D3.3 Implementation of Mitigating Algorithms

Christos Gkartzios[1], **Glykeria Toulina**[1], **Spyridon Tzimas**[1], **Panagiotis Papadakos**[2], **Panagiotis Papapetrou**[3], **Stavros Sintos**[4], **Evaggelia Pitoura**[1], **Panayiotis Tsaparas**[1]

[1] Department of Computer Science and Engineering, University of Ioannina, Greece
[2]Institute of Computer Science (ICS) - Foundation for Research and Technology - Hellas (FORTH), Greece
[3]Department of Computer and Systems Sciences, Stockholm University, Sweden
[4]Department of Computer Science, University of Illinois Chicago, USA

1. Introduction

The overall objective of WP3 is to design and implement bias mitigation algorithms applicable to key Machine Learning and Data Mining tasks relevant to the Themis platform. The work addressed six main problem domains:

- **Community detection:** Community detection is the problem of finding subsets of nodes in a graph that are well connected to each other, while sparsely connected with the rest of the network.
- **Link analysis:** Link anaysis refers to the use of graph links for assessing the importance of nodes in the graph. In our work we consider the fairness of the celebrated Pagerank algorithm.
- **Counterfactual Explanations:** Counterfactual Explanations suggests feature changes for an instance that will result in an alternative classification.
- **Opinion Formation**: Opinion formation models suggest mechanisms for understanding the diffusion of opinions in networks, and they can be used for algorihtmic interventions.
- Clustering: Clustering is the problem of grouping together similar objects (points), while separating them from dissimilar ones.
- k-NN Classification: A k-NN classifier performs classification according to the majority class in the k Nearest Neighbors of a new sample in the training set.

We proposed and implemented algorithms for all of the above problems. In this Deliverable, we make the code for our algorithms in the github repository of the Themis project[1].

2. Algorithm Implementation

For WP3, we designed and implemented algorithms for the problems we outlined above. The detailed description of the algorithms can be found in Deliverables D3.2 and D3.4. The implementation of all our algorithms can be found in the repository **mitigation algorithms** in the Themis github repository[2].

[1]https://elidek-themis.github.io/
[2]https://github.com/elidek-themis/mitigation-algorithms

- **Fair Modularity-based Community Detection.** We propose a novel fair community detection algorithm that incorporates the community modularity fairness metric we introduced. It extends the popular Louvain algorithm. The results of our work were published in the 2025 ACM International Web Conference [1].

  The implementation of the algorithm is available at the github repository **fair-network-communities-through-group-modularity** under the mitigation algorithms folder [3].

- **Fair spectral and deep-NN algorithms for community detection under modularity fairness:** We extended the work on modularity fairness, with the development of fair spectral and deep-learning-based community detection algorithms, which incorporate group-sensitive connectivity either directly at the input level, or through fairness-aware loss functions. This work was publised at the ICDM 2025 conference [2].

  The implementation of the algorithm is available at the github repository **modularity-fair-deep-community-detection**, under the mitigation algorithms folder[4].

- **Fair Label-Propagation.** We designed a fair label propagation algorithm using physics-inspired principles to promote balanced group representation during the propagation process. Sensitive attributes are modeled as electrical charges, so that nodes are attracted from communities with a deficit in nodes from a different group, while repelled from communities with surplus in nodes from the same group. This electrostatic interaction is combined with a pulling force that favors grouping adjacent nodes. The algorithm was presented in the Algorithmic Fairness in Network Science workshop at NetSci 2025, and published at ASONAM 2025 [5].

  The implementation of the algorithm is available at the github repository **flp** under the mitigation algorithms folder[5].

- **Fair Pagerank.** We propose a fair version of the PageRank algorithm, building on the framework in [6]. Our algorithms aim to minimize the cost of the local interventions required for achieving fairness. This work has been accepted for publication to the Workshop on Web & Graphs, Responsible Intelligence, and Social Media, collocated with WSDM 2026 [7].

  The implementation of the algorithm is available at the github repository **fair-pagerank** under the mitigation algorithms folder[6].

- **Debiasing Counterfactual Explanations.** Building on the metrics defined in Deliverable D1.3, we propose an algorithm for debiasing counterfactual explanations over different sensitive groups. This work was published in ICDM 2024 [3]. The paper was selected as one of the best-ranked papers of ICDM 2024, and an extended version of the work was published in Knowledge and Information Systems (KAIS) [4].

  The implementation of the algorithm is available at the github repository **CounterFair** under the mitigation algorithms folder[7].

- **Fair Clustering.** Building upon the Physics-inspired methodology for community detection we extend it to the case of numerical data, where the goal is to cluster points (vectors) in a multidimensional space. We consider two algorithms. The first is a fair variant of the popular k-means algorithm. The second, fair Potential Field Clustering algorithm (F-PFC), is a novel algorithm that explores gravitational and electrostatic forces to group points into clusters.

  The implementation of the fair k-means algorithm is available at the github repository `fair-k-means` under the algorithms folder[8].

[3] https://github.com/fair-network-communities-through-group-modularitymain/fair-network-communities-through-group-modularity
[4] https://github.com/elidek-themis/mitigation-algorithms/tree/main/modularity-fair-deep-community-detection
[5] https://github.com/elidek-themis/mitigation-algorithms/tree/main/flp
[6] https://github.com/elidek-themis/mitigation-algorithms/tree/main/fair-pagerank
[7] https://github.com/elidek-themis/mitigation-algorithms/tree/main/counter-fair
[8] https://github.com/elidek-themis/mitigation-algorithms/tree/main/fair-k-means

The implementation of the F-PFC is available at the github repository **fair-potential-field-clustering** under the mitigation algorithms folder[9].

- **Fair Opinion Formation.** we investigated fairness in the Friedkin–Johnsen model by defining fairness in terms of influence across groups. We designed algorithms that make minimal interventions on node stubbornness to achieve fair influence allocation.

  The implementation of the algorithms is available at the github repository **opinion-fairness** under the mitigation algorihtms folder[10]

- **Fair k-NN Classification.** We considered the problem of fairness for the k-Nearest-Neighbors (k-NN) classification algorithm, and we proposed algorithms that make minimal number of interventions to the training set to achieve fairness. We consider two types of modifications, flipping the label of training points, or removing training points.

  The implementation of the algorithms that flips points is available at the github repository **k-NN-classification-fairness-flip** under the mitigation algorithms folder[11].

  The implementation of the algorithms that removes points is available at the github repository **k-NN-classification-fairness-remove** under the mitigation algorithms folder[12].

3. Conclusions

In this Deliverable, we provide the repositories with the code for the mitigation algorithms we implemented in Themis. Our work spans several different areas of Machine Learning and Data Mining. All code is publicly available.

Acknowledgments

We would like to acknowledge the contributions of Nikos Theologis, Vassilis Vogianou, Manolis Smyrnakis to this work.

References

[1] Christos Gkartzios, Evaggelia Pitoura, and Panayiotis Tsaparas. "Fair Network Communities through Group Modularity". In: *Proceedings of the ACM Web Conference (WWW '25)*. 2025.

[2] Christos Gkartzios, Evaggelia Pitoura, and Panayiotis Tsaparas. "Modularity-Fair Deep Community Detection". In: *ICDM*. 2025.

[3] Alejandro Kuratomi et al. "CounterFair: Group Counterfactuals for Bias Detection, Mitigation and Subgroup Identification". In: *IEEE International Conference on Data Mining (ICDM)*. 2024.

[4] Alejandro Kuratomi et al. "Subgroup fairness based on shared counterfactuals". In: *Knowl. Inf. Syst.* 67.11 (2025), pp. 10863–10901. DOI: 10.1007/S10115-025-02555-7.

[5] Glykeria Toulina and Panayiotis Tsaparas. "A Fair Label Propagation Community Detection Algorithm". In: *ASONAM*. 2025.

[6] Sotiris Tsioutsiouliklis et al. "Fairness-Aware PageRank". In: *Proceedings of the Web Conference 2021*. 2021, pp. 3815–3826. ISBN: 9781450383127.

[7] Spyridon Tzimas et al. "Minimizing the Cost of PageRank Fairness". In: *Workshop on Web & Graphs, Responsible Intelligence, and Social Media (WEB & GRAPH 2026), collocated with WSDM (Accepted for publication)*. 2026.

[9] https://github.com/elidek-themis/mitigation-algorithms/tree/main/fair-potential-field-clustering
[10] https://github.com/elidek-themis/mitigation-algorithms/tree/main/opinion-fairness
[11] https://github.com/elidek-themis/mitigation-algorithms/tree/main/k-NN-classification-fairness-flip
[12] https://github.com/elidek-themis/mitigation-algorithms/tree/main/k-NN-classification-fairness-remove