Information diffusion as a control model

November 14, 2014

1 Introduction

We consider the problem of controlling the information diffusion in social networks. A social network can be seen as a directed graph $G = (V, E)$. Each node in the G can be seen as a user and each directed edge can be seen as a relationship. Let $|V| = n$ and $|E| = m$. We can assume $m = O(n)$ for a social network G (Sparsity assumption).

Information diffusion is the general problem on how information propagates in a social network. Let a message be initially distributed from a set of nodes (small number) in G. It is interesting to study how wide/fast/much cost this message spread over G.

2 Information Diffusion Model

We present our model of information diffusion based on [Rodriguez 14]. We consider the context where a message is distributed from one node to another at discrete time steps $\{t\}_{t \in \mathbb{N}}$. When a message is spread from a node j at time t_j, it can reach each of its neighbor node i at any time afterwards (i is a neighbor node of j if $(j, i) \in E$). This message transmission may or may not succeed at any given time $> t_j$. But when it does succeed at time $\bar{t}_i$, $t_i - 1 < \bar{t}_i \leq t_i$, node i will be ready to spread this message to its own neighbors at the next time step t_i. In that case, we say node i is **activated** at t_i.

For every pair of node (j, i), we model the transmission of message from j to i by a pdf. Namely, let $f(t_i \,|\, t_j; \alpha_{j,i})$ be the pdf that i is activated at t_i by a message from j conditioned on that j is activated at time t_j, where $\alpha_{j,i}$ is a coefficient parameter of $f(t_i \,|\, t_j; \alpha_{j,i})$ representing the transmission rate from j to i. Notice that $f(t_i \,|\, t_j; \alpha_{j,i})$ should be 0 whenever $t_i \geq t_j$

or $\alpha_{j,i} = 0$ (there is no edge from j to i). Three popular models for this transmission distribution are shown below.

1. Exp.
$$f(t_i \,|\, t_j; \alpha_{j,i}) = \begin{cases} \alpha_{j,i} \cdot e^{-\alpha_{j,i}(t_i - t_j)} & \text{if } t_i > t_j \\ 0 & \text{otherwise} \end{cases}$$

2. Pow.
$$f(t_i \,|\, t_j; \alpha_{j,i}) = \begin{cases} \frac{\alpha_{j,i}}{\delta}(\frac{t_i - t_j}{\delta})^{-1-\alpha_{j,i}} & \text{if } t_i > t_j + \delta \\ 0 & \text{otherwise} \end{cases}$$

3. Ray.
$$f(t_i \,|\, t_j; \alpha_{j,i}) = \begin{cases} \alpha_{j,i}(t_i t - j)e^{-\frac{1}{2}\alpha_{j,i}(t_i - tj)^2} & \text{if } t_i > t_j \\ 0 & \text{otherwise} \end{cases}$$

Based on the above probabilistic transmission model, we model the information diffusion by a discrete sequence of $\{X_t\}_{t \in \mathbb{N}}$, where $X_t = [x_1^{(t)}, x_2^{(t)}, \ldots, x_n^{(t)}]^\top$ with each $x_i^{(t)}$ representing the probability that node i **has been activated** at time t (i could be activated at any time $\leq t$).

3 Model with perfect information

Assume an agent has perfect information about this model. That is, the agent knows all $\alpha_{j,i}$, $f(t_i \,|\, t_j; \alpha_{j,i})$ and keeps record of $X_1, \ldots, X_t$. Then, he can precisely compute X_{t+1} by the following algorithm.

1. For each node $i \in [n]$, we have
$$x_i^{(t+1)} = x_i^{(t)} + (1 - x_i^{(t)})P_i^{t+1}, \tag{1}$$
   where P_i^{t+1} denote the probability that node i is activated at $t + 1$.

2. To compute P_i^{t+1}, we need to consider the transmission from each j with $\alpha_{j,i} \neq 0$. For an arbitrary node j, the probability that i is activated at $t + 1$ due to the transmission from j conditioned on that j is activated at time t_j is
$$F^{(t+1)}(i \,|\, t_j; \alpha_{j,i}) = \int_t^{t+1} f(t_i | t_j; \alpha_{j,i}) \, dt_i. \tag{2}$$

Hence, the probability that i is activated at $t+1$ due to the transmission from j is the summation of this conditional probability.

$$F^{(t+1)}(i \mid \alpha_{j,i}) = \Sigma_{t_j=1}^t F^{(t+1)}(i \mid t_j; \alpha_{j,i}) \cdot P_j^{(t_j)}. \qquad (3)$$

We assume each message transmission is independent. Then, P_i^{t+1} can be computed by the following equation.

$$P_i^{(t+1)} = 1 - \Pi_{j:\alpha_{j,i} \neq 0}(1 - F^{(t+1)}(i \mid \alpha_{j,i})). \qquad (4)$$

3. Given X_t and X_{t-1}, each P_i^t can be easily computed by Equation 1. For all t, i, j, $F^{(t)}(i \mid t_j)$ can be precomputed by Equation 2.

4   Model with a controller

Now, let's assume there is an outside controller who wants to manipulate the information diffusion.

First, consider the case when the controller wants to reinforce the information diffusion. The goal is to spread a message to certain degree (e.g. The average probability achieves certain threshold) by minimum utility and time cost. Let t denote the time step that the goal is accomplished. Specifically, we consider the goal that can be written as the following form.

$$AX_t \geq B \qquad (5)$$

1. What the controller can do is to make some of the central nodes (e.g. a social media) in the network "broadcast" the message. Let $I \subseteq V$ be the set of central nodes, where $|I| = k$. Such central node j usually have a lot of neighboring nodes i. When j make a broadcast at time t_j, it acts like it is "reactivated". That is, j will spread the message to each of its neighbor i following some distribution g, where $g(t_i \mid t_j; \alpha_{j,i})$ is the pdf that i is activated at t_i by a message from j conditioned on j is broadcasting at t_j. We assume all broadcasting and regular message transmission operates independently.

2. Making j broadcast the message comes with a utility cost, which depends on its "centrality" γ_j. This centrality γ_j can be simply the degree of j, or any relevant measures that represent its power in terms of spreading the message. We use $u_t(\gamma_j)$ to denote the utility cost of making j broadcast at time t, where u_t is some time-dependent function. Let $b_t(j)$ be the decision indicator of whether or not making node

j broadcast at time t. (As an easy extension, we can allow $b_t(j)$ to be a probabilistic decision, where $b_t(j) \in [0,1]$ indicating the probability of making j broadcast at time t.)

$$b_t(j) = \begin{cases} 1 & \text{if make } j \text{ broadcast at time } t \\ 0 & \text{otherwise} \end{cases}$$

Then, the total utility cost to time t is the follows.

$$\Sigma_{l=1}^{t} \Sigma_{j \in I} u_t(\gamma_j) b_t(j) \tag{6}$$

3. We also consider a general time cost $\mu(t)$, where μ is some monotonically increasing function. The information diffusion controlling problem can be represented by the following optimization problem.

$$\begin{aligned} min & \quad \Sigma_{l=1}^{t} \Sigma_{j \in I} u_t(\gamma_j) b_t(j) + \mu(t) \\ \text{such that} & \quad AX_t \geq B \end{aligned}$$

5 Model with a controller and Imperfect information

In practice, perfect information is often infeasible in social networks. It is very difficult to either access or store all there data. Therefore, the controlling the information diffusion with imperfect information is a more meaningful problem.

In terms of imperfect information, we can assume that the controller only knows a small subset of $\alpha_{j,i}$, $f(t_i \mid t_j; \alpha_{j,i})$, and only keep record of $\{Y_t\}, Y_t \subseteq X_t$, where $|Y_t| << n$.