# Semantic Mining of Social Networks

# Synthesis Lectures on the Semantic Web: Theory and Technology

Editors
**Ying Ding,** ***Indiana University***
**Paul Groth,** ***Elsevier Labs***

Founding Editor Emeritus
**James Hendler,** ***Rensselaer Polytechnic Institute***

Synthesis Lectures on the Semantic Web: Theory and Application is edited by Ying Ding of Indiana University and Paul Groth of Elsevier Labs. Whether you call it the Semantic Web, Linked Data, or Web 3.0, a new generation of Web technologies is offering major advances in the evolution of the World Wide Web. As the first generation of this technology transitions out of the laboratory, new research is exploring how the growing Web of Data will change our world. While topics such as ontology-building and logics remain vital, new areas such as the use of semantics in Web search, the linking and use of open data on the Web, and future applications that will be supported by these technologies are becoming important research areas in their own right. Whether they be scientists, engineers or practitioners, Web users increasingly need to understand not just the new technologies of the Semantic Web, but to understand the principles by which those technologies work, and the best practices for assembling systems that integrate the different languages, resources, and functionalities that will be important in keeping the Web the rapidly expanding, and constantly changing, information space that has changed our lives.
Topics to be included:

- Semantic Web Principles from linked-data to ontology design
- Key Semantic Web technologies and algorithms
- Semantic Search and language technologies
- The Emerging "Web of Data" and its use in industry, government and university applications
- Trust, Social networking and collaboration technologies for the Semantic Web
- The economics of Semantic Web application adoption and use
- Publishing and Science on the Semantic Web
- Semantic Web in health care and life sciences

Semantic Mining of Social Networks
Jie Tang and Juanzi Li
2015

Social Semantic Web Mining
Tope Omitola, Sebastián A. Ríos, and John G. Breslin
2015

Semantic Breakthrough in Drug Discovery
Bin Chen, Huijun Wang, Ying Ding, and David Wild
2014

Semantics in Mobile Sensing
Zhixian Yan and Dipanjan Chakraborty
2014

Provenance: An Introduction to PROV
Luc Moreau and Paul Groth
2013

Resource-Oriented Architecture Patterns for Webs of Data
Brian Sletten
2013

Aaron Swartz's A Programmable Web: An Unfinished Work
Aaron Swartz
2013

Incentive-Centric Semantic Web Application Engineering
Elena Simperl, Roberta Cuel, and Martin Stein
2013

Publishing and Using Cultural Heritage Linked Data on the Semantic Web
Eero Hyvönen
2012

VIVO: A Semantic Approach to Scholarly Networking and Discovery
Katy Börner, Michael Conlon, Jon Corson-Rikert, and Ying Ding
2012

Linked Data: Evolving the Web into a Global Data Space
Tom Heath and Christian Bizer
2011

© Springer Nature Switzerland AG 2022
Reprint of original edition © Morgan & Claypool 2015

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Semantic Mining of Social Networks

Jie Tang and Juanzi Li

ISBN: 978-3-031-79461-2 paperback
ISBN: 978-3-031-79462-9 ebook

DOI 10.1007/978-3-031-79462-9

A Publication in the Springer series
*SYNTHESIS LECTURES ON THE SEMANTIC WEB: THEORY AND TECHNOLOGY*

Lecture #11
Series Editors: Ying Ding, *Indiana University*
Paul Groth, *Elsevier Labs*
Founding Editor Emeritus: James Hendler, *Rensselaer Polytechnic Institute*
Series ISSN
Print 2160-4711 Electronic 2160-472X

# Semantic Mining of Social Networks

Jie Tang and Juanzi Li
Tsinghua University

*SYNTHESIS LECTURES ON THE SEMANTIC WEB: THEORY AND TECHNOLOGY #11*

## ABSTRACT

Online social networks have already become a bridge connecting our physical daily life with the (web-based) information space. This connection produces a huge volume of data, not only about the information itself, but also about user behavior. The ubiquity of the social Web and the wealth of social data offer us unprecedented opportunities for studying the interaction patterns among users so as to understand the dynamic mechanisms underlying different networks, something that was previously difficult to explore due to the lack of available data.

In this book, we present the architecture of the research for social network mining, from a microscopic point of view. We focus on investigating several key issues in social networks. Specifically, we begin with analytics of social interactions between users. The first kinds of questions we try to answer are: What are the fundamental factors that form the different categories of social ties? How have reciprocal relationships been developed from parasocial relationships? How do connected users further form groups?

Another theme addressed in this book is the study of social influence. Social influence occurs when one's opinions, emotions, or behaviors are affected by others, intentionally or unintentionally. Considerable research has been conducted to verify the existence of social influence in various networks. However, few literature studies address how to quantify the strength of influence between users from different aspects. In Chapter 4 and in [138], we have studied how to model and predict user behaviors. One fundamental problem is distinguishing the effects of different social factors such as social influence, homophily, and individual's characteristics. We introduce a probabilistic model to address this problem.

Finally, we use an academic social network, ArnetMiner, as an example to demonstrate how we apply the introduced technologies for mining real social networks. In this system, we try to mine knowledge from both the informative (publication) network and the social (collaboration) network, and to understand the interaction mechanisms between the two networks. The system has been in operation since 2006 and has already attracted millions of users from more than 220 countries/regions.



## KEYWORDS

social tie, strong/weak ties, parasocial interactions, reciprocity, social influence, collective classification, graphical model, social network analysis, social relationship, relationship mining, link prediction, influence maximization, network centrality, user modeling, social action, social theories, social balance, social status, triadic closure, factor graph, influence propagation, conservative influence propagation, non-conservative influence propagation, user behaviour prediction, profile extraction, expert finding, name disambiguation, ArnetMiner

# Contents

**Acknowledgments** . . . . . **xi**

**1 Introduction** . . . . . **1**

1.1 Background . . . . . 2
1.1.1 Social Theories . . . . . 3
1.1.2 Social Tie Analysis . . . . . 4
1.1.3 Social Influence Analysis . . . . . 6
1.1.4 User Modeling and Actions . . . . . 8
1.1.5 Graphical Models . . . . . 8
1.2 Book Outline . . . . . 8

**2 Social Tie Analysis** . . . . . **11**

2.1 Overview . . . . . 11
2.2 Predicting Missing Links . . . . . 12
2.2.1 Similarity Metrics . . . . . 15
2.2.2 Matrix Factorization . . . . . 20
2.3 Inferring Social Ties . . . . . 25
2.3.1 Problem Formulation . . . . . 27
2.3.2 Unsupervised Learning to Infer Social Ties . . . . . 28
2.3.3 Supervised Learning to Infer Social Ties . . . . . 29
2.3.4 Actively Learning to Infer Social Ties . . . . . 40
2.3.5 Inferring Social Ties across Heterogeneous Networks . . . . . 49
2.4 Conclusions . . . . . 68

**3 Social Influence Analysis** . . . . . **69**

3.1 Overview . . . . . 69
3.2 Mining Topic-level Social Influence analysis . . . . . 71
3.2.1 Topical Affinity Propagation . . . . . 72
3.2.2 Dynamic Social Influence Analysis . . . . . 79
3.2.3 Model Application . . . . . 81
3.2.4 Experimental Results . . . . . 82
3.2.5 Summary . . . . . 87

3.3 Mining Topic-level Influence from Heterogeneous Networks . . . . . 87
3.3.1 The Approach Framework . . . . . 89
3.3.2 Influence Propagation and Aggregation . . . . . 94
3.3.3 Conservative and Non-conservative Propagation . . . . . 96
3.3.4 User Behavior Prediction . . . . . 100
3.3.5 Evaluation . . . . . 101
3.3.6 Summary . . . . . 110
3.4 Conclusions . . . . . 110

**4 User Behavior Modeling and Prediction . . . . . 111**
4.1 Overview . . . . . 111
4.2 Approach Framework for Social Action Prediction . . . . . 114
4.2.1 Model Learning . . . . . 117
4.3 Evaluation . . . . . 122
4.3.1 Evaluation Metrics . . . . . 123
4.3.2 Prediction Performance . . . . . 125
4.3.3 Efficiency Performance . . . . . 126
4.3.4 Qualitative Case Study . . . . . 127
4.4 Summary . . . . . 130

**5 ArnetMiner: Deep Mining for Academic Social Networks . . . . . 131**
5.1 Overview . . . . . 131
5.2 Researcher Profile Extraction . . . . . 134
5.2.1 A Unified Approach to Profiling . . . . . 136
5.2.2 Profile Extraction Performance . . . . . 138
5.3 Name Disambiguation . . . . . 138
5.3.1 A Unified Probabilistic Framework . . . . . 141
5.3.2 Name Disambiguation Performance . . . . . 143
5.4 Topic Modeling . . . . . 147
5.4.1 Our Proposed Topic Models . . . . . 148
5.5 Expertise Search . . . . . 153
5.5.1 Data Sets and Evaluation Measures . . . . . 154
5.5.2 Results . . . . . 155
5.6 Academic Social Network Mining . . . . . 163
5.6.1 Mining Advisor-Advisee Relationships . . . . . 164
5.6.2 Measuring Academic Influence . . . . . 165
5.6.3 Modeling Researcher Interests . . . . . 168
5.7 Conclusions . . . . . 168

6 **Research Frontiers** . . . . . 171

A **Resources** . . . . . 173

A.1 Software . . . . . 173

A.2 Data Sets . . . . . 174

**Bibliography** . . . . . 177

**Authors' Biographies** . . . . . 193

# Acknowledgments

This work is supported by the National High-Tech R&D Program (No. 2014AA015103), National Basic Research Program of China (No. 2014CB340506, No. 2012CB316006), Natural Science Foundation of China (No. 61222212, No.61035004), the Tsinghua University Initiative Scientific Research Program (20121088096), a research fund supported by Huawei Inc., and Beijing key lab of networked multimedia.

Jie Tang and Juanzi Li
April 2015