Exploring Fairness in Graph Embeddings for Bipartite Graphs

Rohit Gajawada
rohitgajawada@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Sanjana Garg
sgarg96@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Rohit Mujumdar
rohitmujumdar@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

KEYWORDS

Graph Embeddings, Bipartite Networks, Fairness, Feature Learning

1 INTRODUCTION

1.1 Aim

In our project, we explore the idea of unbiased random walks to improve the fairness in the node embeddings. We plan to experiment with different sampling strategies in the generation of the random walks used in training. We will be focusing on transductive embedding strategies like metapath2vec[**?**] but if time permits we would also like to modify the inductive training strategy of GraphSAGE [**?**] to incorporate fairness.
Our objectives are to explore if existing graph embedding approaches can have their training strategy modified so that they are relatively more unbiased then their vanilla counterparts. We will assess our results by measuring their performance in terms of fairness metrics like statistical parity and downstream task metrics like F1 score.

1.2 Challenges

There has not been any existing work where the notion of unbiased random walks are used to generate embeddings for bipartite graphs as most of the existing work for graph embeddings and fairness have been done for homogeneous networks or through adversarial training. Also to the best of our knowledge, there has not been any work in unbiased inductive methods for bipartite graphs.

1.3 Impact

With the increase in usage of artificially intelligent systems around us, it becomes important that they are ethically fair towards the people who are impacted by it. For example, when developing intelligent analytics using networks representing real-world dynamics between people and the world around them, it is paramount that they do not learn any inherent social and cultural biases as this could adversely affected historically underprivileged or discriminated populations.
If we can develop such node embeddings for networks that are robust to bias and also information rich, they can be stored and can be used for downstream tasks with relative ease when compared to end-to-end learnt systems.

2 LITERATURE SURVEY AND BASELINES

One of the first attempts to study potential bias issues inherent within graph embeddings in a transductive way was when Rahman et. al. proposed Fairwalk [4], a fairness-aware embedding method that extended node2vec [2]. While node2vec uses a random walk over a graph to generate walk traces and then extracts features based on the learned traces, Fairwalk first partitions neighbors into groups based on their sensitive attribute values and gives each group the same probability of being chosen regardless of their sizes. Only then is a node from the chosen group selected at random for the walk. Fairwalk is designed for homogeneous networks and would invariably lose out on the semantics of bipartite (heterogeneous), whereas our algorithm would not treat nodes heterogeneous nodes identically.
metapath2vec [**?**] extends node2vec for heterogenous graphs by conditioning the sampling based on the node types and demonstrates using the heterogenous information provides improvements over naive homogenous approaches that ignore this.

Instead TODO REDUCE CONTENT AND ADD METAPATH2VEC, CLEARLY STATE DELIVERABLES

3 PLAN OF ACTION

3.1 Proposed Methodology for fair embeddings

In the original node2vec paper [2], each neighbour of a node has an equal chance of being sampled. In Fairwalk [4], the neighbours of a node are grouped according to different demographic features and now each group has an equal chance of being samples. If a group is sampled, then a neighbour in that group is chosen at random. We extend this idea to be used in the metapath2vec [**?**] approach.

Assume there are two types of nodes V_1 and V_2 in a bipartite graph G and V_1 has sensitive demographic attributes like gender, race, etc (usually a user) and V_2 has attributes like genre, etc. We can enforce an unbiased step from V_2 to V_1 such that there is an equal likelihood of sampling users from different demographics $v_1 \in N(v_2)$ where $v_1 \in V_1$ in the neighbourhood of the node $v_2 \in V_2$.

For V_1 to V_2:

$$p\left(v_1^{i+1} \mid v_2^{i},\right) = \begin{cases} \frac{1}{|N_{i+1}(v_2^i)|} & \left(v_1^{i+1}, v_2^{i}\right) \in E, \\ 0 & \left(v^{i+1}, v_t^{i}\right) \in E, \phi\left(v^{i+1}\right) \neq t+1 \\ 0 & \left(v^{i+1}, v_t^{i}\right) \notin E \end{cases}$$

For V_2 to V_1:

$$p\left(v_1^{i+1} \mid v_2 t^{i}, \mathcal{P}\right) = \begin{cases} \frac{1}{|N_{t+1}(v_e^i)|} & \left(v^{i+1}, v_t^{i}\right) \in E, \phi\left(v^{i+1}\right) = t+1 \\ 0 & \left(v^{i+1}, v_t^{i}\right) \in E, \phi\left(v^{i+1}\right) \neq t+1 \\ 0 & \left(v^{i+1}, v_t^{i}\right) \notin E \end{cases}$$

Since the Metapath2vec [**?**] approach is a transductive approach and this might not work well on unseen data, having unbiased graph embeddings approaches which are inductive (for example based on GraphSage [**?**] or Graph Attention Network [**?**]) would be nice. node2vec and metapath2vec could be thought of as DFS based approaches and GraphSAGE could be thought of as a BFS based approach.

However, to the best of our knowledge no work has been introduced in exploring unbiased ways for this.
Based on our learnings from node2vec based approaches for fairness, we would like to introduce an unbiased AGGREGATE step in GraphSAGE. We can also add a weighted component to the loss function that encourages similar representations to adjacent nodes where the weights would promote fairness.

3.2 Downstream Task - Recommendation system

We plan to use graph embeddings and evaluate their fairness on the recommendation task. For this we will be adopting a supervised approach, where our classifier will take as input two embedding vectors, one for user and one for item and predict the presence or absence of an edge between them . The presence of an edge implies recommendation and vice versa. We will be using the positive class probability as the recommendation score. An item v is recommended to a user u, if the recommendation score for the pair $(u, v) \notin E$ is within the top $k\%$ of all the scores received by all candidate pairs.The resulting set of recommendations given to a user u is denoted by $\rho(u)$.

3.3 Notations

We have a bipartite graph $\mathbf{G} = (V, E)$ where $V = V_1 \cup V_2$ and $V_1 \cap V_2 = \emptyset$, and the set of edges $E \subseteq \{(u, v) : u \in V_1, v \in V_2\}$.

A sensitive attribute is denoted by $\mathcal{S}$, for example $\mathcal{S} = g$ for gender. $\mathcal{Z}^{\mathcal{S}}$ denotes the set of all possible values of $\mathcal{S}$. The function $\zeta : V_1 \longrightarrow \mathcal{Z}^{\mathcal{S}}$ maps users to their attribute values. For example, for a male u, $\zeta^g(u) =$ "*male*".

A set of items recommended to user u as $\rho : V_1 \longrightarrow 2^{V_2}$.

For measuring fairness, user-item pairs $(u, v) \in V_1 \times V_2$ are partitioned into groups $G^{\mathcal{S}}_{ij}$ based on the attribute values of u, i.e. $G^{\mathcal{S}}_{ij} = \{(u, v) : \zeta(u) = i, u \in V_1, v \in V_2$. The set of all groups based on a sensitive attribute $\mathcal{S}$ is denoted by $\mathcal{G}^{\mathcal{S}}$.

3.4 Metrics

In this section, we explain the metrics we'll be employing for evaluating fairness of existing and our fairness-aware approach for recommendation. We also aim to evaluate the recommendation performance.

3.4.1 *Fairness Metrics.* [4]

- **Statistical Parity**- This measure requires the acceptance (recommendation) rates from two groups to be equal. Let $P(G^{\mathcal{S}}_{ij})$ denote the acceptance rate for group $G^{\mathcal{S}}_{ij}$, then

$$P(G^{\mathcal{S}}_{ij}) = \frac{|\{(u, v) : v \in \rho(u) \wedge (u, v) \in G^{\mathcal{S}}_{ij}\}|}{|G^{\mathcal{S}}_{ij}|}$$

  The bias or statistical parity between two groups is then the difference between their acceptance rates. Extending this to multiple groups it can be calculated as the variance of the acceptance rate of each group.

$$bias(\mathcal{G}^{\mathcal{S}}) = var(\{P(G^{\mathcal{S}}_{ij})\} : G^{\mathcal{S}}_{ij} \in \mathcal{G}^{\mathcal{S}})$$

- **Equality of Representation** - Recommendation systems often suffer from bubble effect where users get recommendations based on their interests, which reduces the diversity of recommendations. As a direct consequence this can lead to recommendations inhibiting minority representation. At the network level this can be measured my measuring bias between different groups, among all recommendations in the network.

$$bias^{ER}(\mathcal{G}^{\mathcal{S}}) = var(\{N(G^{\mathcal{S}}_{ij})\} : G^{\mathcal{S}}_{ij} \in \mathcal{G}^{\mathcal{S}})$$

  $N(G^{\mathcal{S}}_{ij}$ denotes number of recommendations from the group $G^{\mathcal{S}}_{ij}$

Property	Value
Number of nodes	2625
Number of edges	100000
Average degree	76.1905
Radius	3
Diameter	5
Density	0.02903

3.4.2 *Recommendation system metrics.*

We plan to evaluate the performance of recommendation system in terms of precision and recall.

3.5 Risks

A potential risk we anticipate is a drop in the performance when change the walk strategy. then we can do a tradeoff between fairness aormance grouping strategy might need experimentation since we have more than one demographic attribute

3.6 Dataset Description

We will primarily be using the MovieLens dataset [3], which is rating data set collected and made available from the MovieLens web site [] by GroupLens Research []. The data was collected from the website during the seven-month period from September 19, 1997 through April 22, 1998.

We intend to use the MovieLens 100K Dataset version of the dataset, recommended for education and development. The dataset consists of 100,000 ratings on a scale of 1 (lowest) to 5 (highest) from 943 users on 1682 movies. Each user has rated at least 20 movies. Simple demographic information such as age, gender, occupation, zip is also provided for each user, whereas the features provided for each movie include the movie title, release date, video release date, IMDb URL and genre.

We construct an undirected bipartite graph of movies and users as nodes and the edges between representing a user who rated a movie. The constructed graph has the properties described in Table 1 **??**:

If time permits, we could carry further experimentation on Freebase-15K-237 dataset, which is a subset of the Freebase-15K introduced by [1]

3.7 Existing Code

For metapth2vec we intend to use a PyTorch Geometric [] implementation of the algorithm as the codebase : source code from the *nn.models.metapath2vec* module and following example. For GrpahSage we intend to use a PyTorch implementation of the algorithm.

REFERENCES

[1] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, J. Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-relational Data. In *NIPS*.

[2] Aditya Grover and J. Leskovec. 2016. node2vec: Scalable Feature Learning for Networks. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016).

[3] F. M. Harper and J. Konstan. 2015. The MovieLens Datasets: History and Context. *ACM Trans. Interact. Intell. Syst.* 5 (2015), 19:1–19:19.

[4] Tahleen A. Rahman, Bartlomiej Surma, M. Backes, and Y. Zhang. 2019. Fairwalk: Towards Fair Graph Embedding. In *IJCAI*.