1 Experiment 3: Original Vs Removed Edges

Description: Experiment 3 is conducted on statistical tests for original feature and removed edges.

1. Subsection 1.1 explains statistical tests for original feature and removed edges. The original data consists of Netpro2VecMetgraphs, Netpro2VecNDD, Netpro2VecTM1, Netpro2VecTM2, Netpro2VecNDD+TM1, Netpro2VecNDD+TM1+TM2, Graph2Vec, GL2Vec, FeatherGraph, and SF. Meanwhile, the removed edges consists of features Netpro2VecMetgraphs, Netpro2VecNDD, Netpro2VecTM1, Netpro2VecTM2, Netpro2VecNDD+TM1, Netpro2VecNDD+TM1+TM2, Graph2Vec, GL2Vec, FeatherGraph, and SF after randomly deleted 10% edges.

2. In this experiment we carry out statistical tests, including accuracy, precission, recall, and F1-score from all models-based graph embeddings.

3. Subsection 1.1 presents Table 1a until Table 1d ilustrate the statistical test results mentioned.

4. Subsection 1.2 depicts visualizations of removal edges and Normalized removed edges based on the statistical tests result, that provides in following Figure 1a until Figure 1d

RESUME: The comparison resume is based on the percentage difference of original feature compare to the removed edges. Rely on the accuracy, precision, recall, and F1-score, we describe the best and the lowest percentage difference as follow:

1. The best percentage difference of accuracy are GCNs and SVMl with percentage difference of original feature compare to the removed edges, that is about 0.000%. Meanwhile maximum decreasing percentage difference of accuracy is in DT about 0.011%, and maximum increasing percentage difference of accuracy is in SVMnl about 0.014%.

2. The best percentage difference of precisions are in GNB and SVMnl that is about 0.001%, meanwhile remaining models show maximum increasing percentage difference of precision is in GCNs about 0.018%.

3. The best percentage difference of recall is found in GCNs that is about 0.000%. Meanwhile the maximum percentage difference of recall about 0.015% in DT

4. The best percentage difference of F1-score are found in GCNs and SVMl that is about 0.000%. Otherwise the worst percentage difference of F1-score is in SVMnl about 0.021%.

CONCLUSION: Based on statistical tests analyzing the effect of the original features and removed edges, we conclude that the best-performing model is GCNs, followed by SVM (linear). GCNs is recommended as a robust model in the context of perturbation, demonstrating stability when randomly removed 10% of edges in the features.

1.1 Original Vs Removed Edges

Table 1: **Table between between Original and Removed Edges**

(a) Accuracy

Accuracy

Rank	Models	Original	Removed Edges	Percentage diff.
1	GCNs	1	1	0,000
2	SVMl	0,97	0,97	0,000
3	KNN	0,965	0,958	-0,007
4	DT	0,948	0,938	-0,011
5	GNB	0,935	0,933	-0,002
6	RF	0,93	0,933	0,003
7	SVMnl	0,883	0,895	0,014

(b) Precision

Precision

Rank	Models	Original	Removed Edges	Percentage diff.
1	GCNs	0,982	1	0,018
2	SVML	0,972	0,974	0,002
3	KNN	0,952	0,956	0,004
4	DT	0,951	0,947	-0,004
5	GNB/SVMnl	0,945	0,946	0,001
6	RF/GNB	0,933	0,936	0,003
7	SVMnl/RF	0,891	0,895	0,004

(c) Recall

Recall

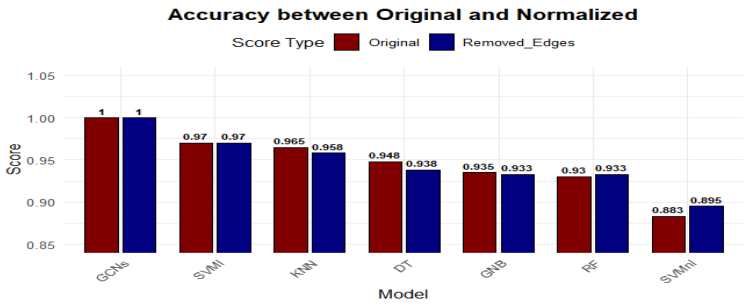
Rank	Models	Original	Removed Edges	Percentage diff.
1	GCNs	1	1	0,000
2	KNN/SVMl	0,965	0,967	0,002
3	SVMl/KNN	0,949	0,957	0,008
4	DT	0,947	0,933	-0,015
5	GNB/SVMnl	0,933	0,931	-0,002
6	RF/GNB	0,93	0,931	0,001
7	SVMnl/RF	0,885	0,897	0,014

(d) F1-Score

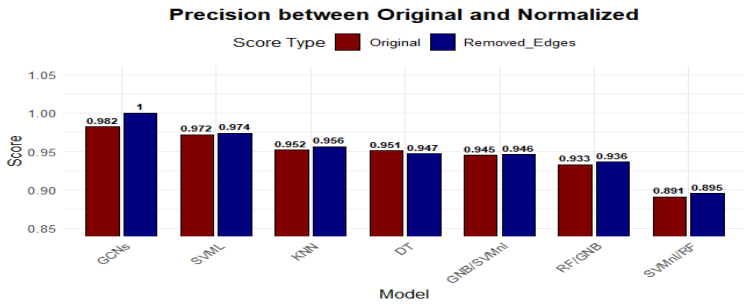
F1

Rank	Models	Original	Removed Edges	Percentage diff.
1	GCNs	1	1	0,000
2	SVML	0,969	0,969	0,000
3	KNN	0,965	0,957	-0,008
4	DT	0,947	0,937	-0,011
5	GNB/SVMnl	0,932	0,93	-0,002
6	RF	0,929	0,927	-0,002
7	SVMnl	0,869	0,887	0,021

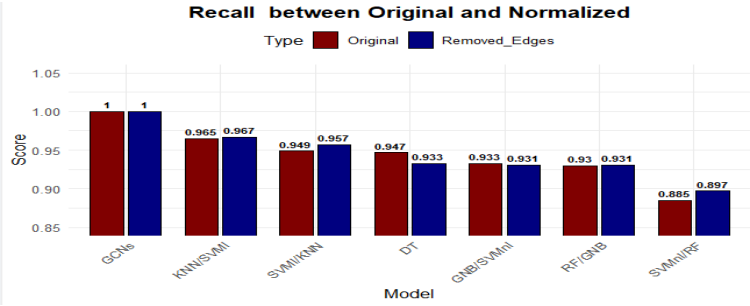
1.2 Figure of Statistical Tests Result



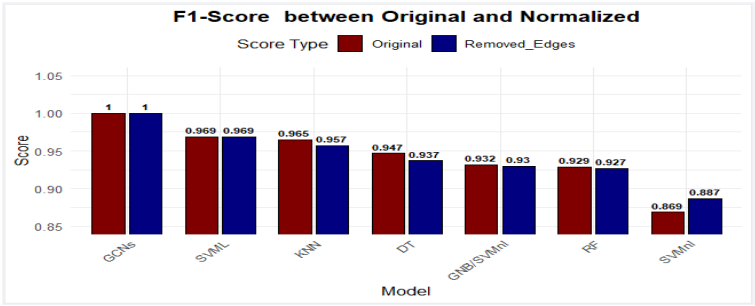
(a) Accuracy



(b) Precision



(c) Recall



(d) F1-Score

Figure 1: **Plot between Original and Removed Edges**