**面向社会网络的新型图搜索和推荐**

随着互联网技术的迅猛发展，非结构化数据占有的比例大幅增加。据美国Forrester分析师在2010年"政府今天所面临的挑战" 报告中估计："数据将会在今后的5年内增加800%；其中非结构化数据在各组织机构的数据中所占份额超过70%到80%，并且这些非结构化数据是结构化数据增长速度的10到50倍"。近年来，研究机构和工业界对图的研究和应用越来越重视。其主要原因是相对于传统的关系模型和XML，图的表达能力丰富，有着广泛的应用，如社交网络和Web网络等。相对于传统的关系数据库和XML，图的表达能力更加丰富，因此图在各个领域中的应用非常广泛。

如社交网络和Web网络的飞速发展，对社会和个人的行为产生了深远的影响。在社交网络中，用户可以看作为图的顶点，用户之间的关系（如朋友等）可以看作为图的边。这样图在社交网络中有着重要的应用，如近邻查询和图压缩等。和社交网络类似，Web网络中的网页可以看作图的顶点，网页之间的链接关系可以看作图的边，因此图在Web网络中也有着广泛的应用，如网页的聚类可以看作图的分类问题，镜像站点检测问题可以看作图的匹配问题等。此外，在公路网络和智能交通等方面图也起到了重要的应用。

目前国内外知名研究院、大学和公司对图数据的相关研究与应用非常重视，如：Microsoft研究院基于云计算Orleans平台和Trinity项目，研究用于数据中心的"Querying Large Distributed Graphs"项目；Google用于处理大图的系统Pregel和Knowledge Graph；Yahoo! 研究院的"Graph Partitioning"项目；Neo4j公司的开源图数据库；Facebook公司的“Graph Search”；UCSB 大学的"Massive Graphs in Clusters"项目；爱丁堡大学的模式匹配和图匹配项目。除了Facebook公司，其他的要么是图数据的管理系统，要么是单独的图查询研究，缺少系统的研究和分析，而Facebook也只是提供了简单的图搜索功能，缺乏系统深入的研究。北京航空航天大学团队也在从事图数据的相关研究，给出了图搜索的统一定义，分析了图搜索的重要性，并研制开了面向社会网络的大图搜索和推荐的一些基础技术。

目前图搜索和推荐在工业界和学术界受到广泛关注，并且图数据在工业界的应用和在学术界的研究是相互促进的。实际上，图搜索和推荐的理论和技术是近年来工业界和学术界共同关注的热点之一。在泛在物联网环境中，数据结构复杂，需要对其中的数据进行关联分析，很多问题都需要研究。面对大规模、动态和不确定的网络环境，（1）如何抽象其中的关系图?（2）在大量、动态和不确定数据中，如何搜索“信息”更快，如何搜索“信息”更准?（3）以及如何通过“知识”增强搜索？（4）通过结合推荐，来提高用户满意度**？**因此，开展图搜索和推荐有着重要的研究意义和应用价值。

**2.工作基础**

**研究基础：** 实验室始终围绕国家战略和信息科学的前沿方向，积极争取和承担国家关键项目，开展原创性的基础研究和创新性的技术突破，服务于国家重大专项和国民经济建设的主战场。实验室的主要研究方向是：软件基础理论与方法论、网络海量信息的物理学特征和复杂性、海量非结构化数据模型与管理系统、物联网软件支撑技术与系统、面向服务的软件开发与运行环境、多媒体协同网络环境与群体软件工程。2007至2011年期间，实验室承担国家973计划、863计划、国家自然科学基金、国防科技计划等各类国家和省部级科研项目共计89项，累计科研经费超过2.2亿元。在上述科研项目的支持下，实验室在一系列基础理论、新技术上取得了创新和突破，开发完成了一批应用软件平台，为国家的信息产业发展的战略需求服务。实验室获得国家科技进步二等奖1项，国家科技发明二等奖1项，国家教学成果二等奖1项，北京市科学技术一等奖1项，北京市科学技术二等奖1项，北京市优秀教学成果二等奖1项；发表学术论文583篇，英文专著1部，中文专著6部，申请国家发明专利393项，已获得授权99项。

近年来，在973计划、863计划和国家自然科学基金等资助下，北航团队开展了面向社会网络的新型图搜索和推荐研究，提出了新型的图搜索查询语言，并研制开发了高效大图搜索和推荐的技术。相关论文发表在SIGMOD, VLDB, ICDE, PODS, IJCAI, TODS, SICOMP, JCSS, VLDB Journal等领域知名会议期刊上，共计30余篇，并获得2010数据库领域国际顶级会议VLDB唯一最佳论文和2013国际知名会议WISE最佳挑战论文奖。

**相关项目：**

1）973计划“网络信息空间大数据计算理论”，2014~2018

2 ) 科技部支撑计划“面向XXX的技术”，2012.10— 2015.10

3 ) 863计划“以支撑电子商务为主的网络操作系统”, 2011.01— 2013.12

4 ) 自然科学基金重大“面向车联网的可信网络系统”， 2012.01— 2015.12

5) 自然科学优秀青年基金“面向社会网络的新型图搜索”， 2014.01— 2016.12

6) 自然科学青年基金“面向社会网络用户需求的推荐系统研究”， 2014.01— 2016.12

7) 教育部留学回国人员科研启动基金“大规模海量图查询技术”，2014/01－2014/12

**依托单位**。依托单位北京航空航天大学计算机学院是计算机科学与技术一级学科、计算机软件与理论国家重点学科单位，拥有国家软件开发环境重点实验室、可信网络计算技术国防重点学科实验室、教育部虚拟现实重点实验室和北京市计算机新技术重点实验室，是计算机理论研究和软件技术的重要研究基地。相关实验室针对国家级科研项目及中青年学术骨干科研工作制定了详细的合作研究制度和政策，能够提供良好的硬件环境和学术氛围，从而有效保证课题研究工作的顺利开展。

**计算机科学与工程创新引智基地：**于2011获批立项建设，依托软件开发环境国家重点实验室、虚拟现实技术与系统国家重点实验室等科研平台，汇聚19位计算机科学与工程领域知名专家（含2名院士）以及来自英国爱丁堡大学、美国普渡大学、宾西法尼亚大学等世界一流高校的15位计算机领域内著名专家（含7名院士），通过引智基地建设，形成以研究生联合培养、人员互访、联合系统开发、联合承担国际相关科研课题、联合举办国际会议等多种形式，着重在计算机基础理论与技术、互联网服务与大规模海量数据管理技术等方向开展高水平国际交流合作研究，推进计算机科学与工程学科建设与发展。

**北航大数据“科学与工程”国际研究中心**：北京航空航天大学计算学院、爱丁堡大学信息学院、香港科技大学计算机系、宾夕法尼亚大学和百度公司联合创建“大数据科学与工程”国际研究中心，旨在以当前互联网和大数据时代新型信息技术为牵引，创造新的学术领域和应用增长点。通过著名高校和互联网企业的强强联合、优势互补，以世界一流为目标，在互联网软件技术和大数据处理技术领域开展广泛的科研合作，整合利用国际优秀计算机教育和企业资源，逐步成为引领“大数据科学与工程”领域科技发展、具有重大国际影响的学术高地。同时，面向“云计算”、“现代服务业”等国家战略性新兴产业发展，成为支撑和引领互联网和数据服务产业技术进步的研发基地。此外，通过学生交换、师资交流、课程开设和联合开展国际研究项目等方式，进一步提升各高校在世界计算机科学领域的地位和影响力，通过融汇世界一流研究人员，贯通计算机科学理论和工程能力，在互联网环境的海量数据处理、软件基础理论、可信分布式计算等方向建立领先研究团队，产生一批具有创新性和应用价值的计算机理论和技术成果，培养具有进取、创新、团结精神和领军领导才能、具有国际视野的高层次研究队伍，为推动信息技术持续服务于社会做出贡献。

**软件开发环境国家重点实验室：**于1988年通过国家计委专家认证，成为用世界银行贷款建设的国家重点实验室，1995年通过教育部和国家自然基金委组织的验收。实验室是在先进网络环境下软件理论、关键技术和开发环境等方面，开展基础、应用基础与前沿高技术研究的开放式科研基地；是国家重大需求和社会重大应用的服务支撑基地；是创新型高层次人才的培养基地。