1 Experiment 4: Normalized Vs Normalized Remove Edges

DESCRIPTION: Experiment 4 is conducted on statistical tests on Normalized feature and Normalized of removal edges features, which is describes as follow:

1. Normalized feature and Normalized of removal edges features. The normalized feature consists of Netpro2VecMetgraphs, Netpro2VecNDD, Netpro2VecTM1, Netpro2VecTM2, Netpro2VecNDD+TM1, Netpro2VecNDD+TM1+TM2, Graph2Vec, GL2Vec, FeatherGraph, and SF. Meanwhile, the removal edges consists of features Netpro2VecMetgraphs, Netpro2VecNDD, Netpro2VecTM1, Netpro2VecTM2, Netpro2VecNDD+TM1, Netpro2VecNDD+TM1+TM2, Graph2Vec, GL2Vec, FeatherGraph, and SF that have been normalized using the min-max method, ensuring that feature values fall within the interval [0,1]. On the other hand, the normalized removed edges refer to the edge removal features, which include Netpro2VecMetgraphs, Netpro2VecNDD, Netpro2VecTM1, Netpro2VecTM2, Netpro2VecNDD+TM1, Netpro2VecNDD+TM1+TM2, Graph2Vec, GL2Vec, FeatherGraph, and SF. These features are derived after randomly removing 10% of Metgraph edges and have been normalized using the min-max method, ensuring that all feature values fall within the interval [0,1].

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 ilustrates Table 1a until Table 1d as the statistical tests result.

4. Subsection 1.2 depicts visualizations of normalized feature and Normalized of removal edges features 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 normalized features and normalized of removal edges features. 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, RF, and GNB with percentage difference of normalized features compare to normalized of removal edges features, that is about 0.000%. Meanwhile maximum decreasing percentage difference of accuracy are in DT and SVMnl that is about 0.011%, and maximum increasing percentage difference of accuracy is in RF about 0.003%.

2. The best percentage difference of precisions is in GCNs that is about 0.000%, meanwhile remaining models show decreasing percentage difference of precision with the maximum decreasing is about 0.009% in SVMnl.

3. The best percentage difference of recall is in GCNs that is about 0.000%, meanwhile remaining models show decreasing percentage difference of precision with the maximum decreasing is about 0.035% in SVMnl.

4. The best percentage difference of F1-score is in GCNs that is about 0.000%, meanwhile remaining models show decreasing percentage difference of precision with the maximum decreasing is about 0.045% in SVMnl.

CONCLUSION: Based on statistical tests analyzing the effect of normalized features compare to normalized of removal edges features, we conclude that the best-performing model is GCNs and for further it is recommended as a robust model in the context of perturbation, demonstrating stability.

1.1 Normalized feature Vs Normalized of removal edges features

Table 1: **Table between Normalized feature Vs Normalized of removal edges features**

(a) Accuracy between Original and Removed Edges

Accuracy				
Rank	Models	Original	Removed Edges	difference (%)
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 between Original and Removed Edges

Precision				
Rank	Models	Original	Removed Edges	difference (%)
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 between Original and Removed Edges

Recall				
Rank	Models	Original	Removed Edges	difference (%)
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 between between Original and Removed Edges

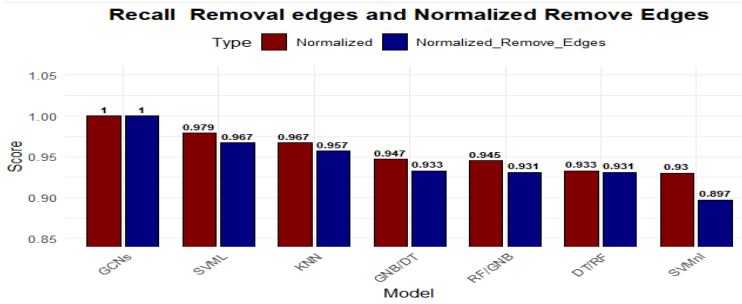| F1				
Rank	Models	Original	Removed Edges	difference (%)
1	GCNs	1	1	0,000
2	SVML	0,969	0,969	0,000
3	KNN	0,965	0,957	-0,008
4	DT	0,947	0,937	-0,011
5	GNB/SVMnl	0,932	0,93	-0,002
6	RF	0,929	0,927	-0,002
7	SVMnl	0,869	0,887	0,021

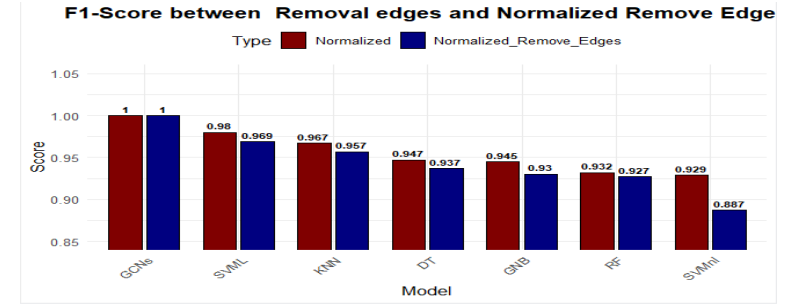## 1.2 Figure of Statistical Tests Result
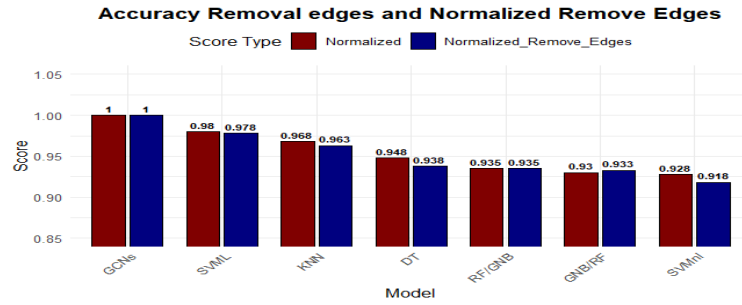
(a) Accuracy
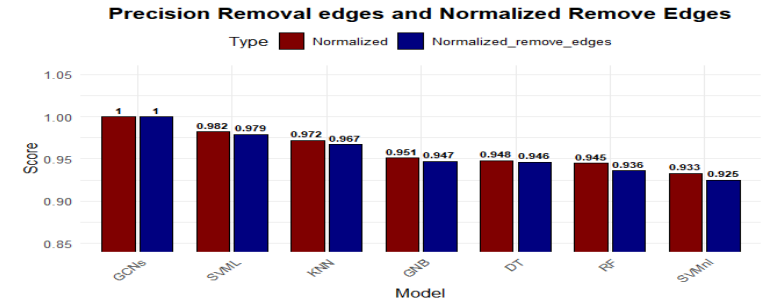
(b) Precision

(c) Recall

(d) F1-Score

Figure 1: Plot between Normalized and Normalized Remove Edges