2016/9/22 ECML-PKDD 2016

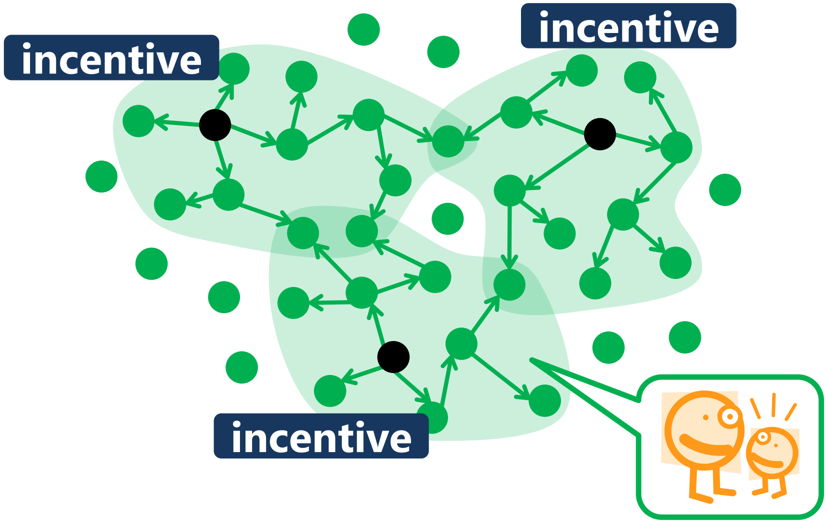
Maximizing Time-decaying Influence in Social Networks

Naoto Ohsaka (UTokyo)
Yutaro Yamaguchi (Osaka Univ.)
Naonori Kakimura (UTokyo)
Ken-ichi Kawarabayashi (NII)
ERATO Kawarabayashi Large Graph Project

ERATO

Influence maximization

[Kempe-Kleinberg-Tardos. *KDD'03*]



Viral marketing application

[Domingos-Richardson. *KDD'01*]

Given

- Social graph $G = (V, E)$
- Diffusion model $\mathcal{M}$
- Integer k

Goal

- maximize $\sigma_{\mathcal{M}}(S) \quad (|S| = k)$

  $\sigma_{\mathcal{M}}(S) \coloneqq \mathbf{E}[\text{size of cascade initiated by } S \text{ under } \mathcal{M}]$

Formulation by
[Kempe-Kleinberg-Tardos. *KDD'03*]

Used classical diffusion models

independent cascade [Goldenberg-Libai-Muller. *Market. Lett.'01*]

linear threshold [Granovetter. *Am. J. Sociol.'78*]

+ Tractable

$\forall X \subseteq Y, v \notin Y$

$\sigma(X \cup v) - \sigma(X) \geq \sigma(Y \cup v) - \sigma(Y)$

Monotonicity & submodularity of $\sigma(\cdot)$

[Kempe-Kleinberg-Tardos. *KDD'03*]

Near-linear time approx. algorithms

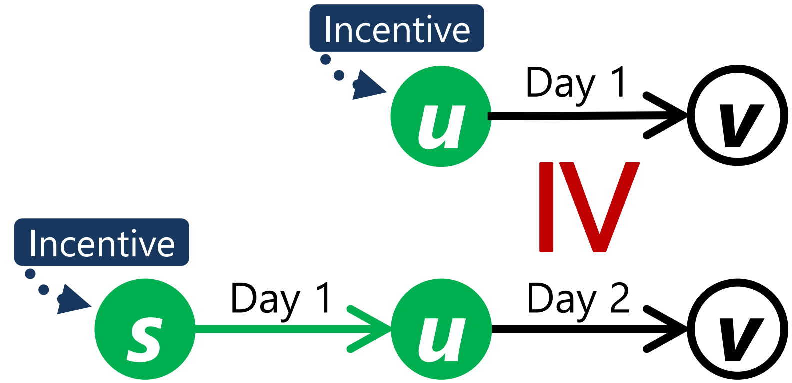
[Borgs-Brautbar-Chayes-Lucier. *SODA'14*] [Tang-Shi-Xiao. *SIGMOD'15*]

− Too simple

No time-decay effects, no time-delay propagation

Our aspect : **Time-decay effects**

The power of influence (rumor)
depends on arrival time
decreases over time



Time-delay propagations have been studied

[Saito-Kimura-Ohara-Motoda. *ACML'09*] [Goyal-Bonchi-Lakshmanan. *WSDM'10*]
[Saito-Kimura-Ohara-Motoda. *PKDD'10*] [Rodriguez-Balduzzi-Schölkopf. *ICML'11*]
[Chen-Lu-Zhang. *AAAI'12*] [Liu-Cong-Xu-Zeng. *ICDM'12*]

Time-decay effects have not much …

[Cui-Yang-Homan. *CIKM'14*] No guarantee for influence maximization

Our question

<u>What if incorporate time-decay effects?</u>

Monotonicity? Submodularity? Scalable algorithm?

Our contribution

Generalize independent cascade & linear threshold with

Time-decay effects & **Time-delay propagation**

Non-increasing probability — Arbitrary length distribution

↳ **Essential**

- Characterization

Diffusion process of TV-IC	=	Reachability on a random shrinking dynamic graph

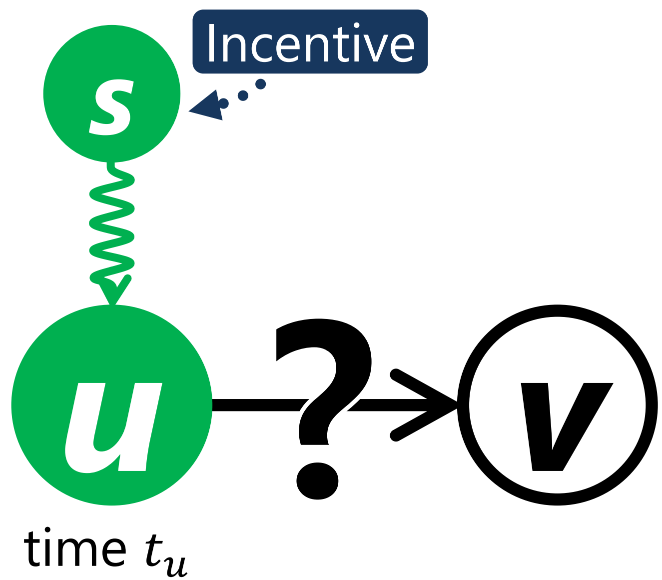
- Monotonicity & submodularity of $\sigma_{\mathrm{TVIC}}(\cdot)$

- Scalable and accurate greedy algorithms
  based on reverse influence set [Tang-Shi-Xiao. *SIGMOD'15*]
  $\mathcal{O}\left(k\left(|E| + \frac{|E|^2}{|V|}\right)\frac{\log^2|V|}{\epsilon^2}\right)$ time & $(1-\mathrm{e}^{-1}-\epsilon)$-approx.

Time-varying independent cascade (TV-IC) model

$G = (V, E)$ Social graph
$\boldsymbol{p_{uv}}$ Non-increasing probability function
$\boldsymbol{f_{uv}}$ Arbitrary length distribution

0. u became **active** at t_u
1. Sample a delay δ_{uv} from $\boldsymbol{f_{uv}}$



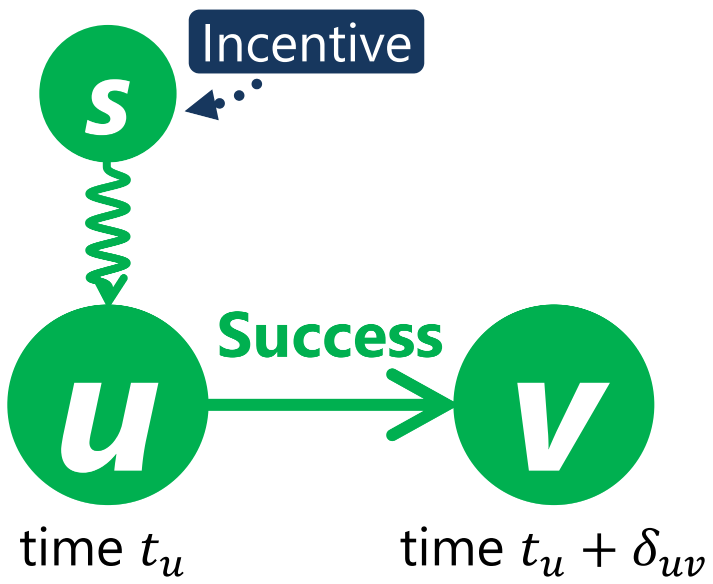
Objective function

$\sigma_{\mathrm{TVIC}}(S) \coloneqq \mathbf{E}[\#$ active vertices initiated by $S]$

Time-varying independent cascade (TV-IC) model

$G = (V, E)$ Social graph
$\boldsymbol{p_{uv}}$ Non-increasing probability function
$\boldsymbol{f_{uv}}$ Arbitrary length distribution

0. u became **active** at t_u
1. Sample a delay δ_{uv} from $\boldsymbol{f_{uv}}$
2A. **Success** w.p. $\boldsymbol{p_{uv}(t_u + \delta_{uv})}$
v becomes **active** at $t_u + \delta_{uv}$



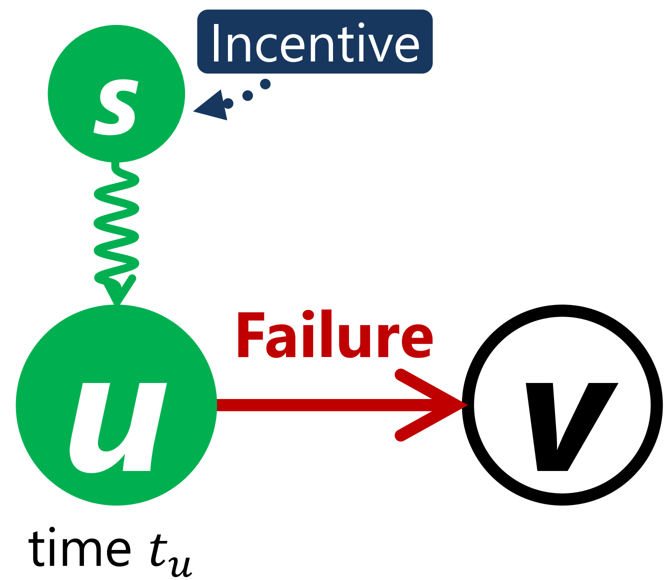
Objective function

$\sigma_{\mathrm{TVIC}}(S) \coloneqq \mathbf{E}[\#\text{ active vertices initiated by } S]$

Time-varying independent cascade (TV-IC) model

$G = (V, E)$ Social graph
$\boldsymbol{p_{uv}}$ Non-increasing probability function
$\boldsymbol{f_{uv}}$ Arbitrary length distribution

0. u became **active** at t_u
1. Sample a delay δ_{uv} from $\boldsymbol{f_{uv}}$
2B. **Failure** w.p. $\mathbf{1} - \boldsymbol{p_{uv}}(\boldsymbol{t_u} + \boldsymbol{\delta_{uv}})$
v remains **inactive**



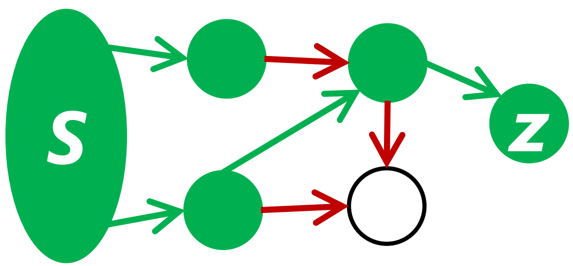
Objective function

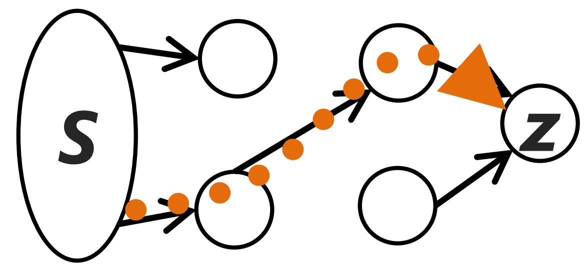
$\sigma_{\mathrm{TVIC}}(S) \coloneqq \mathbf{E}$[# active vertices initiated by S]

Characterization of the IC model

[Kempe-Kleinberg-Tardos. *KDD'03*]

S influences z in the IC process $\Longleftrightarrow$ S can reach z in a random subgraph G'

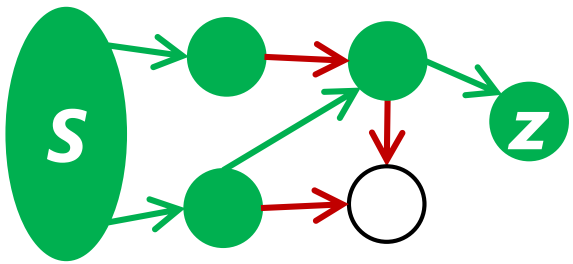


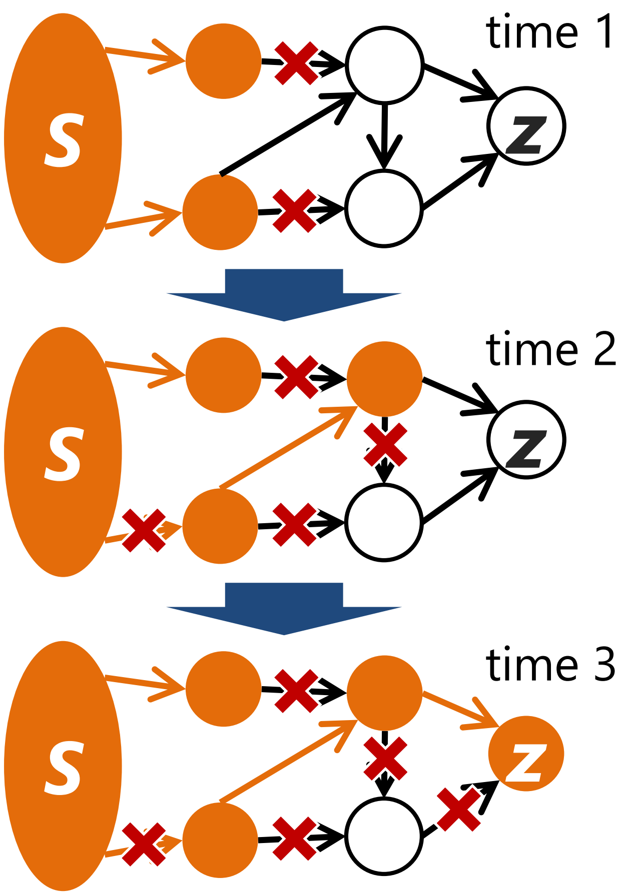


No need to consider time

Characterization of the TV-IC model

S influences z in the TV-IC process ⟺ S can reach z in a random <u>shrinking</u> <u>dynamic</u> graph G^{t}





- <u>Dynamic</u> …
  G^{t} <u>changes</u> over time
  ∵ p_{uv} <u>changes</u> over time
- <u>Shrinking</u> …
  G^{t} has only <u>edge deletions</u>
  ∵ p_{uv} is <u>non-increasing</u>

Monotonicity & submodularity of $\sigma_{\mathrm{TVIC}}(\cdot)$

S influences z in the TV-IC process $\Leftrightarrow$ S can reach z in a random shrinking dynamic graph G^{t}

$$\sigma_{\mathrm{TVIC}}(S) = \mathbf{E}_{G^{\mathrm{t}}}[\text{\# vertices reachable from } S \text{ in } G^{\mathrm{t}}]$$

Our results : Monotone & submodular

Shrinking is a **MUST**

⇝ Greedy algorithm is $(1 - \mathrm{e}^{-1})$-approx.

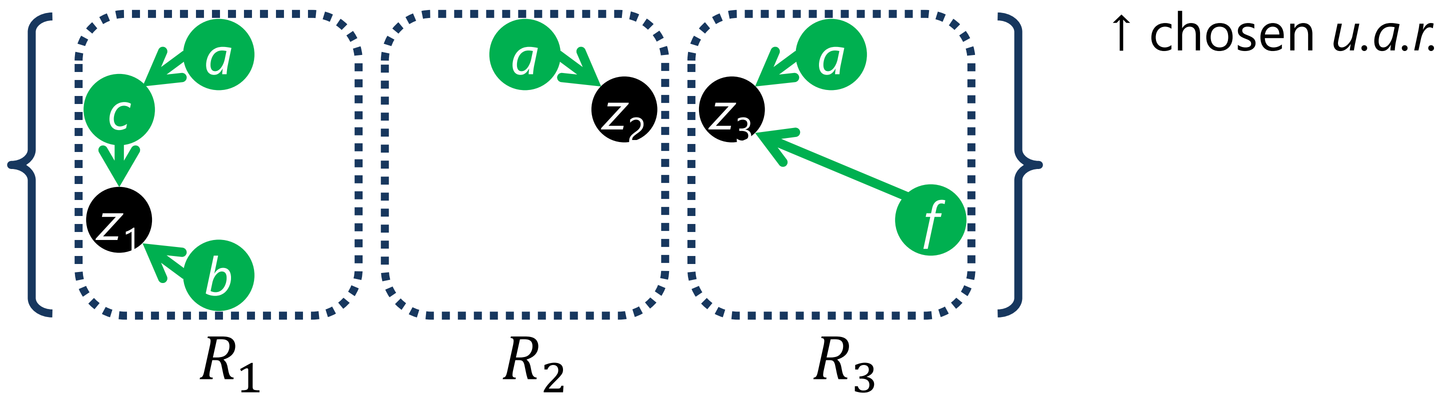
[Nemhauser-Wolsey-Fisher. *Math. Program.*'78]

Efficient approximation of $\sigma_{\mathrm{TVIC}}(\cdot)$

Greedy algorithm requires estimations of $\sigma_{\mathrm{TVIC}}(\cdot)$

Reverse influence set [Tang-Shi-Xiao. *SIGMOD'15*]

$R_i :=$ {vertices that would have influenced z_i}

$\mathcal{R} =$ 

$\sigma(S) \propto \mathbf{E}[\#$ sets in $\mathcal{R}$ intersecting $S]$

BFS-like alg. for the IC [Borgs-Brautbar-Chayes-Lucier. *SODA'14*]

Dijkstra-like alg. for the continuous IC [Tang-Shi-Xiao. *SIGMOD'15*]

☹ Cannot be applied to our model

Naive reverse influence set generation under the TV-IC model

Check each vertex separately

$v \in R_i$ $\Longleftrightarrow$ v can reach z_i in a random shrinking dynamic graph

$\rightsquigarrow$ $\Omega(|V| \cdot |E|)$ time

Efficient reverse influence set generation under the TV-IC model

Key = **latest activation time** $\tau[v]$

max. $\tau[v]$ s.t. (v gets active in $\tau[v]$) $\rightsquigarrow$ (z_i gets active)

$v \in R_i$	$\Leftrightarrow$	$\boldsymbol{\tau[v] \geq 0}$

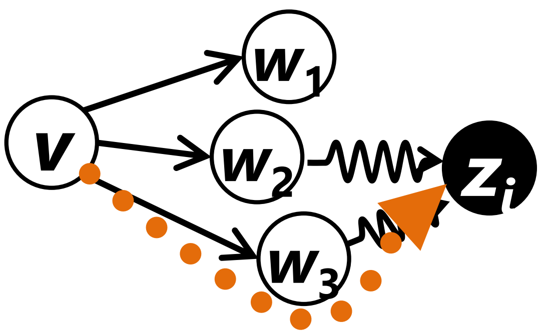
Computing $\tau[v]$

“ v must pass through (v, w) for some w to reach z_i ”

$$\tau[v] = \max_{(v,w) \in E} \min\{\tau[w] - \delta_{vw}, p_{vw}^{-1}(x_{vw})\}$$

Dynamic programming yields

$$+\infty = \tau[z_i] \geq \tau[v_1] \geq \cdots \geq \tau[v_r] \geq 0$$



Performance analysis

Lemma 2 Correctness

Lemma 3 Running time = $\mathcal{O}\left(\frac{|E|\cdot\mathrm{OPT}}{|V|}\log|V|\right)$ in exp.

Theorem 4

Our reverse influence set generation method + *IMM* (Influence Maximization via Martingales) framework [Tang-Shi-Xiao. *SIGMOD'15*]

$(1 - \mathrm{e}^{-1} - \epsilon)$-approx. w.h.p.

$\mathcal{O}\left(k\left(|E| + \frac{|E|^2}{|V|}\right)\frac{\log^2|V|}{\epsilon^2}\right)$ expected time

$\frac{|E|}{|V|}$ is small for real graphs

Experiments: Setting

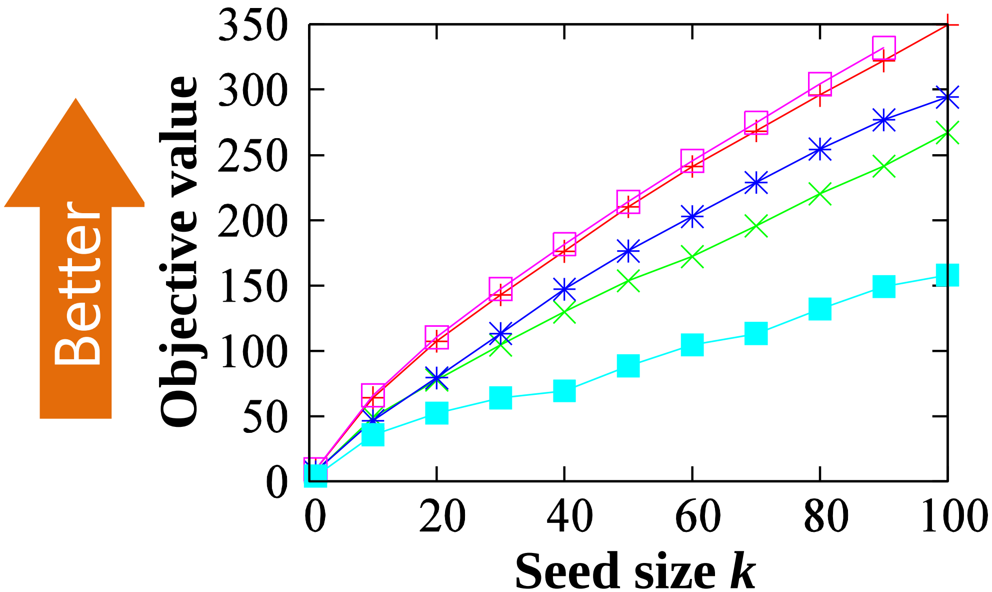
- Comparative algorithms

Algorithm	
This work	
IMM-CTIC [Tang-Shi-Xiao. *SIGMOD'15*]	Efficient algorithms under different models
IMM-IC [Tang-Shi-Xiao. *SIGMOD'15*]	Efficient algorithms under different models
LazyGreedy [Minoux. *Optim. Tech.'78*]	Accurately estimate $\sigma_{\mathrm{TVIC}}(\cdot)$ by simulations
Degree	Simple baseline

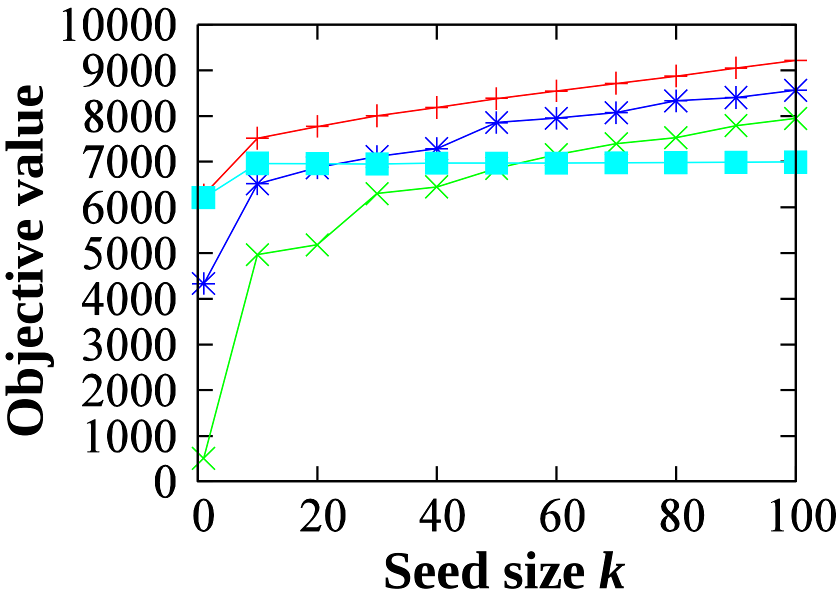
- $p_{uv}(t) = (\text{in-deg}(v) \cdot c \cdot t)^{-1}$
- $f_{uv}(\cdot) = $ (Weibull distribution) [Lawless. '02]

Experiments: Solution quality

$\sigma_{\mathrm{TVIC}}(\cdot)$ of seed sets produced by each method



ca-GrQc $|V|$=5K $|E|$=29K



ego-Twitter $|V|$=81K $|E|$=2,421K

LazyGreedy > **This work** > *IMM-IC* > *IMM-CTIC* > *Degree*

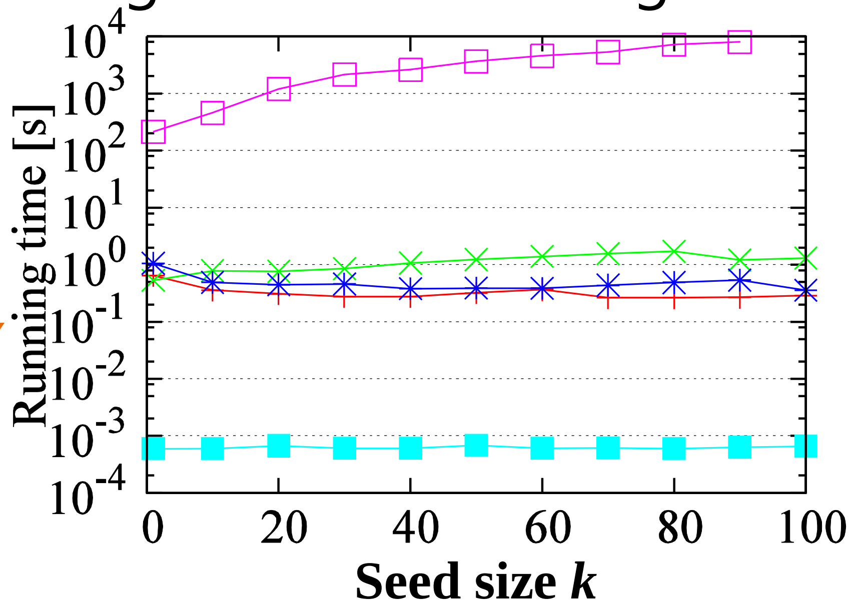
Accuracy guarantee (LazyGreedy, This work)

Ignore time-decay effects (IMM-IC, IMM-CTIC, Degree)

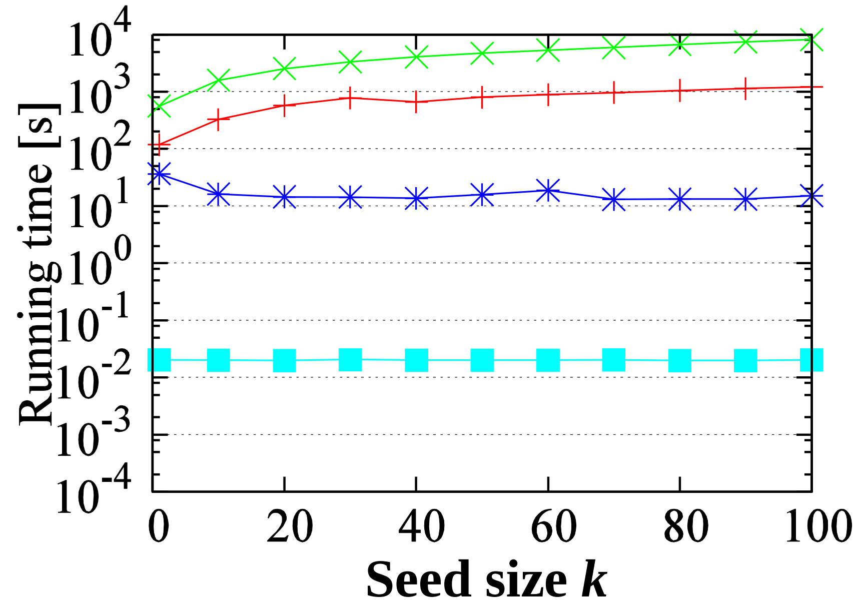
Dataset : Stanford Large Network Dataset Collection. http://snap.stanford.edu/data
Environment : Intel Xeon E5540 2.53 GHz CPU + 48 GB memory & g++v4.8.2 (-O2)

Experiments: Scalability

Running time for selecting seed sets of size k for each method



ca-GrQc $|V|$=5K $|E|$=29K



ego-Twitter $|V|$=81K $|E|$=2,421K

Degree ≫ **This work** ≈ *IMM-IC* > *IMM-CTIC* ≫ *LazyGreedy*

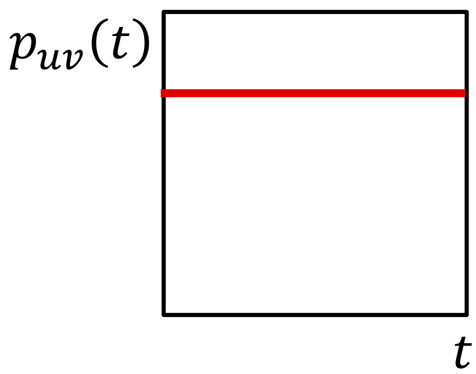
Just sorting — Near-linear time — Quadratic time

Dataset : Stanford Large Network Dataset Collection. http://snap.stanford.edu/data
Environment : Intel Xeon E5540 2.53 GHz CPU + 48 GB memory & g++v4.8.2 (-O2)

Summary & future work

Classical

Constant probability



Monotone & submodular

[Kempe+'03]

Influence maximization

$\mathcal{O}\left(k(|V|+|E|)\frac{\log|V|}{\epsilon^2}\right)$ time

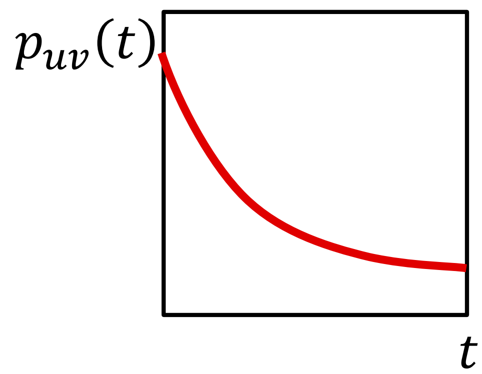
$(1-\mathrm{e}^{-1}-\epsilon)$-approx.

[Tang+'15]

generalize →

Our study

Time-decay probability



Monotone & submodular

Influence maximization

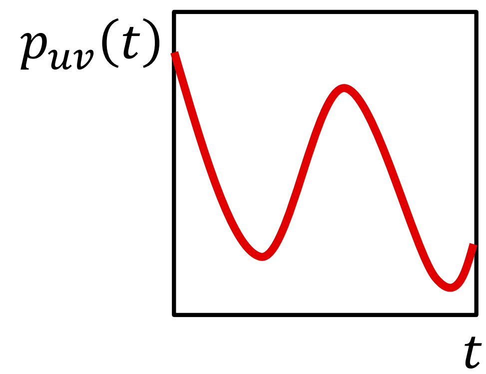
$\mathcal{O}\left(k\left(|E|+\frac{|E|^2}{|V|}\right)\frac{\log^2|V|}{\epsilon^2}\right)$ time

$(1-\mathrm{e}^{-1}-\epsilon)$-approx.

generalize →

More general

Time-varying probability



Simple extension violates submodularity

Efficient influence max.?