社交网络影响力研究综述

丁兆云[1] 贾 焰[2] 周 斌[2] 唐 府[3]
(国防科技大学信息系统与管理学院 长沙 410073)[1] (国防科技大学计算机学院 长沙 410073)[2]
(中国人民解放军77675部队40分队)[3]

摘 要 互联网正逐步演变为无处不在的计算平台和信息传播平台。在线社交网站、微博、博客、论坛、维基等社交网络应用的出现和迅猛发展,使得人类使用互联网的方式产生了深刻变革——由简单信息搜索和网页浏览转向网上社会关系的构建与维护、基于社会关系的信息创造、交流和共享。社交网络中个体间的交互形成影响力,社交网络中的影响力主要依赖个体间的关系强度、个体间的网络距离、时序因子以及网络特征与个体特征等。影响力分析技术的相关研究主要包括个体间的影响强度度量技术、个体影响力度量技术、影响力扩散机制等。

关键词 社会网络,数据挖掘,影响力,意见领袖
中图法分类号 TP391 **文献标识码** A

Survey of Influence Analysis for Social Networks

DING Zhao-yun[1] JIA Yan[2] ZHOU Bin[2] TANG Fu[3]
(School of Information System and Management, National University of Defense Technology, Changsha 410073, China)[1]
(School of Computer, National University of Defense Technology, Changsha 410073, China)[2]
(PLA77675, 40)[3]

Abstract The Internet is gradually evolved into a ubiquitous computing platform and information dissemination platform. Emergence and rapid development of online social networking sites, micro blogging, blogs, forums, wikis and social networking applications, make the form for the human to use the Internet to produce profound changes from simple information search and Web browser to construction and maintenance of online socialration information creation, exchange and sharing based on social relations. Of interaction between individuals in the social network influence, the influence of the social network is mainly dependent on the strength of the relationship between the individual, the network distance between individuals, the timing factor, as well as network characteristics and individual characteristics. Influential analysis technology related research includes individual impact strength measurement technology, individual influence and the diffusion mechanisms of influence.

Keywords Social networks, Data mining, Influence, Opinion leaders

1 引言

基于互联网的社交网络正在成为人类社会中社会关系维系和信息传播的重要渠道和载体,对国家安全和社会发展都会产生深远的影响:(1)社会个体通过各种连接关系在社交网络上构成"关系结构",包括以各种复杂关系关联而成的虚拟社区;(2)基于社交网络的关系结构中,大量网络个体围绕着某个事件而聚合,并相互影响、作用、依赖,从而形成具有共同行为特征的"网络群体"。(3)基于社交网络的关系结构和网络群体中,各类"网络信息"得以快速发布并传播扩散形成社会化媒体,并反馈到现实社会,从而使得社交网络与现实社会间形成互动,并对现实世界产生影响。

社交网络的意见领袖在虚拟社区、网络群体以及信息传播中发挥着巨大作用,能够快速扩散、放大舆论。社交网络意见领袖在激发舆论、推动议题讨论上扮演着越来越重要的角色。社交网络意见领袖针对舆论发表言论,与网民、媒体之间形成互动,其观点往往影响大批粉丝和舆论走向,社交网络中每个用户都能够发布自己的观点,广大网民都有机会通过社交网络形成意见领袖。近两年来,在"打拐"、贫困地区学童"免费午餐"等事件中,意见领袖起到了重要的参与和引导作用,而在拆迁、上访、事故灾难等突发事件上,意见领袖在事件产生、发酵、传播、爆炒等环节中占据重要地位。

如2010年9月江西抚州拆迁自焚事件中,《凤凰周刊》记者邓飞因以社交网络报道事件进展,引发大量粉丝的关注和转发,对事件的传播起着重要作用;2011年3月,日本因地震引发核泄漏事故后,国内东南沿海开始流传吃碘盐可防辐射、

到稿日期:2013-05-01 返修日期:2013-06-10 本文受国家自然科学基金项目(60933005,91124002),国家"八六三"高技术研究发展计划基金项目(2010AA012505,2011AA010702,2012AA01A401,2012AA01A402),国家科技支撑计划基金项目(2012BAH38B04,2012BAH38B06),国家242信息安全计划基金项目(2011A010)资助。

丁兆云(1982—),男,博士,讲师,主要研究方向为数据挖掘、信息安全,E-mail:zyding@nudt.edu.cn。

核泄漏污染了海盐等传闻，社交网络中一位拥有 210 万粉丝的台湾艺人，号召大家“多摄入含有碘的食物”，这条信息转发次数达到 18703 次，对抢盐现象起到了推波助澜的作用；2011 年“7.23 动车追尾”事件中，因意见领袖姚晨的参与发帖，导致该信息的转发量达 37907 条、评论量达 13101 条。

因此，社交网络意见领袖放大舆论，推波助澜的力量不容小视。如何挖掘意见领袖，分析社交网络中用户之间的影响强度以及每个用户的影响力扩散能力，依靠意见领袖积极引导社会舆论，提高新形势下舆情信息的分析能力，及时准确地掌握社会舆情动态，是社交网络所面临的严肃课题与严峻挑战。本文主要分析了社交网络中影响力分析的相关研究现状，主要包括影响强度度量、个体影响力度量与影响力扩散。

影响力分析技术在社会学、通信学、经济学、政治科学等领域被广泛研究，影响力分析技术在市场营销与社会运作，例如在时装推广[1]与政治选举[2]中起着重要作用。1955 年，Katz 等人[3]通过对美国总统选举时选民投票意向的研究，提出了两级传播理论，发现了个体影响力的差异性，小部分“意见领袖”(opinion leaders)或者“影响力个体”影响着大部分普通民众。1962 年，Rogers 等人[4]定义了“影响力个体”(influentials 或者 influencers)即擅长说服其他人的个体。影响力个体通常包括 4 个特点：1)容易将自己的观点传达给其他人；2)代表大多数普通人的观点；3)具有新颖的观点；4)也被称为舆论领袖(opinion leaders)、扩散创新理论的革新者(innovators)、网络中心(hubs)、网络桥节点(connectors)、专家(mavens)等。2003 年，文献[5]定义了社会影响力(social influence)即个人的行为能够直接或者间接地影响他人的想法、感情以及行动。2012 年，文献[6]定义了社会网络中的影响力(influence)，即受到网络中其他用户的影响而导致个人行为的改变，影响力在社会网络中是个普遍的现象。

2 影响强度度量技术

影响强度(influence strength)即是用户之间相互影响的能力。传统的社会网络中影响强度度量方法主要利用边的结构，依靠两个节点之间的共同邻居数目来度量影响强度。随后，学者考虑了社会网络中个体的行为特征与话题特征。个体的行为特征决定个体间的影响强度，比如交互更加频繁的用户之间通常具有更高的影响强度。个体之间在不同话题类别中通常表现为不同的影响强度，目前的相关研究主要利用统计机器学习方法度量个体在不同话题类别中的影响强度。

仅考虑边结构的影响强度度量方法：依靠两个节点之间的共同邻居数目来度量影响强度。针对社会网络中两个节点 A 与 B，其共同邻居数目越多，则影响强度越高，利用杰卡德相似系数(Jaccard coefficient)计算两节点的影响强度。

$$S(A,B)=\frac{|n_A \cap n_B|}{|n_A \cup n_B|}$$

式中，n_A 和 n_B 分别表示节点 A 与 B 的邻居。如果节点 A 与 B 之间拥有大量的共同邻居，则认为 A 与 B 为强关系(strong tie)，否则认为 A 与 B 为弱关系(weak tie)。依靠两节点的共同邻居数目能够比较直观地衡量个体之间的影响强度，但却忽略了个体自身特征之间的相似度以及交互频度，通常相似度越高的个体之间具有越高的影响强度，交互频度高的用户之间具有更高的影响强度。

考虑个体行为的影响强度度量方法：个体的行为特征包括个体特征的相似性、个体之间的交互频度等。文献[7-9]分别考虑了个体行为来度量个体间的影响强度。Goyal 等人[7]利用个体间的行为日志度量影响强度。Xiang 等人[8]在 Facebook 和 LinkedIn 数据集上利用个体之间的交互性和话题相似性，提出了潜在的变分模型来评估个体之间的影响强度。Aral 等人[9]利用 130 万 FaceBook 用户来计算社交网络的影响力与脆弱性。

考虑个体行为使得社会网络中个体间的影响强度度量更加准确，但却不能更加细粒度地度量个体在话题级别的影响强度。

利用统计机器学习方法度量话题级别的影响强度：统计机器学习方法假设用户之间的影响强度为一个隐变量，通过 EM 或者 Gibbs 抽样等迭代方法学习该隐变量。

统计机器学习方法代表性的研究工作为统计语言模型，2003 年，Blei 等人[10]首次提出了 LDA(Latent Dirichlet Allocation)模型，它用一个服从 Dirichlet 分布的 K 维隐含随机变量表示文档话题混合比例，模拟文档产生过程。2007 年，Dietz 等人[11]综合考虑文本内容与网络关系模拟文档的产生过程，假如一篇文档中词汇的产生来自两种途径：1)受到其他个体的影响而引用他人的词汇；2)自己的创新观点，使用独特新颖的词汇。因此个体使用他人的词汇越多，则受到的影响强度越高，作者使用 Gibbs 抽样方法，通过不断迭代计算每个词汇的来源以及对应的概率，以计算个体在每个话题类别中的影响强度。随后，Liu 等人[12,13]将统计语言模型应用于更大规模的学术数据与微博数据来度量话题级别的影响强度。Ding 等人[17]综合考虑了时间间隔与词汇流动性度量微博的个体间的影响强度。

同时，文献[14,15]针对不存在网络关系的新闻或者学术数据，利用统计语言模型度量文档之间的影响强度。Shaparenko 等人[14]针对不存在网络关系的学术数据，提出了概率模型，即利用统计测试方法度量影响力。Gerrish 等人[15]同样针对不存在网络关系的学术数据，提出了动态话题模型来度量论文间的影响力，且利用变分方法推理和评估隐参数。

另外，国内清华大学唐杰等人[16]为了度量个体在话题级别上的影响力，提出了话题近似传播(TAP)模型，并且基于 Map-reduce 思想实现了该模型。

统计机器学习方法能够度量话题级别的影响强度，但却忽略了个体间的时序关系与时间间隔，时间序列相似性高的个体之间通常具有更高的影响强度，个体间的信息时间间隔越短，影响强度越高。

3 个体影响力度量技术

社会网络由一个图 $G=\{V,E\}$ 表示，V 是节点集合，E 是边的集合。个体影响力度量技术的相关研究主要包括点度中心度(degree)、接近中心度(closeness)、中间中心度(betweenness)、HITS、PageRank 及扩展方法等。

点度中心度：指的是该节点的度数，即与该节点直接相连的节点个数。点度中心度用来分析节点直接影响力，即考察个体的直接社会关系。令 A 是网络图的邻接矩阵，$deg(i)$ 为节点 i 的度，则节点 i 的点度中心度 c_i^{DEG} 即为该节点的度数。

$$c_i^{DEG}=\deg(i)$$

节点的出度可以理解为一个节点对他人的影响程度，或该节点的活跃度[18]；节点的入度标志着该节点的受欢迎度[19]。国内乔少杰等人[20]在电子邮件数据中利用用户个性特征的正态分布模型模拟真实的邮件通信行为，发现犯罪网络的核心成员。国外 Nascimento 等人[5]在学术合作网络中将论文数量和引用数量作为衡量一个作者影响力的重要标志。Cha 等人[21]为了度量 Twitter 中个体影响力，分别计算了关注网络、转发网络、提及网络的点度中心度。Pal 等人[22]在 Twitter 数据集上考虑了个体的发帖数、回复数、被转发数、被提及数(mention)和粉丝数目，分别计算个体的转发影响力、被提及影响力和扩散影响力等。

点度中心度比较直观地衡量一个节点的影响力，计算开销相对较小，但针对大规模的微博网络，其将忽略部分影响力个体。

接近中心度：指个体与社交网络中所有其它节点的捷径距离(最短路径)之和。接近中心度用来分析个体通过社交网络对其它个体的间接影响力。节点 i 的接近中心度 c_i^{AO} 被定义如下。

$$c_i^{AO}=e_i^T S1$$

式中，S 为一个矩阵，它的第(i,j)个元素表示从节点 i 到节点 j 的最短路径长度；1 表示所有元素都为 1 的向量。

接近中心度需要计算网络中所有节点对之间的最短路径，计算开销大，优点是能够衡量一个节点的间接影响力。

中间中心度：指的是节点处于其它节点最短路径上的能力。中间中心度用来分析节点对信息传播的影响，即个体在多大程度上处于其他个体的中间，是否发挥出“中介”作用。节点 i 的中间中心度 c_i^{BET} 被定义如下。

$$c_i^{BET}=\sum_{j,k}\frac{b_{jik}}{b_{jk}}$$

式中，b_{jk} 表示节点 j 与 k 之间的最短路径数目；b_{jik} 表示节点 j 与 k 之间，且通过节点 i 的最短路径数目。计算中间中心度的朴素方法为计算所有节点对之间的最短路径，需要 $O(n^3)$ 的时间开销与 $O(n^2)$ 的空间开销。Brandes[23]提出了一种单源最短路径算法，它仅需要 $O(n+m)$ 的空间开销、$O(nm)$ 与 $O(nm+n^2\log n)$ 的时间开销，其中 n 表示节点数目；m 表示边的数目。中间中心度计算开销大，优点是能够找到网络中的“中介”节点。Quercia 等人[24]研究了 Twitter 关注网络结构，发现大多数网络结构洞是意见领袖，且发布各种话题，情感变化丰富。

HITS：由康奈尔大学的 Jon Kleinberg[25]提出，英文全称为 Hypertext Induced Topic Search，最初应用在搜索引擎中，根据一个网页的中心度(Hub)和权威度(Authority)来衡量网页重要性。对网络图中的每个节点 v_i，令 $a(v_i)$ 为该节点的权威度，$h(v_i)$为该节点的中心度。则节点权威度与中心度定义如下。

$$a^{(k+1)}(v_i)=\sum_{v_j\in inlink[v_i]}h^{(k)}(v_j)h^{(k+1)}(v_i)$$
$$=\sum_{v_j\in outlink[v_i]}a^{(k+1)}(v_j)$$

Romero 等人[26]为了度量 Twitter 中的个体影响力，综合考虑了影响力与冷漠性，提出了类 HITS 算法的 IP(influence-passivity)算法。

HITS 算法综合考虑了节点的权威度与中心度，需要迭代计算，但却忽略了节点影响力的划分。

PageRank：由 Google 创始人之一 Larry Page 提出[27]，最初应用在搜索引擎中，根据网页之间的超链接计算网页排名，一个页面的得票数由所有链向它的页面的重要性决定，但随后学者将 PageRank 算法应用到社会网络中，它是个体影响力度量的基础算法。PageRank 算法用一种基于马尔科夫的随机游走思想来模拟用户浏览网页的行为。令 π 为网络中节点影响力得分向量，P 为网络图的转移矩阵，则 PageRank 计算公式如下。

$$\pi=\alpha P^{\mathrm{T}}\pi+(1-\alpha)\frac{1}{n}e,e=(1,1,\cdots,1)^{\mathrm{T}}$$

式中，α 为跳转因子，$\frac{1}{n}e$ 为自重启(restart)向量。

Tunkelang 等人[28]为了度量 Twitter 中的个体影响力，针对 Twitter 中的关注关系构造了一种类似 PageRank 的算法，该算法用粉丝的影响力来衡量个体的影响力，粉丝影响力越高且关注的其他用户越少，则粉丝对该个体影响力的贡献越大。

PageRank 算法考虑了节点影响力的传播，需要迭代计算，但却忽略了节点自身特征，微博中用户行为表现复杂且用户规模数量庞大，仅依靠网络结构将忽略更加细粒度的影响力个体，比如无法发现话题层次的影响力个体。相关学者针对这一问题，在 PageRank 算法基础上，提出了结合个体特征与网络结构的影响力度量技术。

PageRank 算法扩展：社会网络中个体特征主要包括个体发布信息所属话题类别，研究每个话题类别的影响力个体；另外还包括个体发布信息的新颖度、敏感度等，创新能力强的个体通常具有更高的影响力，同时发布敏感信息的个体通常具有更高的影响力。Haveliwala 等人[29]考虑个体用户特征，在 PageRank 算法基础上，提出了 Personalized PageRank 算法，计算公式如下。

$$\pi=\alpha P^{\mathrm{T}}\pi+(1-\alpha)r$$

Personalized PageRank 算法将均分自重启向量$\frac{1}{n}e$ 改为个性化向量 r，比如元素 r_i 表示个体对话题的偏好程度、个体发布信息的新颖程度与敏感程度等。Agarwal 等人[30]在博客数据中综合考虑个体的知名度、活跃度、新颖度以及个体表达能力 4 个因数来衡量个体的影响力。Cai 等人[31]基于博客数据考虑了用户所属不同的兴趣领域，认为用户在不同的兴趣领域往往有不同的影响力。Hui 等人[32]基于博客数据，考虑了用户的情感(sentiment)，认为信誉(credibility)高的用户相对有更高的影响力。Singla[48]、Anagnostopoulos[49]以及 Crandal[50]等人通过分析社会网络中的用户行为，研究了用户行为属性和个体影响力的关系。

相关学者针对个体在不同话题类别中具有不同的影响力，在 PageRank 算法基础上，研究了话题层面的影响力个体，代表性的研究工作有 Weng 等人[33]的 TwitterRank 算法，即在 Twitter 数据集上，根据关注网络和用户兴趣相似性计算个体在每个话题上的影响力。Li 等人[34]依靠微博中的历史消息和社会交互记录，利用统计学习过程构造历史意见和意见影响力，提出了话题级的意见影响力模型，并合并了话题因素与社会影响力。另外学者综合考虑个体发布信息的新颖

度与网络结构，在 PageRank 算法基础上，研究了基于新颖度的影响力个体发现方法，代表性的研究工作有 Song 等人[35]的 InfluenceRank 算法，其在博客数据集上考虑了博文新颖度对网络的贡献，以识别博客中的意见领袖。

另外，Ding 等人[36]针对微博的多交互特性，提出了多关系网络的随机游走模型来度量微博个体影响力。

结合个体特征与网络结构的影响力度量技术综合考虑了个体的自身特性，使个体影响力度量相对更加准确，但却忽略了网络的多关系特性，比如微博中用户之间的交互存在多关系性，网络的多关系特性为个体影响力度量带来了新的挑战。

4 影响力扩散

影响力扩散的相关研究主要集中在影响扩散最大化(influence maximization)，影响扩散最大化最早由 Kempe 等人[37]提出，旨在挖掘社会网络中 K 个节点作为信息传播的初始种子节点，使得扩散的节点数量最大化。Kempe 等人首次利用子模块理论分析了影响力扩散最大化为 NP-hard 问题，以线形阈值模型与独立级联模型为基础给出了使影响力在整个网络中扩散最大化的 4 种启发式算法。

随机算法：从网络中随机选择 K 个节点作为传播种子节点。

基于节点中心度的启发式算法：优先选择节点度排名靠前的 K 个节点作为传播种子节点。

基于节点中间中心度的启发式算法：优先选择与其他所有节点最短路径之和最小的 K 个节点作为传播种子节点。

贪心算法：随机算法推广，随机选择 M 个节点作为传播源模拟传播过程，选择影响传播最大化的一个节点作为传播种子节点，如此迭代 K 次，找到 K 个节点作为传播种子节点。

影响力最大化为 NP-hard 问题，因此在最近几年被广泛研究[38-40]。Lee 等人[41]在 Twitter 数据集上模拟关注网络中的影响力传播，通过计算用户的有效读者数来衡量一个用户的影响力。Gruhl 等人[42]针对博客数据，利用疾病传染模型模拟话题传播过程，同时利用话题传播路径构造影响传播图。Java 等人[43]将博客网络关系图转换为影响扩散图，对博客中的影响传播建模。Bakshy 等人[44]在 Twitter 数据集中，根据每个相同的 URL 构造传播级联树，用种子节点的扩散范围来衡量每个种子节点的影响力。Kitsak 等人[45]研究了复杂网络中的高扩散能力用户发现方法，发现高扩散能力用户位于网络中依靠 k-shell 分解分析法的网络核心位置。Aggarwal 等人[46]提出了一种随机信息流模型来发现 Twitter 中有代表性的权威节点。Steeg 等人[47]利用转移熵理论刻画用户间的信息流，以识别 Twitter 网络中有影响力的链接。

影响力扩散最大化的实际应用难点在于线形阈值模型与独立级联模型的阈值难以确定，且要求信息沿着已知网络传播。微博的自身特点为信息传播带来了良好的滋生环境，使得以前复杂网络中仅能依靠模拟的传播过程得以现实化，但微博的开放性，打破了传统研究中的“封闭式”假设，微博中的信息传播不仅依靠网络结构，还与外部环境、信息本身等因素相关，从而使得传统的信息传播模型不能直接应用到微博中。

结束语 社会网络中影响力分析主要包括个体间的影响强度度量、影响力个体挖掘、影响力扩散能力度量等。影响强度度量研究中，相关学者通常利用图结构中的共同邻居数目和个体间的交互行为来度量个体间的影响强度。在影响力个体挖掘中，通过类 PageRank 的个体间投票算法来决定个体的影响力是一种非常普遍的做法。在影响力扩散能力的度量中，通常认为种子节点具有较强的扩散能力。尽管社会网络中的影响力分析研究已存在相应的进展，但随着微博的迅猛发展，微博的新特性使得微博的影响力研究面临诸多挑战。概括来讲，微博的富含噪音性使得部分垃圾用户表现为影响力个体的特征；微博的社会媒体性使得用户间的交互表现为更加细粒度的词汇生成与流动特性；微博的多关系性使得个体的投票行为表现为多关系网络的随机游走特性；微博的信息扩散快速性使得需要更加细粒度地利用时间因子来度量个体的影响力扩散能力。

1)噪音数据多样化特性。随着微博服务的普及，存在大量以刺探隐私情报、商业推销、推高用户人气、制造与传播舆论等为目的人工垃圾链接和垃圾用户。垃圾用户存在的目的不同，导致其行为特征的差异性与多样性。部分垃圾用户为达到特定目的，其行为特征类似于意见领袖。比如：1)连续发布大量的博文吸引正常用户的关注；2)连续地提及(mention)正常用户而吸引其他用户的关注；3)关注大量正常用户而吸引其他用户的关注。同时部分“虚假”的意见领袖为了提高自己的人气，利用技术手段或者微博服务漏洞，制造大量“僵死粉”，这些“僵死粉”行为特征具有隐蔽性，甚至部分“僵死粉”由特定商业公司操纵，其行为特征更加多样化。

微博中大量垃圾用户为影响力个体发现带来了新的挑战，因此提前准确地发现、过滤垃圾用户，可为影响力的研究降低数据集的噪音。

2)社会媒体特性。韩国科学技术院 Kwak 等人[51]研究表明微博不仅具有社交网络功能，更倾向于具有社会媒体功能，表现为社会媒体特性，用户使用微博通常具有两个目的：1)交友；2)发布与传播有价值的信息。用户受到其好友影响时，通常表现为“复制”其好友博文词汇，即更加细粒度的词汇生成与流动性，如图 1 所示。因此，仅依靠传统的共同邻居数目无法准确地度量用户间的影响强度。

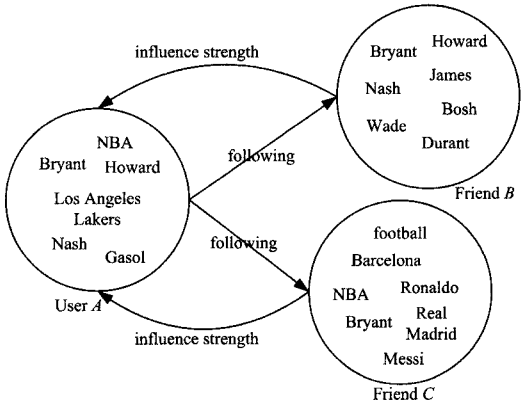

图 1 微博词汇生成与流动性示例

用户复制其好友博文词汇越多，则说明该用户受到其好友影响力越强。但用户发布博文中的词汇可能来自多个好友，例如图 1 中用户 A 的“Bryant”词汇同时来自其好友 B 与 C。因此，根据社会媒体性计算用户间影响强度的关键性挑战问题为如何计算词汇流动性的概率，即用户 A 的“Bryant”

词汇来自其好友 B 与 C 的概率分别为多少。

用户间的影响强度度量能够用来预测信息传播的路径，即当突发话题或者舆论事件发生后，根据已经受到影响的部分用户来预测其好友是否会传播该突发话题或者舆论事件，从而预测该突发话题或者舆论事件的影响范围。

3)个体交互多关系特性。微博中用户间交互复杂，当用户 A 受到其好友 B 影响时，可能表现多种行为：1)用户 A 在自己的博文中使用类似“RT @B” 或者“via @B”，转发用户 B 的博文；2)用户 A 在自己的博文中使用类似“@B”，回复用户 B 的博文；3)用户 A 没有明确地使用类似“RT @B” 或者“via @B”等转发类型标签，复制用户 B 的博文；4)用户 A 阅读用户 B 的博文。因此，微博中用户间的交互存在多关系特性，用户间的影响网络为一个多关系网络(multi-relational network)。如何针对多关系网络计算多关系网络的转移概率，以及如何融合多关系网络，构造多关系网络的随机游走模型，将是微博中影响力个体挖掘面临的挑战性问题。

微博中影响力个体挖掘对舆情控制与引导起着重要作用，微博中影响力个体分布广泛，处于网络中的各个位置，且涉及到各个话题。当突发话题或者舆论事件发生后，影响力个体能够在网络中的各个位置传播自己的信息、发布自己的意见，起到了引发舆论和影响舆论的作用。同时，影响力个体在市场营销与产品推广中也起着重要作用，影响力个体依靠自己的网络位置以及独特的“人格魅力”能够迅速扩散产品，达到市场营销与产品推广的目的。

4)信息扩散快速特性。微博的转发功能(“RT @”)使得信息无限制快速地被转发，从而使得微博具有信息扩散快速特性。因此，不能够简单地认为种子节点具有较强的扩散能力，尽管一个用户首次发布了一条舆论事件，但该信息很久才被少量用户转发，则该用户并不具有较强的扩散能力。用户的影响力扩散能力通常与 4 个因素相关：1)用户在传播级联中的位置，由于微博的信息扩散快速特性，通常认为在舆论事件发生初期参与的用户具有越强的扩散能力；2)用户影响的其他用户数目，通常认为转发一个用户博文的其他用户数目越多，则该用户具有越强的扩散能力；3)用户影响的其他用户自身的扩散能力，通常认为被扩散能力强的用户传播的用户具有较强的扩散能力；4)用户扩散信息的速度，通常认为用户扩散信息速度越快，扩散能力越强。如何综合考虑上述 4 个因素，建立影响力扩散能力度量模型，将是微博中影响力扩散能力度量面临的挑战性问题。

微博中的影响力扩散能力度量使得能够准确地发现舆论事件中的意见领袖，方便梳理舆论事件的传播过程，即发现哪些用户在舆论事件传播过程中起了推波助澜的作用。同时，意见领袖由于自身的活跃性以及对舆论事件的积极参与性，通常能够代表舆论事件的演化和发展趋势，因此应实时监控意见领袖的博文内容以及行为状态，尽可能快速地预测舆论事件的演化和发展趋势。

参 考 文 献

[1] Gladwell M. The Tipping Point: How Little Things Can Make a Big Difference [M]. New York: Little Brown, 2000

[2] Berry J, Keller E. The Influentials: One American in Ten Tells the Other Nine How to Vote, Where to Eat, and What to Buy [M]. New York: The Free Press, 2003

[3] Katz E, Lazarsfeld P. Personal Influence: The Part Played by People in the Flow of Mass Communications [M]. New York: The Free Press, 1955

[4] Rogers E M. Diffusion of Innovations [M]. New York: The Free Press, 1962

[5] Cialdini R B. Influence: Science and Practice [M]. Boston: Allyn and Bacon, 2003

[6] Aggarwal C C. Social Network Data Analytics [M]. New York: Springer, 2012

[7] Goyal A, Bonchi F, Lakshmanan L V S. Learning influence probabilities in social networks [C]// the 3rd ACM International Conference on Web Search and Data Mining (WSDM'10). New York, USA, February 2010: 241-250

[8] Xiang R, Neville J, Rogati M. Modeling relationship strength in online social networks [C]// the 19th International Conference on World Wide Web (WWW'10). Raleigh, USA, April 2010: 981-990

[9] Aral S, Walker D. Identifying influential and susceptible members of social networks [J]. Science, 2012, 337(6092): 337-341

[10] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation [J]. Journal of Machine Learning Research, 2003, 3: 993-1022

[11] Dietz L, Bickel S, Scheffer T. Unsupervised Prediction of Citation Influences [C]// the 24th International Conference on Machine Learning (ICML'07). Corvallis, USA, June 2007: 233-240

[12] Liu L, Tang J, Han J, et al. Mining topic-level influence in heterogeneous networks [C]// the 19th ACM International Conference on Information and Knowledge Management (CIKM'10). Toronto, Canada, October 2010: 199-208

[13] Liu L, Tang J, Han J, et al. Learning influence from heterogeneous social networks [J]. Data Mining and Knowledge Discovery, 2012, 25(3): 511-544

[14] Shaparenko B, Joachims T. Information genealogy: Uncovering the flow of ideas in non-hyperlinked document databases [C]// the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07). San Jose, USA, August 2007: 619-628

[15] Gerrish S M, Blei D M. A language-based approach to measuring scholarly impact [C]// the 26th International Conference on Machine Learning (ICML'10). Haifa, Israel, November 2010: 375-382

[16] Tang J, Sun J, Wang C, et al. Social influence analysis in large-scale networks [C]// the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09). Paris, France, June 2009: 807-816

[17] Ding Z Y, Jia Y, Zhou B, et al. An influence strength measurement via time-aware probabilistic generative model for microblogs[C]// the 15th Asia-Pacific Web Conference. Sydney, Australia, April 2013

[18] Wasserman S, Faust K. Social Network Analysis: Methods and Applications [M]. London: Cambridge University Press, 1994

[19] Aggarwal C, Wang H. Managing and Mining Graph Data [M]. New York: Springer, 2010

[20] 乔少杰，唐常杰，彭京，等. 基于个性特征仿真邮件分析系统挖掘

犯罪网络核心[J]. 计算机学报,2008,31(10):1795-1803

[21] Cha M, Haddadi H, Benevenuto F, et al. Measuring user influence in Twitter: The million follower fallacy [C]// the 4th International AAAI Conference on Weblogs and Social Media (ICWSM'10). Washington, USA, May 2010: 10-17

[22] Pal A, Counts S. Identifying topical authorities in microblogs [C]// the 4th ACM International Conference on Web Search and Data Mining (WSDM'11). Hong Kong, China, February 2011: 45-54

[23] Brandes U. A faster algorithm for betweenness centrality [J]. Journal of Mathematical Sociology, 2001, 25: 163-177

[24] Quercia D, Capra L, Crowcroft J. The social world of Twitter: Topics, geography, and emotions [C]// the 6th International AAAI Conference on Weblogs and Social Media (ICWSM'12). Dublin, Ireland, June 2012: 298-305

[25] Kleinberg J M. Authoritative sources in a hyperlinked environment [J]. Journal of the ACM, 1999, 46(5): 604-632

[26] Romero D M, Galuba W, Asur S, et al. Influence and passivity in social media [C]// the 20th International Conference Companion on World Wide Web (WWW'11). Hyderabad, India, March 2011: 113-114

[27] Page L, Brin S, Motwani R, et al. The PageRank citation ranking: bringing order to the Web [R/OL]. http://ilpubs.stanford.edu:8089/422/, 1999

[28] Tunkelang D. A Twitter analog to PageRank [EB/OL]. http://thenoisychannel.com/2009/01/13/a_twitter_analog_to_pagerank/, 2009

[29] Haveliwala T, Sepandar K, Glen J. An analytical comparsion of approaches to personalizing PageRank [R/OL]. http://ilpubs.stanford.edu:8089/596/, 2003

[30] Agarwal N, Liu H, Lei T, et al. Identifying the influential bloggers in a community [C]// Proc of the 1th ACM International Conference on Web Search and Data Mining. New York, NY: ACM, 2008: 207-217

[31] Cai Y, Chen Y. Mining influential bloggers: From general to domain specific [C]// Proc of the 13th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems. Berlin: Springer, 2009: 447-454

[32] Hui P, Gregory M. Quantifying sentiment and influence in blogspaces [C]// Proc of the 1st Workshop on Social Media Analytics. New York, NY: ACM, 2010: 53-61

[33] Weng J, Lim E P, Jiang J, et al. TwitterRank: Finding topic-sensitive influential twitterers [C]// the 3rd ACM International Conference on Web Search and Data Mining (WSDM'10). New York, USA, February 2010: 261-270

[34] Li D, Shuai X, Sun G, et al. Mining topic-level opinion influence in microblog [C]// the 21st ACM International Conference on Information and Knowledge Management (CIKM'12). Maui, USA, October 2012: 1562-1566

[35] Song X, Yun C, Hino K, et al. Identifying opinion leaders in the blogosphere [C]// the 16th ACM International Conference on Information and Knowledge Management (CIKM'07). Lisboa, Portugal, November 2007: 971-974

[36] Ding Z Y, Jia Y, Zhou B, et al. Mining topical influencers based on the multi-relational network in micro-blogging sites [J]. China Communications, 2013, 10(1): 93-104

[37] Kempe D, Kleinberg J, Tardos E. Maximizing the spread of influence through a social network [C]// the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03). Washington, USA, August 2003: 137-146

[38] Manuel G R, Leskovec J, Krause A. Inferring networks of diffusion and influence [C]// the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'10). Washington, USA, July 2010: 1019-1028

[39] Chen W, Wang C, Wang Y. Scalable influence maximization for prevalent viral marketing in large-scale social networks [C]// the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'10). Washington, USA, July 2010: 1029-1038

[40] Wang C, Chen W, Wang Y. Scalable influence maximization for independent cascade model in large-scale social networks [J]. Data Mining and Knowledge Discovery, 2012, 25(3): 545-576

[41] Lee C, Kwak H, Park H, et al. Finding influentials based on the temporal order of information adoption in Twitter [C]// the 19th International Conference Companion on World Wide Web (WWW'10). Raleigh, USA, April 2010: 1137-1138

[42] Gruhl D, Guha R, Liben-Nowell D, et al. Information diffusion through blogspace [C]// Proc of the 13th International World Wide Web Conference. New York, NY: ACM, 2004: 43-52

[43] Java A, Kolari P, Finin T, et al. Modeling the spread of influence on the blogosphere [R]. Maryland: UMBC, 2006

[44] Bakshy E, Hofman J M, Mason W A, et al. Everyone's an influencer: Quantifying influence on Twitter [C]// the 4th ACM International Conference on Web Search and Data Mining (WSDM'11). Hong Kong, China, February 2011: 65-74

[45] Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks [J]. Nature Physics, 2010, 6: 888-893

[46] Aggarwal C C, Khan A, Yan X. On flow authority discovery in social networks [C]// the 11th SIAM International Conference on Data Mining (SDM'11). Phoenix, USA, April 2011: 522-533

[47] Steeg G V, Galstyan A. Information transfer in social media [C]// the 21st International Conference on World Wide Web (WWW'12). Lyon, France, April 2012: 509-518

[48] Singla P, Richardson M. Yes, there is a correlation -From social networks to personal behavior on the Web [C]// Proc of the 17th International World Wide Web Conference. New York, NY: ACM, 2008: 655-664

[49] Anagnostopoulos A, Kumar R, Mahdian M. Influence and correlation in social networks [C]// Proc of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: ACM, 2008: 7-15

[50] Crandall D, Cosley D, Huttenlocher D, et al. Feedback effects between similarity and social influence in online communities [C]// Proc of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY: ACM, 2008: 160-168

[51] Kwak H, Lee C, Park H, et al. What is Twitter, A social network or a news media? [C]// the 19th International Conference on World Wide Web (WWW'10). Raleigh, USA, April 2010: 591-600