

社会网络节点影响力分析研究^{*}

韩忠明^{1,2}, 陈炎¹, 刘雯¹, 原碧鸿¹, 李梦琪¹, 段大高¹



¹(北京工商大学 计算机与信息工程学院, 北京 100048)

²(食品安全大数据技术北京市重点实验室, 北京 100048)

通讯作者: 韩忠明, E-mail: hanzm@th.btbu.edu.cn

摘要: 社会网络节点影响力研究是社会网络分析的关键问题之一. 过去的 10 多年间, 随着在线社会的快速发展, 研究人员有机会在大量现实社会网络上对影响力进行分析和建模, 并取得了丰硕的研究成果和广泛的应用价值. 分析和总结了近年来社会网络影响力分析的主要成果. 首先介绍了节点影响力的相关定义、作用范围以及表现形式; 接着, 重点分类介绍了节点影响力的度量方法, 通过网络拓扑、用户行为和内容分析这 3 类方法总结了影响力的建模和度量方法; 然后总结了影响力的传播和最大化模型相关成果; 最后介绍了影响力的评价指标和应用. 根据对现有方法的系统总结, 对社会网络影响力的未来研究提出了一些值得关注的方向.

关键词: 社会网络; 节点影响力; 关键节点; 影响力最大化; 信息传播

中图法分类号: TP393

中文引用格式: 韩忠明, 陈炎, 刘雯, 原碧鸿, 李梦琪, 段大高. 社会网络节点影响力分析研究. 软件学报, 2017, 28(1): 84-104. <http://www.jos.org.cn/1000-9825/5115.htm>

英文引用格式: Han ZM, Chen Y, Liu W, Yuan BH, Li MQ, Duan DG. Research on node influence analysis in social networks. Ruan Jian Xue Bao/Journal of Software, 2017, 28(1): 84-104 (in Chinese). <http://www.jos.org.cn/1000-9825/5115.htm>

Research on Node Influence Analysis in Social Networks

HAN Zhong-Ming^{1,2}, CHEN Yan¹, LIU Wen¹, YUAN Bi-Hong¹, LI Meng-Qi¹, DUAN Da-Gao¹

¹(School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China)

²(Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing 100048, China)

Abstract: Research on the influence of social network nodes is one of the key issues in social network analysis. Over the past decade, with the rapid development of online social networks, researchers have the opportunity to analyze and model node influence on many real social networks to achieve fruitful research results which can be applied in a wide range of applications. This paper analyzes and summarizes the main research efforts of social network influence analysis in recent years. First, different definitions of influence, influence functional scope and forms of influence are introduced. Next, models and methods of measuring of node influence are discussed and analyzed in detail with respect to network topology, user behavior and content analysis. Then, literatures about influence spreading and influence maximization model are summarized. Moreover, different indexes for evaluating influence methods are compared, and applications related influence are also presented. Finally, some future research directions on influence analysis are suggested based on the review and analysis of existing research efforts.

Key words: social network; node influence; key node; influence maximization; information propagation

影响力的研究早在 20 世纪初期就受到了社会学家和心理学家的关注, Triplet^[1]在研究社会促进论时, 发现

* 基金项目: 国家自然科学基金(61170112); 教育部人文社会科学研究基金(13YJC860006)

Foundation item: National Natural Science Foundation of China (61170112); Humanities and Social Sciences of Ministry of Education Planning Fund (13YJC860006)

收稿时间: 2016-01-04; 修改时间: 2016-04-25; 采用时间: 2016-06-10; jos 在线出版时间: 2016-10-11

CNKI 网络优先出版: 2016-10-12 16:26:48, <http://www.cnki.net/kcms/detail/11.2560.TP.20161012.1626.013.html>

人们在受到他人关注时会表现得更加突出.到了 20 世纪 50 年代,Kate 等人^[2]发现:无论是在日常生活中还是在政治选举等方面,影响力都起着至关重要的作用.近年来,随着微博等大规模社会网络的兴起,六度分割理论^[3]、四度分割理论^[4]以及小世界理论^[5]等都说明人与人之间的距离变得越来越短,联系也越来越紧密,社会网络为影响力研究提供了实验平台和大量数据.早期的政治家用影响力来赢得选举,商人用影响力来推销商品.社交媒体时代,社会网络的无标度性质^[6]注定了社会网络中少部分人掌握着大部分的话语权.微博意见领袖利用自身的影响力在网上呼风唤雨,很多热门话题、突发事件都能看到他们的“身影”,影响力发挥着越来越显著的作用.对社会网络节点影响力进行分析、度量、建模以及传播的相关研究具有重要的理论和实践价值.

国内外学者在影响力的研究相关领域涌现了大量杰出成果,目前已有学者对此进行了一些总结.Sun 等人^[7]对社交网络拓扑结构的一些统计测量指标以及相似度与影响力的关系进行了综述.AlFalahi 等人^[8]对影响力概率模型进行了综述.刘建国等人^[9]从网络结构和传播动力学角度综述了节点重要性排序.吴信东等人^[10]对社会网络的基础结构、影响力度量以及影响力传播等作了比较全面的综述.有 3 个原因促使我们对社会网络节点影响力的文献进行重新梳理和总结.

- (1) 节点影响力的表现形式多种多样,如何定义节点影响力,节点影响力的作用范畴有哪些,现有综述文献对节点影响力相关定义的梳理和总结较少;
- (2) 虽然社会影响力没有统一的评价指标,但常见的评价方法却有迹可循,现有文献针对影响力的不同定义具有不同的评价.节点影响力的评价需要进行总结;
- (3) 社会网络节点影响力的研究发展得非常快,近几年已出现不少新的重要的研究成果,因此也需要对社会网络节点影响力相关研究进行更加全面的总结.

本文第 1 节介绍社会网络节点影响力的定义和表现形式.第 2 节分别从用户的网络拓扑结构、用户的内容和行为特征等角度介绍影响力的量化指标和模型.第 3 节介绍影响力的传播模型和影响力最大化研究.第 4 节介绍影响力的常见指标.第 5 节提出具有挑战性的问题和进一步的研究方向.最后,第 6 节对全文进行总结.

1 社会网络节点影响力的定义

政治家利用影响力赢得选举,商人利用社会网络上口口相传(word-of-mouth)的影响力将商品推销到整个社会网络,社会舆论的引导和创新理论的传播等都可以借助于社会网络上具有高影响力的个体用户.影响力到底是什么?影响力可以定性分析也可以定量分析,影响力具有不同的作用范围.到目前为止,影响力并没有统一的形式化定义和标准的计算方法.

早期的社会学家都是定性分析影响力.社会学家 Rashotte^[11]把影响力定义为个体在与他人或群体的互动中,导致自身的思想、感觉、态度或行为发生变化的现象.节点影响力是导致一个人在与比他优秀或和他拥有相同爱好的人进行交流时,其行为发生了改变的原因.

Katz 等人^[2]在研究美国总统选举中选民投票意向时,将少部分影响力大的个体定义为“意见领袖(opinion leaders)”,他们影响了大多数普通民众的投票意向,并提出了二级传播理论,以说明个体的影响力存在一定的差异性.Anagnostopoulos 等人^[12]将社会网络上的用户分为权威用户和普通用户,并研究了这两类用户在信息传播中的影响力.Yang 等人^[13]在研究信息扩散时,将社会网络中的用户分为 3 类角色:意见领袖、结构洞(structural hole)和普通用户.Granovetter^[14]和 Krackhardt 等人^[15]提出的社会学经典理论“弱连接”理论说明:不同的连接关系对节点影响力的贡献存在差异,并且弱连接对节点的影响力所产生的作用优于强连接.

社会网络的出现为定义和研究节点影响力提供了定量基础,定量度量节点影响力需要构建一个可测量的指标.个体与个体之间通过各种关系连接形成社会网络拓扑结构.如科学家与科学家合作形成了科学家合作网络,论文与论文之间的引用关系形成了引文网络,微博用户通过关注行为形成了关注网络等.直观分析,社会网络中的节点重要性排序指标可以用来度量节点影响力.节点的度中心性^[16]、介数中心性^[17,18]、紧密中心性^[19]、聚类系数^[5,20,21]等都能一定程度地表示节点影响力.研究者通过 K -核分解^[22]将拓扑结构中的节点从边缘到核心分解到不同层次中.PageRank^[23,24],HITS^[25],LeaderRank^[26]等随机游走算法则可以用对节点打分的排序结果

来区分节点影响力的大小.

影响力可以表达为一个个体的特性,也可以表达为个体之间的作用形式,所以影响力具有全局和局部范围.社会学家定性地分析影响力,指出利用网络统计指标得出的影响力都属于全局影响力.Cha 等人^[27]根据用户在社会网络上的行为特征和交互信息的统计指标来表示影响力,如粉丝数量、该用户的转发数量和提到该用户的次数等.Tang 等人^[28,29]认为,节点在不同话题上的影响力也有所不同,将影响力视为一个隐变量.这些文献将影响力在对象和作用范围上加以区分,都属于局部影响力.

定性影响力的研究将影响力表现为分类结果^[2,12,13]、排序结果^[23-25]或节点之间是否存在影响的概率^[30,31],定量影响力则将影响力通过可测量的大小来体现,如随机变量^[32]、统计指标^[27]、行为传播数量^[33,34]等.社会网络上信息传播速度和范围与用户影响力息息相关,因此,用户信息在社会网络上的传播速度和覆盖范围可以用来表示节点的传播影响力,尤其是在病毒式营销、舆论引导等实际应用上.通过总结不同领域研究者们定性和定量研究影响力的成果,我们用表 1 给出节点影响力的各种表现形式.

Table 1 Representations and representative methods of influence

表 1 影响力的表现形式与代表方法

| 影响力 | 表现形式 | 范围 | 代表性指标或方法 |
|------|-----------|-------|--------------------------|
| 定性分析 | 分类的结果 | 全局 | 两级传播理论,弱连接与强连接理论 |
| | 网络拓扑上的重要性 | 局部/全局 | 介数,紧密中心性,K-核分解等 |
| | 相对排序的结果 | 全局 | PageRank,LeaderRank,HITS |
| 定量分析 | 用户间的影响概率 | 局部 | IDPM,TOIM,ICM,等 |
| | 某种分布的随机变量 | 全局 | LIM,LDM,等 |
| | 话题相关隐变量 | 局部 | TAP,HF-NMF,等 |
| | 统计指标 | 全局 | 度,Cha 等人提出的 3 种影响力 |
| | 传播时激活节点数量 | 全局 | LTM,ICM,等 |

表 1 中的方法说明如下:

- IDPM(influence diffusion probability model)为影响力扩散概率模型;
- TOIM(topic-level opinion influence model)为话题观点影响力模型;
- HF-NMF(hybrid factor non-negative matrix factorization)为混合因子非负矩阵分解模型(见第 2.2.1 节);
- LIM(linear influence model)为线性影响力模型(见第 2.2.2 节);
- ICM(independent cascade model)为独立级联模型;
- LTM(linear threshold model)为线性阈值模型(见第 3.1 节);
- TAP(topical affinity propagation)为局部亲和力传播模型(见第 2.2.1 节).

分析节点影响力相关定义与表现形式可以看出,节点的全局影响力越大,节点对信息、行为在整个社会网络中的传播控制能力越强,社会网络中一小部分最具影响力的节点能够控制整个社会网络中大部分的传播.而一个节点对另一个节点的影响力则属于局部影响力,节点对另一个节点的影响力越大,后者在社会网络中就越会追随和模仿前者的行为.从定量形式度量节点局部影响力角度出发,针对不同应用的整体要求,结合局部影响力和网络结构来定义节点影响力,能够取得较好的效果.

2 社会网络节点影响力的度量

社会网络的拓扑结构、用户交互行为、用户内容构成了社会网络的 3 个要素.拓扑结构能够从宏观层面上刻画节点的影响力,也容易获取,复杂网络中的拓扑结构指标相对成熟,因此,用拓扑结构来度量节点的影响力成为一种常见的做法.然而网络拓扑结构中的连边无法描述节点间的复杂交互关系.如在微博中,用户间不仅有关注关系,还存在着转发、评论等关系,这些交互行为的频率也不同.两个亲密好友和两个萍水相逢的好友在网络拓扑结构中被“一视同仁”,用户行为与交互信息则能很好地反映出用户影响力的形成与变化细节,所以也有研究者综合利用两者的优点进行影响力的度量.社会网络是用户交互的基础,而用户交互的内容则是用户活动

的根本.直观分析,不同领域的用户在各自领域的影响力有所不同,所以也有研究者利用内容来分析影响力.综合3个可能的度量影响力的维度,我们用图1来说明利用社会网络不同角度的特征与用户影响力各种研究之间的关系.

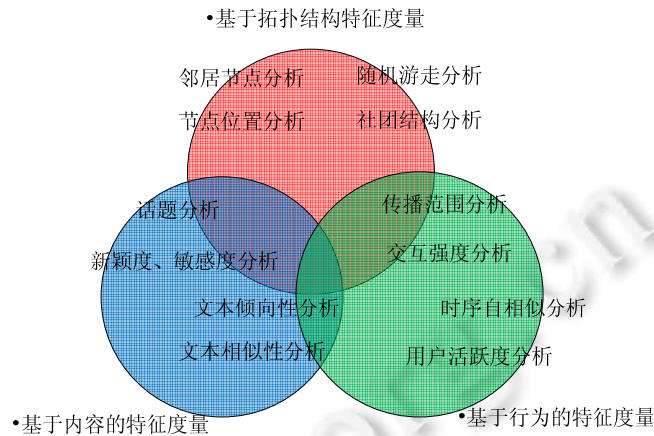


Fig.1 Comparison of measurement methods of influence

图1 影响力度量方法对比

从图1可以看出,从拓扑结构、行为特征、内容特征这3个维度来度量影响力具有不同的研究角度和典型方法.我们将分别从3个维度将相关成果分成3类进行综述和分析.

2.1 基于拓扑结构的度量

最早利用网络拓扑结构度量节点影响力的是社会学相关领域的研究者,随后,其他领域的研究者也进行了研究和改进.本节将从基于局部属性、基于全局属性、基于随机游走、基于社团关系这4个角度进行介绍.

一些社会网络的基础指标在文献[35]中有更为详细的介绍,本文则着重介绍近几年来和影响力紧密相关的研究进展.

2.1.1 基于局部属性的度量

基于局部属性的度量指标最常见的为度中心性(degree centrality)^[16],度定义为节点的邻居节点的数量.度中心性反映的是在整个网络中当前节点的直接影响.例如,微博中拥有大量粉丝的用户可能有更大的影响力,然而只考虑度不考虑节点在网络中的位置显然是不可取的^[36].Chen 等人^[37]提出了局部中心性(local centrality)指标,该指标综合考虑了节点度和其邻居节点的度信息.研究者^[38,39]发现:当网络中的传播率较小时,度中心性对于节点的传播影响力效果较好;而当传播率在达到临界值附近时,特征向量中心性的度量效果较好.此外,Idé 等人^[40]也从动力学角度进行了优美的解释.闵磊等人^[41]在此基础上通过对当前节点的邻居节点度的累加进一步扩展了节点度,提出了扩展度(ExDegree)指标,并分析在不同传播率下,适合信息传播的所需拓展的层数.Fowler 等人^[42]提出的三度影响力原则认为:节点不仅可以影响邻居节点(一度),还能影响邻居节点的邻居节点(二度),甚至能影响到邻居节点的邻居节点的邻居节点(三度),只要在三度以内都属于强连接关系,都有引发行为的可能性.如果超过三度以上,节点相互间的影响则会消失.以上基于度拓展的指标与Fowler 等人的思想不谋而合.基于度中心性及其改进的度量指标简单、直观且时间复杂度低,适用于大规模网络.然而,这类指标只是从节点可能影响其他节点的数量上考虑节点的影响力,并没有考虑影响其他节点的强度之间的差异性,且没有考虑节点在整个网络的位置.

社会网络中,联系紧密的多个好友形成社团的现象在社会网络中很常见.局部聚类系数(local clustering coefficient)用于衡量节点的邻居节点之间联系的紧密程度.局部聚类系数(以下简称为聚类系数)等于节点 v_i 的邻居节点之间连边的数量与邻居节点之间可以连边的最大数量之比^[5].无向图聚类系数计算公式如下所示:

$$C(v_i) = \frac{2|\{e_{jk} : v_j, v_k \in N_{v_i}, e_{jk} \in E\}|}{k_i(k_i - 1)},$$

其中, k_i 为节点 v_i 指向其他节点的连边数量与其他节点指向 v_i 的连边数量的和, 因为在无向图中, 边不区分方向, 因此除以 2. 有向图聚类系数公式如下:

$$C(v_i) = \frac{|\{e_{jk} : v_j, v_k \in N_{v_i}, e_{jk} \in E\}|}{k_i(k_i - 1)}.$$

图 2 给出一个具有 4 个节点的网络实例, 在 3 种结构下, 蓝色节点的聚类系数分别为 0, 1/3 和 1. 研究者^[20,21]结合节点与邻居节点的度和节点的聚类系数度量节点的影响力, 研究发现: 节点的聚类系数对于节点的影响力促进并不明显, 甚至起到抑制作用. Centol^[43]在研究在线社会网络中的行为传播与节点的聚类系数关系时发现, 在社会网络聚类系数较高时, 节点行为的传播速度往往更加迅速. 节点的影响力与节点的聚类系数相关联. Ugander 等人^[44]研究 Facebook 中好友关系的形成与演化规律, 发现节点影响力取决于节点的邻居节点之间形成的连通子图的多少而不是邻居数量的多少. Cui 等人^[45]从信息传播的持续性角度进行研究, 发现: 当信息传播持续较长时间, 信息在树形网络上比在晶格网络上的传播范围更广, 这就意味着聚类系数大的节点并不一定能促进信息的传播, 即, 高聚类节点并不一定影响力大. Ugander^[44]和 Mislove 等人^[46]的研究工作表明: 节点的聚类系数对于节点的影响力是起负面作用的, 聚类系数高的几个关系密切的好友形成了紧密朋友圈. Chen 等人^[21]在科学家合作网和短消息网络上进行了实验, 验证了聚类系数对于节点获取新的邻居节点是起负面作用的, 并结合节点度与聚类系数提出了节点影响力度量模型 ClusterRank. 模型中, 节点聚集系数越高, 节点的影响力越小. Zhang 等人^[47]研究了学术网络中节点的词语使用受他人影响的情况, 在此基础上提出了 Role-Conformity Model(RCM)模型度量节点受其他节点的影响趋势, 发现: 度值大但聚类系数较小的节点易受其他节点的影响. 将邻居间的关系作为影响力的相关因素, 模型提高了精度, 但时间复杂度却有所增加.

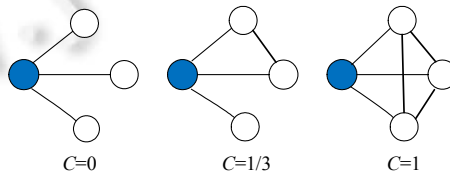


Fig.2 Local clustering coefficient

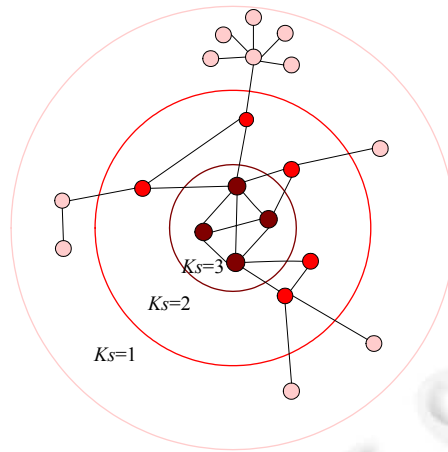
图 2 局部聚类系数

2.1.2 基于全局属性的度量

基于节点全局属性的节点影响力度量指标主要考察节点所在网络的全局网络信息, 这些指标能够较好地反映节点的拓扑特性, 但时间复杂度较高, 多数指标不适用于大规模网络.

介数中心性(betweenness centrality)^[17,18]定义为网络中两个节点之间的最短路径经过当前节点的次数, 介数中心性描述的是信息在社会网络中传播时经过该节点的频率. 该指标值越大, 表示在网络拓扑中该节点越繁忙. 若移除介数大的节点, 则会造成网络拥堵, 不利于信息传播. 紧密中心性(closeness centrality)^[19]则衡量节点达到其他节点的速度, 该指标值越大, 表示当前节点到达另一节点的路径越多且路径长度较短. 该指标可以衡量节点对其他节点的间接影响力. 特征向量中心性(eigenvector centrality)^[48,49]是度量节点全局影响力的一个重要指标, 特征向量中心性不仅考虑邻居节点的数量, 还考虑邻居节点的重要性, 将单个节点的影响力看成其他节点影响力的线性组合. 与特征向量中心性类似, 考虑到邻居节点的不同重要性的还有 Katz 中心性^[50], 但是该方法的相关最优权重系数需要大量实验才能获得, 因此存在一定的局限性.

Kitsak 等人^[22]通过对社交网络、邮件网络等实证研究后发现, 高介数或者 Hubs 节点不一定是影响力最大的节点, 并利用 K -核分解将节点在位置上从边缘层到核心层分为不同的层次, 认为核心节点(K_s 值大的节点)才是影响力大的节点. 如图 3 所示, 通过迭代地减去除去小于等于 K 的节点, 节点被分为 3 层, 其中, K_s 为 3 的节点属于核心节点, 即影响力大的节点. K_s 值为 1 的节点属于边缘层, 影响力较小.

Fig.3 Schematic diagram of K -shell decomposition图3 K -shell 分解示意图

K -核分解的提出,给广大研究者带来了很大的启发.近几年,不少学者针对 K -核分解的一些缺陷进行了改进,进一步提高了它的准确性和适用范围.Liu 等人^[51]发现,具有相同 K_s 值的节点之间影响力差异也有可能很大,随后给出了相同 K_s 值的不同节点的影响力排序结果.Zeng 等人^[52]考虑了 K 核分解的剩余度,提出了一种 K -核分解的改进方法.Moreno 等人^[53]在研究社会网络上的谣言传播时发现, K -核分解的效果并不好.Borge-Holthoefner 等人^[54]在研究节点活跃性时,不同的活跃性分布和分配方式对 K -核分解效果的影响非常明显.Liu 等人^[55]分析了不同真实网络中通过 K -核分解得到的核心结构后发现:由于社会网络中存在紧密连接的小团体,造成了许多社会网络中“伪核心”的存在,考虑到局部小团体存在较低的信息熵,提出了基于网络连接熵的影响力度量指标. H 指数是常见的评价学者或期刊影响力的指标,Lü 等人^[56]研究发现, H 指数在度量社会网络中的节点影响力时同样具有很好的效果.文献[56]巧妙地证明了度、 H 指数以及核数之间的关联性. H 指数对节点影响力的度量效果比度和节点核数有了明显的提升.

2.1.3 基于随机游走的度量

基于随机游走的影响力度量典型的方法有 PageRank^[23],HITS^[25]以及 LeaderRank^[26]等.将节点间的连接看成网页间的链接,可以计算节点的 PageRank 值,这样进行排序就能度量节点影响力.

由于孤立点和不连通子图的存在,原始的 PageRank 存在排序结果不唯一的缺陷.Liu 等人^[26]对此进行了改进,提出了 LeaderRank 算法.在原有网络上增加一个对所有节点双向连接的节点,解决了排序不唯一的问题.Li 等人^[57]通过加权对 LeaderRank 算法进行了改进.HITS 算法则是同时考虑到节点的中心和权威性的排序方法.基于随机游走的影响力度量方法用邻居节点来刻画节点的影响力,虽然避免了噪声,但却忽略了节点自身的性质.

2.1.4 基于社团结构的度量

Granovetter^[14]和 Krackhardt 等人^[15]提出的弱连接经典理论表明,人与人的人际关系存在“弱连接”和“强连接”之分,Granovetter 用认识时间的长度、交互的次数、亲密程度等定义了个体间连接的强度.例如:两个亲密交往的朋友、天天交流工作想法的同事都属于“强连接”关系,而只有偶尔见几次的同学则属于弱连接.Granovetter 认为:弱连接优于强连接,因为弱连接能够在不同群体之间起到“桥”的作用.从社会网络拓扑图来看:关系密切的好友形成社团,弱连接对应着社团之间的稀疏连接,而强连接则对应着社团内部的紧密连接.Shi 等人^[58]研究在各个论坛数据中,用户的参与行为模式与其所参与社团的社团结构和信息传播影响力密切相关.赵之滢等人^[59]在利用社团划分算法对社会网络进行划分后,根据节点所连接的社团数量,在已有节点影响力度量指标上,为了进一步区分节点影响力提出了节点的 V_c 指标.

如图4所示,图中21个节点被分为4个社团,2号连接了4个社团,而5号节点位于1个社团的内部.节点2的 V_c 值为4,节点5的 V_c 值为1,因此,2号节点的影响力比度数同样为5但只在一个社团内部的5号节点影响力要大.胡庆成等人^[60]提出了KSC(K -shell and community centrality)影响力模型,将节点的度数、紧密度、介数和 K -shell等中心化指标定义为节点的内部属性,将节点及其邻居节点所属的社团结构定义为节点的外部属性,综合节点的内部属性和外部属性定义节点的影响力.KSC模型中,用当前节点所在社团的联系紧密程度和社团的规模定义了节点的外部影响力.

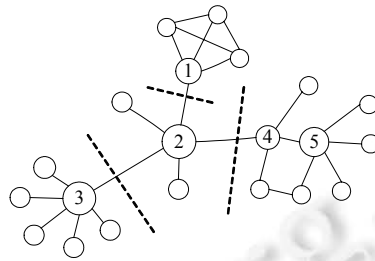


Fig.4 Relationship between node influence and community structure

图4 节点影响力与社团结构的关系

Burt等人^[61,62]提出的“结构洞(structural holes)”是经典的社会学理论.由于结构洞的存在,一些充当中间人的节点可以获得相比于其邻居节点更高的网络收益,即,这些中间节点的重要性更大.以图4中的节点为例,节点2位于结构洞位置.节点2在3个社团间充当“中间人”,因此,节点2在信息控制方面拥有更大的优势.若在节点1、节点3和节点4之间产生连接关系,则节点2的控制能力会大为降低.苏晓萍等人^[63]在结构洞理论的基础上提出了 N -Burt模型来寻找网络中最具影响力的节点,模型考虑节点的邻域结构洞性质,在节点影响力度量中体现出节点所处社团的中心性和连接不同社团的“桥接”性.韩忠明等人^[64]利用结构洞性质,通过ListNet的排序学习方法,有效地融合了包括网络约束系数在内的多种度量指标,在关键节点排序上取得了较好的效果.Lou等人^[65]研究表明:在Twitter上,1%的结构洞节点却能控制着25%的信息传播.Yang等人^[13]在研究信息扩散时将用户分为3类角色:意见领袖、结构洞和普通用户,并分析了当用户作为不同角色在信息扩散时发挥的作用,意见领袖对其粉丝的传播影响力是普通用户对其好友的10多倍,而结构洞节点在不同的群体之间起到桥接作用,少量结构洞节点加入信息传播过程中能促使信息的传播范围迅速扩大.

基于社团结构的节点影响力指标不仅考虑了节点的邻居节点,还考虑了邻居节点的社团性质,优点是个体与群体之间的影响力体现出来,但是对于度量的结果依赖于社会网络的社团性质和社团划分算法,对于社团结构不明显的社会网络,其度量效果并不好.

利用拓扑结构度量节点影响力是最基本的影响力度量方法.这类方法拥有多学科的理论基础,从整个社会网络宏观层面上取得了很好的效果,部分度量指标简单、易算,在大规模网络上拥有较大的优势.然而虚拟网络节点间的关系与真实世界中个体间关系存在很大差异,例如在真实世界中,个体都拥有不同的属性,而在拓扑结构中,每个节点都是一样的,没有区分.两个密切联系的好友和偶然联系了一次的关系在拓扑结构中都是一条边.显然,社会网络拓扑结构只能从宏观层面表示整个社会网络,并不能刻画微观层面上节点对其他节点产生影响力的形成以及演化规律.社会网络拓扑结构对于节点本身的行为和节点对其他节点多种形式的交互行为的利用太少.例如,拓扑结构将微博中一个用户对另一个用户的关注看作一条边,然而微博中用户间还存在着转发、评论以及交互的时间等多种因素.用户对一个每天积极转发评论的粉丝的影响力和对一个只是关注却从不转发评论的粉丝的影响力显然是不一样的.基于网络拓扑结构的影响力度量指标对比以及优缺点可见表2.

Table 2 Influence metrics based on network topology
表 2 基于网络拓扑结构的影响力度量指标

| 指标或方法 | 优点 | 缺点 |
|------------------------|--------------------------------|------------------------|
| Dgree | 简单直观,计算方便 | 只能反映节点局部特征 |
| ExDgree | 对度进行扩展,比度准确性高 | 未考虑邻居节点的位置 |
| Clustering coefficient | 考虑了邻居间的紧密关系 | 不能找到全局影响力大的节点 |
| ClusterRank | 结合了度和聚类系数的优点,进一步提高了准确率 | 不适用于树形网络 |
| Betweenness | 能够找到信息负载能力高的节点 | 不适用于大规模网络 |
| Closeness | 可以计算节点间间接影响力 | 计算节点全局影响力时算法复杂度太高 |
| Eigenvector | 能够体现邻居节点的重要性 | 简单的线性叠加,不考虑结构 |
| Kate | 区分不同邻居对节点的影响力 | 需要大量实验确定权重衰减因子 |
| Ks | 能够表达节点全局网络的位置中心 | 不适用于树形网络 |
| H-index | 影响力度量效果比度和 Ks 明显提高 | 多个 H-index 相同的节点差异可以较大 |
| PageRank | 大规模网络全局排序准确性高 | 忽略了节点自身属性,排序结果不唯一 |
| LeaderRank | 比 PageRank 准确性高,且排序结果唯一,抗噪声能力强 | 不适用于无向网络 |
| HITS | 综合了节点中心性和权威性的优点 | 抗噪声能力差 |
| Vc | 体现了个体对群体的影响力 | 依赖其他指标和社团结构划分算法 |
| KSC | 结合了节点的位置中心性与连接社团的优点 | 不适用于树形网络结构 |

2.2 基于内容与行为特征的度量

2.2.1 基于信息内容的度量

社会网络中不仅有用户间的链接关系,还有用户发布信息的文本内容,信息内容是影响力传播的载体,结合用户的信息内容则有助于分析影响力促进信息传播背后的机理。

Matumura 等人^[66]提出了基于回复信息内容的影响力传播模型 IDM(influence diffusion model),若用户 y 对用户 x 发布的内容进行了回复,则 y 受到 x 影响的程度由两者所发布的内容的相似性所决定.计算公式如下:

$$i_{x,y} = |w_x \cap w_y| / |w_y|,$$

其中, $i_{x,y}$ 代表用户 x 对用户 y 的影响力, w_x 为用户 x 所发布信息内容的信息条目(词语)集合, w_y 为用户 y 对用户 x 所发布信息的回复内容的信息条目集合.若用户 z 继续对用户 y 的内容进行了回复,则用户 x 对用户 z 的间接影响力定义为

$$i_{x,z} = (|w_x \cap w_y \cap w_z| / |w_z|) i_{x,y}.$$

Matumura 等人通过用户间的回复内容相似性来度量用户间的直接和间接影响力,对用户的局部影响力有较好的度量效果.然而, IDM 模型并没有对用户回复的结构进行分析,存在着一定的缺陷。

段松青等人^[67]则根据论坛中帖子回复者的倾向性或情感转变程度决定发帖者对回复者的影响程度,提出了基于倾向性转变的用户影响力排序模型(TTRank),模型度量用户在发帖过程中对他人产生的影响力以及受他人的影响程度,最终得到用户影响力排名结果.樊兴华等人^[30]分析了 IDM 模型中可能存在的回复链中断和内容传递中断以及虚假回复等问题,对 IDM 模型进行了改进,提出了影响力扩散概率模型(influence diffusion probability model,简称 IDPM),并将模型应用于意见领袖的发现中.根据信息内容相似性以及内容的情感态度转变等来度量用户间的影响力.基于内容的影响力度量模型能够分析出用户对他人影响所表示出的具体形式,但对数据文本内容的要求较高,不利于大规模社会网络的用户影响力的度量。

在大规模的社会网络中,用户所发信息内容的新颖程度是影响用户所发信息流程度影响因素之一,而用户发布信息的流行时间和范围也是度量用户影响力的依据. Song 等人^[68]在分析博客中用户的影响力时,综合考虑了用户发布信息内容的新颖度和拓扑结构,对 PageRank 算法进行了拓展,提出了 InfluenceRank 算法用来发现博客社会网络中影响力较大的用户. Agarwal 等人^[69]综合利用了博客数据中博文的内容长度、新颖程度、引用和回复数量来度量用户的影响力. Bakshy 等人^[70]使用 Twitter 上信息扩散的树形结构来度量用户的影响力,并对用户的影响力进行了预测.研究发现,很多流行范围很广的话题是由大影响力用户发起的.然而, Bakshy 等人却并未能准确度量用户的影响力. Peng 等人^[71]在研究微博信息的最终流行度时,发现信息传播的早期转发者的拓扑结构和最终流行程度有很大相关性.若用户所发信息早期的传播深度较广,则有利于信息在社会网络中

的广泛传播。

社会网络中,用户都有各自感兴趣的话题.研究者发现,用户在不同话题上的影响力也不同.度量用户在不同话题上的影响力,则能从不同角度对用户的影响力进行更加细致的刻画.研究者利用统计机器学习方法度量用户在不同话题上的影响力.假设用户对另一个用户的影响力为一个隐变量,机器学习方法通过 Gibbs 采样或者 EM 算法等进行迭代学习出该隐变量.Blei 等人^[72]提出的 LDA(latent dirichlet allocation)模型可以模拟文档的产生过程,从而可以利用 LDA 模型得到用户所发信息内容在不同话题上的分布.Dietz 等人^[73]综合利用用户所发信息的文本内容和用户拓扑结构,假设用户产生文本内容的原因有两种,用户自身创作产生或受他人影响产生,通过 Gibbs 采样迭代计算文本内容的来源及其对应的概率,从而得到用户在不同话题上的影响力.Tang 等人^[28]利用 LDA 得到大规模异构网络中用户在不同话题上的话题分布,结合用户间的话题相似度或者强度提出了 TAP(topical affinity propagation)模型用来度量用户间在不同话题上的影响力强度,并通过 Gibbs 采样进行求解,然后在此基础上对 PageRank 加以拓展,发现在不同话题上的代表性用户,即不同话题上影响力大的用户.

Li 等人^[74]用一个统一的概率模型 TOIM(topic-level opinion influence model)将话题因素、用户观点和节点影响力结合到同一模型中,该模型有两个学习阶段:第 1 阶段结合话题因素和用户影响力产生用户间在不同话题上的影响力关系,第 2 阶段通过 TOIM 模型利用用户历史信息 and 交互关系构造历史观点和邻居观点的影响,从而进一步预测用户在某一话题上的观点.Guo 等人^[75]利用社会网络改进了 LDM 模型,提出了 SRTM (social-relational topic model)话题模型,并将其应用到度量用户的话题影响力上.该模型在学习用户话题分布时,不仅考虑用户自身所发信息内容,而且将其邻居所发话题的内容同样考虑进去.模型用 Gibbs 采样算法进行参数求解,并用逻辑斯蒂回归判定用户的关注行为由被关注用户的话题影响力引起还是由被关注用户的全局影响力引起,从而得到用户的话题影响力和全局影响力.Weng 等人^[76]根据综合 Twitter 上的用户关注网络拓扑结构和用户之间的兴趣相似度,在 PageRank 的基础上提出了 TwitterRank 以度量用户在不同话题上的影响力.Cui 等人^[77]利用社会网络上用户之间的交互信息提出了 HF-NMF(hybrid factor non-negative matrix factorization)模型, HF-NMF 模型可以度量和预测用户间在不同信息条目上的影响力.Tsur^[78]等人将内容和时间因素与网络拓扑结构相结合,使用线性回归预测 Twitter 上标签在一定时间内流行的范围.

也有学者研究了在网络拓扑结构未知的情况下,根据内容度量用户的影响力.Gerrish 等人^[79]在分析没有连接关系的论文数据集中的论文影响力时提出了动态话题模型,并将其应用于度量论文间的影响力.Shaparenko 等人^[80]利用机器学习的概率模型来度量学术影响力.Romero 等人^[81]研究了 Twitter 上标签的流行时间和影响力的范围.

基于信息内容的影响力度量方法和模型能够更加细致地描述用户在影响他人时所表现出来的具体形式,这种影响可能导致他人在信息内容上与用户的相似性和一致性,也有可能是导致他人在某个话题上情感态度发生转变的原因.然而,这类方法忽略了用户间在长期交流过程中形成的相对稳定的影响力.

2.2.2 基于用户行为的度量

对于人类行为的定量分析以及人类行为的时空规律研究已经吸引了复杂性科学、统计物理等相关领域的研究者们,并产生了极大的兴趣,人类行为时空特征研究对社会网络中用户行为的建模、分析以及传播动力学研究起着重要的理论指导意义,从而进一步为通过社会网络中用户的行为来度量用户影响力提供了理论基础.文献[82]对人类行为时空特性分析研究的进展进行了详细的综述.社会网络中的用户发布内容信息,然后通过交互行为将内容传播,通过分析这些行为,不仅可以度量用户之间的影响力强度,还能预测用户行为在社会网络上的传播速度和范围.基于用户行为度量的通常是节点间的影响力,属于局部影响力.局部影响力可以通过加权平均以及随机游走等方法转化成全局影响力.

社会网络中的用户行为以日志的形式被记录,通过这些记录可以度量用户间的影响力强度以及用户信息和行为在社会网络中的传播范围和速度.Yang 等人^[32]在分析 Twitter 上用户影响力随时间的变化规律时,采用了指数函数或幂律函数度量用户在采用某条信息后,在一定时间段内其他用户提及该信息的次数.他们提出了 LIM (linear influence model)模型用来度量用户影响力并预测用户信息在短时间内的传播范围,该模型中未使

用用户间的网络拓扑结构.Tan 等人^[83]综合考虑了拓扑结构和用户历史行为等数据以度量邻居对用户的影响力,从而预测用户的行为.Trusov 等人^[84]分析了社会网络中用户的活跃度与用户影响力之间的关系后发现:若用户在用户活跃期间其好友数量快速增加,则用户的影响力较大.Goyal 等人^[31]从 Flickr 网站用户之间分享图片等行为中用机器学习的方法度量用户间的影响力强度,将行为传播的频率作为度量用户影响力指标,将行为的传播范围作为度量行为影响力的指标.Xiang 等人^[85]利用用户间的交互行为和话题相似性度量用户间的影响力.Zaman 等人^[86]结合博客中的用户间的转发行为和博文内容,利用协同过滤算法度量用户转发博文的影响力.Yang 等人^[87]研究 Twitter 中的用户转发行为,发现用户的转发行为服从幂律分布且受用户的影响力、内容和时间因素的影响.Cha 等人^[27]分别利用关注行为、转发行为和被提及行为度量用户的关注影响力、转发影响和提及影响力,并分析了用户影响力随时间的演化规律,发现被大量用户关注的用户并不一定能引起关注者的转发行为和提及行为.毛佳昕等人^[88]考虑了微博中用户的阅读习惯和转发偏好等用户行为特征以度量用户的影响力.

上述模型从用户的行为出发或度量用户的影响力或度量用户间的影响强度,在用户行为和用户影响力之间建立起了联系并取得了较好的效果.然而,此类模型将一对用户之间的行为与其他用户的行为视为不相关,未考虑用户与其多个有联系的用户之间的复杂关系,对于社会网络的拓扑结构没有利用或利用不充分.因此,综合利用用户的行为特征与社会网络的拓扑结构来提高度量用户影响力模型的准确性和效率,是当前研究者的热门研究趋势之一.

Zhang 等人^[89]研究了微博用户的转发行为,发现用户转发消息的概率受其邻居中已转发此消息邻居之间形成的连通图的个数的影响,并提出了局部节点影响力(social influence locality),局部影响力度量用户朋友圈的结构对转发行为的影响.图 5 给出一个实例,红色点表示已经转发的邻居节点,白色代表未转发,图 5(a)所示的红色节点形成两个连通子图,图 5(b)、图 5(c)分别形成 4 个和 6 个连通子图.根据文献结果可知, V 节点在图 5(a)所示状态下转发消息的概率要远超图 5(b)、图 5(c)所示的状态.

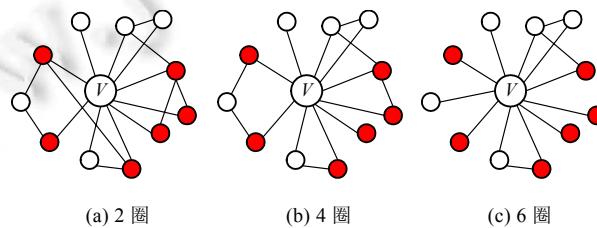


Fig.5 Influence of neighbor node topology on forwarding behavior of nodes^[89]

图 5 邻居节点拓扑结构对节点转发行为的影响^[89]

Romero 等人^[90]在度量用户的影响力时,不仅考虑了用户自身受欢迎的程度和活跃度,还考虑了其好友的被动性(passivity),被动性越强,则越不易受他人影响.研究者^[91-93]在研究社会网络中用户的不同行为时,分析了用户的行为特征和用户的节点影响力之间的联系.Bao 等人^[94]在分析影响微博中用户所发布信息流行度时发现:信息的最终流行度与早期转发此信息的用户的拓扑结构特征有关,将早期转发者中存在关注关系的连接数量与所有可能的连接数量之比定义为连接密度(link density),将所有转发者中到信息发起者的最大距离定义为扩散深度(diffusion depth).其研究表明:早期转发者的连接密度越低,扩散深度越高,信息的最终流行程度较高.用户的影响力与用户发布的信息的流行性息息相关,影响力更大的用户引起流行度较大的话题的可能性更大.上述模型充分分析了用户和其有联系的群体之间的影响力,结合用户的行为对用户的未来行为以及信息传播等进行预测获得了较好的效果.然而,该类模型并没有直接度量用户的影响力.

研究者从不同角度对 PageRank 算法进行了拓展,Haveliwala 等人^[95]综合考虑了用户话题的倾向性和发布信息敏感性与新颖性,在 PageRank 的基础上,提出了 Personalized PageRank 算法以度量用户影响力.Tang 等人^[28]提出了用户在话题上的传播模型 Topical Affinity Propagation(TAP),并结合 PageRank 提出了 PageRank with topic-based influence(TPRI)模型用来度量用户在不同话题上的全局影响力.Xiang 等人^[96]研究了科学家合作网络中用户的权威性和影响力之间的关系,提出了一种线性影响力模型,并通过引入先验知识改进了

PageRank 的用户权威性度量.Weng 等人^[76]提出了 Twiterrank,认为用户拥有多个兴趣话题且在每个话题上的影响力是不同的,综合考虑了用户间的话题相似度和拓扑结构,对 PageRank 进行了拓展,以度量用户在单个话题上的影响力.

丁兆云等人^[97]根据 Twitter 中用户的阅读行为、转发行为、复制行为和回复行为定义了 4 种关系网络,提出了基于多关系网络的话题层次影响力模型(MultiRank),该模型考虑了用户影响力在这 4 种网络上和网络之间的转移概率,在得到用户在单个话题上的影响力的基础上,利用杰卡德相似系数计算用户在多话题上的综合影响力.王晨旭等人^[98]从用户初始影响力、微博消息传播过程中影响力的持久性以及衰减指数 3 个方面综合度量微博中意见领袖的影响力,在研究 Web 网络中的网站的影响力时,与传统研究方法利用网页间的信息流与连接关系有所不同.Li 等人^[99]通过分析用户群体注意力的变化情况,提出一个注意力流(attention flow model)模型,用不同网站之间的动态流动结构与模式来度量网站的影响力并进行网站排名,这为社会网络节点影响力度量提供了一条新颖的思路.

2.2.3 基于时间因素的度量

交互时间、行为时间等时间相关因素在社会网络中的信息传播中扮演着十分重要的角色,将时间因素考虑到影响力度量模型中,能有效提高模型的精度.人类行为中的一些特性,如阵发性与记忆性,在社会网络中同样适用.Huang 等人^[100]分析了微博中用户间转发信息的概率,发现一个用户转发另一个用户的信息距离上次转发该用户信息的时间间隔呈现出短期内阵发、高频和长时间沉默的特点,并服从参数不同的幂律分布,可用贝叶斯学习方法进行建模和预测用户间的信息转发概率.Wang 等人^[101]在度量用户间影响力强度(PFG)模型的基础上,结合时间信息提出了动态因子图 Dynamic Factor Graph(DFG)模型,DFG 模型用来度量用户间的影响力强度随时间的变化情况.赵佳等人^[102]根据节点在时域上的自相似性和交互次数,利用贝叶斯模型度量了用户间的直接影响,并利用半环代数度量用户间的间接影响力.通过在原社会网络上加入“虚假粉丝”的条件,验证了模型所得到的用户影响力排序结果的稳定性.Pan 等人^[103]利用用户的状态信息的变化在用户行为等条件未知的情况下对用户影响力进行建模,分析了用户影响力随时间的变化情况.Huang 等人^[104]分析了用户通过口口相推荐购买商品的节点影响力和购买商品后的评价影响力.Rodriguez 等人^[105]根据历史信息扩散数据,利用信息扩散时的时间延迟推测出潜在的网络拓扑结构和信息在用户间的传播速率,提出 NetRate 算法并用随机凸优化对这一问题进行求解.

2.3 影响力度量方法小结

用户的影响力看不见摸不着,然而用户通过自身影响力使社会网络中其他用户产生的变化却是可以观测到的,这种变化或表现为内容上的相似性和行为上的一致性,或表现为观念态度的转变.基于用户信息和行为特征的影响力度量方法相比于基于拓扑结构的影响力度量方法能够更好地刻画用户与用户之间影响力的形成和发展状况.社会网络中,用户的行为存在多种形式,现有模型大多只考虑其中的一种行为或者将各种行为割裂开来,通过简单加权来度量用户影响力,而没有考虑这些行为以及用户信息内容之间的相互联系和相互区别.若一个用户对另一个用户的影响力很大,则可能表现为交互行为次数很多,也可能表现为两者的行为在时间序列上相似性很高,还有可能表现为两者的行为时间间隔很短.然而,这些行为特征之间的内在联系却很少,在现有用户影响力模型中均可得以体现.进一步理清社会网络中用户各种行为间的复杂关系与用户影响力的联系仍有待研究.我们将 3 类度量影响力方法的优缺点进行对比,见表 3.

Table 3 Advantages and disadvantages of influence measurement methods

表 3 影响力度量方法的侧重点以及缺陷

| 影响力度量方法 | 优点 | 缺点 |
|-------------|---|-----------------------|
| 基于网络拓扑结构的度量 | 数据易获取,多种社会网络适用性较强,宏观层面上具有指导意义,且在大规模网络中有一定优势 | 忽略了用户和用户间关系的多样性 |
| 基于内容特征的度量 | 对影响力的度量更加具体化 | 对数据的预处理要求较高 |
| 基于行为特征的度量 | 预测精度高,考虑了用户行为引起的影响力的传播 | 只适用于特定社会网络,未考虑全局的网络结构 |

3 影响力传播

无论是定性还是定量的影响力度量方法,都是将影响力赋予个体节点.事实上,影响力通常是通过社会网络上节点行为的传播表现出来的,所以探索和分析影响力的传播成为影响力研究的热点方向之一.

信息和影响力在社会网络中的动态传播过程十分复杂,但却可以通过影响力传播模型对影响力在社会网络中的传播过程进行刻画.社会网络中的信息传播模型和影响力传播模型所描述的过程十分类似.独立级联模型(independent cascade model)和线性阈值模型(linear threshold model)是两种最经典的影响力传播模型.

3.1 影响力的传播

独立级联模型(independent cascade model)^[33,34]是一种概率模型,每条边 (v,u) 上的影响概率为一指定常量.IC模型可以描述为:在社会网络 $G=(V,E)$ 中,部分节点在初始时刻处于活跃状态,如果节点 v 在某时刻 t 由不活跃状态被激活变为活跃状态,则节点 v 在 t 时刻有单次机会尝试去激活每一个处于不活跃状态的邻居节点.若节点 u 处于不活跃状态且其有多个邻居节点处于活跃状态,则这些活跃的邻居节点激活 u 节点的顺序是任意的.例如,节点 v 的邻居节点 u 在 t 时刻处于不活跃状态,则节点 v 以概率去激活节点 u ,如果成功激活,则节点 u 从 $t+1$ 时刻起变为活跃状态,无论成功与否,节点 v 不会再试图去激活节点 u .传播过程从初始状态开始迭代直到 G 中没有产生新的活跃节点为止.

线性阈值模型(linear threshold model)^[33,106]体现了影响力的累积过程,该模型可以描述为:在社交网络 $G=(V,E)$ 中,对任意节点 v 都有一个对应阈值表示只有当节点 v 的邻居节点对节点 v 的影响力之和超过阈值时,节点 v 才能被激活.节点 v 对其邻居节点 u 的影响力权重为 w ,且节点 v 对所有邻居节点的影响力权重之和不超过1,表示节点 v 的邻居节点,即:给定初始时刻处于活跃状态的用户集合为 A ,传播过程和独立级联模型类似,不活跃节点 v 状态受其所有活跃状态节点的影响.如果 t 时刻节点 v 处于不活跃状态,且其所有处于活跃状态的邻居节点对节点 v 的影响力之和大于节点 v 的阈值,则节点 v 从 $t+1$ 时刻起变为活跃状态.

Kempe 等人^[33]在独立级联(IC)模型和线性阈值(LT)模型的基础上又进行了改进,提出了通用级联模型,通用阈值模型等使线性阈值中的线性关系变得更加灵活,优化了独立级联中的影响概率,从而得到更大的覆盖范围.更为详细的介绍参考文献[107],上述影响力传播模型中节点被激活后状态就不再改变,显然与真实社会网络中的情况不符.

3.2 影响力最大化

影响力最大化就是找到社会网络中少量的种子节点集合,使影响力在短时间内通过种子节点迅速传遍整个社会网络.Domingos 和 Richardson 等人^[108,109]首次提出在社会网络中引入影响力最大化算法.Kempe 等人^[33]则首先将影响力最大化形式化定义为在特定影响力传播模型中挖掘影响力大的 K 个节点的离散优化问题,并证明了影响力最大化问题是一个 NP-hard 问题.在此基础上,给出了贪心算法,达到了近似最优解的 63%.然而,贪心算法每次迭代获取近似的影响范围都需要调用蒙特卡罗算法^[110],耗时过久,不适用于大规模网络.

影响力最大化的提出引起了大量研究者的关注.Leskovec 等人^[111]利用影响传播模型中的次模函数的特性提出的 CELF 算法比贪心算法的效率提高了几百倍.Chen 等人^[112]进一步基于节点的度提出了度减小(degree discount)优化算法,算法的实验结果与贪心算法相近,然而算法效率却得到了很大的提高.Jung 等人^[113]提出了 IRIE 模型,首先通过全局影响力排序(influence ranking)算法对所有节点影响力进行排序,选择影响力最大的节点作为种子节点,然后用影响力估计(influence estimation)算法对剩余节点影响力重新排序,从而解决了影响力重叠问题.该模型无论是在效率还是在速度上,相比于贪心算法都得到了很大的提高.Goyal 等人^[114]提出了 CD(credit distribution)模型来解决影响力最大化问题,通过用户历史数据直接估计出用户间的影响概率.相比于其他算法,避免了蒙特卡罗算法在学习用户间影响概率时的大量时间消耗.曹玖新等人^[115]基于节点 K 核分解和度提出了核覆盖模型(CCA),在选取新的种子节点时,优先选择与已选节点距离较远且核数较大的节点,若核数相同则优先选取度数较大的节点.Tong 等人^[116]认为:在研究影响力最大化问题时,仅使用静态的网络拓扑结构不足以追踪在真实社会网络中的动态变化,他们提出了动态级联模型(dynamic independent cascade)和自适应贪

心算法作为种子节点选取策略.Zhuang 等人^[117]通过周期性地探测一些节点以更新他们的连接,从而获取整个网络拓扑结构的动态变化,最终得到动态网络中影响力最大化问题的近似最优解.Liu 等人^[118]在研究时间限制下的影响力最大化问题时,用概率分布来表示节点间在不同时间段上的影响力,并提出了影响力传递路径(influence spreading path)以解决影响力最大化问题.Yao 等人^[119]研究了在一部分节点已经被感染的状态下,锁定 K 个未被感染的节点,使最终所有感染节点数达到最小,对影响力最大化问题作了变形.

自影响力最大化问题提出以来一直是研究热点,无论是模型的精度还是效率都得到了不断的提高.然而在传播模型中,用户间的影响力传播概率通常在一定范围内随机取值,而没有考虑真实社会网络中用户之间传播率的差异性,从而导致模型在真实网络中的适用性不高.利用用户的历史数据和机器学习等方法来度量用户间的影响概率,这也是当前影响力最大化研究仍需要解决的问题之一.

4 影响力的评价与应用

由于影响力没有一致的定义和度量方法,造成迄今为止节点影响力研究并没有公认的统一的评价指标.一方面,节点影响力研究模型和指标虽然数量众多,但都是从各自研究角度去分析节点影响力对其他因素所产生的效果,而没有给出节点影响力的形式化定义;另一方面,节点影响力分析带有很强的应用色彩,通过宏观层面的用户影响力大小排序可以找出领域内的专家^[28]、社会网络中的意见领袖^[67];通过微观层面的个体或群体对个体的影响力分析则可以应用于行为与观点的预测^[77]推荐系统^[104]、链接预测^[93]等.节点影响力在病毒式营销^[108]、舆论引导^[2]等应用中则更加侧重于用户的传播影响力.多领域、多应用也是导致用户的节点影响力评价方法和指标缺失的原因之一.虽然没有统一的模型或指标去衡量用户影响力分析模型的好坏,但常见的方法却有迹可循.常见的用户影响力评价方法有基于信息检索的评价方法、基于传播动力学的评价方法、基于鲁棒性和脆弱性的方法以及基于影响力的传播模型.

基于信息检索的评价指标主要包括 $P@N$ ^[28,29,47,51,120], Precision^[28,74,104], Recall^[89,120], $F1$ ^[89,120]等.这些指标都是信息检索中常见的指标,通常用于评价影响力排序和通过影响力进行预测的效果. $P@N$ 表示对根据用户影响力大小进行排序前, N 个用户中被人工判定排序正确的个数. Precision 和 Recall 分布表示实验结果的准确率和召回率, $F1$ 则是实验结果准确率和召回率的综合体现.由于用户影响力研究缺少标准数据集,因此,通过人工对用户的影响力大小进行判定成为了一种常见的方法.通过人工判定的结果去评价用户影响力模型的好坏,在一些用户规模较小的社会网络,诸如科学家合作网络、校园博客、论坛等社会网络中有一定的可行性.然而,人工判定存在着主观性强且不适用于大规模社会网络中的缺点.

基于传播动力学的评价方法主要有传染病模型 SI^[58,121], SIS^[22]和 SIR^[57,122]模型.

- 在 SI 模型中,节点有两种状态:S(susceptible)易感态表示该节点可能被其邻居节点中处于感染态的节点感染,I(Infected)感染态表示节点被感染.节点一旦被感染,就会永远处于感染态.初始时刻,单个节点作为感染源以 p 的概率去感染其节点,可以观测不同时间被感染节点的数量,从而得到该节点的传播速度;
- 在 SIS 模型中,节点可以被反复感染.在 SIR 模型中,节点多了一种状态免疫态 R(recovered),表示被感染节点在一段时间后会变为免疫态,免疫态的节点不会再次被感染,也不会去感染其他节点.

用单个节点作为感染源,最终被感染过的节点数量表示该节点传播覆盖的范围,多次实验取平均值作为节点的传播影响力.对于模型中一些参数以及更为详细的传播动力学模型研究进展的介绍可以参考文献^[123].基于传播动力学的评价方法是对现实世界中人或事物的高度抽象和模拟,具有一定的指导意义.但是此类评价模型受动力学影响很大,由于真实社会网络中个体特征差异很大,导致其在真实社会网络中的适用性较差.

基于影响力传播模型的评价方法通常应用于影响力最大化研究中, ICM^[33,34]和 LIM^[33,106]是公认的影响力传播模型.通过在传播模型上验证不同规模的种子节点集合(影响力最大化模型得到的影响力较大的节点集合),最终能够激活的节点数量可以验证影响力最大化模型的优劣.

基于鲁棒性和脆弱性的评价方法通过原社会网络中增加或删除一定比例的节点后^[20,102],观察社会网络变

动前后节点影响力前后的差值:若差值波动很小,则认为该模型抗干扰能力较好.无论从什么角度去评价影响力模型,都需要考虑模型在大规模社会网络中的时间复杂度,部分模型虽然效果和预测精度都很好,但在大规模网络中需要耗费大量时间.

以上评价方法各有各的侧重点与优势,也有其不足之处.各类评价方法对比见表 4.

Table 4 Evaluation methods for node influence

表 4 节点影响力评价方法对比

| 评价类型 | 评价方法 | 适用影响力形式 | 优势与不足 |
|-----------|---------------------------------------|-----------|----------------------------------|
| 基于信息检索 | 人工标识结果的 P@N, Precision, Recall 和 F1 等 | 分类相对排序概率 | 与真实社会网络情况一致,但不适用于大规模社会网络 |
| 基于传播动力学 | SI, SIR, SIS 等传染病模型 | 网络拓扑上的重要性 | 相对客观且在宏观上具有指导意义,但与真实网络存在差异 |
| 基于影响力传播模型 | ICM, LIM 等模型 | 传播时激活节点数量 | 影响力最大化研究的标准模型.与传染病模型类似,高度抽象,有好有坏 |
| 基于鲁棒性和脆弱性 | 增加或删除一定比例节点 | 多形式 | 考虑了模型度量影响力结果的抗干扰能力.没有验证影响力的大小 |

5 面临的挑战和进一步研究方向

影响力研究已经取得了众多成果,但还缺少基准性的研究.另外,伴随着社会网络、移动网络等技术的快速发展,新的特性对社会网络影响力研究提出了新的挑战.

通过分析研究文献,我们认为,未来值得深入研究的方向主要有:

- (1) 节点影响力评价标准缺失一直是影响力分析度量模型的一大难点.通过影响力度量模型给出了用户的影响力大小,然而用户的影响力并不能直接观测得到,现有解决方法只能通过人工标识重要节点,或通过 SIR 等传染病模型感染的规模和效率来度量用户影响力的效果.这些方法都是从侧面评价影响力,并没有直接给出影响力的统一定义.因此,深入研究影响力的形式化定义,从而进一步给出评价标准;
- (2) 社会网络节点影响力具有动态性,现有研究大多在静态网络拓扑结构上考察用户的影响力或者考察用户影响力在静态拓扑结构上随时间的变化情况.然而,用户的社会网络拓扑结构在变,用户的影响力同时也在变,究竟是用户的影响力变化引起了网络拓扑结构的变化,还是网络拓扑结构的变化导致了用户影响力随之改变,两者之间的关系对于研究社会网络中圈子的形成和演化具有十分重要的意义;
- (3) 个体之间影响力产生作用效果的表现形式很多,如两者的交互行为次数越多,影响力可能越大.影响力越大的节点对的直接交互行为的时间越稳定,某用户做出某个动作,其好友有的很短时间内就对此做出反应,而有的则很长时间才会做出反应.现有研究大多从用户自身行为在时序上的自相似性角度来考虑影响力,即,认为最近处于活跃状态的用户未来短时间内会对他人产生交互的可能性更高,而忽略了用户交互的延迟时间所起的作用.将更多因素考虑到节点影响力度量模型中以提高模型的效果,使模型更加接近真实社会中的影响力,这也是值得关注的研究方向;
- (4) 目前,节点影响力的研究主要集中在个体在全局中的影响力、个体与个体之间的影响力、个体对群体以及群体对个体的影响力上.将联系非常紧密的群体视为一个整体,度量群体与群体之间的影响力以及影响机制,在网络规模达到千万甚至上亿级别时就拥有一定的优势.关于群体影响力的研究非常少见,值得进行深入研究;
- (5) 现有的影响力度量模型大多只考虑社会网络中的一种行为,或将不同行为割裂开,通过简单加权来度量用户影响力,没有考虑不同类型的行为关联以及行为和内容之间的关联.综合利用社会网络用户的多种行为和内容,对影响力统一分析建模,是影响力研究所面临的又一挑战.

6 总 结

社会学家对于个体的影响力产生的原因和个体影响力差异进行了探索.图论与概率统计等理论为度量社会网络中节点的影响力提供了坚实的理论基础.近年来,论坛、博客和微博等大规模社会的出现,为节点影响力研究提供了新的契机,研究者所能获得的社会网络用户数据不仅其规模有了空前提高,而且用户行为和用户内容也变得更加丰富.本文详细介绍了社会网络节点影响力研究的相关成果,首先介绍了节点影响力的定义、理论以及具体表现形式;然后,重点从多角度介绍了节点影响力的度量方法以及影响力在社会网络中的传播;最后介绍了影响力的评价指标和应用.节点影响力的研究和建模不仅能够进一步帮助人们理解社会网络中的个体和群体的行为的演化,还能为制定公共决策和舆情分析提供理论依据,同时也有助于社会、文化、经济等领域的安全和发展,社会网络节点影响力研究具有十分重要的理论价值和应用价值.

References:

- [1] Triplett N. The dynamogenic factors in pacemaking and competition. *The American Journal of Psychology*, 1898,9(4):507-533. [doi: 10.2307/1412188]
- [2] Katz E, Lazarsfeld PF. *Personal Influence, the Part Played by People in the Flow of Mass Communications*. New York: Free Press, 1955. 1-12.
- [3] Milgram S. The small-world problem. *Psychology Today*, 1967,2(1):61-67.
- [4] Backstrom L, Boldi P, Rosa M, Ugander J, Vigna S. Four degrees of separation. In: Contractor N, Uzzi B, eds. *Proc. of the 4th Annual ACM Web Science Conf. (WebSci 2012)*. New York: ACM Press, 2012. 33-42. [doi: 10.1145/2380718.2380723]
- [5] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks. *Nature*, 1998,393(6684):440-442. [doi: 10.1038/30918]
- [6] Barabási AL, Albert R. Emergence of scaling in random networks. *Science*, 1999,286(5439):509-512. [doi: 10.1126/science.286.5439.509]
- [7] Sun J, Tang J. *A Survey of Models and Algorithms for Social Influence Analysis*. New York: Springer-Verlag, 2011. 177-214. [doi: 10.1007/978-1-4419-8462-3_7]
- [8] AlFalahi K, Atif Y, Abraham A. Models of influence in online social networks. *Int'l Journal of Intelligent Systems*, 2014,29(2): 161-183. [doi: 10.1002/int.21631]
- [9] Liu JG, Ren ZM, Guo Q, Wang BH. Node importance ranking of complex networks. *Acta Physica Sinica*, 2013,62(17):178901 (in Chinese with English abstract). [doi: 10.7498/aps.62.178901]
- [10] Wu XD, Li Y, Li L. Influence analysis of online social network. *Chinese Journal of Computers*, 2014,37(4):735-752 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2014.00735]
- [11] Rashotte L. *Social Influence: The Blackwell Encyclopedia of Psychology*. Malden: Blackwell Publishing, 2007. 4426-4427.
- [12] Anagnostopoulos A, Brova G, Terzi E. Peer and authority pressure in information-propagation models. In: *Proc. of the Machine Learning and Knowledge Discovery in Databases*. Heidelberg: Springer-Verlag, 2011. 76-91. [doi: 10.1007/978-3-642-23780-5_15]
- [13] Yang Y, Tang J, Leung C, Sun Y, Chen Q, Li J, Yang Q. RAIN: Social role-aware information diffusion. In: *Proc. of the 29th AAAI Conf. on Artificial Intelligence (AAAI 2014)*. Menlo Park: AAAI Press, 2014. 367-373. <http://www.aaai.org/Library/AAAI/aaai15contents.php>
- [14] Granovetter MS. The strength of weak ties. *American Journal of Sociology*, 1973,78(6):1360-1380.
- [15] Krackhardt D. *The Strength of Strong Ties: The Importance of Philos in Organizations*. Boston: Harvard Business School Press, 1992. 216-239.
- [16] Bonacich P. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 1972,2(1):113-120. [doi: 10.1080/0022250X.1972.9989806]
- [17] Freeman LC. A set of measures of centrality based on betweenness. *Sociometry*, 1977,40(1):35-41. [doi: 10.2307/3033543]
- [18] Newman MEJ. A measure of betweenness centrality based on random walks. *Social Networks*, 2005,27(1):39-54. [doi: 10.1016/j.socnet.2004.11.009]
- [19] Sabidussi G. The centrality index of a graph. *Psychometrika*, 1966,31(4):581-603. [doi: 10.1007/BF02289527]

- [41] Min L, Liu Z, Tang XY, Chen M, Liu SY. Evaluating influential spreaders in complex networks by extension of degree. *Acta Physica Sinica*, 2015,64(8):387–397 (in Chinese with English abstract).
- [42] Fowler JH, Christakis NA. Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham heart study. *BMJ*, 2008,337:a2338. [doi: 10.1136/bmj.a2338]
- [43] Centola D. The spread of behavior in an online social network experiment. *Science*, 2010,329(5996):1194–1197. [doi: 10.1126/science.1185231]
- [44] Ugander J, Backstrom L, Marlow C, Kleinberg J. Structural diversity in social contagion. *Proc. of the National Academy of Sciences*, 2012,109(16):5962–5966. [doi: 10.1073/pnas.1116502109]
- [45] Cui P, Tang M, Wu ZX. Message spreading in networks with stickiness and persistence: Large clustering does not always facilitate large-scale diffusion. *Scientific Reports*, 2014,4:6303. [doi: 10.1038/srep06303]
- [46] Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B. Measurement and analysis of online social networks. In: *Proc. of the 7th ACM SIGCOMM Conf. on Internet Measurement (IMC 2007)*. New York: ACM Press, 2007. 29–42. [doi: 10.1145/1298306.1298311]
- [47] Zhang J, Tang J, Zhuang H, Leung CW, Li J. Role-Aware conformity influence modeling and analysis in social networks. In: *Proc. of the 28th AAAI Conf. on Artificial Intelligence (AAAI 2014)*. Menlo Park: AAAI Press, 2014. 958–965. <http://www.aaai.org/Library/AAAI/aaai14contents.php>
- [48] Stephenson K, Zelen M. Rethinking centrality: Methods and examples. *Social Networks*, 1989,11(1):1–37. [doi: 10.1016/0378-8733(89)90016-6]
- [49] Borgatti SP. Centrality and network flow. *Social Networks*, 2005,27(1):55–71. [doi: 10.1016/j.socnet.2004.11.008]
- [50] Katz L. A new status index derived from sociometric analysis. *Psychometrika*, 1953,18(1):39–43. [doi: 10.1007/BF02289026]
- [51] Liu JG, Ren ZM, Guo Q. Ranking the spreading influence in complex networks. *Physica A: Statistical Mechanics and its Applications*, 2013,392(18):4154–4159. [doi: 10.1016/j.physa.2013.04.037]
- [52] Zeng A, Zhang CJ. Ranking spreaders by decomposing complex networks. *Physics Letters A*, 2013,377(14):1031–1035. [doi: 10.1016/j.physleta.2013.02.039]
- [53] Borge-Holthoefer J, Moreno Y. Absence of influential spreaders in rumor dynamics. *Physical Review E*, 2012,85(2):026116. [doi: 10.1103/PhysRevE.85.026116]
- [54] Borge-Holthoefer J, Meloni S, Gonçalves B, Moreno Y. Emergence of influential spreaders in modified rumor models. *Journal of Statistical Physics*, 2013,151(1-2):383–393. [doi: 10.1007/s10955-012-0595-6]
- [55] Liu Y, Tang M, Zhou T, Do Y. Core-Like groups result in invalidation of identifying super-spreader by k -shell decomposition. *Scientific Reports*, 2015,5:9602. [doi: 10.1038/srep09602]
- [56] Lü L, Zhou T, Zhang QM, Stanley E. The H-index of a network node and its relation to degree and coreness. *Nature Communications*, 2016,7:10168. [doi: 10.1038/ncomms10168]
- [57] Li Q, Zhou T, Lü L, Chen D. Identifying influential spreaders by weighted LeaderRank. *Physica A: Statistical Mechanics and its Applications*, 2014,404:47–55. [doi: 10.1016/j.physa.2014.02.041]
- [58] Shi X, Zhu J, Cai R, Zhang L. User grouping behavior in online forums. In: *Proc. of the 15th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2009)*. New York: ACM Press, 2009. 777–786. [doi: 10.1145/1557019.1557105]
- [59] Zhao ZY, Yu H, Zhu ZL, Wang XF. Identifying influential spreaders based on network community structure. *Chinese Journal of Computers*, 2014,37(4):753–766 (in Chinese with English abstract).
- [60] Hu QC, Yin YS, Ma PF, Gao Y, Zhang Y, Xin XC. A new approach to identify influential spreaders in complex networks. *Acta Physica Sinica*, 2013,62(14):9–19 (in Chinese with English abstract). [doi: 10.7498/aps.62.140101]
- [61] Burt RS. *Structural Holes: The Social Structure of Competition*. Shanghai: Shanghai Century Publishing Co. Ltd., 2008. 9–45.
- [62] Burt RS, Kilduff M, Tasselli S. Social network analysis: Foundations and frontiers on advantage. *Annual Review of Psychology*, 2013,64:527–547. [doi: 10.1146/annurev-psych-113011-143828]
- [63] Su XP, Song YR. Leveraging neighborhood “structural holes” to identifying key spreaders in social networks. *Acta Physica Sinica*, 2015,64(2):020101 (in Chinese with English abstract). [doi: 10.7498/aps.64.020101]

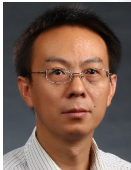
- [64] Han ZM, Wu Y, Tan XS, Duan DG, Yang WJ. Ranking key nodes in complex networks by considering structural holes. *Acta Physica Sinica*, 2015,64(5):429–437 (in Chinese with English abstract). [doi: 10.7498/aps.64.058902]
- [65] Lou T, Tang J. Mining structural hole spanners through information diffusion in social networks. In: *Proc. of the 22nd Int'l Conf. on World Wide Web Steering Committee (IW3C2)*. New York: ACM Press, 2013. 825–836. [doi: 10.1145/2488388.2488461]
- [66] Matsumura N, Ohsawa Y, Ishizuka M. Influence diffusion model in text-based communication. *Trans. of the Japanese Society for Artificial Intelligence*, 2002,17:259–267. [doi: 10.1527/tjsai.17.259]
- [67] Duan SQ, Wu B, Wang B. TTRank: User influence rank based on tendency transformation. *Journal of Computer Research and Development*, 2014,51(10):2225–2238 (in Chinese with English abstract).
- [68] Song X, Chi Y, Hino K, Tseng B. Identifying opinion leaders in the blogosphere. In: *Proc. of the 16th ACM Conf. on Information and Knowledge Management (CIKM 2007)*. New York: ACM Press, 2007. 971–974. [doi: 10.1145/1321440.1321588]
- [69] Agarwal N, Liu H, Tang L, Philip S. Identifying the influential bloggers in a community. In: *Proc. of the 2008 Int'l Conf. on Web Search and Data Mining (WSDM 2008)*. New York: ACM Press, 2008. 207–218. [doi: 10.1145/1341531.1341559]
- [70] Bakshy E, Hofman JM, Mason WA, Watts DJ. Everyone's an influencer: Quantifying influence on Twitter. In: *Proc. of the 4th ACM Int'l Conf. on Web Search and Data Mining (WSDN 2011)*. New York: ACM Press, 2011. 65–74. [doi: 10.1145/1935826.1935845]
- [71] Peng HK, Zhu J, Piao D, Yang R, Zhang Y. Retweet modeling using conditional random fields. In: *Proc. of the 11th IEEE Int'l Conf. on Data Mining Workshops. IEEE (ICDMW 2011)*. Washington: IEEE Computer Society, 2011. 336–343. [doi: 10.1109/ICDMW.2011.146]
- [72] Blei DM, Ng AY, Jordan MI. Latent dirichletallocation. *Journal of Machine Learning Research*, 2003,3:993–1022.
- [73] Dietz L, Bickel S, Scheffer T. Unsupervised prediction of citation influences. In: *Proc. of the 24th Int'l Conf. on Machine Learning (ICML 2007)*. New York: ACM Press, 2007. 233–240. [doi: 10.1145/1273496.1273526]
- [74] Li D, Shuai X, Sun G, Tang J, Ding Y, Lou Z. Mining topic-level opinion influence in microblog. In: *Proc. of the 21st ACM Int'l Conf. on Information and Knowledge Management (CIKM 2012)*. New York: ACM Press, 2012. 1562–1566. [doi: 10.1145/2396761.2398473]
- [75] Guo W, Wu S, Wang L, Tan T. Social-Relational topic model for social networks. In: *Proc. of the 24th ACM Int'l Conf. on Information and Knowledge Management (CIKM 2012)*. New York: ACM Press, 2015. 1731–1734. [doi: 10.1145/2806416.2806611]
- [76] Weng J, Lim EP, Jiang J, He Q. Twiterrank: Finding topic-sensitive influential twitterers. In: *Proc. of the 3rd ACM Int'l Conf. on Web Search and Data Mining (WSDM 2010)*. New York: ACM Press, 2010. 261–270. [doi: 10.1145/1718487.1718520]
- [77] Cui P, Wang F, Liu S, Ou M, Yang S, Sun L. Who should share what?: item-level social influence prediction for users and posts ranking. In: *Proc. of the 34th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR 2011)*. New York: ACM Press, 2011. 185–194. [doi: 10.1145/2009916.2009945]
- [78] Tsur O, Rappoport A. What's in a hashtag? Content based prediction of the spread of ideas in microblogging communities. In: *Proc. of the 5th ACM Int'l Conf. on Web Search and Data Mining (WSDN 2012)*. New York: ACM Press, 2012. 643–652. [doi: 10.1145/2124295.2124320]
- [79] Gerrish S, Blei DM. A language-based approach to measuring scholarly impact. In: *Proc. of the 27th Int'l Conf. on Machine Learning (ICML 2010)*. New York: ACM Press, 2010. 375–382.
- [80] Shaparenko B, Joachims T. Information genealogy: Uncovering the flow of ideas in non-hyperlinked document databases. In: *Proc. of the 13th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2007)*. New York: ACM Press, 2007. 619–628. [doi: 10.1145/1281192.1281259]
- [81] Romero DM, Meeder B, Kleinberg J. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter. In: *Proc. of the 20th Int'l Conf. on World Wide Web (IW3C2 2011)*. New York: ACM Press, 2011. 695–704. [doi: 10.1145/1963405.1963503]
- [82] Zhou T, Han XP, Yan XY, Yang ZM, Zhao ZD, Wang BH. Statistical mechanics on temporal and spatial activities of human. *Journal of University of Electronic Science and Technology of China*, 2013,42(4):481–540 (in Chinese with English abstract). [doi: 10.3969/j.issn.1001-0548.2013.04.001]

- [83] Tan C, Tang J, Sun J, Lin Q, Wang F. Social action tracking via noise tolerant time-varying factor graphs. In: Proc. of the 16th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2010). New York: ACM Press, 2010. 1049–1058. [doi: 10.1145/1835804.1835936]
- [84] Trusov M, Bodapati AV, Bucklin RE. Determining influential users in internet social networks. *Journal of Marketing Research*, 2010,47(4):643–658. [doi: 10.1509/jmkr.47.4.643]
- [85] Xiang R, Neville J, Rogati M. Modeling relationship strength in online social networks. In: Proc. of the 19th Int'l Conf. on World Wide Web (IW3C2 2010). New York: ACM Press, 2010. 981–990. [doi: 10.1145/1772690.1772790]
- [86] Zaman TR, Herbrich R, Van Gael J, Stern D. Predicting information spreading in Twitter. *Workshop on Computational Social Science and the Wisdom of Crowds*, 2010,104(45):17599–601.
- [87] Yang Z, Guo J, Cai K, Stern D. Understanding retweeting behaviors in social networks. In: Proc. of the 19th ACM Int'l Conf. on Information and Knowledge Management (CIKM 2010). New York: ACM Press, 2010. 1633–1636. [doi: 10.1145/1871437.1871691]
- [88] Mao JX, Liu YQ, Zhang HM, Ma SP. Social influence analysis for micro-blog user based on user behavior. *Chinese Journal of Computers*, 2014,37(4):791–800 (in Chinese with English abstract).
- [89] Zhang J, Tang J, Li J, Liu Y, Xing C. Who influenced you? Predicting retweet via social influence locality. *ACM Trans. on Knowledge Discovery from Data (TKDD)*, 2015,9(3):25. [doi: 10.1145/2700398]
- [90] Romero DM, Galuba W, Asur S, Huberman B. Influence and passivity in social media. In: Proc. of the Machine Learning and Knowledge Discovery in Databases. Berlin, Heidelberg: Springer-Verlag, 2011. 18–33. [doi: 10.1007/978-3-642-23808-6_2]
- [91] Singla P, Richardson M. Yes, there is a correlation: From social networks to personal behavior on the Web. In: Proc. of the 17th Int'l Conf. on World Wide Web (IW3C2 2008). New York: ACM Press, 2008. 655–664. [doi: 10.1145/1367497.1367586]
- [92] Anagnostopoulos A, Kumar R, Mahdian M. Influence and correlation in social networks. In: Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2007). New York: ACM Press, 2008. 7–15. [doi: 10.1145/1401890.1401897]
- [93] Crandall D, Cosley D, Huttenlocher D, Kleinberg J, Suri S. Feedback effects between similarity and social influence in online communities. In: Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2008). New York: ACM Press, 2008. 160–168. [doi: 10.1145/1401890.1401914]
- [94] Bao P, Shen HW, Huang J, Cheng X. Popularity prediction in microblogging network: A case study on SinaWeibo. In: Proc. of the 22nd Int'l Conf. on World Wide Web Companion Steering Committee (WWW 2013). New York: ACM Press, 2013. 177–178. [doi: 10.1145/2487788.2487877]
- [95] Haveliwala T, Kamvar S, Jeh G. An Analytical Comparison of Approaches to Personalizing Pagerank. Technical Report, Stanford, 2003. <http://ilpubs.stanford.edu:8090/596/>
- [96] Xiang B, Liu Q, Chen E, Xiong H, Zheng Y, Yang Y. PageRank with priors: An influence propagation perspective. In: Proc. of the 23rd Int'l Joint Conf. on Artificial Intelligence (IJCAI 2013). Menlo Park: AAAI Press, 2013. 2740–2746.
- [97] Ding ZY, Zhou B, Jia Y, Zhang LM. Topical influence analysis based on the multi-relational network in microblogs. *Journal of Computer Research and Development*, 2013,50(10):2155–2175 (in Chinese with English abstract).
- [98] Wang CX, Guan XH, Qin T, Zhou YD. Modeling on opinion leader's influence in microblog message propagation and its application. *Ruan Jian Xue Bao/Journal of Software*, 2015,26(6):1473–1485 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4627.htm> [doi: 10.13328/j.cnki.jos.004627]
- [99] Li Y, Meng XF, Wang CQ. Quantifying the influence of websites based on online collective attention flow. *Journal of Computer Science and Technology*, 2015,30(6):1175–1187. [doi: 10.1007/s11390-015-1592-4]
- [100] Huang J, Li C, Wang WQ, Shen H, Cheng X. Temporal scaling in information propagation. *Scientific Reports*, 2013,4:5334–5334. [doi: 10.1038/srep05334]
- [101] Wang C, Tang J, Sun J, Han H. Dynamic social influence analysis through time-dependent factor graphs. In: Proc. of the Int'l Conf. on Advances in Social Networks Analysis and Mining (ASONAM 2011). New York: ACM Press, 2011. 239–246. [doi: 10.1109/ASONAM.2011.116]

- [102] Zhao J, Yu L, Li JR. Node influence calculation mechanism based on Bayesian and semiring algebraic model in social networks. *Acta Physica Sinica*, 2013,62(13):130201 (in Chinese with English abstract).
- [103] Pan W, Dong W, Cebrian M, Kim T. Modeling dynamical influence in human interaction: Using data to make better inferences about influence within social systems. *Signal Processing Magazine*, 2012,29(2):77–86. [doi: 10.1109/MSP.2011.942737]
- [104] Huang J, Cheng XQ, Shen HW, Zhou T, Jin X. Exploring social influence via posterior effect of word-of-mouth recommendations. In: *Proc. of the 5th ACM Int'l Conf. on Web Search and Data Mining (WSDM 2012)*. New York: ACM Press, 2012. 573–582. [doi: 10.1145/2124295.2124365]
- [105] Rodriguez MG, Leskovec J, Balduzzi D. Uncovering the structure and temporal dynamics of information propagation. *Network Science*, 2014,2(1):26–65. [doi: 10.1017/nws.2014.3]
- [106] Granovetter M. Threshold models of collective behavior. *American Journal of Sociology*, 1978,83(6):1420–1443. [doi: 10.1086/226707]
- [107] Chen W, Lakshmanan LVS, Castillo C. Information and influence propagation in social networks. *Synthesis Lectures on Data Management*, 2013,5(4):1–177. [doi: 10.2200/S00527ED1V01Y201308DTM037]
- [108] Domingos P, Richardson M. Mining the network value of customers. In: *Proc. of the 7th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2001)*. New York: ACM Press, 2001. 57–66. [doi: 10.1145/502512.502525]
- [109] Richardson M, Domingos P. Mining knowledge-sharing sites for viral marketing. In: *Proc. of the 8th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2002)*. New York: ACM Press, 2002. 61–70. [doi: 10.1145/775047.775057]
- [110] MacKay DJC. Introduction to Monte Carlo methods. In: *Proc. of the Learning in Graphical Models*. Berlin: Springer-Verlag, 1998. 175–204. [doi: 10.1007/978-94-011-5014-9_7]
- [111] Leskovec J, Krause A, Guestrin C. Cost-Effective outbreak detection in networks. In: *Proc. of the 13th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2007)*. New York: ACM Press, 2007. 420–429. [doi: 10.1145/1281192.1281239]
- [112] Chen W, Wang Y, Yang S. Efficient influence maximization in social networks. In: *Proc. of the 15th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2009)*. New York: ACM Press, 2009. 199–208. [doi: 10.1145/1557019.1557047]
- [113] Jung K, Heo W, Chen W. Irie: Scalable and robust influence maximization in social networks. In: *Proc. of the 12th Int'l Conf. on Data Mining (ICDM 2012)*. New York: ACM Press, 2012. 918–923. [doi: 10.1109/ICDM.2012.79]
- [114] Goyal A, Bonchi F, Lakshmanan LVS. A data-based approach to social influence maximization. *Proc. of the VLDB Endowment*, 2011,5(1):73–84. [doi: 10.14778/2047485.2047492]
- [115] Cao JX, Dong D, Xu S, Zheng X, Liu B, Luo JZ. A k -core based algorithm for influence maximization in social networks. *Chinese Journal of Computers*, 2015,38(2):238–248 (in Chinese with English abstract).
- [116] Tong G, Wu W, Tang S, Du D. Adaptive influence maximization in dynamic social networks. arXiv, 1506.06294. 2015.
- [117] Zhuang H, Sun Y, Tang J, Zhang J, Sun X. Influence maximization in dynamic social networks. In: *Proc. of the 13th Int'l Conf. on Data Mining (ICDM 2013)*. New York: ACM Press, 2013. 1313–1318. [doi: 10.1109/ICDM.2013.145]
- [118] Liu B, Cong G, Zeng Y, Sun X. Influence spreading path and its application to the time constrained social influence maximization problem and beyond. *IEEE Trans. on Knowledge and Data Engineering*, 2014,26(8):1904–1917. [doi: 10.1109/TKDE.2013.106]
- [119] Yao Q, Shi R, Zhou C, Shi R, Wang P, Guo L. Topic-Aware social influence minimization. In: *Proc. of the 24th Int'l Conf. on World Wide Web Companion Steering Committee (WWW 2015)*. New York: ACM Press, 2015. 139–140. [doi: 10.1145/2740908.2742767]
- [120] Tang J, Zhang C, Cai K, Zhang L, Su Z. Sampling representative users from large social networks. In: *Proc. of the 29th AAAI Conf. on Artificial Intelligence (AAAI 2014)*. Menlo Park: AAAI Press, 2014. 304–310. <http://www.aaai.org/Library/AAAI/aaai15/contents.php>
- [121] Zhou T, Liu JG, Bai WJ, Chen G, Wang B. Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity. *Physical Review E*, 2006,74(5):056109. [doi: 10.1103/PhysRevE.74.056109]
- [122] Chen DB, Xiao R, Zeng A. Predicting the evolution of spreading on complex networks. *Scientific Reports*, 2013,4:6108–6108. [doi: 10.1038/srep061081]
- [123] Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. Epidemic processes in complex networks. *Reviews of Modern Physics*, 2015,87(3):925. [doi: 10.1103/RevModPhys.87.925]

附中文参考文献:

- [9] 刘建国,任卓明,郭强,汪秉宏.复杂网络中节点重要性排序的研究进展.物理学报,2013,62(17):178901. [doi: 10.7498/aps.62.178901]
- [10] 吴信东,李毅,李磊.在线社交网络影响力分析.计算机学报,2014,37(4):735-752. [doi: 10.3724/SP.J.1016.2014.00735]
- [20] 任卓明,邵凤,刘建国,郭强,汪秉宏.基于度与集聚系数的网络节点重要性度量方法研究.物理学报,2013,62(12):128901. [doi: 10.7498/aps.62.128901]
- [30] 樊兴华,赵静,方滨兴,李欲晓.影响力扩散概率模型及其用于意见领袖发现研究.计算机学报,2013,2:360-367. [doi: 10.3724/SP.J.1016.2013.00360]
- [35] 任晓龙,吕琳媛.网络重要节点排序方法综述.科学通报,2014,59(13):1175-1197. [doi: 10.1360/972013-1280]
- [41] 闵磊,刘智,唐向阳,陈矛,刘三妍.基于扩展度的复杂网络传播影响力评估算法.物理学报,2015,64(8):387-397.
- [59] 赵之滢,于海,朱志良,汪小帆.基于网络社团结构的节点传播影响力分析.计算机学报,2014,37(4):753-766.
- [60] 胡天成,尹龔榮,马鹏斐,高旸,张勇,邢春晓.一种新的网络传播中最有影响力的节点发现方法.物理学报,2013,62(14):9-19. [doi: 10.7498/aps.62.140101]
- [63] 苏晓萍,宋玉蓉.利用邻域“结构洞”寻找社会网络中最具影响力节点.物理学报,2015,64(2):020101. [doi: 10.7498/aps.64.020101]
- [64] 韩忠明,吴杨,谭旭升,段大高,杨伟杰.面向结构洞的复杂网络关键节点排序.物理学报,2015,64(5):429-437. [doi: 10.7498/aps.64.058902]
- [67] 段松青,吴斌,王柏.TTRank:基于倾向性转变的用户影响力排序.计算机研究与发展,2014,51(10):2225-2238.
- [82] 周涛,韩筱璞,闫小勇,杨紫陌,赵志丹,汪秉宏.人类行为时空特性的统计力学.电子科技大学学报,2013,42(4):481-540. [doi: 10.3969/j.issn.1001-0548.2013.04.001]
- [88] 毛佳昕,刘奕群,张敏,马少平.基于用户行为的微博用户社会影响力分析.计算机学报,2014,37(4):791-800.
- [97] 丁兆云,周斌,贾焰,张鲁民.微博中基于多关系网络的话题层次影响力分析.计算机研究与发展,2013,50(10):2155-2175.
- [98] 王晨旭,管晓宏,秦涛,周亚东.微博消息传播中意见领袖影响力建模研究.软件学报,2015,26(6):1473-1485. <http://www.jos.org.cn/1000-9825/4627.htm> [doi: 10.13328/j.cnki.jos.004627]
- [102] 赵佳,蒯莉,李静茹.社交网络中基于贝叶斯和半环代数模型的节点影响力计算机理.物理学报,2013,62(13):13-21.
- [115] 曹玖新,董丹,徐顺,郑啸,刘波,罗军舟.一种基于 k -核的社会网络影响最大化算法.计算机学报,2015,38(2):238-248.



韩忠明(1972—),男,山西文水人,博士,副教授,CCF 专业会员,主要研究领域为社会网络,数据挖掘,大数据处理.



原碧鸿(1992—),女,硕士生,CCF 学生会员,主要研究领域为数据挖掘,数据分析.



陈炎(1991—),男,学士,主要研究领域为互联网数据挖掘,机器学习,社会网络.



李梦琪(1993—),女,学士,CCF 学生会员,主要研究领域为深度学习,自然语言处理.



刘雯(1992—),男,学士,CCF 学生会员,主要研究领域为社交网络挖掘.



段大高(1976—),男,博士,副教授,CCF 专业会员,主要研究领域为异构数据挖掘,大数据处理,社会网络.