Réseaux de neurones sur graphes pour l'analyse de signaux EEG

Maria SARKIS Mira RIZKALLAH Saïd MOUSSAOUI

Nantes Université, École Centrale Nantes, LS2N, CNRS UMR 6004, F-44300 Nantes, France

Résumé – Les signaux électro-encéphalographiques (EEG) suscitent un intérêt croissant car il permettent d'accéder à des informations précises sur l'activité cérébrale. Tirant profit de ces informations recueillies, des méthodes d'apprentissage statistique ont été développées durant les dernières années afin de classifier divers états cognitifs. Entre autres, les réseaux de neurones sur graphes apparaissent comme des outils prometteurs, intégrant à la fois la structure sous-jacente entre les nœuds (*i.e* electrodes) et les descripteurs du signal associé à chaque nœud. Cependant, dans ses representations, les position des électrodes reste encore très peu exploitée. Ce travail propose une approche de réseaux de neurones sur graphe, pour la classification à partir des signaux EEG, intégrant la position des électrodes, des mesures d'énergies des signaux et leur degré de connectivité. La classification est effectuée en considérant différentes structures de graphes ou combinaisons de descripteurs. Les résultats obtenus montrent que cette approche surpasse la méthode de référence basée sur du filtrage spatial (Common Spatial Pattern).

Abstract – The analysis of electroencephalography signals using graph theory has gained increasing interest in the recent years, particularly for inference based on these data. Graph Neural Networks have emerged as promising tools, incorporating both the functional multivariate EEG statistics and monovariate signal-related features. In the context of brain connectivity, the precise localization of brain electrical activity plays a crucial role in assessing mental states, yet electrode positions remain largely under-explored. In this paper, we propose a classification pipeline from EEG, based on graph neural networks, integrating electrode positions along with connectivity and energy measures derived from the signals. Results on a public EEG dataset show that our method outperforms conventional approaches based on spatial filtering.

1 Introduction

L'analyse de l'activité cérébrale à partir de signaux électroencéphalographiques (EEG) s'appuie traditionnellement sur des approches incluant des étapes de prétraitement, d'extraction de caractéristiques et de classification. Deux approches bien établies dans ce domaine sont la méthode de filtrage spatial, *Common Spatial Patterns* (CSP), et le classifieur à Distance Minimale à la Moyenne Riemannienne (MDM) [1]. La méthode CSP permet d'extraire des filtres spatiaux optimaux en maximisant la séparabilité entre classes dans des problèmes de classification supervisée. Initialement conçue pour des tâches binaires, elle a été étendue au cadre multi-classe à l'aide de stratégies telles que un-contre-tous(UCT) [2]. De son côté, MDM repose sur une représentation géométrique des signaux EEG via leurs matrices de covariance, classifiées dans un espace riemannien.

Plus récemment, la modélisation des EEG dans un cadre de traitement des signaux sur graphes s'est imposée comme une approche puissante pour analyser la connectivité cérébrale [3]. Ce paradigme permet de représenter les signaux EEG sous forme de graphes, où les nœuds correspondent aux électrodes et les arêtes capturent leurs relations fonctionnelles ou structurelles. La structure sous-jacente du graphe, c'est-à-dire les poids des arêtes, peut être définie à partir de la distance physiques entre les électrodes [4], ou dérivée des signaux EEG acquis. Dans ce cas, il est possible de définir manuellement des mesures de connectivité fonctionnelle en se basant sur les similarités entre les signaux. [5].
Au-delà des informations structurelles au niveau des arêtes, des caractéristiques supplémentaires peuvent être intégrées au niveau des noeuds. En effet, les caractéristiques des nœuds représentent des attributs ou propriétés spécifiques à chaque électrode. Par example, l'extraction de caractéristiques fréquentielles à l'aide des densités spectrales de puissance sur différentes gammes de fréquences [5].

Les graphes ainsi construits, intégrant à la fois des attributs au niveau des arêtes et des nœuds, sont encodés en un vecteur unique à l'aide de réseaux de neurones profonds pour graphes [6] afin de réaliser la tâche de classification. Cependant, dans les approches de modélisation en graphe du signal EEG, les positions des nœuds (électrodes) sont souvent ignorées, ce qui engendre une perte de l'information issue de la localisation exacte des signaux étudiés. Par exemple, dans le contexte de la conduite automobile et aérienne [7], l'étude des signaux physiologiques montre qu'une augmentation de la charge mentale se manifeste par des modifications de l'activité EEG dans certaines bandes de fréquences et régions cérébrales spécifiques.

Dans ce travail, nous proposons d'intégrer l'information de localisation des électrodes dans l'apprentissage sur graphe à partir de données EEG. L'originalité réside dans la représentation des signaux EEG sous forme d'un graphe intégrant à la fois les informations fréquentielles (des signaux mésurés) et spatiales (positions des électrodes). Ceci, couplé à un réseau de neurones sur graphe, conduit à une meilleure classification de trois niveaux de charge mentale, comparé aux méthodes de l'état de l'art sur une base de donnée publique [1].

2 Approche proposée

Considérons un ensemble de signaux EEG acquis par un réseau de N électrodes avec des positions spatiales connues $\mathcal{P}$. Nous souhaitons prédire à partir de chaque fenêtre de t secondes de ces signaux, une classe. Les étapes de traitement que nous proposons sont illustrées dans la Fig.1. Un graphe pondéré

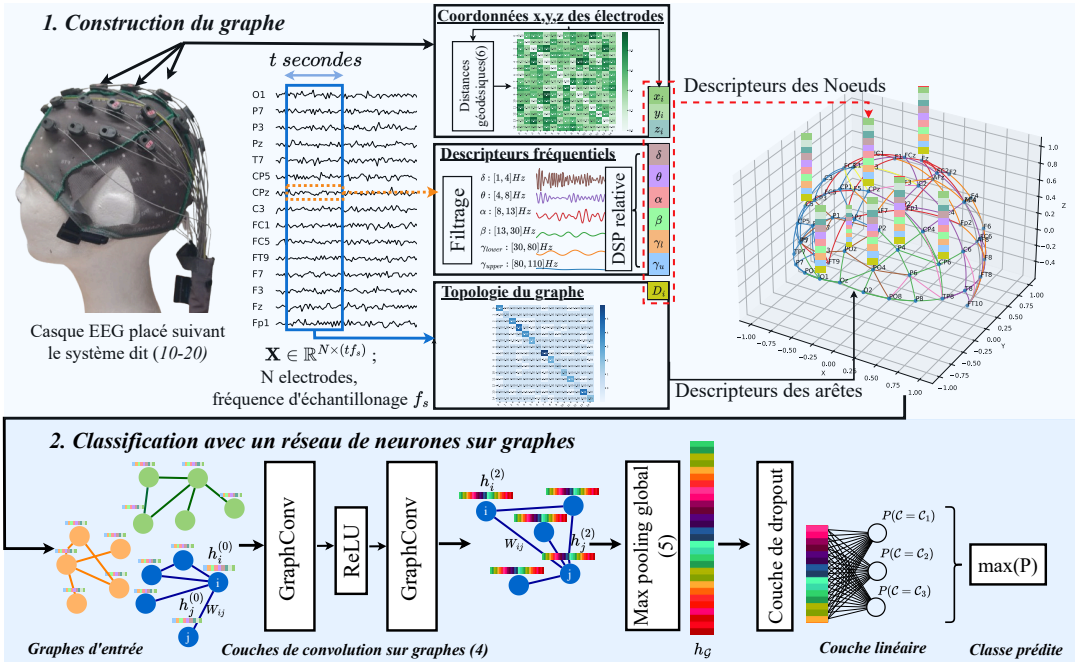

FIGURE 1 : Cadre de classification proposé : à partir de chaque fenêtre EEG, un graphe de connectivité est appris. Des caractéristiques fréquentielles, éventuellement combinées aux positions et à un descripteur topologique, définissent les attributs des nœuds. Les arêtes reflètent la connectivité ou l'inverse de la distance géodésique. Les graphes sont ensuite traités par un GNN pour la classification.

non orienté $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$, associé à ce réseau est défini par un ensemble de nœuds $\mathcal{V}$ reliés par des arêtes $\mathcal{E}$. Les nœuds correspondent aux électrodes et le poids W_{ij} défini sur l'arrête entre deux nœuds, i et j, est calculé à partir des signaux EEG. Des caractéristiques spécifiques aux noeuds sont extraites pour enrichir la représentation du graphe. Le graphe construit est traité par un réseau de neurones sur graphes ayant comme sortie la classe prédite. Les détails de ces différentes étapes sont présentés ci-dessous.

2.1 Construction du graphe

La matrice des poids $\mathbf{W}$ est construite soit manuellement à partir de mesures de connectivité (coefficient de corrélation ou *Phase Locking Value*), estimant directement le poids associé à chaque arête reliant deux électrodes i et j, soit apprise des données multivariées en estimant le laplacien généralisé (3) sous contraintes de parcimonie [8].

Coefficient de corrélation de Pearson : cette pondération quantifie la corrélation entre deux signaux de deux électrodes et est calculée comme suit :

$$W_{ij} = \frac{S_{ij}}{\sqrt{S_{ii} S_{jj}}} \tag{1}$$

où $S_{i,j}$ est la covariance entre les signaux EEG mesurés par les électrodes i et j. Afin de favoriser la parcimonie, la matrice de corrélation est seuillée en utilisant comme seuil la médiane de l'ensemble des valeurs observées sur toutes les fenêtres temporelles.

***Phase Locking Value* (PLV) :** cette pondération quantifie la cohérence des différences de phase entre deux signaux. Elle est définie par :

$$W_{ij} = \left| \frac{1}{T} \sum_{t=1}^{T} e^{-\mathrm{i}(\phi_{\mathbf{x}_i(t)} - \phi_{\mathbf{x}_j(t)})} \right| \tag{2}$$

où T est le nombre d'échantillons, et $\mathbf{x}_i(t)$, $\mathbf{x}_j(t)$ sont les signaux EEG des électrodes i et j, respectivement. L'argument de phase ϕ est obtenu via la transformée de Hilbert [9].

Laplacien généralisé : la matrice de poids $\mathbf{W}$ est apprise à l'aide de la méthode décrite dans [8], que nous avons adaptée à notre contexte comme présenté dans [10]. Cette approche permet d'estimer la structure sous-jacente du graphe, modélisée par un laplacien $\mathbf{L}$, à partir des signaux EEG multivariés, sous des contraintes de parcimonie et de connectivité. Le laplacien reflète alors la probabilité conditionnelle d'un lien entre les électrodes i et j et permet d'exprimer la matrice de poids sous la forme :

$$\mathbf{W} = \mathrm{Diag}(\mathbf{L}) - \mathbf{L} \tag{3}$$

où $\mathrm{Diag}(\mathbf{L})$ désigne une matrice diagonale contenant les éléments diagonaux de $\mathbf{L}$, les autres éléments étant nuls.

Chaque nœud du graphe est décrit par un vecteur de descripteurs pouvant contenir différentes formes d'information. Dans ce travail, nous considérons trois types de descripteurs :

1. Descripteur topologique : afin de quantifier l'importance relative d'un nœud dans le graphe, nous utilisons comme mesure topologique son degré correspondant à la somme des poids des connexions qui le relient aux autres nœuds. Cette valeur, simple mais informative, reflète la connectivité structurelle de chaque électrode dans le réseau.

2. Descripteurs fréquentiels : la densité spectrale de puissance relative (DSP) est utilisée, en raison de son efficacité démontrée dans les tâches de classification

EEG [5, 7]. Pour chaque fenêtre temporelle, la DSP est estimée pour chaque électrode à l'aide de la méthode de Welch, dans les bandes de fréquences suivantes : δ [1–4 Hz], θ [4–8 Hz], α [8–13 Hz], β [13–30 Hz], γ_{bas} [30–80 Hz], et γ_{haut} [80–110 Hz]. Les puissances dans ces bandes sont ensuite normalisées par la puissance totale (1–110 Hz), ce qui permet d'exprimer la contribution relative de chaque bande fréquentielle.

3. Position des électrodes : au-delà de l'activité cérébrale elle-même, la localisation spatiale de cette activité constitue une source d'information complémentaire et pertinente. Ainsi, les coordonnées tridimensionnelles (x, y, z) de chaque électrode sont intégrées comme descripteurs de position au niveau des nœuds.

A l'issue de cette étape, $\mathbf{H}^{(0)} \in \mathbb{R}^{N \times k}$ représente la matrice des caractéristiques des N électrodes. Chaque ligne de $\mathbf{H}^{(0)}$, qu'on note $\mathbf{h}_i^{(0)}$, correspond au vecteur de caractéristiques intial d'un nœud, de dimension k.

2.2 Apprentissage sur graphe

Un réseau de neurones sur graphes (GNN) est une extension des réseaux de neurones artificiels, conçue pour traiter des données structurées sous forme de graphes. Dans ce travail, nous implémentons un GNN pour une tâche de classification de graphes issus de signaux EEG. L'entrée du modèle correspond aux caractéristiques du graphe $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$, construit comme décrit dans la section 2.1. Au niveau des nœuds, chaque couche de *GraphConv* [11] met à jour les représentations des nœuds en agrégeant l'information provenant de leurs voisins. Deux nœuds i et j sont considérés comme voisins si $W_{ij} > 0$. La mise à jour d'un nœud i à la couche l est définie par :

$$\mathbf{h}_i^{(l)} = \mathbf{\Theta}_1 \mathbf{h}_i^{(l-1)} + \mathbf{\Theta}_2 \max_{j \in \mathcal{N}_i} (W_{ij} \cdot \mathbf{h}_j^{(l-1)}) \tag{4}$$

où $\mathbf{h}_i^{(l)}$ est le vecteur de caractéristiques du nœud i à la couche l. $\mathcal{N}_i$ est le voisinage direct du nœud i. $\mathbf{\Theta}_1, \mathbf{\Theta}_2$ sont des transformations linéaires apprises, projetant les vecteurs de caractéristiques vers un espace caché de dimension d. Le paramètre d représente le nombre de couches cachées, et constitue un hyperparamètre de l'architecture. Empiler deux couches de convolutions de graphes permet à chaque nœud de capter l'information de ses voisins à deux sauts. Une fonction d'activation, dite ReLU (unité linéaire rectifiée), est appliquée après la première couche pour introduire de la non-linéarité et permettre l'apprentissage de motifs complexes. Pour obtenir une représentation globale du graphe, les plongements des nœuds sont agrégés via une couche de sortie. Le plongement du graphe est alors calculé en prenant le maximum, pour chaque dimension de caractéristiques, à travers tous les nœuds :

$$\mathbf{h}_{\mathcal{G}}(k) = \max_i \mathbf{h}_i(k), \quad k = 1, \ldots, d. \tag{5}$$

Une opération, dite de *dropout* (abandon), est ensuite appliquée aux plongement du graphe, avec une probabilité donnée, afin de désactiver aléatoirement certains éléments et améliorer la capacité de généralisation du modèle. Enfin, une couche linéaire entièrement connectée est utilisée pour effectuer la tâche de classification à partir du plongement du graphe.

3 Résultats expérimentaux

Notre approche est évaluée dans le cadre d'une tâche de classification intra-sujet et intra-session de la charge mentale (MWL) en utilisant le jeu de données EEG public [1] : "COG-BCI". Ce dernier comprend des enregistrements de 29 participants sur 3 sessions, au cours desquelles ils ont effectué quatre tâches cognitives Pour cette étude, nous testons notre approche sur les enregistrements MATB-II des sessions 1 et 2. Les participants ont réalisé trois tâches indépendantes de cinq minutes chacune, correspondant à trois niveaux de difficulté : faible, moyen et élevé. Les signaux EEG ont été enregistrés à l'aide de 64 électrodes Ag-AgCl suivant le système dit *10-20*, avec une fréquence d'échantillonnage de $f_s = 250$ Hz. Après prétraitement et suppression des artéfacts, comme décrit dans [1], 61 électrodes sont conservées.

Les hyperparamètres d'entrainement du modèle GNN sont fixés après une recherche en grille appliquée à un sujet du jeu de données et fixés pour tous les sujets ensuite. Les valeurs retenues sont les suivantes : un taux d'apprentissage de 10^{-3}, une taille de lot de 30, une probabilité de *dropout* de 0.5, et un nombre de canaux cachés fixé à 65. L'entraînement du modèle est régulé par un mécanisme d'arrêt anticipé après 2000 époques, avec une patience de 100 époques.

La performance du modèle pour chaque session de chaque sujet est évaluée en calculant la moyenne des précisions obtenues lors de 5 plis de validation croisée. Les résultats présentés dans cette étude correspondent à la moyenne des performances (en précision de détection) sur l'ensemble des sujets et de leurs sessions.

3.1 Impact de la structure du graphe

Pour évaluer l'intérêt de la construction du graphe via l'apprentissage du laplacien généralisé, nous utilisons la concaténation des trois descripteurs de nœuds présentés en 2.1, et les attributs sur les arêtes issus de la matrice de poids $\mathbf{W}$ apprise. Les performances sont ensuite comparées à celles obtenues avec des matrices $\mathbf{W}$ construites à partir de mesures manuelles (corrélation de Pearson et PLV). Les résultats reportés dans le Tableau 1 montrent que l'apprentissage d'un graphe parcimonieux à partir des signaux EEG permet d'atteindre les meilleures performances. Contrairement à la corrélation de Pearson et au PLV, qui ne capturent que des connexions par paires sans tenir compte des influences des électrodes voisines, la matrice de précision intègre efficacement ces dépendances. De plus, alors que la corrélation de Pearson et le PLV nécessitent un seuillage explicite pour induire la parcimonie, l'apprentissage de la matrice de précision tient naturellement compte de cette contrainte, sans besoin de définir un seuil.

TABLE 1 : Comparaison entre différentes stratégies de construction du graphes (moyenne $\pm$ écart type).

Structure du graphe	**Précision**
Corrélation de Pearson	0.76 ± 0.08
Phase Locking Value	0.73 ± 0.07
Laplacien généralisé	$\mathbf{0.84 \pm 0.10}$

3.2 Impact des descripteurs des noeuds

Nous fixons ensuite la méthode de calcul du graphe en utilisant le laplacien, et nous étudions l'effet d'un encodage des positions des électrodes sous forme de distances, non plus comme attributs des nœuds, mais comme attributs des arêtes. Pour cela, nous calculons les distances géodésiques d_{ij} entre électrodes. La matrice $\mathbf{R}$ encode alors la structure du graphe en intégrant ces distances, tout en respectant les connexions définies par $\mathbf{W}$, selon la règle suivante :

$$\mathbf{R}_{ij} = \begin{cases} \frac{1}{d_{ij}} & \text{si } W_{ij} > 0 \text{ et } i \neq j, \\ 0 & \text{sinon.} \end{cases} \tag{6}$$

Nous testons également différentes configurations du graphe en variant les descripteurs utilisés pour les nœuds et les arêtes.

Les résultats résumés dans le Tableau 2 montrent que l'encodage de la connectivité fonctionnelle en tant qu'attributs des arêtes apporte davantage d'information que l'encodage basé uniquement sur les distances entre electrodes. Par ailleurs, l'utilisation seule de la DSP ne contribue pas considérablement à la classification. En revanche, la topologie du graphe représentée par le degré, utilisé seul ou combiné aux coordonnées spatiales (x, y, z), mène à de meilleures performances, même lorsqu'il est comparé à des combinaisons intégrant des descripteurs supplémentaires comme la DSP. Enfin, l'ajout de la DSP à l'ensemble des descripteurs n'apporte pas de gain, voire dégrade légèrement les performances, ce qui suggère que, dans ce cas, la DSP ne constitue pas un descripteur pertinent pour la tâche de classification de charge mentale.

TABLE 2 : Comparaison des configurations d'attributs de nœuds et d'arêtes (moyenne ± écart type).

Noeuds	Arêtes	Précision
[DSP]	R	0.68 ± 0.07
[Degré]	R	0.75 ± 0.11
[DSP, degré]	R	0.72 ± 0.07
[DSP]	W	0.69 ± 0.08
[x, y, z, degré]	W	$\mathbf{0.88 \pm 0.09}$
[x, y, z, degré, DSP]	W	0.84 ± 0.10

3.3 Comparaison avec les approches classiques

Enfin, nous comparons l'aproche GNN proposée à deux méthodes de référence exploitant les caractéristiques EEG dans le domaine spatial : CSP-UCT [2] et R-MDM [1]. Dans le cas du CSP-UCT, les caractéristiques extraites sont classifiées à l'aide d'une analyse discriminante linéaire (LDA). Pour les deux approches, une validation croisée en 5 plis est appliquée.

TABLE 3 : Comparison avec les méthodes de l'état de l'art exploitant des descripteurs spatiaux (moyenne $\pm$ écart type).

Descripteurs	Classifieur	Précision
Covariance [1]	R-MDM	0.62 ± 0.08
CSP-UCT [2]	LDA	0.78 ± 0.14
[x, y, z, degré] ; W	GNN	$\mathbf{0.88 \pm 0.09}$
[x, y, z, degré, DSP] ; W	GNN	0.84 ± 0.10

Les résultats présentés dans le Tableau 3 mettent en évidence la supériorité de notre approche par rapport aux méthodes classiques basées sur l'analyse de descripteurs spatiaux. Cela souligne l'importance de modéliser la connectivité cérébrale sous forme de graphes, enrichie par des descripteurs de position et de fréquences, afin d'obtenir une représentation plus globale des signaux dans le cadre de la classification de la charge mentale.

4 Conclusion

Dans cette communication, nous avons proposé une approche de classification des signaux EEG à l'aide de representations basées graphes. Nous avons mis en évidence l'intérêt d'intégrer la position des électrodes dans les descripteurs des noeuds, en plus des caractéristiques fréquentielles et topologiques. Nous soulignons aussi l'importance d'exploiter la nature multivariée des signaux EEG, via un laplacien généralisé, pour créer les liens entre les noeuds. Cette representation couplée à des réseaux de neurones sur graphe, permet d'atteindre de meilleures performances de classification que les approches de référence se basant sur les caractéristiques spatiales des EEG.

Références

[1] M. F. Hinss, E. S. Jahanpour, B. Somon, L. Pluchon, F. Dehais, and R. N. Roy, "Open multi-session and multi-task eeg cognitive dataset for passive brain-computer interface applications," *Scientific Data*, vol. 10, no. 1, p. 85, 2023.

[2] W. Wu, X. Gao, and S. Gao, "One-versus-the-rest (ovr) algorithm : An extension of common spatial patterns (csp) algorithm to multi-class case," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pp. 2387–2390, IEEE, 2006.

[3] L. E. Ismail and W. Karwowski, "A graph theory-based modeling of functional brain connectivity based on eeg : a systematic review in the context of neuroergonomics," *Ieee Access*, vol. 8, pp. 155103–155135, 2020.

[4] P. Zhong, D. Wang, and C. Miao, "Eeg-based emotion recognition using regularized graph neural networks," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1290–1301, 2020.

[5] D. Klepl, F. He, M. Wu, D. J. Blackburn, and P. Sarrigiannis, "Eeg-based graph neural network classification of alzheimer's disease : An empirical evaluation of functional connectivity methods," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 2651–2660, 2022.

[6] D. Klepl, M. Wu, and F. He, "Graph neural network-based eeg classification : A survey," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

[7] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neuroscience & Biobehavioral Reviews*, vol. 44, pp. 58–75, 2014.

[8] H. E. Egilmez, E. Pavez, and A. Ortega, "Graph learning from data under laplacian and structural constraints," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 825–841, 2017.

[9] C. T. Briels, D. N. Schoonhoven, C. J. Stam, H. de Waal, P. Scheltens, and A. A. Gouw, "Reproducibility of eeg functional connectivity in alzheimer's disease," *Alzheimer's research & therapy*, vol. 12, pp. 1–14, 2020.

[10] M. Sarkis, M. Rizkallah, and S. Moussaoui, "Functional graph image representation applied to eeg-based mental workload classification," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1–4, IEEE, 2024.

[11] C. Morris, M. Ritzert, M. Fey, W. L. Hamilton, J. E. Lenssen, G. Rattan, and M. Grohe, "Weisfeiler and leman go neural : Higher-order graph neural networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 4602–4609, 2019.