

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.

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

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

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 12×30 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 12×30 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

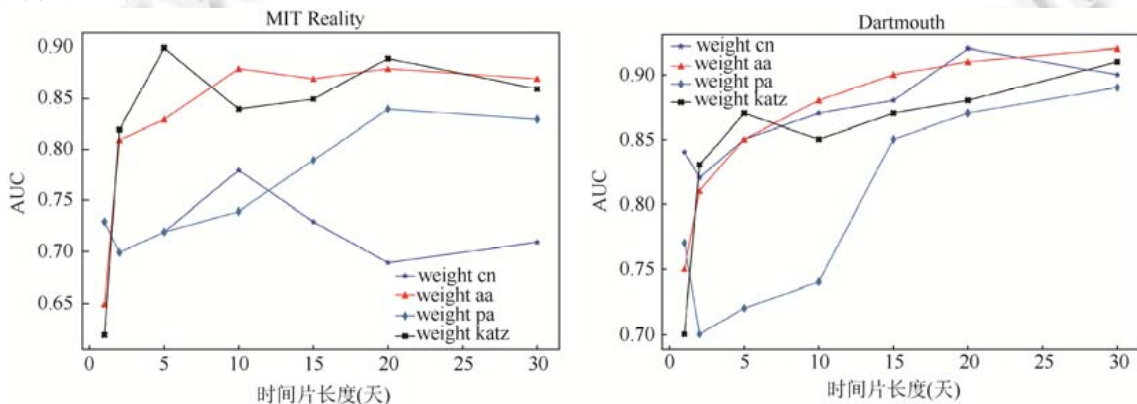
节点社交行为相似性指标结合了关系权重和传统相似性指标, 本文对传统的 CN、AA、PA 及 Katz 指标进行加权, 比较加权前后相似性指标的预测效果. 表 2 是对传统相似性指标加权后的表示形式.

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.

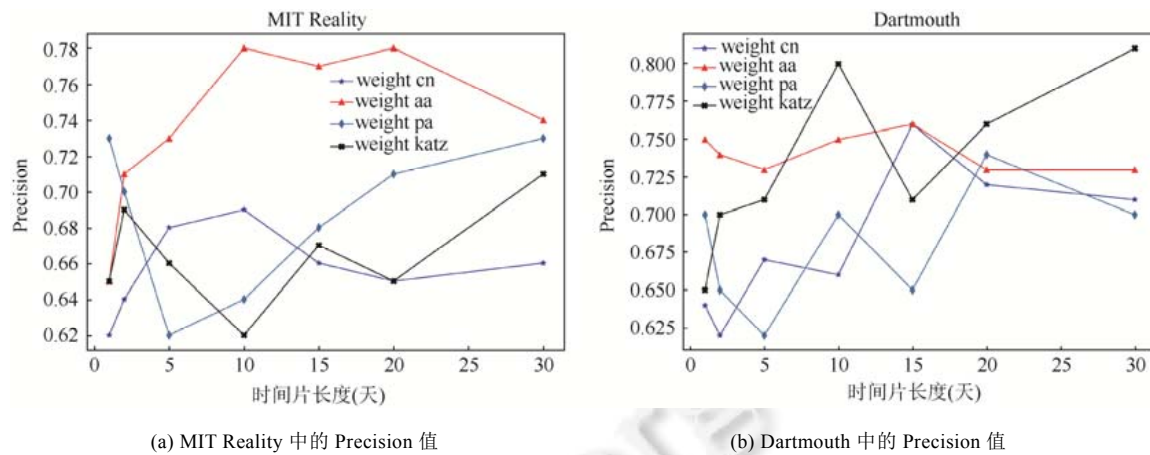


Fig. 11 Precision comparison of weighted similarity indicators at different time slice lengths

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 σ in the MIT Reality data set

表 3 MIT Reality 数据集中不同 σ 下的 AUC 以及 Precision

σ	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 σ in the Dartmouth data set

表 4 Dartmouth 数据集中不同 σ 下的 AUC 以及 Precision

σ	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 σ 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 σ 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 α 和 β ,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.

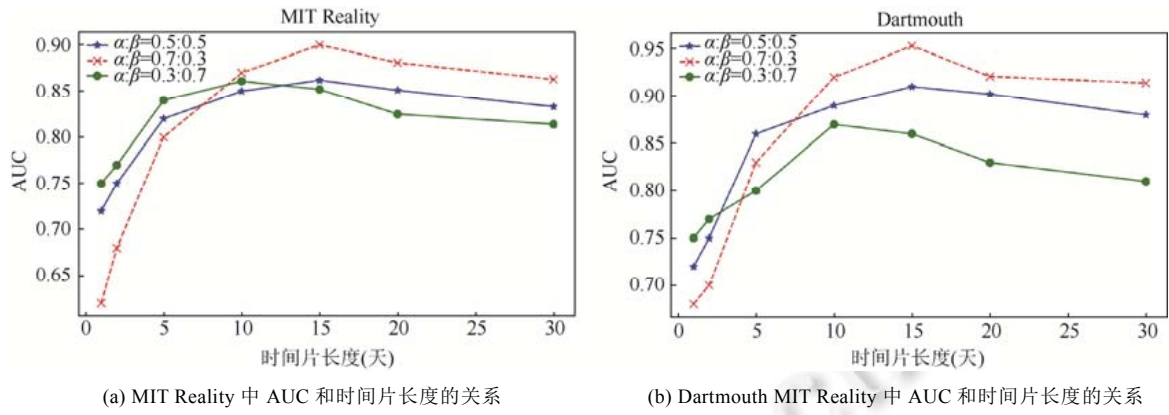


Fig.12 The AUC value of different ratios of α to β at different time slice lengths

图 12 不同 α 与 β 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 σ 取 0.7,实验结果如图 13 和表 5 所示.

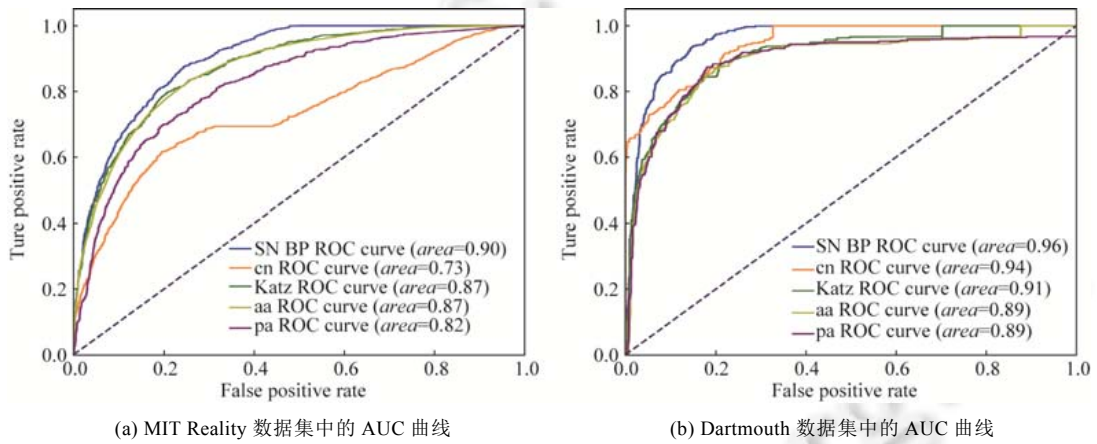


Fig.13 ROC graph of SNBP and traditional similarity indicators

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	0.71
Dartmouth	0.73	0.72	0.69	0.74	0.80

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.