中国计算机学会学术著作丛书

大规模图数据的分布式处理

谷峪 于戈 鲍玉斌 著

清华大学出版社

中国计算机学会学术著作丛书

大规模图数据的分布式处理

谷峪 于戈 鲍玉斌 著

清华大学出版社
北京

内 容 简 介

随着以社交网络为代表的图数据规模高速增长，复杂的查询需求不断涌现，处理这类大规模数据有许多理论问题需要解决。本书结合作者多年的研究积累，系统地介绍了大图分布式处理中基础的数据划分、组织和消息管理技术，以及三角形、最大 k 边连通子图、最小生成树、频繁子图、重叠社区发现等大图查询和分析算法的优化，并对系统实现技术进行了探讨。

本书适合高等院校计算机专业的教师、学生及计算机应用系统的研发人员阅读参考。

图书在版编目(CIP)数据

大规模图数据的分布式处理/谷峪，于戈，鲍玉斌著. --北京：清华大学出版社，2015
(中国计算机学会学术著作丛书)
ISBN 978-7-302-42072-9

Ⅰ. ①大… Ⅱ. ①谷… ②于… ③鲍… Ⅲ. ①分布式数据处理－研究 Ⅳ. ①TP274

中国版本图书馆 CIP 数据核字(2015)第 263558 号

责任编辑：薛　慧
封面设计：傅瑞学
责任校对：刘玉霞
责任印制：沈　露

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　邮　　编：100084
社 总 机：010-62770175　　邮　　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：北京密云胶印厂
经　　销：全国新华书店
开　　本：175mm×245mm　　**印　　张**：15.75　　**字　　数**：306 千字
版　　次：2015 年 12 月第 1 版　　**印　　次**：2015 年 12 月第1 次印刷
印　　数：1～1500
定　　价：58.00 元

产品编号：066763-01

清华大学出版社计算机学术著作出版基金

评审委员会

中国计算机学会学术著作丛书

名誉主任委员：张效祥

主任委员：唐泽圣

副主任委员：陆汝钤

委　　员：（按姓氏笔画为序）

王　珊　吕　建　李晓明

林惠民　罗军舟　郑纬民

施伯乐　焦金生　谭铁牛

丛书序

第一台电子计算机诞生于20世纪40年代。到目前为止，计算机的发展已远远超出了其创始者的想象。计算机的处理能力越来越强，应用面越来越广，应用领域也从单纯的科学计算渗透到社会生活的方方面面：从工业、国防、医疗、教育、娱乐直至人们的日常生活，计算机的影响可谓无处不在。

计算机之所以能取得上述地位并成为全球最具活力的产业，原因在于其高速的计算能力、庞大的存储能力以及友好、灵活的用户界面。而这些新技术及其应用有赖于研究人员多年不懈的努力。学术研究是应用研究的基础，也是技术发展的动力。

自1992年起，清华大学出版社与广西科学技术出版社为促进我国计算机科学技术与产业的发展，推动计算机科技著作的出版，设立了“计算机学术著作出版基金”，并将资助出版的著作列为中国计算机学会的学术著作丛书。时至今日，本套丛书已出版学术专著近50种，产生了很好的社会影响，有的专著具有很高的学术水平，有的则奠定了一类学术研究的基础。中国计算机学会一直将学术著作的出版作为学会的一项主要工作。本届理事会将秉承这一传统，继续大力支持本套丛书的出版，鼓励科技工作者写出更多的优秀学术著作，多出好书，多出精品，为提高我国的知识创新和技术创新能力，促进计算机科学技术的发展和进步作出更大的贡献。

中国计算机学会

2002年6月14日

前言

数据能够有效地反映数据之间普遍存在的联系，具有丰富的表达力，在 Web、社会网络、生物和化学数据库等领域获得了广泛的应用。随着数据获取方式的多样化，图数据规模越来越大，应用也日趋复杂，传统的集中式图查询处理和分析挖掘方法满足不了日益增长的功能和性能上的需求。特别是近年来随着云计算和大数据等概念的兴起，分布式图处理计算也随之得到快速的发展，成为热点的研究领域。本专著系统综述了目前该领域的主要研究进展，并总结和整理了作者近年来在这方面的研究成果，内容囊括大规模图数据分布式处理的主要模型、技术和系统，包括执行机制、数据组织、代表性算法，以及系统实现和典型应用等各个方面。本书试图为读者系统地展现大数据技术高速发展和变革时代大图处理区别于传统数据管理和分布式计算的新技术、新思想、新系统和新挑战。

本书共分为 10 章，第 1 章主要介绍大规模图数据分布式处理的研究背景和问题；第 2 章介绍分布式图计算模型和执行机制；第 3 章和第 4 章分别介绍基础的数据组织问题，包括数据的划分以及存储和索引；第 5 章到第 9 章介绍代表性的大图复杂查询、分析和挖掘算法及其分布式实现技术，包括三角形查询、最大k边连通子图查询、最小生成树搜索、频繁子图挖掘和重叠社区发现；第 10 章对现有的主要分布式大图处理系统和典型应用进行综述。

本书涉及的研究课题得到国家重点基础研究发展计划（“973 计划”）项目（No. 2012CB316201）、国家自然科学基金项目（61472071、61272179、61433008）、教育部-中国移

动科研基金项目(MCM20125021)等资助。

作者指导的部分研究生参与了本书的撰写和相关课题的研发,他们是王志刚、刘金鹏、王文安、杨佳学、张天明、张楠、毕亚辉等,他们为本书付出了辛勤的劳动,在此一并表示衷心的感谢。

该专著主要作为从事图数据管理、分布式计算和大数据分析等相关领域研究开发和管理人员的参考书籍,也可作为高校计算机和大数据等相关专业研究生的补充教材和参考读物。

由于著者水平所限,而本书涉及很多新的技术,因此书中难免有疏漏和错误,恳请读者提出宝贵意见。

作 者

2015 年 5 月 7 日

目录

第 1 章
大规模图数据处理：问题与挑战

1.1 大图数据处理的背景

随着云计算等新技术的快速发展、社交网络等新型互联网应用的兴起和各种电子设备的日益普及，人类获取和存储数据的规模正以前所未有的速度爆炸式增长，与大数据相关的技术变革成为学术界和工业界的热点问题。《中华人民共和国国民经济和社会发展第十二个五年规划纲要》提出要重点研究海量信息处理及知识挖掘的理论和方法，从国家战略层面上强调了对大数据的研究。而大图是大数据研究领域的一个重要分支，具有广泛的理论研究和应用价值。作为计算机科学中最常用的一类抽象数据结构，图能够有效地表达对象之间广泛存在的联系，在结构和语义方面比线性表和树更为复杂，更具有一般性表示能力。

现实世界中的许多应用场景都需要用图结构表示，与图相关的处理和应用几乎无所不在。传统应用如最优运输路线的确定、疾病暴发路径的预测、科技文献的引用关系等；新兴应用如道路交通管理、社交网络分析、语义 Web 分析、生物信息网络分析等。如在蛋白质交互网络中，图的顶点对应着蛋白质，边对应着蛋白质之间的联系；再如化合物的分子结构就可以被抽象为无向标号图，其中图的顶点对应着原子而边则可以表示成原子间的化学键；在社交网络中，图的顶点表示具体的用户，边表示用户之间的好友或者关注关系，相关的属性可以记录在对应的点和边上。此外，信息科学中的资源描述框架（RDF）文件、XML 文件、文本检索，以及机械工程中技术图纸的基本对象建模和通信网

络中集成电路的布局布线等领域也都大量使用了图数据。针对这些快速增长、形式多样且语义丰富的图数据，如何开发有效并且高效的查询分析技术，成为具有重要应用价值的课题。

传统的图数据库和图数据管理技术通常针对彼此独立的“小图”分别进行处理，尽管图的数目可能较多，但通常不需要复杂的迭代过程，也不会产生大量的消息，算法的时间和空间开销一般较低。然而，近年来，随着互联网的普及和Web 2.0技术的推动，以及生物化学等科学数据采集手段的丰富，产生了众多规模巨大且结构复杂的“大图”甚至“超大图”。以互联网和社交网络为例，据报道，Google索引的网页数目早已超过了1万亿(10^{12})幅，Facebook 2012年的活跃用户已经超过10亿。特别地，这种发展的势头非常迅猛，据CNNIC统计，2010年中国网页规模就已经达到600亿，年增长率78.6%，国内如微信、微博等社会媒体的发展也异常迅猛。而在生物信息学领域，人脑级别的图建模已经达到了10^{14}的规模。

真实世界中实体规模的扩张，导致相应图模型的数据规模迅速增长，动辄有数十亿个顶点和上万亿条边。以搜索引擎中常用的PageRank计算为例[1]，网页用图顶点表示，网页之间的链接关系用有向边表示，一个网页的PageRank得分根据网页之间相互的超链接关系计算而得到。假设按邻接表形式存储100亿个图顶点和600亿条边，每个顶点及出度边的存储空间占100字节，那么整个图的存储空间将超过1TB。值得注意的是，庞大的顶点和边数目构成的结构信息只是大图数据规模惊人的冰山一角，复杂应用中的图数据为了表达复杂的语义，因此在顶点和边上往往会附带各类属性信息，这些属性信息内容丰富，需要大量的空间开销。此外，除了静态的结构和属性信息，相比于基于属性的简单查询和搜索，大图上的统计分析算法往往需要基于图的结构进行循环和递归操作，直至达到收敛条件，因此需要频繁地处理并行迭代过程中由于通信交互产生的消息数据等中间结果。面对如此大规模的静态和动态数据，对其存储、索引、查找和分析等处理的时间开销和空间开销远远超出了传统集中式图数据管理的承受能力。对大规模图数据的高效管理和计算，已经成为急需解决的问题，也是一项极具挑战性的工作。

1.2　图数据的表示

作为数学的一个重要分支，图论以图作为研究对象，在简单图的基础上衍生出超图理论、极图理论、拓扑图论等，使图可以从多方面表达现实世界。当前大规模图数据管理，采用的数据模型有多种，按照图中节点的复杂程度分为简单节点图模

型和复杂节点图模型，按照一条边可以连接的顶点数目分为简单图模型和超图模型。不论是简单图模型、超图模型、简单节点模型还是复杂节点模型，它们的顶点和边都可以带有属性。

1. 简单图模型

这里所说的简单图，并不是图论中的简单图，是相对于超图而言的，一条边只能连接两个顶点，可以存在环路。其形式化表示，即 $G=(V, E)$。其中，$V=\{v_1, v_2, \cdots, v_n\}$，$E=\{e_1, e_2, \cdots, e_n\}=\{\{v_1, v_2\}, \{v_3, v_4\}, \cdots, \{v_{n-1}, v_n\}\}$。简单图的存储和处理都比较容易，对于一般的应用，简单图的表达能力完全可以胜任，如 PageRank 计算、最短路径查询等。简单图模型的常用组织存储结构包括邻接矩阵、邻接表、十字链表和邻接多重表等多种方式。不同的系统根据目标不同采用不同的表示方式。

2. 超图模型

一条边可以连接任意数目的图顶点。此模型中图的边称为超边。基于这种特点，超图比上述简单图的适用性更强，保留的信息更多。形式化表示为 $G=(V, E)$，其中，$V=\{v_1, v_2, \cdots, v_n\}$，$E=\{e_1, e_2, \cdots, e_n\}=\{\{v_1, v_2, v_3\}, \{v_3, v_4, v_5\}, \cdots, \{v_{n-2}, v_{n-1}, v_n\}\}$。例如，以图顶点代表文章，每条边代表两个顶点（文章）享有同一个作者。现有三篇文章 v_1（作者 A、B），v_2（作者 A、C），v_3（作者 A、D），三篇文章的作者都有 A。图 1-1 的左图表示了简单图存储模式，边集 $E=\{e_1, e_2, e_3\}=\{\{v_1, v_2\}, \{v_1, v_3\}, \{v_2, v_3\}\}$，无法直接保留作者 A 同时是三篇文章 v_1, v_2, v_3 的作者这一信息。图 1-1 的右图代表了超图存储模式，边集 $E=\{e_1\}=\{\{v_1, v_2, v_3\}\}$，超边 e_1 中直接保留了 A 是三篇文章 v_1, v_2, v_3 的作者这一信息。对于具有复杂联系的应用，可以使用超图模型建模，例如社交网络、生物医学网络等。

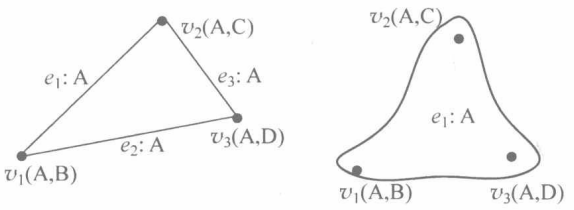
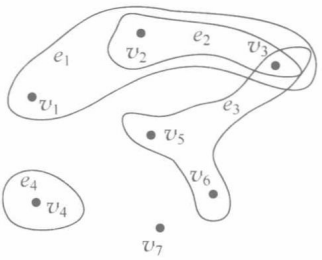


图 1-1 简单图（左）和超图（右）

超图模型的组织方式主要使用关系矩阵，它与邻接矩阵较为相似。图 1-2 展示了超图的关系矩阵表示方法。与邻接矩阵不同的是，关系矩阵的行和列分别表示图顶点编号和超边的编号，关系矩阵中，1 表示一条超边包含某个图顶点，或一个图顶点隶属于某条超边。



	e_1	e_2	e_3	e_4
v_1	1	0	0	0
v_2	1	1	0	0
v_3	1	1	1	0
v_4	0	0	0	1
v_5	0	0	1	0
v_6	0	0	1	0
v_7	0	0	0	0

图 1-2　超图的关系矩阵表示方法

1.3　传统的大图数据管理方法

传统的分布式图处理技术主要基于 NoSQL 数据库，关注的是图数据、特别是小图数据集的管理问题，查询相对简单，具有高效的索引，可以支持数据的更新。当图数据更新时，需要解决在分布式环境下的一致性控制问题，提供事务功能。

特别地，NoSQL 数据库采用的数据模型主要有文档存储(Document Store)模型、列族存储(ColumnFamily Store)模型、Key-Value 存储模型、图形模型等几大类[2]。

Key-Value 存储模型的存储模式简单，在高并发环境下可以提供高效的查询或遍历服务，能存储海量数据，非常适合通过主键进行查询或遍历，但对复杂的条件查询支持度不佳。从图处理的角度看，像 PageRank 计算等并不需要复杂查询，Key-Value 模型完全可以胜任。对于采用邻接表组织的图数据，可以将图顶点及其值作为 Key，将出边或出度顶点列表作为 Value。文献[3]结合语义 Web 和传统的 Key-Value 模型，提出 Key-Key-Value 模型。以社交网络为例，Key-Key-Value 模型将 Alice 和 Bob 之间的好友关系组织为一个三元组〈Alice，Bob，FriendShip〉。该模型存储的信息比传统的 Key-Value 模型更加丰富，可以据此进行数据迁移和合并，以提高时空局部性，使得在查询处理时能减少数据远程跨机读取的次数，因而可以提高数据读取效率。文档存储模型在存储格式方面十分灵活，比较适合存储系统日志等非结构化数据，对以邻接矩阵或邻接表组织的图数据来讲，意义不大。而且文档存储模型为支持灵活性所导致的处理效率的降低也会成为大规模图数据管理的瓶颈。列族存储模型比较适合对某一列进行随机查询处理，但是对于穷举式遍历，反而不如传统的面向行的存储模式。图模型的相关研究目前还不完善，只有少数分布式图数据库，如 Neo4j[4] 等采用这种模型存储图数据。文献[5]从管理数据的规模和模型的复杂性两个维度比较了这 4 种基本存储模型，见图 1-3。

在查询处理方面，传统的分布式图数据库和一些新型的大图在线查询系统往往只支持图的简单查询检索，返回用户感兴趣的信息，而通常不支持大图上的复杂