Missing Data Imputation and Demographic Parity in Graph Neural Networks for Social Network Node Classification

Assignee Research

June 16, 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: How does missing data imputation affect the demographic parity of graph neural networks used for node classification in social network analysis?.

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 7.4/10.

3 Results

15 papers retrieved. 22 claims extracted; 19 independently verified. Quality review score: 7.4/10.

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 neural network algorithms integrate node features and graph edges (topological structure) to enhance node represen	✓	0.23
Social bias in data causes fairness issues in graph algorithms, limiting their applicability in practical applications.	×	0.12
Many datasets contain discrimination and social bias towards sensitive attributes such as region, age, skin color, and g	✓	0.21
Machine learning models trained on biased data can inherit that bias.	×	0.15
Studies on the fairness of GNNs and Node2Vec have reported that these models inherit bias from training data.	✓	0.17
Bias in graph data propagates through edges, which can aggravate fairness issues in graphs.	✓	0.18
Homophily in graph data refers to connections among nodes having the same attribute values.	✓	0.16
In social network graphs, nodes with similar sensitive features tend to connect with each other more than nodes with dif	✓	0.29
In graphs with homophily, aggregating neighbor node features produces representations of similar groups that differ sign	✓	0.22
Missing data, if not dealt with properly, has an adverse effect on fairness in machine learning problems.	✓	0.18
Missing values can cause sensitive attribute imbalance, which implies that missing data can decrease fairness.	✓	0.26
Zhang et al. [45] concluded that missing data imputation produces bias in data, but their study did not include graph da	✓	0.22
Work referenced in [44] addresses accessing fairness under missing data without considering imputation and does not cons	✓	0.18
No prior work has been performed regarding the fairness of data imputation on graph structure data.	✓	0.25
Fairness in graph data is affected by missing data imputation.	✓	0.24
Data imputation methods have an impact on both fairness and accuracy in graph 4ode classifiers.	✓	0.25
Missing data mechanisms have an adverse effect on fairness.	✓	0.20
Most fairness issues in the studied context are associated with sample imbalance.	×	0.12
Missing data affects the accuracy of graph node classifiers.	✓	0.16

References

- http://arxiv.org/abs/2107.08310v5
- http://arxiv.org/abs/1905.01907v2
- http://arxiv.org/abs/2211.00783v1