Robust Influence Maximization

(Extended Abstract)

Meghna Lowalekar, Pradeep Varakantham, Akshat Kumar
School of Information Systems, Singapore Management University
meghnal.2015@phdis.smu.edu.sg,{pradeepv,akshatkumar}@smu.edu.sg

ABSTRACT

Influence Maximization [2] is the problem of finding a fixed size set of nodes, which will maximize the expected number of influenced nodes in a social network. The number of influenced nodes is dependent on the influence strength of edges that can be very noisy. The noise in the influence strengths can be modeled using a random noise or adversarial noise model. It has been shown that all random processes that independently affect edges of the graph can be absorbed into the activation probabilities themselves and hence random noise can be captured within the independent cascade model.

On the other hand, similar to He *et al.* [1], we consider the adversarial noise where influence strength for an edge can belong to any point in the interval: $[\check{p}_{u,v}, \hat{p}_{u,v}]$ and the exact values are chosen by an adversary from this interval. The problems of evaluating robustness of a given solution and computing robust optimal solutions have received scant attention in the literature and are of key interest in this paper. Specifically, we aim to minimize (over all available seed sets) the maximum (over all instantiations of influence strengths) regret. Concretely, the key contributions are: **(1)** We show that maximum regret for a given solution is attained when influence strength on each of the edges is set to one of the extreme values of the influence strength intervals on edges. **(2)** We provide a novel way of considering samples that accounts for the noise in influence strength on all edges. **(3)** We develop a framework which provides an approach to get an optimal regret solution and more importantly a metric to evaluate robustness of a given solution based on the regret optimal solution. **(4)** Finally, we show results on evaluating the robustness of the well known greedy approach. Surprisingly, even without considering noise in influence strengths explicitly, greedy approach achieves highly robust solutions on small-medium scale social network instances.

Keywords

Influence Maximization, Regret, Optimization

1. ROBUST INFLUENCE MAXIMIZATION

The value of influence strength, $p_{u,v}$ from u to v can not be observed directly and is generally inferred from observed behavior which is inherently very noisy. We assume that the actual influence strength for an edge can belong to any point in the interval: $[\check{p}_{u,v}, \hat{p}_{u,v}]$. Our goal is to compute a robust seed set of M nodes that will minimize the maximum regret over the interval noise in influence strengths. Let $\mathbb{S}$ denote the set of all possible node sets of size M and let $\mathcal{P} = \{[\check{p}_e, \hat{p}_e]\}_{e \in E}$ represent the noise in influence strengths. We define regret, maximum regret and minimax regret using the following equations:

$$\delta(S, \mathbf{p}) = \max_{S' \in \mathbb{S}} \ \sigma_{\mathbf{p}}(S') - \sigma_{\mathbf{p}}(S) \quad (1)$$

$$\delta^{\mathbf{MR}}(S, \mathcal{P}) = \max_{p \in \mathcal{P}} \ \delta(S, \mathbf{p}) \quad (2)$$

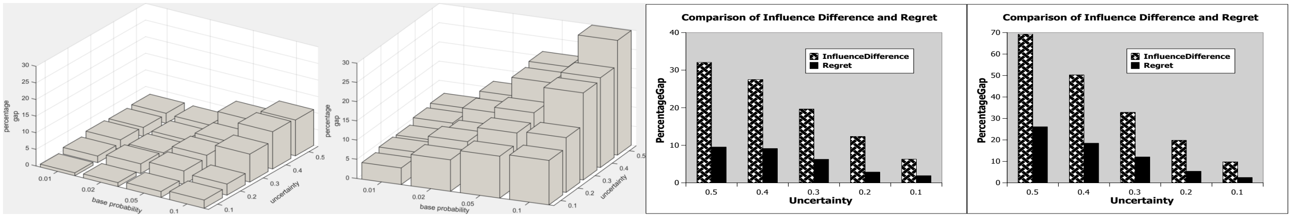
$$\delta^{\mathbf{MMR}}(\mathcal{P}) = \min_{S \in \mathbb{S}} \ \delta^{\mathbf{MR}}(S, \mathcal{P}) \quad (3)$$

where S' and S are seed sets of size M and $\sigma_{\mathbf{p}}(S)$ is the expected influence value of set S for the probability values $\mathbf{p}$. We can then show that maximum regret for any set S is obtained when all edge probabilities are at extreme values.

The proposed optimization model for calculating $\mathbf{MMR}(\mathcal{P})$ is extended on the one by Sheldon *et al.* [3]. The constraints in the formulation ensure that for a given seedset, regret is higher than maximum possible regret over all influence strength vectors in set $\mathcal{P}$ and the objective ensures the computation of least maximum regret over all possible seedsets. Since the set of extreme influence strength vectors grows exponentially with the number of edges, the optimization formulation is not scalable. We address this challenge by using constraint generation.

The key idea is to decompose the formulation of $\mathbf{MMR}(\mathcal{P})$ into two components that are run iteratively until convergence: **(1)** A master component, $\mathbf{MMR}(\mathcal{P}_{sub})$ that computes minimax regret solution, S for a discrete subset, $\mathcal{P}_{sub}$ of probability vectors. **(2)** A slave component, $\mathbf{MR}(S, \mathcal{P})$ that computes the probability vector, p^*, which yields maximum regret given $\mathcal{P}$ and S. If $\mathbf{MR}(S, \mathcal{P}) \leq \mathbf{MMR}(\mathcal{P}_{sub})$, then we stop the process and return S, otherwise p^* is added to $\mathcal{P}_{sub}$ for the master component to compute a new S. The key novelty of the slave formulation is in identifying a probability vector (with extreme values for each edge) for a given seed set that maximizes regret. Unlike in traditional influence maximization and in master component, the activation of an edge is not generated before hand (outside optimization model) based on uniform random numbers, but in the formulation as we have an interval of influence strengths for each edge. Given the influence strength interval, $[\check{p}_e, \hat{p}_e]$ for an edge e, constraints ensure that edge is active with a min-

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*



(a) (b) (c) (d)

Figure 1: Percentage gap of greedy solution (a) SmallW (b) NetScience. In (a) Number Of Samples($|\Xi|$) = 200 and (b) $|\Xi|$ =110. Comparison of Influence Difference and Regret on (c) SmaGri Network; and (d) CSPhD Network.

imum probability of $\check{p}_e$ and a maximum probability of $\hat{p}_e$.

Evaluating Robustness of a Given Solution We now describe a method to compute percentage gap of a given solution from the optimal minimax regret for a given set of samples. We denote $\hat{\mathbf{R}}$ and $\check{\mathbf{R}}$ as the upper and lower bound on regret, if the algorithm is stopped before convergence. We can provide a bound on the difference from regret optimal influence value. Let $\sigma_{\mathbf{p}}(S^*)$ be the optimal influence value for any influence strength vector $\mathbf{p}$,then for the current solution S, the upper and lower bound of influence at $\mathbf{p}$ are:

$$\hat{\sigma}_{\mathbf{p}}(S) = \sigma_{\mathbf{p}}(S^*) - \check{\mathbf{R}}, \check{\sigma}_{\mathbf{p}}(S) = \sigma_{\mathbf{p}}(S^*) - \hat{\mathbf{R}}$$

The percentage gap, Δ for S and $\mathbf{p}$ is given by

$$\Delta_{\mathbf{p}}(S) = \frac{\hat{\sigma}_{\mathbf{p}}(S) - \check{\sigma}_{\mathbf{p}}(S)}{\hat{\sigma}_{\mathbf{p}}(S)} * 100 = \left(1 - \frac{1}{1 + \frac{\hat{\mathbf{R}} - \check{\mathbf{R}}}{\sigma_{\mathbf{p}}(S^*) - \hat{\mathbf{R}}}}\right) * 100 \quad (4)$$

The above expression will be maximum when $\sigma_{\mathbf{p}}(S^*)$ is minimum and the influence value will be minimum when all the edges are at minimum probability. So we calculate $\Delta_{\mathbf{p}}(S)$ at $\mathbf{p} = \{\check{p}_{u,v}, \forall u, v\}$ to obtain $\Delta(S)$ (the worst case percentage gap). In summary, $\Delta(S)$ gives us one number to evaluate the robustness of a given solution S by comparing it against our approach.

2. EXPERIMENTAL RESULTS

In this section, we evaluate the percentage gap of the well known greedy algorithm [2] for varying levels of uncertainty. We are able to demonstrate that percentage gap for greedy is low (<30%) even at very high uncertainty levels (20%-50%). We also show that the existing metric of influence difference proposed by [1] is not suitable to evaluate robustness of solution. We employ the settings used by [1], and vary the base probability to take on the values $\{0.01, 0.02, 0.05, 0.1\}$. The uncertainty interval for an edge (u, v) is: $[(1-\varepsilon) * p_{u,v}, (1+\varepsilon) * p_{u,v}]$, where $p_{u,v}$ is the observed base probability and ε is the uncertainty parameter which takes on the values $\{10\%, 20\%, 30\%, 40\%, 50\%\}$.

We first demonstrate the percentage gap values for greedy heuristic provided by the regret metric. We calculate the maximum regret of this greedy solution S^g by using MaxRegret linear program. Let $\mathbf{R}^g$ be the regret of greedy solution, then we calculate the percentage gap of the greedy solution, $\Delta(S^g)$ from the regret optimal solution by substituting $\mathbf{R}^g$ in place of $\hat{\mathbf{R}}$ in equation (4). The detailed percentage gap results for the greedy algorithm on the 2 data sets [1] are given in Figure 1(a-b). Each 3D graph has base probability on the X-axis, uncertainty interval on the Y-axis and the percentage gap of the greedy based compared to regret optimal (or regret lower and upper bounds) on the Z-axis.

We also observe that as the base probability and uncertainty increases the percentage gap from the optimal solution increases. This is because for higher base probability and uncertainty value convergence is slower. So the percentage gap of greedy solution will improve if we can compute the exact value of optimal regret solution.

For adversarial noise model, the only other metric that has been proposed is by He *et al.* [1]. While "influence difference maximization" is employed to measure the stability of an instance and not robustness of a solution, "influence difference for a given solution" can be employed to evaluate robustness of a solution. Influence difference value for a given solution S is calculated using equation 1 in [1](by fixing the solution S and base probability, θ), i.e., $\lambda(S) = max_{\mathbf{p} \in \mathcal{P}} |\sigma_{\mathbf{p}}(S) - \sigma_{\theta}(S)|$

Specifically, we calculate the percentage gap generated by the greedy solution (for base probability) using both the metrics. Let $\lambda(S^g)$ be the influence difference of greedy solution, S^g. We substitute the value $\lambda(S^g)$ in place of $\hat{\mathbf{R}}$ in equation (4), to calculate percentage gap of solution S^g by using influence difference. Figure 1(c)-(d) shows the comparison of percentage gap of S^g by using influence difference and regret metrics for 2 networks for 0.1 base probability and different uncertainty values. By using regret in place of influence difference, we were able to reduce the bounds by almost 60%. In other words, if influence difference indicates that a solution is not-robust, that may not be correct as it is an upper bound on regret. This shows that our method is able to more accurately evaluate robustness of a solution.

Acknowledgments This research is supported by Singapore National Research Foundation under its International Research Centre @Singapore Funding Initiative and administered by the IDM Programme Office, Media Development Authority (MDA).

[1]http://vlado.fmf.uni-lj.si/pub/networks/data/

REFERENCES

[1] X. He and D. Kempe. Stability of influence maximization. *CoRR*, abs/1501.04579, 2015.

[2] D. Kempe, J. Kleinberg, and E. 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*, KDD '03, pages 137–146, New York, NY, USA, 2003. ACM.

[3] D. Sheldon, B. N. Dilkina, A. N. Elmachtoub, R. Finseth, A. Sabharwal, J. Conrad, C. P. Gomes, D. B. Shmoys, W. Allen, O. Amundsen, and W. Vaughan. Maximizing the spread of cascades using network design. In *UAI, Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence, Catalina Island, CA, USA, July 8-11, 2010*, pages 517–526.