Algorithm to Trade Off between Utility and Privacy Cost of Online Social Search

Yan Li, Zhiyi Lu and Victor O. K. Li
Email: liyanhku@gmail.com
Department of Electrical and Electronic Engineering, The University of Hong Kong

Outline

- Online social search website
  - Seek answers from experts by using the question-and-answer social network website.

  - User looking for information can pose a question and send it to his friends or person recommended by the system.
  - User who get this question may answer it or forward it to others.

The trade-off between utility and privacy cost of online social search

Utility

- OSS can take the advantage of the OSNs to look for experts.
- The users who poses the question may get utility as the question may finally reach the experts and get a great number of responses.

Privacy cost

- When a question is asked and passed around to other users along friendship links, the questioner's personal information may also be exposed.
- The more the number of people who have received the question, the higher the questioner's privacy exposure.

Framework of the utility and privacy cost of online social search

- The network graph: $G = (V, E, L)$
  - $|V| = n$ vertices and $|E| = m$ edges
  - For every edge $(u, v) \in E, p(u, v)$ denotes the probability of the influence from u to v : Independent Cascade model
- The measurement of utility
  - $L = \{l_1, l_2, \dots, l_k\}$ is the set of labels to indicate expertise in various fields. $L = \{computer\ science, economics,\ geogrophy, \dots\}$
  - Each node $u \in V$ has a set of labels $LB(u) \subseteq L$. $L_e \subseteq L$ represents the expertise required.
  - p_{l_i}: utility value $p_{LB(u)} = \sum_{l_i \in LB(u)} p_{l_i}$

Framework of the utility and privacy cost of online social search

- The measurement of privacy cost
  - $PI = \{pi_1, pi_2, \dots pi_m\}$: all kinds of personal information for one person in the system.
  - $PIS \subseteq PI$:any information spread can be regarded as a set of personal information.
  - c_{p_i}: privacy cost of personal information pi_i $\mathrm{C} = \sum_{pi_i \in PIS} c(pi_i)$
- The information diffusion model : Independent Cascade model (IC model)
  - A user may choose a set of seed nodes: $S \subseteq V$
  - S_t: node set newly activated at time t $S_0 = S, S_t \cap S_{t-1} = \emptyset$
  - At time $t+1$, every node $u \in S_t$ tries to activate its neighbors $v \in V \setminus \cup_{0 \le i \le t} S_i$ independently with probability $p(u, v)$
  - $A(S)$: the set of nodes activated by the seed set S $\sigma(S)$: the expected value of $|A(S)|$
  - $\bar{E} = \{i | i \in A(S), LB(i) \cap L_e \neq \emptyset\}$: the set of experts activated by the seed set S

The trade-off of the utility and privacy cost of online social search

- The problem formulation:
  - $U_{Le}(S) = \Sigma_{u \in \bar{E}} p_{LB(u)}$: the utility the questioner may get by choosing the set of seed nodes S.
  - $C\sigma(S)$:the privacy cost of the questioner.
  - How to make a trade-off between the utility and the privacy cost?
- Two properties of the $\sigma(\cdot)$ function ([6])
  - Submodular : $\sigma(S \cup \{v\}) - \sigma(S) \geq \sigma(T \cup \{v\}) - \sigma(T)$ for all $v \in V$ and all subsets S and T with $S \subseteq T \subseteq V$
  - Monotone: $\sigma(S) \leq \sigma(T)$ for all set $S \leq T$
  - For any function $F(\cdot)$ that is both submodular and monotone, it can be proved that the simple greedy algorithm can provide $1 - 1/e$ approximation for maximizing $F(S)$ among all sets S of size k. Besides, many algorithms can be used to solve the influence maximization problem, like Degree Discount Algorithm[6].

Algorithm to trade-off the utility and privacy cost of online social search

- Maximize the ratio between the utility and privacy cost
- $U_{Le}(S)/ C\sigma(S)$
- Not submodular

- If we only consider the utility, then the Labeled Degree Discount heuristic[7] could be used here to find seed nodes.
- Utility Privacy Cost Ratio Discount Algorithm

Utility Degree Discount Algorithm

Algorithm 1 Utility Degree Discount Algorithm

```
Initialize S = ∅
for each node v ∈ 𝒱 do
  compute its degree d_v
  dd_v = d_v
  Initialize |t_v| = 0,|s_v| = 0
  for i = 1 to k do
    Select u = argmax_{v∈𝒱\S}{dd_v}
    S = S ∪ {v}
    for each neighbor v of u and v ∈ 𝒱 \ S do
      s_v = s_v + 1
      if LB(u) = L_e then
        t_v = t_v + 1
      end if
      if LB(v) = L_e then
        dd_v = (1 − p)^{s_v}[1 + (d_v − t_v)]
      else
        dd_v = (1 − p)^{s_v}(d_v − t_v)
      end if
    end for
  end for
end for
return S
```

d_v : the number of neighbors of v who are experts

s_v : the number of neighbors of v who are seeds

t_v : the number of neighbors of v who are seeds and experts

dd_v : degree discount

Utility Privacy Cost Ratio Discount Algorithm

Algorithm 2 Utility Privacy Cost Ratio Discount Algorithm

```
Initialize S = ∅
for each node v ∈ V do
  compute its degree d_v
  dd_v = d_v / dg_v
  Initialize |t_v| = 0, |s_v| = 0
  for i = 1 to k do
    Select u = argmax_{v∈V\S}{dd_v}
    S = S ∪ {v}
    for each neighbor v of u and v ∈ V \ S do
      s_v = s_v + 1
      if LB(u) = L_e then
        t_v = t_v + 1
      end if
      if LB(v) = L_e then
        dd_v = (1 − p)^{s_v}[1 + (d_v − t_v)]/(dg_v − s_v)
      else
        dd_v = (1 − p)^{s_v}(d_v − t_v)/(dg_v − s_v)
      end if
    end for
  end for
end for
return S
```

d_v : the number of neighbors of v who are experts

dg_v : the number of neighbors of v

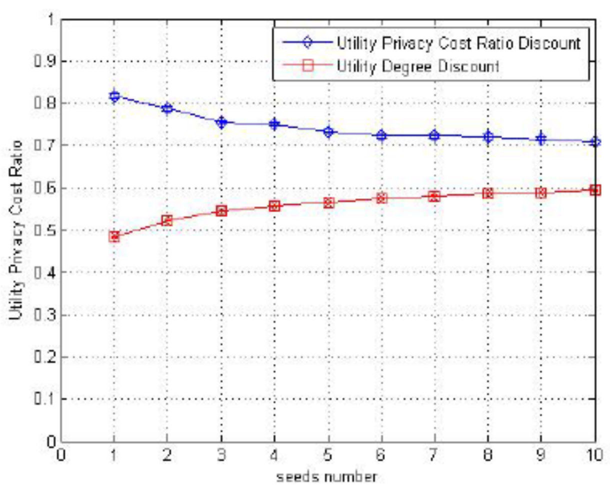
s_v : the number of neighbors of v who are seeds

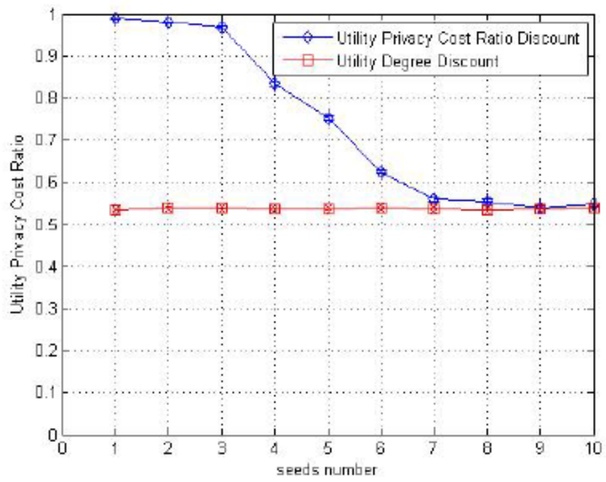
t_v : the number of neighbors of v who are seeds and experts

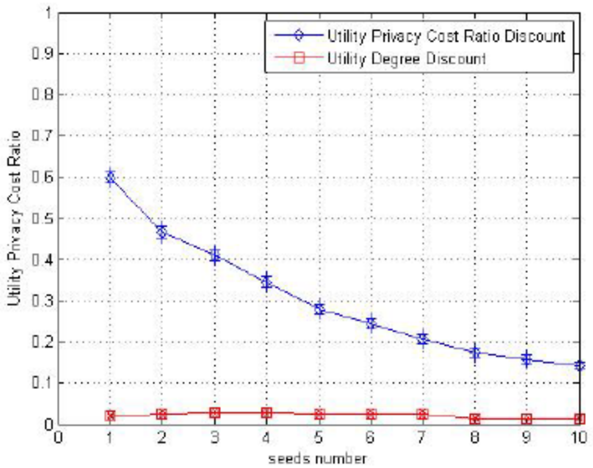
dd_v : degree discount

Evaluation

- Results
  - Utility and cost ratio of three questions, community 1(426 nodes), community 2(400 nodes), community 20(40 nodes), the probability in IC model of the community 20 is 0.05







Thank You!