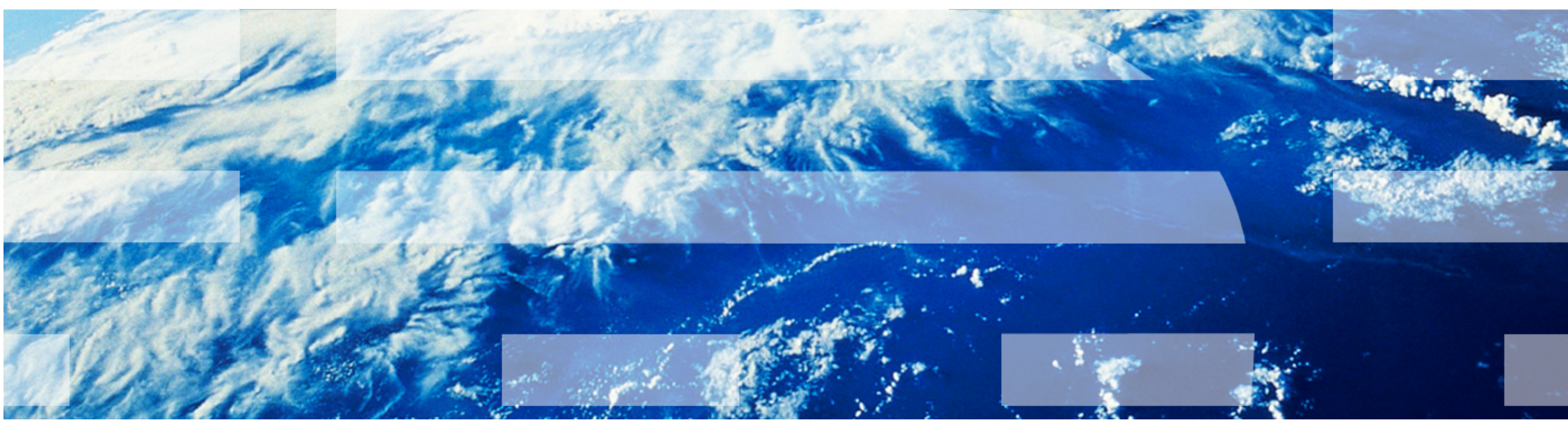
E6895 Advanced Big Data Analytics Lecture 10:

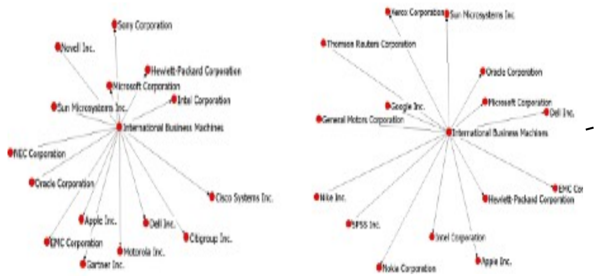
Social and Cognitive Analytics

Ching-Yung Lin, Ph.D.

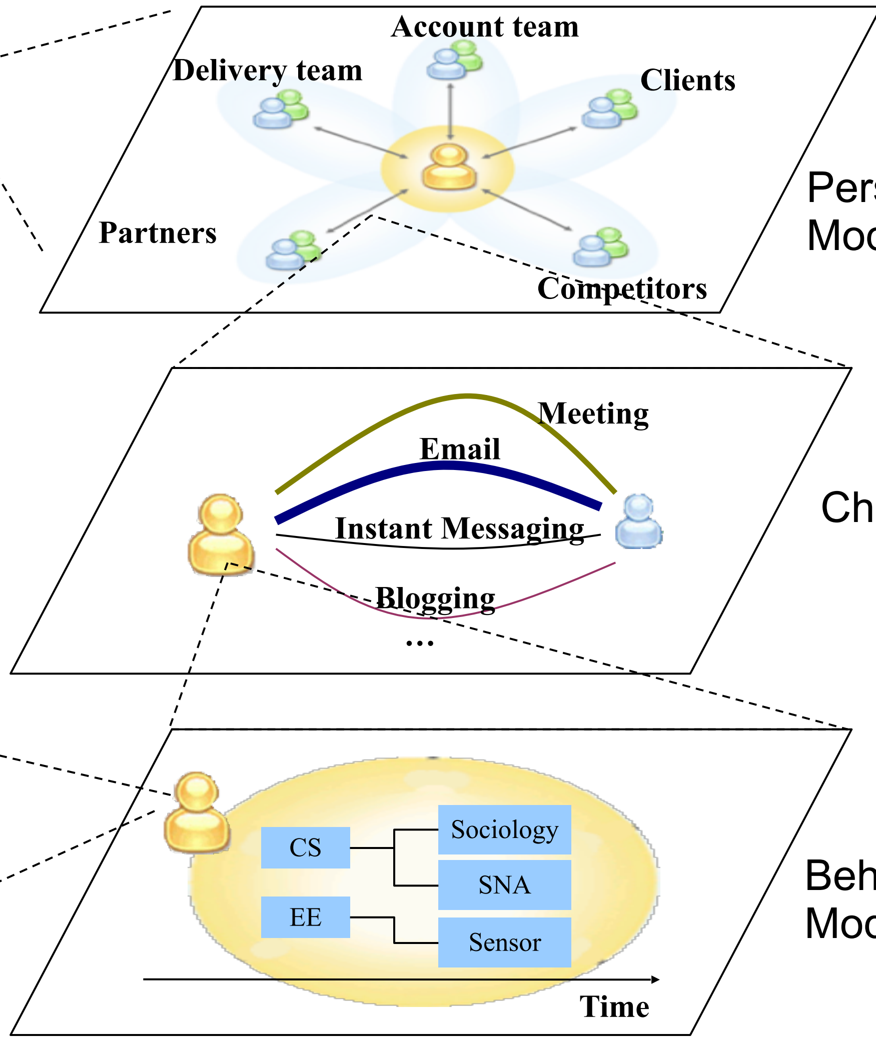
Adjunct Professor, Dept. of Electrical Engineering and Computer Science



Understanding People – from cognitive level to societal level



Organization Networks Modeling

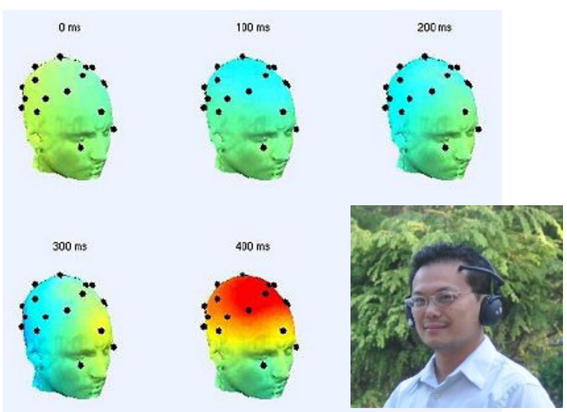


Personal Network Modeling

Channel Modeling

Behavior Evolution Modeling

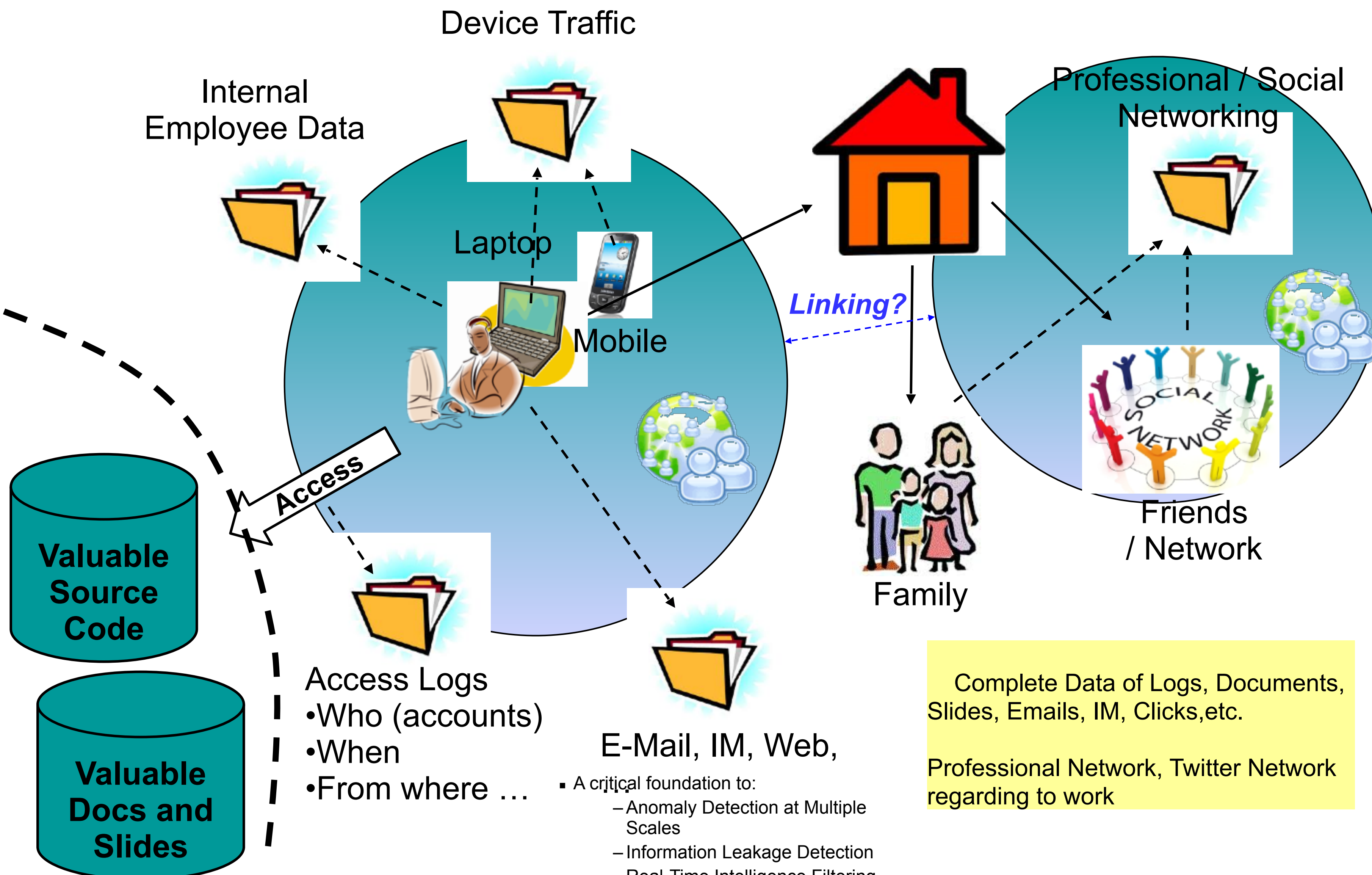
Combining Signal Processing , Data & Graph Mining, and Machine Learning by multi-modality signal sensing & understanding



Cognitive Signal Modeling

Computational Social and Cognitive Sciences

People Analytics → All Aspects of a Person



Access Logs
- Who (accounts)
- When
- From where …

E-Mail, IM, Web,

- A critical foundation to:
  - Anomaly Detection at Multiple Scales
  - Information Leakage Detection
  - Real-Time Intelligence Filtering

Complete Data of Logs, Documents, Slides, Emails, IM, Clicks,etc.

Professional Network, Twitter Network regarding to work

22

Example — our enterprise social analytics system

Emails
Chats
Meetings
Web Page Clicks
DB Server Logs

Social sensors
Click streams capturer
Feed subscription
Database access

Graph analysis
Behavior analysis
Semantics analysis

Multimodality Analysis

15,000 volunteers; 76 countries; 119,000 users

25,000,000 emails & instant messages (incl. content)

1,500,000 Learning click data; 44,000 entities

6,681,000 Knowledge & Sales access data; 240,000 entities

1,687,000 Media Library access data; 105,000 entities

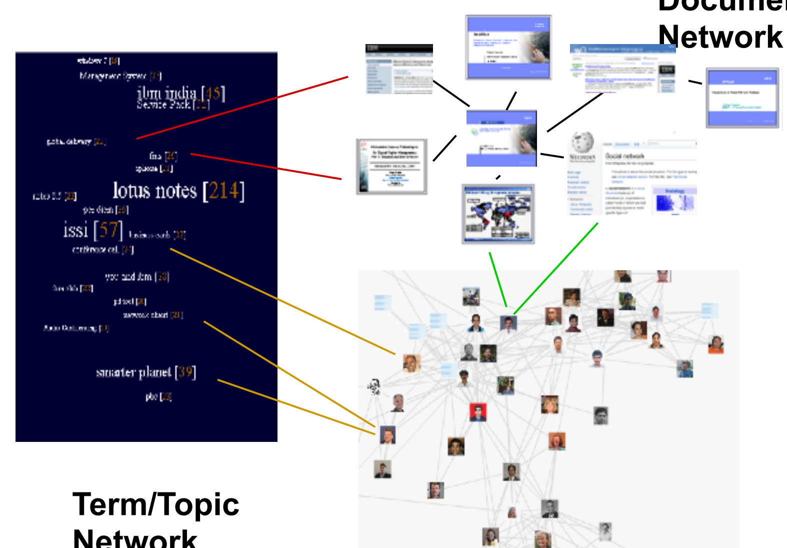
700,000 posts (blogs 3,000, file sharing 210,000, bookmark 450,000, Wiki 11,000) data

200,000 people's consulting financial databases

400,000 organization/demographic data

100,000 intranet searches per day

Document Network

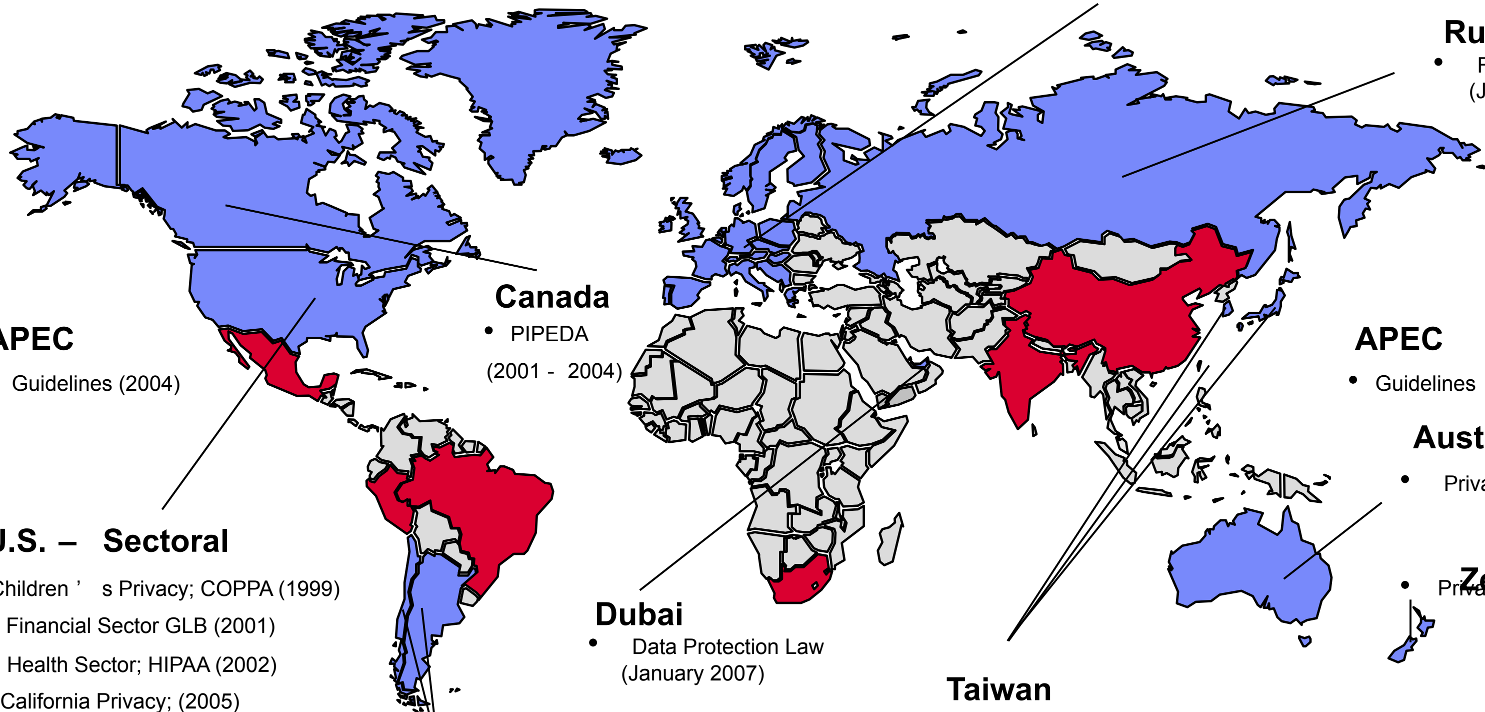


Term/Topic Network

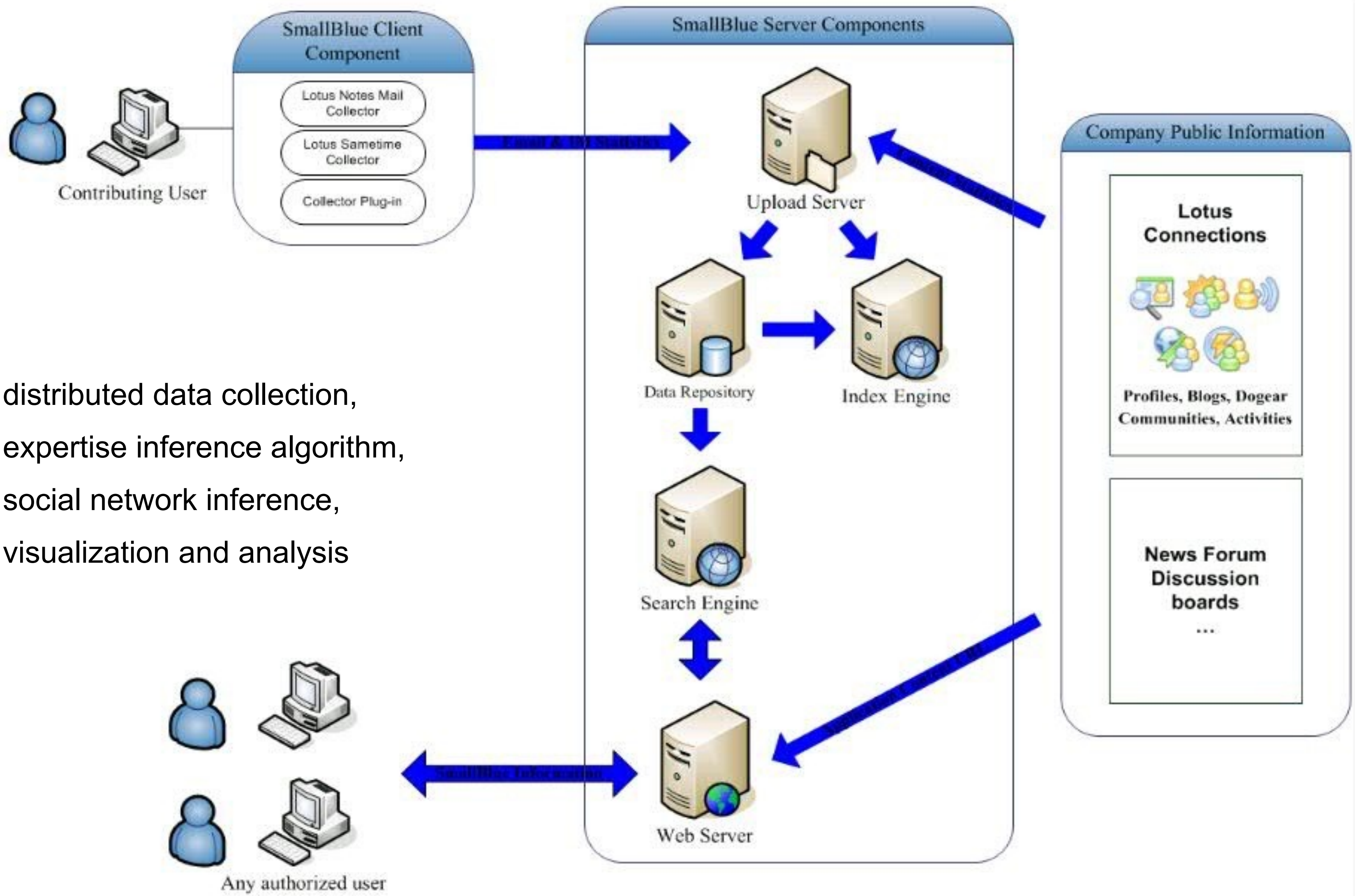
People Network

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

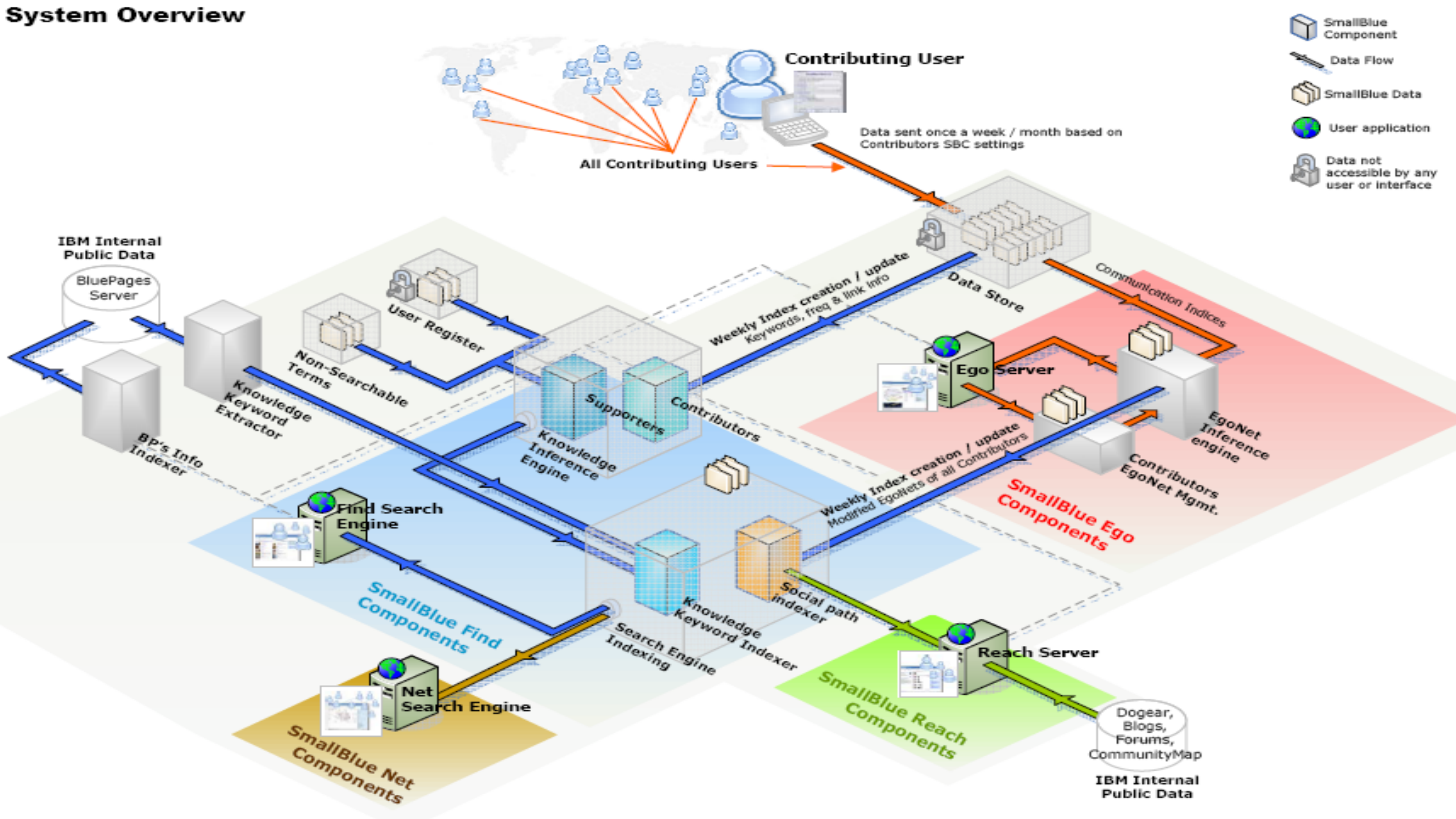


Social Mining Enterprise Architecture

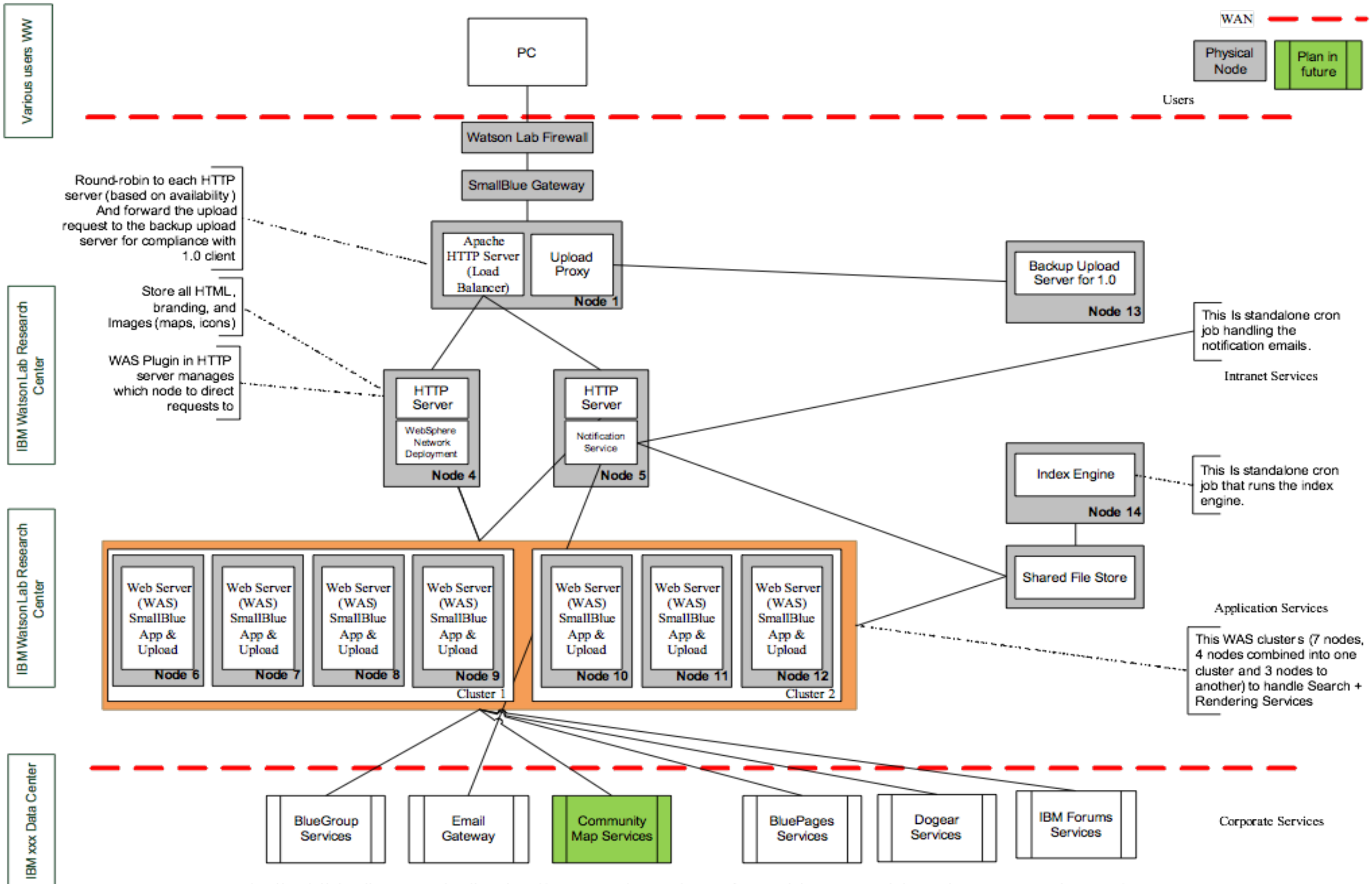


- novel distributed data collection,
- novel expertise inference algorithm,
- novel social network inference,
- novel visualization and analysis

SmallBlue Enterprise 1.1

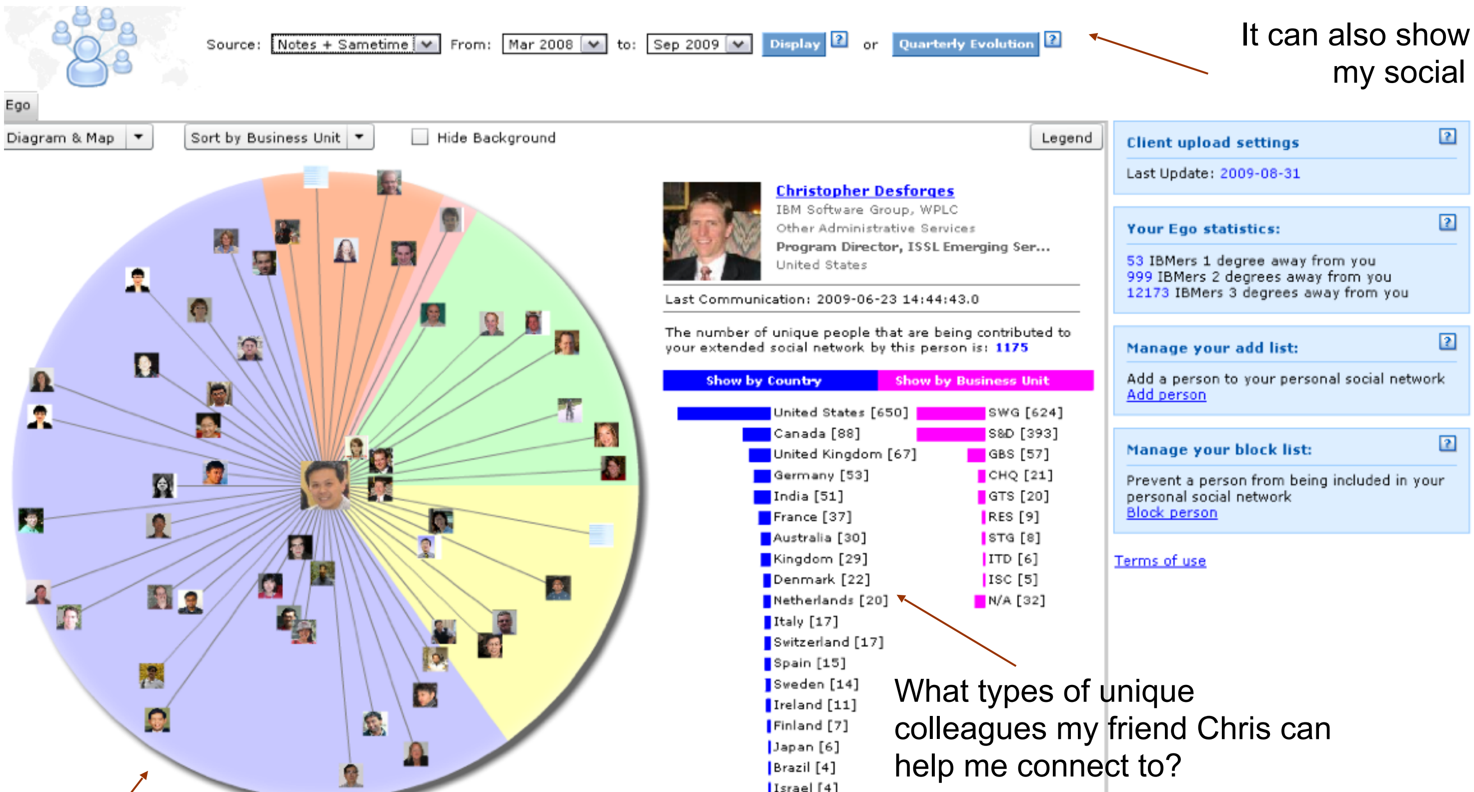


Advanced Social Mining Enterprise 1.1 architecture



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

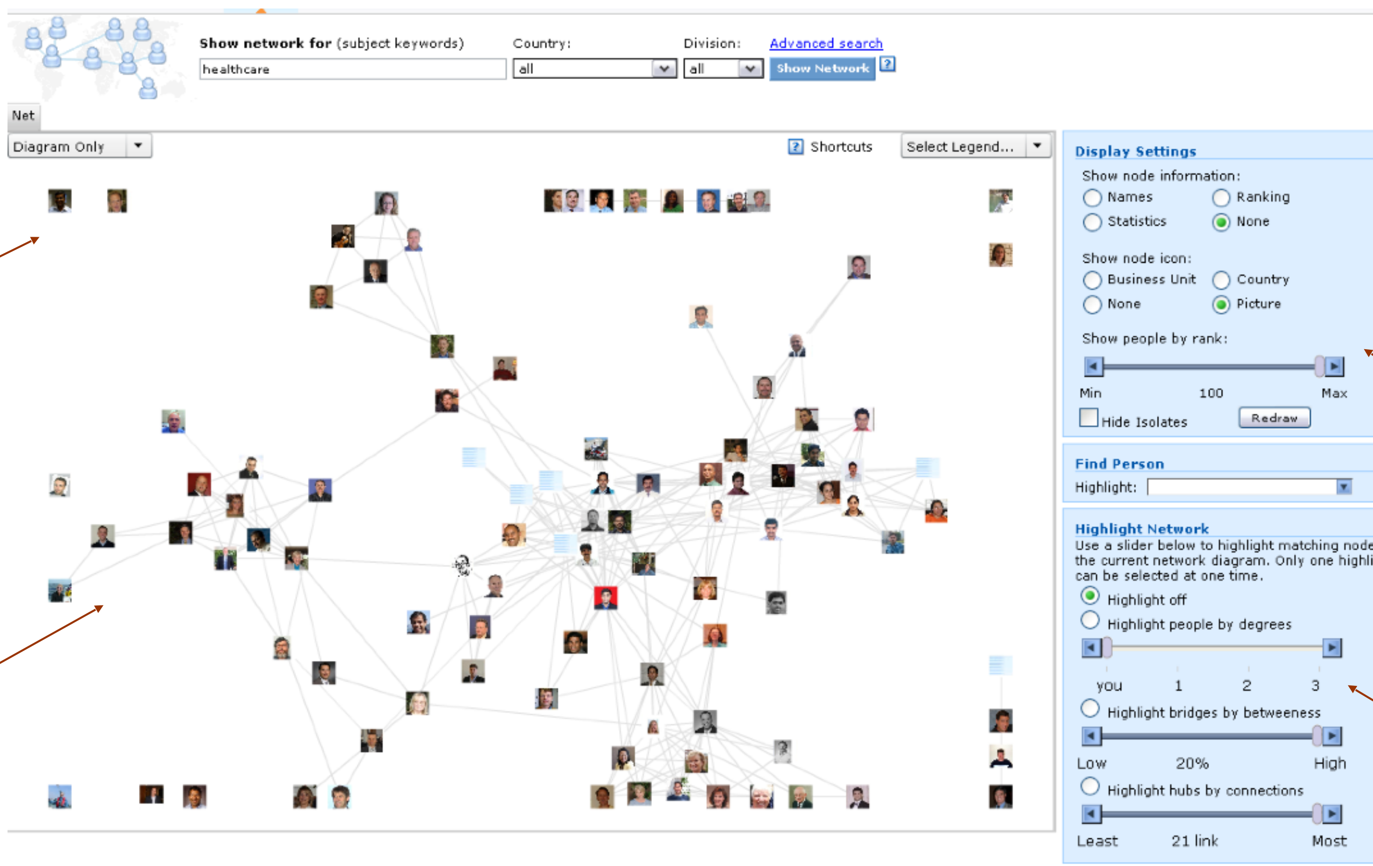
- Adding a person in personal network (i.e., frequent communications), increases $948 yearly revenue for IBM. (***selected by BusinessWeek Magazine as the Top Story of the Week, April 10, 2009***)
- 1% increase in social network diversity is associated with $239.5 in monthly revenue
- 1% increase in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

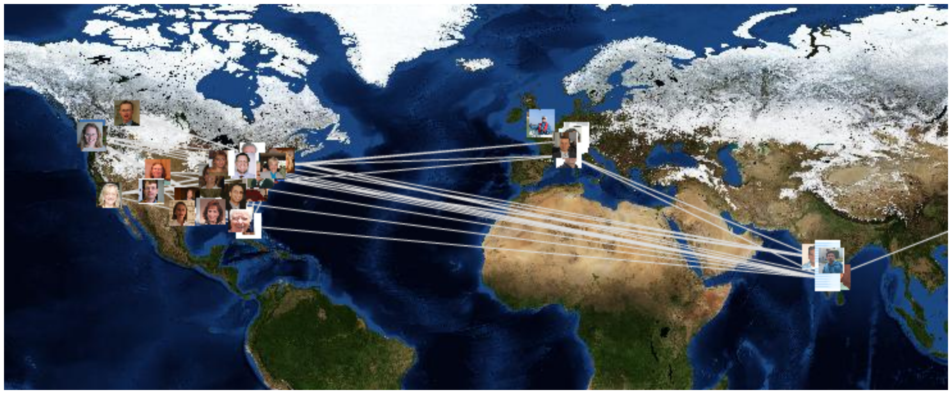
A cluster of healthcare experts



Top-N experts on healthcare

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

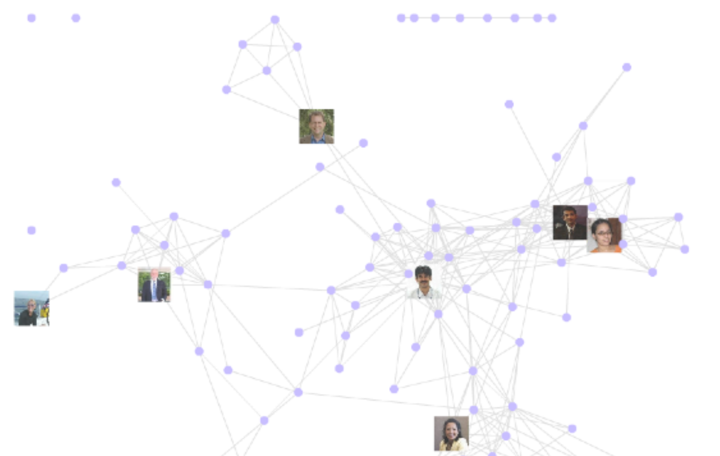
Social Network Analysis (cont'd)



IBM Healthcare-related employees in the world

Connections between different divisions

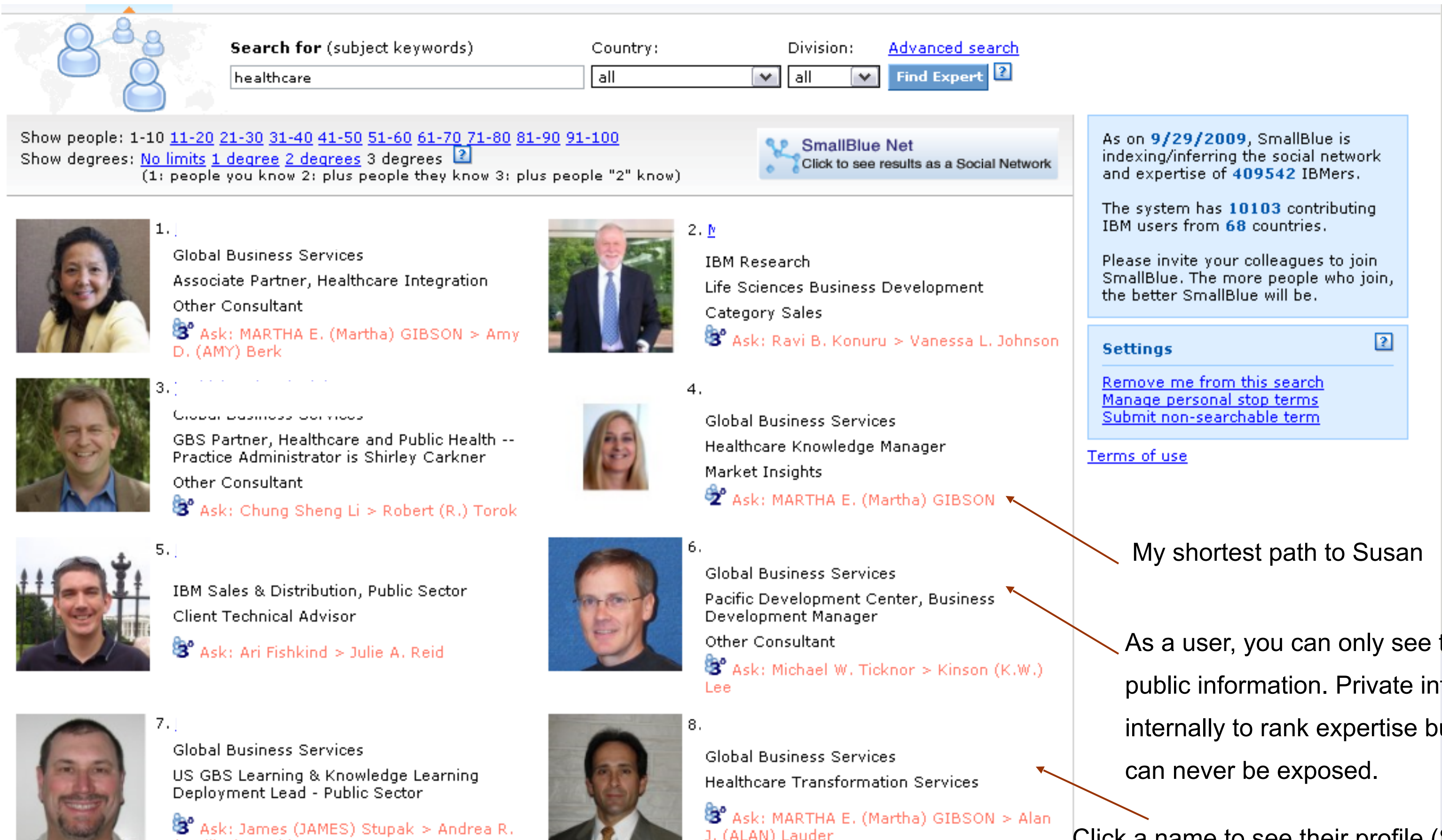
IBM Healthcare-related employees in the U.S.



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)



My shortest path to Susan

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.

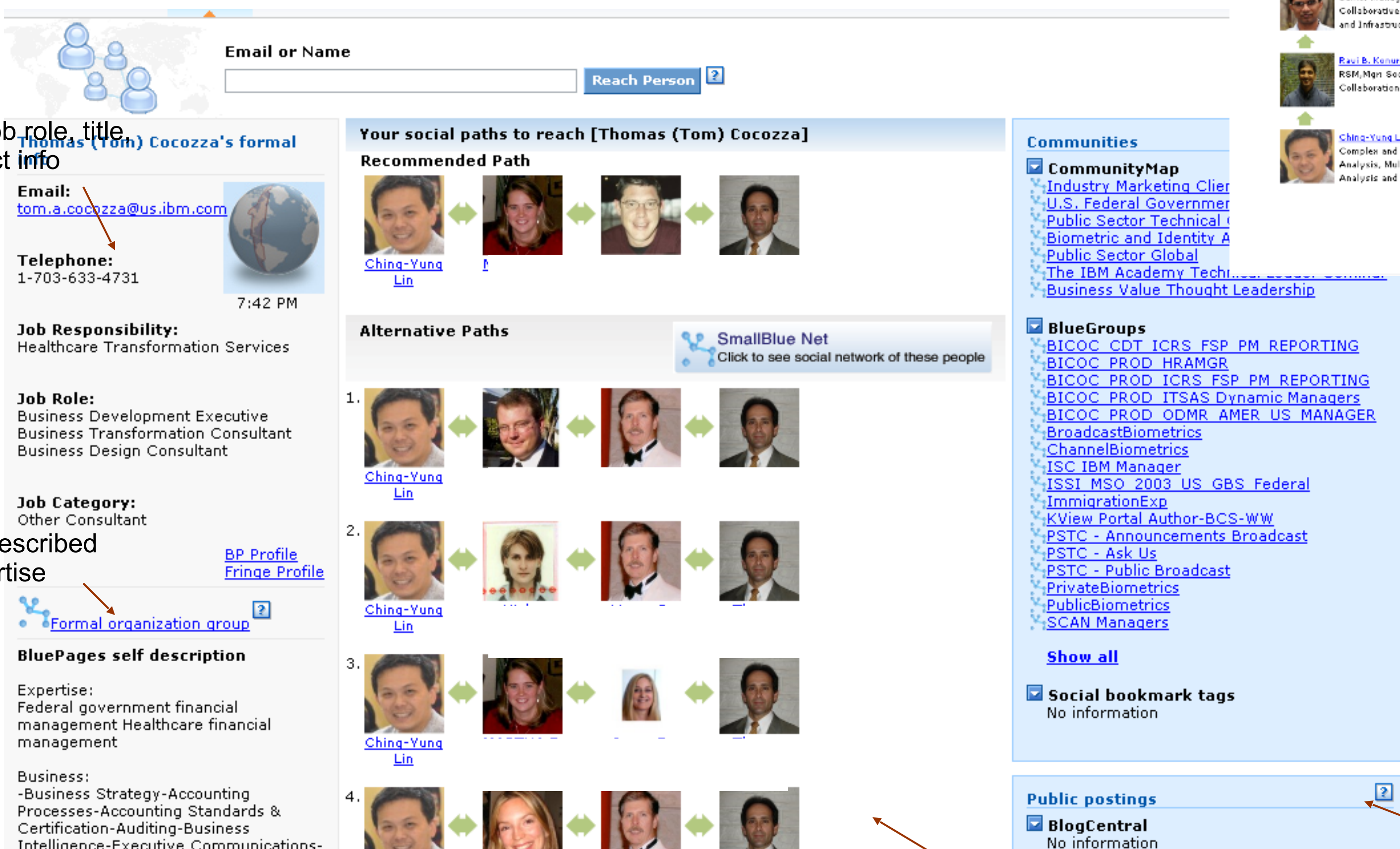
Click a name to see their profile (SmallBlue Reach)

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

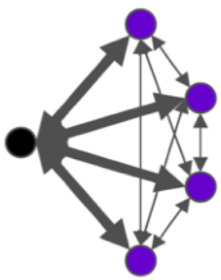
The public interest groups he is in

His blogs, forum, postings..

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

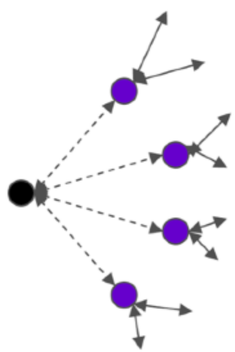
Economic Issues – Network Topology and Worker Productivity

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



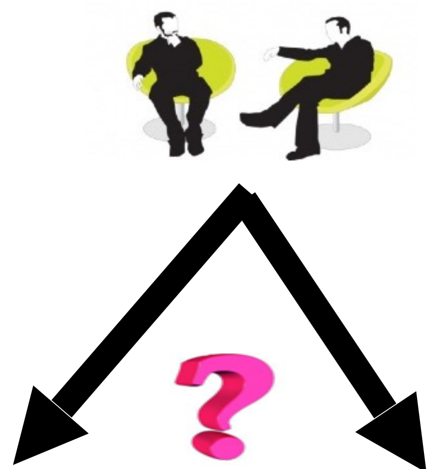
<u>Cohesive Network</u>

- Trust
- Absorptive capacity
- Precision, Reliability



<u>Structurally Diverse Network</u>

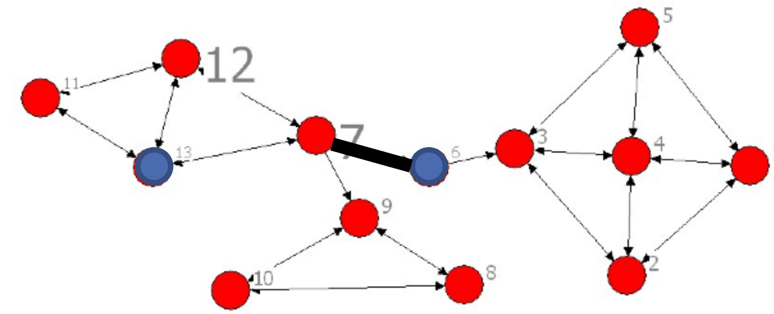
- Brokering position
- Access to many pools of diverse, novel information



<u>What type of network structure is most beneficial in a electronic network for consultants?</u>

- 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 Div12)=0.16	+Transfer complex knowledge - Access diverse knowledge	Diversity→ Strong work performance (?)

Enterprise becomes more successful utilizing Social Network Analysis

- 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 Company 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 Company. (***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).


BusinessWeek
SEARCH SITE
GO
Advanced Search
TOP NEWS | BW MAGAZINE | INVESTING | ASIA | EUROPE | TECHNOLOGY | AUTOS | INNOVATION | SMALL BIZ | B-SCHOOLS | CAREERS
APRIL 10, 2009
Insider Newsletter
A weekly summary of the best in BusinessWeek and BusinessWeek.com
NEWS THIS WEEK'S TOP STORY
Putting a Price on Social Connections
Researchers at IBM and MIT have found that certain e-mail patterns at work correlate with higher revenue production
EDITOR'S MEMO
Researchers at IBM and Massachusetts Institute of Technology say that indulging in certain types of electronic communications makes for higher productivity at work. Our Insider Top Story this week, "Putting a Price on Social Connections," is part of our Special Report, The Value of Virtual Friends. Find out what sort of networking increases your value as an employee.
You've seen the advertisements on TV and in the paper—apparently there has never been a better time to sell Grandma's tiara. But hold on: Click through our slide show, "Selling Your Jewelry," and read the accompanying story first to see what you can get for what you've got. Have you ever had the thing appraised?
Beer sales are down in the mature markets, but one brewer is hoping an untapped market will make up for it. BW correspondent Kerry Capell follows giant SABMiller into Africa. The idea is to make commercial beer that's more affordable, to compete with local brews.
And we have stories on buying shares in distressed sectors and forming nonprofits in this troubled economy.

- 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 corrected.
  - *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.

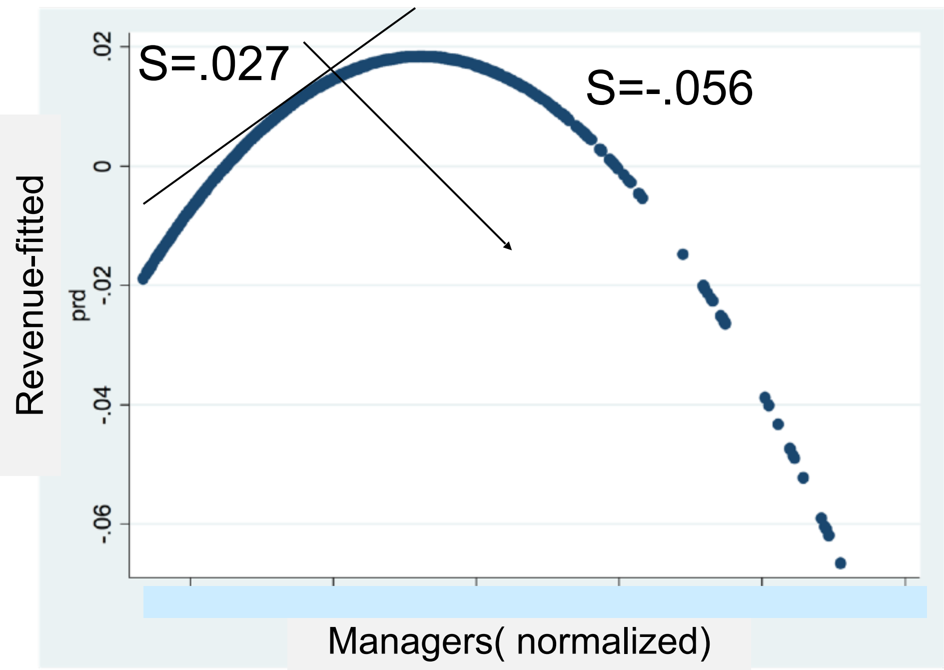
Project Team Composition—Managers

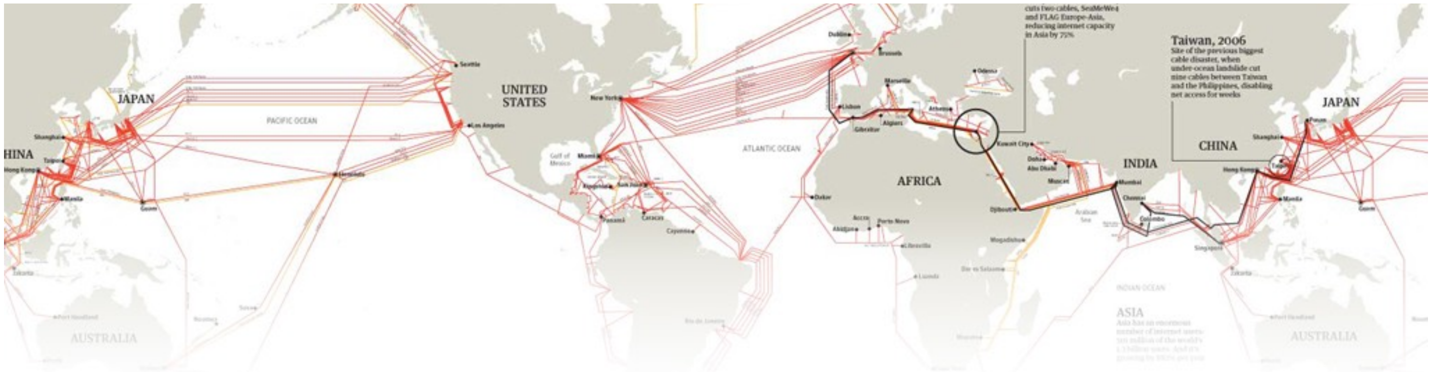
The number of managers in a project exhibit an inverted-U shaped curve.

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.

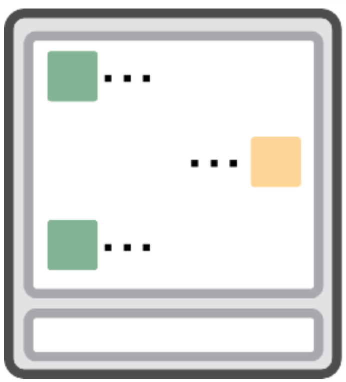
$$revenue = \alpha + \beta_1 \cdot mgr + \beta_2 \cdot mgr^2 + \gamma_1 \cdot otherfactor_1 + \ldots + \gamma_k \cdot otherfactor_k + \varepsilon$$

# Managers in project	β_1	**2733.9*** (537.5)**
(# Managers in project) ^2	β_2	**-682.02*** (215.3)**





Collaborating Globally:

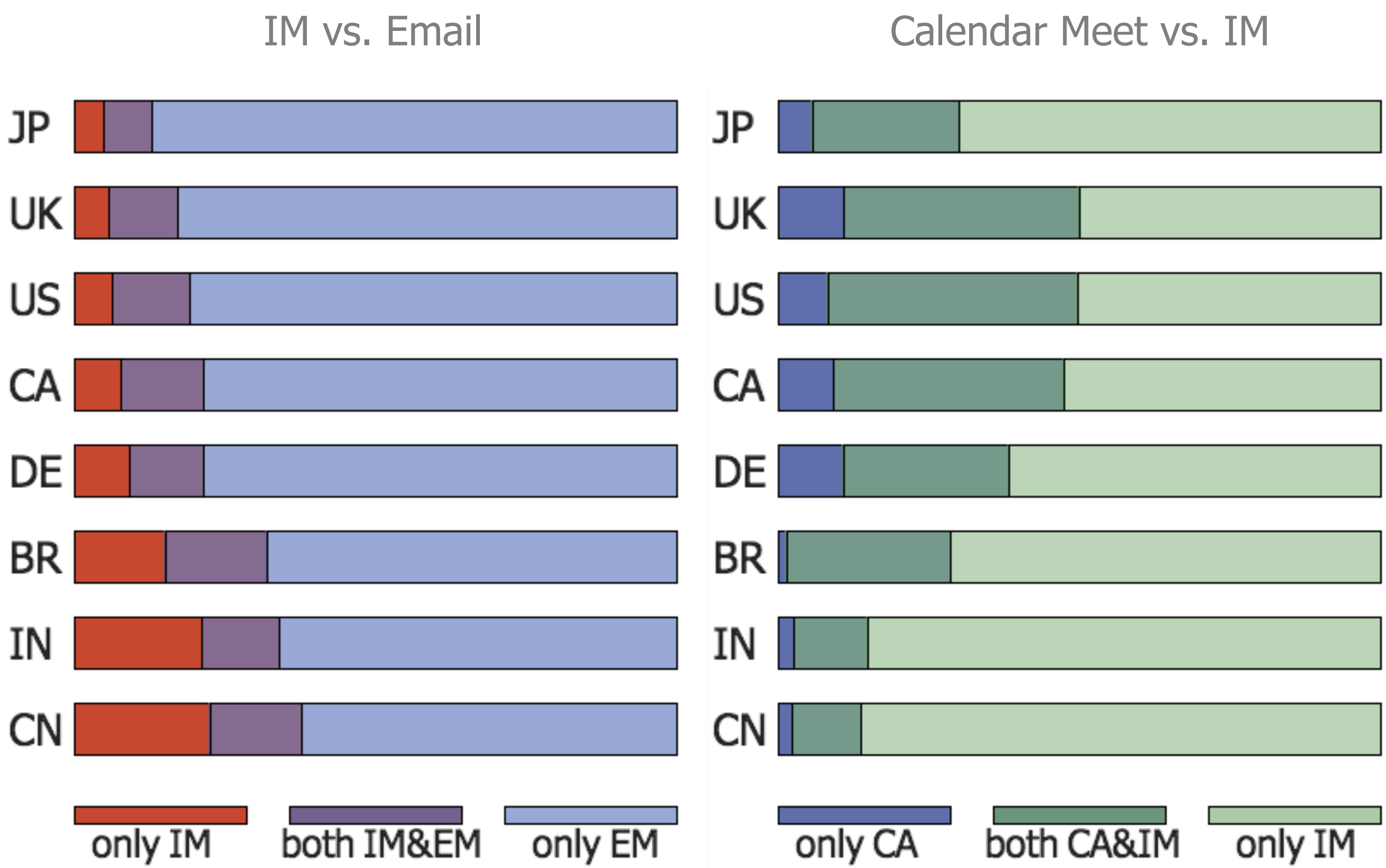


preferences of CMC tools

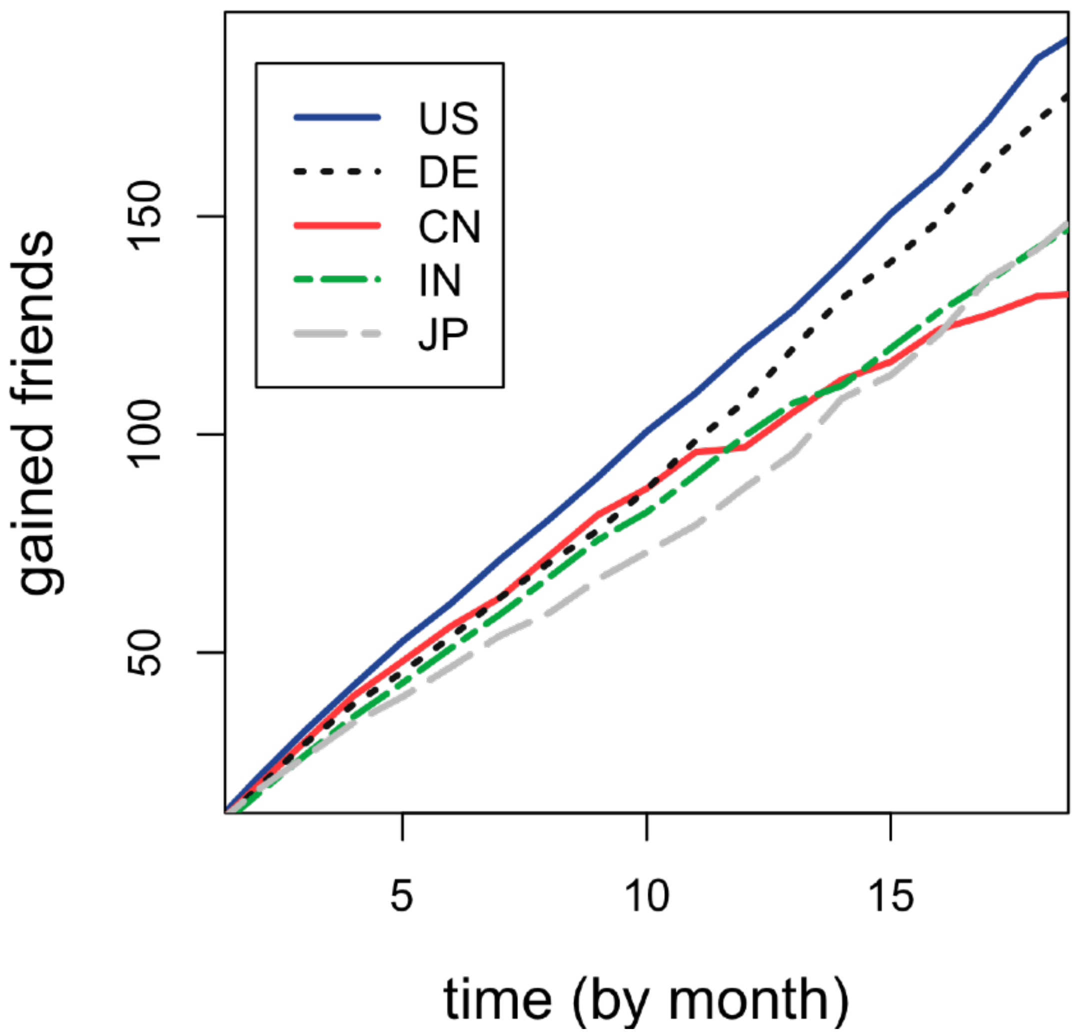
patterns of growing social network

sentiments in conversations

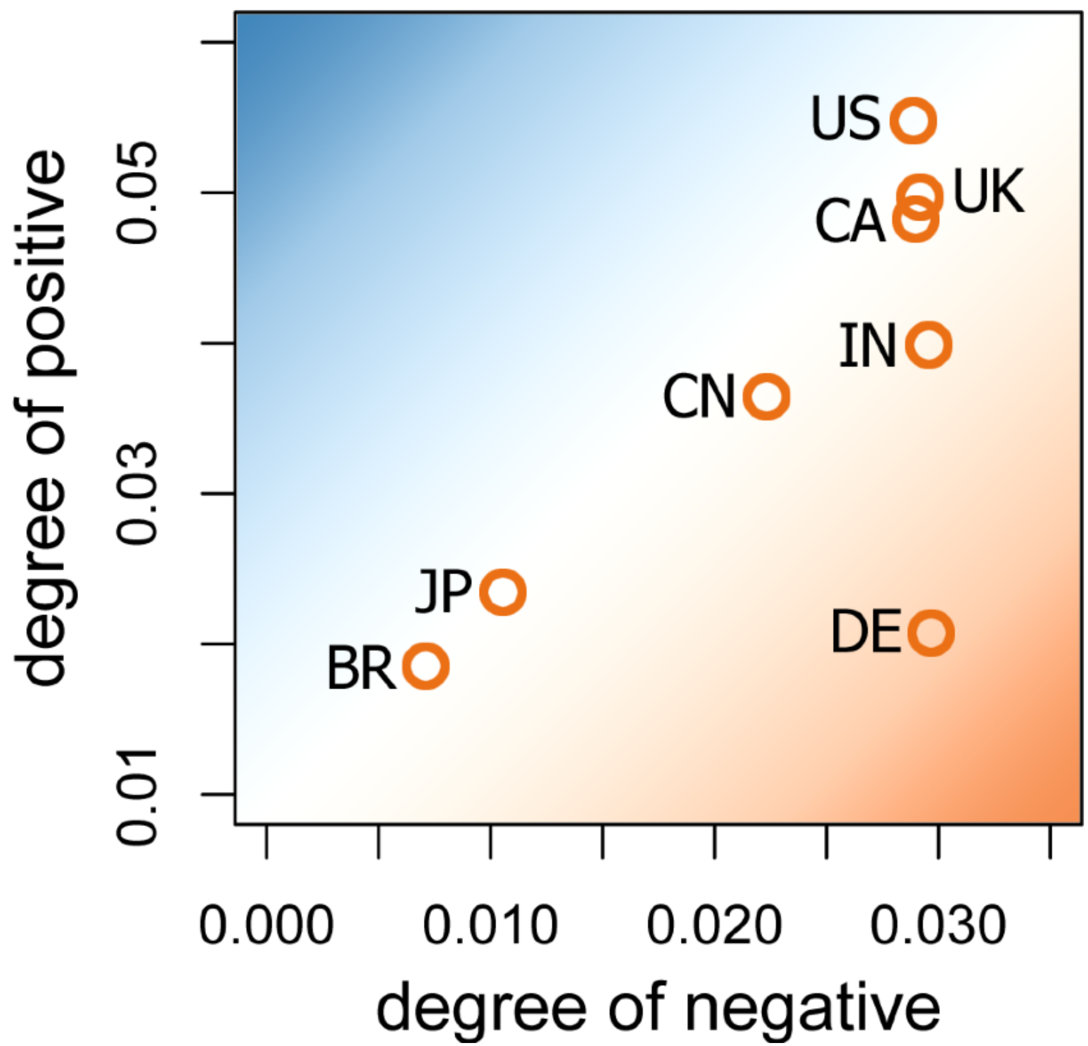
Preferences of CMC Tools



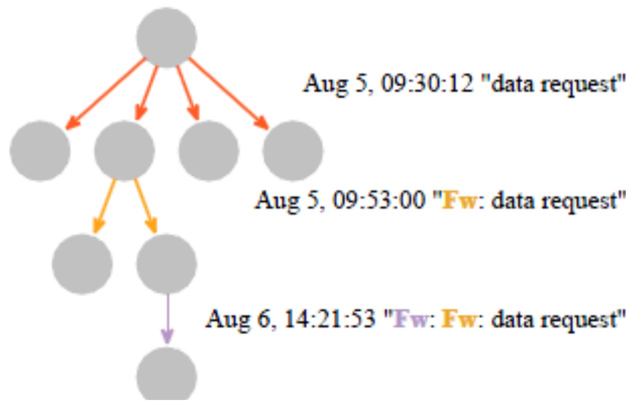
Growing one's Social Networks

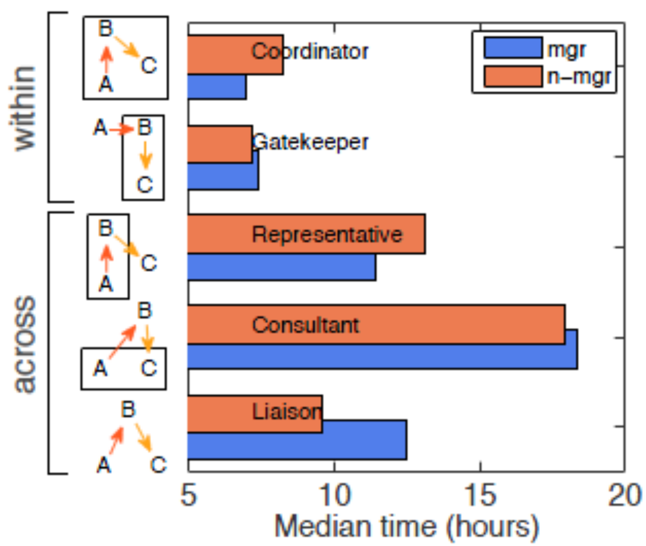


Sentiments in Conversation



Role Analysis

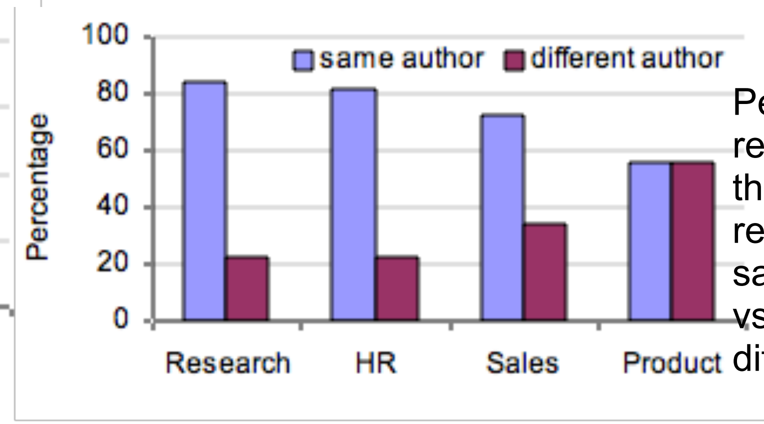




Role difference of normal behavior

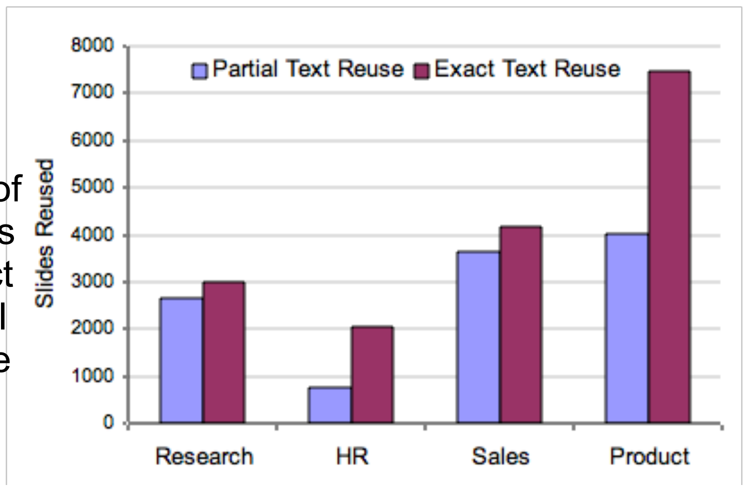
Information Reuse Behavior (CHI '11)

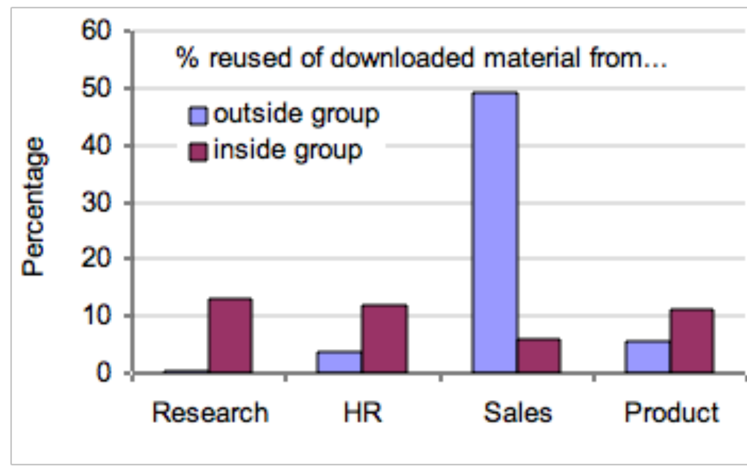
Percen-
tage of
slides
with
reused
content



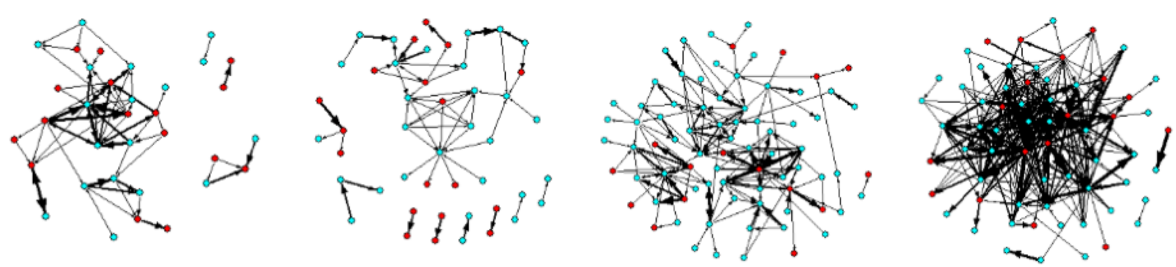
Percentage of
reused slides
that were
reused by the
same author
vs. by a
different author

Number of
slide pairs
with exact
vs. partial
text reuse



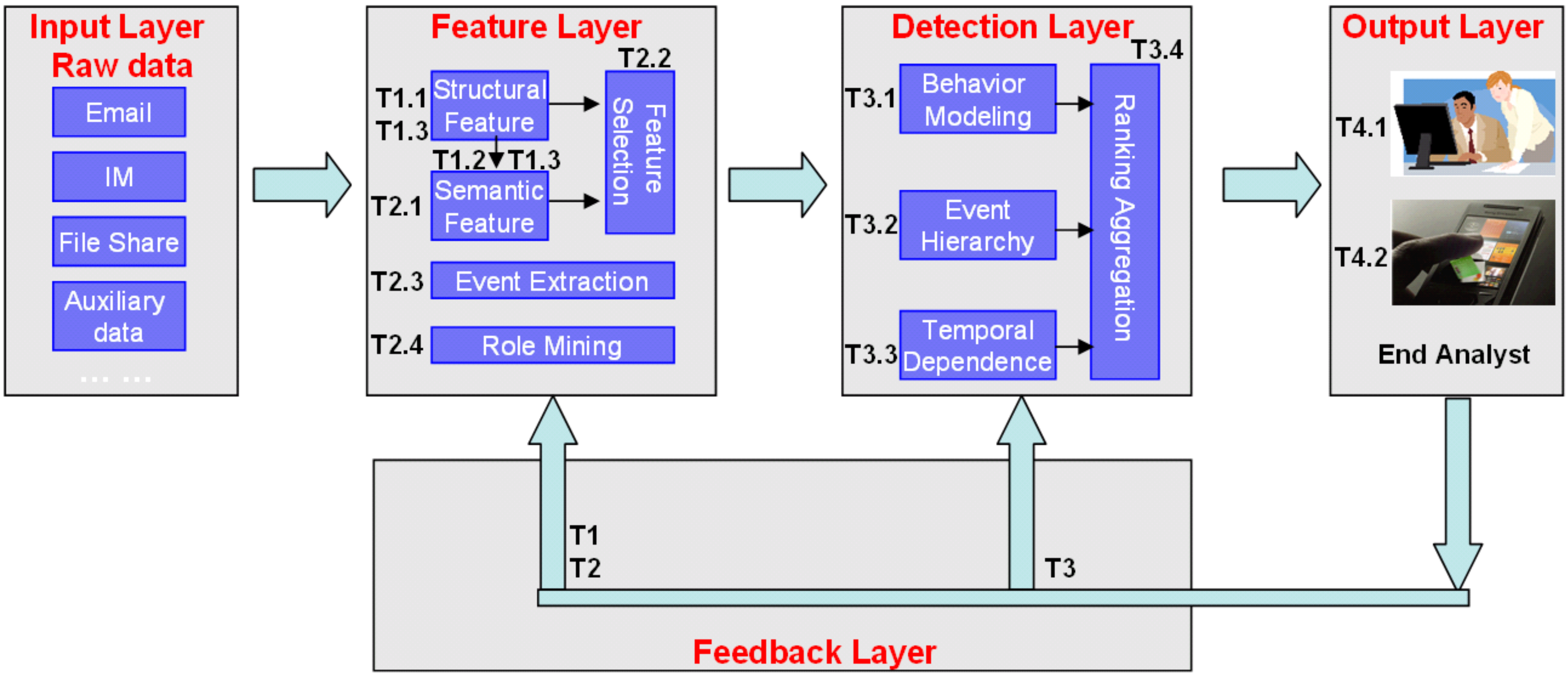


Percentage of
downloaded
material being
reused



(a) Research (b) HR (c) Sales (d) Product

- **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

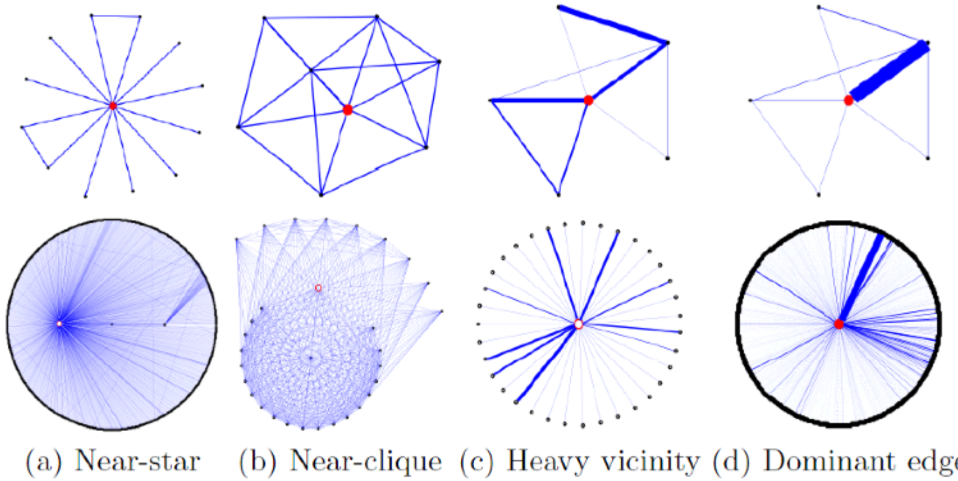
-- Address the 'lack-of-the ground-truth' issue by

(1) Interpretation friendly properties (e.g., non-negativity, sparseness, etc) into the current anomaly detection matrix factorization; and

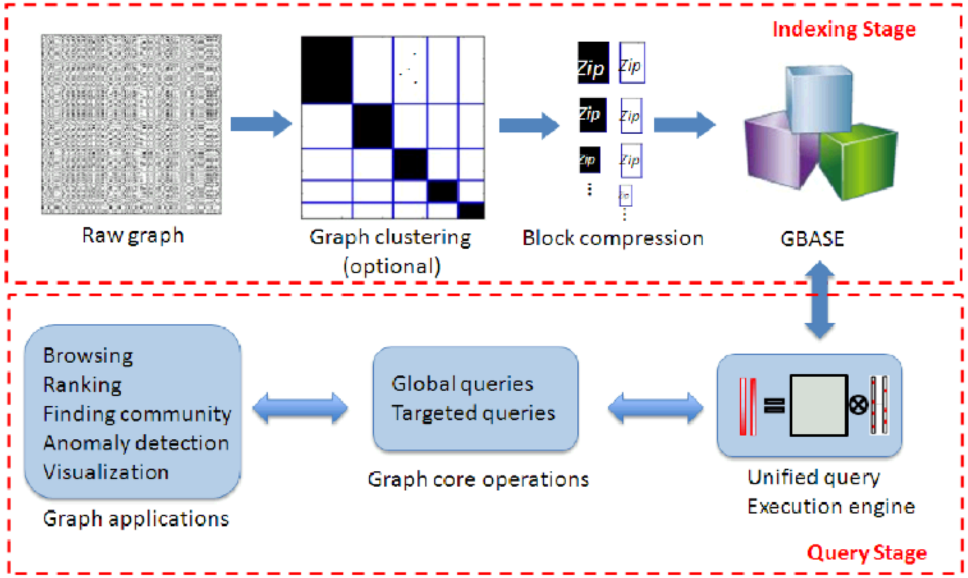
(2) providing some concise summarization to perform anomaly attribution.

▪**Thrust 3:** Infrastructure Support

-- General and scalable graph/network management system to process large network data, especially social and behavior of people / unique IPs.

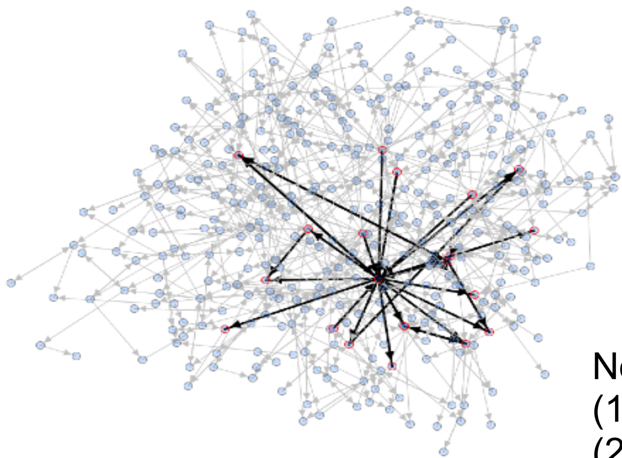


Typical abnormal nodes and their local ego-net structures

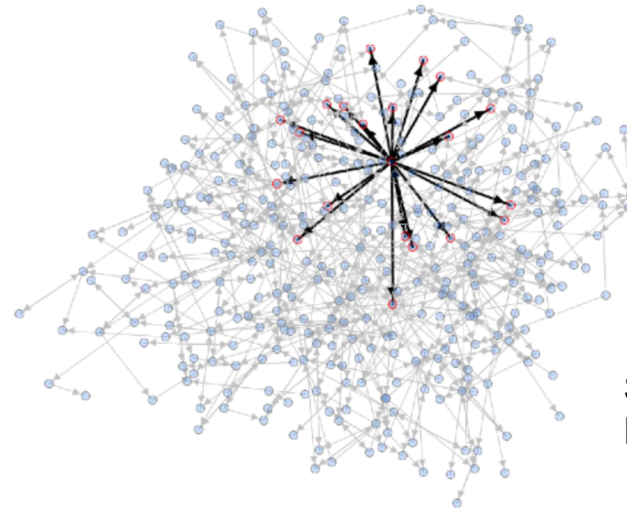


The overall flowchart of the graph management system

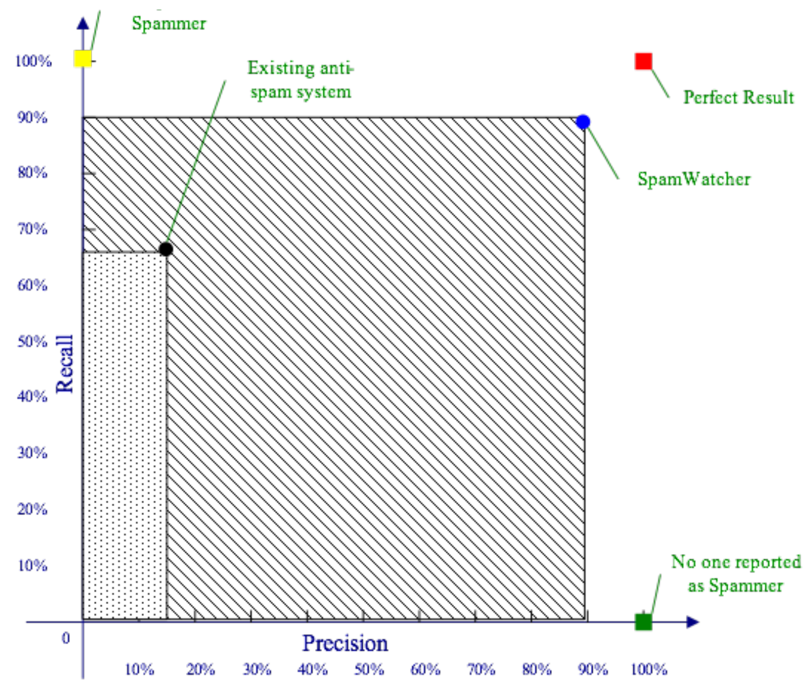
Use Case: Utilizing Social Network Analysis for Spam Detection



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



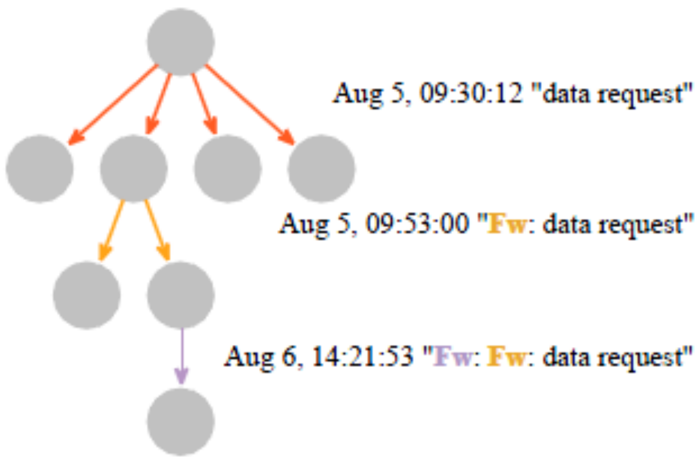
Spamming:
Near-Star



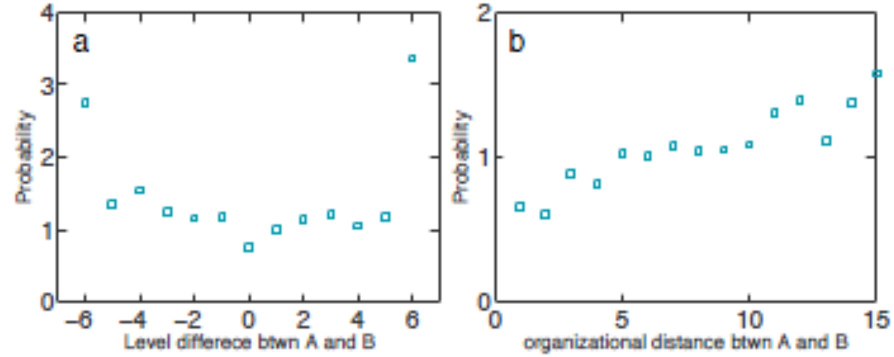
- A pilot project was done by CRL in a telecomm area of 6 million users in 2009.
- In experiment
  - Social Network Analysis is with recall of 89.97% and precision of 88.17% while comparison system is with 66.77% recall and 14.85% precision.
  - SNA's precision/recall area is 8 times larger

Anomaly Detection – information flow-based approach

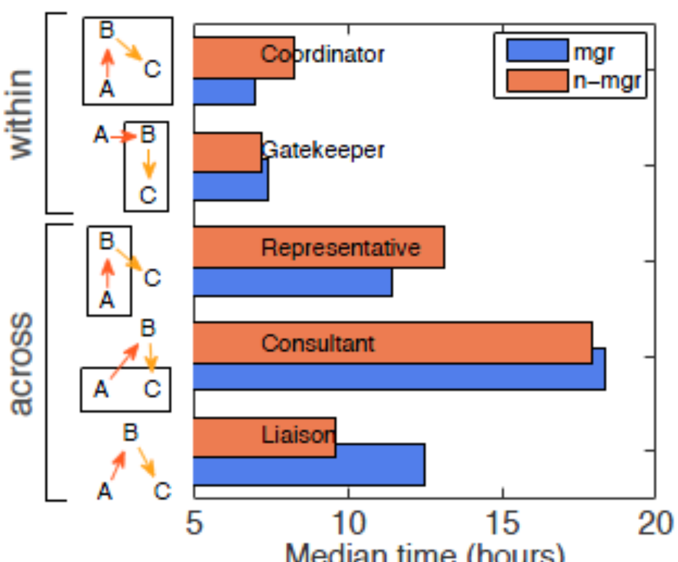
Columbia University

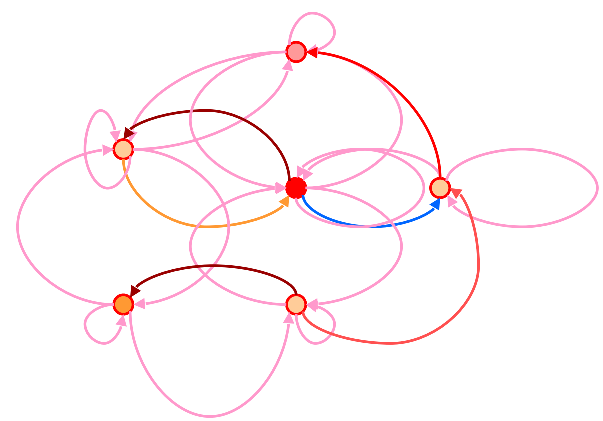


An illustrative example of an informa spreading tree. This tree is of size 8, width 4, depth 3.



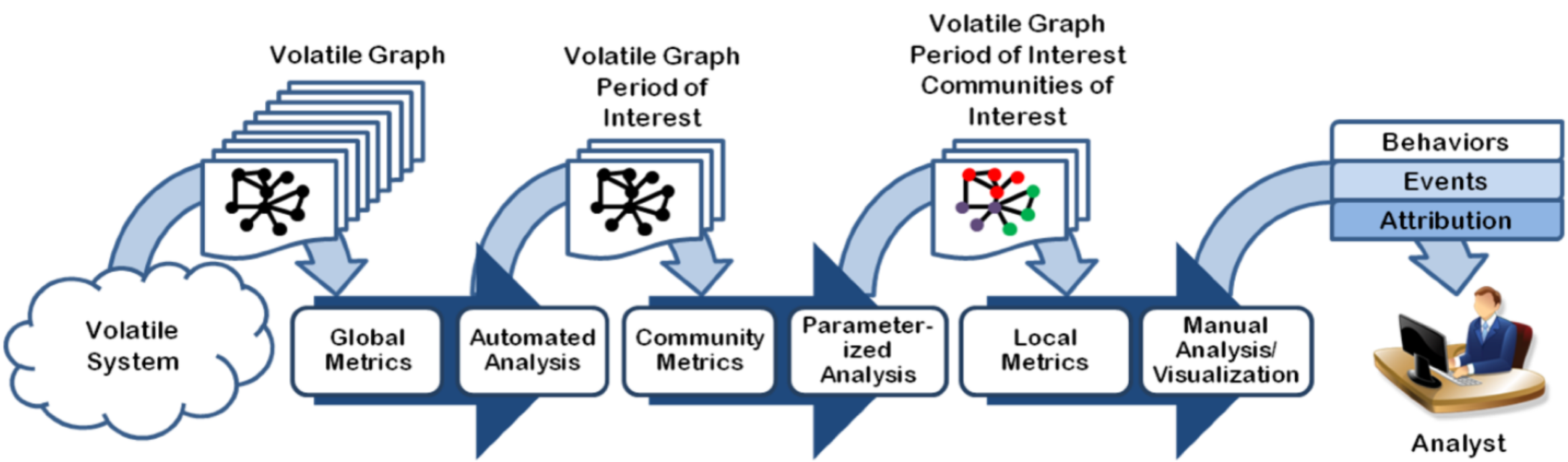
Probability ratio of email forwarding as a function of (a) hierarchical level difference and (b) organizational distance between initiators and spreaders. The information spreading exhibits some -homophily effect.





Large-Scale Graph Analysis

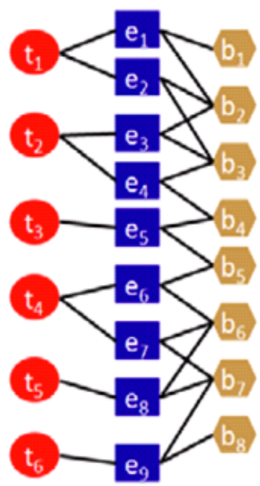
- T 1.1: Structure Feature Extraction
- T 1.2: Initial Filtering
- T 1.3: Scalability



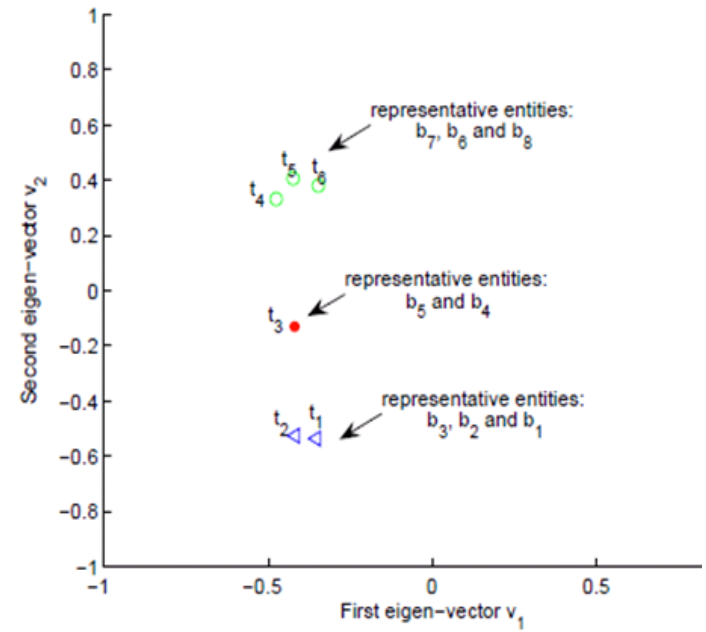
Time-to-Time (T3) Proximity Matrix Analysis

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



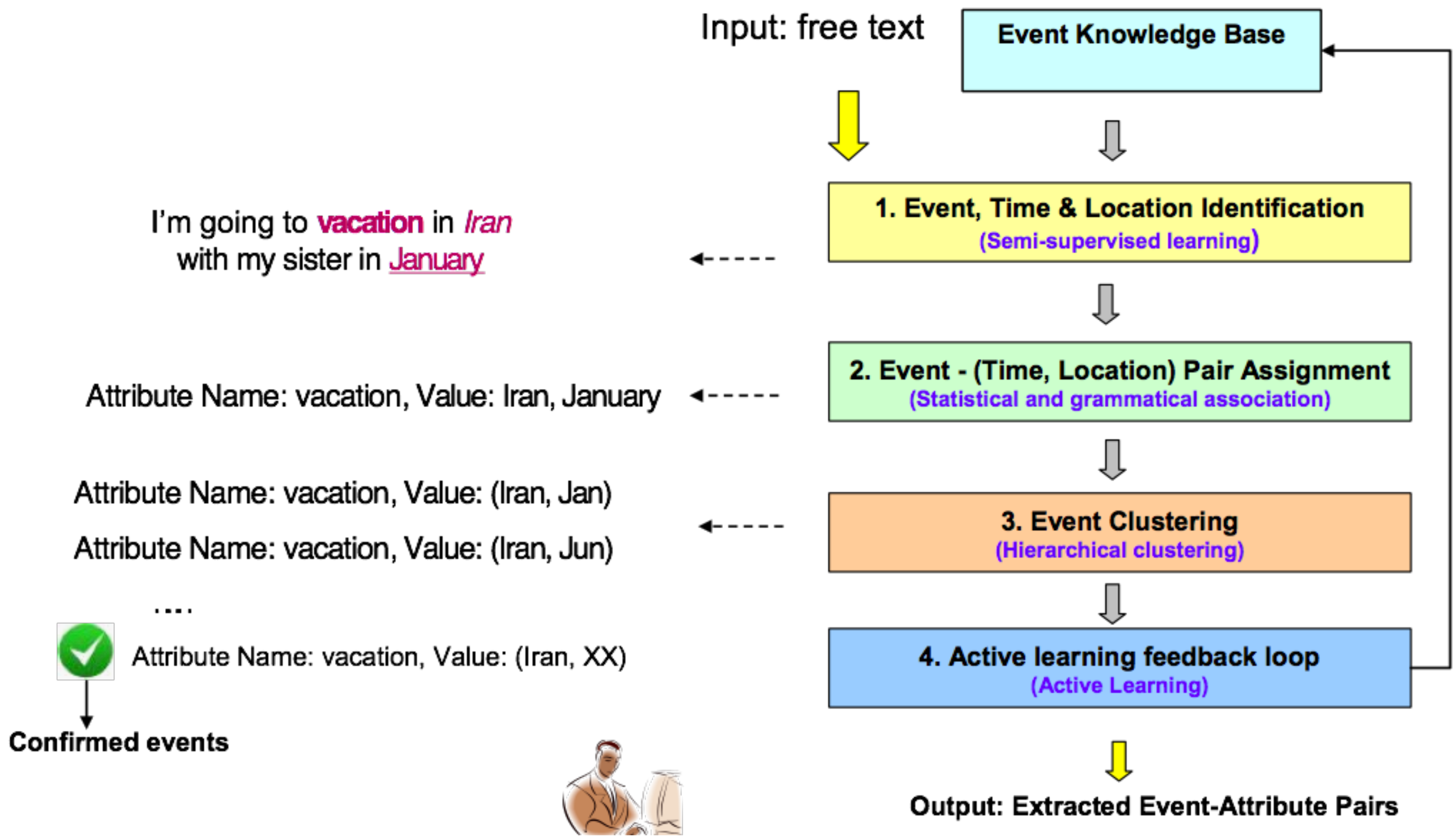
(b) Graph Representation



(c) Output of T3 (at finest level)

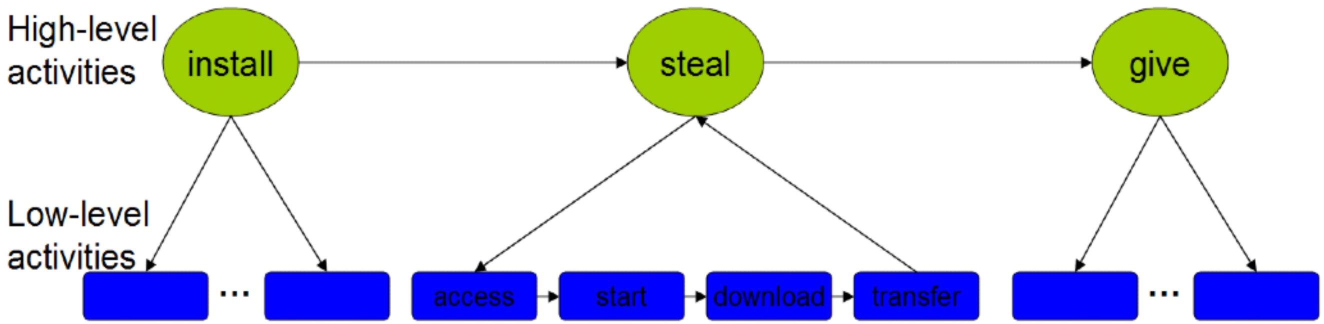
Thrust 2: Semantic Level Analysis

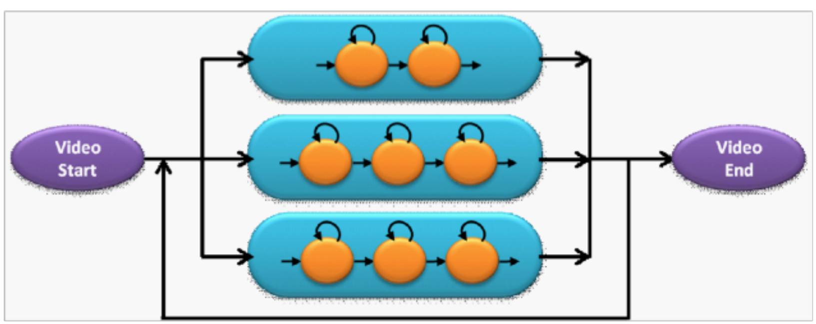
- T 2.1: Semantic Feature Extraction
- T 2.2: Feature Augmentation and Selection
- T 2.3: Abnormal Event Extraction
- T 2.4: Role Mining and Analysis



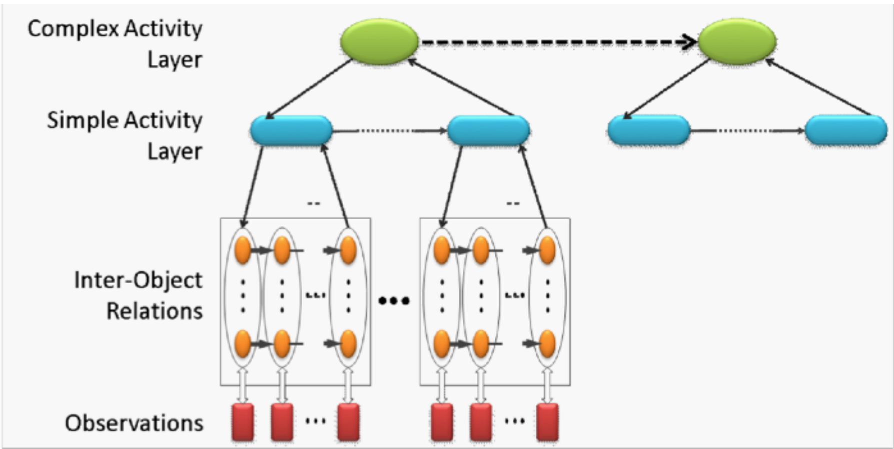
Thrust 3: Ranking and Aggregation

- T 3.1: User Behavior Modeling
- T 3.2: Hierarchical Event Analysis
- T 3.3: Temporal Dependency Analysis
- T 3.4: Anomaly Aggregation



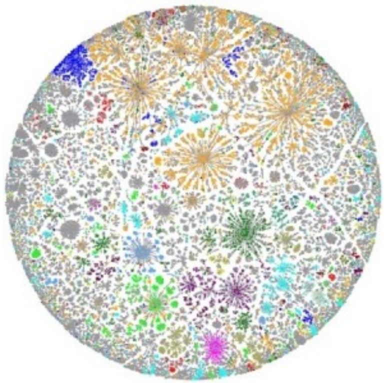


(a) Combined Composite Activity Model

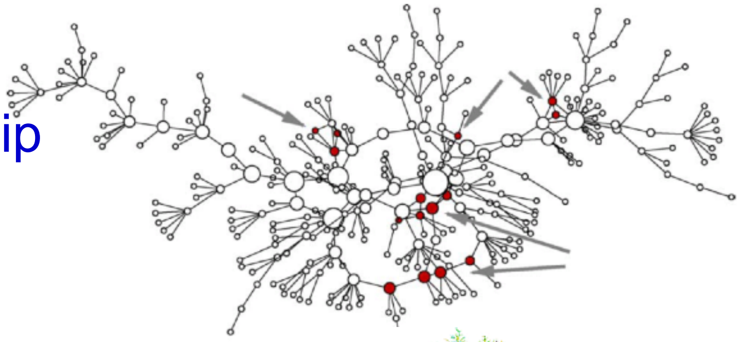


(b) Unrolled Composite Activity Model

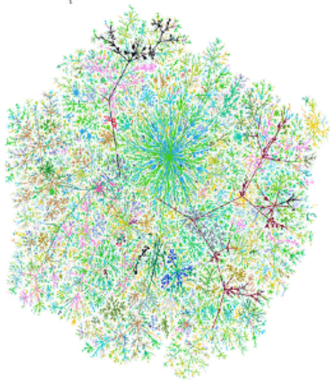
Example 1: Internet Map
Nodes: ISPs; Edges: Connection
(33K Nodes, 290K edges)



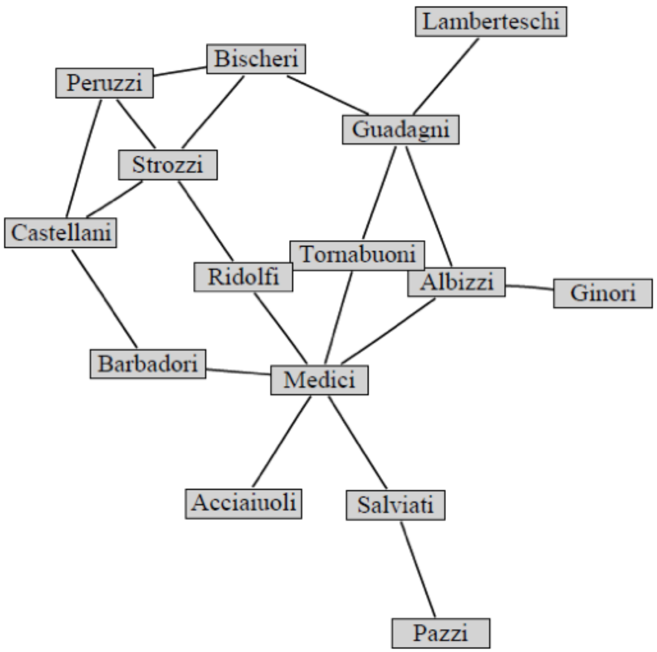
Example 2: Social Network
Nodes: People; Edges: Friendship
(FaceBook has 500M+ Users)



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]

“Who are the most important actors?”

Three centralities

Degree: # of neighbor

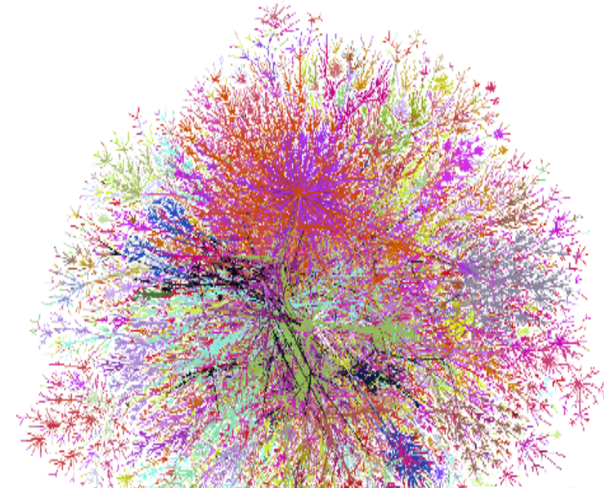
Closeness: avg. shortest path length

Betweenness: # of times a node sits between shortest path

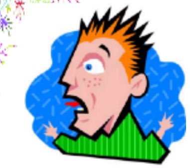
Application

Measuring the financial company value

Network attack monitoring



[Internet Web]

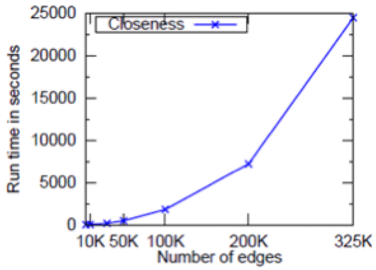


15th Century Florentine Family	Complexity	Internet Web
|V| = 15, ***|E| = 19***		***|V| = Billions***, ***|E| = Billions***
Degree : Easy	*O(|E|)*	***Degree : Easy***
Closeness : Easy	*O(|V|3)*	***Closeness : Hard***
Betweenness : Easy	*O(|V|2log|V|)*	***Betweenness : Hard***

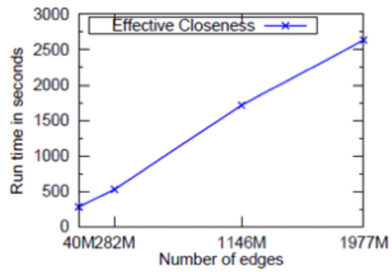
For 2 Billon Edges,
- standard closeness: 30,000 years

Network Analysis -- Effectiveness & Efficiency (GBase)

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



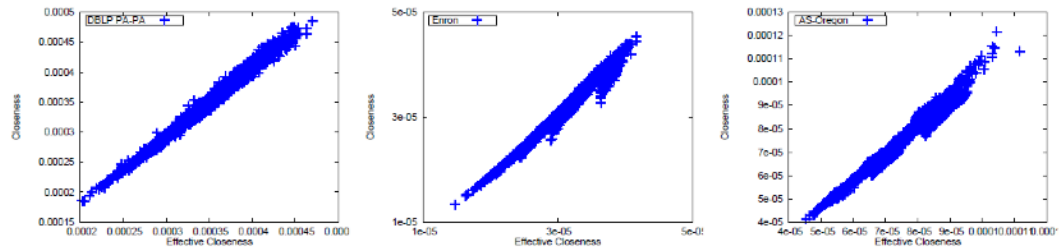
(a) Closeness: running time vs. edges



(c) Effective Closeness: running time vs. edges

Scalability Results

(Near-linear scalability)

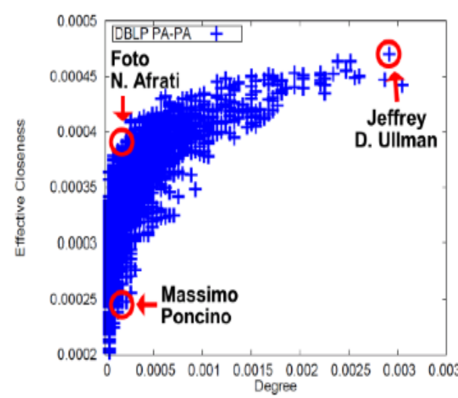


(a) DBLP Authors (b) Enron (c) AS-Oregon

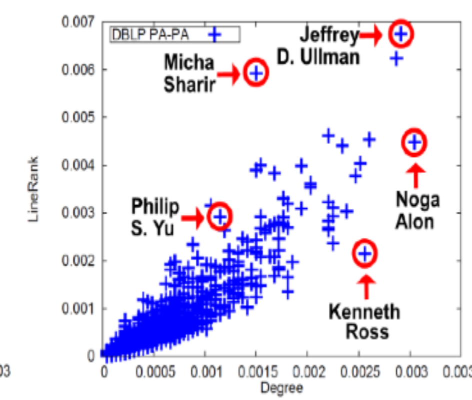
Effective Closeness vs. Closeness

(Near-linear correlation (≥97.8%)

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



(a) DBLP Authors: Effective Closeness vs. Degree

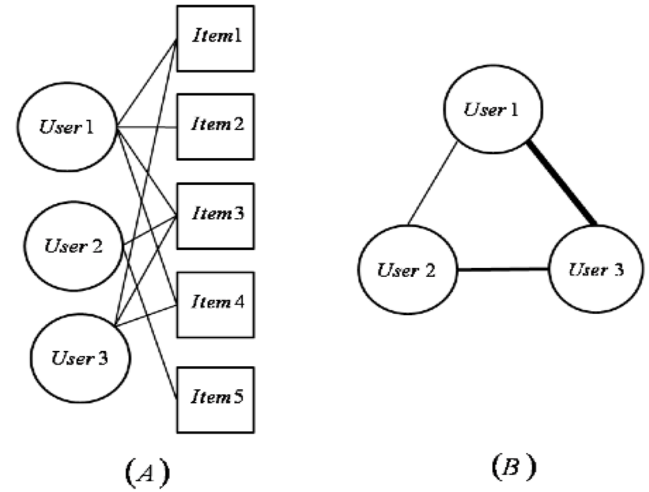


(b) DBLP Authors: LineRank vs. Degree

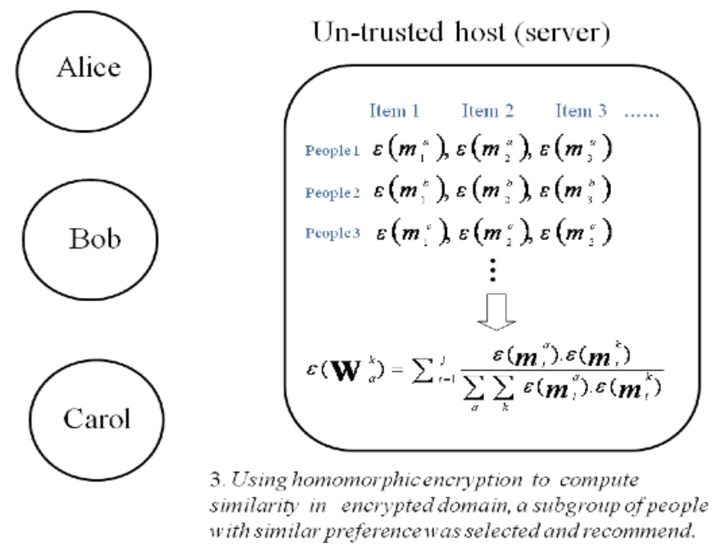
Analysis of Real-World Graph

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

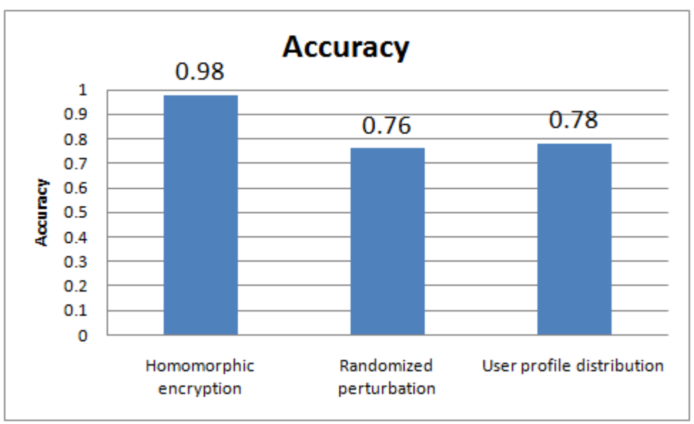
- A novel methodology & system for data mining and content/people recommendations
- New cryptographic method:
  - – Polynomial Ring Homomorphism, derived from Lattice-Based Cryptography
    - •Encrypted domain:
      - Addition
      - Multiplication
      - Division
- Key management protocol for:
  - – Encrypted Multi-Layer Ranking



Soft clustering for recommendations



Encrypted domain computation

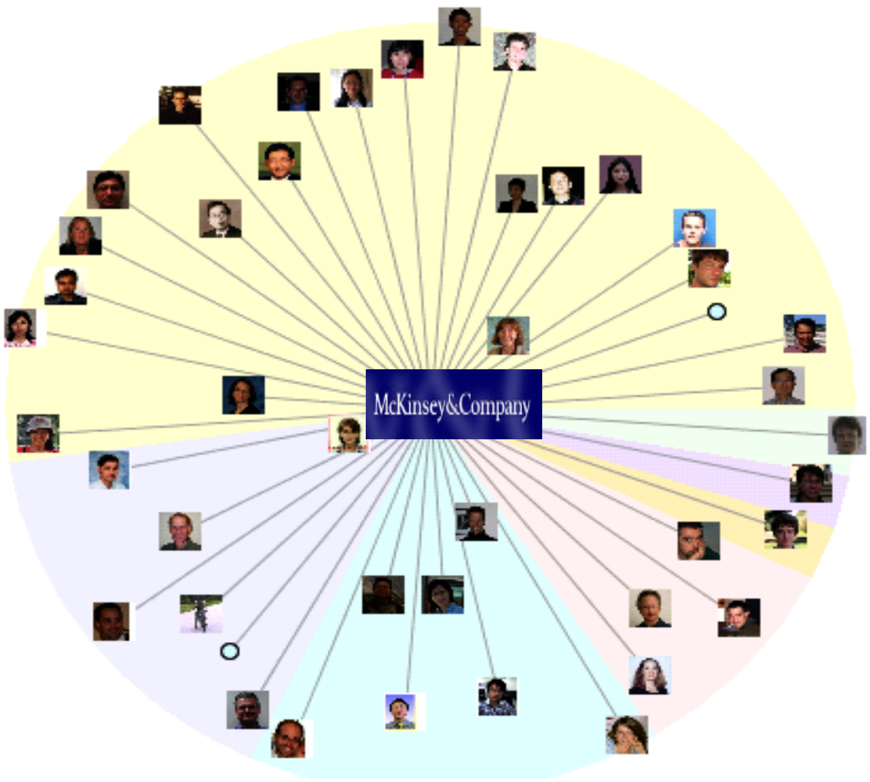


Social Network Analysis for Marketing and Sales

- **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 (VisblePath, 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?**
  - –Expand existing IBM Production System and Commercial Service Asset: SmallBlue / IBM Atlas for large-scale social network analysis, economic analysis, privacy-preserving data collection, marketing and sales expertise search and recommendation, etc, for Unica and Coremetrics platform.
  - –After successful completion, sharing these Analytics with other platforms based on business needs.

An example of utilizing micro- and macro social capital

- Who among Employees 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 Company, in terms of relationship strength, to other companies?



Flow of B2B Marketing using Social Network Analysis and Optimization

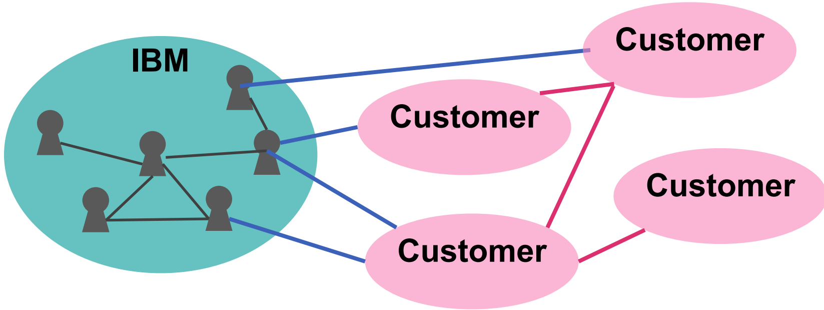
Social Network Analysis

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, Company 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 custome

It is a chance to match company A and diving .

  – 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 Company 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?

Multi-partite Network Analytics

Term Suggestion and Query Expansion

Document-based	Log-based	Ontology-based		Multi-partite network analytics
Influenced by test collection characteristics	Query log, failure for rare queries	WordNet	Wikipedia	Network community -based
	Click log, biased in favor of top ranks	Limited semantic relatedness	Simple concept links only	Extracting human factor
	Not publicly available	Difficult to update		Incorporate expertise

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.

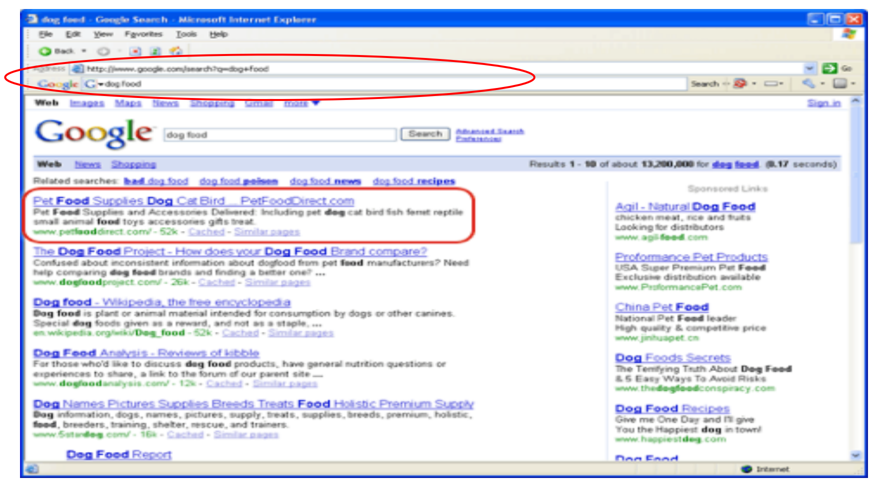
Multi-partite Network Analytics

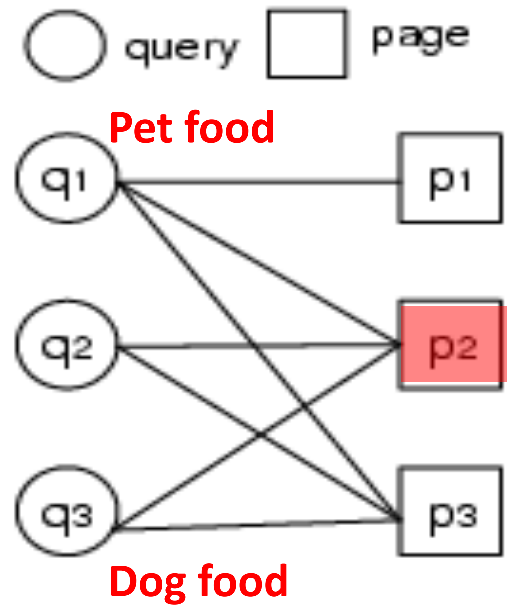
Term Suggestion and Query Expansion

Document-based	Log-based	Ontology-based		Multi-partite network analytics
Influenced by test collection characteristics	Query log, failure for rare queries	WordNet	Wikipedia	Network community -based
	Click log, biased in favor of top ranks	Limited semantic relatedness	Simple concept links only	Extracting human factor
	Not publicly available	Difficult to update		Incorporate expertise

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

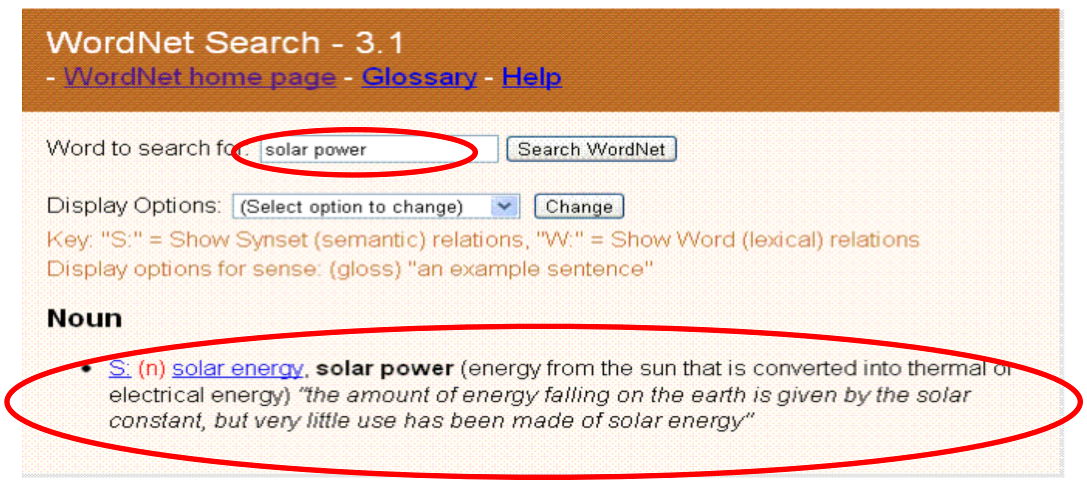
Term Suggestion and Query Expansion

Document-based → Log-based → Ontology-based → Multi-partite network analytics

- **Document-based**
  - Influenced by test collection characteristics
- **Log-based**
  - Query log, failure for rare queries
  - Click log, biased in favor of top ranks
  - Not publicly available
- **Ontology-based**
  - WordNet
    - Limited semantic relatedness
  - Wikipedia
    - Simple concept links only
  - Difficult to update
- **Multi-partite network analytics**
  - Network community -based
  - Extracting human factor
  - Incorporate expertise

WordNet as Ontology

- Manuallyconstructed 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

Log in / create account

Article Talk Read View source View history Search

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

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.[1]

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.

The PS10 concentrates sunlight from a field of heliostats onto a central tower.

Contents [hide]

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

Document-based → Log-based → Ontology-based → Multi-partite network analytics

Document-based

- Influenced by test collection characteristics

Log-based

- Query log, failure for rare queries
- Click log, biased in favor of top ranks
- Not publicly available

Ontology-based

- WordNet
- Wikipedia
- Limited semantic relatedness
- Simple concept links only
- Difficult to update

Multi-partite network analytics

- Network community -based
- Extracting human factor
- Incorporate expertise

Multi-partite Network Analytics

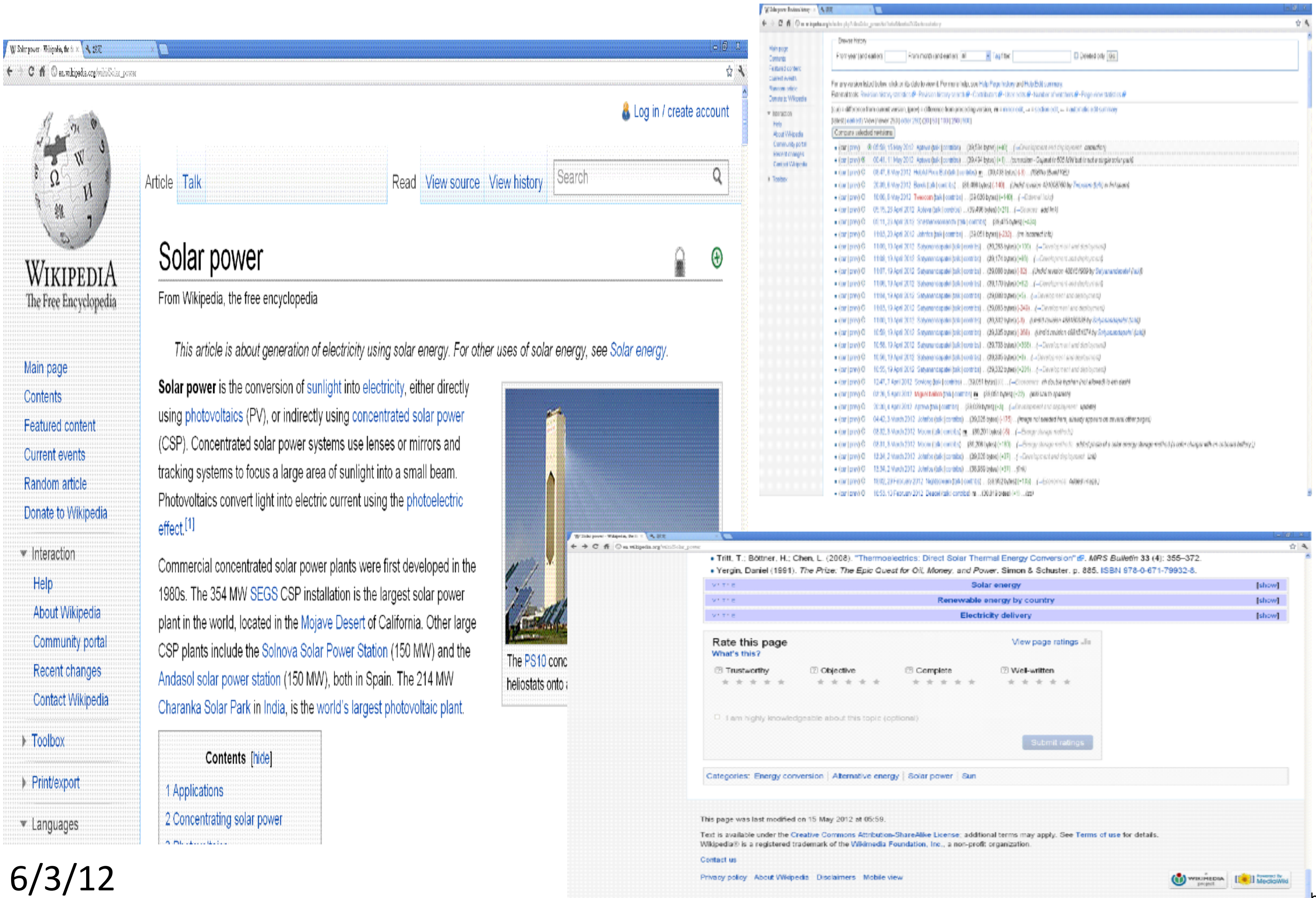
Columbia University

Term Suggestion and Query Expansion

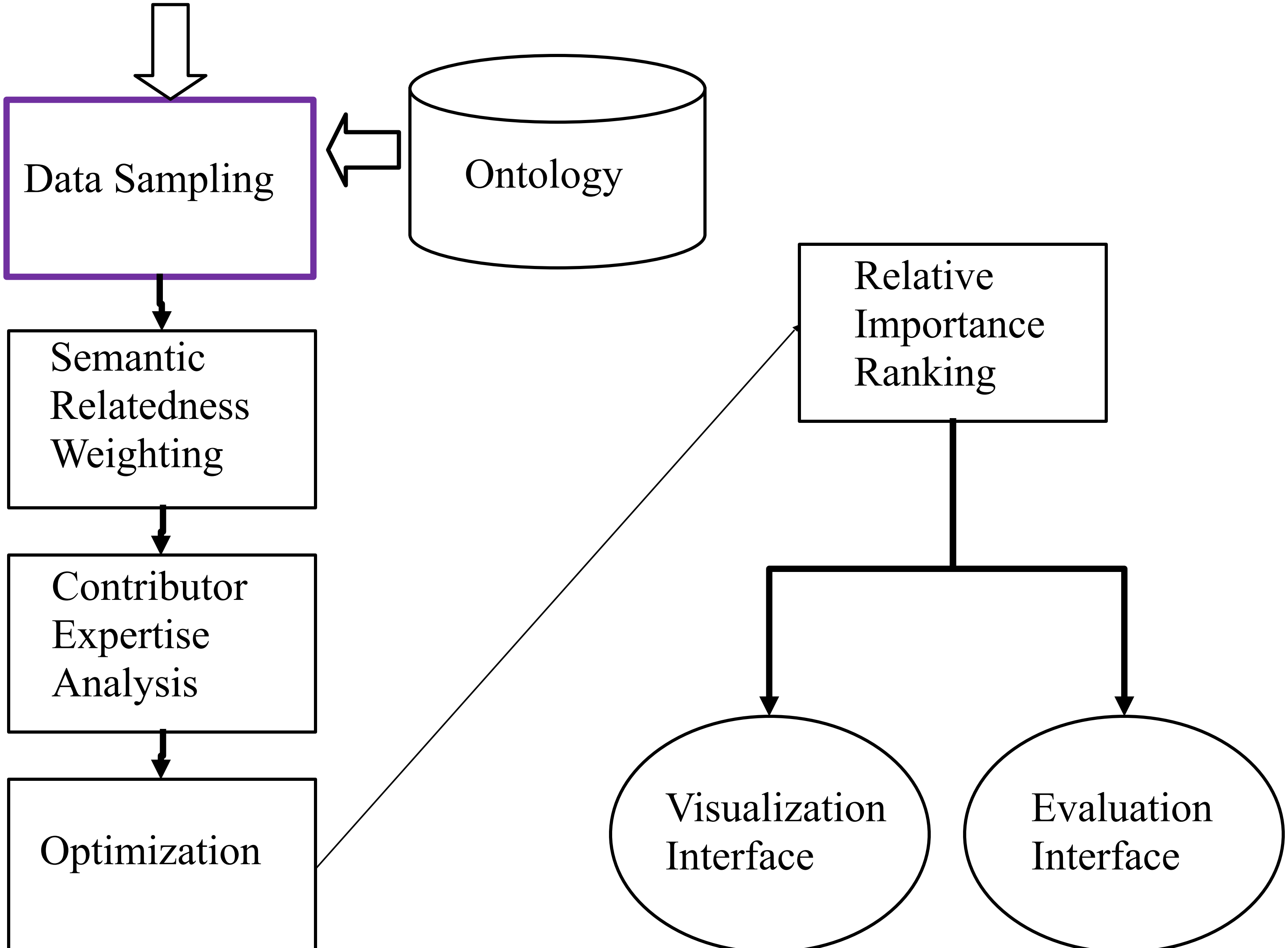
Document-based → Log-based → Ontology-based → Multi-partite network analytics

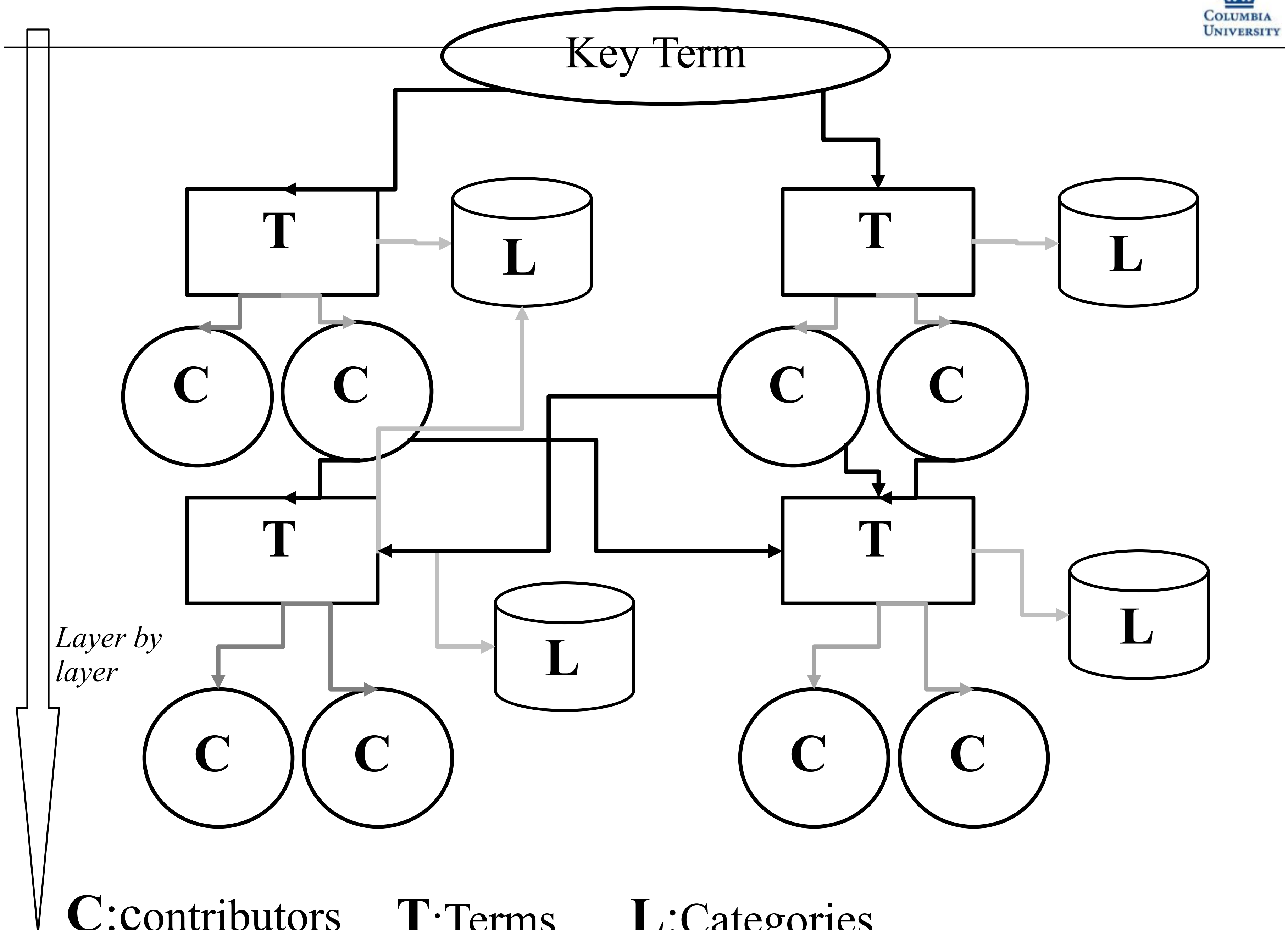
Document-based	Log-based	Ontology-based		Multi-partite network analytics
Influenced by test collection characteristics	Query log, failure for rare queries	WordNet	Wikipedia	Crawling is resource-intensive
	Click log, biased in favor of top ranks	Limited semantic relatedness	Simple concept links only	Human factor modeling
	Not publicly available	Difficult to update		Semantic relatedness difficult to evaluate

Wikipedia as Ontology



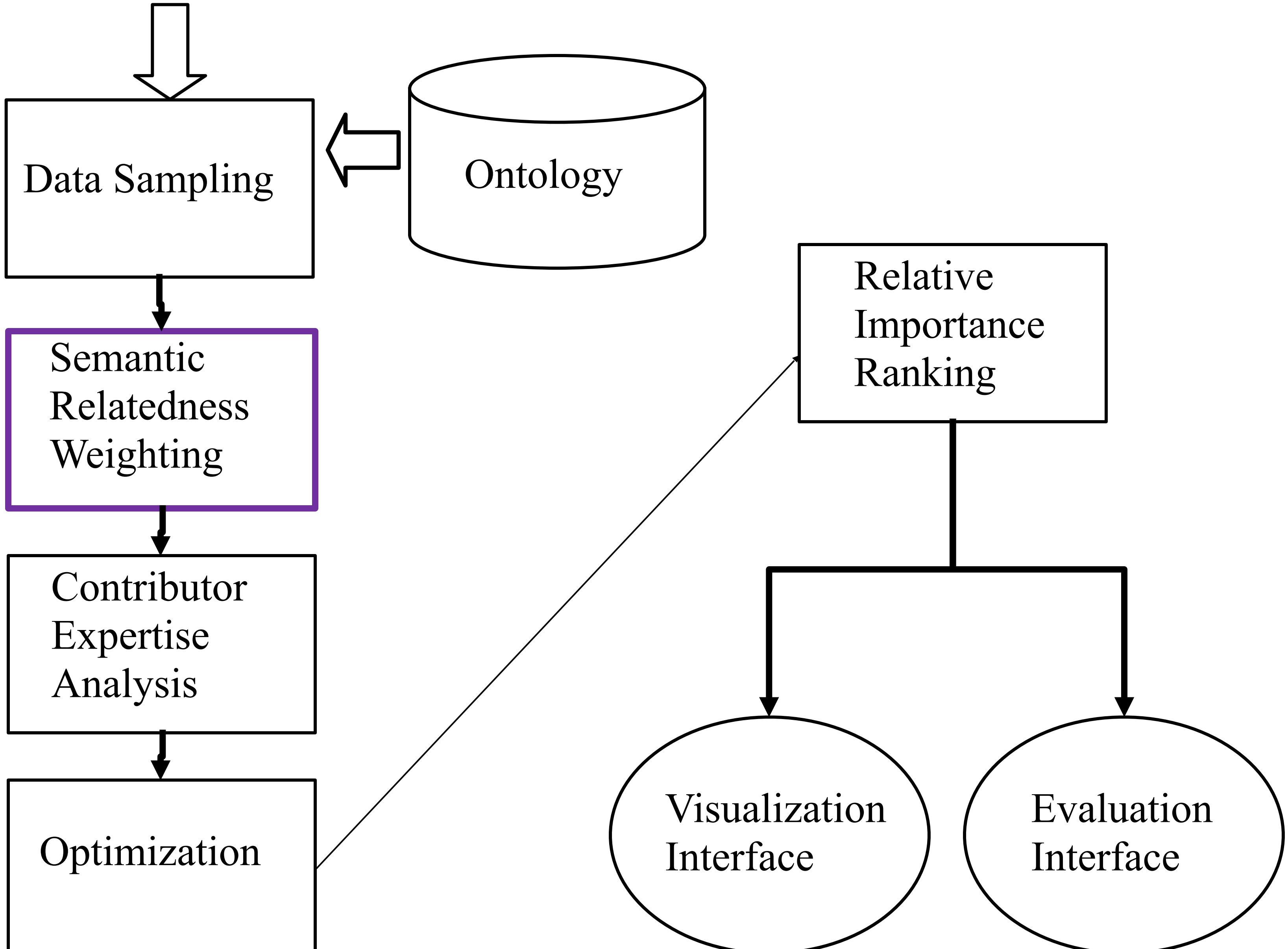
Query

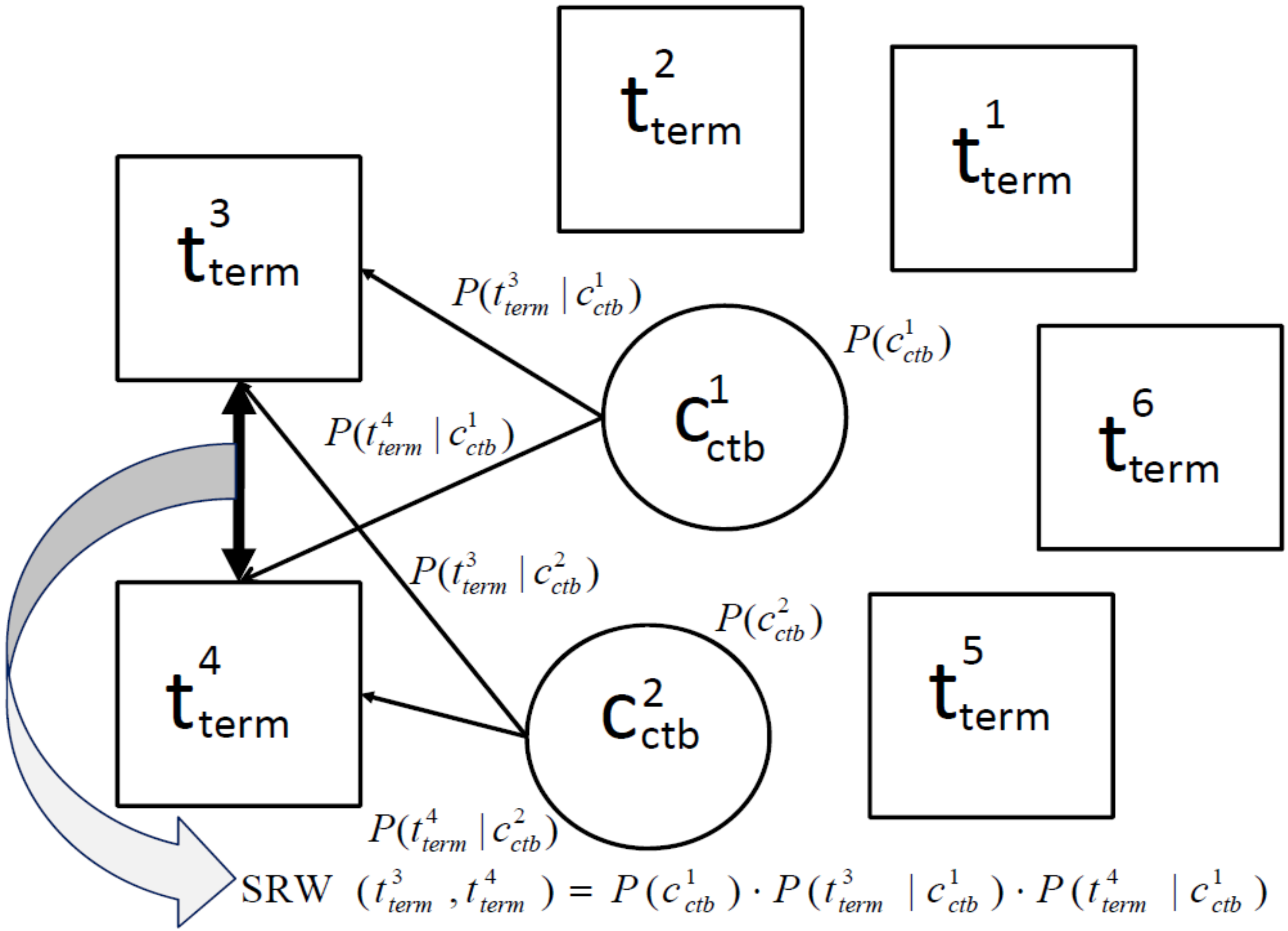

Data Sampling
Ontology
Semantic Relatedness Weighting
Contributor Expertise Analysis
Optimization
Relative Importance Ranking
Visualization Interface
Evaluation Interface



C:contributors **T**:Terms **L**:Categories

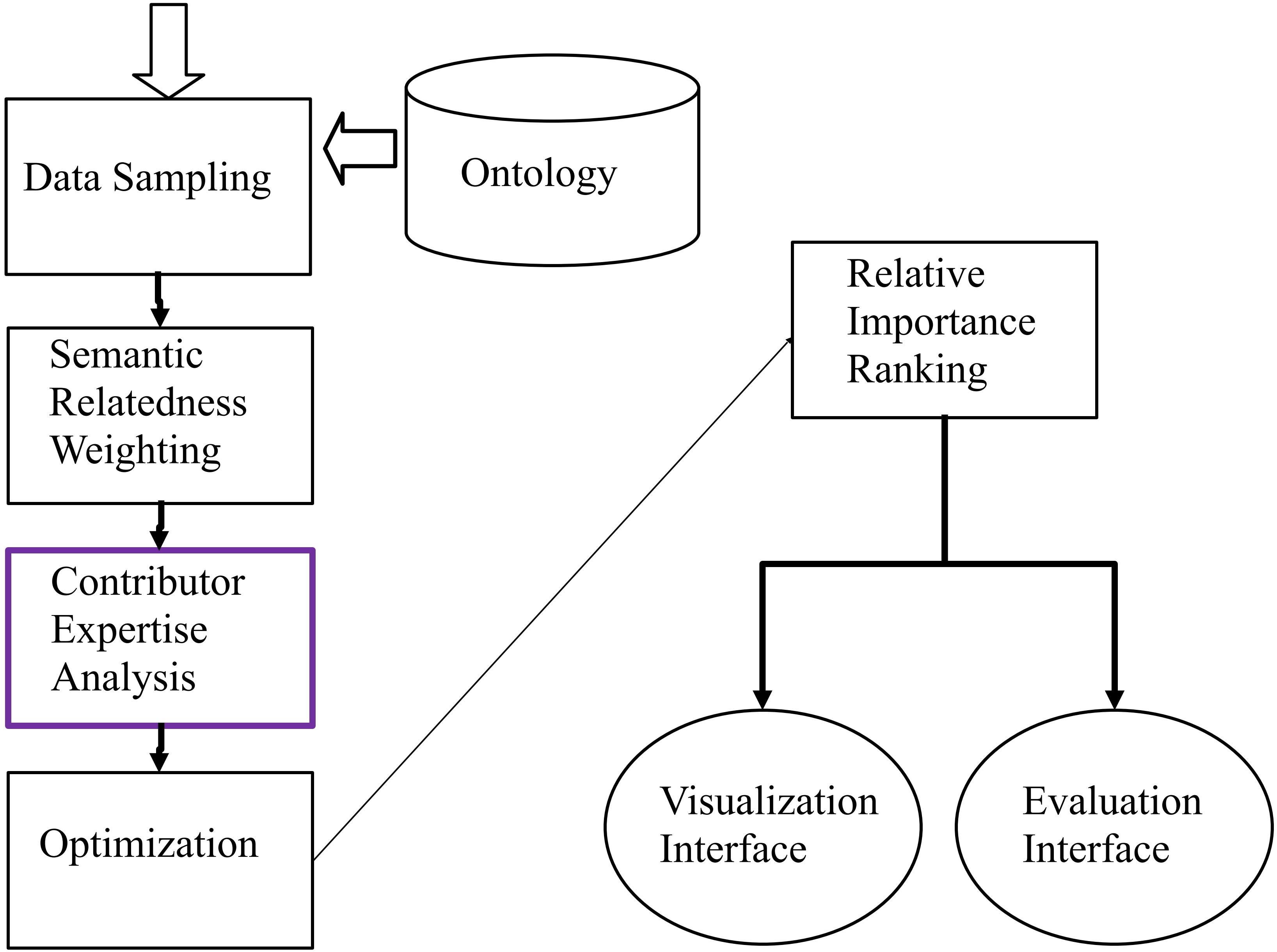
Query

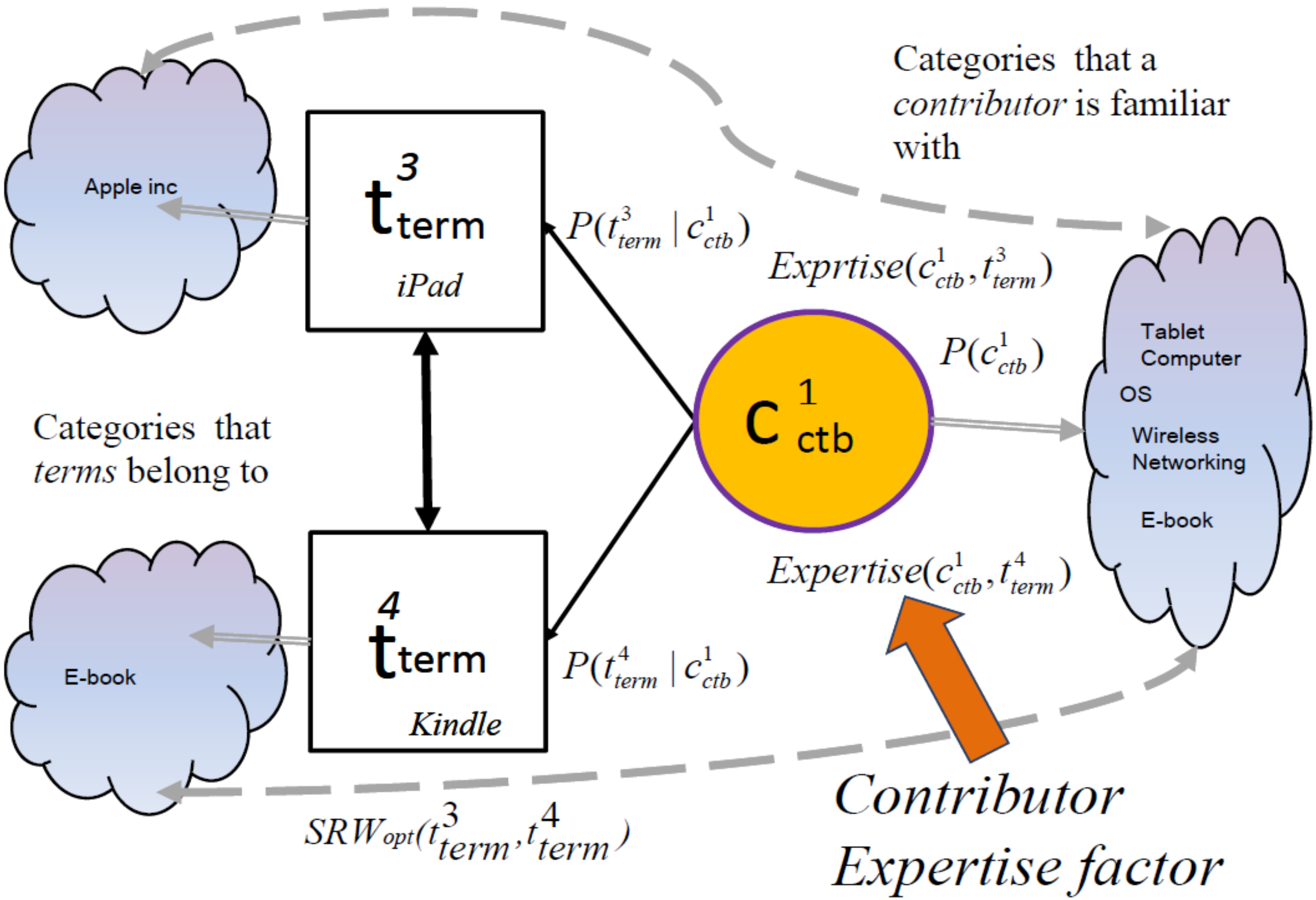

Data Sampling
Ontology
Semantic Relatedness Weighting
Contributor Expertise Analysis
Optimization
Relative Importance Ranking
Visualization Interface
Evaluation Interface

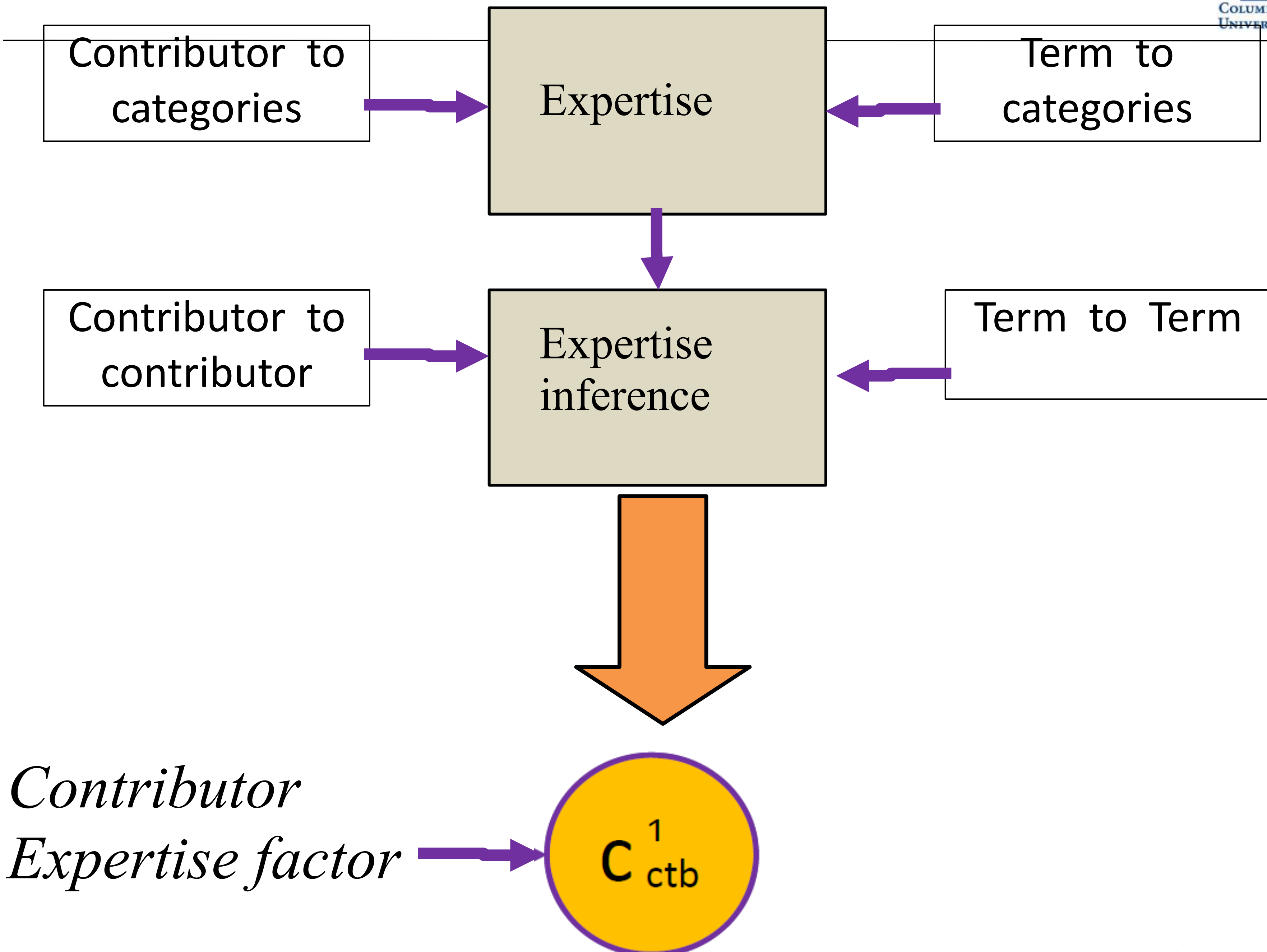


$$\mathrm{SRW}\,(t^3_{term}, t^4_{term}) = P(c^1_{ctb}) \cdot P(t^3_{term} \mid c^1_{ctb}) \cdot P(t^4_{term} \mid c^1_{ctb}) + P(c^2_{ctb}) \cdot P(t^3_{term} \mid c^2_{ctb}) \cdot P(t^4_{term} \mid c^2_{ctb})$$

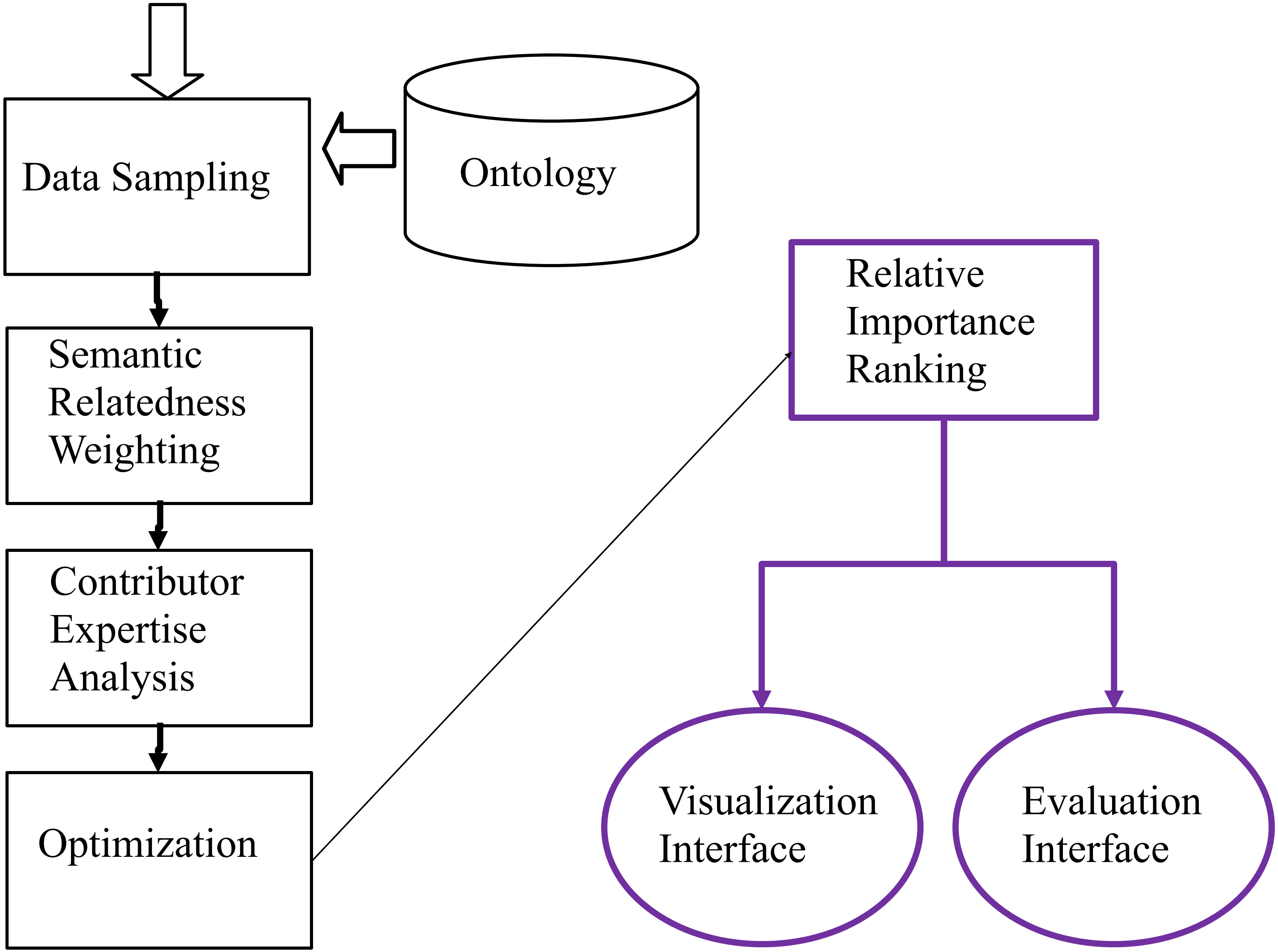
Query


Data Sampling
Ontology
Semantic Relatedness Weighting
Contributor Expertise Analysis
Optimization
Relative Importance Ranking
Visualization Interface
Evaluation Interface

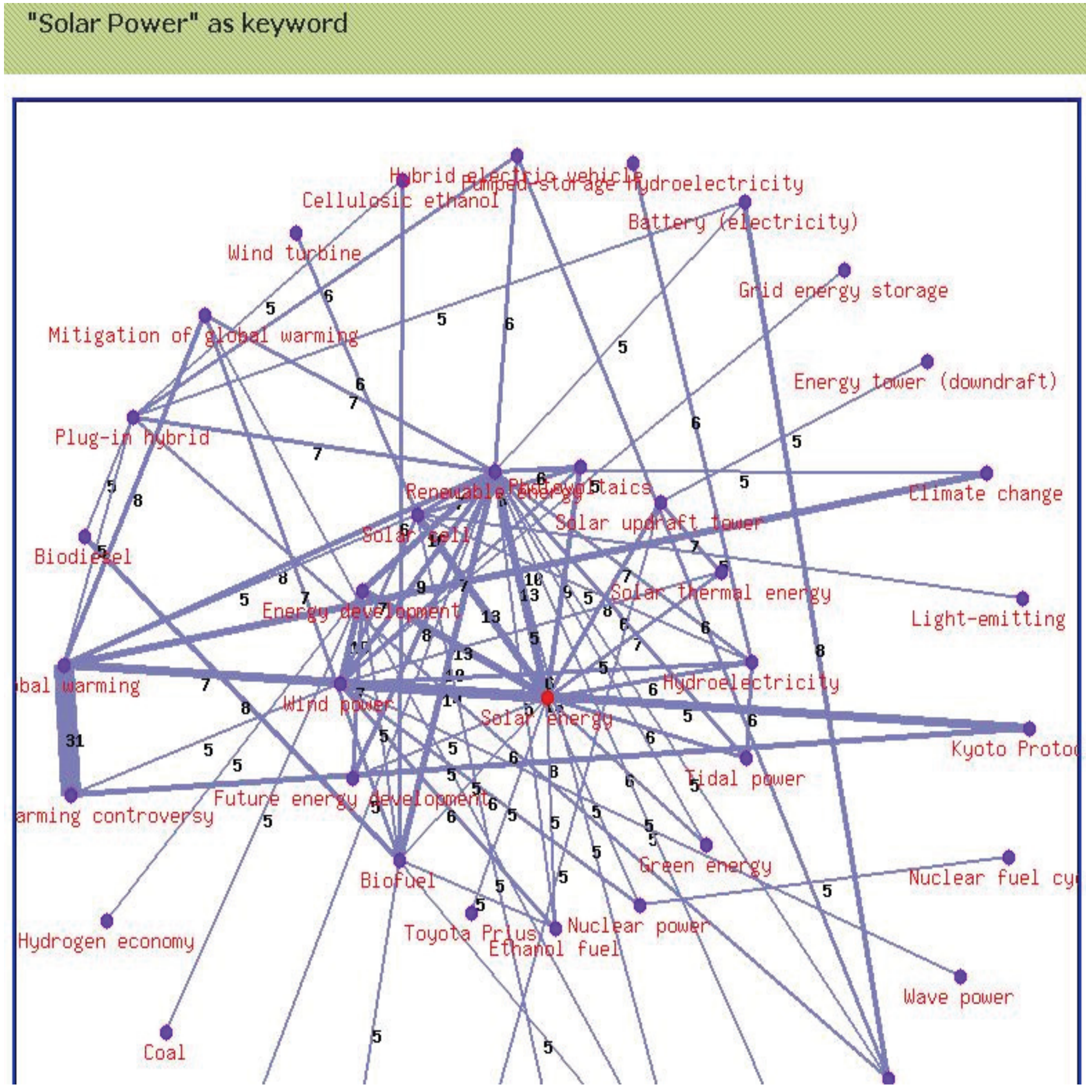

Categories that a *contributor* is familiar with
Apple inc
t^{3}_{term}
iPad
$P(t^{3}_{term} \mid c^{1}_{ctb})$
$Exprtise(c^{1}_{ctb}, t^{3}_{term})$
Tablet Computer
OS
Wireless Networking
E-book
$P(c^{1}_{ctb})$
c^{1}_{ctb}
Categories that *terms* belong to
$Expertise(c^{1}_{ctb}, t^{4}_{term})$
E-book
t^{4}_{term}
Kindle
$P(t^{4}_{term} \mid c^{1}_{ctb})$
$SRW_{opt}(t^{3}_{term}, t^{4}_{term})$
Contributor Expertise factor


Contributor to categories
Expertise
Term to categories
Contributor to contributor
Expertise inference
Term to Term
Contributor Expertise factor
C 1 ctb

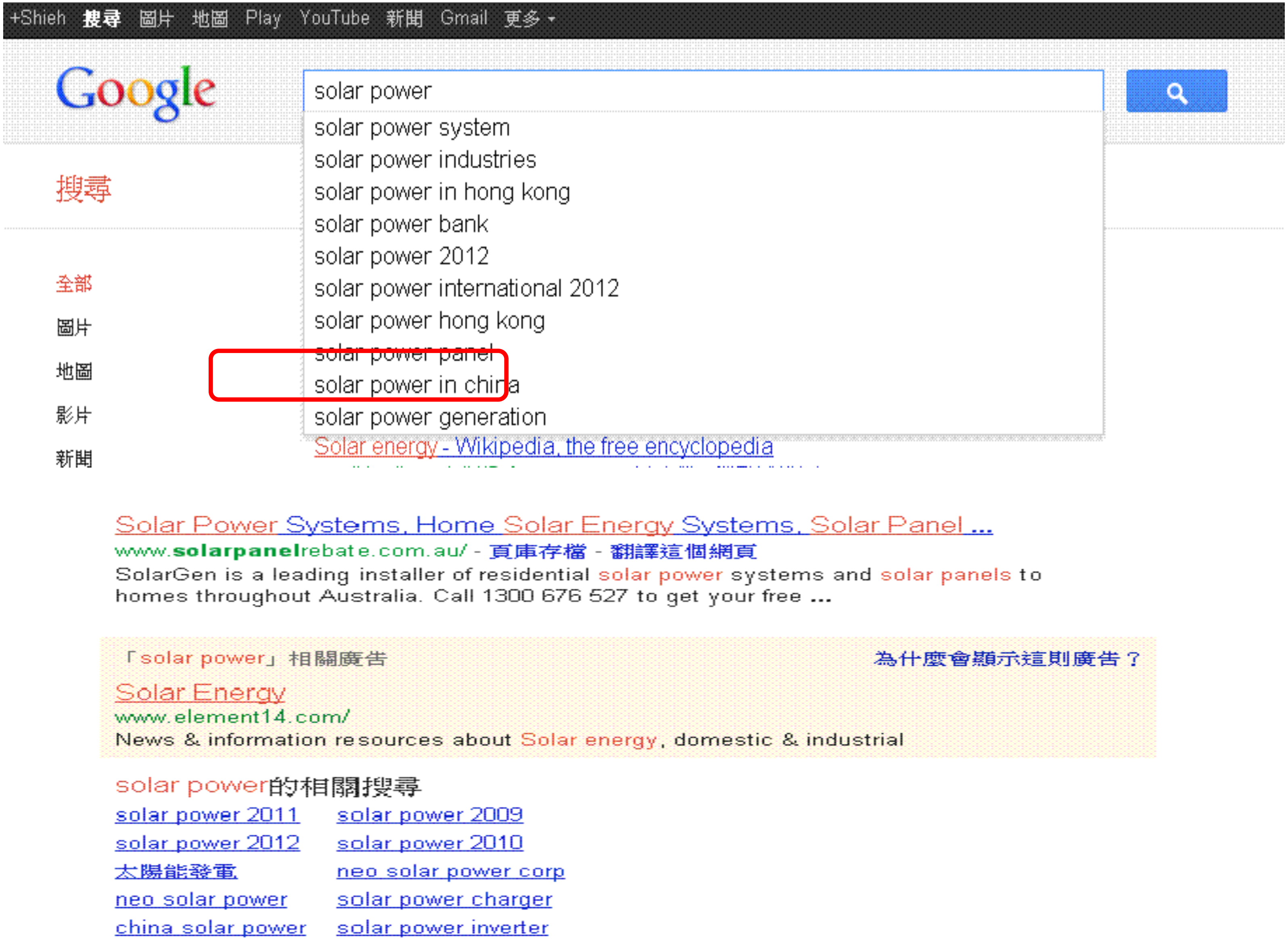
Query


Data Sampling
Ontology
Semantic Relatedness Weighting
Contributor Expertise Analysis
Optimization
Relative Importance Ranking
Visualization Interface
Evaluation Interface

High Semantic Relatedness Term Suggestion from Our System



Word-completion Term Suggestion



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).

