

# 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.

## **Editorial and Programme Advisory Board**

Dan Braha, New England Complex Systems Institute and University of Massachusetts, Dartmouth, USA

Péter Érdi, Center for Complex Systems Studies, Kalamazoo College, USA and Hungarian Academy of Sciences, Budapest, Hungary

Karl Friston, Institute of Cognitive Neuroscience, University College London, London, UK

Hermann Haken, Center of Synergetics, University of Stuttgart, Stuttgart, Germany

Viktor Jirsa, Centre National de la Recherche Scientifique (CNRS), Université de la Méditerranée, Marseille, France

Janusz Kacprzyk, System Research, Polish Academy of Sciences, Warsaw, Poland

Scott Kelso, Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, USA

Markus Kirkilionis, Mathematics Institute and Centre for Complex Systems, University of Warwick, Coventry, UK

Jürgen Kurths, Nonlinear Dynamics Group, University of Potsdam, Potsdam, Germany

Linda Reichl, Center for Complex Quantum Systems, University of Texas, Austin, USA

Peter Schuster, Theoretical Chemistry and Structural Biology, University of Vienna, Vienna, Austria

Frank Schweitzer, System Design, ETH Zurich, Zurich, Switzerland

Didier Sornette, Entrepreneurial Risk, ETH Zurich, Zurich, Switzerland

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

© Springer-Verlag Berlin Heidelberg 2011

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

*Cover design:* SPi Publisher Services

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://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.

# Contents

<b>1</b>	<b>Introduction to Permutations, Markov Chains, and Partitions</b> .....	1
1.1	Permutations and Their Matrix Representations .....	1
1.2	Permutation Orbits and Fixed Points .....	3
1.3	Fixed Points and the Inclusion-Exclusion Principle .....	5
1.4	Finite Markov Chains .....	8
1.5	Birkhoff–von Neumann Theorem .....	9
1.6	Generating Functions .....	10
1.7	Partitions .....	13
1.7.1	Compositions .....	13
1.7.2	Multi-set Permutations .....	14
1.7.3	Weak Partitions .....	15
1.7.4	Integer Partitions .....	16
1.8	Concluding Remarks and Further Reading .....	17
<b>2</b>	<b>Worth Another Binary Relation: Graphs</b> .....	19
2.1	Binary Relations and Their Graphs .....	19
2.2	Representation of Graphs by Matrices .....	20
2.3	Algebraic Properties of Adjacency Operators .....	23
2.4	Perron–Frobenius Theory for Adjacency Matrices .....	24
2.5	Spectral Decomposition of Adjacency Operators .....	26
2.6	Adjacency and Walks on a Graph .....	28
2.7	Principal Invariants of the Graph Adjacency Matrix .....	30
2.8	Euler Characteristic and Genus of a Graph .....	33
2.9	Euler Characteristics and Genus of Complex Networks .....	35
2.10	Coloring a Graph .....	36
2.11	Shortest Paths in a Graph .....	38
2.12	Concluding Remarks and Further Reading .....	41



<b>3</b>	<b>Permutations Sieved Through Adjacency: Graph Automorphisms</b>	43
3.1	Graph Automorphisms	43
3.2	Nontrivial Graph Automorphisms and the Structure of Eigenvectors of the Adjacency Matrix	45
3.3	Automorphism Invariant Linear Functions of a Graph	47
3.3.1	Automorphism Invariant Stochastic Processes	48
3.3.2	Automorphism Invariant Harmonic Functions	49
3.4	Relations Between Eigenvalues of Automorphism Invariant Linear Functions	51
3.5	Summary	54
<b>4</b>	<b>Exploring Undirected Graphs by Random Walks</b>	55
4.1	Graphs as Discrete Time Dynamical Systems	56
4.2	Generating Functions of the Transition Probabilities	57
4.3	Cayley-Hamilton's Theorem for Random Walks	58
4.4	Stationary Distribution and Recurrence Time of Random Walks	59
4.5	Entropy of Random Walks Defined on a Graph	61
4.6	Hyperbolic Embeddings of Graphs by Transition Eigenvectors	64
4.7	Exploring the Shape of a Graph by Random Currents	68
4.8	Summary	72
<b>5</b>	<b>Embedding of Graphs in Probabilistic Euclidean Space</b>	73
5.1	Methods of Generalized Inverses in the Study of Graphs	73
5.2	Affine Probabilistic Geometry of Pseudo-inverses	75
5.3	Reduction to Euclidean Metric Geometry	76
5.4	Probabilistic Interpretation of Euclidean Geometry	77
5.5	Probabilistic Embedding of Simple Graphs	79
5.6	Group Generalized Inverse of the Laplace Operator for Directed Graphs	81
5.7	Summary	83
<b>6</b>	<b>Random Walks and Electric Resistance Networks</b>	85
6.1	Electric Resistance Network and its Probabilistic Interpretation	85
6.2	Dissipation and Effective Resistance in Electric Resistance Networks	87
6.3	Effective Resistance is Bounded Above by the Shortest Path Distance	89
6.4	Kirchhoff and Wiener Indexes of a Graph	90
6.5	Relation Between Effective Resistances and Commute Times	90
6.6	Summary	91

<b>7</b>	<b>Random Walks and Diffusions on Directed Graphs and Interacting Networks</b> .....	93
7.1	Random Walks on Directed Graphs .....	93
7.1.1	A Time Forward Random Walk .....	94
7.1.2	Backward Time Random Walks .....	94
7.1.3	Stationary Distributions of Random Walks on Directed Graphs .....	95
7.2	Laplace Operator Defined on Aperiodic Strongly Connected Directed Graphs .....	96
7.2.1	Bi-orthogonal Decomposition of Random Walks Defined on Strongly Connected Directed Graphs .....	98
7.3	Spectral Analysis of Self-adjoint Operators Defined on Directed Graphs .....	101
7.4	Self-adjoint Operators Defined on Interacting Networks .....	103
7.5	Summary .....	105
<b>8</b>	<b>Structural Analysis of Networks and Databases</b> .....	107
8.1	Structure and Function in Complex Networks and Databases .....	108
8.2	Graph Cut Problems .....	109
8.2.1	Weakly Connected Graph Components .....	110
8.2.2	Graph Partitioning Objectives as Trace Optimization Problems .....	112
8.3	Markov Chains Estimate Land Value in Cities .....	116
8.3.1	Spatial Networks of Urban Environments .....	117
8.3.2	Spectra of Cities .....	118
8.3.3	First-passage Times to Ghettos .....	120
8.3.4	Random Walks Estimate Land Value in Manhattan .....	121
8.4	Unraveling the Tangles of Language Evolution .....	123
8.4.1	Applying Phylogenetic Methods to Language Taxonomies .....	124
8.4.2	The Data Set We Have Used .....	125
8.4.3	The Relations Among Languages Encoded in the Matrix of Lexical Distances .....	126
8.4.4	The Structural Component Analysis on Language Data ..	128
8.4.5	Principal Structural Components of the Lexical Distance Data .....	131
8.4.6	Geometric Representation of the Indo-European Family .....	132
8.4.7	In Search of Lost Time .....	135
8.4.8	Evidence for Proto-Indo-Europeans .....	137
8.4.9	In Search of Polynesian Origins .....	140
8.4.10	Geometric Representation of Malagasy Dialects .....	144
8.4.11	Austronesian Languages Riding an Express Train .....	148

8.5	Markov Chain Analysis of Musical Dice Games .....	152
8.5.1	Musical Dice Game as a Markov Chain .....	153
8.5.2	Encoding of a Discrete Model of Music (MIDI) into a Transition Matrix .....	156
8.5.3	Musical Dice Game as a Generalized Communication Process .....	160
8.5.4	First Passage Times to Notes Resolve Tonality of Musical Dice Games .....	164
8.5.5	First Passage Times to Notes Feature a Composer .....	167
8.6	Summary .....	170
<b>9</b>	<b>When Feedbacks Matter: Epidemics, Synchronization, and Self-regulation in Complex Networks .....</b>	<b>171</b>
9.1	Susceptible-Infected-Susceptible Models in Epidemics .....	172
9.1.1	Dynamical Equation of the Epidemic Spreading in Scale Free Networks .....	172
9.1.2	Simplified Equation for Low Infection Rates .....	174
9.1.3	Stationary Solution of the Epidemic Equation for Low Infection Rates .....	175
9.1.4	Dynamical Solution of the Evolution Equation for Low Infection Rates .....	178
9.2	Epidemic Spreading in Evolutionary Scale Free Networks .....	180
9.3	Transitions to Intermittency and Collective Behavior in Randomly Coupled Map Networks .....	183
9.3.1	The Model of Random Networks of Coupled Maps .....	185
9.3.2	Spatiotemporal Intermittency and Collective Behavior .....	186
9.3.3	The Evolution of $G(N, k)$ with $k$ .....	193
9.4	Thermodynamics of Random Networks of Coupled Maps .....	196
9.5	Large Gene Expression Regulatory Networks .....	202
9.5.1	A Model of a Large Gene Expression Regulatory Networks .....	203
9.5.2	Numerical Analysis of Large Gene Expression Regulatory Networks .....	206
9.6	Mean Field Approach to the Large Transcription Regulatory Networks .....	213
9.7	Summary .....	217
<b>10</b>	<b>Critical Phenomena on Large Graphs with Regular Subgraphs .....</b>	<b>219</b>
10.1	Description of the Model and the Results .....	221
10.2	The Regular Subgraphs Viewed as Riemann Surfaces .....	222
10.3	Nonlinear Diffusions Through Complex Networks .....	224

10.4	Diffusion as a Generalized Brownian Motion .....	229
10.5	Scaling of a Scalar Field Coupled to a Complex Network .....	233
10.6	Summary.....	235
<b>References</b>	.....	<b>237</b>
<b>Glossary of Graph Theory</b>	.....	<b>258</b>
<b>Index</b>	.....	<b>259</b>