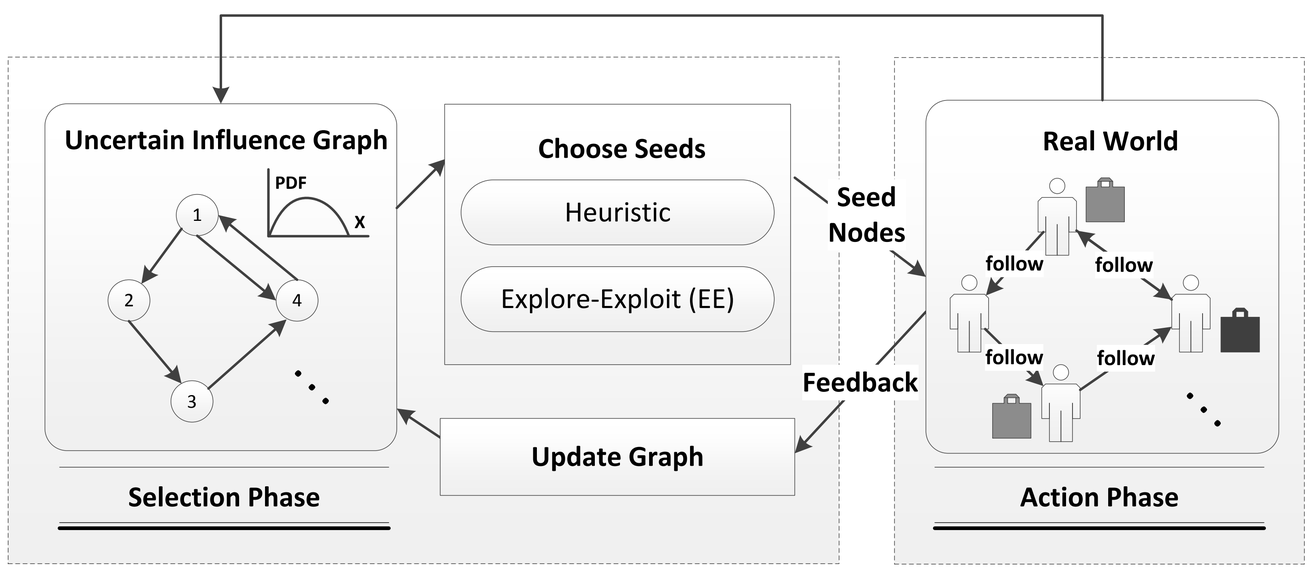
Online Influence Maximization

Siyu Lei **Luyi Mo** **Reynold Cheng**
University of Hong Kong

Silviu Maniu
Noah's Ark Lab, Huawei

Pierre Senellart
Télécom ParisTech & NUS

s



- **Objective**: effective influence maximization (IM) when the influence probabilities between users are unknown
- **Framework**: given a budget (number of trials), maximise influence spread
  - different case from one-step IM
  - classic tradeoff between exploration (refining the model) and exploitation (using the model)

OIM Framework

Algorithm 1 `Framework(G, k, N)`

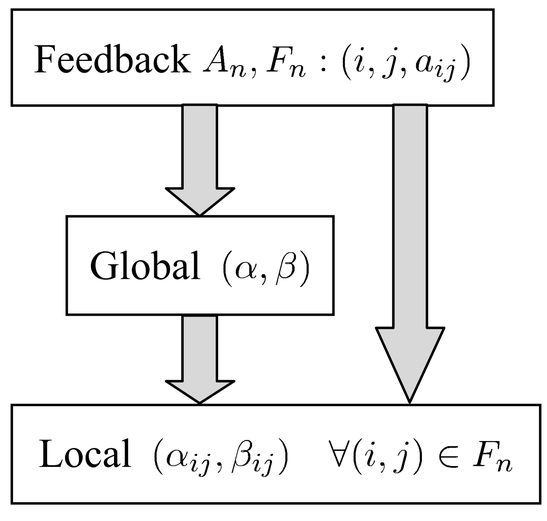
Input: # trials N, budget k, uncertain influence graph G
Output: seed nodes $S_n(n = 1 \dots N)$, activation results A
$A \leftarrow \emptyset$
for $n = 1$ **to** N **do**
  $S_n \leftarrow \texttt{Choose}(G, k)$
  $(A_n, F_n) \leftarrow \texttt{RealWorld}(S_n)$
  $A \leftarrow A \cup A_n$
  $\texttt{Update}(G, F_n)$
return $\{S_n | n = 1 \dots N\}$, A

Model: an uncertain graph of influence probabilities, starting from prior knowledge
Algorithm: iterative process involving selection of seed users, feedback gathering, and update of the model

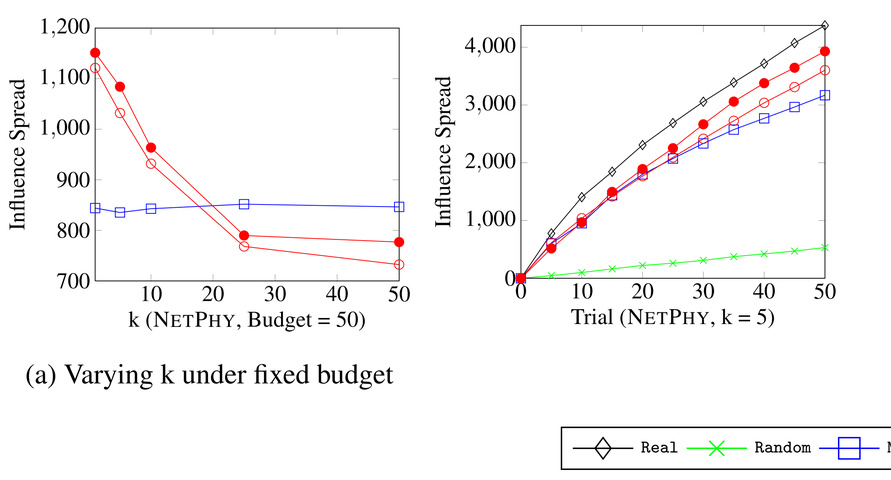
- **Seed selection**: use the uncertain graph for selection, by applying IM black box algorithms and explore—exploit strategies (ε-greedy, confidence bound, exp. gradient)
- **Real-world feedback**: test the chosen seeds in the real-world, and get an activation feedback trace

Crucial component — graph update
After each step and feedback, the uncertain influence graph is updated, using a combination of:

- local update: each edge in the feedback is updated in a Bayesian manner (Beta distributions)
- global update: every edge in the graph is updated using methods such as maximum likelihood or least squares regression



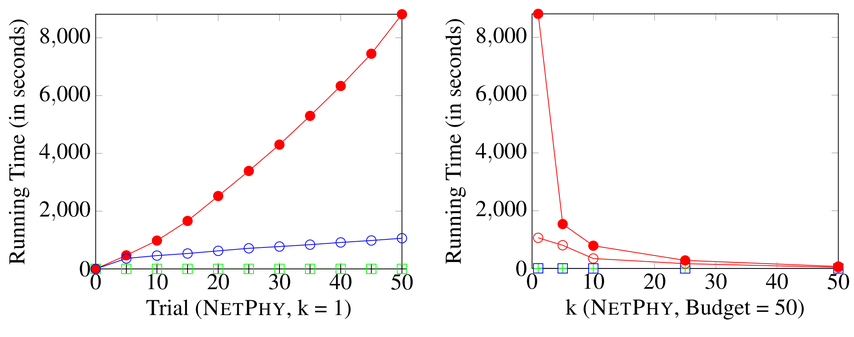
Effectiveness and Efficiency



(a) Varying k under fixed budget

(b) Varying k under fixed trials

- effectiveness: using explore—exploit combined with graph updates increases the influence spread
- efficiency: caching samples is highly effective and is comparable to non-sampling baselines



(a) Time v.s. N

(b) Time v.s. k