

## 基于社会属性的 PSN 消息路由算法

曹玖新, 陈高君, 杨婧, 朱子青, 刘波

(东南大学 计算机科学与工程学院, 江苏 南京 211189)

**摘要:** 在 PSN (pocket switched network) 中实现数据传输成功率与传输延迟的有效平衡是当前有待解决的问题。提出一种基于社会属性的 PSN 路由算法——BridgingCom 算法, 该算法使用带有衰退机制的局部社区识别方法, 引入桥接中心度 (bridging centrality) 作为中继节点的选择依据, 将社会网络中节点重要性与移动社会网络的特征相结合, 考虑了节点社区关系和节点活跃度对路由算法的影响。实验表明, 该算法与现有经典算法相比能够获得较高的传输效率和较低的传输延迟。

**关键词:** PSN; DTN; 路由算法; 社区; 中心度

**中图分类号:** TP393

**文献标识码:** A

## Social-based routing in pocket switched networks

CAO Jiu-xin, CHEN Gao-jun, YANG Jing, ZHU Zi-qing, LIU Bo

(College of Computer Science and Engineering, Southeast University, Nanjing 211189, China)

**Abstract:** In pocket switched network, how to make effective balance between the message delivery ratio and transmission delay is a problem to be solved. A community-based PSN routing algorithm, named BridgingCom, is proposed. It uses local community detection method with aging mechanism and introduces the bridging centrality as the basis of forwarding decisions. The algorithm combines the analysis of node centrality with the structure of social mobile network, makes full use of the community relationship and activeness of nodes in PSN. Simulation results show that BridgingCom routing algorithm has better performance in the message delivery ratio and message average delay compared to a number of previously proposed algorithms.

**Key words:** PSN; DTN; routing algorithm; community; centrality

### 1 引言

随着手机等手持设备的大量普及, 利用手持设备实现信息传输并提供网络服务功能的口袋交换网络 (PSN, pocket switched network) 受到了广泛的关注, 它是一种由人随身携带的手持设备组成的特

殊的延迟容忍网络 (DTN, delay tolerance network) 网络, 由 2005 年剑桥大学和 Intel 研究院提出<sup>[1]</sup>。它利用人类的移动性和局域连通性, 采用“存储—携带—转发”的传输模式, 利用节点移动而进入彼此的通信范围所产生的数据交换机会进行通信, 在移动用户的设备之间传递数据。

收稿日期: 2014-02-20; 修回日期: 2014-03-24

**基金项目:** 国家自然科学基金资助项目 (61272531, 61202449, 61272054, 61370207, 61370208, 61300024, 61320106007, 61472081); 国家重点基础研究发展计划 (“973” 计划) 基金资助项目 (2010CB328104); 国家高技术研究发展计划 (“863” 计划) 基金资助项目 (2013AA013503); 高等学校博士点学科专项科研基金资助项目 (2011009213002); 江苏省科技计划基金资助项目 (SBY2014021039-10); 江苏省网络与信息安全重点实验室基金资助项目 (BM2003201); 计算机网络和信息集成教育部重点实验室 (东南大学) 基金资助项目 (93k-9)

**Foundation Items:** The National Natural Science Foundation of China(61272531, 61202449, 61272054, 61370207, 61370208, 61300024, 61320106007, 61472081); The National Basic Research Program of China (973 Program) (2010CB328104); The National High Technology Research and Development Program of China (863 Program) (2013AA013503); China Specialized Research Fund for the Doctoral Program of Higher Education(2011009213002); Jiangsu Provincial Science and Technology Plan Program (SBY2014020139-10); Jiangsu Provincial Key Laboratory of Network and Information Security(BM2003201); Key Laboratory of Computer network and Information Integration of Ministry of Education of China(93k-9)

设备随着设备携带者移动,而设备携带者的移动往往具有目的性和规律性,因而 PSN 网络具有社会网络特征。研究 PSN 网络中节点之间构成的社会网络有助于发现节点行为规律,预测这些无线通信设备构成的通信网络的变化趋势,有利于更好地发现信息携带节点、实现 PSN 网络中的信息路由。

路由问题作为 PSN 网络中一个独立开放的研究领域,其当前研究的热点是如何更好地选择中继节点以获得更高的传递成功率和更低的网络负载。目前已有越来越多的研究<sup>[2-6]</sup>开始关注通过挖掘 PSN 网络中节点的社会信息来优化路由算法。

通过社交网络分析可知,大部分人的活动范围都是局限于一定的区域或社区。社区 (community) 是一些节点所组成的子集,而它的内部节点之间的联系比社区之间的联系要紧密。为了充分利用 PSN 网络中节点的社会运动规律,本文将社会网络中节点重要性与移动社会网络的特征相结合,考虑节点社区关系和节点活跃度对路由算法的影响,提出了基于社区关系的 BridgingCom 路由算法。算法首先通过每个节点识别出自己所在社区,获得局部社区信息,作为节点关系的判断依据;然后使用桥接中心度 (bridging centrality) 作为选择中继节点的依据,使消息传递更有方向性和目的性,尽快从本地传递到目标节点。

## 2 相关工作

### 2.1 PSN 中的“存储—携带—转发”模型

PSN 网络作为一种特殊的 DTN 具有网络位置稀疏、高延迟、低数据传输率、节点间间歇性连接、消息排队时间较长和节点资源有限等特征。因此,传统移动自组织网络中的路由技术不再适用。为了能够充分利用节点移动带来的连接机会,PSN 在数据传输时采用的是“存储—携带—转发”的模式来克服链接的频繁断开。

图 1 所示为 PSN 路由过程,即一个消息的交付过程。源节点  $S$  欲发送信息至目标节点  $D$ ,然而在  $t_1$  时刻节点  $S$  与  $D$  并未发生接触且节点  $S$  没有路径可以到达  $D$ ;于是节点  $S$  将信息数据转发至邻居节点 1,节点 1 在  $t_2$  将数据转发给节点 3,最后在  $t_3$  时刻节点 3 与目标节点  $D$  相遇并将信息数据发送给  $D$ 。从上面的示例可以看出,“存储—携带—存储”模型是“存储—转发”模型的拓展,它充分利用了节点的移动能力和存储能力,为 PSN 中的信息发布提供了一个非常有效的方法。通过该转发策略,使

源节点不需与目标节点发生接触,转而通过中间节点的运载和转发,成功的将数据发送给目标节点。

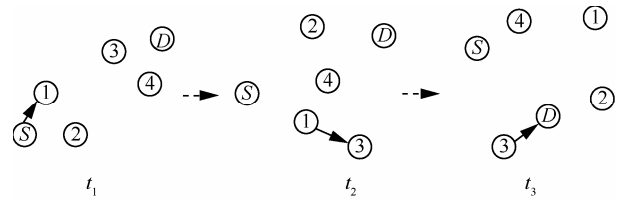


图1 “存储—携带—转发”实例

### 2.2 现有 PSN 网络中的路由算法

由于“存储—携带—转发”模式的存在,中继节点的选择以及副本数的确定显得尤为重要。与传统 Internet 以最小跳数、最短路径为路由目标不同,PSN 中不同的应用场景可能具有不同特征与不同的目标,主要有最小化传输延迟、最大化消息传递成功率以及最小化缓冲、网络带宽和能量消耗等目标。目前已有的 PSN 的路由算法从不同的角度有不同的分类方法<sup>[7]</sup>。按照消息复制策略,已有的 PSN 路由算法可以分为 3 大类:单副本路由、传染路由以及基于知识的路由。

在单副本路由中,任何时间网络中只有一个节点携带同一消息。单副本路由的开销很小,但通常交付延迟较高,可靠性相对较低。

传染路由 (epidemic routing)<sup>[8]</sup> 中每个节点都不进行路由决策,而将消息泛洪给自己所有的邻居节点。如果所有的节点都能够充分移动,传染路由能够几乎将所有的数据都正确分发,由于传染路由穷举了所有可能的传输路径,消耗资源严重,在现实中,能量、带宽、缓冲等资源缺乏而造成资源竞争,传染路由性能会严重降级。

不依赖于网络知识的路由方式中消息传输显得较盲目,为了充分利用资源,基于知识启发的路由决策可使消息传递的方向明确。根据依据的网络知识不同可以细分为基于上下文的路由、基于历史信息的路由和基于社交信息的路由。如基于传输预测的路由算法<sup>[9]</sup>利用历史连接数据计算传递概率,预测节点相遇概率,为每个节点计算到达目标节点的传输预测值,将消息副本向传输预测值大的节点传递。与传统的不考虑历史数据的路由算法相比它们具有更好的性能。

近年来,基于社交信息的 DTN 路由研究受到了广泛的关注,PSN 作为典型的具有社交属性的 DTN 网络表现出小世界<sup>[10]</sup>、有差异的节点中心度<sup>[11,12]</sup>、

有周期规律性的活动、节点分布广等特征。利用 PSN 网络中节点的内在社会关系, 可有效地解决移动社会网络中信息传输效率和准确性较低的问题。SimBet 路由算法<sup>[2]</sup>同时考虑以当前节点为中心的 Ego 网络的中心度 (Ego-centric centrality) 和社会相似性决定消息转发。消息向中心度高的节点转发, 提高成功选择有效中继节点的可能性。文献[13]基于地理位置进行网络划分, 获得节点社区关系, 但这种划分方法根据位置进行划分而不是根据节点本身划分。Bubble Rap 算法<sup>[3]</sup>选择介数中心度 (betweenness centrality) 高的节点和目标节点所在社区节点作为中继。消息到达目标社区前, 相遇节点之间比较全局中心度, 到达目标社区后, 相遇节点比较局部中心度。因此, 消息在传输过程中会涌向中心度高的节点, 这些节点的资源被迅速消耗掉。Hu 等人在 Bubble Rap 算法的基础上加入 SRL (social relation level), 提出了 BiBUBBLE 算法<sup>[14]</sup>, 在中继节点选择时除了考虑节点的介数中心度外, 也考虑节点的 SRL 值, 增加中继节点数目, 提高传输效率。CAF 算法<sup>[15]</sup>在原有路由算法的基础上加入多宿主节点 (multi-homed node) 概念, 在社区之间通过多宿主节点传递消息, 提高消息传递准确性。文献[16]发现具有相似偏好的设备携带者更有可能相遇, 因此文章提出使用  $k$  维向量表示其偏好并将消息转发给具有相似偏好的节点。文献[17]同时考虑中心度及用户偏好, 基于用户偏好信息计算中心度。2012 年 Matthew Orlinski 等人提出了基于 Epidemic 算法的 QualityEpidemic 算法<sup>[18]</sup>, 使用已有的社区识别算法将节点划分为不同社区, 节点只把消息传递给与目标节点相同社区的节点。该算法适用性比较强, 但是否能够将消息传递到目标节点依赖于节点的移动性。文献[19]基于分簇结构将移动规律相似的节点聚合为最近社交圈的节点簇, 并提出基于分簇结构的路由协议, 路由分为簇外喷射、簇间转发和簇内传染 3 个阶段。

由于 PSN 网络中节点移动模式、应用场景、路由优化目标都有所不同, 因此每个 PSN 中路由算法的侧重点都不同。在没有基础设施的情况下, 充分利用便携设备节点的数据存储、数据处理以及网络通信功能, 研究人的移动和相遇规律, PSN 可以更高效地完成数据在不同节点间传输, 实现数据的交换和信息共享。在信息服务、灾难救援、军事领域等很多方面都具有良好的应用前景。

### 3 相关概念

本章首先给出 PSN 形式化网络模型, 然后基于此模型介绍局部社区识别算法中使用到的基本概念, 最后给出节点中心度的定义和计算方法。

#### 3.1 网络模型

为了便于网络中节点关系分析, 方便地表达社会网络的统计性质, 更准确地进行社区关系分析, 这里将 PSN 网络建模为具有时间属性的无向图, 其形式化描述如下。

给定  $G_t(V, E)$ , 其中,  $V$  是网络中的顶点集合,  $|V|=n$ ;  $E$  为网络中 2 个正在通信的节点之间连接边的集合,  $|E|=m$ 。节点  $u, v$  之间的边权重用  $T(u, v)$  表示, 本文使用节点间累计通信时间作为  $T(u, v)$  的度量。

#### 3.2 熟知集合和局部社区

为了便于下文描述局部社区识别算法, 首先定义社区识别算法中的基本概念。

**定义 1** 熟知集合 (familiar set)。对于一个给定的顶点  $v_i$ , 它的完整的熟知集合形式化表示为  $F_i$

$$F_i = \{v_j | T(v_j, v_i) \geq T_{th}, v_j \in N_i\}$$

其中,  $T(v_j, v_i)$  表示  $v_j$  与  $v_i$  之间的累计连接时间,  $T_{th}$  是时间阈值,  $N_i$  表示节点  $v_i$  的邻居节点集合。节点的熟知集合保存的是当前节点的邻居集合中节点的累积连接时间超过阈值  $T_{th}$  的所有节点。

**定义 2** 局部社区 (local community)。对于一个给定的顶点  $v_i$ , 它的局部社区表示为  $C_i$ , 它包括熟知集合  $F_i$  中的所有节点和通过局部社区识别算法获取的节点。

一般社区识别算法需要通过分析整个网络结构, 将网络划分为若干个相互分离的社区<sup>[20]</sup>, 本文中使用的社区识别算法是通过节点局部知识获得节点所在的社区信息, 因此称为局部社区。由于网络动态变化和有限的网络信息知识, 可能本属于同一社区的节点会分配到不同的局部社区。

**定义 3** Ego 网络 (Ego network)<sup>[21]</sup>。对于给定的节点  $v$ , 它的 Ego 网络是指节点  $v$  本身、 $v$  的直接邻居以及它们之间的连接关系组成的子网。

#### 3.3 节点中心度定义

为了使消息能够尽快进入目标节点所在社区, 识别出可以在 2 个社区之间起到沟通桥梁作用的节点成为关键。因此, 本文使用桥接中心度作为节点中心度<sup>[11]</sup>的排名依据。

桥接中心度(bridging centrality)<sup>[22]</sup>能够很好地识别出位于不同社区之间的起到桥梁沟通作用的节点。桥接中心度  $BC$  定义如式(1)所示。

$$BC(v) = C_{\text{soc}}(v)\beta(v) \quad (1)$$

其中,  $C_{\text{soc}}(v)$  表示节点  $v$  的介数中心度<sup>[23]</sup>, 如式(2)所示。

$$C_{\text{soc}}(v) = \sum_{j=1}^N \sum_{k=1}^{j-1} \frac{g_{jk}(p_i)}{g_{jk}} \quad (2)$$

$\beta(v)$  为桥接系数(bridging coefficient), 如式(3)所示。

$$\beta = \frac{1}{d(v)} \sum_{i \in N(v)} \frac{1}{d(i)} \quad (3)$$

其中,  $g_{jk}$  表示  $j$  与  $k$  之间的最短路径数,  $g_{jk}(p_i)$  表示  $j$  与  $k$  之间的最短路径中通过  $i$  的数目,  $d(i)$  表示节点  $i$  的度(degree)值。

$C_{\text{soc}}(v)$  可以衡量网络中通过该节点的最短路径的数目, 表示节点在网络中的社会重要性;  $\beta(v)$  表示节点在信息流中的重要性, 反映了节点位于紧密连接的区域之间的程度, 衡量节点在拓扑结构中的重要性。

在 PSN 中, 网络动态变化, 节点之间没有稳定的连接, 网络拓扑  $G_t(V, E)$  随时发生着改变, 2 个相遇节点很难获得整个网络的拓扑结构。因此, 本文利用当前节点的局部社区信息, 在当前节点的 Ego 网络中计算节点的桥接中心度, 式(4)为本文使用的桥接中心度计算公式。

$$\overline{BC}(v) = \begin{cases} C_{\text{Ego}}(v)\beta(v), d(v) \neq 0 \\ 0, d(v) = 0 \end{cases} \quad (4)$$

考虑到 PSN 中很可能在某一时刻不是一个完全连通图, 网络连接图中有度为 0 的局部独立节点存在, 也就意味着不会有其他节点的消息通过此节点。为了保证在计算  $\beta(v)$  时不发生除以 0 溢出问题, 式(4)设定如果节点  $v$  的度  $d(v)$  为 0, 将  $\overline{BC}(v)$  赋值为 0。

#### 4 基于社会属性的 BridingCom 路由算法

由社交网络分析可知, 大部分人的活动范围都是局限于一定的区域或社区。如果目标节点与源节点并不在同一个活跃社区, 消息在超时之前被传递到目标节点可能需要经过其他中间社区, 成功传递的概率将会受限。如果消息的传播范围被限制在“局部”, 不仅造成消息无法成功投递, 还占用了

中继节点宝贵的缓存空间, 影响其他有可能成功送达目的地的消息接收。有效利用节点的节点历史运动轨迹和社会活动关系可以使消息传递更有方向性和目的性, 提高路由效率, 降低传输延迟。

但是, PSN 网络结构实时发生改变, 并且节点没有办法通过统一平台(如 Internet 等)获得统一服务器提供的网络全局数据, 网络中每一个节点无法获得整个网络的实时结构。目前已有的根据整个网络拓扑对网络社团结构进行的研究<sup>[20]</sup>都不适合 PSN 这种动态变化的网络, 因此本文通过局部社区识别算法获得节点的社区信息。

在局部社区识别的基础上, 本文基于节点历史运动轨迹和社会学研究理论, 提出了一种基于社会信息的 BridingCom 路由算法。该算法的基本思想是通过节点之间的社区关系为不同节点选择不同的路由方向, 防止社区内部消息在社区间的传递以及减少社区间的消息在社区内部的传递, 从而使消息传递更有方向性、降低网络负载。当消息未达到目标节点所在社区, 选择全局中心度高的节点, 增加接触到目标节点的机会; 当消息已经传递到目标节点所在社区时, 选择本地中心度高的节点, 防止消息扩散。

##### 4.1 局部社区识别算法

为了根据社区信息选择合适的中继节点, 节点能够正确、高效地识别出自己所在的社区信息是需要解决的首要问题。为了获得节点的社区信息, 本文从节点局部出发, 采用局部社区识别算法, 根据节点历史通信记录, 寻找节点所在的社区信息并实时更新。

2007 年 Pan Hui 等人提出的 Simple 局部社区识别算法<sup>[24]</sup>通过进入彼此通信范围的节点交换获得彼此的局部信息, 更新局部知识, 计算节点所在社区。在 Simple 局部社区识别算法中, 节点不断累加与邻居节点的连接时间, 随着时间推移, 每个节点存储的历史信息会逐步增加。

长时间未联系的节点信息不仅浪费存储空间, 而且还可能会对未来行为预测产生错误的指导作用。因此, 在 Simple 局部社区识别算法的基础上加入连接老化机制, 算法基本思想如算法 1 所示。

**算法 1** 带有老化机制的局部社区识别算法

输入: 本节点  $s$  的局部社区  $C_{\text{loc}}(s) = \phi$ ;

熟知集合  $F(s)$ ;

连接节点的  $u$  的局部社区  $C_{\text{loc}}(u)$ ;

熟知集合  $F(u)$ 。

输出: 节点  $s$  所在的局部社区  $C_{\text{loc}}(s)$ ;

更新后的熟知集合  $F(s)$ 。

①从源节点  $s$  的开始，初始化  $C_{loc}(s) = \{s\}$ ；

②当  $s$  与节点  $u$  相遇时，交换彼此的本地信息；

③节点  $s$  根据连接老化阈值  $T$  更新局部社区，删除超过时间  $T$  未再遇到的节点；

④若节点  $u$  不在  $s$  的熟知集合  $F(s)$  中， $s$  更新与  $u$  之间的连接时间记录，若连接时间超过设定阈值，将  $u$  并入  $F(s)$  和  $C_{loc}(s)$ ；

⑤节点  $u$  不在  $C_{loc}(s)$  时，若  $|F(u) \cap C_{loc}(s)| / |F(u)| > \lambda$ ，将  $u$  并入到  $C_{loc}(s)$ ，其中  $\lambda$  为并入阈值；

⑥若  $u$  已经通过前几步被加入到  $C_{loc}(s)$ ，则考虑是否合并 2 个局部社区：如果 2 个局部社区的重叠区域中的节点数  $|C_{loc}(s) \cap C_{loc}(u)| > \gamma |C_{loc}(s) \cup C_{loc}(u)|$ ，则合并 2 个局部社区，其中  $\gamma$  为合并阈值。

网络中每个节点需要保存的信息包括需传递的消息列表，以及本节点的熟知集合  $F$  和目前的局部社区标识  $C$  及社区内成员，以便当节点相遇时交换彼此携带的信息，更新节点的局部社区知识。

算法 1 中的连接老化机制主要通过时间阈值  $T$  控制熟知集合和局部社区中节点增删发挥作用。当节点  $s$  遇到其他节点后，首先更新节点的局部社区信息，然后根据相遇节点是否在  $s$  的熟知集合和局部社区集合内做不同的操作：如果节点已经在  $s$  的局部社区内，则根据 2 个节点的熟知集合和局部社区集合的重合度更新节点  $s$  的局部社区集合。这样不仅可以节约缓存空间，而且还可以减弱过时的历史信息对算法的影响，使算法更加准确地反映网络节点实时的传输能力。

## 4.2 BridgingCom 路由算法

在 PSN 网络中，消息是在相遇节点之间不断复制传递的，节点在网络拓扑中的重要性影响了消息传播的效率，例如，处于网络核心地带的节点更容易将消息传递出去，而在网络边缘的节点则不具备这种能力。因此，在 4.1 节中局部社区识别算法获得的局部社区信息的基础上，BridgingCom 算法根据要传递消息的目标节点与相遇节点的社区关系，选择不同的路由策略。

1) 跨社区传递：如果消息携带节点  $s$  及相遇节点都不知道目标节点  $d$  的方位（都不在目标节点所在的社区），将消息发送到桥接中心度高的节点，增加接触到目标节点所在社区的机会。

2) 社区内传递：当目标节点与相遇节点处于相

同社区，表明消息已经接近目标节点，因此将消息推送到社区内重要性高的节点，这里使用节点  $u$  的 Ego 网络中的边数作为局部中心度，使用  $LocalRank(u)$  表示。

图 2 描述了 BridgingCom 算法的基本思想，展示了具有 3 个社区的网络中 2 个消息的传递路径。设网络中的 3 个社区分别为社区 A、社区 B 与社区 C。当社区 A 中节点  $S$  有消息要传递到社区 A 中的节点  $D_1$  时，社区内节点  $S$  与节点  $D_1$  虽然无法直接通信，此时选取中心度高的节点，可以高效地将消息传递给  $D_1$ 。当社区 A 中节点  $S$  有消息要传递到社区 C 中的节点  $D_2$  时，假设最佳路径需要通过社区 B 中节点才能到达社区 C 中的节点  $D_2$ 。此时根据 BridgingCom 算法选取位于 2 个社区之间的“桥节点”，尽快将消息推送出  $S$  所在的社区 A，传递到边界节点  $N_1$ ，然后通过连接社区 A 与社区 B 的节点  $B_1$  传递到社区 B，进而获得到达社区 C 中节点  $D_2$  的路径。

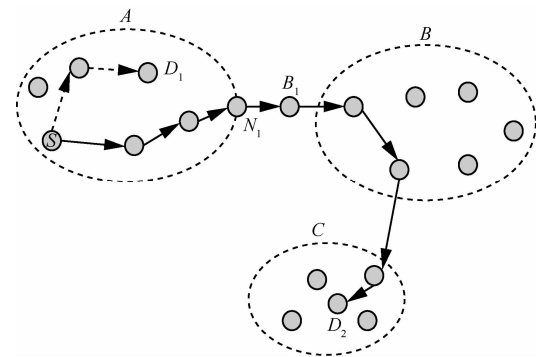


图 2 “消息传递”实例

BridgingCom 路由算法如算法 2 所示。其中，节点的社区信息通过局部社区识别算法获得。

### 算法 2 BridgingCom 路由算法

输入：本地节点  $s$ ；相遇节点  $u$ ；要传递的消息  $m$ ； $m$  的目标节点  $t$

输出：消息是否复制传递成功；

① 当节点  $s$  遇到节点  $u$ ，更新节点  $s$  的局部社区集合  $C(s)$ ；

② 当目标节点  $t$  在节点  $s$  所在社区  $t \in C(s)$ ，如果  $u \in C(s)$ ，则转到③，否则算法结束；

③ 计算节点  $s$  的局部中心度  $LocalRank(s)$  与节点  $u$  的局部中心度  $LocalRank(u)$ ，如果  $LocalRank(s) \leq LocalRank(u)$ ，则将消息复制传递给  $u$ ，否则算法结束；

④ 当目标节点  $t$  不在节点  $s$  所在社区即

$t \notin C(s)$ , 如果相遇节点  $u$  与目标节点  $t$  在同一社区  $t \in C(u)$  时, 则将消息复制传递给  $u$ , 否则转到⑤;

⑤ 如果没有传递消息, 则计算节点  $s$  的桥接中心度  $\overline{BC}(s)$  和节点  $u$  的桥接中心度  $\overline{BC}(u)$ , 如果  $\overline{BC}(u) \geq \overline{BC}(s)$ , 则同样将消息复制传递给  $u$ 。

BridgingCom 路由算法的有效运作依赖于局部社区识别算法和节点重要性度量策略。因此当 2 个节点相遇时, 首先交换彼此存储的局部网络信息, 包括节点熟知集合、局部社区集合, 节点更新自己的局部社区知识和节点中心度。

## 5 实验与分析

为了评价 BridgingCom 算法的性能, 在 The ONE<sup>[25]</sup>仿真平台上比较了本文提出的 BridgingCom 算法与 Direct 算法、Epidemic 算法、Bubble 算法<sup>[3]</sup>、BiBUBBLE 算法<sup>[14]</sup>以及 QualityEpidemic 算法<sup>[18]</sup>在不同的 TTL (time to live) 情况下的传递性能。

Direct 算法使源节点只与目标发生消息传递, 因此消息传递延迟非常大, 不进行消息复制, 网络负载最低。Epidemic 算法与 Direct 算法相反, 将消息复制传递给所有相遇节点, 能够获得较短传输延迟, 但毫无选择地盲目复制使传输能耗巨大, 严重影响网络寿命。Bubble 算法、BiBUBBLE 算法以及 QualityEpidemic 算法为近几年提出的基于节点社会关系进行路由决策的代表算法。

### 5.1 实验目标

PSN 中消息传输失败的原因可能有: 由于 PSN 间歇连通, 消息传输过程中链路中断; 消息在应用能够容忍的时间内不能到达目标节点, 被节点丢弃; 节点为了有效利用有限的缓存或节约能量而选择性的丢弃部分消息。PSN 路由的目标是以较少的资源消耗获得较高的消息交付比率(可靠性)和较低的交付延迟。主要从以下几个方面来评估相关算法的性能。

1) 消息传递成功率 (delivery ratio): 路由算法的终极目标就是获得更高的传递成功率。消息交付比率就是在整个通信过程中目标节点成功接收到数据分组数量与在源节点所传递的所有数据分组比值, 反映了传递成功率。

$$\text{消息传递成功率} = \frac{\text{成功交付到目的地的消息数}}{\text{源节点产生的原始消息总数}}$$

2) 平均延时 (average delay): 消息延时是节点在发送数据时数据块从源节点到达目标节点所需

的时间。平均消息延时体现了网络的实时性, 同时也间接反应了整个网络消息流通是否正常。另外, 尽管 PSN 中应用能够容忍一定的延迟, 但传输延迟越短, 将使得应用的时效性越强, 网络资源将越早得到释放, 网络资源的利用率也将得以提高。

$$\text{平均延时} = \frac{\text{所有成功交付消息的延迟之和}}{\text{所有成功交付的消息总数}}$$

3) 网络传输冗余比 (overhead ratio): 网络传输冗余比是指在 PSN 中副本消息与成功传递的消息之差除以成功递交的消息数目。此参数反应了网络的冗余消息量, 如果比值高, 则说明网络性能差, 网络中充斥着很多没法交付的消息; 若比值低, 说明网络有较好的传输能力, 能将较多的消息成功交付。

网络传输冗余比 =

$$\frac{\text{网络中传输的消息总数} - \text{成功交付的消息数目}}{\text{成功交付的消息数目}}$$

### 5.2 实验数据

为了验证算法在真实场景中的性能, 本文使用以下 2 个数据集, 并对数据进行了处理, 删除连接时间为 0 的无效连接信息, 防止瞬时连接影响实验效果, 数据集基本信息如表 1 所示。

**数据 1** MIT Reality Mining Project<sup>[26]</sup>数据集 (简称 Reality 数据)

Reality Mining 项目包括 75 个来自 MIT Media 实验室学生, 25 个来自 MIT Media 实验室附近的 MIT Sloan Business School 的学生。该数据集中有效节点数为 97。由于 Reality 数据集时间跨度比较长, 因此这里只选用实验数据中 4 段比较稳定而且没有额外的假期干扰的连接数据, 每段数据包含一周的连接信息。

**数据 2** Huggle Project<sup>[27]</sup>数据集 (简称 Infocom2006 数据)

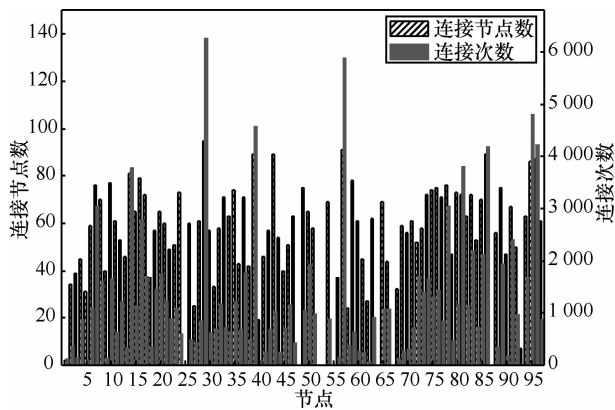
该数据集是 98 个 Infocom2006 的与会人员携带 iMotes 设备产生的记录数据。由于此组数据集集合时间跨度比较短, 因此选用整个数据集合进行实验。

**表 1 Reality 和 Infocom2006 数据集基本信息**

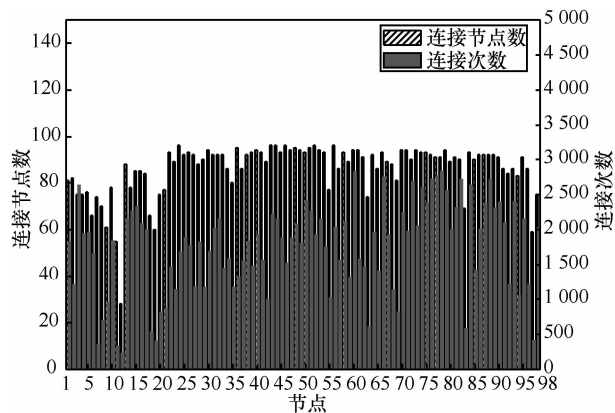
参数	Reality	Infocom2006
设备	Phone	iMote
网络类型	Bluetooth	Bluetooth
持续周期/天	246	3
连接间隔/s	300	120
实验设备数目	97	98
连接事件	54 667	191 336
平均每天每 2 个节点间连接事件数目	0.024	6.7

这 2 组数据的一些基本特征如表 1 所示。Reality 数据集和 Infocom2006 数据集的连接间隔 (granularity) 分别为 300 s 和 120 s, 连接事件 (number of internal contacts) 分别为 54 667 次和 191 336 次, 平均每天每 2 个节点之间发生的连接事件数目分别为 0.024 次和 6.7 次。可以看出, Reality 数据集的持续周期长, 用户活动范围比较广, 连接事件间隔比较长, 连接发生频率比较低; Infocom2006 数据集的持续周期短, 用户活动范围小, 记录周期短, 连接发生频率高。

图 3 为 Reality 数据集和 Infocom 2006 数据集 在实验周期中每个节点与其他节点发生连接的总次数、每个节点曾遇到过的节点总数。从图 3 可以看出节点活动具有社会规律: 节点活动范围具有局限性, 发生的大部分连接都是与某一固定节点集合中节点之间发生的; 节点活跃程度不同, 某些节点表现出明显高于其他节点的活跃度。选择合适的中继节点, 将消息尽快推送出局部社区, 将有助于增加接触到目标节点的机会。



(a) Reality 数据集



(b) Infocom 2006 数据集

图 3 节点连接次数与连接节点数统计

### 5.3 实验方法

本实验采用网络仿真工具 The ONE<sup>[25]</sup>, 此平台是专为评估 DTN 网络路由和应用协议设计的。本文通过 The ONE 提供的导入式移动模式 (external movement) 将处理过的 2 组数据导入到仿真平台。

为了能更好评估算法性能, 模拟时节点缓存大小设为 10 MB, 消息大小为 5~10 kB。采取以下 2 种消息生成机制。

1) 每隔 60~120 s 随机产生一条新消息, 消息的源节点与目标节点随机生成。消息产生位置和目标节点均匀分布在整个网络, 防止由于消息的局部性对实验结果造成影响。

2) 每隔 60~120 s 随机产生一条新消息, 消息的源节点与目标节点之间一跳可达。消息目标节点选择受到源节点的限制, 目标节点距离源节点不远, 减少了“不存在路径”的消息传递对实验结果的影响。

BridgingCom 算法能够准确运行的前提是局部社区识别算法能准确获知节点的社区信息。经过实验发现<sup>[30]</sup>, 连接时间阈值  $F_i$  确定时, 算法 2 的并入阈值  $v_j$  和合并阈值  $\bar{F}_i$  在 0.5 到 0.9 之间变化对获得的局部社区划分结果影响不大, 且在一般情况下推荐阈值设置为 0.6, 因此本文采用如表 2 所示的参数。

表 2 分布式社区识别算法基本参数

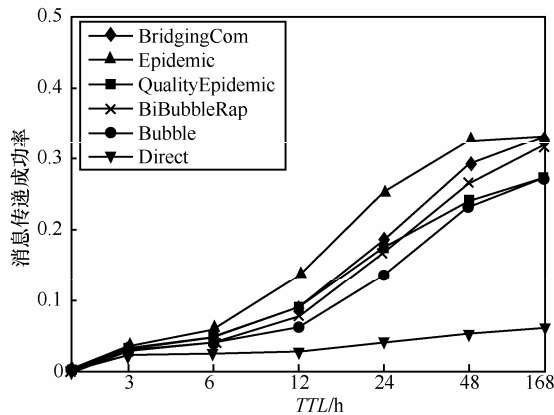
参数	Reality	Infocom2006
连接时间阈值 $F_i/s$	21 600	10 800
并入阈值 $v_j$	0.6	0.6
合并阈值 $F_i$	0.6	0.6
连接老化阈值窗口 $G_i(V, E)/s$	21 600	10 800

### 5.4 结果分析

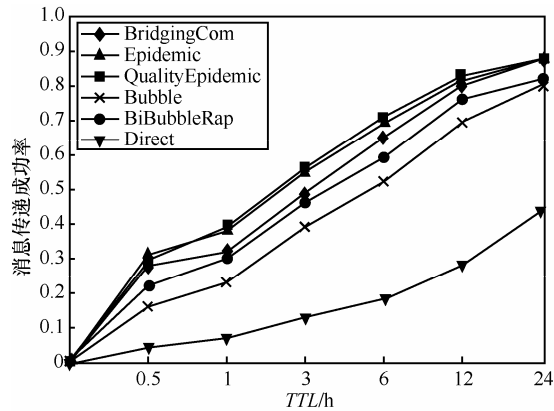
为了模拟真实场景, 使通信网络中一直有新的消息加入, 本文首先采用消息生成机制 1), 每隔 60~120 s 随机选择一个源节点产生一条新消息, 然后随机选择消息的目标节点。在 Reality 数据集和 Infocom 2006 数据集上的不同 TTL (time to live) 下获得仿真结果, 并对各个路由算法的消息传递成功率、网络传输冗余比和消息传递成功的平均延迟时间进行了详细的分析和比较。

从图 4 可以看出, 随着 TTL 增大, 消息将有更大的机会到达目标节点, 因此每种路由的传递成功率都随着 TTL 的变化而逐渐提高。与其他算法相比, 本文提出的 BridgingCom 算法在 TTL 比较小优势并不明显, 各个算法差异不大; 但当 TTL

增大,特别是超过 12 h 后,由于 BridgingCom 算法有策略的限制消息复制,其传递性能优势开始体现,表现出优于其他算法的传递成功率(除 Epidemic 算法外)。另外,当  $TTL$  继续增大时,直到  $TTL$  等于模拟周期(也就是网络中的消息会一直存在,不会因为  $TTL$  超时而被删除)时,各个路由算法的传递成功率增长开始变得缓慢,特别是 Epidemic 算法表现最为明显。



(a) Reality 数据集



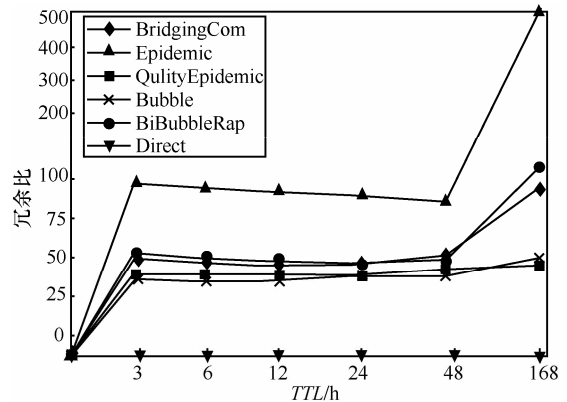
(b) Infocom 2006 数据集

图4 消息交付比率

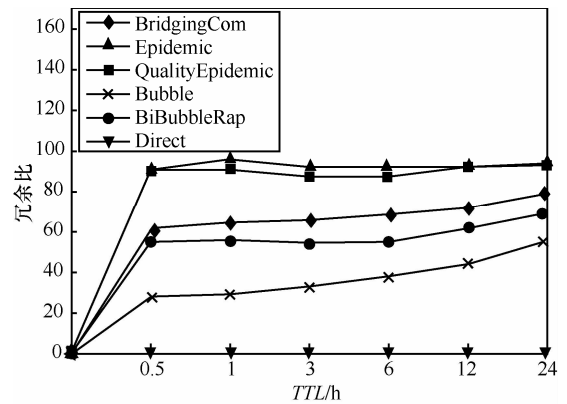
图 4 (b)中结果与图 4 (a)整体趋势一致, BridgingCom 算法获得了优于其他算法的传递成功率。不同的是,在 Infocom2006 数据集中 QualityEpidemic 算法获得了异常高的传递成功率,这是因为 Infocom2006 数据集中节点密度大(节点活动范围小,连接事件发生频繁),基于洪泛的 QualityEpidemic 算法无法有效选择中继节点,产生大量消息副本,可以从图 5(b)看出,虽然消息传递成功率提高,但是网络副本数却也同时大大增加,并没有有效的控制副本数量。

图 5 为各算法的传输冗余比结果。基于洪泛的

Epidemic 算法和 QualityEpidemic 算法在网络中副本数目最多,远远大于其他算法。BridgingCom 算法在不同  $TTL$  下,产生的冗余消息数目稳定,冗余比增长缓慢。原因是 BridgingCom 算法根据目标节点的位置,选择不同的中继节点,与 Epidemic 算法相比,能够显著减少网络开销。



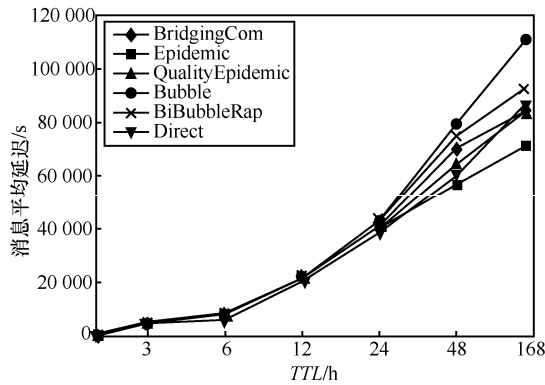
(a) Reality 数据集



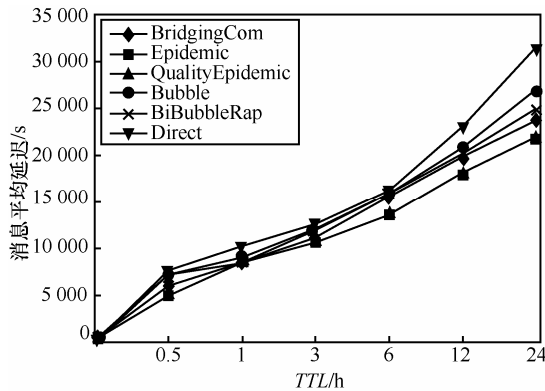
(b) Infocom 2006 数据集

图5 网络传输冗余比

从图 6 可以看到,随着  $TTL$  增大,各个算法的平均延迟时间呈现增长趋势。当  $TTL$  较小时,各个算法的消息平均延迟时间差别不大,但是当在图 6(a)中  $TTL$  增大到 24 h、图 6(b)中  $TTL$  增大到 3h,各个算法产生的消息平均延迟时间差异变大。BridgingCom 路由算法在提高消息传递成功率的同时能有效控制消息的平均传递延迟时间,特别是在网络负载比较高时, BridgingCom 路由算法有相对比较低的延迟时间,如图 6(a)中  $TTL$  大于 48 h 之后,图 6(b)中  $TTL$  大于 6 h 之后。这是由于 BridgingCom 路由算法对有不同目的地的消息选择不同路由策略,能有效控制消息的复制和传递方向。



(a) Reality 数据集合



(b) Infocom 2006 数据集合

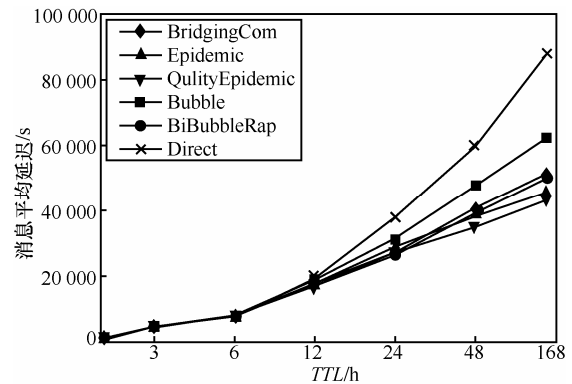
图 6 消息平均延迟时间

可以从图 6(a)观察到, Direct 算法在  $TTL$  为 6 h、12 h 和 24 h 时的平均延时并没有明显大于其他算法。一般认为, Direct 算法没有消息复制, 并且只当源节点遇到目标节点时才传递消息, 应该是获得传递机会最少的一种路由算法, 而消息传递出去机会少, 传递成功的延时应该会变大。

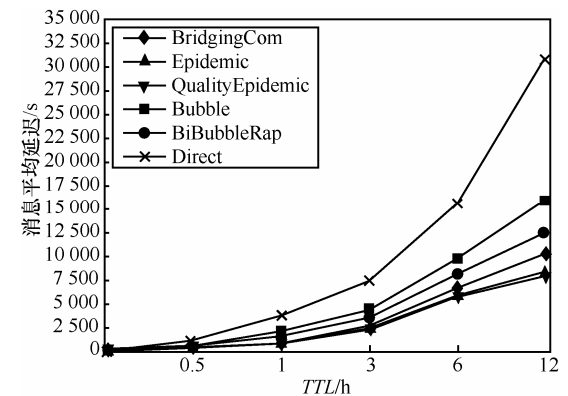
为了找到出现这个“异常结果”的原因, 本文采用消息生成机制 2) 控制消息的目标节点范围, 进行了一组新的实验, 在产生新消息时不再随机设定目标节点, 而是根据源节点不同选择不同的目标节点, 实验结果如图 7 所示。

从图 7 可以发现, 随着  $TTL$  增大, 各个算法的传递成功的消息的平均延时都逐步增大。当  $TTL$  比较小时, 新产生的消息获得的传递机会较小, 这时各个算法之间的差异体现不大, 因此各个算法的传递成功的消息的平均延时差不多。随着  $TTL$  增大, 虽然消息在网络中保存时间增长, 但 Direct 路由算法可供选择的中继节点有限, 获得的传递机会最少, 呈现出较高的消息延时。因此, 出现图 6(a)中的特殊情况的原因是其他算法传递成功的消息不

仅包含直接传递成功的消息还包含大量的、花费较长时间间接传递成功的消息, 大量需要长时间传递的消息将消息的传输延迟平均值提高。



(a) Reality 数据集合



(b) Infocom 2006 数据集合

图 7 消息平均延迟时间

另外, 通过图 4 可以看出, 当完全随机选择新产生消息的源节点和目标节点时 (消息生成机制 1)), 随着  $TTL$  变化, 各个算法在 Reality 数据集合和 Infocom2006 数据集合上的结果虽然整体变化趋势类似, 但也可以看出网络连接密度对网络传输效率的影响: 网络密度大, 节点单位时间接触到的邻居多, 消息有更多的转发机会, 也可能产生更多的消息副本, 增加网络运行时的系统开销和网络资源消耗。

## 6 结束语

本文的贡献是在 PSN 中基于节点移动性和连接时间分析节点社会关系, 引入桥接中心度 (bridging centrality) 作为中继节点的选择依据, 提出了基于社会属性的 BridgingCom 路由算法, 并且在真实数据集合上验证本算法的性能。实验证明

BridgingCom 路由算法能以较低的网络负载获得较高的传递成功率。

PSN 路由算法的研究难点在于如何在连接非持续、传输高延迟及资源受限的情况下, 选定合适的数据传输路径以完成高效的端到端传输, 并实现数据传输成功率、传输能耗以及传输延迟之间的有效平衡。PSN 作为一种具有社会规律性的 DTN, 通过挖掘节点之间的社会关系可以更加可靠地预测节点移动规律, 提高消息传递的效率。在完善现有的方法的同时, 将密切关注当前国内外对 PSN 路由提出的新理论新方法, 深入研究 PSN 中节点社会特征的理论支撑, 考虑如何动态适应网络实时变化调节社区识别的算法的阈值, 以及有效的中继节点度量等。

### 参考文献:

- [1] HUI P, CHAINTREAU A, *et al.* Pocket switched networks and human mobility in conference environments[A]. Proc of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking[C]. ACM, 2005. 244-251.
- [2] DALY E, HAAHR M. Social network analysis for routing in disconnected delay-tolerant manets[A]. Proc of ACM MobiHoc[C]. ACM, 2007. 32-40.
- [3] HUI P, CROWCROFT J, YONEKI E. Bubble rap: social-based forwarding in delay-tolerant networks[J]. Mobile Computing, IEEE Transactions, 2011, 10(11): 1576-1589.
- [4] 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 Communications Surveys & Tutorials, 2012, 15(1): 387-401.
- [5] EAGLE N, PENTLAND A. Reality mining: sensing complex social systems[J]. Personal and Ubiquitous Computing, 2006, 10(4): 255-268.
- [6] DIOT C, *et al.* Hagggle project[EB/OL]. <http://www.hagggleproject.org>, 2004.
- [7] 苏金树, 胡乔林, 赵宝康, 彭伟. 容延容断网络路由技术[J]. 软件学报, 2010, 21(1): 119-132.
- [8] SU J S, HU Q L, ZHAO B K, PENG W. Routing techniques on delay/disruption tolerant networks[J]. Journal of Software, 2010, 21(1): 119-132.
- [9] VAHDAT A, BECKER D. Epidemic routing for partially connected ad hoc networks[R]. Technical Report, 2000.
- [10] ANDERS L, AVRI D, OLOV S. Probabilistic routing in intermittently connected networks[J]. ACM SIGMOBILE Mobile Computing and Communications Review, 2003, 7(3): 19-20.
- [11] MILGRAM S. The small world problem[J]. Psychology Today, 1967, 2: 60-67.
- [12] FREEMAN L C. Centrality in social networks conceptual clarification[J]. Social networks, 1979, 1(3): 215-239.
- [13] DALY E M, HAAHR M. Social network analysis for information flow in disconnected delay-tolerant MANETs[J]. IEEE Transactions on Mobile Computing, 2009, 8(5): 606-621.
- [14] SARAFIJANOVIC-DJUKIC M P N, GROSSGLAUSER M. Island hopping: efficient mobility-assisted forwarding in partitioned networks[A]. Proc of Sensor and Ad Hoc Communications and Networks[C]. IEEE, 2006. 226-235.
- [15] HU T, HONG F, ZHANG X Q. BiBUBBLE: social-based forwarding in pocket switched networks[A]. Proc of UIC/ATC[C]. IEEE, 2010. 195-199.
- [16] MTIBAA A, HARRAS K A. Social forwarding in large scale networks: insights based on real trace analysis[A]. Proc of Computer Communications and Networks[C]. IEEE, 2011. 1-8.
- [17] MEI A, MORABITO G, SANTI P, *et al.* Social-aware stateless forwarding in pocket switched networks[A]. Proc of IEEE INFOCOM[C]. IEEE, 2011. 251-255.
- [18] GAO W, CAO G. User-centric data dissemination in disruption tolerant networks[A]. Proc of IEEE INFOCOM[C]. IEEE, 2011. 3119-3127.
- [19] ORLINSKI M, FILER N. Quality distributed community formation for data delivery in pocket switched networks[A]. Proc of SIMPLEX[C]. ACM, 2012. 31-36.
- [20] LI Z, LI Q M, ZHANG H, *et al.* Closely social circuit based routing in social delay tolerant networks[J]. Journal of Computer Research and Development, 2012, 49(6): 1185-1195.
- [21] 程学旗, 沈华伟. 复杂网络的社区结构[J]. 复杂系统与复杂性科学, 2011(8): 57-70.
- [22] CHENG X Q, SHEN H W. Community structure of complex networks[J]. Complex Systems and Complexity Science, 2011(8): 57-70.
- [23] MARSDEN P V. Egocentric and sociocentric measures of network centrality[J]. Social Networks, 2002(24): 407-422.
- [24] HWANG W, CHO Y, ZHANG A. *et al.* Bridging centrality: identifying bridging nodes in scale-free networks[A]. Proc of ACM SIGKDD[C]. ACM, 2006. 20-23.
- [25] FREEMAN L C. Centrality in social networks conceptual clarification[J]. Social networks, 1979, 1(3): 215-239.
- [26] PAN H, EIKO Y, SHU Y C, JON C. Distributed community detection in delay tolerant networks[A]. Proc of 2nd ACM/IEEE International Workshop on Mobility in the Evolving Internet Architecture[C]. 2007. 27-30.
- [27] ONE: Opportunistic Network Environment[EB/OL]. <http://www.netlab.tkk.fi/tutkimus/dtn/theone/>. 2009.
- [28] EAGLE N, PENTLAND A. Reality mining: sensing complex social systems[J]. Personal and Ubiquitous Computing, 2006, 10(4): 255-268.
- [29] DIOT C, *et al.* Hagggle project[EB/OL]. <http://www.hagggleproject.org>, 2004.

### 作者简介:



曹玖新 (1967-), 男, 河南商丘人, 东南大学教授、博士生导师, 主要研究方向为服务计算、网络安全、社会计算。

陈高君 (1988-), 男, 河南漯河人, 东南大学硕士生, 主要研究方向为社会计算。

杨婧 (1988-), 女, 河南三门峡人, 东南大学硕士生, 主要研究方向为服务计算、社会计算。

朱子青 (1990-), 男, 江苏徐州人, 东南大学硕士生, 主要研究方向为社会计算。

刘波 (1975-), 女, 河南南阳人, 东南大学副教授, 主要研究方向为普适计算、社会计算。