Introduction to Computational Advertising

MS&E 239

Stanford University

Autumn 2011

Instructors: Dr. Andrei Broder and Dr. Vanja Josifovski

Yahoo! Research

General course info

- Course Website: http://www.stanford.edu/class/msande239/
- Instructors
  - **Dr. Andrei Broder**, Yahoo! Research, broder@yahoo-inc.com
  - **Dr. Vanja Josifovski**, Yahoo! Research, vanjaj@yahoo-inc.com
- TA: **Krishnamurthy Iyer**
  - Office hours: Tuesdays 6:00pm-7:30pm, Huang
- Course email lists
  - Staff: msande239-aut1112-staff
  - All: msande239-aut1112-students
  - Please use the staff list to communicate with the staff
- Lectures: 10am ~ 12:30pm Fridays in HP
- Office Hours:
  - After class and by appointment
  - Andrei and Vanja will be on campus for 2 times each to meet and discuss with students. Feel free to come and chat about even issues that go beyond the class.

Course Overview (subject to change)

1. 09/30 Overview and Introduction
2. 10/07 Marketplace and Economics
3. 10/14 Textual Advertising 1: Sponsored Search
4. 10/21 Textual Advertising 2: Contextual Advertising
5. 10/28 Display Advertising 1
6. 11/04 Display Advertising 2
7. 11/11 Targeting
8. 11/18 Recommender Systems
9. 12/02 Mobile, Video and other Emerging Formats
10. 12/09 Project Presentations

Lecture 06: Display Advertising Part 2

Based on

A. David Hallerman, Buying Display Ad Inventory, E-Marketeer Webinar, Aug 25. 201
B. Agarwal, Kota, Agrawal, Khanna: *Estimating Rates of Rare Events with Multiple Hierarchies through Scalable Log-linear Models, KDD 2010*
C. Slideware from many sources
D. Help from many colleagues in particular, Jai Shanmugasundaram

Disclaimers

- This talk presents the opinions of the authors. It does not necessarily reflect the views of Yahoo! inc or any other entity.
- Algorithms, techniques, features, etc mentioned here might or might not be in use by Yahoo! or any other company.
- These lectures benefitted from the contributions of many colleagues and co-authors at Yahoo! and elsewhere. Their help is gratefully acknowledged.

Lecture plan for today

- Review of last lecture
- Non guaranteed display advertising
- Presentation on Mobile advertising [15 min incl discussion]
- Prediction technology

Guaranteed vs.. Non-Guaranteed

- Advertiser can buy the ad space
  1. in advance (1-24 months) as GD
     - Pay a premium
     - Get premium inventory
     - Many targeting attributes
  2. on the spot market (at the time of page view) as NGD

Why GD?

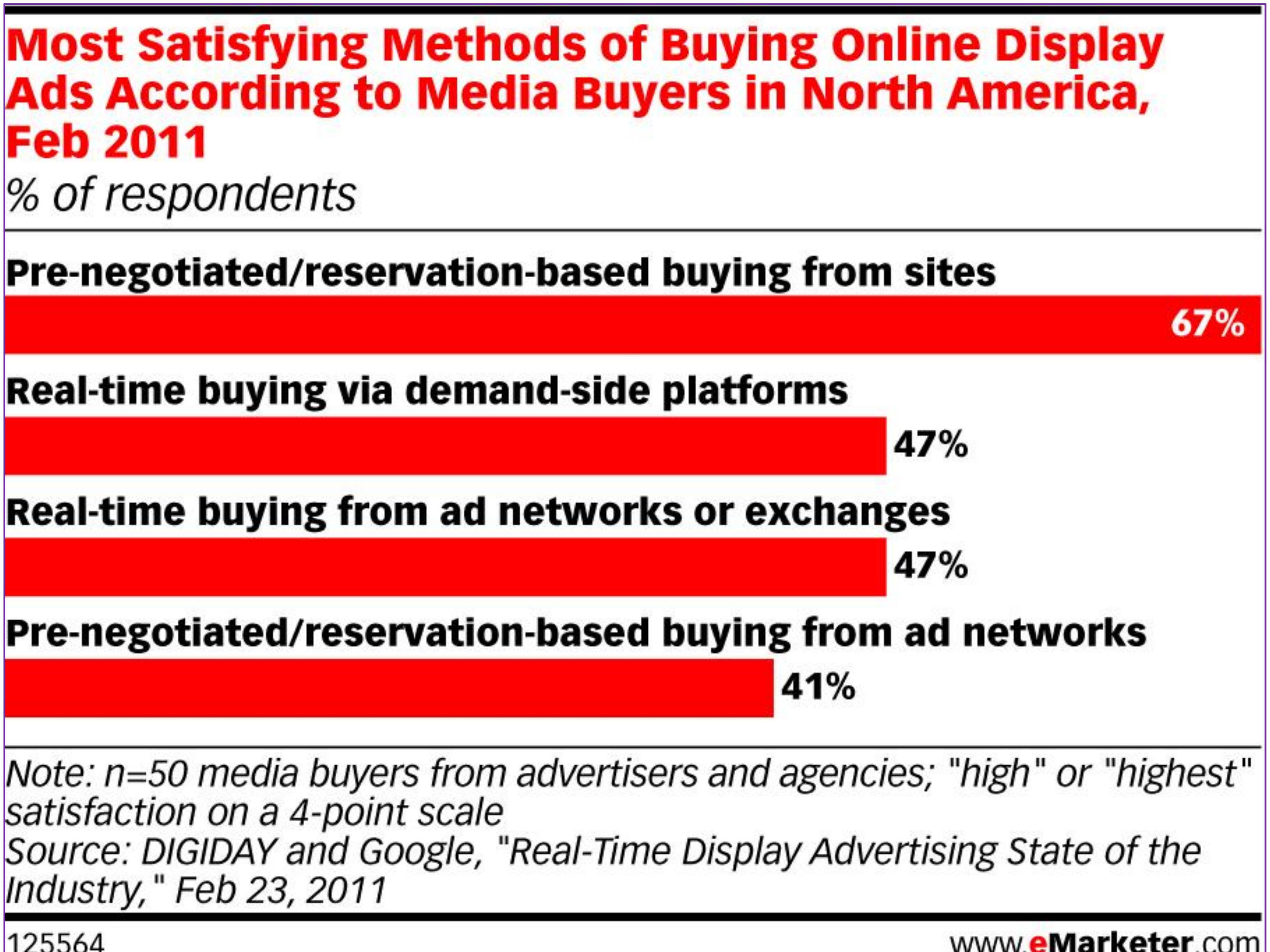
- Currently, GD commands premium prices and is also known as "premium inventory" or "direct buy"
- Bought directly from publisher, via a sales team rather than automated methods
- Advertisers get maximum control over timing, position, property, etc
  - E.g. GM knows will launch a new model targeted to single, young, irresponsible males in Nov 2011…
  - Can buy an exclusive (no competing ads)
- In general: quality of inventory (page views) given to GD contracts > non-guaranteed
  - We have seen how to trade-off quality given to GD vs. revenue

Key points of last lecture

Display advertising

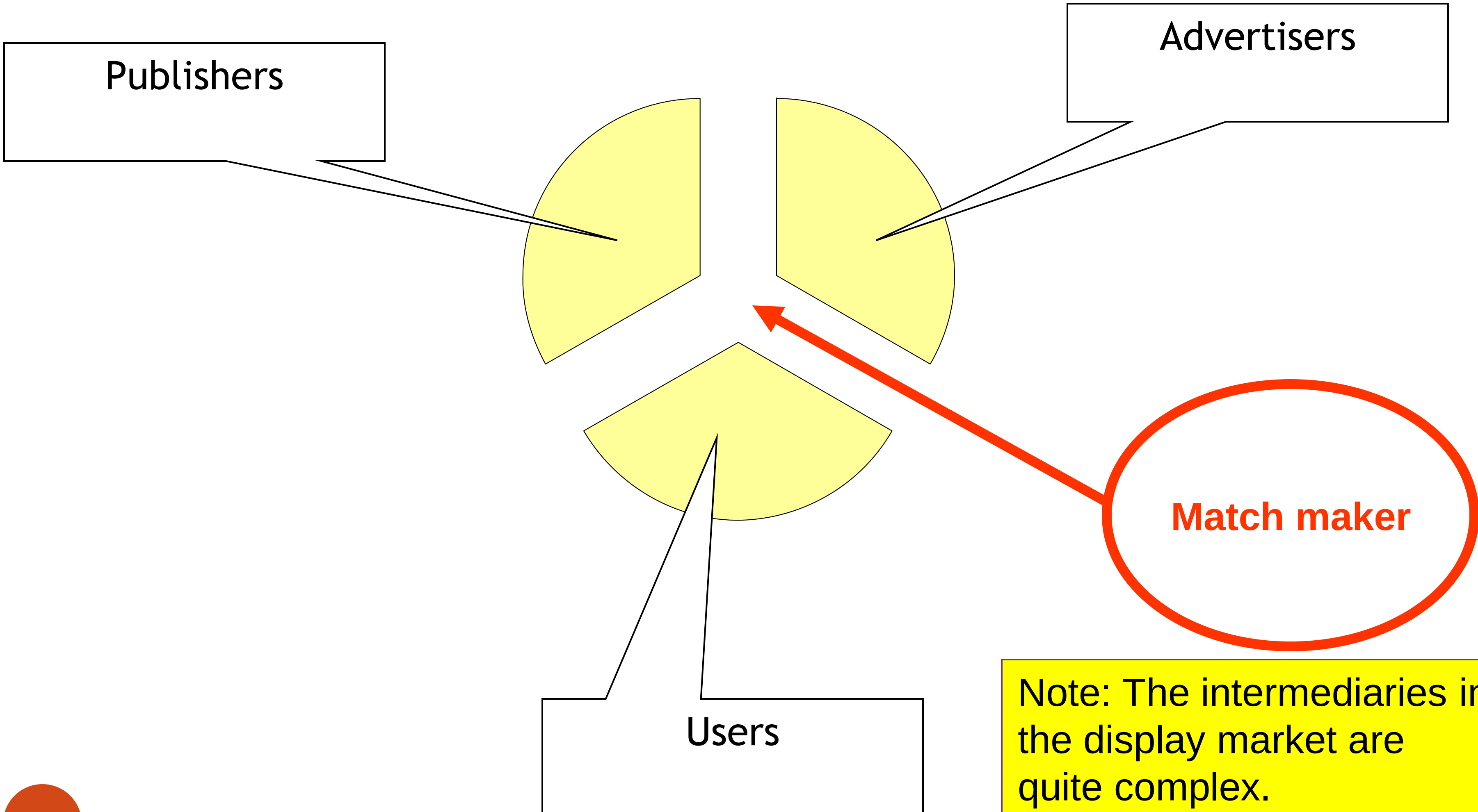
- Complex optimization problem – a lot more math than you might suspect
- Interplay of forecasting, optimization, economics
- Need to have solutions for:
  1. Forecast supply, demand, NGD pricing
  2. Admission control
  3. Pricing
  4. Optimal allocation of impressions to contracts
  5. Ad serving

GD is considered the most desirable inventory



The NGD Marketplace

The three actors: Publishers, Advertisers, Users, & “Match maker”



Note: The intermediaries in the display market are quite complex.

Advertisers: basic principles

- Display advertisers aim to
  - **Reach** audiences of interest with certain frequency
  - Achieve certain **performance** of the campaign

- In marketing terms, display advertising aims for both
  - Brand marketing that raises the awareness for a brand
  - Direct marketing

- For both reach and performance campaigns, the key challenge is to select the right audience, i.e. to **target** the right users
  - Many advertisers have multiple simultaneous campaigns with different goals

- In all cases, ultimate goal is to maximize **ROI (Return on Investment)** or **ROAS (Return On Advertising Spend)**

Advertiser Process

Pick a Marketing Objective

Define Audience:
Demos, Geo, Behaviors, Interests, Purchasing

Develop Creative:
Video, Rich Media, Large Canvas, Banners

Reach Audience:
Guaranteed (GD) or Opportunistic (NGD)

Measure Results

Key Focus: Maximize Return on Ad Spend

Publisher Process

1. Develop Content
2. Attract, Retain Audience
3. Package Content and Audience
4. Monetize: Direct Sale
   Exclusives, Sponsorships, Guarantees, Price
5. Monetize: Indirect Sale
   Ad Networks, Exchanges

Key Focus: Yield while controlling the user experience

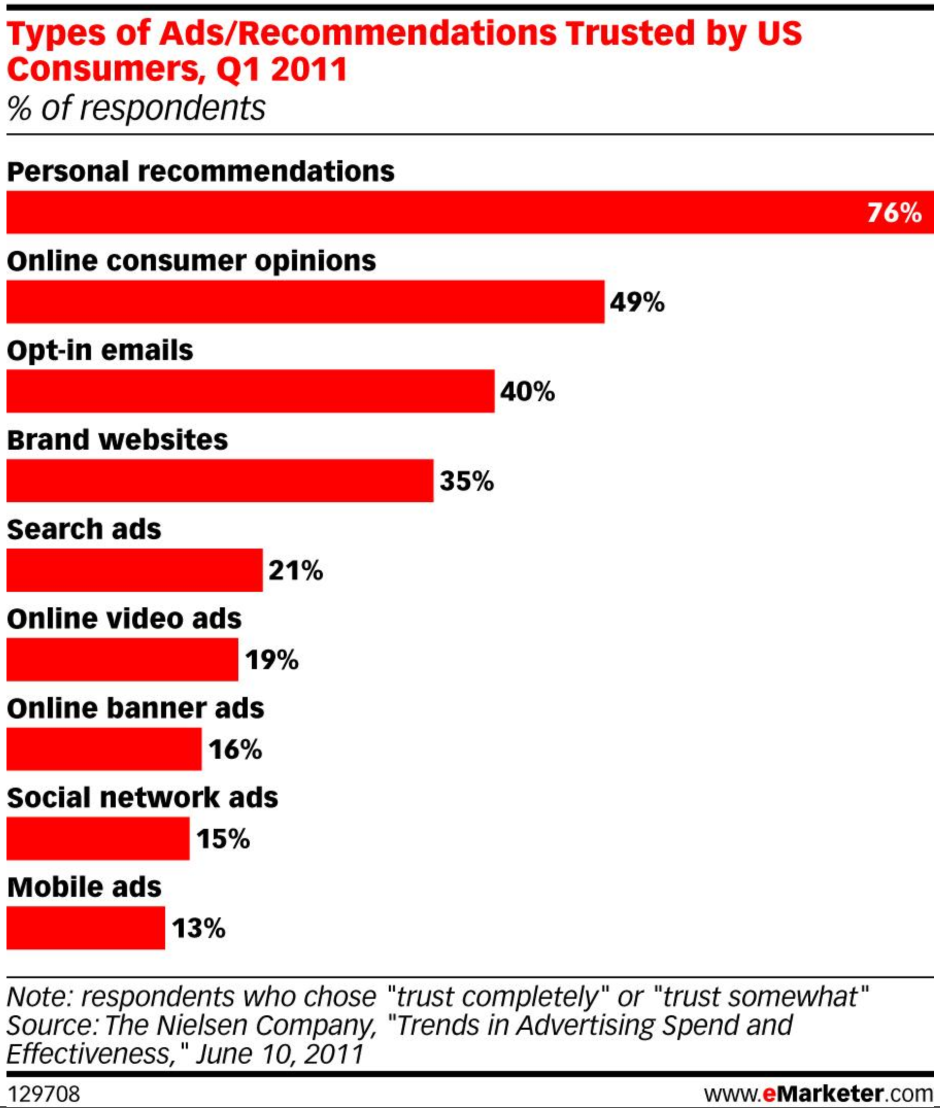
Publishers

- Own and operate the site
- Some impressions more valuable than others – price determined by supply and demand:
  - More competition for "females, 30-50, high income" than for "teenager drop-outs"
  - Also reflected in property "Yahoo horoscopes" less valuable than "Yahoo finance"

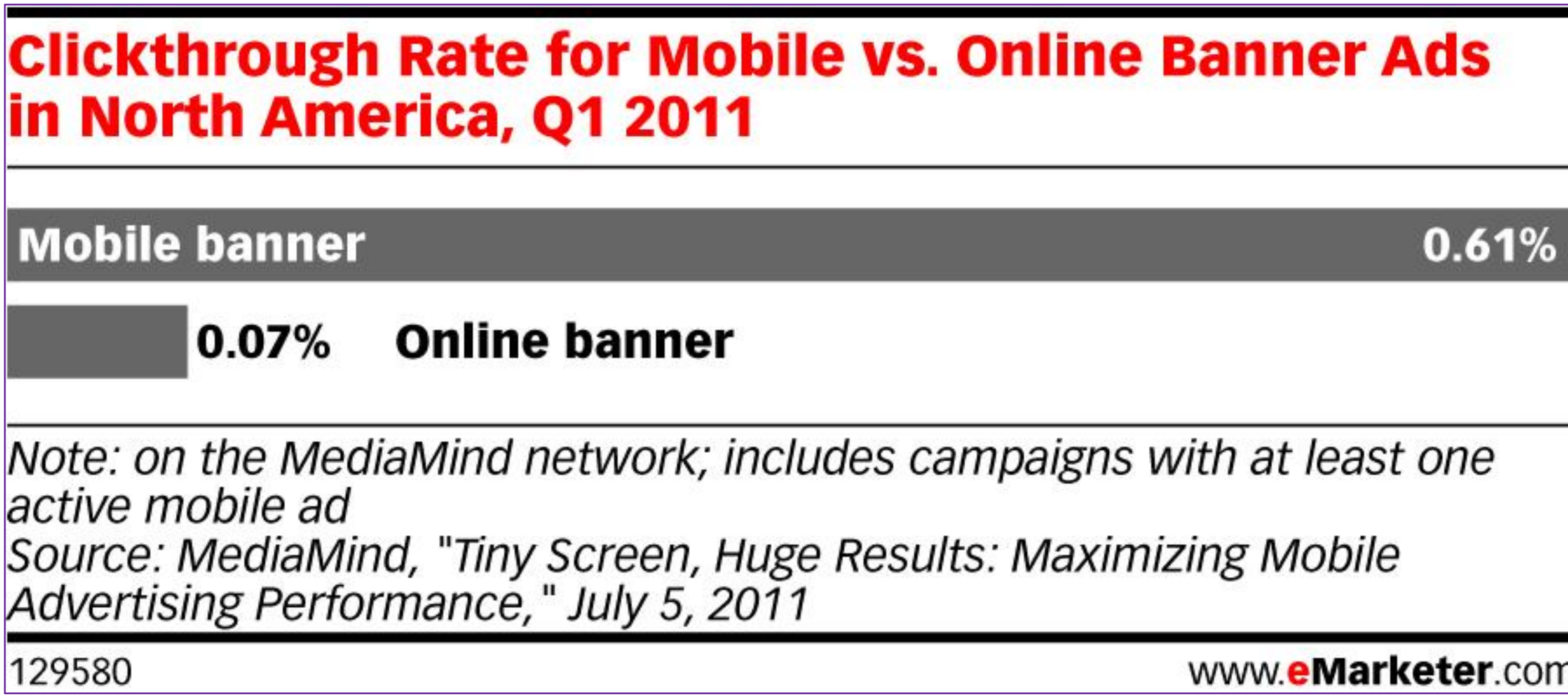
- Sell a portion of the impressions as **premium** e.g. $15 CPM
  - Usually sold on a "guaranteed" base by a publisher's direct sales force
- Rest sold as "**remnant**", "**network**", etc e.g. <$2 CPM
  - Non-guaranteed inventory usually sold via intermediaries

- Maximizing long term revenue is the primary goal
- Want control of:
  - Pricing
  - Targeting information
  - Supply
  - User experience
- Some big publishers: Yahoo!, Facebook, MSN, AOL, etc

Challenge: display advertising is not perceived very trustworthy



Very low CTR


Clickthrough Rate for Mobile vs. Online Banner Ads in North America, Q1 2011
Mobile banner
0.61%
0.07%
Online banner
Note: on the MediaMind network; includes campaigns with at least one active mobile ad
Source: MediaMind, "Tiny Screen, Huge Results: Maximizing Mobile Advertising Performance," July 5, 2011
129580
www.eMarketer.com

On the other hand online ads have a considerable off-line effect!

- See:

  “Does Retail Advertising Work? Measuring the Effects of Advertising on Sales via a Controlled Experiment on Yahoo!” by Randall A. Lewis and David H. Reiley

  http://www.davidreiley.com/papers/DoesRetailAdvertisingWork.pdf

Intermediaries

- Dozens of different companies with different business models
  - Connect the advertisers and the publishers
  - Optimize the connections
  - Package impressions into audiences
  - Match the ads with the users and the context
  - Increase the fluidity of the market
  - Charge for the added value
- Etc

Intermediation & disintermediation

- Intermediation
  - Reduces friction
  - Ads value
  - Decreases transparency – what exactly are you buying?
  - E.g. Buy a house using a real estate agent – you pay but get advice, selection
- Disintermediation
  - Reduces costs
  - Increases transparency
  - E.g. Buy a house on your own

Methods Used by US Ad Agencies to Buy Online Advertising, Q1 & Q2 2011

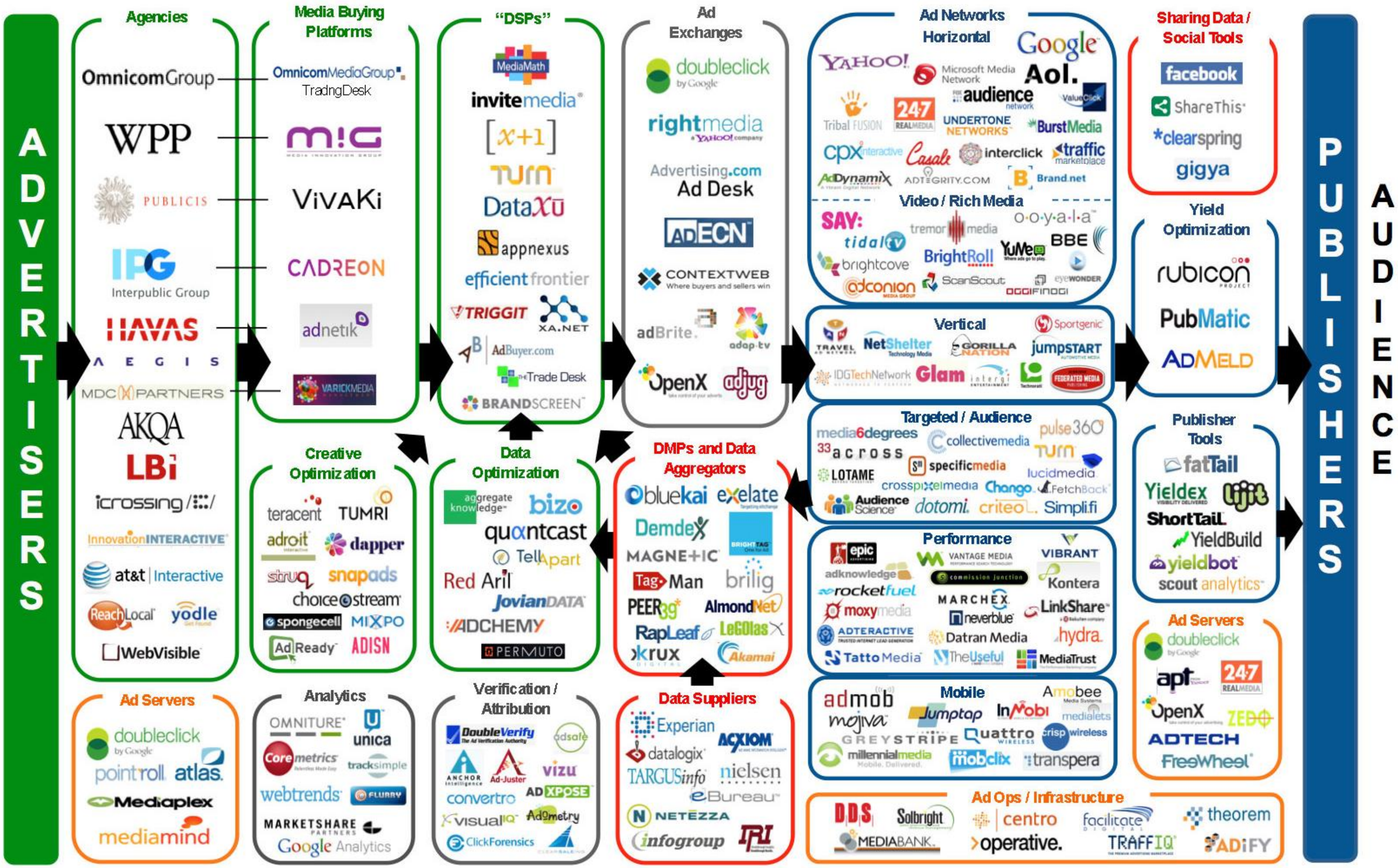
% of respondents

	Q1 2011	Q2 2011
Through an ad network (Google, Yahoo!, Yellowbook, etc.)	66.0%	67.6%
Through traditional media (TV, print, radio stations, etc.)	69.1%	50.9%
Direct from the publisher (Expedia, OpenTable, NCAA, etc.)	41.5%	43.5%
Through a DSP or exchange (Invite Media, Right Media, Acxiom, etc.)	10.6%	19.4%
Through self-services (FatTail's PageGage, Adap.tv's Marketplace, etc.)	2.1%	3.7%
Other	6.4%	8.3%

Source: STRATA, "2nd Quarter 2011 Survey Results," July 26, 2011

130575 www.eMarketer.com

Display advertising ecosystem (Luma Partners, 2011)



Principal components

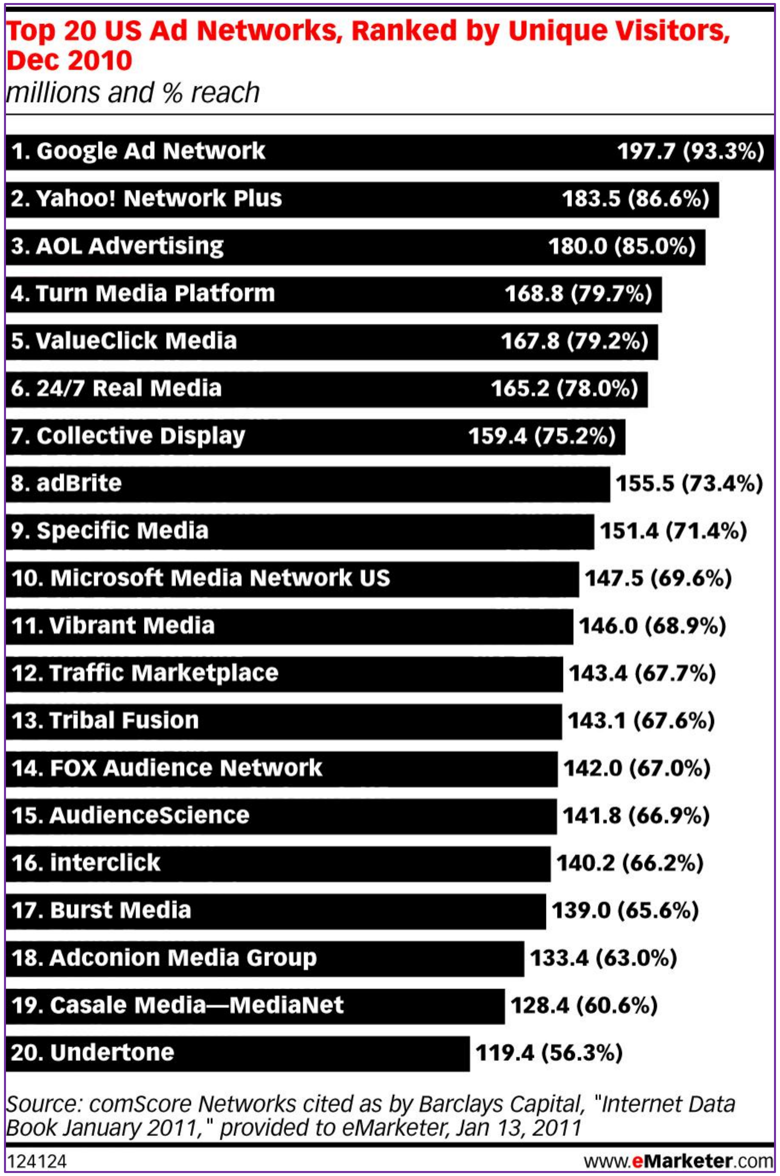
Principal components

- Publishers
- Yield optimization & supply side platforms (SSP)
- Ad networks
- Exchanges
- Demand side platforms (DSP) & Trading Desks
- Advertisers

Ad networks defined

- Companies that aggregate supply from multiple publishers or other intermediaries and matches it with advertiser demand
- Often sells inventory not sold as GD ("remnant") or from small publishers
- Predate the exchanges
- Translate from **publisher audience** (people that visit section X) to **advertiser audience** (people interested in Y)
  - Typically groups ad inventory by categories or demographics
  - Create segments of users that cut across multiple sites
  - **Key proposition that justifies the existence of many of the intermediaries**
- Margin revenue model: a percentage of transactions
- Horizontal, vertical, targeted, international,…
- Estimated > 300 ad networks!!

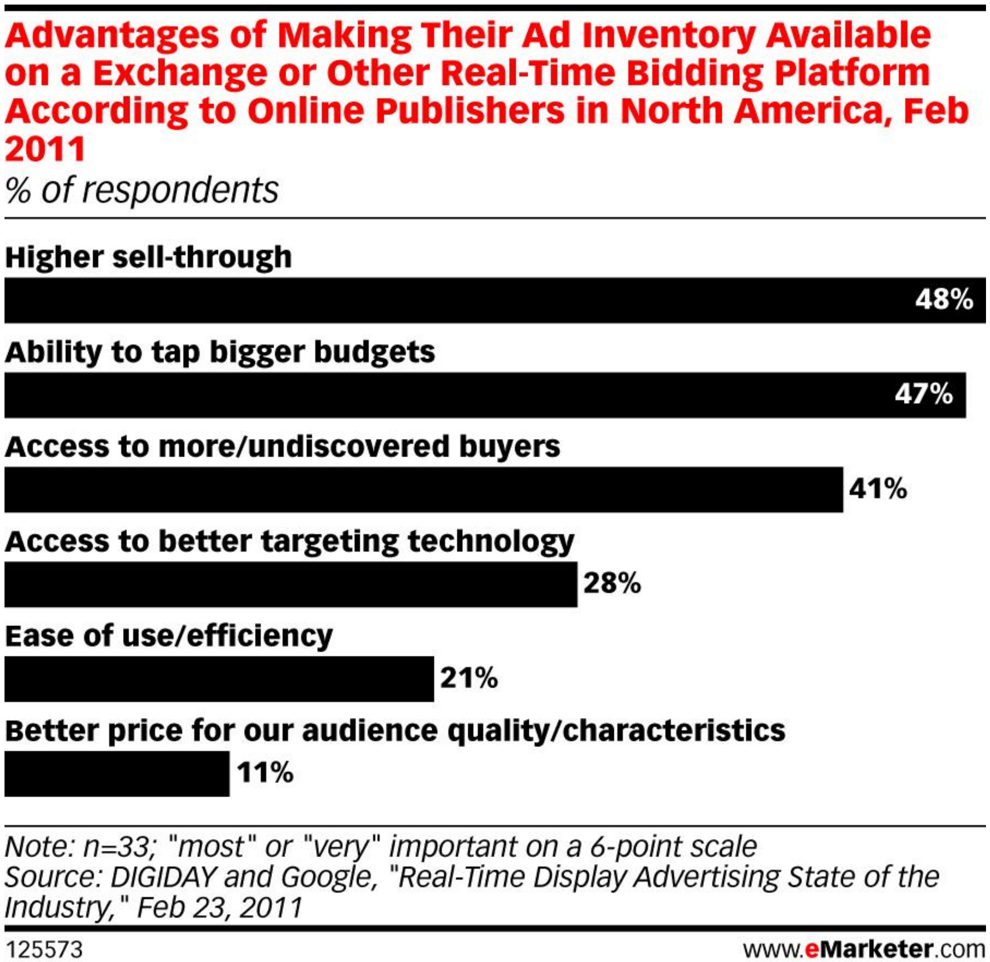
Top ad networks in US & their reach



Exchanges defined

- Marketplace for trading impressions between ad networks and some large advertisers or agencies
  - Similar to stock exchanges
  - Increase the liquidity of the marketplace by aggregating supply and demand
  - Use auction models to sell and charge for impressions
- Charge per transaction (not percentage)
- Buying possibilities:
  - Bulk buying – conditionally buy multiple impressions at once (e.g. offer $10 CPM for any mid age female on an entertainment page, total budget $1000/day)
  - Real time bidding (RTB) -- buy single impression (spot market) – e.g. offer 1c for a particular impression and a particular user
- Not every exchange allows RTB

Exchanges provide value to publishers



Demand Side Platforms (DSP) defined

- Ad exchanges complicated to use:
  - Which exchange?
  - When to bid?
  - What to bid?
- DSP: technology driven optimization for the demand providers
  - Unified access to multiple exchange
  - Integration with the data providers to allow Real Time Bidding (RTB)
  - Bid & budget management and optimization
  - Analysis of results
- Trading desks are essentially DSPs but serve a single agency → no competitive pressures/data leaks

- Audience Science, MediaMath, AdChemy, [x+1], Turn,…

Use of DSP

Methods Used by US Ad Agencies to Buy Online Advertising, Q1 & Q2 2011
% of respondents

	Q1 2011	Q2 2011
Through an ad network (Google, Yahoo!, Yellowbook, etc.)	66.0%	67.6%
Through traditional media (TV, print, radio stations, etc.)	69.1%	50.9%
Direct from the publisher (Expedia, OpenTable, NCAA, etc.)	41.5%	43.5%
Through a DSP or exchange (Invite Media, Right Media, Acxiom, etc.)	10.6%	19.4%
Through self-services (FatTail's PageGage, Adap.tv's Marketplace, etc.)	2.1%	3.7%
Other	6.4%	8.3%

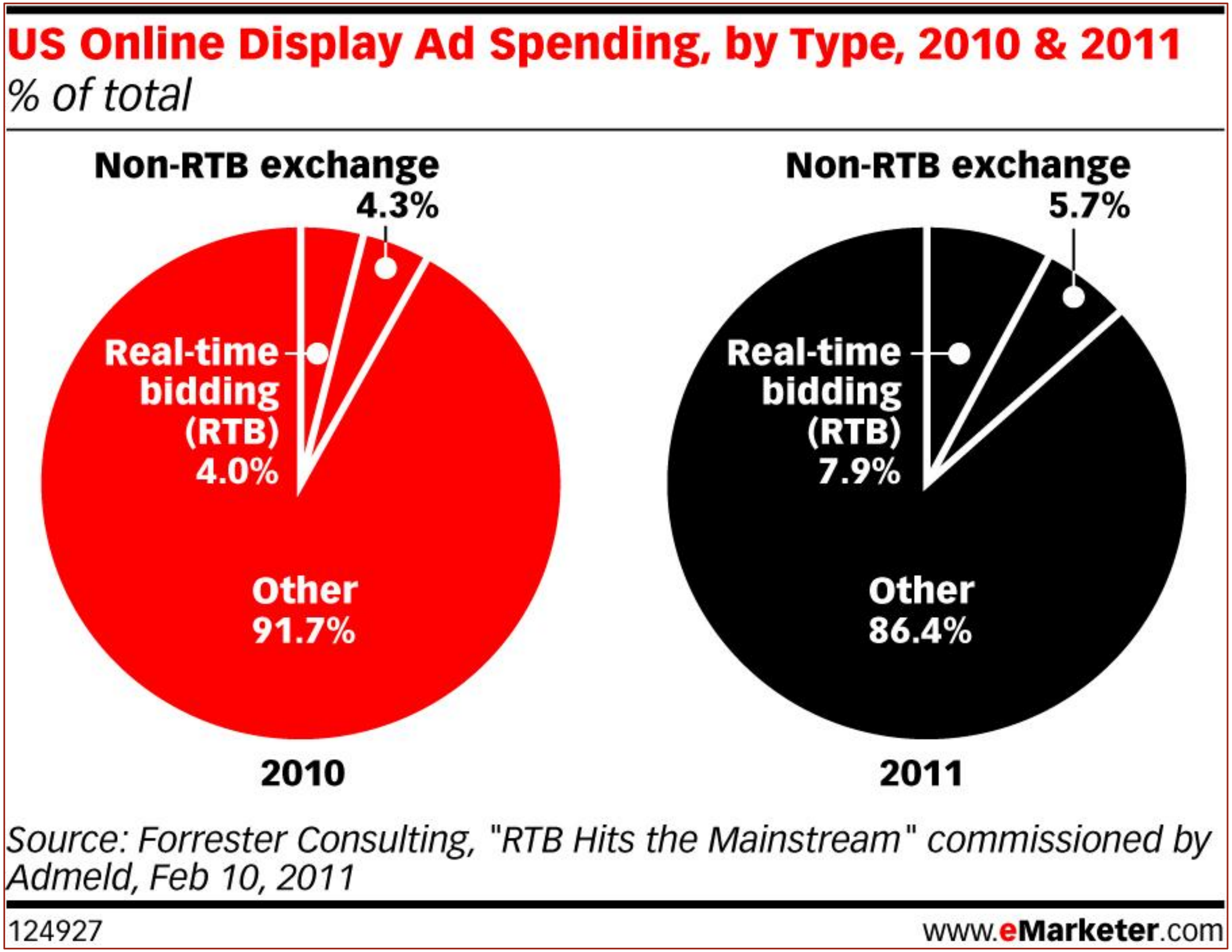
Source: STRATA, "2nd Quarter 2011 Survey Results," July 26, 2011

130575 www.eMarketer.com

Real time bidding defined

- Component of an exchange protocol that allows buyers to bid for each ad impression (context + user) separately in real time
- The buyers use their own data and targeting options + purchased data to decide how much to bid
  - Profiles, location, recent activity, etc. More in the targeting lecture
- RTB bid is usually CPM + first price.

RTB usage: small but growing faster than non RTB exchange



Ad Agencies

- Present in traditional and online advertising
- Define campaigns, and formats of advertising
- Manage budgets for the advertisers and split it between different online and offline formats
- Buy media in bulk from
  - Networks
  - Publishers
  - Exchanges
- Omnicom, WPP, IPG, Aegis,…

Data supply trade and aggregation

- Most publishers have multiple pixels on their pages (cookies or beacons)
- Benefit from this data by improved ad targeting
- Can the data also be a revenue source of it self?
- Data sale is a upcoming industry:
  - Mobile, location, social network
  - Auction based model or direct sale
- Still not regulated, but FTC is starting to note
- A mixture of new and old players:
  - BlueKai, eXelate, Experian, Comscore, Nielsen
- See: http://adage.com/adnetworkexchangeguide09/article?article_id=136003
- Key Quote: Aggregators and Exchanges Aim to Create 'Liquid Market' Based on Users' Activities, Not Their Locations - but Can They Get Past Privacy Concerns?

Putting everything together: Re-targeting

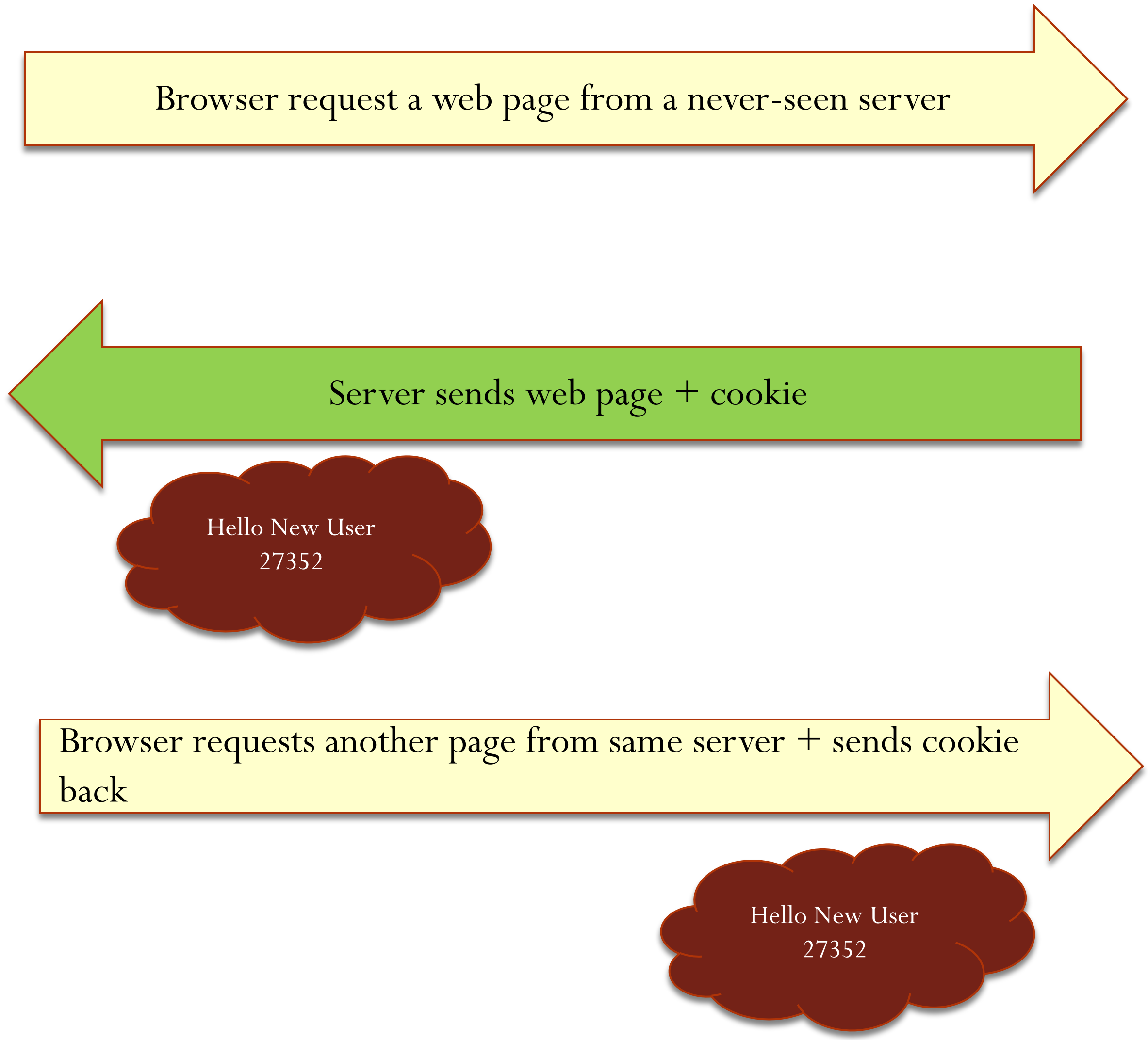
Re-targeting idea

- Use immediate search or browse to target/create ads
  - Examples:
    1. User that has searched for “Prius” sees ads for Prius or Toyota dealer for the next few days on non-search context, e.g. when browsing a complete different site
    2. User that has browsed a fashion site sees ads for shoes when browsing a complete different site or using e-mail, etc
    3. User that has browsed BuyAGizmo.com but did not convert, sees “get back” ads for BuyAGizmo.com on many other sites, maybe + coupon, special discount
- Mostly search re-targeting due to higher intentionality
- Special case of BT (more recent, more specific)
- Companies:, Advertising.com, FetchBack, Real Media, Dapper, Microsoft DrivePM, Audience Science, BlueLithium (Yahoo!),

Detour

Cookies

Cookies


Browser request a web page from a never-seen server
Server sends web page + cookie
Hello New User
27352
Browser requests another page from same server + sends cookie back
Hello New User
27352

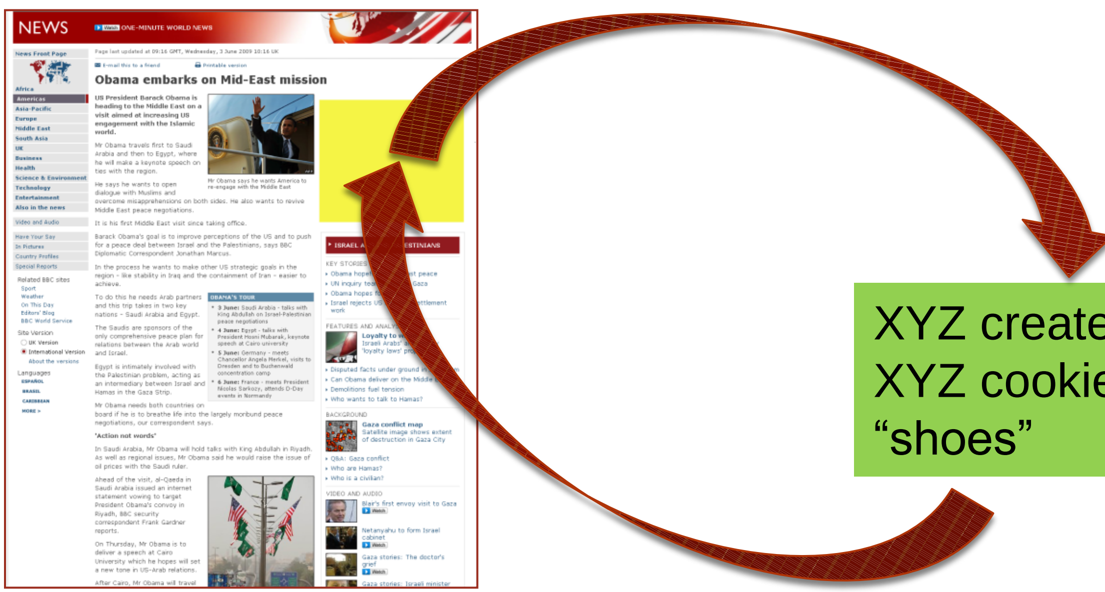
Third party cookies

- First party cookie: cookies associated to the domain shown in the browser's address bar
- Third party cookies: cookies associated to different domains than shown in the browser's address bar.
  - Created because parts of the page are created by http requests to other domains, not the domain in the bar, e.g. an image stored on a different server
  - Extreme case: a beacon = 1 pixel invisible image used to track a user visit
- More in the targeting lecture

Basic search retargeting scheme

- User searches for shoes on the XYZ engine

Site ABC sends ad request + XYZ cookie to XYZ



XYZ creates shoes ad based on XYZ cookie that remembers “shoes”

Basic browse retargeting scheme

- User Joe views skiing site ABC that contains some XYZ produced ad or just "beacon"
- Joe's XYZ cookie captures visit to ABC
- Now on site DEF Joe sees ski ads

Sends ad request + XYZ cookie to XYZ

•XYZ creates skiing ad based on XYZ cookie that remembers "skiing"
•Alternative: XYZ puts ad for ABC

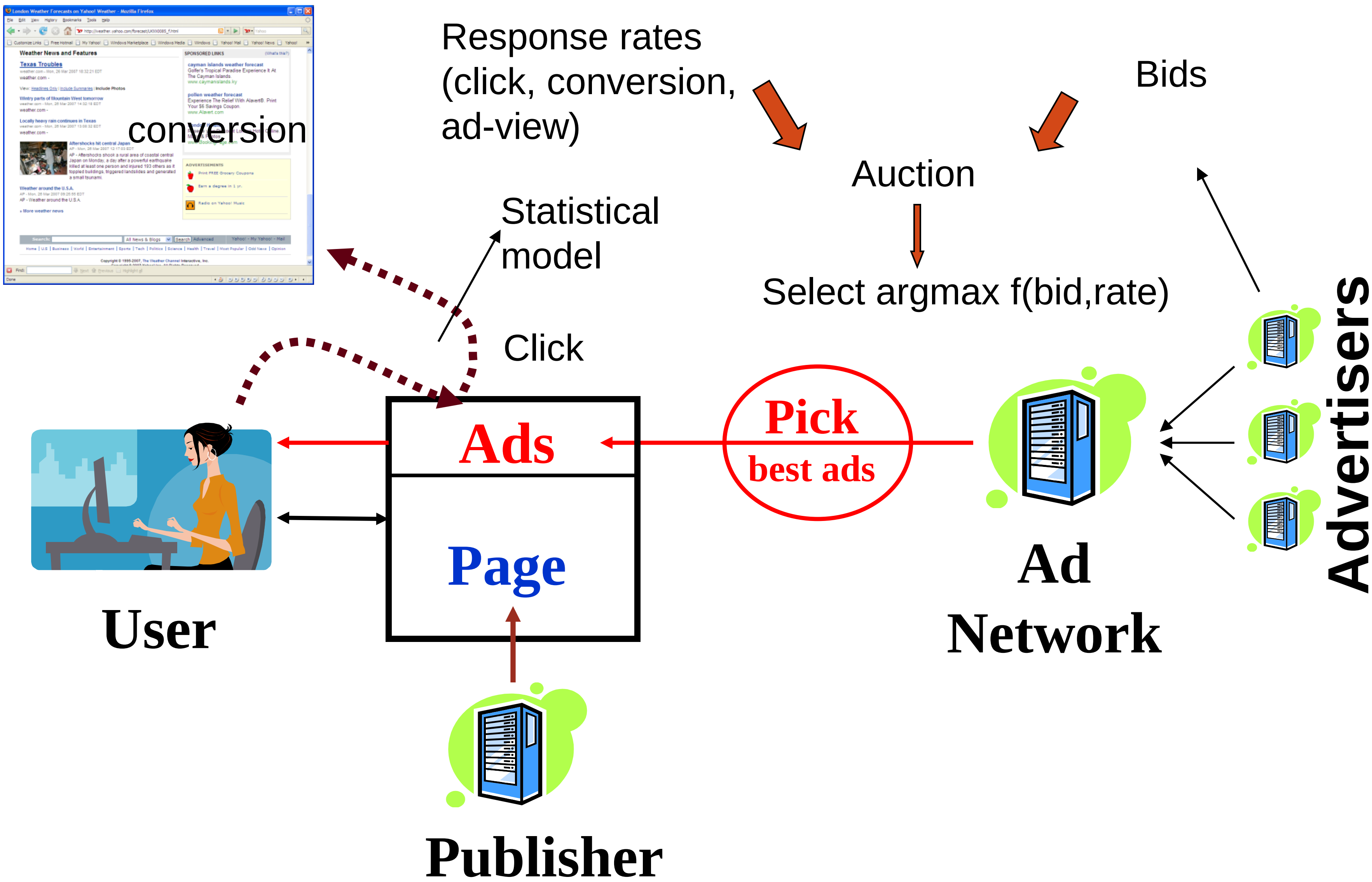
AOL explanation

Estimating Response Rates in Display Advertising through Multi-Hierarchy Smoothing

Deepak Agarwal & Nagaraj Kota ,Yahoo!

IISA 2011

Ad selection via bidding



Exchanges

- Advertisers participate (bid) in different ways
  - CPM (pay by ad-view)
  - CPC (pay per click)
  - CPA (pay per conversion)
  - DCPM (Exchange bids CPM but based on conversion goals. Like CPA/CPC but advertiser takes the risk)
- To conduct an auction, *normalize* across pricing types
  - Compute eCPM (expected CPM)
    - Click-based ---- eCPM = click-rate*CPC
    - Conversion-based ---- eCPM = conv-rate*CPA

Pricing Types

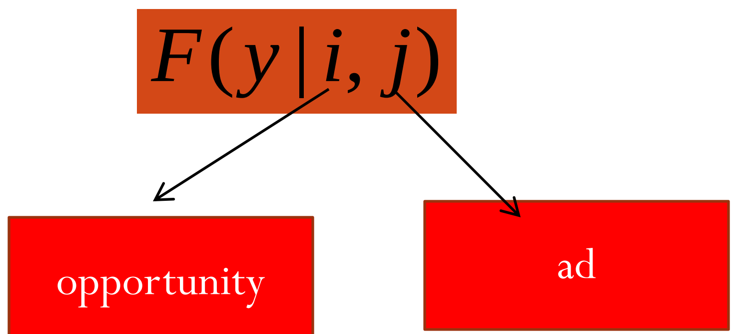
- RevShare
- CPM
- CPM w CPC goal
- CPM w CPA goal
- CPC
- CPC w CPA goal
- CPA
- dCPM (dynamic CPM)
- dCPM w CPC goal
- dCPM w CPA goal

Statistical Issues

- Need to estimate the click or conversion probability

  f(context, user, ad)

- High dimensional density estimation



- Response obtained through interaction among few heavy-tailed categorical variables (opportunity and ad)

Problem Definition

Example applications
Content, Movie,
Advertising,
Shopping,
…..

Item Inventory
Articles, web page,
ads, …

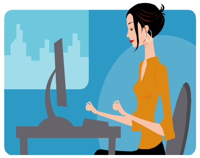
Context
page,
previous item viewed,
…

Construct an automated algorithm to select item(s) to show

Get feedback
(click, time-spent,rating, buy,…)
Refine parameters of the algorithm

Repeat (large number of times)
Optimize metric(s) of interest
(Total clicks, Total revenue,…)

Low Marginal cost per serve,
Efficient and intelligent systems can provide significant improvements



USER

Structure in the Data

- Covariates available for both opportunity and ad
  - Opportunity: Publisher content type, user demographics,…
  - Ad: Industry, text/video, text (if any)

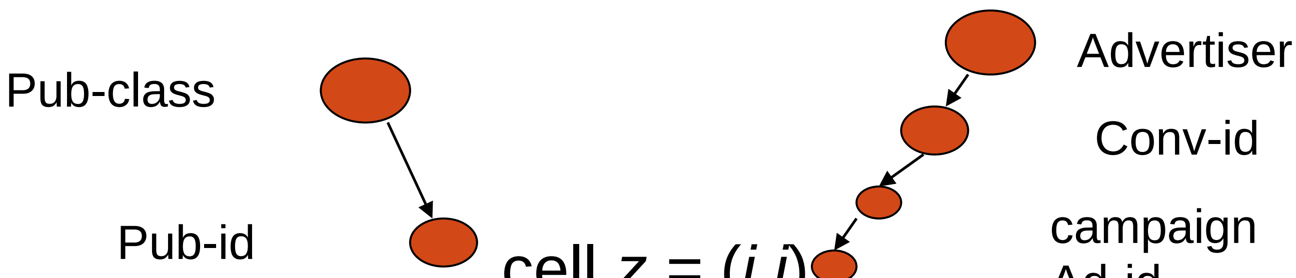
- Hierarchically organized
  - Publisher hierarchy: URL $\rightarrow$ Domain $\rightarrow$ Publisher type
  - Geo hierarchy for users
  - Ad hierarchy: Ad $\rightarrow$ Campaign $\rightarrow$ Advertiser

Statistical Issues

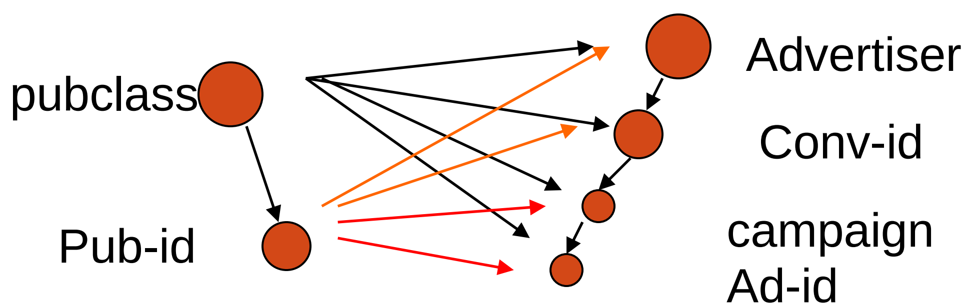
- Key: Modeling residual correlations in residual response rates when interacting variables are organized hierarchically
  - (Actually a DAG is fine, tree not needed)
- High dimensional
  - 100M "cells" from ~ hundred billion auctions
- Data sparsity
  - Large number of zeroes, extreme variation in #Tries
    - Small sample size corrections essential
- Model parsimony
  - Cannot store a large model in our ad-servers, parsimony along with accuracy important
- Temporal smoothing

Hierarchical Smoothing of residuals

- Assuming two hierarchies (Publisher and advertiser)



- Use the well represented interactions for smoothing



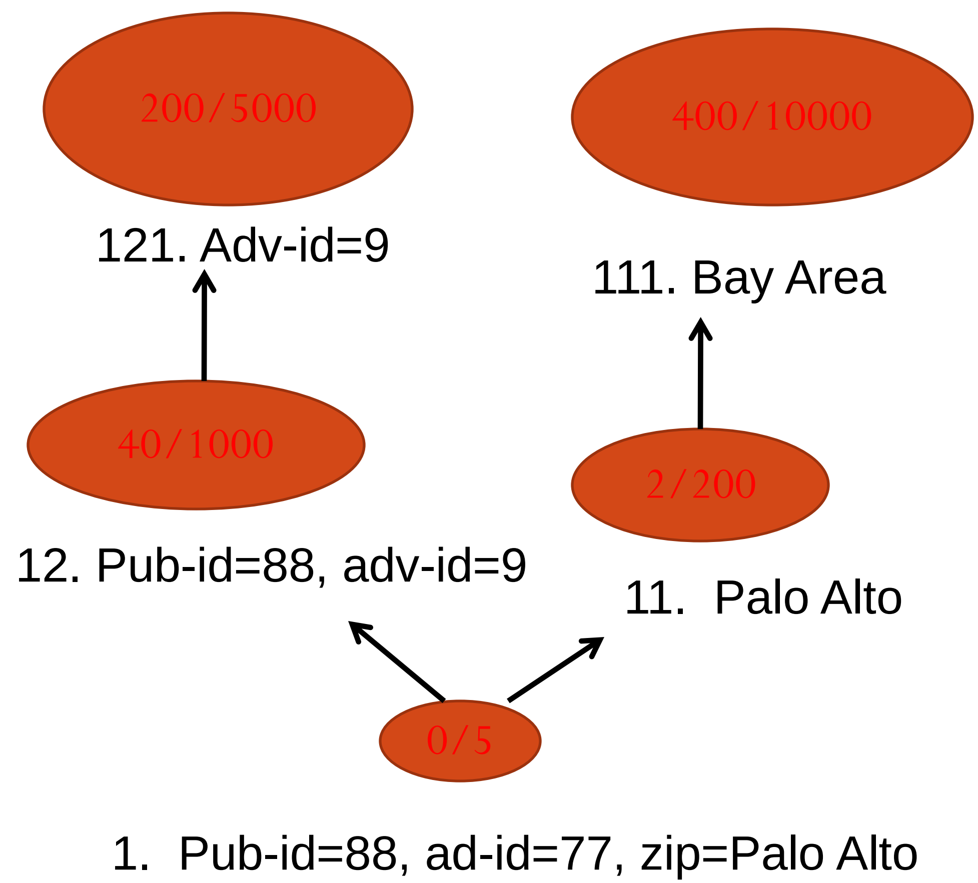
Modeling data at granular resolutions

- Pros of learning things at granular resolutions
  - Better estimates of affinities at event level
    - (ad 77 has high CTR on publisher 88, instead of ad 77 has good CTR on sports publisher)
  - Bias becomes less problematic
    - The more we chop, less prone we are to aggregating dissimilar things, less biased our estimates from non-randomized data
- Challenges
  - Too much sparsity to learn everything at granular resolutions
    - We don't have that much traffic
    - E.g. many ads are not even shown on many publishers
  - Explore/exploit helps but cannot do so much experimentation
  - In advertising, response rates (conversion, click) are too low, further exacerbates the problem

Solution: Go granular but with back-off

- Too little data at granular level, need to borrow from coarse resolutions with abundant data (smoothing, shrinkage)

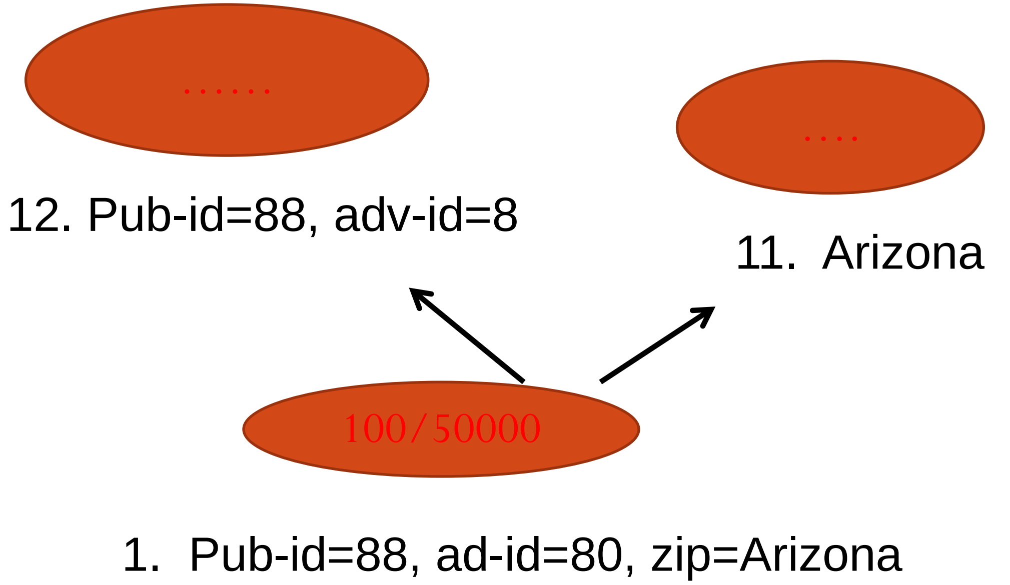
$$CTR(1) = w_1(0/5) + w_{11}(2/200) + w_{12}(40/1000) + w_{121}(200/5000) + w_{111}(400/10000)$$



Sometimes enough data at granular level

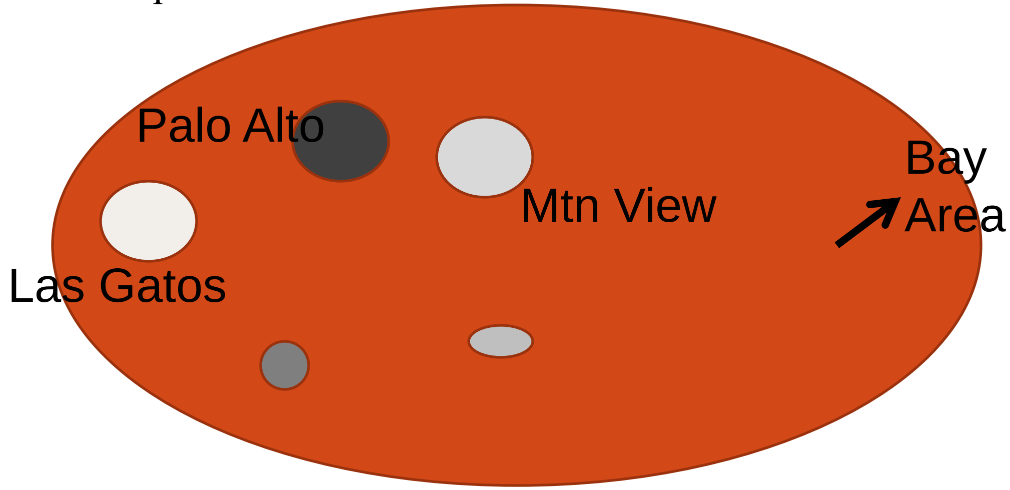
No need to back-off

CTR(1) = 100/50000



How much to borrow from ancestors?

- Learning the weights when there is little data
- Depends on variance in CTRs of small cells
  - Ancestors with similar CTR child nodes are more credible
- E.g. if all zip-codes in Bay Area have similar CTRs, more weights given to Bay Area node
  - Pool similar cells, separate dissimilar ones



Common misconception

- If a parent has large amounts of data, the variance is small and hence it is a good back-off candidate
  - WRONG!
- The back-off depends on the variance in *true* CTRs of small cells nested in a parent node
  - This variance does not go to zero even with infinite data!
- But how do we estimate variance in true CTRs when the goal is to estimate the CTRs themselves?
  - Statistical modeling assumptions can help

Data

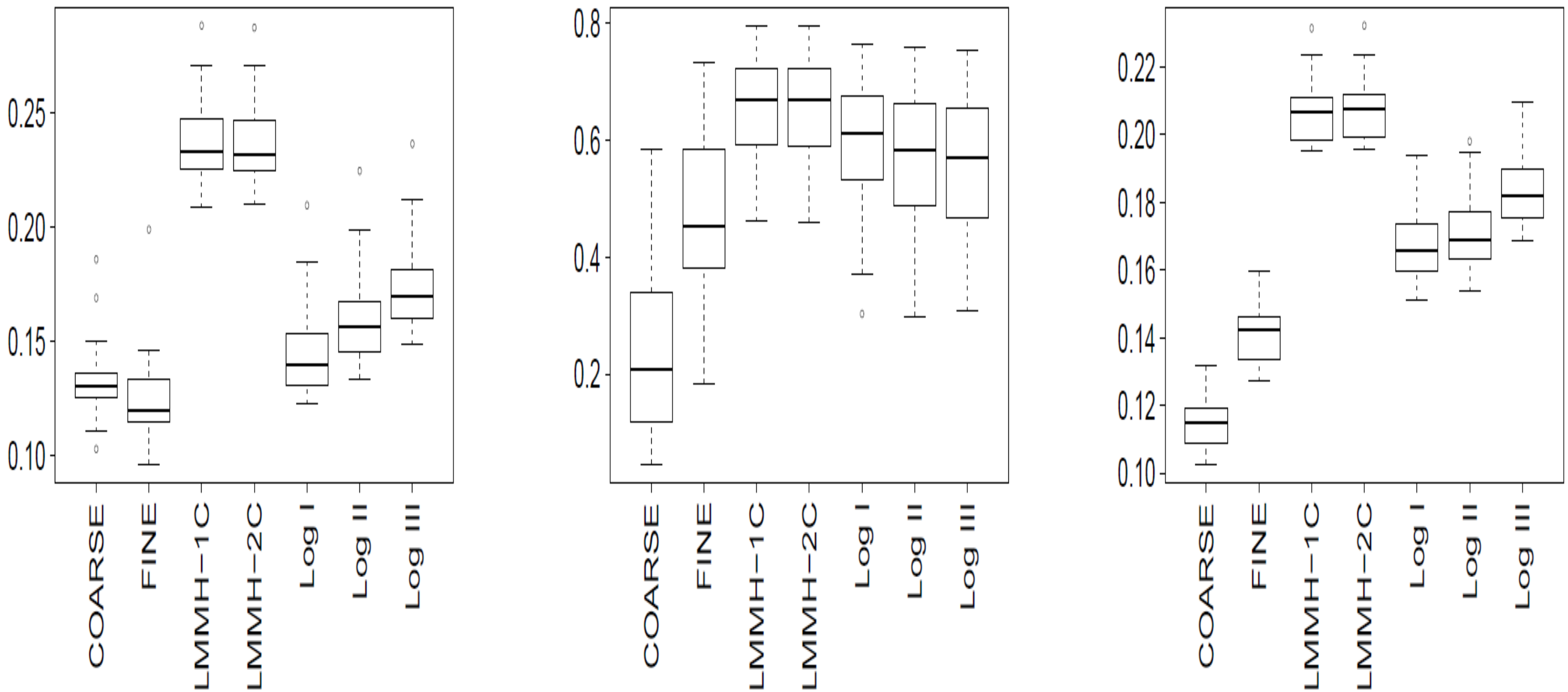
- Two kinds of conversion rates
  - Post-Click conv-rate = click-rate*conv/click
  - Post-View conv-rate = conv/ad-view

- Three response rate models
  - Click-rate (CLICK), conv/click (PCC),
  - post-view conv/view (PVC)

Datasets : Right-Media

- CLICK [~90B training events, ~100M parameters]
- Post Click Conversion(PCC) (~.5B training events,~81M parameters)
- PVC – Post-View conversions (~7B events, ~6M parameters)
  - Cookie gets augmented with pixel, trigger conversion when user visits the landing page
- Features
  - Age, gender, ad-size, pub-class, user fatigue
  - 2 hierarchies (publisher and advertiser)
- Two baselines
  - Pubid x adid [FINE] (no hierarchical information)
  - Pubid x advertiser [COARSE] (collapse cells)

Accuracy: Average test log-likelihood



(a) PCC (b) PVC (c) CLICK

More Details

- Agarwal, Kota, Agrawal, Khanna: *Estimating Rates of Rare Events with Multiple Hierarchies through Scalable Log-linear Models, KDD 2010*

Summary

- Scalable map-reduce log-linear models to precisely estimate rare response rates by exploiting correlation structures with cross-product of hierarchies

- Significantly better than state-of-the-art logistic regression methods widely used in computational advertising

Recent Work

- Mixture of Multi-Hierarchy Models
  - Baseline model using Decision Trees, one Multi-Hierarchy Model per node that are shrunk towards each other

- Hierarchical-Temporal Smoothing through Kalman filters

Summary

Key points

- Structure of the NGD marketplace
- Prediction should take advantage of hierarchical structure of the data

Questions?

We welcome suggestions about all aspects of the course: msande239-aut0910-staff

Thank you!

broder@yahoo-inc.com
vanjaj@yahoo-inc.com

http://research.yahoo.com

Targeting

Scientific perspective on targeting

Targeting from the matchmaker point of view

Targeting as a dual problem of ad selection

Ad Selection ← duality → Targeting

Find optimum ad given(context, user)

Find optimum users given (context, ad)

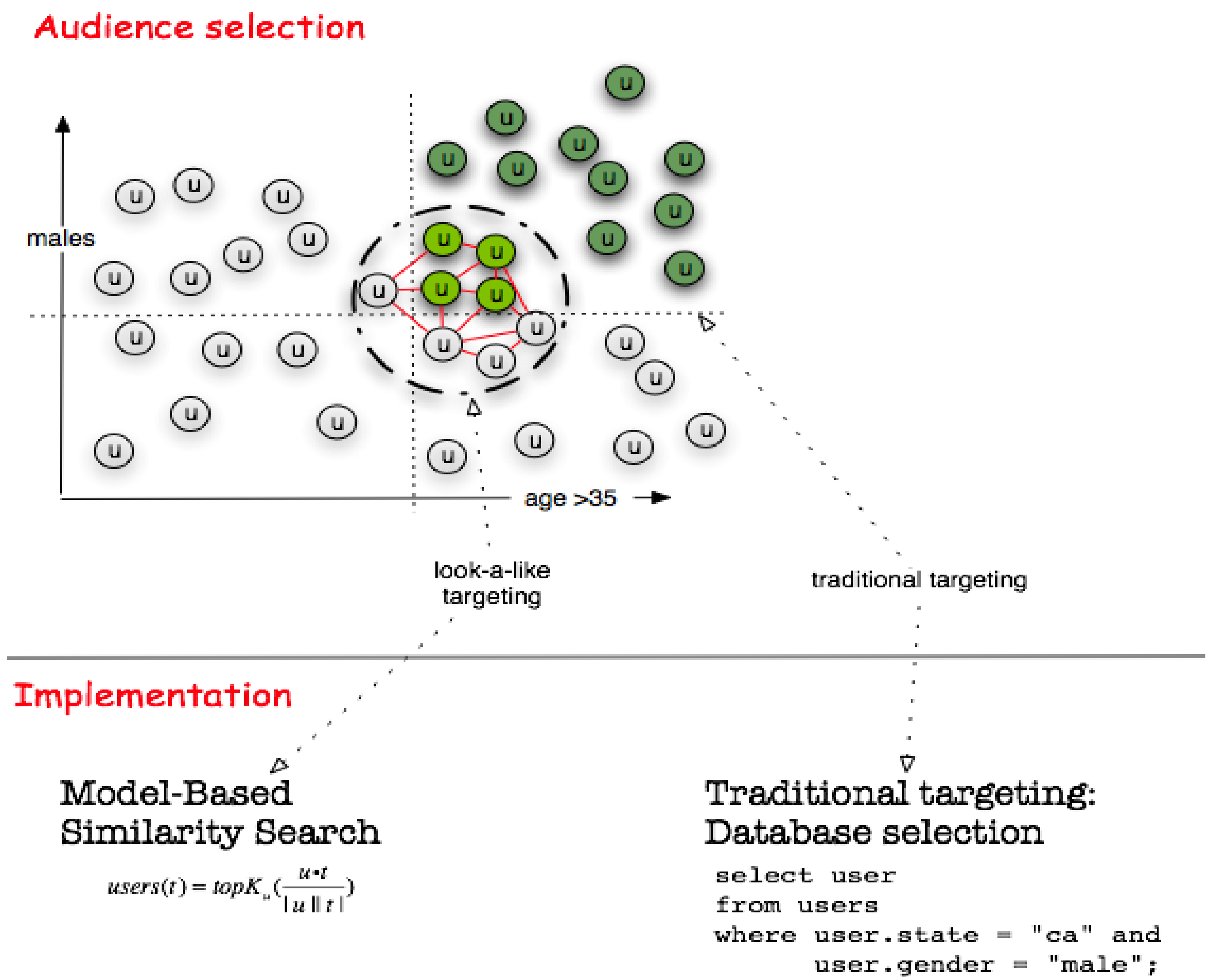
- Targeting is **<u>audience selection</u>**
- Traditionally not understood as such
- Duality not easy to explain to media buyers! ☺

Targeting specification

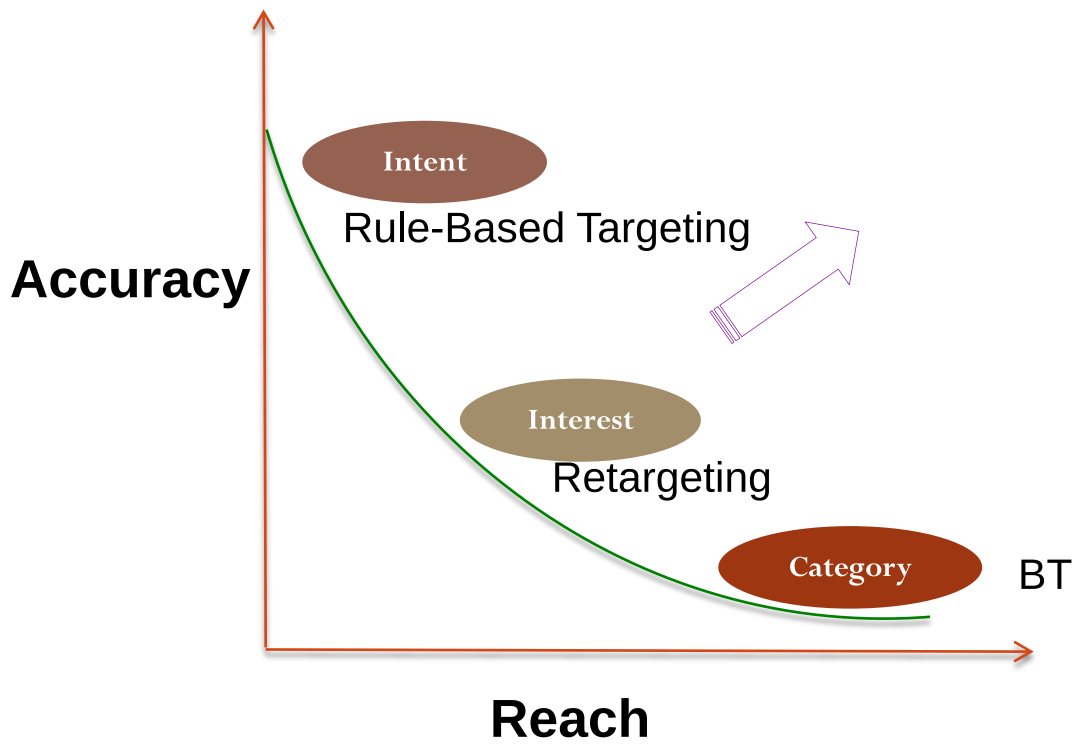
- "Classic" → Rule based
  - Males in California, 25-35 age
  - Focus on a few dimensions in the user profile
  - Boolean selection (database query): the user is either qualified or not

- "Modern" → Model based
  - Determine the weight for all dimensions/features
    - Market research – identify the characteristics of the most responsive users
  - Supervised approach: based on a given set of positive users
    - Users that have already interacted with the advertiser
    - Market research – identify the users that appear most likely to be responsive

Audience selection vs. traditional targeting



Aim of targeting: move the curve up



Current trends in targeting

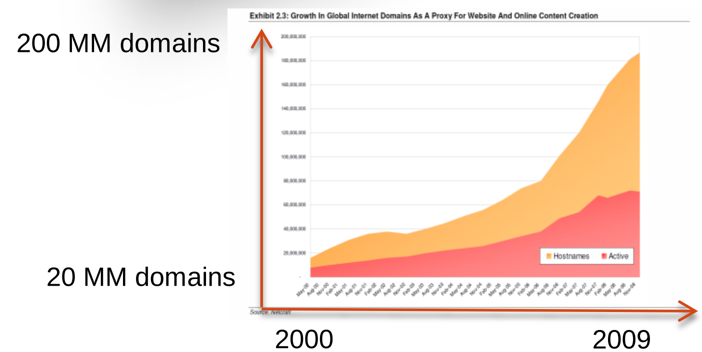
The world from the standpoint of the advertiser

Market Trend: Growing need for deeper targeting

- **Context was proxy for audience** Historically advertisers have bought inventory based on Context (e.g. Gillette advertises on Sports sites) or publisher defined segments

- **Supply has fragmented** Increasing fragmentation of digital media consumption challenges advertisers to find audiences at scale

- **Marketers are looking for highly targeted reach** as reach per unit of buy is going down



Market needs: Advertisers want to target ‘personas’

Targeting Personas: Advertisers want to target a specific persona – that may not be available through a standard Demo or a publisher defined category

For cosmetics company, XYZ, the segments “**Women between 35-54**” or “**Interest Beauty-Cosmetics**” are not narrow enough as they **don’t capture the unique beauty needs of women with different persona** in the same age group

“Frazzled Mom”

≠

“High Flying career woman”

≠

“School Teacher”

More on personas

- Online users perform a sequence of (overlapping) tasks
- Personas share interests and behavior
- Usually pre-defined segments based on real world behavior
- Same user can have multiple personas: **personas are facets of personality**
  - Andrei is a computer geek
  - Andrei is also
    - An eclectic reader
    - An expert cook
    - A non-expert skier
    - Etc
- NB: "is" here means "behaves like"

Market needs: Campaign objectives are varied

Objectives: Every campaign is unique in what it's looking to achieve – based on type of segment, product, purchase funnel, etc.

Large department store	Goal: Drive weekend traffic **Objective: In-store traffic, weekend sales**
Credit card company	Goal: Promote new low introductory rate **Objective: Qualified credit card applications**
Cosmetics company	Goal: Raise awareness of new products **Objective: Reach / Unique users**

Custom ads (dynamic ads, smart ads)

- Use targeting attributes to create custom ads
- Ads are modified at run-time (under the control of the advertiser)
- Contrast with the situation where ads are chosen by the intermediary

Sampler of targeting techniques

Demographic targeting

- Important indicator of people's interest and potential of a conversion
  - Imagine you want to sell a $50K sports car. Who do you target?
- Used widely in traditional advertising:
  - TV, magazines, etc. maintain very detailed statistics of their audience
- Common classic dimensions:
  - Age
  - Gender
  - Income bracket
  - Location
  - Interests ("Golf enthusiast")
  - ....
- Each dimension has multiple values

Geo targeting

- Goal: determine user location
  - Home
  - Current
  - Often wrong ☹
- Inputs
  - Registration data
  - IP (Main source!)
  - Browser default language
  - Search language
  - Etc …
- Lots of papers/results, but no time to discuss …

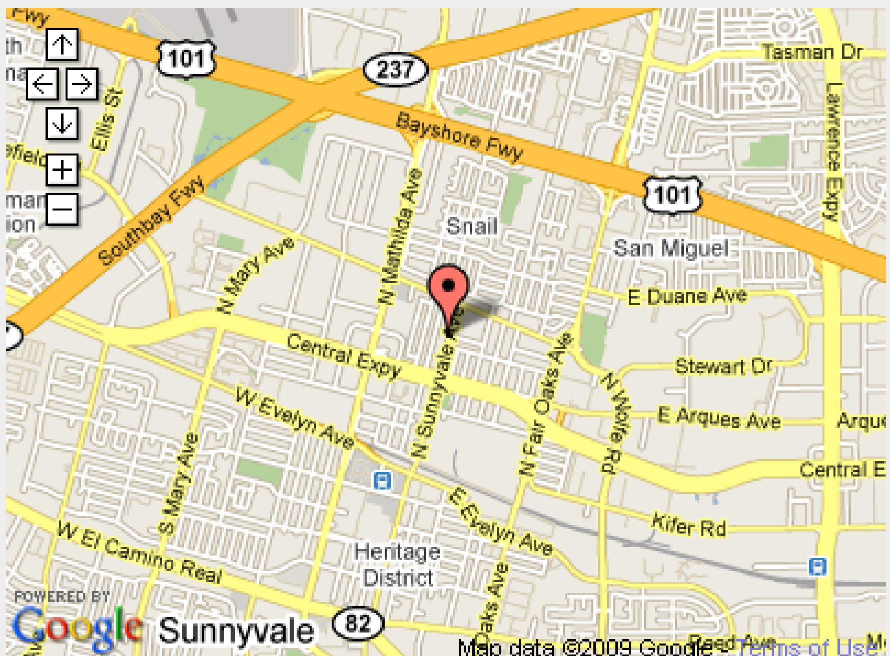
Example: hostip.info

Domain to IP or **Host name lookup**

209.131.62.115 | Go

Host name: **nat-dip6.cfw-a-gci.corp.yahoo.com.**
IP address: **209.131.62.115**
Location: **Sunnyvale, CA, UNITED STATES** (change)

Are you an ISP / host? **Update an entire block**



Live demos

- Message selection based on location and weather data in a demo from Teracent (bought by Google) for Dunkin' Donuts: http://adserver.teracent.net/tase/demo/DunkinDonuts.jsp
- Yahoo! Smart ads simulator http://advertisingcentral.yahoo.com/publisher/sa_simulator
- Visit www.overstock.com . The items you are examining are shown later on ads from overstock.com e.g. on http://dailycaller.com/

Behavioral Targeting (BT) aka "interest-based advertising"

- A technique used by publishers and advertisers to increase campaign effectiveness based on a given user's historical behavior:
  - **Previous searches/search sessions**
  - **Previous browsing activity**
  - **Previous ad-clicks**
  - Previous conversions
  - Declared demographics data
  - Etc.

Main techniques

- Utility – everyone wins! (at least in theory ☺)
  - Advertisers: get a more appropriate/receptive audience, increased conversion rate, better ROI
  - Publishers: can ask for a premium
  - Users: see more interesting ads

How popular is BT?

Source: Datran Media survey of 3,000 execs from Fortune 1000 companies December 2008

Privacy concerns

- BT mostly based on cookies
- Users do not understand the cookie mechanisms
- Difficult to turn off – many sites stop being functional without cookies
- If you accept cookies from XYZ, XYZ can become aware of your visits to any site where XYZ has a visible or invisible presence on the page.
- Many proposals / regulations / "trust-me" solutions
  - E.g. Phorm (http://www.phorm.com/) promises to collect only category data and keep cookies anonymous (not linked to IP, name, etc)
  - Most companies have data retention policies (90 days for Yahoo!)
  - Most companies allow user control over stored data
  - Opt-out BT
  - Etc

Hot off the press: Network advertizing initiative

- Signed on by all major players + trade groups
- Icon on all ads:

- Users click on it and get general “opt out” page

Opt out status

Opt-Out Status

Select all | Clear | Submit

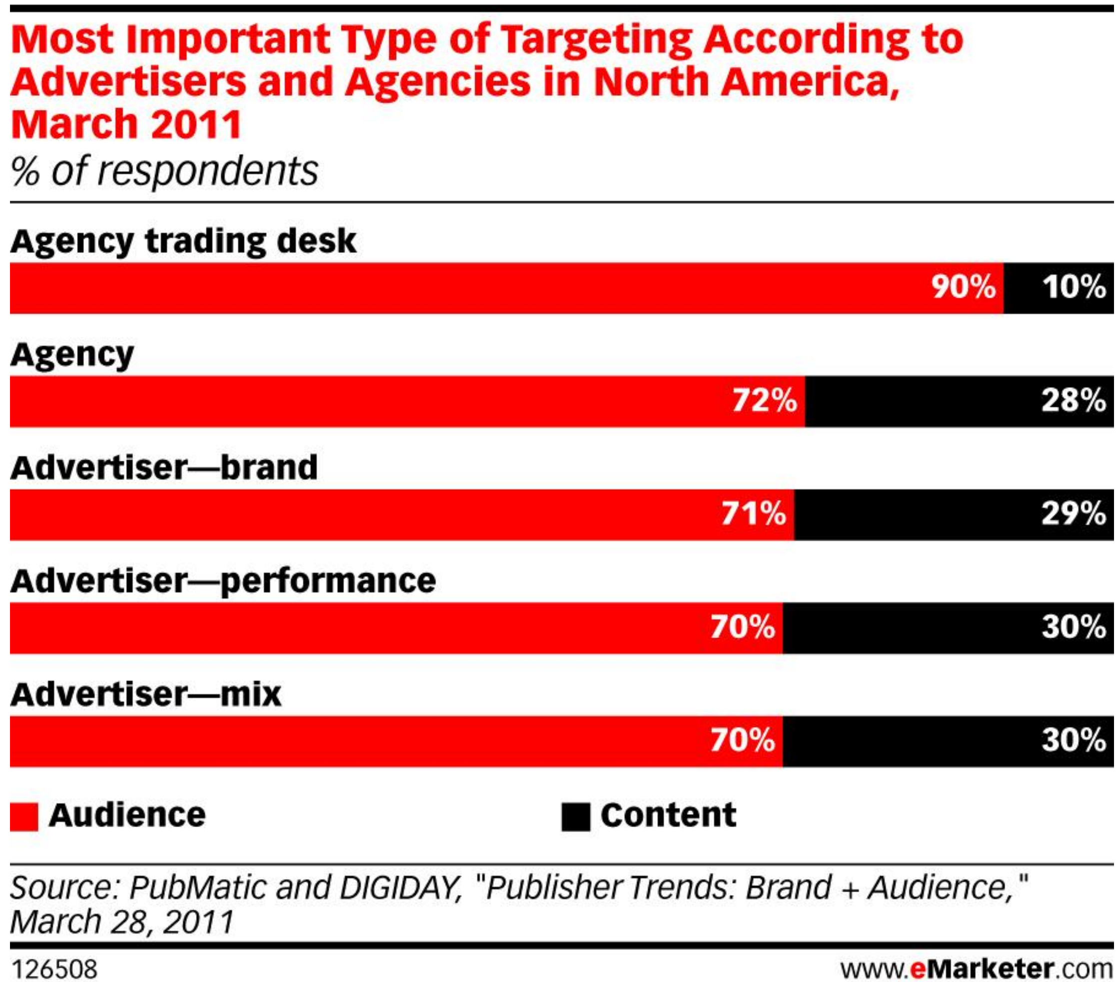
Member Company	Status	Opt-Out
aCerno More Information	**Active Cookie** You have not opted out and you have an active cookie from this network.	Opt-Out ☐
AdBrite More Information	**Active Cookie** You have not opted out and you have an active cookie from this network.	Opt-Out ☐
AdChemy More Information	**No Cookie** You have not opted out and you have no cookie from this network.	Opt-Out ☐
Adconion More Information	**Active Cookie** You have not opted out and you have an active cookie from this network.	Opt-Out ☐
Adara Media More Information	**No Cookie** You have not opted out and you have no cookie from this network.	Opt-Out ☐
Adify Media	**Active Cookie**	Opt-Out ☐

Summary

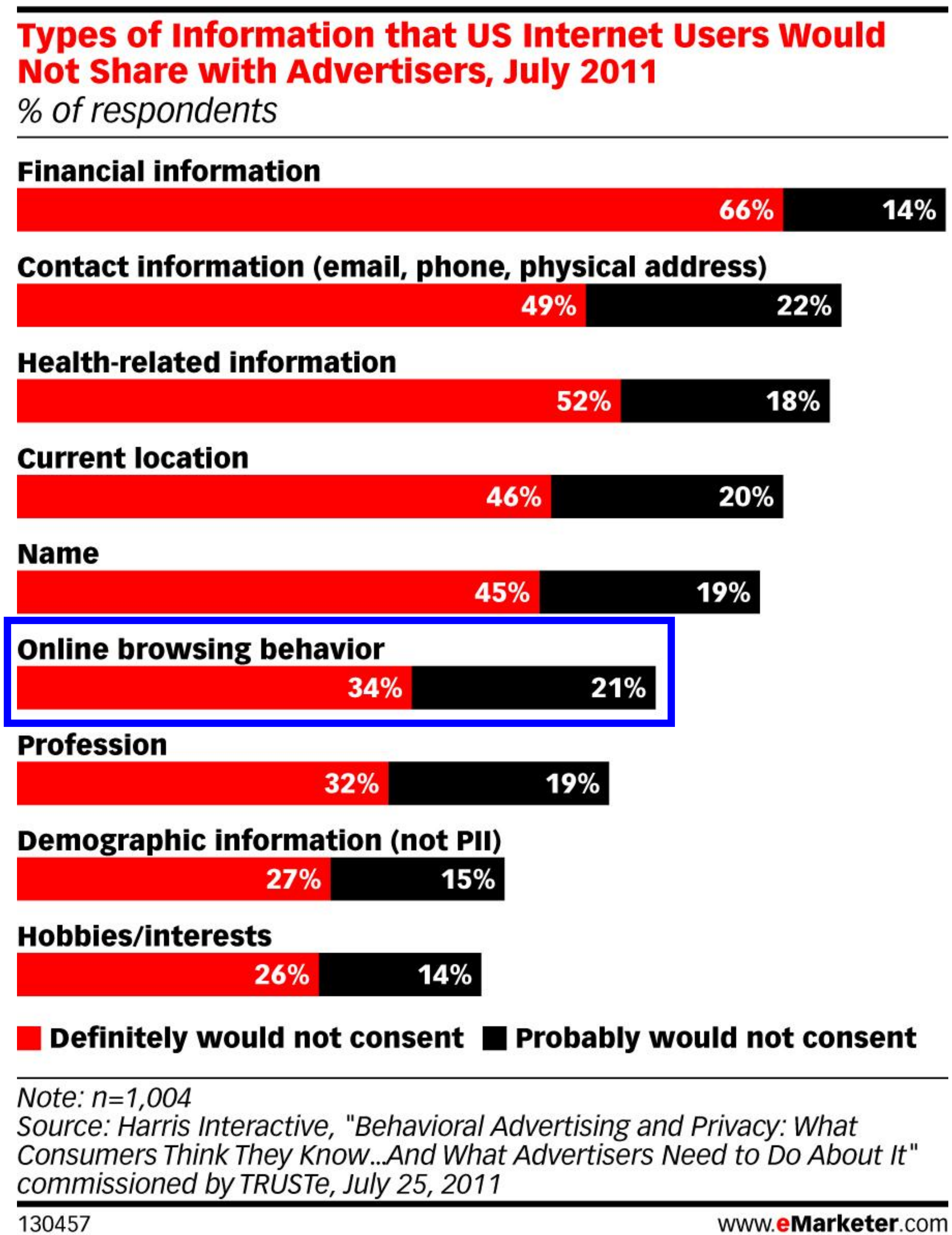
Key points

- Targeting is emerging as a key component of on-line advertising
- Demographic targeting mimics classic advertising, but it can be both declared and inferred
- Behavioral Targeting based on recent search and query behavior appears effective, quantitative aspects still TBD
- Re-targeting is becoming very popular, with new players mushrooming
- Privacy and regulatory concerns

30% of marketers or less cite content targeting as most important vs.. at least 70% who favor audience targeting



Limits of audience data—and therefore targeting—will be shaped increasingly by the types of info that users will NOT share



Ad inventory options, sources and methods

- Publisher direct (premium)
- Ad networks
- Ad exchanges
- Private exchanges
- Demand-side platforms (DSPs)
- Agency trading desks
- Real-time bidding (RTB)

Ad inventory options, sources and methods

- Publisher direct (premium)
- Ad networks
- Ad exchanges
- Private exchanges
- Demand-side platforms (DSPs)
- Agency trading desks
- Real-time bidding (RTB)

Publishers and networks (branding) vs.. exchanges, DSPs, RTB (direct response)

Best Use of Select Display Inventory Sources/Buying Methods According to US Agencies, Dec 2010

% of respondents

1 Fostering engagement
2 Generating awareness
3 Prospecting
4 Direct response
5 Cost-effective

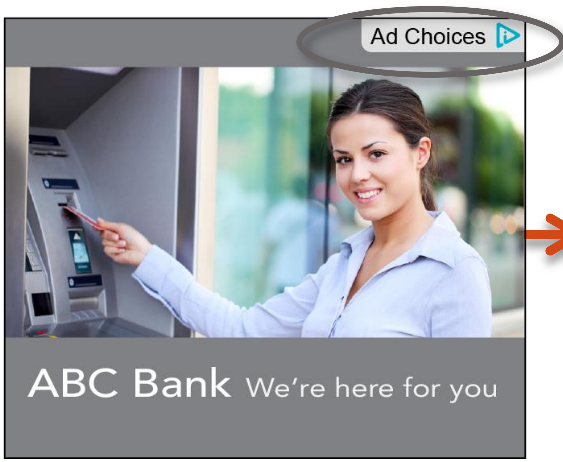
	1	2	3	4	5
Premium publishers	22%	54%	4%	7%	6%
General ad networks	6%	33%	14%	13%	33%
Exchanges, DSPs, RTB	5%	13%	13%	24%	40%

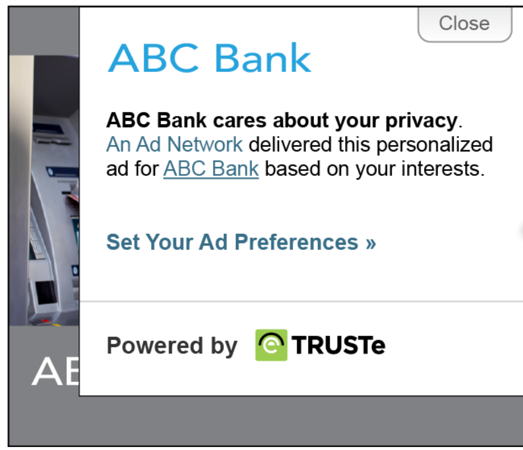
Note: n=109
Source: DataXu and DIGIDAY, "Digital Advertising State of the Industry Survey," Dec 9, 2010

131027 www.eMarketer.com

How a DAA Compliance Solution Works

Advertising User Experience





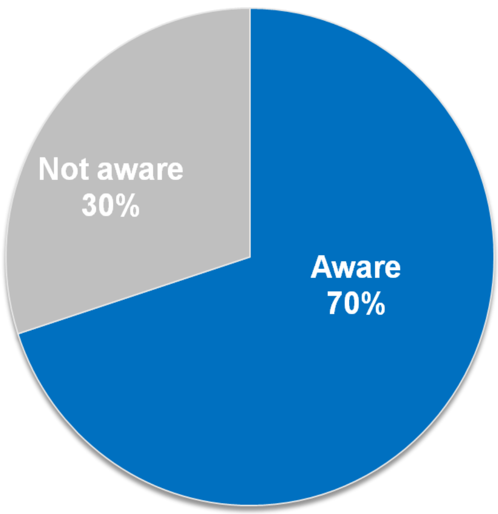
1. A simple ad tag inserts the DAA icon

2. If clicked the icon launches privacy notice inside the ad

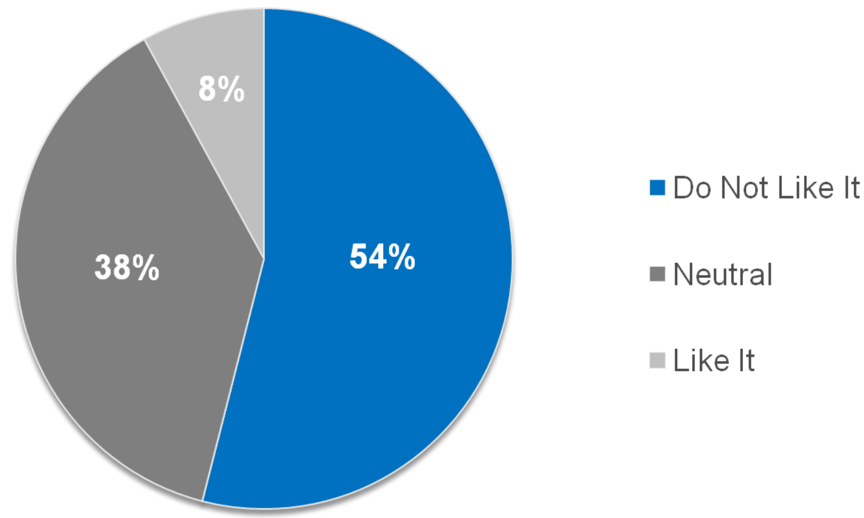
3. Consumers have option to click to Preference Manager and opt out of selected tracking networks

Consumers are Aware and Skeptical of Behavioral Advertising

Awareness of OBA Concept



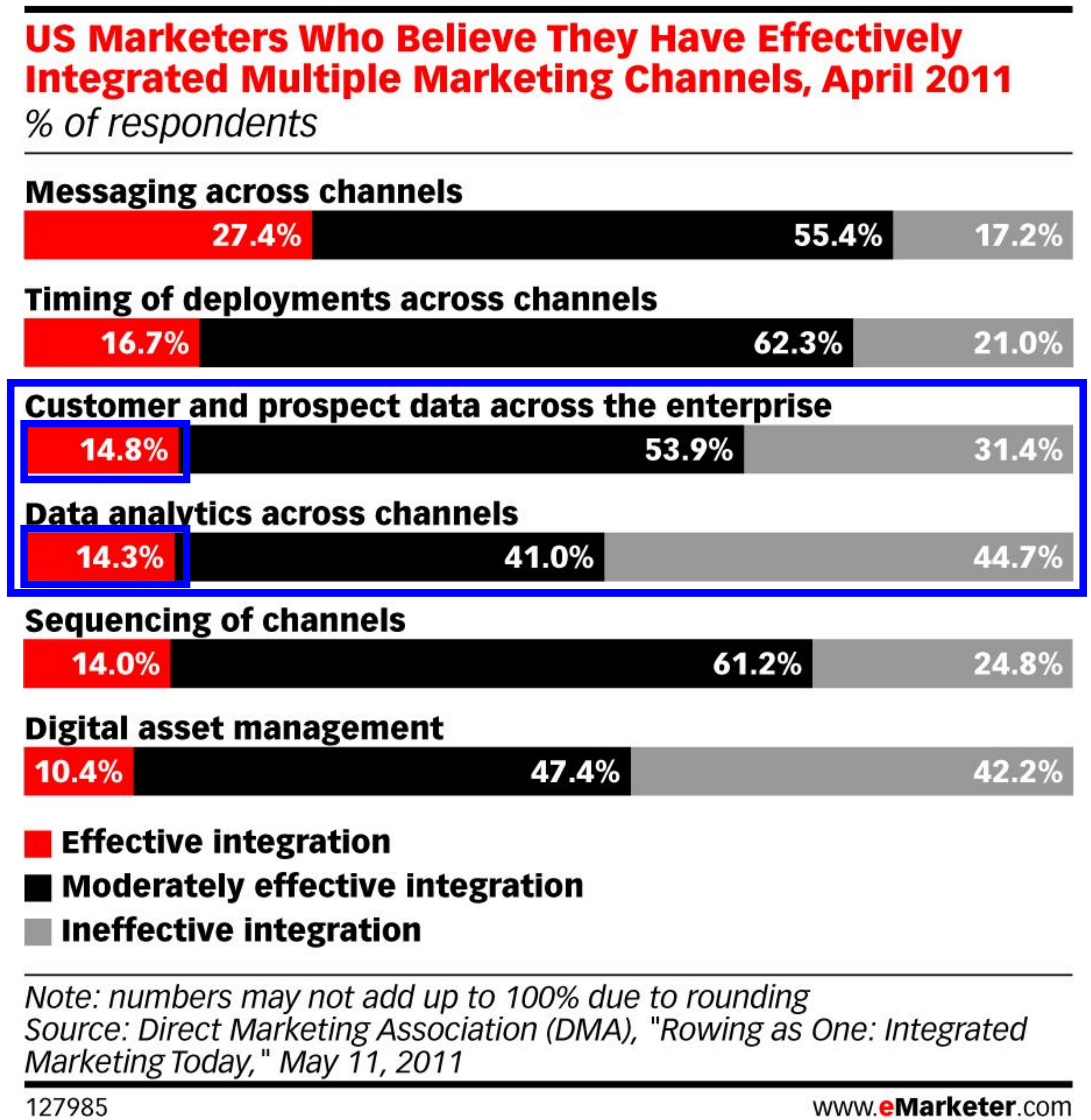
Favorability Towards OBA Concept



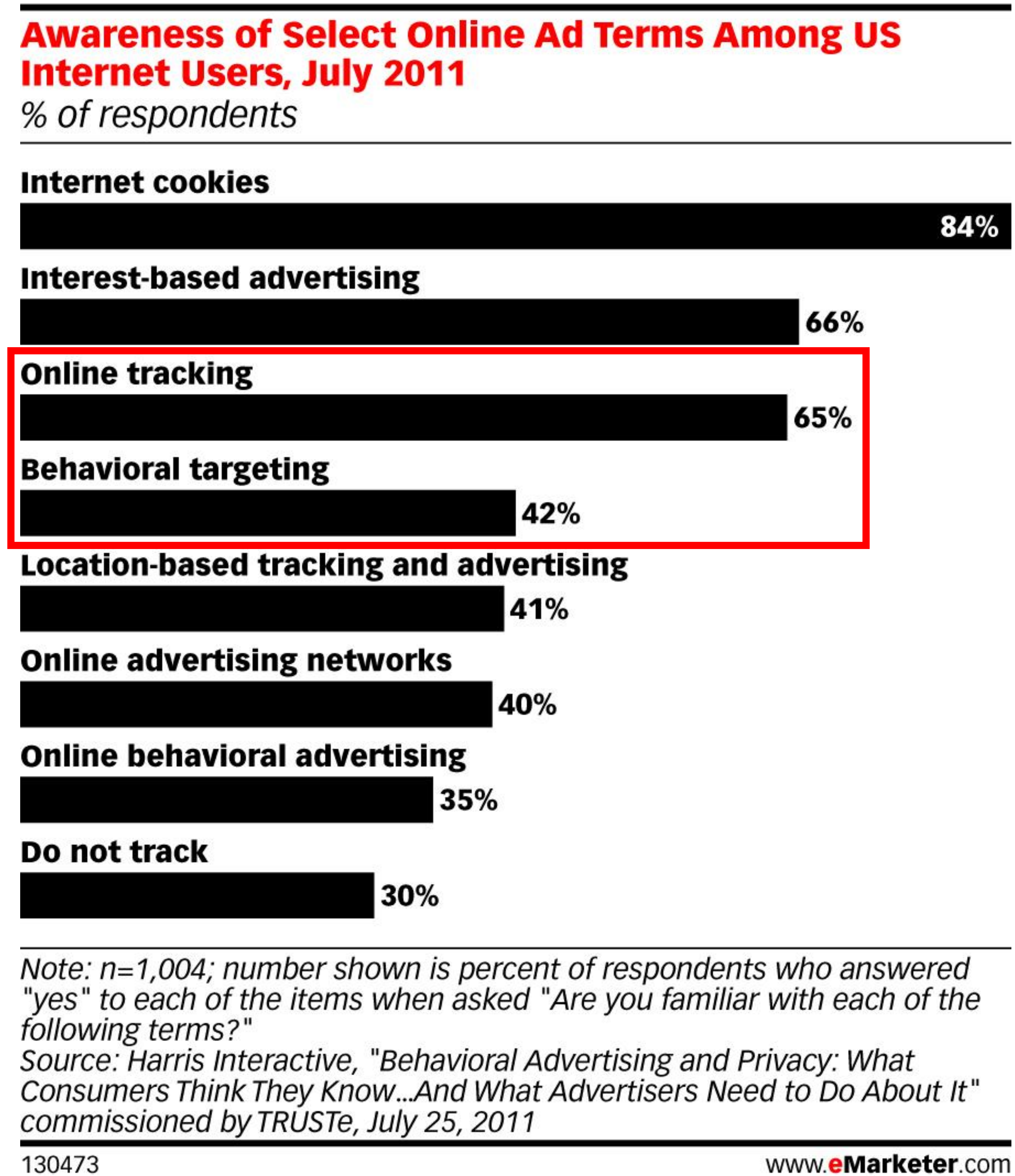
BASE: Total Qualified Respondents (n=1004)
Q710 Are you aware that some advertisers and websites track your browsing activities and show you ads deemed relevant based on your browsing history? This is commonly referred to as Online Behavioral Advertising.

BASE: Total Qualified Respondents (n=1004)
Q715 How do you feel about Online Behavioral Advertising as described above?

Very few marketers believe they've effectively **integrated data** across their company or channels



RTB's audience targeting raises hackles of privacy advocates, and users are ever more aware of being tracked



Estimating Response Rates in Display Advertising through Multi-Hierarchy Smoothing

Deepak Agarwal* Y! Research, Santa Clara, USA

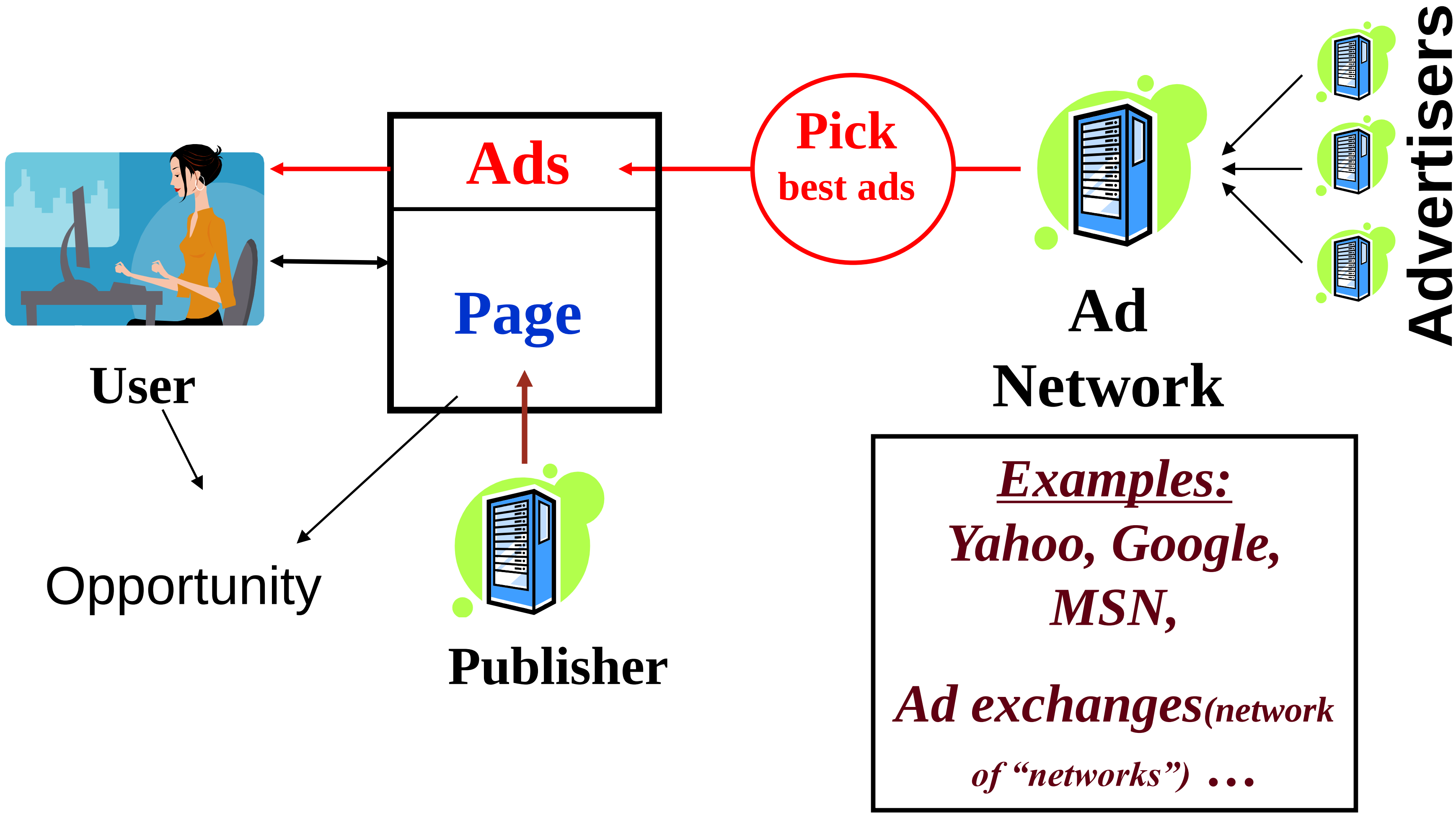
Nagaraj Kota
Y! Labs, Bangalore, India

IISA conference 2011, Raleigh, 23rd April, 2011

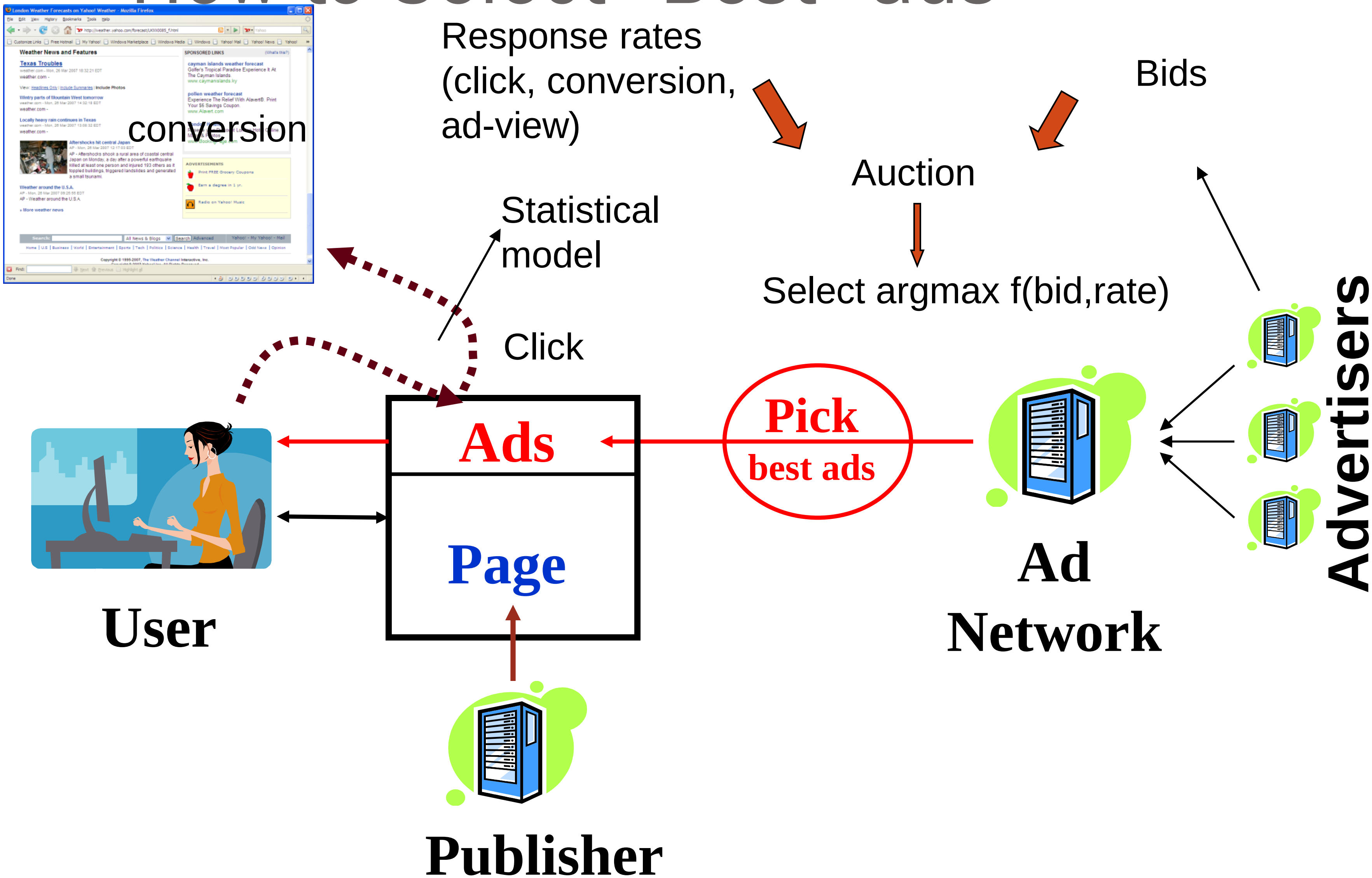
Agenda

- Motivating Example : Computational Advertising
- Problem Definition : Predicting response rates of rare events by exploiting multiple hierarchies
- Our Log-linear model for multiple hierarchies (LMMH)
- Scalable model fitting in a map-reduce framework
- Experiments : Data from Right Media Ad Exchange
- Summary

Computational Advertising: Matching ads to opportunities

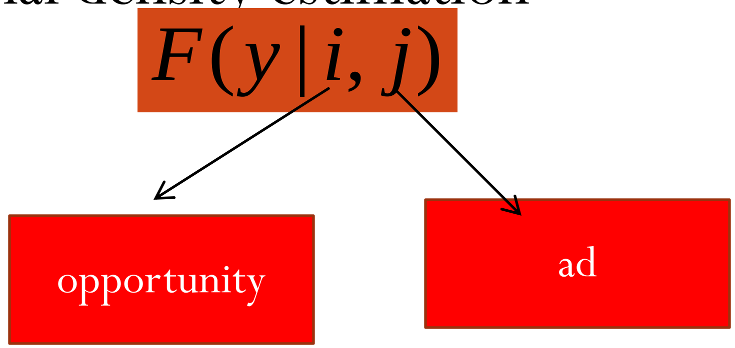


How to Select “Best” ads



Statistical Issues in Conducting Auctions

- f(bid, rate) ---- rate is unknown, needs to be estimated
- Goal: maximize revenue, advertiser ROI
- High dimensional density estimation



- Response obtained through interaction among few heavy-tailed categorical variables (opportunity and ad)
  - #levels : could be millions and changes over time

Other Structure in our Data

- Covariates available for both opportunity and ad
  - Opportunity: Publisher content type, user demographics,…
  - Ad: Industry, text/video, text (if any)

- Hierarchically organized
  - Publisher hierarchy: URL → Domain → Publisher type
  - Geo hierarchy for users
  - Ad hierarchy: Ad → Campaign → Advertiser
- Past empirical analysis (Agarwal et al 2007)
  - Hierarchies: Induces residual correlations
    - E.g. Residual rates of sporty ads shown to Palo Alto users

Statistical Issues

- Key: Modeling residual correlations in residual response rates when interacting variables are organized hierarchically (DAG is fine, Tree not needed)
- High dimensional
  - Data shown here: 100M "cells" from ~ hundered billion auctions
- Data sparsity
  - Large number of zeroes, extreme variation in #Tries
    - Small sample size corrections essential
- Model parsimony
  - Cannot store a large model in our ad-servers, parsimony along with accuracy important
- Temporal smoothing

Model Setup

baseline

$$p_{ijc} = B(\boxed{x_{i,}}\ \boxed{x_{c,}}\ \boxed{x_j})\ \lambda_{ij}$$

residual

i j

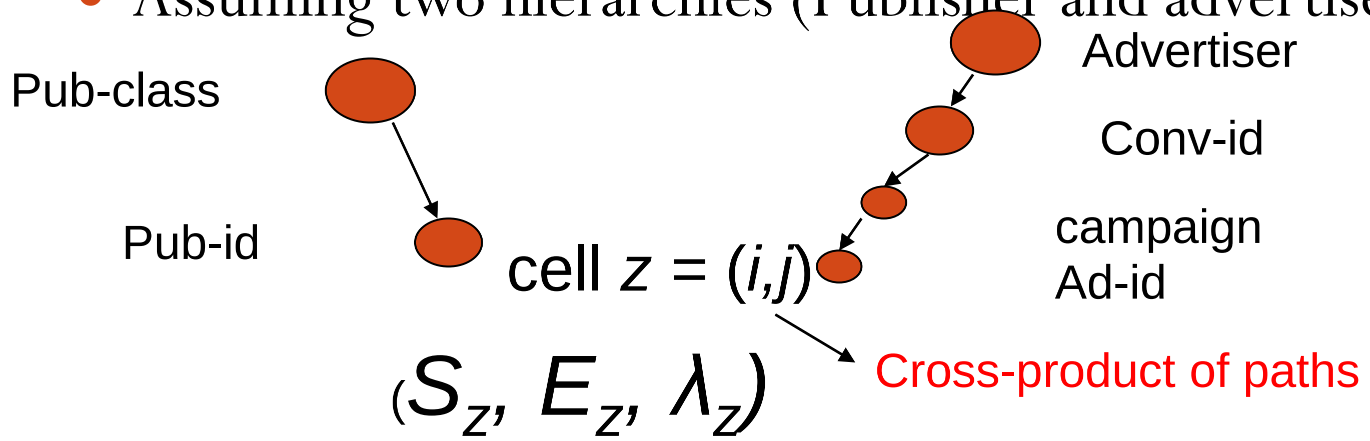
$E_{ij} = \sum_c B(x_i, x_c, x_j)$ (Expected Success)

$S_{ij} \sim \text{Poisson}(E_{ij} \lambda_{ij})$

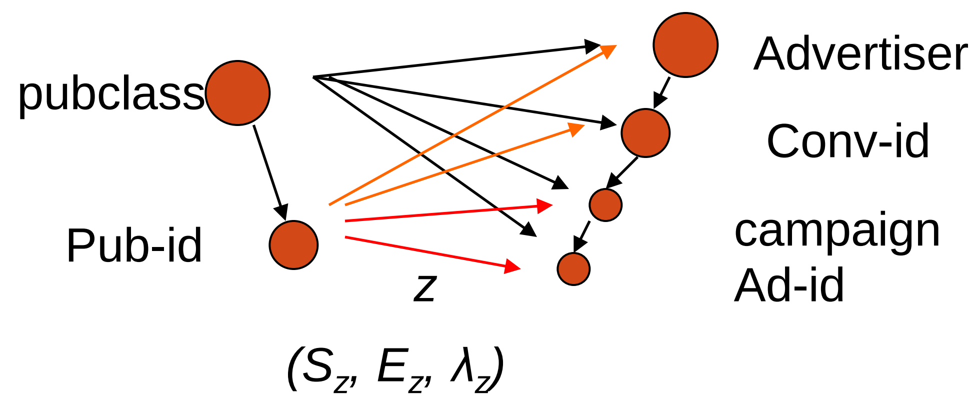
Obviously, MLE does not work

Hierarchical Smoothing of residuals

- Assuming two hierarchies (Publisher and advertiser)



$$\lambda_z = \prod_{s=1}^{m} \prod_{t=1}^{n} \phi_{i_s, j_t}$$



Spike and Slab prior on node States

- Prior on node states: IID Spike and Slab prior

$$\pi(\phi; a, P) = P1(\phi = 1) + (1 - P)\mathrm{Gamma}(\phi; 1, 1/a)$$

- Encourage parsimonious solutions
  - Several cell states have residual of 1
- We choose P = .5 (and choose "a" by cross-validation)
  - a – psuedo number of success

Model Fitting to find posterior mode

- Find a solution that maximizes

$$l(\phi) + \sum_{ij} log(\pi(\phi_{ij}; a, P))$$

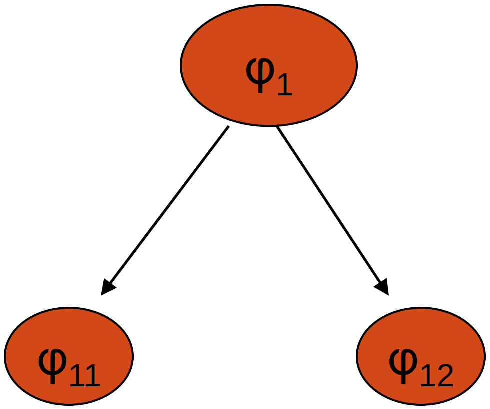
$$[\phi_k | \phi_1^t, \cdots, \phi_{k-1}^t, \phi_{k+1}^{t-1}, \cdots, \phi_M^{t-1}, \mathrm{Data}]$$

Conditional mode – closed form

- Reduces to computing 1-d conditional modes

$$[S|E^*, \phi] \sim \text{Poisson}(E^*\phi)$$
$$[\phi] \sim \pi(\phi; a, P)$$

- E* = Adjusted eSucc aggregating statistics on all paths that include the node being updated



$$\text{Poisson}(S_1, E_1^*\phi_1)\pi(\phi_1) \text{ where } E_1^* = \phi_{11}E_{11} + \phi_{12}E_{12}$$

Conditional model --- closed form

- Threshold estimator : conducts hypothesis test

$$\begin{cases} \tilde{\phi} = 1 \; if \, Q - \log(g(\phi_m; S + a, E^* + a) - g(1; S + a, E^* + a)) \\ \qquad = \phi_m \; otherwise \end{cases}$$

where

$$Q = log\frac{Poisson(S, E^*)}{NB(S; 1, E^*, a)} + log(\frac{P}{1 - P})$$

$$\tilde{\phi}_m = (S + a - 1)/(E^* + a)$$

Scalable Map-reduce implementation

Algorithm 1 Psuedocode for map-reduce implementation

Initialize the global constant a, the state variables $\phi_0^0 = 1$.
Iterate until convergence,
Iterate t over the conjunction of paths $z = (i, j)$ in the data,
Iterate over all node pairs (i_s, j_t), indexed by $k = 1, \ldots, M$. Note that $(k-1)$ is M from $(t-1)$'th iteration, when $k = 1$ and $t > 1$. For 1'st iteration with k=1, $(k-1)$ would be treated as record id and the corresponding parent node state variable as 1.

$$\begin{aligned} Map: (k-1, data, S_z, E_z^*) \bowtie (k-1, \phi_{k-1}^t) \\ \rightarrow (k, \{data, S_z, E_z^* \phi_{k-1}^t\}) \\ Reduce: (k, \{data, S_z, E_z^* \phi_{k-1}^t\}) \bowtie (k, \phi_k^{t-1}) \\ \rightarrow \left\{ \begin{array}{c} (k, \{data, S_z, E_z^* \phi_{k-1}^t / \phi_k^{t-1}\}) \\ (k, \phi_k^t) \end{array} \right\} \end{aligned}$$

where, ϕ_k^t is computed for key k using $\sum S_z, \sum E_z^* \phi_{k-1}^t / \phi_k^{t-1}$, using mode formula described in Theorem 1.

Our data --- Ad- exchange (RightMedia)

- Advertisers participate in different ways
  - CPM (pay by ad-view)
  - CPC (pay per click)
  - CPA (pay per conversion)

- To conduct an auction, *normalize* across pricing types
  - Compute eCPM (expected CPM)
    - Click-based ---- eCPM = click-rate*CPC
    - Conversion-based ---- eCPM = conv-rate*CPA

Data (2)

- Two kinds of conversion rates
  - Post-Click --- conv-rate = click-rate*conv/click
  - Post-View --- conv-rate = conv/ad-view
    - Not important for this talk

- Three response rate models
  - Click-rate (CLICK), conv/click (PCC),
  - post-view conv/view (PVC)

Multiple (K) hierarchies

- Product of ${}^{K}C_{2}$ pair wise hierarchies
- Primarily done to deal with data sparseness
- Ongoing research

  - Find small subset of 3-way, 4-way combinations that are important

  - Main idea is to adjust for multiple tests by "shrinking" Observed/expected from all 2-factor models to detect significant higher order interactions

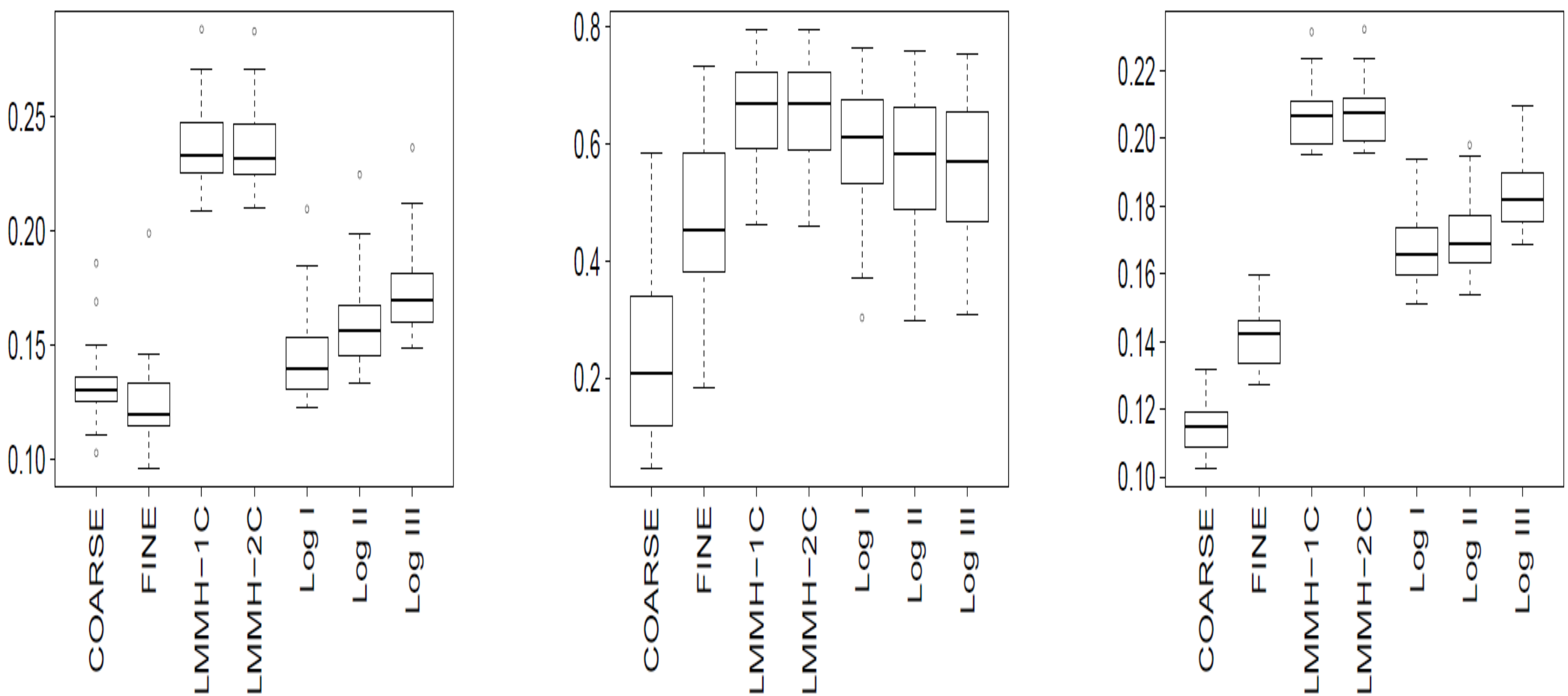
Datasets : RightMedia

- CLICK [~90B training events]
- Post Click Conversion(PCC) (~.5B training events)
  - Conversion only through click
- PVC – Post-View conversions (~7B events)
  - Cookie gets augmented with pixel and triggers success
- Covariates
  - Age, gender, ad-size, pub-class, user fatigue
  - 2 hierarchies (publisher and advertiser)
- Two baselines
  - Pubid x adid [FINE] (no hierarchical information)
  - Pubid x advertiser [COARSE] (collapse cells)

Other baselines: variations of logistic regression

- Log I
  - Main effects using leaf nodes as covariates (size ~ 200-300k)
- Log II
  - Main effects but using both leaf and non-leaf (size ~ 300-700k)
- Log III
  - Interactions added by using large number of sparse random projections of original interaction covariates
    - Added roughly 500K projections
- All three variations run on Map-Reduce using Y! code

Accuracy: Average test log-likelihood



(a) PCC (b) PVC (c) CLICK

Model Parsimony

- With spike and slab prior
- Parsimony

data	#cells	#retained
PCC	~81M	4.4M
PVC	~6M	35K
CLICK	~16.5M	150K

Some rough computation time

- CLICK : 135 mins, 50 reducers
- PVC : 123 minutes, 25 reducers
- PCC: 109 minutes, 20 reducers

- LogI, II, III (CLICK) : 4, 6,7 hours; 80 reducers
  - PVC: 3,4.5,5 hours with 40 reducers
  - PCC: 4.5, 8, 9 hours with 80 reducers

How do we estimate variance for a parent?

- Simple example: CTR of an ad in different zip-codes
- (si, ti): i=1,…,K; emCTRi = si / ti

$$E(s_i / t_i) = p_i$$

$$\text{Backoff depends on } \mathrm{Var}(p_i)$$

- Var(emCTRi) good measure of Var(pi)?
  - Not quite, empirical estimates not good for small ti and(or) si

$$Var(s/t) = Var(E(s/t) \mid p) + E(Var(s/t) \mid p)$$
$$= Var(p) + E(p(1-p)/t)$$
$$\longrightarrow Var(p) \text{ as } t \longrightarrow \infty$$

- Hence, using variance in empirical CTRs can lead to overestimates in variance, will reduce the amount of back-off

How do we estimate variance in unbiased way?

- Simple example: CTR of an ad in different zip-codes

  $(s_i, t_i): i=1,\ldots,K;\ emCTR_i = s_i / t_i$

- Use a model

$$s_i \sim \text{Binomial}(t_i, p_i)$$

$$E(p_i) = \mu; Var(p_i) = \sigma^2$$

- Variance among true CTRs can be estimated in an unbiased way using MLE/MOM

  (Agarwal & Chen, *Latent OLAP, SIGMOD 2011*)

- For more complex data at multiple resolutions, better statistical models needed