

附 1:

江西省研究生创新专项资金项目 申 请 表

项 目 名 称: 社交网络影响力最大化
的种子节点集研究

申 请 人: 梁文

指 导 教 师: 杨书新

培 养 单 位 (签 章): _____

填 报 时 间: 2018 年 4 月 25 日

江西省教育厅制

一、项目申报人基本情况

姓 名	梁文	性别	男		
出生年月	1994.07	籍贯	吉林省梅河口市		
在读学历层次	研究生	入学日期	2017.09.08		
在读专业	计算机技术				
身份证号码	220581199407200379				
指导师姓名	杨书新	研究方向	图数据管理、生物信息学		
本科(硕士)毕业 学校	江西理工大学		专 业	软件工程	
所在院系	信息工程学院		E-mail	541444726@qq.com	
联系电话	15579124322		手 机	15579124322	

一、研究意义

在线社交网络（Online Social Network）的出现缩短了地域空间距离、减少了信息传播时间、使大世界成为小小的地球村，它包罗万象、模拟着人们的真实生活。在社交网络中，以人为节点，以人与人之间关系的连线为边，真实世界则被模拟成一张图。人际关系交互、科研论文合著^[1]、蛋白质合作^[2]、微博谣言追源^[3]、链路预测分析^[4]、流行疾病传播的计算^[5]均可归为社会网络分析的研究领域。在真实社会中，个体与个体、个体与群体以及群体间存在着相互影响的关系，它小到身边人推荐一款手机游戏，大到参与地方人大代表的投票选举，其最终表现并非个体独立决策，而是在周围个体与群体不同程度影响作用下产生的。在线社交网络中影响的含义，是对个体在全局网络中重要程度的量化计算，它具备现实基础，而非空洞的抽象概念。其应用非常广泛，选票分析、广告投放、舆论控制、专家发现、意见领袖挖掘、个性推荐等。认识、洞察与量化计算社会影响力对企业高层有效预测与精准投入，尤其对政府部门掌握意识形态领导权具有不可或缺的意义。

二、研究现状

2.1 影响力最大化问题描述

社交网络中个体影响力的计算由来已久，可追溯至 1940 年美国总统竞选的选票统计。影响最大化问题^[6]也称 Top- k 问题，它要求给定社交网络 $G = (V, E)$ 与传播模型，选择 k 个初始节点作为种子节点集 S ($k = |S|$) 使 S 在全局网络产生的影响覆盖范围最广。它是对销售群体中选择最具推广力、影响力的用户免费试用产品并对外推广以达到产品最佳销售效果的形式化表达。

种子节点表示的可能是微博粉丝数量最多的有影响的活跃用户与明星用户，可能是兴趣社区被投票、与星级评分等级最高的用户。在真实社交网络，找到最具影响力的种子用户不能仅凭靠用户粉丝最多，评分最高，投票最热来考量，种子节点的影响力大小计算要依靠全局或局部拓扑结构、用户在线行为等因素综合分析。

2.2 影响力最大化及种子节点计算研究现状

目前，关于影响力最大化（Influence Maximization, IM）及种子节点的研究仍然是国内外的研究热点。2018 年 Menta 等人^[7]在 IEEE 发表了关于社交网络病毒营销如何有效选拔最具影响力节点的文章。同年，Samadi 等人^[8]发表了社交网络影响最大化的种子节点激活调度的研究，在给定预算下安排种子启动的时间问题。2017 年 G Tong 等人^[9]在 IEEE/ACM 发表了社交网络影响力最大化的自适应种子节点求解方法。Li Xiao 等人^[10]基于位置感知的社交网络中划分社区，挖掘种子节点并计算影响最大化。该问题自 2001 年提出至今从未停歇，不断突破。

关于影响最大化与种子节点的研究现状，本项目主要围绕“贪心算法及其改进、启发法预选种

子节点、同生活实际结合的种子计算”三方面阐述。

2.2.1 贪心算法及其改进算法

贪心策略是最为经典且最早应用于求解影响最大化问题的算法。Kempe 等人^[11]首次将 IM 问题定义为离散优化问题，利用贪心算法求解，并证明最大化的求解是 NP 难的，至此种子节点的求解有了开端。后续研究者对贪心算法的改进主要依据函数的子模特性。Leskovec 依据函数的子模特性提出 CELF^[12]的贪心优化算法，它是原算法运行效率的 700 倍。Goyal^[13]对 CELF 改进，提出较比 CELF 提升 35%-55%的 CELF++优化算法。Chen W^[14]提出了 MixdGreedy 的快速贪心算法，先剪去网络中影响力极低的边缘节点，再使用 CELF 算法求解。Zhou C^[15]通过设定上限减小问题规模，利用 CELF 算法使运行效率得到改善。LNG 算法是 Zhou S^[16]提出的通过局部影响力的计算近似替代全局影响力的算法，研究者将其命名为惰性节点的计算。

另一类主要依靠谷歌 PageRank^[17]算法。如，Hailin W 等人^[18]基于社区结构利用改进的 PageRank 算法对微博网络用户进行 IM 求解。随着研究的不断深入，2017 年 Rico Angell 和 Grant Schoenebeck^[19]发表了名为舍弃贪心算法求解 IM 问题的文章，论述并验证了根据社区结构寻找种子求解 IM 的有效性，合理地替代贪心算法。

2.2.2 启发法预选种子节点

预选种子节点，即预选出种子节点的候选集并在此集合发现种子节点。这类方法可分为两种，一种在全局搜索启发，一种对局部（社区）搜索启发。两种类型各有优劣，每个算法特点也不尽相同，根据文献[22]，如果能找到问题的映射，改进元启发式群体智能算法则是一个新的解决思路。

关于全局搜索启发，Chen W^[20]等人基于有向的网络图提出 Degree Discount 算法，即节点影响力的计算要打折扣，其效果较 Degree 算法求解影响力有更好的影响范围。Chen W 等人^[20; 21]基于独立级联模型，提出了 PMIA 算法，该算法认为节点影响力的传播只会发生在特定路径上，并将这个路径定义为最大影响路径，种子节点在其中产生，并造成全局影响最大化。Weskida 等人^[22]利用遗传算法映射影响最大化问题，利用演化算法启发种子节点，在 LT 模型下运行实验，结果表明速度是贪心算法的 35 倍。

有研究者发现社区结构内节点连接程度更紧密，信息的传播更快捷，远快于社区与社区间信息的传播，基于此特征有了社区结构解决 IM 问题的办法。关于局部搜索启发，Sirag E 等人^[23]根据网络局部结构对预选种子的启发法对比分析，证明了该方法能够近似替代全局网络搜索种子。曹玖新等人^[24]根据 k -核特征及影响半径，提出 CCA 算法。先以 k -核分解对节点分层并标记核数，在选择种子节点时，第一轮选择核数最大的节点当作种子节点，再进行影响最大化的求解。宋甲秀^[25]通过改进 CI (Collective Influence) 算法得到 LCII (Local Collective Influence Index) 局部集体影响指标评估法，并在破坏性网络条件下验证是有效的，但该方法在树形网络中不能得到最优

解。

启发法对传统贪心求解 IM 问题在运行效率方面有了极大突破。

2.2.3 同实际结合的种子计算

关于此类研究,少数研究者更关注种子节点集与生产实际的结合。文献[26]根据营销实际情况,提出种子节点的替补节点方法 GIA 算法,在种子节点不能被激活的条件下预留备选种子。文献[27; 28]提出种子集合最小化概念,求解的是影响力最大化对偶问题,同样依靠贪心算法函数子模特性。Budak C 等人^[29]基于竞争网络环境提出阻断影响最大化问题,通过加强刺激我方种子,使造成的影响最大。文献[30]在竞争网络环境中,基于线性阈值模型对阻断影响最大化的种子节点选择问题开展研究,提出 CLDAG 算法。Leskovec J 等人^[12]在网络中布置特殊节点以实时反映影响传播的效果,这类监控节点与种子节点的效应与选拔方法极为相似,都是利用贪心算法与模型的单调性、子模性,可列为同类问题的求解。可以看出,单独针对影响最大问题的种子节点集研究较少,与生活实际相结合的更是寥寥无几。

三、立论总结

影响力最大化问题评价指标有两项,一是算法的最终运行效率,即时间复杂度;二是种子节点最终在全局网络造成的影响覆盖范围,即种子节点激活全局节点的总个数。前人成果已经非常丰硕,且在两项评价指标上不断进步、突破,但现阶段已有研究针对种子节点与实际销售需求结合的很少。对实际销售情况、激活条件考虑不足,尤其在推广产品过程中通过控制种子节点投放以加强刺激的研究,尚处于空白。本项目立足于此,提出两类结合销售实际的种子节点投放且进一步加强刺激推进全局影响覆盖范围的方法。

参考文献:

- [1] Krasnov F, Dokuka S, Yavorskiy R. The Structure of Organization: The Coauthorship Network Case[M]. 2017: 100-107.
- [2] Sharma P, Bhattacharyya D K, Kalita J K. Centrality analysis in PPI networks[C]. International Conference on Accessibility To Digital World, 2017: 135-140.
- [3] Kboncheva. RDSM 2015 : 1st International Workshop on Rumors and Deception in Social Media: Detection, Tracking, and Visualization (#RDSM2015)[J], 2018.
- [4] Haghani S, Keyvanpour M R. A systemic analysis of link prediction in social network[J]. Artificial Intelligence Review, 2017, (3): 1-35.
- [5] Saeedian M, Azimi-Tafreshi N, Jafari G R, et al. Epidemic spreading on evolving signed networks[J]. Physical Review E, 2017, 95(2-1): 022314.
- [6] Domingos P R M. Mining the network value of customers[J]. In Proceedings of the 2011 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.ACM 2001: 54-66.
- [7] Menta V P T, Singh P K. Efficient selection of influential nodes for viral marketing in social networks[C]. IEEE International Conference on Current Trends in Advanced Computing, 2018.

- [8] Samadi M, Nagi R, Semenov A, et al. Seed Activation Scheduling for Influence Maximization in Social Networks[J]. Omega, 2018, 77.
- [9] Tong G, Wu W, Tang S, et al. Adaptive Influence Maximization in Dynamic Social Networks[J]. IEEE/ACM Transactions on Networking, 2017, 25(1): 112-125.
- [10] Li X, Cheng X, Su S, et al. Community-Based Seeds Selection Algorithm for Location Aware Influence Maximization[J]. Neurocomputing, 2017.
- [11] Kempe D, Kleinberg J. Maximizing the spread of influence through a social network[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003: 137-146.
- [12] Leskovec J, Krause A, Guestrin C, et al. Cost-effective outbreak detection in networks[C]. Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, California, Usa, 2007: 420-429.
- [13] Goyal A, Lu W, Lakshmanan L V S. CELF++:optimizing the greedy algorithm for influence maximization in social networks[C]. International Conference Companion on World Wide Web, 2011: 47-48.
- [14] Chen W, Wang Y, Yang S. Efficient influence maximization in social networks[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009: 199-208.
- [15] Zhou C, Zhang P, Guo J, et al. UBLF: An Upper Bound Based Approach to Discover Influential Nodes in Social Networks[J]. IEEE International Conference on Data Mining, 2013: 907-916.
- [16] Zhou S, Yue K, Fang Q, et al. An efficient algorithm for influence maximization under linear threshold model[C]. 中国控制与决策会议, 2014: 5352-5357.
- [17] Page L. The PageRank Citation Ranking : Bringing Order to the Web[C]. Proc. of the WWW Conference, 1998: 1-14.
- [18] Hailin W U, Amp G S, Co T. An Influence Maximization Algorithm Based on Community Partition and Improved PageRank[J]. Mobile Communications, 2017.
- [19] Angell R, Schoenebeck G. Don't Be Greedy: Leveraging Community Structure to Find High Quality Seed Sets for Influence Maximization[J], 2017: 16-29.
- [20] Chen W, Wang C, Wang Y. Scalable influence maximization for prevalent viral marketing in large-scale social networks[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010: 1029-1038.
- [21] Chen W, Yuan Y, Zhang L. Scalable Influence Maximization in Social Networks under the Linear Threshold Model[C]. IEEE International Conference on Data Mining, 2011: 88-97.
- [22] Weskida M, Michalski R. Evolutionary algorithm for seed selection in social influence process[C]. Ieee/acm International Conference on Advances in Social Networks Analysis and Mining, 2016: 1189-1196.
- [23] Erkol Ş, Yücel G. Influence maximization based on partial network structure information: A comparative analysis on seed selection heuristics[J]. International Journal of Modern Physics C, 2017.
- [24] 曹玖新, 董丹, 徐顺, et al. 一种基于 k-核的社会网络影响最大化算法[J]. 计算机学报, 2015, 38(2): 238-248.
- [25] 宋甲秀, 杨晓翠, 张曦煌. 复杂网络中 Top-k 影响力节点的识别算法[J]. 计算机科学与探索: 1-15.2018.
- [26] 马茜, 马军. 在影响力最大化问题中寻找种子节点的替补节点[J]. 计算机学报, 2017, (03): 674-686.
- [27] Michalski R, Kajdanowicz T, Bródka P, et al. Seed Selection for Spread of Influence in Social Networks: Temporal vs. Static Approach[J]. New Generation Computing, 2014, 32(3-4): 213-235.
- [28] Goyal A. On minimizing budget and time in influence propagation over social networks[J]. Social Network Analysis & Mining, 2013, 3(2): 179-192.

- [29] Budak C, Agrawal D, Abbadi A E. Limiting the spread of misinformation in social networks[C]. International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April, 2011: 665-674.
- [30] He X, Song G, Chen W, et al. Influence Blocking Maximization in Social Networks under the Competitive Linear Threshold Model Technical Report[J]. Computer Science, 2012.

四、研究方案

4.1 研究内容

本项目拟研究社交网络影响力最大化问题的特殊节点，种子节点集。具体研究包括以下三方面。

(1) 快速挖掘种子节点方法

种子节点的计算在启发式相关的研究中，被转化为种子节点候选集的求解问题，笼统的理解为节点影响力大小的排序，详见 2.2.2 一节。如今社交网络的规模越来越大，千级用户网络不能够称之为大规模网络，百万级用户网络已不是少数。找到一种在大规模网络快速挖掘种子节点的方法是现实应用亟待解决的。这种方法的计算不仅依靠拓扑网络结构，也要适当依靠用户的在线行为，要与用户的历史操作记录相结合，对社交网络的种子进行“精准计算”与“精准投放”。有研究者认为度量节点影响力要度量朋友的朋友甚至更多层次才能精准发现种子。但这样的方法效率极低，面对大规模网络根本无法适用。所以，快速挖掘的方法是该领域研究的难点，也是解决下文两个问题的重要手段。

(2) 影响不足再追加新一批的种子节点

假设种子节点被全部投放至社交网络，销售方认为该批种子最造成终影响效果与理想转化率存在较大差距。这里，转化率是产品在社交网络推广后，产品销售的增幅指标。达不成指标，可能存在产品推广力度不足、所选种子节点不完美、销售方对种子的投入力度过小等原因。此时销售方追加投入一批种子，有了影响不足情况下种子节点追加的问题。追加情况包含着种子数量多少为宜的问题，即 k 个种子要满足在拓扑结构中的大幅度激活，也要达到销售方所需的影响与刺激效果。追加问题充分结合生活实际，是本项目一个主要的研究点。

(3) 按批次投放的种子节点

传统影响力最大化方法选出给定的 k 个节点后全部投入社交网络，观察影响效果再进行评估。在实际营销中，销售方选拔出一批最具影响力的 k 个用户，往往不会一次性地全部投入网络进行产品推广，而是将最具影响力的 k 个用户分为 n 个批次，即种子节点会被分组、分批投入社交网络中。这样处理在销售中具有实际意义，便于观察分阶段的推广效果，如在信息传播达到疲惫期加强刺激，延续并扩大影响。在形式上，批次投放是种子节点划分批次的问题，而实际它包含种子节点投放次序的问题，关于投放是按照影响力先大后小更容易积累爆发，还是先小后大更有利

于刺激加强，尚无相关研究。因此，分批次投放种子节点的影响力最大化问题将是本项目重点研究内容。

4.2 研究目标

本项目围绕社交网络影响力最大化及种子节点集求解展开研究，预期达到如下目标：

(1) 设计快速挖掘节点潜在影响力大小的计算方法，用于启发种子节点并求解影响力最大化问题。

(2) 针对影响不足再追加新一批种子节点的问题，设计一种基于群体智能算法的求解办法，以启发合适的节点成为新种子。

(3) 在确定的信息传播模型中，规划出种子节点批次投放问题的批次及每批次节点的数量，对比传统一次性全投放，以证明影响效果更强。

4.3 拟解决关键问题

本项目拟解决关键问题包括如下三点。

(1) **设计一种快速挖掘节点影响力并求解最大化问题的计算方法。**节点影响力的计算方法百家争鸣，找到一种快速求解办法是本项目的核心，是解决种子节点投放问题的关键技术。影响力最大化问题有两项评价指标，这里不再赘述。根据评价指标，要求算法时间复杂度尽可能的小，即快速，另一方面，要求节点影响力计算与筛选后得到的种子节点是准确的，造成大面积的有效影响，且能够在真实的网络数据集上开展实验加以验证。

(2) **关于追加问题，如何在部分已被影响或激活的网络图中求解新的种子节点。**该问题是实现种子追加投放问题的关键所在。可以拆分成两个子问题，一是在该类网络中，如何避免种子节点及其激活的邻居节点的重复计算，另外追加节点时，如何设定启发规则，将新一批种子节点在该类网络中选拔出来，其特征应如何表述。

(3) **批次与投放问题的种子节点合理规划研究。**批次投放种子节点是找到全部种子节点并能保证种子节点全部激活条件下，将种子分批、分组投入社交网络的问题。它可以拆解成两个子问题。第一，采用何种模型描述节点有停顿的、可积累的传播过程；第二，如何规划批次，共划分几个批次，每个批次的节点数量是多少。解决好这两个子问题，并确保该方法较比传统一次性完全投入的办法有效。

五、拟采取的研究方法、技术路线及实验方案

5.1 拟采取的研究方法和技术路线

本项目整体技术方案如图 1 所示

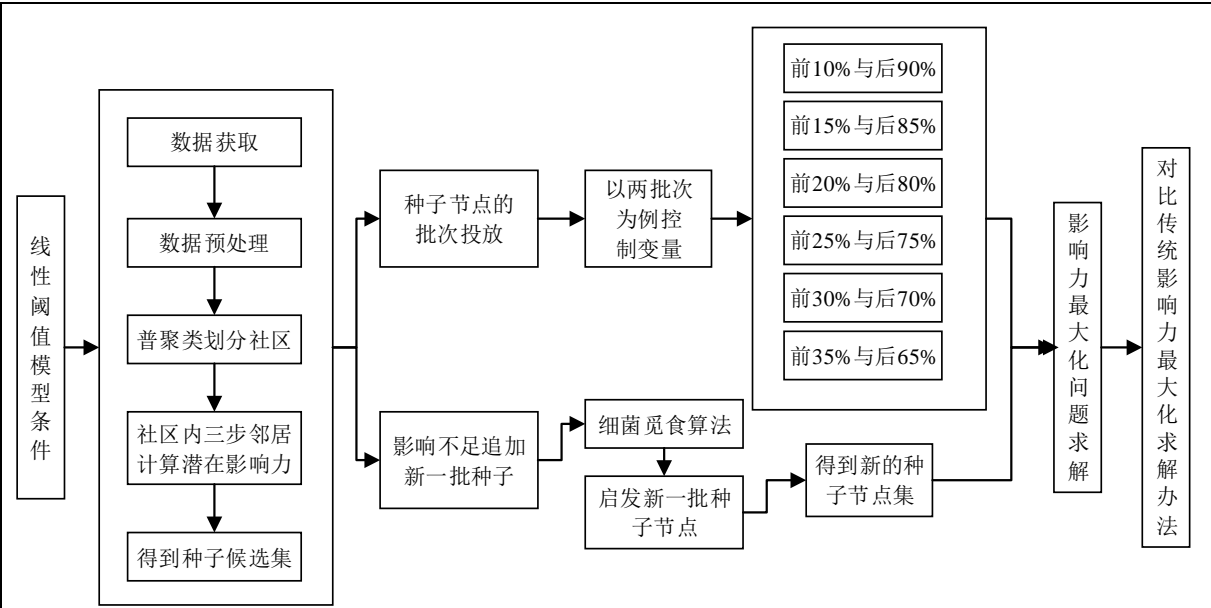


图 1 项目整体技术方案

表 1 是开展研究所需的数据集整理情况。

数据名称	节点	边	数据描述
ArxivPh-Coauthor	6325	28626	Arxiv 的高能物理合作网络
YouTube	22302	30409	在线具有社交功能视频分享网络
Slashdot	20540	50310	可评价喜欢、不喜欢的新闻网
Gowalla	12104	22105	位置信息分享的社交网络
E-mail	1133	5451	某大学师生邮件关系网络
MSN-Blog	3982	6803	MSN 的博客关系网络
PGP	10680	24340	PGP 加密网络
DBLP	14593	32404	DBLP 数据库合作者网络

表 1 实验所需数据整理

(1) 基于社区结构的三步步长法计算影响力

社区是社交网络中成员趋于紧密联系的群，社区类似生活俗称的圈子、团体、社团。之所以被称为社区，是因为社区内个体之间的交往比外部个体的交往更为频繁。

图 2 给出了社区结构的示意图，共有 5 个社区，每个社区内部紧密连接，而社区与社区之间的公共边较稀少。本项目提出一种基于社区结构与有限步邻居以发现种子节点候选集的办法。有限步邻居在社交网络中表示的可能是朋友的朋友。这种算法思想与步骤描述如下：

A) 将社交网络数据采用普聚类方法划分为若干社区，每个社区为一张子图；

B) 在社区结构的子图内以三步邻居步长计算节点潜在影响力；

Step 1: BFS 搜索节点 u 的全部邻居节点；

Step 2: 计算节点 u 的第一步, 第二、三步的节点潜在影响概率;

Step 3: 累加求和得到节点 u 在社区的潜在影响力;

C) 对三步邻居法计算得到的节点按影响力大小排序, 得到种子节点候选集;

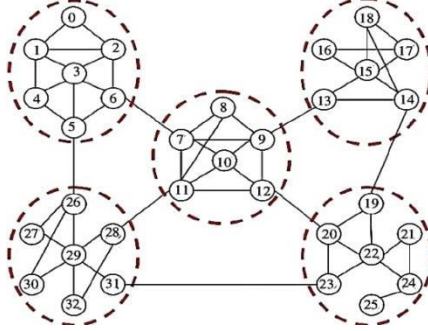


图 2 社区结构示意图¹⁾

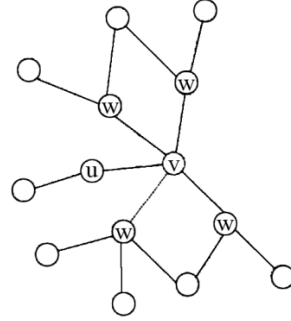


图 3 三步步长计算图示

¹⁾ (图引用: 李浩, 乔东平. 一种复杂集成服务型机械产品的模块化结构建模方法[J] 河北科技大学学报. Vol-34(3):182-187, 2013)

结合图 3, 以 u 为研究节点, 将三步步长节点影响力计算方法形式化定义。 $Inf_i(u)$ 表示第 i 步节点 u 的影响力大小。则:

$$Inf_1(u) = \sum_{v \in N(u)} p_{uv}$$

$$Inf_2(u) = \sum_{v \in N(u)} \sum_{w \in N(v)} p_{uv} p_{vw}$$

节点 u 第三步邻居的影响力计算大小如果依次累加相乘会使算法变得高消耗, 即算法的时间复杂度变高。这里以第三步节点为边界, 先计算节点 w 可能激活的节点数量 $Q(w)$ 。

$$Q(w) = \Phi(w) - \sum_{v \in N(v)} \sum_{v \in N(w)} |v|$$

其中, $\sum_{v \in N(v)} \sum_{v \in N(w)} |v|$ 表示即是节点 v 又是节点 w 的邻居节点数量, $\Phi(w)$ 是节点 w 邻居节点的个数。将第二、三步看作整体, 则 u 对第二、第三步的节点影响力近似为:

$$Inf_{2,3}(u) = Q(w) Inf_2(u)。最终, 节点 u 的潜在影响力 $Inf(u)$ 为:$$

$Inf(u) = \alpha \cdot Inf_1(u) + \beta \cdot Inf_2(u) \cdot Q(w)$, 其中 α 和 β 是两个参数, 用于调整第一步和第二步、第三步的影响概率。第一步的激活是基础, 如果第一步不能激活, 后续的第二第三步很难再被激活。所以引入两个参数进行调整。

采用普聚类方法划分社区是出于算法时间复杂度的考虑。在社区子图内采用三步步长计算节点潜在影响力大小, 是吸收了传统全局网络依据四步及四步以上步长计算节点潜在影响力大小的思想, 将计算量极大简化并限定在社区中, 由于社区内部节点连接十分紧密, 故步长设定为 3 足以描述出节点在社区内的核心位置, 在理论上能够满足节点潜在影响力更大的判断, 且时间复杂度得到了保障。

预期在实验中，比对度中心性、介数中心性及部分典型节点影响力计算方法。再投入至社交网络数据集中，求解影响力最大化问题。

（2）基于改进细菌觅食算法求解影响不足的种子追加问题

已投入一批种子但影响不足情况下全局网络节点已存在种子与被激活节点，根据状态它们都可以被划分为已激活节点，则追加问题的难点在于，如何在这样的“残局”网络中寻找新一批种子节点。本项目拟采用简化后的细菌觅食算法求解该问题。

细菌觅食算法模拟的是细菌吸收影响的趋向、复制、死亡等过程。具有参数少，不易陷入局部最优陷阱等优点。经典的细菌觅食算法将细菌觅食与问题的求解分为“细菌向营养区聚集、细菌彼此吸引、亚健康细菌死亡且健康细菌复制、复制后细菌被随机放置”四个部分，本项目将该算法引入追加问题的求解，简化了细菌觅食向营养区聚集的复杂过程，并舍弃了细菌彼此吸引的繁琐计算，简化后的算法与问题映射关系如图 4 所示。

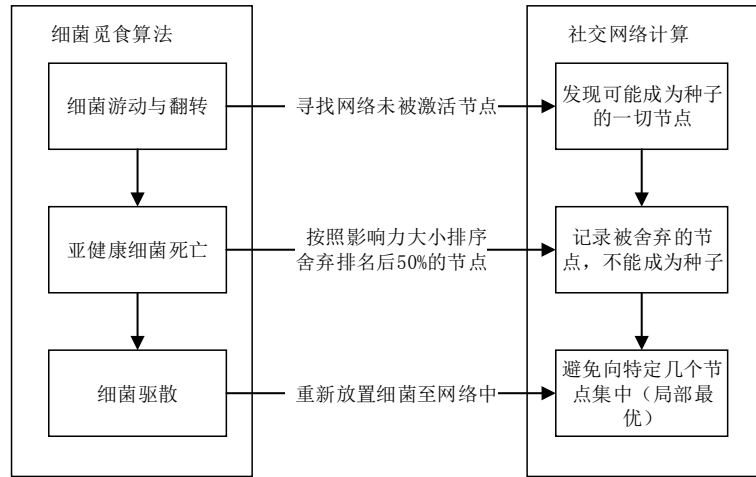


图 4 细菌觅食与种子节点计算的映射关系

细菌的游动与翻转映射节点的 BFS 遍历。细菌健康与亚健康的分类和计算是对节点潜在影响力的估计，其影响力大小按上述 5.1（1）三步步长法评估。复制后的细菌进行驱散，避免了细菌向特定节点集中的局部最优现象。这样按照三步步长方法、节点的激活或未激活状态进，结合改进的细菌觅食算法，在线性阈值模型条件下计算影响力最大化问题。

（3）控制变量法研究分批次投放种子节点的问题

下式将批次投放种子节点问题形式化定义：

$$|S| = \sum_i^n |S_i|$$

其中， $|S|$ 是种子节点个数， $|S_i|$ 表示第 i 批次节点的数量。则该问题可转化为对 n 与 $|S_i|$ 的求解。对该问题设计实验，并采用线性阈值模型，利用其积累的特性保留节点影响力，以保证投放下一批节点，节点影响力仍然存在，而不是一切从新开始。

预计在实验采用线性阈值模型模拟节点间影响传播，以两批次为例，即令 $n=2$ 。种子节点将被分为两组投放至网络中。按照影响力从大到小的排序将种子分别取前 10%、15%、20%、25%、30%、40%、50%，第二批再将对应的余下种子投入。因节点影响力的概率是随机分配的，所以每次得到的结果不尽相同，为消除实验结果存在的偶然性，对每次种子节点的分批投放重复实施 50 次，构造大子样取其平均值。并将节点影响力从小至大排序，重复实验过程，验证种子节点在线性阈值模型条件下，先大后小的投放次序与先小后大的投放次序哪个更能加强刺激，并对比传统一次性全部投放的 50 次重复结果，验证两批次投放的优越性。最后，对影响力最大化最终激活节点数量的问题进行求解。

实验采用的真实数据分为两类，一类是简单随机网络，网络的节点数和边的数量比较少。一类是符合幂律分布的复杂网络，节点数量和边的数量多。预使用的数据集情况见表 1。

六、可行性分析

1. 前期研究

申请人阅读了大量的文献资料，这些资料关于影响传播模型、节点影响力计算、影响力最大化求解与社交网络图论。同时，对基础性研究动手实验，在线性阈值模型、独立级联模型与传染模型环境下利用传统贪心算法实现了基础的影响力最大化求解，掌握以蚁群、蜂群、和声搜索、细菌觅食为代表的群智能算法的理论知识。具备一定基础研究能力。并参与江西省教育厅科学技术研究项目，基于话题感知的社交网络影响最大化研究。

2. 实验条件

申请人所在单位设有高性能计算中心实验室，配备了计算刀片服务器、并行存储系统（含索引控制器和数据控制器）、管理监控网络交换机、胖节点服务器等设备，这些设备为本项目的数据分析和算法测试提供了良好的硬件资源。

且开展试验涉及的数据集是开源免费，在网络中可查询与下载，对数据集的使用也有明确的格式与规范。监督实验的指导教师是该研究方向的资深研究者。

3. 理论论证

本项目涉及到的群智能算法前人已有研究，细菌觅食算法是在国际会议提出的具有典型意义的一类群体智能优化算法，本项目围绕该算法的核心框架进行改进应用于影响不足追加新一批种子的问题。项目中基于社区结构的三步步长计算，是以社会网络分析学科的主流观点，衡量节点影响力要观察对象朋友的朋友的朋友，才能实现准确计算，且信息依靠社区内紧密连接的结构特性在其中能得以快速的传播。

项目的两个问题是依据现实生活与销售实际提出的，具有现实意义。

七、项目创新点与特色

- 1. 项目与生活实际充分结合，首创性提出种子节点批次投放与影响不足的再追加问题。在产品销售推广，舆情控制与传播方面具有极强的现实意义；
- 2. 提出基于社区结构选拔种子节点的三步步长计算方法，摒弃传统全局网络四步及其以上的计算方法，降低时间复杂度并保证有效求解；
- 3. 利用改进细菌觅食算法求解影响力不足追加新一批种子节点的问题。

项目预期成果（发表本项目有关的学术论文）及考核指标（技术、经济指标和社会效益要具有明确的可考核性，1000 字以内）：

- （1）人才培养：培养 1-2 名社会网络分析方向的计算机研究生；
- （2）论文发表：在 CSCD 或以上级别的期刊发表论文 2-3 篇；
- （3）成果表达：提交社交网络影响力最大化的种子节点集研究为题的研究报告。

项目时间进度安排和阶段目标		
	考核时间节点（年/月）	阶段目标（阶段考核指标）
1	2018 年 9 月-10 月	统筹规划，完成项目总体方案论证，包括完成项目的可行性研究、需求分析、项目任务规划，并完成相关文档的编写
2	2018 年 11 月-12 月	完成系统的总体设计和详细设计工作，完成相关文档的编写
3	2019 年 1 月-5 月	设计批次投放种子节点与追加种子节点的实验过程与代码，多次进行实验，并对比结果
4	2019 年 6 月-9 月	整理资料，撰写研究成果报告书，提交结题验收
预期主要成果形式	1. 论文 2. 研究报告 3. 计算机软件著作权一项	

项目负责人主要科研成果：

计算机类主要科研成果：

2017 年，设计开发省心物流信息管理系统，投入江西五环物流有限公司使用。

教育类主要科研成果：

[1] 2015 年，论文《高校毕业生就业跟踪调查手段研究报告》，教育前沿，晏峰，梁文；

[2] 2016 年，论文《以感恩文化体系构建高校校园》，教育前沿，张晓惠，梁文；

[3] 2018 年，论文《牢牢掌握好高校意识形态工作领导权》，在投，梁文，张晓惠；

[4] 2018 年，论文《针对中学生注意力的启发式训练方法研究》，在投，梁文，张晓惠；

同组参与人主要成就：

汤达荣

2017 年江西省数学建模竞赛三等奖。

占炼炼

[1] 天梯赛团体赛国家级三等奖；

[2] 蓝桥杯三等奖；

[3] 华东交通大学“双基杯”一等奖。

项目采取何种方式组织和管理以确保项目完成（200 字以内）

本项目研究者共 4 名，硕士研究生三人，本科生一人。其中，成员汤达荣有过计算机类技术岗位三年的研发经历，最高曾任企业框架设计师。为达成项目预期目标以确保项目顺利完成，项目组拟定四项要求。

首先，每月两次碰头会议对接项目研究进展，将讨论目标一致化，确定化；

其次，积极参与社交网络相关的权威学术会议，吸收先进思想、了解前沿技术。如 2018 年 8 月 5 日预计在哈尔滨工业大学举办的中国中文信息学会的 SMP2018 学术交流会议；

再次，充分发挥指导教师在本学科研究方向的学术优势，指导教师近年来在社交网络影响力计算方面的成果丰硕，有着扎实基础知识与丰富的实验经验；

最后，列举本项目的分工情况，分为“数据收集，数据分析，算法设计，算法实现，实验汇报与论文发表”六个部分，每个部分除组内研究开展外，在算法设计与论文发表阶段要介入指导教师的有效干预，起到监督与纠正作用，确保项目的顺利进行。

指导教师情况（研究方向、学术地位、代表性成果）：

杨书新，男，江西九江，研究方向图数据管理、生物信息学。近年来一直关注图数据的研究及在相关领域应用的最新研究成果，进行图数据及社交网络影响最大化相关方向的研究。在前期图数据相关研究工作中关于社交网络影响最大化开展研究，积累了影响传播模型、节点影响力求解算法的研究基础，取得了一定的成绩。并申请相关技术的国家发明专利。

指导教师主持或参加科研项目（课题）及人才计划项目情况

- [1] 江西省科技厅青年科学基金，20122BAB211035，基于GPU加速的图数据关键词查询技术研究，2012/07-2014/12，2.5万元，已结题，主持
- [2] 江西省教育厅科技项目，GJJ12349，基于图数据挖掘技术的软件故障定位研究，2012/01-2013/12，1.5万元，已结题，主持
- [3] 国家自然科学基金地区项目，61662028，基于全基因组数据的致病SNP位点挖掘，2017/01-2020/12，40万元，在研，参加（排名第二）
- [4] 国家自然科学基金地区项目，41362015，基于不确定数据挖掘的滑坡区域地质灾害危险性评价方法研究，2014/01-2017/12，47万元，已结题，参加（排名第三）

与本项目有关的发明专利情况

- [1] 一种基于图形处理器并行计算的频繁子图挖掘方法，2016.09.07，中国，201310503214.6，授权
- [2] 一种基于商品影响力分析的促销商品组合的求解方法，申请号：201610615590.8，实审
- [3] 基于两阶段启发的社交网络影响最大化问题求解方法，申请号：201610394859.4，实审

发表的期刊论文（近五年）

发表期刊论文15篇（均为第一作者），其中5篇论文被EI检索。

- [1] Complex splitting of context-aware recommendations, International Journal of Computational Science and Engineering (EI 期刊: 20173604113275) ,2017, 15(1-2):32-40
- [2] A Heat Diffusion Model Based Algorithm for Personalized Influence Maximization in Social Network, Journal of High Speed Network (EI期刊: 20171603585029) , 2017,23(2):137-147
- [3] 图数据关键词查询研究进展, 电子学报, 2014, 42 (11): 2260-2267 (检索号: 20145300388104)
- [4] A GRAPHICS PROCESSING UNIT-BASED PARALLELALGORITHM FOR KEYWORD SEARCH ON A DATA GRAPH, Journal of Computational Information Systems, 2015, 11 (20):

7431-7444 (检索号: 20155101706526)

- [5] Results clustering for keyword search over relational database, Journal of Software, 2013, 8 (12): 3188-3193 (检索号: 20140117152927)
- [6] 基于两阶段启发的社交网络影响最大化算法, 小型微型计算机系统, 2017.38 (10): 2268-2274 (CSCD核心)
- [7] 基于影响路径的个性化影响最大化算法, 计算机工程与科学, 2016, 38 (6): 1128-1134 (CSCD核心)
- [8] 一种改进的基于交叉矩阵统计的软件故障定位技术, 计算机工程与科学, 2016, 38(4): 679-685 (CSCD核心)
- [9] 基于贝叶斯方法与聚类的上下文感知推荐, 小型微型计算机系统, 2015, 36 (10): 2262-2265 (CSCD核心)
- [10] 一种基于图结构分解的图近似查询方法研究, 计算机工程与科学, 2015, 37 (2): 359-364 (CSCD核心)
- [11] 基于图挖掘和决策树的软件故障定位研究, 计算机工程与应用, 2015, 51 (20): 67-71 (CSCD核心)
- [12] 并行子树约简方法在软件故障定位中的运用, 计算机应用研究, 2014, 31 (7): 2097-2100 (CSCD核心)
- [13] 结合查询相关性的关键词查询排序方法, 计算机工程与设计, 2013, 34(9): 3136-3140 (CSCD核心)
- [14] 基于数据图的关系数据库关键词查询排序研究, 计算机应用研究, 2013, 31 (2): 440-442, 447 (CSCD核心)
- [15] 基于内容和结构的数据库关键词检索排序策略, 情报杂志, 2013, 32 (3): 127-131 (CSCD核心)

项目主要人员情况				
	姓 名	院 系	专 业	在项目中具体分工
负责人	梁文	信息工程学院	计算机技术	算法设计
项目 成员	汤达荣	信息工程学院	计算机应用技术	数据收集与分析
	朱凯丽	信息工程学院	计算机技术	算法实现
	占炼炼	信息工程学院	计算机科学与技术	算法实现
九、项目经费情况				
支出科目		金 额（元）	计划根据及理由	
材料费		1000	打印机、办公用品等消耗品	
会议费/差旅费		5000	用于学术交流、参加学术会议等	
出版物/文献/信息传播费		4000	论文版面费等	
合计		10000		

十、审查意见	
指导教师 意见	<p>申请人在申请前期已做大量基础性研究，对传统影响最大化算法有着实验基础。学习期间积极参与课题的研究，并阅读了中外的相关文献，了解前沿技术与研究热点。目前，申请人已完成数据集收集与整理工作。申请人在项目中清晰具体地说明了实验方案，提出的方案具有可行性，预期可以达到一定的研究成果。若项目获批，本人将集中精力加强指导教育、有效干预实验开展，提供必要的人力与物力支持。</p> <p style="text-align: center;">同意推荐</p> <p style="text-align: right;">指导教师签名： 年 月 日</p>
培养单位 意见	<p style="text-align: right;">负责人签章： 年 月 日</p>
专家组 评审意见	<p style="text-align: right;">专家组长签名： 年 月 日</p>
省级主管 部门审核 意见	<p style="text-align: right;">签章： 年 月 日</p>
备注	