

Adversarial Graph Embeddings for Fair Influence Maximization over Social Networks

Moein Khajehnejad^{1*}, Ahmad Asgharian Rezaei^{2*}, Mahmoudreza Babaei³,
Jessica Hoffmann⁴, Mahdi Jalili² and Adrian Weller^{5,6}

¹Monash University

²RMIT University

³Max Planck Institute for Software Systems

⁴The University of Texas at Austin

⁵University of Cambridge

⁶Alan Turing Institute

moein.khajehnejad@monash.edu, ahmad.asgharian.rezaei@rmit.edu.au, babaei@mpi-sws.org,
hoffmann@cs.utexas.edu, mahdi.jalili@rmit.edu.au, aw665@cam.ac.uk

Abstract

Influence maximization is a widely studied topic in network science, where the aim is to reach the maximum possible number of nodes, while only targeting a small initial set of individuals. It has critical applications in many fields, including viral marketing, information propagation, news dissemination, and vaccinations. However, the objective does not usually take into account whether the final set of influenced nodes is fair with respect to sensitive attributes, such as race or gender. Here we address *fair influence maximization*, aiming to reach minorities more equitably. We introduce Adversarial Graph Embeddings: we co-train an auto-encoder for graph embedding and a discriminator to discern sensitive attributes. This leads to embeddings which are similarly distributed across sensitive attributes. We then find a good initial set by clustering the embeddings. We believe we are the first to use embeddings for the task of fair influence maximization. While there are typically trade-offs between fairness and influence maximization objectives, our experiments on synthetic and real-world datasets show that our approach dramatically reduces disparity while remaining competitive with state-of-the-art influence maximization methods.

1 Introduction

A subfield of particular interest in network diffusion is influence maximization (IM), in which nodes and edges respectively represent individuals and connections. In its classical formulation, the goal is for the diffusion to reach the maximum number of nodes, while only disseminating the information to a few initial individuals, also called *seeds*.

Influence maximization is crucial in a variety of applications, including adopting new behavior in social networks [Richardson and Domingos, 2002; Kempe *et al.*, 2003], viral marketing [Babaei *et al.*, 2013], propagation of information, disease spread [Banerjee *et al.*, 2013; Mirzasoleiman *et al.*, 2012], and social recommendations [Ye *et al.*, 2012]. However, [Kempe *et al.*, 2003] showed that finding the best seeds is NP-hard. Numerous approximations and heuristics have been proposed. Several approaches identify the most influential nodes to maximize the total number of people influenced [Li *et al.*, 2018; Kourtellis *et al.*, 2013]. Using network structure, they consider central nodes (highest degree, closeness, betweenness, *etc.*) as the most influential nodes. Tackling the task from a different angle, [Kempe *et al.*, 2003] formulated the problem under the framework of discrete optimization. Their results significantly out-perform node selection heuristics, and showed that a provable approximate guarantee is obtainable. In this work, we adopt a fresh perspective: Adversarial Network Embeddings. We show our method is competitive with the previous influence maximization state-of-the-art, while significantly improving *fairness*.

Indeed, our objective is *fair influence maximization*, in which the resulting set of influenced nodes is diverse with respect to sensitive attributes, such as age, gender, race *etc.*. Adding fairness objectives to influence maximization yields an even harder task, which none of the previously mentioned techniques can tackle. However, these constraints are very relevant in today's society. For instance, consider the spread of a job opening, a loan advertisement, or even news. We want to make sure that belonging to a minority does not affect whether or not we would see this job opportunity or this loan offer. Moreover, as discussed in [Babaei *et al.*, 2018], receiving different types of news may be crucial to developing unbiased viewpoints. Making sure that everyone has a chance to see critical news, independently of the communities to which they belong, could be a stepping stone in the fight against fake news.

* Authors contributed equally to this work.

Fair influence maximization has been recently studied [Ali *et al.*, 2019; Tsang *et al.*, 2019; Fish *et al.*, 2019]. However, these methods suffer from a trade-off between fairness and the influence maximization objective. Using our Adversarial Network Embeddings, we achieve a highly diverse set of influenced nodes with respect to multiple attributes, while still achieving state-of-the-art influence maximization objectives on large synthetic and real-world networks.

1.1 Our Contributions

We highlight the following contributions:

1. We use an embedding approach to tackle influence maximization. We believe this is the first time embeddings have been used for fair influence maximization.
2. We introduce Adversarial Network Embeddings to address fair influence maximization. Using an autoencoder coupled with a discriminator in an adversarial setting, we obtain embeddings which are similarly distributed across sensitive attributes.
3. Our approach achieves state-of-the-art results for influence maximization (comparing to the greedy approach, previous state-of-the-art) in experiments on synthetic and real-world datasets.
4. Our method also achieves high diversity in the context of fair influence maximization.

Our work provides a fresh tool that we hope could be helpful to other social network settings where fairness is relevant, such as fair clustering or fair node-level classification.

2 Related Work

Adversarial approaches to fairness were introduced recently [Madras *et al.*, 2018; Adel *et al.*, 2019]. To our knowledge, only a few studies have considered fairness in the context of network diffusion. In this section, we review work on influence maximization, node embedding approaches, and recent investigations of fair influence maximization.

2.1 Influence Maximization

Influence maximization was first introduced as an algorithmic problem by [Richardson and Domingos, 2002] by proposing a heuristic approach to find an initial set of nodes to maximize the number of further adopters. Over the years, extensive research has focused on cascading behavior, diffusion and spreading of ideas, information or diseases, by identifying a set of initial nodes which maximizes the influence through a network [Leskovec *et al.*, 2007; Kempe *et al.*, 2003; Richardson and Domingos, 2002; Gomez-Rodriguez *et al.*, 2010].

Identifying individuals who are good initial seeds for spreading of information is studied in two ways: (i) find the set of most central nodes based on network structural properties [Kourtellis *et al.*, 2013; Kempe *et al.*, 2003]; or (ii) tackling the problem as discrete optimization [Goyal *et al.*, 2013; Kempe *et al.*, 2003; Babaei *et al.*, 2013].

[Kempe *et al.*, 2003] studied influence maximization under different social contagion models such as Linear Threshold (LT) and Independent Cascade (IC) models. They showed

that although finding the optimal solution is NP-hard, submodularity of the influence function can be used to obtain provable approximation guarantees. Since then, there has been a large body of work studying various extensions [Goyal *et al.*, 2013; Carnes *et al.*, 2007] among which [Keikha *et al.*, 2020] takes advantage of a network embedding approach.

2.2 Network Embedding

Learning a low-dimensional embedding of the nodes in the network is at the core of our proposed approach. Generally, the network embedding problem proposes to map nodes to a low dimensional space such that the network structure can be reconstructed. Network embeddings have proven their efficiency in classification and clustering problems [Hamilton *et al.*, 2017b; Cao *et al.*, 2015], and have attracted much attention from the machine learning and data mining communities [Cai *et al.*, 2018; Hamilton *et al.*, 2017b; Tang *et al.*, 2015]. Consequently, several methods have been proposed based on random-walk based models [Khajehnejad, 2019; Grover and Leskovec, 2016], deep learning architecture [Wang *et al.*, 2016], and graph neural networks [Hamilton *et al.*, 2017a]. However, to our knowledge no one has considered network embedding to address the problem of fair influence maximization.

2.3 Contemporary Works

Only a few recent works consider fairness in influence maximization. Recently, several works promote diversity in choosing the seeds [Benabbou *et al.*, 2018; Aghaei *et al.*, 2019]. [Babaei *et al.*, 2016] also show that users are sub-optimal in selecting their sources in social media in the sense of receiving diverse information. But these works do not consider a fairness criterion. The notion of individual fairness (similar individuals should be treated similarly) and group fairness (on average, members of different groups are treated similarly) for influence maximization were proposed by [Fish *et al.*, 2019] and [Tsang *et al.*, 2019], respectively.

These two studies are the most related to our work. However, they suffer from a trade-off between the objectives of influence maximization and fairness. This trade-off becomes worse as the number of sensitive attributes increases. Our method achieves state-of-the-art influence maximization results, even with multiple sensitive attributes, while also increasing fairness. We next present how our Adversarial Network Embeddings achieve these desired properties.

3 Methodology

In the *fair influence maximization* problem, we aim to influence the maximum number of nodes, while ensuring that the fraction of influenced nodes is (approximately) equal across predefined groups (e.g. minorities). For ease of exposition, we assume in this section that there are only two groups A and B , but the results extend naturally to more groups. Let I_A and I_B be the expected total number of influenced nodes belonging to groups A and B , respectively. We say a spreading process is *fair* if, by the end of it, we have:

$$\frac{I_A}{|A|} \approx \frac{I_B}{|B|}$$

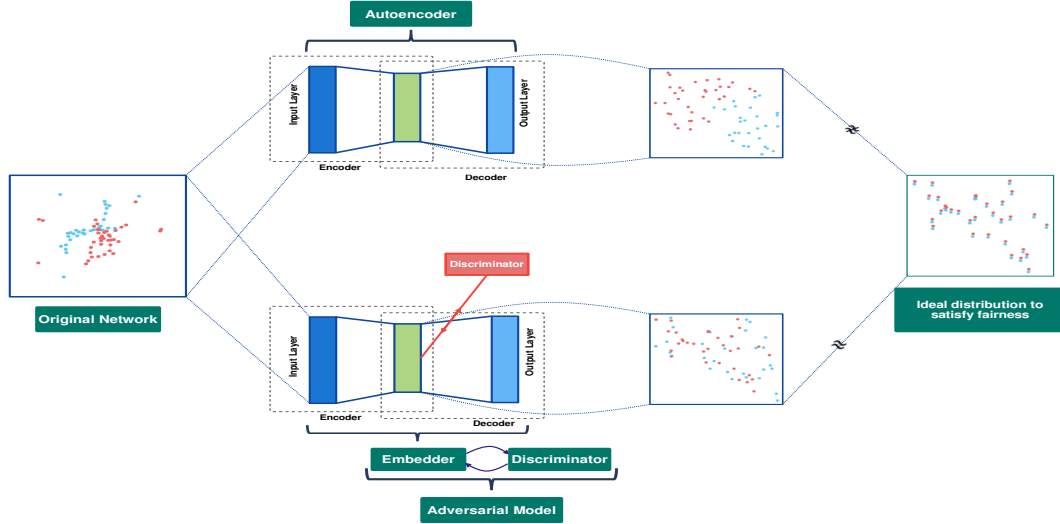


Figure 1: Illustrating the comparison between an auto-encoder and an adversarial system consisting of an auto-encoder interacting with a discriminator. The method is applied on a synthetic graph and the node representations in the new space are compared with an ideal case of equal distribution for both communities. The adversarial setting (bottom pathway) leads to a much more similar distribution to the ideal case.

We now provide the intuition and motivation behind our embedding approach. If we pick S as our set of initial seeds, let $q_S(U)$ be the probability that the set of nodes U was influenced by the end of the spreading process. We also write U_A (resp. U_B) for the set of nodes of U which belong to A (resp. to B), and $\mathcal{P}(S)$ the set of all subsets of a set S . To compute the expected fraction of influenced nodes in A , we sum over all possible disjoint infection patterns:

$$I_A = \sum_{u_a \in \mathcal{P}(A)} \mathbb{P}(u_a) \cdot |u_a|$$

The probability that all nodes of u_a are infected is given by marginalizing over all subsets of B : $\mathbb{P}(u_a) = \sum_{u_b \in \mathcal{P}(B)} q(u_a + u_b)$. Therefore:

$$I_A = \sum_{u_a \in \mathcal{P}(A)} \sum_{u_b \in \mathcal{P}(B)} q(u_a + u_b) \cdot |u_a|$$

Reindexing, and dividing by $|A|$:

$$\frac{I_A}{|A|} = \frac{1}{|A|} \cdot \sum_{U \in \mathcal{P}(\{1, \dots, N\})} q_S(U) \cdot |U_A|$$

This is where our Adversarial Network Embedding becomes useful. If we knew that for the sets U of high probability mass (*i.e.* the most likely sets of influenced nodes, with $q_S(U)$ of highest values), we had $\frac{|U_A|}{|A|} \approx \frac{|U_B|}{|B|}$, then we could claim that $\frac{I_A}{|A|} \approx \frac{I_B}{|B|}$. Intuitively, by matching the distributions of nodes from A and B in the embedding space (to have $\frac{|U_A|}{|A|} \approx \frac{|U_B|}{|B|}$ for as many U sets as possible), and picking seeds in densely populated areas (so as to put more probability mass where the above equality is verified), we design our framework, which we describe next.

To achieve the desired adversarial network embedding for an input network with two communities, inspired by *Generative Adversarial Networks* [Goodfellow et al., 2014], we

design an adversarial setting in which the embedder plays against a discriminator. In our setting, the discriminator distinguishes between the embeddings of the two communities. Concurrently the embedder tries to generate embeddings that are indistinguishable by the discriminator. In other words, the discriminator forces the embedder to generate embeddings for the two communities that are coming from distributions which are as similar as possible. We train the embedder and the discriminator with the following coupled loss functions:

$$\begin{cases} \mathcal{L}_E = \mathcal{L}_{recon} - \beta \mathcal{L}_D \\ \mathcal{L}_D = \mathcal{D}(Z_A) - \mathcal{D}(Z_B) \quad Z_A = E(X_A), Z_B = E(X_B) \end{cases}$$

where X_A, X_B refer to the vector representation of nodes in communities A, B , and Z_A, Z_B are their corresponding vector representation in the embedding space. E and L_E represent the embedder function and its relative loss function. L_{recon} denotes the reconstruction loss of a standard auto-encoder. Similarly, D and L_D refer to the discriminator function and its loss. The discriminator function computes the distance between the distribution of the given embeddings with the initial distribution of the embeddings of nodes of community A . Figure 1 depicts the whole process of our method in a graphical way. Details are shown in Algorithm 1. We can further extend the above setting in order to address fairness in influence maximization with respect to multiple sensitive attributes. In this case, we want our embeddings to have similar distributions for each value of our different sensitive attributes. To do so, we add one discriminator and one extra term to the embedder loss per sensitive attribute. Then during the adversarial training, the embedder plays against a set of discriminators (one discriminator per attribute). For an instance of two sensitive attributes, the embedder and the discriminator's loss functions would be as follows:

$$\begin{cases} \mathcal{L}_E = \mathcal{L}_{recon} - \beta_1 \mathcal{L}_{D_1} - \beta_2 \mathcal{L}_{D_2} \\ \mathcal{L}_{D_1} = \mathcal{D}_1(Z_{A_1}) - \mathcal{D}_1(Z_{B_1}) & Z_{A_1} = E(X_{A_1}), Z_{B_1} = E(X_{B_1}) \\ \mathcal{L}_{D_2} = \mathcal{D}_2(Z_{A_2}) - \mathcal{D}_2(Z_{B_2}) & Z_{A_2} = E(X_{A_2}), Z_{B_2} = E(X_{B_2}) \end{cases}$$

Algorithm 1 ADVERSARIAL GRAPH EMBEDDING: Adversarially trains the *Embedder* function E with a *Discriminator* function \mathcal{D} .

Require: Input network (X) with two communities (X_A, X_B) , *Embedder* (E) and *Discriminator* (\mathcal{D}) functions, with parameters (θ_E, θ_D) .

- 1: Pre-train E on X with $\mathcal{L}_{recon}(X) = \|X - \hat{X}\|_2^2$.
- 2: $Z_A \leftarrow E(X_A)$, $Z_B \leftarrow E(X_B)$
- 3: Pre-train \mathcal{D} on Z_A, Z_B with $\mathcal{L}_{\mathcal{D}}(Z) = \mathcal{D}(Z_A) - \mathcal{D}(Z_B)$
- 4: **for** $epoch \in \{1, \dots, n\}$ **do**
- 5: **for** $X^r \in X$ **do**
- 6: $\theta_E = \theta_E - \nabla_E[\mathcal{L}_{recon}(X^r) - \beta\mathcal{D}(E(X_A^r)) + \beta\mathcal{D}(E(X_B^r))]$
- 7: $Z_A^r \leftarrow E(X_A^r)$, $Z_B^r \leftarrow E(X_B^r)$
- 8: $\theta_D = \theta_D - \nabla_D[\mathcal{D}(Z_A^r) - \mathcal{D}(Z_B^r)]$
- 9: **end for**
- 10: **end for**
- 11: **Return** θ_E, θ_D

After reaching a favorable low-dimensional representation of the network where nodes of different communities are distributed similarly in the embedding space, the final step is to choose the proper seeds. We can safely perform the selection in the embedding space since our embedding method is invertible (the original graph can be reconstructed from the embeddings). For the case of one sensitive attribute, our goal is to choose an initial set of influential seeds S . We examine the two following approaches:

- (i) Normal Selection: This applies a k -means method on the Z space with $k = |S|$ to select the resulting k cluster centroids as initial seeds.
- (ii) Fair Selection: In normal selection (above), depending on the network structure, seeds might come from just one community, leading to disparity. To tackle this concern, we propose an alternative method introduced in Algorithm 2. Assume we want to select $|S| = k \times s$ nodes from Z as the initially influenced nodes. We start by performing a k -means algorithm to group all nodes in Z into k clusters. Then, in each cluster, we select the s nearest neighbours to the centroid and determine whether they are members of community A (N_A) or B (N_B). We also divide all the nodes of each cluster into two sub-clusters with nodes belonging to A or B . Finally, we exploit the k -means algorithm on each of these sub-clusters using $k = |N_A|$ and $k = |N_B|$ respectively and obtain the resulting centroids to have selected $s = |N_A| + |N_B|$ seeds from each of the initial k clusters giving us $|S| = k \times s$ seeds from the whole network.

We now show the proposed method's effectiveness using a small synthetic data set. Figure 1 illustrates the high performance and efficiency of the system in re-creating the network using our model. A standard auto-encoder generates a network where the two communities are completely distinguishable which is representative of completely different distributions. However, with our Adversarial Graph Embeddings, the two distributions mostly overlap - it is hardly possible to separate the two communities. The results are even more striking

Algorithm 2 FAIR SELECTION: Selects $|S|$ initial seeds attempting to choose the most influential ones.

Require: Number of clusters $k = |S|$, Embedding space Z

- 1: $\{g^i\}_{i=1}^k \leftarrow \text{CLUSTERCENTERS}(k, Z)$
- 2: $\{G^i\}_{i=1}^k \leftarrow \text{CLUSTERPOINTS}(k, Z)$
- 3: $s \leftarrow \frac{|S|}{k}$
- 4: **for** $i \in \{1, \dots, k\}$ **do**
- 5: $N^i \leftarrow \text{KNN}(s, g^i, G^i)$
- 6: $N_A^i \cup N_B^i \leftarrow N^i$
- 7: $G_A^i \cup G_B^i \leftarrow G^i$
- 8: $S_0 \leftarrow \text{CLUSTERCENTERS}(|N_A^i|, G_A^i) \cup \text{CLUSTERCENTERS}(|N_B^i|, G_B^i)$
- 9: **end for**
- 10: **Return** S_0

in Figure 2 which shows the embedding space of our proposed method in comparison with the embedding space of a standard auto-encoder for a large synthetic data set. It can be seen that using our algorithm, nodes from different communities show similar distributions in the embedding space.

In Algorithm 2, the function $\text{CLUSTERCENTERS}(k, Z)$ gives the set of k centroids when performing k -means on the Z space and $\text{CLUSTERPOINTS}(k, Z)$ returns the corresponding k cluster of points. Finally, $\text{KNN}(s, g^i, G^i)$ outputs s nearest neighbours of g^i in the space of G^i .

4 Dataset

We evaluate the performance of our approach on real and synthetic data.¹

4.1 Real Dataset

Dataset Description: The real dataset is the *Rice University Facebook* dataset, collected by [Mislove *et al.*, 2010], which represents the friendship relations between students of Rice University. The friendship graph is an undirected graph that has 1205 nodes with 42443 links between them. Nodes in this graph contain information about students such as student id, age (which is a number between 18 to 22), and major.

Experimental Setup: A sub-graph of the *Rice Facebook*

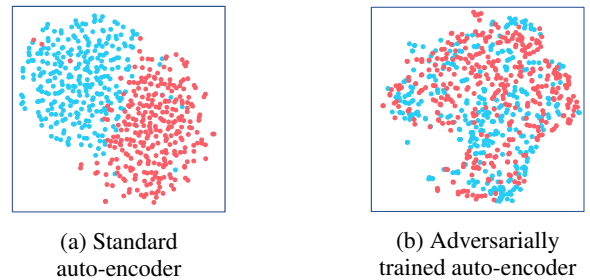


Figure 2: Embedding space for the standard and adversarially trained auto-encoders. Nodes of different communities are mapped to two segregated sub-spaces in (a), while being similarly distributed in the embedding space of (b).

¹For code and details of parameter tunings please refer to: https://github.com/Ahmadreza401/fair_influen_max

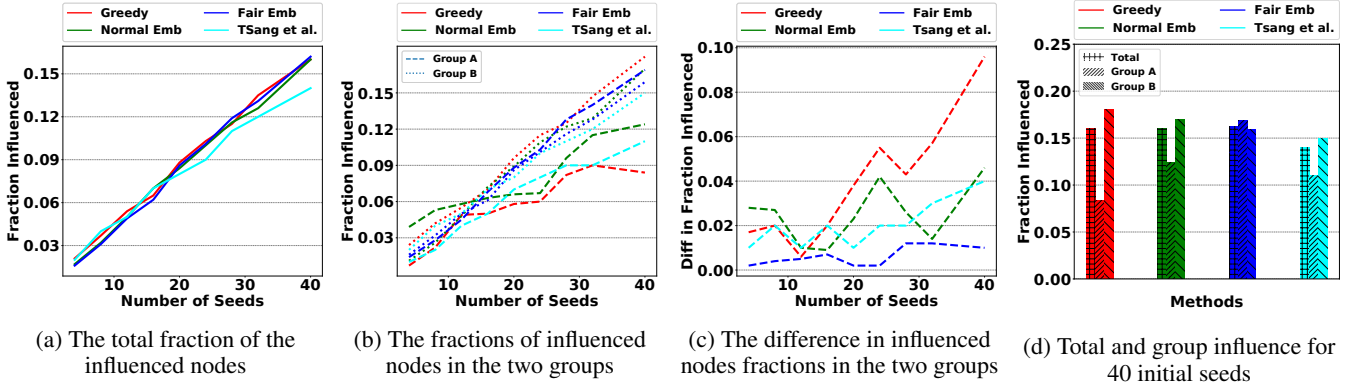


Figure 3: Comparison of the results of the *Normal Embedding* (green) and *Fair Embedding* method (blue), as well as the *Greedy* method (red), and *Tsang et al.* (cyan) as baselines over the Rice-Facebook dataset. It can be seen that the *Fair Embedding* approach influences a fair fraction of both groups while it does not affect the total influence.

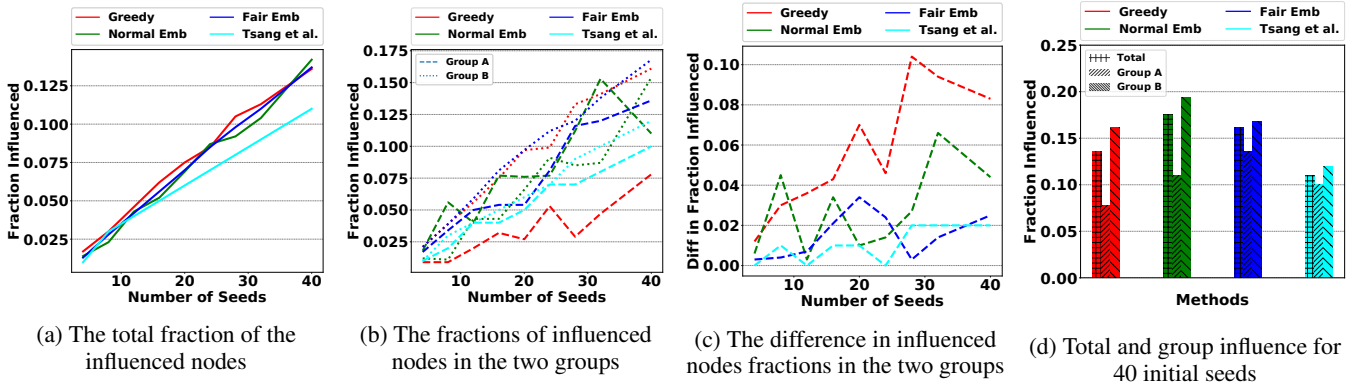


Figure 4: Comparison of the results on the synthetic dataset. Similarly, the *Fair Embedding* (blue) method outperforms the *Normal Embedding* (green), *Tsang et al.* (cyan), and *Greedy* (red) methods in enhancing the fairness while being very close to them in terms of total influence.

dataset is used in the experiments by excluding nodes with a value greater than 20 for their age attribute. We define two communities in the sub-graph. Nodes with a value of 18 or 19 for their age attribute constitute group A or \mathcal{V}_A , and the rest of the nodes form group B , \mathcal{V}_B . Group \mathcal{V}_A has 97 nodes with 513 intra-connections, while group \mathcal{V}_B has 344 nodes with 7441 intra-connections. There are 1779 inter-connections between nodes of the two communities. We assume an activation probability of 0.01 for every link, which is used in the independent cascade model for information propagation.

4.2 Synthetic Dataset

Dataset Description: The synthetic dataset is an undirected graph where each node belongs to either of two groups, \mathcal{V}_A and \mathcal{V}_B . The size of the groups is set based on a parameter ratio r , $0 < r < 1$, where $r|\mathcal{V}|$ nodes belong to group \mathcal{V}_A and $((1 - r)|\mathcal{V}|)$ nodes belong to group \mathcal{V}_B . Nodes are connected based on intra-group, $\mathcal{P}_{intra}^{\mathcal{V}_A}$ or $\mathcal{P}_{intra}^{\mathcal{V}_B}$, and inter-group (\mathcal{P}_{inter}) connection probabilities. To connect two nodes that belong to the same group (lets say \mathcal{V}_A), we do a Bernoulli trial with probability $\mathcal{P}_{intra}^{\mathcal{V}_A}$ and connect the two nodes if the outcome of the trial is 1. Figure 2 shows that the distribution of our Adversarial Graph Embeddings is extremely similar across both groups. Similarly, two nodes that belong to two different groups are connected only if a

Bernoulli trial with probability \mathcal{P}_{inter} is successful. As for the real dataset, there is an activation probability \mathcal{P}_a associated with every link in the network.

Experimental Setup: In our experiments, we set $r = 0.3$ which gives 150 nodes in \mathcal{V}_A and 350 nodes in \mathcal{V}_B . We use the same intra-group connection probability for both groups, $\mathcal{P}_{intra}^{\mathcal{V}_A} = \mathcal{P}_{intra}^{\mathcal{V}_B} = 0.025$. To ensure two almost-separated communities, we set the inter-group connection probability $\mathcal{P}_{inter} = 0.001$ to be less than the intra-group probability. Group \mathcal{V}_A has 134 intra-group connections, while group \mathcal{V}_B has 2843 intra links. There are also 129 inter-group connections between the nodes of the two groups. The activation probability $\mathcal{P}_a = 0.03$ is the same for all links in the network. We tried our method over a range of synthetic data-sets achieving qualitatively similar results. Here we include just the results for the above settings.

5 Results

This section includes the results for our proposed method along with the results of two state-of-the-art methods, including *Greedy* [Kempe et al., 2003], as well as *Tsang et al.* [Tsang et al., 2019] over synthetic and real data-sets. Our experiments follow a two-step process. First, a number of seeds are picked from the input network using our proposed method and the baseline methods. For the *seed selection*, we utilize

the *Fair Selection* approach as it proves to outperform the Normal Selection method. The selected seeds are then passed to an independent cascade model [Kempe *et al.*, 2003] to compute the final influenced nodes. The results reported show the total fraction of nodes influenced, the fraction of nodes influenced from each community, and the difference between the fractions of nodes influenced from the two communities. Figures 3 and 4 show the results over the real and synthetic datasets respectively. In the following, we discuss the technical aspects of the methods that appeared in the results and will explain the results in more details. We conclude the results section by explaining the results for a more general case of two sensitive attributes (Figure 5).

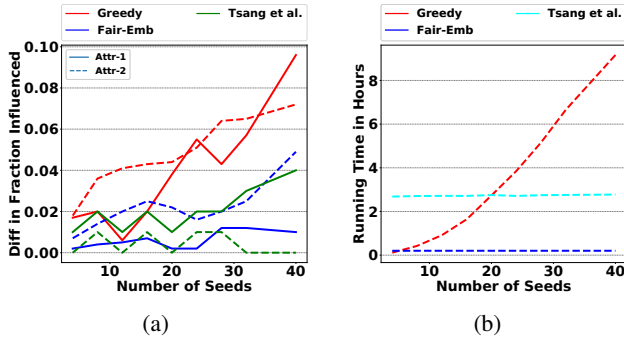


Figure 5: (a) The difference in the fraction influenced by the two groups for the case of two sensitive attributes. (b) Comparison of the running time of our proposed method and the baseline.

Normal Embedding: Every node of the input network is first described by its row in the adjacency matrix. The *Normal Embedding* method learns a low-dimensional representation for each node using an auto-encoder. The auto-encoder ensures that the low-dimensional representation is informative enough to reconstruct the adjacency list by imposing a *cross-entropy* loss that penalizes the model for any dissimilarity between the main and reconstructed adjacency list. After getting the embedding, we cluster the points in the embedding space to four clusters (We tested for different numbers of clusters and observed that considering 4 clusters performs best in terms of fairness) and pick the seeds that are nearest to the center of the clusters. This strategy helps to pick nodes with different properties that are likely to influence a good fraction of nodes in the network. Results shows that *Normal Embedding* influences a comparable fraction of the nodes in both the real (Figure 3(a)) and synthetic (Figure 4(a)) data-sets which is close to the fraction influenced by the *Greedy* method (state-of-the-art classical IM solutions). However, observe that both *Normal Embedding* and *Greedy* are very biased: influencing more nodes from the community of bigger size. In both data-sets, group B is larger and as we see (Figure 3(b, d) and Figure 4(b, d)), the majority of nodes influenced by the *Normal Embedding* and the *Greedy* methods belong to group B.

Fair Embedding: To address fairness concerns in *Normal Embedding*, we adversarially train the auto-encoder with a discriminator. Before adversarial training, the discriminator

is pre-trained to be able to discriminate between embeddings of the two communities. Through the adversarial training, the auto-encoder tries to learn an embedding for nodes of different groups that are indistinguishable by the discriminator. Using this method, the embedding for nodes of group A will have similar distribution to that of group B. The adversarial training is followed by a selection algorithm that picks the seeds from the adversarially trained embedding. The selection algorithms are discussed in full details in the *methodology* section.

The results show that the *Fair Embedding* method gives a boost to the fraction of nodes influenced from the group of smaller size in both real and synthetic data sets (Fig 3(c, d) and Fig 4(c, d)). This happens while the total fraction of nodes influenced remains similar (Fig3(a) and Fig4(a)), the *Fair Embedding* only marginally reduces the fraction of nodes influenced from the larger group (See Fig3(b) and Fig4(b)). Overall, the fractions of nodes influenced from the two communities are close to one another (Fig3(c,d) and Fig4(c,d)), improving fairness compared to the *Normal Embedding* and *Greedy* methods. More importantly, in the real dataset, our method significantly outperforms baselines in both maximizing the numbers of people reached and reducing disparity (Figure 3). In synthetic dataset, it seems that our results are very competitive compared to *Tsang et al.* in terms of fairness, however we outperform *Tsang et al.* in terms of the total number of people influenced, and the fraction of people influenced from both A and B communities.

Two sensitive attributes: Figure 5a shows the results for the case of two sensitive attributes. As shown, our method outperforms *Greedy* in terms of fairness concerning both sensitive attributes. Our method also shows a comparable performance with *Tsang et al.* in terms of fairness, however, it is more efficient in terms of running time (see Figure 5b).

6 Conclusion

In this paper, we proposed a new method to tackle the problem of fair influence maximization. We observed that existing algorithmic methods to address influence maximization in the population usually lead to considerable disparity. While there usually exists a trade-off between fairness objectives and maximization objectives proposed, our proposed approach achieves both: experimental results over synthetic and real-world data-sets demonstrate that our method is competitive with the prior state-of-the-art for maximizing the total number of influenced people, while also significantly decreasing disparity.

While we demonstrated the effectiveness of our Adversarial Network Embeddings in the context of fair influence maximization, we believe our approach could be used in future work to tackle other fairness tasks over social networks such as fair clustering or fair node-level classification.

Acknowledgements

AW acknowledges support from the David MacKay Newton Research Fellowship at Darwin College, The Alan Turing Institute under EPSRC grant EP/N510129/1 & TU/B/000074, and the Leverhulme Trust via the CFI.

References

- [Adel et al., 2019] Tameem Adel, Isabel Valera, Zoubin Ghahramani, and Adrian Weller. One-network adversarial fairness. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2412–2420, 2019.
- [Aghaei et al., 2019] Sina Aghaei, Mohammad Javad Azizi, and Phebe Vayanos. Learning optimal and fair decision trees for non-discriminative decision-making. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1418–1426, 2019.
- [Ali et al., 2019] Junaid Ali, Mahmoudreza Babaei, Abhijnan Chakraborty, Baharan Mirzasoleiman, Krishna P. Gummadi, and Adish Singla. On the fairness of time-critical influence maximization in social networks. *ArXiv*, abs/1905.06618, 2019.
- [Babaei et al., 2013] Mahmoudreza Babaei, Baharan Mirzasoleiman, Mahdi Jalili, and Mohammad Ali Safari. Revenue maximization in social networks through discounting. *Social Network Analysis and Mining*, 3(4):1249–1262, 2013.
- [Babaei et al., 2016] Mahmoudreza Babaei, Przemyslaw Grabowicz, Isabel Valera, Krishna P Gummadi, and Manuel Gomez-Rodriguez. On the efficiency of the information networks in social media. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 83–92, 2016.
- [Babaei et al., 2018] Mahmoudreza Babaei, Jui Kulshrestha, Abhijnan Chakraborty, Fabrício Benevenuto, Krishna P Gummadi, and Adrian Weller. Purple feed: Identifying high consensus news posts on social media. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pages 10–16. ACM, 2018.
- [Banerjee et al., 2013] Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. The diffusion of microfinance. *Science*, 341(6144), 2013.
- [Benabbou et al., 2018] Nawal Benabbou, Mithun Chakraborty, Xuan-Vinh Ho, Jakub Sliwinski, and Yair Zick. Diversity constraints in public housing allocation. In *AAMAS*, pages 973–981, 2018.
- [Cai et al., 2018] Hongyun Cai, Vincent W Zheng, and Kevin Chen-Chuan Chang. A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1616–1637, 2018.
- [Cao et al., 2015] Shaosheng Cao, Wei Lu, and Qiongkai Xu. Grarep: Learning graph representations with global structural information. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 891–900. ACM, 2015.
- [Carnes et al., 2007] Tim Carnes, Chandrashekhar Nagarajan, Stefan M Wild, and Anke Van Zuylen. Maximizing influence in a competitive social network: a follower’s perspective. In *EC*, pages 351–360. ACM, 2007.
- [Fish et al., 2019] Benjamin Fish, Ashkan Bashardoust, danah boyd, Sorelle A. Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. Gaps in Information Access in Social Networks. In *WWW*, 2019.
- [Gomez-Rodriguez et al., 2010] M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring Networks of Diffusion and Influence. In *KDD*, 2010.
- [Goodfellow et al., 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [Goyal et al., 2013] Amit Goyal, Francesco Bonchi, Laks VS Lakshmanan, and Suresh Venkatasubramanian. On minimizing budget and time in influence propagation over social networks. *Social network analysis and mining*, 3(2):179–192, 2013.
- [Grover and Leskovec, 2016] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864. ACM, 2016.
- [Hamilton et al., 2017a] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.
- [Hamilton et al., 2017b] William L Hamilton, Rex Ying, and Jure Leskovec. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584*, 2017.
- [Keikha et al., 2020] Mohammad Mehdi Keikha, Maseud Rahgozar, Masoud Asadpour, and Mohammad Faghih Abdollahi. Influence maximization across heterogeneous interconnected networks based on deep learning. *Expert Systems with Applications*, 140:112905, 2020.
- [Kempe et al., 2003] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *KDD*, 2003.
- [Khajehnejad, 2019] Moein Khajehnejad. Simnet: Similarity-based network embeddings with mean commute time. *PloS one*, 14(8):e0221172, 2019.
- [Kourtellis et al., 2013] Nicolas Kourtellis, Tharaka Alahakoon, Ramanuja Simha, Adriana Iamnitchi, and Rahul Tripathi. Identifying high betweenness centrality nodes in large social networks. *Social Network Analysis and Mining*, 3(4):899–914, 2013.
- [Leskovec et al., 2007] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *KDD*, pages 420–429. ACM, 2007.
- [Li et al., 2018] Yuchen Li, Ju Fan, Yanhao Wang, and Kian-Lee Tan. Influence maximization on social graphs: A survey. *TKDE*, 30(10):1852–1872, 2018.
- [Madras et al., 2018] David Madras, Elliot Creager, Toniann Pitassi, and Richard Zemel. Learning adversarially fair and transferable representations. In *International Conference on Machine Learning*, pages 3384–3393, 2018.
- [Mirzasoleiman et al., 2012] Baharan Mirzasoleiman, Mahmoudreza Babaei, and Mahdi Jalili. Immunizing complex networks with limited budget. *EPL (Europhysics Letters)*, 98(3):38004, 2012.
- [Mislove et al., 2010] Alan Mislove, Bimal Viswanath, Krishna P Gummadi, and Peter Druschel. You are who you know: inferring user profiles in online social networks. In *WSDM*, pages 251–260. ACM, 2010.
- [Richardson and Domingos, 2002] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *KDD*, 2002.
- [Tang et al., 2015] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on world wide web*, pages 1067–1077. International World Wide Web Conferences Steering Committee, 2015.
- [Tsang et al., 2019] Alan Tsang, Bryan Wilder, Eric Rice, Milind Tambe, and Yair Zick. Group-Fairness in Influence Maximization. *arXiv preprint arXiv:1903.00967*, 2019.
- [Wang et al., 2016] Daixin Wang, Peng Cui, and Wenwu Zhu. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1225–1234. ACM, 2016.
- [Ye et al., 2012] Mao Ye, Xingjie Liu, and Wang-Chien Lee. Exploring social influence for recommendation: a generative model approach. In *SIGIR*, pages 671–680. ACM, 2012.