DOTGREEDX: EXPLAINING GNNS APPLIED TO CHEMOINFORMATICS

Mariana Brito Azevedo[1], Jean-Luc Lamotte[1], Luc Brun[1], Pierre Héroux[2]

[1]Université Caen Normandie, ENSICAEN, CNRS, Normandie Univ, GREYC UMR 6072, 14000 Caen, France

[2]Université Caen Normandie, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS, 76000 Rouen, France

mariana.brito-azevedo@unicaen.fr, jean-luc.lamotte@unicaen.fr, luc.brun@unicaen.fr, pierre.heroux@univ-rouen.fr

In collaboration with *Centre d'Études et de Recherche sur le Médicament de Normandie (CERMN)*

GREYC
Electronics and Computer Science Laboratory

1. Background and motivation

Context: development of GNNs in various chemoinformatics applications

Molecular property prediction

Drug design

Biological activity prediction

↓

Problem: GNNs often operate as black boxes, making their decision-making process difficult to understand

Why this prediction? **How** was it made? Can we **trust** it?

↓

Solution: development of XAI techniques for GNN explainability

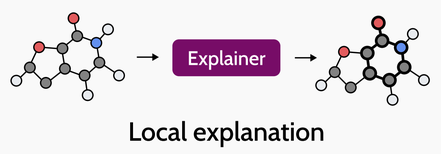
Transparency & Reliability

Error detection & Improvement

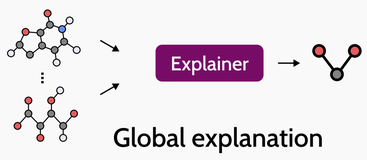
Decision Support & Pattern Discovery

2. Taxonomy of GNN Explanations[1]

Explanations by level: the scope of the graph information



Local explanation

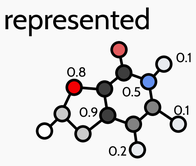


Global explanation

Explanations by modality: how the explanation is represented

Generation

Extraction



Scoring

Disadvantages of existing methods

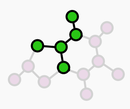
- Fidex-size explanation and need for hyperparameters
- Usually unclear how to identity important elements
- May not reflect a rankable importance score

→ DotGreedX

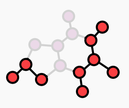
3. DotGreedX : combining scoring-based method and greedy search algorithm

Evaluation metrics[2]

$$Fidelity^{-} = f(\ \ ) - f(\ \ )$$

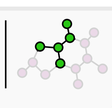


$$Fidelity^{+} = f(\ \ ) - f(\ \ )$$



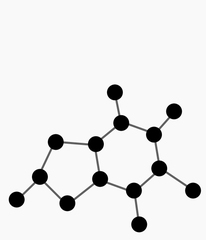
$$Characterization_{score} = \frac{2 \cdot Fidelity^{+} \cdot (1 - Fidelity^{-})}{Fidelity^{+} + (1 - Fidelity^{-})}$$

$$Sparsity = 1 - \frac{|\ \ |}{|\ \ |}$$



$$HSC_{score} = \frac{2 \cdot Characterization_{score} \cdot Sparsity}{Characterization_{score} + Sparsity}$$

Methodology[3]



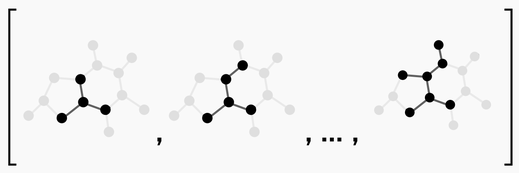
Original graph

→ 1 Model's classification → Correct classification? → YES → 2 DotGreedX scoring → Initial explanation subgraph

STOP ↑ NO

↓

3 Optimal size search → 4 Maximize HSC-Score → Final explanation subgraph



List of candidate explanation subgraphs decreasing size

4. Experimental results

Dataset	Explainer	Exp_{acc}	$Charac_{score}$	$Sparsity$	HSC_{score}	$Time(s)$
Benzene	SA[4]	0.660±0.223	0.748±0.178	0.594±0.032	0.649±0.066	1.96±0.07
	IG[5]	0.944±0.064	0.966±0.035	0.652±0.009	0.778±0.008	49.35±0.17
	CAM[6]	0.949±0.042	0.972±0.022	0.651±0.004	0.780±0.009	**1.58±0.01**
	GNNExp.[7]	0.398±0.371	0.470±0.350	0.193±0.038	0.238±0.082	148.47±0.59
	EiG-Search[8]	0.638±0.099	0.793±0.040	0.577±0.010	0.668±0.019	34.74±0.68
	DotGreedX	**0.999±0.001**	**0.998±0.003**	**0.739±0.036**	**0.848±0.023**	88.69±10.87
Mutage-nicity	SA	0.517±0.171	0.570±0.073	0.573±0.025	0.569±0.039	0.65±0.06
	IG	0.379±0.196	0.528±0.102	0.605±0.048	0.559±0.067	20.55±1.16
	CAM	0.555±0.162	0.614±0.076	0.664±0.016	0.636±0.047	**0.64±0.04**
	GNNExp.	0.481±0.296	0.529±0.143	0.490±0.039	0.498±0.071	62.35±2.25
	EiG-Search	0.863±0.141	0.712±0.037	**0.677±0.017**	0.694±0.021	10.58±0.68
	DotGreedX	**0.928±0.068**	**0.794±0.044**	0.651±0.077	**0.714±0.060**	85.28±21.60

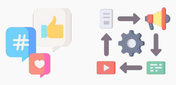
5. Perspectives

Optimizing the runtime of the greedy search algorithm

Extending DotGreedX to other tasks

- Node classification
- Multi-class classification
- Regression tasks

Extending DotGreedX to other types of graphs



GNN Global Explanations

A coherent vision of the model's behavior

Improved results analysis

Pattern recognition and definition of general rules

1. M. Bugueño; R. Biswas; G. Melo. Graph-Based Explainable AI : A Comprehensive Survey (2024).
2. Y. Li et al. A Survey of Explainable Graph Neural Networks : Taxonomy and Evaluation Metrics (2023).
3. M. B. Azevedo et al. DotGreedX : Combining Scoring-Based Technique and Greedy Search for GNN Explainability (2026).
4. F. Baldassarre; H. Azizpour. Explainability Techniques for Graph Convolutional Networks (2019).
5. B. Sanchez-Lengeling et al. Evaluating Attribution for Graph Neural Networks (2020).
6. P. E. Pope et al. Explainability Methods for Graph Convolutional Neural Networks (2019).
7. Z. Ying et al. GNNExplainer : Generating Explanations for Graph Neural Networks (2019).
8. S. Lu et al. EiG-Search : Generating Edge-Induced Subgraphs for GNN Explanation in Linear Time (2024).

8th Symposium MaDICS, June 2026, Avignon, France.