**面向社会网络的图搜索**

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，报告将讨论如何从社交网络等图结构数据中准确、快速搜索信息。

**参考文献：**

1. Bin Shao, Haixun Wang, and Yatao Li. Trinity: A Distributed Graph Engine on a Memory Cloud. In SIGMOD 2013.

2. Shuai Ma, Yang Cao, Wenfei Fan, Jinpeng Huai, and Tianyu Wo, Strong Simulation: Capturing Topology in Graph Pattern Matching.

ACM Transactions on Database Systems (TODS), 39, 1, Article 4, 46 pages, January 2014.

3. Xin Luna Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Kevin Murphy, Shaohua Sun, Wei Zhang: From Data Fusion to Knowledge Fusion. PVLDB 7(10): 881-892 (2014)

4. Wenfei Fan and Jinpeng Huai, Querying Big Data: Bridging Theory and Practice, Journal of Computer Science and Technology, 2014 (invited).

5. Wenfei Fan, Xin Wang, and Yinghui Wu, Querying Big Graphs within Bounded Resources, ACM SIGMOD Conference on Management of Data (SIGMOD), 2014.