

一种关注消息时效性的机会社会网络中 节点传播能力分析模型

蔡青松^{1,4}, 刘 燕², 牛建伟³, 孙利民⁴

(1. 北京工商大学计算机与信息工程学院, 北京 100048;
2. 北京大学软件与微电子学院, 北京 102600; 3. 北京航空航天大学计算机学院, 北京 100191;
4. 中国科学院信息工程研究所物联网信息安全技术北京市重点实验室, 北京 100093)

摘 要: 当前评估节点传播能力的工作大多针对静态网络. 本文采用演化图模型对机会社会网络进行刻画, 通过将传统通路(walk)的概念和静态图中的 Katz 中心性度量扩展到动态网络中, 提出了一种基于历史相遇记录评估节点消息分发能力的方法. 进一步, 针对消息的时效性特性, 本文提出了消息随时间推移效用递减的节点传播能力分析模型, 该模型考虑了消息所有可能经历的空间和时间通路, 并沿时间方向向下加权以描述消息时效性递减效应, 可用于有效计算和预测节点的消息转发能力. 本文的结论通过真实数据得到了验证.

关键词: 机会社会网络; Katz 中心性; 动态通路; 社会相遇记录

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2015)09-1705-09

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2015.09.005

An Analytical Model for Ranking the Candidates of Disseminating Time-Bound Messages in Opportunistic Mobile Social Networks

CAI Qing-song^{1,4}, LIU Yan², NIU Jian-wei³, SUN Li-min⁴

(1. School of Computer Science and Information Engineering, Beijing Technology and Business University, Beijing 100048, China;

2. School of Software and Microelectronics, Peking University, Beijing 102600, China;

3. School of Computer Science and Engineering, Beihang University, Beijing 100191, China;

4. Beijing Key Lab of IoT Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China)

Abstract: Traditional methods for evaluating node importance in sustaining the overall network topology or information dissemination are widely studied, while most of them are not applicable to dynamic settings where connections among nodes change frequently over time. This paper treats an opportunistic mobile social network as a time evolved, dynamic graph and proposes an effective scheme to calculate the relative nodal dissemination capability based on the contact history. In particular, we analyze the node importance in forwarding messages in more general settings where messages are time-dependent and become less important or out of date over time. To this end, we take a dynamic walk counting approach to calculate all possible temporal-spatial routes from a node to any other node by using a method of down-weighting of length. Since the age of a message increases with time, the old walks are discounted to represent the fading influence on the destination nodes. Experiments are conducted based on 4 real-world trace datasets, and the results show that our analytical result is effective at ranking the capabilities of nodes in disseminating or receiving the time-dependent messages.

Key words: opportunistic mobile social networks; Katz centrality; dynamic walk; social contact trace

1 引言

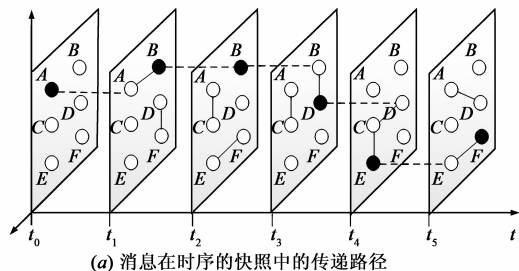
随着技术的进步, 日益普及的便携智能移动终端已经演化为具有丰富计算、感知、通信和存储能力的移动

计算平台, 应用领域随之出现了一些利用这些智能终端的短距离通信接口 (WiFi/Bluetooth/ZigBee 等) 以 P2P (Peer-to-Peer) 方式交换信息的需求 (如移动 P2P 社交), 使得利用这些移动设备进行动态组网的需求日趋迫

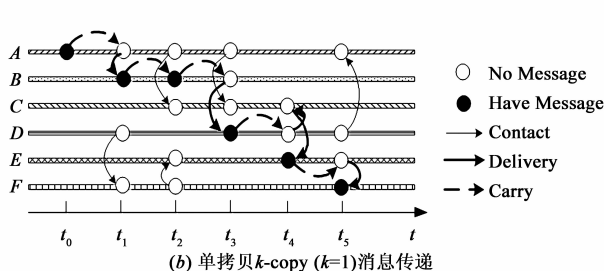
切.机会社会网络^[1] (OMSN, Opportunistic Mobile Social Networks)是借助人的社会性和移动性,利用智能设备的短距离通信接口感知其它设备并以自组织方式交换信息的新型组网技术.基于移动节点的相遇机会进行组网的机会计算技术^[2]在移动 P2P 社交、信息传播、城市规划、社会关系挖掘和分析等领域有广泛的潜在应用价值.

OMSN 是典型的自组织动态演化网络,网络中的个体因其不同的社会属性而具有不同的消息分发和获取能力.在实际中,消息通常都具有时效性(如新闻事件、新产品广告等),消息的效用往往随时间推移呈递减趋势.如何刻画和评估这类网络中节点对最新消息的获取和分发能力?对于给定的历史相遇信息,如何预测哪些节点可获得最新消息并以最快的速度在网络中传播?上述问题的解决对于挖掘 OMSN 中节点的社会属性和社会关系,以及研究社会环境中的消息传播与抑制机制,具有重要意义.

分析 OMSN 中节点的传播能力需要准确刻画这类



(a) 消息在时序的快照中的传递路径



(b) 单拷贝 k -copy ($k=1$) 消息传递

图1 OMSN中从节点A到节点F的消息时空传递路径示意

2 相关工作与现状

虽然许多研究工作^[4,5]提出了不同的机会消息分发算法并研究了人类移动性对转发算法性能的影响^[6],人类复杂的移动性规律使预测网络的拓扑演化和研究网络中节点的消息传播能力成为一类挑战性的问题.随着复杂网络研究的兴起,基于随机图(RG, Random Graph)^[7]的动态网络建模方法日益受到关注.然而随机图样本之间的时间无关性和节点对之间连边的独立性使得 RG 模型无法直接应用于 OMSN.时间演化图^[8](TEG, Time Evolving Graph)或时变图(TG, Temporal Graph)^[9]是一种关注时空尺度的动态网络模型,通过将动态网络刻画为离散时间序列上的快照序列,该模型可在相邻快照间研究拓扑的演化特性而在快照内研究节点的空间拓扑关系.一些研究工作针对静态网络中节点的中心化程度进行刻画,提出了节点的度中心性(DC, Degree Centrality),介数中心性(BC, Betweenness Centrality),紧密度中心性(CC, Closeness Centrality),Katz 中心性(KC, Katz Centrality)^[10]等度量,通过这些度量方法

网络的动态性.当前,一种常用方法是利用时间演化图^[3]模型,即将动态网络刻画为一系列沿时间方向变化的快照.该模型不仅可刻画移动节点间的瞬时空间拓扑关系,也可在时间上刻画相邻拓扑随时间的演变.图1示例了一个具有6个移动节点的时间演化图中消息从节点A沿时空传递路径传递到节点F的过程.由于 OMSN 的频繁断链和割裂特性,且移动节点的资源相对有限(电池能量、内存和缓冲区大小等),为使时效性消息在失效之前被传递给尽可能多的其它节点,消息的传递轨迹不再局限于沿传统的“路径”或“最短路径”进行,而可能多次经过一些相同的节点.在传统图论中,这样的消息传递轨迹被称为通路(walk).

本文将静态图中的通路和节点的 Katz 中心性度量扩展到 OMSN 中,针对消息的时效性,提出了一种计算节点相对传播能力的分析模型.实验结果表明,该模型不仅可针对不同的时效性消息快速计算各节点在网络中的消息传播和获取能力排名,而且可用于有效预测节点未来的消息传播和获取能力.

来刻画节点的重要程度.

研究工作^[11]通过对实验 Trace 数据进行经验分析,揭示了不同节点消息传播能力的差异,但并未提出一种有效的量化模型.研究工作^[12]在 KC 度量的基础上针对社交网络提出了一种计算节点传播能力的方法,并利用 Email 网络和电话呼叫网络数据进行了验证.研究工作^[13]利用时间近邻中心度(TCC, Temporal Closeness Centrality)研究了人类相遇网络中对关键节点免疫以阻止有害信息的传播问题.研究工作^[14]针对 Email 网络研究了信息传播效应随时间递减条件下节点传播能力的刻画方法.研究工作^[15]通过节点的社会相邻关系计算节点的重要性作为下一跳转发节点的选择依据,从而达到网络的性能优化.

不同于上述研究工作,在 OMSN 中,受社会性驱动力的节点相遇以及节点频繁的社会活动使得 OMSN 具有更为显著的动态性.本文工作专注于利用消息通路的概念研究 OMSN 中节点对时效性消息传播能力的度量方法.

3 模型

3.1 基本定义

令 $G = (V, E)$ ($|V| = n, n \in \mathbb{Z}^+, E \subseteq C_n^2$) 表示一个无向无权图, 其邻接矩阵为 $A = (a_{i,j}) (i, j \in V)$. \mathbb{Z}^+ 表示正整数, \mathbb{Z} 为整数集合, C 表示组合, C_n^2 表示从 n 个节点中任取 2 个, 若节点对 (i, j) 之间存在边, 则 $a_{i,j} = 1$, 否则 $a_{i,j} = 0$. 特别的, 令 $a_{i,i} = 0$ 表示图中无自环.

定义 1(路径) 图 G 中任意节点对 (i, j) 之间长度为 l 的路径(path)定义为节点 i 与节点 j 之间不同节点构成的有序集 $i, k_1, k_2, \dots, k_{l-1}, j$, 且满足:

$$a_{i,k_1} \neq 0, a_{k_1,k_2} \neq 0, \dots, a_{k_{l-2},k_{l-1}} \neq 0, a_{k_{l-1},j} \neq 0 \quad (1)$$

邻接矩阵 A 直观地表示了节点对 (i, j) 之间是否存在长度为 1 的路径 ($a_{i,j} = 1$). 由于 $(A^2)_{i,j} = \sum_{k=1}^n a_{i,k} a_{k,j}$, $(A^2)_{i,j}$ 可表示节点对 (i, j) 之间长度为 2 的路径条数. 对于任意正整数 $m (m > 2)$, 由于

$$(A^m)_{i,j} = \sum_{k_1=1}^n \sum_{k_2=1}^n \dots \sum_{k_{m-2}=k_{m-1}=1}^n \sum_{k_{m-1}=1}^n a_{i,k_1} a_{k_1,k_2} \dots a_{k_{m-2},k_{m-1}} a_{k_{m-1},j} \quad (2)$$

可见 $(A^m)_{i,j}$ 计算了从节点 i 出发经 m 个节点 $(k_1, k_2, \dots, k_{m-1}, k_m)$ 所有可能到达节点 j 的“组合路径”的数量, 这些“路径”可能多次经过同一个节点. 在传统图论中这类“路径”通常称为通路(walk), 下面给出通路的一般定义.

定义 2(通路) 令 $V = \{v_0, v_1, \dots, v_n\}$ 为图 G 的节点集, $E = \{e_1, e_2, \dots, e_m\}$ 为 V 上的边集, 节点与边的交替序列 $v_0 e_1 v_2 e_2 \dots e_k v_k$ 称为从节点 v_0 到节点 v_k 的长度为 k 的一条通路(walk), 其中 $e_i = (v_{i-1}, v_i)$.

显然, 定义 2 允许一条通路中多次出现同一个节点或同一条边, 且 v_k 和 v_0 可以是相同节点, 此时称该通路为闭合通路(close walk). 根据式(2), 有如下引理.

引理 1 若图 $G = (V, E)$ 的邻接矩阵为 A , 则 $(A^m)_{i,j} (m \in \mathbb{Z}^+)$ 为节点对 (i, j) 之间长度为 m 的通路数, $(A^m)_{i,i}$ 为长度为 m 闭合通路数.

引理 1 给出了计算从节点 i 到节点 j 之间的长度为 m 的 walk 数量的方法.

3.2 静态网络中的节点传播能力度量

Katz 中心性通过计算节点 $v_i \in V$ 的第 k 阶邻居数量来计算其在网络中的相对重要性程度, 其中第 k 阶邻居对节点 v_i 的“影响”以 $\alpha^k (0 < \alpha < 1, \alpha \in \mathbb{R}^+)$ 递减. \mathbb{R}^+ 表示正实数集, 令 K 表示图 G 的 Katz 中心性矩阵, 根据引理 1, 由任意节点出发到达所有其他节点长度为 $w (w \in \mathbb{Z}^+)$ 的带权通路总数为

$$K = \sum_{k=1}^w \alpha^k A^k \quad (3)$$

当 $\alpha < 1/\rho(A)$ ($\rho(A)$ 为矩阵 A 的谱半径) 且 $w \rightarrow \infty$ 时上式收敛为

$$K = \lim_{w \rightarrow \infty} (\alpha A + \alpha^2 A^2 + \dots + \alpha^w A^w) = (I - \alpha A)^{-1} - I \quad (4)$$

节点的重要性向量计算为

$$k = \sum_{j=1}^n [(I - \alpha A)^{-1} - I] = [(I - \alpha A)^{-1} - I] \mathbf{1} \quad (5)$$

其中 I 为单位矩阵, $\mathbf{1} = (1, 1, \dots, 1)^T$, 节点 i 的传播能力为 k_i . 在实际网络中, 因子 α 可解释为消息沿不可靠链路成功传输到下一节点的概率.

3.3 动态网络中的节点传播能力度量

为将 Katz 中心性度量方法推广到一般动态网络中, 本文需借助时间演化图对 OMSN 建模, 为此先给出了时间演化图的一般定义.

定义 3(时间演化图) 对于任意节点集 $|V| = n (n \in \mathbb{Z}^+)$, 令 $T \subseteq [T_s, T_e]$ 为定义在起始时间 T_s 和终止时间 T_e 上的任意时间集合, $\Delta t = (t_i, t_{i+1}] \in T (i \in \mathbb{Z}^+)$ 为时间间隔, 若令 $G_i = (V, E_i) (E_i \subseteq C_n^2)$ 为 Δt 上观察到的子图, 则 $g := \{G_i\}$ 为定义在 $[T_s, T_e]$ 上的时间演化图.

从定义 3 看出, 演化图 $g := \{G_1, G_2, \dots, G_i, \dots\}$ 由一系列沿时间方向构成的静态快照组成, 其对应的邻接矩阵序列为 $\{A_1, A_2, \dots, A_i, \dots\}$. 为利用 Katz 中心性度量方法, 需要将通路的概念引入到动态图中, 为此给出动态通路的定义.

定义 4(动态通路) 定义在一个非递减的离散时间序列 $t_{r_1} \leq t_{r_2} \leq \dots \leq t_{r_m} (r_m, w \in \mathbb{Z}^+, t_{r_m} \in T)$ 上的边序列 $v_1 \rightarrow v_2, v_2 \rightarrow v_3, \dots, v_w \rightarrow v_{w+1}$ 构成从节点 v_1 到节点 v_{w+1} 的长度为 w 的动态通路, 当且仅当第 r_m 个快照的邻接矩阵满足 $(A_{r_m})_{v_m, v_{m+1}} \neq 0$.

定义 4 表明, 约束条件 $(A_{r_m})_{v_m, v_{m+1}} \neq 0$ 要求自节点 v_1 出发的动态通路在第 r_m 个快照上至少“前进一步”. 在 OMSN 中, 从节点 v_i 出发到达其它节点的动态通路由如下两种“通路片段”构成: (1) 节点 v_i 在某一快照中与其它节点形成静态通路(拓扑相邻), 这部分称为空间通路; (2) 节点 v_i 到其他节点沿时间方向形成通路, 该通路可能穿越若干个快照, 称为时间通路. 对于情形 (1), 可按照引理 1 计算这些通路数量, 对于 (2) 则不能直接使用引理 1 计算. 在 OMSN 中, 为满足定义 4 的约束条件, 当持有消息的节点无法在当前快照将消息转发时, 允许节点将消息转发给自身以表示“携带”. 为此, 需将式(4)加上一个单位矩阵 I 以表示“携带”状态, 式(4)改写为

$$\lim_{w \rightarrow \infty} (I + \alpha A + \alpha^2 A^2 + \dots + \alpha^w A^w) = (I - \alpha A)^{-1} \quad (6)$$

因此, OMSN 中的任意节点对在快照序 $G_1, G_2, \dots, G_i, \dots, G_k$ 中形成的长度为 w 的动态通路总数为

$$\alpha^w \mathbf{A}_1^{m_1} \mathbf{A}_2^{m_2} \cdots \mathbf{A}_i^{m_i} \cdots \mathbf{A}_k^{m_k} \quad (7)$$

其中, $m_1 + m_2 + \dots + m_k = w (m_i \in \mathbb{Z}^+)$. 式(7)直观地表明, 长度为 w 的动态通路位于快照 G_i 中的“一段”为 $\mathbf{A}_i^{m_i}$, 且在快照 G_i 中第一跳的权值为 $\alpha^{m_1 + m_2 + \dots + m_{i-1} + 1}$. 计算所有加权动态通路, 可得传播能力矩阵 \mathbf{D}^k 为

$$\begin{aligned} \mathbf{D}^k &= \sum_{w=0}^{\infty} \alpha^{m_1 + m_2 + \dots + m_k} \mathbf{A}_1^{m_1} \mathbf{A}_2^{m_2} \cdots \mathbf{A}_k^{m_k} \\ &= \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \cdots \sum_{m_k=0}^{\infty} \alpha^{m_1 + m_2 + \dots + m_k} \mathbf{A}_1^{m_1} \mathbf{A}_2^{m_2} \cdots \mathbf{A}_k^{m_k} \\ &= \sum_{m_1=0}^{\infty} \alpha^{m_1} \mathbf{A}_1^{m_1} \sum_{m_2=0}^{\infty} \alpha^{m_2} \mathbf{A}_2^{m_2} \cdots \sum_{m_k=0}^{\infty} \alpha^{m_k} \mathbf{A}_k^{m_k} \\ &= (\mathbf{I} - \alpha \mathbf{A}_1)^{-1} (\mathbf{I} - \alpha \mathbf{A}_2)^{-1} \cdots (\mathbf{I} - \alpha \mathbf{A}_k)^{-1} \\ &= \mathbf{D}^{k-1} (\mathbf{I} - \alpha \mathbf{A}_k)^{-1} \end{aligned} \quad (8)$$

其中 $\mathbf{D}^0 = \mathbf{I}$, 节点 v_i 在动态网络中的传播能力为 $\mathbf{D}^k \mathbf{1}_i$.

3.4 OMSN 中节点的时效性消息传播能力度量

式(8)通过计算所有的加权动态通路数量来刻画节点的传播能力, 然而, 式(8)将发生在任意时间区间内长度相等的动态通路对源节点的传播能力贡献度视为等同, 这在实际应用中存在局限性. 在实际中, 消息往往具有时效性, 即使消息传递的通路长度相同, 但当前或最近发生的事件往往重要性更高, 或人们更重视“最近”发生的事件或“最新的”消息, 而不是传播了相同的通路长度的很久之前发生的事件或传播的新闻. 因此, 需要针对消息的时效性对上述模型进行修正.

设消息 M 具有有限的生存期且其当前年龄为 T_{age} , 为考虑消息的时效性, 一种直观的方法是将消息的效用按“年龄”以负指数 $e^{-\beta T_{\text{age}}} (\beta \in \mathbb{R}^+)$ 递减. 为简化分析, 假设动态通路在每个快照中“至多走一步”, 式(8)可简化为

$$\begin{aligned} \mathbf{D}^1 &= \mathbf{I} + \alpha \mathbf{A}_1 \\ \mathbf{D}^k &= \mathbf{D}^{k-1} (\mathbf{I} + \alpha \mathbf{A}_k) \end{aligned} \quad (9)$$

其中 \mathbf{D}^k 和 \mathbf{D}^{k-1} 分别表示计算到第 k 和 $k-1$ 个快照时节点的传播能力度量矩阵. 类似的, 若令 \mathbf{H}^k 表示网络演化到 G_k 时节点对消息 M 的传播能力矩阵 (令 $\mathbf{H}^0 = \mathbf{0}$), 则 \mathbf{H}^{k-1} 表示网络演化到 G_{k-1} 时的传播能力矩阵. 由于 \mathbf{H}^{k-1} 计算了所有可能在快照序列 $\mathbf{A}_1, \dots, \mathbf{A}_{k-1}$ 中任意一个产生, 并在同一快照或其后续任一快照中结束的所有加权动态通路, 因此有如下推论.

推论 1 给定 TEG $g := \{G_1, G_2, \dots, G_i, \dots, G_k\} (i < k \in \mathbb{Z}^+, k \rightarrow \infty)$ 和对应的快照邻接矩阵序列 $\{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_i, \dots, \mathbf{A}_k\}$, 令 G_k 为当前快照, ΔT_{age} 为距上一次计算动态通路以来的年龄, α, β 分别为动态通路递减因子和消息效用递减因子, 若消息的效用按 $e^{-\beta T_{\text{age}}}$ 递减,

则有

$$\mathbf{H}^k = (\mathbf{I} + e^{-\beta \Delta T_{\text{age}}} \mathbf{H}^{k-1}) (\mathbf{I} - \alpha \mathbf{A}_k)^{-1} - \mathbf{I} \quad (10)$$

证明 为简单起见, 证明过程先沿用式(9)的假定, 即动态通路在每个快照中“至多走一步” ($m_i \leq 1$), 然后我们再归纳到 $m_i \rightarrow \infty$ 的情形.

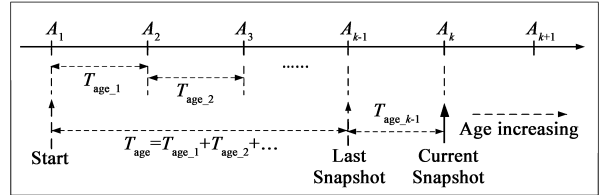


图2 消息当前年龄与当前快照的关系

图2 示例了动态通路到达当前快照和消息当前年龄的关系. 其中, \mathbf{A}_i 为快照 G_i 的邻接矩阵, 起始快照为 \mathbf{A}_1 , 当前快照为 \mathbf{A}_k , 快照 \mathbf{A}_{k-1} 和 \mathbf{A}_k 的时间间隔为 $T_{\text{age}, k-1}$. 下面用归纳法对式(10)进行证明.

(1) 当 $k=1, 2$ 时

$k=1$ 时, 当前快照为 \mathbf{A}_1 . 动态通路在快照 \mathbf{A}_1 中至多“走了一步”, 其年龄为 0, 此时 $\mathbf{H}^1 = \alpha \mathbf{A}_1$.

$k=2$ 时, 当前快照为 \mathbf{A}_2 . 若动态通路在 \mathbf{A}_1 产生并在 \mathbf{A}_1 结束, 消息的当前年龄为 $T_{\text{age}, 1}$, 该通路对源节点传播能力的贡献为 $e^{-\beta T_{\text{age}, 1}} \alpha \mathbf{A}_1$; 若该通路在 \mathbf{A}_1 产生并在 \mathbf{A}_2 结束, 消息的当前年龄仍为 $T_{\text{age}, 1}$, 该通路的贡献为 $e^{-\beta T_{\text{age}, 1}} \alpha^2 \mathbf{A}_1 \mathbf{A}_2$; 若该通路在 \mathbf{A}_2 产生并在 \mathbf{A}_2 结束, 消息年龄为 0, 其贡献为 $\alpha \mathbf{A}_2$. 因此有

$$\begin{aligned} \mathbf{H}^2 &= \alpha \mathbf{A}_2 + e^{-\beta T_{\text{age}, 1}} \alpha \mathbf{A}_1 + e^{-\beta T_{\text{age}, 1}} \alpha^2 \mathbf{A}_1 \mathbf{A}_2 \\ &= (\mathbf{I} + e^{-\beta T_{\text{age}, 1}} \alpha \mathbf{A}_1) (\mathbf{I} + \alpha \mathbf{A}_2) - \mathbf{I} \\ &= (\mathbf{I} + e^{-\beta T_{\text{age}, 1}} \mathbf{H}^1) (\mathbf{I} + \alpha \mathbf{A}_2) - \mathbf{I} \end{aligned}$$

(2) 设式(10)在 $k=k'-1$ 时成立, 即

$$\mathbf{H}^{k'-1} = (\mathbf{I} + e^{-\beta \Delta T_{\text{age}}} \mathbf{H}^{k'-2}) (\mathbf{I} + \alpha \mathbf{A}_{k'-1}) - \mathbf{I}$$

(3) 考虑当 $k=k'$ 时的情形, 此时当前快照为 $\mathbf{A}_{k'}$. 从(1)可知, $\mathbf{H}^{k'-1}$ 计算了所有在 $\mathbf{A}_{k'}$ 之前任意快照产生, 并在同一或其后续任意快照中结束的所有动态通路. 这些通路分三种情况: ① 在当前快照产生并结束的动态通路; ② 所有在 $\mathbf{A}_1, \dots, \mathbf{A}_{k'-1}$ 中产生, 至多在 $\mathbf{A}_{k'-1}$ 结束的动态通路; ③ 所有在 $\mathbf{A}_1, \dots, \mathbf{A}_{k'-1}$ 产生, 并在当前快照 $\mathbf{A}_{k'}$ 继续“前进一步”的动态通路.

对于①这部分动态通路的贡献计算为 $\alpha \mathbf{A}_{k'}$.

对于②根据归纳假定, 这部分动态通路的贡献为 $\mathbf{H}^{k'-1}$, 但其年龄增加了 $T_{\text{age}, k-1}$, 因此贡献度为 $e^{-\beta T_{\text{age}, k-1}} \mathbf{H}^{k'-1}$.

对于③这部分动态通路在②的基础上又“前进一步”, 即其贡献度为 $\mathbf{H}^{k'-1} \alpha \mathbf{A}_{k'}$, 其年龄增加了 $T_{\text{age}, k-1}$, 因此其贡献度为 $e^{-\beta T_{\text{age}, k-1}} \mathbf{H}^{k'-1} \alpha \mathbf{A}_{k'}$.

因此有

$$\begin{aligned} \mathbf{H}^k &= \alpha \mathbf{A}_{k'} + e^{\beta T_{\text{app}, k'-1}} \mathbf{H}^{k'-1} + e^{-\beta T_{\text{app}, k'-1}} \mathbf{H}^{k'-1} \alpha \mathbf{A}_{k'} \\ &= (\mathbf{I} + e^{-\beta T_{\text{app}, k'-1}} \mathbf{H}^{k'-1}) (\mathbf{I} + \alpha \mathbf{A}_{k'}) - \mathbf{I} \end{aligned}$$

释放式(9)的假定,即允许动态通路在任意快照 G_k 中

至多可走 $m_i (m_i \rightarrow \infty)$ 步,将式 $\mathbf{I} + \alpha \mathbf{A}_{k'} + \alpha^2 \mathbf{A}_{k'}^2 + \alpha^3 \mathbf{A}_{k'}^3 + \dots + \alpha^m \mathbf{A}_{k'}^m = (\mathbf{I} - \alpha \mathbf{A}_{k'})^{-1}$ 替代上式中的 $\mathbf{I} + \alpha \mathbf{A}_{k'}$ 项,即得

$$\mathbf{H}^k = (\mathbf{I} + e^{-\beta \Delta T_{\text{app}}} \mathbf{H}^{k'-1}) (\mathbf{I} - \alpha \mathbf{A}_{k'})^{-1} - \mathbf{I}$$

证毕.

在实际应用中,列向量 $\mathbf{H}^k \mathbf{1}$ 和 $(\mathbf{H}^k)^T \mathbf{1}$ 分别可表示各节点传播和获取消息的能力.由于矩阵相乘运算的不可置换性以及 OMSN 中消息传递的时向性,有 $\mathbf{H}^k \mathbf{1} \neq (\mathbf{H}^k)^T \mathbf{1}$.

4 实验验证

4.1 数据集

本文采用 CRAWDDAD^[16] 中的 4 个不同场景下的真实 Trace 数据来验证本文提出的传播能力分析模型的有效性.分别是:(1)Reality Mining.该实验记录了 100 个携带 Nokia 6600 手机的用户自 2004 年 4 月至 2005 年 5 月间利用 Bluetooth 接口相遇的数据;(2)SigComm 2009.记录了 SigComm 2009 会议期间 100 个持 HTC s620 手机的用户通过 Bluetooth 接口相遇的数据;(3)Sassy.记录了自 2008 年 2 月起的 75 天内,圣安德鲁斯大学 27 名携带 T-Mote 设备的不同人员在校园中通过 Zigbee 接口相遇

的信息;(4)UCSD Topology Discovery.记录了 11 周内 275 名携带配置了 WiFi(802.11b)接口的 PDA 用户通过 AP 形成的间接相遇数据.表 1 列出了该 4 个经过预处理后的 Trace 数据集的摘要信息.

表 1 Trace 数据信息

数据集描述	Reality Mining	SigComm 2009	Sassy	UCSD
设备类型	Nokia6600	HTC s620	T-Mote	PDA phone
网络类型	Bluetooth	Bluetooth	ZigBee	WiFi
持续时间(天)	491	5	75	77
有效(总)用户数	96(100)	76(100)	25(27)	257(275)
有效记录数	47,109	14,418	11,225	101,471
相遇类型	Direct	Direct	Direct	Indirect

4.2 实验结果及分析

(1)不同节点在网络中的消息传播速度

首先考察无时效性约束时不同节点的消息传播效率.由于泛洪(flooding)方式在 OMSN 中具有良好的自适应性和冗余分发特性,可以准确地刻画基于通路的消息传递轨迹,因此本实验采用泛洪的方式考察节点的传播能力.为降低网络开销,实验采用 k -copy ($k \leq 2$) 的方式考察 $D^k \mathbf{1}$ 中排名最高、中间和排名最低的节点作为消息源时的消息覆盖速度.图 3 所示的结果表明,在 4 个数据集中 $D^k \mathbf{1}$ 中排名最高节点的传播能力均明显高于平均情况,而 $D^k \mathbf{1}$ 排名最低的节点则明显劣于平均情况.

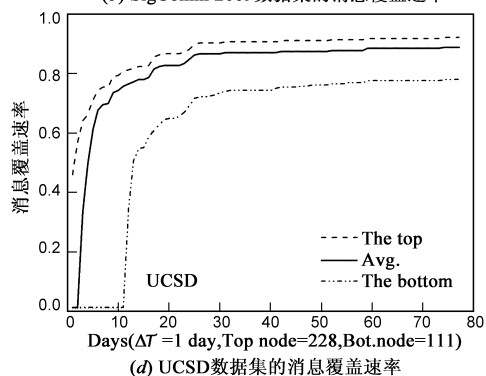
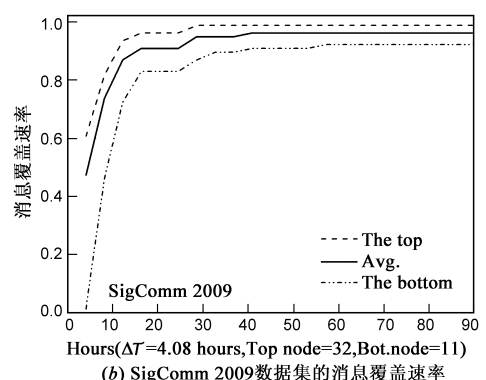
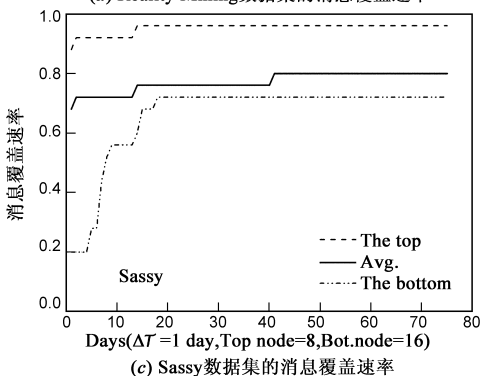
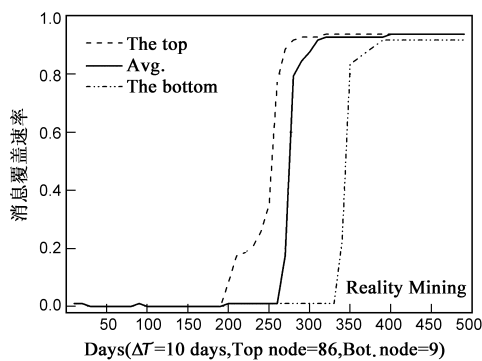


图 3 $D^k \mathbf{1}$ 中最高、平均和最低节点在 OMSN 中的传播速度

(2) 衰减因子 β 对 H^k 的影响

在式(10)中, $e^{-\beta\Delta T_{app}}$ 可解释为消息与年龄相关的效用因子. 当 $\beta \rightarrow 0$ 时 $e^{-\beta\Delta T_{app}} \rightarrow 1$, 此时 $H^k \rightarrow D^k - I$, 即消息随时间推移其效用不递减, 则节点转发该消息并为其传播能力增加贡献的可能性越大, 即 H^k 越大. 当 $\beta \rightarrow \infty$ 时 $e^{-\beta\Delta T_{app}} \rightarrow 0$, 此时 $H^k \rightarrow (I - \alpha A_k)^{-1} - I$, 即消息

至多在当前快照产生并在该快照结束, 之后节点转发该消息并为其传播能力增加贡献的可能性变小, 即 H^k 越小. 因此因子 β 的取值对 H^k 的计算结果有直接影响. 图 4 给出了 4 个数据集中随机选取的节点在 H^k 中的值随 β 取值的变化, 结果显示, 当 β 取值越大 H^k 值变化越不明显, 反之则变化显著.

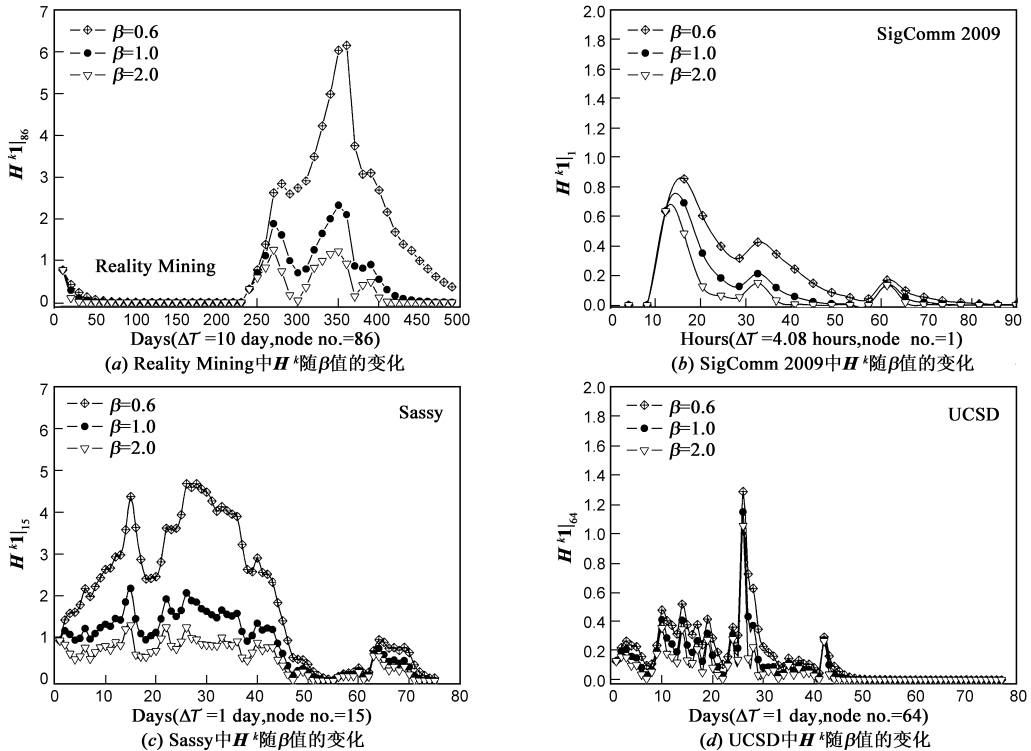


图4 各数据集中不同的 β 值对 H^k 的影响

(3) 不同 β 值对节点传播能力排名的影响

为考虑 β 取不同值时节点传播能力在网络中排名的变化情况, 我们分别在 4 个 Trace 计算的 $D^k \mathbf{1}$ 中排名前 3、居中和后 3 的节点中随机选取一个节点, 分别是 15 号节点, 46 号节点, 21 号节点和 24 号节点 (按表 1 的顺序). 分别对上述 4 个节点考察 β 的不同取值对其传播能力排名的影响. 图 5 的结果显示, β 取值越大, 节点随时间推移其传播能力排名变化剧烈程度增大, 而对于不考虑消息时效性的情形来说, 节点的排名则相对稳定 (纵轴表示传播能力低于当前节点的节点数占总节点的百分比).

(4) 不同 Δt 取值对节点最终排名的影响

通过定义 4 可知, 不同的快照间隔 Δt 将直接影响所计算的快照数量, 进而影响式(10)的计算结果和节点的传播能力排名. 事实上, OMSN 所固有的动态非均匀演化特性使得如何根据具体的网络场景设置合理的 Δt 目前仍是一个尚未解决的挑战性问题. 一般而言, 等间隔划分便于计算, 但过大或过小的 Δt 取值将导致不足或额外的计算, 导致不准确的传播能力排名结果; 非等

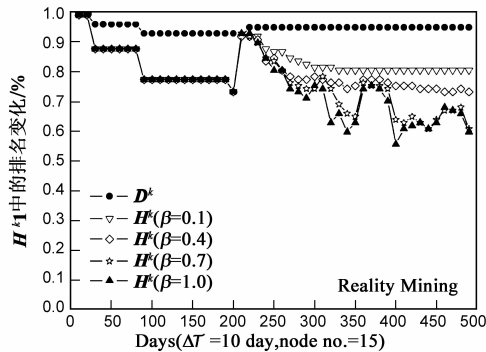
间隔的快照划分则跟实际网络场景紧密相关. 由于通常我们只关注节点基于历史信息“累积”至当前的相对传播能力, 因此本实验考察等间隔快照划分中不同的 Δt 取值对于节点传播能力排名的影响. 图 6 给出了随机选取的 4 个节点随 Δt 增大的传播能力排名变化. 可以看出, Δt 和 β 的取值均对节点的传播能力排名产生不同程度的影响, 呈现出非均匀的变化特性. 当取定 β 时, 4 个节点的排名均随 Δt 呈明显变化. 当选定 Δt 时, 4 个节点的排名也明显受 β 取值的影响.

(5) 节点传播和获取最新消息能力的预测

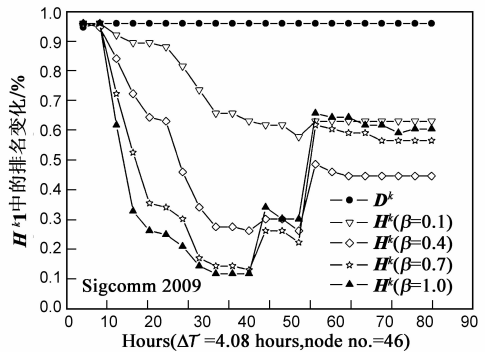
式(10)的迭代形式可以用来预测未来快照中节点的传播能力. 为考察预测的准确度, 本文利用皮尔逊相关系数 P_{corr} (Pearson Correlation Coefficient) 来考察 $H^{k+1} \mathbf{1}$ 和 $(I - \alpha A_{k+1})^{-1} \mathbf{1}$ 的相关性随 β 取值的变化. P_{corr} 越大则预测准确率越高, 反之则较低. 向量 $H^{k+1} \mathbf{1}$ 表示按式(10)预测的节点传播能力, 而向量 $(I - \alpha A_{k+1})^{-1} \mathbf{1}$ 则表示节点的当前实际传播能力. 图 7 给出了 β 取值增大时节点传播和获取最新消息能力的预测结果. 可以看出, β 取值

分别为 0.8, 1.0, 0.8 和 0.6 时, 4 个数据集中节点的传播能力预测准确度分别为 77.65%, 69.27%, 89.50% 和 79.79%, 达到了较高的准确度. 同样, 节点获取消息的预

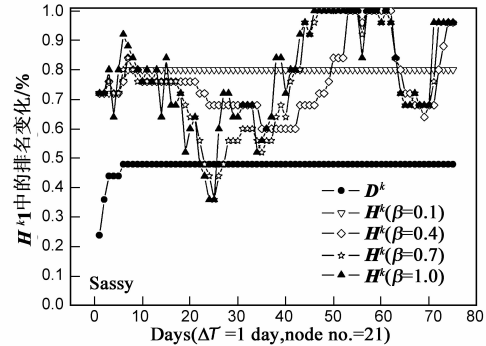
测也达到了较高的准确度. 图 7 也表明, 当 β 值增大时节点的消息传播和获取能力和获取能力均随 β 的取值增大趋向于稳定值, 这与实验(2)的分析结果是一致的.



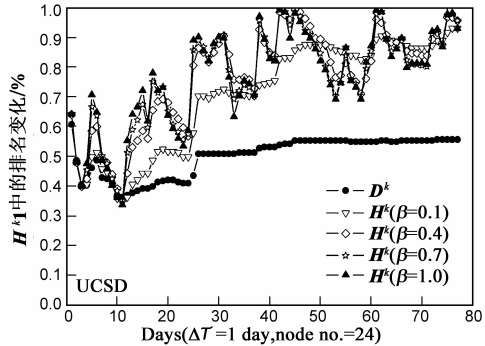
(a) Reality Mining中 β 值对节点排名的影响



(b) SigComm 2009中 β 值对节点排名的影响

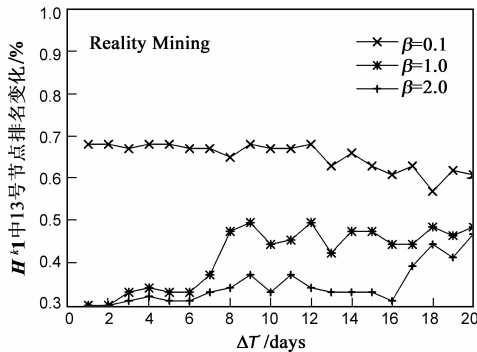


(c) Sassy中 β 值对节点排名的影响

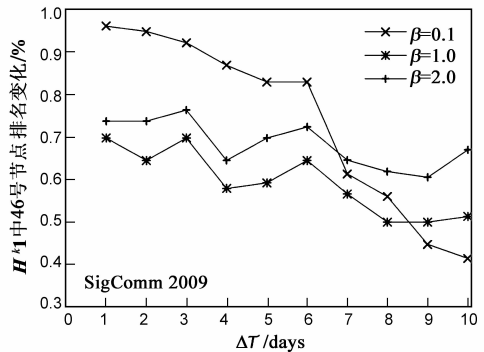


(d) UCSD中 β 值对节点排名的影响

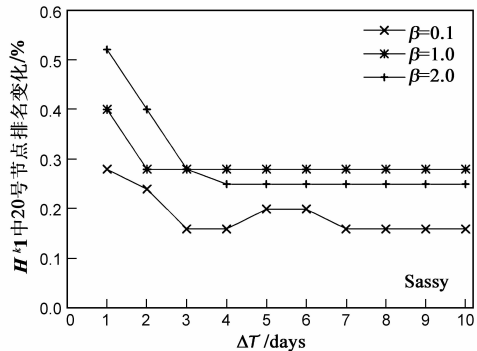
图5 不同 β 值随时间对节点传播能力排名的影响



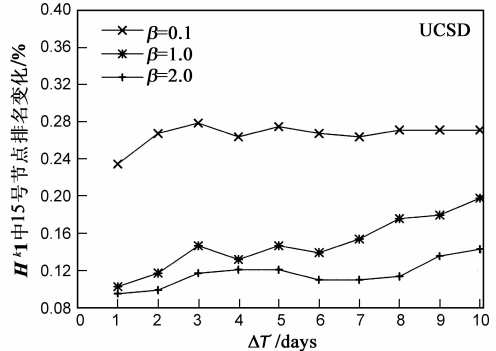
(a) Reality Mining中 Δt 取值对节点排名的影响



(b) SigComm 2009中 Δt 取值对节点排名的影响

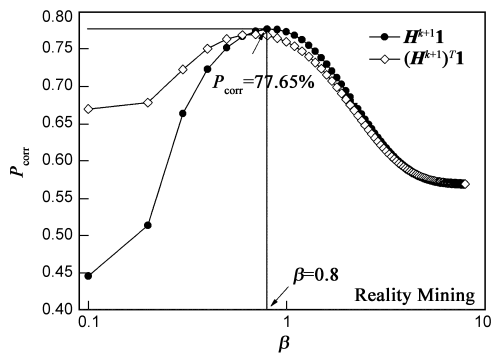
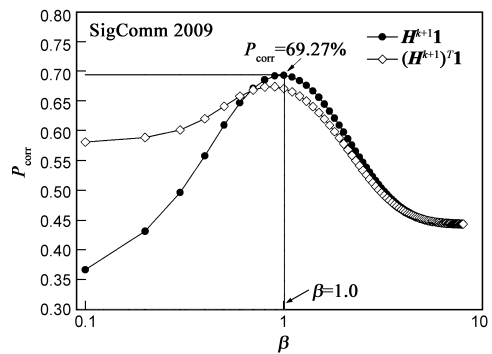
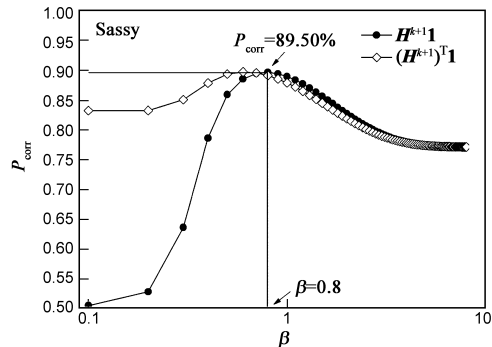
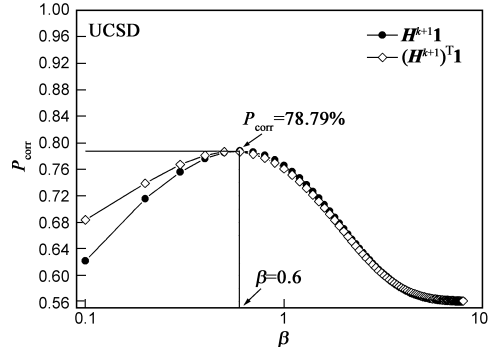


(c) Sassy中 Δt 取值对节点排名的影响



(d) UCSD中 Δt 取值对节点排名的影响

图6 快时间间隔 Δt 对节点传播能力排名的影响

(a) Reality Mining 的预测准确率与 β 的关系(b) SigComm 2009 的预测准确率与 β 的关系(c) Sassy 的预测准确率与 β 的关系(d) UCSD 的预测准确率与 β 的关系图7 取不同值时 $H^{k+1}1$ 和 $(H^{k+1})^T 1$ 预测的准确性

5 结论及进一步工作

本文将静态图中的通路和 Katz 中心性度量扩展到 OMSN 中,提出了一种关注消息时效性的节点传播能力评估分析模型.模型计算了 OMSN 中所有的时间和空间通路,通过对“长”通路向下加权和对消息效用按时间指数递减,给出了关注消息时效性的节点传播能力分析模型.该模型的迭代计算形式不仅便于用来沿时间方向“逐步”快速计算,同时也可用于对节点未来的消息传播和获取能力进行准确预测.

基于本文工作,下一步我们将重点研究基于相遇历史信息的节点间社会关系挖掘方法,研究节点间社会关系演化、预测及消息的定向传播机制.

参考文献

[1] Vastardis N, Kun Y. Mobile social networks: architecture, social properties and key research challenges[J]. IEEE Communications Surveys and Tutorials, 2013, 15(3): 1355 – 1371.

[2] Conti M, Kumar M. Opportunities in opportunistic computing [J]. IEEE Computer, 2010, 43(1): 42 – 50.

[3] Sun J, Faloutsos C, Papadimitriou S, et al. Graphscope: parameter-free mining of large time-evolving graphs[A]. Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [C]. San Jose, CA, USA: ACM, 2007. 687 – 696.

[4] Pelusi L, Passarella A, Conti M. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks[J]. IEEE Communications Magazine, 2006, 44(11): 134 – 141.

[5] Conti M, Crowcroft J, Giordano S, et al. Middleware for Network Eccentric and Mobile Applications[M]. Berlin Heidelberg: Springer, 2009. 121 – 147.

[6] Chaintreau A, Hui P, Crowcroft J, et al. Impact of human mobility on opportunistic forwarding algorithms[J]. IEEE Transactions on Mobile Computing, 2007, 6(6): 606 – 620.

[7] Newman M E J. Handbook of Graphs and Networks: From the Genome to the Internet[M]. New York, NY, USA: Wiley-VCH Verlag GmbH & Co KGaA, 2005. 35 – 68.

[8] Rossi R A, Gallagher B, Neville J, et al. Modeling dynamic behavior in large evolving graphs[A]. Proceedings of the 6th ACM International Conference on Web Search and Data Mining [C]. Dallas, TX, USA: ACM, 2013. 667 – 676.

[9] Holme P, Saramäki J. Temporal networks[J]. Physics reports, 2012, 519(3): 97 – 125.

[10] Newman M. Networks: An Introduction[M]. Oxford: Oxford University Press, 2010.

[11] Yoneki E, Hui P, Crowcroft J. Bio-Inspired Computing and Communication[M]. Berlin Heidelberg: Springer, 2008. 116 – 132.

[12] Grindrod P, Parsons M C, Higham D J, et al. Communicability across evolving networks[J]. Physical Review E, 2011, 83(4): 046120.

- [13] Tang J, Mascolo C, Musolesi M, et al. Exploiting temporal complex network metrics in mobile malware containment [A]. 2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM) [C]. Washington, DC: IEEE Computer Society, 2011. 1 - 9.
- [14] Grindrod P, Higham D J. A matrix iteration for dynamic network summaries[J]. SIAM Review, 2013, 55(1): 118 - 128.
- [15] Mtibaa A, May M, Diot C, et al. PeopleRank: Social opportunistic forwarding[A]. Proceeding of the 29th IEEE Conference on Computer Communications (INFOCOM 2010) [C]. San Diego, CA USA; IEEE Press Piscataway, NJ, USA, 2010. 1 - 5.
- [16] A Community Resource for Archiving Wireless Data [EB/OL]. <http://crawdad.cs.dartmouth.edu/>, 2014-6-10.

作者简介



蔡青松 男, 1973 年生人, 博士, 副教授. 现任职北京工商大学计算机与信息工程学院, 主要研究领域包括: 移动自组网和无线传感器网络, DTN/机会网络, 移动社会网络等.

E-mail: simoncaiqs@gmail.com



刘燕 (通信作者) 女, 1971 年生人, 博士, 副教授. 现任职北京大学软件与微电子学院, 主要研究领域包括: 软件工程, 计算机网络.

E-mail: ly@ss.pku.edu.cn