FairMIA: The heuristic based Fair Influence Maximization

SASIKANTH KOTTI*, MT19AIE308
NIKHILA DHULIPALLA†, MT19AIE270
ADHUN THALEKKARA‡, MT19AIE205

Abstract: As a model for spreading information across social networks, Influence Maximization is gaining traction. This issue has real-world implications that might have an impact on people's lives. In such contexts, algorithmic decision making raises questions regarding societal ramifications. One of these problems, algorithmic bias and fairness, has surprisingly garnered little attention thus far. This issue is important in studying how information spreads in a community of networks, with application areas like marketing, news dissemination, vaccination, and generating online trends. Because historical prejudices may be encoded in human networks, algorithms that use them to automate outcomes may capture and recreate such biases. In this paper, we propose a heuristic-based algorithm for fair influence maximization, called FairMIA (fair-maximum influence arborescence). Ours is the first attempt to explore heuristics in the fair influence maximization problem. The experimental evaluation shows that our algorithm performs satisfactorily via the 'Price of Fairness', 'Fairness Score' and 'Maximin' metrics. We have also developed a model to construct synthetic networks to evaluate fair influence-maximization.

Keywords: *Influence Maximization, Fairness, Social Networks, Synthetic networks*

1 INTRODUCTION

People and their relationships make up a social network. is one of the well-studied issue in online networks and mostly used in commercial propagation's, news dissemination, viral contents in the online, disease transmission, as well as a variety of other areas. The network structure will play a prominent role in the effective dissemination of an idea, technology, or product across a social network: who will possess the information and who they will be related to decides who that information will be reaching. As a result, being a 'early-adopter,' or someone who accepts a product of their own will or gets it for free, is a privileged position because of their social standing and direct effect on the diffusion process. Early adopters spread the word about the contents to their friends, who may or may not embrace it and continue the cycle, culminating in a cascade. Chen et al. [3] People will usually trust the information significantly from their close friends or relatives compared to the information spread from wide advertisement channels such as television, newspapers, and web advertisements, according to research. As a result, many individuals feel that word-of-mouth marketing is the most successful kind of promotion.

Recent work has proposed several definitions of fairness, including maximin fairness Rahmattalabi and Tambe [21] and diversity constraints, to directly embed fairness into influence maximization Tsang et al. [23]. Maximin fairness tries to increase any community's minimal level of influence. Diversity constraints, on the other hand, are based on game theory and ensure that each community is at least as well-off as if they were given a proportional share of resources and distributed internally. Each of these ideas provides a different viewpoint on fairness. They do, however, have disadvantages. Due to its strict need to support the worst-off group as much as possible, maximin fairness, for example, might result in a considerable reduction in overall influence, even if it may be difficult to transmit the influence to other groups due to their scarce connections. While the diversity limitations seek to account for the community's capacity to disseminate influence, they do not explicitly account for the reduction of inequality.

The first contribution to fair influence maximization was initially proposed by Tsang et al. [23]. It begins with a maximin idea inspired by the legal concept of disparate effect, which mandates to maximise the smallest proportion of impacted nodes within each group. While the conventional influence maximization issue is

*Student 1 designed the experiments, performed experiments and wrote 10% of the report.
†Student 2 performed result analysis and wrote 80% of the report.
‡Student 3 performed experiments and wrote 10% of the report.

submodular; nevertheless, fairness considerations result in substantially non-submodular objectives. As a result, typical procedures are rendered useless. Algorithmic contribution is a novel approach for universal multi-objective submodular optimization that offers a significantly higher approximation guarantee and, in many cases, a shorter runtime. After this, several contributions were made, and discussed few of them in Section 3.

The independent cascade model is NP-hard when it comes to spreading influence. It suggests that the greedy method presented previously may have inherent challenges in scaling to big networks. Chen et al. [3] The scalability problem is addressed by presenting a novel heuristic approach that is many orders of magnitude quicker than previous greedy algorithms while maintaining the greedy algorithms' influence spread. The basic concept behind the heuristic technique is to estimate influence propagation using local arborescence structures of each node. The influence spread in the MIA model is submodular, i.e. has a diminishing marginal return property, a simple greedy algorithm that selects one node with the highest marginal influence spread in every round could indeed assure an influence spread within (1-1/e) of the optimum in the MIA model, whereas any higher the ratio estimation is NP-hard. Because of its balanced efficiency and efficacy, the heuristic approach proposed may be used as a generic solution for influence maximization in a variety of large-scale online social networks.

The report is organized as follows: Section 2 describes the problem statement formally and Section 3 reports the related work. Sections 4 and 5 describe our proposed algorithm and experimental results. Finally, Section 6 concludes the research outcome.

2 PROBLEM STATEMENT

Influence maximization is predominantly used to perform viral marketing and spreading information about brands etc. There are a variety of methods, such as greedy algorithm and heuristics for influence maximization. The general influence maximization objective doesn't take into consideration any fairness aspects. In general social networks also consists of minority communities. Hence it is essential to ensure that interventions targeted towards the upliftment of the population need to reach out even to the minority communities. This is even more important for interventions related to health, education and others. In such scenarios, the usual influence maximization may not ensure a fair spread.

There are many techniques proposed to address the Fairness Influence Maximization, but ours is the first attempt by making use of heuristics in the fair influence maximization problem.

3 RELATED WORK

Friends endorse an items to their common friends or mutual connections, who then suggest it to their friends, and the chain goes on by making it as a cascade of recommendations, according to viral marketing study. The item's marketing can grow from a finite count of initial nodes to a presumably much larger group using this method.

Viral marketing on social media was initially framed as an influence maximization issue by Domingos and Richardson [8]. They used Markov random fields to describe the problem and suggested heuristic solutions. Influence maximization is studied as an optimization problem by Kempe et al. [13] and in both the Independent Cascade (IC) and Linear Threshold (LT) Models, they argued that the issue is NP-hard, and they introduced a greedy method to solve it. Leskovec et al. [16] To improve the efficiency of greedy algorithms, it was discovered that not all rest of the nodes required to be assessed in every round, and CELF was provided as a solution. CELF++, which has been demonstrated to run 35 percent to 55 percent quicker than CELF, was also proposed by Goyal A. [11]; CELF and CELF++ both rely on submodularity. Chen [4] looked at optimal influence maximization from two different perspectives. The first is to present New Greedy, which enhances on the initial greedy algorithm, and the second is to provide Degree Discount heuristics, which enhances the diffusion of influence. Also created a number of heuristic techniques for calculating influence spread. The predicted number of other nodes effected by adding node v to the seed set is assessed using Degree Discount based on v's one hop neighbourhood. Chen Y. [7] introduced a novel frame-work called community-based influence maximization (CIM) to address the influence maximization problem with a focus of minimizing downtime. To generate the greatest influence set under the IC and LT models, two approximation techniques, PMIA and LDAG, are introduced in Chen et al. [3] and Chen W. [5]. It has been demonstrated in LDAG that evaluating influence spread in a DAG has a linear time difficulty under the LT model. The technique on the local DAG generation is offered to minimise time complexity.

The time-critical impact maximization issue was introduced independently by Chen W. [6] and Liu [19] in which the propagation model takes into account the time constraint. They demonstrated the time restricted influence spread function's monotonicity and submodularity, and proposed methodology to address the problem. According to Li [17] in more real-world situations, marketers frequently target certain items at specific customer segments. As a result, they introduced the labelled influence maximization approach, which tries to locate a collection of seed nodes in a labelled social network that may cause the greatest spread of influence on the target consumers. Nguyen [20] have introduced the budgeted influence maximization problem (BIM), which entails choosing a set of seed nodes to maximise the total number of nodes impacted in social networks at a total cost no greater than the budget. Influence maximization was investigated by Goyal [10] from a novel data-based paradigm. They introduced a new model called credit distribution that uses the existing propagation traces to understand how influence spreads in a network and uses that information to predict probable influence spread. Using mobile crowd generated data, Li [18] suggested a new network model and influence propagation model that considers influence spreading in both online social networks and the real-world. Chen [2] developed an addition to the Independent Cascade Model (ICM) that takes into account the formation and spread of negative attitudes, which is well-known in the social psychology literature. Recently, Guo [12] delved into a new fascinating challenge of maximising social network influence. Given a target user w, the goal is to determine the top-k most influential nodes for the user. To compute positive influence, Wang [24] develop IMIC-OC, an independent cascade-based model for influence maximization.

In recent times Group-Fairness in Influence Maximization Tsang et al. [23] was proposed to ensure fairness for the spread. This is also extensively studied in the context of feature-awareness by Stoica and Chaintreau [22]. The idea of using randomization to obtain fair influence maximization is proposed and studied by Becker et al. [1]. Khajehnejad et al. [15] utilized adversarial graph embeddings to obtain Fair Influence Maximization over Social Networks. These graph embeddings are obtained by training graph neural networks with adversarial optimization methods.

Similarly Khajehnejad et al. [14] addressed influence maximization using a generic method of CrossWalk. Farnadi et al. [9] solved the problem of fair influence maximization by modelling fairness with constraints and formulating the problem as an integer programming problem. However, so far none of the works attempted to model fairness constraints with heuristics for fair influence maximization.

4 MAIN SECTION OF OUR CONTRIBUTION

In our current work, we attempted to address the problem of fair influence maximization by defining fairness with heuristics.

Following are our contributions :

(1) We proposed a new heuristic based algorithm for fair influence maximization, called FairMIA (fair maximum influence arborescence) which is inspired from MIA algorithm.
(2) We demonstrated that our algorithm performs satisfactorily via the 'Price of Fairness', Fairness Score' and 'Maximin' metrics.
(3) We proposed new 'Fairness Score' metric for determining how equitably a particular attribute has been influenced.

(4) We proposed a model to construct synthetic networks to evaluate fairness influence maximization.

4.1 IC model with greedy algorithm

Edge labels dp: E [0, 1] in a directed graph G = (V,E) and dp(u, v) signifies the edge's diffusion probability. In this probability v gets activated by u across the edge in the next step after u is active, for any edge (u, v) E. Let's focus at the set of nodes that have been activated at a time step that is greater than zero. As a result, every node will activate its out-neighbors with an independent probability of diffusion probability at the next iterative time step (u,v). When all of the active nodes have been saturated, the process draws to a close. Each active node has just one possibility to activate its out-neighbors in the time step immediately after its activation, and each node remains activated after that.

The function has a declining marginal return so, the non-negative real values function on subsets of edges is submodular. Until k seeds are obtained, the greedy algorithm repeatedly picks new seeds that maximises the incremental change of function (submodular and monotone) into the seed set. As a result, the Greedy algorithm solves the influence maximization issue with a 1 - 1/e approximation ratio.

The below algorithm is taken from Chen et al. [3].

Algorithm 1 Greedy Algorithm, $Greedy(k, f)$

```
Initialize seed set S with ∅
for each i = 1 ... k do
    choose u = argmax_{w∈V\S}(f(S ∪ {w}) − f(S))
    S = S ∪ {w}
end for
return seed set S
```

4.2 MIA model, MIIA and MIOA algorithms

Activating all the nodes along the path say P then the probability that u activates v via path P is dp(P). Chen et al. [3] suggested using the maximum influence path (MIP) to assess the influence from one node to another in and around the social network to near the actual expected influence. Diffusion probability is funnelled to a distance weight on every edge in the graph in this case and in the weighted graph, MIP may be considered of as one of the shortest path from u to v. As a result, maximum influence paths and subsequently maximum influence arborescences are straightforward equivalents of shortest paths and shortest-path arborescences, enabling effective methods like the Dijkstra algorithm to generate them.

Maximum Influence In Arborescene (MIIA) is the aggregate of the maximum influence connections to v in order to calculate the influence of many other nodes within the network on v. In similar way, Maximum Influence Out Arborescene (MIOA) is a comparable technique for assessing v's influence on many other nodes. Also as we continuously maintain a relationship in maximum influence paths, the combination of maximum influence paths to a node seems to have no undirected cycles, demonstrating that it is an arborescence. MIIA and MIOA, in other words, will provide local influence regions of v.

We approach the IC model by assuming that the influence from seeds to edges is solely transmitted across edges in MIIA while assessing the set of seeds in a graph as well as the arborescence in MIIA. With this approximation, we can calculate the probability that edges is activated given seeds exactly. We can determine the probability that edges will be activated with seeds precisely while using approximation. The activation probability of each node in MIIA is to represent the probability that the node will indeed be activated whenever the seed set and influence are passed in MIIA. The activation probability is calculated iteratively using the algorithm below taken from Chen et al. [3].

Algorithm 2 Activation probability, $ap(u$, seed set $S, MIIA(v, \emptyset))$

if $u \in$ seed set S **then**
  $ap(u) = 1$
else if $N^{in}(u) = \emptyset$ **then**
  $ap(u) = 0$
else
  $ap(u) = 1 - \Pi_{w \in N^{in}(u)}(1 - ap(w).dp(w, u))$
end if

4.3 Computing Linear coefficients

We can now infer the recursive calculation of linear coefficients from the recursive computation of activation probability in Algorithm 2. It may be turned into an iteration way, with one traversal through MIIA from the root to the leaves computing all linear coefficients. It's vital for calculating a node's incremental influence spread. Consider the MIIA of size t with a given seed set S yet again and if we choose the next seed, the activation probability would increase from 0 to 1. The activation probability of the edge and the new seed chosen have quite a linear connection with the linear coefficient, implying that the new seed has incremental influence on the edge. As a result, only one run of MIIA is required to calculate the activation probability of the new seed, and a second run of MIIA is required to calculate the linear coefficients. This cuts down on the time it takes to compute MIIA's incremental influence spread for all nodes. The below algorithm is taken from Chen et al. [3].

Algorithm 3 Compute $\alpha(v, u)$ with $MIIA(v, \theta)$ and seed set S, after $ap(u$,seed set $S, MIIA(v, \theta))$ for all u in $MIIA(v, \theta)$ are known.

/* recursive algorithm */
if $u = v$ **then**
  $\alpha(v, u) = 1$
else
  Assign w as out-neighbor of u
  if $w \in$ seed set S **then**
    $\alpha(v, u) = 0$
  else
    $\alpha(v, u) = \alpha(v, w).dp(u, w).\Pi_{u' \in N^{in}(w) \setminus \{u\}}(1 - ap(u').dp(u', w))$
  end if
end if

4.4 Fair-Maximum Influence Arborescence

Inorder to account for fair influence maximization the original Maximum Influence Arborescence is modified to include the fairness notion. A defined by Tsang et al. [23] we included the maximin fairness notion in our algorithm. We are calling this algorithm as FairMIA. To the best of our knowledge this is the first time in which the fairness notion for influence maximization is attempted using heuristic based approach. We briefly outline the steps for the algorithm below :

Algorithm 4 FairMIA, $FairMIA(G, k, \theta)$

/* initialization */
Set seed set $S = \emptyset$
set $IncInf(v) = 0$ for all nodes $v \in V$
for each node $v \in V$ **do**
    compute $MIIA(v.\theta)$ and $MIOA(v, \theta)$
    set $ap(u, S, MIIA(v, \theta)) = 0, \forall u \in MIIA(v, \theta$ /* since $S = \emptyset$*/
    compute $\alpha(u, v), \forall u \in MIIA(v, \theta$ (Algo.3)
    for each node $u \in MIIA(v, \theta)$ **do**
        $IncInf(u) += \alpha(v, u).(1 - ap(u, S, MIIA(v, \theta)))$
    end for
    compute subgraph G where node attribute value $\in v$
    compute d dijkstra path between v and random node of subgraph G
    compute subgraph H where $node \in G$ and $v$ompute $MIIA(v, H, \theta)$
    for each node $w \in MIIA(v, H, \theta)$ **do**
        $IncInf(w) += \alpha(v, w).(1 - ap(w, S, MIIA(v, H, \theta)))$
    end for
end for
/* main loop */
for $i = 1 \ldots k$ **do**
    pick $u = argmax_{v \in V \setminus S}\{IncInf(v)\}$
    /* update incremental influence and fair influence spreads*/
    for $v \in MIOA(u, \theta) \setminus S$ **do**
        /* subtract previous incremental influence and fair influence spreads*/
        for $w \in MIIA(v, \theta) \setminus S$ **do**
            $IncInf(w) -= \alpha(v, w).(1 - ap(w, S, MIIA(v, \theta)))$
        end for
        compute $MIIA(v, H, \theta)$
        for each node $w \in MIIA(v, H, \theta)$ **do**
            $IncInf(w) -= \alpha(v, w).(1 - ap(w, S, MIIA(v, H, \theta)))$
        end for
    end for
    $S = S \cup \{u\}$
    for $w \in MIOA(u, \theta) \setminus S$ **do**
        compute $ap(w, S, MIIA(v, \theta)), \forall w \in MIIA(v, \theta)$ (Algo. 2)
        compute$\alpha(v, w), \forall w \in MIIA(v, \theta)$ (Algo. 3)
        /* add new incremental influence and fair influence spreads */
        for $w \in MIIA(v, \theta) \setminus S$ **do**
            $IncInf(w) += \alpha(v, w).(1 - ap(w, S, MIIA(v, \theta)))$
        end for
        compute $MIIA(v, H, \theta)$
        for each node $w \in MIIA(v, H, \theta)$ **do**
            $IncInf(w) += \alpha(v, w)(1 - ap(w, S, MIIA(v, H, \theta)))$
        end for
    end for
end for
return S

Initialization :

- Initially the incremental influence for all the nodes in the graph is zero
- For each node in the graph, obtain the MAXIMUM INFLUENCE IN-ARBORESCENCE structure and calculate the activation probability
- Utilizing the activation probability calculate the incremental influence
- After the above steps are completed for all nodes in the graph, generate a sub-graph consisting of nodes having the attributes, belonging to respective group
- For each node in the graph, span a sub-graph consisting of the current node and the previously generated sub-graph
- As there cannot be a direct path between the current node and the any random node of the sub-graph related to respective group, the path for connection is obtained by dijkstra shortest path algorithm
- For each node in the modified sub-graph, obtain the MAXIMUM INFLUENCE IN-ARBORESCENCE structure with influence threshold as zero and calculate the activation probability
- Utilizing the activation probability calculate the incremental influence
- repeat the above two steps for all nodes in the graph.

Main Loop :
Subtract incremental influence and fair influence spreads :

- Iterate for k times, where k is the number of seeds needed for fair Influence maximization
- Select the node that has the maximum incremental influence
- Obtain MAXIMUM INFLUENCE OUT-ARBORESCENCE structure based on this selected node
- For each node in this MAXIMUM INFLUENCE OUT-ARBORESCENCE structure , subtract the previous incremental influence
- Also compute MAXIMUM INFLUENCE OUT-ARBORESCENCE structure with influence threshold as zero, on this selected node utilizing the modified sub-graph. This modified sub-graph is obtained using this selected node and the nodes related to respective group.
- For each node in this MAXIMUM INFLUENCE OUT-ARBORESCENCE structure , subtract the previous incremental influence. This incremental influence is obtained by only considering the sub-graph and not the entire graph.
- Add the selected node to the seed set

Add new incremental influence and fair influence spreads :

- For each node in this MAXIMUM INFLUENCE OUT-ARBORESCENCE structure , Add the previous incremental influence
- Also compute MAXIMUM INFLUENCE OUT-ARBORESCENCE structure with influence threshold as zero, on this selected node utilizing the modified sub-graph. This modified sub-graph is obtained using this selected node and the nodes related to respective group.
- For each node in this MAXIMUM INFLUENCE OUT-ARBORESCENCE structure , Add the previous incremental influence. This incremental influence is obtained by only considering the sub-graph and not the entire graph.
- After completing k iterations, return the seed set S

4.5 Barabasi Albert Attribute Model Generator

The Barabasi Albert Model is a straightforward approach for generating scale-free networks. These are extensively utilised as they closely resemble real-life social networks. In this paper, we present an addition to the Barabasi Albert that combines user-defined subgroups or selections such as '25-30,'40-50,'male,'female,' with user-defined attribute values such as Age, Gender, Ethnicity. In the standard Barabasi Albert model, we first generate a random network with m nodes, then add m nodes in every iteration, with every node tending to link with high degree nodes until the total number of nodes reaches n. We used the same method and incorporated attributes; the difference is that when we create new m nodes, we will have to choose weighted choices (user defined weights) at random, and the node would have a slightly greater propensity to connect with nodes in the same attribute group, based on the generated attribute group. This propensity is also influenced by a variable known as the dependence index, which is indeed selected by the user during the initialization phase. We designed a class object where users may add each attribute, their options, the value of each option, and the dependency index. After adding all of them, the user could use the generategraph graph method to get a networkx graph.

Algorithm 5 Barabasi Albert Attribute Model Generator

```
Generate a random star graph with m number of nodes
Initialize the nodes with some sample attributes as given by the user
Initialize a list called repeatedItems
for each node in graph do
    add degree times the node to repeatedItems list
end for
source=length of the graph
while source numberOfNodes do
    initialize new key source to the dictionary graphAttribute
    initialize a blank array called repeatedItems
    for each Attribute in Attributes do
        generate a random choice for the attribute of the node
        extend repeatedItems list with dependencyIndex times the node
    end for
    remove the current handling nodes from repeatedItems to avoid self loops
    Add repeatedItems to the repeatedNodes
    get a random subset of repeatedNodes and form connection
    remove the repeatedItems from the repeatedNodes list
    extend the repeatedNodes with the new subset
    source= source+1
end while
```

4.6 Fairness Metrics

4.6.1 Fairness Score. The fairness score is a new metric we've developed for determining how equitably a particular attribute has been influenced. This fairness score ranges from 0 to 1, with 1 indicating the highest fairness and 0 indicating no fairness in the influence maximization approach. The percentage of influence nodes in a group of the attributes is not related to the fairness score. Let's imagine we have two groups of attributes, and only 10 percent of the nodes within every group are influenced; in this case, their fairness score will be one. Even if 90 percent of total of the nodes in the group are influenced, the fairness score will stay at 1. However, if one group has a 10percentage influence and another has a 90percentage influence, then fairness score will be remarkably low. As a result, when you combine the fairness score with the percentage of maximising, you'll have a clear view of the fair influence maximization. To calculate fairness score we make use the below given equation:

$$\text{Fairness score} = \left[\sum_{i=0}^{n} \frac{\left| \mu - \frac{numberOfNodes_i}{numberOfNodesInfluenced_i} \right|}{\mu * n} \right]$$

where;

n is the number of groups in the attributes

numberOfNodes is the total number of nodes in the group i

numberOfInfluencedNodes is the total number of influenced nodes after influence maximization in the group i

and μ is the mean of percentage of influence in each group given by:

$$\mu = \left[\sum_{i=0}^{n} \frac{\frac{numberOfNodes_i}{numberOfNodesInfluenced_i}}{n} \right]$$

Algorithm 6 Fairness Score Calculation

Input
  n:list of number of nodes in each category
  nInfluenced: list of number of nodes influenced in each category
Output
  returns fairness score
$numberOfGroups \leftarrow$ length of n
$percentageOfInfluence \leftarrow$ new list formed by $nInfluenced$ divided by n
Find the mean of percentageOfInfluence
create list called deviation with the distance from mean to percentageOfInfluence
sum the deviation list and devide by mean value
return $(1 - deviation/numberOfGroups)$

4.6.2 Maximin Fairness. Tsang et al. [23] Maximin Fairness encapsulates the simple objective of enhancing the lives of the poorest population. That is, aim to maximise any group's minimal influence as a fraction of its population.

$$U^{Maximin}(A) = min_i \frac{I_{G,C_i}(A)}{|C_i|}$$

$I_G^{Maximin}$ is the predicted set of nodes activated by a seed configuration that optimises the proportionate influence gained by any group with the least level of influence.

4.6.3 Price of Fairness. Tsang et al. [23] calculated the Price of Fairness, which would be the ratio of optimal influence to best achievable influence, to determine the cost of guaranteeing a reasonable outcome for the diversified community.

$$PoF^{Maximin} = \frac{I^{OPT}}{I^{Maximin}}$$

5 EXPERIMENTS AND RESULTS

5.1 Data Collection

We utilized two datasets or social networks in the experiment. The Antelope Valley network from the paper Tsang et al. [23] was the first one we used. The synthetic dataset was generated using our customized attribute-incorporated social network generator, which is an expanded version of the Barabasi-Albert method. One may create any social network with different attributes and choices using this customized algorithm, as well as determine the ratio of attribute choices (for example, US:10 India:90) in the population and the attribute's dependency index (which says how much a node tend to make bond with the same attribute).

5.2 Experimental Setup

We have two datasets on which we have conducted the experiments. To acquire the networkx graph for the Antelope Valley network, we have imported the pickle data from the dataset publication. Our own BA enhanced approach also generates a networkx graph from a synthetic dataset. To create the synthetic dataset, we used the following parameters given in Table 1 to add three option attributes: gender, religion, and age. Then, for each dataset, assign a random integer between 0 and 0.1 to the edge probability for each node in the dataset. For each dataset, we ran MIA, FairMIA, and showed the price of fairness. We also discovered the lowest fractional influence with MIA and a fair MIA.

The definition of Price of Fairness, requires computing optimal influence and influence with fairness. To obtain this values, IC model is employed. 10000 simulations were executed on the IC model to obtain optimal influence and fair influence. The seeds for the optimal influence are obtained by executing MIA algorithm on the graph where seeds for fair influence are obtained by executing fairMIA algorithm defined in the Section 4. A total of 15 seeds were considered and influence threshold was set to 0.01.

Table 1. Parameters considered for our Synthetic dataset

Attribute Name	Attribute Type	Choices	Weights	Dependency Index
Gender	Choice	Male, Female	50, 50	3
Region	Choice	India, US	90, 10	2
Age	Choice	20-25, 50-59	60, 40	1

5.3 Results

Below are the results of our extensive experiments conducted using Antelope Valley network dataset and Sythetic dataset.

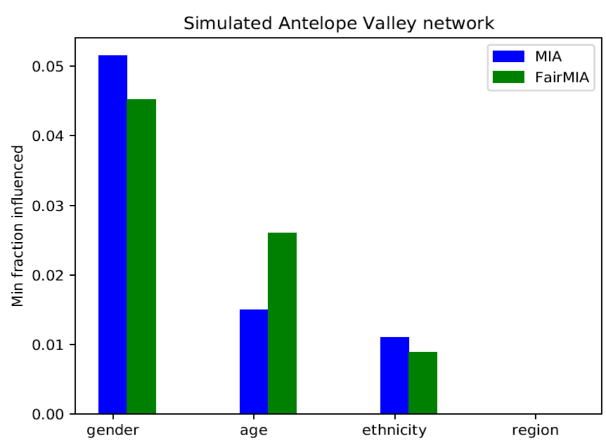

Fig. 1. Simulated Antelope Valley network vs Min Fraction Influenced. As part of our experiments we have considered four attributes - gender, age, ethnicity and region. Here, the proposed Fair MIA performed well on Age attribute and comparatively good on Gender and Ethnicity attributes. But Region attribute is at zero as there was no fairness or MIA detected.

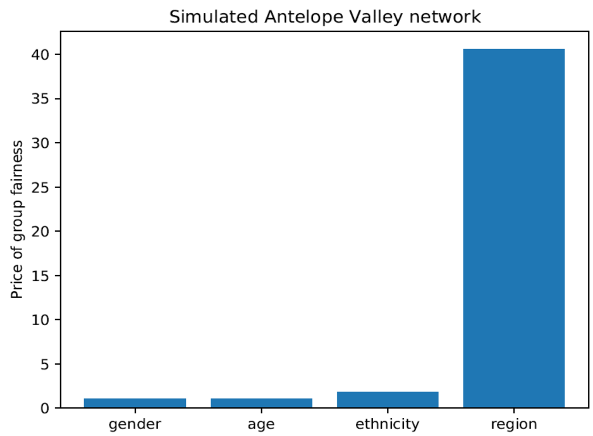

Fig. 2. Simulated Antelope Valley network vs Price of group fairness. As part of our experiments we have considered four attributes - gender, age, ethnicity and region. Here, the PoGF is high for Region attribute compared to other attributes as optimal influence obtained was near to zero so the cost of guaranteeing a reasonable outcome for the diversified community is more.

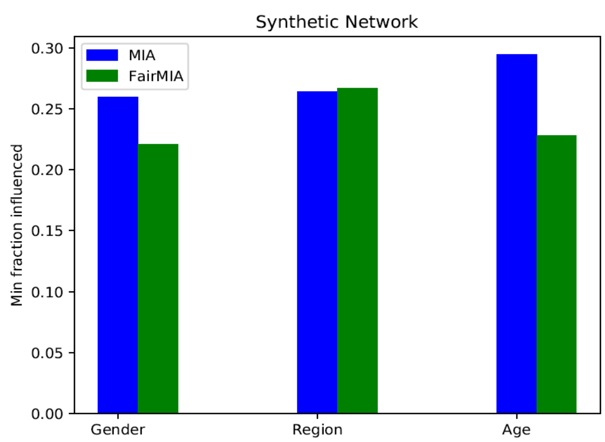

Fig. 3. Synthetic network vs Min Fraction Influenced. As part of our experiments we have considered three attributes - gender, age and region. Here, the proposed Fair MIA performed well on Region attribute and comparatively good on Gender and Age attributes.

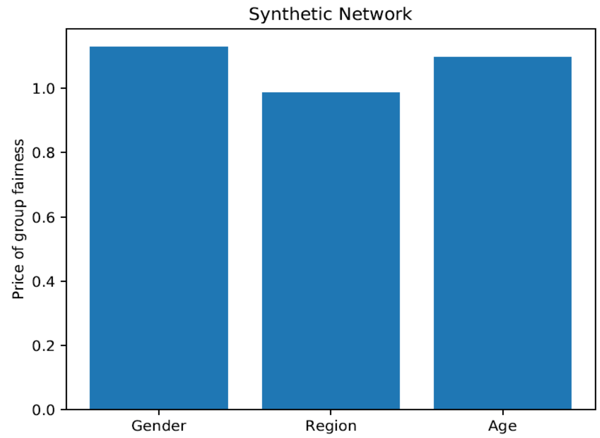

Fig. 4. Synthetic network vs Price of group fairness. As part of our experiments we have considered three attributes - gender, age and region. Here, the PoGF is high for Gender attribute compared to other attributes as optimal influence obtained was less comparatively so the cost of guaranteeing a reasonable outcome for the diversified community is more.

Table 2. Results of the proposed Attribute Fairness Score on datasets used in the proposed Fair MIA

Attribute Name	Simulated Antelope Valley network	Synthetic network
Gender	0.8457	0.7675
Age	0.5892	0.7975
Region	-0.0361	0.9315
Ethnicity	0.5251	-

```
Results on Simulated Antelope Valley network :
Attributes of the graph are : {'gender': ['male', 'female'], 'age': ['30-39', '50-59', '18-24', '65+', '25-29', '40-49', '60-64'], 'ethnicity': ['other', 'asian', 'black', 'white', 'latino'], 'region': ['desert_view_highlands', 'lake_los_angeles', 'quartz_hill', 'littlerock', 'acton', 'palmdale', 'northwest_antelope_valley', 'lancaster', 'leona_valley', 'sun_village', 'northwest_palmdale', 'northeast_antelope_valley', 'southeast_antelope_valley']}
Initialization Completed
Seeds for IM with MIA : [263, 17, 32, 268, 13, 319, 271, 28, 423, 265, 288, 464, 238, 155, 18]
Initialization Completed
Price of group fairness : {'gender': 1.0243039717433224}
Attribute fairness score : {'gender': 0.8457}
Min fraction influenced with fair MIA : {'gender': 0.04521595330739021}
Initialization Completed
Price of group fairness : {'gender': 1.0243039717433224, 'age': 1.082901458192911}
Attribute fairness score : {'gender': 0.8457, 'age': 0.5892}
Min fraction influenced with fair MIA : {'gender': 0.04521595330739021, 'age': 0.026018867924530786}
Initialization Completed
Price of group fairness : {'gender': 1.0243039717433224, 'age': 1.082901458192911, 'ethnicity': 1.7849592399553273}
Attribute fairness score : {'gender': 0.8457, 'age': 0.5892, 'ethnicity': 0.5251}
Min fraction influenced with fair MIA : {'gender': 0.04521595330739021, 'age': 0.026018867924530786, 'ethnicity': 0.008923076923076853}
Initialization Completed
Price of group fairness : {'gender': 1.0243039717433224, 'age': 1.082901458192911, 'ethnicity': 1.7849592399553273, 'region': 40.56411838912598}
Attribute fairness score : {'gender': 0.8457, 'age': 0.5892, 'ethnicity': 0.5251, 'region': -0.0361}
Min fraction influenced with fair MIA : {'gender': 0.04521595330739021, 'age': 0.026018867924530786, 'ethnicity': 0.008923076923076853, 'region': 0.0}
Price of group fairness : {'gender': 1.0243039717433224, 'age': 1.082901458192911, 'ethnicity': 1.7849592399553273, 'region': 40.56411838912598}
Min fraction influenced with fair MIA : {'gender': 0.04521595330739021, 'age': 0.026018867924530786, 'ethnicity': 0.008923076923076853, 'region': 0.0}
Min fraction influence with MIA : {'gender': 0.051510699588477385, 'age': 0.01503448275862024, 'ethnicity': 0.010976923076923169, 'region': 0.0}
/root/fairMIA/utils.py:78: UserWarning: FixedFormatter should only be used together with FixedLocator
  ax.set_xticklabels(keys)
Attribute fairness score with fair MIA: {'gender': 0.8457, 'age': 0.5892, 'ethnicity': 0.5251, 'region': -0.0361}
Results on Synthetic network G :
Attributes of the graph G are : {'Gender': ['male', 'female'], 'Region': ['India', 'US'], 'Age': ['20-25', '50-59']}
Initialization Completed
Seeds for IM with MIA : [9, 23, 24, 31, 76, 3, 12, 48, 88, 6, 99, 11, 34, 1, 60]
Initialization Completed
Initialization Completed
Initialization Completed
Price of group fairness : {'Gender': 1.12847983005883, 'Region': 0.9858330884429136, 'Age': 1.0975268222025953}
Min fraction influenced with fair MIA : {'Gender': 0.22098199999999943, 'Region': 0.26681666666666504, 'Age': 0.22808285714284954}
Min fraction influence with MIA : {'Gender': 0.2594280000000008, 'Region': 0.2640999999999981, 'Age': 0.2945184615384677}
/root/fairMIA/utils.py:78: UserWarning: FixedFormatter should only be used together with FixedLocator
  ax.set_xticklabels(keys)
Attribute fairness score with fair MIA: {'Gender': 0.7675, 'Region': 0.9315, 'Age': 0.7975}
```

Fig. 5. Overall results obtained for the experiments conducted

6 CONCLUSION AND FUTURE WORK

We demonstrated in our work to address the problem of fair influence maximization by defining fairness with heuristics. Ours was the first attempt to induce fairness with heuristics for influence maximization. We could successfully show that our proposed new heuristic-based algorithm for fair influence maximization performed comparatively better than MIA. Our experimental evaluations demonstrated that our proposed algorithm performs satisfactorily via the 'Price of Fairness', 'Fairness Score' and 'Maximin' metrics. The fairness score metric was a new attempt to determine how equitably a particular attribute was influenced. Our novel model constructed to generate synthetic networks for evaluating fair influence maximization has also produced a satisfactory result.

As a future work, the improvements to fairMIA algorithm can be proposed. Along with this additional notions of fairness can be investigated for influence maximization.

CODE

Code for the proposed algorithm and contributions, is provided in the following link.
https://anonymous.4open.science/r/fairMIA-83B6/README.md

ACKNOWLEDGEMENT

This work is in part supported by the Indian Institute of Technology - Jodhpur and we also would like to acknowledge Dr. Suman Kundu, PhD for the support of this project.

REFERENCES

[1] Ruben Becker, Gianlorenzo D'Angelo, Sajjad Ghobadi, and Hugo Gilbert. 2021. Fairness in Influence Maximization through Randomization. In *AAAI*.

[2] A.; Cummings R.; Ke T.; Liu Z.; Rincon D.; Sun X.; Wang Y.; Wei W.; Yuan Y Chen, W.; Collins. 2011. Influence maximization in social networks when negative opinions may emerge and propagate. *Proceedings of the SDM, Mesa, AZ, USA, ; pp. 379–390* (2011).

[3] Wei Chen, Chi Wang, and Yajun Wang. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (2010).

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

[5] Zhang L Chen W., Yuan Y. 2010. Scalable influence maximization in social networks under the linear threshold model. *Proceedings of the 10th IEEE International Conference on Data Mining, Sydney, Australia, pp. 88–97* (2010).

[6] Zhang N Chen W., Lu W. 2012. Time-critical influence maximization in social networks with time-delayed diffusion process. *Proceedings of the Conference on Artificial Intelligence, Toronto, ON, Canada, pp. 1–26* (2012).

[7] Peng W. Lee W. Lee S.Y. Cim Chen Y., Zhu W. 2014. Community-based influence maximization in social networks. *ACM Trans. Intell. Syst. Technol* (2014).

[8] P. Domingos and M. Richardson. 2001. Mining the network value of customers. *Proceedings of the 7th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 57–66* (2001).

[9] G. Farnadi, Behrouz Babaki, and Michel Gendreau. 2020. A Unifying Framework for Fairness-Aware Influence Maximization. *Companion Proceedings of the Web Conference 2020* (2020).

[10] F.; Lakshmanan L Goyal, A.; Bonchi. 2011. A data-based approach to social influence maximization. *Proc. VLDB Endow, 73–84* (2011).

[11] Lakshmanan L. Celf++ Goyal A., Lu W. 2011. Optimizing the greedy algorithm for influence maximization in social networks. *Proceedings of the 20th International Conference Companion on World Wide Web, Hyderabad, India, pp. 47–48* (2011).

[12] P.; Zhou C.; Cao Y.; Guo L Guo, J.; Zhang. 2013. Personalized influence maximization on social networks. *Proceedings of the 22nd ACM International Conference on Conference on Information Knowledge Management, San Francisco, CA, USA ; pp. 199–208* (2013).

[13] D. Kempe, J. M. Kleinberg, and É. Tardos. 2003. Maximizing the spread of influence through a social network. *Proceedings of the 9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 137–146* (2003).

[14] Ahmad Khajehnejad, Moein Khajehnejad, Mahmoudreza Babaei, Krishna P. Gummadi, Adrian Weller, and Baharan Mirzasoleiman. 2021. CrossWalk: Fairness-enhanced Node Representation Learning. *ArXiv* abs/2105.02725 (2021).

[15] Moein Khajehnejad, Ahmad Asgharian Rezaei, Mahmoudreza Babaei, Jessica Hoffmann, Mahdi Jalili, and Adrian Weller. 2020. Adversarial Graph Embeddings for Fair Influence Maximization over Social Networks. In *IJCAI*.

[16] J. Leskovec, A. Krause, C. Faloutsos C. Guestrin, J. VanBriesen, and N. S. Glance. 2007. Cost-effective outbreak detection in networks. *Proceedings of the 13th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 420–429* (2007).

[17] C.; Shan M Li, F.; Li. 2011. Labeled influence maximization in social networks for target marketing. *Proceedings of the International Conference on Privacy, Security, Risk and Trust, 10th IEEE International Conference on Social Computing, Boston, MA, USA, pp. 560–563* (2011).

[18] Z.; Yan M.; Li Y Li, J.; Cai. 2016. Using crowdsourced data in location-based social networks to explore influence maximization. *Proceedings of the 35th Annual IEEE International Conference on Computer Communications, San Francisco, CA, USA, ; pp. 1–9* (2016).

[19] G.; Xu D.; Zeng Y Liu, B.; Cong. 2012. Time constrained influence maximization in social networks. *Proceedings of the IEEE 12th International Conference on Data Mining, Brussels, Belgiuma, pp. 439–448* (2012).

[20] R Nguyen, H.; Zheng. 2013. On budgeted influence maximization in social networks. *IEEE J. Sel. Areas Commun, 1084-1094* (2013).

[21] Vayanos P. Fulginiti A. Rice E. Wilder B. Yadav A. Rahmattalabi, A. and M. Tambe. 2019. Exploring Algorithmic Fairness in Robust Graph Covering Problems. *Advances in Neural Information Processing Systems, 15750–15761* (2019).

[22] Ana-Andreea Stoica and Augustin Chaintreau. 2019. Fairness in Social Influence Maximization. *Companion Proceedings of The 2019 World Wide Web Conference* (2019).

[23] Alan Tsang, Bryan Wilder, Eric Rice, Milind Tambe, and Yair Zick. 2019. Group-Fairness in Influence Maximization. *ArXiv* abs/1903.00967 (2019).

[24] Y.; Lin Z.; Cheng S.; Yang T Wang, Q.; Jin. 2016. Influence maximization in social networks under an independent cascade-based model. *Phys. A Stat. Mech. Appl ; 444, 20–34* (2016).