



Research
Cybersecurity—Review

社会影响力分析——模型、方法和评价

李侃*, 张林, 黄河燕

School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China

ARTICLE INFO

Article history:

Received 10 December 2017

Revised 5 January 2018

Accepted 8 January 2018

Available online 16 February 2018

关键词

社会影响力分析

在线社交网络

社会影响力分析模型

影响力评价

摘要

社会影响力分析 (SIA) 是一个广泛的研究领域,吸引了诸多研究者的兴趣。本文总结了 SIA 相关的模型、方法和评价方面代表性的工作,并将 SIA 模型归纳为两种类型:微观模型和宏观模型。微观模型考虑人与人之间的相互影响和影响的过程,而宏观模型认为每个人具有相同的传播概率和影响力。本文分析了包括影响最大化、影响最小化、影响流和个人影响力等的社会影响力分析方法;介绍了影响力评价指标,并分析了社会影响力评价模型。本文的目标是对社会影响力提供全面的分析,旨在辅助理解社会行为,为舆论影响提供理论基础,并揭示未来的研究方向和潜在的应用。

© 2018 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 引言

在线社交网络,如微博、Twitter和Facebook等,为用户的信息传播提供了有价值的平台。在信息传播过程中,社会影响力伴随一个人的观点、情绪或行为受到其他人的影响而产生[1]。个人的态度、想法、感受或行为会因为与其他人或团体发生交流而改变。社会影响力分析 (SIA) 正在成为社交网络中一个重要的研究领域。SIA主要研究如何模拟网络上影响扩散的过程,以及如何提出一种有效的方法来识别网络中的相应的目标节点[2]。这些研究包括:谁影响了谁、谁受到了影响、谁是最有影响力的人等。SIA具有重要的社会意义,并已在许多领域得到了应用。病毒营销[3–10]、在线推荐[11]、医疗社区[12–14]、专家发现[15–17]、

谣言扩散[18]和其他领域[19–21]都依赖于社会影响效应。分析社会影响力可以帮助理解人的社会行为,为公众决策和影响舆论提供理论支持,并可促进各类活动的交流和传播[22]。

本文从模型、方法、评价等方面对SIA进行了全面的分析。基于此,本文分析研究了现有模型、方法以及社会影响力评价模型的优缺点。首先介绍了现有的社会影响力分析模型;然后总结了社会影响力相关计算方法;最后分析了社会影响力评价的相关工作。

本文的组织如下:第2节讨论了SIA模型;第3节分析了包括影响最大化、影响最小化、影响流和个人影响力等的SIA方法;第4节详细论述了社会影响力评价工作;最后,本文总结了所提到的社会影响力模型和方法,并讨论了这个领域开放的问题。

* Corresponding author.

E-mail address: likan@bit.edu.cn (K. Li)

2. 社会影响力分析模型

SIA模型已经被广泛地研究，本文将这些模型归纳为两类：微观模型和宏观模型。

2.1. 微观模型

微观模型关注人与人之间的相互作用，并考虑影响过程。广泛使用的两个影响力分析模型是独立级联模型 (independent cascade, IC) [23–25] 和线性阈值模型 (linear threshold, LT) [23,26]。这两种模型最早由 Kempe 等 [23] 提出，现在仍主要用于模拟社会影响力扩散过程。

2.1.1. IC 和 LT 模型

IC模型将社会网络抽象为一个有向图 $G=(V, E)$ ，其中 V 代表节点集合， E 代表节点之间的有向边的集合，同时还存在一个种子节点集 $S(S \subseteq V)$ ， $S_t(S_t \subseteq V)$ 表示时间 $t(t \geq 0)$ 被激活的节点集合。在 $t+1$ 时刻，每个节点 v_i 会以概率 P_{ij} 去激活它的非激活状态下的邻居节点。当没有节点被激活时，影响过程结束。在此过程中，每一个节点在被激活后有且只有一个机会去激活他的邻居节点，被激活节点会一直保持激活状态。

LT模型同样将网络抽象为一个有向图 $G=(V, E)$ ，每一个节点与其所有邻居节点之间的影响权重之和满足 $\sum_{v_j \in N_{\text{gact}}} w_{ij} \leq 1$ 。其中 w_{ij} 表示节点 v_i 和它的邻居节点 v_j 之间的影响权重； N_{gact} 表示被节点 v_i 激活的邻居节点集合；节点 v_i 按照 0 和 1 之间的均匀分布随机选择自己的阈值 θ_i 。只有当节点 v_i 激活状态下的邻居节点影响权重之和超过了 v_i 的阈值， v_i 才会被激活。

2.1.2. 扩展的模型

在IC和LT模型中，为了计算某个节点的影响通常需要利用蒙特卡洛模拟过程，而这是一个非常耗时的操作，且不适合大规模社交网络。因此，很多研究者提出了IC和LT模型的改进模型。本文将这些改进模型划分为4类：基于IC的模型、基于LT的模型、基于IC和LT的模型、其他模型。这里只列举出在这些方面有代表性的模型。

(1) 基于IC的模型。一些研究人员在IC模型基础之上，考虑了时间延迟和受时间限制的影响扩散模型。Chen等[27]扩展了IC模型并提出了具有相遇事件的IC模型 (IC model with meeting events, IC-M)。在IC-M

模型中，激活节点以一定概率遇到非激活节点。虽然IC-M模型计算的时间复杂度略高于一般IC模型，但是计算结果更接近实际情况。Feng等[28]基于以前的研究，发现信息重复曝光会减小对用户的影响。因此，他们在IC模型中加入了时间衰减因素，提出了一种基于传播路径的算法来评价种子节点的影响力。在提出的新模型中社会网络的每个有向边有两个值：传播概率和预计传播延迟时间。Mohamadi-Baghmolaei等[29]考虑了时间和信任因素，提出了一种基于信任的延迟感知独立级联模型 (trust-based latency-aware independent cascade, TLIC)。这是首次在经典IC模型中考虑信任因素。在TLIC模型中，信任邻居节点和不信任邻居节点按照不同的概率改变一个节点的状态 (如激活状态或者非激活状态)。Budak等[30]研究了在社会网络中两种影响相互竞争和相互影响抵消的过程，提出了多层IC模型，该模型可以同时模拟两个级联传播的过程。

(2) 基于LT的模型。Liu等[31]从遏制社会网络中竞争对手的影响扩散出发，将LT模型改进为扩散-遏制模型 (diffusion-containment, D-C)。传统的LT模式不适用于涉及扩散和遏制影响的情况。在D-C模型中，某个节点的状态受激活概率的控制，每个节点只会受到传播概率较高的邻居节点的影响。同样每个节点和所有邻居节点之间的传播概率之和小于或等于1。Borodin等[32]分析了基于LT模型的不同模型的竞争影响扩散效果。

(3) 基于IC和LT的模型。Mohammadi等[33]考虑了时间延迟提出了两种扩散模型：延迟独立级联模型 (delayed independent cascade, DIC) 和延迟线性阈值模型 (delayed linear threshold, DLT)。在这两个模型中，节点有3种状态：活动、非活动和潜在活动。为了从非活动状态转变为活动状态，节点必须经过处于潜在活动状态的中间过程。与传统的IC和LT模型相比，该模型中影响扩散的效果更好。但由于存在多个状态，因此计算复杂度相对较高。在IC和LT模型中，信息扩散被视为一系列以同步方式发生的节点状态变化。但是，实际的扩散是以异步方式进行的，并且已知数据的时间戳并不是均匀分布。因此有些模型放弃了传统IC和LT模型的同步假设，并扩展了这两种模型以使状态可以异步变化。Sait等[34]对IC和LT模型进行了异步扩展，提出了异步独立级联模型 (asynchronous IC, AsIC) 和异步线性阈值模型 (asynchronous LT, AsLT)。他们的算法可以从相对较少的信息扩散结果中估计网络中节点的影响程

度,从而避免过度拟合问题。Guille和Hacid[35]也提出了一个异步模型来模拟扩散过程——基于时间的异步独立级联模型(time-based AsIC)。其他研究者[30,32,36-44]也在不同的方向对经典模型进行了改进。Fa等[38]构建了两个模型:机会主义一对一模型(the opportunistic one-activate-one, OPOAO)和确定性一对多模型(the deterministic one-activate-many, DOAM)。在OPOAO模型中每个人只能同时与另外一个人交流,而DOAM模型类似于信息广播过程的机制。对于OPOAO模型,他们使用经典的贪婪算法产生一个 $(1-1/e)$ 的近似解;对于DOAM模型,他们提出了一种基于覆盖的贪婪算法(set cover-based greedy, SCBG)来实现 $O(\ln n)$ 求解(n 是节点数),这两个模型中每个用户的传播概率都是相同的。Galam[45]提出了一个调查意见的动态模型, Lee等[39]对Galam提出的模型进行改进,但是这些模型仅限于口碑信息传播的过程。

(4) 其他模型。一些模型与IC或LT模型及其变体不同,它们从新的角度解决了信息影响扩散问题。Lee等[46]提出了一个数据驱动模型,利用元学习概念求解影响力最大化。但这种模式需要大量的数据,其结果的准确性还需要进一步改善。Golnari等[47]提出了一个热传导模型(heat conduction, HC)。与之前仅考虑逐步传播过程的IC或LT模型完全不同,HC模型考虑非逐步传播过程,传播初始化时有一组种子节点,其他节点可以任意赋值。Wang等[48]研究了在大规模图像社交网络中的情感影响,并提出了情感影响模型。他们设计了一个因子图模型来推断来自社交网络中图像的情感影响。Gao[49]提出了一个读写模型(read-write, RW)来描述意见形成、影响和扩散的详细过程。然而,这个模型需要进一步考虑3个主要问题:许多模型必需的参数怎样推断、有关意见影响和扩散的数据集怎样收集以及适合于此任务的评价指标。

无论是经典IC、LT模型或新模型,社会影响力分析模型的最终目标都是利用更少的内存实现更快的计算和更精确的结果,但是这些微观模型缺乏一些细节。首先,由于个体特征的差异(即教育背景和个人意识),不同的人对同一信息的接受程度不同;其次,不同的人对他人的影响力不同,即信息发布者的权威不同对邻居和社会的影响力也不同;第三,个人将信息传递给其他人的概率并不是恒定的,应该取决于个人对信息的感兴趣程度。而且这些模型都使用二元变量来记录个体是否受到影响,假设一旦一个人受到影响,这个人再也不会改变

其状态。然而,这种假设并不能真实反映出一个人从一种状态向另一种状态的自然变化。

2.2. 宏观模型

宏观模型认为所有人对信息具有相同的兴趣,同时也具有相同的传播概率和相同的影响力。由于宏观模型不考虑个体,所以这些模型计算结果的准确率较低。因此,为了改进这些模型,应该考虑个体之间的差异。宏观模型将节点划分为不同的类别(即状态),并关注每个类别中节点的状态演变。每个类别中节点的百分比用简单的微分方程表示。传染病模型是用于从宏观角度研究社会影响的最常见模型,这些模型主要是为了模拟传染病传播过程,但是它们忽视了社交网络的拓扑特征。每个类别中节点的百分比的计算方式过于简单,不能准确描述这种复杂的演化过程。Daley和Kendall[50]分析了传染病扩散与信息传播之间的相似性,并提出了经典的Daley-Kendall模型。从那以后,研究人员不断改进这些传染病模型来克服它们的缺陷。Daley和Kendall[50]、Moreno等[51]、Nekovee等[52]和Zhou等[53]在他们的方法中考虑网络底层的拓扑特征,然而忽视了影响扩散过程中个人行为的影响。

最近,研究人员开始考虑信息影响扩散过程中人类行为和不同机制的作用。Zhao等[59,60]在同质和非同质网络中提出了susceptible-infected-hibernator-removed模型作为经典susceptible-infected-removed模型(SIR)的延伸,以便纳入遗忘和记忆机制。Wang等[54]在概率模型中考虑了社会的正面或负面影响,分析了社会强化如何影响传播速度提出了一个在线社交网络中的谣言扩散模型的变体。Wan等[55]通过引入节点之间的信任机制提出了一个SIR模型,这种机制可以减少谣言最终影响范围和谣言扩散速度。Xia等[56]考虑到谣言内容的吸引力和模糊性,并且认为谣言越清晰,其影响越小,提出了一个修改后的susceptible-exposed-infected-removed模型(SEIR)用于研究犹豫机制对谣言传播模型的影响。Su等[57]通过考虑用户的不完整阅读行为,提出了microblog-susceptible-infected-removed的信息传播模型。Liu等[58]受到Xia等[56]和Su等[57]工作的激发,在SEIR模型基础上提出了一个新的异构网络SEIR模型用以研究微博的扩散动态。

3. 社会影响力分析方法

社会影响力分析方法用于解决影响最大化、影响最

小化、影响传播和个体影响力等社会网络影响分析的子问题。所有这些问题都涉及影响的扩散，所以在某些情况下，同一个影响传播模型可以用于所有这些问题。但是对于每一个问题，使用该模型的最终目的不同。

3.1. 影响最大化

影响最大化需要找到社交网络中最具影响力的成员集合。Kempe等[23]对这个问题进行了定义：给定以用户为节点的有向图以及预算或阈值数 k ，其中有向图中边的权重代表用户之间相互影响的程度，影响最大化的目的是在社交网络中找到 k 个节点，以便通过激活这些节点得到最大的影响传播范围。这是一个离散优化问题，同时，对于IC和LT模型都是NP难问题。影响最大化是社会影响力分析中研究最为广泛的问题，Leskovec等[10]和Rogers [20]将其当做一个算法问题，并提出了一些概率计算方法。影响最大化算法需要满足计算速度快、结果精度高和计算存储要求低这3个标准。考虑个体差异的大多数算法可以分为贪婪算法或启发式算法。

3.1.1. 贪婪算法

贪婪算法在每次迭代中“贪婪地”选择对现有种子集具有最大边际增益的活动节点。贪婪算法的研究是基于爬山贪婪算法，在这种算法中每次选择都可以利用局部最优解来逼近全局最优解，把影响力最大的节点加入种子集合中。社会影响力分析中很多用于研究个体差异的算法都是基于这些贪婪算法，这些算法的优点是精度比较高，达到了 $(1-1/e-\epsilon)$ ， ϵ 可以是任何大于0的数值，然而算法复杂度高、执行时间长、效率相对较低。Kempe等[23]首次把求解影响最大化作为离散优化问题，并提出了一种贪婪近似算法。Leskovec等[61]提出了一种贪婪的优化方法——cost-effective lazy forward (CELF) 方法。Chen等[27]提出了新的贪婪算法：New-Greedy算法和MixGreedy算法。Zhou等[62]提出了基于上限的延迟转发算法（upper bound-based lazy forward, UBLF），用来发现最具影响力的 k 个节点。

3.1.2. 启发式算法

由于贪婪算法的计算复杂度很高，许多优秀的启发式算法已经被提出用来缩短求解时间并追求更高的算法效率。这些启发式算法基于特定启发信息（如节点的度或PageRank排序结果）迭代地选择节点，而不用每次

迭代过程中计算节点的边际增益。但是启发式算法的缺点是准确性相对较低。最基本的启发式算法是由Kempe等[23]提出的随机算法、基于度的算法和中心启发式算法。Chen等[41]根据基于度的启发式算法提出了独立级联模型（IC）的启发式算法——DegreeDiscount，之后他们还提出了一种新的启发式算法——prefix excluding maximum influence arborescence (PMIA) 算法[63]。对于LT模型，Chen等[63]提出了基于局部有向无环图启发式算法（local directed acyclic graph, LDAG）。还有很多启发式算法也是基于这个启发式思想，如SIMPATH[64]和IRIE[65]。除此之外，Borgs等[66]在IC模型下提出了一个线性时间复杂度用于求解影响最大化的算法，取得了理论突破。Tang等[67]提出了一个两阶段影响最大化的两阶段算法（two-phase influence maximization, TIM），这个算法的时间复杂度是 $O(k+l)(n+m)\log n/\epsilon^2$ ，求解结果以至少 $(1-n^{-l})$ 的概率符合 $(1-1/e-\epsilon)$ 近似。

上述算法的时间复杂度如表1所示。

在这里，本文介绍了一些考虑到个体差异或用户自身属性的代表性研究。Li等[68]为了模拟异构社交网络对解决影响最大化问题的作用，考虑了人类之间的行为关系（即正常和稀有行为）。他们提出了两种基于熵的启发式算法来识别网络中的传播者，然后最大化传播的

表1 不同算法时间复杂度比较

Algorithm	Time complexity
Hill-climbing greedy[23]	$O(knRm)$
CELF[61]	$O(knRm)$
NewGreedyIC[41]	$O(kRm)$
NewGreedyWC[41]	$O(kRTm)$
MixGreedyIC[41]	$O(kRm)$
MixGreedyWC[41]	$O(kRTm)$
DegreeDiscountIC[41]	$O(k \log n + m)$
PMIA[63]	$O[nt_{i\theta} + kn_{o\theta}n_{i\theta}(n_{i\theta} + \log n)]$
LDAG[63]	$O(l_v + n \log l_v)$
SIMPAT[64]	$O(kmn)$
IRIE[65]	$O(kmn)$
Quasilinear time algorithm[66]	$O[k\epsilon^{-2}(m+n) \log n]$
TIM[67]	$O[(k+l)(m+n) \log n/\epsilon^2]$

n : number of vertices in G ; m : number of edges in G ; k : number of seeds to be selected; R : number of rounds of simulations; T : number of iterations; $t_{i\theta}$, $n_{o\theta}$, $n_{i\theta}$: constants decided by θ ; θ : the influence threshold; $n_{i\theta} = \max_{v \in V} \{MIIA(v; \theta)\}$; $n_{o\theta} = \max_{v \in V} \{MIOA(v; \theta)\}$; $MIIA(v; \theta)/MIOA(v; \theta)$: the maximum influence in arborescence/out arborescence of a node v ; $t_{i\theta}$: the maximum running time to compute $MIIA(v; \theta)$; l : number of communities in the network; ϵ : any constant larger than 0; l_v : the volume of LDAG(v ; θ).

影响。尽管基于熵的启发式性能优于Degree [23]和DegreeDiscount[41], 但该方法仍然基于启发式思想, 准确性不足。Subbian等[69]提出个人社会价值的概念, 他们发现现有的影响最大化算法不能模拟个人的社会价值, 但是这通常是相互联系的真实动机。因此他们基于社会资本的概念提出了一个框架, 通过网络中的绑定和桥接连接的数量来计算网络的社会价值。该算法的性能优于PMIA[63], PageRank和基于度的方法。Li等[70]提出了从众意识级联模型 (conformity-aware cascade, C^2) 和从众感知贪婪算法来解决影响最大化问题。这个算法适用于分布式平台时效果很好, 但由于基于贪婪算法并考虑从众意识级联, 所以计算的复杂性仍需要改善。Lee和Chung[71]将影响最大化问题定义为查询处理问题, 即区分特定用户与其他用户。由于影响最大化查询处理是NP难问题, 且求解目标函数也是NP难问题, 所以他们侧重如何有效地近似求解最优种子节点。Deng等[72]发现求解影响最大化问题中在计算不同用户的影响时, 节点特征总是被忽视。因此他们提出了3种量化措施来分别评价节点特征: 用户活跃度、用户敏感度和用户亲和度。他们将节点特征融合到用户的静态影响中, 然后使用连续指数衰减函数来转换两个相邻用户之间用户动态影响的强度。他们还提出了带有节点特征的信用分布模型 (credit distribution with node feature, CD-NF), 该模型重新定义了信用, 并且基于CD-NF模型设计了具有节点特征的贪婪算法 (greedy algorithm with node feature, GNF)。

考虑影响最大化问题时, 个人的属性会对结果产生很大的影响; 因此, 需要考虑到个体差异或用户的属性。从以上关于个体差异研究的描述中可以清楚地看出, 大多数研究都是基于贪婪算法以追求高准确度, 然而由于考虑个体差异, 与原始贪婪算法计算相比, 计算复杂度将进一步提高, 导致算法效率低。因此, 目前需要改善计算的复杂度和结果的准确率, 为了缩短运行时间、提高算法效率, 可以考虑使用启发式算法, 也可以在分布式平台上进行大量工作。

3.2. 影响最小化

假设负面信息在网络 $G=(V, E)$ 中传播, 最初被影响的节点集合为 $S(S \subseteq V)$, 影响最小化的目标是指通过干预阻塞 k 个节点的传播使最终受影响的范围最小。阻塞的节点集合用 D 表示 $D \subseteq V$, 而且 k ($k \ll |V|$)是一个常数, 可以表示为以下优化问题:

$$D^* = \arg \min_{D \subseteq V, |D| \leq k} \sigma(S|V \setminus D) \quad (1)$$

式中, $\sigma(S|V \setminus D)$ 表示当集合 D 中的节点被阻塞后, S 集中节点的影响力。

从这个概念[73]中, 我们知道影响最小化是影响最大化的对偶问题。影响最小化主要用于遏制谣言、监督舆论等。Yao等[73]提出了一种方法, 通过从主题模型的角度阻止有限数量的节点传播来减少网络中的负面影响的传播。当社交网络中出现谣言和其他不良事件时, 已有一些用户受影响, 这个模型的目的是最大限度地减少最终受到影响的最终用户数量。Wang等[74]提出了一个考虑用户感受的动态谣言影响最小化模型。这个模型通过阻止部分节点激活, 最大限度地减少了谣言的影响。Groeber等[75]受社会心理学中认知失调的概念启发设计了一个社会影响力的总体框架, 其中个体将因与邻居在某一社交网络上的不一致而产生的不和谐的先决条件最小化。Chang等[76]提出了一种用于估计成功限制受影响数量低于失控阈值的概率, 它可以允许当局适应性地调整干预成本以满足必要的风险控制, 避免爆发风险。他们还提出了一个影响最小化模型, 以有效防止网络上易发生大规模舆情的扩散。

3.3. 影响流

信息传播会伴随着影响流[77]。近年来, 许多研究人员提出了研究影响流的方法。Subbian等[78]提出了一种具有特定流动有效性约束条件的流模式挖掘方法。Kutskov等[79]提出了一种叫STRIP的流式方法来计算社交网络中每个链接的影响力。Teng等[80]研究了包括美国物理学会、Facebook、Twitter和LiveJournal在内的各种真实平台的信息流, 然后利用用户的行为模式构建虚拟信息影响扩散过程。Chintakunta和Gentimis[81]讨论了社交网络的拓扑结构与其内部信息流之间的关系。然而, 大多数社交网络与微博平台不同, 不能提供足够的上下文来挖掘流模式。

3.4. 个体影响力

个体影响力是一种相对微观的评估, 它可以模拟用户对其他用户或整个社交网络的影响。Chintakunta和Gentimis[82]提出了一种名为SoCap的方法在社交网络中找到有影响力的人, 他们将在社交网络中的发现有影响力的人建模为价值分配问题, 其中分配的价值代表个人社会资本。Subbian等[82]提出了一种在开放式智

能体系统中,使用矩阵分解方法,通过度量网络中节点的影响力来挖掘有影响的智能体的方法。Liu等[83]提出了面向信任的社会影响力方法(trust-oriented social influence, TOSI),该方法考虑了社会背景(即参与者之间的社会关系和社会信任)和用户偏好,来评价个人影响力。TOSI方法在有效性、效率和鲁棒性方面大大优于SoCap。Deng等[84]在评估不同用户的影响时,纳入了时间因素和节点的特征。结果表明,他们的方法对于识别种子节点是合理有效的,并且其对影响扩散的预测结果比原来忽略了扩散过程中节点特征的方法更准确。

总之,考虑更全面的用户特征和用户交互信息会带来更精确的结果。

4. 社会影响力分析评价

4.1. 分析评价指标

运行时间是一个评价模型效率非常直观的度量标准,而且易于计算。一般来说,在相同条件下,模型运行得越快,效果就越好。然而,传统的贪婪算法需要通过大量重复的蒙特卡洛模拟来计算节点的影响力,导致消耗很长的时间,特别是在当前大规模社交网络上,现有的算法无法满足应用对效率的要求。因此,运行时间是衡量的重要手段。

由于求解影响传播问题是NP难问题,因此很难获得目标函数的最优解。现在的大多数算法都依赖函数的单调性和子模性来满足 $(1-1/e)$ 的近似解[23]。然而,尝试实现更高的近似求解从未停止过。Zhu等[85]考虑影响传递性和限制传播距离,在他们的模型中提出了基于半定的算法。

另一个指标是蒙特卡罗模拟的次数。由于无法获得最优解,因此通常使用蒙特卡洛模拟来估计实际价值。现有的贪婪算法需要在初始阶段对每个节点的传播函数进行重复的蒙特卡洛模拟计算,这大大降低了模型的效率。Zhou等[62]提出的UBLF算法可以减少CELF方法中蒙特卡洛模拟数量的95%以上,并且当种子集小时,可以实现了2~10倍的加速。

预期传播程度表示种子集最终可能影响的节点数量,受影响的节点数量越多越好。在现实生活场景中许多需要尽可能最大影响范围的应用,这种应用的典型例子是市场营销和广告。在这两种应用中,最终的预期结果为产品促销的好处或产品的盈利能力。因此,探索高预期的传播算法是社会影响力分析的一个重要研究内容。

鲁棒性是指在维持一些其他性能不变的前提下,允许某些参数(如结构、大小)扰动的特性。Jung等[65]和Liu等[83]在他们的算法中都提到了鲁棒性。Jung等[65]提出的IRIE算法在各种密度网络和不同大小级联下运行时间和内存使用方面更稳健和稳定。实验结果表明,大规模网络中IRIE算法的运行速度比PMIA[63]等现有方法运行速度快两个数量级,并且使用的内存更少。Liu等[83]提出的TOSI评估方法稳定性比SoCap[81]表现得更出色。

可伸缩性是在对现有系统影响很小的情况下能够不断扩展或增强系统的功能,在社会网络中可伸缩性通常是指从小型网络扩展到大型网络的能力,这是用于评估模型质量的常用指标。由于算法复杂度高、运行时间长,目前的算法仅适用于节点数在100万以下的中小型社会网络,鉴于当今的大型社交网络,必须设计具有良好可扩展性的影响力分析算法,以应对大量社会网络数据带来的挑战。

4.2. 社会影响力评价模型

社会影响力评价是一个复杂的过程。社会关系作为一种主观属性,具有动态性、事件差异性、不对称性、传递性等多种特征,而且在社会网络中频繁的用户互动和网络结构的变化使得社会影响力评价更加困难。关于社会影响力评价模型的研究一直都有学者在研究。He等[86]设计了一个关于在线投诉主题的影响度量模型,该模型基于熵权模型、实时监控和分析投诉信息的静态和动态属性。企业可以使用这种模式来管理在线用户的投诉。Wang等[12]提出了一个细粒度的基于特征的社会影响力评价模型(fine-grained feature-based social influence, FBI),该模型探讨了用户的重要性以及用户影响他人的可能性。然后他们通过对朋友的影响力贡献设计一个基于PageRank算法的社会影响力评价模型。FBI评价模型在HEPTH[87]、DBLP[88]和Arnet-Miner[89]3个数据集上验证可以出色识别出影响较大的TOP-K用户。

5. 结论和下一步工作

本文从模型、方法和评价等方面综述了SIA的最新研究进展,分析了当前模型和方法的优缺点,并揭示了未来的研究方向和潜在的应用。

在社会影响力分析模型中,本文归纳了两种类型的

模型：微观模型和宏观模型。微观模型考虑人的相互作用和影响过程；宏观模型认为所有用户有相同的传播概率和相同的影响力，未来的宏观模型应着眼于信息传播过程中如何加入人的行为和实现不同的机制。虽然许多研究者在改进经典模型和从不同视角提出新模型方面做了大量的努力，例如在模型中加入约束和引入竞争影响传播，但仍有改进的余地。

在大多数的现有模型中，考虑的都是单一网络中一个人受其他人的影响，而且影响扩散过程是独立的，然而在现实生活中，人们经常在多网络中与他人交流。在这种情况下，社会影响发生在多网络，不同的单一网络之间的影响会存在合作和竞争关系。如何在多网络中构建信息传播模型是一个有价值的研究课题。同时在动态网络中如何计算跨时间影响传播的问题也值得被研究。除此之外，在大多数实验中数据集涵盖了大约100 000个节点，因此将社交网络分析相关问题如何应用于海量数据集（可能包括数百万或数千万甚至更多节点）仍需要研究。简而言之，在扩展SIA模型以解决效率和可伸缩性等局限性的问题上仍有研究空间。

致谢

本文研究部分得到了中国国家重点基础研究项目（2013CB329605）的支持。感谢李玲玲和商军英对本文的建设性意见和支持。

Compliance with ethics guidelines

Kan Li, Lin Zhang and Heyan Huang declare that they have no conflict of interest or financial conflicts to disclose.

References

- [1] Travers J, Milgram S. The small world problem. *Psychol Today* 1967;1:61–7.
- [2] Chen W, Lakshmanan LV, Castillo C. Information and influence propagation in social networks. San Rafael: Morgan & Claypool; 2013.
- [3] Freeman LC. A set of measures of centrality based on betweenness. *Sociometry* 1977;40(1):35–41.
- [4] Baas F. A new product growth model for consumer durables. *Manage Sci* 1969;15(5):215–27.
- [5] Brown JJ, Reingen PH. Social ties and word-of-mouth referral behavior. *J Consum Res* 1987;14(3):350–62.
- [6] Mahajan V, Muller E, Bass FM. New product diffusion models in marketing: A review and directions for research. *J Mark* 1990;54(1):1–26.
- [7] Domingos P, Richardson M. Mining the network value of customers. In: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2001 Aug 26–29; San Francisco, CA, USA; 2001. p. 57–66.
- [8] Goldenberg J, Libai B, Muller E. Talk of the network: a complex systems look at the underlying process of word-of-mouth. *Mark Lett* 2001;12(3):211–23.
- [9] Richardson M, Domingos P. Mining knowledge-sharing sites for viral marketing. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2002 Jul 23–26; Edmonton, AB, Canada; 2002. p. 61–70.
- [10] Leskovec J, Adamic LA, Huberman BA. The dynamics of viral marketing. *J ACM Trans Web* 2007;1(1):5.
- [11] Pálóvics R, Benczúr AA, Kocsis L, Kiss T, Frigó E. Exploiting temporal influence in online recommendation. In: Proceedings of the 8th ACM Conference on Recommender Systems; 2014 Oct 6–10; Foster City, CA, USA; 2014. p. 273–80.
- [12] Wang G, Jiang W, Wu J, Xiong Z. Fine-grained feature-based social influence evaluation in online social networks. *IEEE Trans Parallel Distrib Syst* 2014;25(9):2286–96.
- [13] Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med* 2007;357(4):370–9.
- [14] Fowler JH, Christakis NA. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *B Med J* 2009;338(7685):23–7.
- [15] Franks H, Griffiths N, Anand SS. Learning influence in complex social networks. In: Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems; 2013 May 6–10; Saint Paul, MN, USA; 2013. p. 447–54.
- [16] Dong W, Pentland A. Modeling influence between experts. In: Proceedings of the ICM 2006 and IJCAI 2007 International Conference on Artificial Intelligence for Human Computing; 2006 Nov 3; Banff, AB, Canada; 2007. p. 170–89.
- [17] Tang J, Sun J, Wang C, Yang Z. Social influence analysis in large-scale networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2009 Jun 28–Jul 1; Paris, France; 2009. p. 807–16.
- [18] He Z, Cai Z, Wang X. Modeling propagation dynamics and developing optimized countermeasures for rumor spreading in online social networks. In: Proceedings of the 2005 IEEE 35th International Conference on Distributed Computing Systems; 2015 Jun 29–Jul 2; Columbus, OH, USA; 2015. p. 205–14.
- [19] Katz E, Lazarsfeld PF. Personal influence: The part played by people in the flow of mass communications. New York: Free Press; 1965.
- [20] Rogers EM. Diffusion of innovations. 5th ed. New York: Free Press; 2003.
- [21] Keller E, Berry J. The influentials: one American in ten tells the other nine how to vote, where to eat, and what to buy. New York: Free Press; 2003.
- [22] Peng S, Yang A, Cao L, Yu S, Xie D. Social influence modeling using information theory in mobile social networks. *Inf Sci* 2017;379:146–59.
- [23] Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. In: Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2003 Aug 24–27; Washington, DC, USA; 2003. p. 137–46.
- [24] Leskovec J, Mcglohon M, Faloutsos C, Glance NS, Hurst M. Patterns of cascading behavior in large blog graphs. In: Proceedings of the 2007 SIAM International Conference on Data Mining; 2007 Apr 26–28; Minneapolis, MN, USA; 2007.
- [25] Gruhl D, Guha R, Liben-Nowell D, Tomkins A. Information diffusion through blogspace. In: Proceedings of the 13th International Conference on World Wide Web; 2004 May 17–20; New York, NY, USA; 2004. p. 491–501.
- [26] Granovetter M. Threshold models of collective behavior. *Am J Sociol* 1978;83(6):1420–43.
- [27] Chen W, Lu W, Zhang N. Time-critical influence maximization in social networks with time-delayed diffusion process. In: Proceedings of the 26th AAAI Conference on Artificial Intelligence; 2012 Jul 22–26; Toronto, ON, Canada; 2012. p. 592–8.
- [28] Feng S, Chen X, Cong G, Zeng Y, Chee YM, Xiang Y. Influence maximization with novelty decay in social networks. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence; 2014 Jul 27–31; Québec City, QC, Canada; 2014. p. 37–43.
- [29] Mohamadi-Baghmolaei R, Mozafari N, Hamzeh A. Trust based latency aware influence maximization in social networks. *J Eng App Artif Intell* 2015;41(C):195–206.
- [30] Budak C, Agrawal D, Abbadi AE. Limiting the spread of misinformation in social networks. In: Proceedings of the 20th International Conference on World Wide Web; 2011 Mar 28–Apr 1; Hyderabad, India; 2011. p. 665–74.
- [31] Liu W, Yue K, Wu H, Li J, Liu D, Tang D. Containment of competitive influence spread in social networks. *Knowl Base Syst* 2016;109(C):266–75.
- [32] Borodin A, Filmus Y, Oren J. Threshold models for competitive influence in social networks. In: Proceedings of the 6th International Conference on Internet and Network Economics; 2010 Dec 13–17; Stanford, CA, USA; 2010. p. 539–50.
- [33] Mohammadi A, Saraee M, Mirzaei A. Time-sensitive influence maximization in social networks. *J Inf Sci* 2015;41(6):765–78.
- [34] Saito K, Ohara K, Yamagishi Y, Kimura M, Motoda H. Learning diffusion probability based on node attributes in social networks. In: Proceedings of the 19th International Conference on Foundations of Intelligent Systems; 2011 Jun 28–30; Warsaw, Poland; 2011. p. 153–62.
- [35] Guille A, Hacid H. A predictive model for the temporal dynamics of information diffusion in online social networks. In: Proceedings of the 21st International Conference on World Wide Web; 2012 Apr 16–20; Lyon, France; 2012. p. 1145–52.
- [36] Bharathi S, Kempe D, Salek M. Competitive influence maximization in social networks. In: Proceedings of the 3rd International Conference on Internet and Network Economics; 2007 Dec 12–14; San Diego, CA, USA; 2007. p. 306–11.
- [37] Carnes T, Nagarajan C, Wild SM, Zuylen AV. Maximizing influence in a com-

- petitive social network: A follower's perspective. In: Proceedings of the 9th International Conference on Electronic Commerce; 2007 Aug 19–22; Minneapolis, MN, USA; 2007. p. 351–60.
- [38] Fan L, Lu Z, Wu W, Thuraisingham B, Ma H, Bi Y. Least cost rumor blocking in social networks. In: Proceedings of the 2013 IEEE 33rd International Conference on Distributed Computing Systems; 2013 Jul 8–11; Philadelphia, PA, USA; 2013. p. 540–9.
- [39] Lee W, Kim J, Yu H. CT-IC: Continuously activated and time-restricted independent cascade model for viral marketing. In: Proceedings of the 2012 IEEE 12th International Conference on Data Mining; 2012 Dec 10–13; Brussels, Belgium; 2013. p. 960–5.
- [40] Kostka J, Oswald YA, Wattenhofer R. Word of mouth: Rumor dissemination in social networks. In: Proceedings of the 15th International Colloquium on Structural Information and Communication Complexity; 2008 Jun 17–20; Villars-sur-Ollon, Switzerland; 2008. p. 185–96.
- [41] Chen W, Wang Y, Yang S. Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2009 Jun 28–Jul 1; Paris, France; 2009. p. 199–208.
- [42] Wang Y, Wang H, Li J, Gao H. Efficient influence maximization in weighted independent cascade model. In: Proceedings of the 21st International Conference on Database Systems for Advanced Applications; 2016 Mar 27–30; Dallas, TX, USA; 2016. p. 49–64.
- [43] Pathak N, Banerjee A, Srivastava J. A generalized linear threshold model for multiple cascades. In: Proceedings of the 2010 IEEE International Conference on Data Mining; 2010 Dec 13–17; Sydney, Australia; 2010. p. 965–70.
- [44] Bharathi S, Kempe D, Salek M. Competitive influence maximization in social networks. In: Proceedings of the 3rd International Workshop on Web and Internet Economics; 2007 Dec 12–14; San Diego, CA, USA; 2007. p. 306–11.
- [45] Galam S. Modelling rumors: the no plane Pentagon French hoax case. *Phys A* 2003;320:571–80.
- [46] Lin SC, Lin SD, Chen MS. A learning-based framework to handle multi-round multi-party influence maximization on social networks. In: Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2015 Aug 10–13; Sydney, Australia; 2015. p. 695–704.
- [47] Golnari G, Asiaee A, Banerjee A, Zhang ZL. Revisiting non-progressive influence models: Scalable influence maximization. In: Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence; 2015 Jul 12–16; Amsterdam, The Netherlands; 2015.
- [48] Wang X, Jia J, Tang J, Wu B, Cai L, Xie L. Modeling emotion influence in image social networks. *IEEE Trans Affect Comp* 2015;6(3):286–97.
- [49] Gao D. Opinion influence and diffusion in social network. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information; 2012 Aug 12–16; Portland, OR, USA; 2012. p. 997.
- [50] Daley DJ, Kendall DG. Epidemics and rumors. *Nature* 1964;204(4963):1118.
- [51] Moreno Y, Nekovee M, Pacheco AF. Dynamics of rumor spreading in complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 2004;69(6 Pt 2):066130.
- [52] Nekovee M, Moreno Y, Bianconi G, Marsili M. Theory of rumor spreading in complex social networks. *Phys A* 2007;374(1):457–70.
- [53] Zhou J, Liu Z, Li B. Influence of network structure on rumor propagation. *Phys Lett A* 2007;368(6):458–63.
- [54] Wang H, Deng L, Xie F, Xu H, Han J. A new rumor propagation model on SNS structure. In: Proceedings of the 2012 IEEE International Conference on Granular Computing; 2012 Aug 11–13; Hangzhou, China; 2012. p. 499–503.
- [55] Wang Y, Yang X, Han Y, Wang X. Rumor spreading model with trust mechanism in complex social networks. *Commun Theor Phys* 2013;59(4):510–6.
- [56] Xia L, Jiang G, Song B, Song Y. Rumor spreading model considering hesitating mechanism in complex social networks. *Phys A* 2015;437:295–303.
- [57] Su Q, Huang J, Zhao X. An information propagation model considering incomplete reading behavior in microblog. *Phys A* 2015;419:55–63.
- [58] Liu Q, Li T, Sun M. The analysis of an SEIR rumor propagation model on heterogeneous network. *Phys A* 2017;469:372–80.
- [59] Zhao L, Wang J, Chen Y, Wang Q, Cheng J, Cui H. SIHR rumor spreading model in social networks. *Phys A* 2012;391(7):2444–53.
- [60] Zhao L, Qiu X, Wang X, Wang J. Rumor spreading model considering forgetting and remembering mechanisms in inhomogeneous networks. *Phys A* 2013;392(4):987–94.
- [61] Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, Glance N. Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2007 Aug 12–15; San Jose, CA, USA; 2007. p. 420–9.
- [62] Zhou C, Zhang P, Zang W, Guo L. On the upper bounds of spread for greedy algorithms in social network influence maximization. *IEEE Trans Knowl Data Eng* 2015;27(10):2770–83.
- [63] Chen W, Wang C, Wang Y. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2010 Jul 25–28, Washington, DC, USA; 2010. p. 1029–38.
- [64] Goyal A, Lu W, Lakshmanan LVS. SIMPATH: An efficient algorithm for influence maximization under the linear threshold model. In: Proceedings of the 2011 IEEE 11th International Conference on Data Mining; 2011 Dec 11–14; Vancouver, BC, Canada; 2012. p. 211–20.
- [65] Jung K, Heo W, Chen W. IRIE: a scalable influence maximization algorithm for independent cascade model and its extensions. *Rev Crim* 2011;56(10):1451–5.
- [66] Borgs C, Brautbar M, Chayes J, Lucier B. Maximizing social influence in nearly optimal time. In: Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms; 2014 Jan 5–7; Portland, OR, USA; 2014. p. 946–57.
- [67] Tang Y, Xiao X, Shi Y. Influence maximization: Near-optimal time complexity meets practical efficiency. In: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data; 2014 June 22–27; Snowbird, UT, USA; 2014. p. 75–86.
- [68] Li CT, Lin SD, Shan MK. Influence propagation and maximization for heterogeneous social networks. In: Proceedings of the 21st International Conference on World Wide Web; 2012 Apr 16–20; Lyon, France; 2012. p. 559–60.
- [69] Subbian K, Sharma D, Wen Z, Srivastava J. Social capital: The power of influencers in networks. In: Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems; 2013 May 6–10; Saint Paul, MN, USA; 2013. p. 1243–4.
- [70] Li H, Bhowmick SS, Sun A. CINEMA: Conformity-aware greedy algorithm for influence maximization in online social networks. In: Proceedings of the 16th International Conference on Extending Database Technology; 2013 Mar 18–22; Genoa, Italy; 2013. p. 323–34.
- [71] Lee JR, Chung CW. A query approach for influence maximization on specific users in social networks. *IEEE Trans Knowl Data Eng* 2015;27(2):340–53.
- [72] Deng X, Pan Y, Shen H, Gui J. Credit distribution for influence maximization in online social networks with node features. *J Intell Fuzzy Syst* 2016;31(2):979–90.
- [73] Yao Q, Zhou C, Shi R, Wang P, Guo L. Topic-aware social influence minimization. In: Proceedings of the 24th International Conference on World Wide Web; 2015 May 18–22; Florence, Italy; 2015. p. 139–40.
- [74] Wang B, Chen G, Fu L, Song L, Wang X. DRIMUX: dynamic rumor influence minimization with user experience in social networks. *IEEE Trans Knowl Data Eng* 2017;29(10):2168–81.
- [75] Groeber P, Lorenz J, Schweitzer F. Dissonance minimization as a microfoundation of social influence in models of opinion formation. *J Math Sociol* 2014;38(3):147–74.
- [76] Chang CW, Yeh MY, Chuang KT. On the guarantee of containment probability in influence minimization. In: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; 2016 Aug 18–21; San Francisco, CA, USA; 2016. p. 231–8.
- [77] Faisan MM, Bhavani SD. Maximizing information or influence spread using flow authority model in social networks. In: Proceedings of the 10th International Conference on Distributed Computing and Internet Technology; 2014 Feb 6–9; Bhubaneswar, India; 2014. p. 233–8.
- [78] Subbian K, Aggarwal C, Srivastava J. Content-centric flow mining for influence analysis in social streams. In: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management; 2013 Oct 27–Nov 1; San Francisco, CA, USA; 2013. p. 841–6.
- [79] Kutzkov K, Bifet A, Bonchi F, Gionis A. STRIP: Stream learning of influence probabilities. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2013 Aug 11–14; Chicago, IL, USA; 2013. p. 275–83.
- [80] Teng X, Pei S, Morone F, Makse HA. Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks. *Sci Rep* 2016;6(1):36043.
- [81] Chintakunta H, Gentimis A. Influence of topology in information flow in social networks. In: Proceedings of the 2016 AsiaSim Conference on Signals, Systems and Computers; 2016 Nov 6–9; Pacific Grove, CA, USA; 2017. p. 67–71.
- [82] Subbian K, Sharma D, Wen Z, Srivastava J. Finding influencers in networks using social capital. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; 2013 Aug 25–28; Niagara Falls, ON, Canada; 2013. p. 592–9.
- [83] Liu G, Zhu F, Zheng K, Liu A, Li Z, Zhao L, et al. TOSI: a trust-oriented social influence evaluation method in contextual social networks. *Neurocomputing* 2016;210:130–40.
- [84] Deng X, Pan Y, Wu Y, Gui J. Credit distribution and influence maximization in online social networks using node features. In: Proceedings of the 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery; 2015 Aug 15–17; Zhangjiajie, China; 2016. p. 2093–100.
- [85] Zhu Y, Wu W, Bi Y, Wu L, Jiang Y, Xu W. Better approximation algorithms for influence maximization in online social networks. *J Comb Optim* 2015;30(1):97–108.
- [86] He J, Hu M, Shi M, Liu Y. Research on the measure method of complaint theme influence on online social network. *Expert Syst Appl* 2014;41(13):6039–46.
- [87] Gehrke J, Ginsparg P, Kleinberg J. Overview of the 2003 KDD cup. *ACM SIGKDD Explor Newslett* 2003;5(2):149–51.
- [88] Yang J, Leskovec J. Defining and evaluating network communities based on ground-truth. *Knowl Inf Syst* 2015;42(1):181–213.
- [89] Tang J, Zhang J, Yao L, Li J, Zhang L, Su Z. ArnetMiner: Extraction and mining of academic social networks. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2008 Aug 24–27; Las Vegas, NV, USA; 2008. p. 990–8.