



AMiner (ArnetMiner)

— Toward understanding big scholar data

<http://aminer.org>

Jie Tang
Tsinghua University

MINING DEEP KNOWLEDGE FROM SCIENTIFIC NETWORKS

Whatever comes to your mind



 Advanced

Hot Topics

Data Mining



Deep Learning



Computer Vision



Database



136,722,706
RESEARCHERS

101,390,721
PUBLICATIONS

7,854,301
CONCEPTS

133,196,029
CITATIONS

Expert Search

Finding experts,
for “data mining”



Demographics: gender, language, location, etc.

Expert **Paper**

Search Results for "data mining"

H-Index :

Gender : Male (934) Female (44)

Language : Chinese (290) English (194) Greek (37) German (27) French (23) Japanese (20) Korean (14) Indian (12)

Location : USA (219) China (141) Taiwan (34) Australia (33) Canada (29) Japan (25) Germany (24) Italy (20) Hong Kong (20) Singapore (20)

Relevance ⓘ H-Index A-Index Activity Diversity Rising Star #Citation #Paper

Jiawei Han (韩家炜) Ⓜ  
H-Index: 126 | #Paper: 790 | #Citation: 90481
Department of Computer Science, University of Illinois at Urbana-Champaign
Professor
Data Mining, Information Extraction, Text Mining
Similar Authors: 
2567 views

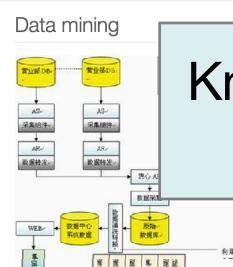
Similar Authors

Philip S. Yu Ⓜ  
H-Index: 124 | #Paper: 838 | #Citation: 69439
Department of Computer Science, University of Illinois Chicago
Professor and Wexler Chair in Information Technology
Distributed System, Query Optimization, Query Processing, Database Systems
Data Mining
321 views

Hillol Kargupta Ⓜ  
H-Index: 40 | #Paper: 141 | #Citation: 6192
Department of Computer Science and Electrical Engineering University of Maryland Baltimore County
Associate Professor
Data Mining, Machine Learning, Data Analysis, Knowledge Discovery
Genetic Algorithms
118 views

Xindong Wu Ⓜ  
H-Index: 45 | #Paper: 331 | #Citation: 9644
Department of Computer Science, University of Vermont
Professor
Machine Learning, Information Extraction, Bayesian Networks, Data Mining
Supervised Learning
35 views

Data mining



Knowledge about “data mining”

Data mining (the analysis step of the “Knowledge Discovery in Databases” process, or KDD), a relatively young and interdisciplinary field of computer science, is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data preprocessing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.

Super Concepts:

Data analysis Data mining Formal sciences Applied sciences Networks Artificial intelligence

Related Concepts:

Data compression Data visualization Natural language processing Data cleansing Distributed computing Informatization Speech recognition Business intelligence Pattern recognition Spatial database Full text search Metadata Computer vision ISAM Biological neural network Database Grid computing Database marketing Parallel computing

数据挖掘

数据挖掘(Data Mining)是通过分析每个数据，从大量数据中寻找其规律的技术，主要有数据准备、规律寻找和规律表示3个步骤。数据挖掘的任务有关联分析、聚类分析、分类分析、异常分析、特异群组分析和演变分析等。

上位词:

资料分析 计算机科学基础理论 决策支持系统 信息管理术语 数据挖掘 形式科学

Researcher Profile

AMiner Whatever comes to your mind Home | 


Upload

Jiawei Han (韩家炜)



Department of Computer Science, University of Illinois at Urbana-Champaign

Professor

(217) 333-6903

hanj@cs.uiuc.edu

http://www.cs.uiuc.edu/~hanj/

External Links



Research Interests

Update

Basic Info.

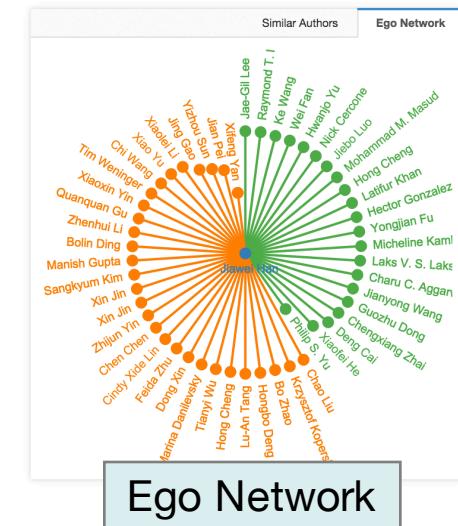
Research Interests

● Data Mining ● Information Extraction ● Data Analysis
● Machine Learning ● Text Mining

Citation Statistics



Ego Network





[Watch +](#)



[Watch +](#)

Bringing Structure to Text: Mining Phrases, Entity Concepts, Topics, and Hierarchies

20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), New York 2014

Mining Massive RFID, Trajectory, and Traffic Data Sets

14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Las Vegas 2008

[About](#)

[Papers](#)

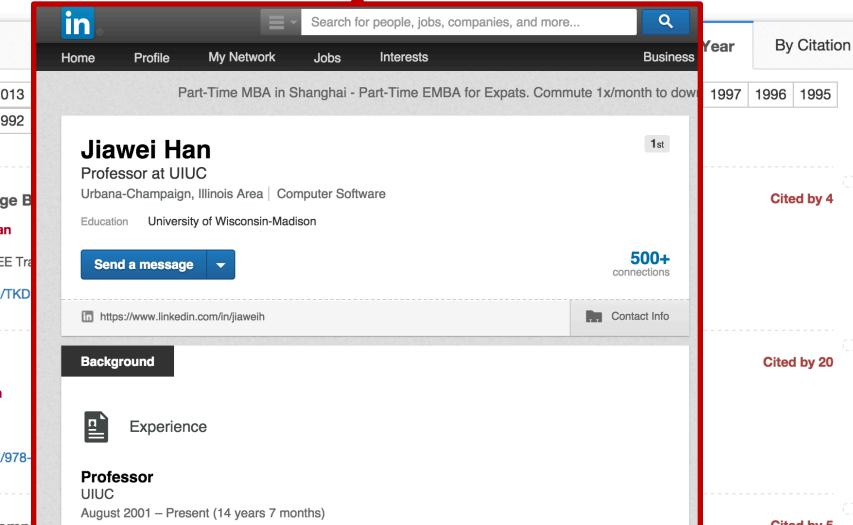
[Lectures](#)

[Patents](#)

[790](#)

[13](#)

[1](#)



Jiawei Han
Professor at UIUC
Urbana-Champaign, Illinois Area | Computer Software
Education University of Wisconsin-Madison

Send a message 

500+ connections

Part-Time MBA in Shanghai - Part-Time EMBA for Expats. Commute 1x/month to down
Cited by 4

Background

Experience

Professor
UIUC
August 2001 – Present (14 years 7 months)

Cited by 20

Troubleshooting interactive comp

Mohammad Maifi Hasan Khan, Hieu Khao Le, Hossein Ahmadi, Tarek F. Abdelzaher, **Jiawei Han**
ACM Transactions on Sensor Networks (TOSN) (2014)
Bibtex <http://dx.doi.org/10.1145/2530290>

787

1

Systems and Methods for Detecting a Novel Data Class

Mohammad Mehedy Masud, Latifur Rahman Khan, Bhavani Marienne Thuraisingham, Qing Chen, Jing Gao, Jiawei Han

Publication-date: 2012-03-01 Application-date: 2011-08-22

4



清华大学
Tsinghua University



AMiner

□ Academic search and mining system – <http://aminer.org>

- Online since 2006
- >136 million researcher profiles
- >101 million publication papers
- >341 million requests
- 100K IP access from 170 countries per month

□ Deep analysis, mining, and search

The screenshot displays the ArnetMiner interface. At the top, a navigation bar includes links for Home, Conference, Collaborator, Geo Search, Topics, Download, Admin, More, and Account. A "Welcome jietang" message is shown.

Profile View: A detailed profile for "Jiawei Han" is shown, featuring a photo, contact information (Position: Professor, Affiliation: Department of Computer Science, University of Illinois at Urbana-Champaign), and a statistics section (H-index: 96, #Papers: 553, #Citations: 55885). Below this is a "Bio" section and a "Research Interest" section listing topics like Efficient Mining, Spatial Data Mining, and Frequent Pattern Mining.

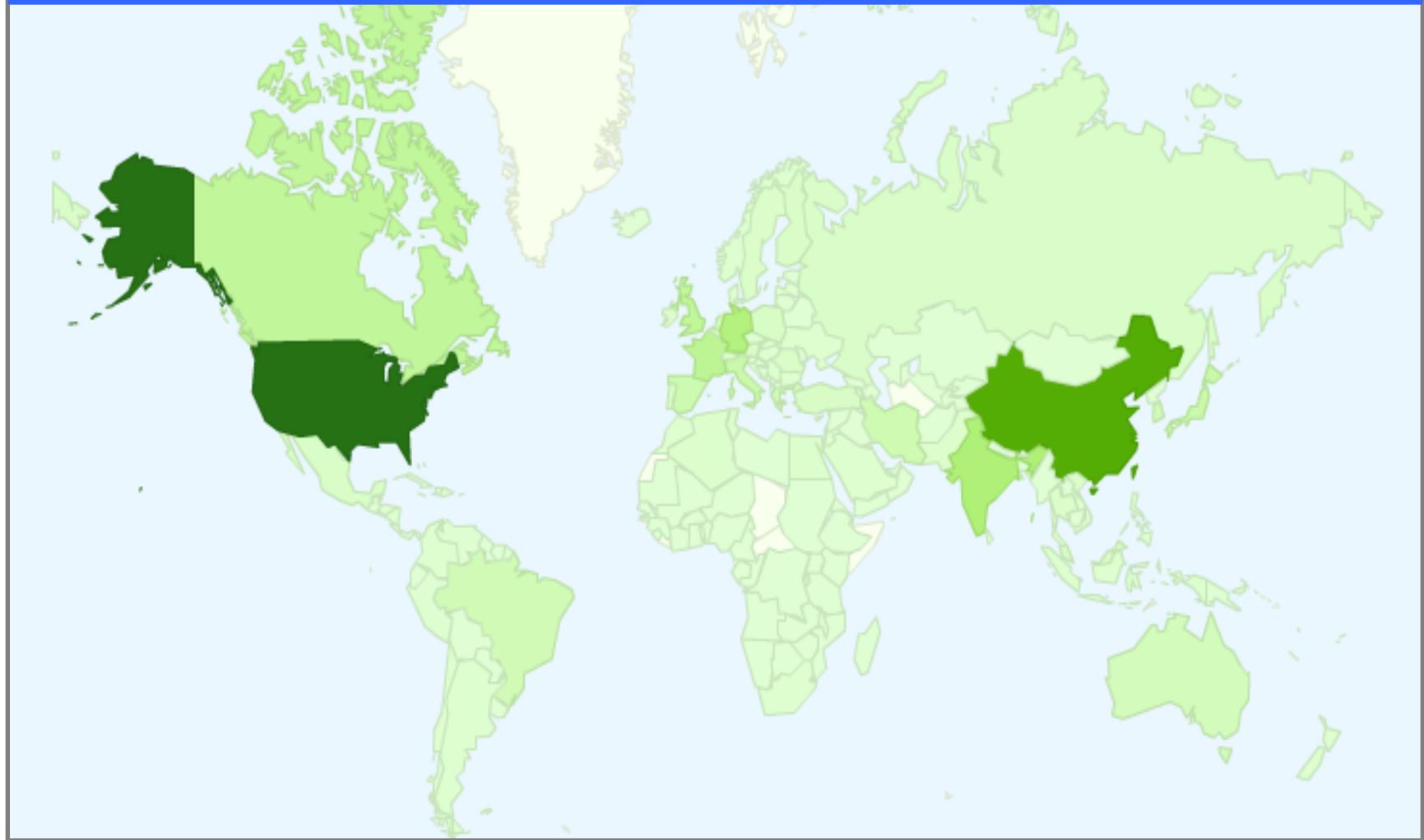
Search Results: A search for "data mining" yields results for researchers such as Jiawei Han, Philip S. Yu, and Mohammed Javed Zaki, each with their own profile card and publication details.

Social Graph: A network graph titled "Social Graph" shows connections between various researchers based on their collaboration history.

Publications: A section titled "Publications" allows users to group results by year or category, with a link to "SEARCH PUBLICATION RESULTS".

User Distribution

8.32 million IP from 220 countries/regions



User Distribution

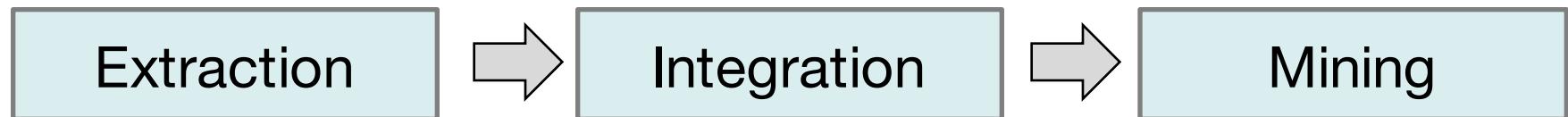
8.32 million IP from 220 countries/regions

Top 10 countries

- | | |
|------------|-----------|
| 1. USA | 6. Canada |
| 2. China | 7. Japan |
| 3. Germany | 8. Spain |
| 4. India | 9. France |
| 5. UK | 10. Italy |

Technologies

— How to populate a **semantic**-based profile database for researchers?



Extracting Semantics from the Web

(ACM TKDD, WWW'12, ISWC'06, ICDM'07, ACL'07)

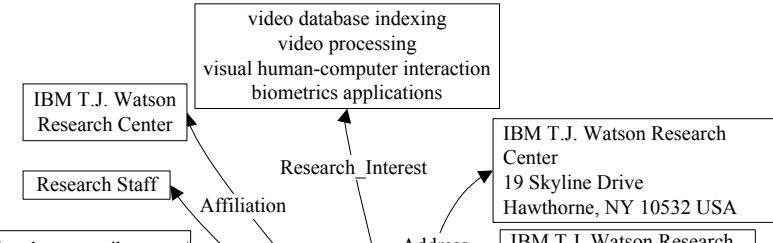


Ruud Bolle 2
Office: 1S-D58

Contact Information

Ruud Bolle 1

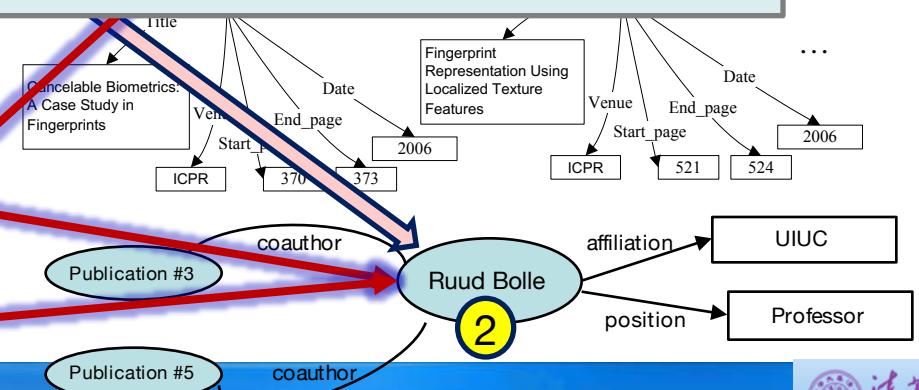
Office: 1S-D58
Letters: IBM T.J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598 USA
Packages: IBM T.J. Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532 USA
Email: bolle@us.ibm.com



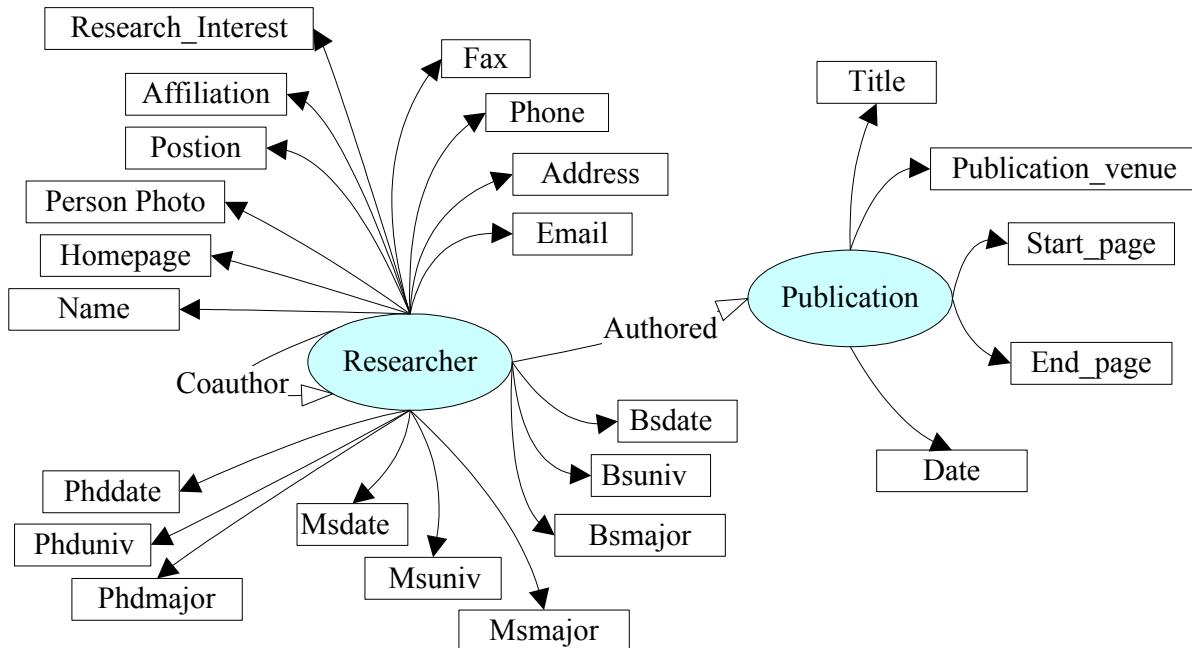
Two questions:

- How to accurately extract the semantic-based profile information from the Web?
- How to deal with the name ambiguity problem?

DBLP: R				
2006				
1	EE	Nalini K. Ratha, Jonathan Connell, Ruud M. Bolle, Sharat Chikkerur: Cancelable Biometrics: A Case Study in Fingerprints. ICPR (4) 2006: 370-373		
2	EE	Sharat Chikkerur, Sharath Pankanti, Alan Jea, Nalini K. Ratha, Ruud M. Bolle: Fingerprint Representation Using Localized Texture Features. ICPR (4) 2006: 521-524		
48	EE	Andrew Senior, Arun Hampapur, Ying-li Tian, Lisa Brown, Sharath Pankanti, Ruud M. Bolle: Appearance models for occlusion handling. Image Vision Comput. 24(11): 1233-1243 (2006)		
2005				
1	EE	Ruud M. Bolle, Jonathan H. Connell, Sharath Pankanti, Nalini K. Ratha, Andrew W. Senior: The Relation between the ROC Curve and the CMC. AutoID 2005: 15-20		
2	EE	Sharat Chikkerur, Venu Govindaraju, Sharath Pankanti, Ruud M. Bolle, Nalini K. Ratha: Novel Approaches for Minutiae Verification in Fingerprint Images. WACV. 2005: 111-116		
...				



Researcher Network Extraction^[1,2]



70.60% of the researchers have at least one homepage or an introducing page

85.6% from universities	14.4% from companies
-------------------------	----------------------

71.9% are homepages	28.1% are introducing pages
---------------------	-----------------------------

40% are in lists and tables	60% are natural language text
-----------------------------	-------------------------------

There are a large number of person names having the ambiguity problem

Even 3 “Yi Li” graduated the author’s lab

70% moved at least one time

[1] J. Tang, L. Yao, D. Zhang, and J. Zhang. A Combination Approach to Web User Profiling. ACM Transactions on Knowledge Discovery from Data (TKDD), (vol. 5 no. 1), Article 2 (December 2010), 44 pages.

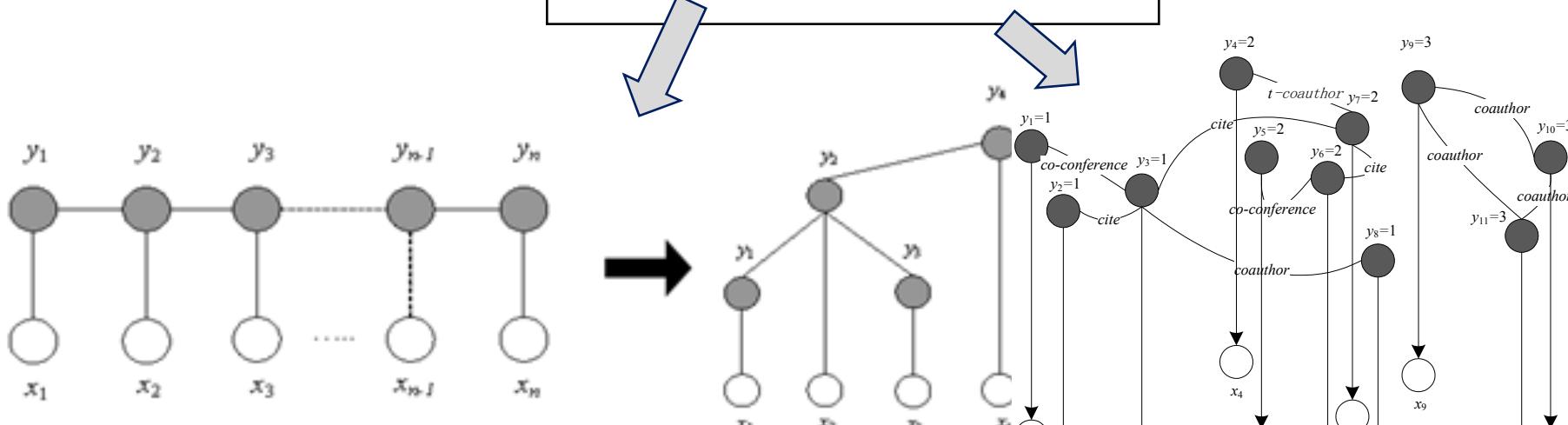
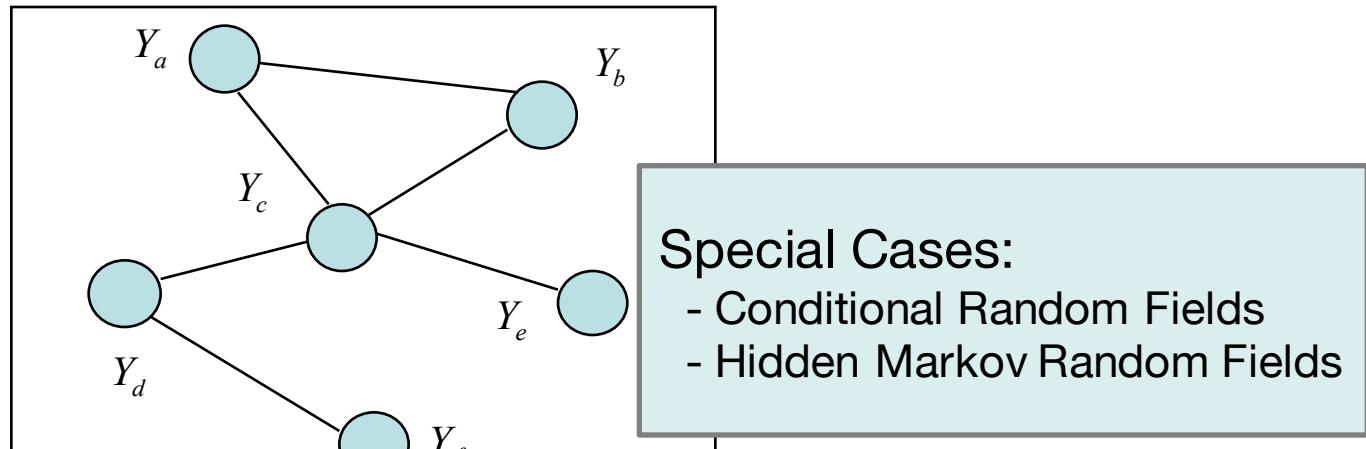
[2] J. Tang, D. Zhang, and L. Yao. Social Network Extraction of Academic Researchers. ICDM’07. pp. 292-301.

Our Approach Picture

– based on Markov Random Field

Markov Property:

$$P(Y_i | \{Y_j | Y_j \neq Y_i\}) \\ = P(Y_i | \{Y_j | Y_j \sim Y_i\})$$



Researcher Profiling

$$\sum_{E^P, E^H, j} \lambda_j t_j(e, y|_e, x) +$$

Name Disambiguation



Researcher Profile Database

ArnetMiner Home Conference Collaborator Geo Search Topics Download Admin More > Welcome jetang Account Search Experts Search

Jiawei Han

Position: Professor
Affiliation: Department of Computer Science, University of Illinois at Urbana-Champaign
Address: 200 N. Goodwin Avenue, Urbana, IL 61801, USA.
Phone: (217) 333-8903
Fax: (217) 265-6494
Email: han@cs.illinois.edu
Links: [Home](#) [Google Scholar](#)

STATISTIC
H-index: 96 Uptrend: 30.46 Diversity: 0.71
#Papers: 553 Activity: 32.04 Sociability: 726.64
#Citations: 55885 Longevity: 26 [More Statistics...](#)

Bio
Jiawei Han is a computer scientist who specializes in research on Data Mining. He was the 2009 winner of the Mcowell Award, the highest technical award made by IEEE. He is currently a professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign. Previously he was a professor in the School of Computing Science at Simon Fraser University. He is an ACM fellow and an IEEE fellow.

Research Interest
Efficient Mining, Spatial Data Mining, Frequent Pattern Mining. [Edit Bio](#)

Education
PhD University [Edit Interest](#)

[FOAF](#) [Follow](#)

M. I. Jordan

ALIAS: Michael I. Jordan, Michael Jordan, Michael Irwin Jordan

Position: Professor
Affiliation: Department of EECS Department of Statistics University of California, Berkeley
Address: University of California, Berkeley EECS Department 731 Soda Hall #1776 Berkeley, CA 94720-1776
Phone: (510) 642-3806
Fax: (510) 642-5775
Email: jordan@stat.berkeley.edu
Links: [Home](#) [Google Scholar](#)

STATISTIC
H-index: 75 Uptrend: 7.2 Diversity: 0.03
#Papers: 242 Activity: 11.12 Sociability: 331.69
#Citations: 44312 Longevity: 23 [More Statistics...](#)

[Edit Profile](#)

H. Garcia

ALIAS: H. Garcia Molina, H. Garcia-Molina, Hector Garcia-Molina, Hector Garcia Molina

Position: Professor
Affiliation: Departments of Computer Science and Electrical Engineering,

See Others:
Andreas Paepcke Coauthor-Count: 32 H-index: 37
Jennifer Widom Coauthor-Count: 24 H-index: 79
A. Barbara Coauthor-Count: 26 H-index: 0

Expertise:
Data (115) Database Systems (60) Database Systems / Automated Data Test (30) Data Mining (23) Mobile Robot / Hybrid Control (22) Library / Information Access

Conference:
KDD (49) ICDE (40) IEEE Trans. Knowl. Data Eng. (36) SIGMOD Conference (32) VLDB (24)

Research Interest
Database Systems, Data Management, Data Warehousing. [Edit Bio](#)

Education
Stanford University [Edit Interest](#)

[FOAF](#) [Follow](#)

ArnetMiner Home Conference Collaborator Geo Search Topics Download Admin More > Welcome Scott Account Search Experts Search

Scott

Position: Professor
Affiliation: University of California, Berkeley
Address: 720 Soda Hall, Berkeley, CA 94720-1776
Phone: (510) 642-5775
Links: [Home](#) [Google Scholar](#)

STATISTIC
H-index: 96 Uptrend: -4.04 Diversity: 0.22
#Papers: 195 Activity: 4.86 Sociability: 407.19
#Citations: 57908 Longevity: 25 [More Statistics...](#)

Expertise:
Wireless network / End-to-end Routing Behavior (80) ATM Networks (21)

Research Interest
Database Systems, Data Management, Data Warehousing. [Edit Bio](#)

Education
University of California, Berkeley [Edit Interest](#)

[Edit Profile](#)

Extracted more than 1,000,000 researcher profiles from the Web



Name Disambiguation

Name	Affiliation
Jing Zhang (26)	Shanghai Jiao Tong Univ.
	Yunnan Univ.
	Tsinghua Univ.
	Alabama Univ.
	Univ. of California, Davis
	Carnegie Mellon University
	Henan Institute of Education

Jing Zhang

List of publications from the DBLP Bibliography Server - [FAQ](#)

[Coauthor Index](#) - [Ask others](#): [ACM DL](#) - [ACM Guide](#) - [CiteSeer](#) - [CSB](#) - [Google](#)

2007

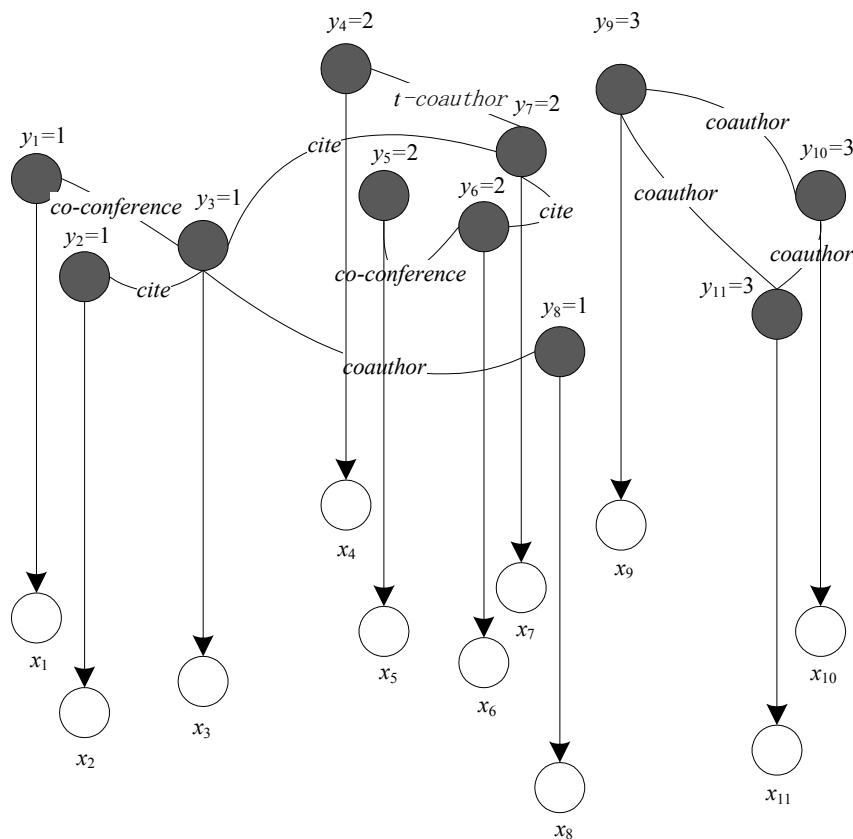
83	EE	Jing Zhang, Guizhong Liu: Hyperspectral images lossless compression by a novel three-dimensional wavelet coding
82	EE	Jing Peng, Dongqing Yang, Changjie Tang, Jing Zhang, Jianjun Hu: CACS: A Novel Classification Algorithm Base
81	EE	Jing Zhang, Xi Chen, Ming Li: Computing Exact p-Value for Structured Motif. <i>CPM 2007</i> : 162-172
80	EE	Jing Zhang, Jie Tang, Juan-Zi Li: Expert Finding in a Social Network. <i>DASFAA 2007</i> : 1066-1069
79	EE	Guojun Chen, Jing Zhang, Xiaoli Xu, Yuan Yin: Real-Time Visualization of Tire Tracks in Dynamic Terrain with LO
78	EE	Jing Zhang, Hai Huang: Federate Job Mapping Strategy in Grid-Based Virtual Wargame Collaborative Environment
77	EE	Maria Wimmer, Michael Gouli, Jing Zhang: Minitrack: E-Government Information and Knowledge Management. <i>HICSS</i> 40: 1-10, 2007
76	EE	Kai Kang, Jing Zhang, Baoshan Xu: Optimizing the Selection of Partners in Collaborative Operation Networks. <i>ICCSN</i> 2007: 1-5
75	EE	Lingshuang Shao, Jing Zhang, Yong Wei, Junfeng Zhao, Bing Xie, Hong Mei: Personalized QoS Prediction for Web Services. <i>ICCSN</i> 2007: 1-5
74	EE	Benyong Liu, Jing Zhang, Xiaowei Chen: Adaptive Training of a Kernel-Based Representative and Discriminative Feature. <i>ICCSN</i> 2007: 1-5
73	EE	Jilong Wang, Jing Zhang: Federation Based Solution for Peer-to-Peer Network Management. <i>International Conference on Computer Science and Software Engineering (ICCSSE)</i> , Volume 2, pages 1-4, 2007
72	EE	Jing Zhang, Fanhuai Shi, Jianhua Wang, Yuncai Liu: 3D Motion Segmentation from Straight-Line Optical Flow. <i>Mobile Data Management (MDM)</i> , Volume 2, pages 1-4, 2007

- How to perform the assignment automatically?
- How to estimate the person number?

[1] J. Tang, A.C.M. Fong, B. Wang, and J. Zhang. A Unified Probabilistic Framework for Name Disambiguation in Digital Library. *IEEE Transaction on Knowledge and Data Engineering (TKDE)* , Volume 24, Issue 6, 2012, Pages 975-987.

[2] X. Wang, J. Tang, H. Cheng, and P. S. Yu. ADANA: Active Name Disambiguation. *ICDM'11*, pages 794-803, 2011.

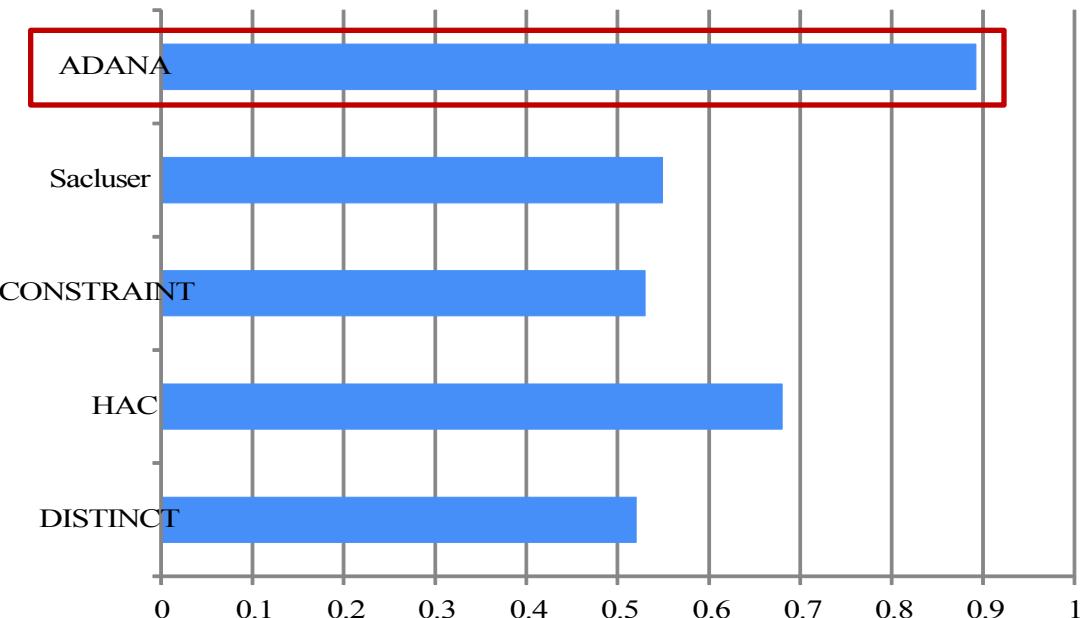
Our Method to Name Disambiguation



- A hidden Markov Random Field model
- Observable Variables X represent publications
- Hidden Variables Y represent the labels of publications
- Paper relationships define the dependencies over hidden variables



Disambiguation Performance



ArnetMiner Home Conference Collaborator Geo Search Topics Download Admin More Welcome jietang Account

Search Experts Search

Name Disambiguation Results for 5 "Hang Li". Show all Hang Li. 1/5

0 H-index: 36 #Papers: 121 #Citation: 3708 1 H-index: 8 #Papers: 9 #Citation: 160 2 H-index: 2 #Papers: 9 #Citation: 10 3 H-index: 0 #Papers: 0 #Citation: 0 4 H-index: 0 #Papers: 0 #Citation: 0 Dept. of Electr. Eng., Te

Hang Li FOAF Follow

Position: Senior Researcher and Research Manager
Affiliation: Information Retrieval and Mining Group, Microsoft Research Asia
Address: No. 5, Da Ling Street, Haidian District, Beijing China, 100080
Phone: (86-10)58963177
Fax: (86-10)82868529
Email: hangli@microsoft.com

Links: [Edit Profile](#)

STATISTIC

H-index:	36	Uptrend:	2.98	Diversity:	1.35
#Papers:	121	Activity:	6.77	Sociability:	204.88
#Citations:	3708	Longevity:	17	More Statistics...	

Bio

I am senior researcher and research manager of Information Retrieval and Mining Group at Microsoft Research Asia. I am also adjunct professor of Peking University, Nanjing University, Xian Jiaotong University, and Nankai University. I am co-director of the MS joint lab at PKU and member of the machine learning lab at NJU.

I joined Microsoft Research in June 2001. Prior to that, I worked at the Research Laboratories of NEC Corporation.

I obtained a B.S. in Electrical Engineering from Kyoto University in 1988 and a M.S. in Computer Science from Kyoto University in 1990. I earned my Ph.D. in Computer Science from the University of Tokyo in 1998.

I am interested in statistical learning, information retrieval, data mining, and natural language processing.

[Edit Bio](#)

Research Interest

Information Retrieval Text Classification Web Search [Edit Interest](#)

See Others:

- Tie-Yan Liu Coauthor-Count: 30 H-index: 25 [Follow](#)
- Tao Qin Coauthor-Count: 13 H-index: 16 [Follow](#)
- Naoki Abe Coauthor-Count: 13 H-index: 22 [Follow](#)

Expertise:

- Information Retrieval / Probabilistic Indexing (36)
- Mobile Robot / Hybrid Control (17)
- Natural Language System / Statistical Machine Translation (14)
- Data mining (11)
- Machine Learning (9)
- Wireless Networks / Algorithmic Complexity (9)

Conference:

- SIGIR (15) CoRR (9)
- Int. Process. Manage. (7) ICML (7)
- SIGIR Forum (6) KDD (6)



Still problems?

- **P1:** error propagation in the profile extraction
 - Homepage finding
 - Profile extraction from the identified pages
- **P2:** the “accurately” extracted semantics are not really “accurate”?



Is this Enough?

AMiner Whatever comes to your mind

Home | Follow | Update

Jeannette Wing

Computer Science Department Carnegie Mellon University

Professor of Computer Science

412-260-8926 (cell)

wing@cs.cmu.edu

<http://www.cs.cmu.edu/~wing/>

External Links

Research Interests

Formal Methods Software Engineering Formal Verification

Embedded Systems Specification Languages

Activity #Papers #Citation H-Index Diversity Sociality C-Index

1982 1985 1990 1995 2000 2005 2009

Similar Authors Ego Network

Carol A. Sied Oleg Mikhalev Ipek Ozkaya Richard Demillo Daniel Jackson B. Craig Meyers Miles McQueen Craig Arthur Damon Mark Flynn Laszlo Belady David Luke Dell Daniel Jackson Nancy Levesque Richard Leme Sherman Eagle Nevin Heintze Cordell Green Peter H. Felber Robert Nord Daniel Plekosh Robert Weeks O'Callahan John Goodenough Rick Kazman Daniel Jackson Cited by 1377

Overview

Papers 24

Merge

All (24) Recent 20

By Year By Citation

Computational thinking



Semantics are distributed...

Homepage

Jeannette M. Wing

President's Professor of Computer Science, Computer Science Department, Carnegie Mellon University

Contact Information at MSR

Address: One Microsoft Way, Microsoft Research, Redmond, WA 98052
Phone: 425-706-4459 (office) 425-786-6017 (cell)
Email: wing@microsoft.com

Contact Information at CMU

Address: Computer Science
Email: wing@cs.cmu.edu

General Remarks

A Vision for the 21st Century of Computational Thinking

Scientific Questions in Computing

PCAST Presentation. Pending transcript.

Research

Publications

Areas of Interest: Foundations of programming methodology

Background

Experience

Corporate Vice President
Microsoft
July 2013 – Present (2 years 1 month) | Redmond
In charge of core research labs at Microsoft Research.

President's Professor of Computer Science
Carnegie Mellon University, Computer Science
1985 – Present (30 years)

Wikipedia

Jeannette Wing

From Wikipedia, the free encyclopedia
Pittsburgh

Jeannette Marie Wing is Corporate Vice President of Microsoft

Research with oversight of its core research laboratories around the world and Microsoft Research Connections.^{[2][3]} Prior to 2013, she was the President's Professor of Computer Science at Carnegie Mellon University, Pittsburgh, Pennsylvania, United States. She also served as assistant director for Computer and Information Science and Engineering at the NSF from 2007 to 2010.^{[4][5][6][7][8][9][10][11][12][13]}

Contents [hide]

AMiner

Jeannette Wing

Whatever comes to your mind

Jeannette Wing

Computer Science Department Carnegie Mellon University

Professor of Computer Science

412-260-8926 (cell)

wing@cs.cmu.edu

http://www.cs.cmu.edu/~wing/

External Links

Research Interests

Background

Experience

Corporate Vice President

President's Professor of Computer Science

Carnegie Mellon University

Overview

Papers 24

Merge

All Recent 20 2009 2008 2006 2004 2001 2000 1999 1993 1991 1990 1987 1985 1982

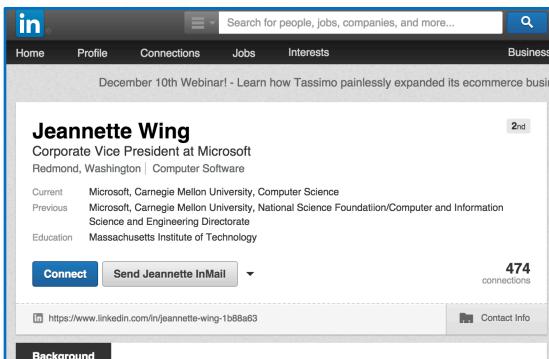
Computational thinking

Cited by 1377

Integrating Semantics across Networks

- Identifying users from multiple heterogeneous networks and integrating semantics from the different networks together.

LinkedIn



December 10th Webinar! - Learn how Tassimo painlessly expanded its ecommerce busin

Jeannette Wing
Corporate Vice President at Microsoft
Redmond, Washington | Computer Software

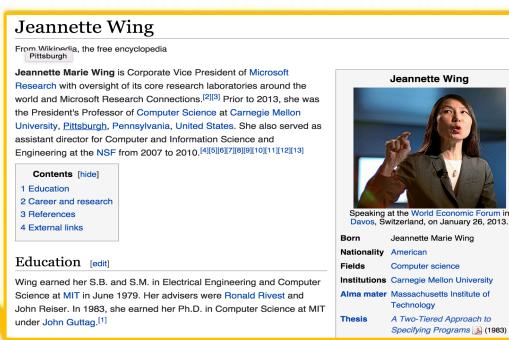
Current: Microsoft, Carnegie Mellon University, Computer Science
Previous: Microsoft, Carnegie Mellon University, National Science Foundation/Computer and Information Science and Engineering Directorate
Education: Massachusetts Institute of Technology

474 connections

Connect Send Jeannette InMail Contact Info

<https://www.linkedin.com/in/jeannette-wing-1b88a63>

Wikipedia



Jeannette Wing
From Wikipedia, the free encyclopedia
Pittsburgh

Jeannette Marie Wing is Corporate Vice President of Microsoft Research with oversight of its core research laboratories around the world and Microsoft Research Connections.^[1] Prior to 2013, she was the President's Professor of Computer Science at Carnegie Mellon University, Pittsburgh, Pennsylvania, United States. She also served as assistant director for Computer and Information Science and Engineering at the NSF from 2007 to 2010.^{[2][3][4][5][6][7][8][9][10][11][12][13]}

Contents [hide]
1 Education
2 Career and research
3 References
4 External links

Education [edit]
Wing earned her S.B. and S.M. in Electrical Engineering and Computer Science at MIT in June 1979. Her advisors were Ronald Rivest and John Reiser. In 1983, she earned her Ph.D. in Computer Science at MIT under John Guttag.^[1]

Born Jeannette Marie Wing
Nationality American
Fields Computer science
Institutions Carnegie Mellon University
Alma mater Massachusetts Institute of Technology
Thesis A Two-Tiered Approach to Specifying Programs (1983)

Same Person



Jeannette Wing
Professor of Computer Science, Oregon State University
Machine Learning · Computational Sustainability · Artificial Intelligence · Reinforcement Learning
Verified email at cs.orst.edu
Homepage

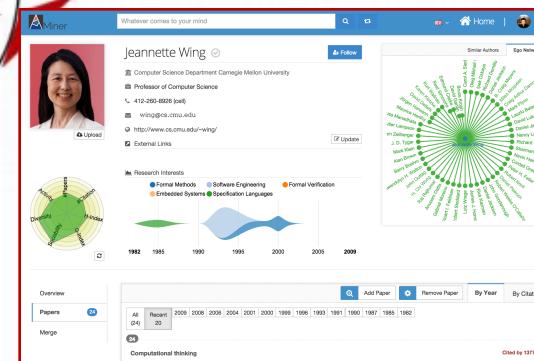
Citation indices
H index: 44 Since 2000: 7369
Citations: 19191
Number: 34
Downloads: 1317
Cited by: 27

Citations to my articles
Follow this author
Follow new articles
Follow new citations
Co-authors
See all 11 co-authors

Recent publications

- Ensemble methods in machine learning
- Solving multiclass learning problems via error-correcting output codes
- Approximate statistical tests for comparing supervised classification learning algorithms
- Neural computation (19.1), 1996-1997
- An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization

Google Scholar



Jeannette Wing ⓘ
Computer Science Department Carnegie Mellon University
Professor of Computer Science
412-268-8205 (cell)
w翼@cs.cmu.edu
<http://www.cs.cmu.edu/~wing/>
External Links

Research Interests
Formal Methods
Embedded Systems
Specification Languages
Formal Verification

Similar Authors
Ego Network

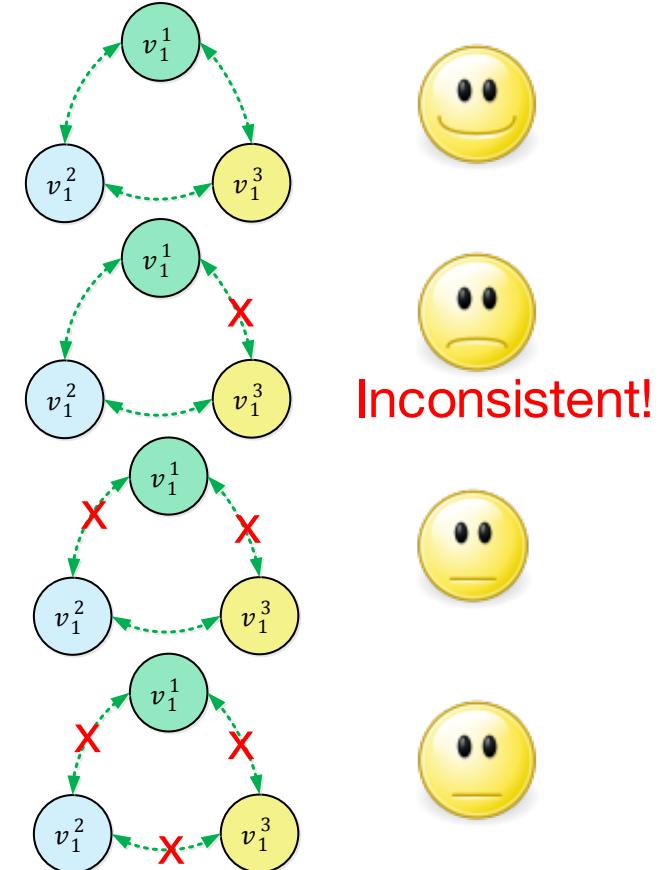
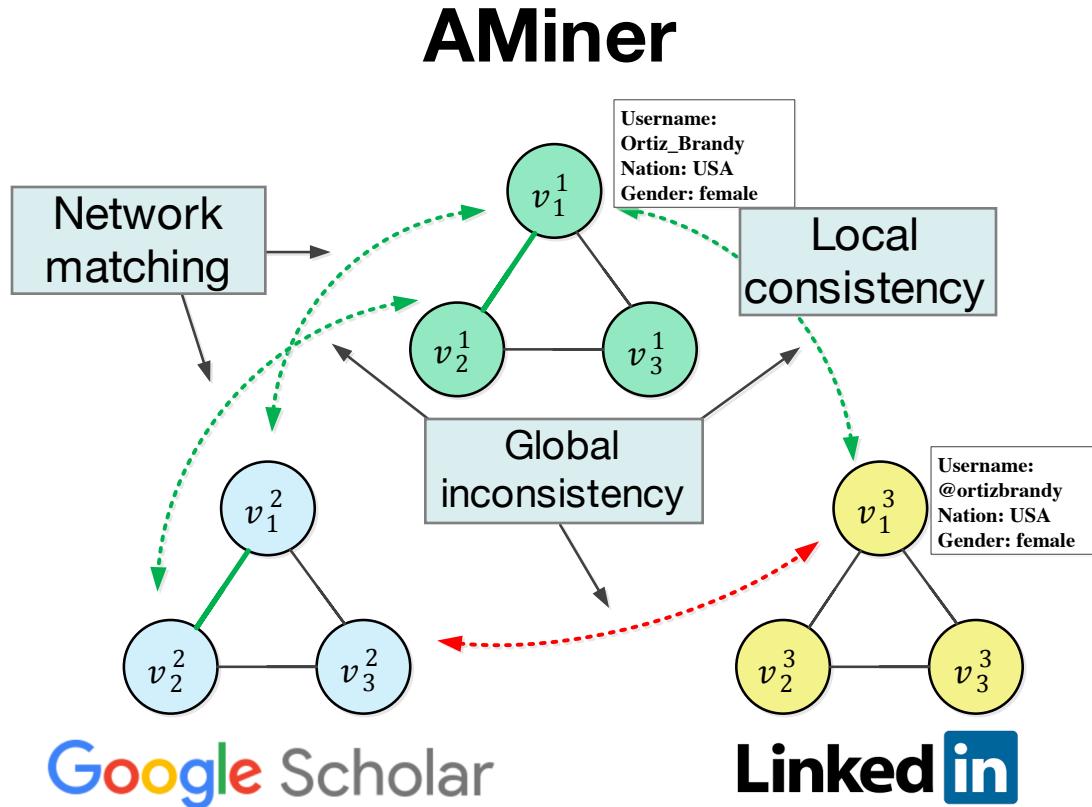
Computational thinking

Overview
Papers Merge

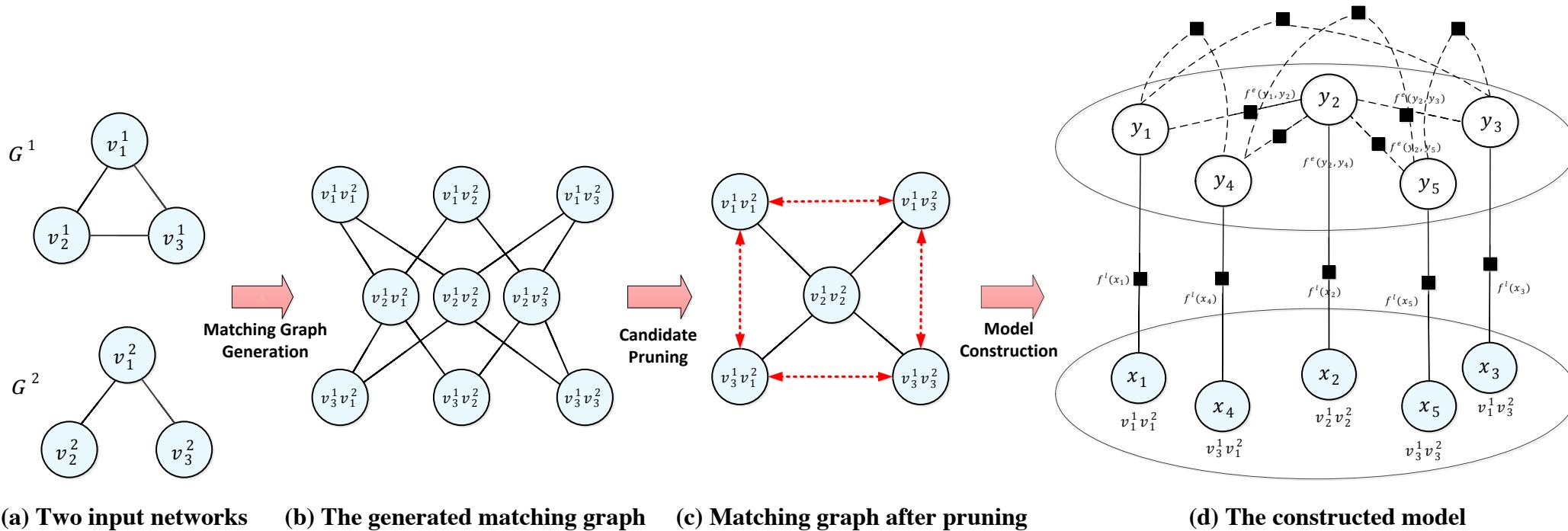
2009 2008 2007 2006 2005 2004 2003 2002 2001 2000 1999 1998 1997 1996 1995 1994 1993 1992 1991 1990 1989 1988 1987 1986 1985 1984 1983 1982 1981 1980 1979 1978 1977

AMiner

Local + Global consistency



COSNET: Connecting Social Networks with Local and Global Consistency



AMiner

Whatever comes to your mind

Jiawei Han (韩家炜)

Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois Area | Computer Software

Professor
(217) 333-6903
hanj@cs.uiuc.edu
<http://www.cs.uiuc.edu/~hanj/>

External Links

Send a message

<https://www.linkedin.com/in/jiaweh>

Background

Experience

Professor
UIUC
August 2001 – Present (14 years 7 months)

500+ connections

Contact Info

Google Scholar

Citation Indices

	All	Since 2011
Citations	113507	56901
H-index	136	96
I10-index	566	485

Co-authors View all...

Xifeng Yan
Philip S. Yu
Xiaofei He
Deng Cai
Jianyong Wang
Yizhou Sun
Jing Gao
Lakshmanan
Guozhu Dong
Charu Aggarwal
Ke Wang
ChengXiang Zhai
Anthony KH Tung

Activity

#Citation

Sociality

H-Index

G-Index

Papers 790

Lectures 13

Patents 1

Add Paper Remove Paper By Year By Citation

All (790) Recent 20 2015 2014 2013 1994 1993 1992

Entity Linking with a Knowledge

Wei Shen, Jianyong Wang, Jiawei Han

Knowledge and Data Engineering, IEEE Transactions on

dx.doi.org/10.1109/TKDE.2015.2432122

Bringing Structure to Text: Mining Phrases, Entity Concepts, Topics, and Hierarchies

20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), New York 2014

Mining Massive RFID, Trajectory, and Traffic Data Sets

14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Las Vegas 2008

Systems and Methods for Detecting a Novel Data Class

Mohammad Mehedy Masud, Latifur Rahman Khan, Bhavani Marienne Thuraisingham, Qing Chen, Jing Gao, Jiawei Han

Publication-date: 2012-03-01 Application-date: 2011-08-22

Troubleshooting interactive complexity bugs in wireless sensor networks using data mining techniques

Mohammad Maifi Hasan Khan, Hieu Khac Le, Hossein Ahmadi, Tarek F. Abdelzaher, Jiawei Han

ACM Transactions on Sensor Networks (TOSN) (2014)

Bibtex <http://dx.doi.org/10.1145/2530290>

787

Search for people, jobs, companies, and more...

Home Profile My Network Jobs Interests Business

1st

500+ connections

Watch +

Watch +

1

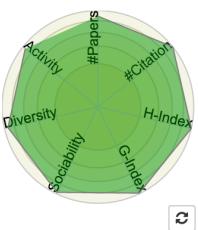
22

Tsinghua University

AMiner



[Upload](#)



About

Papers 790

Lectures 13

Patents 1

Jiawei Han (韩家炜) [Follow](#)

Department of Computer Science, University of Illinois at Urbana-Champaign

Professor

(217) 333-6903

hanj@cs.uiuc.edu

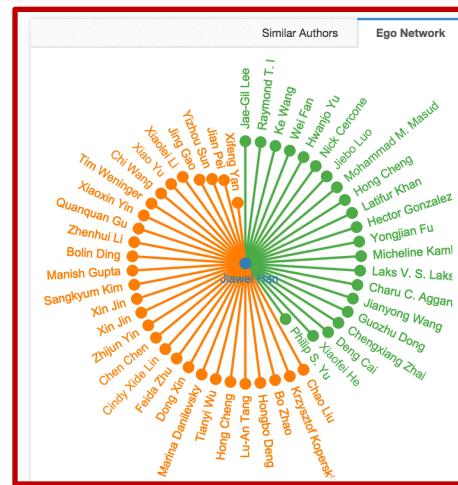
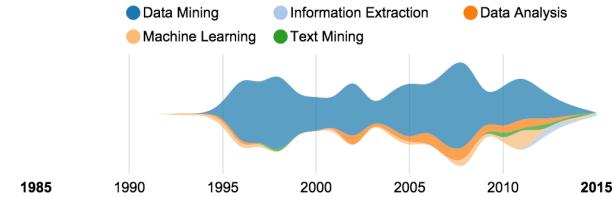
<http://www.cs.uiuc.edu/~hanj/>

[External Links](#)

[Update](#)



Research Interests

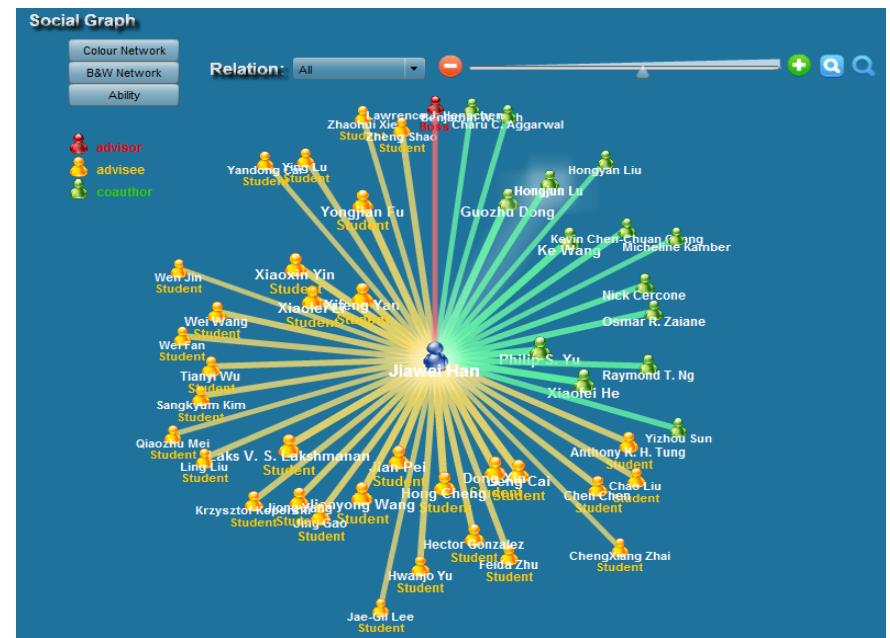
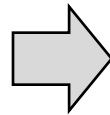
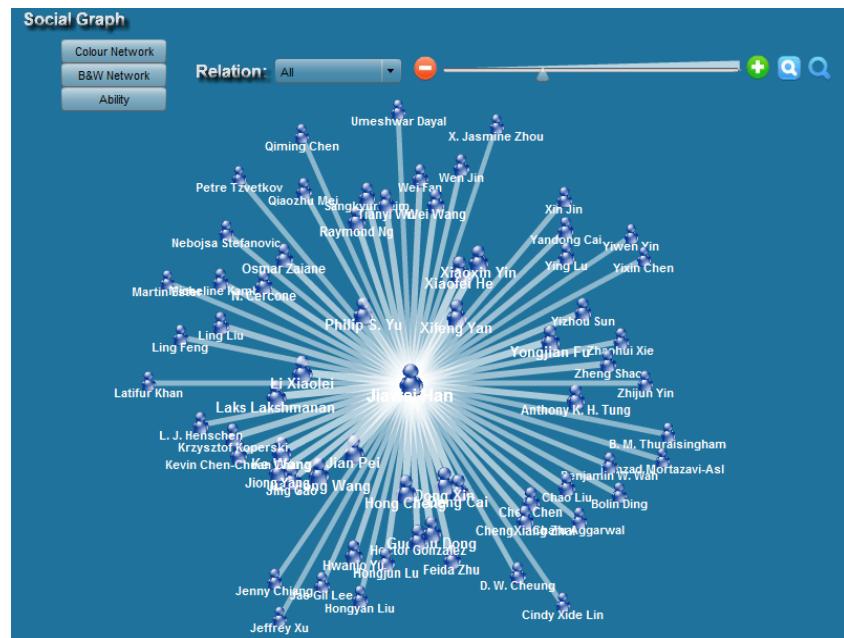


Ego Network

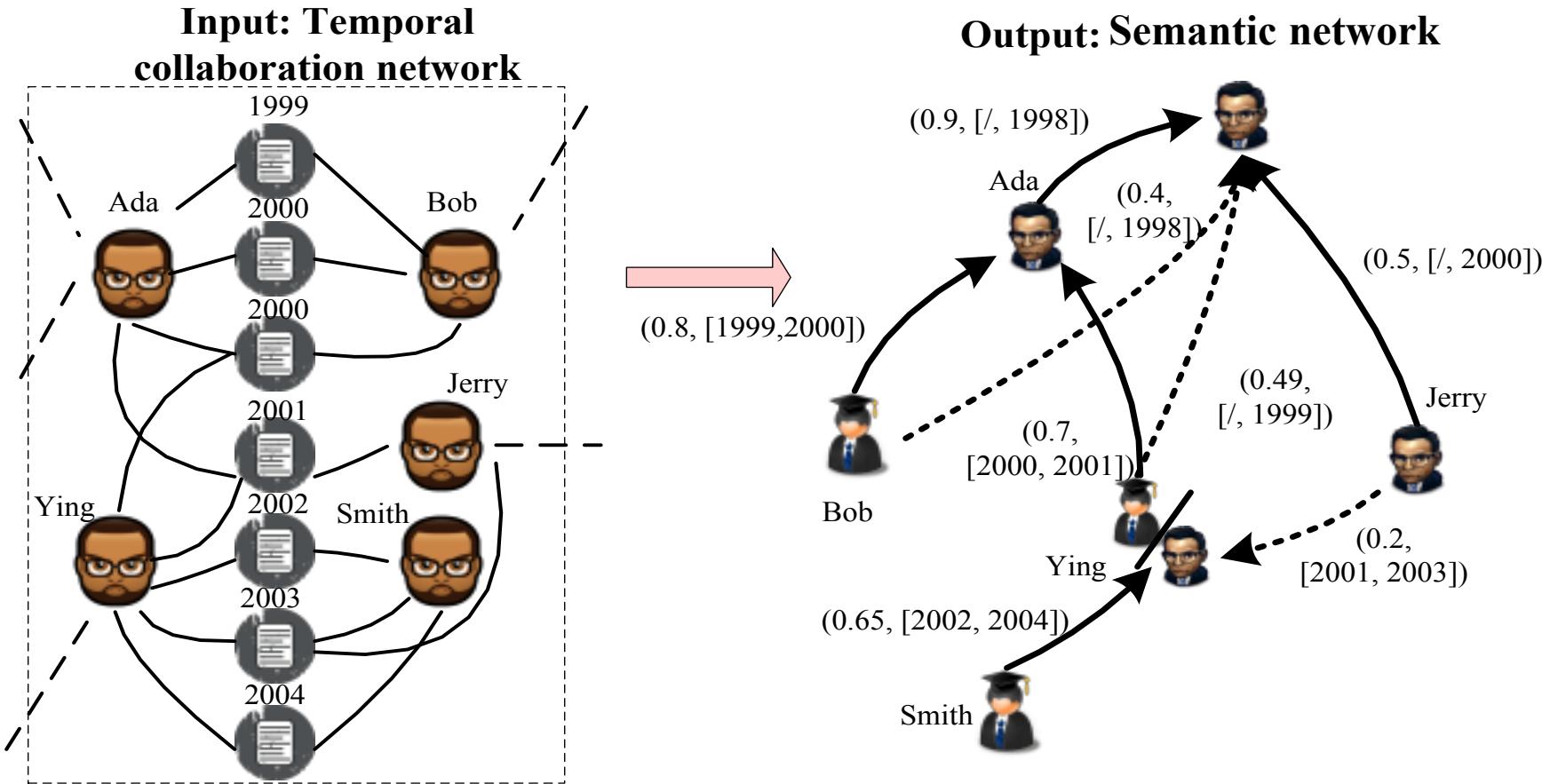
		Add Paper		Remove Paper	By Year	By Citation
All (790)	Recent 20	2015	2014	2013	2012	2011
1994	1993	1992	1991	1990	1989	1988
1987	1986	1985				
790						
Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions Cited by 4						
Wei Shen, Jianyong Wang, Jiawei Han						
Knowledge and Data Engineering, IEEE Transactions (2015)						
BibTex http://dx.doi.org/10.1109/TKDE.2014.2327028						
789						
Mining Graph Patterns. Cited by 20						
Hong Cheng, Xifeng Yan, Jiawei Han						
Frequent Pattern Mining (2014)						
BibTex http://dx.doi.org/10.1007/978-1-4419-6045-0_12						
788						
Troubleshooting interactive complexity bugs in wireless sensor networks using data mining techniques Cited by 5						
Mohammad Maifi Hasan Khan, Hieu Khao Le, Hossein Ahmadi, Tarek F. Abdelzaher, Jiawei Han						
ACM Transactions on Sensor Networks (TOSN) (2014)						
BibTex http://dx.doi.org/10.1145/2530290						
787						

Network Semantics

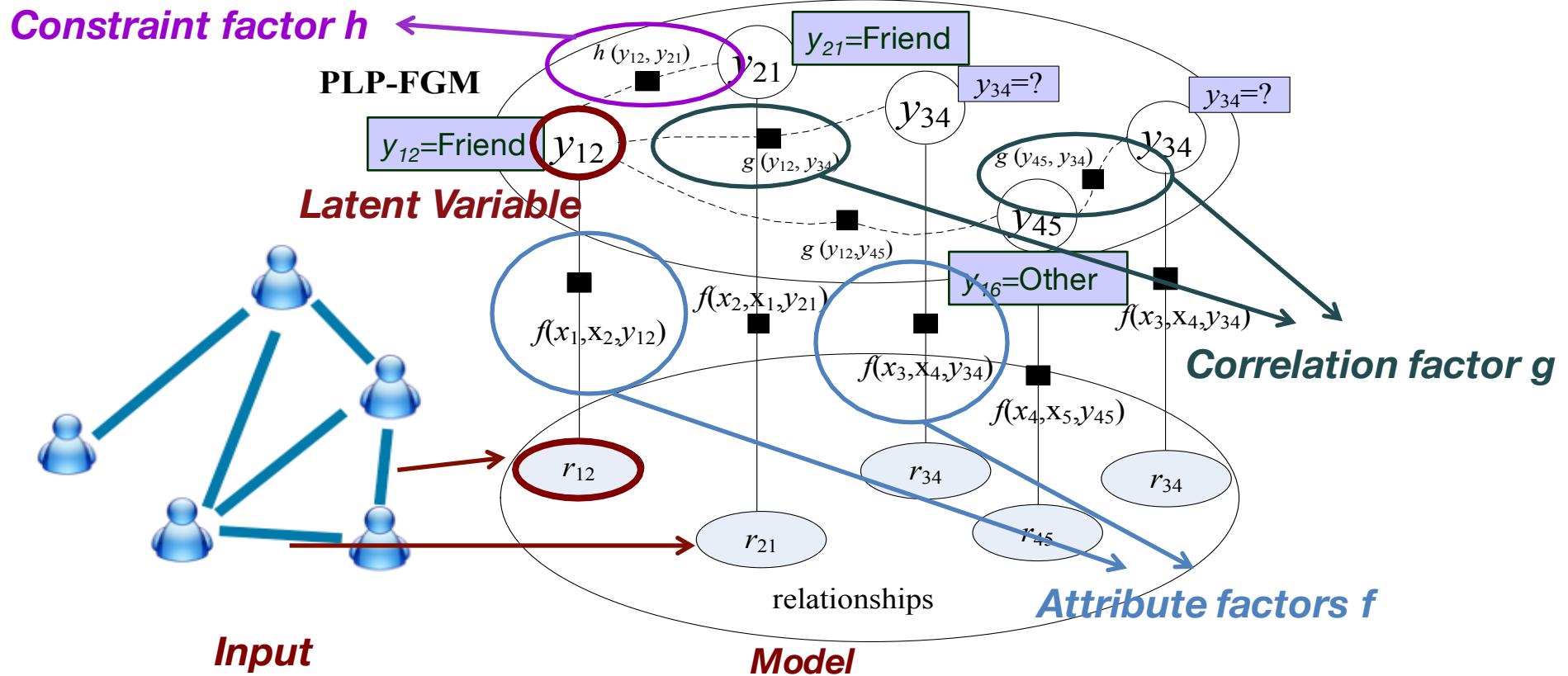
- How to mine semantics within networks?
 - Who are Jiawei Han's students and advisor?



Mining Network Semantics



Partially Labeled Pairwise Factor Graph Model (PLP-FGM)



Map relationship to nodes in model

Example:
#Coaut

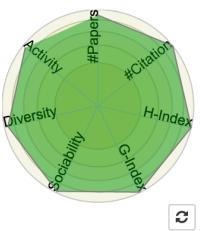
Example:

Author A has a longer publication history than both B and C.

AMiner



[Upload](#)



About

Papers 790

Lectures 13

Patents 1

Jiawei Han (韩家炜)

Department of Computer Science, University of Illinois at Urbana-Champaign

Professor

(217) 333-6903

hanj@cs.uiuc.edu

<http://www.cs.uiuc.edu/~hanj/>

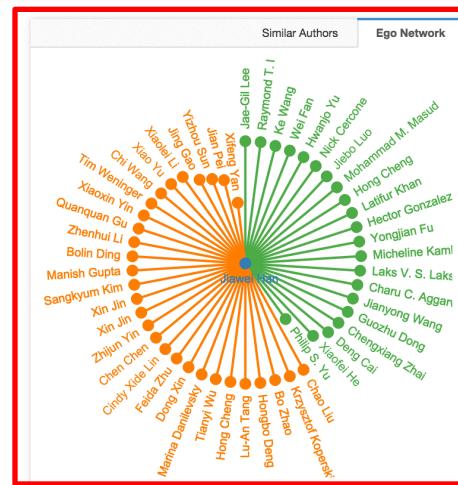
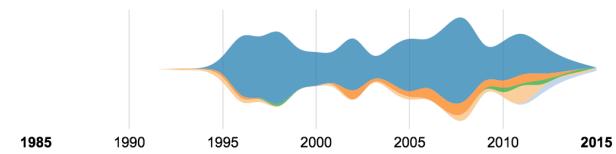
[External Links](#)

[Update](#)



Research Interests

● Data Mining ● Information Extraction ● Data Analysis
 ○ Machine Learning ● Text Mining



Ego Network

		Add Paper	Remove Paper	By Year	By Citation	
All (790)	Recent 20	2015 2014 2013 2012 2011 2010 2009 2008 2007 2006 2005 2004 2003 2002 2001 2000 1999 1998 1997 1996 1995				
790						
789						
788						
787						
Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions						
Wei Shen, Jianyong Wang, Jiawei Han						
Knowledge and Data Engineering, IEEE Transactions (2015)						
http://dx.doi.org/10.1109/TKDE.2014.2327028						
Mining Graph Patterns.						
Hong Cheng, Xifeng Yan, Jiawei Han						
Frequent Pattern Mining (2014)						
http://dx.doi.org/10.1007/978-1-4419-6045-0_12						
Troubleshooting interactive complexity bugs in wireless sensor networks using data mining techniques						
Mohammad Maifi Hasan Khan, Hieu Khao Le, Hossein Ahmadi, Tarek F. Abdelzaher, Jiawei Han						
ACM Transactions on Sensor Networks (TOSN) (2014)						
http://dx.doi.org/10.1145/2530290						

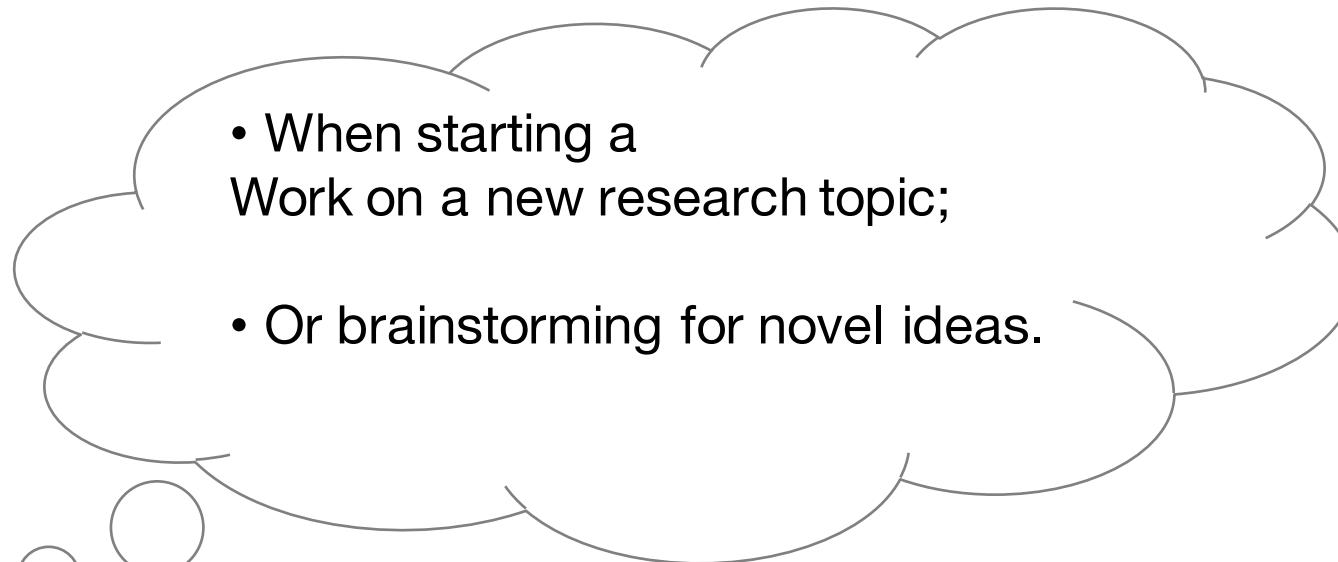


Why you will love AMiner?

—finding collaborations, applying
PhD, reviewing, recruiting...

Related publications: ACM TKDD (2), IEEE TKDE (3), KDD'08-15, WWW'12-15, J. Informetrics, SIGMOD'09, IJCAI'09-15, ICML'14

Examples – Expert search

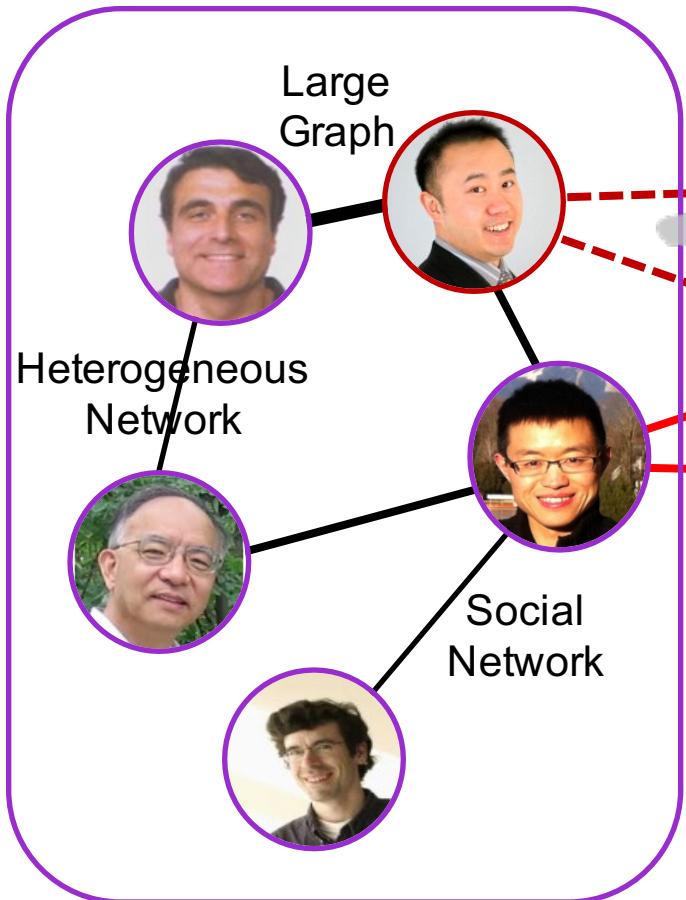


- 
- Who are experts in this field?
 - What are the top conferences in the field?
 - What are the best papers?
 - What are the top research labs?

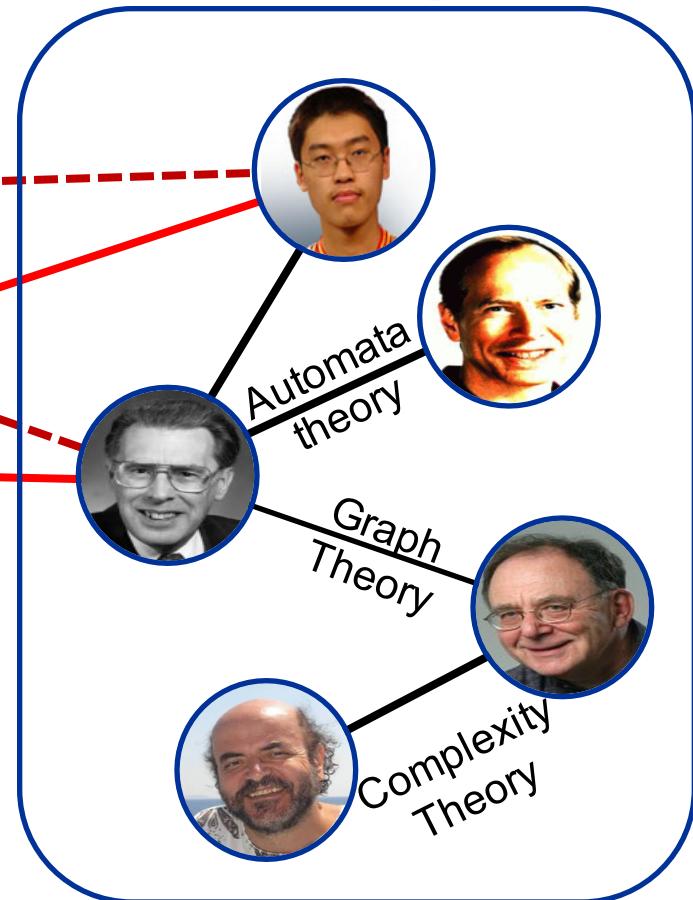
Examples – Collaboration Recommendation



Data Mining



Theory



Examples – More Challenging Questions...

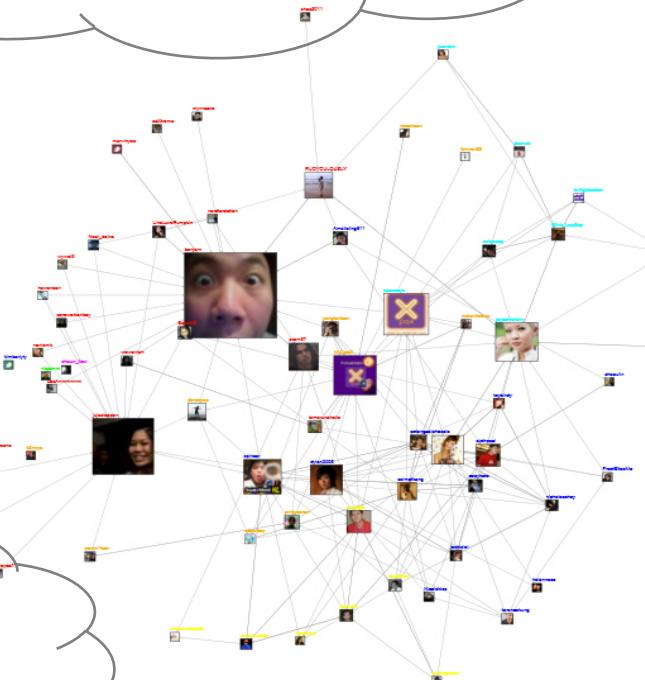
**Organization
or Agency**



- Looking for researchers from the U.S. who can speak Chinese...



- Want to recruit female researchers working on WSDM



Examples – Reviewer Suggestion

Conf. Committee
Journal Editors



Who are **best matching reviewers for each paper?**

Who will **accept the invitation to review the paper**



Paper content



Applications

Data Mining Experts

Global Data Mining Experts

Public

This collection includes experts in data mining field. The basic metric is that if a researcher served as (S)PC at KDD (the best conference in Data Mining) for more than once, she/he will be considered to be included in this collection.

Jie Tang ©2016-02-18

0 ~ 20 of 310 results

[Statistics](#)

[Extract emails](#)

[Export](#)

[Members](#)

Keywords

Name

Organization



[Find All Persons](#)



H-Index : >=60 (13) 50-59 (20) 40-49 (29) 30-39 (62) 20-29 (71) 10-19 (106) <10 (96)

Gender : Male (362) Female (35)

Language : Chinese (113) English (99) Indian (21) Greek (15) German (12) Japanese (8) French (8) Korean (4)

Location : USA (170) China (28) Hong Kong (14) Germany (11) Canada (11) Australia (10) Singapore (9) Belgium (7)

Relevance 

H-Index

Activity

Diversity

Rising Star

#Citation

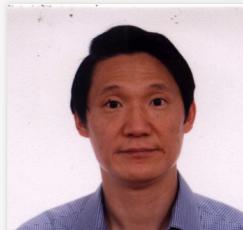
#Paper

AS



Jiawei Han (韩家炜)

Department of Computer Science, University of Illinois at Urbana-Champaign
 Professor



Philip S. Yu

Department of Computer Science, University of Illinois Chicago
 Professor and Wexler Chair in Information Technology



Christos Faloutsos

Dept. of Computer Science Carnegie Mellon University
 Professor



Vipin Kumar

University of Minnesota
 Professor



Rajeev Motwani

Stanford University
 Professor and Director of Graduate Studies

flags
 Male Chinese

flags
 Male Chinese

flags
 Male English

flags
 Male Indian

flags
 Male English



George Karypis



Jian Pei



Gerhard Weikum



Charu C. Aggarwal



Zhi-Hua Zhou (周志华)

<https://aminer.org/datamining-experts/>

NSFC (NSF of China) — proposal reviewer recommendation



Search Results for experts (2568 results)

李亚栋 (Yadong Li)
H-Index: 42 | #Paper: 289 | #Citation: 6237
清华大学
Nanostructures, Hydrothermal Synthesis, Nanocrystals, Crystal Structure, Semiconductors, Self-Assembly, Luminescence, Nanoparticles

魏飞 (Fei Wei)
H-Index: 40 | #Paper: 384 | #Citation: 5806
清华大学
Carbon Nanotubes, Chemical Vapor Deposition, Carbon Nanotube, Fluidized Bed, Propylene, Riser, A. Carbon Nanotubes, Sapo-34

薛其坤 (Qikun Xue)
H-Index: 33 | #Paper: 281 | #Citation: 5792
清华大学
Scanning Tunneling Microscopy, Molecular Beam Epitaxy, Topological Insulator, Scanning Tunneling Microscope, Silicon, Angle-Resolved Photoemission Spectroscopy, Electronic Structure, Superconductivity

王建勇 (Jianyong Wang)
H-Index: 27 | #Paper: 100 | #Citation: 5512
清华大学
Efficient Algorithm, Xml Document, Keyword Search, Efficient Mining, Keyword Query, Web Data, Xml Query, Extensive Experimental Study

李景虹 (Jinghong Li)
H-Index: 38 | #Paper: 166 | #Citation: 5139
清华大学
Graphene, Electrocatalysis, Biosensor, Direct Electrochemistry, Electrochemistry, Dna, Horseradish Peroxidase, Ionic Liquid

唐杰 (Jie Tang)
H-Index: 33 | #Paper: 185 | #Citation: 4596
清华大学
Social Network, Semantic Web, Social Influence, Heterogeneous Network, Ontology Mapping, Name Disambiguation, Topic Model, Unified Approach



Integration with ScholarOne

Find the extension—**reviewer recommender**—from Chrome Web Store

HTML PDF + Supplementary Files Abstract Cover Letter External Searches

Reviewer List

Order	Name	Status	History	Remove
1	Yangarber, Roman	Invited Response Select... Save	Invited: 07-Jul-2015 view full history	X
2	Toussaint, Yannick	Invited Response Select... Save	Invited: 07-Jul-2015 view full history	X
3	Nie, Zaiging Microsoft Research Asia, Web Search and Mining Group	Invited Response Select... Save	Invited: 07-Jul-2015 view full history	X
4	Jia, Yunde	Declined	Invited: 07-Jul-2015 Declined : 07-Jul-2015 view full history	X
5	Weninger, Tim	Invited Response Select... Save	Invited: 07-Jul-2015 view full history	X
6	Jung, Kyomin Computer Science	Declined	Invited: 07-Jul-2015 Declined : 07-Jul-2015 view full history	X

Alternates

Save X top

Author's Preferred / Non-Preferred Reviewers

Name, Keywords, Institution/Organization, Roles	Current / Past 12 Months	Days Since Last Review	Average R-Score	Add
0 people entered.				

Progress

# reviews required to make decision	3
# active selections	4 ✓
# invited	4 ✓
# agreed	0
# declined	2
# returned	0

Save

Reviewer Recommender

Keywords (Separated by ", "):

Text processing , Applications , Pattern Recognition , I Computing Methodologies , Evolutionary computing and genetic algorithms , Miscellaneous , Artificial Intelligence , I Computing Methodologies , Text mining ,

H-index: <10 10~20 20~30 >30

Location: All Language: All

Recommend

Result (20 of 207 experts)

I-Jeng Wang (Assistant Research Professor, Department of Computer Science) H-index: 11 Expertise: Probabilistic Models, Wireless Sensor Networks, Sensor Networks, Distributed Systems, Sensor Network E-mail: iwang@cs.jhu.edu

Steffi Beckhaus (Professor, University of Hamburg (W1, Juniorprofessor), head of im/ve group) H-index: 10 Expertise: Virtual Environment, Virtual Reality, Collision Avoidance, Virtual Environments, Individual Tutorial E-mail: steffi.beckhaus@uni-hamburg.de

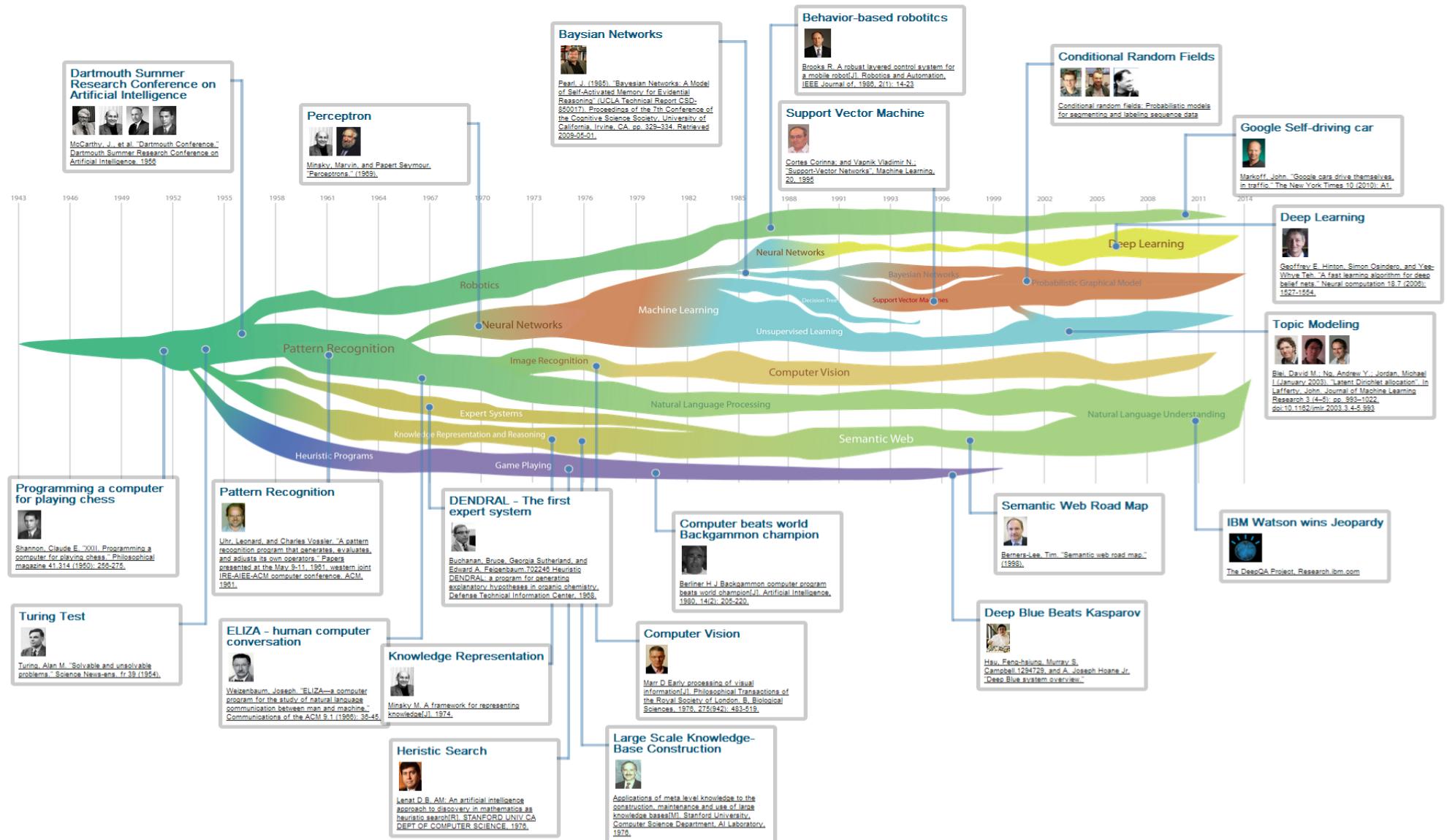
Dick De Ridder (Information & Communication Theory Group Delft University of Technology) H-index: 18 Expertise: Pattern Recognition, Neural Networks, Artificial Neural Networks, Neural Network, Protein Function Prediction E-mail: http://genlab.tudelft.nl/~dick/pic/email.gif

Powered by AMiner

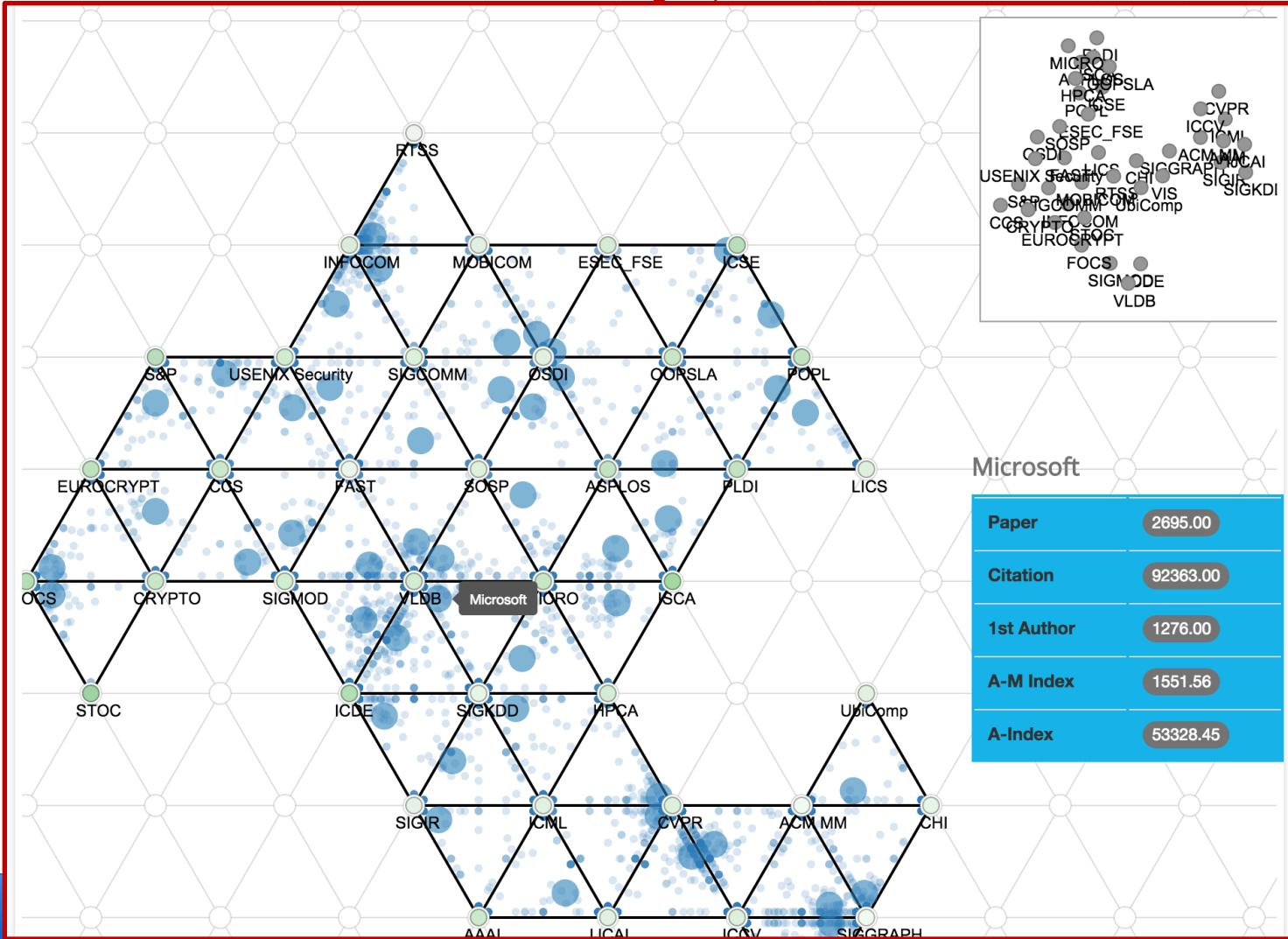
ScholarOne is a submission management system used by many journals such as IEEE Transactions and ACM Transactions.



Knowledge Trend for “AI”



Organization Ranking



Computer Science

Statistics

Visualization

All

High Performance Computing

Computer Network

Net and Information Security

All University Research institute Company

By different metrics

Rank	Organization	#Paper	#FirstAuthor	Credit	#Citations	Weighted_Credit	A-Index	AM-Index
1	Microsoft	2,695	1,276	1,451.82	92,363	52,978.29	53,328.45	1,551.56
2	Carnegie Mellon University	1,456	1,081	983.48	52,700	33,540.08	34,854.62	888.69
3	Massachusetts Institute of Technology	1,110	780	748.87	50,019	33,984.45	34,356.75	823.14



Microsoft

Paper	2695.00
Citation	92363.00
1st Author	1276.00
A-M Index	1551.56
A-Index	53328.45

577.01	50,746	30,910.64	32,000.09	769.55
882.87	49,432	28,605.32	28,968.54	724.06
589.39	42,246	25,936.45	26,413.88	661.09
723.26	37,566	26,671.27	27,062.57	646.47
494.02	32,806	21,883.74	22,296.53	597.85
499.99	22,065	15,803.23	16,064.56	431.86
484.63	23,233	15,669.74	15,877.43	400.24
363.71	21,406	14,344.10	14,576.56	397.31
421.99	18,035	13,690.39	13,879.77	395.41
475.18	20,106	12,130.08	12,426.60	385.67
374.13	22,283	14,180.29	14,588.01	380.17
484.95	19,742	13,084.09	13,232.71	369.72
342.06	19,561	13,801.29	14,134.56	361.08
432.66	18,332	12,104.18	12,315.72	343.56
406.87	15,740	11,095.00	11,279.12	318.02
231.29	18,441	11,764.28	11,821.74	300.50
234.81	27,424	13,648.39	13,945.38	288.10
270.80	16,841	12,821.13	12,796.51	284.73
503.02	13,288	8,929.52	8,977.09	269.57
242.03	17,496	10,056.54	10,345.91	269.57
257.77	16,979	9,058.36	9,071.13	258.45
245.24	14,406	8,208.11	8,169.25	254.69
349.30	11,493	5,888.79	6,129.02	243.98
330.87	17,799	9,487.88	9,628.10	243.48
290.43	12,593	8,032.33	8,238.18	230.03



Widely used..

- The largest publisher:
Elsevier
- Conferences
 - KDD 2010-15
 - WSDM 2011
 - ICDM 2011
 - ICDM 2012
 - SocInfo 2011
 - ICMLA 2011
 - WAIM 2011
 - etc.

The screenshot shows the SciVerse Hub search interface. At the top, there are links for Hub, ScienceDirect, Scopus, SciTopics, and Applications. A message indicates that ArnetMiner is logged in. The search bar contains the query "data mining". Below the search bar, it says "About 54317 results for: ALL(data mining)". On the left, there are filters for "Search within results", "Refine Results" (with options for Limit to and Exclude), "Content Sources" (listing ScienceDirect, Scopus, Patent Offices, Digital Archives, and MD Consult with counts 30766, 3487, 699, 652, and 517 respectively), "Year" (with years from 2010 to 2006), and "Keyword" (with categories like data mining, clustering, classification, neural networks, and algorithms). The main search results area lists various items under these categories.

The image compares the websites for the Fourth ACM International Conference on Web Search and Data Mining (WSDM 2011) and the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-2011).

WSDM 2011: The header features a red background with the WSDM logo and text. It includes links for Home, About, Committee, Authors, Attendees, Program, and Sponsors. The date "9-12 February 2011 Hong Kong" is prominently displayed. A "QUICK LINKS" sidebar lists instructions for authors, registration, accommodations, travel grants information, tutorial information, and workshop information. An "IMPORTANT DATES" sidebar lists deadlines: 1/15 Camera-ready deadline (New), 1/11 Google travel grant deadline, 1/22 Author registration deadline, 2/15 Early-bird registration deadline, 3/15 Google travel grant notification, 2/20 Student travel grant deadline, 3/9 Group rate hotel reservation deadline, and 3/10 Student travel grant notification.

KDD-2011: The header features a blue background with the KDD-2011 logo and text. It includes links for Home, About, Program, Workshops, Tutorials, and KDD Cup. A "QUICK LINKS" sidebar lists instructions for authors, registration, accommodations, travel grants information, tutorial information, and workshop information. An "IMPORTANT DATES" sidebar lists deadlines: 1/15 Camera-ready deadline (New), 1/11 Google travel grant deadline, 1/22 Author registration deadline, 2/15 Early-bird registration deadline, 3/15 Google travel grant notification, 2/20 Student travel grant deadline, 3/9 Group rate hotel reservation deadline, and 3/10 Student travel grant notification.

Organizing Committee: Both sites feature sections for the Organizing Committee, General Chairs, Program Chairs, and Industry and Government Track Chairs. Each section includes a list of names, their institutions, and small profile pictures. The KDD-2011 site also includes sections for Tutorial Chairs, KDD-Cup Chairs, Panel Chairs, Poster Chairs, Best Paper Chairs, Doctoral Dissertation Award Chair, Innovation and Service Award Chair, Test-of-Time Paper Award Chair, Local Arrangements Chairs, Student Travel Award Chairs, Exhibit and Demo Chairs, Media and Publicity Chairs, Publications Chairs, Social Network Chairs, Sponsorship Chair, Registration Treasurer, and Video.



Academic Social Platform

Registration

 Email address First name Last name Gender Position I want to receive news and special offers I agree with the [Terms and Conditions](#)[Register](#)

Jie Tang

News Feed



Recommended publication from Yitai Qian

Synthesis and characterization of hexagonal NiTe₂ nanoplates

Journal of Nanoscience and Nanotechnology (2009)

Single-crystalline NiTe₂ nanoplates have been synthesized on a large scale through a hydrothermal reaction of Ni(CH₃COO)₂ and TeO₂ at 210 °C for 48 h. The NiTe₂ nanoplates have hexagonal shape. The average edge sizes and thicknesses of nanoplates are 350 nm and 85 nm, respec...

[hexagonal](#) [nanoplates](#) [nite 2](#) [single-crystalline](#)

≈ 1 citations

Like

Comment



Recommended publication from Jiawei Han

Mining Thick Skylines over Large Databases

People recently are interested in a new operator, called skyline [3], which returns the objects that are not dominated by any other objects with regard to certain measures in a multi-dimensional space. Recent work on the skyline operator [3,15,8,13,2] focuses on efficient comp...

≈ 51 citations

Like

Comment



Recommended publication from Xichen Zhao

Considerations to the Development of Shenzhen University Town Library

Library Tribune (2005)

介绍了深圳大学城图书馆的概况和特点，阐述了深圳大学城图书馆的办馆

Edit Profile

Following

Followers

LIBRARY

My Bookmarks

My Publications

EXPERT EXPLORER

My Expert Bases

Flags

Red

Yellow

Cyan

Green

Blue

ADMIN

Profile Merge

Data Annotation

Jie Tang

News Feed

Edit Profile

Following

Followers

LIBRARY

My Bookmarks

My Publications

EXPERT EXPLORER

My Expert Bases

Flags

Red

Yellow

Cyan

Green

Blue

ADMIN

Profile Merge

Data Annotation

**Yitai Qian (钱逸泰)****H-Index:** 66 | **#Paper:** 1233 | **#Citation:** 21315

University of Science and Technology of China

✓ Following

**Xichen Zhao (赵洗尘)****H-Index:** 3 | **#Paper:** 4 | **#Citation:** 30

Library of Shenzhen University Town, Shenzhen, China

✓ Following

**Bingkun Zhou (周炳琨)****H-Index:** 15 | **#Paper:** 286 | **#Citation:** 1083

清华大学电子工程系

✓ Following

**Jiawei Han (韩家炜)****H-Index:** 133 | **#Paper:** 818 | **#Citation:** 111048

Department of Computer Science, University of Illinois at Urbana-Champaign

✓ Following

**Yehoshua Sagiv****H-Index:** 50 | **#Paper:** 169 | **#Citation:** 11640

Department of Computer Science at the Technion.

✓ Following

Jie Tang

News Feed

Edit Profile

Following

Followers

LIBRARY

My Bookmarks

My Publications

EXPERT EXPLORER

My Expert Bases

Flags

Red

Yellow

Cyan

Green

Blue

ADMIN

Profile Merge

Data Annotation

Financial Dependence and Growth

RAGHURAM G. RAJAN LUIGI ZINGALES

(1998)

Cited by 6591 Bibtex **Introduction to Data Mining**

Florin Gorunescu

Cited by 3492 Bibtex http://dx.doi.org/10.1007/978-3-642-19721-5_1 **Data mining: concepts and techniques**

Jiawei Han Micheline Kamber

Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial (2000)

Cited by 28914 Bibtex <http://dx.doi.org/10.4114/ia.v10i29.873>**Theory of the firm: Managerial behavior, agency costs and ownership structure**

Michael C. Jensen William H. Meckling

Journal of Financial Economics (1976)

Cited by 55844 Bibtex [http://dx.doi.org/10.1016/0304-405X\(76\)90026-X](http://dx.doi.org/10.1016/0304-405X(76)90026-X)**Applied Cryptography: Protocols, Algorithms, and Source Code in C**

Bruce Schneier

Applied Cryptography: Protocols, Algorithms, and Source Code in C (1993)

Cited by 12693 Bibtex [http://dx.doi.org/10.1016/0167-4048\(94\)90072-8](http://dx.doi.org/10.1016/0167-4048(94)90072-8)**Tags**

- All
- data mining
- hello world



Open Big Scholar Data

Citation Network

Citation Network Dataset

Citation Network Dataset

[Overview](#)[Description](#)[References](#)

Overview

Overview

The data set is designed for research purpose only. The citation data is extracted from DBLP, ACM, and other sources. The first version contains 629,814 papers and 632,752 citations. Each paper is associated with abstract, authors, year, venue, and title.

The data set can be used for clustering with network and side information, studying influence in the citation network, finding the most influential papers, topic modeling analysis, etc.

A larger version will be released soon.

[Citation-network V1](#): 629,814 papers and >632,752 citation relationships (2010-05-15).

[Citation-network V2](#): 1,397,240 papers and >3,021,489 citation relationships (2010-09-13).

[DBLP-Citation-network V3](#): 1,632,442 papers and >2,327,450 citation relationships (2010-10-22).

[DBLP-Citation-network V4](#): 1,511,035 papers and 2,084,019 citation relationships (2011-01-08).

[DBLP-Citation-network V5](#): [download from mirror site] 1,572,277 papers and 2,084,019 citation relationships (2011-01-08).

[DBLP-Citation-network V6](#): 2,084,055 papers and 2,244,018 citation relationships (2013-09-29).

[DBLP-Citation-network V7](#): [download from mirror site] 2,244,021 papers and 4,354,534 citation relationships (2014-05-25).

Data set	#paper	#Citation Relationship	Comment
Citation-network V1	629,814	>632,752	
Citation-network V2	1,397,240	>3,021,489	
DBLP-Citation-network V3	1,632,442	>2,327,450	
DBLP-Citation-network V4	1,511,035	2,084,019	Arnetminer [2011-01-08]
DBLP-Citation-network V5 [download from mirror site]	1,572,277	2,084,019	Arnetminer [2011-02-21]
DBLP-Citation-network V6	2,084,055	2,244,018	Arnetminer [2013-09-29]
DBLP-Citation-network V7 [download from mirror site]	2,244,021	4,354,534	Arnetminer [2014-05-25]

Description

Data Description

<https://aminer.org/citation>



Collaboration Network with Rich Semantics

Extraction and Mining of Academic Social Networks

Overview
Data Description
References

Overview

This dataset is designed for research purpose only.

The content of this data includes paper information, paper citation, author information and author collaboration. **2,092,356** papers and **8,024,869** citations between them are saved in the file [AMiner-Paper.rar](#); **1,712,433** authors are saved in the file [AMiner-Author.zip](#) and **4,258,615** collaboration relationships are saved in the file [AMiner-Coauthor.zip](#).

FileName	Node	Number	Size
AMiner-Paper.rar [download from mirror site]	Paper Citation	2,092,356 8,024,869	509 MB
AMiner-Author.zip [download from mirror site]	Author	1,712,433	167 MB
AMiner-Coauthor.zip [download from mirror site]	Collaboration	4,258,615	31.5 MB



AMiner (ArnetMiner)

– Toward understanding big scholar data

<http://aminer.org>

Semantics

+

Social

+

Open Data

Science of Science!



AMiner (ArnetMiner)

— Toward understanding big scholar data

<http://aminer.org>

Thanks to our partners





AMiner (ArnetMiner)

— Toward understanding big scholar data

<http://aminer.org>



Special thanks to [Yuxiao Dong](#) from ND for helping cook the slides!



Why AMiner.org

Google Scholar search results for "data mining":

- Advances in knowledge discovery and data mining** (UM Fayyad, G Piatesky-Shapiro, P Smyth... - 1996 - citeulike.org)
- Data mining techniques: for marketing, sales, and customer support** (MJ Berry, G Linoff - 1997 - dl.acm.org)
- Data Mining Techniques** (thoroughly acquaints you with the new generation of data mining tools and techniques and shows how to use them to make better business decisions. One of the first practical guides to mining business data, it describes techniques for ... Cited by 2384 Related articles All 2 versions Cite Save More)
- The WEKA data mining software: an update** (M Hall, E Frank, B Pfahringer... - ACM SIGKDD ..., 2009 - dl.acm.org)
- From data mining to knowledge discovery in databases** (U Fayyad, G Piatesky-Shapiro, P Smyth - AI magazine, 1996 - aaai.org)

Microsoft Academic search results for "data mining":

- Data Mining - DM** (Publications: 37,408 | Citation Count: 269,610)
- Definition Context (5)**
- Data Mining(DM) is the process of extracting implicit, valuable, and interesting information from large sets of data. As huge amounts of data have been stored in traffic and transportation databases, data warehouses, geographic information systems, and other information repositories, data mining is receiving substantial interest from both academia and industry...**

Semantic Scholar search results for "data mining":

- Data Mining: Concepts and Techniques** (Jiawei Han, Micheline Kamber - MK - 2000)
 - Citation Velocity: 172
 - View PDF Cite Save
- The WEKA data mining software** (Mark A. Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer)
 - Citation Velocity: 383
 - More than twelve years have elapsed since the first public release of WEKA. In that time, the software has been rewritten entirely from scratch, evolved substantially and now accompanies a text on data mining [35]. These days, WEKA enjoys widespread acceptance in both academic and industrial settings. It is downloaded more than 1.4... (More)
 - View PDF Cite Save
- Introduction to Data Mining** (Pang-Ning Tan, Michael Steinbach, Vipin Kumar - AW - 2005)
 - Citation Velocity: 87
 - Cite Save

“Academic search was treated as document search, but ignored semantics”