Network Interventions to Limit Misinformation Spread on Reddit

Alexandre St-Aubin, Benjamin Movassagh, and Alexandrine Fortier

Abstract— **Social media has become the primary news source in the United States, yet user awareness alone cannot prevent misinformation spread. This study investigates network-based interventions to limit misinformation propagation on Reddit using the FACTOID dataset of users and their interactions. We evaluate seed selection strategies using centrality metrics and implement a competitive cascade model where factual information competes against misinformation. Our key finding addresses the superspreader dilemma: while removing high-influence accounts is topologically optimal, it incurs substantial social costs. We compare two intervention strategies under a super-linear cost model: removing few influential users versus removing many smaller users. Results show that (1) out-degree centrality outperforms other metrics for seed selection, (2) misinformation spreads more efficiently than factual content, and (3) removing many smaller users achieves greater impact than removing few superspreaders under the same budget constraint. These findings suggest that targeting numerous low-influence accounts provides a more cost-effective approach to neutralizing misinformation spread than banning high-profile users.**

I. Introduction

For the first time, social media ranked as the leading news source in the United States in 2025 [1]. Social media provides access to an abundance of varied information, quickly. Although users are aware of the danger of being exposed to misinformation, awareness alone is not enough to prevent it [2]. In fact, users tend to consume and believe news sources that are politically aligned with their beliefs, making it difficult to guarantee access to unbiased information. Because user awareness is not sufficient to counter misinformation, external interventions are needed.

Prior work on network interventions for misinformation has explored counter-seeding strategies where factual information competes with misinformation [3], node immunization approaches that protect specific users from influence [4], and probabilistic edge dropout that randomly removes connections based on user attributes [5]. However, these approaches do not address the strategic removal of nodes. We investigate the effect of removing nodes (e.g., banning users from the platform) on misinformation propagation in the network. Specifically, we address the superspreader [6] dilemma: while removing high-influence accounts (e.g., ”influencers”) is topologically optimal, it incurs substantial social and political costs, including public backlash, accusations of censorship, and concerns about silencing legitimate voices.

We address this gap by comparing two intervention strategies: head removal, which targets few high-influence users, and swarm removal, which targets many smaller users to achieve similar mitigation at lower cost. We model the cost of removing a user as a super-linear function that reflects the increasing social and political difficulty. Using the FACTOID dataset [7] with 4,150 labeled Reddit users and 3.4M interactions, we evaluate these strategies through a competitive cascade framework [3], running factual versus misinformation campaigns.

Our contributions are:

1) We compare four centrality metrics for seed selection and show out-degree consistently outperforms in-degree, PageRank, and degree.
2) We confirm misinformation spreads more efficiently than factual information, achieving equilibrium with fewer seeds.
3) We demonstrate that removing many smaller users (swarm removal via Knapsack optimization) has greater impact than removing few superspreaders (head removal via greedy selection) under the same budget constraint and super-linear cost model.
4) We show that strategic node removal acts as an effective complementary mechanism alongside factual counter-campaigns to neutralize misinformation spread.

II. Background

A. Independent Cascade Model

The Independent Cascade (IC) model was first introduced in [8] as a way to model the influence maximization problem: how to maximize influence while minimizing the initial set of nodes targeted in a social network. More precisely, the IC model is a probabilistic diffusion model where each newly-activated node gets one chance to activate each of its neighbors with some probability, and the process cascades through the network until no further activations occur. It has since been widely used when trying to solve influence maximization problems, in various types of networks and optimizations [9]–[13]. A notable extension of the IC model is the *competitive* cascade model [3]. The competitive aspect allows to compare the propagation of multiple independent campaigns in a single network. In the context of misinformation, this can be framed as a factual campaign competing against a non-factual campaign, with the goal of maximizing the reach of factual information.

B. Centrality Metrics for Seed Selection

In influence propagation models, the choice of initial seed nodes critically determines the final cascade size. Centrality metrics provide a computationally efficient way to select influential nodes in a network by quantifying their structural importance. While optimal seed selection is NP-hard [8], centrality-based heuristics offer practical alternatives that can approach near-optimal performance in many network topologies [3].

We consider four centrality metrics commonly used in influence maximization:

- Out-degree centrality: The number of outgoing edges from a node. In a directed network, this directly measures a node's potential to immediately influence its neighbors in the next step of the diffusion process.
- In-degree centrality: The number of incoming edges to a node. This captures a node's popularity or prestige but does not directly measure its broadcasting capability.
- Degree centrality: In an undirected view of the network, this is simply the total number of connections a node has, treating all edges as bidirectional.
- PageRank: Originally developed for web page ranking [14], PageRank computes a node's importance based on both the number and quality of its incoming connections. A node has high PageRank if it is connected to other high-PageRank nodes, capturing recursive notions of influence.

Different centrality metrics capture different aspects of network structure. Out-degree prioritizes nodes with high broadcasting potential, while PageRank identifies nodes embedded in influential neighborhoods. The choice of metric can significantly impact the final cascade size, as shown in our experiments (subsection V-A).

C. FACTOID Dataset

FACTOID [7] is a user-level dataset that contains 4,150 users and their 3.4M Reddit posts and replies, covering the time period from January 2020 to April 2021. Each user is labeled as either a misinformation spreader or factual based on the news sources they shared, with factuality ratings determined using *mediabiasfactcheck.com* [15]. Of the 4,150 users, 1,086 are labeled as misinformation spreaders. The dataset also provides political bias scores, factuality levels, science belief scores, and satire degrees, although these are not used in this work. We construct a graph where users are nodes and edges are directed replies between users, as detailed in section III.

III. FACTOID Graph Construction

In order to model how misinformation spreads, we construct a directed graph $G=(V,E)$ from the FACTOID dataset, where each node $u \in V$ represents a unique user. A directed edge $(u,v) \in E$ is established if user v has replied to a post or comment made by user u, implying a flow of information or attention from u to v. The resulting graph provides a structural representation of user interactions that underlie the diffusion of information.

To better understand propagation in the graph, **we make the assumption that users labeled as misinformation spreaders will spread misinformation, and that users labeled as factual will spread factual information**. This allows us to look at global patterns, as knowing exactly where misinformation is spread/adopted would require temporal analysis. We observe information propagation (factual or not) through user interactions, where post replies serve as the propagation mechanism.

A. Defining Edge Probabilities

One way in which we could simply define the probabilities between edges is by using the interaction frequency of the users. Since the probabilities of the incoming edges to a node must sum to 1, we can weight each edge by the number of times two users interacted, and then normalize to get the probability of spread. A critical limitation of this method is its agnosticism toward the semantic content of users' interactions. It does not consider that a user could reply to another with varying levels of disagreement. This fails to account for the resistance or susceptibility of a user to influence based on political stance.

To address this, we deploy a user-embedding approach based on [16]. We generate feature vectors z_u using the *all-MiniLM-L6-v2* sentence embedding model from the Sentence-Transformer library [17], based on MiniLM [18]. For each user, we embed each of their posts using *all-MiniLM-L6-v2*, and then use mean pooling to get a global representation of the user. The edge probability is then calculated based on the cosine similarity between users, under the assumption that users are more likely to be influenced by those who are more similar to them.

IV. Experiments

A. Seed Nodes Selection

We investigate which centrality metric yields the most effective seed sets for viral propagation in our specific network. We compare four standard metrics: out-degree, in-degree, PageRank, and undirected degree. For each metric, we select the top-k nodes as seed sets for both Misinformation and Factual campaigns, varying k. We then measure the influence spread (the total number of infected nodes) by simulating a standard (single-campaign) IC model.

HA1. We hypothesize that out-degree will outperform other metrics for maximizing influence spread. Unlike in-degree or degree, which may select sinks or hubs with limited broadcasting capability, out-degree directly measures a node's potential to infect its immediate neighbors. While PageRank captures global structural importance, we expect the local transmission probability in the IC model to favor nodes with high immediate fan-out (out-degree) over those with high recursive authority.

B. Competitive Cascade Model

We investigate the minimum resources required for a counter-campaign to neutralize a dominant viral propagation. We extend the standard IC model to a competitive cascade setting [3]. In this scenario, we model the network as a competition between two opposing cascades: **Fact** (C_F) and **Misinformation** (C_M). Nodes can adopt one of two states

	Vary Factual Seeds	Vary Misinfo Seeds
Misinformation gets priority	**Scenario I** *How many factual seeds are needed to match misinformation?*	**Scenario II** *How cheaply can misinformation overwhelm a fixed defense?*
Factuality gets priority	**Scenario IV** *How efficient is a high-priority factual campaign?*	**Scenario III** *How resilient is factual information with priority?*

TABLE I: Experimental scenarios defined by seed allocation and tie-breaking priority.

(Believer of Fact or Believer of Misinformation) and, once activated by a cascade, can not change their state or belief.

The simulation proceeds in discrete time steps. At each step t, newly activated nodes attempt to activate their neighbors according to the edge probabilities defined in the previous section. For simplicity, we assume that the content from both seed sets travel at the same rate. This symmetry assumption allows our framework to apply not only to settings of misinformation spread, but more broadly to competitive diffusion and campaigning scenarios. Under this propagation model, a node may receive simultaneous activation attempts from neighbors associated with different seed types. To resolve such conflicts, we explicitly define a priority-based tie-breaking rule that determines which seed type successfully activates the node when these attempts occur at the same time step. Rather than fixing this rule for all experiments, we vary which seed type wins in the case of a tie and measure the different propagation dynamics. This priority mechanism attempts to capture asymmetries in attention or trust that affects all users.

Let A denote the *dominant* campaign, defined as having priority over campaign B in our competitive IC simulation (i.e., if both campaigns activate a node v simultaneously, v adopts A). We simulate four distinct competitive scenarios (outlined in Table I) to isolate the effects of propagation priority versus seed set magnitude:

I. ($A=$ Misinfo, $B=$ Factual). Misinformation has priority. We fix $|S_M| = 500$ and vary $|S_F|$ to determine the factual seed size required to match the misinformation reach.

II. ($A=$ Misinfo, $B=$ Factual). Misinformation has priority. We fix $|S_F| = 500$ and vary $|S_M|$ to observe the seed magnitude required for misinformation to dominate a fixed factual defense.

III. ($A=$ Factual, $B=$ Misinfo). Factual information has priority. We fix $|S_F| = 500$ and vary $|S_M|$ to find the misinformation seed size required to breach the factual defense.

IV. ($A=$ Factual, $B=$ Misinfo). Factual information has priority. We fix $|S_M| = 500$ and vary $|S_F|$ to quantify the efficiency of a high-priority factual campaign.

We make the following hypotheses,

HB1. We hypothesize that propagation priority is a stronger determinant of cascade size than seed set magnitude. Specifically, we expect that when Misinformation has priority (Scenario I), the Factual campaign will require a super-linear increase in seed nodes ($|S_F| \gg |S_M|$) to achieve equilibrium. Conversely, when Factual information is prioritized (Scenario III), it will suppress Misinformation with a significantly smaller seed budget ($|S_F| \approx |S_M|$).

HB2. We hypothesize that, independent of propagation priority, Misinformation possesses an intrinsic spreading advantage due to the network topology (e.g., community structure or embedding similarity). Let $|S^{\text{eq}}|_{X,k}$ denote the *equilibrium seed size*, i.e. the size of the variable seed set for campaign X in Scenario k required to generate a cascade size equal to that of the fixed adversary campaign.

We predict that when Misinformation is the dominant campaign (resp. disadvantaged), it requires fewer resources to overwhelm a fixed Factual defense (Scenario II, resp. Scenario III) than dominant (resp. disadvantaged) Factual information requires to overwhelm a fixed Misinformation campaign (Scenario IV, resp. scenario I). Formally, assuming the fixed adversary size is identical in both scenarios ($|S_{F,\text{fixed}}| = |S_{M,\text{fixed}}|$), we hypothesize:

$$|S^{\text{eq}}|_{M,\text{II}} < |S^{\text{eq}}|_{F,\text{IV}}$$

This would indicate that Misinformation spreads more efficiently per unit of seed budget than Factual information, even when propagation priority is held constant.

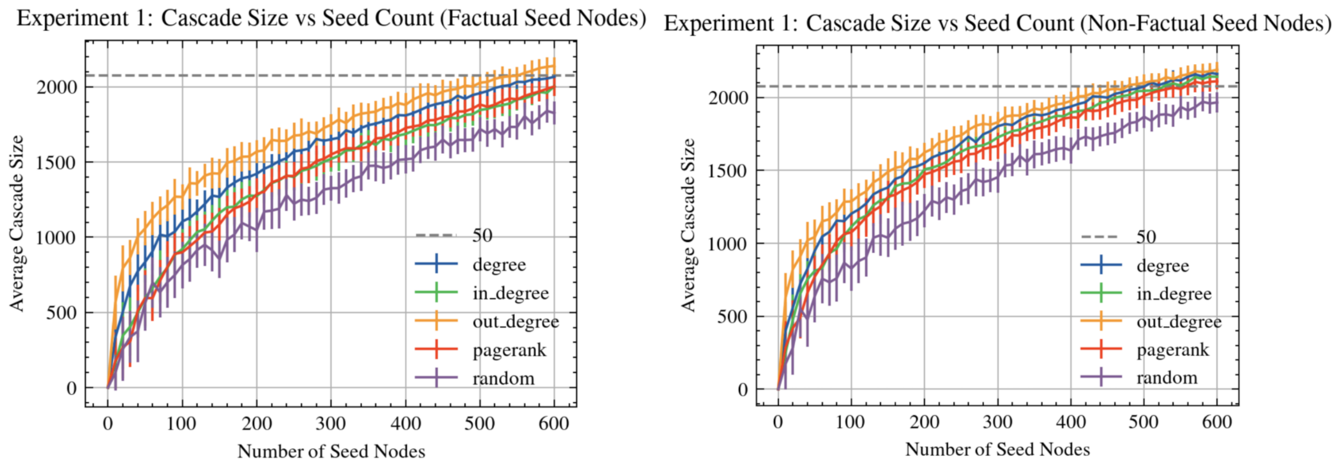
C. Cost-Aware Node Removal

Conventional immunization strategies typically prioritize *Superspreaders* [6] nodes with the highest centrality. However, in practice, interventions against high-profile accounts incur disproportionate social costs, such as public backlash and accusations of censorship. On the other hand, removing smaller, less visible accounts may be less controversial but necessitates a larger volume of interventions to achieve a comparable reduction in network connectivity.

We propose an empirical analysis to compare the cost-efficiency of a *swarm removal* strategy (targeting many low-influence users) against a traditional *head removal* strategy (targeting few high-influence users) within a competitive IC setting. For both strategies, we restrict seed removal to known misinformation spreaders, and users not already in the Misinformation seed set. We also fix the dominant campaign to Misinformation, meaning that the Factual campaign is disadvantaged.

Definition 1 (Removal Cost)**.** We define the cost $C(u)$ of removing a user u as a super-linear function of their network influence (degree centrality), reflecting the increasing difficulty or visibility of the intervention:

$$C(u) = \lfloor (\deg(u))^{\alpha} \rfloor \tag{1}$$



(a) Number of infected nodes as a function of the seed set size for the Factual campaign

(b) Number of infected nodes as a function of the seed set size for the Misinformation campaign.

Fig. 1: Factual and misinformation cascade size evolution for different seed selection strategies. Out-degree and degree metrics outperform the others.

where $\alpha \geq 1$ is a penalty parameter. A higher α disproportionately penalizes the removal of superspreaders. In our experiments, we set $\alpha = 1.2$.

We evaluate two distinct allocation strategies given a fixed intervention budget B,

1) Strategy I: Head Removal (Greedy by Degree): This strategy simulates the standard approach of targeting the most prominent actors. At each iteration, the algorithm selects the unremoved node with the highest removal cost (and thus highest degree) that fits within the remaining budget. This results in a sparse set of removed nodes, consisting exclusively of high-degree users.

2) Strategy II: Swarm Removal (Knapsack Optimization): We formulate the node removal problem as a 0/1 Knapsack Problem. Let N be the set of nodes, where each node i has a weight $w_i = C(i)$ (cost) and a value $v_i = \deg(i)$ (influence). The objective is to maximize the total influence removed:

$$\begin{aligned} \text{maximize} \ & \sum_{i \in N} v_i x_i \\ \text{subject to} \ & \sum_{i \in N} w_i x_i \leq B, \quad x_i \in \{0,1\} \end{aligned}$$

Since the cost function $C(u)$ is convex with respect to degree (for $\alpha > 1$), a single high-degree node is significantly more expensive per unit of influence than multiple low-degree nodes. Consequently, the optimal solution to this Knapsack formulation naturally favors a *swarm* of smaller nodes to maximize the aggregate degree removed. We solve this using dynamic programming.

HC1. We hypothesize that while removing superspreaders is topologically optimal in an unconstrained setting, it is *economically* suboptimal under a super-linear cost regime ($\alpha > 1$). Specifically, we expect the Swarm Removal (Knapsack) strategy to achieve greater viral suppression than Head Removal (Greedy) for the same total budget.

HC2. We hypothesize that the removal of known misinformation spreaders from the graph with *strategy II* will negatively affect the cascade size of the misinformation seeds more than that of the factual seeds in the Competitive IC simulation. This would confirm that removing users from the social network is a viable approach for suppressing the spread of misinformation.

D. Evaluation Metrics

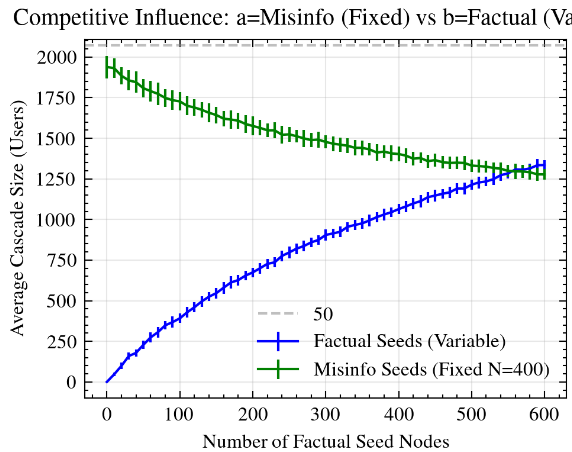
To quantify the efficacy of our seeding and intervention strategies, we employ the **Average Cascade Size** as our primary metric. Due to the probabilistic nature of the IC model, a single diffusion trajectory is insufficient to characterize the influence of a seed set.

Definition 2 (Cascade Size)**.** Let $G = (V, E)$ be the network graph and $S \subseteq V$ be the set of initial seed nodes. The cascade size, denoted as $\sigma(S)$, is defined as the cardinality of the set of all active nodes at the end of the diffusion process,

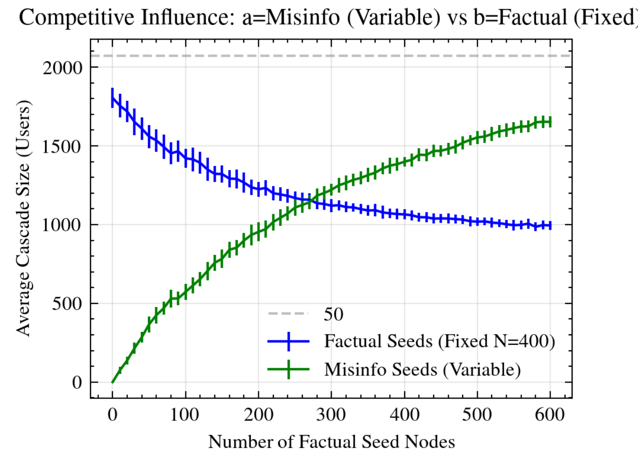
$$\sigma(S) = |\{v \in V \mid v \text{ is activated}\}| \tag{2}$$

This includes both the initial seed set S and all nodes subsequently activated via propagation. The diffusion process terminates once no nodes successfully influence any of their neighbours.

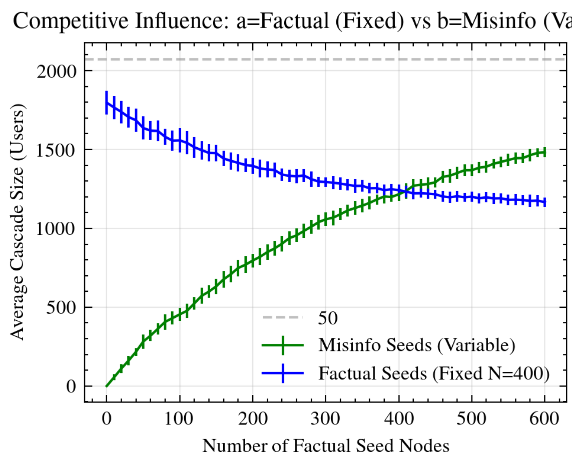
Since calculating the exact expected influence spread is known to be #P-hard, we approximate it using Monte Carlo simulations. For every unique experimental configuration (e.g., specific seed set size or removal budget), we execute $M = 100$ independent simulations of the diffusion process. We report the **Average Cascade Size** across these M iterations, along with the standard deviation (represented as error bars in our plots) to account for variance.



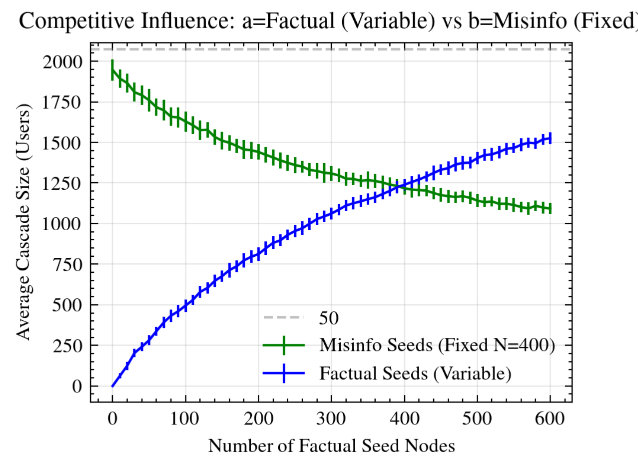
(a) **Scenario I.**



(b) **Scenario II.**



(c) **Scenario III.**



(d) **Scenario IV.**

Fig. 2: Comparison of the four scenarios introduced in section IV-B. The gray line labeled 50 denotes 50% of the network nodes.

V. Results

In this section, we present the empirical findings from our simulations. We evaluate the efficacy of centrality-based seeding, quantify the impact of propagation priority, and analyze the cost-efficiency of node removal strategies.

A. Seed Nodes Selection

We first evaluate which centrality metric produces the most influential seed sets (HA1). Figure 1 illustrates the total number of infected nodes as a function of seed set size k for both Misinformation and Factual campaigns for each seed selection method.

Consistent with **HA1**, the out_degree metric consistently yields the largest cascade sizes for a given k. This validates our hypothesis that direct broadcasting capability (measured by outgoing edges) is more predictive of influence than global prestige (PageRank) or popularity (in-degree) in the IC framework. The out-degree metric also shows the tightest average cascade size difference between the two campaigns. While degree (undirected) performs competitively, it falls short of out-degree, likely because it includes edges that cannot propagate influence (incoming edges). PageRank performs poorly in this specific topology, even worse than in-degree, suggesting that global structural importance does not necessarily translate to local transmission power in the IC framework.

To assess the robustness of our findings, we computed pairwise overlap coefficients between seed selection methods (Figure 3). The degree-based seed set shares at least 48% of its nodes with alternative selection methods, indicating that our main results would hold even with different seed selection strategies. This overlap indicates that the degree-based seed set has substantial agreement with alternative selection methods, suggesting that our results are relatively robust to the specific choice of seed selection strategy.

B. Competitive Cascade Model

Through the competitive cascade model, we analyze the interplay between propagation priority and seed magnitude. In these experiments, we set the fixed seed set size to 400. All

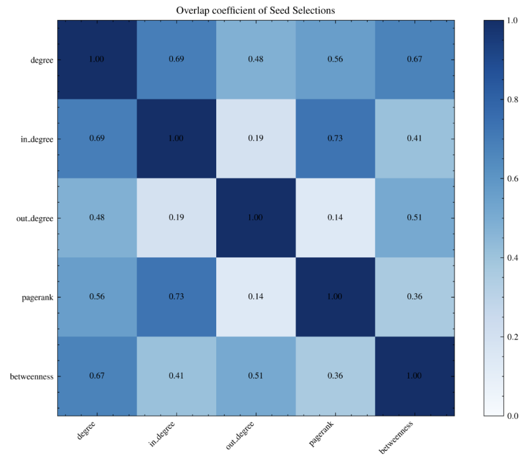


Fig. 3: Overlap coefficient matrix for seed selection methods.

seed sets were selected using the degree metric. The reason for this is primarily based on the empirical results from our seed selection experiments, where the degree metric had the highest average coefficient overlap with all other methods.

1) Propagation Priority (HB1): We compare Scenario I (Misinfo Priority) and Scenario IV (Factual Priority) to test **HB1**.

- Scenario I: When Misinformation has priority, the Factual campaign struggles to achieve equilibrium. As shown in Figure 2a, achieving equilibrium in the number of infected nodes of both campaigns requires a seed set size $550 \approx |S_F| < |S_M| = 400$, validating the hypothesis that overcoming priority requires super-linear resource investment.
- Scenario IV: Conversely, when Factual information is prioritized (Figure 2d), it suppresses Misinformation effectively with nearly equal resources ($|S_F| \approx |S_M|$). This asymmetry confirms that the choice of the structural IC model used makes a great difference in the outcome of experiments.

2) Misinformation Spreading (HB2): To isolate the intrinsic spreading power of Misinformation, we compare the seed resources required to reach equilibrium in symmetric scenarios. Comparing the slopes in Figures 2a and 2c, we observe a distinct asymmetry: Misinformation is significantly more efficient at overcoming a structural disadvantage.

Specifically, in Scenario III (where Factual has priority), Misinformation achieves parity with the dominant Factual campaign using a roughly equal seed count ($|S_M| \approx |S_F|$). This is remarkable given that Misinformation is disadvantaged by the priority rule. In contrast, in Scenario I (where Misinformation has priority), the Factual campaign requires a substantial seed surplus to catch up.

We calculated the equilibrium seed size $|S^{\text{eq}}|$ required to neutralize a fixed adversary of size $N = 400$. We found that $|S^{\text{eq}}|_{M,\text{III}} \approx 400$, whereas $|S^{\text{eq}}|_{F,\text{I}} \approx 550$. The inequality $|S^{\text{eq}}|_{M,\text{III}} < |S^{\text{eq}}|_{F,\text{I}}$ confirms that Misinformation possesses a latent topological advantage allowing it to spread more efficiently than factual content, even when the propagation rules explicitly favor the truth. Thus, we accept **HB2**.

C. Cost-Aware Node Removal

Finally, we evaluate the economic efficiency of node removal strategies under a super-linear cost constraint ($\alpha = 1.2$). In these experiments, we fix the initial seed sizes for both campaigns ($|S_F|$ and $|S_M|$) and vary the removal budget available to the intervention algorithm. Figure 4 illustrates the decline in cascade sizes as the budget increases.

1) Efficiency of Swarm Removal (HC1): We observe that the Swarm Removal (Knapsack) strategy yields a significantly steeper reduction in cascade size compared to the Head Removal (Greedy) strategy. This validates **HC1**: under a convex cost regime, removing many mid-tier nodes is more effective than removing a handful of expensive superspreaders. The Greedy approach quickly exhausts its budget on a few hubs, yielding diminishing returns, whereas the Knapsack optimizer maximizes the total influence removed.

2) Vulnerability of Misinformation (HC2): We also find that Misinformation campaigns are particularly susceptible to this cost-optimized intervention. As shown in Figure 4b (with $|S_F| = 400, |S_M| = 300$), the Knapsack strategy successfully *inverts dominance*, reducing Misinformation reach below that of the Factual campaign, with a budget of only 50,000 (approx. 10% of nodes removed). In contrast, the Greedy strategy requires a four-fold budget increase ($B \approx 200,000$) to achieve the same tipping point.

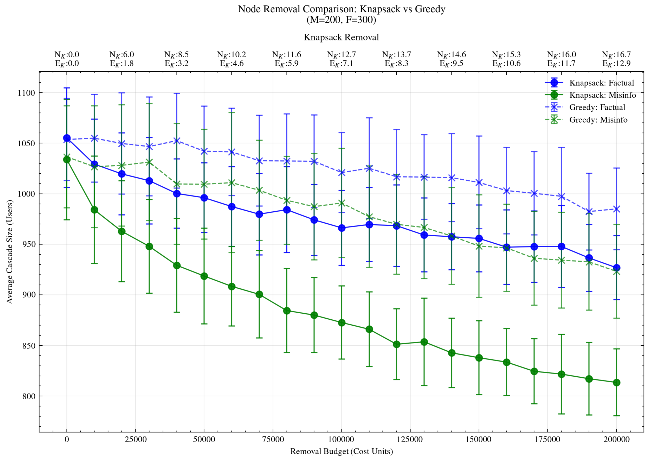
This confirms **HC2**, showing that the removal of identified misinformation spreaders suppresses viral misinformation more than the spread of factual information. This suggests that swarm-based moderation is a useful technique for neutralizing disinformation.

In summary, our results demonstrate three key findings: (1) out-degree centrality provides the most effective seed selection heuristic, (2) misinformation possesses an intrinsic spreading advantage requiring defenders to over-invest in counter-campaigns, and (3) swarm-based removal strategies significantly outperform traditional superspreader targeting under realistic cost constraints.

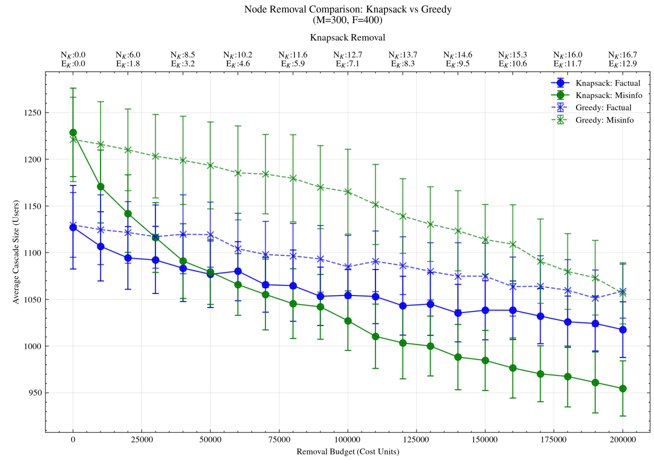
VI. Discussion

The competitive cascade experiments highlight a structural asymmetry between misinformation and factual content. Even when propagation priority favors factual information, misinformation achieves comparable reach with fewer or equal seeds. This suggests that misinformation benefits from latent network properties—such as homophily, clustering, or higher embedding similarity among misinformation spreaders—that amplify diffusion efficiency. From a policy perspective, this implies that purely reactive counter-campaigns are insufficient: defenders must over-invest resources simply to reach parity

A central contribution of this paper is reframing the cost of removal of superspreaders. While classical network immunization theory favors removing high-centrality nodes, our results show that this strategy becomes economically



(a) $|S_M| = 200$, $|S_F| = 300$.



(b) $|S_M| = 300$, $|S_F| = 400$.

Fig. 4: Knapsack (swarm) vs Greedy (head) removal strategies.

suboptimal once super-linear intervention costs are considered. High-influence users are disproportionately expensive to remove due to visibility, backlash, and governance concerns, a reality often ignored in theoretical models.

Under this cost regime, removal of many smaller nodes becomes a more effective and robust alternative. By distributing interventions across many low- and mid-degree misinformation spreaders, the knapsack-based strategy removes more aggregate influence per unit cost, leading to a steeper decline in misinformation cascade size. This suggests that low-visibility moderation (such as coordinated removal of accounts) may be more impactful than bans of prominent figures.

We emphasize that our results do not suggest that high-influence accounts are unimportant for misinformation dynamics. Rather, they show that when intervention costs scale super-linearly with visibility or influence, targeting such accounts may yield diminishing returns compared to coordinated low-profile moderation.

A. Limitations

Several limitations should be noted:

1) Our diffusion model assumes static user labels and does not incorporate temporal dynamics or belief revision; users cannot switch states once activated
2) edge probabilities are derived from embedding similarity rather than direct measurements of persuasion or disagreement, which may oversimplify complex conversational dynamics
3) Edge probabilities are mean pooled over the course of a users history, which implicitly gives a uniform weighting to all of their comments. This does not account for potential changes in views over time that a user may have
4) Interactions are the only edges that are modeled in the graph, however users may only read a post and still be influenced by the content.
5) Our cost model, while motivated by real-world considerations, remains an abstraction; actual social costs may depend on factors beyond degree centrality, such as identity or platform norms.

VII. Conclusion

This work shows that network-based interventions can effectively limit misinformation spread on Reddit. We demonstrate that out-degree centrality outperforms other metrics for seed selection, and that misinformation spreads more efficiently than factual content due to network topology. Our key finding addresses the superspreader dilemma: under super-linear cost constraints, removing many smaller users substantially outperforms removing few influential accounts. Combined with factual counter-campaigns, swarm-based removal offers a cost-effective approach to neutralizing misinformation spread on social media platforms. These results provide a first step toward realistic intervention strategies such as account bans or shadowbanning.

References

[1] N. Newman, "Digital news report 2025," Reuters Institute for the Study of Journalism, Oxford, UK, Tech. Rep., June 2025, accessed: 2025-12-11. [Online]. Available: https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025

[2] E. A. Harris, S. L. DeMora, and D. Albarracín, "The consequences of misinformation concern on media consumption," *Harvard Kennedy School (HKS) Misinformation Review*, June 2024. [Online]. Available: https://misinforeview.hks.harvard.edu/article/the-consequences-of-misinformation-concern-on-media-consumption/

[3] C. Budak, D. Agrawal, and A. El Abbadi, "Limiting the spread of misinformation in social networks," in *Proceedings of the 20th International Conference on World Wide Web*. ACM, 2011, pp. 665–674.

[4] M. A. Manouchehri, M. Helfroush, and H. Danyali, "A theoretically guaranteed approach to efficiently block the influence of misinformation in social networks," *IEEE Transactions on Computational Social Systems*, vol. PP, pp. 1–12, 03 2021.

[5] Y. E. Bayiz and U. Topcu, "Countering misinformation on social networks using graph alterations," 2022. [Online]. Available: https://arxiv.org/abs/2211.04617

[6] D. Zhang, Y. Wang, and Z. Zhang, "Identifying and quantifying potential super-spreaders in social networks," *Scientific Reports*, vol. 9, no. 1, p. 14811, 2019.

[7] F. Sakketou, J. Plepi, R. Cervero, H. J. Geiss, P. Rosso, and L. Flek, "FACTOID: A new dataset for identifying misinformation spreaders and political bias," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, and S. Piperidis, Eds. Marseille, France: European Language Resources Association, June 2022, pp. 3231–3241. [Online]. Available: https://aclanthology.org/2022.lrec-1.345/

[8] D. Kempe, J. Kleinberg, and É. 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*. ACM, 2003, pp. 137–146.

[9] K. Saito, R. Nakano, and M. Kimura, "Prediction of information diffusion probabilities for independent cascade model," in *KES 2008, Part III. LNCS (LNAI)*, vol. 5179. Springer, 2008, pp. 67–75.

[10] Z. Wen, B. Kveton, M. Valko, and S. Vaswani, "Online influence maximization under independent cascade model with semi-bandit feedback," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/7137debd45ae4d0ab9aa953017286b20-Paper.pdf

[11] C. Wang, W. Chen, and Y. Wang, "Scalable influence maximization for independent cascade model in large-scale social networks," *Data Mining and Knowledge Discovery*, vol. 25, no. 3, pp. 545–576, 2012.

[12] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2009, pp. 199–208.

[13] D. Kempe, J. Kleinberg, and É. Tardos, "Influential nodes in a diffusion model for social networks," in *Automata, Languages and Programming*, ser. Lecture Notes in Computer Science, L. Caires, G. F. Italiano, L. Monteiro, C. Palamidessi, and M. Yung, Eds., vol. 3580. Berlin, Heidelberg: Springer, 2005, pp. 1127–1138.

[14] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," Stanford InfoLab, Tech. Rep. 1999-66, 1999.

[15] Media Bias/Fact Check, "Media bias/fact check," 2021, accessed: 2021. [Online]. Available: https://mediabiasfactcheck.com/

[16] S. Feng, G. Cong, A. Khan, X. Li, Y. Liu, and Y. M. Chee, "Inf2vec: Latent representation model for social influence embedding," in *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, 2018, pp. 941–952.

[17] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *ArXiv*, vol. abs/1908.10084, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:201646309

[18] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou, "Minilm: deep self-attention distillation for task-agnostic compression of pre-trained transformers," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, ser. NIPS '20. Red Hook, NY, USA: Curran Associates Inc., 2020.