

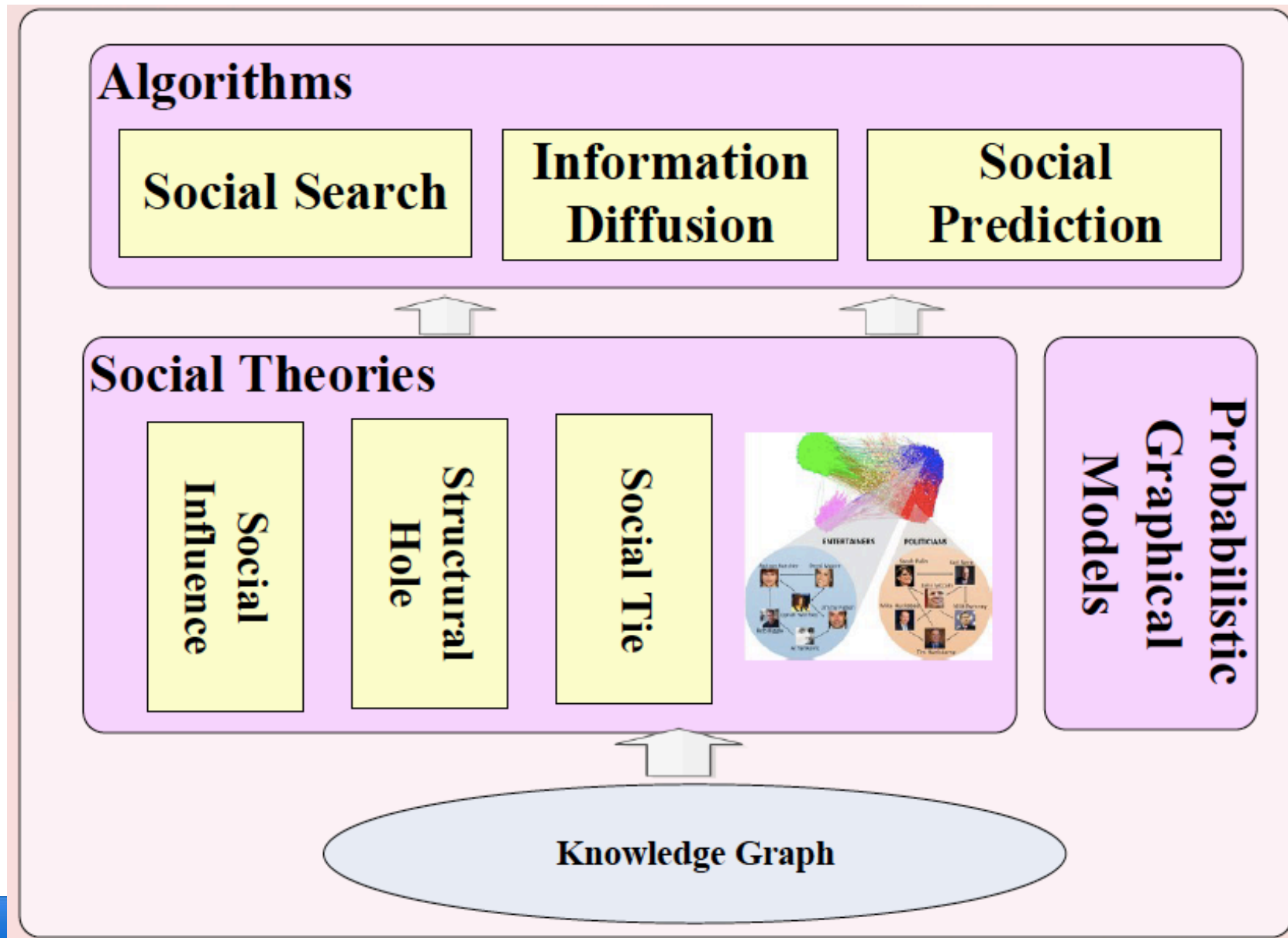


Computational Models for Micro-level Social Network Analysis

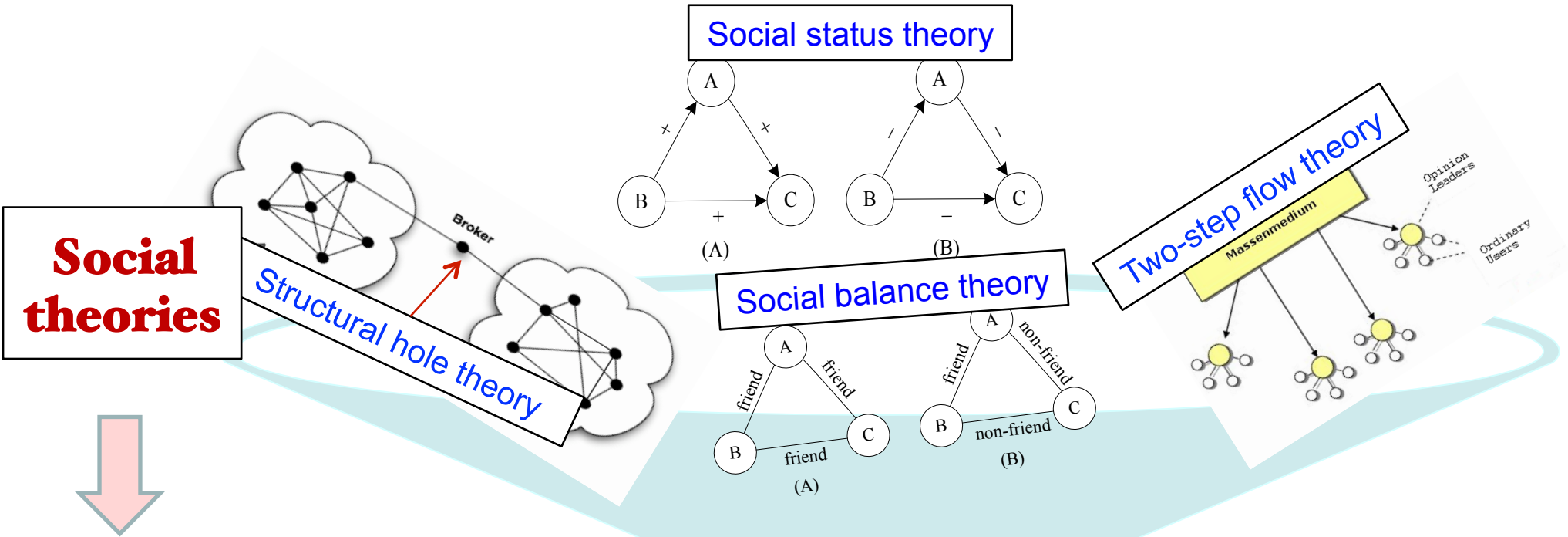
Jie Tang

Tsinghua University, China

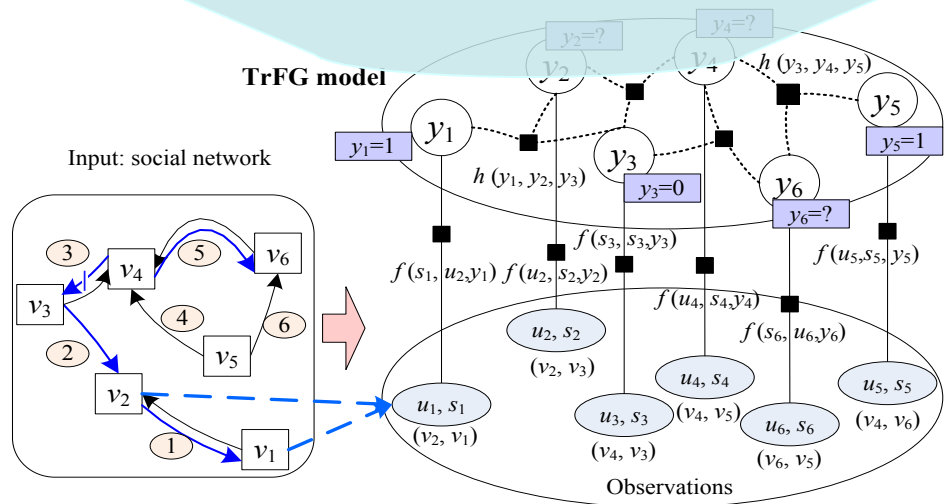
What we do: a big picture



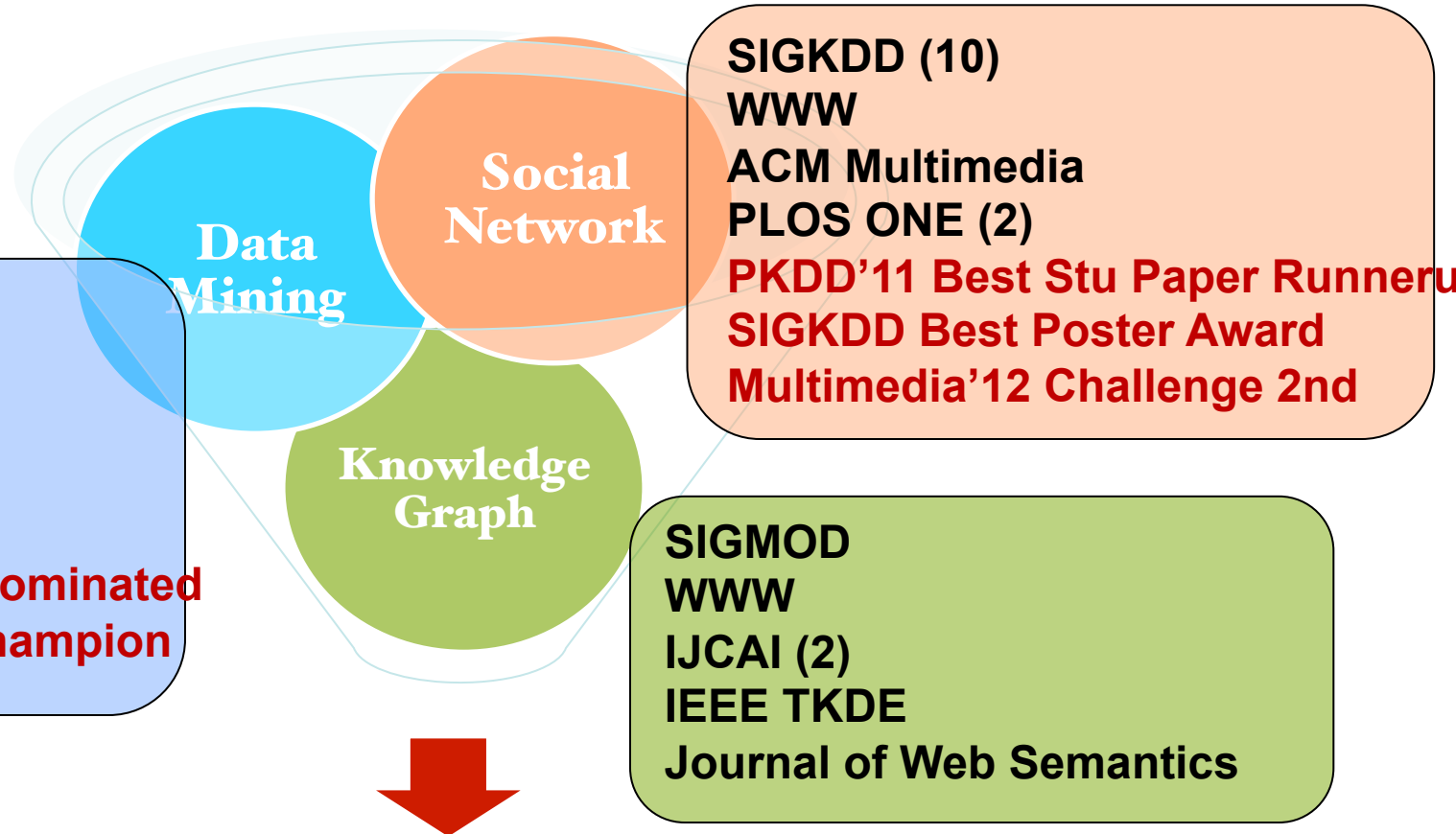
Computational Models for Micro-level Social Network Analysis



Graphical models

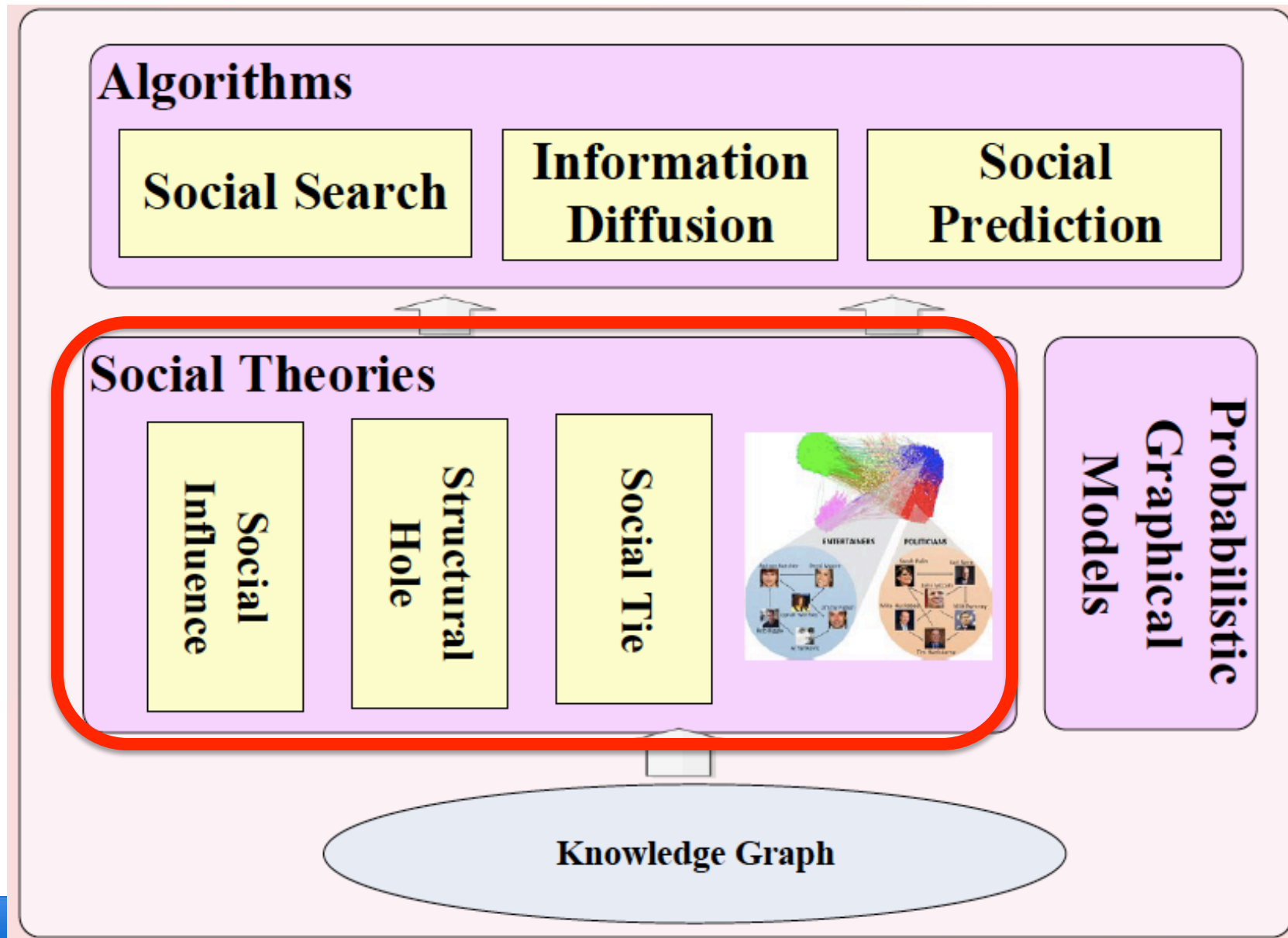


Related Publications



Published more than 100 papers on major conferences and journals

What we do: a big picture

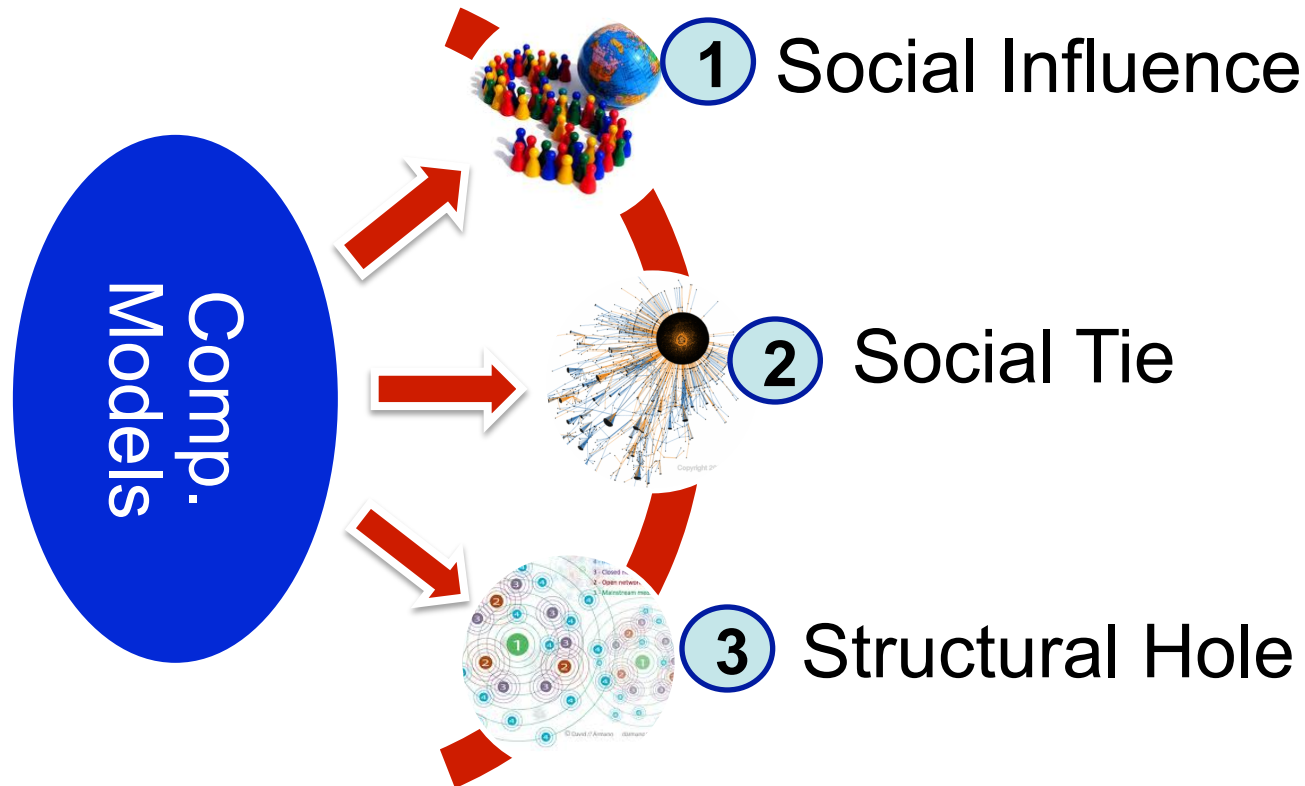


20+ Datasets

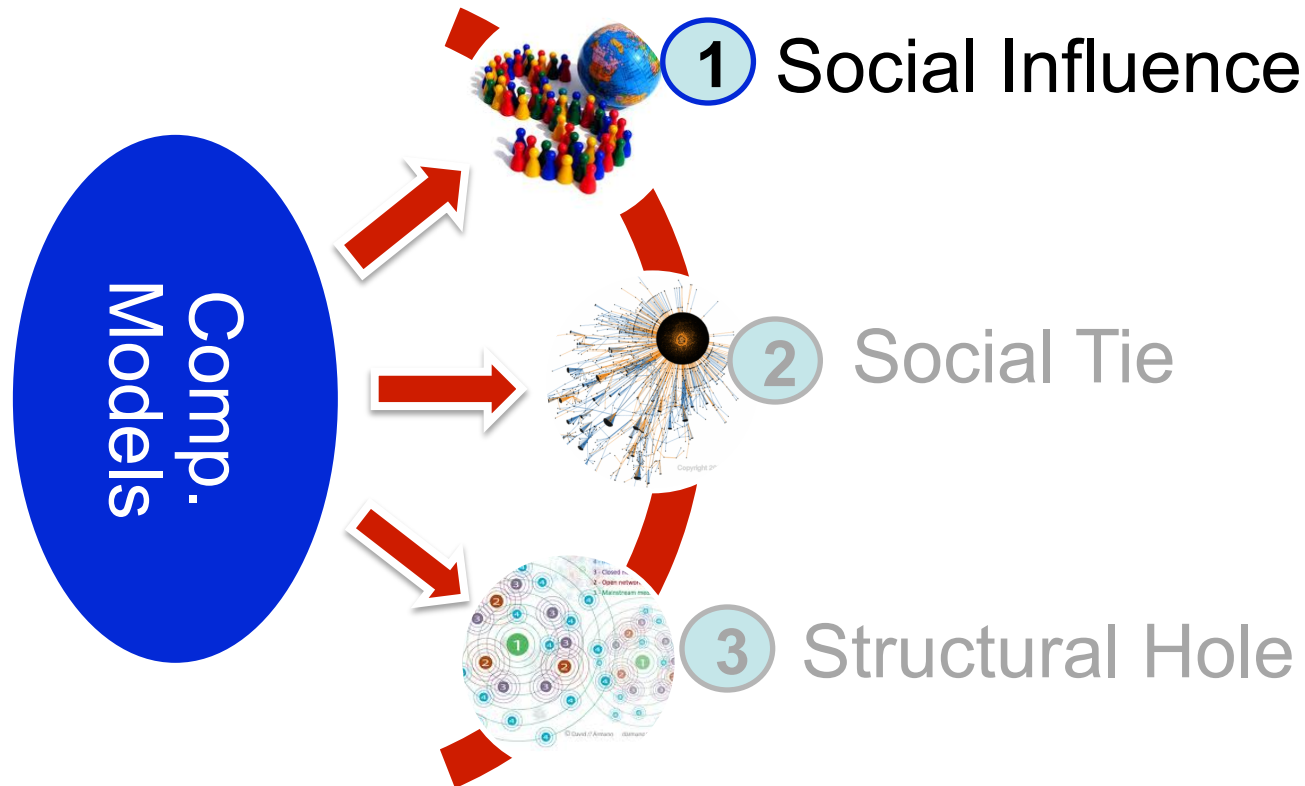
Network	#Nodes	#Edges	Behavior
Twitter-net	111,000	450,000	Follow
Weibo-Retweet	1,700,000	400,000,000	Retweet
Slashdot	93,133	964,562	Friend/Foe
Mobile (THU)	229	29,136	Happy/Unhappy
Gowalla	196,591	950,327	Check-in
ArnetMiner	1,300,000	23,003,231	Publish on a topic
Flickr	1,991,509	208,118,719	Join a group
PatentMiner	4,000,000	32,000,000	Patent on a topic
Citation	1,572,277	2,084,019	Cite a paper
Twitter-content	7,521	304,275	Tweet "Haiti Earthquake"

Most of the data sets are publicly available for research.

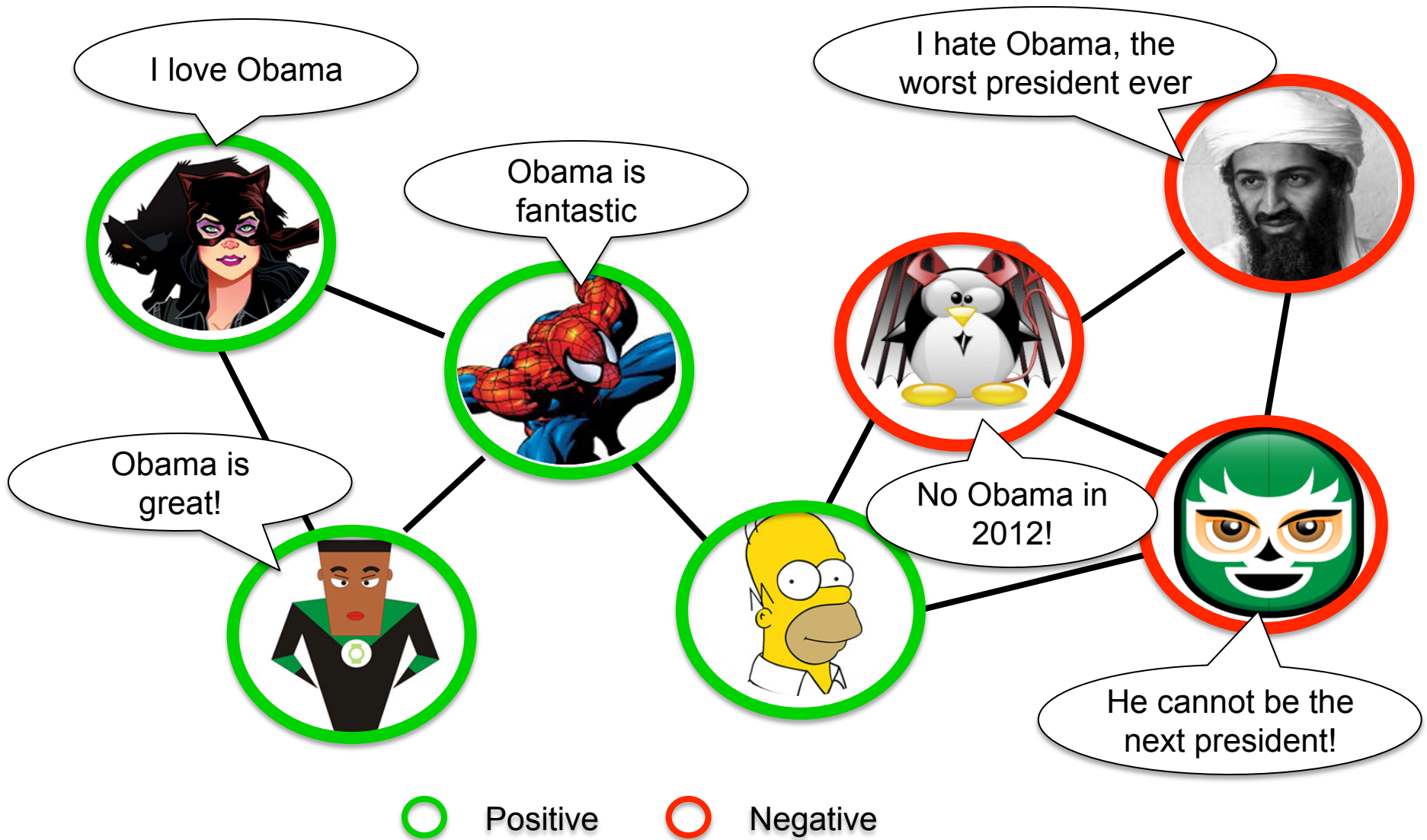
Micro-level Social Network Analysis



Micro-level Social Network Analysis



“Love Obama”

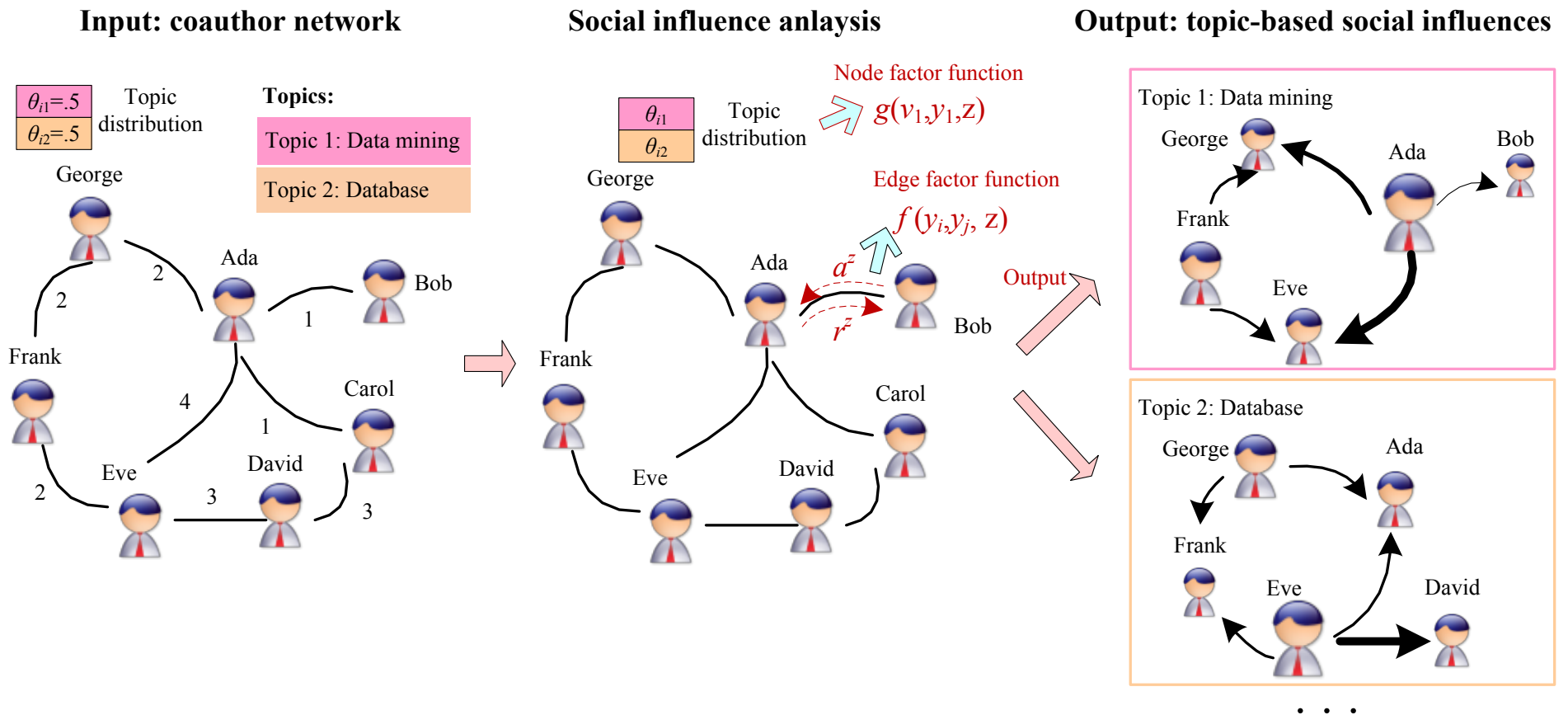


Social Influence Analysis

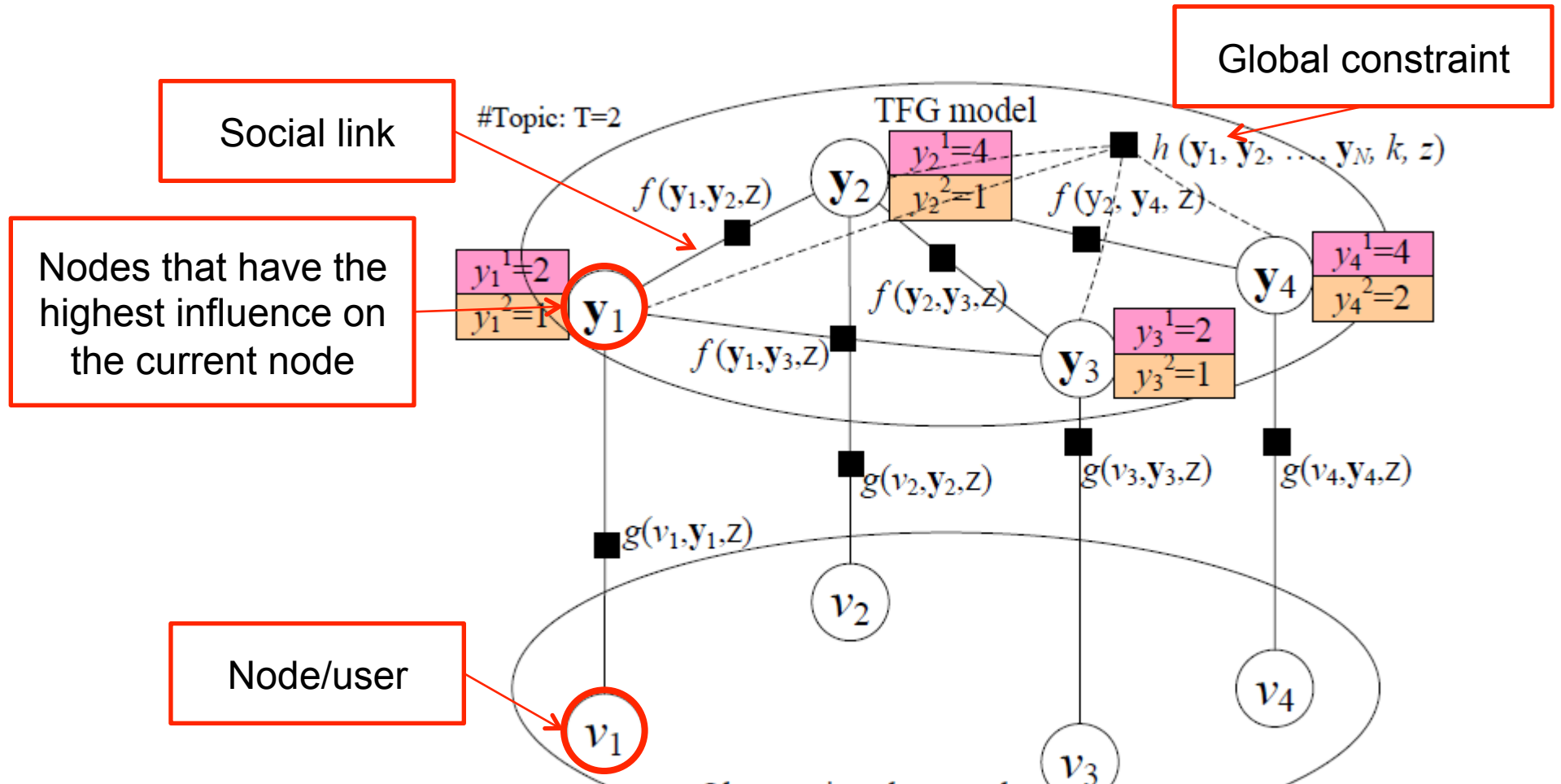
- Influence analysis
 - Topic-based influence measure [Tang-Sun-Wang-Yang 2009]
 - Learning influence distribution [Liu-Tang-Han-Yang 2010&2012]
 - Conformity influence [Tang-Wu-Sun 2013]
- Social influence and behavior prediction
 - Social action tracking [Tan-Tang-Sun-Lin-Wang 2010]
 - User-level sentiment in social networks [Tan-et-al 2011]
 - Emotion prediction in mobile network [Tang-et-al 2012, *spotlight paper*]
 - Inferring affects from images [Jia-et-al 2012, *Grand Challenge 2nd Prize*]

Topic-based Social Influence Analysis

- Social network -> Topical influence network



Topic-based Social Influence Analysis



The problem is cast as identifying which node has the highest probability to influence another node on a specific topic along with the edge.

Topical Factor Graph (TFG)

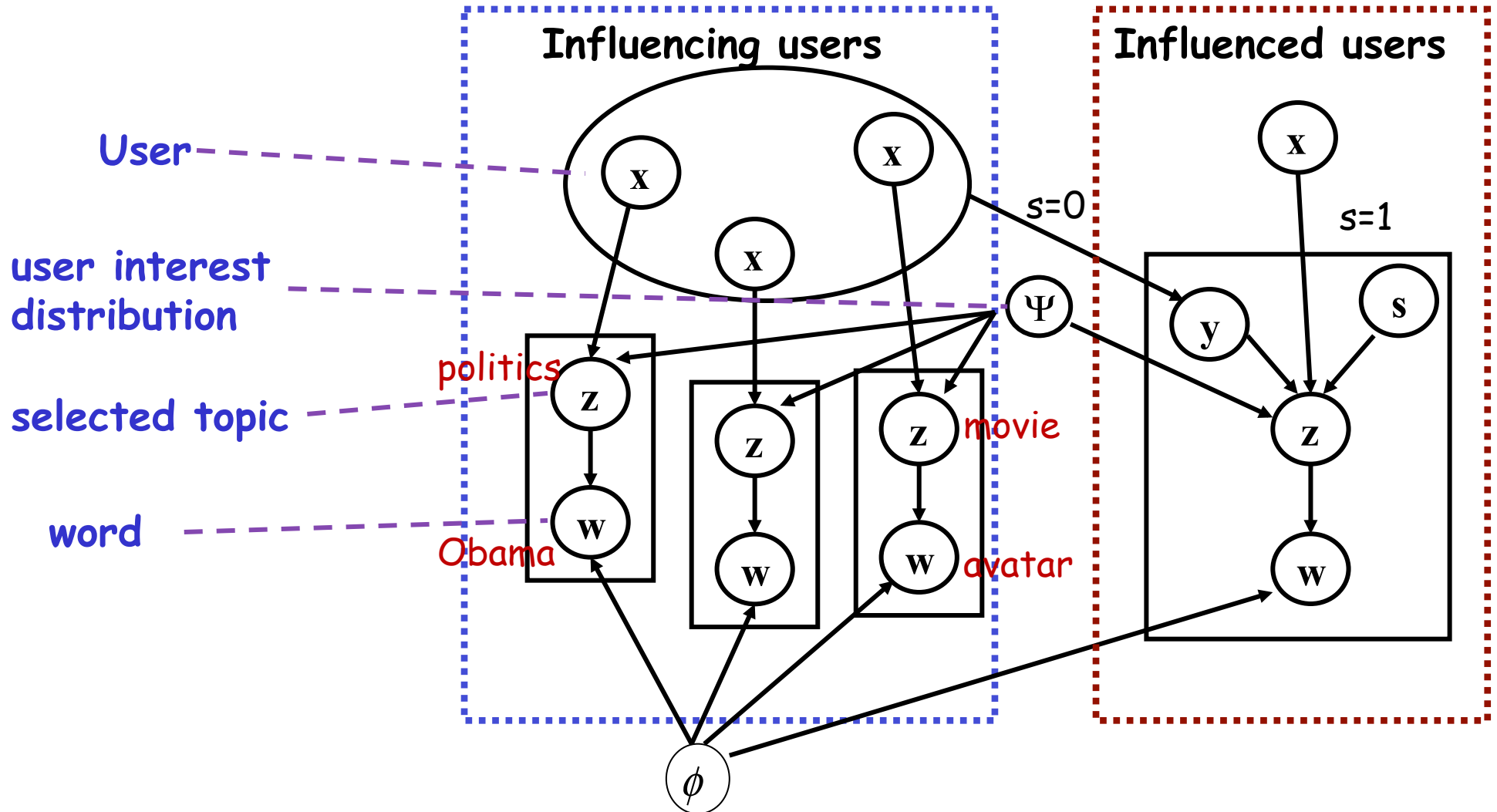
Objective function:

$$P(\mathbf{v}, \mathbf{Y}) = \frac{1}{Z} \prod_{k=1}^N \prod_{z=1}^T h(\mathbf{y}_1, \dots, \mathbf{y}_N, k, z) \prod_{i=1}^N \prod_{z=1}^T g(v_i, \mathbf{y}_i, z) \prod_{e_{kl} \in E} \prod_{z=1}^T f(\mathbf{y}_k, \mathbf{y}_l, z)$$




1. Feature Functions

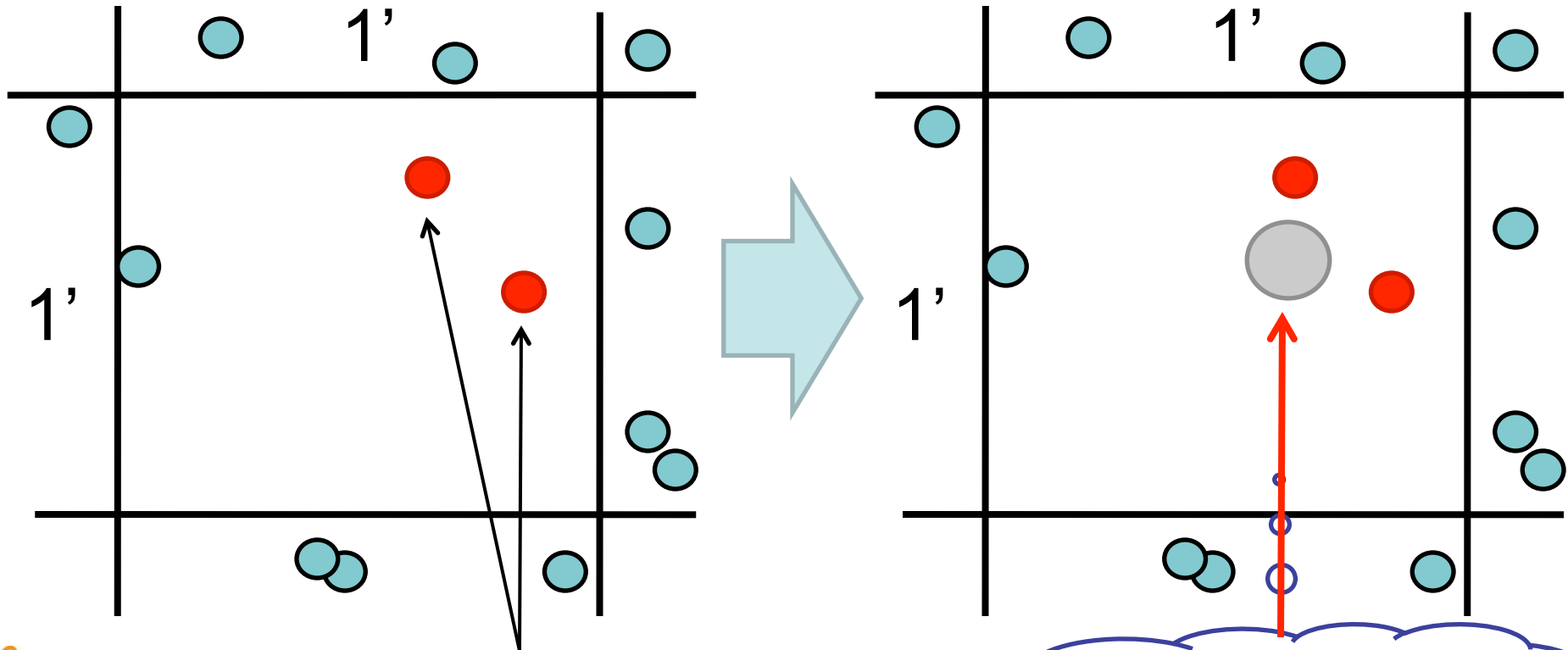
- The learning task is to find a configuration for all $\{\mathbf{y}_i\}$ to maximize the joint probability.

Probabilistic Generative Model



Confluence: Conformity Influence

Legend  Alice  Alice's friend  Other users

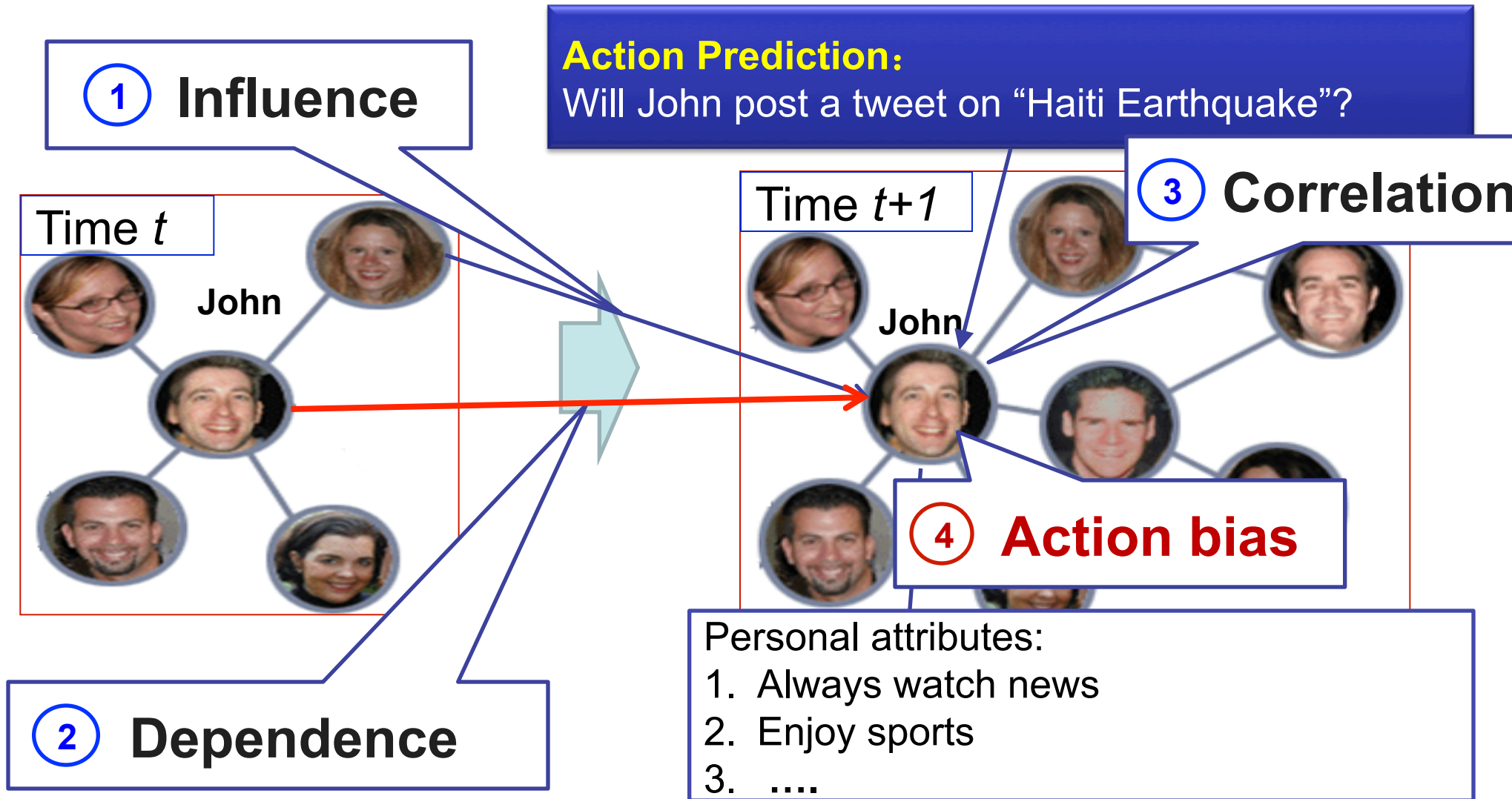


If Alice's friends check in this location at time t

Will Alice also check in nearby?

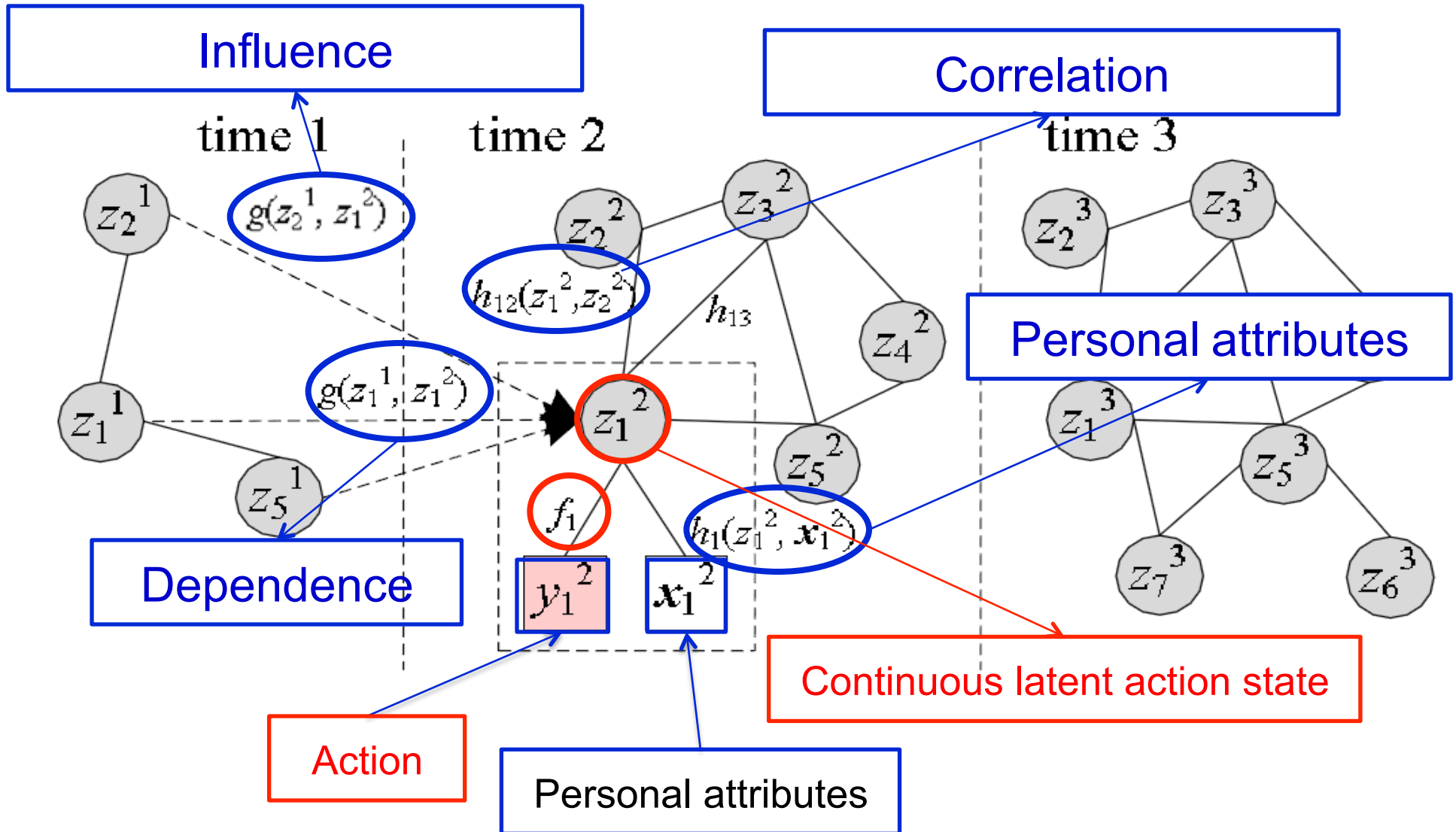


Social Influence & Action Modeling^[1]

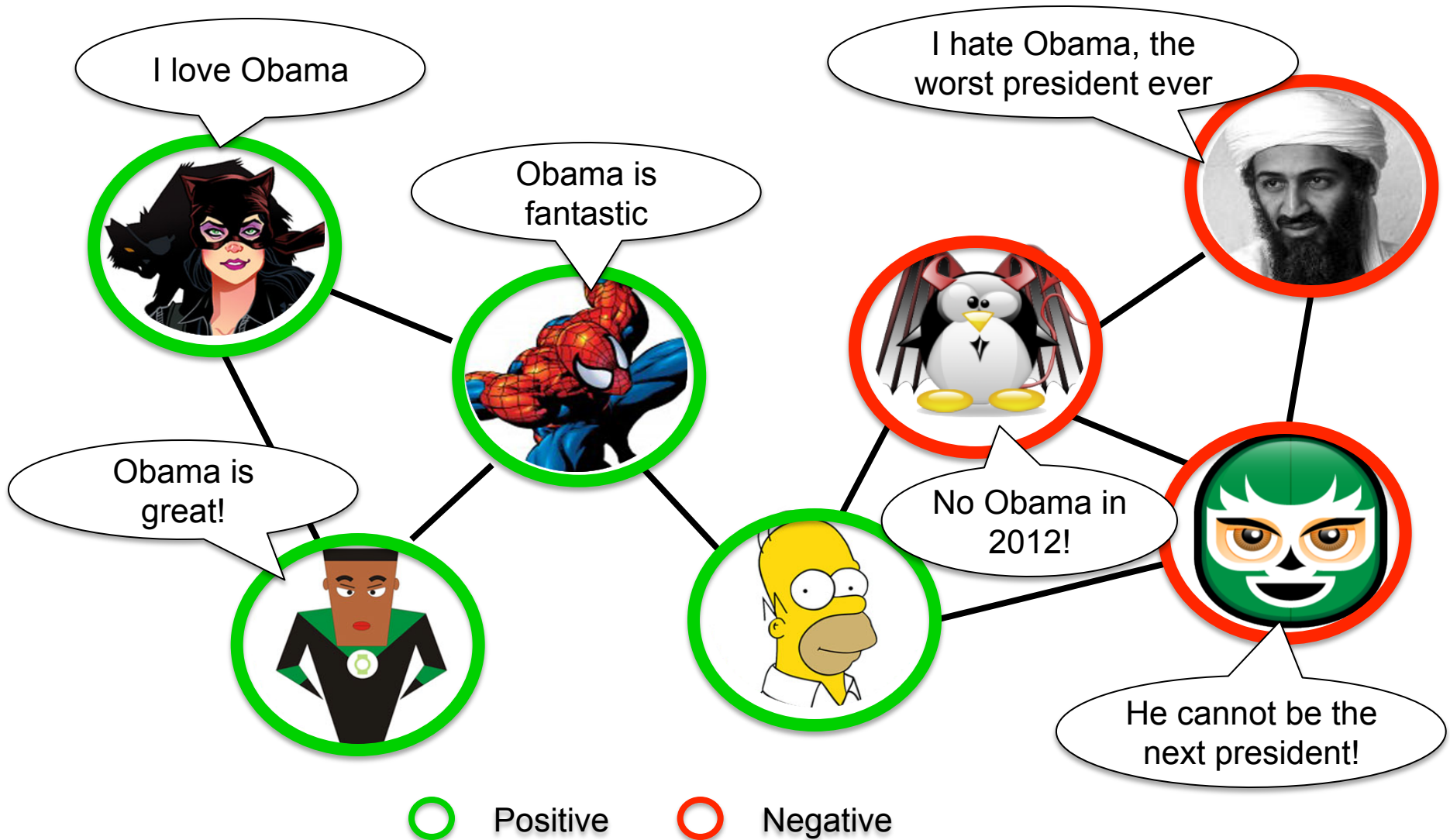


[1] C. Tan, J. Tang, J. Sun, Q. Lin, and F. Wang. Social action tracking via noise tolerant time-varying factor graphs. In KDD'10, pages 807–816.

A Discriminative Model: NTT-FGM



User-level Sentiment Analysis



Happy System—case study

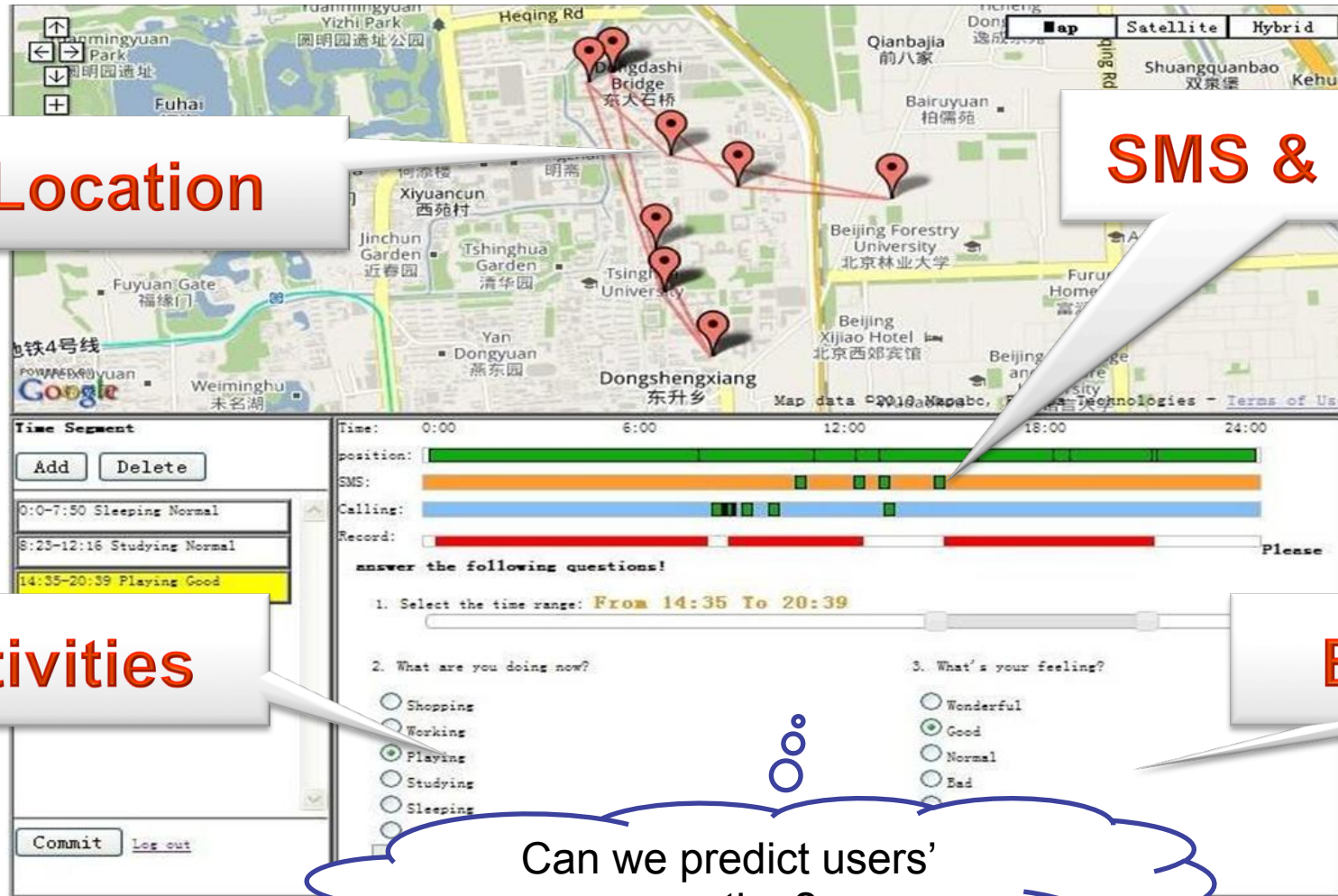
Location

SMS & Calling

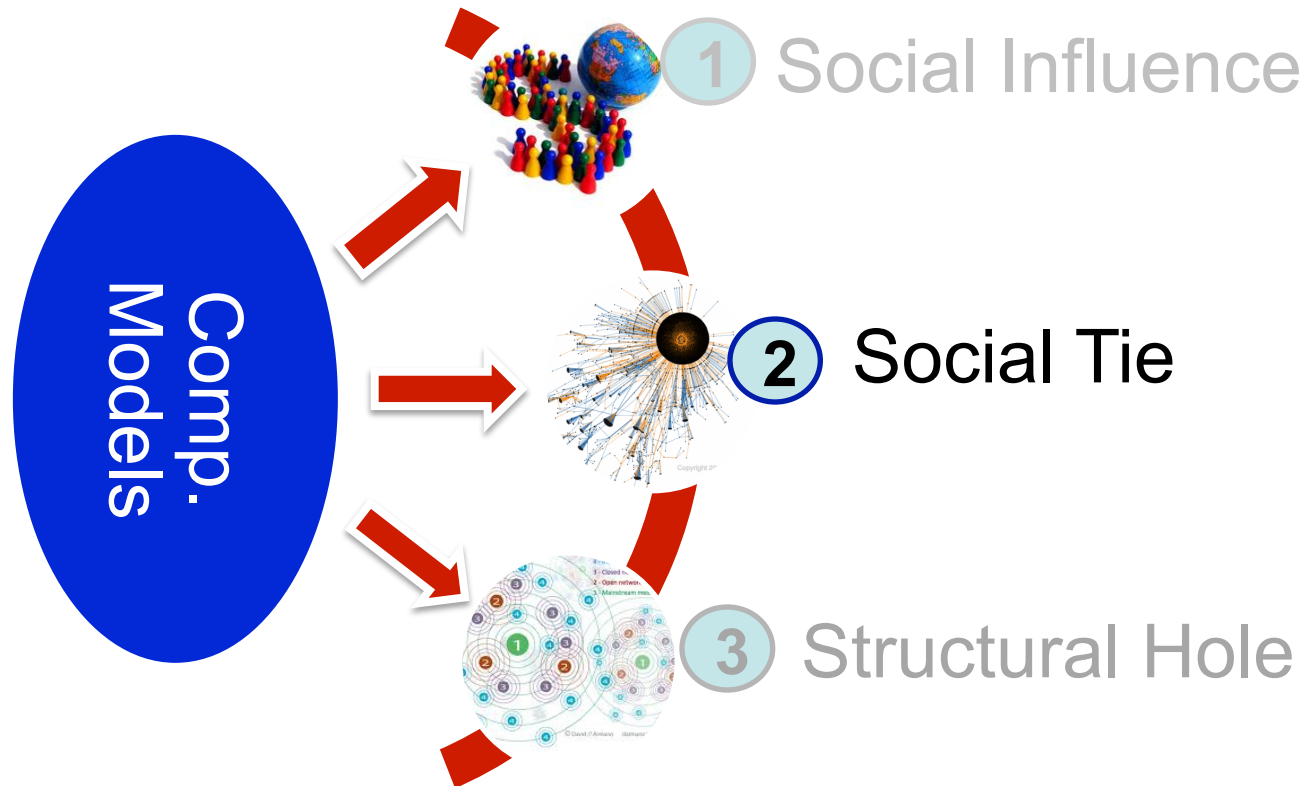
Activities

Emotion

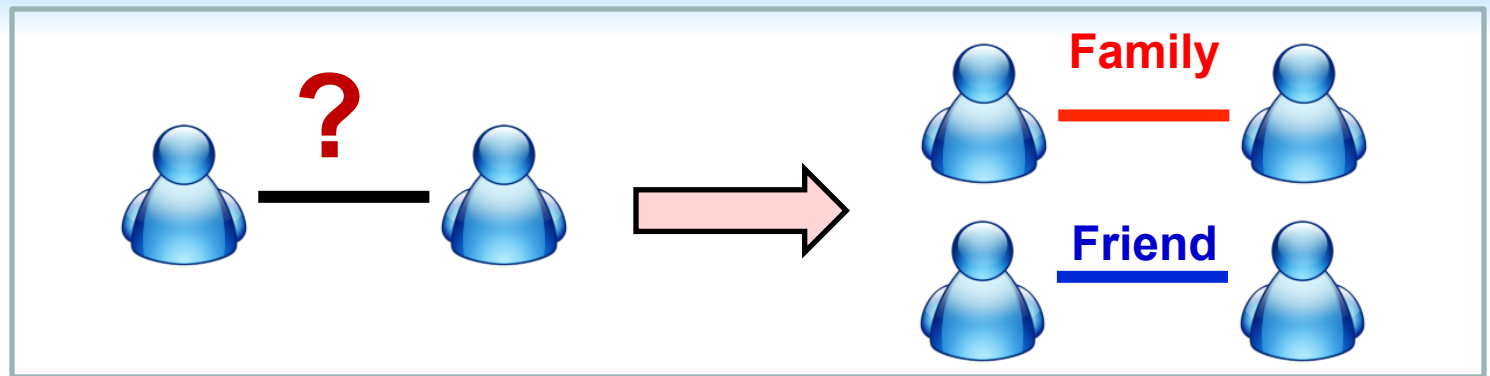
Can we predict users' emotion?



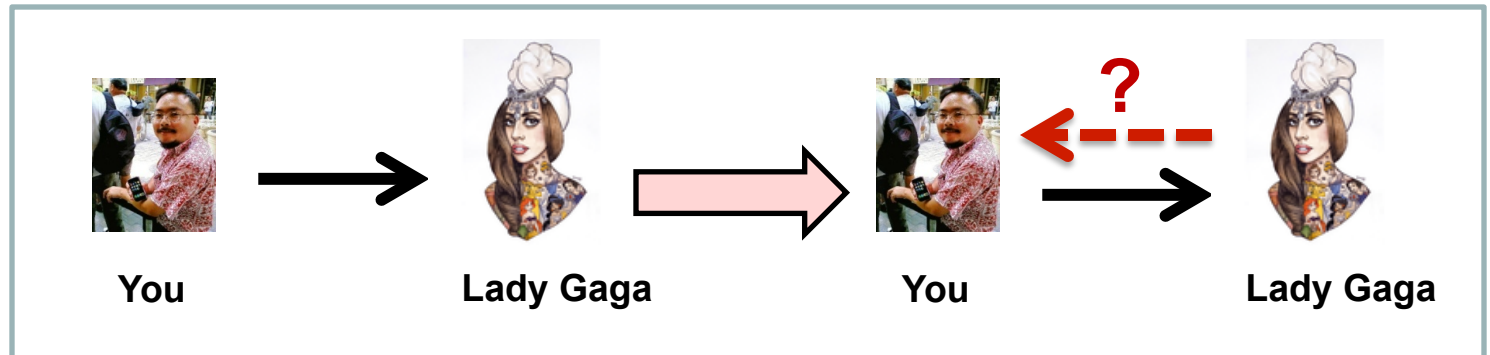
Micro-level Social Network Analysis



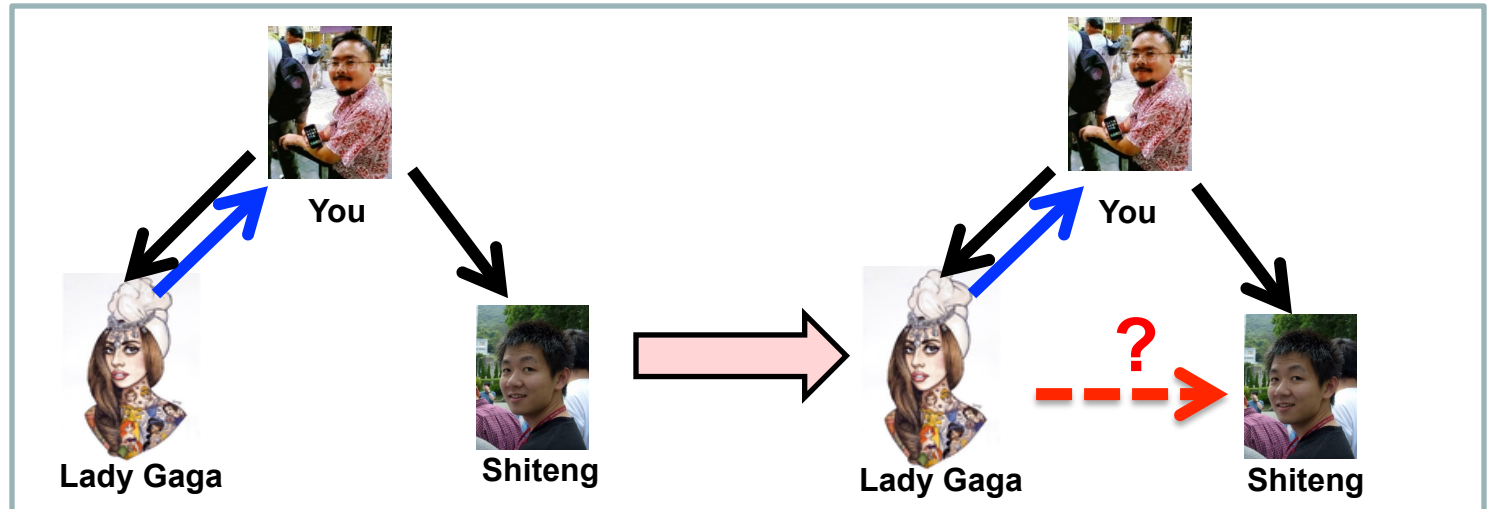
Inferring social ties



Reciprocity



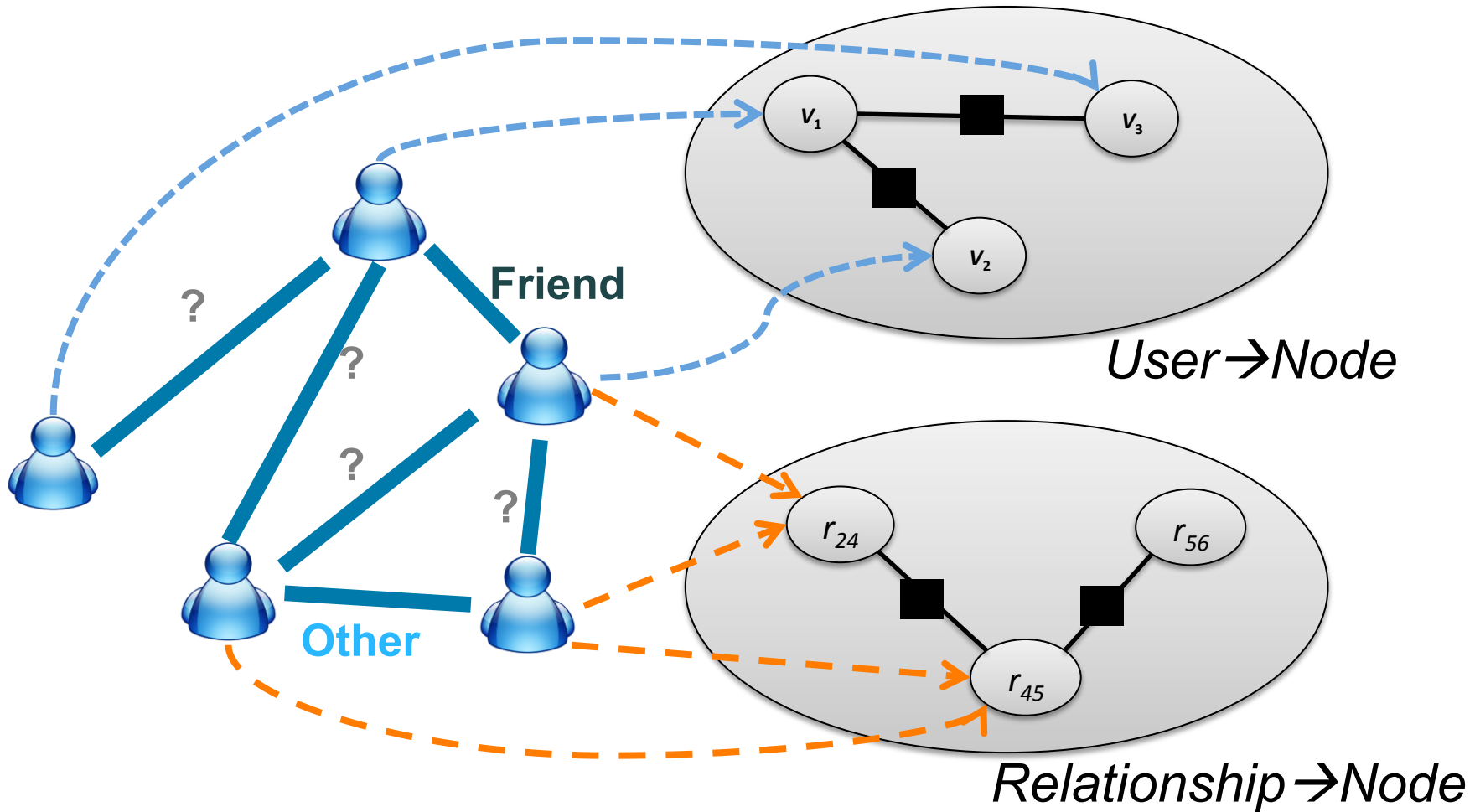
Triadic Closure



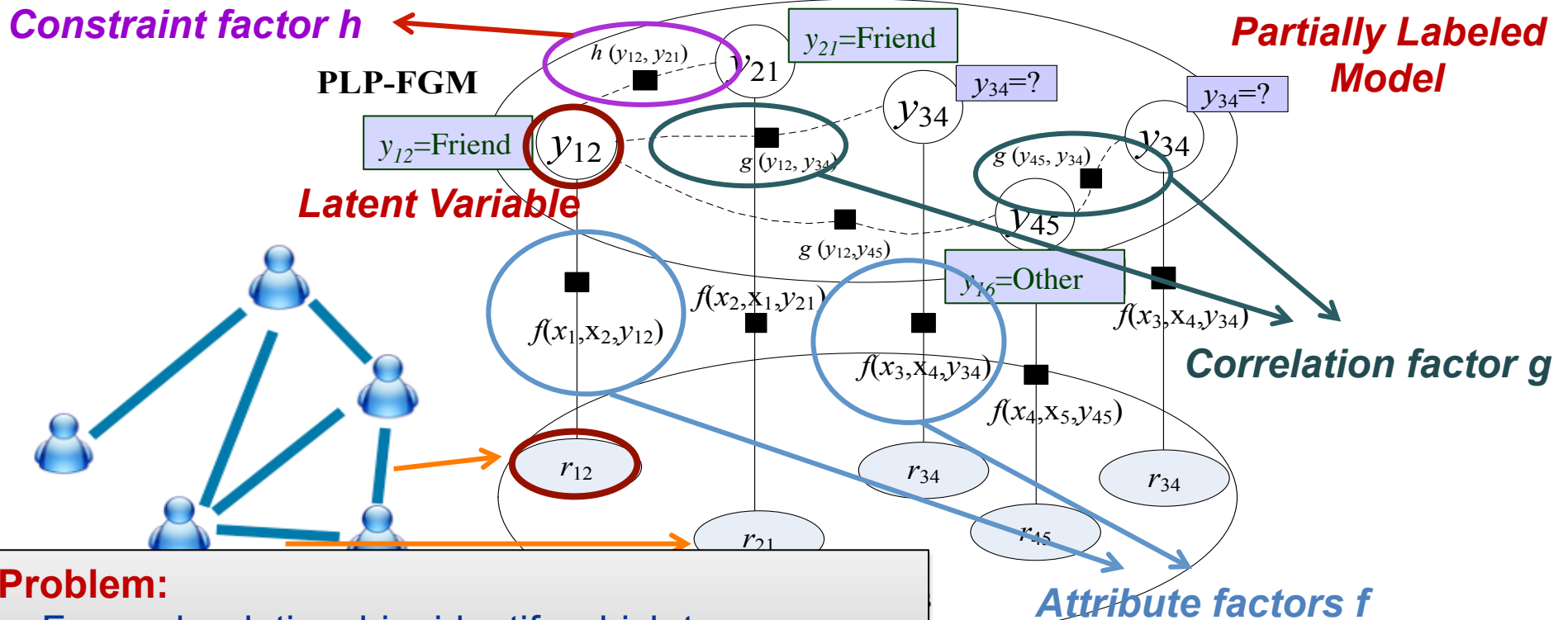
Social Tie Analysis

- Social relationship
 - Mining advisor-advisee relationships [Wang-Han-et-al 2010]
 - Learning to infer social tie [Tang-Zhuang-Tang 2011, *Best Runnerup*]
 - Reciprocity prediction [Hopcroft-Lou-Tang 2011]
 - Inferring social tie across networks [Tang-Lou-Kleinberg 2012]
- Triadic closure
 - Inferring triadic closure [Lou-Tang-Hopcroft-Fang-Ding 2013]
- Community
 - Kernel community [Wang-Lou-Tang-Hopcroft 2011]
 - Community co-evolution [Sun-Tang-Han-Chen-Gupta 2013]

Basic Idea



Partially Labeled Pairwise Factor Graph Model (PLP-FGM)



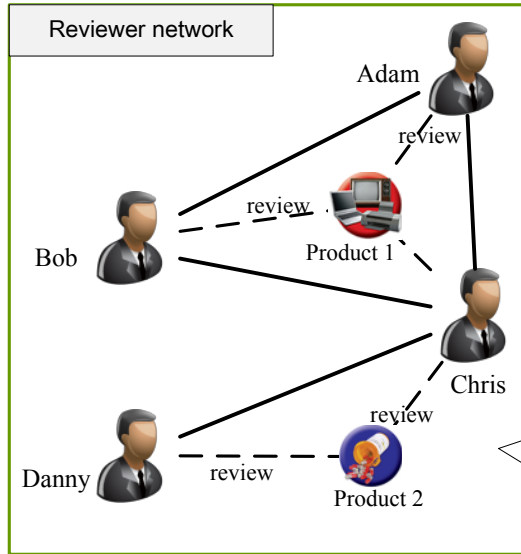
Problem:
For each relationship, identify which type has the highest probability?

Example:
A makes call to B immediately after the call to C.

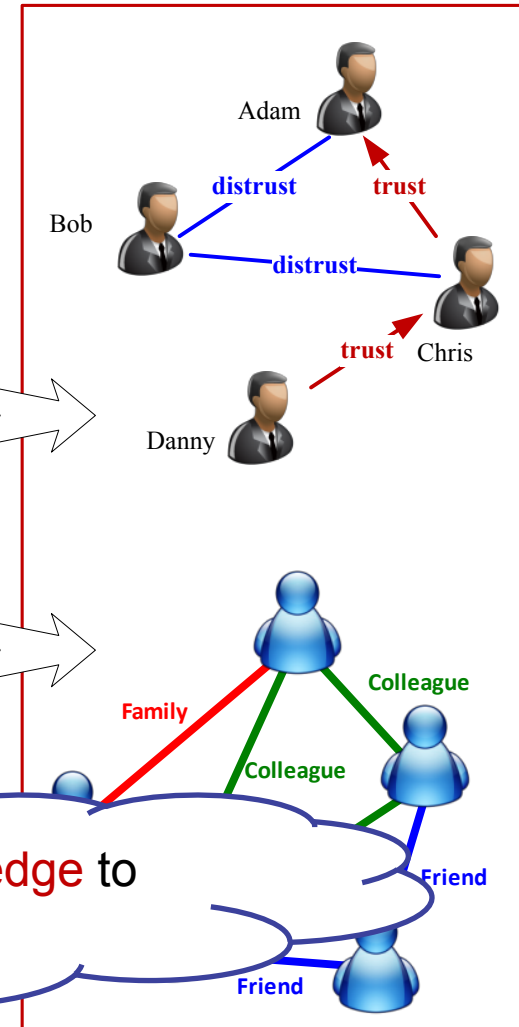
Inferring Social Ties Across Networks

Epinions

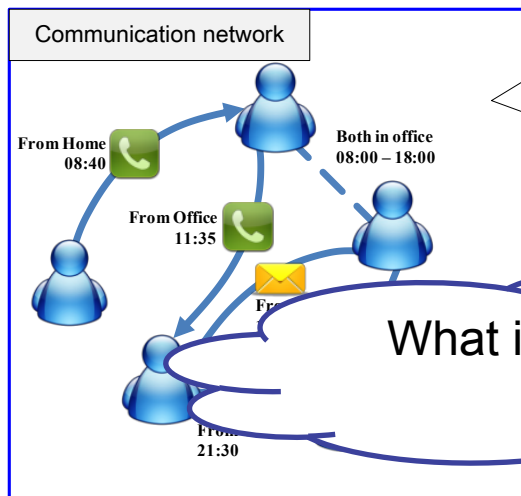
Input: Heterogeneous Networks



Output: Inferred social ties in different networks



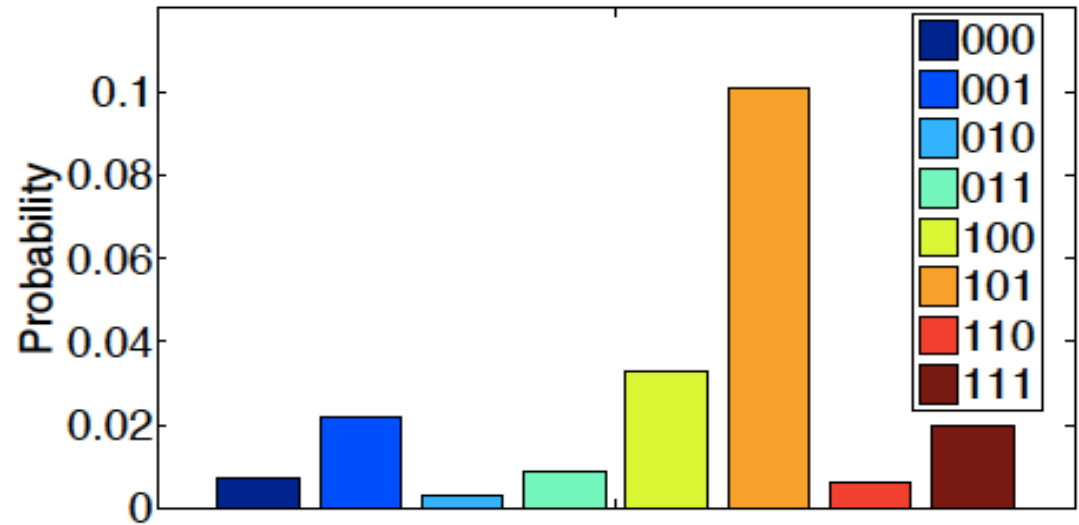
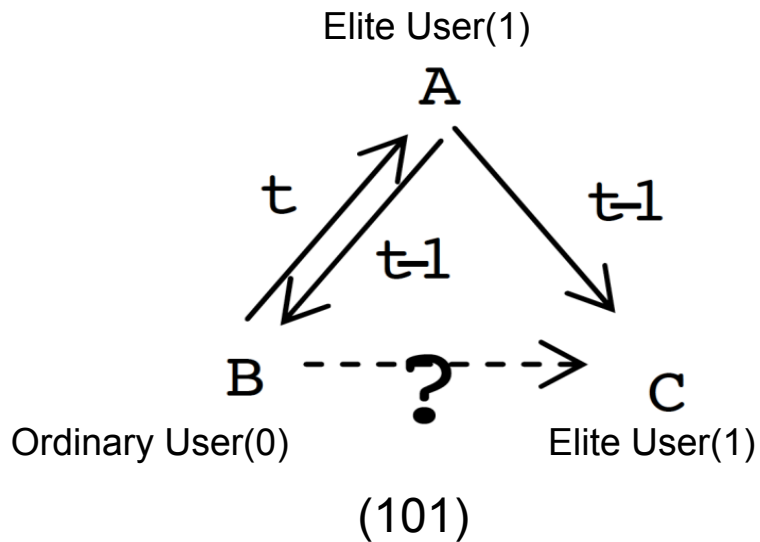
Mobile



Knowledge Transfer for Inferring Social Ties

What is the **knowledge** to transfer?

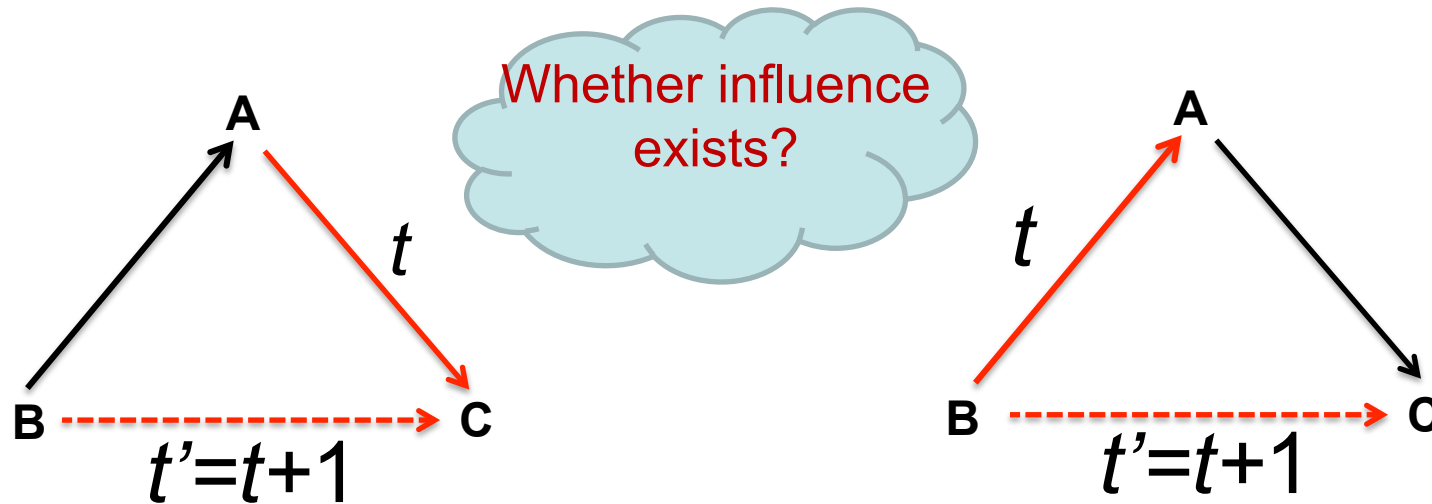
Triad Closure



- $P(1XX) > P(0XX)$. Elites users play a more important role to form the triadic closure. The average probability of $1XX$ is three times higher than that of $0XX$.
- $P(X0X) > P(X1X)$. Low-status users act as a bridge to connect users so as to form a closure triad. The likelihood of $X0X$ is 2.8 times higher than $X1X$.
- $P(XX1) > P(XX0)$. The rich gets richer. This phenomenon validates the mechanism of preferential attachment [Newman 2001].

Following Influence in Triad Formation

Two Categories of Following Influences



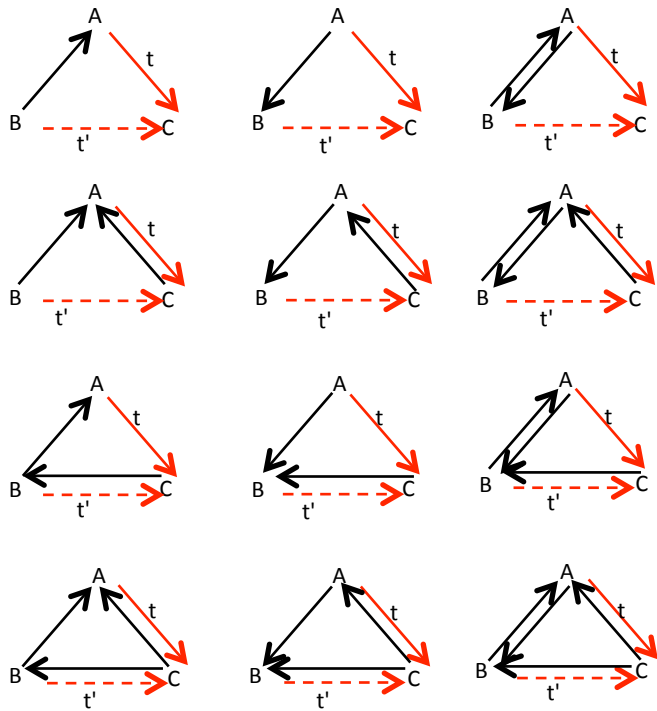
Follower diffusion

Followee diffusion

- >: pre-existed relationships
- >: a new relationship added at t
- >: a possible relationship added at $t+1$

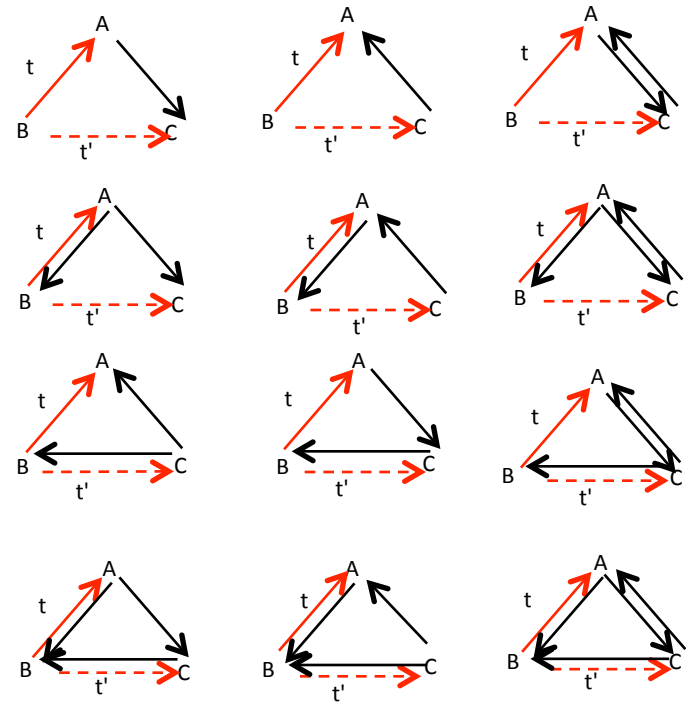
24 Triads in Following Influence

Follower diffusion



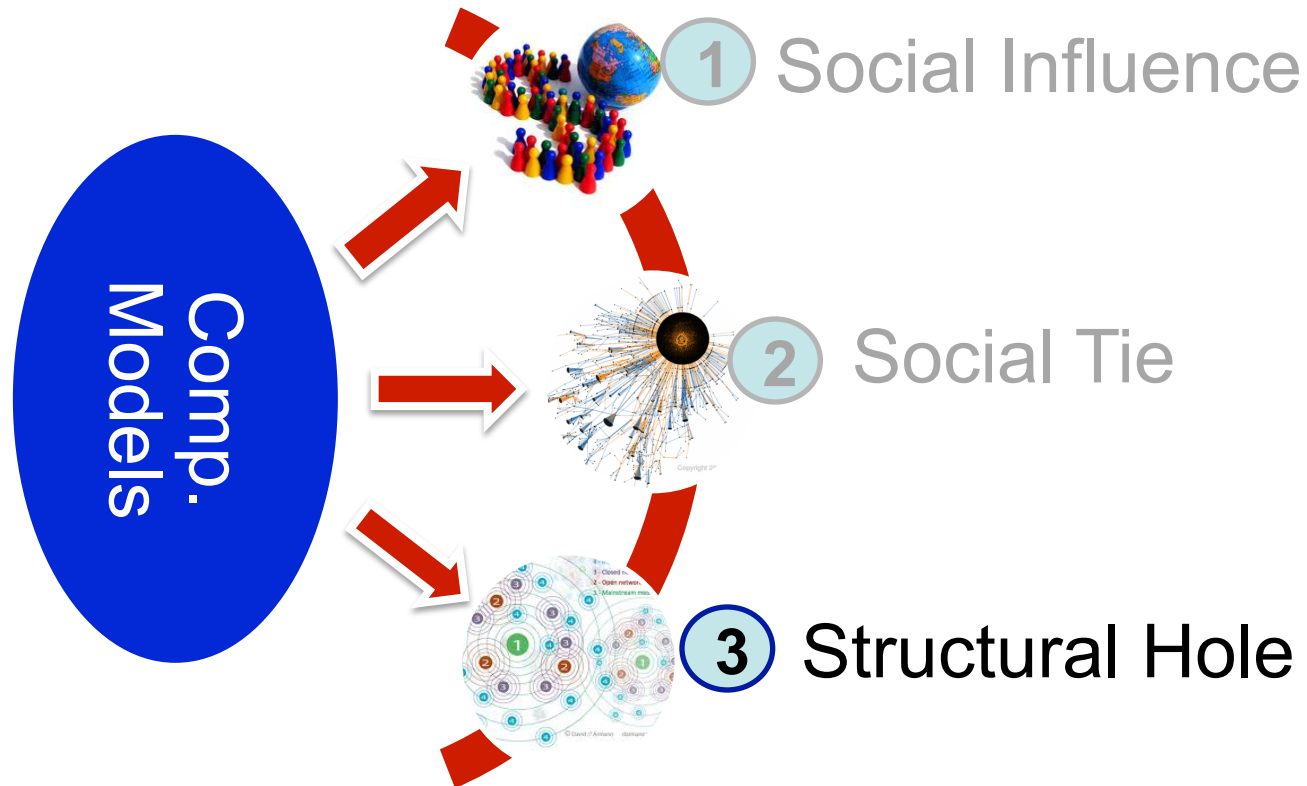
12 triads

Followee diffusion



12 triads

Micro-level Social Network Analysis



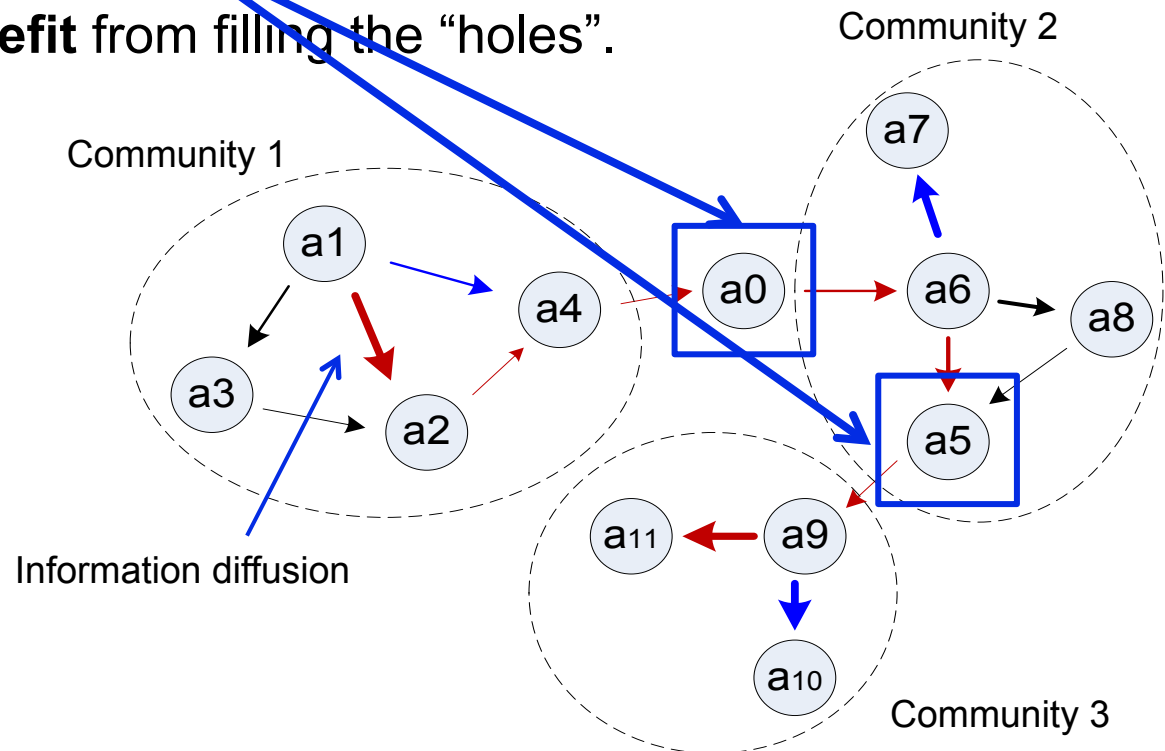
Structural Hole and Information Diffusion

- Structural Hole
 - Mining Top-k structural hole spanners [Lou-Tang 2013]
- Influence diffusion
 - Influence locality in information diffusion [Zhang-et-al 2013]

Mining Structural Hole Spanners

- The theory of Structural Hole [Burt92]:
 - “Holes” exists between communities that are otherwise **disconnected**.
- **Structural hole spanners**
 - Individuals would **benefit** from filling the “holes”.

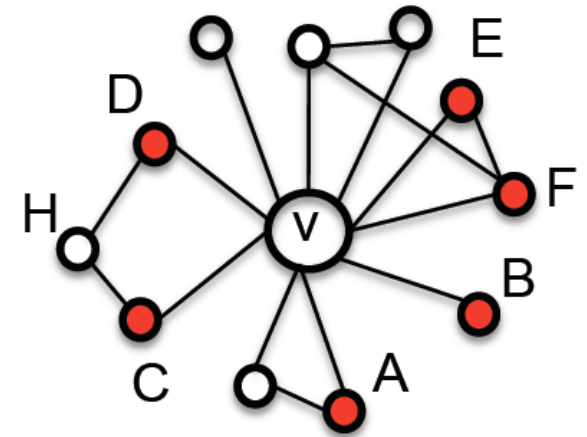
On Twitter, **Top 1%** twitter users control **25%** **retweeting** flow between communities.



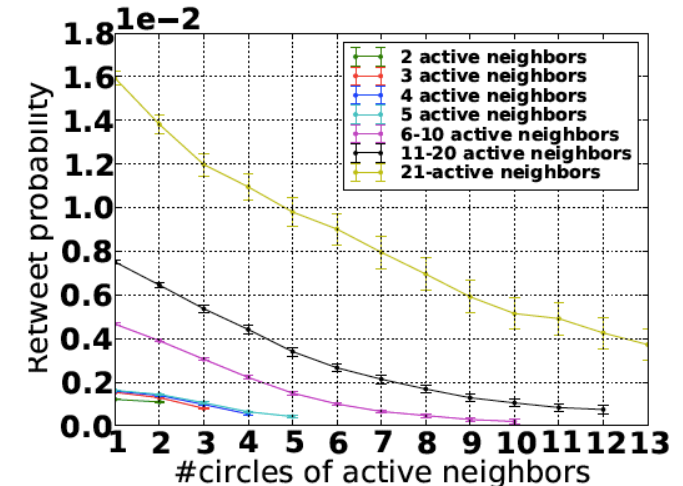
Influence Locality in Information Diffusion

- Randomization test
 - Debiased testing
 - Locality influence indeed exist (t-test, $p \ll 0.01$)
- Locality influence function

Part of v's 2-ego network



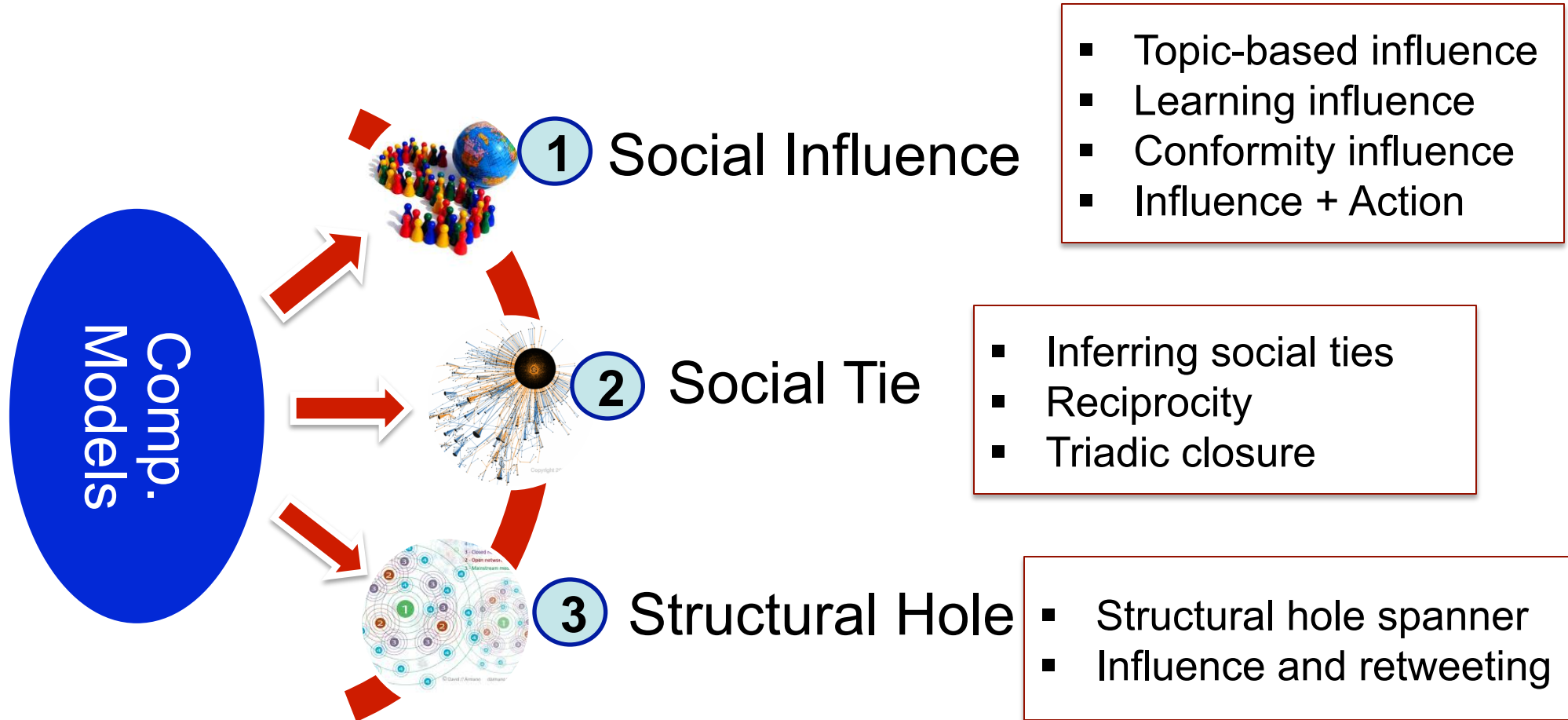
The retweet probability is negatively correlated with the number of circles.



- Help retweet prediction

Model	Prec.	Rec.	F1	Acc.
LRC-B	68.11	74.26	71.05	69.74
LRC-Q	66.82	77.22	71.65	69.44
LRC-BQ	69.89	77.06	73.30	71.93

Micro-level Social Network Analysis





Other Data mining Work—

Recommendation, Emotion Analysis, Expert Finding,
Integrations, Content Analysis

Social Applications & Web Mining

- Social recommendation
 - Cross-domain collaboration recommendation [Tang-et-al 2012, *best poster*]
 - Typicality-based recommendation [Cai-et-al 2013]
 - Patent partner recommendation [Wu-Sun-Tang 2013]
- Expert finding
 - Topic level expertise search [Tang-Zhang-Jin-Yang-Cai-Zhang-Su 2011]
 - Expertise matching with constraints [Tang-Tang-Lei-Tan-Gao-Li 2012]
 - Combining topic model and random walk [Tang-Jin-Zhang 2008]
 - Expert finding in social networks [Zhang-Tang-Li 2007]
 - User profiling [Tang-Zhang-Yao 2007, Tang-Yao-Zhang-Zhang 2010]

Algorithms and Applications (cont.)

- Information integration/alignment
 - Using Bayesian decision for alignment [Tang-et-al 2013]
 - Cross-lingual knowledge linking [Wang-Li-Wang-Tang 2012]
 - Cross-lingual knowledge linking via concept annotation [Wang-et-al 2013]
 - Name disambiguation [Tang-Fang-Wang-Zhang 2012]
 - A dynamic alignment framework [Li-Tang-Li-Luo 2009]
 - Unbalanced alignment [Zhong-Li-Xie-Tang-Zhou 2009]
- Content analysis/text mining
 - Social content alignment [Hou-Li-Li-Qu-Guo-Hui-Tang 2013]
 - Social context summarization [Yang-et-al 2011]
 - Ontology learning from folksonomies [Tang-et-al 2012, *spotlight paper*]
 - Tree-structural CRF [Tang-et-al 2006]

User Profiling

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

Search Experts Search

Jiawei Han FOAF Follow

Position: Professor
 Affiliation: Department of Computer Science, University of Illinois at Urbana-Champaign
 Address: 201 N. Goodwin Avenue, Urbana, IL 61801, USA.
 Phone: (217) 333-6903
 Fax: (217) 265-6494
 Email: han[at]cs.uiuc.edu
 Links: [Home] [Email]

See Others:
 Philip S. Yu Coauthor-Count: 50 H-index: 86 Follow
 Xifeng Yan Coauthor-Count: 54 H-index: 32 Follow
 Jian Pei Coauthor-Count: 37 H-index: 49 Follow

STATISTIC

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

Expertise:
 Data mining (146)
 Mobile Robot / Hybrid Control (119)
 XML Data (110)
 Data Mining / Query Processing (44)
 Database Systems (21)
 Information Retrieval / Probabilistic Indexing (14)

Conference:
 KDD (49) ICDE (40)
 IEEE Trans. Knowl. Data Eng. (36)
 ICDM (33)
 SIGMOD Conference (32)
 VLDB (21)

Bio
 Jiawei Han is computer scientist who specializes in research on Data Mining. He was the 2009 winner of the McDowell 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

Education
 Phd University Univ. Wisconsin-Madison Phd Major Computer Science Phd Date

Basic Info.

Citation statistics

Social Network

Social Graph

Colour Network B&W Network Ability

Relation: All

Publications EDIT DISAMBIGUATION RESULT

Group By Year Group By Category

ALL (560)	RECENT 100	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994
		1993	1992	1991	1990	1989	1988	1987	1986											

560 Query-Driven Discovery of Semantically Similar Substructures in Heterogeneous Networks
 Xiao Yu, Yizhou Sun, Peixiang Zhao, Jiawei Han
 2012, KDD, pp.1500~1503

559 Integrating Meta-Path Selection with User-Guided Object Clustering in Heterogeneous Information Networks
 Yizhou Sun, Brandon Nonck, Jiawei Han, Xifeng Yan, Philip S. Yu, Xiao Yu

Research Interests

Publications

[1] J. Tang, L. Yao, D. Zhang, and J. Zhang. A Combination Approach to Web User Profiling. ACM TKDD, (vol. 5 no. 1), 2010, 44 pages.

Name Disambiguation^[1,2]

Jing Zhang

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

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

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

		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 LC
78	EE	Jing Zhang, Hai Huang: Federate Job Mapping Strategy in Grid-Based Virtual Wargame Collaborative Environme
77	EE	Maria Wimmer, Michael Goul, Jing Zhang: Minitrack: E-Government Information and Knowledge Management. <i>H</i>
76	EE	Kai Kang, Jing Zhang, Baoshan Xu: Optimizing the Selection of Partners in Collaborative Operation Networks. <i>IC</i>
75	EE	Lingshuang Shao, Jing Zhang, Yong Wei, Junfeng Zhao, Bing Xie, Hong Mei: Personalized QoS Prediction for We
74	EE	Benvong Liu, Jing Zhang, Xiaowei Chen: Adaptive Training of a Kernel-Based Representative and Discriminative I
73	EE	Jilong Wang, Jing Zhang: Federation Based Solution for Peer-to-Peer Network Management. <i>International Confer</i>
72	EE	Jing Zhang, Fanhuai Shi, Jianhua Wang, Yuncai Liu: 3D Motion Segmentation from Straight-Line Optical Flow. <i>M</i>

- 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.

Expertise Search

Finding experts, expertise conferences, and expertise papers for "data mining"

The screenshot shows the ArnetMiner search results for the query "data mining". The interface includes a search bar at the top with the query "data mining" and a "Search" button. Below the search bar, there are suggestions: "Do you mean: Mark Mine, Ray Mines, Data Becker". The main results are divided into three sections: "Experts found for 'data mining'", "Conference/Journals", and "Publications".

Experts found for "data mining"
0 - 15 of 8748 experts for data mining (0 seconds)

- Jiawei Han** (FOAF)
Professor, Department of Computer Science, University of Illinois at Urbana-Champaign
H-index: 96, #Papers: 553, #Citations: 55885
Efficient Mining | Spatial Data Mining | Frequent Pattern Mining
- Philip S. Yu** (ALIAS: Philip Yu, Philip Shi-Lung Yu) (FOAF)
Professor and Wexler Chair in Information Technology, Department of Computer Science, University of Illinois Chicago
H-index: 86, #Papers: 683, #Citations: 32506
Data Mining | Data Streams | Data Mining Techniques
- Mohammed Javeed Zaki** (ALIAS: Mohammed J. Zaki, Mohammed Zaki) (FOAF)
Professor, Department of Computer Science, Rensselaer Polytechnic Institute
H-index: 46, #Papers: 722, #Citations: 9895
Data Mining | Web Usage Mining | Parallel Data Mining
- Christos Faloutsos** (FOAF)
Professor, Dept. of Computer Science Carnegie Mellon University
H-index: 81, #Papers: 357, #Citations: 31344
Data Mining | Data Visualization | Large Graphs
- Jian Pei** (FOAF)
Professor, School of Computing Science, Simon Fraser University
H-index: 49, #Papers: 219, #Citations: 15467
Uncertain Data | Frequent Pattern Mining | Frequent Itemset Mining
- H. Mannila** (ALIAS: Heikki Mannila) (FOAF)
Professor, Helsinki University of Technology
H-index: 52, #Papers: 193, #Citations: 14783
Knowledge Discovery | Data Mining | Data Mining Methods
- Charu C. Aggarwal** (ALIAS: Charu Chandra, Charu Chandra Aggarwal, Charu) (FOAF)

Conference/Journals
0 - 10 of 1373 conf/journals (0 seconds)

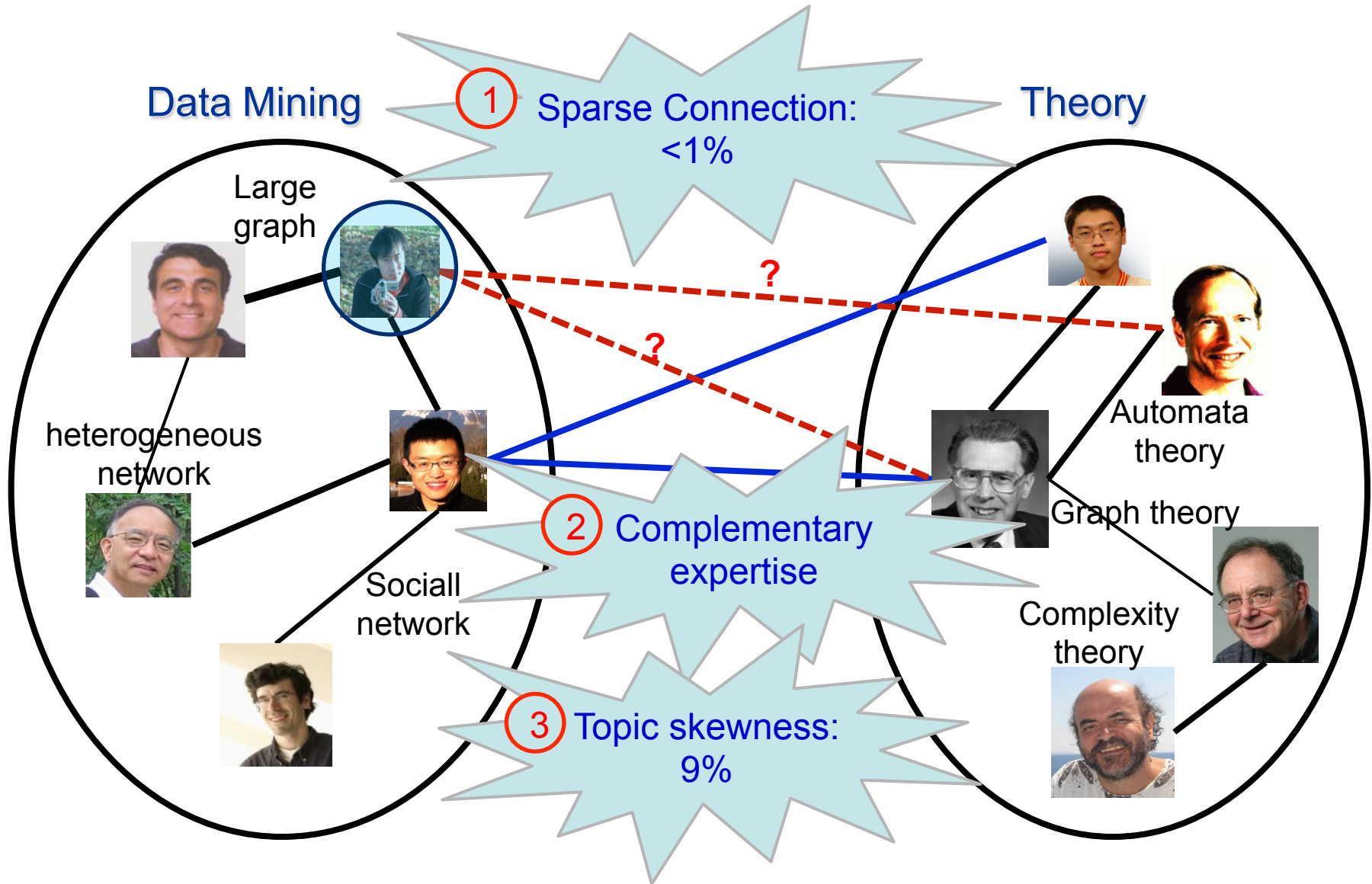
- KDD
- SIGKDD Explorations
- Datenverarbeitungspraxis
- ICDM
- Expert Syst. Appl.
- PAKDD
- ICDM Workshops
- CoRR
- Data Min. Knowl. Discov.
- PKDD

Publications
0 - 10 of 15651 publications (0 seconds)

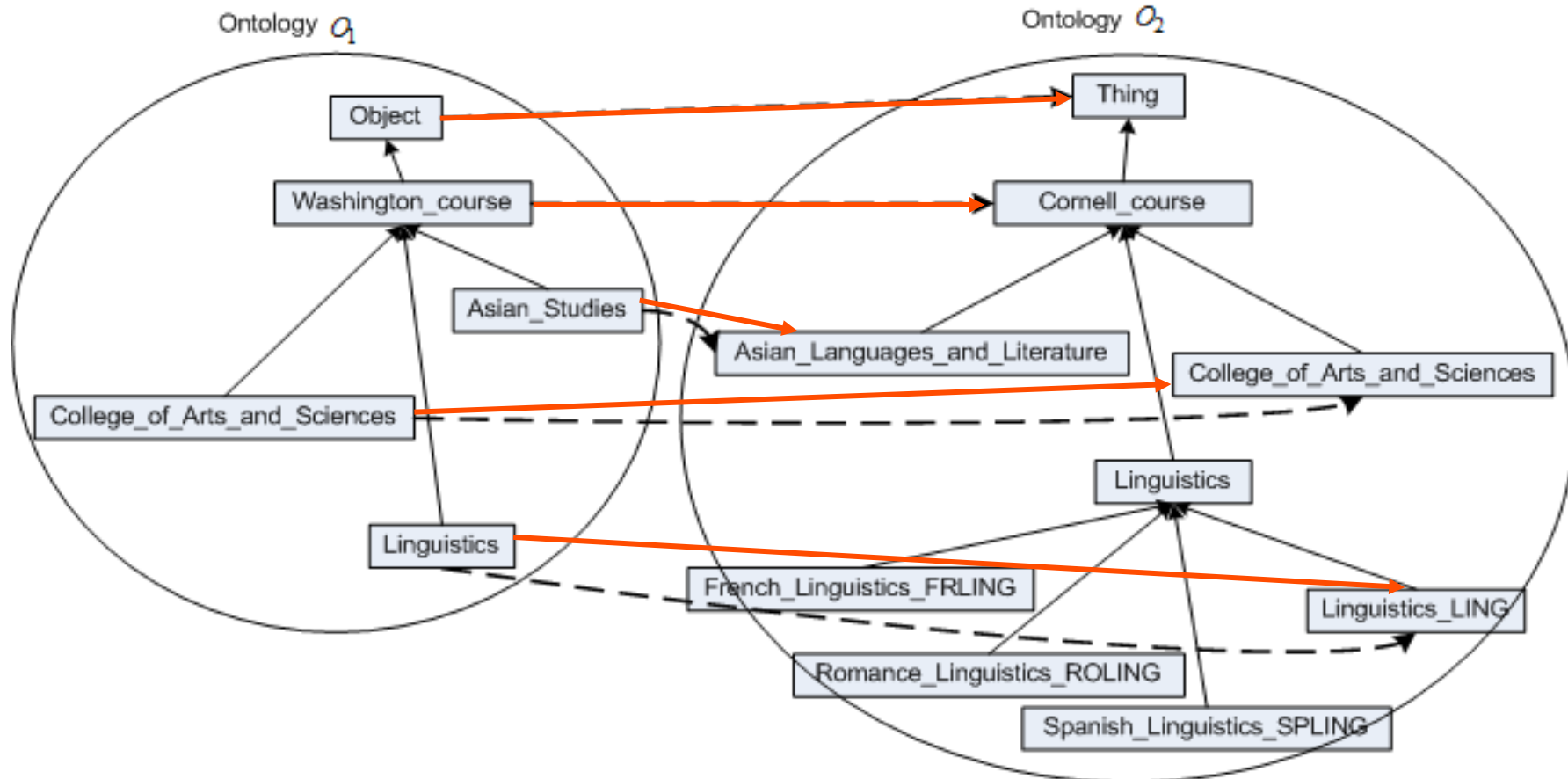
Order by: RELEVANCE | YEAR | CITATION

- Data Mining: Concepts and Techniques**
Authors: Jiawei Han, Micheline Kamber.
Published Year: 2000
CitedBy 13179 | PDF | BIBTEX
- Mining Association Rules between Sets of Items in Large Databases.**
Authors: Rakesh Agrawal, Tomasz Imielinski, Arun N. Swami.
JConf: SIGMOD Conference
Published Year: 1993
CitedBy 11393 | BIBTEX
- From Data Mining to Knowledge Discovery in Databases.**
Authors: Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth.
JConf: AI Magazine
Published Year: 1996
CitedBy 4332 | PDF | BIBTEX
- Introduction to Data Mining**
Authors: Pang-Ning Tan, Michael Steinbach, Vipin Kumar

Cross-domain Collaboration Recommendation



Schema Alignment



Knowledge Linking

The image shows two side-by-side screenshots of Wikipedia and Baidu Baike. The left screenshot is the English Wikipedia page for 'Anaerobic exercise', and the right is the Chinese Baidu Baike page for '无氧运动'. Annotations with colored boxes and arrows highlight specific elements:

- Titles:** A blue box highlights the article title 'Anaerobic exercise' on Wikipedia and '无氧运动' on Baidu Baike.
- Links:** Green boxes highlight internal links within the text, such as 'anaerobic metabolism' and 'aerobic exercise'.
- Categories:** A blue box highlights the 'Categories: Exercise physiology' section at the bottom of the Wikipedia page.
- Authors:** A yellow box highlights the list of recent editors at the bottom of the Wikipedia page, including 'Fcsprinter123', 'Nepenthes', and 'Fayerman'.
- Cross lingual links:** A red box highlights the 'Languages' menu on the left side of the Wikipedia page, showing options like 'Català', 'Deutsch', 'Español', etc.

The Baidu Baike page also features a '百科名片' (Infobox) with an image of a person on a stationary bike and a paragraph of introductory text. A blue box highlights the title '无氧运动' in the infobox area.

Cross lingual links

Authors

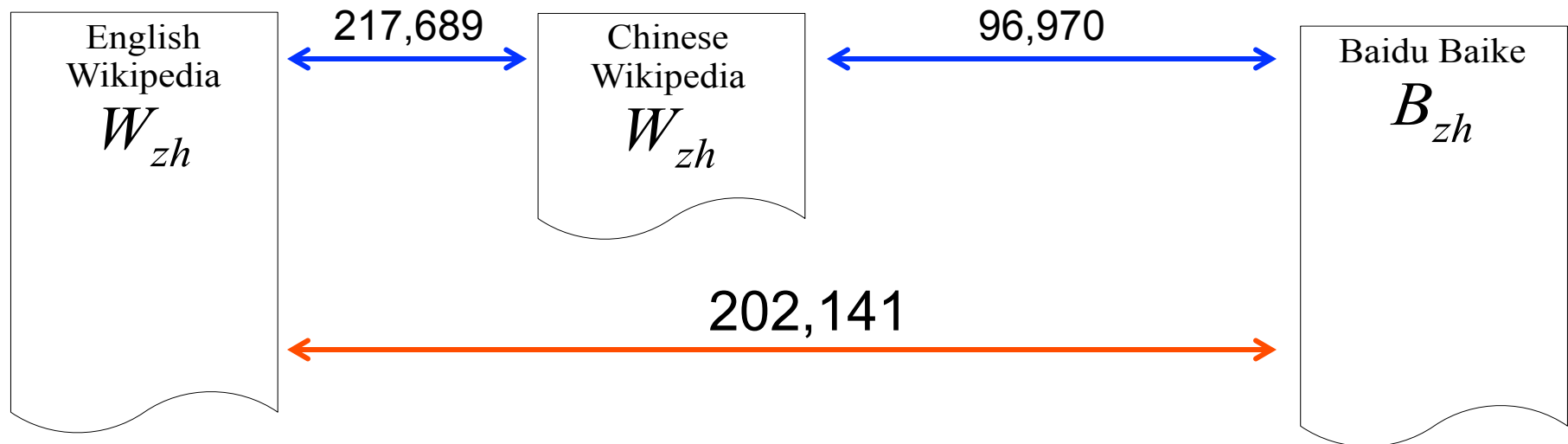
[1] Z. Wang, J. Li, Z. Wang, and J. Tang. Cross-lingual Knowledge Linking Across Wiki Knowledge Bases. In WWW'12, pages 459-468.

[2] Zhichun Wang, Juanzi Li, and Jie Tang. Boosting Cross-lingual Knowledge Linking via Concept Annotation. In IJCAI'13.

Results

- Discover crosslingual links between Wikipedia and Baidu

	Articles	Categories	Authors
Wikipedia	3,786,000	531,771	3,592,495
Baidu	3,941,659	599,463	1,454,204





System

AMiner (ArnetMiner)^[1]

- Academic Social Network Analysis and Mining system—Aminer (<http://aminer.org>)
 - Online since 2006
 - **>1 million** researcher profiles
 - **>131 million** requests
 - **>2.35 Terabyte** data
 - **100K IP access from 170 countries / month**
 - **10% increase** of visits per month
- Key Features
 - Mining semantics from academic data
 - Deep social network analysis
 - Knowledge based search

The screenshot shows the ArnetMiner profile for Jiawei Han. It includes a profile picture, contact information (Position: Professor, Affiliation: Department of Computer Science, University of Illinois at Urbana-Champaign), and a statistics table. The statistics table shows: H-index: 96, Uptrend: 30.46, Diversity: 0.71, #Papers: 563, Activity: 32.04, Sociability: 726.64, #Citations: 55885, Longevity: 26. There is also a 'Bio' section and a 'Research Interest' section.

STATISTIC			
H-index:	96	Uptrend:	30.46
#Papers:	563	Activity:	32.04
#Citations:	55885	Sociability:	726.64
		Longevity:	26

The screenshot shows search results for 'data mining' and a social graph. The search results list several researchers with their profiles and statistics. The social graph shows a network of researchers connected by lines, with a central node and many peripheral nodes. The graph is titled 'Social Graph' and includes a legend for 'Publications'.

Experts found for "data mining":

- Jiawei Han: H-index: 96, #Papers: 563, #Citations: 55885
- Philip S. Yu: H-index: 81, #Papers: 462, #Citations: 32584
- Mohammed Jawad Zaki: H-index: 47, #Papers: 192, #Citations: 9590
- Christos Faloutsos: H-index: 61, #Papers: 357, #Citations: 20344
- Jian Pei: H-index: 43, #Papers: 216, #Citations: 15467
- H. Mannila: H-index: 52, #Papers: 192, #Citations: 14792
- Charu C. Aggarwal: H-index: 47, #Papers: 208, #Citations: 11929

[1] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su. ArnetMiner: Extraction and Mining of Academic Social Networks. In KDD'08. pp.990-998.

AMiner History

2006

- Research profiling
- Integration
- Interest analysis



2009

- Topic analysis
- Course search
- Expert search
- **520K** users from **187** countries



2010

- Association
- Disambiguation
- Suggestion
- **1.02M** users from **202** countries



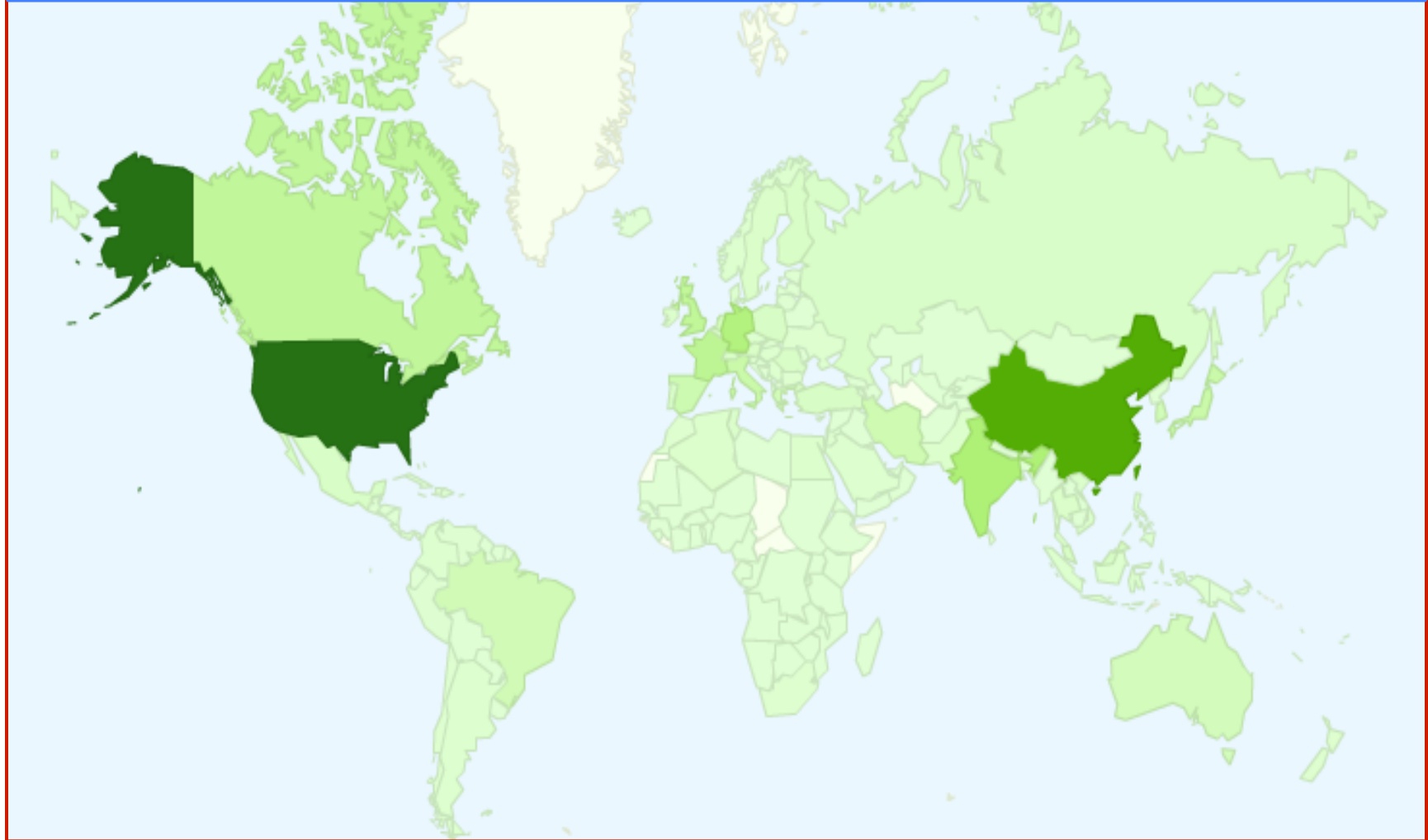
2012

- Geo search
- Collaboration recommendation
- **432M** users from **220** countries



User Distribution

4.32 million IP from 220 countries/regions



User Distribution

4.32 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 |

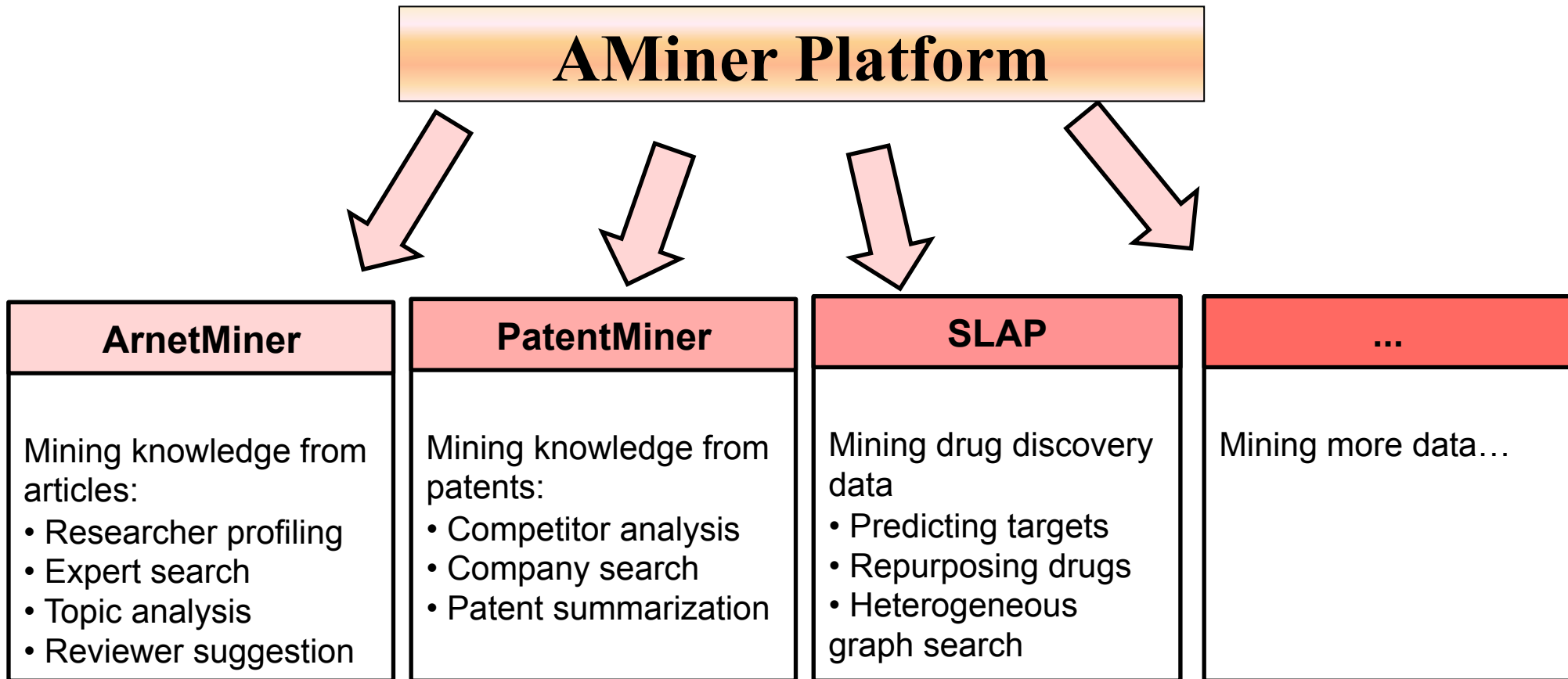
Widely used..

- The largest publisher: Elsevier
- Conferences
 - KDD 2010
 - KDD 2011
 - KDD 2012
 - KDD 2013
 - WSDM 2011
 - ICDM 2011-13
 - SocInfo 2011
 - ICMLA 2011
 - WAIM 2011
 - etc.



The screenshot shows the SciVerse website interface. At the top, there's a navigation bar with 'SciVerse Hub' and 'BETA' labels. Below that, a search bar contains 'data mining' with a search button and a search tips link. The search results show 'About 54317 results for: ALL:(data mining)'. On the left, there are filters for 'Search within results', 'Refine Results', and 'Content Sources'. The main content area features a banner for 'The 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining KDD-2011 San Diego, CA • August 21-24, 2011'. Below the banner, there are sections for 'Organizing Committee' (General Chair: Irwin King), 'Conference Organizers' (General Chair: Chidanand Apte, Program Co-Chairs: Joydeep Ghosh and Padhraic Smyth, Industry Track Co-Chairs: Ted E. Senator and Michael Zeller, Industry Practice Expo Co-Chairs: Ying Li and Rajesh Parekh), and 'Sponsors' (Platinum: Microsoft Research, Gold: Google, Technical: Arnetminer).

AMiner Platform...





PatentMiner: Topic-driven Patent Analysis and Mining

TECHNOLOGY | August 25, 2012, 1:41 p.m. ET

Apple Wins Big in Patent Case

Jury Finds Samsung Mobile Devices Infringed Six Apple Patents, Awards \$1.05 Billion in Damages

- A court decision in 08/2012: Samsung's Galaxy smart phone infringed upon a series of patents of Apple's iPhone, besides 4 appearance design patents, 3 software patents so-called 381, 915, and 163 are included, respectively cover "bounce back", "pinch-to-zoom", and "tap-to-zoom".
- The above 3 software patents all belong to the following three patent categories: active solid-state devices (touch screen), computer graphics processing (graph scaling), and selective visual display systems (tap to select).



PatentMiner: Topic-driven Patent Analysis and Mining

Jie Tang, Bo Wang, Yang Yang, Po Hu, Yanting Zhao, Xinyu Yan, Bo
Gao, Minlie Huang, Peng Xu, Weichang Li, and Adam K. Usadi

Tsinghua University

ExxonMobil Research and Engineering Company

<http://www.pminer.org>

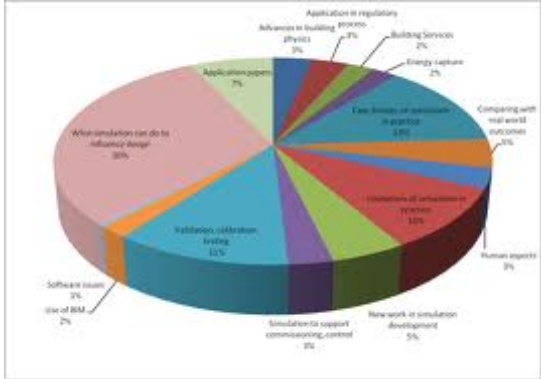
Knowledge and Network Integration



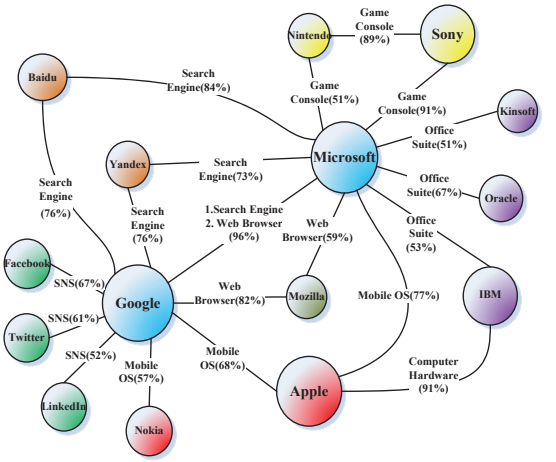
Patents

A Uniform Patent Search and Analytic System

PatentMiner



Hidden Topics



Competitive Network

Patent Search

PatentMiner

All Domain data mining

Summary of "data mining":

Result Categories (U.S. Class) Close

Tips: left click to focus, right click to return.

Mining or in situ disintegration of hard material				Hydraulic and earth engineering	
Hard material disintegrating machines	Processes	Cutter tooth or tooth head	Automatic control; signaling or indicating	Earth treatment or control	
			In situ conversion of solid to fluid		

Patent Found for "data mining"

Order by **relevance** date authority Found 3211 patents, used time 0.012 seconds

Extensible data mining framework

Inventors: Raman S. Iyer, Ioan Bogdan Crivat, C. James MacLennan, Scott C. Oveson, Rong J. Guan, ZhaoHui Tang, Pyungchul Kim, Irina G. Gorbach [1 reviews] 79%

Company: Microsoft Corporation

Issued Date: 2008-06-03

The subject disclosure pertains to extensible data mining systems, means, and methodologies. For example, a data mining system is disclosed that supports programmatic integration of information from multiple data sources and third parties, such ...

Patent data mining

Inventors: R. Bharat Rao, Sathyakama Sandilya, Christopher Jude Amies, Radu Stefan Niculescu, Arun Kumar Goel, Thomas R. Warrick [0 reviews] 69%

Company: Siemens Medical Solutions USA, Inc.

Issued Date: 2009-11-10

The present invention provides a data mining framework for mining high-quality structured clinical information. The data mining framework includes a data miner that mines medical information from a computerized patient record (CPR) based on domain-sp ...

Clustering module for data mining

Inventors: Marcos M. Campos [0 reviews] 73%

Company: Oracle International Corporation

Issued Date: 2006-08-15

A system, software module, and computer program product for performing clustering based data mining that improved performance in model building, good integration with the various databases throughout the enterprise, flexible specification and adjustm ...

Inventors	# query related / all
William R. Kennedy	# patents: 38/46
John M. Kennedy	# patents: 37/43
Gerhard Merten	# patents: 23/64
Walter Weirich	# patents: 21/49
Rakesh Agrawal	# patents: 21/171
John C. Stankus	# patents: 19/51
Kunibert Becker	# patents: 17/22
Raymond L. Wright	# patents: 17/20
Maurice K. LeBegue	# patents: 16/30
James J. Fallon	# patents: 14/24

Prev 1 / 451 Next

Companies	# query related / all
International Business Machines Corporation	# patents: 152/60180
Gewerkschaft Eisenhutte Westfalia	# patents: 147/349
The United States of America as represented by the Secretary of the Navy	# patents: 140/23714
Coal Industry	# patents: 59/283

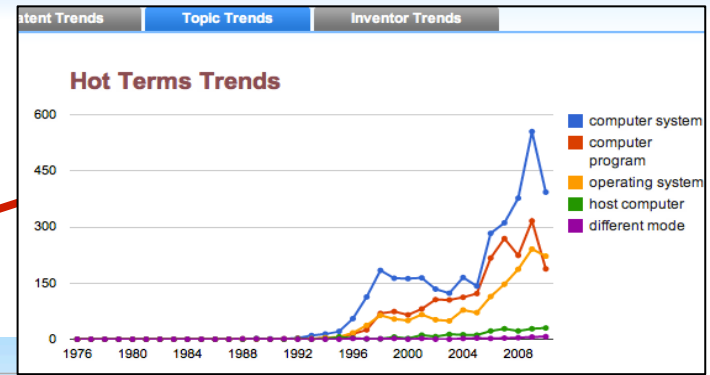
Topics of search results

Top Patents

Top Inventors

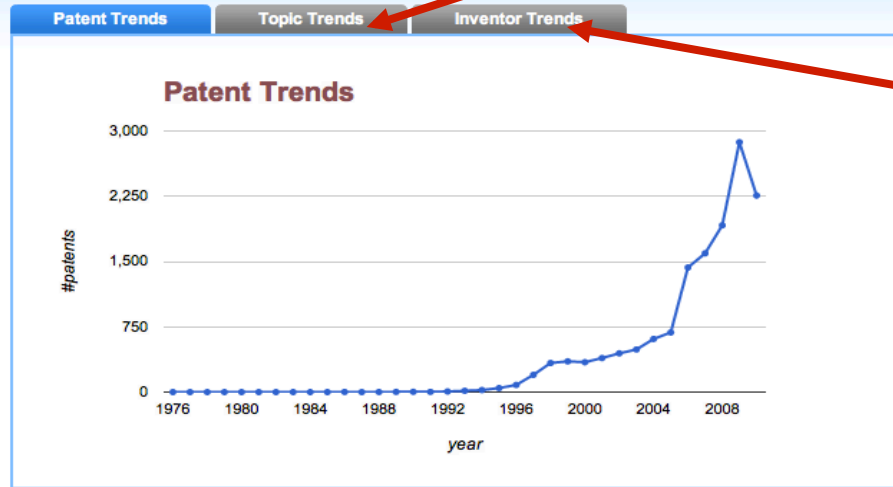
Top Companies

Topic-based Analysis for "Microsoft"



Microsoft Corporation

Trend Analysis:



Microsoft

Founded:
Albuquerque, New Mexico
1975

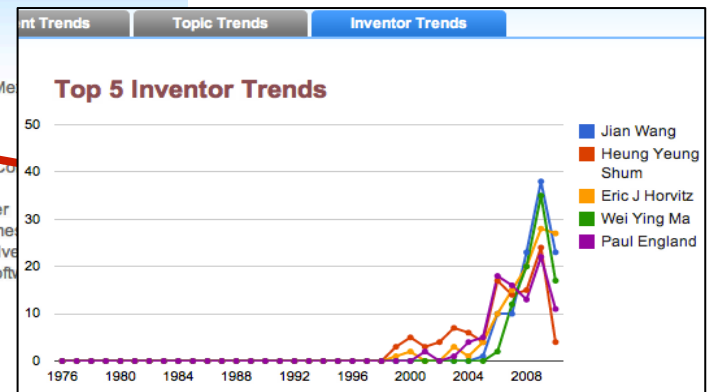
Industry:
Computer software; Consumer electronics; Digital distribution; Computer hardware; Video game consulting; Online advertising; Automotive software

Employees:
89,000 (2010)

Products:
See products listing

Website:
<http://www.microsoft.com>

Tags:
computer system, user interface



Competitors:

Global	Evolution	Topic	Topic Evolution
<p>International Business Machines Corporation # patents: 60180</p> <p>home compare</p>			
<p>Hewlett-Packard Development Company, L.P. # patents: 21015</p> <p>home compare</p>			

Competitors:

Global	Evolution	Topic
<p>International Business Machines Corporation # patents: 60180</p> <p>Hewlett-Packard Development Company, L.P. # patents: 21015</p> <p>Apple Computer, Inc. # patents: 3136</p> <p>Sun Microsystems, Inc. # patents: 7291</p> <p>Intel Corporation # patents: 18682</p> <p>Hitachi, Ltd. # patents: 38863</p> <p>SAP AG # patents: 1281</p>	<p>Microsoft (14100 Patents)</p> <p>Apple Computer (3136 Patents)</p> <p>Worldwide Area</p> <p>Computer software, Consumer electronics, Digital distribution, Computer hardware, Video games, IT consulting, Online advertising, Retail stores, Automotive software</p> <p>Albuquerque, New Mexico (April 4, 1975)</p> <p>\$86.113 billion (2010) Assets</p> <p>\$46.175 billion (2010) Equity</p>	<p>Apple Computer (3136 Patents)</p> <p>Worldwide Area</p> <p>Computer hardware, Computer software, Consumer electronics, Digital distribution</p> <p>Cupertino, California, U.S. (April 1, 1976 (1976-04-01))</p> <p>\$75.18 billion (FY 2010)[4] Assets</p> <p>\$47.79 billion (FY 2010)[4] Equity</p>

What is PMiner?^[1]

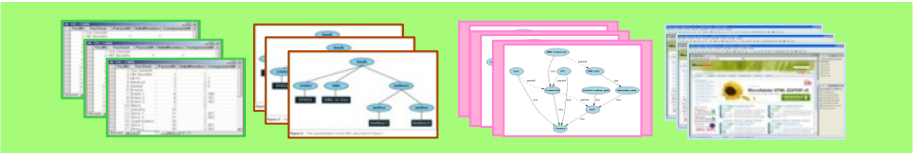
- Current patent analysis systems focus on search
 - Google Patent, WikiPatent, FreePatentsOnline
- PMiner is designed for an *in-depth* analysis of patent activity at the topic-level
 - Topic-driven modeling of patents
 - Heterogeneous network co-ranking
 - Intelligent competitive analysis
 - Patent summarization

- * Patent data:
 - > 3.8M patents
 - > 2.4M inventors
 - > 400K companies
 - > 10M citation relationships

- * Journal data:
 - > 2k journal papers
 - > 3.7k authors

The crawled data is increasing to >300 Gigabytes.

Opportunity: exploiting social network and data mining in the real-world



Web, relational data, ontological data, social data

Data Mining and Social Network techniques

Scientific Literature

Users cover >180 countries
>600K researcher
>3M papers

Arnetminer.org
(NSFC, 863)

Social search & mining

Social extraction
Social mining

IBM, Huawei

Advertisement

Advertisement
Recommendation

Sohu, Tencent

Mobile Context

Mobile search
& recommendation

Nokia, GM

Energy trend analysis

Energy product
Evolution
Techniques
Trend

Oil Company

Large-scale Mining

Scalable algorithms
for message tagging
and community
Discovery

Google, Baidu

Search, browsing, complex query, integration, collaboration, trustable analysis, decision support, intelligent services,

Future Directions

- Modeling user lifecycle and structure changes
- Building role-aware information diffusion models
- Mining the fundamental difference between different networks



Related Publications

- Tiancheng Lou and Jie Tang. Mining Structural Hole Spanners Through Information Diffusion in Social Networks. In **WWW'13**, pages 837-848, 2013.
- Jing Zhang, Biao Liu, Jie Tang, Ting Chen, and Juanzi Li. Social Influence Locality for Modeling Retweeting Behaviors. In **IJCAI'13**.
- Jie Tang, Sen Wu, and Jimeng Sun. Confluence: Conformity Influence in Large Social Networks. In **KDD'2013**.
- Tiancheng Lou, Jie Tang, John Hopcroft, Zhanpeng Fang, Xiaowen Ding. Learning to Predict Reciprocity and Triadic Closure in Social Networks. In **TKDD**, 2013.
- Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social Influence Analysis in Large-scale Networks. In **KDD'09**, pages 807-816, 2009.
- Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. ArnetMiner: Extraction and Mining of Academic Social Networks. In **KDD'08**, pages 990-998, 2008.
- Chenhao Tan, Jie Tang, Jimeng Sun, Quan Lin, and Fengjiao Wang. Social action tracking via noise tolerant time-varying factor graphs. In **KDD'10**, pages 807–816, 2010.
- Chi Wang, Jiawei Han, Yuntao Jia, Jie Tang, Duo Zhang, Yintao Yu, and Jingyi Guo. Mining Advisor-Advisee Relationships from Research Publication Networks. In **KDD'10**, pages 203-212, 2010.
- Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. User-level sentiment analysis incorporating social networks. In **KDD'11**, pages 1397–1405, 2011.
- Tiancheng Lou, Jie Tang, John Hopcroft, Zhanpeng Fang, Xiaowen Ding. Learning to Predict Reciprocity and Triadic Closure in Social Networks. In **TKDD**.
- Lu Liu, Jie Tang, Jiawei Han, and Shiqiang Yang. Learning Influence from Heterogeneous Social Networks. In **DMKD**, 2012, Volume 25, Issue 3, pages 511-544.

Thanks to all of our collaborators!

John Hopcroft, Jon Kleinberg, Lillian Lee, Chenhao Tan (**Cornell**)

Jiawei Han, Chi Wang, Yizhou Sun, and Duo Zhang (**UIUC**)

Tiancheng Lou (**Google**)

Jimeng Sun (**IBM**)

Wei Chen, Ming Zhou, Long Jiang (**Microsoft**)

Jing Zhang, Zhanpeng Fang, Zi Yang, Sen Wu, Jia Jia (**THU**)

.....

Jie Tang, KEG, Tsinghua U,
Download all data & Codes,

<http://keg.cs.tsinghua.edu.cn/jietang>
<http://arnetminer.org/download>