

Lecture Notes in Social Networks

Şule Gündüz-Öğüdücü
A. Şima Etaner-Uyar *Editors*

Social Networks: Analysis and Case Studies

 Springer

Lecture Notes in Social Networks

Şule Gündüz-Öğüdücü
A. Şima Etaner-Uyar *Editors*

Social Networks: Analysis and Case Studies

 Springer

Lecture Notes in Social Networks

Lecture Notes in Social Networks (LNSN)

Series Editors

Reda Alhajj
University of Calgary
Calgary, AB, Canada

Uwe Glässer
Simon Fraser University
Burnaby, BC, Canada

Advisory Board

Charu Aggarwal, IBM T.J. Watson Research Center, Hawthorne, NY, USA
Patricia L. Brantingham, Simon Fraser University, Burnaby, BC, Canada
Thilo Gross, University of Bristol, United Kingdom
Jiawei Han, University of Illinois at Urbana-Champaign, IL, USA
Huan Liu, Arizona State University, Tempe, AZ, USA
Raúl Manásevich, University of Chile, Santiago, Chile
Anthony J. Masys, Centre for Security Science, Ottawa, ON, Canada
Carlo Morselli, University of Montreal, QC, Canada
Rafael Wittek, University of Groningen, The Netherlands
Daniel Zeng, The University of Arizona, Tucson, AZ, USA

For further volumes:
<http://www.springer.com/series/8768>

Şule Gündüz-Öğüdücü • A. Şima Etaner-Uyar
Editors

Social Networks: Analysis and Case Studies

 Springer

Editors

Şule Gündüz-Öğüdücü
Computer and Informatics Department
Istanbul Technical University
Istanbul
Turkey

A. Şima Etaner-Uyar
Computer Engineering Department
Istanbul Technical University
Istanbul
Turkey

ISSN 2190-5428

ISBN 978-3-7091-1796-5

DOI 10.1007/978-3-7091-1797-2

Springer Wien Heidelberg New York Dordrecht London

ISSN 2190-5436 (electronic)

ISBN 978-3-7091-1797-2 (eBook)

Library of Congress Control Number: 2014944729

© Springer-Verlag Wien 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, 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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

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

To Selim and Emre—Ş.G.Ö
To Krie et al.—Ş.E.U

Foreword

Network-based representation and analysis is rapidly growing into an effective technique for data analysis for various applications. The development in the technology directly and tremendously influenced the area of social network analysis and mining by allowing the construction of comprehensive frameworks for social networking and by facilitating the automated analysis of social networks. This has recently attracted considerable attention in the society almost at all levels from practitioners, to researchers, to developers, to end users, etc. Indeed, the amount of research outcome and the number of research groups working on social networks have rapidly increased over the past decades due to the recognition of the social network methodology as an effective model to develop unique solutions for some vital problems. This edited book entitled *Social Networks: Analysis and Case Studies* by Şule Gündüz-Öğüdücü and Şima Etaner-Uyar addresses the need by including a number of chapters written by experts in the field. It aims at helping the reader whether researcher or practitioner to grasp the state of the art in this rapidly growing networking era.

The chapter entitled *Ranking Authors on the Web: A Semantic AuthorRank* by Lule Ahmedi, Lavdim Halilaj, Gëzim Sejdiu, and Labinot Bajraktari introduces the CO-AUTHORONTO as an extended FOAF ontology. The CO-AUTHORONTO is further extended with PageRank and AuthorRank metrics for ranking authors based on their co-author links. The chapter entitled *Detecting Neutral Nodes in a Network of Heterogeneous Agent Based System* by Fatemeh Hendijani Fard and Behrouz H. Far proposes a technique that can identify neutral nodes, i.e., nodes that can never be considered as violating or interesting nodes because they show similar behavior in all of their interactions. David B. Skillicorn and Quan Zheng wrote the chapter entitled *Global Structure in Social Networks with Directed Typed Edges*. The chapter entitled *Social Networks and Group Effectiveness: The Role of External Network Ties* by Fabiola Bertolotti and Maria Rita Tagliaventi explores the relationship between group effectiveness and social networks. Alessia Amelio and Clara Pizzuti wrote a survey about *Overlapping Community Discovery Methods*. Zehra Çataltepe and Abdullah Sönmez address *Classification in Social Networks*. The chapter entitled *Experiences Using BDS: A Crawler for Social Internetworking*

Scenarios by Francesco Buccafurri, Gianluca Lax, Antonino Nocera, and Domenico Ursino argues that a crawling strategy which is good for social networks cannot be expected to remain valid in a Social Internetworking scenario. Accordingly, the authors defined a new crawling strategy specifically conceived for SISs. Lei Chen and Mingxuan Yuan address *Privacy and Ethical Issues in Social Network Analysis*. Finally, the chapter entitled *Social Media: The Evolution of E-health Services* by Tiziana Guzzo, Alessia D'Andrea, Fernando Ferri, and Patrizia Grifoni analyzes e-health services provided by different Social Media and introduces a Hybrid Cloud E-health Services architecture able to provide open, interoperable, scalable, and extensible services for all the e-health activities.

At the end it is important to emphasize this book as a valuable source for people involved in various applications in security, health, multiagent systems, ranking, social network structure and community analysis, among others. The reader will benefit from the content to build a clear vision in the field.

Calgary, AB, Canada
March 2014

Reda Alhadj

Preface

With the increasing popularity of Web 2.0, social media has become a widely used communication platform. Parallel to this development, Social Network Analysis gained in importance as a research field, while opening up many opportunities in different application domains. The book is a comprehensive resource for practitioners and researchers alike, who are new to the field as well as those who have years of experience. The book consists of chapters on Social Network Analysis theory and methods with a focus on current applications and case studies applied in various domains such as mobile networks, security, machine learning, and e-health.

This book was conceived as a result of the TUBITAK-EEEAG project 110E027 titled “A Semantic Recommendation System for Social Bookmarking Sites” supported by the Scientific and Technological Research Council of Turkey. We would like to thank our authors for their valuable contributions, without whom this book would not have been possible. They have met our deadlines and then patiently awaited this book to appear in print. We would also like to thank Reda Alhajj for writing the foreword, Tansel Özyer for his continuous support, and Nagehan Ilhan for her help during the preparation of the final manuscript. Finally, special thanks go to the staff at Springer, in particular Annelies Kersbergen.

We hope that you enjoy reading this book.

Istanbul, Turkey
February 2014

Şule Gündüz-Öğüdücü
A. Şima Etaner-Uyar

Glossary

Adjacency matrix A representation of a social network with n nodes using an $n \times n$ matrix where a non-zero entry at position ij indicates the presence of an edge, possible directed, from node i to node j . The magnitude of the entry may be used to represent the strength of the relationship between nodes i and j .

Agent-based system or Multiagent system A system consists of software agents.

AKT A broad ontology, referred to as the AKT Reference Ontology, for describing an academic computer science community. It consists of several sub-ontologies: Support, Portal, etc.

Association for Computing Machinery A premier membership organization for computing professionals. Unlike the IEEE, the ACM is solely dedicated to computing.

Assortativity index A measure of how similar are the labels of connected nodes.

Assortative networks The networks in which nodes with large degree connect to nodes with large degree.

AuthorRank The PageRank algorithm adopted to rank authors.

Blog Blog is an Internet discussion site that consists of discussion posts.

Bridge In a Social Internetworking Scenario, a bridge is a user who has accounts in different social networks and explicitly has specified links among these accounts.

Cliques A clique of a graph G is a complete subgraph of G , and the clique of largest possible size is referred to as a maximum clique. A maximal clique is a clique that cannot be extended by including one more adjacent vertex, meaning it is not a subset of a larger clique. Maximum cliques are therefore maximal cliques (but not necessarily vice versa).

Cloud-Computing Cloud-Computing is a term used to describe different types of computing concepts that involve various computers connected through Internet.

Co-authorOnto An ontology which extends FOAF to support modeling co-authorship networks on the Web.

Co-authorship networks An important class of social networks modeling co-author relations to, say, analyze the trends of research collaborations, or determine the status of individual researchers.

Collaboration diagram Shows the possible interaction path between different types of agents and how they can communicate with each other.

Collaboration matrices Matrix to show the relationships between different agent types containing agents as the nodes and weight of edges between them as the entries of the matrix.

Collective inference Iterative procedures, which classify related instances simultaneously. The content and link information for both training and test data are available. First, based on the available training content, link and label information, models are trained. Then, those models are used to label the test data simultaneously and iteratively where each test sample is labeled based on its neighbors.

Communication Network Network involving all the relationships through which organizational actors share resources such as information, help, and guide related to the execution of their work.

DBLP An on-line repository of bibliographic data on major computer science publications well known to its community.

Degree distribution The degree distribution is the probability distribution of node degrees over a network.

Degrees of nodes The degree of a node is the number of edges connected to the node. In terms of the adjacency matrix A , the degree for a node indexed by i in an undirected network is $k_i = \sum_j a_{i,j}$, where the sum is over all nodes in the network. In a directed network, each node has two degrees. The out-degree is the number of outgoing edges emanating from a node $k_i^{out} = \sum_j a_{j,i}$ and the in-degree is the number of incoming edges onto a node $k_i^{in} = \sum_j a_{i,j}$. The total degree of the node is the sum of its in- and out-degree $k_i^{tot} = k_i^{in} + k_i^{out}$.

Dendrogram A dendrogram is a tree diagram frequently used to illustrate the arrangement of the clusters produced by hierarchical clustering. The top row of nodes represent data (individual observations), and the remaining nodes represent the clusters to which the data belong, with the arrows representing the distance (dissimilarity). The distance between merged clusters is monotone increasing with the level of the merger: the height of each node in the plot is proportional to the value of the intergroup dissimilarity between its two daughters (the top nodes representing individual observations are all plotted at zero height).

Disassortative networks The networks in which large degree nodes tend to connect to small degree nodes.

E-health E-health indicates healthcare practice supported by electronic communication and processes.

Edge prediction Predicting when an edge ‘should’ be present in a graph, even though it is not, based on the entire structure of the graph.

Edge weight A positive weight associated with the edge in a social network representing the intensity of the relationship between the nodes at its ends.

Edit distance The edit distance between two strings is the number of single character manipulations (insertion, deletion, or replacement) that are needed to transform one string into the other. The edit distance thus describes how similar

or dissimilar two strings are by the number of steps it takes to turn from one into the other, where a step is defined as a single character change.

Eigendecomposition A way of transforming the representation of a matrix to use new axes that are aligned along directions of greatest uncorrelated variation.

When applied to the representation of a graph, an eigendecomposition is the basis for an embedding into a geometry. This geometry can be used to draw the graph, or to measure the similarity of any pair of its nodes by their distance apart.

Erdős number An early example of co-authorship networks, in which a “collaborative distance,” i.e., the smallest number of co-authorship links between any individual author and the mathematician Paul Erdős is calculated.

Friend of a Friend An RDF schema for machine-readable modeling of homepage-like profiles and social networks.

Generalization error The expected test error of a model on any test set that come from the same input distribution as the training set.

Greedy algorithm A greedy algorithm is any algorithm that follows the problem solving metaheuristic of making the locally optimal choice at each stage with the hope of finding the global optimum.

Group effectiveness Multidimensional construct that entails how: groups must produce outputs considered as adequate by those who receive, or who are in charge of evaluating; group members are able to work together in successive tasks; group members derive satisfaction from the execution of their tasks in the group.

Hamming In information theory, the Hamming distance between two strings of equal length is the number of positions at which the corresponding symbols are different. Put another way, it measures the minimum number of substitutions required to change one string into the other, or the number of errors that transformed one string into the other.

Healthcare Healthcare is the prevention and treatment of illness, disease, injury, and other mental and physical impairments in human beings.

Heterogeneous Agent-Based System Agent-based system in which the agents are heterogeneous.

Heterophily The tendency of entities not to be related to other similar entities.

Homophily The tendency of entities to be related to other similar entities.

Indegree The number of times an actor is drawn into an interaction by others.

Institute of Electrical and Electronics Engineers The world’s largest professional association dedicated to advancing technological innovation and excellence for the benefit of humanity.

Inter-type relationship Shows the interactions of different agent types.

Intra-type relationship Considers the communications between instances of the same type.

Laplacian matrix A transformation of an adjacency matrix by negating the existing entries and placing the row sums on the diagonal. A Laplacian acts as a kind of normalization, centering the cloud of points corresponding to the nodes around the origin.