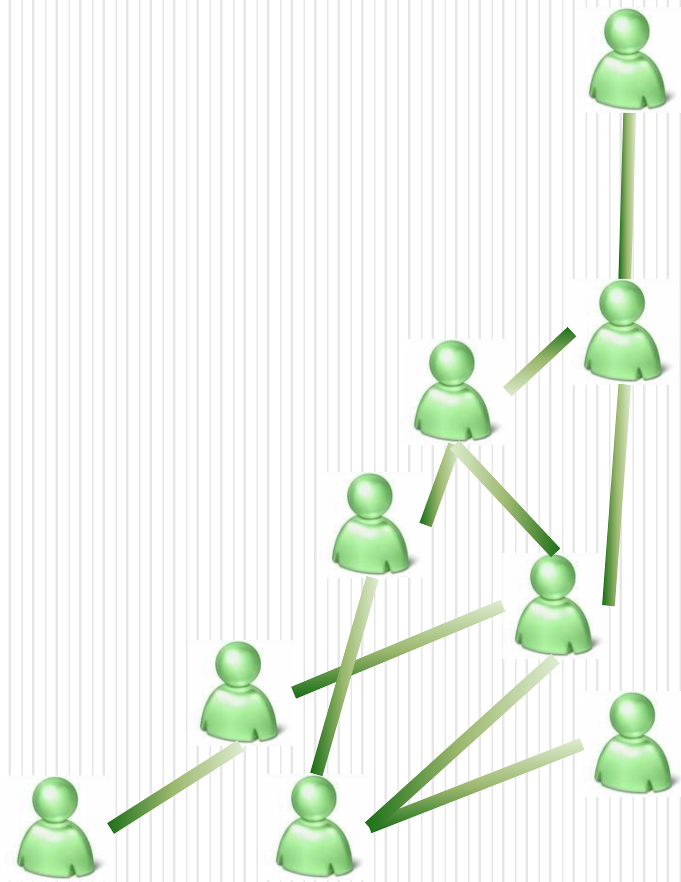
Influence Diffusion Modeling and Optimizations in Social Networks

Wei Chen
陈卫
Microsoft Research Asia



Math Programming Conference, May 11, 2014

Social influence (人际影响力)

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

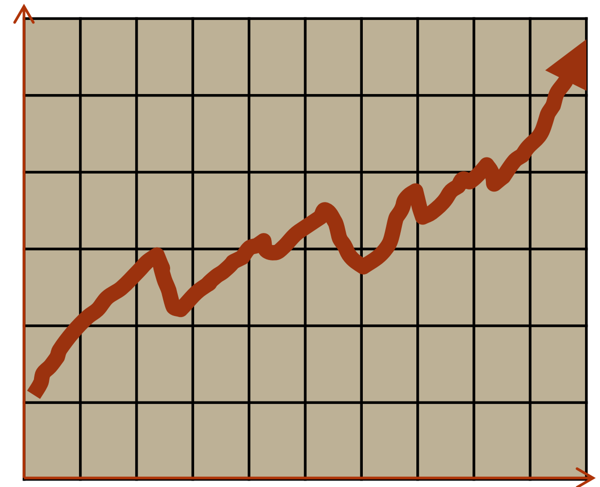
WIKIPEDIA
The Free Encyclopedia

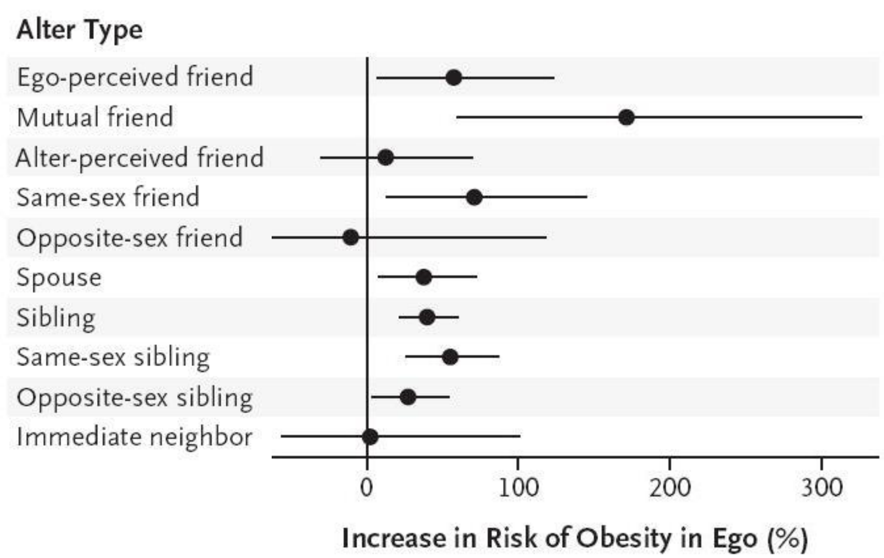

싸이6甲
PSY
GANGNAM
STYLE

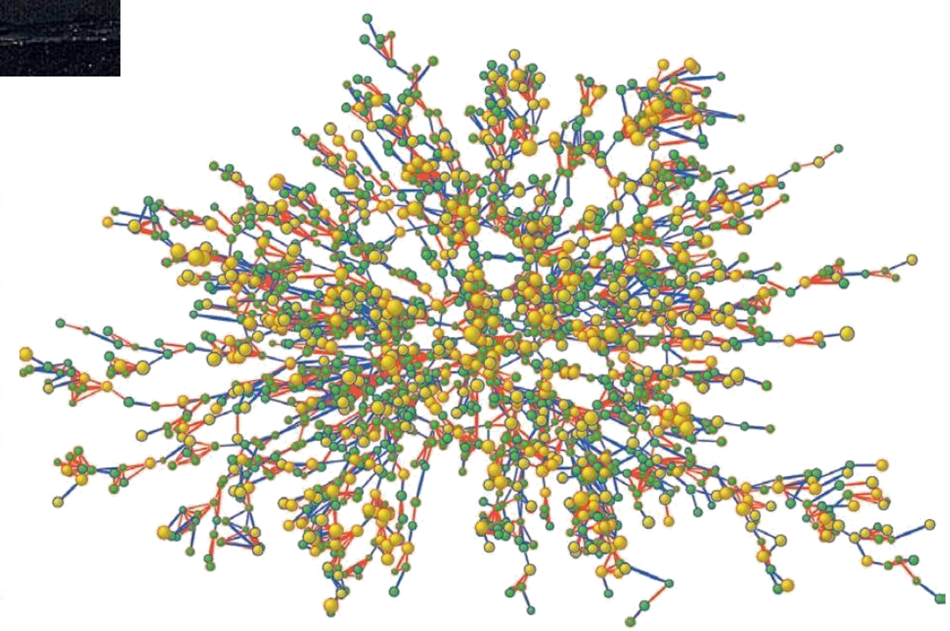

별에서 온
그대


爸爸
去哪儿
Where
Are We
Going?
Dad


northern rock

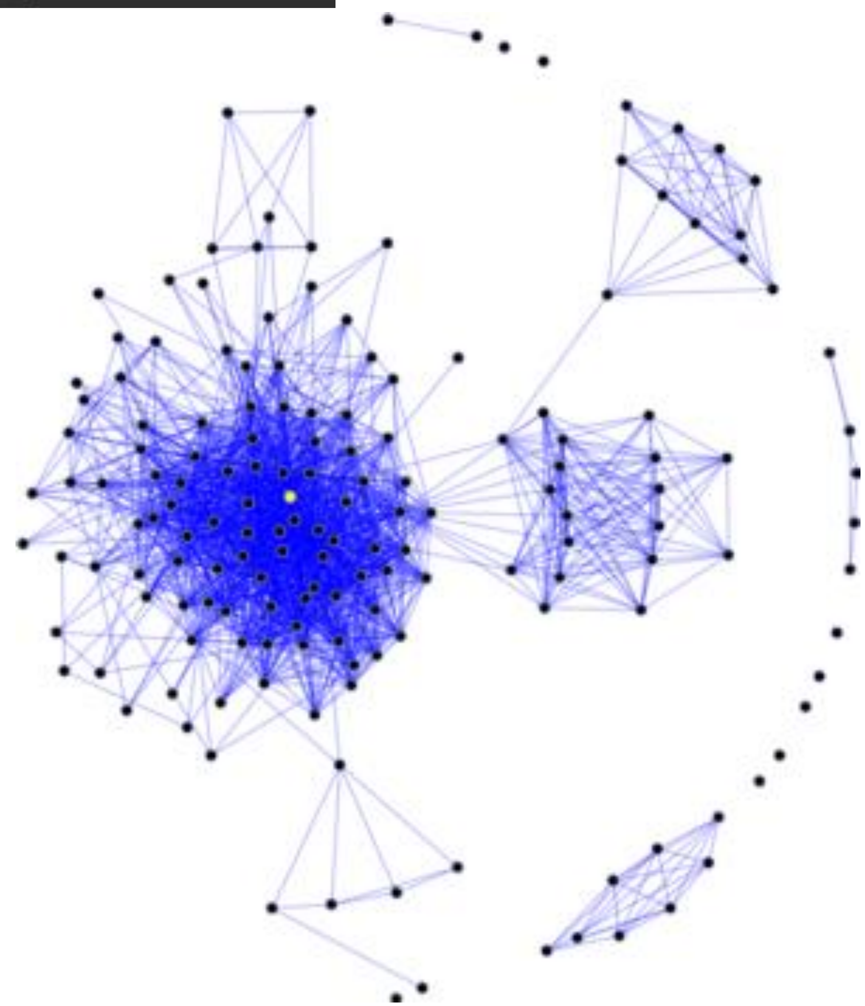






[Christakis and Fowler, NEJM’07,08]

Booming of online social networks

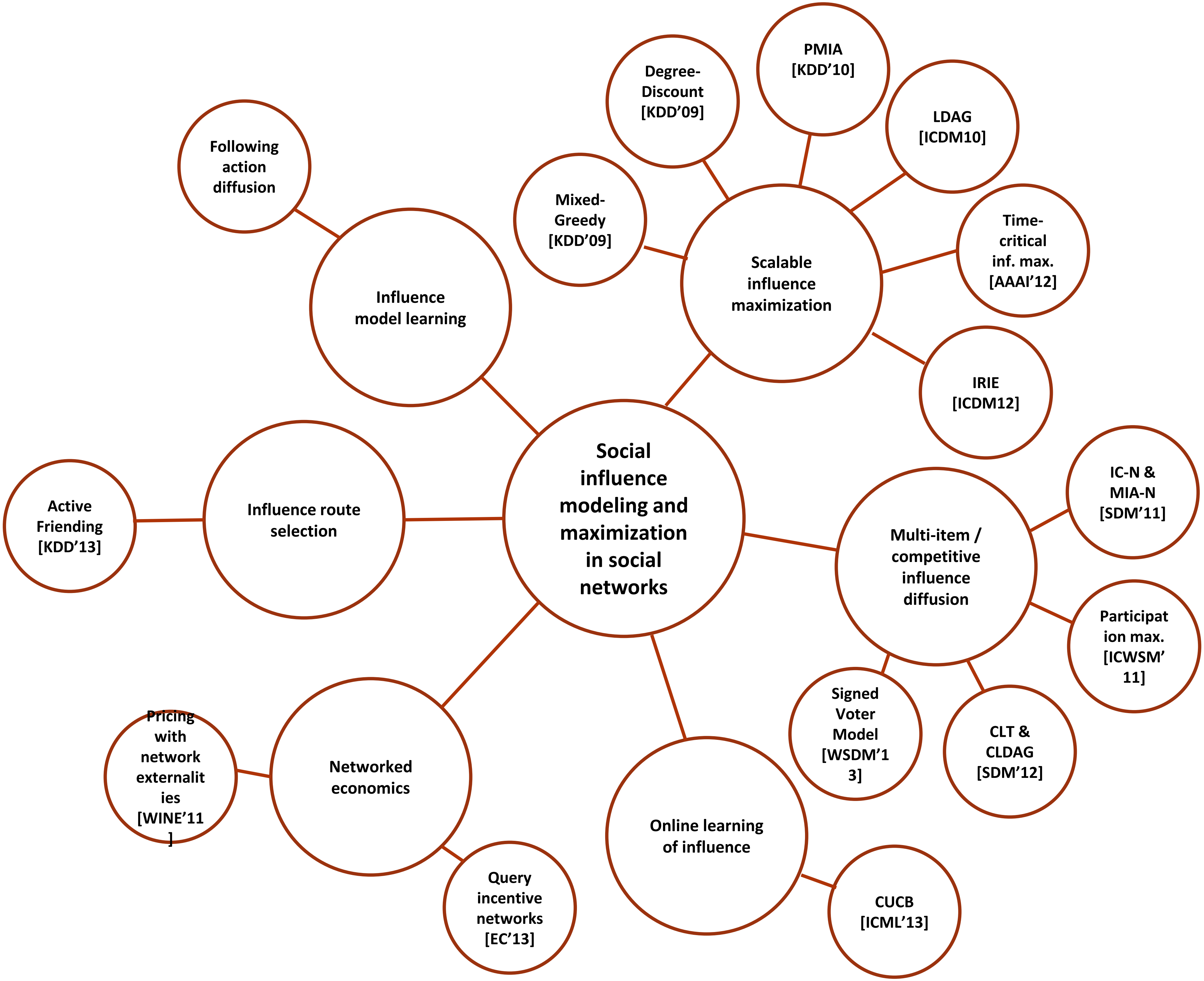


Opportunities on online social influence research and applications

- massive data set, real time, dynamic, open
- help social scientists to understand social interactions, influence, and their diffusion in grand scale
- help providing new social network services (e.g. identifying influencers)
- help health care, business, political, and economic decision making

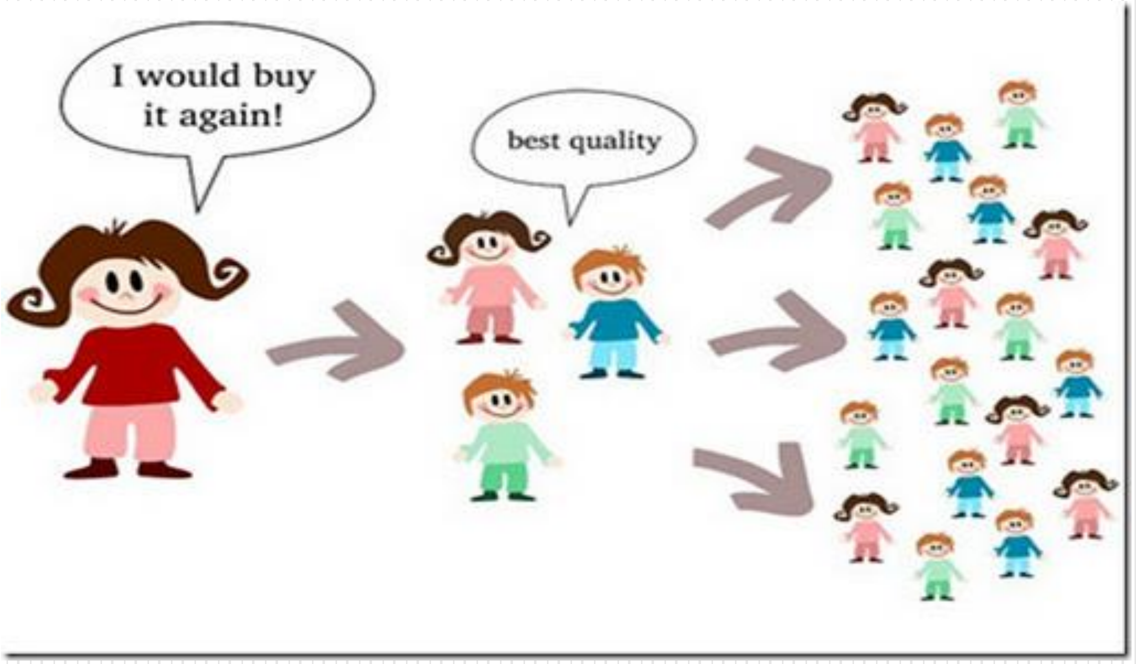
Computational Social Influence Research

- How does information/influence propagate in social networks?
  - Influence diffusion modeling
- How to learn the strength of influence relationship?
  - Model learning from real data
- How to benefit from information/influence diffusion?
  - Viral marketing: maximize influence of a new technology
  - Effective competition: minimize the influence of the competitor or negative influence
  - Crowd sourcing: mobilize crowds (perhaps with incentives)
  - Targeted influence: increase the influence to a target individual or a group of individuals

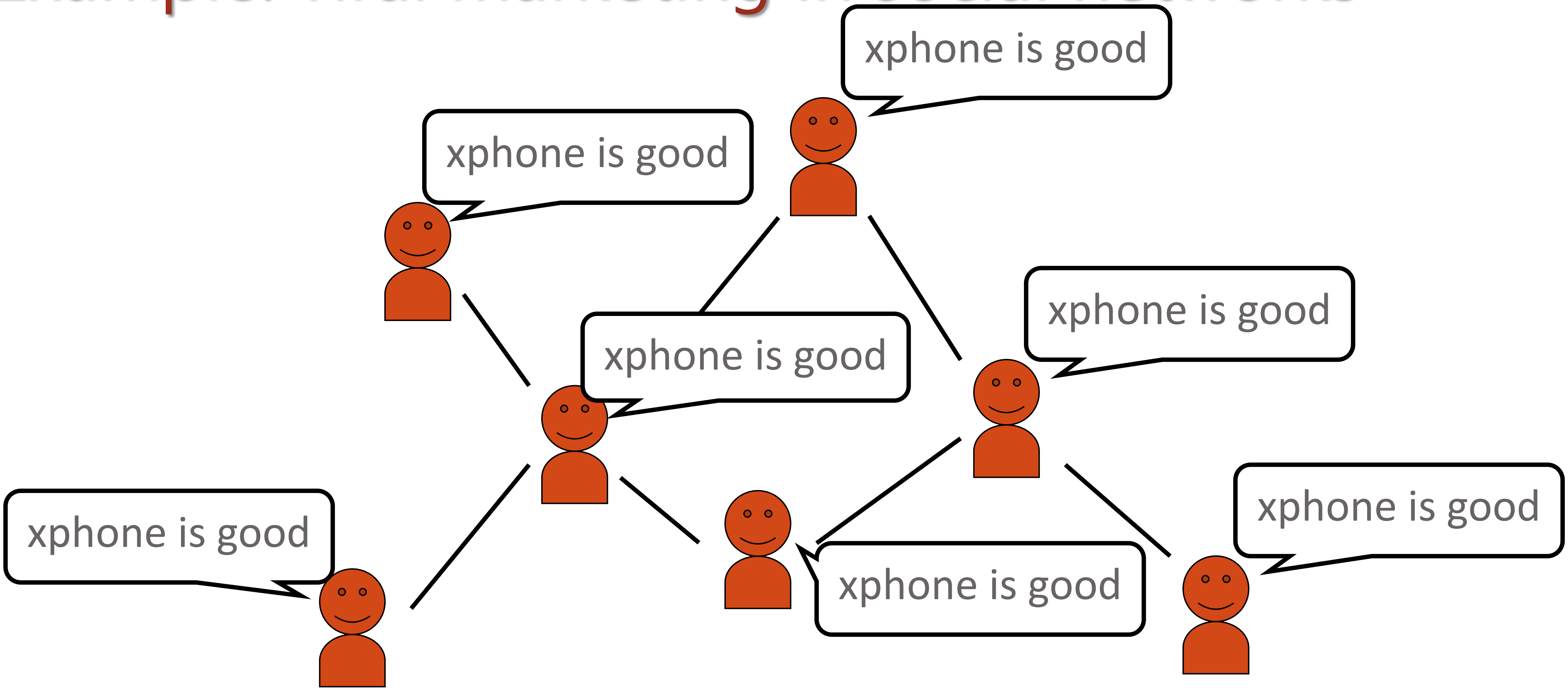

Social influence modeling and maximization in social networks
Influence model learning
Following action diffusion
Scalable influence maximization
Mixed-Greedy [KDD'09]
Degree-Discount [KDD'09]
PMIA [KDD'10]
LDAG [ICDM10]
Time-critical inf. max. [AAAI'12]
IRIE [ICDM12]
Influence route selection
Active Friending [KDD'13]
Multi-item / competitive influence diffusion
IC-N & MIA-N [SDM'11]
Participation max. [ICWSM'11]
Signed Voter Model [WSDM'13]
CLT & CLDAG [SDM'12]
Networked economics
Pricing with network externalities [WINE'11]
Query incentive networks [EC'13]
Online learning of influence
CUCB [ICML'13]

Influence Maximization

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



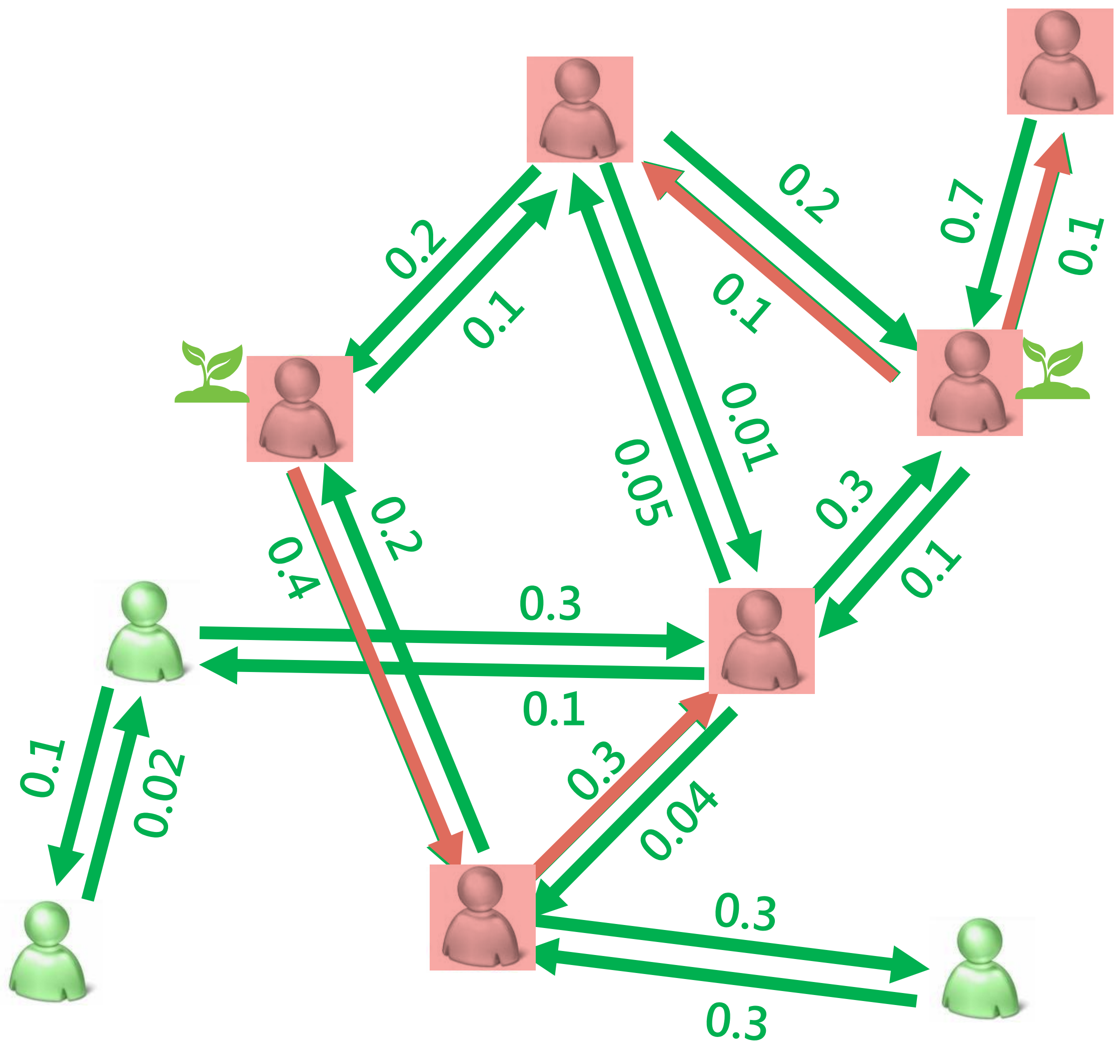
Example: viral marketing in social networks



- Viral (word-of-mouth) marketing is believed to be a promising advertising strategy.
- Increasing popularity of online social networks may enable large scale viral marketing

Influence diffusion model

- Directed graph $G = (V, E)$
  - V: set of nodes, representing users
  - E: set of directed edges, representing influence relationships
- Influence probabilities on edges
  - $p(u, v)$: the probability that u activates v
- Independent cascade model
  - Initially nodes in a seed set S are activated
  - At step t, each node u activated at step $t - 1$ has one chance to activate each of its out-going neighbor v, with success probability $p(u, v)$



Influence maximization

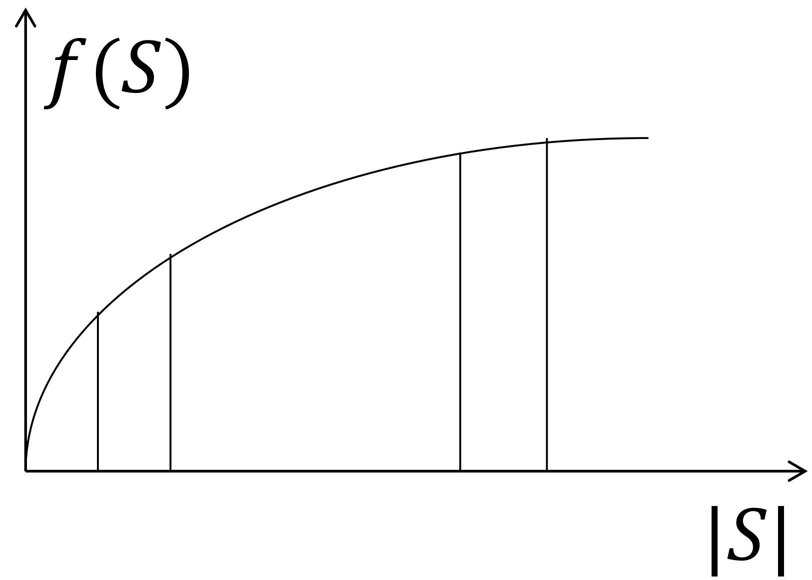
- Given a social network, a diffusion model with given parameters, and a number k, find a seed set S of at most k nodes such that the influence spread of S is maximized.
- May be further generalized:
  - Instead of k, given a budget constraint and each node has a cost of being selected as a seed
  - Instead of maximizing influence spread, maximizing a (submodular) function of the set of activated nodes

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)
- Several subsequent studies improved the running time
- Serious drawback:
  - very slow, not scalable: > 3 hrs on a 30k node graph for 50 seeds

Optimizing submodular functions

- **Sumodularity** of set functions $f: 2^{V} \to R$
  - for all $S \subseteq T \subseteq V$, all $v \in V \setminus T$, $f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$
  - diminishing marginal return
- **Monotonicity** of set functions f: for all $S \subseteq T \subseteq V$, $f(S) \leq f(T)$
- Submodular function maximization
  - Greedy algorithm
  - $1 - 1/e$ approximiation

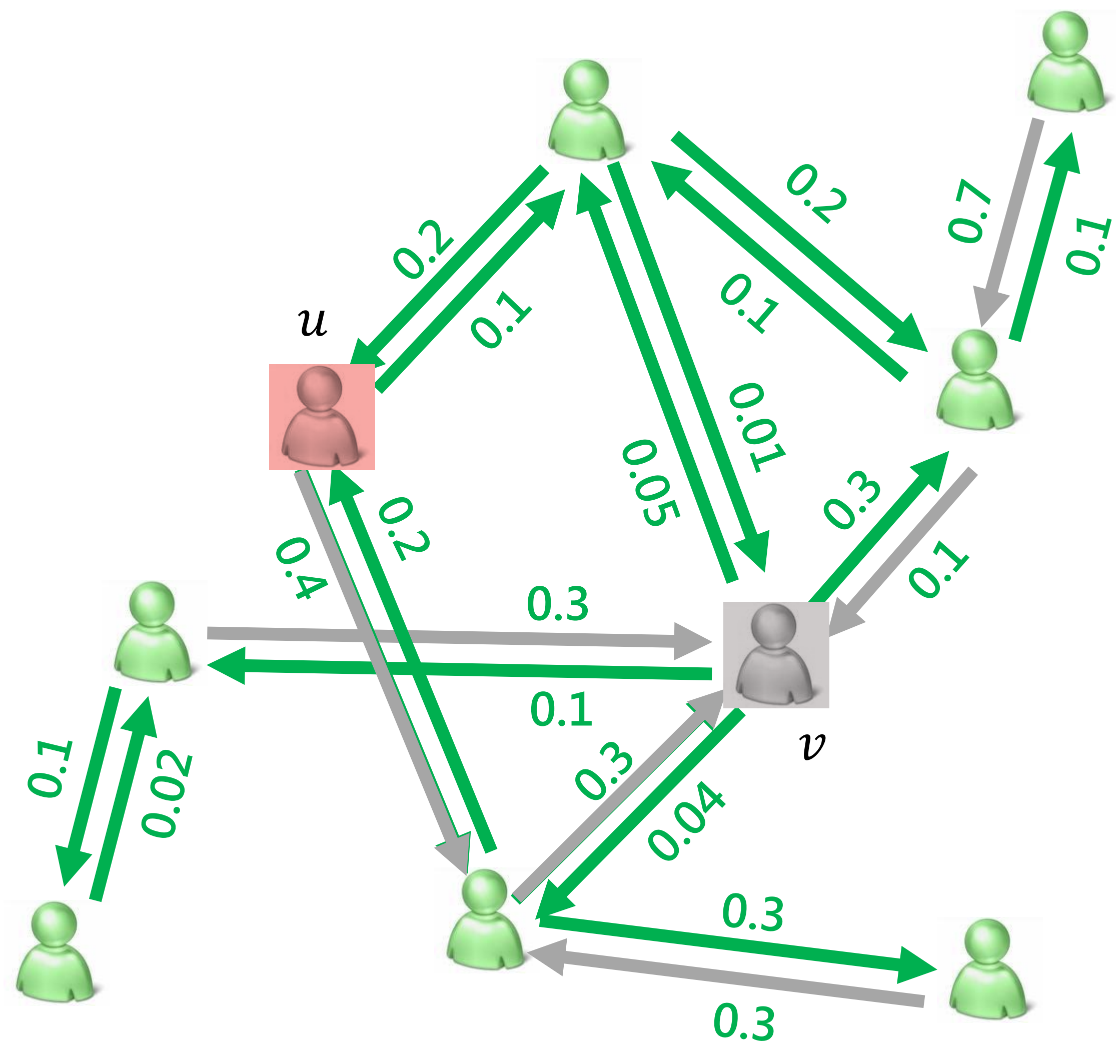


Our work

- 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]

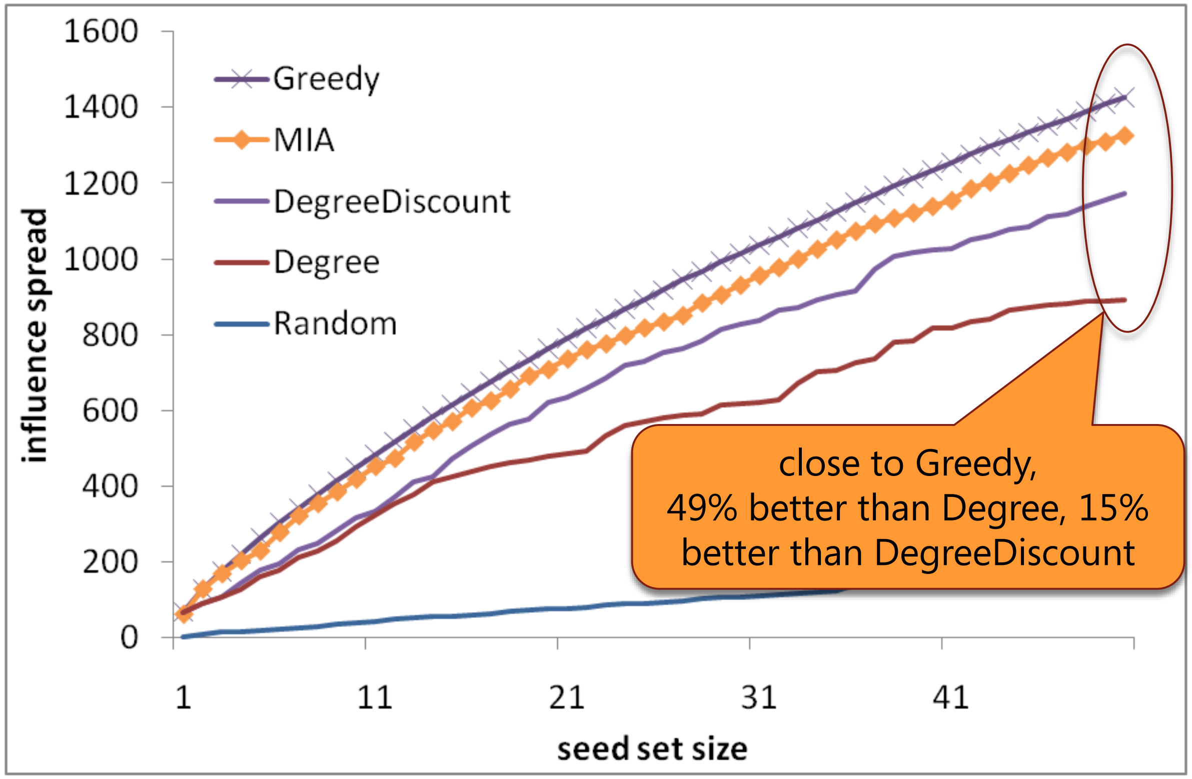
Features of Maximum Influence Arborescence (MIA) heuristic

- Based on greedy approach
- Use local tree structure
  - Dijkstra shortest path alg. for tree creation
  - Recursive computation for influence computation on trees
- linear batch update on marginal influence spread

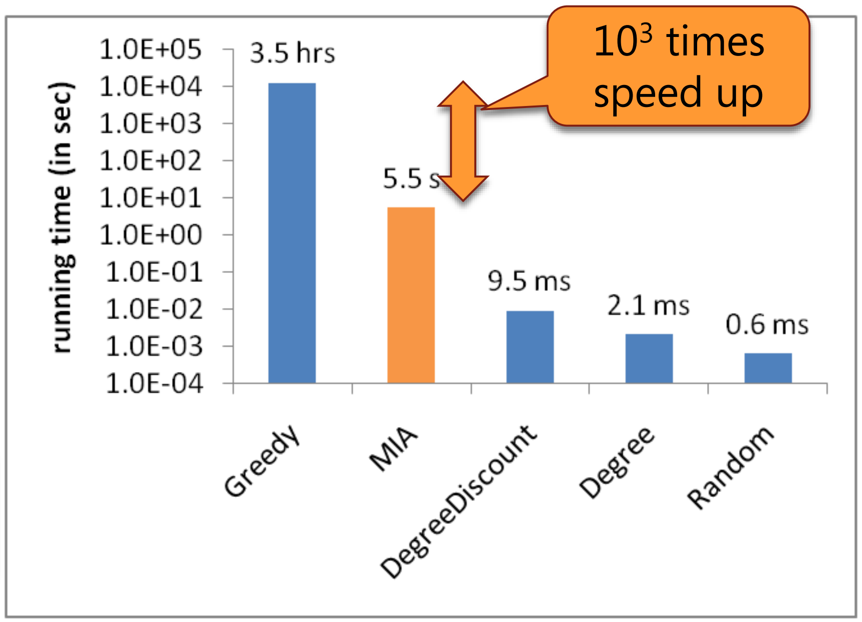


Experiment results on MIA heuristic

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

Influence Blocking Maximization in Competitive Diffusion Models

[SDM'12]

Motivation

- Multiple source of influence are propagating in networks
  - Often competitive
- Question:
  - How to maximize my own influence in face of competition?
  - How to minimize the competitor's influence?
    - E.g. stop rumor diffusion

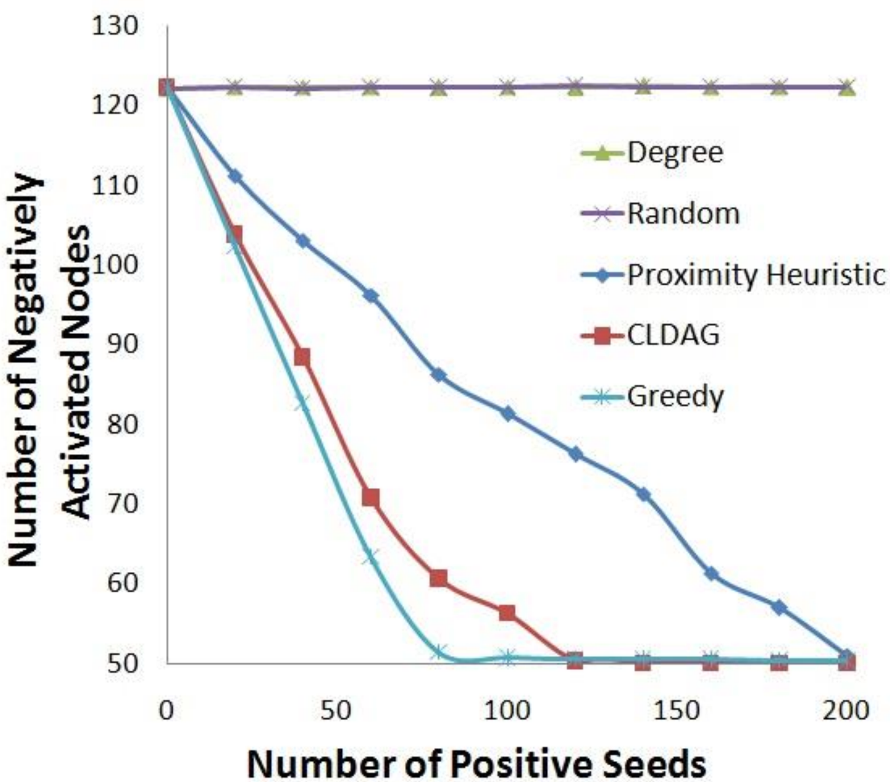
Influence Blocking Maximization

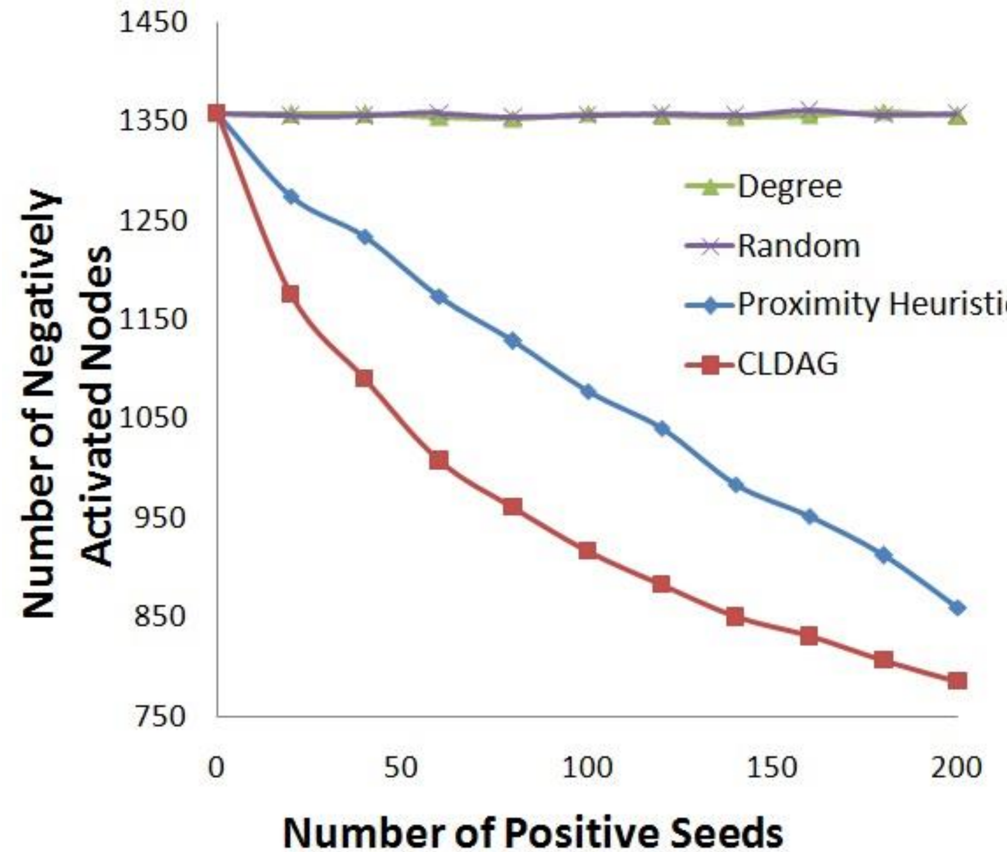
- Competitive linear threshold model
  - positive and negative influence each follows LT model
  - when competing on a node at the same step, negative influence wins with a fixed probability
- Influence blocking maximization
  - Given the negative activation status
  - find k positive seeds
  - minimize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

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

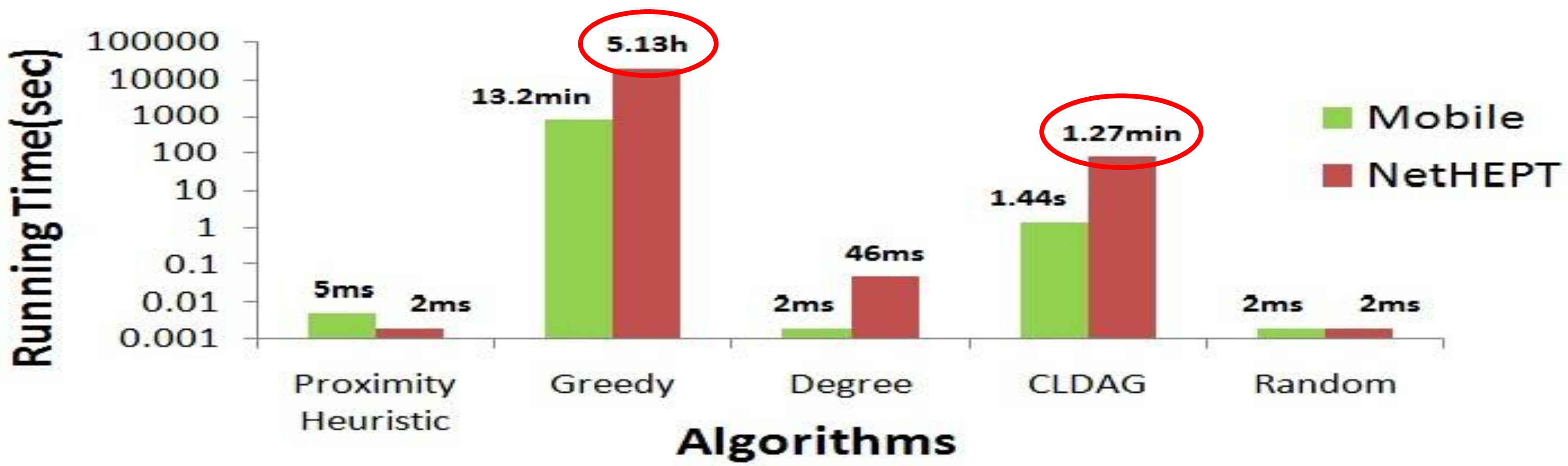




- with Greedy algorithm
- 1000 node sampled from a mobile network dataset
- 50 negative seeds with max degrees

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

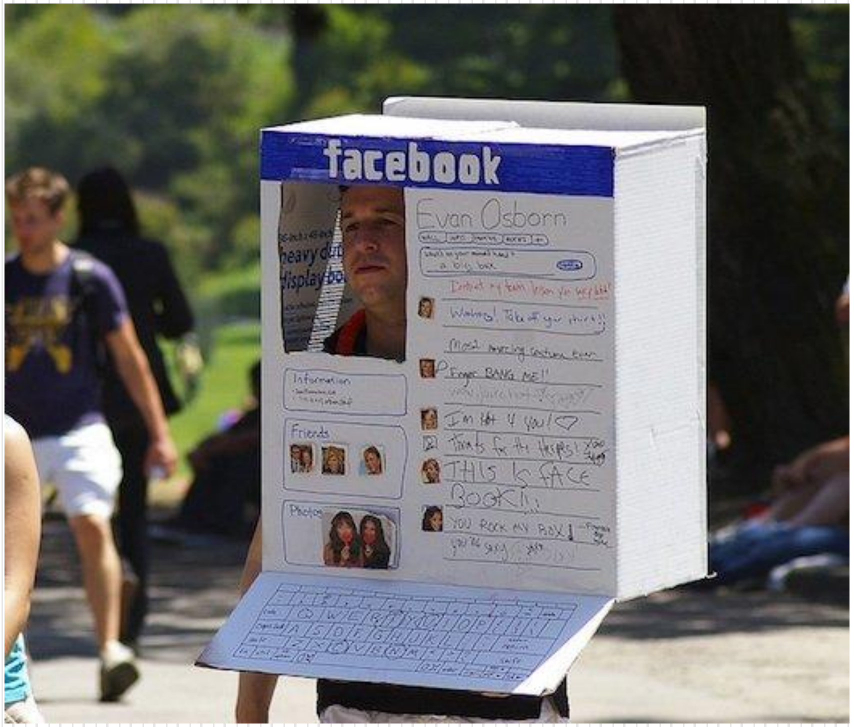
Scalability—Real dataset



Scalability Result for subgraph with greedy algorithm

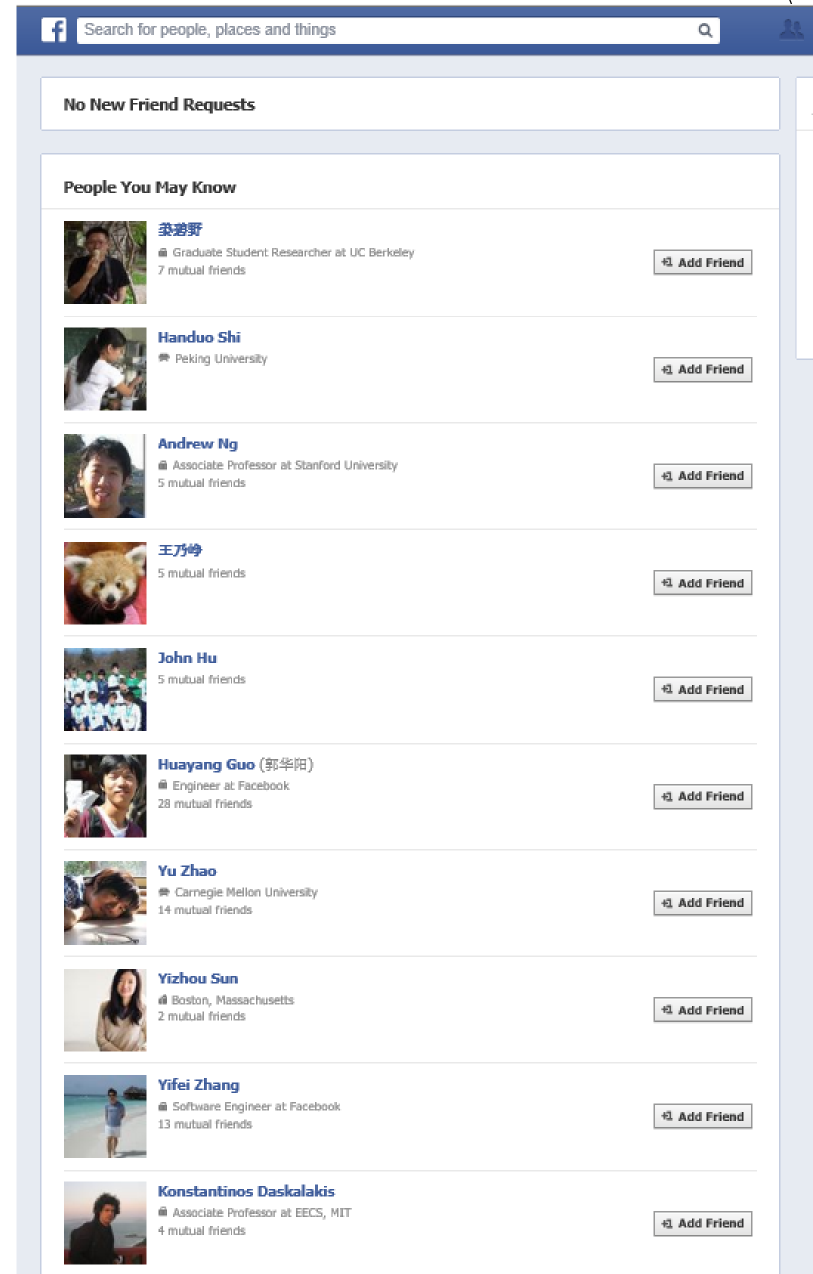
Active Friending

[KDD’13]



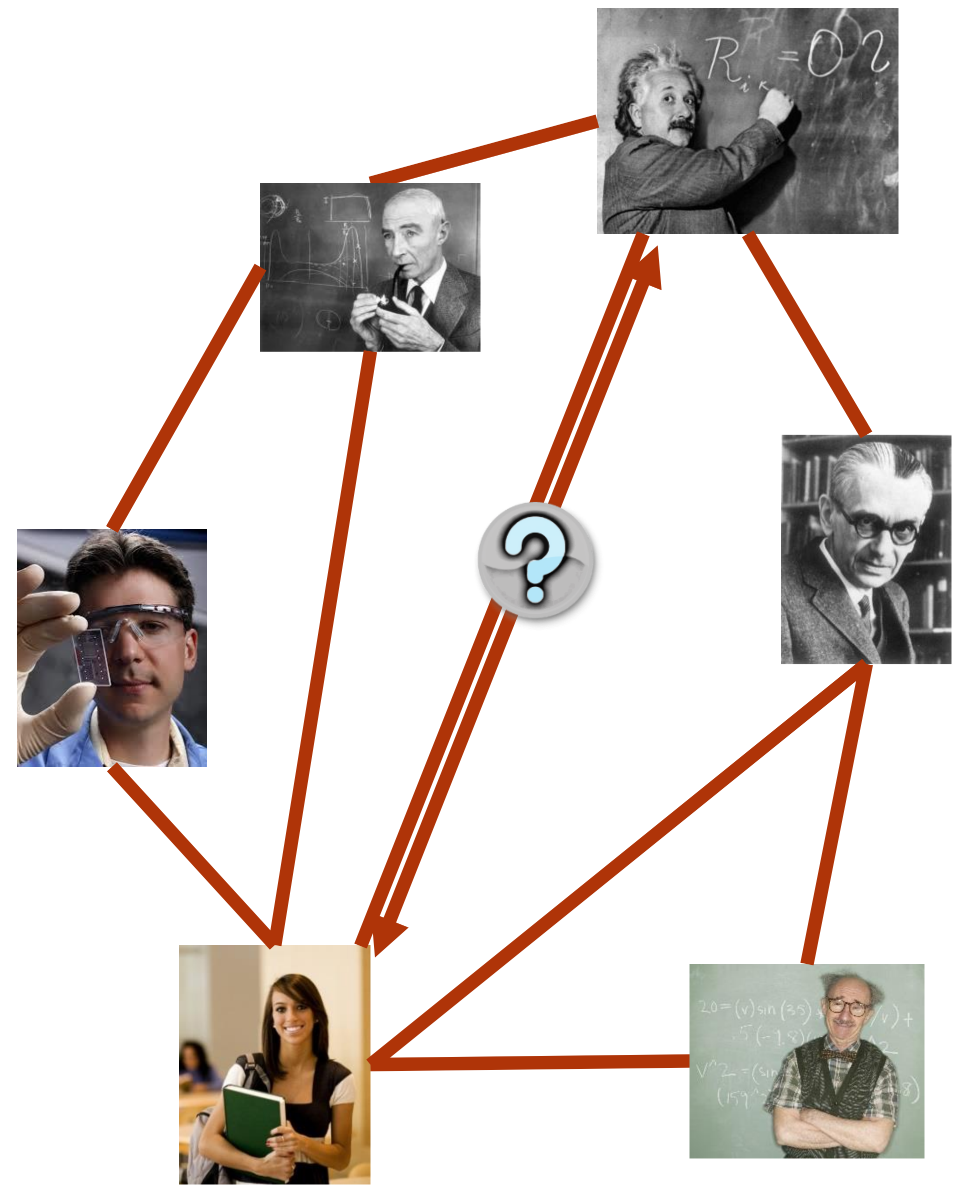
Passive friending

- A user is given a list of friend recommendations --- “people you may know”
- Select people from the list to send friending request
- Many online social networking sites provide this service



Active friending

- A user wants to friend with a particular person
- How to increase the chance of the target person accepting the friending request?
- No service exists today

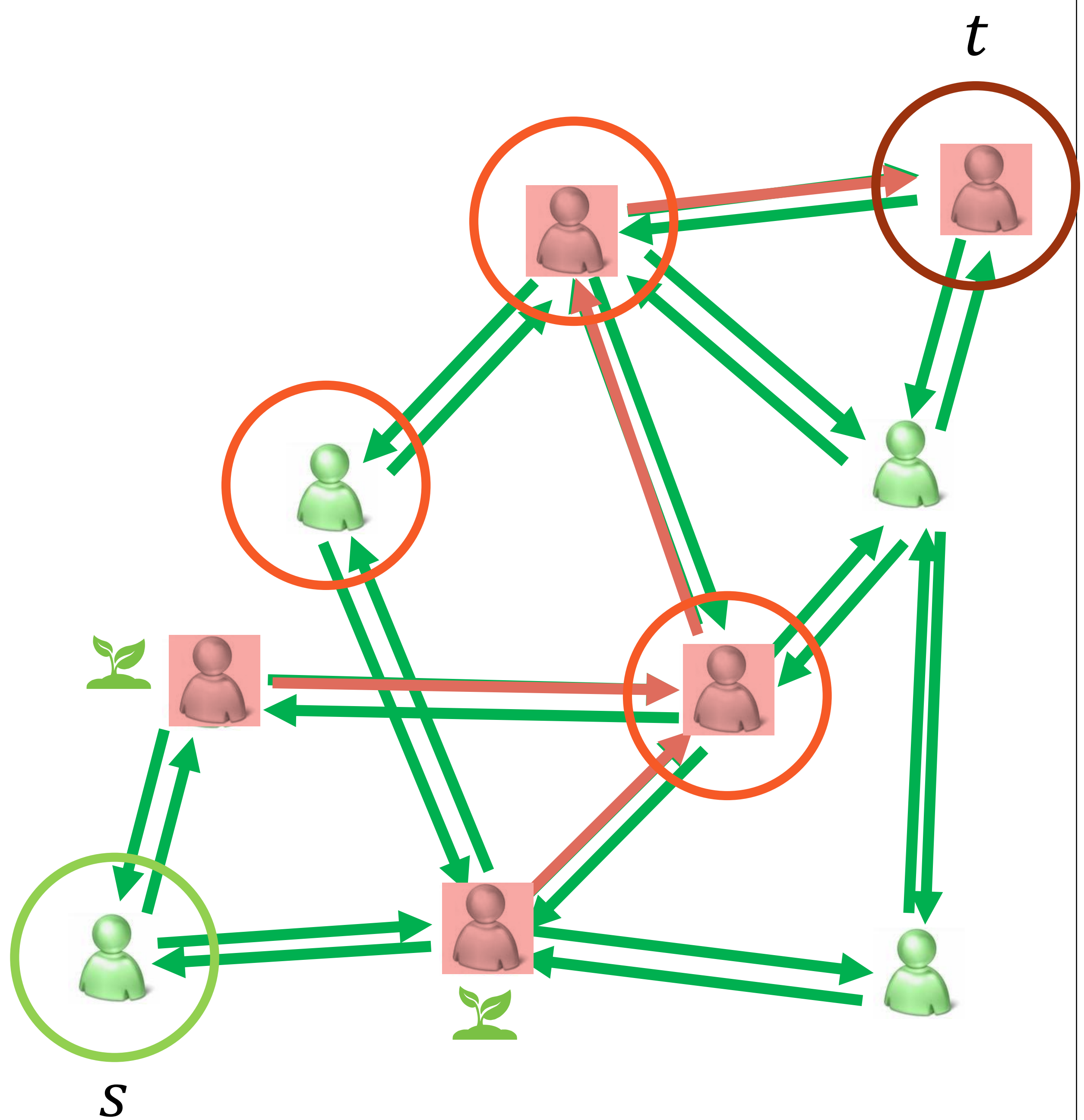


Acceptance Probability Maximization (APM) for active friending

- Given
  - Graph $G = (V, E)$ with influence probabilities
  - Initiator s, target t
  - Budget r_R
- Select r_R non-friend nodes R (including t) s.t.
  - The out-neighbors of s is the seed set
  - Influence diffusion only passes through nodes in R
  - Need to maximize the probability of t being activated

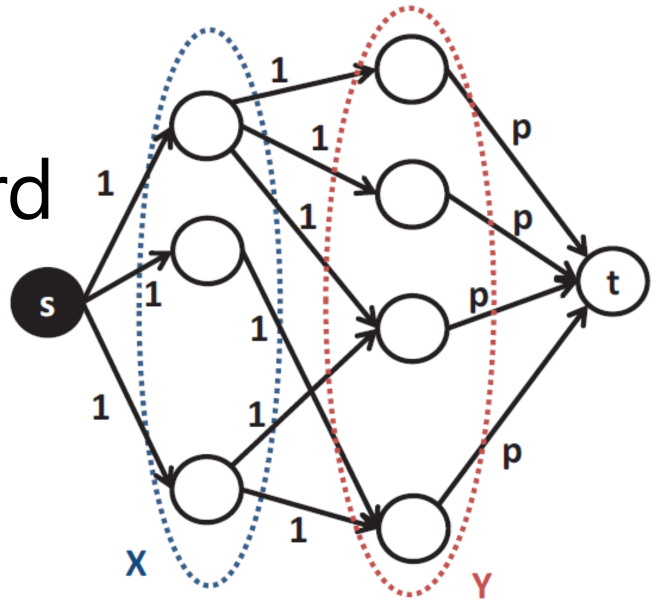
Illustration

- Selected nodes in R as recommendation candidates
  - Recommended iteratively
- A friend u triggers candidate v means v accepts request from s due to u



Hardness result

- Given R, calculating acceptance probability is #P-hard
- APM in general graph is NP-hard

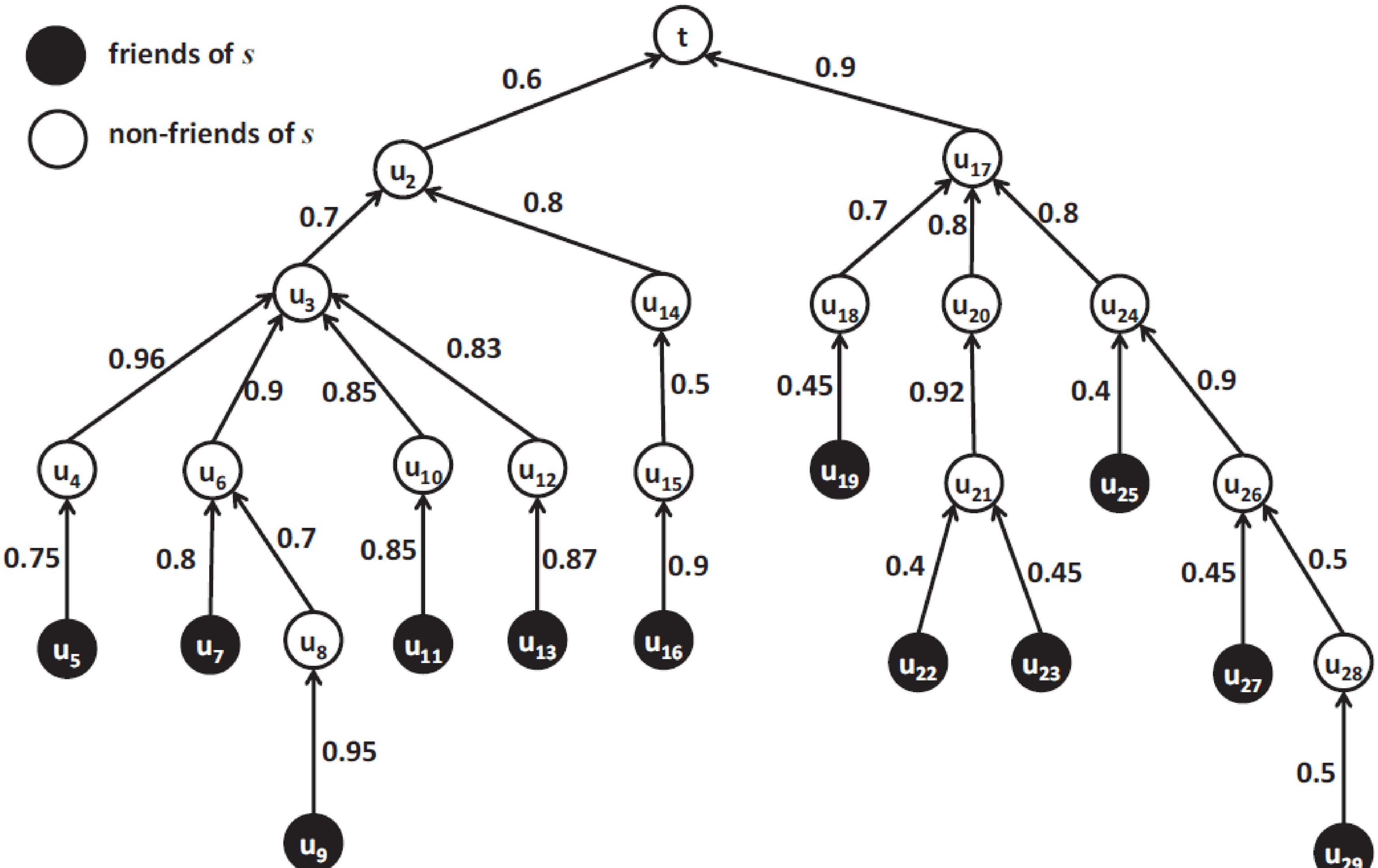


- Acceptance probability of t as a set function of R is not submodular (not having diminishing marginal return property)

Our approach

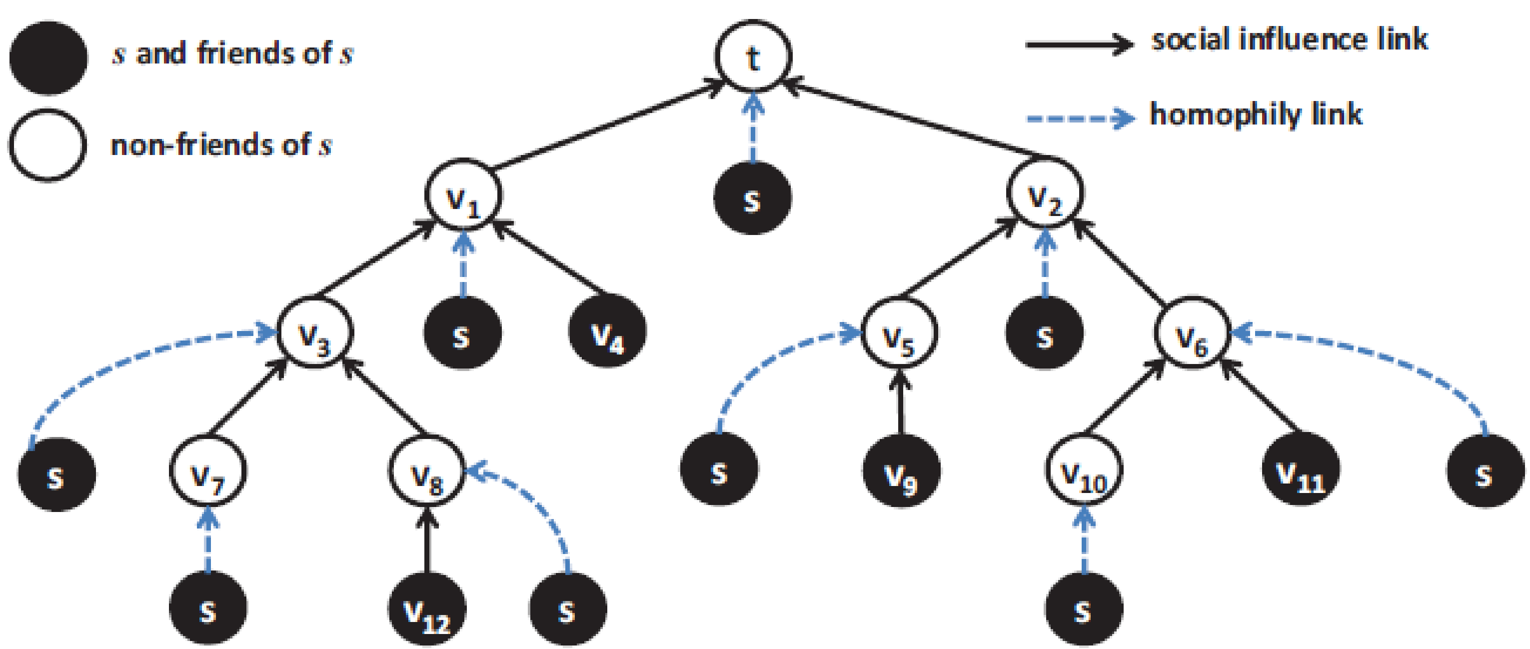
- Following MIA approach [C, Wang & Wang, 2010]
- Approximate influence diffusion on a MIA tree rooted at target t
- Algorithm SITINA: Apply dynamic programming to solve APM problem on the tree

Example of MIA tree



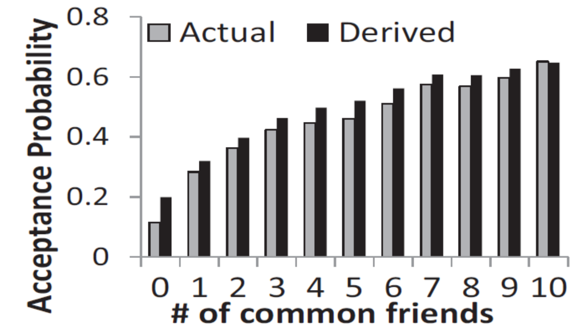
Incorporating homophily effect

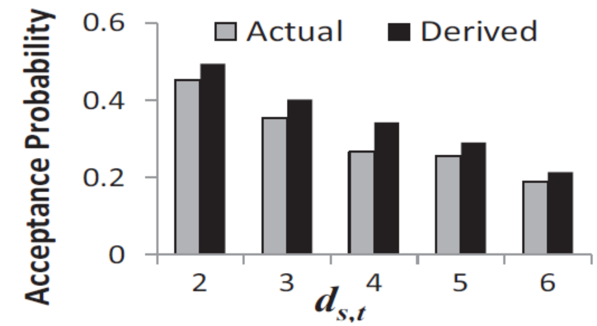
- Node v accepting request from s may not be due to common friend influence
- Homophily: may be v and s are similar
- Add direct link from s to node v for homophily

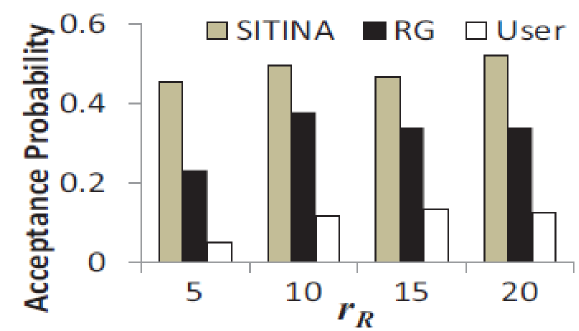


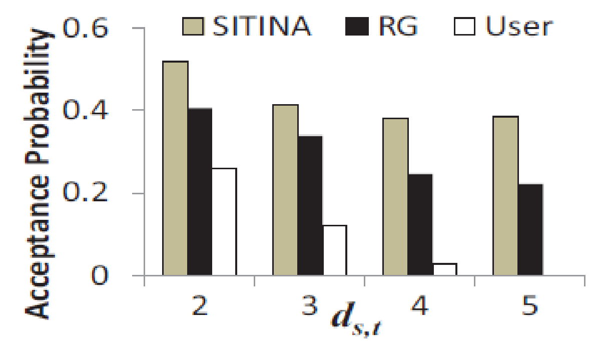
User study

- App on Facebook
  - 169 people participated
- Learn influence probabilities by acceptance behavior
- Results
  - Acceptance probability estimated by MIA model is accurate
  - SITINA recommendation works better than users' own selection



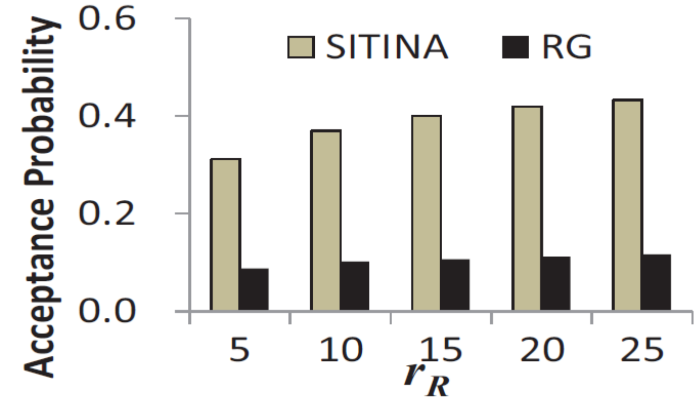




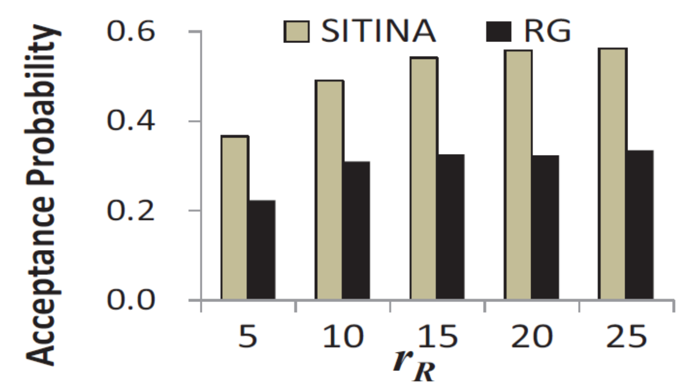


Simulation study

- Datasets
  - Facebook (60K users, 1.5M links)
  - Flickr (1.8M users, 22.6M links)
- Influence probabilities use distribution from the user study
- Results
  - SITINA leads to higher acceptance probability



Facebook



Flickr

(sec.)	SITINA	SITA	RG
UserStudyData	0.0002	0.4472	2E-05
FacebookData	0.0592	> 7 days	0.0002
FlickrData	1.8315		0.0026

Running time
($r_R = 20$)

Summary and Future Directions

- Stochastic modeling of influence/information diffusion
- Many optimization tasks on diffusion
- New directions to explore
  - Non-submodular optimization tasks
  - Dynamic graphs + influence diffusion
  - Online learning + influence maximization
  - Topic-aware influence maximization

Further resources

Search "Wei Chen Microsoft"

- Monograph: "Information and Influence Propagation in Social Networks", Morgan & Claypool, 2013
- KDD'12 tutorial on influence spread in social networks
- my papers