# Information Dissemination in Socially Aware Networks Under the Linear Threshold Model

Srinivasan Venkatramanan
Department of Electrical Communication Engineering
Indian Institute of Science
Bangalore 560012, India
Email: vsrini@ece.iisc.ernet.in

Anurag Kumar
Department of Electrical Communication Engineering
Indian Institute of Science
Bangalore 560012, India
Email: anurag@ece.iisc.ernet.in

***Abstract*—We provide new analytical results concerning the spread of information or influence under the linear threshold social network model introduced by Kempe et al. in [1], in the information dissemination context. The seeder starts by providing the message to a set of initial nodes and is interested in maximizing the number of nodes that will receive the message ultimately. A node's decision to forward the message depends on the set of nodes from which it has received the message. Under the linear threshold model, the decision to forward the information depends on the comparison of the total influence of the nodes from which a node has received the packet with its own threshold of influence. We derive analytical expressions for the expected number of nodes that receive the message ultimately, as a function of the initial set of nodes, for a generic network. We show that the problem can be recast in the framework of Markov chains. We then use the analytical expression to gain insights into information dissemination in some simple network topologies such as the star, ring, mesh and on acyclic graphs. We also derive the optimal initial set in the above networks, and also hint at general heuristics for picking a good initial set.**

## I. Introduction

Social networks are used to model interactions or interdependencies between individuals or organizations. Each individual is represented by a node, and it is connected through links to all the nodes with which it can potentially interact. Social networks play a fundamental role as a medium for the spread of information, ideas and influence among its members. Network diffusion processes have been investigated extensively in the past, with focus on spread of epidemics, diffusion of innovation and decision models.

We can imagine a social network sitting atop a communication network. The communication network governs the connectivity between any two nodes, while the social network determines whether the nodes actually leverage the connectivity. Consider the scenario in which we are interested in spreading a global information to all the nodes in the network. We seed a chosen set of initial nodes in the network with the message. The nodes in the network are socially aware i.e., a node chooses to forward the message only to those nodes with which it has a link in the social network. A node is said to be active, if it is ready to forward the message to its neighbours. The initial seed set is active by default. Other nodes will get activated only after a certain condition is met by the set of neighbours who have already forwarded the message to it, described by the activation process. The activation process could wait till a sufficient number of neighbours have sent the packet or till a highly trusted neighbour has forwarded the packet. The level of trust a node has on each of its neighbours is indicated by the weight of the edge from that neighbour, and is a measure of influence of that neighbour on the node. With this preferential forwarding and activation, it is interesting to study the process of information dissemination through the network. One quantity of interest is the expected number of nodes that would ultimately receive the message, provided we start by seeding a given initial set of nodes. Following the terminology used in [1], the term *influence* is used to refer to the potential of the set to reach out to other nodes in the network and the set that maximizes the spread is referred to as *the most influential set*. The problem of choosing the optimal initial set to maximize the spread of information, will be referred to as the *influence maximization problem*.

One possible choice for activation process is the Linear Threshold(LT) model. The concept of using threshold models to explain collective behaviour was first put forward by Granovetter in [2], where he discusses the spread of binary decisions among a group of rational agents, as in voting models. In LT model, a node will become activated only if the sum of edge weights from all the forwarding neighbours exceeds a particular threshold chosen by the node at the beginning.

This paper extends upon the body of literature in the area of viral marketing, where the *word-of-mouth* effects help spread the information to a wider set of individuals. Domingos and Richardson [3] were the first to study information diffusion under the viral marketing perspective, and were also the first to pose the combinatorial optimization problem of choosing the initial set of customers to maximize the net profits. Kempe et al. [1] studied the problem of choosing the most influential initial set using two different activation processes, and showed that the problem is NP-hard and the objective function is *submodular*. They proposed a *greedy approximation algorithm* that was shown to achieve an approximation factor of $(1-1/e)$. In [4] the authors propose a general framework for optimal sensor deployment for cost effective outbreak detection and show that the influence maximization problem is a special case of the outbreak detection problem.

**Our Contributions:** We build upon the Linear Threshold model studied by Kempe et al. [1]. Our major contributions are as follows:

- We derive recursive expressions for the expected influence of a given initial set and provide an interpretation via Acyclic Path Probabilities in Markov chains.
- We use the analytical expression on simple networks such as star, ring and mesh to obtain optimal initial sets, and provide a better understanding of the model itself.

## II. The Network Model

### A. The Communication Link Model

Here we assume that the nodes are connected by a communication infrastructure, e.g., the short message service over a cellular network. This enables any of the nodes to potentially interact with any other node in the network. An alternative is a Delay Tolerant Network (DTN), where nodes are able to transfer messages only on meeting, which also requires a mobility model to be defined. Making the communication infrastructure non-restrictive allows us to deal with restrictions in message forwarding only in the social network layer.

### B. The Social Network Model

The social network is a weighted directed graph $\mathcal{N} = (V, E)$, where the edge weight $w_{i,j}$ gives a measure of influence of node $i$ on node $j$. The activation process begins with an initial set of active nodes $\mathcal{A}_0$ and takes place in discrete time steps. Each active node forwards the message to each of its inactive neighbours. By the activation process, some of the neighbours become activated, and will forward the message in the next step. At the end of each step there are three sets of nodes: nodes that were just activated in that step (also referred to as *infectious* nodes), active nodes that are no longer infectious and the set of inactive nodes. The set of active nodes (both infectious and non-infectious) at time step $k$ is denoted by $A_k$ and the set of infectious nodes is denoted by $D_k$. The activation process stops when there are no more infectious nodes, i.e., $D_S = \emptyset$ and a *terminal set* $A_S$ is reached, from where the activation process cannot proceed further. We also assume that once a node has become active, it cannot become inactive (progressive case).

### C. Activation Models

There are two widely used activation models, namely, the *Linear Threshold model* and the *Independent Cascade model*. We will be interested in the Linear Threshold(LT) model. In the *Linear Threshold model*, $\sum_{i \neq j} w_{i,j} \leq 1$. In this model, each node $j$ randomly chooses a threshold $\Theta_j$ uniformly from [0,1] at the beginning. A given inactive node, receives the message from all its active neighbours, and gets activated once the net influence of the nodes that sent it the message exceeds the chosen threshold. In other words, a node $j$ gets activated in step $k$ if, it had been inactive until step $k-1$, i.e. does not belong to $A_{k-1}$ and

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

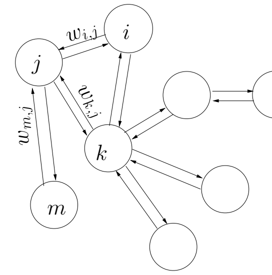


Fig. 1. Linear Threshold model

### D. Problem Statement

Given the initial set $\mathcal{A}_0$, and the LT activation process, let $A_k$ denote the set of all active nodes at time $k$. Since we are dealing with the progressive case, it is clear that $A_0 \subset A_1 \subset \ldots \subseteq \mathcal{N}$. $D_k$ denotes the set of infectious nodes at time $k$, i.e., $D_k = A_k \backslash A_{k-1}$ and $D_0 = \mathcal{A}_0$. Let $S$ denote the random stopping time at which the activation process stops, i.e., $S = \arg\min_k \{A_k = A_{k-1}\}$. Then we define $\sigma^{(\mathcal{N},\mathcal{A}_0)} = \mathbb{E}^{(\mathcal{N},\mathcal{A}_0)}[|A_S|]$ to be the expected size of the *terminal set* $A_S$, starting with $\mathcal{A}_0$ as the initial set in the network $\mathcal{N}$.

The influence maximization problem [1] is then formulated as follows. Given a $K \geq 1$,

$$\max \sigma^{(\mathcal{N},\mathcal{A}_0)} \qquad (1)$$

$$\text{s.t. } \mathcal{A}_0 \subset \mathcal{N}$$

$$|\mathcal{A}_0| = K$$

In [1], a greedy hill-climbing approach was proposed for choosing the most influential set. This involves starting with an empty set $X_0$ , and at each stage $t$ add the node that gives maximum marginal contribution to $X_{t-1}$. It is proved that this achieves an approximation factor of $(1 - 1/e)$, and the proof involves the submodularity and monotonicity of $\sigma^{(\mathcal{N},\mathcal{A})}$.

### Glossary of 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
- $\Theta_j$: 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 $k$
- $S$: = $\arg\min_k \{A_k = A_{k-1}\}$, Random time at which the activation process stops
- $g_j^{(\mathcal{N},\mathcal{A})}(k)$: = $\mathbb{P}^{(\mathcal{N})}(j \in D_k | \mathcal{A}_0 = \mathcal{A})$
- $g_j^{(\mathcal{N},\mathcal{A})}$: = $\mathbb{P}^{(\mathcal{N},\mathcal{A})}(j \in A_S)$
- $\sigma^{(\mathcal{N},\mathcal{A})}$: = $\mathbb{E}^{(\mathcal{N},\mathcal{A})}[|A_S|]$

## III. Recursive Expression for $\sigma^{(\mathcal{N},\mathcal{A}_0)}$

To the best of our knowledge, there is no work on analytically characterising $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ for the models introduced in

[1]. Moreover, $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ is generally obtained by simulating the activation process. In this section, we derive an expression for $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ in recursive form. Such an expression can be useful, since it helps us decide on the optimality for several special cases, and might also help in development of better heuristics to choose the optimal initial set, for a general case.

We have, by definition,

$$\sigma^{(\mathcal{N},\mathcal{A}_0)} = \mathbb{E}^{(\mathcal{N},\mathcal{A}_0)}[|A_S|]$$

Note that, since $D_k$'s are disjoint, and $\bigcup_{k=0}^{\infty} D_k = A_S$, we can write,

$$\begin{aligned}\sigma^{(\mathcal{N},\mathcal{A}_0)} &= \sum_{k=0}^{|\mathcal{N}|} \mathbb{E}^{(\mathcal{N},\mathcal{A}_0)}[|D_k|] \\ &= \sum_{k=0}^{|\mathcal{N}|} \mathbb{E}^{(\mathcal{N},\mathcal{A}_0)}[\sum_{j\in\mathcal{N}} I_{\{j\in D_k\}}] \\ &= \sum_{k=0}^{|\mathcal{N}|} \sum_{j\in\mathcal{N}} g_j^{(\mathcal{N},\mathcal{A}_0)}(k) \qquad (2)\end{aligned}$$

In the above expressions, $I_{\{E\}}$ denotes the indicator variable for the event $E$, and we also use the fact that the total number of time steps of the activation process is bounded above by $|\mathcal{N}|$. We define $g_j^{(\mathcal{N},\mathcal{A}_0)}(k)$ to be the probability that node $j$ is activated at the time step $k$, given that we start with $\mathcal{A}_0$ as the initial set in the network $\mathcal{N}$. The following lemma provides a recursive characterisation of $g_j^{(\mathcal{N},\mathcal{A}_0)}(k)$.

*Lemma 3.1:* In a network $\mathcal{N}$ with influence matrix $W$, starting with $\mathcal{A}_0$ as the initial set, we have,

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

*Proof:* Note that 1(a) and 2(a) are obvious, since $D_0 = \mathcal{A}_0$, chosen deterministically. 1(b) follows from 1(a) and the observation that $\sum_{k=0}^{\infty} g_j^{(\mathcal{N},\mathcal{A}_0)}(k) \le 1$ by definition. For 2(b), since $\mathcal{A}_0 \subset \mathcal{N}\setminus\{j\}$ and recalling the definition of $b_j(.)$,

$$g_j^{(\mathcal{N},\mathcal{A}_0)}(k) = \mathbb{P}^{(\mathcal{N}\setminus\{j\},\mathcal{A}_0)}\left(b_j(A_{k-2}) < \Theta_j \le b_j(A_{k-1})\right)$$

Here, $\Theta_j$ is independent of the processes $A_k$ and $D_k$, since $j$ is excluded from the network. Since $D_{k-1} = A_{k-1}\setminus A_{k-2}$, and $\Theta_j$ is chosen uniformly from [0,1], we can write,

$$\begin{aligned}g_j^{(\mathcal{N},\mathcal{A}_0)}(k) &= \mathbb{E}^{(\mathcal{N}\setminus\{j\},\mathcal{A}_0)}[b_j(D_{k-1})] \\ &= \mathbb{E}^{(\mathcal{N}\setminus\{j\},\mathcal{A}_0)}\left[\sum_{l\in\mathcal{N}\setminus\{j\}} I_{\{l\in D_{k-1}\}} w_{l,j}\right] \\ &= \sum_{l\in\mathcal{N}\setminus\{j\}} g_l^{(\mathcal{N}\setminus\{j\},\mathcal{A}_0)}(k-1)\ w_{l,j}\end{aligned}$$

■

*Remark:* It can be seen that this result depends crucially on the fact that the activation threshold of a node is distributed uniformly on $[0, 1]$.

*Remark:* Now, substituting recursively, by using Lemma 3.1 and suitably rearranging terms, we can write, For $j \notin \mathcal{A}_0$,

$$\begin{aligned}g_j^{(\mathcal{N},\mathcal{A}_0)}(1) &= \sum_{i\in\mathcal{A}_0} w_{i,j} \\ g_j^{(\mathcal{N},\mathcal{A}_0)}(2) &= \sum_{i\in\mathcal{A}_0} \sum_{\substack{k_1\neq j\\ k_1\notin\mathcal{A}_0}} w_{i,k_1} w_{k_1,j} \\ g_j^{(\mathcal{N},\mathcal{A}_0)}(3) &= \sum_{i\in\mathcal{A}_0} \sum_{\substack{k_1\neq j\\ k_1\notin\mathcal{A}_0}} \sum_{\substack{k_2\neq j,k_1\\ k_2\notin\mathcal{A}_0}} w_{i,k_1} w_{k_1,k_2} w_{k_2,j}\end{aligned}$$

and so on. Note that the terms in $g_j^{(\mathcal{N},\mathcal{A}_0)}(k)$ can be understood as representing the influence of nodes $i \in \mathcal{A}_0$ reaching node $j$ through an acyclic path of $k$ steps, *without passing through any other node in* $\mathcal{A}_0$. For each such acyclic path from $\mathcal{A}_0$ to $j$, we have a term on the right hand side, which is just the product of edge weights along that path. These can also be understood as acyclic path probabilities in a Discrete Time Markov Chain(DTMC) derived from the social network by reversing the edges. For more details on the equivalent DTMC framework and how it can be used, refer to [5].

### A. Singleton Initial Set

*Theorem 3.1:* 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

$$\sigma^{(\mathcal{N},i)} = 1 + \sum_{j\in\mathcal{N}\setminus\{i\}} w_{i,j}\sigma^{(\mathcal{N}\setminus\{i\},j)} \qquad (3)$$

*Remark:*According to theorem, under LT model, the total influence of any node $i$ in the network, is one (for the node $i$ itself) plus the weighted sum of the influences of its neighbours in the network without $i$. It is interesting to see that (Figure 2), this allows us to decompose the problem into those involving $i$'s neighbours. It is to be noted that the uniform distribution of the activation threshold is critical in allowing us to add up all influences exerted by $i$ through its different neighbours.

*Proof:* From Equation (2) we have,

$$\begin{aligned}\sigma^{(\mathcal{N},i)} &= \sum_{k=0}^{\infty}\sum_{j\in\mathcal{N}} g_j^{(\mathcal{N},i)}(k) \\ &= 1 + \sum_{j\in\mathcal{N}\setminus\{i\}} g_j^{(\mathcal{N},i)}(1) + \sum_{j\in\mathcal{N}\setminus\{i\}} g_j^{(\mathcal{N},i)}(2) + \cdots \\ &= 1 + \sum_{j\in\mathcal{N}\setminus\{i\}} w_{i,j} + \sum_{j\in\mathcal{N}\setminus\{i\}} \sum_{k_1\in\mathcal{N}\setminus\{i,j\}} w_{i,k_1} w_{k_1,j} + \cdots\end{aligned}$$

By changing variables and rearranging summations, this is equivalent to

$$\sigma^{(\mathcal{N},i)} = 1 + \sum_{k_1\in\mathcal{N}\setminus\{i\}} w_{i,k_1}\Bigg[1 + \sum_{k_2\in\mathcal{N}\setminus\{i,k_1\}} w_{k_1,k_2}\Bigg[\cdots$$

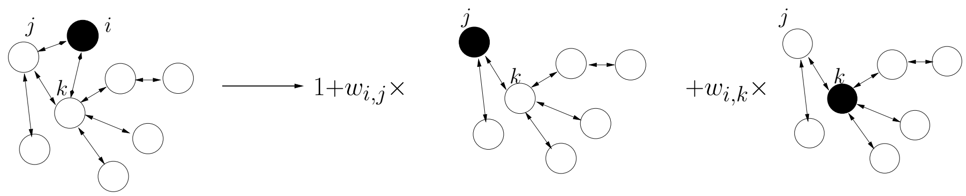


Fig. 2. Influence of a single node evaluated through its neighbours (Theorem 3.1)

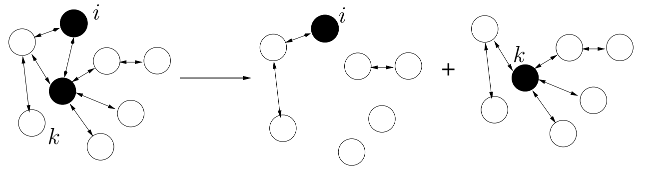


Fig. 3. Evaluating set influences through individual influences (Theorem 3.2)

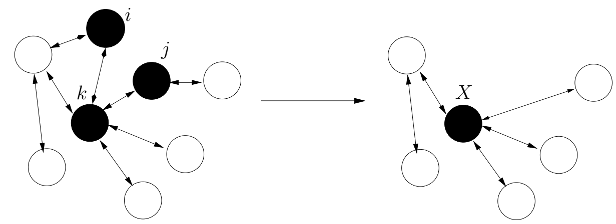


Fig. 4. Replacing the initial set with a supernode (Theorem 3.3)

Note that this equation is recursive in nature, and hence we get the above theorem. ∎

### B. Initial Set $\mathcal{A}_0$

*Theorem 3.2:* Given a social network $\mathcal{N}$ with influence matrix $\mathbf{W}$ and an initial set $\mathcal{A}_0$, define sub-networks $\mathcal{N}_i^{\mathcal{A}_0}$, for all $i \in \mathcal{A}_0$, such that,

$$\mathcal{N}_i^{\mathcal{A}_0} = \{\mathcal{N} \backslash \mathcal{A}_0\} \cup \{i\}$$

Then the influence of the initial set $\mathcal{A}_0$ is given by,

$$\sigma^{(\mathcal{N},\mathcal{A}_0)} = \sum_{i \in \mathcal{A}_0} \sigma^{(\mathcal{N}_i^{\mathcal{A}_0},i)} \tag{4}$$

*Remark:* It is interesting to see that some nodes might be left with a very small part of the network to directly influence. In Figure 3, by including $k$ in the initial set, we have essentially restricted $i$'s spread of influence to only two other nodes. To reach any other node, the influence has to pass through $k$, which is already included in the initial set.

*Proof:* The proof follows by substituting the $g_j^{(\mathcal{N},\mathcal{A}_0)}(k)$ expressions and noting that the edge weights $\{w_{i,j}, j \in \mathcal{A}_0\}$ do not have any effect on $\sigma^{(\mathcal{N},\mathcal{A}_0)}$. This allows us to split the problem of evaluating influences into $K$ sub-problems each involving only one node from $\mathcal{A}_0$. ∎

### C. Replacing an initial set with a supernode

*Theorem 3.3:* Given a social network $\mathcal{N}$, with influence matrix $\mathbf{W}$, a given initial set $\mathcal{A}_0$ can be replaced by a supernode $X$ where,

$$w_{X,v} = \sum_{i \in X} w_{i,v}$$

$$w_{v,X} = \sum_{i \in X} w_{v,i}$$

and, $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ is equal to the influence of $X$ in the modified network, with $X$ being counted as $|X|$ nodes instead of 1.

*Remark:* The theorem allows us to reduce the initial problem of evaluating influence of a set into a problem of evaluating a node's influence(see Figure 4). This is possible because of the uniform distribution of the activation threshold and the fact that influence spreads in an acyclic manner branching out from the initial set.

*Proof:* The proof follows directly by using Theorem 3.1 for each sub-network in Theorem 3.2 and combining the common terms. ∎

## IV. Examples

In this section, we will use the $\sigma^{(\mathcal{N},\mathcal{A}_0)}$ expressions on some simple networks, and obtain insights into the spread of influence and the optimal initial set in such networks.

### A. Star Topology

Let us consider the star topology $\mathcal{N}$ with $N$ nodes including the hub (see Figure 5). Let $\alpha$ be the influence exerted by the hub node on each of the peripheral nodes and let $\beta$ be the influence of any one peripheral node on the hub. By LT model, $\alpha \leq 1$ and $\beta \leq \frac{1}{N-1}$. Such a setting is possible in authorization based social networks, where any data transfer between two nodes, has to pass through a central authority.

Given $K$ as the size of initial set, we can either have one hub node and $K-1$ peripheral nodes, or $K$ peripheral nodes. Call the former set $H(K)$ and the latter $\tilde{H}(K)$. The influence functions can be readily written as follows using Theorem 3.3.

$$\sigma^{(\mathcal{N},H(K))} = K + \alpha(N-K)$$

$$\sigma^{(\mathcal{N},\tilde{H}(K))} = K + K\beta\big(1 + \alpha(N-K-1)\big)$$

Given $K$ and the values of $\alpha$ and $\beta$, one might be interested in knowing which of $H(K)$ or $\tilde{H}(K)$ is optimal, i.e. whether the hub node belongs to the optimal initial set. We see that there exists $\alpha^\star$, such that $\tilde{H}(K)$ is more *influential* than $H(K)$ for $\alpha < \alpha^\star$.

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

It is interesting to see that, there are cases where it would be wiser to leave out the hub node from the initial set, in order to maximize the spread. This might seem counter-intuitive since any message has to pass through the hub node. But, to see why $\alpha$ has to be sufficiently large, consider $\beta = \frac{1}{N-1}$ and $K = N - 1$. Then by picking $K$ peripheral nodes we get influence of exactly $N$, whereas if $\alpha < 1$,then picking $K-1$ peripheral nodes and the hub node will give us an influence strictly less than $N$.

### *B. Ring Topology*

Consider a ring topology $\mathcal{N}$ with $N$ homogeneous nodes(Figure 6). Such a topology could arise in proximity based social networks. Let $\alpha \leq 0.5$ denote the influence of any node on each of its two neighbours.

In such a network, let $A(K)$ be a set of $K$ nodes chosen, and denote the indices of the nodes in the ring by $(a_1, a_2, \dots a_K)$. Let $(l_1, l_2, \dots, l_K)$ be the number of nodes in between the chosen nodes in the ring, i.e. $l_1 = a_2 - a_1 - 1$, $l_2 = a_3 - a_2 - 1$ and so on, with $\sum_{i=1}^{K} l_i = N - K$. Then we can write the influence of $A(K)$ as,

$$\sigma^{(\mathcal{N}, A(K))} = K + 2\frac{\alpha}{1-\alpha}\big(K - \sum_{i=1}^{K} \alpha^{l_i}\big)$$

Note that this expression is maximized only if all $l_i$'s are equal. For simplicity if we assume that $K$ divides $N$, then this means that in the optimal set, $K$ nodes are equally separated along the ring. Also note that with $\alpha = 0.5$, for large $N$ and small $K$, the influence of $A(K)$ grows as $3K$. This result is not evident from the way the network is constructed and arises directly from the analytical expression.

### *C. Node Degree based Model*

We now look at a specific set of models, where the edge weight depends on the degree of the destination node. In such a case, each node weighs all its neighbours equally, and will become activated only if a sufficient number of copies of the message have been received in comparison to chosen threshold count. In this class of models, we start with an undirected graph without self-loops, whose adjacency matrix is given by $A$. Define $W$ as follows.

$$w_{i,j} = a_{i,j}/d_j \tag{5}$$

where $d_j = \sum_i a_{i,j}$ is the degree of the node $j$. Let us restrict our attention to *acyclic graphs*. We then have the following theorem.

*Theorem 4.1:* Consider an acyclic undirected graph $\mathcal{N}$ represented by the adjacency matrix $A$. Let the influence matrix be generated by Equation (5). Then, for any node $i \in \mathcal{N}$,

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

The proof of the theorem follows from applying Theorem 3.1 and noting that each edge yields exactly an expected influence of 1 irrespective of the network beyond that edge (see example in Figure 7). Note that, the local property (degree) of a node governs its global effect (the total influence of that node on the network). We direct the reader to the arXiv report[5] for a complete proof.

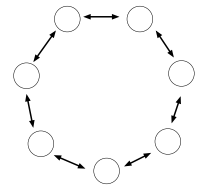

Fig. 5. Ring topology

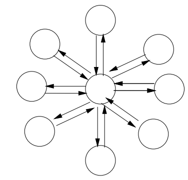

Fig. 6. Star topology

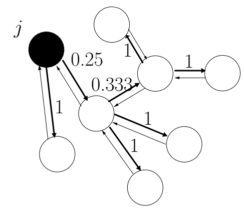


Fig. 7. Node degree based model. Example network for Theorem 4.1. $\sigma^{(\mathcal{N},j)} = 1 + 1 + 0.25(1 + 1 + 1 + 0.333(1 + 1 + 1)) = 3 = d_j + 1$

We have also studied completely connected social networks, where the edge weight $w_{i,j}$ is determined by influence level of node $i$ and susceptance level of node $j$. We have derived closed form expression for influence and identified optimal initial sets in certain special cases. We also have drawn interesting parallels between the optimal initial set and the nodes with high stationary probability values in the DTMC obtained by reversing the edges of the social network. We refer the reader to the arXiv report [5] for more details.

## V. CONCLUSION

In this paper, we have studied the information dissemination problem in a social network under the LT model and derived analytical expressions for the expected information spread achieved by an initial set. We have used it to gain insights into some simple network topologies. Several extensions are possible for this work. Firstly, we can replace the present communication link model by a DTN, where nodes are mobile, and can transfer messages 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

[1] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *ACM SIGKDD*, 2003.
[2] M. Granovetter, "Threshold models of collective behaviour," *American Journal of Sociology*, 1978.
[3] P. Domingos and M. Richardson, "Mining the network value of customers," in *ACM SIGKDD*, 2001.
[4] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. Vanbriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *ACM SIGKDD*, 2007.
[5] S. Venkatramanan and A. Kumar, "New insights from an analysis of social influence networks under the linear threshold model," http://arxiv.org/abs/1002.1335.