

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

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	3889	1186	1256	1977	599	637
AMLSim 1/5	6338	2036	2133	3226	1039	1090
AMLSim 1/10	12330	4239	4182	6366	2184	2160
AMLSim 1/20	24283	7945	8036	12931	4222	4267

Table 2: Class proportions per dataset

Dataset	# legal	% legal	# illicit	% illicit
AMLSim 1/3	1317	0.666	660	0.334
AMLSim 1/5	2566	0.795	660	0.205
AMLSim 1/10	5706	0.896	660	0.104
AMLSim 1/20	12271	0.949	660	0.051

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 importances

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.

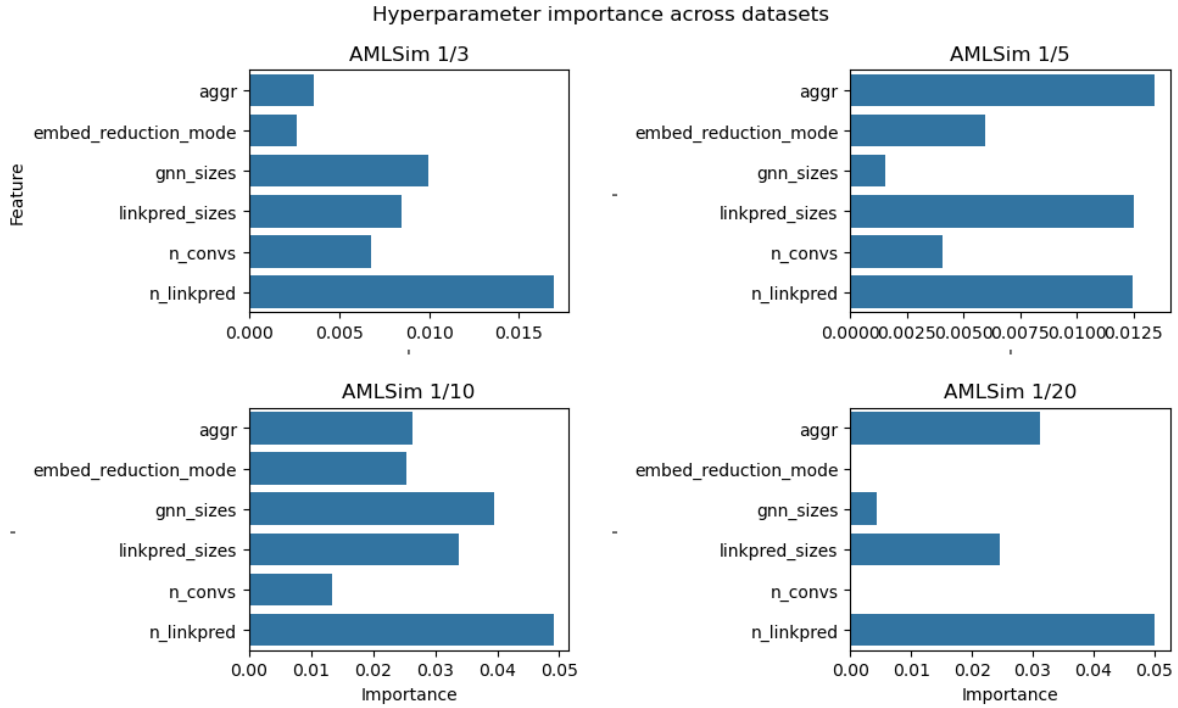


Figure 1: PED-ANOVA hyperparameter importances across datasets.

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	Significance
Conv. Dim.	0	3	0	0	1	0	No
No. Conv. Layers	-0	2	-0	-0	1	-0	No
Aggr.	-0	4	-0	-0	1	-0	No
Embed Reduction	0.0555	1	0.0555	3629.56	0	0.9959	Yes
Conv. Dim. * No. Conv. Layers	0	6	0	0	1	0	No
Conv. Dim. * Aggr.	14.0643	12	1.172	76630.3	0	1	Yes
No. Conv. Layers * Aggr.	-0	8	-0	-0	1	-0	No
Conv. Dim. * Embed Reduction	0	3	0	0	1	0	No
No. Conv. Layers * Embed Reduction	0.1503	2	0.0752	4914.07	0	0.9985	Yes
Aggr. * Embed Reduction	-0	4	-0	-0	1	-0	No
Conv. Dim. * No. Conv. Layers * Aggr.	20.9588	24	0.8733	57097.9	0	1	Yes
Conv. Dim. * No. Conv. Layers * Embed Reduction	0.1418	6	0.0236	1544.69	0	0.9984	Yes
Conv. Dim. * Aggr. * Embed Reduction	5.7815	12	0.4818	31500.9	0	1	Yes
No. Conv. Layers * Aggr. * Embed Reduction	0.7407	8	0.0926	6053.83	0	0.9997	Yes
Conv. Dim. * No. Conv. Layers * Aggr. * Embed Reduction	6.6888	24	0.2787	18222.3	0	1	Yes
Residual	0.0002	15	0	nan	nan	nan	No

Table 4: Anova results for AMLSim 1/5.

Source	SS	DF	MS	F	p-unc	np2	Significance
Conv. Dim.	-0	3	-0	-0	1	-0	No
No. Conv. Layers	-0	2	-0	-0	1	-0	No
Aggr.	0	4	0	0	1	0	No
Embed Reduction	0	1	0	0	1	0	No
Conv. Dim. * No. Conv. Layers	0	6	0	0	1	0	No
Conv. Dim. * Aggr.	0	12	0	0	1	0	No
No. Conv. Layers * Aggr.	0	8	0	0	1	0	No
Conv. Dim. * Embed Reduction	0	3	0	0	1	0	No
No. Conv. Layers * Embed Reduction	-0	2	-0	-0	1	-0	No
Aggr. * Embed Reduction	0	4	0	0	1	0	No

Table 4: Anova results for AMLSim 1/5.

Source	SS	DF	MS	F	p-unc	np2	Significance
Conv. Dim. * No. Conv. Layers *	24.1042	24	1.0043	94.3579	0	0.9934	Yes
Aggr.							
Conv. Dim. * No. Conv. Layers *	4.4824	6	0.7471	70.1874	0	0.9656	Yes
Embed Reduction							
Conv. Dim. * Aggr. * Embed	7.5262	12	0.6272	58.9242	0	0.9792	Yes
Reduction							
No. Conv. Layers * Aggr. *	1.6184	8	0.2023	19.0058	0.0001	0.9102	Yes
Embed Reduction							
Conv. Dim. * No. Conv. Layers *	4.029	24	0.1679	15.7718	0	0.9619	Yes
Aggr. * Embed Reduction							
Residual	0.1597	15	0.0106	nan	nan	nan	No

Table 5: Anova results for AMLSim 1/10.

Source	SS	DF	MS	F	p-unc	np2	Significance
Conv. Dim.	0.0868	3	0.0289	0.9192	0.4504	0.1267	No
No. Conv. Layers	0	2	0	0	1	0	No
Aggr.	0.0886	4	0.0222	0.7037	0.5991	0.129	No
Embed Reduction	-	1	-	-	1	-	No
	0.0101		0.0101	0.3214		0.0172	
Conv. Dim. * No. Conv. Layers	0	6	0	0	1	0	No
Conv. Dim. * Aggr.	0	12	0	0	1	0	No
No. Conv. Layers * Aggr.	0	8	0	0	1	0	No
Conv. Dim. * Embed Reduction	-	3	-	-	1	-	No
	0.0035		0.0012	0.0367		0.0058	
No. Conv. Layers * Embed	-0	2	-0	-0	1	-0	No
Reduction							
Aggr. * Embed Reduction	0	4	0	0	1	0	No
Conv. Dim. * No. Conv. Layers *	17.479	24	0.7283	23.1337	0	0.9669	Yes
Aggr.							
Conv. Dim. * No. Conv. Layers *	0.5281	6	0.088	2.7957	0.0682	0.4689	No
Embed Reduction							
Conv. Dim. * Aggr. * Embed	1.4988	12	0.1249	3.9674	0.0237	0.7148	Yes
Reduction							
No. Conv. Layers * Aggr. *	0.0882	8	0.011	0.3502	0.901	0.1285	No
Embed Reduction							
Conv. Dim. * No. Conv. Layers *	4.919	24	0.205	6.5104	0.0001	0.8916	Yes
Aggr. * Embed Reduction							

Table 5: Anova results for AMLSim 1/10.

Source	SS	DF	MS	F	p-unc	np2	Significance
Residual	0.5982	19	0.0315	nan	nan	nan	No

Table 6: Anova results for AMLSim 1/20.

Source	SS	DF	MS	F	p-unc	np2	Significance
Conv. Dim.	-0	3	-0	-0	1	-0	No
No. Conv. Layers	0.0072	2	0.0036	1.2612	0.31	0.1362	No
Aggr.	0	4	0	0	1	0	No
Embed Reduction	0.0028	1	0.0028	0.9897	0.3346	0.0583	No
Conv. Dim. * No. Conv. Layers	0	6	0	0	1	0	No
Conv. Dim. * Aggr.	-0	12	-0	-0	1	-0	No
No. Conv. Layers * Aggr.	0	8	0	0	1	0	No
Conv. Dim. * Embed Reduction	-0	3	-0	-0	1	-0	No
No. Conv. Layers * Embed Reduction	0	2	0	0	1	0	No
Aggr. * Embed Reduction	0	4	0	0	1	0	No
Conv. Dim. * No. Conv. Layers * Aggr.	14.7529	24	0.6147	215.651	0	0.9969	Yes
Conv. Dim. * No. Conv. Layers * Embed Reduction	0.7525	6	0.1254	43.9988	0	0.9429	Yes
Conv. Dim. * Aggr. * Embed Reduction	1.579	12	0.1316	46.162	0	0.9719	Yes
No. Conv. Layers * Aggr. * Embed Reduction	3.7461	8	0.4683	164.276	0	0.988	Yes
Conv. Dim. * No. Conv. Layers * Aggr. * Embed Reduction	3.9741	24	0.1656	58.0913	0	0.9887	Yes
Residual	0.0456	16	0.0029	nan	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.

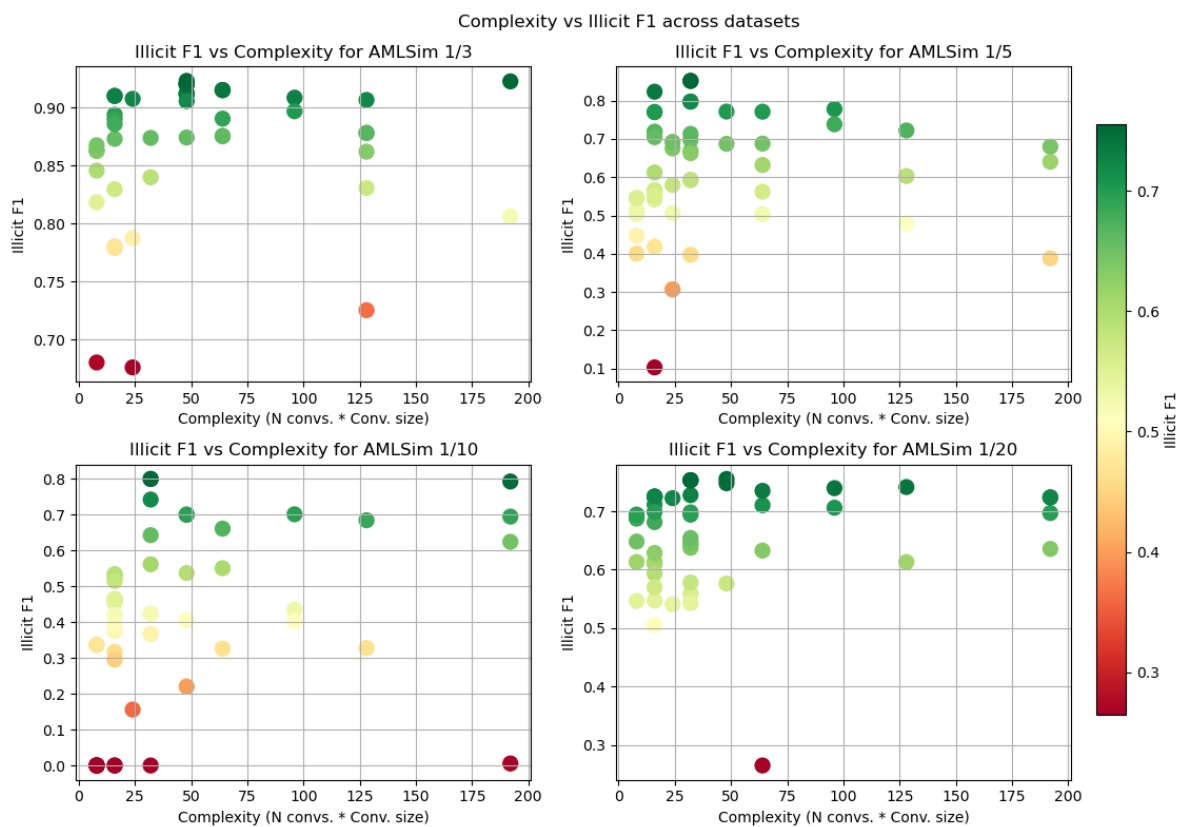


Figure 2: F1 score vs complexity across datasets.

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.

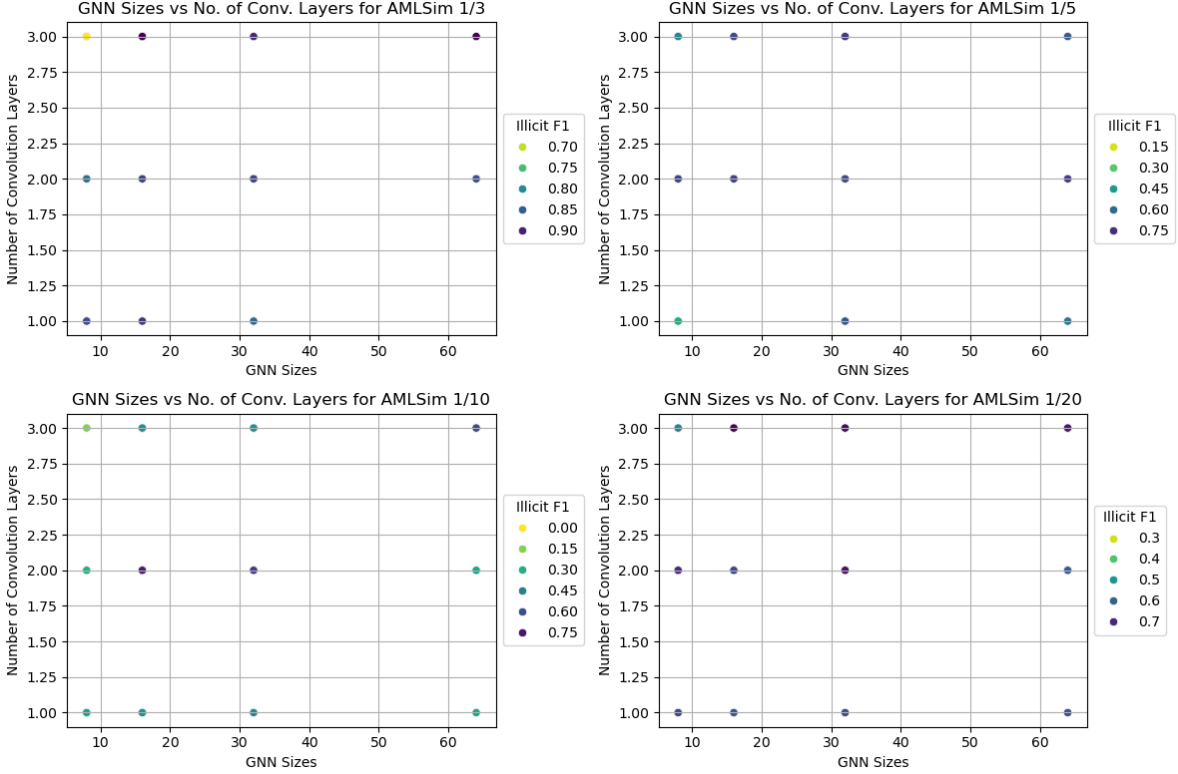


Figure 3: GNN sizes vs number of convolutions across datasets.

Model structure

This section presents the details of the final convolution layers selected for each dataset. Tables Table 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.

Table 7: Model summary for AMLSim 1/3.

Layer	Input Shape	Output Shape	#Param
MAGICModel	[3889, 6], [2, 1977], [1977, 8]	[3889, 8]	480
(convs)ModuleList	—	—	480
(0)MAGICConv	[3889, 6], [2, 1977], [1977, 8]	[3889, 8]	207

Table 7: Model summary for AMLSim 1/3.

Layer	Input Shape	Output Shape	#Param
(1)MAGICConv	[3889, 8], [2, 1977], [1977, 8]	[3889, 8]	273

Table 8: Model summary for AMLSim 1/5.

Layer	Input Shape	Output Shape	#Param
MAGICModel	[6338, 6], [2, 3226], [3226, 8]	[6338, 16]	1,288
(convs)ModuleList	–	–	1,288
(0)MAGICConv	[6338, 6], [2, 3226], [3226, 8]	[6338, 16]	359
(1)MAGICConv	[6338, 16], [2, 3226], [3226, 8]	[6338, 16]	929

Table 9: Model summary for AMLSim 1/10.

Layer	Input Shape	Output Shape	#Param
MAGICModel	[12330, 6], [2, 6366], [6366, 8]	[12330, 16]	1,640
(convs)ModuleList	–	–	1,640
(0)MAGICConv	[12330, 6], [2, 6366], [6366, 8]	[12330, 16]	455
(1)MAGICConv	[12330, 16], [2, 6366], [6366, 8]	[12330, 16]	1,185

Table 10: Model summary for AMLSim 1/20.

Layer	Input Shape	Output Shape	#Param
MAGICModel	[24283, 6], [2, 12931], [12931, 8]	[24283, 8]	1,728
(convs)ModuleList	–	–	1,728
(0)MAGICConv	[24283, 6], [2, 12931], [12931, 8]	[24283, 32]	663
(1)MAGICConv	[24283, 32], [2, 12931], [12931, 8]	[24283, 8]	1,065

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*.