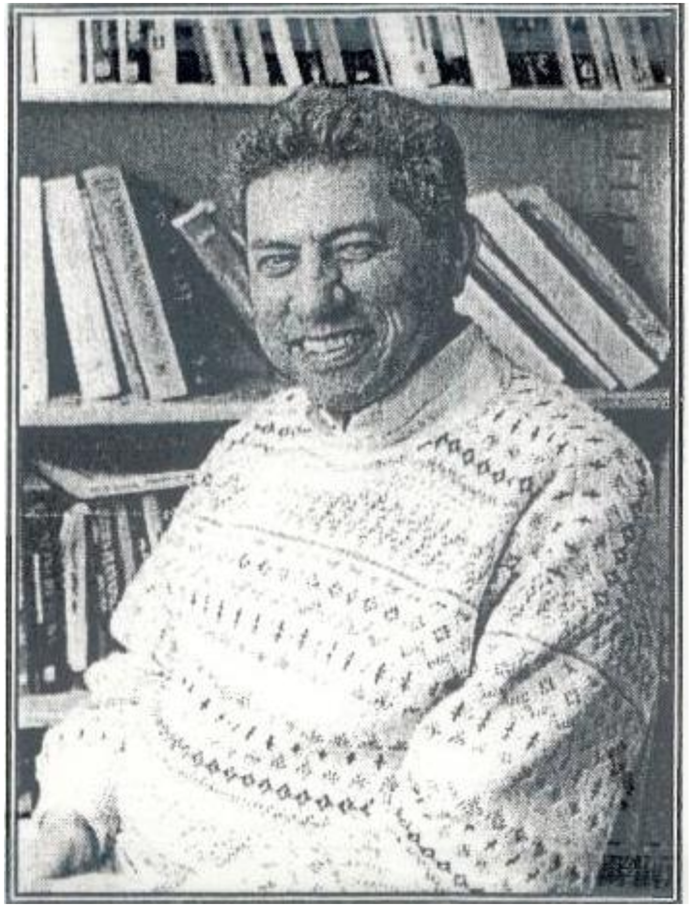
OM Research: From Problem Driven to Data Driven Research

David Simchi-Levi
E-mail: dslevi@mit.edu

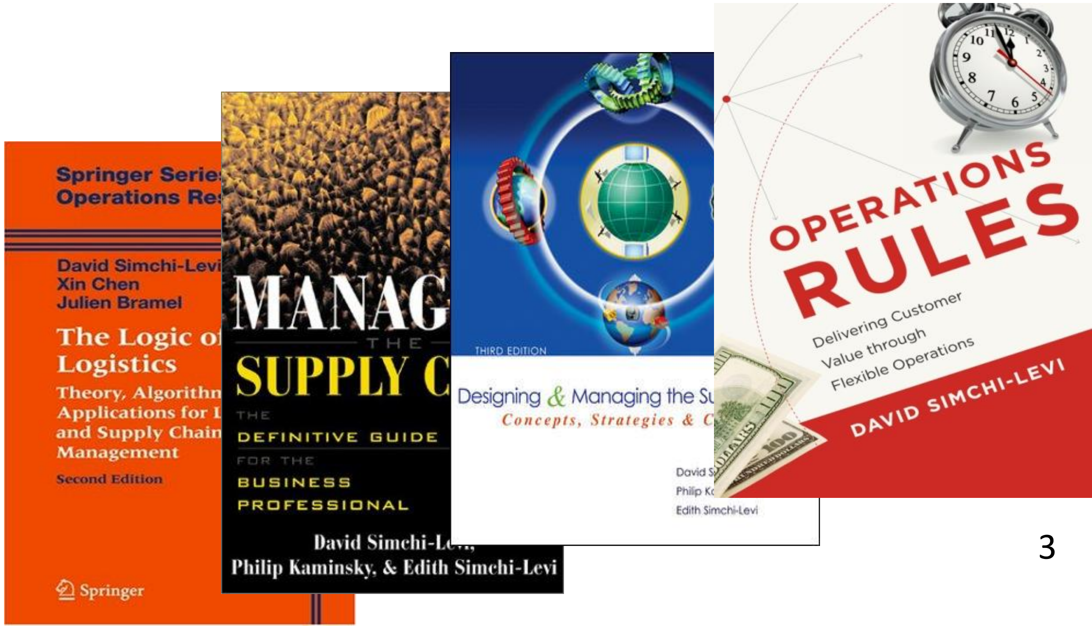
OPERATIONS
RESEARCH
CENTER

Meir Rosenblatt Memorial Lecture



A Few Stories …

- **From Theory to Practice**
  - School Bus Routing in NYC
- **From Practice to Theory**
  - Flexibility at PepsiCo
- **Merging Theory and Practice**
  - Online Retailing
- **Conclusions**
  - Data Driven Models

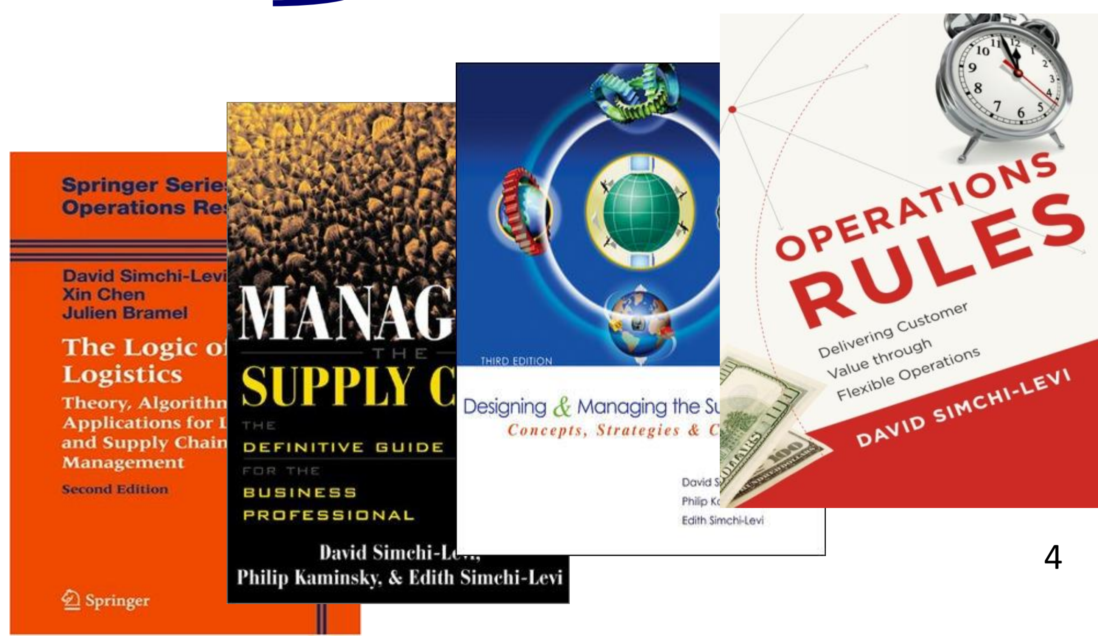


A Few Stories …

- **From Theory to Practice**
  - School Bus Routing in NYC
- **From Practice to Theory**
  - Flexibility at PepsiCo
- **Merging Theory and Practice**
  - Online Retailing
- **Conclusions**
  - Data Driven Models

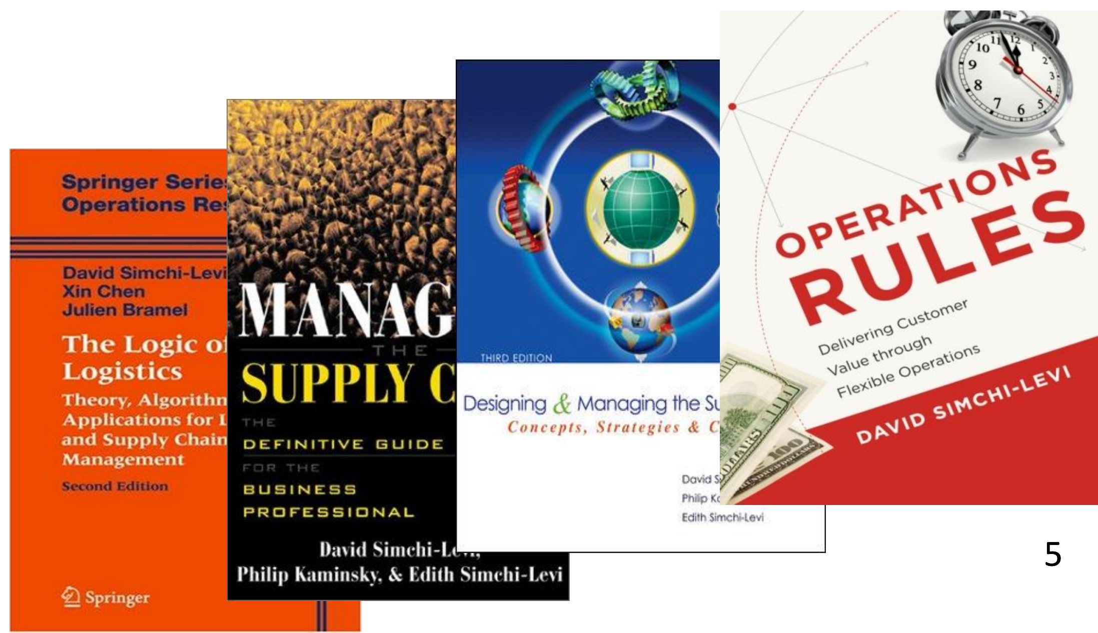
Problem Driven Research

Data Driven Research



A Few Stories …

- **From Theory to Practice**
  - School Bus Routing in NYC
- **From Practice to Theory**
  - Flexibility at PepsiCo
- **Merging Theory and Practice**
  - Online Retailing
- **Conclusions**
  - Data Driven Models



Algorithms and Probability

- **Karp, R. M. (1977), Probabilistic Analysis of Partitioning Algorithms for the Traveling Salesman Problem. *Math. Oper. Res.***

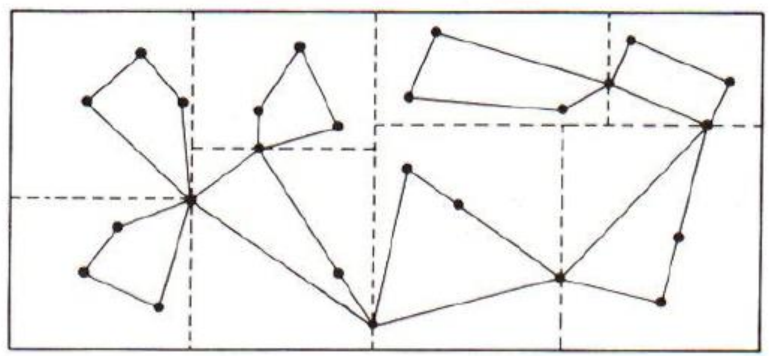


FIGURE 3. Walk Created by Algorithm 1.

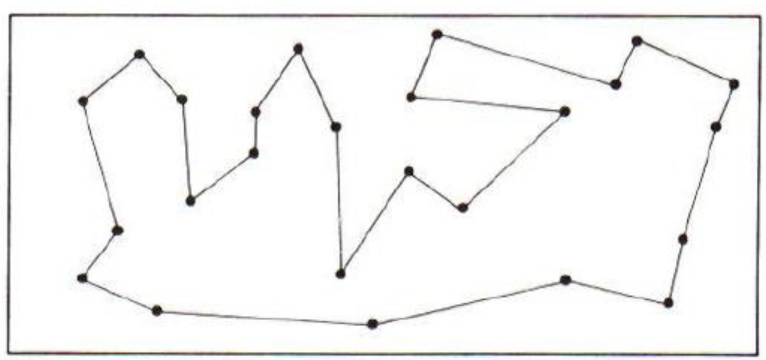


FIGURE 4. Tour Obtained Using the Loop and Pass Operations.

Asymptotically optimal heuristics

Algorithms and Probability

- **Haimovich, M. and A. H. G. Rinnooy Kan (1985), Bounds and Heuristics for Capacitated Routing Problems. *Math. Oper. Res.***
  - **Key Assumption: Equal Customer Demand**

- **The Challenge: Identify asymptotically optimal algorithms for general vehicle routing problems**
  - **Unequal customer demand**
  - **Time window constraints**

General Vehicle Routing Problems

- **Bramel and Simchi-Levi (1996), Probabilistic Analysis and Practical Algorithms for the Vehicle Routing Problem with Time Windows. *Operations Research***
- **Cost is dominated by redial distance and optimal packing of customers.**

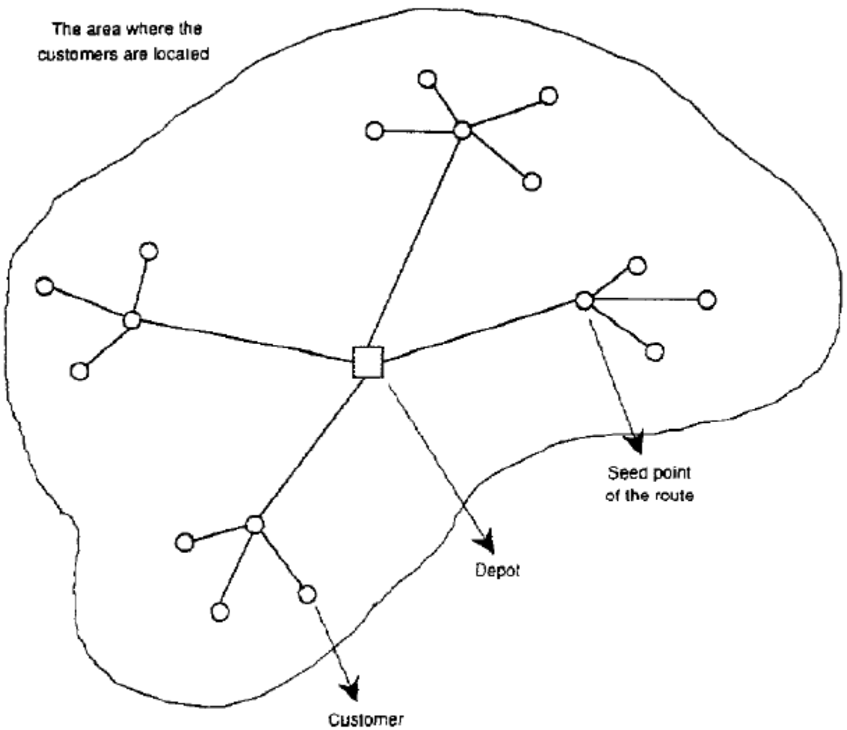


Figure 1. Tour used to construct heuristic.

- **Leads to modeling routing problems as "Capacitated Location Problems."**
- **Asymptotically optimal heuristic**

Weaknesses and Strengths

- **Two very strong assumptions**
  - **Large size problems**
  - **Independent customer behavior**

- **Provides insight into the structure of efficient algorithms:**
  - **Cost and Service**
  - **Computational time**

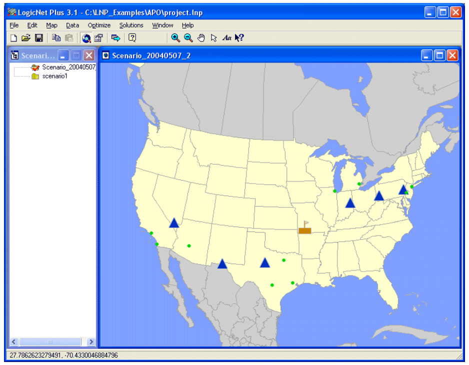
NYC--School Bus Routing System

- 1500 buses
- $100K per bus and driver per year
- The “Manhattan Project”
- 30-40% savings

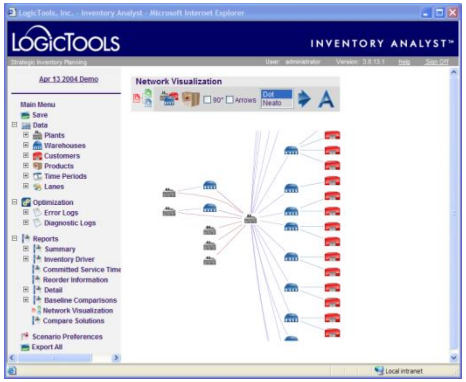
- Computerized System for School Bus Routing
- Combine Large Database, Visualization and Analytics
- First Place Prize in Windows World Open Competition, 1994

LogicTools, Inc. – Corporate Overview

- Industry Leading Company
  - Founded in 1996 by David Simchi-Levi, professor at MIT and Edith Simchi-Levi
  - The market leader in supply chain planning systems that integrate state-of-the-art optimization technology and easy-to-use interfaces
  - HQ in Chicago, IL
    - Offices: Boston MA, Eugene OR and Munich Germany
  - Over 350 companies (70+ Fortune 500) in many different industries use and benefit from our solutions



- LogicTools is an SAP Software Partner since 2004
  - SAP recognizes LogicTools' thought leadership in supply chain planning
  - Peak Performance Initiative with Microsoft

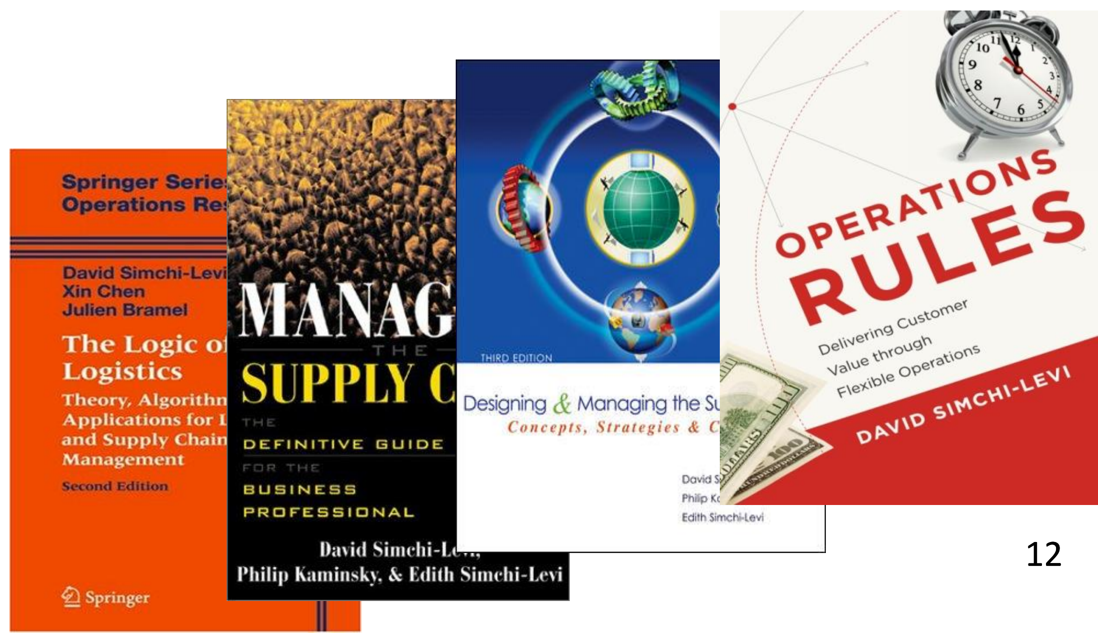


- Acquired in April 2007
  - Now part of IBM Business Analytics Solutions
  - LogicTools technology is "used by over 50% of the world's largest supply chains," according to IBM.

Confidential
LogicTools © 2006, All Rights Reserved.

A Few Stories …

- **From Theory to Practice**
  - School Bus Routing in NYC
- **From Practice to Theory**
  - Flexibility at PepsiCo
- **Merging Theory and Practice**
  - Online Retailing
- **Conclusions**
  - Data Driven Models



The Pepsi Bottling Company (A Division of PepsiCo)

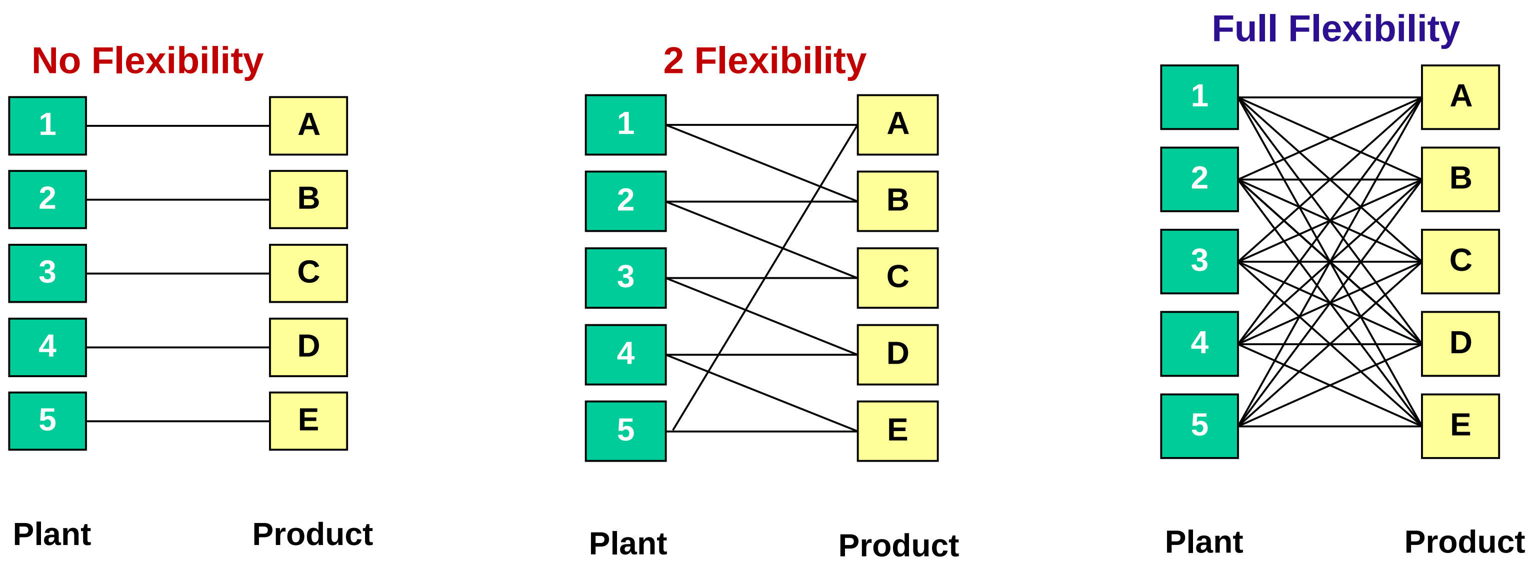
Make	Sell	Deliver	Service
Operates 57 Plants in the U.S. and 103 Plants Worldwide	7 Business units in the U.S. each responsible for local demand	240,000 Miles are Logged Every Day to Meet the Needs of Our Customers	Strong Customer Service Culture

The Challenge (Beginning of 2006):

- Shifting consumer preference
  - From carbonated to non-carbonated drinks
  - From cans to bottles
- Produced these products in limited plants
- Service problems during periods of peak demand

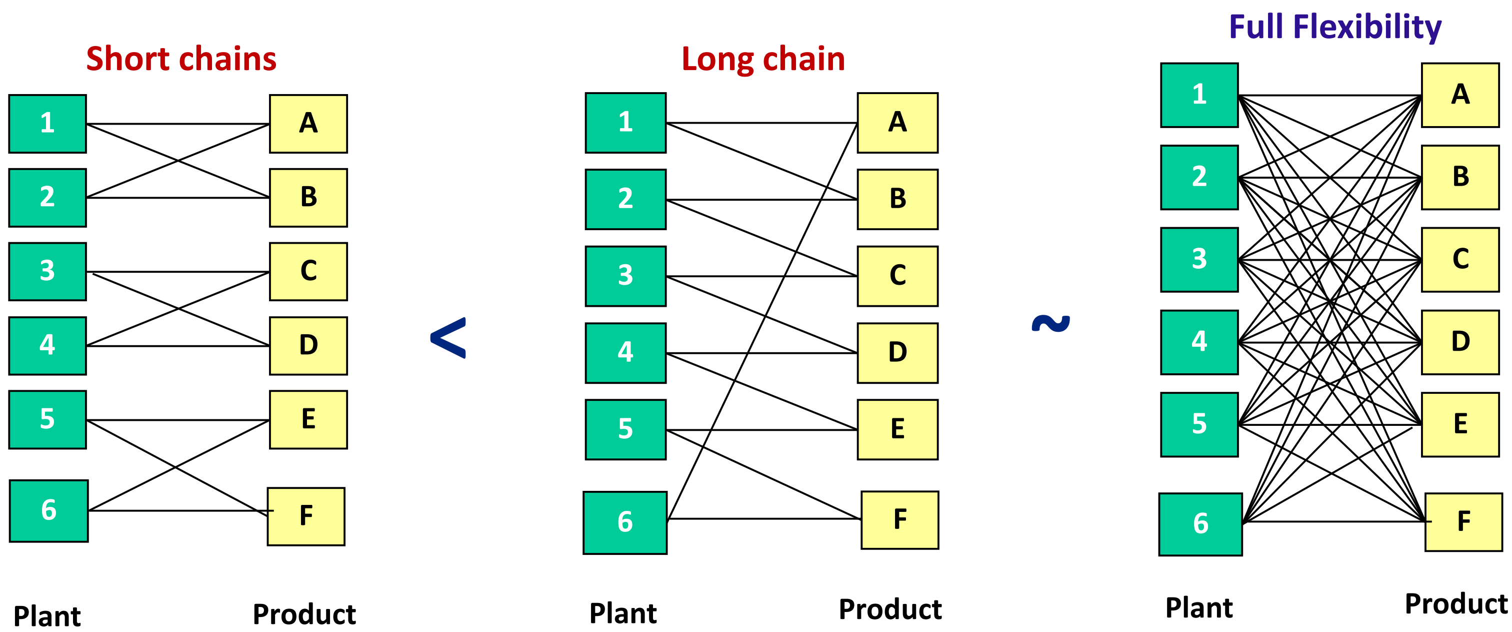
Process Flexibility

- Balance transportation and manufacturing costs
- Match supply with demand
- Better utilize resources



Chaining Strategy (Jordan & Graves 1995)

- **Focus: maximize the amount of demand satisfied**
- **Simulation study**



Applications to different settings:

[Sheikhzadeh et al. 1998], [Graves & Tomlin 2003], [Hopp et al. 2004], [Gurumurthi & Benjaafar 2004], [Bish et al. 2005], [Iravani et al. 2005], [Wallace & Whitt 2005], [Chou et al. 2010a]

PepsiCo's Press Release, 2008—The Impact

- Creation of regular meetings bringing together Supply chain, Transport, Finance, Sales and Manufacturing functions to discuss sourcing and pre-build strategies
- Reduction in raw material and supplies inventory from $201 to $195 million
- A 2 percentage point decline in in growth of transport miles even as revenue grew
- An additional 12.3 million cases available to be sold due to reduction in warehouse out-of-stock levels

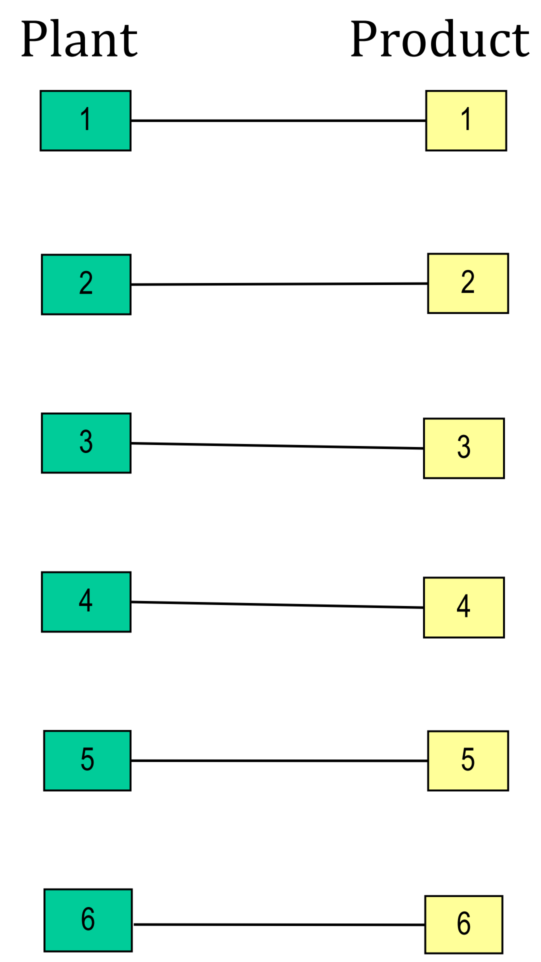
To put the last result in perspective, the reduction in warehouse out-of-stock levels effectively added one and a half production lines worth of capacity to the firm's supply chain without any capital expenditure.

Motivating Example

Fixed plant capacity: 10000

IID Normal product demand:
mean = 10000
stdev = 3300

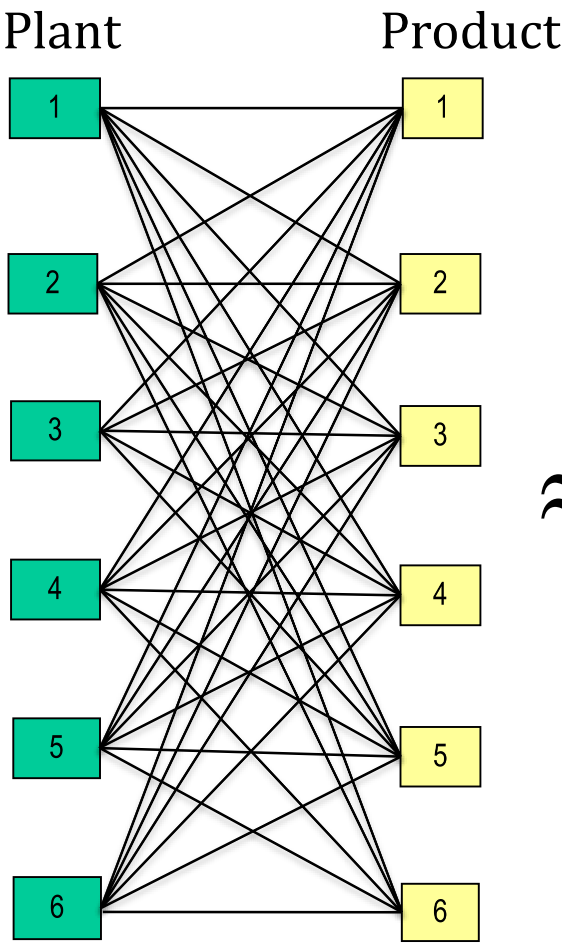
No Flexibility



Expected Sales≈ 52034

Adding Flexibilities

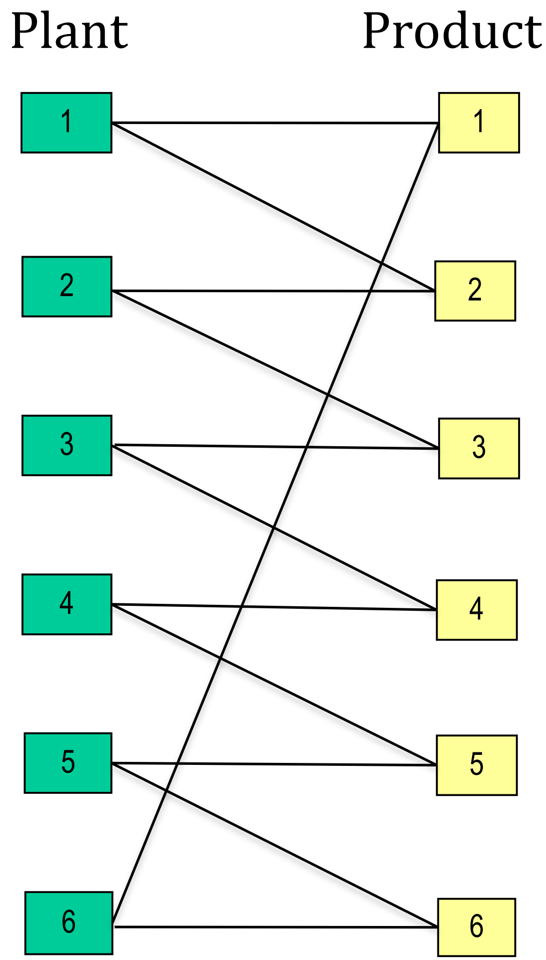
Full Flexibility



Expected Sales ≈ 56757
Improvement ≈ 9.08%

≈

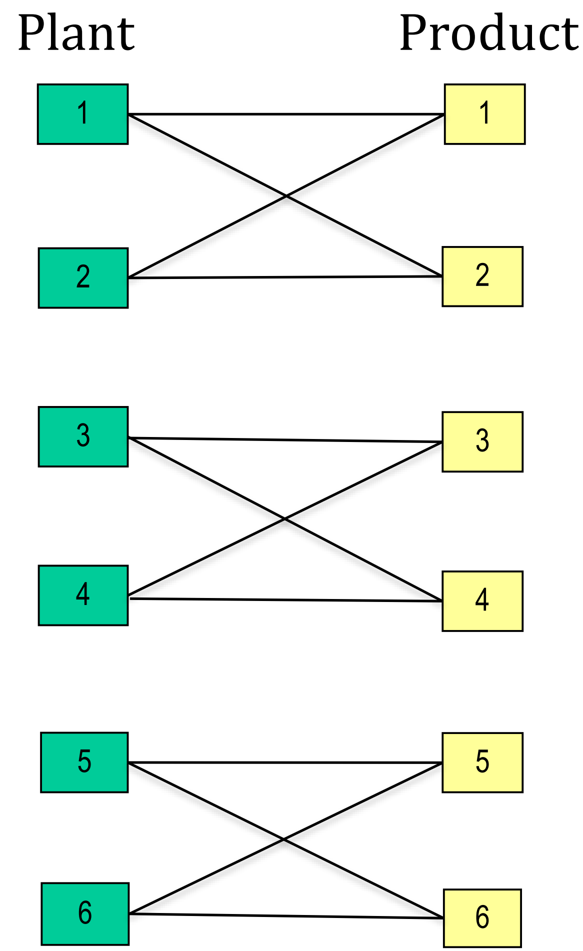
Long Chain



Expected Sales ≈ 56735
Improvement ≈ 9.03%

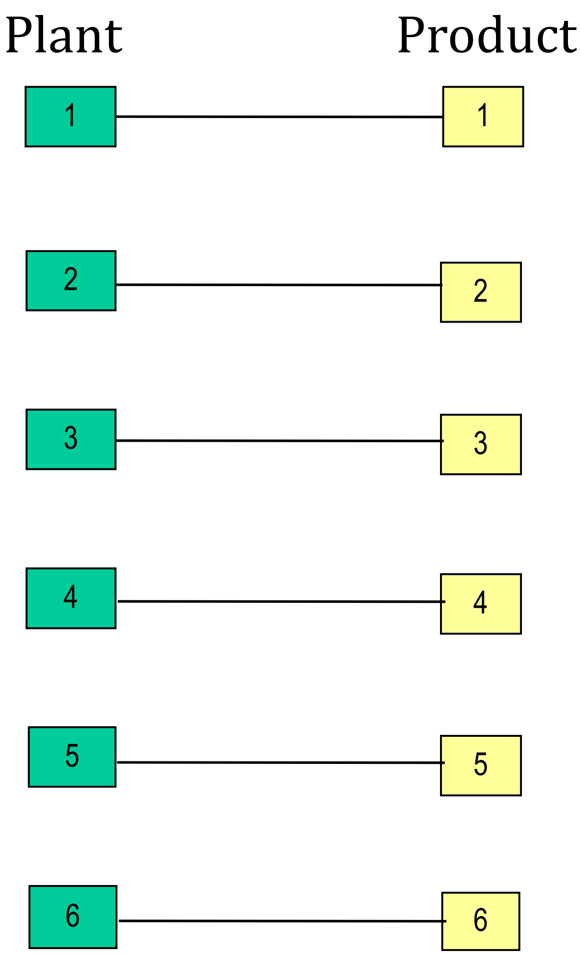
≥

Short Chains



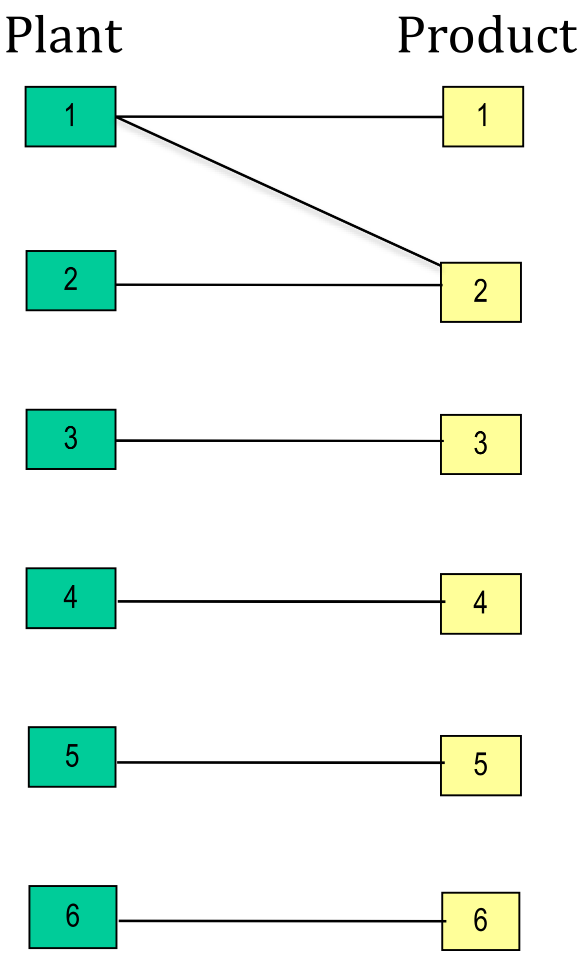
Expected Sales ≈ 54374
Improvement ≈ 4.50%

Increasing Incremental Benefit



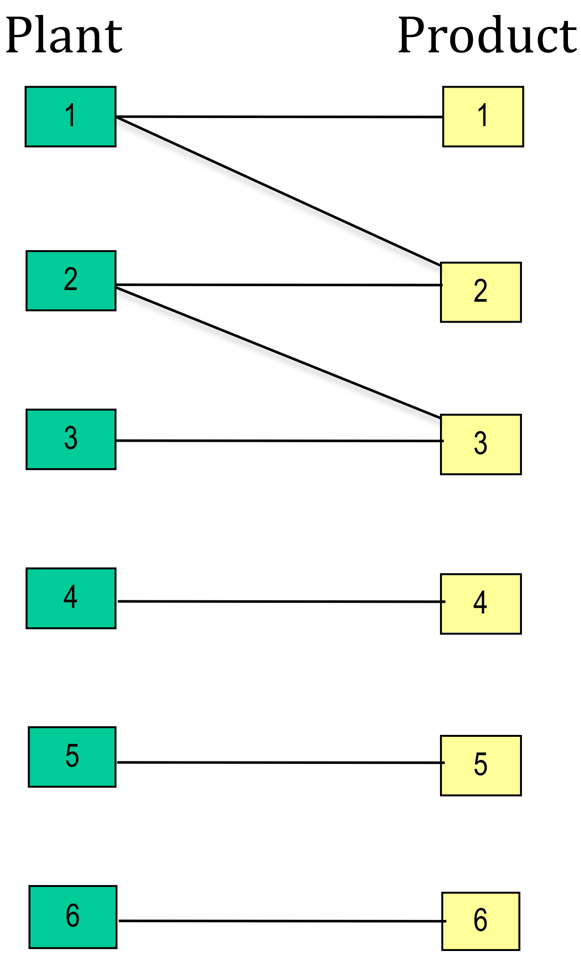
Design	Expected Sales	Marginal Benefit
No Flexibility	52034	
Add Arc (1,2)		
Add Arc (2,3)		
Add Arc (3,4)		
Add Arc (4,5)		
Add Arc (5,6)		
Add Arc (6,1)		

Increasing Incremental Benefit



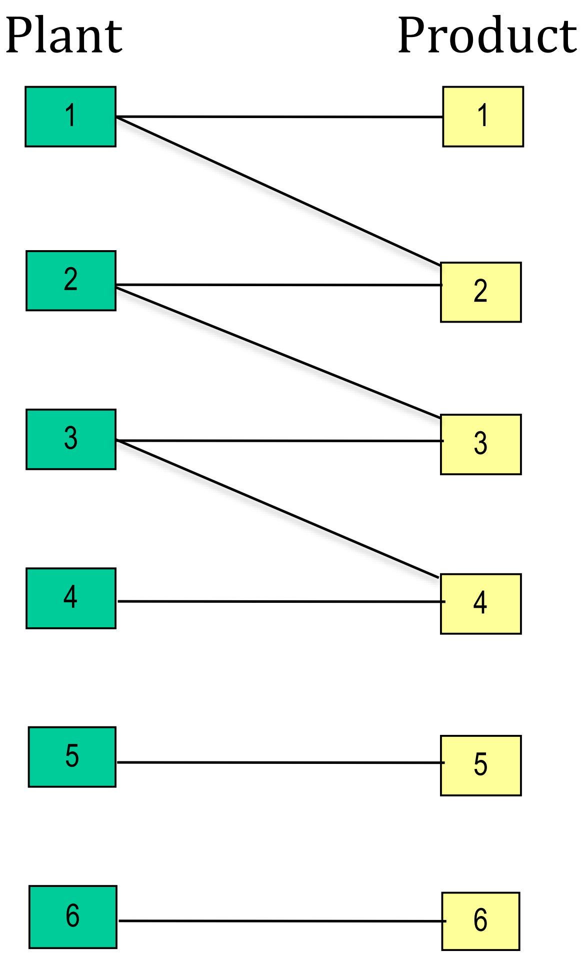
Design	Expected Sales	Marginal Benefit
No Flexibility	52034	
Add Arc (1,2)	52424	390
Add Arc (2,3)		
Add Arc (3,4)		
Add Arc (4,5)		
Add Arc (5,6)		
Add Arc (6,1)		

Increasing Incremental Benefit



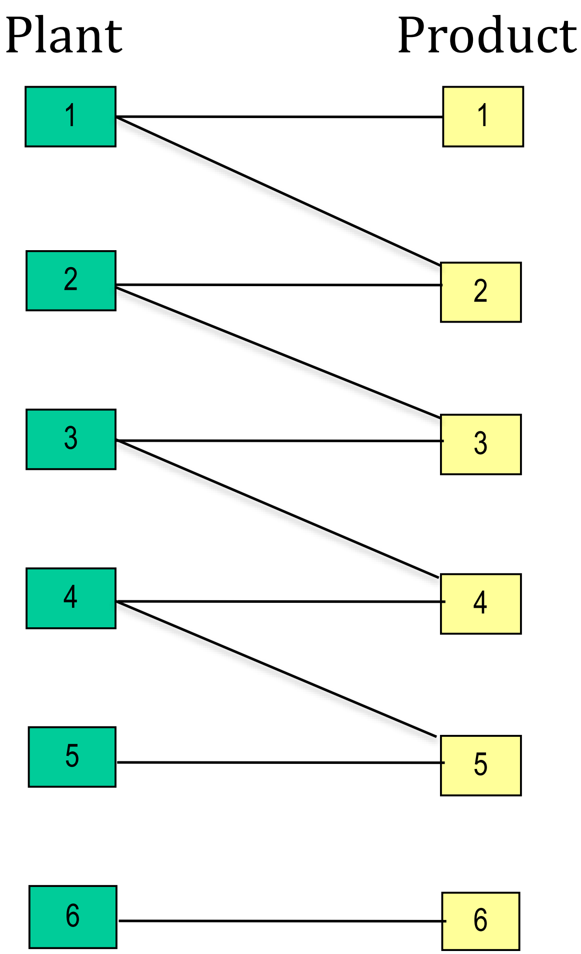
Design	Expected Sales	Marginal Benefit
No Flexibility	52034	
Add Arc (1,2)	52424	390
Add Arc (2,3)	52982	558
Add Arc (3,4)		
Add Arc (4,5)		
Add Arc (5,6)		
Add Arc (6,1)		

Increasing Incremental Benefit



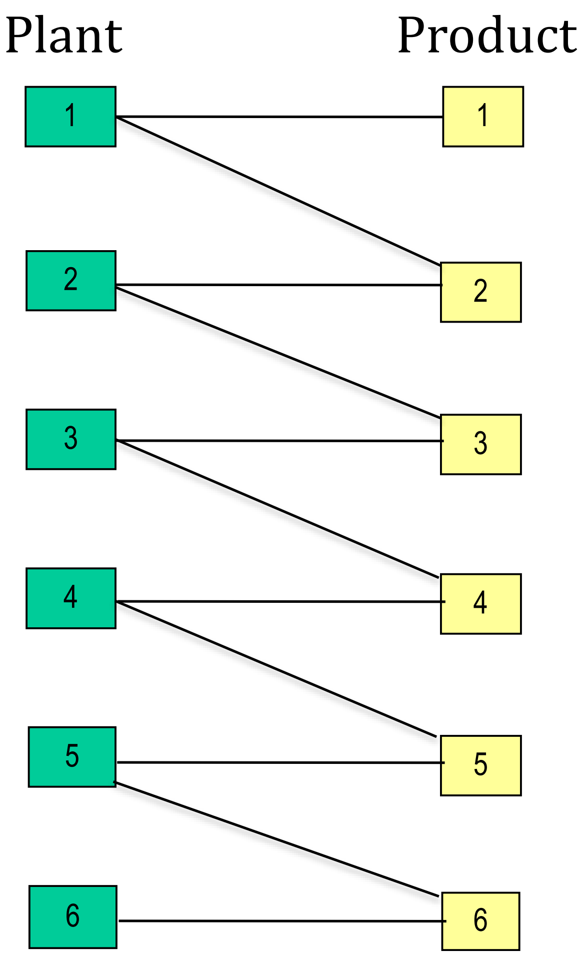
Design	Expected Sales	Marginal Benefit
No Flexibility	52034	
Add Arc (1,2)	52424	390
Add Arc (2,3)	52982	558
Add Arc (3,4)	53647	665
Add Arc (4,5)		
Add Arc (5,6)		
Add Arc (6,1)		

Increasing Incremental Benefit



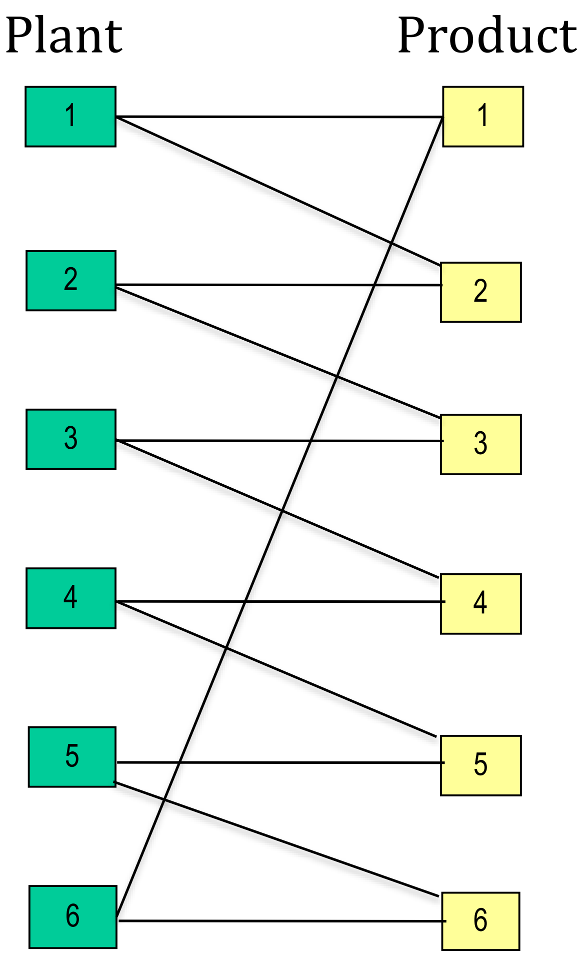
Design	Expected Sales	Marginal Benefit
No Flexibility	52034	
Add Arc (1,2)	52424	390
Add Arc (2,3)	52982	558
Add Arc (3,4)	53647	665
Add Arc (4,5)	54378	731
Add Arc (5,6)		
Add Arc (6,1)		

Increasing Incremental Benefit



Design	Expected Sales	Marginal Benefit
No Flexibility	52034	
Add Arc (1,2)	52424	390
Add Arc (2,3)	52982	558
Add Arc (3,4)	53647	665
Add Arc (4,5)	54378	731
Add Arc (5,6)	55163	785
Add Arc (6,1)		

Increasing Incremental Benefit

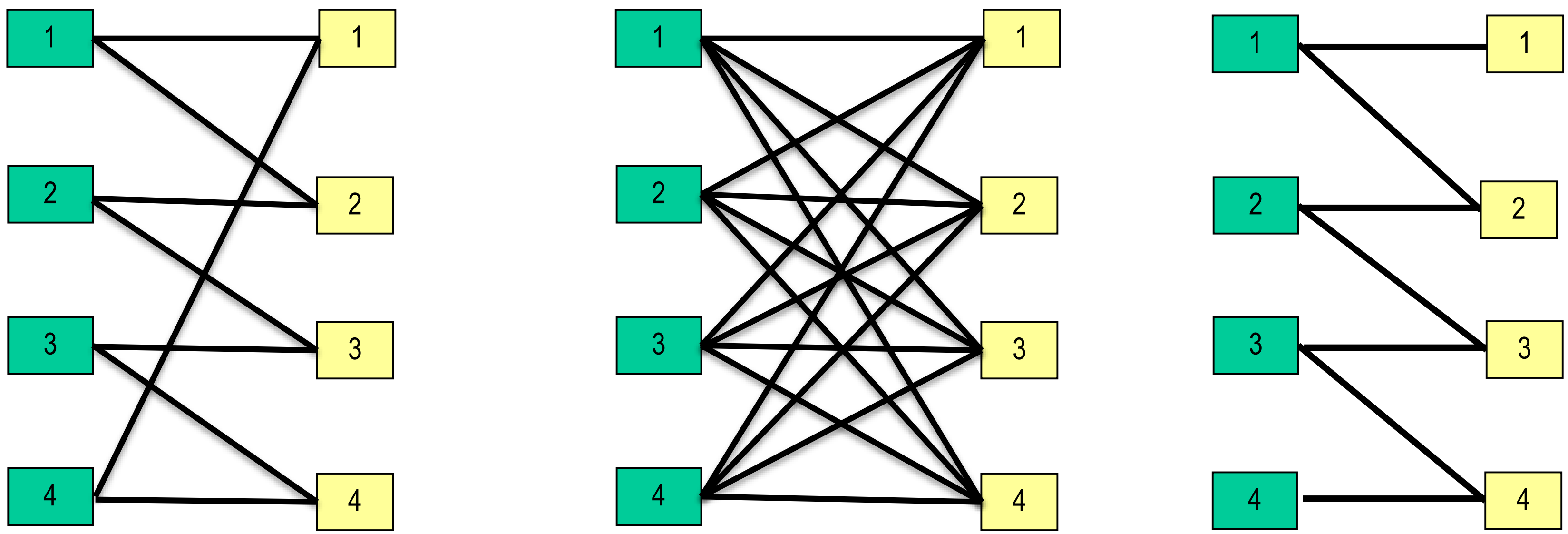


Design	Expected Sales	Marginal Benefit
No Flexibility	52034	
Add Arc (1,2)	52424	390
Add Arc (2,3)	52982	558
Add Arc (3,4)	53647	665
Add Arc (4,5)	54378	731
Add Arc (5,6)	55163	785
Add Arc (6,1)	56735	1572

Observed by papers such as [Hopp et al. 2004] & [Graves 2008]

Flexibility Structures

Long Chain | Full Flexibility | Open Chain



C_n | F_n | L_n

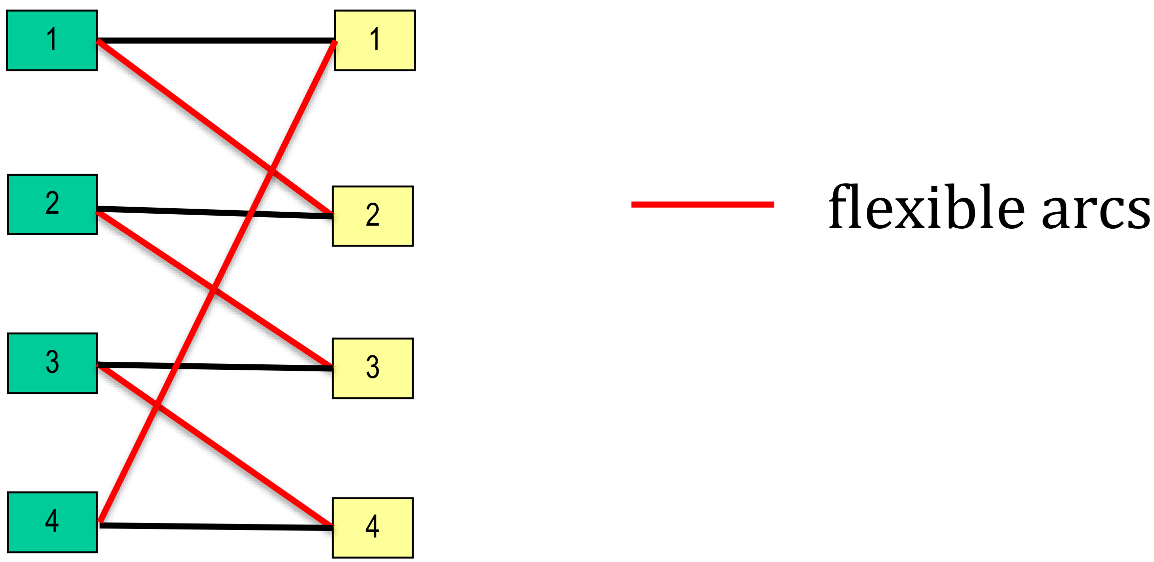
- **Assumptions :** # plants =# products; plant capacity is 1
- **Notations:** For a demand realization **d**, we use P (**d**, A) to denote the sales of flexibility structure A

Supermodularity

Theorem 1 (Supermodularity of Long Chain)

Given a fixed demand instance $\mathbf{d}$, for any α and γ that are flexible arcs in C_n, and any flexibility structure $A \subseteq C_n$,

$P(\mathbf{d}, A \cup \{\alpha, \gamma\}) - P(\mathbf{d}, A \cup \{\gamma\}) \geq P(\mathbf{d}, A \cup \{\alpha\}) - P(\mathbf{d}, A)$.



C_n

Supermodularity

Theorem 1 (Supermodularity of Long Chain)

Given a fixed demand instance $\mathbf{d}$, for any α and γ that are flexible arcs in C_n, and any flexibility structure $A \subseteq C_n$,

$P(\mathbf{d}, A \cup \{\alpha, \gamma\}) - P(\mathbf{d}, A \cup \{\gamma\}) \geq P(\mathbf{d}, A \cup \{\alpha\}) - P(\mathbf{d}, A)$.

- $P(\mathbf{d}, A \cup \{\alpha\}) - P(\mathbf{d}, A)$
  - increase in sales when we add α to A.
- $P(\mathbf{d}, A \cup \{\alpha, \gamma\}) - P(\mathbf{d}, A \cup \{\gamma\})$:
  - increase in sales when we add α to $A \cup \{\gamma\}$.

Supermodularity

Theorem 1 (Supermodularity of Long Chain)

Given a fixed demand instance **d**, for any α and γ that are flexible arcs in C_n, and any flexibility structure $A \subseteq C_n$,

$P(\mathbf{d}, A \cup \{\alpha, \gamma\}) - P(\mathbf{d}, A \cup \{\gamma\}) \geq P(\mathbf{d}, A \cup \{\alpha\}) - P(\mathbf{d}, A)$.

Application of [Gale and Politof 1981].

Corollary 1 (Increasing in Marginal Benefits)

The marginal benefits as the long chain is constructed is always increasing.

Observed by papers such as [Hopp et al. 2004] & [Graves 2008]

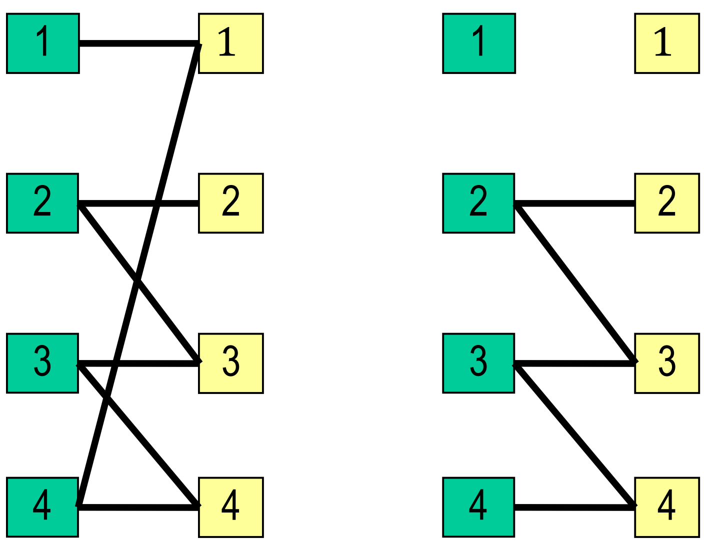
Decompose the Sales of the Long Chain

Theorem 2 (Decomposition of Long Chain)
Given a demand instance **d**,

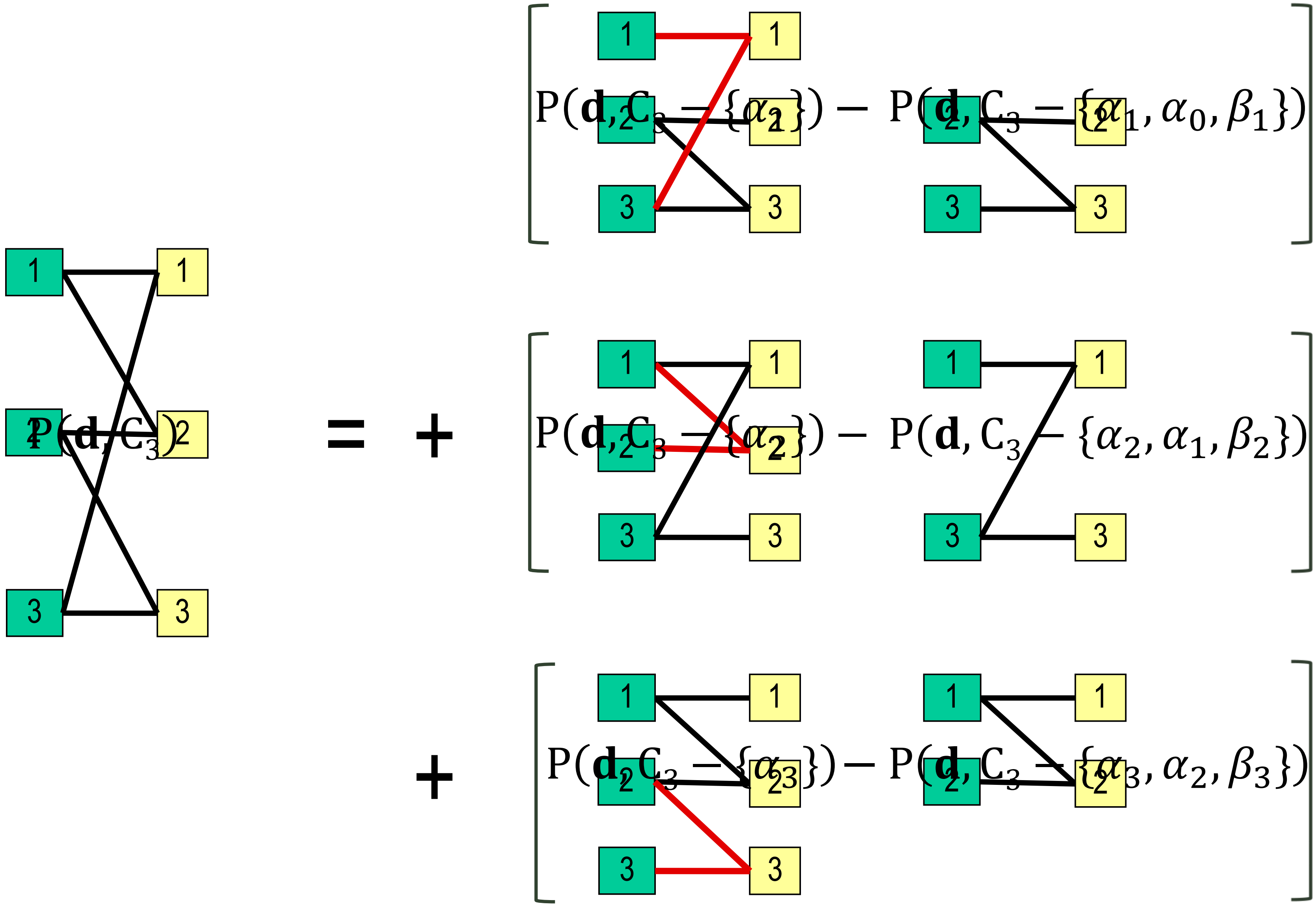
$$P(\mathbf{d}, C_n) = \sum_{i=1}^{n} \left(P(\mathbf{d}, C_n - \{\alpha_i\}) - P(\mathbf{d}, C_n - \{\alpha_i, \alpha_{i-1}, \beta_i\})\right)$$

where $\alpha_i=(i,i+1)$ for $i=1,\dots,n-1$, $\alpha_0= \alpha_n=(n,1)$ and $\beta_i=(i,i)$ for $i=1,\dots,n$.

Example ($n=4$, $i=1$): $P(\mathbf{d}, C_4 - \{\alpha_1\})$ $P(\mathbf{d}, C_4 - \{\alpha_1, \alpha_0, \beta_1\})$



Illustrating the Decomposition

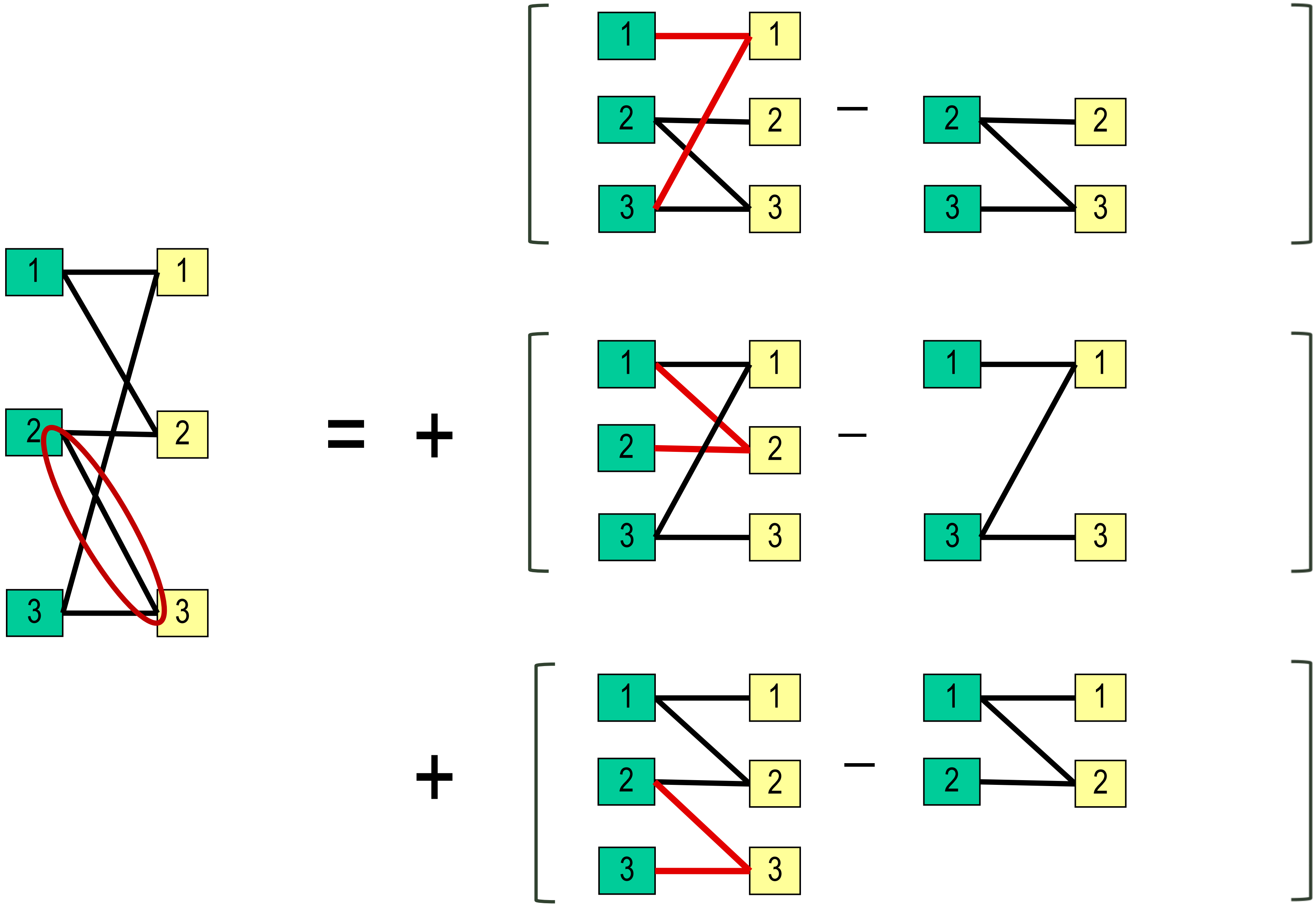


$$P(\mathbf{d}, C_3) = + \left[P(\mathbf{d}, C_3 = \{\alpha_1\}) - P(\mathbf{d}, C_3 = \{\alpha_1, \alpha_0, \beta_1\}) \right]$$

$$+ \left[P(\mathbf{d}, C_3 = \{\alpha_2\}) - P(\mathbf{d}, C_3 = \{\alpha_2, \alpha_1, \beta_2\}) \right]$$

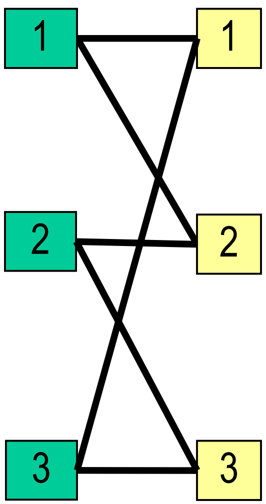
$$+ \left[P(\mathbf{d}, C_3 = \{\alpha_3\}) - P(\mathbf{d}, C_3 = \{\alpha_3, \alpha_2, \beta_3\}) \right]$$

Key Idea of the Proof



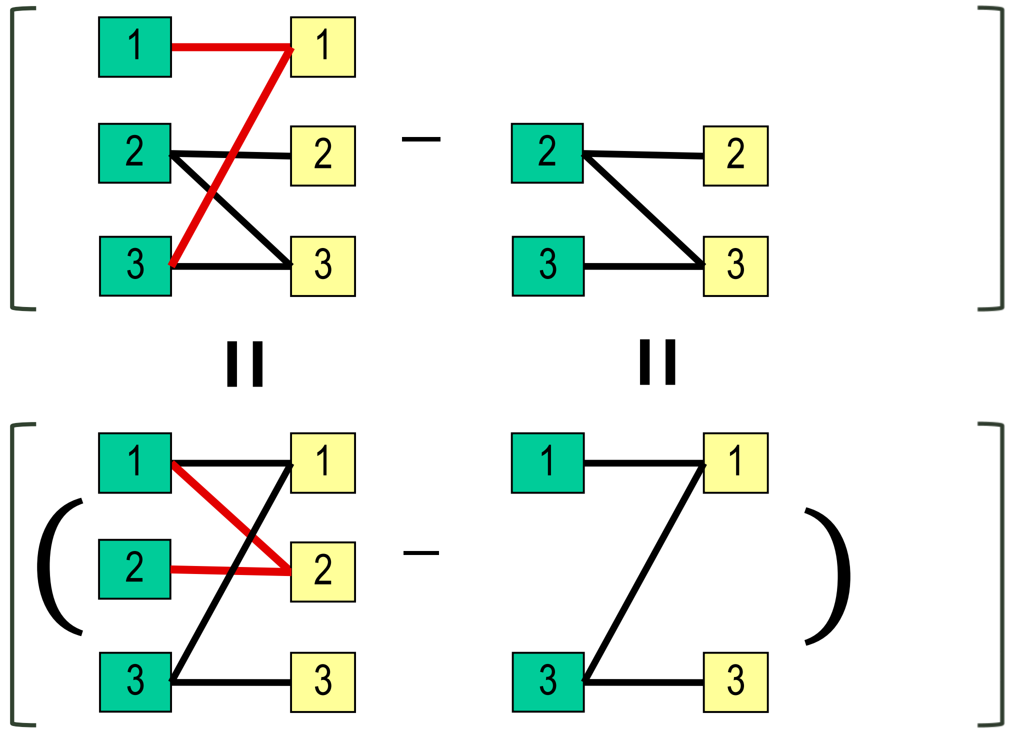
Under IID Demand

Under IID demand,



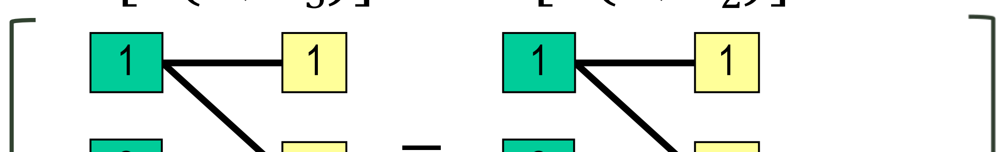
$E[P(\mathbf{D}, C_3)]$

$=$ + 3X



$E[P(\mathbf{D}, L_3)]$ $E[P(\mathbf{D}, L_2)]$

+



Proposition 1 (Expected Performance of Long Chain)
For IID demand, $E[P(\mathbf{D}, C_n)]/n = E[P(\mathbf{D}, L_n) - P(\mathbf{D}, L_{n-1})]$.

Impact

Corollary 2 (Risk Pooling of Long Chain)
$E[P(\mathbf{D}, C_n)]/n$ is increasing with n.

Corollary 3 (Optimality of Long Chain)
The expected sales of the long chain is greater than or equal to that of the any 2-flexibility structures.

Corollary 4 (Exponential Convergence of the Fill Rate)
$E[P(\mathbf{D}, C_n)]/n$ converges exponentially quickly with n.

A collection of several disjoint large chains can work just as well as the long chain.

Impact (Cont.)

Corollary 5 (Effectiveness of Long Chain)
For any integer $n \geq 2$,

$$\frac{E[P(\mathbf{D}, F_n)]}{n} - \frac{E[P(\mathbf{D}, C_n)]}{n} \leq \frac{E[P(\mathbf{D}, F_{n+1})]}{n+1} - \frac{E[P(\mathbf{D}, C_{n+1})]}{n+1} \leq 1 - u,$$

where $u = \lim_{k\to\infty} \frac{E[P(\mathbf{D}, C_k)]}{k}$.

E.g. when $D_1 = N(1,0.33)$, $\lim_{k\to\infty} \frac{E[P(\mathbf{D},C_k)]}{k} \approx 0.96$, [Chou et al 2010].

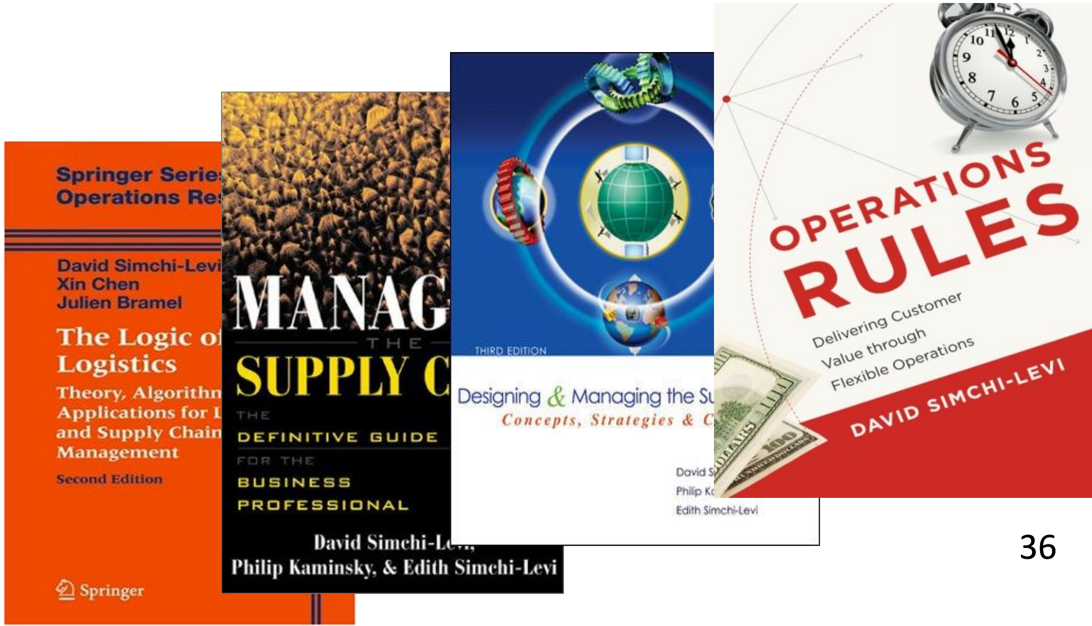
$$\frac{E[P(\mathbf{D}, F_n)]}{n} - \frac{E[P(\mathbf{D}, C_n)]}{n} \leq 0.04 \text{ and } \frac{E[P(\mathbf{D}, C_n)]}{E[P(\mathbf{D}, F_n)]} \geq 0.9568.$$

Supply chain resiliency with process flexibility and inventory:

- Implementation at Ford –Recently Awarded the INFORMS Daniel H. Wagner Prize for Excellence in Operations Research Practice

A Few Stories …

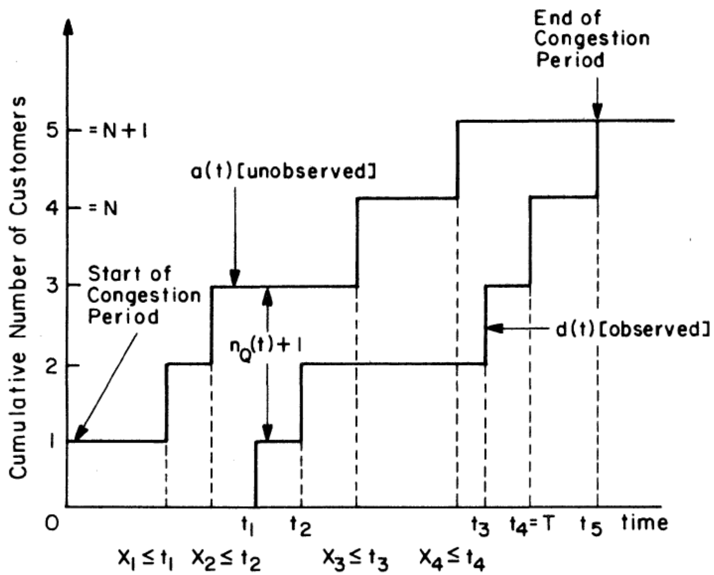
- **From Theory to Practice**
  - School Bus Routing in NYC
- **From Practice to Theory**
  - Flexibility at PepsiCo
- **Merging Theory and Practice**
  - Online Retailing
- **Conclusions**
  - Data Driven Models



Data Driven Research--The Opportunity

- **Extensive use of data to identify models that drive decisions and actions**
  - **Spans Statistics, Computational Science and Operations Research techniques**

The Queue Inference Engine: Deducing Queue Statistics from Transactional Data. Richard C. Larson *Management Science*, (1990) Vol. 36, pp. 586-601



Given customers transaction times, can you Infer the number of customers waiting to use the machine?

Online Retailing: Online Fashion Sample Sales Industry

- Offers extremely limited-time discounts (“flash sales”) on designer apparel & accessories
- Emerged in mid-2000s and has had nearly 50% annual growth in last 5 years

Snapshot of Rue La La’s Website

From the Reserve: Watches by Rolex & Cartier ▸

CLOSING IN 2 DAYS, 19:47:42

Judith Ripka Jewelry & Watches ▸

CLOSING IN 2 DAYS, 19:47:42

Check Off His List: Gift Ideas Under $100 ▸

CLOSING IN 2 DAYS, 19:47:42

Saucony Women ▸

CLOSING IN 1 DAY, 19:47:42

Furs by Christian Dior & More: Picks by WGACA ▸

CLOSING IN 1 DAY, 19:47:42

Saucony Men ▸

CLOSING IN 1 DAY, 19:47:42

"Style"

Saucony "Triumph 10" Running Shoe

~~$130.00~~ **$79.90**

Saucony "Progrid Guide 6" Running Shoe

~~$110.00~~ **$65.90**

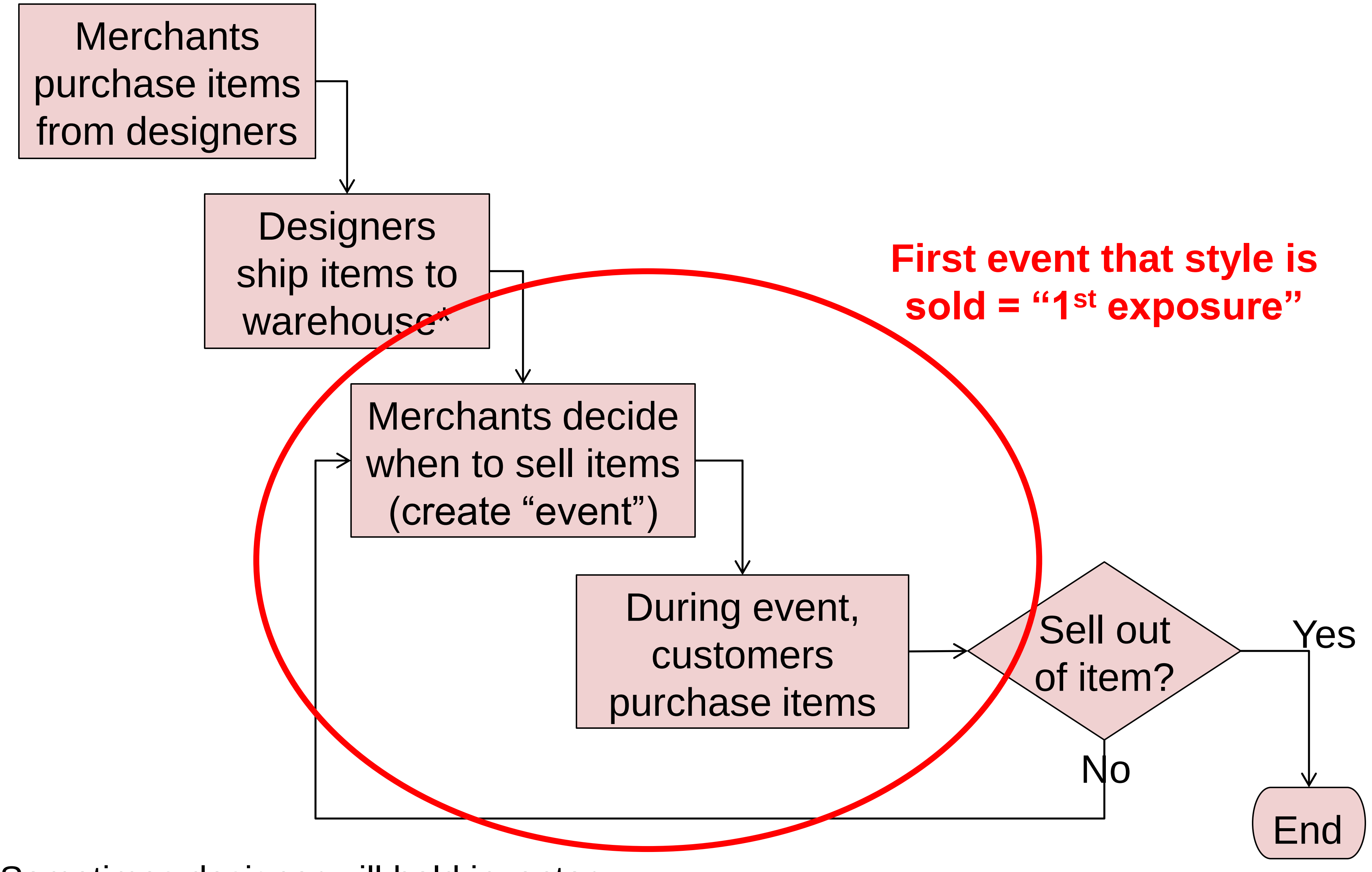
Saucony "Triumph 10" Running Shoe

~~$130.00~~ **$79.90**

“SKU”

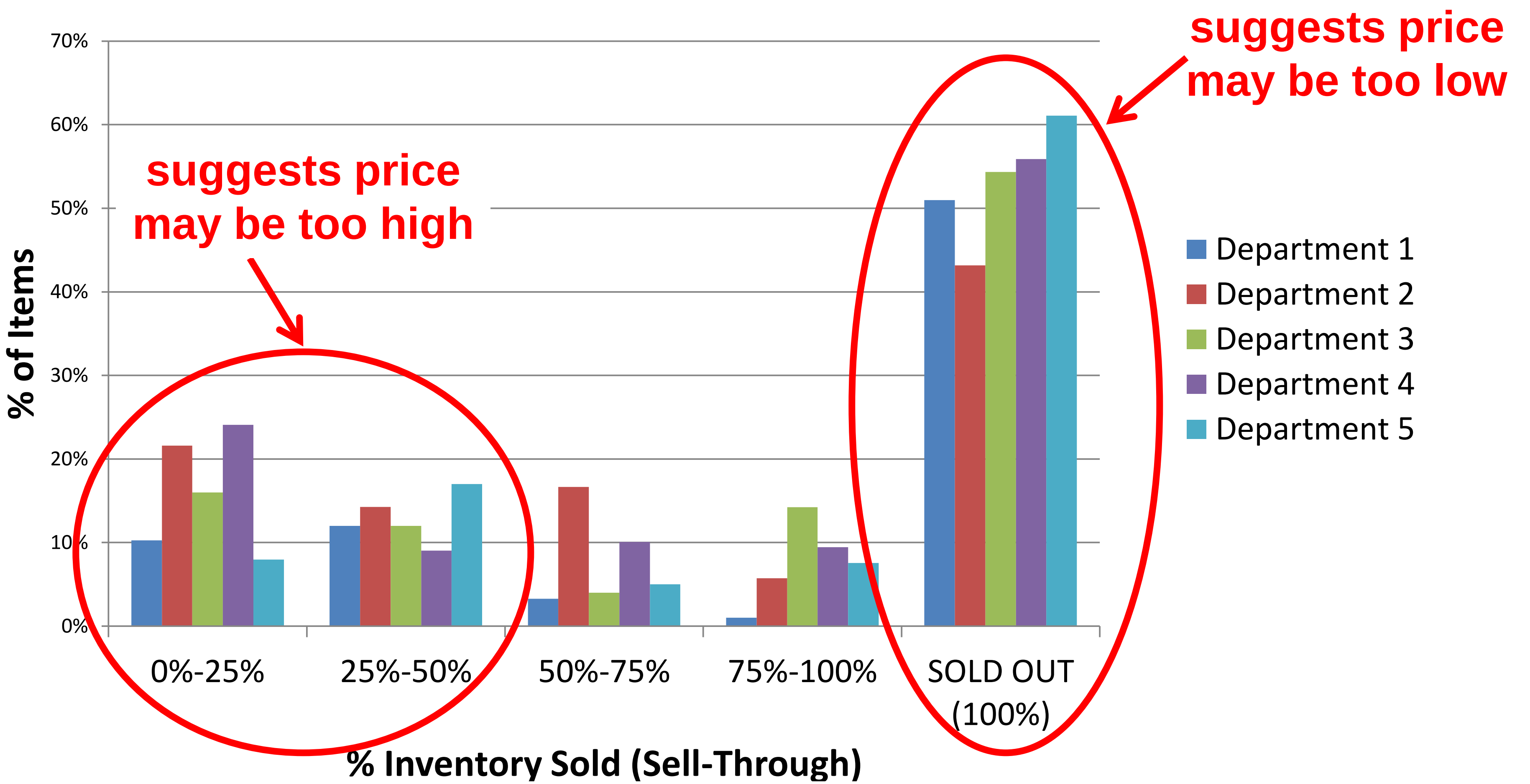
Saucony "Progrid Guide 6" Running Shoe
$110.00 $65.90
Size
5 5.5 6 6.5 7 7.5 8 8.5 9
9.5 10 10.5 11
Quantity
1
ADD TO BAG
Sign up for Quick! Buy It.
Never miss out on something you love.

Flash Sales Operations



*Sometimes designer will hold inventory

Sell-Through Distribution of New Products



*Data disguised to protect confidentiality

Approach

Goal: Maximize expected revenue from 1st exposure styles

Demand Forecasting

Challenges:

- Predicting demand for items that have never been sold before
- Estimating lost sales

Techniques:

- Clustering
- Machine learning models for regression

Price Optimization

Challenges:

- Structure of demand forecast
- Demand of each style is dependent on price of competing styles → exponential # variables

Techniques:

- Novel reformulation of price optimization problem
- Creation of efficient algorithm to solve daily

Forecasting Model:
Explanatory Variables Included

Products

- Department
- Class
- Color Popularity
- Size Popularity
- Brand Type A/B
- Brand Popularity

Combination

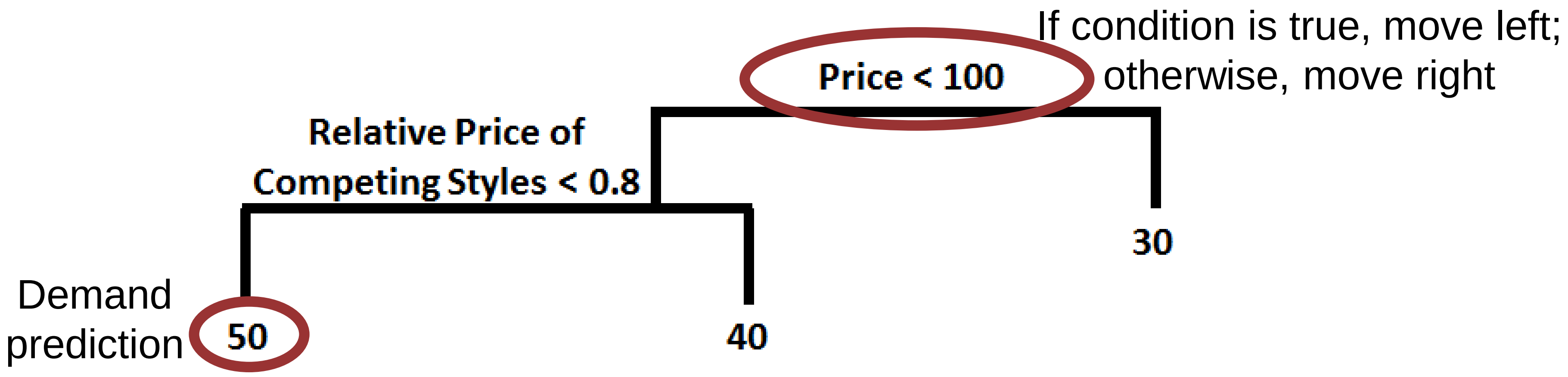
- Price
- % Discount = (1 – Price / MSRP)
- # Concurrent Events in Department
- # Styles Sold in Same Subclass and Event (i.e. # Competing Styles)
- Relative Price of Competing Styles
- # Branded Events in Previous 12 Months

Events

- Year
- Month
- Week Day / Time
- Event Type
- Event Length

- Tested several machine learning techniques
  - Regression trees performed best

Regression Tree – Illustration



Approach

Goal: Maximize expected revenue from 1^{st} exposure styles

Demand Forecasting

Challenges:

- Predicting demand for items that have never been sold before
- Estimating lost sales

Techniques:

- Clustering
- Machine learning models for regression

Price Optimization

Challenges:

- Structure of demand forecast
- Demand of each style is dependent on price of competing styles → exponential # variables

Techniques:

- Novel reformulation of price optimization problem
- Creation of efficient algorithm to solve daily

Price Optimization Complexity

- Three features used to predict demand are associated with pricing
  - Price
  - % Discount = $\dfrac{1-\text{Price}}{\text{MSRP}}$
  - Relative Price of Competing Styles =

$$\frac{\text{Price}}{\text{Avg. Price of Competing Styles}}$$

- Pricing must be optimized concurrently for all competing styles
  - Would be impractical to calculate revenue for all potential combinations of prices
- We developed an efficient algorithm to solve on a daily basis

Field Experiment

- Goal: to identify whether or not raising prices would decrease sales
- Set lower bound on price = legacy price (cost + markup)
  - Model only recommends price increases (or no change)
- Identified ~6,000 styles where tool recommended price increases
- **Overall impact = 10% increase in revenue**

  → Much larger profit margin impact

Dynamic Pricing

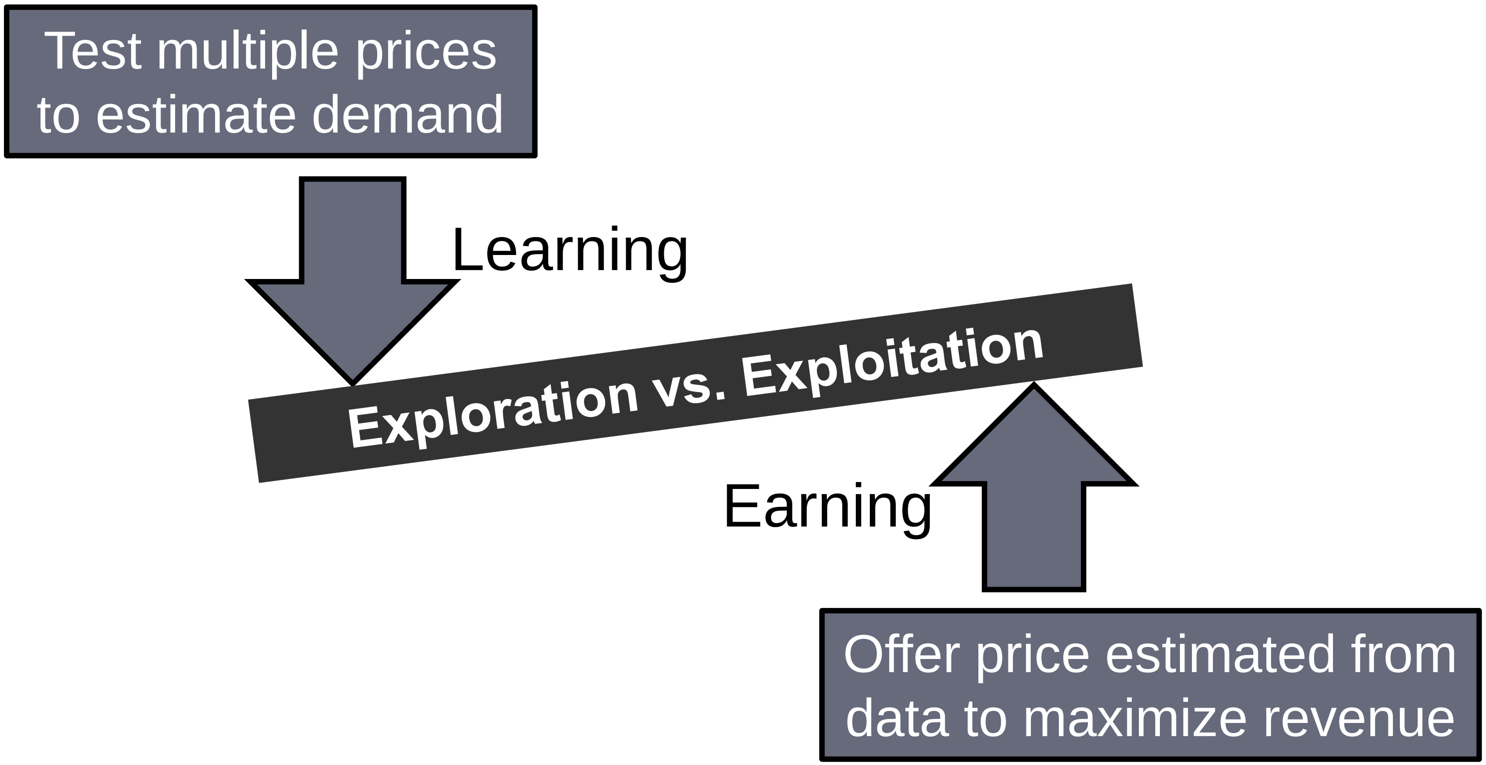
- What if you can change a style's price throughout the event?

{\$24.90, \$29.90, \$34.90, \$39.90} possible price set, p_i

d_1 d_2 d_3 d_4 mean demand, d_i (purchase probability)

- Given unlimited inventory and known demand, select price with highest revenue = $p_i * d_i$
- Challenges
  1. Unknown demand
  2. Limited inventory
  3. Finite selling season

Exploration vs. Exploitation Tradeoff



Demand Hypotheses

- m demand function hypotheses:

$$\{\varphi_1, \varphi_2, \ldots, \varphi_m\}$$

$$\varphi_i : P \to [0,1]$$

- If hypothesis i holds, customer buys with probability $\varphi_i(p)$
- The retailer *does not* know the true hypothesis

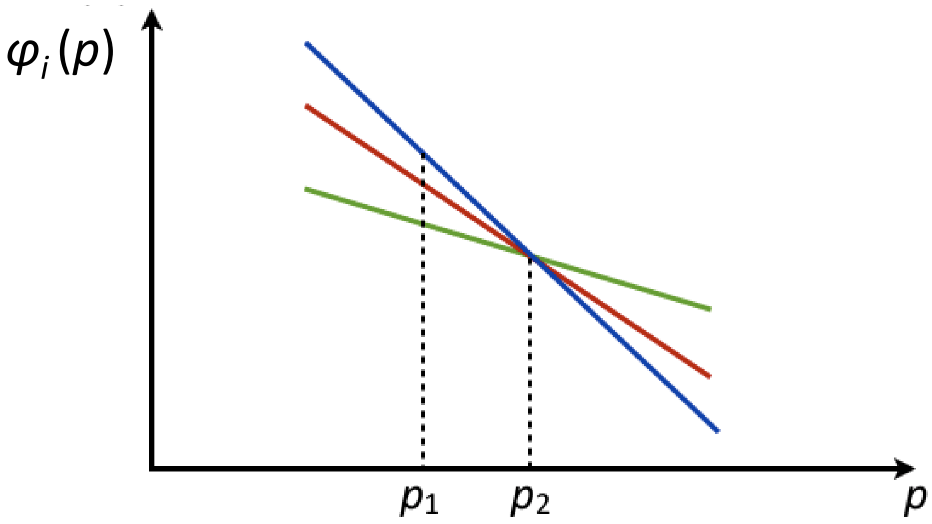


Fig: An example with 3 hypotheses

Unlimited Inventory, Finite Selling Season

- Suppose there are n customers.
- In a fixed price strategy: Regret $\sim O(n)$
  - Regret = revenue of an oracle who knows the true hypothesis – revenue of the retailer
- If the retailer can change price once, the regret $\sim \mathbf{O}(\boldsymbol{log}\ \boldsymbol{n})$
- In general, if the retailer can change price *k* times, the regret $\sim O(\underbrace{log\ log \ldots log}_{k \text{ times}}\ n)$
- Best possible heuristic—a matching lower bound
- Hence, Regret$(\boldsymbol{n}) = \Theta(\log \ldots \log \boldsymbol{n})$

Limited Inventory, Finite Selling Season

- Continuous exploration & exploitation
- Learns demand at each price to maximize revenue
- Model as Multi-armed bandit problem with inventory
- Thompson Sampling plus Linear Optimization

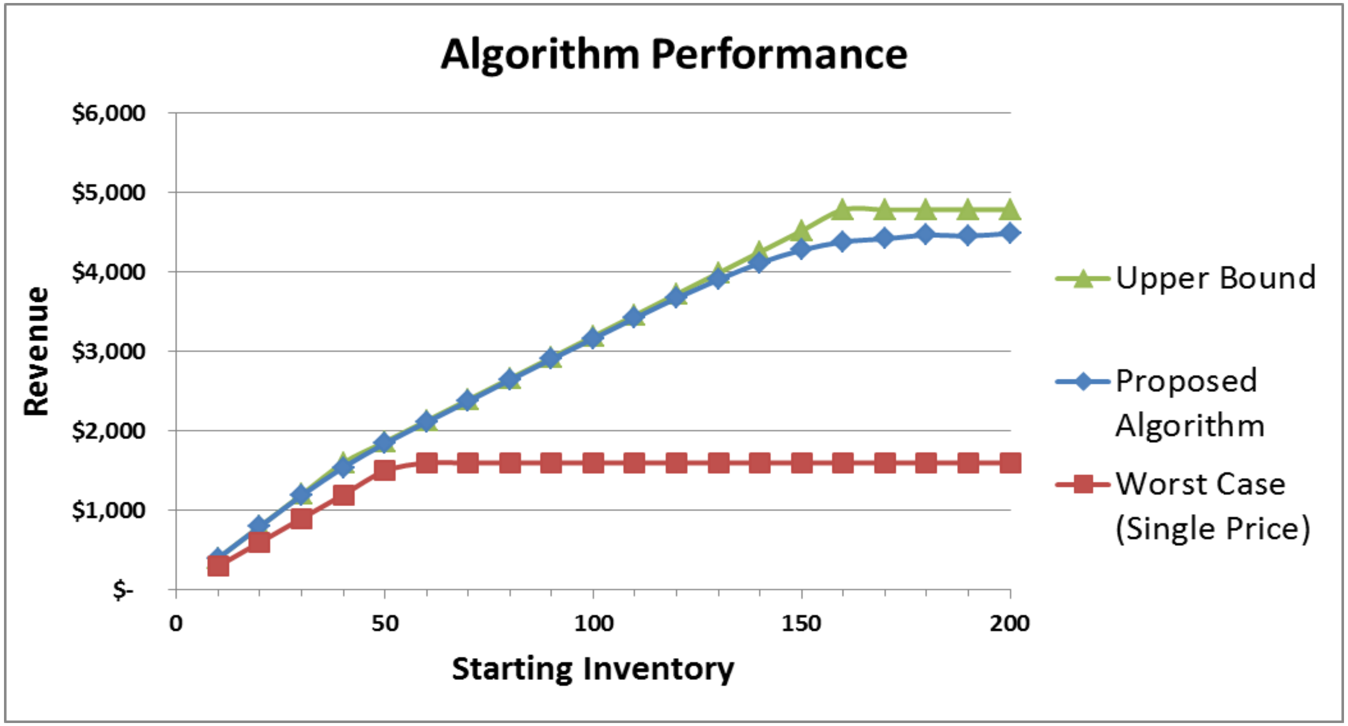
Theoretical Results

- Using the modified algorithm, we have

$$Regret(T) \leq O(\sqrt{T \log^2 T}),$$

- Matches lower bound $\Omega(\sqrt{T})$

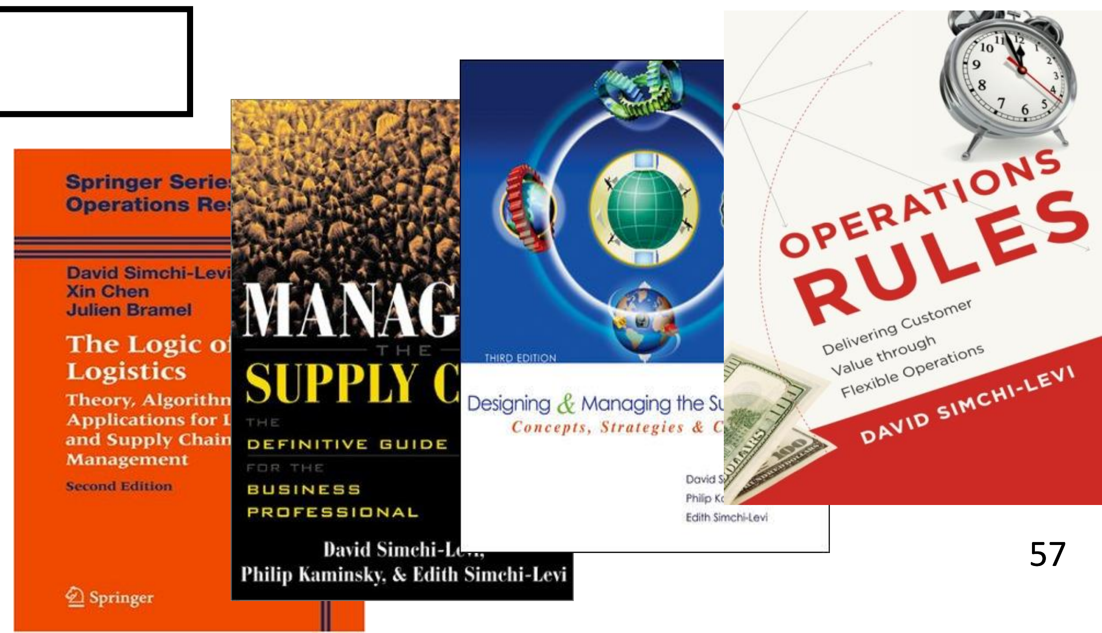
Algorithm Performance: Simulations



price = [29.9,39.9]; demand = [0.008,0.002]; time = 20k; # simulations = 1,000

A Few Stories …

- **From Theory to Practice**
  - School Bus Routing in NYC
- **From Practice to Theory**
  - Flexibility at PepsiCo
- **Merging Theory and Practice**
  - Online Retailing
- **Conclusions**
  - Data Driven Models



Two Types of Data Driven Research

- **Type I: Focus on a specific goal**
  - Examples: Increase Revenue; Decrease Cost; Reduce the spread of an epidemic
  - Challenge: Let the data identify the specific issues, opportunities and models
  - Impact: Data Driven Models

- **Type II: Open-ended search for correlations and relationships without any clear goal in mind**
  - Typically the objective of data mining: Uncover economic or other relationships by analyzing huge data sets

Data Driven Models (DDM)

- **Two Examples: ATM Model; Online Retailing**
- **DDM is linked with Decision Making**
  - **Fits with our unique set of skills and tools**
- **Allows to distinguish our profession from economics and statistics**
  - **Apply data mining and focus on Type II**
- **It can be different than "empirical research"**
- **The early history of OR focused on DDM**
  - **Methods of Operations Research** by Philip M. Morse and George E. Kimball, published in 1951

Example (Philip M. Morse and George E. Kimball, 1951)

- **Mail order delivery, selling to low-income rural families using COD (Cash on Delivery) agreement**
    - **Many customers refuse product upon arrival**
- **Statistical analysis showed high correlation between COD refusal and the time original order was made by the family and delivery time by the mailman**
    - **If item does not arrive at a certain time, money spent elsewhere and COD item was refused**
- **Solution: Limit market area covered by the delivery service-- Network Design model**
    - **Impact: “considerable reduction in lost sales”**

The Future of OM Research

- **Emphasize data driven in research and teaching**
  - Today, there is too little reliance on data in formulating models and identifying research opportunities
  - Systems involving people can be difficult to analyze unless you have data about behavior
- **Develop new engineering and scientific methods that explain, predict and change behavior**
- **Need to develop an open source data repository**
  - Example: MIT wide competition with data from a public organization

References…

- Simchi-Levi, D. (2014). OM Research: From Problem Driven to Data Driven Research. ***Manufacturing and Service Operations Management, Vol. 16, pp. 1-10.***
- Johnson, K., Lee, B. H. A., and D. Simchi-Levi (2014). Analytics for an Online Retailer: Demand Forecasting and Price Optimization. **2014 INFORMS Revenue Management and Pricing Section Practice Award.**
- Cheung W-C, Simchi-Levi D., Wang H. (2014). Dynamic Pricing and Demand Learning with Limited Price Experimentation. Submitted.
- Johnson, K., Simchi-Levi, D., and H. Wang (2014). Inventory-Constrained Demand Learning and Dynamic Pricing.
- Simchi-Levi, D., W. Schmidt and Y. Wei (2014), Managing Unpredictable Supply Chain Disruptions. **HBR,** January-February 2014, pp. 96-101. **2014 INFORMS Daniel H Wagner Prize for Excellence in Operations Research Practice**