基于3-layer中心度的社交网络影响力最大化算法

王 俊 余 伟 胡亚慧 李石君
(武汉大学计算机学院 武汉 430072)

摘 要 社交网络影响最大化问题是指如何寻找网络中有限的初始节点,使得影响的传播范围最广。一些贪心算法可以得到较好的影响范围,但是因时间复杂度太高而不适用于大型社交网络。基于度中心性的启发式算法简单但准确度不高;基于介数中心性、接近中心性等全局指标的启发式算法可以较好地识别影响力最大的节点,但计算复杂度也过高。考虑网络节点深层次结构对影响扩散的作用并权衡计算复杂度与准确度,定义了3-layer局部中心度,以计算节点的潜在影响力值。基于线性阈值模型,启发选择一部分种子节点:每一次都选取潜在影响力最大的节点作为种子节点进行激活;运用贪心算法选取剩下的一部分种子节点:每一次都选取具有最大影响增量的节点作为种子节点进行激活。实验表明,该混合算法具有很好的激活范围以及非常低的时间复杂度。

关键词 社交网络,影响力最大化,启发式算法,3-layer局部中心度,贪心算法
中图法分类号 TP311.13 **文献标识码** A

Heuristic Algorithm Based on 3-layer Centrality for Influence Maximization in Social Networks

WANG Jun YU Wei HU Ya-hui LI Shi-jun
(School of Computer, Wuhan University, Wuhan 430072, China)

Abstract Influential maximization in a social network is the problem of finding the limited initial nodes which can maximize the spread of influence. Some greedy algorithms can get wide spread of influence, but have too high cost to be applied in large-scale social networks. The heuristic method based on degree centrality is simple but of low accuracy. Heuristic algorithms based on global metrics such as betweenness centrality and closeness centrality can better identify the most influential nodes, but the computational complexity of which is much high too. As a tradeoff between the low-accurate degree centrality and other time-consuming measures, we defined the 3-layer local centrality for computing the potential influence of nodes. Based on linear threshold model, we selected some seed nodes through the heuristic algorithm in which the node with maximal potential influence is selected at each step, and we chose another seed nodes by the greedy algorithm in which the node with the largest influence increment is chosen at each time. The experimental results show that our hybrid algorithm has good spread of influence and low time complexity.

Keywords Social network, Influence maximization, Heuristic algorithm, 3-layer local centrality, Greedy algorithm

1 引言

一个人使用了一种新的产品后可能会告诉他的朋友们也去使用这种产品,他的朋友们又把这种产品推荐给他们的朋友,于是产生了级联推荐的效果,这就是“口碑效应”。商家计划推出新产品时,就可以利用这种效应:商家可以先找到一些具有影响力的个体当试验人,让他们免费试用产品并把产品推荐给他们的朋友们,以使得产品在人群中最大范围地被认可。在资金有限的情况下,如何找到有限的最大影响力试验人就是商家们所要考虑的问题。这种“口碑效应”的营销模式充分利用了顾客之间的关系,被形象地称为“病毒营销”[1,2]。它可以使商家利用有限的资金得到更多的回报,越来越受到商家们的推崇。

对于谣言在社会中的扩散、计算机病毒在计算机网络上的蔓延、传染病在人群中的流行等不良传播,我们要找出其中影响力最大的传播种子,然后重点将这些传播种子去除或阻隔,以扼制这些不良影响的快速传播势头。

近年来,由于大型社交网络如Facebook、人人网、微博等的兴起,各种信息通过社交网络被广泛传播。如何选取有限的初始传播种子,从而使信息的最终传播范围最大,称为“社交网络影响力最大化问题”。社交网络影响力最大化问题是社交网络研究领域的一个热点问题。

Kemp和Kleinberg等人证明了社交网络影响力最大化问题是一个NP难问题,并提出了一个爬山贪心算法KK算法[3]。KK算法能保证在1−1/e的范围内接近最优解,可以得到较好的影响力最大化节点,但是因为计算复杂度太高而

到稿日期:2013-06-18 返修日期:2013-07-17 本文受国家自然科学基金(61272109)资助。
王 俊(1986−),男,博士,主要研究方向为社交网络、Web数据挖掘,E-mail:wjwj@whu.edu.cn;余 伟(1987−),男,讲师,主要研究方向为Web数据挖掘、社交网络,E-mail:yuwei@whu.edu.cn(通信作者);胡亚慧(1980−),女,博士,主要研究方向为数据挖掘、人工智能;李石君(1964−),男,教授,博士生导师,主要研究方向为数据挖掘、数据库。

不适用于大规模社交网络。

田家堂等人提出了一种结合启发算法和贪心算法的混合的社交网络影响力最大化算法 HPG[4]。HPG 算法在启发阶段以节点的度中心性为主要的度量标准来计算节点的潜在影响力，而节点的度中心性没有涉及网络的更深层次结构，度数高的节点不一定是最有潜在影响力的节点。陈浩等人对 HPG 算法进行改进，提出了一种基于动态阈值的混合算法 TBH[5]。当启发阶段选择的节点数量相同时，TBH 算法可以得到比 HPG 算法更大的影响范围，而且在完全不利用 KK 算法的情况下，也能得到和 KK 算法很接近的激活范围，同时时间复杂度相对 KK 算法更低。

多种中心性指标被提出以衡量节点的影响力[6]，常用的中心性指标包括度中心性、介数中心性、接近中心性[7,8]和特征向量中心性[9,10]。Chen Duanbing 等人提出了半局部中心度[11]的概念，半局部中心度相对于介数中心性、接近中心性等指标具有更低的计算复杂度，相对于度中心性指标具有更好的识别结果。吕宁媛等人提出了 LeaderRank[12]算法，该算法能很好地识别有向网络中的影响力最大化节点，但是对无向网络效果不大好，PageRank[13]算法也有类似的缺陷。Bonan Hou 等人定义了全方位距离(all-around distance[14])来衡量节点的影响力，这比单一指标更精确和更稳定，但是计算复杂度相当大。D. Wei 等人基于证据理论提出了一种新的中心性指标[15]，权衡了加权网络中每个节点的度和强度，被证明能有效地识别加权网络中影响力最大的节点。社交网络中社区结构的拓扑属性有助于识别最具影响力的节点[16,17]。另外，动态社交网络的影响力最大化问题已成为一个研究热点和难点[18]。

本文提出了基于 3-layer 局部中心度的启发式算法，再与贪心算法组合成新的混合算法 LPG。在不同的数据集上进行实验，结果显示 LPG 算法比 HPG 算法和 KK 贪心算法能更好地识别社交网络影响力最大节点，具有很好的影响范围以及非常低的时间复杂度。

2 背景知识

2.1 问题描述

用网络图 $G(V,E)$来描述社交网络，其中 V 表示网络上的节点集，节点代表社交网络中的组成单元；E 表示节点之间存在的边集合，代表单元之间的影响关系。当一个节点被成功影响时，称此节点被激活。每个节点有两种初始状态，即激活和未激活，只有处于激活状态的节点才对它指向的节点具有影响力。每个节点只能由未激活状态转化为激活状态。一个处于激活状态的节点对其所有能产生影响效果的节点尝试激活后，自身仍保持激活状态，但已不具备影响力了，即不会出现重复影响的情况。

给定社交网络 $G(V,E)$，正整数 k，$A\subset V$，令 $\sigma(A)$为以节点集 A 为初始节点在影响扩散过程结束后激活节点的个数。社交网络影响最大化问题就是要找出其中影响力最大的 k 个节点，即找到节点集 S，使得

$$S=\arg\max_{|A|=k}\sigma(A) \tag{1}$$

2.2 传播模型

社交网络影响扩散的传播模型中，最常见的是独立级联(Independent Cascade)模型和线性阈值(Linear Threshold)模型。

2.2.1 独立级联模型[3,19,20](IC)

IC 模型传播机制中的要点是用概率来表示一个节点对另一个节点的激活情况，概率越大表示激活成功的可能性越大。

当节点 v 变成激活状态，w 为 v 指向的处于未激活状态的节点，节点 v 对节点 w 影响成功的概率是 $p(v,w)$。节点 w 已被激活的邻居节点对节点 w 的影响顺序是随机的。如果节点 w 被成功激活，则它对其指向的节点开始具有影响力并尝试激活它们。

IC 模型中 $p(v,w)$与其它节点对节点 w 的影响无关，当节点 v 激活 w 失败后，影响效应不会累积。

2.2.2 线性阈值模型[3,21](IT)

LT 模型传播机制中的关键点是用阈值表示一个节点被激活的难易程度，阈值越大表示节点越难被激活。只有当节点的已激活邻居节点对该节点的影响力之和达到或者超过这个阈值时，该节点才被激活。

$\forall v\in V$，θ_V 为 v 的激活阈值，对它有影响的邻居节点集合记为 A_v。$\forall u\in A_V$，b_{uv} 为节点 u 对节点 v 的影响力，满足 $\sum_{u\in A_V} b_{uv}\leqslant 1$。如果节点 v 的已激活邻居节点 $B_v(B_v\subset A_v)$对 v 的累积影响力之和达到或超过 v 的激活阈值，即 $\sum_{u\in B_v} b_{uv}\geqslant\theta_v$，那么节点 v 会由未激活状态变为激活状态。节点 v 变为激活状态后又会对它所指向的节点产生影响。

如果已激活节点 u 对 v 激活失败，则影响力 b_{uv} 被积累下来，在后续其它节点对 v 的激活过程中 b_{uv} 仍发挥作用，即 IT 模型具有“影响累积”效应。

2.3 KK 算法

Kemp 和 Kleinberg 提出了一种贪心算法，这里简称为 KK 算法[3]。令 $\sigma(S)(S\subset V)$为以 S 为初始节点集在影响扩散过程结束后已激活节点的数量，则 KK 贪心算法的描述如下：

```
1. S=∅
2. FOR i=1 TO k
3.    v=arg max_{u∈V\S}(σ(S∪{u})−σ(S))
4.    S=S∪{v}
5. END FOR
```

最后得到的 S 即为 KK 算法识别出的影响力最大的 k 个节点的集合。KK 算法每一步都要寻找影响力增量最大的节点，可以得到较好的影响范围，但显然很耗时，对于大规模社交网络不适用。

2.4 HPG 算法

田家堂等提出了一种启发和贪心相结合的混合算法 HPG[4]。启发阶段主要基于节点的度中心性计算未激活节点的潜在影响力(potential influence)PI。设 u 为网络中的未激活节点，$inf(u)$为 u 对它指向的未激活节点的影响力之和，即 $inf(u)=\sum_{v\in out1(u),active(v)=0} b_{uv}$，则在 HPG 算法中，$u$ 的潜在影响力定义如下：

$$PI(u)=outdegree(u)+(1-e^{-inf(u)})$$

这里 $outdegree(u)$为 u 的出度。HPG 算法首先基于 PI 每一步选择潜在影响力最大的节点进行激活，得到一部分种子节点，然后贪心地选取另外一部分节点。HPG 算法的构思很巧妙，有效利用了线性阈值模型的影响力累积效应，与 KK 算法相比，在大幅度降低了时间复杂度的同时具有更好的影响范围。

3 基于 3-layer 局部中心度的影响力最大化算法（LPG 算法）

3.1 3-layer 局部中心度

如图 1 所示，虽然节点 11 比节点 1 有更低的度，但它的影响力很可能比节点 1 大，也就是说以度中心性作为潜在影响力的度量指标精确度不高。HPG 算法以度作为衡量节点潜在影响力的主要指标，精确度有待提高。为此本文提出了 3-layer 局部中心度的概念。

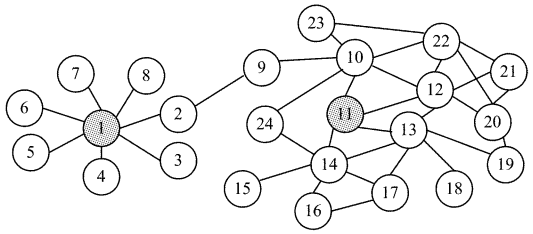

图 1　24 个节点组成的网络图

对于网络 $G=(V,E)$，$\forall v\in V$，节点 v 的邻居节点集为 $out1(v)$，$out1(v)$ 中所有节点的邻居节点集的总集合为 $out2(v)$，$out2(v)$ 中所有节点的邻居节点集的总集合为 $out3(v)$；设 $deg1(v)$、$deg2(v)$、$deg3(v)$ 分别为 v 的相应层次邻居节点集 $out1(v)$、$out2(v)$、$out3(v)$ 的影响度。

对于无向网络，单个节点 u 的邻居节点指与 u 有边相连的节点；对于有向网络，单个节点 u 的邻居节点为 u 所指向的节点。

对于无符号网络，影响度定义为邻居节点集的元素个数，即

$$deg1(v)=|out1(v)|$$
$$deg2(v)=|out2(v)|$$
$$deg3(v)=|out3(v)|$$

对于带符号网络，节点 u 到节点 v 的边 (u,v) 的符号用 $sign(u,v)$ 表示，$sign(u,v)=\pm1$，符号为正表示促进传播，符号为负表示抑制传播，令

$$V_1=\{u|sign(v,u)=1,u\in out1(v)\}$$

则 $out2(v)$ 为 V_1 中所有节点的邻居节点集；令

$$V_2=\bigcup_{p\in out1(v)}\{u|sign(p,u)=1,u\in out2(v)\}$$

则 $out3(v)$ 为 V_2 中所有节点的邻居节点集。

邻居节点集 $out1(v)$、$out2(v)$、$out3(v)$ 的影响度分别对应 v、V_1、V_2 出边中符号为正的边数减去符号为负的边数，即

$$\begin{aligned}deg1(v)&=|\{m|sign(v,m)=1,m\in out1(v)\}|-|\{m|sign(v,m)=-1,m\in out1(v)\}|\\&=\sum_{m\in out1(v)}sign(v,m)\\deg2(v)&=|\{m|sign(v,m)=1,v\in V_1,m\in out2(v)\}|-|\{m|sign(v,m)=-1,v\in V_1,m\in out2(v)\}|\\&=\sum_{m\in out2(v),v\in V_1}sign(v,m)\\deg3(v)&=|\{m|sign(v,m)=1,v\in V_2,m\in out3(v)\}|-|\{m|sign(v,m)=-1,v\in V_2,m\in out3(v)\}|\\&=\sum_{m\in out3(v),v\in V_2}sign(v,m)\end{aligned}$$

定义 1　对网络 $G=(V,E)$，若 G 为无符号网络，则可以看作符号全为正的带符号网络，这里将 G 统一看作符号网络。设 M_+ 为网络中符号为正的边的总数目，令 $c=M_+/N$，即 c 为网络中每个节点的正出边的平均数目，则 c^2 为每个节点对应的每个正出边节点的正出边的平均数目。令 $deg1(v)$、$deg2(v)$、$deg3(v)$ 分别为 v 的 3 个层次(layer)邻居节点集的影响度。定义节点 v 的 3-layer 局部中心度 $deg(v)$ 如下：

$$deg(v)=w1\cdot deg1(v)+w2\cdot deg2(v)/c+w3\cdot deg3(v)/c^2 \tag{2}$$

其中，$w1$、$w2$、$w3$ 分别为各个层次邻居节点集的度权重，可以根据具体情况预定义。因为信息传播的不确定性，层次越深，节点对信息传播的影响度越低，所以应该满足 $0\leqslant w3<w2<w1\leqslant1$。

这里分别与 $deg1(v)$、$deg2(v)$、$deg3(v)$ 对应的节点之间可能有重复节点，但是考虑到 IC 模型的影响力累加效应，我们认为重复节点同时累加了更多重影响力，所以在计算影响度时不排除重复节点。因为要使 $deg2(v)$ 起作用，v 需通过其正出边激活尽可能多的邻居节点，而 $deg2(v)$ 是否起作用是有很大随机性的，所以将 $deg2(v)$ 除以 c 并根据具体情况调整权重 w_2，同理将 $deg3(v)$ 除以 c^2 并根据具体情况调整权重 w_3。当然也可以定义包含更深层次的局部中心度，但是综合考虑计算复杂度和影响相关度，这里只定义 3-layer 局部中心度。

3.2 节点间影响力

根据网络的拓扑结构，计算网络中两两节点之间的影响力。假设 $G(V,E)$ 为有向图，无向图的边可以当双向边处理。令 b_{uv} 为图 G 中节点 u 对 v 的影响力，(u,v) 为从 u 到 v 的边，$in(v)$ 为所有指向 v 的节点集。

3.2.1 加权图

对于无符号加权图 $G(V,E)$，G 中任意边 (u,v) 的权重为 W_{uv}，有

$$b_{uv}=\frac{W_{uv}}{\sum_{w\in in(v)}W_{wv}},\forall u\in in(v)$$

而对于带符号加权图，有

$$b_{uv}=\frac{sign(u,v)W_{uv}}{\sum_{w\in in(v)}W_{wv}},\forall u\in in(v)$$

3.2.2 无权图

对于无权图 $G(V,E)$，我们可以认为 G 中每条边的权重都为 1，令 $ind(v)=|in(v)|$ 为节点 v 的入度，则

对于无符号网络，有

$$b_{uv}=\frac{1}{ind(v)},u\in in(v)$$

而对于带符号网络，有

$$b_{uv}=\frac{sign(u,v)}{ind(v)},u\in in(v)$$

3.3 算法框架

3-layer 局部中心度刻画了节点更深层次的拓扑结构，相对节点的度在保持低时间复杂度的同时能更好地衡量节点的潜在影响力。我们对 HPG 算法进行改进，基于 3-layer 局部中心度提出新的混合式算法 LPG。LPG 算法分为启发和贪心两个阶段。

3.3.1 启发阶段

启发阶段利用 LT 模型的影响累积特性来启发式寻找最具潜在影响力的 $\lceil fk\rceil$ $(0\leqslant f\leqslant1)$ 个节点作为种子节点。虽然这些节点不要求像 KK 算法一样达到局部最优，但是在后续的激活过程中它们的潜在影响力会被激发出来，最终将得到比 KK 算法更大的影响范围。

若节点 v 处于激活状态，记 $active(v)=1$；否则记 $active$

$(v)=0$。令 $inf(u)$ 为 u 对所有未激活出边邻居节点的影响力之和：

$$inf(u)=\sum_{v\in out1(u),active(v)=0} b_{uv}$$

定义 2 节点 u 的潜在影响力值 PI 定义为：

$$PI(u)=deg(u)+w_0(1-e^{-inf(u)}) \quad (3)$$

这里 $deg(u)$ 为 u 的 3-layer 局部中心度，w_0 为累积影响权重，w_0 可以视具体网络选取和修正。

启发阶段每一步都选取最具潜在影响力的节点作为种子节点，进行激活过程后再选择下一个种子节点。

3.3.2 贪心阶段

以启发阶段已激活的节点为初始节点，利用 KK 贪心算法局部最优的特性来选取剩余的 $k-\lceil fk\rceil$ 个种子节点。在贪心阶段，启发阶段选取的 $\lceil fk\rceil$ 个种子节点的潜在影响力得到激活，最终得到很好的影响范围。

3.3.3 LPG 算法框架

输入：网络图 G(V,E)，激活阈值 θ_v(v∈V)，种子节点个数 k，启发比例系数 f，权重 w_1、w_2、w_3、w_0

输出：k 个影响力最大节点的集合 S，启发阶段激活节点个数 k_1，贪心阶段激活节点个数 k_2，最终激活节点总个数 k_0

1. S=ϕ
2. FOR i=1 TO ⌈fk⌉
3. 基于 3-layer 中心度和节点间的累积影响计算当前还未被激活的节点的潜在影响力 PI，找出 PI 最大的节点 v
4. S=S∪{v}
5. 进行激活过程，直到没有节点被激活为止
6. END FOR
7. 统计已激活节点个数 k_1
8. FOR i=(⌈fk⌉+1) TO k
9. 计算每一个未激活节点若被激活之后的影响增量，选择影响力增量最大的节点 v
10. S=S∪{v}
11. 进行激活过程，直到没有节点被激活为止
12. END FOR
13. 统计已激活节点个数 k_0
14. $k_2=k_0-k_1$

3.4 算法复杂度分析

假设网络 $G(V,E)$ 为有向带符号网络(无向网络可以当双向网络处理，无符号网络可以认为是边符号全为正的符号网络)，G 的边数为 M，符号为正的边数为 M_+。设启发阶段 LPG 算法和 HPG 算法每个种子节点的平均激活节点数分别为 A_1、A_2；贪心阶段 LPG 算法、HPG 算法、KK 算法每个种子节点的平均激活范围分别为 T_1、T_2、T_3。则有

LPG 算法的复杂度为：

$$O((1+M_+/N+(M_+/N)^2)\times M/N\times A_1)\times\lceil fk\rceil+O(N\times T_1)\times(k-\lceil fk\rceil) \quad (4)$$

HPG 算法的复杂度为：

$$O(M/N\times A_2)\times\lceil fk\rceil+O(N\times T_2)\times(k-\lceil fk\rceil) \quad (5)$$

KK 算法的复杂度为：

$$O(N\times T_3)\times k \quad (6)$$

A_1、A_2、T_1、T_2、T_3 是比较接近的有限值，而 $M_+/N\leqslant M/N\ll N$，其中“≪”表示“远小于”，所以我们可以推断 LPG 算法和 HPG 算法的复杂度比较接近，而比 KK 算法的复杂度低很多。

4 实验

4.1 实验数据集

实验数据集的统计信息如表 1 所列。

表 1 数据集信息

序号	数据集	节点数	边数(正/负数)	平均度	图类型
1	co-operate	7343	11898	3.2	无向、带权
2	Wiki-vote	7115	103689	26.6	无权、有向
3	Slashdot	77357	396378/120197	13.4	带符号、有向

第 1 个数据集是计算几何作者合作网络，边上的权重代表了作者之间合作的次数；

第 2 个数据集是 Wikipedia 的投票历史网络，节点 u 到节点 v 的有向边指用户 u 把票投给了用户 v；

第 3 个数据集是 Slashdot 网站 2008 年 11 月的数据，是一个朋友敌人网络，带符号，节点 u 到节点 v 的有向边表示用户 u 把用户 v 标注为朋友或敌人(若符号为正，则为朋友，反之为敌人)。

数据集 1 来自：Jones, B., Computational Geometry Database, February2002; FTP/HTTP。数据集 2 和数据集 3 来自 Stanford 大学的大型网络数据搜集网站(http://snap.stanford.edu/)。

4.2 实验结果与分析

通过大量实验表明：当 $w_1=0.9$，$w_2=0.4$，$w_3=0.1$，$w_0=1.5$ 时，LPG 算法具有相对其它权重更好的识别结果。在下面的实验中权重统一如此取值。

4.2.1 加权网络

实验采用数据集 1 的作者合作网络，所有节点初始激活阈值取 $\theta=0.5$。在不同的 f 下，变化种子结点数，比较 LPG 算法、HPG 算法、KK 算法激活的结点数目 k_0，结果如图 2 所示。从中可以看出，当 $f=1$ 时，即在完全不采用贪心算法的情况下，LPG 算法比 HPG 算法具有更好的影响范围，且接近 KK 算法的影响范围。如果完全不采用贪心算法，启发阶段积累的潜在影响力就不能被充分激活，最终影响范围不广，所以一般不这样做。当 f 不为 1 时，LPG 算法几乎总是比 HPG 算法和 KK 贪心算法的影响范围要好。k 越大时 LPG 算法与 KK 算法的影响范围差距越明显，这是因为此时 LPG 算法有更多的节点来积累潜在影响力，并且所累积的潜在影响力在贪心阶段也有更足够的机会被激发。

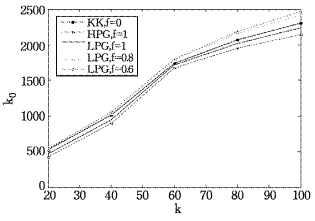

图 2 不同算法在不同 k 下的激活范围

4.2.2 有向网络

实验采用数据集 2 的维基投票网络，所有节点的初始激活阈值取 $\theta=0.6$。在相同的 k 下，变化 f，比较 LPG 算法和 HPG 算法激活的结点数目 k_0，当 $k=70$ 时，结果如图 3 所示，显示 LPG 算法比 HPG 算法具有更好的影响范围，而且 f 越大差异越显著。这是因为 LPG 算法能更好地找到最具潜在影响力的节点，最终就能激活更多的节点；而且随着 f 的增大，启发阶段结束后 LPG 算法与 HPG 算法已激活节点累积的潜在影响力差距也越大，最终影响范围的差距就相应变大。

当 $k=70$ 时，结果如图 4 所示，也可得出 LPG 算法具有比 HPG 算法更好的影响范围，而且除了 $f=1$ 的情况外，f 越大差异越明显。因为当 $f=1$ 时，虽然 LPG 算法较 HPG 算法能找到更有潜在影响力的节点，但是此时两算法都没有贪心阶段，潜在影响力不能充分地发挥作用，所以 LPG 算法与 HPG 算法的差距不一定变大。

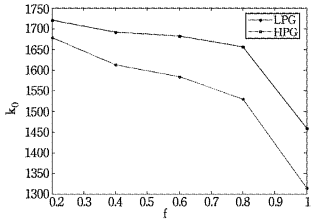

图 3 $k=40$ 时 LPG 算法和 HPG 算法在不同 f 下的激活范围

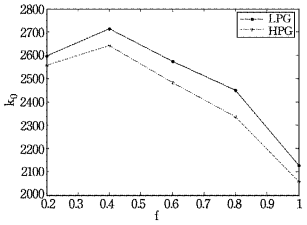

图 4 $k=70$ 时 LPG 算法和 HPG 算法在不同 f 下的激活范围

4.2.3 符号网络

实验采用数据集 3 的资讯科技网络，所有节点的初始激活阈值取 $\theta=0.9$。HPG 算法被证明当 $f=0.5$ 时效果相对较好，所以我们在相同的 $f=0.5$ 下，变化 k，比较 LPG 算法和 HPG 算法的运行时间 t 以及激活节点数 k_0，并与 KK 算法作比较。运行时间结果如图 5 所示，显示 LPG 算法与 HPG 算法的运行时间很接近，但是大幅低于 KK 算法。激活节点结果如图 6 所示，可以看出 LPG 算法总是比 HPG 算法有更好的激活范围，而且除了较小的 $k=10$ 外，也总是优于 KK 算法。当 k 越大时，HPG 算法的优势越明显。

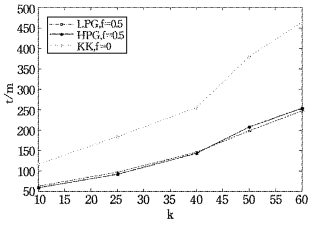

图 5 不同算法的运行时间

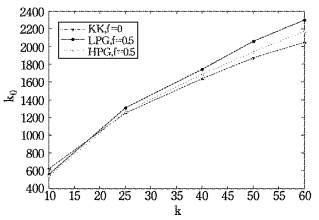

图 6 不同算法的激活范围

结束语 本文定义了 3-layer 局部中心度的概念，先以 3-layer 中心度为主要指标启发式选择部分潜在影响力最大的节点，再用贪心算法选择剩下的部分种子节点，从而提出一种新的混合式社交网络影响力最大化算法 LPG。在几个典型的数据集上进行实验，结果显示 LPG 算法更优于 HPG 算法和 KK 贪心算法，在保持较低时间复杂度的情况下，具有更好的影响范围。权衡复杂度和相关度，这里只探讨了 3-layer 局部中心度，我们将研究不同 layer 局部中心度的情况并将算法推广到 K-layer 局部中心度。另外权重值的变化对识别结果的影响满足什么规律也有待探讨。未来我们还打算将网络的社区结构融入此算法，并将其扩展到动态社交网络的影响最大化问题中。

参 考 文 献

[1] Leskovec J, Adamic L A, Huberman B A. The dynamics of viral marketing[J]. ACM Transactions on the Web, 2007, 1(1)

[2] Richardson M, Domingos P. Mining knowledge-sharing sites for viral marketing[C]// Preceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD, 2002: 61-70

[3] Kempe D, Kleinberg J, Tardos E. Maximizing the spread of influence through a social network[C]// Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, USA, 2003: 137-146

[4] Tian Jia-tang, Wang Yi-tong, Feng Xiao-jun. A new hybrid algorithm for influence maximization in social networks[J]. Chinese Journal of Computers, 2011, 34(10): 1956-1965

[5] Chen Hao, Wang Yi-tong. Threshold-Based Heuristic Algorithm for Influence Maximization[J]. Journal of Computer Research and Development, 2012, 49(10): 2181-2188

[6] Nicosia V, Criado R, Nunes M, et al. Controlling centrality in complex networks[J]. Scientific Reports, 2012(2): 218-223

[7] Okamoto K, Chen W, Li X-Y. Ranking of closeness centrality for large-scale social networks[C]// Proceedings of the 2nd Annual International Workshop on Frontiers in Alogorithmics, FAW'08. 2008: 186-195

[8] Freeman L. Centrality in social networks conceptual clarification [J]. Social Networks, 1979(1): 215-239

[9] Bonacich P, Lloyd P. Eigenvector-like measures of centrality for asymmetric relations[J]. Social Networks, 2001(23): 191-201

[10] kosch D, Lehmann K A, Peeters L, et al. Centrality indices[J]. Network, 2005, 3418: 16-61

[11] Chen Duan-bing, Lü Lin-yuan, Shang Ming-sheng, et al. Identifying in fluential nodes in complex networks[J]. Physica A: Statistical Mechanics and its Applications, 2012(391): 1777-1787

[12] Lü Lin-yuan, Zhang Yi-cheng, Yeung Chi-Ho, et al. Leaders in social networks, the delicious case[J]. PloSOne, 2011, 6(6): e21202

[13] Brin S, Page L. The anatomy of a large-scale hyper textual web search engine[J]. Computer Networks and ISDN Systems, 1998 (30): 107-117

[14] Hou Bo-nan, Yao Yi-ping, Liao Dong - sheng. Indentifying all-round nodes for spreading dynamics in complex networks[J]. Physica A, 2012(391): 4012-4017

[15] Wei Dai-jun, Deng Xin-yang, Zhang Xiao-ge, et al. Identifying influential nodes in weighted nodes in weighted networks based on evidence theory[J]. Physica A: Statistical Mechanics and its Applications, 2013, 392(10): 2564-2575

[16] Chen Yi-cheng, Chang Su-hua, Chou Chien-li, et al. Exploring Community Structures for Influence Maximization in Social Networks[C]// The 6th SNA-KDD Workshop'12 (SNA-KDD'12)

[17] Zhang Xiao-hang, Zhu Ji, Wang Qi, et al. Identifying influential nodes in complex networks with community structure [J]. Knowledge-Based Systems, 2013(42): 74-84

[18] Aggarwal C C, Lin Shu-yang, Yu P S. On influential node discovery in dymamic social networks[C]// Proceedings of the Eleventh SIAM International Conference on Data Ming, SDM 2011. Mesa, Arizona, USA, 2011: 636-647

[19] Yong H P. The diffusion of innovations in social networks[C]// Blume L, Durlauf S. The Ecomomy as a Complex System III. USA: Qoford University Press, 2003: 1-19

[20] Watts D J. A simple model of global cascades on random networks[J]. National Academy of Sciences, 2002, 99(9): 5766-5571

[21] Glodenberg J, Libai B, Muller E. Talk of the network: A complex systems look at the underlying process of word-of-mouth [J]. Marketing Letters, 2001, 12(3): 211-223