The Science behind on-line Advertising

Rushi Bhatt

Yahoo! Labs, Bangalore

December13, 2011

Agenda

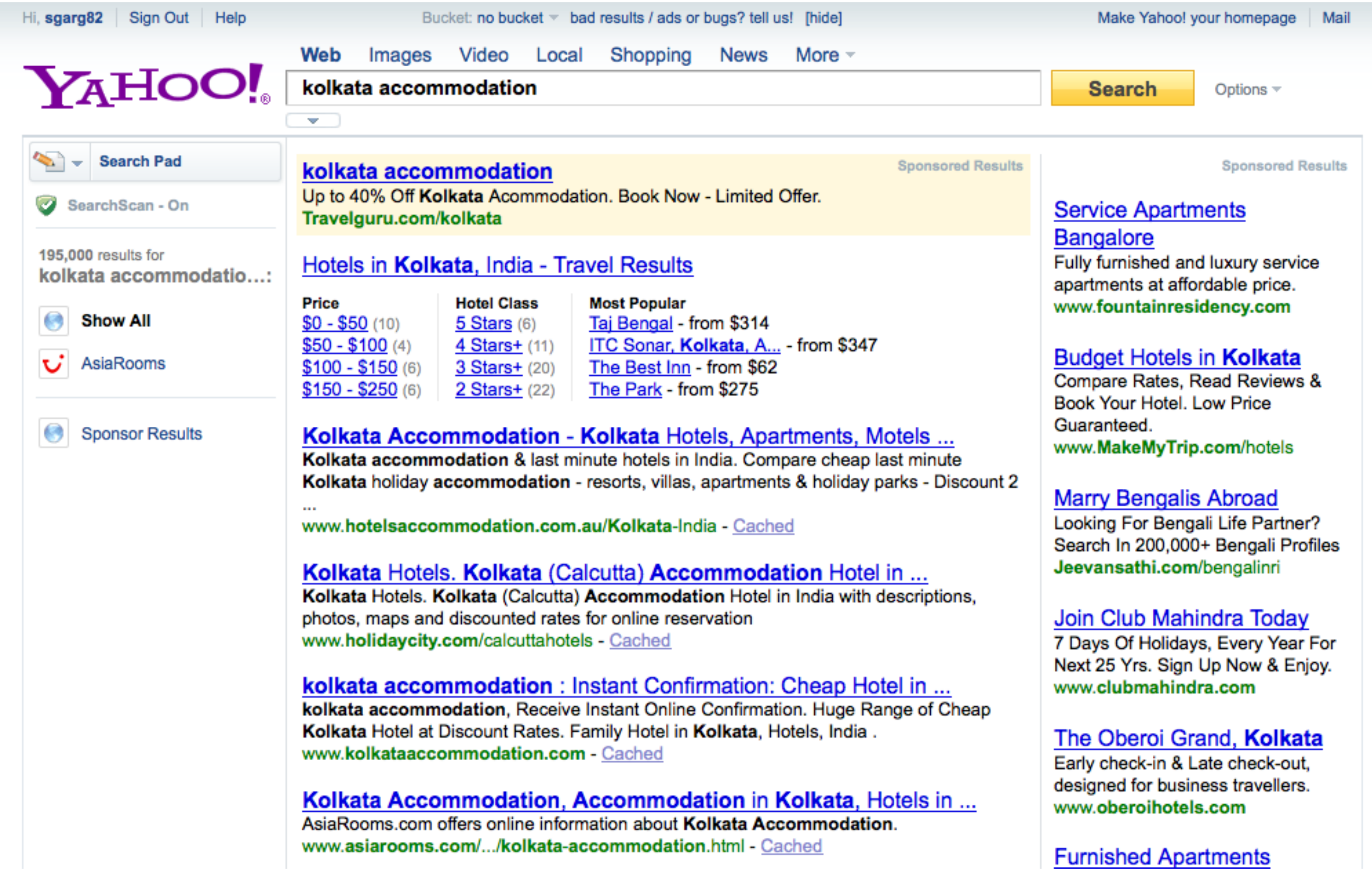
- Introduce Web monetization
  - Search and Display Advertisement
- Auctions
  - Participation in Auctions
  - Pricing
- Exchanges
  - Ranking, response prediction
- “Next gen” monetization
  - Social Targeting
  - Chunked-rewards

A Free Web!

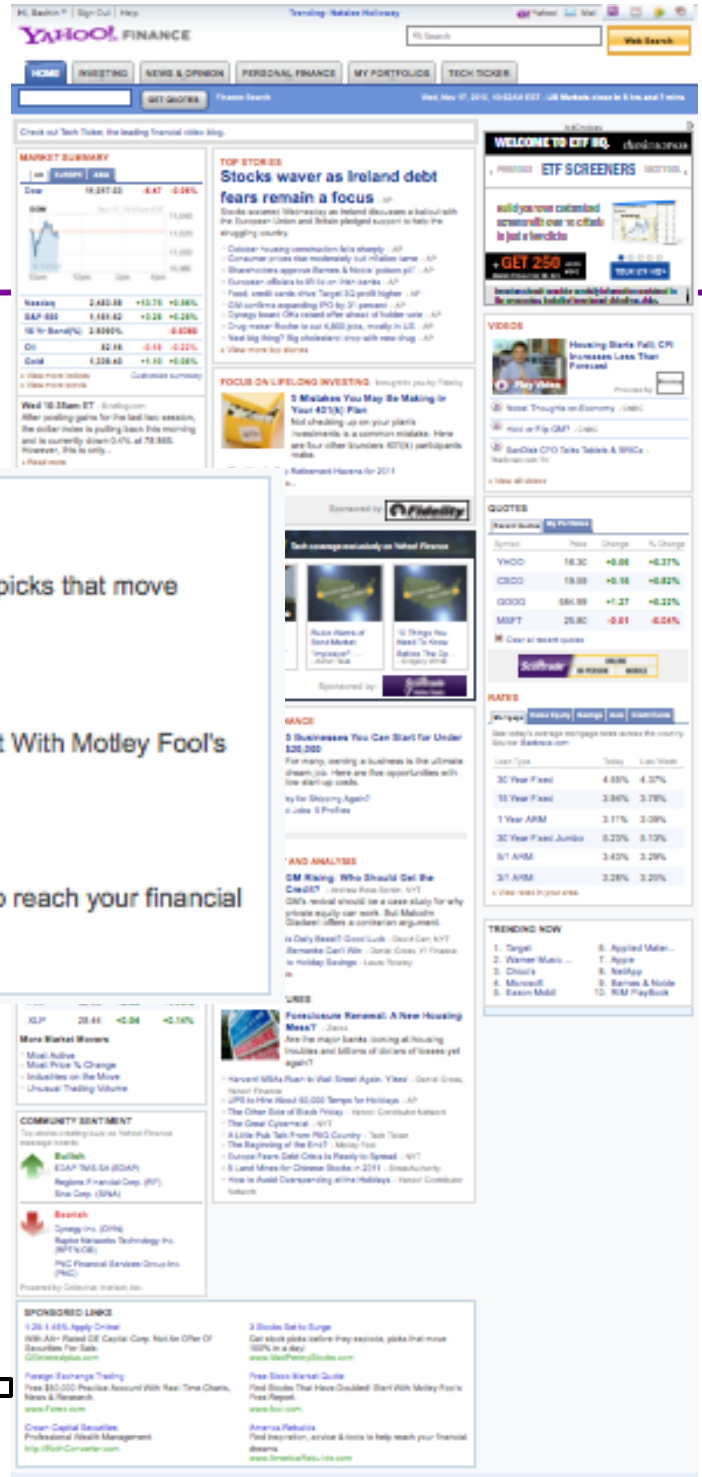
- The WWW comes free for users
  - Browse, search, communicate, chat, socialize, all free!
- Web is funded by advertisers!
  - To a large extent (Yahoo!, Google, Facebook, Bing)

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
YAHOO!
Web Images Video Local Shopping News More
kolkata accommodation
Search
Options
Search Pad
SearchScan - On
195,000 results for
kolkata accommodatio...:
Show All
AsiaRooms
Sponsor Results
Sponsored Results
kolkata accommodation
Up to 40% Off Kolkata Acommodation. Book Now - Limited Offer.
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
Taj 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 - 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 - 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 - Cached
Kolkata Accommodation, Accommodation in Kolkata, Hotels in ...
AsiaRooms.com offers online information about Kolkata Accommodation.
www.asiarooms.com/.../kolkata-accommodation.html - Cached
Sponsored Results
Service Apartments
Bangalore
Fully furnished and luxury service
apartments at affordable price.
www.fountainresidency.com
Budget Hotels in Kolkata
Compare Rates, Read Reviews &
Book Your Hotel. Low Price
Guaranteed.
www.MakeMyTrip.com/hotels
Marry Bengalis Abroad
Looking For Bengali Life Partner?
Search In 200,000+ Bengali Profiles
Jeevansathi.com/bengalinri
Join Club Mahindra Today
7 Days Of Holidays, Every Year For
Next 25 Yrs. Sign Up Now & Enjoy.
www.clubmahindra.com
The Oberoi Grand, Kolkata
Early check-in & Late check-out,
designed for business travellers.
www.oberoihotels.com
Furnished Apartments

Contextual Ads



SPONSORED LINKS

1.25-1.45% Apply Online!
With AA+ Rated GE Capital Corp. Not An Offer Of Securities For Sale.
GEinterestplus.com

3 Stocks Set to Surge
Get stock picks before they explode, picks that move 100% in a day!
www.MadPennyStocks.com

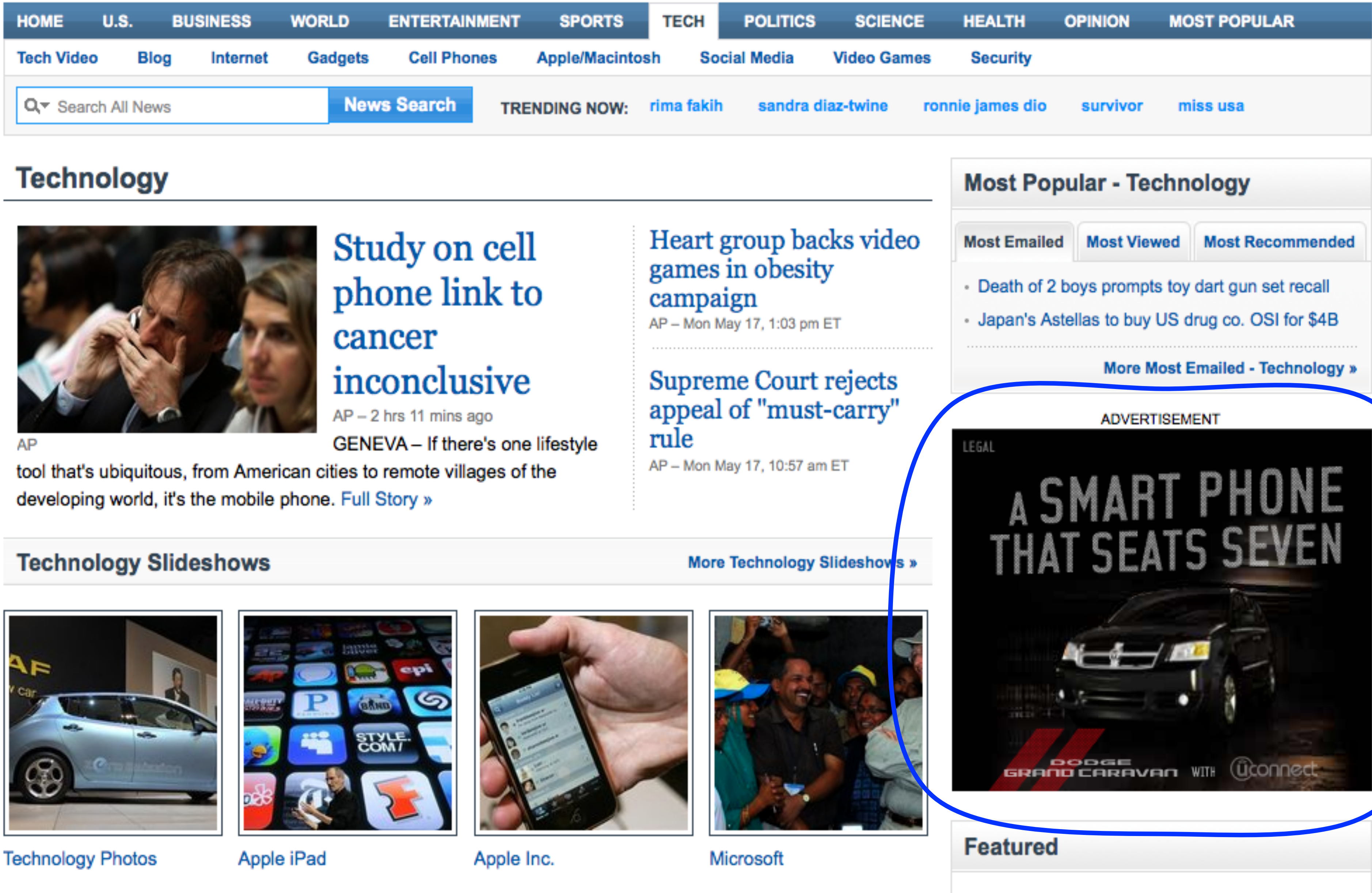
Foreign Exchange Trading
Free $50,000 Practice Account With Real-Time Charts, News & Research.
www.Forex.com

Free Stock Market Quote
Find Stocks That Have Doubled! Start With Motley Fool's Free Report.
www.fool.com

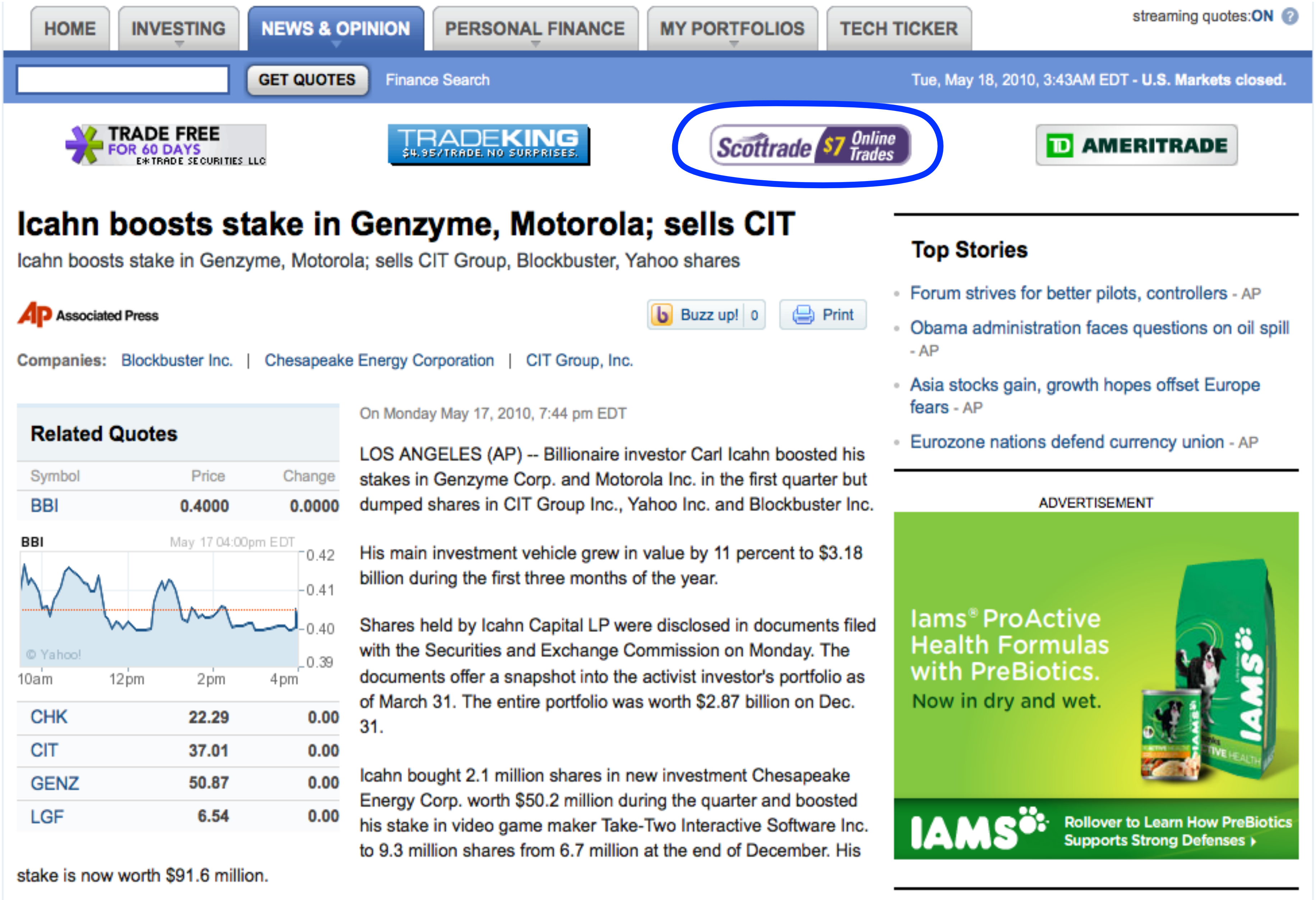
Crown Capital Securities
Professional Wealth Management
http://RothConverter.com

America Rebuilds
Find inspiration, advice & tools to help reach your financial dreams.
www.AmericaRebuilds.com

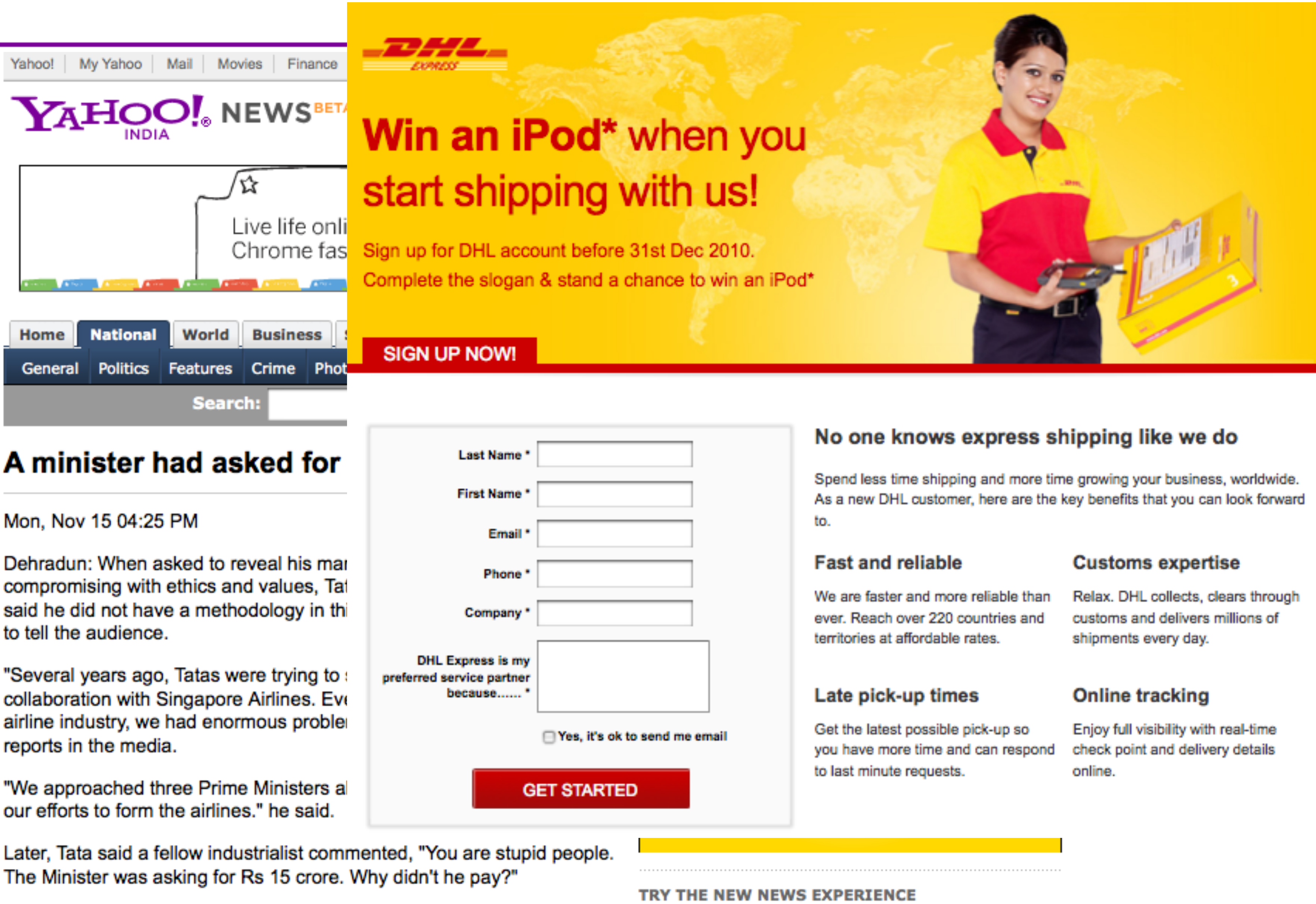
Display Ad: Impression



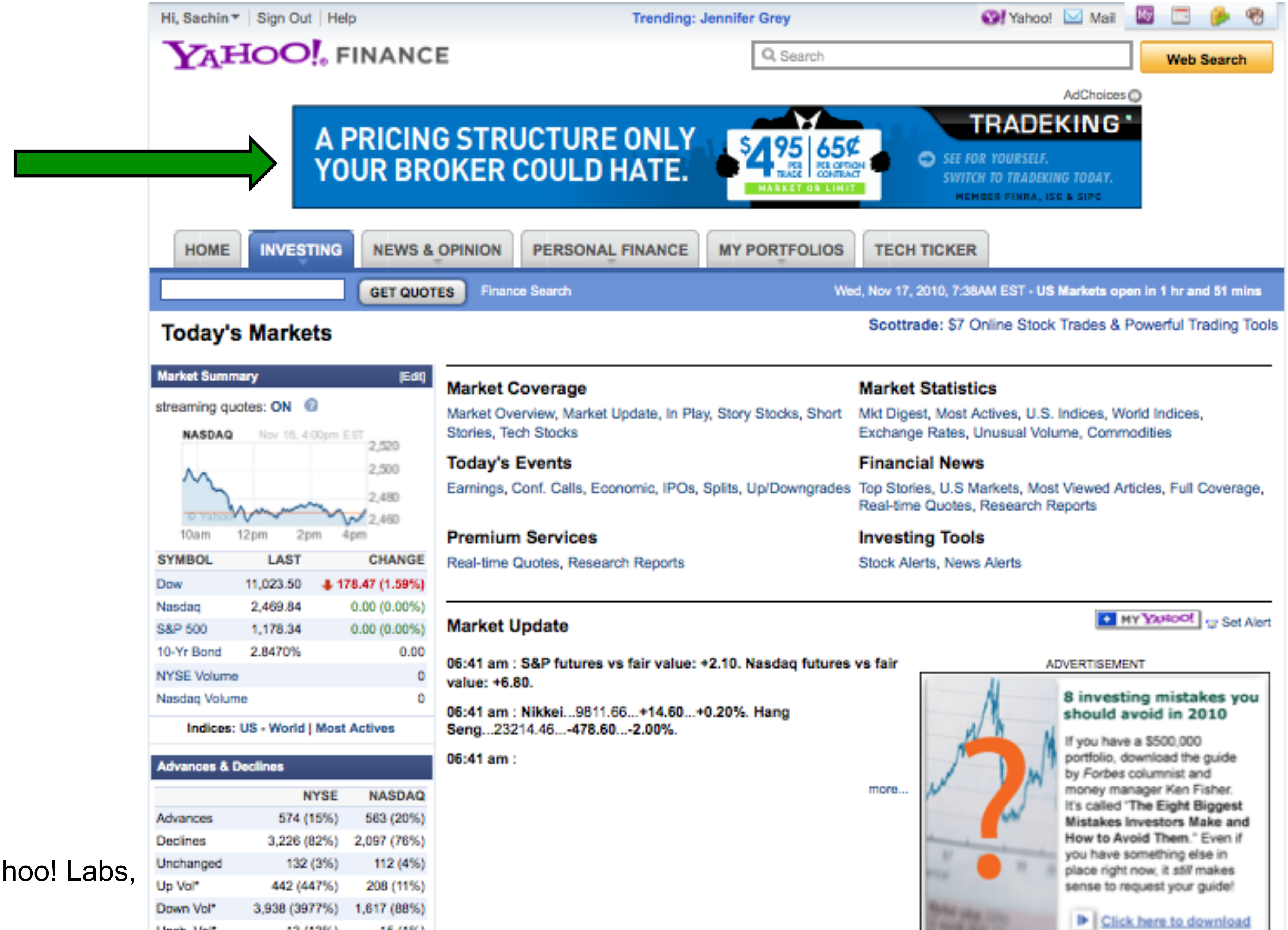
Display Ad: Click



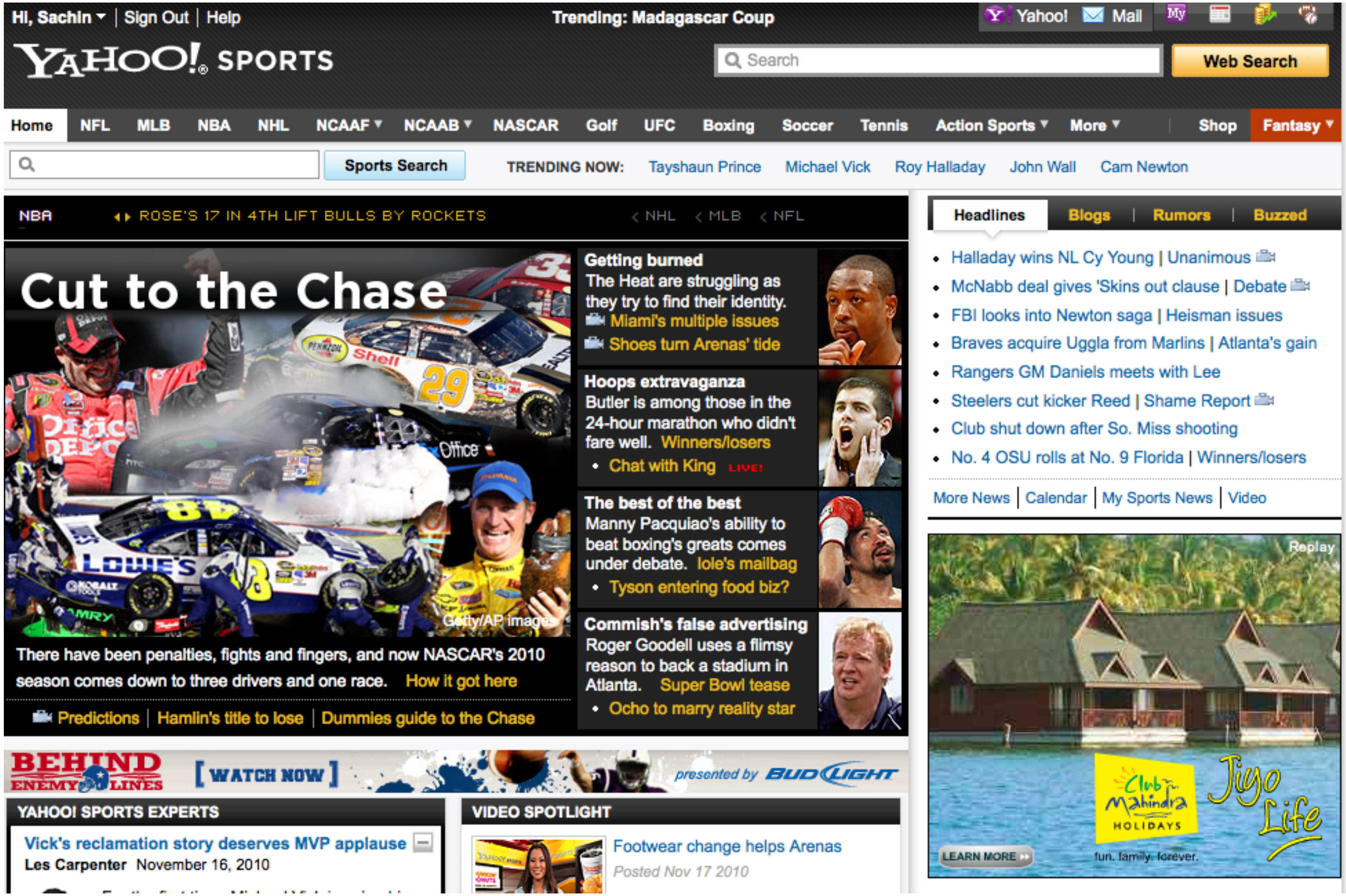
Display Ad: Conversion



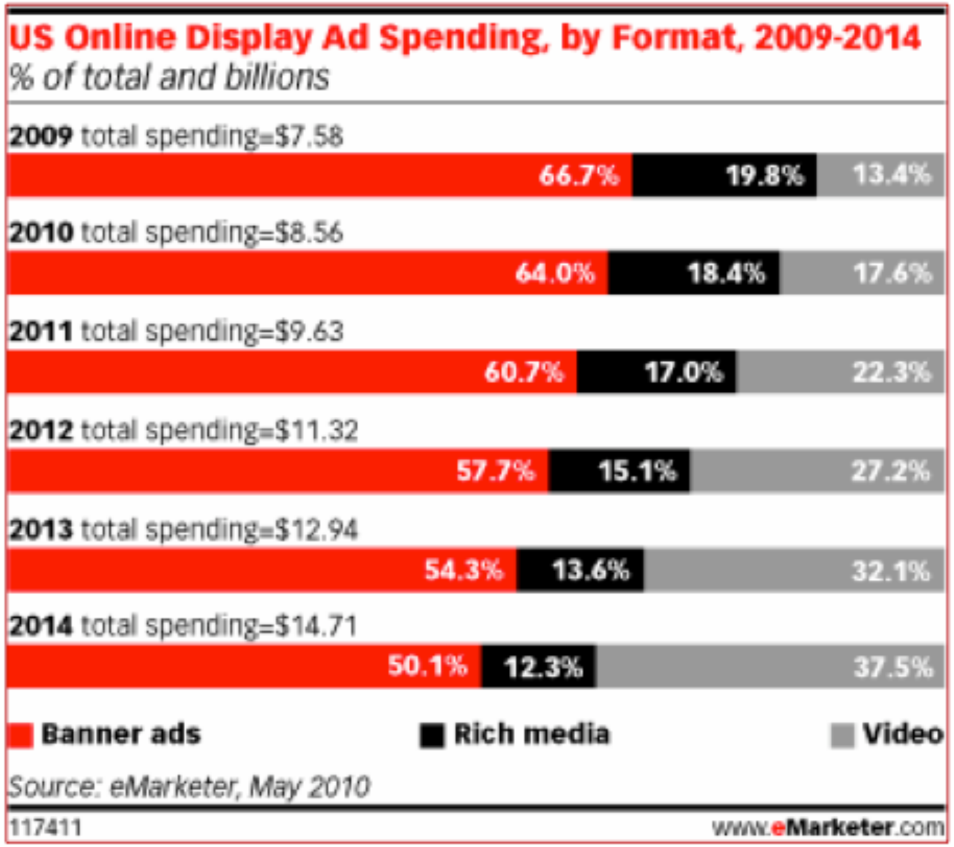
Display Ad: Publisher targeting

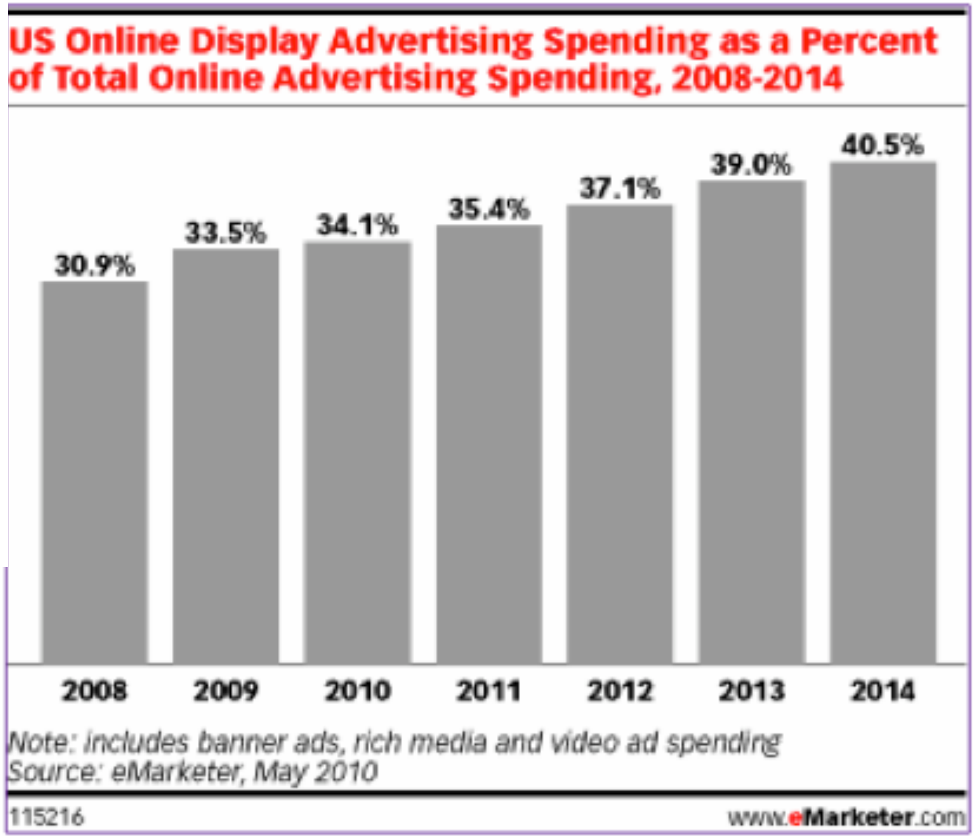


Display Ad: Location targeting



Display Ads is big business





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

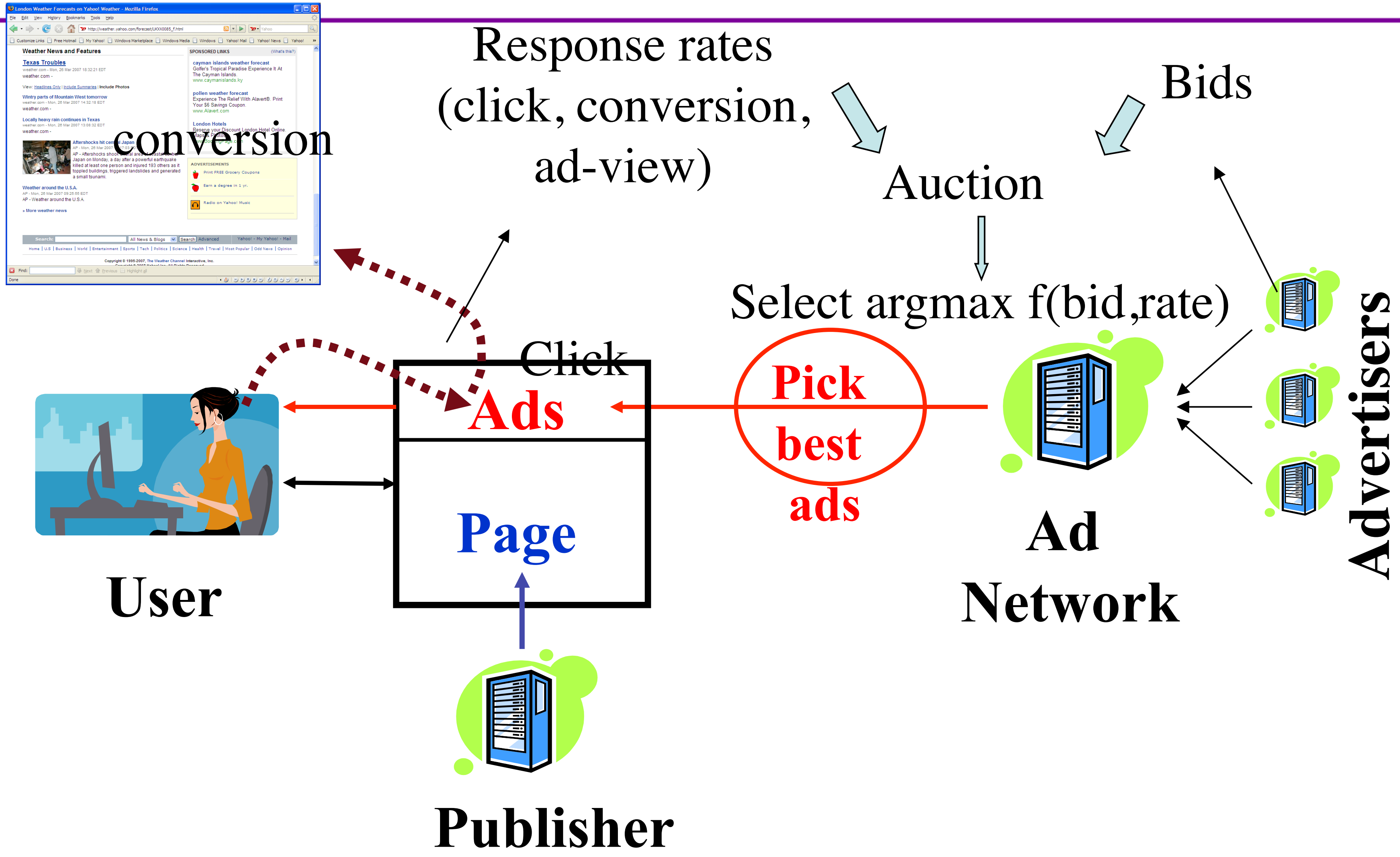
- Advertiser: Wants high RoI (e.g. cost per lead or conversion)
  - Demand: Buyer of “ad impressions” from publishers
  - Targets users (e.g. Males, from California) and pages (e.g. sports)
  - May require guaranteed impressions (e.g. 10 Million in 30 days)
  - Or competes in an on-line auction to win 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 ?



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 * $P(Click|u,p)$= $0.003, (if $P(Click|u,p)$ = 0.001)
    - Ad 3 = $90 * $P(Conv|u,p)$= $0.0009, (if $P(Conv|u,p)$ = 0.0001)
  - Auction conducted on expected revenue: highest one wins
  - Ad 2 wins and gets shown to the user

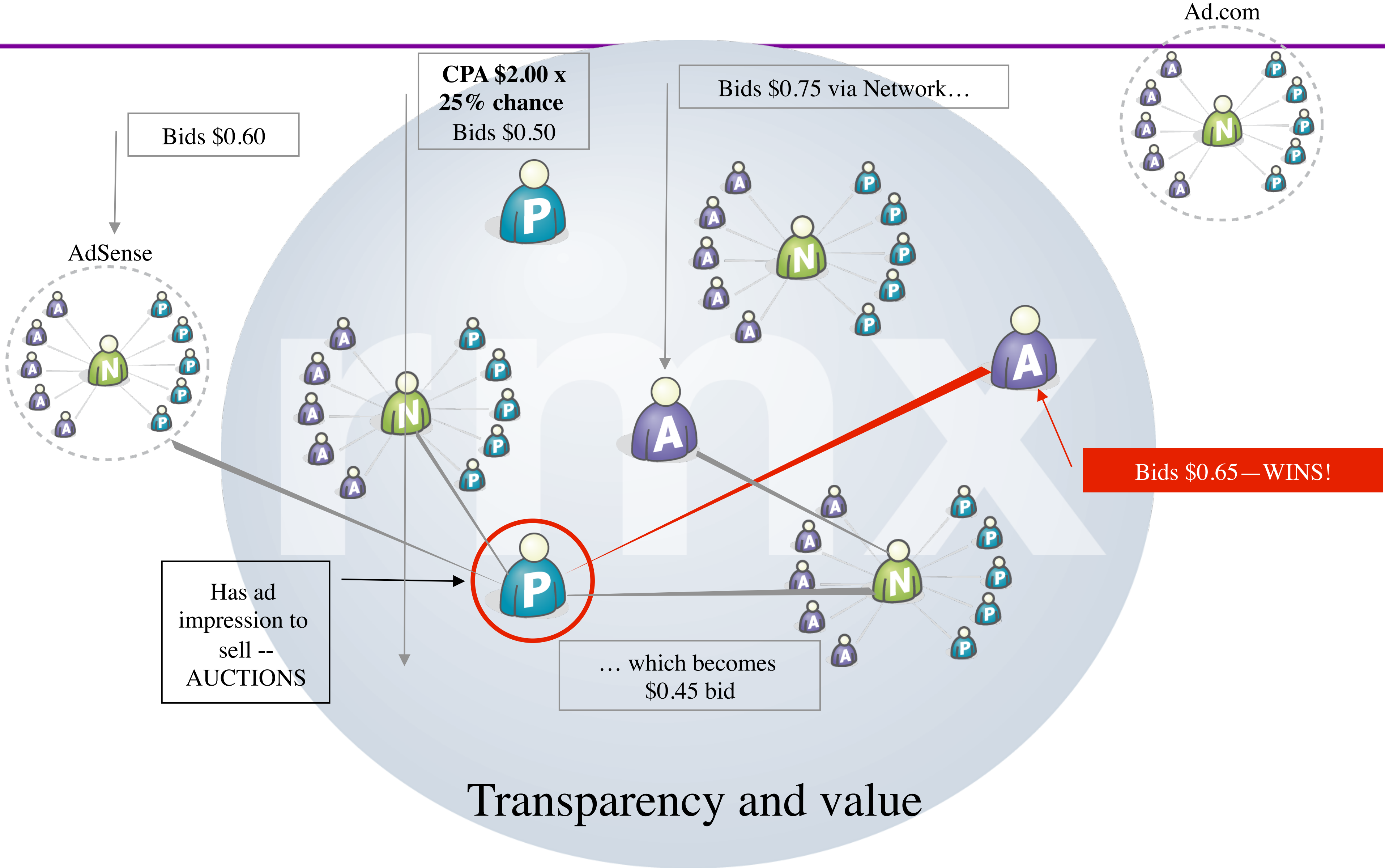
Ad Selection: Prediction

- True $P(Click|u,p)$, $P(Conv|u,p)$ are unknown
- Need to predict
  - Say $\hat{P}(Click|u,p)$ for Ad 2 = 0.0006 (true = 0.001), then Ad 1 wins $\Longrightarrow$ Loss of \$0.001
  - Say $\hat{P}(Conv|u,p)$ for Ad 3 = 0.0004 (true = 0.0001), then Ad 3 wins $\Longrightarrow$ 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

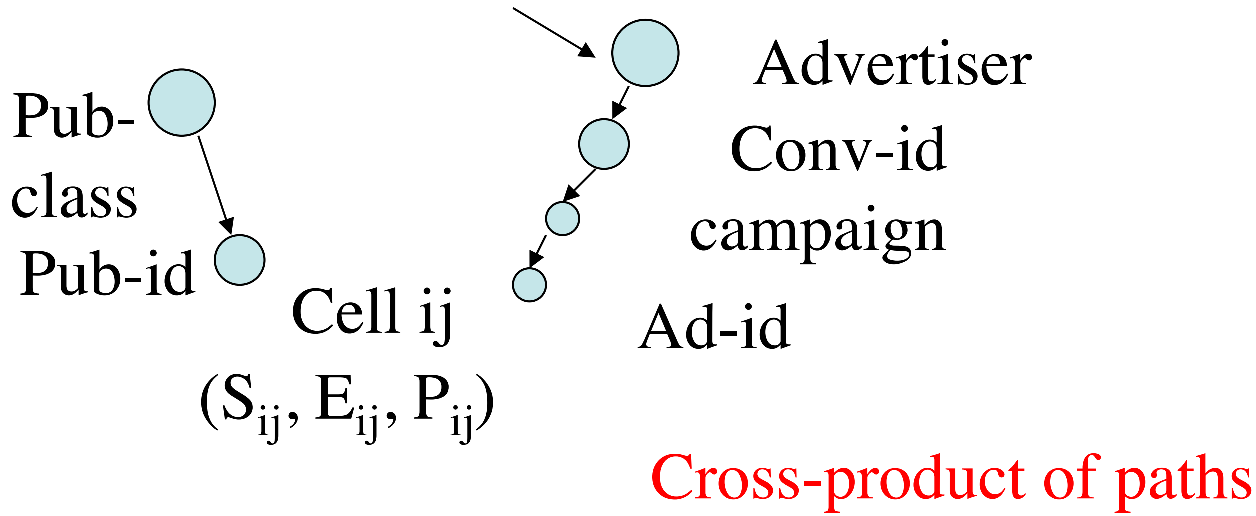
- Sparsity:
  - Too many combinations, most cells have $N_{ijk} = 0$
    - High dimensional categorical variables, e.g. In CLICK data, 100M cells
  - ‘cold-start’ problem
  - If not zero, most cells have small tries
- Rare response:
  - Response rates are extremely rare
  - 0 in 100 is not the same as 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)



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

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

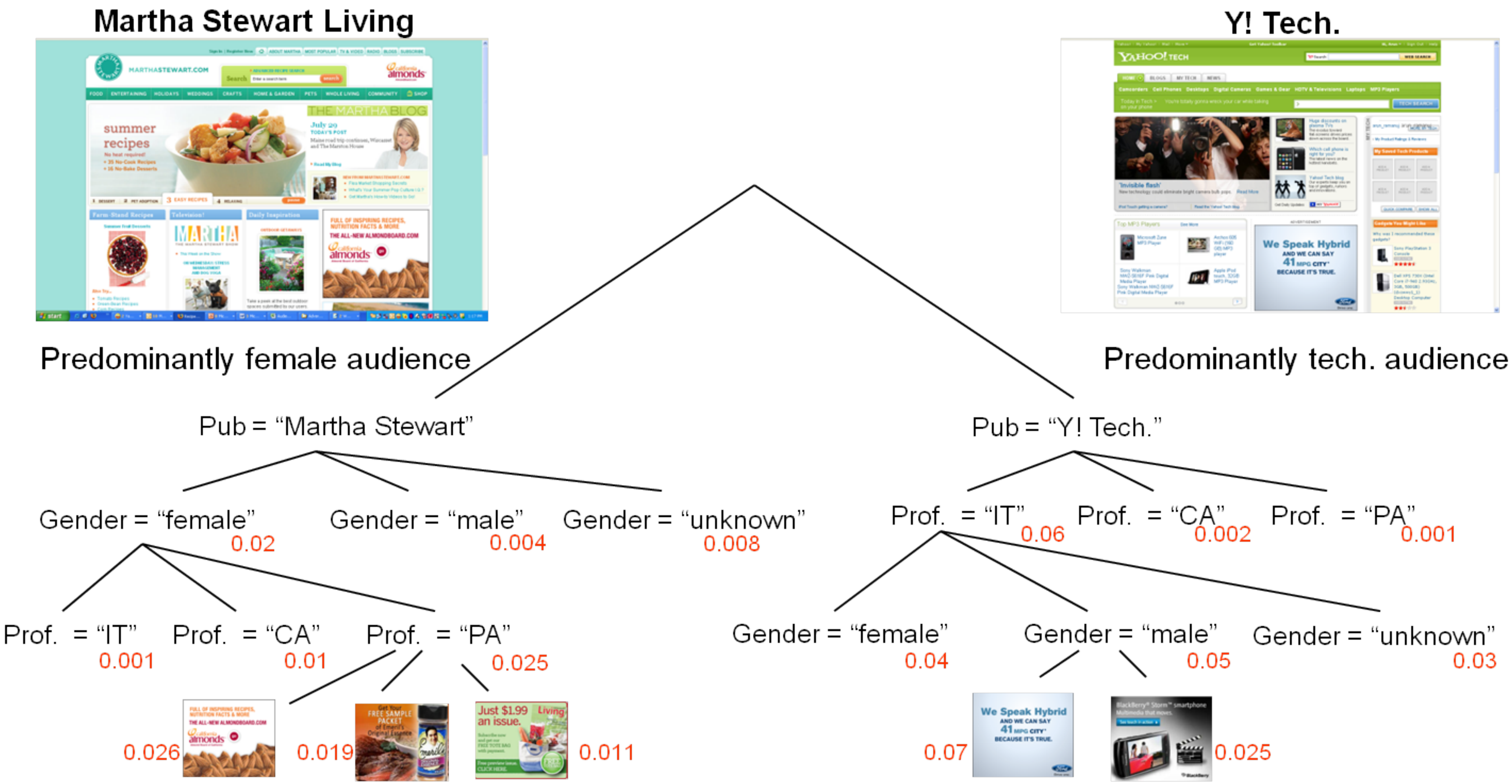
- Use singleton features only
- Use conjunction features
  - Need hashing trick to reduce dimensionality [Weinberger et. Al, ICML 2009]

- 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



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

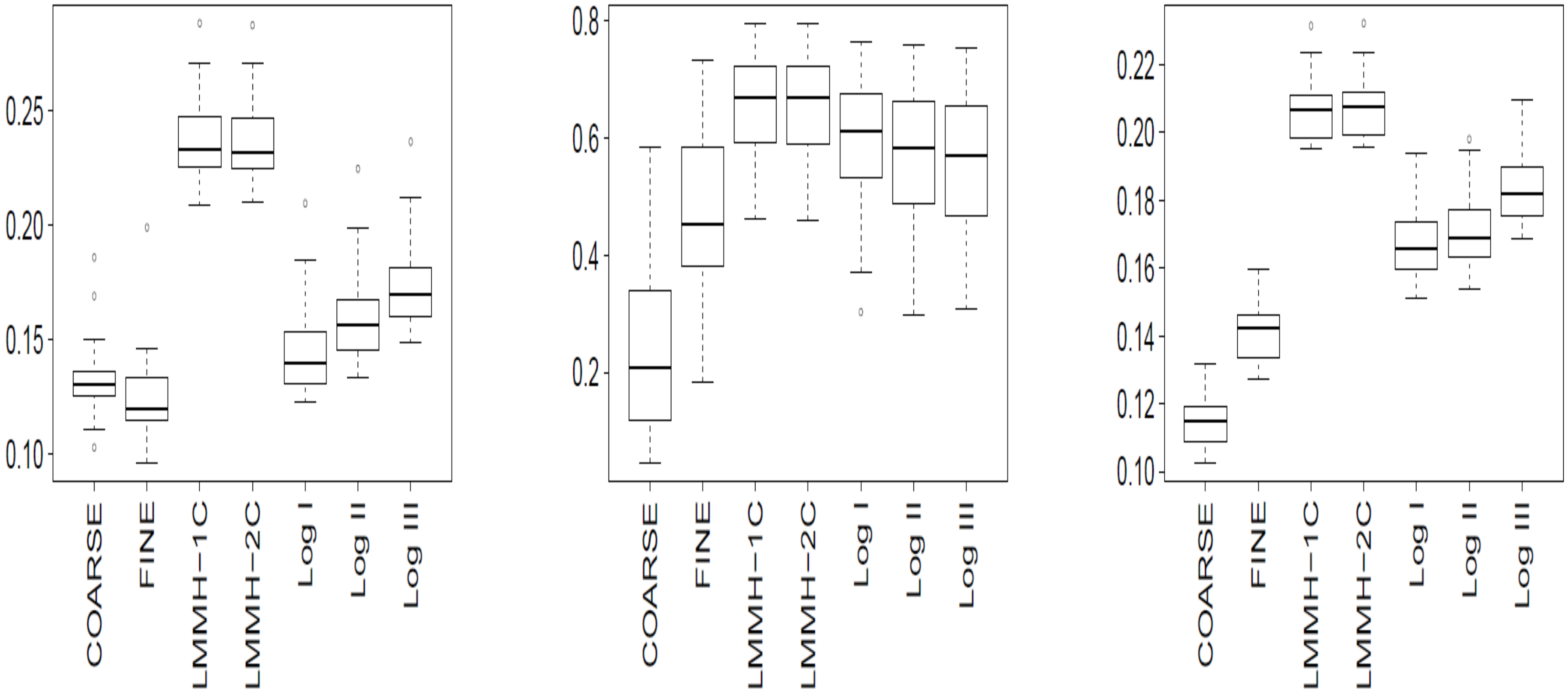
LMMH (Agarwal et. Al: ACM SIGKDD 2010)

- 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

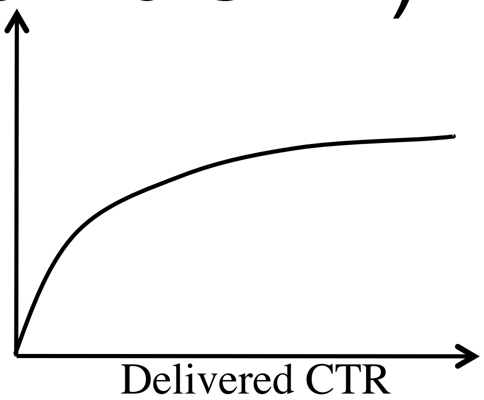
What if prediction still goes wrong?

- 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

Questions

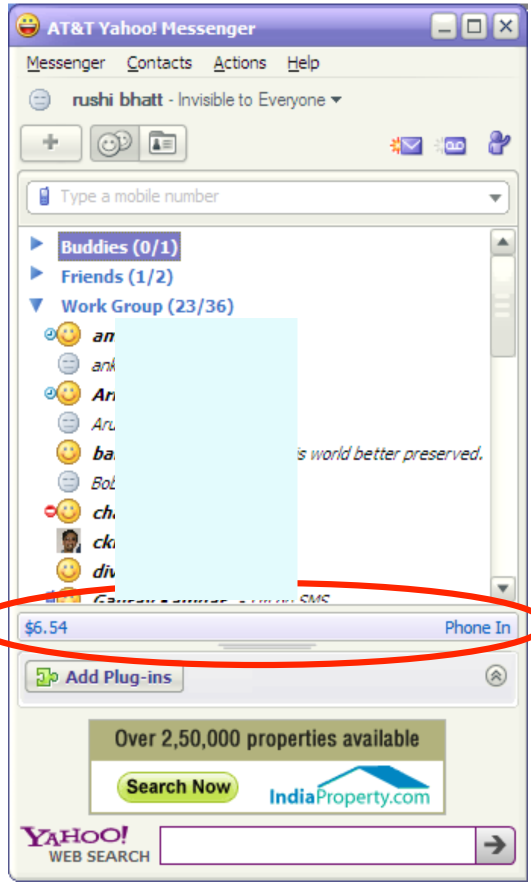
- What are the correlates of adoption spread?
  - Are high-degree users “different”?
  - Are neighborhoods predictive?
- How do we improve uptake?
  - Do social attributes provide a lift?
  - Is *neighborhood targeting* a good idea?

Spread of Premium Service Adoption Through a Network

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

Data Sources

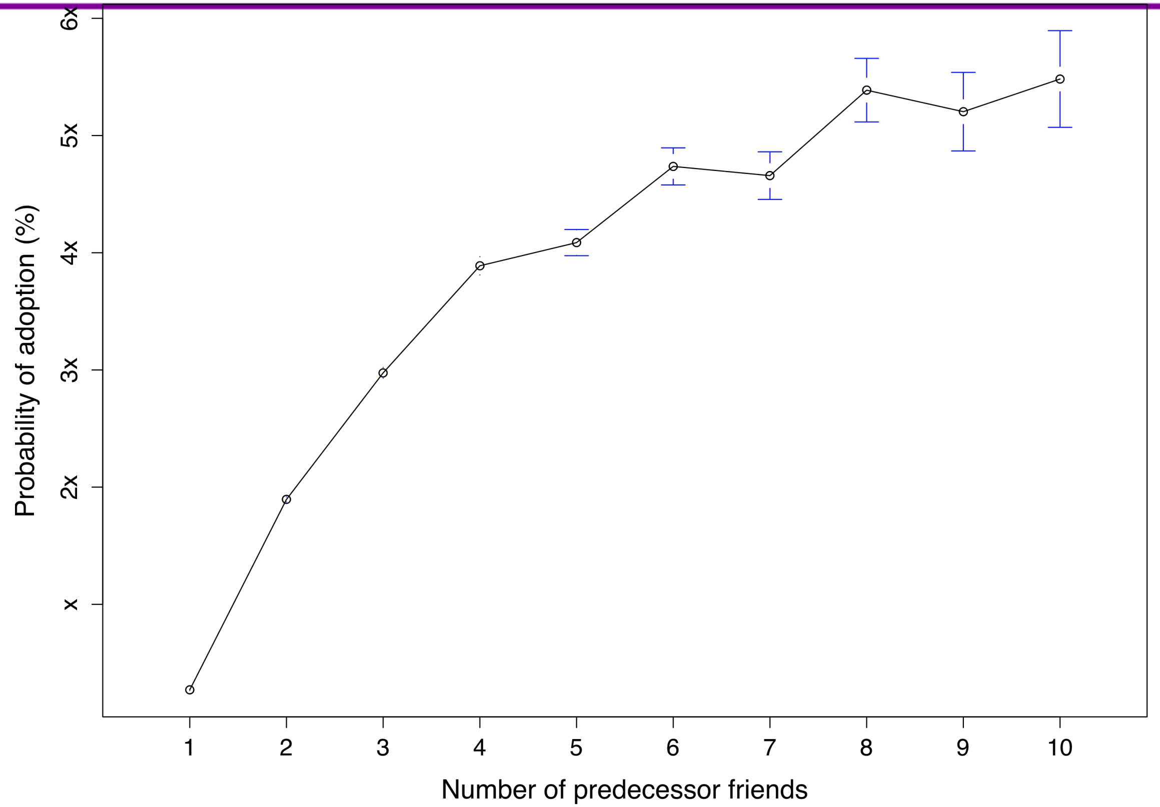
PC to Phone
Premium Service



- 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

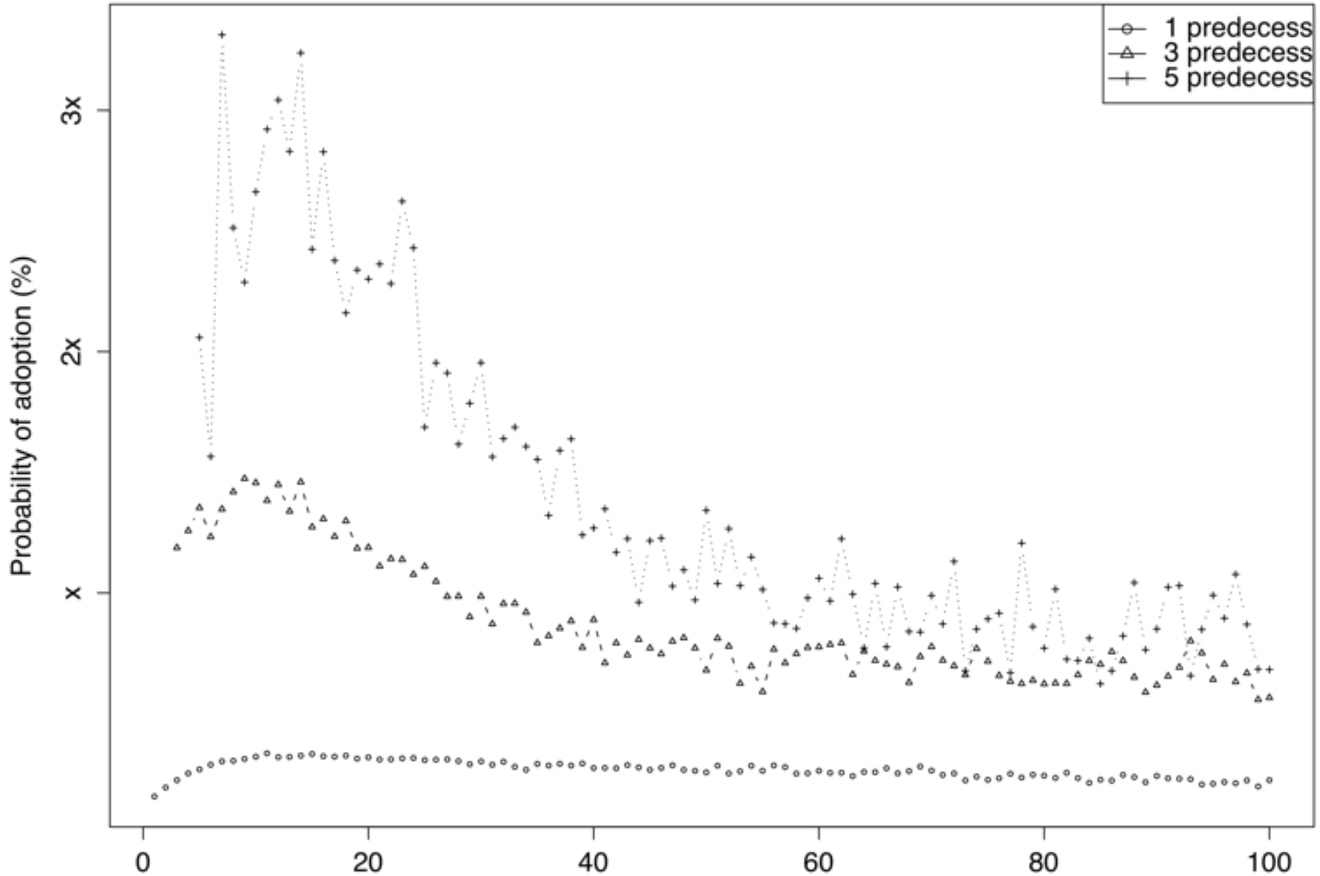
Low baseline adoption rates

Social Network Effect Exists



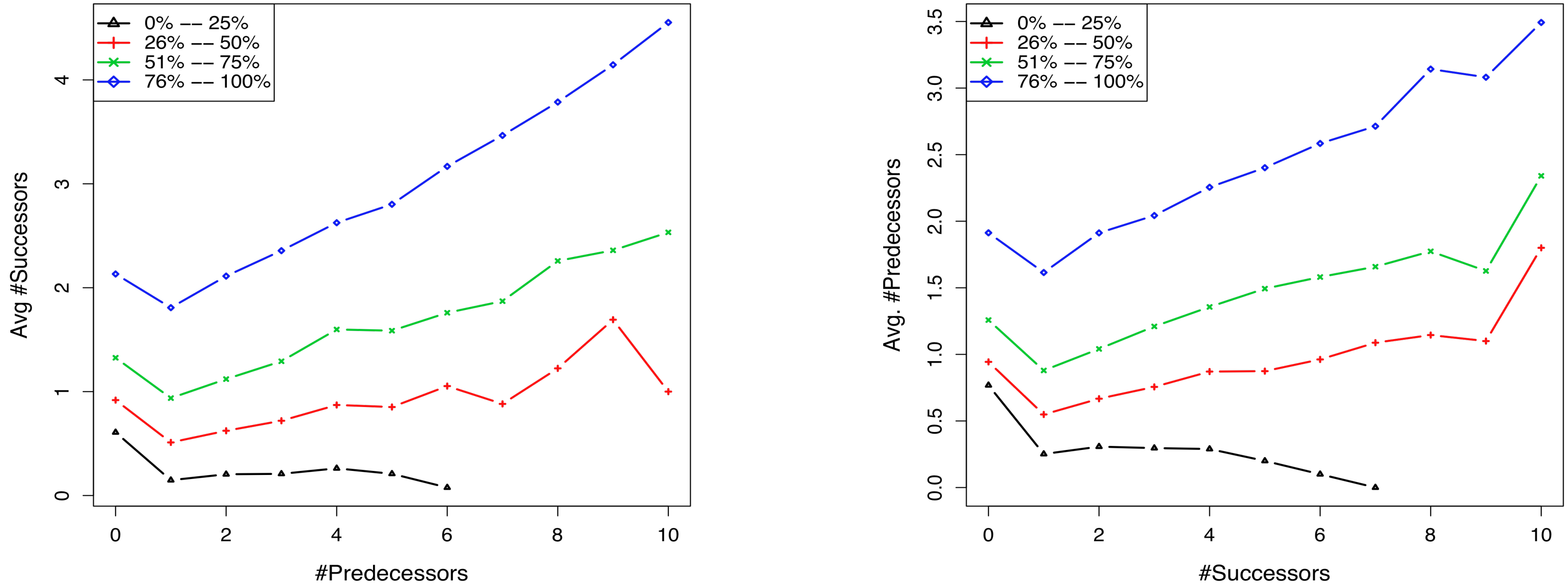
More friends adopting improves chances of adoption

High-degree users are harder to convert



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

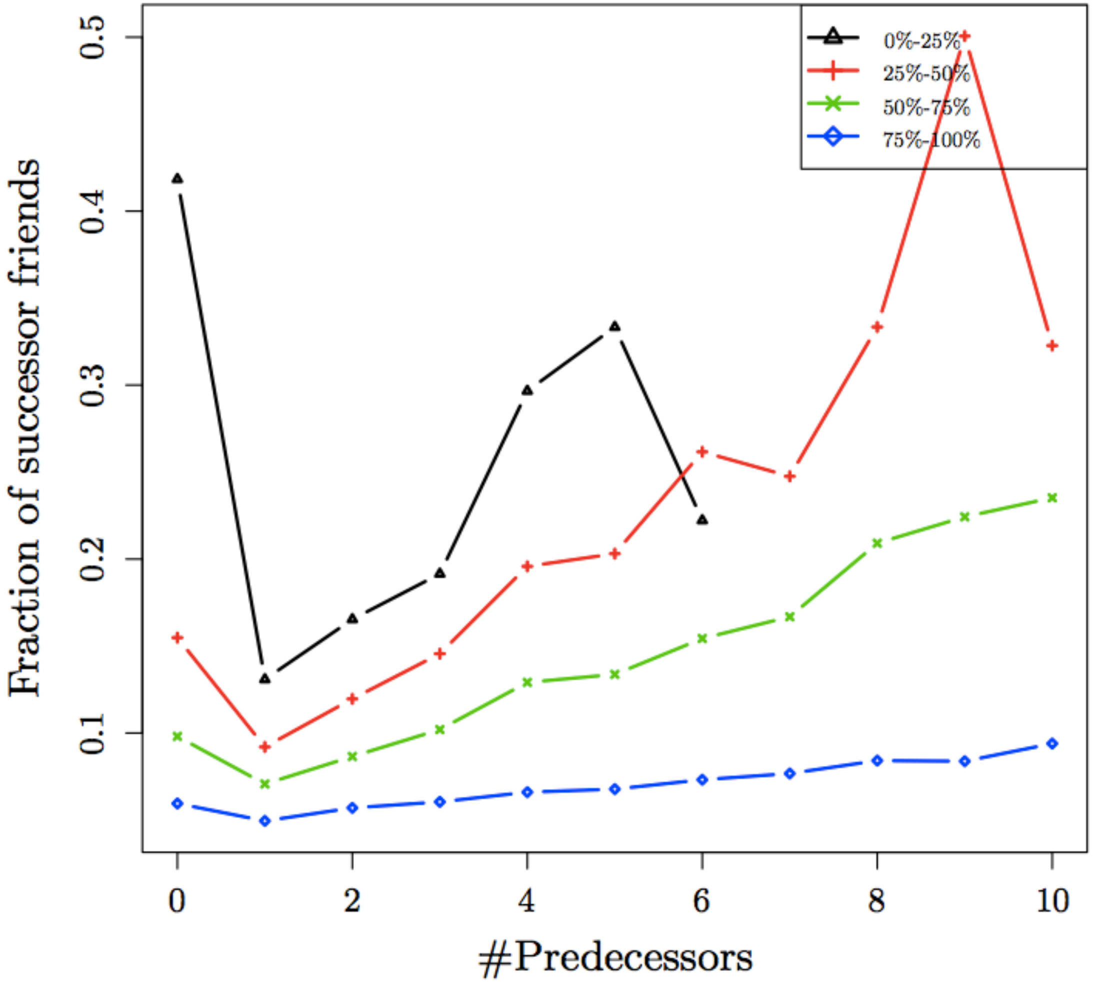
Are High Degree Users Influencers?



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

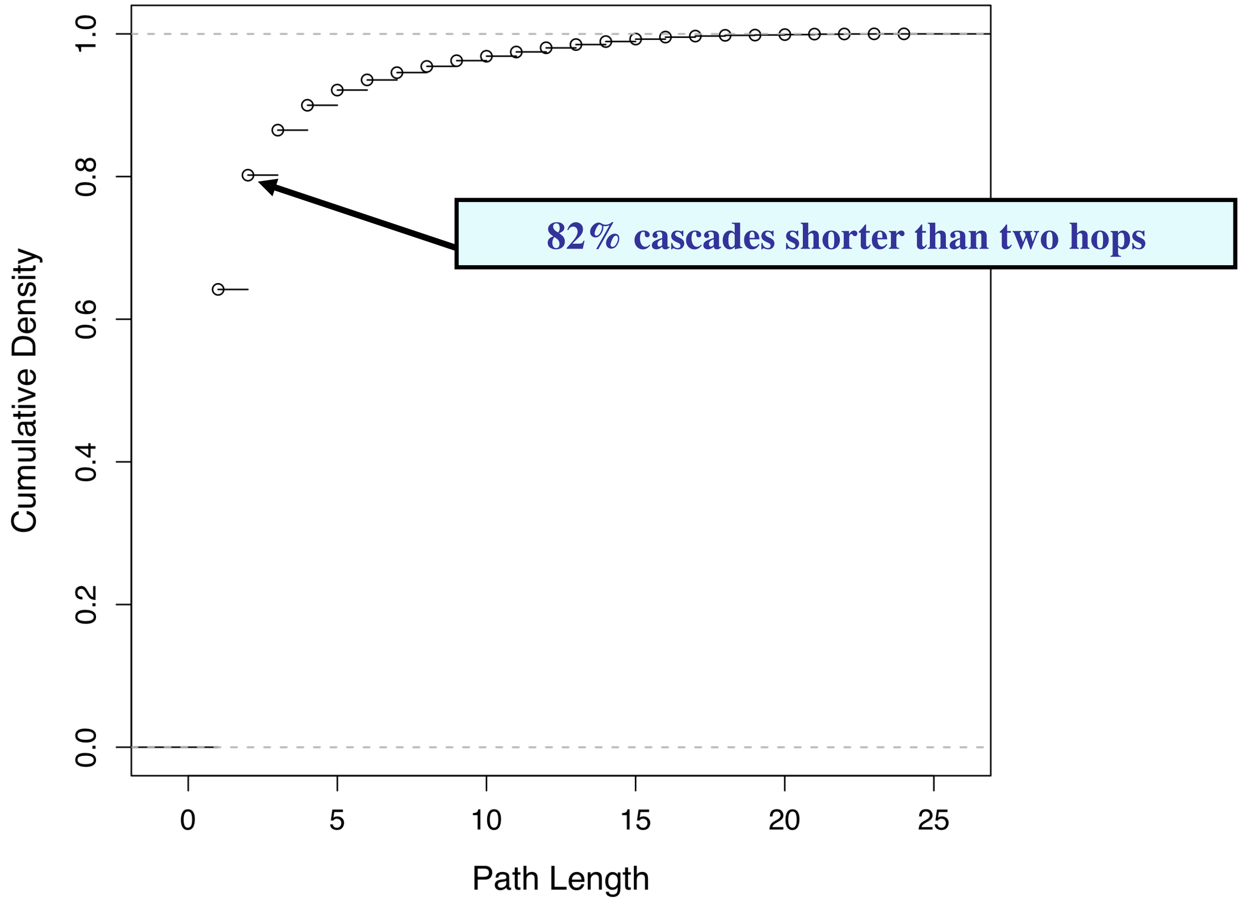
Also, Anagnostopoulos et al. '08, Dodds & Watts '07

High Degree Users "Underperform"



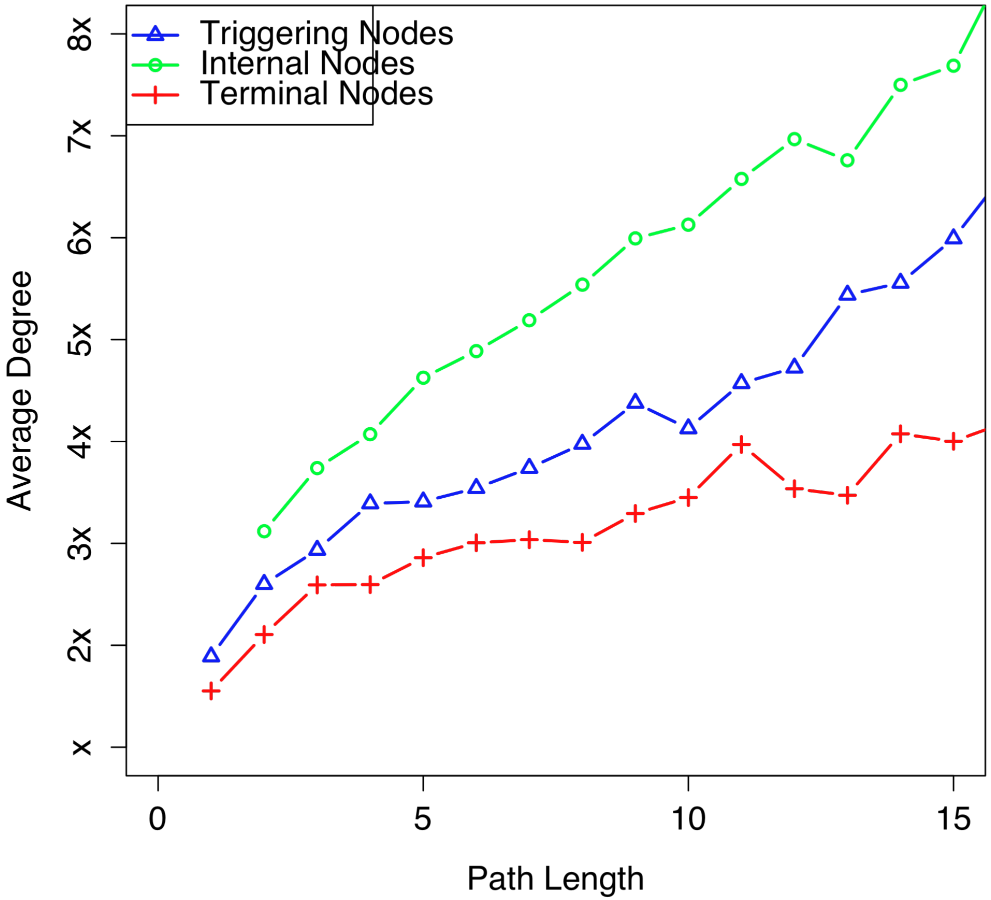
Successors per user reached lower for high degree users [Dodds & Watts '07, also Kitsak et al. '10]

Adoption Spread is Mostly Local Influence is not Far Reaching



Longer cascades do exist, but very few

Adoption Spread by Internal Nodes



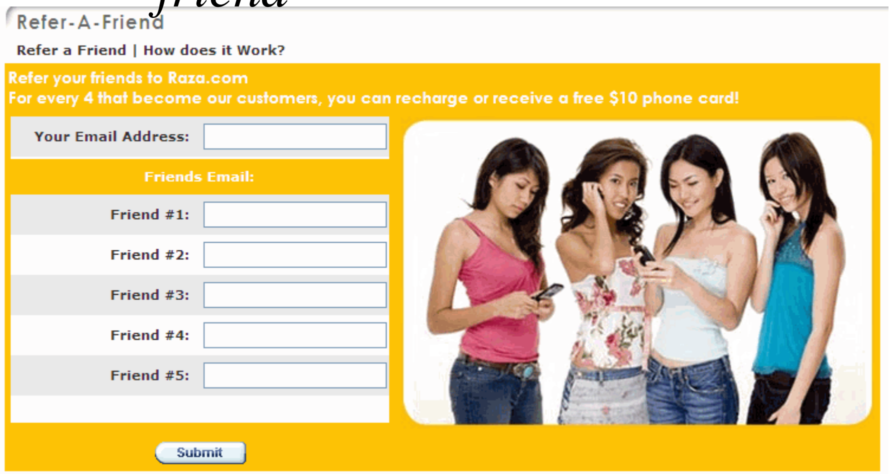
Longer cascades triggered by high degree users
Intermediate nodes sustain the cascades

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*"



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

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

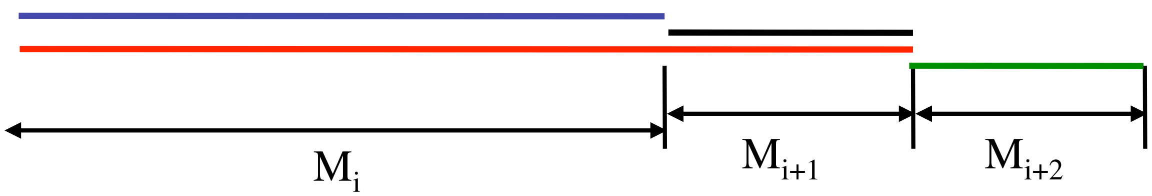
Demographic
- Age
- Gender

Geographic
- Originating Country

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

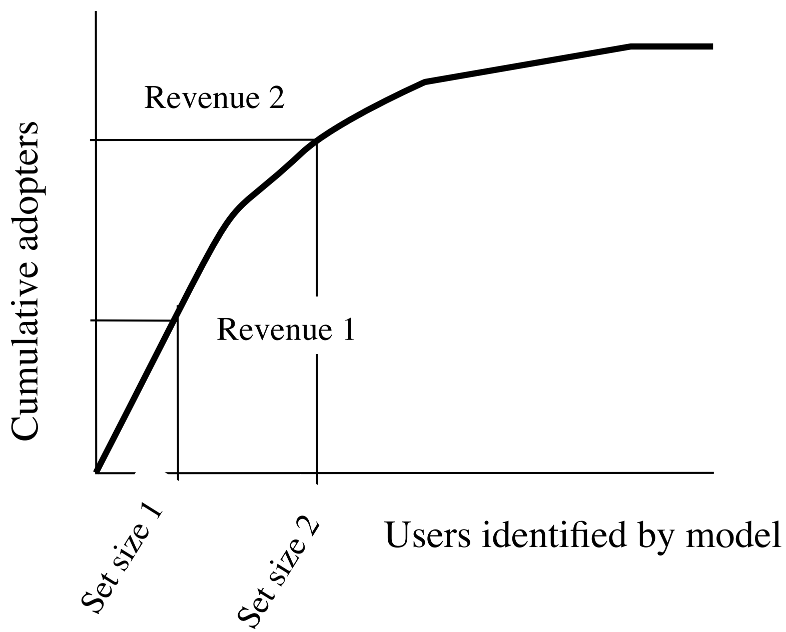
Training Period:
User features till M_i
Adoption in M_{i+1}

Testing Period:
User features till M_{i+1}
Adoption in M_{i+2}

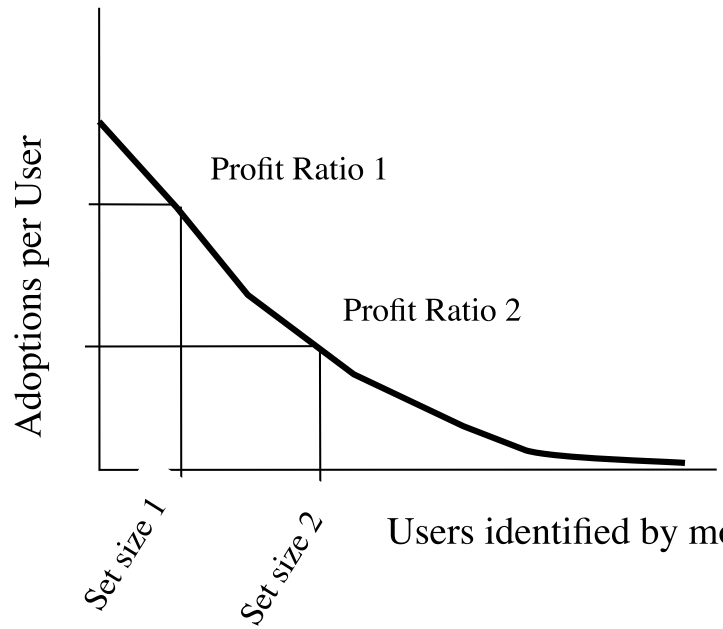


Model Evaluation

- 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



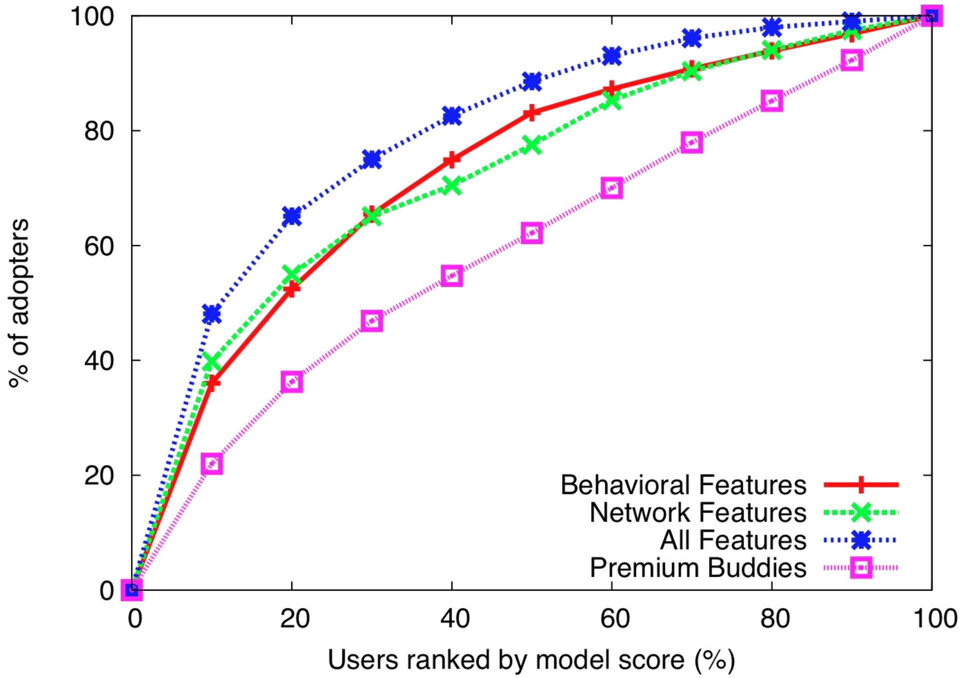
- Metric 2: Adoptions per user
  - Good when we have the highest adoptions per targeted user
  - Factors in the cost of targeting to each user
  - Helps to decide the *right* incentives for each targeted user



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

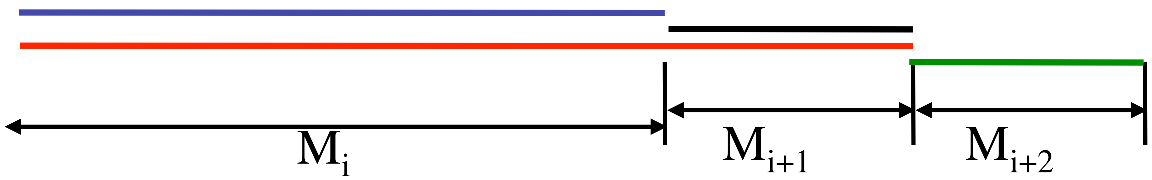
- Combined user and behavioral features best for direct targeting [Also, Hill et al. '06]

Social Targeting

- Given a set A of adopters
  - Define *σ(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 *σ(A)*
- Approach
  - Train a regression model (Gradient Boosted Decision Tree) to predict *σ(A)*
  - Rank order adopters in descending order of *σ(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}

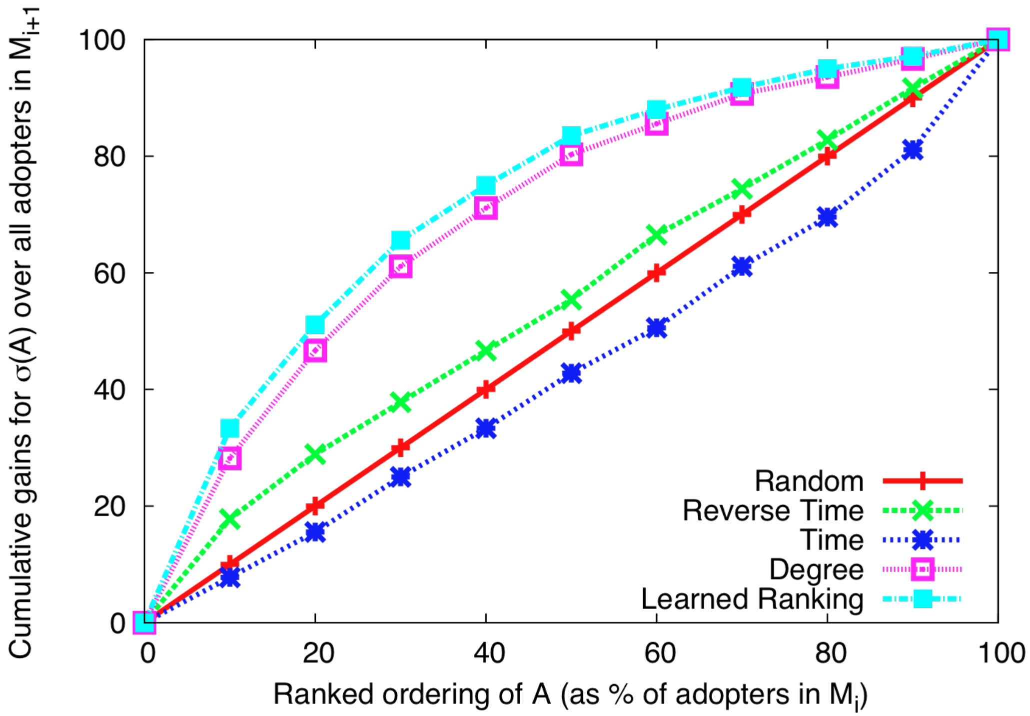


Methods Compared

- 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

Q&A


AS THE MARKETING MANAGER FOR SOCIAL MEDIA, MY JOB IS TO USE THESE TWO WORDS A LOT.
FACEBOOK
TWITTER
Dilbert.com DilbertCartoonist@gmail.com
MARKETING THROUGH SOCIAL MEDIA IS LIKE HERDING CATS. AND JUST TO MAKE IT INTEREST-ING, MANY OF THE CATS ARE DRUNK AND STUPID.
9-14-10 © 2010 Scott Adams, Inc./Dist. by UFS, Inc.
BURN.
I AM TOTALLY DEFRIENDING THAT WITCH.

Thank you!

Ad Selection with a Chunked Price Model

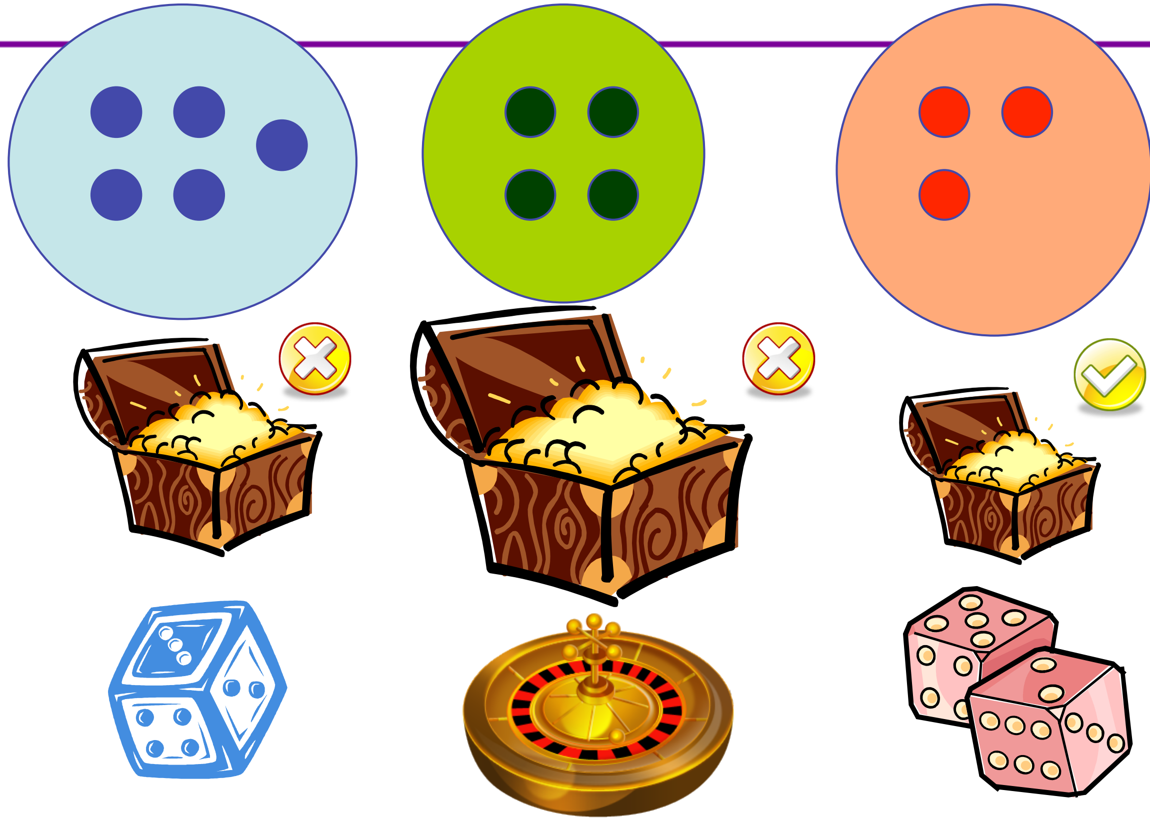
Narayan Bhamidipati, Rushi Bhatt, Michael Grabchak (Cornell)

Chunked Price Model: The Setup

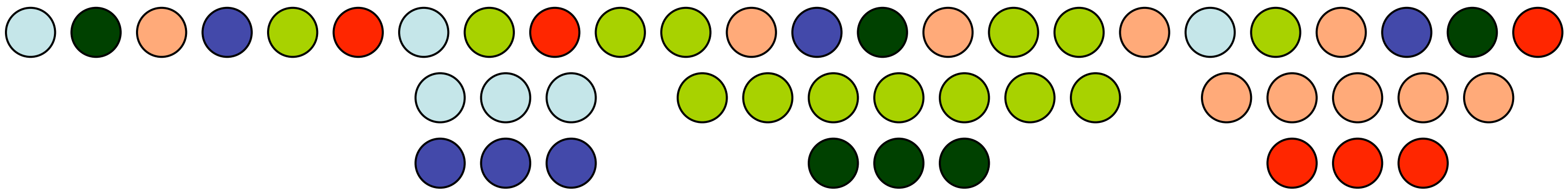
Goal:
#successes needed

Reward:
Payout on achieving goal

Probability:
of success in an attempt



Time:
#attempts



Objective: Revenue Maximization

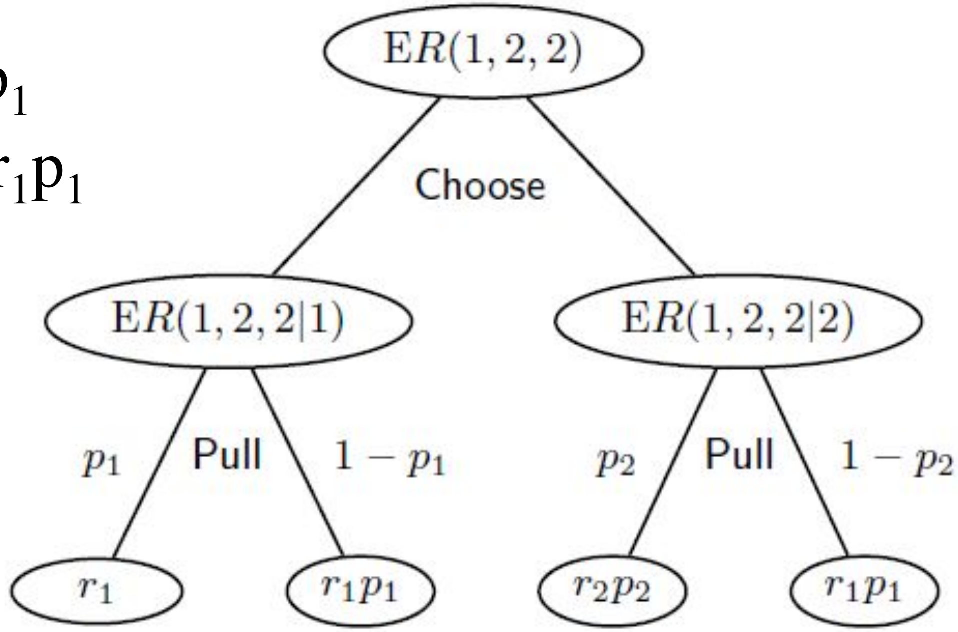
- At each time $1 \le t \le T$
  - decide which ad to show
  - observe the outcome (click or no click)
  - revise the goals
  - repeat
- In such a way that total (expected) revenue is maximized at time T
- Obviously, once a goal for an ad is met, that ad is never shown again

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

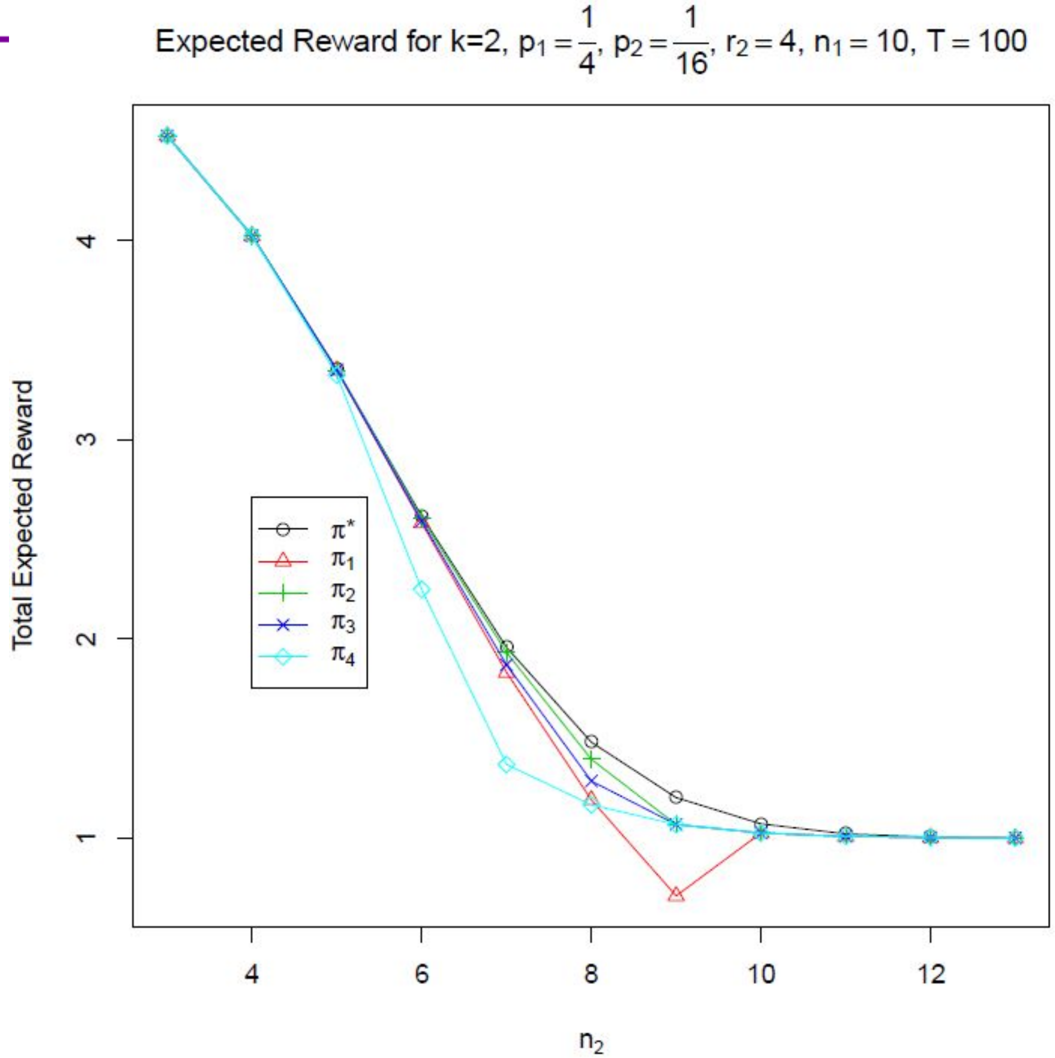
- For each ad, define an index that
  - Considers only ads with unattained but attainable goals
  - Increases with p_i and r_i, decreases with n_i.
  - Is a scalar multiple of r_i.
- Some indices:
  - $r_i p_i / n_i$: continues showing the same ad
  - $r_i p_i / n_i$ $P(n_i \text{ in } T)$: depends on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

Theoretical Guarantees

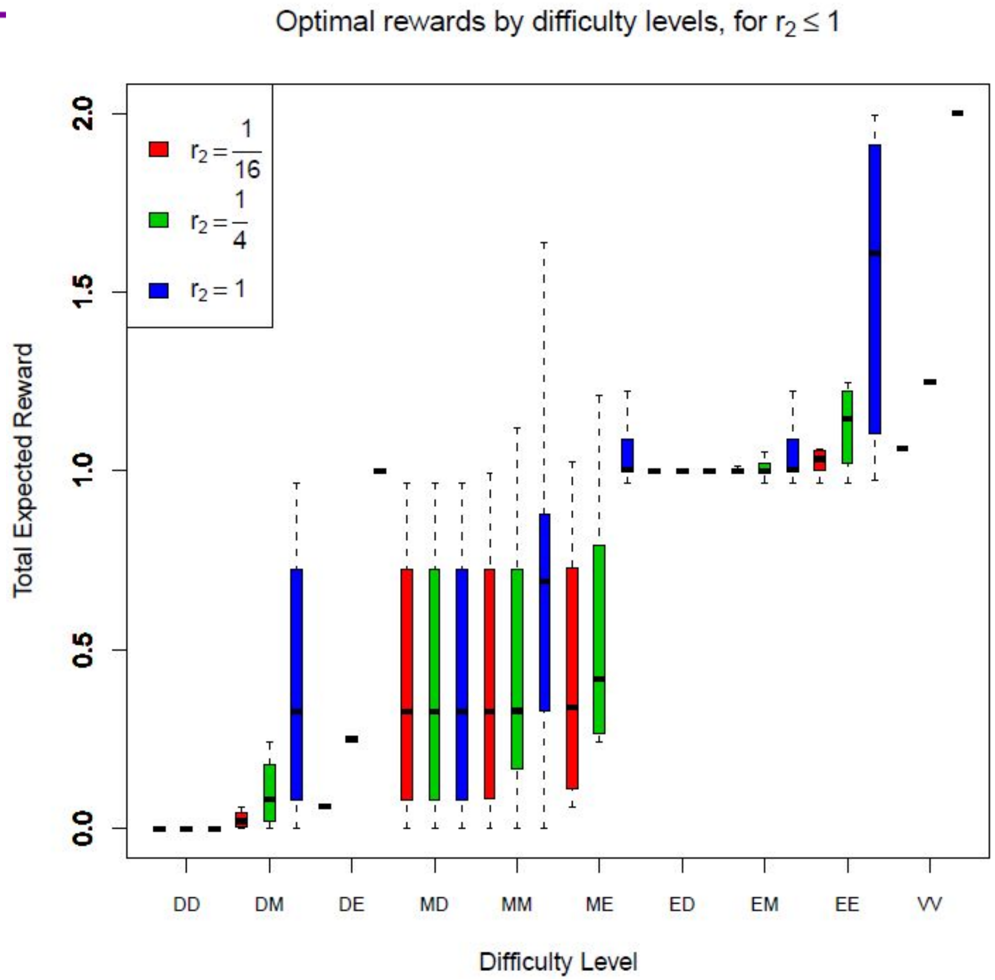
- Make use of the parallels with Stochastic Knapsack Framework
  - Items = ads,
  - random weight W_i = #attempts to attain goal
- Introduce artificial policies like
  - γ_0 = always show the ad with max exp reward
  - γ_7 = choose the better of γ_0 and $r_i p_i / n_i$.
- Obtain a 3-approximation
  - under mild assumptions: $P(n_i \text{ in } T) \geq ½$

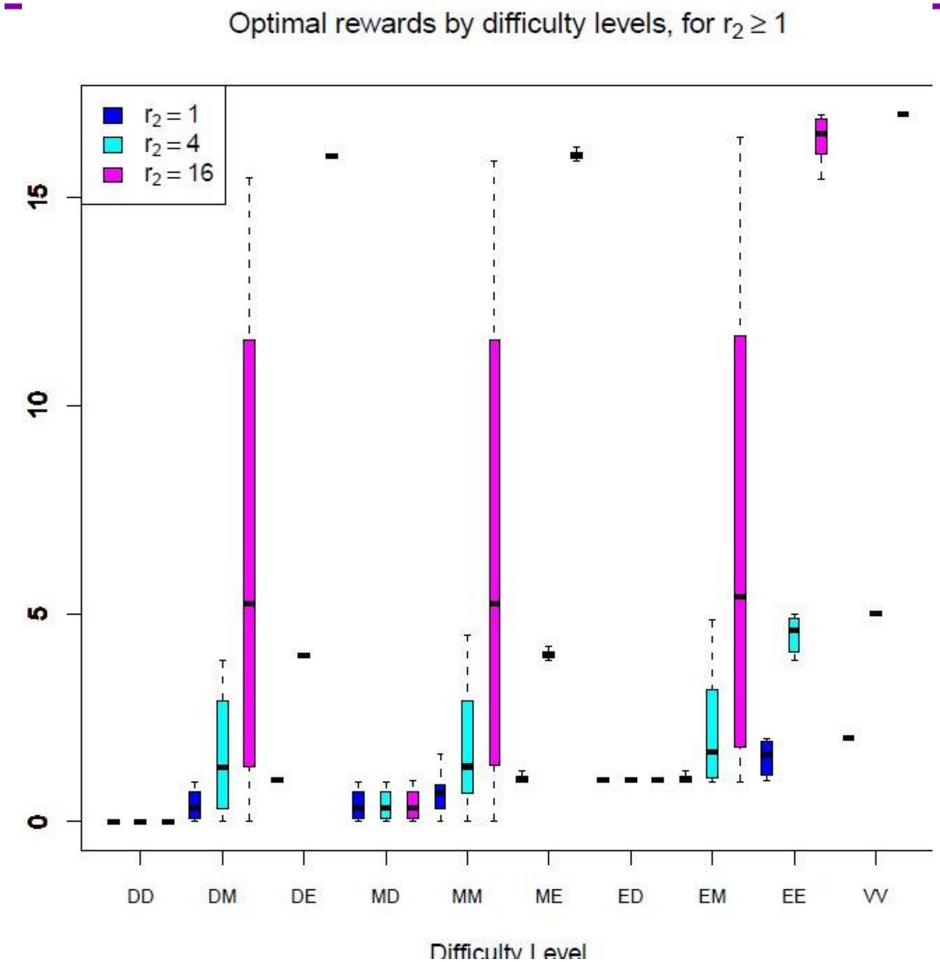
Optimal vs. Greedy Performance

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



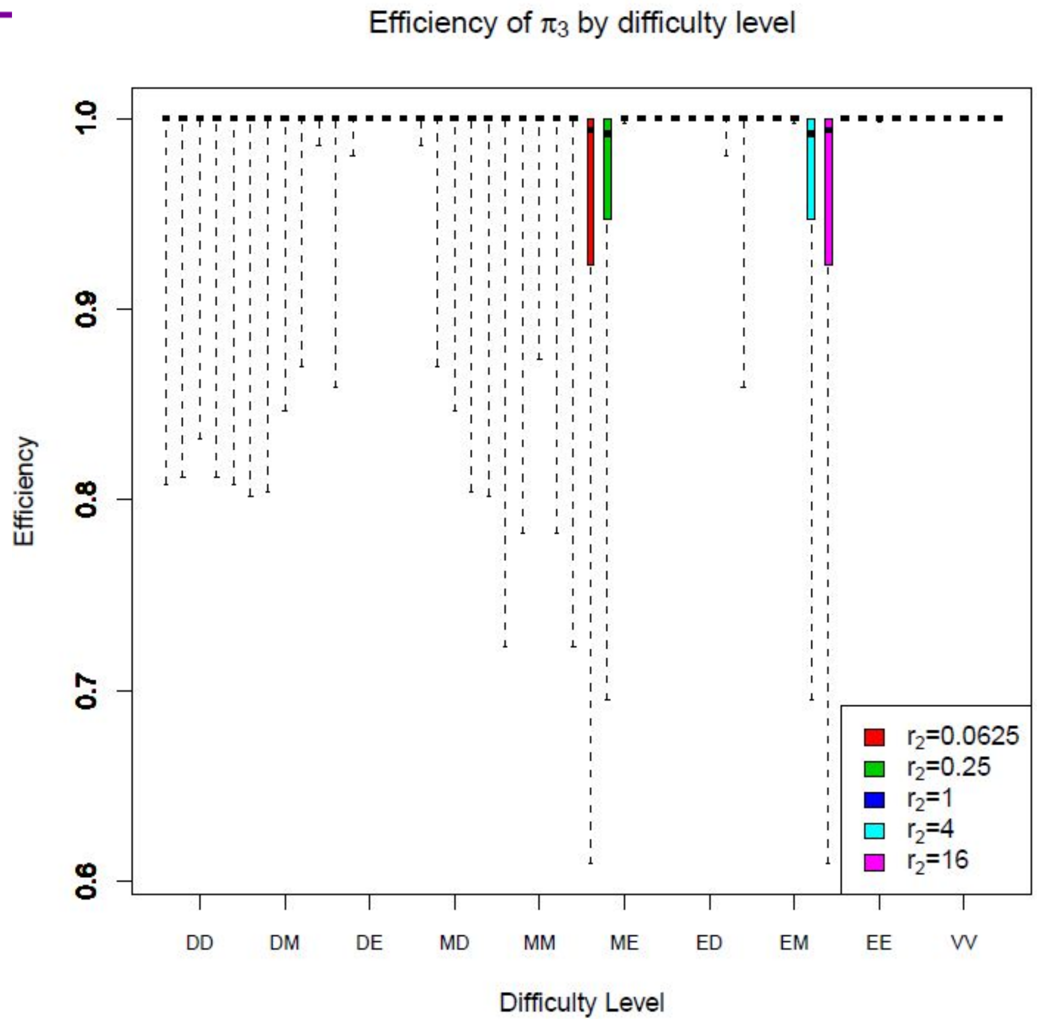
Complete Enumeration: Optimal





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