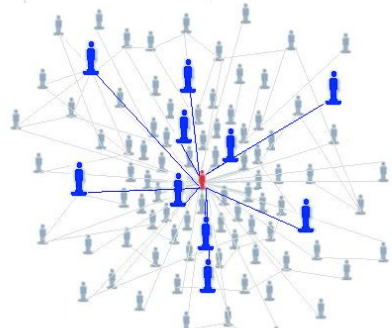
NYU STERN
NEW YORK UNIVERSITY · LEONARD N. STERN SCHOOL OF BUSINESS

Machine Learning for Display Advertising

Foster Provost
New York University and Coriolis Ventures

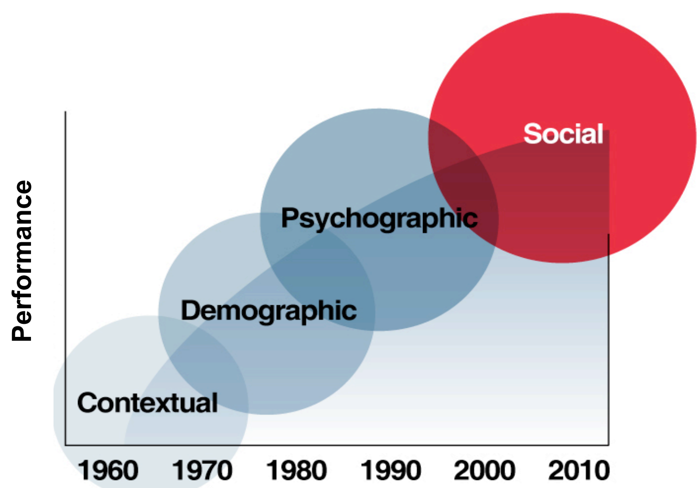
Keynote for MLOAD-2010 -- December 10, 2010



Work done with Brian Dalessandro, Rod Hook, Xiaohan Zhang, Alan Murray, and Claudia Perlich
Thanks to Media6Degrees
Thanks to the Marketing Science Institute for a research award
Thanks to INFORMS for the 2009 Design Science Award

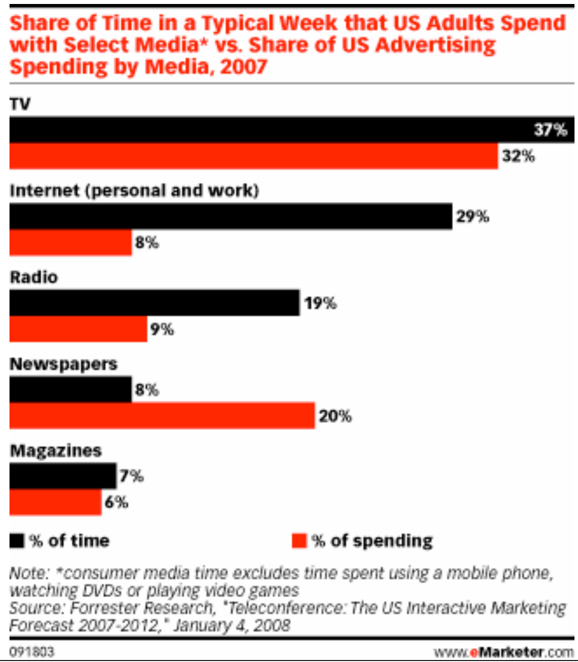
The organizers asked me to talk a little about Media6Degrees…

The Idea: Social Targeting for Online Advertising

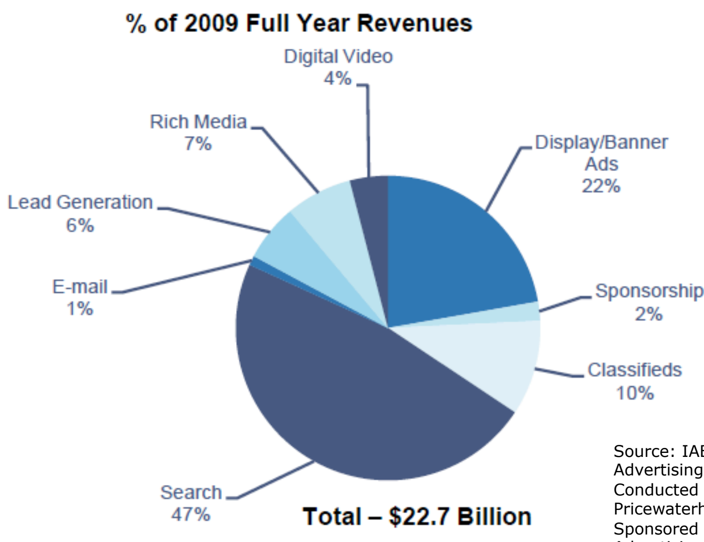


media6degrees

Current ad spending seems disproportionate…



Online Advertising Spending Breakdown (2009)



Source: IAB Internet Advertising Revenue Report Conducted by PricewaterhouseCoopers and Sponsored by the Interactive Advertising Bureau (IAB) April 2010

Two different goals for display advertising

- Drive conversions (short term)
- <u>Brand</u> advertising (longer term)

Online Brand Advertising

- goal: to deliver brand message to selected audience
- contrast with "direct marketing" online advertising
  - for brand advertising, goal is not necessarily clicks or online conversions
- key: selecting audience
  - example strategy (traditional): find audience based on published content (tv shows, magazines) or location (billboards, etc.)
- traditional brand advertising strategy applies on line:
  - premium display slots or remnants (e.g., on espn.com, etc.)
  - contextual targeting (e.g., Google AdSense)
- alternative strategy: *<u>identify **members** of the target audience</u> and target them anywhere on the web (e.g., bid for them on ad exchanges – the non-premium display market)*

- Non-premium display ad market predicted to grow significantly faster than the rest of online advertising (e.g., sponsored search, premium display, contextual)
  - (Coolbrith 2007)
  - largely due to the stabilization of the technical ad-serving infrastructure based on the consolidation into a small number of ad exchanges (e.g., DoubleClick, Right Media)

- There is evidence that display brand advertising increases purchases (online and offline), and improves search advertising as well (Manchanda et al. 2006, Comscore 2008, Atlas Institute 2007, Fayyad personal communication, Klaassen 2009, Lewis & Reiley 2009)
  - other (older) work shows display ads lead to increased ad awareness, brand awareness, purchase intention, and site visits (see cites in Manchanda)

Two different goals for display advertising

- Drive conversions (short term)
- <u>Brand</u> advertising (longer term)

- Both are important
  - (in the off-line world most ad spending is on brand ads)
  - Our KDD-2009 paper focused on online brand advertising
  - Today I'll meld the two together

- What I'm <u>**not**</u> interested in is <u>clicks</u> on display ads
  - we'll return to that later

Main points for this morning

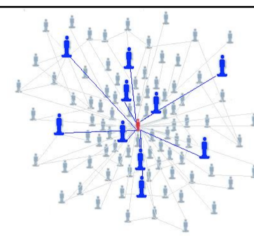
1. Machine learning can be used as the basis for <u>effective</u>, <u>***privacy friendly***</u> targeting for online advertising

2. Important to consider carefully the <u>target</u> variable used for training

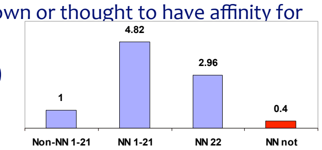
3. Question: should machine learning researchers be spending more time considering the <u>*effectiveness*</u> of advertising?

Prior work:

Social network targeting



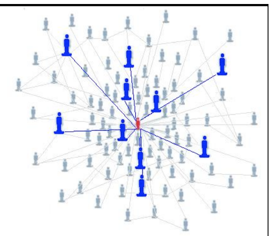
- Defined Social Network Targeting
  --> cross between viral marketing and traditional
    - target "network neighbors" of existing customers
    - based on direct communication between consumers
    - this could expand "virally" through the network without any word-of-mouth advocacy, or could take advantage of it.
- Example application:
  - Product: new communications service
  - Firm with long experience with targeted marketing
  - Sophisticated segmentation models based on data, experience, and intuition
    - e.g., demographic, geographic, loyalty data
    - e.g., intuition regarding the types of customers known or thought to have affinity for this type of service
- <u>Results</u>: tremendous lift in response rate (2-5x)

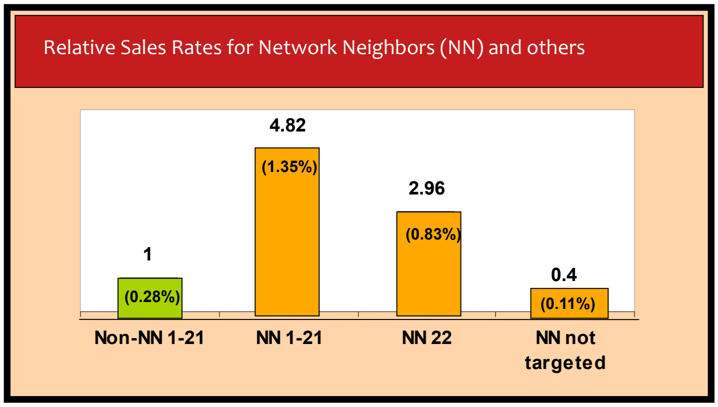


Hill, Provost, and Volinsky. "Network-based Marketing: Identifying likely adopters via consumer networks. " *Statistical Science* 21 (2) 256–276, 2006.

Sales rates are substantially higher for network neighbors

(Hill, Provost, Volinsky Stat. Sci. 2006)





1-21 are targeted marketing segments;
22 comprises NNs not deemed good targets by traditional model

Is such "guilt-by-association" targeting justified theoretically?

Thanks to (McPherson, et al., 2001)

- *Birds of a feather, flock together*
  – attributed to Robert Burton (1577-1640)
- (*People) love those who are like themselves*
  -- Aristotle, *Rhetoric* and *Nichomachean Ethics*
- *Similarity begets friendship*
  -- Plato, *Phaedrus*
- *Hanging out with a bad crowd will get you into trouble*
  -- Foster's Mom

December 6, 2007

Apologetic, Facebook Changes Ad Program

By LOUISE STORY

Mark Zuckerberg, founder and chief executive of the social networking site Facebook, apologized to the site's user
yesterday about the way it introduced a controversial new advertising feature last month.

Facebook also introduced a way for members to avoid the feature, known as Beacon, which tracks the actions of i
members when they use other sites around the Internet.

Mr. Zuckerberg's apology — in the form of a blog post on Facebook — followed weeks of criticism from members,
groups and advertisers.

"I'm not proud of the way we've handled this situation, and I know we can do better," Mr. Zuckerberg wrote.

Facebook has also been meeting with advertising agencies in recent days and discussing their concerns about Bea
according to one executive who was invited.

Facebook originally presented Beacon to the advertising community as an opt-in program that its members woul
choose to use. It planned to sell ads alongside the messages sent to people's friends about their purchases and act
on other sites. Some advertisers like Coca-Cola have expressed surprise that Beacon then required users to take a
they did not want the messages sent out.

"Privacy" online?

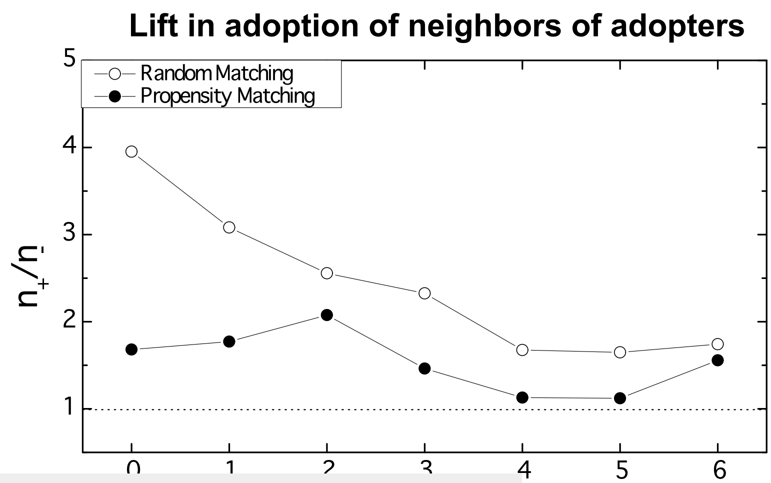
Where would we like firms to operate on the spectrum between the two unacceptable extremes:

→Are there points between the extremes that give us acceptable tradeoffs between "privacy" and efficacy?

I'll discuss an attractive one. ML provides many possibilities.
Room for more research…

Seeming adoption influence between network neighbors can be largely explained by homophily

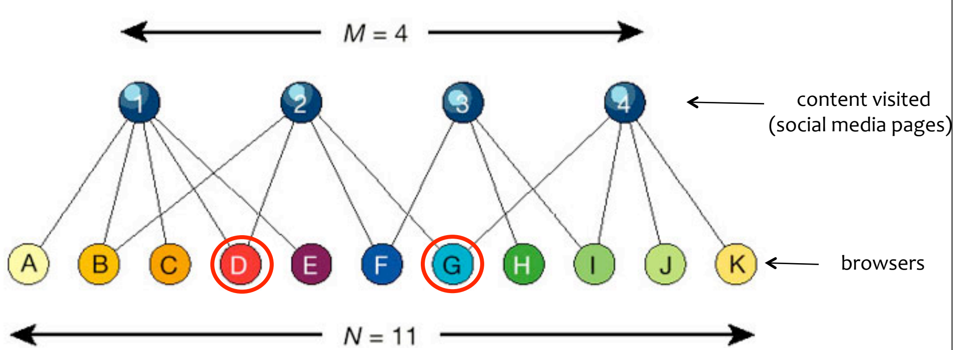
from (Aral et al. PNAS 2009)



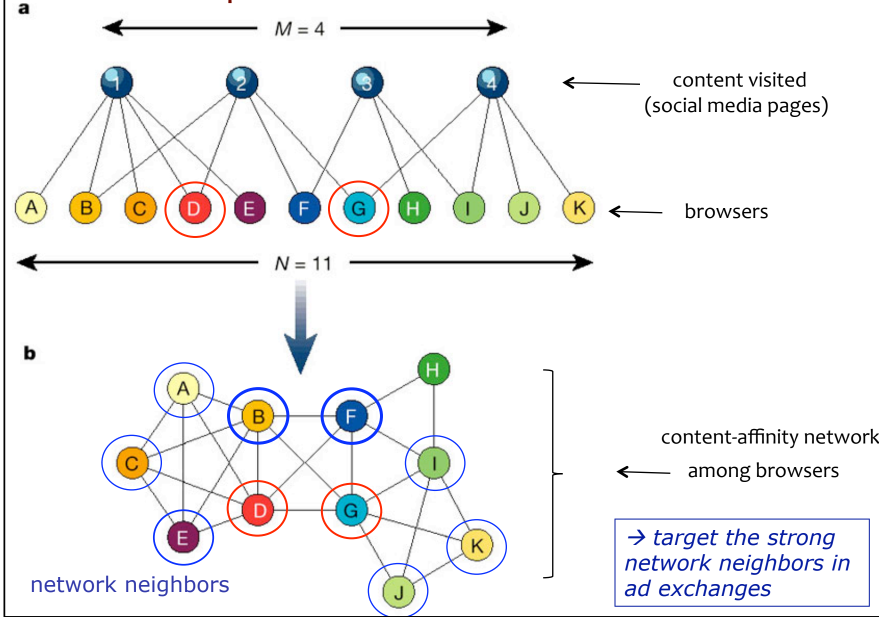
Social connections reveal deep similarity
- profiles, interests, attitudes,...

Media6's social targeting is different...

Doubly-anonymized bipartite content-affinity network

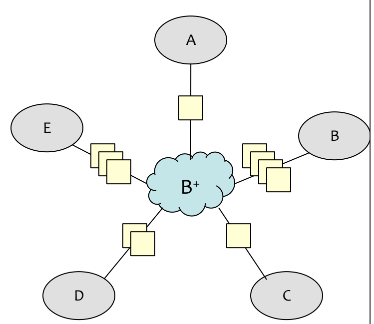


From doubly-anonymized bipartite content-affinity network to quasi-social network



Some brand proximity measures

- POSCNT
  - number of unique content pieces connecting browser to B^+
- MATL
  - maximum number of content pieces through which paths connect browser to some particular action taker (i.e., seed node in B^+)
- minEUD
  - minimum Euclidean distance of normalized content vector to a seed node
- maxCos
  - maximum cosine similarity to a seed node
- ATODD
  - "odds" of a neighbor being an action taker (i.e., seed node in B^+).



Plus, multivariate statistical models using these as features

See: **Audience Selection for On-line Brand Advertising: Privacy-friendly Social Network Targeting.** Provost, F., B. Dalessandro, R. Hook, X. Zhang, and A. Murray.. In *Proceedings of the Fifteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD 2009).

Multivariate model

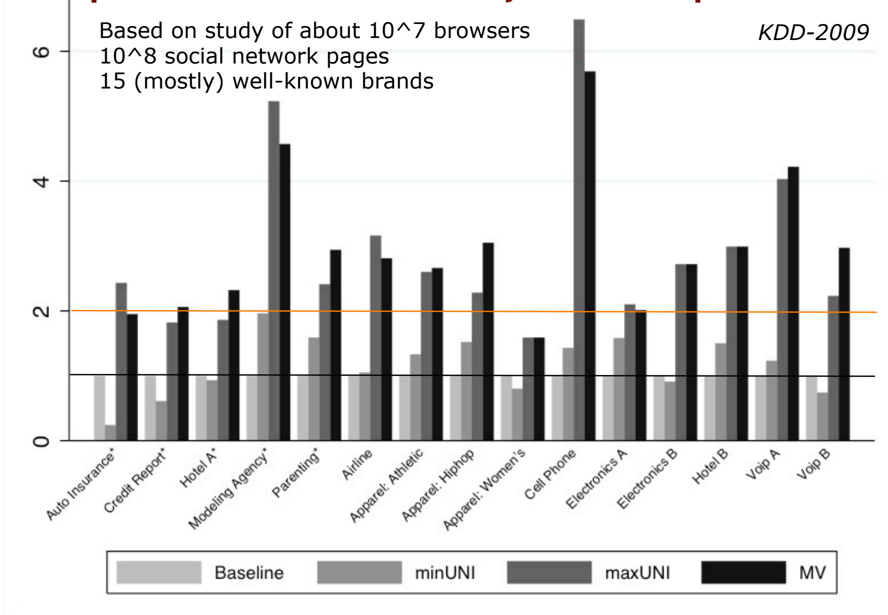
- For each browser b_i, a feature vector φ_{bi} can be composed of the various brand proximity measures
- The different evidence can be combined via a ranking function $f(\varphi_{bi})$
- We let f(.) be a multivariate logistic function, trained via standard MLE logistic regression (not regularized)
- Training is based on a held-out training set

Initial Study: Data

KDD-2009

- a sample of about 10 million anonymized browsers
- all of their observed visits to social media content over 90 days (here: from several of the largest SN sites)
- bipartite graph:
  - $10^7 \times 10^8$ with ~2.5×10^8 non-zero entries
- quasi-social network:
  - 10^7 nodes with 20-40 neighbors each (on average)
- more than a dozen well-known brands:
  - Hotel A, Hotel B, Modeling Agency, Cell Phone, Credit Report, Auto Insurance, Parenting, VOIP A, VOIP B, Airline, Electronics A, Electronics B, Apparel:Athletic, Apparel:Women's, Apparel:HipHop
  - on average ~100K seed nodes per brand

Example result from initial study: Lift for top 10% of NNs



Network neighbors often show similar demographics

For one campaign (Cell Phone) we asked Quantcast.com for demographic profiles of the seed browsers and their close network neighbors:

Demographic	Seeds	Neighbors
Gender	Female	Female
Ethnicity	Hispanic	Hispanic
Age	Young	Young
Income	Low	Low
Education	No College	No College

but this belies the advantage of network neighbor targeting: all people who share demographics don't share interests AND all people who share interests don't share demographics

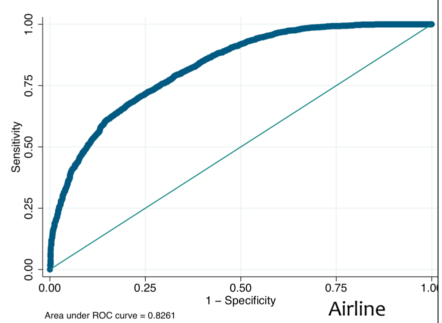
Social vs. Quasi-Social

The content-affinity network embeds a friendship network?

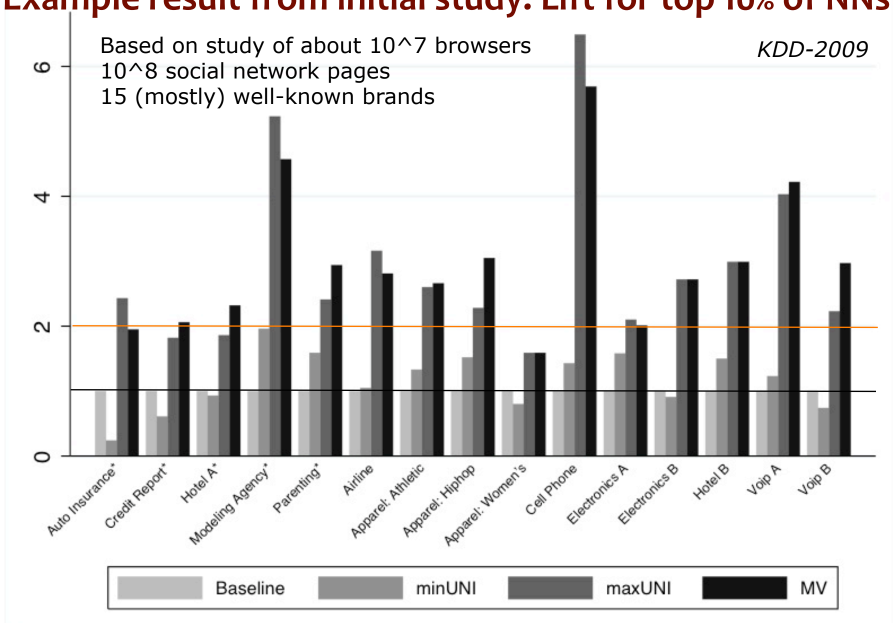
- estimate each browser's home page based on techniques analogous to author id based on citations (Hill & Provost, 2003)
- estimate "friends" to be those who visit each other's home page
- Ask: do brand proximity measures rank brand actors' friends highly?

- F-AUC measures probability that a known-friend is ranked higher than a browser not-known-to-be-friend

Brand	F-AUC on all B	F-AUC on *N* only
Hotel A	0.96	0.79
Modeling Agency	0.98	0.84
Credit Report	0.93	0.79
Parenting	0.94	0.80
Auto Insurance	0.97	0.81
…		
15 Brand Average	0.96	0.81

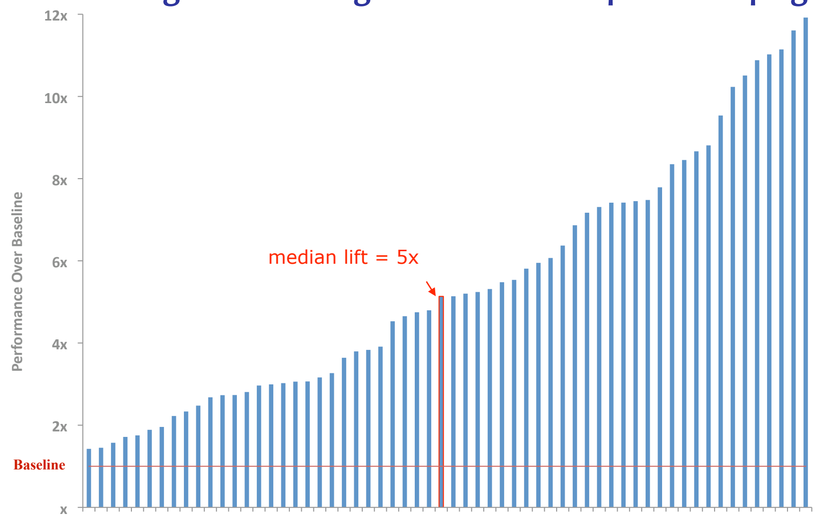


Example result from initial study: Lift for top 10% of NNs



Fast forward 1.5 years… "In vivo" performance (Media6Degrees production targeting)

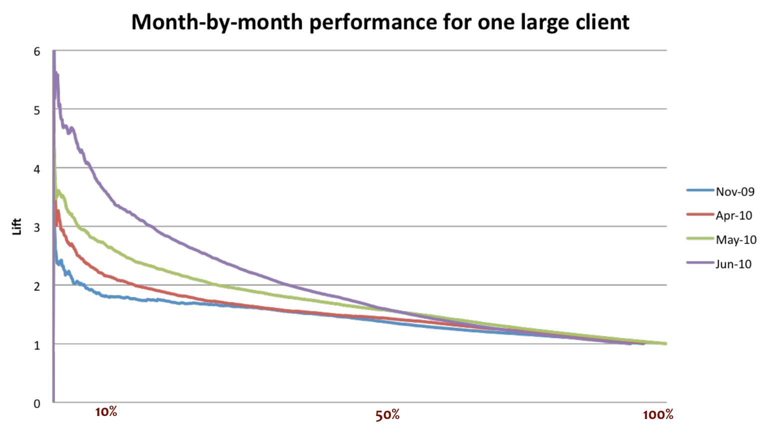
Lift for strong network neighbors across sample of campaigns



Each Bar = Lift for one Media6Degrees client during the month of December

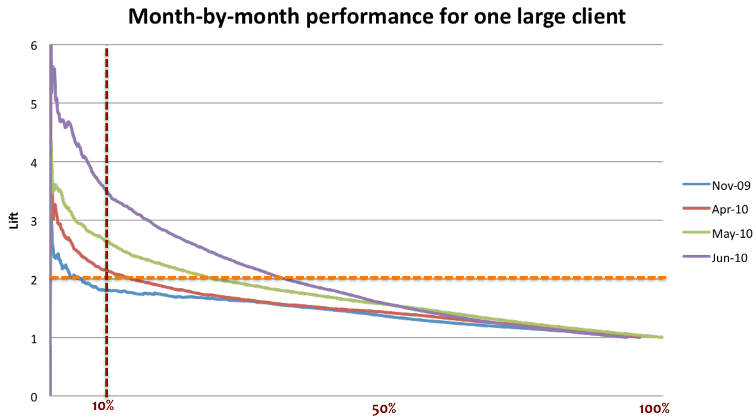
Lift = freq. that targeted M6D browsers visit site / freq. that baseline browsers visit site

Performance improves <u>substantially</u> with more data..



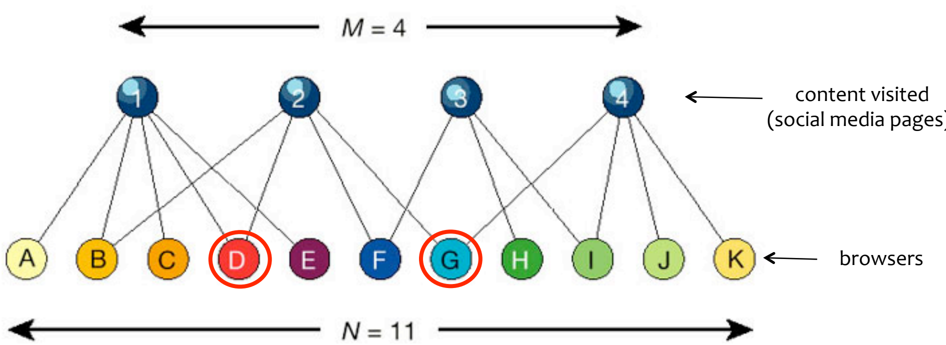
- Shows lift for a particular size targeted population (%)
- Left-to-right decreases targeting threshold

Performance improves substantially with more data..



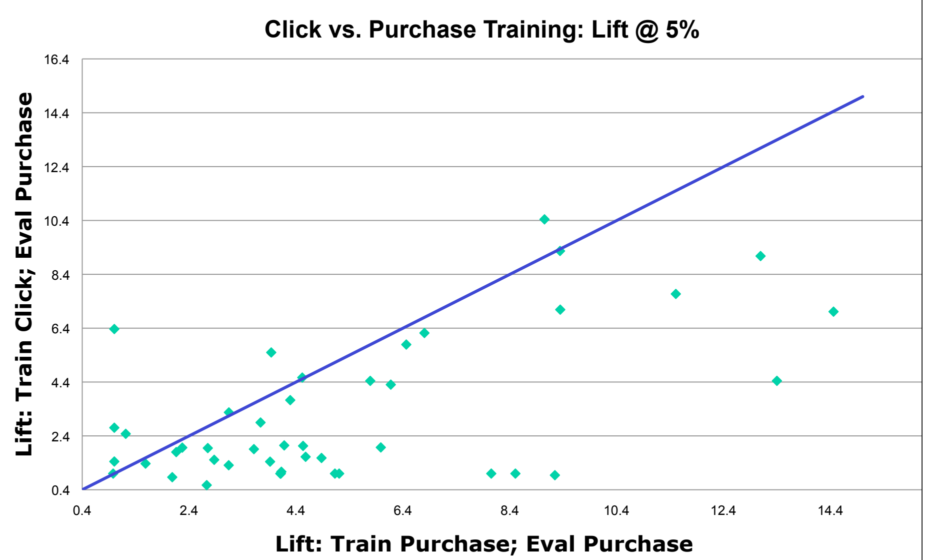
- The orange horizontal line intersects the curves at the % of the population we can reach to get a 2x lift.
- The maroon vertical line intersects the curves at the lift multiple for the best 10% of each population.

Potential stumbling block:
… what do those red circles represent again?

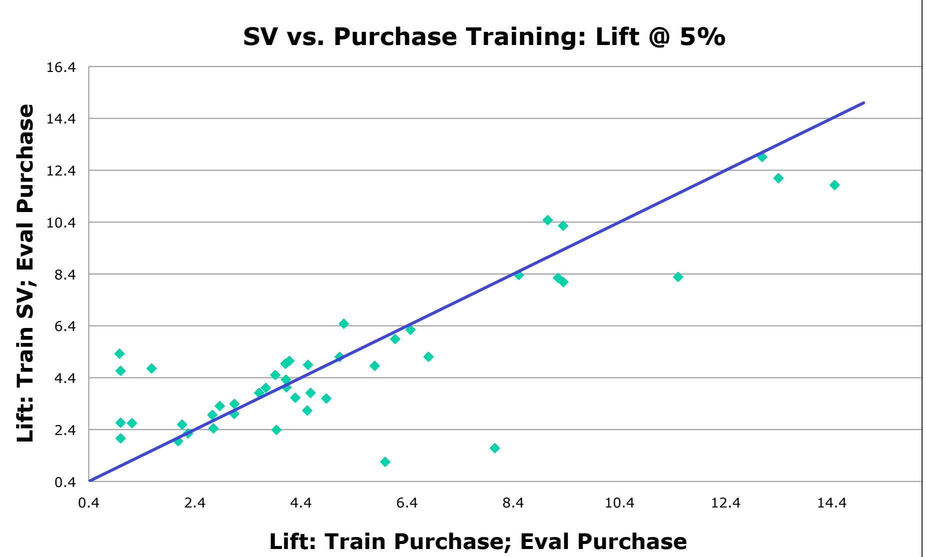


If there are not very many conversions,
how can we build effective predictive models?

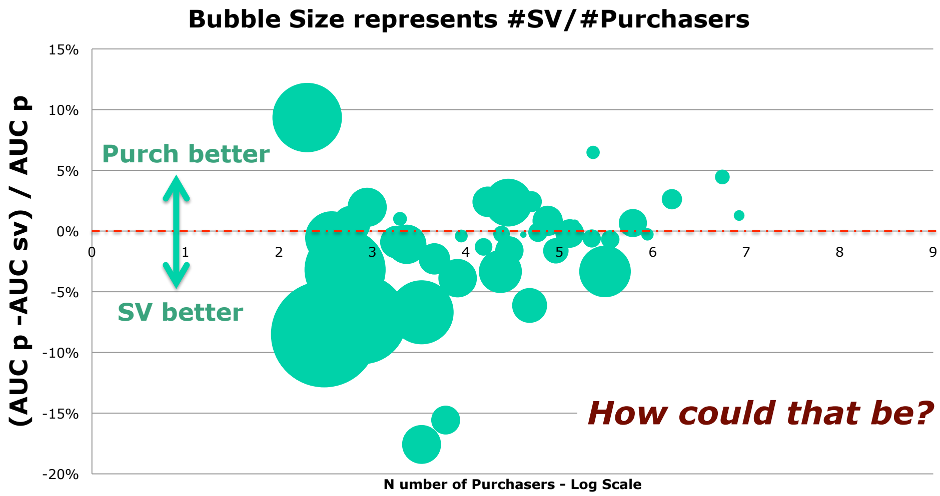
What not to do: use clicks as a surrogate for conversions



But there *is* a good, easy-to-obtain surrogate: site visits



Generally, site visits is better than conversions when there are relatively few conversions



Summary of main points

1. Machine learning can be the basis for effective *privacy friendly* targeting for online advertising – effective from several different angles, clearly improves with more data

2. Important to consider carefully the target used for training *– conversions are good if you can get them; site visits can be a surprisingly good surrogate; clicks generally are not a good surrogate*

3. Question: should machine learning researchers be spending more time considering the *effectiveness* of advertising? *– initial evidence shows surprisingly strong influence of seeing an online advertising impression. This deserves more study.*