

# Proposition de sujet de Master

## Explicabilité des décisions d'un GNN, application à la chémoinformatique

Novembre 2023

### 1 Introduction

Le GREYC mène une collaboration soutenue avec le laboratoire CERMN autour de la fouille de données appliquée aux molécules thérapeutiques.

Le GREYC entretient par ailleurs une collaboration avec le LITIS sur la thématique de l'apprentissage profond sur graphes. Cette collaboration s'est ces dernières années concrétisée à travers le projet AGAC (2018-2019) soutenu par la Région Normandie. Le travail commun et complémentaire des deux laboratoires sur la thématique de l'apprentissage profond sur graphes a également été reconnu par la sélection et le financement par l'ANR du projet CoDeGNN (2022-2025).

La chémoinformatique offre également pour le LITIS un cadre applicatif support de problématiques scientifiques qui ont émané des collaborations qu'il entretient notamment avec le COBRA. Ainsi, les méthodes développées dans le cadre du présent projet pourront également trouver des applications, par exemple, dans l'identification des caractéristiques moléculaires que doivent observer les polymères (à synthétiser) devant disposer de certaines propriétés physico-chimiques.

Les équipes CODAG et Image du GREYC ainsi que l'équipe Apprentissage du LITIS ont décidé de combiner leurs compétences pour un projet sur l'explicabilité des réseaux convolutionnels sur graphes (GCN) en faisant une demande de thèse auprès de l'université de Caen ou de la région. Ce projet implique les personnes suivantes :

- GREYC : Jean-Luc Lamotte, Luc Brun, Bertrand Cuissart
- LITIS : Pierre Héroux, Paul Honeine
- CERMN : Ronan Bureau, Alban Lepailleur

Ce projet nécessite plusieurs compétences. Les jeux de données sur lesquels nous comptons travailler étant des jeux d'interaction ligand/protéines, une expertise forte en chémoinformatique est nécessaire. De plus, permettre à des GCN d'expliquer leurs résultats nécessite une compréhension fine de leurs comportements. Nous souhaitons donc initier ce projet au travers d'un stage de M2 qui sera co-encadré par :

- Jean-Luc Lamotte qui apporte à ce projet ses compétences en fouille de données et en chémoinformatique,
- Luc Brun du GREYC qui apportera ses compétences en GNN et plus particulièrement en pooling sur graphes,

- Pierre Héroux du LITIS participera également à l’encadrement où il apportera ses compétences en GNN et en méthodes de convolution.

L’analyse des premiers résultats réclamera une interaction forte avec Ronan Bureau et Alban Lepailleur du CERMN. Notons que ce projet permettra à Jean-Luc Lamotte nommé en septembre 2022 au GREYC d’approfondir ses contacts avec l’équipe Apprentissage du LITIS.

Cette collaboration GREYC/LITIS, permettra également de renforcer l’action transverse Graphes de NormaSTIC et plus généralement les collaborations promues par la fédération.

## 2 Sujet du stage

### 2.1 Contexte scientifique

L’apprentissage profond a révolutionné de nombreux domaines tels que l’analyse d’images ou le traitement automatique des langues. Pourtant, l’incapacité de ces systèmes à justifier leurs décisions est très vite apparue comme une limite forte. Des méthodes permettant de mesurer la saillance de chaque donnée d’entrée ont donc été définies. Ces méthodes peuvent se décomposer en deux grandes familles : Les méthodes dites ”boîte blanche” qui supposent le réseau connu et se basent donc sur toutes les informations de celui-ci : topologie, poids appris et gradient calculé par une rétropropagation. Inversement, les méthodes dites ”boîte noire”, considèrent le réseau comme une fonction inconnue et mesurent l’importance de chaque élément en masquant aléatoirement des parties du signal d’entrée. L’importance de chaque élément et alors définie comme la moyenne des réponses du système sur l’ensemble des masques sur lesquels l’élément n’est pas masqué.

La chimoinformatique, domaine relatif au traitement de l’information chimique, a été fortement impactée par l’arrivée de l’apprentissage profond et des réseaux convolutionnels (CNN). Un Graph Neural Network (GNN) est l’équivalent d’un CNN opérant sur des graphes. La convolution y est remplacée par une convolution sur graphe et des opérations de pooling peuvent également être définies. Une différence notable entre CNN et GNN se situe au niveau du pooling. Dans les CNN, le pooling opère sur des topologies fixes et utilise des fonctions de réduction généralement non apprises comme max pooling ou average pooling. Dans les GNN, le pooling apprend à la fois des regroupements de sommets de topologie arbitraire et des fonctions de réduction sur ces regroupements.

L’action thérapeutique d’une molécule, dépend de la capacité de celle-ci à se fixer sur les protéines ciblées. Cette capacité de la molécule est déterminée par ses conformations les plus probables ainsi que par la présence de différents groupes fonctionnels, appelés aussi caractéristiques pharmacophoriques, qui vont permettre à la molécule de se fixer. On peut ainsi distinguer deux types de représentations de molécules pour ce type de prédiction : Le graphe squelettique (sommets et arêtes représentent respectivement les atomes et leurs liaisons covalentes) et le graphe des caractéristiques pharmacophoriques où chaque sommet code un groupe fonctionnel de la molécule et la distance (topologique ou topographique) entre deux caractéristiques est codée par une arête. Dans ce cas, le graphe est un un graphe complet avec beaucoup moins de sommets que dans la représentation squelettique [11, 5].

L’explicabilité des réseaux est fondamentale dans la prédiction des propriétés thérapeutiques de molécules. En effet, celle-ci permet :

1. de valider les prédictions avant d’engager des efforts et des moyens sur les synthèses d’une série moléculaire

2. d'obtenir une intuition sur les propriétés physico-chimiques clés que doit posséder une molécule pour avoir une action biologique ciblée.

Les méthodes expliquant l'action des réseaux dans le cadre des graphes se situent dans la prolongation des méthodes conçues pour les CNN. On peut notamment appliquer la distinction boîte noire [15, 19, 13, 13, 6]/boîte blanche [14, 2, 4] mais également subdiviser les approches en méthodes locales (telles que les méthodes précédemment citées), qui fournissent une explication de la prédiction de chaque entrée et les méthodes globales [18, 10, 16] qui tentent de mettre en avant une explication globale des motifs clés pour la prédiction. Notons que les deux approches ne sont pas orthogonales. Par exemple, [1] agglomère les explications locales sur un jeu de données pour fournir une explication globale. D'autres critères de décomposition sont également possibles. On peut par exemple décomposer les approches en méthodes factuelles, par exemple [2], qui cherchent des motifs (noeuds, arêtes, sous-graphes) expliquant au mieux la prédiction et des modèles contre-factuels [9, 3] cherchant au contraire des modifications minimales des données d'entrées qui changent la prédiction. Parmi les méthodes factuelles, on peut également distinguer les méthodes qui intègrent l'extraction de motif à la prédiction. Ces méthodes sont donc auto-interprétables [12, 17] et les méthodes qui combinent le modèle de prédiction à une autre méthode pour générer des explications (par ex. [2]). Le lecteur intéressé peut se référer au surveys suivants pour d'avantage d'information [8, 7].

Notons toutefois que les méthodes type boîte blanche se sont focalisées sur des réseaux de type GNN sans pooling. L'utilisation du pooling pour l'explicabilité n'a à notre connaissance pas encore été explorée.

### 3 Déroulé du stage

Notre étude commencera par une étude des méthodes GNN permettant de prédire les propriétés de nos jeu de données. L'étude sera ciblée sur la prédiction des interactions protéines/ligands à partir des graphes complets de pharmacophores. A cette occasion, des méthodes basées GCN, au sens large, et GCN+pooling seront étudiées.

Si cette étape est validée, nous aborderons une comparaison des méthodes de la littérature permettant d'expliquer les résultats de ces GNNs. Les résultats produits (en termes d'explication) seront évalués en utilisant plusieurs critères [7] tels que l'accuracy, l'aire sous la courbe, la fidélité, la parcimonie, . . . Nous espérons identifier à partir de cette étude des sous structures pharmacophoriques pertinentes pour les propriétés à prédire.

Nous essaierons, dans un troisième temps, d'appliquer ces méthodes sur les graphes moléculaires squelettiques. Il s'agira de comparer sous l'angle de l'explicabilité les descriptions moléculaires "brutes" et celles produites en intégrant une expertise du domaine via le graphe pharmacophorique.

#### 3.1 Profil du candidat

Le candidat doit être inscrit en dernière année d'un Master ou d'un diplôme d'ingénieur dans un domaine lié à l'informatique ou aux mathématiques appliquées, et posséder de solides compétences en programmation. Une expérience en informatique pour la Science des Données, apprentissage profond, notamment sur graphes, sera un plus.

## 3.2 Conditions du stage

L'appel à candidature sera largement communiqué et le stage sera effectué au GR-EYC ou au LITIS en fonction du lieu de recrutement de l'étudiant. Il débutera en février ou mars 2024 pour une durée de 6 mois et bénéficiera d'une gratification au tarif minimum réglementaire pour les stages.

## 4 Demande financière

- Indemnité de stage : 3600€
- Nous comptons également solliciter NormaSTIC pour des déplacements Caen/Rouen. Ces frais ne devraient pas excéder 1000 €

## References

- [1] Steve Azzolin, Antonio Longa, Pietro Barbiero, Pietro Liò, and Andrea Passerini. Global explainability of gns via logic combination of learned concepts. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023.
- [2] Federico Baldassarre and Hossein Azizpour. Explainability techniques for graph convolutional networks. *CoRR*, abs/1905.13686, 2019.
- [3] Ziheng Chen, Fabrizio Silvestri, Jia Wang, Yongfeng Zhang, Zhenhua Huang, Hongshik Ahn, and Gabriele Tolomei. GREASE: generate factual and counterfactual explanations for gnn-based recommendations. *CoRR*, abs/2208.04222, 2022.
- [4] Qizhang Feng, Ninghao Liu, Fan Yang, Ruixiang Tang, Mengnan Du, and Xia Hu. DEGREE: decomposition based explanation for graph neural networks. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.
- [5] Damien Geslin, Alban Lepailleur, Jean-Luc Manguin, Nhat-Vinh Vo, Jean Luc Lamotte, Bertrand Cuissart, and Ronan Bureau. Deciphering a pharmacophore network: A case study using BCR-ABL data. *J. Chem. Inf. Model.*, 62(3):678–691, 2022.
- [6] Qiang Huang, Makoto Yamada, Yuan Tian, Dinesh Singh, and Yi Chang. Graphlime: Local interpretable model explanations for graph neural networks. *IEEE Trans. Knowl. Data Eng.*, 35(7):6968–6972, 2023.
- [7] Jaykumar Kakkad, Jaspal Jannu, Kartik Sharma, Charu C. Aggarwal, and Sourav Medya. A survey on explainability of graph neural networks. *CoRR*, abs/2306.01958, 2023.
- [8] Yiqiao Li, Jianlong Zhou, Sunny Verma, and Fang Chen. A survey of explainable graph neural networks: Taxonomy and evaluation metrics. *CoRR*, abs/2207.12599, 2022.
- [9] Ana Lucic, Maartje A. ter Hoeve, Gabriele Tolomei, Maarten de Rijke, and Fabrizio Silvestri. Cf-gnnexplainer: Counterfactual explanations for graph neural networks. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *International Conference on Artificial Intelligence and Statistics, AISTATS 2022, 28-30 March 2022, Virtual Event*, volume 151 of *Proceedings of Machine Learning Research*, pages 4499–4511. PMLR, 2022.

- [10] Lucie Charlotte Magister, Dmitry Kazhdan, Vikash Singh, and Pietro Liò. Gcexplainer: Human-in-the-loop concept-based explanations for graph neural networks. *CoRR*, abs/2107.11889, 2021.
- [11] Jean-Philippe Métivier, Bertrand Cuissart, Ronan Bureau, and Alban Lepailler. The pharmacophore network: A computational method for exploring structure–activity relationships from a large chemical data set. *Journal of Medicinal Chemistry*, 61(8):3551–3564, 04 2018.
- [12] Siqi Miao, Mia Liu, and Pan Li. Interpretable and generalizable graph learning via stochastic attention mechanism. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato, editors, *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 15524–15543. PMLR, 2022.
- [13] Tamara A. Pereira, Erik Nascimento, Lucas E. Resck, Diego Mesquita, and Amauri A. Souza. Distill n’ explain: explaining graph neural networks using simple surrogates. In Francisco J. R. Ruiz, Jennifer G. Dy, and Jan-Willem van de Meent, editors, *International Conference on Artificial Intelligence and Statistics, 25-27 April 2023, Palau de Congressos, Valencia, Spain*, volume 206 of *Proceedings of Machine Learning Research*, pages 6199–6214. PMLR, 2023.
- [14] Phillip E. Pope, Soheil Kolouri, Mohammad Rostami, Charles E. Martin, and Heiko Hoffmann. Explainability methods for graph convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 10772–10781. Computer Vision Foundation / IEEE, 2019.
- [15] Minh N. Vu and My T. Thai. Pgm-explainer: Probabilistic graphical model explanations for graph neural networks. In Hugo Larochelle, Marc’Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.
- [16] Han Xuanyuan, Pietro Barbiero, Dobrik Georgiev, Lucie Charlotte Magister, and Pietro Liò. Global concept-based interpretability for graph neural networks via neuron analysis. In Brian Williams, Yiling Chen, and Jennifer Neville, editors, *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 10675–10683. AAAI Press, 2023.
- [17] Junchi Yu, Tingyang Xu, Yu Rong, Yatao Bian, Junzhou Huang, and Ran He. Graph information bottleneck for subgraph recognition. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.
- [18] Hao Yuan, Jiliang Tang, Xia Hu, and Shuiwang Ji. XGNN: towards model-level explanations of graph neural networks. In Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash, editors, *KDD ’20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pages 430–438. ACM, 2020.

- [19] Yue Zhang, David DeFazio, and Arti Ramesh. Relex: A model-agnostic relational model explainer. In Marion Fourcade, Benjamin Kuipers, Seth Lazar, and Deirdre K. Mulligan, editors, *AIES '21: AAAI/ACM Conference on AI, Ethics, and Society, Virtual Event, USA, May 19-21, 2021*, pages 1042–1049. ACM, 2021.