Influence Spread in Large-Scale Social Networks – A Belief Propagation Approach

Huy Nguyen and Rong Zheng

Department of Computer Science

University of Houston, TX, USA

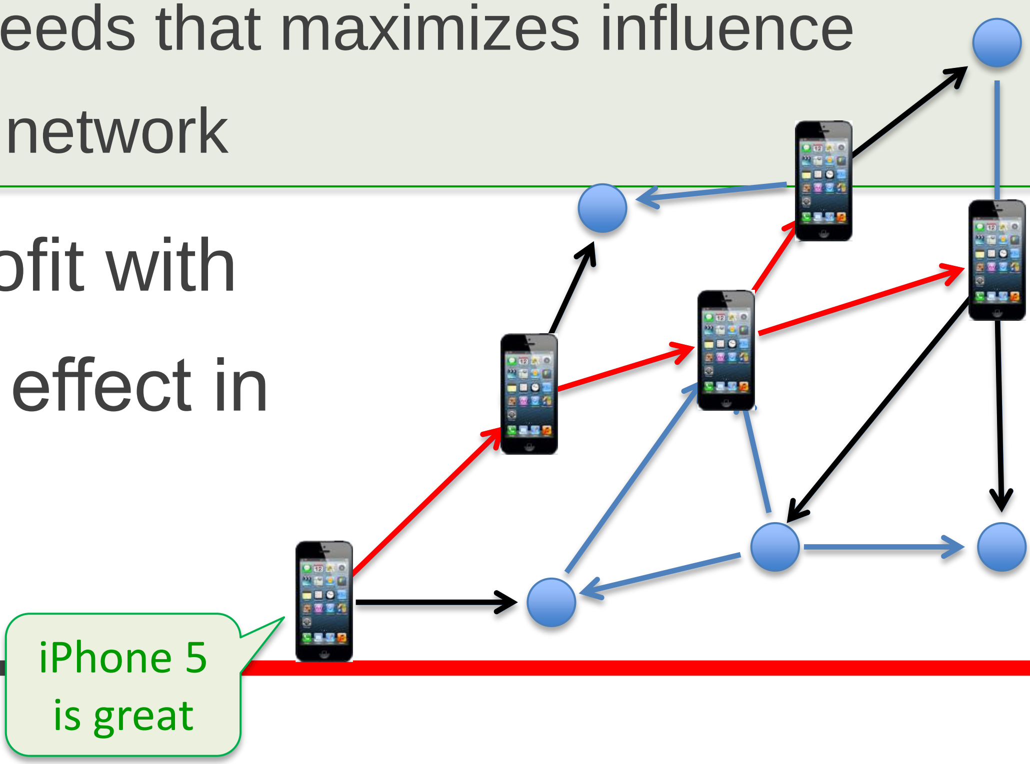
The social media network

Agenda

- Influence Maximization in social networks
- Spread computation on DAGs
- Seed selection algorithm
- Evaluation
- Conclusion and Future Work

Influence maximization (IM) problem

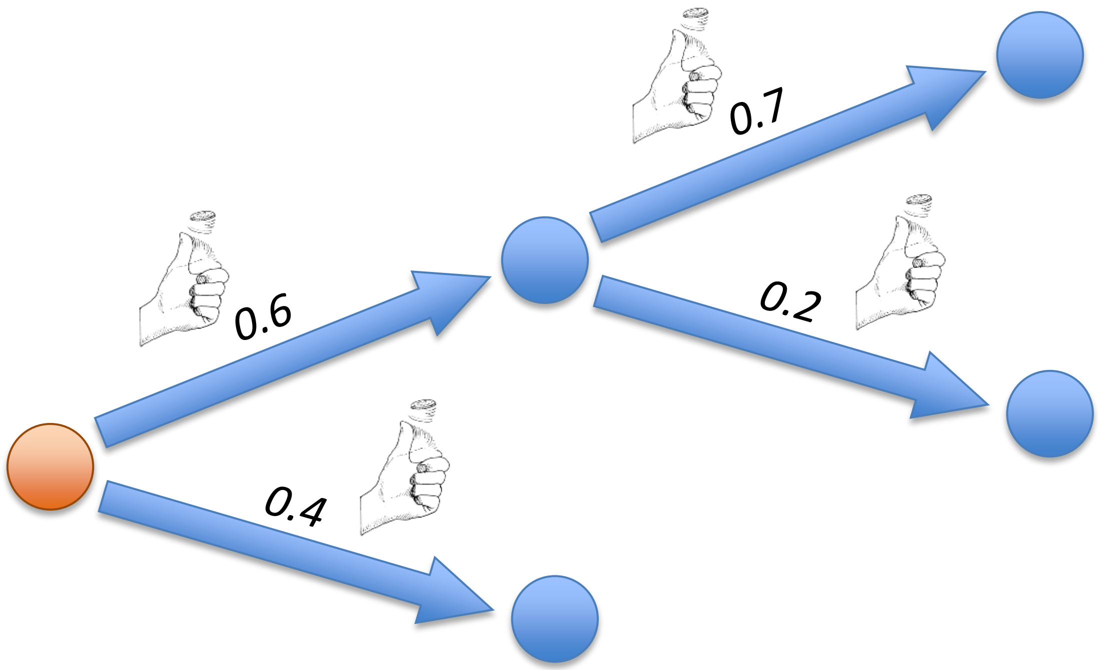
- Users influence each other in a social network
  - Spreading opinion, idea, information, action …
- Influence maximization problem (#P-Hard)
  - Find a set of k seeds that maximizes influence spread over the network
- Maximize the profit with "word-of-mouth" effect in Viral Marketing



WiSeR Wireless System Research Group

Independent cascade model

- Spread probability associated with each edge



- Influence spread = expected number of influenced nodes

Wireless System Research Group

Traditional solution

Greedy seed selection scheme [Kempe et al. 2003]

1. Seed set $S = \emptyset$
2. Calculate incremental spread of $v, \forall v \in V$
3. Select u = node with max incremental spread
4. $S = S \cup u$
5. Return to step 2 until $|S| = k$

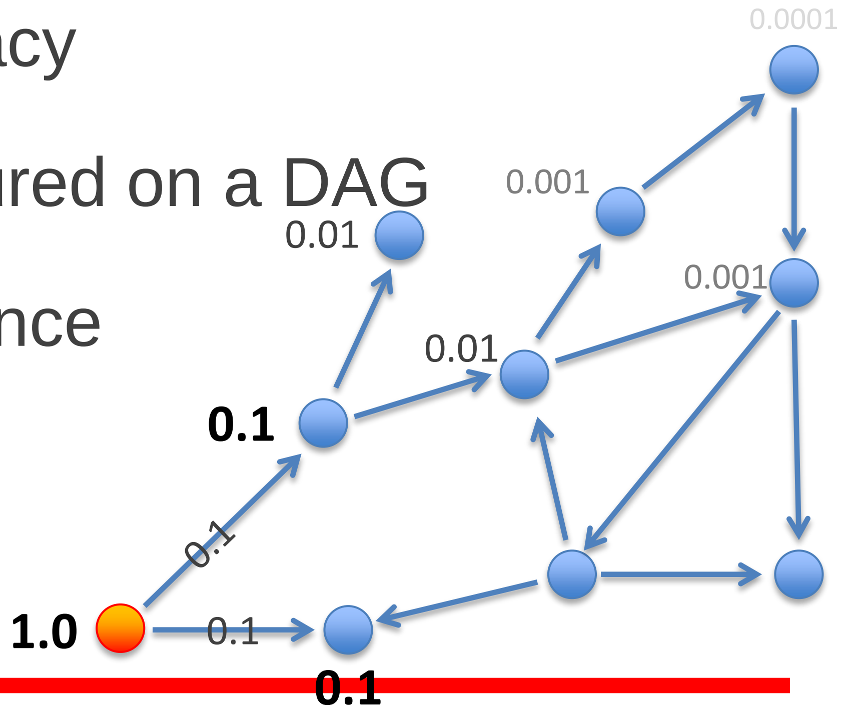
- As good as ~63% of the optimal solution
- Problem
  - Influence spread computation
  - Too many evaluations after each iteration

Our contributions

- Solutions to both aforementioned problems
- Too many evaluations after each iteration
  - Localizing the influence region from a node – modeled by directed acyclic graphs (DAGs)
  - Minimizing the number of nodes to be evaluated
- Influence spread computation
  - Spread computation using belief propagation algorithm on Bayesian Network

Localizing spread region

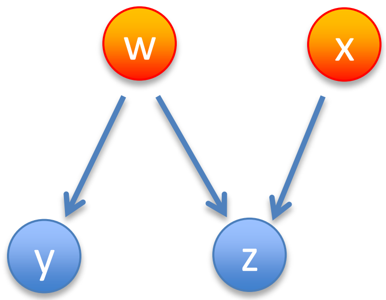
- Influence spread decays quickly with distance from the source
- Localizing spread region make computation much faster while retaining accuracy
- Most influence can be captured on a DAG
- DAG structure makes influence computation much easier



Wireless System Research Group

Belief propagation

- Technique invented by Pearl in 1982 to calculate marginals / most likely states in Bayes nets.
- Given
  - Bayes net
    $P(w,x,y,z) = P(w)P(x)P(y|w)P(z|w,x)$
  - Observed variables: w, x
  - Hidden variables: y, z

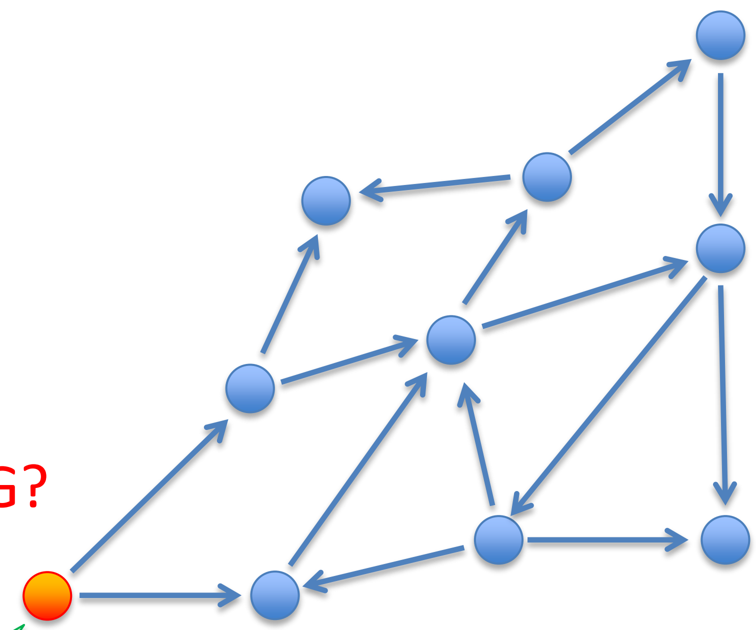


- Find: $P(y), P(z)$
- Neighbors passing “messages”: I (w) think that you (z) belong in states … with likelihood …
- Messages passed from observed to hidden variables
  - Marginal probabilities (**beliefs**) could be estimated

Spread computation on DAGs

- Exact computation of influence spread is hard (#P-complete even on DAGs)
- Belief propagation algorithms calculate marginal distribution from a set of seeds
- Two BP algorithms used
  - Loopy: slow – more accurate
  - Single-pass: fast – less accurate

Wait, how do we convert a graph to a DAG?

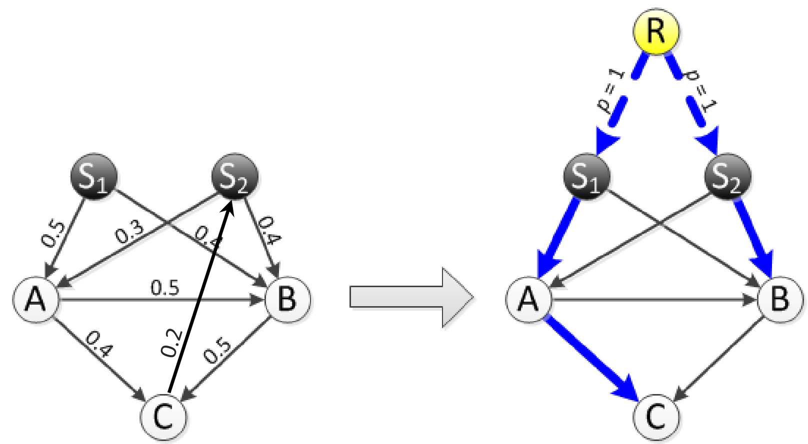


Expectedly, how many people can I persuade?

DAG 1

- Any DAG has at least one topological order
- Order can be obtained from node's "distance" from a seed (a.k.a. node rank)

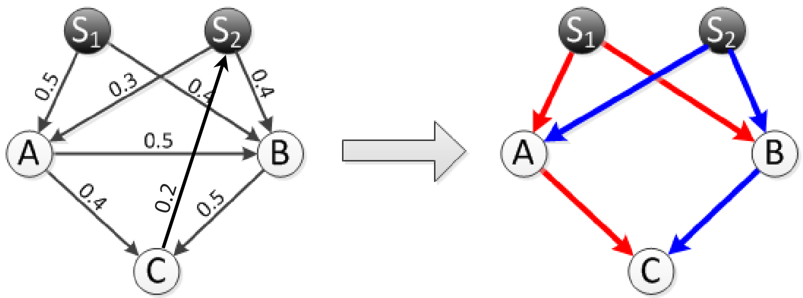
1. Introduce a super root R connected to all seeds with p = 1
2. Calculate a Dijkstra tree T from R
3. Calculate rank of all nodes on T
4. Augment T with edges from a lower to a higher ranked node



Node	S_1	S_2	A	B	C
$r(\text{Node})$	0	0	0.301	0.398	0.699

DAG 2

- Build Dijkstra trees from seed nodes
- DAG 2 = union of all Dijkstra trees
- Comparing to DAG 1:
  - DAG 2 is built faster
  - Same set of nodes
  - Subset of edges



- Spread computation problem is converted to an instance of BP on a Bayesian network

Seed selection algorithm

Greedy seed selection scheme [Kempe et al. 2003]

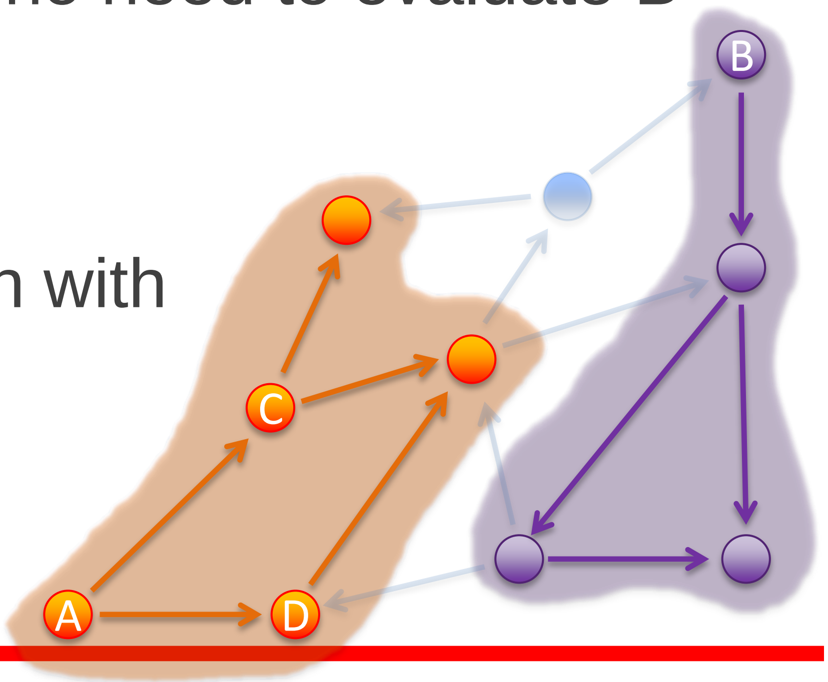
1. Seed set $S = \emptyset$
2. Calculate incremental spread of v, $\forall v \in V$
3. Select u = node with max incremental spread
4. $S = S \cup u$
5. Return to step 2 until $|S| = k$

Estimated with BP algorithm on DAGs

Candidate set is reduced with Lazy Forward mechanism [Leskovec et al. 2007]. However, it can be further improved.

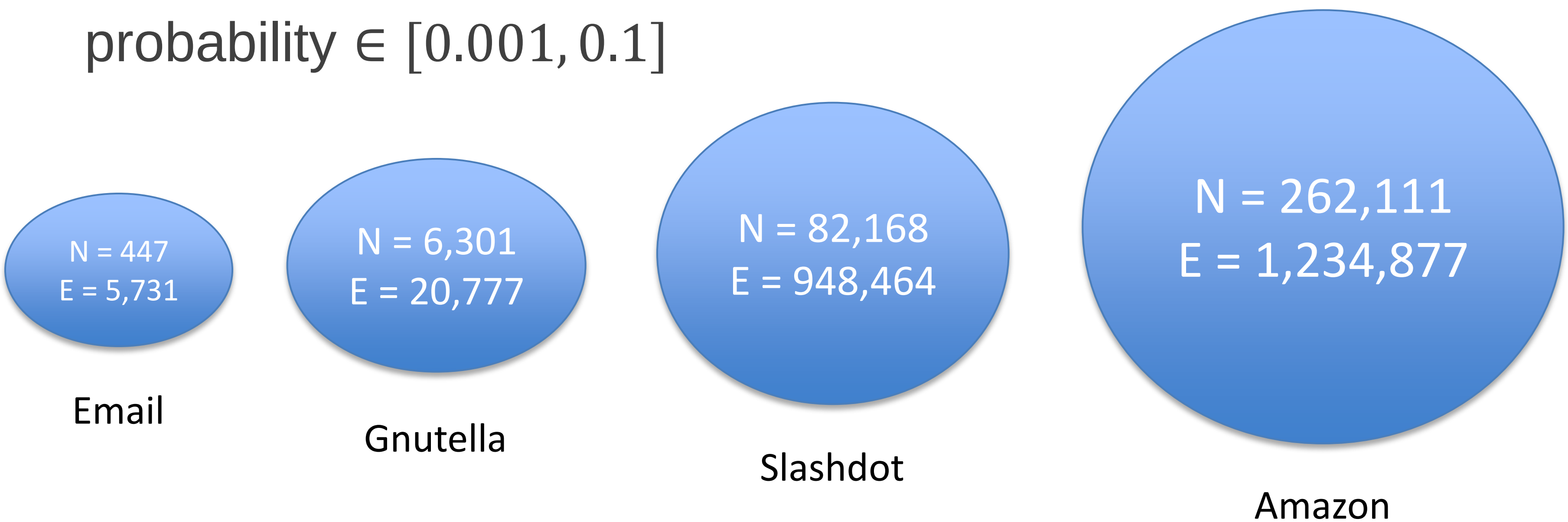
Seed selection algorithm (2)

- Only need to evaluate nodes that have overlapping influence regions with the new seed
- A is selected as a seed → no need to evaluate B again
- Can be used in conjunction with Lazy Forward mechanism

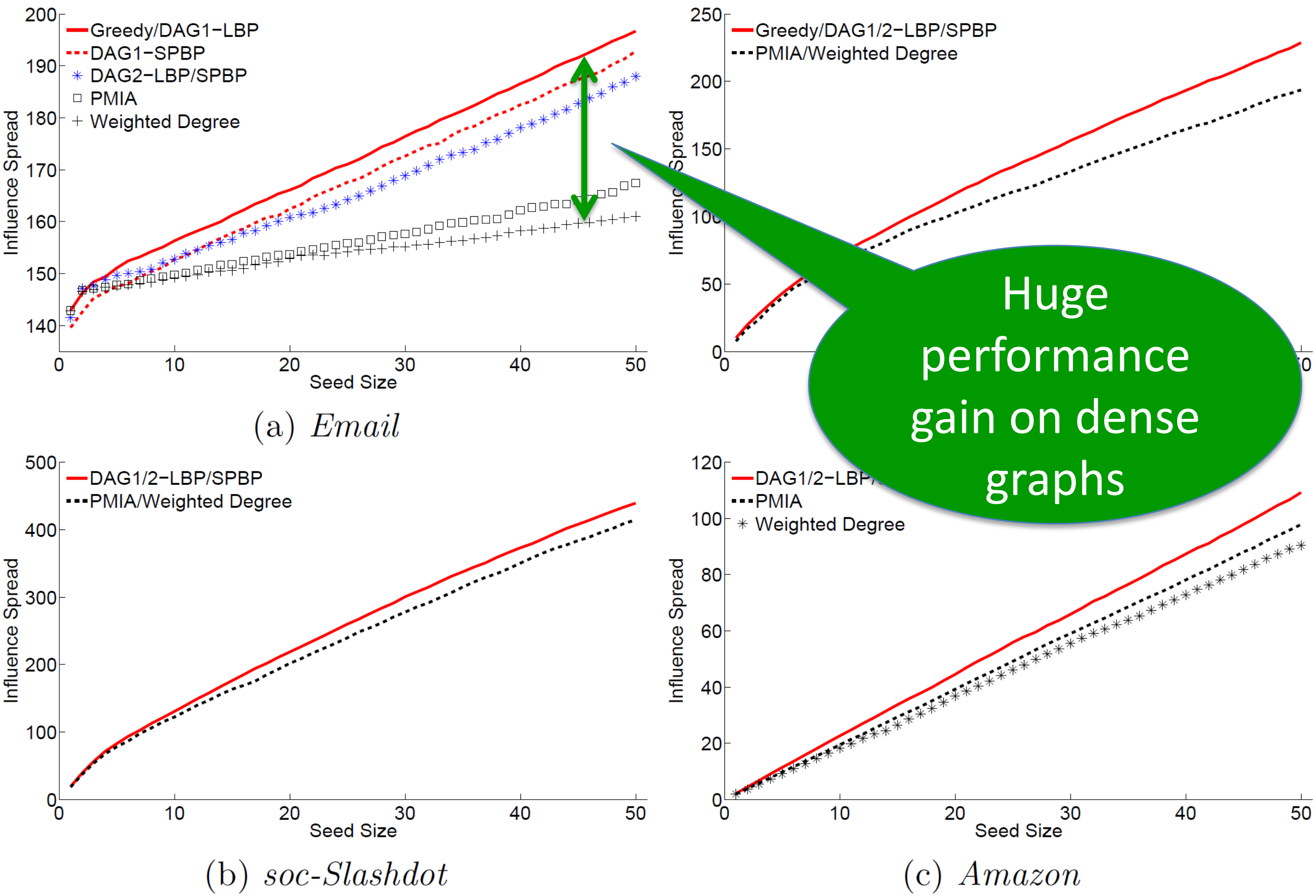


Evaluation overview

- Our approach vs. state-of-the-art solutions: PMIA [Chen et al. 2009], CELF [Leskovec et al. 2007], and Weighted Degree
- Network edge is assigned random propagation probability $\in [0.001, 0.1]$



Influence spread result

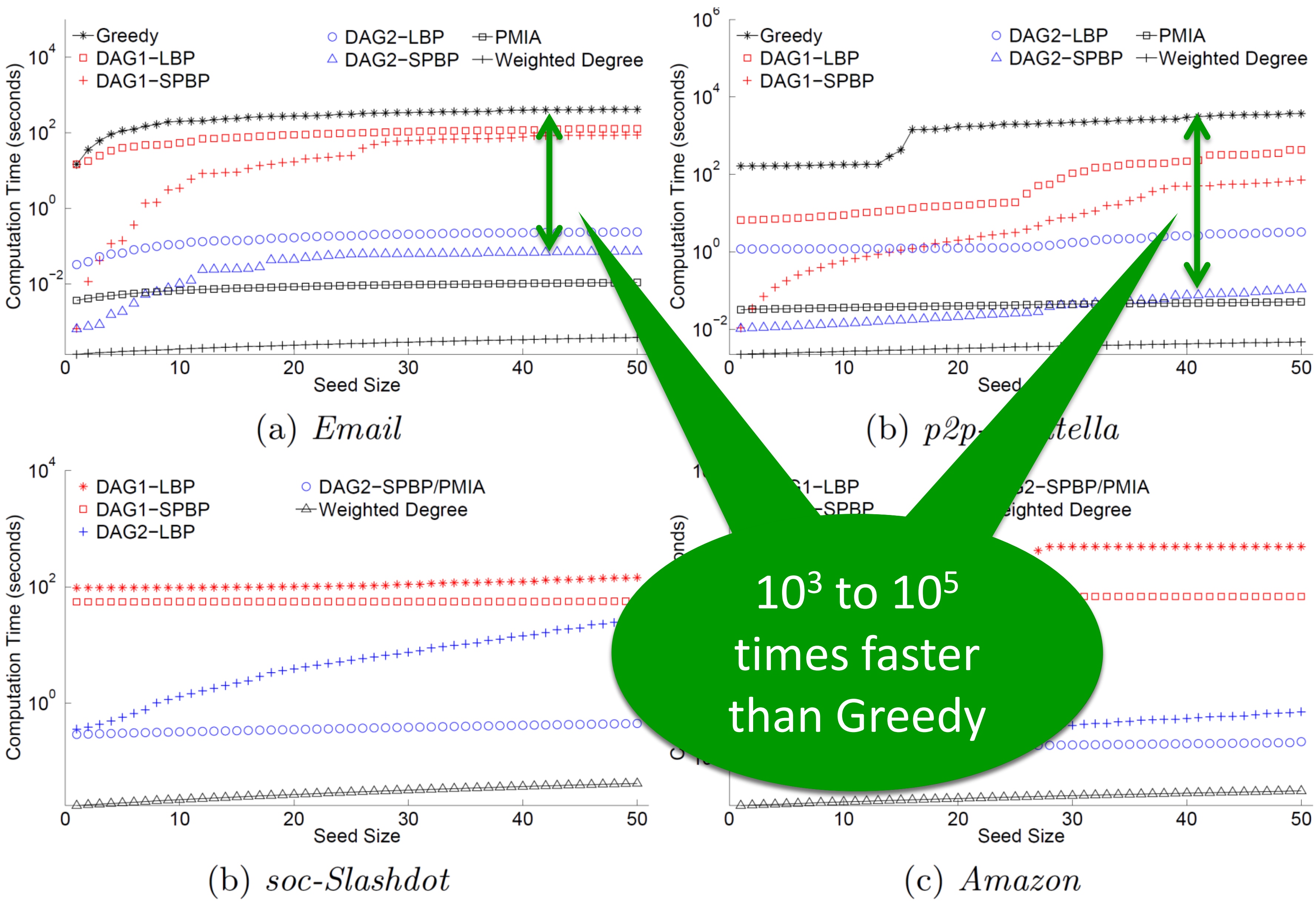


(a) *Email*

(b) *soc-Slashdot*

(c) *Amazon*

Running time result



Conclusion and future work

- New framework to solve IM problem in social networks with BP algorithms
- Application flexibility

	DAG 1	DAG 2
Loopy	**Best performance - slow**	Better than DAG2-SPBP
Single-pass	Very close to DAG1-Loopy	**Acceptable performance - fastest**

- Future study
  - Impact of graph structure on IM algorithm selection
  - IM problem with incomplete network data

Thank you for your attention

nahuy@cs.uh.edu

http://www2.cs.uh.edu/~nahuy/