Threshold Influence Model for Allocating Advertising Budgets

Nao Kakimura

University of Tokyo

Joint work with Takuro Fukunaga, Atsushi Miyauchi, Yuni Iwamasa

Shonan 2016

河原林巨大グラフプロジェクト
ERATO Kawarabayashi Large Graph Project

2016/04/12

Topic: Budget Allocation for Marketing

- Modelling information diffusion
- Simple algorithm for optimization

TV Newspaper DM Social ads

… Media sources

Q. How can we allocate budgets for effective advertisement?

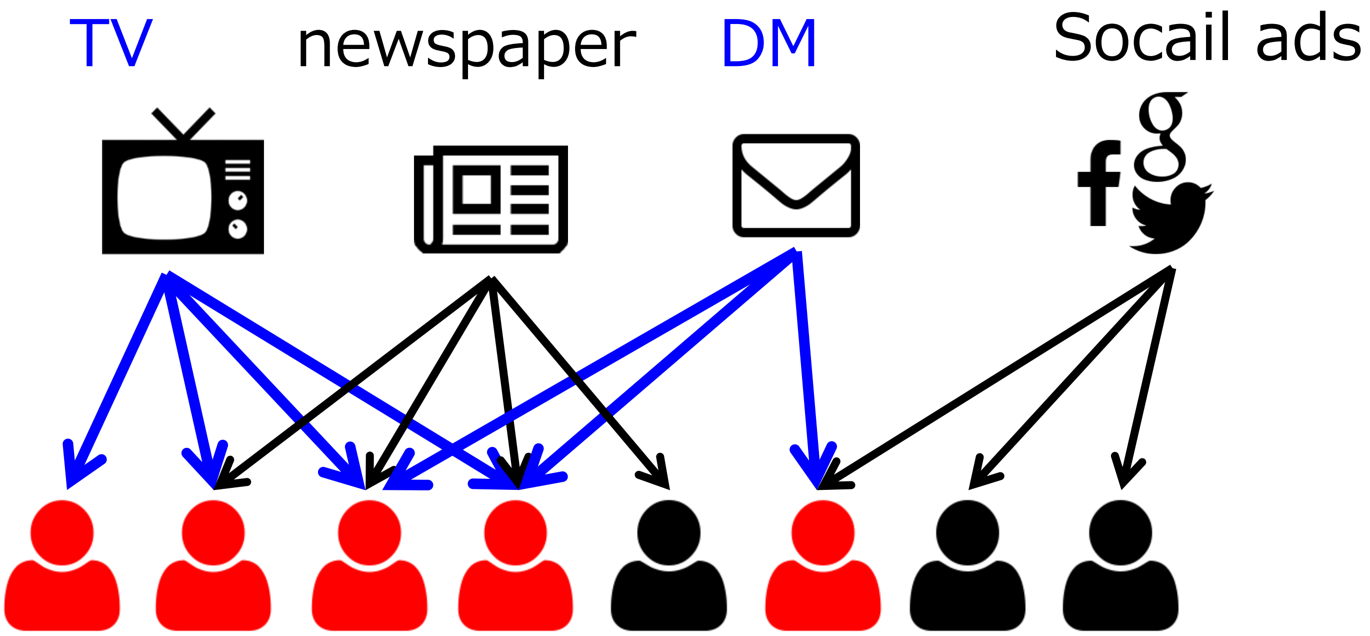
Target customers

Simple Model: Maximum Coverage

- Maximize #customers watching ads

$$f(X) = \#\{\, t \,:\, |N(t) \cap X| \geq 1 \,\}$$

$$\begin{aligned} \text{max.} \quad & f(X) \\ \text{s.t.} \quad & |X| \leq k \end{aligned}$$



TV and DM → 5 people are covered

Two Influence Models[Alon et al.12]

$$\text{max.} \quad f(X) = \sum_{t \in T} \Pr[t]$$
$$\text{s.t.} \quad |X| \leq k$$

Expected #influenced customers

- Source-side model ⊆ max. monotone submodular fns
  - Each **source** has a probability

$$\Pr[\ t \text{ is influenced }] = 1 - \prod_{s \in N(t) \cap X} (1 - p_s)$$

- Target-side model ≈ max. thresholds coverage
  - Each **target** has a probability

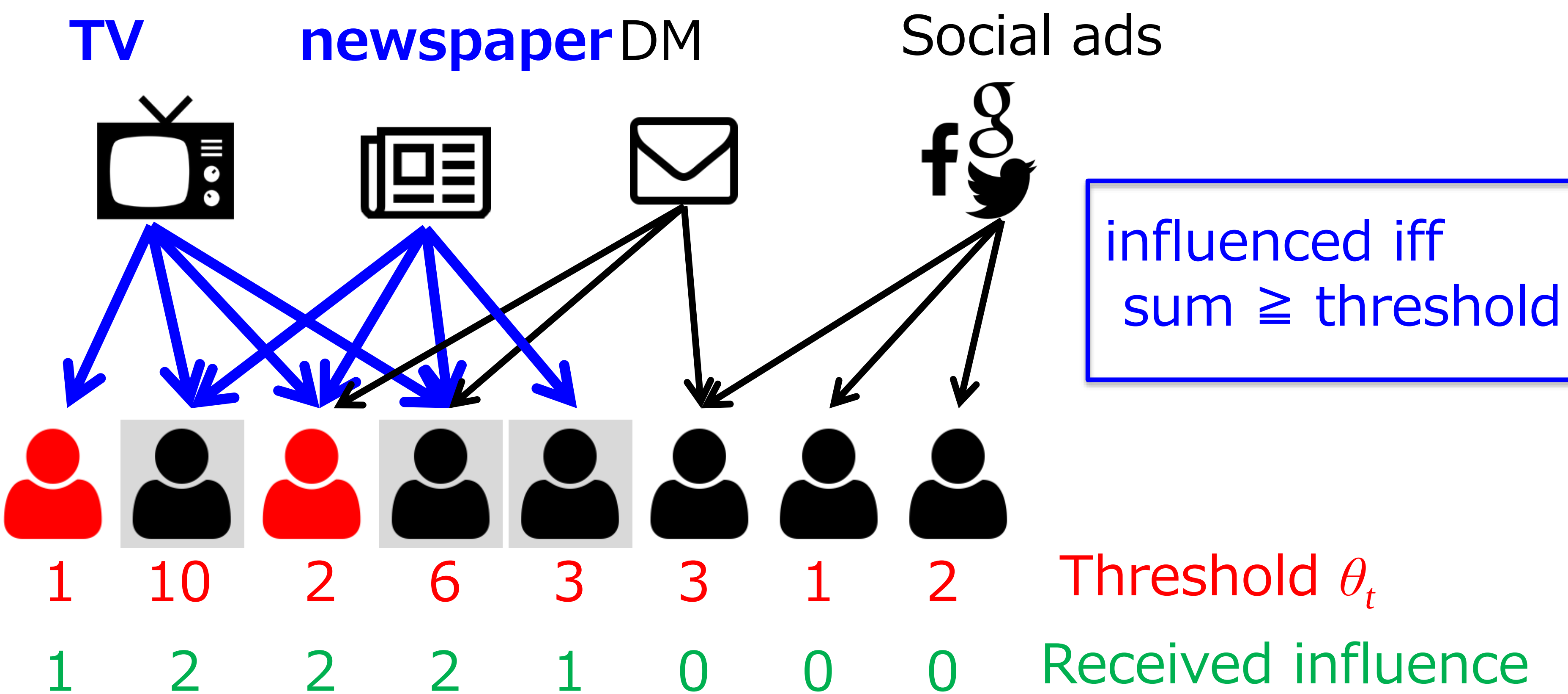
$$\Pr[\ t \text{ is influenced }] = 1 - (1 - p_t)^{|N(t) \cap X|}$$

Maximum Thresholds Coverage

- Each customer has a threshold

$$f(X) = \#\{ t : |N(t) \cap X| \geq \theta_t \}$$

$$\text{max.} \quad f(X) \qquad \text{s.t.} \quad |X| \leq k$$



Maximum Thresholds Coverage

$$\max. \quad f(X) \quad \text{s.t.} \quad |X| \leqq k$$

#customers whose received influence ≧ θ_t

- Observation
  - Each threshold is 1 = maximum coverage
    - (1-1/e)-approximation
  - Harder than the densest k-subgraph problem
    - f is not submodular or supermodular in general
    - If θ is randomly chosen, (1-1/e)-approximation

→ **Today**: Different objective on the same model

This Talk: Problem with Different Objective

- Maximizing **cost-effectiveness**
  - performance of budgets without constraint

$$\max \frac{f(X)}{|X|}$$

customers influenced

the cost spent

- NP-hard

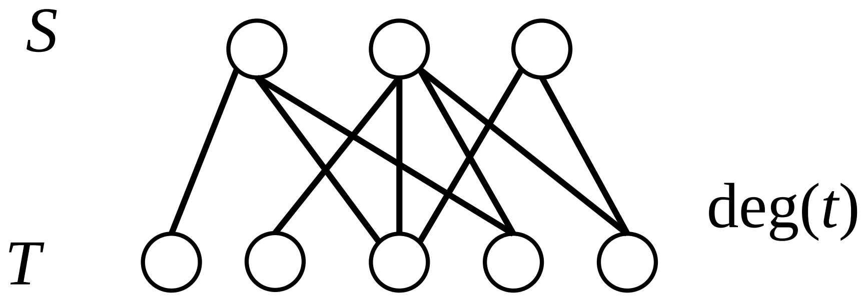
Result: Cost-Effectiveness Approximation

1) Decremental greedy algorithm

- Repeatedly remove one source w./ min. contribution
- Approx. factor = $\max \deg(t)$
- Works for a more general case
  - when influence is submodular
  - when sources/targets have weights

2) LP-based algorithm

- Approx. factor = $\max\,(\deg(t) - \theta_t + 1)$



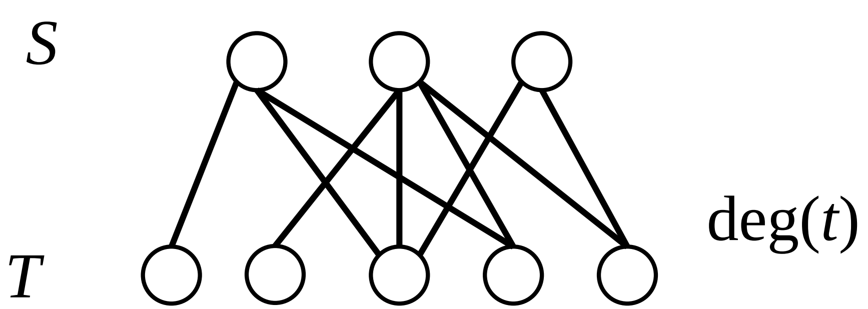
Result: Cost-Effectiveness Approximation

1) Decremental greedy algorithm

- Repeatedly remove one source w./ min. contribution
- Approx. factor = $\max \deg(t)$
- Works for a more general case
  - when influence is submodular
  - when sources/targets have weights

2) LP-based algorithm

- Approx. factor = $\max\,(\deg(t) - \theta_t + 1)$



LP for Threshold Model

- Generalizing LP for densest subgraph [Charikar00]

  ($\theta_t = \deg(t) = 2$ for any t)

  - Each source s : x_s
  - Each target t : y_t

$$\max \sum_{t \in T} y_t$$

sub to

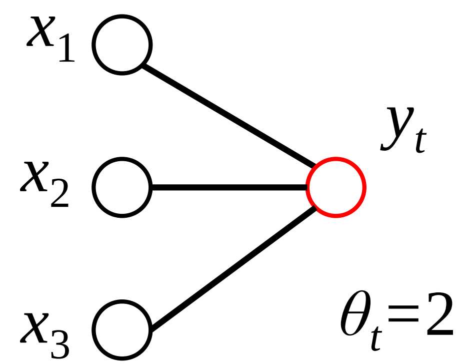
Covering constraints

At least θ_t of neighboring x_s's are 1
$\Leftrightarrow y_t = 1$

It is not good to have only $\sum_{s:\text{neighbor of } t} x_s \geq \theta_t y_t$

Constraint choosing 2 out of 3

- At least 2 of x_s's are 1 ⇔ y_t =1
  - ⇔ Sum of the 2 smallest ones is ≧1



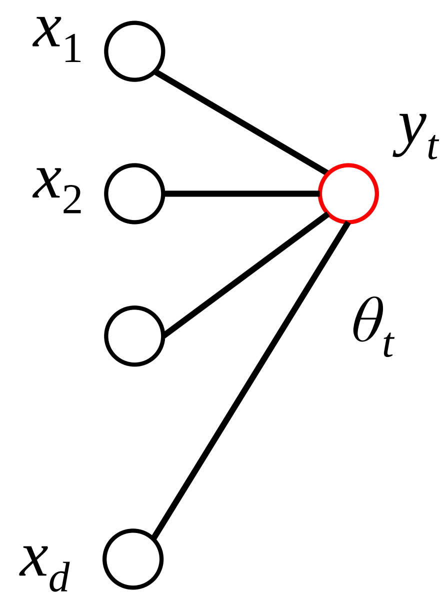
$$(1 \quad 1 \quad 1) \quad (1 \quad 1 \quad 0) \quad (1 \quad 0 \quad 0) \quad (0 \quad 0 \quad 0)$$

⇕

$$x_1 + x_2 \quad 3 \; y_t$$
$$x_2 + x_3 \; 3 \; y_t$$
$$x_1 + \quad x_3 \; 3 \; y_t$$

Constraint choosing θ_t out of d=deg(t)

- At least θ_t of x_s's are 1 ⇔ y_t =1
  - ⇔ Sum of the p_t smallest ones is ≧1



$$p_t = \deg(t) - \theta_t + 1$$

$$\begin{pmatrix} \underbrace{1 & \cdots & 1}_{\theta_t} & 0 & \cdots & 0 \end{pmatrix}$$

⇕

$$\begin{cases} x_1 + x_2 + \cdots + x_{p_t} \geq y_t \\ x_2 + x_3 + \cdots + x_{p_t+1} \geq y_t \\ \vdots \\ \text{for any choice of } p_t\text{-size subsets} \end{cases}$$

LP for Threshold Model

$$\max \sum_{t \in T} y_t$$

Exponential #consts

$$\sum_{s \in U} x_s \geq y_t \quad \forall t \in T, \quad \forall U \subseteq N(t), |U| = p_t$$

$$\sum_{s \in S} x_s = 1 \quad \leftarrow \text{Normalization}$$

- Rem. This LP can be solved in poly-time
  - target t is valid if the sum of minimum p_t values $\geqq y_t$

LP for Threshold Model

Exponential #consts

$$\max \sum_{t \in T} y_t$$

$$\sum_{s \in U} x_s \ge y_t \qquad \forall t \in T \quad \forall U \subseteq N(t), |U| = p_t$$

$$\sum_{s \in S} x_s = 1 \qquad \leftarrow \text{Normalization}$$

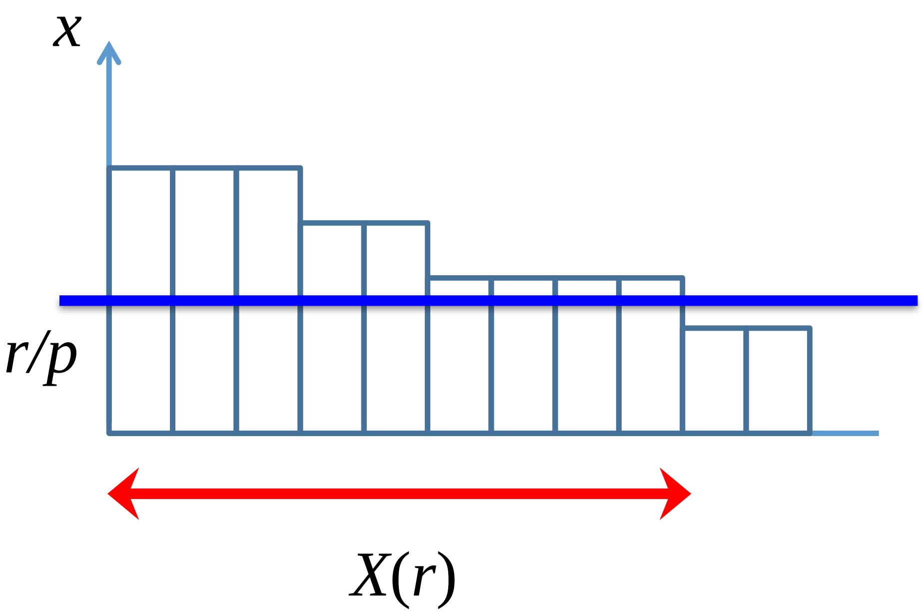
[Thm]

LP-optimal value ≧ optimal cost-effectiveness

Rounding Algorithm

$$p = \max_{t \in T} p_t$$

- Find an LP-optimal solution $\boldsymbol{x}, \boldsymbol{y}$
- For a number r in $[0,1]$, define $X(r) = \{ s \mid x_s \geqq r/p \}$
- Return the best $X(r)$ by changing r



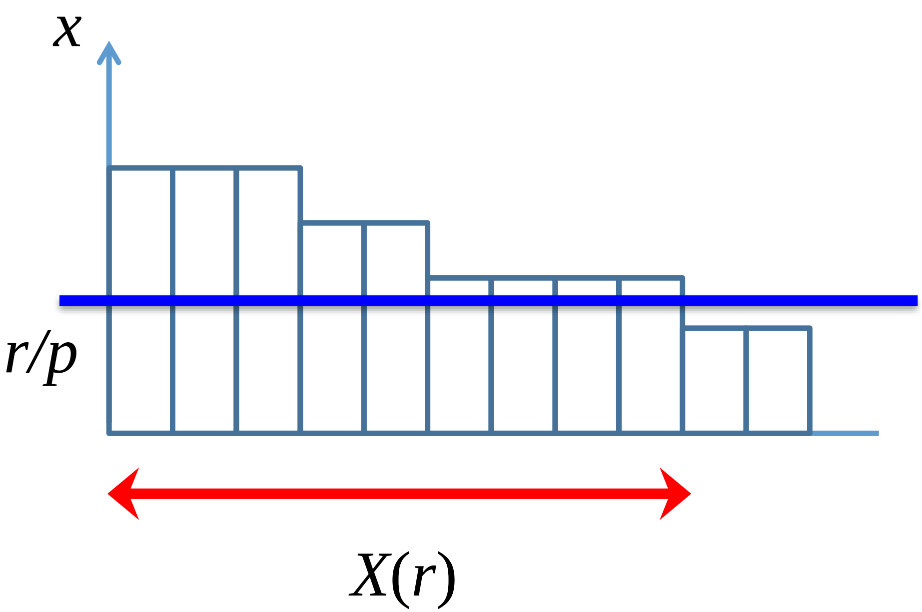
[Thm]

Rounded solution $X(r)$ is p-approximation

Rounding Algorithm

$p = \max_{t \in T} p_t$

- Find an LP-optimal solution $\boldsymbol{x}, \boldsymbol{y}$
- For a number r in $[0,1]$, define $X(r) = \{ s \mid x_s \geqq r/p \}$
- Return the best $X(r)$ by changing r



Covering at least

$Y(r) = \{ t \mid y_t \geqq r \}$

Sum of the p_t smallest ones ≧ r

⇔ The θ_tth smallest ones ≧ r/p

⇔ θ_t of x_s's are in $X(r)$

[lemma]

$$\exists r \qquad \frac{|Y(r)|}{|X(r)|} \geq \frac{LPOPT}{p}$$

Proof Idea of Lemma

$$\max \sum_{t \in T} y_t$$

$$\sum_{s \in U} x_s \geq y_t$$

$$\sum_{s \in S} x_s = 1$$

- If every r satisfies $|Y(r)| < \frac{LP}{p} |X(r)|$

$$\int_0^1 |Y(r)|\, dr < \frac{LP}{p} \int_0^1 |X(r)|\, dr$$

[lemma]

$$\exists r \quad \frac{|Y(r)|}{|X(r)|} \geq \frac{LPOPT}{p}$$

Proof Idea of Lemma

$$\max \sum_{t \in T} y_t$$

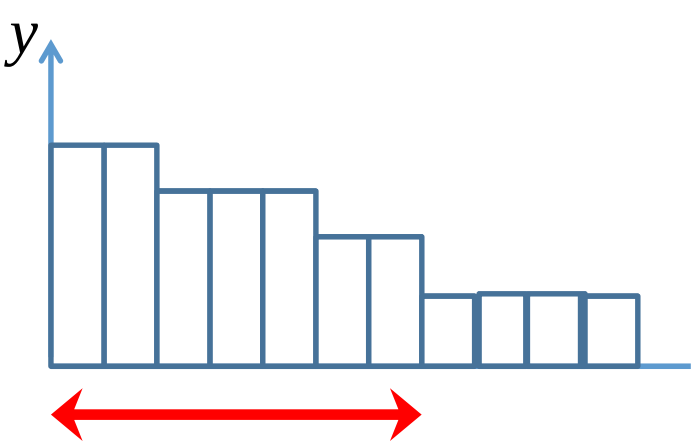
$$\sum_{s \in U} x_s \geq y_t$$

$$\sum_{s \in S} x_s = 1$$

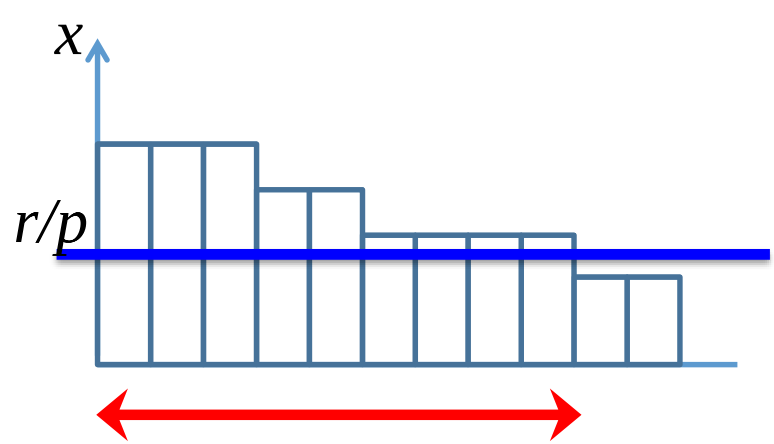
- If every r satisfies $|Y(r)| < \frac{LP}{p} |X(r)|$

$$\int_0^1 |Y(r)|\,dr < \frac{LP}{p} \int_0^1 |X(r)|\,dr$$

$$\sum_t y_t = LP \qquad \leq \sum_s p x_s = p \qquad \to \text{contradiction}$$



$Y(r) = \{\, t \mid y_t \geqq r \,\}$



$X(r)$

Summary

$$\max \frac{f(X)}{|X|}$$

- $\max(\deg(t) - \theta_t + 1)$-approximation
  - Q. Better approximation?
    - P when $\theta_t = 1$ for any t (maximum coverage)
      - when f is submodular, opt is a singleton
    - P when $\theta_t = \deg(t)$ for any t (densest subhypergraph)
      - f is supermodular

Q. Fractional optimization for other problems?

- Densest k-subgraph vs Densest subgraph

Q. Approx. for size-constrained problem?

- LP for densest k-subhypergraph gives deg(s)-approximation [Arulselvan14]

$$\begin{aligned} \text{max.} \quad & f(X) \\ \text{s.t.} \quad & |X| \leqq k \end{aligned}$$

Thank you for your attention