# Heuristics Approaches for the Influence Maximization Problem on Hypergraphs

(Discussion Paper)

Vincenzo Auletta[1], Francesco Cauteruccio[1,*] and Diodato Ferraioli[1]

[1]*Department of Information Engineering, Electrical Engineering and Applied Mathematics (DIEM), University of Salerno, I84084, Fisciano, Italy*

**Abstract**

While the Influence Maximization (IM) problem has been extensively studied in graph topologies, with numerous algorithms and heuristics proposed, there has been relatively little focus on exploring this problem in the context of hypergraphs, which, despite being more complex, offer greater expressiveness. In this paper, we consider the current IM scenario considering hypergraph topologies, and we discuss two families of algorithms for the IM problem. The first family uses node importance measures that are specifically defined for hypergraphs, leveraging both topological characteristics and concepts from cooperative game theory. The second family addresses the problem through two well-known metaheuristic approaches, namely, hill climbing and evolution strategy. An initial experimental evaluation demonstrates that these approaches frequently outperform the leading algorithms currently proposed in the literature.

**Keywords**

Influence Maximization, Hypergraphs, Shapley Value, Hill Climbing, Evolution Strategies



## 1. Introduction

The advent of social media and viral marketing has highlighted the critical need to understand how information, behaviors, and innovations propagate through networks. Central to this exploration is the influence maximization (IM) problem, a key concept in network theory, where the objective is to identify a group of key individuals in a network whose combined influence is maximized [1, 2]. This problem has attracted substantial attention in different areas, and its implications are broad, impacting various applications, including opinion formation, combating misinformation, link recommendations, and even influencing election outcomes on social networks [3, 4, 5, 6, 7]. The primary goal is to identify the smallest set of nodes (or seeds) in a graph that can maximize the spread of information under specific propagation models. Traditionally, the IM problem has been studied within the context of standard networks with dyadic relationships [2, 8], with the seminal work in [1] establishing hardness results and introducing an algorithm with a $(1-1/e)$ approximation guarantee under the independent cascade (IC) and linear threshold (LT) diffusion models. Over time, a variety of approaches tackling the IM problem have been developed [9, 10, 11, 12]. However, recent advancements in hypernetwork science, a field that explores higher-order interactions within complex systems [13], have introduced new opportunities. Hypergraphs, the primary modeling tool in this field, extend traditional graph theory by enabling edges (which we now call hyperedges) to connect multiple vertices simultaneously, offering a richer representation of real-world phenomena. While the IM problem has been extensively studied in ordinary networks, its adaptation to hypergraphs remains a relatively nascent area of research. Early efforts focused on transforming hypergraphs into simpler structures, such as bipartite graphs, to leverage existing algorithms [14]. However, this approach often loses critical information encoded in the higher-order structure of hypergraphs. The computational challenges of IM in hypergraphs were rigorously

*AIxIA 2024 Discussion Papers - 23rd International Conference of the Italian Association for Artificial Intelligence, Bolzano, Italy, November 25–28, 2024*

*Corresponding author.
auletta@unisa.it (V. Auletta); fcauteruccio@unisa.it (F. Cauteruccio); dferraioli@unisa.it (D. Ferraioli)
0000-0002-7875-3366 (V. Auletta); 0000-0001-8400-1083 (F. Cauteruccio); 0000-0002-7962-5200 (D. Ferraioli)

explored in [15], which demonstrated the problem's NP-hardness and proposed an approximation framework with a $(1 - 1/e - \epsilon)$ guarantee. Building on this foundation, recent studies have developed specialized algorithms tailored to specific diffusion models. For instance, [16] introduced heuristics under the LT diffusion model, while [17] proposed a ranking-based approach for the HyperCascade model. Algorithms like HADP [18] and MEI [19] have sought to minimize overlap between seeds to enhance influence spread, though both rely on transforming hypergraphs into simpler graph representations. Direct approaches to IM on hypergraphs have also gained traction. Similarly, [20] modeled hypergraphs as electrostatic fields, introducing a novel perspective for assessing node influence. Other contributions, such as [21], extended message-passing techniques from ordinary networks to hypergraphs, focusing on collective influence within hyperedges. Research on weighted hypergraphs, as in [22], has introduced adaptive dissemination models to better capture real-world dynamics. Despite these advancements, many existing methods either specialize in a single diffusion process or require structural transformations that limit their generality.

Furthermore, it is worth noting that, to the best of our knowledge, several aspects remain understudied in this context. First, traditional types of IM algorithms have yet to be thoroughly explored for hypergraph topologies. Generally, classical IM approaches are categorized into four main types: *(i)* simulation-based, *(ii)* proxy-based, *(iii)* sketch-based, and *(iv)* intelligent optimization-based approaches. The first category includes algorithms that utilize techniques such as Monte Carlo simulations to model information propagation across individual nodes. A notable example is the algorithm proposed in [1], which has also been extended to hypergraph topologies. The second category adopts proxy models to approximate the influence spread of a given seed set, thereby avoiding potentially time-consuming simulations. Several algorithms in this category, such as those proposed in [18, 19], have been adapted for hypergraph topologies. The third category encompasses algorithms that evaluate influence spread by computing sketches based on the given graph and a specific diffusion model. While some classical approaches, such as reverse influence sampling, have been explored in the literature [18], their application to and formalization for hypergraph topologies remains an ongoing area of research. Finally, the fourth category involves the use of intelligent optimization algorithms, such as metaheuristic methods, to address the IM problem. Although a plethora of such approaches have been proposed in the classical setting, relatively few have been developed specifically for hypergraph topologies [23].

In this discussion paper, we illustrate an ongoing work focusing on the design and development of two families of approaches for tackling the IM problem on hypergraphs. These include node properties-based algorithms and metaheuristics-based algorithms. The former involves an algorithm that selects seed nodes by considering different centrality values, some of them also based on cooperative game theory. The latter consists of two metaheuristics algorithms based on hill climbing and evolutionary strategy, respectively. Also, we highlight an initial experimental evaluation, in which these algorithms are used, and we conclude by discussing several research directions that, in our opinion, are worth to be studied in the future.

The outline of the paper is as follows. In Section 2, we provide the background definitions and the problem statement. In Section 3, we discuss the two families of algorithms, and we briefly highlight an initial experimental evaluation assessing their performance. Finally, in Section 4, we draw our conclusion and discuss a series of future directions regarding this context.

## 2. Problem Statement

A *hypergraph* $H = (V, E)$ is a pair consisting of a set $V = \{v_1, \ldots, v_n\}$ of elements called *nodes*, and a family of sets $E = (e_1, \ldots, e_m)$ called *hyperedges*. A hyperedge represents a relation among a subset of vertices in $V$, i.e., $e_j \subseteq V$, for all $j = 1, \ldots, m$. The *order* of a hypergraph is its number of nodes, i.e., $n = |V|$, while the *size* of a hypergraph is $m = |E|$. A node $v_i \in V$ belongs to a hyperedge $e_j \in E$ if $v_j \in e_j$. The *degree* of a node $v_i$ is the number $d(v_i)$ of neighbors of $v_i$; a node $v_l \in V$ is the neighbor of node $v_i$ iff there exists at least one hyperedge $e_j$ which $v_l$ and $v_i$ both belong to. The *hyperdegree* of a node $v_i$ is the number $d^H(v_i)$ of hyperedges to which $v_i$ belongs. The *line graph* $L(H)$ of $H$ is the graph on node set $E' = \{e'_1, \ldots, e'_m\}$ and edge set $V' = \{\{e'_i, e'_j\} : e_i \cap e_j \neq \emptyset \text{ for } i \neq j\}$ [24]. In other words, $L(H)$ is the graph where nodes represent the hyperedges, and there is an edge between two nodes if the two hyperedges share at least one common node in $H$. Furthermore, given a set $V$ of nodes and a function $\nu \colon 2^V \to \mathbb{R}$, that assigns a measure of importance to each subset of nodes, the *Shapley value* of $v \in V$ [25] with respect to $\nu$ is defined as the average of the marginal contribution of $v$ to the subsets at which she belongs, i.e., how much she increases the importance of these groups. The Shapley value can be efficiently computed for the following specific choices of $\nu$ [26], namely: *(i)* $\nu_{deg}(S)$, that measures the importance of a subset $S$ of nodes as its size and the number of neighbors; (ii) $\nu_{close}(S)$, that measures the importance of a subset $S$ of nodes as the inverse of the minimum distance between nodes outside $S$ from nodes in $S$. Given a hypergraph $H = (V, E)$, a value $k \in \mathbb{Z}_{>0}$, and a diffusion process model on hypergraph $\sigma_H$, the Influence Maximization (IM) problem on hypergraphs consists in finding a subset $S^* \subseteq V$ of $k$ nodes, called *seed node set*, such that the expected number of infected nodes is maximized. Formally, $S^* = \arg\max_{S \subseteq V, |S|=k} \sigma_H(S)$, where $\sigma_H(S)$ indicates the expected influence (i.e., the number of reached nodes) of the seed node set $S$ at the end of the process. In line with the literature [18, 19], we use the Susceptible-Infected (SI) model with Contact Process (CP) dynamics on hypergraphs (SICP [18]). In this model, a node can be either in a susceptible (S) or infected (I) state. An S-state node can be infected by each of its neighbors in the I-state with a given infection rate $\beta$. The model works as follows: *(i)* nodes in the seed set are set to be infected (I-state), and the remaining nodes are susceptible (S-state); *(ii)* at each time step $t$, we find the I-state nodes. For each I-state node $v_i$, we find all hyperedges $E_i$ containing the node $v_i$. Then, a hyperedge $e_j$ is chosen from $E_i$ uniformly at random. Then, each of the S-state nodes in $e_j$ will be infected by $v_i$ with probability $\beta$; *(iii)* the process terminates after $T$ steps, and we set $\sigma_H(S)$ to be the number of nodes in I-state.

## 3. Algorithms and Results

We discuss two families of algorithms to tackle the IM problem on hypergraphs. The first one consists of an algorithm called SmartPROPS. It leverages node centrality to construct an optimal seed set, based on a node property function $\phi \colon V \to \mathbb{R}$ and a threshold function $\rho \colon V \to \mathbb{R}$. The idea behind the algorithm is to use $\phi$ to sort nodes, and then iteratively select the most important one based on $\rho$, which ensures that nodes with a considerable number of overlapping hyperedges are discarded. We propose four different variants of SmartPROPS, namely: *(i)* SmartDEG, in which the seeds coincide with the top-$k$ highest degree nodes [2], and $\phi$ is the degree centrality; *(ii)* SmartHYPERDEG, similar as the previous one, it exploits the hyperdegree of each node; *(iii)* SmartSHAPDEG, the Shapley Degree value, computed on $L(H)$, is used as the node property; *(iv)* SmartSHAPCLOSE, here the Shapley Closeness is used as the node property instead. The second family of algorithms is metaheuristics-based, and includes two algorithms, HC and ES. The former is based on a random-restart steepest ascent hill climbing approach. It is a simple yet powerful and versatile metaheuristic optimization algorithm that iteratively seeks to improve a solution with respect to a given measure of quality; it has been used extensively in disparate contexts [27, 28, 29]. The algorithm begins by randomly selecting an initial solution and iteratively improves it by generating neighbor solutions, via a perturbation function, and upgrading it accordingly to the calculated expected influence. Eventually, a global best solution is obtained when no further improvements are possible. Four variants can be defined: $\text{HC}_1$, where a node from $R$ is replaced with one of its neighbors chosen randomly from the largest hyperedge containing the former; $\text{HC}_2$, similarly as $\text{HC}_1$ but the node to be replaced is the one with the smaller degree value; $\text{HC}_3$ and $\text{HC}_4$ replace the node having respectively the smallest Shapley Degree value and the smallest Shapley Closeness value. Instead, ES draws inspiration from a population-based metaheuristic inspired by the principles of biological evolution, called evolution strategy [30], and it is based in particular on a variant called $(\mu + \lambda)$ [31], where $\mu$ is the number of candidate solutions in the parent generation, and $\lambda$ is the number of candidate solutions obtained from the parent generation. The algorithm iterates for a given number of generations. At each generation, the best $\mu$ solutions are kept from the $\lambda$ candidates and their parents. At the start, ES initializes a population of $\mu$ individuals, chosen uniformly at random.

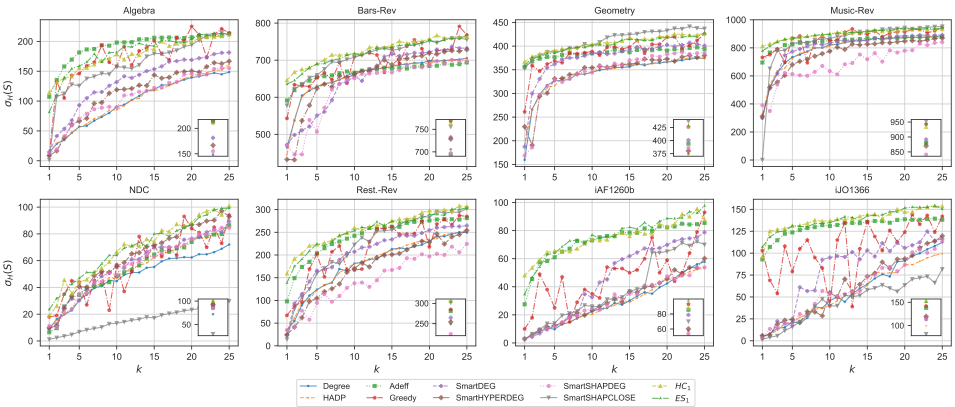


**Figure 1:** Expected influence $\sigma_H(S)$ obtained with different values of $k$ (from 1 to 25), averaged over 100 runs, and with the SICP parameters being $\beta = 0.01$ and $T = 25$.

In each generation, the algorithms iterates over $\lambda$ to generate the same amounts of new solutions via a mutation operator. This operator implements the same variations proposed for the HC algorithm, leading to the $ES_1$, $ES_2$, $ES_3$, and $ES_4$ variants.

We now highlight some preliminary results of the proposed algorithms on eight real-world hypergraphs used as benchmarks in the literature [18, 19]. Different baselines are considered, namely, Degree, Greedy, HADP [18], and Adeff [19][1]. The first is a common baseline based on selecting the top-$k$ nodes with the maximal degree. The second one is drawn by the approach in [1], and selects the node with maximal influence in each iteration. The last two are recently proposed algorithms for the problem. For the discussed algorithms, we set parameters as follows: for HC, we set the number of restarts to 25, while for ES, we set $\mu = 4$, $\lambda = 10$, and the maximum number of generations to 25. The obtained influence spread curve is presented in Figure 1, averaged over 100 runs. For variants, we only show $HC_1$ and $ES_1$. The $x$-axis refers to the value of $k$, while the $y$-axis reports the average expected influence $\sigma_H(S)$. Inside each plot, a smaller one reports the obtained value at $k = 25$. We can see how the metaheuristics-based algorithms generally perform better than the remaining ones. Also, they achieve the largest expected influence when $k = 1$. Except for some cases where the Greedy reaches similar values, no other algorithms perform in the same manner. As far as $k = 25$ is concerned, we observe almost identical behavior to the previous one. Here, $HC_1$ and $ES_1$ perform effectively, together with the Greedy one.

## 4. Conclusion and Future Directions

This discussion paper explores the Influence Maximization (IM) problem within hypergraph topologies, an area that remains largely understudied despite its significant potential for capturing higher-order interactions in complex systems. In this paper, we critically discuss two families of algorithms, namely, *(i)* node properties-based, *(ii)* and metaheuristics-based ones, designed to address the challenges of the IM problem on hypergraph topologies. The first family involves an algorithm that leverages hypergraph-specific features, such as Shapley-based centrality measures, to select the seed set, while the second one includes metaheuristics algorithms, and in particular hill climbing and evolutionary strategies, to tackle the problem. Our initial analysis, supported by experimental observations, provides insight into the potential of these approaches while highlighting areas where further refinement is needed.

As this is an ongoing effort, several directions for future research emerge. First, we aim to deepen our understanding of the problem by performing a comprehensive evaluation across diverse hypergraph structures and diffusion models. This includes investigating settings with different conditions, such as hypergraphs evolving over time, and hypergraphs presenting weights and other dynamic features.

[1]All algorithms have been implemented in Python 3.10

Second, a critical area of future work lies in designing scalable algorithms suitable for large-scale hypergraphs. Indeed, such algorithms could require exploring parallelization techniques and designing more efficient heuristics. Furthermore, incorporating adaptive mechanisms to optimize parameter selection dynamically could improve algorithmic robustness and applicability. Third, an understudied aspect is investigating practical applications of the IM problem in domains such as social influence campaigns, biological networks, and knowledge graph propagation. Bridging the gap between theoretical discussions and real-world deployment is essential for demonstrating the utility of hypergraph-based IM approaches. Finally, future work could also explore the integration of these algorithms with machine learning models to predict influence spread, potentially combining traditional IM techniques with data-driven approaches.

## Acknowledgments

This work is supported by the PNRR project FAIR—Future AI Research (PE00000013) under the NRRP MUR program funded by NextGenerationEU.

## References

[1] D. Kempe, J. Kleinberg, E. Tardos, Maximizing the spread of influence through a social network, in: Proc. of the of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD), 2003, pp. 137–146. ACM.

[2] S. Banerjee, M. Jenamani, D. K. Pratihar, A survey on influence maximization in a social network, Knowledge and Information Systems 62 (2020) 3417–3455. Springer.

[3] V. Auletta, D. Ferraioli, G. Greco, On the complexity of reasoning about opinion diffusion under majority dynamics, Artificial Intelligence 284 (2020) 103288. Elsevier.

[4] M. Castiglioni, D. Ferraioli, N. Gatti, G. Landriani, Election manipulation on social networks: Seeding, edge removal, edge addition, Journal of Artificial Intelligence Research 71 (2021) 1049–1090. AI Access Foundation.

[5] G. Bonifazi, F. Cauteruccio, E. Corradini, M. Marchetti, A. Pierini, G. Terracina, D. Ursino, L. Virgili, An approach to detect backbones of information diffusers among different communities of a social platform, Data & Knowledge Engineering 140 (2022) 102048. Elsevier.

[6] W. Chen, C. Wang, Y. Wang, Scalable influence maximization for prevalent viral marketing in large-scale social networks, in: Proc. of the ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), 2010, pp. 1029–1038. ACM.

[7] B. Wilder, L. Onasch-Vera, J. Hudson, J. Luna, N. Wilson, R. Petering, D. Woo, M. Tambe, E. Rice, End-to-end influence maximization in the field, in: Proc. of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS), volume 18, 2018, pp. 1414–1422.

[8] P. Liu, L. Li, Y. Wen, S. Fang, Identifying influential nodes in social networks: Exploiting self-voting mechanism, Big Data 11 (2023) 296–306. Mary Ann Liebert.

[9] A. Goyal, W. Lu, L. V. Lakshmanan, Celf++ optimizing the greedy algorithm for influence maximization in social networks, in: Proc. of the 20th international conference companion on World Wide Web (WWW), 2011, pp. 47–48. ACM.

[10] C. Wang, W. Chen, Y. Wang, Scalable influence maximization for independent cascade model in large-scale social networks, Data Mining and Knowledge Discovery 25 (2012) 545–576. Springer.

[11] Y. Tang, Y. Shi, X. Xiao, Influence maximization in near-linear time: A martingale approach, in: Proc. of the 2015 ACM International Conference on Management of Data (SIGMOD), 2015, pp. 1539–1554. ACM.

[12] W. Chen, Y. Yuan, L. Zhang, Scalable influence maximization in social networks under the linear threshold model, in: Proc. of the 2010 IEEE International Conference on Data Mining (ICDM), 2010, pp. 88–97. IEEE.

[13] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, G. Petri, Networks beyond pairwise interactions: Structure and dynamics, Physics Reports 874 (2020) 1–92. Elsevier.
[14] F. Amato, V. Moscato, A. Picariello, G. Sperlí, Influence maximization in social media networks using hypergraphs, in: Proc. of the International Conference on Green, Pervasive, and Cloud Computing (GPC), 2017, pp. 207–221. Springer.
[15] J. Zhu, J. Zhu, S. Ghosh, W. Wu, J. Yuan, Social influence maximization in hypergraph in social networks, IEEE Transactions on Network Science and Engineering 6 (2018) 801–811. IEEE.
[16] A. Antelmi, G. Cordasco, C. Spagnuolo, P. Szufel, Social influence maximization in hypergraphs, Entropy 23 (2021) 796. MDPI.
[17] A. MA, A. Rajkumar, Hyper-imrank: Ranking-based influence maximization for hypergraphs, in: Proc. of the 5th Joint International Conference on Data Science & Management of Data (9th ACM IKDD CODS and 27th COMAD), 2022, pp. 100–104. ACM.
[18] M. Xie, X.-X. Zhan, C. Liu, Z.-K. Zhang, An efficient adaptive degree-based heuristic algorithm for influence maximization in hypergraphs, Information Processing & Management 60 (2023) 103161.
[19] X. Gong, H. W. X., Wang, C. Chen, W. Zhang, Y. Zhang, Influence maximization on hypergraphs via multi-hop influence estimation, Information Processing & Management 61 (2024) 103683.
[20] S. Li, X. Li, Influence maximization in hypergraphs: A self-optimizing algorithm based on electrostatic field, Chaos, Solitons & Fractals 174 (2023) 113888. Elsevier.
[21] R. Zhang, X. Qu, Q. Zhang, X. Xu, S. Pei, Influence maximization based on threshold models in hypergraphs, Chaos: An Interdisciplinary Journal of Nonlinear Science 34 (2024). AIP Publishing.
[22] Q. Pan, H. Wang, J. Tang, Z. Lv, Z. Wang, X. Wu, Y. Ruan, T. Yv, M. Lao, Eioa: A computing expectation-based influence evaluation method in weighted hypergraphs, Information Processing & Management 61 (2024) 103856. Elsevier.
[23] S. Genetti, E. Ribaga, E. Cunegatti, Q. F. Lotito, G. Iacca, Influence maximization in hypergraphs using multi-objective evolutionary algorithms, in: Proc. of the 18th International Conference on Parallel Problem Solving from Nature (PPSN), 2024, pp. 217–235. Springer.
[24] S. G. Aksoy, C. Joslyn, C. O. Marrero, B. Praggastis, E. Purvine, Hypernetwork science via high-order hypergraph walks, EPJ Data Science 9 (2020) 16. Springer.
[25] L. Shapley, A value for n-person games, Contributions to the Theory of Games (1953) 307–317. Princeton University Press.
[26] T. Michalak, K. Aadithya, P. Szczepanski, B. Ravindran, N. Jennings, Efficient computation of the shapley value for game-theoretic network centrality, Journal of Artificial Intelligence Research 46 (2013) 607–650. AI Access Foundation.
[27] M. Gendreau, J.-Y. Potvin, Handbook of metaheuristics, volume 2, Springer, 2010.
[28] H. Li, Y. Liu, S. Yang, Y. Lin, Y. Yang, J. Yoo, An improved hill climbing algorithm for graph partitioning, The Computer Journal 66 (2023) 1761–1776.
[29] C. Stamile, F. Cauteruccio, G. Terracina, D. Ursino, G. Kocevar, D. Sappey-Marinier, A model-guided string-based approach to white matter fiber-bundles extraction, in: Proc. of 8th International Conference on Brain Informatics and Health (BIH 2015), London, UK, August 30-September 2, 2015. Proceedings 8, 2015, pp. 135–144. Springer.
[30] T. Back, Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms, Oxford University Press, 1996.
[31] D. Simon, Evolutionary optimization algorithms, John Wiley & Sons, 2013.