Query Incentive Networks

Jon Kleinberg
Cornell University

Prabhakar Raghavan
Yahoo! Research

FOCS '05, pages 132-141

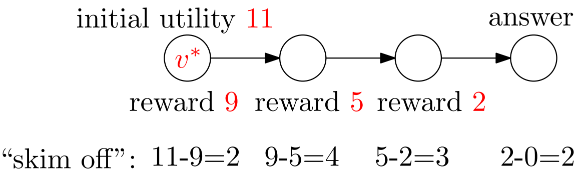
Presented by Jian XIA

for COMP670O: Game Theoretic Applications in CS
Spring 2006, HKUST

May 8, 2006

Incentive Query Propagation

- Root (seeking information or services) issues queries, together with a reward.
- Each node is offered a reward by its parent, and offers a (smaller) reward to its children for continued propagation.
- The propagation down each branch stops when the reward reaches 0, or when a node with an answer is reached.
- The reward is only paid if on the chosen path to the answer.



Motivations: Peer-to-Peer networks, social-networking systems

Basic Questions

- How much reward is needed in order for the root to achieve a reasonable probability of obtaining an answer?
- What is the effect of network topology?
- What is the strategic behavior of each node?
  — How much can it skim off and yet appear on the chosen path to the answer?

Game Formulation

For each node (player) v,

- (integer-valued) utility: $r > 0$ (reward form its parent)
- (integer-valued) strategy function: $f_v(r) < r$ (reward to its children)

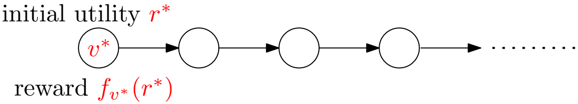
$$\text{reward } r \longrightarrow v \longrightarrow \text{reward } f_v(r)$$

$$\text{“skim off”:} \qquad r - f_v(r)$$

- Each node holds the answer with probability $1 - p > 0$
  — answer rarity: $n = \frac{1}{1-p}$
  — one of n nodes holds the answer on average

First Attack: Line Model

- An infinite line L starting from the root v^*



Auxiliary functions:

- $\alpha_v(\mathbf{f}, x)$: the probability the nodes after v yield the answer, given that v offers reward x and that v does not possess the answer, where $\mathbf{f} = \{f_v : v \in L\}$
- $\beta_v(\mathbf{f}, x) = 1 - \alpha_v(\mathbf{f}, x)$
- $\beta_v(\mathbf{f}, x) = p \cdot \beta_w(\mathbf{f}, f_w(x))$, for any v and its child w

Existence of the Nash Equilibrium

Definition of Nash equilibrium $\mathbf{g}$
— all functions g_v are the same

- base: $g_v(1) = 0$ for all nodes v
- induction: assume that $g_v(x)$ has been defined for all v and all $x < r$, then

$$g_v(r) = \arg\max_x \, (r - x)\alpha_v(\mathbf{g}, x)$$

Theorem

The set of functions $\mathbf{g}$ *is a Nash equilibrium.*

Existence of the Nash Equilibrium

Idea: $(r-x)\alpha_v(\mathbf{g},x)$ is proportional to the portion of the payoff over which v "has control".

Proof: Given the initial utility r^* at the root, and a choice of functions $\mathbf{f}$ at each node, consider the following four events:

- C: the query reaches v
- B: an answer is found in the sub-line starting from v (including v)
- A: the reward is propagated down to v
- D: v holds the answer

Define a random variable $Y_{\mathbf{f},r^*}$ denoting the payoff to v, given $\mathbf{f}$ and r^*, then it can be shown that

$$\begin{aligned}
E[Y_{\mathbf{f},r^*}] =& E[Y_{\mathbf{f},r^*}|A\cap B\cap C\cap D]\cdot \Pr[A\cap B\cap C\cap D] \\
&+ E[Y_{\mathbf{f},r^*}|A\cap B\cap C\cap \overline{D}]\cdot \Pr[A\cap B\cap C\cap \overline{D}] \\
&= r\cdot \Pr[A\cap B\cap C\cap D] \\
&+ (r-f_v(r))\cdot \Pr[A|B\cap C\cap \overline{D}]\cdot \alpha_v(\mathbf{f}, f_v(r))\cdot \Pr[C\cap \overline{D}]
\end{aligned}$$

Breakpoint Structure of Rewards

- hop function: $h(r)$, the number of hops the initial reward r can be pushed further
  — the number of times we have to iterate the function g to reduce an initial reward of r down to 0
  — if the first node holding the answer is within $h(r)$ hops from the root, everyone from the root to this node gets paid (else, no one gets paid)
- $\hat{\phi}_j$: the probability that no node in the first j hops has the answer, given that the root does not
  — $\beta_{v^*}(\mathbf{g}, r) = \hat{\phi}_{h(r)}$
  — $\hat{\phi}_j = p \cdot \hat{\phi}_{j-1} = p^j$
  — $g_v(r) = \arg\max_x \, (r - x)(1 - p^{h(x)})$
- breakpoint: u_j, the minimum initial reward r to push j hops
  — $h(u_j) = j$

Growth rate of the breakpoints

- $\Delta_j = u_j - u_{j-1}$

Since $g_v(r) = \arg\max_x \, (r - x)(1 - p^{h(x)})$, we have,

$$\Delta_j(1 - p^{j-1}) \geq (\Delta_j + 1)(1 - p^{j-2})$$

therefore,

$$\Delta_j \geq \frac{1}{1-p}(\frac{1}{p^{j-2}} - 1) = n((\frac{n}{n-1})^{j-2} - 1)$$

For rarity n, the distance from the root to the nearest answer is $O(n)$ with high probability in the line model. It follows,

Theorem

In line model, the utility required for v^* to find the answer with constant probability is exponential in n.

Second Attack: Branching Process Model

- infinite complete d-ary tree T rooted at node r^*
- each node in T is active independently with probability q
- random subtree T' of T: the set of all nodes reachable from the root using paths consisting entirely of active nodes
  — T' is a Galton-Watson tree generated from an offspring distribution that produces j children with probability $\binom{d}{j} q^j (1-q)^{d-j}$ (binomial distribution)
  — expected number of offspring per node (branching factor): $b = qd$
  - $b < 1$: T' is almost surely finite
  - $b > 1$: there is a positive probability of obtaining an infinite tree T'

Structural lower bound: For an infinite random tree T', the distance in T' to the nearest answer is $O(\log n)$ with high probability.

Main Results

(Sufficient) competition makes incentive networks efficient.

- When $b < 2$, the utility required for v^* to find the answer with constant probability σ is $\Omega(n^c)$.
  — exponential in the path length from the root to the answer
  — knowing fewer than 2 people is expensive
- When $b > 2$, the utility required for v^* to find the answer with constant probability σ is $O(\log n)$.
  — linear in the path length from the root to the answer

Existence of the Nash Equilibrium

Auxiliary functions:

- $\alpha_v(\mathbf{f}, x)$: the probability the subtree of T' below v yields the answer, given that v offers reward x and that v does not possess the answer, where $\mathbf{f} = \{f_v : v \in T\}$
- $\beta_v(\mathbf{f}, x) = 1 - \alpha_v(\mathbf{f}, x)$
- $\beta_v(\mathbf{f}, x) = \prod_{w \text{ child of } v} [1 - q(1 - p\beta_w(\mathbf{f}, f_w(x)))]$

Definition of Nash equilibrium $\mathbf{g}$

— all functions g_v are the same

- base: $g_v(1) = 0$ for all nodes v
- induction: assume that $g_v(x)$ has been defined for all v and all $x < r$, then $g_v(r) = \arg\max_x (r - x - 1)\alpha_v(\mathbf{g}, x)$

  — For technical reason

  — place unit cost on the effort of establishing the "connection" along the path from the root to the chosen node

Theorem

The set of functions $\mathbf{g}$ *is a Nash equilibrium.*

Breakpoint Structure of Rewards

- hop function $h(r)$ — the number of times we have to iterate the function g to reduce an initial reward of r down to 0
- $\hat{\phi}_j$: the probability that no node in the first j hops has the answer, given that the root does not
  — $\beta_{v^*}(\mathbf{g}, r) = \hat{\phi}_{h(r)}$
  — $\hat{\phi}_{j+1} = (1 - q(1 - p\hat{\phi}_j))^d$
  — $g_v(r) = \arg\max_x \, (r - x - 1)(1 - \hat{\phi}_{h(x)})$
- breakpoint: u_j, the minimum initial reward r to push j hops
  — $h(u_j) = j$

Breakpoint Structure of Rewards

Inductive definition of u_j

- base: $u_1 = 1, u_2 = 2$ (Since $g(1) = 0, g(2) = 1$)
- induction: assume that $u_1, \ldots, u_j$ has been defined,
  — for a given utility r, define the following linear functions of r:
  $$\ell_i(r) = (r - u_i - 1)(1 - \hat{\phi}_i),$$
  the payoff to the root when it offers reward u_i with utility r
  — for all $r \geq u_{j-1}$, $\ell_{j-1}(r) > \ell_{j-2}(r) > \cdots > \ell_1(r)$
  — at $r = 1 + u_j$, we have $\ell_j(r) = 0$, and hence $\ell_j(r) < \ell_{j-1}(r)$
  — $u_{j+1} = \lceil y_{j+1} \rceil$, where y_{j+1} is the reward value of the cross point of ℓ_j and ℓ_{j-1}, the value y for which
  $$(y - u_j - 1)(1 - \hat{\phi}_j) = (r - u_{j-1} - 1)(1 - \hat{\phi}_{j-1})$$
- Define $\Delta'_j = y_j - u_{j-1}$ and $\Delta_j = u_j - u_{j-1}$, we have
  $$1 + \frac{\Delta_j}{\Delta'_{j+1} - 1} = \frac{1 - \hat{\phi}_j}{1 - \hat{\phi}_{j-1}}$$

Growth rate of the breakpoints

- $R_\sigma(n,b)$: the minimum utility needed by v^* in order for the query process to yield an answer with probability at least σ — given an answer rarity $n = (1-p)^{-1}$, a success probability σ, and a branching factor $b = qd$
  — Which is the asymptotic dependence of $R_\sigma(n,b)$ on n and b?
  — assumption: $\sigma \gg n^{-1}$

Useful lemmas for function $t(x) = (1 - q(1-px))^d$
Claim: Fix ε such that $\frac{1}{dn} < \varepsilon < 1$. If $x \in [1-\varepsilon, 1]$, then $t'(x) \in [pb(1-2bd\varepsilon), pb]$.
Claim: $1 - \frac{b}{n} \leq t(1) \leq 1 - \frac{1}{dn}$.
Claim: Suppose p, b and ε are such that $pb(1-2bd\varepsilon) > 1$, and let $0 < \gamma_0 < \gamma_1 \leq \varepsilon$. Let $N(\gamma_0, \gamma_1)$ denote the number of iterations of the function t needed to reduce $1-\gamma_0$ to a quantity that is $\leq 1-\gamma_1$. Then $N(\gamma_0,\gamma_1) = \Theta(\log(\gamma_1/\gamma_0))$.

The case when $b < 2$

For a fixed $b < 2$, consider the sequence of $\hat{\phi}_j$ values up to the point at which it drops below $1 - \sigma_0$, for a small constant $\sigma_0 < \sigma$ — $pb(1 - 2bd\varepsilon) > 1$: $\sigma_0 < \sigma$ small enough, and n large enough

- first segment: the set I_1 of indices j for which $\hat{\phi}_j \geq 1 - \kappa_0/n$ for a constant κ_0
- second segment: the set I_2 of indices j for which $1 - \kappa_0/n > \hat{\phi}_j \geq 1 - \sigma_0$

Lemma: The first segment has length $O(1)$ and the second segment has length $\Theta(\log n)$.

Lemma: There is a constant $b_1 < 2$ such that for all j in the second segment we have $\frac{1-\hat{\phi}_{j+1}}{1-\hat{\phi}_j} \leq b_1$.

The case when $b < 2$

Theorem

There is a constant $c > 1$, depending on b, so that if $\hat{\phi}_j < 1 - \sigma$, then $u_j \geq n^c$. Hence $R_\sigma(n, b) \geq n^c$.

Proof.

Since $\Delta_j = u_j - u_{j-1}$, we have $\Delta_j = u_1 \prod_{i=3}^{j} \frac{\Delta_i}{\Delta_{i-1}}$.

Since $1 + \frac{\Delta_j}{\Delta'_{j+1} - 1} = \frac{1 - \hat{\phi}_j}{1 - \hat{\phi}_{j-1}}$, if j belongs to the second segment of indices, then

$$\frac{\Delta_{j+1}}{\Delta_j} \geq \frac{\Delta'_{j+1}}{\Delta_j} \geq \frac{\Delta'_{j+1} - 1}{\Delta_j} = \frac{1}{\frac{1 - \hat{\phi}_{j+1}}{1 - \hat{\phi}_j} - 1} \geq \frac{1}{b_1 - 1} > 1.$$

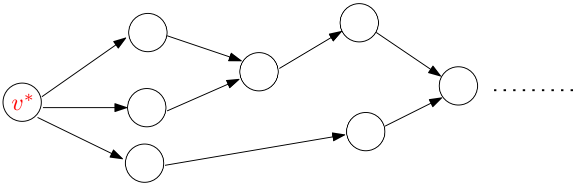
Let $c_0 = \frac{1}{b_1 - 1}$. Since the second segment has length $\geq \tau \log n$ for a constant $\tau > 0$, we have

$$u_j \geq \Delta_j = u_1 \prod_{i=3}^{j} \frac{\Delta_i}{\Delta_{i-1}} \geq \prod_{i \in I_2} \frac{\Delta_i}{\Delta_{i-1}} \geq c_0^{|I_2|} \geq c_0^{\tau \log n} = n^{\tau \log c_0}.$$

Third Attack? : Directed Acyclic Graph Model

Model of real competition
— in tree model, limited form of competition exists

- directed acyclic graph with a single root node r^*
- Node may receive different reward offers from different parents
- Node must take into account that its children can get competing offers



More Attacks...

- What if we have to piece together multiple answers?
- What if the overall utility is a formula on individual answers? — a single-player case studied in Charikar *et al.* 2000
- What if the answer is enhanced along the path back to the root?

Thank You!

- based on Dr. Raghavan's PowerPoint Presentation (www.la-web.org/spire2005/raghavan.pdf)
- thanks to Yan ZHANG for discussions on line model.