A Pressure-Based Diffusion Model for Influence Maximization on Social Networks

Curt Stutsman[1], Eliot W. Robson[2*], Abhishek K. Umrawal[1]

[1]University of Illinois Urbana-Champaign, Urbana, Illinois 61801, USA
[2]Narmi Inc., New York City, New York 10016, USA
curtiss8@illinois.edu, eliot.robson24@gmail.com, aumrawal@illinois.edu

Abstract

In many real-world scenarios, an individual's local social network carries significant influence over the opinions they form and subsequently propagate. In this paper, we propose a novel diffusion model — **the Pressure Threshold model (PT)** — for dynamically simulating the spread of influence through a social network. This model extends the popular Linear Threshold (LT) model by adjusting a node's outgoing influence in proportion to the influence it receives from its activated neighbors. We examine the Influence Maximization (IM) problem under this framework, which involves selecting seed nodes that yield maximal graph coverage after a diffusion process, and describe how the problem manifests under the PT model. Experiments on real-world networks, supported by enhancements to the open-source network-diffusion library CyNetDiff, reveal that greedy IM under PT can yield seed sets distinct from those under LT. Furthermore, the analyses show that densely connected networks amplify pressure effects far more strongly than sparse networks.

1 Introduction

1.1 Motivation

Diffusion processes within human populations aim to imitate how behaviors, information, diseases, and more propagate through social networks. Diffusion models provide the underlying mathematics that dictate how attributes spread from one node to the next. Early diffusion models provided critical insights into these processes, paving the way for numerous practical applications in epidemiology and the social sciences. With the advent of digital platforms and the increasing significance of viral marketing around the turn of the twenty-first century (Evans and McKee 2010), diffusion models became crucial for solving the Influence Maximization (IM) problem, first formulated by Domingos and Richardson (2001). The goal of the IM problem is to select a set of k seed nodes in a network such that the expected spread of influence is maximized after a diffusion process. The choice of diffusion model significantly impacts how influence propagates, thus determining the effectiveness of seed node selection.

*Part of this work was completed while the author was a Ph.D. student at the University of Illinois Urbana-Champaign.

The influence maximization (IM) problem was popularized by Kempe, Kleinberg, and Tardos (2003), who proposed a greedy approximation algorithm for solving it. Building on threshold-based models of collective behavior (Granovetter 1978) and cascade models from viral marketing (Goldenberg, Libai, and Muller 2001), they formalized the Linear Threshold (LT) and Independent Cascade (IC) models as graph-based diffusion processes that remain widely used in IM research. These models seek to not only mathematically define the propagation of influence but also accurately represent genuine human interactions in real-world communities. However, despite their utility, traditional models often fail to capture key interpersonal phenomena such as reinforcement within echo chambers, where repeated exposure amplifies beliefs and behaviors (Bakshy, Messing, and Adamic 2015). Empirical work shows that online diffusion is shaped by local reinforcement: false or novel items spread farther, faster, and deeper than truth on Twitter (Vosoughi, Roy, and Aral 2018), users self-organize into homophilic clusters that bias transmission toward like-minded peers (Cinelli et al. 2021; Brugnoli et al. 2019), and adoption probabilities increase with multiple confirming exposures, i.e., complex contagion (Mønsted et al. 2017).

To address these shortcomings and better represent realistic social dynamics, this research introduces the Pressure Threshold model (PT), an extension of the LT model. The PT model introduces dynamic feedback mechanisms, wherein nodes amplify their outgoing influence proportional to the incoming influence from activated neighbors. This mechanism directly operationalizes the empirically observed reinforcement pattern: when a node activates under substantial 'pressure,' it becomes a stronger advocate in subsequent steps, thereby capturing within-community amplification consistent with observed diffusion regularities (Vosoughi, Roy, and Aral 2018; Cinelli et al. 2021). Such reinforcement reflects realistic social scenarios. For example, individuals are more likely to strongly advocate a product when many of their friends or peers have already adopted it (Gunawan, Rahmania, and Kenang 2023). By making this feedback explicit while retaining LT's additive semantics, PT offers a simple and interpretable way to model reinforcement beyond static LT/IC assumptions, aligning the diffusion mechanism with documented online behaviors and providing new leverage for IM analyses.

1.2 Literature Review

Diffusion models have a long history of use, extending beyond the IM problem. We now review the history of diffusion models, their role in Influence Maximization, and dynamic approaches related to the Pressure Threshold model.

Early Diffusion Models: The study of diffusion models originated in epidemiology and sociology. Early foundational work by Kermack and McKendrick (1927) introduced the Susceptible-Infected-Recovered (SIR) epidemiological model, setting the stage for understanding contagion processes. In social networks, Granovetter (1978)'s threshold model became foundational, conceptualizing adoption based on cumulative peer influence. Both the Linear Threshold (LT) and Independent Cascade (IC) models were later formalized by Kempe, Kleinberg, and Tardos (2003), and remain standard in IM research. These models provide the basis for most algorithmic work on influence spread.

Influence Maximization: The IM problem was explicitly formulated by Domingos and Richardson (2001) in the context of viral marketing, aiming to maximize influence spread through optimal seed selection. Kempe, Kleinberg, and Tardos (2003) advanced IM research by proving NP-hardness, submodularity of influence, and introducing a greedy approximation algorithm, sparking extensive follow-up work.

Improvements to IM: Subsequent research refined IM algorithms for scalability and efficiency. Leskovec et al. (2007) introduced CELF using lazy forward evaluations, reducing complexity while preserving guarantees, later improved as CELF++ by Goyal, Lu, and Lakshmanan (2011). Borgs et al. (2014) developed Reverse Influence Sampling (RIS) for near-optimal scalability. Chen, Wang, and Yang (2009) proposed the degree discount heuristic for faster, though less optimal, solutions. Recent work explores community-aware heuristics (Umrawal and Aggarwal 2023; Robson and Umrawal 2025) and learning-based approaches (Kumar et al. 2022). Beyond the original problem, variants include partial incentives (Chen, Wu, and Yu 2020; Umrawal, Aggarwal, and Quinn 2023; Bhimaraju et al. 2024), online settings (Agarwal et al. 2022; Nie et al. 2022; Xu and Umrawal 2026), fairness (Becker et al. 2022; Lin et al. 2023; Nguyen et al. 2022), and deep learning methods (Kumar et al. 2022). See (Li et al. 2023) for a detailed survey.

Dynamic Diffusion Models: Unlike the static LT and IC models, dynamic models capture evolving network structure and influence weights. Zhuang et al. (2013) addressed edge dynamics over time, and Xie et al. (2015) proposed DynaDiffuse, combining continuous-time propagation with dynamic edge weights. These approaches share similarities with PT, which introduces adaptive influence updates.

Predictive Comparison of Diffusion Models: Some studies compare diffusion models against real-world cascades. Kuo et al. (2011) evaluated predictive accuracy, finding IC superior for direct prediction tasks, while LT performed better in indirect comparisons. Ananthasubramaniam et al. (2024) analyzed Twitter hashtag propagation, showing that accuracy depends on network topology and content type. Aral and Walker (2012) validated cumulative influence in product adoption, supporting threshold-based assumptions.

Shortcomings of Existing Models: Despite progress, existing models often overlook reinforcement dynamics evident in online networks and misinformation cascades. Bakshy, Messing, and Adamic (2015) showed that ideological reinforcement and echo chambers amplify diffusion on Facebook. Vosoughi, Roy, and Aral (2018) demonstrated that false news spreads faster and deeper than truth, driven by repeated exposure and amplification. Such findings underscore the need for models that incorporate local reinforcement, a gap addressed by the PT model.

In summary, while prior work has advanced IM theory and algorithms, gaps remain in modeling dynamic feedback and empirically observed reinforcement. Our research bridges these gaps by introducing a diffusion model that explicitly accounts for local social reinforcement, supported by empirical evidence.

1.3 Contribution

This paper makes four primary contributions:

1. **Pressure Threshold (PT) Model.** We introduce the Pressure Threshold diffusion model, an extension of the Linear Threshold (LT) model that incorporates local reinforcement: when a node activates, it increases its outgoing influence in proportion to the cumulative incoming pressure that triggered its activation. This mechanism captures empirically observed amplification within communities while preserving the additive semantics of LT.
2. **Theoretical Properties.** We prove that influence maximization under PT is NP-hard (by reduction from LT). We also show by counterexample that PT is neither monotone nor submodular when the amplification parameter $\alpha > 0$, clarifying the scope of classical greedy-approximation guarantees.
3. **Implementation.** We enhance the CyNetDiff (Robson, Reddy, and Umrawal 2024) Python library for accelerated simulations, leveraging Cython-based kernels to support large-scale Monte Carlo evaluation across diffusion models, including PT. These improvements enable practical experiments on real-world networks at scale.
4. **Empirical Analysis.** Across real and synthetic networks, PT (i) can produces seed sets distinct from LT and (ii) often achieves larger final spreads for comparable budgets. Although submodularity does not hold in general, we observe approximate submodular behavior in large networks for small α, under which greedy strategies remain effective. Analyses on synthetic topologies highlight regimes where PT's reinforcement is especially beneficial (e.g., denser structures with more opportunities for amplification).

Collectively, these results indicate that explicitly modeling local reinforcement yields diffusion dynamics and seed-selection behavior that are not captured by static LT/IC formulations, while remaining computationally tractable for influence maximization workflows.

1.4 Organization

The remainder of the paper is organized as follows. Section 2 introduces preliminaries and formally defines the Pressure Threshold model, including the NP-hardness argument. Section 3 provides a small-scale walk-through of PT dynamics and presents our analysis of non-monotonicity and non-submodularity using counterexamples. Section 4 describes our experimental setup, implementation details (CyNetDiff), datasets, and results on both real-world and synthetic networks. Section 5 concludes with a summary of contributions, real-world implications, and avenues for future work.

2 Preliminaries and Model Formulation

In this section, we formally establish foundational concepts and clearly define the Pressure Threshold model (PT). We discuss existing diffusion models, formalize the proposed PT model, and analyze its computational complexity.

2.1 Linear Threshold Model and Influence Maximization

Diffusion models mathematically describe how information, ideas, behaviors, or diseases propagate through a network. One of the most widely studied models in the context of influence maximization is the Linear Threshold (LT) model, which models adoption via cumulative social influence and classic threshold-based theories (Granovetter 1978).

Given that the PT model is a generalization of the LT model, we will now give a formal definition of the LT model. Let $G = (V, E)$ be a directed graph where V is the set of nodes and $E \subseteq V \times V$ is the set of edges. Each edge $(u, v) \in E$ is assigned a weight $w_{uv} \in [0, 1]$ such that for each node v, $\sum_{u \in N(v)} w_{uv} \leq 1$, where $N(v)$ denotes the in-neighbors of v. The edge weight is meant to quantify the degree to which node u can influence node v. Additionally, each node v has a threshold θ_v drawn uniformly at random from the interval $(0, 1]$. Initially, a set of seed nodes $S \subseteq V$ is activated. In each subsequent discrete time step, an inactive node v becomes active if the total weight of its active in-neighbors meets or exceeds its threshold:

$$\sum_{u \in P(v)} w_{uv} \geq \theta_v,$$

where $P(v) = \{u \in A_t : (u, v) \in E\}$ denotes the set of active parent nodes (or in-neighbors) of v. This process continues until no new nodes can be activated, and the diffusion ceases. For clarity, we consistently use $N(v)$ for the (time-invariant) set of in-neighbors of v, and $P(v)$ for the (time-dependent) subset of $N(v)$ that are active at time t.

We define the influence of a seed set $S \subseteq V$ under the Linear Threshold (LT) model as the expected number of nodes activated by the diffusion process initiated from S, where the expectation is taken over the random thresholds $\theta_v \in (0, 1]$ for all $v \in V$. More generally, for a diffusion model M, we denote the influence of S by $\sigma_M(S)$. The influence maximization (IM) problem is then to select, given a directed graph $G = (V, E)$, a diffusion model M, and an integer budget k, a seed set $S \subseteq V$ with $|S| \leq k$ that maximizes $\sigma_M(S)$ (Kempe, Kleinberg, and Tardos 2003). Formally, $\arg\max_{S \subseteq V} \sigma_M(S),$ such that $|S| \leq k.$

2.2 Pressure Threshold Model Definition

The PT model is defined on the same graph structure as the LT model and shares the same node activation conditions. Specifically, let $G = (V, E)$ be a directed graph, where V denotes the set of nodes and $E \subseteq V \times V$ denotes the set of directed edges. Each edge $(u, v) \in E$ is assigned a weight $w_{uv} \in [0, 1]$ such that for each node v, $\sum_{u \in N(v)} w_{uv} \leq 1$, where $N(v)$ denotes the in-neighbors of v. Unless otherwise stated, each weight w_{uv} is initialized with a value $1/\text{in-degree}(v)$. Furthermore, each node $v \in V$ is assigned a threshold $\theta_v \in (0, 1]$, which specifies the minimum cumulative influence required from its incoming neighbors for the node to become active.

The diffusion process unfolds in discrete time steps and begins with an initial seed set $S \subseteq V$, where the nodes in S are assumed to be active at time step $t = 0$. Let $A_t \subseteq V$ denote the set of active nodes at time t, with the initial condition $A_0 = S$. At each subsequent time step, the activation of new nodes is governed by the same rule as the classical Linear Threshold (LT) model: a node $v \in V \setminus A_t$ becomes active at time $t + 1$ if the cumulative influence from its currently active in-neighbors meets or exceeds its threshold θ_v. Formally, the activation condition is given by:

$$\sum_{u \in P(v)} w_{uv} \geq \theta_v,$$

where $P(v) = \{u \in A_t : (u, v) \in E\}$ denotes the set of active parent nodes (or in-neighbors) of v.

While the activation rule mirrors that of the LT model, the PT model introduces a key modification through a two-phase update mechanism. Specifically, each round of the diffusion process is composed of the following phases:

1. **Activation Phase**: All nodes that satisfy the activation condition based on the current edge weights and active neighbor states become active.
2. **Influence Adjustment Phase**: The influence exerted by newly activated nodes on their neighbors is recalibrated based on the influence they received during activation.

The rationale for the influence adjustment phase is to model a kind of ancestral amplification: a node that was heavily influenced by its neighbors becomes a more influential source in turn. A node's outbound influence is amplified only once during the timestep of its activation, which prevents recursive feedback loops that could otherwise occur in denser regions of the network. Consequently, the adjustment phase updates only edges toward inactive neighbors, since active nodes remain permanently activated, reducing redundant computation and improving efficiency. To formalize this, let $v \in V$ be a node that becomes active at time step t. Let I_v denote the total incoming influence received by v at the time of its activation:

$$I_v = \sum_{u \in P(v)} w_{uv}.$$

The PT model introduces a tunable amplification parameter $\alpha \geq 0$, which controls the extent to which this received influence translates into increased outgoing influence. For each newly activated node v and for each of its outgoing neighbors s such that $(v, s) \in E$ and $s \notin A_t$, the influence weight is updated according to:

$$w'_{vs} = \min(1, w_{vs} + \alpha \cdot I_v),$$

where w'_{vs} is the new, adjusted influence weight from v to s. The use of the minimum ensures that the updated weights respect the upper bound of 1, consistent with the range of node thresholds. Because the influence added to each edge is directly proportional to αI_v and bounded by the $\min(1, \cdot)$ constraint, the amplification remains small and localized. The additive form of the amplification scheme reflects the additive nature of the threshold activation function in both the LT and PT models. This design ensures that the model remains consistent with the underlying interpretation of influence aggregation. Like other standard diffusion models, this process continues until no new nodes can be activated and is progressive, meaning nodes may go from inactive to active, but never the other way. Algorithm 1 summarizes the complete PT diffusion procedure.

Conceptually, α serves as a reinforcement coefficient that encodes how strongly social feedback amplifies an individual's influence on their peers. Small values of α correspond to diffusion processes that behave similarly to the Linear Threshold model, while larger values produce stronger social reinforcement dynamics and faster saturation. Because α modulates the intensity of behavioral feedback rather than the network structure itself, in practice it can be treated as a tunable hyperparameter that reflects the degree of social reinforcement present in a given domain and could be estimated or calibrated from empirical diffusion traces, similar to contagion strength in epidemiological models or persuasion intensity in social influence studies (Czuba et al. 2024; Kalimeris et al. 2018).

Remark. As in the LT model, the influence of a seed set $S \subseteq V$ under the PT model is defined as the expected number of activated nodes by the diffusion process initiated from S, where the expectation is taken over the random thresholds.

Algorithm 1: Pressure Threshold (PT) Diffusion

Require: Directed graph $G = (V, E)$; weights $w_{uv} \in [0, 1]$; thresholds $\theta_v \in (0, 1]$; seed set S; amplification $\alpha \geq 0$

$t \leftarrow 0;\ A_t \leftarrow S$ ▷ initial active set
repeat
  $N_t \leftarrow \emptyset$ ▷ newly activated at time t
  for all $v \in V \setminus A_t$ **do** ▷ activation phase
    $P_t(v) \leftarrow \{\, u \in A_t : (u, v) \in E \,\}$
    $I_v \leftarrow \sum_{u \in P_t(v)} w_{uv}$
    if $I_v \geq \theta_v$ **then**
      $N_t \leftarrow N_t \cup \{v\}$
    end if
  end for
  if $N_t = \emptyset$ **then**
    break ▷ no new activations; diffusion halts
  end if
  for all $v \in N_t$ **do** ▷ influence adjustment phase (one-shot)
    for all $(v, s) \in E$ **with** $s \notin A_t$ **do** ▷ only toward currently inactive neighbors
      $w_{vs} \leftarrow \min\big(1,\ w_{vs} + \alpha \cdot I_v\big)$
    end for
  end for
  $A_{t+1} \leftarrow A_t \cup N_t;\ t \leftarrow t + 1$
until false
return A_t

2.3 Computational Complexity

Theorem 1. *Influence maximization under the Pressure Threshold (PT) diffusion model (PT-IM) is NP-hard.*

Proof. The classical influence maximization problem under the Linear Threshold (LT) model (LT-IM) is NP-hard (Kempe, Kleinberg, and Tardos 2003). We reduce LT-IM to PT-IM in polynomial time. Given an LT instance (G, k), construct the PT instance $(G, k, \alpha = 0)$. When $\alpha = 0$, the PT process is identical to LT because the adjustment phase is inactive. Formally, for every seed set S we have:

$$\sigma_{\mathrm{PT}}(S; \alpha = 0) = \sigma_{\mathrm{LT}}(S).$$

Thus, an optimal solution for PT-IM with $\alpha = 0$ is an optimal solution for the original LT-IM instance. Thus, any polynomial-time algorithm for PT-IM would yield one for LT-IM, giving a polynomial-time many-one reduction. Therefore, PT-IM is NP-hard. □

Remark (Decision vs. optimization). The theorem above is stated for the optimization problem (maximizing expected spread). For the decision version ("Does there exist $S \subseteq V$, $|S| \leq k$, with expected spread at least T?"), the same restriction $\alpha = 0$ yields an identical polynomial-time reduction, implying NP-hardness of the decision problem as well.

Corollary 1 (Inapproximability inherited at $\alpha = 0$)**.** *Under $\alpha = 0$, PT reduces to LT. It is NP-hard to approximate optimal influence for LT within a factor better than $1 - 1/e + \varepsilon$ for any $\varepsilon > 0$ unless* P = NP *(Kempe, Kleinberg, and Tardos 2003). Consequently, PT-IM inherits the same inapproximability barrier.*

This computational complexity result highlights the necessity of efficient approximation or heuristic approaches for practical applications.

3 Properties of the Proposed Model

3.1 Model Demonstration

To build intuition for the reinforcement dynamics introduced by PT, we present a compact visualization contrasting PT and LT on a small controlled instance. The objective is not to claim generality from a toy example, but to provide a clear, step-by-step illustration of how additive amplification upon activation changes propagation relative to the classical, static-weight LT evolution on the same graph.

We illustrate the PT dynamics using a small-scale example with $\alpha = 0.25$ on a 9-node network. Figures 1 and 2 show the early and later stages of PT diffusion alongside LT on the same graph and seed set. A larger value of α is used to better highlight the distinction between the two models.

Both models begin with two identical seeds at $t = 0$. Activated nodes are shown in red, newly activated nodes are outlined, and edges whose weights are amplified during the adjustment phase are highlighted. In the earliest timesteps, PT and LT behave similarly since the initial weights are identical and no adjustment has yet occurred. Differences emerge once reinforcement activates: newly activated PT nodes increase their outgoing influence, allowing neighbors that remain inactive under LT to cross their thresholds. As these effects accumulate, PT achieves full coverage, whereas LT terminates with only five activated nodes.

3.2 Monotonicity Analysis

While the LT model is known to yield a non-decreasing influence spread function (Kempe, Kleinberg, and Tardos 2003), no analogous guarantee holds under the PT model.

A set function $f : 2^V \to \mathbb{R}$ is *monotone* (non-decreasing) if for all $S \subseteq T \subseteq V$, $f(S) \leq f(T)$.

Common illustrative graph. The following two propositions and remarks refer to the same directed network $G = (V, E)$, where $V = \{a, b, c, d, e\}$ and $E = \{(b, d), (b, c), (c, a), (a, b), (e, b)\}$, with corresponding edge weights $w_{bd} = w_{bc} = w_{ca} = 1$ and $w_{ab} = w_{eb} = 0.5$.

Proposition 1. *There exists a graph G, edge weights, and a parameter $\alpha > 0$ such that the influence $\sigma_{\mathrm{PT}}(\cdot)$ under the PT model is not monotone.*

Proof. Let $S = \{c\}$ and $T = \{a, c\}$, so that $S \subseteq T$. Set $\alpha = 0.5$. By estimating $\sigma_{\mathrm{PT}}(\cdot)$ via averaging cascade sizes over $1,000$ random threshold realizations, we obtain $\sigma_{\mathrm{PT}}(S) = 4 > 3.019 = \sigma_{\mathrm{PT}}(T)$, violating monotonicity. □

Remark (Conceptual clarification). The non-monotonicity in Proposition 1 arises from comparing seed sets, not from a single cascade. For any fixed seed set S, the active set grows monotonically over time, nodes never deactivate, and the cascade process itself remains progressive.

Remark (Pointwise illustration). For the same graph (with the same $\alpha = 0.5$), non-monotonicity can be seen for a fixed threshold realization. Let the thresholds be $\theta_a = 0.6$, $\theta_b = 0.7$, $\theta_c = 0.4$, $\theta_d = 0.6$, and $\theta_e = 0.2$. For $S = \{c\}$, node c is seeded and hence active initially. Node a receives influence $w_{ca} = 1 > \theta_a$ and activates. Since a was activated through propagation and not seeded, PT amplification applies to its outgoing edges, so the weight on (a, b) is amplified to $w'_{ab} = \min\{1, w_{ab} + \alpha w_{ca}\} = \min\{1, 0.5 + 0.5 \cdot 1\} = 1$. Thus, node b receives influence $w'_{ab} = 1 > \theta_b$ and activates, after which node b activates d via $w_{bd} = 1 > \theta_d$. The final active set is $\{a, b, c, d\}$, of size 4. For $T = \{a, c\}$, nodes a and c are both seeded and hence active initially. Since PT amplification applies only to nodes activated through propagation, amplification does not apply to outgoing edges of a, and node b receives only the unamplified influence $w_{ab} = 0.5 < \theta_b$, remaining inactive. No further activations occur, and the final active set is $\{a, c\}$, of size 2. Since $4 > 2$, this threshold realization shows pointwise non-monotonicity.

3.3 Submodularity Analysis

While the LT model is known to produce a submodular influence function (Kempe, Kleinberg, and Tardos 2003), the activation-triggered amplification mechanism of the PT model can violate the diminishing returns property.

A set function $f : 2^V \to \mathbb{R}$ is *submodular* if for all $S \subseteq T \subseteq V$ and $v \notin T$, $f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$.

Proposition 2. *There exists a graph G, edge weights, and a parameter $\alpha > 0$ such that the influence $\sigma_{\mathrm{PT}}(\cdot)$ under the PT model is not submodular.*

Proof. Using the same graph and amplification parameter $\alpha = 0.5$, let $S = \{c\}$, $T = \{a, c\}$, and $v = b$. By estimating $\sigma_{\mathrm{PT}}(\cdot)$ via averaging cascade sizes over 1000 independent random threshold realizations, we obtain $\sigma_{\mathrm{PT}}(S \cup \{b\}) = 4$, $\sigma_{\mathrm{PT}}(S) = 4$, $\sigma_{\mathrm{PT}}(T \cup \{b\}) = 4$, and $\sigma_{\mathrm{PT}}(T) = 3.019$. Thus, $\sigma_{\mathrm{PT}}(S \cup \{b\}) - \sigma_{\mathrm{PT}}(S) = 0 < 0.981 = \sigma_{\mathrm{PT}}(T \cup \{b\}) - \sigma_{\mathrm{PT}}(T)$, violating submodularity. □

Remark (Approximate submodularity). Due to pressure effects, the PT influence function σ_{PT} is generally neither monotone nor submodular, and therefore the classical $(1 - 1/e)$ greedy guarantee (Nemhauser, Wolsey, and Fisher 1978) does not apply. For non-monotone submodular maximization under cardinality constraints, the optimal polynomial-time approximation factor is $1/e$ (Feige, Mirrokni, and Vondrák 2011). Thus, without monotonicity, one cannot, in general, expect a $(1 - 1/e)$ guarantee. Nevertheless, it is still informative to quantify how far σ_{PT} departs from submodularity via the submodularity ratio $\gamma_{U,k}(f)$ (Das and Kempe 2018), defined (for any set function f) as

$$\gamma_{U,k}(f) = \min_{\substack{L \subseteq U \\ S:\, |S| \leq k,\ S \cap L = \varnothing}} \frac{\sum_{x \in S} \big(f(L \cup \{x\}) - f(L)\big)}{f(L \cup S) - f(L)}.$$

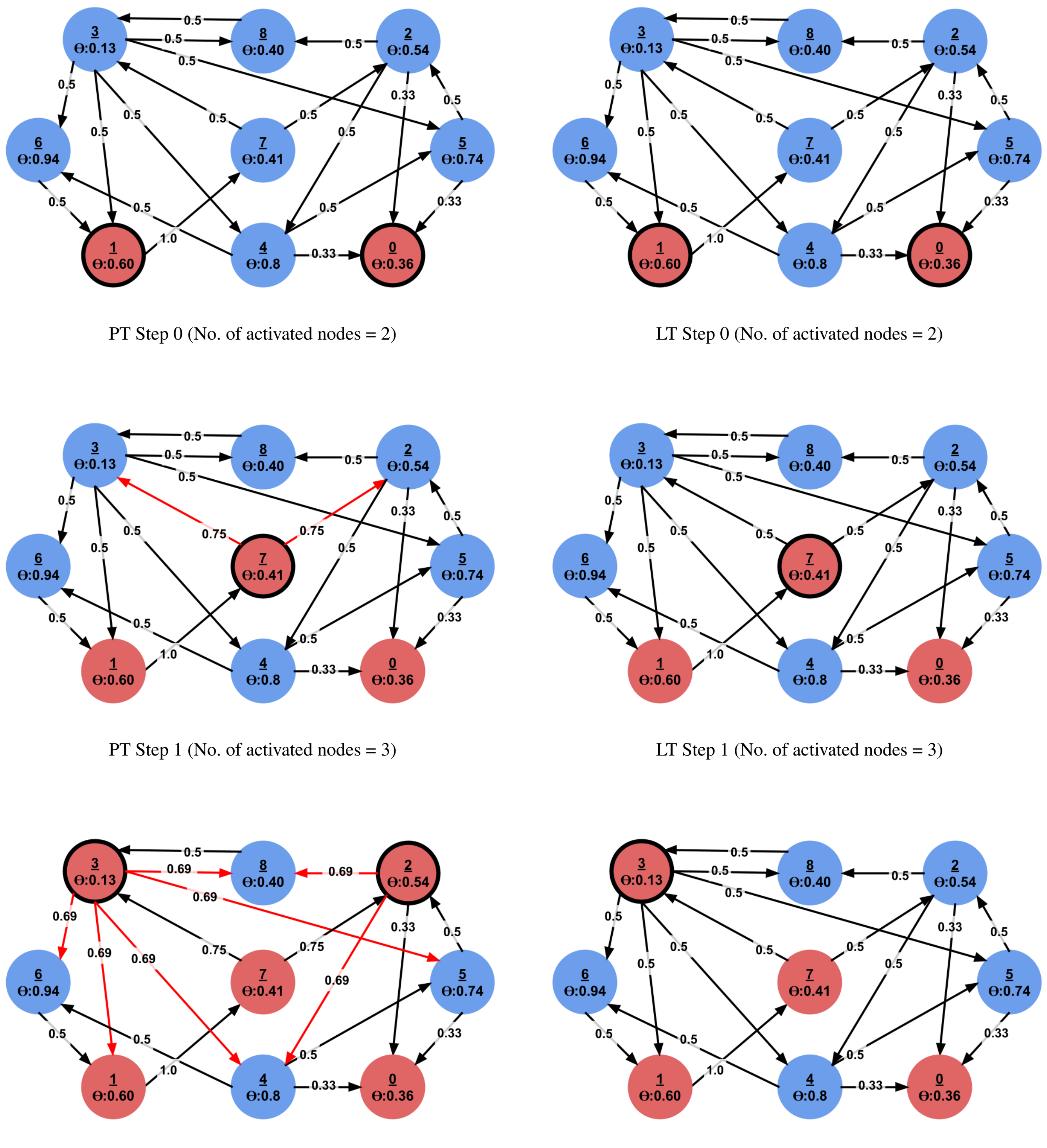
When $\gamma_{U,k} = 1$, the function is submodular.

In the PT model, the submodularity ratio satisfies

$$\gamma_{U,k}(\sigma_{\mathrm{PT}}) \geq \underline{\gamma}(G, \alpha), \qquad \text{with} \qquad \underline{\gamma}(G, \alpha) \xrightarrow[\alpha \to 0]{} 1,$$

which recovers the classical LT setting and its $(1 - 1/e)$ approximation guarantee in the zero-pressure limit. A precise, data-dependent characterization of $\underline{\gamma}(G, \alpha)$ as a function of the graph and pressure is left for future work.

Remark (Pointwise illustration). Using the same graph (with the same $\alpha = 0.5$), and threshold realization as in the pointwise non-monotonicity remark, consider the marginal effect of adding b. For $S = \{c\}$, the cascade proceeds exactly as in that remark: c activates a, amplification applies to (a, b), and the resulting cascade activates b and then d, yielding a final active set of size 4. Adding b to S produces no additional activations, since b is now seeded and amplification does not apply to (b, d). For $T = \{a, c\}$, node b does not activate in the base cascade, as $w_{ab} = 0.5 < \theta_b$. Adding b activates d directly via $w_{bd} = 1 > \theta_d$, strictly increasing the cascade size. Thus, the marginal gain of adding b is larger for T than for S, violating submodularity pointwise for this realization.



PT Step 0 (No. of activated nodes = 2)

LT Step 0 (No. of activated nodes = 2)

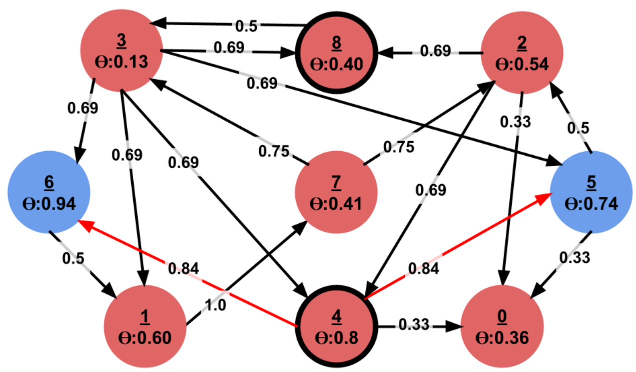
PT Step 1 (No. of activated nodes = 3)

LT Step 1 (No. of activated nodes = 3)

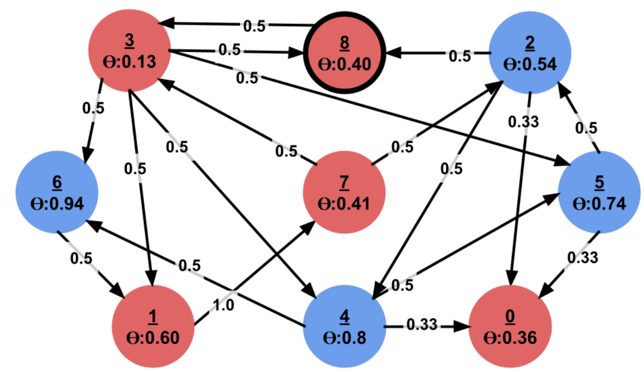
PT Step 2 (No. of activated nodes = 5)

LT Step 2 (No. of activated nodes = 4)

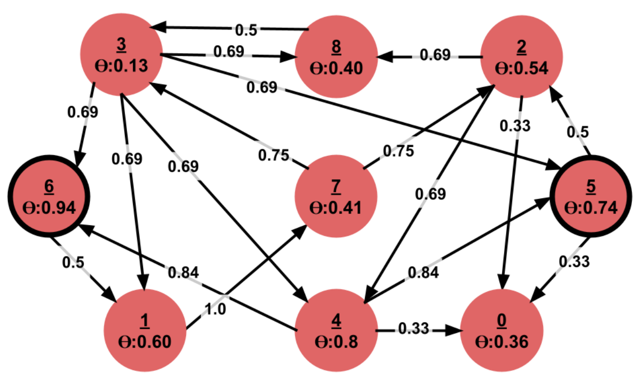
Figure 1: Early diffusion stages under PT and LT on the same graph using an identical seed set. Visualizations show the initial evolution of the diffusion process, during which the reinforcement effects are present but have not yet accumulated strongly. Activated nodes are shown in red, newly activated nodes are highlighted in red with a black outline, and unactivated nodes are shown in blue. Newly amplified edges resulting from PT's influence adjustment mechanism are highlighted in red. Node labels display the node index (underlined) together with the corresponding activation threshold θ.



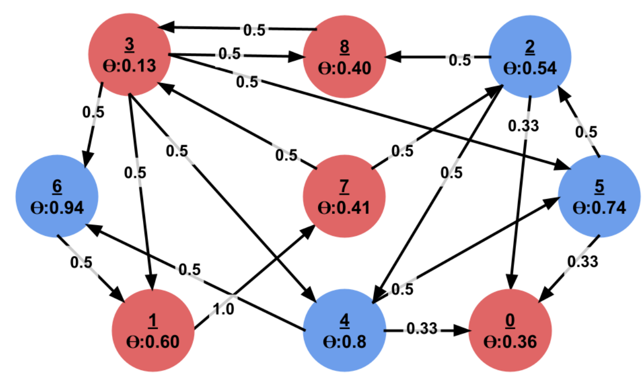
PT Step 3 (No. of activated nodes = 7)



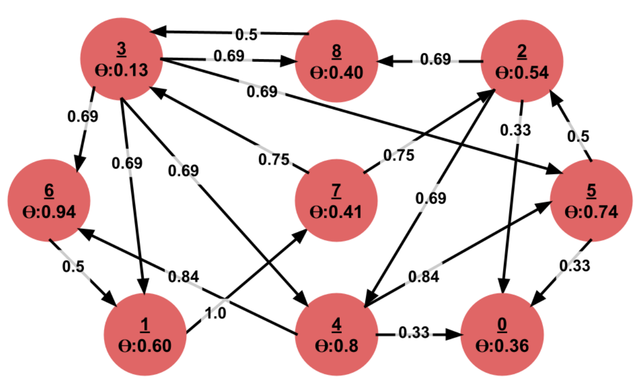
LT Step 3 (No. of activated nodes = 5)



PT Step 4 (No. of activated nodes = 9)



LT Final Step (No. of activated nodes = 5)



PT Final Step (No. of activated nodes = 9)

Figure 2: Later diffusion stages under PT and LT on the same graph using an identical seed set. As the diffusion evolves, the reinforcement effects intensify, PT continues amplifying outgoing influence along activated edges and attains full coverage, whereas LT halts after activating fewer nodes. Activated nodes are shown in red, newly activated nodes are highlighted in red with a black outline, and unactivated nodes are shown in blue. Newly amplified edges due to PT's reinforcement mechanism are highlighted in red. Node labels display the node index (underlined) together with the corresponding activation threshold θ.

4 Experiments

4.1 Network Data

We evaluate the Pressure Threshold (PT) model against the Linear Threshold (LT) model on four networks spanning distinct topologies and semantics. Three of these networks are real-world graphs obtained from the Stanford Large Network Dataset Collection (SNAP) (Leskovec and Krevl 2014): (i) Facebook social circles (Leskovec and Mcauley 2012), (ii) Wikipedia voting (Leskovec, Huttenlocher, and Kleinberg 2010a,b), and (iii) Bitcoin OTC trust (Kumar et al. 2016, 2018). To complement these with a purely synthetic baseline, we also include (iv) an Erdős–Rényi random network $G(n, p)$ (Erdős and Rényi 1959). The number of nodes, number of directed edges, and edge-to-node ratios are reported in Table 1. Throughout, we study progressive cascades (nodes, once activated, remain active) and keep the underlying graph structure fixed during each diffusion run.

Network	Nodes	Edges	Edge/Node Ratio
Facebook	4,039	88,234	21.846
Bitcoin	5,881	35,592	6.052
Wikipedia	7,115	103,689	14.573
Erdős–Rényi	5,000	62,597	12.519

Table 1: Basic statistics of the networks used in experiments.

Graph construction and directionality. The Facebook circles dataset is originally undirected, with an edge indicating a mutual friendship relation. Following standard practice in LT-based diffusion studies (Kempe, Kleinberg, and Tardos 2003), we convert each undirected edge $\{u, v\}$ into two directed edges (u, v) and (v, u) to preserve symmetric influence opportunities. The Bitcoin OTC trust network is inherently directed: an edge (u, v) represents a rating or trust assignment from user u to user v on a peer-to-peer trading platform. The Wikipedia voting network is also directed: an edge (u, v) indicates that user u voted for user v in an admin election, producing a directed endorsement flow aligned with influence propagation. Similar to the Facebook network, for the Erdős–Rényi random graph, we generate $G(n, p)$ with $n = 5{,}000$ and $p = 0.005$, and convert each undirected edge into two directed arcs.

Edge weights (weighted-cascade normalization). Unless stated otherwise, we use the weighted-cascade convention (Kempe, Kleinberg, and Tardos 2003), assigning for each directed edge (u, v) the weight:

$$w_{uv} \;=\; \frac{1}{\text{in-degree}(v)}.$$

This ensures that for every node v, the incoming weights sum to one, i.e., $\sum_{u \in N(v)} w_{uv} = 1$. In the PT model, these weights may be updated according to the additive adjustment rule upon activation events (Section 2), while respecting the upper bound $w'_{uv} \leq 1$ via the $\min(1, \cdot)$ cap.

Thresholds and randomness. In all runs, node thresholds θ_v are drawn i.i.d. uniformly from $(0, 1]$ at the beginning of each Monte Carlo (MC) simulation, consistent with the LT literature. For paired comparisons (PT vs. LT) under the same seed budget k, we reuse the same threshold instantiation to isolate model effects from sampling noise. All reported spreads are averages over independent MC trials, with the number of trials specified per experiment.

4.2 Experiment Details

We perform three complementary experiments to probe (i) the qualitative effect of PT's reinforcement on seed selection (Experiment 1), (ii) the quantitative impact on influence spread versus budget across heterogeneous graphs (Experiment 2), and (iii) the sensitivity of PT to the reinforcement parameter α (Experiment 3). We use the CELF algorithm (Leskovec et al. 2007) for seed selection in Experiments 1 and 2 to maintain consistency with standard IM practice.

Experiment 1: Seed selection under PT vs. LT. This experiment examines whether PT and LT identify the same influential nodes for the same budget k. We run CELF for the Facebook network with budget $k = 20$. To evaluate the marginal gains within CELF, we estimate spreads via 1,000 MC simulations per evaluation. For the PT runs, we set $\alpha = 0.1$, which empirically balances (i) reinforcement leading to altered seed rankings and (ii) avoidance of near-instant saturation. For consistency, the same set of random threshold instantiations is shared between LT and PT, ensuring that observed differences are attributable to the model rather than randomness. The order of seed node selection is preserved for side-by-side comparison.

Experiment 2: Influence vs. budget. To quantify the downstream effect of PT's reinforcement, we compute the influence spread as a function of the seed budget k across the four graphs in Table 1. For each network and each budget $k \in \{1, 2, \ldots, 60\}$, we run CELF under three diffusion settings: LT, PT with $\alpha = 0.001$, and PT with $\alpha = 0.005$. Each evaluation of the influence of a set within CELF averages 1,000 MC simulations. These α values were selected empirically to illustrate qualitative differences between PT and LT while ensuring stable and interpretable diffusion dynamics. Pilot tests with larger α rapidly saturated the denser graphs (Facebook, Wikipedia), while substantially smaller α yielded behavior nearly identical to LT.

Remark (Alternative activation threshold priors). We also tested $\text{Normal}(0.5, 0.1)$, $\text{Beta}(2, 5)$, and $\text{Beta}(5, 2)$ activation threshold priors; all yielded PT–LT spread gaps similar to the uniform prior, so redundant plots are omitted.

Experiment 3: Sensitivity of PT to α. We probe how PT's final spread scales with the reinforcement parameter by fixing a seed set size ($k = 10$) on Facebook and sweeping α over a fine grid $[10^{-4}, 10^{-1}]$ with step 10^{-4}. For each α, we perform 1,000 MC simulations and report the average spread. To reduce high-frequency variability due to stochastic thresholds, we apply a centered moving average with a window of 9 grid points; each plotted value is the average of the focal α and its four nearest neighbors.

4.3 Results

The results for Experiment 1 appear in Table 2 and Figure 3. The results for Experiment 2 are shown in Figures 4–7. The results for Experiment 3 are given in Figure 8.

Model	Seeds selected (in order)
Linear Threshold	107, 351, 1821, 352, 0, 348, 1831, 349, 366, 1490, 1373, 1827, 2154, 1285, 838, 2806, 1149, 2130, 2403, 2088
Pressure Threshold	107, 351, 352, 1821, 0, 348, 1490, 1215, 1831, 1285, 1426, 1373, 2806, 366, 2403, 3234, 349, 2145, 1827, 2832

Table 2: Ordered seed sets returned by CELF for the Facebook network $k = 20$. The overlap in early positions and divergence in later positions reveal how PT's reinforcement can elevate otherwise secondary vertices.

4.4 Discussion

Figure 3 compares the seed sets that CELF selected under the LT and PT models for a budget of $k = 20$ for the Facebook network. The first column lists the LT seeds, whereas the second lists the PT seeds. Lines connect the vertices common to both seed sets. This graph shows that solving the IM problem under PT yields different seed sets than LT, highlighting a key distinction between the models. The models agree on most early selections (those with the largest marginal gains), but diverge more and more in ordering and in selection as the iteration proceeds. From step 8 onward, PT introduces vertices (e.g., 1215) that never appear in the LT list, indicating that amplification can elevate otherwise secondary nodes. Conversely, several LT-specific vertices never surface under PT. Because PT applies an additive effect to edge weights rather than a diminishing one, these omissions imply that other vertices experience larger marginal-gain increases and overtake them in priority. We also observe that the PT model's seed selection tends to favor nodes embedded within denser regions of the network, where local clustering and mutual reinforcement create more opportunities for amplification compared to nodes with large but sparsely connected neighborhoods. This behavior arises because the PT process implicitly places greater value on a node's in-degree, reflecting the cumulative pressure it can receive from active neighbors rather than solely on out-degree, as in traditional LT and IC models. Table 2 presents the ordered lists in compact form.

Figures 4–7 show the influence of diffusions across the four networks as the budget varies. As expected, network diffusions with the PT model result in a larger influence spread than the LT model. What is interesting about this experiment is the margin of variation under the PT model between the different network types. For example, the Facebook and Wikipedia networks result in much larger influence spreads for the PT model than the LT model when compared to the Bitcoin network, whose influence under the PT model is not much larger than under the LT model. This is most likely due to the larger edge-to-node ratios of the former, as given in Table 1. Intuitively, this makes sense, as more edges yield more opportunity for the influence amplification to take place (phase 2 in the PT diffusion). The amplification was so large that some networks were fully diffused before reaching the max budget of $k = 60$, such as the Erdős–Rényi network reaching full coverage at $k = 22$. Despite these scale differences, the influence curves for LT and PT maintain similar shapes, suggesting that PT adds an approximately constant offset to the baseline diffusion rate. The logarithmic shape of the PT curves, which matches the diminishing returns of the LT curve, also suggests practical submodularity of the influence spread function on the datasets tested.

Figure 8 shows the influence spread of a PT diffusion for the Facebook network as a function of the parameter α. Because of the large amount of data points, we applied a centered moving average with a window size of 9 to smooth the α-vs-influence curve. In the plot, each point is replaced by the average of the point itself and the four points to its right and left. As expected, increasing the α value increases the influence amplification at each node, which increases the total influence spread.

5 Conclusion

5.1 Summary of Contributions

In this work, we introduced the Pressure Threshold (PT) model, a diffusion framework that captures reinforcement effects and cumulative influence in social networks. PT extends the Linear Threshold (LT) model by amplifying a node's outgoing influence according to the pressure it receives from neighbors. We showed that influence maximization under PT remains NP-hard and that PT is neither monotone nor submodular when $\alpha > 0$, implying that classical greedy guarantees may not fully apply. Experiments on real and synthetic networks demonstrated that PT yields seed sets and spreads distinct from LT, with stronger gains in denser networks and more modest effects in sparse ones. These findings highlight PT's ability to model reinforcement-driven diffusion not captured by LT.

5.2 Real-World Implications

The Pressure Threshold model has notable implications for applications such as viral marketing, political campaigning, and information dissemination. By capturing reinforcement phenomena like echo chambers and social amplification, PT can improve predictions of influence spread and help organizations select influential seed nodes more effectively. Consequently, it can inform policymaking for combating misinformation and optimizing public awareness campaigns, offering decision-makers a clearer understanding of how local network pressures shape overall influence dynamics.

5.3 Future Work

Future work includes PT-specific approximation analyses. For non-monotone PT objectives, the $1/e$ approximation factor provides a natural baseline. An important direction is

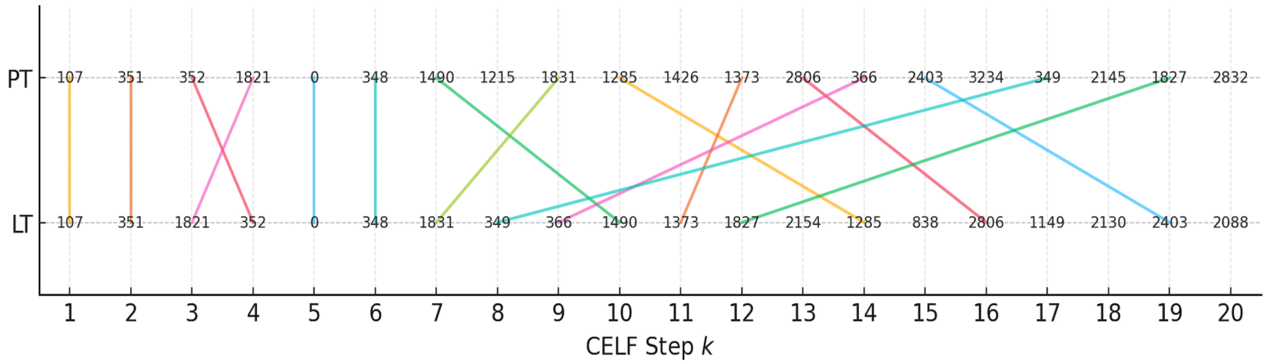


Figure 3: CELF seed-selection timeline for the Facebook network (PT vs. LT, $k = 20$). Horizontal position shows the selection step; vertical links mark common seeds. Strong early overlap is present, with divergence emerging as PT favors reinforcement.

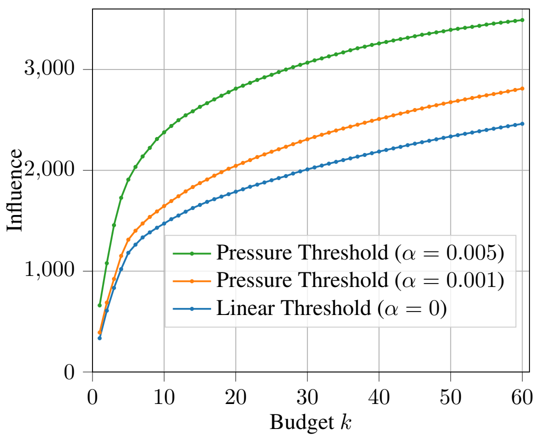


Figure 4: Influence vs. budget for the Facebook network. PT outperforms LT, especially for larger α.

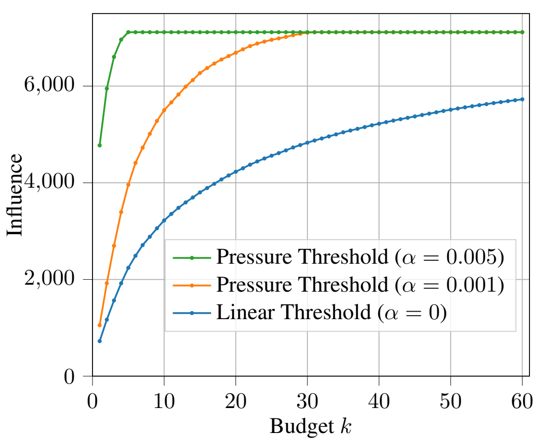


Figure 5: Influence vs. budget for the Wikipedia network, with PT consistently exceeding LT.

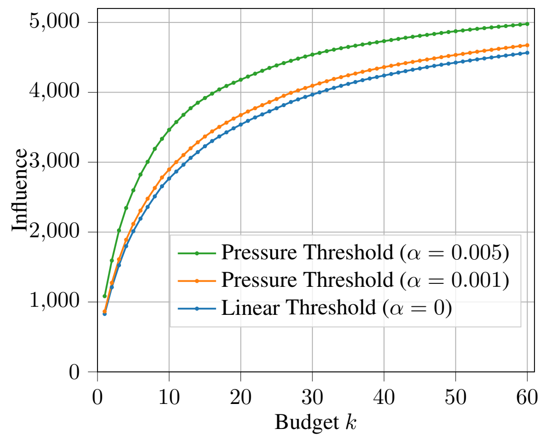


Figure 6: Influence vs. budget for the Bitcoin network. PT gains are modest due to network sparsity.

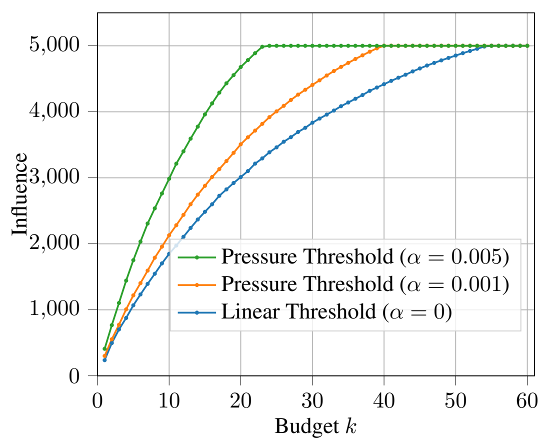


Figure 7: Influence vs. budget for the Erdős–Rényi random network. Amplification accelerates full coverage.

to characterize regimes with limited non-monotonicity, enabling improved, data-dependent guarantees and recovery of the classical $(1 - 1/e)$ limit as $\alpha \to 0$. In addition, validating the PT model on large-scale, time-stamped diffusion datasets and estimation of the pressure parameter α from empirical traces are also important directions.

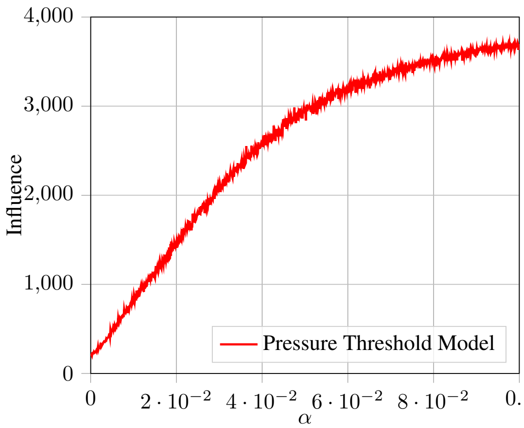


Figure 8: Effect of α on PT influence (smoothed using a centered moving average with window size of 9) for the Facebook network ($k = 10$); influence increases monotonically with reinforcement strength.

Ethical Statement

While the PT model provides a framework for understanding and optimizing information diffusion, such methods may also be applied to influence operations or targeted persuasion. We recognize the dual-use nature of this research and emphasize transparency, accountability, and fairness. When used in public communication or policy contexts, amplification and seed-selection mechanisms should be audited to prevent manipulative or discriminatory outcomes. All data used in this study are public and anonymized; no personally identifiable information was processed.

Acknowledgments

We thank Camille Woods for valuable discussions on the PT model's theoretical properties.

References

Agarwal, M.; Aggarwal, V.; Umrawal, A. K.; and Quinn, C. J. 2022. Stochastic Top K-Subset Bandits with Linear Space and Non-Linear Feedback with Applications to Social Influence Maximization. *ACM/IMS Transactions on Data Science (TDS)*, 2(4): 1–39.

Ananthasubramaniam, A.; Zhu, Y.; Jurgens, D.; and Romero, D. 2024. The Role of Network and Identity in the Diffusion of Hashtags. *arXiv preprint arXiv:2407.12771*.

Aral, S.; and Walker, D. 2012. Identifying Influential and Susceptible Members of Social Networks. *Science*, 337(6092): 337–341.

Bakshy, E.; Messing, S.; and Adamic, L. 2015. Political science. Exposure to ideologically diverse news and opinion on Facebook. *Science (New York, N.Y.)*, 348.

Becker, R.; D'Angelo, G.; Ghobadi, S.; and Gilbert, H. 2022. Fairness in Influence Maximization through Randomization. *Journal of Artificial Intelligence Research*, 73: 1251–1283.

Bhimaraju, A.; Robson, E. W.; Varshney, L. R.; and Umrawal, A. K. 2024. Fractional Budget Allocation for Influence Maximization under General Marketing Strategies. In *Proceedings of 33rd ACM International Conference on Information and Knowledge Management*, 3627–3631.

Borgs, C.; Brautbar, M.; Chayes, J.; and Lucier, B. 2014. Maximizing social influence in nearly optimal time. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, 946–957.

Brugnoli, E.; Cinelli, M.; Quattrociocchi, W.; and Scala, A. 2019. Recursive Patterns in Online Echo Chambers. *Scientific Reports*, 9: 20118.

Chen, W.; Wang, Y.; and Yang, S. 2009. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 199–208. Association for Computing Machinery. ISBN 9781605584959.

Chen, W.; Wu, R.; and Yu, Z. 2020. Scalable lattice influence maximization. *IEEE Transactions on Computational Social Systems*, 7(4): 956–970.

Cinelli, M.; De Francisci Morales, G.; Galeazzi, A.; Quattrociocchi, W.; and Starnini, M. 2021. The Echo Chamber Effect on Social Media. *Proceedings of the National Academy of Sciences*, 118(9): e2023301118.

Czuba, M.; Nurek, M.; Serwata, D.; Qiu, Y.-X.; Jia, M.; Musial, K.; Michalski, R.; and Bródka, P. 2024. Network Diffusion Framework to Simulate Spreading Processes in Complex Networks. *Big Data Mining and Analytics*, 7(3): 637–654.

Das, A.; and Kempe, D. 2018. Approximate submodularity and its applications: subset selection, sparse approximation and dictionary selection. 19(1): 74–107.

Domingos, P.; and Richardson, M. 2001. Mining the network value of customers. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 57–66. ISBN 158113391X.

Erdős, P.; and Rényi, A. 1959. On random graphs I. *Publicationes Mathematicae Debrecen*, 6(290-297): 18.

Evans, D.; and McKee, J. 2010. *Social Media Marketing: The Next Generation of Business Engagement*. John Wiley & Sons.

Feige, U.; Mirrokni, V. S.; and Vondrák, J. 2011. Maximizing non-monotone submodular functions. *SIAM Journal on Computing*, 40(4): 1133–1153.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Goldenberg, J.; Libai, B.; and Muller, E. 2001. Talk network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters*, 12(3): 211–223.

Goyal, A.; Lu, W.; and Lakshmanan, L. V. 2011. CELF++: optimizing the greedy algorithm for influence maximization in social networks. WWW '11, 47–48. Association for Computing Machinery. ISBN 9781450306379.

Granovetter, M. 1978. Threshold models of collective behavior. *American Journal of Sociology*, 83(6): 1420–1443.

Gunawan, C. M.; Rahmania, L.; and Kenang, I. H. 2023. The influence of social influence and peer influence on intention to purchase in e-commerce. *Review of Management and Entrepreneurship*, 7(1): 61–84.

Kalimeris, D.; Singer, Y.; Subbian, K.; and Weinsberg, U. 2018. Learning Diffusion using Hyperparameters. In Dy, J.; and Krause, A., eds., *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, 2420–2428. PMLR.

Kempe, D.; Kleinberg, J.; and Tardos, É. 2003. Maximizing the spread of influence through a social network. In *Proceedings of 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 137–146. ACM.

Kermack, W. O.; and McKendrick, A. G. 1927. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772): 700–721.

Kumar, S.; Hooi, B.; Makhija, D.; Kumar, M.; Faloutsos, C.; and Subrahmanian, V. 2018. Rev2: Fraudulent user prediction in rating platforms. In *Proceedings of Eleventh ACM International Conference on Web Search and Data Mining*, 333–341. ACM.

Kumar, S.; Mallik, A.; Khetarpal, A.; and Panda, B. 2022. Influence maximization in social networks using graph embedding and graph neural network. *Information Sciences*, 607: 1617–1636.

Kumar, S.; Spezzano, F.; Subrahmanian, V.; and Faloutsos, C. 2016. Edge weight prediction in weighted signed networks. In *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, 221–230. IEEE.

Kuo, T.-T.; Hung, S.-C.; Lin, W.-S.; de Lin, S.; Peng, T.-C.; and Shih, C. C. 2011. Assessing the Quality of Diffusion Models Using Real-World Social Network Data. *2011 International Conference on Technologies and Applications of Artificial Intelligence*, 200–205.

Leskovec, J.; Huttenlocher, D.; and Kleinberg, J. 2010a. Predicting positive and negative links in online social networks. In *Proceedings of 19th International Conference WWW*, 641–650.

Leskovec, J.; Huttenlocher, D.; and Kleinberg, J. 2010b. Signed networks in social media. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, 1361–1370.

Leskovec, J.; Krause, A.; Guestrin, C.; Faloutsos, C.; VanBriesen, J.; and Glance, N. 2007. Cost-effective outbreak detection in networks. In *Proceedings of 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 420–429.

Leskovec, J.; and Krevl, A. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data.

Leskovec, J.; and Mcauley, J. J. 2012. Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems*, 539–547.

Li, Y.; Gao, H.; Gao, Y.; Guo, J.; and Wu, W. 2023. A survey on influence maximization: From an ml-based combinatorial optimization. *ACM Transactions on Knowledge Discovery from Data*, 17(9): 1–50.

Lin, M.; Sun, L.; Yang, R.; Liu, X.; Wang, Y.; Li, D.; Li, W.; and Lu, S. 2023. Fair Influence Maximization in Large-scale Social Networks Based on Attribute-aware Reverse Influence Sampling. *Journal of Artificial Intelligence Research*, 76: 925–957.

Mønsted, B.; Sapieżyński, P.; Ferrara, E.; and Lehmann, S. 2017. Evidence of Complex Contagion of Information in Social Media: An Experiment Using Twitter Bots. *PLOS ONE*, 12(9): e0184148.

Nemhauser, G. L.; Wolsey, L. A.; and Fisher, M. L. 1978. An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14(1): 265–294.

Nguyen, B.-N. T.; Pham, P. N. H.; Le, V.-V.; and Snášel, V. 2022. Influence Maximization under Fairness Budget Distribution in Online Social Networks. *Mathematics*, 10(22).

Nie, G.; Agarwal, M.; Umrawal, A. K.; Aggarwal, V.; and Quinn, C. J. 2022. An Explore-then-Commit Algorithm for Submodular Maximization Under Full-bandit Feedback. In *The 38th Conference on Uncertainty in Artificial Intelligence*.

Robson, E. W.; Reddy, D.; and Umrawal, A. K. 2024. CyNetDiff: A Python Library for Accelerated Implementation of Network Diffusion Models. *Proceedings of the VLDB Endowment*, 17(12): 4409–4412.

Robson, E. W.; and Umrawal, A. K. 2025. A Community-Aware Framework for Influence Maximization with Explicit Accounting for Inter-Community Influence. *arXiv preprint arXiv:2512.23973*.

Umrawal, A. K.; and Aggarwal, V. 2023. Leveraging the Community Structure of a Social Network for Maximizing the Spread of Influence. *ACM SIGMETRICS Performance Evaluation Review*, 50(4): 17–19.

Umrawal, A. K.; Aggarwal, V.; and Quinn, C. J. 2023. Fractional Budget Allocation for Influence Maximization. In *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE.

Vosoughi, S.; Roy, D.; and Aral, S. 2018. The Spread of True and False News Online. *Science*, 359(6380): 1146–1151.

Xie, M.; Yang, Q.; Wang, Q.; Cong, G.; and Melo, G. 2015. DynaDiffuse: A Dynamic Diffusion Model for Continuous Time Constrained Influence Maximization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1).

Xu, J.; and Umrawal, A. K. 2026. LOFA: Online Influence Maximization under Full-Bandit Feedback using Lazy Forward Selection. *arXiv preprint arXiv:2601.00933*.

Zhuang, H.; Sun, Y.; Tang, J.; Zhang, J.; and Sun, X. 2013. Influence Maximization in Dynamic Social Networks. In *2013 IEEE 13th International Conference on Data Mining*, 1313–1318.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, the study is methodological and computational, using publicly available, anonymized datasets commonly used in research on network diffusion and influence maximization.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes, the abstract and introduction describe the Pressure Threshold (PT) model and its theoretical analysis and comparison with the Linear Threshold (LT) model.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, Sections 2 and 3 define the PT model and analyze its properties and complexity, and Section 4 provides experimental validation.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, Section 4.4 discusses amplification differences between dense and sparse networks, highlighting network-specific artifacts.
   (e) Did you describe the limitations of your work? Yes, limitations are discussed in Section 3.2 on non-monotonicity, Section 3.3 on loss of submodularity, and Section 5.3 on future work.
   (f) Did you discuss any potential negative societal impacts of your work? Yes, these are discussed in Section 5.4 on ethical considerations.
   (g) Did you discuss any potential misuse of your work? Yes, these are discussed in Section 5.4 on ethical considerations.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, all datasets are publicly available and anonymized (e.g., SNAP (Leskovec and Krevl 2014)). The CyNetDiff library (Robson, Reddy, and Umrawal 2024) is open-source, and experiments are described in sufficient detail to support reproducibility.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes, the study relies only on publicly available, anonymized datasets and conforms to applicable ethics review guidelines.
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Yes, Section 2.2 states all model assumptions, including node thresholds, edge weights, and the amplification parameter α.
   (b) Have you provided justifications for all theoretical results? Yes, Section 2.3 provides an NP-hardness proof, and Section 3 analyzes monotonicity and submodularity with proofs and counterexamples.
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes, Section 1.2 reviews alternative diffusion models, including the Linear Threshold (LT), Independent Cascade (IC), and dynamic diffusion models.
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes, Section 4.4 compares PT and LT across networks and attributes differences to structural properties such as network density.
   (e) Did you address potential biases or limitations in your theoretical framework? Yes, Sections 3.2 and 3.3 show that the PT model is non-monotone and non-submodular when the reinforcement parameter $\alpha > 0$, limiting the theoretical guarantees.
   (f) Have you related your theoretical results to the existing literature in social science? Yes, Section 1.1 motivates the PT model using reinforcement and echo chamber effects from prior social science work.
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes, Section 5.2 discusses implications for misinformation mitigation, awareness campaigns, and policymaking.
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? Yes, Section 2.2 defines the assumptions underlying the PT model.
   (b) Did you include complete proofs of all theoretical results? Yes, proofs are included for NP-hardness in Section 2.3, non-monotonicity in Section 3.2, and non-submodularity in Section 3.3.
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA, as the experiments involve diffusion simulations rather than training machine learning models.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA.
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA.
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA.

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes, SNAP datasets (Leskovec and Krevl 2014) and their original sources (Leskovec and Mcauley 2012; Leskovec, Huttenlocher, and Kleinberg 2010a,b; Kumar et al. 2016, 2018) are properly cited.
   (b) Did you mention the license of the assets? Yes, all datasets follow the licenses provided by SNAP (Leskovec and Krevl 2014) and the original sources.
   (c) Did you include any new assets in the supplemental material or as a URL? No new datasets or assets are released.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? NA, as all datasets are pre-existing, public, and anonymized.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes, Section 5.4 clarifies that the datasets are anonymized, structural, and do not contain personally identifiable information.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? NA.
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? NA.
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? NA, as no human participants are involved.
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA.
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA.
   (d) Did you discuss how data is stored, shared, and deidentified? Yes, Section 5.4 clarifies that the study relies exclusively on publicly available datasets that have been anonymized, and none of the data used contain personal, sensitive, or identifying information.