

The AAI-05 Workshop on Link Analysis

LinkAnalysis-2005

Preface

Link Analysis has been developed over the past 20 years in various fields including Discrete Mathematics (Graph Theory), Social Sciences (Social Network Analysis) and Computer Science (graph as a data structure). Recently this area has attracted a wider attention for its applicability in law enforcement investigations (e.g., terrorism, money laundering), fraud detection (e.g., insurance, banking), WWW analysis (e.g., search engines, marketing), telecommunications (e.g., routers, traffic, connectivity), and similar.

Particularly interesting are problems and issues that fall within the intersection of Link Analysis and fields such as Web and Text Mining, Relational Data Mining, and more general, Data Mining. Typical examples are in the area of trend analysis, community identification, Web user profiling, media clipping, marketing, etc., where Link Analysis complements other fields of research and derives additional value from information processing. Another interesting scenario is extraction of information from unstructured data, representation of extracted data in the graphical form, and further analysis of the resulting graph structure to derive and discover new knowledge.

Publications included in the LinkAnalysis-2005 Workshop cover a range of topics, attesting to the richness and versatility of this research area.

Methods for manipulation of graphs are covered in several papers, including research on pattern matching efficiency of semantic graphs using higher-order constructs and extraction of relevant semantic sub-graphs to facilitate learning using Bayesian Networks.

Data analysis workflows generally include a variety of tools in order to solve a problem. Graph analysis is typically just one of the components used. As expected, different application areas impose different types of workflow for analysts and thus the tools need to be flexible in receiving and transferring intermediary analysis results. The paper on PHERL takes a step in the direction of specifying a representation language to enable exchange of patterns, hypotheses, and evidence among analysis tools. It aspires to incite the rest of the research community to contribute further requirements for PHERL.

Research results from one area often inspire novel solutions in another. Similarly, a combination of techniques from several fields yield novel approaches. The Workshop includes a several papers that show the cross application benefits.

For example, results from social networks research are applied to the problem of ranking autonomous systems on the Internet. Another paper addresses the management of servers based on the power law relationships observed in the network topologies themselves. This enables us to look at the problem area from different perspectives.

Furthermore, the discovery of link structure and exploitation of graph properties is becoming a common trend in information retrieval. A paper on topic-specific scoring of documents combines the standard PCA method with link properties of the topics structure to enhance document retrieval. Similarly, the paper on summarization of broadcast news video exploits the link analysis of named entities.

These and related techniques can be boosted by the availability and quality of a domain ontology and thus the issue of automatic extraction and structuring of domain specific terms is of great importance. The Workshop includes a paper on HITS, a graph based ranking algorithm that identifies domain keywords and exploits the dependencies among terms to structure them as concepts and attributes.

Applications often challenge the standard research methods. Such is the application of relational graphs analysis to the problem of detecting tax fraud. The sheer size of data requires a re-engineering effort to implement custom data structure and identify optimal strategies by combining graph representation and cost based optimization. The presented paper illustrates how such a research effort resulted into a prototype that has been deployed in practice and demonstrated its benefits.

Program Chairs:

Dunja Mladenic, email: <dunja@cs.cmu.edu>

J. Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

Carnegie Mellon University, School of Computer Science, Pittsburgh, USA,

Natasa Milic-Frayling, email: <natasamf@microsoft.com>

Microsoft Research Ltd, Roger Needham Building, 7 J J Thomson Avenue

Cambridge CB3 0FB, United Kingdom

Marko Grobelnik, email:<marko.grobelnik@ijs.si>

J. Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia

Program Committee:

- Vladimir Batagelj, University Ljubljana, Slovenia
- Janez Brank, J.Stefan Institute, Ljubljana, Slovenia
- Ronen Feldman, ClearForest, Israel
- Rayid Ghani, Accenture Technology Labs - Research, Chicago, USA
- Marko Grobelnik, J.Stefan Institute, Ljubljana, Slovenia
- Rosie Jones, Yahoo! Research, USA
- Jure Leskovec, Carnegie Mellon University, Pittsburgh, USA
- Natasa Milic-Frayling, Microsoft Research Ltd, Cambridge, UK
- Dunja Mladenic, J.Stefan Institute, Ljubljana, Slovenia
- Ion Muslea, Language Weaver, Inc., Marina del Rey, CA, USA

Table of Contents

- **The Role of Higher-Order Constructs in the Inexact Matching of Semantic Graphs**, Michael Wolverton, Jerome Thomere
- **Knowledge Representation and Bayesian Inference for Response to Situations**, Rakesh Gupta, Vasco Calais Pedro
- **Efficiency and Fairness of Load-distribution Based on Scale-free Property**, Kensuke Fukuda, Shin-ya Sato, Osamu Akashi, Kazuhiro Kazama, Toshio Hirotsu, Satoshi Kurihara, Toshiharu Sugawara
- **A Social Network Approach for the Ranking of the Autonomous Systems of the Internet**, Fabrice Clérot, Quang Nguyen
- **Relational Graph Analysis with Real-World Constraints: An Application in IRS Tax Fraud Detection**, Eric Bloedorn, Neal J. Rothleder, David DeBarr, Lowell Rosen
- **Domain Term Extraction and Structuring via Link Analysis**, Nasreen AbdulJaleel, Yan Qu
- **Topic-Specific Scoring of Documents with Discrete PCA**, Wray Buntine, Jaakko Lofstrom, Sami Perttu, Kimmo Valtonen
- **Summarization of Broadcast News Video through Link Analysis of Named Entities**, Norman Papernick, Alexander G. Hauptmann
- **PERL: an Emerging Representation Language for Patterns, Hypotheses, and Evidence**, Kenneth Murray, Ian Harrison, John Lowrance, Andres Rodrigue Rodriguez, Jerome Thomere, Michael Wolverton