3. Theoretical Results

The theoretical contribution of this paper is as follows: first, Section 3.1 outlines a scheme which predicts the diffusion process of a Linear Threshold model from the perspective of a single node. Then, Section 3.2 details how this model can be utilized to optimally manipulate social connections, when the ultimate goal is to control one's own probability of being affected by an outside influence. Section 3.3 gives reasoning for using a much simpler approximation to the metric established in 3.2, which results in a faster algorithm. Experimental confirmations of the model on real-world networks are presented in Section 4.

3.1. The Predictive Model

Assume a Linear Threshold model of influence. We use an iterative approach to model the diffusion process. First, we figure out the spread of influence at each time-step. Recall from Section 2.1 that S_t is the set of activated nodes at time t. Let $q_{i,t}$ be the probability that $v_i \in S_t$. Let $p_{i,t}$ be the probability that $v_i \in S_t$, given that $v_i \notin S_{t-1}$. We see that $p_{i,t}$ almost captures the probability that v_i is activated at time t ($v_i \in S_t - S_{t-1}$), except it is conditioned on $v_i \notin S_{t-1}$, which happens with probability $1 - q_{i,t-1}$. Therefore, the probability that v_i is activated at time t is $p_{i,t}(1 - q_{i,t-1})$. We thus have that $q_{i,t} = q_{i,t-1} + (1 - q_{i,t-1})p_{i,t}$. All that remains is to find $p_{i,t}$.

To find $p_{i,t}$, we need to find the influence exerted upon v_i at time t. By Equation 2.2 from Section 2.1, the exact value for the total influence on v_i at time t is given by $\min\bigl(1, \sum_{v_j \in S_{t-1}} w_{j\to i}\bigr)$. However, this value cannot be calculated, as it is impossible to know exactly which nodes are in S_{t-1}. Even approximating the spread over all thresholds θ_i would require many iterations, as mentioned in [7]). Therefore, we must instead predict which nodes are expected to be in S_{t-1}. Using our definition of $q_{i,t}$, we find that the expected influence exerted upon v_i at time t is then $\min\bigl(1, \sum_{v_j \in V} q_{j,t-1} w_{j\to i}\bigr)$. Let this value be $\Omega_{i,t}$.

In the Linear Threshold Model, $p_{i,t}$ is the probability that $\Omega_{i,t} \geq \theta_i$, given that $\Omega_{i,t-1} \not\geq \theta_i$. Since θ_i is originally a uniform random variable from 0 to 1, the condition that $\Omega_{i,t-1} \not\geq \theta_i$ means that θ_i becomes a uniform random variable bounded between $\Omega_{i,t-1}$ and 1.

Thus, we have that:

$$p_{i,t} = \frac{\Omega_{i,t} - \Omega_{i,t-1}}{1 - \Omega_{i,t-1}} = \frac{\min\bigl(1, \sum_{v_j \in V} q_{j,t-1} w_{j\to i}\bigr) - \sum_{v_j \in V} q_{j,t-2} w_{j\to i}}{1 - \sum_{v_j \in V} q_{j,t-2} w_{j\to i}}$$

$$= \min\Bigl(1, \frac{\sum_{v_j \in V} (q_{j,t-1} - q_{j,t-2}) w_{j\to i}}{1 - \sum_{v_j \in V} q_{j,t-2} w_{j\to i}}\Bigr) \tag{3.1}$$

In equation 3.1, $\sum_{v_j \in V} q_{j,t-2} w_{j\to i}$ could exceed 1, so we must cap it to eliminate issues with nonpositive denominators. We therefore end with two equations:

$$q_{i,t} = q_{i,t-1} + (1 - q_{i,t-1})p_{i,t} \tag{3.2}$$

$$p_{i,t} = \min\Big(1, \frac{\sum_{v_j \in N_i} p_{j,t-1}(1 - q_{j,t-2})w_{j\to i}}{1 - \min(1-\epsilon, \sum_{v_j \in N_i} q_{j,t-2}w_{j\to i})}\Big) \tag{3.3}$$

In the numerator of equation 3.3, $q_{j,t-1} - q_{j,t-2}$ has been simplified to $(1 - q_{j,t-2})p_{j,t-1}$ by applying equation 3.2 at time $t - 1$. The sums in 3.3 have been simplified from $v_j \in V$ to $v_j \in N_i$, as $w_{j\to i} = 0$ for $v_j \notin N_i$.

Equations 3.2 and 3.3 completely capture the expected outcome of the diffusion process, given $p_{i,0}$ for each v_i and $w_{j\to i}$ for each pair (v_i, v_j). We now state the following theorem:

Theorem 1. *As $t \to \infty$, equations 3.2 and 3.3 predict diffusion in Linear Threshold influence networks in a manner consistent with equation 2.3, adjusted with a constant.*

Proof. First, we see that equation 3.2 is still true if one replaces $q_{i,t}$ and $q_{i,t-1}$ with $\Omega_{i,t}$ and $\Omega_{i,t-1}$. This allows us to conclude that, when $w_{j\to i}, p_{i,0} \ll 1$, $q_{i,t} \approx \Omega_{i,t}$. If $w_{j\to i}, p_{i,0} \not\ll 1$, then $q_{i,t}$ is close to 1, which we account for at the very end in Equation 3.8. Therefore, we proceed with the former case, where $w_{j\to i}, p_{i,0} \ll 1$.

From before, we have that $\Omega_{i,t} = \min\big(1, \sum_{v_j \in V} q_{j,t-1}w_{j\to i}\big) \approx \sum_{v_j \in V} \Omega_{j,t-1}w_{j\to i}$. Therefore, $\Omega_{i,t} - \Omega_{i,t-1} \approx \sum_{v_j \in V}(\Omega_{j,t-1} - \Omega_{j,t-2})w_{j\to i}$. The incremental difference in Ω at time t is thus roughly proportional to the incremental differences at time $t - 1$. For $w_{j\to i} \ll 1$, we see that the incremental differences generally decrease and approach 0 for all nodes over time, even if some nodes have a lot of neighbors. Therefore, Ω must converge over time. As $q_{i,t} \approx \Omega_{i,t}$, this suggests that for $w_{j\to i} \ll 1$, as $t \to \infty$, $q_{i,t}$ converges to some $Q < 1$.

This is consistent with the dynamics of the diffusion process for $w_{j\to i} \ll 1$, as the spread of influence will eventually stop without activating every node. Thus, the activation of nodes v_i will depend on θ_i, and this probabilistic factor is captured in $q_{i,t}$ over time.

Assuming that Q exists, we have $\sum_{v_j \in N_i} q_{j,t-2}w_{j\to i} = \Omega_{i,t-1} < 1$. Furthermore, as $q_{i,t} - q_{i,t-1} \to 0$, we have that $p_{i,t} \to 0$ by Equation 3.2. Therefore, we may eliminate the minimum functions in Equation 3.3. Plugging this simplified expression for $p_{i,t}$ into Equation 3.2 then yields:

$$q_{i,t} = q_{i,t-1} + (1 - q_{i,t-1})p_{i,t} \approx q_{i,t-1} + (1 - q_{i,t-1})p_{i,t+1} \tag{3.4}$$

$$\frac{q_{i,t} - q_{i,t-1}}{1 - q_{i,t-1}} \approx p_{i,t+1} = \frac{\sum_{v_j \in N_i}(q_{j,t} - q_{j,t-1})w_{j\to i}}{1 - \sum_{v_j \in N_i} q_{j,t-1}w_{j\to i}} \tag{3.5}$$

$$\sum_{v_j \in N_i} (q_{j,t} - q_{j,t-1}) w_{j \to i} = \Big(\frac{1 - \sum_{v_j \in N_i} q_{j,t-1} w_{j \to i}}{1 - q_{i,t-1}}\Big)(q_{i,t} - q_{i,t-1})$$

$$\approx \Big(\frac{1 - \sum_{v_j \in N_i} q_{j,t} w_{j \to i}}{1 - q_{i,t}}\Big) q_{i,t} - \Big(\frac{1 - \sum_{v_j \in N_i} q_{j,t-1} w_{j \to i}}{1 - q_{i,t-1}}\Big) q_{i,t-1} \tag{3.6}$$

Now, let $F_{i,t} = \Big(\frac{1 - \sum_{v_j \in N_i} q_{j,t} w_{j \to i}}{1 - q_{i,t}}\Big) q_{i,t} - \sum_{v_j \in N_i} q_{j,t} w_{j \to i}$. By Equation 3.6, we have that $F_{i,t} = F_{i,t-1}$. As we are taking $t \to \infty$, $F_{i,t}$ must be equal to some constant. This means that, for some C_i, we have that:

$$(1 - \sum_{v_j \in N_i} q_{j,t} w_{j \to i}) q_{i,t} - (1 - q_{i,t}) \sum_{v_j \in N_i} q_{j,t} w_{j \to i} = -C_i$$

$$q_{i,t} + C_i = \sum_{v_j \in N_i} q_{j,t} w_{j \to i} \tag{3.7}$$

which is precisely equation 2.3 adjusted for having seed nodes. When Q does not exist, all we must do is cap off probabilities, as mentioned in Section 2.1:

$$q_{i,t} + C_i = \min\big(1, \sum_{v_j \in N_i} q_{j,t} w_{j \to i}\big). \tag{3.8}$$

$\square$

3.2. Social Leverage

Consider when the goal is to maximize or minimize the probability that any individual node v_i is activated ($q_{i,t}$) through changing that node's connections ($w_{j \to i}$ for $j \in N_i$). In most settings, we want maximum control over $q_{i,t}$ at minimum cost, or minimum change in $w_{j \to i}$.

To manipulate $q_{i,t}$ efficiently, we must quantify the impact of a node v_i changing $w_{j \to i}$ (for $j \in N_i$). Rigorously, for some $v_a \in N_i$, define "social leverage" to be the quantity $\frac{\partial q_{i,t}}{\partial w_{a \to i}}$. We want to maximize social leverage by choosing the right neighbor(s).

Note that in this section we assume $\sum_{v_j \in N_i} q_{j,t-2} w_{j \to i} < 1$, else $q_{i,t} = 1$ and $\frac{\partial q_{i,t}}{\partial w_{a \to i}}$ is meaningless.

We now calculate social leverage with the equations from Section 3.1:

$$\frac{\partial q_{i,t}}{\partial w_{a \to i}} = \frac{\partial q_{i,t-1}}{\partial w_{a \to i}}(1 - p_{i,t}) + (1 - q_{i,t-1}) \frac{\partial p_{i,t}}{\partial w_{a \to i}}. \tag{3.9}$$

The $\frac{\partial q_{i,t-1}}{\partial w_{a \to i}}$ is then expanded using at time $t-1$, and so on, until we arrive at $\frac{\partial q_{i,0}}{\partial w_{a \to i}} = \frac{\partial p_{i,0}}{\partial w_{a \to i}} = 0$. Expanding and simplifying yields:

$$\frac{\partial q_{i,t}}{\partial w_{a\to i}} = \sum_{m=1}^{t}(1-q_{i,m-1})\frac{\partial p_{i,m}}{\partial w_{a\to i}}\prod_{n=m+1}^{t}(1-p_{i,n}). \tag{3.10}$$

By the definition of $q_{i,t}$, $\frac{\partial q_{i,t}}{\partial w_{a\to i}} \geq 0$ always, which is true in Equation 3.10. The underlying assumption in 3.10 is that $w_{a\to i}$ is modified at time 0, the same time as the start of the diffusion process. Modifying $w_{a\to i}$ at any other time changes only the bounds for m, and all analysis and algorithms in Sections 3.2 and 3.3 maintain their validity regardless of those bounds.

We calculate each term in 3.10 separately. We have that:

$$\frac{\partial p_{i,t}}{\partial w_{a\to i}} = \frac{(1-\sum_{v_j\in N_i} q_{j,t-2}w_{j\to i})p_{a,t-1}(1-q_{a,t-2}) + q_{a,t-2}\sum_{v_j\in N_i} p_{j,t-1}w_{j\to i}(1-q_{j,t-2})}{(1-\sum_{v_j\in N_i} q_{j,t-2}w_{j\to i})^2}. \tag{3.11}$$

All sums in 3.11 are over $v_j \in N_i$, but if we instead make all sums be over $v_j \in N_i - \{v_a\}$, the numerator of 3.11 does not change, as the $q_{a,t-2}p_{a,t-1}w_{a\to i}(1-q_{a,t-2})$ term cancels out. We make the reasonable assumption that the graphs we work on are locally tree-like [18], which means that $\frac{\partial q_{j,t}}{\partial w_{a\to i}} \approx 0$ for $v_j \in N_i$ (including v_a). Thus, as we change $w_{a\to i}$, the numerator of 3.11 remains unchanged. Meanwhile, as $w_{a\to i}$ decreases, the denominator of 3.11 ($(1-\sum_{v_j\in N_i} q_{j,t-2}w_{j\to i})^2$) must increase, so $\frac{\partial q_{i,t}}{\partial w_{a\to i}}$ necessarily decreases. Similarly, as $w_{a\to i}$ increases, $\frac{\partial q_{i,t}}{\partial w_{a\to i}}$ also increases.

We now consider two cases: where the ultimate goal is to maximize $q_{i,t}$, and where the ultimate goal is to minimize it.

If we wish to increase $q_{i,t}$ with minimal changes to $w_{j\to i}$, we focus on the $v_a \in N_i$ with largest $\frac{\partial q_{i,t}}{\partial w_{a\to i}}$. Call this the "closest" neighbor of v_i. As we increase $w_{a\to i}$, v_a only grows "closer" to v_i, so we do not need to consider any other neighbor of v_i. This makes sense intuitively: we should increase the influence of the friend who is already most influential. This shows that the metric of $\frac{\partial q_{i,t}}{\partial w_{a\to i}}$ is an appropriate measure of impact.

It is more complicated if the goal is to decrease $q_{i,t}$. Previous analysis reveals that $\frac{\partial q_{i,t}}{\partial w_{a\to i}}$ decreases with $w_{a\to i}$. Therefore, the neighbor v_a with the largest value of $\frac{\partial q_{i,t}}{\partial w_{a\to i}}$ may change. Always choosing the same neighbor, as we did before, is then almost never optimal.

At any given time, to find the most optimal $w_{a\to i}$ to reduce, we still find the closest neighbor of v_i. Again, let this neighbor be v_a. As we decrease $w_{a\to i}$, $\frac{\partial q_{i,t}}{\partial w_{a\to i}}$ also decreases. However, in order to determine when v_a is no longer the closest neighbor, we must compare $\frac{\partial q_{i,t}}{\partial w_{a\to i}}$ between different neighbors. If v_i has many neighbors, this is not easy to solve. Therefore, we instead run an algorithm which decrements $w_{a\to i}$ by a very small discrete amount δ, and recalculates which neighbor of v_i is closest to v_i. This algorithm is only perfectly optimal as $\delta \to 0$, but for small δ the error is still minimal.

With the simplifying assumption that the cost for decreasing $w_{a\to i}$ is linear, we formalize the algorithm as follows:

Algorithm 1 Effective Reduction of Influence

procedure REDUCTION(v_i, t) ▷ t is the time limit
  Calculate p, q up to time t based on weights
  for $v_j \in N_i$ **do**
    Calculate $\frac{\partial q_{i,t}}{\partial w_{j\to i}}$
  $v_a \leftarrow v_j$ with largest value of $\frac{\partial q_{i,t}}{\partial w_{j\to i}}$ ▷ find the closest neighbor
  $w_{a\to i} \leftarrow w_{a\to i} - \delta$ ▷ weights are updated

3.3. *A Potential Simplification*

Calculating the values in Equation 3.10 can be very time-consuming. We attempt to simplify Equation 3.10 so that Algorithm 1 can be made more efficient. Specifically, notice that the expected amount of influence contributed by a specific $v_j \in N_i$ to v_i by time t is approximately $q_{j,t-1}w_{j\to i}$. This motivates us to approximate $\frac{\partial q_{i,t}}{\partial w_{a\to i}} \approx q_{a,t-1}$. We formalize this in the following theorem:

Theorem 2. *If* $\sum_{v_j\in N_i} q_{j,m-2}w_{j\to i} \ll 1$*, then* $q_{a,t-1}$ *accurately approximates* $\frac{\partial q_{i,t}}{\partial w_{a\to i}}$.

Proof. Consider the numerator on the right hand side of Equation 3.11:

$$(1 - \sum_{v_j\in N_i} q_{j,t-2}w_{j\to i})p_{a,t-1}(1 - q_{a,t-2}) + q_{a,t-2}\sum_{v_j\in N_i} p_{j,t-1}w_{j\to i}(1 - q_{j,t-2}).$$

If we have that $\sum_{v_j\in N_i} q_{j,m-2}w_{j\to i} \ll 1$, then $p_{j,t-1}, w_{j\to i}$ are even smaller, so the second term of this expression is negligible. We therefore have that:

$$\frac{\partial p_{i,t}}{\partial w_{a\to i}} \approx \frac{p_{a,t-1}(1 - q_{a,t-2})}{(1 - \sum_{v_j\in N_i} q_{j,t-2}w_{j\to i})} = \frac{(q_{a,t-1} - q_{a,t-2})}{(1 - \sum_{v_j\in N_i} q_{j,t-2}w_{j\to i})}. \tag{3.12}$$

Now, separately consider each term in Equation 3.10. As $p_{j,t} \ll 1$, we use a linear approximation for the products. Putting everything together, we have that:

$$\prod_{n=m+1}^{t}(1-p_{i,n}) \approx 1-\sum_{n=m+1}^{t} p_{i,n} \approx 1-\sum_{n=m+1}^{t} p_{i,n}(1-q_{i,n-1})$$

$$= 1-\sum_{n=m+1}^{t} q_{i,n}-q_{i,n-1} = 1-(q_{i,t}-q_{i,m}). \tag{3.13}$$

$$\frac{\partial q_{i,t}}{\partial w_{a\to i}} = \sum_{m=1}^{t}(1-q_{i,m-1})\frac{\partial p_{i,m}}{\partial w_{a\to i}}\prod_{n=m+1}^{t}(1-p_{i,n})$$

$$\approx \sum_{m=1}^{t}(1-q_{i,m})\frac{\partial p_{i,m}}{\partial w_{a\to i}}\big(1-(q_{i,t}-q_{i,m})\big)$$

$$\approx \big(1-q_{i,t}\big)\sum_{m=1}^{t}\frac{(q_{a,m-1}-q_{a,m-2})}{(1-\sum_{v_j\in N_i} q_{j,m-2}w_{j\to i})}. \tag{3.14}$$

When $\sum_{v_j\in N_i} q_{j,m-2}w_{j\to i} \ll 1$, the denominators in the last sum of Equation 3.14 have little to no impact on the value of $\frac{\partial q_{i,t}}{\partial w_{a\to i}}$. Since the only pertinent action involving $\frac{\partial q_{i,t}}{\partial w_{a\to i}}$ is comparing it across different v_a, we can ignore the denominators completely. Therefore, the last sum in Equation 3.14 simply becomes $\sum_{m=1}^{t}(q_{a,m-1}-q_{a,m-2})$, which telescopes and equals $q_{a,t-1}$. Therefore, $q_{a,t-1}$ is a good approximation of $\frac{\partial q_{i,t}}{\partial w_{a\to i}}$ when $\sum_{v_j\in N_i} q_{j,t-2}w_{j\to i} \ll 1$. □

Since $\sum_{v_j\in N_i} q_{j,t-2}w_{j\to i} \ll 1$ very often, we propose a new algorithm which is optimal in most cases and reduces running calculation time significantly.

Algorithm 2 Effective Reduction of Influence - Approximation

procedure ReductionApprox(v_i, t, D) ▷ t is the time limit, $D \ll 1$ is a constant
  Calculate p, q up to time t based on weights
  for $v_j \in N_i$ **do**
    if $\sum_{v_j\in N_i} q_{j,t-2}w_{j\to i} < D$ **then** ▷ if $\sum_{v_j\in N_i} q_{j,t-2}w_{j\to i} \ll 1$
      $b_j \leftarrow q_{j,t-1}$
    else
      $b_j \leftarrow \frac{\partial q_{i,t}}{\partial w_{j\to i}}$
  $v_a \leftarrow v_j$ with largest value of b_j ▷ find the closest neighbor
  $w_{a\to i} \leftarrow w_{a\to i} - \delta$ ▷ weights are updated