Ranking of Classification Association Rules through Graph Convolutional Networks

Maicon Dall'Agnol

Summary

INTRODUCTION TO ASSOCIATIVE CLASSIFIERS

- Classification Association Rules is a special type of Association Rules
  - Used to generate a Classifier
  - Associative Classifier (AC)
  - Same format in AR **(A ⇒ B)**, but **B is a class**

- AC vs other Interpretable approaches
  - **Accuracy** competitive with other interpretable models
  - **Usability**, no need for complete retraining
  - **Natural interpretability**
  - **Lower** accuracy compared to black-box models

- **Trade-off** between **Interpretability vs Accuracy**
  - Discussed in Rudin (2019)

ASSOCIATIVE CLASSIFIER STRUCTURE

- CBA (Classification Based on Association) baseline and the first algorithm in AC family

- 3 steps are necessary to build an AC:
  - **Extraction** (association rule induction)
  - **Sorting (Ranking)** and pruning CARs
    - Work in the area has shown great impact
  - **Prediction** (classification algorithm)

- Ranking method (Rule 1 precedes Rule 2) if:
  - – Confidence [Rule 1] > Confidence [Rule 2]
  - – (Tiebreaker 1) Support [Rule 1] > Support [Rule 2]
  - – (Tiebreaker 2) ID [Rule 1] > ID [Rule 2]

OBJECTIVE MEASURES

- Objective Measures (OMs) are math functions to compute values to a rule

- Formulated from the combination of the frequency of A, B and AB, being described in probability **[P(A), P(B) and P(AB)]**

- All steps of AC use OMs

- Confidence (Conf) and Support (Sup) are the main ones
  - **Conf = P(AB)/P(A)**
  - **Sup = P(AB)**

- More than 60 OMs are available in the literature [6]
  - Examples: Lift, Confidence, Complement Class Support, Leverage,Conditional Entropy, etc.
  - **This work use 44 OMs**

GRAPH CONVOLUTIONAL NETWORKS

- "We consider the problem of classifying nodes (such as documents) in a graph (such as a citation network), where labels are **only available for a small subset of nodes**." (Kipf, Thomas N., and Max Welling, 2016)
- Semi-supervised approach
- Connected graph used to "propagate" labels through the network
- Convolution is performed over the connections of the graph

CONTEXT AND MOTIVATION

- Doesn’t have best measure for all cases. Literature works tried:
  - Create new measures
  - Aggregate measures (combine many OMs)
  - Several Ranking methods **do not take into account the relationship between rules**

- Recent work shows gains in approaches that explore several equations at the same time (different ways of ranking)

- Graph Convolutional Networks
  - Recent approach
  - Not applied in AC
  - Explores the relationship between RACs and different ways of ranking at the same time

- Compare the results with other aggregation methods in literature

RELATED WORKS

- Dall'Agnol and Carvalho (2023) present the use of Borda's methods to ranking CARs in contrast to the use of CSC method inside CBA algorithm
  - Good results comparing to other algorithms like in Silva and Carvalho (2018), and Dall'Agnol and Carvalho (2020)

- **AC.Rank$_A$** (Dall'Agnol and Carvalho, 2024) expands the 2023 works to news methods and more OMs groups, also expanding the experimental setup to other algorithms besides CBA
  - Good results compared to CBA and others configurations
  - Improve the model size while maintaining the F1 score without significant difference with the others

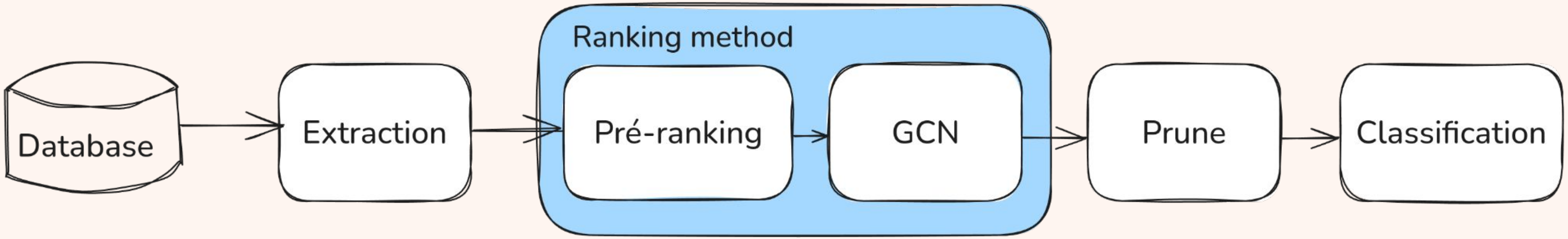
- Bui-Thi, Meysman and Laukens (2022):
  - MO aggregation in ranking via multi-objective optimization modeling.
  - Use Neural network to build an Objective Measure for each database
  - Good results in Model size, but with bad results at runtime

PROPOSED METHOD

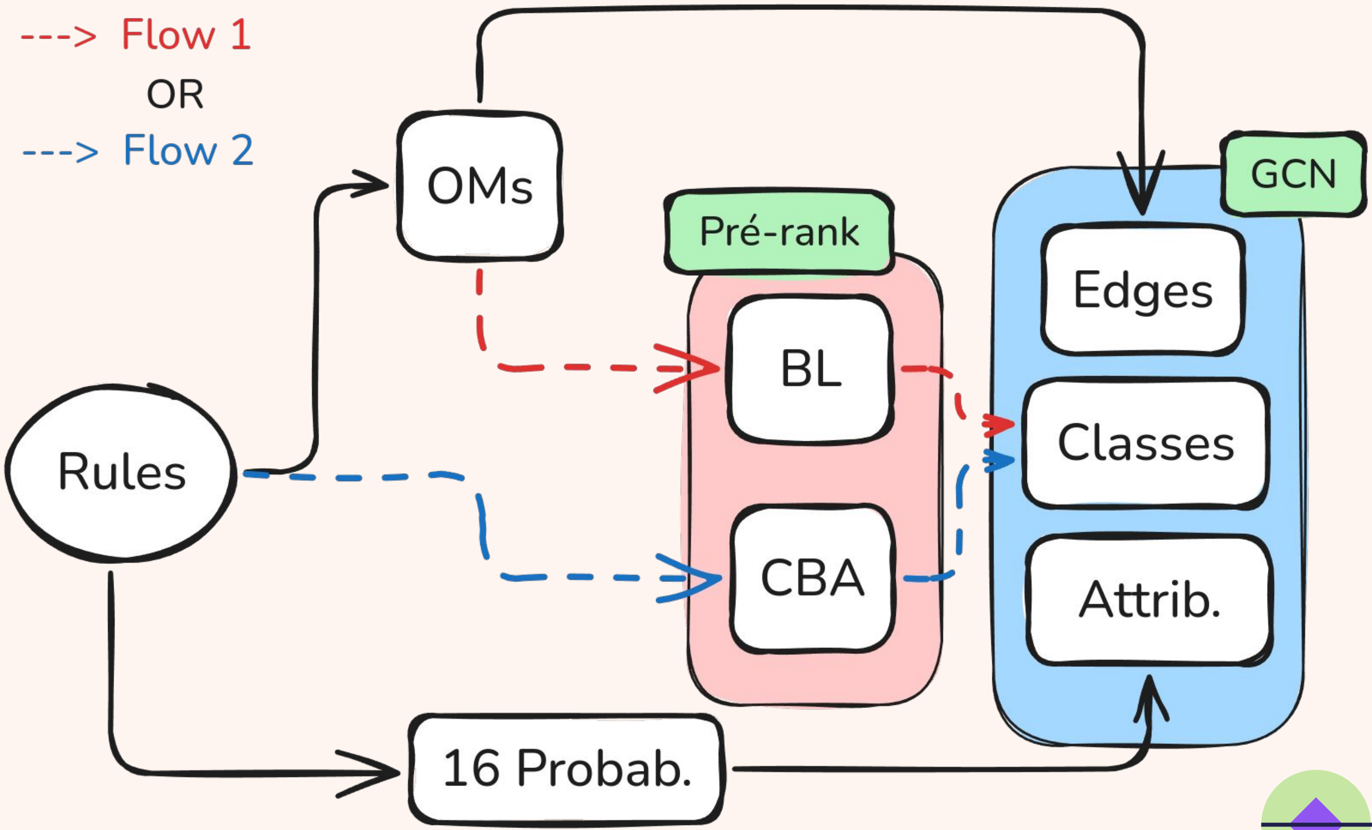
- Use **GCN for re-rank rules** inside CBA algorithm
  - Use **multiple OMs** as attributes in GCN (attributes)
  - Connect rules (nodes) by **similarity between the OMs of rules**(edges)
  - Use **pre-ranking method to classify some rules** (semi-supervised classes)

- These 3 parts are needed to build GCN
  - Some classes
  - Attributes for nodes
  - Edges between nodes

PROPOSED METHOD



POSSIBLE FLOWS



ATTRIBUTES

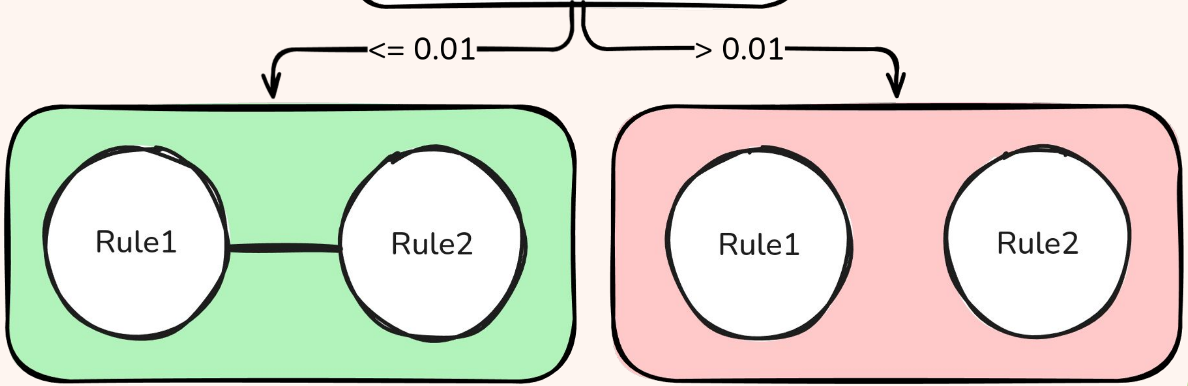
P(A)	A
P(B)	B
P(~B)	1 - B
P(~A)	1 - A
P(A~B)	A - AB
P(~AB)	B - AB
P(~A~B)	1 - B - A + AB
P(B|A)	AB/A
P(A|B)	AB/B
P(B|~A)	~AB/~A
P(A|~B)	A~B/~B
P(~B|~A)	1 - B_notA
P(~A|~B)	1 - A_notB
P(~B|A)	1 - B_A
P(~A|B)	1 - A_B

+60 OMs

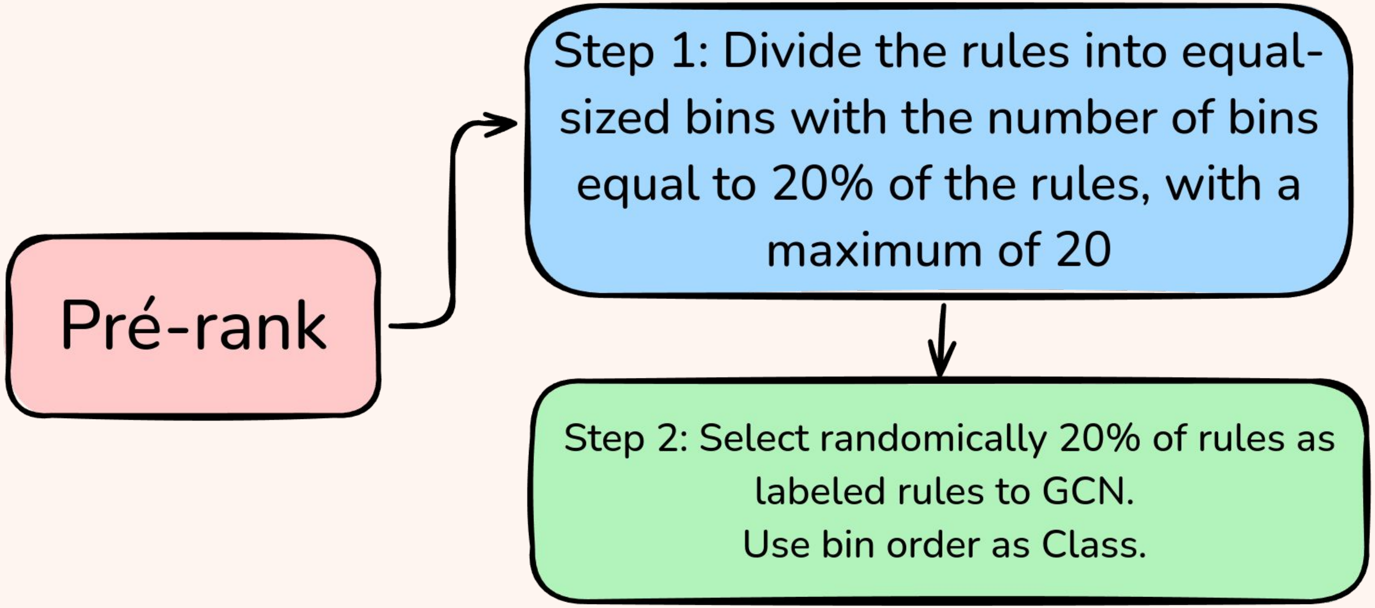
EDGES

	OM1	OM2	OM3
R1	0.13	0.6	0.15
R2	0.17	0.5	0.98

$$d(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^{n} |x_i - y_i|}{n}$$



CLASSES

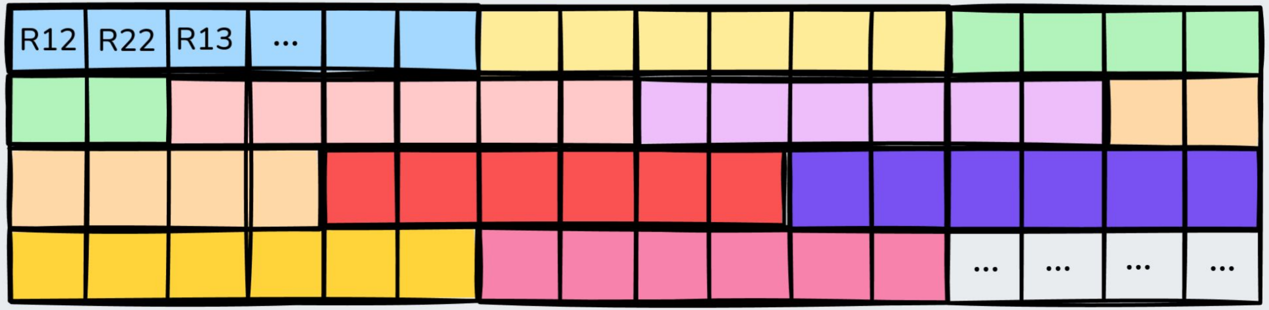

Pré-rank
Step 1: Divide the rules into equal-sized bins with the number of bins equal to 20% of the rules, with a maximum of 20
Step 2: Select randomically 20% of rules as labeled rules to GCN.
Use bin order as Class.

CLASSES EXAMPLE

Example

128 rules x 20% = 25,6 bins.

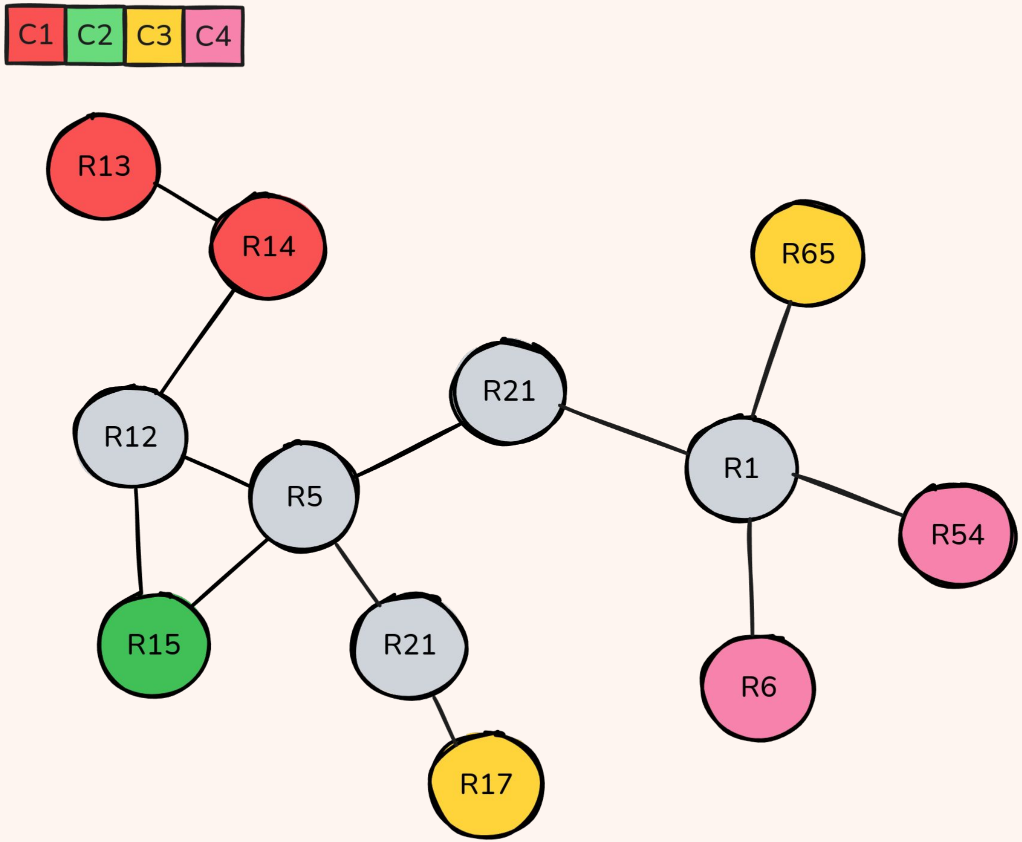
Max is 20 bins. Thats result in 20 bins of 6 or 7 values (128/20 == 6.4)



Select 20% of rules randomically (25,6 rules -> 25 rules)

R13	R31	R17	R25	R10	R19	...
C1	C2	C3	C4	C5	C5	...

FINAL EXEMPLE



EXPERIMENTAL SETUP

- **42 datasets**
  - Evaluation of **F1-Macro** and **Model Size** using 10-fold Cross-Validation
  - Discretized by Fayyad and Irani (1993)
  - Balanced sets (1:2.5)

- Basic parameters for extract rules:
  - Variable support per set (maximum of 10,000 rules)
  - 0% Minimum confidence for extraction
  - Maximum rule size: 5, except in four sets

- Evaluation of F1-score Macro and Model size
  - Average values of 10-fold CV
  - **Friedman** test and **Nemenyi test** with Critical Difference plot

EXPERIMENTAL SETUP

- 5 OMs groups
  - **CS** - Confidence and Support (from CBA)
  - **TW** - Tew's Group (Tew et al., 2014)
  - **GF** - Guangfei's Group (Yang e Cui, 2015))
  - **G1** - G1 from AC.Rank$_A$ (Dall'Agnol et al., 2024)
  - **G2** - G2 from AC.Rank$_A$ (Dall'Agnol et al., 2024)

- 2 Methods in Pré-ranking step
  - BL and CBA

- **GCN Parameters**
  - Learning Rate: 0.01
  - Number of neurons in the hidden layer: 32
  - Number of epochs: 200
  - Output equal to the number of classes used (limit of 20)

RESULTS

	[CBA]	[CS]			[G1]			[G2]			[GF]			[TW]		
	[CBA]	AC.Rank [BL]	GCN [BL]	GCN [CBA.r]	AC.Rank [BL]	GCN [BL]	GCN [CBA.r]	AC.Rank [BL]	GCN [BL]	GCN [CBA.r]	AC.Rank [BL]	GCN [BL]	GCN [CBA.r]	AC.Rank [BL]	GCN [BL]	GCN [CBA.r]
Australian	0,8659	0,8701	0,8429	0,8703	0,8658	0,8687	0,8659	0,8661	0,8662	0,8626	0,8615	0,8703	0,8601	0,8692	0,8549	0,8641
Banana	0,7417	0,6812	0,6	0,6495	0,7456	0,7253	0,6854	0,7363	0,6747	0,6955	0,7398	0,7225	0,6492	0,6817	0,6846	0,6974
Bands	0,5768	0,5262	0,4596	0,5831	0,587	0,6179	0,6304	0,5936	0,6076	0,6109	0,6092	0,6153	0,6043	0,553	0,4649	0,5854
Breast	0,6188	0,6224	0,5939	0,6101	0,6138	0,6188	0,6188	0,6251	0,6559	0,6067	0,6067	0,6282	0,6147	0,6241	0,6461	0,627
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Wine	0,905	0,905	0,9195	0,9276	0,905	0,905	0,905	0,9343	0,9448	0,9144	0,9115	0,9147	0,9028	0,9268	0,9201	0,9215
Wisconsin	0,9493	0,9602	0,9412	0,939	0,9493	0,9493	0,9493	0,9436	0,9443	0,899	0,951	0,9537	0,9537	0,9491	0,7074	0,937

INITIAL DISCUSSION

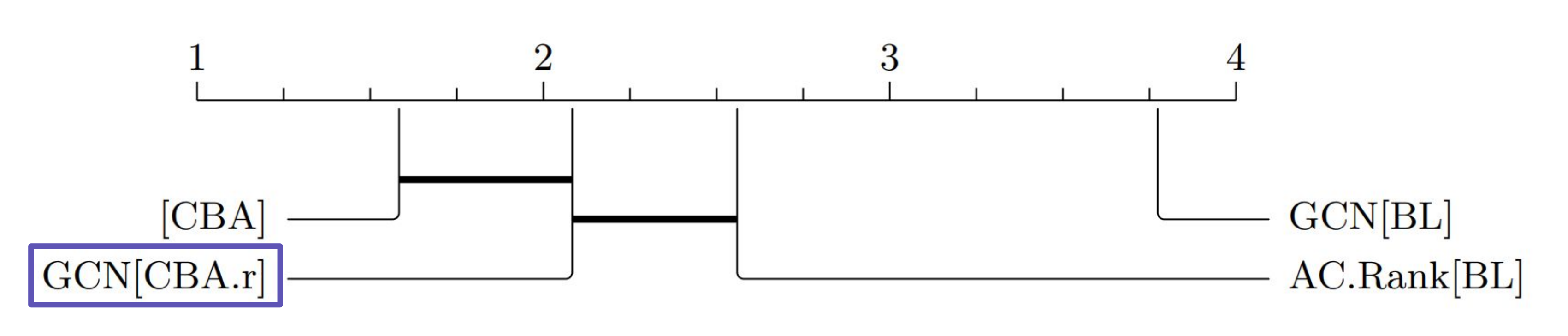
- Friedman test for all cases
  - All values < 0.05

- Nemenyi test with Critical Plot diagram plotted

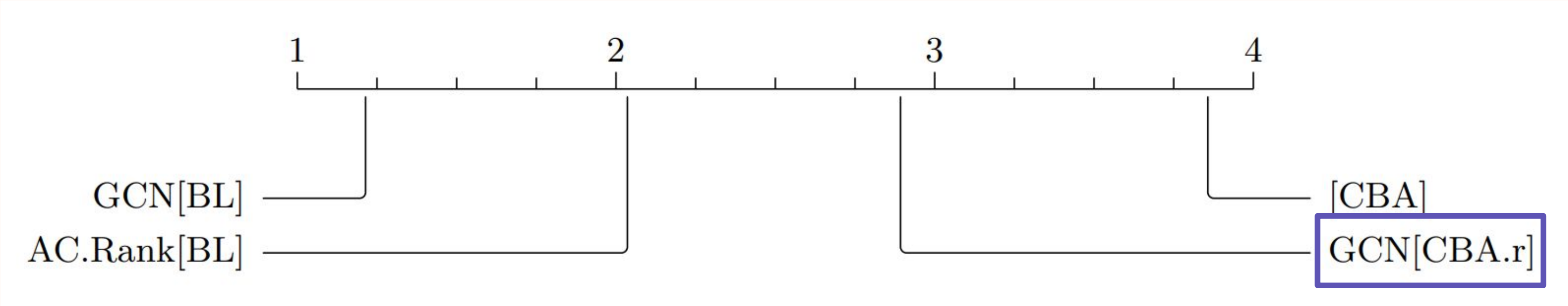
- Values calculated over mean for 10-fold CV and for the 42 datasets

DISCUSSION - [CS] group

- CD for Nemenyi test for F1-score

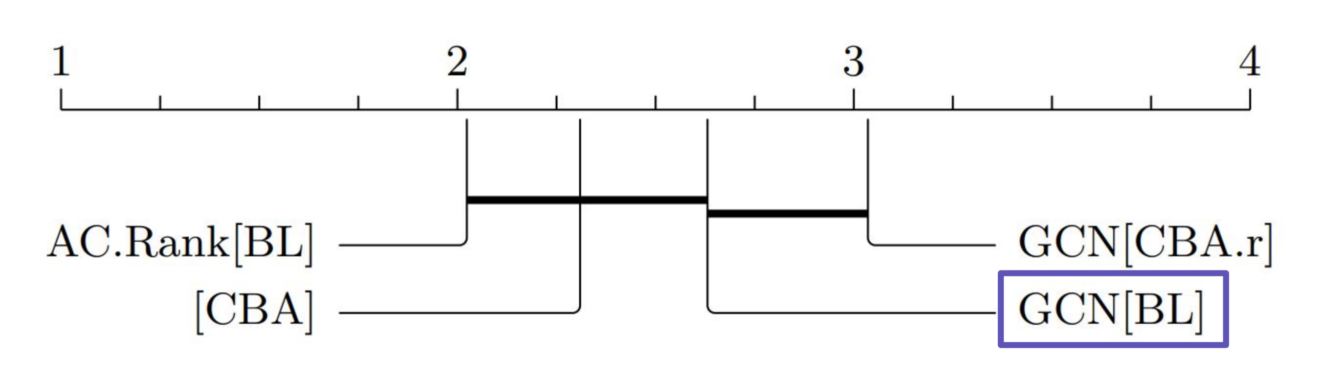


- CD for Nemenyi test for Size

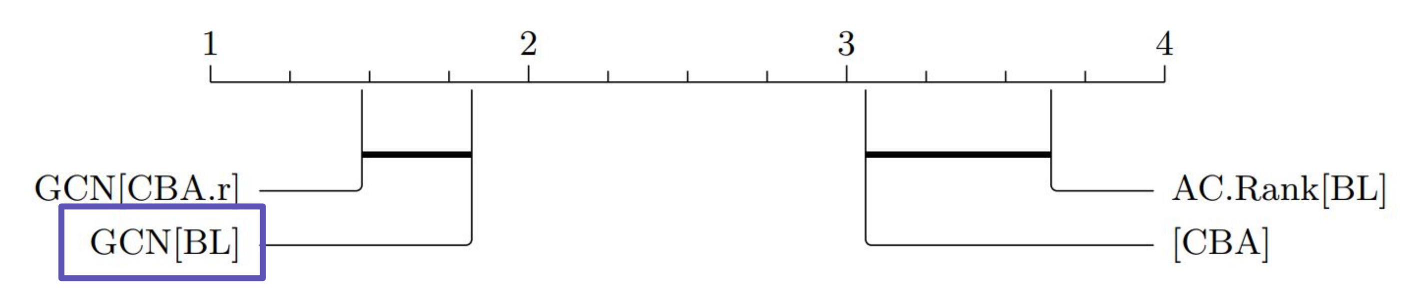


DISCUSSION - [G1] group

- CD for Nemenyi test for F1-score

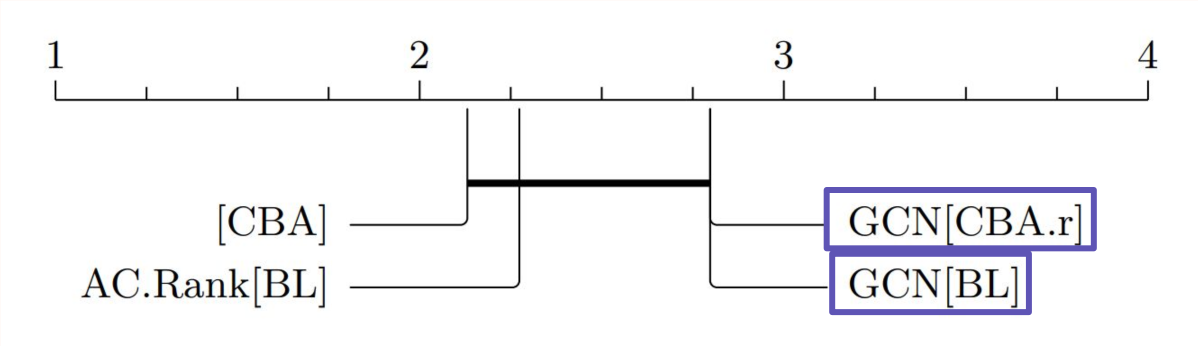


- CD for Nemenyi test for Size

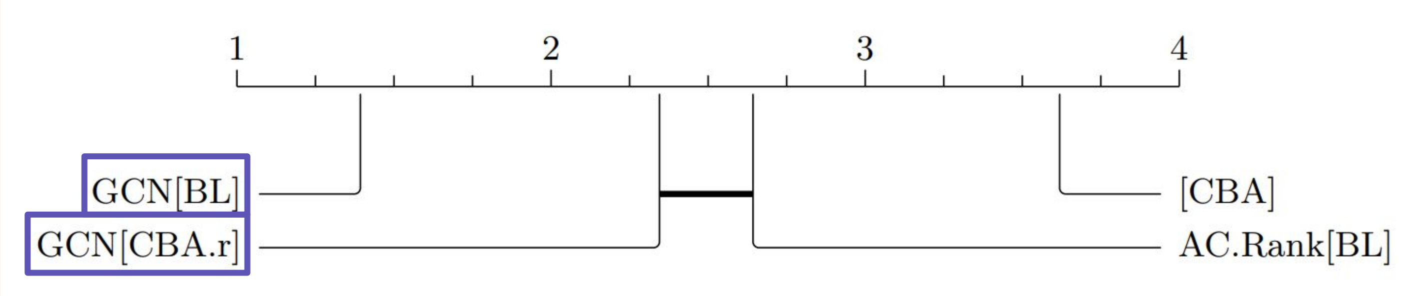


DISCUSSION - [G2] group

- CD for Nemenyi test for F1-score

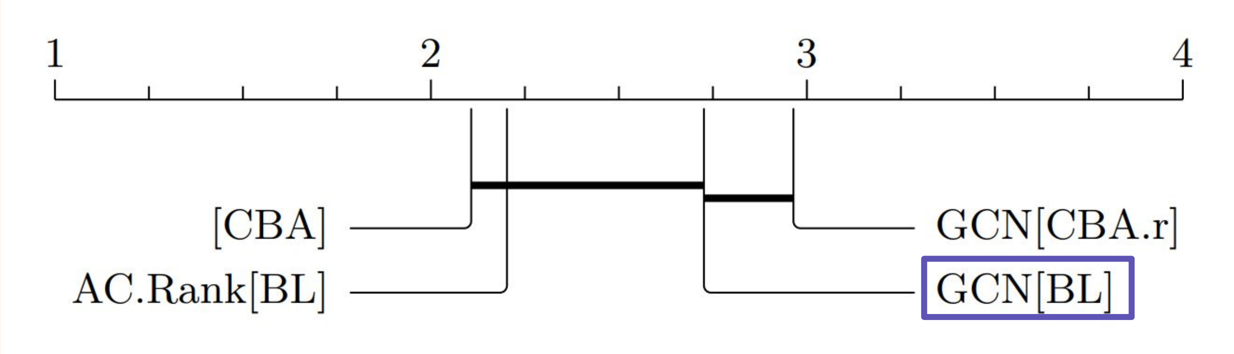


- CD for Nemenyi test for Size

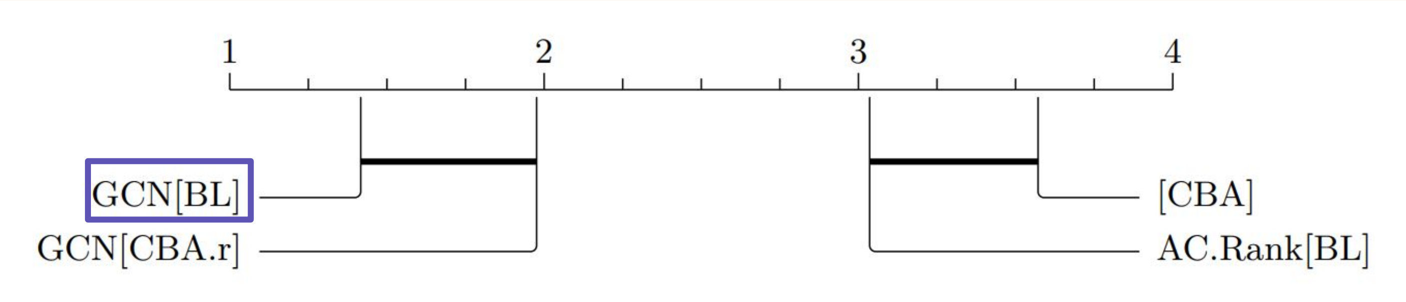


DISCUSSION - [GF] group

- CD for Nemenyi test for F1-score

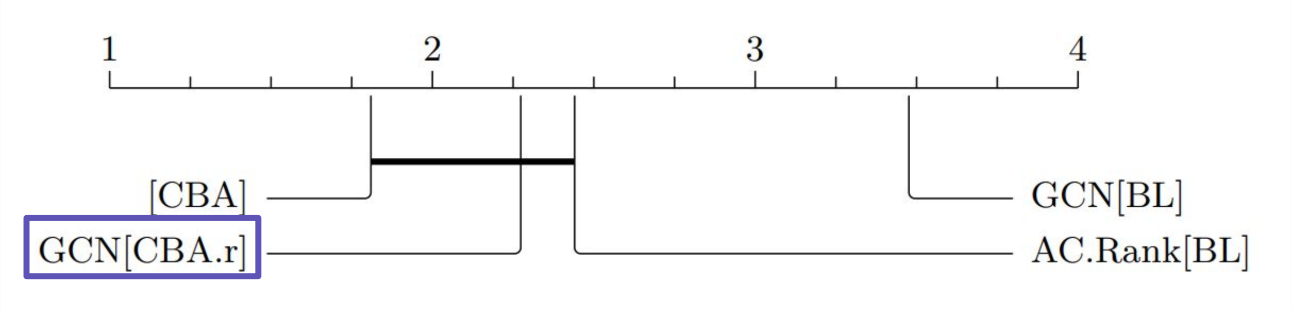


- CD for Nemenyi test for Size

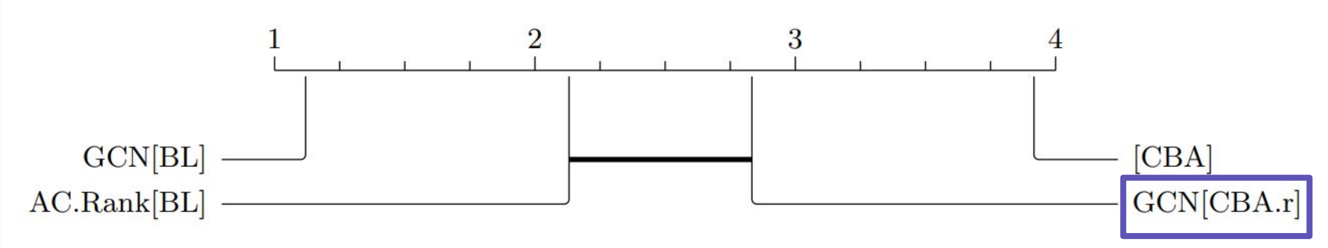


DISCUSSION - [TW] group

- CD for Nemenyi test for F1-score



- CD for Nemenyi test for Size



SUMMARY OF RESULTS

- First value F1
- Second indicate Model size

- Dot represents algorithm in same group

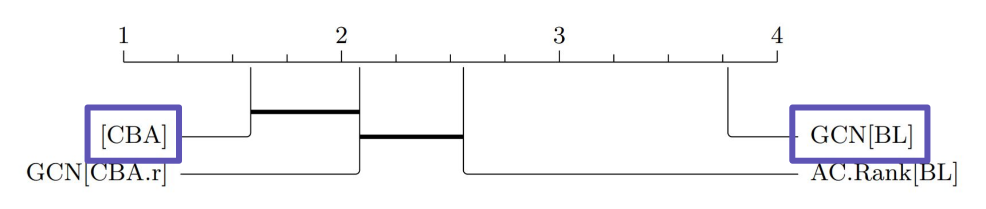
- Arrow points to algorithm with difference and to the better

- Exemple:
  - CBA > GCN[BL] in F1
  - GCN[BL] > [CBA] in M.S.

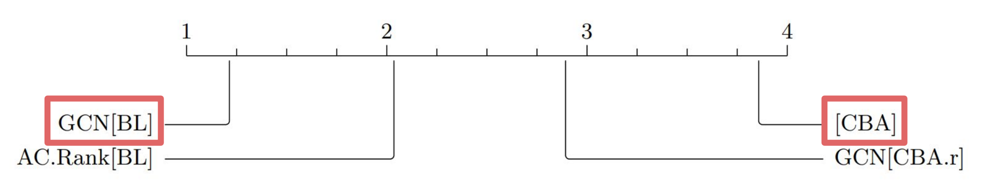
- Viewing only half the matrix is enough to see all the comparisons

		GCN[BL]	GCN[CBA.r]	AC.Rank[BL]
[CS]	[CBA]	⬅️⬆️	●⬆️	⬅️⬆️
	GCN[BL]		⬆️⬅️	⬆️⬅️
	GCN[CBA.r]			●⬆️

- F1



- Model size



SUMMARY OF RESULTS

- GCN[BL] and GCN[CBA.r] have advantage over [CBA] in 3 OMs group
  - F1 without difference and better model size

- GCN[BL] have advantage over AC.Rank[BL] 3 times
- and 1 over GCN[CBA.r]

- AC.Rank[BL] have one advantage over GCN[CBA.r]

- All other comparisons have equal group or one arrow pointing for each algorithm

		GCN[BL]	GCN[CBA.r]	AC.Rank[BL]
[CS]	[CBA]	⬅️⬆️	•⬆️	⬅️⬆️
	GCN[BL]		⬆️⬅️	⬆️⬅️
	GCN[CBA.r]			•⬆️
[G1]	[CBA]	•⬆️	⬅️⬆️	••
	GCN[BL]		••	•⬅️
	GCN[CBA.r]			⬆️⬅️
[G2]	[CBA]	•⬆️	•⬆️	•⬆️
	GCN[BL]		•⬅️	•⬅️
	GCN[CBA.r]			••
[GF]	[CBA]	•⬆️	⬅️⬆️	••
	GCN[BL]		••	•⬅️
	GCN[CBA.r]			⬆️⬅️
[TW]	[CBA]	⬅️⬆️	•⬆️	•⬆️
	GCN[BL]		⬆️⬅️	⬆️⬅️
	GCN[CBA.r]			••

SUMMARY OF RESULTS

- GCN[BL] shows better performance over others
- Best result in G2
    - Result similar to Dall'Agnol (2024)
- GCN[BL] have good performance in decrease model size in all comparisons

		GCN[BL]	GCN[CBA.r]	AC.Rank[BL]
[CS]	[CBA]	⬅️⬆️	•⬆️	⬅️⬆️
	GCN[BL]		⬆️⬅️	⬆️⬅️
	GCN[CBA.r]			•⬆️
[G1]	[CBA]	•⬆️	⬅️⬆️	••
	GCN[BL]		••	•⬅️
	GCN[CBA.r]			⬆️⬅️
[G2]	[CBA]	•⬆️	•⬆️	•⬆️
	GCN[BL]		•⬅️	•⬅️
	GCN[CBA.r]			••
[GF]	[CBA]	•⬆️	⬅️⬆️	••
	GCN[BL]		••	•⬅️
	GCN[CBA.r]			⬆️⬅️
[TW]	[CBA]	⬅️⬆️	•⬆️	•⬆️
	GCN[BL]		⬆️⬅️	⬆️⬅️
	GCN[CBA.r]			••

CONCLUSIONS

- Results in GCN[BL] and GCN[CBA.r] shows a good performance with advantage over [CBA] tradicional ranking method

- GCN methods have better performance in model size, decreasing the size several times

- GCN[BL] have the best results against [CBA], GCN[CBA.r] and AC.Rank[BL]
  - 7 times of 14 possibibles
  - 7 wins
  - 2 draws type same F1 and Model size
  - 2 draws type better F1
  - 3 draws type better Model size

- Execution time could be a problem, with GCN with much longer training time than [CBA] and AC.Rank[BL]

BIBLIOGRAPHY

- J. Filip and T. Kliegr, “Classification based on associations (CBA) - a performance analysis,” EasyChair Preprint no. 501, 2018.

- M. Dall’Agnol and V. O. De Carvalho, “Ranking rules in associative classifiers via borda’s methods,” in 2023 18th Iberian Conference on Information Systems and Technologies (CISTI), 2023, pp. 1–6.

- M. Dall’Agnol and V. O. Carvalho, “Objective measures ensemble in associative classifiers,” in Proceedings of the 22nd International Conference on Enterprise Information Systems - Volume 1: ICEIS, Insticc. SciTePress, 2020, pp. 83–90.

- M. Dall’Agnol and V. O. d. Carvalho, "AC.RankA: Rule Ranking Method via Aggregation of Objective Measures for Associative Classifiers," in IEEE Access, vol. 12, pp. 88862-88882, 2024, doi: 10.1109/ACCESS.2024.3419130.

- T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” arXiv preprint arXiv:1609.02907, 2016.

Thank you!

Any question?