

Variable-Length Evolutionary Search for Influence Maximization with Influence, Cost and Time

Mingyang Feng

11856010@mail.sustech.edu.cn

University of Birmingham

Qi Zhao

Southern University of Science and Technology

Shan He

University of Birmingham

Yuhui Shi

Southern University of Science and Technology

Research Article

Keywords: Influence Maximization, Multiobjective, Evolutionary Algorithm

Posted Date: December 4th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-8152524/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Variable-Length Evolutionary Search for Influence Maximization with Influence, Cost and Time

Mingyang Feng^{a,b}, Qi Zhao^a, Shan He^b, Yuhui Shi^{a,*}

^aDepartment of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055, China

^bSchool of Computer Science, The University of Birmingham, Birmingham, B15 2TT, UK

Abstract

Influence maximization has been studied for social network analysis, such as viral marketing (advertising), rumor prevention, and opinion leader identification. However, most studies neglect the interplay between influence spread, cost efficiency, and temporal urgency. In practical scenarios such as viral marketing and information campaigns, jointly optimizing Influence, Cost, and Time is essential, yet remaining largely unaddressed in current literature. To bridge the gap, this paper proposes a new multi-objective influence maximization problem that simultaneously optimizes influence, cost, and time. We show the intuitive and empirical evidence to prove the feasibility and necessity of this multi-objective problem. We also develop an evolutionary variable-length search algorithm that can effectively search for optimal node combinations. The proposed EVEA algorithm outperforms all baselines, achieving up to 19.3% higher hypervolume and 25–40% faster convergence across four real-world networks, while maintaining a diverse and balanced Pareto front among influence, cost, and time objectives. We release the source code on GitHub: <https://github.com/fmyzckj>.

Keywords: Influence Maximization, Multiobjective, Evolutionary Algorithm

1. Introduction

Influence Maximization (IM) is a foundational problem in social network analysis, focusing on selecting a small set of seed users to maximize information spread across the network via mechanisms like social influence or viral propagation. Pioneered by Kempe et al. [1] with the Independent Cascade (IC) and Linear Threshold (LT) models, IM has become pivotal in both academia and practice. Its applications span viral marketing, where businesses leverage influential users to amplify product adoption [2]; public health, aiding in identifying super-spreaders for targeted interventions [3]; and social recommendation systems, enhancing engagement through peer influence [4].

In this paper, we formalize a novel multi-objective influence maximization problem that simultaneously optimizes three conflicting criteria: influence coverage, activation cost, and propagation time (termed as IM-ICT). To demonstrate the necessity of this formulation, consider the scenario in Figure 1 where a supermarket promotes a time-sensitive sale through a social network of 10 users (A-J). The shopkeeper must select two seed users under constrained budgets and promotional deadlines. We define node activation costs proportional to their degrees (reflecting influencer pricing) and propagation time as the longest path from seeds to influenced nodes. Figure 1(a) illustrates a maximal influence strategy selecting high-degree seeds A (degree=2) and B (degree=3), achieving full network coverage at

*Yuhui Shi is the corresponding author

Email addresses: 11856010@mail.sustech.edu.cn (Mingyang Feng), zhaoq@sustech.edu.cn (Qi Zhao), s.he@cs.bham.ac.uk (Shan He), shiyh@sustech.edu.cn (Yuhui Shi)

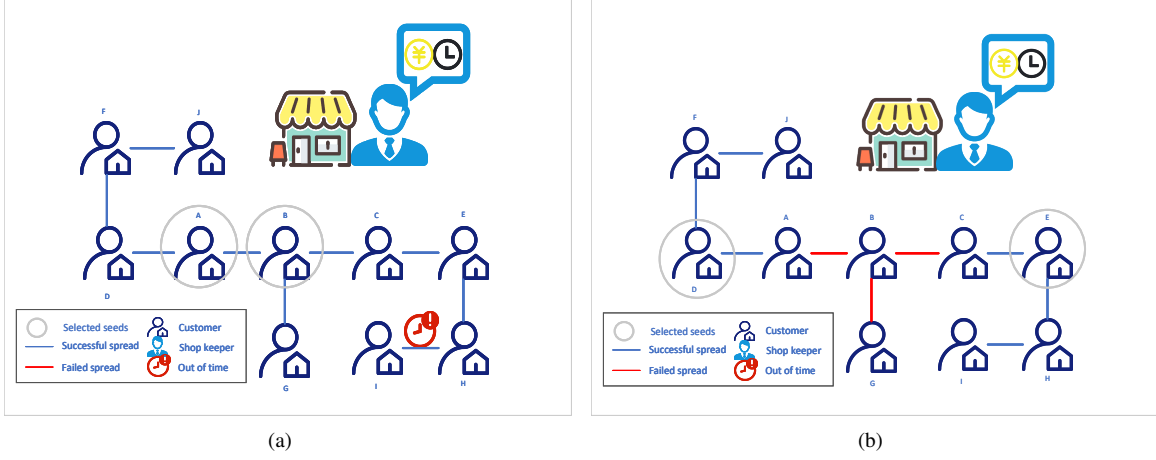


Figure 1: Supermarket

a total cost of 5 units. However, the critical path $B \rightarrow C \rightarrow H \rightarrow I$ requires 4 time steps, exceeding the permissible 3-step threshold. By contrast, Figure 1(b) demonstrates a Pareto-optimal solution choosing seeds D (degree=2) and E (degree=2), which reduces both cost (4 units) and time (2 steps) while sacrificing partial coverage (nodes B and G remain unaffected). This contrast highlights fundamental trade-offs between the objectives: high influence demands premium seeds and extended propagation horizons, whereas cost-time efficiency necessitates strategic coverage compromises. Beyond commercial promotions, our IM-ICT formulation also enables nuanced decision-making in public health (e.g., balancing vaccine distribution costs, immunization speed, and population coverage) and cybersecurity (e.g., minimizing containment costs while maximizing malware eradication speed and scope).

The optimization of the proposed IM-ICT problem incurs two algorithmic challenges. First, we rigorously prove the proposed IM-ICT’s NP-hardness via a polynomial-time reduction from the Set Cover problem. In addition to the NP-hard challenge, the length of solutions requires variability to enable the algorithm to naturally evolve seed sets with different sizes to suit the diverse optimization pressures from conflicting objectives, leading to a more complete and expressive Pareto front. To address these challenge, we propose an Embedding-aligned Variable-length Evolutionary Algorithm (EVEA), which introduces a graph embedding-driven crossover operator and enables variable-length search by aligning the nodes with embedding vectors.

In summary, our contributions are:

1. The new IM-ICT problem formulation with real-world constraints. Unlike prior single- or bi-objective formulations, our model explicitly integrates propagation time as the longest path length from seeds to influenced nodes, reflecting real-world urgency requirements (e.g., limited-duration sales or epidemic containment). We provide both theoretical justification (NP-hardness proof) and empirical evidence (case studies in Figures 1–2) to validate the necessity of this new formulation.
2. The Embedding-aligned Variable-length Evolutionary Algorithm for solving IM-ICT. The proposed algorithm’s core innovation lies in its crossover operator, which aligns parent solutions using node embedding vectors. This enables the recombination of seed sets with varying lengths while preserving topological semantics. Such operations enhance search space exploration and accelerate convergence.
3. Comprehensive Empirical Validation. Experiments on real-world networks (Twitter, DBLP) demonstrate EVEA’s superiority over state-of-the-art multi-objective optimizers.

The remainder of this paper is organized as follows: Section 2 reviews related work on IM and multi-objective optimization. Section 3 formally defines the proposed IM-ICT problem and analyze its complexity. Section 4 details EVEA’s architecture, focusing on the embedding-aligned crossover mechanism. Section 5 presents experimental results, and Section 6 concludes with future directions.

2. Related Works

This section provides a comprehensive overview of the IM problem, its foundational formulations, and recent advances in both single-objective and multi-objective settings. We first introduce the problem definition and diffusion models, followed by a discussion of classical single-objective approaches, and finally review the state-of-the-art in multi-objective influence maximization.

2.1. Problem Formulation: Influence Maximization in Networks

IM is a fundamental problem in network science, originally proposed by Kempe et al. [1], which aims to select a small set of seed nodes within a social network to maximize the expected number of activated nodes under a chosen diffusion model.

Formally, consider a directed graph $G = (V, E)$, where V represents the set of nodes and E the set of edges. Each edge (u, v) has an associated propagation probability $p_{uv} \in [0, 1]$. The goal is to select a seed set $S \subseteq V$ of size k that maximizes the expected influence spread, defined as:

$$S^* = \arg \max_{S \subseteq V, |S|=k} \sigma(S) \quad (1)$$

where $\sigma(S)$ denotes the expected number of activated nodes when the diffusion process starts from S .

The most widely used diffusion models in IM are the *IC model* and the *LT model*. In the IC model, diffusion proceeds in discrete time steps. Each active node u has a single chance to activate each of its inactive neighbors v , with probability p_{uv} . The process continues until no further activations occur. The expected influence spread under the IC model is formally defined as:

$$\sigma(S) = \mathbb{E} \left[\sum_{v \in V} \mathbb{I}(v \text{ is activated}) \right] \quad (2)$$

where $\mathbb{I}(\cdot)$ is the indicator function.

In the LT model, each node v is assigned a threshold $\theta_v \in [0, 1]$. A node becomes active when the sum of incoming weights from its active neighbors exceeds its threshold. The LT model captures cumulative influence effects, whereas the IC model emphasizes independent, probabilistic activations.

To address time-sensitive scenarios, recent extensions like the *Latency-Aware Independent Cascade (LAIC) model* incorporate temporal delays into the diffusion process. In this model, each activation attempt from node u to node v introduces a delay δt sampled from a distribution $P_u(\delta t)$. The propagation time $T(S)$ of a seed set S can be expressed as the maximum activation time across all influenced nodes:

$$T(S) = \max_{v \in V} \left\{ \min_{u \in S} (t(u) + d(u, v) \cdot \delta t) \right\} \quad (3)$$

where $d(u, v)$ is the shortest path length between u and v in the network, and $t(u)$ is the activation time of the seed node u (typically $t(u) = 0$).

The Influence Maximization problem is known to be NP-hard under both the IC and LT models, and approximating the expected spread $\sigma(S)$ is #P-hard. Consequently, solving IM requires either approximation algorithms or heuristic approaches that balance accuracy and efficiency.

2.2. Influence Maximization: Single-Objective Approaches

Since its introduction, Influence Maximization has been predominantly studied as a single-objective optimization problem, focusing on maximizing the spread of influence. Kempe et al. [1] proposed a greedy algorithm that achieves a $(1 - 1/e)$ -approximation guarantee under submodular diffusion models, but the method's high computational cost—due to the need for repeated Monte Carlo simulations—prompted subsequent research into more efficient techniques. Leskovec et al. [5] introduced the CELF algorithm, which improves greedy selection efficiency by exploiting submodularity properties to minimize redundant computations. Further advances include sketch-based methods such as TIM and IMM [6, 7], which approximate influence spread via reverse reachable (RR) sets, significantly reducing computation time while maintaining high solution quality.

Beyond exact or approximate methods, heuristic strategies like degree centrality, PageRank, and community detection have been widely explored for their scalability [8, 9, 10, 11, 12], albeit at the expense of accuracy and theoretical guarantees. In parallel, evolutionary and swarm-based algorithms—such as particle swarm optimization (PSO) [13, 14] and biogeography-based optimization (BBO) [15, 16]—have been proposed to balance influence spread and computational cost by exploring large solution spaces using population-based search strategies. These methods have demonstrated success in handling larger networks but remain limited by their focus on a single objective: maximizing the number of influenced nodes, without considering factors like seed acquisition costs, propagation delays, or fairness constraints that often arise in real-world applications.

2.3. Multi-Objective Influence Maximization: Recent Advances

Recognizing the limitations of single-objective formulations, researchers have increasingly focused on Multi-Objective Influence Maximization, where influence spread is optimized alongside other conflicting objectives such as seed cost, propagation latency, and fairness [17, 18, 19, 20, 21]. The trade-off between influence and cost has been a primary focus, as high-impact seed nodes—often those with high degree or centrality—are typically associated with higher acquisition costs in practical scenarios. Several studies have proposed algorithms to address this trade-off. Wang and Zhang [22] introduced a Multi-Objective Crow Search Algorithm (MOCSA) that simultaneously maximizes influence spread and minimizes seed costs, demonstrating superior performance across real-world networks. Yang and Liu [14] proposed a multi-objective particle swarm optimization framework for influence maximization under budget constraints, offering diverse Pareto-optimal solutions that balance influence and cost. De et al. [16] developed a biogeography-based multi-objective algorithm that optimizes influence spread while reducing budget expenditure, and Gershtein et al. [20] presented IM-Balanced, a system allowing users to specify trade-offs between objectives, enabling tailored solutions for different applications.

Another major research direction in MOIM is the inclusion of propagation latency as an optimization objective. In time-sensitive scenarios, such as viral marketing campaigns or emergency information dissemination, the speed at which influence spreads is as critical as the extent of coverage. Chen et al. [17] were among the first to explore time-critical influence maximization, incorporating propagation delays into the IC and LT models. Li et al. (2020) extended this line of work by proposing a latency-aware influence maximization model that optimizes both influence and diffusion speed. Wang et al. [23] developed a multi-objective evolutionary algorithm that minimizes both latency and seed cost while maintaining competitive influence spread.

Fairness in influence propagation has also gained attention in recent years. In heterogeneous networks, ensuring equitable influence distribution across different groups—such as demographic segments, geographic regions, or interest communities—is crucial for ethical and practical reasons. Tsang et al. [18] introduced a group fairness-aware influence maximization framework that ensures balanced coverage across diverse user groups. Staib and Jegelka [24] studied constrained submodular maximization problems, applying fairness constraints to influence maximization settings. Farnad et al. [25] designed algorithms that jointly optimize for influence spread and fairness metrics, such as equal opportunity, while Rahmattalabi et al. [26] proposed fairness-constrained formulations to address exposure disparities across communities.

Beyond these primary objectives, recent studies have explored multi-objective influence maximization in dynamic networks, competitive settings, and under additional constraints such as privacy, energy efficiency, and physical distance [27, 21, 28, 29, 30]. Despite these advances, challenges remain in designing scalable algorithms that effectively balance multiple objectives while adapting to real-world complexities.

3. Proposed IM-ICT Problem

3.1. Objectives

Building upon the conventional IM definition in Section 2.1, we formalize our proposed model with three optimization objectives:

Objective 1: Influence Spread Maximization: Consistent with traditional IM, our primary objective aims to maximize the number of influenced nodes in the network. This can be expressed as:

$$f_1 = \sigma(S) = \mathbb{E} \left[\sum_{v \in V} \mathbb{I}_{(v \text{ activated})} \right] \quad (4)$$

where node v becomes activated through multi-round diffusion if any predecessor $u \in \mathcal{N}^-(v)$ successfully influences it through edge (u, v) .

Objective 2: Seed Cost Minimization: Under the IM framework that assumes activation of k seed nodes at initial time, we introduce a cost set C representing activation expenses for each node (extending the network definition to $G = (V, E, C)$). The secondary objective minimizes the total seed acquisition cost:

$$f_2 = B(S) = \sum_{v_i \in S} c_i \quad (5)$$

where c_i represents the activation cost of node v_i in seed set S , and $B(S)$ denotes the total budget required for seed set S .

Objective 3: Propagation Latency (Time) Minimization: Under the Latency-Aware Independent Cascade (LAIC) model, we define propagation latency as the latest activation time among all influenced nodes (To simplify this, we call this objective time in this paper):

$$f_3 = T(S) = \max_{v_i \in S, v_j \in V} \left[t_i + \sum_{k=1}^{d(v_i, v_j)} \delta_{t_k} \right] \quad (6)$$

where t_i indicates the initial activation time of seed v_i (typically $t_i = 0$), $d(v_i, v_j)$ represents the shortest path length from v_i to v_j , and δ_{t_k} denotes the activation delay at the k -th hop sampled from distribution $P_u^{lat}(\delta_t)$.

As analyzed in Section I, inherent conflicts exist among these objectives:

1. Influence-Cost Conflict: Maximizing influence spread typically requires selecting high-impact opinion leaders as seeds, which substantially increases activation costs.
2. Influence-Time Tradeoff: Broad coverage often necessitates longer propagation paths, thereby increasing time delays.
3. Cost-Time Dilemma: Cost minimization strategies tend to select peripheral nodes, which may compromise temporal effectiveness due to their marginal network positions.

3.2. IM-ICT

Therefore, the IM-ICT problem can be formulated as follows:

$$\begin{aligned} \max F &= (f_1, -f_2, -f_3) \\ \text{s.t. } & S \subseteq V \text{ \& } |S| \geq 1 \\ & c_i \geq 0, \quad \forall v_i \in V \\ & p_{uv} \in [0, 1] \\ & \delta_t \sim P_u^{lat}(\delta_t) \text{ \& } \delta_t \geq 0 \end{aligned} \quad (7)$$

where p_{uv} denotes the propagation probability from node u to v .

4. Proposed Embedding-aligned Variable-length Evolutionary Algorithm

In this section, we present the design and working principles of our proposed algorithm, the Embedding-aligned Variable-length Evolutionary Algorithm (EVEA), developed to tackle the IM-ICT problem under multiple conflicting criteria.

4.1. Overall framework

EVEA is a population-based evolutionary algorithm that optimizes influence maximization solutions by balancing competing objectives including influence spread, activation cost, and time. Unlike traditional approaches, EVEA adopts a variable-length representation for candidate solutions, allowing the number of seed nodes in each solution to

Algorithm 1: Embedding-aligned Variable-length Evolutionary Algorithm

Input: $G = (V, E, C)$, Generation G_{max} , Network size N

Output: Pareto front p_f of population P_t

```
1 Generation counter  $G_t = 0$ 
2  $P^G \leftarrow \text{Initialization}()$ 
3 while stopping criterion not met do
4    $P^G \leftarrow \text{Tournament\_Selection}(P^G)$ 
5    $Q \leftarrow \text{Embedding - Aligned\_Crossover}(Q)$ 
6    $Q \leftarrow \text{Variable - Length\_Mutation}(Q)$ 
7    $P_G \leftarrow \text{Environmental\_Selection}(P_G \cup Q)$ 
8    $G_t = G_t + 1$ 
9  $P_t = P_G$ 
10  $p_f \leftarrow P_t$ 
```

adaptively evolve during the search process. This flexibility enables the algorithm to explore a diverse range of trade-offs and achieve a well-distributed Pareto front. Additionally, the algorithm integrates embedding-based knowledge to enhance solution recombination, further promoting efficiency and convergence in the multi-objective setting.

The pseudo-code of EVEA is shown in Algorithm 1. The algorithm starts with the initialization of a population of candidate solutions, each representing a distinct seed set. At each generation, a tournament selection is performed to select parent solutions. The selected parents undergo an embedding-aligned crossover to produce offspring, followed by a variable-length mutation operation that modifies the seed set size and composition. The combined parent and offspring populations are then subjected to a non-dominated sorting selection (as in NSGA-II [31]) to maintain Pareto-optimal solutions. The algorithm iterates until a termination condition is met (e.g., a fixed number of generations). In the following, we detail the core innovations of the algorithm, i.e., the embedding aligned crossover and variable-length mutation, respectively.

4.2. Embedding Aligned Crossover

The core innovation of EVEA lies in its embedding-aligned crossover operator (Algorithm 2). Unlike traditional crossover operators that require equal-length representations, this operator aligns nodes from two parent solutions using their graph embedding vectors (e.g., derived from Node2Vec [32] or DeepWalk [33], in this paper, we choose to use Node2Vec). Specifically, given two parent solutions S_1 and S_2 , we compute the pairwise Euclidean distances between their nodes' embeddings. Nodes with the smallest distances are paired, and each pair has a probability p_c of swapping their membership across solutions.

To further clarify the embedding-aligned crossover process, we provide a visual illustration in Figure 2, which demonstrates the key steps of the operator. In this figure, part (a) depicts two parent solutions represented in a two-dimensional embedding space, where nodes are positioned according to their topological and semantic similarity. Part (b) shows the alignment process, where nodes from different parents are paired based on their proximity in the embedding space. Part (c) illustrates the actual crossover operation, where aligned node pairs are exchanged with a certain probability, facilitating the recombination of seed sets. Finally, part (d) presents the offspring solutions generated through this crossover, highlighting how the embedding-guided approach preserves structural semantics while enabling diversity in solution construction.

By leveraging the embedding-based alignment, this crossover strategy allows EVEA to recombine solutions of varying seed set sizes effectively, avoiding the limitations of index-based or fixed-length operators. This design enhances the algorithm's ability to explore the solution space meaningfully, promoting better convergence toward high-quality Pareto fronts.

4.3. Variable Length Mutation

To further promote diversity and adaptability, EVEA incorporates a variable-length mutation operator (Algorithm 3). At each mutation step, one of the following strategies is randomly selected with equal probability:

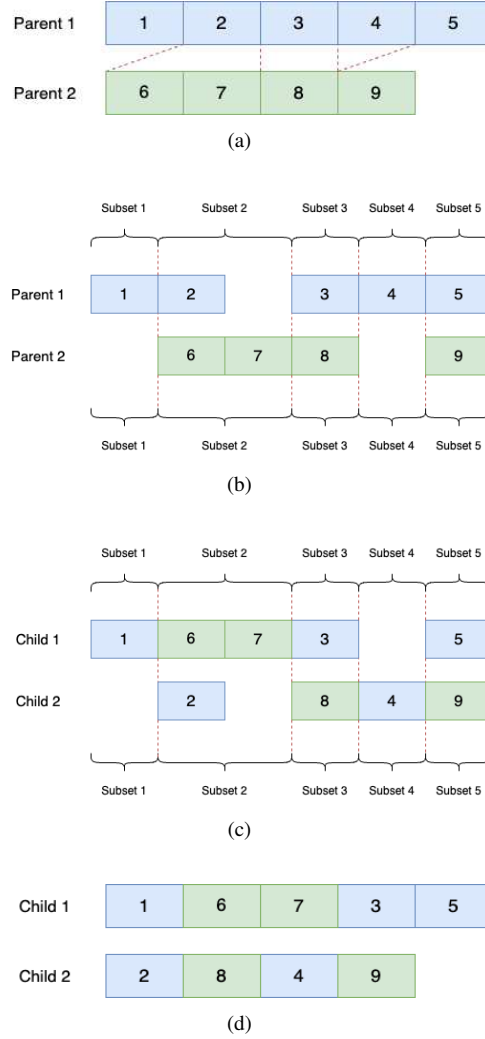


Figure 2: Embedded crossover

- Add: A random node not currently in the solution is added to the seed set.
- Delete: A random node from the solution is removed, reducing the seed set size.
- Replace: A random node in the solution is swapped with another node outside the solution.

These operations enable EVEA to flexibly adjust the size of the seed set, allowing the algorithm to navigate complex trade-offs and better adapt to the multi-objective nature of the problem.

4.4. Time Complexity Analysis

The time complexity of the proposed algorithm is primarily determined by the multi-objective selection mechanism, which ensures the preservation of high-quality and diverse solutions across generations. In each iteration, the algorithm evaluates a population of N candidate solutions and applies a selection strategy based on Pareto dominance relations and crowding distance calculations to maintain a well-distributed Pareto front. This selection step has a time complexity of $O(N^2)$.

Additionally, the embedding-aligned crossover operator involves at most $\min(k_1, k_2)$ pairwise distance computations between seed nodes in the parent solutions, where k_1 and k_2 are the lengths of the respective parent solutions.

Algorithm 2: Embedding Aligned Crossover

Input: Parents S_1, S_2 , Embedding EM , Crossover Rate p_c

Output: Children C_1, C_2

- 1 Align nodes in S_1, S_2 by embedding (Euclidean Distance)
 - 2 For each paired nodes, exchange nodes with probability p_c
 - 3 Output C_1, C_2
-

Algorithm 3: Variable length Mutation

Input: Solution S , Mutation rate p_m

Output: Solution S^*

- 1 $Strategy \leftarrow Random(\{add, delete, replace\})$
 - 2 **if** $Strategy = add$ **then**
 - 3 $v \leftarrow G - S$
 - 4 $S^* \leftarrow S \cup \{v\}$
 - 5 **else if** $Strategy = delete$ **then**
 - 6 $v \leftarrow S$
 - 7 $S^* \leftarrow S \setminus \{v\}$
 - 8 **else if** $Strategy = replace$ **then**
 - 9 $v_G \leftarrow G$
 - 10 $v_S \leftarrow S$
 - 11 $S^* \leftarrow S \cup \{v_G\} \setminus \{v\}$
 - 12 Output S^*
-

Since $k_1, k_2 \ll |V|$ (with $|V|$ being the number of nodes in the network), this step introduces negligible overhead relative to the overall selection process.

Therefore, the overall time complexity of each generation is $O(N^2)$, ensuring scalability and efficiency while maintaining the algorithm’s ability to effectively balance trade-offs between influence spread, activation cost, and time.

5. Experimental Results

This section presents the experimental results designed to evaluate the effectiveness and robustness of the proposed algorithm. We conduct experiments on four real-world network datasets and compare EVEA with several representative baselines. The results are analyzed from four perspectives: objective conflict demonstration, quantitative performance comparison, convergence behavior, and distributional analysis.

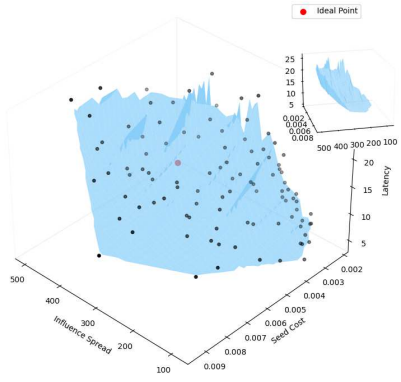
5.1. Experimental Settings

The experiments are conducted on four benchmark datasets - Facebook, GRQC, and Gnutella, all of which are obtained from the Stanford Network Analysis Project (SNAP) [34]. Each network exhibits different topological characteristics in terms of size, connectivity, and clustering. Table 1 summarizes the key statistics of these datasets.

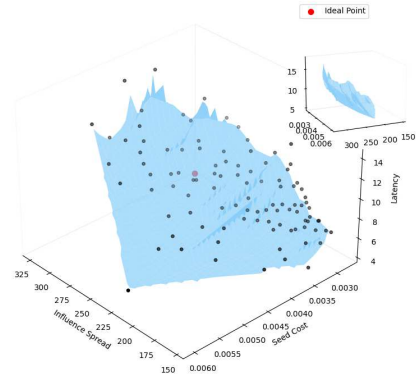
We compare EVEA against three variants of NSGA-II: (1) standard NSGA-II [31], (2) NSGA-II with variable-length crossover [35], and (3) NSGA-II with variable-length mutation [36]. All algorithms are run for 1000 generations, and the population size is fixed across methods for fairness. The hypervolume (HV) indicator is used to assess the quality of the obtained Pareto fronts. We also report the average HV and standard deviation over multiple runs.

5.2. Demonstrating Objective Conflicts

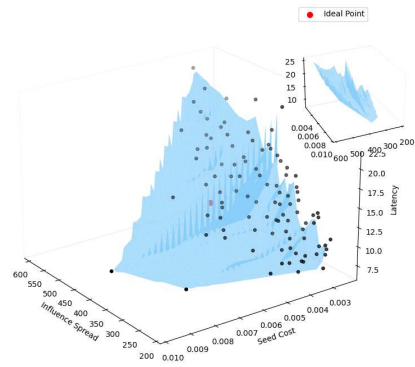
To clearly illustrate the conflicting nature of the three objectives—maximizing influence spread, minimizing seed cost, and minimizing time—we present 3D Pareto front visualizations for four representative networks: Facebook,



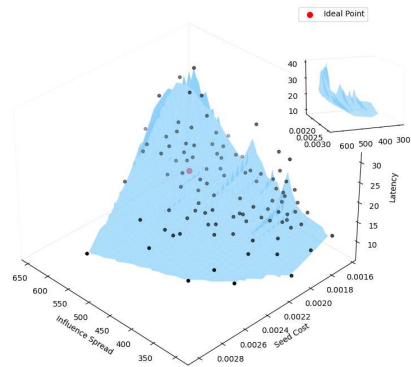
(a) Facebook



(b) Grqc



(c) Gnutella



(d) Wiki

Figure 3: The Pareto-front of EVEA

Table 1: Network Description

Network Name	Nodes#	Edges#	Description
Facebook	4039	88234	Social circles from Facebook (anonymized)
GRQC	5242	14496	Collaboration network of Arxiv General Relativity
Gnutella	6301	20777	Gnutella peer-to-peer network from August 8 2002
Wiki	7115	103689	Wikipedia who-votes-on-whom network

GRQC, Gnutella, and Wiki (Figure3). In each case, the surface formed by the non-dominated solutions reveals a continuous and well-structured Pareto front in three-dimensional space. These surfaces are not trivial point clouds, but rather demonstrate smooth trade-off regions, indicating that the solutions are not concentrated at extreme points but instead spread across a spectrum of compromises among the objectives. The geometry of the Pareto surfaces, particularly the curvature and spread in each plot, further confirms that no single solution simultaneously optimizes all three objectives. Enhancing one objective, such as influence spread, inevitably results in a higher seed cost or longer time. Therefore, the presence of stable and interpretable Pareto surfaces across all four datasets validates the multi-objective nature of the problem and highlights the necessity of using a Pareto-based approach such as EVEA.

5.3. Performance Comparison on Benchmark Networks

We report the quantitative performance of EVEA and the baseline methods on four representative networks using the hypervolume (HV) indicator. The results, shown in Table 2, reflect the average HV over multiple independent runs, with standard deviations provided in parentheses. The highest mean HV in each network is boldfaced, and statistically significant improvements by EVEA (based on the Wilcoxon signed-rank test at $p < 0.05$) are marked with an asterisk (*).

Overall, EVEA consistently outperforms all baseline methods across all four networks. On the **Facebook** network, EVEA achieves an HV of 1.0920, which is significantly higher than NSGA-II (0.9528), its crossover variant (0.8873), and the mutation variant (1.0137). This corresponds to relative improvements of 14.6%, 23.1%, and 7.7%, respectively. A similar trend is observed on the **GRQC** network, where EVEA obtains the highest HV of 1.1023, compared to 0.9237 for NSGA-II, 0.7687 for the crossover variant, and 1.0090 for the mutation variant, reflecting margins of 19.3%, 43.3%, and 9.2%.

On the **Gnutella** network, EVEA again yields the highest HV (1.1100), outperforming the second-best mutation variant (1.0628) by 4.4% and clearly surpassing NSGA-II (0.9705) and its crossover-enhanced version (0.7969) by 14.3% and 39.3%, respectively. The advantage of EVEA is even more pronounced on the **Wiki** network, where it reaches an HV of 1.1659 with minimal variance, while the other methods remain notably behind—NSGA-II (1.0536), VM (1.1096), and VC (0.8987)—resulting in improvement margins of 10.6%, 5.1%, and 29.7%.

These consistent improvements confirm the superiority of EVEA in discovering well-distributed and high-quality Pareto-optimal solutions across varying network topologies. The integration of embedding-aligned variable-length mutation and crossover operators enhances the flexibility of the search process, allowing the algorithm to adaptively explore diverse solution lengths and configurations with high robustness and effectiveness.

Table 2: Hypervolume (HV) Results with Standard Deviation on Four Networks. Values marked with * indicate statistical significance ($p < 0.05$) based on the Wilcoxon signed-rank test compared to all baselines.

Algorithm	Facebook	GRQC	Gnutella	Wiki
EVEA	1.0920 $\pm$ 0.0020*	1.1023 $\pm$ 0.0299*	1.1100 $\pm$ 0.0136*	1.1659 $\pm$ 0.0004*
NSGA-II	0.9528 $\pm$ 0.0167	0.9237 $\pm$ 0.0299	0.9705 $\pm$ 0.0472	1.0536 $\pm$ 0.0183
NSGA-II + Crossover	0.8873 $\pm$ 0.0269	0.7687 $\pm$ 0.0343	0.7969 $\pm$ 0.0304	0.8987 $\pm$ 0.0436
NSGA-II + Mutation	1.0137 $\pm$ 0.0244	1.0090 $\pm$ 0.0330	1.0628 $\pm$ 0.0196	1.1096 $\pm$ 0.0231

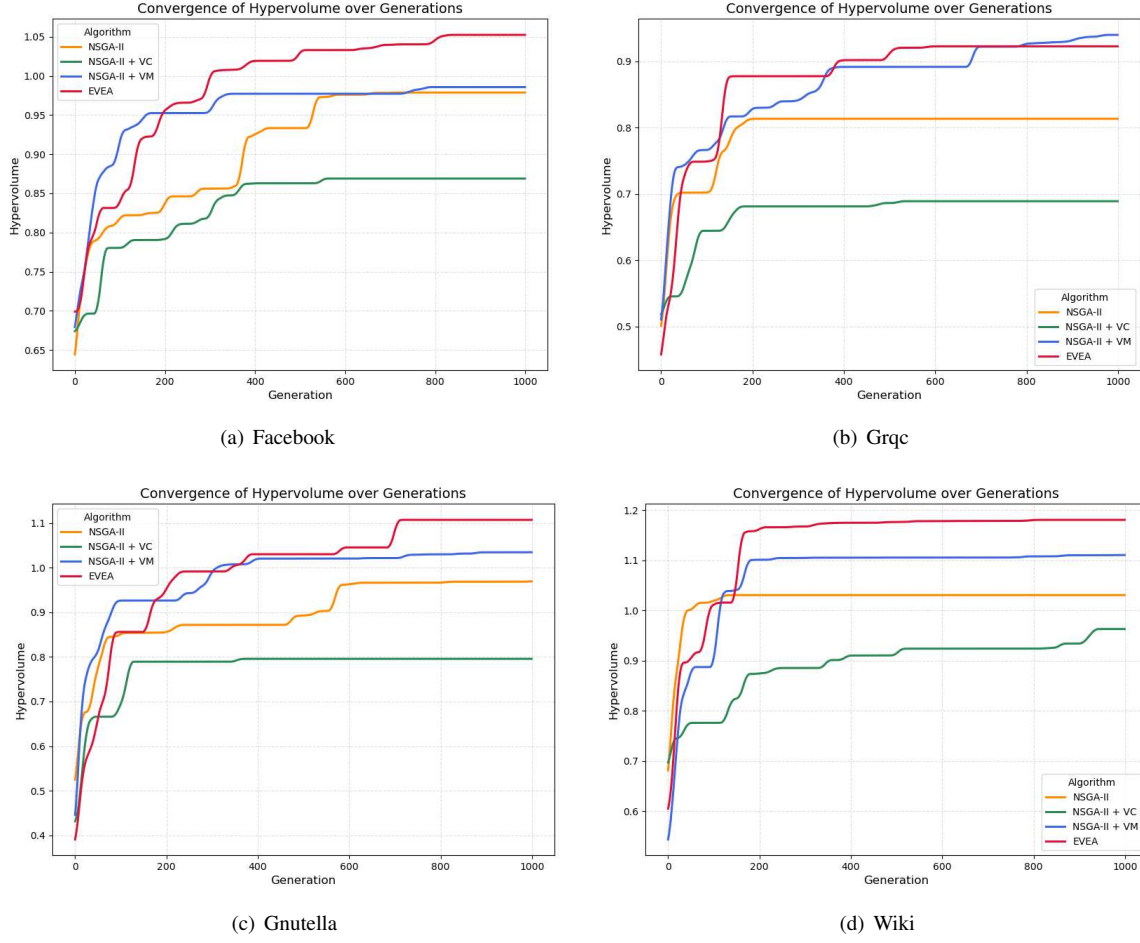


Figure 4: Convergence curves (HV vs. generations) of four algorithms across four networks.

5.4. Convergence Analysis

To assess the convergence behavior of the algorithms, we plot the evolution of hypervolume (HV) values over generations. The convergence curves, shown in Figure 4, illustrate the progression of Pareto front quality over time for each method across four benchmark networks.

EVEA consistently demonstrates both faster convergence and higher final HV values than all baseline algorithms. On the **Facebook** network, EVEA achieves over 90% of its final HV within the first 50 generations, while NSGA-II and its variants require more than 100 generations to reach comparable levels. For example, at generation 50, EVEA reaches an HV of approximately 1.05, which is 11% higher than that of NSGA-II (0.947). This trend continues, and by generation 100, EVEA has reached a stable HV near its maximum of 1.0920, clearly surpassing all baselines.

A similar pattern appears on the **GRQC** network, where EVEA steadily ascends to an HV of 1.1023, while NSGA-II converges more slowly and plateaus earlier. Notably, the HV of EVEA at generation 100 is 12.4% higher than that of NSGA-II and nearly 44% higher than that of NSGA-II + crossover. EVEA also maintains low variance throughout the run, highlighting its stability and robustness in complex search landscapes.

On the **Gnutella** network, EVEA reaches an HV of 1.09 by generation 100 and ultimately stabilizes at 1.1100. Compared to NSGA-II + crossover (0.7969), this represents a 39.3% relative improvement. Meanwhile, NSGA-II + mutation performs better than the standard NSGA-II, but still trails EVEA by a significant margin.

Lastly, for the **Wiki** network, EVEA shows the most stable and rapid convergence among all methods, attaining an HV of 1.1659 with negligible variance (± 0.0004). Other methods lag noticeably, with NSGA-II and its crossover

variant showing both slower ascent and lower final values.

In summary, EVEA achieves superior convergence characteristics across all datasets—both in speed and solution quality—demonstrating its effectiveness in navigating the complex multi-objective search space and delivering high-quality, stable Pareto fronts in fewer generations.

6. Conclusion

In this paper, we proposed a novel formulation of the Influence Maximization problem that simultaneously optimizes three conflicting objectives: influence spread, seed cost, and time. This tri-objective formulation, referred to as IM-ICT, captures the practical trade-offs faced in real-world scenarios such as viral marketing and time-sensitive information campaigns. We demonstrated that these objectives are inherently in conflict through comprehensive empirical analysis on multiple real-world networks.

To solve the IM-ICT problem effectively, we developed an Embedding-aligned Variable-length Evolutionary Algorithm (EVEA). EVEA introduces two key innovations: a variable-length representation that enables flexible seed set sizes, and an embedding-guided crossover operator that preserves topological semantics while enhancing search diversity. These design choices allow EVEA to efficiently explore the solution space and maintain a well-distributed Pareto front.

Experimental results on four benchmark networks validated the superiority of EVEA over several baselines in terms of convergence speed, solution quality, and hypervolume performance. The 3D Pareto surfaces further illustrated that EVEA can uncover a wide spectrum of trade-off solutions, confirming the multi-objective nature of the problem and the importance of Pareto-based optimization.

For future work, we plan to extend our model to incorporate spatial-temporal constraints. In the current framework, time is modeled independently of node location. However, in many real-world applications, physical distance and network delay are closely coupled. Therefore, incorporating geospatial information and location-aware diffusion dynamics into influence maximization remains a promising and challenging research direction.

References

- [1] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.
- [2] Wei Chen, Chi Wang, and Yajun Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1029–1038, 2010.
- [3] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925–979, 2015.
- [4] Jun Li, Chaochao Chen, Huiling Chen, and Changfei Tong. Towards context-aware social recommendation via individual trust. *Knowledge-Based Systems*, 127:58–66, 2017.
- [5] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 420–429, 2007.
- [6] Youze Tang, Xiaokui Xiao, and Yanchen Shi. Influence maximization: Near-optimal time complexity meets practical efficiency. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 75–86, 2014.
- [7] Youze Tang, Yanchen Shi, and Xiaokui Xiao. Influence maximization in near-linear time: A martingale approach. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, pages 1539–1554, 2015.
- [8] Y. Shi, J. Wu, and H. Chen. A pagerank-inspired heuristic scheme for influence maximization in social networks. *International Journal of Web Service Research*, 12(4):10–25, 2015.
- [9] X. Liu, Y. Zheng, and W. Zhang. Influence maximization with consideration of pagerank centrality. In *Proceedings of SEKE 2021*, pages 214–221, 2021. PRTH + PRDD algorithms under IC model.
- [10] Zahra Aghaei and Afsaneh Fatemi. An influence maximization algorithm based on community detection using topological features. In *Proceedings of the 11th International Conference on Computer Engineering and Knowledge (ICCCKE)*, Isfahan, Iran, 2021.
- [11] Abhishek K. Umrawal, Christopher J. Quinn, and Vaneet Aggarwal. A community-aware framework for social influence maximization. *arXiv preprint arXiv:2207.08937*, 2022.
- [12] Masoud Jalayer, Morvarid Azheian, and Mehrdad M. Kermani. A hybrid algorithm based on community detection and multi-attribute decision-making for influence maximization. *arXiv preprint arXiv:2105.09507*, 2021.
- [13] Maoguo Gong, Jianan Yan, Bo Shen, Lijia Ma, and Qing Cai. Influence maximization in social networks based on discrete particle swarm optimization. *Information Sciences*, 367:600–614, 2016.
- [14] Jie Yang and Jing Liu. Influence maximization-cost minimization in social networks based on a multiobjective discrete particle swarm optimization algorithm. *IEEE Access*, 6:2320–2329, 2017.

- [15] Abbas Salehi and Behrooz Masoumi. Katz centrality with biogeography-based optimization for influence maximization problem. *Journal of Combinatorial Optimization*, 40(1):205–226, 2020.
- [16] Sagar S. De and Satchidananda Dehuri. Multi-objective biogeography-based optimization for influence maximization-cost minimization in social networks. In *BITMDM 2019, LAIS 10*. Springer, 2020. Combines BBO and NSGA-II for multiobjective IM–cost problem.
- [17] Wei Chen, Wei Lu, and Ning Zhang. Time-critical influence maximization in social networks with time-delayed diffusion process. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
- [18] Alan Tsang, Bryan Wilder, Eric Rice, Milind Tambe, and Yair Zick. Group-fairness in influence maximization. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5997–6005, 2019.
- [19] Junaid Ali, Mahmoudreza Babaei, Abhijnan Chakraborty, Baharan Mirzasoleiman, Krishna P. Gummadi, and Adish Singla. On the fairness of time-critical influence maximization in social networks. *IEEE Transactions on Knowledge and Data Engineering*, 2021. Combines time-critical IM with group fairness.
- [20] Shay Gershtein, Tova Milo, and Brit Youngmann. Multi-objective influence maximization. In *Proceedings of the 24th International Conference on Extending Database Technology (EDBT)*, 2021. Allows balancing multiple group-specific influence objectives.
- [21] Elia Cunegatti, Leonardo Lucio Custode, and Giovanni Iacca. Many-objective evolutionary influence maximization: Balancing spread, budget, fairness, and time. *arXiv preprint arXiv:2403.18755*, 2024. Optimizes spread, budget, fairness, communities, and time objectives.
- [22] Ping Wang and Ruisheng Zhang. A multi-objective crow search algorithm for influence maximization in social networks. *Electronics*, 12(8):1790, 2023.
- [23] Ping Wang and Ruisheng Zhang. Influence maximization in social networks using an improved multi-objective particle swarm optimization. *The Computer Journal*, 67(6):2137–2150, 2024.
- [24] Matthew Staib and Stefanie Jegelka. Distributionally robust submodular maximization. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2018. Applies distributional robustness to submodular maximization; relevant to fair/robust influence maximization.
- [25] Golnoosh Farnadi, Behrouz Babaki, and Michel Gendreau. A unifying framework for fairness-aware influence maximization. In *WWW '20 Companion*, pages 714–715, 2020. Defines and jointly optimizes influence spread and fairness metrics (equality, equity, maximin, diversity) via MIP.
- [26] Aida Rahmattalabi, Shahin Jabbari, Himabindu Lakkaraju, Phebe Vayanos, Max Izenberg, Ryan Brown, Eric Rice, and Milind Tambe. Fair influence maximization: A welfare optimization approach. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence*, 2021. Proposes isoelastic welfare-based fairness to address exposure disparities across communities in influence maximization.
- [27] Zhu et al. Influence maximization based on prediction and replacement in dynamic online social networks. *Mathematics*, 10(8):1341, 2021.
- [28] S. Genetti and E. Ribaga. Influence maximization in hypergraphs using multi-objective evolutionary algorithms. In *Proceedings of [Conference]*, 2025, 2025. Multi-objective EA on hypergraphs: maximize spread, minimize seed set size.
- [29] Shixun Huang, Wenqing Lin, Zhifeng Bao, and Jiachen Sun. Influence maximization in real-world closed social networks. *arXiv preprint arXiv:2209.10286*, 2022. Addresses privacy constraints via limited friend-edge recommendations.
- [30] Guangmo Tong, Weili Wu, Shaojie Tang, and Ding-Zhu Du. Adaptive influence maximization in dynamic social networks. *arXiv preprint arXiv:1506.06294*, 2015. Introduces dynamic IC model with seed selection over time.
- [31] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.
- [32] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 855–864, 2016.
- [33] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 701–710, 2014.
- [34] Jure Leskovec and Rok Sosič. Snap: A general-purpose network analysis and graph-mining library. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(1):1–20, 2016.
- [35] Tarun K Biswas, Alireza Abbasi, and Ripon K Chakraborty. An improved clustering based multi-objective evolutionary algorithm for influence maximization under variable-length solutions. *Knowledge-Based Systems*, 256:109856, 2022.
- [36] Stefano Genetti, Eros Ribaga, Elia Cunegatti, Quintino F Lotito, and Giovanni Iacca. Influence maximization in hypergraphs using multi-objective evolutionary algorithms. In *International Conference on Parallel Problem Solving from Nature*, pages 217–235. Springer, 2024.