

文章编号:1007-5321(2017)02-0097-05

DOI:10.13190/j.jbupt.2017.02.016

校园移动社交网络中基于种子的数据分发算法

王青山^{1,2}, 汤 颖¹, 王 琦¹, 王 翟¹, 陈莉君¹

(1. 合肥工业大学 数学学院, 合肥 230009; 2. 南京大学 计算机软件新技术国家重点实验室, 南京 210093)

摘要: 为了研究如何利用节点间间歇性连接传输数据,提出了校园移动社交网络中基于种子的数据分发算法,其主要思想是为每一个社区选择一个种子节点,并利用种子节点来进行数据分发. 仿真实验表明,与著名的 Epidemic、PS 和 SGBR 算法相比,该算法可明显地降低网络开销,同时接近 Epidemic 算法达到的最大传递率.

关 键 词: 数据分发; 课程表; 网络开销; 移动社交网络

中图分类号: TP393

文献标志码: A

Seed-Based Data Dissemination in Campus MSNets

WANG Qing-shan^{1,2}, TANG Ying¹, WANG Qi¹, WANG Di¹, CHEN Li-jun¹

(1. School of Mathematics, Hefei University of Technology, Hefei 230009, China;

2. Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)

Abstract: In mobile social networks (MSNets), the packet is propagated through the intermittently connectivity which is established when two human-carried wireless-enabled devices are within communication range of each other. In order to use the intermittently connectivity to disseminate the packet, a seed-based data dissemination (SDD) algorithm was proposed in campus MSNets. The main idea is to choose a seed node for each community and use the seed node in data dissemination. Simulations show that, compared with the Epidemic algorithm, the publish/subscribe system and social groups based routing, SDD algorithm could reduce network overhead significantly, and near to the maximum delivery ratio obtained by the Epidemic algorithm.

Key words: data dissemination; student schedule; network overhead; mobile social networks

容迟网络(DTNs, delay tolerant networks)^[1]是一类节点间歇性连接的无线网络,DTNs 应用广泛,包括移动社交网络(MSNets, mobile social networks)、车载自组织网络、传感器网络和星际网络. 随着智能手机等移动终端的广泛使用,移动社交网络越来越受到关注. 移动社交网络中节点稀疏、电池能量受限等使源节点到目的节点很难建立一条端到端的路径. 因此,路由算法成为一个关键问题.

Epidemic 路由^[2]是容迟网络中首先被提出的

路由算法,携带数据包的节点通过移动遇到其他没有此数据包的节点,则复制数据包给该节点. 为了减少网络中数据包的拷贝数,国内外学者提出一些路由算法,包括基于概率的路由算法^[3]、基于历史轨迹的路由算法^[4]、基于节点或社区的社会属性的路由算法^[5-10]. 但是,针对校园移动社交网络这类特殊容迟网络中的数据分发问题,很少有人研究. 校园移动社交网络中学生的活动范围基本在教室、食堂、寝室、商店、电影院等场所,体现了规律性和可

收稿日期: 2016-08-27

基金项目: 国家自然科学基金项目(61571179,61401144); 南京大学计算机软件新技术国家重点实验室开放基金项目(KFKT2014B22); 中央高校基本科研业务费项目(2015HGZX0019)

作者简介: 王青山(1975—),男,副教授,硕士生导师, E-mail: qswang@hfut.edu.cn.

预测性. Srinivasan 等^[11]用学生课表做为学生联系模式,并使用 Epidemic 路由算法在校园移动社交网络中分发数据,实验结果表明,数据包可以快速蔓延. 为了减少多余的拷贝,提出了基于种子的数据分发算法(SDD, seed-based data dissemination),它的主要思想是为每一个社区选择一个种子节点来扩散消息,当前种子节点负责拷贝数据包给遇到对该包感兴趣节点以及为每个下时段社区选择一个种子节点,同时,在课间处于同一热点区域内节点相互传包以加速数据包在不同社区传输速度.

1 数据收集和模型

笔者从学校教学管理数据库中收集了主校区2014上半年共1万个学生的课程表,作为研究的数据. 这些学生来自14个学院,例如仪器学院、机械学院等. 数据包括这些学生的所有必修课程和选修课程. 将学生模型化为节点,通过节点间机会连接形成移动社交网络. 通过对数据分析得出星期一到星期五学生上课的人数都超过70%,因此这个移动社交网络可以有效地传输数据.

社区的定义如下.

定义1 2个学生属于同一社区当且仅当他们在同一间教室上同一节课.

社区是由教室和上课时间共同决定,在某时刻,一个学生至多属于一个社区,因为一个学生在该时刻只能在一个教室上课或者没有课. 上午和下午各有2节大课,每节大课包含2节小课;晚上有1节大课,包含3节小课. 因此总共有11节小课. 为了简化起见,在不引起混淆的地方都称小课为课. 每节小课表示为 $[t_{i,start}, t_{i,end}]$ ($1 \leq i \leq 11$),课间都有休息时间. 假设: 1) 学生严格按照自己的课程表上课; 2) 根据自己的课程表,每个学生知道任意时刻自己所属社区.

学生在日常活动中往往对感兴趣的信息具有相对固定的属性,用一个兴趣集合表示. 例如篮球、跑步、看书. 同理,对网络中传输的数据包都有一个兴趣标签. 数据包的兴趣标签表示某一种兴趣. 如果它是某个学生兴趣集合中的一个元素,则说明这个学生对该数据包感兴趣并希望收到该包. 例如,有一个关于篮球比赛的数据包,其兴趣标签是篮球,刚好学生A的兴趣集合中包含篮球这一元素,则学生A对此包感兴趣并希望接收该包. 本研究期望兴趣集合包含数据包兴趣标签的节点都能收到此包.

假设课间学生有可能离开教室去每个楼层 N_{hall} 个热点区域,如卫生间、饮水区域等. 学生课间从教室走到同一层的某个热点的概率为 p_{out} . 假设在同一时刻位于同一间教室或者同一热点区域的节点之间可以相互通信.

2 基于种子的数据转发算法(SDD)

定义2 在教室中负责转发数据包的数据包携带者称为种子节点.

源节点是最初的种子节点并且一直是种子节点,其中源节点是从所有节点中随机选取出来的. 另一方面,种子节点将选择第一次遇到的下节大课与自己不在同一教室的节点为下节大课该社区的种子节点,并且复制包给该节点. 每个种子节点有一张种子选择表来储存自己选择的种子并且不断更新,其中种子选择表包括该所选种子编号和所选种子对应的下节大课所在社区编号.

为了减少网络中的拷贝数目,每个社区只有一个种子节点. 当多个种子出现在当前社区时,需要执行种子相互吸收动作. 种子节点 i 吸收种子节点 j ,意味着节点 i 仍是种子节点,但是节点 j 变成普通节点. 吸收方法如下: 1) 源节点种子节点吸收其他种子节点; 2) 同为非源节点种子节点,对此包感兴趣种子节点吸收对此包不感兴趣的种子节点; 3) 否则,节点编号大的种子吸收编号小的种子. 种子节点 i 吸收种子节点 j 执行的动作如下: 1) 将节点 j 为下节大课选择的种子节点信息拷贝到节点 i 的种子选择表中; 2) 将节点 j 设置为普通节点.

SDD 算法的基本思想是为每个社区选择一个种子节点,并利用种子节点来扩散数据包. 社区内和社区间的数据分发算法分别见算法1和算法2. 在算法1中,为了减少多余数据包拷贝,只有种子节点拷贝数据包给下面2种相遇节点: 1) 对数据包感兴趣的节点; 2) 下节大课与自己不是同一社区的且是该社区的种子节点. 如算法1所示,在(1.2)、(2.1.2.a)、(2.2.1)和(2.2.2.a)步骤中,节点 i 拷贝数据包给节点 j ,在(2.1.1.b)、(2.1.2.b)和(2.2.2.b)步骤中,节点 i 选择节点 j 作为节点 j 下节大课的种子节点,在(1.1)和(2.1.1.a)步骤中,节点 i 和节点 j 之间进行吸收动作.

算法1 社区内数据分发

当种子节点 i 与节点 j 相遇:

1) 若节点 j 和节点 i 下节大课在同一社区;

(1.1) 若节点 j 也是种子节点, 则进行吸收动作;

(1.2) 若节点 j 没有拷贝但对包感兴趣, 则节点 i 拷贝包给节点 j ;

2) 否则

(2.1) 若节点 j 和节点 i 下节大课不在同一社区;

(2.1.1) 若节点 j 有拷贝;

(2.1.1.a) 若节点 j 是种子, 则进行吸收动作;

(2.1.1.b) 若节点 i 没有为节点 j 的下节大课选择种子, 则节点 i 选择 j 为其下节大课所在社区种子;

(2.1.2) 否则

(2.1.2.a) 若节点 j 对包感兴趣, 则节点 i 拷贝给节点 j ;

(2.1.2.b) 若节点 i 没有为节点 j 的下节大课选择种子, 则节点 i 选择节点 j 为其下节大课所在社区种子;

(2.2) 否则

(2.2.1) 若节点 i 下节大课有课, 节点 j 没课但对包感兴趣, 节点 i 拷贝包给节点 j ;

(2.2.2) 若节点 j 下节大课有课, 但节点 i 没课;

(2.2.2.a) 若节点 j 对包感兴趣, 则节点 i 拷贝包给节点 j ;

(2.2.2.b) 若节点 i 没有为节点 j 的下节大课选择种子, 则节点 i 选择节点 j 作为其下节大课所在社区种子;

3) 结束.

为了提高数据包在社区间的传播速度, 下课期间, 走出教室来到热点区域的数据包携带者或种子节点拷贝包给感兴趣的节点, 并且为下个时段社区选择种子. 社区间数据分发过程如算法 2 所示, 步骤 (2.1) 表示携带包的节点或种子节点复制包给感兴趣的节点, 步骤 (1.1) 和 (2.2) 表示为下个时段社区选择种子节点.

算法 2 社区间数据分发

当数据包携带者或者种子节点 i 与不在同一社区节点 j 相遇;

1) 若节点 j 有拷贝

(1.1) 若节点 j 下节大课有课, 节点 i 没有为节点 j 的下节大课所在社区选种子, 则节点 i 选择节

点 j 作为节点 j 下节大课所在社区种子;

2) 否则

(2.1) 若节点 j 对包感兴趣, 则节点 i 拷贝包给节点 j ;

(2.2) 若节点 j 下节大课有课, 节点 i 没有为节点 j 的下节大课所在社区选种子, 则节点 i 选择节点 j 为其下节大课所在社区种子;

3) 结束.

3 仿真实验

在仿真实验中, 将学校教学管理数据库^[12]中主校区 2014 上半年共 1 万个学生课程表, 作为网络中节点的移动模式. 假设这些学生严格按照课表进行活动. 采用 C++ 编写仿真程序. 仿真过程中选择著名的 Epidemic^[2]、PS (publish/subscribe system)^[7]和 SGBR (social groups based routing)^[5]算法与提出的 SDD 算法进行多次比较. 通过分析下面 3 个方面来评估 SDD 算法的网络性能: 1) 数据包传递率, 即数据包被成功传递到目的节点的百分比; 2) 代价, 定义为网络中拥有数据包拷贝的节点数平摊到成功收到包的感兴趣节点总数而得到的比值; 3) 传递延迟, 即成功传递的数据包在网络中平均延迟时间. 实验参数设置如下: 包的生存期 (TTL, time-to-live) 为 48 h; 在提出的 SDD 算法中, $N_{\text{hub}} = 2$, $p_{\text{out}} = 0.2$; 在 SGBR 算法中, 老化常数为 0.98, 更新因子为 0.45, 连通阈值为 0.5. 所有仿真结果是 600 次模拟实验的平均值.

首先, 为了研究学生上课的周数对 SDD 算法的影响, 改变发包周数从第 1 周 ~ 第 16 周, 每天发包时刻是 08:00:00, 实验结果如图 1 所示. 针对第 1 周的仿真结果得出它们的数据包传递率在置信度为 95% 时的置信区间为 $[0.852\ 437\ 4, 0.853\ 661\ 6]$, 传递延迟在置信度为 95% 时的置信区间为 $[81\ 072.13, 83\ 265.26]$, 拷贝数在在置信度为 95% 时的置信区间为 $[2.805\ 484, 2.834\ 619]$, 同时 P 值小于 0.05, 说明 600 次实验具有代表性. 图 1(a) 表明, 随着时间的增加, 所有算法的传递率先增加而后减少, 原因是在第 4 ~ 8 周学生上课人数比其他周要多, 因此相应数据包传递到感兴趣的节点概率较大. 进一步, SDD 算法的传递率比 PS 和 SGBR 算法大而且接近达到最高传递率的 Epidemic 算法. 因为 Epidemic 算法使用洪泛策略, 所以传递率达到上界. 对于代价来说, 如图 1(b) 所示, SDD 算法分别比 Epi-

demic、PS 和 SGBR 算法低 42.75%、39.80% 和 27.60%。图 1(c) 显示 SDD 的传递延迟比 Epidemic 稍大,其中 Epidemic 传递延迟由于采用洪泛策略达到下界。

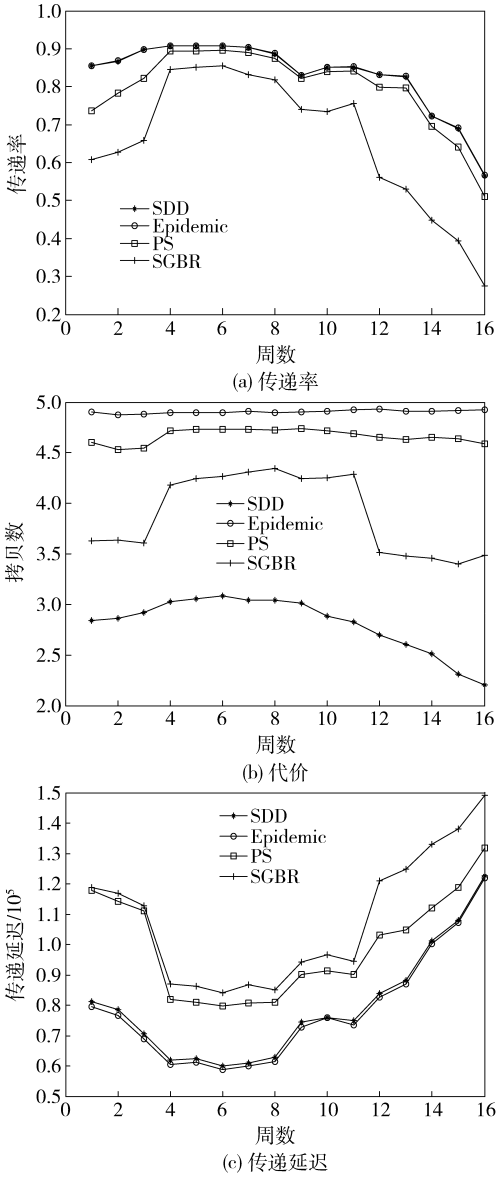


图 1 改变发包周数对网络性能的影响

接着,改变发包日期从星期一到星期五,发包时刻均为 08:00:00,仿真结果如图 2 所示. 图 2(a) 表示 SDD 算法的传递率比 PS 和 SGBR 算法大且只稍低于 Epidemic 算法. 同时,因为数据包在星期六不能得到分发,所以在星期五每个算法的传递率都较低. 在图 2(b)中 SDD 算法的代价低于其他两个算法,这表明利用种子节点扩散数据包能有效地节省网络资源. 图 2(c)表明 SDD 的传递延迟较 PS 和 SGBR 降低很多,只是比 Epidemic 稍大.

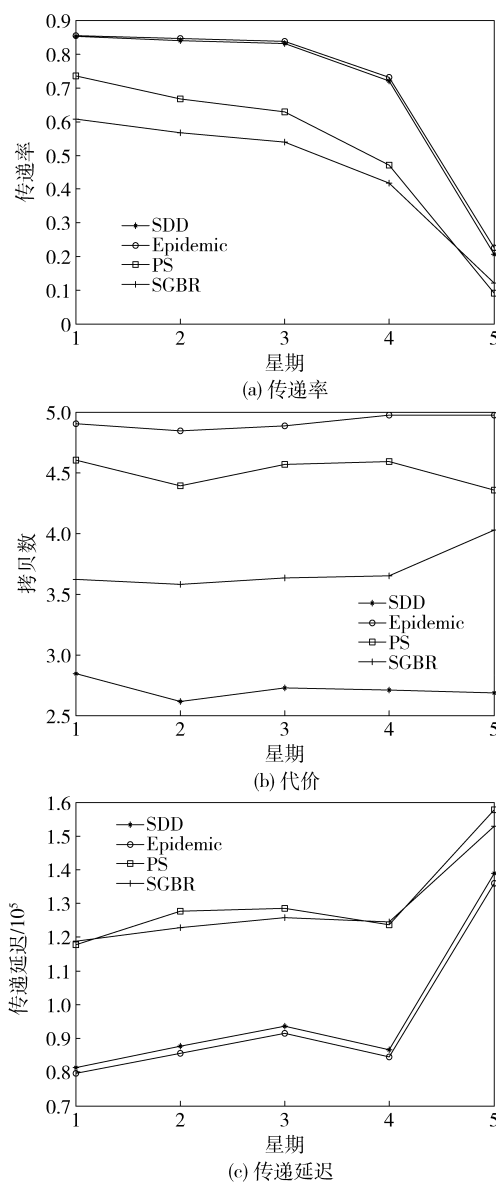


图 2 改变发包日期对网络性能的影响

4 结束语

讨论了校园移动社交网络中数据分发问题. 采用学生课表作为学生移动和联系模式,提出了基于种子的数据分发算法 SDD. 在 SDD 中,为了控制网络中数据包的数量,为每一个社区只选择一个种子节点并负责复制数据包给遇到的且对该包感兴趣节点和为下一时段社区选择种子节点来扩散数据包. 在课间休息位于同一热点区域内,数据包携带节点或者种子节点复制数据包给相遇感兴趣节点和为未来各社区选择种子节点. 最后,在仿真实验中,与著名的 Epidemic、PS 和 SGBR 算法相比,SDD 算法可以明显地降低网络开销,同时取得接近 Epidemic 算

法达到的最大传递率.

参考文献:

- [1] Zhu Ying, Xu Bin, Shi Xinghua, et al. A survey of social-based routing in delay tolerant networks: positive and negative social effects [J]. IEEE Communications Surveys & Tutorials, 2013, 15(1): 387-401.
- [2] Vahdat A, Becker D. Epidemic routing for partially connected ad hoc networks[R]. Duke University, 2000.
- [3] Lindgren A, Doria A, Schelén O. Probabilistic routing in intermittently connected networks [J]. Acm Sigmobile Mobile Computing and Communications Review, 2003, 7(3): 239-254.
- [4] 陶桦, 许艺凡, 王笑笑, 等. 一种运行轨迹驱动的移动自组织网络路由算法 [J]. 北京邮电大学学报, 2014, 37(6): 125-128.
Tao Hua, Xu Yifan, Wang Xiaoxiao, et al. A trace-driven routing algorithm in mobile ad hoc networks[J]. Journal of Beijing University of Posts and Telecommunications, 2014, 37(6): 125-128.
- [5] Hui P, Crowcroft J. How small labels create big improvements[C]//2007 IEEE PerCom Workshop on Smart Environments. Hawaii: IEEE Press, 2007: 65-70.
- [6] Abdelkader T, Naik K, Nayak A, et al. SGBR: a routing protocol for delay tolerant networks using social grouping[J]. IEEE Trans on Parallel and Distributed Systems, 2013, 24(12): 2472-2481.
- [7] Gao W, Cao G H, Porta T L, et al. On exploiting transient social contact patterns for data forwarding in delay-tolerant networks [J]. IEEE Trans on Mobile Computing, 2013, 12(1): 151-165.
- [8] Zhao Y X, Wu J. Socially-aware publish/subscribe system for human networks[C]//2010 IEEE Wireless Communications and Networking Conference (WCNC). Francisco, USA: IEEE Press, 2010: 1-6.
- [9] Lin C J, Chen C W, Chou C F. Preference-aware content dissemination in opportunistic mobile social networks[C]//2012 IEEE International Conference on Computer Communications (INFOCOM). Orlando: IEEE Press, 2012: 1960-1968.
- [10] Wei Kaimin, Guo Song, Zeng Deze, et al. Exploiting small world properties for message forwarding in delay tolerant networks[J]. IEEE Trans on Computers, 2015, 64(10): 2809-2818.
- [11] Srinivasan V, Motani M, Ooi W T. Analysis and implications of student contact patterns derived from campus schedules[C]//2006 12th Annual International Conference on Mobile Computing and Networking (MobiCom). Los Angeles: ACM, 2006: 86-97.