Transactions on Soft Computing

www.tsc.reapress.com

Trans. Soft. Comput. Vol. 1, No. 3 (2025), pp. 190–202.

Paper Type: Original Article

Submodular Influence Maximization with Hesitant Fuzzy Edge Probabilities

Sudipta Kotal[1,*], **Tarasankar Pramanik**[1]

[1]Department of Technical Sciences, Algebra Bernays University, Gradišćanska 24, 10000 Zagreb, Croatia; kosudipta@gmail.com; tarasankar.math07@gmail.com.

Citation:

Received: 05 February 2025 *Revised: 17 May 2025* *Accepted: 21 July 2025*	Kotal, S., & Pramanik, T. (2025). Submodular influence maximization with hesitant fuzzy edge probabilities. *Transactions on Soft Computing*, *1*(3), 190-202.

Abstract

Influence maximization is usually studied by assigning a single propagation probability to each directed edge. In many real-world situations, however, experts and repeated calibration provide several possible probability values instead of one fixed choice. This uncertainty is represented by hesitant fuzzy elements and is interpreted through two models: a weighted finite scenario model and an independent edge-wise hesitation model. For the weighted scenario model, the expected influence spread is proved to be normalized, monotone, and submodular because it is formed as a nonnegative combination of live-edge reachability functions. Hence, the classical greedy algorithm retains its $(1 - 1/e)$ approximation guarantee. For the independent edge-wise model, marginalization is shown to be exactly equivalent to the independent cascade model with the weighted mean probability of each hesitant element. Lower and upper hesitant envelopes provide monotone submodular spread bounds and computable certificates for every admissible edge-probability realization. A Lipschitz sensitivity bound relates changes in expected spread to variations in edge probabilities. An explicit counterexample shows that the pointwise minimum of correlated scenario spreads need not be submodular, so robust optimization requires additional structure. The proposed framework preserves hesitant uncertainty while maintaining mathematically valid approximation guarantees.

Keywords: Hesitant fuzzy set, Influence maximization, Submodular function, Independent cascade, Greedy algorithm, Robust envelope.

1 | Introduction

Influence maximization is studied to identify a small set of seed vertices that can maximize the spread of information in a directed network. This problem has been given significant

Corresponding Author: kosudipta@gmail.com

https://doi.org/10.48314/tsc.v1i3.68

Licensee System Analytics. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

attention in social networks, recommendation systems, epidemic analysis, and viral marketing, because the overall diffusion process is strongly affected by the selection of suitable seed vertices. Under the independent cascade and linear threshold models, the expected influence spread has been shown to be monotone and submodular, which allows efficient approximation to be achieved through greedy optimization [1, 9].

Several scalable algorithms have later been developed to improve the efficiency of influence estimation on large networks. The computational cost has been greatly reduced by reverse-reachable sampling and related optimization techniques, while theoretical guarantees have been preserved [3, 4]. However, a single propagation probability is assumed by most existing methods to be assigned to every directed edge, and this assumption may not be considered accurate enough for representing uncertain real-world environments.

In many soft-computing applications, propagation probabilities are not known precisely. Instead, multiple plausible values may be obtained from expert opinions, repeated observations, different time periods, or calibration procedures. A natural mathematical representation for this type of uncertainty is provided by hesitant fuzzy sets, because several membership values can be associated with the same element [5]. Distance measures, aggregation operators, and decision-making methods for hesitant fuzzy information have also been extensively investigated [6, 8]. Nevertheless, useful uncertainty information is removed when hesitant values are replaced by their arithmetic mean, whereas every possible combination becomes computationally expensive to evaluate.

A theoretical framework is developed in this paper to integrate hesitant fuzzy information with influence diffusion. Two different semantics are considered: a weighted scenario model and an independent edge-wise hesitation model. The preservation of monotonicity and submodularity is established, exact edge-wise marginalization is proved, and lower and upper endpoint networks are constructed to bound every admissible realization. Mathematically valid approximation guarantees are provided by the proposed framework, while the uncertainty represented by hesitant fuzzy edge probabilities is maintained.

1.1 | Motivation

Influence maximization models usually assume that every directed edge has one fixed propagation probability. In practice, however, this assumption is often unrealistic because social interactions are influenced by uncertain human behavior, changing environments, and incomplete observations. As a result, several plausible probability values may be obtained for the same edge instead of one reliable estimate.

Hesitant fuzzy sets provide a natural representation for this uncertainty by allowing multiple possible values to be associated with a single edge. However, most existing influence maximization methods either replace these values by their average or evaluate all possible combinations, which may lead to the loss of uncertainty information or excessive computational cost.

This work is motivated by the need for a mathematical framework that preserves hesitant uncertainty while maintaining the theoretical properties of influence diffusion. Therefore, weighted scenario modeling, independent edge-wise hesitation, and endpoint spread bounds are studied within a unified influence maximization framework.

1.2 | Preliminaries

This section introduces the basic notation and mathematical concepts used throughout the paper. These definitions provide the foundation for the hesitant fuzzy diffusion model and the theoretical results developed in the following sections.

Let $D = (V, E)$ be a finite directed graph, where V denotes the set of vertices and E denotes the set of directed edges. The total number of vertices is represented by $n = |V|$. A directed edge $(u, v) \in E$ describes a possible influence from vertex u to vertex v. Throughout this paper, the graph is assumed to be finite, and self-loops and parallel edges are not considered.

The influence maximization problem is formulated by selecting a seed set $S \subseteq V$. All vertices in S are activated at the beginning of the diffusion process, and the objective is to maximize the expected number of activated vertices after the propagation process is completed.

A hesitant fuzzy element (HFE) is defined as a non-empty finite subset of values in the interval $[0, 1]$. Each value represents a possible propagation probability associated with a directed edge. Instead of assigning one fixed probability, several plausible values are allowed to describe uncertainty.

$$H_e = \{p_{e1}, \dots, p_{er_e}\} \subseteq [0, 1].$$

If importance weights are available, they are denoted by $\omega_{eh} \geq 0$ and satisfy $\sum_h \omega_{eh} = 1$. For an unweighted hesitant fuzzy element, equal weights are assigned, that is, $\omega_{eh} = 1/r_e$. These values are interpreted as propagation probabilities only when a probabilistic diffusion model is considered; otherwise, they remain hesitant fuzzy membership grades.

A live-edge realization is obtained by retaining each directed edge independently according to its propagation probability. The resulting graph is denoted by L. For a given seed set S, the set of all vertices reachable in L is written as $R_L(S)$.

Let $p = (p_e)_{e \in E}$ be a fixed edge-probability vector. Under the independent cascade model, each directed edge is given only one opportunity to activate its neighboring vertex after its source vertex becomes active. The propagation process continues until no further activation is possible. This live-edge interpretation is mathematically equivalent to the independent cascade model [1].

The expected influence spread of a seed set is defined by

$$\sigma_p(S) = \mathbb{E}_p[\,|R_L(S)|\,].$$

Here, $\sigma_p(S)$ represents the expected number of activated vertices, while $R_L(S)$ denotes the reachable set in the live-edge realization.

For convenience, the main notation used throughout the paper is summarized in Table 1.

A set function $f : 2^V \to \mathbb{R}$ is called monotone if its value never decreases when additional vertices are included in the seed set. It is called submodular if it satisfies the diminishing-returns property, where the marginal gain of adding a new vertex becomes smaller as the seed set becomes larger [9].

Table 1. Basic notation used throughout the paper.

Symbol	Description
$D = (V, E)$	Directed graph
V	Vertex set
E	Directed edge set
S	Seed set
H_e	Hesitant fuzzy element of edge e
p_e	Propagation probability of edge e
L	Live-edge realization
$R_L(S)$	Reachable vertex set
$\sigma_p(S)$	Expected influence spread

$$f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B),$$

for $A \subseteq B$ and $v \notin B$.

Monotonicity guarantees that adding new seed vertices cannot reduce the expected influence spread, whereas submodularity explains why greedy seed selection achieves strong approximation guarantees. These properties form the mathematical basis of the theoretical framework developed in the following sections.

2 | Theoretical Framework

2.1 | Weighted Hesitant Scenarios

Let $\Omega = \{1, \ldots, M\}$ denote a finite collection of coherent scenarios. In each scenario r, a complete edge-probability vector $p^{(r)}$ is assigned to the network. The scenario weights satisfy $\theta_r \geq 0$ and $\sum_r \theta_r = 1$. The hesitant expected influence spread is defined as

$$F_\theta(S) = \sum_{r=1}^{M} \theta_r \sigma_{p^{(r)}}(S). \tag{1}$$

This formulation allows correlations between edges to be preserved. For example, one scenario may represent a high-engagement environment in which many edges simultaneously receive larger propagation probabilities.

Theorem 2.1. *The hesitant expected influence spread is normalized, monotone, and submodular. In particular, $F_\theta(\emptyset) = 0$, every seed set satisfies $F_\theta(S) \geq |S|$, and the diminishing-returns property holds for F_θ.*

Proof. For a fixed scenario, let L be a live-edge graph. The deterministic reachability function $g_L(S) = |R_L(S)|$ is monotone because adding new seed vertices cannot reduce the reachable set. The marginal gain of a vertex is equal to the number of additional vertices that become reachable, and this gain can only decrease when the seed set becomes larger. Therefore, g_L is submodular.

Since expectation preserves linear inequalities, the spread function $\sigma_{p^{(r)}}$ is also monotone and submodular. Finally, F_θ is obtained as a nonnegative weighted sum of these spread functions, so both properties are preserved. The normalization and self-activation results follow directly. □

Given a seed budget k, the greedy algorithm starts with the empty seed set $S_0 = \emptyset$. At each step, the vertex that gives the largest increase in the expected influence spread is selected. The selected vertex is added to the current seed set until k vertices have been chosen.

$$v_t \in \arg\max_{v \notin S_{t-1}} \left(F_\theta(S_{t-1} \cup \{v\}) - F_\theta(S_{t-1}) \right). \tag{2}$$

Theorem 2.2. *Let S_g be the seed set obtained by the greedy algorithm, and let S^* be the optimal seed set with $|S| \leq k$. Then the greedy solution satisfies the classical $(1 - 1/e)$ approximation guarantee.*

Proof. At each iteration, submodularity ensures that the remaining gap between the greedy solution and the optimal solution is bounded by the sum of at most k marginal gains. Therefore, the greedy algorithm captures at least $1/k$ of the remaining gap in every step. By repeatedly applying this argument, the required recurrence is obtained, which directly proves the $(1 - 1/e)$ approximation guarantee [9]. □

When influence spread is estimated by Monte Carlo simulation, an estimation error is introduced. If every estimated marginal gain differs from its true value by at most ε, uniformly over the candidates evaluated at each step, approximate greedy satisfies $F_\theta(S_g) \geq (1 - 1/e)F_\theta(S^*) - 2k\varepsilon$.

2.2 | Independent Edge-Wise Hesitation

In the independent edge-wise model, each edge selects one value from its hesitant fuzzy element independently according to the assigned weights. After the selection is completed, the independent cascade process is performed on the resulting network. The weighted mean probability of each edge is defined as

$$\bar{p}_e = \mathbb{E}[P_e] = \sum_h \omega_{eh} p_{eh}. \tag{3}$$

Theorem 2.3. *Under independent edge-wise hesitation, the unconditional live-edge indicators remain independent Bernoulli random variables with parameters $\bar{p}_e$. Consequently, the expected influence spread is exactly equal to the spread obtained from the mean probability vector.*

$$\mathbb{E}_P[\sigma_P(S)] = \sigma_{\bar{p}}(S)$$

for every seed set S.

Proof. Let X_e denote the live-edge indicator for edge e. Conditioned on P_e, the variable X_e follows a Bernoulli distribution with parameter P_e. Taking expectation gives $\Pr(X_e = 1) = \mathbb{E}[P_e] = \bar{p}_e$. Since the edge selections are independent, the joint distribution of all live-edge indicators is identical to that of independent Bernoulli variables with parameters $\bar{p}_e$. Therefore, the complete live-edge process is equivalent to the independent cascade model with the mean probability vector. □

This result shows that edge-wise averaging preserves the expected influence spread under the independence assumption. However, uncertainty, risk, and correlations between edges are not preserved by averaging. When hesitant selections are correlated, the weighted scenario model provides the correct mathematical representation.

The dispersion of a hesitant fuzzy element is defined by

$$d_e^2 = \sum_h \omega_{eh}(p_{eh} - \bar{p}_e)^2.$$

Although the dispersion does not change the expected spread, it can be used as an uncertainty measure or as part of a risk-sensitive objective. Such objectives require a separate analysis because subtracting a variance term does not automatically preserve submodularity.

2.3 | Lower and Upper Hesitant Envelopes

Let

$$p_e^- = \min H_e, qquad p_e^+ = \max H_e.$$

Every admissible deterministic selection p satisfies $p^- \leq p \leq p^+$ coordinatewise.

Theorem 2.4. *For every $S \subseteq V$ and every $p \in \prod_e H_e$,*

$$\sigma_{p^-}(S) \leq \sigma_p(S) \leq \sigma_{p^+}(S). \tag{4}$$

Both endpoint functions are monotone and submodular in S.

Proof. The proof is obtained by coupling all live-edge graphs through independent random variables $U_e \sim \text{Uniform}[0,1]$. An edge is declared live whenever $U_e \leq q_e$ for a given probability vector q. Since $p^- \leq p \leq p^+$ coordinatewise, the corresponding live-edge graphs satisfy

$$L(p^-) \subseteq L(p) \subseteq L(p^+).$$

The reachable sets preserve the same inclusion. Therefore, taking expectations directly gives the spread bounds. The monotonicity and submodularity of both endpoint functions follow from Theorem 2.1. □

For the Cartesian product uncertainty set, the common lower endpoint attains the minimum for every fixed seed set:

$$\min_{p^- \leq p \leq p^+} \sigma_p(S) = \sigma_{p^-}(S).$$

Consequently, maximizing the lower-endpoint spread is exactly the max–min problem over the product set, and greedy retains the $(1 - 1/e)$ guarantee for that robust objective. This conclusion does not extend automatically to an arbitrary correlated scenario set.

For a correlated scenario set, define $F_{\min}(S) = \min_r \sigma_{p^{(r)}}(S)$. Although each scenario spread is monotone and submodular, their pointwise minimum is not necessarily submodular. Therefore, the classical greedy approximation guarantee cannot be applied directly to a robust worst-scenario objective.

To see this within the influence model, let $V = \{a, b, x, y\}$ and consider two deterministic scenarios. In scenario 1 the only live edge is $a \to x$; in scenario 2 the only live edge is $b \to y$. Restrict attention to seed choices from $\{a, b\}$. Then

$$F_{\min}(\{a\}) = F_{\min}(\{b\}) = 1, \qquad F_{\min}(\{a, b\}) = 3, \qquad F_{\min}(\varnothing) = 0.$$

Thus $F_{\min}(\{a\}) + F_{\min}(\{b\}) = 2 < 3 = F_{\min}(\{a, b\}) + F_{\min}(\varnothing)$, which violates the two-set submodularity inequality.

2.4 | Sensitivity to Hesitant Grades

The following result provides a simple distribution-free bound for measuring the sensitivity of influence spread to changes in edge probabilities.

Theorem 2.5. *For any two probability vectors $p, q \in [0, 1]^E$ and seed set S,*

$$|\sigma_p(S) - \sigma_q(S)| \leq n \sum_{e \in E} |p_e - q_e|. \tag{5}$$

Proof. A shared-uniform coupling is used for both probability vectors. For each edge, the probability that the live-edge state differs is exactly $|p_e - q_e|$. By applying the union bound, the probability that at least one edge differs is bounded by the sum of these values. Whenever both live-edge graphs are identical, the reachable sets are also identical. Otherwise, their sizes can differ by at most n. Taking expectation gives the required Lipschitz bound. □

A tighter bound can be obtained by replacing n with the maximum number of downstream vertices influenced by each edge. However, the Lipschitz bound is useful because it immediately shows that if the total change in edge probabilities is at most ϵ, then the expected influence spread changes by at most $n\epsilon$.

2.5 | Curvature-refined Approximation

Proposition 2.6. *Let F be a normalized monotone submodular function with total curvature κ. For the cardinality-k problem and $\kappa > 0$, greedy satisfies*

$$F(S_g) \geq \frac{1}{\kappa}\left(1 - \left(1 - \frac{\kappa}{k}\right)^k\right) F(S^*) \geq \frac{1 - e^{-\kappa}}{\kappa} F(S^*).$$

The limit at $\kappa = 0$ equals one.

Proof. The proof follows from the standard greedy recurrence together with the curvature inequality, which preserves a larger fraction of the remaining marginal contribution at each iteration. By repeatedly applying this recurrence, the curvature-refined approximation bound is obtained [14]. □

Although the exact computation of total curvature is expensive for large networks, it can be estimated from endpoint or sampled diffusion models. Such estimates help determine whether the classical greedy bound is conservative for a given network.

In practical applications, different vertices may have different selection costs. Let $c_v > 0$ denote the cost of selecting vertex v, and let B be the available budget. The optimization problem is extended by maximizing the expected influence spread while keeping the total seed cost within the given budget.

$$\max\{F_\theta(S) : \sum_{v\in S} c_v \le B\}. \tag{6}$$

A cost-aware greedy strategy selects vertices according to the ratio of marginal gain to cost. However, this strategy alone may ignore an individually valuable but expensive vertex. A stronger $(1 - 1/e)$ approximation is obtained by combining partial enumeration with the greedy method under a single knapsack constraint [13]. The same guarantee also applies to the lower and upper endpoint spread functions because both remain monotone and submodular.

When costs are also represented by hesitant or interval values, the optimization should be performed using a declared upper-cost realization. Using only the average cost may produce a seed set that exceeds the actual budget in practice.

Expected spread alone may not fully describe the performance of a seed set across different coherent scenarios. Therefore, three complementary measures are considered for any seed set S:

$$\mu_S = \sum_r \theta_r \sigma_r(S), \qquad v_S = \sum_r \theta_r (\sigma_r(S) - \mu_S)^2, \quad q_S = \min_r \sigma_r(S).$$

The triplet $(\mu_S, \sqrt{v_S}, q_S)$ represents the expected spread, spread dispersion, and worst-case performance of a seed set. These measures provide a practical way to compare candidate solutions without changing the optimization objective. Thus, a reliable optimization guarantee is preserved while a broader assessment of uncertainty is also obtained.

3 | Applications

3.1 | Exact Five-Vertex Example

To demonstrate the proposed model, a directed acyclic graph with five vertices is considered. The network contains the directed edges

$$1 \to 2,\ 1 \to 3,\ 2 \to 4,\ 3 \to 4,\ 4 \to 5.$$

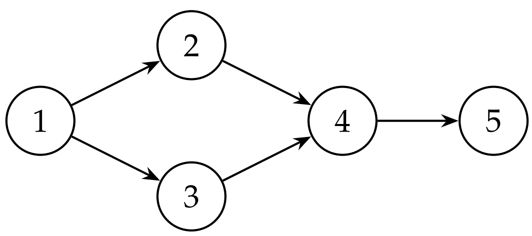

Figure 1. Five-vertex directed acyclic graph used in the exact influence spread example.

In the displayed edge order, assign the unweighted hesitant fuzzy elements

$$H_{12} = \{.4, .5, .6\}, \quad H_{13} = \{.3, .4, .5\}, \quad H_{24} = \{.5, .6, .7\},$$

$$H_{34} = \{.2, .4, .6\}, \qquad H_{45} = \{.6, .7, .8\}.$$

Under the independent edge-wise hesitation model, Theorem 2.3 gives the mean probability vector

$$\bar{p} = (0.5, 0.4, 0.6, 0.4, 0.7).$$

For the seed set $S = \{1\}$, vertices 2 and 3 are activated with probabilities 0.5 and 0.4, respectively. Vertex 4 becomes active whenever at least one of the two independent paths, $1 \to 2 \to 4$ or $1 \to 3 \to 4$, is live. Because the two paths use disjoint edge sets,

$$\Pr(4) = 1 - (1 - .5 \cdot .6)(1 - .4 \cdot .4) = .412.$$

Vertex 5 is activated only if vertex 4 is active and edge $(4, 5)$ is live. Hence, its activation probability is obtained by multiplying the activation probability of vertex 4 by 0.7. The expected influence spread will be

$$\sigma_{\bar{p}}(\{1\}) = 1 + .5 + .4 + .412 + .2884 = 2.6004. \tag{7}$$

The same network is evaluated using the lower and upper endpoint probability vectors. The resulting spreads verify that the mean-model spread always lies between the two endpoint values, confirming the spread envelope established in Theorem 2.4.

At the lower endpoint $p^- = (.4, .3, .5, .2, .6)$,

$$\Pr(4) = 1 - (1 - .20)(1 - .06) = .248, \quad \Pr(5) = .1488,$$

so $\sigma_{p^-}(\{1\}) = 2.0968$. At the upper endpoint $p^+ = (.6, .5, .7, .6, .8)$,

$$\Pr(4) = 1 - (1 - .42)(1 - .30) = .594, \quad \Pr(5) = .4752,$$

and $\sigma_{p^+}(\{1\}) = 3.1692$. Equation (4) is verified numerically:

$$2.0968 \leq 2.6004 \leq 3.1692.$$

Table 2 records exact mean-model spread for several one-vertex seeds. Because the graph is acyclic and small, each number follows by elementary path-event calculations; no simulation is required. Thus vertex 1 is the optimal single seed for the mean model.

Table 2. Mean-model spread for single seeds.

Seed	Expected spread	Calculation structure
$\{1\}$	2.6004	Eq. (7)
$\{2\}$	$1 + .6 + .6(.7) = 2.0200$	path $2 \to 4 \to 5$
$\{3\}$	$1 + .4 + .4(.7) = 1.6800$	path $3 \to 4 \to 5$
$\{4\}$	$1 + .7 = 1.7000$	edge $4 \to 5$
$\{5\}$	1.0000	no outgoing edge

A robust report should additionally compare the lower-envelope values rather than assuming that this ranking persists automatically.

3.2 | Optimization Consequences

The five-vertex example permits exact two-seed calculations. With mean probabilities, seed set $\{1,4\}$ activates vertices 1 and 4 deterministically, vertices 2 and 3 with probabilities .5 and .4, and vertex 5 with probability .7. Hence

$$\sigma_{\bar{p}}(\{1,4\}) = 2 + .5 + .4 + .7 = 3.6000.$$

The marginal contribution of vertex 4 to the empty seed set is

$$\sigma_{\bar{p}}(\{4\}) - \sigma_{\bar{p}}(\varnothing) = 1.7,$$

whereas its contribution after selecting vertex 1 is

$$\sigma_{\bar{p}}(\{1,4\}) - \sigma_{\bar{p}}(\{1\}) = .9996.$$

The reduction is a concrete instance of diminishing returns: much of vertex 4's downstream effect is already reached from vertex 1.

Other pairs can be checked in the same way. For $\{1,2\}$, vertex 3 activates with probability .4. Vertex 4 activates either directly from seeded vertex 2 with probability .6 or through the independent two-edge path $1 \to 3 \to 4$ with probability $.4(.4) = .16$. Therefore

$$\Pr(4 \mid \{1,2\}) = 1 - (1 - .6)(1 - .16) = .664,$$

and

$$\sigma_{\bar{p}}(\{1,2\}) = 2 + .4 + .664 + .664(.7) = 3.5288.$$

For $\{2,3\}$, vertex 4 is reached through either outgoing edge with probability $1 - (1 - .6)(1 - .4) = .76$, so the spread is $2 + .76 + .76(.7) = 3.292$. These values are exact and can serve as unit tests for an implementation.

Submodularity guarantees the classical $(1 - 1/e)$ approximation bound, but a stronger guarantee can be obtained when the overlap between vertex contributions is smaller. This improvement is measured by the total curvature, which represents the amount of redundancy in the influence function. A value of $\kappa = 0$ indicates a modular function, whereas values close to 1 indicate that some vertex contributions become highly redundant in the presence of other vertices.

$$\kappa = 1 - \min_{v \in V} \frac{F(V) - F(V \setminus \{v\})}{F(\{v\})}, \qquad 0 \le \kappa \le 1. \tag{8}$$

When $\kappa = 0$, the function is modular; when κ is near one, some vertex contribution can become almost redundant in the presence of all others.

3.3 | Computational Considerations

The exact evaluation of $\sigma_p(S)$ is generally #P-hard, and influence maximization is NP-hard [1]. Monte Carlo simulation is commonly used to estimate the expected influence spread, while the weighted scenario model increases the computational cost because each scenario must be evaluated separately. Reverse-reachable sampling provides an efficient alternative by sampling a scenario first and then generating the corresponding live-edge graph. Under independent edge-wise hesitation, Theorem 2.3 removes the need to sample hesitant fuzzy elements, since the expected spread is computed directly from the mean probability vector.

The efficiency of the greedy algorithm can be improved through lazy marginal evaluation, where previously computed marginal gains remain valid upper bounds because of submodularity. The CELF++ implementation reduces the running time without changing either the optimization objective or its approximation guarantee [15].

For a threshold stopping rule, let $\hat{\Delta}(v \mid S)$ denote the estimated marginal gain of vertex v. Hoeffding's inequality is used to control the estimation error because the cascade size is bounded in $[0, n]$ [12]. For reproducible experiments, the number of Monte Carlo simulations, confidence level, random seed, and stopping rule should always be reported.

4 | Results and Interpretation

The analytical results confirm that hesitant fuzzy uncertainty can be incorporated into influence maximization without losing the fundamental properties of graph diffusion. Both the weighted scenario model and the independent edge-wise hesitation model preserve monotonicity and submodularity under their respective assumptions, allowing the greedy algorithm to maintain its theoretical approximation guarantee.

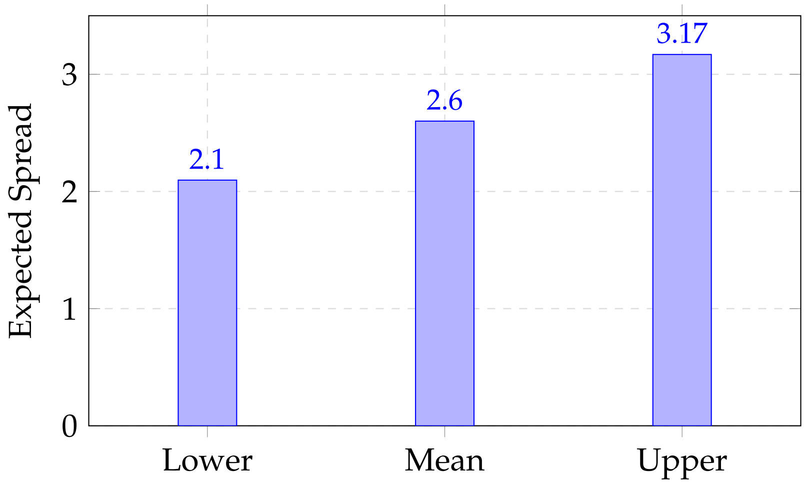

Figure 2. Comparison of lower, mean, and upper expected influence spread for the seed set $S = \{1\}$.

Figure 2 compares the expected influence spread obtained from the lower, mean, and upper probability models. The mean-model spread lies between the two endpoint values, directly illustrating the spread envelope in Theorem 2.4. The increase from the lower to the upper model also shows the effect of uncertain edge probabilities on influence diffusion.

5 | Conclusion

The proposed framework demonstrates that hesitant fuzzy edge probabilities can be incorporated into influence maximization without losing its fundamental mathematical properties. Two different interpretations of hesitant uncertainty were studied: the weighted scenario model and the independent edge-wise hesitation model. Both models preserve valid diffusion semantics under their respective assumptions.

Theoretical results show that the expected influence spread remains monotone and submodular for weighted scenarios, allowing the classical greedy algorithm to retain its($1-1/e$ approximation guarantee. For independent edge-wise hesitation, exact marginalization proves that the mean probability vector produces the same expected spread as the original hesitant model.*Proceedings*

The five-vertex analytical example verifies the proposed theory through exact path-event calculations and confirms the lower, mean, and upper spread bounds. In addition, optimization and computational analyses illustrate the practical applicability of the framework for uncertainty-aware influence diffusion.

Future work may extend this model to dynamic social networks, competitive diffusion processes, and data-driven estimation of hesitant edge probabilities from real-world network observations.

Author Contributions

All authors contributed equally.

Funding

The funding statement must be confirmed by all listed authors before publication.

Data Availability

No external dataset was used. The complete synthetic graph and hesitant elements are included in the text.

Competing Interests

The competing-interests statement must be confirmed by all listed authors before publication.

Generative-AI Assistance

Generative AI assisted with language, LaTeX preparation and representation of the paper.

References

[1] Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 137-146). ACM Digital Library. https://doi.org/10.1145/956750.956769

[2] Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., & Glance, N. (2007). Cost-effective outbreak detection in networks. *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 420-429). ACM Digital Library. https://doi.org/10.1145/1281192.1281239

[3] Borgs, C., Brautbar, M., Chayes, J., & Lucier, B. (2014). Maximizing social influence in nearly optimal time. *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms* (pp. 946-957). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611973402.70

[4] Tang, Y., Shi, Y., & Xiao, X. (2015). Influence maximization in near-linear time: A martingale approach. *Proceedings of the 2015 ACM SIGMOD international conference on management of data* (pp. 1539-1554). ACM Digital Library. https://doi.org/10.1145/2723372.2723734

[5] Torra, V. (2010). Hesitant fuzzy sets. *International journal of intelligent systems, 25*(6), 529-539. https://doi.org/10.1002/int.20418

[6] Xu, Z., & Xia, M. (2011). Distance and similarity measures for hesitant fuzzy sets. *Information sciences, 181*(11), 2128-2138. https://doi.org/10.1016/j.ins.2011.01.028

[7] Xia, M., & Xu, Z. (2011). Hesitant fuzzy information aggregation in decision making. *International journal of approximate reasoning, 52*(3), 395-407. https://doi.org/10.1016/j.ijar.2010.09.002

[8] Liao, H., Xu, Z., & Xia, M. (2014). Multiplicative consistency of hesitant fuzzy preference relation and its application in group decision making. *International journal of information technology & decision making, 13*(01), 47-76. https://doi.org/10.1142/S0219622014500035

[9] Nemhauser, G. L., Wolsey, L. A., & Fisher, M. L. (1978). An analysis of approximations for maximizing submodular set functions—I. *Mathematical programming, 14*(1), 265-294. https://doi.org/10.1007/BF01588971

[10] Fujishige, S. (2005). *Submodular functions and optimization* (Vol. 58). Elsevier. https://books.google.com/books?id=gdcRXdoV89QC&dq=Fujishige,+S.+(2005)

[11] Krause, A., Singh, A., & Guestrin, C. (2008). Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of machine learning research, 9*(2), 235-284. https://jmlr.org/papers/v9/krause08a.html

[12] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American statistical association, 58*(301), 13-30. https://doi.org/10.1080/01621459.1963.10500830

[13] Sviridenko, M. (2004). A note on maximizing a submodular set function subject to a knapsack constraint. *Operations research letters, 32*(1), 41-43. https://doi.org/10.1016/S0167-6377(03)00062-2

[14] Conforti, M., & Cornuéjols, G. (1984). Submodular set functions, matroids and the greedy algorithm: Tight worst-case bounds and some generalizations of the Rado-Edmonds theorem. *Discrete applied mathematics, 7*(3), 251-274. https://doi.org/10.1016/0166-218X(84)90003-9

[15] Goyal, A., Lu, W., & Lakshmanan, L. V. (2011). Celf++ optimizing the greedy algorithm for influence maximization in social networks. *Proceedings of the 20th international conference companion on World wide web* (pp. 47-48). ACM Digital Library. https://doi.org/10.1145/1963192.1963217