Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems—cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial or functional structures. Models of such systems can be successfully mapped onto quite diverse "real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The three major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations, and the "Springer Briefs in Complexity" which are concise and topical working reports, case studies, surveys, essays and lecture notes of relevance to the field. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

Indexed by SCOPUS, INSPEC, zbMATH, SCImago.

Series Editors

Understanding Complex Systems

Founding Editor: Scott Kelso

Hernán A. Makse · Marta Zava

The Science of Influencers and Superspreaders

Using Networks and Artificial Intelligence to Understand Fake News, Pandemics, Markets, and the Brain

Springer

Hernán A. Makse
Department of Physics
City University of New York
New York, NY, USA

Marta Zava
Department of Finance
Bocconi University
Milan, Italy

ISSN 1860-0832 ISSN 1860-0840 (electronic)
Understanding Complex Systems
ISBN 978-3-031-78057-8 ISBN 978-3-031-78058-5 (eBook)
https://doi.org/10.1007/978-3-031-78058-5

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

To Laura, Loris, Ana María and Milena.

Preface

This is a book about influencers. This term is as an umbrella term to describe essential, critical, central, superspreader, and core nodes in any type of complex network. We will review network theoretical frameworks and algorithms to identify and understand the emergence of influencers not only in society but also across a spectrum of systems ranging from biology to markets. Each chapter in this book is dedicated to study a different type of influencer, in turn. We will start with the canonical influencer problem:

- **Influence Maximization Problem**: Finding the optimal set of influencers that maximizes the spreading of information in a social network.

And then apply similar concepts to describe and identify:

- Superspreaders of disease in pandemics,
- Genetic influencers in gene regulatory networks,
- Neural influencers in the brain,
- Keystone species in ecosystems, and
- Financial influencers in markets.

The influencers' theory lays out that each big data problem is unique. There is no silver-bullet solution for all big data challenges. All networks have a different structure, and therefore, influencers vary depending on the problem. In other words, a new mathematical algorithm will have to be crafted for each dataset. Some algorithms work better for social media, others for genetic networks, and others for ecological networks. In this book, we will touch upon the different applications of AI and develop tailored solutions for every significant big data system.

Organization of the Book

The book is organized as follows.

In Chap. 1, we explain the standard mathematical theory behind influencer dynamics. We explain the simplest algorithms to identify influencers for any complex network. These are based on the topological properties of the network, called centralities, which are commonly used to identify and rank essential network nodes. These include centralities involving the number of connections (degree), k-cores, PageRank, eigenvector and eigenvalue-based centralities, betweenness, closeness, and many others. We will then mathematically define the influence maximization problem, trying to capture the commonalities among the different influencer problems. We will explore the different approaches in the literature employed to solve this optimization problem, including concepts from percolation theory, network dismantling and decycling, spectral theory, spin glasses, greedy algorithms, and others. We draw upon the work published in the vast literature encompassing social, computer, mathematical, and physical sciences to survey the theoretical approaches for influencer search and identification.

In Chap. 2, social media influencers are discussed. We show how combining network theory, AI, and alternative social media data can help predict and understand opinion trends. By collecting billions of data points from social media like Twitter and Facebook, we show how a few influencers—not the traditional media—can drive the conversation and potentially change the opinions of millions of people. We will show how fake news spreads in social networks and how influencers can drive the conversation across the spectrum spanning from genuine to fake news. We show how AI is used to classify the actors and simultaneously predict opinion trends, including election outcomes with machine learning, and show how influencers can shape the outcome of elections. For instance, influencers appeared organically in social media and enormously impacted the US presidential elections in 2016 and 2020.

In Chap. 3, we will lay out how influencer marketing is done using network theories for influencer identification. We will describe how influencers emerge organically to influence people to adopt specific products and brands. Anecdotal evidence suggests that the economic status of an individual is reflected in the structure of their social ties. Here, we will show that an individual's influence on the structural integrity of the social network can be used to infer their economic wellness. This leads to targeted marketing strategies based on influencer theory, reporting a three-fold increase in response rate compared to usual targeting campaigns.

In Chap. 4, we focus on superspreaders of disease during the COVID-19 pandemic. We show how the theory of influencers fed with big data from GPS geolocalization of people obtained from mobile smartphones can be used to design digital contact tracing protocols to implement intelligent quarantines to stop the spread of COVID-19, and at the same time, minimize the disruptions to society. The spread of the coronavirus disease caused by the SARS-CoV-2 virus became a worldwide problem with devastating consequences. To slow down the spread of the disease, massive quarantines were implemented across the world, provoking enormous social and

economic disruptions. Influencer theory can contribute to ameliorating this problem by providing intelligent quarantine protocols based on big data. We show that influencer theory applied to contact networks obtained from digital contact tracing can readily identify superspreader events. An optimal targeted quarantine then follows by isolating only those superspreaders (influencers) that optimally dismantle the disease's transmission chain upon isolation. This could greatly reduce the disruption to society. Thus, optimization theory applied to influencers can become a crucial tool to combat pandemics.

Chapter 5 focuses on genetic influencers in gene regulatory networks. We represent the information flow between nodes in biological networks to decipher how cellular function emerges from their interactions. In this case, influencers take the form of either essential genes for integration that hold the network together or building blocks crucial to synchronizing information. We introduce a particular form of symmetry, symmetry fibration, to identify these building blocks. Using these symmetries, we identify input trees that deconstruct the network into functional building blocks processing equivalent coexpression of genes. This has significant ramifications for the identification of disease genes and for finding molecular targets for drug development.

In Chap. 6, we focus on neural influencers in the brain. The study of the human brain, as the elemental paradigm of an efficient, robust modular system interconnected as a network, will help scientists predict the map of neural collective influencers. In this chapter, we elaborate on models of brain networks to shape the pattern of brain activations. We explore how influencer theory is applied to these networks to predict and subsequently target nodes essential for global integration in a memory network in rodents. We further show how the theory predicts the localization of essential functional areas in the human brain that are used to guide surgical interventions in the human brain with minimal impacts to cognition.

Chapter 7 studies how influencer theory can help to identify the causes for the collapse and instability of dynamical systems. These systems include ecosystems, where influencers are the keystone species, human societies, and financial systems. When keystone species located at the maximum k-core of the mutualistic ecosystem go extinct, the system reaches the tipping point of its collapse. Monitoring the k-core then proves to be a valuable method to anticipate the catastrophic collapse of many ecosystems.

Chapter 8 studies financial networks. We trace the evolution of venture capital networks, study the AI's transformative impact on mergers and acquisitions, examine institutional investors' networks in the initial public offers, and explain novel applications of natural language processing in stock market analysis. Through an interdisciplinary lens, we demonstrate how the integration of network science and AI enhances analytical capabilities and reshapes the very nature of financial markets.

How to Read This Book

The approach of this book has a strong algorithmic, computational, and physics flavor. It is based on the premise that complex human behavior cannot be inferred by studying a single actor or a small number of them. As the *leitmotiv* of complexity science enunciated by Phil Anderson in his Nobel lecture says: '*more is different*'. Human behavior is an emergent collective phenomenon arising from the interactions of their many actors, and this idea also applies to complex biological systems composed of genes, proteins, and neurons. These interactions are inevitably captured by networks, and their emergent properties appear only when we observe many of these interacting units.

Thus, understanding the emergence of collective human behavior is only possible when we observe, analyze, and predict the properties of millions or billions of individuals. This is the domain of big-data science applied to human behavior, for which a computational algorithmic approach is presented in this book. However, the foundations of the present work are well-rotted in sociology, especially in the subfield of computational sociology, studying the spread of influence and innovations. Most concepts we will develop in this book have been discussed in the sociological literature for decades and continue to be addressed.

From a social science perspective, there are excellent books that deal with the collective nature of influence. We can recommend a few, starting from the seminal book of Rogers on '*Diffusion of innovations*', to the most cited paper in sociology (by Web of Science) by Granovetter on the '*Strength of weak ties*' in social network theory, to more recent accounts in terms of modern social network data by Centola on '*How Behavior Spreads: The Science of Complex Contagions*'.

Here, we will explore emergent collective behaviors from a different viewpoint, inspired by a statistical physics approach, where emergent properties arise at what physicists call the 'thermodynamics limit'; at massively large populations. Because the topic spans a broad spectrum of research, we cannot exhaustively report every relevant work. However, we tried to organize the book to cover recent developments in several fields of broad interest.

This book is intended to inform readers at three different levels:

- First, those interested in the mathematically rigorous theories of influencers can concentrate on Chap. 1, where we explain the mathematics behind the algorithms to identify influencers.
- A second level informs data scientists interested in applying these algorithms in their own fields and day-to-day work. For those readers, we provide Chaps. 2–8 where we explain the applications of the theory of influencers to social networks (Chaps. 2 and 3), epidemiology (Chap. 4), biology (Chap. 5), neuroscience (Chap. 6), ecology (Chap. 7), and markets (Chap. 8). Appendix A provides links to freely available algorithms with ready-made codes organized by chapters to be used with Python, R, and Julia through various Github repositories.
- Third, this book is also intended to reach layman audiences in the social media, financial, marketing and politics fields, as well as audiences keen on learning

how big data, influencers, and AI can contribute to a better decision-making based on mathematically proven algorithms and advanced analytics in their field and business of interest. These readers are encouraged to glance over the more technical Chap. 1 and focus on the relevant chapters for their businesses.

New York, USA — Hernán A. Makse
Milan, Italy — Marta Zava

Acknowledgements

This book could not have been written without the input of countless students, postdocs, and collaborators from all sorts of fields of life. Our interest in complex networks started with conversations with then-graduate student Chaoming Song in the early 2000s about the emergence of fractal properties in networks, and it has led to the development of influencer algorithms, theories, and applications in different fields, as discussed in this book. In this journey, we have benefited from enjoyable collaborations with José Soares Andrade Jr, Francesca Arese Lucini, Paolo Boldi, Alexander Bovet, Kate Burleson-Lesser, Guido Caldarelli, Santiago Canals, Gino Del Ferraro, Martina Del Gaudio, Lazaros Gallos, Martin Golubitsky, Alireza Hashemi, Shlomo Havlin, Andrei I. Holodny, Maksim Kitsak, Ian Leifer, Qiongee Li, Omar Lizardo, Shaojun Luo, Isotta Magistrali, Flaviano Morone, David Phillips, Lucas Parra, Sen Pei, Luca Pitta, Bernardo Principi, Saulo Reis, Diego Rybski, Hernán Rozenfeld, Matteo Serafino, Mariano Sigman, Francesco Sorrentino, H. Eugene Stanley, Ian Stewart, Boleslaw K. Szymanski, Matías Travizano, Lorenzo Timpone, Giovanni Valle, Stefan Wuchty, and Zhenkun Zhou.

Disclaimer

The studies of information spreading in social media reproduced in this book typically use classifications of news outlets, social media posts, users, and influencers given in terms of different categories ranging from fake news and extremely biased news to conspiracy theories, and political orientations from left to right. These categories and classifications are a matter of opinion rather than a statement of fact. The categorizations and labels assigned to the corresponding classes are those used in the cited scientific papers and are reproduced in this book. These results have originated publicly available datasets from fact-checking organizations as credited in this book. Results are reproduced as published in scientific journals and are based on classifications of news and others from fact-checking organizations and curated lists of fake and genuine websites and outlets that have not been generated nor curated by the

authors of this book. These classifications are subject to subjectivity. The views and conclusions contained in the cited articles and this book, as well as the classification of news media and individuals into fake news, extremely biased news, conspiracy theories, influencers, and categories of political orientation from the extreme left, left-leaning, left, center, right, right-leaning, to extreme right should not be interpreted as representing the opinions of the authors of this book, the publishers, nor the funders of the work. Partial funding for this work has been provided by NIH Grant R01 EB028157 and NSF Grant 2214217.

Contents