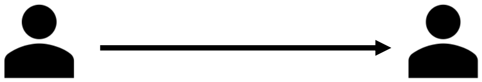
CAPACITY CONSTRAINED INFLUENCE MAXIMIZATION IN SOCIAL NETWORKS

Shiqi Zhang, Yiqian Huang, Jiachen Sun, Wenqing Lin
Xiaokui Xiao, Bo Tang

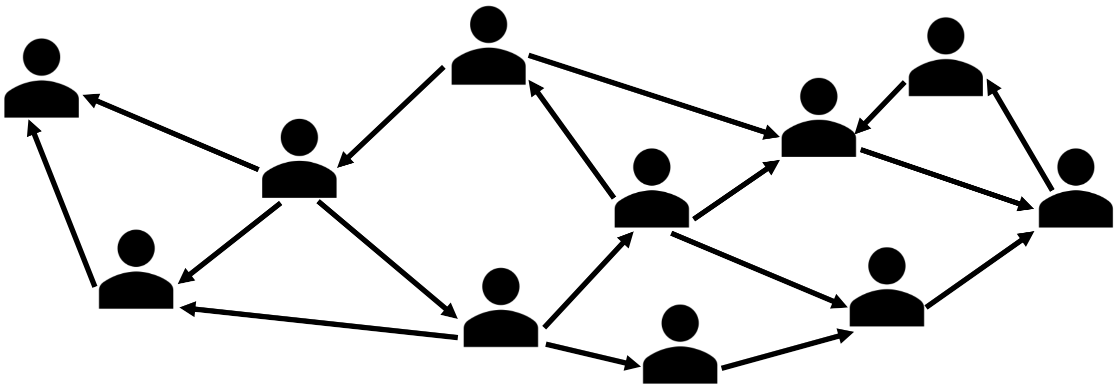
August 2023

INFLUENCE MAXIMIZATION (IM)

inactive user:
not adopt the product yet



followee can influence follower



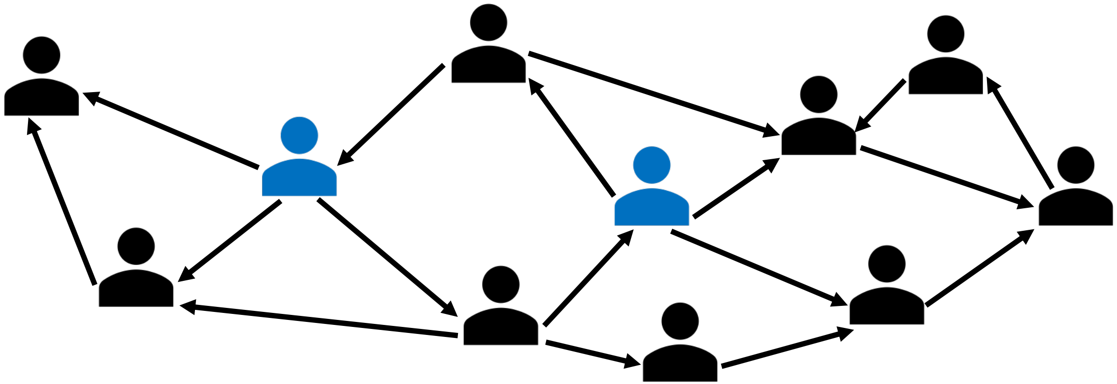
social network

- Motivated by viral marketing in social networks

INFLUENCE MAXIMIZATION (IM)

inactive user:
not adopt the product yet

initial adopter:
paid by the merchant



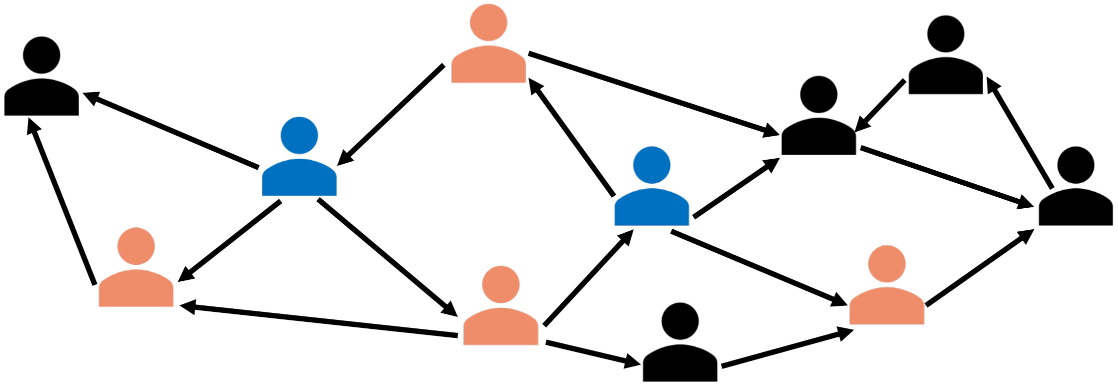
social network

- Motivated by viral marketing in social networks
  - pay k individuals

INFLUENCE MAXIMIZATION (IM)

initial adopter:
paid by the merchant

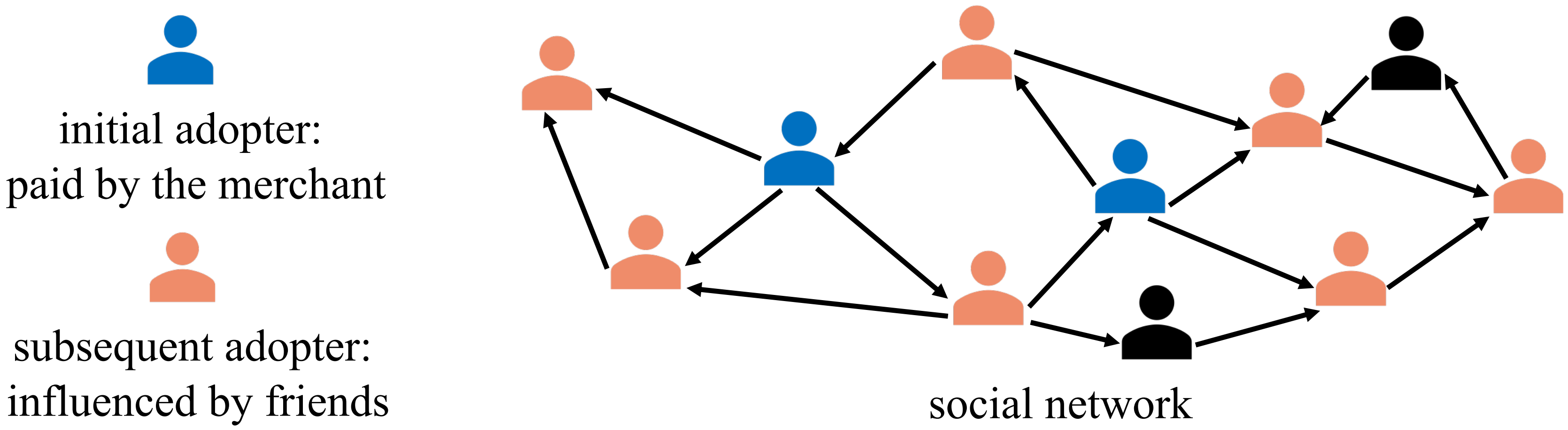
subsequent adopter:
influenced by friends



social network

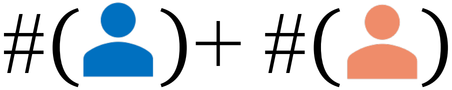
- Motivated by viral marketing in social networks
  - hope word-of-mouth promotes the given product

INFLUENCE MAXIMIZATION (IM)

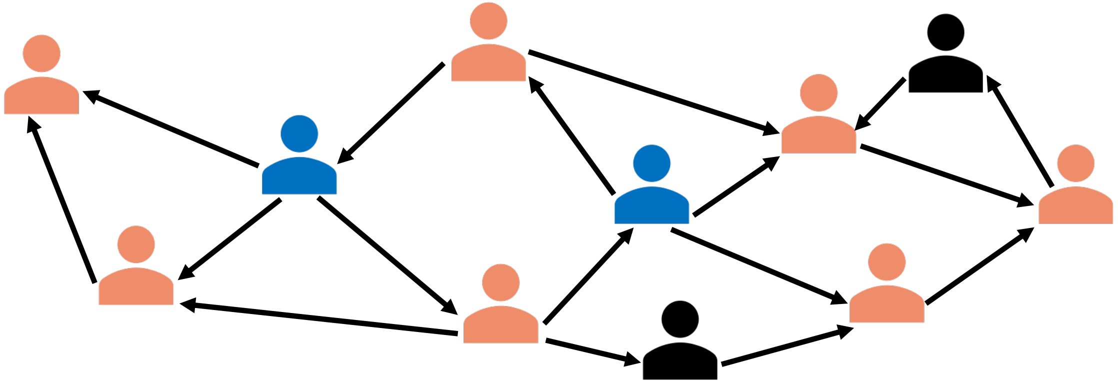


- Motivated by viral marketing in social networks
  - hope word-of-mouth promotes the given product
  - create a *cascade of influence*

INFLUENCE MAXIMIZATION (IM)



influence spread

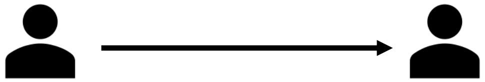


social network

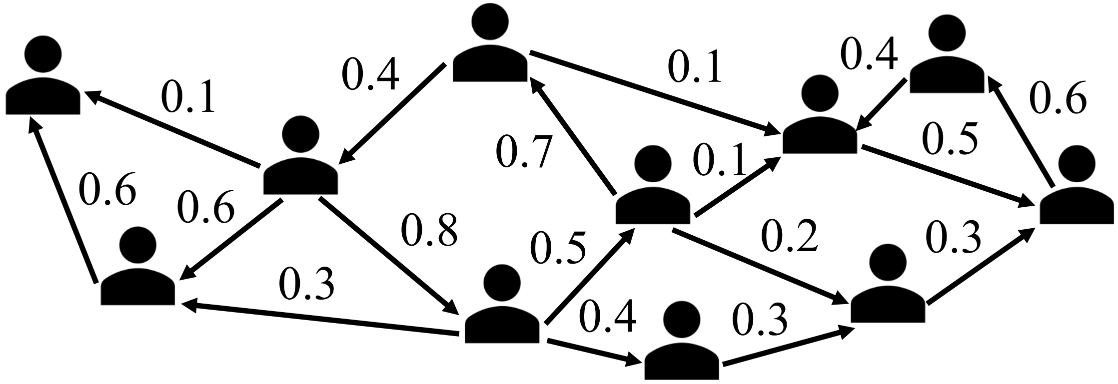
- Problem: for a fixed k, how to pick k individuals for the merchant such that the eventual influence spread is maximized?

EXISTING SOLUTIONS FOR IM

- Adopt a stochastic model M to simulate the propagation
  - i.e., under what condition will a user be influenced
- Example: Independent Cascade (IC) model



IC model:
u independently influences v
with probability $p_{u,v}$



social network

EXISTING SOLUTIONS FOR IM

- Adopt a stochastic model M to simulate the propagation
  - i.e., under what condition will a user be influenced
- Generate samples of the social network based on M
- Identify k influencers by using
  - the samples to estimate the spread
  - and the greedy algorithm over samples

LIMITATION #1 OF IM

- Conventional IM
  - considers the cost factor
  - ignores individual's capacity for spending efforts on consuming the promoting content
- User's capacity
  - is crucial as it determines the adoption of the product
  - is limited on online platforms
    - e.g., while playing e-games with friends

LIMITATION #2 OF IM

- Conventional IM
  - assumes influencers unconditionally become initial adopters
- Observation from real-world scenarios
  - influencers tend to be the friends of initial adopters

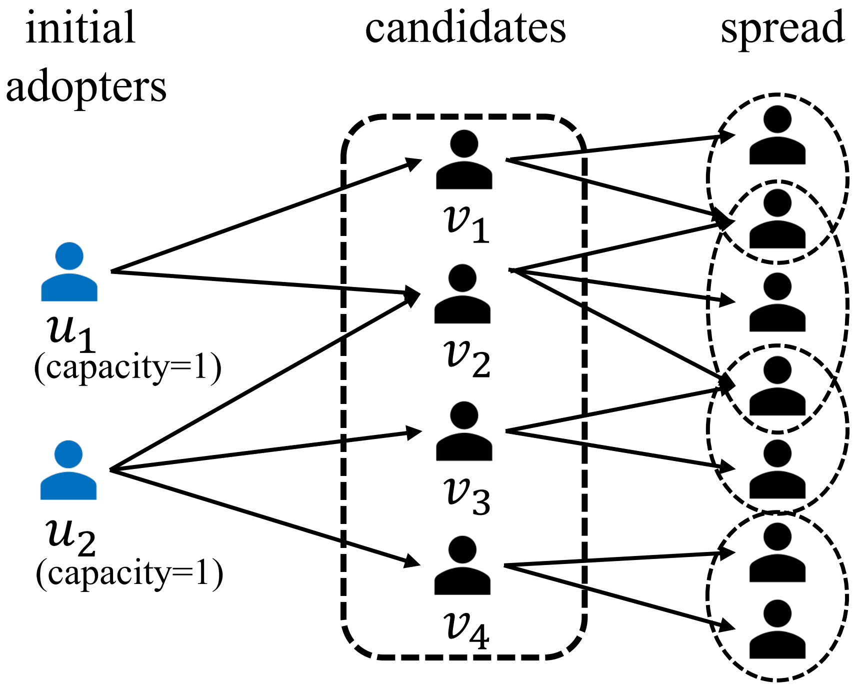
source: https://zootopia.fandom.com/wiki

CAPACITY CONSTRAINED IM (CIM)

- Input
  - social network G and stochastic model M
  - d initial adopters and capacity constant k
- Output
  - k influential friends (seeds) for each of d initial adopters
- Objective
  - maximize the spread of the set of all selected seeds
- CIM is NP-hard

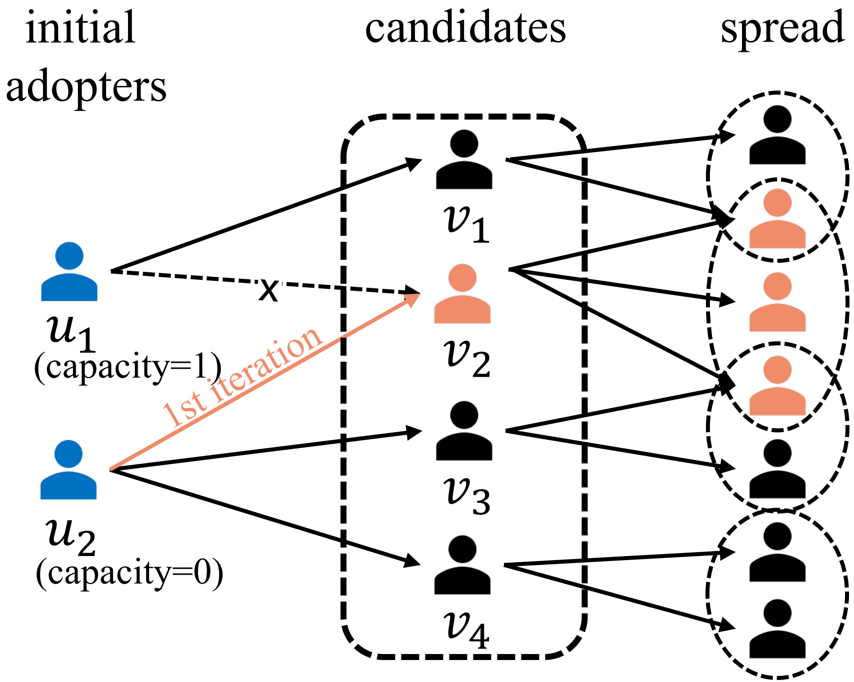
IDEA OF GREEDY ALGORITHMS

- MG-Greedy: select a user v from candidates as the next if
  - adding v to current seed set yields the largest lift
  - and, existing a v's friend in initial adopters that remains capacity
- $\frac{1}{2}$-approximate if known spread



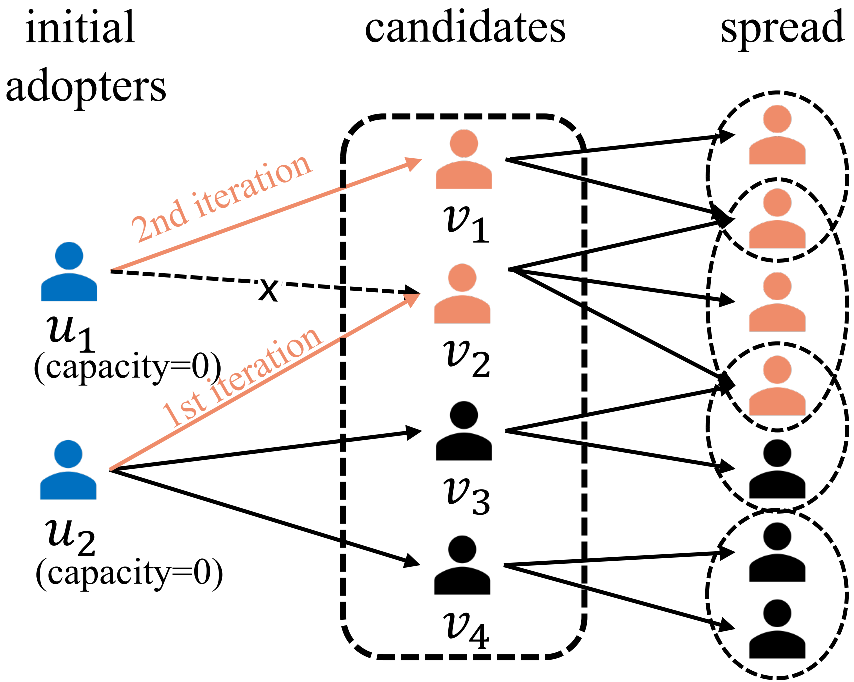
IDEA OF GREEDY ALGORITHMS

- MG-Greedy: select a user v from candidates as the next if
  - adding v to current seed set yields the largest lift
  - and, existing a v's friend in initial adopters that remains capacity
- $\frac{1}{2}$-approximate if known spread



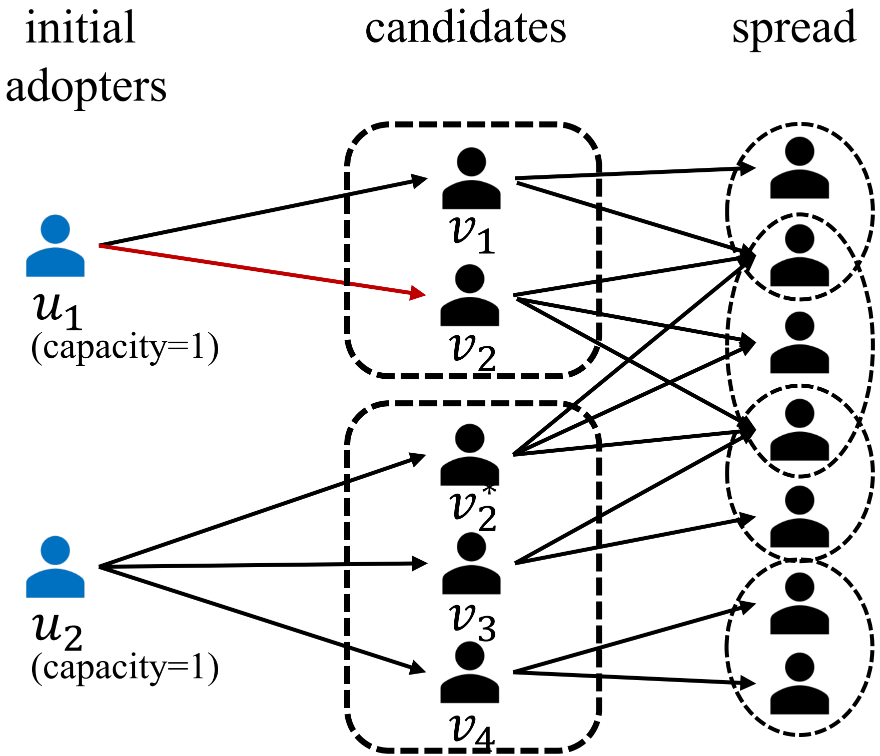
IDEA OF GREEDY ALGORITHMS

- MG-Greedy: select a user v from candidates as the next if
  - adding v to current seed set yields the largest lift
  - and, existing a v's friend in initial adopters that remains capacity
- $\frac{1}{2}$-approximate if known spread



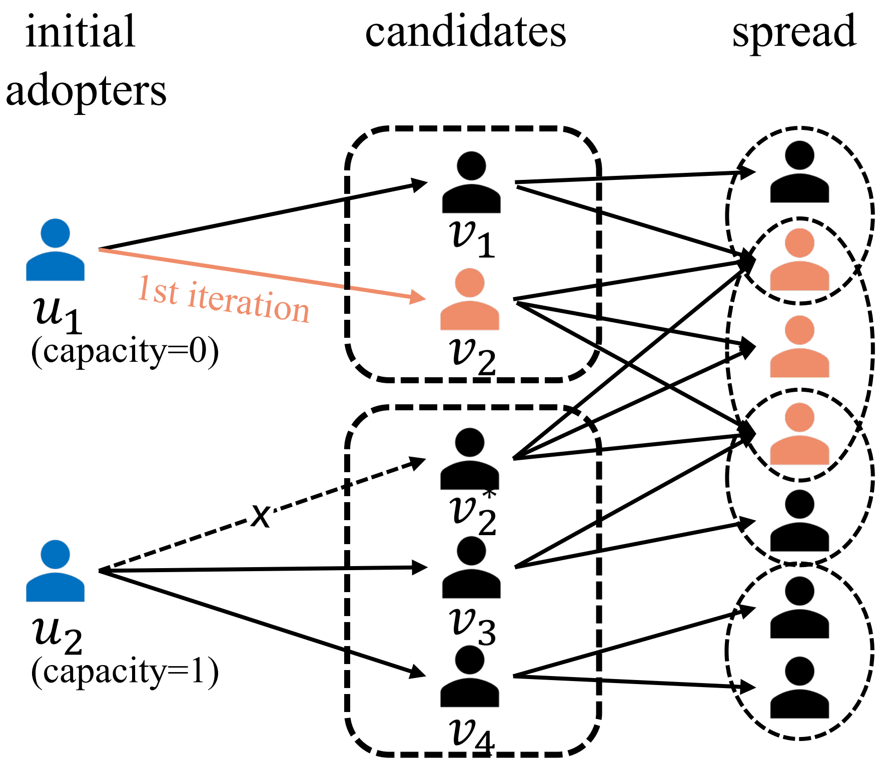
IDEA OF GREEDY ALGORITHMS

- RR-Greedy: select in a round-robin manner
  - choose an initial adopter u remaining capacity
  - select from u's candidates
  - add v to current seed set if yielding the largest lift
- $\geq \frac{1}{2}$-approximate if known spread



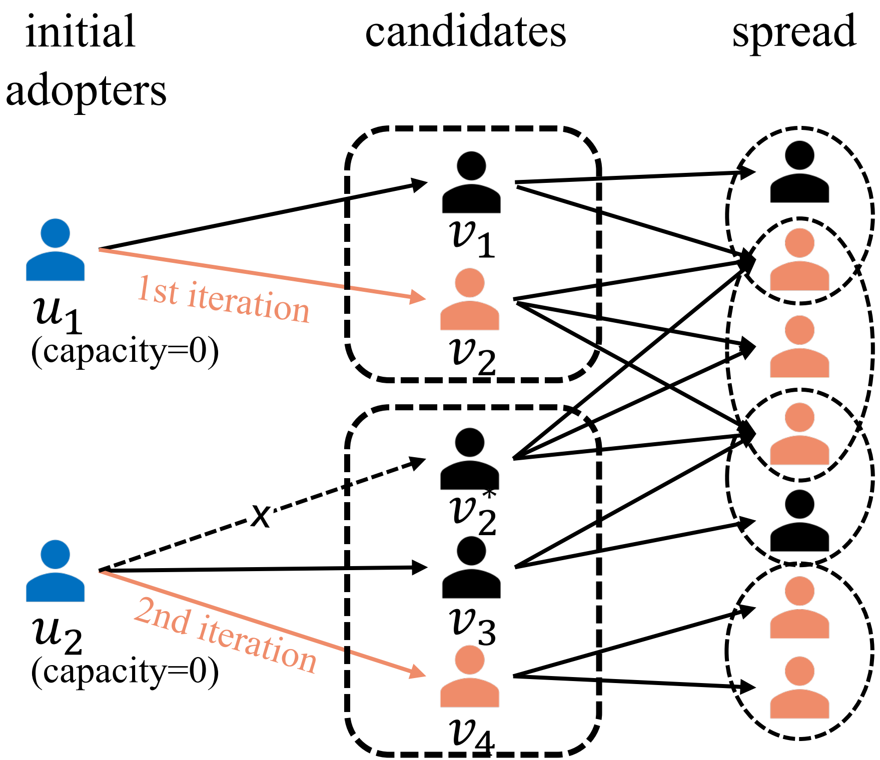
IDEA OF GREEDY ALGORITHMS

- RR-Greedy: select in a round-robin manner
  - choose an initial adopter u remaining capacity
  - select from u's candidates
  - add v to current seed set if yielding the largest lift
- $\geq \frac{1}{2}$-approximate if known spread



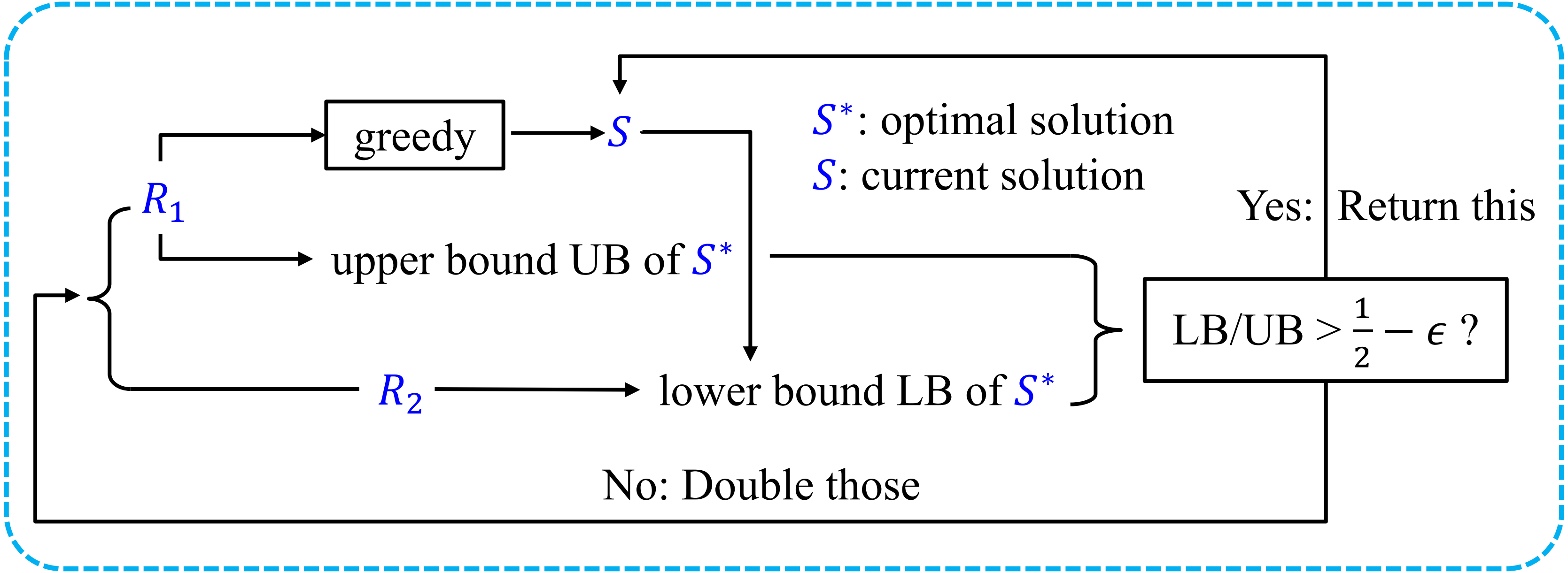
IDEA OF GREEDY ALGORITHMS

- RR-Greedy: select in a round-robin manner
  - choose an initial adopter u remaining capacity
  - select from u's candidates
  - add v to current seed set if yielding the largest lift
- $\geq \frac{1}{2}$-approximate if known spread



SCALABLE IMPLEMENTATIONS

- Borrow OPIM-C framework in SIGMOD'18
- Generate two equal-size sets of samples: R_1 and R_2



SCALABLE IMPLEMENTATIONS

- Borrow OPIM-C framework in SIGMOD'18
- Redesign each parameter by rigorous theoretical analysis
- Implementations:
  - MG-OPIM: greedy = MG-Greedy
  - RR-OPIM: greedy = RR-Greedy
  - RR-OPIM+: RR-OPIM with an optimized UB
- Result: $(\frac{1}{2}-\epsilon)$-approximate in near-linear running time

APPROACHES

- Local competitors: independently select k friends for each initial adopter
  - based on a heuristic score: Degree, PageRank
  - based on a SOTA IM solver: IMM, OPIM-C
- Greedy solutions
  - MG-Greedy, RR-Greedy
- Scalable implementations
  - MG-OPIM, RR-OPIM, RR-OPIM+

PUBLIC DATASETS

- Various types of datasets
  - *DNC*
  - *Blog*
  - *Twitch*
  - *Orkut*
  - *Twitter*

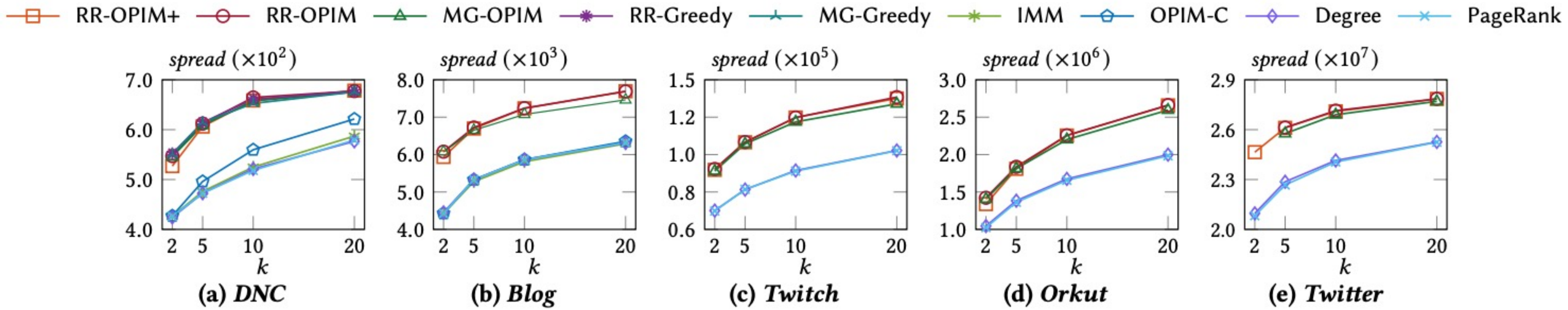
Name	**#nodes (n)**	**#edges (m)**
DNC	$0.9K$	$24.2K$
Blog	$10.3K$	$668.0K$
Twitch	$168.1K$	$13.6M$
Orkut	$3.1M$	$234.2M$
Twitter	$41.7M$	$2.9B$

Dataset statistics ($K = 10^3, M = 10^6, B = 10^9$)

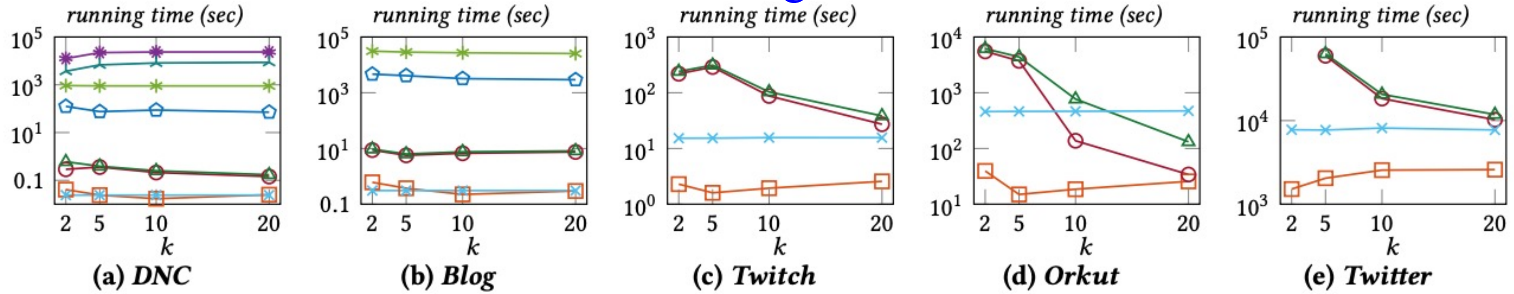
PERFORMANCE ON PUBLIC DATASETS

- Final solution RR-OPIM+ outperforms all other solutions

Large spread



Low running time



GAME DATASET WITH ACTUAL SPREAD*

- For offline evaluation: a Tencent role-playing game with
  - 243.4 thousand month-active users
  - 11.8 million corresponding friendships
  - 0.8 thousand initial adopters
  - 1.7 thousand candidate friends with ground-truth spread
- For online deployment: a Tencent battle-royale game with
  - 88.2 million quarter-active users
  - 3.2 billion corresponding friendships

*The data is collected and used under the local regulations and privacy protections

TENCENT GAME

- Common data platform: CDP 公共数据平台部
  - provide data support to game business
- Trusted by many games in Tencent

ACTUAL SPREAD ON TENCENT GAME

- Final solution RR-OPIM+ outperforms
  - all solutions during offline evaluation

Solution	RR-OPIM+	MG-OPIM	RR-OPIM	Degree	PageRank
Spread	**1,632**	1,625	1,609	1,488	1,471

  - control group during online deployment

Solution	Treatment	Control
Spread	**60.69K**	58.28K

SUMMARY

- CIM: a new problem for real-world viral marketing
- MG/RR-Greedy: effective greedy algorithms for CIM
- RR-OPIM+: scalable greedy implementation for CIM
- Have been deployed on Tencent gaming platforms

THANK YOU!

Q&A