On Threshold Behavior in Query Incentive Networks

Esteban Arcaute\textsuperscript{1}  Adam Kirsch\textsuperscript{2}  Ravi Kumar\textsuperscript{3}  David Liben-Nowell\textsuperscript{4}  Sergei Vassilvitskii\textsuperscript{1}

\textsuperscript{1}Stanford University  \textsuperscript{2}Harvard University  \textsuperscript{3}Yahoo! Research  \textsuperscript{4}Carleton College

The 8th ACM Conference on Electronic Commerce EC’07
Motivation

Trusted answers? Ask your friends!
Online friends? Use incentives!

Model

Mathematical Formulation
Branching Process and Framework
Objective

Results

Previous Results
Our Results
Discussion
Current Research
Model introduced by Kleinberg and Raghavan [FOCS ’05]

- Assume that a user, say $u$, of a social network has a question (e.g. Where to find a good physician?)
- Suppose that some subset of users have an answer
- How would $u$ retrieve an answer from those individuals?
An Answer or The Answer

Differences

To get an answer, \( u \) could:
- use a search engine; or
- ask friends.

What’s the difference?
- Search engine: many answers \textit{but} may not be reliable
- Friends: trusted answers \textit{but} may not have any

Not enough friends? Reach friends’ friends!
\[ \Rightarrow \text{“web of trust”}. \]
Ask Your Friends, Please

- Reaching friends’ friends through incentives
- Offer payment for answers
  \[ \leftrightarrow \text{utility transfer} \]
- Users act as strategic agents

Natural question: how much should \( u \) offer?
Informal Description

Key Ideas to Model

Key features from Kleinberg and Raghavan’s model.

- **Nodes and answers:**
  - all answers are created equal
  - each person, independently, has an answer with probability $\frac{1}{n}$
- Users aware of only local topology
  - model with a random graph
- Providing incentives to answer, *not* creating a market
Network, Agents and Incentives

- Underlying network: complete $d$-ary tree ($d > 1$)
- Root: special node with query (question)
Network, Agents and Incentives

- Underlying network: complete $d$-ary tree ($d > 1$)
- Root: special node with query (question)
- Realized network: each node has (independently) $0 \leq i \leq d$ children with distribution $C$
  identities of nodes chosen uniformly at random
Network, Agents and Incentives

- Underlying network: complete $d$-ary tree ($d > 1$)
- Root: special node with query (question)
- Realized network: each node has (independently) $0 \leq i \leq d$ children with distribution $C$

identities of nodes chosen uniformly at random
Completing the Model

For the incentives:

- parent node offers reward for answer to children
- if agent has an answer, communicates it to parent
- if there are many answers, choose one uniformly at random
- if providing answer, pay unit cost
Completing the Model

For the incentives:

- parent node offers reward for answer to children
- if agent has an answer, communicates it to parent
- if there are many answers, choose one uniformly at random
- if providing answer, pay unit cost

Formally, if a node is offered $r$ and doesn’t have an answer

Tradeoff faced by the node: if it offers $f(r)$,

- amount it keeps $r - f(r) - 1$
- probability of finding an answer in subtree increases with $f(r)$

Solution concept: Nash Equilibrium
On Threshold Behavior in Query Incentive Networks

Arcaute
Kirsch
Kumar
Liben-Nowell
Vassilvitskii

Motivation
Trusted answers? Ask your friends!
Online friends? Use incentives!

Model
Mathematical Formulation
Branching Process and Framework
Objective

Results
Previous Results
Our Results
Discussion
Current Research

Schema of Incentives

- offer $r$
- offer $f(r)$
- offer $f(f(r))$

Mathematical Formulation

- payoffs: $f(r) - f(f(r)) - 1$
- payoffs: $f(f(r)) - 1$
- payoffs: $r - f(r) - 1$
Motivation
Trusted answers? Ask your friends!
Online friends? Use incentives!

Model
Mathematical Formulation
Branching Process and Framework
Objective

Results
Previous Results
Our Results
Discussion
Current Research

Schema of Incentives

offer $r$

offer $f(r)$

offer $f(f(r))$

payoffs:

$r - f(r) - 1$

$f(r) - f(f(r)) - 1$

$f(f(r)) - 1$
Model as Branching Process
Parameters

- $\mathcal{C}$ distribution with support \{0, ..., $d$\}
  let $b$ be its expectation
- Realized network: realization of branching process according to $\mathcal{C}$
- identities of nodes chosen uniformly at random

\[
b > 1 \implies \text{infinite network with constant probability}
\]

- Average number of nodes in the first $k$ layers:
  \[
  \frac{1 - b^{k+1}}{1 - b} = \Theta\left(b^k\right)
  \]

- In $\Theta(\log n)$ layers, one answer with constant probability
Objective

Given

- probability of success $1 > \sigma > 0$;
- the distribution $C$;
- the rarity of the answer $n$; and
- agents play a Nash Equilibrium given by the function $f$

Find minimum offer $R_{\sigma,C}(n)$ to get answer with probability at least $\sigma$

Study dependency of $R_{\sigma,C}(n)$ on $C$ and $\sigma$
Kleinberg and Raghavavan
Main Result

Setting:
- each child present independently at random
  \( \leftrightarrow \mathcal{C} \) is a binomial distribution
- expected number of children \( b \)
- \( \sigma \) is a constant

Results:
- If \( 1 < b < 2 \), then \( R_{\sigma,\mathcal{C}}(n) = n^{\Omega(1)} \)
- If \( b > 2 \), then \( R_{\sigma,\mathcal{C}}(n) = O(\log n) \)

Phase transition for rewards, but nothing obvious happening from a structural perspective!
Summary of Results

In this paper, we consider the robustness of Kleinberg and Raghavan’s original result with respect to

- the distribution $C$: result is robust; and
- the success probability $\sigma$: result is not robust
Robustness with respect to $\mathcal{C}$

Given:

- $\sigma = O(1)$
- $d = O(1)$
- an arbitrary distribution $\mathcal{C}$ with support $\{0, 1, \ldots, d - 1, d\}$

Theorem

For all $\sigma$, $d$ and distributions $\mathcal{C}$ as defined above, we have that

- If $1 < b < 2$, then $R_{\sigma,\mathcal{C}}(n) = n^{\Theta(1)}$
- If $b > 2$, then $R_{\sigma,\mathcal{C}}(n) = O(\log n)$
High Probability Case: Vanishing Threshold

- We want $\sigma = 1 - o(1)$

**Given:**

- $\sigma_0 = 1 - \frac{1}{n}$
- $d = O(1)$
- an arbitrary distribution $\mathcal{C}$ with support $\{1, \ldots, d - 1, d\}$

**Theorem**

*For all $\sigma > \sigma_0$, $d$ and distributions $\mathcal{C}$ as defined above, we have that*

- If $1 < b < 2$, then $R_{\sigma, \mathcal{C}}(n) = n^{\Theta(1)}$
- If $b > 2$, then $R_{\sigma, \mathcal{C}}(n) = n^{\Theta(1)}$
Discussion of Results

Let $\ell$ be the expected path length to an answer.

For $\sigma$ constant:
- $\ell = \Theta(\log n)$
- $2 > b > 1$, reward exponential in $\ell$
- $b > 2$, reward of same order as $\ell$

For $\sigma \geq 1 - \frac{1}{n}$:
- $2 > b > 1$, still exponential in $\ell$
- $b > 2$, also exponential in $\ell$ but blowup occurs in the last $O(\log \log n)$ steps
Current Research and Open Problems

Many open directions remain:

- Different network topology
- Aggregate answers

Most important open problem: probabilistic interpretation/proof of results.
Thank you