

Lecture Notes in Social Networks

Series editors

Reda Alhajj, University of Calgary, Calgary, AB, Canada

Uwe Glässer, Simon Fraser University, Burnaby, BC, Canada

Huan Liu, Arizona State University, Tempe, AZ, USA

Rafael Wittek, University of Groningen, Groningen, The Netherlands

Daniel Zeng, Tucson, AZ, USA

Advisory Board

Charu C. Aggarwal, Yorktown Heights, NY, USA

Patricia L. Brantingham, Simon Fraser University, Burnaby, BC, Canada

Thilo Gross, University of Bristol, Bristol, UK

Jiawei Han, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Raúl Manásevich, University of Chile, Santiago, Chile

Anthony J. Masys, University of Leicester, Ottawa, ON, Canada

Carlo Morselli, School of Criminology, Montreal, ON, Canada

More information about this series at <http://www.springer.com/series/8768>

Mehmet Kaya • Jalal Kawash • Suheil Khoury
Min-Yuh Day
Editors

Social Network Based Big Data Analysis and Applications

Editors

Mehmet Kaya
Department of Computer Engineering
Firat University
Elazig, Turkey

Jalal Kawash
Department of Computer Science
University of Calgary
Calgary, AB, Canada

Suheil Khoury
American University of Sharjah
Sharjah, United Arab Emirates

Min-Yuh Day
Tamkang University
Taipei, Taiwan

ISSN 2190-5428

ISSN 2190-5436 (electronic)

Lecture Notes in Social Networks

ISBN 978-3-319-78195-2

ISBN 978-3-319-78196-9 (eBook)

<https://doi.org/10.1007/978-3-319-78196-9>

Library of Congress Control Number: 2018940255

© Springer International Publishing AG, part of Springer Nature 2018

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.

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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Preface

Social networks (SN) have brought an unprecedented revolution in how people interact and socialize. SN are used not only as a lifestyle but also in various other domains, including medicine, business, education, politics, and activism. The number of SN amounts to billions of users. At the beginning of 2016, Twitter claimed to have 313 million monthly active users. As of the third quarter of 2017, Facebook had slightly more than 2 billion monthly active users. Online social media (OSM), media produced by SN users, has offered a real and viable alternative to conventional mainstream media. OSM is likely to provide “raw,” unedited information, and the details can be overwhelming with the potential of misinformation and disinformation. Yet, OSM is leading to the democratization of knowledge and information. OSM is allowing almost any citizen to become a journalist reporting on specific events of interest. This is resulting in unimaginable amounts of information being shared among huge numbers of OSM participants. For example, Facebook users are generating several billion “likes” and more than 100 million posted pictures in a single day. Twitter users are producing more than 6000 tweets per second. The size of the data generated presents increasing challenges to mine, analyze, utilize, and exploit such content. This book includes eleven contributions that examine several topics related to data analysis and social networks. Applications include sentiment dictionaries, malicious content identification, video recapping, cancer biomarkers, face detection, pattern detection, and cell phone subscription predictions. What follows is a quick summary of each of these chapters.

Nuno Guimarães, Luís Torgo, and Álvaro Figueira complement traditional sentiment dictionaries with a system for lexicon expansion, extracting and classifying domain- and time-specific terms with sentiment based on public opinion. Domain- and time-specific lexicons improve the performance of sentiment analysis methods on short informal texts, such as tweets. The proposed system can generate dictionaries, on a daily basis, to complement the more traditional sentiment lexicons.

Prateek Dewan, Shrey Bagroy, and Ponnuram Kumaraguru address the issue of identifying malicious content on Facebook, such as publishing untrustworthy information, misleading content, adult and child unsafe content, and scams. The

identified 627 malicious pages revealed through spatial and temporal analysis dominant presence of politically polarized entities engaging in spreading content from untrustworthy domains. Multiple supervised learning algorithms and multiple feature sets are evaluated, and they find that artificial neural networks trained on a fixed sized bag-of-words perform the best in identifying such malicious pages.

Automatic generation of video recaps and summaries is the subject of the chapter by Xavier Bost, Vincent Labatut, Serigne Gueye, and Georges Linarès. They propose narrative smoothing, a method for the extraction of dynamic social networks of video characters. They introduce an algorithm to estimate verbal interactions from a sequence of spoken segments. The data used are a corpus of 109 TV series episodes from three popular TV shows: *Breaking Bad*, *Game of Thrones*, and *House of Cards*.

Gabriela Jurca, Omar Addam, Jon Rokne, and Reda Alhajj study the assessment of candidates for academic positions or for promotion. They employ social network analysis and community detection to measure the influence and diversity of members, within the Department of Computer Science at the University of Calgary. Different measures between various ranks in the department are presented and discussed.

In another chapter, Gabriela Jurca, Omar Addam, Jon Rokne, and Reda Alhajj study biomarkers used to diagnose prostate cancer. They used text mining to provide a tool to examine whether biomarkers are emerging or decreasing in terms of publication popularity. They also provide a tool to examine the increasing or decreasing popularity of gene families with respect to prostate cancer research. Selected biomarkers which have been labeled as emerging in qualitative reviews are then evaluated.

The spread of influence in complex networks is the subject of the chapter by Arun Sathanur, Mahantesh Halappanavar, Yalin Sagduyu, and Yi Shi. They consider the problem of modeling the spread of influence and the identification of influential entities in a complex network with nodal activation, intrinsic or external through neighbors. They approach mining for the influential nodes through influence maximization. One of the findings is how influential content creators can drive engagement on social media platforms.

Yingbo Zhu, Zhenhua Huang, Zhenyu Wang, Linfeng Luo, and Shuang Wu revisit Spiral of Silence in the context of social networks with real information diffusion data. They analyze four information diffusion tree metrics: width, depth, message sentiment, and modularity. Based on Spiral of Silence, polarity prediction of users' review without considering semantic meaning of content is proposed and discovered. Their results indicate that opinions of people in propagation are impacted by the social environment. The Anti-Spiral of Silence is also found to play a significant role in leading rational public opinion and revealing truth in social networks.

Prediction of mobile service subscription types is entertained by Yongjun Liao, Wei Du, Márton Karsai, Carlos Sarraute, Martin Minnoni, and Eric Fleury, specifically the behavioral differences between prepaid and postpaid customers. The findings are used to provide methods that detect the subscription type of customers

by using information about their personal call statistics and their egocentric networks. This allows this classification problem to be treated as a problem of graph labeling, which can be solved by max-flow, min-cut algorithms. The chapter also aims at inferring the subscription type of customers, using node attributes, and a two-ways indirect inference method based on observed hemophilic structural correlations.

Konstantinos F. Xylogiannopoulos, Panagiotis Karampelas, and Reda Alhajj take on real-time detection of all repeated patterns in a big data stream. A new data structure is introduced: LERP Reduced Suffix Array with a new detection algorithm. This allows the detection of all repeated patterns in a string in a very short time. Specifically, their results show analysis of one million data points and a sliding window of groups of three subsequences of the same size simultaneously with detection in about 300 ms.

Cold start in a dating recommendation service is addressed by Mo Yu, Xiaolong Zhang, Dongwon Lee, and Derek Kreager. They approach this challenge by proposing a novel community-based recommendation framework. Detecting communities to which existing users belong and by matching new users to these communities, the proposed method improves on existing recommendation methods.

The last chapter by Salim Afra and Reda Alhajj studies the performance of face clustering approaches using different feature extraction techniques. Best practices for face recognition of terrorists and criminals are entertained. Performance evaluation for various feature extraction techniques and clustering algorithms using four datasets is also studied.

To conclude this preface, we would like to thank the authors who submitted papers and the reviewers who provided detailed constructive reports which improved the quality of the papers. Various people from Springer deserve great credit for their help and support in all the issues related to publishing this book.

Elazig, Turkey
Calgary, AB, Canada
Sharjah, United Arab Emirates
Taipei, Taiwan
November 2017

Mehmet Kaya
Jalal Kawash
Suheil Khoury
Min-Yuh Day

Contents

Twitter as a Source for Time- and Domain-Dependent Sentiment Lexicons	1
Nuno Guimarães, Luís Torgo, and Álvaro Figueira	
Hiding in Plain Sight: The Anatomy of Malicious <i>Pages</i> on Facebook	21
Prateek Dewan, Shrey Bagroy, and Ponnurangam Kumaraguru	
Extraction and Analysis of Dynamic Conversational Networks from TV Series	55
Xavier Bost, Vincent Labatut, Serigne Gueye, and Georges Linarès	
Diversity and Influence as Key Measures to Assess Candidates for Hiring or Promotion in Academia	85
Gabriela Jurca, Omar Addam, Jon Rokne, and Reda Alhajj	
Timelines of Prostate Cancer Biomarkers	105
Gabriela Jurca, Omar Addam, Jon Rokne, and Reda Alhajj	
Exploring the Role of Intrinsic Nodal Activation on the Spread of Influence in Complex Networks	123
Arun V. Sathanur, Mahantesh Halappanavar, Yi Shi, and Yalin Sagduyu	
Influence and Extension of the Spiral of Silence in Social Networks: A Data-Driven Approach	143
Yingbo Zhu, Zhenhua Huang, Zhenyu Wang, Linfeng Luo, and Shuang Wu	
Prepaid or Postpaid? That Is the Question: Novel Methods of Subscription Type Prediction in Mobile Phone Services	165
Yongjun Liao, Wei Du, Márton Karsai, Carlos Sarraute, Martin Minnoni, and Eric Fleury	
Dynamic Pattern Detection for Big Data Stream Analytics	183
Konstantinos F. Xylogiannopoulos, Panagiotis Karampelas, and Reda Alhajj	

Community-Based Recommendation for Cold-Start Problem: A Case Study of Reciprocal Online Dating Recommendation	201
Mo Yu, Xiaolong (Luke) Zhang, Dongwon Lee, and Derek Kreager	
Combining Feature Extraction and Clustering for Better Face Recognition	223
Salim Afra and Reda Alhajj	
Index.....	243