

---

# Digital Advertising: Going from Broadcast to Personalized Advertising

**James G. Shanahan**  
*Independent Consultant*  
*Church and Duncan Group Inc.*

*EMAIL: James\_DOT\_Shanahan\_AT\_gmail.com,*

**Machine Learning in Online Advertising MLOAD Workshop**  
**NIPS 2010, Whistler**

December 12, 2010

*NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan*

1

## Abstract

- Online advertising is a form of promotion that uses the Internet and World Wide Web for the expressed purpose of delivering marketing messages to attract customers. Examples of online advertising include text ads that appear on search engine results pages, banner ads, in-text ads, or Rich Media ads that appear on regular web pages, portals or applications. Since its inception over 15 years ago, online advertising has grown rapidly and currently accounts for 13% of the overall advertising spend (which is approximately \$700 billion worldwide). A large part of the more recent success in this field has come from the following key factors:
- Personalization: offline advertising (via broadcast TV, radio, newspaper etc.) is largely a broadcast form of communication whereas digital advertising is much more targeted and thus enables a personalized, and possibly informative, message to consumers.
- Interactivity: internet advertising is becoming increasingly interactive with the advent of new forms of advertising such as social advertising; this enables advertisers and consumers to operate in a more conversant manner.
- Engagement: consumers are spending more time online than with any other form of media thereby enabling a broader reach and deeper connection with consumers.
- Explainability: advertisers are beginning to understand their consumers better.
- This shift in focus in digital advertising from location (i.e., publisher web pages) to personalization has brought with it numerous challenges some of which have received a lot of research attention in the DIK communities (Data, Information and Knowledge communities) over the past 10-20 years. In this talk I will review, along the dimensions outlined above, some of these key technical problems and challenges that arise when advertising becomes personal. This will be done within the context of the elaborate (and ever-evolving) ecosystems of modern day digital advertising where one has to capture, store, and process petabytes of data within the constraints of a, sometimes, sequential workflow. The ultimate goal is to provide millisecond-based decision-making at each step of this workflow that enables customizable and engaging consumer experiences.

*NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan*

2

## Brief Bio James G. Shanahan

- **20 years in the field AI and information management**
  - Principal and Founder, Boutique Data Consultancy
    - Clients include: Digg, AT&Ti, Ancestry, SearchMe, SkyGrid, MyOfferPal
  - Lecturer, University of California Santa Cruz (UCSC)
  - Chief Scientist, Turn Inc. (A CPX ad network → DSP)
  - Principal Scientist, Clairvoyance Corp (CMU spinoff; sister lab to JRC)
  - Research Scientist, Xerox Research
  - Research Engineer, Mitsubishi Group
  - PhD in machine learning (1998), University of Bristol, UK;  
B.Sc. Comp. Science (1989), Uni. of Limerick, Ireland
- **Now: Machine Learning Consultant (San Francisco)**
  - IF *(you have large **data problems** and need a consultant)*  
THEN *{email me at [James.Shanahan AT gmail.com](mailto:James.Shanahan@gmail.com)}*
  - Where **problems**  $\in$  *{web search, online advertising, machine learning, ranking, user modeling, optimization, social networks}*

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

3

## Disclaimer

- **The Author retains all rights, including copyrights and distribution rights.**
- **No publication or further distribution in full or in part permitted without explicit written permission from the author**
- **Living vicariously!**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

4

## Outline

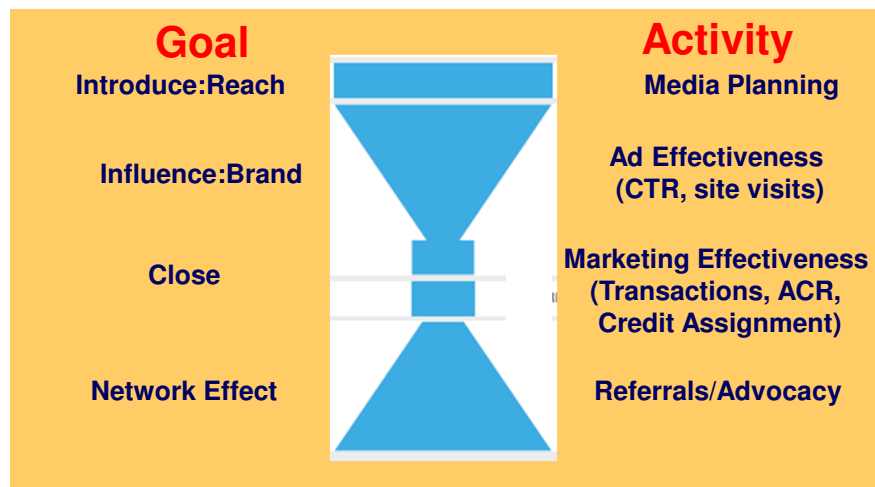
- **Background: digital advertising, the 0.2% opportunity (of GDP)**
- **3<sup>rd</sup> Generation DA (Personalization)**
  - Explainability: trading desks for the advertiser
  - Behavioral targeting
  - Interactivity: thru social ads
  - Engagement: transformative ad formats
- **Summary**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

5

## What marketers want?

- **Deliver marketing messages and attract customers and sell products/services (long term vs. short term)**



NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

6

## It annoys but it yet it works.....

- **Online advertising is a form of advertising utilizing the Internet and World Wide Web in order to deliver marketing messages and attract customers** [wikipedia.com]
- **Advertising annoys people! Advertising works!**
  - "Half the money I spend on advertising is wasted; the trouble is, I don't know which half." - [John Wanamaker](#), father of modern advertising. [Credit assignment]
  - "I do not regard advertising as entertainment or an art form, but as a medium of information..."; "Ogilvy on Advertising" by [David Ogilvy](#)
- **Goals of Online advertising**
  - A** – Deliver/push an advertiser's message to the consumer interest
  - A+P** – Generate ROI for the advertiser and **P** for the publisher
  - P+C** – Enable ads as a *medium of information* (true in the case of search)!

•Credit Assignment  
•Annoying  
•and there is privacy

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

7

## Advertising makes up ~2% of US GDP

Despite its problems (lack of credit assignment etc.)

- **US GDP = \$14.1 Trillion (Global \$56 Trillion,  $56 \times 10^{12}$ )**
- **US Advertising Spend**
  - ~\$275 Billion across all media (2% of GDP since the early 1900s)
  - ~\$23 Billion in Digital Advertising (8.4% of overall spend)
- **In 2008, Worldwide online advertising was \$65B**
- **I.e., about 10% of all ad spending across all media [IDC, 2008]**

<http://en.wikipedia.org/wiki/Advertising>

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

8



## Why online?: ROI versus CPM

|                  | CPM         | ROI per \$1 Spent |
|------------------|-------------|-------------------|
| Outdoor          | \$1-5       |                   |
| Cable TV         | \$5-8       | \$6.81            |
| Radio            | \$8         | \$8.60            |
| Online           |             | \$19.83           |
| -Display \$5-30  | \$5-\$30    | \$19.78           |
| -Contextual      | \$1-\$5     |                   |
| -Search          | \$1 - \$200 | \$21.84           |
| Social           | \$0.2-\$5   | \$12.57           |
| Mobile           | \$0.2-\$6   | \$7.50            |
| Email            |             | \$44.93           |
| Network/Local TV | \$20        | \$6.81            |
| Magazine         | \$10-30     | \$10.11           |
| Newspaper        | \$30-35     | \$12.77           |
| Direct Mail      | \$250       | \$7.34-\$15.28    |

**TV = 7:1 ROI**  
**Online = 20:1 ROI**

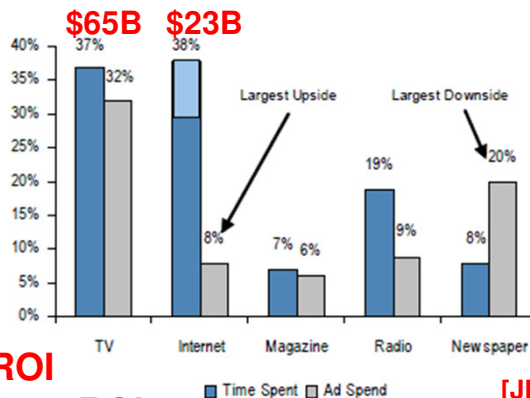
[ROI numbers based on  
 DMA Power of direct  
 marketing 2010 Edition]

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

9

## Big lag in ad dollars spend online

- Typically ad dollars have followed media consumption
- Despite people spending more time on the internet, there is a \$40 billion deficit (in US)



**TV = 7:1 ROI**  
**Online = 20:1 ROI**

[JP Morgan, 2010]

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

10



## Big questions

---

- **\$40 Billion online gap**
- **\$80 billion display gap**

## Why the big lag in ad dollars spend online?

---

- **Advertisers and agencies are old fashioned**
  - Don't use mathematically sophisticated systems; human experts
  - Advertisers are broadcast oriented (TV, Newspaper, Radio etc)
  - Huge culture change
    - Real-time versus weekly/monthly
    - They speak the language of audiences and panels (not data mining, machine learning); social scientists
- **Accountability: Internet is still relatively new and largely unknown**
  - Apart from sponsored search (2% of online activity) other media types are largely not accountable and underexploited
  - Poor targeting
  - Guaranteed markets at spot market prices
- **Huge opportunity if history is anything to go by and ....**

# Outline

- **Background: digital advertising, the 0.2% opportunity (of GDP)**
- **3<sup>rd</sup> Generation DA (Personalization)**
  - Explainability: trading desks for the advertiser
  - Behavioral targeting
  - Interactivity: thru social ads
  - Engagement: transformative ad formats
- **Summary**

# From Mad Men To Wall Street and beyond!

- Set in New York City, *Mad Men* begins in 1960 at the fictional Sterling Cooper advertising agency on New York City's Madison Avenue.

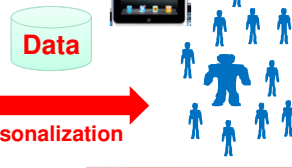
2007



Increasingly



Personalization



Human Intensive  
Lots of guess work  
Forward Market

Technology  
Driven  
Forward Market  
Spot Markets

**Double digit growth**

Advertisers still in broadcast mode



1<sup>st</sup> Generation

2<sup>nd</sup> Generation

3<sup>rd</sup> Generation

# Audience Participation



NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan





## Audience Participation



## NYC's Sexiest Billboard Doesn't Care For Dumbphones



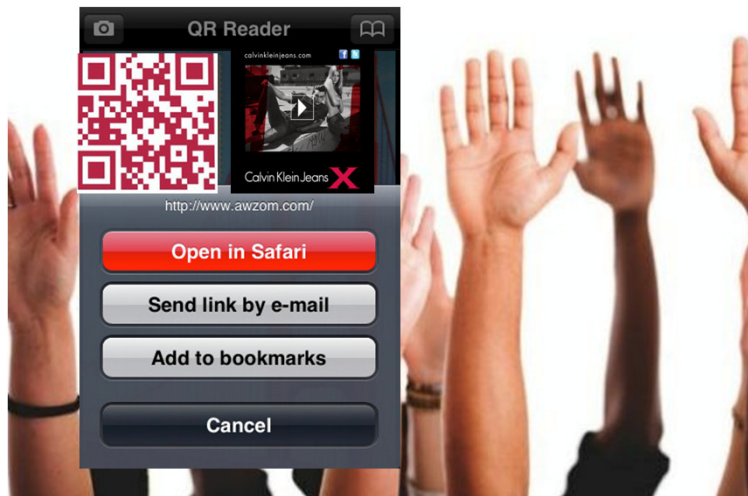
Quick Response Codes

<http://www.gizmodo.com.au/2010/07/nvcs-sexiest-billboard-doesnt-care-for-dumbphones/#more-406771>

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

21

## Audience Participation



NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

22

## NYC's Sexiest Billboard Doesn't Care For Dumbphones



<http://www.gizmodo.com.au/2010/07/nycs-sexiest-billboard-doesnt-care-for-dumbphones/#more-406771>

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

23

## From Mad Men To Wall Street and beyond!

- Set in New York City, *Mad Men* begins in 1960 at the fictional Sterling Cooper advertising agency on New York City's Madison Avenue.

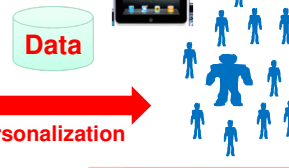
2007



Increasingly



Personalization



Human Intensive  
Lots of guess work  
Forward Market

Technology  
Driven  
Forward Market  
Spot Markets

Double digit growth

Advertisers still in  
broadcast mode



1<sup>st</sup> Generation

2<sup>nd</sup> Generation

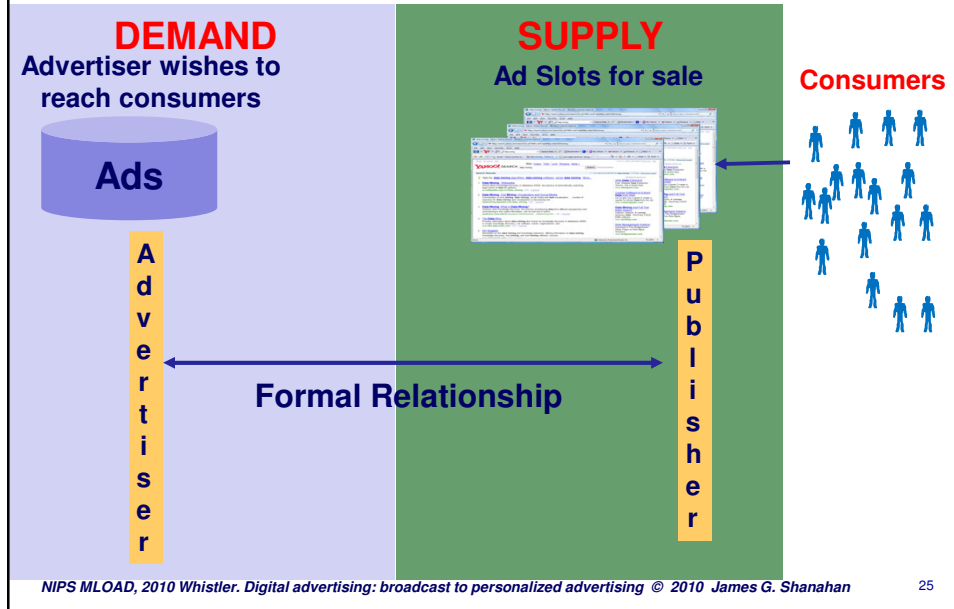
3<sup>rd</sup> Generation

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

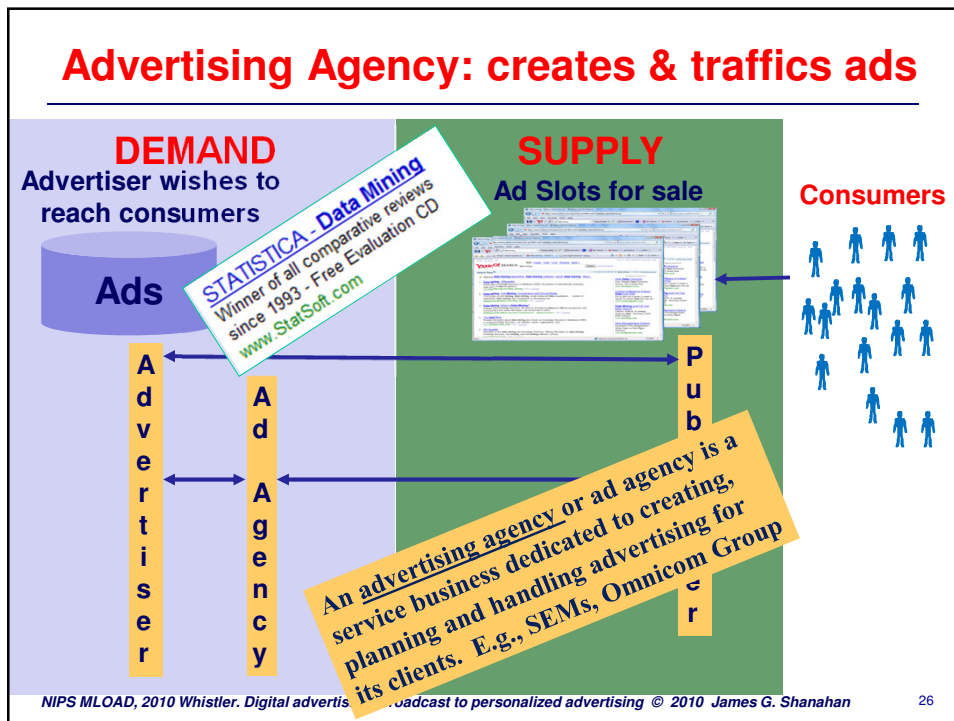
24



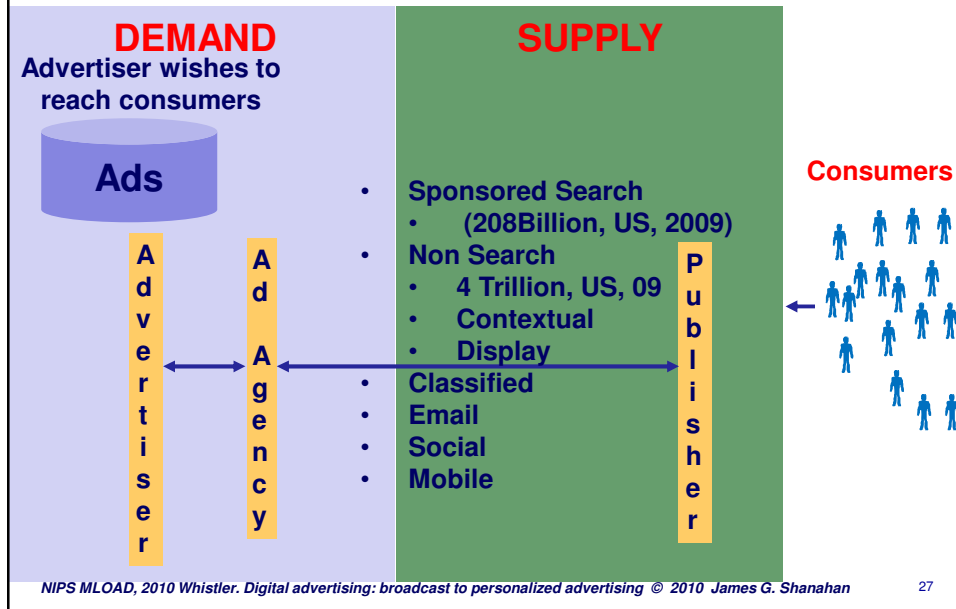
## Advertising: a supply-demand marketplace



## Advertising Agency: creates & traffics ads



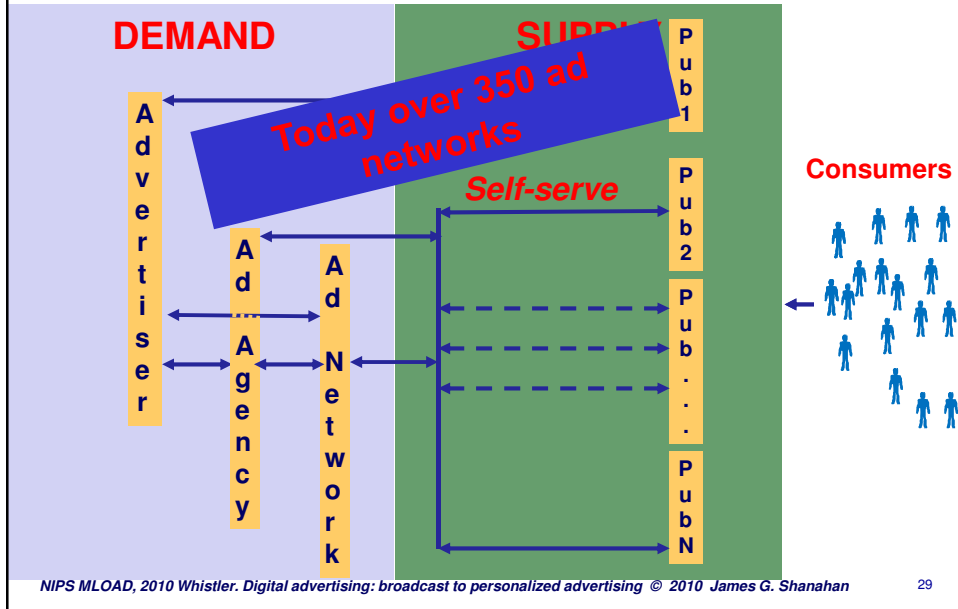
## Advertising Agency: creates & traffics ads



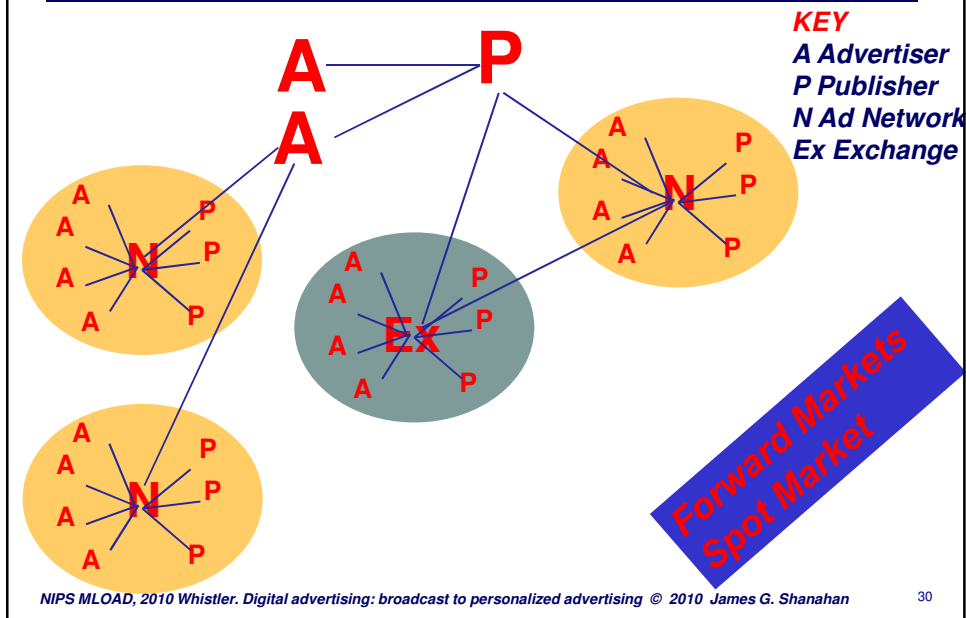
## 2<sup>nd</sup> Generation

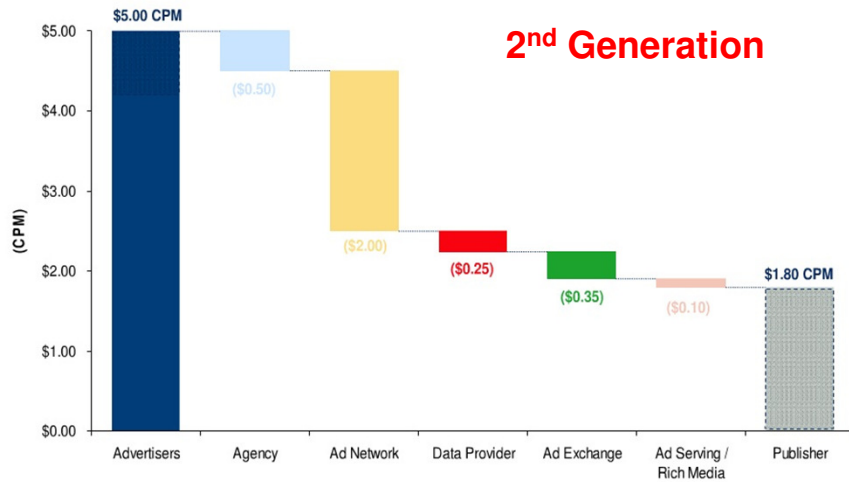
- **CPC, CPA**
- **Quant driven and quant support**
- **Supply can be fragmented → Ad Networks**
  - Outside of search supply can be fragmented
  - Publishers maybe small and not have a sales team

## Advertising Network: Aggregates Publishers



## Online Advertising is a Frenemic Network Play

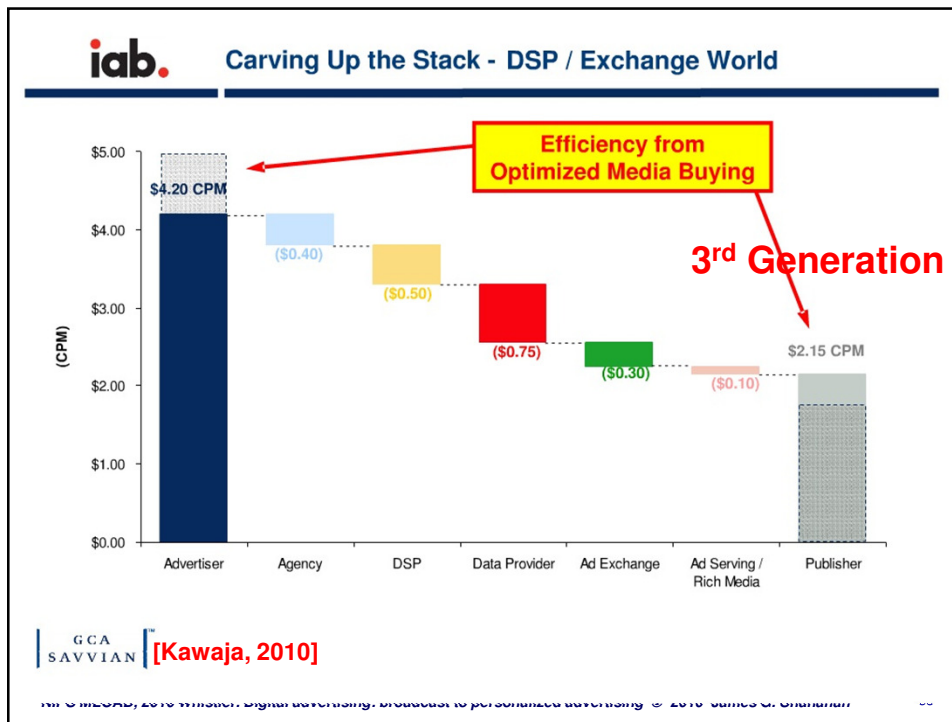




GCA SAVVIAN [Kawaja, 2010]

## 3<sup>rd</sup> Generation

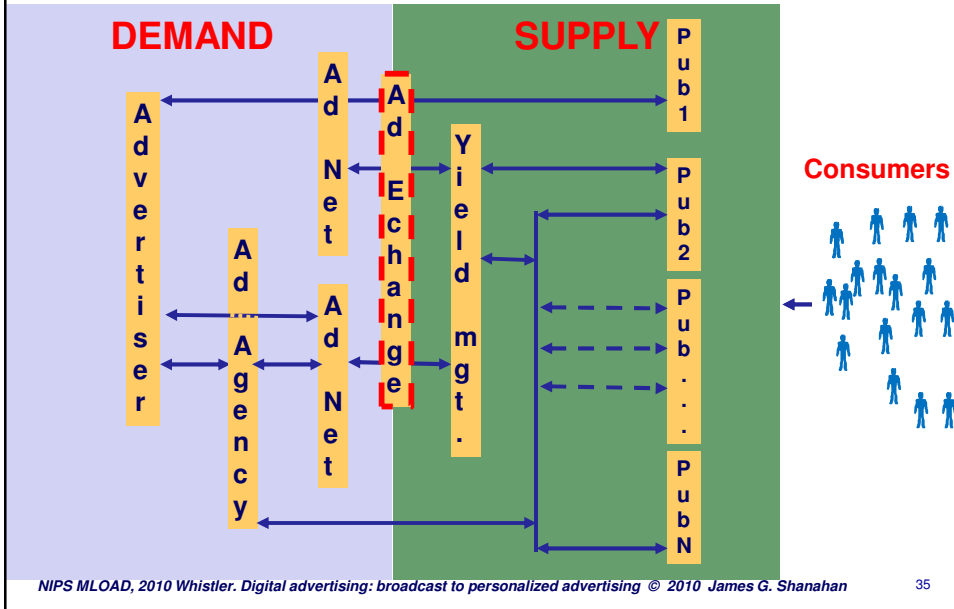
- **New more efficient market places**
  - Ad Exchanges
  - Data exchanges
- **Audience-based targeting**
- **Very complex pipeline**
  - Yield mgt and Demand side platforms



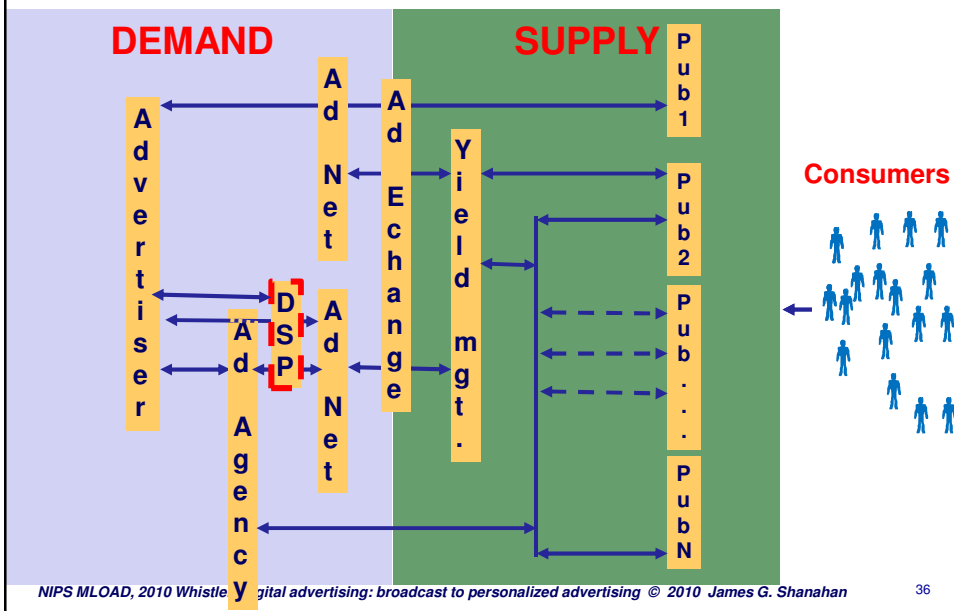
## Ad Exchanges: a new SD Marketplace

- **The ad exchange is a real time marketplace**
  - with an auction-based system where the participants - advertisers and publishers – transact on a common platform to purchase and sell online graphical advertising.
- **Currently, publishers sell remnant inventory**
  - on the exchange for advertisers to purchase through bidding on a user-friendly interface.
- **Ad Exchanges do not compete with ad networks**
  - targeting technologies, or publishers, but rather serve as a more efficient way for the exchange of inventory within these groups
- **Googles acquired DoubleClick, Yahoo acq RightMedia, etc.. \$11 in M&A in 2007**

## Ad Exchange: auctioneer-centric marketplace

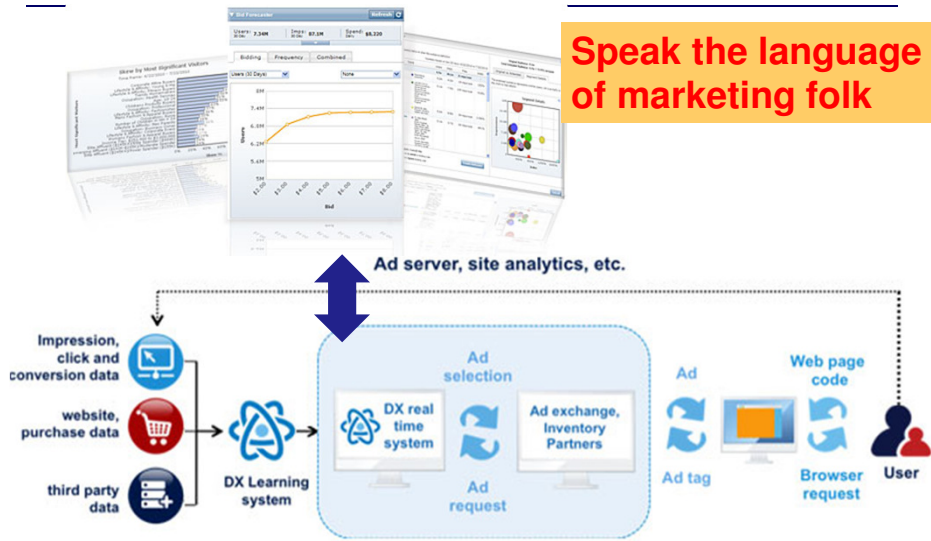


## Demand-Side Platform: A trading desk for Adv.



## Demand side platforms

Speak the language of marketing folk

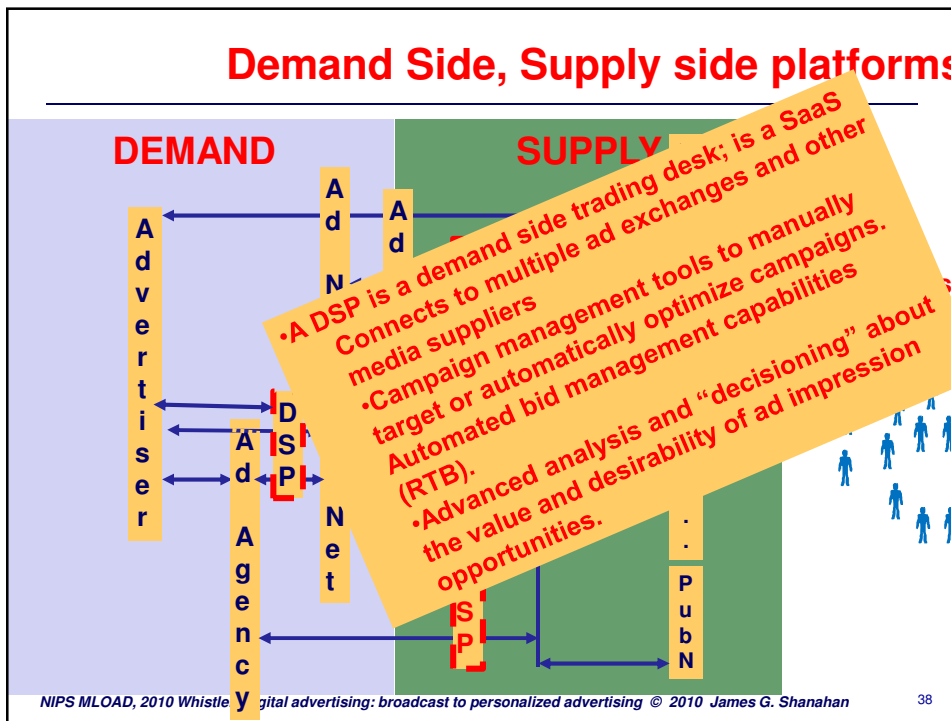


[E.g., Turn, DataXu]

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

37

## Demand Side, Supply side platforms



NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

38

## Key Features of DSP

---

- **Advanced and accurate audience targeting capabilities**
- **Easy-to-use inventory control**
- **Bidding dashboards**
- **Ability to set frequency caps on the ads being served**
  - reaching the "right consumer" too many times can lead to a significant decline in interest

## Open research Areas

---

- **Forecasting**
- **Targeting**
- **Mechanism design**
- **Realtime bidding**
  - [Selective call out and real time bidding, Chakraborty et al., 2010]



## Look-a-like modeling: explainable Knowledge

- **Rule-based targeting**
  - Advertiser/Agency specifies knowledge of target audience (lives in metropolitan area, 25-44, income, reads business sections)
- **Data driven from consumers who bought already**
  - Observe individuals who just transacted; find more who look-like these
  - Positive and negative examples
  - **Active learning**: Instead of assuming that all of the training examples are given at the start, active learning algorithms interactively collect new examples, typically by making queries to a human user. Often, the queries are based on unlabeled data, which is a scenario that combines semi-supervised learning with active learning.  
[Settles, Dum \(2008\), "Active Learning Literature Survey", TR 1548, University of Wisconsin-Madison, http://pages.cs.wisc.edu/~bsettles/pub/settles.activelearning.pdf](http://pages.cs.wisc.edu/~bsettles/pub/settles.activelearning.pdf)
- **Forecasting (see CIKM 2010 papers and posters)**
  - How many of uniques? What-if I increase by bid price?

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

41

## Look-a-like modeling: Challenges

- **Modeling user based on interaction patterns**
- **Data e.g., from QuantCast**
  - Millions of partner sites
  - 10 Billion weblog records (ad tag firing events from publisher); 250 Billion per month
  - 1 Billion users globally
  - 15 terabytes per day of new data
- **Advertisers need to define, refine their models of a typical consumer daily, hourly, secondly**
- **Forecasting (CIKM 2010)**
- **Map-Reduce for offline, real-time modeling**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

42

## Outline

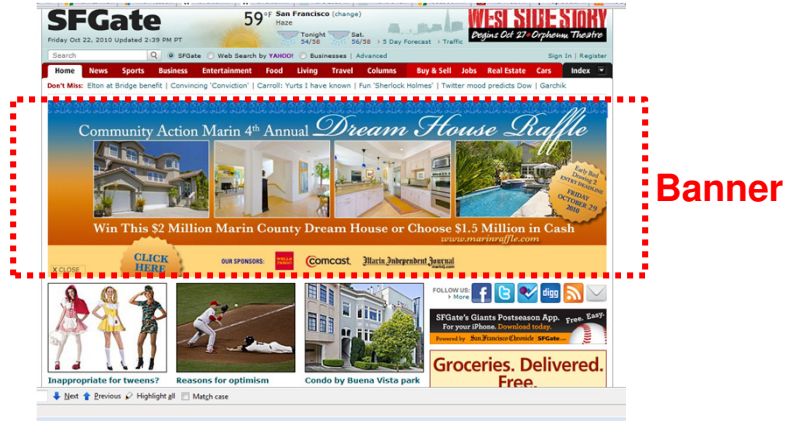
- **Background: digital advertising, the 0.2% opportunity (of GDP)**
- **3<sup>rd</sup> Generation DA (Personalization)**
  - Explainability: trading desks for the advertiser
  - Behavioral targeting
  - Interactivity: thru social ads
  - Engagement: transformative ad formats
- **Summary**

## Advent of WWW: media became personal

- **Advent of the World Wide Web media consumption starts to become more personalized**
  - people could pull the media they wanted to read/watch/listen
  - Media pushed also
- **And the Internet banner ad was a marketing revolution.**
  - We could not only measure the size of the audience that saw our ads but we had a real-time (or near real-time) feedback mechanism in the click-through rate
- **WSJ.com, for example, charges advertisers as much as \$64.60 to show a banner ad to 1,000 viewers. → \$5 CPM**

[A Pricing Revolution Looms in Online Advertising, Business Week, 4/2009]

## The dreaded banner is a big money winner



WSJ.com, for example, charges advertisers as much as \$64.60 to show a banner ad to 1,000 viewers. → \$5 CPM

[A Pricing Revolution Looms in Online Advertising, Business Week, 4/2009]

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

45

## But banner ads are mass marketing

- Measure the size of the audience and CTRs
- However, various studies showed banner blindness
- E.g., comScore/Fox study
  - 84 percent of Web users never click on banner ads.
  - 12 percent not very active
  - 4% percent of Web users that click on banner ads, produce 67 percent of all clicks.

90-9-1 Rule

[http://www.comscore.com/Press Events/Press Releases/2008/02/Display Ad Click-Through Behavior](http://www.comscore.com/Press%20Events/Press%20Releases/2008/02/Display%20Ad%20Click-Through%20Behavior)

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

46

## Behavioral Targeting: Modeling The User

- **Target ads based on user's browsing behavior**
- **Commonly used by:**
  - Website owners (e.g., e-commerce websites)
  - Ad networks
- **Key players include:**
  - E-commerce websites such as Amazon
  - [Blue Lithium](#) (acq by Yahoo!, \$300M), [Tacoda](#) (acq by AOL, \$275M), [Burst](#), [Phorm](#) and [Revenue Science](#), Turn.com, and others...

[ For more background see: [http://en.wikipedia.org/wiki/Behavioral\\_targeting](http://en.wikipedia.org/wiki/Behavioral_targeting) ]

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

47

## Personalization via BT

- **Intuition:**
  - the users who share similar Web browsing behaviors will have similar preference over ads
- **Selling Audiences (and not sites)**
  - Traditionally did this based on panels (user surveys or using Comscore/NetRatings); very broad and not very accurate
  - Through a combination of cookies and log analysis BT enables very specific segmentation
- **Domains of Application**
  - Sponsored search
  - Non-Sponsored search (e.g., contextual, display)

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

48

## Behavioral Targeting: Main Types

---

- **Website/web page visited**
  - E.g., Users who frequent Orbitz.com and Expedia.com or the travel section of USAToday.com would become part of the “Travel Shoppers” segment. Then, these users are re-targeted when they are found on other more general content type sites
- **Keyword profile**
  - using recent searches or content that was read
- **Retarget past visitors to your website (surrogate modeling)**
  - The goal is to use BT to let you send the appropriate message to each user based on where they are in your product sales cycle.

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

49

## BT Technology

---

- **Segmenting users**
  - Clustering, data mining, classification
  - Rule-based system, hybrid systems
  - Segmenting publisher real estate into to categories
- **Collaborative filtering**
  - People who bought this also bought X...
  - Yehuda Koren, Factorization meets the neighborhood: a multifaceted collaborative filtering model. KDD 2008 426-434
  - Remember NetFlix problem??

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

50

# Organic versus Sponsored Clicks

**User Query**

**Organic Clicks**

**Sponsored Clicks**

[How much the Behavioral Targeting can Help Online Advertising? Jun Yan, et al., WWW 2009]

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan 51

# Represent user behavior: terms or click'd URLs

- Represent the users by their behaviors using BT strategies
  - **Queries:** leading to a User by Query Term (query unigram) matrix
    - Use all user queries (in unigram fashion)
    - Stem; remove stopwords and words with a frequency *Less than or Eq 1*

| User × Term       | Q Term <sub>1</sub>  | Term <sub>2</sub> | ..... | Term <sub>i</sub>    |
|-------------------|----------------------|-------------------|-------|----------------------|
| User <sub>1</sub> | TF_IDF <sub>11</sub> |                   | ....  |                      |
| User <sub>2</sub> |                      |                   |       |                      |
| .....             |                      |                   |       |                      |
| User <sub>n</sub> |                      |                   |       | TD_IDF <sub>ni</sub> |

- **Clicked page:** represent user in terms of URLs of pages she clicked

| User × Term       | URL <sub>1</sub> | URL <sub>2</sub> | ..... | URL <sub>i</sub> |
|-------------------|------------------|------------------|-------|------------------|
| User <sub>1</sub> | U <sub>11</sub>  |                  | ....  |                  |
| User <sub>2</sub> |                  |                  |       |                  |
| .....             |                  |                  |       |                  |
| User <sub>n</sub> |                  |                  |       | U <sub>ni</sub>  |

where

$$U_{ij} = (\text{Log}(\# \text{User}_i \text{ click URL}_j) + 1) \times \text{IDF}(\text{URL}_j)$$

## Cluster users based on query unigrams

| User × Term       | Q Term <sub>1</sub>  | Term <sub>2</sub> | ..... | Term <sub>i</sub>    |
|-------------------|----------------------|-------------------|-------|----------------------|
| User <sub>1</sub> | TF_IDF <sub>11</sub> |                   | ....  |                      |
| User <sub>2</sub> |                      |                   |       |                      |
| .....             |                      |                   |       |                      |
| .....             |                      |                   |       |                      |
| .....             |                      |                   |       |                      |
| .....             |                      |                   |       |                      |
| .....             |                      |                   |       |                      |
| User <sub>n</sub> | .....                |                   |       | TD_IDF <sub>ni</sub> |

**Cluster1**

**Cluster20**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan 53

## Add ad clicks to matrix measure CTRs

| User × Term       | Q Term <sub>1</sub>  | Term <sub>2</sub> | ..... |  |  | Ads      |      |                 |
|-------------------|----------------------|-------------------|-------|--|--|----------|------|-----------------|
|                   |                      |                   |       |  |  | Ad1      | .... | Ad <sub>m</sub> |
| User <sub>1</sub> | TF_IDF <sub>11</sub> |                   | ....  |  |  | ClickCnt |      |                 |
| User <sub>2</sub> |                      |                   |       |  |  | ClickCnt |      |                 |
| .....             |                      |                   |       |  |  | No Click |      |                 |
| .....             |                      |                   |       |  |  | No click |      |                 |
| .....             |                      |                   |       |  |  | No click |      |                 |
| .....             |                      |                   |       |  |  | No click |      |                 |
| User <sub>n</sub> | .....                |                   |       |  |  | No click |      |                 |

**C1**

**C20**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan 54

## Add ad clicks to matrix: compare CTRs

|                   |                      |                   |       |  |  | Ads                          |      |                 |
|-------------------|----------------------|-------------------|-------|--|--|------------------------------|------|-----------------|
| User × Term       | Q Term <sub>1</sub>  | Term <sub>2</sub> | ..... |  |  | Ad1                          | .... | Ad <sub>m</sub> |
| User <sub>1</sub> | TF_IDF <sub>11</sub> |                   | ....  |  |  | ClickCnt                     |      |                 |
| User <sub>2</sub> |                      |                   |       |  |  | <b>CTR<sub>AD1,c1</sub></b>  |      | <b>C1</b>       |
|                   | .....                |                   |       |  |  | No Click                     |      |                 |
|                   |                      |                   |       |  |  | No click                     |      |                 |
|                   |                      |                   |       |  |  | NO click                     |      |                 |
|                   |                      |                   |       |  |  |                              |      |                 |
|                   | .....                |                   |       |  |  |                              |      |                 |
| User <sub>n</sub> |                      |                   |       |  |  | <b>CTR<sub>AD1,c20</sub></b> |      | <b>C20</b>      |

**CTR<sub>AD1,c1</sub> >> CTR<sub>AD1,c20</sub>**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan 55

## Cluster organic behaviors using K-means

- **Represent the users by their organic behaviors**
  - Queries terms
  - Clicked URLs
- **Cluster users based on organic behaviour**
  - using k-means or CLUTO
- **Evaluate clusters using clicks on sponsored ads**
  - Is there a tendency to group all clicks into one (or small number of organic clusters)?
  - Evaluate how much BT can help online advertising by delivering ads to good user segments.
- **Metrics**
  - Inter and intra cluster, CTR, F-measure, Ads Click entropy



## Short/long term AND Query/Page

---

- **7 Days**

- LP: using Long term user behavior all through the seven days and representing the user behavior by Page-views;  
**Page model**
- LQ: using Long term user behavior all through the seven days and representing the user behavior by Query terms;  
**Query model**

- **1 Day**

- SP: using Short term user behavior (1 day) and representing user behavior by Page-views;  
**Page model**
- SQ: using Short term user behavior (1 day) and representing user behavior by Query terms.  
**Query model**

## BT Dataset

---

- **Focused on English queries**
- **6,426,633 unique users 17,901 ads**
  - Originally 335,170 unique ads within the seven days.
  - Remove anyone who has more than 100 clicks a day
  - But filter out all the ads that have less than 30 clicks within these seven days,

**Table 1. Format of click-through log used in our study.**

|           |                                                                                                |                                                                                                                                     |
|-----------|------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| UserID    | UID030608473X                                                                                  | A user ID for each unique user.                                                                                                     |
| QueryText | xbox                                                                                           | The detailed query text used by the user                                                                                            |
| QueryTime | 08-06-03 21:15:47                                                                              | The time when the query was issued                                                                                                  |
| ClickTime | 08-06-03 21:16:02                                                                              | The time when the click occurred after the query was issued                                                                         |
| ClickURL  | http://www.xbox365.com                                                                         | The URL which has been clicked by the user                                                                                          |
| IsAd      | 0                                                                                              | A Boolean value to show the clicked URL is an ad or not                                                                             |
| NumberAd  | 3                                                                                              | The number of ads displayed in the search results                                                                                   |
| DisplayAd | http://video-games.half.ebay.com/<br>http://accessories.us.dell.com/<br>http://www.gamefly.com | The URL list of all the ads that displayed by the query. (To save space, we only reserve top domain of the ad URL in this example.) |

## Log File Format

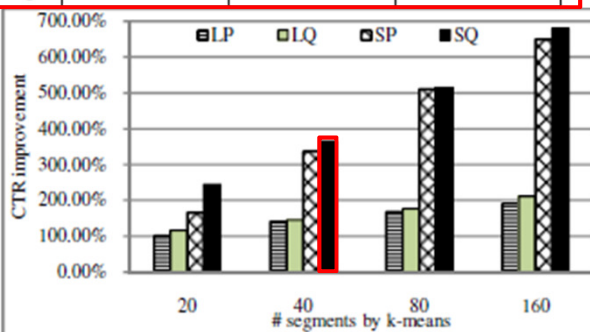
James G. Shanahan

59

## Day-based query model works best

**Table 2. Within- and between- ads user similarity.**

|                                   | $S_w$ | $S_b$  | $R$    |         |
|-----------------------------------|-------|--------|--------|---------|
| 7Days of URLs<br>7Days of queries | LP    | 0.1417 | 0.0252 | 28.9217 |
| 1Days of URLs                     | LQ    | 0.2239 | 0.0196 | 44.2908 |
| 1Days of queries                  | SP    | 0.1532 | 0.0281 | 24.5086 |
| 1Days of queries                  | SQ    | 0.2594 | 0.0161 | 91.1890 |



(a) User clustering by k-means

NIPS MLOAD, 2010 Whistler. Digital advertising: precision performance advertising © 2010 James G. Shanahan

60

## Paper Conclusions

- **Representing short term behaviors (1 day) is better than 7 days**
  - Query words, URLs models are similar; A user's interest tend to diverge over long periods so the shorter the period the better
- **Results suffer from some biases**
  - Post-hoc analysis; therefore targeting was not part of study
  - Results measured using 7 days of historical logs (not an AB test)
  - Queries have a strong correlation to the clicked ads but URLs have less of a correlation
  - (Ads are targeted by keywords so users clustered using clicked ads and search terms would yield similar clusters)
- **Having said that, very strong signal: 670% CTR improvement**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

61

## How to accomplish BT?

- **BT Scenario**
  - Given positive examples (query terms, clicked results, ignored results)
- **Much research**
  - Filtering [TREC]
  - Adaptive Filter [TREC]
  - Interactive tracks
- **Can use adaptive approaches**
  - Rocchio: weight terms (using base + positive exs + negative exs)
  - Active Learning (base machine learning algorithm with explore+exploit capabilities)
    - Bandit problem; Bayesian updating
  - Offline categorization (and then constrain results set); e.g., SVMs
  - WWW, SIGIR, CIKM, TREC, EC etc.
  - 100 milliseconds

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

62

## BT at Google AdSense + AdWords

- **Ad network#**
  - Google segments users along 20 categories and nearly 600 subcategories.
  - Assign users based on users browser behavior
    - It does not plan to associate the cookie of users with search data or with information from other Google services, like Gmail.
- **Query history to target ads on Google SERPs\$**
  - When determining which ads to show on a Google search result page, the AdWords system evaluates some of the user's previous queries during their search session as well as the current search query.
  - If the system detects a relationship, it will show ads related to these other queries, too.

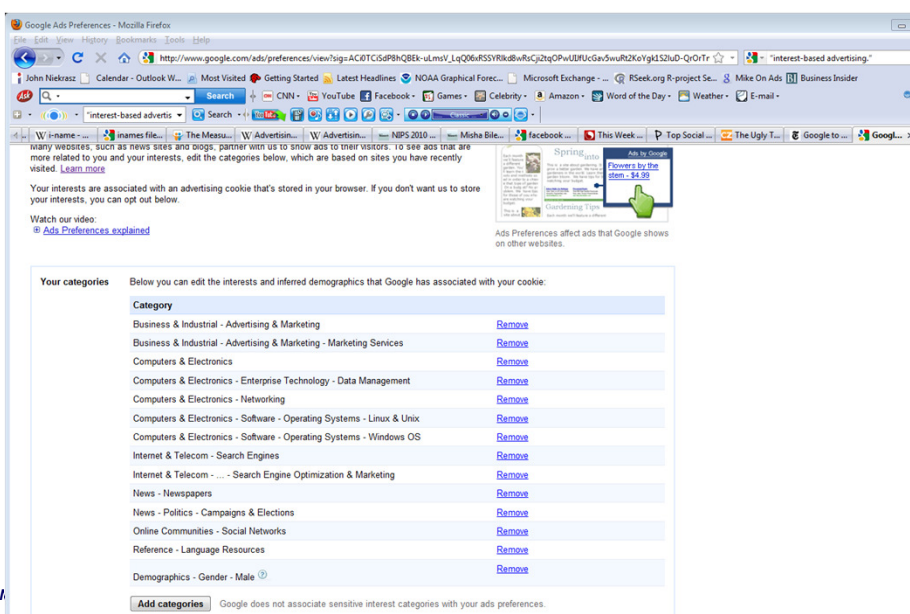
#<http://www.nytimes.com/2009/03/11/technology/interest/11google.html> March 11, 2009 NyTimes.com

<http://adwords.google.com/support/aw/bin/answer.py?hl=en&answer=74246>

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

63

## Google: editing my BT interests



The screenshot shows the Google Ads Preferences page in a Mozilla Firefox browser. The page title is "Google Ads Preferences - Mozilla Firefox". The address bar shows the URL: [http://www.google.com/ads/preferences/view?sig=AC6TCSdPBHQBEL-uLmV\\_LqQ06rFSSYRkdbwRzCj2tqOPwLEUcGa5wvR2KoYgk1S2uD-QrTr](http://www.google.com/ads/preferences/view?sig=AC6TCSdPBHQBEL-uLmV_LqQ06rFSSYRkdbwRzCj2tqOPwLEUcGa5wvR2KoYgk1S2uD-QrTr). The page content includes a section titled "Your categories" with the following text: "Below you can edit the interests and inferred demographics that Google has associated with your cookie:". Below this text is a list of categories, each with a "Remove" link:

| Category                                                              | Remove                 |
|-----------------------------------------------------------------------|------------------------|
| Business & Industrial - Advertising & Marketing                       | <a href="#">Remove</a> |
| Business & Industrial - Advertising & Marketing - Marketing Services  | <a href="#">Remove</a> |
| Computers & Electronics                                               | <a href="#">Remove</a> |
| Computers & Electronics - Enterprise Technology - Data Management     | <a href="#">Remove</a> |
| Computers & Electronics - Networking                                  | <a href="#">Remove</a> |
| Computers & Electronics - Software - Operating Systems - Linux & Unix | <a href="#">Remove</a> |
| Computers & Electronics - Software - Operating Systems - Windows OS   | <a href="#">Remove</a> |
| Internet & Telecom - Search Engines                                   | <a href="#">Remove</a> |
| Internet & Telecom - Search Engine Optimization & Marketing           | <a href="#">Remove</a> |
| News - Newspapers                                                     | <a href="#">Remove</a> |
| News - Politics - Campaigns & Elections                               | <a href="#">Remove</a> |
| Online Communities - Social Networks                                  | <a href="#">Remove</a> |
| Reference - Language Resources                                        | <a href="#">Remove</a> |
| Demographics - Gender - Male                                          | <a href="#">Remove</a> |

At the bottom of the list, there is an "Add categories" button and a note: "Google does not associate sensitive interest categories with your ads preferences."

## Interest Categories

The screenshot displays the Google Ads Preferences interface. A modal window titled "Add categories" is open, allowing users to manage their interest categories. The dialog contains a list of categories with corresponding "Add" or "Remove" buttons:

| Category                | Action |
|-------------------------|--------|
| Arts & Entertainment    | Add    |
| Autos & Vehicles        | Add    |
| Beauty & Fitness        | Add    |
| Books & Literature      | Add    |
| Business & Industrial   | Add    |
| Computers & Electronics | Remove |
| Computer Security       | Add    |
| Antivirus & Malware     | Add    |

The background shows the "Your categories" section of the Google Ads Preferences page, which lists various categories and their associated actions (e.g., "Business & Industrial - Advertising & Marketing" with a "Remove" button).

## Personalized Sponsored Search via BT

- **Google uses query history to target ads on Google SERPs**
  - When determining which ads to show on a Google search result page, the AdWords system evaluates some of the user's previous queries during their search session as well as the current search query.
  - If the system detects a relationship, it will show ads related to these other queries, too.
- **This feature is an enhancement of broad match.**
  - It works by generating similar terms for each search query based on the content of the current query and, if deemed relevant, the previous queries in a user's search session.
  - Your ad will potentially show if one of your broad-matched keywords matches any of these similar terms.
  - **Win-win**; quality score not affected; Google accepts the cost

<http://adwords.google.com/support/aw/bin/answer.py?hl=en&answer=74246>

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

66

## BT had a bad name

- **Gaining respect**
  - Google began showing ads on 3/11/2009 to people based on their previous online activities in a form of advertising known as behavioral targeting
  - Interest-based advertising
- **Privacy Protection**
  - Edit BT profiles
    - give users the ability to see and edit the information that it has compiled about their interests for the purposes of behavioral targeting
  - Opt-out easily
    - Google and many others is a participating member of the [Network Advertising Initiative](#).
  - Remove cookies

<http://www.nytimes.com/2009/03/11/technology/internet/11google.html> March 11, 2009 NyTimes.com

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

67

## The Network Advertising Initiative (NAI)

A consortium of approximately 30 companies that use BT technology. Opt-out easily!!

The screenshot shows the NAI website's 'Opt Out of Behavioral Advertising' page. The page includes a navigation menu, a sidebar with links like 'Overview', 'Principles Overview', 'Get-Out', 'Enforcement', 'Opt-out Problems', and 'FAQs'. The main content area explains the purpose of the tool and provides instructions on how to use it. Below the text is an 'Opt-Out Status' section with a table listing member companies and their current status.

| Member Company                                                      | Status                                                                                    | Opt-Out                          |
|---------------------------------------------------------------------|-------------------------------------------------------------------------------------------|----------------------------------|
| <a href="#">aCerno</a><br><a href="#">More Information</a>          | No Cookie<br>You have not opted out, and you have no cookie from this network.            | Opt-Out <input type="checkbox"/> |
| <a href="#">Advertising.com</a><br><a href="#">More Information</a> | Active Cookie<br>You have not opted out, and you have an active cookie from this network. | Opt-Out <input type="checkbox"/> |
| <a href="#">Akamai</a><br><a href="#">More Information</a>          | Active Cookie<br>You have not opted out, and you have an active cookie from this network. | Opt-Out <input type="checkbox"/> |

On the right side of the page, there is a list of participating networks with 'More Information' links for each: Google, interCLICK, Media6degrees, Mindset Media, NextAction, Safecount, SpecificMEDIA, Traffic Marketplace, Turn, 24/7 Real Media, Undertone Networks, [x+1] (formerly Poinde), and Yahoo! Ad Network.

NIPS MLOAD

68

## BT Works

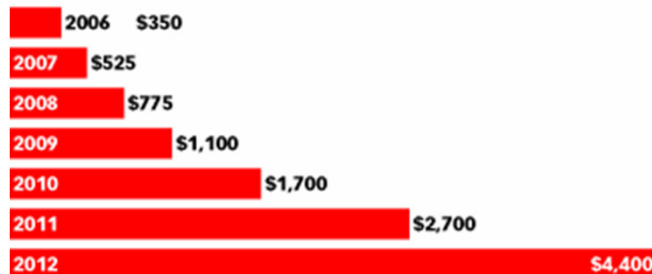
- **Behavioral targeting tries to model user needs and intents based on past Web-surfing behavior**
  - Sponsored search, display, and contextual advertising
- **Advertiser reduce CPMs**
  - Reduce ad costs from \$60 to \$3 CPM, a 95% savings
- **Publisher can generate more and higher RPMs**
  - Find unsold segments of readers and sell them to advertisers that previously couldn't access them.
    - Use data on past reader "click streams" from other Web sites to help advertisers reach the consumers who are most interested in their products.

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

69

## Behavioral Targeting is growing

### US Behaviorally Targeted Online Advertising Spending, 2006-2012 (millions)



Note: advertising displayed to a select audience whose interests or intentions are revealed by Web site or ISP tracking data, audience segmentation and/or predictive analysis; excludes ads targeted using adware

Source: eMarketer, June 2008

094973

www.eMarketer.com

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

70

## Behavioural Targeting Summary

- **BT works for both sponsored search and non-sponsored search**
  - CTRs >> E.g., 11% to 840% higher than average
- **Holds huge promise to monetize longtail of mostly remnant inventory (such as Web2.0)**
  - \$8 Billion US Display ad market
  - Network effect will increase the value of this inventory (reach more people)
- **Privacy concerns need to be addressed before consumers, advertisers and publishers embrace this direction in a big way**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

71

## Personalized Ads Example



- “The RHS ad was customized and chosen from thousands of different creative elements, automatically and in real-time, by machine-learning algorithms developed by [Teracent](#), a San Mateo, California startup”

<http://googleblog.blogspot.com/2009/11/displaying-best-display-ad-with.html>

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

72



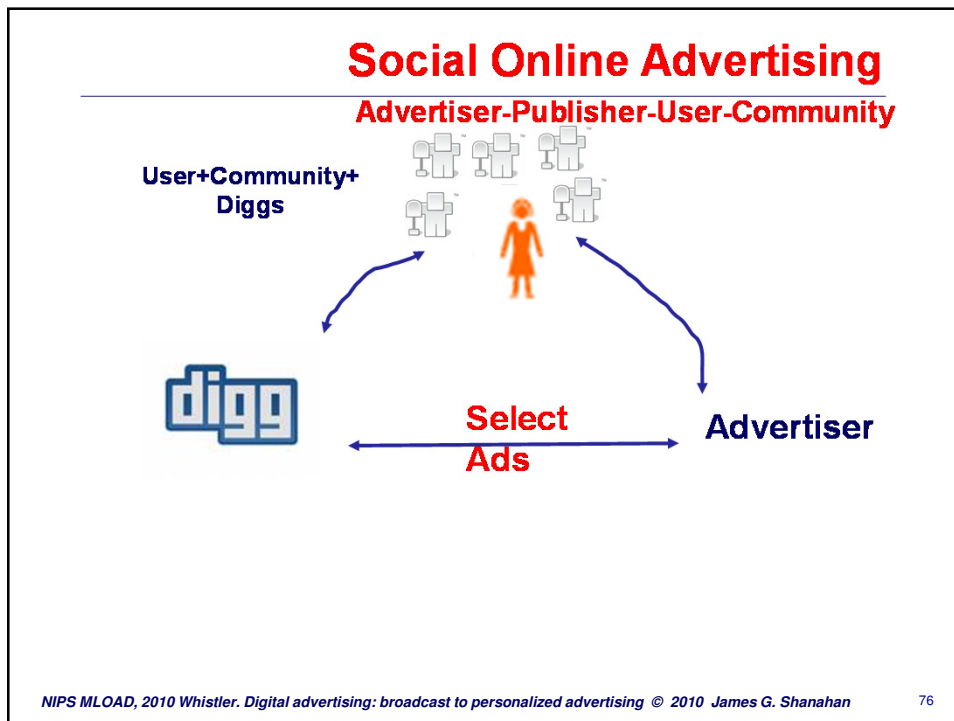
## Outline

- **Background: digital advertising, the 0.2% opportunity (of GDP)**
- **3<sup>rd</sup> Generation DA (Personalization)**
  - Explainability: trading desks for the advertiser
  - Behavioral targeting
  - Interactivity: thru social ads
  - Engagement: transformative ad formats
- **Summary**

## Social Advertising

- **Introduced in the context of community-oriented websites**
  - such as Facebook and Digg. And more recently Twitter
- **Leverages historically "offline" dynamics**
  - such as peer-pressure, recommendations, and other forms of social influence
- **In the case of Digg, Digg or Bury the ad**
  - users determine what ads appear on the website news streams by voting up or "burying" ads,
  - much the way they can digg or "bury" organic news items
  - recently they have rolled this out for graphic ads also

**This enables advertisers and consumers to operate in a more conversant manner.**



# Interactive Social Graphic Ads

The screenshot shows the Digg website interface. At the top, there are navigation tabs for 'My News', 'Top News', and 'Upcoming'. A search bar and 'Join Digg!' button are also visible. The main content area is titled 'Top News' and features several news items with 'digg' counts and brief descriptions. A large, eye-catching advertisement for 'Jackass 3D' is highlighted with a red dashed border. The ad features the 'Jackass 3D' logo, a skull and crossed axes, and the text 'now playing get tickets & showtimes experience it in real 3D and digital 3D'. Below the main news feed, there is a 'Top News in All Topics' section with a list of trending items.

**Jackass 3D**  
now playing  
get tickets & showtimes  
experience it in real 3D and digital 3D

Jackass 3D Might Be The Reason 3D Was Invented.

Top News in All Topics

- 438 Seven inches is enough, RIM tells Jobs
- 383 9 Actresses Who Get By On Their Looks
- 310 Which Party You Will Vote for in the Next Election (Flowchart)
- 280 The Most Amazing Kill Ever: LoopZook [VID]
- 278 Sarah Palin & Fox News Declare War On NPR
- 252 10 geek sins that will get your geek card revoked

NIPs MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan 77

# Social Ads Facebook

The screenshot shows a Facebook news feed for a user named James G. Shanahan. The feed includes a notification to update the email address, a search bar, and several news items. A large advertisement for Janet Reilly is highlighted with a red border. The ad features a photo of Janet Reilly and the text 'NO on Janet Reilly for District 2'. Below the photo, it says 'Janet Reilly supported Prop. H that would have cost SF \$4 billion. Learn about her gamble and how it would have put SF services at risk.' There is a 'Like' button at the bottom of the ad.

facebook

Please update your email address

Our systems have detected that james.shanahan@turn.com is no longer a valid email. Facebook requires all users to maintain an active contact email. Please enter and confirm a new contact email below:

New Email:  Send Confirmation

If you believe you have received this message in error, please reconfirm your current email.

News Feed

What's on your mind?

Antonietta Grasso via Mario Ferrandi: w le ultrasettantenni  
«Lo spinnello? Fa dormire meglio» - Corriere della Sera  
www.corriere.it  
Ornella Vanoni alla radio: una confessione che farà discutere

2 hours ago · Comment · Like · Share

Michael Witbrock started a trip to Ljubljana  
See the trip on Dopplr  
1453m by plane.

16 hours ago via Dopplr: Where next? · Comment · Like · See the trip on Dopplr

Mario Grobernik likes this.

Marco Zagħa Ljubljana on Dopplr? Can I buy a vowel?  
7 hours ago · Like

Write a comment...

Kevin Collins-Thompson is way behind on pretty much everything...

NO on Janet Reilly for District 2

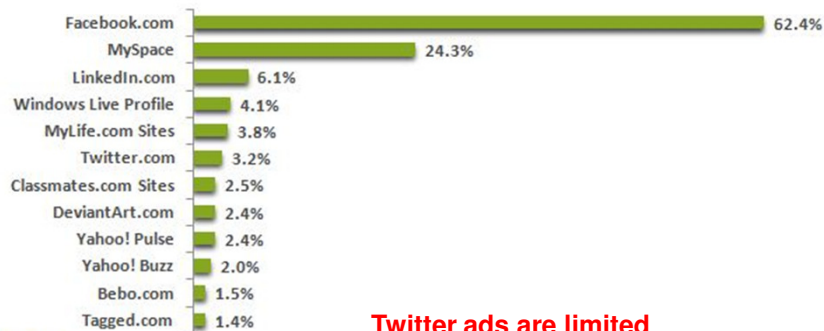
Janet Reilly supported Prop. H that would have cost SF \$4 billion. Learn about her gamble and how it would have put SF services at risk

Like

## Ad Reach within top Social Networks

Top Social Network Publishers by Advertising Exposed Reach (%) of U.S. Online Population

Source: comScore Ad Metrix, September 2010



Twitter ads are limited to search streams



NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

79

## Social Advertising Summary

- **Social advertising has the potential to be more conversant with the consumer**
  - thereby increase the advertisers trust in display-advertising in particular
- **Still very new but with huge reach**
- **Social ads successfully deployed**
  - much more popular than pure display ads
- **500 million users on Facebook, \$1 Billion in revenue**
  - \$2 per user revenue

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

80

## Outline

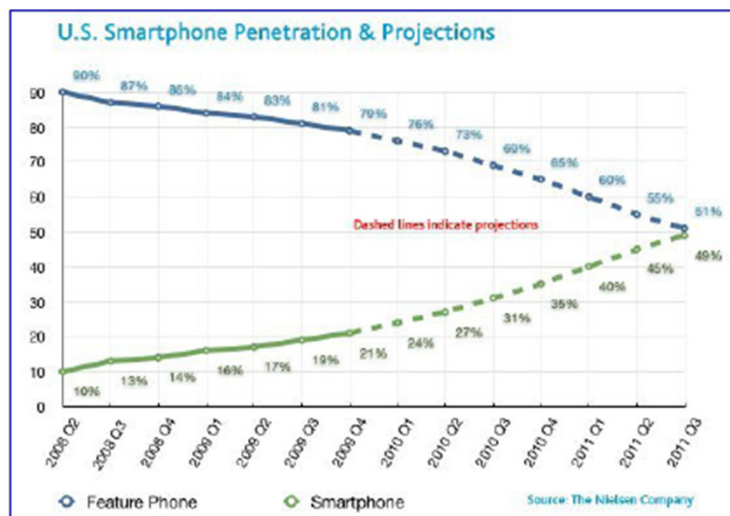
- **Background: digital advertising**
- **Personalization thru:**
  - Explainability: trading desks for the advertiser
  - Behavioral targeting
  - Interactivity: thru social ads
  - Engagement: the untethered consumer
    - Smarter and more portable devices (3G, 4G, broadband)
    - Social networking is becoming more pervasive
    - Location based services (LBS), mobile video and apps, augmenting websites with social capabilities
    - Transformative advertising
    - Augmented reality and quick response (QR) codes
- **Summary**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

81

## Smartphones 50% share in mid2011

USA only



NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

82

## Mobile Ad

- Click/Tap to land in AP Store

Mobile Ad



[\[http://www.admob.com/marketing/pdf/CapitalRadio\\_AdMobCaseStudy.pdf\]](http://www.admob.com/marketing/pdf/CapitalRadio_AdMobCaseStudy.pdf)

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan 83

- Mobile Ad
- Click to land in AP Store

Mobile Ad



NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan 84



# Pandora

The diagram on the left illustrates the AdMob ecosystem. It shows a cycle where 'Your App' (1) shows ads for other apps, which then show their own ads (2), leading to users downloading the app from the App Store (3). The Pandora app interface on the right shows a video player for 'Brindley Brothers Hurricane Filled With Fire' with a 'BECK'S' advertisement overlay. The interface includes playback controls and a progress bar.

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

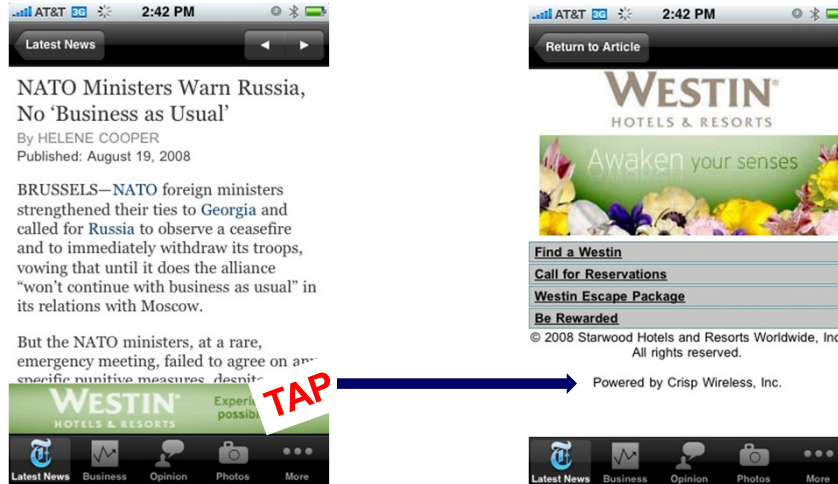
# App: UrbanSpoon

The image shows two screenshots of the UrbanSpoon app. The left screenshot displays a search result for 'Basilico Gourmet Pizza & Restaurant' in Spring Valley, NY, with a 'TAP' callout pointing to the 'Add Photo' button. The right screenshot shows the restaurant's detail page with buttons for 'Place order', 'Find Store', and 'Get Deals'. A red arrow points from the 'Add Photo' button in the first screenshot to the 'Place order' button in the second.

<http://searchengineland.com/google-adsense-ads-making-their-way-to-iphone-apps-16969>

NIPS MLOAD, 2010 Whistler. Dig

## Mobile Web



NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

87

## Tablet PCs

- **Large Format Benefit**
  - Enhanced mobile apps
  - Increased content display opportunities
  - Video optimization
  - Interactive graphics
- **Growth**
  - Total media tablets device market
    - 15MM units in 2010 growing
    - 28MM in 2011 (ABI, 2010; Barclays Capital, 2010)
    - 46MM units in 2014 (IDC, May 10, 2010)
  - Apple expected to sell ~20MM iPads in 2011 (Barclays Capital, July 2010)



NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

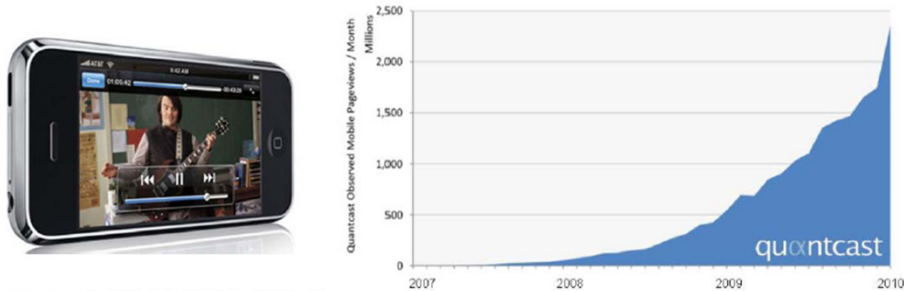
88



## Mobile Use & Video

### Smartphones driving video explosion

- 51% mobile video growth since 2009 (Nielsen, 2010)
- # of mobile video users will rise by 34% annually to 95MM in 2015 (Coda Research Consultancy, 2010)
- By 2015, US mobile consumers are expected to use 327,000 terabytes of mobile data monthly, rising at a CAGR of more than 117% (Coda Research Consultancy, 2010)
- Mobile video revenue set to exceed \$2B in 2013, thanks to 4G mobile broadband (ABI Research, 2010)



NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

89

## Tipping Points

- **The true turning point for the industry was the introduction of the smartphone, heightened by Apple's iPhone launch in mid-2007.**
  - Tearing down the carrier wall
- **Better connectivity**
  - 3G, 4G mobile phones led to better connection speeds
  - Wi-Fi connectivity and the rise of mobile Internet browsing.
- **Richer Content**
  - Apps, For mobile web pages, Video, TV, Games
- **In developing markets, mobile is the primary gateway to the Internet not PC!**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

90

## Mobile Monetization

- All the internet monetization models are applicable to mobile.
- New monetization models that leverage unique mobile features e.g. **pay per call**.

|            | Search   | Contextual | Display  |
|------------|----------|------------|----------|
| CPM        |          |            | X        |
| CPC        | X        | X          | X        |
| CPA        | X        | X          | X        |
| <b>PPC</b> | <b>X</b> | <b>X</b>   | <b>X</b> |

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

91

## Select Best Ads to Display

- **Requires:**
  - Ads to select from (i.e., ads with budget remaining)
  - Ability to determine best ad to show (see below);
  - Serve ad or not (social responsibility)
- **Selecting Best Ad(s)**
  - Estimate the CTR of each ad using available features.
  - A number of techniques for estimating CTR:
    - Regularized Regression models, Multi-armed bandit models
    - Non parametric: collaborative filtering, nearest neighbor methods.
  - Features for estimating CTR:
    - Location based features,
    - Social network features
    - Site characteristics
    - Ad Characteristics
    - User Characteristics, user behavior

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

92

## Mobile Ad Spend will triple by 2012

### Mobile ad spend:

- Projected \$800MM+ in 2011, \$1.1B in 2012
- Marketers investing in mobile ads

(\$ millions)

|                           | 2008            | 2009            | 2010            | 2011            | 2012            |
|---------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Paid search               | \$7,910         | \$9,492         | \$11,044        | \$13,364        | \$15,368        |
| Display                   | \$4,089         | \$4,212         | \$4,507         | \$4,957         | \$5,453         |
| Classifieds               | \$3,127         | \$3,221         | \$3,285         | \$3,581         | \$3,903         |
| Internet video/rich media | \$1,854         | \$2,206         | \$2,691         | \$3,229         | \$3,940         |
| Social media              | \$386           | \$558           | \$728           | \$948           | \$1,232         |
| Mobile                    | \$232           | \$391           | \$561           | \$802           | \$1,147         |
| Internet radio            | \$200           | \$230           | \$258           | \$286           | \$315           |
| Podcast                   | \$25            | \$28            | \$31            | \$36            | \$40            |
| <b>Total</b>              | <b>\$17,823</b> | <b>\$20,338</b> | <b>\$23,106</b> | <b>\$27,204</b> | <b>\$31,399</b> |

Note: at current prices; numbers may not add up to total due to rounding  
Source: ZenithOptimedia, provided to eMarketer, Jul 19, 2010

117815

www.eMarketer.com

### Mobile ad spend:

- Projected 43% growth in 2011 and 2012
- Largest digital media ad spend growth rate

(% change)

|                           | 2008         | 2009         | 2010         | 2011         | 2012         |
|---------------------------|--------------|--------------|--------------|--------------|--------------|
| Mobile                    | 93.3%        | 68.5%        | 43.5%        | 43.0%        | 43.0%        |
| Social media              | 31.6%        | 44.6%        | 30.6%        | 30.2%        | 30.0%        |
| Internet video/rich media | 40.5%        | 19.0%        | 22.0%        | 20.0%        | 22.0%        |
| Paid search               | 22.0%        | 20.0%        | 16.4%        | 21.0%        | 15.0%        |
| Internet radio            | -            | 15.0%        | 12.0%        | 11.0%        | 10.0%        |
| Podcast                   | 15.0%        | 11.9%        | 10.6%        | 15.0%        | 11.0%        |
| Display                   | 15.0%        | 3.0%         | 7.0%         | 10.0%        | 10.0%        |
| Classifieds               | 15.0%        | 3.0%         | 2.0%         | 9.0%         | 9.0%         |
| <b>Total</b>              | <b>22.8%</b> | <b>14.1%</b> | <b>13.6%</b> | <b>17.7%</b> | <b>15.4%</b> |

Note: based on figures at current prices  
Source: ZenithOptimedia, provided to eMarketer, Jul 19, 2010

117810

www.eMarketer.com

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

93

## Outline

- **Background: digital advertising**
- **Personalization thru:**
  - Behavioral targeting
  - Interactivity: thru social ads
  - Engagement: transformative ad formats
    - Smarter and more portable devices (3G, 4G, broadband)
    - Social networking is becoming more pervasive
    - Location based services (LBS), mobile video and apps, augmenting websites with social capabilities
    - Transformative advertising
    - Augmented reality and quick response (QR) codes
  - Explainability: trading desks for the advertiser
- **Summary**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

94



## Social Network Ad Spend Cools Off

### US Online Social Network Advertising Spending, 2008-2013 (millions and % change)

|      |                 |
|------|-----------------|
| 2008 | \$1,175 (33.8%) |
| 2009 | \$1,295 (10.2%) |
| 2010 | \$1,335 (3.1%)  |
| 2011 | \$1,420 (6.3%)  |
| 2012 | \$1,515 (6.7%)  |
| 2013 | \$1,640 (8.3%)  |

Source: eMarketer, December 2008

100080

www.eMarketer.com

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

97

## Outline

- **Background: digital advertising**
- **Personalization thru:**
  - Behavioral targeting
  - Interactivity: thru social ads
  - Engagement: transformative ad formats
    - Smarter and more portable devices (3G, 4G, broadband)
    - Social networking is becoming more pervasive
    - Location based services (LBS), mobile video and apps, augmenting websites with social capabilities
    - Transformative advertising
    - Augmented reality and quick response (QR) codes
  - Explainability: trading desks for the advertiser
- **Summary**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

98

# Top Tier Location Based Social Networks

|                                  | foursquare                                             | Gowalla                                                | brightkite                                             | loopt                                            | yelp                                                        | where                                                  | Rooyah!                                                    | Facebook Places                                                        |
|----------------------------------|--------------------------------------------------------|--------------------------------------------------------|--------------------------------------------------------|--------------------------------------------------|-------------------------------------------------------------|--------------------------------------------------------|------------------------------------------------------------|------------------------------------------------------------------------|
| Users                            | 2.6 MILLION                                            | 390 THOUSAND                                           | 2.2 MILLION                                            | 4 MILLION                                        | 2 MILLION                                                   | 3 MILLION                                              | 2.5 MILLION                                                | Just Launched                                                          |
| Game Mechanics                   | Points & Badges                                        | Pin & Home                                             | Badge levels                                           | Achievements                                     | Badges and Loyalty Levels                                   | None Currently                                         | Monopoly like with Points, locations, Buildings & Products | None Currently                                                         |
| API?                             | YES<br>Limited to check in data                        | YES<br>Limited                                         | YES<br>Limited                                         | NO                                               | YES<br>Limited to Review and Ratings                        | YES<br>In beta - limited                               | NO                                                         | YES<br>Read & Write                                                    |
| Business or Retailer Dashboards? | YES                                                    | NO                                                     | NO                                                     | NO                                               | YES                                                         | NO                                                     | NO                                                         | YES                                                                    |
| Platforms                        | iPhone, Android, BlackBerry, Palm, Symbian, Java, iPad | iPhone, Android, BlackBerry, Palm, Symbian, Java, iPad | iPhone, Android, BlackBerry, Palm, Symbian, Java, iPad | AppStore                                         | iPhone, Android, BlackBerry, Palm, Symbian, Java, iPad      | iPhone, Android, BlackBerry, Palm, Symbian, Java, iPad | AppStore, BlackBerry, Palm, Symbian, Java, iPad            | iPhone, Android, BlackBerry, Palm, Symbian, Java, iPad                 |
| Primary User Benefits and Perks  | Coupons & Freebies to Mayors                           | Free Gifts, coupons & specials                         | Reward based programs                                  | Instant trend meet-up, retail discounts          | Coupons, comprehensive restaurant reviews and rating guide. | Comprehensive business reviews, mobile coupons         | Great real world game, some product discounts              | Think tag your friends to check it, see what businesses are hot or not |
| CEO                              | Denis Crowley                                          | Jack Williams                                          | Jonathan Lerner                                        | Sam Altman                                       | Jeremy Stoppelman                                           | Walt Doyle                                             | Gaith Lee                                                  | Mark Zuckerberg                                                        |
| In Ten words or less             | Get rewarded for exploring your city                   | Discover new places and earn rewards                   | Location based social recommendation service           | Find friends, go places, rate them, get rewarded | Share local business reviews and tips with friends          | Mobile Citysearch and recommendation service           | Where Monopoly & SimCity & real life meet via mobile       | Check in to local businesses and share moments                         |

Source: Mark Fidelman @ markfidelman.com SeekOmega.com Technorati CloudAve.com

## Layering on Social: Google Connect

- “Google Friend Connect instantly awakens and strengthens the community that visits your site by enriching it with social features”.

The screenshot shows a recipe page for 'Guac of the day' with a social overlay. The recipe includes ingredients like tomatoes, onions, and avocados, and directions for preparation. The social overlay features a 'Google Friend Connect BETA' logo, a list of friends who are members, a 'Members of guac' list, and a review section with a 5-star rating and a post from user 'mussie' stating 'This is the best guac EVAR'.

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

## Add social gadgets to your website

- Choose from a range of social gadgets that let users post comments, post links, rate, review, and more.
- The gadget gallery includes gadgets created by Google and OpenSocial developers.

The screenshot shows the 'Guitar Universe' website with several social gadgets. Red boxes highlight the following features:

- Join:** A 'Join this site' button with a 'Join' label next to it.
- Community members:** A 'Members (46)' section showing a grid of member avatars.
- Community Poll:** A poll titled 'What is my guitar skill level?' with options for 'Advanced', 'Medium', and 'Beginner'.
- Filtered content:** A 'Featured Content' section with a red box around the text 'Guitar.com - A place for guitarists to call'.

At the bottom right of the screenshot, the text 'James G. Shanahan' and the number '101' are visible.

## Simply add HTML code to webpage

The screenshot shows a Mozilla Firefox browser window displaying the 'Google Friend Connect Overview' page. The page is titled 'Create the HTML code' and provides instructions for generating HTML code for a gadget. A 'Generate code' button is visible, and a code editor shows the resulting HTML code. The code includes various attributes for content, links, and colors. At the bottom of the browser window, the text 'Transferring data from v16.lscache6.youtube.com...' and 'zotero' are visible.

## Hot Areas: Model network behavior

- **Need new tools to model information (product) diffusion in social networks (email, Facebook, MySpace, Digg etc.)**
  - The very small world of the well-connected. Shi, Bonner, Adamic, Gilbert, *ACM Hypertext*, 2008.
- **Influence models**
  - Investigate the direct peer influence on information diffusion among social networks.
  - Explore different features, both associated with the users or the information, correlate with information diffusion.
  - [User grouping behavior in online forums. Shi, Zhu, Cai, Zhang, *KDD'09*.]
  - [Information diffusion in computer science citation networks, Shi, Tseng, Adamic, *ICWSM*, 2009]
- **Other Diffusion models**
  - [Inferring networks of diffusion and influence](#) Jure Leskovec, M. Gomez-Rodriguez and A. Krause. *KDD 2010*.

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

103

## Vertex-Importance Graph Synopsis (VIGS)

[The very small world of the well-connected. Shi, Bonner, Adamic, Gilbert, *ACM Hypertext*, 2008]

- **Create sub-graphs of important vertices to study both key vertices and the entire graph**
- **Which vertices are important?**
  - People with most friends on facebook? The most quoted blogs?
  - Standard, well-established measures: Degree, Betweenness, Closeness, PageRank

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

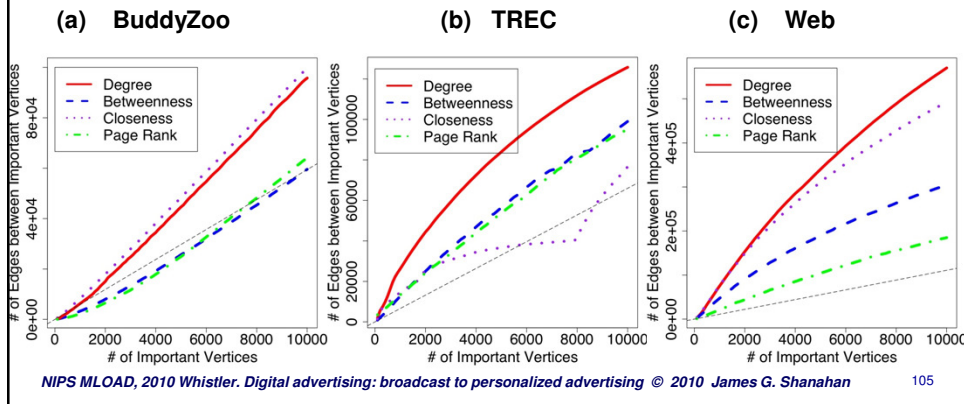
104



## VIGS in real networks: graph density

[The very small world of the well-connected. Shi, Bonner, Adamic, Gilbert, *ACM Hypertext*, 2008]

Important vertices in real networks are densely connected, no matter which importance measure we use.



## Properties of VIGS in real networks

Compare the properties of important vertices in subgraphs and in the original graphs

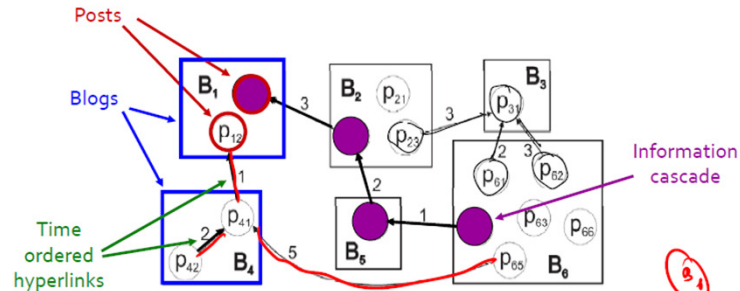
|                       | Erdos-Renyi                                       | BoddyZoo           | TREC               | Web                |
|-----------------------|---------------------------------------------------|--------------------|--------------------|--------------------|
| Density               | Less dense ↓                                      | Denser ↑           | Denser ↑           | Denser ↑           |
| Connectivity          | Disconnected when the number of vertices is small | Almost connected   | Always connected   | Always connected   |
| Average Shortest Path | Much longer ↑                                     | Almost the same 🇲🇳 | Almost the same 🇲🇳 | Almost the same 🇲🇳 |
| Rank of Importance    | Very different 🇲🇳                                 | Similar 🇲🇳         | Similar 🇲🇳         | Similar 🇲🇳         |

Get good idea of relationships between important vertices using only the subgraphs.

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

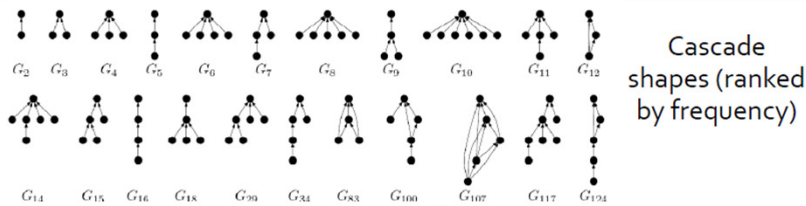
106

# Diffusion in Blogs



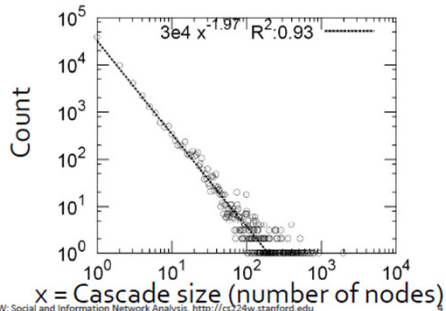
- Data – Blogs:
  - We crawled 45,000 blogs for 1 year
  - 10 million posts and 350,000 cascades

# Blogs: Cascade shapes



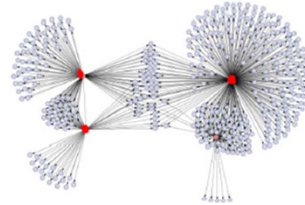
The probability of observing a cascade on  $n$  nodes follows a Zipf distribution:  

$$p(n) \sim n^{-2}$$



## Next: Creating big cascades

- Blogs – information epidemics
  - Which are the influential/infectious blogs?
- Viral marketing
  - Who are the trendsetters?
  - Influential people?
- Disease spreading
  - Where to place monitoring stations to detect epidemics?



10/13/2009

Jure Leskovec, Stanford CS224W: Social and Information Network Analysis, <http://cs224w.stanford.edu>

10

Ni

## Diffusion in partially observable networks

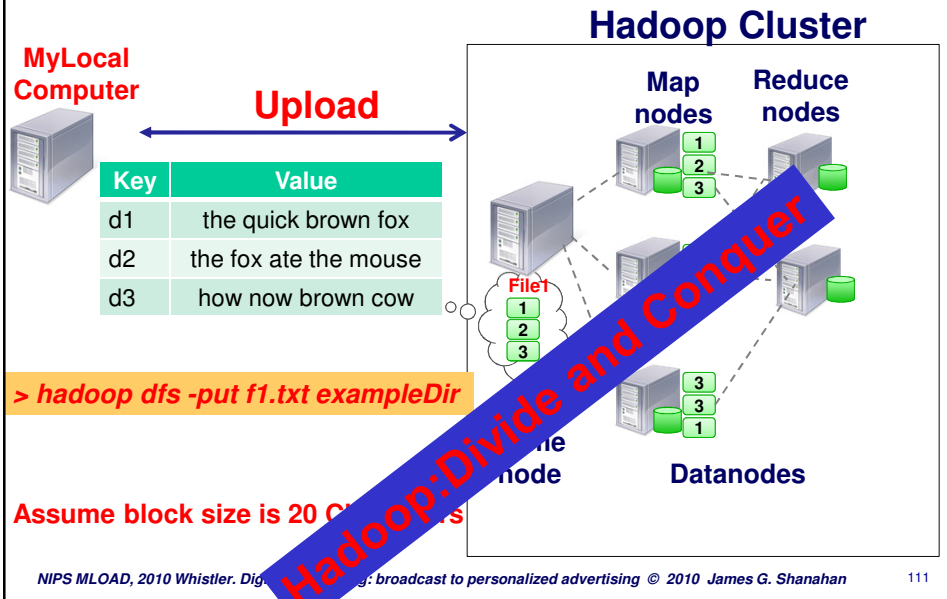
- In many applications, the underlying network over which the diffusions and propagations spread is actually *unobserved*.
- Develop a method for tracing paths of diffusion and influence through networks and inferring the networks over which contagions propagate.
- Given the times when nodes adopt pieces of information or become infected, identify the optimal network that best explains the observed infection times.
- This optimization problem is NP-hard to solve exactly
  - develop an efficient approximation algorithm that scales to large datasets and in practice gives provably near-optimal performance.

[Inferring networks of diffusion and influence Jure Leskovec, M. Gomez-Rodriguez and A. Krause. KDD 2010.]

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

110

## Graph Algs are Computational Expensive



## Outline

- **Background: digital advertising**
- **Personalization thru:**
  - Behavioral targeting
  - Interactivity: thru social ads
  - Engagement: transformative ad formats
    - Smarter and more portable devices (3G, 4G, broadband)
    - Social networking is becoming more pervasive
    - Location based services (LBS), mobile video and apps, augmenting websites with social capabilities
    - Transformative advertising
      - Augmented reality and quick response (QR) codes
  - Explainability: trading desks for the advertiser
- **Summary**



## Quick Response Codes (QR)

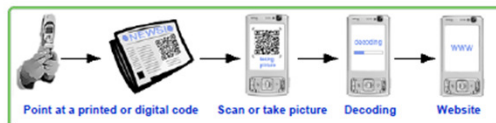
What is a QR code?

- A QR Code is a matrix barcode (or two-dimensional code), readable by QR scanners, mobile phones with camera, and smartphones



How do people use QR codes?

- Mobile users scan or take a picture of a QR code to trigger an action:
  - Display web or mobile pages
  - Save a contact
  - Make a voice call
  - Send an SMS/MMS
  - Send an e-mail
  - Promotion entry/data collection
  - Multi-action menu
  - Calendar entry
  - Send tweets
  - Display coupons/offers
  - Generate video
  - Register for/enter loyalty programs



**Download the Ray-Ban Virtual Mirror, and virtually try on the latest Ray-Ban styles.**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan 115

## Augmented reality (AR)

- **AR is a layering technology; used to overlay tagged images, video files, applications and data to display information over video, image, and camera displays on PC's, mobile phones and TV/monitors**
  - E.g., NFL game broadcasts use the yellow electronic line to indicate where the first down is located
- **AR can use GPS and location-based targeting, as well as image and facial recognition**
- **Different formats commonly used?**
  - Print-Based
    - Leverages print to drive consumer and brand engagement
    - 3-D Image & Video Use
- **Smartphone Layering**
  - Utilizes smartphone camera to
  - Geo-location content



NIPS MLOAD, 2010 Whistler. Digital advertising: b



## AR +QR: Virtual Dressing Room

117

## Engagement

- **Smarter and more portable devices (3G, 4G, broadband)**
- **Social networking is becoming more pervasive**
  - location based services (LBS), mobile video and apps, augmenting websites with social capabilities
- **Transformative advertising**
  - augmented reality and quick response (QR) codes
- **How many more rewrites**
  - Basic webpages
  - Web2.0
  - Mobile
  - IPTV

## Outline

---

- **Background: digital advertising**
- **Personalization thru:**
  - Behavioral targeting
  - Interactivity: thru social ads
  - Engagement: transformative ad formats
  - Explainability: trading desks for the advertiser
- **Other issues**
  - Credit assignment
- **Summary**

*NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan*

119

## Credit Assignment is a new area...

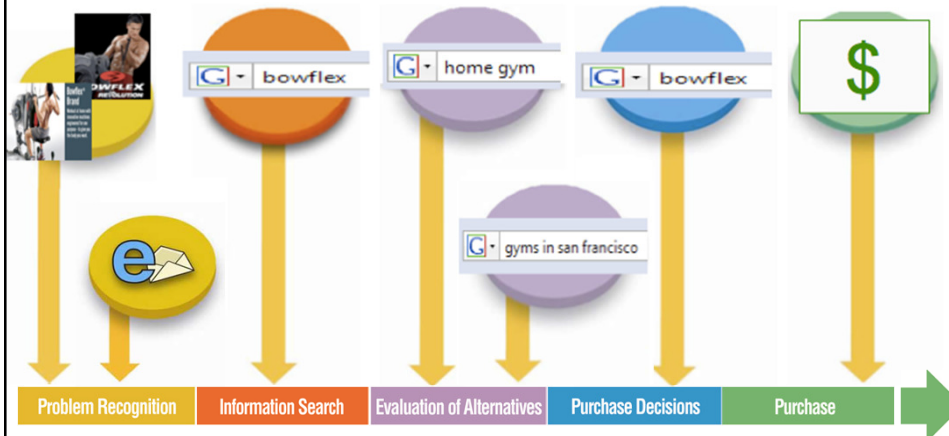
---

- **Giving credit where credit is due (offline and online)**
- **Marketers know that customers rarely travel along a purchase path in a neat, tidy, linear fashion.**
  - Typically, they can do up to five searches before getting close to a page to convert, and in the middle of those searches, can look at an e-mail blast promotion, go to a store and look at product in person, and perhaps visit influential social media sites along the way.
  - Which of these activities had a true effect on a purchase and the relative influence of each.
  - In the past, analytics relied on a “last click” model, assigning 100% of the purchasing influence to the last click a consumer made before purchasing.

<http://www.dmnews.com/attribution-management-impacting-the-bottom-line/article/163537/>



## Credit Assignment: The Customer Buying Cycle



[Figure from Goodyear & ClearSaleing Study, DMA 2010 San Francisco]

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan 121

## Outline

- **Background: digital advertising, the 0.2% opportunity (of GDP)**
- **3<sup>rd</sup> Generation DA (Personalization)**
  - Explainability: trading desks for the advertiser
  - Behavioral targeting
  - Interactivity: thru social ads
  - Engagement: transformative ad formats
- **Summary**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

122

## Key Enablers

- **Offline advertising is largely broadcast**
  - (via broadcast TV, radio, newspaper etc.) is largely a broadcast form of communication where as digital advertising is much more targeted and thus enables a personalized, and possibly informative, message to consumers.
- **“Always connected” has graduated from desktop to handheld**
  - Bridging offline with online
  - Smartphones (>50% will have one next year), Tablet
  - Life-logging capabilities
  - IP-based TV
- **Storage and Data requirements**
  - Store, Process data
  - Algorithms make all this happen

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

123

## Always connected → Lifelogging



<http://www.imrc.kist.re.kr/wiki/LifeLog>

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

124

## Research Opportunities

- Data sparsity
- Feature engineering
- Optimization theory and objectives
- Metrics
- Data storage and throughput
- User modeling
- Credit assignment
- Economic modeling, game theory, mechanism design

## From Mad Men To Wall Street and beyond!

- Set in New York City, *Mad Men* begins in 1960 at the fictional Sterling Cooper advertising agency on New York City's Madison Avenue.

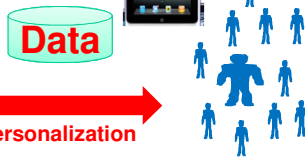
2007



Increasingly



Personalization



Human Intensive  
Lots of guess work  
Forward Market

Technology  
Data Driven  
Forward Market  
Spot Markets

Advertisers still in  
broadcast mode



1<sup>st</sup> Generation

2<sup>nd</sup> Generation

3<sup>rd</sup> Generation

## Executive Summary: Digital Advertising

- **3<sup>rd</sup> Generation Online advertising: Location → personalization**
  - Advertisers deliver a message to consumers via publishers
  - \$23.4 billion in USA (2008), \$65B worldwide (10% of overall ad spend)
  - Online advertising was location, location and location (\$11B M&A, 2007)
  - But it is more about personalization, personalization, personalization
    - Behavioral Targeting, Engaging Ads, AR, Trading Benches
    - ....and data, features, objectives, optimization, and emergent behaviors
  - **Machine learning is a key driver in the success of online advertising**
- **As a discipline**
  - **Business informatics/engineering/analytics**
  - **Bottom up:** Economics, IR, ML, OR, Game Theory, Stats, Social Sciences
  - Top-down: create (under)grad programs
    - UC Santa Cruz (ISM209, ISM250, ISM251), Stanford University

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

127

## ICWSM (Social Media) 2011 in Barcelona

Co-located with IJCAI



Fifth International AAAI Conference on Weblogs and Social Media

17-20 July 2011, Barcelona, Spain

|                   |                                                                                                                                        |
|-------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| <b>ICWSM 2011</b> | The International Confer unique venue that bring of NLP, Social Psycholog to increase our understa Research that blends so encouraged. |
| Home              |                                                                                                                                        |
| Schedule          |                                                                                                                                        |
| Important Dates   | The 2011 meeting will b government innovators media to increase transp citizenry.                                                      |
| Call for Papers   |                                                                                                                                        |

**IMPORTANT DATES:**  
**Sponsorship immediately!!**  
**Tutorial proposals due: Feb 18, 2011**  
**Workshop proposals due: Feb 18, 2011**  
**Papers due: Jan 31, Feb 6, 2010**  
**Conference July 17-20**

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

128

---

# Thanks

**EMAIL:**  
***James\_DOT\_Shanahan\_AT\_gmail\_DOT\_com***

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

129

## **Digital advertising is a frenemy culture**

---

- Digital advertising is “frenemic” (players are simultaneously a partner and a competitor).

NIPS MLOAD, 2010 Whistler. Digital advertising: broadcast to personalized advertising © 2010 James G. Shanahan

130