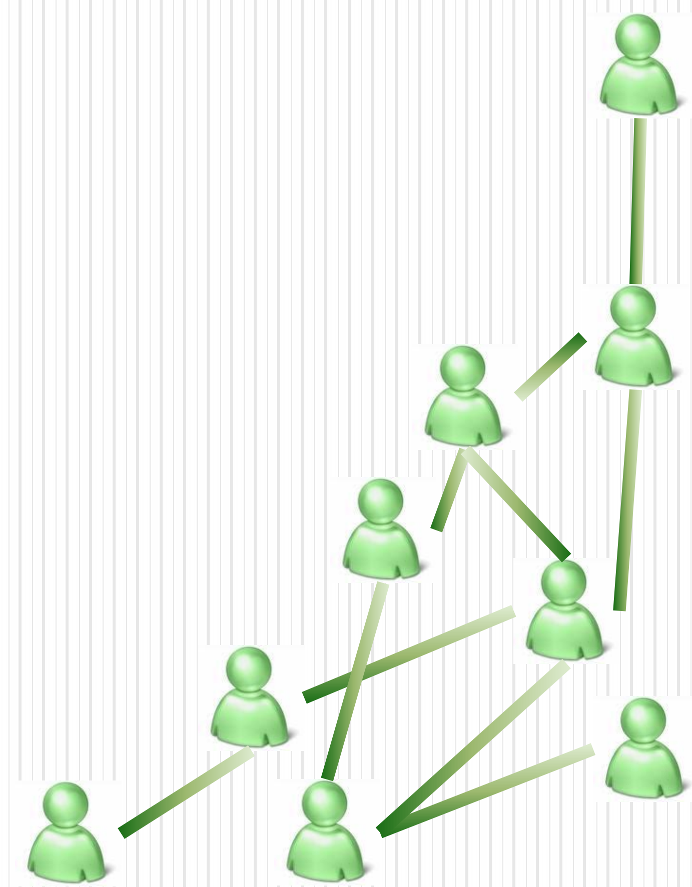
Influence diffusion dynamics and influence maximization in complex social networks

Wei Chen
陈卫
Microsoft Research Asia



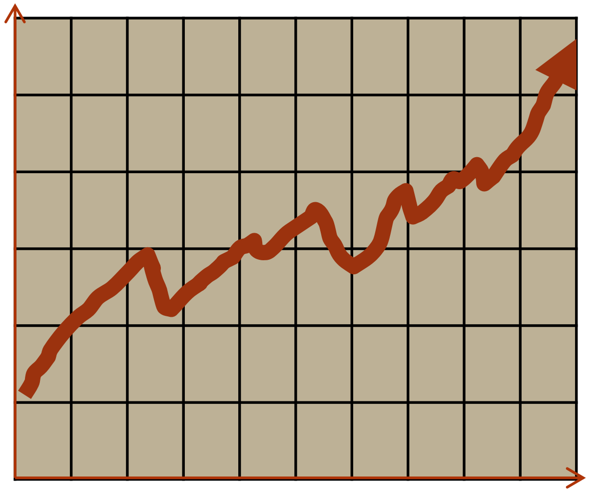
Social influence (人际影响力)

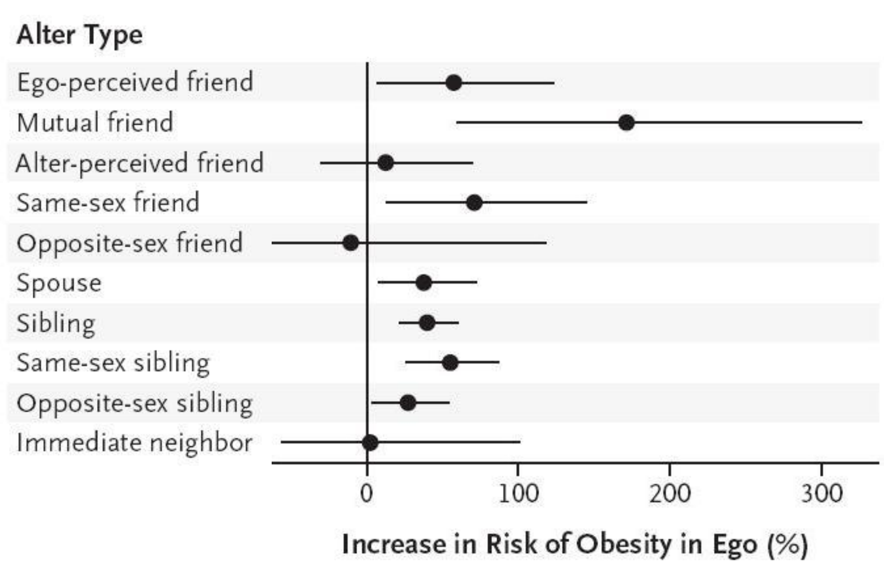
- **Social influence** occurs when one's emotions, opinions, or behaviors are affected by others.

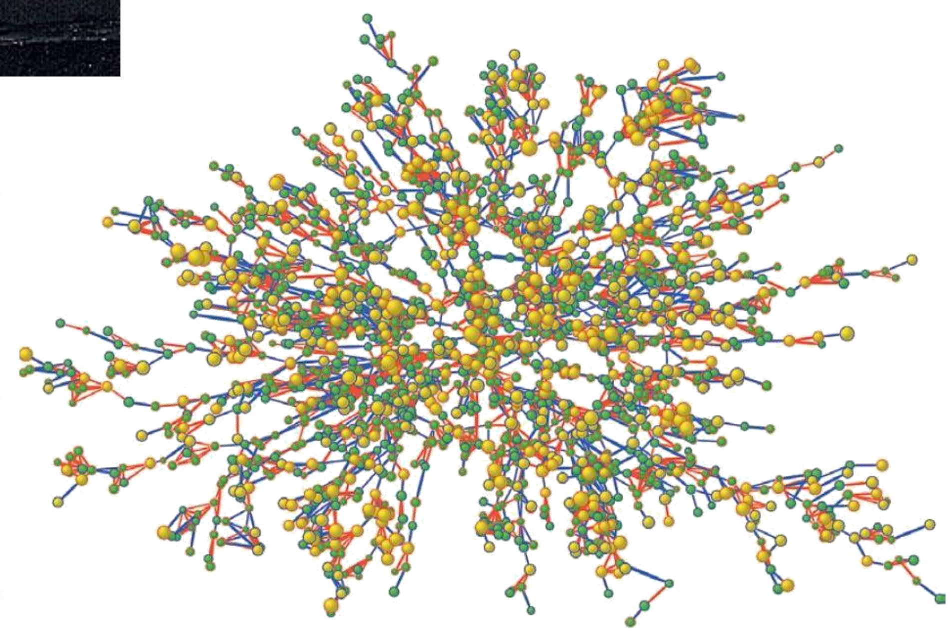
- **Social influence** is when the actions or thoughts of individual(s) are changed by other individual(s).


PSY
GANGNAM
STYLE


northern rock

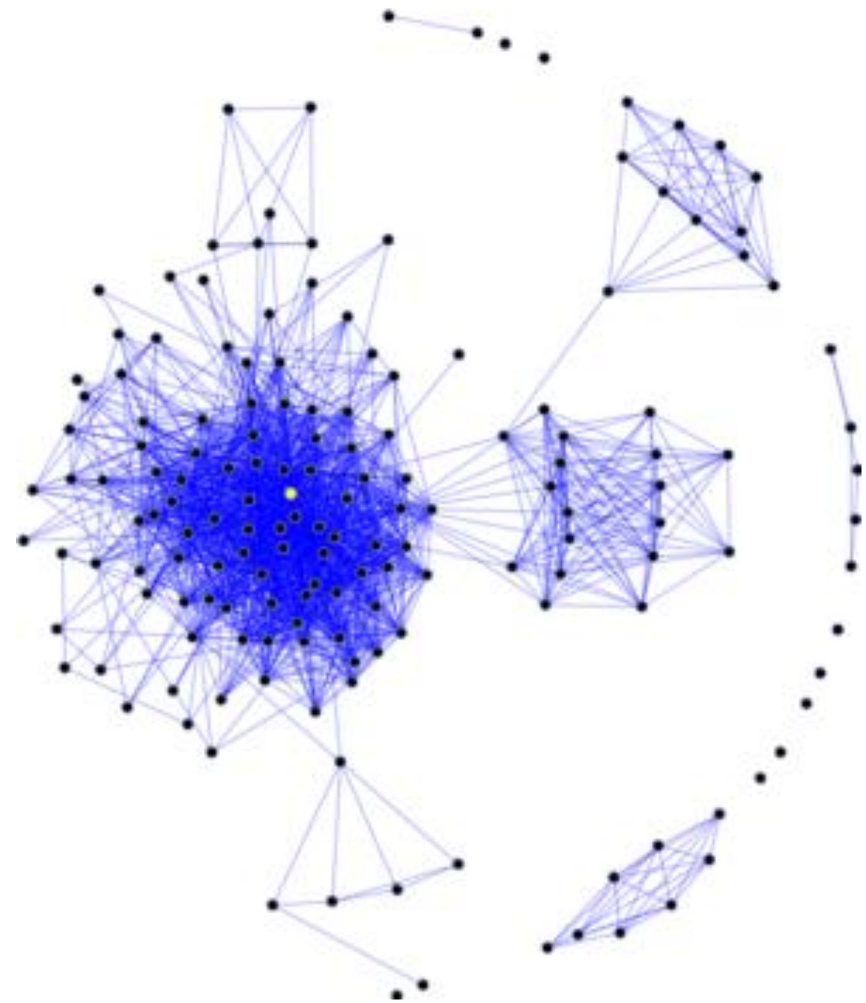






[Christakis and Fowler, NEJM'07,08]

Booming of online social networks



Opportunities on online social influence research and applications

- massive data set, real time, dynamic, open
- help social scientists to understand social interactions, influence, and their diffusion in grand scale
- help identifying influencers
- help health care, business, political, and economic decision making

Examples of recent studies

- Influential and susceptible members of social networks [Aral & Walker, Science'2012]
  - installing a facebook app and automatic notification to friends
  - men are 49% more influential than women, women influence men 46% more than they influence other women
  - younger users are more susceptible than older users
  - influential people are likely to be clustered
- Voting mobilization [Bond et al, Nature'2012]
  - show a facebook msg. on voting day with faces of friends who voted
  - generate 340K additional votes due to this message, among 60M people tested

Challenges on the research of online social networks

- Data mining and modeling: mining large-scale social network data, and building realistic models of social influence diffusion patterns, both macro level and micro level
- Algorithm: Scalable algorithm design on influence computation, influencer ranking, and influence maximization
- System: Graph-based data storage and processing, for both offline and online data analysis

Outline of this talk

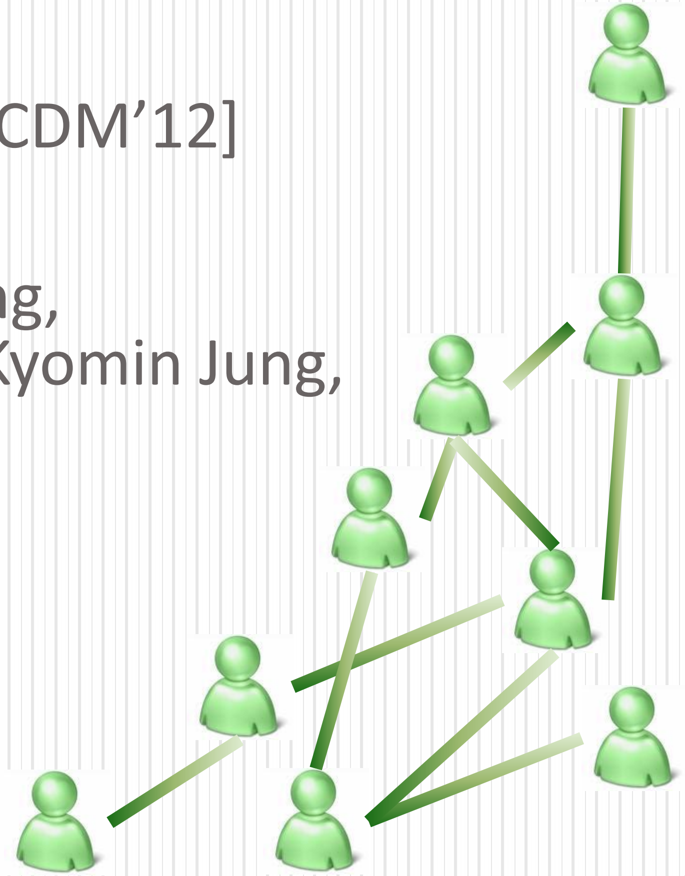
- Scalable influence maximization
- Competitive influence dynamics and influence blocking maximization

Scalable Influence Maximization in Social Networks

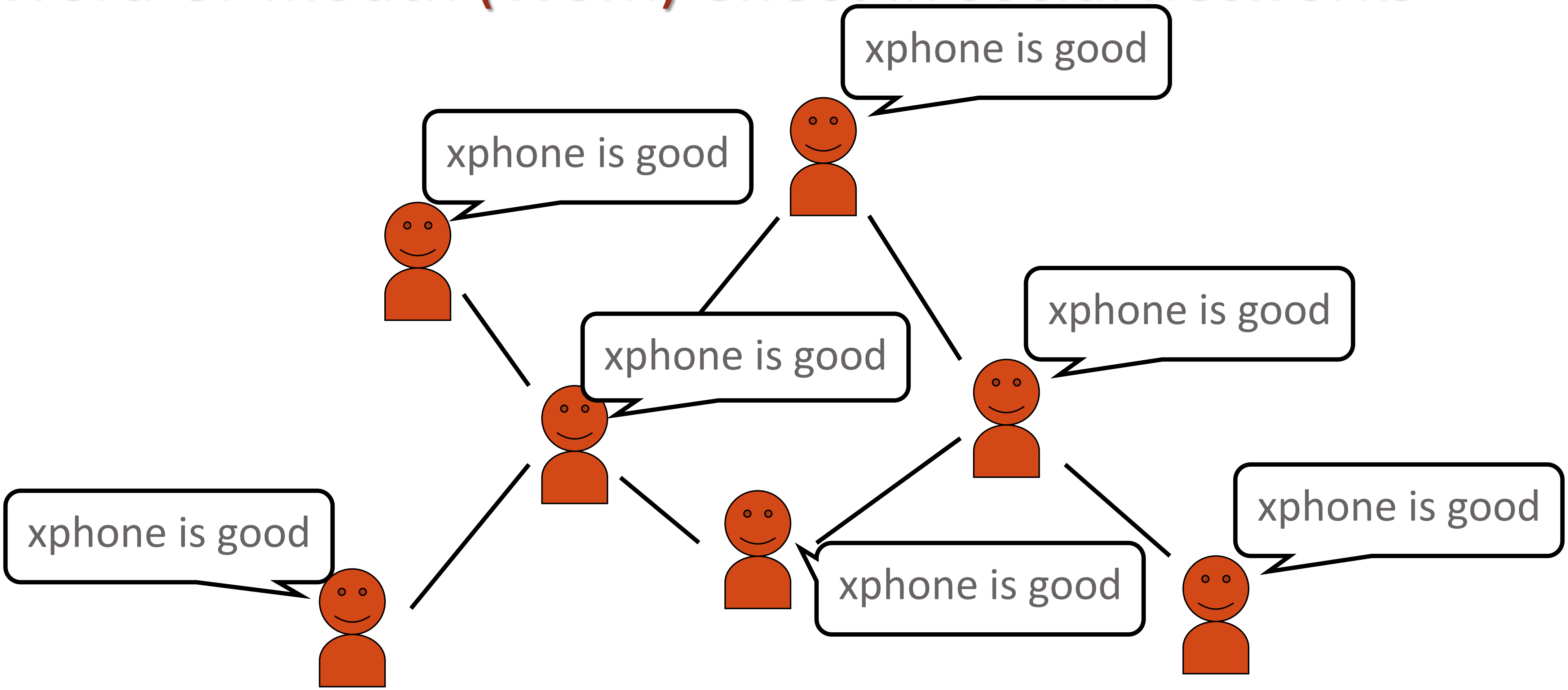
[KDD'09, KDD'10, ICDM'10, AAAI'12, ICDM'12]

Collaborators:

Yajun Wang, Siyu Yang, Chi Wang,
Yifei Yuan, Li Zhang, Wei Lu, Ning Zhang, Kyomin Jung,
Wooram Heo



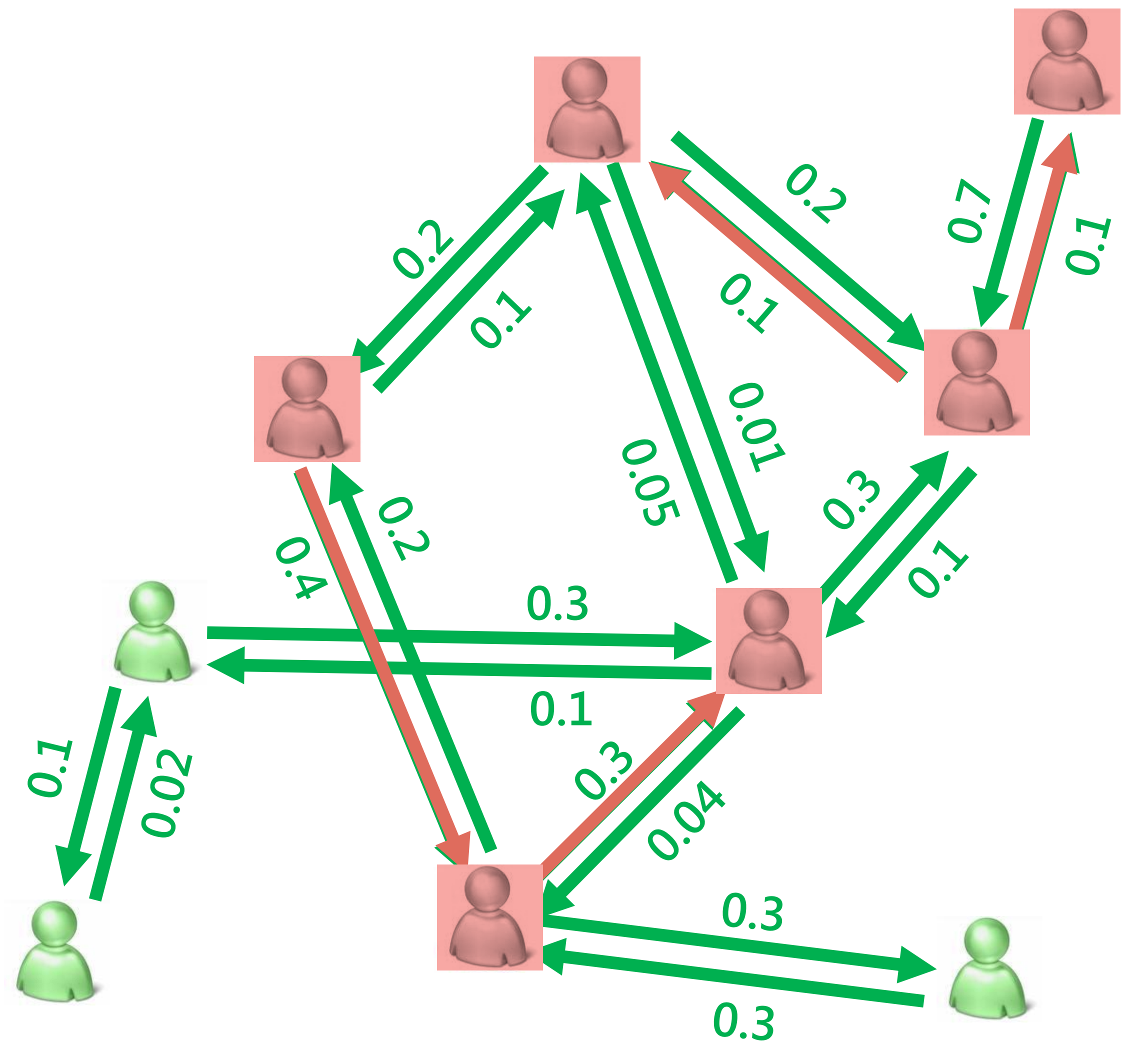
Word-of-mouth (WoM) effect in social networks



- Word-of-mouth effect is believed to be a promising advertising strategy.
- Increasing popularity of online social networks may enable large scale WoM marketing

The Problem of Influence Maximization

- Social influence graph
  - vertices are individuals
  - links are social relationships
  - number p(u,v) on a directed link from u to v is the probability that v is activated by u after u is activated
- Independent cascade model
  - initially some *seed* nodes are activated
  - At each step, each newly activated node u activates its neighbor v with probability p(u,v)
- Influence maximization:
  - find *k* seeds that generate the largest expected influence



Prior work

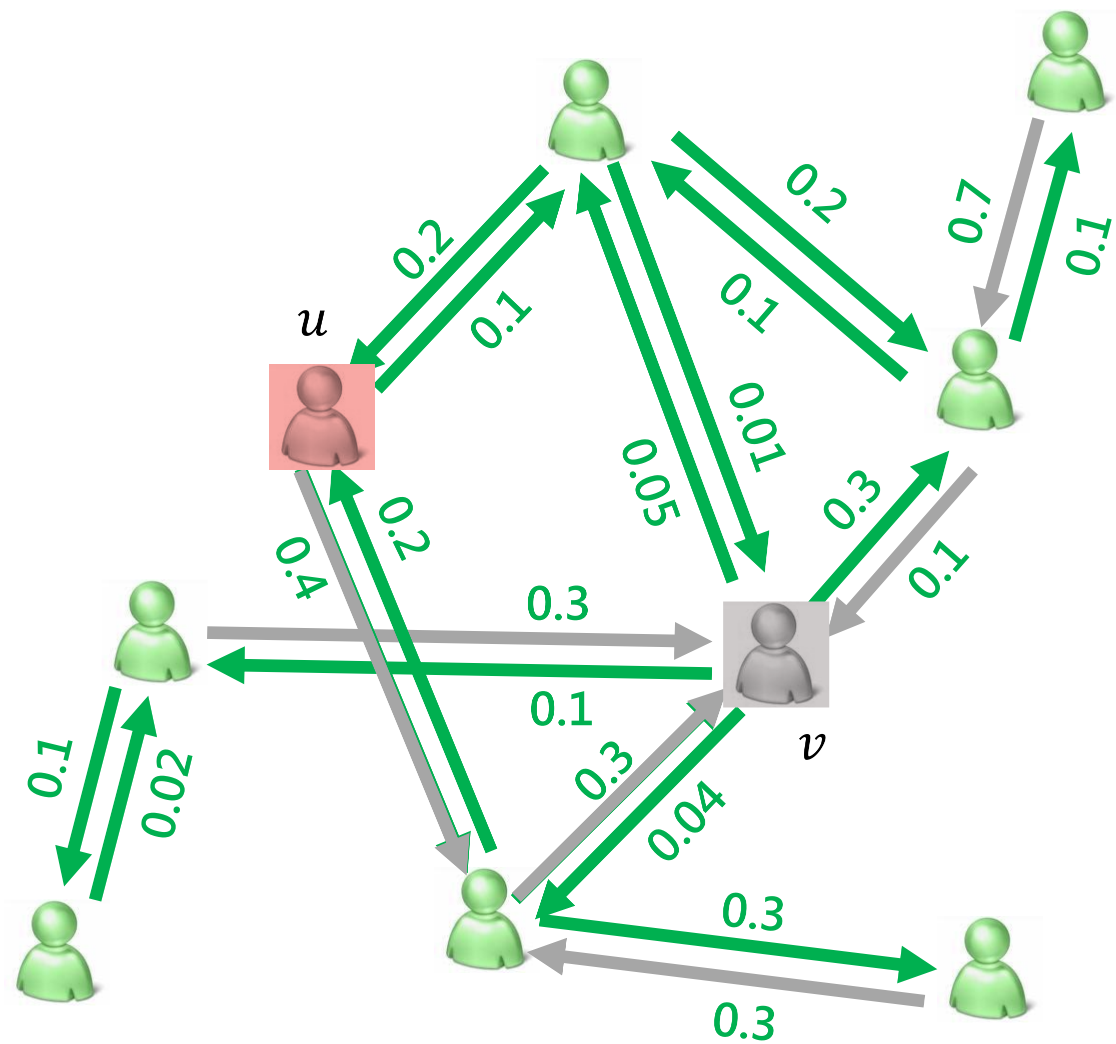
- Influence maximization as a discrete optimization problem proposed by Kempe, Kleinberg, and Tardos, 2003
  - Introduce Independent Cascade (IC) and Linear Threshold (LT) models
  - Finding optimal solution is provably hard (NP-hard)
  - Greedy approximation algorithm, 63% approximation of the optimal solution (based on submodularity)
    - select k seeds in k iterations
    - in each iteration, select one seed that provides the largest marginal increase in influence spread
- Several subsequent studies improved the running time
- Serious drawback:
  - very slow, not scalable: > 3 hrs on a 30k node graph for 50 seeds

Our work

- Exact influence computation is #P hard, for both IC and LT models --- computation bottleneck [KDD’10, ICDM’10]
- Design new heuristics
  - MIA for general IC model [KDD’10]
    - 10^3 speedup --- from hours to seconds
    - influence spread close to that of the greedy algorithm of [KKT’03]
  - Degree discount heuristic for uniform IC model [KDD’09]
    - 10^6 speedup --- from hours to milliseconds
  - LDAG for LT model [ICDM’10]
    - 10^3 speedup --- from hours to seconds
  - IRIE for IC model [ICDM’12]
    - further improvement with time and space savings
- Extend to time-critical influence maximization [AAAI’12]

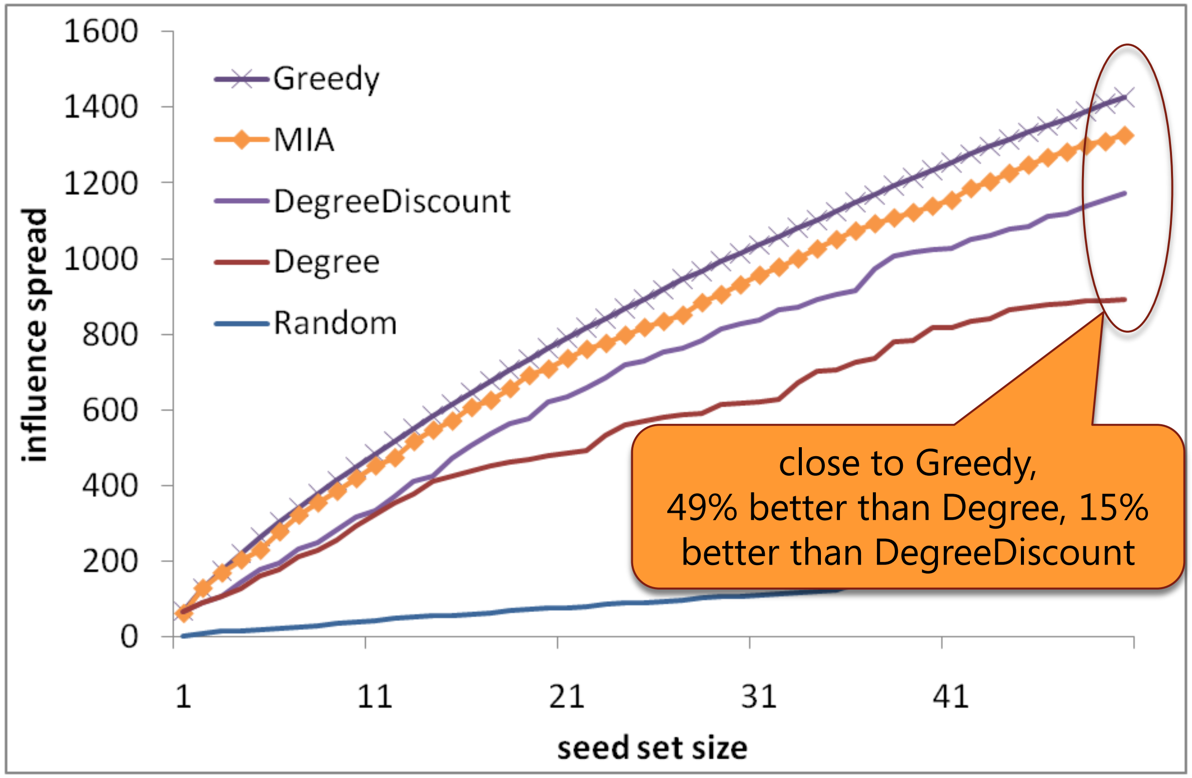
Features of Maximum Influence Arborescence (MIA) heuristic

- Based on greedy approach
- Localize computation
- Use local tree structure
  - easy to compute
- linear batch update on marginal influence spread

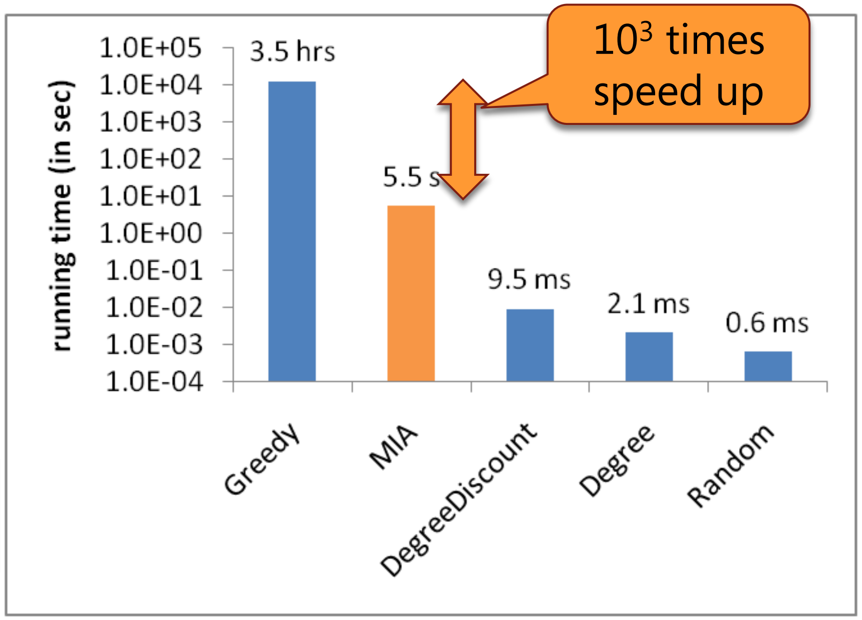


Experiment results on MIA heuristic

Influence spread vs. seed set size



running time



Experiment setup:

- 35k nodes from coauthorship graph in physics archive
- influence probability to a node v = 1 / (# of neighbors of v)
- running time is for selecting 50 seeds

Summary

- Scalable influence maximization algorithms
  - MixedGreedy and DegreeDiscount [KDD’09]
  - PMIA for the IC model [KDD’10]
  - LDAG for the LT model [ICDM’10]
  - IRIE for the IC model [ICDM’12]: further savings on time and space
  - MIA-M for IC-M model [AAAI’12]: include time delay and maximization within a short deadline
- PMIA/LDAG have become state-of-the-art benchmark algorithms for influence maximization

Competitive Influence

[SDM'11, SDM'12, others under submission]

Alex Collins, Rachel Cummings, Te Ke, Zhenming Liu, David Rincon, Xiaorui Sun, Yajun Wang, Wei Wei, Yifei Yuan, Xinran He, Guojie Song, Qingye Jiang, Yanhua Li, Zhi-Li Zhang

Competitive influence diffusion

- Exogenous competition: rival products compete for social influence in the social network
  - CLT model and CLDAG algorithm for influence blocking maximization [SDM'12]
- Endogenous competition: bad opinions about a product due to product defect competes with positive opinions
  - IC-N model and MIA-N algorithm [SDM'11]
- Influence diffusion in networks with positive and negative relationships
  - voter model in signed networks with exact inf. max. algorithm

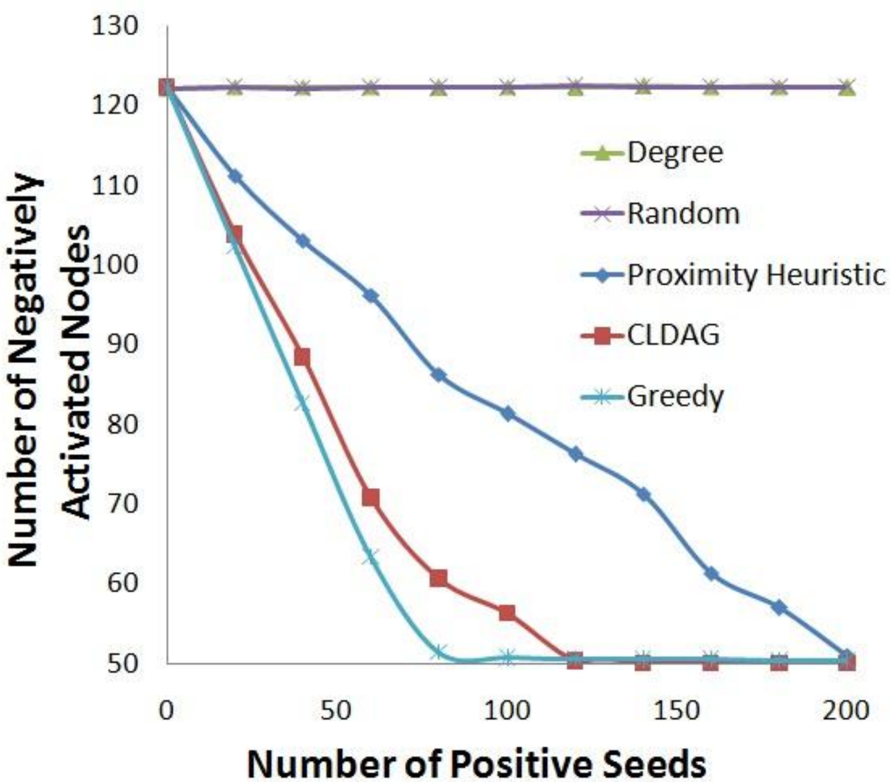
Exogenous competition

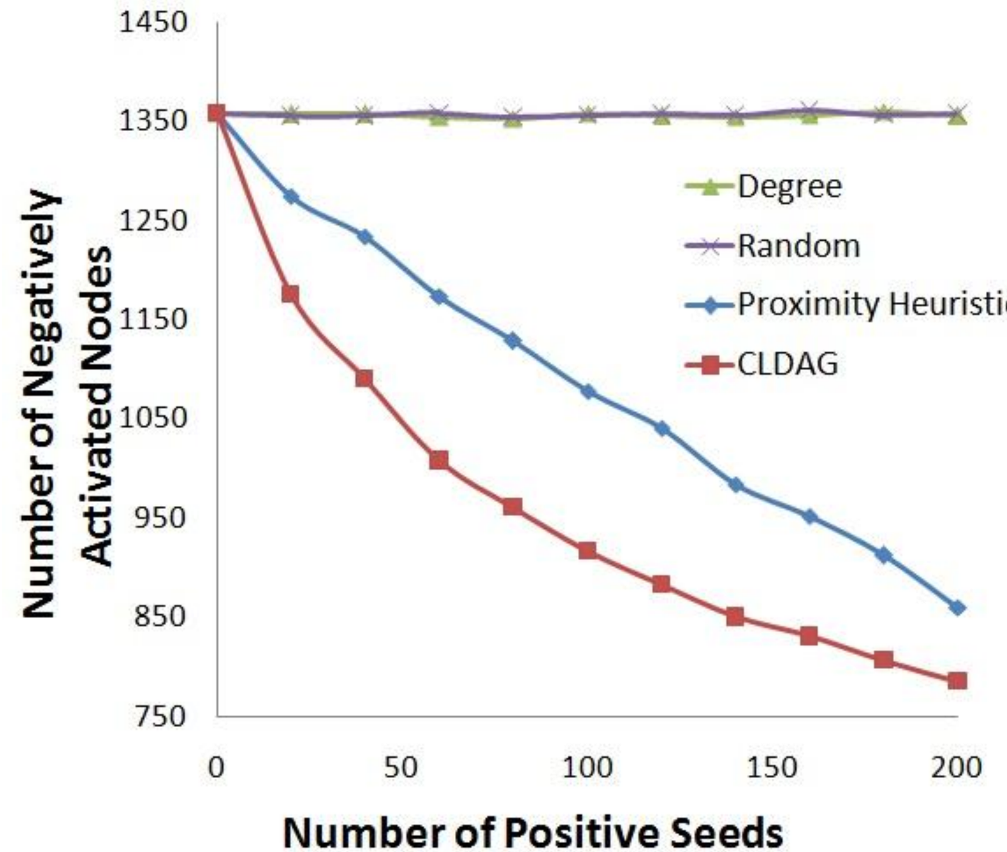
- Competitive linear threshold model
  - positive and negative influence each follows LT model
  - when competing on a node at the same step, negative influence wins with a fixed probability
- Influence blocking maximization
  - Given the negative activation status
  - find k positive seeds
  - minimize the further negative influence, or maximize the expected number of "saved" or "blocked" nodes from negative influence --- *negative influence reduction*
  - application: rumor control

Influence blocking maximization under CLT

- Negative influence reduction is submodular
- Allows greedy approximation algorithm
- Fast heuristic CLDAG:
  - reduce influence computation on local DAGs
  - use dynamic programming for LDAG computations

Performance of the CLDAG

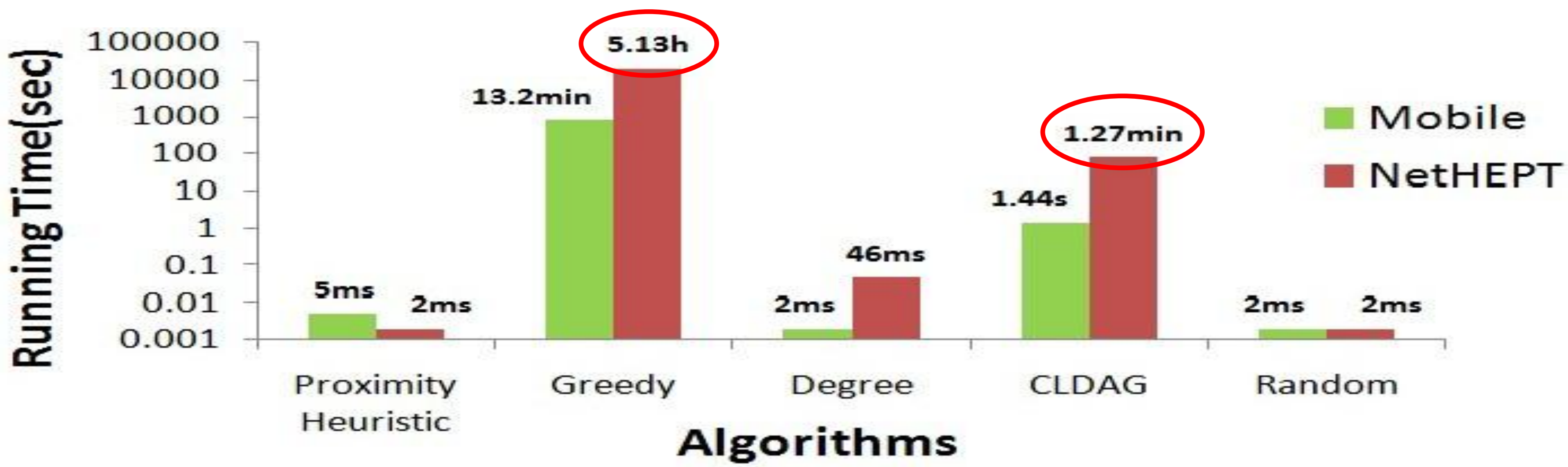




- with Greedy algorithm
- 1000 node sampled from a mobile network dataset
- 50 negative seeds with max degrees

- without Greedy algorithm
- 15K node NetHEPT, collaboration network in arxiv
- 50 negative seeds with max degrees

Scalability—Real dataset



Scalability Result for subgraph with greedy algorithm

Ongoing and future research directions

- Model validation and influence analysis from real data
- Even faster heuristic algorithms
- Fast approximation algorithms
- Online and adaptive algorithms
- Game theoretic settings for competitive diffusion
- Incentives for information / influence diffusions

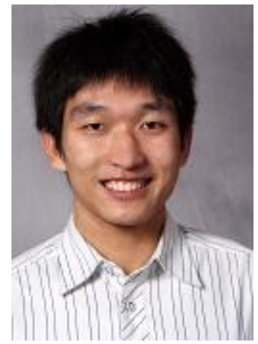
Grand challenge

- Understand the true viral diffusion scenarios, online and offline
- Apply social influence research to explain, predict, and control viral phenomena
- New focus of network science in the next decade

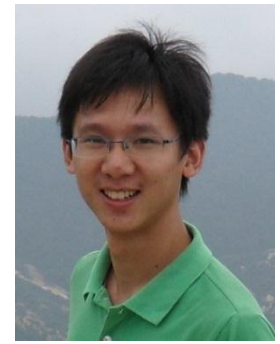
Acknowledgments

Yajun Wang
MSRA

Siyu Yang
PhD student
Princeton



Chi Wang
PhD Student
UIUC



Yifei Yuan
PhD student
UPenn

Li Zhang
MSR-SVC

Rachel Cummings
PhD student
Northwestern U.

Alex Collins
Google

Te (Tony) Ke
PhD student
UC Berkeley

Xiaorui Sun
PhD student
Columbia

Zhenming Liu
post-doc
Princeton

Acknowledgments

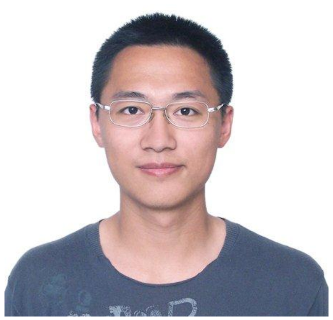
Wei Wei
PhD student
CMU

David Rincon
Universitat
Politècnica
de Catalunya

Wei Lu
PhD Student
UBC

Ning Zhang
Purdue U.

Guojie Song
Peking U.



Qingye Jiang
Master student
Columbia U.

Xinran He
PhD student
USC

Yanhua Li
PhD student
U. of Minnisoda

Zhili Zhang
U. of Minnisoda

Kyomin Jung
KAIST

Questions?

Additional materials on my homepage:

Search “Wei Chen Microsoft”

- KDD’12 tutorial on influence spread in social networks
- my papers