Boosting Information Spread: An Algorithmic Approach

Yishi Lin *, Wei Chen †, John C.S. Lui *

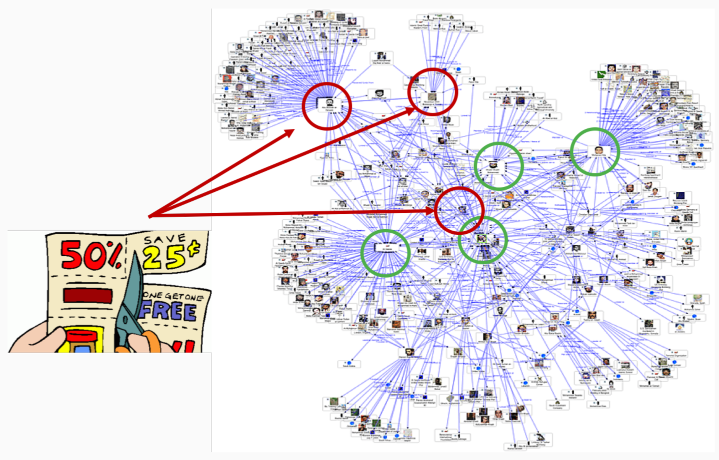
* The Chinese University of Hong Kong, † Microsoft Research

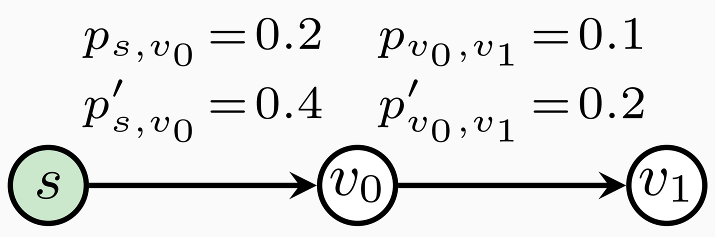
Problem Overview

Motivation

Some marketing strategies <u>“boost”</u> customers so that they are *more likely to be influenced by friends or to influence their friends*.

- E.g., customer incentive programs, social media advertising, referral marketing





B	$\sigma_S(B)$	$\Delta_S(B)$
$\emptyset$	1.22	0.00
$\{v_0\}$	1.44	0.22
$\{v_1\}$	1.24	0.02
$\{v_0, v_1\}$	1.48	0.26

Figure 1: Influence Boosting Model

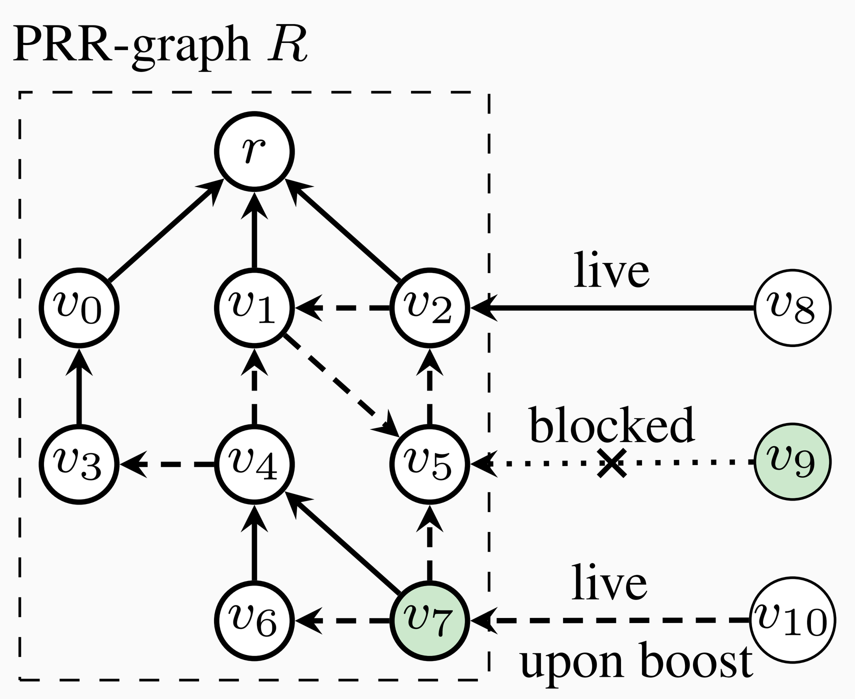
Problem Definition

- **Problem**: Given a graph $G = (V, E)$ with influence probabilities on edges, and a set $S \subseteq V$ of seeds, find a boost set $B \subseteq V$ with k nodes, such that the <u>boost of influence spread</u> of B denoted by $\Delta_S(B)$ is maximized.
- **Hardness**: NP-hard. The computation of $\Delta_S(B)$ given S and B is #P-hard.
- **Submodularity**: $\Delta_S(B)$ is neither submodular nor supermodular.

Boosting on General Graphs

Building blocks

- Potentially Reverse Reachable Graphs (PRR-graphs)
  - Usage: Estimate boost of influence and its lower bound
- State-of-the-art influence maximization techniques
  - Usage: Sampling PRR-graphs
- Sandwich approximation strategy
  - Approx. ratio: $\Delta_S(B_{\text{sa}}) \geq \frac{\mu(B^*)}{\Delta_S(B^*)} \cdot (1-1/e-\varepsilon) \cdot OPT$



- Estimating the boost
  - $f_R(\emptyset) = 0$
  - $f_R(\{v_1\}) = 1$
  - $f_R(\{v_3\}) = 1$
  - $f_R(\{v_2, v_5\}) = 1$
- Critical nodes
  - $C_R = \{v_1, v_3\}$
- Estimating the lower bound
  - $\mu_R(B) = \mathbb{I}(B \cap C_R \neq \emptyset)$

Figure 2: Example of a PRR-graph and related concepts

Boosting: Algorithm Design

Steps of PRR-Boost/PRR-Boost-LB

1. Sampling PRR-graphs for estimating the boost
2. Node selection according to estimated lower bound
   - `PRR-Boost-LB` returns here
3. Node selection according to estimated boost
4. Return the “better” solution
   - according the to estimated boost of influence

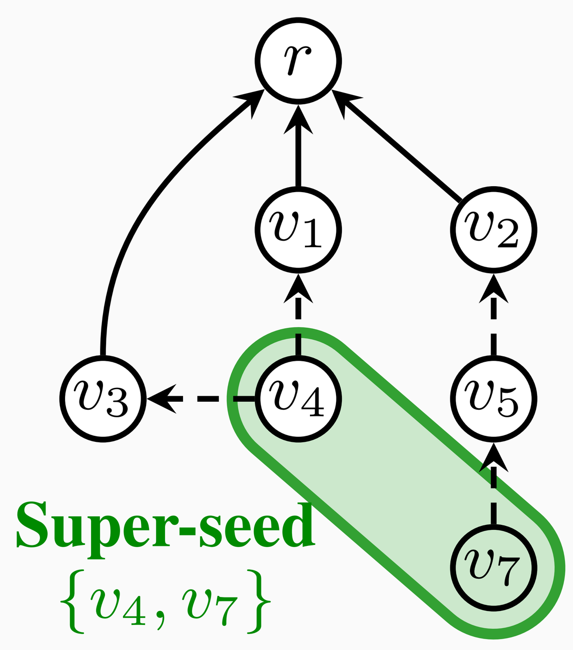
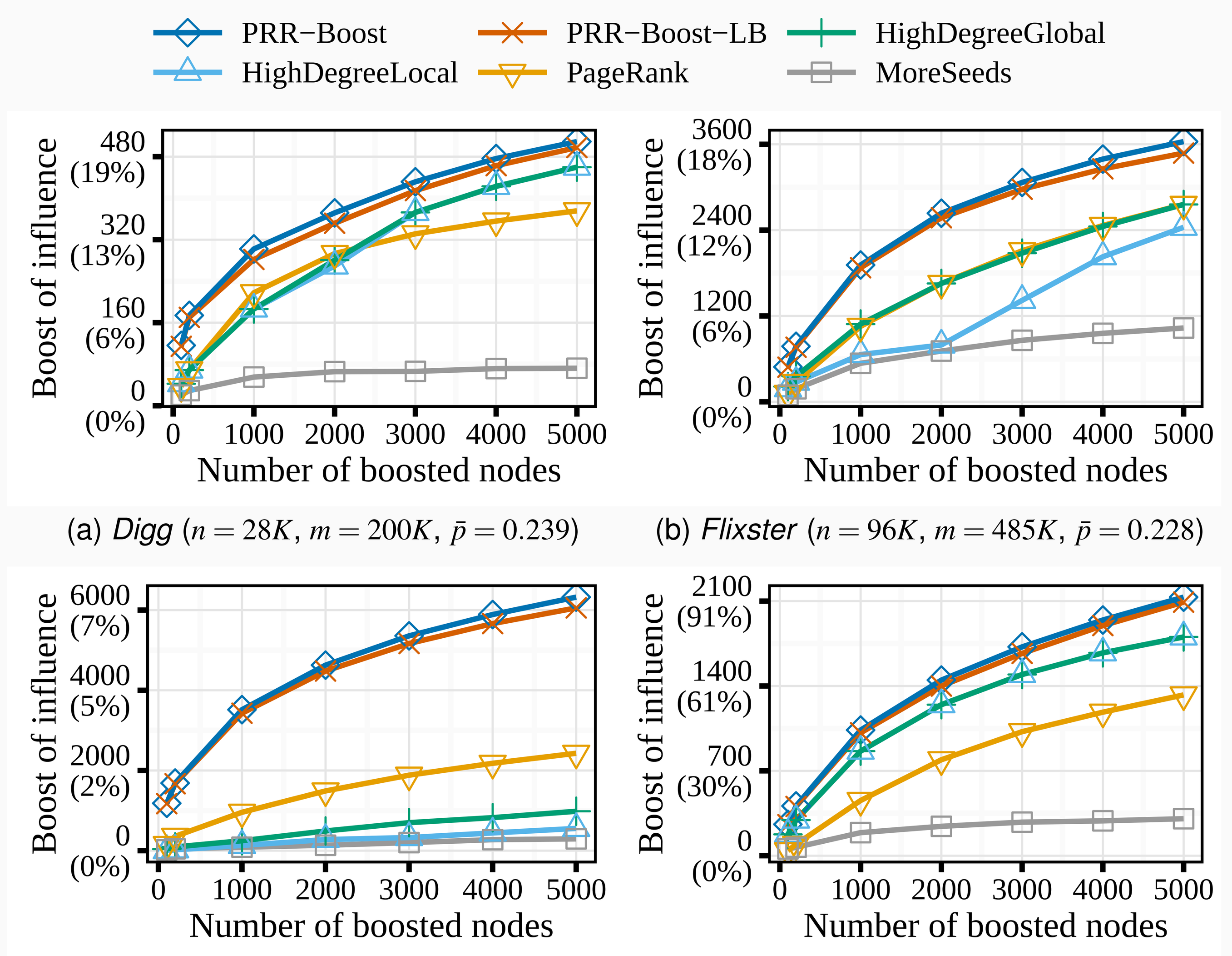


Figure 3: Example of a compressed PRR-graph

With a prob. of at least $1-n^{-\ell}$, `PRR-Boost`/`PRR-Boost-LB`

- returns a $(1-1/e-\varepsilon) \cdot \frac{\mu(B^*)}{\Delta_S(B^*)}$-approximate solution;
- has expected running time $O\left(\frac{EPT}{OPT_\mu} \cdot k(k+\ell)(n+m)\log n/\varepsilon^2\right)$.

Evaluation of PRR-Boost and PRR-Boost-LB



(a) *Digg* ($n = 28K$, $m = 200K$, $\bar{p} = 0.239$) (b) *Flixster* ($n = 96K$, $m = 485K$, $\bar{p} = 0.228$)

(c) *Twitter* ($n = 323K$, $m = 2.14M$, $\bar{p} = 0.608$) (d) *Flickr* ($n = 1.45M$, $m = 2.15M$, $\bar{p} = 0.013$)

Figure 4: Boost of the influence versus k (50 influential seeds).

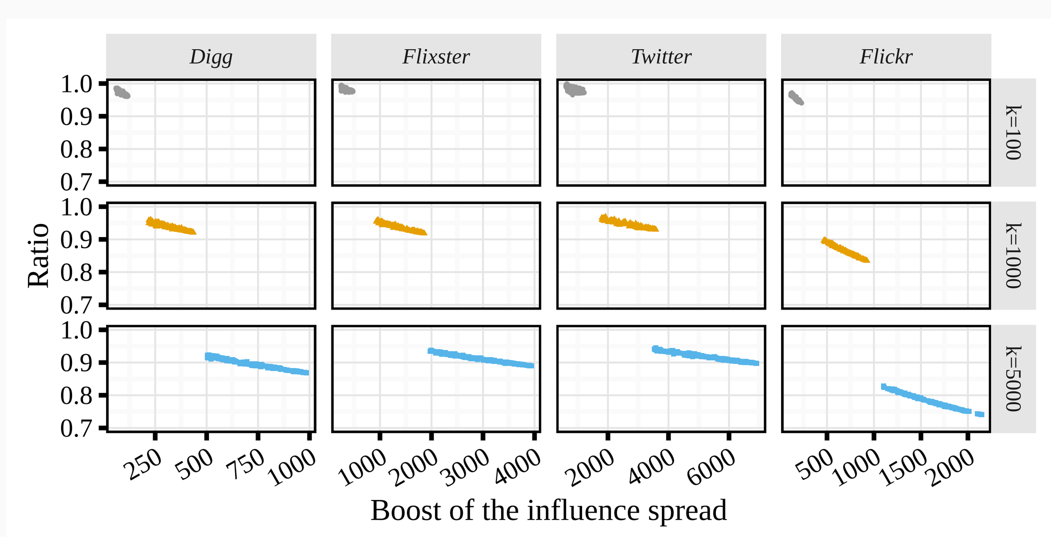
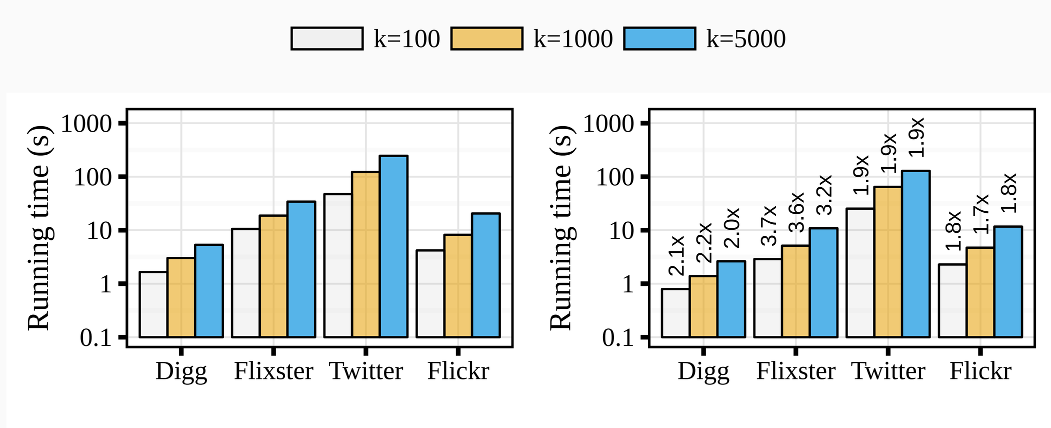


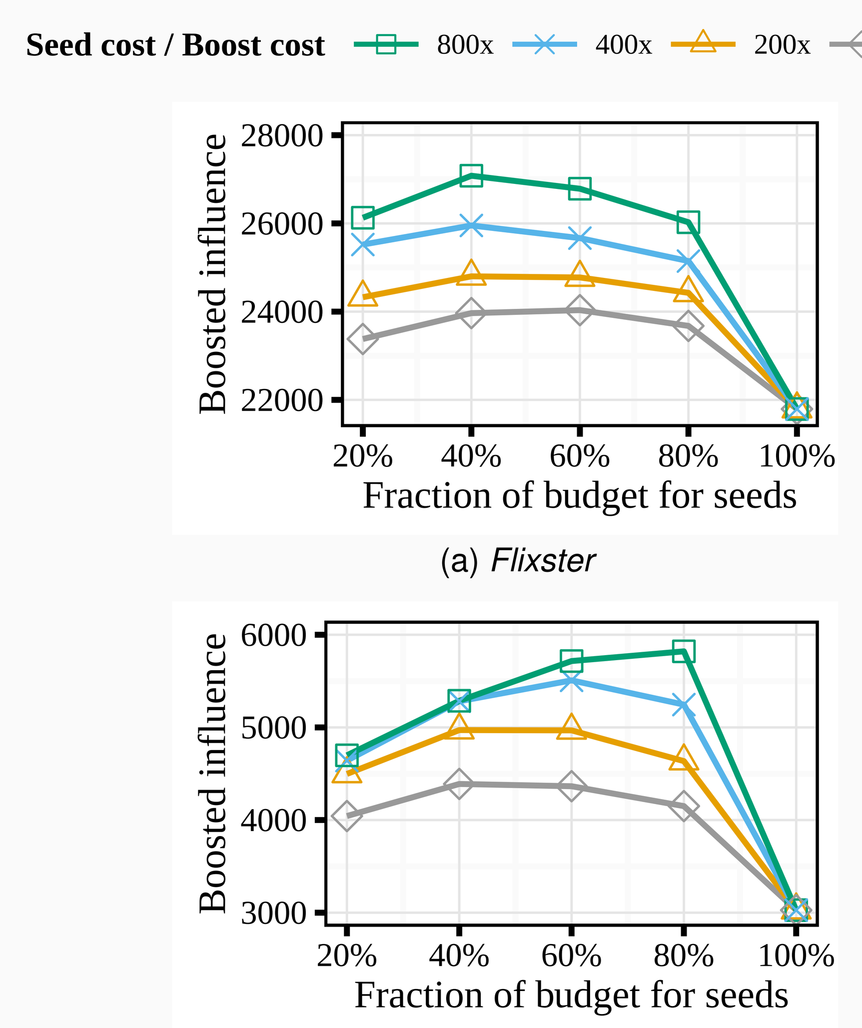
Figure 5: Sandwich Approximation: $\frac{\mu(B)}{\Delta_S(B)}$ (50 influential seeds).



(a) `PRR-Boost` (b) `PRR-Boost-LB`

Figure 6: Running time (50 influential seeds).

Budget Allocation



(a) *Flixster*

(b) *Flickr*

Figure 7: Budget allocation between seeding and boosting.

LaTeX TikZposter