

机会社交网络中基于时变兴趣社区的查询消息路由算法

毕俊蕾¹, 李致远^{2,3}

(1. 江苏大学信息化中心, 江苏 镇江 212013;
2. 江苏大学计算机科学与通信工程学院, 江苏 镇江 212013;
3. 江苏大学新一代信息技术产业研究院, 江苏 镇江 212013)

摘 要: 针对机会社交网络环境下资源查询成功率低的问题, 提出一种基于时变兴趣社区的查询消息路由算法。首先, 分析了 2 个重要的移动社交网络数据集, 发现了用户节点行为在时间和空间上的关联性和规律性; 然后, 基于用户行为在时空上的关联和规律性, 构建了时变兴趣社区; 最后, 在时变兴趣社区的基础上, 设计新的查询消息路由算法。理论分析结果表明, 所提算法时间复杂度是 $O(n \log n)$; 仿真实验结果表明, 所提算法在查询成功率、平均查询时延、查询跳数及系统开销方面均优于与同类型算法。

关键词: 机会社交网络; 查询路由; 移动社交数据分析; 时变兴趣社区

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2019177

Time-variant interest community based query message routing algorithm in opportunity social network

BI Junlei¹, LI Zhiyuan^{2,3}

1. Information Center, Jiangsu University, Zhenjiang 212013, China
2. School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China
3. New-Generation Information Technology Industry Institute, Jiangsu University, Zhenjiang 212013, China

Abstract: To solve the problem that the successful rate of resource query is low in opportunity social network, a query message routing algorithm based on time-variant interest community was proposed. Two well-known mobile social datasets were firstly analyzed, and then the temporal and spatial correlations of users were found. The time-variant interest communities based on temporal and spatial correlations of users were constructed. Next, a dynamic query routing on the constructed time-variant interest communities was proposed. Theoretical analysis shows that the proposed routing can run in $O(n \log n)$ time. The comparisons between the proposed routing and state-of-the-art message delivery algorithms show that the proposed routing can keep high query success rate, reduce the average query latency and the hop count of a query and maintain low system overhead.

Key words: opportunistic social network, query routing, mobile social data analysis, time-variant interest community

1 引言

机会社交网络 (OSN, opportunistic social networking) 是一种新型的移动社交网络^[1]。与基于移动通信网络 (如 3G、4G 等) 的在线移动社交应用

不同^[2], OSN 网络中节点是利用短距离无线通信技术 (如 Wi-Fi、蓝牙等) 进行机会式多跳通信。OSN 在应用过程中所面临的重要问题之一是查询消息路由算法的研究。查询消息路由算法的设计目标是高效地发送查询消息到目的节点 (即能提供请求响

收稿日期: 2019-01-16; 修回日期: 2019-06-28

基金项目: 江苏省自然科学基金资助项目 (No. BK20170558); 中国博士后科学基金资助项目 (No. 2015M570469)

Foundation Items: The Natural Science Foundation of Jiangsu Province (No. BK20170558), China Postdoctoral Science Foundation of China (No.2015M570469)

应的节点)。本文将现有的查询消息路由方法分为两类: 扩散式查询路由方法^[3-4]和基于社会特性查询的路由方法^[5-13]。

扩散式查询消息路由方法^[3]最早出现在连接易中断的无线网络中, 当该方法应用到 OSN 环境时, 需要每个节点携带一张消息列表。当 2 个节点相遇时, 相互交换各自的消息列表。节点对比对方的消息列表后, 相互交换自身所没有的消息数据。显然, 扩散式查询消息路由方法可以获得最优的查询成功率, 但其指数级增加的消息副本数量使该方法的可扩展性较差。基于社会特性的查询消息路由方法是利用相遇历史频度高或中心度高的节点帮助进行查询消息转发^[5]。此外, 为了提高查询算法的可扩展性, 社会化社区结构和理论也常被采用。事实上, 相同兴趣社区内部的节点较社区外部的节点有更高的相遇概率, 然而, 这种方案存在的问题在于实际应用中构造逻辑社会社区中节点间的逻辑连接关系与物理网络中节点间的物理连接关系不匹配, 这种拓扑适配现象会导致查询消息无法及时到达目的节点。

为了解决上述查询路由方法中存在的各种局限性, 本文提出了基于时变兴趣社区的查询消息路由算法 (TIMER, time-variant interest community based query message routing algorithm), 主要工作如下。

1) 对 2 个知名的移动社交网络数据集进行了分析, 发现移动节点在时间和空间上的关联性和规律性。

2) 在 1) 的基础上, 提出了时变兴趣社区模型, 并描述了时变兴趣社区的建立过程。

3) 在 2) 提出的时变兴趣社区结构上, 设计了时间复杂度为 $O(n \log n)$ 查询消息路由算法, 并对算法的性能进行了评估。

2 相关工作

本节主要对扩散式查询消息路由算法和基于社会特性的查询消息路由算法的原理进行描述, 并分析其优缺点。

1) 扩散式查询消息路由算法

文献[3]首次在部分连通的 Ad Hoc 网络中提出扩散式查询消息路由算法 Epidemic。Epidemic 算法利用节点的移动来传递消息数据, 即当节点相遇时, 它们将自身携带的消息数据传送给对方。Epidemic 算法通过消息快速复制的方式将查询消

息扩散到全网, 一定程度上保证了消息的投递成功率, 但同时也造成了消息副本拷贝的泛滥, 致使 OSN 网络资源利用率降低, 查询消息的中途丢失率增加。针对 Epidemic 算法的不足, SW (Spray&Wait) 路由算法有效地控制了网络中消息副本的数量^[4]。其原理如下。设 S 发送消息 m 给目的节点 D , 算法的输入是消息在网络中的最大副本数 L 。当 S 遇到节点 G 时, 就转发消息 m 给 G , 并将最大副本数 L 减 1。如果 G 为目的节点, 则消息投递成功, 此次路由结束; 如果 G 中继节点, 则重复上述过程, 直到消息传递到目的节点为止。Binary SW 算法是 SW 算法的改进版本, 它的原理是每次将消息传递给 G 时, 将自身的最大副本数降低为 $\left\lfloor \frac{L}{2} \right\rfloor$, 而 G 的副本数

同样设为 $\left\lfloor \frac{L}{2} \right\rfloor$, 直到将查询消息传递给目的节点为止。SW 算法的问题在于副本数 L 的设置, 副本数 L 设置较大则会存在 Epidemic 算法的问题, 若设置较小, 则会降低消息的投递成功率, 增大投递的时延。

2) 基于社会特性的查询消息路由算法

基于社会特性的查询消息路由算法思想是利用节点间历史交互信息对未来的节点行为进行预测。这类路由算法中的代表性算法包括 PRoPHET (probabilistic routing protocol using history of encounters and transitivity)^[5]、SimBet (similarity and betweenness centrality)^[6]和 ICR (interest community routing)^[7]。PRoPHET 通过节点间的历史相遇信息来估计传输概率, 它使用传输预测值 $P_{(a,b)}$ 作为节点 a 和 b 的路由度量, 表示将消息从节点 a 传递到节点 b 的成功概率。若目标节点为 d , 节点 a 与 b 相遇, 仅当 $P_{(b,d)} \geq P_{(a,d)}$ 时, 节点 a 才将消息转发给节点 b 。SimBet 路由节点的社会性度量是由节点相似性和局部介数中心性加权求和得到的。其中, 节点相似性表示当前节点与目的节点的共同邻居数, 共同邻居数越高, 节点间的相似性越大; 局部介数中心性则反映了节点在网络中重要程度, 节点介数越大作为中继转发时成功的概率越高。当节点 b 的 SimBet 值大于节点 a 的 SimBet 值时, 节点 a 将消息传递给节点 b 。ICR 研究了节点的兴趣, 给出了节点的兴趣相似度计算模型, 将兴趣相似的节点组成兴趣社区, 然后在社区内和社区间分别设计了消息路由算法。评估结果显示社区组织结构确实可以

有效地提高消息路由的性能。然而，PRoPHET、SimBet 和 ICR 理论模型和实验评估都是基于节点的随机运动展开，忽略了节点行为在时间和空间上的关联关系，模型的性能受制于节点移动的随机性，从而使节点相遇所需的等待时延仍比较高。

3 移动用户行为时空规律性和关联性分析

本文研究的机会社交网络运行的场景是在学术机构、办公场所等区域，这些场景下的科研人员及其他工作人员构建起了本文研究的机会社交网络。为了研究上述场景下节点的行为特征、分析节点的行为规律，选取了与上述物理场景相近的 2 个代表性数据集——copelabs/usense^[14] 和 upb/hyccups^[15] 进行统计分析。copelabs/usense 数据集是以人们日常活动为周期（一个周期时长 24 h），参与的移动节点数为 72 个，采集终端为三星 Galaxy S3 智能手机，数据采集和预处理软件是安装在智能手机上的应用软件 NSense，节点的组网采用无线 Ad Hoc 方式（Wi-Fi 和蓝牙），节点的活动范围是办公区域（实验室和会议讨论区）、休闲娱乐区域及家中，数据采集的时间间隔为 1 min，数据收集时间总时长为 50 h。upb/hyccups 数据集来源于布加勒斯特理工大学校园内，参与的移动节点数为 78 个，采集的终端采用 Android 智能手机（Android 系统版本号分别是 4.2 和 5.1）数据收集应用软件为基于 Wi-Fi 的 AllJoyn 框架，数据采集的时间间隔为 5 min，数据收集时长为 63 d。

图 1~图 3 分别展示了 24 h 内用户在时间和空间维度上的关联和规律性，数据来自于文献[14]。图中横坐标表示用户 ID，纵坐标表示一天时长，即 24 h。图 1 是用户在办公区域的在线时间数据统计结果。如图 1 所示，大多数节点的工作时间为 8:00~18:00，少部分节点的工作时间为晚上和下午 2 个时段，这和科研人员的作息是相关的。图 2 表示的是节点在休闲娱乐区域的在线时间统计结果，从图中不难发现，大多数节点在白天时段的休息时间为 12:00~13:00 或 18:00~22:00，其中，中午休息时间较短，通常为 1 h，少数节点选择在凌晨活动。图 3 为节点在家中出现的时间，大多数节点在家中时间为 00:00~8:00。

综上所述，对于 8:00~17:30 大多数节点出现在办公区域，如会议室和办公讨论区。对于晚上时间段，部分节点选择在办公区域，部分节点选择在家

中，部分节点选择去休闲娱乐区域，其逗留时间比较灵活，有长有短，依赖于节点的生活作息习惯。

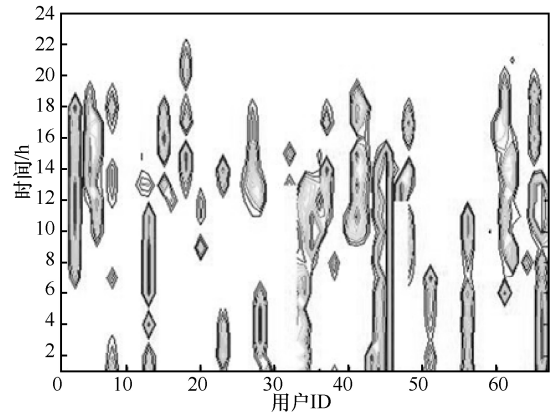


图 1 节点在办公区域的时段统计

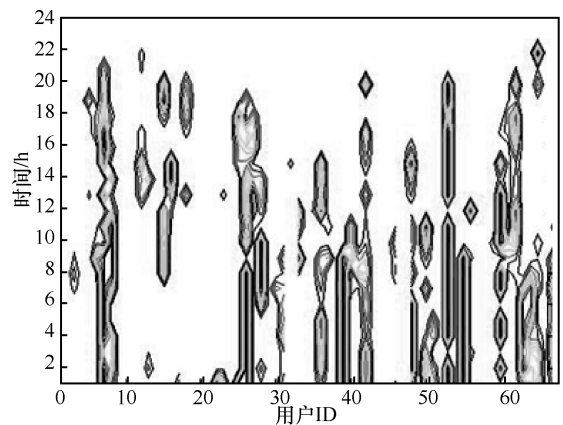


图 2 节点在休闲娱乐区的时段统计

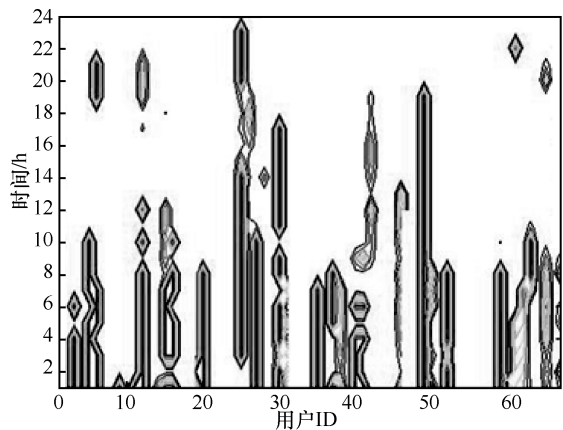


图 3 节点在家中的时段统计

图 4 和图 5 分别展示了布加勒斯特理工大学校区内节点的兴趣及其演变过程，数据来自于文献[15]。图 4 中的横坐标和纵坐标均表示用户 ID。当用户属于朋友关系且属于同一个兴趣组时，标记为 1，即图中的圆点。经过统计，发现从数据采集

开始到数据采集结束，因为兴趣相同而成为朋友关系的比例增加了 32.4%。图 5 是图 4 节点关系的演变，每一个点表示移动用户，每一条边表示节点之间的关系。有着共同兴趣的用户更倾向于加入同兴趣的社区，并成为朋友关系，之后会相互共享彼此感兴趣的信息和资源。通过兴趣聚类，可以有效地提高查询消息分发的效率。

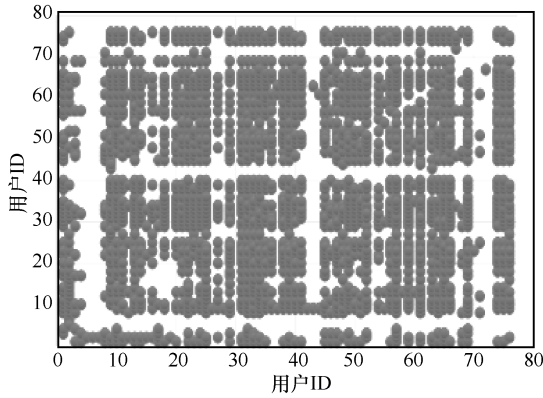


图 4 节点间兴趣关系统计结果

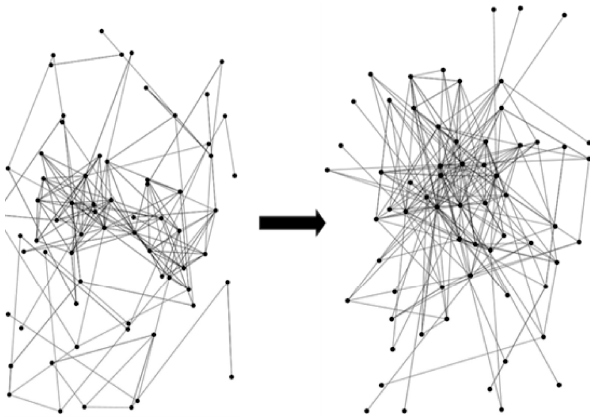


图 5 节点兴趣关系图演变

综合以上实验结果发现，移动节点的拥有者赋予了移动节点在兴趣维度上的关联关系。本文的消息路由算法研究更加关注节点的行为规律性，这对消息路由算法的中继节点选择起到了关键作用。同时表明移动节点在时间、空间和兴趣维度上具有一定的关联和规律性。这些关联和规律性便于引导本文有效地构造时变兴趣社区，并设计出高效的查询消息路由方法。

4 基于时变模型的查询消息路由算法

4.1 TIMER 路由算法概述

TIMER 路由算法在高校校园的物理场景中的应用，如图 6 所示，图中圆内的数字表示节点标号，

C_i 表示 i 类兴趣社区。其中，图 6(a)为高校校园物理场景，包含图书馆、实验室、餐厅和宿舍，具体见 4.3 节；图 6(b)为利用节点在时间维度上的规律性所构建的三类兴趣社区结构，具体见 4.2 节；图 6(c)为利用节点在时空维度上的关联性所构建的时变兴趣社区。中间的三角形区域为本文所设计的查询消息路由算法，它是基于上述兴趣社区拓扑结构的，具体见 4.4 节。

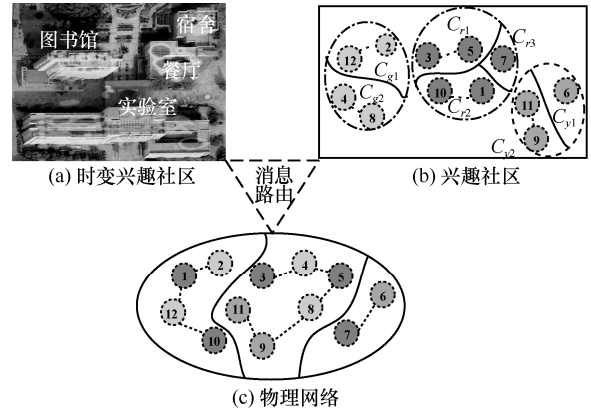


图 6 基于时变模型的查询路由算法实现原理

4.2 基于社会亲密度的兴趣社区的构建

首先给出社会亲密度的定义。节点间的社会亲密度包含了节点的连接时间、等待时间及通信频率。由于网络中部分节点有直接联系、部分节点没有直接联系，将社会亲密度分为直接社会亲密度和间接社会亲密度。

定义 1 直接社会亲密度是节点 i 与 j 之间平均间隔联系时间的倒数，其表达式如式(1)所示。它的物理意义是 2 个节点等待下一次相遇的时间越短，说明它们之间的社会亲密度越高。

$$\lambda_{ij} = \frac{1}{T_{ij}} = \frac{k_{ij}}{\int_0^T f_{ij}(t) dt} \quad (1)$$

其中， λ_{ij} 表示节点 i 与 j 之间的直接社会亲密度， T_{ij} 表示节点 i 与 j 之间的平均间隔联系时间， $f_{ij}(t)$ 表示时刻 t 之后 2 个节点下一次相遇所需等待时间的函数， k 表示节点在 $[0, T]$ 的时间范围内联系的频率。

定义 2 间接社会亲密度是节点 i 与 j 通过其他节点 w 进行联系，它们之间的平均间隔联系时间的倒数，其表达式如式(2)所示。

$$\lambda_{ij}^* = \prod_{w \in \text{relay}}^{w \neq i, j} \lambda_{iw} \lambda_{wj} = \prod_{w \in \text{relay}}^{w \neq i, j} \frac{k_{iw} k_{wj}}{\int_0^T f_{iw}(t) dt \int_0^T f_{wj}(t) dt} \quad (2)$$

其中, w 表示节点 i 与 j 之间的中继联络节点, k_{iw} 表示节点 i 与 w 之间的联系频率, k_{wj} 表示节点 w 与 j 之间的联系频率, $f_{iw}(t)$ 表示时刻 t 之后节点 i 与 w 下一次相遇所需等待时间的函数, $f_{wj}(t)$ 表示时刻 t 之后节点 w 与 j 下一次相遇所需等待时间的函数。

设网络中有 n 个节点, 基于节点的通信历史数据, 节点 n_i 根据定义 1 和定义 2 计算与其他节点的社会亲密度, 从而得到字节的矩阵 \mathbf{M} 用来存储网络中节点间的社会亲密度数值。之后, 将 \mathbf{M} 作为社会亲密度数值矩阵的输入, 使用 K -means 聚类算法^[16-17] ($K=3$) 得到图 6(b) 所示的兴趣社区。其中 C_g 表示第 g 类兴趣社区, C_r 表示第 r 类兴趣社区, C_y 表示第 y 类兴趣社区。之后, 根据图 8-1 中物理连接关系重组兴趣社区, 得到 C_{g1} 、 C_{g2} 、 C_{r1} 、 C_{r2} 、 C_{r3} 、 C_{y1} 和 C_{y2} 。

4.3 基于时变特性的动态兴趣社区构建

假设兴趣社区在 3 个时间状态之间相互转换, 这 3 个时间状态分别是初始状态、迁移状态和空状态 (Null)。具有时变特性的动态兴趣社区构建过程如下。首先定义 2 个状态矩阵 T 和 S , 如式(3)所示。

$$\mathbf{T} = \begin{pmatrix} 0 & t_{12} & \cdots & t_{1j} \\ t_{21} & 0 & \cdots & t_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ t_{i1} & t_{i2} & \cdots & 0 \end{pmatrix}, \mathbf{S} = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1(\alpha-t_{ij})} \\ S_{21} & S_{22} & \cdots & S_{2(\alpha-t_{ij})} \\ \vdots & \vdots & \ddots & \vdots \\ S_{i1} & S_{i2} & \cdots & S_{i(\alpha-t_{ij})} \end{pmatrix} \quad (3)$$

其中, \mathbf{T} 表示社区的状态转移时间矩阵, 矩阵中的元素 t_{ij} 表示社区从状态 i 到状态 j 的转移时间; \mathbf{S} 是布尔矩阵, 表示社区是否迁移成功。当 $\alpha - t_{ij} \geq 0$ 时, \mathbf{S} 中的元素 $S_{i(\alpha-t_{ij})}$ 为 1, 其物理意义为社区从状态 i 成功地迁移到了状态 j ; 否则, $S_{i(\alpha-t_{ij})}$ 为 0, 其物理意义是社区仍然滞留在状态 i 。

那么, 在 α 单位时间 (单位时间可以是小时等) 之后, 社区从状态 i 到状态 j 的转移概率如式(4)所示。

$$L_{ij}(\alpha) = P(X_\alpha = j | X_0 = i) = p_{ij} S_{i(\alpha-t_{ij})} \quad (4)$$

其中, p_{ij} 表示社区从状态 i 到状态 j 的转移概率, $S_{i(\alpha-t_{ij})}$ 的数值从 \mathbf{S} 中获得。

在 α 单位时间 (单位时间可以是小时等) 之后, 社区从状态 i 经过状态 r 到状态 j 的转移概率如式(5)所示。

$$L_{ij}(\alpha) = \prod_{r \neq i, j} \prod_{\tau=1}^{\alpha-1} L_{ir}(\tau) L_{rj}(\alpha - \tau) \quad (5)$$

综上, 可以得到社区的状态转移概率的表达式如式(6)所示。

$$L_{ij}(\alpha) = \begin{cases} p_{ij} S_{i(\alpha-t_{ij})}, & r = 0 \\ \prod_{r \neq i, j} \prod_{\tau=1}^{\alpha-1} p_{ir} p_{rj} S_{i(\tau-t_{ir})} S_{r(\alpha-\tau-t_{rj})}, & r \neq 0 \end{cases} \quad (6)$$

基于时变特性的兴趣社区迁移如图 7 所示。在 10:00 时, 在实验室构建了专业技术社区, 在图书馆和宿舍分别构建了学习社区和综合性社区, 餐厅无社区形成。在 12:00 时, 学生和教职工进入餐厅, 此时, 综合性社区在餐厅形成, 之前 2 h 所构建的兴趣社区解散后进入空状态。在 22:00 时, 学生和教职工回到宿舍休息, 综合性社区在宿舍区形成, 之前的兴趣社区解散后转入安防社区, 该社区用于保护公共财产安全。

4.4 基于动态兴趣社区的查询路由算法

基于兴趣社区的查询路由算法 TIMER 包括兴趣社区间的查询和兴趣社区内的查询。兴趣社区内的查询采用经典的 Binary SW 算法。下面重点阐述基于节点移动轨迹相似度的兴趣社区间查询算法。

定义 3 当前位置。节点 i 的当前位置用 (x_i, y_i, t_i) 表示, 其中 x_i 和 y_i 分别表示节点 i 的经纬度坐标, t_i 表示当前的时刻。

定义 4 滞留时间和滞留区域

滞留时间 T_p 等于 $(t_i - t_s)$, 其中 t_s 为节点进入某区域的时间, t_i 为节点离开某区域的时间。

滞留区域为节点在滞留时间 T_p 内的活动范围 $\max\{d(p, q)\}$ 。其中, $p(x_i, y_i, t_i)$ 为节点在 t_i ($t_s \leq t_i \leq t_i$) 时刻的位置, $q(x_j, y_j, t_j)$ 为节点在 t_j ($t_s \leq t_j \leq t_i$) 的位置, $d(p, q)$ 表示 p 、 q 两点之间的欧几里得距离。

定义 5 移动轨迹

节点的移动轨迹是由节点所经过的滞留位置所组成, 如式(7)和式(8)所示。

$$T_r = \left[A_1(T_{p(1)}, \delta_1, \eta_1), A_2(T_{p(2)}, \delta_2, \eta_2), \dots, A_n(T_{p(n)}, \delta_n, \eta_n) \right] \quad (7)$$

$$\delta_i = \frac{\eta_i}{\sum_{i=1}^n \eta_i} \quad (8)$$

其中, T_r 表示节点的移动轨迹; A_i 为节点所经过的滞留区域 i ; $T_{p(i)}$ 为节点在区域 A_i 的滞留时间; δ_i 为权重, 表示区域 A_i 对节点 i 的重要性; η_i 表示节点访问区域 A_i 的频度。

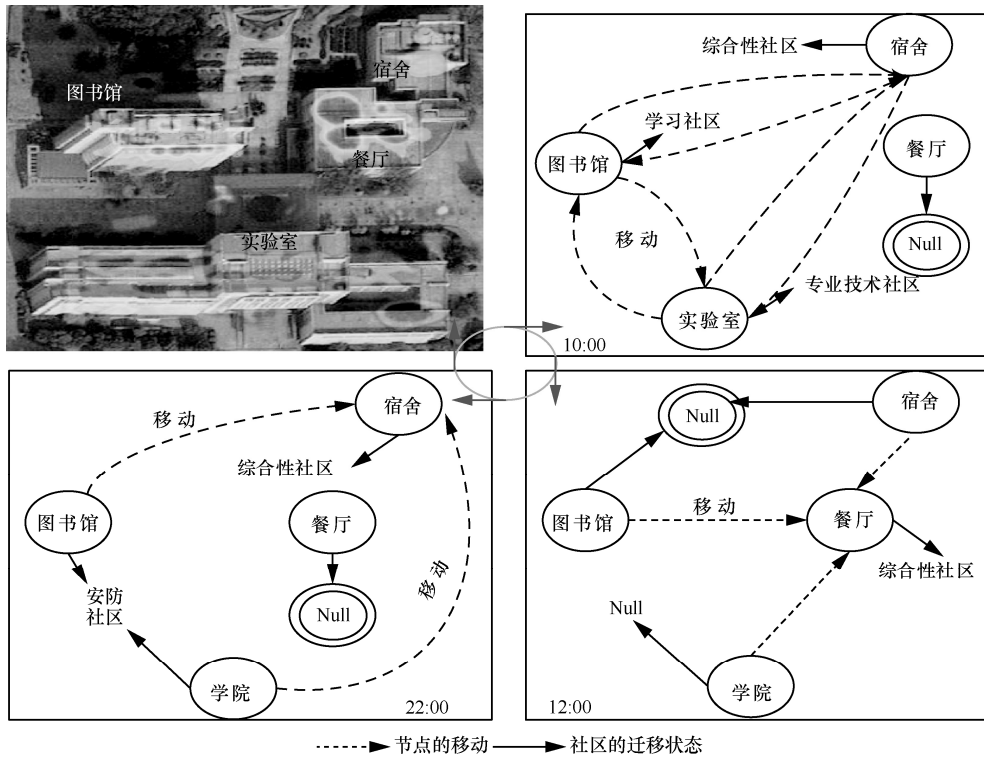


图7 基于时变特性的兴趣社区迁移示意

接下来，按照区域对节点的重要性更新节点的移动轨迹序列。2个节点移动轨迹中所包含的相同滞留区域越多，表明2个节点的兴趣度越相近，将兴趣度相近且频繁游离于不同兴趣社区的节点作为大使节点辅助消息转发。基于节点移动轨迹相似度的兴趣社区间查询算法如算法1所示。

算法1的执行如图8所示，兴趣社区1中的请求节点R在其查询消息无响应时，找到2个大使节点N₁和N₂。之后，通过比较与大使节点N₁和N₂的移动轨迹相似度后发现自身与节点N₂的移动轨迹相似度更高，于是将查询消息发送给大使节点N₂，N₂将查询消息携带到兴趣社区3中，采用Binary SW算法进行转发，最终查询消息达到了响应节点H。

算法1 基于节点移动轨迹相似度兴趣社区间查询路由算法

输入 社区中n个节点的移动区域，节点到达过的区域集合m

输出 消息响应节点集合

- 1) begin
- 2) for (i=0, i < n, i++)
- 3) for (j=0, j < m, j++) {
- 4) if 节点n_i的滞留区域与区域集合中的m_j相同{
- 5) η_j += 1

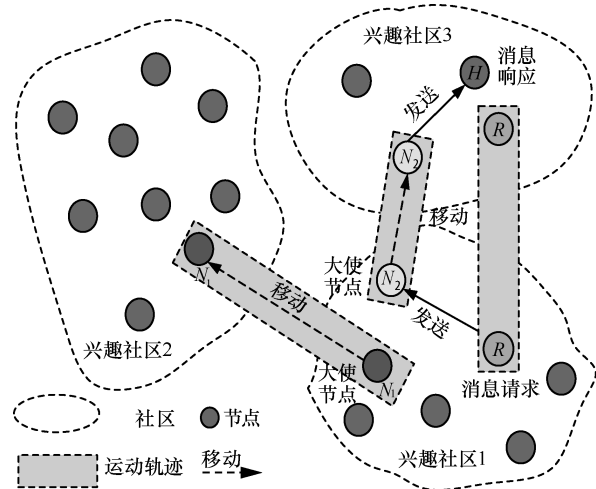


图8 基于移动轨迹相似度的查询算法

- 6) 添加 $A_j(m_j, \delta_j, \eta_j)$ 到节点 n_i 的移动轨迹中
- 7) }end if
- 8) }
- 9) end for
- 10) end for

11) 当社区内无节点响应时，统计节点所到达的区域，若节点到达区域大于区域集合中的半数，直接选为大使节点；否则通过对比节点轨迹的相似度来选择大使节点

- 12) 大使节点携带查询消息往返于各社区之间
- 13) 当消息达到其他社区后, 启用 Binary SW 算法
- 14) 返回消息响应节点集合
- 15) end begin

TIMER 时间复杂度的分析如下。假设网络场景中有 m 个位置, 每个查询节点产生 L 份查询消息拷贝。社区内部采用 Binary SW 算法, 该算法的时间复杂度为 $O(n + \log L)$, 其中 L 是常数, 因此 Binary SW 算法时间复杂度为 $O(n)$ 。算法 1 中移动轨迹生成和大使节点选择代码段的时间复杂度是 $O(mn)$; 请求节点计算最优大使节点的最坏情况下时间复杂度为 $O(n \log n)$ 。综上, TIMER 的执行时间复杂度为 $O(2mn + n \log n + n)$ 。由于 $2mn + n \log n + n \leq 2(m+1)n \log n$, 而 m 是常数, 因此, TIMER 的时间复杂度是 $O(n \log n)$ 。

5 TIMER 性能评估

5.1 仿真环境和参数设置

本文选用随机网络仿真器 ONE (opportunistic network environment) 作为算法性能评估的仿真平台。仿真场景地图为江苏大学校园一角, 利用 OPEN JUMP 采集相应的节点移动轨迹, 在仿真程序中设定了 4 个兴趣社区, 即嵌入式开发兴趣社区、电气工程兴趣社区、化学兴趣社区和理学兴趣社区。TIMER 评估的实验参数配置如表 1 所示, 仿真模拟场景如图 9 所示。

表 1 实验参数配置

参数	描述	默认值
world size	仿真实验场景面积	100 m × 200 m
EndTime/天	仿真时间	7
NrofHosts	用户节点个数	200
NrofHostGroups	分组数	4
Interface.type	通信接口类型	SimpleBroadcast
transmit speed/(Mbit·s ⁻¹)	传输带宽	2
transmit range/m	传输距离	10~20
MovementModel	组内节点移动模型	ShortestPathMap BasedMovement
mobility speed/(m·s ⁻¹)	节点的移动速度	0.5~1.5
Message.Size/KB	消息的大小	500~1024
Message.Interval/s	消息产生的间隔	1~2
Message.TTL/min	消息的生存周期	300
L	消息副本数	6



图 9 模拟仿真实验场景

5.2 实验结果分析

在实验评估中, 将本文提出的 TIMER 与同类型的查询消息路由算法 EPIDEMIC (EPIDEMIC routing for partially connected Ad Hoc network)^[3]、PRoPHET (probabilistic routing protocol using history of encounters and transitivity)^[5] 和 ICR (interest community routing for opportunistic network)^[7] 在消息投递成功率、平均查询时延、查询消息的平均跳数和查询成功的系统开销这 4 个方面进行比较, 具体定义如下。

定义 6 消息投递成功率是仿真期间成功传输的消息数占总消息数的比例。

定义 7 平均查询时延是从消息产生到消息响应的平均时延。

定义 8 成功查询的平均跳数是指从消息产生到消息响应所经过的平均节点数。

定义 9 查询成功的系统开销是指一次成功查询所需要产生的查询消息的平均副本数。

图 10~图 13 的横坐标为交互次数, 它指的是仿真期间节点所建立的连接数统计。本文将 EPIDEMIC 方案作为性能评估的基准^[5,7]。

1) 消息投递成功率

如图 10 所示, 随着交互次数的增多, EPIDEMIC、PRoPHET、ICR 和 TIMER 这 4 种算法的消息投递成功率均逐渐增加, 随后趋于稳定。这是因为消息投递成功率是建立在节点交互次数的基础上, 交互次数越多, 投递成功的概率就越高。尽管 ICR 比 PRoPHET 的性能略好, 但是与 EPIDEMIC 的性能尚有差距。这是由于 PRoPHET 为了控制系统中的消息副本数, 采用了概率转发, 因此实际的投递次数低于节点交互的次数。尽管 ICR 采用了基于兴趣社区的消息投递方法, 但是所构建的兴趣社区是静态的, 无法适应移动网络实际的变化。TIMER 提出了时变社区的概念, 时变社区较好地适应了网络拓扑的变化。此外, TIMER 也充分利用了节点的交互次数。TIMER 收

敛后的消息投递成功率在 90%左右。

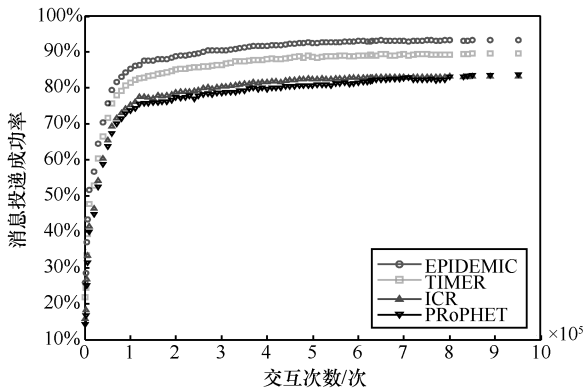


图 10 消息投递成功率评估

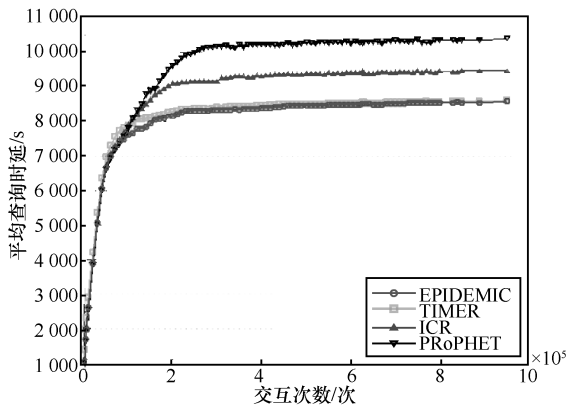


图 11 平均查询时延评估

2) 平均查询时延

平均查询时延是消息路由算法设计的一个重要衡量指标，这是因为平均查询时延越长，查询消息在网络中的副本数越多，在节点缓存中存放的时间越长。因此，降低平均查询时间对消息路由而言是十分必要的。如图 11 所示，EPIDEMIC 的平均查询时间最低，这是因为 EPIDEMIC 方法产生了多个查询消息副本，平均查询时延则是统计从消息产生到消息响应的时间，因此实际上它是将多副本查询消息中最短的查询时间作为本次的查询时延。TIMER 所构建的时变社区是有效的，查询节点总能通过一跳或多跳检索到响应节点。ICR 比 TIMER 的平均查询时延要差，但是比 PRoPHET 的平均查询时延要好，这说明 PRoPHET 通过节点的历史通信记录进行概率计算的方法尽管控制了消息副本数量的增长，但是却牺牲了节点的查询时延，与 EPIDEMIC 方法相比，多了 2 000 s 的时延。ICR 与 EPIDEMIC 相比，多了 1 000 s 的时延。TIMER 与 EPIDEMIC 相比，仅多了几十秒到百秒的时延。

3) 成功查询平均跳数

成功查询的平均跳数与平均查询时延都是评估算法性能的指标，这 2 个评估指标很相近，所不同的是查询时延表示的是数据的传输时间，而平均跳数是一次成功查询跳过的节点数。如图 12 所示，这个实验结果和上面评估的平均查询时延结果是一致的。EPIDEMIC 方法一次成功查询所需的平均跳数是 2.28，TIMER 算法一次成功查询所需的平均跳数是 2.68，ICR 算法一次成功查询所需的平均跳数是 4.25，PRoPHET 方法一次成功查询所需的平均跳数是 5.24。需要说明的是，在交互次数低于 5×10^5 时，平均跳数会增加，这种情况发生在网络中节点所处的位置比较分散，此时节点只能通过移动寻找未来与其他节点相遇的机会，从而增加投递成功的机率。

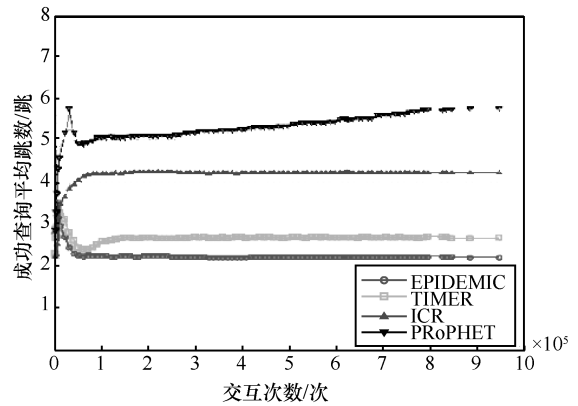


图 12 成功查询的平均跳数对比

4) 查询成功的系统开销

各种消息路由算法一次成功查询所需要产生的副本数，如图 13 所示。EPIDEMIC 方法的副本数最高，达到了平均使用 23 条副本消息才能完成一次成功查询。其他几种消息路由算法都有针对性地控制了查询消息副本的数量，基本控制在 5 条副本以下。TIMER 算法所需要的查询副本数最低，基本通过 2~3 条消息副本就可以发现资源响应节点。这就表明本文提出的 TIMER 算法在控制消息副本数增长方面效果最好。

综合对以上性能指标的实验评估，表明 TIMER 算法所构建的兴趣社区是有效的，节点基本上在兴趣社区内就可以发现请求响应节点。即使社区内不存在请求响应节点，请求消息也可通过最优的大使节点将查询消息传递给当前兴趣社区之外的请求响应节点。因此，TIMER 算法具有一定的可扩展性，适当地扩大了搜索范围，提高了查询消息转发的成功率。

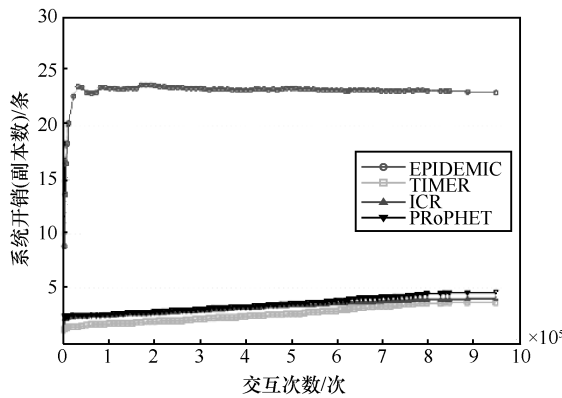


图 13 成功查询所需的系统开销评估

6 结束语

本文以高校校园为应用场景，研究校园环境下的机会社交网络，并在此基础上提出一种基于时变兴趣社区的查询消息路由算法 **TIMER**。本文对与科研工作场景相关的 2 个重要移动社交数据集(**copelabs/usense** 和 **upb/hyccups**)进行了分析，通过分析发现这些移动节点具有一定的时空关联性和规律性。基于这一发现，本文提出了基于社会亲密度的兴趣社区构建方法和时变兴趣社区模型，并在此基础上给出了社区内和社区间的查询路由算法。对算法的理论分析表明，**TIMER** 算法的时间复杂度为 $O(n \log n)$ 。在仿真实验平台上，将 **TIMER** 算法与同类算法在消息投递成功率、平均查询时延、成功查询平均跳数和查询成功的系统开销四方面进行比较，实验结果表明 **TIMER** 算法完成一次成功查询所需的查询时延、平均跳数、平均跳数低于 **PRoPHET** 和 **ICR** 查询消息路由算法，消息投递成功率和可扩展性优于 **PRoPHET** 和 **ICR** 查询消息路由算法，与基准算法 **EPIDEMIC** 在数值上接近。然而，本文提出的 **TIMER** 算法在查询消息副本数上要明显优于基准算法 **EPIDEMIC**。综上，**TIMER** 算法在网络规模和消息传播数据量增大时各方面综合表现是最优的。下一步工作是针对机会社交网络中存在自私节点情况下的消息路由算法研究。

参考文献:

[1] PIETILÄNEN A K, DIOT C. Dissemination in opportunistic social networks: the role of temporal communities[C]// The 13th ACM International Symposium on Mobile Ad Hoc Networking and Computing. ACM, 2012:165-174.

[2] 李超, 殷丽华, 耿魁,等. 面向移动社交网络内容分享的位置隐私保护方法[J]. 通信学报, 2016, 37 (11) :31-41.

LI C, YIN L H, GENG K, et al. Location privacy preservation Approach towards to content sharing on mobile online social network: survey and perspectives[J]. Journal on Communications, 2016, 37 (11):31-41.

[3] VAHDAT A, BECKER D. Epidemic routing for partially connected ad Hoc networks: technical report: CS-200006[R]. North Carolina: Duke University,2000.

[4] SPYROPOULOS T, PSOUNIS K, RAGHAVENDRA C S. Spray and wait: an efficient routing scheme for intermittently connected mobile networks[C]// The 21th ACM SIGCOMM Workshop on Delay-Tolerant Networking. ACM, 2005: 252-259.

[5] LINDGREN A, DORIA A, DAVIES E, et al. Probabilistic routing protocol for intermittently connected networks: RFC 6693[Z]. IETF Datatracker, 2012.

[6] DALY E M, HAAHR M. Social network analysis for routing in disconnected delay-tolerant manets[C]// The 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing. ACM, 2007: 32-40.

[7] LIU Q, HU C, LI Y, ZHAO W, et al, An interest community routing scheme for opportunistic networks[C]// The 54th IEEE Globecom Workshops. IEEE, 2013: 4366-4371.

[8] ZHENG Z J, WANG T Y, SONG L Y, et al. Social-aware multi-file dissemination in device-to-device overlay networks[C]// The 33rd IEEE International Conference on INFOCOM. IEEE, 2014: 219-220.

[9] LI Y, LI Q, JIN D P, et al. Contact duration aware evaluation for content dissemination delay in mobile social network [J]. Wireless Communications & Mobile Computing, 2015, 15 (3): 527-537.

[10] MANSILLA A T, ESTEVA J L. Question waves: a multicast query routing algorithm for social search[J]. Information Sciences, 2013, 253(1): 1-25.

[11] YOU L, LI J B, WEI C J. MPAR: a movement pattern-aware optimal routing for social delay tolerant networks[J]. Ad Hoc Networks, 2015, 24(3): 228-249.

[12] PAGANI E, VALERIO L, ROSSI G P. Weak social ties improve content delivery in behavior-aware opportunistic networks[J]. Ad Hoc Networks, 2015: 25(7): 314-329.

[13] LI Y, WU D, XUA J, et al, Spatial-aware interest group queries in location-based social networks[J]. Data & Knowledge Engineering, 2014, 92(1): 20-38.

[14] FIRDOSE S, LOPES L, MOREIRA W, et al. The copelabs/usense dataset: Crawdad [Z]. (2017-01-27)[2018-12-16].

[15] CIOBANU R I, DOBRE C. The upb/hyccups dataset: Crawdad [Z].(2016-10-17)[2018-12-16].

[16] 徐健锐, 詹永照. 基于 Spark 的改进 K-means 快速聚类算法[J].江苏大学学报 (自然科学版), 2018, 39 (3) :316-323.

U J R, ZHAN Y Z. Improved K-means fast clustering algorithm based on Spark[J]. Journal of Jiangsu University (Natural Science Editions), 2018, 39 (3):316-323.

[17] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: machine learning in Python[J]. Journal of machine learning research, 2011, 32(1): 2825-2830.

[作者简介]



毕俊蔷 (1981-)，女，河南南阳人，博士，江苏大学讲师，主要研究方向为移动社交网络、机会网络路由算法。

李致远 (1981-)，男，河南开封人，博士，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、物联网和软件定义网络及其安全机制。