

基于朋友关系变迁演化的社区划分算法*

杨荣刚¹, 张书奎^{1,2†}, 李克清³, 林政宽^{1,2}

(1. 苏州大学 计算机科学与技术学院, 江苏 苏州 215006; 2. 江苏省无线传感网高技术研究重点实验室, 南京 210003; 3. 苏州市职业大学 计算机工程学院, 江苏 苏州 215000)

摘要: 为了获得一种能够适用于机会网络的社区划分算法, 提出了一种基于朋友关系变迁演化的社区划分算法。首先, 定义了描述节点间朋友关系强弱的朋友关系度量标准; 然后, 利用朋友关系度量标准和节点自身的相遇历史信息, 将网络中的节点划分到不同的朋友社区中。在两种不同数据集下的实验结果表明, 利用朋友关系进行社区划分可有效降低消息在传输过程中的路由开销, 同时保持较高的传输成功率。

关键词: 机会网络; 朋友关系; 社区划分; 数据集

中图分类号: TP301.6 **文献标志码:** A **文章编号:** 1001-3695(2018)05-1337-05

doi: 10.3969/j.issn.1001-3695.2018.05.012

Community division algorithm based on evolution of friendships

Yang Ronggang¹, Zhang Shukui^{1,2†}, Li Keqing³, Lin Zhengkuan^{1,2}

(1. School of Computer Science & Technology, Soochow University, Suzhou Jiangsu 215006, China; 2. Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing 210003, China; 3. School of Computer Engineering, Suzhou Vocational University, Suzhou Jiangsu 215000, China)

Abstract: In order to obtain a community division algorithm that can be suitable for opportunistic networks, this paper proposed a community division algorithm based on the evolution of friendships. First, it defined the friendship metric that described the strength of the relationships between nodes. Then, utilizing this metric and the encounter history of nodes, it divided the nodes in the networks into different friendship communities. Experiments were conducted under two different sets of data, and the experimental results show that routing overhead reduces effectively and the transmission rate still keeps a high value by using friendships to divide communities.

Key words: opportunistic networks; friendship; community division; data set

0 引言

机会网络 (opportunity network)^[1] 源于容忍延迟网络^[2] (delay-tolerant network, DTN) 和移动自组织网 (mobile Ad hoc network, MANET), 可以视为是两者的子类。机会网络的主要特点是不需要源节点与目的节点之间存在一条完整路径, 而是利用节点移动带来的相遇机会进行网络通信。在实际应用中, 组成机会网络的各种设备节点大多是由人携带的, 因此节点具有人的社会属性; 同时机会网络也具有社会网络的一些特点, 如小世界效应^[3]、无标度特性^[4] 以及社区结构特性^[5] 等。

社会网络是指社会个体成员之间因为互动而形成的相对稳定的关系体系。在社会网络中, 具有相似社会行为的节点 (人) 会形成各种社会结构, 即社区结构, 它代表了各种社会关系, 如朋友、同事、同学、生意伙伴、种族信仰关系等。再通过这些社会关系, 将网络中的节点 (人) 串连起来。

社区结构具有相对稳定性, 在同一社区内节点之间接触频繁, 而在不同社区内节点之间很少接触。属于一个社区的节

点, 也可能属于另一个社区, 即一个节点可以属于多个社区。例如, 在人际关系中, 每个人都可能属于家庭、工作单位、朋友圈子等。通过对社区的研究, 能够有助于分析和挖掘社区之间的动态关系、模式和功能, 找到隐藏的社会关系和社会行为。因此, 对社区的研究对社区的划分具有重要的理论意义和实际的应用前景。现有的社区划分算法大多是基于社会网络的分析和研究, 其主要思路是将社会网络抽象成一个图, 图中的节点表示人, 而边 (或链接) 表示人与人之间的关系, 再通过物理学、数学和计算机科学等领域的理论和技术, 将图划分为不同的社区。这使得现有的传统社区划分算法无法直接应用于机会网络中, 其主要原因如下:

a) 现有传统社区划分算法, 只能在获得了网络全局信息的前提下对全局信息进行处理。而在机会网络中, 由于没有基础的通信设施, 且节点的无线通信能力有限, 无法获得网络的全局信息。

b) 现有传统社区划分算法, 只能对静态数据进行处理, 无法对动态数据进行处理。而机会网络是利用节点移动带来的

收稿日期: 2016-12-30; **修回日期:** 2017-02-20 **基金项目:** 国家自然科学基金资助项目 (61672370); 苏州市前瞻性应用研究计划项目 (SYG201730); 江苏省“六大人才高峰”高层次人才选拔培养资助项目 (2014-WLW-010); 苏州市融合通信重点实验室资助项目 (SKLCC2013XX)

作者简介: 杨荣刚, 男, 江苏人, 硕士, 主要研究方向为无线传感器网络; 张书奎, 男 (通信作者), 内蒙古人, 教授, 博士, 主要研究方向为物联网、无线通信网络、移动计算、网络安全、分布式计算、智能信息处理 (zhangsk@suda.edu.cn); 李克清, 男, 湖北荆门人, 教授, 博士, 主要研究方向为无线传感器网络、信息安全、优化设计; 林政宽, 男, 台湾人, 副教授, 博士, 主要研究方向为并行计算系统、交换网络、无线网络、无线传感器网络、算法设计与分析。

相遇机会进行通信,节点的频繁移动,会使得网络的拓扑结构不断变化。

c) 现有传统社区划分算法,其计算量和时间开销比较大。而机会网络中,由人携带的节点设备其能量、存储资源和计算资源是极其有限的。

本文利用机会网络中节点(人)的社会属性,提出一种基于朋友关系的朋友社区划分算法。为了知道网络中节点之间是否具有朋友关系,本文定义了节点之间是否具有朋友关系的度量标准。再通过该度量标准计算网络中节点间的朋友关系强度,将网络划分成不同的朋友社区。考虑到在不同时间段,节点所属朋友社区会发生变化,本文利用节点在不同时间段的相遇历史记录信息,将节点划分到不同的社区中。

1 相关工作

最早的社区划分算法是由 Girvan 等人^[5]提出的 GN 算法。该算法的原理是将整个网络抽象成图,先找出社区之间的所有链接,然后将它们全部删除,最后每个连通子图即为一个社区。针对 GN 算法计算速度慢、计算开销过大和时间复杂性高等缺点,研究者们提出了一些改进方法。Tyler 等人^[6]将统计方法引入 GN 算法,提出了一种近似 GN 算法。Radicchi 等人^[7]利用链接聚类系数对 GN 算法进行了优化。随着应用领域的不同,网络的类型和复杂度也不同。为了适应各种不同的网络,Newman 等人^[8]又提出了用于刻画网络社区结构优劣的量化标准——模块性函数 Q 。在 2004 年,Newman^[9]提出了第一个局域模块性优化的社区划分算法(fast Newman, FN)。基于这一思想,Guimerà 等人^[10]提出了一种基于模拟退火的模块度优化算法,Blondel 等人^[11]提出了快速模块度优化算法等。以上这些算法都是将网络划分为若干独立的社区,并且每个节点只隶属于一个社区。而在实际应用中,网络中的社区并不是独立的,一个节点可以隶属于多个社区。例如在实际的人际关系中,一个人可以属于家庭、学校、工作以及朋友圈等不同的社会团体。因此,研究者们提出重叠社区的概念。目前具有代表性的重叠社区划分算法有基于团渗理论的算法、基于链接划分的算法和基于局部扩展的算法等。

在重叠社区的划分算法中,最具代表性的算法是由 Palla 等人^[12]提出的团渗算法(clique percolation method, CPM)。该算法的主要思想是将社区看做是由一些团(全连通子图)构成的集合,这些团之间通过共享节点而紧密连接。在 2009 年,Evans^[13]对网络中的链接进行了划分,以产生节点的重叠社区结构,即通过“用边表示节点,用节点形成边”的方法,将网络转换为线图,然后选择已有的节点划分算法获取链接社区结构。Lancichinetti 等人^[14]最早提出了基于局部扩展的重叠社区划分算法(local fitness measure, LFM),其主要方法是先随机选择一个节点作为出发节点,再不断向外扩张划分社区,直到社区划分达到局部最优为止。

以上重叠社区的划分仍然需要在获得全局的网络拓扑结构的前提下,将网络转换成图,再将图划分成若干连通子图。虽然重叠社区划分算法能够将节点划分到不同的社区,但社区一旦划分完毕就不再变化,无法体现出社区的演变过程。

在机会网络中,经典的具有代表性的路由算法有 Epidemic^[15]、Spray and Wait^[16]和 Prophet^[17]。在 Epidemic 算法中,每个节点都维护一个消息队列,当任何两个节点相遇时,节点

间彼此交换各自没有的消息,因而它们拥有相同的消息队列。其主要优点是通过泛洪的方式来增加网络中消息副本的数量,从而最大化消息传输的成功率,减少传输延迟;缺点是网络中存在大量的消息副本,会消耗太多的网络资源。

为了减少网络中消息副本的数量,研究者在路由算法中采取了一些限制消息副本的策略,如 Spray and Wait 和 Prophet 算法等。Spray and Wait 算法将消息的路由分为 Spray 和 Wait 两个阶段。Spray 阶段源节点向网络中扩散固定数量的消息副本,Wait 阶段节点使用 Direct Delivery^[18]算法将消息传递到目标节点。该算法在限制消息副本数量的同时仍能够获得较高的传输成功率。Prophet 算法是一种基于单拷贝方式的路由算法。通过对现实中自组织网络的观察和网络中节点移动行为的分析,作者认为节点的移动行为模型具有重复性(如果一个节点曾多次到过某个地方,则它很有可能再次去往这个地方),并且在这种移动行为模型下,节点的移动行为是可以预测的。根据分析,Lindgren 等人^[17]提出了一种基于预测的消息转发算法。该算法定义了一个传输预测值,用来描述节点成功传输消息的概率,当两个节点相遇时,节点更新彼此的传输预测值,并通过该传输预测值进行消息的转发决策,使消息沿着传输预测值高的方向传输。

在经典的路由算法中,大多数算法都假设节点的移动行为比较简单(大多使用随机移动模型 Random Way Point^[19]),而在真实场景下,设备节点大多是由人携带,由这些设备节点组成的自组织网络具有明显的社会网络特性。同时,为了能够更适应实际应用需求,Hui 等人^[20]提出一种基于社区的路由算法 Bubble Rap。在该算法中,每个节点维护两种排序表,即全局排序表(global ranking)和社区内排序表(local ranking)。前者表示节点在整个网络中的活跃度排名,后者表示节点在所属社区内活跃度的排名。在进行消息的转发决策时,先将消息转发给全局排序表中活跃度高的节点,直到遇到与目标节点在同一社区的其他节点;然后再将消息转发给社区内排序表中活跃度高的节点,直到遇到目标节点,完成消息的传输。该算法的优点是利用了机会网络具有社区结构的特性,使消息尽量在社区内部进行传输,可在有效限制网络中消息副本的前提下,保持较高的传输成功率。该算法存在的缺陷是需要提前获得网络的拓扑结构,再对网络进行社区划分,且社区一旦划分就固定不变,网络中的节点始终属于一个特定的社区,这与实际情况是不相符的。受社会网络中社区划分算法和现有路由方式的启发,结合机会网络的特点,本文提出了一种新的社区划分算法。首先定义了一个新的度量指标,用来表示节点间朋友关系的强弱;然后对社区的演变过程进行分析;最后给出朋友社区划分的算法详细设计方法。

2 算法设计与实现

2.1 朋友关系的度量标准

在文献[21]中 Chen 等人通过观察发现,节点间的朋友关系在不同时间段内变化很小,具有一定的稳定性。因此,在一天中的不同时间段内,朋友节点间的聚集会形成朋友社区,在朋友社区内朋友节点之间会频繁相遇。例如在现实生活中,通常认为相互认识并且频繁联系的人很可能是朋友。在机会网络中,由人携带的节点具有人的社会属性。因此,可以认为相遇频繁、接触时间长且相遇具有规律性的节点之间同样具有朋

友关系。在机会网络中,当节点之间需要进行通信时,如果利用网络中的朋友节点进行消息的传递,将会大大提高消息转发的成功率,减少网络中消息副本的数量,降低消息在网络中的传输开销。因此,当节点通过移动带来通信机会时,需要对节点之间能够成功传递消息的概率进行评估。

在以往的研究中,一些度量标准被用来描述节点之间成功传递消息的概率,如节点间的相遇频率、总的接触时间、平均接触时间和平均相遇间隔时间等。然而这些度量标准在描述节点成功传递消息的概率上都不够准确。如图1所示,节点*i*和*j*在不同时间段内的六次相遇历史记录,阴影部分表示在时间间隔*T*内节点间相遇的持续时间。比较图1(a)和(b),节点间的相遇频率相同,但在(b)中节点间的接触持续时间更长,因此(b)比(a)具有更好的转发机会;比较(b)与(c),节点间总的接触持续时间相同,但是相遇频率不同,而节点的频繁相遇能够使节点具有更多的消息转发机会,因此(c)比(b)具有更好的消息转发机会。

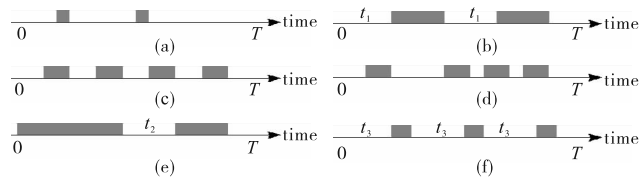


图1 节点*i*与*j*在不同时间段 $[0, T]$ 内的六次相遇记录

另外,在上述度量标准中,如图1(a)与(b),节点间相遇频率的高低不能够表示成功传递消息的概率大小;如(b)与(c),节点间总的接触时间长短不能够表示成功传递消息的概率大小。虽然在(a)(b)和(c)中,节点间的平均相遇间隔时间能够表示节点间成功传递消息的概率大小,但是在其他情况下则不能。比较(c)与(d),节点间的总接触时间、相遇频率和平均相遇间隔时间相同,但是由于(c)的相遇是均匀分布的,所以(c)比(d)具有更好的转发机会。此外,对于(e)与(f),节点间的平均相遇间隔时间不能准确表示成功传递消息的概率大小。如果 $t_1 = t_2$,平均相遇间隔时间就不能区分(b)和(e),由于(e)中节点间具有更长的接触时间,所以(e)有更好的转发机会(如果 $t_1 < t_2$,由于(b)的平均相遇间隔时间少,所以(b)具有更好的转发机会)。同理,如果 $t_1 = t_3$,平均相遇间隔时间就不能区分(b)和(f),即使(b)会提供更好的转发机会。

综合以上分析,结合机会网络中节点的社会特性,为了获得一个能更准确地描述节点之间成功传递消息的概率(即消息的转发机会),本文综合考虑了朋友关系的三个特征:相遇频率高、接触时间长和相遇具有规律性。换句话说,即被认为具有朋友关系的两个节点,它们会频繁相遇,接触时间会比较长,且它们的接触具有一定的规律性。这里的规律性表示节点间的相遇在时间上具有一定的规律性。本文将综合考虑以上三个特征,提出一种朋友关系度量标准 CFM (close friendship metric):

$$CFM(x) = - \sum_{i=1}^n P(x_i) \log_2 P(x_i) \quad (1)$$

其中: $P(x_i)$ 表示节点间的相遇持续时间和当前所在时间段总时长的比值, $0 \leq P(x_i) \leq 1$, $P(x_i) = \frac{f(x)}{T}$, $f(x)$ 表示两个节点一次相遇的持续时间。它有以下几个重要的性质:

性质1 非负性。由于 $0 \leq P(x_i) \leq 1$,所以 $-P(x_i) \log_2 P(x_i) > 0$,即 $CFM(x) = - \sum_{i=1}^n P(x_i) \log_2 P(x_i) > 0$ 。

性质2 对称性。 $CFM(x)$ 为每一项 $-P(x_i) \log_2 P(x_i)$ 的求和,当 $P(x_i)$ 的顺序变化时, $CFM(x)$ 的求和顺序也发生变化,但是并不会影响求和的结果。

$$\begin{aligned} CFM(x) &= CFM(P(x_1), P(x_2), \dots, P(x_{n-1}), P(x_n)) = \\ &= CFM(P(x_2), \dots, P(x_{n-1}), P(x_n), P(x_1)) = \dots = \\ &= CFM(P(x_n), P(x_1), P(x_2), \dots, P(x_{n-1})) \end{aligned} \quad (2)$$

性质3 递增性。当任意一个 $P(x_i)$ 划分成 m 个值 $q_1, q_2, q_3, \dots, q_{m-1}, q_m$,且 $\sum_{i=1}^m q_i = P(x_i)$,则 $CFM(x)$ 随值的分解而增加。

$$\begin{aligned} CFM(P_1, P_2, P_3, \dots, P_{n-1}, q_1, q_2, q_3, \dots, q_m) &= \\ CFM(P_1, P_2, P_3, \dots, P_{n-1}, P_n) + P_n CFM\left(\frac{q_1}{P_n}, \frac{q_2}{P_n}, \dots, \frac{q_m}{P_n}\right) \end{aligned} \quad (3)$$

通过图1可再次验证该度量标准的有效性,对图1中的所有情况,该度量标准都能够准确地描述在哪种情况下能够提供更好的消息转发机会。需要说明的是,对于图1(c)与(d),通过计算, CFM 值相等,即(c)与(d)对消息的转发机会相等,与前面讨论不同。但在一段时间内,节点相遇频率和每次相遇持续时间都相同的情况下,对于计算节点之间朋友关系强度几乎没有影响。

2.2 朋友社区的演变

在现实生活中,早晨在家时,属于家庭社区;到公司后,属于工作社区;下班后参加朋友聚会时,属于朋友社区。在一天中的不同时间段,会属于不同的社区。在不同社区里,遇到的人基本是固定不变的,与他们会频繁相遇,并且这种社区的演化具有一定的规律性,每天重复发生。

通过对 MIT 真实数据集^[22]的分析,给出了一天中节点85和28与其他节点的相遇分布图,如图2所示。从图2中可以看出,节点间的彼此相遇发生在一天当中的特定时间段内。例如节点28与38经常在下2点到下午10点的时间段内相遇,而它与节点15经常在下7点到下午10点的时间段内相遇;又如节点85和其他节点的相遇几乎都集中在上午12点到下午9点的时间段内。

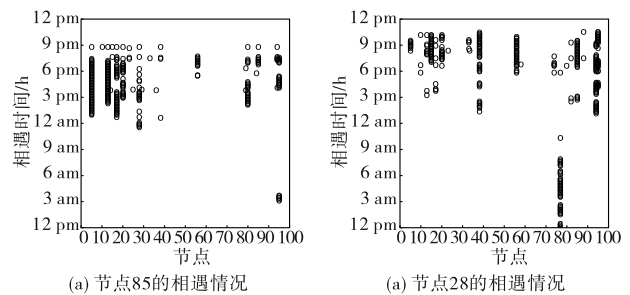


图2 MIT真实数据集中节点85和28与其他节点的相遇分布图

考虑到人的日常活动具有一定的规律这一事实,出现如图2中的现象是符合常理的。例如节点*i*和*j*可能是教室里上课的同学、同一实验室的同学,又或者是宿舍里的舍友,相应地他们的相遇时间和相遇次数也会不同。此外,在完整的一天中,如果他们一直保持接触,则节点*i*与*j*既可能是同一实验室里的同学,又可能是宿舍里的舍友。

2.3 朋友社区的划分算法

文献^[23,24]的研究表明大多数节点对之间的相遇时间符合对数正态分布,因此,它们在未来的相遇时间与它们过去的相遇时间有关(正态分布具有非记忆性)。此外,文献^[25]的研究指出真实个体的移动行为具有弱周期性,也就是说个体在未来的相遇时间与过去相遇有关。受这些研究的启发,可以

利用节点与其他节点的相遇历史记录来预测节点与其他节点相遇。在机会网络中,利用节点移动带来的通信机会进行通信时,节点设备的无线通信能力、存储资源、计算资源和能量等是极其有限的,所以无法获知网络的全局信息,且在进行社区划分时,算法的计算量不能太大。如果因感知网络的全局信息和进行复杂计算时消耗过多的能量,则节点的寿命会大大缩短,这样会进一步加快网络通信质量的下降。因此,利用节点间的相遇历史记录计算节点间的朋友关系强度,将网络中的节点划分到不同的朋友社区,具有实际的可行性。

机会网络中节点具有人的社会属性,因此本文利用节点在不同时间段属于特定社区的特点,结合节点与其他节点的相遇历史信息,将网络中的节点划分为不同的朋友社区。算法1给出了在特定时间内朋友社区划分的详细方法。

算法1 朋友社区划分算法

输入:节点的相遇历史记录 record,时间段 T 。

输出:朋友社区内的节点列表 friendList。

for each node i in record(T) /* 节点 i 的相遇历史记录 */

$t_{cur}^{start} \leftarrow 0$ /* 一次相遇的起始时间 */

$t_{cur}^{end} \leftarrow 0$ /* 一次相遇的结束时间 */

$P \leftarrow 0$

CFM $\leftarrow 0$

for Record(T) $_i$.hasnext() do

$t_{cur}^{start} \leftarrow \text{Record}(T)_i.\text{next}().\text{start}$

$t_{cur}^{end} \leftarrow \text{Record}(T)_i.\text{next}().\text{end}$

$P \leftarrow \frac{t_{cur}^{end} - t_{cur}^{start}}{T}$

CFM $\leftarrow -P \log_2 P + \text{CFM}$

end for

if CFM $_i > \tau$ then

friendList.add(i) /* 加入到朋友社区列表 */

end if

end for

由于节点(人)的移动具有规律性,在一天中的特定时间段内会属于某一特定社区,且社区内的节点成员几乎是固定不变的,所以当要获知在某一特定时间段内节点属于哪一社区及社区内的节点成员时,可以根据节点在前一天相应时间段内的相遇历史记录,通过社区划分算法计算得到。首先获得节点自身与节点 i 在时间段 T 内的相遇记录,再根据朋友关系度量标准计算节点间的 CFM 值。当 CFM 值大于阈值 τ 时,将节点 i 加入到朋友社区的节点列表 friendList 中。遍历完该时间段内与其他所有节点的相遇历史记录,就能得到节点在时间段 T 内所属朋友社区及社区内所有的节点成员。

利用节点自身的相遇历史记录,每个节点都能够计算出与其他节点的 CFM 值。当 CFM 的值大于一定的阈值 τ 时,就认为它们属于同一朋友社区。节点间的朋友关系会随时间的变化而变化,这与节点一天中所在不同时间段有关。

3 实验与结果分析

本文利用 The ONE^[26] 仿真平台进行仿真实验,验证所提出的朋友社区划分算法对机会网络的网络通信能力的影响。在实验中分别使用了真实环境下采集的数据集和综合移动模型产生的数据集。

3.1 数据集

3.1.1 真实的机会网络数据集

在实验中使用 MIT 的真实数据集(MIT reality data set),可以从 CRAWAD^[27] 下载。MIT reality data set 是由 97 名 MIT 的学生和教职工携带 97 部诺基亚智能手机产生的移动轨迹组

成。实验中采用从二月初到四月底三个月内的链接日志,这段时间是一学年的第二学期,其特点是人之间的关系相对稳定的,对校园活动的参加也比较积极^[28]。

3.1.2 WDM 移动模型产生的数据集

WDM(working day movement)模型^[29]通过对人在工作日内的三种主要行为活动进行建模,使 WDM 模型产生的节点移动轨迹更接近现实生活中节点的移动轨迹。在工作日内,人的三种主要行为活动为在家休息、在公司工作及晚上和朋友聚会,这三种行为活动分别对应一种子模型,节点移动特性根据一天当中的不同时间段在三种模型之间进行切换。

除了三种行为活动,WDM 模型中还包含了步行、开车和搭公交三种不同的出行方式,且它们的移动速度都不相同。WDM 模型还引入了社区和社会关系,而这些是无法通过简单的移动模型(如 RWP^[19]模型)实现。在同一个办公室工作、在晚上参加同一个活动或住在一起的节点构成社区。WDM 模型产生的间接连接时间和连接时间的分布与真实世界的测量情况相接近^[29]。

3.2 度量指标

在实验过程中,将经典的机会网络中的路由算法 Epidemic、Prophet 和 Spray and Wait,在没有朋友社区和使用了朋友社区的情况下进行比较。在使用了朋友社区时,使消息仅在朋友社区内进行传输,并通过传输成功率、传输延迟和路由开销三个度量标准进行评估。传输成功率是指成功传输到目的节点的消息总数与源节点产生的消息总数的比值。传输延迟是指成功传输的消息从源节点到目的节点所需的平均时间。路由开销是指所有节点转发的消息总数与源节点产生的消息总数的比值。

3.3 实验结果

在利用 MIT 真实数据集进行实验时,将数据集的 1/5 用于训练,让节点记录它们初始的连接历史。训练时间过后,在网络中每 6 min 产生一条消息,每条消息的源节点和目的节点都随机产生,总共产生 5 000 条。同时,假设每个节点都有足够的空间来缓存节点所接收到的消息,且有足够带宽来传递相遇节点间需要传递的消息。此外,对每种算法进行 10 次实验,并在同一数据集下,对每种算法的运行结果进行统计,实验结果为多次实验的平均值。

在利用 WDM 模型进行实验时,在网络中同样产生 5 000 条消息,每 15 s 产生一条,其他设置与上述相同。对实验场景的主要参数设置如表 1 所示。对于其他一些设置采用与文献[20]中相同的默认参数值,如区域划分(住宅区、办公区等)、位置设置(家的位置信息、公司位置信息、购物地点等)、节点行为特征(上班时间、下班后去购物及朋友聚会的概率、工作时喝咖啡的时间等)、交通工具(公交行驶路线、每一站的停留时间、公交数量)等。同样,对每种算法进行 10 次实验,实验结果为多次实验的平均值。

3.3.1 传输成功率比较

图 3、4 给出了各算法在两种不同数据集下传输成功率的比较。从图 3 可以看出,当阈值小于 0.9 时,传输成功率逐渐下降,之后几乎不再变化。从图 4 可以看出,当阈值小于 0.1 时,传输成功率会缓慢下降,之后几乎不再变化。这种不同的变化主要是由于不同数据集中节点数量不同所导致的。随着阈值的增大,朋友社区中的节点数量逐渐减少,最后趋于稳定。

由于 MIT 数据集中节点稀疏,节点间的相遇频率较低,对传输成功率影响较大。但在节点密度相对较高时,如在 WDM 模型下,阈值变化对传输成功率的影响较小。通过以上分析可以得出,在节点密度适中的前提下,一定范围内的阈值大小对传输成功率的影响较小。

表1 WDM模型的相关参数设置

类型	参数	值
场景特征	模拟时间	48 h
	区域范围	4 500 m × 3 400 m
	模拟背景城市	Helsinki
节点特征	个数	200
	移动速度	0.8 ~ 1.4 m/s
	传输速度	250 KBps
	节点传输范围	10 m
	节点缓存大小	不限
消息特征	大小	500 KB ~ 1 MB
	数量	5 000
	产生频度	15 s

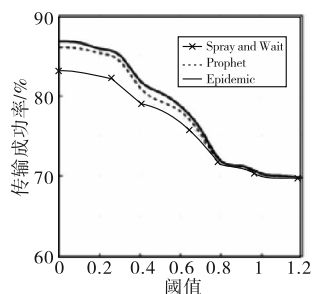
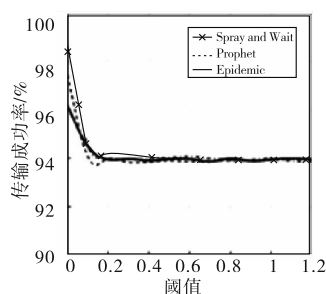


图3 MIT数据集下的传输成功率 图4 WDM模型下的传输成功率



3.3.2 传输延迟比较

图5、6给出了各算法在两种不同数据集下传输延迟的比较。从实验结果可以看出,在两种不同的数据集下,随着阈值的增加,传输延迟不断增加,最后趋于不变。阈值的增加导致了朋友社区中节点数量的减少,降低了对消息的转发频率,必然导致消息的传输延迟增加。

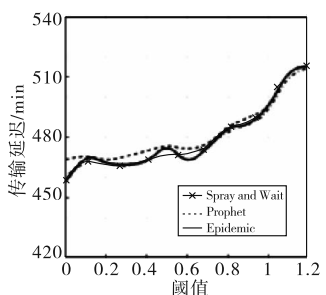
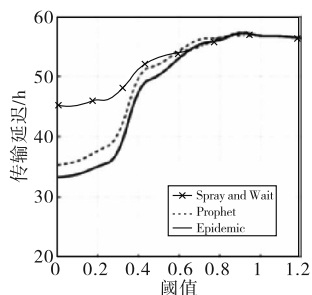


图5 MIT数据集下的传输延迟 图6 WDM模型下的传输延迟

3.3.3 路由开销比较

图7、8给出了各算法在两种不同数据集下路由开销的比较。从实验结果可以看出,在两种数据集下,随着阈值的增大,路由开销逐渐下降。阈值的增加导致了朋友社区中节点数量的减少,能够对消息进行转发的节点数量逐渐减少,消息在网络中的副本数量也逐渐减少,必然导致路由开销的逐渐降低。

通过以上三种度量标准下实验结果的分析可以看出,在不同节点密度下,通过设置合适的阈值,使消息在朋友社区内进行转发,可有效降低路由开销,对传输延迟影响较小,并可保持较高传输成功率。

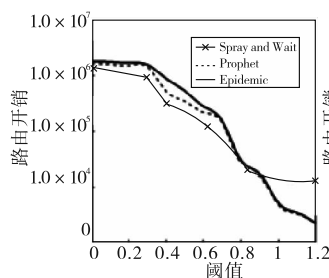


图7 MIT数据集下的路由开销

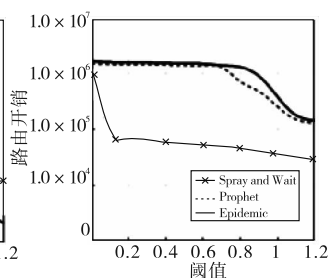


图8 WDM模型下的路由开销

4 结束语

本文对社会网络中传统的社区划分算法和机会网络中的路由特点进行分析,发现传统的社区划分算法并不适用于机会网络。为了找到适合机会网络的社区划分算法,本文提出了基于朋友关系的朋友社区划分算法。通过分析机会网络中节点间的关系,定义了节点间的朋友关系度量标准,并利用该度量标准,对机会网络进行朋友社区的划分。将经典的机会网络路由算法,在没有进行朋友社区划分和进行了朋友社区划分的两种不同的数据集下分别进行实验。实验结果表明,通过设置节点间朋友关系的阈值,对网络进行朋友社区划分,使消息在朋友社区内进行转发,可有效降低路由开销,同时保持较高的传输成功率。

参考文献:

- [1] 熊永平,孙利民,牛建伟,等. 机会网络[J]. 软件学报,2009,20(1):124-137.
- [2] Fall K. A delay-tolerant network architecture for challenged internets [C]//Proc of ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. New York:ACM Press,2003:27-34.
- [3] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks[J]. Nature,1998,393(6684):440-442.
- [4] Barabasi A L, Albert R. Emergence of scaling in random networks [J]. Science,1999,286(5439):509-512.
- [5] 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.
- [6] Tyler J R, Wilkinson D M, Huberman B A. E-mail as spectroscopy: automated discovery of community structure within organizations [M]//Communities and Technologies. Berlin: Springer, 2003: 143-153.
- [7] Radicchi F, Castellano C, Cecconi F, et al. Defining and identifying communities in networks[J]. Proceedings of the National Academy of Sciences of the United States of America,2004,101(9):2658-2663.
- [8] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. Physical Review E: Statistical Nonlinear & Soft Matter Physics,2004,69(2):026113.
- [9] Newman M E J. Fast algorithm for detecting community structure in networks[J]. Physical Review E: Statistical Nonlinear & Soft Matter Physics,2004,69(6):066133.
- [10] Guimerà R, Amaral L A N. Functional cartography of complex metabolic networks[J]. Nature,2005,433(7028):895-900.
- [11] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics Theory & Experiment,2008,2008(10):155-168.

工训练语料难以获取的问题,提出了 SVM 与 TBL 相结合进行词性标注的方法,并应用于轨道交通工程招中标领域自适应的任务中,验证本文研究成果的有效性。通过分析招中标领域词汇形态特征,对有关的语法特点和语言现象进行总结,设计 SVM 的词性标注特征模板和 TBL 规则模板,并采用核心属性渗透方法标注合成词的词性。在该语料库的构建方面,基于通用基础训练语料,并辅助以小规模的人工标注的领域数据语料,使用领域自适应技术提升合成词的自动标注的准确率。在未来的研究工作中,将针对涉及影响标注结果的特征、规则等因素进行更深入的研究,以期能找到一些更有实用价值的特征信息以及分辨能力更强的转换规则,使本文方法的整体标注性得到进一步的提升,并对领域自适应的方法进行更深入的探索。

参考文献:

- [1] 姜维,王晓龙,关毅,等. 基于多知识源的中文词法分析系统[J]. 计算机学报,2007,30(1):137-145.
- [2] 梁喜涛,顾磊. 中文分词与词性标注研究[J]. 计算机技术与发展,2015,25(2):175-180.
- [3] 赵岩,王晓龙,刘秉权,等. 融合聚类触发对特征的最大熵词性标注模型[J]. 计算机研究与发展,2006,43(2):268-274.
- [4] 余昕聪,李红莲,吕学强. 最大熵和 HMM 在中文词性标注中的应用[J]. 无线互联科技,2014(11):122-124.
- [5] 袁里驰. 基于改进的隐马尔可夫模型的词性标注方法[J]. 中南大学学报:自然科学版,2012,43(8):3053-3057.
- [6] 姜芳,李国和,岳翔,等. 基于粗分和词性标注的中文分词方法[J]. 计算机工程与应用,2015,51(6):204-207,265.
- [7] 洪铭材,张阔,唐杰,等. 基于条件随机场(CRFs)的中文词性标注方法[J]. 计算机科学,2006,33(10):148-151.
- [8] 王艺帆,王希杰. 基于双层条件随机场的汉语词性标注方法研究[J]. 安阳师范学院学报,2016(5):87-91.
- [9] 王丽杰,车万翔,刘挺. 基于 SVMTool 的中文词性标注[J]. 中文信息学报,2009,23(4):16-21.
- [10] Giménez J, Mürquez L. SVMTool: a general POS tagger generator based on support vector machines[C]//Proc of the 4th International Conference on Language Resources and Evaluation. 2004.
- [11] Florian R, Ngai G. Fast transformation-based learning toolkit[EB/OL]. [2008-09-10]. <http://nlp.cs.jhu.edu/~rflorian/fnlbl/documentation.html>.
- [12] Brill E. Transformation-based error-driven learning and natural language processing: a case study in part of speech tagging[J]. Computational Linguistics, 1995, 21(4):543-565.
- [13] 李文坤,张仰森,陈若愚. 基于词内部结合度和边界自由度的新词发现[J]. 计算机应用研究,2015,32(8):2302-2304,2342.
- [14] Song Yan, Xia Fei. Using a goodness measurement for domain adaptation: a case study on Chinese word segmentation[C]//Proc of the 8th International Conference on Language Resources and Evaluation. 2012:3853-3860.
- [15] Che Wanxiang, Li Zhenghua, Liu Ting. LTP: a Chinese language technology platform[C]//Proc of the 23rd International Conference on Computational Linguistics: Demonstrations. Stroudsburg, PA: Association for Computational Linguistics, 2010:13-16.
- [16] 杨梅. 现代汉语合成词构词研究[D]. 南京:南京师范大学,2006.
- [17] Palla G, Derényi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. Nature, 2005, 435(7043):814-818.
- [18] Evans T S. Clique graphs and overlapping communities[J]. Journal of Statistical Mechanics Theory & Experiment, 2010, 2010(12):257-265.
- [19] Lancichinetti A, Fortunato S, Kertész J. Detecting the overlapping and hierarchical community structure of complex networks[J]. New Journal of Physics, 2009, 11(3):19-44.
- [20] Vahdat A, Becker D. Epidemic routing for partially-connected Ad hoc networks[D]. Durham:Duke University, 2000.
- [21] Spyropoulos T, Psounis K, Raghavendra C S. Spray and wait: an efficient routing scheme for intermittently connected mobile networks[C]//Proc of ACM SIGCOMM Workshop on Delay-Tolerant Networking. New York:ACM Press, 2005:252-259.
- [22] Lindgren A, Doria A, Schelén O. Probabilistic routing in intermittently connected networks[J]. ACM SIGMOBILE Mobile Computing & Communications Review, 2004, 7(3):239-254.
- [23] Grossglauser M, Tse D N C. Mobility increases the capacity of Ad hoc wireless networks[J]. IEEE/ACM Trans on Networking, 2002, 10(4):477-486.
- [24] Bettstetter C, Hartenstein H, Pérezcosta X. Stochastic properties of the random waypoint mobility model[J]. Wireless Networks, 2004, 10(5):555-567.
- [25] Hui Pan, Crowcroft J, Yoneki E. Bubble rap: social-based forwarding in delay tolerant networks[J]. IEEE Trans on Mobile Computing, 2011, 10(11):1576-1589.
- [26] Chen Kang, Shen Haiying. SMART: lightweight distributed social map based routing in delay tolerant networks[C]//Proc of the 20th IEEE International Conference on Network Protocols. Piscataway, NJ: IEEE Press, 2012:1-10.
- [27] Eagle N, Pentland A, Lazer D. Inferring social network structure using mobile phone data[J]. Proceedings of the National Academy of Sciences of the United States of America, 2009, 106(36):15274-15278.
- [28] Srinivasa S, Krishnamurthy S. CREST: an opportunistic forwarding protocol based on conditional residual time[C]//Proc of IEEE Communications Society Conference on Sensor, Mesh and Ad hoc Communications and Networks. Piscataway, NJ:IEEE Press, 2009:342-350.
- [29] Tournoux P U, Leguay J, Benbadis F, et al. The accordion phenomenon: analysis, characterization, and impact on DTN routing[C]//Proc of IEEE INFOCOM. Piscataway, NJ: IEEE Press, 2009:1116-1124.
- [30] Liu Cong, Wu Jie. Routing in a cyclic mobispace[C]//Proc of the 9th ACM International Symposium on Mobile Ad hoc Networking and Computing. New York:ACM Press, 2008:351-360.
- [31] Keränen A, Ott J, Kärkkäinen T. The ONE simulator for DTN protocol evaluation[C]//Proc of the 2nd International Conference on Simulation Tools and Techniques. Brussels, Belgium: ICST, 2009: Article No. 55.
- [32] CRAWDAD data set[EB/OL]. <http://crawdad.cs.dartmouth.edu>.
- [33] Hui Pan, Crowcroft J. Predictability of human mobility and its impact on forwarding[C]//Proc of the 3rd International Conference on Communications and Networking. Piscataway, NJ:IEEE Press, 2008: 543-547.
- [34] Ekman F, Keränen A, Karvo J, et al. Working day movement model[C]//Proc of the 1st ACM SIGMOBILE Workshop on Mobility Models. New York:ACM Press, 2008:33-40.

(上接第1341页)