Fairness in Link Prediction Through Active Learning

Madhura Pawar
14386739

Abstract

Social media has become a prominent carrier of media and information. A major downside of these platforms is that they are suspected of creating "echo-chambers" due to their people recommender systems. The echo chambers lead to the amplification of extreme ideological opinions among like-minded individuals while marginalizing minor communities. We use active learning to steer the link recommender algorithm (Node2Vec) to generate link predictions in a way that minorities are well-represented within the community. We design active learning loop intending to improve the ability of minority nodes to become major information carriers. To assess the fairness towards minority nodes due to active learning we introduce two metrics to evaluate how important minority nodes become over time and how much they disperse within local communities to improve their representation. We find that in a strong majority homophilic network, active learning successfully intervenes to improve the representation of minority nodes. However, in a strongly minority homophilic network, active learning relies on minority node-size intervention to increase fairness towards minority nodes.

1 Introduction

Online Social Networks such as LinkedIn, Facebook, and Twitter are widely used by people. These platforms employ people recommender systems to increase the engagement of users ultimately affecting the growth of the network (Guy and Pizzato [2016]). These recommender algorithms are suspected to create "echo-chambers" which leads to the development and magnification of extreme ideological opinions among like-minded individuals. This can lead to a detrimental effect as it might affect the mindset of people due to the potential spreading of misinformation and targeting marginal communities. Cinelli et al. [2021] finds this effect to exist in the Facebook network. The formation of an echo chamber can be attributed to the fact that the recommender systems predict more homophilic links than heterophilic links. To unbias such a machine learning model, we have to assume the availability of (biased) labeled data in sufficient quantity (Barocas et al. [2018]). Since obtaining accurate labeled data is expensive, we might even obtain it in a limited amount. The main aim of the paper is to steer the recommendation model to generate more diverse recommendations so that people are exposed to varied viewpoints. Hence, we want to improve the fairness of the recommender model while keeping in note that the budget for obtained labeled data is bounded. Active learning is an ideal strategy for such a scenario. It sequentially chooses the unlabeled instances where their labeling is the most beneficial for the performance improvement of the ML model.

Active learning for link prediction: For social networks, in the case of a link prediction problem, the focus of active learning would be to query an oracle to obtain unconnected links so that the underlying link recommendation model improves its performance. The focus of this paper is thus to incorporate active learning and link prediction model, to not improve the accuracy of the link prediction model but to improve its fairness. The main research question we tackle is: *Can we improve the fairness of a popular link prediction algorithm - Node2Vec using active learning?*

We tackle this research question by first designing our components of active learning - how should we generate an unconnected edge pair through the oracle which might lead to fairer link-predictions, and what model is suitable to simulate an expert oracle? As our social networks, we use scale-free directed networks with adjustable homophily and minority group size (Espín-Noboa et al. [2022]). We then systematically design a feedback loop to observe the evolution of the social network with time and quantify the changes in connectivity for minority group over time. Ferrara et al. [2022] shows how the fairness towards minority nodes due to recommendation algorithms is dependent on homophily (the tendency to connect to similar others) and the proportion of minorities in the network.

To this end, we break down our main research question of whether active learning can improve fairness in link predictions into two sub-questions:

RQ1 Is the process of generating fairer link predictions by active learning affected by the homophily of the network?
RQ2 Does the size of the minority group in the network affect the active learning process?

Our contribution is as follows: (1) We design our active learning loop while taking into account that querying through expert should lead to improvement in the fairness of minority nodes, (2) We quantify *fairness* by introducing two metrics which measure how important minority nodes become over time and how well they disperse into locally-formed communities, (3) We find that in a network where majority nodes prefer to connect with other majority nodes (a driving factor to creation of echo-chambers), active learning successfully improves representation of minority nodes.

We discuss relevant literature and useful background concepts in Section 2. We talk about our active learning methodology and evaluation metrics in Section 3. Finally, we discuss results after the application of active learning on social networks in Section 4. We conclude the discussion of our results and possible future directions of this work in Section 5.

2 Background & Related Work

We organize the related work in two parts. We introduce the relevant literature on different mechanisms that incorporate active learning in link prediction tasks. Next, we discuss the existing literature that studies the effects of link prediction algorithms on fairness towards the minority nodes.

2.1 Related Work

In active learning, the aim is to select unlabeled data samples that improve the model's performance the most. This selection is done by an *acquisition function* that decides which data points to ask an oracle to label. It selects one or more points from a pool of unlabeled data points. These samples are then labeled by an oracle and then used to retrain the model. The unlabeled data sample is chosen by acquisition function by measuring the model's uncertainty in making predictions for that sample. Katsimpras and Paliouras [2023] uses active learning to select highly uncertain unlabeled nodes to be labeled to improve the learning of Graph Neural Networks. Ostapuk et al. [2019] employs active learning to incrementally train deep neural networks for link prediction on knowledge graphs. Chen et al. [2020] proposed an active learning framework to improve the accuracy of link prediction models by proposing various query strategies based on network embedding. Chen et al. [2014] proposes a hybrid active learning approach for link prediction problem that exploits the local structure and the global structure of the network. The focus of all the above works is improving the accuracy of the link-recommender model. However, we focus on improving the fairness of the link recommender model. Our work sits at the intersection of three topics: active learning, link prediction, and fairness. Our notion of *fairness* shall try to increase the involvement of minority nodes in the process of formation of links over time in the social network. The more links pass through minority nodes, the more authoritative gain they can have in the network wrt information being transmitted over the edges. This can help them to improve their representation in the network. Here, we devise our active learning strategy such that our acquisition function outputs minority nodes such that their role as an information carrier is strengthened. The oracle then would use their expert knowledge to "rewire" these minority nodes in the network to ideally, increase their representation.

Cinus et al. [2021] shows that people recommenders lead to a significant increase in echo chambers on social graphs generated by random network models with tunable homophily. Fabbri et al. [2021] propose a model to simulate the feedback loop created by multiple rounds of interactions between users and a link recommender in a social network. Their results show that a minority group, if homophilic enough, can get a disproportionate advantage in exposure from all link recommenders.

Table 1: Table summarizing information about the generated graphs

Graph	Scenario	h_{MM}	h_{mm}
G0	Hete/Homo	0.2	0.8
G1	Neutral	0.5	0.5
G2	Homo/Hete	0.8	0.2

Instead, when it is heterophilic, it gets under-exposed. To truly understand how effective is our fairness-driven active learning strategy, we test its intervention along different combinations of homophily and minority sizes. We quantify the notion of *fairness* based on its influence: how much influence do minority nodes have, to enable communication between disconnected groups of nodes?

As part of the background, we describe the networks used in our experiments, the link prediction algorithm we employ (Node2Vec), and some relevant concepts that we use to design our fairness metrics.

2.2 Background

Synthetic Networks: We consider node-attributed directed networks, $G = (V, E, C)$ where $\{v_1, v_2, ..v_n\}$ is a set of n nodes. $E \in V \times V$ is a set of unweighted directed edges, and $C : V \to \{0, 1\}$ is a function that maps each node v_i into its group (or class) membership c. We consider binary group membership hence the function C, divides the nodes into two groups, a minority, called m, and a majority, called M. We refer to the fraction of the minority group in the network as f_m.

Homophily: Homophily can be defined as the tendency of nodes to connect with similar other nodes McPherson et al. [2001]. Thus, the probability that a source node v_i connects to a target node v_j is driven by the homophily between their classes c_i and c_j. Homophily values range from 0.0 to 1.0. If the homophily value is high, that means that nodes of the same class are attracted to each other more often than nodes of different class. We stick to the same notation of homophily as used by Espín-Noboa et al. [2022] where h_{ab} denotes the homophily to connect nodes from class a to class b and h_{aa} is the homophily of connecting nodes in the same class a. Nodes of the same class with homophily $h_{aa} = 0.5$ are referred to as neutral (i.e., they connect randomly to either class), otherwise, they are heterophilic if $h_{aa} < 0.5$ (i.e., more likely to connect to the other class), or homophilic when $h_{aa} > 0.5$ (i.e., more likely to connect to the same class). We note that in- and between-class homophily values are complementary: $h_{mm} = 1 - h_{mM}$ and $h_{MM} = 1 - h_{Mm}$.

To systematically create networks we employ the DPAH model (Espín-Noboa et al. [2022]). This model allows to generate scale-free bi-populated directed networks with adjustable homophily (for each group), minority size, and edge density. The probability of creating a directed link from v_i to v_j is defined as:

$$\mathbb{P}(v_i \to v_j) = \frac{h_{ij} k_j^{in}}{\sum_{l=1}^{n} h_{il} k_l^{in}} \tag{1}$$

where k_j^{in} is the in-degree of v_j, and h_{ij} is the homophily between v_i and v_j. We modify the homophily within groups and the size of the minority. In particular, to answer **RQ1** we generate 3 networks for combinations of homophily parameters as per Table 1 and fix the number of nodes $N = 1000$, the size of the minority $f_m = 0.3$, and the edge density $d = 0.03$. We further adjust the size of the minority $f_m \in \{0.2, 0.3, 0.4, 0.5, 0.6\}$ to answer **RQ2**.

Link Prediction: For each node $v_i \in V$, the recommendation algorithm suggests a ranked list of k nodes that v_i is not yet connected with. The ranked list is sorted in descending order in terms of relevance scores. We create a direct link $v_i \to v_j$ for each top-1 of these recommendations. By doing so, we create a new out-link for each node v_i. For every addition, we remove a random out-link. We do this to prevent a significant increase in the edge density of the network as done by Cinus et al. [2021]. We use Node2Vec (N2V), a popular embedding algorithm that maps nodes to a low-dimensional space of features. The low-dimensional space is found by the use of random walks in the graph with the objective of maximizing the likelihood of preserving nodes' neighborhoods

(Grover and Leskovec [2016]). N2V recommends to each node v_i the most similar node in the embedding space, according to the cosine similarity of the embedded vectors.

Betweenness Centrality: Betweenness centrality measures the extent to which a vertex lies on paths between other vertices (Freeman [1978]). The vertices with the highest betweenness are also the ones whose removal from the network will most disrupt communications between other vertices because they lie on the largest number of paths taken by messages. We use betweenness centrality to design our acquisition function and to find the importance of minority nodes in Sections 3.1 and 3.2.1 respectively.

Community Detection: Community detection involves identifying groups of nodes or entities within a network that are more densely connected to each other than to the rest of the network. The Louvain algorithm (Blondel et al. [2008]) is a popular and efficient method used for community detection. The Louvain algorithm is based on the idea of optimizing a measure called modularity. Modularity quantifies the quality of a community structure by comparing the number of edges within communities to the expected number of edges if the network were randomly connected. We use the Louvain algorithm for community detection to find how much minorities spread in different communities over time (Section 3.2.2).

3 Method(s)

3.1 Active Learning

To reiterate, our aim for using active learning is to steer our link recommender to generate fairer recommendations. In the case of active learning where the focus is to improve the model's performance, the idea is to design the acquisition function in such a way, so that the oracle annotates a datalabel the model finds the most uncertain. Here we aim to design the acquisition function in such a way that the oracle annotates the **least-important** minority nodes. We assume that the least-important nodes are the ones that have the least contribution to transmitting information, i.e., the ones that lie on the least number of shortest paths between any two vertices. Thus, our acquisition function finds the minority nodes with the least betweenness centrality, giving us:

$$\{u_1, u_2.., u_B\} = \arg\min \mathcal{A}(\mathcal{G}) = \arg\min \left(\text{bet}(u_i)\ \forall c_i = 1, i \in [1, N]\right) \tag{2}$$

where $\text{bet}(u)$ is the betweenness centrality of node u and nodes for which class $c = 1$ are the minority nodes.

We use the batch active learning setting where instead of labeling one unlabeled sample, we label B samples at every iteration. The acquisition function $\mathcal{A}$ thus outputs B minority nodes. Now we want to generate links for these B nodes which are the most unimportant ones, and improve their representation. We then ask the human annotator to connect these B nodes to other nodes in such a way (preferably to very centric majority nodes), that their representation in the network improves. In a real-life setting, maybe we have an oracle who is provided with user-interface with graph visualizations so that they can pick the best vertices to connect our least-important vertices to. Since we do not have a real-human annotator, we simulate one using Fairwalk (Rahman et al. [2019]).

Fairwalk: Fairwalk modifies the random walk procedure from the original node2vec. They conduct a random walk for a target node by partitioning neighbors into groups based on their sensitive attribute values (in this case, the sensitive attribute of a node is belonging to a majority or minority class) and give each group the same probability of being chosen. Then a random node from the chosen group is selected for the jump. Rahman et al. [2019] demonstrates the existence of bias in node2vec when used for friendship recommendation. We need our oracle to help guide node2vec towards fairer recommendations, hence we choose this particular fairness-aware embedding method to simulate the oracle.

Fairwalk predicts links for every node in the social network. N2V, which is our original link recommender model then learns embeddings on this graph and predicts links for every node. We then apply our evaluation metric from Section 3.2 to assess how better N2V is generating recommendations. This process is then repeated for $\mathcal{T}$ timesteps. Algorithm 1 summarizes the steps for one whole iteration of active learning. Figure 1 sketches out the steps sequentially followed by our active learning procedure.

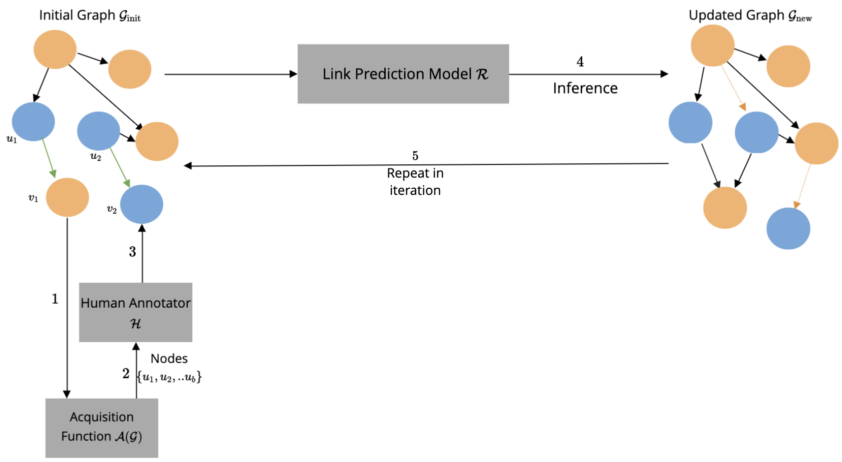


Figure 1: Workflow Diagram for Fairness-Driven Active Learning.The numbering near arrows denote the sequence of the steps. Arrows in green denote the sample edges generated by annotator $\mathcal{H}$. Arrows in dotted-orange denote the edges predicted by our link prediction model $\mathcal{R}$.

Algorithm 1: Fairness-Driven Active Learning

Data: An initial directed graph $\mathcal{G}_{\text{init}} = (\mathcal{V}, \mathcal{E})$. Link Recommder Model $\mathcal{R}$. Batch size B. Time-steps $\mathcal{T}$.
Acquisition Function $\mathcal{A} : G \rightarrow \{u_1, u_2.., u_B\}$.
Human Annotator / Oracle $\mathcal{H} : \{u_1, u_2.., u_B\} \rightarrow \{(u_1, v_1), (u_2, v_2), .., (u_B, v_B)\}$
Result: An updated graph with links predicted by recommender $\mathcal{R}$

1 $\mathcal{A}$ computes a set of minority nodes $U = \{u_1, u_2.., u_B\}$ given $\mathcal{G}_{\text{init}}$ as input ;
2 $\mathcal{H}$ is queried and it generates annotations by adding an edge with source node u_i and target node $v_i, \quad \forall i \in B$ in $\mathcal{G}_{\text{init}}$;
3 Recommender Model $\mathcal{R}$ predicts 1-top link for every node in $\mathcal{G}_{\text{init}}$, new links are added giving new updated graph $\mathcal{G}_{\text{new}}$;
4 The new graph $\mathcal{G}_{\text{new}}$ is now set as the base graph $\mathcal{G}_{\text{init}}$ ($\mathcal{G}_{\text{init}}$ = $\mathcal{G}_{\text{new}}$) ;
5 Repeat steps 1-4 for more $(\mathcal{T} - 1)$ timesteps ;

Hyper-parameters: For our batch active learning setting, we set the batch size, $B = 75$. We experimented with various batch sizes $B \in \{50, 75, 100, 200\}$ and found that the batch size of 75 was the minimum batch size on which we see improvement in fairness. In a real-life setting, we would prefer working with a minimal budget to obtain queries by annotators, so we chose $B = 75$. In N2V and Fairwalk, we use the default values for the dimensions of the embedding space *dimensions* = 64, the number of visited nodes in each random walk *walk _length* = 10, and the number of random walks to be generated from each node in the graph *num_walks* = 200.

3.2 Evaluation Metrics

3.2.1 Importance of Minority Nodes

We assume that a node is "important" if it has considerable influence within a network by its control over information passing i.e., it has high betweenness. If minority nodes have high betweenness over time, then it would make them one of the major information carriers in the network, thus allowing them to improve their representation. We calculate the importance at time t as the average betweenness centrality of minority nodes:

$$\mathcal{B}_t = \frac{\sum_{i=1}^{m_0} \text{bet}(u_{it})}{m_0} \tag{3}$$

where m_0 is the number of minority nodes and $\text{bet}(u_{it})$ denotes the betweenness centrality of i^{th} minority node u at time t. To further understand the trends in the change of importance of minority nodes over time, we rely on the edge-link ratio introduced by Ferrara et al. [2022]. The edge-link ratio would help us to understand how predictions are changing over time which affects the importance of minority nodes. For two classes a, b the in-link ratio for class a is the fraction of links within that particular group, given as:

$$\mathcal{I}_a = \frac{e_{aa}}{e_{aa} + e_{ab}} \tag{4}$$

whereas the out-link ratio for class a is the fraction of links formed with a different group, class b in this case:

$$\mathcal{O}_a = \frac{e_{ab}}{e_{aa} + e_{ab}} \tag{5}$$

3.2.2 Dispersion of Minorities in Communities

A community is defined as a subset of nodes within the graph such that connections between the nodes are denser than connections with the rest of the network (Radicchi et al. [2003]). With an increase in the link predictions over time, there is a possibility that some nodes become densely connected, leading to the formation of communities. It would be ideal if minority nodes were dispersed more into the communities of the majority nodes so that locally they can increase their importance, as well as have influence on information transmission within that community. To assess whether active learning helps to increase minorities' visibility within the community, (1) we apply the Louvain community detection algorithm which assigns each node a community, (2) If $\{C_1, C_2, ..C_p\}$ are p communities formed we find the percent of minorities within say one community C_p as:

$$\%C_p = \frac{\text{\# of minority nodes in } \ C_p}{\text{\# of total nodes in } \ C_p} \times 100.0 \tag{6}$$

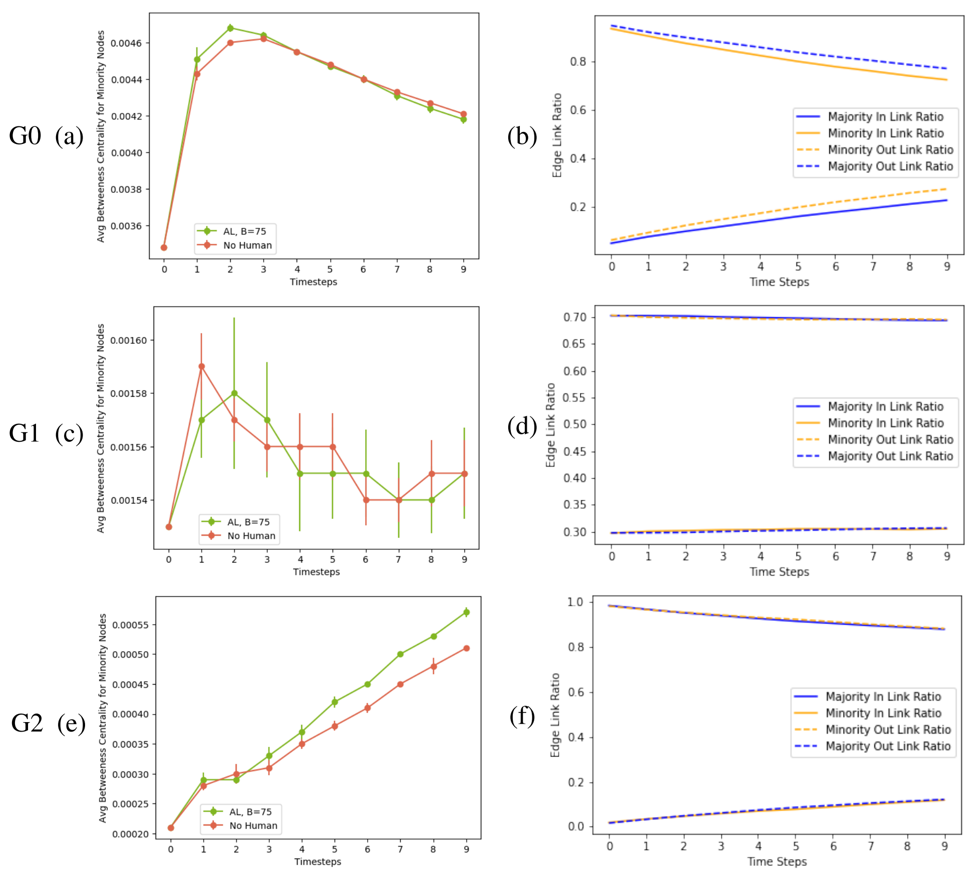


Figure 2: Left column signifies the variation of the importance of nodes with time. The Red Line denotes recommendations without active learning, while the green line denotes when recommendations are done with active learning. f_m is set to 0.3 The right column shows in and out edge-link ratios for recommendations generated with active learning.

4 Experimental Results

To answer **RQ1**, we vary the configurations of homophily of majority and minority (h_{MM} and h_{mm}) for three settings as per Table 1 for a fixed minority fraction, $f_m = 0.3$. We run each experiment for two configurations - without using active learning and with using active learning. To answer **RQ2**, we vary the size of minority $f_m \in \{0.2, 0.3, 0.4, 0.5\}$. We run each experiment across three random seeds. The experimental work for this paper can be found at `https://github.com/madhu221b/himl-link-prediction`.

4.1 RQ1: Active Learning And Homophily

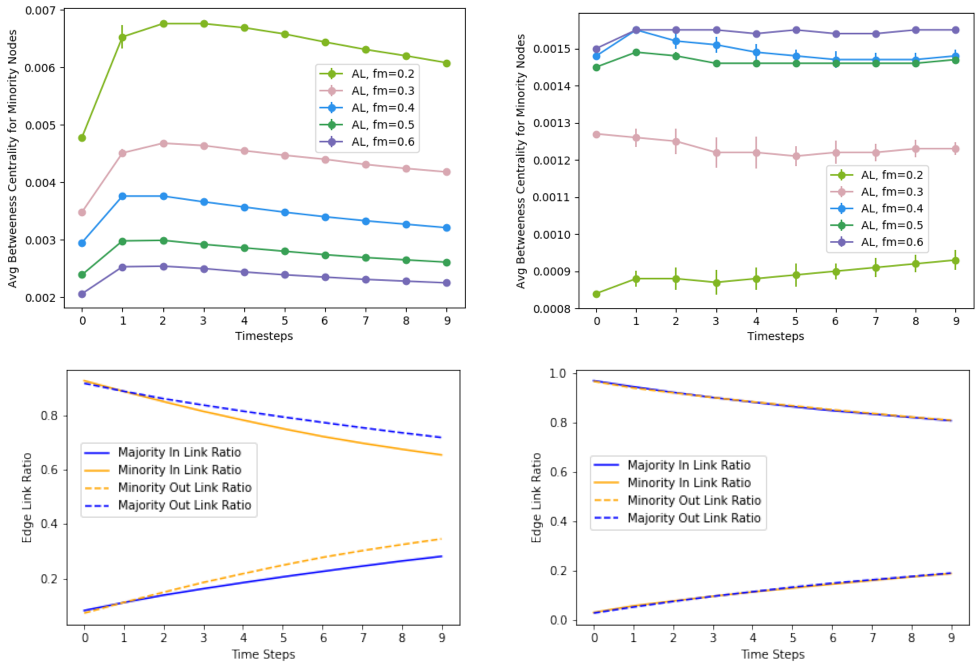
The importance which is the betweenness centrality is dependent on: (1) how many majority nodes connect with minority nodes (majority out-link ratio), (2) how much minority nodes connect within their own group as opposed to connecting to majority nodes (minority out and in link ratio). The importance of nodes would be affected by the initial homophily of the graph. We expect that ideally, active learning should be able to improve the importance of minority nodes if they are underrepresented in a particular network. We now analyze the results of variation of average betweenness of minority nodes (importance of minority nodes) with time. We also stuy how dispersed minorities become in local communities through active learning (Section 3.2.2).

The minority is homophilic and majority is heterophilic for G0 ($h_{MM} = 0.2, h_{mm} = 0.8$). We see that in this case, the importance of minority nodes decreases for both the scenarios: no-active learning setting and active learning setting as seen in Figure 2(a). As seen in Figure 2(b), we see that active learning helps with decreasing the minority in-links ($m \to m$) and increasing minority out-links ($m \to M$). However, this change is quite gradual and slow over the $\mathcal{T}$ timesteps. Thus, the corrective behavior is incorporated by the active learning but the initial strong homophily of the minority prevents the annotator from generating more fairer links. Our finding is corroborated by looking at the percent of minorities in local communities for G0 in Figure 4(h) for $T = 9$ timestep. We see that we have a high accumulation of minorities in local communities (high-density count for communities with more than 30% minority population). That means active learning did not help with the dispersion of minorities. Thus, we conclude that **for a strong minority homophilic scenario, active learning did not help to generate fairer link predictions**. Can fairness intervention then be helped by changing the initial conditions of minority groups to improve their representation? We explore this question further by varying the sizes of minority nodes for G0 and discuss the results in Section 4.2.

In G1 ($h_{MM} = h_{mm} = 0.5$), nodes of either group have an equal probability of forming links with nodes of either group. Our simulated human annotator (Fairwalk) biases its random walks in such a way that a node from every group has an equal probability of getting visited. Due to this similar impartial setting of both the Fairwalk and neutral conditions of homophily, we do not expect Fairwalk to debias embeddings generated by N2V. This can be seen in Figure 2(c) where the average betweenness by active learning does not significantly change compared to the average betweenness of a no active-learning setting. We also see this wrt dispersion of minorities in the communities for G1 (Figures 4(l) and 4(p)) where both active and non-active settings have the minorities within communities balanced out.

The majority is homophilic and minority is heterophilic for G2 ($h_{MM} = 0.8, h_{mm} = 0.2$). In such a scenario, it is very prone for the network to form echo chambers as majority nodes prefer connecting to other majority nodes rather than minority nodes. Ideally, active learning should help try to steer this homophilic behavior of majority nodes, towards encouraging more out-links from majority to minority. We see in Figure 2(e), that active learning helps to increase the importance of minority nodes over time. Our finding is also supported by seeing how the minorities are getting dispersed and not accumulated in different local communities as seen in Figures 4(w) - 4(x). We thus conclude that **for a strongly majority homophilic scenario, active learning leads to fairer link predictions.**

(a) $h_{MM} = 0.2, h_{mm} = 0.8$ (b) $h_{MM} = 0.8, h_{mm} = 0.2$



(c) $f_m = 0.2$ (d) $f_m = 0.5$

Figure 3: The top row denotes variation of minority size $f_m \in \{0.2, 0.3, 0.4, 0.5, 0.6\}$ for G0, and G2. The bottom row denotes the edge-link ratio for G0, and G2 for $f_m = 0.2$, 0.5 respectively.

4.2 RQ2: Active Learning and Variation in Minority Size

We vary minority size f_m for the two extreme scenarios - G0 (with strong minority homophily) and G1 (with strong majority homophily).

We saw in Section 4.1 that for G0, active learning does not help to generate fairer recommendations with time. We then vary the size of the minority in G0, to assess whether changing f_m helps active learning. In Figure 3(a), we observe that decreasing f_m in G0 helped active learning to improve the importance of minority nodes. We see a steeper increase in the rate of minority out-links ($m \rightarrow M$) predicted over time in Figure 3(c). To reduce the effect of strong homophily within the minority, we reduced its fraction in the network which helped active learning to increase its heterophily behavior towards majority nodes. Thus, we conclude that for a **strong minority homophily scenario, active learning requires node-size intervention to generate fairer link predictions**.

In Figure 3(b) we see that increasing f_m, leads active learning to increase the importance of minority nodes more than what it already improved when $f_m = 0.3$. This might be because in a majority homophilic network, increasing f_m might help active learning to be more representative towards the minority as it has more opportunities to generate links for more minority nodes.

5 Conclusions, Limitation, & Future Work

To summarise our results, we see active learning help steer the link recommender model to generate fairer link predictions in a strongly majority homophilic network. This finding is quite important because of its potential to de-bias an exacerbating partiality towards the majority group. Active learning can help to reduce the echo chambers systematically and raise the minority representation in the network. However, we also observe that active learning did not prove fruitful to a strongly minority homophilic network. It wasn't able to decouple the minority-minority edges and encourage heterophilic links in a substantial amount. It was after we dabbled with the initial configuration of the network (decreasing f_m), which helped boost active learning to improve the visibility of

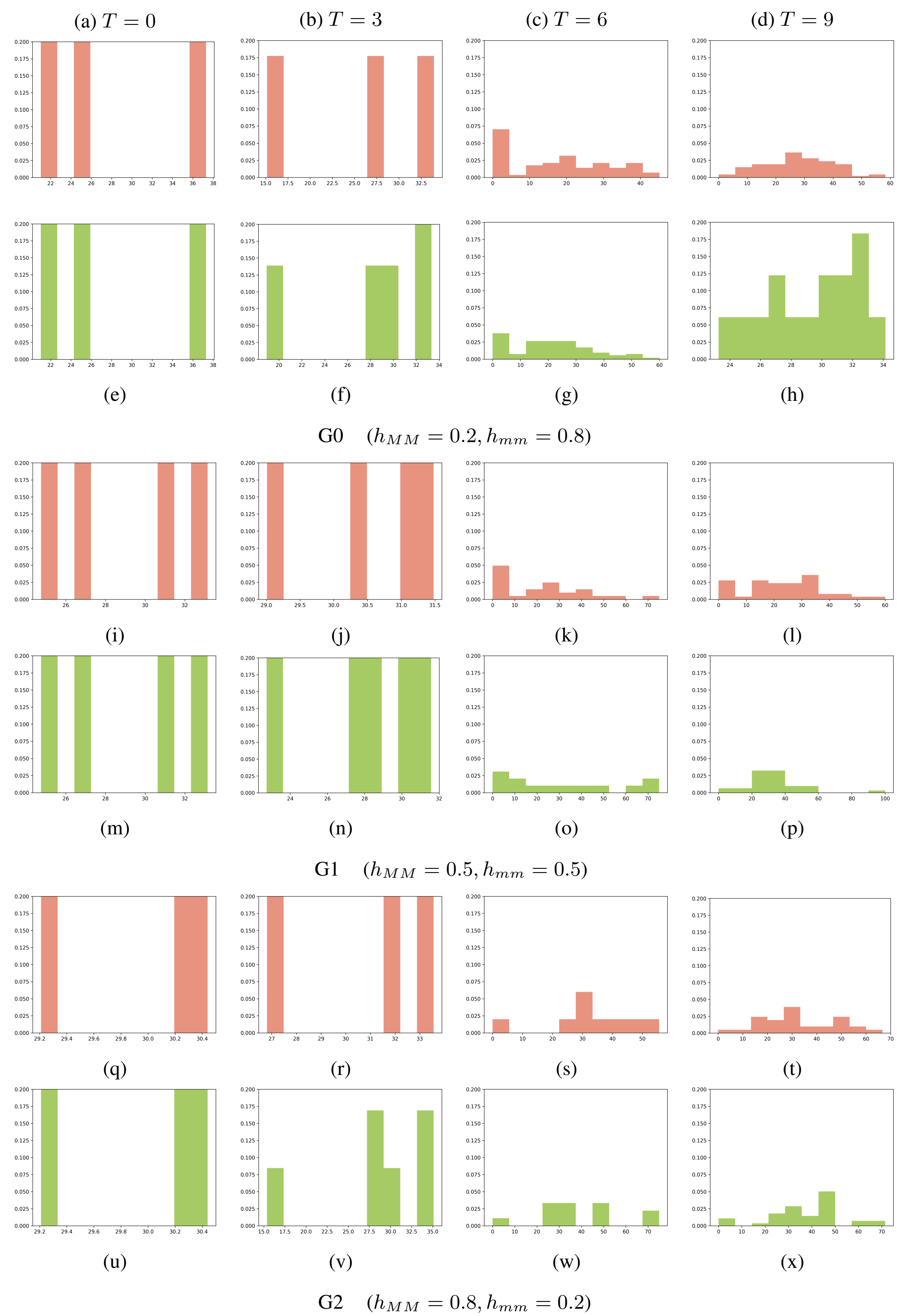


Figure 4: Histogram of the percent of minorities in communities. The $x-$axis denotes value of $\%$ of minorities (C_p) for p communities while the $y-$axis denotes the density-count of (C_p) for p communities with T. The histograms in red are the graphs that evolved without active learning while green are the ones that used the active learning component.

minority nodes. Thus, we see that there is no one-size-fits-all approach for different configurations of homophily for social networks. As future work, our acquisition function can be studied to make it more robust in different settings of homophily. Also, the right balance between the extent of active learning and that of the node-size intervention can be further studied too.

We see a promising avenue of incorporating active learning for link predictions to improve the representation of minorities in online social networks. Currently, we are conducting our experiments on synthetic networks. It would be interesting as well as important to test out this active learning methodology on real-time social network datasets too. Also, since our focus was on addressing the fairness of the link-recommender model, an insightful task would be to conduct an accuracy-fairness tradeoff while using active learning. We used a distance-based recommendation model (cosine similarity to find top-1 recommendation). To assess the accuracy of link recommender, we can use a model-based recommendation system where we generate node embeddings through N2V and train a binary classifier to predict links.

References

Solon Barocas, Moritz Hardt, and Arvind Narayanan. Fairness and machine learning limitations and opportunities. 2018. URL `https://api.semanticscholar.org/CorpusID:113402716`.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008 (10):P10008, October 2008. ISSN 1742-5468. doi: 10.1088/1742-5468/2008/10/p10008. URL `http://dx.doi.org/10.1088/1742-5468/2008/10/P10008`.

Ke-Jia Chen, Jingyu Han, and Yun Li. Hallp: A hybrid active learning approach to link prediction task. *J. Comput.*, 9:551–556, 2014. URL `https://api.semanticscholar.org/CorpusID:12610062`.

Xi Chen, Bo Kang, Jefrey Lijffijt, and Tijl De Bie. Alpine: Active link prediction using network embedding. *ArXiv*, abs/2002.01227, 2020. URL `https://api.semanticscholar.org/CorpusID:211020989`.

Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118, 2021. URL `https://api.semanticscholar.org/CorpusID:232038574`.

Federico Cinus, Marco Minici, Corrado Monti, and Francesco Bonchi. The effect of people recommenders on echo chambers and polarization. In *International Conference on Web and Social Media*, 2021. URL `https://api.semanticscholar.org/CorpusID:244773073`.

Lisette Espín-Noboa, Claudia Wagner, Markus Strohmaier, and Fariba Karimi. Inequality and inequity in network-based ranking and recommendation algorithms. *Scientific Reports*, 12(1), February 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-05434-1. URL `http://dx.doi.org/10.1038/s41598-022-05434-1`.

Francesco Fabbri, Maria Luisa Croci, Francesco Bonchi, and Carlos Castillo. Exposure inequality in people recommender systems: The long-term effects. *ArXiv*, abs/2112.08237, 2021. URL `https://api.semanticscholar.org/CorpusID:245144692`.

Antonio Ferrara, Lisette Espin-Noboa, Fariba Karimi, and Claudia Wagner. Link recommendations: Their impact on network structure and minorities. In *14th ACM Web Science Conference 2022*, WebSci '22. ACM, June 2022. doi: 10.1145/3501247.3531583. URL `http://dx.doi.org/10.1145/3501247.3531583`.

Linton C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1: 215–239, 1978. URL `https://api.semanticscholar.org/CorpusID:751590`.

Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016. URL `https://api.semanticscholar.org/CorpusID:207238980`.

Ido Guy and Luiz Pizzato. People recommendation tutorial. In *Proceedings of the 10th ACM Conference on Recommender Systems*, RecSys '16, page 431–432, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450340359. doi: 10.1145/2959100.2959196. URL `https://doi.org/10.1145/2959100.2959196`.

Georgios Katsimpras and Georgios Paliouras. Improving graph neural networks by combining active learning with self-training. *Data Mining and Knowledge Discovery*, 2023. URL `https://api.semanticscholar.org/CorpusID:260852655`.

Miller McPherson, Lynn Smith-Lovin, and James M. Cook. Birds of a feather: Homophily in social networks. *Review of Sociology*, 27:415–444, 2001. URL `https://api.semanticscholar.org/CorpusID:2341021`.

Natalia Ostapuk, Jie Yang, and Philippe Cudré-Mauroux. Activelink: Deep active learning for link prediction in knowledge graphs. *The World Wide Web Conference*, 2019. URL `https://api.semanticscholar.org/CorpusID:86386713`.

Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101 9:2658–63, 2003. URL `https://api.semanticscholar.org/CorpusID:2980810`.

Tahleen A. Rahman, Bartlomiej Surma, Michael Backes, and Yang Zhang. Fairwalk: Towards fair graph embedding. In *International Joint Conference on Artificial Intelligence*, 2019. URL `https://api.semanticscholar.org/CorpusID:199466038`.