

# Lecture Notes in Artificial Intelligence 7250

Subseries of Lecture Notes in Computer Science

LNAI Series Editors

Randy Goebel

*University of Alberta, Edmonton, Canada*

Yuzuru Tanaka

*Hokkaido University, Sapporo, Japan*

Wolfgang Wahlster

*DFKI and Saarland University, Saarbrücken, Germany*

LNAI Founding Series Editor

Joerg Siekmann

*DFKI and Saarland University, Saarbrücken, Germany*

Michael R. Berthold (Ed.)

# Bisociative Knowledge Discovery

An Introduction to Concept, Algorithms,  
Tools, and Applications

## Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada  
Jörg Siekmann, University of Saarland, Saarbrücken, Germany  
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

## Volume Editor

Michael R. Berthold  
University of Konstanz  
Department of Computer and Information Science  
Konstanz, Germany  
E-mail: michael.berthold@uni-konstanz.de

## Acknowledgement and Disclaimer

The work reported in this book was funded by the European Commission in the 7th Framework Programme (FP7-ICT-2007-C FET-Open, contract no. BISON-211898).

ISSN 0302-9743 e-ISSN 1611-3349  
ISBN 978-3-642-31829-0 e-ISBN 978-3-642-31830-6  
DOI 10.1007/978-3-642-31830-6  
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012941862

CR Subject Classification (1998): I.2, H.3, H.2.8, H.4, C.2, F.1

LNCS Sublibrary: SL 7 – Artificial Intelligence

© The Editor(s) (if applicable) and the Author(s) 2012. The book is published with open access at Springer-Link.com.

OpenAccess. This book is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

All commercial rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for commercial use must always be obtained from Springer. Permissions for commercial 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, 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.

*Typesetting:* Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

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

# Foreword

We have all heard of the success story of the discovery of a link between the mental problems of children and the chemical pollutants in their drinking water. Similarly, we have heard of the 1854 Broad Street cholera outbreak in London, and the linking of it to a contaminated public water pump. These are two high-profile examples of bisociation, the combination of information from two different sources.

This is exactly the focus of the BISON project and this book. Instead of attempting to keep up with the meaningful annotation of the data floods we are facing, the BISON group pursued a network-based integration of various types of data repositories and the development of new ways to analyze and explore the resulting gigantic information networks. Instead of finding well-defined global or local patterns they wanted to find domain-bridging associations which are, by definition, not well defined since they will be especially interesting if they are sparse and have not been encountered before.

The present volume now collects the highlights of the BISON project. Not only did the consortium succeed in formalizing the concept of bisociation and proposing a number of types of bisociation and measures to rank their “bisociativeness,” but they also developed a series of new algorithms, and extended several of the existing algorithms, to find bisociation in large bisociative information networks.

From a personal point of view, I was delighted to see that some of our own work on finding structurally similar pieces in large networks actually fit into that framework very well: Random walks, and related diffusion-based methods, can help find correlated nodes in bisociative networks. The concept of bisociative knowledge discovery formalizes an aspect of data mining that people have been aware of to some degree but were unable to formally pin down. The present volume serves as a great basis for future work in this direction.

May 2012

Christos Faloutsos

# Table of Contents

## Part I: Bisociation

Towards Bisociative Knowledge Discovery . . . . .	1
<i>Michael R. Berthold</i>	
Towards Creative Information Exploration Based on Koestler's Concept of Bisociation . . . . .	11
<i>Werner Dubitzky, Tobias Kötter, Oliver Schmidt, and Michael R. Berthold</i>	
From Information Networks to Bisociative Information Networks . . . . .	33
<i>Tobias Kötter and Michael R. Berthold</i>	

## Part II: Representation and Network Creation

Network Creation: Overview . . . . .	51
<i>Christian Borgelt</i>	
Selecting the Links in BisoNets Generated from Document Collections . . . . .	54
<i>Marc Segond and Christian Borgelt</i>	
Bridging Concept Identification for Constructing Information Networks from Text Documents . . . . .	66
<i>Matjaž Juršič, Borut Sluban, Bojan Cestnik, Miha Grčar, and Nada Lavrač</i>	
Discovery of Novel Term Associations in a Document Collection . . . . .	91
<i>Teemu Hynönen, Sébastien Mahler, and Hannu Toivonen</i>	
Cover Similarity Based Item Set Mining . . . . .	104
<i>Marc Segond and Christian Borgelt</i>	
Patterns and Logic for Reasoning with Networks . . . . .	122
<i>Angelika Kimmig, Esther Galbrun, Hannu Toivonen, and Luc De Raedt</i>	

## Part III: Network Analysis

Network Analysis: Overview . . . . .	144
<i>Hannu Toivonen</i>	
BiQL: A Query Language for Analyzing Information Networks . . . . .	147
<i>Anton Dries, Siegfried Nijssen, and Luc De Raedt</i>	

Review of BisoNet Abstraction Techniques . . . . .	166
<i>Fang Zhou, Sébastien Mahler, and Hannu Toivonen</i>	
Simplification of Networks by Edge Pruning . . . . .	179
<i>Fang Zhou, Sébastien Mahler, and Hannu Toivonen</i>	
Network Compression by Node and Edge Mergers . . . . .	199
<i>Hannu Toivonen, Fang Zhou, Aleksi Hartikainen, and Atte Hinkka</i>	
Finding Representative Nodes in Probabilistic Graphs . . . . .	218
<i>Laura Langohr and Hannu Toivonen</i>	
(Missing) Concept Discovery in Heterogeneous Information Networks . . .	230
<i>Tobias Kötter and Michael R. Berthold</i>	
Node Similarities from Spreading Activation . . . . .	246
<i>Kilian Thiel and Michael R. Berthold</i>	
Towards Discovery of Subgraph Bisociations . . . . .	263
<i>Uwe Nagel, Kilian Thiel, Tobias Kötter, Dawid Piątek, and Michael R. Berthold</i>	
<b>Part IV: Exploration</b>	
Exploration: Overview . . . . .	285
<i>Andreas Nürnberger</i>	
Data Exploration for Bisociative Knowledge Discovery: A Brief Overview of Tools and Evaluation Methods . . . . .	287
<i>Tatiana Gossen, Marcus Nitsche, Stefan Haun, and Andreas Nürnberger</i>	
On the Integration of Graph Exploration and Data Analysis: The Creative Exploration Toolkit . . . . .	301
<i>Stefan Haun, Tatiana Gossen, Andreas Nürnberger, Tobias Kötter, Kilian Thiel, and Michael R. Berthold</i>	
Bisociative Knowledge Discovery by Literature Outlier Detection . . . . .	313
<i>Ingrid Petrič, Bojan Cestnik, Nada Lavrač, and Tanja Urbančič</i>	
Exploring the Power of Outliers for Cross-Domain Literature Mining . . .	325
<i>Borut Sluban, Matjaž Juršič, Bojan Cestnik, and Nada Lavrač</i>	
Bisociative Literature Mining by Ensemble Heuristics . . . . .	338
<i>Matjaž Juršič, Bojan Cestnik, Tanja Urbančič, and Nada Lavrač</i>	

## Part V: Applications and Evaluation

Applications and Evaluation: Overview .....	359
<i>Igor Mozetič and Nada Lavrač</i>	
Biomine: A Network-Structured Resource of Biological Entities for Link Prediction .....	364
<i>Lauri Eronen, Petteri Hintsanen, and Hannu Toivonen</i>	
Semantic Subgroup Discovery and Cross-Context Linking for Microarray Data Analysis .....	379
<i>Igor Mozetič, Nada Lavrač, Vid Podpečan, Petra Kralj Novak, Helena Motaln, Marko Petek, Kristina Gruden, Hannu Toivonen, and Kimmo Kulovesi</i>	
Contrast Mining from Interesting Subgroups.....	390
<i>Laura Langohr, Vid Podpečan, Marko Petek, Igor Mozetič, and Kristina Gruden</i>	
Link and Node Prediction in Metabolic Networks with Probabilistic Logic .....	407
<i>Angelika Kimmig and Fabrizio Costa</i>	
Modelling a Biological System: Network Creation by Triplet Extraction from Biological Literature .....	427
<i>Dragana Miljkovic, Vid Podpečan, Miha Grčar, Kristina Gruden, Tjaša Stare, Marko Petek, Igor Mozetič, and Nada Lavrač</i>	
Bisociative Exploration of Biological and Financial Literature Using Clustering .....	438
<i>Oliver Schmidt, Janez Kranjc, Igor Mozetič, Paul Thompson, and Werner Dubitzky</i>	
Bisociative Discovery in Business Process Models .....	452
<i>Trevor Martin and Hongmei He</i>	
Bisociative Music Discovery and Recommendation .....	472
<i>Sebastian Stober, Stefan Haun, and Andreas Nürnberger</i>	
<b>Author Index</b> .....	485