Information and Influence Propagation in Social Networks

Synthesis Lectures on Data Management

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

Synthesis Lectures on Data Management is edited by Tamer Özsu of the University of Waterloo. The series will publish 50- to 125 page publications on topics pertaining to data management. The scope will largely follow the purview of premier information and computer science conferences, such as ACM SIGMOD, VLDB, ICDE, PODS, ICDT, and ACM KDD. Potential topics include, but not are limited to: 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.

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 2014

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.

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

ISBN: 978-3-031-00722-4 paperback
ISBN: 978-3-031-01850-3 ebook

DOI 10.1007/978-3-031-01850-3

A Publication in the Springer series
SYNTHESIS LECTURES ON DATA MANAGEMENT

Lecture #37
Series Editor: M. Tamer Özsu, *University of Waterloo*
Series ISSN
Synthesis Lectures on Data Management
Print 2153-5418 Electronic 2153-5426

Information and Influence Propagation in Social Networks

Wei Chen
Microsoft Research Asia

Laks V.S. Lakshmanan
University of British Columbia

Carlos Castillo
Qatar Computing Research Institute

SYNTHESIS LECTURES ON DATA MANAGEMENT #37

ABSTRACT

Research on social networks has exploded over the last decade. To a large extent, this has been fueled by the spectacular growth of social media and online social networking sites, which continue growing at a very fast pace, as well as by the increasing availability of very large social network datasets for purposes of research. A rich body of this research has been devoted to the analysis of the propagation of information, influence, innovations, infections, practices and customs through networks. Can we build models to explain the way these propagations occur? How can we validate our models against any available real datasets consisting of a social network and propagation traces that occurred in the past? These are just some questions studied by researchers in this area. Information propagation models find applications in viral marketing, outbreak detection, finding key blog posts to read in order to catch important stories, finding leaders or trendsetters, information feed ranking, etc. A number of algorithmic problems arising in these applications have been abstracted and studied extensively by researchers under the garb of *influence maximization*.

This book starts with a detailed description of well-established diffusion models, including the independent cascade model and the linear threshold model, that have been successful at explaining propagation phenomena. We describe their properties as well as numerous extensions to them, introducing aspects such as competition, budget, and time-criticality, among many others. We delve deep into the key problem of influence maximization, which selects key individuals to activate in order to influence a large fraction of a network. Influence maximization in classic diffusion models including both the independent cascade and the linear threshold models is computationally intractable, more precisely #P-hard, and we describe several approximation algorithms and scalable heuristics that have been proposed in the literature. Finally, we also deal with key issues that need to be tackled in order to turn this research into practice, such as learning the strength with which individuals in a network influence each other, as well as the practical aspects of this research including the availability of datasets and software tools for facilitating research. We conclude with a discussion of various research problems that remain open, both from a technical perspective and from the viewpoint of transferring the results of research into industry strength applications.

KEYWORDS

social networks, social influence, information and influence diffusion, stochastic diffusion models, influence maximization, learning of propagation models, viral marketing, competitive influence diffusion, game theory, computational complexity, approximation algorithms, heuristic algorithms, scalability.

To our families

Jian, Joice Yitao, and Ellie Yiqing

Sarada, Sundaram, Sharada, and Kaavya

Fabiola and Felipe

Contents

Acknowledgments . . . . . **xv**

1 Introduction **1**

1.1 Social Networks and Social Influence 1
1.1.1 Examples of Social Networks 1
1.1.2 Examples of Information Propagation 2
1.2 Social Influence Examples 2
1.3 Social Influence Analysis Applications 4
1.4 The Flip Side 5
1.5 Outline of This Book 6

2 Stochastic Diffusion Models **9**

2.1 Main Progressive Models 10
2.1.1 Independent Cascade Model 10
2.1.2 Linear Threshold Model 15
2.1.3 Submodularity and Monotonicity of Influence Spread Function 19
2.1.4 General Threshold Model and General Cascade Model 20
2.2 Other Related Models 25
2.2.1 Epidemic Models 25
2.2.2 Voter Model 29
2.2.3 Markov Random Field Model 30
2.2.4 Percolation Theory 32

3 Influence Maximization **35**

3.1 Complexity of Influence Maximization 36
3.2 Greedy Approach to Influence Maximization 38
3.2.1 Greedy Algorithm for Influence Maximization 39
3.2.2 Empirical Evaluation of (G, k) 44
3.3 Scalable Influence Maximization 48
3.3.1 Reducing the Number of Influence Spread Evaluations 48
3.3.2 Speeding Up Influence Computation 51
3.3.3 Other Scalable Influence Maximization Schemes 66

4 Extensions to Diffusion Modeling and Influence Maximization 67
4.1 A Data-Based Approach to Influence Maximization 67
4.2 Competitive Influence Modeling and Maximization 71
4.2.1 Model Extensions for Competitive Influence Diffusion 72
4.2.2 Maximization Problems for Competitive Influence Diffusion 75
4.2.3 Endogenous Competition 86
4.2.4 A New Frontier – The Host Perspective 88
4.3 Influence, Adoption, and Profit 94
4.3.1 Influence vs. Adoption 95
4.3.2 Influence vs. Profit 97
4.4 Other Extensions 101

5 Learning Propagation Models 105
5.1 Basic Models 105
5.2 IC Model 106
5.3 Threshold Models 109
5.3.1 Static Models 111
5.3.2 Does Influence Remain Static? 113
5.3.3 Continuous Time Models 114
5.3.4 Discrete Time Models 115
5.3.5 Are All Objects Equally Influence Prone? 116
5.3.6 Algorithms 117
5.3.7 Experimental Validation 119
5.3.8 Discussion 120

6 Data and Software for Information/Influence: Propagation Research 123
6.1 Types of Datasets 123
6.2 Propagation of Information "Memes" 124
6.2.1 Microblogging 124
6.2.2 Newspapers/blogs/etc. 125
6.3 Propagation of Other Actions 126
6.3.1 Consumption/Appraisal Platforms 126
6.3.2 User-Generated Content Sharing/Voting 127
6.3.3 Community Membership as Action 127
6.3.4 Cross-Provider Data 128
6.3.5 Phone Logs 128
6.4 Network-Only Datasets 128

6.4.1 Citation Networks . . . 129
6.4.2 Other Networks . . . 129
6.5 Other Off-Line Datasets . . . 129
6.6 Publishing Your Own Datasets . . . 130
6.7 Software Tools . . . 130
6.7.1 Graph Software Tools . . . 131
6.7.2 Propagation Software Tools . . . 131
6.7.3 Visualization . . . 132
6.8 Conclusions . . . 132

7 Conclusion and Challenges . . . 135
7.1 Application-Specific Challenges . . . 135
7.1.1 Prove Value for Advertising/Marketing . . . 135
7.1.2 Learn to Design for Virality . . . 136
7.1.3 Correct for Sampling Biases . . . 136
7.1.4 Contribute to Other Applications . . . 137
7.2 Technical Challenges . . . 137
7.3 Conclusions . . . 139

A Notational Conventions . . . 141

Bibliography . . . 143

Authors' Biographies . . . 157

Index . . . 159

Acknowledgments

The material presented here has been deeply influenced by the research we conducted with many wonderful colleagues, students, and post-doctoral fellows as well as the numerous stimulating discussions we have had with them. We would like to express our gratitude to our collaborators, including: Smriti Bhagat, Francesco Bonchi, Alex Collins, Rachel Cummings, Aristides Gionis, Amit Goyal, Xinran He, Dino Ienco, Qingye Jiang, Te Ke, Yanhua Li, Zhenming Liu, Wei Lu, Michael Mathioudakis, David Rincón, Guojie Song, Xiaorui Sun, Antti Ukkonen, Suresh Venkatasubramanian, Chi Wang, Yajun Wang, Wei Wei, Siyu Yang, Yifei Yuan, Li Zhang, and Zhi-Li Zhang. We are grateful to Lewis Tseng, Yajun Wang, and Cheng Yang for helpful discussions on some technical sections of the book, and to Tian Lin for running a few additional evaluation tests on top of what previously published papers had and for generating relevant plots. Thanks are due to Qiang Li, Tian Lin, Wei Lu, Lewis Tseng, Yajun Wang, and Cheng Yang for their careful reading of and excellent comments on the manuscript which greatly improved our presentation. We appreciate Diane Cerra and Tamer Ozsu for their assistance throughout the preparation of this manuscript and for their patience. Tamer's editorial comments were especially helpful in improving the readability. Last but not the least, we are indebted to our families whose patience and support throughout this project has been invaluable.

Wei Chen, Laks V.S. Lakshmanan, and Carlos Castillo
August 2013