Going Viral with the Subscription Business Model

Giovanni Viglietta
JAIST
johnny@jaist.ac.jp

1 Introduction

Companies with a *limited promotional budget* that want to introduce a new product in a *social network* know the importance of carefully picking a *target set*: a set of individuals to whom the product is directly marketed by the company. The idea is that customers do not only have an *individual value*, but also a *network value*: the initial set of adopters will spontaneously advertize the product to their friends, starting a cascade of influence that will spread through the network, causing larger numbers of people to adopt the product [3].

Several models of *influence spreading* in social networks have been studied, starting with the one in [5], where the network is a graph whose nodes have *thresholds*: as soon as the number of *active* neighbors of a node reaches the threshold, the node will become active as well. More advanced models introduce a limited *time window* during which an active node exerts influence on its neighbors [4], *incentives* to reduce node thresholds [2], influence whose intensity *decays* over time, etc. It is known that the problem of choosing a minimum-size target set that will cause the entire network to adopt the product is hard even to approximate [1], but several efficient algorithms are known for special classes of networks in various models [2, 4, 5].

My interest in these problems began in 2015, when I was a Postdoctoral Researcher in Ottawa under Paola Flocchini and Nicola Santoro: we had some research meetings with Joseph Peters, who visited from Simon Fraser University and introduced us to the subject. In the summer of 2016 I traveled to Vancouver for CCCG, and I met Joseph Peters again, who invited me to spend a few days at SFU. During that period, we developed some ideas and studied alternative models of influence spreading: this paper contains my recollections of this experience and a summary of some of our findings.

I will focus on networks with *non-persistent influence*, motivated by the fact that some products or services follow the *subscription business model*: access is given only to subscribers, who periodically decide whether to renew their subscription or not. Here the nodes of the network can not only become active, but also *inactive*, if they do not receive enough influence from their neighbors. However, every time a node is activated, it stays active at least for a *time window*, no matter what the activation state of its neighbors is.

During my stay at SFU, I also wrote a *simulator* supporting several influence-spreading models, including the one with non-persistent influence. The simulator and its source code can be found here: `http://giovanniviglietta.com/files/influence/influence.zip`.

2 Model Definition

Let $G = (V, E)$ be a finite graph modeling the *network*, and let $N(v)$ denote the set of neighbors of a node $v \in V$ (by definition, $v \notin N(v)$). Let $t\colon V \to \mathbb{N}$ be the *threshold function*, and let $\lambda \in \mathbb{N}$ be the size of the *time window*. For every $S \subseteq V$, we define two functions $a_S : V \times \mathbb{N} \to \{0, 1\}$ and $w_S : V \times \mathbb{N} \to \mathbb{N}$ such that, for every $\ell \in \mathbb{N}$,

$$\begin{aligned}
a_S(v, 0) &= \begin{cases} 1 & \text{if } v \in S \\ 0 & \text{if } v \notin S \end{cases} \\
a_S(v, \ell + 1) &= \begin{cases} 1 & \text{if } \sum_{u \in N(v)} a_S(u, \ell) \geq t(v) \\ 1 & \text{if } a_S(v, \ell) = 1 \text{ and } w_S(v, \ell + 1) < \lambda \\ 0 & \text{otherwise} \end{cases} \\
w_S(v, 0) &= 0 \\
w_S(v, \ell + 1) &= (w_S(v, \ell) + 1) \cdot a_S(v, \ell)
\end{aligned}$$

S represents the *target set*, i.e., the set of initially active nodes. $a_S(v, \ell)$ is the *activation function*, which is 1 if and only if v is active at time ℓ, and $w_S(v, \ell)$ keeps count of the number of consecutive rounds v has been active before time ℓ.

According to the definition of a_S, after the initial round, a node becomes active whenever it receives influence from a number of active neighbors that is at least as large as its threshold, and then it remains active for λ rounds. After this time window, the node becomes inactive as soon as the number of its active neighbors gets smaller than its threshold.

We call a set of nodes $W \subseteq V$ *self-supporting* if each node of $v \in W$ has a number of neighbors in W that is at least as large as its threshold $t(v)$. Such sets have interesting properties: clearly, the union of self-supporting sets is again self-supporting, and whenever all the nodes of a self-supporting set are active, they are guaranteed to remain active forever. Picking a self-supporting target set has a desirable effect: in the subsequent rounds, the active nodes will form a weakly growing chain of self-supporting sets, and therefore no influence will ever be "lost".

3 Influencing the Whole Network

A typical question when studying influence-spreading problems is how to influence the whole network with the minimum budget. In other terms, we seek the *smallest target set* that will eventually cause *all* the nodes in the network to be *permanently active*.

Unfortunately, even finding approximate solutions to this problem is hard. In fact, we can give a simple approximation-preserving reduction from Set Cover, which is known to be NP-hard to approximate within a factor of $(1 - \varepsilon) \ln n$, for any $\varepsilon > 0$. As shown in Figure 1, we model each element of the universe set U as a node with threshold 1. Then, for every given set S_i in the Set Cover instance, we add a node connected to the elements of U that are in S_i, and we set the threshold of this new node to be equal to its degree. Clearly, the most efficient way to choose a target set is to ignore the nodes in U and just select some S_i's. If we manage to cover all of U with these S_i's, then the entire network will be influenced within 3 rounds (the case $\lambda = 1$ is an exception, because we also have to select all nodes in U).

Proposition 1 *The minimum size of a target set that will cause all n nodes in a network to be permanently active is NP-hard to approximate within a factor of* $(1 - \varepsilon) \ln n$*, for any* $\varepsilon > 0$.

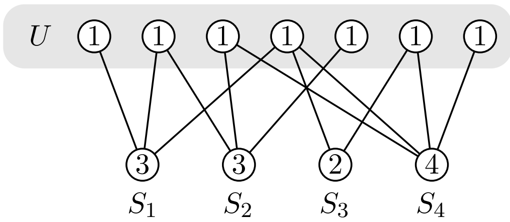

図 1: Approximation-preserving reduction from the Set Cover instance with universe set $U = \{1, 2, 3, 4, 5, 6, 7\}$ and subsets $S_1 = \{1, 2, 4\}$, $S_2 = \{2, 3, 5\}$, $S_3 = \{4, 6\}$, and $S_4 = \{3, 4, 6, 7\}$. The numbers inside nodes denote their thresholds.

We showed that spreading influence in general networks is hard, but what if we restrict our analysis to a smaller class of graphs? The above reduction produces a *bipartite graph*, so bipartite networks are hopeless, but what about *complete graphs*? It turns out that here a minimum-size target set can be found efficiently: the crucial observation is that, if $\lambda > 1$, the number of active nodes will be monotonic (increasing or decreasing) after the initial λ rounds. Therefore, after $O(n)$ rounds, the configuration of active nodes will be stable. Also, it is always convenient to pick the highest-threshold nodes as the target set. As a consequence, after sorting the nodes by their thresholds in $O(n \log n)$ time, simulating $O(n)$ rounds from a given target set of highest-threshold nodes takes only $O(n)$ time. Hence the optimal target set can be found by binary search in $O(\log n)$ iterations of $O(n)$-time simulations.

Proposition 2 *In a complete network of size n, a minimum-size target set that will cause all nodes to be permanently active can be computed in time* $O(n \log n)$.

What if the network is a *path* or a *cycle*? Here the observation is that, if two neighboring nodes with threshold 2 are inactive at the same time, they will never be activated. So, at least one of them must be in the target set, and the other one must become active within λ rounds. Actually, if $\lambda = 1$, both of them and their neighbors must be in the target set: this suggests a linear-time algorithm for selecting a minimum-size target set in this special case. Moreover, if $\lambda > 1$, activating a single node in a subpath of nodes of threshold 1 will cause the whole chain to become active. This also leads to a linear-time algorithm.

Proposition 3 *In a network whose graph is a path or a cycle, a minimum-size target set that will cause all nodes to be permanently active can be computed in linear time.*

4 Activation Periods

Observe that any given network of n nodes with a time window of size λ has only a finite number of "configurations": at each round, for each node, the only relevant parameter is how many rounds have passed since the node was last inactive. This number can range from 0 (i.e., the node is currently inactive) to λ. Indeed, in our model of activation, any value larger than λ can be treated as being λ (cf. the definitions in Section 2, where the counter w_S is only ever compared with λ). This yields a total number of configurations of $(\lambda + 1)^n$ for the whole network.

Since the states of the network are finitely many, after some rounds a state is certainly repeated, and therefore the sequence of states is *eventually periodic*, because the network evolves in a deterministic way. From a theoretical standpoint, we may ask ourselves a fundamental question: What periods are possible for a given λ?

First of all, observe that the period must be either 1 or greater than λ. Indeed, $\lambda + 1$ is the minimum time required for any node to become active and then inactive. Also, a period of 1 is trivial to obtain in every network by picking the empty target set. If $\lambda = 1$, a period of 2 is also easy to produce: take the network consisting of two neighboring nodes with threshold 1, and activate only one of them. It can be proven that periods greater than 2 are impossible if $\lambda = 1$. However, if $\lambda > 1$, any period greater than λ is obtainable: Figure 2 shows an example that can be easily generalized.

Proposition 4 *If* $\lambda = 1$, *the possible periods are only* 1 *and* 2. *If* $\lambda > 1$, *all periods are possible, except* 2, 3, …, λ.

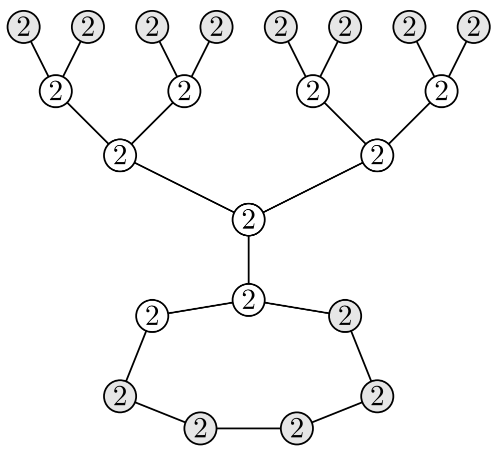

図 2: If $\lambda = 5$, the activation sequence of this network has period 7. All nodes have threshold 2, and the shaded nodes constitute the target set.

What if we consider special classes of networks? Let us assume $\lambda > 1$. From the discussion preceding Proposition 2, we already know that *complete graphs* must have period 1: indeed, the configuration of active nodes must become stable after $O(n)$ rounds. Also, if the network is a *cycle*, it is easy to see that certain areas must become permanently active or permanently inactive, and therefore act as "walls" for the propagation of influence. It follows that the period is 1 in cycles, as well. A similar phenomenon is observed in *tree* networks, where a chain of deactivation events cannot "rebound" off leaves.

Proposition 5 *If $\lambda > 1$, the activation sequence of a network whose graph is either complete or a cycle or a tree has period* 1.

Propositions 4 and 5 acquire a suggestive meaning when a social network with non-persistent influence is reinterpreted as a crude model for a *neural network*: nodes now represent *neurons* which can be activated or deactivated based on a threshold function whose inputs are the signals received from neighboring neurons. Under this shift of perspective, λ models the "persistence of information" within single neurons.

Now, Proposition 4 expresses the fact that *memory is not an emergent behavior*, in that neurons with no persistence of information (i.e., with $\lambda = 1$) give rise to networks without memory (i.e., which are either stable or oscillate between two states). On the other hand, even the smallest amount of persistence (i.e., $\lambda = 2$) allows for arbitrarily large periods, and hence complex collective behaviors of the network. Moreover, Proposition 5 suggests that such complex behaviors also require some structural complexity in the network, no matter how large the persistence of information in the single neurons is: simple-to-describe network structures such as complete graphs, cycles, and trees cannot exhibit complex activation patterns.

An interesting direction for future research is studying the model where two or more *competing products* are marketed in the same social network. Whoever adopts a product cannot adopt any other, and different companies want to choose their target sets (perhaps one node at a time, in a round-robin fashion) in such a way as to maximize the number of adopters. I conjecture this game-like problem to be PSPACE-hard in general networks.

参考文献

[1] N. Chen. On the Approximability of Influence in Social Networks. *SIAM Journal on Discrete Mathematics*, 23:1400–1415, 2009.

[2] G. Cordasco, L. Gargano, J. Peters, A. Rescigno, and U. Vaccaro. Time-Bounded Influence Diffusion with Incentives. In *Proceedings of SIROCCO'18*, 280–295, 2018.

[3] P. Domingos and M. Richardson. Mining the Network Value of Customers. In *Proceedings of KDD'01*, 57–66, 2001.

[4] L. Gargano, P. Hell, J. Peters, and U. Vaccaro. Influence Diffusion in Social Networks under Time Window Constraints. *Theoretical Computer Science*, 584:53–66, 2015.

[5] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of KDD'03*, 137–146, 2003.