

《社交大数据环境下影响力最大化节点挖掘技术及其应用》

项目总结报告

一、项目概况

项目基本信息：本项目是由东南大学和苏州米谷网络科技有限公司联合承担，立项时间为2014年9月，项目编号为BY2014127-10，项目负责人为东南大学计算机科学与工程学院曹玖新。本项目研究经费150万元，其中省拨款30万元，单位自筹120万元。

计算机技术的快速发展和社交方式的多样化催生了很多大规模的在线社交网站。这些大型在线社交网络的成员数目庞大、结构错综复杂、交互行为数据繁多。它们的出现对社交网络的研究提出了巨大的机遇和挑战。通过研究影响最大化问题不仅有助于人们理解影响传播的动态过程，而且在社会心理学、传播学、公共关系学、广告学、营销学、消费行为学等领域也具有重要的应用，比如市场营销、疾病预防、舆论引导和社会安定等。

最大影响力节点挖掘的目标是挖掘社交网络社区中具有一定影响力的节点，这些节点的交互行为对整个网络的信息传播趋势具有重大影响。因此，最大影响力节点的挖掘有助于网络舆情的监控与引导。微博是目前主流的社交网络，信息流动速度快，用户交互行为频繁，以内容简洁、发布实时、互动性强、自由开放和高容纳性吸引了越来越多的用户，成为虚拟全民社会。通过对一些意见领袖的监控，可以有效地从源头抑制谣言的流传与扩散。另外，通过对意见领袖的舆论引导，如微博中通过意见领袖的转播或评论行为使热点新闻消息得到爆炸式传播，产生一定的社会反响，引起有关部门的高度关注，目前政府部门竞相开通政务微博，微博、BBS等社交媒体成为事件曝光、网络反腐、政务公开的一个重要网络平台。作为传播媒介的社交网络，在社会个体之间相互影响，信息传播和思想引导方面，发挥着重大作用。研究表明，通过朋友或其他信任源的推荐可以获得更高的满足度。这就提出了如何充分利用个体之间的相互关系来达到最好影响效果问题。例如一家创业公司新开发了一个新型且很实用的产品，想通过社交网络对产品做推广营销，然而推广资金有限，例如厂家只能给小部分用户发放免费产品去使用它，如果其中有用户满意他们就会购买，并且推荐给他们的朋友也去购买，这些受到影响的朋友又会把该产品推荐给他们自己的朋友，同样的过程持续进行，就产生了级联推荐的效果，这就是营销领域的“口碑效应”(work-of-mouth)现象和“病毒式营销”(viral marketing)。

综上所述，针对以上问题，东南大学与苏州米谷网络科技有限公司联合承担

了江苏省产学研前瞻性联合研究项目“社交大数据环境下影响力最大化节点挖掘技术及其应用”的研究工作，去研究真实社交网络中影响力最大化节点挖掘技术，即在大规模社交网络中挖掘一个较小的节点集合使得该集合在社交网络中的影响力最大化，通过高效且准确的影响力最大化节点挖掘算法，实现企业广告在社交网络中的精准营销与投放、社会舆论的监控等。具体研究内容包括六个方面：

(1) 基于话题的最大影响力节点挖掘：本部分旨在基于社交网络用户的交互行为和信息传播机制，对社交网络的节点影响力进行深度挖掘。通过建立合适的用户节点属性模型，包括结构属性和话题属性，设计有效的挖掘算法，解决能适用于真实社交环境中的意见领袖识别算法，从而为微博类新型社交网络的舆情分析与传播控制提供技术支撑。

(2) 基于拓扑结构的最大影响力节点挖掘：本部分首先对真实社交网络数据集进行抽象，分析网络中的结构特性和节点属性；然后对所有数据集提供一个抽象的接口，以供其他模块使用，而无需关注具体的数据集特征；再次，基于具体的传播模型和网络接口，对节点的影响力进行建模；最后，基于节点核数概念提出高效的影响力节点集合发现算法；同时，提出一种综合启发式和贪心策略的混合算法，它既有着贪心算法那样稳定且高的传播影响力，又没有传统贪心算法那样高的时间复杂度，在时间效率和影响效果上达到良好的平衡。

(3) 基于节点主题属性的最大影响力节点挖掘：通过对比发现，本部分研究的问题与基于拓扑结构的影响最大化问题的最大区别就在于网络中加入了节点主题属性信息，最终选取的节点在具有特定主题属性的节点集内影响最大化。节点主题属性对信息传播的影响从深层次来说，是节点主题属性对节点间影响概率的影响，这是由于节点主题属性影响节点对信息的偏好程度。因此，解决基于节点主题属性的影响最大化问题的直观思路就是：根据节点主题属性对节点间影响概率建模，在此基础上研究高效的影响最大化算法挖掘 k 个影响力节点。

(4) 基于成本效益的最大影响力节点挖掘：传统的影响最大化问题的研究在选择影响力节点时忽略了节点之间的差异性，没有考虑到选择节点所要花费的成本，而实际营销中的推广活动都有预算的约束，选择不同的用户需要不同的花费。基于以上考虑，本项目给出了对节点成本的定义方法，并提出了基于成本效益的影响最大化问题。针对这个问题，本部分综合考虑网络拓扑和传播模型的特点，提出了基于概率覆盖范围的启发式算法，并在此基础上利用子模函数特性和惰性计算技术提出基于概率覆盖范围的惰性节点选择算法。

(5) 基于成本的广告投放策略研究：本部分同时考虑了选择节点所需要的成本和节点主题属性两个因素，提出了基于成本的广告投放问题，旨在广告预算有限的情况下选择合适的初始种子节点来达到尽可能好的广告效应，这个问题本质上是考虑了成本效益和节点主题属性的影响最大化问题。针对这个问题，本部分

首先结合网络营销平台和现实情况对节点成本建模；然后利用节点对某信息的偏好程度(节点主题属性与广告主题间的相似度)和节点之间关系亲密程度(节点间的关注度)这两个因素对节点间影响概率 p_{uv} 进行建模；最后根据节点间影响概率 p_{uv} 和节点的成本提出新的度量节点影响力的指标：平均概率，之后设计一种基于平均概率的种子节点选择算法。

(6) 社交网络中影响力最大化节点挖掘系统的实现：根据上述研究内容，采用本项目的研究成果，实现社交网络中影响力最大化节点挖掘系统。该系统主要由以下几个模块组成：数据输入与处理模块、算法实现模块、传播模拟模块和结果可视化模块。

二、项目实施情况

经东南大学及苏州米谷网络科技有限公司双方协作，本项目成立了以项目负责人为首的校企联合研发团队共计 15 人，其中副教授 2 人，讲师 1 人，企业工程师 5 人，博硕士研究生 7 人。

经过两年多的产学研合作研究，较系统地研究了社交大数据环境下影响力最大化节点挖掘技术，成功实现了社交网络中影响力最大化节点挖掘原型系统，圆满完成了研发和产业化任务。具体表现在以下几方面：

(1) 在基于话题的最大影响力节点挖掘方面，通过社交网络提供的开放接口获得我们所需要的数据集，包括用户信息，微博内容等。在此基础上提取评论集进行话题识别，得到网络兴趣社区。在网络兴趣社区中对于每个话题形成的用户发帖和回帖网络进行用户的观点挖掘，得到话题的意见评价情况。最后，在网络兴趣社区中结合观点挖掘得到的结果以及网络兴趣社区的结构特征与话题特征识别某个兴趣社区中的意见领袖，为话题传播的舆情控制与引导提供决策支持。

(2) 在基于拓扑结构的最大影响力节点挖掘方面，基于社交网络的拓扑结构和节点之间的等概率传播，提出一种基于节点核数的核覆盖算法，k-核解析是一种研究大规模网络的层次性方法，通过 k-核解析，网络逐渐趋于核心的区域，节点核数越大，其连通性也越强，并在节点之间存在大量不同的路径，再通过限制节点间的距离，使得影响力能够有效的传播。同时，提出一种综合启发式和贪心策略的混合算法，它既有着贪心算法那样稳定且高的传播影响力，又没有传统贪心算法那样高的时间复杂度，在时间效率和影响效果上达到良好的平衡。

(3) 在基于节点主题属性的最大影响力节点挖掘方面，选择影响力节点时加入了节点主题属性信息。节点主题属性对信息传播的影响从深层次来说，是节点主题属性对节点间影响概率的影响，这是由于节点主题属性影响节点对信息的偏好程度。因此，针对这个问题，首先根据节点主题属性对节点间影响概率建模，在此基础上研究高效的影响最大化算法挖掘 k 个影响力节点。

(4) 在基于成本效益的最大影响力节点挖掘方面，选择影响力节点时考虑了

节点成本因素，并提出了基于成本效益的影响最大化问题。本项目对节点成本建模使得影响最大化问题更贴近真实，同时综合考虑网络拓扑和传播模型的特点，提出了基于概率覆盖范围的启发式算法，并在此基础上利用子模函数特性和惰性计算技术提出基于概率覆盖范围的惰性节点选择算法。

(5) 在基于成本的广告投放策略研究方面，选择影响力节点时同时考虑了节点成本和节点主题属性。在对节点成本建模时结合了网络营销平台和现实的情况，对节点间影响概率 p_{uv} 建模时又利用了节点对某信息的偏好程度和节点之间关系亲密程度这两个因素，最后根据节点间影响概率 p_{uv} 和节点的成本提出新的度量节点影响力的指标：平均概率，然后设计了基于平均概率的种子节点选择算法。

(6) 在社交网络中影响力最大化节点挖掘系统实现方面，根据本课题的理论研究成果，实现了社交网络中影响力最大化节点挖掘系统，该系统主要包括数据输入与处理模块、算法实现模块、传播模拟模块和结果可视化模块。这样可以将整个研究成果更直观地呈现给用户，使得用户能更方便地使用它，同时也可以将复杂数据转换成图形或图像在屏幕上显示出来，并与用户进行交互。

(7) 项目共申请发明专利 3 项；软件著作权 1 项；在国内外核心期刊和会议上发表论文 6 篇，EI 收录论文 6 篇。

(8) 项目研究过程中，培养博士研究生 1 名(周涛)，硕士研究生 7 名(徐顺、董丹、吴江林、赵钊、王浩然、冯雪艳、袁堂飞)。

三、项目技术情况

本项目主要围绕社交大数据环境下影响力最大化节点挖掘技术，从基于话题的最大影响力节点挖掘、基于拓扑结构的最大影响力节点挖掘、基于节点主题属性的最大影响力节点挖掘、基于成本效益的最大影响力节点挖掘、基于成本的广告投放机制以及社交网络中影响力最大化节点挖掘系统六个方面进行研究与开发，具体技术情况汇报如下：

1. 基于话题的最大影响力节点挖掘

(1) 解决的关键技术

通过社交网络提供的开放接口获得我们所需要的数据集，包括用户信息，微博内容等。在此基础上提取评论集进行话题识别，得到网络兴趣社区。在网络兴趣社区中对于每个话题形成的用户发帖和回帖网络进行用户的观点挖掘，得到话题的意见评价情况。最后，在网络兴趣社区中结合观点挖掘得到的结果以及网络兴趣社区的结构特征与话题特征识别某个兴趣社区中的意见领袖，为话题传播的舆情控制与引导提供决策支持。

(2) 研究方法及技术路线

基于兴趣的话题识别。通过对用户产生的话题进行文本分析，采用聚类算法

进行话题识别。得到话题集后进行话题-用户映射，识别出主题社区。

多维度属性挖掘和观点挖掘。意见领袖挖掘建立在网络兴趣社区中，为了获得意见领袖的影响力和权威度，有必要从评论内容角度分析周围用户对意见领袖的态度趋势，同时还需要挖掘其他体现用户权威度的特征，比如用户的粉丝数，参与话题的活跃度，所发表微博获得的转发数与评论数等。因此，本项目不仅分析节点的结构特征和话题特征，而且考虑用户之间的评价，基于情感分析，从评论内容角度分析用户对话题的意见态度趋势。

意见领袖节点挖掘。在社交网络中，用户的大多数行为发生在一个社区中，不同社区之间的用户交互行为较少。为了更好地研究用户群中的意见领袖，根据网络兴趣社区和观点挖掘得到的网络模型，以及社会网络的动态特征，在分析用户节点意见趋势的基础上，基于正向和负向评价提出意见领袖识别算法获得意见领袖，有效地提高意见领袖识别的准确性。

(3) 取得成果及创新点

本项目基于话题研究社交网络中的最大化影响力问题。从社交网络的拓扑结构与信息内容角度出发，基于真实的社交网络数据，对产生的微博内容进行话题识别，构建主题社区。分析了用户的意见趋势以及影响力在主题社区中的分布规律，建立结构特征、话题特征和情感类特征多维度体系，利用这些影响力影响因子挖掘主题社区的意见领袖。这同时兼顾了动态特征以及用户对话题的情感倾向，可以更加准确地挖掘具有较大影响力的节点。

2. 基于拓扑结构的最大影响力节点挖掘

(1) 解决的关键技术

本项目针对目前影响最大化算法中存在的不足，提出以降低算法时间复杂度和提高影响范围为目标的影响最大化新算法。因此，首先从网络的拓扑结构出发，分析节点之间的关联关系且对节点影响力建模；其次提出基于核数的节点选择策略和结合启发式和贪心策略的混合算法来求解影响最大化问题；最后在多个数据集和多种传播模型上，分析对比各种算法，以验证所提出算法的有效性。

(2) 研究方法及技术路线

节点影响力建模。基于节点核数来定义节点间影响力并设计影响最大化算法。

提出一种基于节点核数的核覆盖种子节点选择算法。 k -核解析是一种研究大规模网络的层次性方法，通过 k -核解析，网络逐渐趋于核心的区域，节点核数越大，其连通性也越强，并在节点之间存在大量不同的路径。再通过限制节点间的距离，使得影响力能够有效的传播。

提出一种综合启发式和贪心策略的混合种子节点选择算法。它既有着贪心算法稳定且高效的传播影响力，又低于传统贪心算法的时间复杂度，在时间效率和

影响效果上达到良好的平衡。

(3) 取得成果及创新点

本项目的主要工作是针对现有影响最大化算法所存在的问题，从网络的拓扑结构出发，分析节点之间的关联关系且对节点影响力建模，并提出了 CCA 和 MHG 算法，CCA 启发式算法利用核数高的影响力节点又使它们保持一定的距离，从而减少重叠激活的可能性；MHG 算法通过把节点边际效益的计算限定在某些最具有潜力的候选节点中，因此能比贪心算法取得多个数量级的加速比和匹配的影响范围。

3. 基于节点主题属性的最大影响力节点挖掘

(1) 解决的关键技术

本项目在原始影响最大化问题基础上加入节点主题属性信息，提出基于节点主题属性的影响最大化问题，该问题就是寻找网络中 k 个影响力节点，使得受到影响的节点与商品目标受众属性相似度最大，为社交网络中的窄告提供相应的技术。通过对比发现，新问题与原始影响最大化问题的最大区别就是在网络中加入了节点主题属性信息，最终选取的影响力节点一般是不同的。节点主题属性对信息传播的影响从深层次来说，是由于节点属性影响节点间的影响概率 p_{uv} ，这是因为节点的自身属性影响节点对信息的偏好，也就是影响节点受影响的概率。因此，本项目先根据节点主题属性对节点间影响概率 p_{uv} 建模，在此基础上研究高效的影响最大化算法去寻找 k 个影响力节点。

(2) 研究方法及技术路线

根据节点主题属性对节点间影响概率建模。在独立级联模型下求解影响最大化问题时，学者们往往将节点间的影响概率 p_{uv} 视为一个固定的系统变量，但在实际生活中，人与人之间的关系存在强弱之分，而且在社交网络中用户与用户之间的关系种类(如朋友、同学、同事、家人等)是不同的，这种关系强弱和关系种类的差异都会影响节点之间影响概率 p_{uv} ，因此一种拓展独立级联模型被提了出来，它是独立级联模型的一种特殊的变形，其与独立级联模型的唯一区别是节点间影响 p_{uv} 是不相同的。节点受某一信息影响的概率不仅仅受节点之间关系亲密程度的影响，还取决于该节点对该信息的偏好程度。

提出一种在拓展独立级联模型下高效的影响最大化算法。通过对节点间影响概率 p_{uv} 的建模得到的节点间影响概率 p_{uv} 是不一致的，这就使得现有的在独立级联模型下的启发式算法在该模型下变得不适用了。因此本项目在拓展独立级联模型下提出一种基于概率和覆盖的启发式算法。

基于节点主题属性和网络拓扑结构的社区识别。复杂网络(如社交网络、蛋白质网络、科学家合作网、科学引文网等)除了拥有小世界、无标度这两个基本特性，还存在社区结构这一特性。本项目要解决的问题是找出在特定属性社团中

具有较大影响力的节点,那么一个很自然的想法就是基于节点主题属性和网络拓扑结构进行社区识别,识别出的社区中节点具有相似属性并且任意两节点在全局物理拓扑结构上是连通的,再在社区的基础上利用基于概率和的影响最大化算法求解影响力节点。

在符合目标条件的 $n(n \geq 1)$ 个社区中寻找 k 个影响力节点。本项目首先基于节点主题属性和物理拓扑结构进行社区识别,识别出的社区不一定在一个连通图内,而存在多个连通图,那么如何更加合理地在多个社区中分配 k 个影响力节点将是一个必须要解决的问题。

(3) 取得成果及创新点

本项目将病毒式营销手段和窄告这种个性化的广告模式结合起来,提出了基于主题属性的影响最大化问题。针对该问题,首先通过微博开放接口抓取了大量真实的数据并对抓取的数据进行预处理;其次,基于抓取的大量真实数据,考虑节点间交互强度和节点主题属性与商品目标受众属性间相似度这两个因素,对节点间影响概率建模;最后提出一种基于概率和的算法 SoPCA 和两种基于局部社区的启发式算法 CBA 和 ICBA。

4. 基于成本效益的最大影响力节点挖掘

(1) 解决的关键技术

本文以社会计算中影响最大化问题为基础,针对现有研究的一些不足,提出基于成本效益的影响最大化问题,把选择节点所需要花费的成本考虑进来,目的是在推广预算有限的情况下选择适当的初始受众节点,达到尽可能广的传播范围。为实现上述研究思路,本文将首先对节点成本建立模型,使影响最大化问题更加符合现实意义,然后在推广预算有限的情况下提出高效的算法来求解这一新问题,最后分析对比各种算法,以验证所提出算法的有效性。

(2) 研究方法及技术路线

从现实出发对节点成本建模。综合对一些网络营销平台的观察,得知广告商在利用社交平台做推广时,通常根据用户的粉丝数或好友数来支付相应的报酬,那么根据实际情况映射到抽象的图中,本项目采用度数作为选择节点所要付出的成本。

提出一种基于节点概率覆盖范围的启发式算法。节点概率覆盖范围不仅考虑节点自身属性,而且也综合考虑传播模型的特点和网络拓扑结构,能够较精确的反映节点可能的影响范围。通过计算节点概率覆盖范围,再结合节点的成本,以单位成本下节点概率覆盖范围为指标,采取性价比最优的选择策略获得初始节点集合,在预算有限的情况下使得影响力能够有效的传播。

提出一种基于概率覆盖范围的惰性节点选择算法。对节点概率覆盖范围的计算方法稍作修改,使其满足子模函数特性,并利用这一特性和惰性计算技术,进

一步提高时间效率和影响效果。

(3) 取得成果及创新点

本项目针对现有研究的一些不足，提出了基于成本效益的影响最大化问题，把选择节点所需要花费的成本考虑进来，目的是在推广预算有限的情况下选择适当的初始受众节点，达到尽可能广的传播范围，并据此提出了高效的影响力节点选择算法 ProbCover。为了避免节点影响力重叠，又提出了一种惰性节点选择算法 ProbCoverLF，使其满足子模函数特性，并利用这一特性和惰性计算技术进一步提高了时间效率和影响效果。

5. 基于成本的广告投放机制

(1) 解决的关键技术

本项目以社会计算中原始影响最大化问题为基础，针对现有研究的不足，根据社交网络本身的特性和其空间结构分布特征提出社交网络中基于成本的广告投放问题，把选择节点所需花费的成本以及节点主题属性信息考虑进来，旨在广告预算有限的情况下选择适当的初始种子节点，达到尽可能广的传播范围，使得最终达到尽可能好的广告效应，这一问题本质上是考虑了成本效益和节点主题属性的影响最大化问题，为解决新提出的问题，首先对节点成本建立模型，之后考虑节点主题属性对选择影响力节点带来的差异，使基于成本的广告投放问题更加符合现实意义，然后在广告预算有限的情况下设计高效的种子节点选择算法。

(2) 研究方法及技术路线

对节点成本建模。在已有的影响最大化问题的研究中，学者们忽略了在选择初始种子节点时各个节点都有成本且成本不同的问题。多数研究在选择初始种子节点时，认为对于不同节点所要付出的成本是相同的，然而这与事实不符，请名人和普通人做产品推广，这两者需要的推广费用相差巨大，并且不同的名人根据其影响力的大小产生的推广费也千差万别。为了更真实地反应现实情况，节点成本的定义更加符合现实意义，将节点成本建模为节点的度数。

对节点间影响概率建模。一般信息都会有主题，那么该信息影响到某节点的概率就同时取决于节点之间关系的亲密程度和某节点对该信息的偏好程度，也就是说如果某用户对该信息比较感兴趣的话，即使先前该用户与其他用户的亲密程度不强，他也很容易被影响到。因此，本项目用节点间的关注度来表示节点之间关系的亲密程度，用节点的属性与广告主题间的相似度来表征不同节点对不同主题的广告信息的偏好差异。

提出一种基于平均概率的启发式算法。通过对比发现，本项目研究的问题与原始影响最大化问题的区别就是考虑了节点的选择成本和在网络中加入了节点主题属性信息，使得最终选取的影响力节点不同。算法的选择策略不同也会造成选择初始种子节点集合所需要的成本差别很大，并且节点主题属性对信息传播的

影响从深层次来说，是因为节点主题属性影响节点对信息的偏好，即影响节点受影响的概率，从而影响节点间的影响概率。因此，针对新提出的问题，综合节点的成本和节点间影响概率这两个因素，提出新的度量节点影响力的指标：平均概率，之后提出一种基于平均概率的启发式算法，在广告预算有限的条件下去选择合适的初始种子节点。

(3) 取得成果及创新点

本项目用新浪微博数据作为数据集，以社会计算中原始影响最大化问题为基础，针对现有研究的一些不足，根据社交网络本身的特性和其空间结构分布特征提出社交网络中基于成本的广告投放问题，它是将选择节点时所需要的不同成本与节点主题属性同时引入到初始种子节点的选择过程中，以此利用量化的影响力评价标准，设计了一种基于平均概率的种子节点选择算法 AvePA，在广告预算有限的情况下选择满足条件的初始种子节点，使得最终达到尽可能好的广告效应。因此在社交网络中投放广告时可以进一步缩小广告受众的范围，使得投放结果更精准、深入。

6. 社交网络中影响力最大化节点挖掘系统

(1) 解决的关键技术

本课题分别从以下五个不同的侧重点展开研究：1. 基于话题的影响最大化问题；2. 基于拓扑结构的影响最大化问题；3. 基于节点主题属性的影响最大化问题；4. 基于成本效益的影响最大化问题；5. 基于成本的广告投放问题。针对每个研究点都提出了高效的解决方案，基于研究成果，设计并实现了社交网络中影响力最大化节点挖掘系统。

(2) 研究方法及技术路线

数据输入和处理模块。该模块主要是进行数据的输入并对数据进行处理，包括广告预算的输入、读取数据集文件、建图、解析图、选择种子节点选择算法以及各种参数的设置，从而完成系统实现的前期工作并得到算法实现模块所需要的数据；

算法实现模块。该模块主要是实现针对各个研究点所提出的高效算法和多种已有的具有代表性的影响最大化算法，包括 Degree 算法、DegreeDiscount 算法、PageRank 算法、IRIE 算法、CCA 算法和 SoPCA 算法；

传播模拟模块。该模块主要是基于拓扑图实现特定的影响力传播模型，并根据各个算法得到的种子节点集合来模拟影响力在整个拓扑中的传播过程，从而得到种子节点集合的影响效果；

结果可视化模块。该模块主要是对各个算法得到的种子节点数以及传播模拟过程结束后得到的处于激活态的节点数和最终的影响效果进行可视化。

(3) 取得成果及创新点

根据本课题的理论研究成果，实现了社交网络中影响力最大化节点挖掘系统，该系统主要包括数据输入与处理模块、算法实现模块、传播模拟模块和结果可视化模块。这样可以将整个课题的研究成果更直观地呈现给用户，使得用户能更方便地使用它，同时也可以将复杂数据转换成图形或图像在屏幕上显示出来，并与用户进行交互。

四、合同任务指标完成情况

在江苏省产学研前瞻性研究项目的资助下，东南大学联合苏州米谷网络科技有限公司开展了紧密的产学研合作研究，取得了以下研究成果：

1. 主要的技术指标完成情况

(1) 深入研究了基于话题的最大影响力节点挖掘技术：从社交网络的拓扑结构与信息内容角度出发，基于真实的社交网络数据，对产生的微博内容进行话题识别，构建主题社区。分析了用户的意见趋势以及影响力在主题社区中的分布规律，建立结构特征、话题特征和情感类特征多维度体系，利用这些影响力影响因子挖掘主题社区的意见领袖。这同时兼顾了动态特征以及用户对话题的情感倾向，可以更加准确地挖掘具有较大影响力的节点。

(2) 深入研究了基于拓扑结构的最大影响力节点挖掘技术：本项目的主要工作是针对现有影响最大化算法所存在的问题，从网络的拓扑结构出发，分析节点之间的关联关系且对节点影响力建模，并提出了 CCA 和 MHG 算法，CCA 启发式算法利用核数高的影响力节点又使它们保持一定的距离，从而减少重叠激活的可能性；MHG 算法通过把节点边际效益的计算限定在某些最具有潜力的候选节点中，因此能比贪心算法取得多个数量级的加速比和匹配的影响范围。

(3) 深入研究了基于节点主题属性的最大影响力节点挖掘技术：本项目将病毒式营销手段和窄告这种个性化的广告模式结合起来，提出了基于主题属性的影响最大化问题。基于抓取的大量真实数据，考虑节点间交互强度和节点主题属性与商品目标受众属性间相似度这两个因素，对节点间影响概率建模，并提出一种基于概率和的算法 SoPCA 和两种基于局部社区的启发式算法 CBA 和 ICBA。

(4) 深入研究了基于成本效益的最大影响力节点挖掘技术：本项目针对现有研究的一些不足，提出了基于成本效益的影响最大化问题，把选择节点所需要花费的成本考虑进来，目的是在推广预算有限的情况下选择适当的初始受众节点，达到尽可能广的传播范围，并据此提出了高效的影响力节点选择算法 ProbCover。为了避免节点影响力重叠，又提出了一种惰性节点选择算法 ProbCoverLF，从而进一步提高了时间效率和影响效果。

(5) 深入研究了基于成本的广告投放机制：本项目以社会计算中原始影响最大化问题为基础，根据社交网络本身的特性和其空间结构分布特征提出社交网络中基于成本的广告投放问题，它同时考虑了节点的成本与节点主题属性。为了在

社交网络中投放广告时可以进一步缩小广告受众的范围,使得投放结果更精准、深入,设计了一种基于平均概率的种子节点选择算法 AvePA,在广告预算有限的情况下选择满足条件的影响力节点,使得最终可以达到尽可能好的广告效应。

(6) 东南大学和苏州米谷网络科技有限公司结合社交网络的特点,基于本课题的理论研究成果,设计并实现了社交网络中影响力最大化节点挖掘系统。该系统主要包括数据输入与处理模块、算法实现模块、传播模拟模块和结果可视化模块。这样可以将整个课题的研究成果更直观地呈现给用户,使得用户能更方便地使用它,同时也可以将复杂数据转换成图形或图像在屏幕上显示出来,并与用户进行交互。

(7) 项目共申请发明专利 3 项;软件著作权 1 项;在国内外核心期刊和会议上发表论文 6 篇, EI 收录论文 6 篇。

(8) 项目研究过程中,培养博士研究生 1 名(周涛),硕士研究生 7 名(徐顺、董丹、吴江林、赵钊、王浩然、冯雪艳、袁堂飞)。

2. 申请的国家发明专利

- [1] 曹玖新, 薛凌云, 赵钊, 刘波. 一种基于成本效益的影响最大化方法. 申请号: 201610319986.8, 申请日: 2016-05-13.
- [2] 曹玖新, 马卓, 陈巧云, 刘波, 周涛. 一种基于 LDA 主题模型的 OSN 社区发现方法. 申请号: 2015106114551, 申请日: 2015-09-23.
- [3] 曹玖新, 马卓, 王瑶, 刘波, 陈高君. 一种基于随机森林和用户关系的 OSN 用户情感分析方法. 申请号: 2015106114104 申请日: 2015-09-23.

3. 软件著作权登记证书

曹玖新, 马卓等. 在线社交网络特定用户观察与分析软件. 登记号: 2016SR033494, 登记日期: 2016-02-18.

4. 论文发表

- [1] Jiuxin Cao, Tao Zhou, Dan Dong, Shuai Xu, Ziqing Zhu, Zhuo Ma, and Bo Liu. Attribute-based Influence Maximization in Social Networks. WISE 2016, Part I, LNCS 10041:3-18.
- [2] 曹玖新, 闵绘宇, 徐顺, 刘波. 基于启发式和贪心策略的社交网络影响最大化算法. 东南大学学报(自然科学版), 2016, 46(5): 950-956.
- [3] 曹玖新, 陈高君, 吴江林, 刘波, 周涛, 胥帅, 朱子青. 基于多维特征分析的社交网络意见领袖挖掘. 电子学报, 2016 年 4 月, 44(4): 898-905.
- [4] Tao Zhou, Jiuxin Cao, Bo Liu, Shuai Xu, Ziqing Zhu, Junzhou Luo. Location-Based Influence Maximization in Social Networks. Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM'15). October 19-23, 2015, Melbourne, VIC, Australia:

1211-1220. (EI: 20160801984087).

[5] 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, 28(2): 238-248(EI: 20151200650587).

[6] 曹玖新, 吴江林, 石伟, 刘波, 郑啸, 罗军舟. 新浪微博网信息传播分析与预测. 计算机学报, 2014, 37(4): 779-790.

5. 人才培养

在人才培养方面, 本项目共培养研究生 8 名, 其中博士生 1 名、硕士生 7 名(5 人已毕业)。未毕业研究生正对本项目进行后续的研究, 以期对理论和技术进行完善, 对原型进行扩展。

项目培养的博、硕士研究生

姓名	学位类型	论文题目	备注
周涛	博士	异质社交网络信息融合关键技术与应用	预计答辩时间 2018 年 6 月
徐顺	硕士	社交网络中影响力节点选择机制研究	已毕业 2013 年 6 月
董丹	硕士	社交网络中基于属性的影响力节点挖掘	已毕业 2014 年 6 月
吴江林	硕士	社交网络中基于话题的意见领袖挖掘研究	已毕业 2014 年 6 月
赵钊	硕士	基于成本效益的影响最大化算法分析与设计	已毕业 2015 年 6 月
王浩然	硕士	社交网络中基于竞争的影响力最大化研究	已毕业 2016 年 6 月
冯雪艳	硕士	社交网络中基于成本的广告投放策略的设计与实现	预计答辩时间 2017 年 5 月
袁堂飞	硕士	社交网络中基于主题的限时影响最大化算法与系统实现	预计答辩时间 2018 年 5 月

五、项目绩效分析

项目的顺利实施不仅可以有效提高企业自身的发展速度, 如苏州米谷网络科技有限公司的 B2C、C2C 虚拟交易服务, 积极推动商业模式与信息技术的联合创新, 实现产业技术改造及企业业务在互联网上的推广, 极大促进我省产业健康、快速、稳定地发展; 同时通过加大项目成果产业化、商业化和规模化应用力度,

使得生产经营管理方式向网络化、数字化、集约化方向发展，有效促进新兴产业结构调整和优化升级，加速形成广告联盟推广产业集群；有助于提高我省市场经济活动效率，有利于中小企业的健康快速成长，为我省长期的经济活力和产业竞争力奠定良好的基础。

六、存在问题、有关建议及下一步研究设想

1. 存在的不足

本项目得到江苏省科技厅的资助，但由于研究时间较紧，项目研究内容多，涉及面较广。虽然在项目承担单位的共同努力下，取得了较大进展，但基础理论研究深度还需进一步提升。

2. 下一步研究设想

本项目组围绕社交网络中影响最大化问题进行了长期的研究工作，在同苏州米谷网络科技有限公司合作研发期间，我们在相关领域进行了一些思考和开创性的预研工作。结合社交网络的相关特性与实际应用，我们在未来阶段将会开展以下方向的研究：

(1) 在未来的研究工作中，可以综合标签、用户的性格特性以及价值观，利用数据挖掘技术来获取用户的兴趣爱好，使广告投放效果更精确，同时，也应该考虑到用户兴趣的动态变化这一特性。

(2) 在更多复杂的市场营销行为中，其中活跃的节点在后续的步骤中可能变成不活跃的节点，并且活跃的节点也可能在能接受的范围内多次激活不活跃的节点。所以，在以后的影响最大化问题研究中，影响力传播模型建模时可以将这一点考虑进去。

(3) 注意到社交网络中一则广告的寿命是很短的，根据 WebTrends 的调查研究，用户对一则社交网络广告的兴趣在三到五天后就将消退，并且在传播过程中，会存在传播时延，所以，在未来的社交网络广告投放研究中，可以将时间因素考虑进去，将其作为一个限制因素，使得最终研究成果的通用性更好。