时间马尔可夫链模型的定义。鉴于此，采用半马尔可夫链。



(a) Huggle Infocom 06 Trace



(b) MIT Reality Trace

图 3 节点逗留时间的分布

3.2.2 基于半马尔可夫链的节点移动模型构建

首先，将节点滞留的区域抽象为马尔可夫链的状态。节点在不同滞留区域间的移动具有马氏性，即节点  $e$  从一个滞留区域到另一个滞留区域的概率仅与当前的转移概率有关，与之前的转移概率无关。于是，有节点移动的离散时间的半马尔可夫链，如式(11)所示。

$$\begin{aligned}
 p_{ij}(e) &= P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} \\
 &= P\{X_{n+1} = j | X_n = i\}
 \end{aligned}
 \tag{11}$$

其中， $p_{ij}(e)$  表示节点  $e$  从滞留区域  $i$  到滞留区域  $j$  的转移概率；离散时间集合  $T = \{0, 1, 2, \dots, L\}$ ，以  $s$  为单位；离散状态集  $I = \{i_0, i_1, i_2, \dots, i_n\}$ ， $i_n$  表示节点运动的区域。

节点  $e$  的一步转移概率  $p_{ij}$  的矩阵为  $P_e$ ，如式(12)所示。

$$P_e = \{p_{ij}(e)\}
 \tag{12}$$

假设状态个数为 4，则一步转移概率矩阵  $P$  如式(13)所示。

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix}, p_{ij} = 0, i, j = 0, \dots, 4 \quad (13)$$

$$\sum_{j=0}^4 p_{ij} = 1, i = 0, 1, \dots, 4$$

将 4 个状态分别定义为家、图书馆、餐厅和办公室，则节点  $e$  移动的马尔可夫模型如图 4 所示。

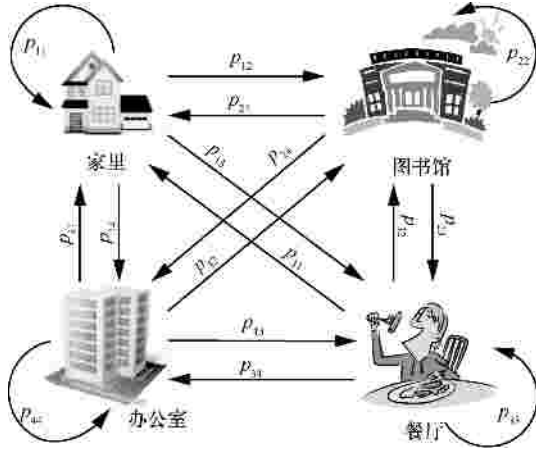


图 4 移动节点的马尔可夫模型

$p_{ij}(e)$  的计算如式(14)所示，其中  $N_{ij}$  表示节点  $e$  从位置  $i$  进入位置  $j$  的次数， $N_i$  表示节点  $e$  从位置  $i$  进入其他位置的总次数。这些统计数据可从 Infocom 06 和 MIT Reality 这 2 个数据集中获得。

$$p_{ij}(e) = \frac{N_{ij}}{N_i} \quad (14)$$

由 Chapman-Kolmogorov 方程式可得节点  $e$  的  $n$  步转移概率表示为  $p_{ij}^{(n)}(e)$ ，见式(15)。

$$p_{ij}^{(n)}(e) = \sum_{k \in I} p_{ik}^{(l)} p_{kj}^{(n-l)} \quad (15)$$

通过马尔可夫模型对节点运动规律的预测可以完成对虚拟兴趣社区中节点关系的动态维护。这样，在进行动态资源发现时，可以极大地提高资源发现的成功率、降低搜索时延和通信开销。

### 3.3 基于虚拟兴趣社区的动态资源发现机制

#### 3.3.1 局部社区内搜索

资源请求节点首先计算请求的资源向量与局部社区的兴趣向量的相似度。如果计算出来的相似度低于阈值  $T_{sim}^2$ ，表明所请求的资源存在于该局部兴趣社区的可能性较低，那么请求消息将会被发送给簇首管理节点，由簇首管理节点发起全局社区内搜索。否则，请求资源存在于该局部兴趣社区的概率较高，请求节点将会计算资源向量与同处该局部兴趣社区的邻居节点的相似度，并选择  $k$  个较高相似度的邻居节点进行搜索。在未发现资源的情况下，该过程将一直持续下去。因此，为了控制搜索深度，避免产生过多负载，为消息设置 TTL 值。当消息每被转发一次，TTL 值减 1，直到 TTL 等于 0，停止搜索。

#### 3.3.2 全局社区内搜索

当请求消息未在局部社区内搜索到资源时，请求的资源向量将会被发送给簇首管理节点，由簇首管理节点发起全局社区内的搜索。图 5 表示了全局社区内的搜索过程。簇首节点  $A$  首先选择出  $k$  个距离较近的局部社区 1-2、1-3、1-4，并通过大使节点的移动携带或者转发方式（大使节点 1 和大使节点 2 为移动携带方式，大使节点 3 为转发方式），将请求消息转发给相应的局部兴趣社区，并在内部进行局部社区内的搜索。

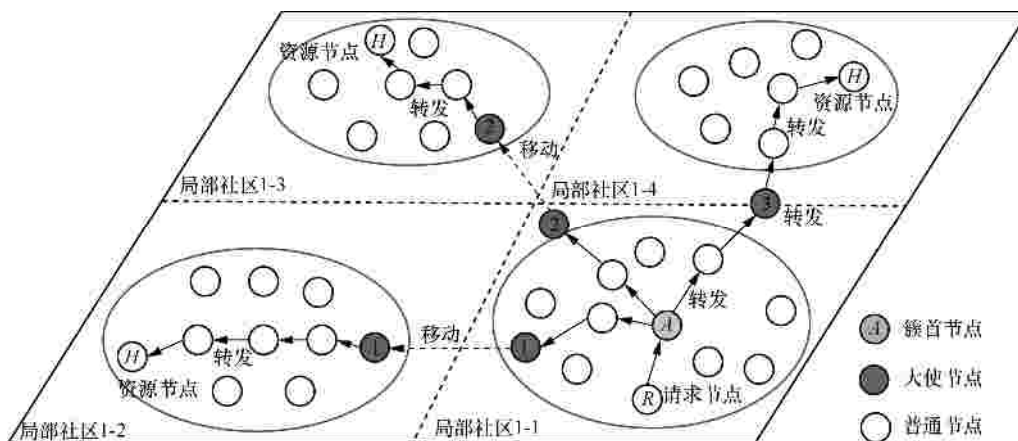


图 5 全局社区内资源搜索示意

### 3.3.3 全局社区间搜索

当请求消息与局部社区兴趣相似度小于阈值  $T_{sim}^2$  时, 请求的资源向量被发送给簇首管理节点, 由簇首管理节点发起全局社区间的搜索。如图 6 所示为全局社区间的搜索过程。簇首管理节点 A 首先通过计算自身与外部全局社区向量的兴趣相似度, 选出  $k$  个相似度较高的全局社区 2,3 和 4, 并通过大使节点的移动—携带—转发功能将消息转发给相应的全局社区, 并在相应的全局社区内部进行局部社区内和全局社区内的搜索。

## 4 IBRD 模型的仿真实验与性能评估

### 4.1 仿真实验环境及参数设置

本文选用随机网络仿真器(ONE, opportunistic network environment)<sup>[15]</sup>作为 IBRD 模型性能评估的仿真平台。人的兴趣是多样的, IBRD 构建的社区是一种可重叠的社区, 它所构建的虚拟社区个数是由用户兴趣所决定的, 是动态变化的。例如一个对计算机与篮球感兴趣的人会同时属于计算机与篮球 2 个虚拟社区, 节点在进行资源查找中根据请求信息选择所属相应社区进行独立的资源查找, 因此整个资源查找过程是一个完全并行独立的过程。用户兴趣社区个数的设置不会对性能评估带来影响。

本次仿真实验的参数如表 1 所示。

阈值  $T_{sim}^1$  与局部兴趣社区的构建结果关联紧密, 大量的计算和实验结果显示, 阈值  $T_{sim}^1$  设置过高使一个或少量用户即成为一个社区, 社区数量增加, 破坏了移动用户之间的社会联系, 造成了关系图的分割。此外, 阈值  $T_{sim}^1$  设置过高给社区的合并与维护带来了更大的开销。在本实验中将  $T_{sim}^1$  设置为 0.3。对于局部兴趣社区内资源发现的重要阈值  $T_{sim}^2$ , 如果设置较高,

资源列表中的资源节点数量较少, 甚至为空; 如果设置较低, 则在资源列表中会出现大量无用的资源节点。在本实验中将  $T_{sim}^2$  设置为 0.5。

表 1 仿真实验参数

参数	描述	默认值
MovementModel.worldSize	仿真实验场景面积	600 m × 800 m
Scenario.endTime	仿真时间	12 h
Scenario.nrofHosts	用户节点个数	100
Scenario.nrofHostGroups	分组数	5
BtInterface.transmitSpeed	传输带宽	250 kbit/s
BtInterface.transmitRange	传输距离	10~20 m
Group.movementModel	组内节点移动模型	ShortestPathMapBasedMovement
MobilitySpeed	节点的行走速度	0.5~1.5 m/s
Message.Interval	产生消息的间隔	1~2 s
Message.TTL	消息的生存周期	300 min
$r$	提取节点兴趣时的实验参数	0.5
$T_{sim}^1$	构建局部兴趣社区的阈值	0.3
$T_{sim}^2$	局部兴趣社区内资源发现的阈值	0.5
$a$	兴趣相似度的权重	0.6
$b$	运动模式相似度的权重	0.4

参数  $a$  和  $b$  分别是式(10)中的兴趣相似度的权重和运动模式相似度的权重。本文提出的 IBRD 模型利用兴趣进行资源发现, 在整个资源发现过程中, 兴趣是资源成功发现的关键。而运动模式则是为了保障虚拟社区构建的本地性, 避免拓扑失配所导致的消息无法成功投递的问题。为了保障 IBRD 模型的资源发现效率, 将兴趣相似度的权重  $a$  设置高于权重  $b$ 。同时, 为了避免过多的查询消息被投递到连通覆盖范围之外的节点上, 权重  $b$  的数值也

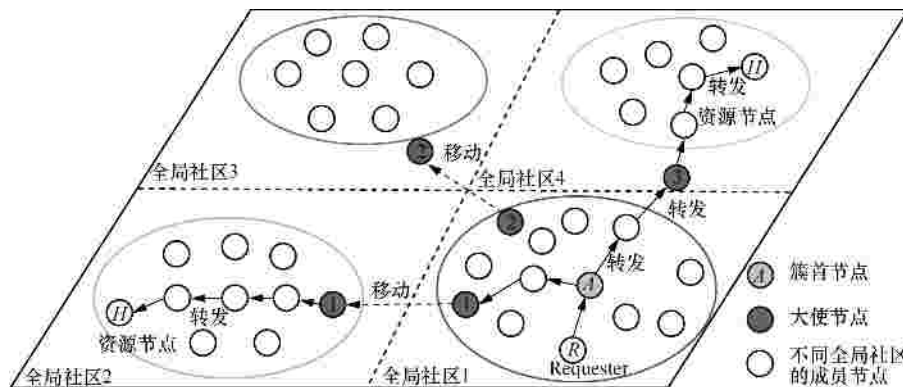


图 6 全局社区间资源搜索示意

不能设置过低。在此，将权重  $a$  设置为 0.6，权重  $b$  设置为 0.4。

本文提出的 IBRD 模型将从资源发现的成功率、资源发现的平均时延和通信开销 3 个方面与经典的基于地理信息感知的移动社交网络资源发现方案 SIG<sup>[10]</sup>和基于内容相似性的移动社交网络资源发现方案 SPOON<sup>[13]</sup>进行比较。

定义 3 资源发现成功率等于成功查询次数与发起查询的总次数的比值。

定义 4 资源发现的平均时延等于所有成功查询所花费时间的平均值。

定义 5 通信开销等于每成功发现一个资源需要对查询消息转发的平均次数。

### 4.2 性能评估

#### 4.2.1 资源发现成功率

如图 7 所示，IBRD 模型的性能最优，其次是 SPOON 模型，最弱的是 SIG 模型。这主要是由于 IBRD 和 SPOON 模型都更加注重将兴趣相同的节点聚合成为虚拟社区，偏好和兴趣相同，必然使这些节点在资源发现获得较高的成功率。而 SIG 模型是利用了节点的移动相似性来构造虚拟社区，不完全符合“物以类聚，人以群分”的哲学观点。比如到同一个行政部门上班的公务人员，或者是在一间教室中上课的同班同学，他们在一天中有较多的时间在一起，但不表明他们有着共同的爱好和兴趣。在资源的查找过程中，因为缺少兴趣驱动导致 SIG 模型的资源发现成功率低于 IBRD 与 SPOON。对于 SPOON 而言，它总是挑选兴趣爱好相似的移动节点组成社区，却未曾考虑到节点间通信距离的影响。这使得在资源发现时，SPOON 方案总是会将查询消

息向量发送到那些与该消息向量相似度高的节点上去，而该节点往往与查询发起节点有较大的通信距离，使查询消息常常无法成功投递到目的节点，致使资源发现成功率下降。本文提出的 IBRD 模型在虚拟社区构建和社区维护方面，始终与节点的地理位置相关联，在保证消息可达条件下，充分利用节点的兴趣，从而保障资源搜索的效率。

#### 4.2.2 资源发现的平均时延

从图 8 的实验结果看，SPOON 模型的平均时延在三者中最高，其次是 SIG 方案，IBRD 模型进行成功资源发现时所消耗的时间最低。这主要是由于 SPOON 模型在组建虚拟社区时更多考虑的是节点的兴趣和偏好，而忽略节点的物理位置。此外，在资源的查找过程中，始终选择兴趣相似度高的节点进行查询，使大量的查询消息被投递到兴趣相似度高而物理距离较远的节点上，这些因素带来了较高的资源发现时延。SPOON 模型更多考虑的是提升资源共享的成功率，因此，必然会造成资源发现平均时延的增加。随着查询消息数量的增长，SPOON 模型所消耗的时延也在快速增长，便证明了这一点。若节点之间的无线通信始终是连通，那么 SPOON 模型所消耗的时延还是可以容忍的，但是移动社交网络是一种连接经常中断、时延容忍的移动网络，这使 SPOON 模型在平均时延方面的性能表现最差。SIG 方案是一种基于节点地理位置的方案。尽管这种方案在资源发现成功率上不如 SPOON 方案，图 8 的实验结果已经证明了这一点，但它总是选择临近的社交网络进行转发使它的平均时延小于 SPOON。本文提出的 IBRD 模型综合两者之长，在资源搜索过程中优先选择近距离且兴趣相似

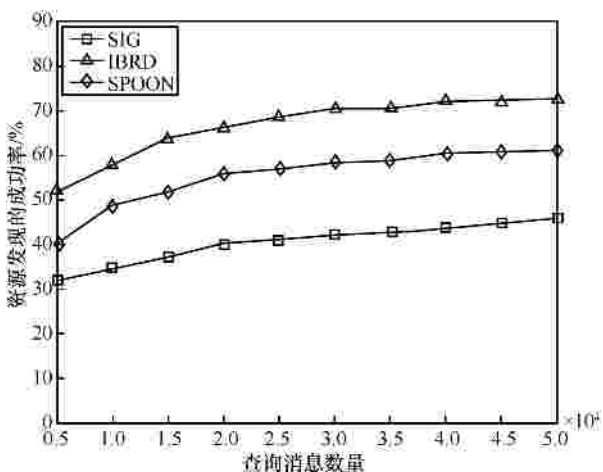


图 7 资源发现成功率

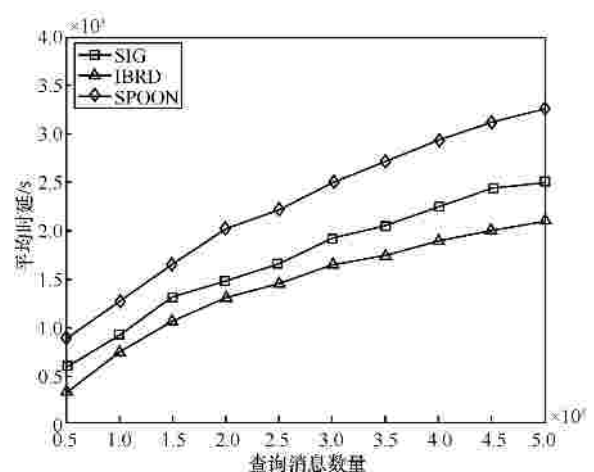


图 8 资源发现的平均时延

度高的节点进行查询,同时引入了节点的行为预测,可以动态地预测节点的运动轨迹,避免产生不必要的无限期等待,从而在保障资源发现成功率的同时,降低了消息平均时延。

#### 4.2.3 通信开销

从图 9 的实验结果看, SIG 模型产生的通信开销最高, SPOON 模型次之, IBRD 模型产生的通信开销最低。这主要是由于 SIG 模型根据节点的运动模式构建初始虚拟社区时,并未考虑到节点的爱好兴趣,简单地通过运动模式的交叉重叠把节点聚合在一起,导致节点在资源发现过程中有很强盲目性,这是致使资源发现成功率降低且通信开销增加的根本原因。对于 SPOON 模型而言,采用了基于内容的虚拟社区构建方法,使节点之间的逻辑距离与物理距离相差较大,往往导致查询发起节点与查询向量相似度高的目的节点相距较远,加之移动社交网络环境下存储—携带—转发的机会传输模式,使查询消息被转发的次数大大增加,带来更大的通信开销。本文提出的 IBRD 模型在虚拟社区的初始化构建以及社区动态维护方面都有相应的机制保障虚拟拓扑中社交用户之间的拓扑关系和物理世界中移动节点之间的拓扑关系的一致性,大大降低了拓扑失配带来的通信开销方面的影响。同时,在资源查找过程中优先选择近距离节点进行查询,也极大地降低了通信开销。

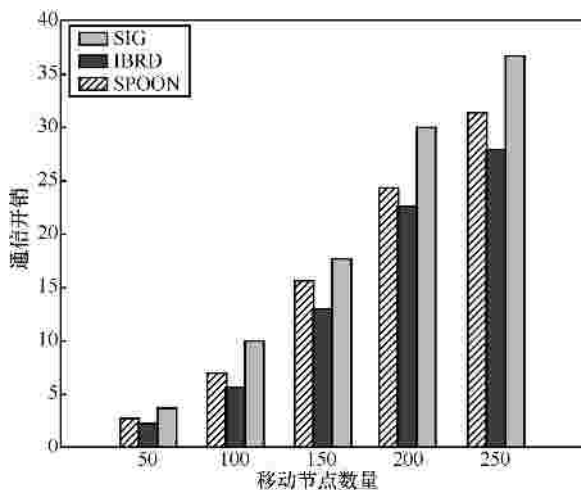


图 9 通信开销

## 5 结束语

本文在三维环境下提出一种基于兴趣和行为预测的移动社交网络动态资源发现机制 IBRD。该

模型首先采用分词统计的方法从用户的文件资源和信息表中提取兴趣向量,然后在节点间计算它们的兴趣向量的余弦相似度,之后采用局部兴趣社区构建算法构造局部社区,通过局部兴趣社区合并获得初始的全局虚拟兴趣社区。之后,采用半马尔可夫链构建预测模型以预测节点的行为和运动轨迹。依据模型的预测结果,实现虚拟兴趣社区中节点间连接关系的动态维护。然后,基于动态的虚拟兴趣社区构建高效的资源发现策略。最后在随机网络仿真器 ONE 上对 IBRD 与同类方法进行了实现,并从资源发现成功率、资源查找的平均时延以及通信开销 3 个方面对这些方案进行了性能评估。实验结果表明, IBRD 与同类资源发现方案相比,具有较高的资源发现成功率、较低的资源查找平均时延和通信开销。

然而,移动社交网络环境是开放的,网络中存在有恶意节点和不合作节点,因此移动社交网络环境是不可信的。仅靠现有的资源发现方案在该环境下进行动态资源发现时,找到可信资源节点的成功率很低,这必将影响移动社交应用的商业化进程,这部分内容将在以后的工作中进行深入的研究。

#### 参考文献:

- [1] WU J, XIAO M, HUANG L. Homingspread: community home-based multi-copy routing in mobile social networks[C]//The 32nd IEEE Conference on Computer Communications (INFOCOM). Turin, Italy, c2013: 2319-2327.
- [2] ZHU Y, XU B, SHI X, et al. A survey of social-based routing in delay tolerant networks: positive and negative social effects[J]. IEEE Communication Surveys & Tutorials, 2013, 15(1): 387-401.
- [3] 张国印, 李军. 移动对等网络覆盖网[J]. 软件学报, 2013, 24(1): 139-152.  
ZHANG G Y, LI J. Overlays in mobile P2P networks[J]. Journal of Software, 2013, 24(1): 139-152.
- [4] EAGLE N, PENTLAND A, LAZER D. Inferring social network structure using mobile phone data[J]. Proc Nat'l Academy of Sciences of USA, 2009, 106(36): 15274-15278.
- [5] A European Union funded project in situated and autonomic comm[EB/OL]. <http://www.haggleproject.org>, 2012.
- [6] SRINIVASAN, S, MOGHADAM, A, HONG S G. 7DS - node cooperation and information exchange in mostly disconnected networks[C]//IEEE International Conference on Communications. Glasgow, United Kingdom, c2007: 1-4.
- [7] SUN H X, WU C. Epidemic forwarding in mobile social

- networks[C]//The 11th IEEE International Conference on Communications (ICC 2012). Ottawa, Canada, c2012: 1421-1425.
- [8] TCHAKAROV J B, VAIDYA N H. Efficient content location in wireless ad hoc networks[C]//The IEEE International Conference on Mobile Data Management (MDM'04). Berkeley, CA, United States, c2004: 74-85.
- [9] HOH C, HWANG R. P2P file sharing system over MANET based on swarm intelligence: a cross-layer design[C]//The IEEE Wireless Communications and Networking Conference (WCNC '07). Hong Kong, China, c2007: 2674-2679.
- [10] LI Y, WU D, XU J, et al. Spatial-aware interest group queries in location-based social networks[J]. Data & Knowledge Engineering, 2014, (92): 20-38.
- [11] YOU L, LI J, WEI C, et al. MPAR: a movement pattern-aware optimal routing for social delay tolerant networks[J]. Ad Hoc Computing, 2015, (24): 228-249.
- [12] EYUPHAN B, BOLESZAW K S. Exploiting friendship relations for efficient routing in mobile social networks[J]. IEEE Transactions on Parallel and Distributed Systems, 2012, 23(12): 2254-2265.
- [13] CHEN K, SHEN H Y, ZHANG H. Leveraging social networks for P2P content-based file sharing in disconnected MANET[J]. IEEE Transactions on Mobile Computing, 2014, 13(2): 235-249.
- [14] LIU G X, SHEN H Y, WARD L. An efficient and trustworthy P2P and social network integrated file sharing system[J]. IEEE Transactions on Computers, 2015, 64(1): 54-70.
- [15] ARI K, JRG O, TEEMU K. The one simulator for DTN protocol

evaluation[C]//The 2nd Int Conf on Simulation Tools and Techniques (ICST'09). c2009: 10-35.

#### 作者简介：



李致远 (1981-), 男, 河南开封人, 江苏大学副教授、硕士生导师, 主要研究方向为移动社交网络、移动 P2P 网络和物联网的资源发现与资源共享。



陈汝龙 (1989-), 男, 山东潍坊人, 江苏大学硕士生, 主要研究方向为移动社交网络的资源发现和资源共享等。



王汝传 (1943-), 男, 安徽合肥人, 南京邮电大学教授、博士生导师, 主要研究方向为计算机软件、计算机网络和网格、信息安全、无线传感器网络、移动代理等。