Advertising market and Social Correlation in Social networks

Nikos Salamanos
(salaman@aueb.gr)

Db-Net Research Group

Athens University of Economics and Business
Outline

1. Advertising Network Formation
   • Motivation & Problem Statement
   • Preliminaries
   • Stochastic Diffusion Market Search
   • Experimental Evaluation

2. Correlation and Influence in Social Networks
   • Social Correlation and Influence
   • Test for the source of correlation
   • A proposed case study: Correlation in Facebook
Advertising Network Formation

Our Goal:
- A agent based method for online Advertising market formation
  - The Advertising matching has to be formed by the economic and social interactions of the agents

Contribution:
- We apply a Swarm Intelligent model in online advertising context
- We propose a dynamic method that converges to a stable state and is economic efficient
Online Advertising

- Major source of income for search engines

- Basic forms:
  1. Sponsored Search:
     - The advertisements displayed in slots alongside search results
  2. Online Advertising Network
Online Advertising Network

- **Ad-Network**: A network that connects advertisements to web sites (publishers)

- **Ad-publishers**: Web-sites that sell advertising space
  - Are willing to publish ads at his web-site

- **Advertisers**: Consumers, are willing to buy advertising space

- **Ad-words**: The advertisement

- Example: Google’s **Ad-Sense** network
Our Approach

- **Google’s Ad-Sense network:**
  - Google has the total control of the advertising matching
  - There is no direct communication between Advertisers and Ad-Publishers
  - **Pricing Schema:** Online Auctions

- **Our Model:**
  - Dynamic and decentralized advertising matching
    - **Consumers’ Behavioral model** based on *Stochastic Diffusion Search* a Swarm Intelligent model
    - **Pricing Schema** based on algorithmic *market equilibria* theory
Problem Statement

- A set of consumers enter to an unknown market
  - Consumers = The advertisers = The ads
  - The advertisers do not have predefined independence preferences for the publishers / goods

- The market consists of a set of goods
  - Goods = The advertising slots = The publishers

Problem: Propose an economic efficient model that maps ads to publishers and the ad-matching is generated by the entities of the network
Stochastic Diffusion Search (SDS)

- Swarm Intelligent method
  - Proposed by Mark Bishop (1989) as pattern matching heuristic

- Solve a class of Optimization Problems
  - The objective function is decomposable into components that can be evaluated independently

- Population Based method
  - Agents operate synchronously searching for the optimal solution of a given optimization problem
1. **Initialize Agents**
   - *Hypothesis* = candidate solution

**Repeat:**

2. **Test – Agents**
   - If test *True*
     - Then *Active (Satisfied)*

3. **Diffuse – Agents**
   - Inactive agent select an agent at random
   - If selected agent *active*
     - Then copy hypothesis
Market Equilibria

Fisher Market: Special case of exchange market
  • Entities: buyers (money) and goods
  • Linear Fisher Market: Buyers’ utility is a linear function

Equilibrium State:
  • Market Clearing Prices: The supply is equal to demand.
  • All the goods have been sold and there is no money to be spent
Market Equilibria (cont.)

- Equilibrium Linear Fisher Market
  - Unique vector of prices
  - Unique optimal allocation of goods
  - Fair allocation
    “If you want a good you get it”
  - Competition monotonicity
    “Increasing your budget only you could benefit from that”

- Fleischer et al. (2008) proposed an polynomial time algorithm which computes an \((1 + \varepsilon)\)-approximate market equilibria.
  - \((1 + \varepsilon)\) - approximation: The demand is at most \((1 + \varepsilon)\) times the supply
Stochastic Diffusion Market Search

Advertisers \rightarrow Hypothesis \rightarrow Market Equilibria Algorithm

Supply Estimation

Publishers

Ad-Network

Test True?

YES

NO

Advertisers’ Social Network

Find Friends

Updated Hypothesis
Ad Market Formulation

Basic notions:

- **Impression**: A single appearance of an advertisement on a web-site

- **Cost per Impression (CPI)**: The price of a single appearance

- **Click Through Rate (CTR)**: A quality metric.

\[
CTR = \frac{\text{Number of user clicks on an ad}}{\text{Number of times the ad was published}}
\]
Ad Market Formulation (cont.)

- A *Linear Fisher Market* defined in every period

- Two group of agents: Publishers and Advertisers

- Ad-publisher = Ad-slot = Good
  - Divisible good
    - Quantity = number of impressions for sell

\[ f_j^t = \text{Number of visitors in period } t-1 \]
Advertiser = Consumer

- **Budget** = The amount of money he is willing to spend

- **Assumption**: Utility function linear
  - **Utility per impression**:
    \[ u_i^t(J) = CTR_i^{t-1}(J) \]

- The utility of advertiser i when obtains one impression of the publisher J
Input:
1. Consumers’ hypothesis for the optimal set of goods

2. Supply: $\text{impressions for sell}_j^t = \text{visitors}_j^{t-1}$
SDMS-Test Phase (cont.)

- Market Equilibria Alg.
  - Output:
    - Advertising matching = Basket of Goods
    - Cost Per Impression Prices
SDMS-Test Phase (cont.)

- **Ad-Network** $\rightarrow$ **Realized CTR** values

- **Test Criterion:**
  - If "Yes" $\rightarrow$ **Active** advertiser $\rightarrow$ **Retest** Hypothesis
  - Else $\rightarrow$ **Inactive** advertiser $\rightarrow$ go to **Diffusion Phase**
Test Criteria

1. **Quality**: “I want the best goods”

   - **Active if:**
     \[
     \text{CTR}_i^t \geq \theta \cdot \text{CTR}^t, \text{ where } 0 < \theta \leq 1
     \]

     \[
     \text{CTR}_i^t = \frac{\text{total clicks on i's ad}}{\text{total impressions consumed by i}}
     \]

     \[
     \text{CTR}^t = \frac{\text{total clicks on ads}}{\text{total impressions consumed in the market}}
     \]

   - The distance from the average utility of the market
   - i’s average utility
   - Market average utility
Test Criteria (cont.)

2. **Bang per Buck = utility per unit amount of money spent**

\[
BpB_i^t = \frac{\text{total clicks on i's ad}}{i's \text{ budget}}
\]

- **Active if:**
  1) \(BpB_i^t \geq \theta \cdot BpB_{\text{market}}^t\)
  2) \(BpB_i^t \geq \theta \cdot BpB_{\text{friends}}^t\)
  3) \(BpB_i^t \geq \theta \cdot \max BpB_{\text{friends}}^t\)
Find Friends $\rightarrow$ Consumers’ Social Network

- Choose a Friend at random
  - If Friend “Active” AND with better basket of goods
    - then: adopt friend's preferences
  - Else: choose at random one new good
Weight each “friend” \( k \)

- **Social Criterion:**

\[
w_i(k) = \alpha \cdot (\text{in degree}(k) + 1) + (1 - \alpha) \cdot (\#\text{common - friends} + 1)
\]
Experimental Evaluation

- Network of 100 X 100 agents
- Synthetic data set:
  - Publishers’ Visitors:
    - We pick with uniform probability an integer from the interval [100, 1000]
  - Advertisers’ Budgets
    - We pick with uniform probability an integer from the interval [1000, 3000]

- Realized CTR( i , J): Zipf distribution
  - We generate for each advertiser 100 Zipf values (Zipf exponent 1.0)
Experimental Evaluation (cont.)

- **Advertisers’ Friends:**
  - For each advertiser:
    - We assign with uniform probability a integer from the interval [0, 9]. Represents the *color* of advertiser
    - Advertisers with the same *color* defined as *friends*
    - 10 groups of *friends*

- **Observe** that this process generates groups of friends with different realized CTR values on the same set of publishers

- **Assumptions:**
  - Constant number of visitors during the periods
  - Constant realized CTR values during the periods
**Convergence State**

(a) **Convergence State**: Number of Active agents during 110 periods

![Graph showing the convergence state]
b) **Social Welfare**: Total utility value in the market during 110 periods
CPI Prices

- After 110 periods

a) Avg.CTR_market

b) Avg.BpB_market

c) Avg.BpB_friends

d) max-BpB_friends
Full information: Advertisers are aware of the realized CTR values.

c) Distribution of CPI prices in SDMS in the 110th period

d) Distribution of CPI prices in the full information case after 110 periods
The distribution of

\[
\text{ratio} = \frac{\text{total utility in SDMS}}{\text{Total utility in Full information case}}
\]
Future Study

- Synthetic Data set where the realized CTR values are correlated with each group of friends
- Other economic criteria
- Other Social metrics
  - Page rank?
- A variation of the model
  - Split the consumer into several buyers with small budget
  - Diffuse the new buyers
Correlation in Social Networks

- **Social Correlation**: user's actions can be correlated to his social affiliations

- **Social influence**: Actions of the user induce his friends to behave in a similar way.
  - Technologies, ideas, customs diffuse through the social network

- Identifying social influence is difficult
  - Unobserved variables induce statistical correlation between the actions of friends

- Applications: Viral Marketing
Social Influence (cont.)

- Aris Anagnostopoulos et al. (KDD 2008)

  - Define a general model of Social Correlation
  
  - Propose two tests that can identify social influence as a source of social correlation
  
  - Case study: Tagging behavior on Flickr
    - Prove that:
      - Significant social correlation
      - No presence of social influence
The model

- **Social Network**: A directed graph $G$.
  - Nodes: the agents
  - Edges: connect friends

- **Active**: An agent becomes active when it performs an action
  - Time period: $[0, T]$  
  - $W$: The set of active agents at the end of the time period

- **Social Correlation**: Activation of $i$ and $j$ are correlated events
Source of Correlation

- **Homophily:** Agents choose friends with similar characteristics
  1. Choose the set $W$ according to a distribution
  2. Pick graph $G$ from a distribution that depends on $W$

- **Confounding:** External influence
  - Example: Two agents become friends because they live in the same city. $\rightarrow$ They perform similar actions
  1. Define a confounding variable $X$
  2. Pick graph $G$ and set $W$ from distributions that are correlated with $X$
Source of Correlation (cont.)

- **Correlation model**: Generalization of Homoplily and Confounding
  1. Pick G and W according to a joint probability distribution
  2. Pick the time of activation in W according to a distribution on $[0, T]$ 

- **Influence**:
  1. Pick G according to a probability distribution
  2. In each time step 1, ..., T each inactive agent decides whether to become active
  3. $\text{Prob}_{\text{agent active}} = \text{function of the number of friends that are already active}$
Methodology

- **Logistic Regression:**
  - Prob. of activation:
    \[ p(\alpha) = \frac{e^{a \ln(\alpha + 1) + b}}{1 + e^{a \ln(\alpha + 1) + b}} \]
  - \( \alpha \) active friends (independent variable)
  - \( a, b \) coefficients
  - Large value of \( a \) indicates large degree of correlation
  - Compute \( a, b \) using maximum logistic regression
The Shuffle Test

- Intuition: If the influence is not the source of correlation then:
  - The activation depends on the number of active friends
  - but the timing of activation is independent from the timing of other agents

1. G graph and \( W = \{w_1, \ldots, w_L\} \) active users during period \([0, T]\)
   - Let \( w_i \) activated at time \( t_i \)
   - Estimate coefficient \( a \)

2. Generate a second problem:
   - Random Permutation \( \pi \) of \( 1, \ldots, L \)
   - Set \( t'_i = t_{\pi(i)} \)
   - Estimate \( a' \)

- If \( a \approx a' \) then no social influence
Social Correlation on Facebook

- Huge social network
- Available personal data
- Fast spread of norms of behavior

Goal:
- Identifying the source of correlation
- Patterns of graph evolution
We create a pseudo identity “Maria”
Intuition (cont.)

- “Maria” subjects the “Free Greek TV Online” (Spam-Group)

- **118.938 members!**

- Several spam members join the group
  - Homoplply?
Our approach

- Methodology…
- Experiments…
- Tools…
Conclusion

- We apply a Swarm Intelligent schema in online advertising context

- Good properties: Convergence to a stable state and economically efficient

- We propose the study of social correlation in Facebook
Thank You!