

# 一种基于节点行为模式的相似性指标\*

舒坚<sup>1,2</sup>, 马玉良<sup>1,2</sup>, 林伟杰<sup>1,2</sup>, 刘琳岚<sup>1,3</sup>



<sup>1</sup>(南昌航空大学 物联网技术研究所, 江西 南昌 330063)

<sup>2</sup>(南昌航空大学 软件学院, 江西 南昌 330063)

<sup>3</sup>(南昌航空大学 信息工程学院, 江西 南昌 330063)

通讯作者: 刘琳岚, E-mail: liulinlan@nchu.edu.cn

**摘要:** 手持设备网络(PSN)是一种利用节点移动和机会性相遇进行数据交互的延迟容忍网络(DTN).由于 PSN 具有网络拓扑不断变化、节点间链接时变性等特点,导致常用的节点间相似性评价方法不能很好地被应用于 PSN 的链路预测.针对 PSN 网络节点的行为特点,分析节点间产生连接次数、连接持续时间以及节点访问区域的规律,将 PSN 网络中节点的相似性分为节点社交行为相似性和移动行为相似性.结合社交行为相似性以及节点移动行为相似性,比较其在不同加权系数下的 AUC 和 Precision,提出基于节点行为模式的相似性指标(SNBP).在 MIT Reality 以及 Dartmouth Campus 数据集上的结果表明:与 CN、AA 和 Katz 相似性指标相比,所提出的相似性指标具有更好的预测准确率.

**关键词:** 链路预测;手持设备网络;相似性指标;节点行为模式

中文引用格式: 舒坚,马玉良,林伟杰,刘琳岚.一种基于节点行为模式的相似性指标.软件学报,2018,29(Suppl.(1)):92-104.  
<http://www.jos.org.cn/1000-9825/18010.htm>

英文引用格式: Shu J, Ma YL, Lin WJ, Liu LL. Similarity index based on node behavior patterns. Ruan Jian Xue Bao/Journal of Software, 2018,29(Suppl.(1)):92-104 (in Chinese). <http://www.jos.org.cn/1000-9825/18010.htm>

## Similarity Index Based on Node Behavior Patterns

SHU Jian<sup>1,2</sup>, MA Yu-Liang<sup>1,2</sup>, LIN Wei-Jie<sup>1,2</sup>, LIU Lin-Lan<sup>1,3</sup>

<sup>1</sup>(Internet of Things Technology Institute, Nanchang HangKong University, Nanchang 330063, China)

<sup>2</sup>(School of Software, Nanchang HangKong University, Nanchang 330063, China)

<sup>3</sup>(School of Information Engineering, Nanchang HangKong University, Nanchang 330063, China)

**Abstract:** Pocket switched networks (PSN) is a kind of delay tolerant network (DTN) which transmits messages based on node movement and opportunity encounter. The traditional similarity index based prediction methods are not suitable for the link prediction of PSN due to the frequent topology change and time-various link in PSN. According to the characteristics of PSN node behavior, this paper analyzes connection time, connection duration, and the law of node accessing areas. Node similarity is categorized into sociality behavior similarity and movement behavior similarity. After comparing AUC and Precision with different weight for sociality behavior similarity and movement behavior similarity, similarity index based on node behavior patterns (SNBP) is proposed. The experiment results on MIT Reality and Dartmouth Campus datasets show that comparison with CN, AA, and Katz similarity index, the proposed similarity index has better precision.

**Key words:** link prediction; pocket switched network; similarity index; node behavior pattern

\* 基金项目: 国家自然科学基金(61762065, 61363015, 61262020); 江西省自然科学基金(20171ACB20018, 20171BAB202009, 20171BBH80022); 江西省研究生创新专项资金(YC2018093)

Foundation item: National Natural Science Foundation of China (61762065, 61363015, 61262020); Natural Science Foundation of Jiangxi Province, China (20171ACB20018, 20171BAB202009, 20171BBH80022); Innovation Foundation for Postgraduate Student of Jiangxi Province (YC2018093)

收稿时间: 2018-05-01; 采用时间: 2018-08-30

手持设备网络(pocket switched network,简称 PSN)来源于延迟容忍网络(delay tolerant network,简称 DTN),网络节点基于“存储-携带-转发”的传输模式进行数据传输,从而导致网络拓扑不断变化,网络中的链路处于时而连接时而断开的状态.PSN 网络借助节点移动完成通信.如图 1 所示,源节点  $S$  在  $T_1$  时刻欲将信息传输至目标节点  $D$ ,但此时  $D$  节点不在  $S$  节点的通信范围内,因此  $S$  节点首先会把消息发送至其通信范围内的 1 号节点.经过一段时间到达  $T_2$  时刻, $D$  节点也不在 1 号节点通信范围内,因此 1 号节点将消息转发至其通信范围内的 4 号节点.直到  $T_3$  时刻 4 号节点移动到  $D$  节点的通信范围内,4 号节点把消息传送给  $D$  节点,至此完成一次 PSN 网络的通信过程.

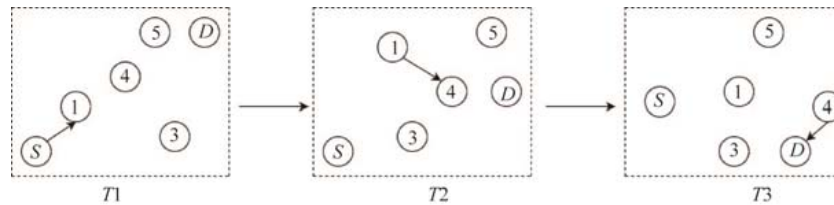


Fig.1 Pocket switched network message forwarding schematic

图 1 PSN 网络消息转发示意图

本文研究 PSN 网络节点间建立连接的规律和节点携带者之间的关系,构建基于节点行为模式的相似性指标(similarity based on node behavior pattern,简称 SNBP).本文第 1 节为相关研究.第 2 节分析 PSN 网络中的节点行为模式.第 3 节介绍 SNBP 相似性指标的构建过程.第 4 节进行实验分析.第 5 节总结全文.

## 1 相关研究

本文主要通过分析 PSN 网络中的节点行为研究其节点的相似性,由于 PSN 网络中的节点携带者是人,所以其相似性与社会网络中节点的相似性相关.

传统相似性指标大多针对静态拓扑结构网络<sup>[1]</sup>,从不同的角度分析网络结构,可将该类相似性指标分为基于共同邻居的相似性指标、基于路径的相似性指标、基于随机游走的相似性指标<sup>[2]</sup>.最为简单的是共同邻居指标(common neighbor,简称 CN),即两节点间的相似性与它们的共同邻居个数成正比<sup>[3]</sup>;基于路径的相似性指标中最具代表性的是 Katz 指标,Katz 指标考虑的是节点对之间路径数以及对应的路径长度来计算其节点相似值;基于随机游走的相似性指标,如 SimRank,是一种通过基于图的拓扑结构来衡量图中任意两个点的相似程度的方法,其核心思想是如果两个节点的邻居节点比较相似,那么这两个节点也应该比较相似,它们的邻居节点是否相似又由其邻居的邻居是否相似决定,如此迭代最终求出节点对之间的相似性<sup>[4]</sup>.

传统的基于网络结构的相似性指标常常通过加权的方式提高其预测性能.文献[5]通过比较基于局部信息的相似性指标在加权前后的预测性能,认为加权可提升基于局部信息的相似性指标预测性能;文献[6]比较了 CN、AA(adamic-Adar)、RA(resource allocation)、Salton、Jaccard 和 LP(local path)指标加权与不加权的形式在复杂网络进行链路预测时的效果,进一步验证了在复杂网络链路预测中,加权确实能够提升相似性指标的预测性能,因此如何选取合适的相似性指标权重,对于加权相似性指标预测性能的提升显得尤为重要;文献[7]提出将时间维度加入到相似性指标中,将不同时间段对应的网络聚集系数和 CN 指标结合,使得节点间相似性指标能够随着时间的推移而发生变化;文献[8]认为,节点对的相遇频率是预测节点对将来能否产生连接的重要依据,并将节点间的相遇频率(contact frequency)和 AA 指标结合作为节点对间的特征值,通过核函数进行链路预测并得到良好的效果;文献[9]认为,评估节点相似性不能仅仅只考虑节点局部或全局信息,还要对这两种信息进行综合考量,它利用信息的传输路径来结合这两种信息,从而构建出基于多路径传输节点相似性指标,对此定义出节点强度和链路强度,通过计算消息传递过程中链路的贡献度来计算节点强度和链路强度;文献[10]把社交网络看作是一个多关系并存的复杂网络,考虑节点对之间关系对节点相似性产生的影响,它将置信因子赋予网络中每个节点对,以此计算消息在某种关系下传输至另一个节点的概率,最终得到节点对在各类关系下的置信向量.

## 2 PSN 网络中的节点行为模式

由于 PSN 网络的节点携带者是人,因此网络中节点在行为模式上会受到节点本身的社会属性的影响.文献[11]提出一种机会网络路由协议,利用节点社会属性对节点间的相似性进行判断,以此作为协议的路由依据,文中指出,由于机会网络中缺少固定的网络结构信息,节点的携带者之间也不存在长时间稳定的链路,因此节点在通信时存在自私性,这意味着节点携带者与熟悉节点的通信几率远大于与陌生节点的通信几率,导致网络消息传递率下降;文献[12]对节点的周期性相遇状态用活跃度来进行描述,通过评估节点活跃度把问题转化成状态预测问题;文献[13]利用节点的自私性将 PSN 网络划分成若干消息交换簇,一个节点可能同时存在于几个消息交换簇中,源节点会把消息发送至与目的节点处于同一消息交换簇中的节点,并将这种选择中继节点的方式作为网络路由选择策略;文献[14]使用 PSN 网络节点携带者对应的社交网络信息,利用马氏距离计算节点相似性.上述研究都利用了 PSN 网络节点携带者的社交网络信息,但是个人隐私问题导致信息收集困难,同时携带者社交关系发生变化时必须同步更新社交信息,给消息转发传递带来额外开销.

与其他 DTN 网络,如 Ad hoc 网络一样,PSN 网络具有节点移动、网络拓扑不断变化等特点,且其节点移动具有规律性和目的性.文献[15]通过研究 PSN 网络节点移动方式,认为节点移动轨迹不完全随机,从而提出莱维行走模型(levy walk model);文献[16]则是通过计算 PSN 网络节点在社交网络的编辑距离从而评估节点相似性,实验证实,将其作为路由选择依据能够提高网络消息投递成功率.

文献[17]研究发现,Dartmouth 数据集中节点选择移动区域具有偏好性,即节点携带者会习惯性地出现几个固定区域位置,并且在这几个区域位置的访问概率远大于在其他区域位置的访问概率,实验结果如图 2 所示;文献[18]将节点访问次数最多的节点作为 HOME 节点,选择与目的节点相同的 HOME 节点作为中继节点,以此作为路由选择策略能够提高网络消息投递成功率.

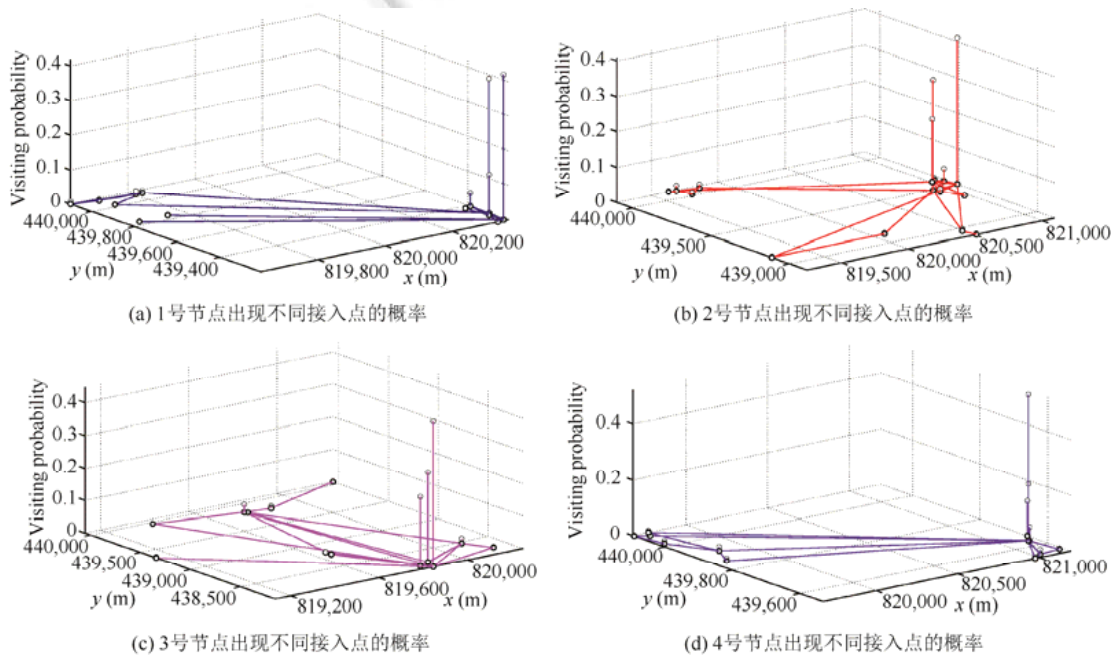


Fig.2 Randomly select 4 nodes in the Dartmouth data set and count their probability of appearing at different access points<sup>[17]</sup>

图2 Dartmouth 数据集中随机选取 4 个节点并统计其出现在不同接入点的概率<sup>[17]</sup>

综上所述,文献[5-10]表明,在对结构复杂且拓扑不断变化的网络进行相似性指标定义时,可以利用对相似性指标进行加权的方式提升其预测性能;文献[11-18]表明,综合考虑 PSN 网络节点移动规律及连接历史信息有

助于相似性计算.因此本文从社交行为和移动行为两个方面对节点行为模式进行分析:一方面通过研究节点间建立连接规律和节点携带者之间的关系,提出衡量节点间关系权重的方法,并将其代入传统相似性指标中,以此得出节点社交行为相似性指标;另一方面针对节点携带者访问区域具有规律性和集中性的特点,提出适用于 PSN 网络的节点移动行为相似性指标.通过加权系数将二者结合,从而构建得到适用于 PSN 网络的 SNBP 相似性指标.最终通过实验对节点社交行为、移动行为及行为模式这 3 种相似性指标进行验证.

### 3 基于节点行为模式的相似性指标

基于节点行为模式的相似性由节点移动行为的相似性和节点社交行为的相似性两部分组成.

#### 3.1 社交行为的相似性

PSN 网络中节点间的社交行为是指节点间数据交换和连接建立.由于 PSN 网络中节点位置和连接关系会随着节点的移动发生变化,在对 PSN 网络进行相似性指标计算时常常需要将数据进行时间片的划分,并对时间片内的网络结构和节点间关系进行分析.然而,这种方式在一定程度上忽略了节点间链路的持续时间、时间片内产生的连接次数等信息,这会使得在数据的处理上存在片面性.

为了验证网络节点数据交换和连接建立能够很好地表征节点社交属性,利用 SIGCOMM 2009 数据集分析网络节点建立连接与社交网络的对应情况.SIGCOMM 数据集记录了 3 天 SIGCOMM 会议中 76 个参会人员携带节点的活动情况,此外还记录了他们在 Facebook 上的社交信息.如图 3(a)所示,1 号节点与 22 号节点在会议前就存在社交关系(Facebook 互为好友),因此在会议期间连接频繁.与此相反,如图 3(b)所示,1 号节点和 40 号节点因为一直没有建立社交关系,致使会议期间只建立过一次连接.

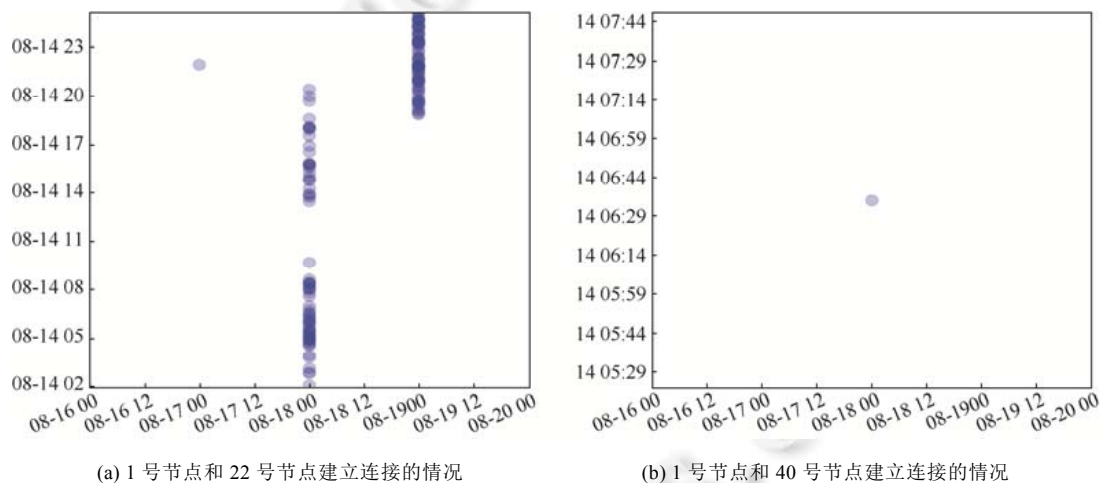


Fig.3 Comparison of the number of connections between friend nodes and non-friend nodes

图 3 朋友节点和非朋友节点间建立连接次数对比

为了验证节点间连接时长对节点社交属性的表征程度,分析 1 号与 69 号节点连接信息,该节点对在会议前互不相识,会议第 2 天(8 月 19 日下午 2 点)在 Facebook 上互加好友建立了社交关系.如图 4 所示,可以看出,在建立社交关系前,节点间联系虽然密集但是持续时间短,社交关系建立后,连接更加频繁且连接持续时间更长,说明节点间关系紧密程度和连接持续时间密切相关.

为不失一般性,本文在 76 个节点中随机抽取 5 个节点,并对比它们与好友和非好友节点间建立连接的次数、平均时长,实验数据如图 5 所示.结果表明,朋友节点对产生连接的次数、平均持续时长要远高于非朋友节点对.进一步验证了节点间关系紧密程度与连接建立次数、持续时间密切相关.

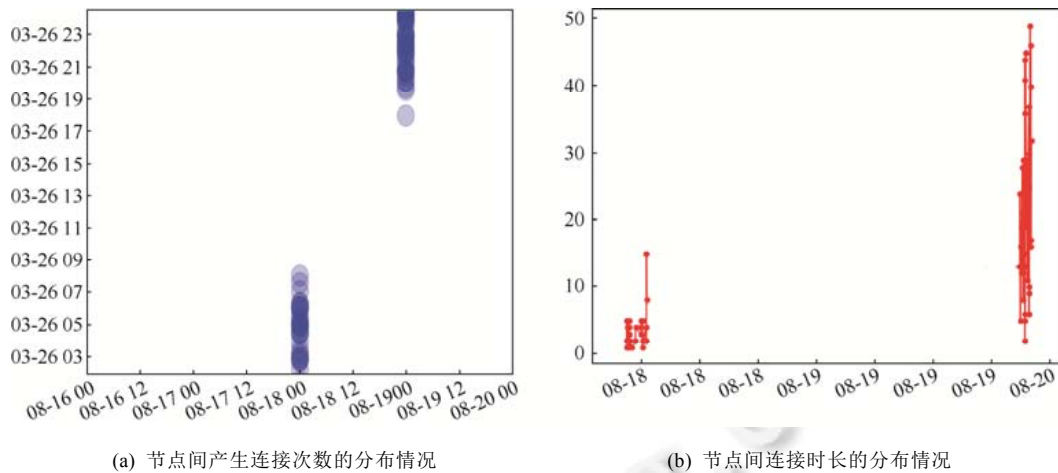


Fig.4 Comparison of the number of connections between nodes and the distribution of connection duration

图 4 节点间产生连接的次数以及连接时长分布情况对比

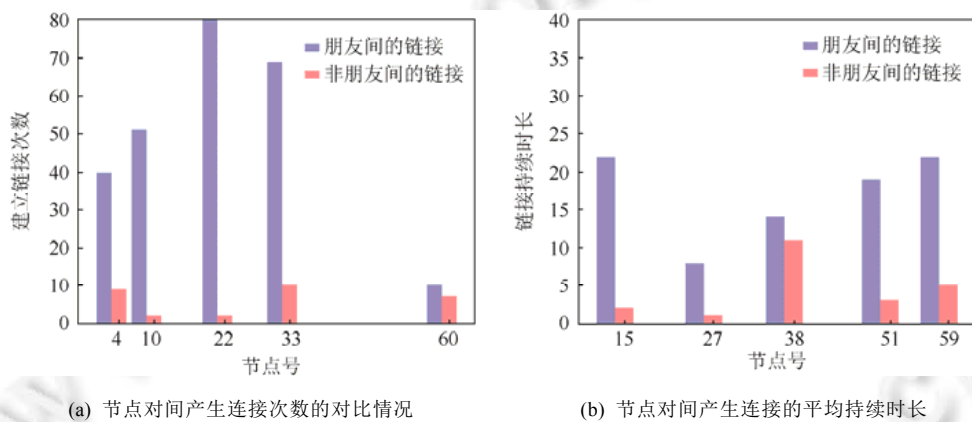


Fig.5 Comparison of the number of connections and duration between randomly selected pairs of nodes

图 5 随机选取的节点间产生连接次数和持续时长的对比

为了更好地定义 PSN 网络中节点连接时长和连接次数对节点间关系紧密程度的影响,本文做出如下假设.如图 6 所示,取时间片长度为  $T$  的时间片  $[T_S, T_N]$ ,  $T_S$  和  $T_N$  分别是时间片开始和结束时刻,  $T_{1S}$ 、 $T_{2S}$  与  $T_{1N}$ 、 $T_{2N}$  分别是节点对  $(A, B)$  在该时间片内两次连接开始与断开时刻.结合 PSN 网络的特点,网络中节点对  $(A, B)$  间的关系权重如式(1)所示.

$$\omega(A, B) = \sum_{i=1}^N e^{t_i}, i \in (1, 2, \dots, N) \quad (1)$$

其中,  $N$  为节点对  $(A, B)$  产生连接的次数,  $t_i$  是第  $i$  次连接对应的连接时长.指数函数  $e^{t_i}$  意味着节点对关系权重会随连接时长的增加而增加,反映了节点对紧密程度与连接时长的关系.

得到节点对关系权重后,将 PSN 网络的基于节点社交行为相似性指标看作是加权社交网络相似性指标,权重就是 PSN 网络节点对的关系权重,为保证文章的连贯性,这里先得出基于社交行为的相似性指标  $SS$ ,该指标由关系权重和  $AA$  相似性指标结合得到,如公式(2)所示.

$$SS = S_{AA} = \sum_{z \in \Gamma(A) \cap \Gamma(B)} \frac{\omega_{Az} + \omega_{Bz}}{\log(1 + S(z))} \quad (2)$$

其中,  $Z$  为节点对  $(A, B)$  的共同邻居集合,  $\omega$  为节点对关系权重,  $S(z)$  为节点  $z(z \in Z)$  与其所有邻居权重之和.

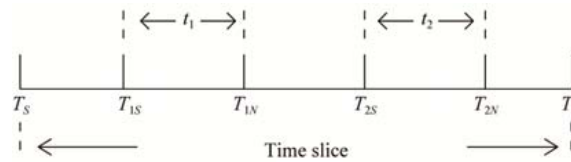


Fig.6 Schematic diagram of node connection duration and number of connections  
图6 节点连接时长和连接次数示意图

### 3.2 移动行为的相似性

文献[17,18]表明,网络节点移动具有惯性,即节点会习惯性地访问几个特定的区域位置.统计并分析节点对活动区域重合个数与平均连接次数的关系,如图7所示,实验结果表明,节点对间平均建立连接次数会随着活动区域重合个数的增加而增加.说明节点对间连接可能性会随着活动区域的重合访问接入点(access point,简称AP)个数的增加而提高.

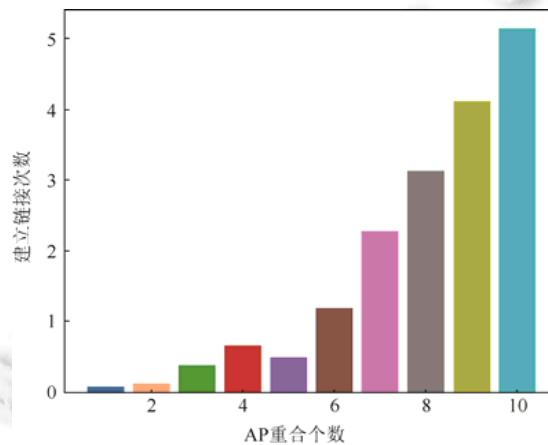


Fig.7 Relationship between coincidence number of access point and average number of connections of nodes  
图7 AP重合个数与节点平均连接次数之间的关系

从前面对PSN网络节点社会性分析来看,网络节点数据交换和连接建立可以很好地表征节点社交属性,而这些又与节点访问区域密切相关,下面对节点访问区域与节点相似性的关系进行研究.

在MIT Reality Mining数据集中,对节点活动区域按访问次数进行排序并取前10个最常访问的区域,统计节点在这些区域中与其他节点连接的概率,找出节点相似性与活动区域的关系.图8为累计分布图,纵坐标为累计分布函数,横坐标为节点访问区域个数.

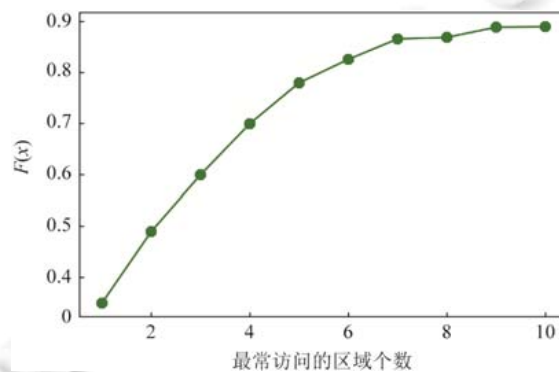


Fig.8 The most frequently visited area and the probability of making a connection in the area  
图8 最常访问区域与在区域中产生连接的概率

可以看出,有 90%的连接是在这些区域中建立的,前 2 个区域甚至产生了一半的连接.从而验证了最常访问区域个数与节点相似性成正比的关系,因此节点对最常访问区域重合个数越多,节点对相似性就越高.

此外,根据节点携带者社交关系把 MIT Reality Mining 数据集中节点对分为朋友、非朋友节点对.分析数据集中节点对活动频率和节点重合度的关系,分析结果如图 9 所示,可以看出,朋友节点对最常活动区域重合度高于非朋友节点对,证明节点相似性与活动区域重合度存在联系.

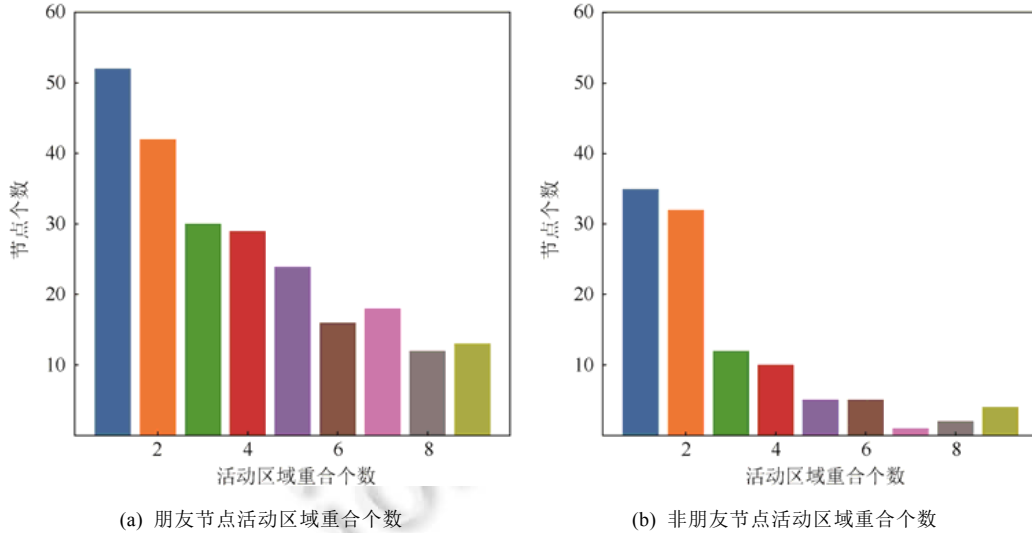


Fig.9 Comparison of number of coincidence regions between friend node and non friend node activity area

图 9 朋友节点和非朋友节点活动区域重合度个数比较

为便于描述,假设节点  $A$ 、 $B$  对各区域访问次数分别为  $(a_1, a_2, \dots, a_N)$  和  $(b_1, b_2, \dots, b_N)$ , 定义其移动区域特征如下.

**定义 1.** 移动区域特征  $V$ .

移动区域特征是一个有限向量集合,包含所有共同访问区域的移动区域特征向量,定义为

$$V = \{V_1, V_2, \dots, V_n\} = \{(a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)\}, \quad n \leq \min\{M, N\} \quad (3)$$

其中,  $n$  表示共同访问区域个数,  $V_i = (a_i, b_i)$  表示第  $i$  个区域的移动区域特征向量,  $a_i$ 、 $b_i$  分别表示节点  $A$ 、 $B$  访问第  $i$  个区域的次数.

本文使用 Cosine Similarity<sup>[14]</sup>来衡量两个节点之间的相似程度. Cosine Similarity 思想是首先通过欧几里德点积公式得出向量  $\mathbf{a}$ 、 $\mathbf{b}$  间的余弦值,如式(4)所示.

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos\theta \quad (4)$$

然后由点积和向量长度得到余弦相似性,如公式(5)所示.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} = \frac{\sum_{i=1}^n a_i \times b_i}{\sqrt{\sum_{i=1}^n a_i^2} \times \sqrt{\sum_{i=1}^n b_i^2}} \quad (5)$$

其中,  $a_i$  和  $b_i$  表示向量  $\mathbf{a}$  和  $\mathbf{b}$  的分量.

借鉴 Cosine Similarity 思想由移动区域特征  $V$  得到节点对  $(A, B)$  的基于节点移动行为相似性指标  $SL$ , 如公式(6)所示.

$$SL = \frac{\sum_{i=1}^n a_i \cdot b_i}{\sqrt{\sum_{i=1}^n a_i^2} \times \sqrt{\sum_{i=1}^n b_i^2}} \quad (6)$$

其中,  $n$  表示共同访问区域个数,  $a_i$  和  $b_i$  分别表示节点  $A$  和  $B$  访问第  $i$  个区域的次数.

计算节点移动行为的相似性指标使用的是所有时间内的访问次数,而不仅仅只统计某个时间片内的访问次数,这与计算社交行为的相似性不同.

### 3.3 行为模式的相似性

PSN 网络在节点对处于同一通信范围的情况下才可能发生信息交换,所以社交行为相似性指标存在不足,它没有考虑节点所处位置带来的影响,同时,若在某时间片内节点对未连接或者无共同邻居,则该指标预测性能会很低.另一方面,基于节点移动行为相似性指标虽然很好地表征了节点携带者按自己意愿出现在某些固定区域的特点,但是有些区域本身就具有被经常访问的特点,如人流量大的超市、广场等,节点在上述场景中相似性指标趋近,导致预测性能无明显提升,只有记录更多的共同访问区域信息才能获得较为精准的预测结果.

综上所述,节点行为模式要综合考量社交行为相似性和移动行为相似性,通过线性加权的方式结合两个指标得到 SNBP 相似性指标,如式(7)所示.

$$SNBP = \sigma \cdot SL + (1 - \sigma) \cdot SS \quad (7)$$

$$\text{其中, } SL = \frac{\sum_{i=1}^n a_i \cdot b_i}{\sqrt{\sum_{i=1}^n a_i^2} \times \sqrt{\sum_{i=1}^n b_i^2}}, SS = S_{AA} = \sum_{z \in \Gamma(A) \cap \Gamma(B)} \frac{\omega_{Az} + \omega_{Bz}}{\log(1 + S(z))}, \sigma \text{ 为对应的权重.}$$

## 4 实验分析

本节使用 Dartmouth/Campus 和 MIT Reality 数据集,比较 SNBP 与 CN、AA、Katz 相似性指标的 Precision 和 AUC,验证 SNBP 能否有效反映 PSN 网络中链路变化的情况.

### 4.1 实验数据集

研究节点行为模式要利用网络节点连接情况和所处区域位置变化情况,因此,采用含有这两种信息的 MIT Reality 和 Dartmouth/campus 数据集作为实验数据集进行研究更具有针对性.数据集见表 1.

Table 1 Experimental data set

表 1 实验数据集

数据集	MIT	Dartmouth
实验时间	2002 年	2009 年
设备	蓝牙	无线传感器
节点个数	97	5 349
时间长度(天)	246	365
采样周期(秒)	300	120

其中,MIT Reality 数据集是麻省理工大学 97 名实验人员携带蓝牙设备进行通信的情况.由于校园范围大以及学生的作息习惯致使晚上记录极为稀疏,因此选用该数据集中一个月内每天 8 点~20 点共  $12 \times 30$  个小时的记录作为实验数据集,避免实验数据过于稀疏从而导致无法对比预测性能;Dartmouth/Campus 数据集相比 MIT Reality 数据集规模更大,包含 5 346 个节点和 532 个 AP 数据.同样,为了解决数据集过于稀疏问题,选取数据集中 2004 年 4 月份每天 8 点~20 点共  $12 \times 30$  个小时的记录作为实验数据集.

### 4.2 评价指标

本文使用受试者工作特征曲线下的面积(area under the curve,简称 AUC)和准确率 Precision 作为节点相似性的评价指标.

#### 4.2.1 AUC

AUC 可以简单理解为从测试集和不存在边集中各取一条测试边和不存在边,如果测试边相似性指标分数值大于不存在边,则加 1 分,等于加 0.5 分.AUC 定义如式(8)所示.



$$AUC = \frac{n' + 0.5n''}{n} \tag{8}$$

其中,  $n$  为独立比较次数,  $n'$  次测试边分数大于不存在边分数,  $n''$  次分数相同.

#### 4.2.2 Precision

Precision 只关心前  $L$  条边中预测正确的边所占比例. Precision 定义如式(9)所示.

$$precision = \frac{m}{L} \tag{9}$$

其中,  $m$  表示预测正确边个数.

可以看出, 如果两个相似性指标的 AUC 差不多, Precision 较大的相似性指标更好, 因为  $L$  相同的情况下, Precision 较大的相似性指标  $m$  更大, 意味着它只要较少的边就能找到更多正确的预测.

### 4.3 实验分析

通过实验对节点行为相似性、加权系数以及 SNBP 相似性指标进行验证, 并得到最佳的基于节点行为相似性指标和加权系数.

#### 4.3.1 节点社交行为相似性的验证

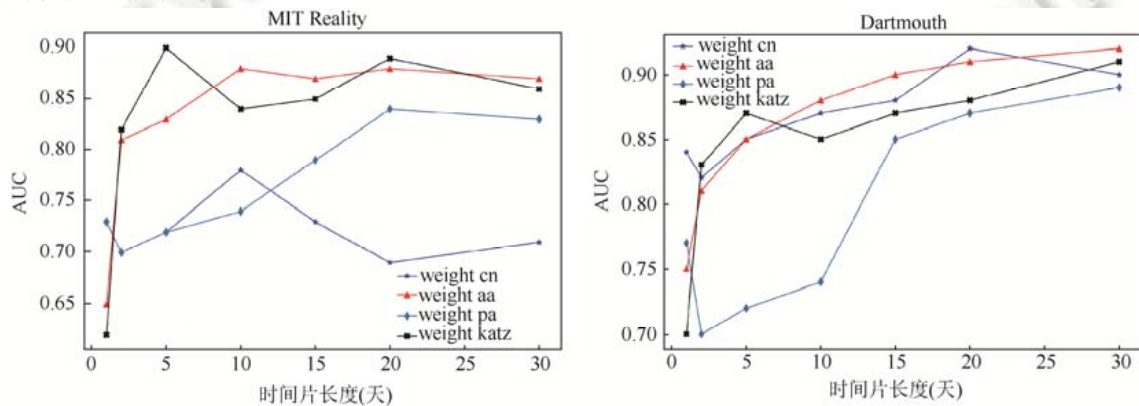
节点社交行为相似性指标结合了关系权重和传统相似性指标, 本文对传统的 CN、AA、PA 及 Katz 指标进行加权, 比较加权前后相似性指标的预测效果. 表 2 是对传统相似性指标加权后的表示形式.

**Table 2** Similarity index formula after weighting

表 2 加权后的相似性指标公式

WCN	$\sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{\omega_{xz} + \omega_{zy}}{2}$	WPA	$\sum_{i \in \Gamma(x)} \omega_{ix} \times \sum_{j \in \Gamma(y)} \omega_{jy}$
WAA	$\sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{\omega_{xz} + \omega_{zy}}{\log(1 + S(Z))}$	Wkatz	$\sum_{l=1}^L a^l \cdot  \omega'_{(x,y)} $

选取 1、2、5、15、20 及 30 天作为时间片长度, 验证 PSN 网络中基于节点社交行为相似性的有效性, 实验结果如图 10 所示.



(a) MIT Reality 中的 AUC 值

(b) Dartmouth 中的 AUC 值

Fig.10 AUC comparison of weighted similarity indicators at different time slice lengths

图 10 加权相似性指标在不同时间片长度下的 AUC 对比情况

传统的相似性指标经过加权后预测性能都有提升. 从图 10 可以看出, 加权后的 AA 和 Katz 指标结果更为理想. 为了避免 AUC 在数据不平衡情况下各类节点行为相似性指标的比较效果不够明显, 对节点行为相似性指标在 Precision 上的表现进行了验证, 实验结果如图 11 所示.

从两个数据集的表现来看, 加权的 AA 指标在不同时间片长度下表现得更稳定, 有较好的 Precision 表现和

AUC 表现,表明加权后的 AA 相似性指标能够更好地适应 PSN 网络,更适合用来构建基于社交行为的相似性指标.

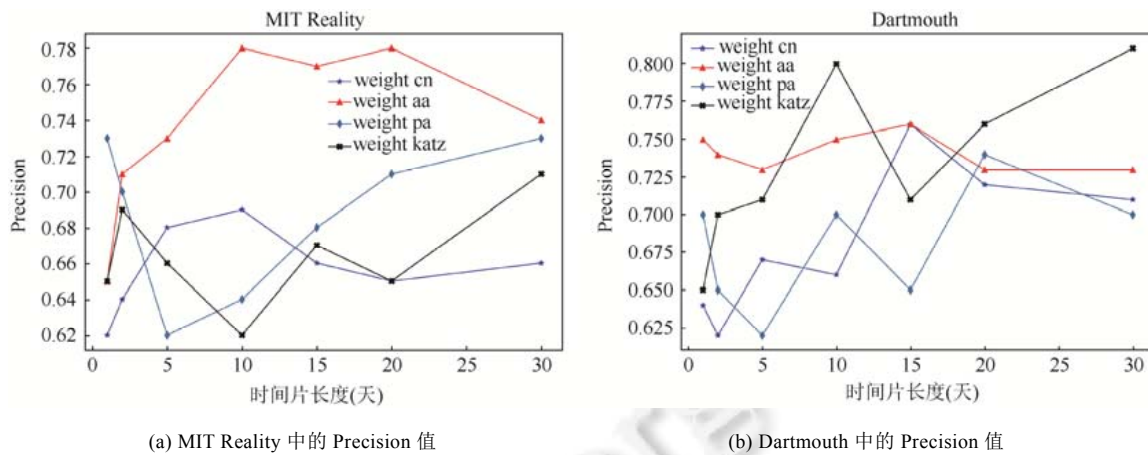


Fig. 11 Precision comparison of weighted similarity indicators at different time slice lengths

图 11 加权相似性指标在不同时间片长度下的 Precision 对比情况

#### 4.3.2 加权系数的选取

SNBP 相似性指标由节点社交行为、移动行为两个相似性指标线性加权组成,加权系数尤为重要.为了研究节点社交行为、移动行为两个相似性指标所占比重,在相同时间片下对不同加权系数的 SNBP 相似性指标进行实验,实验结果见表 3 和表 4.

**Table 3** AUC and precision under different  $\sigma$  in the MIT Reality data set

表 3 MIT Reality 数据集中不同  $\sigma$  下的 AUC 以及 Precision

$\sigma$	0.1	0.3	0.5	0.7	0.9	1
AUC	0.781	0.82	0.85	0.89	0.872	0.88
Precision	0.28	0.37	0.49	0.52	0.44	0.47

**Table 4** AUC and precision under different  $\sigma$  in the Dartmouth data set

表 4 Dartmouth 数据集中不同  $\sigma$  下的 AUC 以及 Precision

$\sigma$	0.1	0.3	0.5	0.7	0.9	1
AUC	0.87	0.89	0.9	0.91	0.88	0.84
Precision	0.52	0.541	0.55	0.56	0.521	0.41

从表中可以看出,当移动相似性指标权重  $\sigma$  取 0.7 时,SNBP 相似性指标在 MIT Reality 和 Dartmouth 数据集中有较好表现.

对 SNBP 相似性指标中移动相似性权重  $\sigma$  作进一步验证,将节点移动行为相似性指标及社交行为相似性指标权重分别定义为  $\alpha$  和  $\beta$ ,且  $\alpha + \beta = 1$ .在不同时间片长度下,将节点移动行为相似性指标和社交行为相似性指标的比重分别设置为相同 ( $\alpha: \beta = 0.5: 0.5$ )、节点移动行为相似性指标所占比重大 ( $\alpha: \beta = 0.7: 0.3$ ) 以及节点移动行为相似性指标所占比重小 ( $\alpha: \beta = 0.3: 0.7$ ) 这 3 种情况进行比较,所得结果如图 12 所示.

在时间片长度较短时,SNBP 相似性指标中节点社交行为相似性指标权重越大对应的 AUC 值越高,但随着时间的增加,SNBP 相似性指标的 AUC 值反而明显下降.在时间长度较大的情况下,SNBP 相似性指标中节点移动行为相似性指标权重高时对应的 AUC 值更高.说明在时间长度较大的情况下,若节点移动行为相似性指标权重重大,则 SNBP 相似性指标预测效果更为理想.

综上所述,SNBP 相似性指标中节点移动行为的相似性指标权重重大其预测效果更为理想,且取节点移动行为相似性指标权重为 0.7 时 SNBP 相似性指标最为理想.

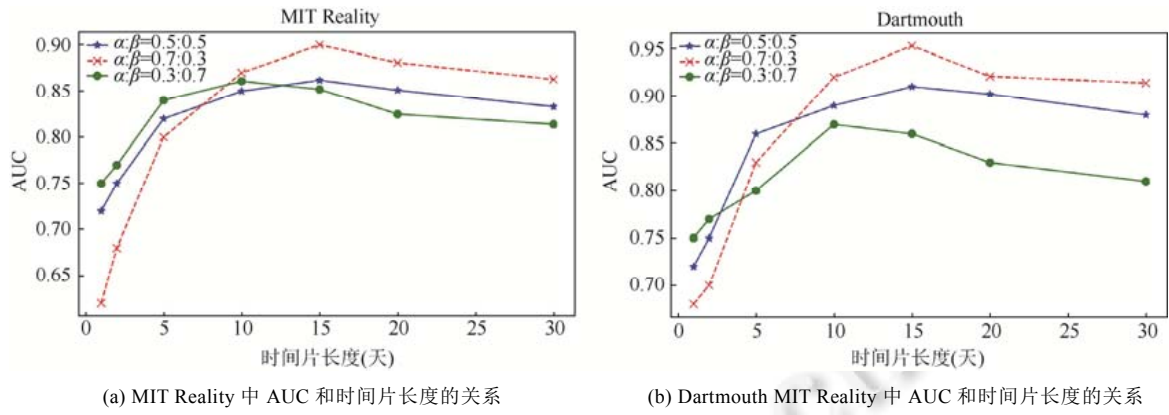


Fig.12 The AUC value of different ratios of  $\alpha$  to  $\beta$  at different time slice lengths

图 12 不同 $\alpha$ 与 $\beta$ 的比值在不同时间片长度下的 AUC 数值

4.3.3 指标有效性验证

在 MIT Reality 和 Dartmouth 数据集中,比较本文提出的 SNBP 相似性指标与 CN、AA、Katz 及 PA 等传统相似性指标的 AUC 和 Precision 值.实验参数 $\sigma$ 取 0.7,实验结果如图 13 和表 5 所示.

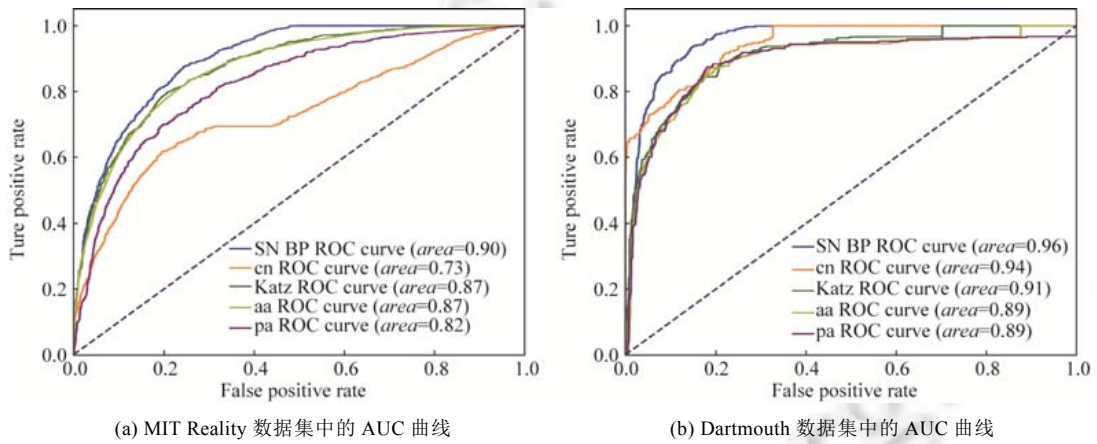


Fig.13 ROC graph of SNBP and traditional similarity indicators

图 13 SNBP 与传统相似性指标的 ROC 曲线图

Table 5 Precision comparison between SNBP and traditional similarity indicators

表 5 SNBP 与传统相似性指标的 Precision 对比情况

	CN	Katz	AA	PA	SNBP
MIT Reality	0.64	0.67	0.70	0.66	<b>0.71</b>
Dartmouth	0.73	0.72	0.69	0.74	<b>0.80</b>

从图 13 和表 5 可以看出,本文提出的 SNBP 相似性指标优于传统的相似性指标.从而证明 SNBP 相似性指标确实能够更好地对 PSN 网络中的节点相似性进行评估.

5 总结

本文针对 PSN 网络中节点具有社会属性和节点携带者的移动具有规律性的特点,提出 PSN 网络中节点社交行为的相似性指标和移动行为的相似性指标,综合考虑节点社交行为的相似性和移动行为的相似性,得到节点行为模式相似性指标 SNBP.两个真实数据集下的实验结果表明,与 CN、AA、Katz 及 PA 相比,SNBP 具有更

好的预测性能和预测准确率,更适用于 PSN 网络.本文工作可为 PSN 网络链路预测方法的深入研究提供参考.

#### References:

- [1] Rao J, Wu B, Dong YX. Parallel link prediction in complex network using mapreduce. *Ruan Jian Xue Bao/Journal of Software*, 2012,23(12):3175–3186 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4206.htm> [doi: 10.3724/SP.J.1001.2012.04206]
- [2] Niu XZ, Niu JJ, Su DZ, She K. Community detection based on weighted content-structural network and random walks. *Acta Electronica Sinica*, 2017,45(9):2135–2142 (in Chinese with English abstract). [doi: 10.3969/j.issn.0372-2112.2017.09.012]
- [3] Hu WB, Wang H, Yan LP, Qiu ZY, Xiao L, Du B. Hybrid quantum swarm intelligence indexing for event detection in social networks. *Ruan Jian Xue Bao/Journal of Software*, 2016,27(11):2747–2762 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4910.htm> [doi: 10.13328/j.cnki.jos.004910]
- [4] Lu J, Gong Z, Lin X. A novel and fast SimRank algorithm. *IEEE Trans. on Knowledge & Data Engineering*, 2017,29(3):572–585. [doi: 10.1109/TKDE.2016.2626282]
- [5] Liu Q, Li J, Xie Z, Zhang PY. An improvement of link prediction by combining local information and betweenness. In: *Proc. of the 11th Int'l Conf. on Natural Computation (ICNC)*. Zhangjiajie, 2016. 456–461. [doi: 10.1109/ICNC.2015.7378032]
- [6] Liu J. Analysis of link prediction method in mobile social network. In: *Proc. of the 7th Int'l Conf. on Measuring Technology and Mechatronics Automation*. Nanchang, 2015. 145–148. [doi: 10.1109/ICMTMA.2015.41]
- [7] Sharma S, Singh A. An efficient method for link prediction in complex multiplex networks. In: *Proc. of the 11th Int'l Conf. on Signal-Image Technology & Internet-Based Systems (SITIS)*. Bangkok, 2016. 453–459. [doi: 10.1109/SITIS.2015.93]
- [8] Huang D, Zhang S, Hui P, Chen Z. Link pattern prediction in opportunistic networks with kernel regression. In: *Proc. of the 7th Int'l Conf. on Communication Systems and Networks*. Bangalore, 2015. 1–8. [doi: 10.1109/COMSNETS.2015.7098684]
- [9] Liu MM, Guo JF, Luo X. Link prediction based on the similarity of transmission nodes of multiple paths in weighted social networks. *Journal of Information Hiding and Multimedia Signal Processing*, 2016,7(4):771–780.
- [10] Dai C, Chen L, Li B, Li Y. Link prediction in multi-relational networks based on relational similarity. *Information Sciences*, 2017,394-395:198–216. [doi: 10.1016/j.ins.2017.02.003]
- [11] Yao L, Man Y, Huang Z, Deng J, Wang X. Secure routing based on social similarity in opportunistic networks. *IEEE Trans. on Wireless Communications*, 2016,15(1):594–605. [doi: 10.1109/TWC.2015.2476466]
- [12] Li F, Si YL, Chen Z, Shen LM. Lightweight opportunistic routing forwarding strategy based on Markov chain. *Journal on Communications*, 2017,38(5):108–120 (in Chinese with English abstract). [doi: 10.11959/j.issn.1000-436x.2017105]
- [13] Rasul K, Makaroff D, Stanley KG. Hybrid community-based forwarding: A complete energy efficient algorithm for pocket switched networks. In: *Proc. of the 40th IEEE Local Computer Networks Conference Workshops (LCN Workshops)*. Clearwater Beach, 2015. 99–100. [doi: 10.1109/LCNW.2015.7365925]
- [14] Mei A, Morabito G, Santi P, Stefa J. Social-Aware stateless routing in pocket switched networks. *IEEE Trans. on Parallel & Distributed Systems*, 2015,26(1):252–261. [doi: 10.1109/TPDS.2014.2307857]
- [15] Rhee I, Shin M, Hong S, *et al.* On the levy-walk nature of human mobility. *IEEE/ACM Trans. on Networking*, 2011,19(3):630–643. [doi: 10.1109/TNET.2011.2120618]
- [16] Lin Y, Chen Z, Wu J. Efficient forwarding strategy for opportunistic network based on node similarity. *Theoretical Computer Science*, 2017,768:86–100. [doi: 10.1007/978-981-10-6893-5\_7]
- [17] Liu Y, Yang F, Zhang S, Zhou W. Impact of social features on the performance of pocket switched network. In: *Proc. of the 79th IEEE Vehicular Technology Conf. (VTC Spring)*. Seoul, 2015. 1–5. [doi: 10.1109/VTCSpring.2014.7022940]
- [18] Huang J, Cheng X, Bi J, Chen B. Wireless relay selection in pocket switched networks based on spatial regularity of human mobility. *Sensors*, 2016,16(1):94. [doi: 10.3390/s16010094]

#### 附中文参考文献:

- [1] 饶君,吴斌,东昱晓.MapReduce 环境下的并行复杂网络链路预测. *软件学报*, 2012,23(12):3175–3186. <http://www.jos.org.cn/1000-9825/4206.htm> [doi: 10.3724/SP.J.1001.2012.04206]

- [2] 牛新征,牛嘉郡,苏大壮,余堃.基于加权内容-结构网络和随机游走的社团划分算法.电子学报,2017,45(9):2135-2142. [doi: 10.3969/j.issn.0372-2112.2017.09.012]
- [3] 胡文斌,王欢,严丽平,邱振宇,肖雷,杜博.混合指标量子群智能社会网络事件检测方法.软件学报,2016,27(11):2747-2762. <http://www.jos.org.cn/1000-9825/4910.htm>. [doi: 10.13328/j.cnki.jos.004910]
- [12] 李峰,司亚利,陈真,申利民.基于马尔可夫链的轻量级机会路由转发策略.通信学报,2017,38(5):108-120. [doi: 10.11959/j.issn.1000-436x.2017105]



舒坚(1964—),男,江西南昌人,教授,CCF 高级会员,主要研究领域为物联网,软件测试,机器学习.



马玉良(1994—),男,学士,CCF 学生会会员,主要研究领域为机会网络.



林伟杰(1993—),男,硕士,CCF 学生会会员,主要研究领域为机会网络.



刘琳岚(1968—),女,教授,CCF 专业会员,主要研究领域为无线传感器网络,软件工程.