

# People and Data: Understanding Customer Behavior

Andreas S. Weigend, Ph.D.  
former chief scientist, amazon.com

[weigend.com](http://weigend.com)



## Outline 提纲

- Making decisions based on experiments (A-B tests)  
以A-B实验为基础进行决策
  - ▶ Three ingredients for innovation  
革新的三大因素
  - ▶ Revealed vs stated preferences  
被揭示 vs 被说出的偏好

- The iterative process of modeling  
反复的建模过程

▶ Define 定义 → Measure 度量 → Describe 描述 → Predict 预测 → Act 行动

- Some insights into online customer behavior  
一些对在线客户行为的理解
  - ▶ Levels of analysis and actionability
    - Personalization vs Occasionalization
  - ▶ Behavioral economics

Amazon Exclusive!! Order a Segway now! It's only at Amazon.

amazon.com

VIEW CART | WISH LIST | YOUR ACCOUNT | HELP



WELCOME YOUR STORE BOOKS APPAREL & ACCESSORIES ELECTRONICS TOYS & GAMES MAGAZINE SUBSCRIPTIONS COMPUTER & VIDEO GAMES MUSIC SEE MORE STORES Your Gold Box

SEARCH MUSIC BROWSE STYLES CLASSICAL TOP SELLERS NEW & FUTURE RELEASES FREE DOWNLOADS ESSENTIAL CDs USED MUSIC

Search: Popular Music GO Browse: All styles: GO




Add \$16.01 to your order to qualify for FREE Super Saver Shipping! Some restrictions apply.

Customers who bought *Kind of Blue* also bought:

 <b>Time Out</b> ~ Dave Brubeck, et al Our Price: <b>\$8.99</b> Used & new from <b>\$5.79</b> <a href="#">Add to cart</a>	 <b>A Love Supreme</b> ~ John Coltrane Our Price: <b>\$13.49</b> <a href="#">Add to cart</a>	 <b>Birth of the Cool</b> ~ Miles Davis Our Price: <b>\$9.99</b> Used & new from <b>\$9.50</b> <a href="#">Add to cart</a>
--	--	---

[Explore similar items](#)

Customers who bought items in your Shopping Cart also bought:

 <b>Sketches of Spain</b> ~ Miles Davis Our Price: <b>\$11.98</b> Used & new from <b>\$2.00</b> <a href="#">Add to cart</a>	 <b>Blue Train</b> ~ John Coltrane Our Price: <b>\$9.99</b> Used & new from <b>\$5.99</b> <a href="#">Add to cart</a>	 <b>Saxophone Colossus</b> ~ Sonny Rollins Our Price: <b>\$8.99</b> Used & new from <b>\$6.90</b> <a href="#">Add to cart</a>
--	--	--

Customers who shopped for *Kind of Blue* also shopped for:


**YOUR SHOPPING CART**

[Proceed to Checkout](#)

Wait! Add \$16.01 to your order to qualify for FREE Super Saver Shipping. Some restrictions apply.

Show gift options during checkout

Added to your Shopping Cart:

  
**Kind of Blue** ~ Miles Davis, et al  
Our Price: **\$8.99**  
Quantity: 1  
Subtotal: **\$8.99**  
[Edit shopping cart](#)

[Proceed to Checkout](#)

[Sign in](#) to turn on 1-Click ordering.

Items in your Shopping Cart always reflect the most recent price displayed on their product pages.

It's only at Amazon

WELCOME YOUR STORE BOOKS APPAREL & ACCESSORIES ELECTRONICS TOYS & GAMES KITCHEN & HOUSEWARES HOME & GARDEN MUSIC SEE MORE STORES Your Gold Box

SEARCH MUSIC BROWSE STYLES CLASSICAL TOP SELLERS NEW & FUTURE RELEASES FREE DOWNLOADS ESSENTIAL CDs USED MUSIC

Search: Popular Music GO Browse: All styles: GO

Add \$16.01 to your order to qualify for FREE Super Saver Shipping! Some restrictions apply.


**YOUR SHOPPING CART**

[Proceed to Checkout](#)

Wait! Add \$16.01 to your order to qualify for FREE Super Saver Shipping. Some restrictions apply.

Show gift options during checkout

Added to your Shopping Cart:




  
**Kind of Blue** ~ Miles Davis, et al  
Our Price: **\$8.99**  
Quantity: 1  
Subtotal: **\$8.99**  
[Edit shopping cart](#)

[Proceed to Checkout](#)

[Sign in](#) to turn on 1-Click ordering.




Items in your Shopping Cart always reflect the most recent price displayed on their product pages.

Customers who bought *Kind of Blue* also bought:

 <b>Time Out</b> ~ Dave Brubeck, et al Our Price: <b>\$8.99</b> Used & new from <b>\$5.79</b> <a href="#">Add to cart</a>	 <b>A Love Supreme</b> ~ John Coltrane Our Price: <b>\$13.49</b> <a href="#">Add to cart</a>	 <b>Birth of the Cool</b> ~ Miles Davis Our Price: <b>\$9.99</b> Used & new from <b>\$9.50</b> <a href="#">Add to cart</a>
--	--	---

[Explore similar items](#)

Customers who bought items in your Shopping Cart also bought:

 <b>Sketches of Spain</b> ~ Miles Davis Our Price: <b>\$11.98</b> Used & new from <b>\$2.00</b> <a href="#">Add to cart</a>	 <b>Blue Train</b> ~ John Coltrane Our Price: <b>\$9.99</b> Used & new from <b>\$5.99</b> <a href="#">Add to cart</a>	 <b>Saxophone Colossus</b> ~ Sonny Rollins Our Price: <b>\$8.99</b> Used & new from <b>\$6.90</b> <a href="#">Add to cart</a>
--	--	--

Customers who shopped for *Kind of Blue* also shopped for:

## Result: Right vs Left 结果: 左右对比的结果

- Metrics  
度量准则

- ▶ **Conversion:** Percentage of visits placing an order  
变化: 下订单者增加的比例
- ▶ **Order size:** Additional items (from the second page) put in cart  
订单规模: (第2页开始) 更多附加物品被放入购物车

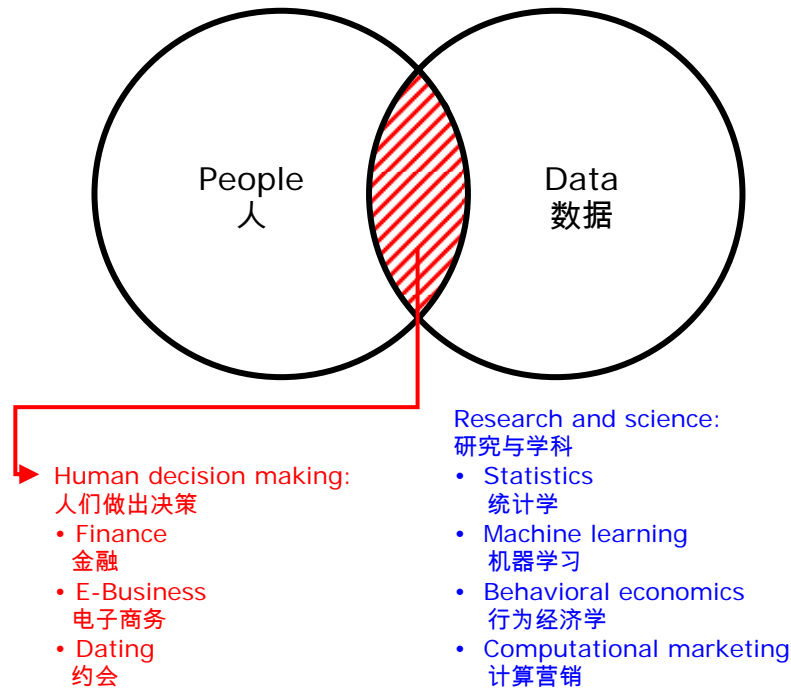
- Some details  
细节

- ▶ **Relative increase:** Blue band on right compared with blue band on left  
相对增加: 左右两边的蓝条相比

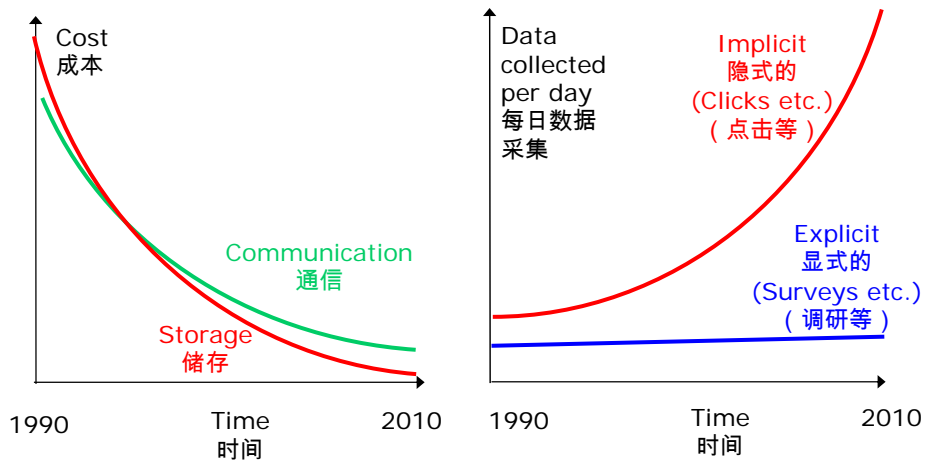
<i>All customers</i> 所有的消费者	<i>Existing customers</i> 现有消费者
Cart-adds from 2 <sup>nd</sup> page: 在第二页购物车物品增加: +0.6%	Cart-adds from 2 <sup>nd</sup> page: 购物车从第二页增加: +0.8%
Wishlist-adds: 购物预期增加: +1.4%	DVD Cart-adds: DVD 购物车: +0.8%
DVD (\$): +1.1%	DVD (\$): +1.0%

6

© 2004 by Weigend Associates LLC | www.weigend.com



Why Now? 为什么是现在?

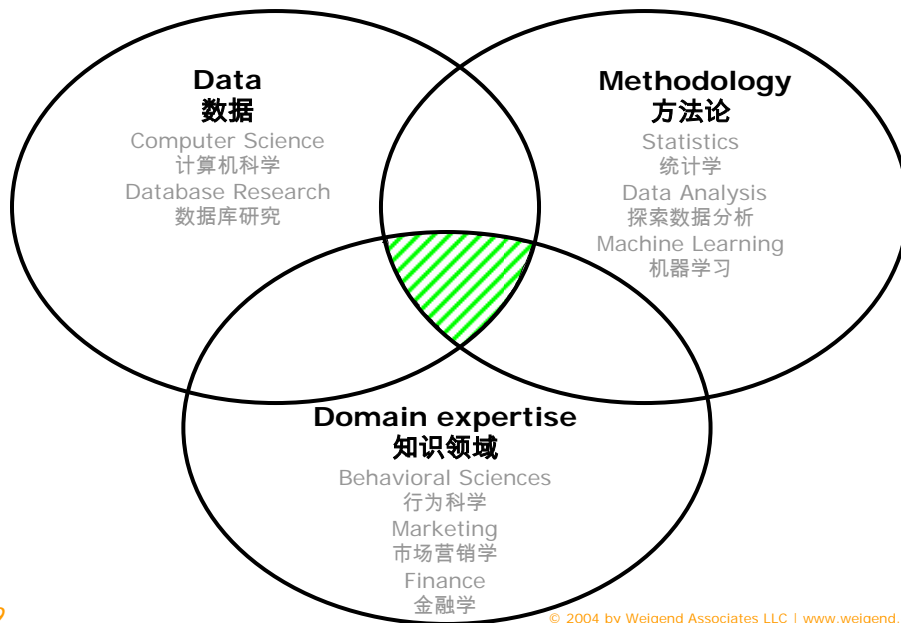


- Data collected implicitly: Dramatic growth over time  
隐式地采集数据：时刻保持急剧增长
- Data collected explicitly / tacitly: Constant over time  
显式地采集数据/无声的：时刻持续进行

8

© 2004 by Weigend Associates LLC | www.weigend.com

Three Ingredients For Innovation 革新的三大因素



9

© 2004 by Weigend Associates LLC | www.weigend.com

- Characterize **paths** through website  
分析通过网站的路径的特征
  - "Modeling Online Browsing and Path Analysis Using Clickstream Data"  
by Alan L. Montgomery, Shibo Li, Kannan Srinivasan, and John C. Liechty.  
*Marketing Science* (2004).
- Understand and influence **conversion**  
理解并影响转变
  - "A Model of Web Site Browsing Behavior Estimated on Clickstream Data"  
by Randolph E. Bucklin and Catarina Sismeiro, *J of Marketing Research* 40 (2003).
  - "Dynamic Conversion Behavior at E-Commerce Sites"  
by Wendy W. Moe and Peter S. Fader. *Management Science* (2004).
- Predict **intention and modality** of the visit  
预测访问意图、目标和状态
  - "Seize the Occasion"  
by Horacio D. Rozanski, Gerry Bollman, and Martin Lipman.  
*Strategy and Business* (2001).
- Compute and apply **customer network value**  
测定并应用客户网络价值
  - "Mining the Network Value of Customers"  
by Pedro Domingos and Matt Richardson, KDD-2001. ACM Press.

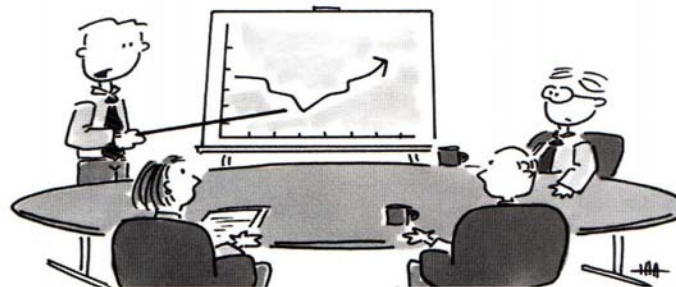
10

© 2004 by Weigend Associates LLC | www.weigend.com

Example: Invest \$10M to Improve Customer Satisfaction  
案例：投资1000万美元提高客户满意度

- Base decision on analysis of behavioral data  
以行为数据分析为基础的决策

- ▶ Quantify  
定量
- ▶ Model  
模型
- ▶ Act  
行为



"So, as you can see, customer satisfaction is up considerably since phasing out the complaint forms."

所以，如大家所见，自从我们停用投诉表以后，用户满意度得到了显著的上升

- Consider 思索
  - ▶ Increase selection? 增加选择？
  - ▶ Increase availability? 提高可用性？
  - ▶ Reduce clutter on web site? 减少网站复杂性？
  - ▶ Improve product search algorithms? 改进产品搜索算法？

All results for: blue pants

Search: Amazon.com for blue pants

Refine your search:

Find blue pants in these categories:

- [Books](#) (41,731)
- [Apparel](#) (2,017)
- [Sports & Outdoors](#)
- [Beta \(what is this?\)](#) (351)
- [Health & Personal Care \(what is this?\)](#) (4)
- [Home & Garden](#) (2)
- [Kitchen & Housewares](#) (2)
- [Toys & Games](#) (1)
- [Industrial Supplies](#) (67)
- [Lifestyle & Gifts](#) (38)



**Giorgio Sant Angelo Silk Dress Slacks**  
**Apparel:** [See all 2,017 items](#) ([Rate this item](#))  
**Buy new:** ~~\$49.99~~ **Sale: \$7.49**



**ruffle capris** (The Children's Place)  
**Apparel:** [See all 2,017 items](#) ([Rate this item](#))  
**Buy new:** ~~\$14.50~~ **Sale: \$6.99**

**Sponsored Links:** [What's this?](#)

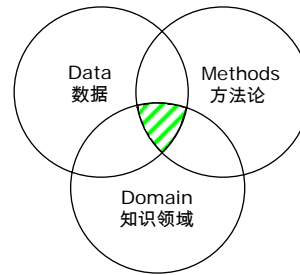
**Buy Jeans - The Buckle** -- Lucky Brand, Silver, Mavi, Ecko, BKE & Shady. -- [www.buckle.com](http://www.buckle.com)

**Blue Jeans** -- Shop for deals on Clothing here! Simply Fast Savings -- [www.Shopping.com](http://www.Shopping.com)

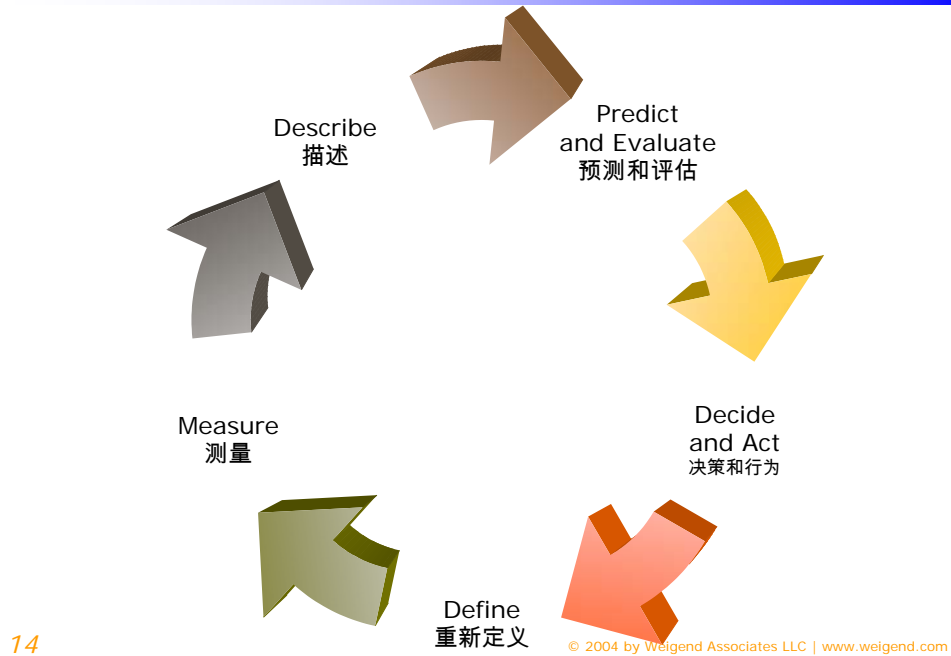
[The boys in the sky-blue pants : the men and events at Camp](#)

Outline 提纲

- Make decisions based on experiments (A-B tests)  
以A-B实验为基础进行决策
  - ▶ The need for a scientific framework  
我们需要一个科学的框架
  - ▶ Three ingredients for innovation  
革新的三大因素
  - ▶ Revealed vs stated preferences  
被揭示 vs 被说出的偏好
- The iterative process of modeling  
反复的建模过程
- Some insights into online customer behavior  
一些对在线客户行为的理解
  - ▶ Levels of analysis and actionability
    - Personalization vs Occasionalization
  - ▶ Behavioral economics



## The Iterative Process of Modeling 模拟反复过程

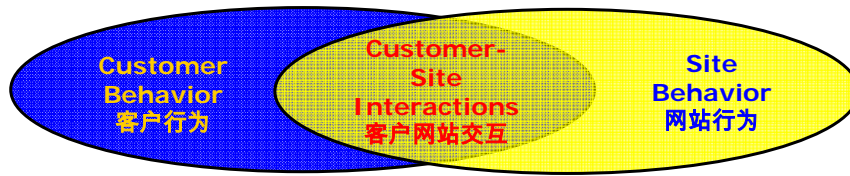


### 1. Define Objectives 定义目标

- ▶ Stock price  
股票价格
- ▶ Profit  
收益
- ▶ Number of items sold  
销售数量
- ▶ Number of visits  
访问量
- ▶ Rate of conversion  
转化率
- ▶ Customer acquisition  
客户获得
- ▶ Customer retention  
客户保持
- ▶ Customer satisfaction  
客户满意度

15

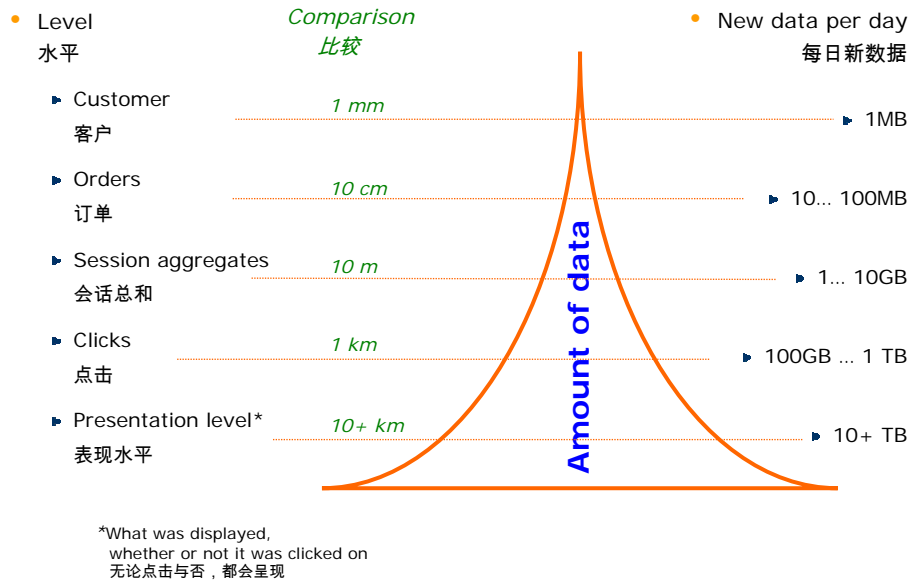
© 2004 by Weigend Associates LLC | [www.weigend.com](http://www.weigend.com)



- ▶ Orders 订货
  - ▶ Overall use of the site 网站的综合利用
    - Buying vs selling 买者vs卖者
    - Searching vs browsing 搜索vs浏览
    - Writing reviews, lists, etc.
  - ▶ Customer service contacts 客户服务联系
    - E-mail, phone 电子邮件, 电话
  - ▶ Surveys 调研
    - Intentions / Goals / Modalities 意图/目标/形式
    - Satisfaction 满意度
  - ▶ Customer service response 客户服务回复
    - Resolution 结果 (Free replacement, Refund 免费重置, 退款)
  - ▶ Delivery date vs promised date 交货日期vs允诺日期
  - ▶ Page generation time 页面产生时间
  - ▶ Search response 搜索回复
    - Number of search results 搜索结果数量
  - ▶ E-mail campaigns and responses 电子邮件广告和回复
- © 2004 by Weigend Associates LLC | www.weigend.com

16

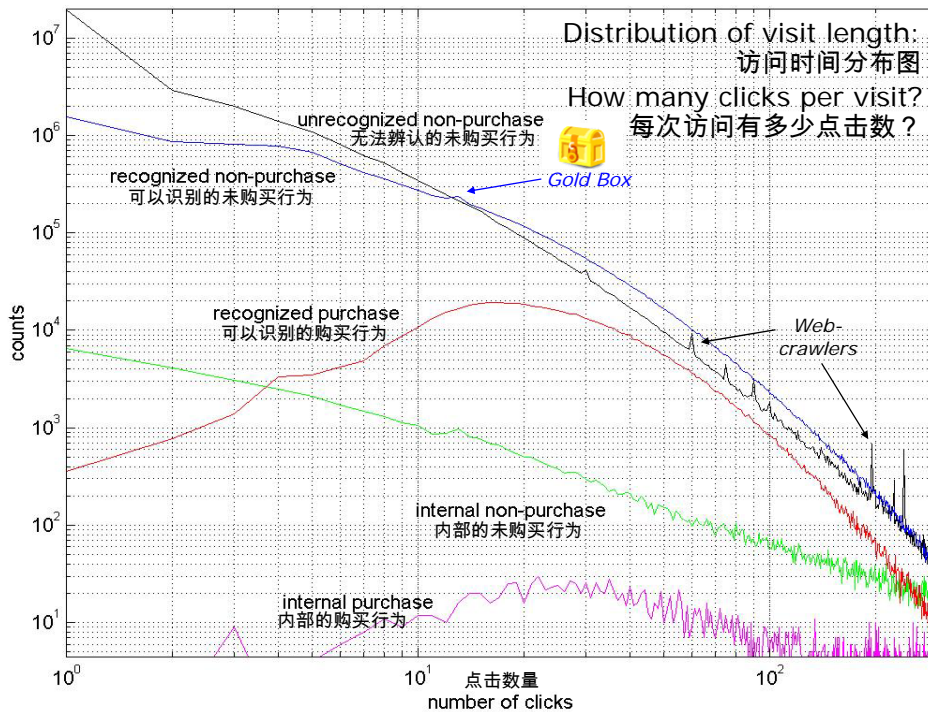
Amount of Data Created Per Day 每天产生的数据量



17

© 2004 by Weigend Associates LLC | www.weigend.com





#### 4. Building and Evaluating Predictive Models 预测模型的建立与评估

- Tasks: Predict, e.g.,  
目标：预测
  - ▶ Probability (buy in this visit *without discount*) vs Prob (buy in this visit *with discount*)  
本次访问用优惠券购物与不用优惠券购物
  - ▶ Probability (current page is last page requested in this visit)  
可能性：该页是本次访问的最后一页。
- Use models from different model classes (different statistical assumptions)  
利用不同类别的模型（不同的统计假设）
  - ▶ Baseline, e.g., Poisson (independent, unconditional)  
基线，例如，泊松（独立的，无条件的）
  - ▶ First order Markov  
一阶马尔可夫过程
  - ▶ Beginning-of-visit information  
访问开始的相关信息
    - HTTP-referrer HTTP转发
    - Search vs Browse 搜索与浏览
  - ▶ Aggregate visit so far (but time ordering ignored)  
访问集合(但是忽略定期订购)

## Building More Complex Probabilistic Models 创建更复杂的概率模型

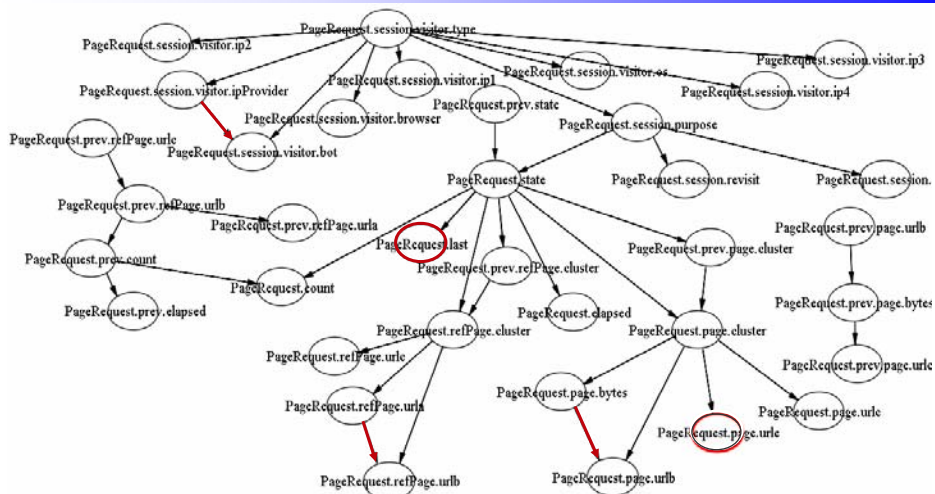
Joint work with Bruce D'Ambrosio, CleverSet Inc.  
与Bruce D'Ambrosio, CleverSet Inc合作

- Add synthetic variables  
增加综合变量
  - ▶ Combine observed variables (automatically generated)  
与观察变量结合(自动产生)
- Add hidden variables  
增加隐藏变量
  - ▶ Unobserved / hidden states  
不可观察/隐藏状态
- Add relational structure  
增加相关结构
  - ▶ E.g., use information from the products table, rather than only product ID  
如产品表而非仅仅产品标识
- Evaluate out-of-sample accuracy  
评估脱离例子分析的准确性
  - ▶ Standard: Area under ROC curve  
ROC曲线下面的区域

21

© 2004 by Weigend Associates LLC | www.weigend.com

### Extended Relational Structure 扩展的关系结构



#### Model

First order Markov 一阶马尔可夫过程  
Hidden Markov 不可见马尔可夫过程  
Basic relational 基础关系  
Extended relational 扩展关系

#### Performance (Area under ROC curve)

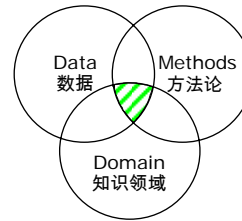
.334  
.513  
.728  
.777

22

© 2004 by Weigend Associates LLC | www.weigend.com

## Outline 提纲

- Making decisions based on A-B experiments  
以A-B实验为基础进行决策
- The need for a scientific framework  
我们需要一个科学的框架



- The iterative process of modeling  
反复的建模过程



- Some insights into online customer behavior  
一些对在线客户行为的理解
  - Levels of analysis and actionability
    - Personalization vs Occasionalization
  - Behavioral economics

23

© 2004 by Weigend Associates LLC | [www.weigend.com](http://www.weigend.com)

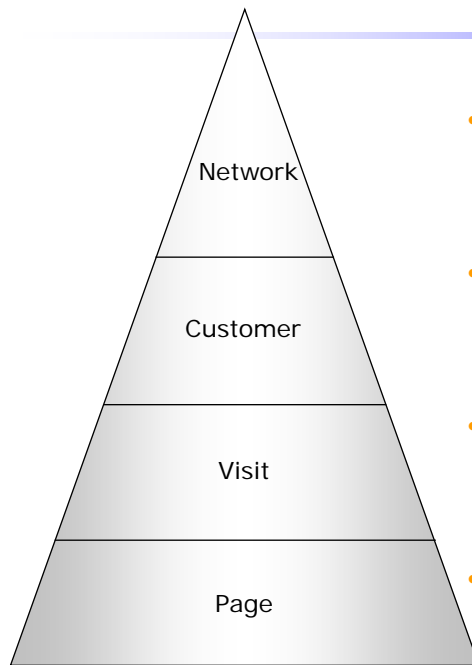
## Insights 理解

- What works best for recommendations, and why?  
什么使推荐更有效，为什么？
  - Key distinction: individual-history free (current click only, same for everybody), vs history-dependent  
主要区别：个体历史无关 vs 历史相关
- Spectrum of information, generating different recommendations and actions  
不同类别的信息产生不同的建议和行动
  - **Demographics** (traditional marketing)  
人口统计 (传统市场营销)
  - **Behavior** (traditional collaborative filtering, customer signatures)  
行为 (传统的协作过滤, 客户签名)
  - **Current intentions and modalities** (based only on information of current visit)  
目前的动机和形态 (仅基于本次访问的信息)

24

© 2004 by Weigend Associates LLC | [www.weigend.com](http://www.weigend.com)

## Levels of Analysis and Actionability



25

- D. Network of customers
  - ▶ Model: Apply social network research
  - ▶ Action: Discounts, better service
- C. Customer
  - ▶ Model: Demographics, behavior
  - ▶ Action: Personalization
- B. Visit
  - ▶ Model: Predict intention, situation, mode
  - ▶ Action: Occasionalization
- A. Page
  - ▶ Model: Content
  - ▶ Action: Show ads

© 2004 by Weigend Associates LLC | www.weigend.com

## Using Social Networks for Customer Acquisition 利用社会网络获得客户

- Social Networks 社会网络
  - ▶ Research 调查
    - STL proximity vs geographic/ZIP proximity (Mark Handcock)
    - Customer Lifetime Value: Intrinsic + Network (Pedro Domingos)
    - 客户终身价值：内质 + 网络
  - ▶ Applications 应用
    - Customer acquisition 客户的获得
    - Compute Customer Network Value 网上客户网络价值

- Data: "Share-the-Love"

- ▶ From the Website: 从网站上

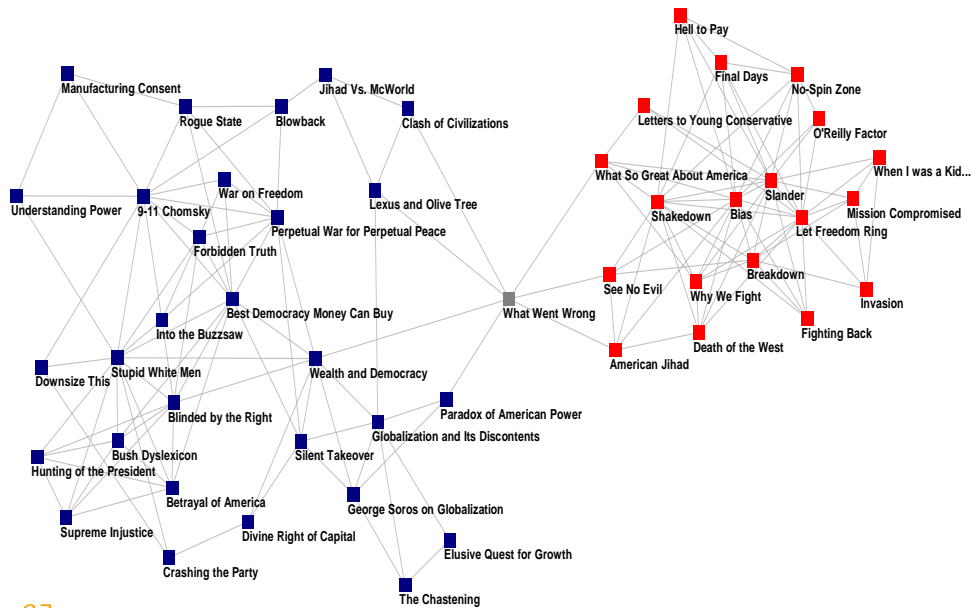


- "Each time you place an order for books, music, DVDs, or videos with us, we'll offer you the chance to e-mail your friends and give them an additional 10% off the items you bought. (You select which items, of course.)" 每次你订购一本书, 音乐, DVD, 或者录像, 我们将给你一个机会, 你可以用电子邮件推荐一件商品给你的朋友, 我们将给你朋友10%的优惠。
- "If any of those people purchases one of those items within a week, you'll receive a credit to use the next time you shop with us!" 如果以上任何人在一周之内购买以上任何一个产品, 你将得到积分下次购物时可以使用。
- "Your credit will equal the dollar amount of your friend's 10% discount." 这个积分与你朋友10%的折扣价值同等。

26

© 2004 by Weigend Associates LLC | www.weigend.com

## Social Network of Books based on *Customers who Bought Also Bought* (from Valdis Krebs)



27

© 2004 by Weigend Associates LLC | www.weigend.com

## Behavioral Economics 行为经济学

- ▶ Measure systematic deviations (heuristics, bounded rationality) in actual behavior 在实际活动中发现并使用系统偏差 (启发式的、理性约束的)
  - = 2<sup>nd</sup> order / descriptive, empirical 描述性、经验性的是第二阶的
  - vs traditional, neo-classical economics = 1<sup>st</sup> order / normative, theoretical 传统的古典经济学标准化、理论的为第一阶次
- ▶ Hypotheses 假设
  - Choice set (Itamar Simonson) 选择设定
  - Sweepstakes (Dan Ariely) 赌金

	Choice: 选择:	Limited 有限的	Extended 扩展的
Number of jams 果酱种类		6	24
Customers stopping by 停下来挑选的客户		40%	60%
Conversion (buy rate) 转变成买者的比率		30%	3%

### ← Jam experiment 果酱试验

When Choice is Demotivating: Can One Desire Too Much of a Good Thing?  
Sheena S. Iyengar and Mark Lepper.  
*Journal of Personality and Social Psychology* 79 (2000) 995.  
当没的选择的时候,人们还能够渴望获得太多的好东西吗?(希娜和马克,人物杂志和社会心理学)

28

© 2004 by Weigend Associates LLC | www.weigend.com

- Contact: **Dr. Andreas S. Weigend**
- Phone: **+1 (917) 697-3800**
- Web: **[www.weigend.com](http://www.weigend.com)**

© 2004 by Weigend Associates LLC | [www.weigend.com](http://www.weigend.com)