计算机科学
COMPUTER SCIENCE

基于大数据的进化网络影响力分析研究综述

何强，尹震宇，黄敏，王兴伟，王源田，崔硕，赵勇

引用本文

何强，尹震宇，黄敏，王兴伟，王源田，崔硕，赵勇. 基于大数据的进化网络影响力分析研究综述[J]. 计算机科学，2022，49(8): 1-11.

HE Qiang, YIN Zhen-yu, HUANG Min, WANG Xing-wei, WANG Yuan-tian, CUI Shuo, ZHAO Yong. Survey of Influence Analysis of Evolutionary Network Based on Big Data[J]. Computer Science, 2022, 49(8): 1-11.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210700240

基于大数据的进化网络影响力分析研究综述

何 强[1,2] 尹震宇[1] 黄 敏[4] 王兴伟[3] 王源田[2] 崔 硕[2] 赵 勇[3]
1 中国科学院沈阳计算技术研究所有限公司 沈阳 110168
2 东北大学医学与生物信息工程学院 沈阳 110169
3 东北大学计算机科学与工程学院 沈阳 110169
4 东北大学信息科学与工程学院 沈阳 110819
(heqiang@bmic.neu.edu.cn)

摘 要 社交影响力分析能够在社交网络中执行复杂行为分析,是现代信息和服务行业最重要的技术之一,越来越多的社交网络研究者把关注点放在社交影响力上。真实的社交网络是不断演化的而非静态的,进化网络的提出也带来了新的挑战和机遇,同时进化网络中海量的社交信息也为大数据分析技术的快速发展提供了强有力的支撑。文中对进化网络和影响最大化问题进行了论述,并讨论了社交影响力分析问题的传播模型和基于社交网络大数据的影响力分析方法,同时进一步整理了一些应用广泛的影响力算法。此外,还论述了大数据、进化网络与社交影响力最大化的关系。文中的目标是通过大规模社交网络中的影响力分析,帮助其他研究人员更好地理解现有的工作,为社交网络影响力分析提供新的思路。

关键词:进化网络;社交影响力;大数据;机器学习
中图法分类号 TP393

Survey of Influence Analysis of Evolutionary Network Based on Big Data

HE Qiang[1,2], YIN Zhen-yu[1], HUANG Min[4], WANG Xing-wei[3], WANG Yuan-tian[2], CUI Shuo[2] and ZHAO Yong[3]
1 Shenyang Institute of Computing Technology Co. Ltd., CAS, Shenyang 110168, China
2 College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110169, China
3 College of Computer Science and Engineering, Northeastern University, Shenyang 110169, China
4 College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

Abstract One of the most important technologies in modern information and service industry is social influence analysis. More and more researchers in social networks focus on social influence. Real social networks are evolving rather than static. The proposal of evolutionary network also brings new challenges and opportunities. At the same time, the massive social information in the evolutionary network also provides strong support for the rapid development of big data analysis technology. In this paper, evolutionary network and influence maximization are discussed. It also discusses the diffusion model of social influence analysis and the influence analysis method based on social network big data. At the same time, some widely used influence algorithms are further sorted out. In addition, this paper also discusses the relationship between big data, evolutionary networks, and social influence maximization. This paper aims to help other researchers to better understand the existing work and provide new ideas for the influence analysis of social networks through the influence analysis of large-scale social networks.

Keywords Evolutionary network, Social influence, Big data, Machine learning

在信息爆炸的时代,人们的日常生活每时每刻都在产生、传递以及交换信息,社交网络大数据时代已经来临。社交影响力分析在社交网络研究中仍旧占据主要地位,被广泛应用于网络和图形挖掘,如病毒营销、推荐系统、源检测、在社区中寻找领导者、网络拓扑优化和谣言抑制等。与此同时,基于大数据社交网络的应用为社交影响力分析带来了机遇和挑战。

社交网络中的影响力分析问题是社交网络中的研究热点之一,其核心是测量社交网络中用户的影响力,提取有影响力

到稿日期:2021-07-25 返修日期:2022-05-08
基金项目:国家重点研发计划(2021YFC3300300);辽宁省博士启动项目(2021-BS-055);中国博士后科学基金(2021M693318);中央高校基本业务费(N2119004,N2119007)
This work was supported by the National Key Research and Development Program(2021YFC3300300), Doctor Startup Foundation of Liaoning Province(2021-BS-055), China Postdoctoral Science Foundation(2021M693318) and Fundamental Funds for the Central Universities(N2119004, N2119007).
通信作者:尹震宇(1164041005@qq.com)

的用户。我们通常用关系图描述社交网络，图中的点代表社交网络中的用户，边代表网络中用户之间的联系。越来越多的社会学家研究社交网络，以便更好地理解使人们相互之间建立关系并产生影响的因素；研究者和公众人物也利用Twitter和Facebook等社交媒体发布消息，从而产生效应；营销专家利用社交网络获取自己需要的市场信息，并通过获取影响力大的节点来投递有针对性的广告进而增加效益。

现阶段的社交网络中影响力分析的研究大多假设社交网络是静态的，而社交网络通常是动态演化而非静止的。真实的社交网络中存在节点的添加和删除、边的添加和删除，还有边权重的修改，即两个点之间关系强度的变化等。也就是说，网络的结构随着时间的变化而变化，即进化网络。在实际应用中，这样的例子随处可见，如分析一个新员工来到公司之后的融入情况，就需要分析这个员工在一段时间内建立的社交关系，以及关系的远近，进而可以得到这段时间内这名员工的交友是否有效、在公司的社交是否正常等。而进化网络研究的难点在于网络的演化中只能获取已知的部分信息，剩下的信息甚至是隐藏的。本文首先对进化网络、大数据分析以及影响力分析的相关内容及其联系进行阐述。同时，为了更好地分析进化网络中的影响力分析问题，论述了影响力分析如何在进化网络中高效地识别和维护有影响力的节点、使用的传播模型以及算法等。

1 相关内容概述

1.1 进化网络中的社交影响力分析

近年来，社交网络范围不断扩大，影响力分析随之成为了社交网络中的研究热点。影响力分析问题[1]在实际生活中有着广泛的应用，如谣言传播模型中的源头检测、传染病模型中的污染和爆发检测等。这些模型在分析节点的影响力时，通常将有影响力的种子集识别出来并保存，输入到网络中进行分析，最后输出最终的分析结果[2-3]，具体流程如图1所示。这些模型在走向实际应用时面临的最大问题就是如何将影响力最大化的成本降至最低[4]。一些研究人员根据部分群体的行为做出分析，然而样本中人群本身也会受其他有影响力的人影响，但是样本并没有涵盖所有高影响力的用户[5-6]。为此，国外的最新研究引入了一种新颖的反向影响力最大化方法[7]来估计成本。这种方法的工作原理就是确定激活网络中一组给定的目标节点所需的最小邻居数量，以与影响力最大化问题相反的方式工作。

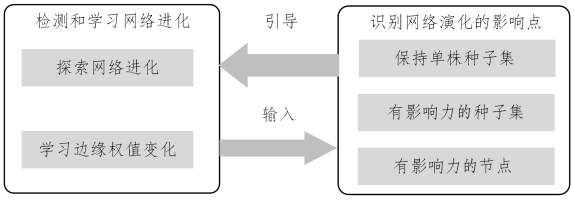

图1 进化网络中的影响力分析

Fig. 1 Influence analysis in evolutionary network

随着网络中节点的增加以及网络随时间不断演化，网络的复杂性不断提高[8]，社交网络等静态网络中的节点影响力分析并不适合进化网络，网络中有影响力的节点的分析面临着巨大挑战。因此，如何在进化网络中提取出关键节点进行分析至关重要。如图2所示，在影响力节点提取的整个过程中，首先对收集到的数据进行预处理，再评估节点的影响力。目前大多数节点重要性评估算法与网络拓扑参数息息相关[9]，如中心度和结构孔。中心度算法节点的影响力依赖于邻居节点的数量，优点是计算复杂度较低，适用于如今节点较多的社交网络，缺点是忽略了邻居节点拓扑的影响，导致测量得到的结果的准确度不高。为了改善其缺点，先前的研究者提出了结构孔的算法。该算法注重节点所在的位置，节点的影响通过节点邻居的拓扑结构来表达。该算法的中心思想是，节点在结构孔中的约束越多，节点的影响就越小[2]。结构孔算法的优点是重要的连接节点可以被精准地识别出来，缺点是该算法识别聚类中心的能力较弱。

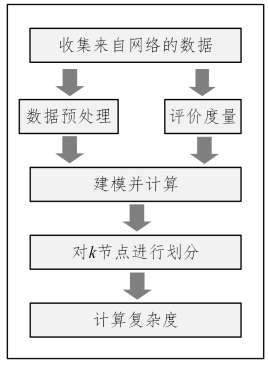

图2 影响力节点提取流程

Fig. 2 Extraction process of influential nodes

1.2 大数据分析

1.2.1 大数据定义

若要分析大数据，首先需将大数据的定义描述清晰，目前在大数据方面人们达成共识的是大数据的定义应该从它的特征入手[10-13]。其中一种描述方法是将容量、多样性和速度作为大数据管理的3个挑战[14]。这种描述方法被称为“3个V”，是最常用的大数据概念框架。

其中，数据流的大小被归为容量。通常来说，大数据的大小经常为兆字节和千兆字节。随着信息时代的到来，数据量呈现爆炸式增长，因此对大数据的量的定义是相对的，会根据时间和数据类型的不同而做出不同的定义[15]。或许，今天被定义为大数据的量在未来的信息中根本达不到大数据的阈值，因为在未来，存储容量会随着时代发展而增加，在这种条件下人们可以获取比现在更大的数据集。多样性讨论则是大数据中的数据类型，数据类型同样也定义了大数据中“大”的含义。两个一样大小的数据集因为类型不同可能需要不同的数据管理技术，就像表格数据与视频数据，因此不同的行业同样可以定义大数据的多样性[16]。从结构来看，多样性代表数据集中结构的异质性。随着技术的不断成熟，各大公司逐步开始使用各种类型的结构化、半结构化和非结构化数据[17]。

大数据中的速度指生成数据的速度以及分析和处理数据的速度。信息时代人们对智能移动设备的需求大大增加，信息流的增加速度也越来越快，这也同时使大数据实时分析和规划技术更加热门，甚至一些传统行业也在快速生成需要分析的实时数据，比如今天的零售业[18]。人们每天从智能手机

这类移动设备发出大量流经移动程序的数据，这些数据背后涵盖了大量的信息，用于分析之后可以针对每个用户提供个性化的服务[19]。因为这些数据涵盖了用户有价值的信息，如地理位置、购买偏好以及模式，有效利用这些信息可以提升用户服务，以在经济市场中创造更大的价值[20]。

本节的大数据处理包括文本、音频和视频的处理。网络中这些内容种类丰富且信息量巨大，给社交影响力分析带来了很大的挑战，但同时也提供了很多机会。纵观信息时代的发展，首先是社交网络的出现和发展创造了社交网络中的大数据，庞大的数据流又加速了社交影响力分析、机器学习和云计算等大数据技术的发展[21]。为了应对庞大的数据流，各社交平台安排了各自不同的更新方案，如图 3 所示。

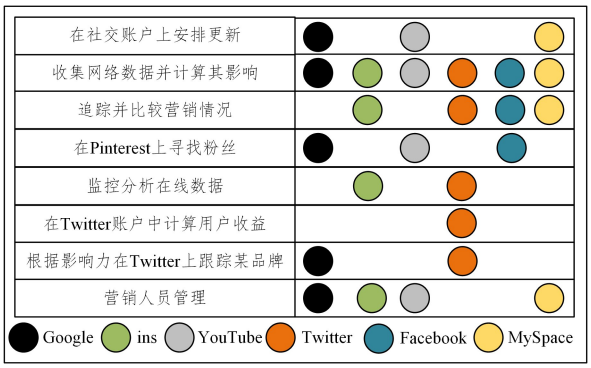

图 3　社交平台上的更新安排

Fig. 3　Updating schedule on social platforms

1.2.2　大数据中的文本分析

文本分析也被称为文本挖掘[21]。顾名思义，文本分析就是从网络中的文本数据中提取需要的信息。文本数据作为基础的数据类型，种类十分丰富，如被订阅的社交网络、电子邮件、博客、各种分类的论坛、调查的回复、公司保存的文档、文本版的新闻和呼叫中心日志等。为了通过文本分析得到更重要的数据，现在人们常用到统计分析的知识、计算语言学和机器学习等[22-25]。文本分析的意义在于省去了大量的人工，可以直接将大数据中的文本数据转换为人们所需要的简练的摘要，从而辅助决定基于大数据的各种决策，这种方法在企业中的收效甚好。例如，现在的技术支持从金融新闻的文本信息中提取人们关心的信息，可以用来预测相关的股票市场。文本分析方法主要依赖信息抽取技术，即从非结构化文本中抽取结构化的数据[6]。例如，利用算法人们可以从非结构化的医疗处方中提取需要的结构化信息，如药物名称、剂量和频率都是经常被提取的；时间宝贵的急诊室可以在文本信息中直接查找患者姓名，并将其分类为预定义的类别，如人员、日期、地点和组织[26]。

从文本信息中提取的文本摘要是根据已有信息自动生成一个或多个文档的简明摘要，生成的文档简要地传达了原始数据中的关键信息。文本摘要主要应用在各类科学文章、新闻文章以及博客。从意义上讲，简要概括了文本遵循的两种方法：提取方法和抽象方法[27]。提取方法指生成的摘要都是提取原始文本数据中的单元，文本数据单元通常指原始数据中的句子，生成的摘要是原始文档的子集。具体方法是首先确定一个本文的单元，并将不同的单元进行连接，在这个过程中，对文本单元重要性的判断尤为重要，目前人们都是通过记录并分析它们在文本中的位置和频率来评估的[23]。这种提取方法不要求算法对文本数据中的信息进行“理解”，而另一种抽象方法则要求从文本中提取语义信息，也就是抽象摘要包含不一定出现在原始文本中的文本单元。为了更好地解析原始文本并生成抽象摘要，抽象方法对自然语言处理技术有一定的要求，因此抽象系统生成的摘要比提取系统更加连贯。由于大数据的信息量十分庞大，提取系统耗时更少，计算量更小，因此更容易被采用[24]。

1.2.3　基于大数据的预测

预测分析就是基于历史以及目前会更新的数据预测未来可能出现的结果的技术。这项技术的实用性极高，几乎可以应用于所有学科——从基于卫星传回数据预测天气，到根据顾客的消费数据推荐他们可能心仪的产品，这其中的核心技术就是找到数据更新的模式以及数据之间的关系。

预测分析技术大体上也可以分为两种：一种是尝试从发现结果的模式来推测未来数据的技术，如移动平均线；另一种则是尝试从变量入手，解释变量之间的相互依赖关系从而预测结果，如线性回归[28]。

从技术上也可以分为两组：一种是基于基础方法的回归技术（如多项逻辑模型）和机器学习技术（如神经网络）；另一种是基于结果变量的类型，如线性回归等技术，用于处理连续的结果变量（如房屋销售价格），而随机森林等其他技术适用于离散的结果变量（如肿瘤良恶性分析）。预测分析技术的基础就是统计方法，在传统的统计方法中，其表现出了很强的统计显著性：从总体中提取一个样本并将它与结果进行比较，以验证特定关系的显著性，将得到的结论推广到整个人口[29]。但是处理大数据时，因为数据量过于巨大，几乎可以认为代表了绝大多数人口，所以传统的显著性的概念与大数据并不相关。在计算效率方面，传统统计方法中提取样本的方法并不适用于大数据。大数据固有的显著特征是：异质性、噪声累积、虚假相关性和偶然内生性。特性的不同和方法的不适用也导致大数据推动了新统计方法的开发[29]。

1.3　进化网络、影响力分析以及大数据之间的关系

由于信息时代的到来，数据量的增长速度达到了巅峰，数据量越来越大，数据的类型也越来越多[30]。面对现在的大数据集，人们需要处理越来越多的动态信息，信息维数也越来越高，这就需要用到结合大数据的云端信息处理技术，例如创建多属性的大数据聚类挖掘模型，来提高数据分析的准确度。这项技术对基于大数据的进化网络有着极其重大的意义。在如今的应用中，人们通常采用网络爬虫进行数据收集，利用搭建好的挖掘模型进行数据分析[31]。

在如今信息爆炸的时代，人们的日常生活完全离不开大数据，社交网络大数据时代已经来临。而在社交影响分析中，社交网络仍旧占据主要地位，大数据在社交网络中的应用为社交影响力分析带来了机遇和挑战。社交网络结合信息技术和社交网络理论，面对每天产生的巨大信息量，提出了不同的方法以应对各种应用领域的挑战[32]，如病毒营销、预测股市、部分领域的专家分析、传染病模型以及关键节点识别，即有影响力的用户识别。

社交网络中提到的有影响力指社交影响，映射到现实中就是社会中的个体在他人的影响下改变自己的行为或决定。社交网络影响力的因素有很多，如彼此之间的关系、现实或者网络距离、时间效应、网络和个人的特征等。病毒式营销、实时广告、产品推荐和很多种社交应用程序可以通过数据得出对一个人或一群人定性和定量的影响力评估。这项技术可以最大程度上利用社交影响，进而从中受益。目前大量研究人员在基于社交网络大数据解决现实世界的问题这一方向上努力[33]，以求得到效率更高的分析方案。在研究人员的努力和大规模社交网络的发展势头下，大量新的数据分析工具和算法促进了可扩展、可访问和可持续的数据基础设施的发展，使人类能更适应如今的信息社会[34]。

而社交网络与进化网络之间的关系，可以从定义出发进行分析。社交网络指通过网络将人与人连接起来，形成一个具有某一特点的团体，而进化网络可以被定义为一种随时间而变化的网络[35]，可以将进化网络视为社交网络的一种自然延伸，进化网络的概念建立在既有的网络理论之上。几乎所有的社交网络都随时间而演化，也就是社交网络中的节点或者节点之间的连接增减依赖于用户结交或者失去朋友，即一些人成为新的社交网络的一员而另一些人离开了原来的社交网络[36]。除此之外，节点的影响力权重值也在随时间变化，而这些过程几乎都是同时发生的。

基于大数据的进化网络的应用远不止如此，例如 Google 著名的 Alphago 围棋人工智能项目打败了世界上许多著名的围棋高手，一时引起了各方的关注。在游戏中，人们同样可以扩展它的应用[37]，例如通过捕捉并挖掘游戏训练中的规则步骤和大部分玩家的操作，自己构建一种基于进化网络的数据分析训练策略，能够对游戏选手起到很好的辅助效果。

2 基于大数据的进化网络影响力最大化问题

2.1 影响力最大化问题

影响力最大化(Influence Maximization，IM)首先被建模为一种算法问题，该算法由 Kempe 等提出并进行了相关研究[1]。2003 年以后这个问题被引入网络进行处理，常表示为图 $G=(V,E)$的社交网络。其中，V 是 G 中的节点集合(即用户)，E 是集合 G 中的(有向/无向)边的边界(即用户之间的社交连接)。IM 问题的目标是在 G 中找到数目为 k 的节点集合，使其具有最大影响力[38]。

任何种子集产生的影响都是基于信息在用户之间扩散传播。市场营销方式就是以信息传播为基础，有效的营销和广告投放方式是影响力最大化和数据挖掘最主要的应用。例如，一个公司希望在市场上推销自己的一款新产品，在产品投放之前，产品经理可以通过受众网络的社交连接找到具有影响力的用户，并将其作为初级用户进行第一轮营销，从而在之后达到提高产品营业额的目的。在此之前，市场营销一直以直接营销为主，在这种传统的大众营销中，新推出的一种产品被推销给所有的顾客。而在新型的网络营销中，找到最有潜力和利润的顾客成为了初始营销的首要目的[6]。虽然直销是增加利润的一个很好的工具，但它只针对个人，而忽略了用户之间的关系。针对此类问题，为了量化信息扩散，文献[39]正式定义了扩散模型以及在该模型下的影响扩散。

进化网路中基于大数据的影响力最大化面临的最大问题就是，在数据集过于庞大的背景下，计算的复杂度直线上升，虽然中心度判断节点影响力的方法计算较为简单，但是因为对多聚类情况的识别不够精准，在大数据中人们通常不采用这种方法。针对这个问题，不少研究人员对现有的贪婪算法进行了不同的改进，从而求得扩散模型的最优解，尽管该算法的精度较高，但是相应的计算量极大，用于实际应用中同样存在困难[40]。

2.2 扩散模型

随着社交网络的引入，信息扩散的话题引起了研究界的兴趣。信息扩散模型描述了信息和数据是如何在社交网络中传播和传输的。信息传播不仅限于社交网络，还作用于人群。多年来，人们一直在研究相关主题，以更好地理解疾病、谣言等是如何作用于人群并在人群中传播的，或者受污染的水会如何影响供水管网。现阶段，已有大量关于设计的文献，如数据挖掘、数据库和流行病学，对扩散模型进行了阐述。本文首先将简单回顾 IM 的常用模型。

首先，研究人员提出了一个通用的扩散框架，该框架中每个节点($u\in V$)的状态为非活动或活动。最初，它会查看各个点的状态，选取作为种子集($S\subseteq V$)的一组用户并将其激活[41]，此时 V 中的其他节点处于非活动状态。然后，基于社交图 G 考虑 S 中的种子影响其他节点的扩散过程，S 中尝试激活其邻居节点，新激活的节点可以进一步激活它们的邻居，以此类推，直到这种扩散中没有新的节点被激活时，扩散进程终止。具体来说，这种框架对前面提到的“激活过程”进行了建模[42]。

2.2.1 独立级联模型

该模型的基本假设是，节点 u 试图激活其邻接节点 v 的行为能否成功是一个概率为 $p(u,v)$的事件，且一个处于非活跃状态的节点被刚进入活跃状态的邻居节点激活的概率独立于之前曾尝试激活该节点的邻居的活动。此外，该模型还做出了这样的假设：网络中任意的节点 u 只有一次机会尝试激活其邻居节点 v，无论成功与否，在以后的时刻中，u 本身虽然仍保持活跃状态，但它已经不再具备影响力，这一类节点被称为无影响力的活跃节点。算法步骤如下：

(1)初始化活跃节点集合 A。在 t 时刻，新被激活的节点 u 对它的邻接节点 v 产生影响，影响成功的概率为 $p(u,v)$。若 v 有多个邻居节点都是新被激活的节点，那么这些节点将以任意顺序尝试激活节点 v。

(2)如果节点 v 被激活成功，则在 $t+1$ 时刻，节点 v 转为活跃状态，将对其邻接非活跃节点产生影响；否则节点 v 在 $t+1$ 时刻状态不发生变化。

(3)不断重复该过程，直到网络中不存在有影响力的活跃节点时，传播过程结束[43]。

2.2.2 线性阈值模型

在线性阈值模型中，每条有向边$(u,v)\in E$ 上都有一个权重 $w(u,v)\in[0,1]$，直观反映了节点 u 在节点 v 的所有入度邻居中影响力的重要性占比。每个节点 v 还有一个被影响阈值$\theta_v\in[0,1]$，这个阈值在 0 到 1 的范围内均匀、随机地选取，

其一旦确定在传播过程中就不再改变[47]。与独立级联模型类似，算法步骤如下：

(1)在 $t=0$ 时刻有且仅有种子集合 S_0 中的节点被激活。

(2)之后每个时刻 $t\geqslant 1$，每个不活跃节点都需要依据它所有已激活的入邻居得到它的线性加权和是否已达到它的被影响值来判断自己是否被激活；若是，则节点 v 在时刻 t 被激活($v\in S_t$)；否则节点 v 仍然保持不活跃状态。

(3)当某一时刻不再有新的节点被激活时，传播过程结束[45]。

2.2.3　触发模型

触发模型是独立级联模型和线性阈值模型的推广，其给定任意用户 v，模型定义了一个分布，该分布给 v 设定一个概率，该概率表示邻接子集影响 v 的概率[46]。触发模型的主要思想在于：对于给定的节点 v，分配触发节点集 T_v 代替线性阈值模型中的阈值，即如果 T_v 中有节点在步骤 $t-1$ 被激活，则 v 在步骤 t 中被激活。与独立级联模型和线性阈值模型一样，触发模型的整个传播过程也是在时域上离散的。传播过程的描述如下：

(1)首先为网络图 G 中的每个节点随机选择一个触发节点集 T_v，T_v 是 v 的所有邻居的集合中的一个子集，然后选择一个已激活节点的集合作为初始的种子集；

(2)步骤 t 中，对于还未被激活的节点 v 而言，只要 T_v 中有一个节点 u 在步骤 $t-1$ 中被激活，则 v 被激活；

(3)重复寻找未被激活的节点直到没有新节点能被激活，此时整个传播过程结束[47]。

2.2.4　时间感知扩散模型

独立级联模型、线性阈值模型和触发模型都是没有引入时间变量的模型，其扩散过程无法在不能激活更多节点时及时终止。然而，现实中的传播活动往往需要考虑时间变量，要求在有限的时间内最大限度地扩大影响力[48]。为了满足这种需求，时间感知模型应运而生。现有的研究大致可以分为两类：离散时间模型和连续时间模型。

离散时间模型将独立级联模型扩展成模拟从一个节点到另一个节点的扩散过程。将其作为不同时间步长上的离散随机变量[49]。然而，这些模型在本质上和独立级联模型是一样的，原因是其扩散只发生在离散的步骤中。

在现实中，一个节点影响另一个节点的过程是连续的，因此相关研究引入了连续时间扩散的模型，而此时需假设节点间的成对传播是时间的连续分布。

2.3　影响最大化算法

现阶段，研究社交网络中基于贪心算法的影响力最大化问题已经十分成熟，然而动态演化的社交网络中的影响力最大化问题的研究还处于初始阶段。进化网络中面对节点或边的添加或删除等问题，针对影响力最大化算法如何适应动态变化的网络，存在以下 3 个主要问题：1)节点、边的添加或删除时节点的影响力是否发生变化？若发生变化，如何对节点的影响力进行更新？2)种子集合中节点是否受到影响。受到影响时如何选取新的种子节点。3)进化网络中网络的结构随时间不断地变化，保证算法的时效性以及有效性十分有必要，考虑如何降低算法的时间和空间损耗。本节从基于社交网络的影响力最大化算法的特性来探讨其是否适用于进化网络。

2.3.1　贪婪算法

研究者提出的原始贪婪方法为图中影响最大化问题提供了一个可靠的解决方案，从离散优化的角度提出如何选择有影响力的种子问题，并证明了该问题是 NP-hard 问题，同时证明了传播影响函数的次模性，并提出了一种近似比为 $(1-1/e-\varepsilon)$ 的贪婪算法，其中 ε 是与社交图 G 和 r 相关的常数，r 是计算 $E[I(S)]$ 估计值时的测量次数。该算法基于独立级联模型和线性阈值模型，从初始化 S 为空集开始，通过种子集 S 筛选传播影响范围内期望 $E[I(S)]$ 增长最大的节点，直到种子集 S 中包含 k 个节点，此时算法结束[50]。

该算法为影响力最大化问题提供了近似比为 $\left(1-\frac{1}{e}\right)$ 的近似最优解。每次选择一个使种子节点集合影响力扩展度增量最大的节点加入种子节点集合，直到种子节点的数量达到预算值为止。另外，在具体的计算中，每次选择新节点都需要精确计算它的影响力扩展度。然而，存在有向无环图的独立级联模型和线性阈值模型的影响力扩展度计算都比较复杂，因此可以使用蒙特卡洛模拟来近似计算影响力扩展度。此蒙特卡洛贪心算法为影响力最大化问题提供了近似比为 $(1-1/e-\varepsilon)$ 的近似最优解。

(1)首先，$|S|=1$，在所有的点中选择一个点在模型中运行，将感染数量最多的节点加入 S，此时运行了 n 次模型；

(2)在剩下的点中选一个加入 S 后结果最好的点加入 S，此时运行了 $n-1$ 次模型；

(3)重复操作(1)和操作(2)，直到 S 扩散的结果能够覆盖所有的点[51]。

随着网络规模的扩大，贪婪算法的时间和空间的消耗呈指数倍增加，为增加算法的有效性以及可扩展性，研究人员在原始贪婪算法的基础上进行优化，常见的贪心的变体算法如表 1 所列。

表 1　常见贪心算法概述

Table 1　Overview of common greedy algorithms

算法	原理	参考文献
Basic Greedy	将问题转化为组合优化问题，对含 k 个影响力用户的种子集 S 进行 k 步迭代，并在每次迭代中使用扩散模型来估计该集合的影响力	文献[1]
CELF	使当前迭代中的影响力传播增益不能大于其在前几次迭代中的边际增益，大幅度减少影响估计函数的评估次数，使运行时间显著缩短	文献[52]
CELF++	利用社会影响函数的子模块性提出了一个优化版的 CELF，避免迭代中的重复计算，CELF++ 的渐近复杂度保持不变，但其速度比 CELF 快 35%～55%	文献[53]
SKIM	该算法改进了基本贪婪算法，确保在每次迭代中，选择添加到种子集的节点具有足够高的概率或预期的边际增益，接近于最大值	文献[54]

贪心算法的核心思想是每次迭代选取最具有影响力的节点。社交网络中每个节点都参与了影响力的评估，研究者们通过数学方法来不断减少算法的时间和空间损耗。贪心算法在每次迭代过程中会重新计算全部节点或部分节点的影响力，当社交网络中的节点或边被删除或添加时，计算新节点的影响力和更新部分节点的影响力就能够重新进行种子节点的

选取。基于上述内容，社交网络中原始的贪心算法以及CELF、CELF++、SKIM等贪心算法的变体都能很好地适用于进化网络，同时也不会过多地增加时间和空间的损耗。

2.3.2 启发式算法

贪婪算法及其优化算法虽然能一定程度地提高算法的效率，但在实际应用中算法的时间复杂度仍偏高。启发式算法没有给出影响扩散的任何最坏情况的边界，但是，与贪心算法相比，大多数算法具有更高的可扩展性和更短的运行时间。基于中心的启发式算法只考虑网络的拓扑结构，因此在大多数情况下所获得的传播影响相比其他最先进的方法要小得多。然而，度折扣启发式算法的表现比其他中心性启发式算法稍好，因为它对两个相邻节点的选择有一些限制。如前文所述，虽然启发式方法相比贪婪算法更加复杂，但它们有一个优势，即这类解决方案是高度可扩展的，可以很容易地应用于非常大的图形。

表2列出了一些经典的启发式方法，这些方法已经用于在影响力最大化问题中寻找种子节点集。启发式算法大多根据节点在社交网络中的拓扑信息选取种子集的节点，当社交网络中的节点或边发生变化时，网络中节点的拓扑信息也随之发生变化。面对不断演化的社交网络，如何选取合适的种子集合是关键问题，因此在选取种子集合时可以通过维护一个候选集合来应对不断变化的社交网络，当网络结构发生变化时，种子集合可以从候选集合中对种子集合进行更新，同时，相比全部节点的更新，维护种子候选集合所消耗的时间和空间更少。基于以上所述，社交网络中的启发式算法能够很好地适用于进化网络。

表2 常见启发式算法概述

Table 2 Overview of common heuristic algorithms

算法	原理	参考文献
Degree centrality	节点拥有的邻居越多，其影响力就越大，k 个度最大的节点作为种子集	文献[55]
Degree discount	算法迭代 k 次，每次迭代选择次数最高的节点加入种子集，忽略了节点与其他网络节点之间的边	文献[56]
Degree punishment	最有影响力的节点被包含在种子集中，其邻居和二阶邻居被惩罚，减小了被选为种子的概率	文献[55]
Heuristic clustering	计算节点之间的相似性，将节点聚类为 k 个簇，并将每个簇的中心作为影响节点	文献[57]

2.3.3 元启发式算法

许多元启发式优化算法是基于进化计算技术发展起来的，并没有给出影响扩散的最坏情况边界。一些算法根据应用背景定义了相应的适应函数，将问题转化为优化问题求解近似解的方法，大大降低了时间复杂度。群智能优化算法被广泛地应用于影响力最大化问题的求解过程，如遗传算法、模拟退火算法、粒子群算法等。相比之前的启发式算法，元启发式算法的传播性能在得到提高的同时，运行时间也大大缩短。

Bucur等[58]用遗传算法解决了SIM问题，他们证明，用一个简单的遗传算子有可能在可行的运行时间内找到影响扩散的近似解。在大多数情况下，他们的方法得到的影响扩散与Kempe等[1]提出的基本贪心算法相当。Jiang等[59]提出了基于模拟退火的算法来解决IC模型下的SIM问题。报道的结果表明，他们提出的方法比文献中现有的启发式方法快2～3倍。Tsai等[60]将遗传算法与Chen等[56]提出的新贪心算法相结合，开发了求解IC模型下SIM问题的遗传新贪心算法。他们报告的结果表明，GNA的影响力传播比遗传算法多10%。Gong等[61]提出了求解SIM问题的离散粒子群优化算法。他们使用Chen等[56]提出的程度折扣启发式方法来初始化种子集和局部影响估计函数，以近似两跳的影响。他们还引入了网络特定的局部搜索策略，以快速收敛他们提出的算法。报告的结果显示，这种方法优于最先进的CELF++，计算时间更短。

2.3.4 其他算法

(1)单周期影响最大化

单周期影响最大化是一种新方法，其基于Kempe等提出的独立级联法，是原算法的一种更简单但精度较低的版本。其使用BSP范例来实现，BSP(Bulk Synchronous Parallelism)模型又被称为块同步并行模型，该模型是由哈佛大学的Leslie Valiant提出的一种基于超级步和全局“栅障”同步的并行模型，对高性能领域的发展起到了不可估量的作用。在BSP模型中，整个计算过程是由一系列使用全局同步分开的周期为 L 的计算部分组成。它通过对所有节点模拟一次独立级联过程，在一次次的迭代中寻找影响范围最大的节点。每个节点都包含它所影响的节点数量的信息，以及它所影响的节点列表[62]。

当所有节点的状态都变为非活动状态或者达到节点数的限制时，程序结束。每个节点在第一个超级台阶向它的所有邻居发送自己的ID(身份信息)，因为每个节点必须知道是谁激活了它，所以必须获取ID而不是布尔值，之后存储它们的ID，并向它们发送一条消息[63]。在第一次迭代中，节点激活它们的邻居的过程都将在受影响的节点上进行。在过程结束时，每个节点都有一个值来指示其预期的扩散，顶部的 k 个节点被报告为最佳初始种子。虽然这种方法很快，但也有一些缺点：1)新节点的激活仅通过一个基于边权重的随机过程来完成；2)所有节点都独立于其他节点执行激活的过程，而根据网络结构，这可能会对最终结果产生一定的影响，由于每个节点都发送自己的ID和受其影响的节点的ID，因此可能会产生网络流量。缺点1)表明，不同运行之间的结果可能不同，因为人们只运行一次模拟。在最初的贪婪算法中，每次进行迭代时，一个节点被添加到当前的初始集合中，并且扩展被重新计算，然后选择下一个添加扩展最多的节点。但是，这里所有的节点都在孤岛中运行，但忽略了网络效应。例如，其中两个相邻节点在所有节点中分别具有最高的分布，算法会在最终结果中报告这两个节点，但实际上选择两个相邻节点并不是一种有效的方法，因为受一个节点影响的节点通常也会受到其他邻居的影响。关于网络流量的后一点，在特殊情况下效率会相当低，甚至会导致整个集群瘫痪。然而，在真正的大图处理中几乎不存在这种情况，因为首先大多数大图非常稀疏，传播概率非常小，以至于节点ID几乎不需要移动超过几个节点。然而，较大图形的测试揭示了这种方法存在另一个缺陷，即由于节点必须在内存中存储激活它们的其他节点的列表，因此在大图中，这通常会导致程序因内存已满而崩溃。

(2)基于特征令牌的实现

基于特征令牌的实现是一种基于单循环算法的新方法，其考虑了独立级联模型的模拟方面。在该算法中，每个节点都将尝试多次，而不是只尝试激活它的邻居一次，节点将保留许多具有唯一 ID 的标记，每个标记都可以追溯到节点本身。在每个步骤中，一个节点将尝试用每个令牌激活它的邻居一次，如果使用令牌激活了一个节点，则它将通知原始源节点，以便该节点可以为该特定令牌增加其内部计数器。一旦没有新节点被激活，程序就结束，所有节点将报告令牌计数器的平均值作为它们的预期分布，然后选择前 k 个节点作为初始种子，每个节点的标记数量决定了算法的准确性。

因此，选择最佳的令牌数量就是在更好的解决方案和更快的算法之间进行权衡。与以前的方法类似，根据图的结构，由于节点之间的通信量大，这种方法可能会产生大量的网络开发成本。此外，它还需要更大的存储空间，因为每个节点都必须为许多不同的模拟分布内存，所以选择模拟的次数越大不仅会减慢过程的执行速度，还会增加系统崩溃的风险。与单循环方法类似，基于令牌的方法也不能在非常大的图上进行操作[64]。

(3)边缘采样

边缘采样和独立级联过程的处理是不同的，但这种方法仍然使用一个或几个蒙特卡洛模拟，因此理论上它不如贪婪方法精确，但可以获得更好的结果，影响到更多的节点，因为过程非常快，这意味着任何现有的边缘都将传播消息。这里，人们可以将报告的顶部的 k 个节点作为解。而盲目选择具有最高估计扩散的节点的一个缺点是，相邻或处于相同连接组件中的节点对最终扩散的贡献不大，因为在初始种子中同时拥有这两个节点是多余的。在这种方法中，由于人们知道每个节点属于哪个连通分量，因此人们可以报告来自前 k 个连通分量的一个节点，即其中节点数最多的连通分量，而不是报告前 k 个节点。该算法还可以通过使用多个采样子图并平均来自所有样本的最佳节点来降低随机性的噪声。与其他方法一样，必须进行一些修改，以使该方法更适合并行框架。使用 BSP 模型可以高效地并行计算连接的组件，它已经作为 GraphX 库的一部分得到了实现。为了进一步优化算法，对 GraphX 的连通分量算法进行了修改。由于对于每个节点，人们只需要它对应的连通分量的大小，因此可以修改原来的算法来代替报告连通分量 ID，它将报告连通分量的大小。新的连接组件算法的结果将是节点列表，每个节点携带其 ID 和相应的连接组件大小。算法从初始化所有节点的列表开始，开始时，每个节点的预期扩散为 0。在模拟的每次迭代中，尺寸会根据连接的元件尺寸进行更新，更新可以简单地通过连接来完成。节点连接在 GraphX 中非常有效，并且可以非常快速地完成。增加模拟次数会提高算法精度，但也会减慢过程[65]。

2.4　基于大数据的算法优化

2.4.1　数据准备

为了对大数据的信息进行实时分析，人们需要具有更低计算复杂度的相关算法。

首先，为了使相关算法能够更好地进行计算，人们需要对数据进行一些初步的处理。处理的首要任务就是确定需要在图的哪个部分对节点进行排序，因为相关节点的局部排名并不等价于整个图的排名，所以排序既可以在整个图形中进行，也可以只在一组节点中进行。在处理具体问题时，人们需要知道 N 中的哪 m 个节点可以吸引更多的 N(节点集)节点到 $N+$(分类为正的节点集)[66]。为了减少操作次数，使运算变快，本地排名将在 N 端进行。为了得到 N，首先需要计算最小切割。在获得 N 的子集之后，人们才能对各个节点进行排序。在节点排名阶段，人们要选取可以吸引更多节点的节点，而这个特征一定与节点本身的性质有关，因此节点排名可以为数据处理做准备，以此获得的排名最高的 m 个节点可以是额外的种子节点。为了减少运算次数，熵排序只在 N 个子图中计算，而排序的复杂度与图中的节点数直接相关。

2.4.2　图形切割

人们需要对图形进行切割。在切割阶段，使用最小切割将图形分成两组：$N+$ 和 N。因为最小切割是在本地计算的，所以需要建立源节点和宿节点。这里本地的最小切割指获得这些节点之间的局部最小切割，而不是全局最小切割。切割后对节点进行分类，每个节点 $I\in N+$ 分类为正，每个节点 $j\in N$ 被分类为负。当被吸引的节点是 $N+$ 边的成员时，其也包括在正节点集中，因此被吸引的节点不添加到正节点集[67]。

在这种切割算法中有一个特殊的操作，它将 $N+$ 侧的所有节点组合在一个节点中。超节点是图论中减少运算次数的一种算法。在这种算法中，由一条边连接到 N 个节点的 $N+$ 中的节点被称为切割节点。该技术将正割节点之间的所有边融合为负切割节点，并将它们融合为同一条边，即所有割节点的权重之和。对于链接到正切割节点的负切割节点也是如此。$N+$ 节点之间的所有权重都被忽略，因为分析 $N+$ 交互没有意义。并且，超级节点永远是源节点。算法时间复杂度由最小切割系数组成。将排名最高的节点添加到 $N+$ 侧后，排名第二的节点也会被吸引，因此不需要将这个节点作为额外的种子。

2.4.3　优化结果

此时人们引入了贪婪算法，虽然贪婪算法可能会导致人们花费更多的时间来获取结果，甚至还可能导致低质量的结果，但人们还是选择利用贪婪方法，目的是将即时消息的组合部分简化为一个优化问题[68]。因此，人们可以修剪组合的数量来计算高质量的种子。假设 $m+m'$ 分别为排名靠前的节点和测试出的能吸引更多节点的节点，其中 m' 是通过一个额外的搜索空间而寻找到的最好的种子，m' 的值可以通过实验测试来确定[46]。

3　基于大数据的进化网络影响力最大化研究的机遇与挑战

3.1　机遇

现有的商业宣传大多通过社交媒体进行营销，利用有影响力的客户和创建社区来增加产品的曝光率。社交媒体的

快速发展导致客户以指数级的速度增长，而将组织活动与社交媒体上的竞争对手进行比较是至关重要的，这样可以采取必要的措施来更好地推广他们的产品[69]。同时，进化网络中用户的增加意味着网络中每时每刻都在产生社交数据，而如何从海量的社交数据中获取有价值的信息同样至关重要。基于大数据的社交影响力分析的优势是，可以让人们对当前趋势有更清晰的认识，在竞争时能有清晰的定位。

进化社交网络的另一优势在于，部分平台提供了脱敏的大型公开社交网络数据集，能够用于各种研究项目，包括文本分类和图形分析等。同时，这些社交网络数据集中包含了大量的社交数据以及信息。社交影响力分析的目的是从这些海量的数据中挖掘出有价值的信息，并更好地利用这些信息去进行下一阶段的工作。有人专门从各种平台或渠道收集数据，从数据中检索有用信息并对获得的数据进行分析处理。社交媒体公司试图从这些网站获取数据，并将其出售给其他分析公司[70]。Facebook 已经开始收集用户接受新条款的数据，包括个人数据、照片、位置、智能手机旅行位置以及 Facebook 服务器上的类似兴趣，目的是向公司提供特定事物的个人兴趣，从而达到营销或推广最大化，增加公司的盈利。

社交网络因其强大的传染性，已经成为目前信息传播的主要渠道，如产品推广、谣言抑制、新闻发布、影响扩散等。从产品推广的角度看，社交网络的用户分享关于产品的帖子或推文，一定程度上能够引导其朋友或邻居访问产品的网站并购买它，提供了一种投资回报的途径。社交网络数据分析对于如何留住客户也至关重要，尤其在目前市场上竞争对手不断增长的背景下，留住客户更是一项挑战，也是机遇。对社交数据的敏锐分析是目前研究中的一大热点，既要保证社交影响分析的准确性，又要保证用户的隐私。基于对用户社交数据的分析，为用户提供个性化的推荐，特定的产品被推荐到它们的适用人群，在提高销量的同时大大提高了用户的满意度，部分满意的用户会把产品推荐给别人，逐步增加产品的销量。通过社交媒体提高客户的口碑对企业来说是非常有益的，对客户来说也是一个令人满意的选择。但是，进化网络中对用户的分析极大程度地依赖于用户个人信息以及行为信息，使得进化网络中用户的身份信息更容易造成隐私泄露，从而导致各种不当行为中的身份滥用[68]。

证券方面也紧跟时代的步伐，目前已经有很多公司意识到了使用大数据分析来代替之前数据挖掘的重要性，利用大数据进行分析以预测证券市场走向这项技术目前可以用于研究市场走向与社会经济新闻之间的关系、市场中人际关系的流动、投资的回报和风险、股票的走向预测等。然而，该领域的研究成果处于刚起步的阶段[6]，现在还很难真正应用到市场中，因此大数据分析在该领域的表现还有很大的提升空间。

3.2 研究挑战

如何平衡好算法的准确性和数据集大小之间的联系，是大数据时代接下来面临的一大挑战。由于云计算和物联网使用的增加以及社交网络平台的整合，网络数据集越来越丰富，但用于构建社交数据以从中推断知识的工具很少，当前研究人员缺少用于理解分析大数据的可用的工具[41]。此外，相比过去效率不高的推销方法，现在需要更加精准地进行推荐。目前利用决策树已经可以支持在众人的数据中挖掘出可用的信息，并基于喜好为他们提供更合适的产品或者对他们的行为进行预测。但当面对信息时代庞大的数据集时，决策树算法也有其局限性，即数据集过大从而导致类别过多时，准确性会有相应的降低。为了解决这个问题，不少研究人员提出了一些解决方案[71]，如用改进的启发式方法构造二元决策树等来提高准确性。

进化网络中种子集合的准确性和计算时间之间的平衡是目前面临的另一大挑战。影响最大化问题被证明是 NP-hard 的，这意味着无法在可行的运行时间内得到最优的种子集合。在这种情况下，直观的方法可能是使用一些启发式方法来选择种子节点，选取近似种子集以减少时间开销。但是，种子节点生成的受影响节点的数量也可以任意减少。在这种情况下，如何设计算法就成为了一个重要的问题，算法要在可承受的时间内运行，并且算法所选择的种子集与最优传播之间的差距要尽可能小。

影响最大化算法的扩展性也是当下不得不考虑的问题。一般情况下，影响最大化问题需要做很多假设，如每一个选择的种子在传播过程中都会达到预期的效果、对网络中每个节点的影响同等重要等。在某些情况下，这种假设可能是不现实的。假设在目标广告的情况下，不是选择所有节点，而是选择一组目标节点，目的是使目标节点内的影响最大化。另一方面，由于扩散的概率性，种子节点在影响扩散过程中可能不能达到预期的效果，有些影响最大化算法适用于特定的情景，无法迁移，大大降低了算法的可用性。

为了进一步提高社交大数据的研究成果，需要考虑更多的方法，包括学习统计分析、机器学习技术，如关联规则、分类、聚类、时间序列分析和回归分析，以确定社交和行为的变化。在研究过程中需要利用社交网络中节点、研究平台、第三方系统和数据库之间的复杂关系来理解数据，还需要应对来自开放空间网络的地理空间和时空数据带来的挑战，因为它们以不同的格式表示。用户在地球上的位置可以从地理空间数据中推断出来，并且全球定位系统数据可以从支持全球定位系统的无处不在的设备中获得。嵌入式传感器通过时间戳和空间位置连接物理世界和数字世界。分析时空数据[72]提供了用户在特定时间间隔的特定模式，这将有助于确定他们未来的行动和异常行为。

结束语 随着进化网络问题的提出，影响力最大化问题引起了学术界的广泛关注。这是一个寻找一组节点的问题，这些节点可以最大化信息在整个网络中的传播。本文旨在研究影响力最大化，特别是其在市场营销中的应用以及目前的解决方案。本文主要讨论了一些传统的模型、算法以及改进后的算法。

目前提出的大多数贪婪解决方案即使在节点数目较小的图上也需要大量的时间和资源，贪婪算法存在可伸缩性问题，因此贪婪算法很难在大数据平台上实现。启发式算法只需要图中的局部信息，因此程序可以对所有节点并行运行。本文介绍的边缘采样方法是贪婪算法和启发式算法的结合。

虽然它需要一些迭代来获得更好的解决方案，但它是可并行的，可以应用于规模较大和较小的图形。目前提出的解决方案并不适用于所有的应用环境且在不同的图中结果不一致，一个合理方案是在所需图形上测试所有方法，并选择提供最佳解决方案。

此外，通过贪婪方法寻找确定性函数（不依赖于蒙特卡洛模拟的函数）是一个仍在研究的领域。基于主题的影响最大化，即用户对不同主题的影响不同是本文没有涉及的另一个领域。影响退化是一个涉及动态影响概率的研究领域，随着时间的推移会发生变化，更智能的种子节点选择是可以进一步改进的方向。

参 考 文 献

[1] KEMPE D, KLEINBERG J, TARDOS É. Maximizing the spread of influence through a social network[C]// Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2003: 137-146.

[2] NEWMAN M E J. The structure of scientific collaboration networks[J]. Proceedings of the National Academy of Sciences, 2001, 98(2): 404-409.

[3] GAO N, WANG B, LU K, et al. Teaching-learning-based optimization of an ultra-broadband parallel sound absorber[J/OL]. Applied Acoustics, 2021, 178: 107969. https://doi. org/10. 1016/j. apacoust. 2021. 107969.

[4] BANERJEE A V. A simple model of herd behavior [J]. The Quarterly Journal of Economics, 1992, 107(3): 797-817.

[5] DAS K, PACHORI R B. Schizophrenia detection technique using multivariate iterative filtering and multichannel EEG signals[J/OL]. Biomedical Signal Processing and Control, 2021, 67: 102525. https://doi. org/10. 1016/j. knosys. 2019. 07. 004.

[6] LESKOVEC J, KRAUSE A, GUESTRIN C, et al. Cost-effective outbreak detection in networks[C]// Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2007: 420-429.

[7] NGUYEN H T, THAI M T, DINH T N. Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks[C]// Proceedings of the 2016 International Conference on Management of Data. 2016: 695-710.

[8] DAGUM P, KARP R, LUBY M, et al. An optimal algorithm for Monte Carlo estimation[J]. SIAM Journal on Computing, 2000, 29(5): 1484-1496.

[9] LV J, YANG B, YANG Z, et al. A community-based algorithm for influence blocking maximization in social networks[J]. Cluster Computing, 2019, 22(3): 5587-5602.

[10] WANG X G. A new algorithm for the influence maximization problem in dynamic networks or traffic sensor networks[J]. Multimedia Tools and Applications, 2016, 75(8): 4833-4844.

[11] LI W M, LI Z, LUVEMBE A M, et al. Influence maximization algorithm based on Gaussian propagation model[J/OL]. Information Sciences, 2021, 568: 386-402. https://doi. org/10. 1016/j. ins. 2021. 04. 061.

[12] TANG J, ZHANG R, WANG P, et al. A discrete shuffled frog-leaping algorithmto identify influential nodes for influence maximization in social networks[J/OL]. Knowledge-Based Systems, 2020, 187: 104833. https://doi. org/10. 1016/j. knosys. 2019. 07. 004.

[13] WU Y Z. Research on sequential regression technology based on evolutionary algorithm[D]. Heifei: University of Science and Technology of China, 2016.

[14] WANG X, ZHANG Y, ZHANG W, et al. Efficient distance-aware influence maximization in geo-social networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 29(3): 599-612.

[15] LIU B, CONG G, XU D, et al. Time constrained influence maximization in social networks[C]// 2012 IEEE 12th International Conference on Data Mining. IEEE, 2012: 439-448.

[16] MOHAMMADI A, SARAEE M, MIRZAEI A. Time-sensitive influence maximization in social networks[J]. Journal of Information Science, 2015, 41(6): 765-778.

[17] BARBIERI N, BONCHI F, MANCO G. Topic-aware social influence propagation models[J]. Knowledge and Information Systems, 2013, 37(3): 555-584.

[18] HE Q, FANG H, ZHANG J, et al. Dynamic Opinion Maximization in Social Networks[J/OL]. IEEE Transactions on Knowledge and Data Engineering, 2021. https://ieeexplore. ieee. org/abstract/document/9423621.

[19] SHU R, CHENG P, CHEN G, et al. Direct Universal Access: Making Data Center Resources Available to {FPGA}[C]// 16th {USENIX} Symposium on Networked Systems Design and Implementation({NSDI} 19). 2019: 127-140.

[20] HUNTER II T. Advanced microservices: a hands-on approach to microservice infrastructure and tooling[M]. Apress, 2017.

[21] HE Q, WANG X, ZHAO Y, et al. Reinforcement Learning Based Competitive Opinion Maximization Approach in Signed Social Networks[J/OL]. IEEE Transactions on Computational Social Systems, 2021. https://ieeexplore. ieee. org/abstract/document/9611781.

[22] HUDSON N, KHAMFROUSH H. Behavioral Information Diffusion for Opinion Maximization in Online Social Networks[J]. IEEE Transactions on Network Science and Engineering, 2020, 8(2): 1259-1268.

[23] MA H, LYU M R, KING I. Learning to recommend with trust and distrust relationships[C]// Proceedings of the Third ACM Conference on Recommender Systems. 2009: 189-196.

[24] MORRIS M, KRETZSCHMAR M. Concurrent partnerships and transmission dynamics in networks[J]. Social networks, 1995, 17(3/4): 299-318.

[25] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: Bringing order to the web[R]. Stanford InfoLab, 1999.

[26] HE Q, SUN L, WANG X, et al. Positive opinion maximization in signed social networks[J/OL]. Information Sciences, 2021, 558: 34-49. https://doi. org/10. 1016/j. ins. 2020. 12. 091.

[27] HE Q, WANG X, HUANG M, et al. Multi-stage opinion maximization in social networks[J]. Neural Computing and Applications, 2021, 33(19): 12367-12380.

[28] LI H, PAN L, WU P. Dominated competitive influence maximization with time-critical and time-delayed diffusion in social networks[J/OL]. Journal of computational science, 2018, 28: 318-327. https://doi.org/10.1016/j.jocs.2017.10.015

[29] GOMEZ-RODRIGUEZ M, SCHÖLKOPF B. Influence Maximization in Continuous Time Diffusion Networks[J]. arXiv: 1205.1682, 2012.

[30] LI J, SELLIS T, CULPEPPER J S, et al. Geo-social influence spanning maximization[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(8): 1653-1666.

[31] NEMHAUSER G L, WOLSEY L A, FISHER M L. An analysis of approximations for maximizing submodular set functions—I[J]. Mathematical Programming, 1978, 14(1): 265-294.

[32] ALI K, WANG C Y, CHEN Y S. A novel nested q-learning method to tackle time-constrained competitive influence maximization[J/OL]. IEEE Access, 2018, 7: 6337-6352. https://ieeexplore.ieee.org/abstract/document/8584421.

[33] QIU L Q, YU J F, FAN X, et al. Analysis of influence maximization in temporal social networks[J/OL]. IEEE Access, 2019, 7: 42052-42062. https://ieeexplore.ieee.org/abstract/document/8620999.

[34] CHEN W, LIN T, YANG C. Real-time topic-aware influence maximization using preprocessing[J]. Computational social networks, 2016, 3(1): 1-19.

[35] ZHANG Z K, LIU C, ZHAN X X, et al. Dynamics of information diffusion and its applications on complex networks[J/OL]. Physics Reports, 2016, 651: 1-34. https://doi.org/10.1016/j.physrep.2016.07.002.

[36] LIN Y, LUI J C S. Analyzing competitive influence maximization problems with partial information: An approximation algorithmic framework[J/OL]. Performance Evaluation, 2015, 91: 187-204. https://doi.org/10.1016/j.peva.2015.06.012.

[37] ZHANG Y, BURER S, NICK STREET W, et al. Ensemble Pruning Via Semi-definite Programming[J]. Journal of Machine Learning Research, 2006, 7(7): 1315-1338.

[38] RICHARDSON M, DOMINGOS P. Mining knowledge-sharing sites for viral marketing[C]// Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2002: 61-70.

[39] MYERS S A, LESKOVEC J. On the convexity of latent social network inference[J]. arXiv: 1010.5504, 2010.

[40] WANG L, ERMON S, HOPCROFT J E. Feature-enhanced probabilistic models for diffusion network inference[C]// Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2012: 499-514.

[41] JACCARD P. The distribution of the flora in the alpine zone. 1[J]. New Phytologist, 1912, 11(2): 37-50.

[42] GOEMANS M X, WILLIAMSON D P. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming[J]. Journal of the ACM (JACM), 1995, 42(6): 1115-1145.

[43] DU N, SONG L, YUAN M, et al. Learning networks of heterogeneous influence[J/OL]. Advances in Neural Information Processing Systems, 2012, 25: 2780-2788. https://dl.acm.org/doi/abs/10.5555/2999325.2999445.

[44] DU N, LIANG Y, BALCAN M, et al. Influence function learning in information diffusion networks[C]// International Conference on Machine Learning. PMLR, 2014: 2016-2024.

[45] GOMEZ-RODRIGUEZ M, LESKOVEC J, KRAUSE A. Inferring networks of diffusion and influence[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2012, 5(4): 1-37.

[46] NETRAPALLI P, SANGHAVI S. Learning the graph of epidemic cascades[J]. ACM SIGMETRICS Performance Evaluation Review, 2012, 40(1): 211-222.

[47] D'ASPREMONT A, EL GHAOUI L, JORDAN M I, et al. A direct formulation for sparse PCA using semidefinite programming[J]. SIAM Review, 2007, 49(3): 434-448.

[48] LANCKRIET G R G, CRISTIANINI N, BARTLETT P, et al. Learning the kernel matrix with semidefinite programming[J]. Journal of Machine Learning Research, 2004, 5(Jan): 27-72.

[49] FUJISAWA K, FUKUDA M, KOJIMA M, et al. SDPA-C (semidefinite Programming Algorithm-Completion Method). User's Manual-Version 6-10[M]. Inst. of Technology, 2004.

[50] LAGRÉEP, CAPPÉO, CAUTISB, et al. Algorithms for online influencer marketing[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2018, 13(1): 1-30.

[51] COHEN E, DELLING D, PAJOR T, et al. Sketch-based influence maximization and computation: Scaling up with guarantees[C]// Proceedings of the 23rd ACM International Conference on Information and Knowledge Management. 2014: 629-638.

[52] GOYAL A, LU W, LAKSHMANAN L V S. Celf++ optimizing the greedy algorithm for influence maximization in social networks[C]// Proceedings of the 20th International Conference Companion on World Wide Web. 2011: 47-48.

[53] COHEN E, DELLING D, PAJOR T, et al. Sketch-based influence maximization and computation: Scaling up with guarantees[C]// Proceedings of the 23rd ACM International Conference on Information and Knowledge Management. 2014: 629-638.

[54] BAO Z K, LIU J G, ZHANG H F. Identifying multiple influential spreaders by a heuristic clustering algorithm[J]. Physics Letters A, 2017, 381(11): 976-983.

[55] WANG X, SU Y, ZHAO C, et al. Effective identification of multiple influential spreaders by Degree Punishment[J/OL]. Physica A: Statistical Mechanics and its Applications, 2016, 461: 238-247. https://doi.org/10.1016/j.physa.2016.05.020.

[56] CHEN W, WANG Y, YANG S. Efficient influence maximization in social networks[C]// Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2009: 199-208.

[57] FREEMAN L C. Centrality in social networks conceptual clarification[J]. Social Networks, 1978, 1(3): 215-239.

[58] BUCUR D, IACCA G. Influence maximization in social networks with genetic algorithms[C]// European Conference on the Applications of Evolutionary Computation. Cham: Springer, 2016: 379-392.

[59] JIANG Q，SONG G，GAO C，et al. Simulated annealing based influence maximization in social networks[C]// Twenty-fifth AAAI Conference on Artificial Intelligence. 2011.

[60] TSAI C W，YANG Y C，CHIANG M C. A genetic newgreedy algorithm for influence maximization in social network[C]// 2015 IEEE International Conference on Systems，Man，and Cybernetics. IEEE，2015：2549-2554.

[61] GONG M，YAN J，SHEN B，et al. Influence maximization in social networks based on discrete particle swarm optimization[J/OL]. Information Sciences，2016，367：600-614. https://doi.org/10.1016/j.physa.2016.05.020.

[62] ERDOS P，RÉNYI A. On the evolution of random graphs[J]. Publ. Math. Inst. Hung. Acad. Sci，1960，5(1)：17-60.

[63] GOLDENBERG J，LIBAI B，MULLER E. Using complex systems analysis to advance marketing theory development：Modeling heterogeneity effects on new productgrowth through stochastic cellular automata[J]. Academy of Marketing Science Review，2001，9(3)：1-18.

[64] GOMEZ-RODRIGUEZ M，LESKOVEC J，KRAUSE A. Inferring networks of diffusion and influence[J]. ACM Transactions on Knowledge Discovery from Data(TKDD)，2012，5(4)：1-37.

[65] DANESHMAND H，GOMEZ-RODRIGUEZ M，SONG L，et al. Estimating diffusion network structures：Recovery conditions，sample complexity & soft-thresholding algorithm[C]// International Conference on Machine Learning. PMLR，2014：793-801.

[66] LESKOVEC J，BACKSTROM L，KLEINBERG J. Meme-tracking and the dynamics of the news cycle[C]// Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2009：497-506.

[67] NAKATA K，FUJISAWA K，FUKUDA M，et al. Exploiting sparsity insemidefinite programming via matrix completion II：Implementation and numerical results[J]. Mathematical Programming，2003，95(2)：303-327.

[68] DU N，SONG L，GOMEZ-RODRIGUEZ M，et al. Scalable influence estimation in continuous-time diffusion networks[J/OL]. Advances in Neural Information Processing Systems，2013，26：3147. https://dl.acm.org/doi/abs/10.5555/2999792.2999963.

[69] BISWAS T K，ABBASI A，CHAKRABORTTY R K. An MC-DM integrated adaptive simulated annealing approach for influence maximization in social networks[J/OL]. Information Sciences，2021，556：27-48. https://doi.org/10.1016/j.ins.2020.12.048.

[70] 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.

[71] SAITO K，KIMURA M，OHARA K，et al. Learning continuous-time information diffusion model for social behavioral data analysis[C]// Asian Conference on Machine Learning. Berlin：Springer，2009：322-337.

[72] YANG G，CAO Y，TAO H. A method for multi-objective optimization and application in automobile impact[J/OL]. Journal of Physics：Conference Series. 2021，1802(3)：032129. https://doi.org/10.1088/1742-6596/1802/3/032129.

HE Qiang，born in 1991，Ph.D，associate professor，is a member of China Computer Federation. His main research interests include social networks and machine learning.

YIN Zhen-yu，born in 1979，Ph.D，is a senior member of China Computer Federation. His main research interests include industrial embedded systems，industrial Internet of things，etc.

（责任编辑：喻藜）