

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

William Aiello Andrei Broder
Jeannette Janssen Evangelos Milios (Eds.)

Algorithms and Models for the Web-Graph

Fourth International Workshop, WAW 2006
Banff, Canada, November 30 - December 1, 2006
Revised Papers

Volume Editors

William Aiello
University of British Columbia
Vancouver, Canada
E-mail: aiello@cs.ubc.ca

Andrei Broder
Yahoo! Research
Sunnyvale, CA 94089, USA
E-mail: broder@yahoo-inc.com

Jeannette Janssen
Dalhousie University
Halifax, B3H 3J5, Canada
E-mail: janssen@mathstat.dal.ca

Evangelos Milios
Dalhousie University
Halifax, B3H 1W5, Canada
E-mail: eem@cs.dal.ca

Library of Congress Control Number: 2008922735

CR Subject Classification (1998): H.5.1, H.5.3, H.5.4

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN	0302-9743
ISBN-10	3-540-78807-7 Springer Berlin Heidelberg New York
ISBN-13	978-3-540-78807-2 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, 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 German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media
springer.com

© Springer-Verlag Berlin Heidelberg 2008
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12246403 06/3180 5 4 3 2 1 0

Preface

The World Wide Web has become part of our everyday life and information retrieval and data mining on the Web are now of enormous practical interest. The algorithms supporting these activities combine the view of the Web as a text repository and as a graph, induced in various ways by links among pages, links among hosts, or other similar networks.

The aim of the 4th Workshop on Algorithms and Models for the Web-Graph (WAW 2006) was to further the understanding of these Web-induced graphs and stimulate the development of high-performance algorithms and applications that use the graph structure of the Web. The workshop was meant both to foster an exchange of ideas among the diverse set of researchers already involved in this topic and to act as an introduction for the larger community to the state of the art in this area. The workshop program included invited keynote talks by Fan Chung-Graham (UCSD), Soumen Chakrabarti (IITB), Walter Willinger (ATT Research) and Filippo Menczer (Indiana).

WAW 2006 took place on November 30 and December 1 at the Banff International Research Institute (BIRS), in Banff, Alberta (Canada). It was the fourth in a series of very successful workshops on the Web graph. WAW 2002 (Vancouver) and 2004 (Rome) were held in conjunction with the Annual IEEE Symposium on Foundations of Computer Science (FOCS). WAW 2003 (Budapest) was held in conjunction with the 12th International World Wide Web Conference.

In response to the call for abstracts we received 28 submissions. Almost all submissions were relevant to the topic of the workshop and contained interesting ideas. The Organizing Committee selected 12 contributors to present their work at the workshop, while 12 others were invited to prepare a poster about their work, which was exhibited at the workshop. Those authors selected for contributed talks were invited to prepare a paper for the proceedings. The poster presenters were given an opportunity to explain their work during a poster session. All participants were invited to judge the posters in a contest for best poster. The three posters that received the most votes were also invited to submit their work to the proceedings. The papers were then submitted to a refereeing process by the Program Committee, and the final versions are included in this volume. The volume also includes abstracts of the invited talks.

We would like to thank all those that helped to make this workshop a success, with special thanks to the staff at BIRS. We thank the MITACS (Mathematics of Information Technology and Complex Systems) Network of Centres of Excellence (NCE) for financial and organizational support. We also owe many thanks to Andrei Voronkov for providing the EasyChair conference system. With this system, managing the refereeing process and the collation of the proceedings became almost effortless. Finally, we would like to thank all participants

in the workshop, all authors of the contributed papers, and especially the invited speakers for their contribution in making WAW 2006 an amicable and stimulating forum for the exchange of ideas about the Web graph.

November 2007

William Aiello
Andrei Broder
Jeannette Janssen
Evangelos Milios

Organization

Organizing Committee

William Aiello	University of British Columbia
Andrei Broder	Yahoo! Inc.
Jeannette Janssen	Dalhousie University
Evangelos Milios	Dalhousie University

Sponsoring Institutions

The Mathematics of Information Technology and Complex Systems (MITACS)
Yahoo! Inc.
GenieKnows.com

Table of Contents

Modelling and Mining of Networked Information Spaces	1
<i>William Aiello, Andrei Broder, Jeannette Janssen, and Evangelos Milios</i>	
Workshop on Algorithms and Models for the Web Graph	18
<i>William Aiello, Andrei Broder, Jeannette Janssen, and Evangelos Milios</i>	
Expansion and Lack Thereof in Randomly Perturbed Graphs	24
<i>Abraham D. Flaxman</i>	
Web Structure in 2005	36
<i>Yu Hirate, Shin Kato, and Hayato Yamana</i>	
Local/Global Phenomena in Geometrically Generated Graphs	47
<i>Ross M. Richardson</i>	
Approximating PageRank from In-Degree	59
<i>Santo Fortunato, Marián Boguñá, Alessandro Flammini, and Filippo Menczer</i>	
Probabilistic Relation between In-Degree and PageRank	72
<i>Nelly Litvak, Werner R.W. Scheinhardt, and Yana Volkovich</i>	
Communities in Large Networks: Identification and Ranking	84
<i>Martin Olsen</i>	
Combating Spamdexing: Incorporating Heuristics in Link-Based Ranking	97
<i>Tony Abou-Assaleh and Tapajyoti Das</i>	
Traps and Pitfalls of Topic-Biased PageRank	107
<i>Paolo Boldi, Roberto Posenato, Massimo Santini, and Sebastiano Vigna</i>	
A Scalable Multilevel Algorithm for Graph Clustering and Community Structure Detection	117
<i>Hristo N. Djidjev</i>	
A Phrase Recommendation Algorithm Based on Query Stream Mining in Web Search Engines	129
<i>M. Barouni-Ebrahimi and Ali A. Ghorbani</i>	
Characterization of Graphs Using Degree Cores	137
<i>John Healy, Jeannette Janssen, Evangelos Milios, and William Aiello</i>	

Web Structure Mining by Isolated Stars	149
<i>Yushi Uno, Yoshinobu Ota, and Akio Uemichi</i>	
Representing and Quantifying Rank - Change for the Web Graph	157
<i>Akrivi Vlachou, Michalis Vazirgiannis, and Klaus Berberich</i>	
Author Index	167