Rebuttal Letter for reviewer C92L

Anonymous Author(s)

1 RELATION BETWEEN IMBALANCED AGGREGATION AND OVERLOOKING

We observed that existing noticeability measures show near-zero noticeability when the attack rate is low, i.e., the number of attack edges ≫ the number of real edges (***overlooking***). To overcome this problem, HideNSeek aggregate the edges scores in an imbalance-aware way, using the Area Under the Receiver Operating Characteristic Curve (AUROC). AUROC is a popular method that can measure the imbalance binary classification performance, such as anomaly detection (detailed computation is in the main paper, Appendix A).

2 EXTREME CASE

In table 5, we measured LEO's performance on attack graph with 50% of attack rate. We used Cora dataset with five attack methods used in this paper. Cosine Sim. was selected as a baseline, which is independent of the attack rate.

Table 1: [exp. 2] LEO's performance in 50% of attack rate scenario

	Random	DICE	PGD	Structack	**Metattack**
Cosine Sim.	0.803	0.814	0.794	0.825	0.808
LEO	0.879	0.881	0.804	0.900	0.815

3 ADAPTIVE ATTACK

We designed an adaptive attack in a gradient-based way and conducted

Table 2: [exp. 3.1] LEO's performance against adaptive attack

	Random	PGD	Adaptive	Clean
LEO (AUROC))	0.894	0.874	0.912	-
Node classification (%)	0.803	0.689	0.797	0.814

Table 3: [Robustness - GUARD baseline]

	GCN	GCN + HideNSeek
Adaptive	0.786	0.793

4 GNNGUARD BASELINE

Table 4: [LEO - GUARD baseline]

	Random	DICE	PGD	Structack	Metattack
GUARD	0.767	0.730	0.737	0.861	0.760
LEO	0.894	0.899	0.874	0.941	0.894

Table 5: [Robustness - GUARD baseline]

	Random	DICE	PGD	Structack	Metattack
GCN	0.791	0.782	0.685	0.773	0.446
GCN+GUARD	0.787	0.771	0.699	0.774	0.482
GCN+HideNSeek	0.793	0.782	0.710	0.784	0.639