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 two major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, and the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

For further volumes:
http://www.springer.com/series/712

Springer Series in Synergetics

Founding Editor: H. Haken

The Springer Series in Synergetics was founded by Herman Haken in 1977. Since then, the series has evolved into a substantial reference library for the quantitative, theoretical and methodological foundations of the science of complex systems.

Through many enduring classic texts, such as Haken's *Synergetics* and *Information and Self-Organization*, Gardiner's *Handbook of Stochastic Methods*, Risken's *The Fokker Planck-Equation* or Haake's *Quantum Signatures of Chaos,* the series has made, and continues to make, important contributions to shaping the foundations of the field.

The series publishes monographs and graduate-level textbooks of broad and general interest, with a pronounced emphasis on the physico-mathematical approach.

Philippe Blanchard • Dimitri Volchenkov

Random Walks and Diffusions on Graphs and Databases

An Introduction

Springer

Philippe Blanchard
Dimitri Volchenkov
Universität Bielefeld
Fakultät für Physik
Universitätsstr. 25
33615 Bielefeld
Germany
blanchard@physik.uni-bielefeld.de
dima427@yahoo.com

ISSN 0172-7389
ISBN 978-3-642-19591-4 e-ISBN 978-3-642-19592-1
DOI 10.1007/978-3-642-19592-1
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011929690

Cover design: SPi Publisher Services

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Dedicated to our parents with thanks for their care and love.

Preface

Classical graph theory has been developed in regard to its applications in urban planning, transport, energetics, and many other fields. The general optimization mindset dominating these researches has addressed to graph theory the questions which were often related to finding the *shortest path* between nodes, as being of the minimum time delay for information transmission and of the minimum cost for connection maintenance. Not surprisingly, the very definition of distance between two vertices in a graph is given as the geodesic distance, i.e., the shortest path connecting them. With respect to the graph metric, a complex network of weighted edges is rather considered as a minimum weight spanning tree of the underlying graph, i.e., a subset of paths that has no cycles but still connects to every vertices at the lowest total cost. However, in many problems of practical interest found in everyday life the existence of many paths of different lengths as well as a nexus of cycles traversing the nodes in many complex networks do matter!

In contrast to classical graph theory paying attention to the shortest paths of least cost, in the approach that we discuss in our book, *all possible paths* between the two vertices in a connected graph are taken into account, although some paths shall be more preferable than others. Such a formulation of graph theory can be called as of a "*path integral*", since "integral" means "to include all." Random walks respecting all graph symmetries assign a probability to each path in the graph to be traversed by a random walker. Then, in order to find the expected first-passage distance between the two vertices, one integrates over all possible paths of the system in between them. Consequently, each vertex is characterized with respect to the entire structure of the graph by its own "path integral" vector accounting for the sum of the probability amplitudes for every possible path leading to that from a randomly chosen vertex. Perhaps, the most interesting fact about such a "path integral" approach to graphs is that the probabilistic distance naturally induces a Euclidean metric on a graph (sometimes called the 'diffusion metric', or the 'effective resistance metric') allowing for a geometric representation of the relationships between vertices in a graph, in terms of distances and angles, as in Euclidean geometry of everyday intuition. Vertexes of graphs and units of data bases that cast in the same mold with respect to the individual data features are revealed by geometric proximity in Euclidean space that might be either exploited visually, or accounted analytically. High-dimensional Euclidean representations of graphs and databases are characterized by the rank-ordering of data traits providing us with the natural geometric framework for dimensionality reduction facilitating the data analysis and further interpretation of results.

Perhaps, Lagrange was the first scientist who investigated a simple dynamical process (diffusion) in order to study the properties of a graph (Lagrange 1867). He calculated the spectrum of the Laplace operator defined on a chain (a linear graph) of N nodes in order to study the discretization of the acoustic equations. Nowadays it is well known that random walks could be used in order to investigate and characterize how effectively the nodes and edges of large networks can be covered by different strategies (see Tadic 2002; Yang 2005; Costa and Travieso 2007 and many others).

In this book, we follow the interdisciplinary lecture course on the stochastic analysis of complex networks and databases delivered by us at the University of Bielefeld (Germany) during the Fall semester 2008 and the Spring semester 2009 and targeted to bring about a more interdisciplinary approach across diverse fields of research including complex network theory and data analysis, as well as sociology, bio-informatics, urban planning and linguistics. The book contains a wealth of material generously equipped with suggestions for further reading and the glossary of term and concepts in graph theory that is helpful for those at the beginning of their acquaintance with the subject.

In the subsequent ten chapters of this book, we describe a fascinating journey from the elementary discrete mathematics (Chaps. 1, 2) to the elements of algebraic graph theory (Chap. 3), to a detailed analysis of complex multicomponent systems and databases (Chaps. 4–7), to the applications of random walk methods for the components analysis of complex networks and databases (Chap. 8). In the Chap. 9, we discuss the dynamical processes in models containing a large number of positive and negative feedbacks such as epidemic spreading, synchronization, and self-regulation in complex genetic networks. Finally, in the Chap. 10, we consider strongly non-linear transport phenomena in large complex networks containing regular subgraphs.

Many colleagues helped over the years to clarify many points throughout the book. Our thanks go to Sven Banisch, Bruno Cessac, Pierre Collet, Jean René Dawin, Sergey Dorogovtsev, Jürgen Jost, Dmitri Krioukov, Andreas Krüger, Tyll Krüger, Ricardo Lima, Zhi-Ming Ma, Rui Vilela Mendes, Walter Pauls, Filippo Petroni, Helge Ritter, Gabriel Ruget, Maurizio Serva, Ludwig Streit, Sören Wichmann.

We are further indebted to Dr. Christian Caron for competent advices and assistance in the completion of the final manuscript. His assistance is gratefully acknowledged.

Bielefeld *Philippe Blanchard*
Dimitri Volchenkov

Contents