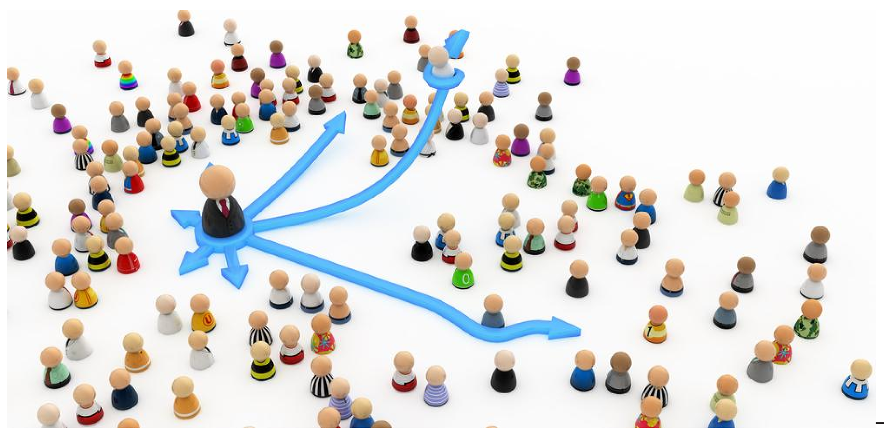


Holistic IM: Scalability & Efficiency with Opinion-Aware Models

Sainyam Galhotra[a], Akhil Arora[a], and Shourya Roy

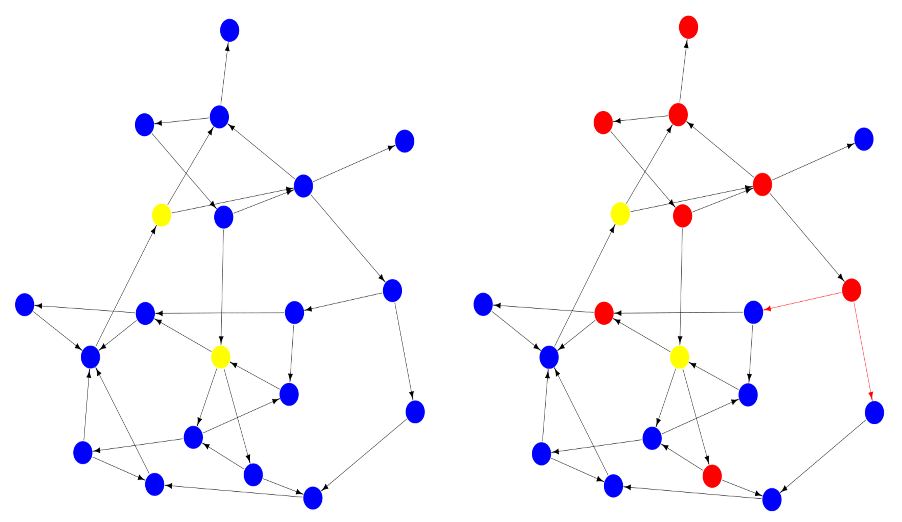
Text & Graph Analytics, Xerox Research Centre India

sainyam@cs.umass.edu, {akhil.arora,shourya.roy}@xerox.com

[a]Equal Contribution

Motivation

- Need for Influence based Modelling??
- Answer: Interpret Real-world processes
  - Spread of information/diseases
  - Traffic and its propagation



- Real-World Applications
  - Product/Topic/Event Promotions
  - Managing Celebrity/Political Campaigns
  - Blog Selection/Viral Ad-Targeting
  - Detect Outbreaks/Epidemics/Rumours
  - Many more ...

Influence Maximization

- **Given:** A model for information diffusion
- **Task:** Identify the most-influential set of nodes
- **Constraints:** Budget ($k = |S|$)
- $\sigma(S) = \mathbb{E}[\mathbb{F}(S)]$: Expected number of activated nodes, if S is targeted for initial activation
- More formally, given a budget k, select a set S of $k = |S|$ nodes, so as to maximize $\sigma(S)$

Unaddressed Challenges!

- Classical models fail to capture real-world phenomena
- How to model diffusion of opinion?? [4, 5]
  - Most models don't incorporate change of opinion; except the OC model
- Extensive study on **run-time** efficiency and efficacy – [1, 2, 3], KDD'07, SODA'14, CIKM'14
- **Scalable** solutions (catering to both running-time and memory-consumption) are **non-existent**

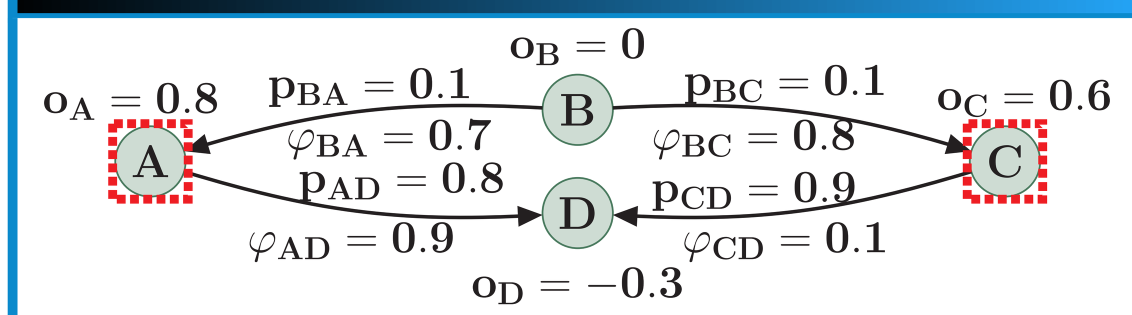
Propagation Models

- Opinion Aware models
  - Opinion-based Cascading (OC) model
  - Opinion-cum-Interaction (OI) model
- Opinion Oblivious models
  - IC, WC and LT Models

OI Model

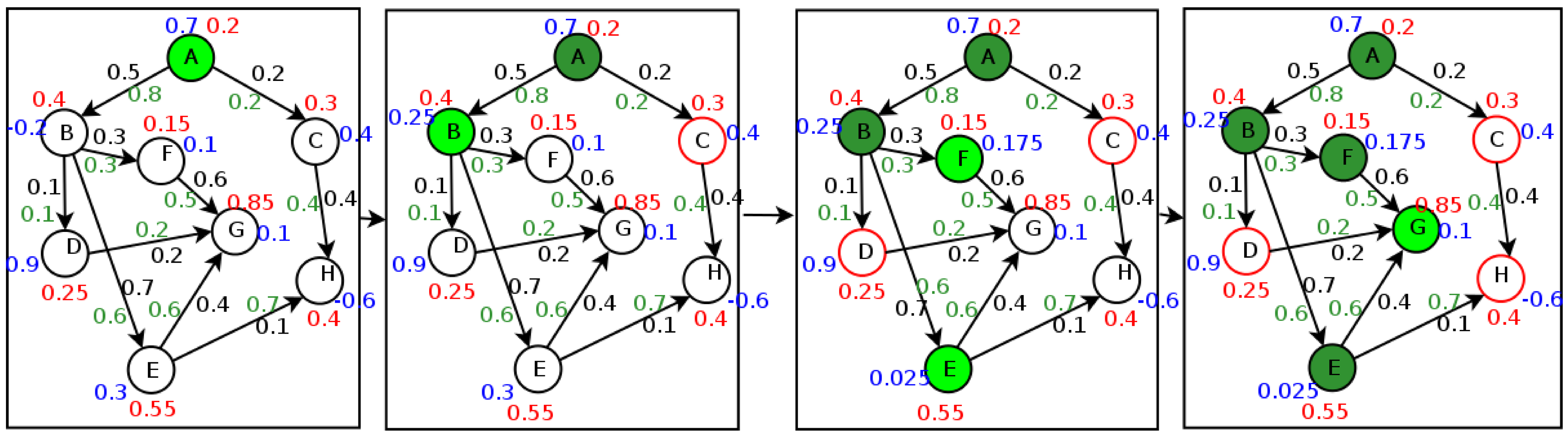
- Added a layer to model change of opinions
- Opinion of a newly activated node
  - is computed using: (1) Its personal opinion, (2) opinion of activating node, and (3) their interaction probability

Need for Opinion-Aware IM



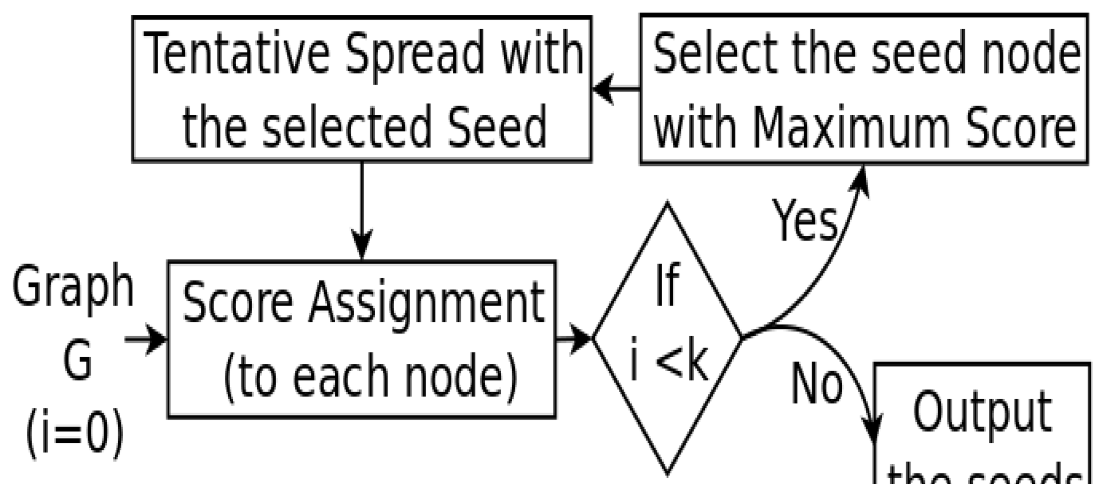
- $\sigma(A) = 0.8$, $\sigma(B) = 0.3628$, $\sigma(\mathbf{C}) = \mathbf{0.9}$ and $\sigma(D) = 0$
- $\sigma^o(A) = p_{AD}\big(\varphi_{AD}(o_D+o_A)/2+(1-\varphi_{AD})(o_D-o_A)/2\big) = 0.136$
- $\sigma^o(B) = -0.022564$, $\sigma^{\mathbf{o}}(\mathbf{C}) = -\mathbf{0.351}$ and $\sigma^o(D) = 0$

OI under LT Model



- Each node possesses an **activation threshold** $\theta_v \in [0, 1]$
- **Edge Weights:** $w_{(u,v)} \in [0, 1]$ & Interaction Probability ($\varphi_{(u,v)}$)

Overview of our Approach

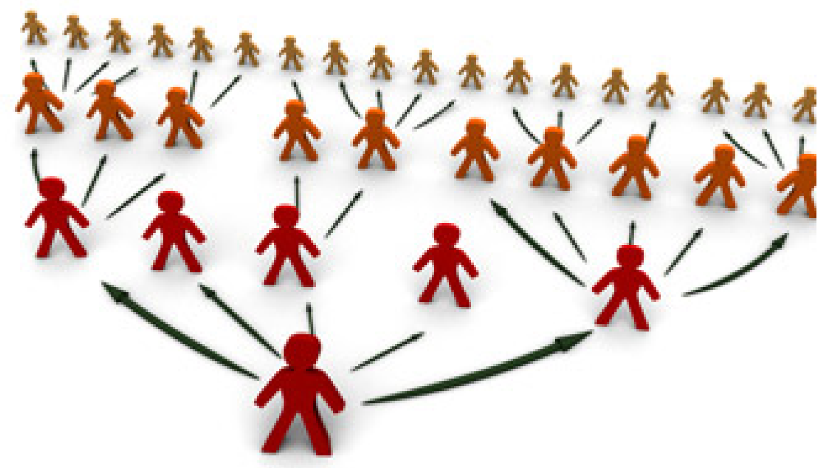


Analysis

- **Time Complexity** (score assgn) – $\mathbf{O(l(m+n))}$
- Total time taken for k seeds – $\mathbf{O(kl(m+n))}$
- **Memory Complexity** – $\mathbf{O(n)}$

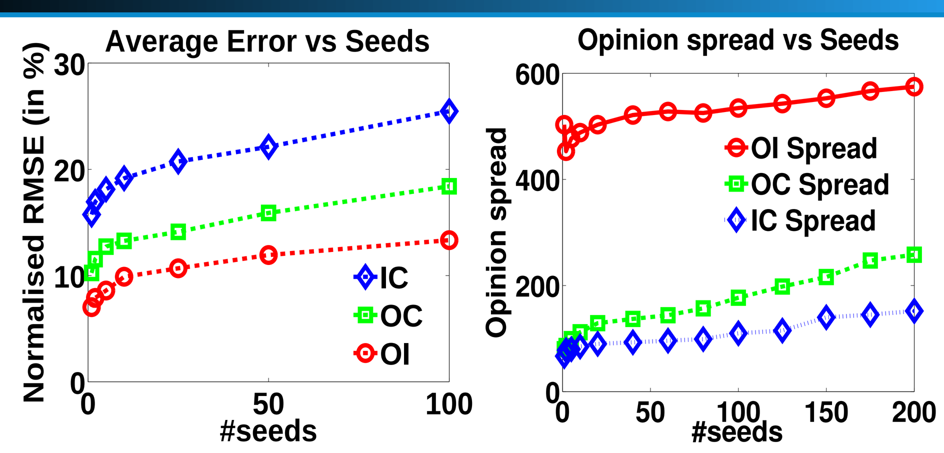
EaSyIM

- **EaSyIM** assigns a score to each node (u) of the graph
- **Intuition:** The probability of a node v to get activated by a seed node u is dependent upon all possible simple paths from u to v in G.



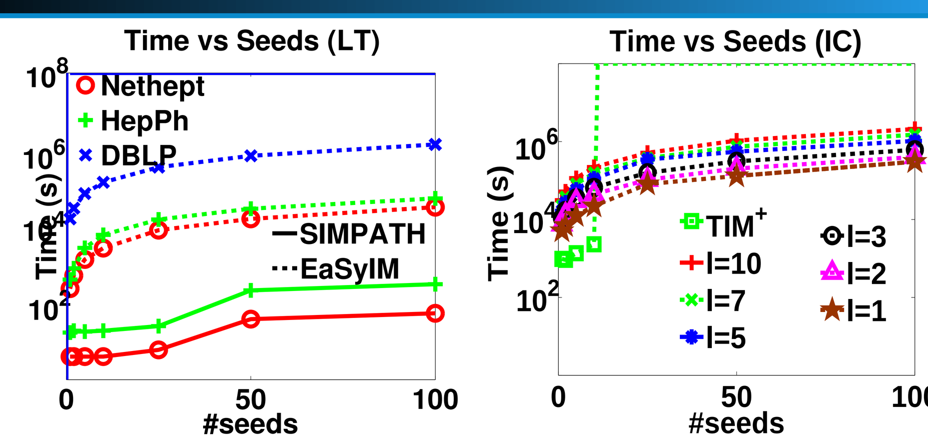
- $\Delta^l(u)$ ($\forall u \in V$) is defined as the weighted sum of the number of simple paths of length at most (l) starting from u

Motivation: Opinion Spread



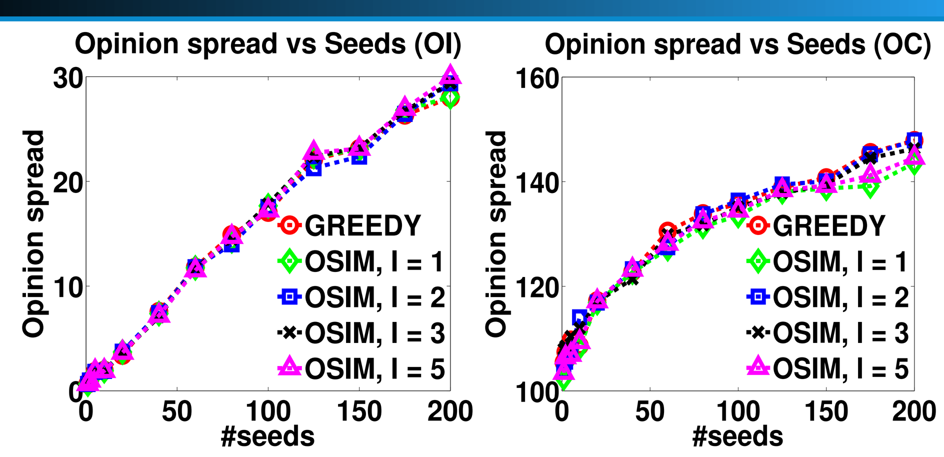
Twitter Dataset   PAKDD dataset

Efficiency



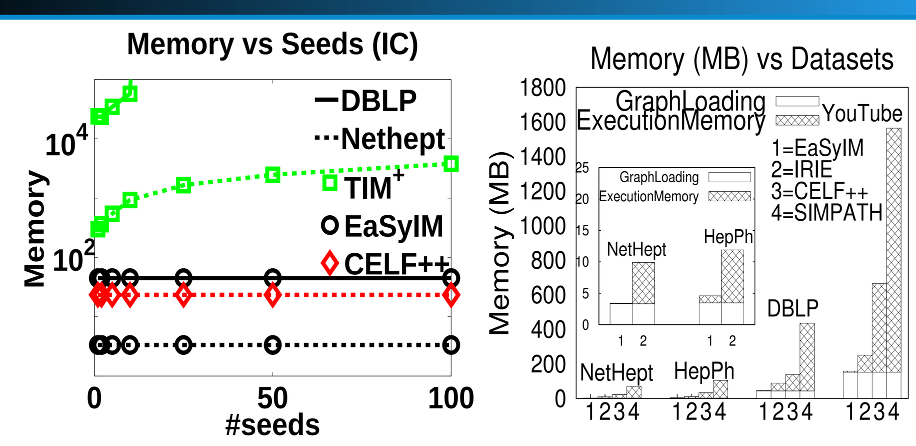
Medium Datasets   DBLP Datasets

Quality: Opinion-Aware



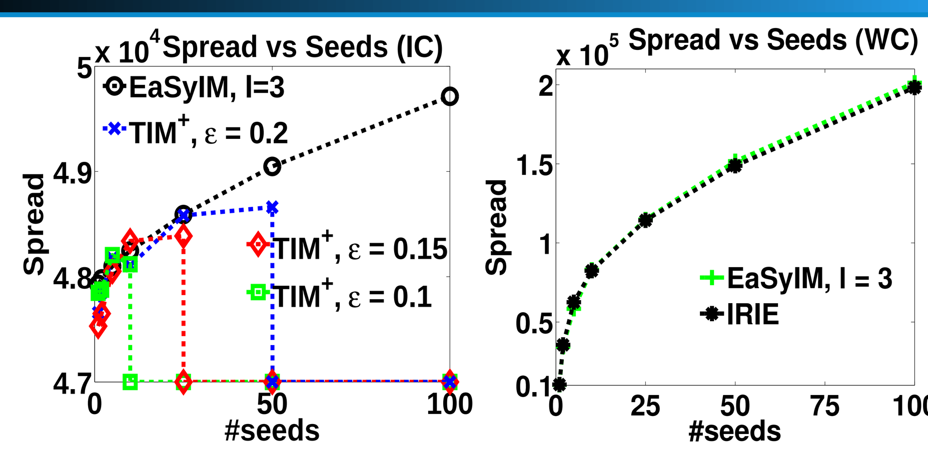
NetHEPT Dataset   HepPh dataset

Scalability



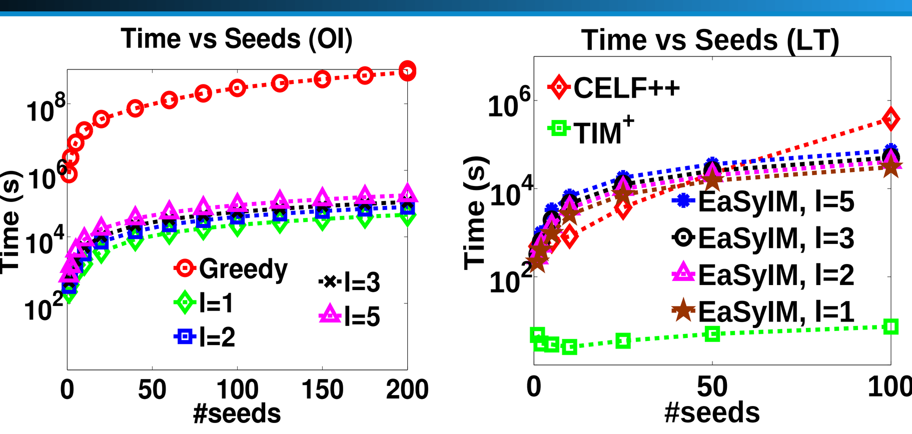
NetHEPT and DBLP   Medium Datasets

Quality: Opinion-Oblivious



DBLP Dataset   YouTube Dataset

Efficiency



NetHEPT Datasets

Result Summary

Dataset	Running Time (min)			Memory (MB)		
	CELF++	EaSyIM	Gain	CELF++	EaSyIM	Gain
NetHEPT	5352.25	118	**45.35x**	23.26	3.39	6.86x
HepPh	9746.74	230	**41x**	24.60	3.47	7.08x
DBLP	NA	5071.67	**∞**	NA	44.73	∞

Dataset	Running Time (min)			Memory (MB)		
	TIM^+	EaSyIM	Gain	TIM^+	EaSyIM	Gain
DBLP	783.1	2183	0.36x	35234.75	46.5	**758x**
YouTube	NA	5089.5	∞	NA	158.3	∞
socLive	NA	15433.33	**∞**	NA	974.94	∞

References

[1] Goyal et al. CELF++: Optimizing the greedy algorithm for influence maximization in social networks. In *WWW (Companion Volume)*, 2011.

[2] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD*, 2003.

[3] Y. Tang, X. Xiao, and Y. Shi. Influence maximization: Near-optimal time complexity meets practical efficiency. In *SIGMOD*, 2014.

[4] H. Zhang, T.N. Dinh, and M.T. Thai. Maximizing the Spread of Positive Influence in Online Social Networks. In *ICDCS*, 2013.

[5] Chen et al. Influence Maximization in Social Networks When Negative Opinions May Emerge and Propagate In *SDM*, 2011.