

# The Science behind on-line Advertising



Rushi Bhatt

Yahoo! Labs, Bangalore

December 13, 2011



# Agenda

---

- Introduce Web monetization
  - Search and Display Advertisement
- Auctions
  - Participation in Auctions
  - Pricing
- Exchanges
  - Ranking, response prediction
- “Next gen” monetization
  - Social Targeting
  - Chunked-rewards



# A Free Web!

---

- The WWW comes free for users
  - Browse, search, communicate, chat, socialize, all free!
- Web is funded by advertisers!
  - To a large extent (Yahoo!, Google, Facebook, Bing)

How does online advertising work ?



# Paid Search

Hi, **sgarg82** | Sign Out | Help

Bucket: no bucket ▾ bad results / ads or bugs? tell us! [hide]

Make Yahoo! your homepage | Mail

Web Images Video Local Shopping News More ▾

# YAHOO!

kolkata accommodation

Search

Options ▾

Search Pad

SearchScan - On

195,000 results for  
kolkata accommodatio....:

Show All

AsiaRooms

Sponsor Results

## [kolkata accommodation](#)

Sponsored Results

Up to 40% Off **Kolkata** Accommodation. Book Now - Limited Offer.

[Travelguru.com/kolkata](http://Travelguru.com/kolkata)

## [Hotels in Kolkata, India - Travel Results](#)

Price

[\\$0 - \\$50](#) (10)

[\\$50 - \\$100](#) (4)

[\\$100 - \\$150](#) (6)

[\\$150 - \\$250](#) (6)

Hotel Class

[5 Stars](#) (6)

[4 Stars+](#) (11)

[3 Stars+](#) (20)

[2 Stars+](#) (22)

Most Popular

[Tai Bengal](#) - from \$314

[ITC Sonar, Kolkata, A...](#) - from \$347

[The Best Inn](#) - from \$62

[The Park](#) - from \$275

## [Kolkata Accommodation - Kolkata Hotels, Apartments, Motels ...](#)

**Kolkata accommodation** & last minute hotels in India. Compare cheap last minute

**Kolkata holiday accommodation** - resorts, villas, apartments & holiday parks - Discount 2

...

[www.hotelsaccommodation.com.au/Kolkata-India](http://www.hotelsaccommodation.com.au/Kolkata-India) - [Cached](#)

## [Kolkata Hotels. Kolkata \(Calcutta\) Accommodation Hotel in ...](#)

**Kolkata Hotels. Kolkata (Calcutta) Accommodation** Hotel in India with descriptions, photos, maps and discounted rates for online reservation

[www.holidaycity.com/calcuttahotels](http://www.holidaycity.com/calcuttahotels) - [Cached](#)

## [kolkata accommodation : Instant Confirmation: Cheap Hotel in ...](#)

**kolkata accommodation**, Receive Instant Online Confirmation. Huge Range of Cheap **Kolkata** Hotel at Discount Rates. Family Hotel in **Kolkata**, Hotels, India .

[www.kolkataaccommodation.com](http://www.kolkataaccommodation.com) - [Cached](#)

## [Kolkata Accommodation, Accommodation in Kolkata, Hotels in ...](#)

AsiaRooms.com offers online information about **Kolkata Accommodation**.

[www.asiarooms.com/.../kolkata-accommodation.html](http://www.asiarooms.com/.../kolkata-accommodation.html) - [Cached](#)

Sponsored Results

## [Service Apartments Bangalore](#)

Fully furnished and luxury service apartments at affordable price.

[www.fountainresidency.com](http://www.fountainresidency.com)

## [Budget Hotels in Kolkata](#)

Compare Rates, Read Reviews & Book Your Hotel. Low Price Guaranteed.

[www.MakeMyTrip.com/hotels](http://www.MakeMyTrip.com/hotels)

## [Marry Bengalis Abroad](#)

Looking For Bengali Life Partner? Search In 200,000+ Bengali Profiles

[Jeevansathi.com/bengalinri](http://Jeevansathi.com/bengalinri)

## [Join Club Mahindra Today](#)

7 Days Of Holidays, Every Year For Next 25 Yrs. Sign Up Now & Enjoy.

[www.clubmahindra.com](http://www.clubmahindra.com)

## [The Oberoi Grand, Kolkata](#)

Early check-in & Late check-out, designed for business travellers.

[www.oberoihotels.com](http://www.oberoihotels.com)

## [Furnished Apartments](#)



# Contextual Ads

The screenshot shows the Yahoo! Finance homepage with various sections:

- Market Summary:** A table showing market indices like Dow, S&P 500, and Nasdaq with their current values and percentage changes.
- Top Stories:** A section with headlines such as "Stocks waver as Ireland debt fears remain a focus" and "3 Mistakes You May Be Making in Your 401(k) Plan".
- Market Data:** A table of stock quotes for YHOO, CSCO, GOOG, and MSFT, including price, change, and percentage change.
- Community Sentiment:** A section with "Bullish" and "Bearish" sentiment indicators for various stocks like AAPL and MSFT.
- Sponsored Links:** A large section in the middle of the page containing several advertisements for financial services like GEinterestplus.com, Forex.com, RothConverter.com, MadPennyStocks.com, and AmericaRebuilds.com.
- Analysis and News:** A section titled "AND ANALYSIS" with articles like "GM Rising: Who Should Get the Credit?" and "Foreclosure Renewal: A New Housing Mess?".

## SPONSORED LINKS

**1.25-1.45% Apply Online!**  
 With AA+ Rated GE Capital Corp. Not An Offer Of Securities For Sale.  
[GEinterestplus.com](http://GEinterestplus.com)

**Foreign Exchange Trading**  
 Free \$50,000 Practice Account With Real-Time Charts, News & Research.  
[www.Forex.com](http://www.Forex.com)

**Crown Capital Securities**  
 Professional Wealth Management  
<http://RothConverter.com>

**3 Stocks Set to Surge**  
 Get stock picks before they explode, picks that move 100% in a day!  
[www.MadPennyStocks.com](http://www.MadPennyStocks.com)

**Free Stock Market Quote**  
 Find Stocks That Have Doubled! Start With Motley Fool's Free Report.  
[www.fool.com](http://www.fool.com)

**America Rebuilds**  
 Find inspiration, advice & tools to help reach your financial dreams.  
[www.AmericaRebuilds.com](http://www.AmericaRebuilds.com)





# Display Ad: Impression

HOME U.S. BUSINESS WORLD ENTERTAINMENT SPORTS **TECH** POLITICS SCIENCE HEALTH OPINION MOST POPULAR

Tech Video Blog Internet Gadgets Cell Phones Apple/Macintosh Social Media Video Games Security

Q Search All News **News Search** TRENDING NOW: rima fakih sandra diaz-twine ronnie james dio survivor miss usa

## Technology



### Study on cell phone link to cancer inconclusive

AP – 2 hrs 11 mins ago  
GENEVA – If there's one lifestyle tool that's ubiquitous, from American cities to remote villages of the developing world, it's the mobile phone. [Full Story »](#)

### Heart group backs video games in obesity campaign

AP – Mon May 17, 1:03 pm ET

### Supreme Court rejects appeal of "must-carry" rule

AP – Mon May 17, 10:57 am ET

## Most Popular - Technology

Most Emailed Most Viewed Most Recommended

- Death of 2 boys prompts toy dart gun set recall
- Japan's Astellas to buy US drug co. OSI for \$4B

[More Most Emailed - Technology »](#)

## Technology Slideshows

[More Technology Slideshows »](#)



Technology Photos



Apple iPad



Apple Inc.



Microsoft

ADVERTISEMENT

LEGAL

**A SMART PHONE THAT SEATS SEVEN**

DODGE GRAND CARAVAN WITH

## Featured





# Display Ad: Click

[HOME](#)[INVESTING](#)[NEWS & OPINION](#)[PERSONAL FINANCE](#)[MY PORTFOLIOS](#)[TECH TICKER](#)

streaming quotes:ON ?

[GET QUOTES](#)

Finance Search

Tue, May 18, 2010, 3:43AM EDT - U.S. Markets closed.



## Icahn boosts stake in Genzyme, Motorola; sells CIT

Icahn boosts stake in Genzyme, Motorola; sells CIT Group, Blockbuster, Yahoo shares

**AP** Associated Press

Buzz up! 0

Print

Companies: [Blockbuster Inc.](#) | [Chesapeake Energy Corporation](#) | [CIT Group, Inc.](#)

### Related Quotes

Symbol	Price	Change
BBI	0.4000	0.0000



CHK	22.29	0.00
CIT	37.01	0.00
GENZ	50.87	0.00
LGF	6.54	0.00

On Monday May 17, 2010, 7:44 pm EDT

LOS ANGELES (AP) -- Billionaire investor Carl Icahn boosted his stakes in Genzyme Corp. and Motorola Inc. in the first quarter but dumped shares in CIT Group Inc., Yahoo Inc. and Blockbuster Inc.

His main investment vehicle grew in value by 11 percent to \$3.18 billion during the first three months of the year.

Shares held by Icahn Capital LP were disclosed in documents filed with the Securities and Exchange Commission on Monday. The documents offer a snapshot into the activist investor's portfolio as of March 31. The entire portfolio was worth \$2.87 billion on Dec. 31.

Icahn bought 2.1 million shares in new investment Chesapeake Energy Corp. worth \$50.2 million during the quarter and boosted his stake in video game maker Take-Two Interactive Software Inc. to 9.3 million shares from 6.7 million at the end of December. His

stake is now worth \$91.6 million.

### Top Stories

- Forum strives for better pilots, controllers - AP
- Obama administration faces questions on oil spill - AP
- Asia stocks gain, growth hopes offset Europe fears - AP
- Eurozone nations defend currency union - AP

ADVERTISEMENT

IAMS® ProActive Health Formulas with PreBiotics. Now in dry and wet.

**IAMS** Rollover to Learn How PreBiotics Supports Strong Defenses ▶



# Display Ad: Conversion

Yahoo! | My Yahoo | Mail | Movies | Finance

YAHOO! NEWS INDIA



Home National World Business General Politics Features Crime Photos Search:

## A minister had asked for

Mon, Nov 15 04:25 PM

Dehradun: When asked to reveal his methodology in compromising with ethics and values, Tata said he did not have a methodology in the to tell the audience.

"Several years ago, Tatas were trying to collaboration with Singapore Airlines. Even in the airline industry, we had enormous problem reports in the media.

"We approached three Prime Ministers about our efforts to form the airlines." he said.

Later, Tata said a fellow industrialist commented, "You are stupid people. The Minister was asking for Rs 15 crore. Why didn't he pay?"

**DHL EXPRESS**

**Win an iPod\* when you start shipping with us!**

Sign up for DHL account before 31st Dec 2010.  
Complete the slogan & stand a chance to win an iPod\*

**SIGN UP NOW!**

Last Name \*

First Name \*

Email \*

Phone \*

Company \*

DHL Express is my preferred service partner because..... \*

Yes, it's ok to send me email

**GET STARTED**

### No one knows express shipping like we do

Spend less time shipping and more time growing your business, worldwide. As a new DHL customer, here are the key benefits that you can look forward to.

#### Fast and reliable

We are faster and more reliable than ever. Reach over 220 countries and territories at affordable rates.

#### Customs expertise

Relax. DHL collects, clears through customs and delivers millions of shipments every day.

#### Late pick-up times

Get the latest possible pick-up so you have more time and can respond to last minute requests.

#### Online tracking

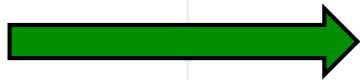
Enjoy full visibility with real-time check point and delivery details online.

TRY THE NEW NEWS EXPERIENCE





# Display Ad: Publisher targeting



Hi, Sachin | Sign Out | Help | Trending: Jennifer Grey | Yahoo! | Mail | AdChoices

**YAHOO! FINANCE** | Search | Web Search

**A PRICING STRUCTURE ONLY YOUR BROKER COULD HATE.**

**\$4.95 PER TRADE** | **65¢ PER OPTION CONTRACT**

**TRADEKING**

SEE FOR YOURSELF. SWITCH TO TRADEKING TODAY. MEMBER FINRA, SIPC & SIPC

HOME | **INVESTING** | NEWS & OPINION | PERSONAL FINANCE | MY PORTFOLIOS | TECH TICKER


GET QUOTES | Finance Search | Wed, Nov 17, 2010, 7:38AM EST - US Markets open in 1 hr and 51 mins

Scotttrade: \$7 Online Stock Trades & Powerful Trading Tools

### Today's Markets

Market Summary [Edit]

streaming quotes: ON



SYMBOL	LAST	CHANGE
Dow	11,023.50	↓ 178.47 (1.59%)
Nasdaq	2,469.84	0.00 (0.00%)
S&P 500	1,178.34	0.00 (0.00%)
10-Yr Bond	2.8470%	0.00
NYSE Volume	0	
Nasdaq Volume	0	

Indices: US - World | Most Actives

#### Advances & Declines

	NYSE	NASDAQ
Advances	574 (15%)	583 (20%)
Declines	3,226 (82%)	2,097 (78%)
Unchanged	132 (3%)	112 (4%)
Up Vol*	442 (447%)	208 (11%)
Down Vol*	3,938 (3977%)	1,617 (88%)

#### Market Coverage

Market Overview, Market Update, In Play, Story Stocks, Short Stories, Tech Stocks

#### Today's Events

Earnings, Conf. Calls, Economic, IPOs, Splits, Up/Downgrades

#### Premium Services

Real-time Quotes, Research Reports

#### Market Update

06:41 am : S&P futures vs fair value: +2.10. Nasdaq futures vs fair value: +6.80.

06:41 am : Nikkei...9811.66...+14.60...+0.20%. Hang Seng...23214.46...-478.60...-2.00%.

06:41 am :

#### Market Statistics

Mkt Digest, Most Actives, U.S. Indices, World Indices, Exchange Rates, Unusual Volume, Commodities

#### Financial News


Top Stories, U.S Markets, Most Viewed Articles, Full Coverage, Real-time Quotes, Research Reports

#### Investing Tools

Stock Alerts, News Alerts

MY YAHOO! | Set Alert

ADVERTISEMENT



**8 investing mistakes you should avoid in 2010**

If you have a \$500,000 portfolio, download the guide by Forbes columnist and money manager Ken Fisher. It's called "The Eight Biggest Mistakes Investors Make and How to Avoid Them." Even if you have something else in place right now, it still makes sense to request your guide!

Click here to download



# Display Ad: Location targeting

Hi, Sachin | Sign Out | Help Trending: Madagascar Coup Yahoo! Mail My

**YAHOO! SPORTS**

Home **NFL** MLB NBA NHL NCAAF NCAAB NASCAR Golf UFC Boxing Soccer Tennis Action Sports More Shop Fantasy

TRENDING NOW: [Tayshaun Prince](#) [Michael Vick](#) [Roy Halladay](#) [John Wall](#) [Cam Newton](#)

**NBA** [ROSE'S 17 IN 4TH LIFT BULLS BY ROCKETS](#) [NHL](#) [MLB](#) [NFL](#)

## Cut to the Chase

There have been penalties, fights and fingers, and now NASCAR's 2010 season comes down to three drivers and one race. [How it got here](#)

[Predictions](#) | [Hamlin's title to lose](#) | [Dummies guide to the Chase](#)

### Getting burned

The Heat are struggling as they try to find their identity. [Miami's multiple issues](#)

### Shoes turn Arenas' tide

### Hoops extravaganza

Butler is among those in the 24-hour marathon who didn't fare well. [Winners/losers](#)

- [Chat with King](#) **LIVE!**

### The best of the best

Manny Pacquiao's ability to beat boxing's greats comes under debate. [Iole's mailbag](#)

- [Tyson entering food biz?](#)

### Commish's false advertising

Roger Goodell uses a flimsy reason to back a stadium in Atlanta. [Super Bowl tease](#)

- [Ocho to marry reality star](#)

#### Headlines

- [Halladay wins NL Cy Young | Unanimous](#)
- [McNabb deal gives 'Skins out clause | Debate](#)
- [FBI looks into Newton saga | Heisman issues](#)
- [Braves acquire Uggla from Marlins | Atlanta's gain](#)
- [Rangers GM Daniels meets with Lee](#)
- [Steelers cut kicker Reed | Shame Report](#)
- [Club shut down after So. Miss shooting](#)
- [No. 4 OSU rolls at No. 9 Florida | Winners/losers](#)

[More News](#) | [Calendar](#) | [My Sports News](#) | [Video](#)

**BEHIND ENEMY LINES** [\[WATCH NOW\]](#) presented by BUD LIGHT

**YAHOO! SPORTS EXPERTS**  
[Vick's reclamation story deserves MVP applause](#)  
Les Carpenter November 16, 2010

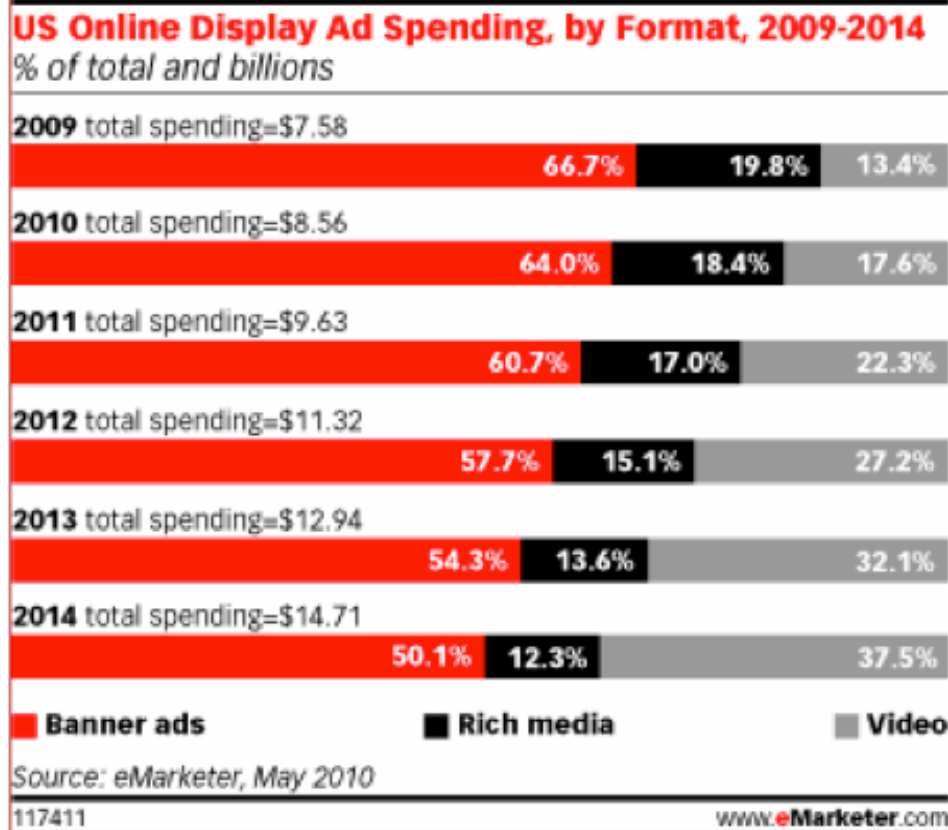
**VIDEO SPOTLIGHT**  
[Footwear change helps Arenas](#)  
Posted Nov 17 2010

**Club Mahindra HOLIDAYS** *Jijo Life*  
fun. family. forever.

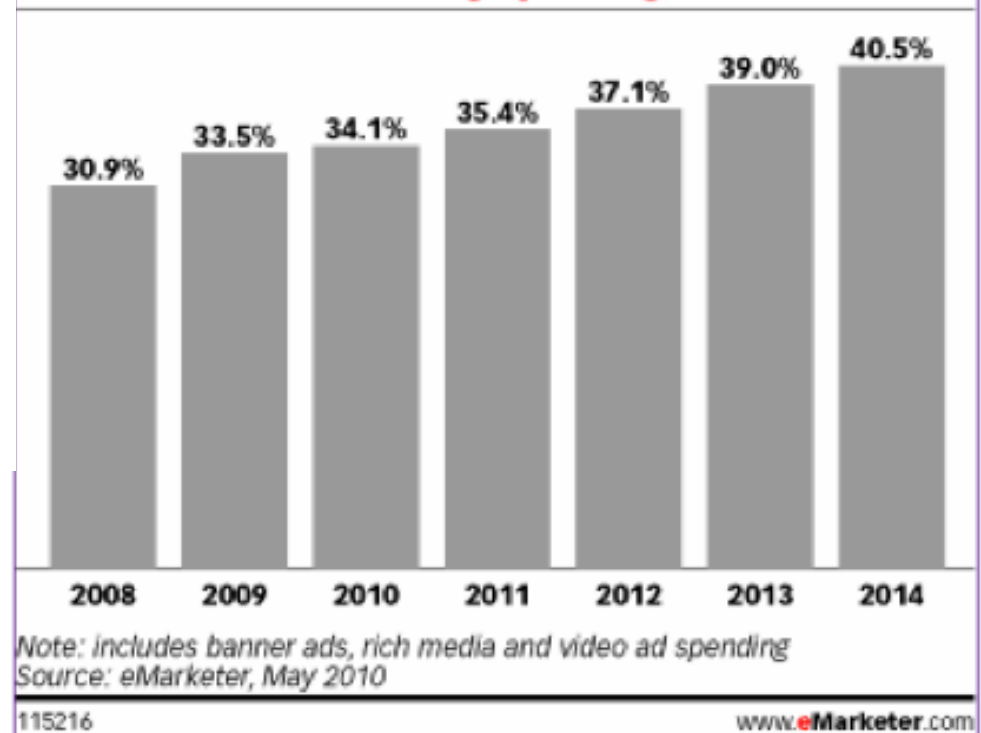
[LEARN MORE](#)



# Display Ads is big business



**US Online Display Advertising Spending as a Percent of Total Online Advertising Spending, 2008-2014**





# Massive scale

## US Online Display Advertising Metrics, 2001-2011

	Pages viewed per user per day	Total pages viewed (billions)	Impressions per page	Total impressions (billions)	CPM	Revenue per 1,000 pages	Total revenues (billions)
2001	35	1,980	0.41	812	\$6.35	\$2.60	\$5.16
2002	37	2,238	0.36	806	\$4.67	\$1.68	\$3.76
2003	38	2,484	0.34	844	\$3.65	\$1.24	\$3.08
2004	40	2,707	0.36	975	\$4.03	\$1.45	\$3.93
2005	42	3,024	0.37	1,119	\$4.25	\$1.57	\$4.76
2006	45	3,341	0.50	1,671	\$3.50	\$1.75	\$5.85
2007	47	3,608	0.60	2,165	\$3.31	\$1.99	\$7.17
2008	49	3,868	0.62	2,398	\$3.32	\$2.06	\$7.95
2009	51	4,120	0.61	2,492	\$3.39	\$2.05	\$8.45
2010	52	4,338	0.62	2,689	\$3.50	\$2.17	\$9.41
2011	54	4,563	0.63	2,875	\$3.62	\$2.28	\$10.41

*Note: includes static, rich media and video advertising*

*Source: JPMorgan and company reports, "Online Advertising Forecast," provided to eMarketer, November 3, 2008*

099290

www.eMarketer.com





# Quick Recap: Types of ads

---

- **Textual** (3 – 4 lines of text)
  - Paid Search (e.g.,
    - Appear on Search results page
    - Selected based on search keywords
    - Advertiser pays per click
  - Contextual Ads
    - Appear on web-pages
    - Keywords constitute content of the page
    - Advertiser pays per click
- **Graphical** (image, video, animation)
  - Appear on web-pages
  - Targeting criteria specified by the advertiser and optimized by the Ad-network
  - Advertiser chooses between paying per impression, per-click or per-conversion





# 4 Players in Display Ads

---

- **Advertiser:** Wants high RoI (e.g. cost per lead or conversion)
  - Demand: Buyer of “ad impressions” from publishers
  - Targets users (e.g. Males, from California) and pages (e.g. sports)
  - May require guaranteed impressions (e.g. 10 Million in 30 days)
  - Or competes in an on-line auction to win impressions
  - Might specify frequency caps (e.g. < 10/user/day)
  - Might specify budget caps (e.g. < \$200 daily)
  - Bidding Examples
    - Lipstick maker: \$2 CPM on 30-40 year old females, who visit fashion pages
    - Insurance seller: \$2 CPA on 40-50 year old males, who visit finance pages
- **Publisher:** Wants revenue to be maximized, but not at the cost of user dissatisfaction
  - Supply: Seller of “ad impressions” to advertisers
  - Specify types of ads that can be placed (e.g. only travel related ads)
  - Might specify desired payment type of the ads



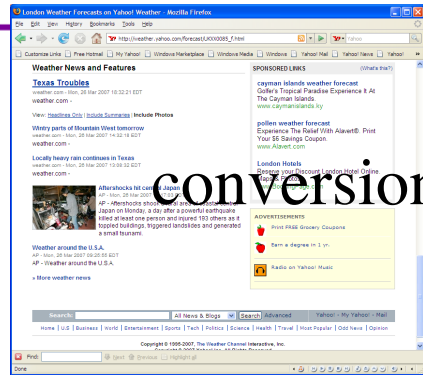
## 4 Players – Contd.

---

- **User:** Wants useful ads
  - Browses
  - Can specify interests / hobbies / likes
  - Views, clicks, converts, buys
- **Ad-Network / Exchange** (e.g. Y!, Google, MSN):  
Wants to maximize revenue
  - The “matchmaker”
  - Operates infrastructure for match-making
  - Takes a cut for each payment from advertiser to publisher
  - In many cases, dual role as publisher (like Yahoo!)



# How it all works ?



conversion

Response rates  
(click, conversion,  
ad-view)

Auction

Bids

Select  $\text{argmax } f(\text{bid}, \text{rate})$

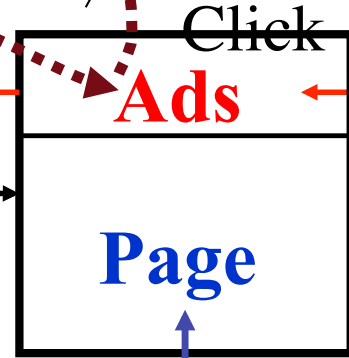
Pick  
best  
ads

Advertisers

Ad  
Network

Publisher

User





# Ad Selection: Simple example

---

- Advertiser 1: Bids \$2 CPM on 30-40 year, males, from Karnataka
- Advertiser 2: Bids \$30 CPC on 25-35 year old males from India
- Advertiser 3: Bids \$90 CPA on males
- Advertiser 4: Bids \$3 CPM on females from Karnataka
  
- A 30 year old male from Karnataka (user  $u$ ) browses (page  $p$ ), causing an opportunity
  
- Which ad will he end up viewing?
  - Match based on targeting (Ads 1, 2 and 3 are eligible)
  - Compute expected revenue from each
    - Ad 1 = \$0.002
    - Ad 2 = \$30 \*  $P(\text{Click}|u, p)$  = \$0.003, (if  $P(\text{Click}|u, p) = 0.001$ )
    - Ad 3 = \$90 \*  $P(\text{Conv}|u, p)$  = \$0.0009, (if  $P(\text{Conv}|u, p) = 0.0001$ )
  - Auction conducted on expected revenue: highest one wins
  - Ad 2 wins and gets shown to the user



# Ad Selection: Prediction

---

- True  $P(\textit{Click}|u, p)$ ,  $P(\textit{Conv}|u, p)$  are unknown
- Need to predict
  - Say  $\hat{P}(\textit{Click}|u, p)$  for Ad 2 = 0.0006 (true = 0.001), then Ad 1 wins  $\implies$  Loss of \$0.001
  - Say  $\hat{P}(\textit{Conv}|u, p)$  for Ad 3 = 0.0004 (true = 0.0001), then Ad 3 wins  $\implies$  loss of \$0.0021
- In-accurate prediction causes ‘wrong’ ad to be shown
- Accurate prediction is crucial to revenue
- Ranking problem versus absolute prediction





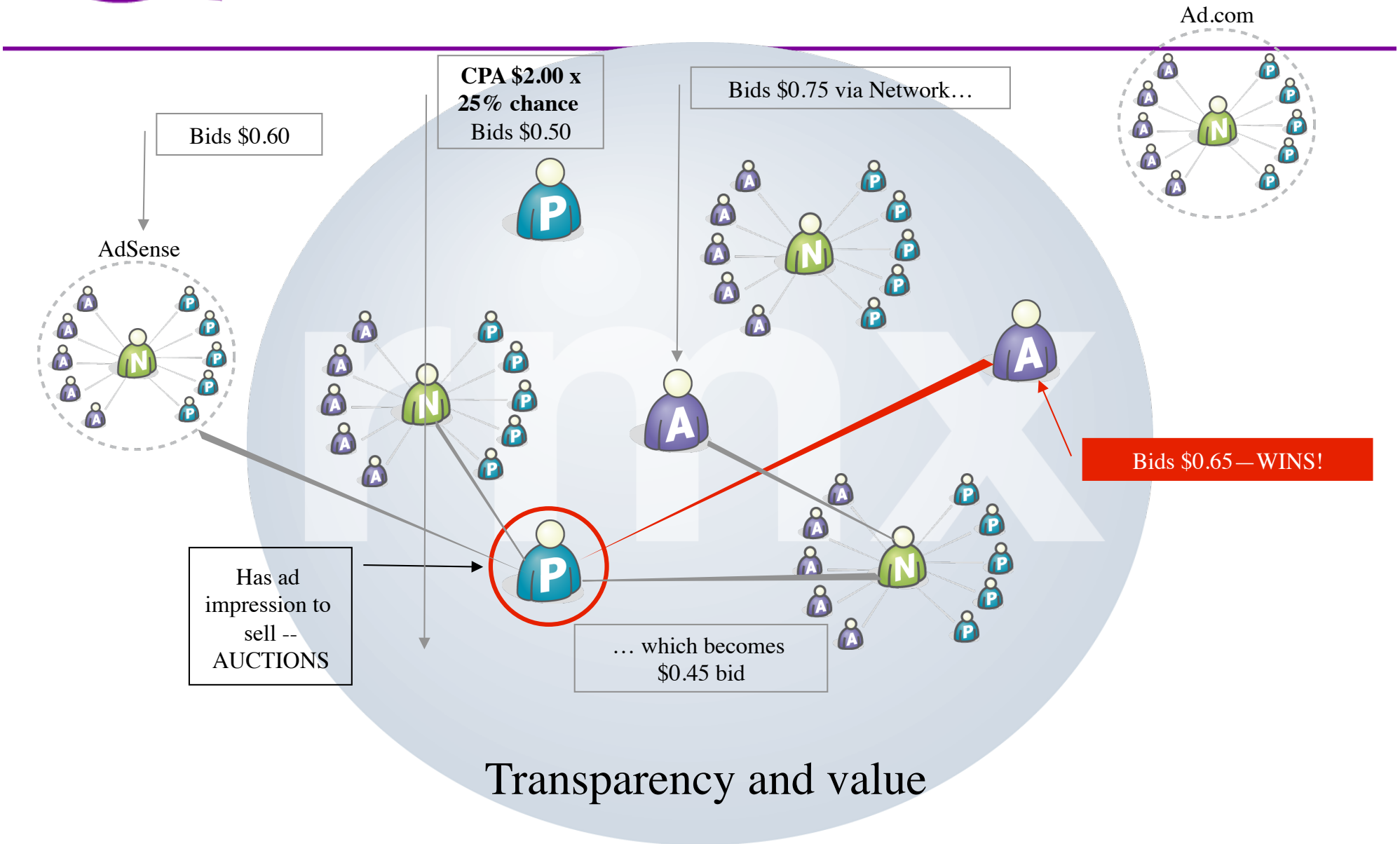
# Digression

---

- Determine eligible ads “Matching Problem”
  - Satisfy targeting criteria
  - Other constrains
    - Budget remaining
    - Frequency caps per user / day
- Auction rules
  - Allocation rule
  - Payment rule
- All this needs to happen in tens of milliseconds



# The Open Exchange





# A bit about data

---

- Users ~ tens of millions
- Pages ~ hundreds of millions
- Ads ~ hundreds of thousands
- Responses are of three types
  - Click, Post-click conversion, Post-view conversion
  - Each needs a separate model
- Billions of <user, page, ad> records per day along with response observed
- Privacy policy limits amount of historical data to be used in prediction
- Response rates are non-stationary
  - Trust recent history more



# Response Prediction Problem

---

- Notation:
  - Page:  $(i, X_p)$   $X_p$  = page features
  - Ad:  $(j, X_a)$   $X_a$  = Ad features
  - User:  $(k, X_u)$   $X_u$  = user features
  - Response: Tries =  $N_{ijk}$ , Success =  $S_{ijk}$
- Goal is to predict response rates for each cell  $P(j|i, k)$
- MLE:  $\hat{P}(j|i, k) = \frac{S_{ijk}}{N_{ijk}}$



# Challenges

---

- Sparsity:
  - Too many combinations, most cells have  $N_{ijk} = 0$ 
    - High dimensional categorical variables, e.g. In CLICK data, 100M cells
  - ‘cold-start’ problem
  - If not zero, most cells have small tries
- Rare response:
  - Response rates are extremely rare
  - 0 in 100 is not the same as 0 in 100,000
- Imbalanced sample size
  - $N_{ijk}$  in cells have huge variation
- Smoothing to perform small sample corrections is important
- How do we perform such corrections in a scalable way?





## Basic Ideas

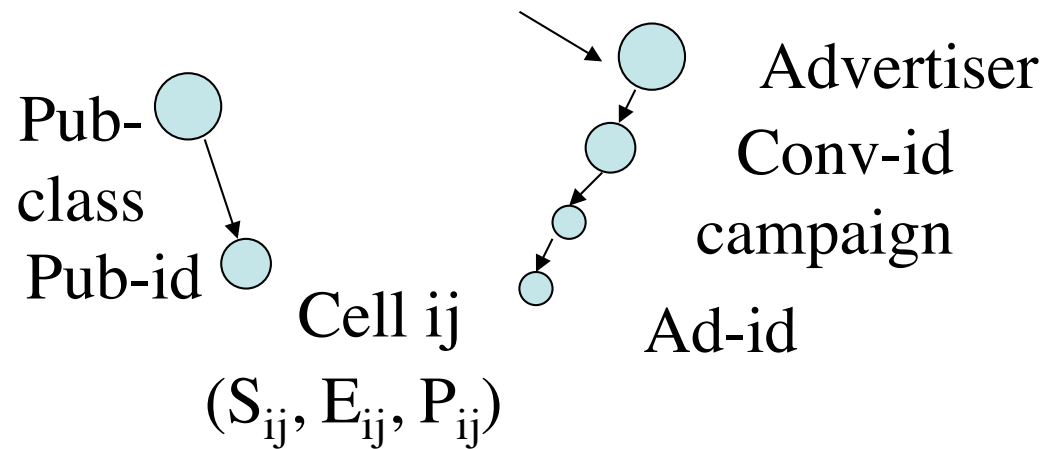
---

- When  $N_{ijk}$  is ‘sufficiently’ large, trust MLE
- For small  $N_{ijk}$ ,
  - Collapse cells based on features and predict from the aggregates
  - Use hierarchical information for aggregation and predict by “falling back”
  - Use smoothing
  - Other statistical corrections



# Hierarchical structure

- Assuming two hierarchies (Publisher and advertiser)



Cross-product of paths



# Three models

---

- Baseline
  - Logistic Regression
- Decision Tree
  - Learnt hierarchy based on predictive-ness of attributes, then smoothing / corrections
- LMMH [*D. Agarwal et. al., ACM SIGKDD 2010*]
  - Natural hierarchy of ads and pages
- Collaborative filtering with hierarchies



# Logistic Regression

---

$$\log \frac{p}{1-p} = \sum_{i,j,k} w X_{u,p,a}$$

- Use singleton features only
- Use conjunction features
  - Need hashing trick to reduce dimensionality [Weinberger et. Al, ICML 2009]
- Performs poorly: Approximating in the non-linear region



# Decision Tree

---

- Each record is labeled 'succ' or 'no-succ'
- Tree induction with gain ratio as the splitting criterion
- Shrinkage: Child's estimate is shrunk towards the parent
- Return from the parent for cold-start
- Runs on Grid (Map-Reduce), model refreshed periodically



# DT Example

### Martha Stewart Living

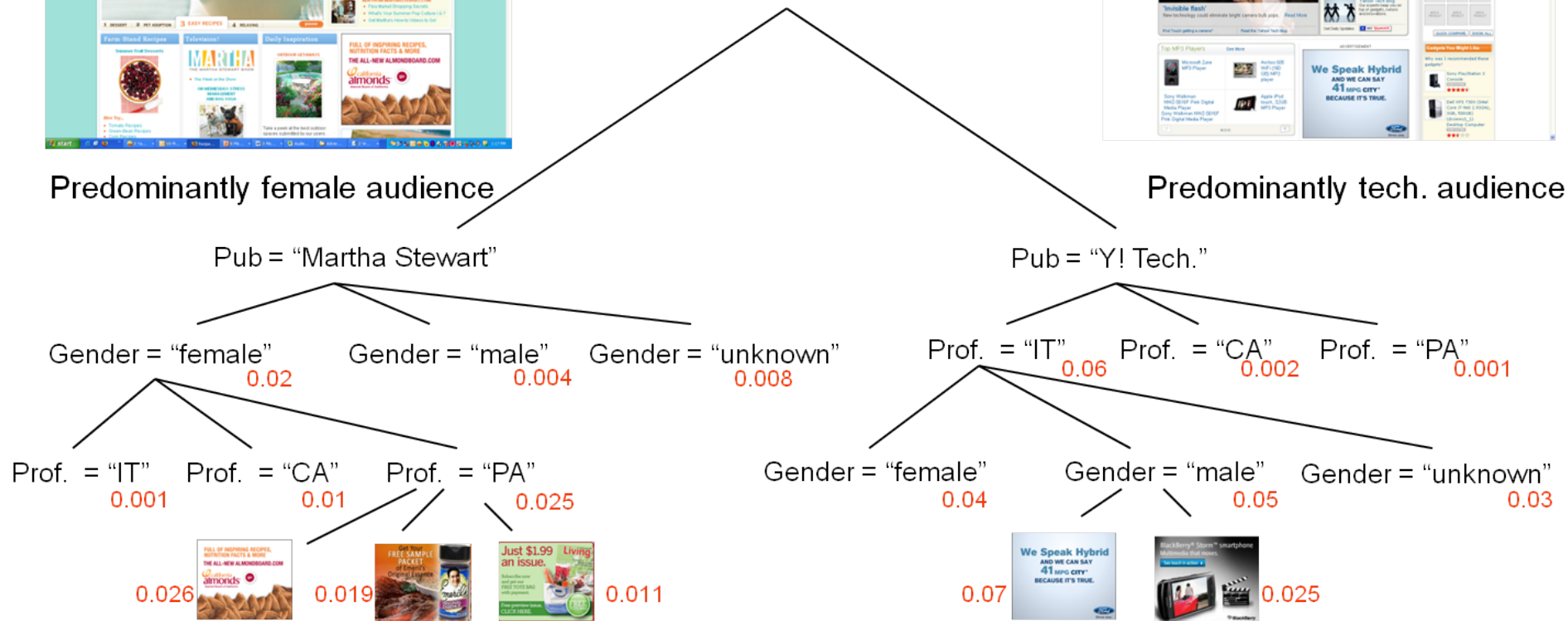


### Y! Tech.



Predominantly female audience

Predominantly tech. audience



DT - Learned hierarchy , tree induction on ad, publisher, user attributes



# LMMH (Agarwal et. Al: ACM SIGKDD 2010)

---

- 3-stage log-linear model
  - Stage 1: Feature based only, uses GLMM
    - e.g. pub category, creative category, daypart,...
  - Stage 2: Clustering data matrix elements
    - through multi-clustering : extension of SIGKDD 07 work
    - E.g. features – 1(pub clust =1, creative cluster=2)
  - Stage 3: corrections using multiple hierarchies
    - Sparse solutions through a new penalty
    - E.g Publisher x Adv hierarchy + daypart x Adv hierarchy





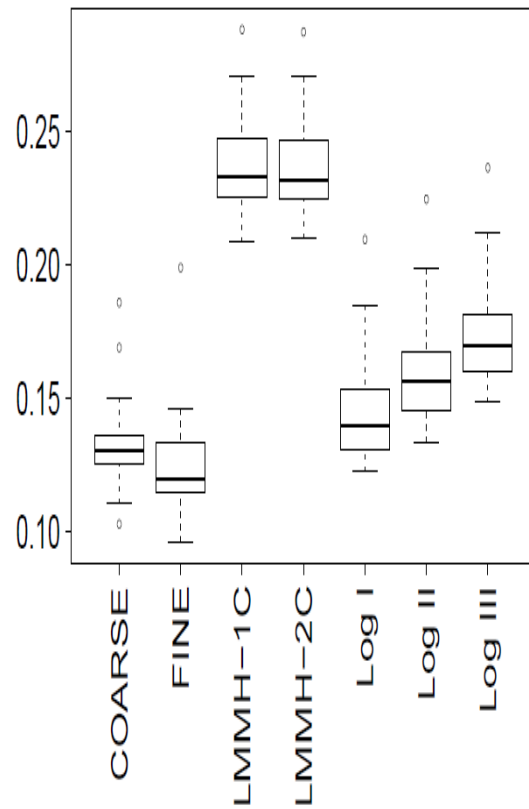
# Real world Data sets

---

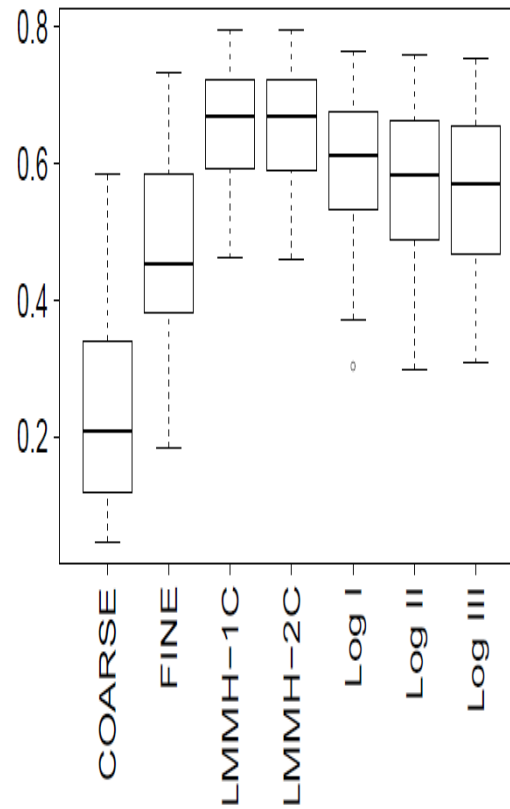
- CLICK [~90B training events]
- PCC (~.5B training events)
  - Conversion only through click
- PVC – Post-View conversions (~7B events)
  - Cookie gets augmented with pixel and triggers success
- Features
  - User, pub and ad features
  - 2 hierarchies (publisher and advertiser)



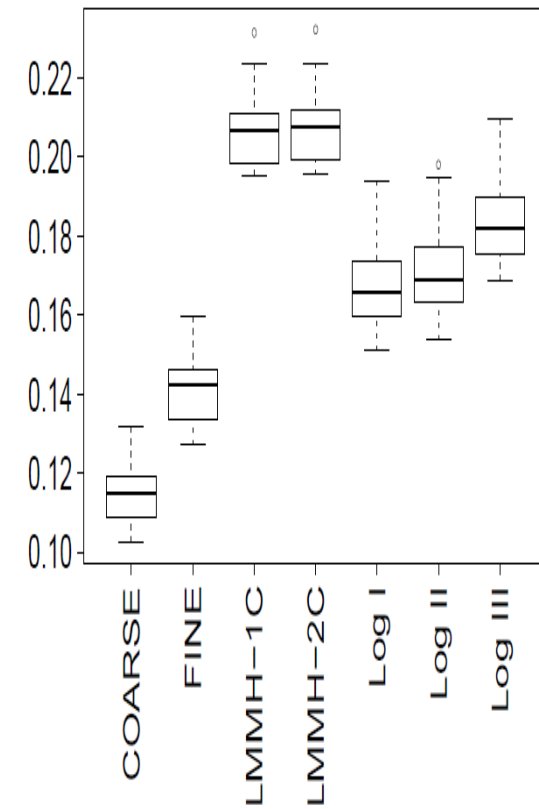
# Accuracy: Average test log-likelihood



(a) PCC



(b) PVC



(c) CLICK



# What if prediction still goes wrong?

---

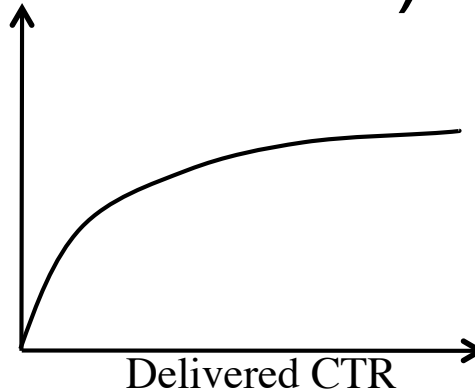
- Erroneous prediction can cause auction to degenerate
- Risk mitigation is needed
- Allocate a budget proportional to historical number of successes
  - Decrement every time the ad is shown
  - Stop when there is no budget left
  - Refresh budget periodically
- Principle: “Throttle” un-tested ads until they (slowly) prove themselves
- Note: Allocation rule no longer just (max eCPM) in auction



# Pricing models

---

- We talked about
  - CPM (advertiser bears the risk)
  - CPC, CPA (Publisher bears the risk)
- Any scheme which can balance this?
  - dCPM (dynamic CPM)





# Things we did not talk about

---

- Guaranteed delivery display!
- How do advertisers bid?
- Behavioral targeting
- Social targeting
- Layout optimization
- Social sharing of Ads
  - Do influencers affect product buying?  
{Bhatt et. Al – CIKM 2010}
- Many many others



# “Next Gen”: Social Targeting, Chunked Rewards

---



# Product Adoption in Large-Scale Social Networks

Rushi Bhatt, Vineet Chaoji, Rajesh Parekh



- What are the correlates of adoption spread?
  - Are high-degree users “different”?
  - Are neighborhoods predictive?
- How do we improve uptake?
  - Do social attributes provide a lift?
  - Is *neighborhood targeting* a good idea?



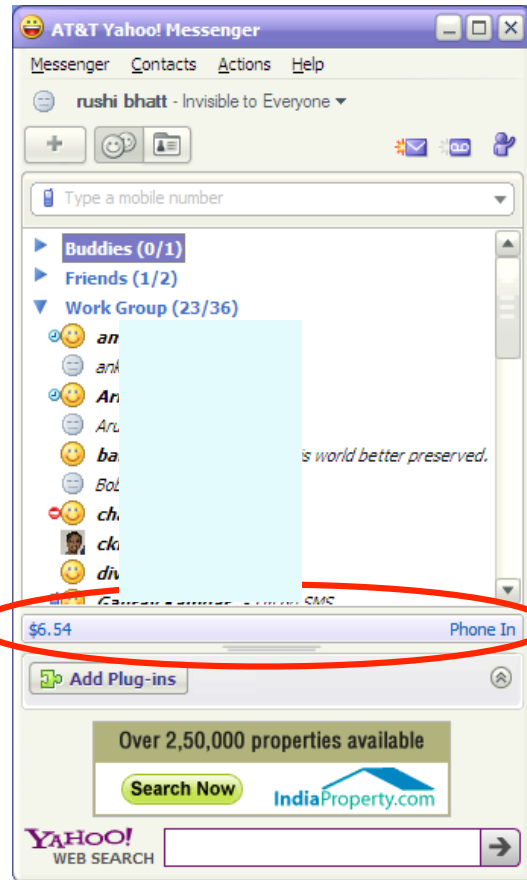
# Spread of Premium Service Adoption Through a Network

---

**Case Study:  
Adoption Spread in  
the PC to Phone Premium Service**



# Data Sources



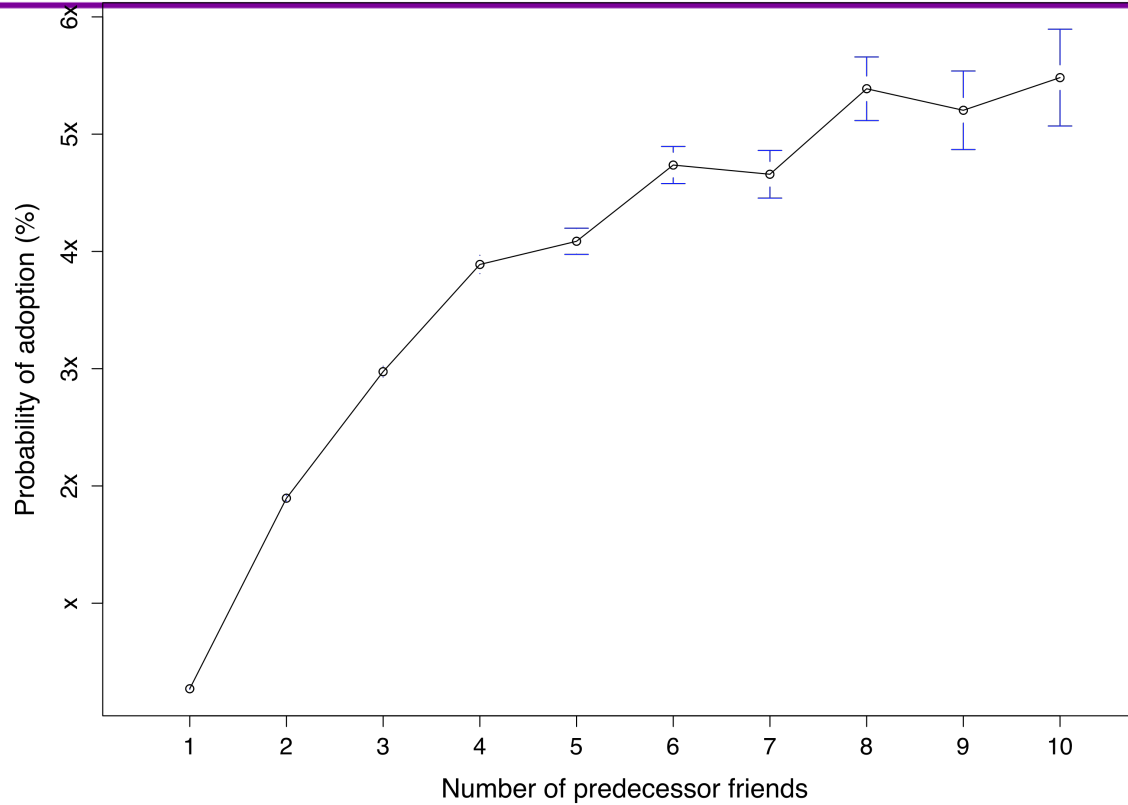
PC to Phone  
Premium Service

- Social Graph
  - IM users and their friendship network
  - $O(100M)$  nodes,  $O(1B)$  edges
- Behavior
  - Granular events: page views, searches, search result clicks, ad views, ad clicks
  - IM usage: messages sent, login days, ...
- Demographic
  - Gender, Age
- Geographic (from IP address)
  - Login country
  - Granular DMA level information
- Premium Service
  - PC 2 Phone subscription

**Low baseline adoption rates**



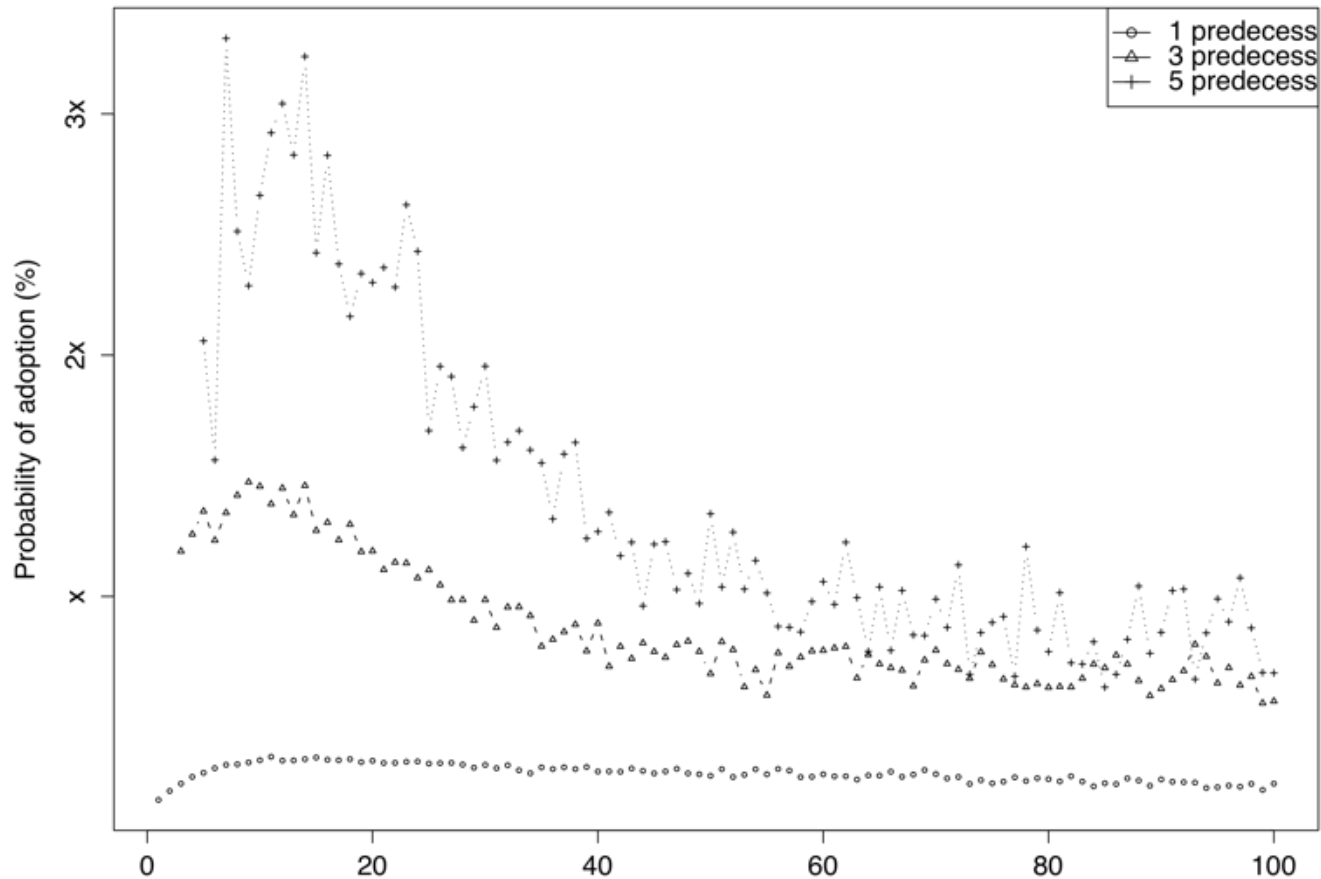
# Social Network Effect Exists



**More friends adopting improves chances of adoption**

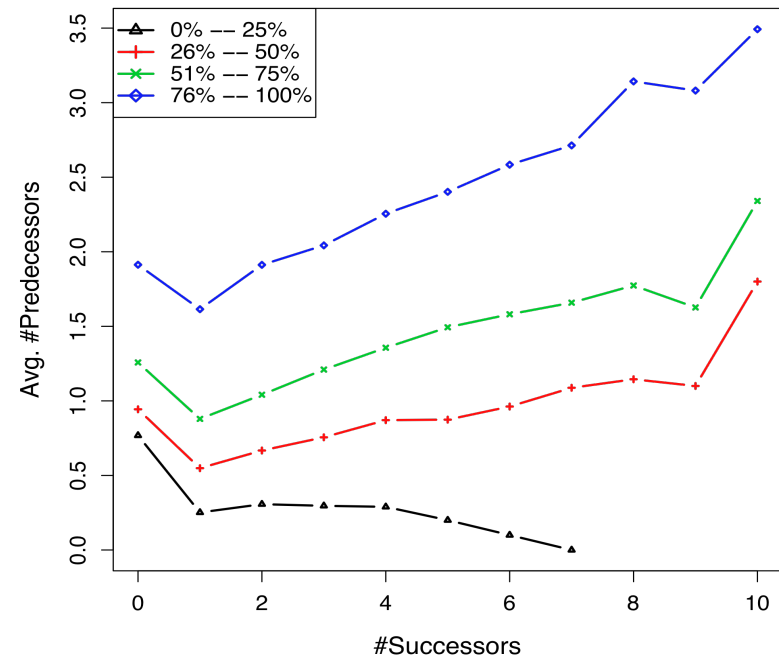
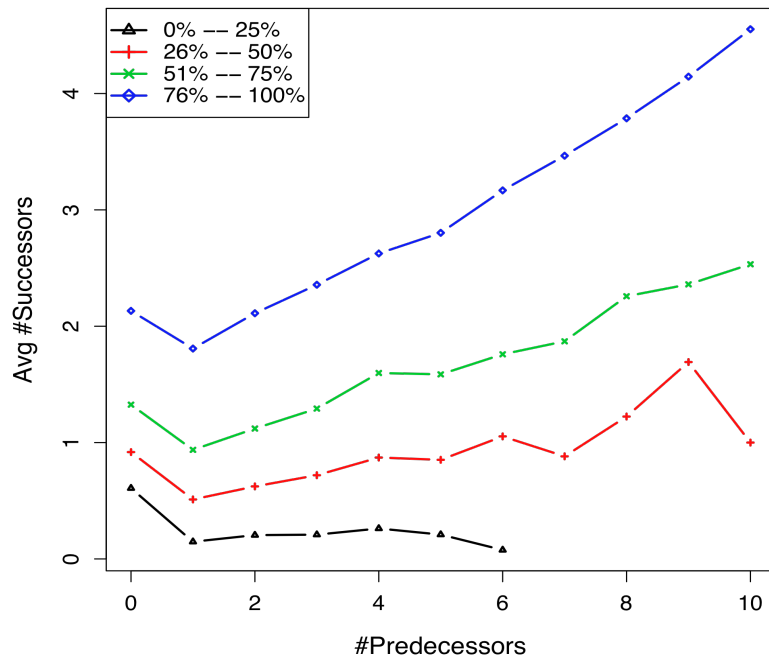


# High-degree users are harder to convert



**Threshold model – For the same #predecessors, highly connected users are less likely to adopt**

# Y! Are High Degree Users Influencers?

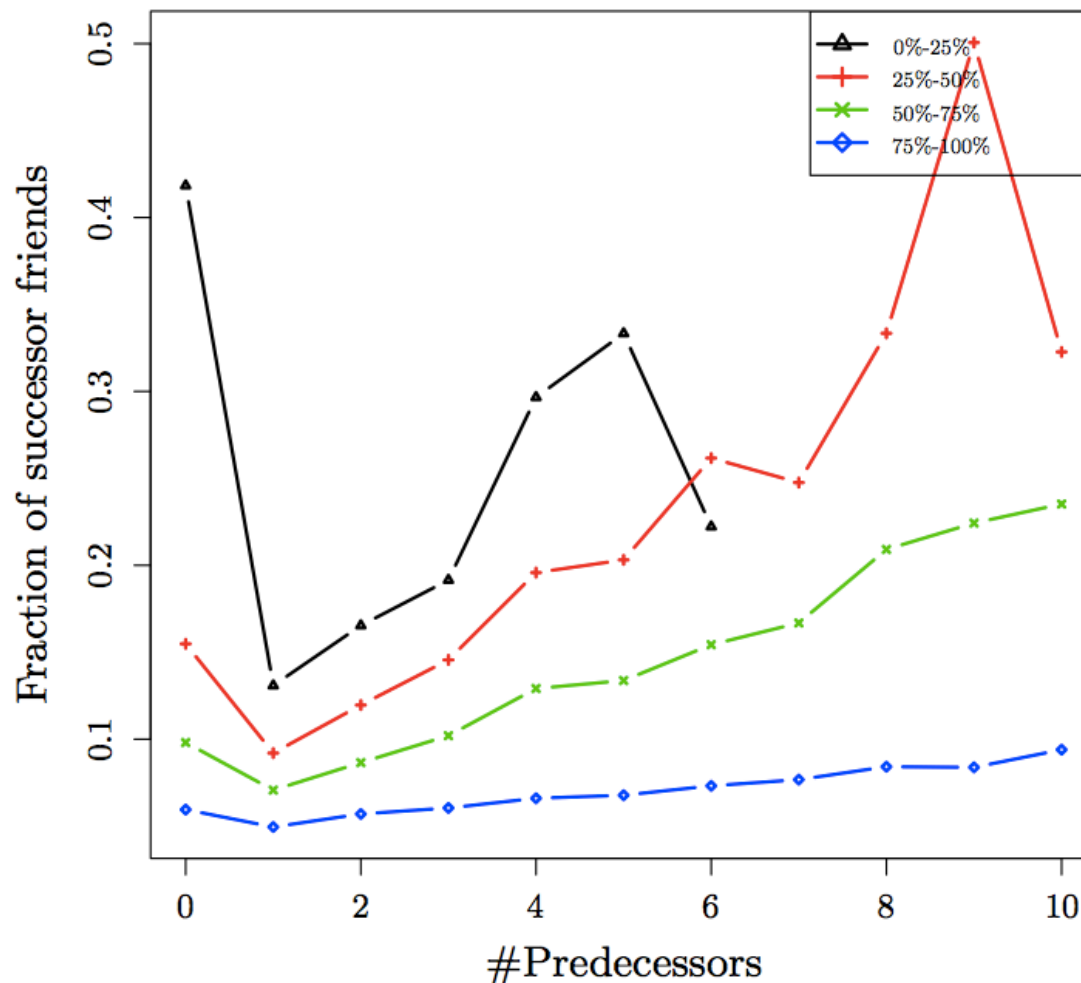


**If high degree → influencer, then there should be difference in #predecessors & #successors**

**Also, Anagnostopoulos et al. '08, Dodds & Watts '07**



# High Degree Users “Underperform”

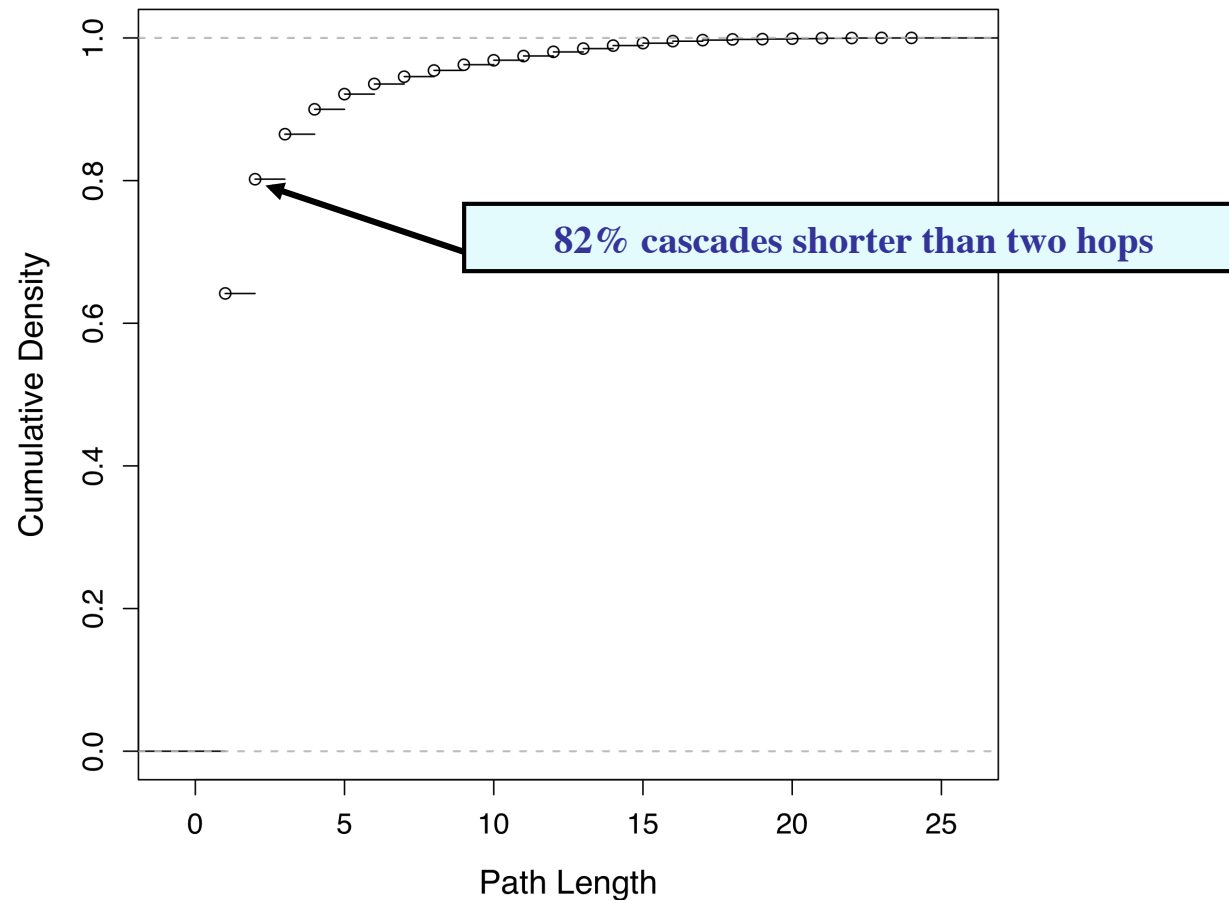


Successors per user reached lower for high degree users  
[Dodds & Watts '07, also Kitsak et al. '10]





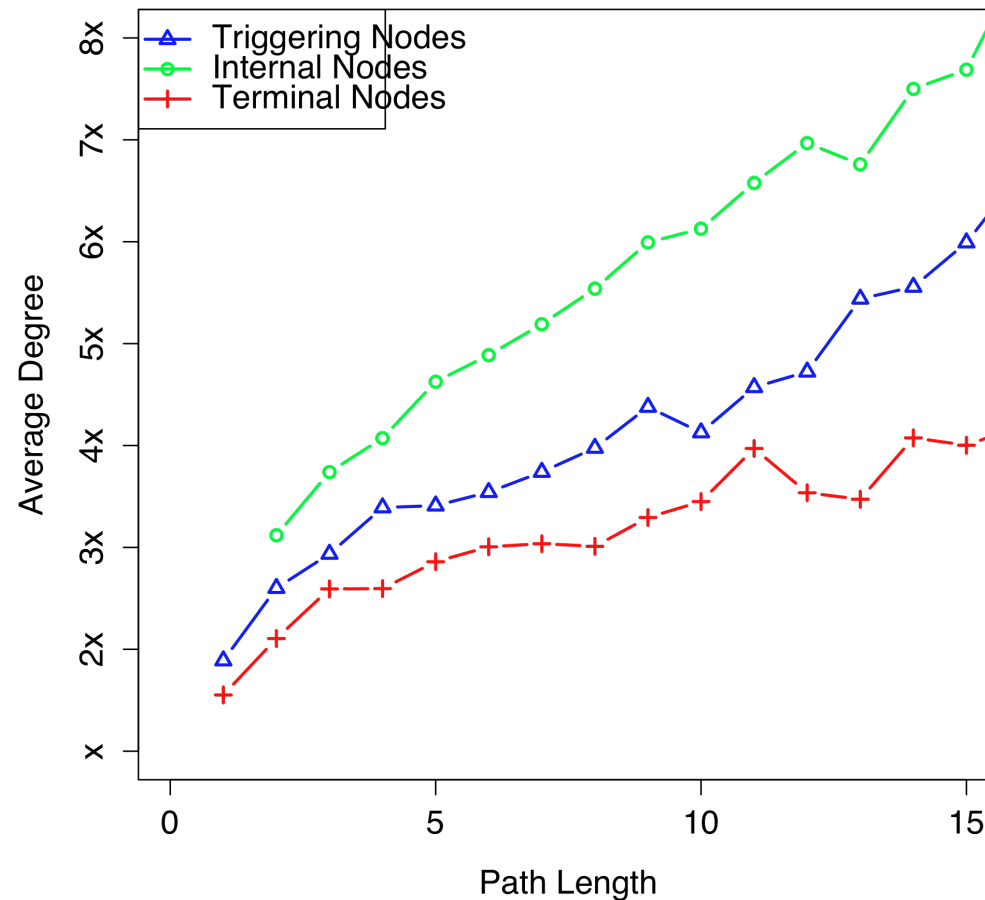
# Adoption Spread is Mostly Local Influence is not Far Reaching



**Longer cascades do exist, but very few**



# Adoption Spread by Internal Nodes



**Longer cascades triggered by high degree users**

**Intermediate nodes sustain the cascades**



# Predicting Future Adoption

- Predict future adoptions, target the likely adopters through messaging

- **Direct:**

- Identify individuals with high probability of adoption
- Message: “*Sign-up now to receive 100 free minutes*”



- **Social Neighborhood:**

- Identify adopters in *prime* social neighborhoods
- Message: “Refer a friend: *Get 100 free minutes per adopting friend*”

Refer-A-Friend

Refer a Friend | How does it Work?

Refer your friends to Raza.com  
For every 4 that become our customers, you can recharge or receive a free \$10 phone card!

Your Email Address:

Friends Email:

Friend #1:

Friend #2:

Friend #3:

Friend #4:

Friend #5:



# Model to Predict Adoption

## Social Neighbourhood:

- # Premium friends
- # Premium friends that are linked
- Total number of friends
- Number of different countries your friends belong to

- **Target variable:**

- Direct marketing: Binary variable indicating adoption during training period
- Train a Decision Tree

## Activity:

- # PC-to-PC Calls
- IMs Sent
- # Friends added
- # logins

## Demographic

- Age
- Gender

## Geographic

- Originating Country

## Training Period:

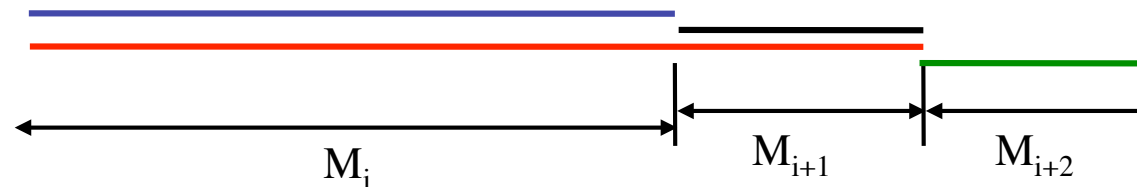
User features till  $M_i$

Adoption in  $M_{i+1}$

## Testing Period:

User features till  $M_{i+1}$

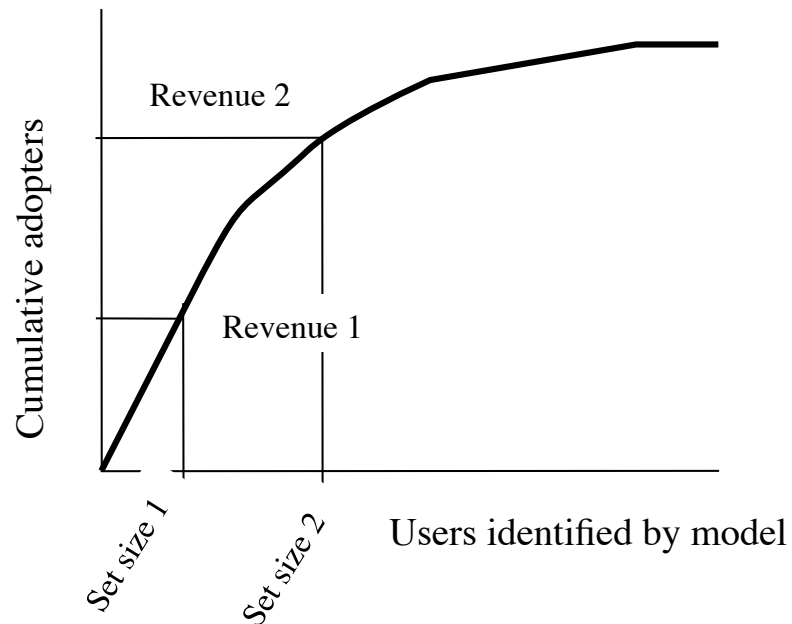
Adoption in  $M_{i+2}$



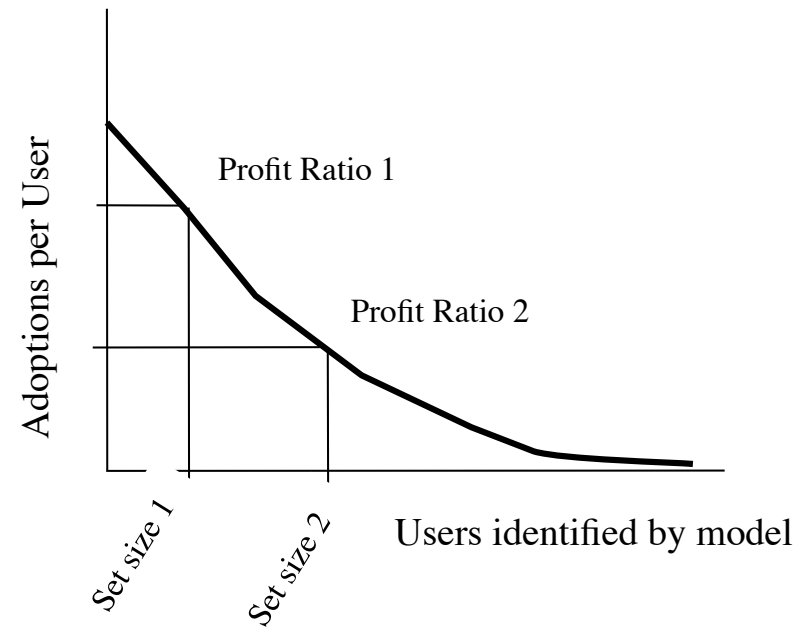


# Model Evaluation

- Metric 1: Cumulative coverage
  - Good if there is a large number of adoptions from a small target pool
  - Does not factor in the cost of targeting users



- Metric 2: Adoptions per user
  - Good when we have the highest adoptions per targeted user
  - Factors in the cost of targeting to each user
  - Helps to decide the *right* incentives for each targeted user





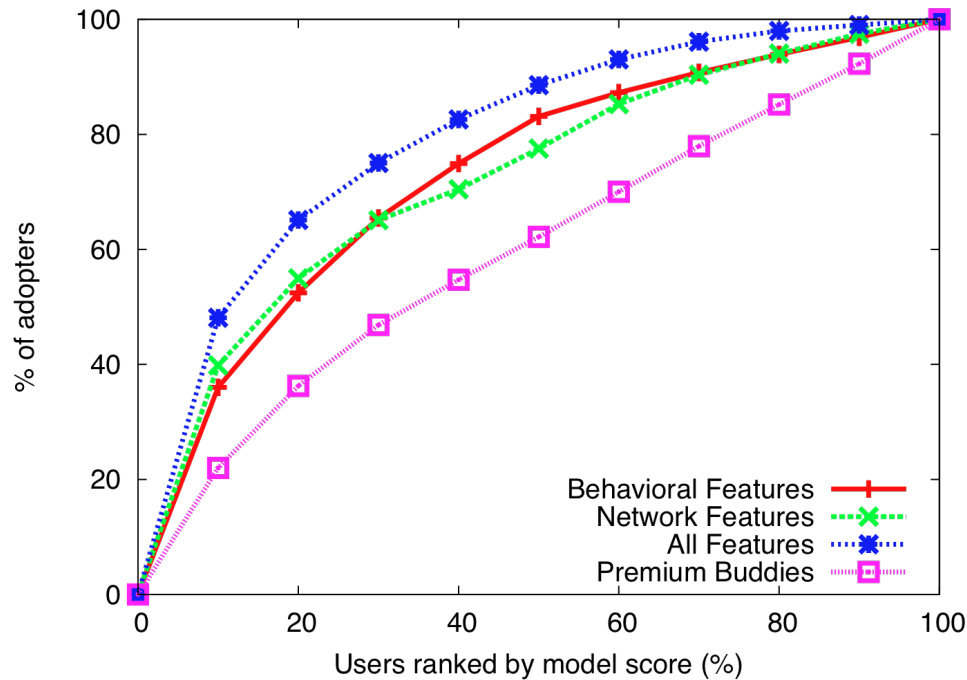
# Methods compared

---

- # Premium predecessors
- Ego-features: User's own behavior
- Social-features: Behavior of one's friends
- Ego+Social



# Direct Targeting



Feature	Relative Importance Score
country	44.76
pc2pc	25.81
prem_bdy	6.32
fringe	6.03
ten_cat	4.51
gender	2.25
age_cat	2.20
n_logins	1.92
n_friends	1.51
buddy_countries	1.27
reach_bdy	1.00

- Combined user and behavioral features best for direct targeting [Also, Hill et al. '06]





# Social Targeting

- Given a set  $A$  of adopters
  - Define  $\sigma(A)$  as the cumulative number of future adoptions among friends of  $A$  (over a specified time period)
- Target variable: Identify the set  $A$  of adopters that maximizes  $\sigma(A)$
- Approach
  - Train a regression model (Gradient Boosted Decision Tree) to predict  $\sigma(A)$
  - Rank order adopters in descending order of  $\sigma(A)$

**Training Period:**

**Users adopting till  $M_i$**

**Friends adopting in  $M_{i+1}$**

**Testing Period:**

**Users adopting till  $M_{i+1}$**

**Friends adopting in  $M_{i+2}$**





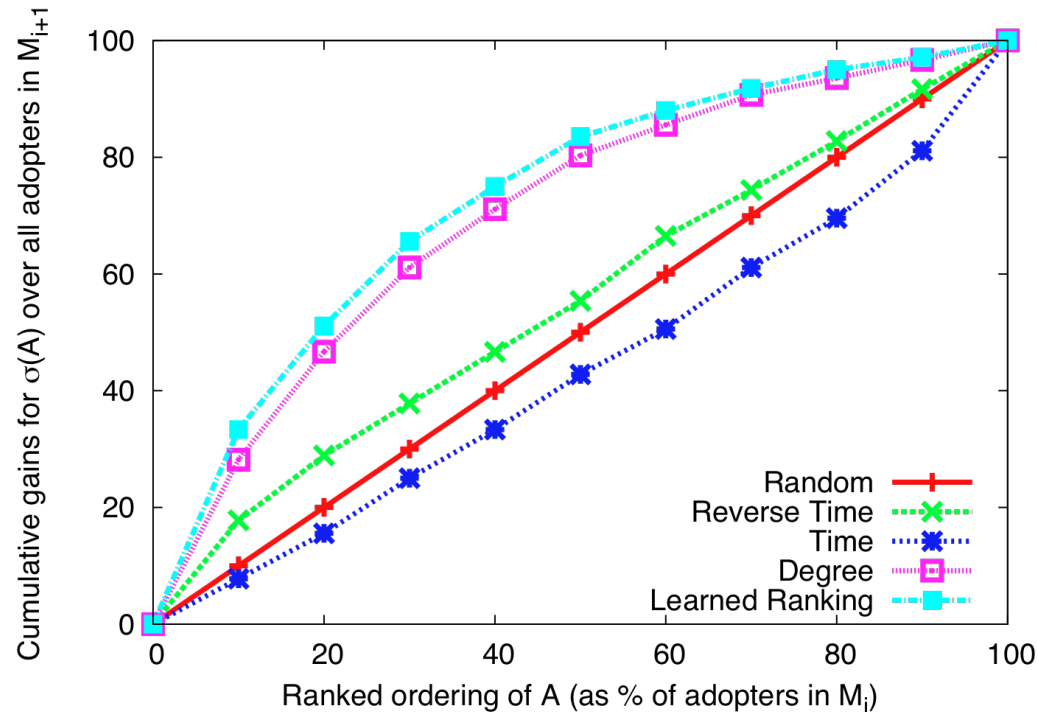
# Methods Compared

---

- Earliest-first
- Latest-first
- Most-connected (!)
  - Enjoys advantage: even random adoption around them will yield highest #successors
- Learned estimates



# Social Targeting



- Learned ranking better than all heuristics



# Caveats

---

- Correlation or causation?
  - Strong correlation in adoption of friends observed from data.  
Does it mean adoptions are induced by social neighborhood?
- Response rates (probability of adoption) are assumed to be same for direct vs. social neighborhood targeting
  - Adoptions per targeted user may be different for the two schemes
- Social neighborhood targeting expects users to recruit their friends
  - Target identification is “crowd sourced” to selected users in the neighborhood



## Conclusion

---

- Neighborhoods, not Individuals (!)
  - Most models already allow this (e.g., IC, Threshold)
- Behavior, Demographics, Geography, Social Neighborhood:  
All matter
- Both targeting methods better than well-established heuristics
  - Social targeting: Assumes that users will **select** and recruit friends with right incentive
  - Direct targeting: Useful to “start off” neighborhoods that are sparse in adoption

# Y! Q&A



Thank you!



# Ad Selection with a Chunked Price Model

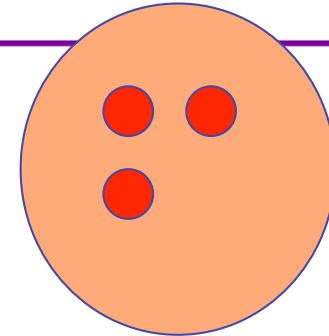
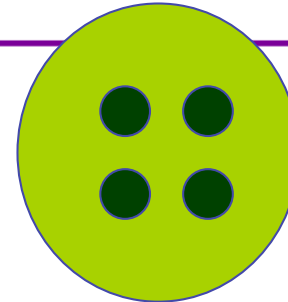
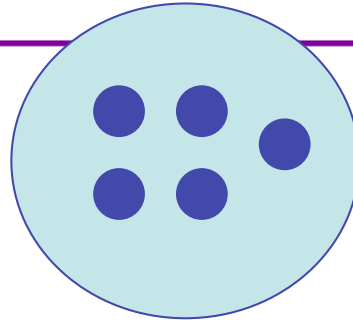
**Narayan Bhamidipati, Rushi Bhatt, Michael Grabchak (Cornell)**



# Chunked Price Model: The Setup

**Goal:**

#successes needed



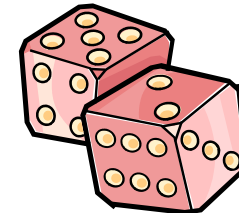
**Reward:**

Payout on achieving goal



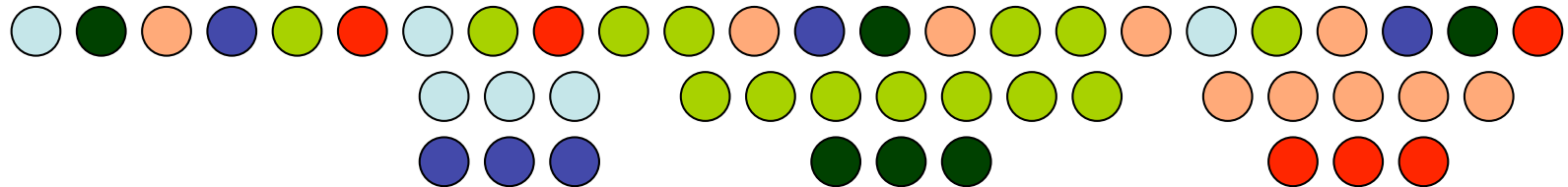
**Probability:**

of success in an attempt



**Time:**

#attempts







# Objective: Revenue Maximization

---

- At each time  $1 \leq t \leq T$ 
  - decide which ad to show
  - observe the outcome (click or no click)
  - revise the goals
  - repeat
- In such a way that total (expected) revenue is maximized at time  $T$
- Obviously, once a goal for an ad is met, that ad is never shown again



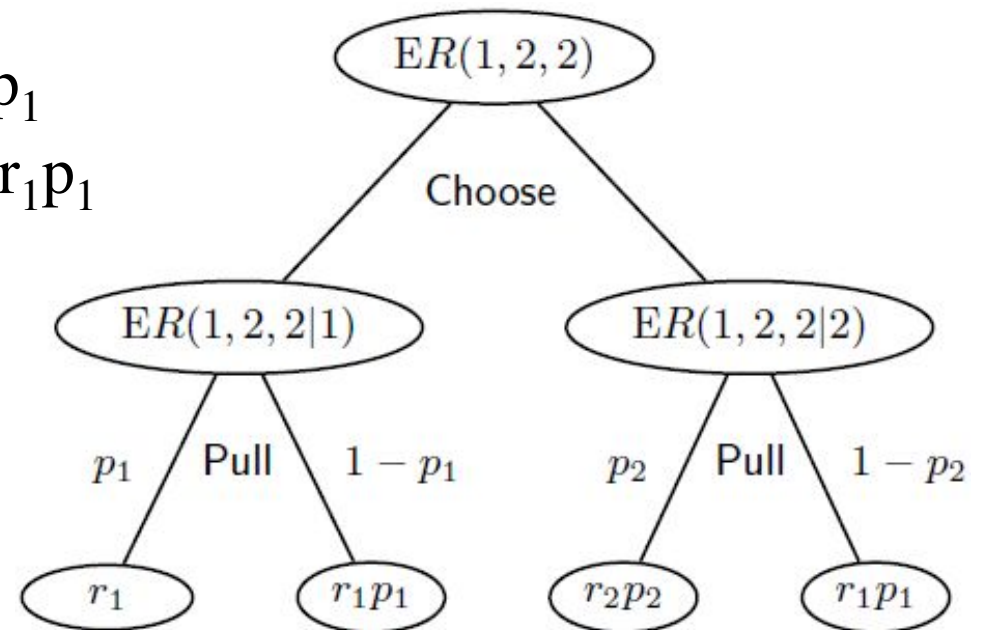
# Optimal Algorithm

- Let  $n_1=1$ ,  $n_2=2$ , and  $T=2$ ,
- Also, let  $r_1p_1 < r_2p_2$
- $ER(1,2,2) = \max\{ ER(1,2,2|1), ER(1,2,2|2)\}$

➤ where,

- $ER(1,2,2|1) = p_1r_1 + (1-p_1)r_1p_1$
- $ER(1,2,2|2) = p_2r_2p_2 + (1-p_2)r_1p_1$

- Complexity:  $O(4^T)$





# Greedy Algorithms

---

- For each ad, define an index that
  - Considers only ads with unattained but attainable goals
  - Increases with  $p_i$  and  $r_i$ , decreases with  $n_i$ .
  - Is a scalar multiple of  $r_i$ .
- Some indices:
  - $r_i p_i / n_i$  : continues showing the same ad
  - $r_i p_i / n_i P(n_i \text{ in } T)$  : depends on  $T$  also
  - $r_i P(n_i \text{ in } T)$  : simplified, yet feasible



# Theoretical Guarantees

---

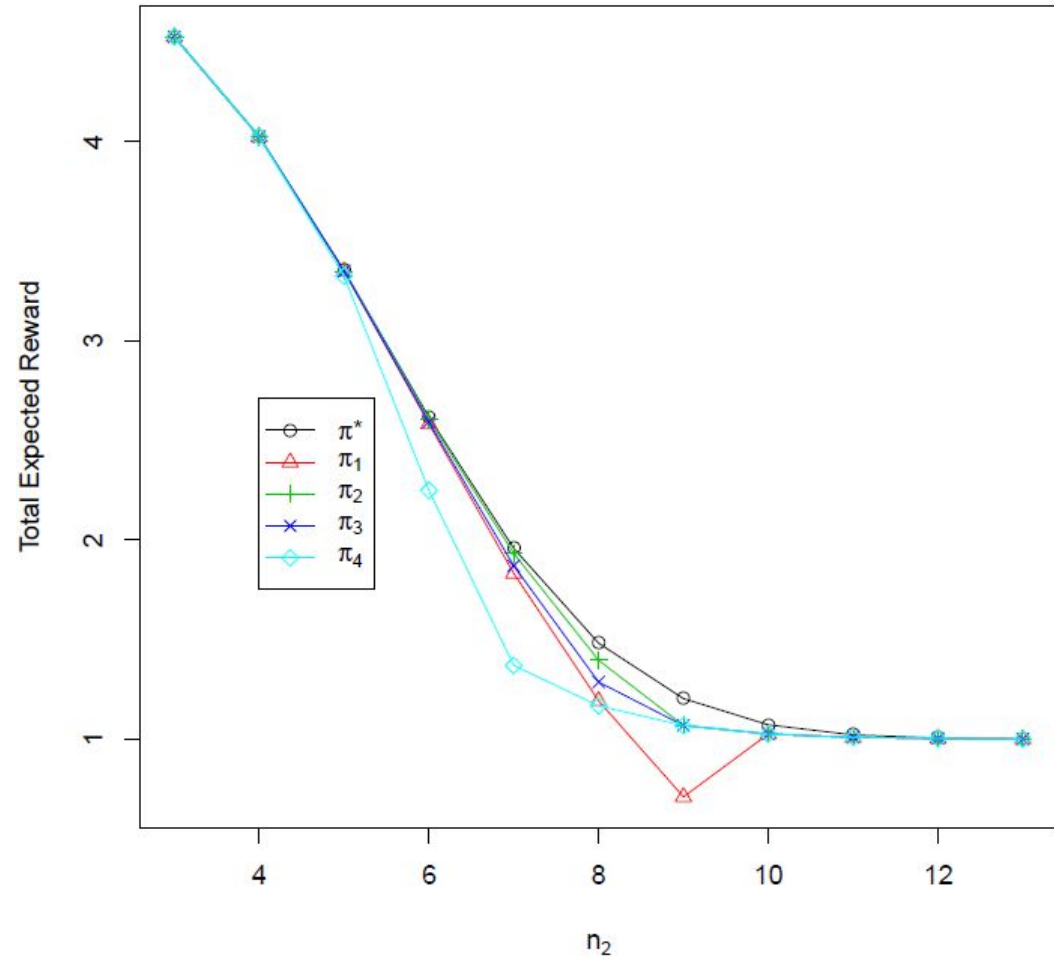
- Make use of the parallels with Stochastic Knapsack Framework
  - Items = ads,
  - random weight  $W_i = \# \text{attempts to attain goal}$
- Introduce artificial policies like
  - $\gamma_0 =$  always show the ad with max exp reward
  - $\gamma_7 =$  choose the better of  $\gamma_0$  and  $r_i p_i / n_i$ .
- Obtain a 3-approximation
  - under mild assumptions:  $P(n_i \text{ in } T) \geq 1/2$



# Optimal vs. Greedy Performance

- Algorithms are much closer to optimal
- Identical to optimal in certain regions
- Significantly lower for some cases

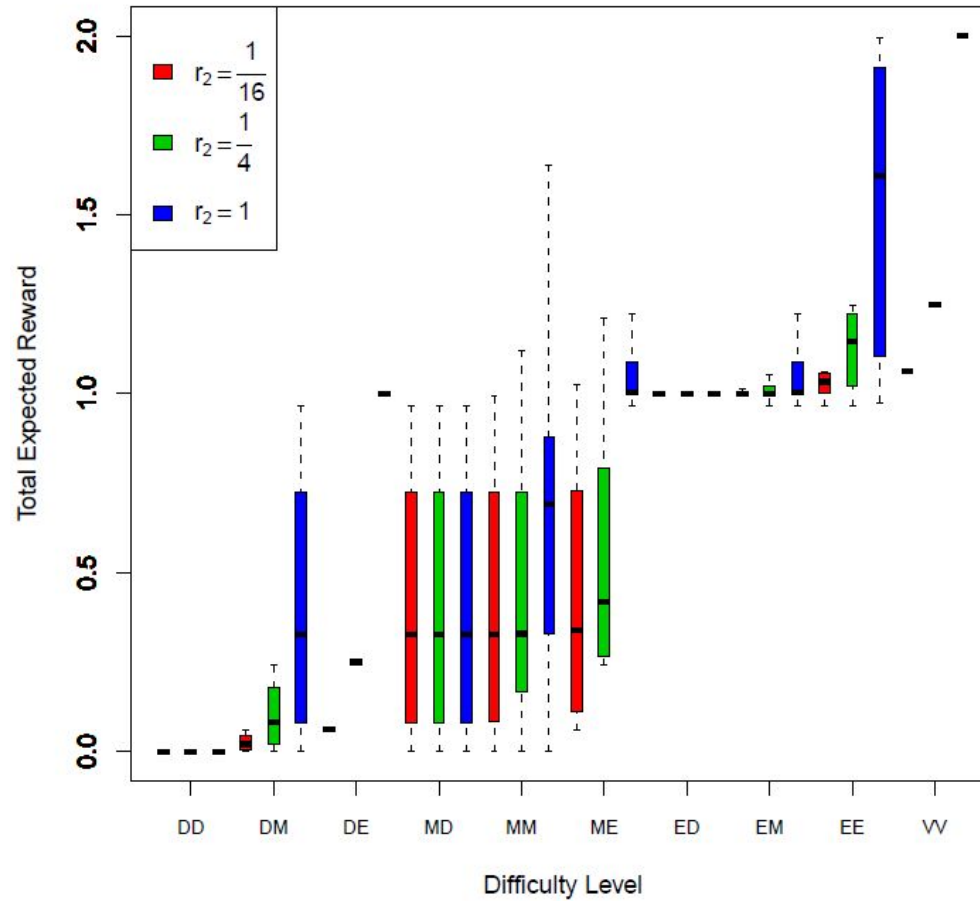
Expected Reward for  $k=2$ ,  $p_1 = \frac{1}{4}$ ,  $p_2 = \frac{1}{16}$ ,  $r_2 = 4$ ,  $n_1 = 10$ ,  $T = 100$



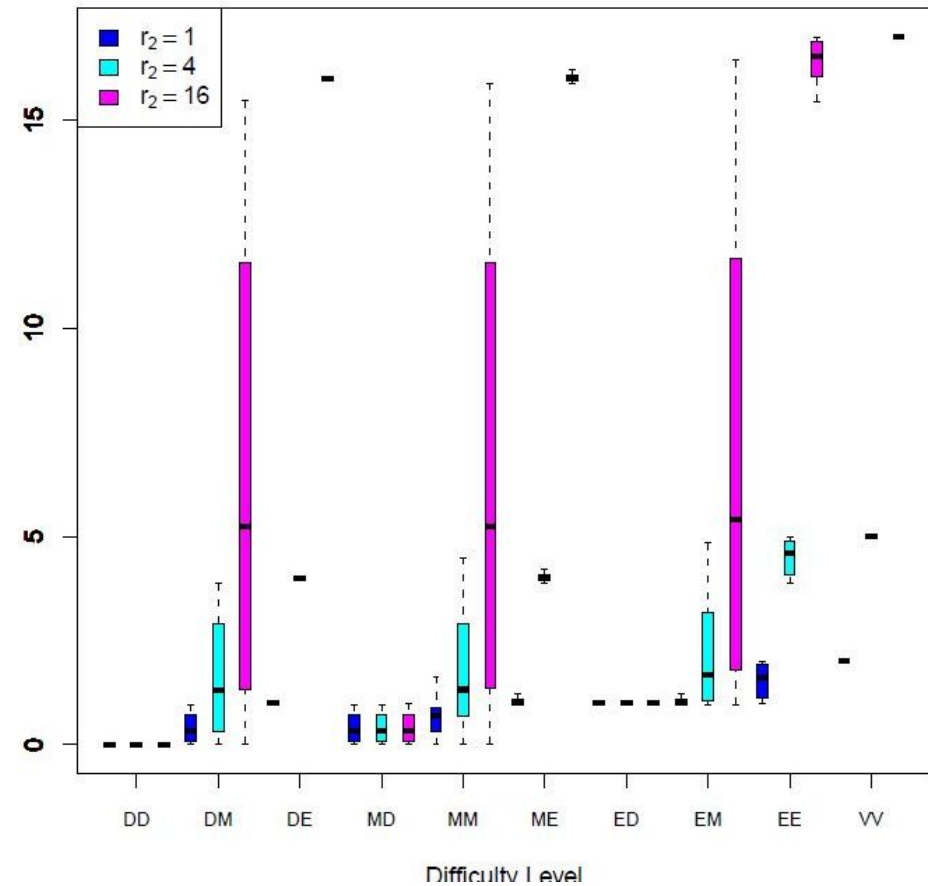


# Complete Enumeration: Optimal

Optimal rewards by difficulty levels, for  $r_2 \leq 1$



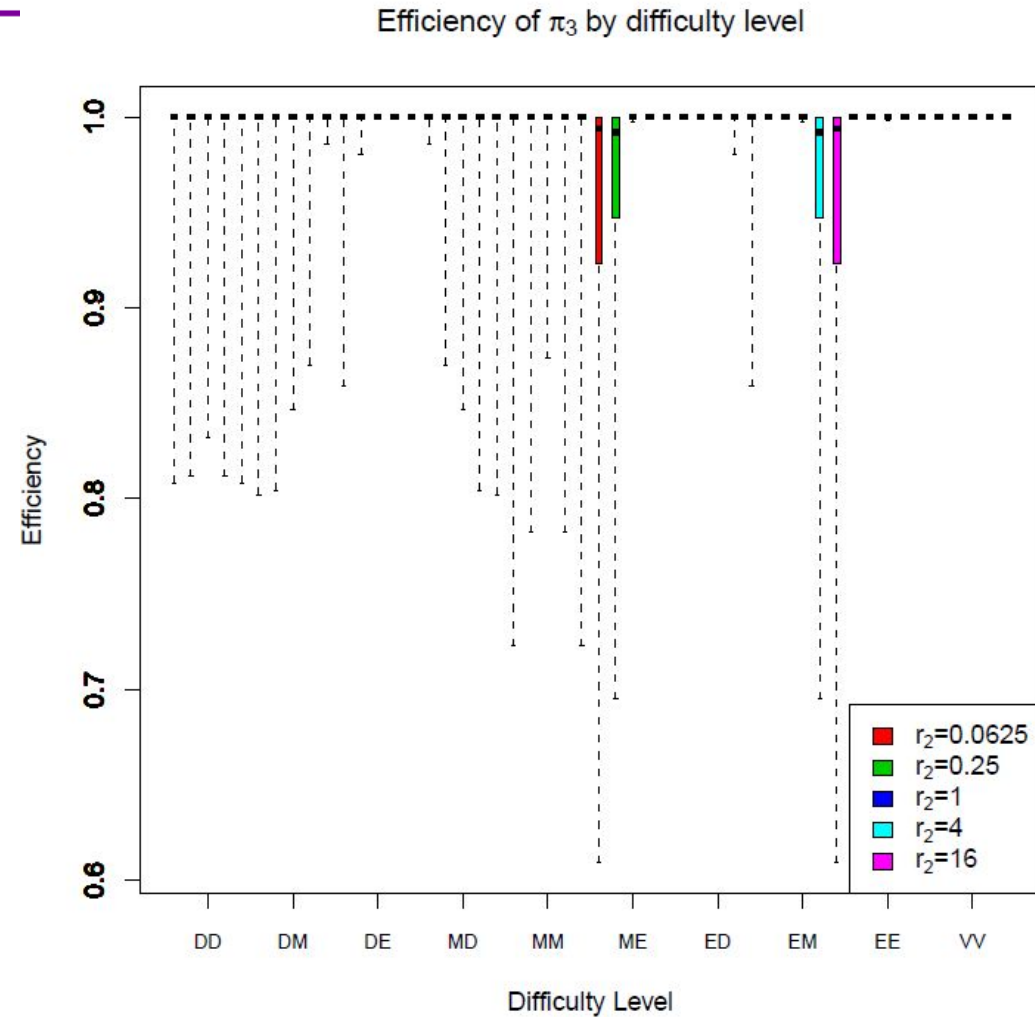
Optimal rewards by difficulty levels, for  $r_2 \geq 1$





# Relative Performance of Greedy

- Two ads
- Several combinations of
  - Goals
  - Rewards
  - Probabilities
- Segregated by difficulty levels





## Real Time Bidding (Current work!)

---

- Once the next ad to be shown is determined
  - Need to bid for it in the NGD system
  - Bid should be proportional to the value the ad is supposed to bring
  - A missed opportunity would imply one less attempt available
  - Maximize profits by optimizing bid prices