Information Dissemination in Socially Aware Networks Under the Linear Threshold Model

Srinivasan Venkatramanan and Anurag Kumar

Department of Electrical Communication Engineering,
Indian Institute of Science,
Bangalore 560 012.
{vsrini,anurag} @ece.iisc.ernet.in

30 January, 2011

Outline of Tutorial

Outline of Talk

1. Socially Aware Networks
2. Mathematical Model
3. Recursive Analytical Expressions
4. Examples
5. Final Remarks

Social Networks

- A collection of individuals with pair-wise relationships between them
  - Naturally modelled as a graph
  - Edges could be directed/undirected and weighted/unweighted
  - Traditionally of interest to sociologists and economists
- Recently much interest in the context of information and communication technologies
  - World-wide-web, online social networks, email-contact networks
  - Citation networks and coauthorship networks
  - Newman'01 [1], Brin et al. '99 [2]

Social Networks

- A collection of individuals with pair-wise relationships between them
  - Naturally modelled as a graph
  - Edges could be directed/undirected and weighted/unweighted
  - Traditionally of interest to sociologists and economists
- Recently much interest in the context of information and communication technologies
  - World-wide-web, online social networks, email-contact networks
  - Citation networks and coauthorship networks
  - Newman'01 [1], Brin et al. '99 [2]

Why Social Networks for a Communication Engineer?

- Example Scenario: Mobile Opportunistic Networks
  - People carry mobile communication devices and links form when such nodes come in "contact"
  - Data moves by *Store-and-Carry-Forward* - "Pocket Switched Networks (PSN)" (Hui et al.'08 [3])
- Mobility models need to incorporate the social behaviour of the human carriers (Chaintreau et al.'07 [4])
- Contact Models
  - On physical contact, communication link formation depends on existence of a *social link*
- Forwarding models
  - Message forwarding depends on who has been met and his/her relation to the destination in the social network
  - Also when forwarding costs are involved, nodes will be "socially selfish" - forward messages only *for* their friends (Li et al.'10 [5])

Why Social Networks for a Communication Engineer?

- Example Scenario: Mobile Opportunistic Networks
  - People carry mobile communication devices and links form when such nodes come in "contact"
  - Data moves by *Store-and-Carry-Forward* - "Pocket Switched Networks (PSN)" (Hui et al.'08 [3])
- Mobility models need to incorporate the social behaviour of the human carriers (Chaintreau et al.'07 [4])
- Contact Models
  - On physical contact, communication link formation depends on existence of a *social link*
- Forwarding models
  - Message forwarding depends on who has been met and his/her relation to the destination in the social network
  - Also when forwarding costs are involved, nodes will be "socially selfish" - forward messages only *for* their friends (Li et al.'10 [5])

Social Overlay on Communication Networks

- Imagine a social network (SN) sitting atop a communication network (CN)
  - CN governs the physical connectivity, while SN determines whether the nodes actually leverage the link
  - Can be exploited for efficient content distribution, like news broadcast, advertising, etc. (Ioannidis et al.'09 [6])
  - In our problem, we assume no mobility, and hence the nodes are always in contact, dictated by the social network

Social Overlay on Communication Networks

- Imagine a social network (SN) sitting atop a communication network (CN)
  - CN governs the physical connectivity, while SN determines whether the nodes actually leverage the link
  - Can be exploited for efficient content distribution, like news broadcast, advertising, etc. (Ioannidis et al.'09 [6])
  - In our problem, we assume no mobility, and hence the nodes are always in contact, dictated by the social network

Information Spread

Information spread occurs (in discrete time) as follows:

- We start with a few initial nodes seeded ("activated") with the message.
- At a given time instant k, a node in the network can be either:
  - Active and Infectious ($i \in D_k \subseteq A_k$) - the node has just been activated and is "ready" to forward the message to its neighbours
  - Active but not Infectious ($i \in A_k \backslash D_k$) - the node has forwarded the message to its neighbours.
  - Inactive ($i \notin A_k$) - the node is not yet "ready" to forward the message (though it might have received some copies of the message)
- An Infectious node stays infectious for exactly one time slot, forwards the message to all its neighbours and becomes non-infectious
- An inactive node becomes infectious once it has received the message from "sufficient" set of neighbours (depending on their weights of influence)

Information Spread

Information spread occurs (in discrete time) as follows:

- We start with a few initial nodes seeded ("activated") with the message.
- At a given time instant k, a node in the network can be either:
  - Active and Infectious ($i \in D_k \subseteq A_k$) - the node has just been activated and is "ready" to forward the message to its neighbours
  - Active but not Infectious ($i \in A_k \setminus D_k$) - the node has forwarded the message to its neighbours.
  - Inactive ($i \notin A_k$) - the node is not yet "ready" to forward the message (though it might have received some copies of the message)
- An Infectious node stays infectious for exactly one time slot, forwards the message to all its neighbours and becomes non-infectious
- An inactive node becomes infectious once it has received the message from "sufficient" set of neighbours (depending on their weights of influence)

Information Spread

Information spread occurs (in discrete time) as follows:

- We start with a few initial nodes seeded ("activated") with the message.
- At a given time instant k, a node in the network can be either:
  - Active and Infectious ($i \in D_k \subseteq A_k$) - the node has just been activated and is "ready" to forward the message to its neighbours
  - Active but not Infectious ($i \in A_k \backslash D_k$) - the node has forwarded the message to its neighbours.
  - Inactive ($i \notin A_k$) - the node is not yet "ready" to forward the message (though it might have received some copies of the message)
- An Infectious node stays infectious for exactly one time slot, forwards the message to all its neighbours and becomes non-infectious
- An inactive node becomes infectious once it has received the message from "sufficient" set of neighbours (depending on their weights of influence)

Information Spread

Information spread occurs (in discrete time) as follows:

- We start with a few initial nodes seeded ("activated") with the message.
- At a given time instant k, a node in the network can be either:
  - Active and Infectious ($i \in D_k \subseteq A_k$) - the node has just been activated and is "ready" to forward the message to its neighbours
  - Active but not Infectious ($i \in A_k \backslash D_k$) - the node has forwarded the message to its neighbours.
  - Inactive ($i \notin A_k$) - the node is not yet "ready" to forward the message (though it might have received some copies of the message)
- An Infectious node stays infectious for exactly one time slot, forwards the message to all its neighbours and becomes non-infectious
- An inactive node becomes infectious once it has received the message from "sufficient" set of neighbours (depending on their weights of influence)

Information Spread

Information spread occurs (in discrete time) as follows:

- We start with a few initial nodes seeded ("activated") with the message.
- At a given time instant k, a node in the network can be either:
  - Active and Infectious ($i \in D_k \subseteq A_k$) - the node has just been activated and is "ready" to forward the message to its neighbours
  - Active but not Infectious ($i \in A_k \backslash D_k$) - the node has forwarded the message to its neighbours.
  - Inactive ($i \notin A_k$) - the node is not yet "ready" to forward the message (though it might have received some copies of the message)
- An Infectious node stays infectious for exactly one time slot, forwards the message to all its neighbours and becomes non-infectious
- An inactive node becomes infectious once it has received the message from "sufficient" set of neighbours (depending on their weights of influence)

Information Spread

Information spread occurs (in discrete time) as follows:

- We start with a few initial nodes seeded ("activated") with the message.
- At a given time instant k, a node in the network can be either:
  - Active and Infectious ($i \in D_k \subseteq A_k$) - the node has just been activated and is "ready" to forward the message to its neighbours
  - Active but not Infectious ($i \in A_k \backslash D_k$) - the node has forwarded the message to its neighbours.
  - Inactive ($i \notin A_k$) - the node is not yet "ready" to forward the message (though it might have received some copies of the message)
- An Infectious node stays infectious for exactly one time slot, forwards the message to all its neighbours and becomes non-infectious
- An inactive node becomes infectious once it has received the message from "sufficient" set of neighbours (depending on their weights of influence)

Literature Survey

- Threshold models to explain collective behaviour were first put forward by Granovetter in [7], where he discussed the spread of binary decisions, among a group of rational agents (e.g. voting models)
- Domingos and Richardson [8] studied information diffusion under the viral marketing framework, and proposed the combinatorial optimization problem of finding the most influential nodes
- Kempe et al. [9] studied the influence spread problem under the LT and IC (independent cascade) model, showed it is NP-hard and gave greedy approximations for the optimal initial set
- Recent works include a general framework for cost effective outbreak detection [10] which generalizes the influence maximization problem
- In this paper we consider the LT model, derive analytical expressions and apply them to some network topologies (star, ring and tree)

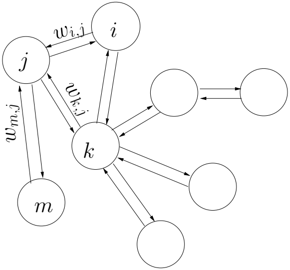
Outline of Talk

Activation Model: The Linear Threshold Model

- The social network is modeled by a directed weighted graph $\mathcal{N} = (V, E)$, with edge weights $w_{i,j}$
- We require that $\sum_i w_{i,j} \leq 1$
- At the beginning, each node j randomly chooses a threshold Θ_j distributed uniformly over [0,1]
  - The thresholds are independent across the nodes
- At step k, a node j gets activated if, it had been inactive until step $k-1$ and

$$\sum_{i \in A_{k-1}} w_{i,j} \geq \Theta_j$$

  - i.e., the total influence of the active nodes into j exceeds the threshold Θ_j



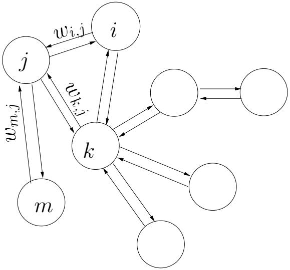
$\Sigma_i w_{i,j} \leq 1, \forall j$

Activation Model: The Linear Threshold Model

- The social network is modeled by a directed weighted graph $\mathcal{N} = (V, E)$, with edge weights $w_{i,j}$
- We require that $\sum_i w_{i,j} \leq 1$
- At the beginning, each node j randomly chooses a threshold Θ_j distributed uniformly over [0,1]
  - The thresholds are independent across the nodes
- At step k, a node j gets activated if, it had been inactive until step $k-1$ and

$$\sum_{i \in A_{k-1}} w_{i,j} \geq \Theta_j$$

  - i.e., the total influence of the active nodes into j exceeds the threshold Θ_j



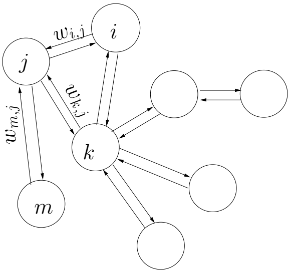
$\Sigma_i w_{i,j} \leq 1, \forall j$

Activation Model: The Linear Threshold Model

- The social network is modeled by a directed weighted graph $\mathcal{N} = (V, E)$, with edge weights $w_{i,j}$
- We require that $\sum_i w_{i,j} \leq 1$
- At the beginning, each node j randomly chooses a threshold Θ_j distributed uniformly over [0,1]
  - The thresholds are independent across the nodes
- At step k, a node j gets activated if, it had been inactive until step $k-1$ and

$$\sum_{i \in A_{k-1}} w_{i,j} \geq \Theta_j$$

  - i.e., the total influence of the active nodes into j exceeds the threshold Θ_j



$\Sigma_i w_{i,j} \leq 1, \forall j$

Notation

$\mathcal{N}$: weighted directed graph of the entire social network

$w_{i,j}$: edge weights of $\mathcal{N}$ indicating influence from i to j

$\mathbf{W}$: influence matrix with $w_{i,j}$ as entries

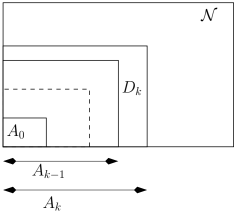
Θ_j: random threshold chosen by j uniformly from $[0,1]$

$b_j(A)$: $= \sum_{i \in A} w_{i,j}$, total influence into node j from set A

$\mathcal{A}_0$: Initial active set

A_k: Set of all active nodes at time step k

D_k: Set of infectious nodes at time step k



S: $= \arg\min_k \{A_k = A_{k-1}\}$

$g_j^{(\mathcal{N},\mathcal{A})}(k)$: $= \mathbb{P}^{(\mathcal{N})}(j \in D_k | \mathcal{A}_0 = \mathcal{A})$

$g_j^{(\mathcal{N},\mathcal{A})}$: $= \mathbb{P}^{(\mathcal{N},\mathcal{A})}(j \in A_S)$

$\sigma^{(\mathcal{N},\mathcal{A})}$: $= \mathbb{E}^{(\mathcal{N},\mathcal{A})}[|A_S|]$

Notation

$\mathcal{N}$: weighted directed graph of the entire social network

$w_{i,j}$: edge weights of $\mathcal{N}$ indicating influence from i to j

$\mathbf{W}$: influence matrix with $w_{i,j}$ as entries

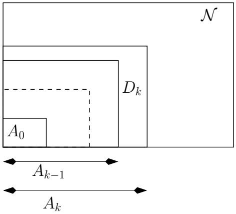
Θ_j: random threshold chosen by j uniformly from $[0,1]$

$b_j(A)$: $= \sum_{i \in A} w_{i,j}$, total influence into node j from set A

$\mathcal{A}_0$: Initial active set

A_k: Set of all active nodes at time step k

D_k: Set of infectious nodes at time step k



S: $= \arg\min_k \{A_k = A_{k-1}\}$

$g_j^{(\mathcal{N},\mathcal{A})}(k)$: $= \mathbb{P}^{(\mathcal{N})}(j \in D_k \mid \mathcal{A}_0 = \mathcal{A})$

$g_j^{(\mathcal{N},\mathcal{A})}$: $= \mathbb{P}^{(\mathcal{N},\mathcal{A})}(j \in A_S)$

$\sigma^{(\mathcal{N},\mathcal{A})}$: $= \mathbb{E}^{(\mathcal{N},\mathcal{A})}[|A_S|]$

Notation

$\mathcal{N}$: weighted directed graph of the entire social network

$w_{i,j}$: edge weights of $\mathcal{N}$ indicating influence from i to j

$\mathbf{W}$: influence matrix with $w_{i,j}$ as entries

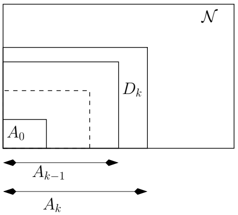
Θ_j: random threshold chosen by j uniformly from $[0, 1]$

$b_j(A)$: $= \sum_{i \in A} w_{i,j}$, total influence into node j from set A

$\mathcal{A}_0$: Initial active set

A_k: Set of all active nodes at time step k

D_k: Set of infectious nodes at time step k



S: $= \arg\min_k \{A_k = A_{k-1}\}$

$g_j^{(\mathcal{N},\mathcal{A})}(k)$: $= \mathbb{P}^{(\mathcal{N})}(j \in D_k \big| \mathcal{A}_0 = \mathcal{A})$

$g_j^{(\mathcal{N},\mathcal{A})}$: $= \mathbb{P}^{(\mathcal{N},\mathcal{A})}(j \in A_S)$

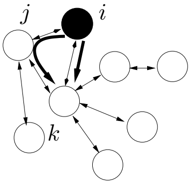
$\sigma^{(\mathcal{N},\mathcal{A})}$: $= \mathbb{E}^{(\mathcal{N},\mathcal{A})}[|A_S|]$

Outline of Talk

Activation Probabilities $g_j^{(\mathcal{N},\mathcal{A})}(k)$

Lemma

1. $j \in \mathcal{A}_0$,
   (a) $g_j^{(\mathcal{N},\mathcal{A}_0)}(0) = 1$
   (b) $g_j^{(\mathcal{N},\mathcal{A}_0)}(k) = 0$, *for all* $k > 0$
2. $j \notin \mathcal{A}_0$,
   (a) $g_j^{(\mathcal{N},\mathcal{A}_0)}(0) = 0$
   (b) $g_j^{(\mathcal{N},\mathcal{A}_0)}(k) = \sum_{l \in \mathcal{N} \setminus \{j\}} g_l^{(\mathcal{N} \setminus \{j\},\mathcal{A}_0)}(k-1)\ w_{l,j}$, *for all* $k > 0$



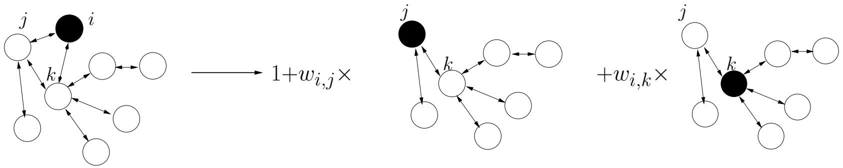
$$g_k^{(\mathcal{N},i)} = w_{i,k} + w_{i,j} \times w_{j,k}$$

Total Expected Activation due to $\{i\}$: $\sigma^{(\mathcal{N},i)}$

Theorem

Given a social network $\mathcal{N}$, with influence matrix $\mathbf{W}$, the total influence of any node i in the network under the LT model is given by

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



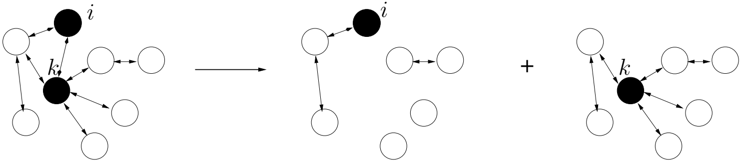
Singleton Initial Set $\{i\}$

Total Expected Activation due to Set $\mathcal{A}_0$: $\sigma^{(\mathcal{N},\mathcal{A}_0)}$

Theorem

Given a network $\mathcal{N}$ *with influence matrix* $\mathbf{W}$ *and an initial set* $\mathcal{A}_0$*, for all* $i \in \mathcal{A}_0$*, define sub-networks* $\mathcal{N}_i^{\mathcal{A}_0} = \{\mathcal{N} \setminus \mathcal{A}_0\} \cup \{i\}$*.*
Then the expected influence of the initial set $\mathcal{A}_0$ *is given by,*

$$\sigma^{(\mathcal{N},\mathcal{A}_0)} = \sum_{i \in \mathcal{A}_0} \sigma^{(\mathcal{N}_i^{\mathcal{A}_0},i)} \tag{2}$$

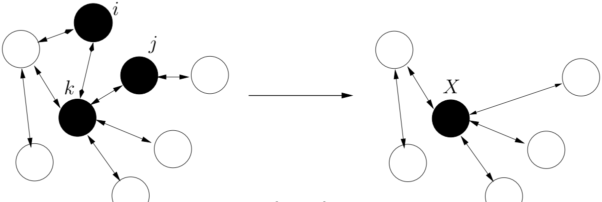


Initial Set $\mathcal{A}_0 = \{i, k\}$

Replacing an Initial Set with a “Supernode”

Theorem

Given a network $\mathcal{N}$, with influence matrix $\mathbf{W}$, an initial set $\mathcal{A}_0$ can be replaced by a supernode X as shown below. Then, $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ is equal to the influence of X in the modified network, with X being counted as $|X|$ nodes instead of 1.



$$X \equiv \{i, j, k\}$$

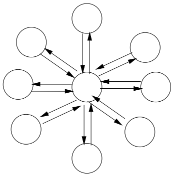
$$\forall v \notin X,$$
$$w_{X,v} = w_{i,v} + w_{j,v} + w_{k,v}$$
$$w_{v,X} = w_{v,i} + w_{v,j} + w_{v,k}$$

Replacing the initial set with a supernode

Outline of Talk

Star Topology

- Consider the star topology network $\mathcal{N}$ with N nodes including the hub
  - e.g., a central authority through which all influence/information must be disseminated
- $\alpha(\leq 1)$: influence of the hub on peripheral nodes
- $\beta(\leq \frac{1}{N-1})$: influence of each peripheral node on the hub
- We show that given K and β, there exists $\alpha^\star$ such that for $\alpha < \alpha^\star$, hub does not feature in the optimal initial set

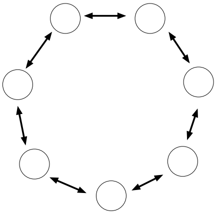


Star topology

$$\alpha^\star = \frac{K}{(N-K)^{\frac{1}{\beta}} - K(N-K-1)}$$

Ring Topology

- Consider the ring topology $\mathcal{N}$ with N homogeneous nodes
- $\alpha \leq 0.5$ denotes the influence of any node on each of its neighbours
- By using the expressions we have proved that
  - In order to maximize the influence, the K nodes should be distributed uniformly over the ring
  - Also, with $\alpha = 0.5$, for large N and small K, the influence of $A(K)$ grows as $3K$



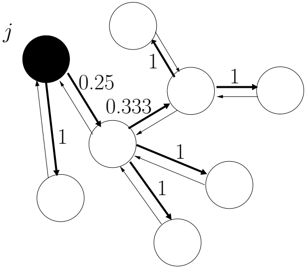
Ring topology

Degree Based Model: General Result on Influence

Theorem

Consider an acyclic undirected network $\mathcal{N}$. For $(i, j) \in E$, define $w_{i,j} = \frac{1}{\texttt{degree}(j)}$. Then, for any node $i \in \mathcal{N}$,

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



$$\sigma^{(\mathcal{N},j)} = 1 + 1 + 0.25(1 + 1 + 1 + 0.333(1 + 1 + 1)) = 3 = d_j + 1$$

Outline of Talk

Other Related Work

- We have also provided an equivalent interpretation of information spread using Acyclic path probabilities in the DTMC obtained by reversing the edges of the social network
- We have also given a heuristic algorithm (G1-sieving) for *Influence maximization* problem, based on insights from the recursive expression, which performs on par with the Greedy algorithm
- We have also studied completely connected uniform influence models and used it to derive fluid limit approximations of the information spread

Future Work

- We can consider mobile nodes (as in a Delay-Tolerant Network), where nodes are able to transfer influence only on meeting
- The problem can also be generalized to edge weights and threshold functions varying with time
- Finally, we can also study information dissemination with different activation processes, and on more generic networks to gain insights into the underlying mechanisms of information dissemination

References I

M. E. J. Newman.
The structure of scientific collaboration networks.
PNAS, 98(2):404–409, 2001.

L. Page, S. Brin, R. Motwani, and T. Winograd.
The pagerank citation ranking: Bringing order to the web.
Technical report, Stanford University Computer Science Department, 1998.

P. Hui, J. Crowcroft, and E. Yoneki.
Bubble rap: Social-based forwarding in delay tolerant networks.
In *MobiHoc*, 2008.

A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott.
Impact of human mobility on opportunistic forwarding algorithms.
In *infocom*.

References II

Q. Li, S. Zhu, and G. Cao.
Routing in socially selfish delay tolerant networks.
In *InfoComm*, 2010.

S. Ioannidis and A. Chaintreau.
Optimal and scalable distribution of content updates over a mobile social network.
2009.

Mark Granovetter.
Threshold models of collective behaviour.
American Journal of Sociology, 1978.

Matt Richardson and Pedro Domingos.
Mining knowledge-sharing sites for viral marketing.
In *ACM SIGKDD*, 2002.

References III

D. Kempe, J. Kleinberg, and E. Tardos.
Maximizing the spread of influence through a social network.
In *ACM SIGKDD*, 2003.

C. Guestrin C. Faloutsos J. Vanbriesen J. Leskovec, A. Krause and N. Glance.
Cost-effective outbreak detection in networks.
In *ACM SIGKDD*, 2007.