

Money Laundering Detection with Multi-Aggregation Custom Edge GIN Network - Supplementary Material

Filip Wójcik

The data

This study builds upon the datasets and methodological framework established by (Silva, Correia, and Maziero 2023). Their research utilized data generated by the financial transaction simulation tool “AMLSim”, developed by IBM (Weber et al. 2018). AMLSIm extends the capabilities of the PaySim architecture and was designed to simulate transactional patterns characteristic of money laundering processes. This simulator stands out for its ability to fine-tune a wide range of parameters, such as the duration of the simulation in time steps (days), transaction value ranges, and class imbalance ratio, among other statistical variables.

In their study, Silva, Correia, and Maziero (2023) generated a dataset representing daily transaction activities for the year 2020, encompassing a total of 365 time steps. This dataset includes account identifiers, predetermined suspicion indicators, details on the sending and receiving accounts, and transaction types. The main elements for their analysis included the transaction value (*“base_amt”*), transaction timestamp (*“tran_timestamp”*), and transaction type (*“tx_type”*), along with flags indicating potential illicit activities (*“is_sar”*), and data on the source or destination accounts (*“orig_acct”*, *“bene_acct”*). These elements were used to construct the transaction graphs, supplemented by initial deposit information (*“initial_deposit”*) and transaction behaviour categorizations (*“tx_behavior_id”*) for each account. That dataset was divided into several portions with different class imbalance rates.

Table 1 summarizes the key elements of the datasets used in this study - number of nodes and edges in train/validation/test splits. Table 2 presents the class proportions for the different datasets.

Table 1: Dataset statistics

Dataset	# train nodes	# val nodes	# test nodes	# train edges	# val edges	# test edges
AMLSim 1/3	3657	1387	1287	1858	705	650
AMLSim 1/5	6336	2070	2101	3230	1056	1069
AMLSim 1/10	12341	4199	4211	6367	2169	2174
AMLSim 1/20	24066	8073	8125	12800	4291	4329

Table 2: Class proportions per dataset

Dataset	# legal	% legal	# illicit	% illicit
AMLSim 1/3	1259	0.678	599	0.322
AMLSim 1/5	2631	0.815	599	0.185
AMLSim 1/10	5768	0.906	599	0.094
AMLSim 1/20	12201	0.953	599	0.047

Hyperparameter optimization

While a traditional ablation study was not conducted, the contribution of each hyperparameter to model performance was systematically evaluated using the PED-ANOVA Importance. This method serves as a possible alternative to ablation studies, particularly in cases where the number of tuned parameters is large and their interactions form subspaces (e.g., the interplay between the number of layers, layer size, and aggregation methods per layer). By leveraging ANOVA-based calculations, PED-ANOVA quantifies how strongly each hyperparameter contributes to achieving performance above a baseline, which is determined by fitting Parzen estimators to the results of completed trials (Watanabe, Bansal, and Hutter 2023).

The feature importance analysis revealed that no single hyperparameter dominated performance across all datasets. Instead, the relative importance varied depending on dataset properties, particularly the class imbalance ratio and dataset size. For less-imbalanced cases (AMLSim 1/3 and 1/5), the number and size of convolutional layers, as well as aggregation functions, were relatively more influential, though their importance scores were comparable. In contrast, in highly imbalanced scenarios (AMLSim 1/10 and 1/20), embedding reduction modes became more critical, alongside aggregation functions.

This variability highlights the importance of adaptive hyperparameter tuning tailored to dataset-specific properties. The flexibility of the MAGIC architecture further supports effective performance under a wide range of configurations, including the use of multiple, learnable aggregations.

Sections below present details on subsequent aspects of the hyperparameter optimization results - jointly and individually for each dataset.

Hyperparameter importance

Figure 1 illustrates the relative importance of each hyperparameter across the different datasets. The importance scores were calculated using the PED-ANOVA method, which quantifies the contribution of each hyperparameter to the model’s performance above a baseline. The baseline was determined by fitting Parzen estimators to the results of completed trials.

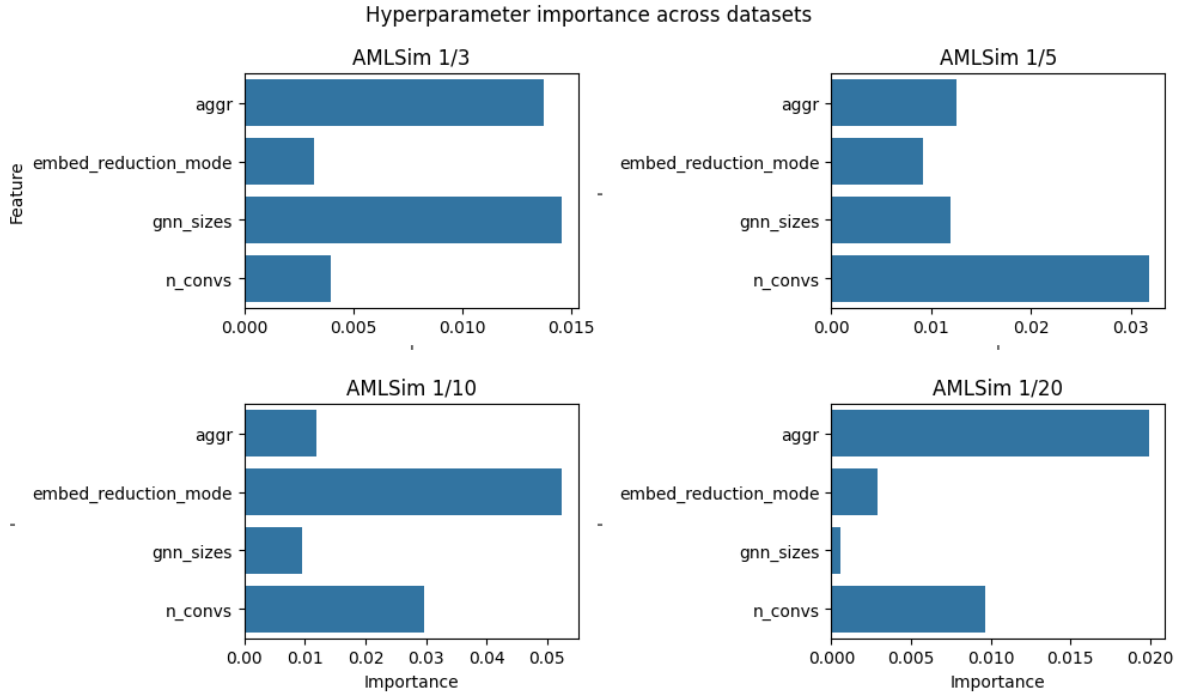


Figure 1: PED-ANOVA hyperparameter importances across datasets.

Subsequent tables Table 3, Table 4, Table 5, and Table 6 present the detailed results of the PED-ANOVA analysis for each dataset. The tables include the hyperparameter name, the importance F-score, and the corresponding p-value.

Table 3: Anova results for AMLSim 1/3.

Source	SS	DF	MS	F	p-unc	np2	Signif.
Conv. Dim.	0	3	0	5.87004e+15	0	1	Yes
No. Conv. Layers	-0	3	-0	-	1	1	No
				1.59139e+15			
Aggr.	-0	2	-0	-	1	1	No
				1.46753e+15			
Embed Reduction	0	1	0	4.18975e+15	0	1	Yes
Conv. Dim. * No. Conv. Layers	0	9	0	1.46271e+16	0	1	Yes
Conv. Dim. * Aggr.	-0	6	-0	-	1	1	No
				3.44301e+16			
No. Conv. Layers * Aggr.	-0	6	-0	-	1	1	No
				2.36161e+18			
Conv. Dim. * Embed Reduction	-0	3	-0	-	1	1	No
				1.88095e+15			
No. Conv. Layers * Embed Reduction	-0	3	-0	-	1	1	No
				4.92538e+15			
Aggr. * Embed Reduction	0	2	0	1.90361e+15	0	1	Yes
Conv. Dim. * No. Conv. Layers * Aggr.	26.2518	18	1.4584	8.39156e+29	0	1	Yes
Conv. Dim. * No. Conv. Layers * Embed Reduction	2.7078	9	0.3009	1.73112e+29	0	1	Yes
Conv. Dim. * Aggr. * Embed Reduction	1.7425	6	0.2904	1.67102e+29	0	1	Yes
No. Conv. Layers * Aggr. * Embed Reduction	2.2757	6	0.3793	2.18234e+29	0	1	Yes
Conv. Dim. * No. Conv. Layers * Aggr. * Embed Reduction	5.2919	18	0.294	1.69157e+29	0	1	Yes
Residual	0	11	0	nan	nan	nan	No

Table 4: Anova results for AMLSim 1/5.

Source	SS	DF	MS	F	p-unc	np2	Signif.
Conv. Dim.	-0	3	-0	-	1	1	No
				2.27451e+14			
No. Conv. Layers	0	3	0	3.39466e+14	0	1	Yes
Aggr.	0	2	0	8.45349e+15	0	1	Yes
Embed Reduction	-0	1	-0	-	1	1	No
				1.91665e+14			
Conv. Dim. * No. Conv. Layers	0	9	0	4.82411e+16	0	1	Yes
Conv. Dim. * Aggr.	0	6	0	1.12366e+15	0	1	Yes
No. Conv. Layers * Aggr.	-0	6	-0	-5.4478e+13	1	1	No
Conv. Dim. * Embed Reduction	0	3	0	2.34125e+14	0	1	Yes
No. Conv. Layers * Embed Reduction	0	3	0	1.01e+14	0	1	Yes
Aggr. * Embed Reduction	0	2	0	5.6504e+13	0	1	Yes
Conv. Dim. * No. Conv. Layers * Aggr.	46.65	18	2.5917	7.76747e+29	0	1	Yes
Conv. Dim. * No. Conv. Layers * Embed Reduction	0.929	9	0.1032	3.09383e+28	0	1	Yes
Conv. Dim. * Aggr. * Embed Reduction	0.6203	6	0.1034	3.09837e+28	0	1	Yes
No. Conv. Layers * Aggr. * Embed Reduction	0.6203	6	0.1034	3.09837e+28	0	1	Yes
Conv. Dim. * No. Conv. Layers * Aggr. * Embed Reduction	10.8115	18	0.6006	1.80017e+29	0	1	Yes
Residual	0	15	0	nan	nan	nan	No

Table 5: Anova results for AMLSim 1/10.

Source	SS	DF	MS	F	p-unc	np2	Signif.
Conv. Dim.	0	3	0	1.15492e+15	0	1	Yes
No. Conv. Layers	0	3	0	2.50744e+14	0	1	Yes
Aggr.	0	2	0	2.25528e+14	0	1	Yes
Embed Reduction	-0	1	-0	-	1	1	No
				9.44726e+14			
Conv. Dim. * No. Conv. Layers	0	9	0	2.37395e+15	0	1	Yes
Conv. Dim. * Aggr.	0	6	0	4.62302e+15	0	1	Yes
No. Conv. Layers * Aggr.	-0	6	-0	-	1	1	No
				2.09523e+15			
Conv. Dim. * Embed Reduction	-0	3	-0	-	1	1	No
				2.19239e+15			
No. Conv. Layers * Embed Reduction	0	3	0	1.83824e+15	0	1	Yes
Aggr. * Embed Reduction	0	2	0	5.7239e+15	0	1	Yes
Conv. Dim. * No. Conv. Layers * Aggr.	12.5444	18	0.6969	4.51291e+29	0	1	Yes
Conv. Dim. * No. Conv. Layers * Embed Reduction	1.5593	9	0.1733	1.12192e+29	0	1	Yes

Table 5: Anova results for AMLSim 1/10.

Source	SS	DF	MS	F	p-unc	np2	Signif.
Conv. Dim. * Aggr. * Embed Reduction	2.2543	6	0.3757	2.43303e+29	0	1	Yes
No. Conv. Layers * Aggr. * Embed Reduction	1.0259	6	0.171	1.10719e+29	0	1	Yes
Conv. Dim. * No. Conv. Layers * Aggr. * Embed Reduction	3.8106	18	0.2117	1.37088e+29	0	1	Yes
Residual	0	14	0	nan	nan	nan	No

Table 6: Anova results for AMLSim 1/20.

Source	SS	DF	MS	F	p-unc	np2	Signif.
Conv. Dim.	-0	3	-0	-	1	1	No
No. Conv. Layers	0	3	0	7.33744e+16	0	1	Yes
Aggr.	0	2	0	4.55234e+15	0	1	Yes
Embed Reduction	0	1	0	8.27329e+15	0	1	No
Conv. Dim. * No. Conv. Layers	-0	9	-0	-	1	1	No
Conv. Dim. * Aggr.	0	6	0	3.15026e+16	0	1	Yes
No. Conv. Layers * Aggr.	-0	6	-0	2.67417e+15	1	1	No
Conv. Dim. * Embed Reduction	-0	3	-0	1.22412e+16	1	1	No
No. Conv. Layers * Embed Reduction	-0	3	-0	3.60972e+16	1	1	No
Aggr. * Embed Reduction	-0	2	-0	4.70239e+14	1	1	No
Conv. Dim. * No. Conv. Layers * Aggr.	17.2683	18	0.9593	2.00577e+15	0	1	Yes
Conv. Dim. * No. Conv. Layers * Embed Reduction	2.1706	9	0.2412	1.95979e+30	0	1	Yes
Conv. Dim. * Aggr. * Embed Reduction	1.0447	6	0.1741	4.92684e+29	0	1	Yes
No. Conv. Layers * Aggr. * Embed Reduction	0.6272	6	0.1045	3.55686e+29	0	1	Yes
Conv. Dim. * No. Conv. Layers * Aggr. * Embed Reduction	3.7122	18	0.2062	2.1353e+29	0	1	Yes
Residual	0	14	0	4.21298e+29	nan	nan	No

Complexity and quality

The tuning process revealed a consistent trade-off between model complexity and illicit F1 scores. Simpler models, characterized by fewer layers and smaller convolutional dimensions, matched or outperformed more complex variants every time. Figure 2 illustrates the relationship between model complexity and illicit F1 scores across all datasets. As a result, simpler architectures were selected whenever performance differences were negligible.

Figure 3 presents only the interaction between complexity (number of layers and their size) and illicit F1 scores. For every dataset simpler models exhibit similar or better performance than more complex ones.

Model structure

This section presents the details of the final convolution layers selected for each dataset. Tables 7, Table 8, Table 9, and Table 10 summarize the final model configurations for the AMLSim 1/3, 1/5, 1/10, and 1/20 datasets, respectively.

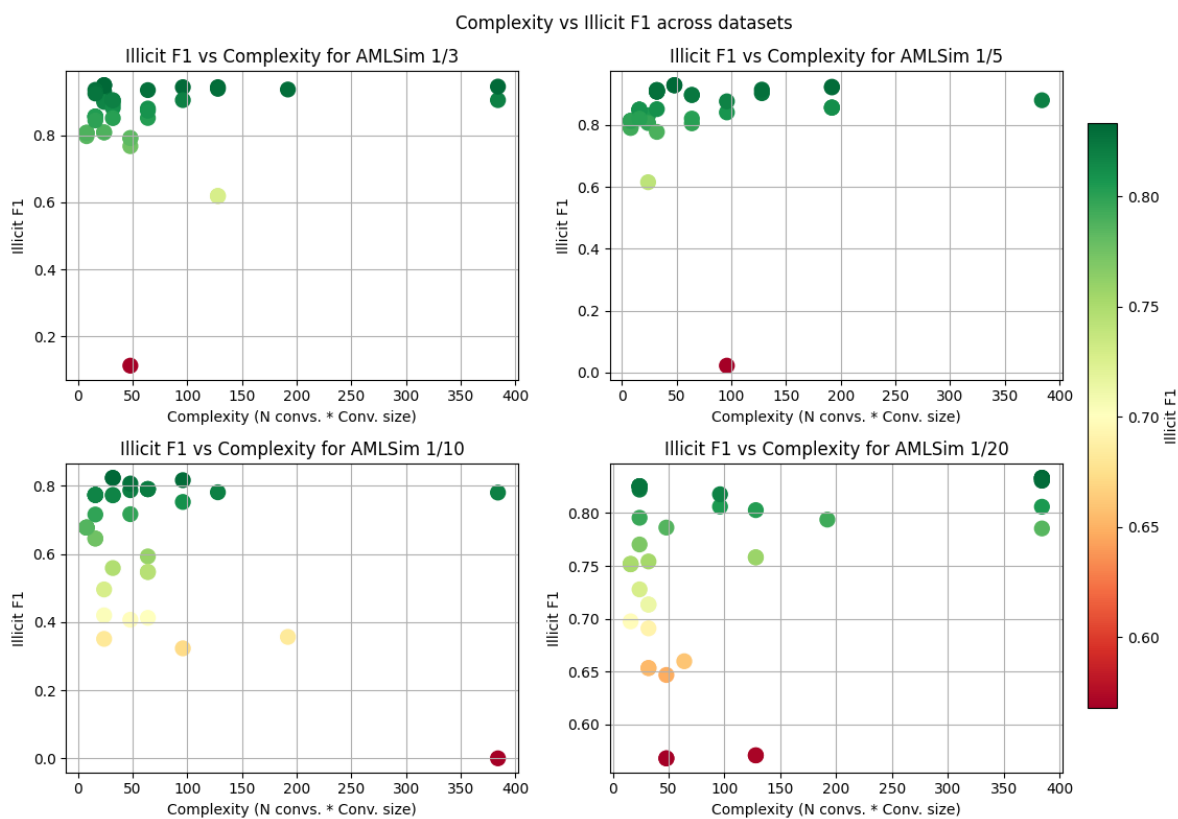


Figure 2: F1 score vs complexity across datasets.

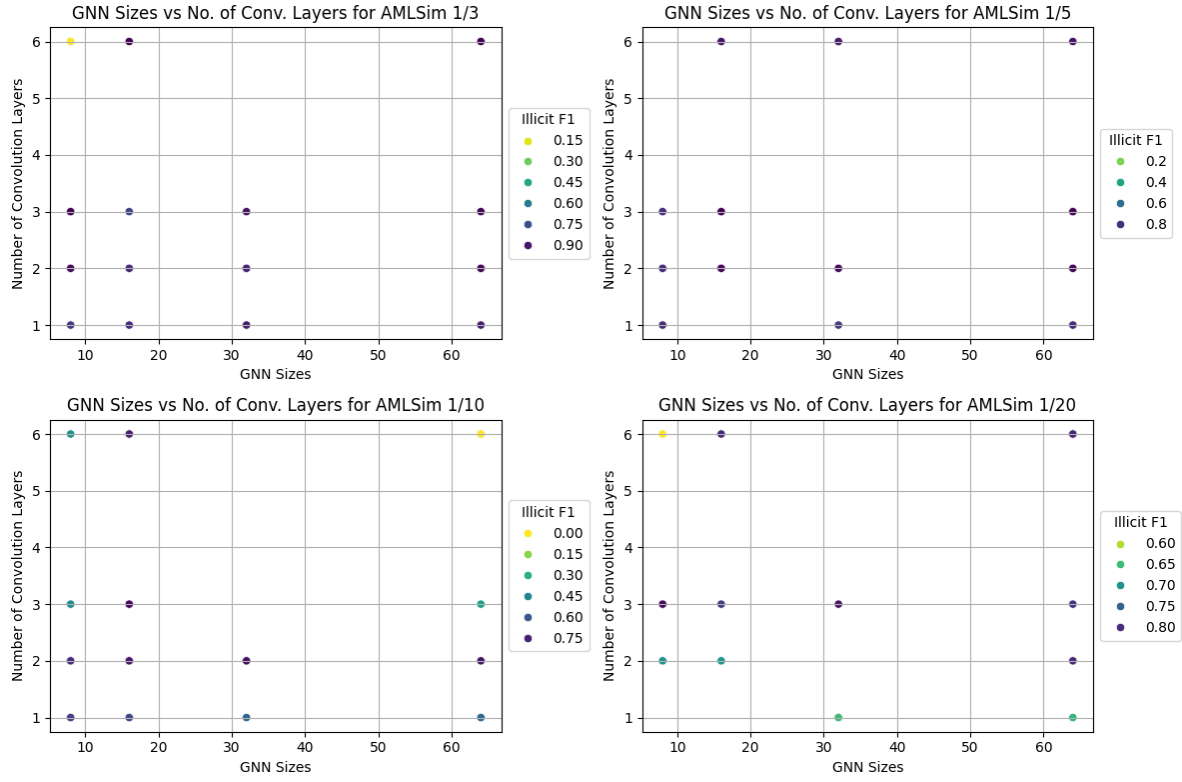


Figure 3: GNN sizes vs number of convolutions across datasets.

Table 7: Model summary for AMLSim 1/3.

Layer	Input Shape	Output Shape	#Param
MAGICModel	[3657, 6], [2, 1858], [1858, 8]	[3657, 16]	2,426
(convs)ModuleList	—	—	2,426
(0)MAGICConv	[3657, 6], [2, 1858], [1858, 8]	[3657, 16]	536
(1)MAGICConv	[3657, 16], [2, 1858], [1858, 8]	[3657, 16]	946
(2)MAGICConv	[3657, 16], [2, 1858], [1858, 8]	[3657, 16]	946

Table 8: Model summary for AMLSim 1/5.

Layer	Input Shape	Output Shape	#Param
MAGICModel	[6336, 6], [2, 3230], [3230, 8]	[6336, 8]	1,145
(convs)ModuleList	—	—	1,145
(0)MAGICConv	[6336, 6], [2, 3230], [3230, 8]	[6336, 8]	327
(1)MAGICConv	[6336, 8], [2, 3230], [3230, 8]	[6336, 8]	409
(2)MAGICConv	[6336, 8], [2, 3230], [3230, 8]	[6336, 8]	409

Table 9: Model summary for AMLSim 1/10.

Layer	Input Shape	Output Shape	#Param
MAGICModel	[12341, 6], [2, 6367], [6367, 8]	[12341, 16]	2,825
(convs)ModuleList	—	—	2,825
(0)MAGICConv	[12341, 6], [2, 6367], [6367, 8]	[12341, 16]	455
(1)MAGICConv	[12341, 16], [2, 6367], [6367, 8]	[12341, 16]	1,185
(2)MAGICConv	[12341, 16], [2, 6367], [6367, 8]	[12341, 16]	1,185

Table 10: Model summary for AMLSim 1/20.

Layer	Input Shape	Output Shape	#Param
MAGICModel	[24066, 6], [2, 12800], [12800, 8]	[24066, 16]	2,825
(convs)ModuleList	—	—	2,825
(0)MAGICConv	[24066, 6], [2, 12800], [12800, 8]	[24066, 16]	455
(1)MAGICConv	[24066, 16], [2, 12800], [12800, 8]	[24066, 16]	1,185
(2)MAGICConv	[24066, 16], [2, 12800], [12800, 8]	[24066, 16]	1,185

References

- Silva, Ítalo Della Garza, Luiz Henrique Andrade Correia, and Erick Galani Maziero. 2023. “Graph Neural Networks Applied to Money Laundering Detection in Intelligent Information Systems.” *Proceedings of the XIX Brazilian Symposium on Information Systems*, May, 252–59. <https://doi.org/10.1145/3592813.3592912>.
- Watanabe, Shuhei, Archit Bansal, and Frank Hutter. 2023. “PED-ANOVA: Efficiently Quantifying Hyperparameter Importance in Arbitrary Subspaces.” *arXiv Preprint arXiv:2304.10255*.
- Weber, Mark, Jie Chen, T. Suzumura, A. Pareja, Tengfei Ma, H. Kanezashi, Tim Kaler, C. Leiserson, and T. Schardl. 2018. “Scalable Graph Learning for Anti-Money Laundering: A First Look.” *ArXiv: 1812.00076*.