Fair-LS: A Group-Specific Clamping Factor for Fair Label Spreading in Graph-Based Semi-Supervised Learning

Willian Dihanster Gomes de Oliveira[1], **Lilian Berton**[1]

[1]Institute of Science and Technology
Federal University of São Paulo - UNIFESP
São José dos Campos – SP – Brazil

{*dihanster.willian, lberton*}@unifesp.br

Abstract. *Graph-based semi-supervised learning methods, such as Label Spreading, can propagate bias when the graph structure reflects group disparities. We propose Fair-LS, an extension of Label Spreading that introduces group-specific clamping factors to control label diffusion differently for privileged and unprivileged groups. Experiments on benchmark datasets, Adult Census Income, Compas Recidivism, and German Credit, show that Fair-LS could improve fairness metrics (Disparate Impact and Average Absolute Odds Difference) with minimal reduction in AUC-ROC. The results suggest that adapting propagation dynamics by group is an effective strategy to reduce bias in semi-supervised learning.*

1. Introduction

Fairness in machine learning has become a critical area of research due to the increasing deployment of automated decision systems in domains such as finance, healthcare, and criminal justice [Barocas et al. 2023, Rabonato and Berton 2025]. These systems often rely on historical data that may reflect structural inequalities, leading to biased predictions that disproportionately affect protected groups defined by attributes such as gender, race or age, for example.

The intersection of semi-supervised learning (SSL) and fairness remains a relatively underexplored area within machine learning research [Zhang et al. 2023, de Oliveira and Berton 2024]. The core issue lies in the potential for the limited labeled data to contain inherent biases, which can then be amplified and propagated throughout the unlabeled data during the semi-supervised training process. This can lead to discriminatory outcomes where the model's predictions disproportionately harm certain demographic groups.

In particular, Graph-Based Semi-Supervised Learning (GSSL) algorithms [Chong et al. 2020, Song et al. 2022], which leverage both labeled and unlabeled data connected via similarity graphs, are susceptible to propagating and even amplifying such biases. The topology of real-world graphs often exhibits imbalances, such as group-level homophily or uneven degree distributions, which can distort label propagation and lead to disparate model behavior across groups [Yang et al. 2023, de Oliveira and Berton 2023].

Label Spreading, a classical GSSL algorithm, propagates label information through the graph using a fixed clamping factor that determines how strongly nodes retain their initial labels versus adapt to their neighbors [Zhou et al. 2003]. However, applying a single clamping value uniformly to all nodes ignores structural and statistical disparities between groups, potentially reinforcing unfair outcomes.

In this work, we propose Fair-LS, a fairness-aware extension of Label Spreading that introduces group-specific clamping factors. By allowing different propagation dynamics for privileged and unprivileged groups, Fair-LS offers a flexible mechanism to reduce bias in label propagation. We evaluate our method on three widely used fairness benchmarks, Adult Census Income, Compas Recidivism, and German Credit, across multiple protected attributes. Results show that Fair-LS could improve fairness metrics (Disparate Impact and Average Absolute Odds Difference) while maintaining competitive classification performance (AUC-ROC), supporting the use of group-adaptive propagation in fair semi-supervised learning.

The remainder of the work is organized as follows: Section 2 details the related works; Section 3 presents concepts of graph theory, graph-based semi-supervised learning, the proposed method and experimental methodology employed in this work; Section 4 presents the results obtained, as well as a general discussion; Finally, Section 5 presents the conclusions obtained and future work.

2. Related Work

Previous studies have focused on integrating fairness considerations into semi-supervised learning (SSL) frameworks, which are crucial when labeled data is scarce. [Yang et al. 2023] present a fair semi-supervised learning framework that addresses the challenge of maintaining fairness when only a few, potentially biased, labeled examples are available. Authors propose a representation learning architecture that achieves both fair and accurate classification, even under data scarcity. [Zhang et al. 2020] propose an optimization-based approach that balances classifier accuracy, label propagation, and fairness constraints across both labeled and unlabeled data. Their method, applied to Logistic Regression and SVMs, is evaluated using fairness metrics like Disparate Impact and Disparate Mistreatment. [Chakraborty et al. 2021] propose using situation testing to filter out potentially unfair labels and build a balanced initial labeled set in semi-supervised learning. While this approach helps reduce label bias by identifying more reliable labeled examples, it depends on external predictive models. As a result, it may lead to the exclusion of potentially useful data, particularly in low-label scenarios, and its effectiveness can vary depending on the choice and performance of the models involved.

Given that Label Spreading operates on graph structures, efforts to imbue graph-based algorithms with fairness are particularly relevant. [Sulaiman and Roy 2025] introduces GFLC (Graph-based Fairness-aware Label Correction), a novel method designed to address the challenge of label noise in biased training data while maintaining demographic parity in machine learning models. [Miyazaki and Sumikawa 2018] presents a graph-based multi-label classification algorithm designed to improve collaborative labeling of digital data, even when some relevant labels are missing. The proposed method updates label weights by referencing the top-k most similar instances in each step. It is not fairness-focused but explored a modification of the clamping mechanism in the label propagation step.

Beyond fairness, modifications to the core mechanics of Label Propagation have been explored to enhance its performance or adaptability. [Tsang et al. 2019] introduce formal fairness definitions grounded in legal and game-theoretic principles and propose an algorithmic framework that enforces fairness constraints while optimizing influence. Using real-world data from an HIV prevention program for homeless youth, they demonstrate that traditional methods often marginalize smaller demographic groups, whereas their approach significantly reduces disparities by ensuring more equitable influence distribution.

Our proposed Fair-LS method extends these lines of research by specifically introducing group-adaptive clamping factors into the Label Spreading algorithm, allowing for nuanced control over label propagation dynamics to address fairness directly within semi-supervised learning. This approach dynamically recalculates potentially biased labels through the adjustable parameter, enabling correction without discarding valuable examples.

3. Materials and Methods

3.1. Graph Fairness

Fairness has become an increasingly important concern in machine learning, especially in scenarios where automated decisions impact sensitive groups defined by attributes such as sex, race, age or socioeconomic status. In graph-based tasks, including semi-supervised node classification, community detection, and recommendation systems, the underlying network structure can amplify existing social biases, making the study of fairness in graphs particularly critical.

In graph learning, the connectivity between nodes directly influences how information propagates and how models generalize. However, real-world networks often exhibit topological bias, such as higher intra-group edge density (homophily) among privileged groups. This can result in biased information propagation, where performance disparities emerge between sensitive groups.

3.2. Graph-Based Semi-Supervised Learning

Graph-Based Semi-Supervised Learning (GSSL) is a family of methods that leverage both labeled and unlabeled data by modeling their relationships as a graph. In GSSL, each instance (labeled or unlabeled) is represented as a node, and edges encode similarity or affinity between instances, typically defined through distance metrics or domain-specific knowledge. The core assumption in GSSL is the manifold smoothness or label smoothness principle: nearby nodes in the graph (i.e., highly similar instances) are likely to share the same label. This enables the model to propagate information from a small set of labeled nodes to a large set of unlabeled ones through the graph structure.

Label Spreading (LS) [Zhou et al. 2003] is a graph-based semi-supervised learning algorithm designed to propagate label information from a small set of labeled nodes to a larger set of unlabeled ones using the structure of a similarity graph. It builds on the principles of Label Propagation [Zhu et al. 2002], but introduces additional regularization to improve stability and mitigate overfitting to the initial labels. The method operates by constructing a graph where each node represents a data instance, and edges encode pairwise similarity, typically computed using a radial basis function (RBF) or a k-nearest neighbors (KNN) scheme.

Formally, let W be the affinity matrix of the graph, normalized to ensure proper scaling of the influence between nodes. The algorithm iteratively updates a soft label distribution matrix F, where each row corresponds to a node and each column to a class. The update rule is defined as:

$$F^{(t+1)} = \alpha S F^{(t)} + (1 - \alpha)Y,$$

where:

- S is a normalized similarity matrix (often computed as the symmetric normalized Laplacian);
- Y is the initial label matrix;
- $\alpha \in (0, 1)$ controls the trade-off between the propagation of information and the retention of the original labels.

Unlike Label Propagation, which keeps labeled nodes fixed throughout the process (hard clamping), Label Spreading allows a soft diffusion of label information even from labeled nodes. This soft clamping improves flexibility and reduces sensitivity to noisy labels.

3.3. Proposed Method - Fair-LS

In standard Label Spreading, the clamping factor α is applied uniformly to all nodes, balancing the influence of initial labels and neighboring nodes. However, this uniformity can reinforce unfairness when structural differences exist between protected groups, such as unequal connectivity, degree, or label availability.

We proposed a modification to the Label Spreading algorithm, called Fair Label Spreading (Fair-LS) which introduces a group-specific clamping factor a_p and a_u, for privileged and unprivileged group respectively, which allows the algorithm to adapt the label propagation dynamics per group, which can:

- Correct for topological imbalances, reducing dominance of well-connected groups;
- Compensate for label scarcity in underrepresented groups by increasing reliance on neighborhood information;
- Mitigate fairness disparities by introducing group-aware regularization;
- Encourage balanced influence, preventing the majority group from disproportionately shaping the label space.

Thus, the label update rule for Fair-LS will be given as:

$$F^{(t+1)} = \begin{cases} \alpha_p S F^{(t)} + (1 - \alpha_p)Y & \text{if } i \text{ are from privileged group} \\ \alpha_u S F^{(t)} + (1 - \alpha_u)Y & \text{if } i \text{ are from unprivileged group} \end{cases} \tag{1}$$

3.4. Experimental Methodology

To evaluate our method, we used the AUC-ROC metric for classification, and Disparate Impact (DI) and Absolute Average Odds Difference (AAOD) for Fairness. AUC-ROC measures the ability to distinguish between classes, with values closer to 1 indicating better performance. DI assesses the ratio of favorable outcomes between protected and unprotected groups, ideally equal to 1, while AAOD measures disparities in true and false positive rates across groups, with lower values (ideally 0) indicating greater fairness.

For datasets we used three widely studied benchmark datasets in fairness research: Adult Census Income, Compas Recidivism, and German Credit. For simplicity, we refer to these datasets in the following sections as “Adult”, “Compas”, and “German”, respectively. The datasets are detailed below.

- **Adult Census Income** [Becker and Kohavi 1996]: This dataset, extracted from the U.S. Census, contains demographic and employment-related attributes used to predict whether an individual’s income exceeds 50K per year. The favorable class (income $>$ 50k) accounts for 25% of the data. We consider sex and race as protected attributes. Males represent 68% of the population and females 32%, where female represents the unprivileged group with a disparate impact of 0.36. For race, white individuals comprise 86% and non-white individuals (the unprivileged group) 14%, with a disparate impact of 0.60.
- **Compas Recidivism** [Angwin et al. 2016]: This dataset contains information on criminal defendants, used to predict recidivism risk. It includes features such as prior offenses and charge degree. The favorable outcome (no recidivism) represents 54% of the data. We consider sex and race as protected attributes. Females, the privileged group, represent 19% of the population and males 81%, with a disparate impact of 0.80. For race, caucasians make up 34% of the population and non-caucasians, the unprivileged group, 66% with a disparate impact of 0.84.
- **German Credit** [Hofmann 1994]: This dataset includes financial and personal attributes for loan applicants, with the goal of predicting credit risk. The favorable class (good credit) makes up 70% of the data. We consider sex and age as protected attributes. Males, the privileged group, account for 69% and females 31% of the data, with a disparate impact of 0.90. For age, individuals older than 25 and younger than 60 represent 76% of the data and are considered the privileged group, while those aged 25 or less or 60 or more represent 24%, with a disparate impact of 0.84.

The datasets used in our experiments were obtained from the AIF360 library [Bellamy et al. 2018], which already includes pre-processing steps such as the removal of missing values and one-hot encoding of categorical variables. Additionally, we applied standard normalization to features to ensure zero mean and unit variance. For the Adult dataset we used a sample of 10000 rows to reduce computational cost.

Figure 1 presents the experimentation steps of our method. This methodology is repeated for each dataset and protected attribute, varying the seed of the random split of the dataset, 10 times. In detail:

- The dataset was divided between train and test in the proportion 80%-20%;
- The train dataset was divided between a labeled part (10%) and an unlabeled part (90%);
- The Label Spreading and Fair Label Spreading algorithms were applied to the labeled and unlabeled data.
  - The clamping parameter was varied for both algorithms with values from 0.05 to 0.95, in increments of 0.05.

  - The number of neighbors was defined as the heuristic $k = \sqrt{n}$ ([Nadkarni 2016]), where n is the size of the training dataset (labeled and unlabeled examples).
- The labeled and pseudo-labeled training data are used to train a LightGBM classifier, a tree-based gradient boosting framework optimized for speed and scalability [Shi et al. 2025]. We only set the parameter class_weight = "balanced".
- The classification and fairness metrics are calculated.

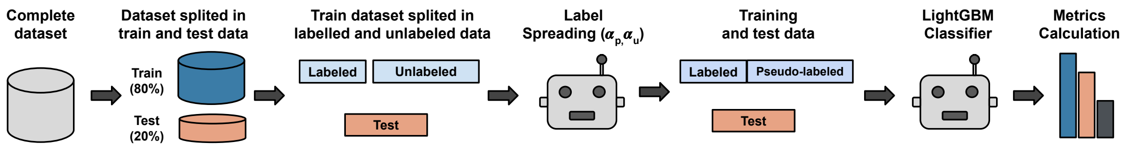

Figure 1. Overview of the experimental methodology pipeline

4. Results

This section details the results obtained for all datasets, protected attributes and metrics considered. First, in section 4.1 a study of the effect of the hyperparameters α_p and α_u of the proposed method, Fair-LS, on the metrics of interest is presented. In section 4.2 the average result of the set of parameters that led to the best score for LS and Fair-LS (individually) is presented, in order to have a more direct comparison between the methods.

4.1. Effect of α_p and α_u Parameters

Figure 2 presents a hexagonal heatmap detailing the results obtained by varying the parameters α_p and α_u for all datasets and attributes. The secondary diagonal of the heatmap matrix approximately represents the behavior of the traditional Label Spreading algorithm. Along this diagonal, the values of α_p and α_u are nearly equal, whereas in the actual Label Spreading algorithm they are exactly the same. It is possible to observe the effect of the parameters on the results and an idea of whether the ideal value is (equal to, greater than, or less than the other). Furthermore, it is noted that different α values between groups (values further away from the secondary diagonal) can generally benefit fairness metrics.

In the Adult dataset, considering AUC-ROC it is possible to notice that lower α_p and α_u values yielded good results. While very large α_p yielded worse results. In DI and AAOD it seems that the best results are obtained with higher α_p, both for race and sex.

For Compas, in AUC-ROC, the best results seem to be obtained with higher values of α_p and α_u. For DI and AAOD, in race, higher values of α_p and lower α_u yielded better results. For sex, lower values of α_p and higher values of α_u yielded the best results.

Finally, for German, in AUC-ROC most values of α_p and α_u yielded good results, except for low α_p and high α_u values. In DI, on the contrary, there are better values for low α_p and high α_u values. In AAOD also most values led to good results, except for very low or very higher values of α_p.

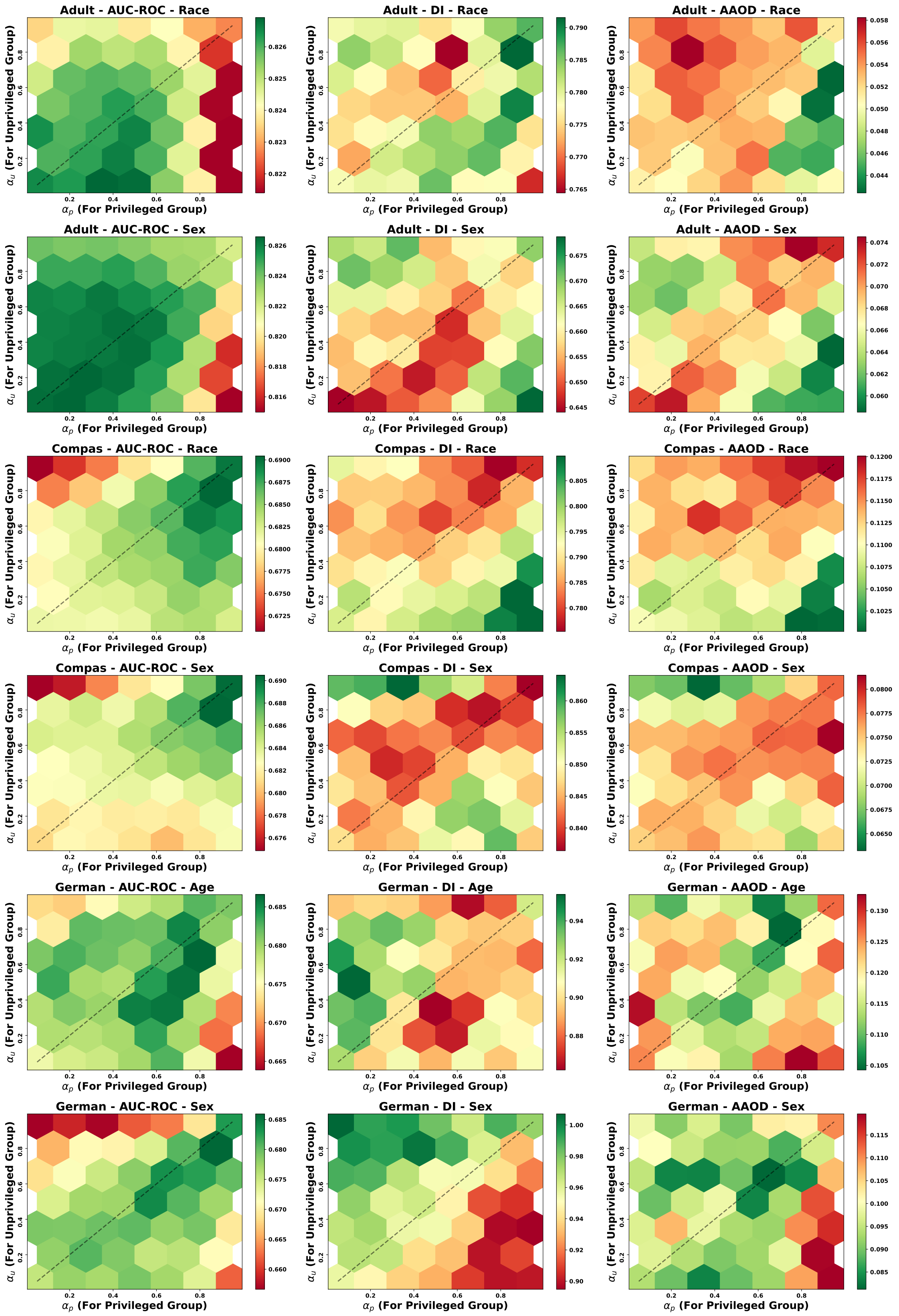

Figure 2. Hexagonal heatmap showing the variation of the metrics with respect to the parameters α_p and α_u

4.2. Overall Results

Figure 3 presents the results obtained by both models: standard Label Spreading (LS) using a single clamping factor α, and Fair-LS using group-specific factors α_p and α_u. The comparison is based on the best-performing parameter configurations for each model, evaluated across three metrics, aggregated using the following scoring function:

$$Score = AUC\text{-}ROC - |DI - 1| - AAOD \tag{2}$$

The selected parameters correspond to those that achieved the highest overall score, reflecting the best combined performance across the three metrics, with equal weight assigned to each. Overall, across all datasets, we generally observe a slight drop in AUC-ROC, which is often compensated by an increase in DI and/or a reduction in AAOD.

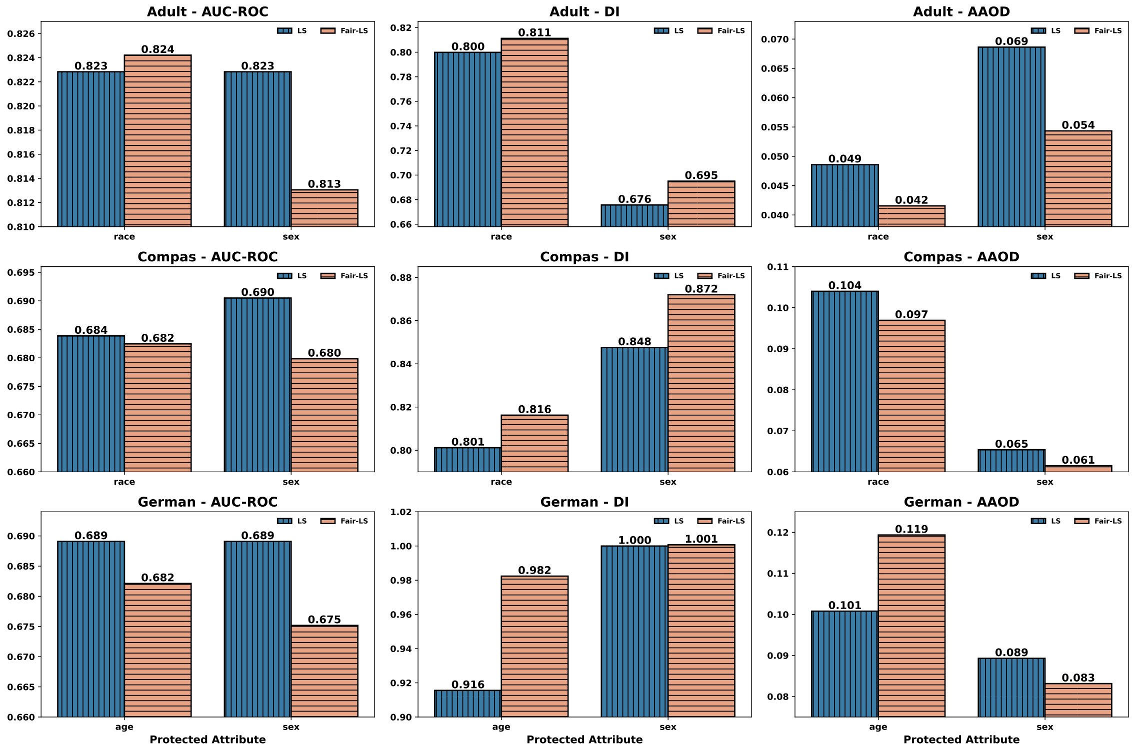

Figure 3. Bar chart of the classification and fairness metrics using the best parameters (from Equation 4.2) for LS and Fair-LS

Focusing on the Adult dataset, regarding the race attribute, AUC-ROC values were nearly identical between LS (0.823) and Fair-LS (0.824), while Fair-LS achieved a gain of 0.011 in DI and a 0.007 reduction in AAOD. For sex, Fair-LS presented a slight AUC-ROC decrease of 0.010, but improved fairness with a 0.019 increase in DI and a 0.015 decrease in AAOD.

In the case of the Compas dataset, Fair-LS led to an AUC-ROC drop of 0.002 for race and 0.010 for sex, but improved DI by 0.015 and 0.024, respectively. AAOD also decreased by 0.007 for race and 0.004 for sex.

Finally, for the German dataset, Fair-LS resulted in a loss in AUC-ROC: 0.007 for age and 0.014 for sex. DI improved by 0.066 for age, while values for sex were nearly identical (1.01). AAOD decreased by 0.006 for sex, but worsened by 0.018 for age.

4.3. Discussion

The experimental results suggest that introducing group-specific clamping factors in the Label Spreading algorithm (Fair-LS) can yield improvements in fairness metrics without severely compromising predictive performance. Across all datasets (Adult, Compas, and German), Fair-LS was able to reduce disparities measured by Disparate Impact (DI) and Average Absolute Odds Difference (AAOD), in many cases with only a minimal drop in AUC-ROC.

In the Adult dataset, fairness improved notably for race, with minor AUC-ROC loss; for sex, performance remained stable. In Compas, both race and sex showed consistent fairness gains with slight AUC-ROC reduction. For German, results were mixed: Age benefited in both fairness metrics, while for sex there was a trade-off, emphasizing the need for careful parameter tuning depending on the attribute and dataset context.

Overall, the findings support the hypothesis that allowing different propagation strengths for privileged and unprivileged groups introduces a valuable degree of flexibility. This adjustment helps mitigate structural imbalances in the graph, such as uneven connectivity or label scarcity, contributing to fairer semi-supervised learning.

As a limitation, the proposed fairness mitigation was applied exclusively to the Label Spreading component, which represents only one stage of the overall learning pipeline. Future work could explore combining this approach with pre-processing and/or post-processing techniques, either during pseudo-label generation or in the supervised classification step, to further enhance both predictive performance and fairness outcomes. Thus, more expressive results in classification and fairness can be obtained.

5. Conclusions

In this work, we proposed Fair-LS, a modified version of the Label Spreading algorithm that introduces group-specific clamping factors to address fairness concerns in graph-based semi-supervised learning. By allowing different propagation strengths for privileged and unprivileged groups, our approach aims to mitigate structural imbalances in the graph and reduce disparities in classification outcomes.

We evaluated Fair-LS on three well established fairness benchmark datasets, Adult, Compas, and German, covering domains such as income prediction, criminal recidivism, and credit risk. Although tested in these contexts, Fair-LS can be applied to a wide range of problems, particularly those where data can be represented as a graph, even when the graph is derived from tabular datasets. The evaluation considered Disparate Impact (DI) and Average Absolute Odds Difference (AAOD) as fairness metrics, alongside AUC-ROC for classification performance. The results show that Fair-LS improved fairness in most scenarios, often with only minimal loss in predictive performance.

These findings suggest that group-aware control over label propagation is a promising direction for improving fairness in semi-supervised graph learning, since controlling for different clamping factors between groups seems to have increased DI values toward 1 and reduced AAOD. In the future, additional experiments could include multi-objective hyperparameter optimization (covering the clamping factors α_p and α_u, k parameter of the KNN graph, and the LightGBM hyperparameters) together with the integration of pre-processing and/or post-processing fairness techniques that could also be applied to the supervised stage of the pipeline.

Acknowledgements

The authors would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Conselho Nacional de Desenvolvimento Científico (CNPq).

References

Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). How we analyzed the compas recidivism algorithm.

Barocas, S., Hardt, M., and Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT press.

Becker, B. and Kohavi, R. (1996). Adult. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5XW20.

Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilovic, A., Nagar, S., Ramamurthy, K. N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., and Zhang, Y. (2018). AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias.

Chakraborty, J., Tu, H., Majumder, S., and Menzies, T. (2021). Can we achieve fairness using semi-supervised learning? *arXiv preprint arXiv:2111.02038*.

Chong, Y., Ding, Y., Yan, Q., and Pan, S. (2020). Graph-based semi-supervised learning: A review. *Neurocomputing*, 408:216–230.

de Oliveira, W. D. G. and Berton, L. (2023). A systematic review for class-imbalance in semi-supervised learning. *Artificial Intelligence Review*, 56(Suppl 2):2349–2382.

de Oliveira, W. D. G. and Berton, L. (2024). A comparative study of pre-processing algorithms for fair classification in few labeled data context. *AI and Ethics*, pages 1–21.

Hofmann, H. (1994). Statlog (German Credit Data). UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5NC77.

Miyazaki, T. and Sumikawa, Y. (2018). Label propagation using amendable clamping. In *IUI Workshops*.

Nadkarni, P. (2016). Chapter 10 - core technologies: Data mining and “big data”. In Nadkarni, P., editor, *Clinical Research Computing*, pages 187–204. Academic Press.

Rabonato, R. T. and Berton, L. (2025). A systematic review of fairness in machine learning. *AI and Ethics*, 5(3):1943–1954.

Shi, Y., Ke, G., Soukhavong, D., Lamb, J., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., Titov, N., and Cortes, D. (2025). *lightgbm: Light Gradient Boosting Machine*. R package version 4.6.0.99.

Song, Z., Yang, X., Xu, Z., and King, I. (2022). Graph-based semi-supervised learning: A comprehensive review. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11):8174–8194.

Sulaiman, M. and Roy, K. (2025). Gflc: Graph-based fairness-aware label correction for fair classification. *arXiv preprint arXiv:2506.15620*.

Tsang, A., Wilder, B., Rice, E., Tambe, M., and Zick, Y. (2019). Group-fairness in influence maximization. *arXiv preprint arXiv:1903.00967*.

Yang, S., Cerrato, M., Ienco, D., Pensa, R. G., and Esposito, R. (2023). Fair swirl: fair semi-supervised classification with representation learning. *Machine Learning*, 112(9):3051–3076.

Zhang, T., Zhu, T., Han, M., Chen, F., Li, J., Zhou, W., and Yu, P. S. (2023). Fairness in graph-based semi-supervised learning. *Knowledge and Information Systems*, 65(2):543–570.

Zhang, T., Zhu, T., Han, M., Li, J., Zhou, W., and Yu, P. S. (2020). Fairness constraints in semi-supervised learning. *arXiv preprint arXiv:2009.06190*.

Zhou, D., Bousquet, O., Lal, T., Weston, J., and Schölkopf, B. (2003). Learning with local and global consistency. *Advances in neural information processing systems*, 16.

Zhu, X., Ghahramani, Z., and Lafferty, J. (2002). Learning from labeled and unlabeled data with label propagation. In *CMU CALD Tech Report*.