techvista
BANGALORE 2010

Microsoft Research

# Spread of Influence in Social Networks

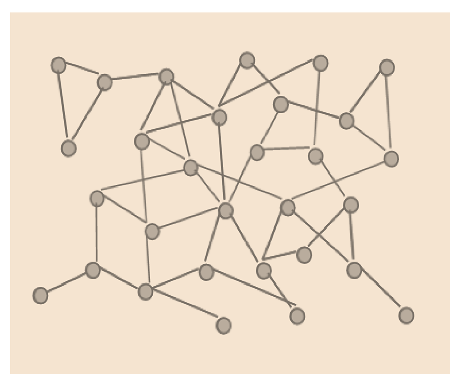
**Srinivasan Venkatramanan[1], Anurag Kumar[2]**

**Department of ECE, Indian Institute of Science, Bangalore – 560 012**

[1]vsrini@ece.iisc.ernet.in [2]anurag@ece.iisc.ernet.in

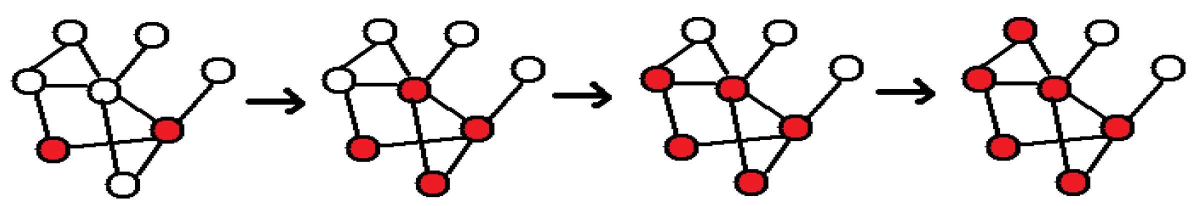
## Introduction

- **Social Network Analysis (SNA) has attracted recent interest**
  - ❖due to proliferation of online social networking sites
  - ❖due to availability of large data sets
- **SNA also helps in better understanding of**
  - ❖Effectiveness of viral marketing
  - ❖Collective human behavior
  - ❖Spread of epidemics



## The System Model

- The underlying social network is usually undirected, but the influence graph is assumed to be directed
- **Influence matrix** $\mathbf{W}=[w_{ij}]$, $\sum_i w_{i,j} \leq 1$ (usually asymmetric)
- Activation model: **Linear Threshold**
  - ❖Each node chooses a random threshold $\theta_j \sim U[0,1]$
  - ❖ Initial set $A_0$
  - ❖ $j \in A_{k+1}$ if
    $j \in A_k$ or
    $j \notin A_k$ and $\sum_{i \in A_k} w_{i,j} \geq \theta_j$
  - ❖$A_s$ is the terminal set, from which no activation can take place



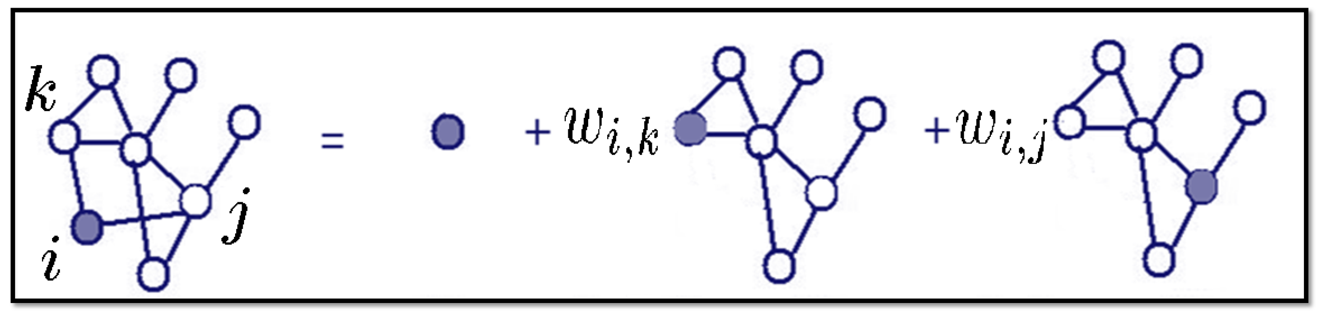
**Given an influence matrix, identify the optimal initial set A of K nodes such that the expected total influence $\sigma^{(\mathcal{N},A)} = E[|A_s|]$ is maximized.**

## Existing Work

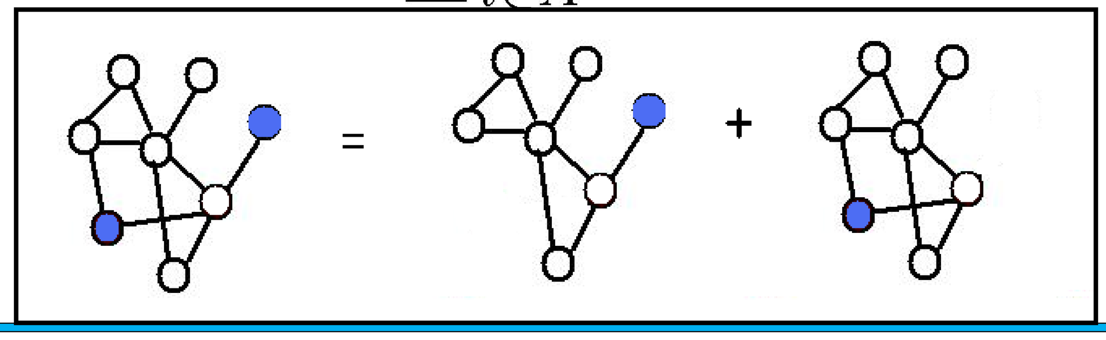
- **[Kempe et al. 2003]**
  - ❖The influence maximization problem is NP-hard
  - ❖The influence function is monotonic and submodular in A.
  - ❖Greedy algorithm provides (1-1/e) approximate solution
- Cooperative game theoretic solutions (Shapley value), and graph theoretic heuristics (degree, distance) have also been suggested

## Recursive expression for $\sigma^{(\mathcal{N},A)}$

$$\sigma^{(\mathcal{N},i)} = 1 + \sum_{j \neq i} w_{i,j} \sigma^{(\mathcal{N}\setminus\{i\},j)}$$



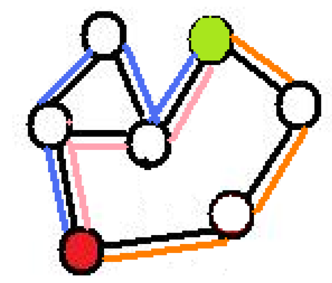

$$\sigma^{(\mathcal{N},A)} = \sum_{i \in A} \sigma^{((\mathcal{N}\setminus A)\cup\{i\},i)}$$




## DTMC Interpretation

Reversing the edges in the influence graph results in a DTMC

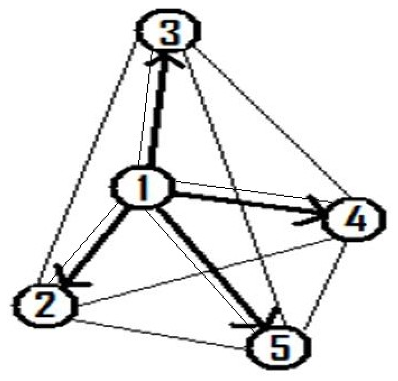
| Probability of j being activated by A in the influence process | = | Probability of reaching A through a simple path (in the DTMC) starting from j |
|---|---|---|



In the adjacent figure, there are three ways in which the influence of the red node can reach the green node.

## Examples (Complete graphs)

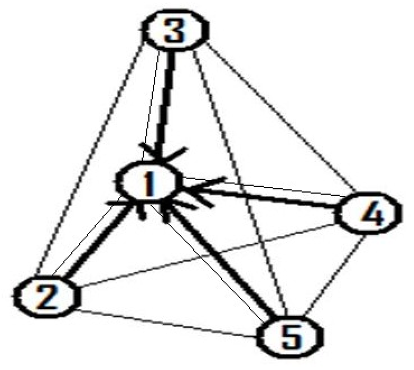
**UILT**




All bold edges have weight $\alpha_1$

Optimal initial set:
Pick nodes with **Top-K $\alpha_i$**

**USLT**



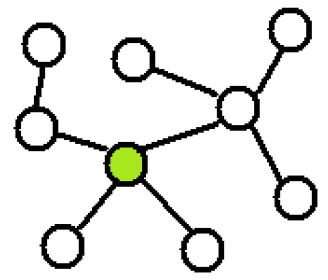

All bold edges have weight $\beta_1$

Optimal initial set:
Pick nodes with **Least-K $\beta_i$**

## Examples (Acyclic graphs)

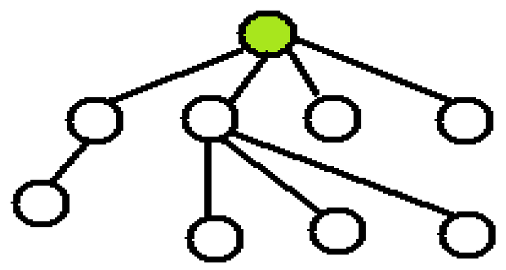
Consider an acyclic graph with $w_{i,j}$ = 1/degree(j).

We get the following result:



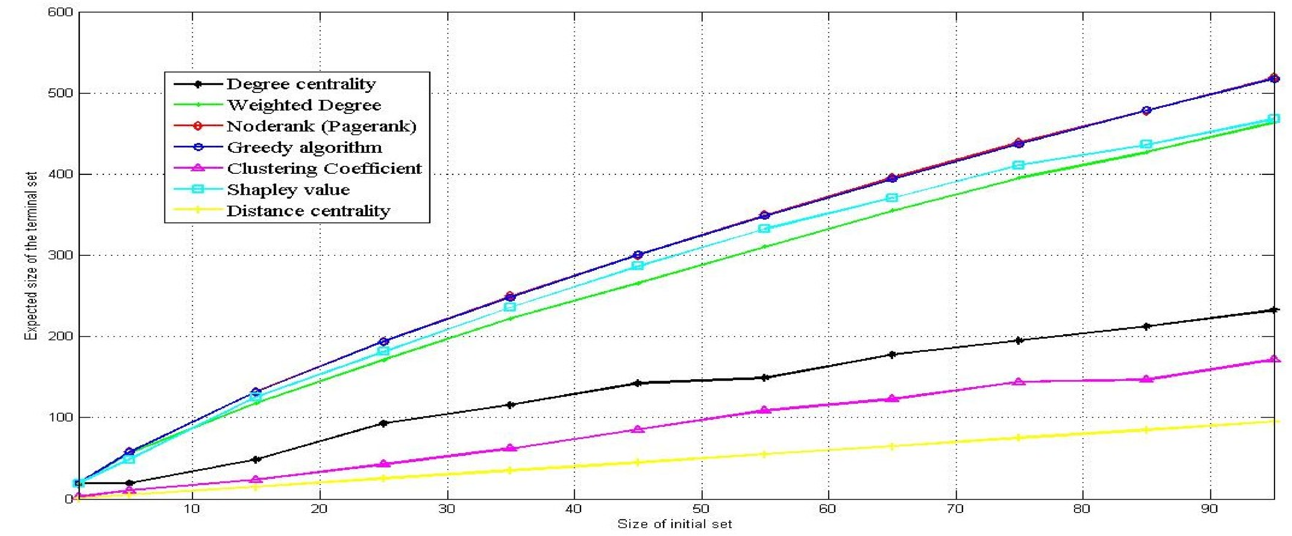
$$\sigma^{(\mathcal{N},i)} = degree(i) + 1$$

Also, one can show that the greedy solution is to pick the nodes in the order of decreasing degree.



## Simulation Results

In the above examples, the solutions provided also matches the Noderanking solution inspired by Pagerank.




**Data source: Co-authorship network data from NetScience**

## Future Work

- Understand influence spread with other activation models
- Look at other problem variants such as:
  - ❖ finding critical links for information spread
  - ❖ finding critical nodes for immunization in the context of epidemic spread

## References

**[1]** P. Domingos and M. Richardson, "Mining the Network Value of Customers", SIGKDD'01

**[2]** D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the Spread of Influence through a Social Network," SIGKDD'03