Intelligent Systems Reference Library

Volume 148

The aim of this series is to publish a Reference Library, including novel advances and developments in all aspects of Intelligent Systems in an easily accessible and well structured form. The series includes reference works, handbooks, compendia, textbooks, well-structured monographs, dictionaries, and encyclopedias. It contains well integrated knowledge and current information in the field of Intelligent Systems. The series covers the theory, applications, and design methods of Intelligent Systems. Virtually all disciplines such as engineering, computer science, avionics, business, e-commerce, environment, healthcare, physics and life science are included. The list of topics spans all the areas of modern intelligent systems such as: Ambient intelligence, Computational intelligence, Social intelligence, Computational neuroscience, Artificial life, Virtual society, Cognitive systems, DNA and immunity-based systems, e-Learning and teaching, Human-centred computing and Machine ethics, Intelligent control, Intelligent data analysis, Knowledge-based paradigms, Knowledge management, Intelligent agents, Intelligent decision making, Intelligent network security, Interactive entertainment, Learning paradigms, Recommender systems, Robotics and Mechatronics including human-machine teaming, Self-organizing and adaptive systems, Soft computing including Neural systems, Fuzzy systems, Evolutionary computing and the Fusion of these paradigms, Perception and Vision, Web intelligence and Multimedia.

More information about this series at http://www.springer.com/series/8578

Miloš Savić · Mirjana Ivanović
Lakhmi C. Jain

Complex Networks in Software, Knowledge, and Social Systems

Springer

Miloš Savić
Faculty of Sciences, Department of Mathematics and Informatics
University of Novi Sad
Novi Sad
Serbia

Mirjana Ivanović
Faculty of Sciences, Department of Mathematics and Informatics
University of Novi Sad
Novi Sad
Serbia

Lakhmi C. Jain
Centre for Artificial Intelligence, Faculty of Engineering and Information Technology
University of Technology Sydney
Sydney, NSW
Australia

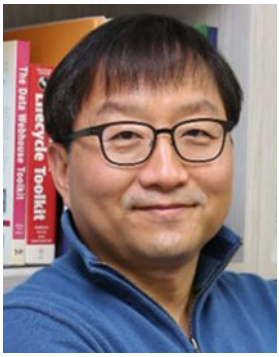
ISSN 1868-4394 ISSN 1868-4408 (electronic)
Intelligent Systems Reference Library
ISBN 978-3-319-91194-6 ISBN 978-3-319-91196-0 (eBook)
https://doi.org/10.1007/978-3-319-91196-0

Library of Congress Control Number: 2018940621

Printed on acid-free paper

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

Foreword



We are living in the information age being surrounded by diverse types of complex networks. The study of complex networks has gained a significant research interest in recent years, mostly because of their ubiquitous presence in nature and society, leading to an inter-disciplinary research field involving researchers from all major scientific disciplines.

This monograph deals with three types of complex networks describing the structure of software systems, the semantic web ontologies, and the self-organized social structure of research collaboration. In Chaps. 1 and 2, the authors give an overview of fundamental concepts, metrics, methods, and models important in studying real-world complex networks. As the main research contribution of the monograph, they propose and empirically validate several novel methods to analyze complex networks in which nodes are enriched with the domain-independent structural metrics and the metrics from a particular domain (i.e., software metrics, ontology metrics, and metrics of research performance, respectively).

Software networks are directed graphs that represent the dependencies among software entities present in a complex software system. One software system can be represented by several software networks reflecting its structure at different

granularity levels. For example, the design structure of an object-oriented software system is typically described by three different kinds of software networks that depict dependencies among methods, classes, and packages (i.e., modules or namespaces). The applications of software networks are numerous, including the analysis of software systems using graph-based methods, computation of software design metrics, program comprehension and visualization, reverse engineering of software systems, identification of key software components, identification of design flaws in source code, analysis of change impact, and prediction of defects in software systems.

The authors give a comprehensive overview of previous empirical studies of software networks in Chap. 3. In the same chapter, they introduce a novel methodology to examine coupling and cohesion in software systems, found on enriched software networks. The authors also propose domain-independent graph clustering evaluation metrics for measuring the cohesiveness of software entities indicating their benefits over commonly used software cohesion metrics. The case studies presented here show that the proposed methodology has both theoretical and practical relevance. It enables a deeper understanding of phenomena that are commonly considered as indicators of poorly designed software systems (i.e., high coupling, low cohesion, and large cyclic dependencies). Additionally, it can be utilized for software engineering practitioners to identify keys, distinctive features of highly coupled software entities, software entities involved in cyclic dependencies, and software entities causing low cohesion providing valuable information for software development, testing, and maintenance activities.

The ontology formally describes the concepts and relationships in a domain of discourse. Ontologies have a prominent role in the development of the semantic web where they serve as shared and agreed-upon knowledge models enabling information reuse and interoperability. Ontologies and networks are very closely related—an ontology is a set of axioms inducing a semantic network of ontological entities present in the ontology. In Chap. 4, the authors show that modular semantic web ontologies represented by enriched ontology networks can be studied and evaluated in the same way as software systems represented by enriched software networks.

The last four chapters of the monograph are devoted to co-authorship networks. Co-authorship networks are social networks in which nodes represent researchers and links do research collaborations among them. In Chap. 5, the authors first discuss several graph-based representations of research collaboration and several ways to quantify its strength. In Chap. 6, they focus on the author name disambiguation problem appearing when extracting a co-authorship network from a bibliographic database in which authors are not uniquely identified. They provide a comprehensive overview of existing heuristic and machine learning approaches to solving the author name disambiguation. Then, the authors propose a novel supervised network-based method for disambiguating author names in bibliographic data.

Research collaboration is one of the fundamental determinants of contemporary science. The study of co-authorship networks is thus crucial for understanding the social structure and evolution of research communities. In Chap. 7, the authors give a thorough overview of existing empirical studies of co-authorship networks and identify their common structural and evolutionary properties. In Chap. 8, they propose a novel methodology based on enriched co-authorship networks to analyze the structure and evolution of research collaboration. The accompanying case study shows that the proposed methodology enables an in-depth analysis of research collaboration and its relationships with other indicators of research performance.

In my opinion, researchers and students interested in complex networks may benefit a lot from this monograph in two ways. First, the monograph provides a comprehensive and up-to-date overview of studies of complex networks from three important domains. Second, it introduces new methods to study complex networks enriched with domain-dependent metrics that are empirically validated with relevant and interesting case studies. The monograph may be also useful for researchers and practitioners in software engineering, ontology engineering, and scientometrics since it gives a network-based perspective on important issues from those three disciplines. I have recognized the significance of the original research contributions presented in the monograph and thus expect that they will motivate further research directions and novel applications.

Seoul, Korea

Prof. Sang-Wook Kim
Hanyang University

Preface

A wide variety of complex natural, engineered, conceptual, and social systems of high technological and scientific importance can be represented by networks—structures that describe relations, dependencies, and interactions between constituent parts of a complex system. Well-known examples of complex networked systems include technological systems such as Internet, power grids, telecommunication, and transportation networks; social systems such as academia, corporations, markets, and online communities; biological systems such as brain, metabolic pathways, and gene regulatory networks; and ecological systems such as food chains. In order to understand, control, or improve a complex system composed out of a large number of inter-related parts, it is necessary to quantify, characterize, and comprehend the structure and evolution of underlying complex networks.

The focus of this monograph is on complex networks from three domains: (1) networks extracted from source code of computer programs that represent the design of software systems, (2) networks extracted from source code of semantic web ontologies that describe the structure of shared and reusable knowledge, and (3) networks extracted from bibliographic databases that reflect scientific collaboration. In the monograph, we present novel methods for analyzing *enriched* software, ontology, and co-authorship networks, i.e., complex networks in which nodes are enriched with both domain-dependent metrics (software, ontology, and metrics of research performance, respectively) and domain-independent metrics used in complex network analysis.

The monograph is intended primarily for researchers, teachers, and students interested in complex networks and data analysis and mining. Additionally, it may also be interesting for researchers dealing with software engineering, ontology engineering, and scientometrics since it addresses topics from those disciplines within the framework of complex networks.

The monograph consists of three major parts entitled “Introduction”, “Software and Ontology Networks: Complex Networks in Source Code”, and “Co-authorship Networks: Social Networks of Research Collaboration”.

Part I. In Chap. 1, we make an introduction to complex networks and outline our main research contributions presented in this monograph. The next chapter, Chap. 2, presents fundamental complex network measures, algorithms, and models. Those two chapters contain the necessary theoretical background and preliminaries used in the rest of the monograph.

Part II. The second part of the monograph is devoted to software and ontology networks. Those two types of complex networks, although representing two different kinds of complex man-made systems, have one important thing in common —they show dependencies between entities present in a system described in a formal language. In Chap. 3, after presenting an overview of the literature investigating software networks, we propose and empirically evaluate a novel methodology to study the structure of enriched software networks. In Chap. 4, we apply the same methodology to study the design of a large-scale modularized ontology.

Part III. The last part of the monograph is focused on co-authorship networks. This part contains four chapters. In Chap. 5, we discuss different models of co-authorship networks, different schemes to quantify the strength of research collaboration, different types of co-authorship networks, and their main applications. Chapter 6 is devoted to the extraction of co-authorship networks from bibliographic databases. We start with an overview of existing approaches to the author name disambiguation problem and their actual utilization in empirical studies analyzing co-authorship networks. In the same chapter, we study the performance of various string similarity metrics for identifying name synonyms in bibliographic records. We present a novel network-based method to disambiguate author names and investigate the impact of author name disambiguation to the structure of co-authorship networks. A comprehensive overview of studies dealing with the analysis of co-authorship networks is given in Chap. 7. Finally, in Chap. 8, we propose a novel methodology to study the structure and evolution of enriched co-authorship networks and demonstrate it on a case study in the domain of intra-institutional research collaboration.

Novi Sad, Serbia — Miloš Savić
Novi Sad, Serbia — Mirjana Ivanović
Sydney, Australia — Lakhmi C. Jain

Contents

About the Authors

Dr. Miloš Savić is an Assistant Professor at the Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Serbia, where he received his B.Sc., M.Sc., and Ph.D. degrees in computer science. His research interests are related to complex network analysis with focus on social, information, ontology, and software networks. He is co-author of 30 research papers published in international journals and proceedings of international conferences. During his studies, he received the faculty award "Aleksandar Saša Popović" for outstanding research work in the field of Computer Science. He is also a Senior Teaching Associate at the Petnica Science Center. From 2014, he serves as an Editorial Assistant for the *Computer Science and Information Systems* (ComSIS) journal.

Dr. Mirjana Ivanović holds the position of Full Professor at Faculty of Sciences, University of Novi Sad. She is a member of the University Council for Informatics. She is author or co-author of 14 textbooks, several monographs, and more than 350 research papers on multi-agent systems, e-learning, and intelligent techniques, most of which are published in international journals and conferences. She is/was a member of Program Committees of more than 230 international conferences, participant of numerous international research projects and principal investigator of more than 15 projects. She delivered several keynote speeches at international conferences, and visited numerous academic institutions all over the world as visiting researcher

and teacher. Currently, she is Editor-in-Chief of the Computer Science and Information Systems journal.

Dr. Lakhmi C. Jain, Ph.D., ME, BE(Hons), Fellow (Engineers Australia) is with the Faculty of Education, Science, Technology and Mathematics at the University of Canberra, Australia and University of Technology Sydney, Australia. He founded the KES International for providing a professional community the opportunities for publications, knowledge exchange, cooperation, and teaming. Involving around 5,000 researchers drawn from universities and companies worldwide, KES facilitates international cooperation and generates synergy in teaching and research. KES regularly provides networking opportunities for professional community through one of the largest conferences of its kind in the area of KES. www.kesinternational.org.

His interests focus on the artificial intelligence paradigms and their applications in complex systems, security, e-education, e-healthcare, unmanned air vehicles, and intelligent agents.

Acronyms

ABST	Package Abstractness
AEC	Afferent–Efferent Coupling
AEXPR	Average Expression Complexity
AP	Average Population of Classes
AVGODF	Average Out-Degree Fraction
AXM	Number of Axioms
BA	Barabási-Albert
BET	Betweenness Centrality
BFS	Breadth First Search
CBO	Coupling Between Objects
CC	Clustering Coefficient
CC	Cyclomatic Complexity (Chap. 3)
CCD	Complementary Cumulative Distribution
CCN	Class Collaboration Network
CDF	Cumulative Distribution Function
CIG	Computer Intelligence in Games
CK	Chidamber–Kemerer
CLO	Closeness Centrality
COLL	Number of Collaborators
COMP	Internal Connectedness
COND	Conductance
CR	Class Richness
CRIS	Current Research Information System
CSCW	Computer Supported Cooperative Work
CUTR	Cut ratio
DBE	Department of Biology and Ecology
DC	Department of Chemistry, Biochemistry, and Environmental Protection
DD	Dominant Department
DEG	Degree

DEN	Internal Density
DFS	Depth First Search
DG	Department of Geography, Tourism, and Hotel Management
DIT	Depth of Inheritance Tree
DMI	Department of Mathematics and Informatics
DP	Department of Physics
EAND	Eager Associative Name Disambiguation
EB	Edge Betweenness
EC	Evolutionary Computation
ECIS	European Conference on Information Systems
ECOLL	Number of External Collaborators
ECST	Enriched Concrete Syntax Tree
ECTEL	European Conference on Technology Enhanced Learning
EM	Expectation–Maximization
ER	Erdős–Renyi
EVC	Eigenvector Centrality
EXP	Expansion
FODF	Flake Out-Degree Fraction
FS-UNS	Faculty of Sciences—University of Novi Sad
GCE	Graph Clustering Evaluation
GDN	General Dependency Network
GMO	Greedy Modularity Optimization
GWCC	Giant Weakly Connected Component
HCI	Human–Computer Interaction
HDIFF	Halstead Difficulty
HITS	Hyperlink-Induced Topic Search
HITSA	HITS Authority Score
HITSH	HITS Hub Score
HK	Henry–Kafura Complexity
HM	Hitz–Montazeri
HVOL	Halstead Volume
ICIS	International Conference of Information Systems
ICN	International Collaboration Network
IM	InfoMap
IMDb	Internet Movie Database
IN	In-degree
IR	Information Retrieval
IRI	Internationalized Resource Identifier
IS	Information Systems
KS	Kolmogorov–Smirnov
LAND	Lazy Associative Name Disambiguation
LCC	Loose Class Cohesion
LCOLL	Number of Local Collaborators
LCOM	Lack of Cohesion in Methods
LIS	Library and Information Science

LOC	Lines of Code
LV	Louvain (community detection algorithm)
MAXODF	Maximum Out-Degree Fraction
MCL	Markov Cluster Algorithm
MCN	Method Collaboration Network
MLE	Maximum Likelihood Estimation
MR	Mathematical Reviews
MWU	Mann—Whitney U
NCLASS	Number of Classes
NEC	Number of External Classes
NINST	Number of Instances
NMI	Normalized Mutual Information
NOC	Number of Children
NUMA	Number of Attributes
NUME	Number of Entities
NUMM	Number of Methods
OCN	Ontology Class Network
ODF	Out-Degree Fraction
OMN	Ontology Module Network
ONGRAM	Ontology Graphs and Metrics
OO	Object-Oriented
OON	Ontology Object Network
OSN	Ontology Subsumption Network
OUT	Out-degree
OWL	Web Ontology Language
PCN	Package Collaboration Network
PMF	Probability Mass Function
PNAS	Proceedings of the National Academy of Sciences
PR	Page Rank
PROF	Productivity (Fractional Counting)
PRON	Productivity (Normal Counting)
PROS	Productivity (Straight Counting)
PS	Probability of Superiority
RDF	Resource Description Framework
REC	References to External Classes
RR	Relationship Richness
RS	Radicchi Strong
RSNC	Reference Similarity Network Clustering
SBM	Stochastic Block Model
SCC	Strongly Connected Component
SCG	Static Call Graph
SICRIS	Slovenian Current Research Information System
SIGIR	Special Interest Group on Information Retrieval
SIGMOD	Special Interest Group on Management of Data
SLAND	Self-training Lazy Associative Name Disambiguation

SNA	Social Network Analysis
SNEIPL	Software Networks Extractor Independent of Programming Language
SOM	Spectral Optimization of Modularity
SQALE	Software Quality Assessment based on Lifecycle Expectations
SRCI	Serbian Research Competency Index
SVM	Support Vector Machine
SW	Small-World
SWEET	Semantic Web for Earth and Environmental Terminology
TCC	Tight Class Cohesion
TDNS	Top Degree Node Set
TEXPR	Total Expression Complexity
TF-IDF	Term Frequency—Inverse Document Frequency
TOT	Total Degree
UNS	University of Novi Sad
W3C	World Wide Web Consortium
WBET	Weighted Betweenness Centrality
WCC	Weakly Connected Component
WCLO	Weighted Closeness Centrality
WCOLL	Strength of Collaboration
WCRE	Working Conference on Reverse Engineering
WDEG	Weighted Degree Centrality
WECOLL	Strength of External Collaboration
WLCOLL	Strength of Local Collaboration
WMC	Weighted Methods Per Class
WS	Watts–Strogatz
WT	Walktrap
WWW	World Wide Web