Lecture Notes in Computer Science 5498

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Lee Giles Marc Smith John Yen
Haizheng Zhang (Eds.)

Advances in Social Network Mining and Analysis

Second International Workshop, SNAKDD 2008
Las Vegas, NV, USA, August 24-27, 2008
Revised Selected Papers

Springer

Volume Editors

Lee Giles
John Yen
Pennsylvania State University, College of Information Science and Technology
University Park, PA 16802, USA
E-mail: {giles, jyen}@ist.psu.edu

Marc Smith
Microsoft Research, One Microsoft Way, Redmond, WA 98002, USA
E-mail: masmith@microsoft.com

Haizheng Zhang
Amazon.com., Seattle, WA, USA
E-mail: haizhengzhang@gmail.com

Library of Congress Control Number: 2010931753

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

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

ISSN 0302-9743
ISBN-10 3-642-14928-6 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-14928-3 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

This year's volume of Advances in Social Network Analysis contains the proceedings for the Second International Workshop on Social Network Analysis (SNAKDD 2008). The annual workshop co-locates with the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). The second SNAKDD workshop was held with KDD 2008 and received more than 32 submissions on social network mining and analysis topics. We accepted 11 regular papers and 8 short papers. Seven of the papers are included in this volume.

In recent years, social network research has advanced significantly, thanks to the prevalence of the online social websites and instant messaging systems as well as the availability of a variety of large-scale offline social network systems. These social network systems are usually characterized by the complex network structures and rich accompanying contextual information. Researchers are increasingly interested in addressing a wide range of challenges residing in these disparate social network systems, including identifying common static topological properties and dynamic properties during the formation and evolution of these social networks, and how contextual information can help in analyzing the pertaining social networks. These issues have important implications on community discovery, anomaly detection, trend prediction and can enhance applications in multiple domains such as information retrieval, recommendation systems, security and so on.

The second SNAKDD workshop focused on knowledge discovery and data mining in social networks, such as contextual community discovery, link analysis, the growth and evolution of social networks, algorithms for large-scale graphs, techniques that can be used for recovering and constructing social networks from online social systems, search on social networks, multi-agent-based social network simulation, trend prediction of social network evolution, and related applications in other domains such as information retrieval and security. The workshop was concerned with inter-disciplinary and cross-domain studies spanning a variety of areas in computer science including graph and data mining, machine learning, computational organizational and multi-agent studies, information extraction and retrieval, and security, as well as other disciplines such as information science, and social science.

In the first paper "Leveraging Label-Independent Features for Classification in Sparsely Labeled Networks: An Empirical Study," Brian Gallagher and Tina Eliassi-Rad study the problem of within-network classification in sparsely labeled networks. The authors present an empirical study and show that the use of LI features produces classifiers that are less sensitive to specific label assignments and can lead to significant performance improvement.

In the second paper "Community Detection Using a Measure of Global Influence," Rumi Ghosh and Kristina Lerman, define "influence" as the number of paths, of any length, that exist between two nodes and argue that this gives a better measure of network connectivity. The authors use the influence metric to partition a network into groups or communities by looking for regions of the network where nodes have more influence over each other than over nodes outside the community.

The third paper, "Communication Dynamics of Blog Networks," by Mark Goldberg, Malik Magdon-Ismil, Stephen Kelley, Konstantin Mertsalov, and William Wallace, studies the communication dynamics of Blog networks by exploring the Russian section of LiveJournal. The two fundamental questions that this paper is concerned with include (1) what models adequately describe such dynamic communication behavior; and (2) how does one detect changes in the nature of the communication dynamics. The paper leverages stable statistics in their research in disclosing the dynamics of the networks and characterizing the locality properties.

In the fourth paper, "Finding Spread Blockers in Dynamic Networks," Habiba, Yintao Yu, Tanya Berger-Wolf, and Jared Saia extend standard structural network measures to dynamic networks. The authors also compare the blocking ability of individuals in the order of ranking by the new dynamic measures. The authors found that overall, simple ranking according to a nodes static degree, or the dynamic version of a nodes degree, performed consistently well. Surprisingly the dynamic clustering coefficient seems to be a good indicator, while its static version performs worse than the random ranking.

The fifth paper, "Social Network Mining with Nonparametric Relational Models," by ZZhao Xu, Volker Tresp, Achim Rettinger, and Kristian Kersting, discusses how the infinite hidden relational model (IHRM) can be used to model and analyze social networks, where each edge is associated with a random variable and the probabilistic dependencies between variables are specified by the model based on the relational structure. The hidden variables are able to transport information such that non-local probabilistic dependencies can be obtained. The IHRM provides effective relationship prediction and cluster analysis for social networks. The experiments demonstrate that this model can provide good prediction accuracy and capture the inherent relations among social actors.

In the sixth paper, "Using Friendship Ties and Family Circles for Link Prediction," the authors (Elena Zheleva, Lise Getoor, Jennifer Golbeck, and Ugur Kuter) investigate how networks can be overlaid and propose a feature taxonomy for link prediction. The authors show that the accuracy of link prediction can be improved when there are tightly-knit family circles in a social network. Their experiments demonstrate significantly higher prediction accuracy compared to using traditional features such as descriptive node attributes and structural features.

The last paper, "Information Theoretic Criteria for Community Detection" by Karl Branting, studies the resolution limit problem with two compression-based algorithms that were designed to overcome such limits. The author

identifies the aspect of each approach that is responsible for the resolution limit and proposes a variant, SGE, that addresses this limitation. The paper demonstrates on three artificial data sets that (1) SGE does not exhibit a resolution limit on graphs in which other approaches do, and that (2) modularity and the compression-based algorithms, including SGE, behave similarly on graphs not subject to the resolution limit.

We would like to thank the authors of all submitted papers for both the joint workshop and this proceedings volume. We are further indebted to the Program Committee members for their rigorous and timely reviewing. They allowed us to make this workshop a major success.

Lee Giles
Marc Smith
John Yen
Haizheng Zhang

Organization

Program Chairs

Lee Giles	Pennsylvania State University, USA
Marc Smith	Microsoft, USA
John Yen	Pennsylvania State University, USA
Haizheng Zhang	Amazon.com, USA

Program Committee

Lada Adamic
Aris Anagnostopoulos
Arindam Banerjee
Tanya Berger-Wolf
Yun Chi
Aaron Clauset
Isaac Councill
Tina Eliassi-Rad
Lise Getoor
Mark Goldberg
Larry Holder
Andreas Hotho
Gueorgi Kossinets
Kristina Lerman
Wei Li
Yi Liu
Ramesh Nallapati
Jennifer Neville
Cheng Niu
Dou Shen
Bingjun Sun
Jie Tang
Andrea Tapia
Alessandro Vespigani
Xuerui Wang
Michael Wurst
Xiaowei Xu

Referees

Vladimir Barash
Mustafa Bilgic
Matthias Broecheler
Guihong Cao
Bin Cao
Sanmay Das
Anirban Dasgupta
Robert JÃschke
Liu Liu
Galileo Namata
Evan Xiang
Xiaowei Xu
Limin Yao
Jing Zhang
Yi Zhang

Table of Contents

Leveraging Label-Independent Features for Classification in Sparsely Labeled Networks: An Empirical Study 1
Brian Gallagher and Tina Eliassi-Rad

Community Detection Using a Measure of Global Influence 20
Rumi Ghosh and Kristina Lerman

Communication Dynamics of Blog Networks 36
Mark Goldberg, Stephen Kelley, Malik Magdon-Ismail, Konstantin Mertsalov, and William (Al) Wallace

Finding Spread Blockers in Dynamic Networks 55
Habiba, Yintao Yu, Tanya Y. Berger-Wolf, and Jared Saia

Social Network Mining with Nonparametric Relational Models 77
Zhao Xu, Volker Tresp, Achim Rettinger, and Kristian Kersting

Using Friendship Ties and Family Circles for Link Prediction 97
Elena Zheleva, Lise Getoor, Jennifer Golbeck, and Ugur Kuter

Information Theoretic Criteria for Community Detection 114
L. Karl Branting

Author Index 131