COMPLEX NETWORKS

An Algorithmic Perspective



Kayhan Erciyes



COMPLEX NETWORKS

An Algorithmic Perspective

Kayhan Erciyes



CRC Press is an imprint of the Taylor & Francis Group, an **Informa** business

CRC Press Taylor & Francis Group 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742

© 2015 by Taylor & Francis Group, LLC CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed on acid-free paper Version Date: 20140728

International Standard Book Number-13: 978-1-4665-7166-2 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Erciyes, Kayhan.

Complex networks: an algorithmic perspective / author, Kayhan Erciyes.
pages cm

Includes bibliographical references and index.

ISBN 978-1-4665-7166-2 (hardback)

1. System analysis---Mathematics. 2. Computational complexity. 3. Algorithms. I. Title.

T57.E73 2014

003'.72--dc23 2014028247

Visit the Taylor & Francis Web site at http://www.taylorandfrancis.com

and the CRC Press Web site at http://www.crcpress.com

Dedication

To my family and all lifelong learners.

Preface

Recent technological advances in the last two decades have provided availability of enormous amounts of data about large networks consisting of hundreds and thousands of nodes. These so-called *complex networks* have non-trivial topological features and can vary from technological networks to social networks to biological networks. The study of complex networks, sometimes referred to as *network science*, has become a fundamental research area since then in various disciplines such as mathematics, statistics, computer science, physics and biology.

These seemingly unrelated networks experimentally have been shown to have common properties such as low average distance between their nodes, high local densities and degree distributions with few high degree nodes and many low degree nodes. Modeling and analysis of these networks based on experiments and evaluations has become an active and attractive area of research with many potential results. Graphs have been widely and successfully used to model computer networks, and it seems graph theory is a promising tool also for complex networks. Although there has been considerable amount of study and research on the modeling and analysis of complex networks, the algorithms for these networks are relatively less investigated.

Whether a complex network is man-made such as the Internet or not such as a protein interaction network, predicting its behavior is not a trivial task. Understanding the functionality of complex networks provides us with insight to predict their behavior and once we can estimate the behavior of a complex network based on its functionality, we may be able to control its functionality. For example, if we can understand the spreading pattern of an epidemic disease which in fact is a complex network, we can estimate where it will most likely spread and can then take precautions to stop it. In summary, controlling the functionality of a complex network is one of the fundamental reasons to study these networks.

As a first step in their study, we need to specify and classify the properties of complex networks. We can then use analytical tools to identify and analyze these properties to understand them better. As an example, a group of entities that make the complex network may be more closely related to each other than the rest of the network. These groups called the *clusters* may have important processing effects on

xxiii

the overall functionality of the network. If we can detect clusters in a social network for example, we can locate these intense regions of activity in that network after which we can investigate the role of these clusters in the functioning of the whole network. Detection of these properties may be visually possible in a small network of few tens of nodes but for a complex network of hundreds of thousands of nodes, we need analytical tools and computational methods. Properties of complex networks such as clustering depend on their topological properties and study of topological properties of these networks provides insight to their functioning.

This book is about specifying, classifying, designing and implementation of mostly sequential, and also parallel and distributed algorithms that can be used to analyze mostly the static properties of complex networks. Our aim has been to identify and describe a repertoire of algorithms that may be of use for any complex network. The starting point was to identify fundamental and mostly topological properties which are static in general and evaluation of these properties which requires efficient algorithms. The problems encountered are NP-hard in many cases and we need to rely on approximation algorithms where sub-optimal solutions in polynomial time can be found. Sometimes, using heuristic algorithms may be the only choice and extensive tests are needed to support that the algorithm works for most of the cases. Parallel algorithms aim at performance and provide efficiency for computation intensive tasks to be performed and we present several parallel algorithms. Distributed algorithms reach a decision by local information and are usually the only choice in computer networks. Design and implementation of parallel and distributed algorithms received little attention for complex networks in the past and are promising areas for potential research in these networks.

An important static topological property of a complex network is its *centrality* measure which shows the importance of a node or an edge in the network. *Clustering* or *community detection* is another fundamental topological complex network property and provides information about groups of nodes in the complex networks which have closer relations among them than the rest of the network. Discovery of *motifs* which are patterns occurring more than any other patterns, possibly indicating a basic function in the network is another important property of the complex networks. Evaluation of such measures using sequential, approximation, heuristic, parallel and distributed algorithms provides us with significant information about a particular network. We can then improve the modeling of the network, understand its function better and possibly predict the behavior of the network.

The style we have adopted is to keep everything as simple as possible, to be able to guide a beginning researcher or a student with virtually no background in the field of complex networks. The language used is mathematical rather than descriptive most of the time; however, a basic discrete mathematics and algorithms background at undergraduate level is sufficient to follow the material. Again, to aid the beginner in the field, most of the algorithms are provided in ready-to-be-executed form to test.

The book is divided into three parts. Part I provides the basic background in terms of the graph theory; algorithms and complexity and the specification of the parameters for the analysis of complex networks. In Part II, we provide a survey of important algorithms for the analysis of complex networks, starting with distance and centrality

algorithms. We then describe algorithms to construct and detect special subgraphs in complex networks, which may be used for other tasks such as clustering. A survey of data and clustering algorithms is also presented and this part concludes by the description of the network motif discovery algorithms. Part III is about case studies of complex networks and we show the implementation of some of the algorithms we have described in real-life networks such as the protein interaction networks, the social networks and the computer networks.

I would first like to thank graduate students at Izmir University who were concurrently taking a related course at the time of the writing of this text and were presented part of the material. I would like to thank Esra Ruzgar and Can Ileri for their feedback and especially Vedat Kavalci for proofreading of several chapters. I would like to thank CRC Press publisher Rich O'Hanley who has always been very kind, supportive and encouraging. I also thank Stephanie Morkert who was prompt and ever willing to help in the editing process and Michele Dimont for final editing.

K. Erciyes Izmir, Turkey

Contents

Lis	t of F	igures
Lis	t of T	ables
Pr	eface	
1	Intro	duction
	1.1	Overview
	1.2	Real-world Complex Networks
		1.2.1 Technological Networks
		1.2.2 Information Networks
		1.2.3 Social Networks
		1.2.4 Biological Networks
	1.3	Topological Properties of Complex Networks
	1.4	Algorithmic Challenges
	1.5	Outline of the Book
	Refe	rences
SI	ECT	ON I: BACKGROUND
2	Gra	ph Theory
	2.1	Basics
	2.2	Subgraphs
	2.3	Graph Isomorphism
	2.4	Types of Graphs
	2.5	Paths and Cycles
	2.6	Connectivity
	2.7	Trees
	2.8	Graph Representations

	2.9	openin reprine or surpris	19
		2.9.1 Eigenvalues and Eigenvectors	20
		2.9.2 The Laplacian Matrix	21
	2.10		22
	Refer	rences	24
3	Algo	rithms and Complexity	27
	3.1		27
	3.2		29
	3.3		31
	3.4		32
	3.5		33
	2.0		34
		3.5.2 Depth-first Search	36
	26		
	3.6		37
	3.7	, ,	39
	3.8		41
			41
			43
			44
			45
		1	45
		3.8.2.4 Independent Set to Clique	46
	3.9	Coping with NP Completeness	48
		3.9.1 Backtracking	48
			49
	3.10		50
			53
			54
			54
	3.12		57
			59
			61
			01
4	Ana	lysis of Complex Networks	63
	4.1	Introduction	63
	4.2	Vertex Degrees	64
		4.2.1 Degree Sequence	65
		4.2.2 Degree Distribution	66
	4.3	Communities	67
		4.3.1 Clustering Coefficient	68
	4.4	The Matching Index	70
	4.5	Centrality	70
	4.6	Network Motifs	71
	4.7	Models	7.1

	4.8 Refe	4.7.1 4.7.2 4.7.3 Chapter	Classical Random Networks Small World Networks Scale-Free Networks r Notes	72 72 73 75 77
SI	ECTI	ON II	: ALGORITHMS	79
5	Dista	ance and	d Centrality	81
	5.1		iction	81
	5.2		g Distances	81
		5.2.1	Average Distance	83
		5.2.2	Dijkstra's Single Source Shortest Paths Algorithm	83
		5.2.3	Floyd-Warshall All Pairs Shortest Paths Algorithm	84
	5.3	Centra	lity	87
		5.3.1	Degree Centrality	87
			5.3.1.1 A Distributed Algorithm for k-hop Degree Cen-	
			trality	89
		5.3.2	Closeness Centrality	91
		5.3.3	Stress Centrality	92
		5.3.4	Betweenness Centrality	93
			5.3.4.1 Newman's Algorithm	94
			5.3.4.2 Brandes' Algorithm	95
		5.3.5	Eigenvalue Centrality	96
	5.4	Chapte	er Notes	97
	Refe			99
6	Spe	cial Sub	graphs	101
	6.1	Introd	uction	101
	6.2		nal Independent Sets	102
	6.3	Domir	nating Sets	103
		6.3.1	A Greedy MDS Algorithm	105
		6.3.2	Guha-Khuller First MCDS Algorithm	106
		6.3.3	Guha-Khuller Second MCDS Algorithm	106
	6.4	Match		108
		6.4.1	A Maximal Unweighted Matching Algorithm	109
		6.4.2	A Maximal Weighted Matching Algorithm	109
	6.5	Vertex	Cover	110
		6.5.1	A Minimal Connected Vertex Cover Algorithm	112
		6.5.2	A Minimal Weighted Vertex Cover Algorithm	113
	6.6	A Dist	tributed Algorithm for MWVC Construction	114
	6.7	Chapte	er Notes	116
	Ref		************	118

7.2 Types of Data Clustering 7.3 Agglomerative Hierarchical Clustering 7.4 k-means Algorithm 7.5 Nearest Neighbor Algorithm 7.6 Fuzzy Clustering 7.7 Density-based Clustering 7.8 Parallel Data Clustering 7.9 Chapter Notes References 8 Graph-based Clustering 8.1 Introduction 8.2 Graph Partitioning 8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3.1 MST-based Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.2.2 The Second Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm	7	Data	Cluster	ing												121
7.2 Types of Data Clustering 7.3 Agglomerative Hierarchical Clustering 7.4 k-means Algorithm 7.5 Nearest Neighbor Algorithm 7.6 Fuzzy Clustering 7.7 Density-based Clustering 7.8 Parallel Data Clustering 7.9 Chapter Notes References 8 Graph-based Clustering 8.1 Introduction 8.2 Graph Partitioning 8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3.1 MST-based Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.2.2 The Second Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm 9.3.2.1 Ullman's Algorithm		7.1	Introduc	ction				 of 5		X II	v				(6)	121
7.3 Agglomerative Hierarchical Clustering 7.4 k-means Algorithm 7.5 Nearest Neighbor Algorithm 7.6 Fuzzy Clustering 7.7 Density-based Clustering 7.8 Parallel Data Clustering 7.9 Chapter Notes References 8 Graph-based Clustering 8.1 Introduction 8.2 Graph Partitioning 8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3.1 MST-based Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.2.2 The Second Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		7.2														122
7.4 k-means Algorithm 7.5 Nearest Neighbor Algorithm 7.6 Fuzzy Clustering 7.7 Density-based Clustering 7.8 Parallel Data Clustering 7.9 Chapter Notes References 8 Graph-based Clustering 8.1 Introduction 8.2 Graph Partitioning 8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.2.2 The Second Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		7.3														123
7.5 Nearest Neighbor Algorithm 7.6 Fuzzy Clustering 7.7 Density-based Clustering 7.8 Parallel Data Clustering 7.9 Chapter Notes References 8 Graph-based Clustering 8.1 Introduction 8.2 Graph Partitioning 8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3.1 MST-based Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		7.4	~~													127
7.6 Fuzzy Clustering 7.7 Density-based Clustering 7.8 Parallel Data Clustering 7.9 Chapter Notes References 8 Graph-based Clustering 8.1 Introduction 8.2 Graph Partitioning 8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3.1 MST-based Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.2.2 The Second Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		7.5														132
7.7 Density-based Clustering 7.8 Parallel Data Clustering 7.9 Chapter Notes References 8 Graph-based Clustering 8.1 Introduction 8.2 Graph Partitioning 8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3.1 MST-based Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motif Significance 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm			Fuzzy (Clustering			x x	 ,		,				3 3		132
7.8 Parallel Data Clustering 7.9 Chapter Notes References 8 Graph-based Clustering 8.1 Introduction 8.2 Graph Partitioning 8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3.1 MST-based Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		7.7	-													134
7.9 Chapter Notes References 8 Graph-based Clustering 8.1 Introduction 8.2 Graph Partitioning 8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3 Graph Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2 I Ullman's Algorithm		7.8	-													137
References 8 Graph-based Clustering 8.1 Introduction 8.2 Graph Partitioning 8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3.1 MST-based Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm																138
8.1 Introduction 8.2 Graph Partitioning 8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3.1 MST-based Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm																141
8.1 Introduction 8.2 Graph Partitioning 8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3.1 MST-based Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm	0			101 4 1												1.42
8.2 Graph Partitioning 8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3 Graph Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm	8	-														143
8.2.1 BFS-based Partitioning 8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3.1 MST-based Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm																143
8.2.2 Kernighan-Lin Algorithm 8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.2.5 Parallel Partitioning 8.3 Graph Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		8.2														144
8.2.3 Spectral Bisection 8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.2.6 Graph Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.2 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm						-										145
8.2.4 Multi-level Partitioning 8.2.5 Parallel Partitioning 8.3.6 Graph Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm				_	_											147
8.2.5 Parallel Partitioning. 8.3 Graph Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm				1												
8.3 Graph Clustering 8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm																
8.3.1 MST-based Clustering 8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		0.0			_											
8.3.2 Clustering with Clusterheads 8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		8.3		_												
8.4 Discovery of Dense Subgraphs 8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm																
8.4.1 Definitions 8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References Polymer Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		0.4			_											
8.4.2 Clique Algorithms 8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		8.4			_											
8.4.2.1 The First Algorithm 8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm																
8.4.2.2 The Second Algorithm 8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm			8.4.2													
8.4.3 k-core Algorithm 8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm						-										
8.5 Chapter Notes References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm																
References 9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		0 =														
9 Network Motif Discovery 9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm			Α.													2.00
9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		Refe	erences			W 18 10 10 10						 41	٠.	٠		167
9.1 Introduction 9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm	9	Netv	work M	otif Disco	very			 								169
9.2 Network Motifs 9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		9.1	Introdu	action				 	,			 2	v 4			169
9.2.1 Measures of Motif Significance 9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm		9.2														
9.2.2 Generating Null Models 9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm																
9.2.3 Hardness of Motif Discovery 9.3 Subgraph Isomorphism 9.3.1 Vertex Invariants 9.3.2 Algorithms 9.3.2.1 Ullman's Algorithm			9.2.2													
9.3 Subgraph Isomorphism			9.2.3													
9.3.1 Vertex Invariants		9.3	Subgra													
9.3.2.1 Ullman's Algorithm																
9.3.2.1 Ullman's Algorithm																
· ·																
9.3.2.2 <i>Nauty</i> Algorithm				9.3.2.2		-										
9.3.2.3 VF2 Algorithm																
9.3.2.4 BM1 Algorithm																

	11.4	Commu	nity Detec	ction Algorithms	. 233
				weenness-based Algorithm	
			-	Networks	
		11.4.3	Random V	Walk Centrality	. 238
				y-based Algorithm	
	11.5				
12				Web	
				(*** • (* * * * * * * * * * * * * * * *	
	12.2	The Int			
		12.2.1	Services.		
			12.2.1.1	Services of Connection	
			12.2.1.2	Circuit and Packet Switching	
			12.2.1.3	Internet Protocol Suite	. 247
		12.2.2	Analysis		. 248
	12.3	The We	eb		. 249
		12.3.1	The Web	Graph	. 250
			12.3.1.1	Properties	
			12.3.1.2	Evolving Model	
			12.3.1.3	Copying Model	
			12.3.1.4	Growth-deletion Model	
			12.3.1.5	Multi-layer Model	
			12.3.1.6	Cyber Community Detection	
		1232		llysis	
		14.3.4	12.3.2.1	Hubs and Authorities	
				Page Rank Algorithm	
	12.4	Chanto			
		Chapte rences			
	IXCIC	rences			. 265
13				vorks	
	13.2			ithms	
		13.2.1	Lowest-I	D Algorithm	. 271
		13.2.2	Dominat	ing Set-based Clustering	. 273
		13.2.3	Spanning	Tree-based Clustering	. 275
	13.3			etworks	
				ture	
		13.3.2	Commun	nity Detection	
		13.3.3		are	
			13.3.3.1	MobiSoC	
			13.3.3.2	MobiClique	
			13.3.3.3	SAMOA	
			13.3.3.4	Yarta	
	13.4	Chante	er Notes	Taita	
	200	Cimple	110000	* * * * * * * * * * * * * * * * * * * *	. 400

	Contents	xiii
References	* * * * * *	 287
Index		 291

List of Figures

1.1	A graph representing a network	2
1.2	The telephone network	3
1.3	An example citation network	4
2.1		12
2.2	a) A graph with four vertices. b) Its complement. c) A clique of order 5	12
2.3	a) A graph. b) Its subgraph which is not induced. c) Its spanning	14
Lund		13
2.4	a) First graph. b) Second graph. c) Union of two graphs. d) Inter-	
	section of two graphs	14
2.5	Two isomorphic graphs	14
2.6	a) A 2-3 bipartite graph. b) A complete 3-4 bipartite graph	15
2.7	a) A weighted graph. b) A directed graph	16
2.8	Paths and cycles in a graph where $\{a, g, f, d, e\}$ is a path;	
	$\{b,c,f,g,b\}$ is a cycle; edge (c,g) is a chord. The $\{a,b,c,d,e,f,g,h,a\}$	
	cycle is Hamiltonian, and $\{a, h, g, b, c, f, e, d\}$ is an Eulerian path .	17
2.9	a) A graph with two components. b) A cutset of a graph	
	$\{(b,c),(f,g)\}$	17
2.10	a) A bridge ($\{ae\}$) of a graph. b) A cutpoint ($\{c\}$) of a graph	18
2.11	a) A spanning tree. b) A minimum spanning tree rooted at vertex e	19
2.12	a) An example graph. b) Its adjacency list. c) Its adjacency matrix.	
	d) Its incidence matrix	20
2.13	The example graph for Ex. 1	23
2.14	The example graph for Ex. 2	23
2.15	The example graph for Ex. 3	23
2.16	The example graph for Ex. 4 and 10	24
3.1	O(g(n)) function	3(

3.2	$\Omega(g(n))$ function	30
3.3 3.4	$\Theta(g(n))$ function	31
2.4	step	34
3.5	BFS algorithm execution	35
3.6	DFS algorithm execution	37
3.7	Prim_MST example	40
3.8	Relations between P, NP, NPC and NPH	43
3.9	CSAT example	44
3.10	Reduction from 3-SAT to IND	46
3.11	Reduction from independent set to vertex cover. a) An independent	
	set. b) A vertex cover from this independent set	46
3.12	Reduction from independent set to clique	47
3.13	Diagram showing reductions between NPC problems	47
3.14	Backtracking algorithm example	49
3.15	Approximate vertex cover example	52
3.16	The greedy heuristic	53
3.17	Parallel calculation of PI	56
3.18	Example graph for Ex. 2 and Ex. 8	60
3.19	Example graph for Ex. 3	60
3.20	Example graph for Ex. 6	60
4.1	Degree sequence of graphs	65
4.2	Graph isomorphism. Graphs in (a) and (b) are isomorphic but (a)	
	and (c), and (b) and (c) are not isomorphic	66
4.3	Degree distribution of graphs. a) The degree distribution of a ho-	
	mogeneous graph where all vertices have degree r . b) The degree	
	distribution of a random graph. c) The degree distribution of a scale-	
	free network such as the world wide web	67
4.4	Clusters in a network	67
4.5	Clustering coefficients of graphs	69
4.6	The matching index	70
4.7	Network motifs with three nodes	71
4.8	WS-network. a) The original regular network. b) Few rewiring pre-	
	serves clustering and results in low average path length. c) Many	
	rewiring approaches ER-model with low clustering and low aver-	
	age path length	73
4.9	The preferential attachment rule	74
4.10	Example for Ex. 1	76
4.11	Example graphs for Ex. 2 and 3	77
4.12	Example graph for Ex. 5	77
5.1	Diameter and eccentricities in a graph	82
5.2	Diameters and radii of complete graphs	

5.3	Dijkstra_SSSP algorithm execution. The vertices on the shortest	
	path to source vertex a are shown by arrows	85
5.4	FW_APSP execution example	87
5.5	Degree centrality example	88
5.6	Edge and vertex disjoint paths	89
5.7	One-rank centrality example	90
5.8	Output of Two_rank algorithm	90
5.9	Closeness centrality calculation	91
5.10	Closeness centrality in a weighted graph	92
5.11	Betweenness centrality algorithm example	95
5.12	Example graph for Ex. 1 and 2	97
5.13	Example graph for Ex. 3	98
5.14	Example graph for Ex. 4	98
5.15	Example graph for Ex. 5	98
5.16	Example graph for Ex. 6	99
6.1	Independent set examples. a) An independent set of size 2 which is	
	also maximal. b) A maximum independent set of size 4 in the same	
	graph	102
6.2	MIS_Alg execution example	103
6.3	Linear network independent set example	104
6.4	Dominating set examples. a) A minimal dominating set. b) A mini-	
	mum dominating set in the same graph	104
6.5	Execution of MDS_Alg1 in two iterations	106
6.6	Execution of MCDS_Alg1 in three iterations	107
6.7	Execution of MCDS_Alg2 in three iterations	107
6.8	Matching examples. a) A maximal matching of a graph. b) A max-	
	imum matching of the same graph	108
6.9	The maximal unweighted matching algorithm example	109
6.10	The maximal and maximum weighted matching algorithm example	110
6.11	Vertex covers in the same graph. a) A minimum vertex cover of size	
	3. b) A minimum connected vertex cover of size 4	111
6.12	a) A minimum unweighted. b) A minimum weighted vertex cover	
	of the same graph where weights are shown next to vertices	112
6.13	MCVC_Alg execution in a sample graph	113
6.14	MWVC_Alg execution in a sample graph	114
6.15	The three rounds of the distributed vertex cover algorithm \dots .	116
6.16	Example graph for Ex. 1	117
6.17	Example graph for Ex. 2	117
6.18	Example graph for Ex. 5	118
6.19	Example graph for Ex. 6	118
7.1	a) The nested clusters. b) The corresponding dendogram	123
7.2	a) Single linkage. b) Complete linkage. c) Average linkage	124
7.3	Hierarchical algorithm first iteration	126