Impact of Missing Data Imputation Techniques on Demographic Parity and Equalized Odds in Graph-Based Node Classification

Assignee Research

June 14, 2026

Abstract

Analysis of the fairness of machine learning (ML) algorithms recently attracted many researchers' interest. Most ML methods show bias toward protected groups, which limits the applicability of ML models in many applications like crime rate prediction etc. Since the data may have missing values which, if not appropriately handled, are known to further harmfully affect fairness. Many imputation methods are proposed to deal with missing data. However, the effect of missing data imputation on fairness is not studied well. In this paper, we analyze the effect on fairness in the context of graph dat

1 Introduction

This paper examines: Impact Of Missing Data Imputation On The Fairness And Accuracy Of Graph Node Classifiers. Research question: What is the impact of different missing data imputation techniques on the demographic parity and equalized odds of graph-based node classification models?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.5/10.

3 Results

15 papers retrieved. 25 claims extracted; 18 independently verified. Quality review score: 7.5/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Graph node classifiers that integrate both node features and graph edges information enhance node representation and imp	✓	0.23
Social bias in data can cause fairness issues in graph node classifiers.	✓	0.19
Many datasets contain discrimination and social bias towards sensitive attributes like region, age, skin color, gender,	✓	0.22
Machine learning models trained on biased data can inherit bias.	×	0.14
Fairness issues in GNNs and Node2Vec have been reported in recent studies.	×	0.12
Bias in data can propagate through the edges in graph data, aggravating fairness issues.	×	0.10
Homophily is common in social network graphs, where nodes with similar sensitive features tend to connect with each othe	✓	0.27
Aggregating neighbor node features in homophilic graphs can lead to severe bias in graph-specific algorithms.	×	0.14
Missing data is common in machine learning problems and can have an adverse effect on fairness if not dealt with properl	✓	0.18
Imputation is the most common approach to dealing with missing data.	✓	0.16
Different statistical and machine learning imputation methods have been proposed to deal with missing data.	✓	0.24
Missing values can cause sensitive attribute imbalance, decreasing fairness.	✓	0.16
Limited research work exists regarding the impact of missing data imputation on fairness.	✓	0.21
Missing data contributes to bias in machine learning algorithms.	×	0.13
Zhang et al. investigated the impact of missing data imputation on fairness and concluded that it produces bias in the d	✓	0.20
Another work reported that accessing fairness under missing data is challenging.	×	0.11
No work has been performed regarding the fairness of data imputation on graph structure data.	✓	0.28
Fairness in graph data is affected by missing data imputation.	✓	0.23
Data imputation methods have an impact on fairness and accuracy.	✓	0.22
Missing data mechanisms have an adverse effect on fairness.	✓	0.20
Most fairness issues are associated with sample imbalance.	✓	0.15
Missing data affects the accuracy of the classi-	×	0.09

References

- http://arxiv.org/abs/1905.01907v2
- http://arxiv.org/abs/2211.00783v1
- http://arxiv.org/abs/2501.12571v2