Surveys and Tutorials in the Applied Mathematical Sciences
7

Haiyan Wang
Feng Wang
Kuai Xu

Modeling Information Diffusion in Online Social Networks with Partial Differential Equations

Springer

Surveys and Tutorials in the Applied Mathematical Sciences

Volume 7

Featuring short books of approximately 80-200pp, Surveys and Tutorials in the Applied Mathematical Sciences (STAMS) focuses on emerging topics, with an emphasis on emerging mathematical and computational techniques that are proving relevant in the physical, biological sciences and social sciences. STAMS also includes expository texts describing innovative applications or recent developments in more classical mathematical and computational methods.

This series is aimed at graduate students and researchers across the mathematical sciences. Contributions are intended to be accessible to a broad audience, featuring clear exposition, a lively tutorial style, and pointers to the literature for further study. In some cases a volume can serve as a preliminary version of a fuller and more comprehensive book.

More information about this series at http://www.springer.com/series/7219

Haiyan Wang • Feng Wang • Kuai Xu

Modeling Information Diffusion in Online Social Networks with Partial Differential Equations

Springer

Haiyan Wang
School of Mathematical & Natural Sciences
Arizona State University
Phoenix, AZ, USA

Feng Wang
School of Mathematical & Natural Sciences
Arizona State University
Phoenix, AZ, USA

Kuai Xu
School of Mathematical & Natural Sciences
Arizona State University
Phoenix, AZ, USA

ISSN 2199-4765 ISSN 2199-4773 (electronic)
Surveys and Tutorials in the Applied Mathematical Sciences
ISBN 978-3-030-38850-8 ISBN 978-3-030-38852-2 (eBook)
https://doi.org/10.1007/978-3-030-38852-2

Mathematics Subject Classification (2010): 05C50, 05C60, 34A34, 35-02, 35C07, 35R35, 35B36, 35K57, 35B32, 35Q94, 35R02, 62H30, 65F15, 68-02, 68P01, 68R10, 68U35, 91C20, 92D25, 91D30, 94A15, 94C15

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To our families.

Preface

For many researchers, the rise of tremendously influential social media such as Twitter and Facebook presents both challenges and opportunities. For example, a better understanding of the information diffusion process over online social networks can effectively predict and coordinate online social activities. However, even though the increasing availability of unprecedented quantities of data has accelerated research on information diffusion in online social networks, because of the complexity of social interactions and rapid changes in social media, the mechanism of information diffusion in online social networks remains elusive. In the literature various mathematical models have been proposed to study information diffusion on online social networks. However, these mathematical models, particularly, deterministic dynamical models, are largely based on ordinary differential equations (ODEs) that deal with collective social processes over time.

The goal of this book is to introduce a new dynamic modeling approach to the use of partial differential equations (PDEs) for describing temporal-spatial patterns in information diffusion over social media. The PDE-based models are reaction-diffusion equations built on intuitive social distances between communities (clusters) of online users. In particular, leveraging clustering analysis of spatial big data will dramatically expand applications of the PDE models to many health and social problems including influenza prediction. The PDE approach advocates a paradigm shift for modeling information diffusion in online social networks and lays the theoretical groundwork for many spatio-temporal modeling problems in the big data era.

This book stems from lecture notes for a course, Social Media and Mathematics, given by the first author, at Arizona State University and other institutions, and the authors' numerous published and unpublished works on modeling information diffusion over online social networks with PDEs.

Background preparations and necessary references for social graphs and information diffusion are also included to ensure the book is accessible to mathematicians and computer scientists as well as general researchers in social media.

Phoenix, AZ, USA Haiyan Wang
Phoenix, AZ, USA Feng Wang
Phoenix, AZ, USA Kuai Xu
October 29, 2019

Acknowledgments

We would like to acknowledge the contributions to the papers from our collaborators: Guowei Dai, Xiaohua Jia, K. Hazel Kwon, Chengxia Lei, Zhigui Lin, Ruyun Ma, Chuan Peng, Jingli Ren, Yufang Wang, Jianhong Wu, Hongyong Zhao, Shuhua Zhang, Dandan Zhu, Linhe Zhu. We would thank several colleagues including Guoping Jiang, Zhen Jin, Maoxing Liu, Yuan Lou, Yurong Song, and a number of students for stimulating discussions. Adrian Avram, Jaime Chon, Ross Raymond, Paul Wagenseller, Daniel Langley, Shaun Fuller, and a number of students helped Twitter data collection and Matlab simulations. Dingyong Bai and Xiaoling Han helped organize the references. We would like to thank Kristina Lerman for making the Digg 2009 data set available to the research project.

We want to thank the reviewers for their inputs as well as the editorial staff of Springer who were involved in the oversight of the book. The Springer editors encouraged and supported the project from the beginning. We are grateful for their encouragement and support.

Finally, we would like to acknowledge that this project has been supported by two research grants from the U.S. National Science Foundation (CNS-1218212 and DMS-1737861).

Contents

Chapter 1
Introduction

Abstract Online social networks (OSNs) such as Twitter and Facebook, emerging as the "model organism" of Big Data, have gained tremendous popularity for the platforms they provided for information exchange. Much of prior work on information diffusion over online social networks has been based on empirical and statistical approaches. The majority of dynamical models arising from information diffusion over online social networks are ordinary differential equations (ODEs). Recently, the authors proposed to use partial differential equations (PDEs) to model information diffusion in online social networks and introduced a new transdisciplinary architecture for modeling information diffusion. These studies demonstrate fascinating connections between advanced mathematics and online social networks.

Online social networks (OSNs) such as Twitter and Facebook, emerging as the "model organism" of Big Data, have gained tremendous popularity for the platforms they provided for information exchange. Much of prior work on information diffusion over online social networks has been based on empirical and statistical approaches. The majority of dynamical models arising from information diffusion over online social networks are ordinary differential equations (ODEs). Recently, the authors proposed to use partial differential equations (PDEs) to model information diffusion in online social networks and introduced a new transdisciplinary architecture for modeling information diffusion. These studies demonstrate fascinating connections between advanced mathematics and online social networks.

A significant body of research about online social networks has focused on analysis of such networks with empirical approaches that use data mining and statistical modeling schemes [9, 18, 19, 25, 29, 34, 35, 41, 43, 47, 50, 55, 60–62, 68, 69, 72, 73, 85, 88, 89, 105, 111, 120, 121, 138–141, 144]. Mathematical models have played a significant role in understanding and predicting

H. Wang et al., *Modeling Information Diffusion in Online Social Networks with Partial Differential Equations*, Surveys and Tutorials in the Applied Mathematical Sciences 7, https://doi.org/10.1007/978-3-030-38852-2_1

information diffusion in online social networks over time. In particular, epidemiological models have influenced the research on information diffusion [8, 51, 80, 89, 90, 124, 145, 146]. However, the deterministic models proposed for online social networks in the literature are largely based on ordinary differential equations (ODEs) that deal with collective social processes over time.

In a paper [128], the authors proposed to use partial differential equations (PDEs) built on intuitive cyber-distance among online users to study both temporal and spatial patterns of information diffusion process in social media. One of the simple, yet fundamental questions that the models address is this: for a piece of given information m initiated from a particular user called *source* s, what is the density of influenced users at network distance x from the source at *any time* t. We validate the models with real datasets collected from two popular social media sites, Twitter and Digg. The experiment results show that the models can achieve over 90% accuracy and effectively predict the density of influenced users.

This paper [128] is the first attempt to propose a PDE-based model for characterizing and predicting the temporal and spatial patterns of information diffusion over online social networks, which is also indicated in a survey by Zhang et al. [146]. In Guille et al.'s survey [38] on information diffusion over online social networks, the PDE model in [128] is reported as one of the three non-graph predictive models: epidemiological models, linear influence model (LIM), and PDE approach. The LIM approach developed in [140] focuses on predicting the temporal dynamics of information diffusion through solving non-negative least squares problems. The PDE-based models are dynamic systems that take into account the influence of the underlying network structure as well as information contents for predicting information diffusion over both temporal and spatial dimensions. We shall propose a number of spatio-temporal epidemiological models to describe information diffusion in online social networks in this book.

The book lies at the interface of mathematics, social media analysis, network science, and data science. A key challenge of this interdisciplinary research is to integrate big data from online social networks into the framework of partial differential equations. We use clustering analysis from data mining to aggregate big data for the validation of the PDE models. There are many clustering methods such as k-means clustering. We focus on spectral clustering analysis in this book as our data are graph-structured. The book integrates the research efforts carried out collaboratively by mathematicians specialized in partial differential equations, computer scientists focused on network theory and data mining, and researchers in social media. The extension of partial differential equations into online social networks presents new opportunities and challenges for mathematicians as well as computer scientists and researchers in social media.

Chapter 2
Ordinary Differential Equation Models on Social Networks

Abstract In this chapter we consider a number of ordinary differential equation models for diffusion of innovation and epidemiological models. We discuss the classical theory on diffusion of innovation, emphasizing online social networks and analyzing several ordinary differential equation models for innovation diffusion. We also present a number of basic compartment epidemiological models and their applications in online social networks, and finally we discuss SIR models and their extensions when the total population is not constant.

2.1 Introduction

Information diffusion over online social networks has become a fast growing research domain encompassing techniques from a plethora of sciences, among them mathematics, computer science, communications and marketing, etc. In addition, the method for studying the spread of infectious diseases among a population has been applied to understanding information spreading patterns. Information diffusion has become a subject with disparate views of what is an information diffusion process. We focus here on a particular case in which information diffuses in online social networks, and we define information diffusion as the process by which a piece of information (knowledge) spreads and reaches individuals through interactions in a network.

The theory of information diffusion can be traced back to the research on the diffusion of innovation over a population, advanced by a pioneering mass communication scholar, E. M. Rogers. Information diffusion in online social networks becomes a premier example to revisit innovation diffusion. In this chapter, we will focus on ODE models for diffusion of innovation

H. Wang et al., *Modeling Information Diffusion in Online Social Networks with Partial Differential Equations*, Surveys and Tutorials in the Applied Mathematical Sciences 7, https://doi.org/10.1007/978-3-030-38852-2_2

and epidemics. Both diffusion of innovations and epidemic models provide a global view of how an innovation (e.g., news, a product) or a type of disease spreads through a population even when interactions among individuals are unavailable. We will extend innovation diffusion and epidemic models to spatial models where both local and global information between clusters in a network are available.

2.2 Diffusion of Innovations

Diffusion of innovations is a theory that seeks to explain how new ideas and technologies spread through cultures. E. M. Rogers was a professor of communication studies who popularized the theory with his book "Diffusion of Innovations" [104]. The origins of the diffusion of innovations theory span various disciplines including anthropology, early sociology, rural sociology, education, industrial sociology, and medical sociology. Rogers [104] theorized that information diffusion is the social process through which an innovation is communicated through certain channels over time among the participants in a social system. He identified four key elements that influence diffusion of a new idea: the innovation itself, communication channels, time, and a social system. The theory of diffusion of innovation has been applied to numerous contexts, including marketing, communications, health promotion, organizational studies, and complexity studies.

The rise of social media has provided a new platform to study diffusion of innovation. Rogers [104] defined an innovation as an idea, practice, or object if it is perceived as novel by an individual or other unit of adoption. A piece of news being reposted many times in online social networks is a typical example of innovations diffusing across online social networks. The theory of diffusion of innovations can be applied to online social networks to answer why and how information spreads fast and reaches a broad audience. It also can be used to reveal some key characteristics of information diffusion such as its motivations. With appropriate models, it can be used to further predict the rate at which ideas spread. In this section, we review some key characteristics of innovation diffusion that are important for the PDE modeling of information diffusion in online social networks. Finally, we present mathematical models that can be used to describe the process of innovation diffusion.

2.2.1 Characteristics of Innovation Diffusion

Rogers [104] explored many characteristics of innovations, individual adopters, and organizations. The nature of networks and the roles opin-