Project Report:
Fair Node Classification

By Johanna Nagel

Lecture: Online Social Network and Media

Date: 14.02.2024

Abstract

Machine learning (ML) is integral for predictions and classifications but confronts challenges tied to bias, particularly affecting underrepresented groups. These encompass diverse ethnicities, foreigners, and individuals with disabilities. ML's role in network analysis, focusing on graph structures, introduces embedding methods representing nodes and relationships in a lower-dimensional space. However, these algorithms may encode biases, necessitating research to identify fair embedding methods.

This experiment investigates the fairness of four embedding methods across five diverse datasets, utilizing metrics proposed by Verma and Rubin with a primary focus on gender-related fairness. Beyond predictive accuracy, the study employs benchmarks to ensure comprehensive examinations. Insights from 5x4 experiments aim to contribute to fair and unbiased decision-making algorithms, promoting inclusivity in ML applications, particularly addressing gender-related considerations.

The results showed that each embedding method exhibited different levels of fairness across the diverse datasets. However, two of the methods were slightly more significantly fairer than the others in terms of gender considerations. The two methods are Node2Vec and DeepWalk. There achieved the most consistent and best predicted results.

Table of Contents

1 Introduction

Machine learning (ML) plays a vital role in prediction and classification tasks, but it often grapples with bias and unfairness issues, especially impacting underrepresented groups like minorities, foreigners, and individuals with disabilities. Addressing these challenges is not just a technological necessity but also crucial for promoting equity and justice. ML algorithms are widely used across various domains, including network analysis, where understanding graph structures is essential. Embedded within this field are embedding methods, which represent nodes and their relationships in a lower-dimensional space.

However, even these algorithms are susceptible to encoding biases implicitly or unintentionally, leading to unfair outcomes. Hence, it's crucial to research and identify fair embedding methods. In this experiment, we aim to investigate whether different embedding methods produce fair predictions. We'll select four embedding methods and apply them to five diverse datasets, using resulting embeddings for label classification and evaluating the outcomes.

To assess fairness in graph-based models, we'll utilize appropriate metrics and benchmarks, ensuring a comprehensive examination beyond predictive accuracy. This holistic approach is crucial to ensure models not only perform well but also adhere to fairness standards, treating diverse groups equitably. We'll utilize fairness metrics proposed by (Verma and Rubin, 2018, p. 1 ff.) with a primary focus on determining if embedding methods exhibit unfair behavior towards women.

Through this experiment, we aim to contribute to ongoing efforts in developing fair and unbiased decision-making algorithms, fostering inclusivity and fairness in ML applications, particularly concerning gender-related considerations.

As a result of these experiments, we find that each method can predict fair results, but based on the datasets, it is easier for Node2Vec and DeepWalk to produce fair results.

2 Methodology

2.1 Procedure

To address unfairness in graph-based machine learning classification, particularly concerning gender, we begin by selecting five datasets as our investigation foundation. Our objective is to binary classify nodes based on these datasets, such as determining loan approval based on gender in the credit dataset (Dataset 2). After selecting datasets, we generate networks from edge files, assigning nodes the gender attribute (0,1) and few additional attributes were considered as well. Next, we choose various embedding methods (refer to Chapter 2.3) and apply them to the datasets to project nodes into a vector space for classification and analysis.

Following method application, we obtain vector representations for nodes in the graph, learned through model training to capture patterns and structure. These embeddings serve as features for predicting labels using Support Vector Machines for binary classification, leveraging their effectiveness in higher-dimensional spaces. We partition embedding vectors into training (70%) and test sets (30%) and train the model, accordingly, assessing fairness towards women across all datasets using specified evaluation metrics (detailed in subsection 2.4).

Post-evaluation, we aim to fine-tune parameters to address fairness concerns, employing a Grid search algorithm to optimize parameters and minimize false positive rates, ultimately striving for improved results in mitigating unfairness.

2.2 Datasets

As mentioned earlier, we have chosen 5 different datasets on which we will apply the respective embedding methods.

How did we come to the selection of these datasets? We explored existing graph-based datasets from Snap.py and those that have already been used in fairness literature (cf. Verma and Rubin, 2018, 1. ff). These datasets will be adapted, if necessary.

2.2.1 Pokec Social Network

The Pokec dataset (Takac and Zabovsky, 2012) is about the social network in Slovakia. The dataset consists of 1.6 million nodes, representing users of the network. The edges between the nodes/users describe the friendships between Person A and Person B in the network. The dataset contains profile data for each node, including gender, age, public profiles, hobbies, interests, education, and much more. Further statistics can be seen in Table 1.

The goal of this dataset is to predict whether Person A has a public profile. Subsequently, we want to demonstrate that the probability of having a public profile is identical for both women and men.

For the analysis we reduce these datasets and the following datasets to get a better time performance.

Table 1: Pocek Dataset Statistics

DATASET STATISTICS	
NODE ATTRIBUTES	Yes
EDGE ATTRIBUTES	No
NODE LABELS	Yes, binary labeled (public account =1))
TEMPORAL	No
NODES	1 632 803 (before data reduction)
EDGES	30 622 564
NODES IN LARGEST WCC	1 632 803 (1.000)
EDGES IN LARGEST WCC	30 622 564 (1.000)

2.2.2 German Credit Dataset

The German Credit (Agarwal et al.; cf. Agarwal and Lakkaraju, 2024) dataset describes credit accounts for each customer. It includes information about creditworthiness, gender, age, loan duration, purpose of the loan, loan amount, and more. The dataset consists of approximately 1000 nodes representing customers (see Table 2). Edges represent the similarity of their credit accounts. The edges weren't represented, to we use a similarity measure to calculate the similarity.

The goal here is to predict whether Person A is creditworthy or not and whether the assignment of credit membership was influenced by gender.

Table 2: German Credit Dataset Statistics

DATASET STATISTICS	
NODE ATTRIBUTES	Yes (e.g. Gender)
EDGE ATTRIBUTES	No
NODE LABELS	Yes, binary labeled (credit worthy =1))

TEMPORAL	No
NODES	1000
EDGES	460113

2.2.3 Employee Dataset

The Employee Dataset (cf. Elemetwally) describes employees of a company. This dataset is not a graph-based dataset, but it will be transformed it to one by treating each employee as a node and the edge between the employees represents the work relationship between them. It means for example Node A works with node B. The edges weren't part of the Dataset, so we created random edges. The dataset contains multiple information about Gender, Age, Education, City, Joining Year, and more information's. More Network information can you find in the table below.

The goal of this dataset is to predict if employees want to leave or staying at the job.

Table 3: Employee Dataset Statistics

DATASET STATISTICS	
NODE ATTRIBUTES	Yes (e.g. Gender, Age, Education, City, Joining Year etc.)
EDGE ATTRIBUTES	No
NODE LABELS	Yes, binary labeled (will leave the job =1))
NODES	4653 (before Data reduction)
EDGES	10822

The other two Datasets can be found in the appendix A2.

2.3 Embeddings

The goal of the project is to determine whether embedding methods are fair. For this analysis, we are using four different embedding methods. These embedding methods include Node2Vec, DeepWalk, AE Embedding, and SINE Embedding.

We chose these four embeddings because they are used in the literature and have also been partially introduced in the lecture.

2.3.1 Node2Vec

Node2Vec is a graph-based semi-supervised learning algorithm used for scalable feature learning in networks (cf. ACM Conferences, 2024, p. 2). It generates numerical representations from graphs, preserving the structure of the graph during the transformation (cf. Lau, 2021). These transformations are performed using a 2nd order (biased) random walk.

2.3.2 DeepWalk

DeepWalk is an algorithm designed to create embeddings for nodes in a graph, emphasizing the notion that nearby nodes should have similar embeddings to capture structural information effectively (cf. Bryan Perozzi et al., p. 1ff). It operates in two phases: first, by conducting random walks to uncover local structures within the network, and second, by employing SkipGram, a model trained to predict common contexts shared by nodes in these random walks, thereby learning embeddings that encode node similarities (cf. ÖZçelik, 2019).

2.3.3 Scalable Incomplete Network Embedding

Scalable Incomplete Network Embedding (SINE) is an attributed Network embedding which learns a low-dimensional vector representation for nodes with attributes and features (cf. Rozemberczki et al., 2019, pp. 1 ff). It's designed to handle incomplete or noisy networks efficiently.

SINE works by first generating random walks on the input network, then extracting shorter sequences called walklets from these walks to capture local structural information. Next, node features are transformed into a compatible format and combined with the walklets. A word embedding model is then trained on this combined data to learn continuous representations of nodes in a vector space. These learned embeddings encode both structural and feature information of nodes, facilitating downstream tasks like node classification or link prediction (cf. Sarkar Rik, 2021, pp. 1 ff; (Zhang et al., 2018).

2.3.4 Attributed Network Embedding

Attributed Network Embedding involves learning low-dimensional representations of nodes in a network while considering both structural connectivity and node attributes. By incorporating information from the network topology and node attributes, these methods aim to capture the underlying patterns in the data more effectively. These embeddings facilitate various downstream tasks such as node classification and link prediction by leveraging both structural and attribute information simultaneously (cf. Rozemberczki et al., 2019, pp. 1 ff).

2.4 Metrics

In the paper by Verman and Rubin (2018), various fairness definitions are presented using a sample dataset. For the fairness analysis, we will introduce specific fairness definitions that we used for the experiment.

Before doing so, let's briefly discuss the validation of classification results. For the traditional validation of predictions, a Confusion Matrix is employed. The Confusion Matrix, shown in the following figure, compares predicted labels with the actual values. It indicates the number of correctly predicted true negatives (TN), false positives (FP), and false negatives (FN) of the model's predictions. This confusion matrix forms the basis for fairness definitions/metrics.

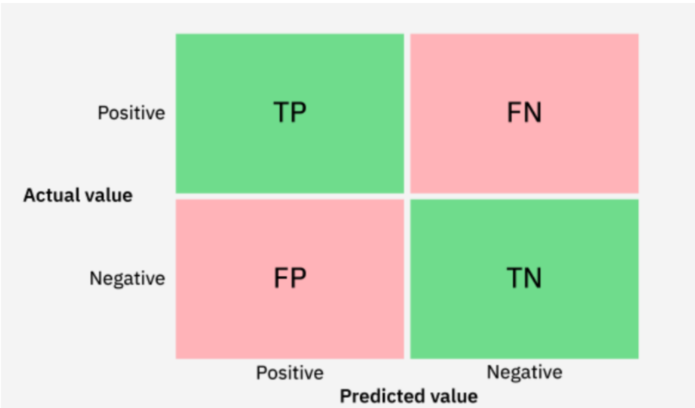

Figure 1: Confusion Matrix (cf. IBM, 2024)

2.4.1 Group Fairness

Group Fairness describes a classifier meeting the fairness definition when subjects from both the protected and unprotected groups have an equal probability of being assigned to a positive predicted class. In our case, this means that women and men have an equal probability of being assigned to a positive class (cf. Verma and Rubin, 2018, p. 3).

2.4.2 False Positive Error Rate Balance

A classifier is fair if the protected and unprotected groups have the same False Positive Rate (FPR). In our case this implies that men and women with an actual negative prediction have an equal probability of receiving a predicted positive value (cf. Verma and Rubin, 2018, p. 4).

2.4.3 Predictive Parity

A classifier is fair if the protected and unprotected groups have the same Positive Predictive Value (PPV). In our case, this means that men and women with a positive prediction have an equal probability of being assigned to a positive class (cf. Verma and Rubin, 2018, p. 3).

2.4.4 Equalized Opportunity:

A classifier is fair if the protected and unprotected groups have the same False Negative Rate (FNR). In our case, this means that men and women with an actual positive prediction have an equal probability of receiving a predicted negative value (cf. Verma and Rubin, 2018, p. 4).

2.4.5 Equalized Odds

A classifier is fair if the protected and unprotected groups have the same FPR and True Positive Rate (TPR) (cf. Verma and Rubin, 2018, p. 4).

After introducing the Methodology, we want to introduce the results of the Experiment.

3 Results

After conducting the experiment to determine which embedding method is the most fair based on the defined fairness metrics, we discovered that each method yielded different results. Here, we will delve into one or two examples for each method that performed the best compared to the other datasets.

3.1 Node2Vec

The best results in the Node2Vec Embedding occurred in the Twitch Dataset and Pokec Dataset. In the following Figure 2, the fairest results were achieved with the best parameter combination ('dimensions': 16, 'num_walks': 25, 'walk_length': 30, 'workers': 4). Based on the fairness definitions in chapter 2.4, Node2Vec is only fair considering the equalized opportunity with scores of 0.81 and 0.85.

	Female	Male
Group Fairness	0.666667	0.333333
Predictive Parity	0.500000	1.000000
False Positive Error Rate	0.111111	0.000000
Equalized Opportunity	0.818182	0.857143
Equalized Odds PPV	0.181818	0.142857
Equalized Odds FPR	0.111111	0.000000

Figure 2: Node2Vec results, Twitch Dataset

When examining the results of Node2Vec based on the Soc Pokec Dataset (see figure 3), it becomes evident that nearly every fairness definition holds true. However, it's crucial to note that the false positive rate (FPR) for both female and male categories is exceptionally high, as is the positive predictive value (PPV). Surprisingly, this outcome surpasses the previous one in terms of overall fairness.

	Female	Male
Group Fairness	0.0	0.000000
Predictive Parity	0.5	0.333333
False Positive Error Rate	1.0	1.000000
Equalized Opportunity	0.0	0.000000
Equalized Odds PPV	1.0	1.000000
Equalized Odds FPR	1.0	1.000000

Figure 3: Node2Vec results, Pokec Dataset

3.2 DeepWalk

The DeepWalk embedding method shows similar results to Node2Vec. In addition, DeepWalk embedding shows fair results for almost every metric based on gender, as shown in Figure 4.

	Female	Male
Group Fairness	0.0	0.000000
Predictive Parity	0.5	0.666667
False Positive Error Rate	1.0	1.000000
Equalized Opportunity	0.0	0.000000
Equalized Odds PPV	1.0	1.000000
Equalized Odds FPR	1.0	1.000000

Figure 4: DeepWalk results, Soc Pokec Datset

3.3 SINE and AE

The figure depicted below illustrates the outcomes derived from applying the SINE and AE algorithms to the German dataset. Notably, across various definitions, excluding Group Fairness and Predictive Parity, the results exhibit similarities. Consequently, based on these fairness definitions, it can be inferred that both SINE and AE algorithms demonstrate fairness. However, when evaluating Group Fairness and Predictive Parity, it becomes apparent that the embeddings do not meet the criteria for fairness.

Group Fairness	0.346667	0.653333
Predictive Parity	0.663462	0.709184
False Positive Error Rate	1.000000	1.000000
Equalized Opportunity	0.000000	0.000000
Equalized Odds PPV	1.000000	1.000000
Equalized Odds FPR	1.000000	1.000000

Figure 5: Results of the German Credit Dataset by applying SINE and AE

4 Discussion

In the preceding section, we observed the outcomes of each experiment, revealing that each embedding method gets different results regarding the dataset and are for different dataset useful. The Datsets Soc Pokec, Twitch and Geman Credit Card dataset have the best results regarding their embedding methods. But it seems that Node2Vec and DeepWalk are the most effective embedding method based on this experiment. However, this does not imply that using these methods eliminates the need to address fairness issues.

The literature may introduce superior embedding methods not explored in this experiment due to the multitude of available techniques.

Another reason to contemplate fairness outcomes is the inherent bias present in the utilized dataset, potentially harboring multiple biased pieces of information. This, in turn, could lead to unfairness. Additionally, it is essential to note that this experiment employed only a binary classifier, the Support Vector Machine. Evaluating the performance of the best classifiers (binary or multi-class) across various embeddings should be explored. Furthermore, investigating whether incorporating more attributes in datasets yields better results is a consideration.

In summary, the research area poses numerous unanswered questions, providing ample opportunities for further exploration.

5 Conclusion

This paper examines a range of embedding methods across diverse datasets with the goal of determining the fairest approach. The findings reveal that the fairness of results varies considerably depending on the datasets, with Node2Vec and DeepWalk emerging as the most consistent methods.

6 References

ACM Conferences (2024) *node2vec | Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* [Online] (Accessed 14 February 2024).

Agarwal, C. and Lakkaraju, H. (2024) *German Credit Dataset* [Online]. Available at https://github.com /chirag126/nifty/tree/main (Accessed 3 February 2024).

Agarwal, C., Lakkaraju, H. and Zitnik, M. 'Towards a Unified Framework for Fair and Stable Graph Representation Learning', *PMLR 161:2114-2124* [Online]. Available at http://arxiv.org/pdf/ 2102.13186.pdf.

Bryan Perozzi, Rami Al-Rfou and and Steven Skiena 'DeepWalk: Online Learning of Social Representations' [Online]. Available at http://www.perozzi.net/publications/14_kdd_deepwalk.pdf (Accessed 14 February 2024).

Elemetwally, T. *Employee Dataset: Employee data for classification task* [Online]. Available at https:// www.kaggle.com/datasets/tawfikelmetwally/employee-dataset?resource=download.

IBM (2024) *What is a confusion matrix?* [Online]. Available at https://www.ibm.com/topics/ confusion-matrix (Accessed 14 February 2024).

Lau, R. (2021) 'Node2vec explained graphically - Towards Data Science', *Towards Data Science*, 6 July [Online]. Available at https://towardsdatascience.com/node2vec-explained-graphically-749e49b7eb6b (Accessed 14 February 2024).

ÖZçelik, R. (2019) 'An Intuitive Explanation of DeepWalk - Analytics Vidhya - Medium', *Analytics Vidhya*, 13 October [Online]. Available at https://medium.com/analytics-vidhya/an-intuitive-explanation-of-deepwalk-84177f7f2b72 (Accessed 14 February 2024).

Rozemberczki, B., Allen, C. and Sarkar Rik (2019) *Multi-scale Attributed Node Embedding* [Online]. Available at https://snap.stanford.edu/data/github-social.html.

Takac, L. and Zabovsky, M. (2012) *Data Analysis in Public Social Networks* [Online], Poland, International Scientific Conference & International Workshop Present Day Trends of Innovations. Available at https://snap.stanford.edu/data/soc-Pokec.html.

Verma, S. and Rubin, J. (2018) 'Fairness definitions explained', *Proceedings of the International Workshop on Software Fairness.* Gothenburg Sweden, 29 05 2018 29 05 2018. New York, NY, ACM, pp. 1–7.

Zhang, D., Yin, J., Zhu, X. and Zhang, C. (2018) *SINE: Scalable Incomplete Network Embedding* [Online]. Available at https://arxiv.org/pdf/1810.06768.pdf (Accessed 11 February 2024).

A - Appendix

A1 – Datasets

A1.1 Twitch Gamers Social Network

This Dataset (Twitch Gamers Social Network) (cf. Rozemberczki & Sarkar Rik, 2021) outlines the mutual follower relationships among Twitch gamers, with gamers representing the nodes in the social network. The entire network constitutes a single strongly connected component. One of the primary objectives is to predict whether gamers are affiliated (affiliate = 1) or not. For this prediction, we will leverage multiple features.

Since the dataset lacks a gender attribute, we plan to synthetically incorporate gender information (1 = female, 0 = male) into the dataset. In the table below, you'll find key details about the entire dataset, indicating that the graph comprises 168,114 nodes. However, for the prediction task, only 1,000 nodes will be used, selected randomly.

Table 4: Twitch Gamers Social Network Dataset Statistics

DATASET STATISTICS	
NODE ATTRIBUTES	Yes
EDGE ATTRIBUTES	No
NODE LABELS	Yes, Binary labeled (affililate=1))
TEMPORAL	No
NODES	168 114 (before data reduction)
EDGES	6 797 557
DENSITY	0,0005
TRANSITIVITY	0,0184

A1.2 GitHub Social Network

The GitHub Social Network (cf. Rozemberczki et al., 2019) delineates mutual follower relationships among developers on GitHub. Nodes in the network represent developers with more than 10 repositories. The objective is to predict whether developers specialize in machine learning (ML) or web development (0 = web, 1 = ML). To enhance the predictive accuracy, multiple attributes will be utilized in the prediction process. Additionally, synthetic gender information will be introduced into the dataset, and the number of nodes will be randomly reduced. Further statistics are presented in the table below.

Table 5: GitHub Social Network Dataset Statistics

DATASET STATISTICS	
NODE ATTRIBUTES	Yes
EDGE ATTRIBUTES	No
NODE LABELS	Yes, Binary labeled (ML=1, Web=0)

TEMPORAL	No
NODES	37 700 (after data reduction 10 000)
DENSITY	0,001
TRANSITIVITY	0,013