Diffusion Source Localization in Large Networks

Synthesis Lectures on Communication Networks

Editor
R. Srikant, ***University of Illinois at Urbana-Champaign***

Founding Editor Emeritus
Jean Walrand, ***University of California, Berkeley***

Synthesis Lectures on Communication Networks is an ongoing series of 75- to 150-page publications on topics on the design, implementation, and management of communication networks. Each lecture is a self-contained presentation of one topic by a leading expert. The topics range from algorithms to hardware implementations and cover a broad spectrum of issues from security to multiple-access protocols. The series addresses technologies from sensor networks to reconfigurable optical networks.
The series is designed to:

- Provide the best available presentations of important aspects of communication networks.
- Help engineers and advanced students keep up with recent developments in a rapidly evolving technology.
- Facilitate the development of courses in this field

Diffusion Source Localization in Large Networks
Lei Ying and Kai Zhu
2018

Communications Networks: A Concise Introduction, Second Edition
Jean Walrand and Shyam Parekh
2017

BATS Codes: Theory and Practice
Shenghao Yang and Raymond W. Yeung
2017

Analytical Methods for Network Congestion Control
Steven H. Low
2017

Advances in Multi-Channel Resource Allocation: Throughput, Delay, and Complexity
Bo Ji, Xiaojun Lin, and Ness B. Shroff
2016

A Primer on Physical-Layer Network Coding
Soung Chang Liew, Lu Lu, and Shengli Zhang
2015

Sharing Network Resources
Abhay Parekh and Jean Walrand
2014

Wireless Network Pricing
Jianwei Huang and Lin Gao
2013

Performance Modeling, Stochastic Networks, and Statistical Multiplexing, Second Edition
Ravi R. Mazumdar
2013

Packets with Deadlines: A Framework for Real-Time Wireless Networks
I-Hong Hou and P.R. Kumar
2013

Energy-Efficient Scheduling under Delay Constraints for Wireless Networks
Randall Berry, Eytan Modiano, and Murtaza Zafer
2012

NS Simulator for Beginners
Eitan Altman and Tania Jiménez
2012

Network Games: Theory, Models, and Dynamics
Ishai Menache and Asuman Ozdaglar
2011

An Introduction to Models of Online Peer-to-Peer Social Networking
George Kesidis
2010

Stochastic Network Optimization with Application to Communication and Queueing Systems
Michael J. Neely
2010

Scheduling and Congestion Control for Wireless and Processing Networks
Libin Jiang and Jean Walrand
2010

Performance Modeling of Communication Networks with Markov Chains
Jeonghoon Mo
2010

Communication Networks: A Concise Introduction
Jean Walrand and Shyam Parekh
2010

Path Problems in Networks
John S. Baras and George Theodorakopoulos
2010

Performance Modeling, Loss Networks, and Statistical Multiplexing
Ravi R. Mazumdar
2009

Network Simulation
Richard M. Fujimoto, Kalyan S. Perumalla, and George F. Riley
2006

© 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.

Diffusion Source Localization in Large Networks

Lei Ying and Kai Zhu

ISBN: 978-3-031-79284-7 paperback
ISBN: 978-3-031-79285-4 ebook
ISBN: 978-3-031-79286-1 hardcover

DOI 10.1007/978-3-031-79285-4

A Publication in the Springer series
SYNTHESIS LECTURES ON COMMUNICATION NETWORKS

Lecture #21
Series Editor: R. Srikant, *University of Illinois at Urbana-Champaign*
Founding Editor Emeritus: Jean Walrand, *University of California, Berkeley*
Series ISSN
Print 1935-4185 Electronic 1935-4193

Diffusion Source Localization in Large Networks

Lei Ying and Kai Zhu
Arizona State University

SYNTHESIS LECTURES ON COMMUNICATION NETWORKS #21

ABSTRACT

Diffusion processes in large networks have been used to model many real-world phenomena, including how rumors spread on the Internet, epidemics among human beings, emotional contagion through social networks, and even gene regulatory processes. Fundamental estimation principles and efficient algorithms for locating diffusion sources can answer a wide range of important questions, such as identifying the source of a widely spread rumor on online social networks. This book provides an overview of recent progress on source localization in large networks, focusing on theoretical principles and fundamental limits. The book covers both discrete-time diffusion models and continuous-time diffusion models. For discrete-time diffusion models, the book focuses on the Jordan infection center; for continuous-time diffusion models, it focuses on the rumor center. Most theoretical results on source localization are based on these two types of estimators or their variants. This book also includes algorithms that leverage partial-time information for source localization and a brief discussion of interesting unresolved problems in this area.

KEYWORDS

source localization, diffusion processes, epidemic processes, Jordan infection center, rumor center, large networks, social networks

To Ethan, Lingfang, and my parents—L.Y.

To Tu, my parents, and grandparents—K.Z.

Contents

Preface . . . . . **xiii**

Acknowledgments . . . . . **xv**

1 Motivation and Background **1**

2 Source Localization under Discrete-Time Diffusion Models **5**

2.1 The Independent Cascade (IC) Diffusion Model on Tree Networks 5
2.1.1 Models 5
2.1.2 Problem Formulation 6
2.1.3 The Short-Fat Tree Algorithm 6
2.1.4 The ML Estimator 10
2.2 Single-Source Localization on the ER Random Graph 17
2.3 Jordan Infection Center: The Universal Sample-Path-Based Source Estimator 25
2.4 NETSLEUTH: Approximate ML Estimator for General Networks 32
2.4.1 Mean Field Approximation 32
2.5 Numerical Evaluation 36
2.5.1 Evaluation Metrics 37
2.5.2 Binomial Trees 37
2.5.3 The ER Random Graph 38
2.5.4 The Internet Autonomous System Network 41
2.5.5 Running Time vs. Performance 41
2.6 Locating Multiple Diffusion Sources 44

3 Source Localization under Continuous-Time Diffusion Models **45**

3.1 Rumor Centrality: The ML Estimator 45
3.1.1 Models 45
3.1.2 Problem Formulation 45
3.1.3 Regular Trees 48
3.1.4 General Tree Networks 48
3.1.5 General Networks 50

3.2 Accuracy of Rumor Centers 52
3.2.1 Line Graphs 52
3.2.2 3-Regular Trees 55

4 Source Localization with Partial Timestamps 59
4.1 EIF: A Greedy Algorithm 61
4.1.1 Cost-Based and Tree-Based Ranking 65
4.2 Gaussian Heuristic 65
4.3 Performance Evaluation 66

5 Open Questions 71

Bibliography 73

Authors' Biographies 79

Preface

We are living in an ever-connected world because of the proliferation of the Internet, mobile devices, and online social networks. Mathematical modeling and analysis of diffusion processes in large networks can help us understand many of the phenomena and problems in this networked world—such as epidemic threshold (i.e. the condition under which an epidemic will break out) and influence minimization/maximization on social networks—and have therefore become a research topic of great interest in recent years. This book studies retrospective analysis of diffusion in networks, in particular the question of how to find the source or sources of a diffusion process given incomplete information on both the network and diffusion. This book is a result of exciting research by the authors since 2011. It is suited for graduate students and researchers who are interested in this problem of large network diffusion. The goal of the book is to provide an overview of key theoretical discoveries on this topic. We feel it is more important for readers to understand the intuition behind the results than to know the detailed proofs that, in some cases, can be quite tedious. Therefore, throughout this book, we use examples to explain the analysis and present simplified proofs based on additional assumptions.

ORGANIZATION OF THE BOOK

We start this book by discussing the motivation and background of the source localization problem in Chapter 1. Chapter 2 focuses on discrete-time diffusion models and source estimators based on the Jordan infection center. Starting from tree networks and independent cascade models, we introduce the short-fat tree (SFT) algorithm based on the Jordan infection center and show that the estimator identified by the SFT is the maximum likelihood (ML) estimator of the diffusion source. Then, we consider the Erdos-Renyi random graph and present conditions under which the source is the unique Jordan infection center with high probability. We further introduce NETSLEUTH, which is a heuristic source localization algorithm for general networks.

Chapter 3 introduces rumor centrality and rumor center for continuous-time diffusion models. We first prove that the rumor center is the ML estimator of the diffusion source on regular tree networks, and then present source localization algorithms based on rumor centrality for general networks. Fundamental limits on the detection probability of rumor center are also included.

In Chapter 4, we introduce algorithms that use partial timestamps—the times at which nodes were infected—for locating the source. Besides providing a source estimator, these algo-

rithms rank the nodes according to their likelihood of being the source, which can be very useful in practice. Finally, in Chapter 5, we briefly discuss several unresolved problems in the field.

Lei Ying and Kai Zhu
June 2018

Acknowledgments

The authors would like to thank Prof. R. Srikant for inviting us to write this book.

Lei Ying and Kai Zhu
June 2018