Whom to Ask?

Jury Selection for Decision Making Tasks on Micro-blog Services

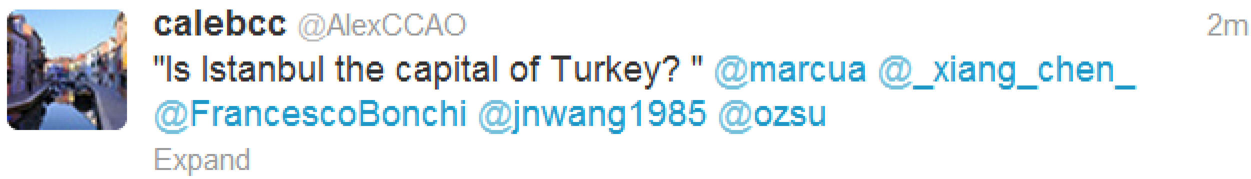
Caleb Chen CAO, Jieying SHE, Yongxin TONG, Lei CHEN

The Hong Kong University of Science and Technology

“Is Istanbul the capital of Turkey?”

Social Network/Media services

the **virtualization** and **digitalization** of people’s social activities



- Minor as dressing for a banquet
- Major as prediction of macro economy trends

“two-option decision making tasks”

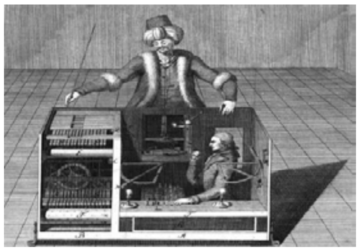
Wisdom of Crowd

"The basic argument there, drawing on a long history of intuition about markets, is that the **aggregate behavior of many people**, each with limited information, **can produce very accurate beliefs**." –D. Easley, J. Kleinberg, "Networks, Crowds, and Markets"

Crowdsourcing-powered DB Systems

- Qurk, "*Human powered Sorts and Joins*", VLDB'2012(MIT)
- Deco, *"A System for Declarative Crowdsourcing"*, VLDB'2012(Stanford)
- CrowdDB, "*Answering Queries with Crowdsourcing*", SIGMOD'2011(Berkeley)

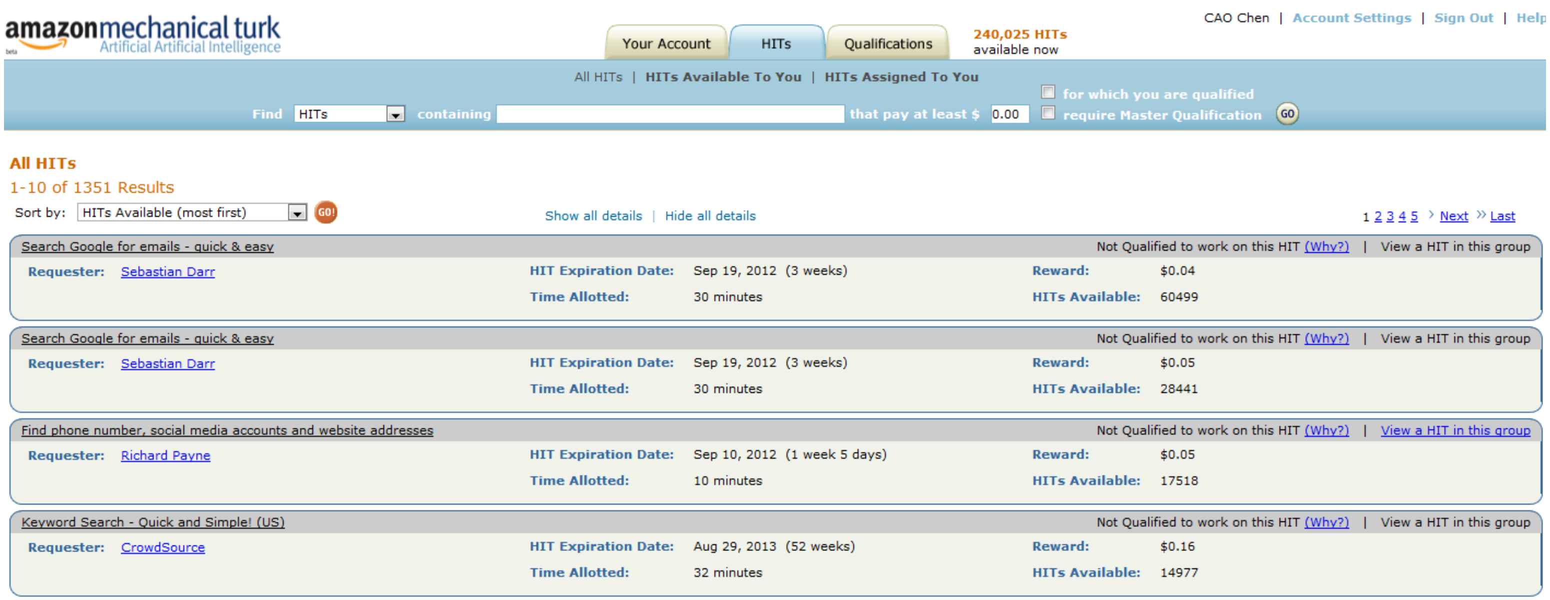
General Crowdsourcing Platforms



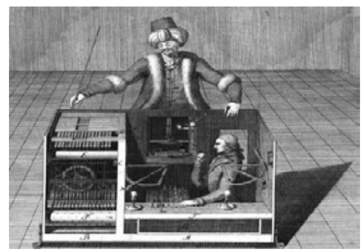
AMT

oDesk

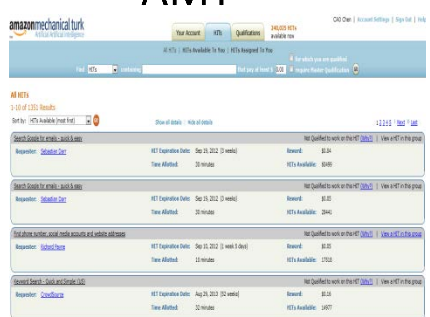
CrowdFlower



Requesters



AMT



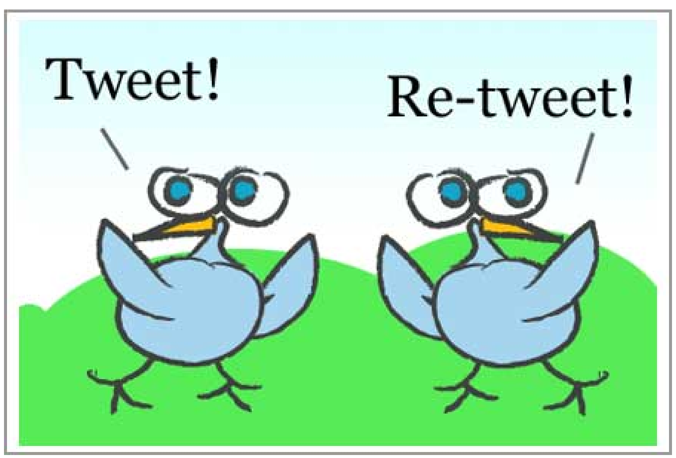
MTurk workers
(Photo By Andrian Chen)

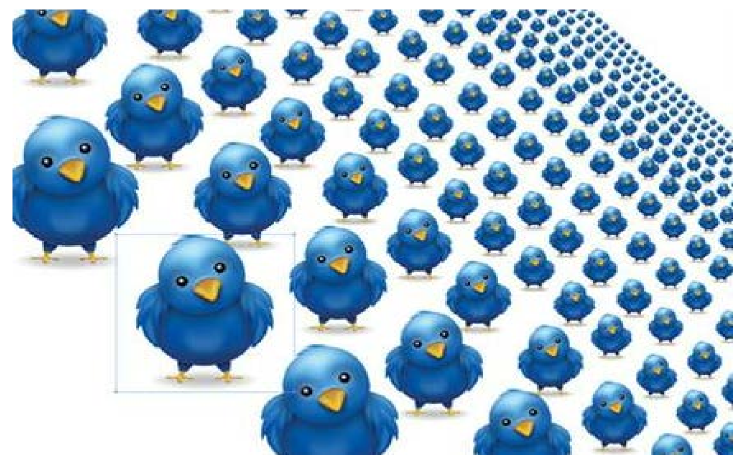
Can we extend the magic power of **Crowdsourcing** onto **social network**?

Microblog Users

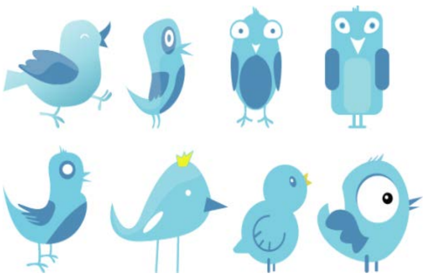
- Simple
  - 140 characters
  - ‘RT’ + ‘@’

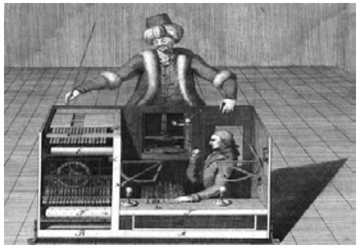
- But comprehensive
  - Large network
  - Various backgrounds of users





Why Microblog Platform?


Twitter

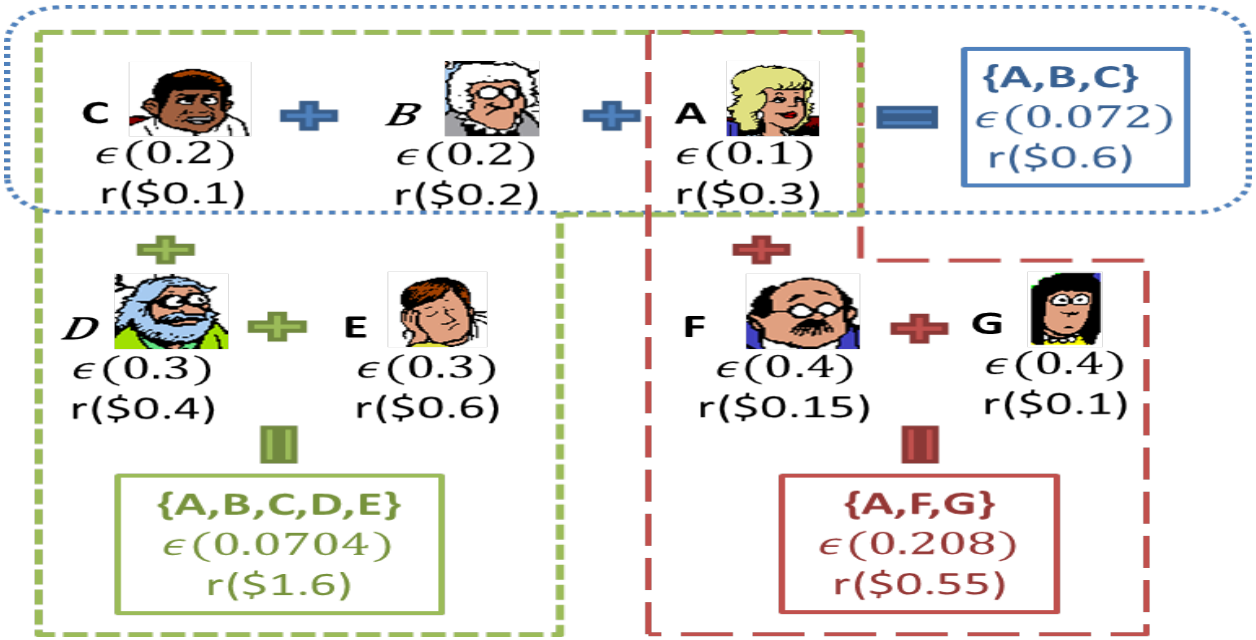

AMT

	Social Media Network	General Purpose Platform
Accessibility	Highly convenient, on all kinds of mobile devices	Specific online platform
Incentive	Altruistic or payment	Mostly monetary incentive
Supported tasks	Simple task as decision making	Various types of tasks
Communication Infrastructure	'Tweet' and 'Reply' are enough	Complex workflow control mechanism
Worker Selection	**Active**, Enabled by '@'	Passively, No exact selection

Outline

- **Running Example**
- Problem Definition
- Jury Selection Algorithms
- Evaluation

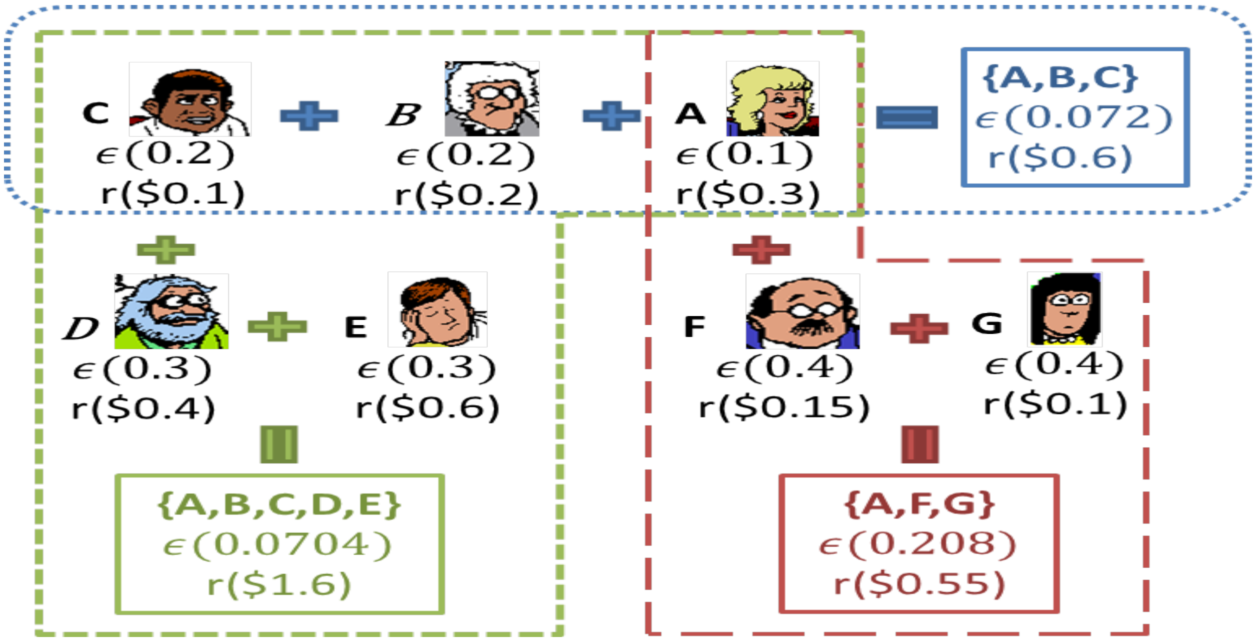
Motivation – Jury Selection Problem Running Case(1)



Is Döner Kebab available in Hong Kong?

- Given a decision making problem, with budget $1, **whom should we ask**?

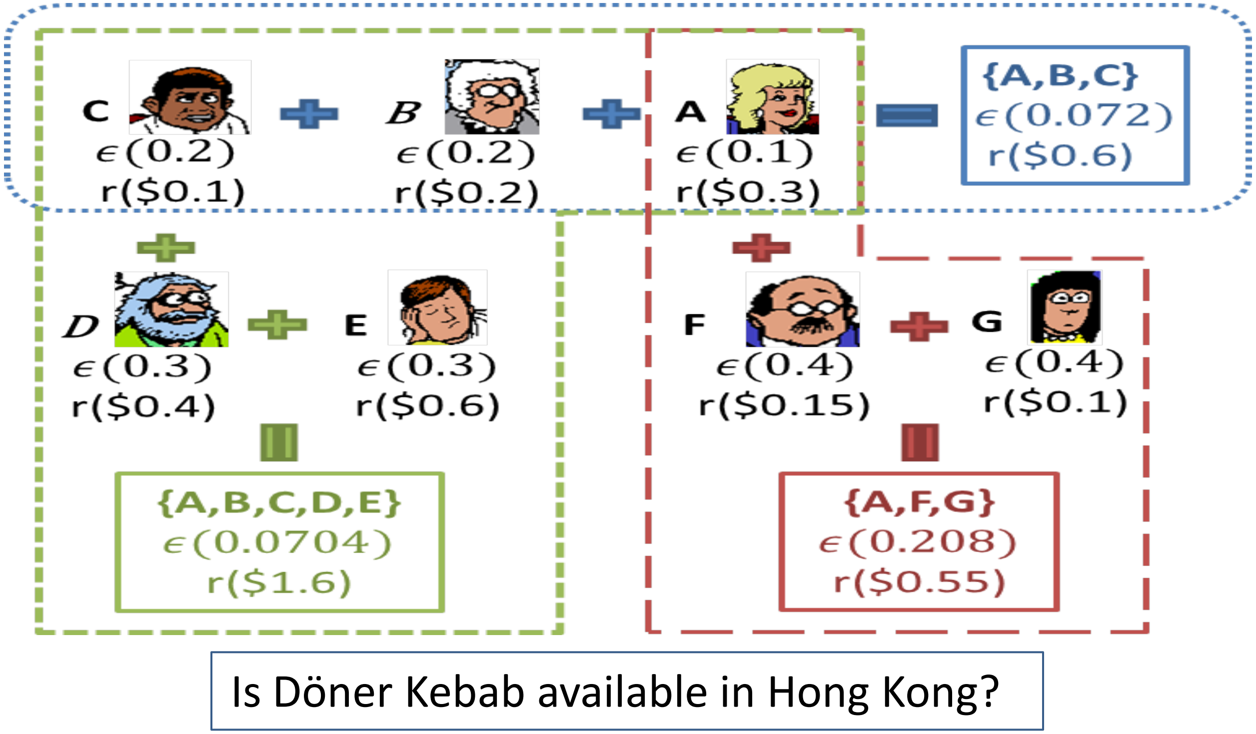
Motivation – Jury Selection Problem Running Case(2)



Is Döner Kebab available in Hong Kong?

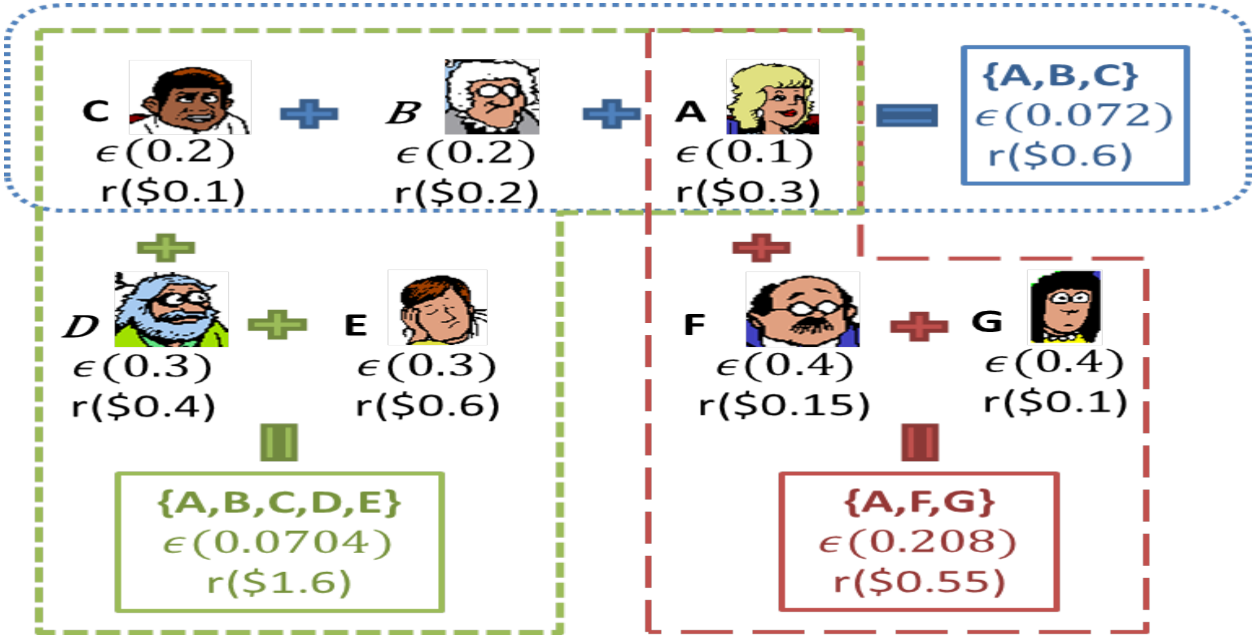
- ϵ: error rate of an individual
- r: requirement of an individual, can be virtual
- Majority Voting to achieve final answer

Motivation – Jury Selection Problem Running Case(2)



- Worker : Juror
- Crowds : Jury
- Data Quality : Jury Error Rate

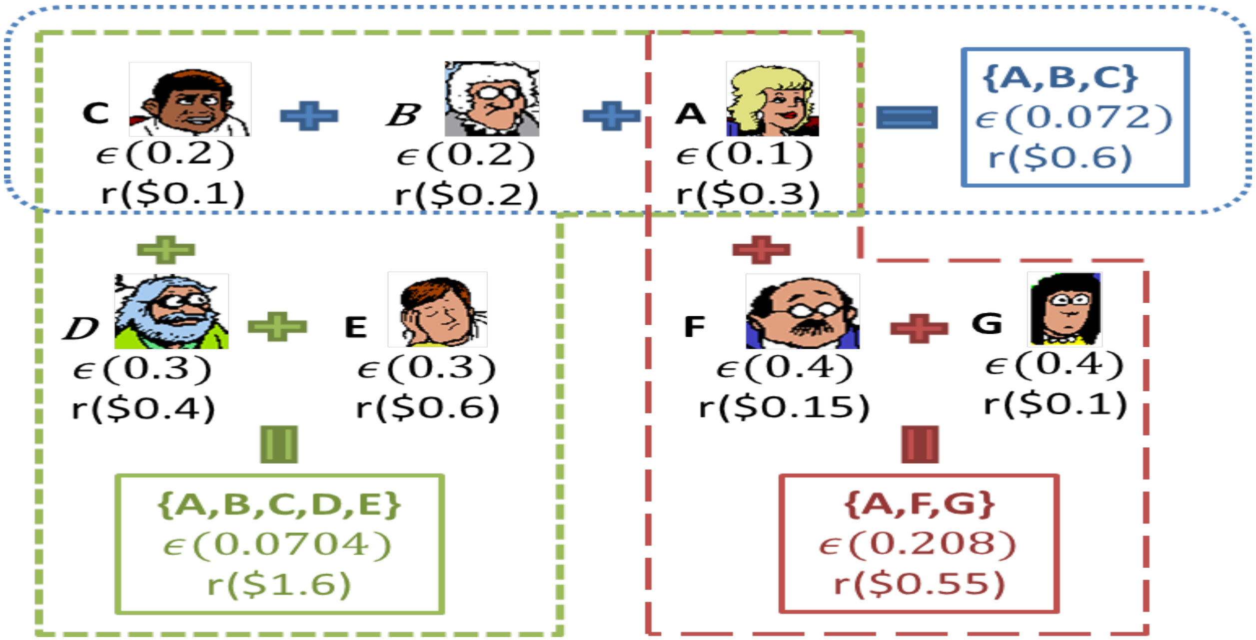
Motivation – Jury Selection Problem Running Case(3)



Is Döner Kebab available in Hong Kong?

- If (A, B, C) are chosen(Majority Voting)
  - JER(A,B,C) = 0.1*0.2*0.2 + (1 - 0.1)*0.2*0.2 + 0.1* (1 - 0.2)*0.2 + 0.1*0.2*(1 - 0.2) = 0.072
  - Better than A(0.1), B(0.2) or C(0.2) individually

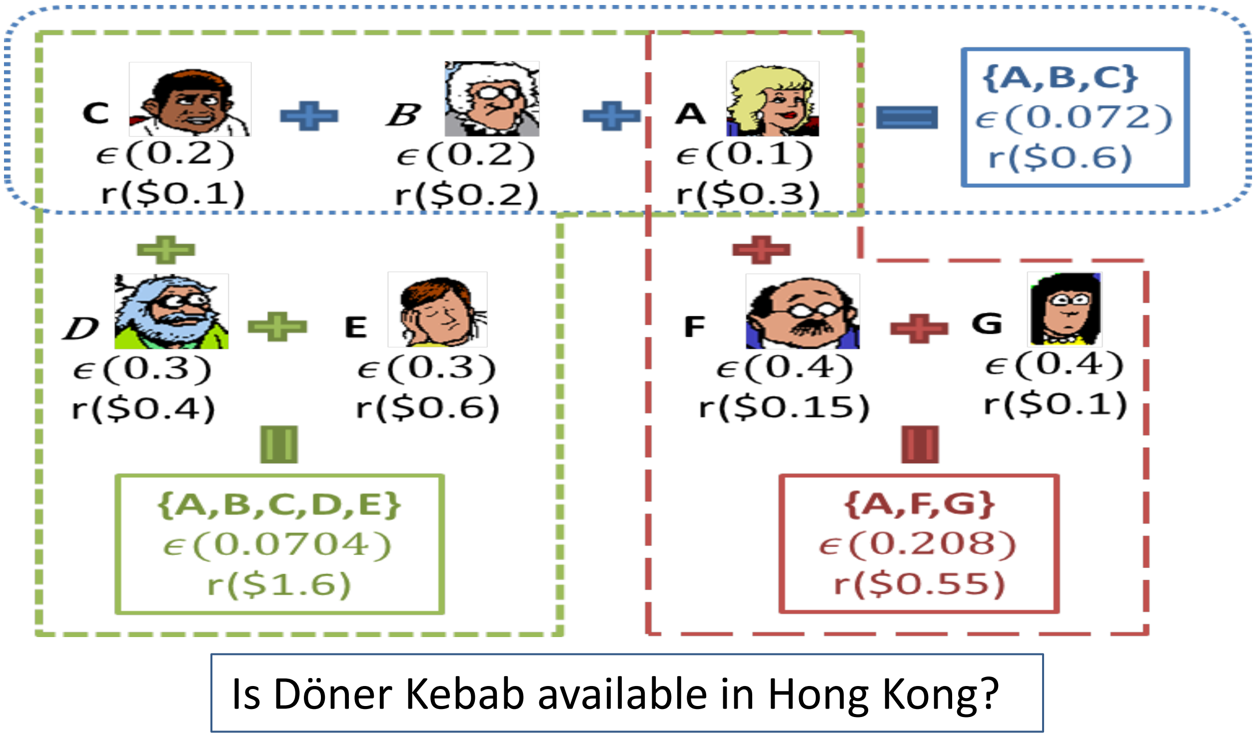
Motivation – Jury Selection Problem Running Case(4)



Is Döner Kebab available in Hong Kong?

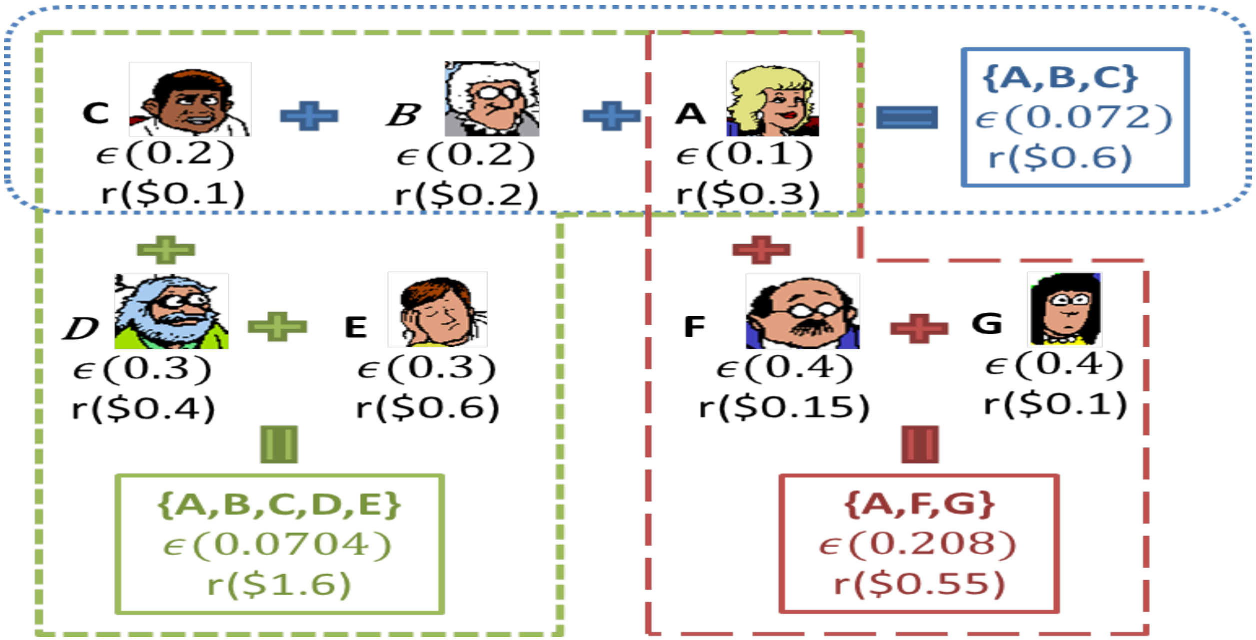
- What if we enroll more
  - JER(A,B,C,D,E) = 0.0704 < JER(A,B,C)
  - The more the better?

Motivation – Jury Selection Problem Running Case(5)



- What if we enroll even more?
  - JER(A,B,C,D,E,F,G) = 0.0805 > JER(A,B,C,D,E)
  - Hard to calculate JER

Motivation – Jury Selection Problem Running Case(6)



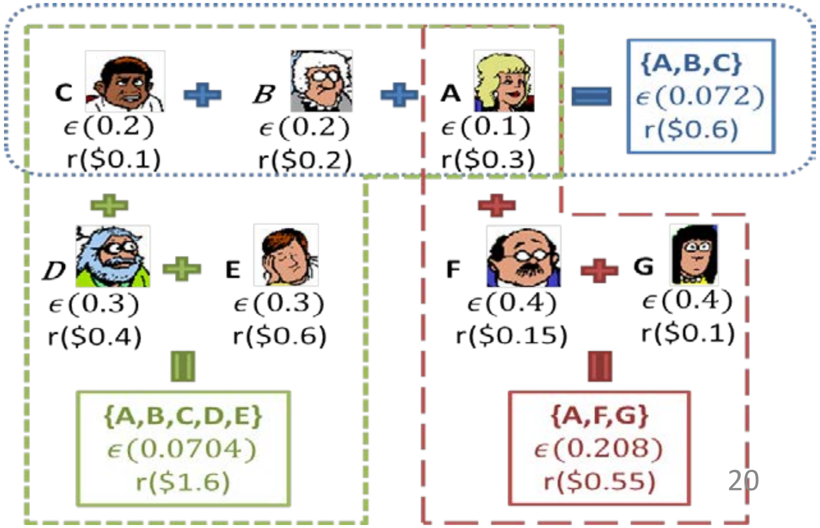
Is Döner Kebab available in Hong Kong?

- So just pick up the best combination?
  - JER(A,B,C,D,E)=0.0704
  - R(A,B,C,D,E) = \$1.6 > budget(\$1.0)

Motivation – Jury Selection Problem Running Case(7)

Crowd	Individual Error-rate	Jury Error-rate
C	0.2	0.2
A	0.1	0.1
C,D,E	0.2,0.2,0.3	0.174
A,B,C	0.1,0.2,0.2	0.072
A,B,C,D,E	0.1,0.2,0.2,0.3,0.3	0.0703
A,B,C,D,E,F,G	0.1,0.2,0.2,0.3,0.3,0.4,0.4	0.0805

Worker selection for maximize the quality of a particular type of product: **the reliability of voting.**



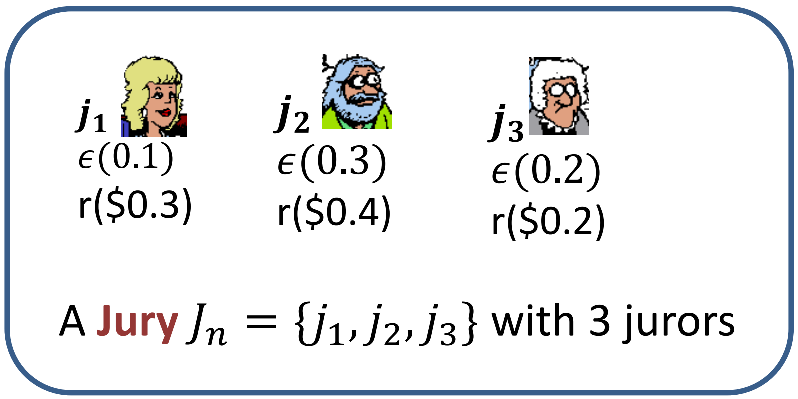
Outline

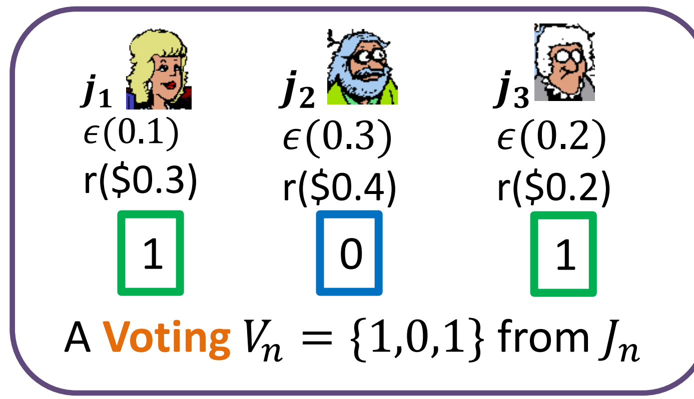
- Motivation
- **Problem Definition**
- Jury Selection Algorithms
- Evaluation

Problem Definition

- Jury and Voting

DEFINITION 1 (JURY). *A jury $J_n = \{j_1, j_2, \cdots, j_n\} \subseteq S$ is a set of jurors with size n that can form a voting.*





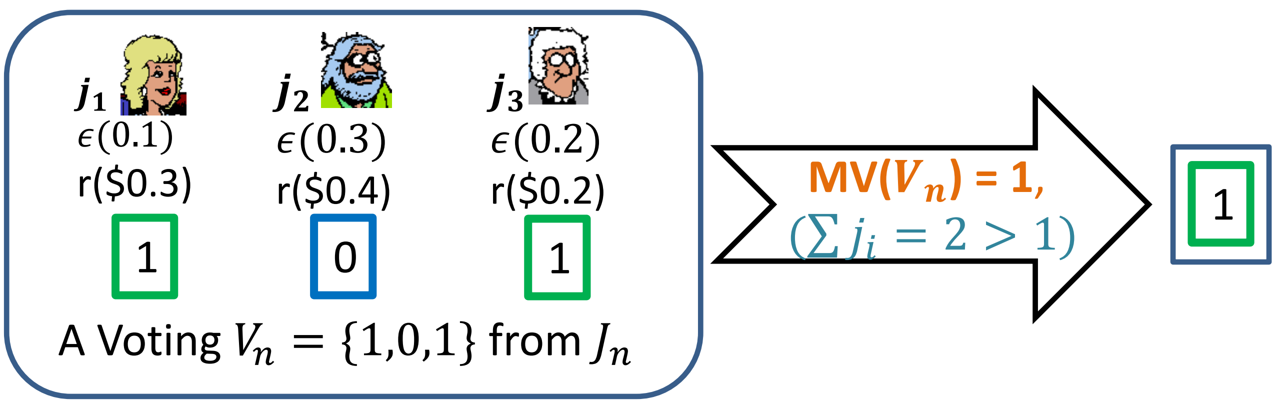
DEFINITION 2 (VOTING). *A voting V_n is a valid instance of a jury J_n with size n, which is a set of binary values.*

Problem Definition

- Voting Scheme

DEFINITION 3 (MAJORITY VOTING - MV). *Given a voting V_n with size n, Majority Voting is defined as*

$$MV(V_n) = \begin{cases} 1 & \text{if } \sum j_i \geq \frac{n+1}{2} \\ 0 & \text{if } \sum j_i \leq \frac{n-1}{2} \end{cases}$$

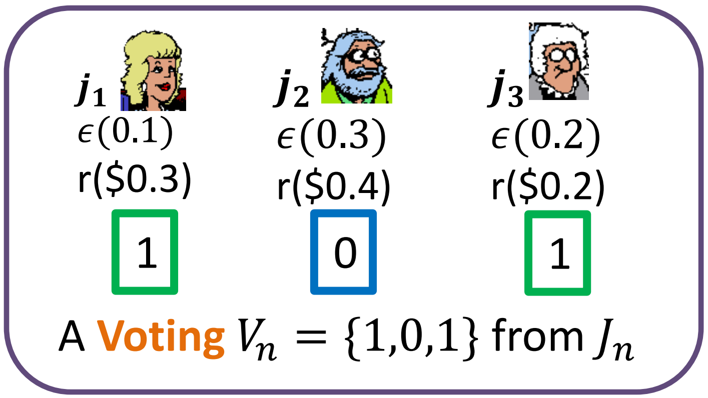


Problem Definition

- Invididual Error-rate

DEFINITION 4 (INDIVIDUAL ERROR RATE - ϵ_i). *The individual error rate ϵ_i is the probability that a juror conducts a wrong voting. Specifically*

$$\epsilon_i = Pr(vote\ otherwise | a\ task\ with\ ground\ truth\ A)$$



DEFINITION 5 (CARELESSNESS - C). *The Carelessness C is defined as the number of mistaken jurors in a jury J_n during a voting, where $0 \leq C \leq n$.*

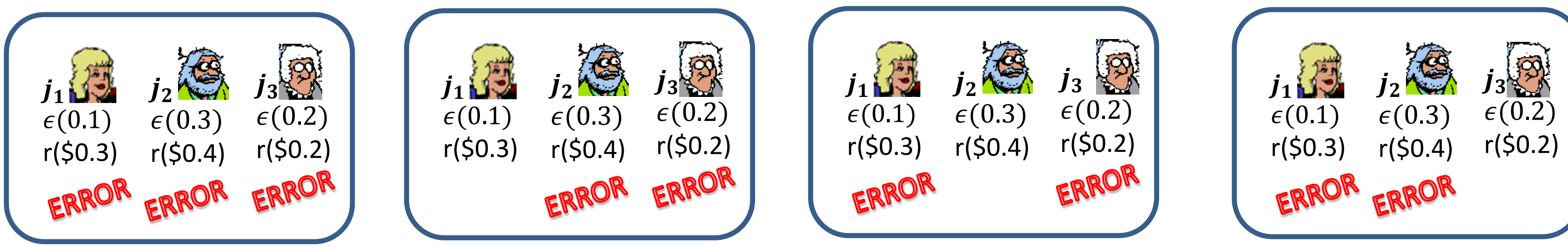
Problem Definition

Definition 6 (Jury Error Rate - $JER(J_n)$). *The jury error rate is the probability that the Carelessness C is greater than $\frac{n+1}{2}$ for a jury J_n, namely*

$$JER(J_n) = \sum_{k=\frac{n+1}{2}}^{n} \sum_{A \in F_k} \prod_{i \in A} \epsilon_i \prod_{j \in A^c} (1-\epsilon_j)$$

$$= \Pr(C \geq \frac{n+1}{2} | J_n)$$

where F_k is all the subsets of S with size k and ϵ_i is the individual error rate of juror j_i.



$JER(J_3)$ = 0.1*0.3*0.2 + (1-0.1)*0.3*0.2 + 0.1*(1-0.3)*0.2 + 0.1*0.3*(1-0.2)
= 0.029

Problem Definition

- Crowdsourcing Models(model of candidate microblog users)

DEFINITION 7 (ALTRUISM JURORS MODEL - ALTRM). *While selecting a jury J from all candidate jurors (choosing a subset $J \subseteq S$), any possible jury is allowed.*

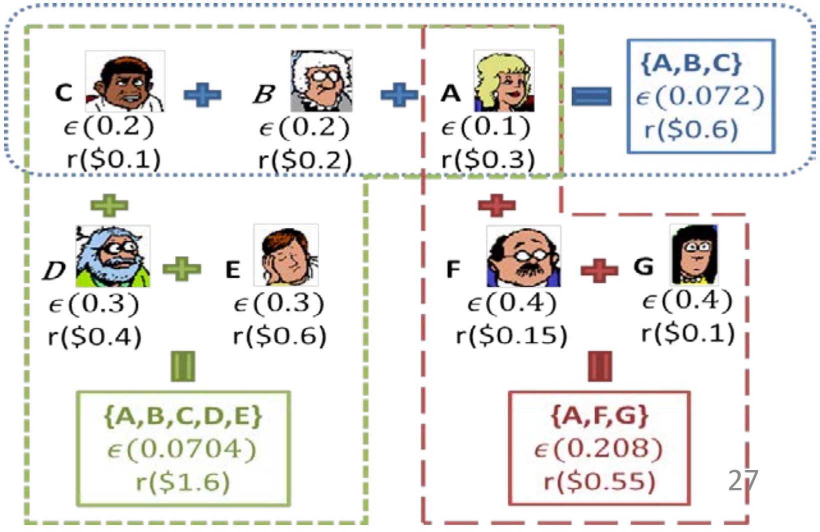
DEFINITION 8 (PAY-AS-YOU-GO MODEL - PAYM). *While selecting a jury J from all candidate jurors (choosing a subset $J \subseteq S$), each candidate juror j_i is associated with a payment requirement r_i where $r_i \geq 0$, the possible jury J is allowed when the total payment of J is no more than a given budget B, namely $\sum_{\forall j_i \in J} r_i \leq B$.*

Problem Definition

- Jury Selection Problem(JSP)

DEFINITION 9 (JURY SELECTION PROBLEM - JSP). *Given a candidate juror set S with size $|S| = N$, a budget $B \geq 0$, a crowdsourcing model(AltrM or PayM), the Jury Selection Problem(JSP) is to select a jury $J_n \subseteq S$ with size $1 \leq n \leq N$, that J_n is allowed according to crowdsourcing model and $JER(J_n)$ is minimized.*

We hope to form a Jury J_n, allowed by the budget, and with lowest JER



Outline

- Motivation
- Problem Definition
- **Jury Selection Algorithms**
- Evaluation

Computation of Jury Error Rate

- The number of careless jurors(*Carelessness*-*C*) is a random variable following Poisson Binomial Distribution

$$JER(J_n) = \sum_{k=\frac{n+1}{2}}^{n} \sum_{A \in F_k} \prod_{i \in A} \epsilon_i \prod_{j \in A^c} (1 - \epsilon_j)$$
$$= \Pr(C \geq \frac{n+1}{2} | J_n)$$

- The naïve computation of JER is exponentially increasing

Computation of Jury Error Rate(2)

- Alg1: Dynamic Programming to compute JER in $O(n^2)$

LEMMA 1. *The calculation of JER of Jury with size n can be split into smaller ones:*

$$
\begin{aligned}
&\Pr(C \geq L | J_n) \\
=&\Pr(C \geq L-1 | J_{n-1}) \cdot \epsilon_n + \Pr(C \geq L | J_{n-1}) \cdot (1-\epsilon_n)
\end{aligned}
$$

where

$$
\begin{aligned}
\Pr(C \geq 0 | J_m) = 1 \qquad &\forall \qquad 0 \leq m \leq n \\
\Pr(C \geq m | J_n) = 0 \qquad &\forall \qquad m > n \qquad \square
\end{aligned}
$$

Computation of Jury Error Rate(3)

- Alg2: Convolution-based to compute JER in $O(nlog^2n)$
  - Treat probability distribution as coefficients of polynomials
  - Divide larger jury in two smaller juries
  - Merge by polynomial multiplication
    - Can be speeded up by using FFT

Computation of Jury Error Rate(4)

- Alg2: Convolution-based to compute JER in $O(nlog^2n)$

Algorithm 1 Convolution-based Algorithm(CBA)

Input: A jury J_n
Output: the vector of distribution of C, D_C

if $n = 1$ **then**
    $D_C[0] = 1 - \epsilon_1$;
    $D_C[1] = \epsilon_1$;
    return D_C;
else
    Dividing J_n into two parts: J_{n1} and J_{n2}, where $|J_{n1}| = \lfloor \frac{n}{2} \rfloor$ and $|J_{n2}| = \lceil \frac{n}{2} \rceil$;
    $D_{C1} = CBA(J_{n1})$;
    $D_{C2} = CBA(J_{n2})$;
    D_C =convolution of D_{C1} and D_{C2};
end if
return D_C;

Divide into two smaller juries

Merge, using FFT to speed up convolution

Computation of Jury Error Rate(5)

- Alg3: lower bound of JER in $O(n)$ time
  - Paley-Zygmund inequality

LEMMA 3 (LOWER BOUND-BASED PRUNING). *Given a jury with size n, the lower bound of $JER(J_n)$ is shown as follows,*

$$JER(J_n) \geq \frac{(1-\gamma)^2\mu^2}{(1-\gamma)^2\mu^2 + \sigma^2}$$

where $\mu = \sum_{i=1}^{n} \epsilon_i, \sigma^2 = \sum_{i=1}^{n}(1-\epsilon_i)\epsilon_i$, *and* $\gamma = (\frac{n+1}{2}/\mu) \in (0, 1)$.

JSP on AltrM(1)

- Monotonicity with given jury size on varying individual error-rate

LEMMA 4. *The lowest JER originates from the Jurors with lowest individual error-rate among the candidate jurors set S.*

PROOF. W.l.o.g, we pick one j_i of the n jurors in a given Jury J_n with size n. Then $JER(J_n)$ can be transformed as below:

$$
\begin{aligned}
&JER(J_n) = \Pr(C \geq \frac{n+1}{2}|J_n) \\
=&\epsilon_i(\Pr(C \geq \frac{n+1}{2} - 1|J_{n-1})) + (1-\epsilon_i)(\Pr(C \geq \frac{n+1}{2}|J_{n-1})) \\
=&\epsilon_i(\Pr(C = \frac{n+1}{2} - 1|J_{n-1}) + (\Pr(C \geq \frac{n+1}{2}|J_{n-1})) \\
=&\epsilon_i \cdot A + B
\end{aligned}
$$

- In English: “best jury comes from best jurors”
- Decide the size

JSP on AltrM(2)

- Algorithm for JSP on AltrM

Alg_AltrM{

1. Sort according to error-rate;
2. Starting from 1 to n, increase the jury size by two; //keep the size odd
    1. Compute JER;
    2. Update best current jury;
3. Output best jury;

}

Might be **convex**, future work

JSP on PayM(1)

- Budget is a constraint
- Objective function is JER
- NP-hardness
  - Reduce to an *nth-order 0-1 Knapsack Problem*

$$optimize \quad \underbrace{\sum_{i_1 \in n} \sum_{i_2 \in n} \cdots \sum_{i_n \in n}}_{n} V[i_1, i_2, \ldots, i_n] \cdot x_1 x_2 \ldots x_n$$

Given an instance of traditional KP, we can construct an $nOKP$ instance by defining the profit n-dimensional vector as $V[i, i, \ldots, i] = p_i$ and $V[otherwise] = 0$ for all i, where p_i is the profit in traditional KP. The weight vector and objective value remain the same. $\square$

JSP on PayM(2)

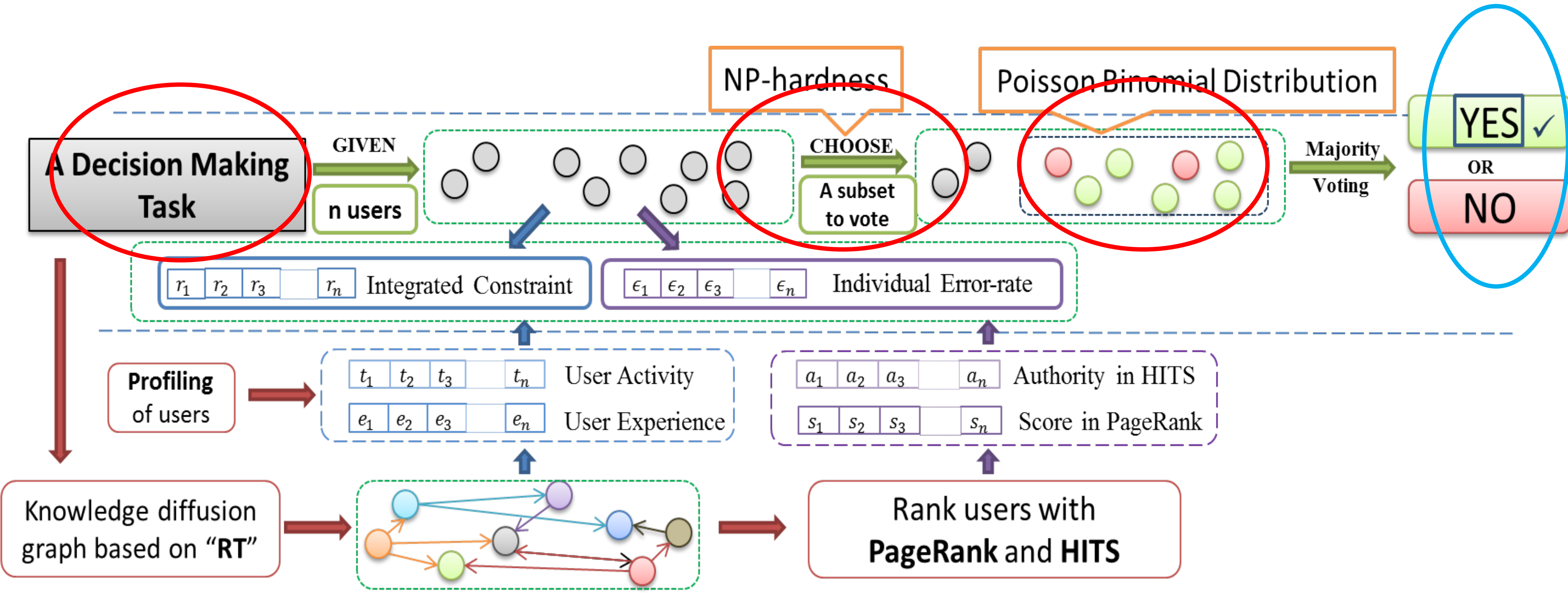
- Approximate Algorithm

Alg_PayM{

1. Sort according to *(requirement * error-rate);*
2. Starting from 1 to n, increase the jury size by two;
    1. Keep track of pair; //increment might be conducted by size of 1
    2. Check whether adding new juror will exceed budget;
    3. If so, compute and compare JER;
    4. Update best current jury;
3. Output best jury;

}

Framework



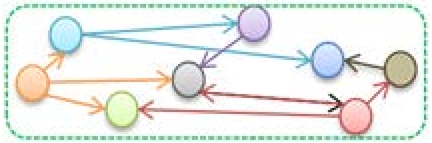
NP-hardness
Poisson Binomial Distribution
A Decision Making Task
GIVEN
n users
CHOOSE
A subset to vote
Majority Voting
YES ✓
OR
NO
r_1 r_2 r_3 r_n Integrated Constraint
ϵ_1 ϵ_2 ϵ_3 ϵ_n Individual Error-rate
Profiling of users
t_1 t_2 t_3 t_n User Activity
e_1 e_2 e_3 e_n User Experience
a_1 a_2 a_3 a_n Authority in HITS
s_1 s_2 s_3 s_n Score in PageRank
Knowledge diffusion graph based on "RT"
Rank users with PageRank and HITS

Outline

- Motivation
- Problem Definition
- Jury Selection Algorithm
- **Evaluation**

Parameter Estimation

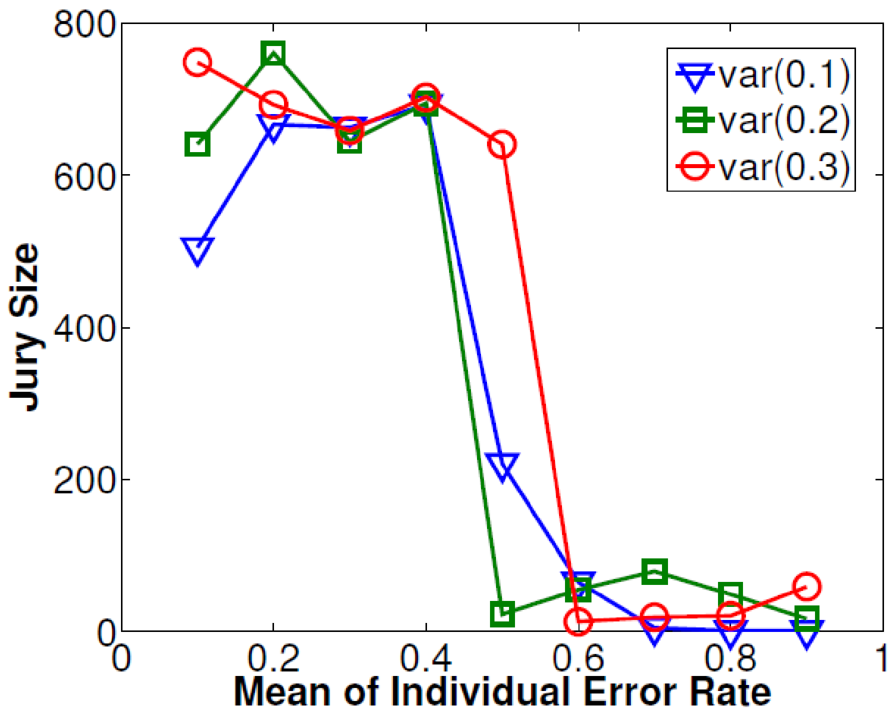
- How to estimate such parameter is itself a research topic
- Individual Error Rate (ϵ) -- 'RT' graph
  - **PageRank** and **HITS**
  - The score in rank is normalized to be the individual error rate



- Integrated requirement*(r)* – account info
  - ***Account Age*** and ***Account Activity***

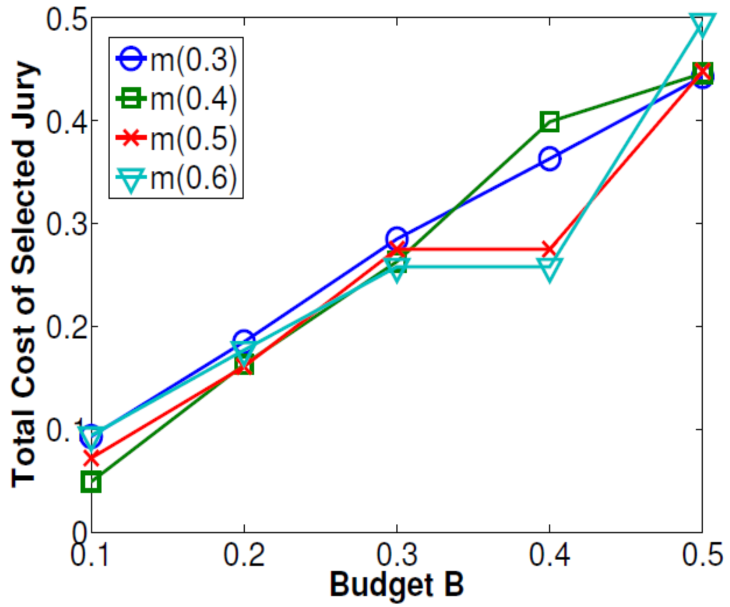
Data Preparation

- We test our algorithms on both synthetic data and real Twitter data
- Varying
  - Size
  - Mean
  - Variance
- 3.4GHz Win7 PC, programmed in C++

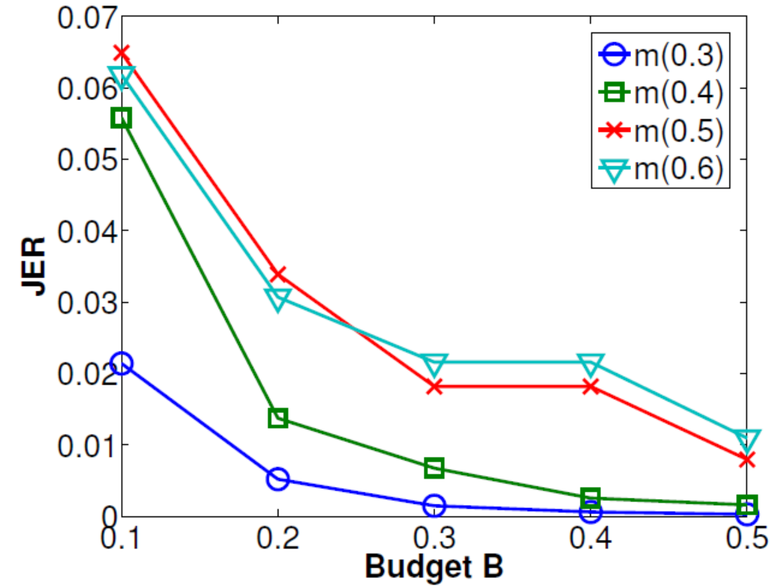


(a) Jury Size v.s. Individual Error-rate

- Mean = 0.5 is the turning point
- On the right side, “truth rests in the hands of a few.”

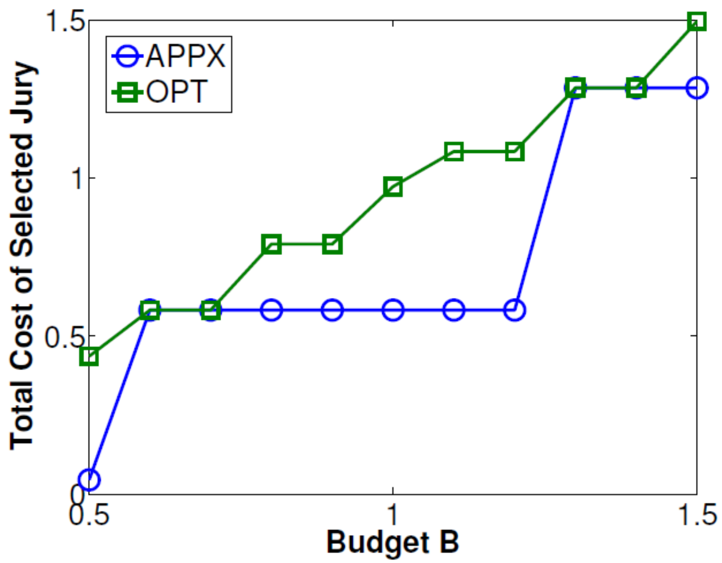


(c) Budget v.s. Total Cost

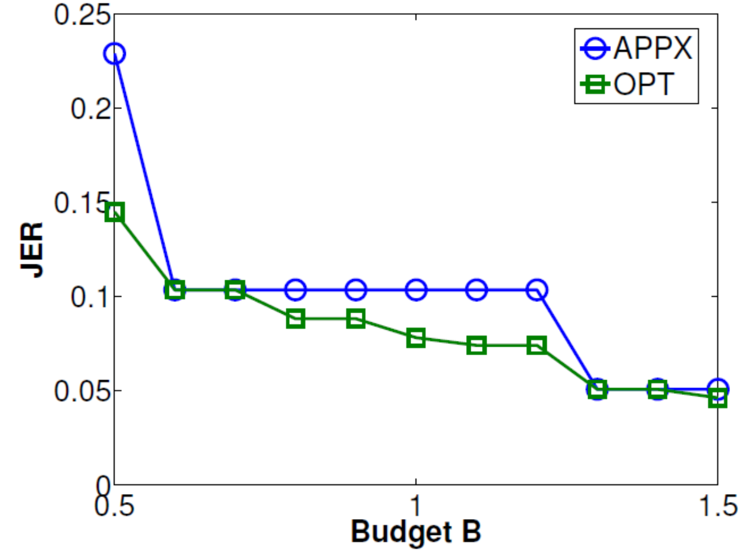


(d) Budget v.s. JER

- While the budget increases
  - The total cost also increase
  - The jury error rate decreases



(e) APPX v.s. OPT on Total Cost



(f) APPX v.s. OPT on JER

- Green – Accurate Algorithm (test with N=20)
- Blue – approximation algorithm
  - O(nlogn)
  - Good approximation on JER

Take-away and Future Work

- Take-away
  - Cultivate a pool of candidate jurors
  - JER deceases very fast according to the size of jury

- Future Work
  - Beyond direct payment
    - Prediction Market
  - Beyond decision making
    - Campaign Boosting

Thank You

- Q & A

Is Döner Kebab available in Hong Kong?