# A Framework for Complex Influence Propagation based on the $F^2DLT$ Class of Diffusion Models

***(Discussion Paper)***

Antonio Caliò and Andrea Tagarelli

Dept. Computer Engineering, Modeling, Electronics, and System Engineering,
University of Calabria, Rende (CS), Italy
{a.calio, tagarelli}@dimes.unical.it

**Abstract.** What are the key-features that enable an information diffusion model to explain the inherent complex dynamics of real-world propagation phenomena? To answer the above question, we discuss a novel class of stochastic Linear Threshold (LT) diffusion models, which are designed to capture the following aspects in influence propagation scenarios: trust/distrust in the user relationships, changes in adopting one or alternative information items, hesitation towards adopting an information item over time, latency in the propagation, time horizon for the unfolding of the diffusion process, and multiple cascades of information that might occur competitively. Around all such aspects, our defined Friend-Foe Dynamic LT ($F^2DLT$) class comprises a non-competitive model as well as two competitive models, which are able to represent semi-progressivity and non-progressivity, respectively, in the propagation process. The above key-constituents embedded in our models make them unique in the literature of diffusion models, including epidemic models. To validate our models through real-world networks, we also discuss different strategies for the selection of the initial influencers to mimic non-competitive and competitive diffusion scenarios, inspired by the widely-known problem of limitation of misinformation spread. Finally, we describe a web-based simulation environment for testing the proposed diffusion models.

## 1 Introduction

Since the early applications in viral marketing, the development of information diffusion models and related optimization methods has provided effective support to address a variety of influence propagation problems. Nowadays, due to the shrinking boundary between real and online social life [2] and the presence of multiple, competitive spreaders over the Web, which could also act towards *misinformation*, deciding whether a source of information is reliable or not has become a critical task. The difficulty in assessing the reliability and trustworthiness of the source generating or sharing a piece of information increases the likelihood of people to be deceived by a spreading information. In addition to this, the tendency to access information from like-minded sources [5] leads users to be trapped in information bubbles, eventually causing forms of network polarization [4]. Polarization can be contrasted via actions devoted to fact-checking or misinformation debunking, but time plays a crucial role in this game, as cognitive phenomena of *confirmation bias* may easily arise.

Understanding the trustworthiness of the source of an information item is often challenging, as it requires tracking the information item of its originator, which could be unfeasible in many cases. Therefore, it beomes essential to capture the effect of *trust/distrust relationships* on both the user behavior and propagation dynamics. One big question hence arises:

> *What are the key-features that make a diffusion model able to explain the inherent dynamic, and often competitive, nature of real-world propagation phenomena?*

**Contributions.** To answer the above question, we discuss a class of stochastic diffusion models, named *Friend-Foe Dynamic Linear Threshold Models* (for short, $F^2DLT$), which has been originally proposed in [3]. Our models are inspired by the classic Linear Threshold (LT) model, whereby an individual can decide to take an action as a result of the exposure to multiple sources of influence. Major key-features of our models is that the information diffusion graph is defined on top of a *trust network*, where trust is encoded into the influence probabilities, the response of a user to the influencing attempts is described by a time-varying activation threshold function, and also a quiescence function is introduced to model the latency or delay in the propagation.

Our $F^2DLT$ models are designed to deal with *non-competitive* and *competitive* propagation scenarios. For competitive campaign scenarios, one model is *semi-progressive*, which assumes that a user, once activated, is only allowed to switch to a different campaign, and another model is *non-progressive*, i.e., it requires a user to have always the support of her/his in-neighbors to keep the activation state with a certain campaign.

To evaluate our models, we also devised four *seed selection* strategies, which mimic different, realistic scenarios of influence propagation. Experimental evaluation conducted on real-world networks, also including comparison with stochastic epidemic models, has provided interesting findings on the meaningfulness and uniqueness of our proposed models.

## 2 The $F^2DLT$ class of diffusion models

### 2.1 Overview

Figure 1 sketches the conceptual architecture of a framework for information diffusion and influence propagation based on our proposed models. Key-constituents

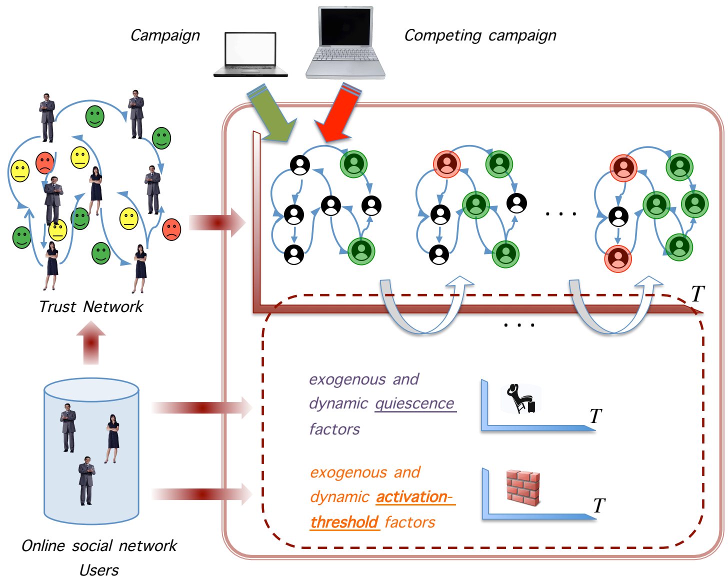


Fig. 1: Illustration of the $F^2DLT$ framework

of this framework are the following: (i) a set of online social network users; (ii) a *trust network*, possibly inferred from a social network; (iii) user-behavior characteristics that provide information for incorporating two main aspects into the diffusion process, namely *activation-threshold* and *quiescence*; (iv) information related to one or multiple competing *campaigns*.

Also, the information diffusion process is supposed to have a certain *time horizon*, and before its expiration, users respond to the network's stimuli which may lead to being active in favor of one campaign or the opposing one. According to their own behavior, users may have the possibility of switching from the adoption of a campaign's item to that of another one.

Putting it all together, our $F^2DLT$ based framework embeds all the above aspects that are essential to explain complex propagation phenomena, i.e., competitive diffusion, non-progressivity, time-aware activation, delayed propagation, and trust/distrust relations.

Notably, in our setting, we tend to reject as true in general, the principle "I agree with my friends' idea and disagree with my foes' idea" (which also resembles the adage "the enemy of my enemy is my friend"), since this would imply that the behavior of a user should be completely determined by the stimuli coming from her/his neighbors. In other terms, friends should be responsible for the influenceability and activation of a user, as opposed to foes, which should instead impact on delayed propagation. Therefore, in our models, the trusted connections and distrusted connections play different roles: only friends can exert a degree of (positive) influence, whereas foes can only contribute to increase the user's hesitation to commit with the propagation process.

One assumption of our framework is the availability of trust relationships between the users involved in the information diffusion context. Although this may not hold in practice, a few studies have been recently developed to infer a trust network from the time-varying interactions observed in evolving social networks, such as [6].

**Activation-threshold function.** Given a directed, weighted network representing an information diffusion graph built on top of a trust network, every node $v$ is associated with an activation-threshold, $\theta_v \in (0, 1]$, which corresponds to the effort of activating the node which is needed in terms of cumulative influence. The *activation-threshold* function $g$, defined over the set of nodes $V$ and the time interval of diffusion $T$, enhances this concept. For each node $v$ at time $t \in T$, it is defined as:

$$g(v, t) = \theta_v + \vartheta(\theta_v, t).$$

Thus, a node $v$ is activated when its perceived influences exceed the threshold $\theta_v$, plus the time-evolving activation term, $\vartheta(\cdot, \cdot)$, which models the dynamic response of users at any activation attempt. Our proposed models can in principle embrace any definition for $\vartheta(\cdot, \cdot)$; here, we focus on two main scenarios.

*Biased*. Modeled as a non-decreasing monotone function, it captures the tendency of a user to consolidate her/his belief, according to the *confirmation-bias* principle [1]. We define this function as follows:

$$g(v, t) = \theta_v + \vartheta(\theta_v, t) = \theta_v + \delta \times \min\left\{\frac{1 - \theta_v}{\delta}, t - t_v^{last}\right\} \tag{1}$$

where $t_v^{last}$ denotes the last time $v$ was activated and $\delta \geq 0$ denotes the increment in the value of $g(v, t)$ for consecutive time steps. Clearly, the value increases as a node keeps staying in the same active state.

*Unbiased*. In applications such as customer retention or churn prediction, a user could revise her/his uncertainty to activate over time. We model this in such a way that, for each $v$, the value of the function is maximum (i.e., 1) just after the activation, i.e., at time $t = t_v^{last} + 1$, then it starts to decrease from the following time steps:

$$g(v, t) = \theta_v + \vartheta(\theta_v, t) = \theta_v + \exp(-\delta(t - t_v^{last} - 1)) - \theta_v \mathbb{I}[t - t_v^{last} = 1] \tag{2}$$

**Quiescence function.** The *quiescence* value quantifies the latency in the propagation through a particular node. It is defined as a non-decreasing and monotone function $q : V, T \mapsto T$, such that for every $v \in V, t \in T$, with $v$ activated at time $t$:

$$q(v, t) = \tau_v + \psi(N_{-}^{in}(v), t),$$

where $\tau_v \in T$ represents an exogenous term modeling the user's hesitation in being committed with the propagation process, and $\psi(N_{-}^{in}(v), t)$ determines an

We assume the second additive term in Eq. (1) is zero if $\delta = 0$.

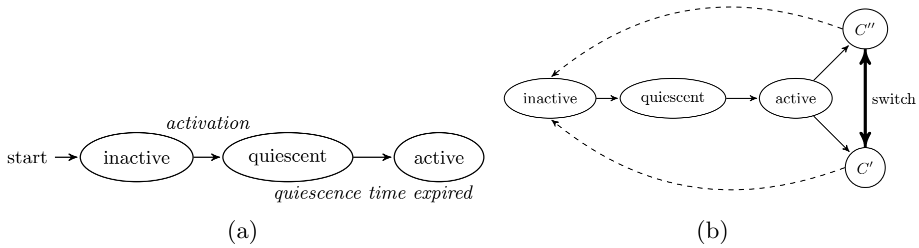


Fig. 2: Life-cycle of a node for the non-competitive model (a) and the competitive models (b). Straight lines in (b) represent the transitions common to both *spC-*$F^2DLT$ and *npC-*$F^2DLT$, while dashed lines refer to *npC-*$F^2DLT$ only.

additional delay proportional to the amount of $v$'s neighbors that are distrusted and active, when the activation attempt is performed by the $v$'s trusted neighbors. Here we focus on the following function:

$$q(v,t) = \tau_v + \psi(N_{-}^{in}(v), t) = \tau_v + \exp\left(\lambda \times \sum_{u \in S_{t-1}} |w_{uv}|\right) \tag{3}$$

where $\lambda \geq 0$ quantifies the user sensitivity towards negative influence.

### 2.2 Diffusion Models

The $F^2DLT$ class includes three diffusion models [3], which are now briefly and informally described. The first one refers to a single-item propagation scenario, while the other two models can dela with competitive scenarios, i.e., two or multiple campaigns spreading informative items over the network in a competitive fashion. Figure 2 shows the life-cycle of a node in the non-competitive scenario and in competitive ones, respectively. It should be noted that all the models share a similar two-phase activation process. In fact, when a node is activated by its neighbors, it always becomes quiescent. Its permanence in this state depends on Eq. 3. After the quiescent time is expired, a node can be regarded as active, regardless of its activation campaign, and is considered fully committed to the propagation process. While the non-competitive model, dubbed *nC-*$F^2DLT$, is a progressive model (i.e., once a node becomes active cannot be deactivated in subsequent times), the semi-progressive competitive model, *spC-*$F^2DLT$, allows a node to switch from a campaign to another, even multiple times. Finally, the fully non-progressive competitive model, *npC-*$F^2DLT$, additionally allows for node deactivation.

**Theoretical properties of the models.** The non-competitive, progressive model *nC-*$F^2DLT$ is proven to be equivalent to LT with Quiescence Time i.e., the activation function in *nC-*$F^2DLT$ is monotone and submodular. On the other hand, the competitive *spC-*$F^2DLT$ and *npC-*$F^2DLT$ can be reduced, via graph serialization, to a progressive Homogeneous Competitive LT, with monotone, non-submodular activation function i.e., the activation function in *spC-*$F^2DLT$ and *npC-*$F^2DLT$ is monotone but not submodular [3].

## 3 Evaluation methodology

**Data.** We used four real-world, publicly available networks, namely: *Epinions*, *Slashdot*, *Wiki-Conflict*, and *Wiki-Vote* — please refer to [3] for details about the datasets. All networks are originally directed and signed; in addition, the two Wikipedia-based networks also have timestamped edges. The influence probability are derived so that to a network having a higher fraction of positive edges will correspond to an equivalent network where users are more willing to be involved in the propagation process.

**Seed selection strategies.** We defined four seed selection strategies, each of which mimics a different, realistic scenario of influence propagation.

*Exogenous and malicious sources of information.* This method, hereinafter referred to as M-Sources, aims at simulating the presence of multiple sources of malicious information within the network. Here, an exogenous source is meant as a node without incoming links. Formally, given a budget $k$, the method selects the top-$k$ users in a ranking solution determined as $r(v) = (\bar{W}^-/(\bar{W}^- + \bar{W}^+)) \log(|N^{out}(v)|)$ for every $v$ such that $N^{in}(v) = \emptyset$. Here, $\bar{W}^+$, resp. $\bar{W}^-$ denote the sum of trust, resp. distrust outgoing weights.

*Exogenous and influential trusted sources of information.* Analogously to the previous method, this one, dubbed I-Sources, is concerned with exogenous sources with emphasis on trusted links. Hence, the ranking function is: $r(v) = (\bar{W}^+/(\bar{W}^- + \bar{W}^+)) \log(|N^{out}(v)|)$.

*Stress triads.* This strategy is based on the notion of *structural balance* in triads. In a typical stress-triad configuration there is a node $v$ with two incoming connections, from a node $z$ with negative weight and from a node $u$ with positive weight, and such that $z$ is connected to $u$ via a positive link. In this case, $z$ is regarded as a *stress-node* since it could activate $v$ through $u$, despite being negatively linked to $v$. The Stress-Triads strategy selects seeds based on the number of stress-triads that are involved in as stress-nodes.

*Newcomers.* We call a node $v \in V$ as a *newcomer* if all of its incoming edges are timestamped as less recent than its oldest outgoing edge. The *start-time* of $v$ is the oldest timestamp associated with its incoming edges. Within the set of newcomers nodes, we further define two separate strategies: Least-New and Most-New. Each one consists in the selection of the the top-$k$ newcomers with highest out-degree among those nodes with the oldest and the newest start-time, respectively. Both strategies require temporal information upon edges.

**Major Findings and Usage Recommendations.** We pursued different goals of evaluations depending on the particular diffusion model. More specifically, we investigated the impact of the negative influence on the final active set in the non-competitive scenario. For the competitive setting, our main goal was to understand the effect of the confirmation-bias effect, in the context of misinformation spreading. Therefore, we will have a good and a malicious actor. Finally, we conducted a comprehensive comparative evaluation of our non-competitive model against the Independent Cascade model as well as stochastic individual-contact epidemic models SIR and SEIR [3].

Here we summarize the major findings emerged from the experiments:

**F1** The setting of the dynamic activation-threshold function and quiescence function plays a crucial role in positive-influence propagation and negative-influence/misinformation limitation.

**F2** The average user's sensitivity in the negative influence perceived from distrusted neighbors ($\lambda$) makes the seed identification more aware of the negative influence spread.

**F3** The confirmation-bias effect ($\delta$) may lead the "stronger" campaign to increase its spread capability.

**F4** The non-progressive competitive model $npC\text{-}F^2DLT$ appears to be less sensitive to the increase of $\delta$ and tends to favor deactivation events (for users previously activated by the weaker campaign) over switched events.

**F5** If compared with classic diffusion models as IC, SIR, and SEIR, the $nC\text{-}F^2DLT$ model tends to favor a slower diffusion, since the propagation process lasts consistently longer than IC and the epidemic models.

**F6** Threshold models are more suitable when it is required to model propagation phenomena where it is important to capture the difference in behavior of each individual. On the contrary, epidemic models does not exhibit this important capability, as they are more suitable to represent single contagions.

**F7** I-Sources reveals higher spread capability, followed by Stress-Triads. On the contrary, the newcomers-based strategies show a significant spread capability, coupled with a negligible effect on the negative influence.

The above results provide evidence that our class of trust-aware models offers an opportunity to better represent the complexity of real-world propagation phenomena. As a consequence, our models can pave the way for the development of sophisticated methods to solve *misinformation spread limitation* and related optimization problems. Also, our models can profitably be used in a variety of applications whereby there is an emergence *to predict the time required to debunk fake information*, or *to estimate how people are affected by the spread of competitive opposite opinions* through a social network.

## 4 Web-based Simulation Environment

As part of our contributions to the research project NextShop PON Grant No. F/050374/01-03/X32, we developed a web application of our diffusion models, available at `http://people.dimes.unical.it/andreatagarelli/f2dlt/` . It is a simulation environment that allows a user to execute a diffusion model in the $F^2DLT$ class, and get an interactive view of a propagation process. As input, an influence graph can be either uploaded from a file or synthetically generated; in the latter case, the user is asked to provide the following information: (i) number of nodes, (ii) number of edges, (iii) percentage of negative edges, and (iv) graph-generator model, by choosing among the Erdős–Rényi, Watts-Strogatz, and Barabási-Albert models. Once the network is created, it is displayed with positive (trust) edges represented in black, and negative (distrust) edges represented in red. Also, after selecting a $F^2DLT$ model, both the activation threshold function and the quiescence function can be configured by setting

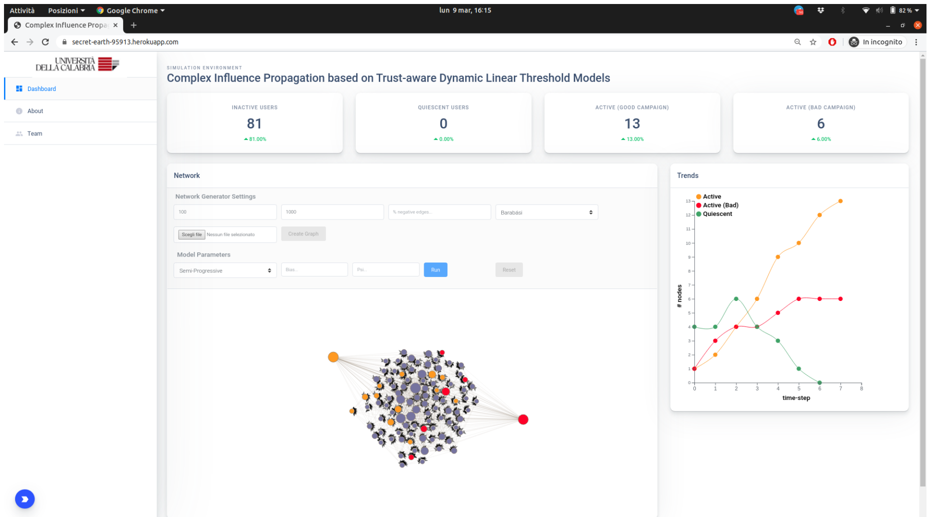


Fig. 3: Screenshot of our web-based simulation environment

$\vartheta(\cdot,\cdot)$,and $\psi(\cdot,\cdot)$. Finally, the early adopters, i.e., the seeds of the propagation process, can be selected by clicking upon specific nodes of the displayed network. While the diffusion process is running, the application updates the layout of the network, and nodes will dynamically change their color accordingly to their activation status. Moreover, the application will show a number of statistics about the propagation process, as well as the temporal evolution of the propagation process.

## References


1. A. Anagnostopoulos, A. Bessi, G. Caldarelli, M. Del Vicario, F. Petroni, A. Scala, F. Zollo, and W. Quattrociocchi. Viral misinformation: The role of homophily and polarization. In *Proc. WebConf*, pp. 355–356, 2015.
2. A. Bessi, G. Caldarelli, M. Del Vicario, A. Scala, and W. Quattrociocchi. Social determinants of content selection in the age of (mis)information. In *Proc. SocInfo*, pp. 259–268, 2014.
3. Antonio Caliò and Andrea Tagarelli. Complex influence propagation based on trust-aware dynamic linear threshold models. *Applied Network Science*, 4(1):14:1–14:41, 2019.
4. K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis. The effect of collective attention on controversial debates on social media. In *Proc. ACM Web Science*, pp. 43–52, 2017.
5. D. Koutra, P. N. Bennett, and E. Horvitz. Events and controversies: Influences of a shocking news event on information seeking. In *Proc. WebConf*, pp. 614–624, 2015.
6. Domenico Mandaglio and Andrea Tagarelli. Generalized preference learning for trust network inference. *IEEE Access*, 7:174583–174596, 2019.