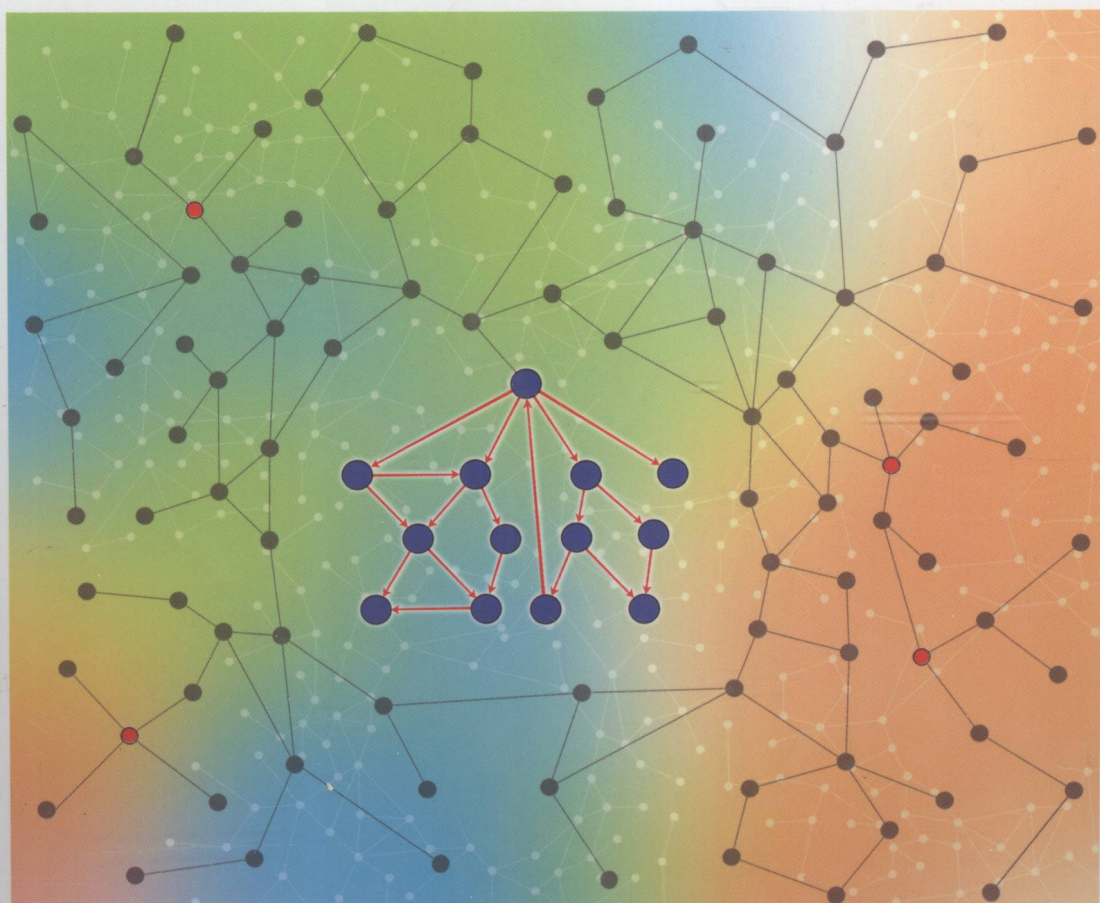


Edited by Matthias Dehmer  
and Frank Emmert-Streib

 WILEY-  
BLACKWELL

# Analysis of Complex Networks

From Biology to Linguistics



017  
A532

# Analysis of Complex Networks

From Biology to Linguistics

*Edited by*

*Matthias Dehmer and Frank Emmert-Streib*



E2009002729

WILEY-VCH Verlag GmbH & Co. KGaA

## The Editors

### **PD Dr. habil. Matthias Dehmer**

Vienna University of Technology  
Discrete Mathematics and Geometry  
Wiedner Hauptstraße 8–10  
1040 Vienna  
Austria  
and

University of Coimbra  
Center for Mathematics  
Apartado 3008  
3001-454 Coimbra  
Portugal

### **Prof. Dr. Frank Emmert-Streib**

Computational Biology and Machine Learning  
Center for Cancer Research and Cell Biology  
School of Medicine, Dentistry and Biomedical  
Sciences Queen's University Belfast  
97 Lisburn Road  
Belfast, BT9 7BL  
UK

■ All books published by Wiley-VCH are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

**Library of Congress Card No.:** applied for

### **British Library Cataloguing-in-Publication**

**Data:** A catalogue record for this book is available from the British Library.

### **Bibliographic information published by the Deutsche Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.d-nb.de>.

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form by photoprinting, microfilm, or any other means nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Printed in the Federal Republic of Germany  
Printed on acid-free paper

**Cover design** Adam Design, Weinheim  
**Typesetting** le-tex publishing services oHG, Leipzig  
**Printing** Strauss GmbH, Mörlenbach  
**Bookbinding** Litges & Dopf Buchbinderei GmbH, Heppenheim

**ISBN** 978-3-527-32345-6

**Analysis of Complex Networks**

From Biology to Linguistics

*Edited by*

*Matthias Dehmer and*

*Frank Emmert-Streib*

## ***Related Titles***

B.H. Junker, F. Schreiber

### **Analysis of Biological Networks**

2008

ISBN 978-0-470-04144-4

F. Emmert-Streib, M. Dehmer (Eds.)

### **Analysis of Microarray Data A Network-Based Approach**

2008

ISBN 978-3-527-31822-3

E. Keedwell, A. Narayanan

### **Intelligent Bioinformatics The Application of Artificial Intelligence Techniques to Bioinformatics Problems**

2005

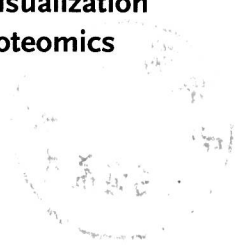
ISBN 978-0-470-02175-0

F. Azuaje, J. Dopazo (Eds.)

### **Data Analysis and Visualization in Genomics and Proteomics**

2005

ISBN 978-0-470-09439-6



## Preface

This book, *Analysis of Complex Networks: From Biology to Linguistics*, presents theoretical and practical results on graph-theoretic methods that are used for modeling as well as structurally investigating complex networks. Instead of focusing exclusively on classical graph-theoretic approaches, its major goal is to demonstrate the importance and usefulness of network-based concepts for scientists in various disciplines. Further, the book advocates the idea that theoretical as well as applied results are needed to enhance our knowledge and understanding of networks in general and as representations for various problems. We emphasize methods for analyzing graphs structurally because it has been shown that especially data-driven areas such as web mining, computational and systems biology, chemical informatics, and cognitive sciences profit tremendously from this field.

The main topics treated in this book can be summarized as follows:

- Information-theoretic methods for analyzing graphs
- Problems in quantitative graph theory
- Structural graph measures
- Investigating novel network classes
- Metrical properties of graphs
- Aspects in algorithmic graph theory
- Analytic methods in graph theory
- Network-based applications

*Analysis of Complex Networks: From Biology to Linguistics* is intended for an interdisciplinary audience ranging from applied discrete mathematics, artificial intelligence, and applied statistics to computer science, computational and systems biology, cognitive science, computational linguistics, machine learning, mathematical chemistry, and physics. Many colleagues, whether consciously or unconsciously, provided us with input, help, and support before and during the development of the present book. In particular we would like to thank Andreas Albrecht, Rute Andrade, Gökhan Bakır, Alexandru T. Balaban, Subhash Basak, Igor Bass, Natália Bebian, and

Danail Bonchev, Stefan Borgert, Mieczyslaw Borowiecki, Michael Drmota, Abdol-Hossein Esfahanian, Bernhard Gittenberger, Earl Glynn, Elena Konstantinova, Dmitrii Lozovanu, Alexander Mehler, Abbe Mowshowitz, Max Mühlhäuser, Arcady Mushegian, Paolo Oliveira, João da Providência, Horst Sachs, Heinz Georg Schuster, Helmut Schwegler, Chris Seidel, Fred Sobik, Doru Stefanescu, Thomas Stoll, John Storey, Kurt Varmuza, Bohdan Zelinka, and all the coauthors of this book and apologize to all those whose names have been mistakenly omitted. We would also like to thank our editors Andreas Sendtko and Gregor Cicchetti from Wiley-VCH; they were always available and extremely helpful. Last but not least, we would like to thank our families for their support and encouragement throughout the writing of the book.

Finally, we hope that this book helps the reader to understand that the presented field is multifaceted in depth and breadth and as such is inherently interdisciplinary. This is important to realize because it allows one to pursue a problem-oriented rather than field-oriented approach to efficiently tackling state-of-the-art problems in modern sciences.

Vienna and Belfast,  
March 2009

*Matthias Dehmer*  
*Frank Emmert-Streib*

## List of Contributors

### **Michael J. Barber**

Austrian Research Centers – ARC  
Division Systems Research  
Donau-City-Straße 1  
1220 Vienna  
Austria

### **Danail Bonchev**

Virginia Commonwealth University  
Center for the Study of  
Biological Complexity  
P.O. Box 842030  
Richmond, VA 23284-2030  
USA

### **Horst Bunke**

University of Bern  
Institute of Computer Science  
and Applied Mathematics  
Neubrückstraße 10  
3012 Bern  
Switzerland

### **Qiong Cheng**

Georgia State University  
Department of Computer Science  
Atlanta, GA 30303  
USA

### **Michel Deza**

École Normale Supérieure  
45 rue d'Ulm  
75005 Paris  
France  
and  
Japan Advanced Institute of  
Science and Technology  
1-1 Asahidai  
Nomi  
Ishikawa  
Japan

### **Dimitris Dimitropoulos**

European Bioinformatics Institute  
Genome Campus, Hinxton  
Cambridge CB10 1SD  
UK

### **Michael Drmota**

Vienna University of Technology  
Institute of Discrete Mathematics  
and Geometry  
Wiedner Hauptstraße 8/104  
1040 Vienna  
Austria



**Ernesto Estrada**

University of Strathclyde  
Institute of Complex Systems  
at Strathclyde  
Department of Physics and  
Department of Mathematics  
Glasgow G1 1XH  
UK

**Maria Fonoberova**

AIMdyn, Inc.  
Santa Barbara, CA 93101  
USA

**Bernhard Gittenberger**

Vienna University of Technology  
Institute of Discrete Mathematics  
and Geometry  
Wiedner Hauptstraße 8/104  
1040 Vienna  
Austria

**Adel Golovin**

European Bioinformatics Institute  
Genome Campus, Hinxton  
Cambridge CB10 1SD  
UK

**Ivan Gutman**

University of Kragujevac  
Faculty of Science  
P.O. Box 60  
34000 Kragujevac  
Serbia

**M. John**

European Bioinformatics Institute  
Genome Campus, Hinxton  
Cambridge CB10 1SD  
UK

**Marcus Kaiser**

Newcastle University  
School of Computing Science  
Newcastle-upon-Tyne NE1 7RU  
UK  
and  
Newcastle University  
Institute of Neuroscience  
Newcastle-upon-Tyne NE2 4HH  
UK

**Alexander Kononov**

Russian Academy of Sciences  
Sobolev Institute of Mathematics  
Novosibirsk  
Russia

**Matjaž Kovše**

University of Maribor  
Faculty of Natural Sciences  
and Mathematics  
Koroška cesta 160  
2000 Maribor  
Slovenia  
and  
Institute of Mathematics,  
Physics and Mechanics  
Jadranska 19  
1000 Ljubljana  
Slovenia

**Eugene Krissinel**

European Bioinformatics Institute  
Genome Campus, Hinxton  
Cambridge CB10 1SD  
UK

**Xueliang Li**

Nankai University  
Center for Combinatorics  
LPMC-TJKLC  
Tianjin 300071  
P.R. China

**Dmitrii Lozovanu**

Moldovan Academy of Sciences  
Institute of Mathematics  
and Computer Science  
Academiei str., 5  
Chisinau, MD-2005  
Moldova

**Alexander Mehler**

Goethe-Universität  
Frankfurt am Main  
Abteilung für geisteswis-  
sensschaftliche Fachinformatik/  
Department for Computing  
in the Humanities  
Georg-Voigt-Straße 4  
60325 Frankfurt am Main  
Germany

**Valia Mitsou**

The City University of New York  
Doctoral Program in Computer  
Science  
365 Fifth Avenue  
New York, NY 10016  
USA

**Abbe Mowshowitz**

The City College of New York  
Department of Computer Science  
Convent Avenue at 138th Street  
New York, NY 10031  
USA  
and  
The City University of New York  
Doctoral Program in Computer  
Science  
365 Fifth Avenue  
New York, NY 10016  
USA

**Manfred Paier**

Austrian Research Centers – ARC  
Division Systems Research  
Donau-City-Straße 1  
1220 Vienna  
Austria

**Christian M. Reidys**

Nankai University  
Center for Combinatorics  
LPMC-TJKLC  
Tianjin 300071  
P.R. China

**Kaspar Riesen**

University of Bern  
Institute of Computer Science and  
Applied Mathematics  
Neubrückstraße 10  
3012 Bern  
Switzerland

**Thomas Scherngell**

Austrian Research Centers – ARC  
Division Systems Research  
Donau-City-Straße 1  
1220 Vienna  
Austria

**Sergey Sevastyanov**

Russian Academy of Sciences  
Sobolev Institute of Mathematics  
Novosibirsk  
Russia

**Mikhail Shtogrin**

Steklov Mathematical Institute  
Gubkina str. 8  
117966 Moscow  
Russia

***Mathieu Dutour Sikirić***

Institute Rudjer Bošković  
Group for Satellite Oceanography  
10000 Zagreb  
Croatia

***Jennifer Simonotto***

Newcastle University  
School of Computing Science  
Newcastle-upon-Tyne NE1 7RU  
UK  
and  
Newcastle University  
Institute of Neuroscience  
Newcastle-upon-Tyne NE2 4HH  
UK

***Stefan Thurner***

Medical University of Vienna  
Complex Systems Research Group  
Währinger Gürtel 18–20  
1090 Vienna  
Austria  
and  
Santa Fe Institute  
1399 Hyde Park Road  
Santa Fe, NM 87501  
USA

***Alexander Zelikovsky***

Georgia State University  
Department of Computer Science  
Atlanta, GA 30303  
USA

***Jianbin Zhang***

Nankai University  
Center for Combinatorics  
LPMC-TJKLC  
Tianjin 300071  
P.R. China

## Contents

**Preface** XIII

**List of Contributors** XV

<b>1</b>	<b>Entropy, Orbits, and Spectra of Graphs</b>	<b>1</b>
	<i>Abbe Mowshowitz and Valia Mitsou</i>	
1.1	Introduction	1
1.2	Entropy or the Information Content of Graphs	2
1.3	Groups and Graph Spectra	4
1.4	Approximating Orbits	11
1.4.1	The Degree of the Vertices	13
1.4.2	The Point-Deleted Neighborhood Degree Vector	13
1.4.3	Betweenness Centrality	15
1.5	Alternative Bases for Structural Complexity	19
	References	21
<b>2</b>	<b>Statistical Mechanics of Complex Networks</b>	<b>23</b>
	<i>Stefan Thurner</i>	
2.1	Introduction	23
2.1.1	Network Entropies	25
2.1.2	Network Hamiltonians	27
2.1.3	Network Ensembles	28
2.1.4	Some Definitions of Network Measures	30
2.2	Macroscopics: Entropies for Networks	31
2.2.1	A General Set of Network Models Maximizing Generalized Entropies	32
2.2.1.1	A Unified Network Model	32
2.2.1.2	Famous Limits of the Unified Model	35
2.2.1.3	Unified Model: Additional Features	35
2.3	Microscopics: Hamiltonians of Networks – Network Thermodynamics	35

2.3.1	Topological Phase Transitions	36
2.3.2	A Note on Entropy	37
2.4	Ensembles of Random Networks – Superstatistics	39
2.5	Conclusion	42
	References	43
<b>3</b>	<b>A Simple Integrated Approach to Network Complexity and Node Centrality</b>	<b>47</b>
	<i>Danail Bonchev</i>	
3.1	Introduction	47
3.2	The Small-World Connectivity Descriptors	49
3.3	The Integrated Centrality Measure	52
	References	53
<b>4</b>	<b>Spectral Theory of Networks: From Biomolecular to Ecological Systems</b>	<b>55</b>
	<i>Ernesto Estrada</i>	
4.1	Introduction	55
4.2	Background on Graph Spectra	56
4.3	Spectral Measures of Node Centrality	58
4.3.1	Subgraph Centrality as a Partition Function	60
4.3.2	Application	61
4.4	Global Topological Organization of Complex Networks	62
4.4.1	Spectral Scaling Method	63
4.4.2	Universal Topological Classes of Networks	65
4.4.3	Applications	68
4.5	Communicability in Complex Networks	69
4.5.1	Communicability and Network Communities	71
4.5.2	Detection of Communities: The Communicability Graph	73
4.5.3	Application	74
4.6	Network Bipartivity	76
4.6.1	Detecting Bipartite Substructures in Complex Networks	77
4.6.2	Application	80
4.7	Conclusion	80
	References	81
<b>5</b>	<b>On the Structure of Neutral Networks of RNA Pseudoknot Structures</b>	<b>85</b>
	<i>Christian M. Reidys</i>	
5.1	Motivation and Background	85
5.1.1	Notation and Terminology	87
5.2	Preliminaries	88
5.3	Connectivity	90

5.4	The Largest Component	93
5.5	Distances in $n$ -Cubes	105
5.6	Conclusion	110
	References	111
<b>6</b>	<b>Graph Edit Distance – Optimal and Suboptimal Algorithms with Applications</b>	<b>113</b>
	<i>Horst Bunke and Kaspar Riesen</i>	
6.1	Introduction	113
6.2	Graph Edit Distance	115
6.3	Computation of GED	118
6.3.1	Optimal Algorithms	118
6.3.2	Suboptimal Algorithms	121
6.3.2.1	Bipartite Graph Matching	121
6.4	Applications	125
6.4.1	Graph Data Sets	125
6.4.2	GED-Based Nearest-Neighbor Classification	129
6.4.3	Dissimilarity-Based Embedding Graph Kernels	129
6.5	Experimental Evaluation	132
6.5.1	Optimal vs. Suboptimal Graph Edit Distance	133
6.5.2	Dissimilarity Embedding Graph Kernels Based on Suboptimal Graph Edit Distance	136
6.6	Summary and Conclusions	139
	References	140
<b>7</b>	<b>Graph Energy</b>	<b>145</b>
	<i>Ivan Gutman, Xueliang Li, and Jianbin Zhang</i>	
7.1	Introduction	145
7.2	Bounds for the Energy of Graphs	147
7.2.1	Some Upper Bounds	147
7.2.2	Some Lower Bounds	154
7.3	Hyperenergetic, Hypoenergetic, and Equienergetic Graphs	156
7.3.1	Hyperenergetic Graphs	156
7.3.2	Hypoenergetic Graphs	157
7.3.3	Equienergetic Graphs	157
7.4	Graphs Extremal with Regard to Energy	162
7.5	Miscellaneous	168
7.6	Concluding Remarks	169
	References	170

<b>8</b>	<b>Generalized Shortest Path Trees: A Novel Graph Class by Example of Semiotic Networks</b>	<b>175</b>
	<i>Alexander Mehler</i>	
8.1	Introduction	175
8.2	A Class of Tree-Like Graphs and Some of Its Derivatives	178
8.2.1	Preliminary Notions	178
8.2.2	Generalized Trees	180
8.2.3	Minimum Spanning Generalized Trees	186
8.2.4	Generalized Shortest Path Trees	190
8.2.5	Shortest Paths Generalized Trees	193
8.2.6	Generalized Shortest Paths Trees	195
8.2.7	Accounting for Orientation: Directed Generalized Trees	198
8.2.8	Generalized Trees, Quality Dimensions, and Conceptual Domains	204
8.2.9	Generalized Forests as Multidomain Conceptual Spaces	208
8.3	Semiotic Systems as Conceptual Graphs	212
	References	218
<b>9</b>	<b>Applications of Graph Theory in Chemo- and Bioinformatics</b>	<b>221</b>
	<i>Dimitris Dimitropoulos, Adel Golovin, M. John, and Eugene Krissinel</i>	
9.1	Introduction	221
9.2	Molecular Graphs	222
9.3	Common Problems with Molecular Graphs	223
9.4	Comparisons and 3D Alignment of Protein Structures	225
9.5	Identification of Macromolecular Assemblies in Crystal Packing	229
9.6	Chemical Graph Formats	231
9.7	Chemical Software Packages	232
9.8	Chemical Databases and Resources	232
9.9	Subgraph Isomorphism Solution in SQL	232
9.10	Cycles in Graphs	235
9.11	Aromatic Properties	236
9.12	Planar Subgraphs	237
9.13	Conclusion	238
	References	239
<b>10</b>	<b>Structural and Functional Dynamics in Cortical and Neuronal Networks</b>	<b>245</b>
	<i>Marcus Kaiser and Jennifer Simonotto</i>	
10.1	Introduction	245
10.1.1	Properties of Cortical and Neuronal Networks	246
10.1.1.1	Modularity	247
10.1.1.2	Small-World Features	247

10.1.1.3	Scale-Free Features	248
10.1.1.4	Spatial Layout	250
10.1.2	Prediction of Neural Connectivity	252
10.1.3	Activity Spreading	254
10.2	Structural Dynamics	255
10.2.1	Robustness Toward Structural Damage	255
10.2.1.1	Removal of Edges	256
10.2.1.2	Removal of Nodes	257
10.2.2	Network Changes During Development	258
10.2.2.1	Spatial Growth Can Generate Small-World Networks	258
10.2.2.2	Time Windows Generate Multiple Clusters	259
10.3	Functional Dynamics	260
10.3.1	Spreading in Excitable Media	260
10.3.1.1	Cardiac Defibrillation as a Case Study	261
10.3.1.2	Critical Timing for Changing the State of the Cardiac System	261
10.3.2	Topological Inhibition Limits Spreading	262
10.4	Summary	264
	References	266
<b>11</b>	<b>Network Mapping of Metabolic Pathways</b>	<b>271</b>
	<i>Qiong Cheng and Alexander Zelikovsky</i>	
11.1	Introduction	271
11.2	Brief Overview of Network Mapping Methods	273
11.3	Modeling Metabolic Pathway Mappings	275
11.3.1	Problem Formulation	277
11.4	Computing Minimum Cost Homomorphisms	277
11.4.1	The Dynamic Programming Algorithm for Multi-Source Tree Patterns	278
11.4.2	Handling Cycles in Patterns	280
11.4.3	Allowing Pattern Vertex Deletion	281
11.5	Mapping Metabolic Pathways	282
11.6	Implications of Pathway Mappings	285
11.7	Conclusion	291
	References	291
<b>12</b>	<b>Graph Structure Analysis and Computational Tractability of Scheduling Problems</b>	<b>295</b>
	<i>Sergey Sevastyanov and Alexander Kononov</i>	
12.1	Introduction	295
12.2	The Connected List Coloring Problem	296
12.3	Some Practical Problems Reducible to the CLC Problem	298
12.3.1	The Problem of Connected Service Areas	298
12.3.2	No-Idle Scheduling on Parallel Machines	300



12.3.3	Scheduling of Unit Jobs on a p-Batch Machine	301
12.4	A Parameterized Class of Subproblems of the CLC Problem	302
12.5	Complexities of Eight Representatives of Class $CLC(\mathcal{X})$	304
12.5.1	Three NP-Complete Subproblems	304
12.5.2	Five Polynomial-Time Solvable Subproblems	305
12.6	A Basis System of Problems	317
12.7	Conclusion	320
	References	322
<b>13</b>	<b>Complexity of Phylogenetic Networks: Counting Cubes in Median Graphs and Related Problems</b>	<b>323</b>
	<i>Matjaž Kovše</i>	
13.1	Introduction	323
13.2	Preliminaries	324
13.2.1	Median Graphs	325
13.2.1.1	Expansion Procedure	328
13.2.1.2	The Canonical Metric Representation and Isometric Dimension	328
13.3	Treelike Equalities and Euler-Type Inequalities	330
13.3.1	Treelike Equalities and Euler-Type Inequalities for Median Graphs	330
13.3.1.1	Cube-Free Median Graphs	332
13.3.1.2	$Q_4$ -Free Median Graphs	333
13.3.1.3	Median Grid Graphs	333
13.3.2	Euler-Type Inequalities for Quasi-Median Graphs	334
13.3.3	Euler-Type Inequalities for Partial Cubes	335
13.3.4	Treelike Equality for Cage-Amalgamation Graphs	336
13.4	Cube Polynomials	337
13.4.1	Cube Polynomials of Cube-Free Median Graphs	339
13.4.2	Roots of Cube Polynomials	340
13.4.2.1	Rational Roots of Cube Polynomials	340
13.4.2.2	Real Roots of Cube Polynomials	341
13.4.2.3	Graphs of Acyclic Cubical Complexes	341
13.4.2.4	Product Median Graphs	342
13.4.3	Higher Derivatives of Cube Polynomials	342
13.5	Hamming Polynomials	343
13.5.1	A Different Type of Hamming Polynomial for Cage-Amalgamation Graphs	344
13.6	Maximal Cubes in Median Graphs of Circular Split Systems	345
13.7	Applications in Phylogenetics	346
13.8	Summary and Conclusion	347
	References	348