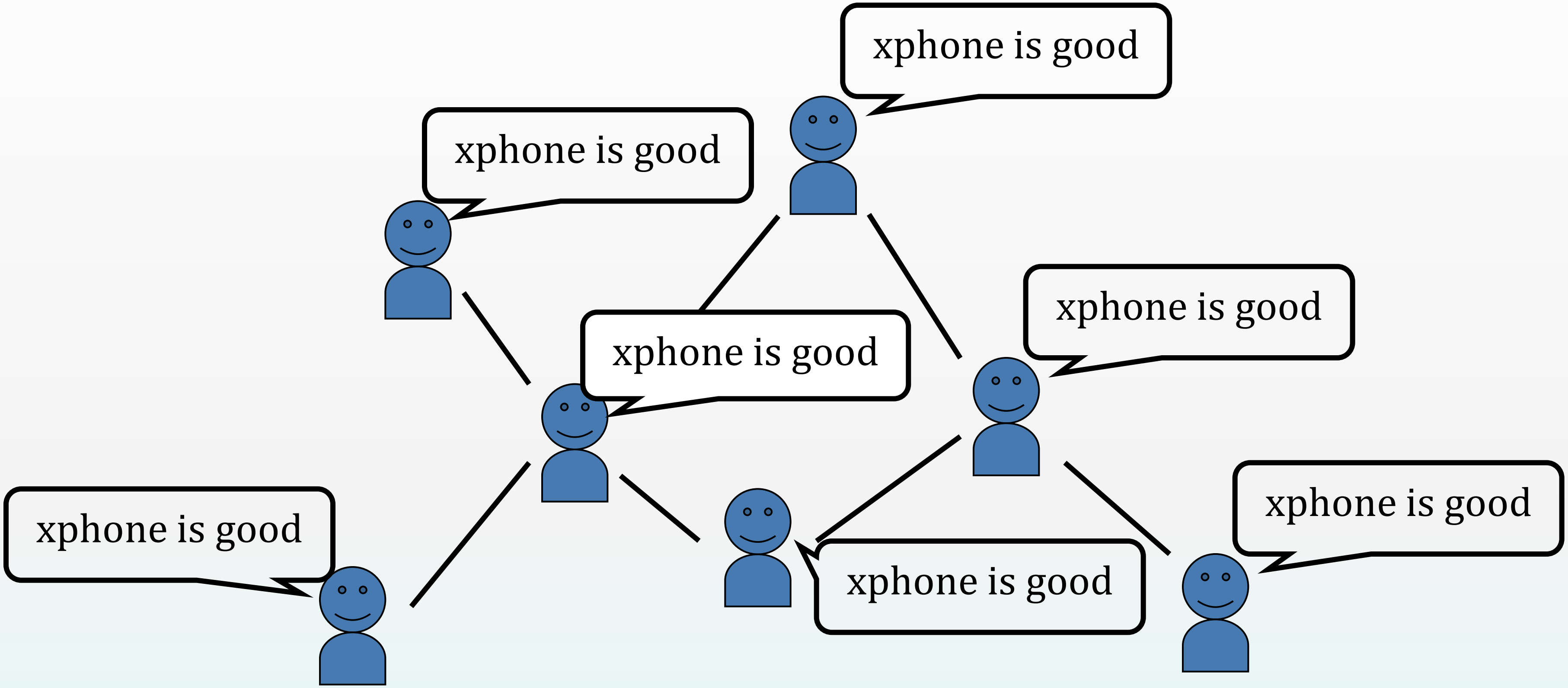
Influence Maximization When Negative Opinions May Emerge and Propagate

Wei Chen

Microsoft Research Asia

with *Alex Collins, Rachel Cummings, Te Ke, Zhenming Liu, David Rincon, Xiaorui Sun, Yajun Wang, Wei Wei, Yifei Yuan*

Word of mouth (WoM) effect in social networks

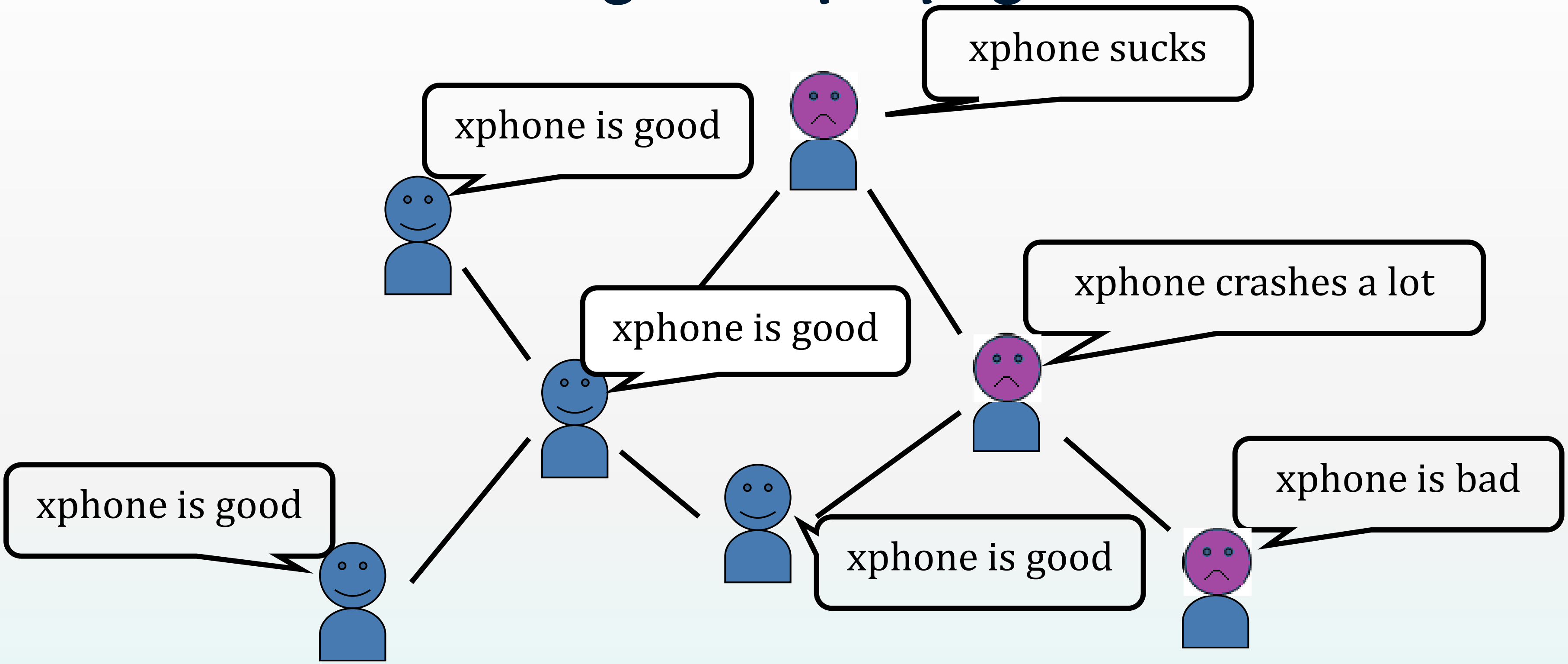


- Word of mouth (WoM) effect is believed to be a promising advertising strategy.
- Increasing popularity of online social networks may enable large scale WoM marketing

Existing influence maximization model

- Social network as a (directed) graph
  - Nodes represent individuals.
  - Edges are social relations.
  - Edge weights ($p(u, v)$) measure the strength of influence
- Independent cascade model [Kempe et.al 03]
  - 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: finding top k seeds that generates the largest *influence spread* (i.e. expected number of activated nodes)

However, negative opinions may emerge and propagate



- Negative opinions originates from poor product/service quality
- Negative opinions may be more contagious --- *negativity bias*

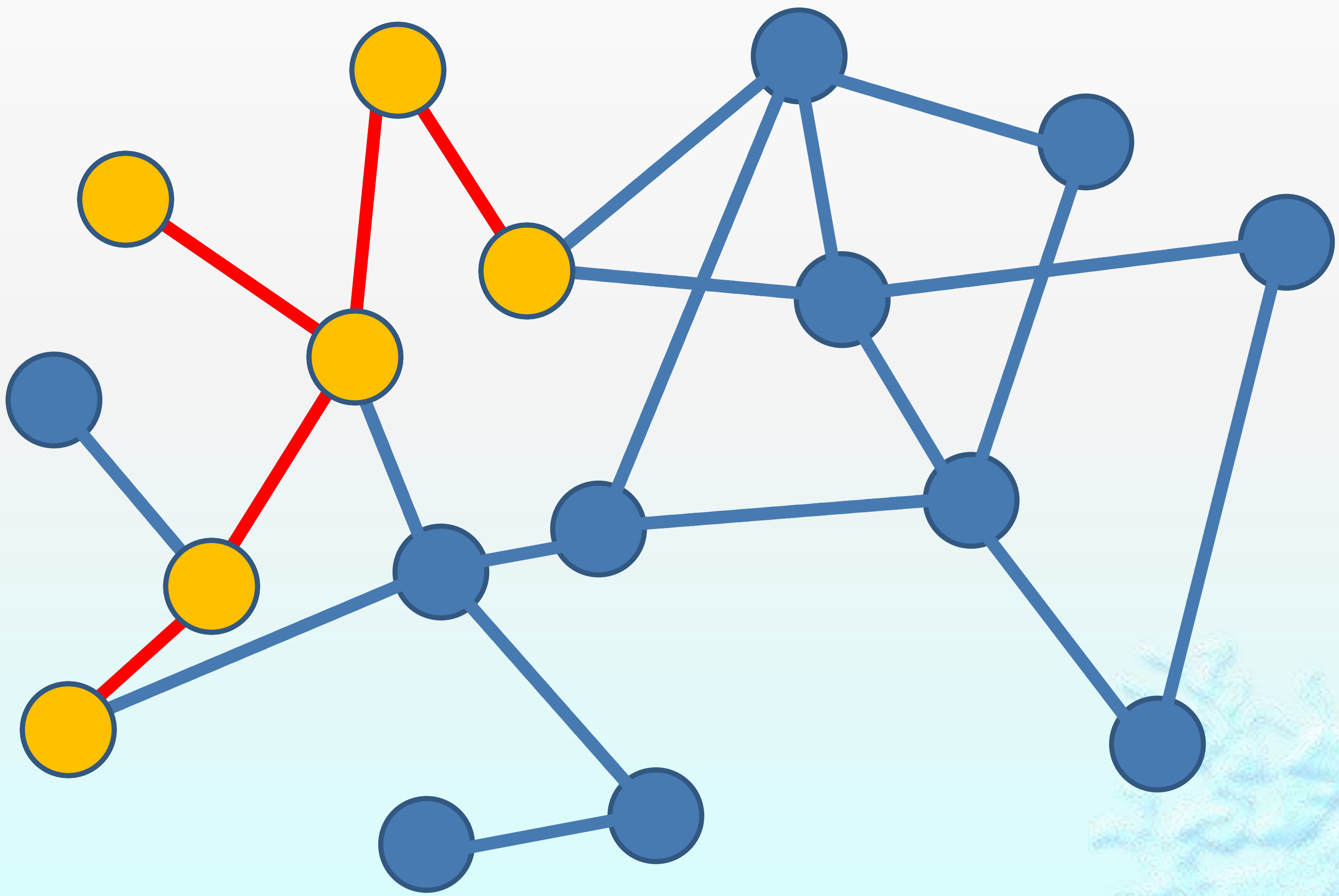
Influence maximization model with negative opinion

- Need to consider the effect of negative opinions to deploy influence maximization strategy.
  - model the emergence and propagation of negative opinion
  - consider negativity bias
  - study influence maximization with negative opinions

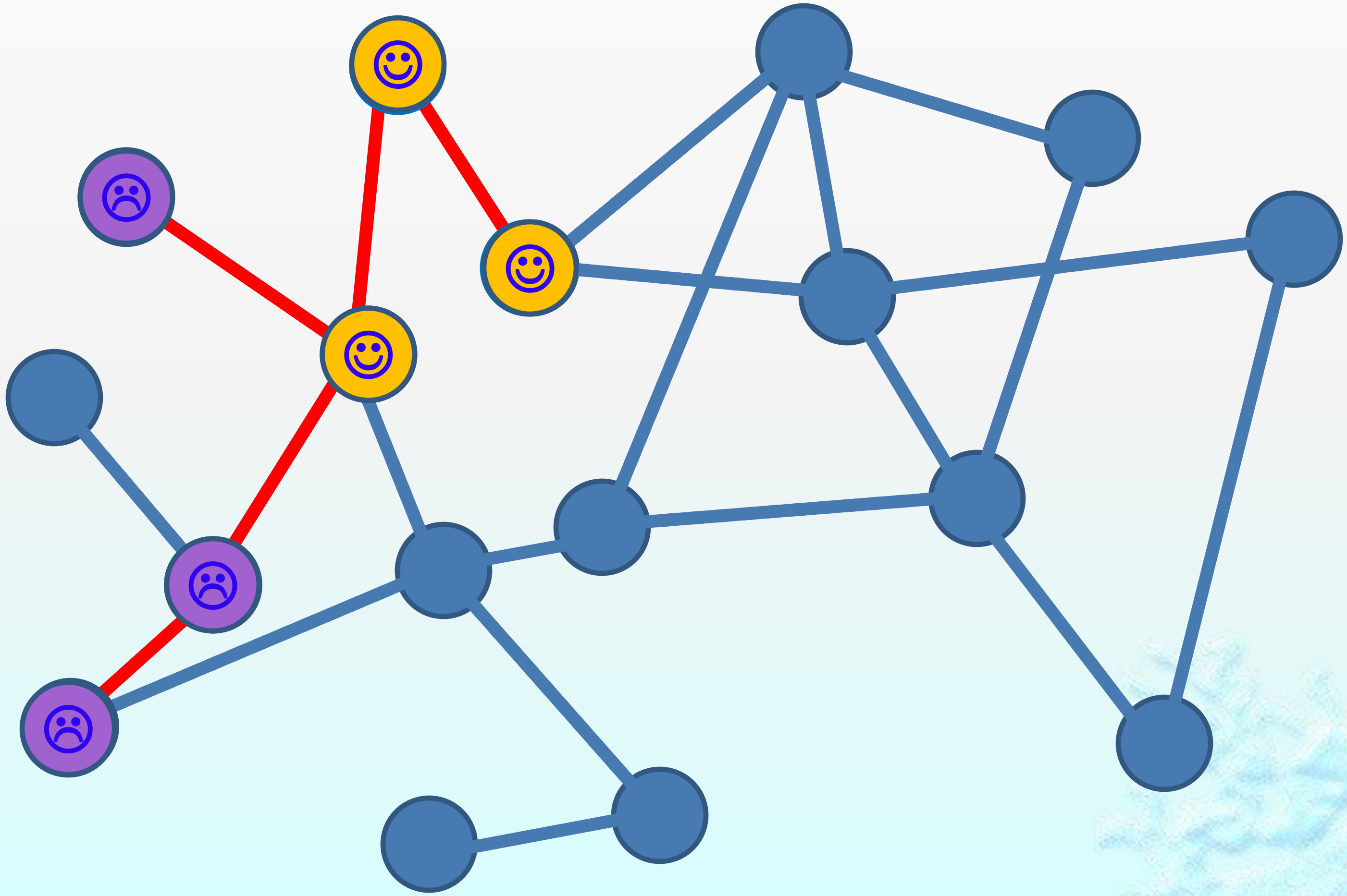
Negative opinion model

- Extention of the independent cascade model
- The quality of the product to be advertised is characterized by the quality factor (QF) $q \in [0,1]$.
- Each node could be in 3 states
  - Inactive, positive, and negative.
- When node v becomes active,
  - If the influencer is negative, the activated influencee is also negative (negative node generates negative opinions).
  - If the influencer is positive, the activated influencee
    - is positive with prob. q.
    - is negative with prob. $1 - q$.
  - If multiple activations of a node occur at the same step, randomly pick one
  - Asymmetric --- negativity bias

Independent Cascading Process (without considering QF).



Independent Cascading Process (when considering QF).



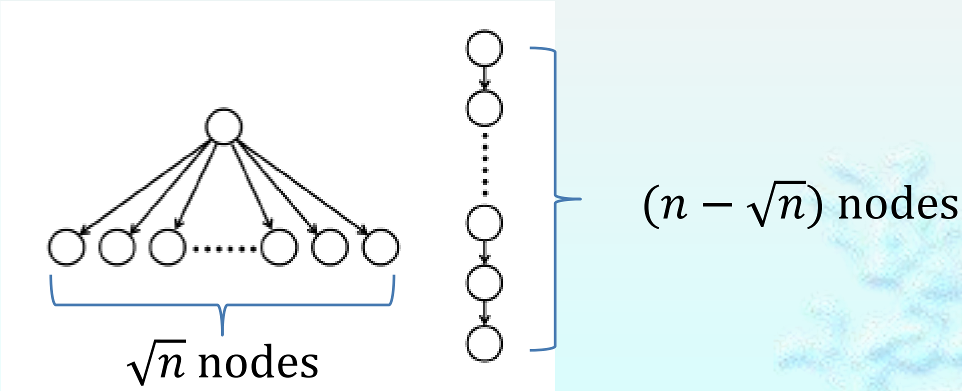
Our results (1)

- Complexity and approximation algorithm results

Scenario	Objective function	Algorithm result	Negative result
General directed graphs	Maximize expected positive nodes	$(1-\frac{1}{e}-\varepsilon)$-approx alg, due to submodularity	Exact sol. is NP hard.
Directed trees (arborescences)	Maximize expected positive nodes	Exists an efficient $(1-\frac{1}{e})$-approx. alg	Same as above
General directed graphs	Maximize expected (positive – negative) nodes.	Exists an $(1-\frac{1}{e}-\varepsilon)$-approx alg. Only when q is sufficiently large	Same as above
Directed graphs with different q for different people	Maximize expected positive nodes	NA	Objective is non-submodular

Our results (2)

- Q: is the knowledge of quality factor important?
  - guess a "universally good" value q so that regardless of the actual quality factor, the seeds are good?
  - No: ∃ social networks s.t. a wrong guess of q could lead to a much worse result than the optimal one. ($\Theta(\sqrt{n/k})$)
  - Intuition: which one seed to select in the following graph?



Our results (3) --- Main focus

- Q: what is the bottleneck of the approx. alg.
  - Given a specific seed set *S*, can we evaluate the expected number of positive nodes.
    - In general, #P-hard; can use Monte Carlo to approximate.
    - But exists efficient exact algorithm for arborescence (trees).
  - Developed scalable heuristic based on influence calculation alg. for arborescences.

Greedy algorithms for influence maximization

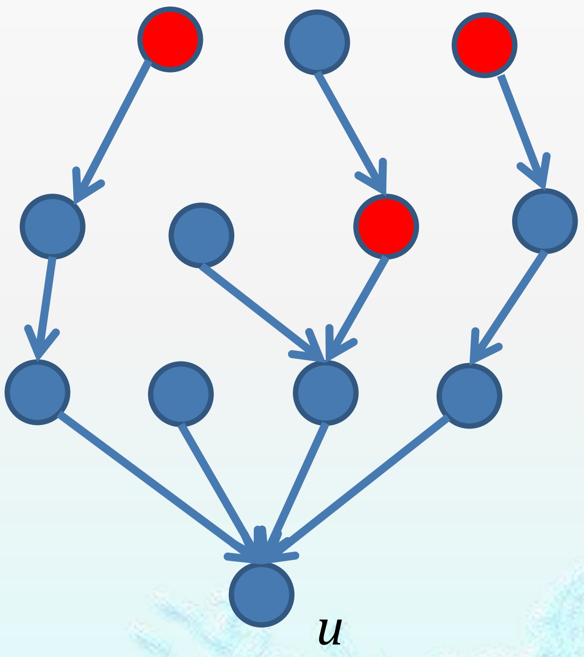
- Def. Let G be an influence graph. Let $S \subseteq V$ be a seed set. Let $\sigma(S)$ = expected # of positive nodes
- Theorem: $\sigma(S)$ is submodular.
- Greedy algorithm works:
  - Step 1. Set $S \leftarrow \emptyset$
  - Step 2. *for* $i \leftarrow 1$ to k
    - Find u in the remaining non-seeds s.t. $\sigma(\{u\} \cup S)$ maximized
    - Set $S \leftarrow S \cup \{u\}$
  - provide $\left(1 - \frac{1}{e}\right)$ approximation guarantee
  - Computing $\sigma(\cdot)$ function is costly

To overcome the drawback of existing greedy algorithm

- Design efficient algorithm computing $\sigma(S)$ for trees
- Utilize the algorithm for trees to design scalable heuristics for general graphs.

Computation in directed trees (in-arborescences)

- Without negative opinions, a simple recursion computes the activation probability of u:
  - $ap(u) = 1 - \prod_{w \in N^{in}(u)} (1 - ap(w)p(w,u))$
- Difficulty with negative opinions: needs to know whether the neighbors of u is positive or negative --- because of negativity bias



Solutions for in-arborescences

- Step 1: compute activation probability of u at step t (via dynamic programming):

$$ap(u,t) = \begin{cases} 1 & t = 0 \wedge u \in S, \\ 0 & t = 0 \wedge u \notin S, \\ 0 & t > 0 \wedge u \in S, \\ \prod_{w \in N^{in}(u)}[1 - \sum_{i=0}^{t-2} ap(w,i)p(w,u)] \\ \quad - \prod_{w \in N^{in}(u)}[1 - \sum_{i=0}^{t-1} ap(w,i)p(w,u)] & t > 0 \wedge u \notin S. \end{cases}$$

- Step 2: compute positive activation probability of u at step t:

$$pap(u,t) = ap(u,t) \cdot q^{t+1}.$$

Influence computation for general graphs

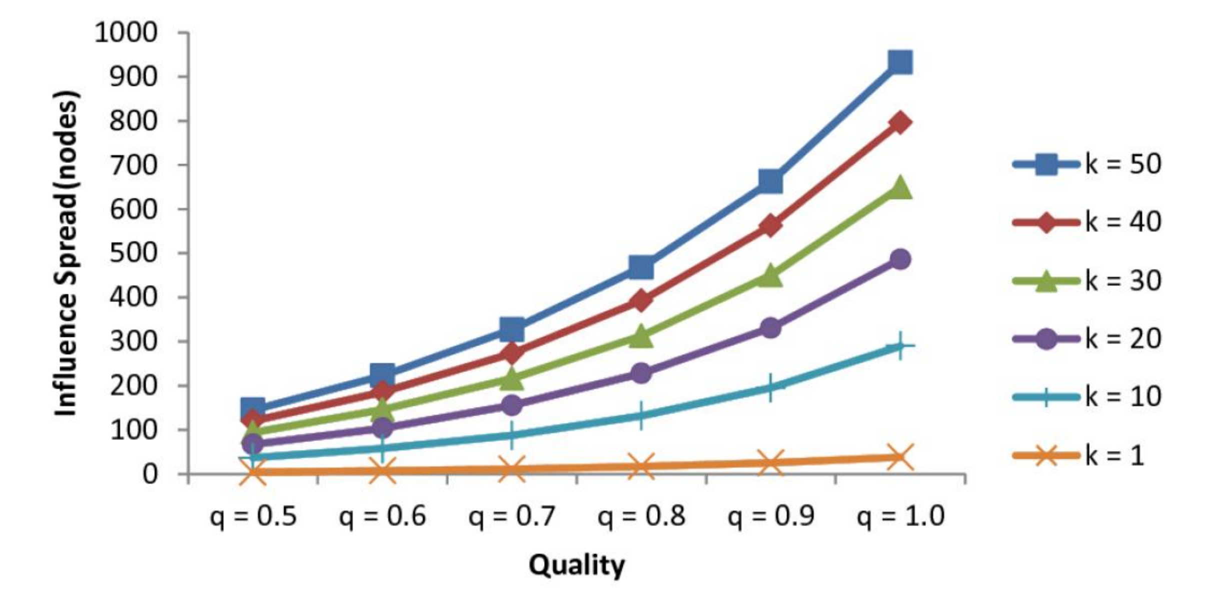
- Utilize influence computation for trees
- Heuristic 1: restrict influence to a node v to a local region --- far-away influence is negligible
- Heuristic 2: “sparcify” the local region of node v to an in-arborescence by finding only the strongest influence path from other nodes to v.

Experiments

- **NetHEPT**: academic collaboration network on high energy physics extracted from arXiv.
- **WikiVote**: interactions among Wikipedia users.
- **Epinions**: extraction of a social network from a website. Contains trust-ness information.

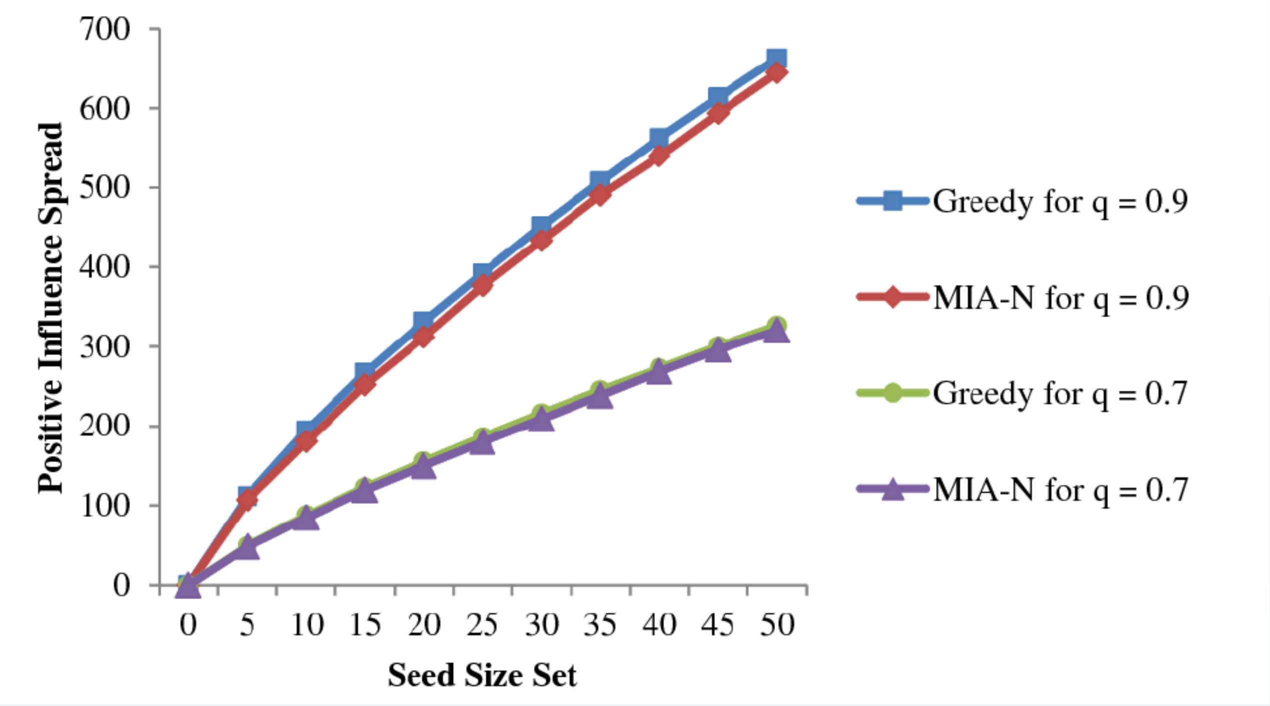
Date set	NetHEPT	WikiVote	Epinions
# of nodes	15,000	7,000	76,000
# of edges	31,000	101,000	509,000
Avg. degree	4.12	26.64	13.4
Max. degree	64	1065	3079

Influence spread and QF



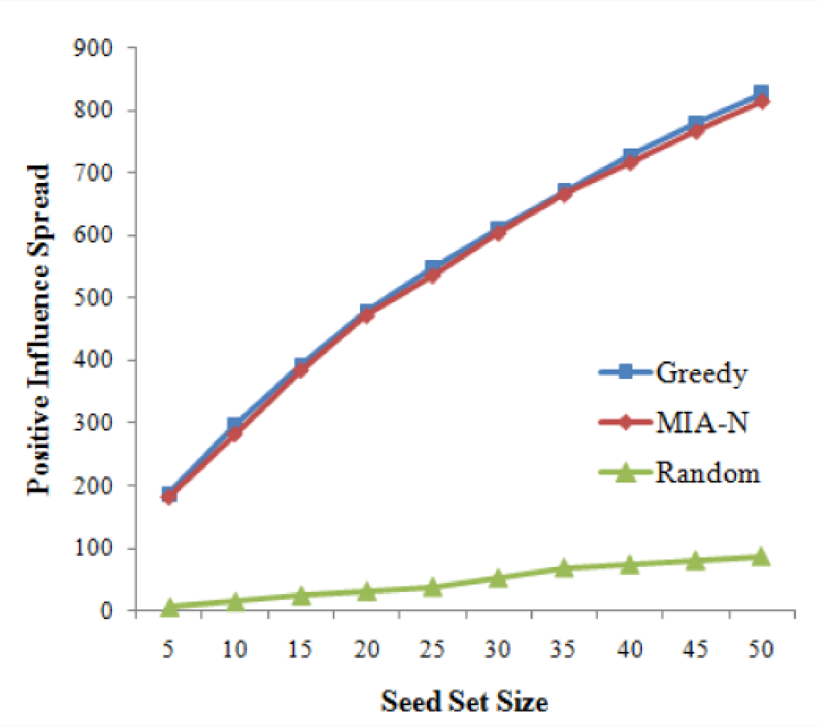
Convex function because of the asymmetric spreading model

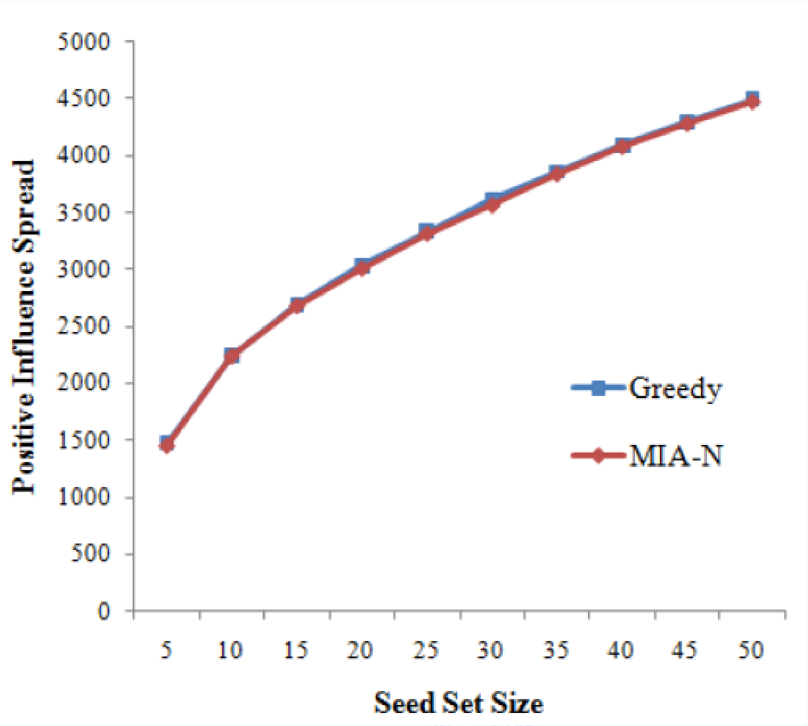
Performance of the heuristic



- Results on NetHEPT.
- MIA-N is the heuristic, performs nearly as good as the original greedy algorithm.

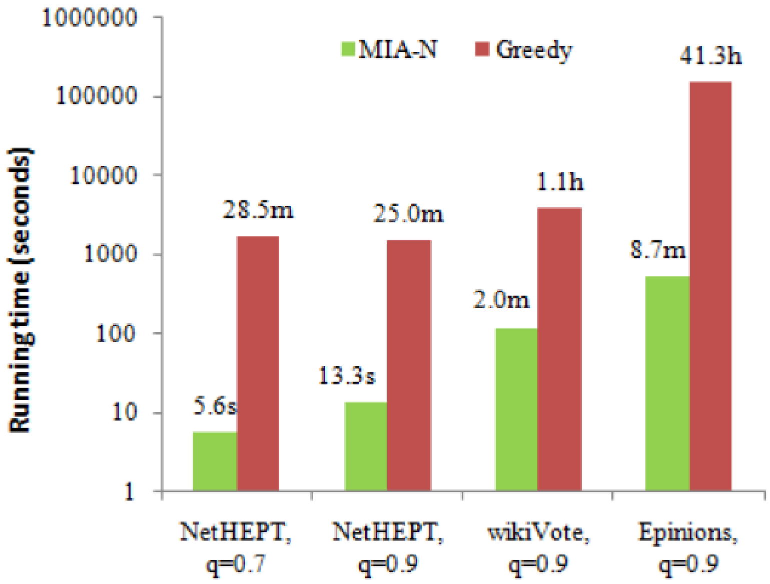
Performance of the heuristic

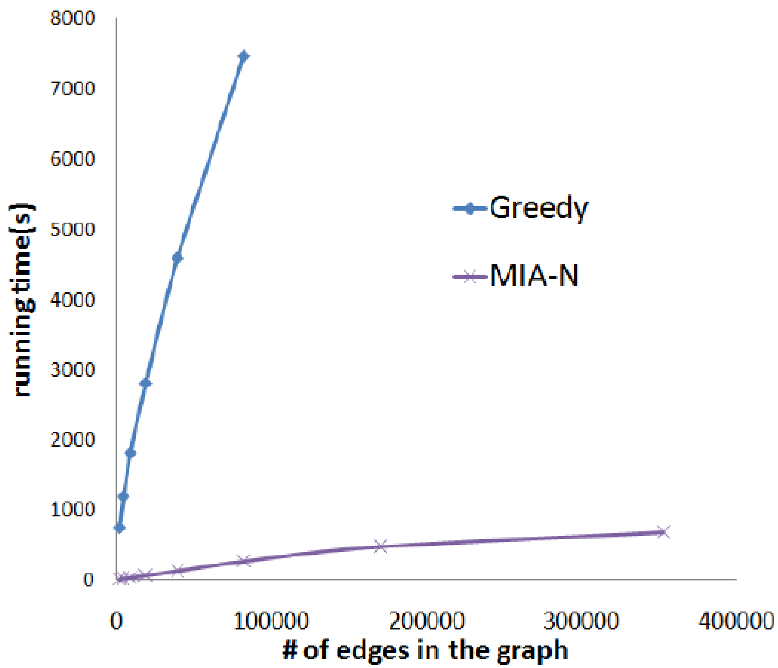




- Results on Wikivote and Epinions for q = 0.9.

Scalability





Future directions

- Consider other sources of negative opinion propagations
  - e.g. from competitors
- Validation of propagation models with negative opinions

Questions?