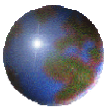
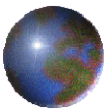


Opportunities for Statistics

Dennis Lin
University Distinguished Professor
Department of Supply Chain & Information Systems
The Pennsylvania State University



Business Intelligent

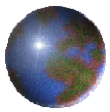
What is the same?

What is new?

What is the Same?

Elements of the exchange process

- A buyer
- A seller
- Buyer and seller can find each other
  (with some way to "authenticate" the other party)
- Each has something of value to offer the other
- They can voluntarily complete the exchange



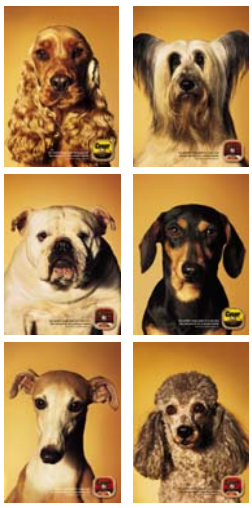
What is New?

- *e-business*
- *Globalization*

Link Analysis

- Owner

- Dog



What are the issues here?

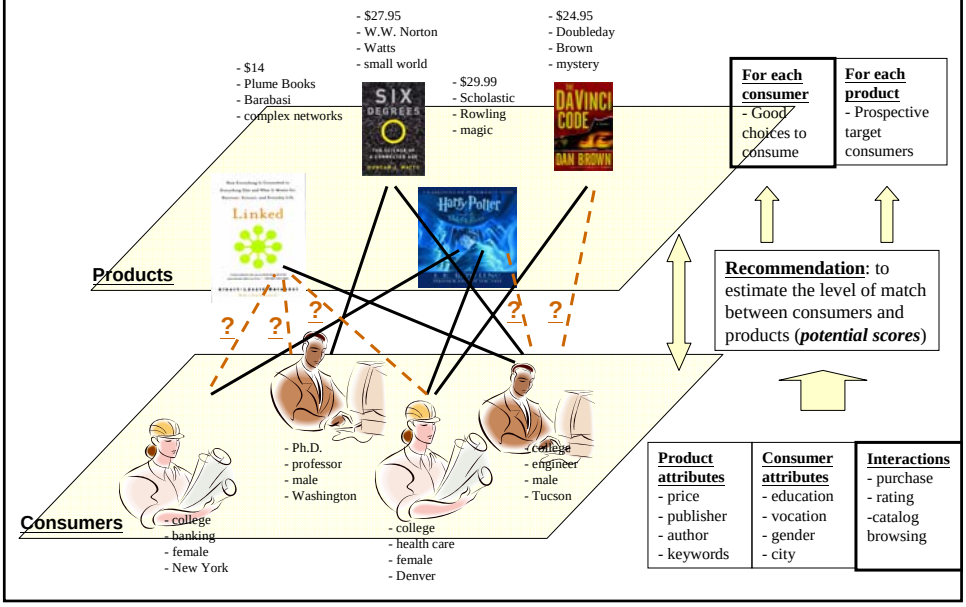
- How do you "quantify" (measure) your observations (people or dog)?
- How do you "characterize" your observations?
- How do you classify (match) them?
- Others?

Recommender Systems

Zan Huang

- Recommender systems
  - Automatically recommend items to users based on product attributes, consumer attributes, and consumers' implicit or explicit feedback on products [Resnick et al. 1994; Shardanand and Maes 1995; Hill et al. 1995; Resnick and Varian 1997]
  - Products: Discussion postings, webpages, movies, jokes, news, research papers, books, etc.
  - Consumers: Information seekers, online shoppers, etc.

The Recommendation Problem



How to measure/characterize a unipartite graph?

Theoretical Prediction of Unipartite Graphs Projected from Random Bipartite Graphs

- Average degree

$$z = G_0'(1)$$

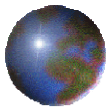
- Average path length

$$L = 1 + \frac{\ln(N / G_0'(1))}{\ln\left(\left(\frac{f_0''(1)}{f_0'(1)}\right)\left(\frac{g_0''(1)}{g_0'(1)}\right)\right)}$$

- Triangle clustering coefficient

$$C_\Delta = \frac{M}{N}\frac{g_0'''(1)}{G_0''(1)}$$

- The predictions for the product graph can be derived similarly by interchanging f and g and interchanging M and N



How to measure/characterize a bipartite graph?

None we know of!!!

How to measure/characterize a bipartite graph?

- Common Practices
- Projection into two uni-partites
  - Is this a right way to do?
  - Projection loss?

Time Series: Univariate Time Series

135, 143, 127, 129, …, 172 What's next?

$y_1, y_2, \ldots, y_T$ What's $\hat{y}_{T+1}$?

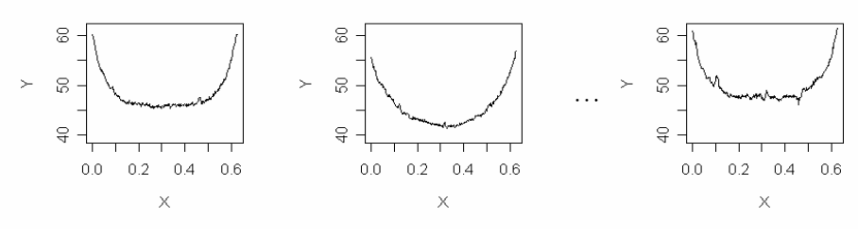
Model Building and Forecasting (short term/long term)
Monitoring (Quality Assurance)

Time Series: Multivariate Time Series

$$\begin{pmatrix} 5 \\ 143 \\ 85 \\ 46 \end{pmatrix}, \begin{pmatrix} 12 \\ 174 \\ 77 \\ 39 \end{pmatrix}, \ldots, \begin{pmatrix} 16 \\ 191 \\ 81 \\ 41 \end{pmatrix}$$ What's next?

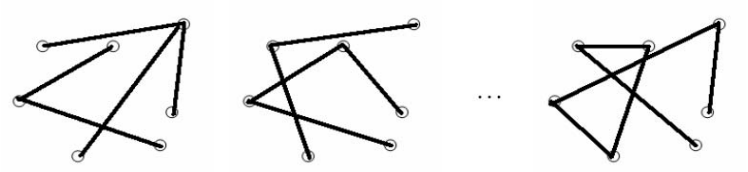
$\vec{y}_1, \vec{y}_2, \ldots, \vec{y}_T$ What's $\hat{\vec{y}}_{T+1}$?

Time Series: Functional Data (Profile)



$f_1, f_2, \ldots, f_T$ What's $\hat{f}_{T+1}$?

Time Series: Graphic/Network (Link Prediction)



$G_1, G_2, \ldots, G_T$ What's $\hat{G}_{T+1}$?

Process Mining

case identifier	task identifier
case 1	task A
case 2	task A
case 3	task A
case 3	task B
case 1	task B
case 1	task C
case 2	task C
case 4	task A
case 2	task B
case 2	task D
case 5	task E
case 4	task C
case 1	task D
case 3	task C
case 3	task D
case 4	task B
case 5	task F
case 4	task D

Table 1. A process log.

Process Mining

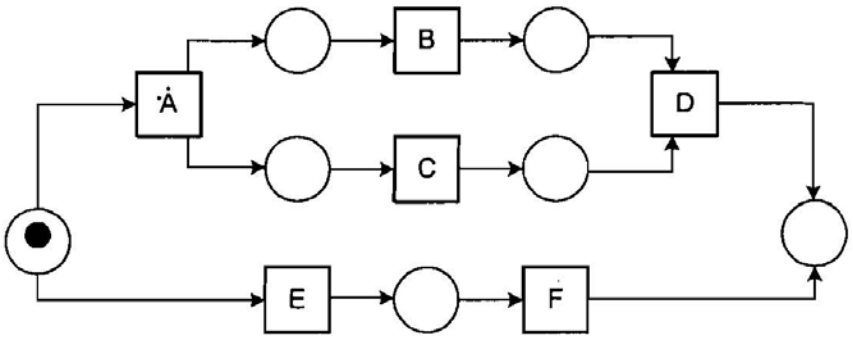


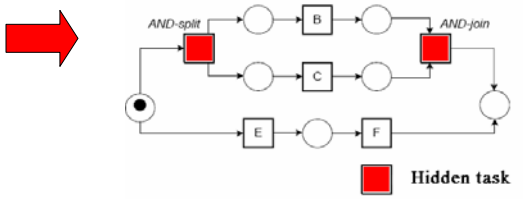
Fig. 1. A process model corresponding to the process log.

Discovering Processes from Event Logs

- Challenges of process mining
  - ***Mining process models***
    1. Mining hidden tasks
    2. Mining duplicate tasks
    3. Mining non-free-choice constructs
    4. Mining loops
  - Robust mining
    5. Dealing with noise
    6. Dealing with incompleteness
  - Process analysis
    7. Using time
    8. Mining different perspectives
    9. Delta analysis
  - Others
    10. Gathering data from heterogeneous sources
    11. Visualizing results

Mining Process Model- Mining Hidden Tasks

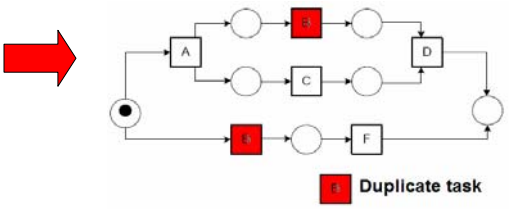
Case identifier	Task identifier
~~Case 1~~	~~Task A~~
~~Case 2~~	~~Task A~~
~~Case 3~~	~~Task A~~
Case 3	Task B
Case 1	Task B
Case 1	Task C
Case 2	Task C
~~Case 4~~	~~Task A~~
Case 2	Task B
~~Case 2~~	~~Task D~~
Case 5	Task E
Case 4	Task C
~~Case 1~~	~~Task D~~
Case 3	Task C
~~Case 3~~	~~Task D~~
Case 4	Task B
Case 5	Task F
~~Case 4~~	~~Task D~~



Even task A is removed from the log, it is clear that there has to be an AND-split if we assume tasks B and C to be in parallel. Similar for D.

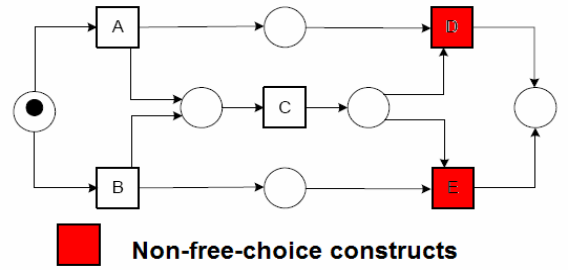
Mining Process Model- Mining Duplicate Tasks

Case identifier	Task identifier
Case 1	Task A
Case 2	Task A
Case 3	Task A
Case 3	Task B
Case 1	Task B
Case 1	Task C
Case 2	Task C
Case 4	Task A
Case 2	Task B
Case 2	Task D
Case 5	Task E → B
Case 4	Task C
Case 1	Task D
Case 3	Task C
Case 3	Task D
Case 4	Task B
Case 5	Task F
Case 4	Task D



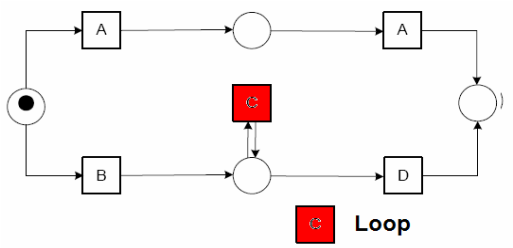
Questionable process model

Mining Process Model-Mining Non-free-choice Constructs



The choice between task D and task E is decided not only by their immediately precedent, but also some earlier choices (A and B). Such constructs are difficult to mine since the choice is non-local and the mining algorithm has to ‘‘remember’’ earlier events.

Mining Process Model-Mining loops



Mining loops can be difficult if the loops includes many tasks (long span) and involve splits/joins.

Algorithms for Process Mining

- Most of existing data mining techniques can't apply to process mining directly.
- So far, the emphasis is on mining process models.
- The inductive bias during process mining algorithm
  - Affected by process modeling languages because of representation limitation of the languages (PetriNet, UML, Block Diagram, EPC, WSBPEL, etc.).
- Local/global strategies
  - local strategies primarily based on a step by step building of the optimal process model based on very local information.
  - Global strategies primarily based on a one strike search for the optimal model.

Potential Subjects

- Finding General Structure/Network
- Finding systematic pattern
- Identify Unusual Observation (Transaction)
- Model Building

  Data/Table ←→ Model/Graph

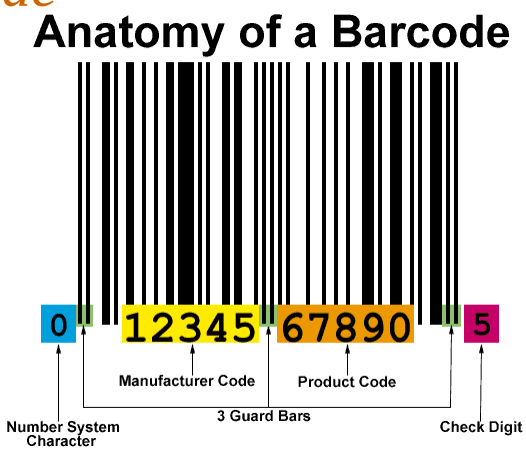
Lin's Recent Work on Process Mining

- Construction algorithm
  - Model Building
- Minimal Chart
- Flow Chart Construction, when there are noises
- Multiple Process Mining

Two Specific Example:

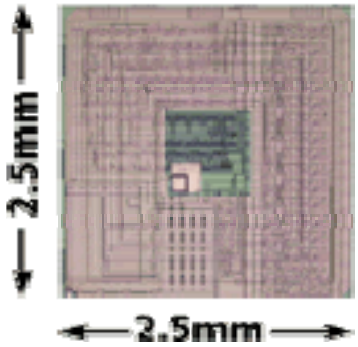
RFID &
Search Engine

Bar Code



RFID:

Radio Frequency Identification



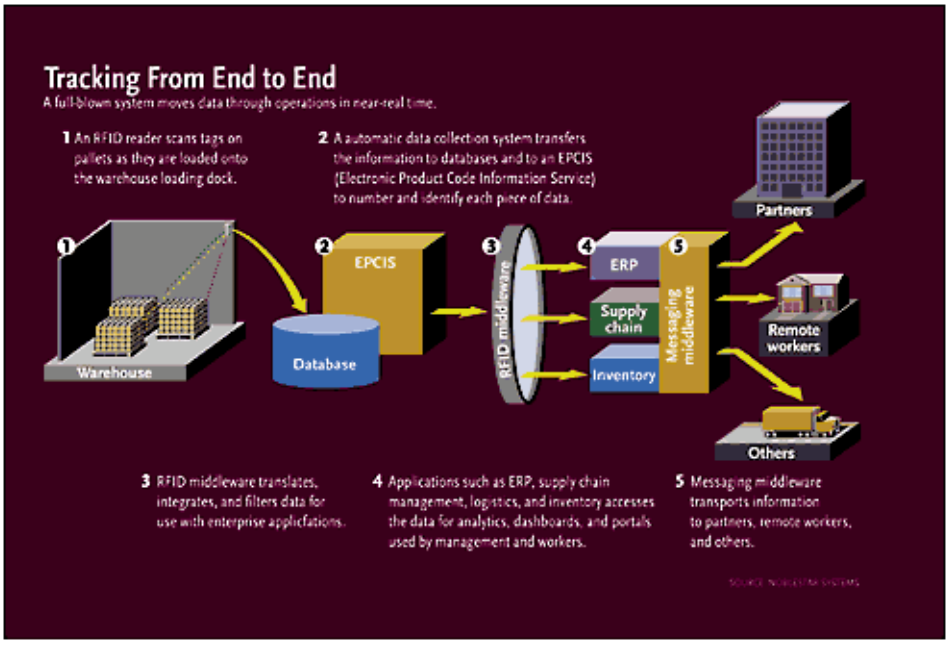
The components of a 96-bit electronic product code (in Hex)

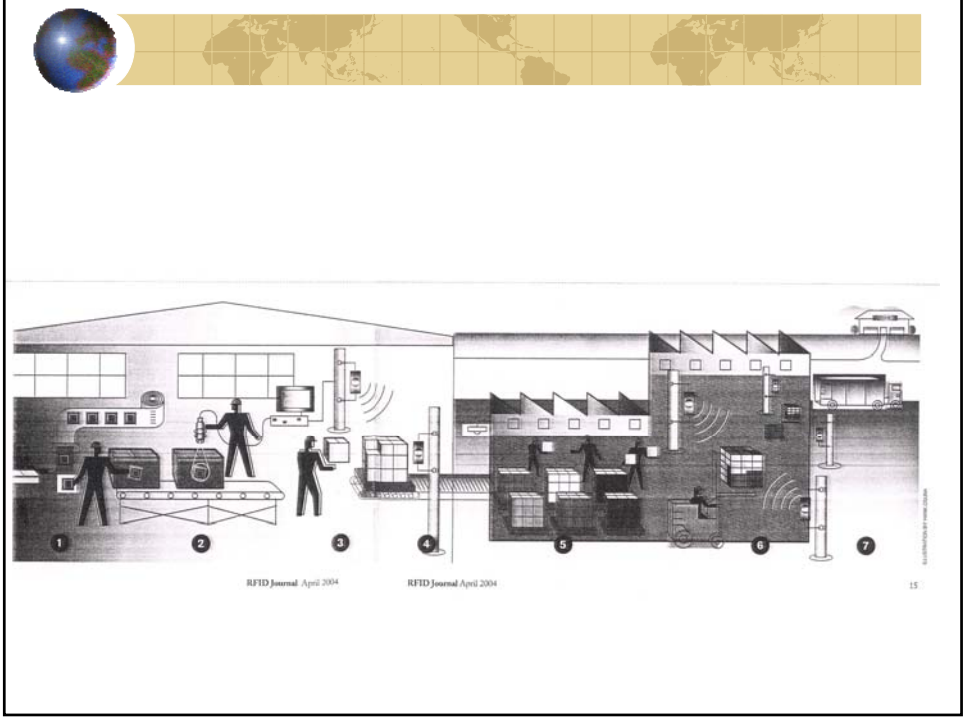
01	0000ABC	000123	000056789
Header 8 bits	EPC Manager (will be assigned to companies) 28 bits	Object Class - Product Code 24 bits	Serial Number 36 bits

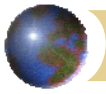
The RFID tag responds to the reader by broadcasting its EPC, which is a 96-bit code consisting of:

· 8 bits of header information
· 28 bits identifying the organization that assigned the code
· 24 bits identifying the type of product
· 36 bits representing serialization information for the product

Source: Avicon white paper, 2003


Tracking From End to End
A full-blown system moves data through operations in near-real time.
1 An RFID reader scans tags on pallets as they are loaded onto the warehouse loading dock.
2 A automatic data collection system transfers the information to databases and to an EPCIS (Electronic Product Code Information Service) to number and identify each piece of data.
Warehouse
EPCIS
Database
RFID middleware
ERP
Supply chain
Inventory
Messaging middleware
Partners
Remote workers
Others
3 RFID middleware translates, integrates, and filters data for use with enterprise applications.
4 Applications such as ERP, supply chain management, logistics, and inventory accesses the data for analytics, dashboards, and portals used by management and workers.
5 Messaging middleware transports information to partners, remote workers, and others.


RFID Journal April 2004
RFID Journal April 2004



RF Communication

- Electromagnetic waves modulated to carry data/signals
- Two different ways to generate ways
  - Inductive coupling
    - Close proximity electromagnetic wave
  - Propagating electromagnetic waves
- The fundamental RF communication theories apply—nothing new.
- New: the cost, size, signal processing capability.

- LF: 100-150 KHz
  門禁, 汽車防盜, 畜牧, 養殖, 工廠流水線
- HF: 13.56 MHz
  - 14443
    - 14443A—小額儲值卡, 公共交通卡
    - 14443B—證件, 護照, 銀聯卡
  - 15693—10cm 物品單體管理 (加密)
- UHF (1m)
  - 915MHz: EPC ISo1800-6C, 物流 倉儲
  - 2.45 GHz
  - 5.8 GHz: 高速移動物件管理

RFID Case Studies

- 電子標籤在血庫管理中的應用
- 电子标签在压力容器中的管理应用
- 机车防盗介绍
- 畜牧业管理
- 停车场方案
- 智能化大楼一卡通
- 電子標簽在酒類防僞中的應用
- 会议无障碍身份识别方案
- 图书馆
- 车辆管理方案

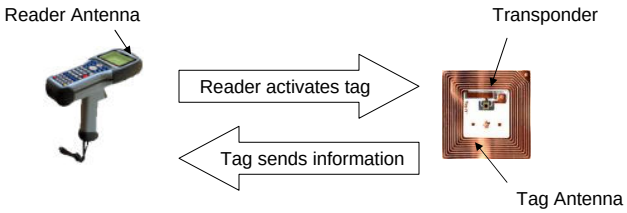
RFID Case Studies

- 電子標籤在血庫管理中的應用
- 电子标签在压力容器中的管理应用
- 机车防盗介绍
- 畜牧业管理
- 停车场方案
- 智能化大楼一卡通
- 電子標簽在酒類防僞中的應用
- 会议无障碍身份识别方案
- 图书馆
- 车辆管理方案

A
Data Explosion
is Coming!

Are You Ready?

RFID



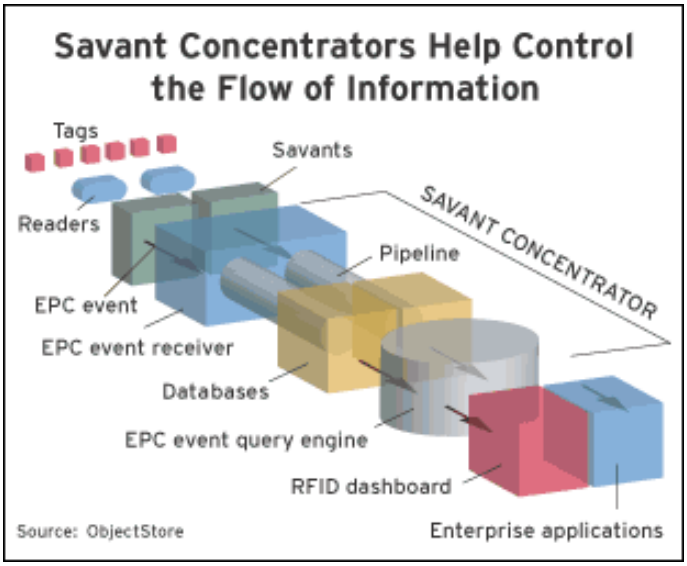
Reader ***Information System*** **Tag**

Features of Savant

Software at each level of supply chain will gather, store and act on information and interact with other Savants

- Smoothing.
- Reader Co-ordination.
- Forwarding.
- Storage.

pallet- + case- + item-level data = 100X or 1000X current volume



Control flow and provide filtering and aggregation of EPC event streams

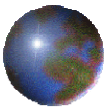
Architecture

ONS: object name server
Maps EPC→ URL

EPC Information service
Higher level service for apps

Gather data from readers:
Smoothing, coordination, forwarding,etc.

Source: Chawathe, et al, VLDB Conference proceedings, 2004



RFID vs. Barcode

Lin, Dennis K.J. and Wadhwa, Vijay

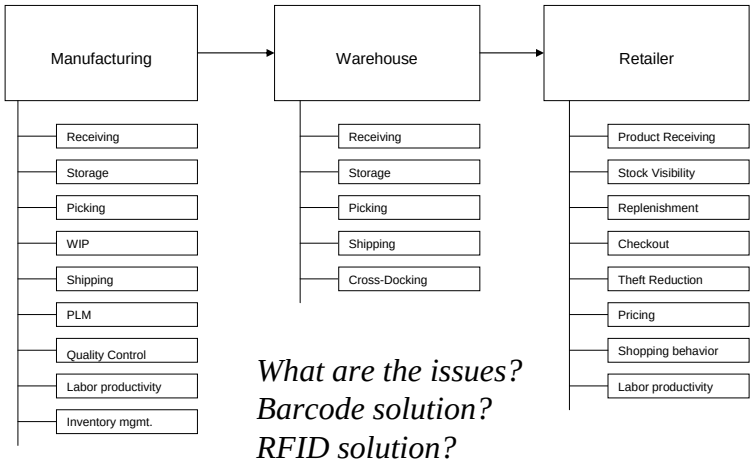
Bar-coding

- Bar coding is one of the most popular and cost effective forms of automated data collection systems.
- Fixed Barcode: Contains static information that is the same for all products of the same brand and type. An example is a UPC bar code on a 12 oz. can of coke.
- Variable barcode: A variable data bar code contains data that identifies a single product and changes for each separate product.

RFID

- RFID uses radio waves to automatically identify people or objects.
- Old technology but increased affordability, scalability, data processing capability.
- Capable of identifying each and every object uniquely in a supply chain.

Basic Structure



What are the issues?
Barcode solution?
RFID solution?
Comparisons!!!

S.I.T. Space

- State
- ID
- TimeStamp

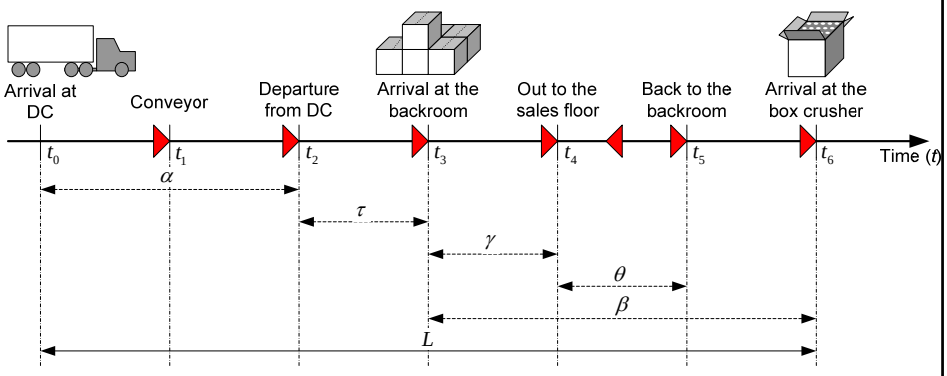
- Namely, {where, which, when}

- Particularly interested in the difference between *Execution* & *Planning*
  (Sense & Response)

Sample RFID Data

Location	EPC	Date/time	Reader
DC 123	0023800.341813.500000024	08-04-05 23:15	inbound
DC 123	0023800.341813.500000024	08-09-05 7:54	conveyor
DC 123	0023800.341813.500000024	08-09-05 8:23	outbound
ST 987	0023800.341813.500000024	08-09-05 20:31	inbound
ST 987	0023800.341813.500000024	08-09-05 20:54	sales floor
ST 987	0023800.341813.500000024	08-10-05 1:10	sales floor
ST 987	0023800.341813.500000024	08-10-05 1:12	backroom
ST 987	0023800.341813.500000024	08-11-05 15:01	sales floor
ST 987	0023800.341813.500000024	08-11-05 15:47	sales floor
ST 987	0023800.341813.500000024	08-11-05 15:49	box crusher

The timeline in a supply chain



An IBM RFID Example on Analysis:

What you're looking for in sensor data?

Lin, Shu and Wadhwa

Preliminary Analysis on Wal-Mart RFID data

Raw Data

The data contains the following fields in the order as it appears:

- EPC Number:
- Item Number:
- Item Description:
- UPC (Universal Product Code)
- GTIN:
- EPC Time Stamp
- Store Number:
- Location:
- Location Description:
- EPC status Description:

Flow-Analysis: Process Mining

- For each EPC there are two ways a location is classified as unique:
  - if the previous location scanned is not same as the current location (sorted by datetime).
  - If the time difference between two consecutive location is greater than a user specified value.
  - For each EPC we vary the user specified time difference to see the process flow obtained.

Flow-Time Analysis

- Flow time of the part through the system is the difference between the time the part enters the system and the time the part leaves the system.
- Flow time can be related to demand; hence it could be used in managing the inventory at various locations.

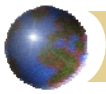
Future Directions & Issues

- Need to understand how the data is gathered and the process behind it.
- Efficiency at various stores/DC can be computed using the time lags between the RFID scans.
- Infer why the actual process differs from the benchmarked process.

Low-Storage Single-Pass Density Estimation

- Univariate
  - Certain (say, 20) representative quantiles
  - (Cubic Smoothing) Spline fitting to these quantiles
  - Statistical inference
- Multivariate
  - Convex Hulls Peeling
  - Certain (say, 20) representative convex hulls
  - Thin Plate Splines fitting to these quantiles
  - Statistical inference

Jim McDermott

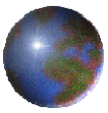


IIE Transactions (2007) **39**, 581–591
Copyright © "IIE"
ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170600899599

Quantile contours and multivariate density estimation for massive datasets via sequential convex hull peeling

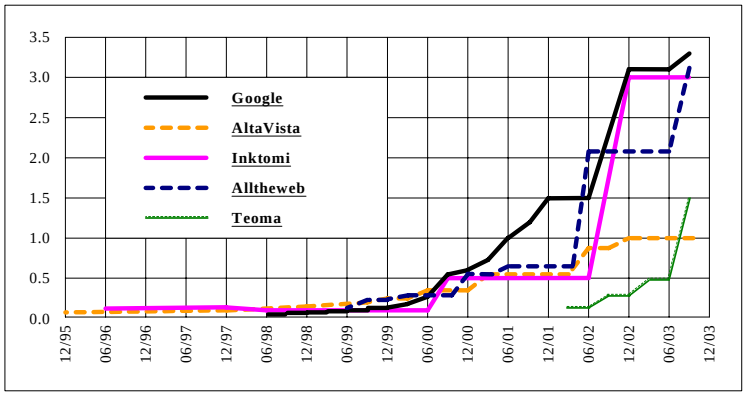
JAMES P. McDERMOTT[1] and DENNIS K. J. LIN[2,*]
[1] *Department of Statistics and* [2] *Department of Supply Chain and Information Systems, The Pennsylvania State University, University Park, PA 16802, USA*
E-mail: DKL5@psu.edu

Received March 2004 and accepted April 2006



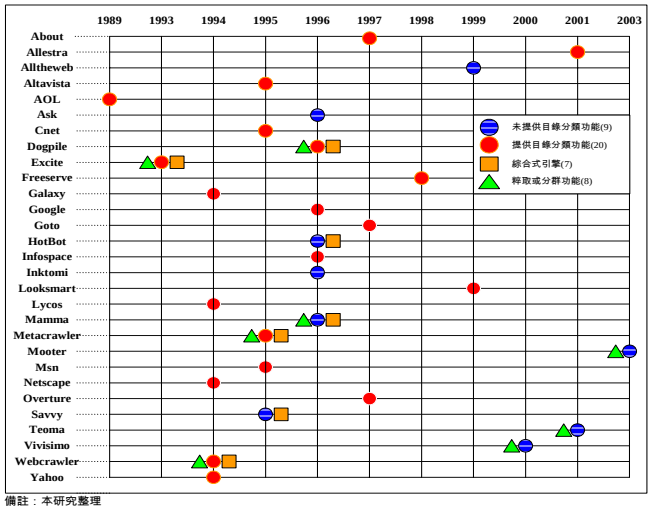
Search Engine & Citation Index

New Era for Search Engine



Unit: One Billion

History of Search Engine



備註：本研究整理

New Aspect of Damping Factor in Google Page Rank

Fu, H.H.
Lin, Dennis K.J.
Tsai, H.T.

Google's Page Rank formula

- The PageRank of a page A is given as follows:

$$PR_1(A) = (1-d) + d \times \left(\frac{PR(T_1)}{C(T_1)} + \frac{PR(T_2)}{C(T_2)} + \ldots + \frac{PR(T_n)}{C(T_n)} \right)$$

- PR(A) is the PageRank of page A;
- PR(Ti) is the PageRank of pages Ti which link to page A;
- C(Ti) is the number of outbound links on page Ti;
- d is a damping factor which can be set between 0 and 1; usually set to **0.85**
- n is the total number of all pages which link to page A.

Markov Chains

- Matix A

$$a_{ij} = \frac{(1-d)}{N} + d\frac{g_{ij}}{c_j} \qquad d=0.85$$

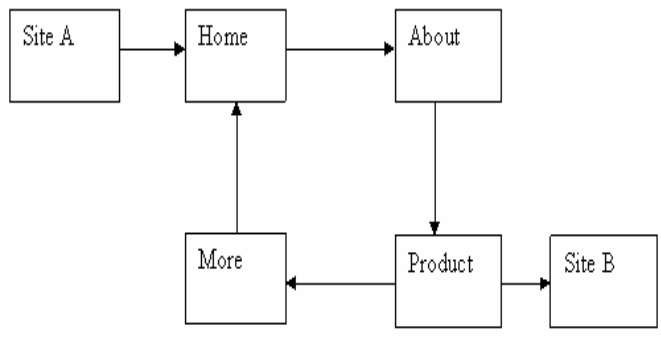
- Matix A max eignvalue =1

$$Ax=x \qquad \sum_i x_i = 1$$

- Matix A eignvector = PageRank(k)

$$x_k = \sum_{j=1}^{N} a_{kj} x_j = \frac{(1-d)}{N} + d\sum_{g_{kj}=1} \frac{x_j}{c_j}$$

Example 5



Example 5

$$H = (1-d) + d(\frac{M}{1} + \frac{SA}{1})$$

$$A = (1-d) + d(\frac{H}{1})$$

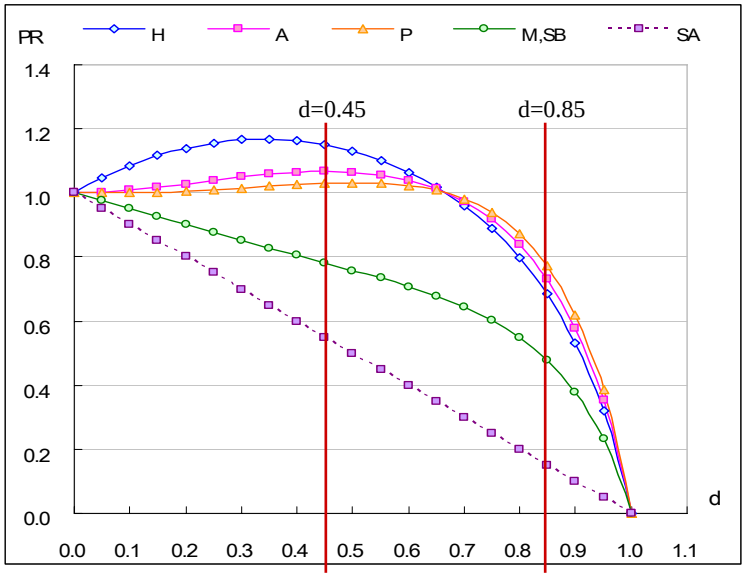
$$P = (1-d) + d(\frac{A}{1})$$

$$M = (1-d) + d(\frac{P}{2})$$

$$SA = (1-d)$$

$$SB = (1-d) + d(\frac{P}{2})$$

Example 5

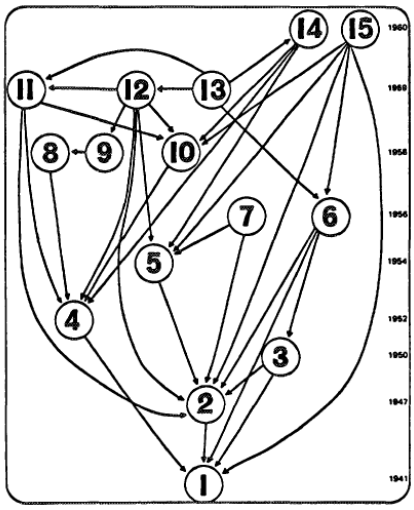


How to Increase your PageRank?

How to Increase Your Paper Citation?

- Individual article
- Individual Author
- Journal
- How is this (Impact Factor) related to PageRank?

Paper Citation: *Garfiled (1972)*



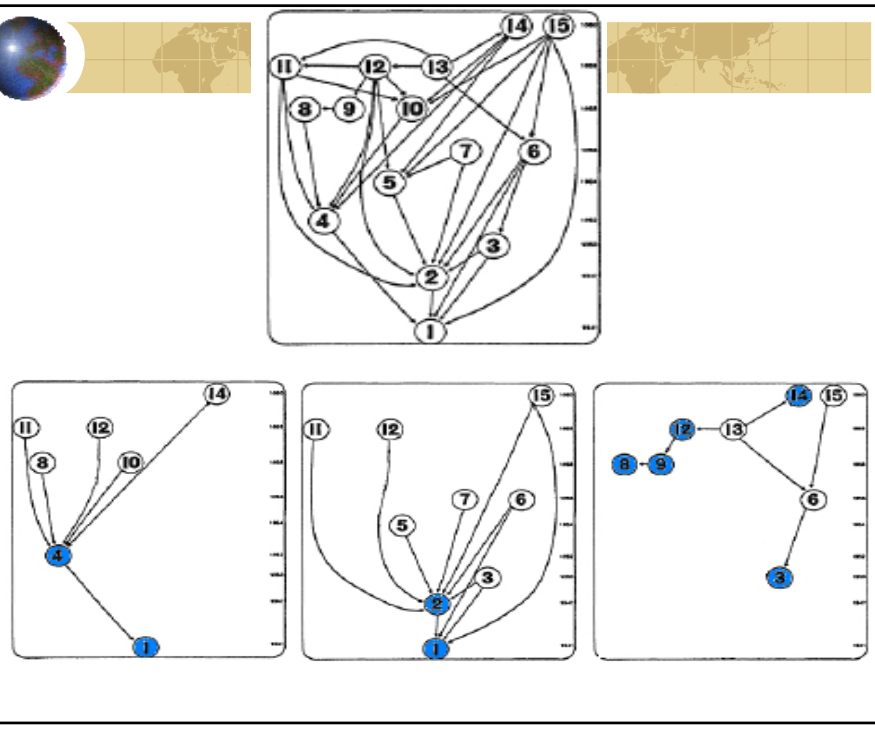
Number of Citations:

Paper#1: 5
Paper#2: 7
Paper#4: 5

Garfiled (*Science*, 1972)

Table 1. Comparison between SCI and CCI for the citation network in Figure 1(*a*)

Paper	SCI	CCI (β = 0.3)	SCI Ranking	CCI Ranking	Ranking Change
1	5	10.16	2	1	1
2	7	8.88	1	2	-1
3	1	1.20	8	9	-1
4	5	7.06	2	3	-1
5	4	4.15	4	5	-1
6	2	2.00	6	7	-1
7	0	0.00	13	13	0
8	1	1.32	8	8	0
9	1	1.05	8	10	-2
10	4	4.36	4	4	0
11	2	2.05	6	6	0
12	1	1.00	8	11	-3
13	0	0.00	13	13	0
14	1	1.00	8	11	-3
15	0	0.00	13	13	0



Some Observations

- Paper 1 has direct influence on Paper 4 as well as indirect influence on Paper 4's citing papers (as shown in Figure 1(*b*)). Such indirect influence should be added to Paper 1's overall influence.
- Paper 2 cites Paper 1 and almost half of Paper 2's citing papers also cite Paper 1 (as shown in Figure 1(*c*)). This implies that Paper 1 has both direct and indirect influence on those citing papers of Paper 2.
- SCI ranks Papers 1 and 4 the same and ranks Paper 2 higher than Paper 1. But based on (1) and (2), it is likely that Paper 1 is more influential than Papers 2 and 4. This observation is confirmed by the CCI rankings in Table 1.
- SCI has the same ranking for Papers 3, 8, 9, 12, and 14 that all have one direct citation (as shown in Figure 1(*d*)), but CCI ranks some of them differently. These differences can be explained by the fact that those papers are cited by papers that have different influences. For example, the CCI ranking of Paper 3 is higher than that of Paper 12, because Paper 3's citing paper (i.e., Paper 6 with CCI = 2.00) is more influential than Paper 12's citing paper (i.e., Paper 13 with CCI = 0).
- This example shows that CCI has higher resolution than SCI and is capable of representing importance of different citations. This distinctive feature of CCI is useful for identifying the different influences of papers that have the same or similar number of citations. This feature of CCI is similar to that of PageRank7 applied to search engines. PageRank considers that in a network, each incoming link is different and an incoming link has more value if it comes from a more important node.

Table 2. Top 10 most influential papers published in *Management Science* between 1954 and 2003*

ID	Title	SCI	CCI	SCI Ranking	CCI Ranking	Ranking Change
1	A New Product Growth for Model Consumer Durables	575	1062.6	7	7	0
2	A Suggested Computation for Maximal Multi-Commodity Network Flows	44	111.6	320	121	199
3	Dynamic Version of the Economic Lot Size Model	269	337.1	24	24	0
4	Games with Incomplete Information Played by 'Bayesian' Players, I: The Basic Model	307	618.6	22	13	9
5	Information Distortion in a Supply Chain: The Bullwhip Effect	443	694.7	11	11	0
6	Jobshop-Like Queueing Systems	262	494.9	26	17	9
7	Linear Programming under Uncertainty	163	277.6	56	33	23
8	Models and Managers - Concept of a Decision Calculus	138	225.0	71	49	22
9	Optimal Policies for a Multi-echelon Inventory Problem	274	528.2	23	16	7
10	The LaGrangian-Relaxation Method for Solving Integer Programming-Problems	378	639.2	15	12	3
	Average			57.5	30.3	

Table 3. Top 10 papers ranked by SCI and top 10 papers ranked by CCI in a citation network in the field of biology**

SCI Top 10					CCI Top 10				
Paper ID & SCI Ranking	SCI	CCI	CCI Ranking	Ranking Change	Paper ID & CCI Ranking	CCI	SCI	SCI Ranking	Ranking Change
SCI-1(CCI-3)	988	8867	3	-2	CCI-1	14307	199	426	425
SCI-2	987	1030	39	-37	CCI-2(SCI-4)	9109	978	4	2
SCI-3(CCI-5)	981	6839	5	-2	CCI-3(SCI-1)	8867	988	1	-2
SCI-4(CCI-2)	978	9109	2	2	CCI-4	7892	560	36	32
SCI-5	977	1240	28	-23	CCI-5(SCI-3)	6839	981	3	-2
SCI-6(*a*)	976	2181	13	-7	CCI-6	6277	519	47	41
SCI-6(*b*)	976	1164	31	-25	CCI-7	5431	108	1240	1233
SCI-8(CCI-8)	953	4639	8	0	CCI-8(SCI-8)	4639	953	8	0
SCI-9	937	1699	17	-8	CCI-9	4097	804	16	7
SCI-10	891	1168	30	-20	CCI-10	2914	536	42	32

Detailed information of the papers listed above

Paper ID	Title	Author
SCI-1(CCI-3)	Regulation of the mevalonate pathway	Goldstein, J.L. & Brown, M.S. (1990)
SCI-2	Insulin-like growth factors and their binding proteins: biological actions	Jones, J.I. & Clemmons, D.R. (1995)
SCI-3(CCI-5)	Apolipoprotein E: cholesterol transport protein with expanding role in cell biology	Mahley, R.W. (1988)
SCI-4(CCI-2)	The NF-kappaB AND IkappaB PROTEINS: New Discoveries and Insights	Baldwin, A.S. (1996)
SCI-5	Inflammation and Atherosclerosis	Libby, P. & Ridker, P.M. (2002)
SCI-6(*a*)	The Effect of Pravastatin on Coronary Events after Myocardial Infarction in Patients with Average Cholesterol Levels	Sacks, F. M. et al. (1996)
SCI-6(*b*)	C-Reactive Protein and Other Markers of Inflammation in the Prediction of Cardiovascular Disease in Women	Ridker, P.M. et al. (2000)
SCI-8(CCI-8)	The pathogenesis of atherosclerosis: a perspective for the 1990s	Ross, R. (1993)
SCI-9	Nuclear Factor-B-A Pivotal Transcription Factor in Chronic Inflammatory Diseases	Barnes, P.J. & Karin, M. (1997)
SCI-10	Functions of Lipid Rafts in Biological Membranes	Brown, D.A. & London, E. (1998)
CCI-1	A receptor-mediated pathway for cholesterol homeostasis	Brown, M.S. & Goldstein, J.L. (1986)
CCI-4	Atherosclerosis: Basic Mechanisms Oxidation, Inflammation, and Genetics	Berliner, J.A. et al. (1995)
CCI-6	Cloning, structure, and expression of the mitochondrial cytochrome P-450 sterol 26-hydroxylase	Andersson, S. et al. (1989)
CCI-7	Studies on the mechanism of hormone action	Sutherland, E.W. (1972)
CCI-9	Coronary Plaque Disruption	Falk, E. et al. (1995)
CCI-10	Structures and Functions of Multiligand Lipoprotein Receptors: Macrophage Scavenger Receptors and LDL Receptor-Related Protein (LRP)	Krieger, M. & Herz, J. (1994)

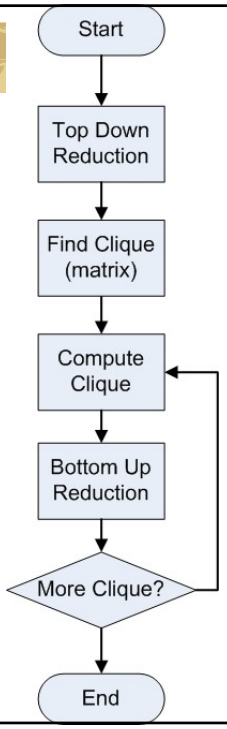
Mathematical Formulation

$$x_i = |J_i| + \beta \sum_{j \in J_i} \frac{x_j}{r_j} = \sum_{j \in J_i} (1 + \beta \frac{x_j}{r_j})$$

$$\mathbf{x} = \mathbf{He} + \beta \mathbf{Gx} = \begin{pmatrix} h_{11} & \dots & \dots & h_{1n} \\ h_{21} & \dots & & h_{2n} \\ \dots & & \dots & \dots \\ h_{n1} & \dots & \dots & h_{nn} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix} + \beta \begin{pmatrix} g_{11} & \dots & \dots & g_{1n} \\ g_{21} & \dots & & g_{2n} \\ \dots & & \dots & \dots \\ g_{n1} & \dots & \dots & g_{nn} \end{pmatrix} \mathbf{x}$$

Algorithm

- Bottom-up reduction. It reduces the size of the M-Matrix by computing the CCI values of the papers whose citing papers' CCI values have been calculated. The reduction is performed as follows:
  - Construct a (***n*** **+ 2**) x (***n*** **+ 1**) matrix , where **H** is the citation network matrix, **R** is a vector of the row sum of **H**, **V** is the initial values (which are 0's) of paper influences, and ***C*** is a row vector of the column sum of **H**.
  - Evaluate **R**. If *rj* = 0, then check **C**, and if *cj* > 0:
    - For all *hij* = 1, *vi* = *vi* + 1 + *βvj* /*cj*.
    - Set *hij* = 0, and update *rj* = *rj* − 1.
    - Move row *i* to the bottom of **H** and move column *j* to the right of **H**.
    - Let **H** be a new matrix without row *i* and column *j*.
  - If **R** > 0, stop and no further reduction is needed; otherwise, repeat steps (*a*) and (*b*).
- Top-down reduction. It reduces the size of the M-Matrix by moving papers without citing papers into a temporary matrix. This process is similar to Bottom-up reduction except for using the vector of row sum instead of column sum.
- Compute Cliques. After Bottom-up and Top-down reductions, the remaining papers can be divided into smaller sets (called Cliques) of papers involving loop citations. Each set can be solved as linear equations that are much smaller than the original problem. After computing each linear equation, perform Bottom-up reduction. The algorithm stops when no more Cliques exist.



STILL QUESTION?

Send $500 to

- Dennis Lin

  University Distinguished Professor

  483 Business Building

  Department of Supply Chain & Information Systems

  Penn State University

- +1 814 865-0377 (phone)
- +1 814 863-7076 (fax)
- DKL5@psu.edu

(Customer Satisfaction or your money back!)