

基于泛洪算法与亲疏度的社会关系检索分析研究

王 月¹, 滕桂法², 王春山¹, 张 璠¹

(1 河北农业大学 信息科学与技术学院, 河北 保定 071001; 2 河北农业大学 研究生学院, 河北 保定 071001)

摘 要: 现有的社会关系检索算法大多忽视人与人之间的亲疏度, 检索最短路径或最优路径, 导致检索结果不能很好地满足人们的要求. 针对此类问题, 创新性地提出一种基于泛洪算法与亲疏度的检索算法(flooding algorithm and intimacy search, FIS), 可以检索到源节点到目标节点的所有路径, 并能根据亲疏度大小将路径排序. 算法将亲疏度映射为边权值, 采用邻接表存储社会关系图, 利用泛洪算法进行路径检索. 仿真实验结果表明, 基于泛洪算法与亲疏度的社会关系检索算法综合性能较好, 适用性强. 同时也发现, 关系路径长度小的路径, 亲疏度未必大; 相反, 关系路径长度大的路径, 亲疏度未必小.

关键词: 泛洪算法; 亲疏度; 社会关系网络; 关系路径

中图分类号: TP312

文献标识码: A

文章编号: 1000—7180(2016)10—0102—04

Study of Social Relations Search Based on Flooding Algorithm and Intimacy

WANG Yue¹, TENG Gui-fa², WANG Chun-shan¹, ZHANG Fan¹

(1 College of Information Science and Technology, Agricultural University of Hebei, Baoding 071001, China;

2 College of Postgraduate, Agricultural University of Hebei, Baoding 071001, China)

Abstract: An innovative algorithm based on flooding algorithm and intimacy in social relations search is put forward on the problem that intimacy is ignored and only can search the shortest or the optimal path. All paths from source node to destination node can be searched in the algorithm. What's more, paths are ordered by intimacy. The intimacy is mapped to edge weight. Social relational graph is stored in adjacency list. Flooding algorithm is used to search paths. The simulation results show that this algorithm is performed with high comprehensive and strong adaptability. At the same time it is showed that the path with short path length, is not necessarily has larger intimacy. On the contrary, the path with long path length, is not necessarily has smaller intimacy.

Key words: flooding algorithm; intimacy; social relations network; relation paths

1 引言

社会关系检索的好坏直接关系到社会关系网络的质量和发展的. 社会关系检索技术已经成为当今国内外研究的一个热点问题^[1-3]. 最短路径 Dijkstra 算法的时间和空间复杂度都是 $O(n^2)$, 郝春梅在此基础上进行改进, 使用邻接表存储关系图, 大大节省了存储空间, 利用堆排序算法寻找下一个路径节点, 缩短了一定的运行时间^[3]. 随机游走算法是从自己的邻居节点中随机

选取一个节点进行下一步检索, 重复这个过程直到检索到目标节点^[4]. 最大度检索算法是从自己的邻居节点中选取度最大的节点进行检索, 直到检索到目标节点^[5]. 沈敬欣将随机游走与最大度检索相结合, 用于复杂网络中检索最短路径^[6]. 毛红梅提出了一种求解源节点到目标节点所有路径的算法, 该算法通过逐次求得前驱节点的所有路径来实现检索目的, 算法易于理解, 但操作繁琐, 时间复杂度较高^[7].

现有检索算法忽视了社会关系网络中人与人之

间的亲疏程度(亲疏程度可以用节点之间的亲疏度来衡量).忽视亲疏度的基于路径长度的检索结果,很多时候不能满足用户需求.针对这个问题,笔者研究一种基于泛洪算法与亲疏度的社会关系检索算法,能够求出用户与想要联系好友间的所有关系路径,并且按照亲疏度高低将关系路径排序,给予用户足够的选择空间.

2 泛洪算法与亲疏度

2.1 泛洪算法

泛洪算法是无线网络中一种最基本的、最简单的路由算法^[8].由于它具有操作简单、容错性高等优点,成为信息传播、文件查询等应用的主要手段.泛洪算法的基本思想是:源节点发送一个检索请求给它的所有邻居节点,这些邻居节点再分别将检索请求转发给它们自己的邻居节点,直到找到期望的资源或者达到网络允许的最大检索跳数后结束检索.检索模型的示意如图1所示,当节点1发送一个检索请求,先把该请求发送到邻居节点2,3,4,节点2再将请求发送给邻居节点6,依次类推,当检索请求到达节点5,由于节点5存储节点1所需的检索资源,就会响应节点1完成检索.

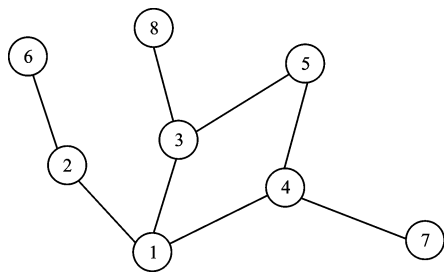


图1 泛洪算法检索模型

2.2 亲疏度

2.2.1 亲疏度介绍

社会关系网络是由一系列节点和关系构建的一种社会结构,可以将其转化为图论中节点和边的关系.节点代表社会关系网络中的个体,可以单指某个人,也可以是组织、社区、企业等具有社会意义的实体.边是节点间的一种关系的连接形式^[9],涵盖着形形色色的社会关系,既包括偶然相识的点头之交,又包括血肉相连的亲人.亲疏度是将关系的亲疏远近程度映射到 $(0,1]$ 区间的定量描述,越接近于1,表示关系越亲近,办事成功的概率越高;越接近于0,表示关系越疏远,办事成功的概率也就越低.它的大小与互动频率、感情力量、亲密程度、互惠交换^[10]等

因素相关.图论中涉及到的相关定义如下所示.

定义1 设社会关系图 $G = (V, E)$ 由节点集 $V(G)$ 和边集 $E(G)$ 组成,其中 $V(G) = \{v_i \mid 1 \leq i \leq n, n \text{ 为结点总数}\}$, $E(G) = \{e_i \mid e_i = (v_k, v_j), v_k, v_j \in V, \text{即 } e_i \text{ 是图中结点 } v_k \text{ 与结点 } v_j \text{ 间的边}\}$.

定义2 关系路径,使两个节点间连通的有效路径称为关系路径,通常用 RL 表示, $RL = \{v_0, v_1, v_2, \dots, v_{l-1}, v_l\}$, l 为关系路径的长度.

定义3 亲疏度,社会关系网中相邻节点间关系亲疏程度的量化值称为亲疏度,用 a 表示, $0 < a \leq 1$.

定义4 累积亲疏度,将一条关系路径上每个相邻节点的亲疏度按照一定计算形式进行计算,即得到这条关系路径的累积亲疏度,用 A 表示.

2.2.2 累积亲疏度的计算

累积亲疏度的大小与关系路径中每个相邻节点的亲疏度大小相关,反映了一条关系路径中源节点与目标节点的亲疏远近程度.求得累积亲疏度必须满足以下两个条件.

条件1:计算出的累积亲疏度的值仍然在 $(0,1]$ 范围内,即 $A \in (0,1]$.

条件2:累积亲疏度不能大于该条关系路径中每个相邻节点的亲疏度,即 $A_{0l} \leq \{a_{01}, a_{12}, a_{23}, \dots, a_{(l-1)l}\}$.

经过这两个条件限制,求得的累积亲疏度与现实生活情况是相符的.关系的累积亲疏度 A_{0l} 计算公式如下:

$$A_{0l} = \delta \prod_{i=0, j=i+1}^{l-1} a_{ij} \quad (1)$$

式中, $\prod$ 是连乘符号; a_{ij} 代表关系路径中每个相邻节点间的亲疏度; l 为关系路径长度; δ 为衰减因子,它的计算公式为

$$\delta = 1 - \frac{l}{M} \quad (2)$$

M 代表可调参数,此处 $M = 50$.

3 FIS 算法与模拟分析

3.1 FIS 算法的检索策略

人与人之间的关系种类复杂多样、关系程度不尽相同.因此,社会关系网络不仅有小世界、无标度的特性,还具有自身特殊的复杂特征.在一个规模庞大的网络中,连接两个节点间的关系路径可能存在多条.如果能够将这些路径全部检索出来,为用户联系好友提供足够的选择性,无疑不是个良好的检索策略.泛洪算法具有实现简单、检索范围广、健壮性高等优势,是检索所有

路径的不二法门. 以往的检索算法大多检索最短或最优路径, 忽视了亲疏度的影响, 致使检索结果不能很好地满足人们的要求. 为此充分考虑亲疏度这个影响因素, 将其映射为边权值, 提出了一种基于泛洪算法与亲疏度的社会关系检索算法, 即 FIS 算法.

FIS 算法思想: 网络中的每个节点保存自己邻居节点的信息, 包括节点的属性和其相连的边的亲疏度. 源节点以泛洪查询的方式检索自己的每个邻居节点, 检索时按照其与邻居节点的亲疏度从大到小的顺序检索, 因为亲疏度越大, 通过该节点找到目标节点的概论也就相对大些. 如果邻居节点中包含目标节点, 就将该条关系路径和路径上的累积亲疏度记录到路径表中; 如果不包含目标节点, 则继续检索邻居节点的邻居节点, 直到检索出网络中的所有关系路径或检索到设置的阈值 TTL 范围内的关系路径. TTL 是根据实际情况设置的最大关系路径长度, 其大小满足式(3):

$$TTL \leq \ln(N) \quad (3)$$

式中, N 是网络中的节点总数. 算法在此基础上, 将检索到的关系路径按照累积亲疏度的大小进行排序, 更科学更合理地满足用户交友要求. FIS 算法具体的实现过程如下.

Neighbor: 邻居节点的集合, 其中 $\text{Neighbor}[i]$ 表示节点 i 的邻居节点的集合.

Linklist: 检索关系路径时, 临时存储节点的单链表.

* front: 单链表中, 指向当前检索到的节点.

* p : 将关系路径存储到路径表时, 用于查询当前节点的前驱节点.

Path: 路径表, 记录已经检索到的关系路径的集合.

TTL: 检索过程中设置的路径长度, 如果没有设置, 则检索整个网络.

s : 设置的源节点.

d : 设置的目标节点.

初始化时, 将源节点 s 插入单链表中.

while(Linklist! = NULL)

{ $a \leftarrow * \text{front}$

if(节点 a = 目标节点 d)

{ 调用 $\text{Get_path}(G, \text{Linklist})$ 将该条路径存入 Path 中
* front = * front + 1 }

else

{ 节点 a 的每一个邻居节点 $b \in \text{Neighbor}[a]$

if($TTL > 0$ 且 $b! = a$ 的前驱节点)

{ 每一个邻居节点 b 按照亲疏度大小 $\text{En_LinkList}(\text{Linklist}, b, W_{ab})$ 插入单链表中 }

* front = * front + 1 }

}

3.2 FIS 算法模拟分析

如图 2 所示, 模拟一个小型的社会关系网络, 该网络由 6 个节点及 7 组关系构成. 节点 a 想要结识节点 f , 需要检索从 a 到 f 的关系路径, 帮助其建立连接.

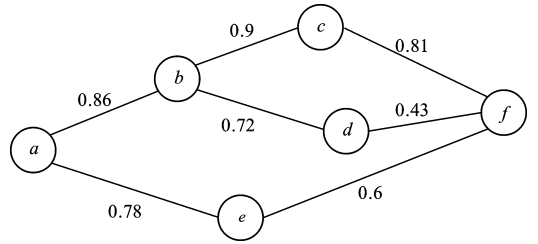


图 2 模拟的小型社会关系网络

初始化时, 将节点 a 插入单链表, 检索节点 a 邻居节点 b, e , 不满足检索请求, 则将 b, e 插入单链表; 节点 b 检索邻居节点 c, d , 不满足检索请求, 则将 c, d 插入单链表; 节点 e 检索邻居节点 f , 满足请求, 记录关系路径 $a \rightarrow e \rightarrow f$, 按照式(2)、(3) 计算得出 $A_{af} = 0.4493$; 节点 c 检索邻居节点 f , 满足请求, 记录关系路径 $a \rightarrow b \rightarrow c \rightarrow f$, $A_{af} = 0.5893$; 节点 d 检索邻居节点 f , 满足请求, 记录关系路径 $a \rightarrow b \rightarrow d \rightarrow f$, $A_{af} = 0.2503$, 此时已经检索整个网络, 检索结束. 按照亲疏度排序后, 得到如下结果:

(1) $a \rightarrow b \rightarrow c \rightarrow f \quad A_{af}^1 = 0.5893$;

(2) $a \rightarrow e \rightarrow f \quad A_{af}^2 = 0.4493$;

(3) $a \rightarrow b \rightarrow d \rightarrow f \quad A_{af}^3 = 0.2503$.

由上述分析可知, 该算法可以检索网络中或设置的 TTL 范围内源节点到目标节点的所有关系路径, 且按照亲疏度的大小顺序依次列出. 通过结果可知, l 小的关系路径, A 未必大, 即关系未必亲近; 相反, l 大的关系路径, A 未必小. 这与实际生活中的交友情况是相符的. 算法的时间主要集中在单链表中指针查找的过程, 因此时间复杂度取决于单链表中节点的个数. 查找时大多到不了链尾就可以找到所需节点, 所以在最坏的情况下, 该算法的时间复杂度为 $O(e)$, e 是网络中边的总数.

4 仿真实验

4.1 关系路径长度与亲疏度关系实验仿真

为了探究关系路径长度与亲疏度的关系, 需要进行一次实验. 实验选取真实的社会关系网络, 网络规模 $N = 1\,000$, 每个节点代表一个单独的个体, 无

重复、重名等现象. 根据用户间的互动频率、感情力量、亲密程度等因素赋予适当亲疏度. 随机选取 10 组源节点和目标节点, 进行路径检索, 分别记录每组中的关系路径及亲疏度, 并且按照关系路径长度进行分组, 最终获取每个关系路径长度值的平均亲疏度. 相应的计算如下:

$$A(k)=\frac{\sum_{i=1}^{10}\sum_{j=1}^{N_i}A^{ij}}{\sum_{i=1}^{10}N_i}\tag{4}$$

式中, $A(n)$ 表示 $l=k$ 的平均亲疏度, $k\in[3,12]$ 的整数; N_i 表示第 i 组中 $l=k$ 的关系路径条数. A^{ij} 表示第 i 组中第 j 条关系路径的累积亲疏度. 按照(4) 得到数据如图 3 所示.

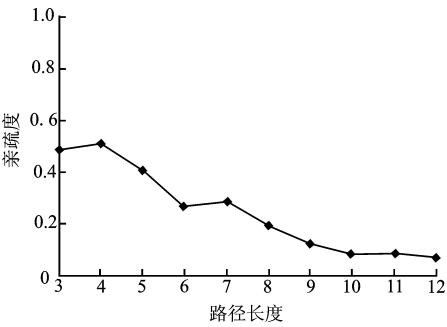


图 3 亲疏度与关系路径长度

从图 3 可以看出, 随着路径长度的增长, 亲疏度总体呈现下降的趋势, 但具体到某一路径长度值时, 路径长度小, 亲疏度未必大. 例如: 图中 $l=3$ 时的亲疏度就略小于 $l=4$ 时的亲疏度.

4.2 FIS 算法综合性能实验仿真

在进行综合性能仿真前, 需要介绍两个检索领域的概念: 查准率(Precision ratio)和查全率(Recall ratio), 二者是反映检索效果的重要指标. 相关计算公式如下:

$$P=\frac{s}{S}\times100\%\tag{5}$$

$$R=\frac{s}{s+s'}\times100\%\tag{6}$$

式中, s 表示检索到的正确的路径条数; S 表示检索到的总的路径条数; s' 表示网络中存在的未被检索到的源节点到目标节点的路径条数.

为了进一步验证算法的有效性, 从查全率、查准率、检索时间方面将随机游走算法、最大度检索算法、文献[7]算法和 FIS 算法进行对比, 用来衡量不同算法的检索效果. 网络仍然选取上节所述的真实社会关系网络. 选取一组源节点和目标节点, 进行

10 次实验, 计算平均值, 以便减小实验产生的误差. 对比结果如表 1 所示.

表 1 各个算法性能参数表

算法	查准率/%	查全率/%	检索时间/s
随机游走算法	100	12.5	0.21
最大度算法	100	12.5	0.14
文献[7]算法	99.8	100	0.62
FIS 算法	100	100	0.45

从表 1 可以看出, 4 种算法在查准率上都比较理想, 都可以检索出从源节点到目标节点的关系路径. 但由于随机游走和最大度检索算法只能检索一条关系路径, 查全率比较低. 文献[7]算法和 FIS 算法在查全率和查准率方面不相上下, 但在检索时间上, 文献[7]算法差于 FIS 算法, 这主要由于文献[7]算法在检索目标节点的路径时, 必须先检索其前驱节点的所有路径, 耗费大量的检索时间. 综上实验可以说明, FIS 算法的综合性能比较理想, 适用性强.

5 结束语

社会关系网络检索中, 忽视了节点之间亲疏度的影响, 而且大部分算法只能检索到一条关系路径. 鉴于此, 本文提出了基于泛洪算法和亲疏度的社会关系检索算法. 该算法用亲疏度来衡量节点之间的关系程度, 可以检索到从源节点到目标节点的所有路径, 并根据亲疏度的大小将路径排序. 研究结果发现, 关系路径长度小的路径, 亲疏度未必大; 相反, 关系路径长度大的路径, 亲疏度未必小. 同时也证明了, 该算法的查全率、查准率和检索时间均比较理想, 适用性较强.

参考文献:

[1] 傅安国, 郑剑虹. 人际关系网络对事业生涯发展影响的质性研究[J]. 青年研究, 2012, 30(3): 63-74.

[2] 杜一鸣, 滕桂法, 马建斌. 社会网络搜索关键技术研究概述[J]. 现代图书情报技术, 2010, 26(2): 68-73.

[3] 郝春梅. 一种改进的 Dijkstra 算法的分析及程序实现[J]. 计算机与现代化, 2011, 26(1): 36-38.

[4] Noh J D, Rieger H. Random walks on complex networks[J]. Phy Rev Lett, 2004, 92(11): 87-93.

[5] Adamic L A, Lukose R M, Huberman B A. Local search in unstructured networks[M]. Berlin: Wiley-VCH, 2003.

[6] 沈敬欣. 结合最大度与随机游走策略的复杂网络搜索技术研究[D]. 辽宁: 大连海事大学, 2012.

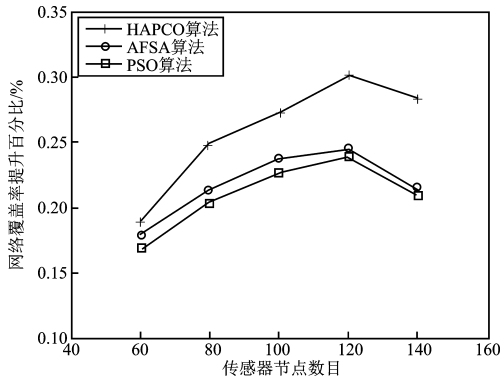


图8 节点数目对覆盖率的影响

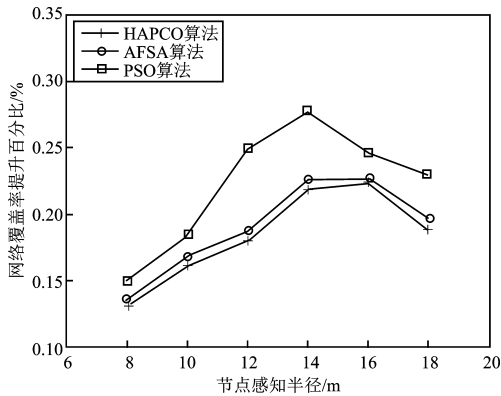


图9 感知半径对覆盖率的影响

6 结束语

本文针对 WMSN 中存在覆盖重叠区和覆盖盲区的问题,采用了人工鱼群与粒子群融合覆盖算法(HAPCO)对监测区域进行覆盖优化,利用人工鱼群全局性好的优点先对传感器节点的位置和方向进行调整,使节点分布均匀,然后通过粒子群算法局部调整传感器节点的位置和方向,优化网络覆盖效果,最后对人工鱼群与粒子群融合覆盖算法进行了一系列的仿真实验,通过改变迭代次数、传感器节点半径、节点数目对算法的性能进行分析,实验表明该算

法较大地提升了网络覆盖率.

参考文献:

- [1] 文政颖, 翟红生. 基于混沌人工蜂群算法的无线传感器网络覆盖优化[J]. 计算机测量与控制, 2014, 22(5): 1609-1612.
- [2] 刘维亭 范洲远. 基于混沌粒子群算法的无线传感器网络覆盖优化[J]. 计算机应用, 2011, 31(2): 338-340.
- [3] Peng L, Liu K, Liu G. Research on wireless sensor networks fault-tolerant coverage algorithm base on particle swarm optimization[C] // IET International Conference on Wireless Sensor Network 2010 (IET-WSN 2010). [s. l.]: IEEE, 2010: 286-290.
- [4] 郭剑, 孙力娟, 韩崇, 等. 基于改进 PSO 的 WMSNs 覆盖增强算法[J]. 微电子学与计算机, 2012, 29(5): 1-5.
- [5] Tao D, Tang S, Liu L. A coverage enhancement algorithm based on constrained artificial fish-swarm in directional sensor networks [J]. Journal of Internet Technology, 2014, 15(1): 43-52.
- [6] 黄瑜岳, 李克清. 基于人工鱼群算法的无线传感器网络覆盖优化[J]. 计算机应用研究, 2013, 30(2): 554-556.
- [7] 李丽, 李会, 张天丽, 等. 基于改进鱼群算法的 WSN 覆盖优化策略[J]. 微电子学与计算机, 2013, 30(2): 83-86.
- [8] Lee D T, Lin A K. Computational complexity of art gallery problems [J]. Information Theory, IEEE Transactions on, 1986, 32(2): 276-282.

作者简介:

王秀全 男, (1990-), 硕士研究生. 研究方向为无线传感器网络. E-mail: 1252764291@qq.com

冯秀芳 女, (1966-), 博士, 教授. 研究方向为无线传感器网络、人工智能.

郑惠月 女, (1988-), 硕士研究生. 研究方向为无线传感器网络.

(上接第 105 页)

- [7] 毛红梅, 甘晟科. 求关系图中源点到各结点所有路径的一种实用算法[J]. 微电子学与计算机, 2009, 26(3): 128-130.
- [8] 李方敏, 刘新华, 旷海兰. 无线传感器网络中一种高效低延时的泛洪算法研究[J]. 通信学报, 2007, 28(8): 46-53.
- [9] Sundaresan N, Yi J. Mining the Web for relations[J]. Computer Networks, 2000, 233(1-6): 699-711.
- [10] Granovetter. The strength of weak ties[J]. The American Journal of Sociology, 1973, 78(6): 1360-1380.

作者简介:

王月 女, (1989-), 硕士研究生. 研究方向为计算机网络与数据库.

滕桂法(通讯作者) 男, (1963-), 教授, 博士生导师. 研究方向为计算机网络与智能信息处理研究.

E-mail: tguifa@hebau.edu.cn.

王春山 男, (1978-), 副教授. 研究方向为智能农业装备.

张璠 女, (1983-), 讲师. 研究方向为计算机网络与智能信息处理.