Unveiling Environmental Sensitivity of Individual Gains in Influence Maximization

Xinyan Su, Zhiheng Zhang
Jiyan Qiu, Zhaojuan Yue,
Jun Li

NEURAL INFORMATION PROCESSING SYSTEMS

Abstract

Influence Maximization (IM) seeks to identify a seed set to maximize information dissemination in a network. In the prevailing literature, these individual gains are typically assumed to remain constant throughout the cascade process, and are solvable using explicit formulas based on the node's characteristics and network topology. However, this assumption is not always feasible due to two key challenges:

- ***Unobservability***: The individual gains of each node are primarily evaluated by the difference between the outputs in the activated and non-activated states. In practice, we can only observe one of these states, with the other remaining unobservable post-propagation.
- ***Environmental sensitivity***: In addition to the node's inherent properties, individual gains are also sensitive to the activation status of surrounding nodes, which changes dynamically during propagation, even if the network topology remains fixed.

we introduce a Causal Influence Maximization (CauIM) framework, leveraging causal inference techniques to model dynamic individual gains. We propose two algorithms, **G-CauIM and A-CauIM**, where the latter incorporates a novel acceleration technique.

Introduction

Traditional Influence Maximization

Due to the power of the "word-of-mouth" phenomenon, influence spread has been demonstrated as a necessity in various applications, such as viral marketing[1], HIV prevention[2] and recommendations[3]. The problem of selecting the seed set to maximize information spread is known as the **I**nfluence **M**aximization (**IM**)[4]

Target Problem

Researchers endeavor to address the question: *how limited resources can be utilized to maximize total gains?* This challenge manifests in various network scenarios, like awareness dissemination and product promotion. For example, when targeting users with varying purchasing power in product promotion, these users exhibit diverse purchasing behaviors resulting in varying profits for the seller. Here, regarding purchasing power as individual gains, the goal is to identify specific users for product advertising and and **optimize the overall difference in profit gains pre- and post-product promotion dissemination.**

Ilustration of individual gains during a certain propagation iteration

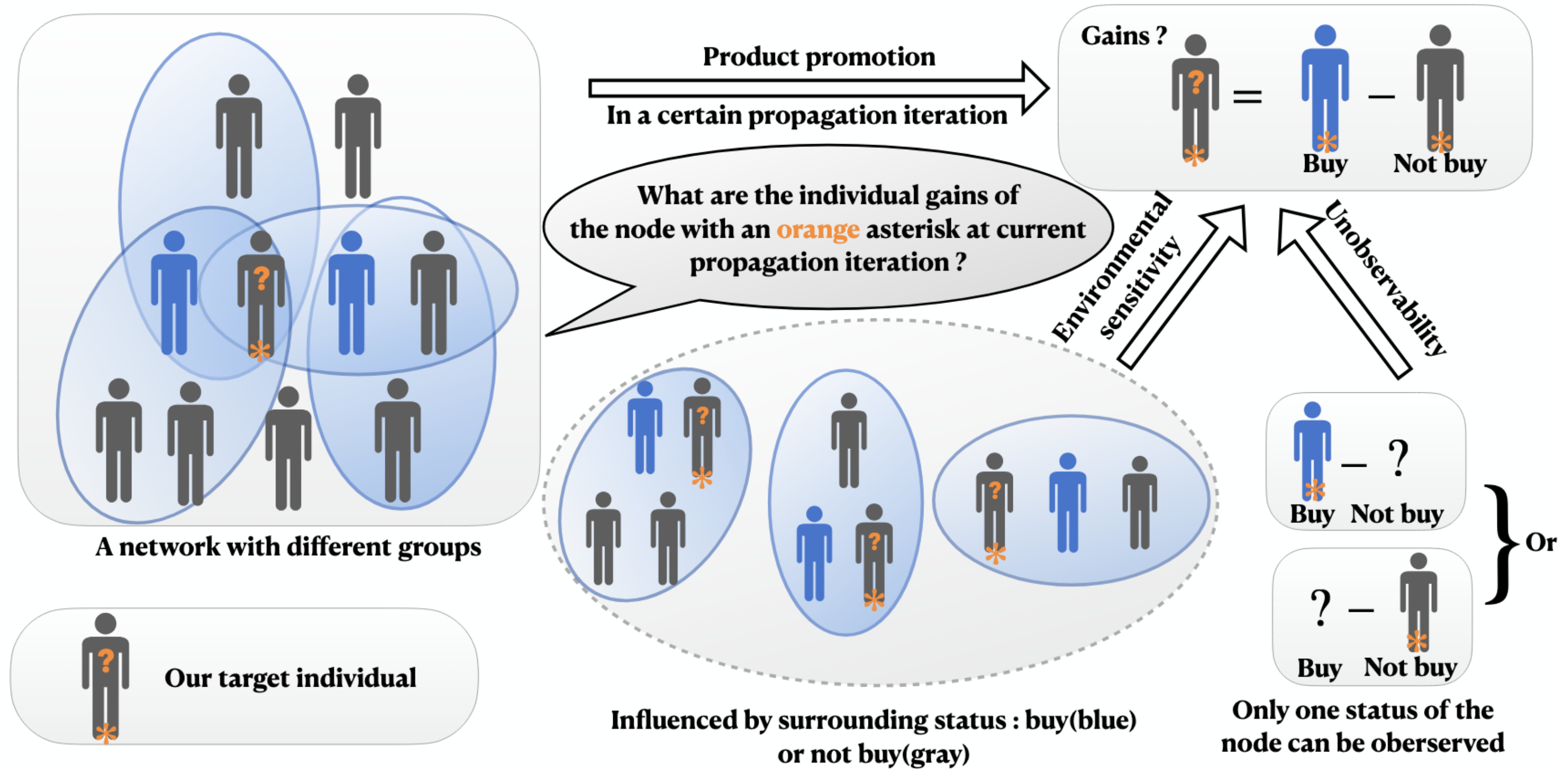


Figure1: we focus on the analysis of the starred (*) node. The leftmost diagram represents the current iteration, showing nodes in different states within the network: blue nodes indicate activation (purchase), grey nodes remain inactive, and "?" represents an unknown status. The actual individual gains is defined as the difference between the profit of the node in its activated state and its non-activated state.

$$v := \arg\max_{v \notin S} \left\{ \frac{\partial\left(\sigma(S)\right)}{\partial\left(ap(v;S)\right)} * ap(v;S) \right\}. \tag{7}$$

A-CauIM Algorithm

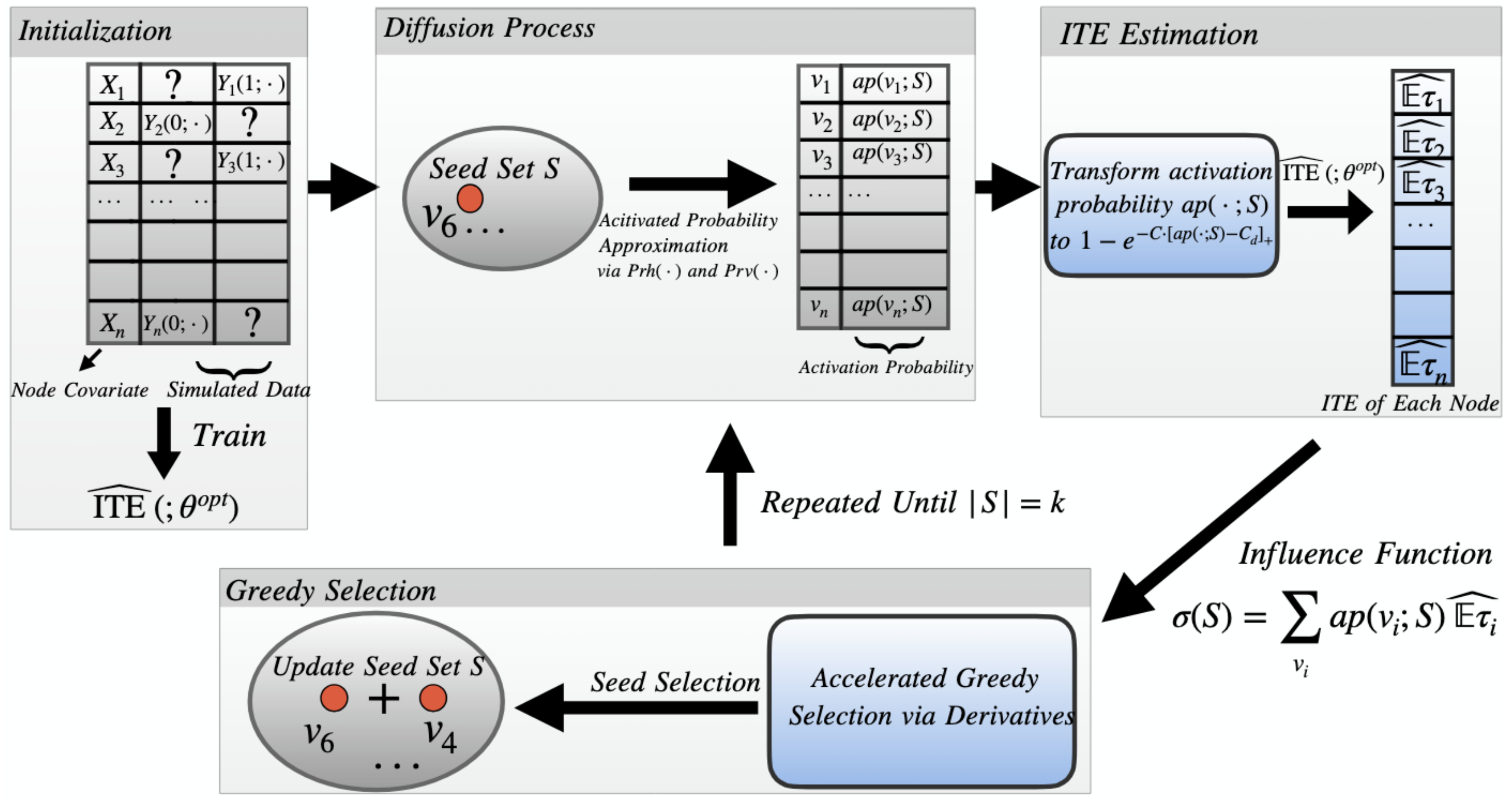


Figure 2: A-CauIM. Compared with G-CauIM (Alg. 1), we add a storage table for activation probabilities $ap(;)$ and then simplify the complex greedy selection (Eq. 5) into more efficient derivative operations (Eq. 7). In addition, we we transform $ap(;)$ into continuous values closing to $0,1$ to signify the activated states T_i of each node on average. And by this procedure, we obtain $\widehat{\mathbb{E}\tau_i}$ which is the approximation of the expectation on unobserved τ_i.

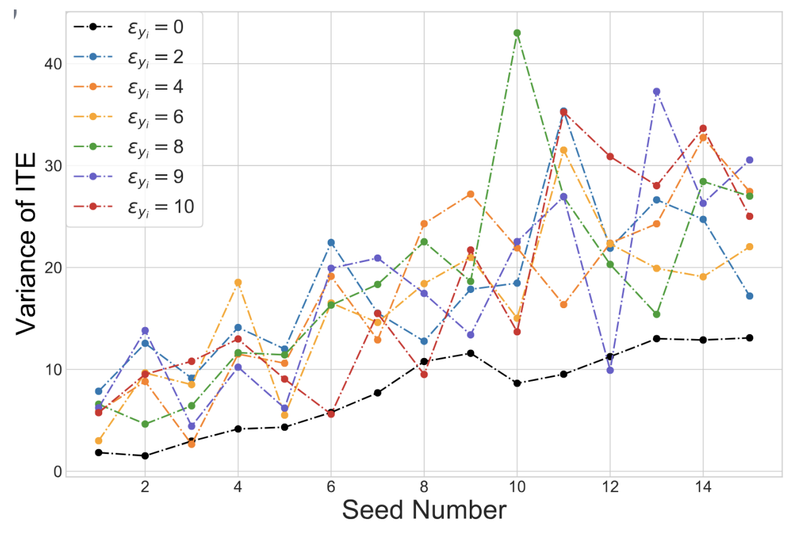
Algorithm 1: G-CauIM

```
Input: $\mathcal{G}(\mathscr{V},\mathscr{H},\mathbb{H})$; seed number $K$; $X_i$, initial treatment $T_i$ and $\boldsymbol{T}_{-i}$ of each node $v_i$; observational data
       $D=\{Y_i(t;\cdot)\}_{v_i\in\mathscr{V}}$, where $t=T_i$; the bound $M$ for ITE.
Output: Deterministic seed set $S^*$ with $|S^*|=K$.
Function $\widehat{\mathrm{ITE}}(X_i, T_i, \boldsymbol{T}_{-i}, \mathcal{G}; \theta)$:
    Compute the representation $Z_i$ of $X_i$ via representation learning;
    Compute the high-order interference representation $O_i := \mathrm{ENV}(\mathbb{H}, \boldsymbol{T}_{-i}, \boldsymbol{Z}_{-i}; \theta)$ (Assumption 2.1 and
     Assumption 2.2);
    Concatenate $Z_i$, $Q_i$ and feed them into a Multi-Layer Perception (MLP):
     $\{\hat{Y}_i(1;\cdot), \hat{Y}_i(0;\cdot)\} \sim \mathrm{MLP}([Z_i||O_i])$;
    Compute the ITE $\hat{\tau}_i = \hat{Y}_i(1;\cdot) - \hat{Y}_i(0;\cdot)$ for $v_i$;
    return $\hat{\tau}_i$;
Function Main:
    (Initialization) $S^* = \emptyset$; Loss $= 0$;
    (Training): Compute the cumulative loss by $D$: Loss $= \sum_{v_i\in\mathscr{V},t=0,1} \left|\left(\hat{Y}_i(t;\cdot) - Y_i(t;\cdot)\right)\mathbb{I}(T_i = t)\right|$
     via the above $\widehat{\mathrm{ITE}}(\cdot;\theta)$ function;
    Get the optimal $\theta^{opt} := \arg\min\{\mathrm{Loss} : \forall v_i, \hat{\tau}_i \le M\}$;
    for $|S^*| < K$ do
        Conduct propagation under current seed set $S$, generate $\hat{\tau}_i = \widehat{\mathrm{ITE}}(X_i, T_i, \boldsymbol{T}_{-i}, \mathcal{G}; \theta')$ for $v_i \notin S$,
         where $T_i$ is changed to its current activated state( 0 or 1), and $\boldsymbol{T}_{-i}$ is changed based on other
         nodes' activated states, $\theta' := \theta^{opt} + \triangle_\theta$, $\triangle_\theta := \min\{\|\theta_q\| : \exists\delta \le \|\theta_q\|, \widehat{\mathrm{ITE}}(\cdot;\theta+\delta) \le M\}$,
         repeat the process and get the mean;
        $v_0 = argmax_{v\notin S^*}\{\sigma(S^*\cup\{v\}) - \sigma(S^*)\}$;
        $S^* = S^* \cup \{v_0\}$;
return $S^*$.
```

Experiments

Table 1: RQ1: Performance comparison of four different methods under four datasets (seed number=15). Our methods gain general improvements compared with baselines: Traditional Greedy (donated as "Baseline") and Random Selection.

Methods	GoodReads	Contact	Email-Eu	SD-100
Baseline	297.56	68.12	735.28	138.91
Random	45.86	66.51	590.67	145.97
G-CauIM	**330.25**	**69.53**	**804.28**	151.59
A-CauIM	302.17	66.78	802.41	**160.49**



(c) Analysis on noise ϵ

We aim to answer the following questions. 1) To calculate the max sum of node ITE that represents the overall individual gains, can our G-CauIM and A-CauIM outperform the traditional IM methods and maintains high efficiency? 2) If our ITE estimation is not accurate enough, can CauIM perform more robustly? The robustness means in perturbations, our approach will achieve an approximate result close to the normal state.

References

[1] Chen, W., Wang, C., & Wang, Y. (2010, July). Scalable influence maximization for prevalent viral marketing in large-scale social networks. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1029-1038).

[2] Wilder, B., Onasch-Vera, L., Hudson, J., Luna, J., Wilson, N., Petering, R., ... & Rice, E. (2018, July). End-to-End Influence Maximization in the Field. In AAMAS (Vol. 18, pp. 1414-1422).

[3] Coró, F., D'angelo, G., & Velaj, Y. (2021). Link recommendation for social influence maximization. ACM Transactions on Knowledge Discovery from Data (TKDD), 15(6), 1-23.

[4] Kempe, D., Kleinberg, J., & Tardos, É. (2003, August). Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 137-146).