

# E6895 Advanced Big Data Analytics Lecture 6:

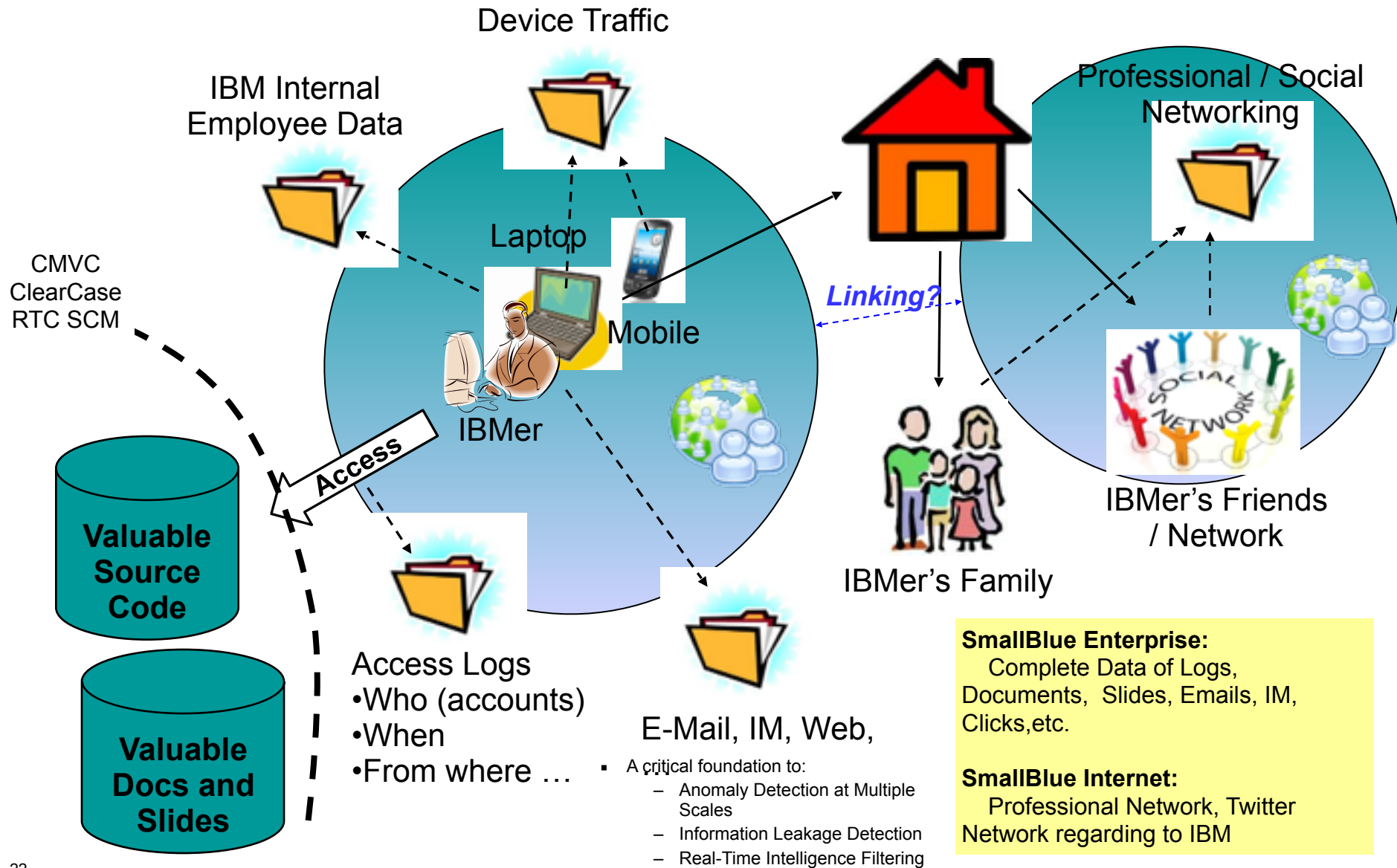
## *Social and Cognitive Analytics (II)*

Ching-Yung Lin, Ph.D.

Adjunct Professor, Dept. of Electrical Engineering and Computer Science

Mgr., Dept. of Network Science and Big Data Analytics, IBM Watson Research Center

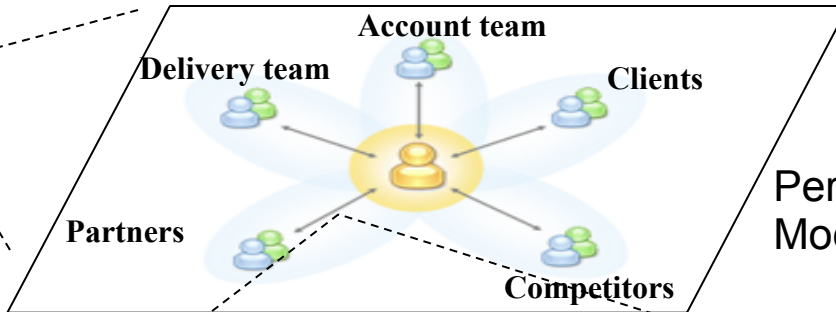




# Understanding People – from cognitive level to societal level

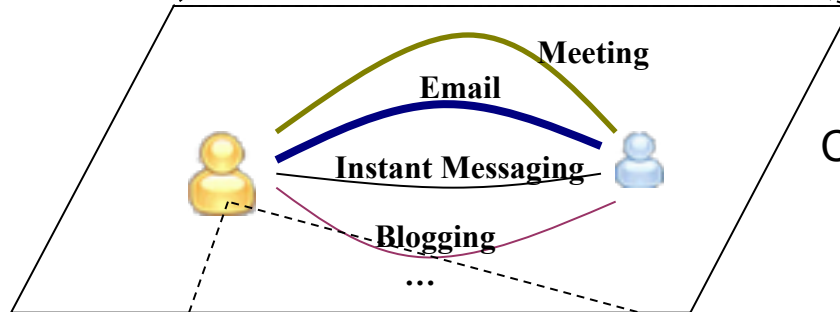


Organization Networks Modeling

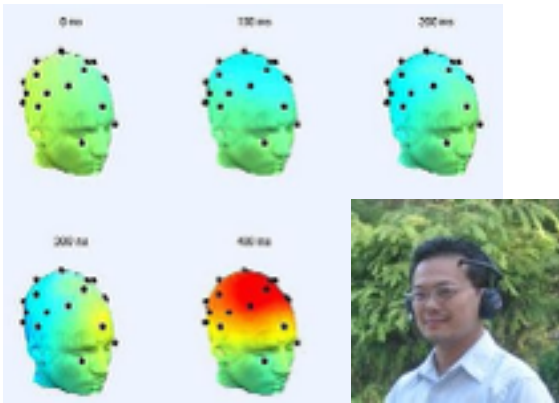


Personal Network Modeling

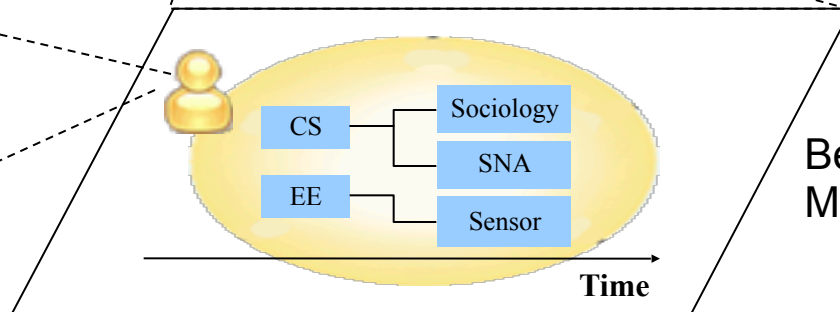
**Combining Signal Processing , Data & Graph Mining, and Machine Learning by multi-modality signal sensing & understanding**



Channel Modeling



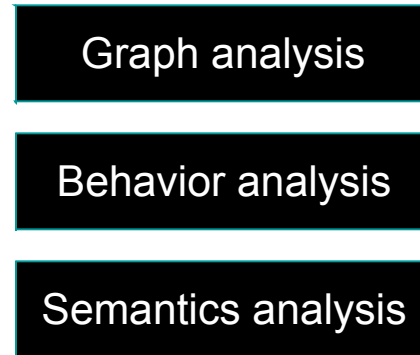
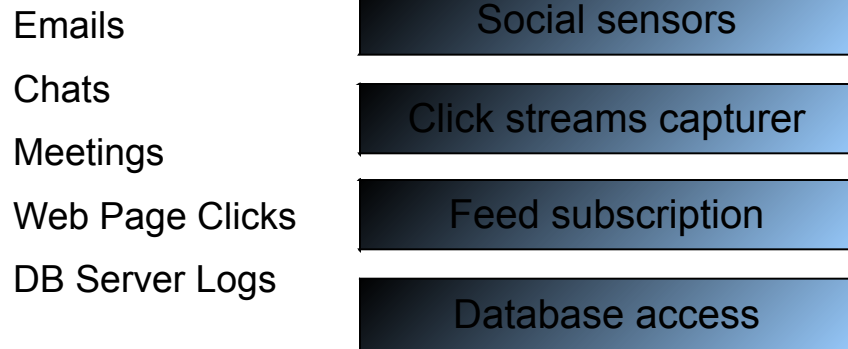
Cognitive Signal Modeling



Behavior Evolution Modeling

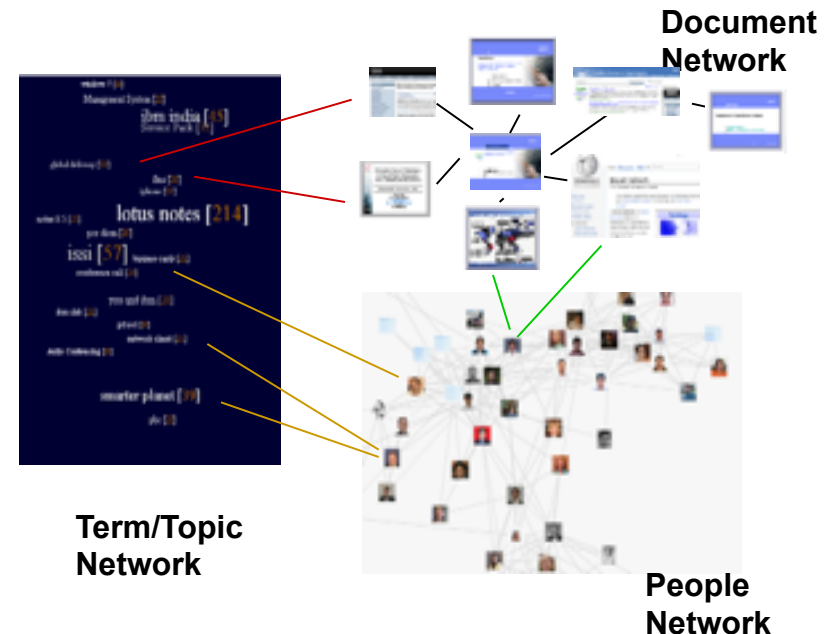
**Computational Social and Cognitive Sciences**

# Example — our enterprise social analytics system (SmallBlue)



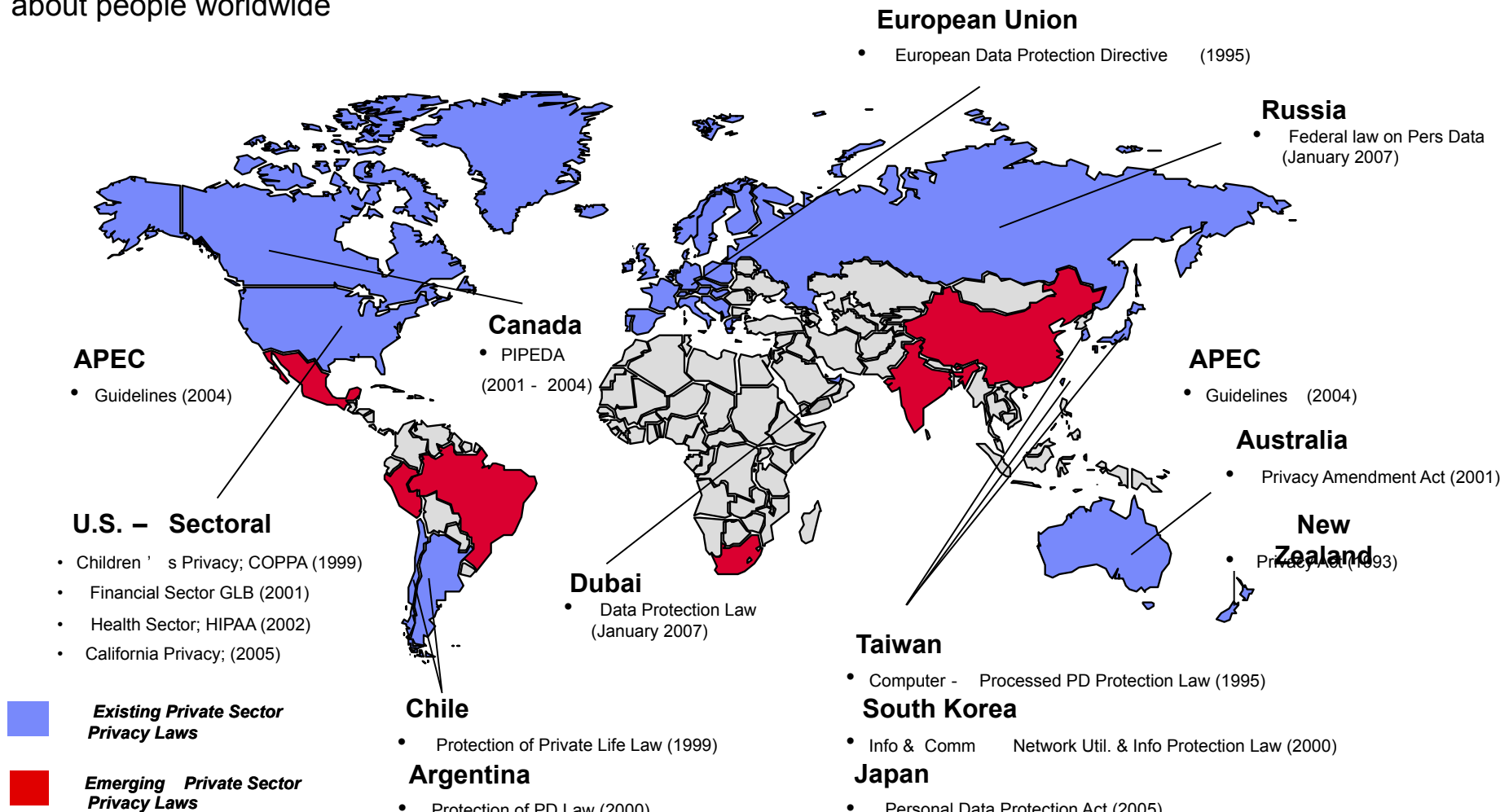
## IBM Deployment: Live Data, Production System

- 15,000 SmallBlue volunteers; 76 countries; 119,000 users
- 25,000,000 emails & SameTime messages (incl. content)
- 1,500,000 Learning click data; 44,000 entities
- 6,681,000 Knowl. & Sales access data; 240,000 entities
- 1,687,000 Media Library access data; 105,000 entities
- 700,000 Lotus Connections (blogs 3,000, file sharing 210,000, bookmark 450,000, Wiki 11,000) data
- 200,000 people's consulting financial databases
- 400,000 organization/demographic data
- 100,000 intranet w3 searches per day



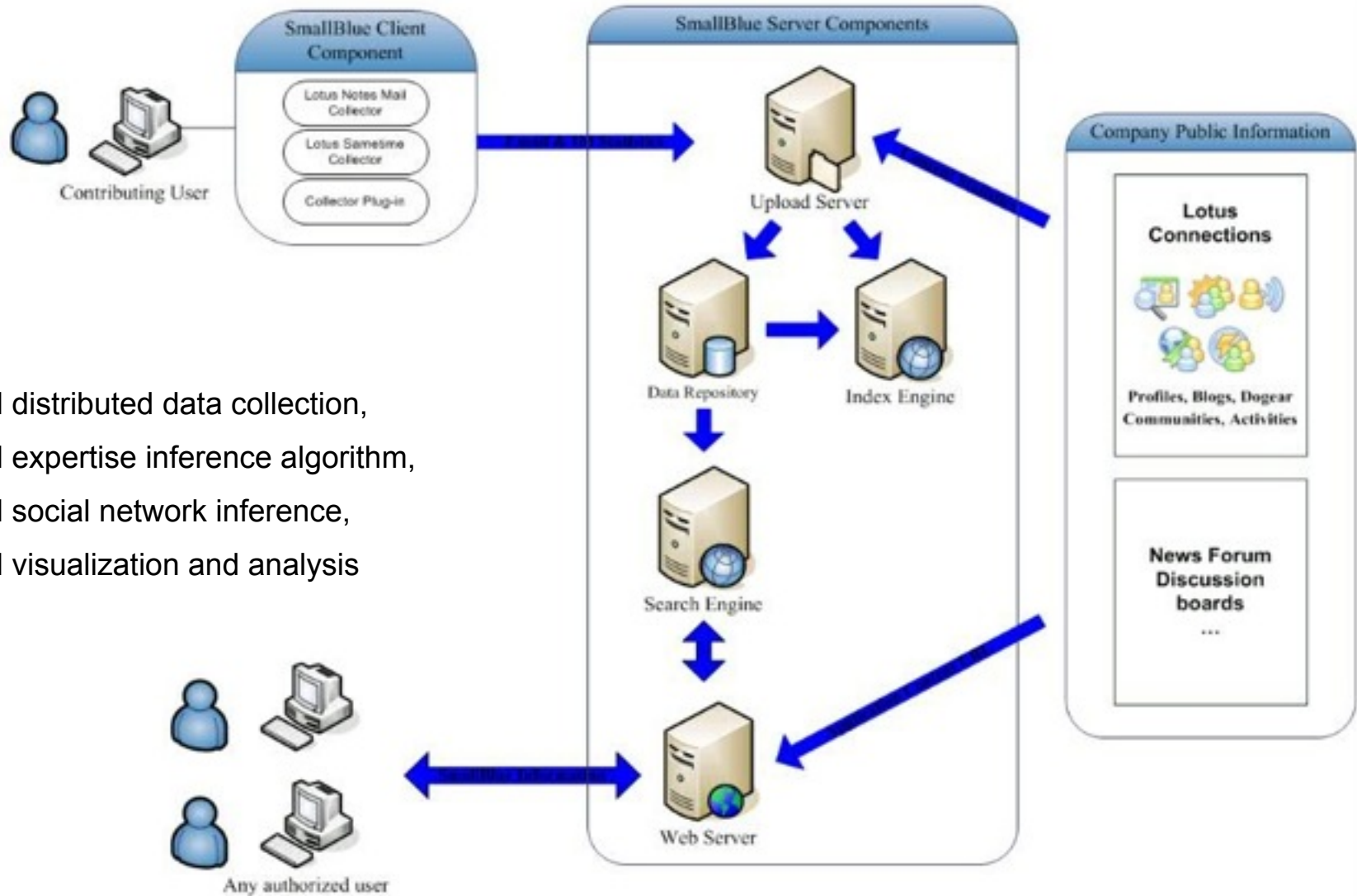
# Privacy – adaptive features for global privacy laws

- Privacy features and worked with GBS to go through 2-year global privacy review with privacy officers and labor union approval to make SmallBlue a deployable production system
- A unique large-scale social network capturing and process system that is lawful & user-aware system about people worldwide





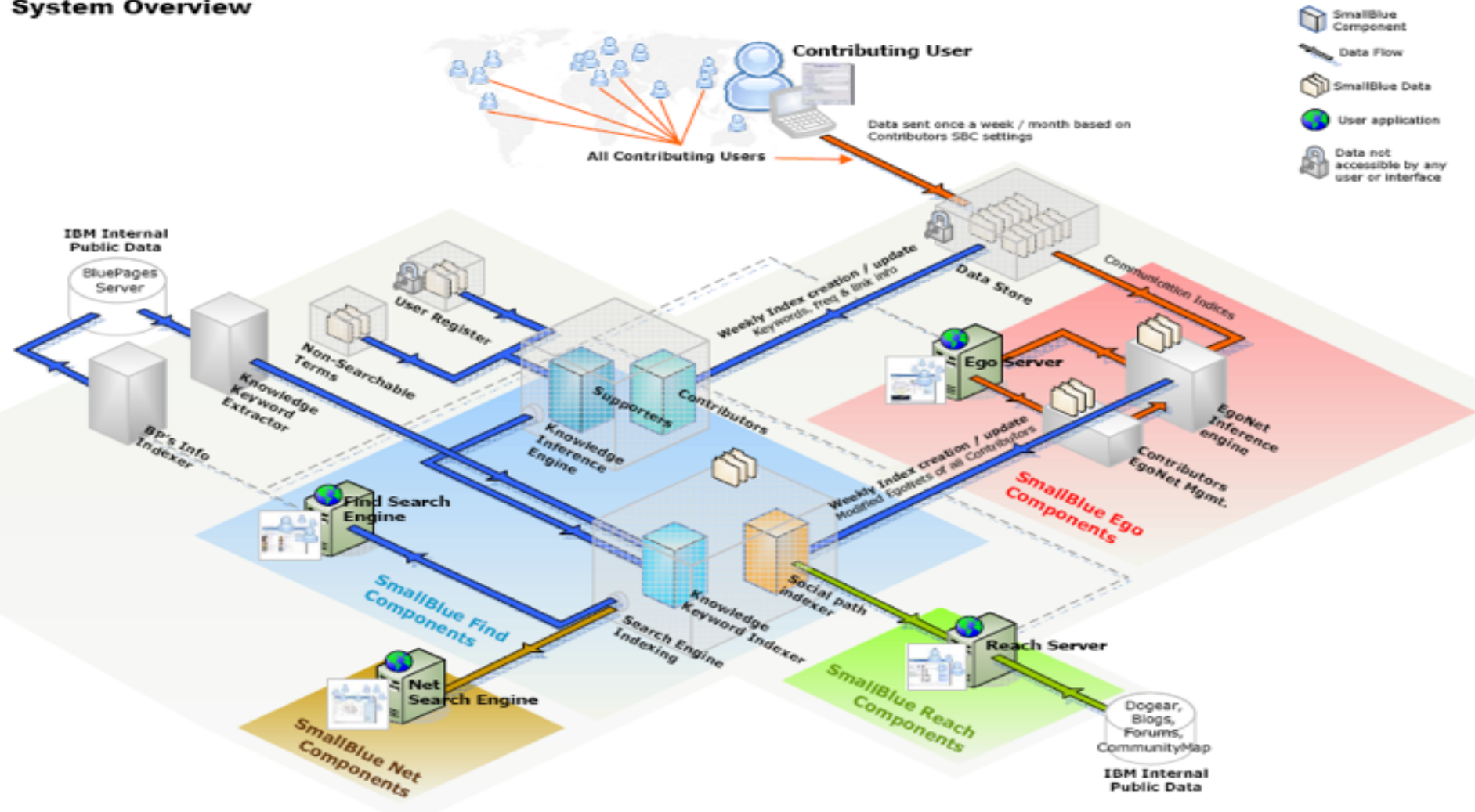
# SmallBlue Enterprise Architecture 1.1 (now v3.6)



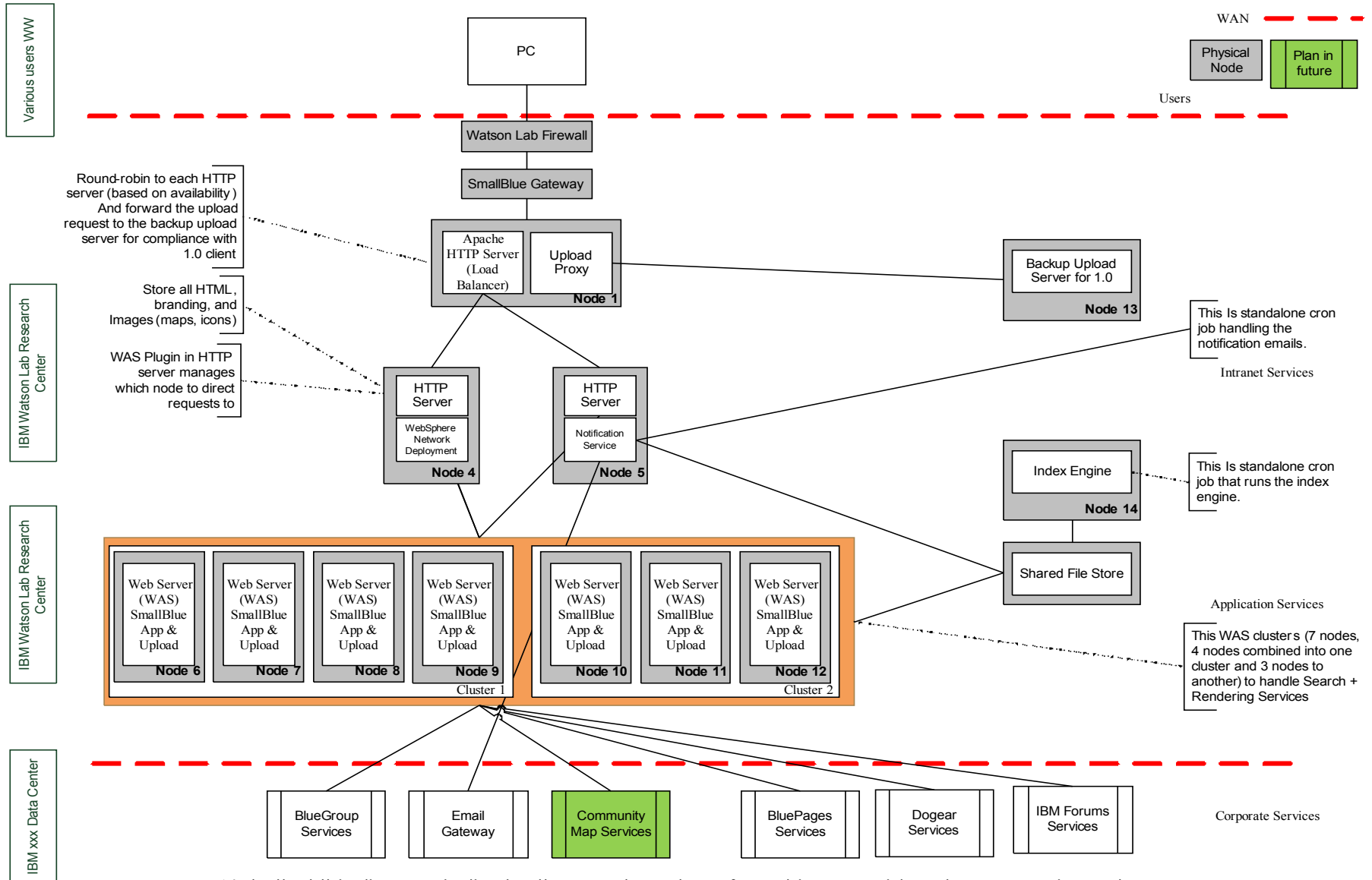
- novel distributed data collection,
- novel expertise inference algorithm,
- novel social network inference,
- novel visualization and analysis

# SmallBlue Enterprise 1.1 ( outdated drawing; current version 2.4)

## System Overview



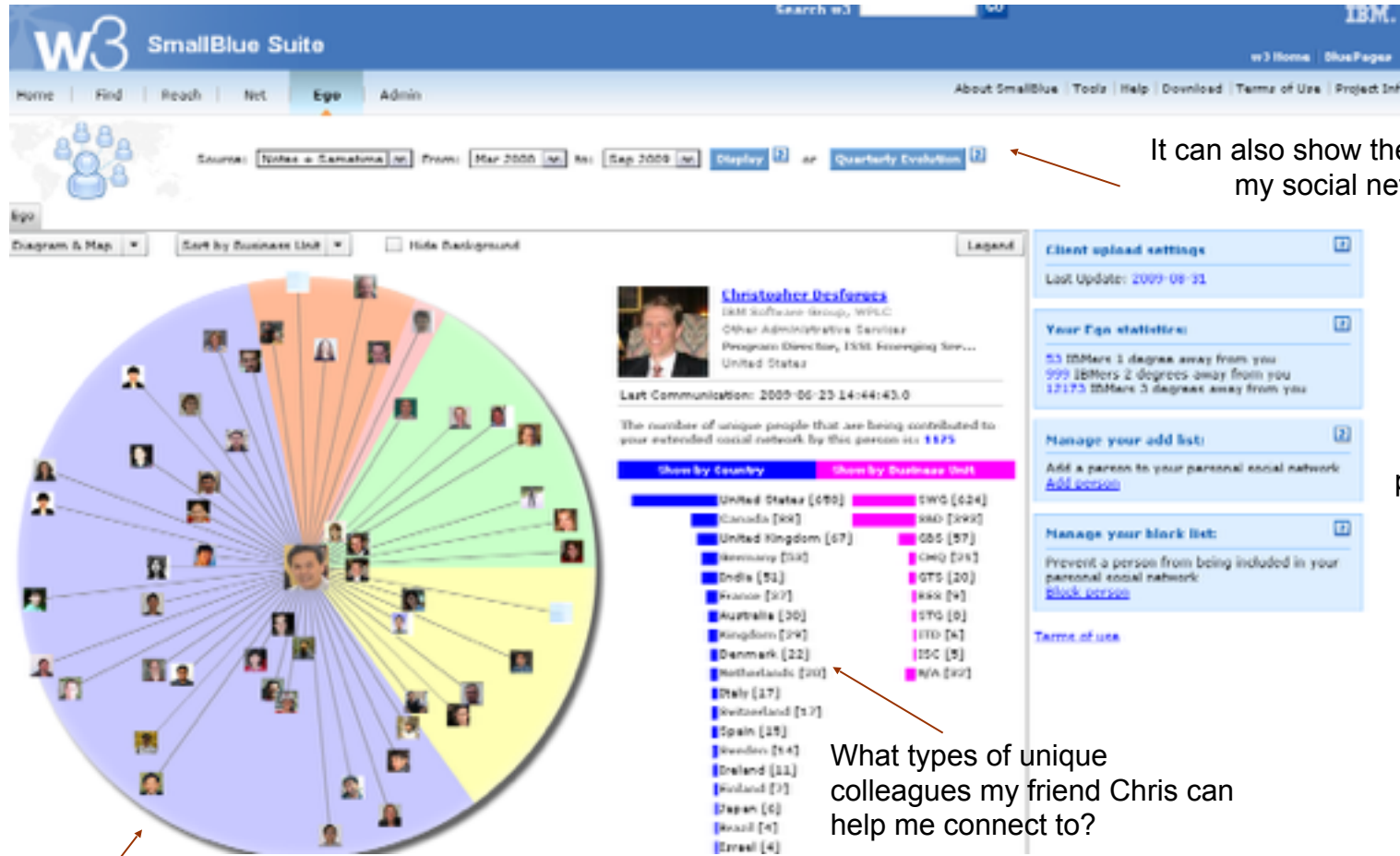
# SmallBlue Enterprise 1.1 architecture (not including Whisper & Synergy)



Note that this diagram indicates the exact number of machines considered necessary in each location type. The community map service will be added in future.



# Personal Social Network Capital



It can also show the evolution of my social network..

How many people in my personal networks?

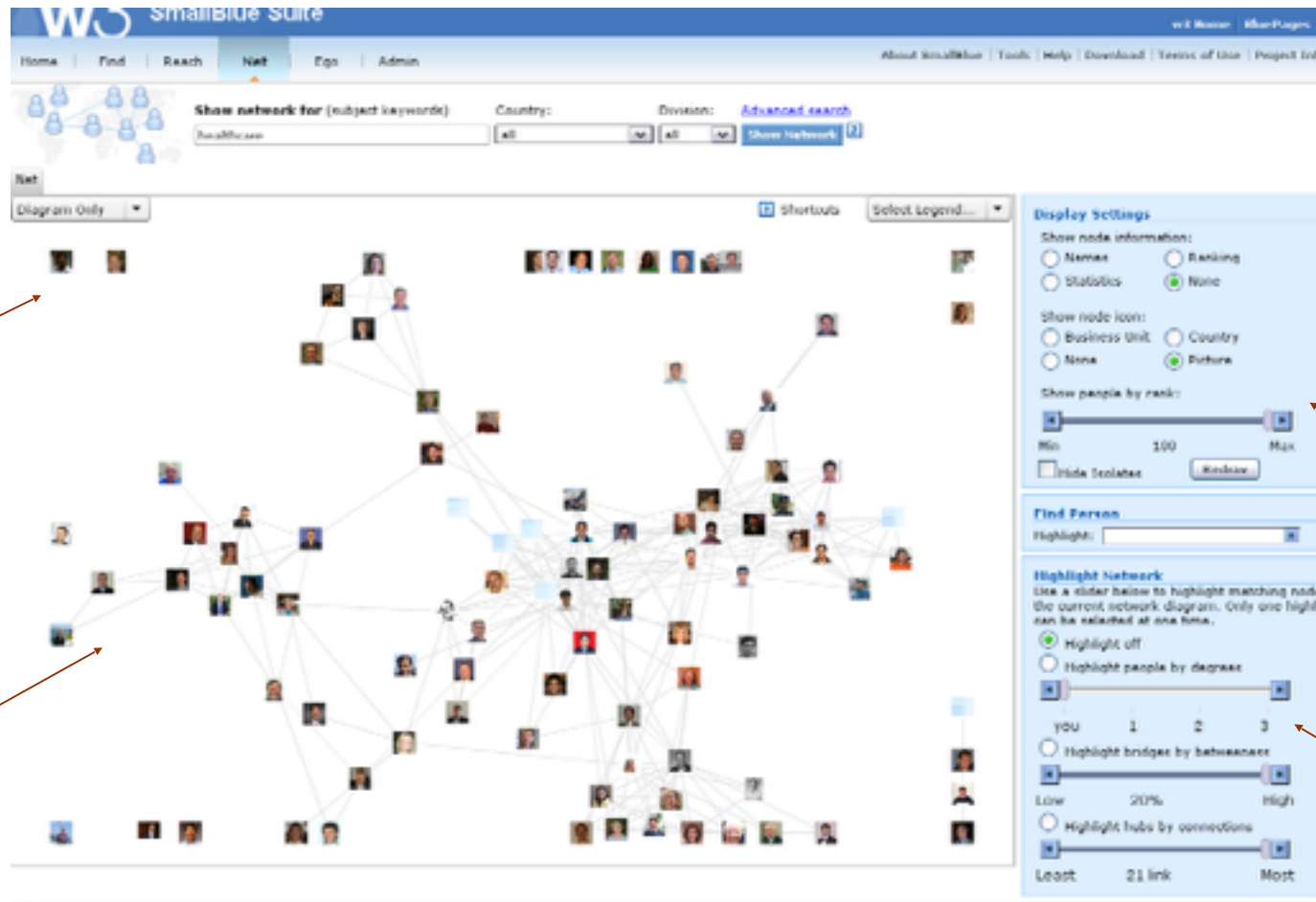
What types of unique colleagues my friend Chris can help me connect to?

My personal social network automatically found by SmallBlue with social distance

- Adding a person in personal network (i.e., frequent communications), increases \$948 yearly revenue for IBM. (selected by BusinessWeek Magazine as the Top Story of the Week, April 10, 2009)
- 1% increase in social network diversity is associated with \$239.5 in monthly revenue
- 1% increase in social network diversity is associated with an increase of 11.8% in job retention.

# Visualizing Social Network Analysis

- How are company's employees communicating 'healthcare' linking with each other? Who are the key bridges? Who have the most connections? How do these people cluster? It can be extended to analyze relationship of customers

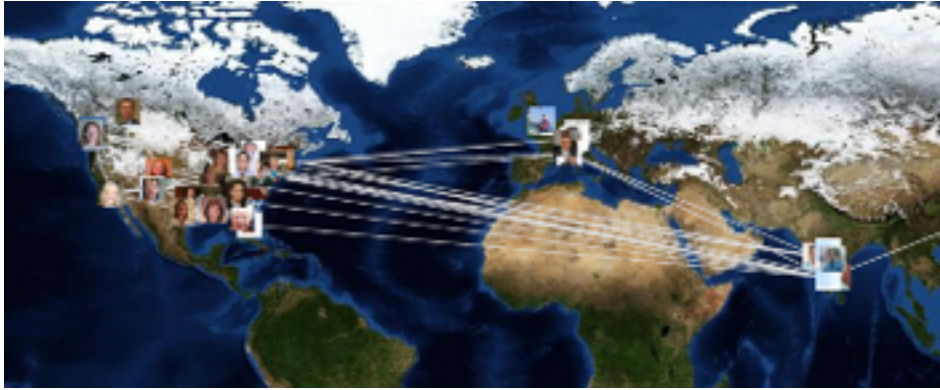


Independent experts on healthcare

A cluster of healthcare experts

Top-N experts on healthcare

Highlight experts based on my social proximity, the number of experts she connects, or the 'social bridges' importance



IBM Healthcare-related employees in the world



Connections between different divisions




IBM Healthcare-related employees in the U.S.



Key social bridges

# Finding expertise

E.g.: Search for the most knowledgeable colleagues within my 3-degree network for who knows 'healthcare'. (or within a country, a division, a job role, or any group/community)



The screenshot shows the SmallBlue Suite search interface. The search results are as follows:

Rank	Name	Company/Role	Ask
1.	Patricia (Pattie) Okita	Global Business Services Associate Partner, Healthcare Integration Other Consultant	Ask: MARTHA E. (Martha) GIBSON > Amy D. (AMY) Berk
2.	Michael Hebenberger	IBM Research Life Sciences Business Development Category Sales	Ask: Ravi B. Konuru > Vanessa L. Johnson
3.	Todd (T.M.) Kalrniuk	Global Business Services GBS Partner, Healthcare and Public Health -- Practice Administrator is Shirley Carkner Other Consultant	Ask: Chung Sheng Li > Robert (R.) Torok
4.	Susan E. (SUSAN) Rivers	Global Business Services Healthcare Knowledge Manager Market Insights	Ask: MARTHA E. (Martha) GIBSON
5.	H.C. (Mark) Eflingham	IBM Sales & Distribution, Public Sector Client Technical Advisor	Ask: Ari Fishkind > Julie A. Reid
6.	Paul (P.F.) Van Angelen	Global Business Services Pacific Development Center, Business Development Manager Other Consultant	Ask: Michael W. Ticknor > Kinson (K.W.) Lee
7.	Eric S. (ERIC) Minkoff	Global Business Services US GBS Learning & Knowledge Learning Deployment Lead - Public Sector	Ask: James (JAMES) Stupak > Andrea R.
8.	Thomas (Tom) Cocozza	Global Business Services Healthcare Transformation Services	Ask: MARTHA E. (Martha) GIBSON > Alan J. (ALAN) Lauder

Annotations on the screenshot:

- Red arrow pointing to Susan E. (SUSAN) Rivers: "My shortest path to Susan"
- Red arrow pointing to Paul (P.F.) Van Angelen: "As a user, you can only see their public information. Private info is used internally to rank expertise but private data can never be exposed."
- Red arrow pointing to Thomas (Tom) Cocozza: "Click a name to see their profile (SmallBlue Reach)"



# Social Paths



Is Tom a right person to me? Shortest Social Paths to any person within 6-degrees..

His official job role, title, contact info

His self-described expertise

His public communities

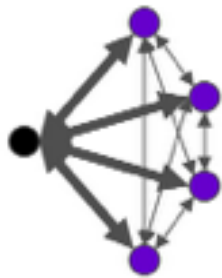
The public interest groups he is in

His blogs, forum, postings..

My various paths to Tom. SmallBlue can show the paths to any colleagues up to 6-degree away

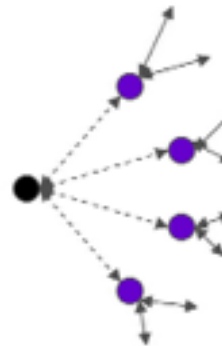


- Topological point of views
  - What type of network structure is beneficial?



## Cohesive Network

- Trust
- Absorptive capacity
- Precision, Reliability



## Structurally Diverse Network

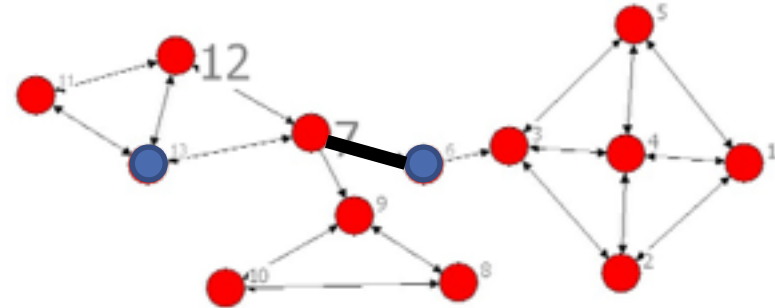
- Brokering position
- Access to many pools of diverse, novel information



What type of network structure is most beneficial in a electronic network for consultants?

- Importance of Direct Contacts?
- Importance of Indirect Contacts?
- Constrained vs. unconstrained?

# Network Topology Measures



## Direct Contacts

Size(7) = 4  
Size(12) = 3

+ No information distortion  
- High maintenance cost

Network size → strong work performance (?)

## Indirect Contacts

Btw(7) = 33  
Btw(12) = 6  
3steps(7) = 11  
3steps(12) = 8

+ Access diverse information  
- Information distortion

Btw-centrality → Strong work performance (?)  
3-step Reach → Strong work performance (?)

## Structural Diversity

Div(7) = .53  
Div(12) = 0.16

+ Transfer complex knowledge  
- Access diverse knowledge

Diversity → Strong work performance (?)

- MIT studied 2,038 IBM Global Business Consultants for 2 years, it was found that:
  - After a consultant started using SmallBlue, his social network/capital obviously grew and his monthly billable revenue for IBM increased by **\$584.15** (i.e., \$7,010 per year)

- Joint analysis of social capital and economic capital:
  - Adding a person in personal network (i.e., someone with frequent communications), increases **\$948** yearly revenue for IBM. *(selected by BusinessWeek Magazine as the Top Story of the Week, April 8, 2009)*
  - 1% increase in social network diversity is associated with **\$239.5** in monthly revenue (i.e., \$2,874 revenue increase per year).
  - 1% increase in social network diversity is associated with an increase of 11.8% in job retention (i.e., surviving layoff).

- IBM Research Achievement – “SmallBlue made Millions of Contribution to GBS in 2009”



SmallBlue / Atlas was featured in 120+ news articles, including 4 times by *BusinessWeek* (Jan and May 2008, April and June 2009)

# Observations from Personal Social Networks vs. Revenue

- Structural Diverse networks with abundance of structural holes are associated with higher performance.
  - *Having diverse friends helps.*
- Betweenness is negatively correlated.
  - *Being a bridge between a lot of people is not helpful.*
- Network reach are highly correlated.
  - *The number of people reachable in 3 steps is positively correlated with higher performance.*
- Having too many strong links — the same set of people one communicates frequently is negatively correlated with performance.
  - *Perhaps frequent communication to the same person may imply redundant information exchange.*
    - Future textual analysis can be done to confirm this.

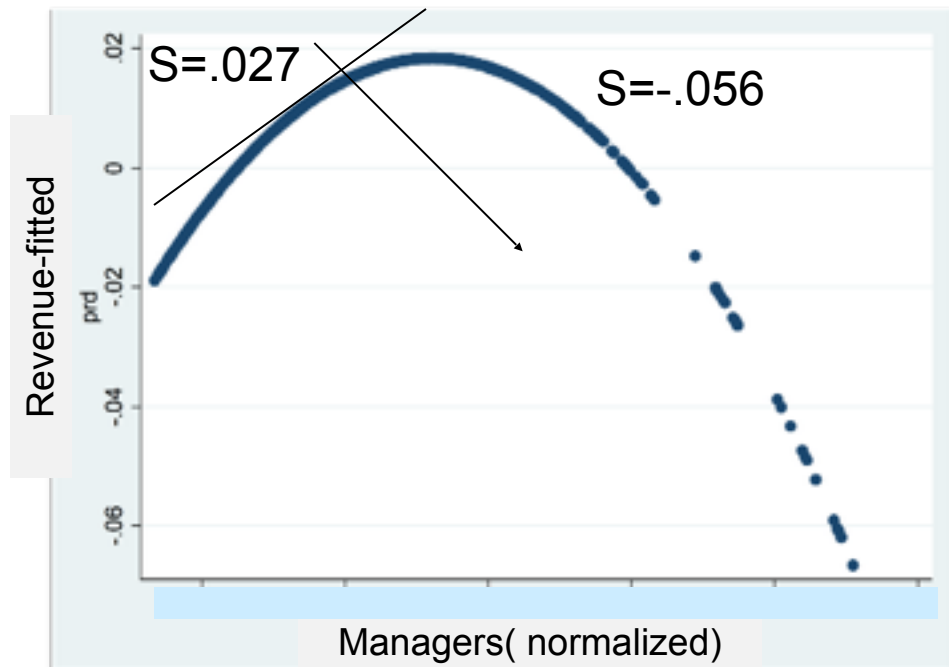
# Project Team Composition—Managers

The number of managers in a project exhibit an inverted-U shaped curve.

1. Having managers in a project is correlated with team performance initially.
2. Too many managers in a project is negatively associated with team performance.

$$revenue = \alpha + \beta_1 \cdot mgr + \beta_2 \cdot mgr^2 + \gamma_1 \cdot otherfactor_1 + \dots + \gamma_k \cdot otherfactor_k + \varepsilon$$

# Managers in project	$\beta_1$	2733.9*** (537.5)
(# Managers in project) <sup>2</sup>	$\beta_2$	-682.02*** (215.3)





# Culture Factor in CMC-based Communications



## Collaborating Globally:



preferences of CMC tools

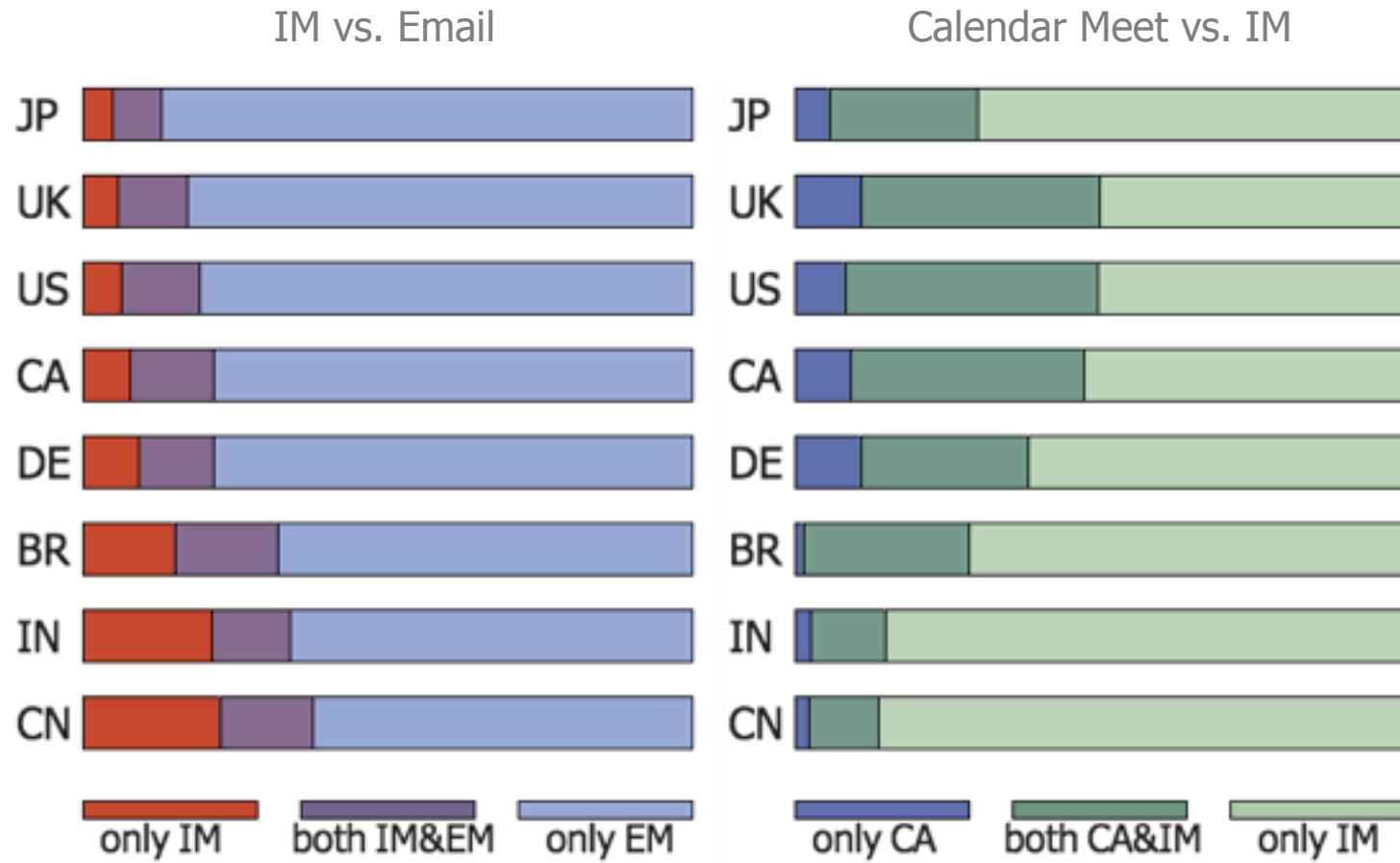


patterns of growing social network

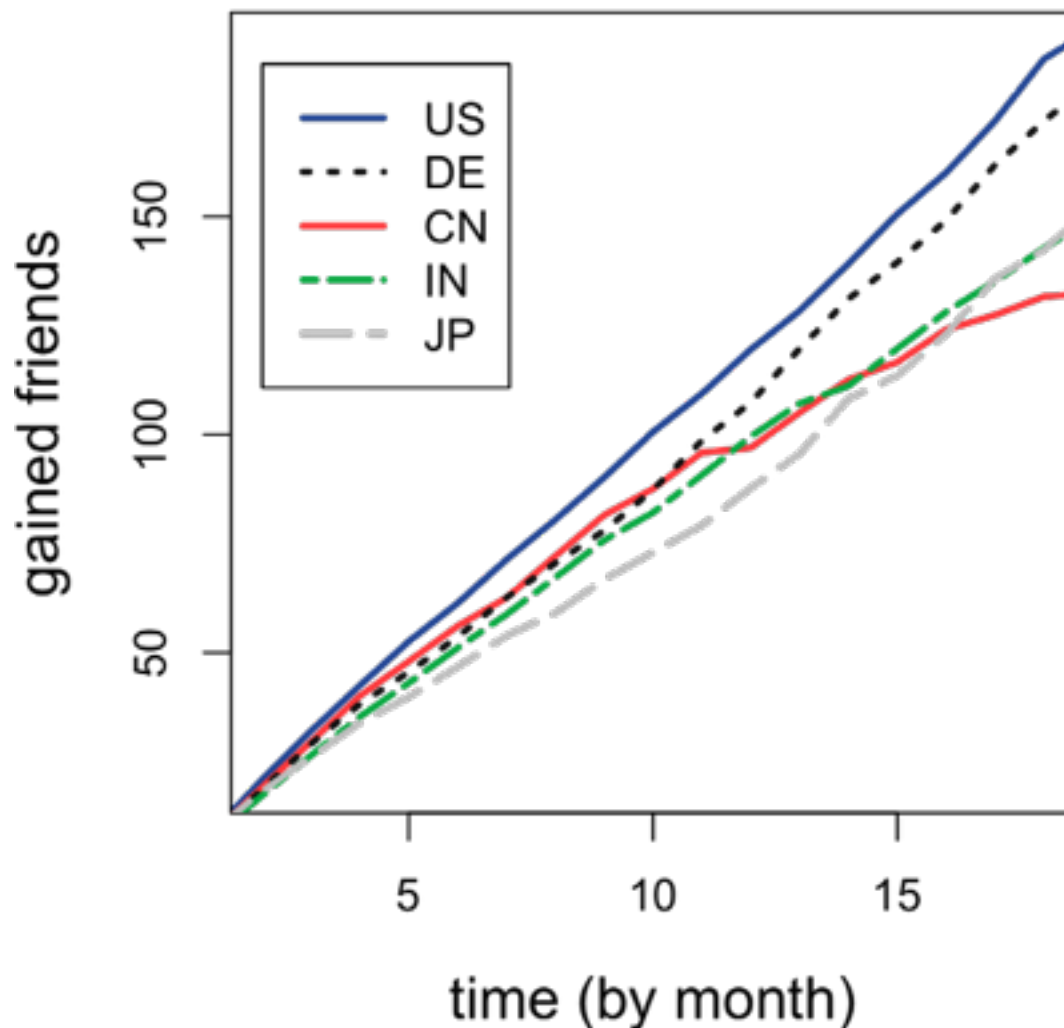


sentiments in conversations

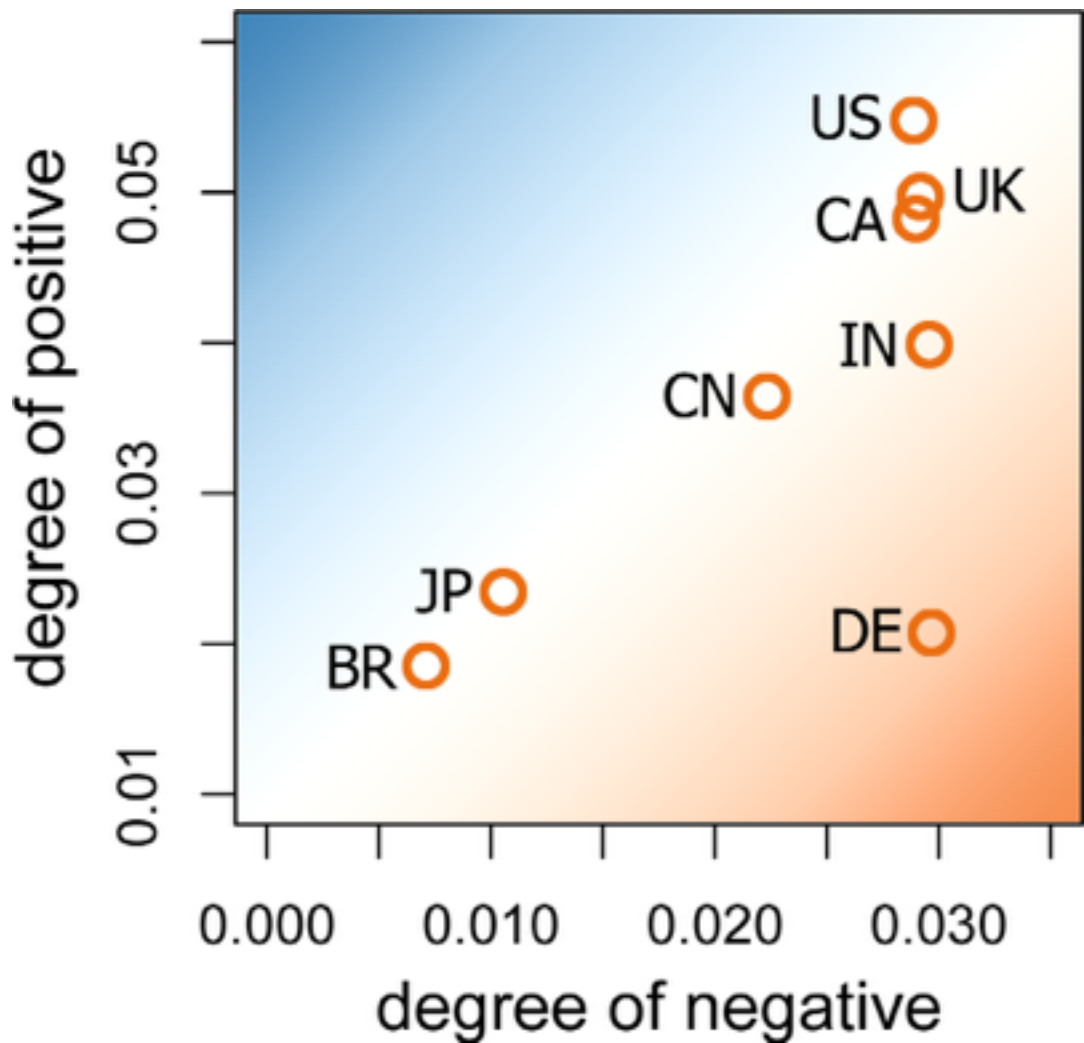
## Preferences of CMC Tools

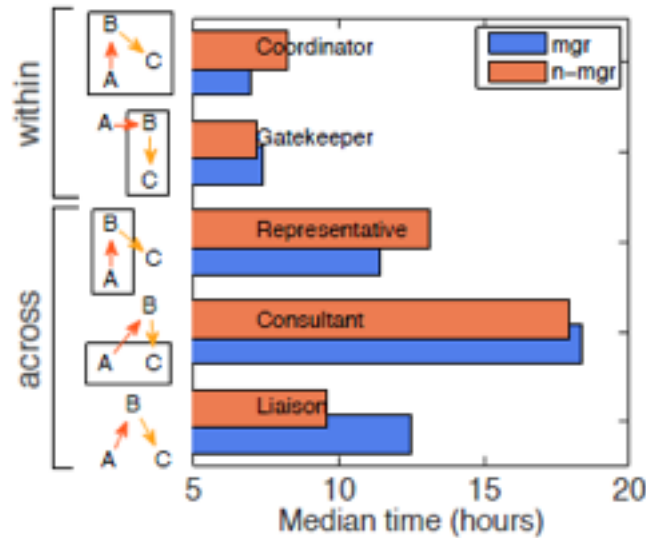
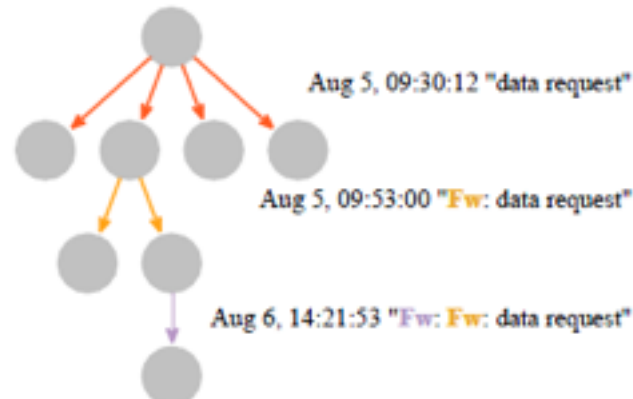


## Growing one's Social Networks



## Sentiments in Conversation



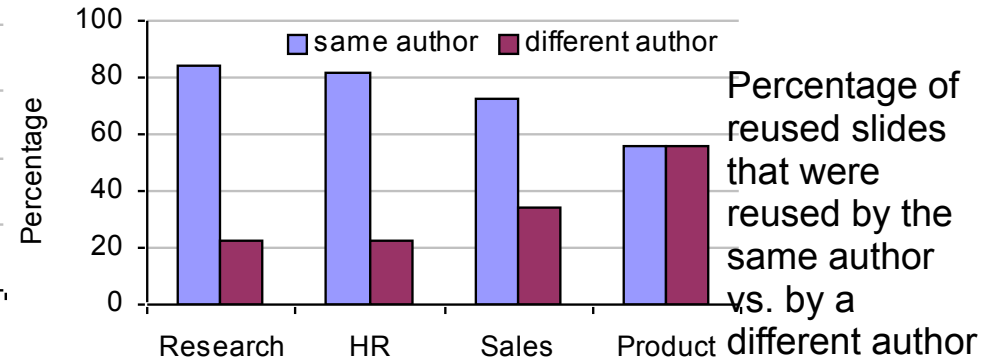


Role difference of normal behavior

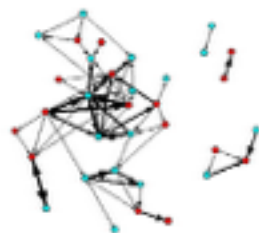
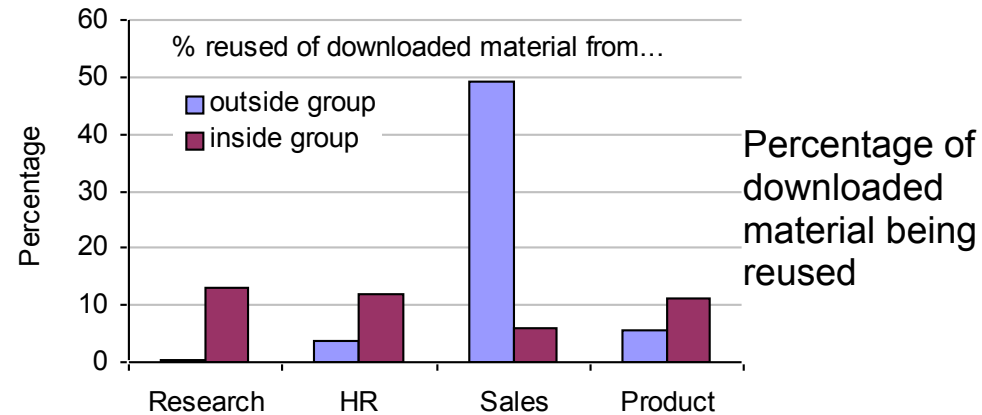
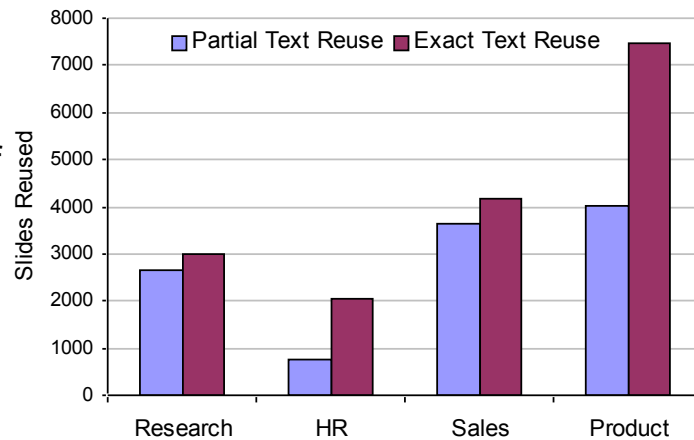


# Information Reuse Behavior (CHI '11)

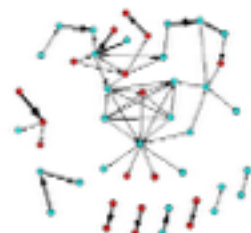
Percentage of slides with reused content



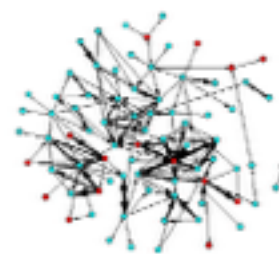
Number of slide pairs with exact vs. partial text reuse



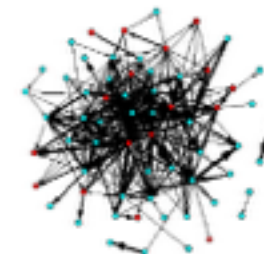
(a) Research



(b) HR



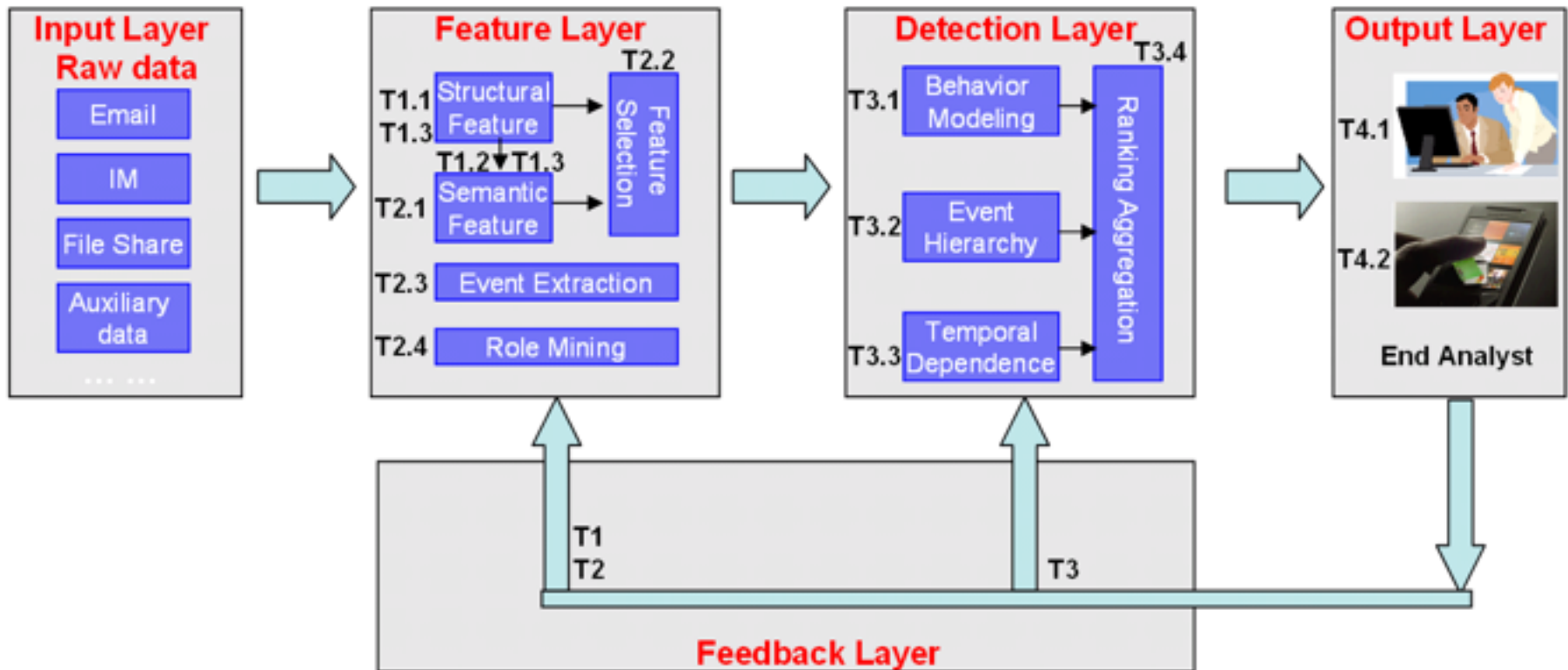
(c) Sales



(d) Product

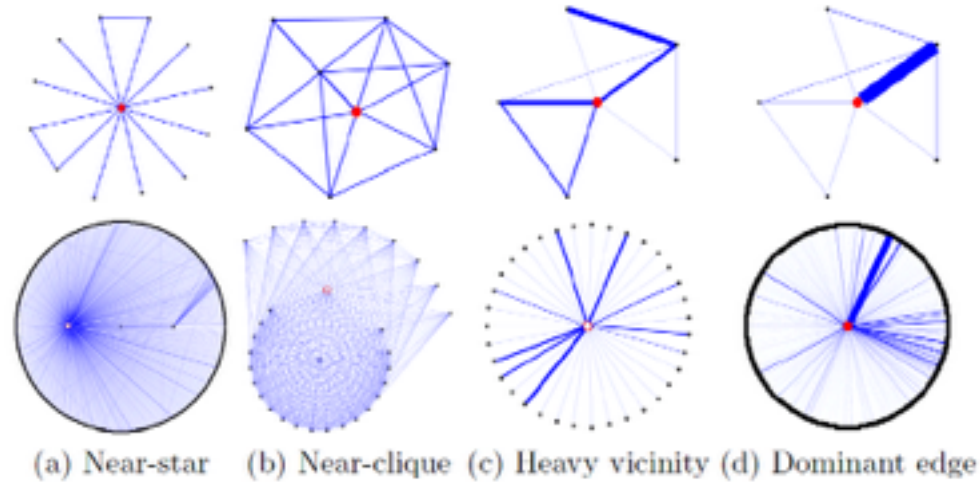
# Behavior Detection

- **Overall Flowchart:** Network Science + Machine Learning + Role Mining + Visualization

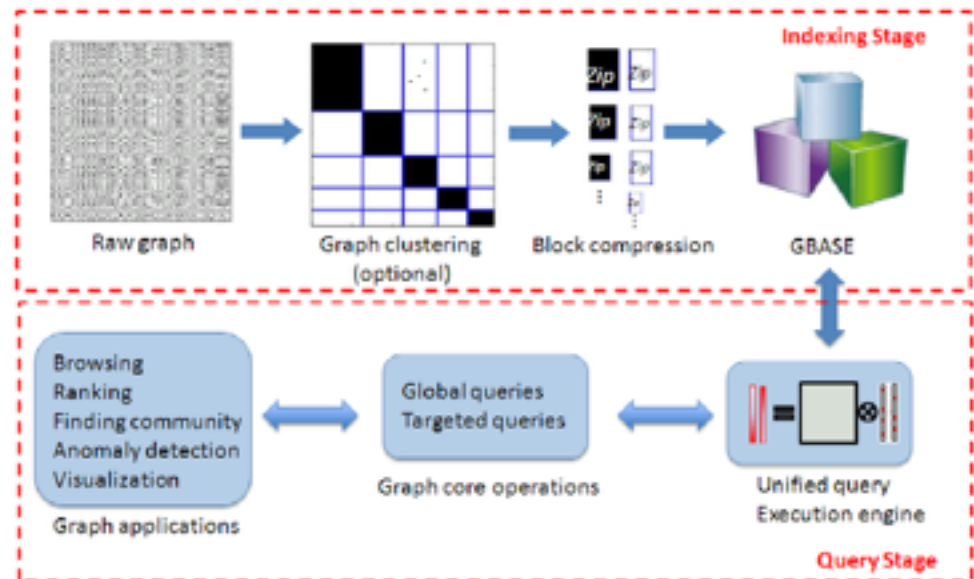


# Anomaly Detection – algorithms and infrastructure

- **Thrust 1:** Anomaly Detection Algorithms
  - New algorithms to detect abnormal humans (nodes) as well as abnormal contacts (edges) from social networks.
  - Explore the structure feature and incorporate content (semantic) features.
  
- **Thrust 2:** Anomaly Usability
  - Address the ‘lack-of-the ground-truth’ issue by
    - (1) Interpretation friendly properties (e.g., non-negativity, sparseness, etc) into the current anomaly detection matrix factorization; and
    - (2) providing some concise summarization to perform anomaly attribution.
  
- **Thrust 3:** Infrastructure Support
  - General and scalable graph/network management system to process large

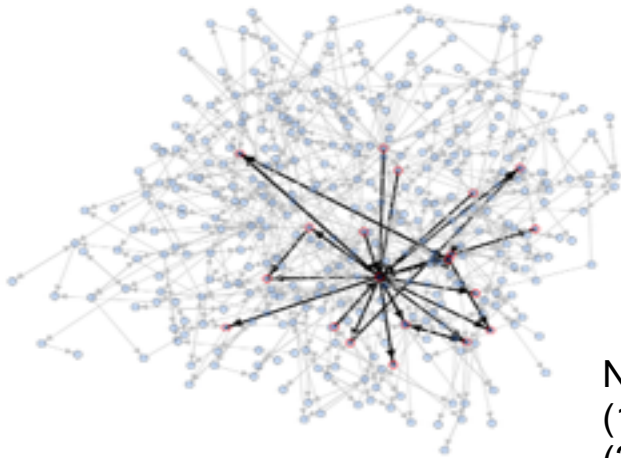


Typical abnormal nodes and their local ego-net structures

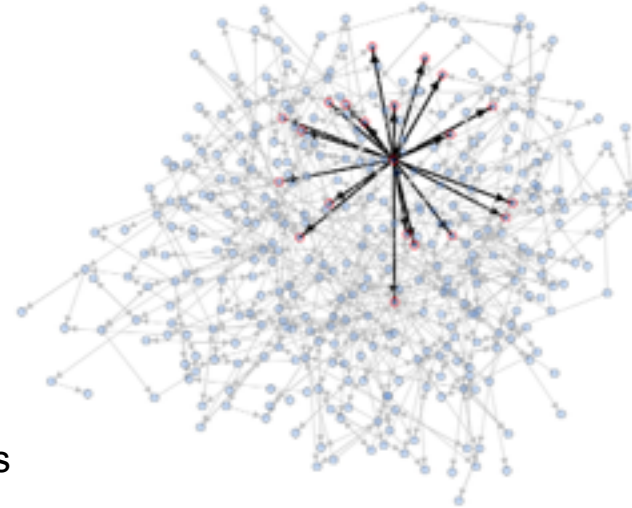


The overall flowchart of the graph management system

# Use Case: Utilizing Social Network Analysis for Spam Detection

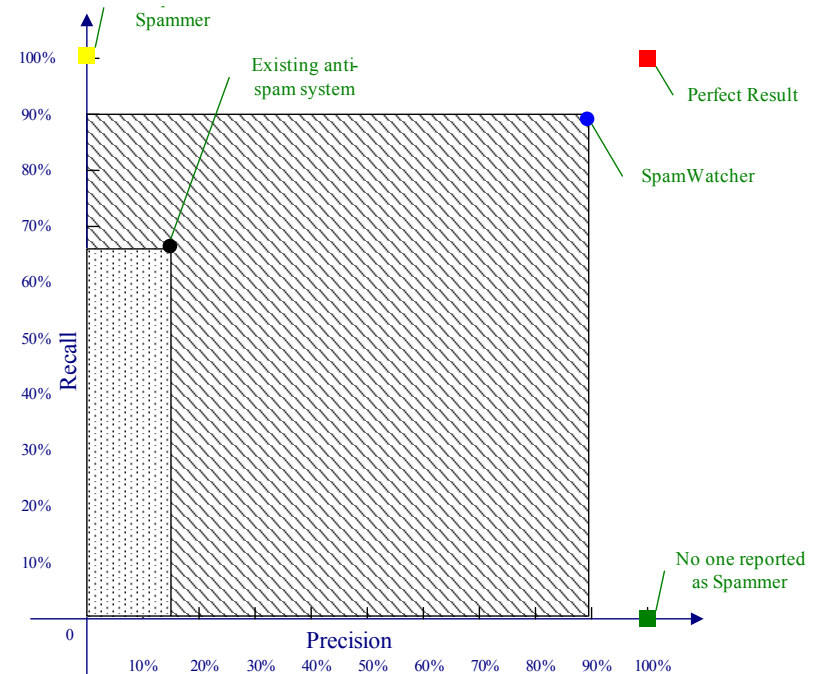


Normal:  
 (1) Clique-like  
 (2) Two-way links

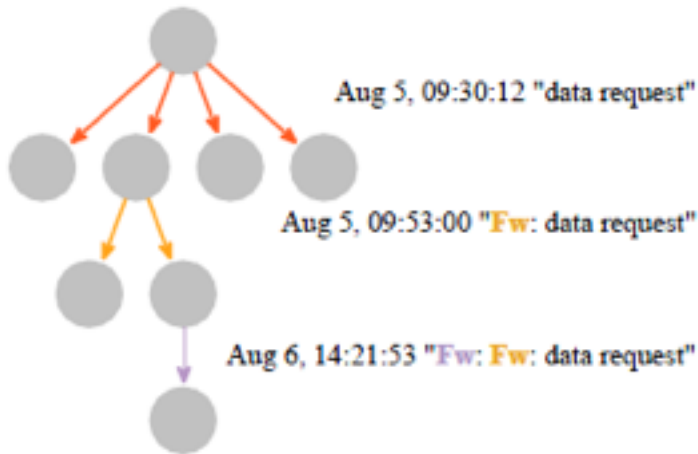


Spamming:  
 Near-Star

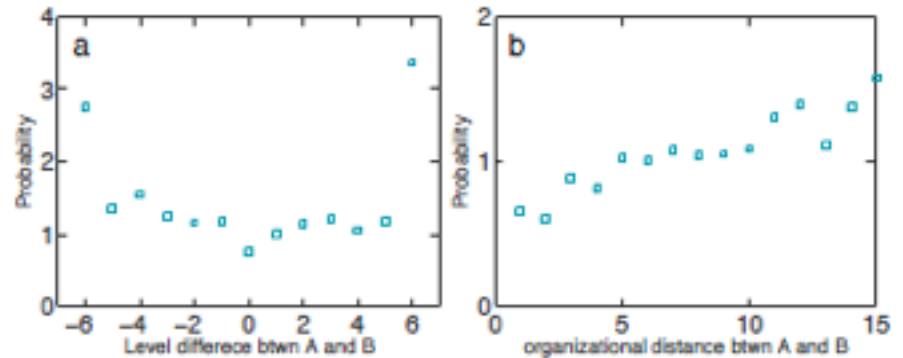
- A pilot project was done by CRL in a telecomm area of 6 million users in 2009.
- In experiment
  - Social Network Analysis is with recall of 89.97% and precision of 88.17% while comparison system is with 66.77% recall and 14.85% precision.
  - SNA’s precision/recall area is 8 times larger



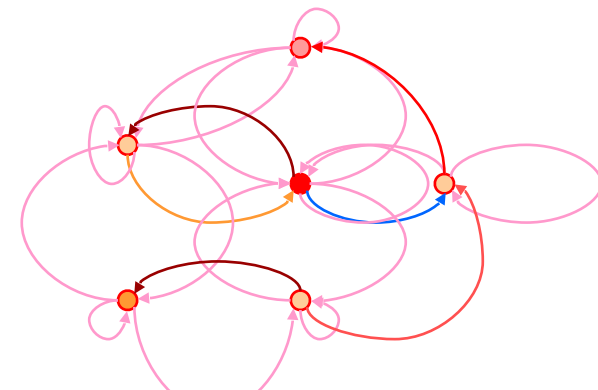
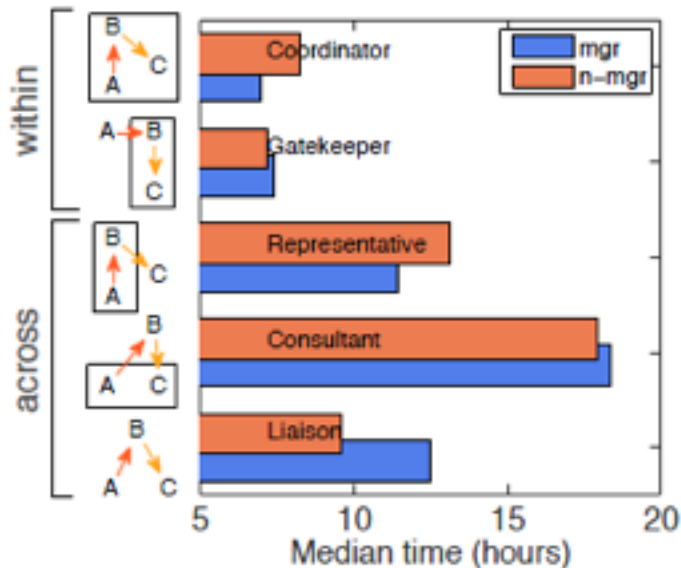
# Anomaly Detection – information flow-based approach



An illustrative example of an information spreading tree. This tree is of size 8, width 4, depth 3.



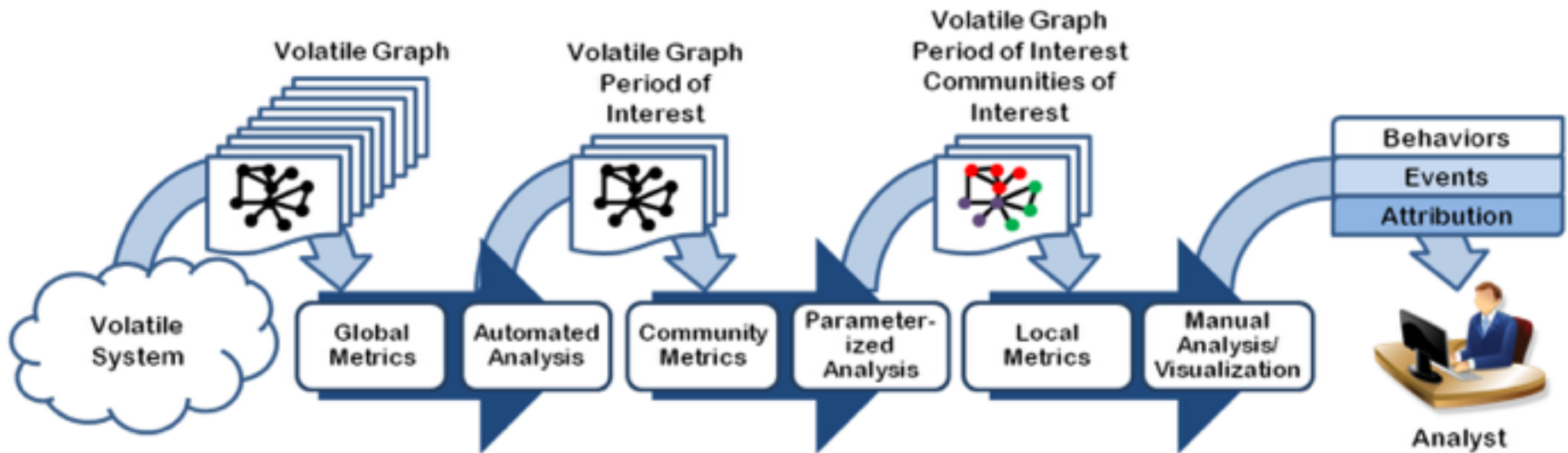
Probability ratio of email forwarding as a function of (a) hierarchical level difference and (b) organizational distance between initiators and receivers. The information spreading exhibits some homophily effect.



Video demo: <http://smallblue.research.ibm.com/demos/>

# Large-Scale Graph Analysis

- T 1.1: Structure Feature Extraction
- T 1.2: Initial Filtering
- T 1.3: Scalability

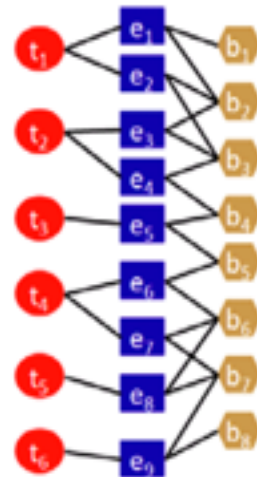




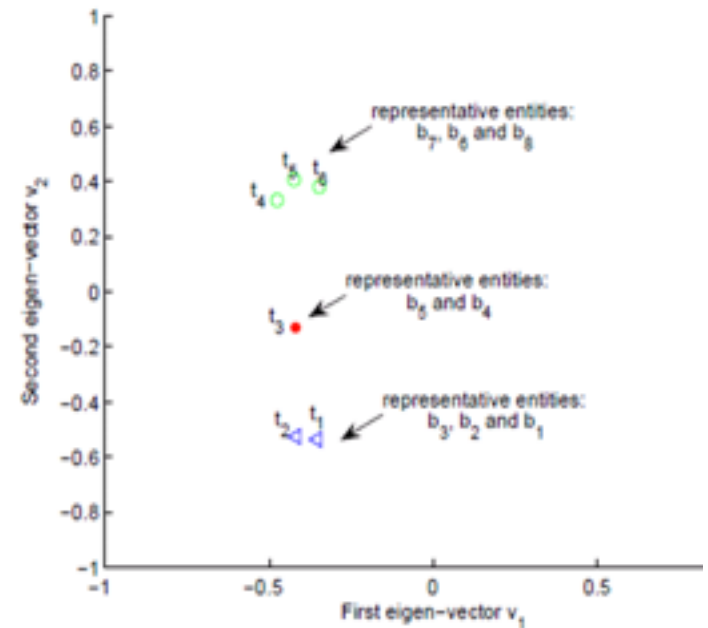
# Time-to-Time (T3) Proximity Matrix Analysis

Time Steps	Events	Entities
$t_1$	$e_1$	$b_1, b_2$
	$e_2$	$b_2, b_3$
$t_2$	$e_3$	$b_2, b_3$
	$e_4$	$b_3, b_4$
$t_3$	$e_5$	$b_4, b_5$
$t_4$	$e_6$	$b_5, b_6$
	$e_7$	$b_6, b_7$
$t_5$	$e_8$	$b_6, b_7$
$t_6$	$e_9$	$b_7, b_8$

(a) Input of T3



(b) Graph Representation

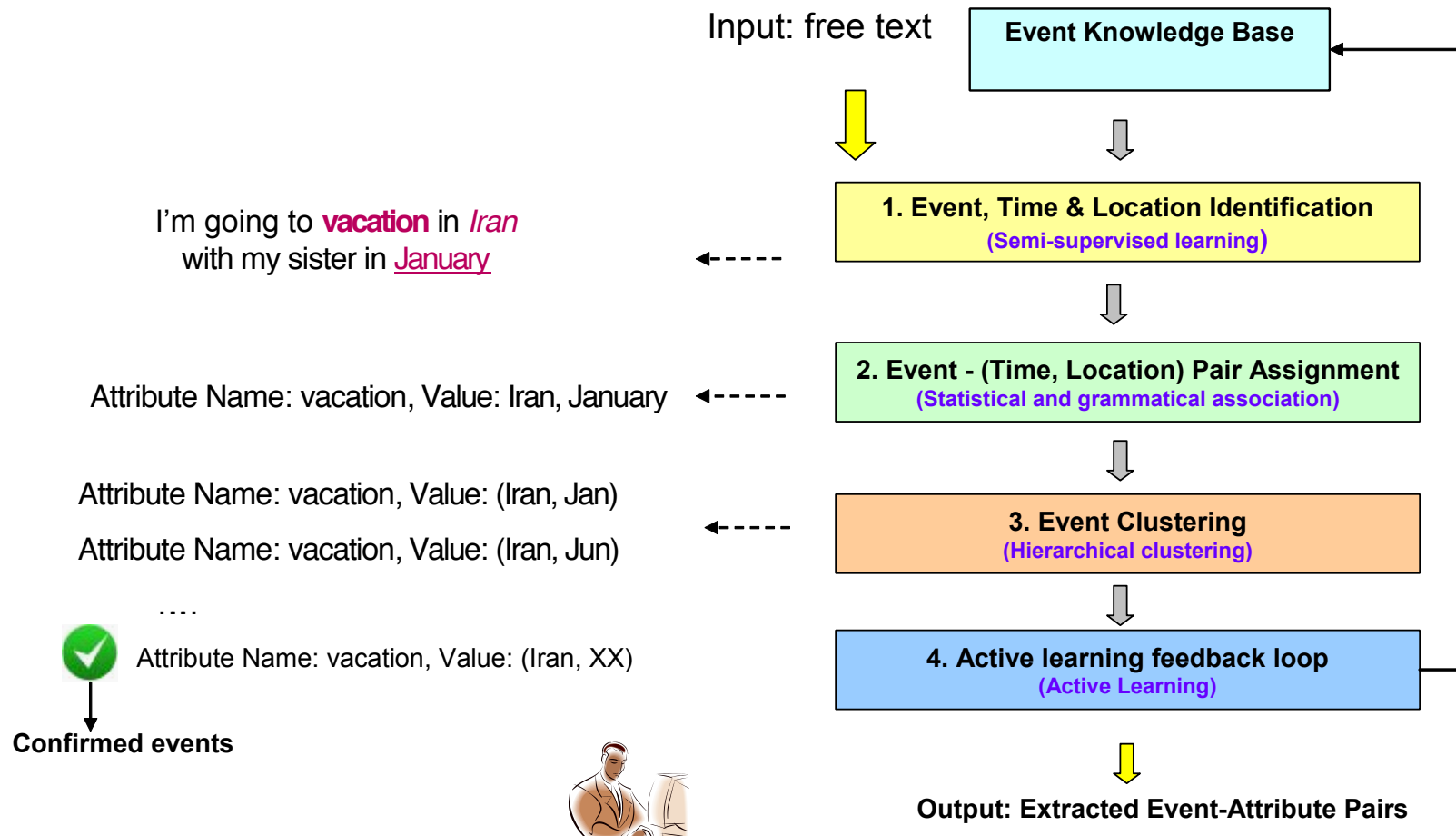


(c) Output of T3 (at finest level)

# Thrust 2: Semantic Level Analysis

- T 2.1: Semantic Feature Extraction
- T 2.2: Feature Augmentation and Selection

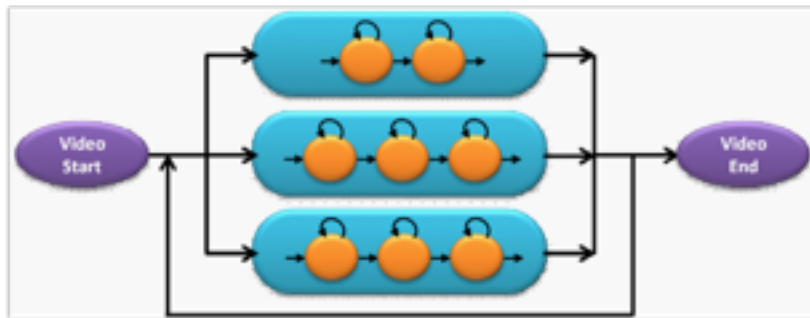
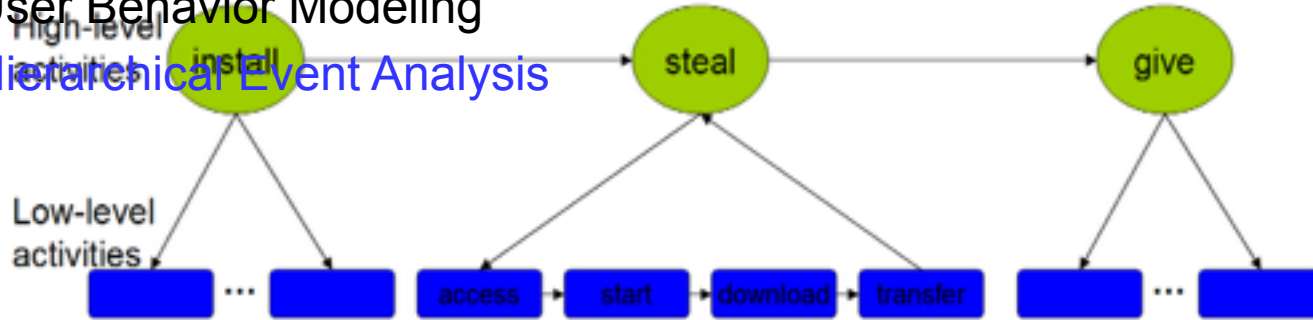
- T 2.3: Abnormal Event Extraction
- T 2.4: Role Mining and Analysis



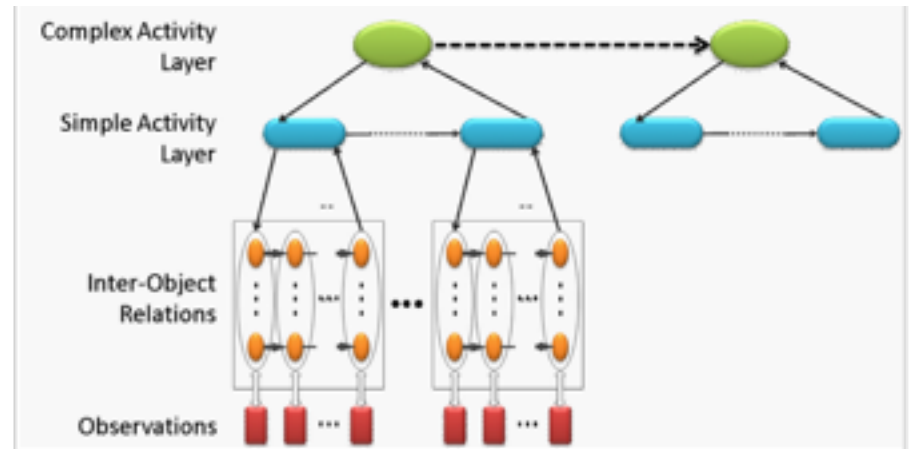
# Thrust 3: Ranking and Aggregation

- T 3.3: Temporal Dependency Analysis
- T 3.4: Anomaly Aggregation

- T 3.1: User Behavior Modeling
- T 3.2: Hierarchical Event Analysis

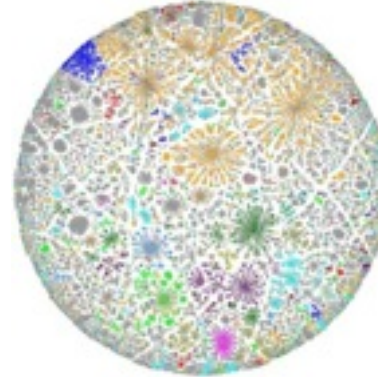


(a) Combined Composite Activity Model

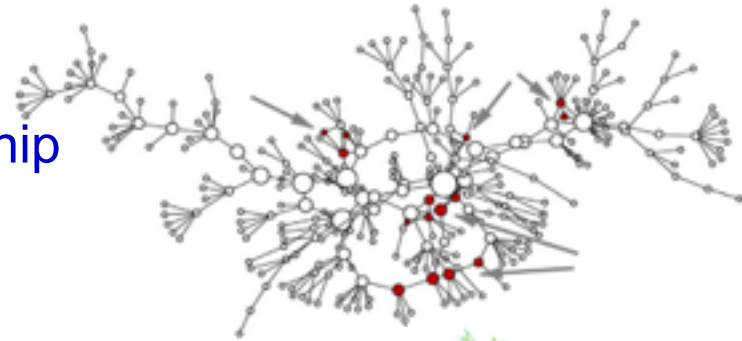


(b) Unrolled Composite Activity Model

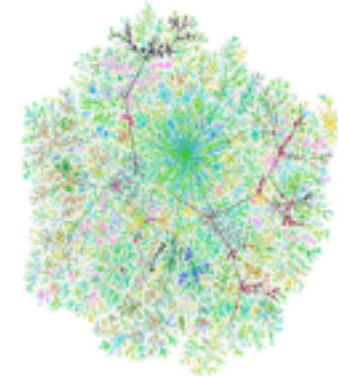
Example 1: Internet Map  
Nodes: ISPs; Edges: Connection  
**(33K Nodes, 290K edges)**



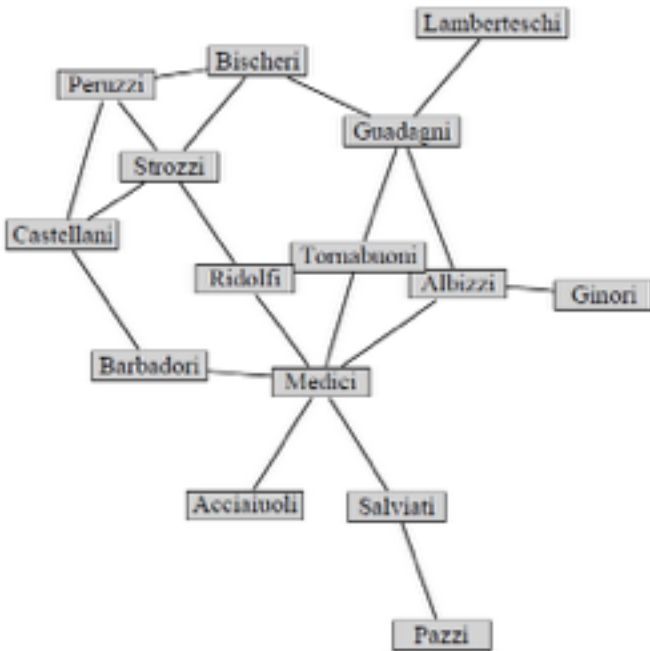
Example 2: Social Network  
Nodes: People; Edges: Friendship  
**(FaceBook has 500M+ Users)**



Example 3: Web Graph  
Nodes: Web Pages; Edges: Hyperlinks  
**(Yahoo Web: 1.4B nodes, 6.6B edges)**



Multiple Scales, Multiple Disciplines



[15th Century Florentine Family]

$|V| = 15$

$|E| = 19$

**Degree : Easy**

**Closeness : Easy**

**Betweenness : Easy**

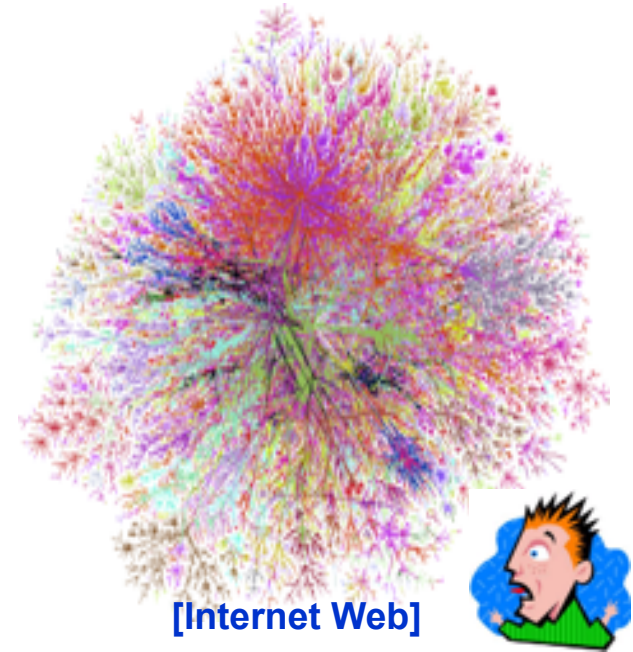
**“Who are the most important actors?”**

**Three centralities**

- Degree: # of neighbor
- Closeness: avg. shortest path length
- Betweenness: # of times a node sits between shortest path

**Application**

- Measuring the financial company value
- Network attack monitoring



[Internet Web]

$|V| = \text{Billions}$

$|E| = \text{Billions}$

**Degree : Easy**

**Closeness : Hard**

**Betweenness : Hard**

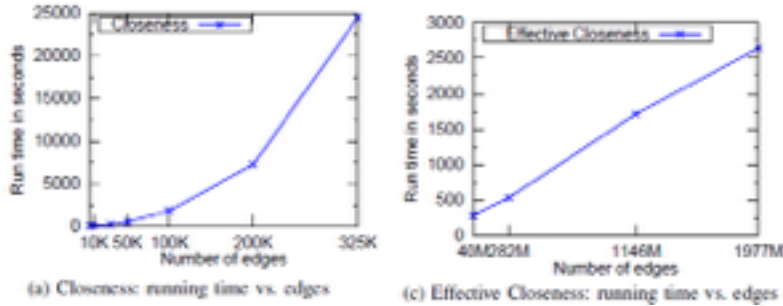
$O(|E|)$

$O(|V|^3)$

$O(|V|^2 \log |V|)$

**For 2 Billion Edges,  
- standard closeness: 30,000 years**

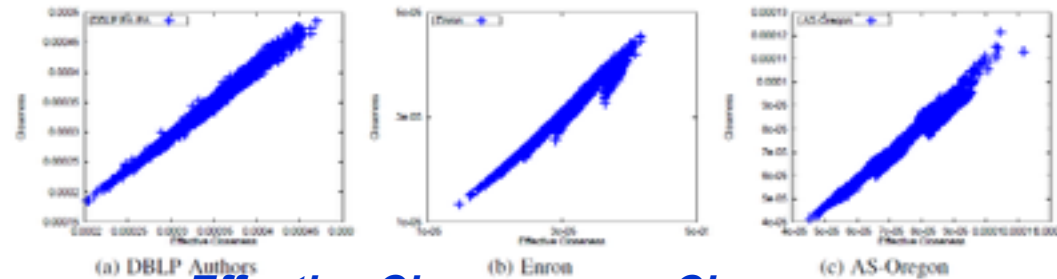
- Example -- we proposed two new centralities ('effective closeness' and 'LineRank'), and efficient large scale algorithms for billion-scale graphs.



## Scalability Results

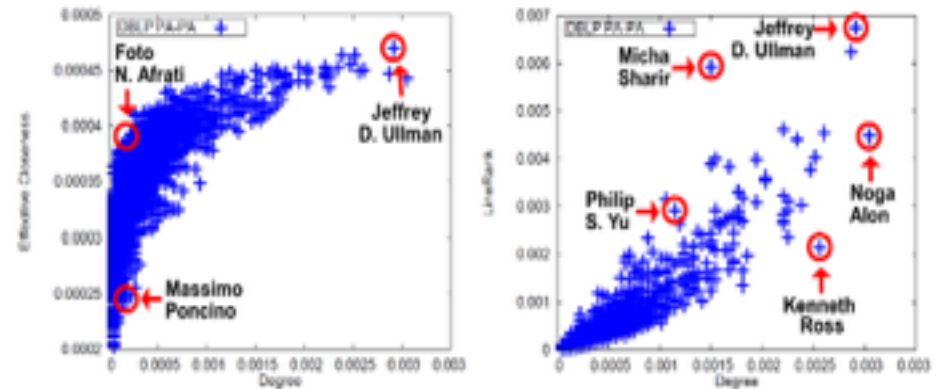
(Near-linear scalability)

**For 2 Billion Edges,**  
 - standard closeness: 30,000 years  
 - effective closeness: ~ 1 day!  
**1,000,000 times faster!**



## Effective Closeness vs. Closeness

(Near-linear correlation ( $\geq 97.8\%$ ))



## Analysis of Real-World Graph



# Privacy – cryptography and key management approach (CIKM '11)

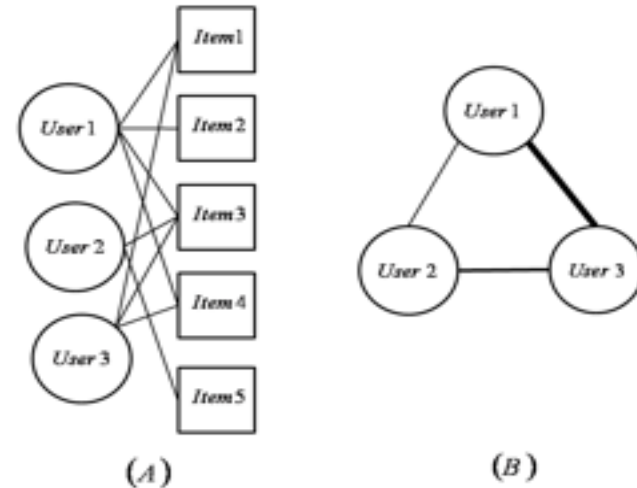
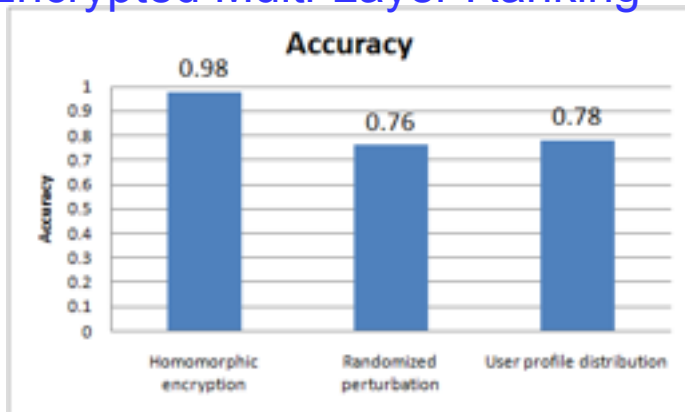
- A novel methodology & system for data mining and content/people recommendations
- New cryptographic method:
  - Polynomial Ring Homomorphism, derived from Lattice-Based Cryptography
  - Encrypted domain:

Addition

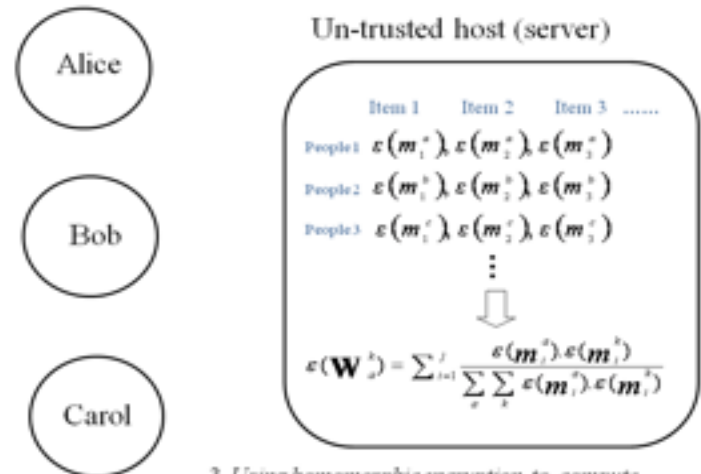
Multiplication

Division

- Key management protocol for:
  - Encrypted Multi-Layer Ranking



Soft clustering for recommendations



3. Using homomorphic encryption to compute similarity in encrypted domain, a subgroup of people with similar preference was selected and recommend.

Encrypted domain computation

# Social Network Analysis for Marketing and Sales

## •What is the problem?

- Past studies showed significant success on utilizing social relationships for sales & marketing.
  - McKinsey (2009) surveyed 190+ firms in all industry sectors utilizing ‘social selling’:
    - the transaction increase consideration by avg. 19%;
    - the average yield increase conversion by avg. 17%.
  - Krackhardt (Carnegie Mellon U, 2005) showed that companies with strong informal networks perform 5 or 6 times better than those with weak networks.
  - Brydon (VisiblePath, 2006) showed that the performance gains of companies utilizing relationships are 16x in sales; 4x in marketing; and 10x in hiring
- How to utilize Social Network Analysis for Marketing and Sales?

## •What is the solution?

- Conduct social graph analysis, human capital analysis, and economic analysis to quantify micro- and macro- social capital of each company (B2B) or each individual (B2C).
- Large-scale Data Mining for social capital calculation through distributed social sensors, sales records, communications, web & social media activities, etc. .
- Inject historical leads and sales records to train machines to associate casualty of social capital and economic gains..
- Optimize collective social & human capitals for marketing strategies and team forming.

## •What are the related assets in Smarter Commerce solutions stack?

- Unica Leads, NetInsight, Detect, CustomerInsight, and PredictiveInsight
- CoreMetrics Continuous Optimization Platform

## •What remains to be done?

# An example of utilizing micro- and macro social capital

- Who among IBMers are the closest to McKinsey? What is the shortest path for me to reach McKinsey through my colleagues? Who should join the team for McKinsey to send this kind of marketing message?
- How strong is IBM, in terms of relationship strength, to other companies?



# Flow of B2B Marketing using Social Network Analysis and Optimization

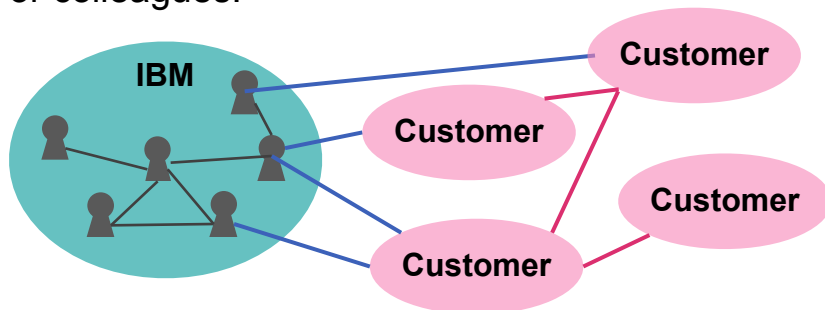
## Social Network Analysis

### 1) Social Network Mining

- Among customer companies
  - From web, news articles, stocks, etc.
- Between employee and customers
  - From intranet and sales force data
- Among employee
  - Already available in SmallBlue

### 2) Quantifying Social Capital

- Calculate the economic value of each person's social network.
- Analyze each person's influence on customers or colleagues.



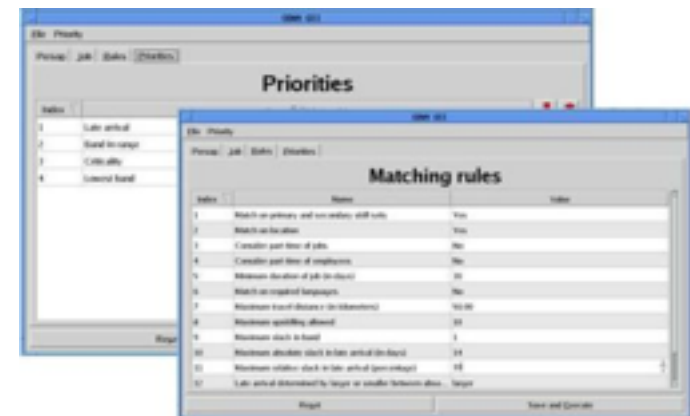
## Optimization

### 3) Define Constraints for Matchmaking

- B2B (customer & customer), employee & customer
- Condition lists for desired team should be determined based on survey for the sales force.

### 4) Optimize Matches based on Constraints

- Obtain optimal or near optimal matches using Constraint Programming.



# Use Case of B2B Sales based on SNA and Optimization

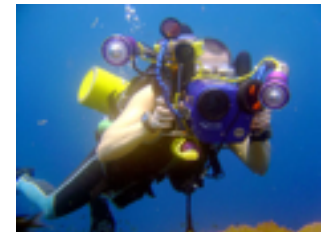
## ■ Main Steps

1. Extract a set of solution candidates (team forming) by social network analysis (SmallBlue).
2. Specify solutions that satisfy constraints by optimization from a set of solutions.

## ■ Example: Waterproof camera manufacturer

- In the past, IBM succeeded to consult company A's to sell their waterproof camera to Best Buy.
  - Which one was the best selling model?
  - What kind of strategy lead the project to be succeeded?
- Now we have a scuba diving school B who wants to buy waterproof camera for their classes as e-marketing customer.

It is a chance to match company A and diving school B.



- Here, this problem is defined as follows:

- ➔ 1. Find social relationships between company A and diving school B.
  - Situation of relationship between A and B would be cleared by social network mining and analysis on news articles, stock markets, blogs and so on.
2. Find following experts from IBM using SmallBlue's social mining techniques.
  - 1 person who knows company A, 1 person who knows diving school B and 1 person who knows how to process campaign in gym, tennis school, etc.
3. Specify appropriate people as a sales team using Tonkawa's optimization techniques.
  - The team members are determined from a set of experts extracted by SmallBlue based on constraints like each person's schedule, skills, expected level of their contribution, relationship among team, etc.

If we could know about available data, more scenarios would be produced.

## Major technical steps for SNA for Marketing and Sales

- Make innovate Sales & Marketing software prototype (especially for B2B)
  
- Technical Approach
  - Foundation:
    - Privacy-Preserving Large Scale Data Mining
    - Large-Scale Network Analysis
    - Large-Scale Graph Management, Storage, Index and Retrieval
    - Large-Scale Optimization
    - Quantifiable economic and financial analysis for sales & marketing optimization strategy on graphs and networks
  - Applications:
    - Quantifying Social Capital of Customer Companies and their people:  
Finding social networks inside and outside companies by extending SmallBlue mining technology
    - Converting Social Capital into Economic Gain for B2B Marketing & Sales:  
Which employee has the shortest ‘social path’ to reach a customer company, or a specific person in a customer company? Who is the right person to send match message to customer?  
Finding matches between customer companies or between customer and employee team based on their social capitals and constraints using optimization technology
  
- Major Research Challenge:
  - First prototype system to quantize Social Capital, and utilize it for B2B Marketing & Sales
  - Significant amount of new system design, social & economic analytics and optimization techniques.

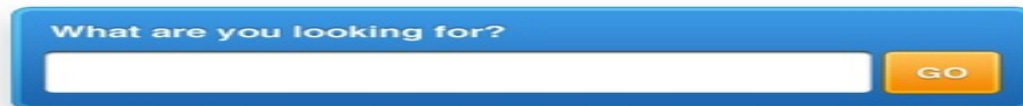


# Relational Term-Suggestion

Q.

*What keywords should I put in the search box to get the information I really want?*

amazon.com

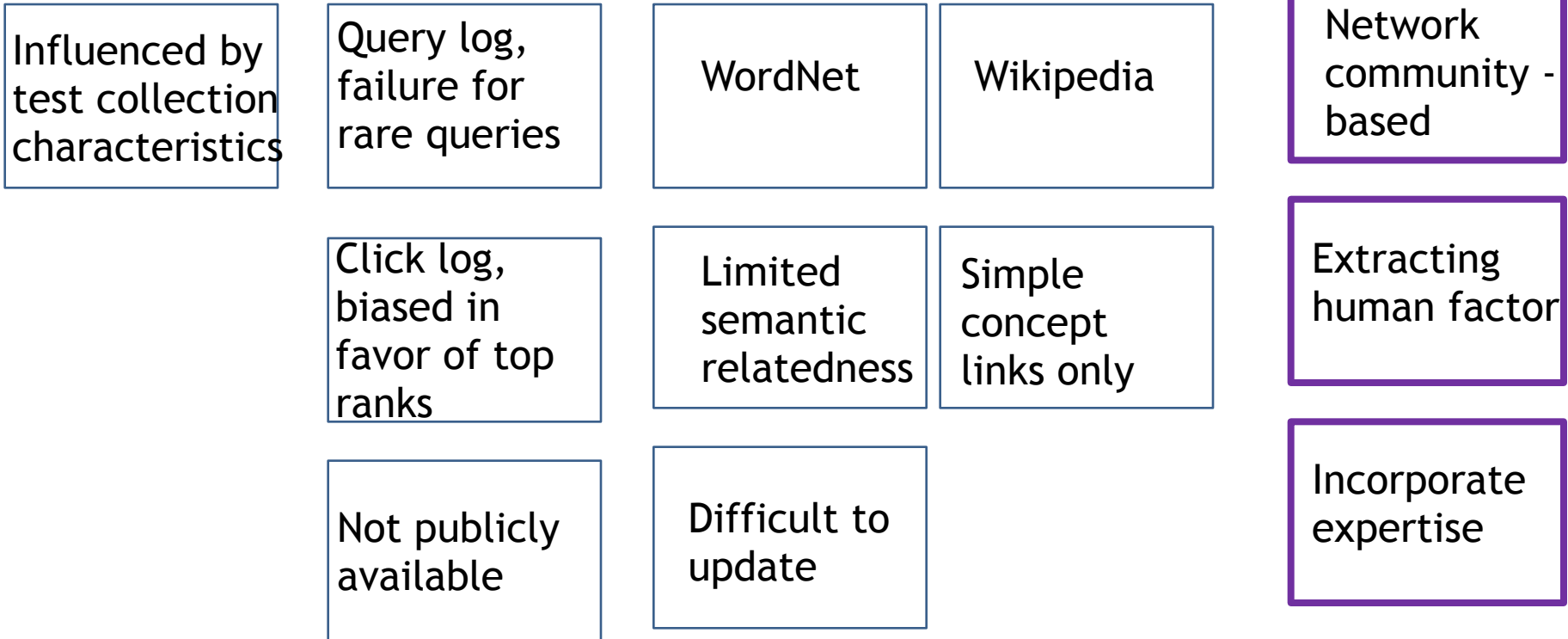
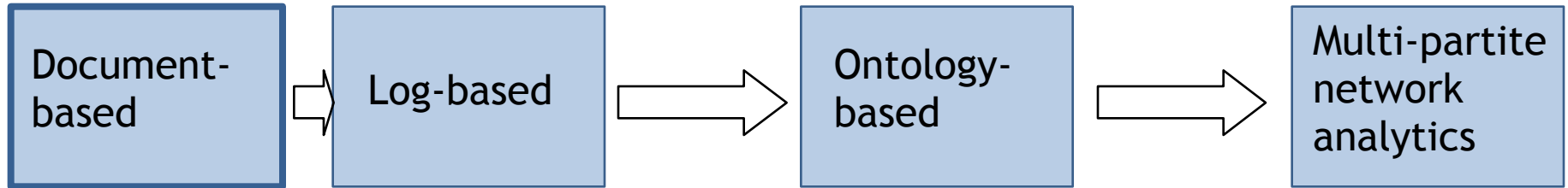


What are you looking for?

GO

# Multi-partite Network Analytics

## Term Suggestion and Query Expansion



# Document-based

- Influenced by test collection characteristics
- No consideration of key terms that are highly semantically related but do not frequently co-occur.



apple juice  
apple tree



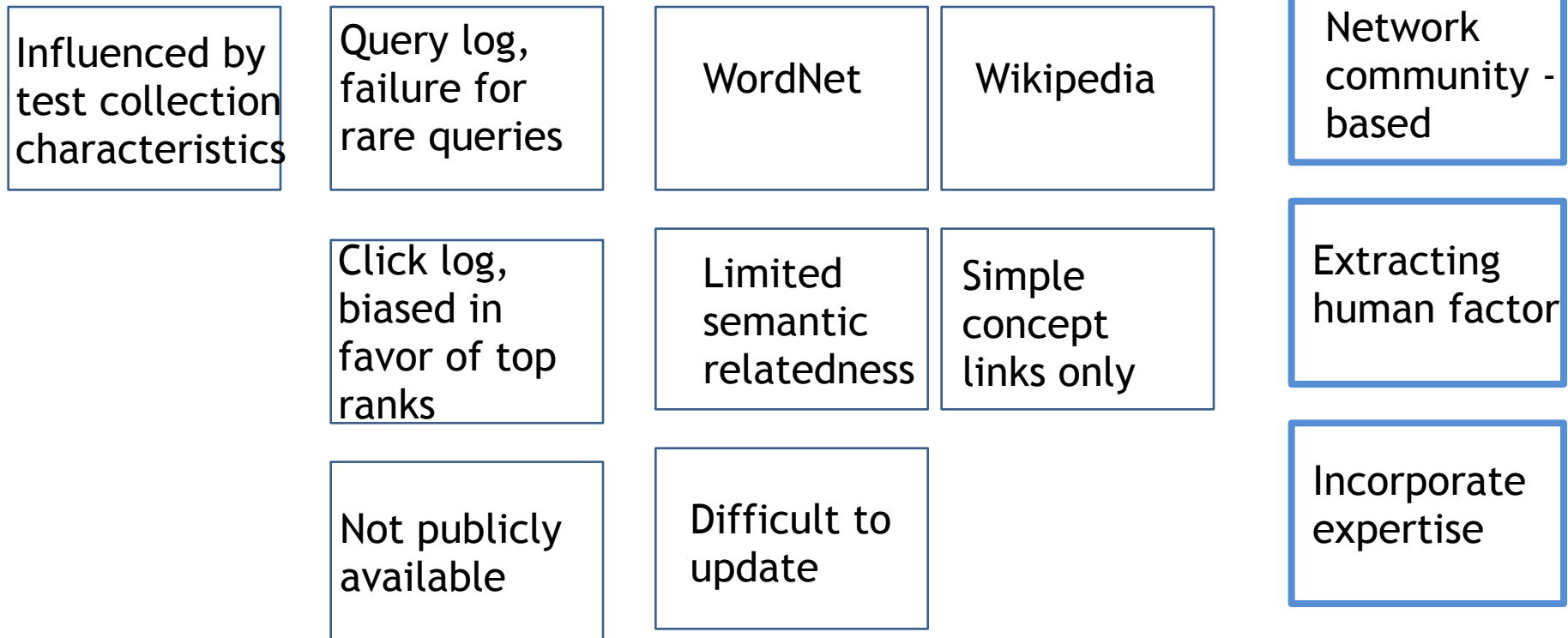
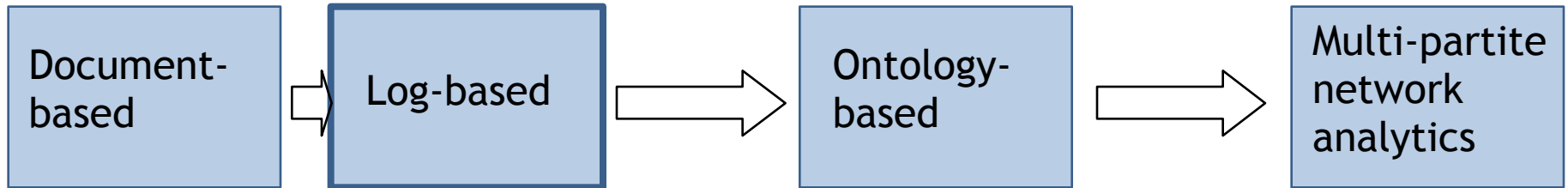
apple store  
apple TV



Kim, M. AND Choi, K. A. 1999. Comparison of collocation-based similarity measures in query expansion. *Information Processing and Management* 35 (1999), 19-30.

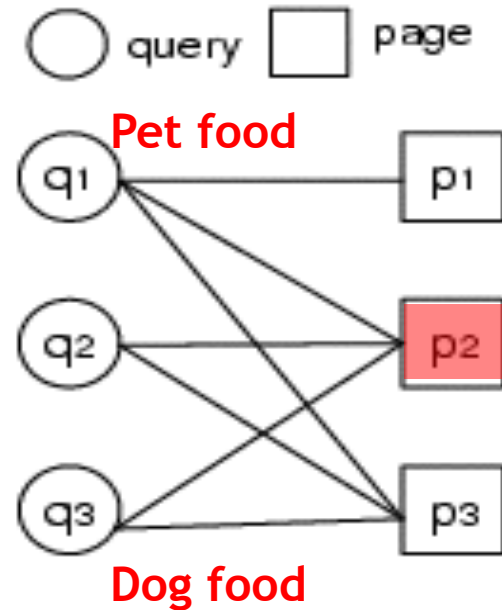
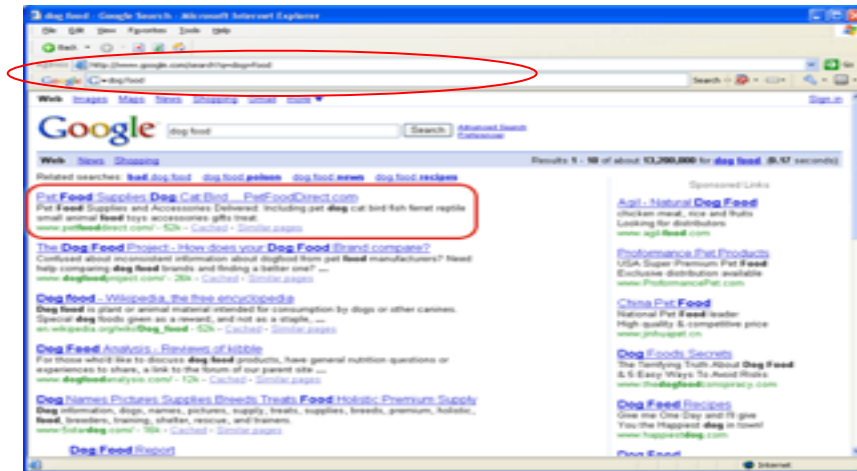
# Multi-partite Network Analytics

## Term Suggestion and Query Expansion



# Log-based

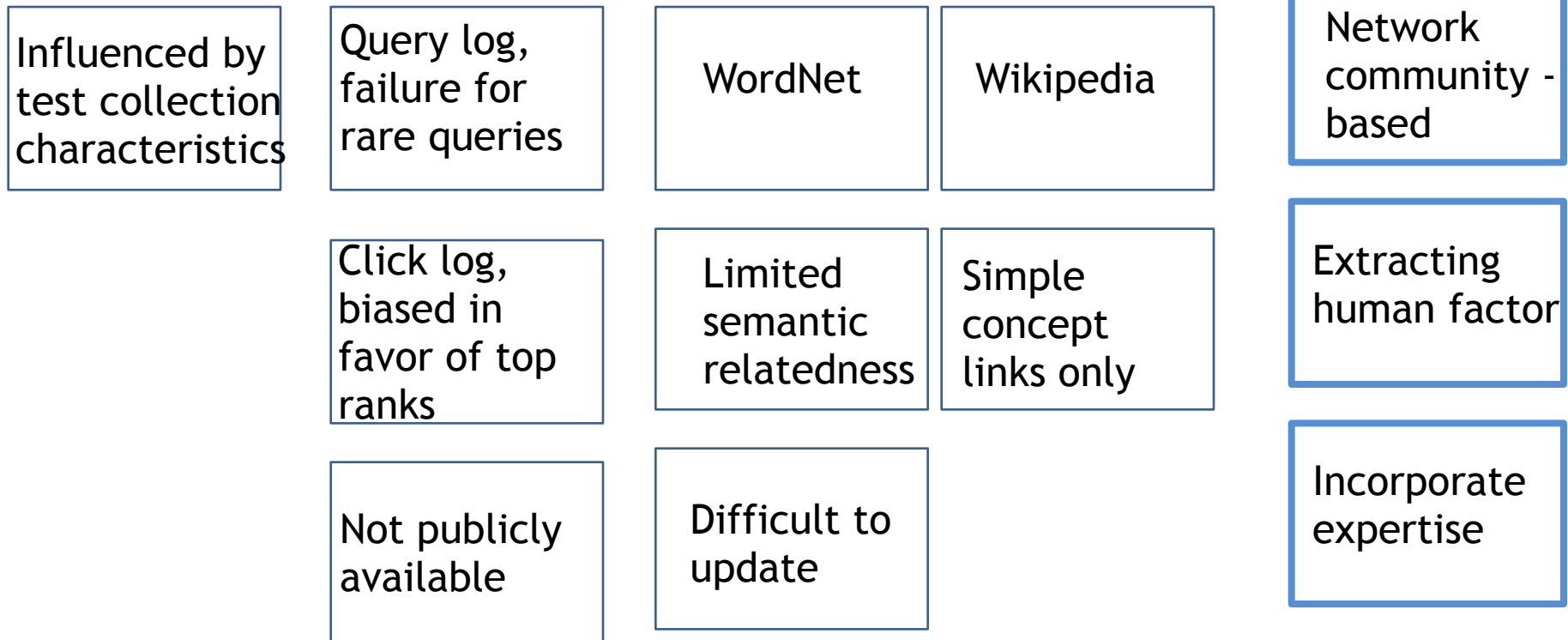
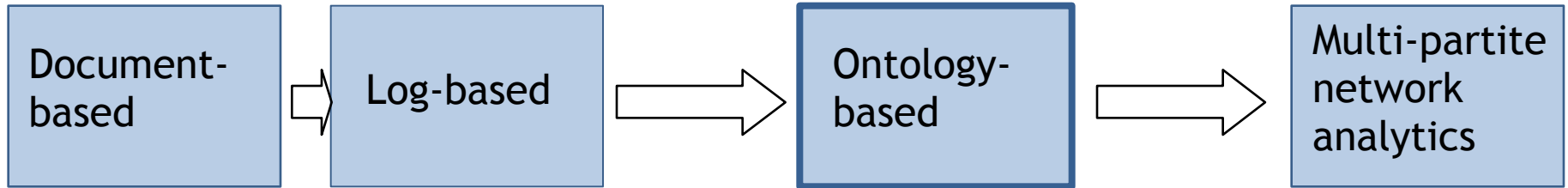
- Cluster queries with similar clicked URLs
- Identifying the mapping between queries and clicked URLs



BAEZA-YATES, R., AND TIBERI, A. 2007. Extracting Semantic Relations from Query Logs. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2007)*, 76-85.

# Multi-partite Network Analytics

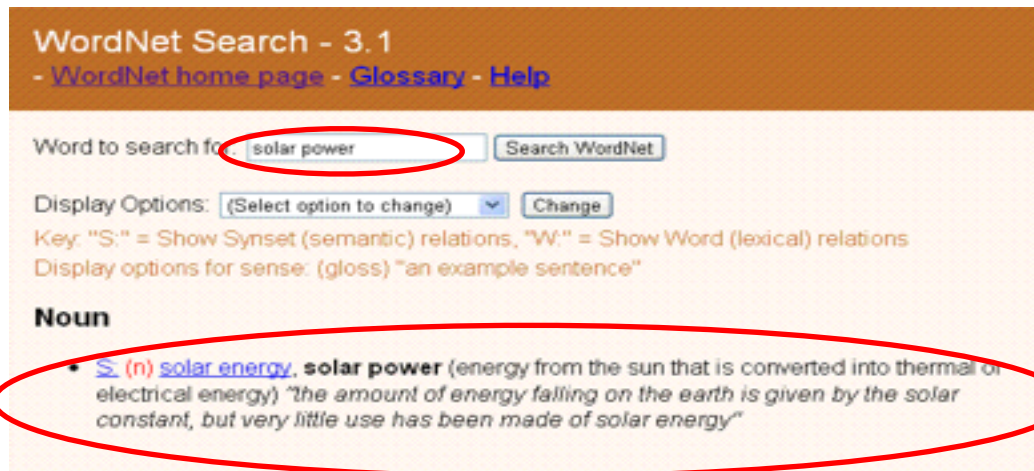
## Term Suggestion and Query Expansion





## WordNet as Ontology

- Manually constructed system based on individual words benefit will be limited
- System is not easily updated



Pedersen, T, Patwardhan, S and Michelizzi, J. "WordNet::Similarity - Measuring the Relatedness of Concepts" 2004 In *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI-2004)* pp. 1024-1025.

# Wikipedia as Ontology



The screenshot shows the Wikipedia article for "Solar power". The page layout includes a left sidebar with navigation links, a main content area with text and a table of contents, and a right sidebar with an image and a related article link.

**WIKIPEDIA**  
The Free Encyclopedia

Main page  
Contents  
Featured content  
Current events  
Random article  
Donate to Wikipedia

Interaction  
Help  
About Wikipedia  
Community portal  
Recent changes  
Contact Wikipedia

Toolbox  
Print/export  
Languages

Article **Talk**

## Solar power

From Wikipedia, the free encyclopedia


*This article is about generation of electricity using solar energy. For other uses of solar energy, see [Solar energy](#).*

**Solar power** is the conversion of [sunlight](#) into [electricity](#), either directly using [photovoltaics](#) (PV), or indirectly using [concentrated solar power](#) (CSP). Concentrated solar power systems use lenses or mirrors and tracking systems to focus a large area of sunlight into a small beam. Photovoltaics convert light into electric current using the [photoelectric effect](#).<sup>[1]</sup>

Commercial concentrated solar power plants were first developed in the 1980s. The 354 MW [SEGS](#) CSP installation is the largest solar power plant in the world, located in the [Mojave Desert](#) of California. Other large CSP plants include the [Solnova Solar Power Station](#) (150 MW) and the [Andasol solar power station](#) (150 MW), both in Spain. The 214 MW [Charanka Solar Park](#) in India, is the world's largest photovoltaic plant.

**Contents** [\[hide\]](#)

- Applications
- Concentrating solar power



The [PS10](#) concentrates sunlight from a field of heliostats onto a central tower.

[Renewable energy](#)

---

## Wikipedia as Ontology

- Wikipedia is a web-based free encyclopedia that anyone can edit.
- The English Wikipedia edition
  - 2.4 million articles
  - 1 billion words.
- Wikipedia relies on the power of collective intelligence
  - by peer-reviewed approaches rather than the authority of individual.
  - high quality,
  - almost noise free.

## Previous Approaches

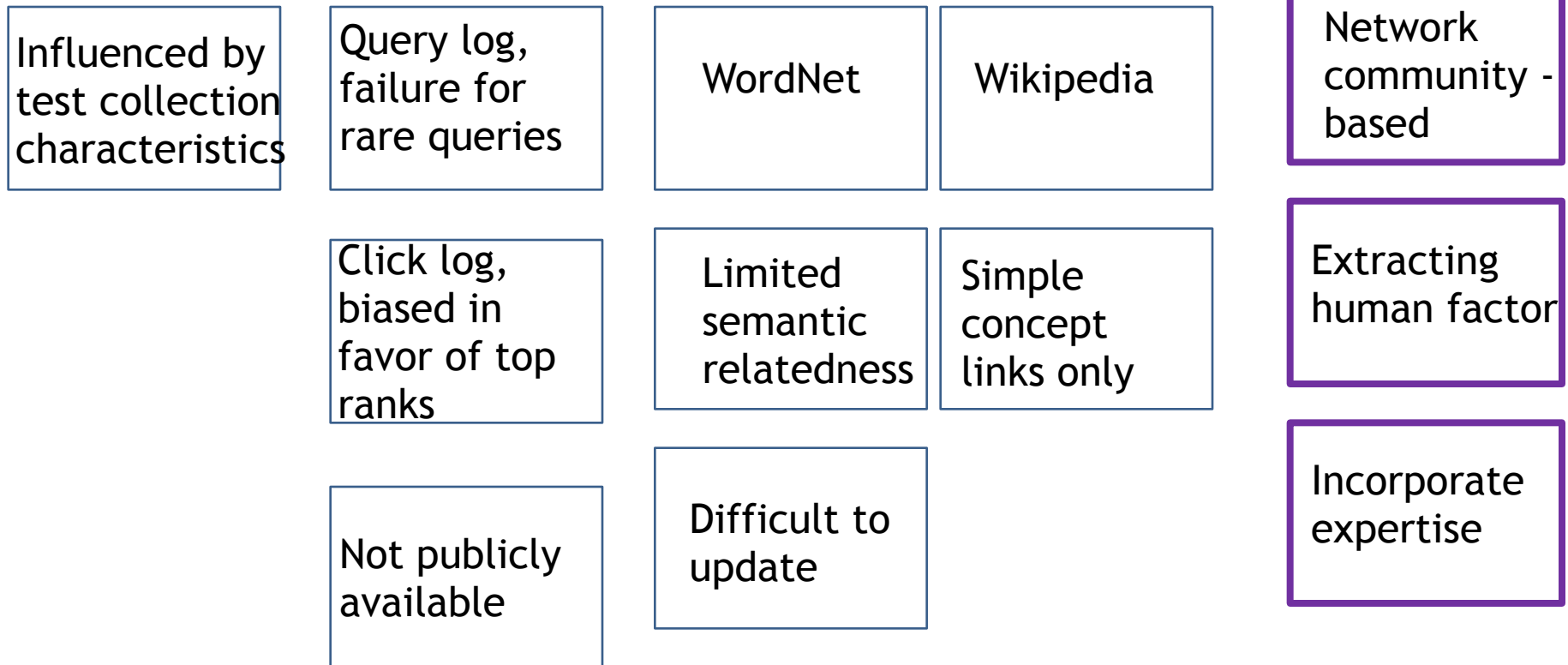
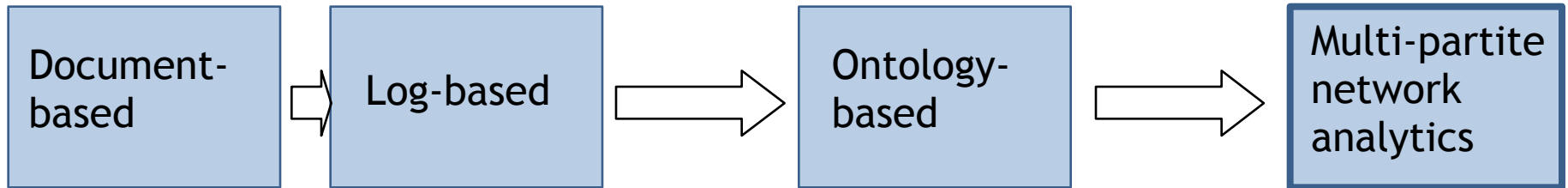
---

- Merely as an online dictionary and utilize it only as a structured knowledge database
- Using associated hyperlinks

MILNE, D., WITTEN, I. H., AND NICHOLS, D. 2007. A Knowledge-Based Search Engine Powered by Wikipedia. In Proceedings of the 16th ACM Conference on Information and Knowledge Management (CIKM 2007), 445-454..

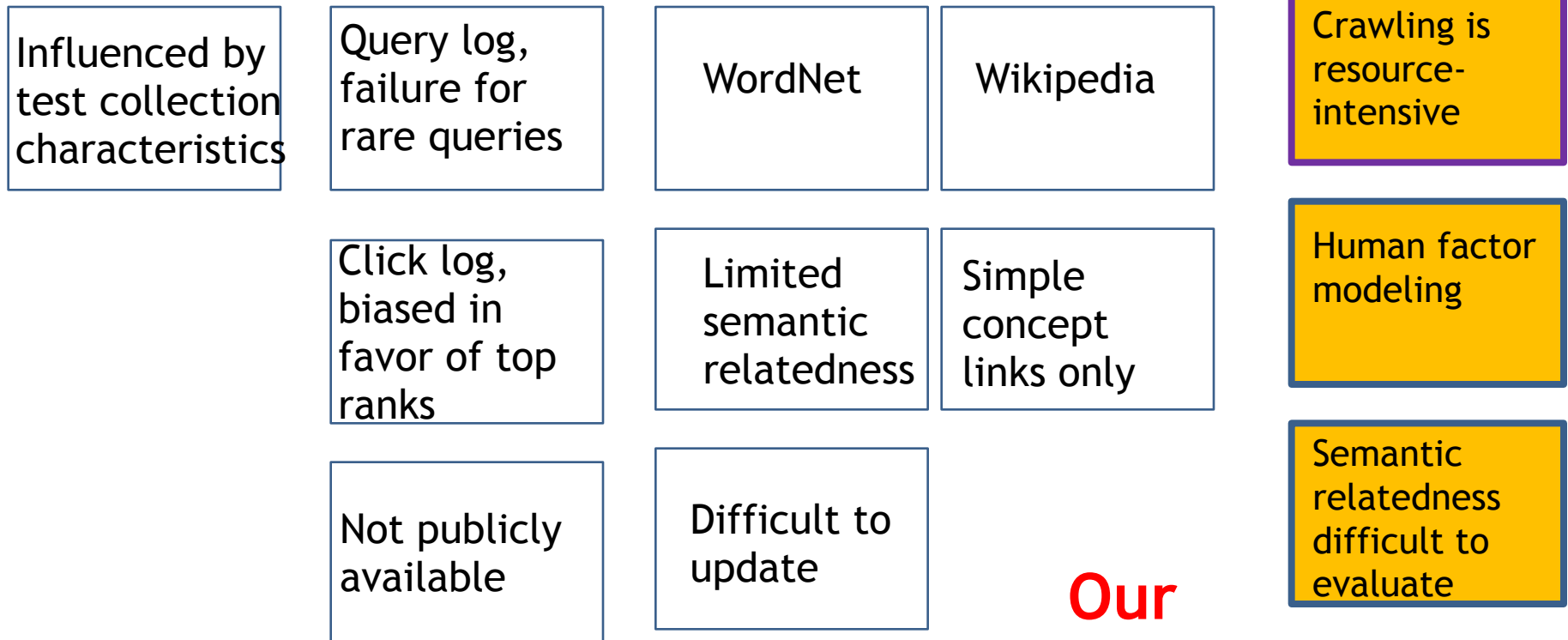
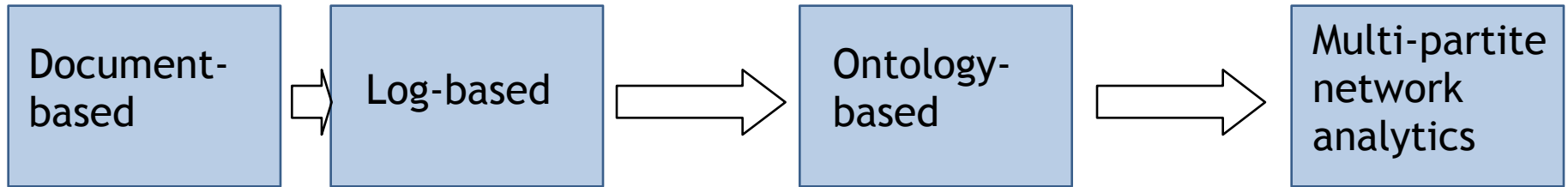
# Multi-partite Network Analytics

## Term Suggestion and Query Expansion



# Multi-partite Network Analytics

## Term Suggestion and Query Expansion



**Our  
Challenge**

# Wikipedia as Ontology

Wikipedia: Solar power

Log in / create account

Article Talk Read View source View history Search

## Solar power

From Wikipedia, the free encyclopedia

*This article is about generation of electricity using solar energy. For other uses of solar energy, see Solar energy.*

**Solar power** is the conversion of **sunlight** into **electricity**, either directly using **photovoltaics** (PV), or indirectly using **concentrated solar power** (CSP). Concentrated solar power systems use lenses or mirrors and tracking systems to focus a large area of sunlight into a small beam. Photovoltaics convert light into electric current using the **photoelectric effect**.<sup>[1]</sup>

Commercial concentrated solar power plants were first developed in the 1980s. The 354 MW **SEGS** CSP installation is the largest solar power plant in the world, located in the **Mojave Desert** of California. Other large CSP plants include the **Solnova Solar Power Station** (150 MW) and the **Andasol solar power station** (150 MW), both in Spain. The 214 MW **Charanka Solar Park** in India, is the world's largest photovoltaic plant.

**Contents** [hide]

- Applications
- Concentrating solar power



The PS10 conc heliostats onto

Wikidata: Solar energy

Property: **Solar energy** (P1000)

Instance of: **Renewable energy** (P1000)

Subclass of: **Energy conversion** (P1000)

Related to: **Solar power** (P1000)

Related to: **Solar thermal energy** (P1000)

Related to: **Solar photovoltaics** (P1000)

Related to: **Solar water heating** (P1000)

Related to: **Solar tower** (P1000)

Related to: **Solar pond** (P1000)

Related to: **Solar still** (P1000)

Related to: **Solar furnace** (P1000)

Related to: **Solar collector** (P1000)

Related to: **Solar panel** (P1000)

Related to: **Solar cell** (P1000)

Related to: **Solar battery** (P1000)

Related to: **Solar inverter** (P1000)

Related to: **Solar charge controller** (P1000)

Related to: **Solar microinverter** (P1000)

Related to: **Solar tracker** (P1000)

Related to: **Solar tower** (P1000)

Related to: **Solar pond** (P1000)

Related to: **Solar still** (P1000)

Related to: **Solar furnace** (P1000)

Related to: **Solar collector** (P1000)

Related to: **Solar panel** (P1000)

Related to: **Solar cell** (P1000)

Related to: **Solar battery** (P1000)

Related to: **Solar inverter** (P1000)

Related to: **Solar charge controller** (P1000)

Related to: **Solar microinverter** (P1000)

Related to: **Solar tracker** (P1000)

Wikipedia: Solar energy

Rate this page

What's new?

Freshness  Clarity  Completeness  Word-count

I am helping Wikipedia about this topic (optional)

Submit changes

Categories: **Energy conversion** | **Renewable energy** | **Solar power** | **Sun**

This page was last modified on 11 May 2012 at 05:55.

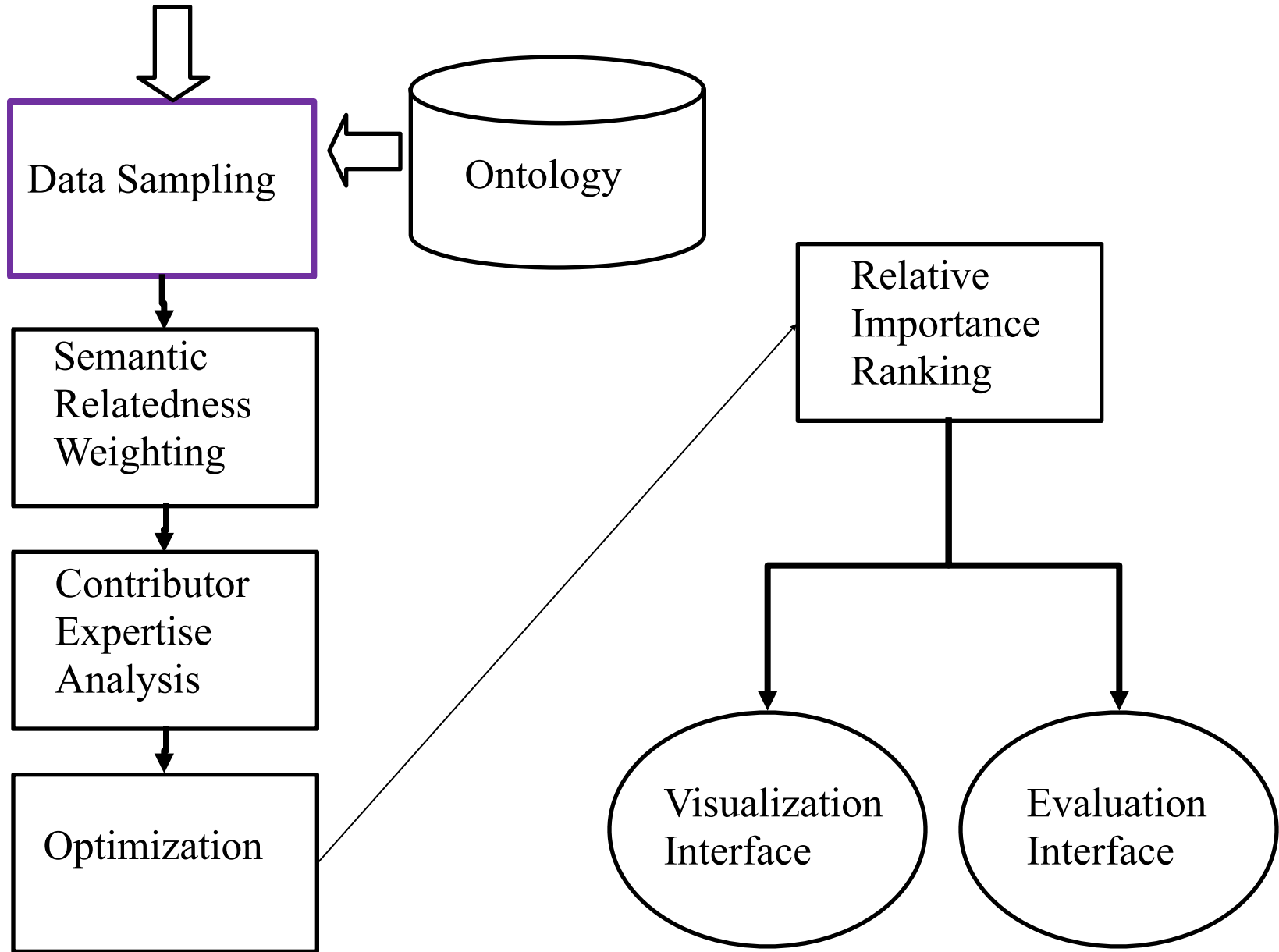
This text is available under the **Creative Commons Attribution-ShareAlike license**; additional terms may apply. See **Terms of use** for details. Wikipedia® is a registered trademark of the **Wikimedia Foundation**, Inc., a nonprofit organization.

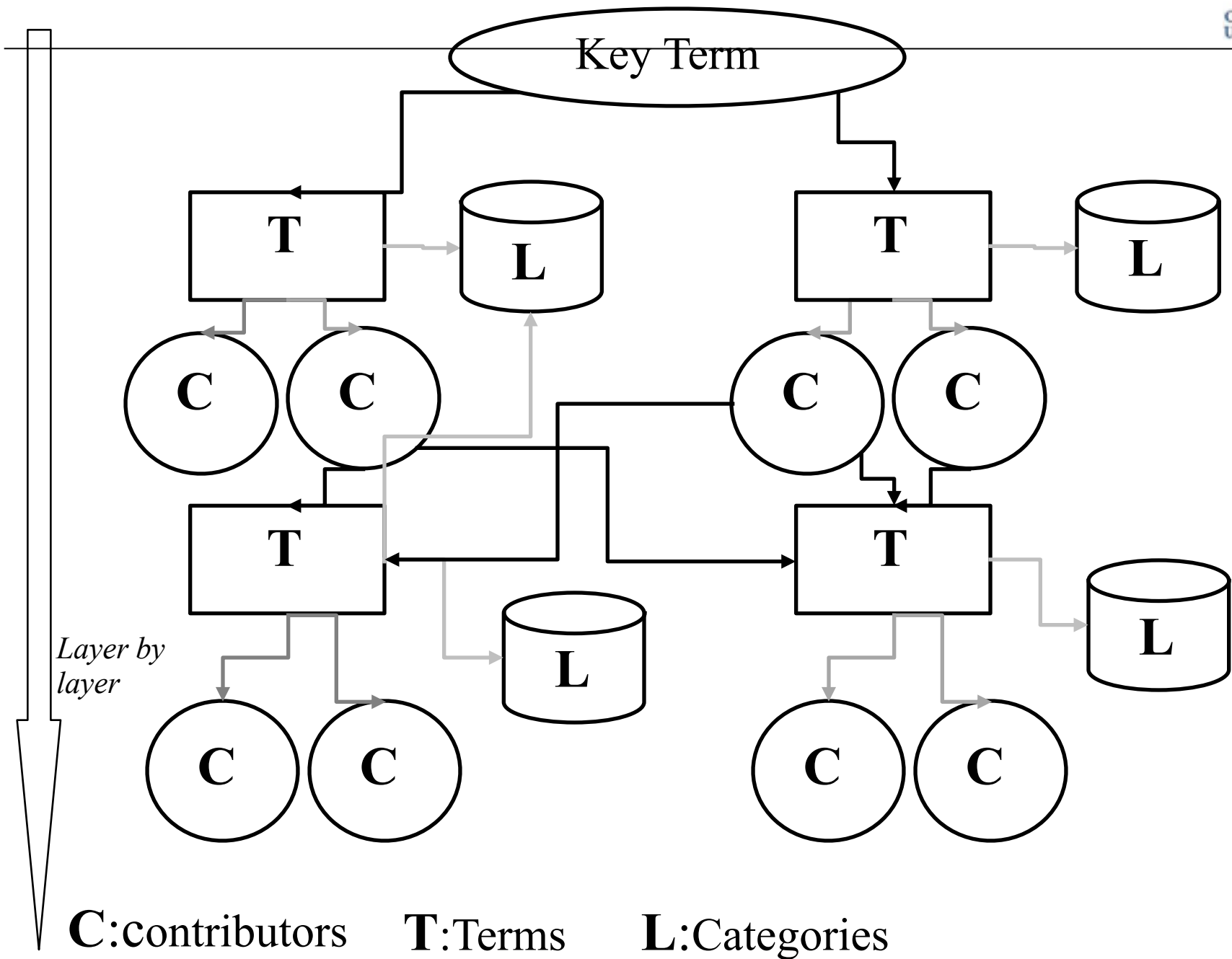
Contact us

Privacy policy | About Wikipedia | Disclaimers | Mobile view

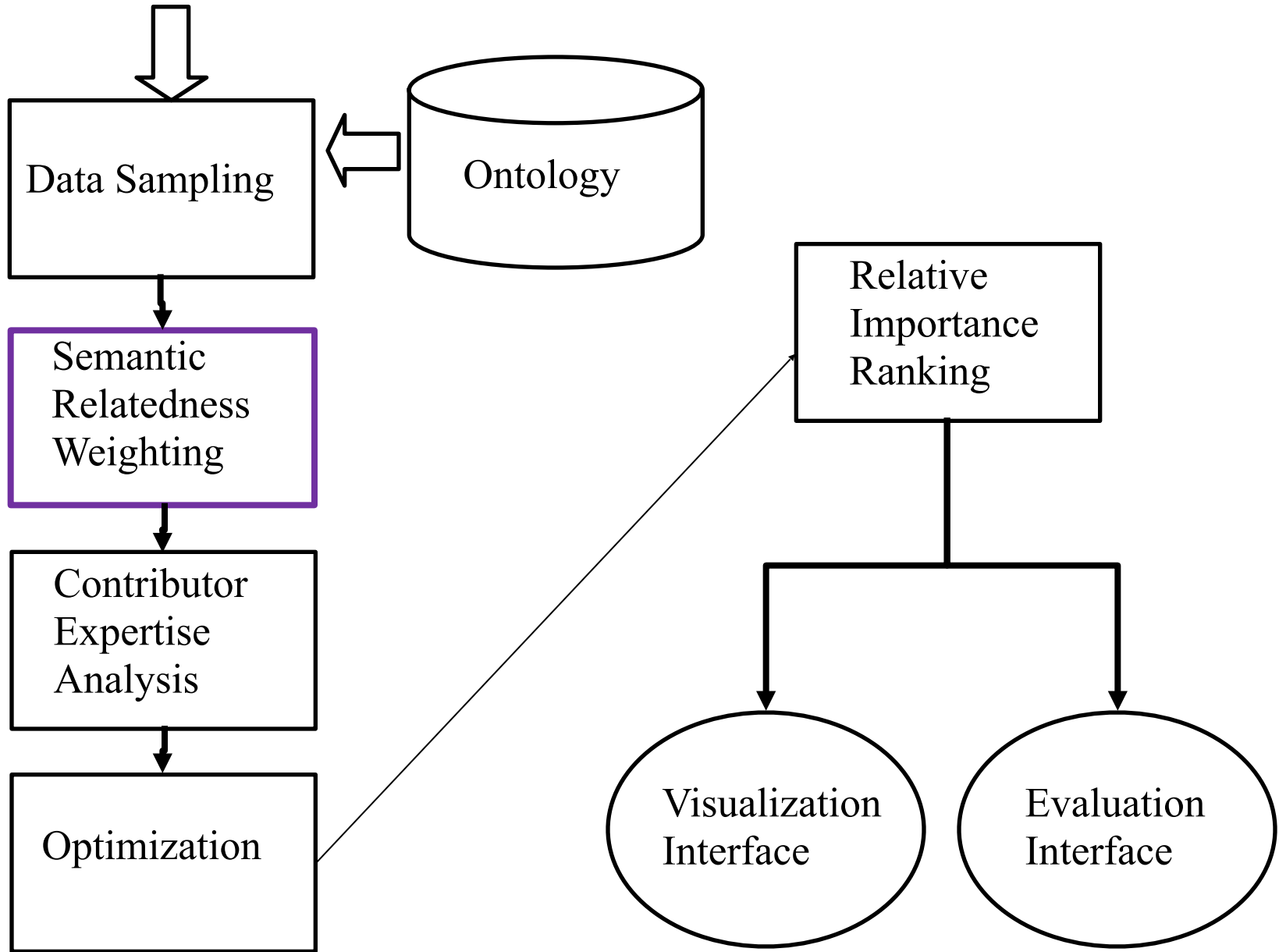


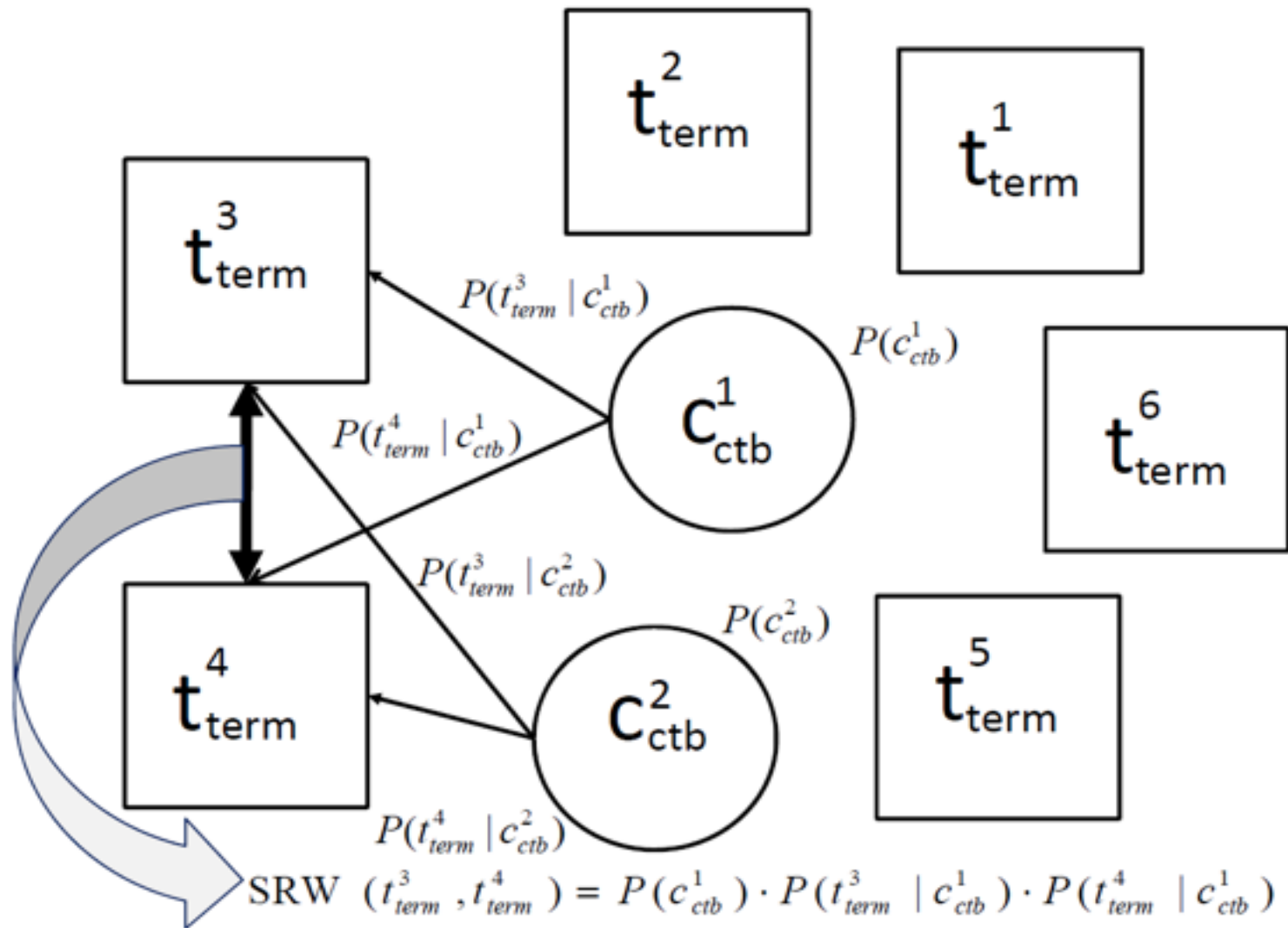
# Query



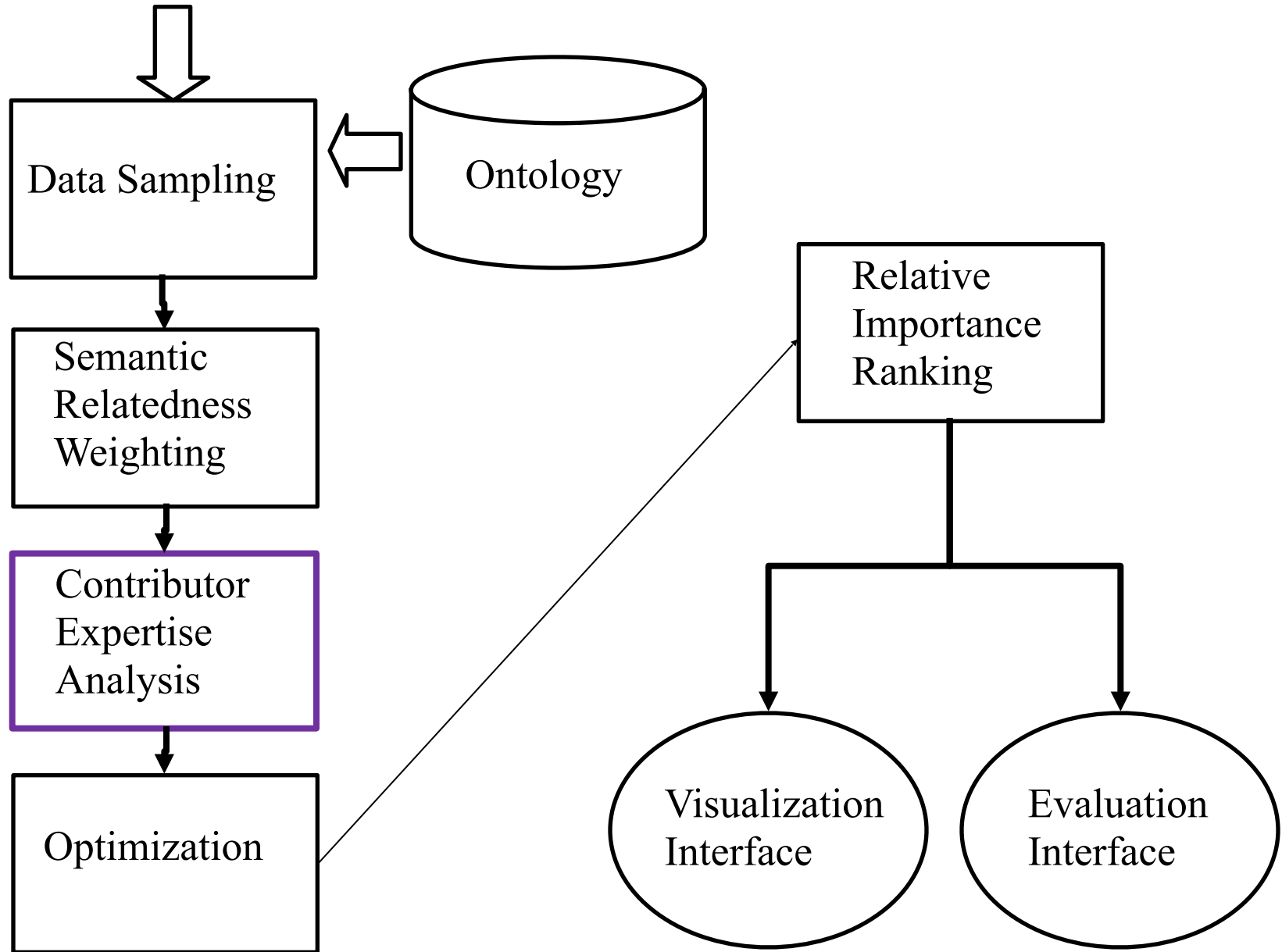


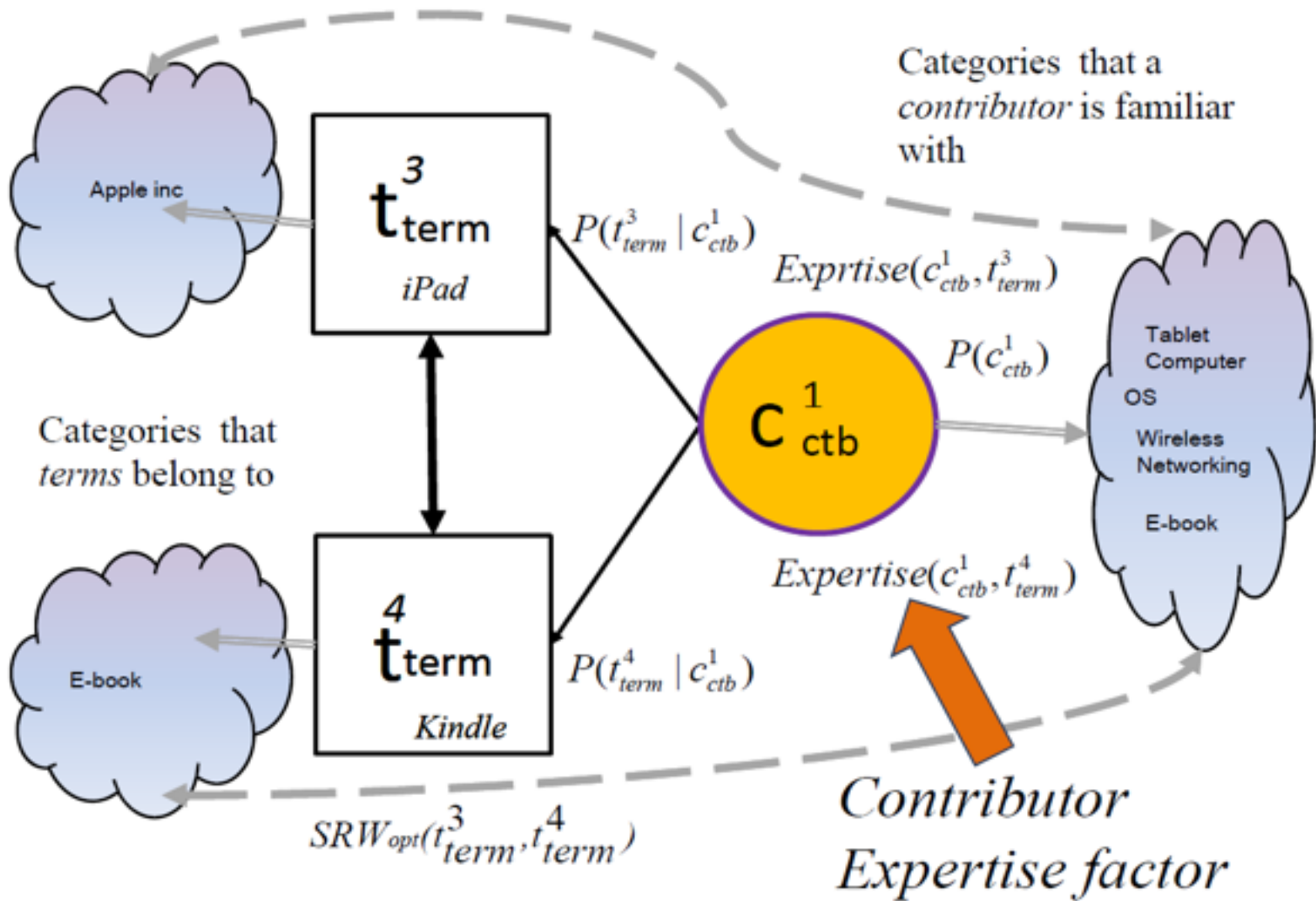
# Query

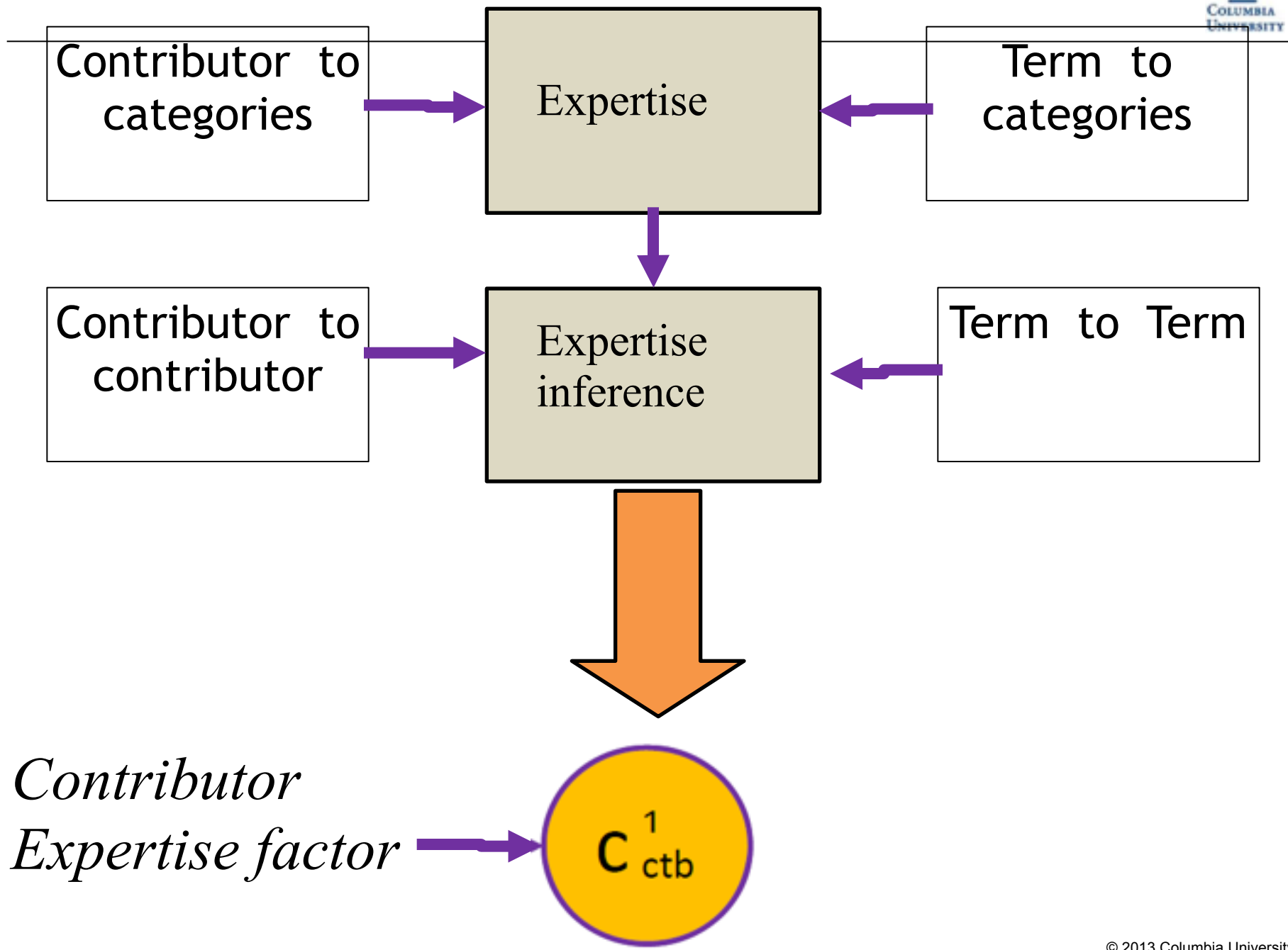




# Query

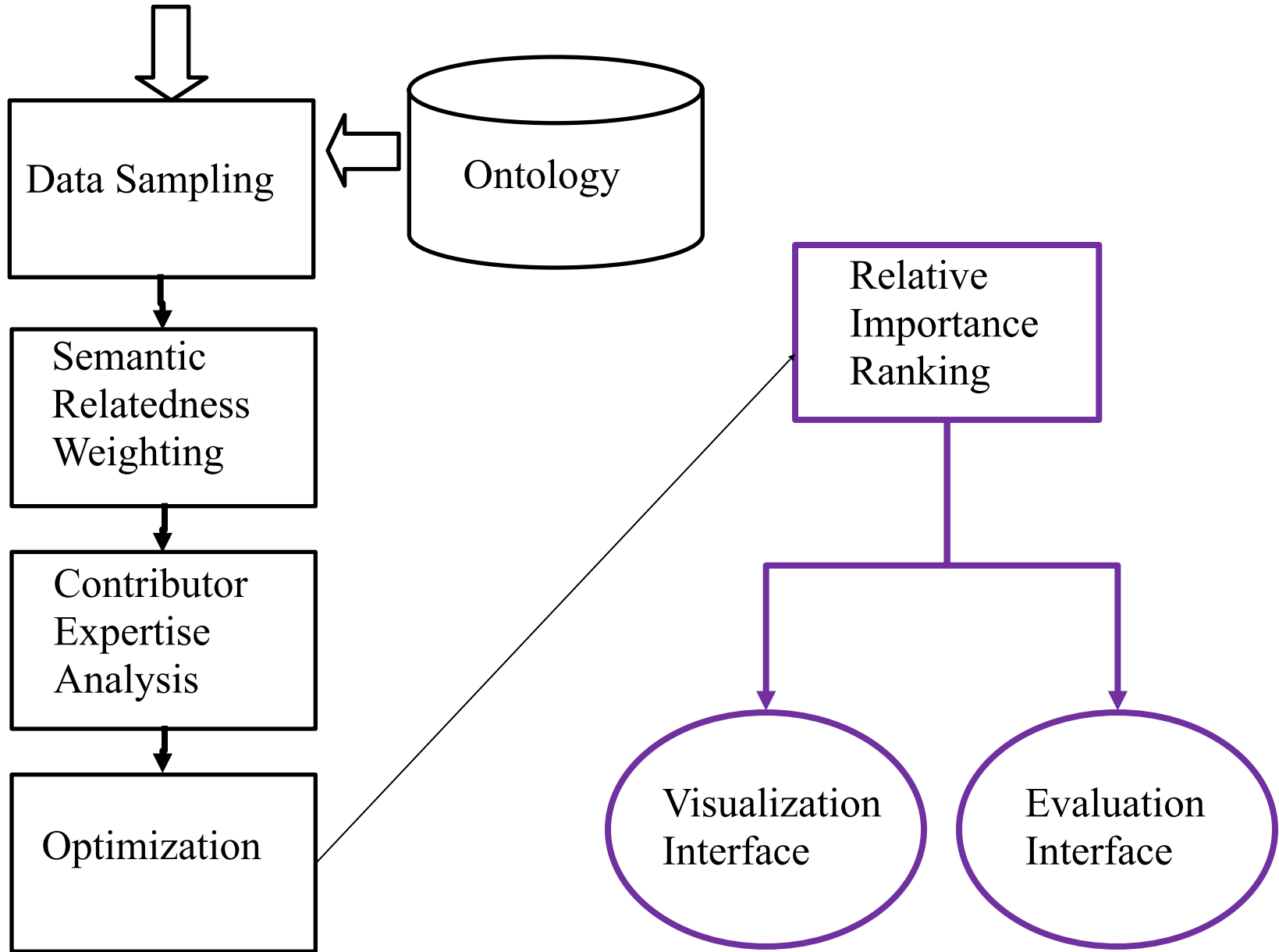






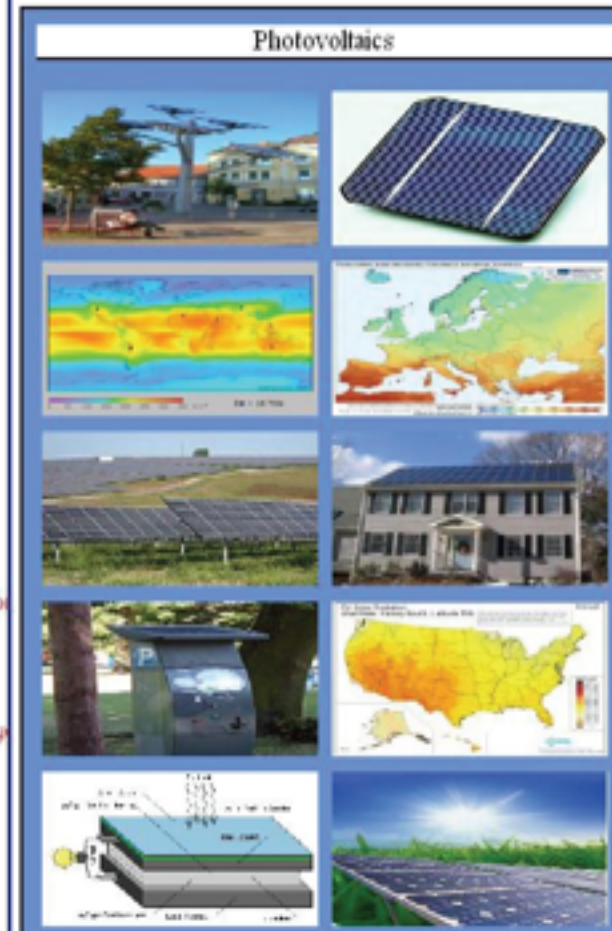
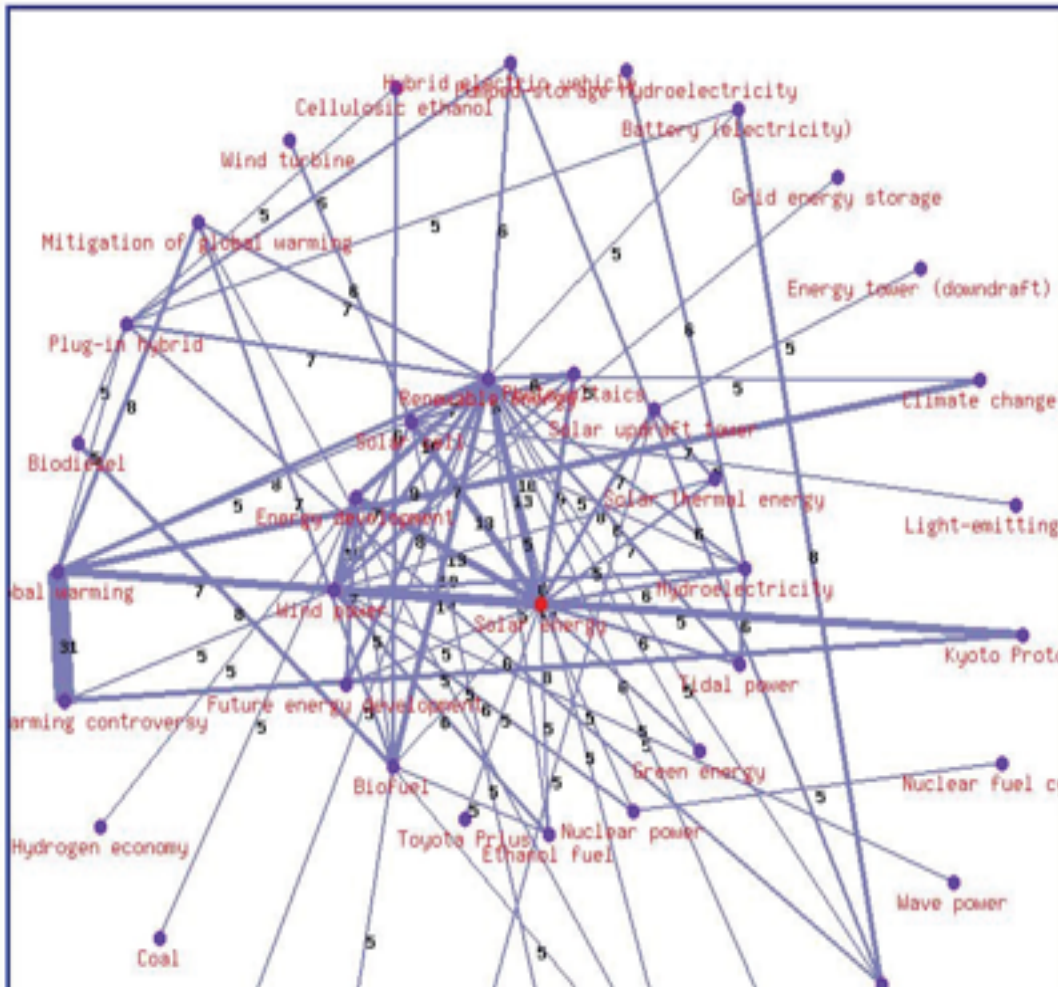


# Query



# High Semantic Relatedness Term Suggestion from Our System

"Solar Power" as keyword



# Word-completion Term Suggestion

+Shieh 搜尋 圖片 地圖 Play YouTube 新聞 Gmail 更多 -

Google

搜尋

全部  
圖片  
地圖  
影片  
新聞

solar power  
solar power system  
solar power industries  
solar power in hong kong  
solar power bank  
solar power 2012  
solar power international 2012  
solar power hong kong  
**solar power panel**  
solar power in china  
solar power generation

[Solar energy - Wikipedia, the free encyclopedia](#)

[Solar Power Systems, Home Solar Energy Systems, Solar Panel ...](#)

[www.solarpanelrebate.com.au/](#) - 頁車存檔 - 翻譯這個網頁

SolarGen is a leading installer of residential **solar power** systems and **solar panels** to homes throughout Australia. Call 1300 676 527 to get your free ...

「**solar power**」相關廣告

為什麼會顯示這則廣告？

[Solar Energy](#)

[www.element14.com/](#)

News & information resources about **Solar energy**, domestic & industrial

**solar power**的相關搜尋

[solar power 2011](#)

[solar power 2009](#)

[solar power 2012](#)

[solar power 2010](#)

[太陽能發電](#)

[neo solar power corp](#)

[neo solar power](#)

[solar power charger](#)

[china solar power](#)

[solar power inverter](#)

# Experiment I

	<b>P@1</b>	<b>P@5</b>	<b>S@5</b>	<b>S@20</b>	<b>MRR</b>
Simple link	0.3736	0.3039	0.6017	0.6231	0.4023
+Contributor	0.6151	0.3917	0.8031	0.8116	0.4125
+Expertise	0.6693	0.4412	0.8297	0.9620	0.5919

Performance Comparison for Different Relationship Levels.  
Using BibSonomy Dataset

# Experiment II – Accuracy on different categories

	Wordnet	Bag of words	Our algorithm
Literature	62.0% ± 5%	62.7% ± 4%	76.8% ± 6%
Natural science	60.7% ± 4%	65.6% ± 6%	73.3% ± 3%
Sociology	72.1% ± 5%	62.9% ± 5%	72.5% ± 7%
Business	60.4% ± 6%	58.5% ± 8%	67.1% ± 7%
Law	52.2% ± 9%	50.4% ± 8%	66.3% ± 6%
Engineering	54.0% ± 6%	68.3% ± 5%	66.2% ± 4%
Electrical & Computer Eng.	77.0% ± 4%	68.0% ± 3%	82.3% ± 3%
Life Science	73.1% ± 6%	70.9% ± 6%	81.4% ± 7%
Agriculture	72.6% ± 5%	65.1% ± 6%	72.3% ± 5%
Medical	63.0% ± 8%	65.6% ± 7%	61.6% ± 8%

ODP-based precision evaluation results increase 12.5% in average

## Precision Comparison With Paraphrase Detection System

	<b>Synonyms</b>	<b>Hyponymy</b>	<b>Antonyms</b>	<b>Paraphrase</b>
Zhao et al.	-	-	-	0.7444
<b>Our approach</b>	<b>0.2197</b>	<b>0.3665</b>	<b>0.2313</b>	-

82% of the suggested terms are reported as related, *i.e.*, synonyms (22%), hyponyms (37%) or antonyms (23%)

## References

---

- Jyh-Ren Shieh**, Ching-Yung Lin, Shun-Xuan Wang, Ja-Ling Wu, “Relational Term-Suggestion Graphs Incorporating Multi-Partite Concept and Expertise Networks,” *ACM Transactions on Intelligent Systems and Technology* (2012).
- Jyh-Ren Shieh**, Ching-Yung Lin, Shun-Xuan Wang, Ja-Ling Wu, “ Building Multi-Modal Relational Graphs for Multimedia Retrieval,” *International Journal of Multimedia Data Engineering and Management (IJMDEM)*: pp. 19-41 (2011). Best paper award nomination.
- Jyh-Ren Shieh**, Yung-Huan Hsieh, Yang-Ting Yeh, Tse-Chung Su, Ching-Yung Lin, Ja-Ling Wu, “Building term suggestion relational graphs from collective intelligence,” *World Wide Web Conference (WWW 2009)* pp. 1091-1092 (2009).
- Jyh-Ren Shieh**, Yang-Ting Yeh, Chih-Hung Lin, Ching-Yung Lin and Ja-Ling Wu, “Using Semantic Graphs for Image Search,” *IEEE International Conference on Multimedia & Expo (ICME 2008)*, pp. 105-108 (2008).



## Homework #2

**Based on the systems implemented in Homework #1 to implement a stock prediction system. Reference the methodology introduced in Lecture 2, with any additional method/tool/algorithm/data of your choice.**

**The grading of Homework #2 will depend on the prediction accuracy — the date after your submission.**

**TA will announce a list of companies to be predicted and the way to submit your result.**

**You can use any tool. If you need a graph/network database, analytics, and visualization tool, a beta version of IBM System G Cloud is available at: <http://systemg.research.ibm.com/cloud.html> (If you face some bug issue, you can send email to Dr. Yinglong Xia ([yxia@us.ibm.com](mailto:yxia@us.ibm.com)) for help.)**