On Uncertain Graphs

Synthesis Lectures on Data Management

Editor
H.V. Jagadish, ***University of Michigan***

Founding Editor
M. Tamer Özsu, ***University of Waterloo***

Synthesis Lectures on Data Management is edited by H.V. Jagadish of the University of Michigan. The series publishes 80–150 page publications on topics pertaining to data management. Topics include query languages, database system architectures, transaction management, data warehousing, XML and databases, data stream systems, wide scale data distribution, multimedia data management, data mining, and related subjects.

On Uncertain Graphs
Arijit Khan, Yuan Ye, and Lei Chen
2018

Answering Queries Using Views
Foto Afrati and Rada Chirkova
2017

Databases on Modern Hardware: How to Stop Underutilization and Love Multicores
Anatasia Ailamaki, Erieta Liarou, Pınar Tözün, Danica Porobic, and Iraklis Psaroudakis
2017

Instant Recovery with Write-Ahead Logging: Page Repair, System Restart, Media Restore, and System Failover, Second Edition
Goetz Graefe, Wey Guy, and Caetano Sauer
2016

Generating Plans from Proofs: The Interpolation-based Approach to Query Reformulation
Michael Benedikt, Julien Leblay, Balder ten Cate, and Efthymia Tsamoura
2016

Veracity of Data: From Truth Discovery Computation Algorithms to Models of Misinformation Dynamics
Laure Berti-Équille and Javier Borge-Holthoefer
2015

Datalog and Logic Databases
Sergio Greco and Cristina Molinaro
2015

Big Data Integration
Xin Luna Dong and Divesh Srivastava
2015

Instant Recovery with Write-Ahead Logging: Page Repair, System Restart, and Media Restore
Goetz Graefe, Wey Guy, and Caetano Sauer
2014

Similarity Joins in Relational Database Systems
Nikolaus Augsten and Michael H. Böhlen
2013

Information and Influence Propagation in Social Networks
Wei Chen, Laks V.S. Lakshmanan, and Carlos Castillo
2013

Data Cleaning: A Practical Perspective
Venkatesh Ganti and Anish Das Sarma
2013

Data Processing on FPGAs
Jens Teubner and Louis Woods
2013

Perspectives on Business Intelligence
Raymond T. Ng, Patricia C. Arocena, Denilson Barbosa, Giuseppe Carenini, Luiz Gomes, Jr., Stephan Jou, Rock Anthony Leung, Evangelos Milios, Renée J. Miller, John Mylopoulos, Rachel A. Pottinger, Frank Tompa, and Eric Yu
2013

Semantics Empowered Web 3.0: Managing Enterprise, Social, Sensor, and Cloud-based Data and Services for Advanced Applications
Amit Sheth and Krishnaprasad Thirunarayan
2012

Data Management in the Cloud: Challenges and Opportunities
Divyakant Agrawal, Sudipto Das, and Amr El Abbadi
2012

Query Processing over Uncertain Databases
Lei Chen and Xiang Lian
2012

Foundations of Data Quality Management
Wenfei Fan and Floris Geerts
2012

Incomplete Data and Data Dependencies in Relational Databases
Sergio Greco, Cristian Molinaro, and Francesca Spezzano
2012

Business Processes: A Database Perspective
Daniel Deutch and Tova Milo
2012

Data Protection from Insider Threats
Elisa Bertino
2012

Deep Web Query Interface Understanding and Integration
Eduard C. Dragut, Weiyi Meng, and Clement T. Yu
2012

P2P Techniques for Decentralized Applications
Esther Pacitti, Reza Akbarinia, and Manal El-Dick
2012

Query Answer Authentication
HweeHwa Pang and Kian-Lee Tan
2012

Declarative Networking
Boon Thau Loo and Wenchao Zhou
2012

Full-Text (Substring) Indexes in External Memory
Marina Barsky, Ulrike Stege, and Alex Thomo
2011

Spatial Data Management
Nikos Mamoulis
2011

Database Repairing and Consistent Query Answering
Leopoldo Bertossi
2011

Managing Event Information: Modeling, Retrieval, and Applications
Amarnath Gupta and Ramesh Jain
2011

Fundamentals of Physical Design and Query Compilation
David Toman and Grant Weddell
2011

Methods for Mining and Summarizing Text Conversations
Giuseppe Carenini, Gabriel Murray, and Raymond Ng
2011

Probabilistic Databases
Dan Suciu, Dan Olteanu, Christopher Ré, and Christoph Koch
2011

Peer-to-Peer Data Management
Karl Aberer
2011

Probabilistic Ranking Techniques in Relational Databases
Ihab F. Ilyas and Mohamed A. Soliman
2011

Uncertain Schema Matching
Avigdor Gal
2011

Fundamentals of Object Databases: Object-Oriented and Object-Relational Design
Suzanne W. Dietrich and Susan D. Urban
2010

Advanced Metasearch Engine Technology
Weiyi Meng and Clement T. Yu
2010

Web Page Recommendation Models: Theory and Algorithms
Sule Gündüz-Ögüdücü
2010

Multidimensional Databases and Data Warehousing
Christian S. Jensen, Torben Bach Pedersen, and Christian Thomsen
2010

Database Replication
Bettina Kemme, Ricardo Jimenez-Peris, and Marta Patino-Martinez
2010

Relational and XML Data Exchange
Marcelo Arenas, Pablo Barcelo, Leonid Libkin, and Filip Murlak
2010

User-Centered Data Management
Tiziana Catarci, Alan Dix, Stephen Kimani, and Giuseppe Santucci
2010

Data Stream Management
Lukasz Golab and M. Tamer Özsu
2010

Access Control in Data Management Systems
Elena Ferrari
2010

An Introduction to Duplicate Detection
Felix Naumann and Melanie Herschel
2010

Privacy-Preserving Data Publishing: An Overview
Raymond Chi-Wing Wong and Ada Wai-Chee Fu
2010

Keyword Search in Databases
Jeffrey Xu Yu, Lu Qin, and Lijun Chang
2009

© Springer Nature Switzerland AG 2022

Reprint of original edition © Morgan & Claypool 2018

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

On Uncertain Graphs

Arijit Khan, Yuan Ye, and Lei Chen

ISBN: 978-3-031-00732-3 paperback
ISBN: 978-3-031-01860-2 ebook
ISBN: 978-3-031-00087-4 hardcover

DOI 10.1007/978-3-031-01860-2

A Publication in the Springer series

SYNTHESIS LECTURES ON DATA MANAGEMENT

Lecture #48
Series Editor: H.V. Jagadish, *University of Michigan*
Founding Editor: M. Tamer Özsu, *University of Waterloo*
Series ISSN
Print 2153-5418 Electronic 2153-5426

On Uncertain Graphs

Arijit Khan
Nanyang Technological University, Singapore

Yuan Ye
Northeastern University, China

Lei Chen
Hong Kong University of Science and Technology, Hong Kong

SYNTHESIS LECTURES ON DATA MANAGEMENT #48

ABSTRACT

Large-scale, highly interconnected networks, which are often modeled as graphs, pervade both our society and the natural world around us. Uncertainty, on the other hand, is inherent in the underlying data due to a variety of reasons, such as noisy measurements, lack of precise information needs, inference and prediction models, or explicit manipulation, e.g., for privacy purposes. Therefore, uncertain, or probabilistic, graphs are increasingly used to represent noisy linked data in many emerging application scenarios, and they have recently become a hot topic in the database and data mining communities. Many classical algorithms such as reachability and shortest path queries become #**P**-complete and, thus, more expensive over uncertain graphs. Moreover, various complex queries and analytics are also emerging over uncertain networks, such as pattern matching, information diffusion, and influence maximization queries. In this book, we discuss the sources of uncertain graphs and their applications, uncertainty modeling, as well as the complexities and algorithmic advances on uncertain graphs processing in the context of both classical and emerging graph queries and analytics. We emphasize the current challenges and highlight some future research directions.

KEYWORDS

uncertain graphs, reliability, pattern matching, similarity search, influence maximization

Contents

Acknowledgments . . . . . **xiii**

1 Introduction to Uncertain Graphs **1**

1.1 Data as Uncertain Graphs 1
1.2 Modeling of Uncertain Graphs 5
1.3 Challenges in Processing Uncertain Graphs 7

2 Reliability Queries **11**

2.1 Reliability 11
2.2 Shortest Path 19
2.3 Nearest Neighbors 21

3 Graph Pattern Matching Queries **25**

3.1 The Pattern Matching Problem 25
3.2 Filtering-and-Verification Framework 27
3.3 Probabilistic Pruning 28
3.4 Verification 31
3.4.1 Basic Sampling 32
3.4.2 Tree-based Sampling 32
3.4.3 Hybrid Sampling 33

4 Graph Similarity Search Queries **37**

4.1 The Similarity Search Problem 37
4.2 Probabilistic Subgraph Similarity Query Processing 41
4.3 Probabilistic Supergraph Similarity Query Processing 47

5 Influence Maximization **55**

5.1 Information Diffusion Models 55
5.2 The Influence Maximization Problem 56
5.3 Competitive Influence Maximization 58
5.4 Influence Maximization as a Service 59
5.5 Topic-aware Influence Maximization 60

6 Major Open Problems 63

Bibliography 65

Authors' Biographies 79

Acknowledgments

We would like to thank our families for their love and support during the busy time spent in writing this book.

For their collaborations, Arijit Khan would like to thank Xifeng Yan, Donald Kossmann, Charu Aggarwal, Francesco Bonchi, Gustavo Alonso, Aris Gionis, Sameh Elnikety, Chengkai Li, Gao Cong, Sourav Bhowmick, Lei Chen, Cong Yu, Prasenjit Mitra, Shu Tao, Yinghui Wu, Francesco Gullo, Aamir Cheema, Vishwakarma Singh, Xiangyu Ke, Leroy Lim Hong Quan, Siyuan Liu, Pratanu Roy, Lorenzo Severini, Vijaya Krishna Yalavarthi, Sixing Yan, Tenindra Abeywickrama, Michelle Teo Wan Teng, Rojin Rezvan, Andreas Nufer, Gustavo Segovia, Thomas Wohler, Benjamin Zehnder, Bojana Dimcheva, and Nandish Jayaram.

The research of Arijit Khan is supported by AcRF MOE Tier-1 RG83/16 and NTU M4081678. Any opinions, findings, and conclusions in this publication are those of the authors and do not necessarily reflect the views of the funding agencies.

Sincerely,

Arijit Khan, Yuan Ye, and Lei Chen
July 2018