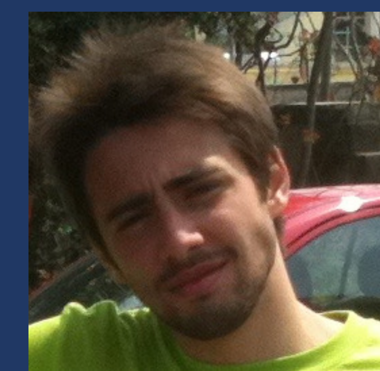
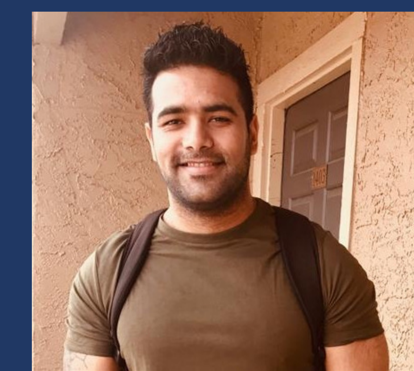
# Generative CLausal Networks: Relational Decision Trees as Probabilistic Circuits

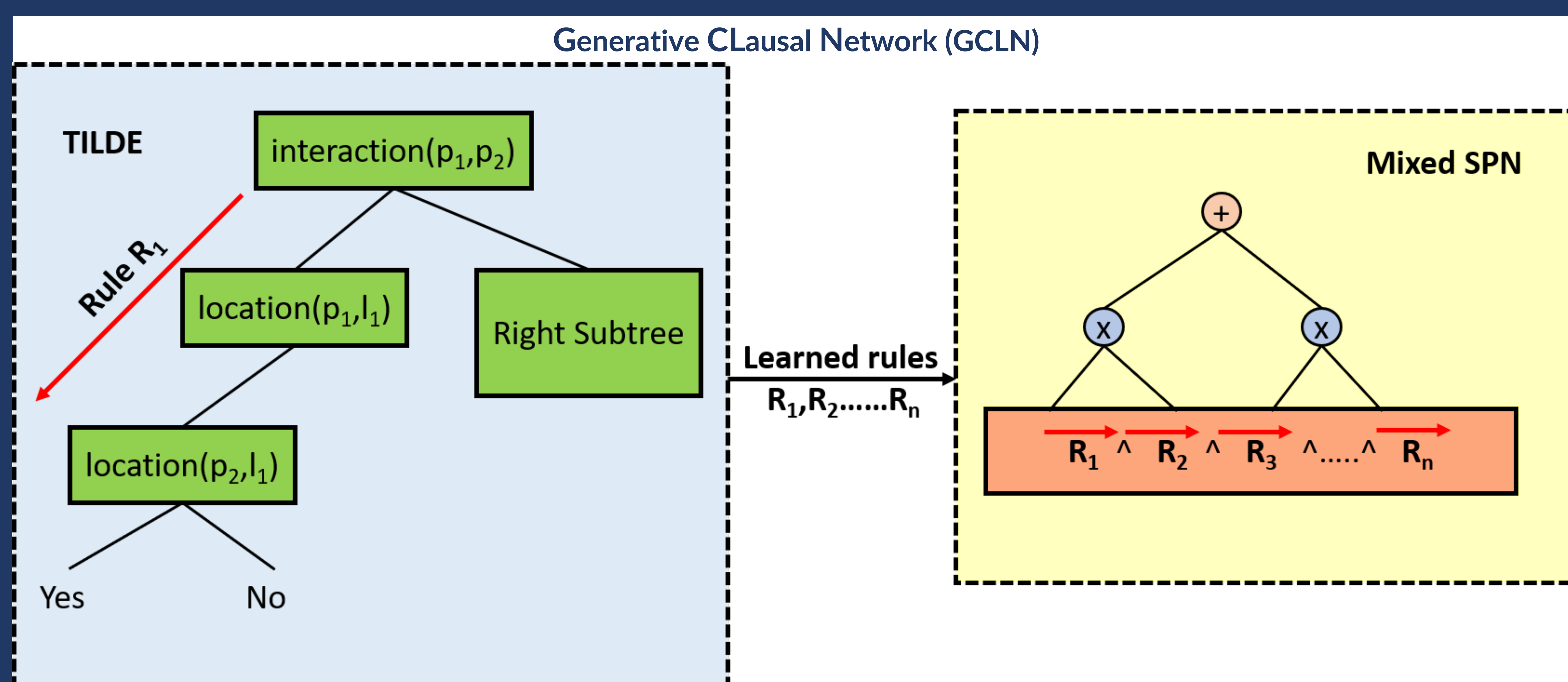





Fabrizio Ventola[1], Devendra Singh Dhami[1], Kristian Kersting[1,2]
[1]Department of Computer Science, TU Darmstadt, Darmstadt, Germany
[2]Centre for Cognitive Science, TU Darmstadt, and Hessian Center for AI (hessian.AI)

**GCLN** is the first connection between relational rule models and probabilistic circuits, specifically, **Sum-Product Networks** (SPNs) for **hybrid** domains [1]. **GCLN** equips a relational rule model with a tractable joint distribution over the class and rules fire counts to learn an **accurate** probabilistic **classifier** which outperforms propositional and (statistical) relational ones, also when **data** is **missing** or **partially observed**. Thanks to its generative power and inference capabilities it can also perform **out-of-domain detection**, **missing data imputation**, and provide clear model **interpretations**.



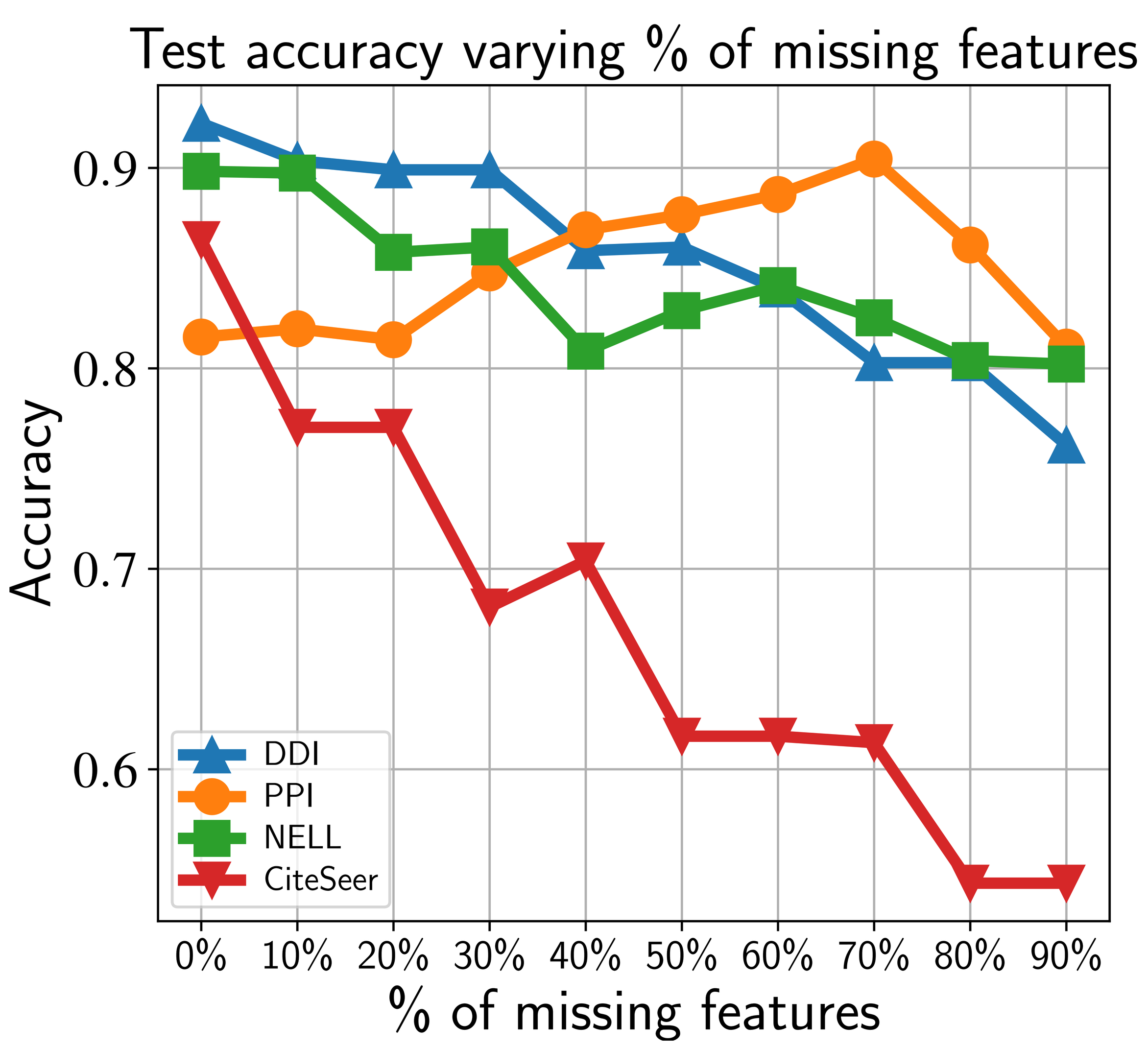

## I. Accurate Classifier

| Data | Methods | Accuracy | AUC-ROC | AUC-PR | Data | Methods | Accuracy | AUC-ROC | AUC-PR |
|---|---|---|---|---|---|---|---|---|---|
| *DDI* | LR | **93.79** | **87.62** | 83.64 | *NELL* | LR | 83.96 | 58.90 | 28.27 |
| | GB | 86.77 | 86.17 | 67.95 | | GB | 87.69 | 75.28 | 48.07 |
| | NN | 87.54 | 85.03 | 69.21 | | NN | 88.26 | 74.42 | 48.35 |
| | DT | 85.52 | 83.22 | 65.12 | | DT | 87.27 | 71.77 | 45.71 |
| | TILDE | 72.52 | 73.43 | 70.79 | | TILDE | 81.48 | 86.27 | 78.23 |
| | RDN-B | 75.54 | 82.87 | 83.13 | | RDN-B | 81.26 | 88.47 | 83.41 |
| | MLN-B | 63.80 | 79.83 | 78.40 | | MLN-B | 60.54 | **89.44** | **85.30** |
| | GCLN-B | 85.05 | 70.57 | 70.84 | | GCLN-B | 80.74 | 38.72 | 23.15 |
| | GCLN-P | 92.22 ↑ | 87.53 ↑ | **83.81** ↑ | | GCLN-P | **89.83** ↑ | **89.44** ↑ | 56.39 ↑ |
| *PPI* | LR | 78.13 | 81.54 | 52.44 | *CiteSeer* | LR | 76.33 | 83.78 | 53.30 |
| | GB | 77.21 | 78.25 | 49.54 | | GB | 96.05 | **97.21** | 87.24 |
| | NN | 76.94 | 75.75 | 47.49 | | NN | **96.50** | 96.88 | 88.10 |
| | DT | 76.47 | 77.52 | 48.71 | | DT | 95.33 | 96.79 | 85.38 |
| | TILDE | 62.20 | 62.87 | 58.27 | | TILDE | 91.47 | 83.33 | 73.45 |
| | RDN-B | 67.15 | 72.84 | 74.02 | | RDN-B | 94.72 | 97.11 | **89.23** |
| | MLN-B | 54.87 | 74.39 | 73.34 | | MLN-B | 81.98 | 94.67 | 80.54 |
| | GCLN-B | 79.72 | 82.75 | 59.43 | | GCLN-B | 77.18 | 71.34 | 41.20 |
| | GCLN-P | **81.56** ↑ | **91.16** ↑ | **80.08** ↑ | | GCLN-P | 86.42 ↑ | 71.57 ↑ | 42.57 ↑ |

**GCLNs outperform in classification accuracy (statistical-)relational and propositional classifiers including NNs and ensemble methods.**

## II. Robust to Missing and Partially Observed Data

| Data | Accuracy | AUC-ROC | AUC-PR |
|---|---|---|---|
| *DDI* | $89.90 \pm 0.035$ | $79.23 \pm 0.036$ | $75.00 \pm 0.040$ |
| *PPI* | $84.77 \pm 0.025$ | $89.18 \pm 0.024$ | $79.83 \pm 0.040$ |
| *NELL* | $86.06 \pm 0.054$ | $79.96 \pm 0.063$ | $47.30 \pm 0.092$ |
| *CiteSeer* | $68.09 \pm 0.103$ | $50.69 \pm 0.016$ | $32.51 \pm 0.013$ |




| Data | Accuracy | AUC-ROC | AUC-PR |
|---|---|---|---|
| *DDI* | $85.93 \pm 0.042$ | $94.11 \pm 0.020$ | $90.83 \pm 0.015$ |
| *PPI* | $42.65 \pm 0.025$ | $87.70 \pm 0.063$ | $83.09 \pm 0.074$ |
| *NELL* | $85.76 \pm 0.058$ | $63.02 \pm 24.26$ | $30.32 \pm 0.092$ |
| *CiteSeer* | $72.96 \pm 0.014$ | $50.00 \pm 0.000$ | $27.04 \pm 0.089$ |

**GCLNs are also robust to missing data and to partially observed data even when altering the firing counts of most discriminative rules.**

## III. Out-of-Domain Detection

| | GCLN | | Discrete Flows | |
|---|---|---|---|---|
| Data | In-domain | OOD | In-domain | OOD |
| *DDI* | $-9.02 \pm 0.431$ | $-21.55 \pm 6.124$ | $-6.92 \pm 0.194$ | $-6.95 \pm 0.194$ |
| *PPI* | $-7.91 \pm 2.056$ | $-8.61 \pm 0.956$ | $-6.98 \pm 0.409$ | $-6.93 \pm 0.415$ |
| *NELL* | $-21.37 \pm 4.969$ | $-23.02 \pm 3.131$ | $-11.81 \pm 0.775$ | $-12.24 \pm 0.750$ |
| *CiteSeer* | $-5.17 \pm 0.122$ | $-124.42 \pm 23.824$ | $-4.26 \pm 0.200$ | $-4.36 \pm 0.200$ |

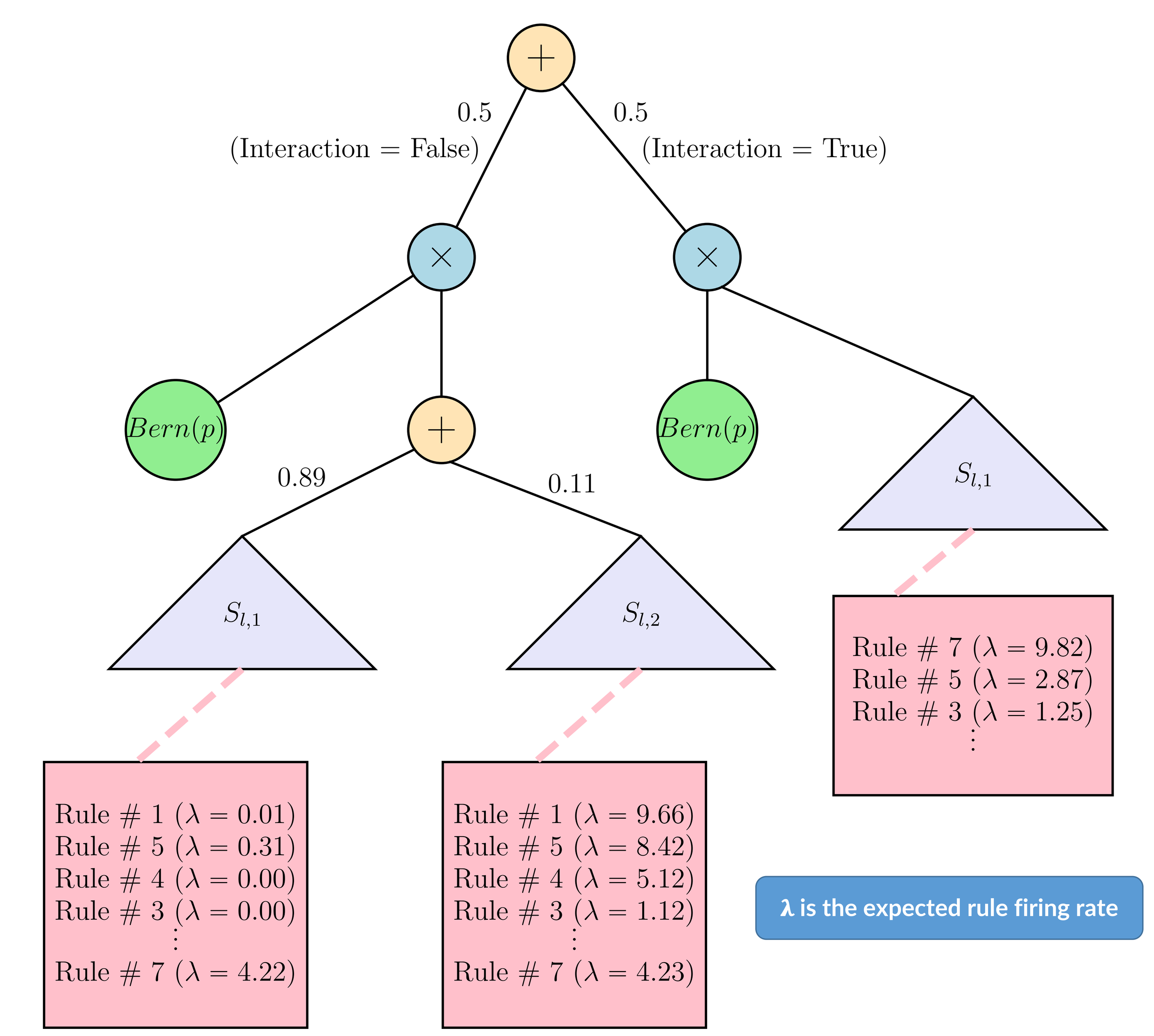
**GCLNs can perform out-of-domain detection outperforming SOTA neural density estimators like Discrete Flows [2].**

## IV. Missing Data Imputation

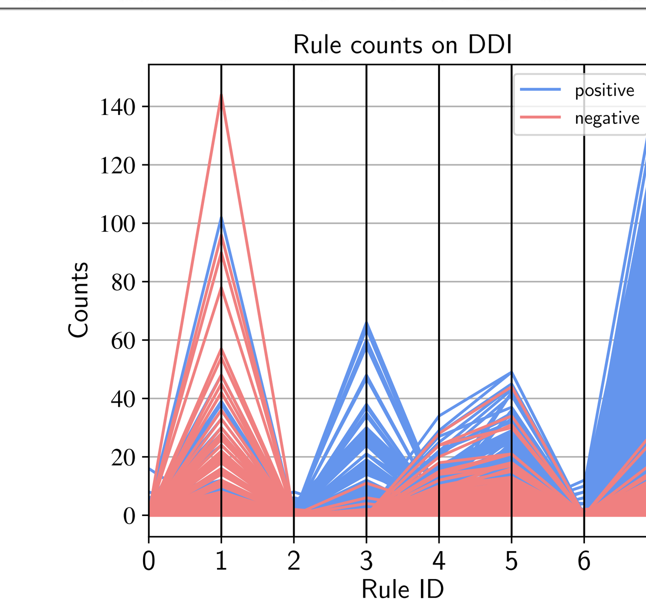
| Data | kNN | MICE | GCLN |
|---|---|---|---|
| *DDI* | $37.41 \pm 3.751$ | $37.25 \pm 3.451$ | $40.72 \pm 3.794$ |
| *PPI* | $2.40 \pm 3.451$ | $1.44 \pm 1.748$ | $3.22 \pm 4.320$ |
| *NELL* | $18.53 \pm 28.900$ | $10.80 \pm 16.644$ | $42.58 \pm 40.612$ |
| *CiteSeer* | $0.76 \pm 0.090$ | $0.65 \pm 0.011$ | $1.02 \pm 0.039$ |

**GCLNs can impute missing data being on-par with task-specific methods such as MICE [3].**

## V. Interpretable Models for XAI




**Rule #1**: Transporter(C, A) ∧ Transporter(C, B) $\implies$ Interacts(A,B)
**Rule #3**: Transporter(C, A) ∧ EnzymeInhibitor(A, D) ∧ EnzymeSubstrate(B, D) ∧ TransporterSubstrate(A, C) ∧ TransporterInducer(A, E) $\implies$ Interacts(A,B)
**Rule #7**: TargetAntagonist(B, C) ∧ EnzymeSubstrate(A, C) $\implies$ Interacts(A,B)




**GCLNs structure concepts hierarchically and can take the best out of the most discriminative rules.**

[1] Molina et al., Mixed sum-product networks: A deep architecture for hybrid domains (AAAI 2018)
[2] Tran et al., Discrete flows: Invertible generative models of discrete data (NeurIPS 2019)
[3] van Buuren et al., mice: Multivariate imputation by chained equations in r (J. Stat. Softw. 2011)