Influence Maximization with Fairness

Ke Han Xiao, Junyang Zhou, Shan Liu

November 1st, 2021

1 Introduction

Influence maximization in social networks is a well-known and studied problem that can be widely applied in many different subjects and fields. Its goal is to maximize the influence by selecting and activating a seed set of limited size. Each activated node then activates its connected child nodes with some pre-defined or pre-learned probability, and the influence is evaluated by the number or proportion of nodes activated.

There has been a lot of work trying to solve the influence maximization problem, but this basic form of the problem does not take into account the fairness of influence spread, which can be crucial in some applications. For example, when the information spread to maximize can improve social welfare, we would prefer all different groups and communities within the society can receive a fair amount of influence.

2 Goals

When we are doing a literature review on IM fairness, we aim to provide one or more definitions (s) of fairness in the context of influence maximization. Then, we will propose an algorithm that can efficiently approximate the influence maximization problem with our definition of fairness. We will also measure the trade-off between influence maximization and fairness, as well as the trade-off between efficiency and fairness. Lastly, we will measure the time complexity and the approximation factor and accuracy of the algorithm. During our research, we aim to experiment with the algorithm on the dataset(s) and compare its performance with existing algorithms.

3 Related research

In recent years, some works have addressed fairness in the context of the influence maximization problem. Maximin fairness [TZ19] measures and improves fairness by maximizing the minimum node activation ratio (number of nodes activated / number of nodes in the group) among all groups in a social network. Therefore, it imposes a lower bound on the share of influence received by each group, but Maximin [TZ19] does not take into account the differences of node activation ratios between groups. In other words, when using Maximin's [TZ19], the maximum and minimum node activation ratios can be very different. This problem is captured in demographic parity ([RA20]), which requires setting an upper bound for the difference of node activation ratios between any two groups, but it can also be hard to find a good value for this upper bound. Therefore, Rahmattalabi et al.[RA20] proposed a welfare fairness optimization framework based on social welfare theory, in which the tradeoff between fairness consideration and efficiency can be controlled by a parameter.

4 Approach

In this section, we will first discuss the different objectives of our problem, which are around the central idea of "Influence Maximization with Fairness". Then, for each of the objectives, we will define a corresponding notion of fairness, w.r.t some approximation algorithms to achieve it. Finally, we will compare their effect w.r.p to our real topic.

4.1 Maximize the sub-influence within each community. ([TZ19])

Objective: Disregarding the total influence, by distributing a fixed amount of seeds for each community, we are trying to maximize the partial influence for all of them.
Notion of fairness: For each community C_i, Given a graph $G = (V, E)$ and a seed set $A \subseteq V$,

$$Maxmin(A) = min_i \frac{I_{G,C_i}(A)}{|C_i|} \tag{1}$$

Such that $I_{G,C_i}(A)$ is the maximum expected number of vertices from C_i that can be activated by the seed set A.
However, in the paper "Group-Fairness in Influence Maximization", [TZ19] the author mentioned this notion of fairness will not preserve the submodularity. Therefore, the author provides an approximation algorithm that maximizes the sub-influence within each community.

Algorithm 1 sub-influence maximization

With a seed constraint k, distribute seeds k_i to each community C_i, such that $k_i = k\frac{|C_i|}{|V|}$. This guarantees the number of seeds will be fairly allocated to the group w.r.t the group size.
for *each* C_i **do**
Proceed the Influence Maximization w.r.t the sub-graph induced by the community C_i by using k_i seeds.
end for

By the author's construction, this algorithm is able to give a $(1 - 1/e)$ approximation within each sub-community. Since we disregard the total influence across the whole graph, the author also argued that the overall influence is inapproximable [KR08].
Thus, to introduce the notion of total influence, we introduce another target.

4.2 Maximize the total influence with seed fairness

Objective: We will still focus to maximize the total influence of the entire graph, however, with fairness distribute the number of seeds from each community.

Notion of fairness: Inherited from the section above, for each community C_i, Given a graph $G = (V, E)$ and a total seed number k, we define the number of seeds k_i distributed to each C_i by:

$$k_i = k\frac{|C_i|}{|V|} \tag{2}$$

Based on the assumption of 4.1, we can also assume that the community members are more tightly connected with each other than the nodes outside of the community.

Note that, by pre-defining the notion of fairness, we can still preserve the submodularity of the set. Thus, we will use an advanced greedy algorithm to approxi1mate our solution:

Algorithm 2 Naive approach for seed fairness

With a seed constraint k, distribute seeds k_i to each community C_i, such that $k_i = k\frac{|C_i|}{|V|}$. This guarantee the number of seeds will be fairly allocated to the group w.r.t the group size.
$S \leftarrow \emptyset$
$|S_i| \leftarrow$ number of seeds belongs to C_i in S
while Not all $|S_i| = k_i$ **do**
  Find the marginal vertex $v_i \in G$ (the vertex gives the most marginal gain), W.O.L.G, let $v_i \in C_i$
  if $|S_i| = k_i$ **then**
    Ignore v and continue
  else
    Add v into S, $|S_i| = |S_i| + 1$
  end if
end while

However, the solution given by this algorithm is still hard to compare with the optimal solution, and even this algorithm follows the submodularity, due to our seed number constraint, it still does not guarantee every node added to S is considered to be "good".

For example, we could add a vertex $v_i \in C_i$ to S, but the most vertex v_i influenced are from community C_j. In this case, to reach a marginal vertex $v_j \in C_j$, we might need to ignore lots of sub-optimal vertices and reach the very end of our enumerating process due to the effect of v_i. In this case, we will consider v_j to be a bad node.

To avoid this condition, we will do a backward subs-traction for our set S:

Algorithm 3 Algorithm for Seed fairness

With a seed constraint k, distribute seeds k_i to each community C_i, such that $k_i = k\frac{|C_i|}{|V|}$. This guarantee the number of seeds will be fairly allocated to the group w.r.t the group size.
$S \leftarrow \emptyset$
$|S_i| \leftarrow$ number of seeds belongs to C_i in S
while Not all $|S_i| \geq k_i$ **do**
  Find the marginal vertex $v_i \in G$ (the vertex gives the most marginal gain), W.O.L.G, let $v_i \in C_i$
  Add v into S, $|S_i| = |S_i| + 1$
end while
while Not all $|S_i| = k_i$ **do**
  Find the reverse marginal vertex $v_i \in G$ (the vertex gives the least marginal loss), W.O.L.G, let $v_i \in C_i$
  if $|S_i| = k_i$ **then**
    Ignore v and continue
  else
    remove v into S, $|S_i| = |S_i| - 1$
  end if
end while

In this case, since the edge condition above could be avoided, and since the set still preserves the submodularity, we are able to reach a fairly good approximation w.r.t the optimal solution. Intuitively, we may be able to bound the marginal loss during the back-deleting process.
Cons: By using this notion of seed fairness, it is hard to approximate the real activated node ratio (w.r.t each community) through our solution.

4.3 Attempt: Maximize the total influence w.r.t the activated node fairness

Not done yet!
Objective: In this section, to improve the previous section, we are trying to make an attempt to maximize both the global influence w.r.t the activated node fairness notion. However, we need to mention the trade-off, such that we need to pre-define the notion of fairness to simplify the problem. Otherwise, with dynamically checking and fixing the fairness of our result, it will be hard for us to maximize the overall influence. **Notion of fairness:**

$$Maxmin(A) = min_i \frac{I_{G,C_i}(A)}{|C_i|} \tag{3}$$

We might switch to other definitions which could facilitate our approximation of fairness.
The algorithm comes from the reverse reachability algorithm ([TSX15]).
Sketch of the algorithm:
1. We pre-calculate some reverse reachable sets for arbitrary vertices. (May need to remove some outlier sets, especially the set with very small reachable nodes.)

Algorithm 4 Algorithm for Weighted RR-Set

With an influence maximization problem on a graph, we defines a weighted RR-Set algorithm to achieve a effect of gap reduction while maximizing the total influence among the graph.

Given:

1) An graph $G = (V, E)$, in this algorithm is undirected.

2) A finite set of communities $\{C_1, C_2, ..., C_x\}$, such that each node in G participate in at least 1 of the communities.

I. Define the first weight, the **Community size weight** $w1$ by: **Foreach** Community C_i, define its weight as $w1_i = \frac{|C_i|}{|V|}$.

II. For the second weight, the **connectivity weight** $w2$, we first apply the notion of **Diversity Constraint** to each of the r-hop bounded Communities C_i^r, such that we calculate the optimal ki seeds on C_i^r that gives a maximal influence on the community C_i (ki $\approxeq k * w1_i$). Then, we find the maximum K-Core (or K-Truss) on C_i^r, such that it contains all the ki optimal seeds. **Define** the second weight $w2_i$ as, **foreach** community C_i, let its weight $w2_i = K$

III. Apply the random root selection with reverse reachable path on the graph G (similar to the RR-Set algorithm) to obtain the RR-Sets.

Then, for each of the RR-Set with the root v_r, if v_r belongs to some communities $C_{i1}, C_{i2}, ..., C_{ih}$, we set its weights to be $w1_{vr} = min_{1 \leq j \leq h} w1_{ij}^{\alpha 1}$, $w2_{vr} = min_{1 \leq j \leq h} w2_{ij} \alpha 2$, which $\alpha 1$ and $\alpha 2$ are the hyper-parameters used to control the trade-off between IM and fairness, which can be defined by the user them-selves.

IV. Finally, we do the greedy weighted set cover problem on those weighted RR-Sets with a constraint of k seeds.

2. For each trialed set, we pre-assign it a community belonging (label), which is the community that the majority of the vertices in the set belong.
3. With these labeled sets, hopefully, we might be able to achieve some approximation by adding some constraint to the LP of the set cover problem (which diverges...), weighted set cover problems (set the weight to be the inverse community size ratio) or some other greedy algorithms.

5 Current status

We've researched different authors in influence maximization fairness and fairness definition in a different academic setting. When we will continue on section 4.3 to find a feasible solution to our proposed problem.

6 Datasets

Currently, we sampled datasets such as Twitter's network dataset, and others from Kaggle. Based on our problem setting, we are still unsure if the available dataset will fit our problem. We will keep iterating on our algorithm approach. And we might create our synthetic dataset if necessary.

7 Issues and Problems

As discussed above, we are currently facing a trade-off between submodularity and the notion of fairness, such that without pre-define the constraints of fairness. We need to dynamically check and fix the fairness of our result within our iterative process. This brings hardness to maximize the global influence of the graph.
We found a paper that discussed how to maximize the objectives of multiple submodular functions ([CC21]), but due to the time limit, we are not finishing reading it.

References

[CC21] C Feng, C Qian. *Multi-Objective Submodular Maximization by Regret Ratio Minimization with Theoretical Guarantee.* AAAI Conference on Artificial Intelligence, 2021.

[KR08] A. Krause, B. McMahan, C. Guestrin, and A. Gupta. *Robust submodular observation selection.* Jour. of ML Research, 9(Dec):2761–2801, 2008.

[TSX15] ouze Tang, Yanchen Shi, Xiaokui Xiao *Influence Maximization in Near-Linear Time: A Martingale Approach.* ACM SIGMOD International Conference on Management of Data. Pages 1539–1554. 2015.

[TZ19] T.Alan, Tsang, A.; Wilder, B.; Rice, E.; Tambe, M.; and Zick, Y. *Group-Fairness in Influence Maximization.* In 28th International Joint Conference on Artificial Intelligence, 5997-6005. 2019.

[RA20] Rahmattalabi, Aida, et al. *Fair influence maximization: A welfare optimization approach.* arXiv preprint arXiv:2006.07906, 2020.