



Décimation hiérarchique pour l'apprentissage sur graphes

Stevan Stanovic



Table des matières

- 1 Introduction et contexte
- 2 État de l'art
- 3 Impact du pooling sur l'over-smoothing et l'over-squashing
- 4 Méthodes proposées et ensembles indépendants maximaux
- 5 Résultats expérimentaux
- 6 Conclusion



Table des matières

- 1 Introduction et contexte
- 2 État de l'art
- 3 Impact du pooling sur l'over-smoothing et l'over-squashing
- 4 Méthodes proposées et ensembles indépendants maximaux
- 5 Résultats expérimentaux
- 6 Conclusion



Réseaux de neurones sur les images

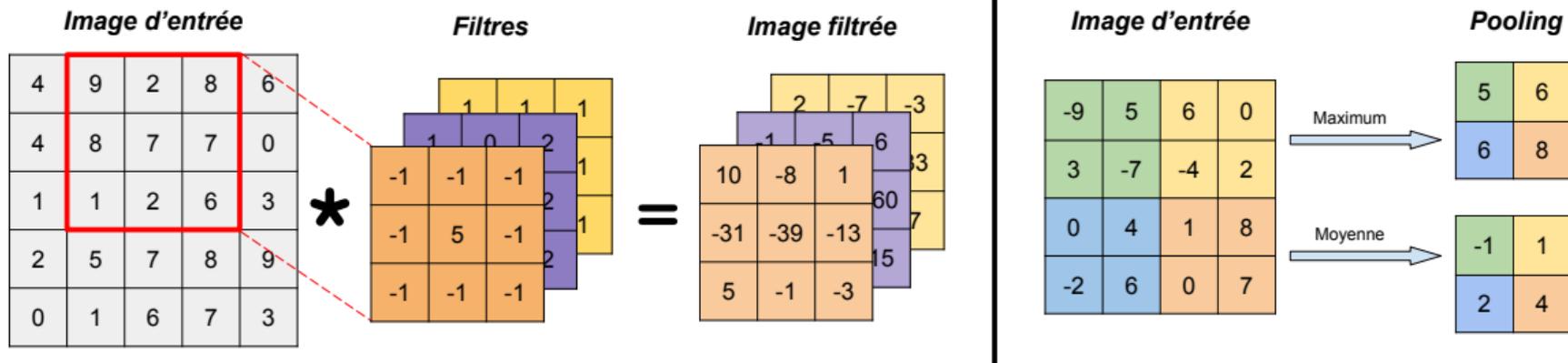
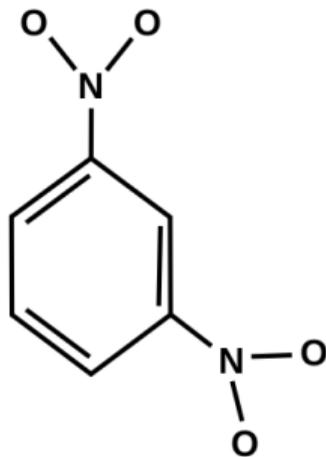
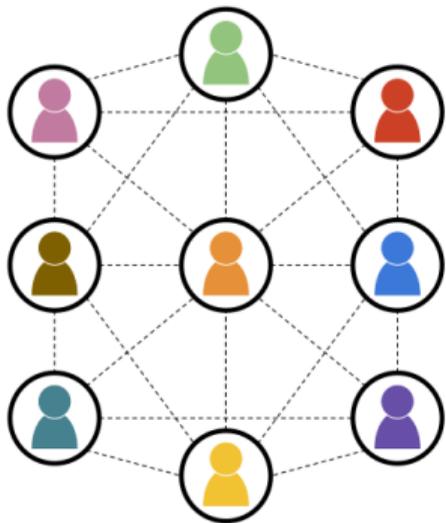


Figure – Schéma de convolution et pooling sur image

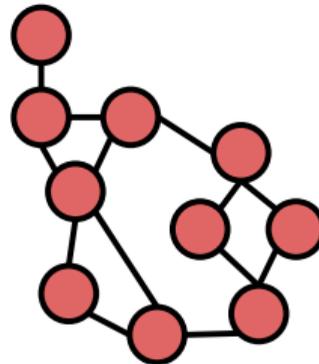
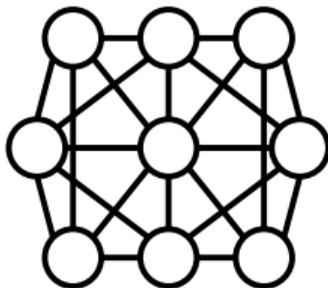
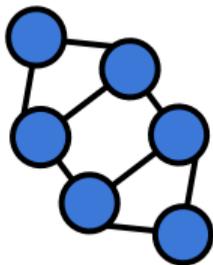
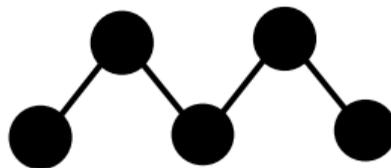
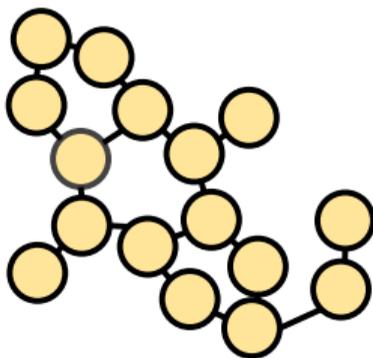
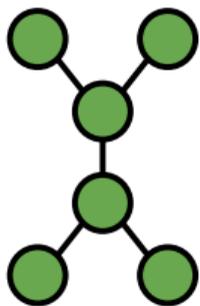


Champs d'application des graphes



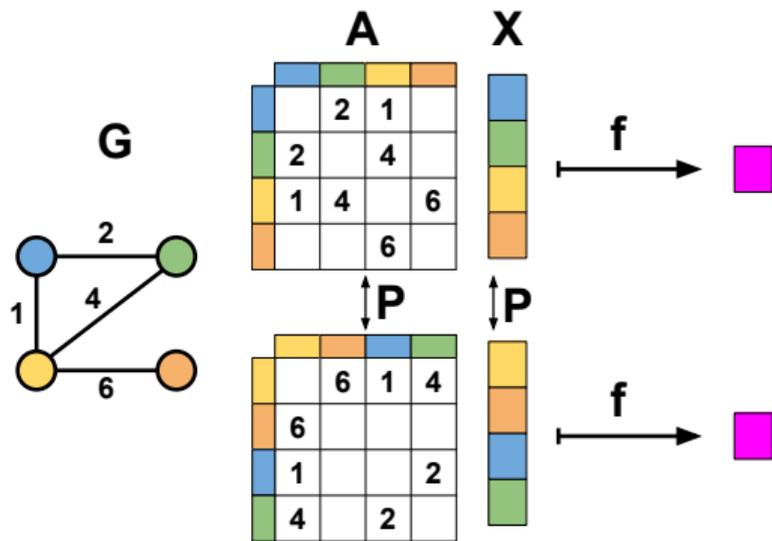


Invariance au nombre de sommets et au nombre de voisins

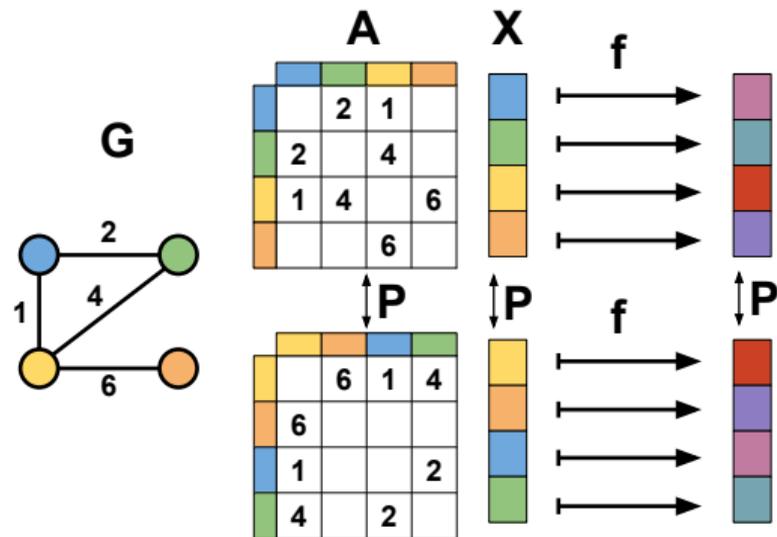




Invariance et équivariance aux permutations



(a) Invariance aux permutations



(b) Équivariance aux permutations

Convolution

Pooling



Contexte de la thèse dans l'ANR



Décimation et Pooling sur graphes

Stevan STANOVIC



Yann TIRARD-GATEL
Réseaux de neurones sur
graphe pour des données
spatio-temporelles



Jason PIQUENOT
Convolution sur graphes



Table des matières

- 1 Introduction et contexte
- 2 État de l'art**
- 3 Impact du pooling sur l'over-smoothing et l'over-squashing
- 4 Méthodes proposées et ensembles indépendants maximaux
- 5 Résultats expérimentaux
- 6 Conclusion



Réseaux de neurones sur graphes

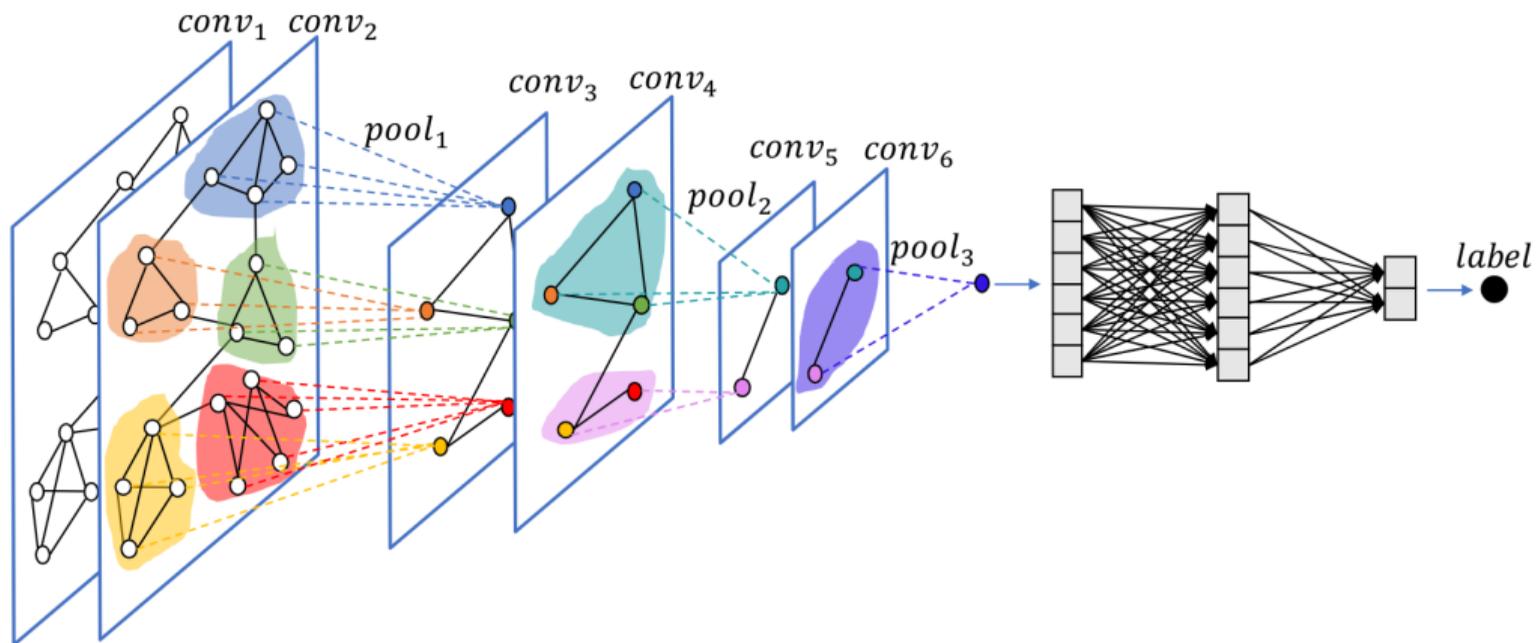


Figure – Prédiction d'une tâche sur l'entièreté d'un graphe (Y. Ma et al. 2019).



Convolution sur graphes

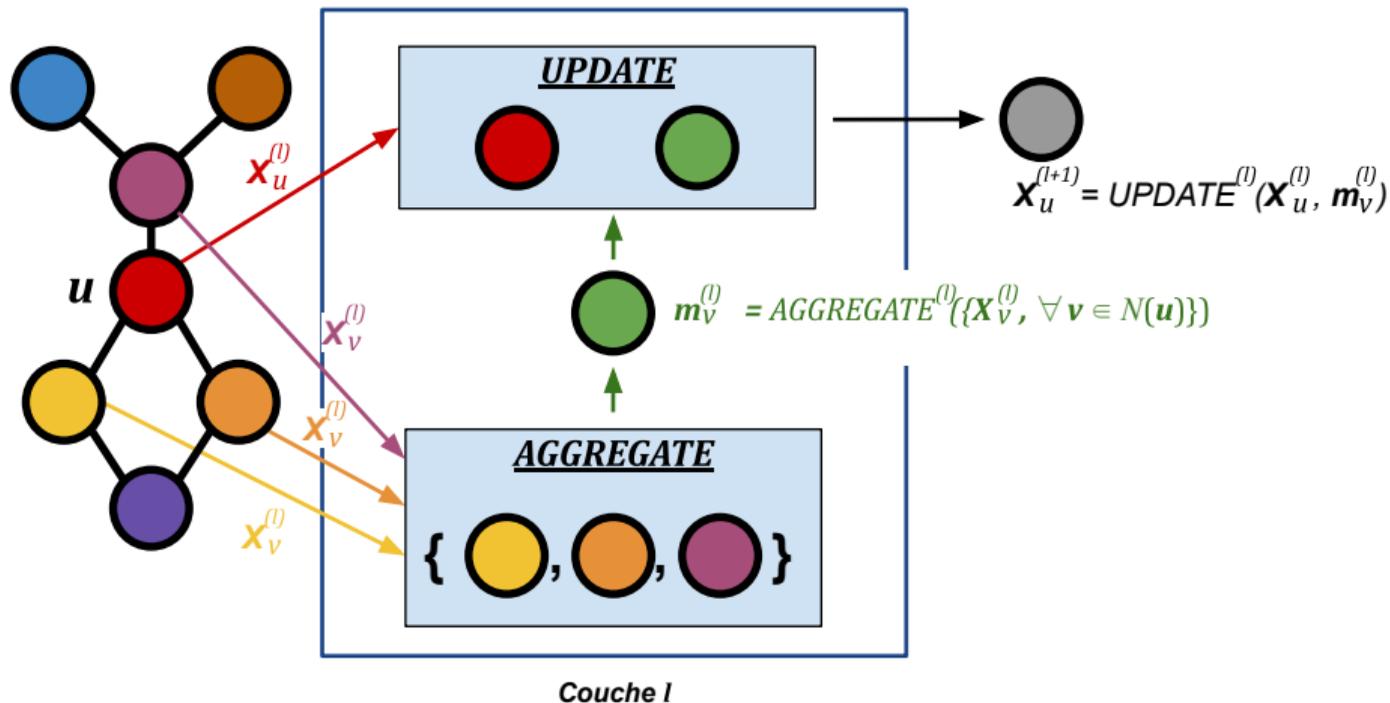


Figure – Illustration du Message Passing Neural Networks (MPNN) (Gilmer et al. 2017).



Convolution sur graphes

$$\mathbf{X}^{(l+1)} = \sigma(\mathbf{C}^{(l)} \mathbf{X}^{(l)} \mathbf{W}^{(l)})$$

- $\mathbf{C}^{(l)} \in \mathbb{R}^{n \times n}$ (support convolutionnel)
- $\mathbf{X}^{(l)} \in \mathbb{R}^{n \times f_l}$ (matrice des attributs des sommets)
- $\mathbf{W}^{(l)} \in \mathbb{R}^{f_l \times f_{l+1}}$ (matrice de poids apprise)



L'opération de pooling sur graphes peut être principalement décrite par la combinaison de trois fonctions :

Sélection

Réduction

Connexion



Pooling sur graphes

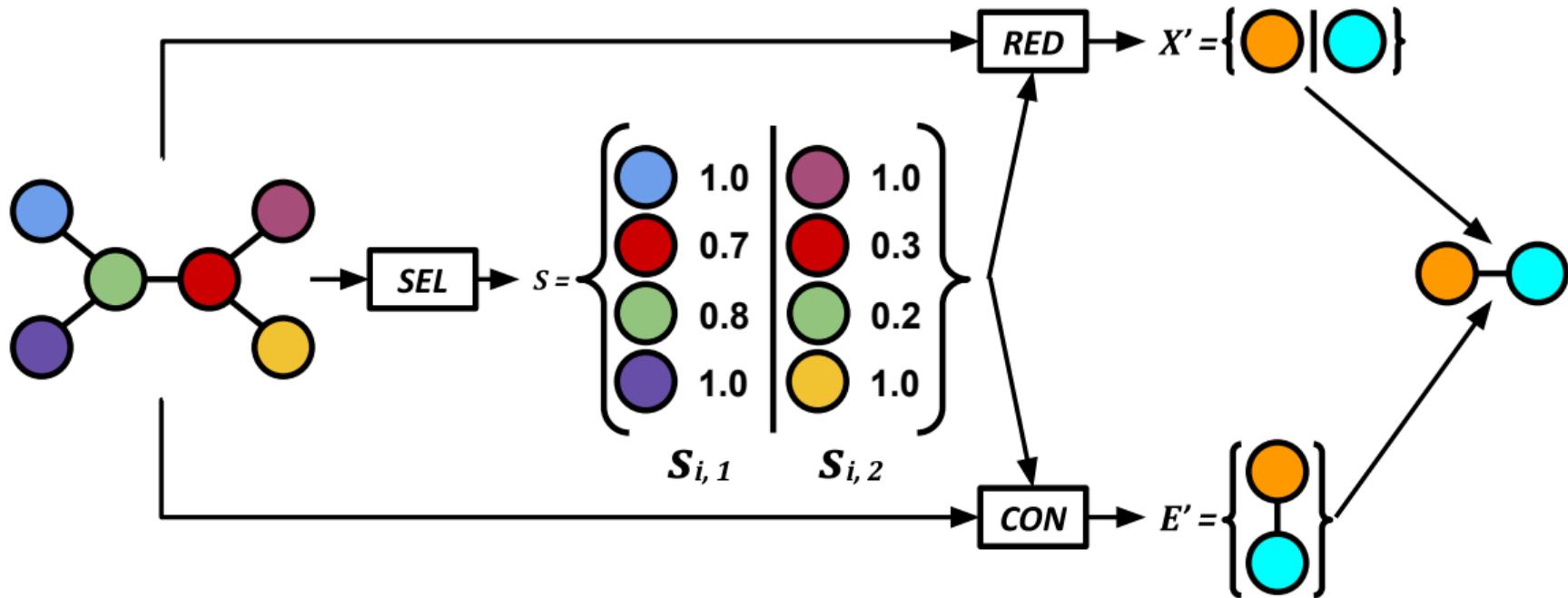


Figure – Schéma d'un pooling avec les fonctions SRC (Grattarola et al. 2024).



Pooling sur graphes

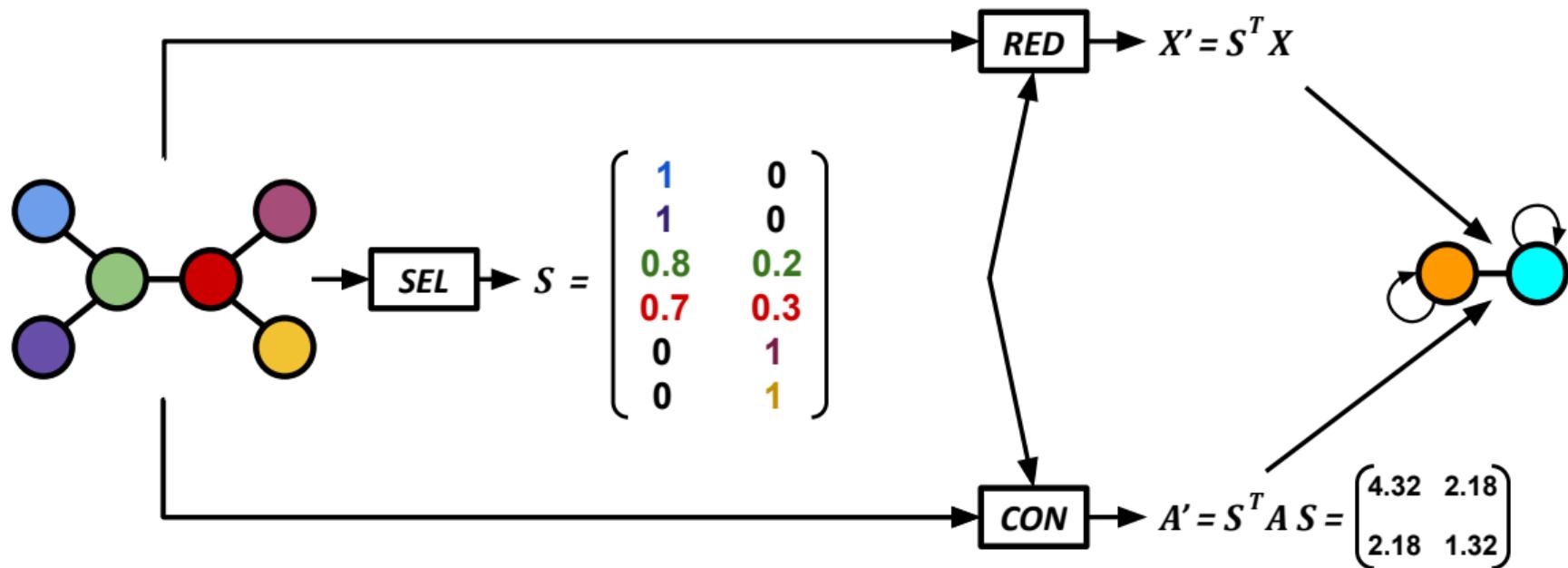
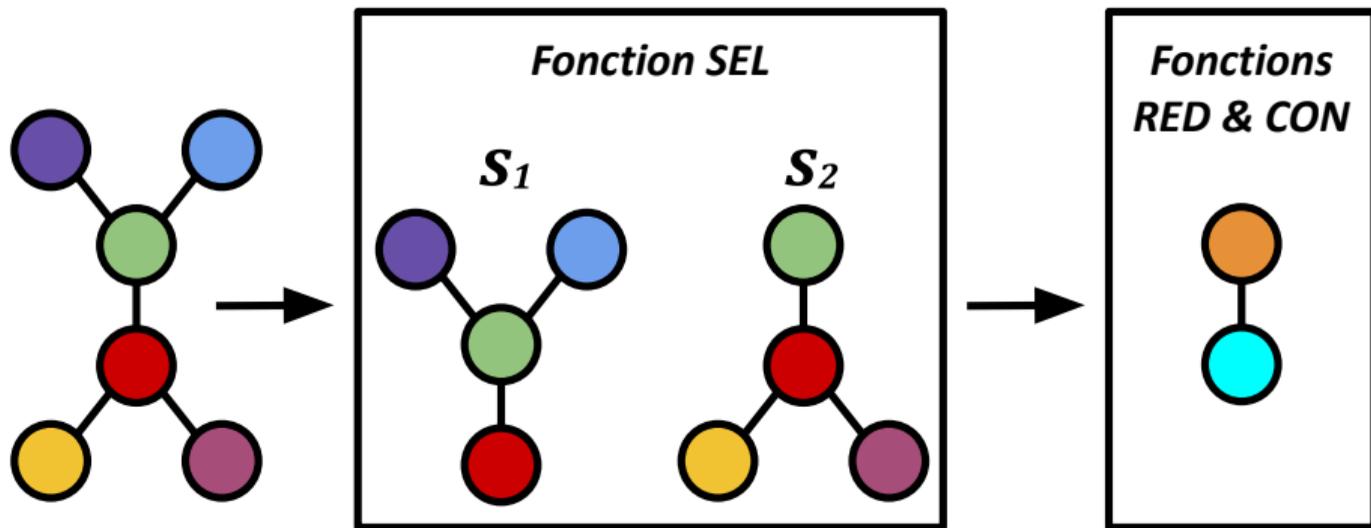


Figure – Schéma d'un pooling avec les fonctions SRC (Grattarola et al. 2024).





Méthode au moins un

Méthodes au moins un les plus connues :

- **DiffPool** (YING et al. 2018)
- **MinCutPool** (BIANCHI et al. 2020)



Méthode au moins un

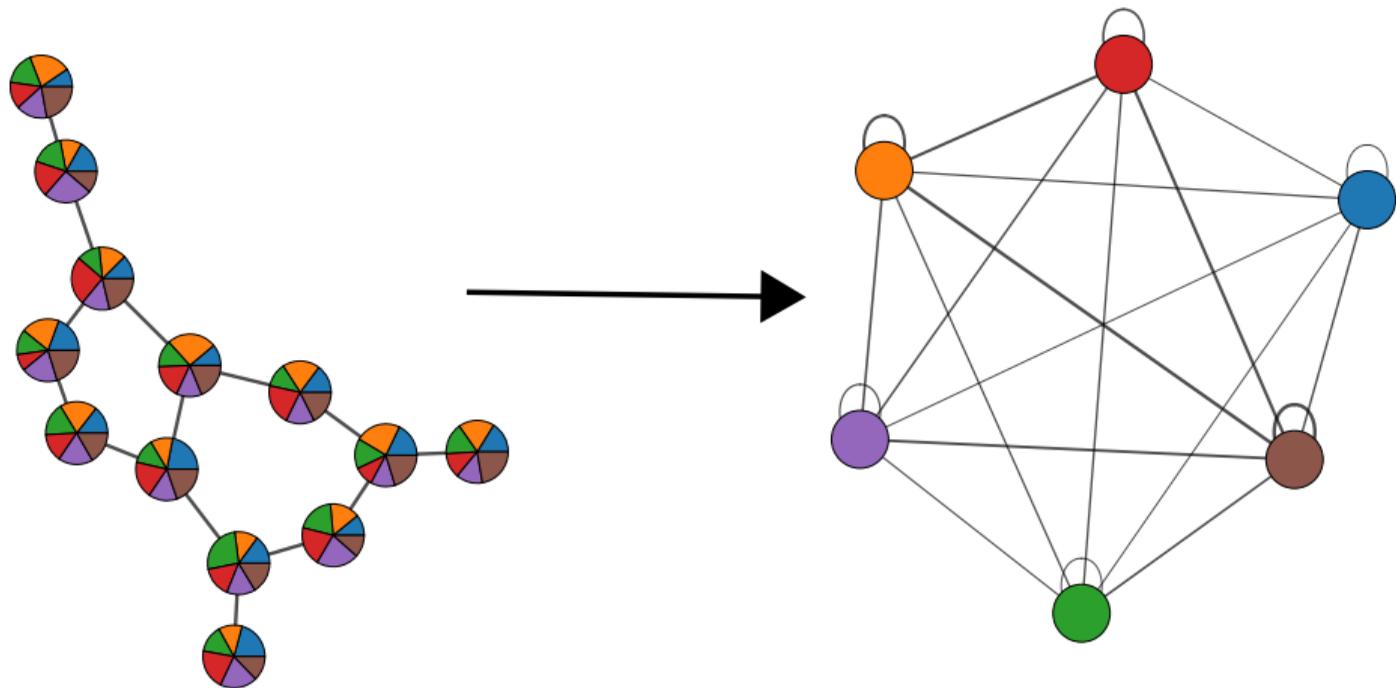
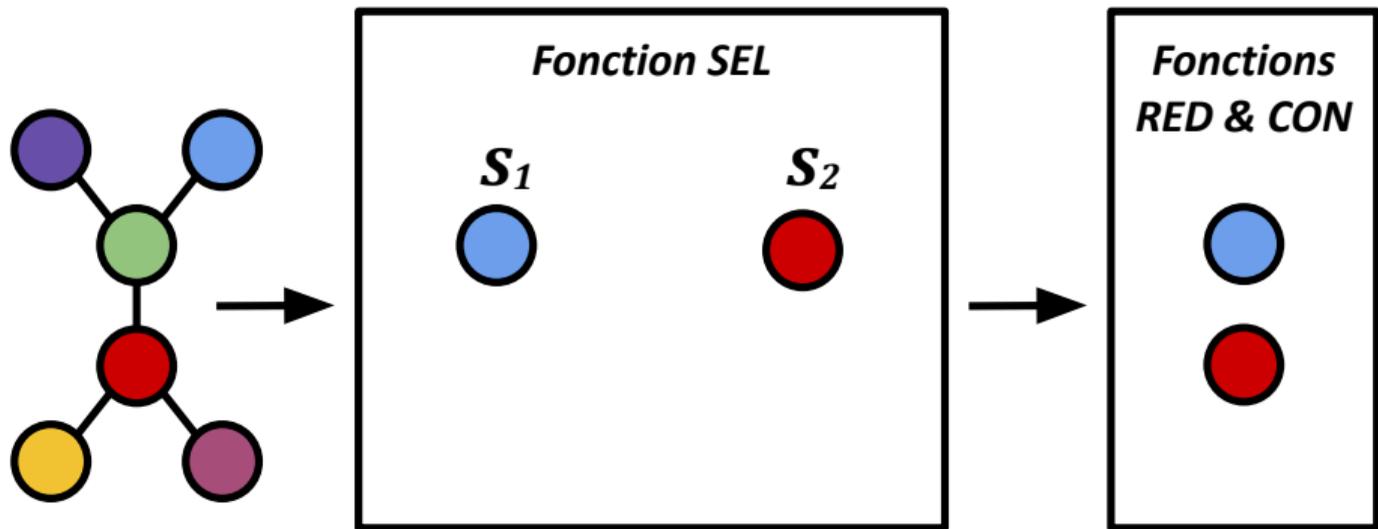


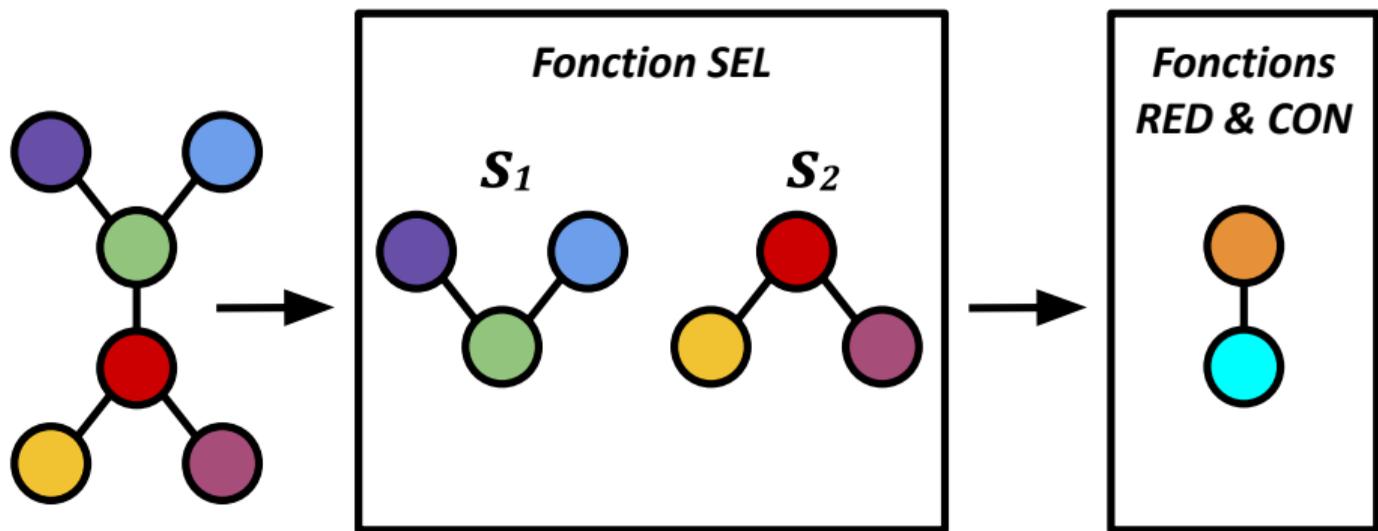
Figure – Exemple d'un graphe réduit après l'application d'une méthode au moins un.





Méthodes au plus un les plus connues :

- **Méthodes Top-k** (GAO et al. 2019; CANGEA et al. 2018; KNYAZEV et al. 2019; LEE et al. 2019; RANJAN et al. 2020; Z. ZHANG, BU, ESTER, J. ZHANG, YAO et al. 2019; Z. ZHANG, BU, ESTER, J. ZHANG, LI et al. 2023; HUANG et al. 2019; Z. MA et al. 2020)





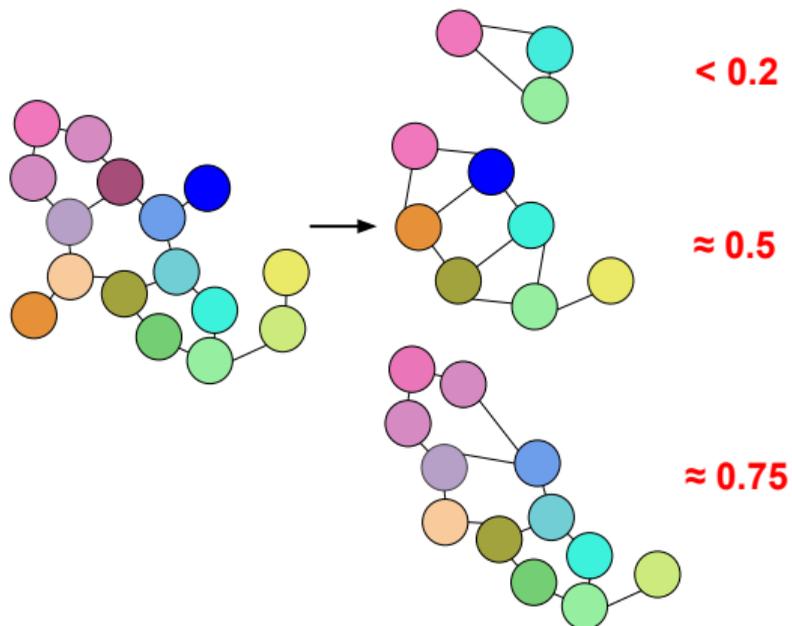
Méthodes exactement un les plus connues :

- **EdgePool** (DIEHL 2019; DIEHL et al. 2019; DHILLON et al. 2007)



Ratio de préservation

Ratio de préservation : $\frac{|\mathcal{G}^{(1+1)}|}{|\mathcal{G}^{(1)}|}$





Convolution :

- Schéma MPNN (GILMER et al. 2017)
- Filtre passe-bas \Rightarrow Redondances dans les attributs

Pooling :

- Schéma SRC (GRATTAROLA et al. 2024)
- Trois familles de méthodes de pooling
- Méthodes exactement un présentent les meilleures propriétés



Néanmoins, ces réseaux souffrent de limitations importantes comme :

Over-smoothing

Over-squashing



Over-smoothing

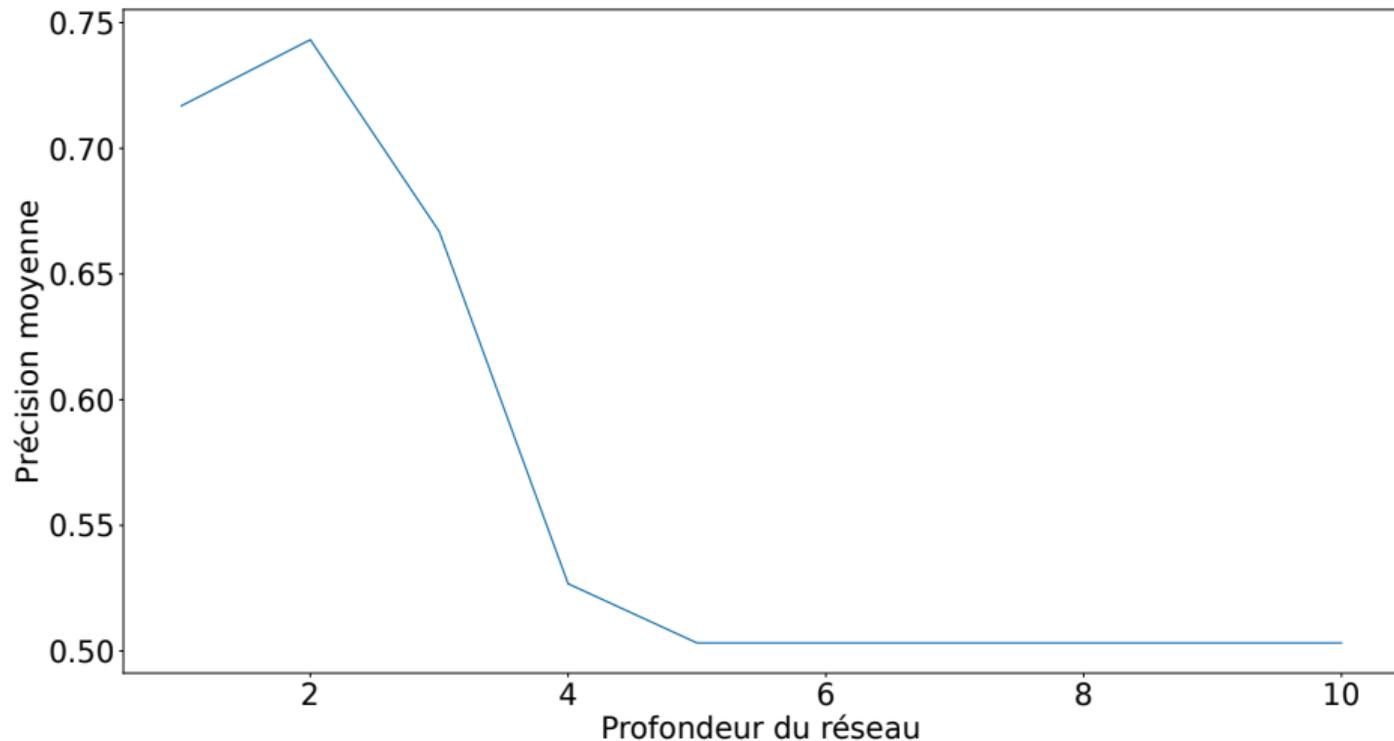


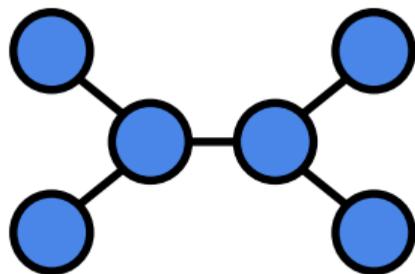
Figure – Évolution de la précision moyenne en fonction de la profondeur du réseau sur NCI109.



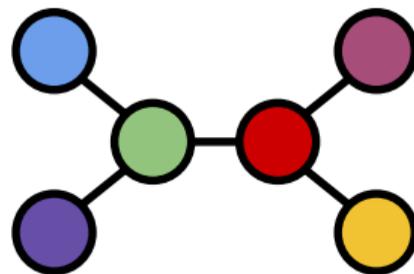
Définition (Énergie de Dirichlet)

Soient \mathbf{X} la matrice des attributs des sommets et \mathbf{L} la matrice laplacienne associée au support convolutionnel \mathbf{C} de la convolution sur graphes. L'énergie de Dirichlet est donnée par :

$$\mathbf{E}(\mathbf{X}) = \frac{1}{|\mathcal{G}^{(l)}|} \text{Trace}(\mathbf{X}^T \mathbf{L} \mathbf{X}) = \frac{1}{|\mathcal{G}^{(l)}|} \sum_{v_i \in \mathcal{V}} \sum_{v_j \in \mathcal{N}(v_i)} \mathbf{C}_{ij} \|\mathbf{X}_i - \mathbf{X}_j\|_2^2$$



$$\mathbf{E}(\mathbf{X}) = 0$$



$$\mathbf{E}(\mathbf{X}) > 0$$



Soit \mathbf{C} un support convolutionnel **symétrique** (et **diagonalisable**), tel que :

- valeurs propres $\lambda_1, \dots, \lambda_m$ (avec λ_1 **dominante**, c'est-à-dire $|\lambda_1| > |\lambda_j|$ pour tout $j > 1$),
- vecteurs propres associés v_1, \dots, v_m .

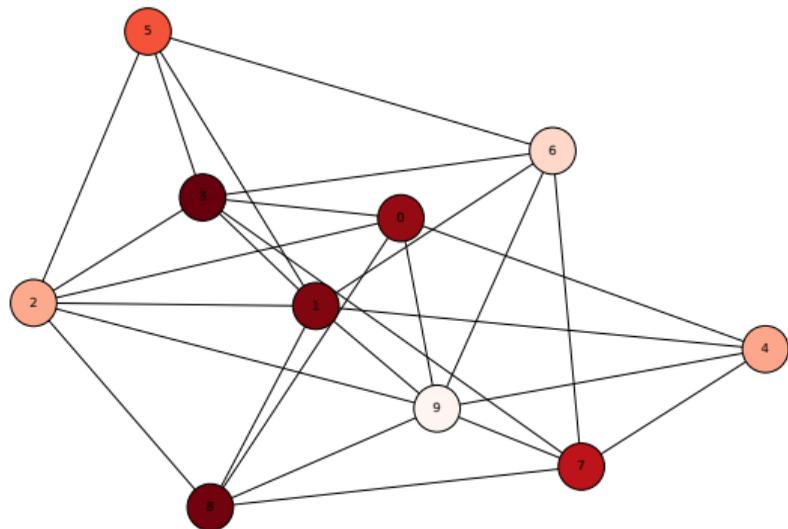
Pour un signal $f = \sum_{i=1}^m c_i v_i$ avec $c_1 \neq 0$, on a :

$$\mathbf{C}^k f \rightarrow c_1 \lambda_1^k v_1 \quad \text{lorsque } k \rightarrow \infty$$

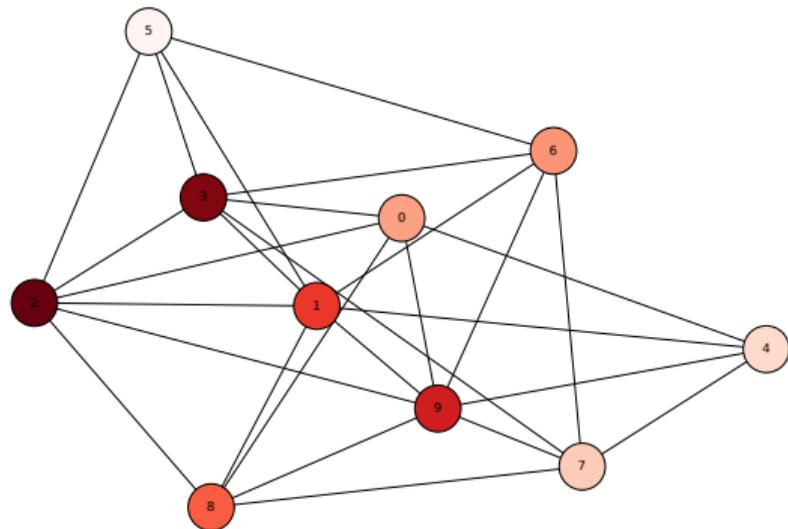
- **Après de nombreuses convolutions, seule la composante dominante (v_1) subsiste.**
- Si \mathbf{C} symétrique normalisée alors : $\mathbf{C}^k f \rightarrow c_1 \mathbf{D}^{\frac{1}{2}} \mathbb{1}$.



Over-smoothing



(a) Graphe initial

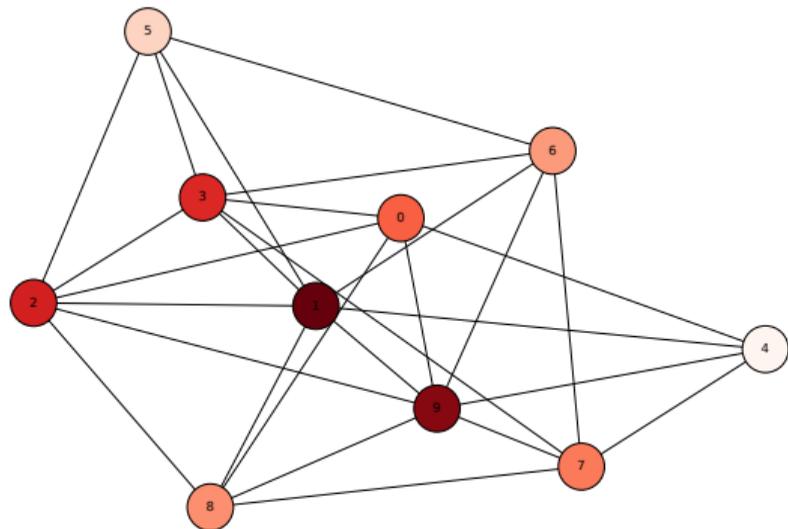


(b) Après 1 convolution

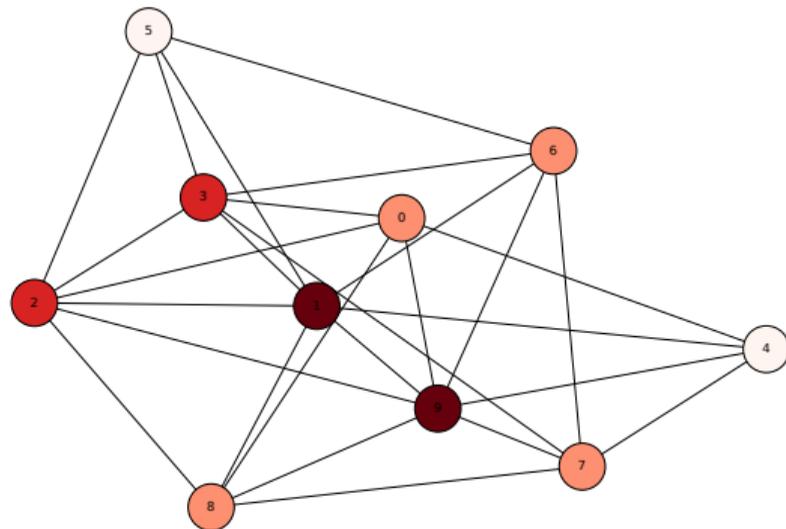
Figure – Exemple d'over-smoothing avec une convolution symétrique normalisée arbitraire.



Over-smoothing



(a) Après 2 convolutions



(b) Racine carrée des degrés

Figure – Exemple d'over-smoothing avec une convolution symétrique normalisée arbitraire.



Néanmoins, ces réseaux souffrent de limitations importantes comme :

Over-smoothing

Over-squashing



Over-squashing

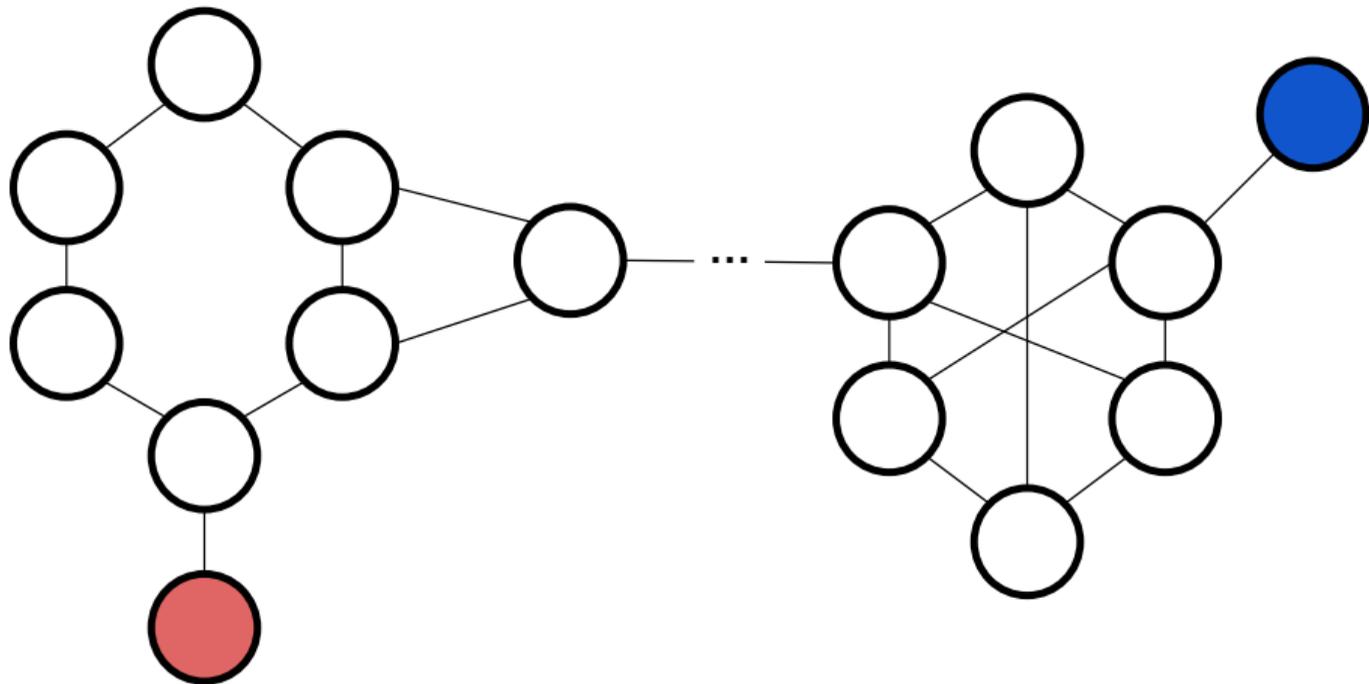


Figure – Exemple de sommets éloignés dans un graphe.



Over-squashing

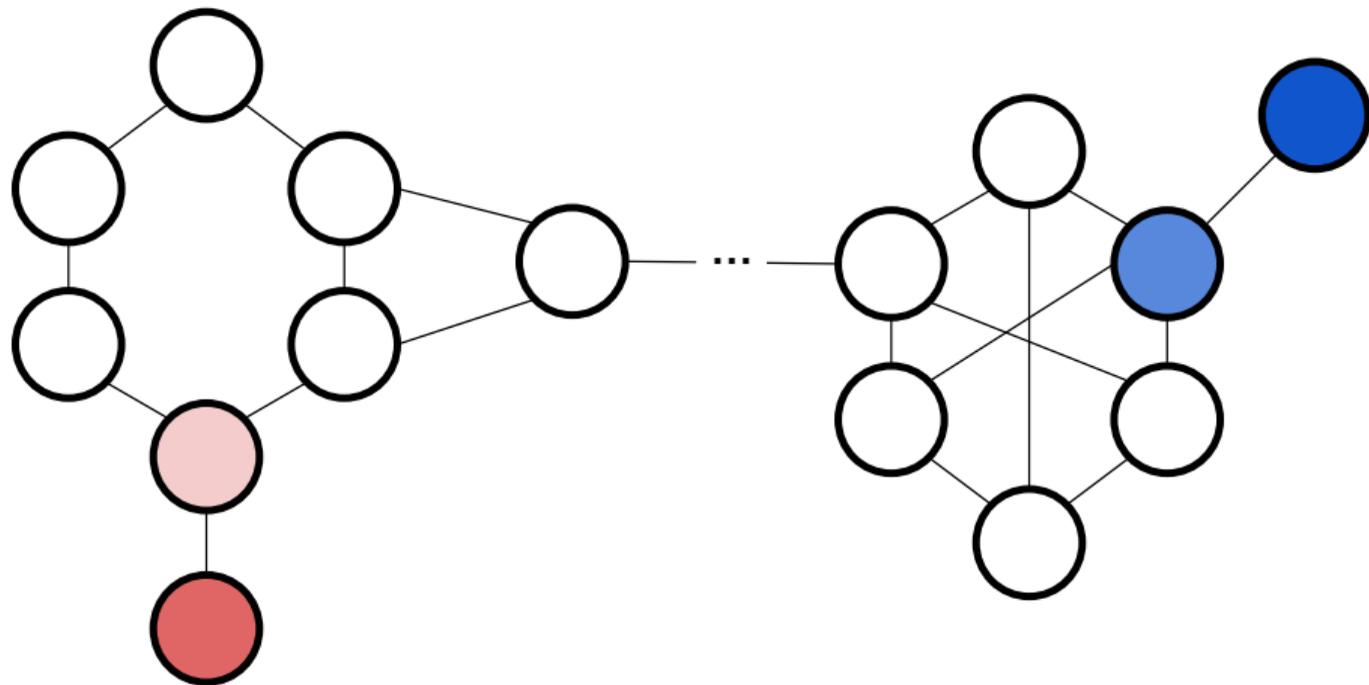


Figure – Exemple de sommets éloignés dans un graphe.



Over-squashing

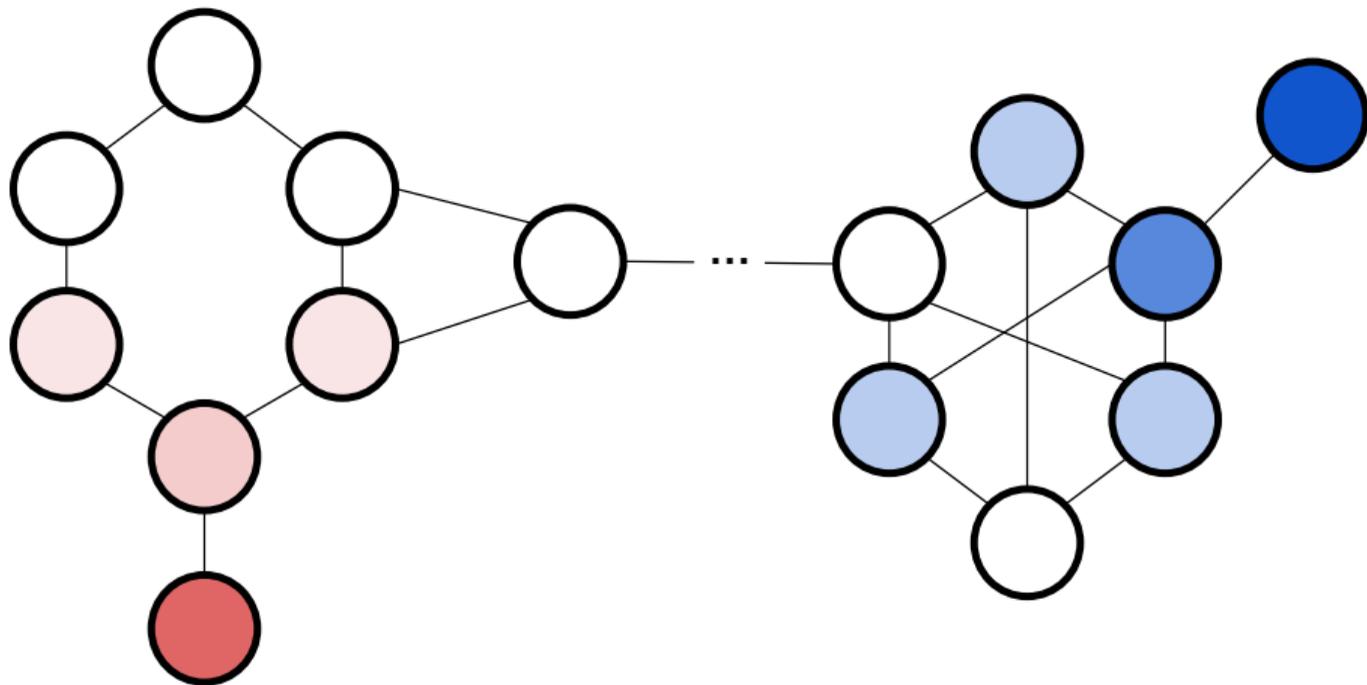


Figure – Exemple de sommets éloignés dans un graphe.



Over-squashing

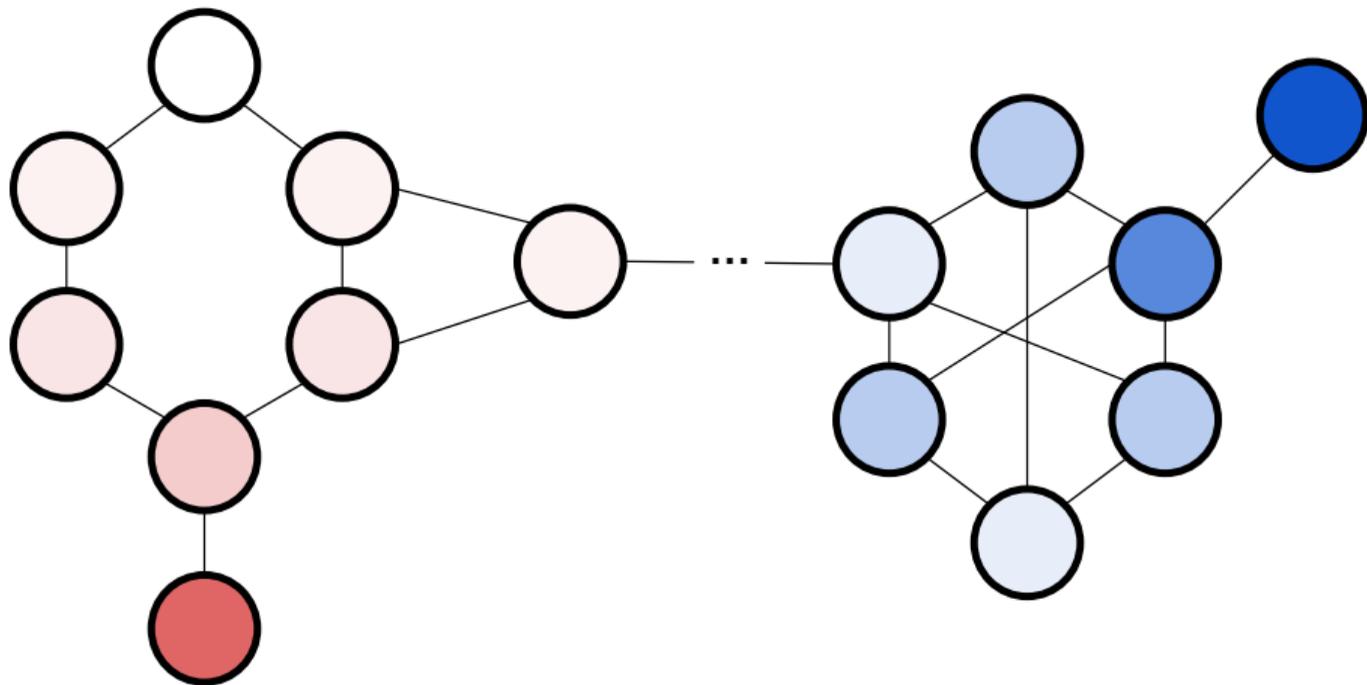
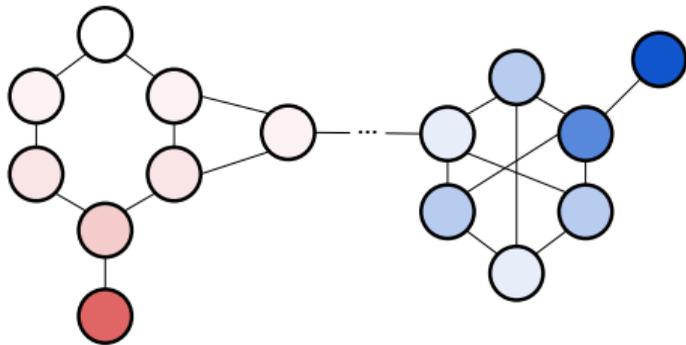


Figure – Exemple de sommets éloignés dans un graphe.



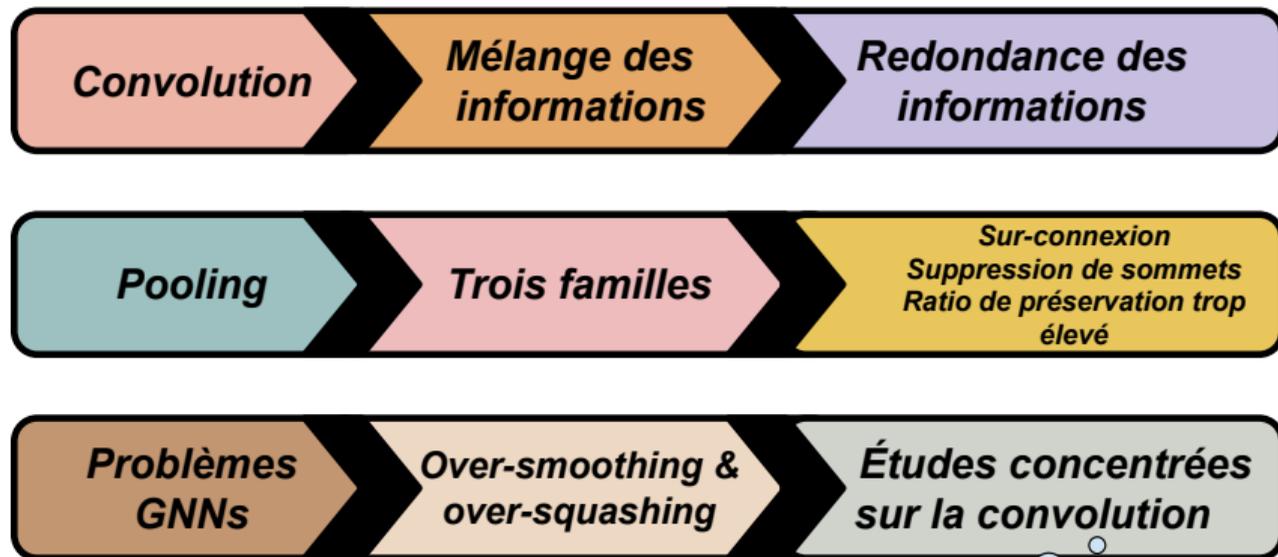
Over-squashing

L'over-squashing correspond au fait qu'un GNN ne transmet pas efficacement les informations entre sommets distants.





Récapitulatif sur l'état de l'art



Quel est l'impact des opérations de pooling sur l'over-smoothing et l'over-squashing ?



Récapitulatif sur l'état de l'art

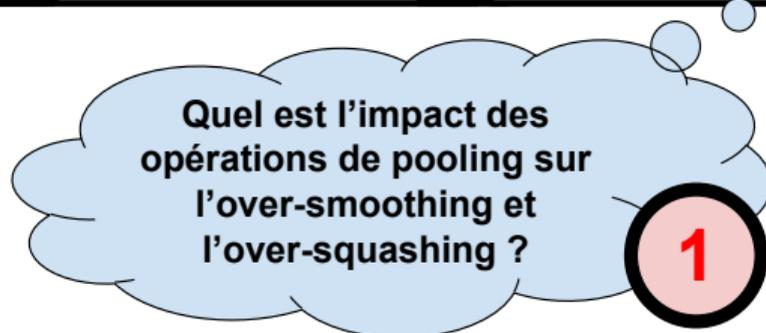
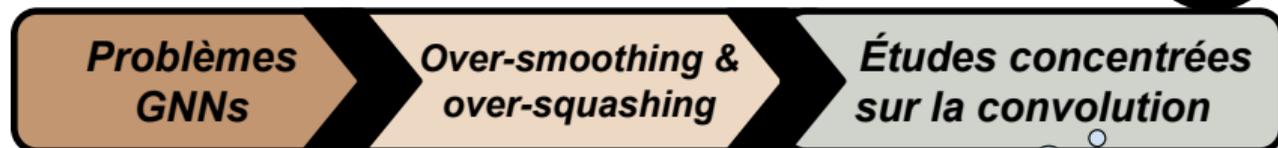
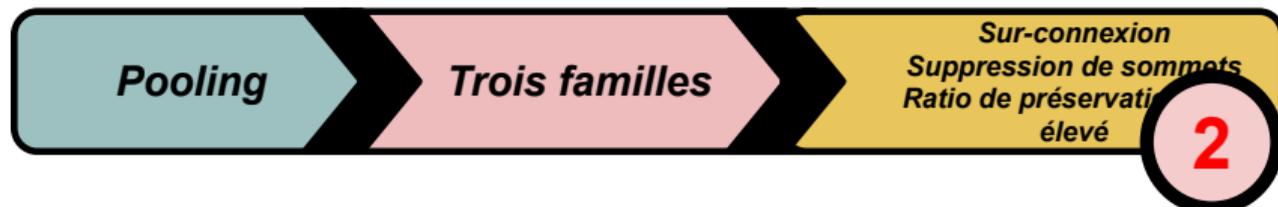


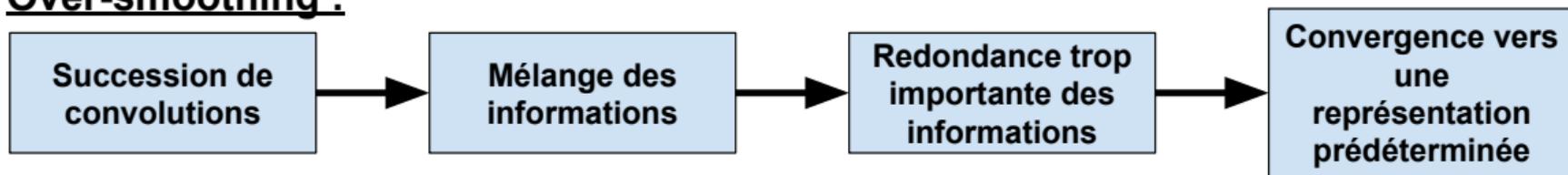


Table des matières

- 1 Introduction et contexte
- 2 État de l'art
- 3 Impact du pooling sur l'over-smoothing et l'over-squashing**
- 4 Méthodes proposées et ensembles indépendants maximaux
- 5 Résultats expérimentaux
- 6 Conclusion

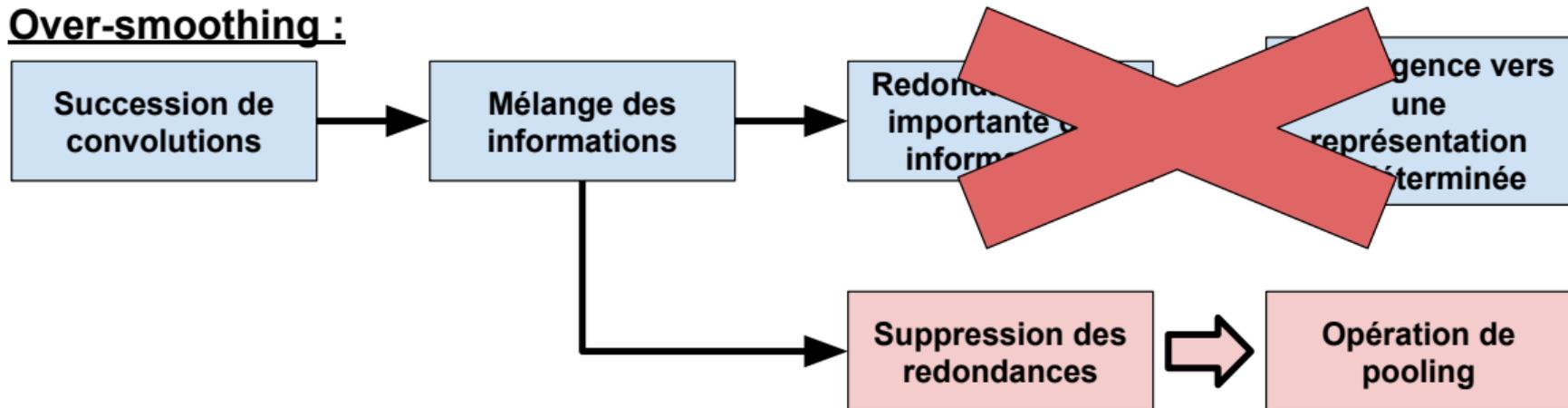


Over-smoothing :



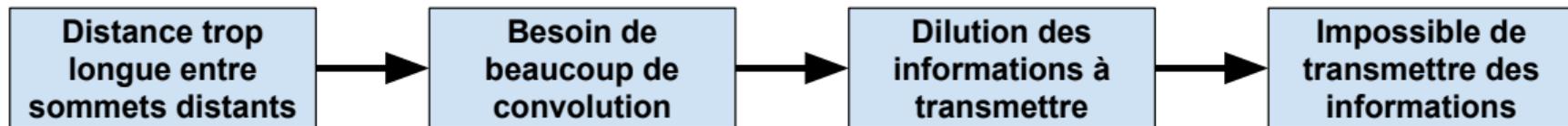


Over-smoothing :



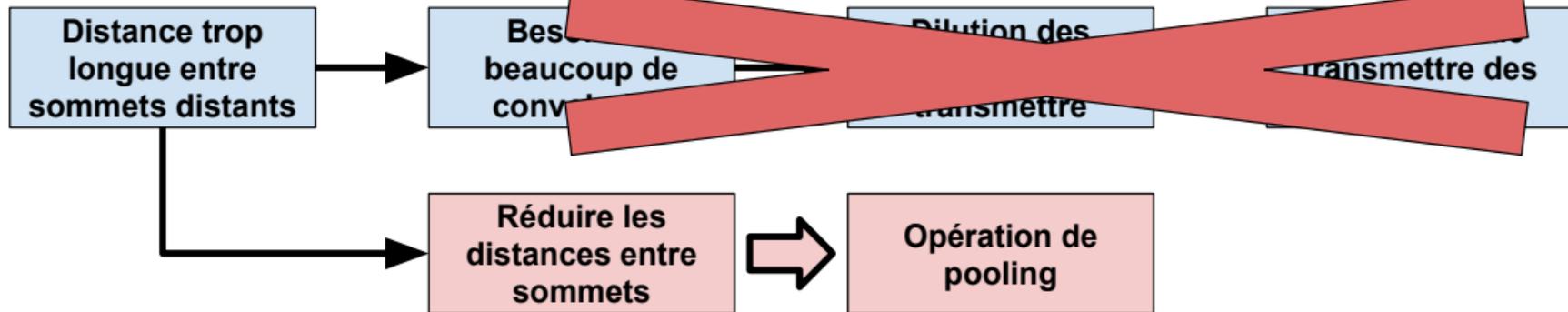


Over-squashing :





Over-squashing :

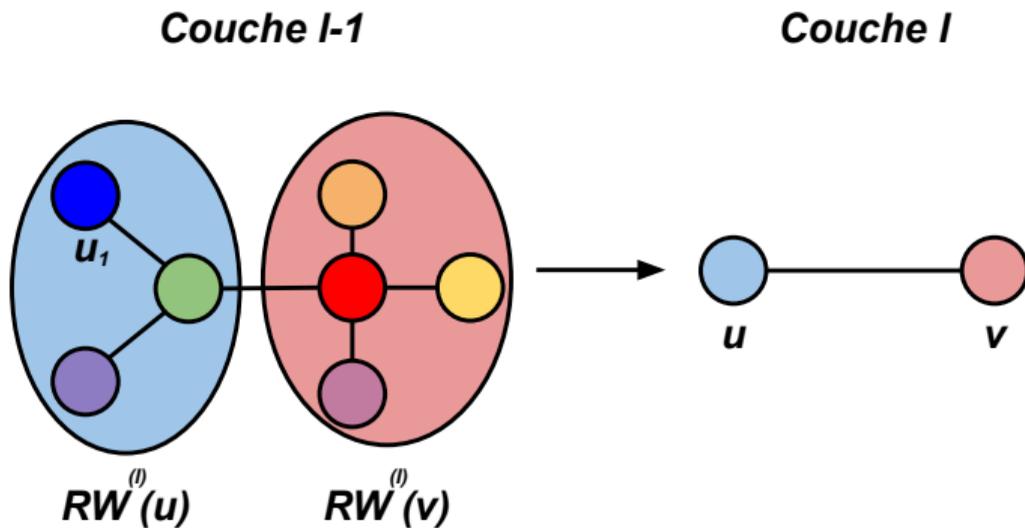




Fenêtres de réduction et champs réceptifs

Fenêtre de réduction $RW^{(l)}(v)$: l'ensemble des sommets contractés pour former v entre les couches $l - 1$ et l

$$\Rightarrow RW^{(0)}(v) = \{v\}$$

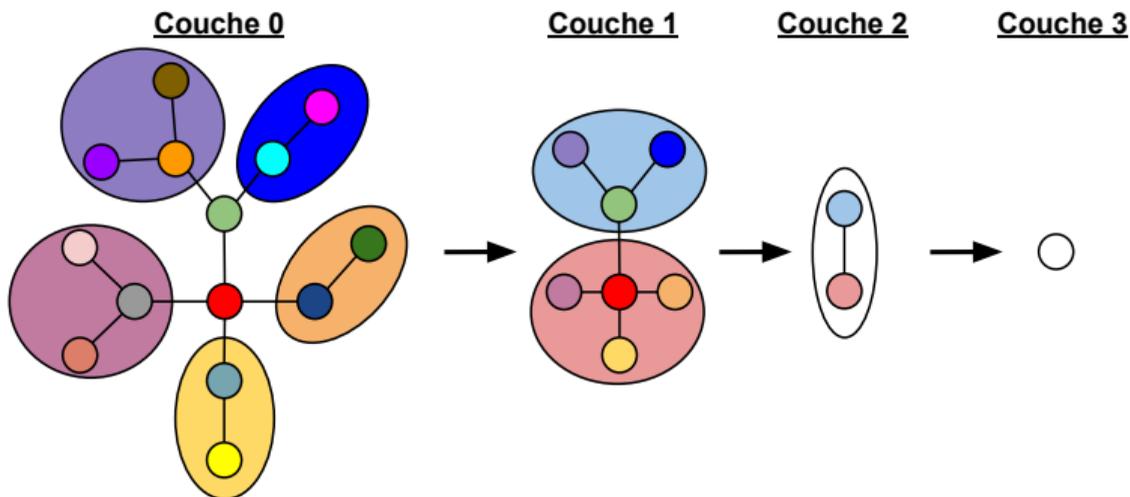




Fenêtres de réduction et champs réceptifs

Champs réceptif $RF^{(l)}(v)$: l'ensemble des sommets contractés pour former v entre les couches 0 et l

$$\Rightarrow RF^{(l)}(v) = \bigcup_{u \in RW^{(l)}(v)} RF^{(l-1)}(u) \text{ avec } RF^{(0)}(u) = RW^{(0)}(u)$$





Definition (Schéma de décimation)

Nous supposons que toutes les fenêtres de réduction étudiées satisfont les équations suivantes pour tout sommet $w \in \mathcal{V}^{(l)}$:

$$\begin{cases} RW^{(l)}(w) = \{w\} \text{ ou} \\ RW^{(l)}(w) = \{w, v_1, \dots, v_n\} \mid \forall i \in \{1, \dots, n\} d_{\mathcal{G}^{(l-1)}}(w, v_i) = 1 \end{cases}$$

où $d_{\mathcal{G}^{(l)}}$ est la distance au sein du graphe $\mathcal{G}^{(l)}$.



Impact du pooling sur l'over-squashing

En utilisant **notre schéma de décimation** et en supposant **une réduction constante γ** des distances entre les sommets, on a le résultat suivant :

$$\log_3 \left(\frac{d_{\mathcal{G}^{(0)}}(i,j)+1}{2} \right) \leq m \leq \frac{\log(d_{\mathcal{G}^{(0)}}(i,j))}{\log(\gamma)} + 1$$

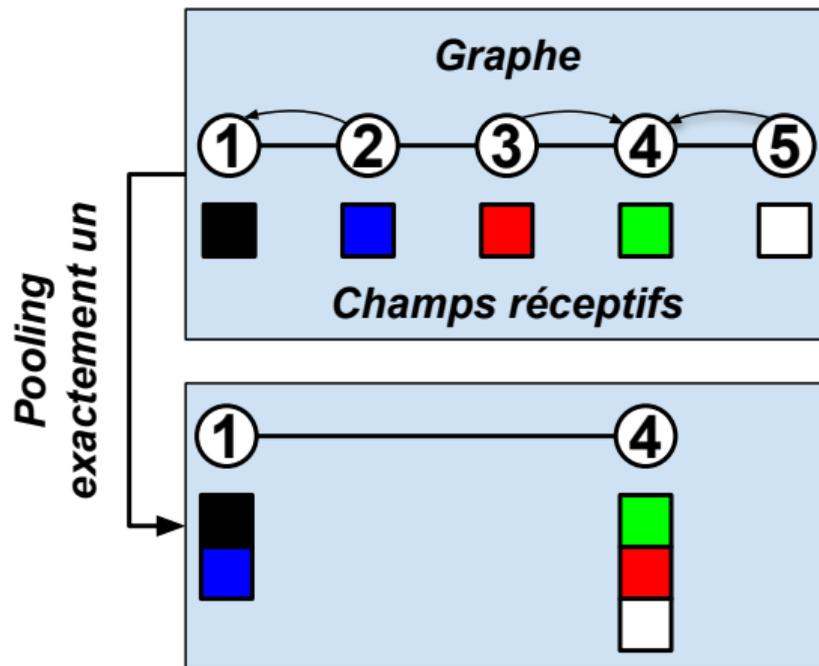
où m est le **nombre de couches nécessaires** pour l'intersection.

Par conséquent, une **relation logarithmique est nécessaire** pour que les attributs de deux sommets s'intersectent :

$$m \approx \log(d_{\mathcal{G}^{(0)}}(i, j))$$

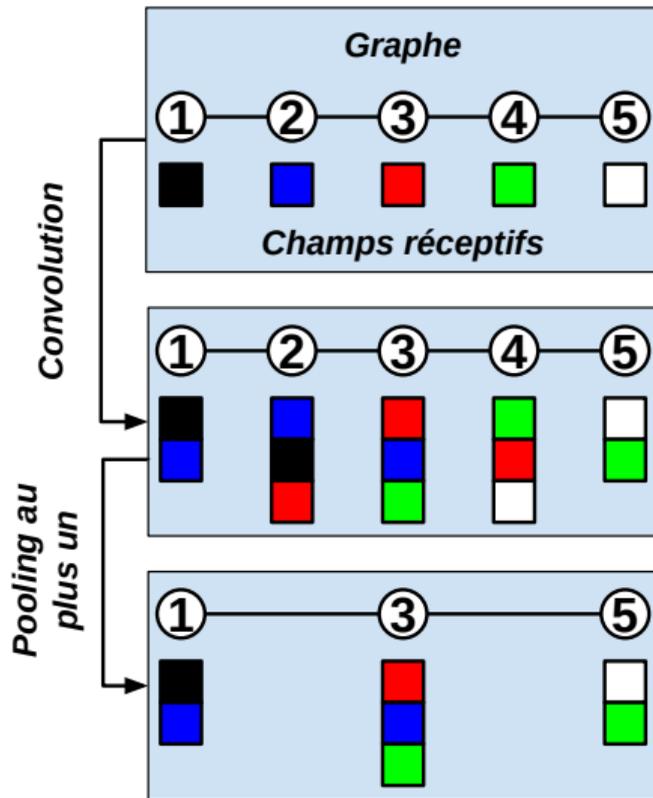


Impact du pooling exactement un sur l'over-smoothing





Impact du pooling au plus un avec convolution sur l'over-smoothing





Récapitulatif des résultats théoriques

Contraintes / Résultats	Pooling exactement un	Pooling au plus un avec convolution
Schéma de décimation	✓	✓
GNN sans convolution	✓	
GNN alternance convolution/pooling		✓
Garantie une partition des sommets	✓	
Over-smoothing	Insensible	Réduit
Over-squashing	Réduit	Réduit



Table des matières

- 1 Introduction et contexte
- 2 État de l'art
- 3 Impact du pooling sur l'over-smoothing et l'over-squashing
- 4 Méthodes proposées et ensembles indépendants maximaux**
- 5 Résultats expérimentaux
- 6 Conclusion



Rappel de l'état de l'art

- Considérer la structure initiale du graphe lors de la réduction
- Préserver les informations des attributs
- Réduire l'information redondante des convolutions
- Ratio de préservation ≈ 0.5



Famille de quatre méthodes exactement un basée sur les ensembles indépendants maximaux (MIS) :

MIVSPool

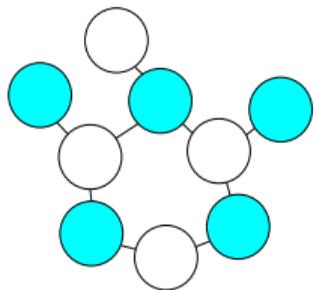
MIESPpool

MIESCUTpool

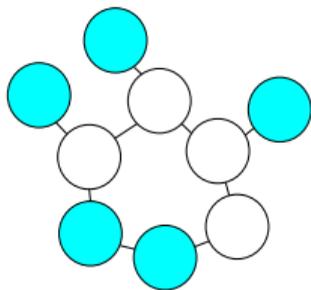
MIDESPpool



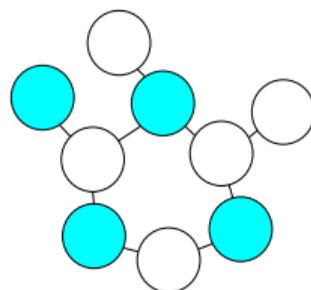
Maximal Independent Set (MIS)



Ensemble de sommets indépendants maximaux



Première propriété non respectée



Seconde propriété non respectée

Propriétés

- 1) Deux sommets survivants ne peuvent être adjacents/voisins :

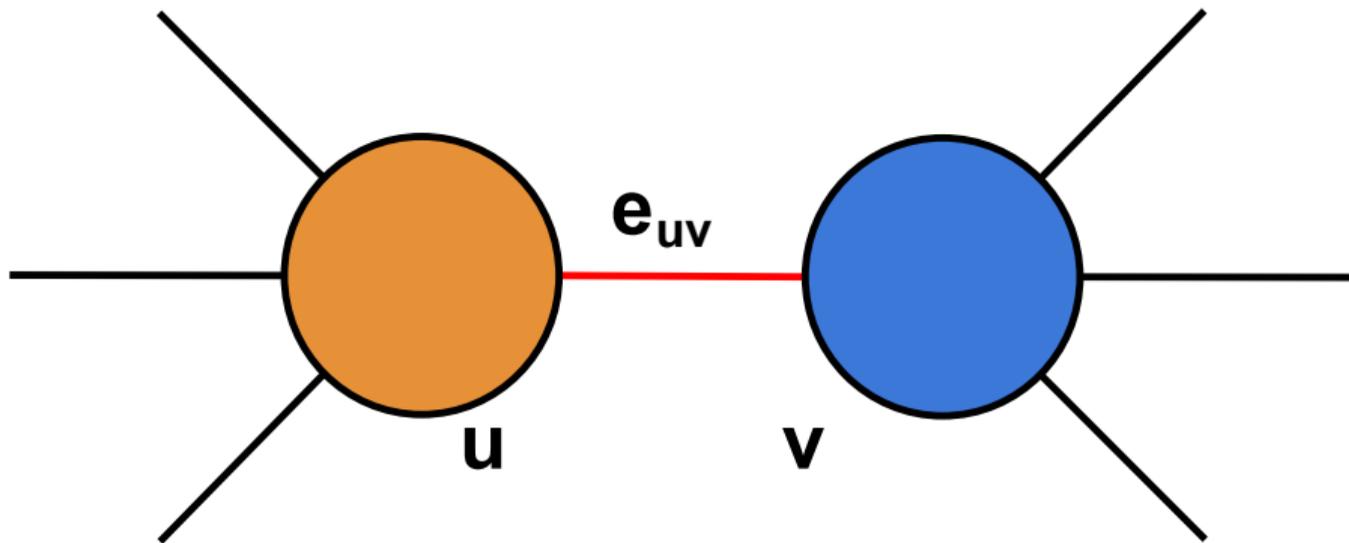
$$\forall (v, v') \in (\mathcal{V}^{(l+1)})^2 : (v, v') \notin \mathcal{E}^{(l)}$$

- 2) Un sommet non-survivant possède au moins un voisin survivant :

$$\forall v \in \mathcal{V}^{(l)} - \mathcal{V}^{(l+1)}, \exists v' \in \mathcal{V}^{(l+1)} : (v, v') \in \mathcal{E}^{(l)}$$



Maximal Independent Set (MIS)





Maximal Independent Set (MIS)

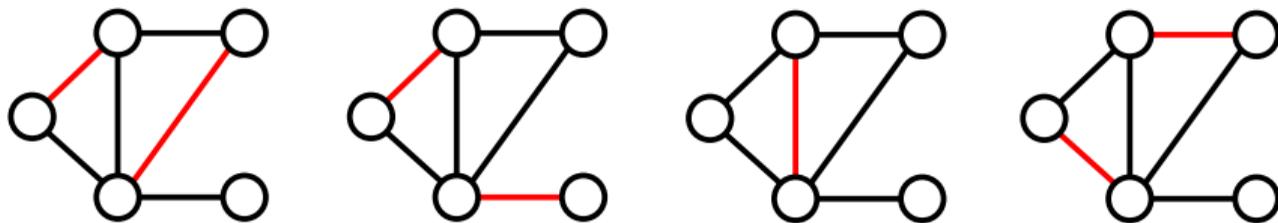
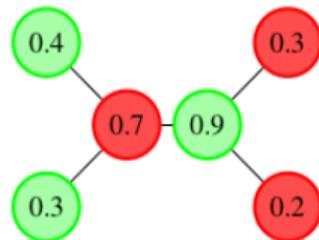
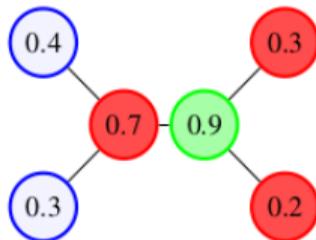
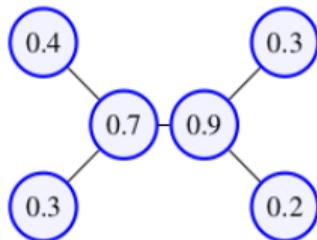


Figure – Appariement maximal (Maximal Matching)



Algorithme de Meer (MEER 1989)



Propriétés

1) Deux sommets survivants ne peuvent être adjacents/voisins :

$$\forall (v, v') \in (\mathcal{V}^{(l+1)})^2 : (v, v') \notin \mathcal{E}^{(l)}$$

2) Un sommet non-survivant possède au moins un voisin survivant :

$$\forall v \in \mathcal{V}^{(l)} - \mathcal{V}^{(l+1)}, \exists v' \in \mathcal{V}^{(l+1)} : (v, v') \in \mathcal{E}^{(l)}$$



MIVSPool : Maximal Independent Vertex Set Pooling

Famille de quatre méthodes exactement un basée sur les ensembles indépendants maximaux (MIS) :

MIVSPool

MIESPpool

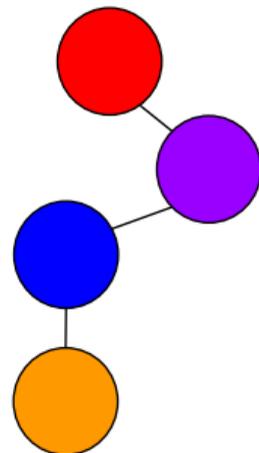
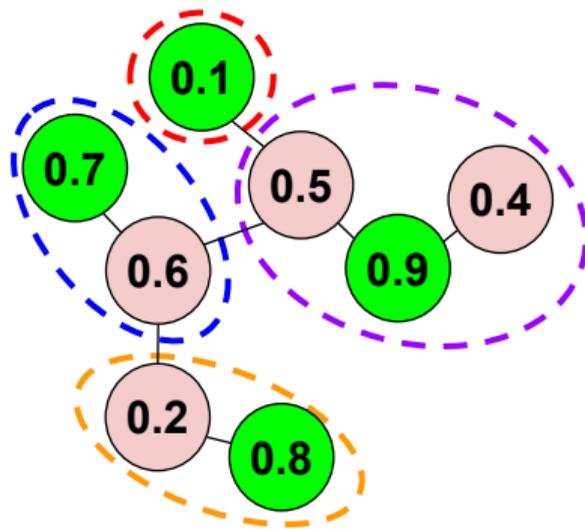
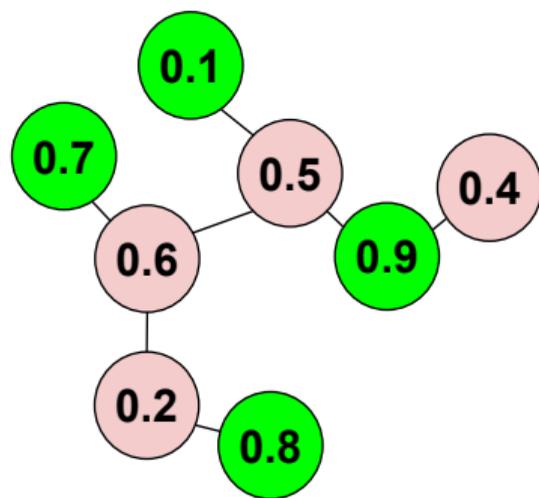
MIESCUTpool

MIDESPpool



MIVSPool : Maximal Independent Vertex Set Pooling

- Calcul de l'importance des sommets
- Calcul du MIS avec l'algorithme de Meer (Meer 1989)
- Affectation des sommets non-survivants





MIVSPool : Maximal Independent Vertex Set Pooling

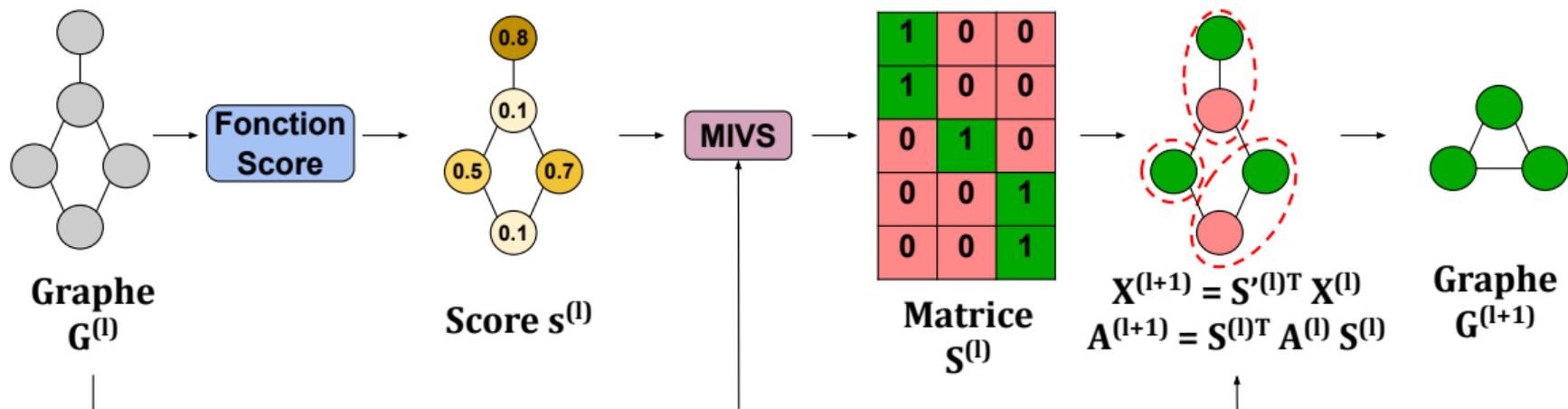


Figure – Schéma récapitulatif de MIVSPool.



Problématique du calcul d'un score sur les sommets

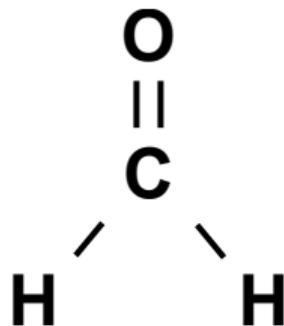


Figure – Molécule avec double liaison.



MIESPool : Maximal Independent Edge Set Pooling

Famille de quatre méthodes exactement un basée sur les ensembles indépendants maximaux (MIS) :

MIVSPool

MIESPool

MIESCUTPool

MIDESPool

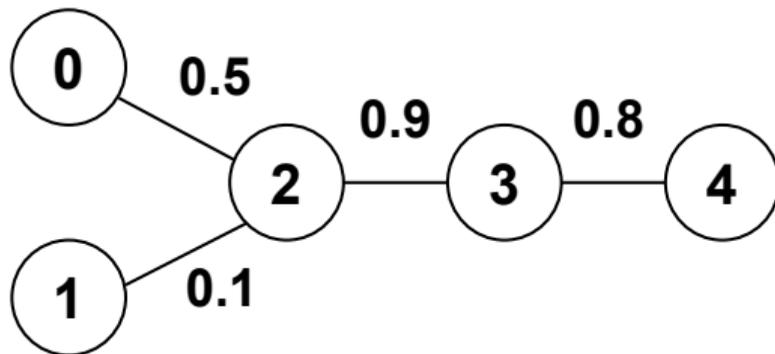


Maximal Independent Edge Set Pooling (MIESPool)

- Appariement maximal
- Fonction de similarité symétrique : $\exp(-\|\mathbf{W}^{(l)} \cdot (x_u - x_v)\|)$

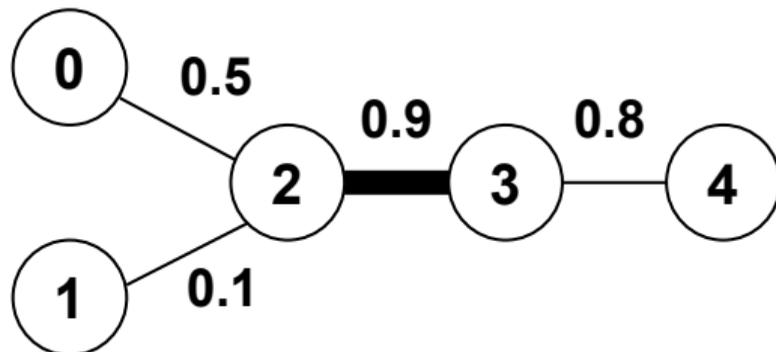


MIESPool : Maximal Independent Edge Set Pooling



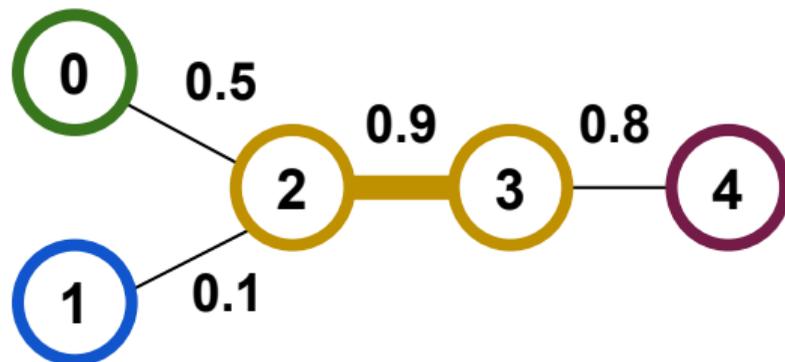


MIESPool : Maximal Independent Edge Set Pooling





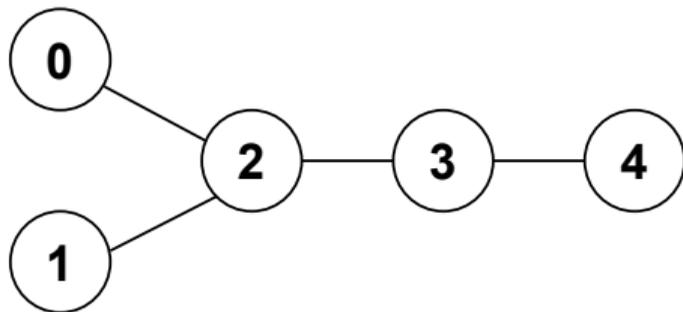
MIESPool : Maximal Independent Edge Set Pooling



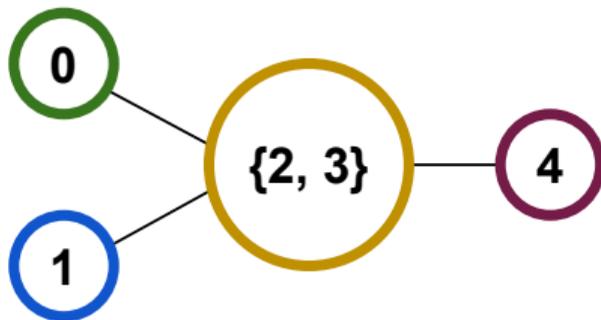


MIESPool : Maximal Independent Edge Set Pooling

Avant



Après





MIESPool : Maximal Independent Edge Set Pooling

Maximal Independent Edge Set Pooling (MIESPool)

- Ratio de préservation élevé **X**



MIDESPool : Maximal Independent Directed Edge Set Pooling

Famille de quatre méthodes exactement un basée sur les ensembles indépendants maximaux (MIS) :

MIVSPool

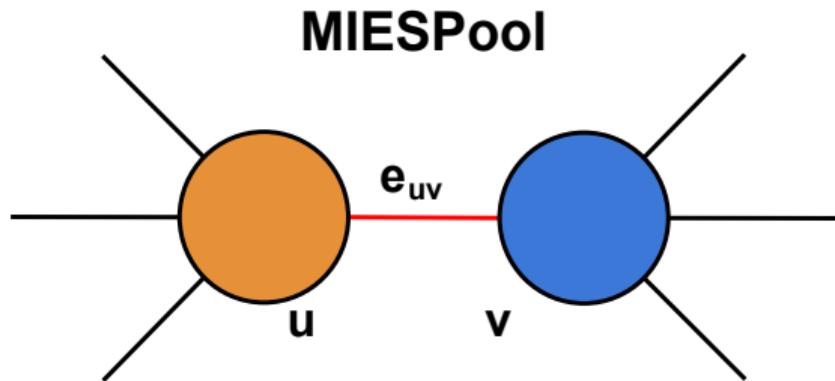
MIESPpool

MIESCUTPool

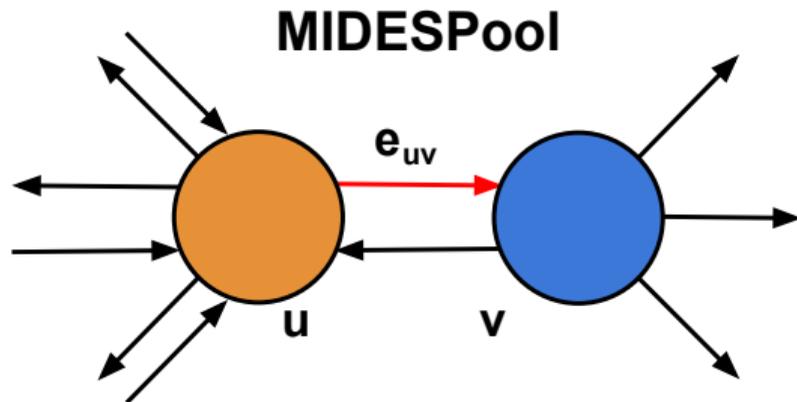
MIDESPpool



MIDESPool : Maximal Independent Directed Edge Set Pooling



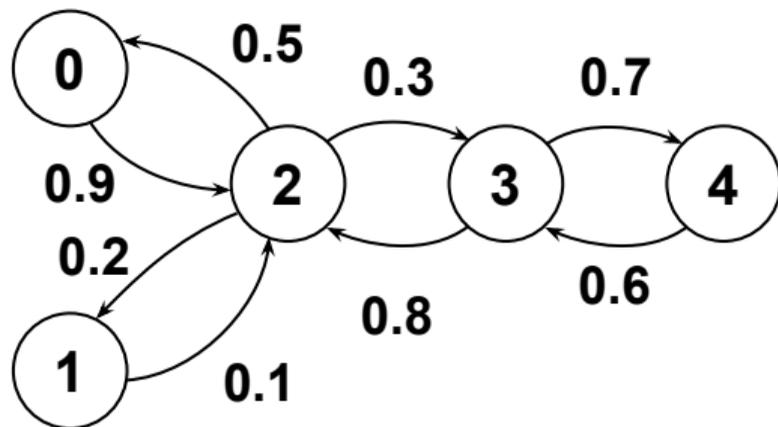
$$\exp(-\|\mathbf{W}^{(l)} \cdot (x_u - x_v)\|)$$



$$\exp(-\|\mathbf{W}^{(l)} \cdot (x_u - x_v) + b\|)$$

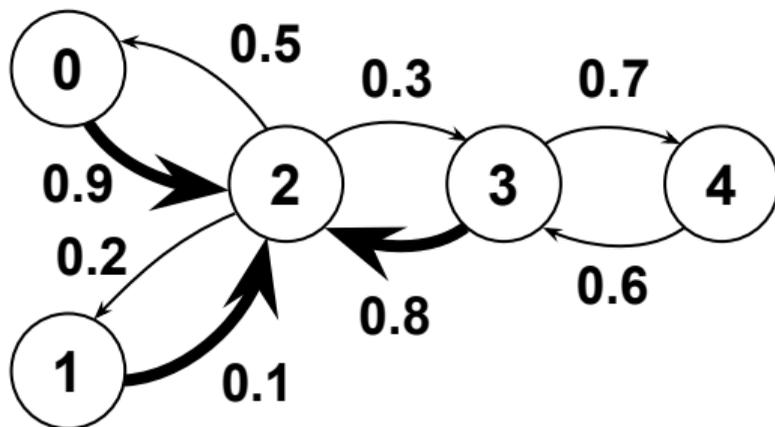


MIDESPool : Maximal Independent Directed Edge Set Pooling



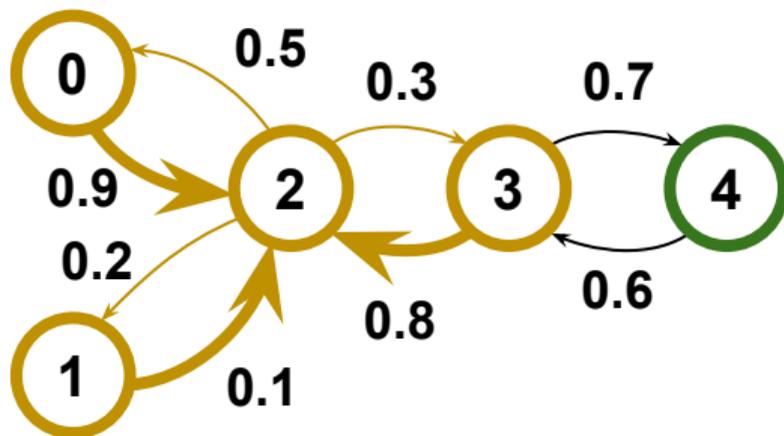


MIDESPool : Maximal Independent Directed Edge Set Pooling





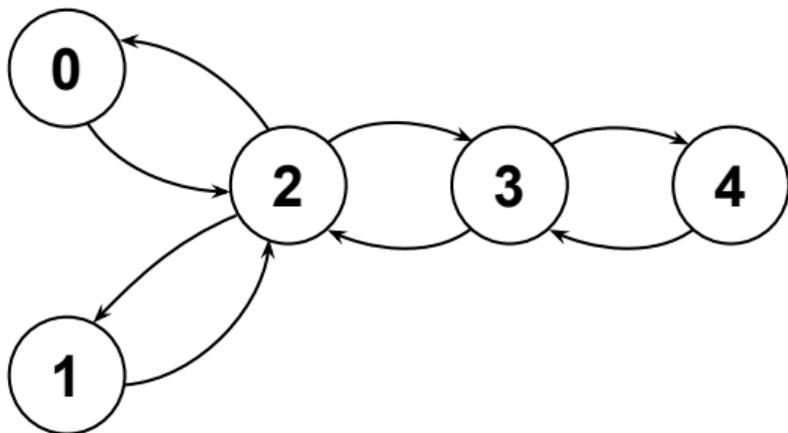
MIDESPool : Maximal Independent Directed Edge Set Pooling



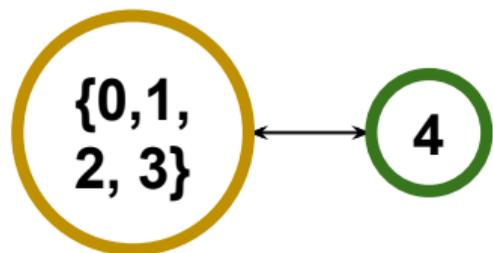


MIDESPool : Maximal Independent Directed Edge Set Pooling

Avant



Après





MIDESPool : Maximal Independent Directed Edge Set Pooling

Maximal Independent Directed Edge Set Pooling (MIDESPool)

- Ratio de préservation ≈ 0.5 ✓



Famille de quatre méthodes exactement un basée sur les ensembles indépendants maximaux (MIS) :

- Considérer la structure initiale du graphe lors de la réduction
- Préserver les informations des attributs
- Réduire l'information redondante des convolutions
- Ratio de préservation ≈ 0.5



Table des matières

- 1 Introduction et contexte
- 2 État de l'art
- 3 Impact du pooling sur l'over-smoothing et l'over-squashing
- 4 Méthodes proposées et ensembles indépendants maximaux
- 5 Résultats expérimentaux**
- 6 Conclusion



Résultats théoriques sur l'over-squashing

- Nombre de couches nécessaire suit une relation logarithmique
- Quelle quantité d'informations ?



Étude expérimentale sur l'over-squashing

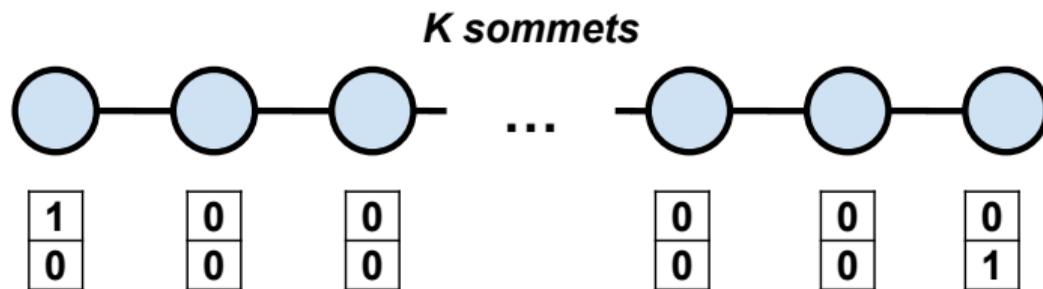


Figure – Schéma des graphes utilisés pour l'expérience d'over-squashing.



Étude expérimentale sur l'over-squashing

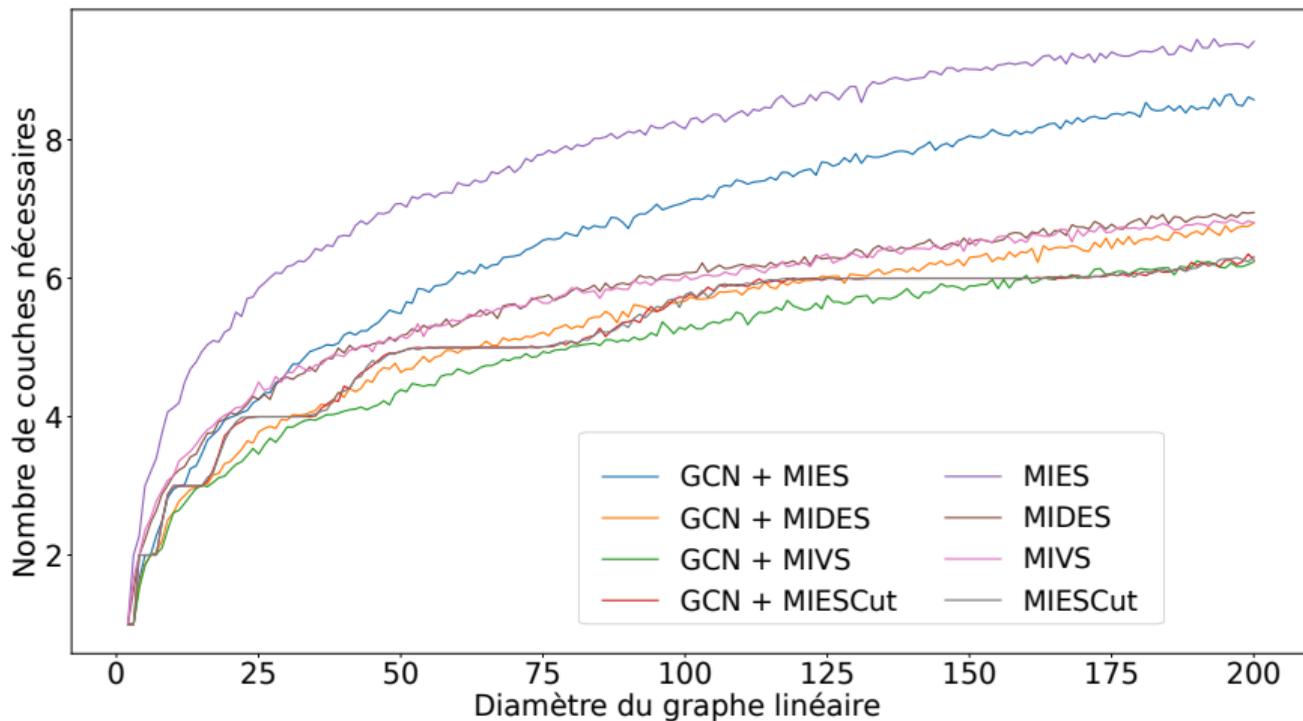


Figure – Nombre de couches nécessaire pour combiner les attributs des sommets distants.



Étude expérimentale sur l'over-squashing

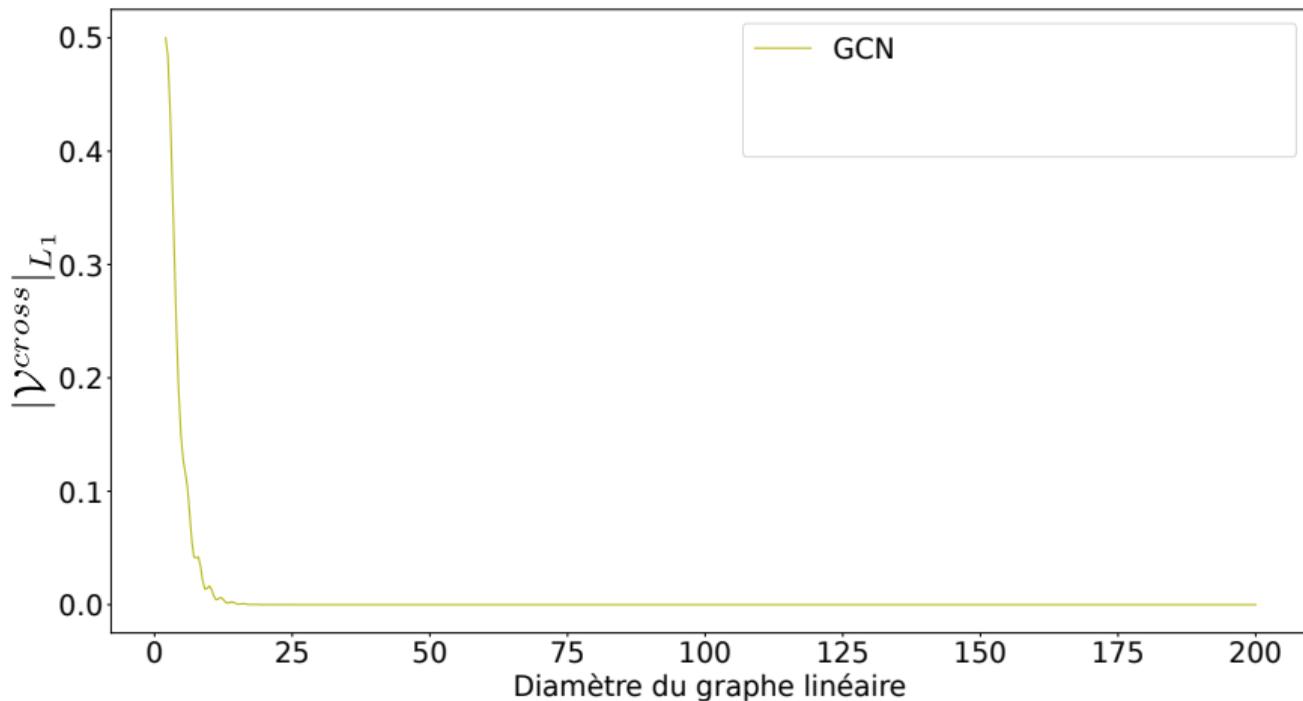


Figure – Quantité moyenne d'attributs transmise lors de la combinaison.



Étude expérimentale sur l'over-squashing

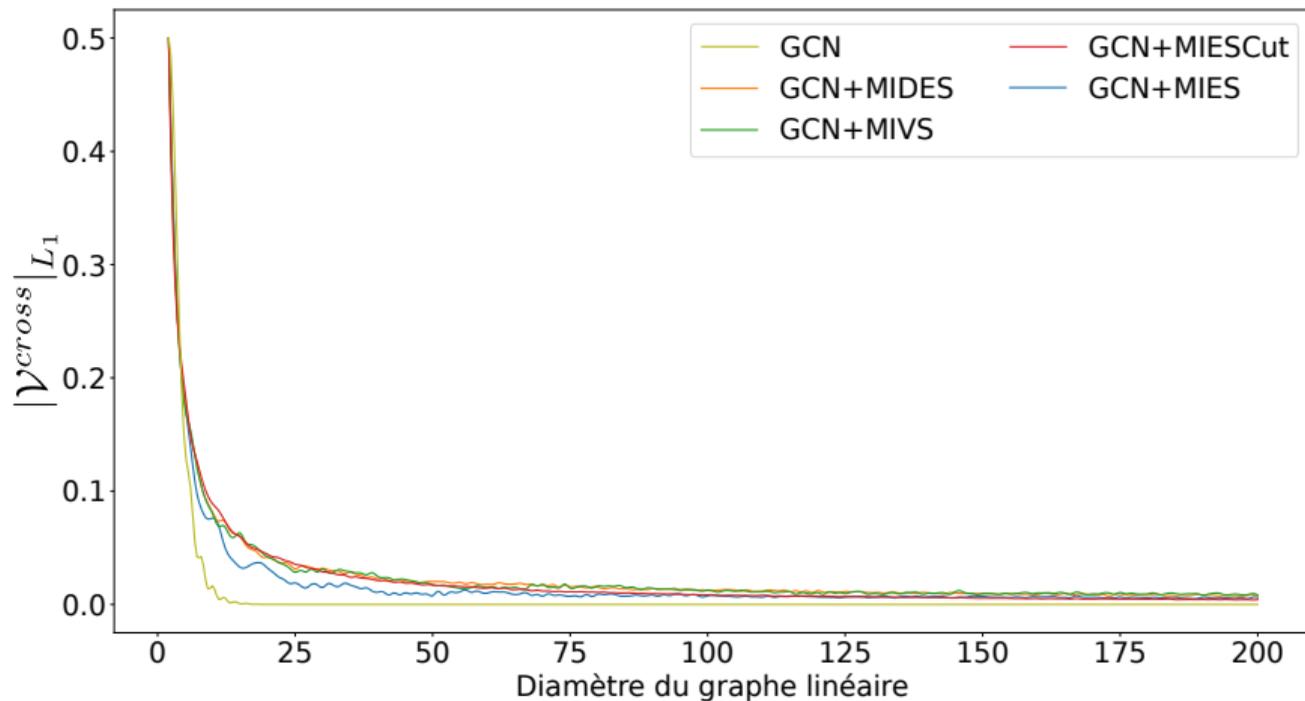


Figure – Quantité moyenne d'attributs transmise lors de la combinaison.



Étude expérimentale sur l'over-squashing

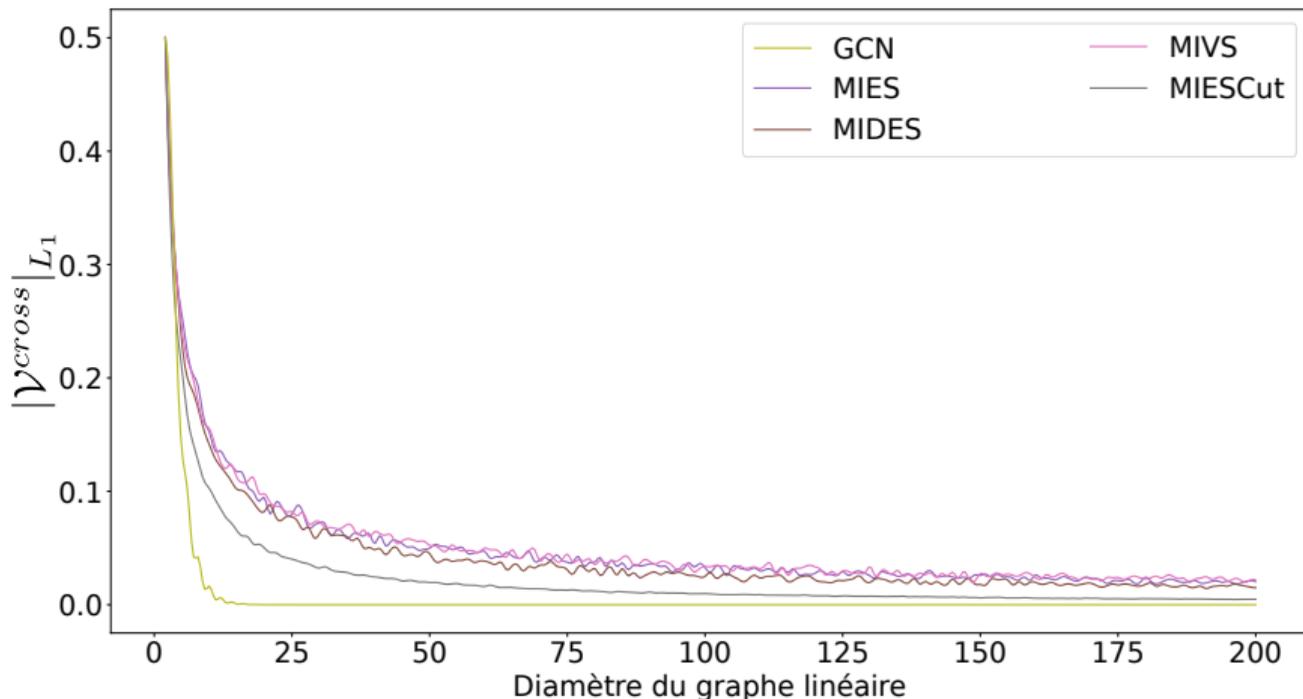


Figure – Quantité moyenne d'attributs transmise lors de la combinaison.



Résultats théoriques sur l'over-smoothing

- Réduction de l'effet d'over-smoothing

⇒ Étude de la variation de l'énergie de Dirichlet par rapport à $\frac{|\overline{RF}_{\mathcal{G}}^{(l)}|}{|\mathcal{G}|}$



Étude expérimentale sur l'over-smoothing

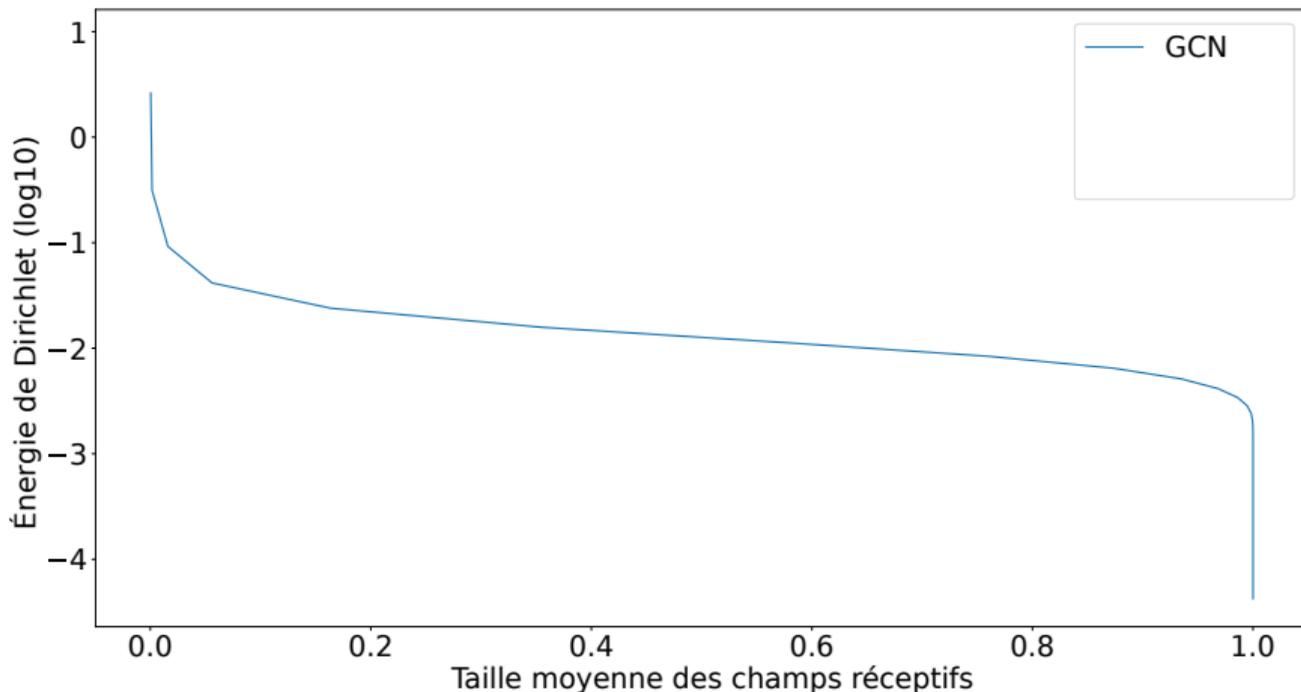


Figure – Énergie de Dirichlet calculée avec GCN et nos méthodes purement pooling sur la composante principale de Cora.



Étude expérimentale sur l'over-smoothing

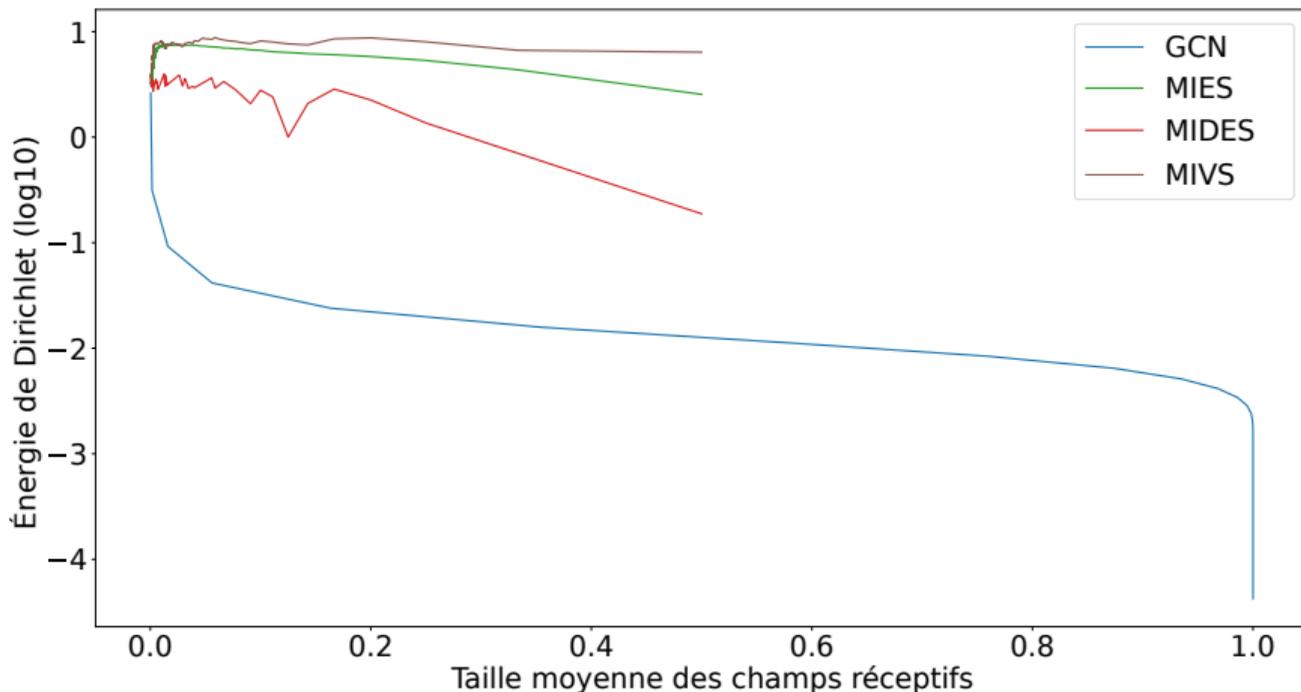


Figure – Énergie de Dirichlet calculée avec GCN et nos méthodes purement pooling sur la composante principale de Cora.



Étude expérimentale sur l'over-smoothing

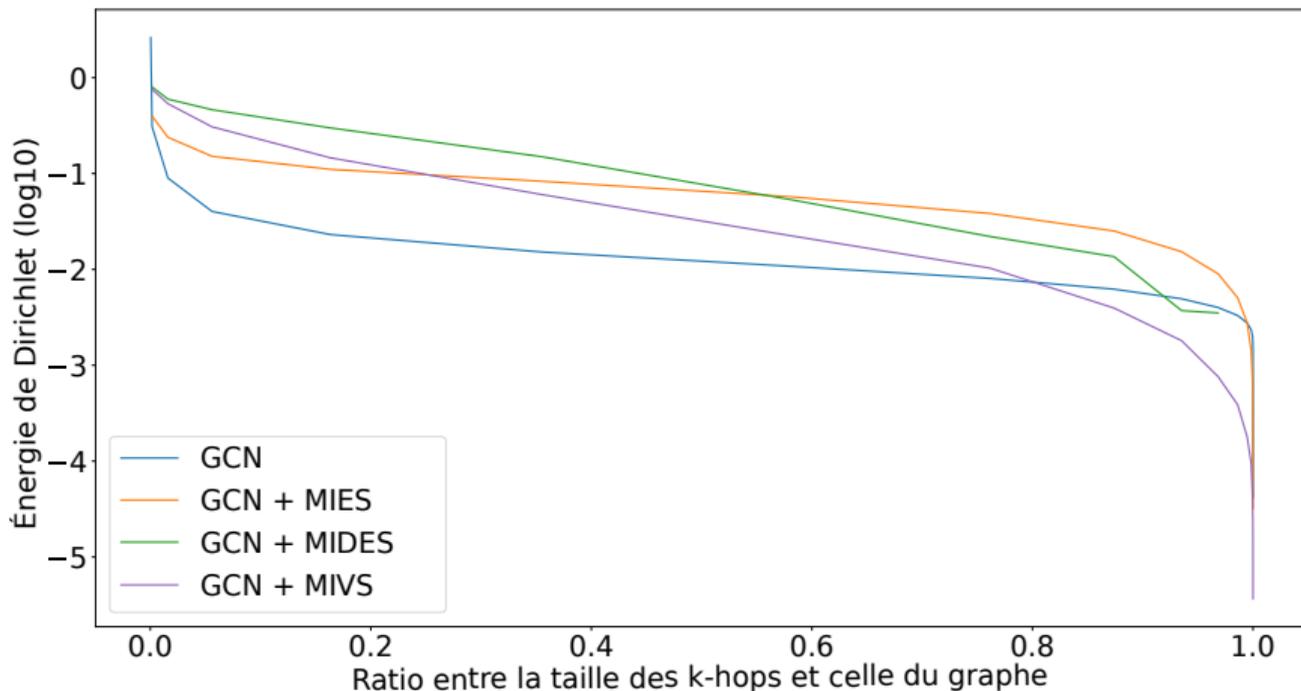


Figure – Énergie de Dirichlet calculée en alternant une convolution GCN et nos méthodes de pooling sur la composante principale de Cora.



Étude expérimentale sur l'over-smoothing

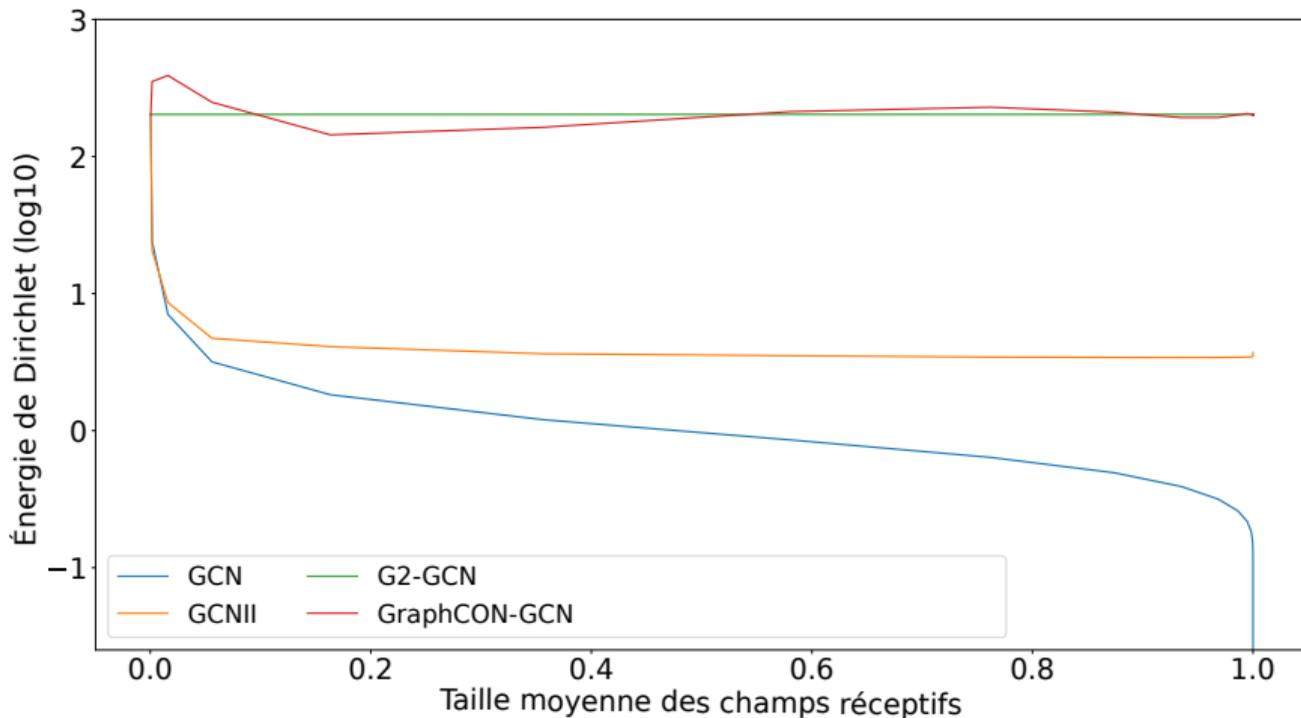


Figure – Comparaison sur la composante principale de Cora entre nos méthodes de pooling et des méthodes de convolutions sur graphes conçues pour réduire l'over-smoothing.



Étude expérimentale sur l'over-smoothing

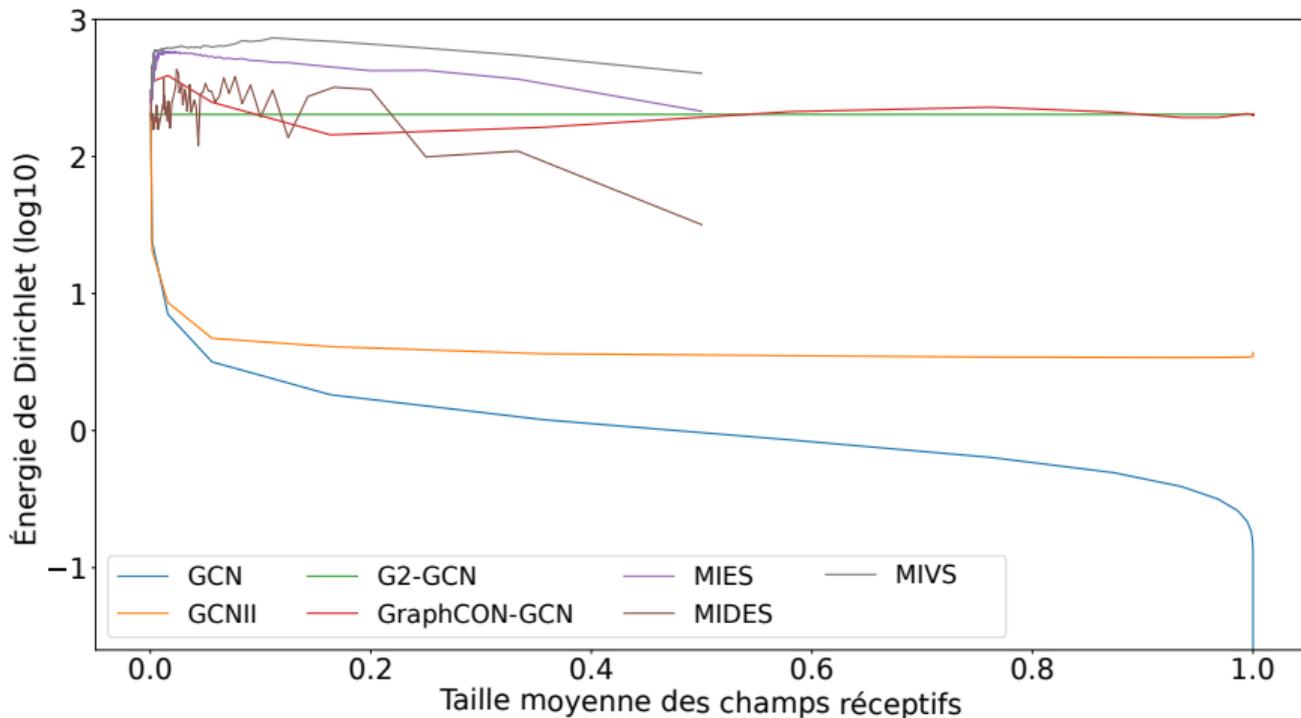


Figure – Comparaison sur la composante principale de Cora entre nos méthodes de pooling et des méthodes de convolutions sur graphes conçues pour réduire l'over-smoothing.



Classification de graphes

- Les étapes de convolution sont-elles complémentaires vis-à-vis des étapes de pooling sur la tâche de classification de graphes ?
- Est-ce que nos approches de pooling sont compétitives vis-à-vis de l'état de l'art ?



Résultats de nos méthodes de pooling sur la tâche de classification

Nos Méthodes	D&D	NCI109	REDDIT-BINARY	REDDIT-5K	REDDIT-12K
Baseline	76.29 ± 2.33	73.37 ± 1.90	87.07 ± 4.72	52.34 ± 2.97	47.45 ± 2.89
Lin+MIVSPool	74.11 ± 2.69	70.13 ± 2.04	86.85 ± 5.17	53.12 ± 1.39	45.96 ± 2.95
Lin+MIESPpool	75.95 ± 3.14	68.86 ± 2.78	87.38 ± 4.33	52.74 ± 1.56	45.79 ± 2.95
Lin+MIESCUTPool	75.72 ± 2.77	68.95 ± 1.92	83.62 ± 6.83	53.42 ± 1.11	46.39 ± 3.37
Lin+MIDESPpool	75.75 ± 1.89	66.59 ± 2.06	85.77 ± 5.21	52.64 ± 1.33	44.89 ± 3.08

Table – Précisions moyennes de nos méthodes pour la tâche de classification de graphes.



Résultats de nos méthodes de pooling sur la tâche de classification

Nos Méthodes	D&D	NCI109	REDDIT-BINARY	REDDIT-5K	REDDIT-12K
Baseline	76.29 ± 2.33	73.37 ± 1.90	87.07 ± 4.72	52.34 ± 2.97	47.45 ± 2.89
Lin+MIVSPool	74.11 ± 2.69	70.13 ± 2.04	86.85 ± 5.17	53.12 ± 1.39	45.96 ± 2.95
Lin+MIESPpool	75.95 ± 3.14	68.86 ± 2.78	87.38 ± 4.33	52.74 ± 1.56	45.79 ± 2.95
Lin+MIESCUTPool	75.72 ± 2.77	68.95 ± 1.92	83.62 ± 6.83	53.42 ± 1.11	46.39 ± 3.37
Lin+MIESPpool	75.75 ± 1.89	66.59 ± 2.06	85.77 ± 5.21	52.64 ± 1.33	44.89 ± 3.08
GCN+MIVSPool	76.35 ± 2.09	72.09 ± 1.27	88.73 ± 4.43	52.17 ± 1.90	46.50 ± 3.04
GCN+MIESPpool	77.17 ± 2.33	73.43 ± 2.27	88.08 ± 4.55	54.40 ± 1.27	47.33 ± 3.37
GCN+MIESCUTPool	77.74 ± 2.85	71.81 ± 2.19	86.47 ± 4.57	53.45 ± 0.91	47.51 ± 3.05
GCN+MIESPpool	76.52 ± 2.21	71.49 ± 1.77	88.40 ± 4.74	53.62 ± 1.35	46.51 ± 3.40

Table – Précisions moyennes de nos méthodes pour la tâche de classification de graphes.



Nos Méthodes	D&D	NCI109	REDDIT-BINARY	REDDIT-5K	REDDIT-12K
Baseline	76.29 ± 2.33	73.37 ± 1.90	87.07 ± 4.72	52.34 ± 2.97	47.45 ± 2.89
GCN+gPool	75.61 ± 2.74	67.78 ± 1.61	84.37 ± 7.82	50.41 ± 1.44	44.00 ± 2.92
GCN+SagPool	76.15 ± 2.88	68.40 ± 1.59	85.63 ± 6.26	50.84 ± 0.93	44.64 ± 2.82
GCN+EdgePool	72.59 ± 3.59	72.69 ± 2.47	87.25 ± 4.78	48.90 ± 2.12	43.25 ± 2.45
Notre meilleure méthode	77.74 ± 2.85 (GCN+MIESCut-Pool)	73.43 ± 2.27 (GCN+MIESPool)	88.73 ± 4.43 (GCN+MIVSPool)	54.40 ± 1.27 (GCN+MIESPool)	47.51 ± 3.05 (GCN+MIESCut-Pool)

Table – Précisions moyennes de classification de diverses méthodes de pooling par rapport à notre meilleure méthode.



Table des matières

- 1 Introduction et contexte
- 2 État de l'art
- 3 Impact du pooling sur l'over-smoothing et l'over-squashing
- 4 Méthodes proposées et ensembles indépendants maximaux
- 5 Résultats expérimentaux
- 6 Conclusion**



Contributions

- Nouvelle taxonomie : exactement un, au moins un, au plus un
- Étude de l'impact du pooling sur l'over-squashing et l'over-smoothing
 - Réduction de l'over-squashing
 - Réduction de l'over-smoothing
- Famille de méthodes exactement un basé sur les MIS
 - MIVSPool
 - MIESPool, MIESCutPool, MIDESPool

Perspectives

- Opération de réduction (RED) plus complexe
- Développement de GNN à base de pooling



Merci pour votre attention





MIESCUTPool : Maximal Independent Edge Set Pooling avec coupe

Famille de quatre méthodes exactement un basée sur les ensembles indépendants maximaux (MIS) :

MIVSPool

MIESPpool

MIESCUTPool

MIDESPool

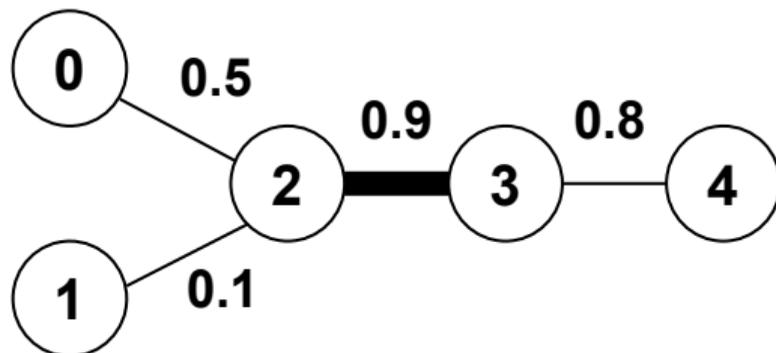


Maximal Independent Edge Set Pooling avec coupe (MIESCutPool)

- Début de l'algorithme comme MIESPool
- Ratio de préservation ≈ 0.5
 - Les sommets isolés sont rattachés au regroupement les plus similaires
 - Les arbres de longueur 3 sont divisés en deux regroupements

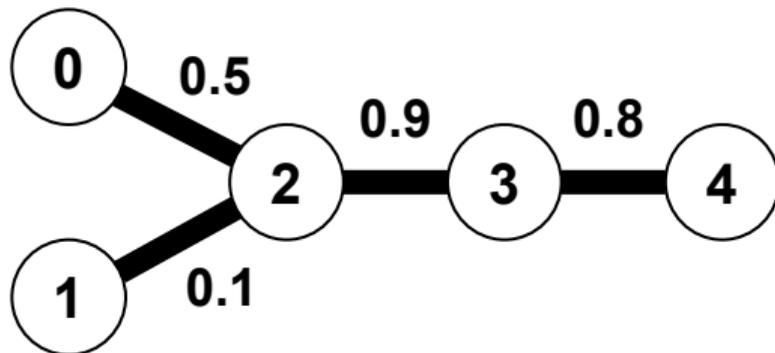


MIESCutPool : Maximal Independent Edge Set Pooling avec coupe



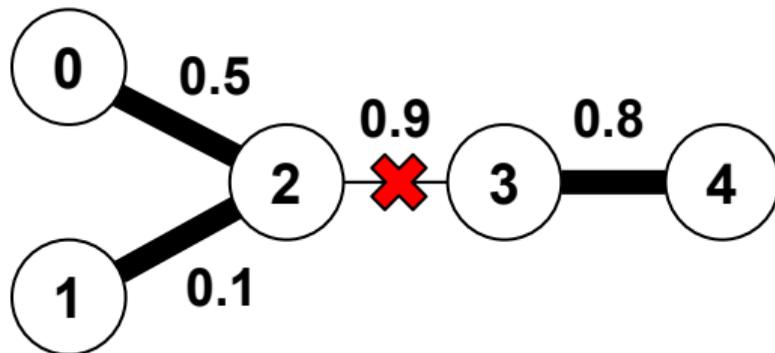


MIESCutPool : Maximal Independent Edge Set Pooling avec coupe



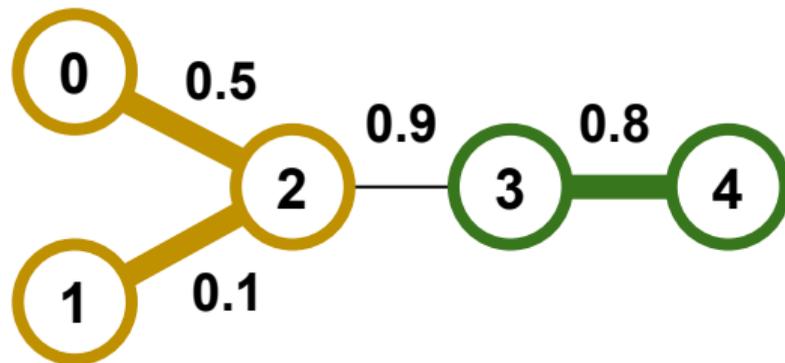


MIESCutPool : Maximal Independent Edge Set Pooling avec coupe





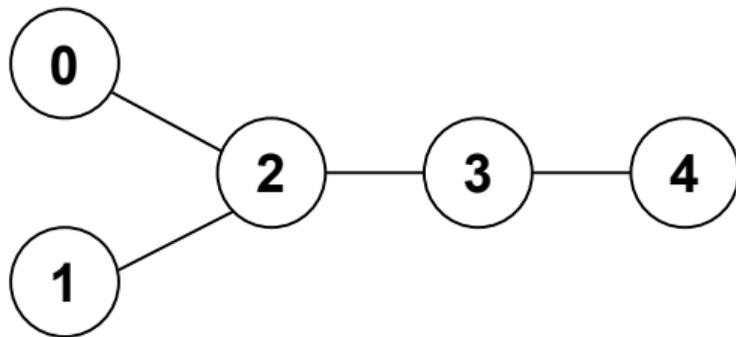
MIESCutPool : Maximal Independent Edge Set Pooling avec coupe



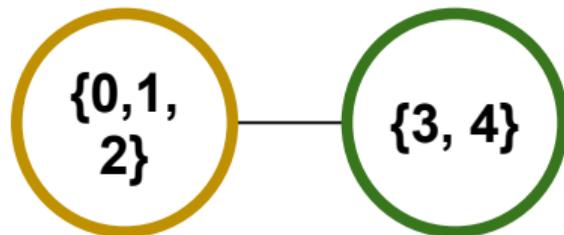


MIESCutPool : Maximal Independent Edge Set Pooling avec coupe

Avant



Après

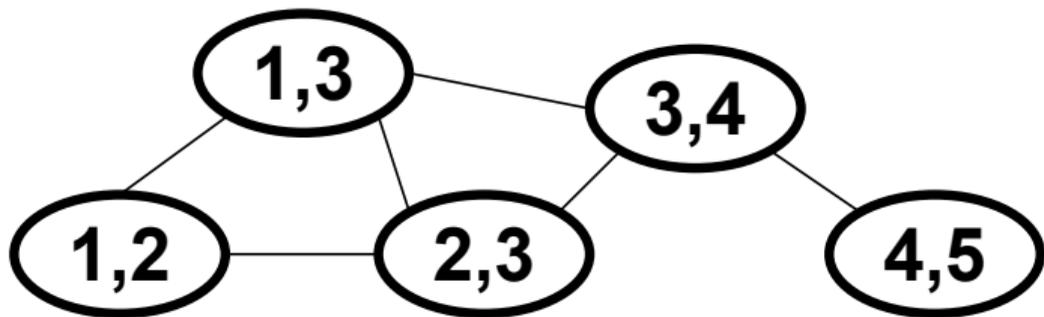
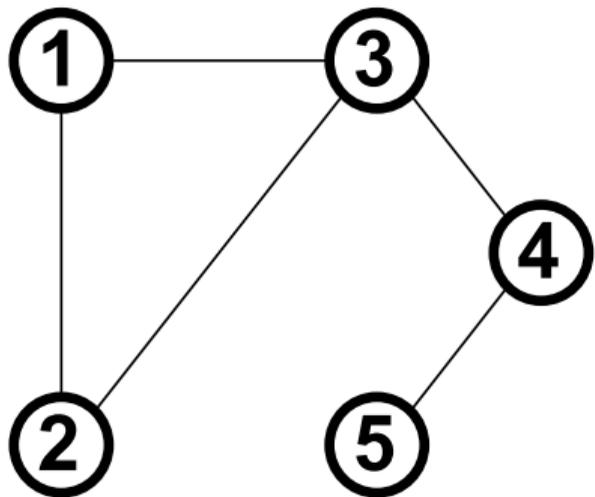




MIESCutPool : Maximal Independent Edge Set Pooling avec coupe

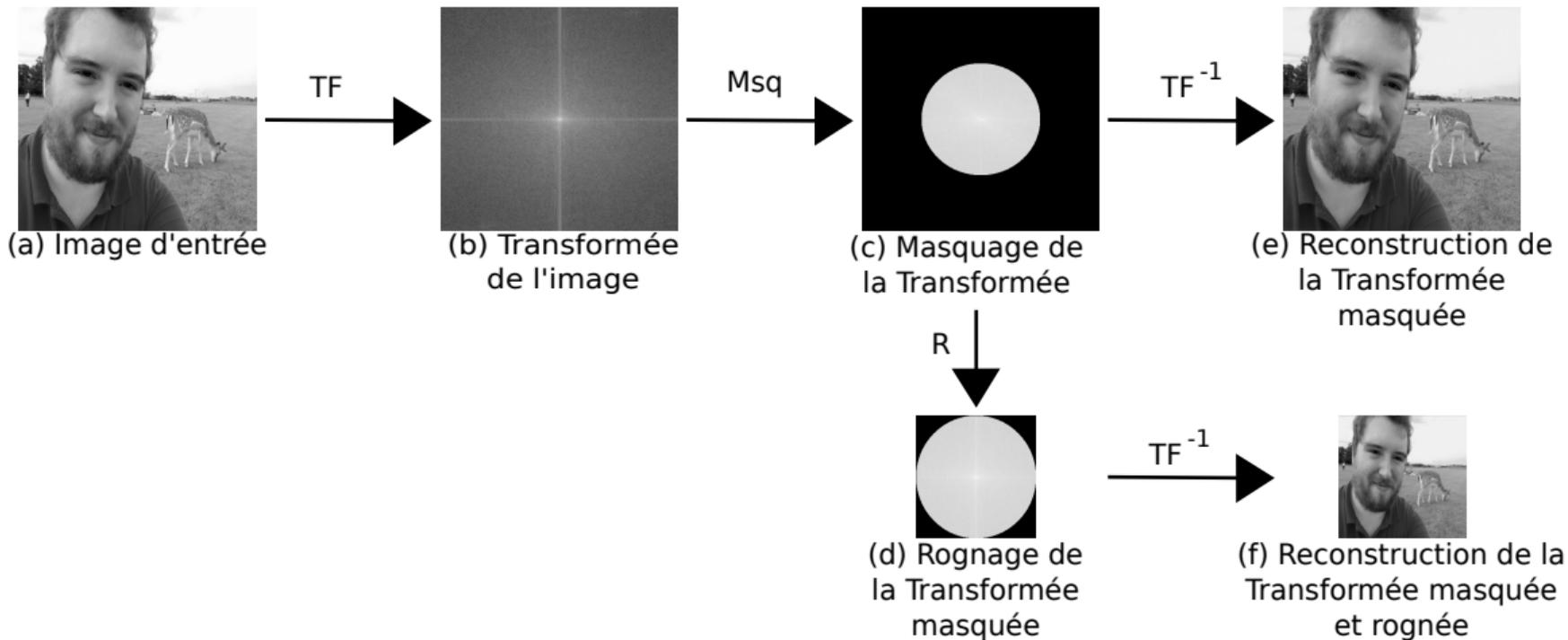
Maximal Independent Edge Set Pooling avec coupe (MIESCutPool)

- Ratio de préservation ≈ 0.5 ✓
⇒ 3 opérations nécessaires ✗
- Peut couper des sommets similaires ✗



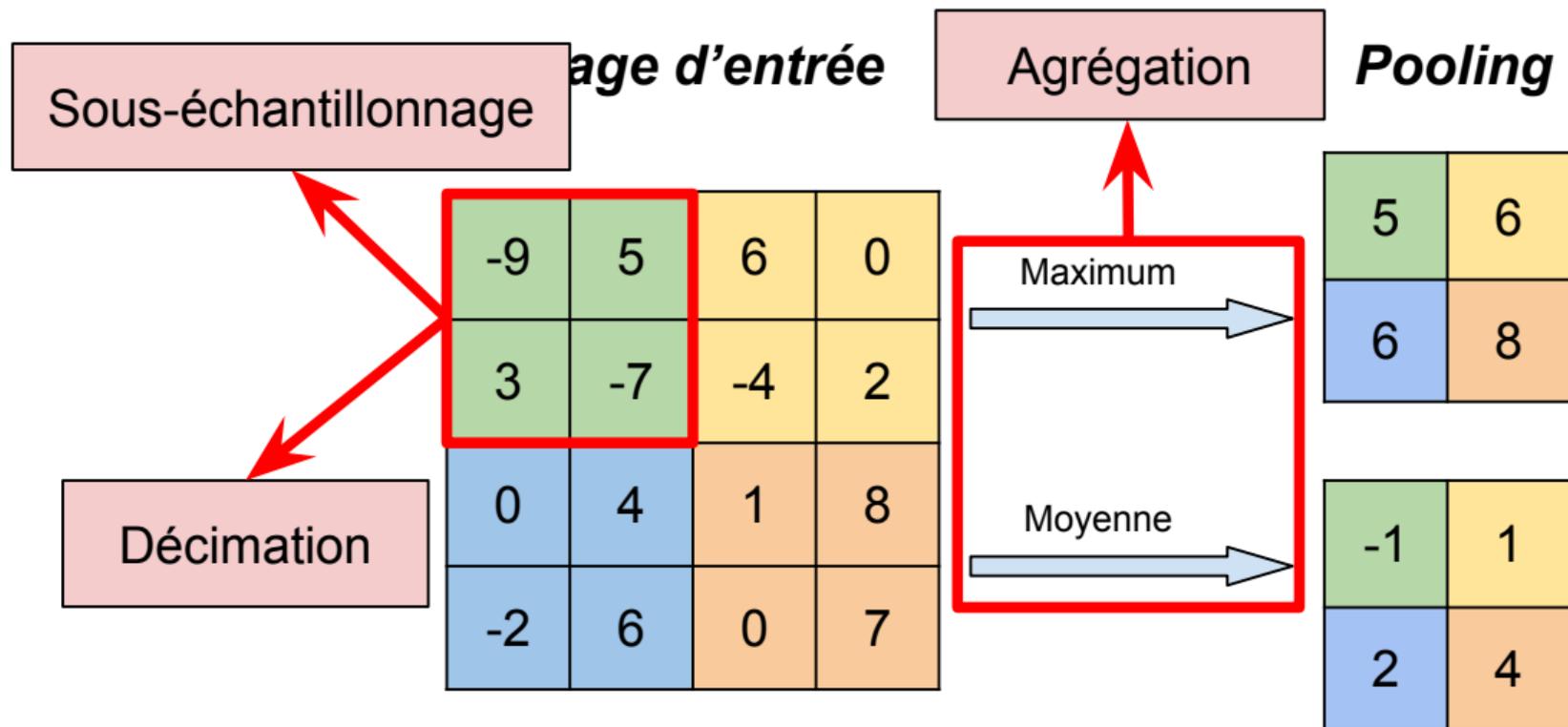


Redondance des convolutions sur les images





Détails du pooling sur image



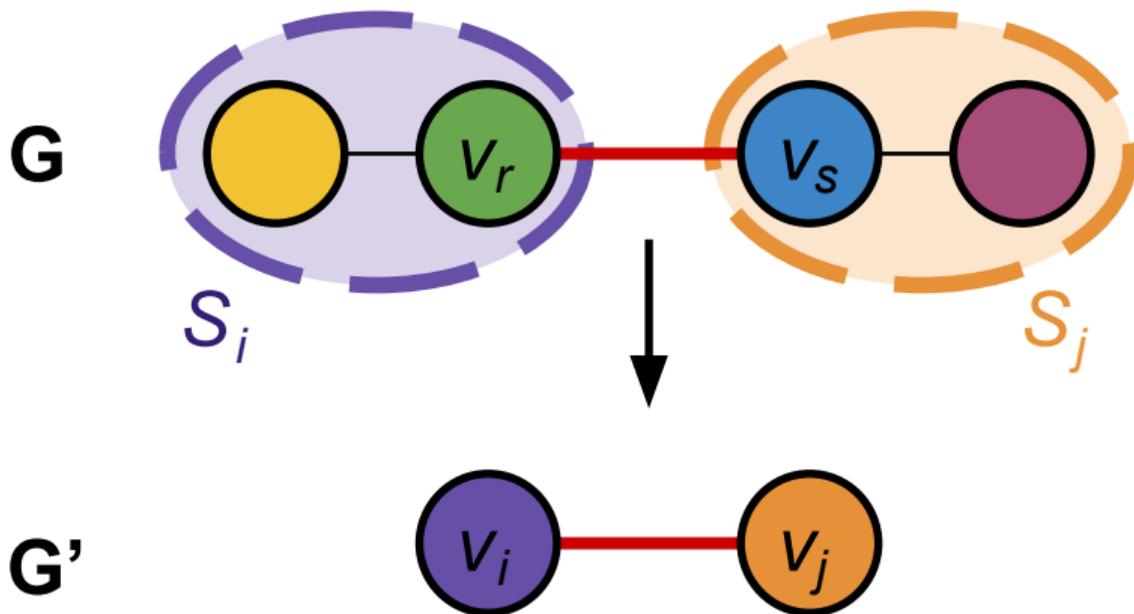


Lien entre spectral et spatial

$$\mathbf{C}^{(=)} \mathbf{U} \mathbf{F}(\Lambda) \mathbf{U}^T$$



Création des arêtes du graphe réduit



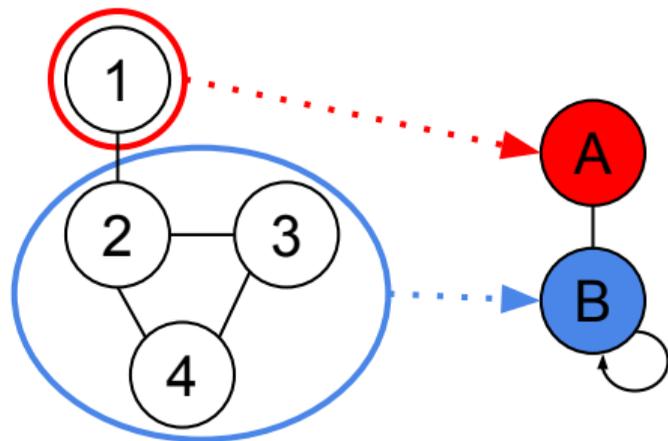


Création des arêtes du graphe réduit

$$\mathbf{A}'_{ij} = \sum_{r,s}^N \mathbf{A}_{rs} \mathbf{S}_{ri} \mathbf{S}_{sj}$$



Autre exemple pour la matrice S



$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

A

$$\begin{array}{l} \xrightarrow{X' = S^T X} \\ \xrightarrow{A' = S^T A S} \end{array}$$

$$\begin{pmatrix} 0 & 1 \\ 1 & 6 \end{pmatrix}$$

A'

$$\begin{array}{cc} \text{A} & \text{B} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \end{array}$$

Matrice S



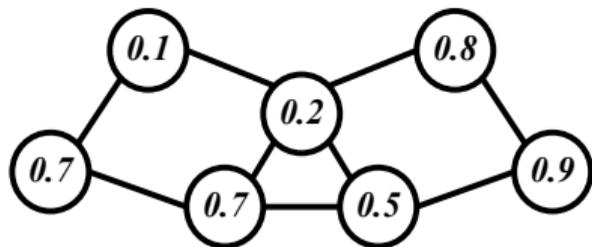
Équivariance du pooling hiérarchique

$$\begin{aligned}\mathbf{S}'^T \mathbf{X}' &= \mathbf{S}^T \mathbf{P}^T \mathbf{P} \mathbf{X} = \mathbf{S}^T \mathbf{X} \\ \mathbf{S}'^T \mathbf{A}' \mathbf{S}' &= \mathbf{S}^T \mathbf{P}^T \mathbf{P} \mathbf{A} \mathbf{P}^T \mathbf{P} \mathbf{S} = \mathbf{S}^T \mathbf{A} \mathbf{S}\end{aligned}$$

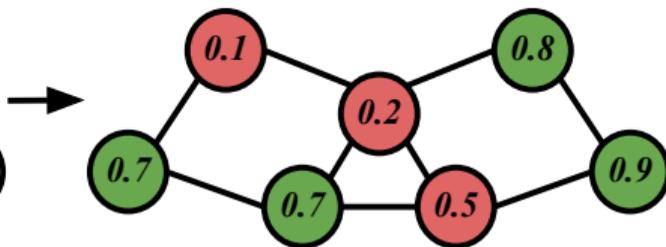


Schéma Top-k

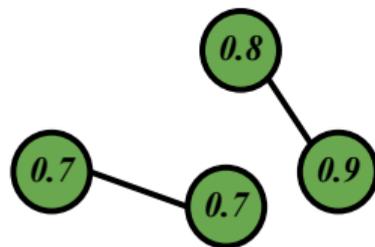
Calcul des scores par sommet



Application du Top-k et sélection des sommets survivants



Graphe réduit





Definition (Énergie de Dirichlet)

Soient \mathbf{X} la matrice des attributs des sommets et \mathbf{L} la matrice laplacienne associée au support convolutionnel \mathbf{C} de la convolution sur graphes. L'énergie de Dirichlet est donnée par :

$$\mathbf{E}(\mathbf{X}) = \text{Trace}(\mathbf{X}^T \mathbf{L} \mathbf{X}) = \sum_{v_i \in \mathcal{V}} \sum_{v_j \in \mathcal{N}(v_i)} \mathbf{C}_{ij} \|\mathbf{X}_i - \mathbf{X}_j\|_2^2$$

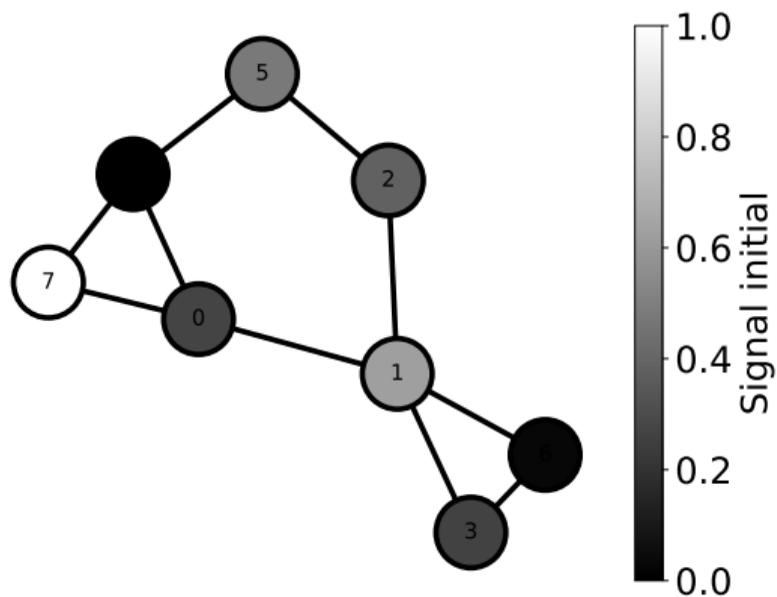
Sa forme normalisée est définie comme suit :

$$\tilde{\mathbf{E}}(\mathbf{X}) = \text{Trace}(\mathbf{X}^T \mathbf{L}_{\text{norm}} \mathbf{X}) = \sum_{v_i \in \mathcal{V}} \sum_{v_j \in \mathcal{N}(v_i)} \mathbf{C}_{ij} \left\| \frac{\mathbf{X}_i}{\sqrt{d_i}} - \frac{\mathbf{X}_j}{\sqrt{d_j}} \right\|_2^2$$

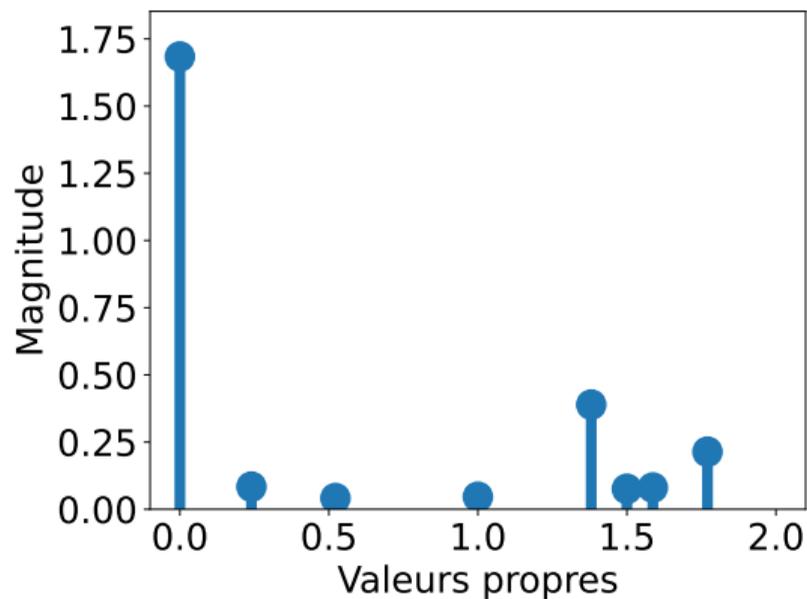
où \mathbf{L}_{norm} correspond à la matrice laplacienne symétrique normalisée associée au support convolutionnel \mathbf{C} et d_i est le degré du sommet i induit par la matrice \mathbf{C} .



Illustration de la Proposition 4.1.1



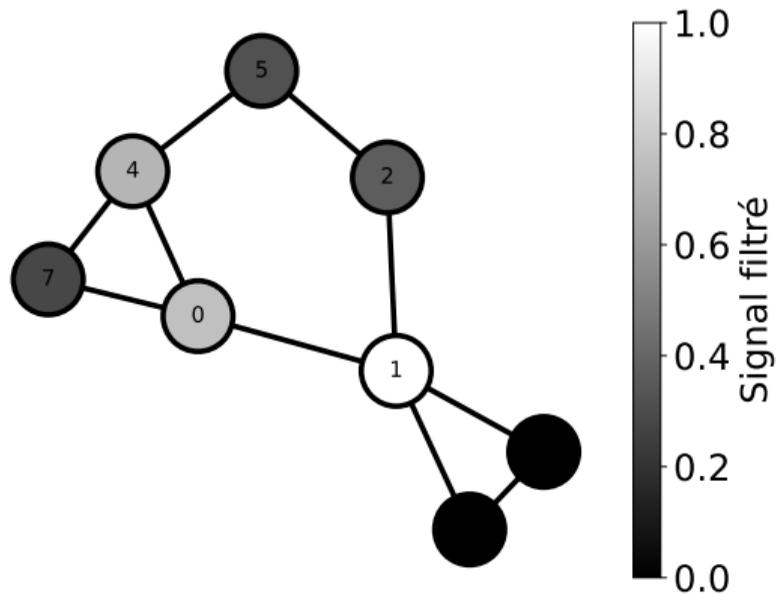
(a) (a) Graphe original



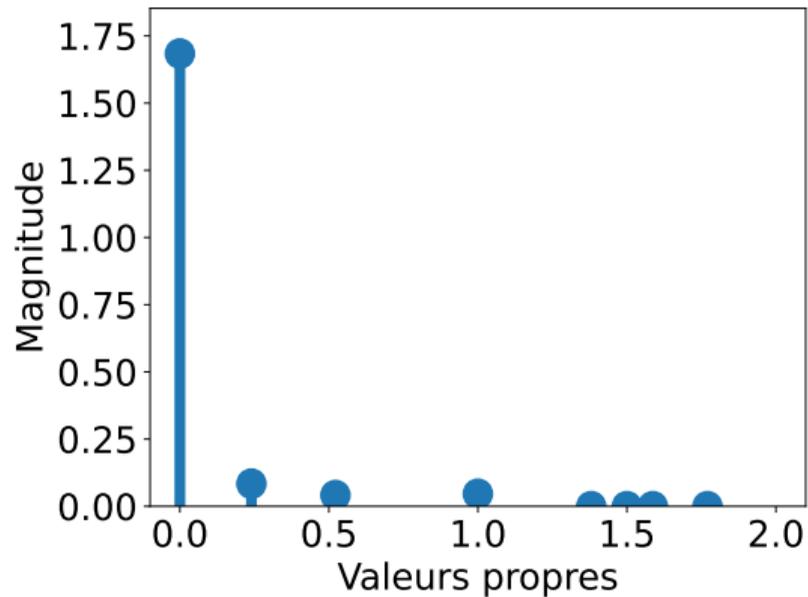
(b) (b) Spectre du signal initial



Illustration de la Proposition 4.1.1



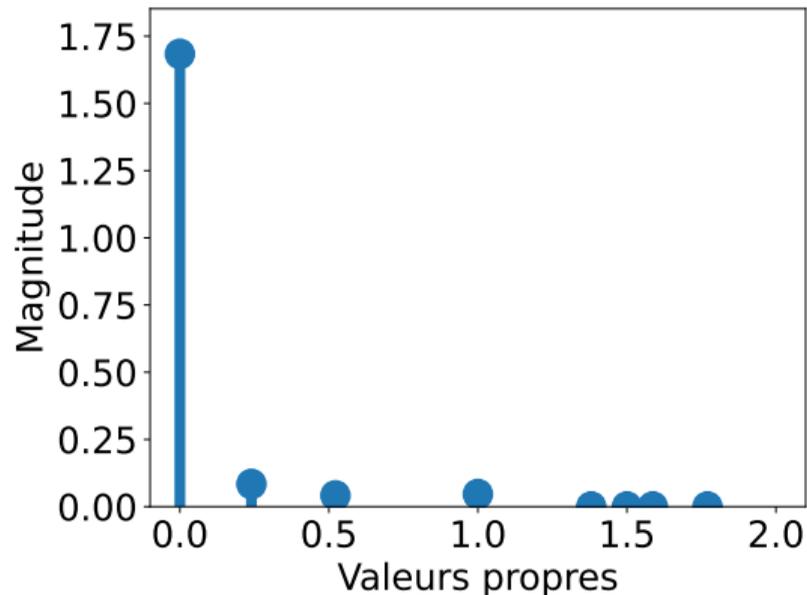
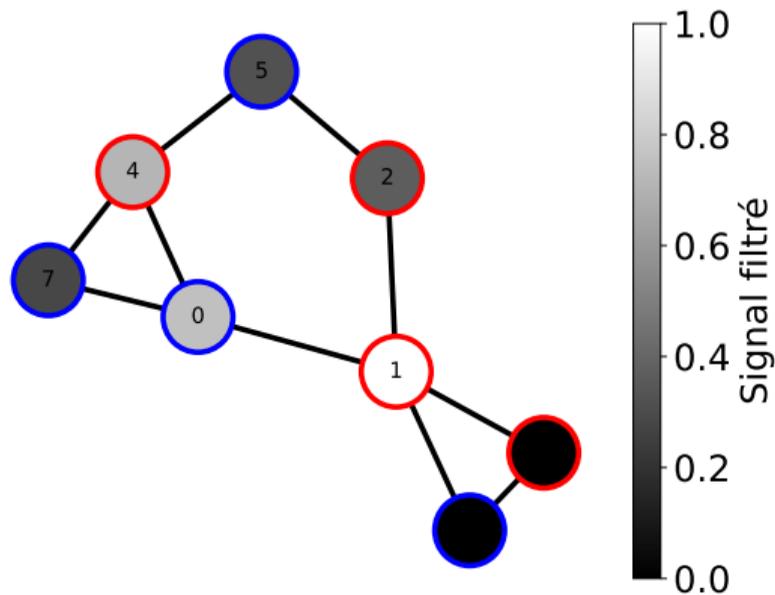
(a) (c) Signal limité en bande



(b) (d) Spectre du signal limité en bande



Illustration de la Proposition 4.1.1

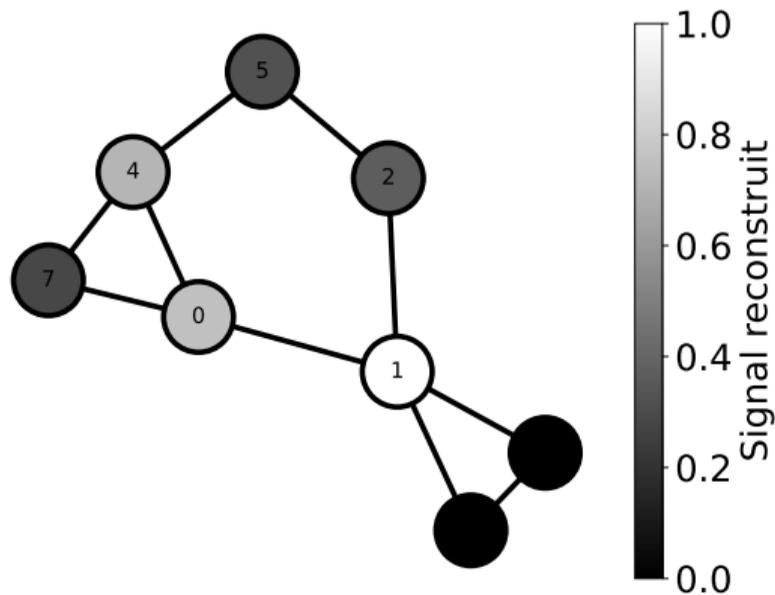


(a) (e) Signal limité en bande avec ensemble d'unicité

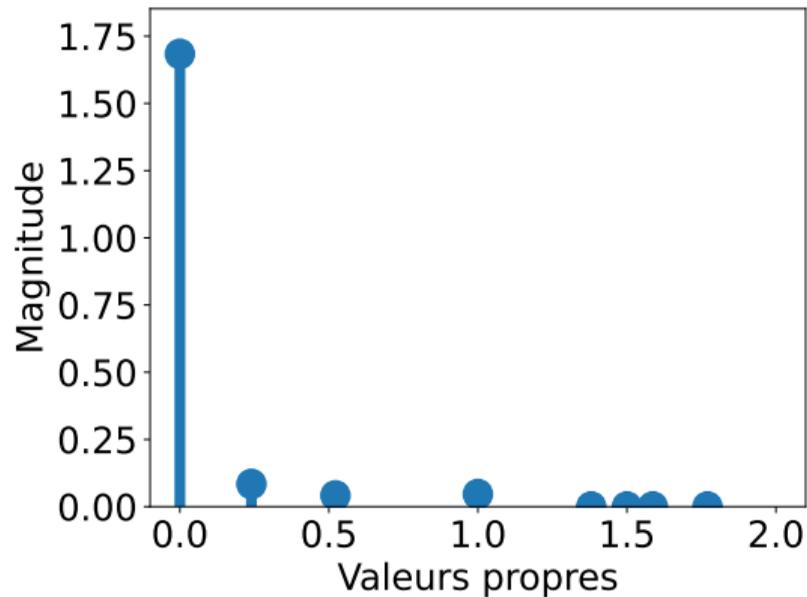
(b) (f) Spectre du signal limité en bande



Illustration de la Proposition 4.1.1



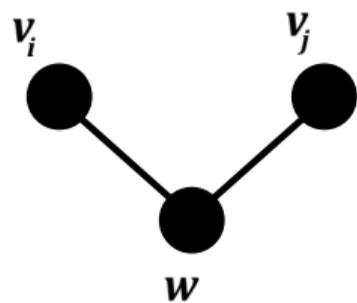
(a) (g) Signal reconstruit à partir de l'ensemble d'unicité



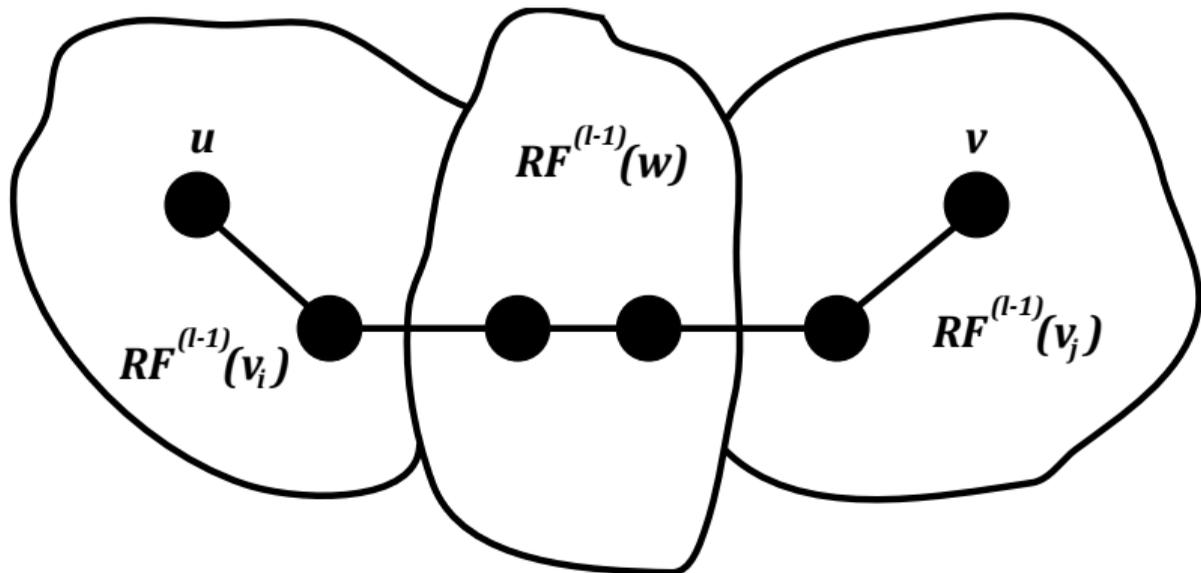
(b) (h) Spectral du signal reconstruit



Connexion entre les champs réceptifs à la couche 0



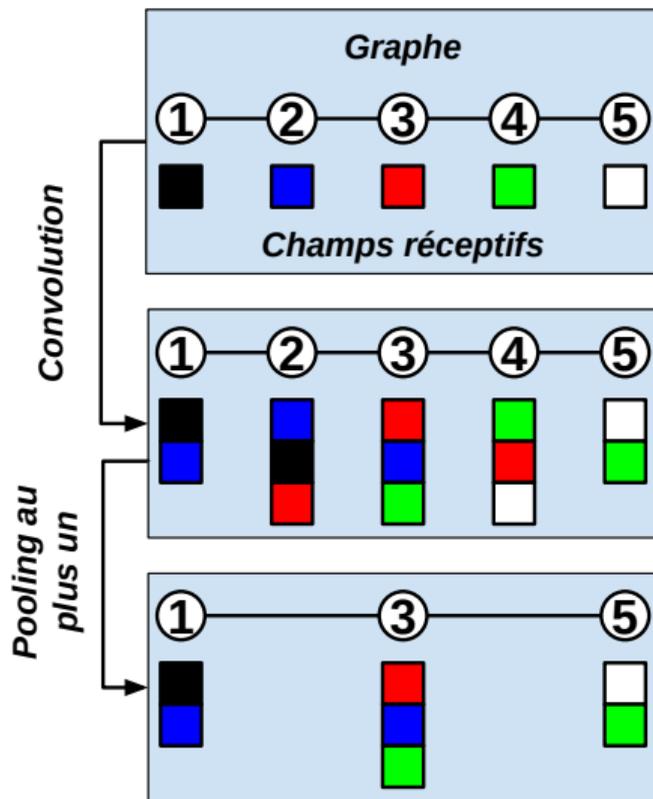
$G^{(l-1)}$



$G^{(0)}$

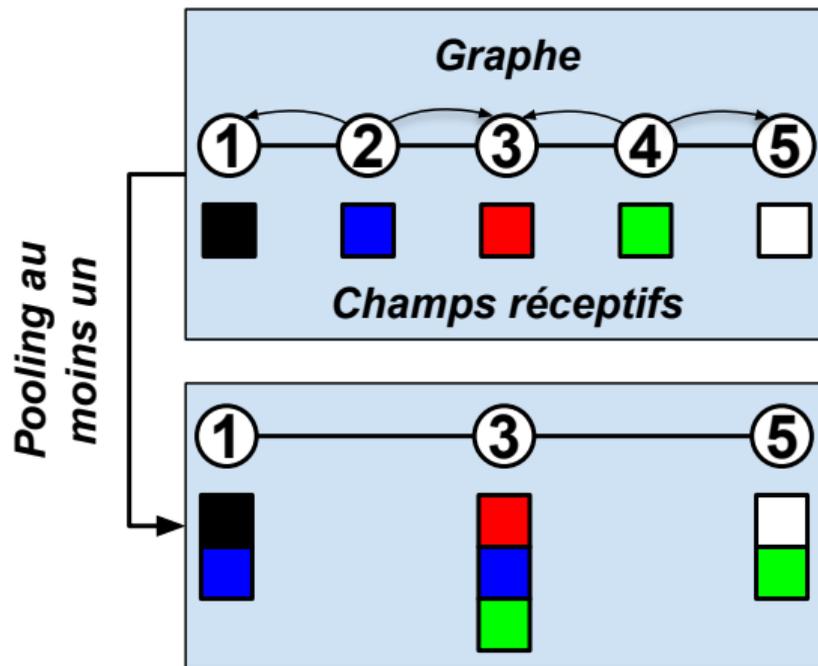


Méthode au moins un (Over-smoothing)





Méthode au moins un (Over-smoothing)





Méthode au moins un (Over-smoothing)

Contraintes / Résultats	Pooling exactement un	Pooling au plus un avec convolution	Pooling au moins un
Schéma de décimation	✓	✓	✓
GNN sans convolution	✓		✓
GNN alter conv/pool		✓	
Partition des sommets	✓		
Matrice d'adjacence diagonale non nulle		✓	✓
Restriction fonc. activa°		✓	✓
Over-smoothing	Insensible	Réduit	Réduit
Over-squashing réduit	✓	✓	✓



$$\mathbf{G}^{(l)} = GNN(\mathbf{X}^{(l)}, \mathbf{A}^{(l)})$$
$$s^{(l)} = \max_{j \in \{1, \dots, p\}} \left| \frac{\mathbf{G}^{(l)} \cdot \mathbf{W}_{:,j}^{(l)}}{\|\mathbf{G}^{(l)}\| \cdot \|\mathbf{W}_{:,j}^{(l)}\|} \right|$$

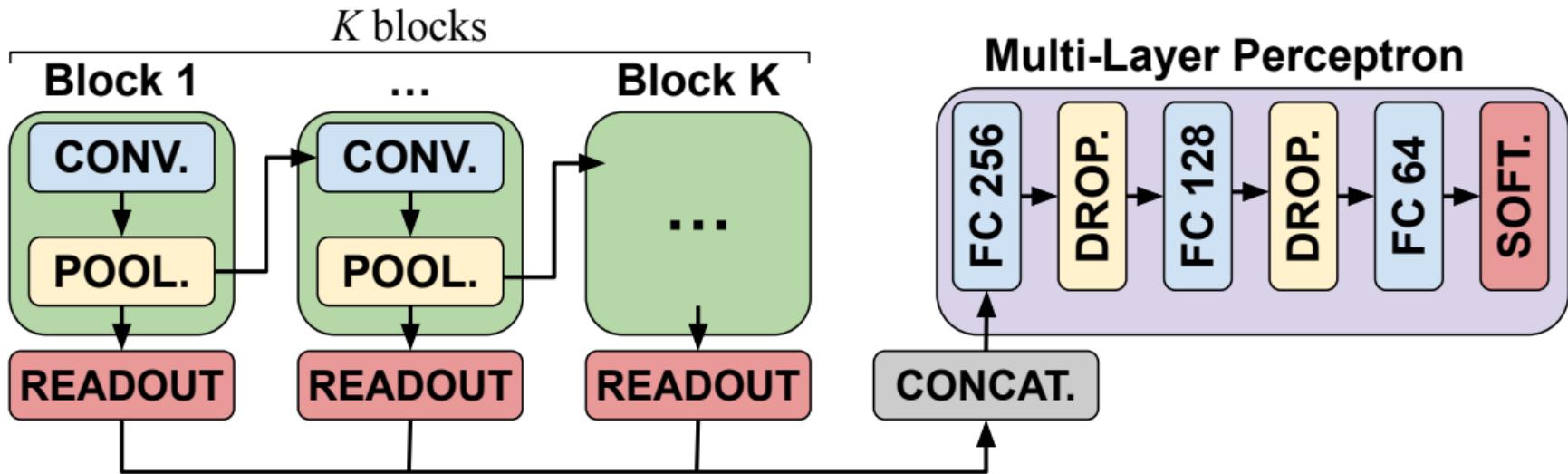


Jeux de données	Nombres de graphes	Nombre de classes	Nombre moyen de sommets	Nombre moyen d'arêtes	Degré moyen	Diamètre moyen
D&D	1178	2	284 ± 272	716 ± 694	4.98 ± 0.59	19.90 ± 7.72
NCI109	4127	2	30 ± 14	32 ± 15	2.16 ± 0.12	13.12 ± 5.31
REDDIT-BINARY	2000	2	430 ± 554	498 ± 623	2.34 ± 0.31	9.72 ± 3.13
REDDIT-5K	4999	5	509 ± 453	595 ± 567	2.25 ± 0.20	11.96 ± 2.57
REDDIT-12K	11929	11	391 ± 429	457 ± 518	2.28 ± 0.24	10.90 ± 3.06



Hyper-paramètres

Hyper-paramètres	Valeurs
Taux d'apprentissage	10^{-3}
Diminution des poids (régularisation L_2)	10^{-4}
Taille des mini-lots	512
Patience	100
Nombre de blocs K	{1, 2, 3, 4, 5}
Dimension des couches intermédiaires	{64, 128}
Dropout	{0.2, 0.5}





Nombre moyen d'itérations de l'algorithme de Meer

Méthodes	Étapes	D&D	NCI109	REDDIT-BINARY	REDDIT-5K	REDDIT-12K
MIVS	1	4.11 ± 0.80	3.00 ± 0.79	2.89 ± 0.84	2.84 ± 0.60	2.64 ± 0.78
	2	2.87 ± 0.68	2.33 ± 0.64	2.56 ± 0.71	3.03 ± 0.99	2.36 ± 0.86
	3	2.30 ± 0.64	1.73 ± 0.56	2.11 ± 0.57	2.25 ± 0.45	1.52 ± 0.72
	4	1.51 ± 0.55	1.22 ± 0.42	1.95 ± 0.55	2.18 ± 0.40	1.30 ± 0.49
	5	1.17 ± 0.37	1.03 ± 0.18	1.76 ± 0.55	1.66 ± 0.53	1.34 ± 0.48
MIES	1	3.81 ± 0.69	2.22 ± 0.61	2.71 ± 0.83	2.94 ± 0.86	2.94 ± 0.97
	2	3.15 ± 0.69	2.03 ± 0.58	2.27 ± 0.86	2.90 ± 1.01	3.11 ± 0.97
	3	2.94 ± 0.68	1.67 ± 0.59	2.29 ± 0.84	3.02 ± 0.95	2.72 ± 0.97
	4	2.78 ± 0.75	1.42 ± 0.53	2.29 ± 0.89	2.75 ± 0.97	2.39 ± 0.97
	5	2.84 ± 0.83	1.17 ± 0.39	2.37 ± 1.05	2.39 ± 0.91	2.38 ± 0.97



Nombre moyen d'itérations de l'algorithme de Meer

Méthodes	Étapes	D&D	NCI109	REDDIT-BINARY	REDDIT-5K	REDDIT-12K
MIESCut	1	3.71 ± 0.67	2.25 ± 0.61	2.63 ± 0.88	3.08 ± 0.89	3.01 ± 0.90
	2	3.12 ± 0.67	2.02 ± 0.67	2.54 ± 0.97	3.10 ± 1.09	2.71 ± 1.09
	3	2.80 ± 0.71	1.31 ± 0.50	2.50 ± 1.21	3.02 ± 1.26	2.54 ± 1.11
	4	2.49 ± 0.79	1.05 ± 0.22	2.36 ± 1.22	2.75 ± 1.39	2.37 ± 1.20
	5	2.02 ± 0.81	1.01 ± 0.07	1.84 ± 0.94	1.97 ± 1.00	1.83 ± 0.90
MIDES	1	3.75 ± 0.63	2.20 ± 0.57	2.85 ± 0.88	3.06 ± 0.80	3.21 ± 0.92
	2	3.35 ± 0.66	2.19 ± 0.57	2.81 ± 0.92	3.01 ± 0.94	3.10 ± 0.95
	3	2.78 ± 0.71	1.57 ± 0.56	2.47 ± 0.91	2.71 ± 0.88	2.74 ± 0.96
	4	2.19 ± 0.71	1.20 ± 0.42	2.14 ± 0.86	2.35 ± 0.8	2.20 ± 0.75
	5	1.57 ± 0.64	1.17 ± 0.38	1.75 ± 0.72	1.96 ± 0.65	1.64 ± 0.68



Évolution densité/degré/diamètre D&D

Méthodes	Étapes	Diamètre moyen	Ratio des diamètres entre deux étapes
MIVSPool	Initial	19.90 ± 7.72	—
	Pooling 1	10.22 ± 3.97	1.95
	Pooling 2	5.08 ± 2.03	2.01
	Pooling 3	2.59 ± 1.16	1.96
	Pooling 4	0.86 ± 0.93	3.01
	Pooling 5	0.41 ± 0.75	2.10
MIESPool	Initial	19.90 ± 7.72	—
	Pooling 1	15.22 ± 5.83	1.31
	Pooling 2	11.72 ± 4.56	1.30
	Pooling 3	8.35 ± 3.20	1.40
	Pooling 4	5.93 ± 2.30	1.41
	Pooling 5	4.37 ± 1.68	1.36



Évolution densité/degré/diamètre D&D

Méthodes	Étapes	Diamètre moyen	Ratio des diamètres entre deux étapes
MIESCutPool	Initial	19.90 ± 7.72	—
	Pooling 1	13.43 ± 5.21	1.48
	Pooling 2	8.84 ± 3.50	1.52
	Pooling 3	5.23 ± 2.19	1.69
	Pooling 4	3.00 ± 1.47	1.74
	Pooling 5	1.61 ± 1.07	1.86
MIDESPpool	Initial	19.90 ± 7.72	—
	Pooling 1	11.20 ± 4.37	1.78
	Pooling 2	6.76 ± 2.55	1.66
	Pooling 3	4.02 ± 1.65	1.68
	Pooling 4	2.44 ± 1.19	1.65
	Pooling 5	1.46 ± 1.08	1.67



Évolution densité/degré/diamètre NCI109

Méthodes	Étapes	Diamètre moyen	Ratio des diamètres entre deux étapes
MIVSPool	Initial	13.12 ± 5.31	—
	Pooling 1	6.77 ± 2.89	1.94
	Pooling 2	3.39 ± 1.57	2.00
	Pooling 3	1.54 ± 1.01	2.20
	Pooling 4	0.38 ± 0.65	4.06
	Pooling 5	0.07 ± 0.33	5.43
MIESPool	Initial	13.12 ± 5.31	—
	Pooling 1	8.54 ± 3.62	1.54
	Pooling 2	5.60 ± 2.44	1.53
	Pooling 3	3.64 ± 1.72	1.54
	Pooling 4	2.32 ± 1.32	1.57
	Pooling 5	1.33 ± 1.07	1.74



Évolution densité/degré/diamètre NCI109

Méthodes	Étapes	Diamètre moyen	Ratio des diamètres entre deux étapes
MIESCutPool	Initial	13.12 ± 5.31	—
	Pooling 1	6.51 ± 3.04	2.02
	Pooling 2	3.29 ± 1.77	1.98
	Pooling 3	1.09 ± 1.00	3.02
	Pooling 4	0.10 ± 0.34	10.9
	Pooling 5	0.00 ± 0.03	—
MIDESPpool	Initial	13.12 ± 5.31	—
	Pooling 1	6.45 ± 2.82	2.03
	Pooling 2	3.74 ± 1.88	1.72
	Pooling 3	1.87 ± 1.25	2.00
	Pooling 4	0.64 ± 0.85	2.92
	Pooling 5	0.13 ± 0.41	4.92

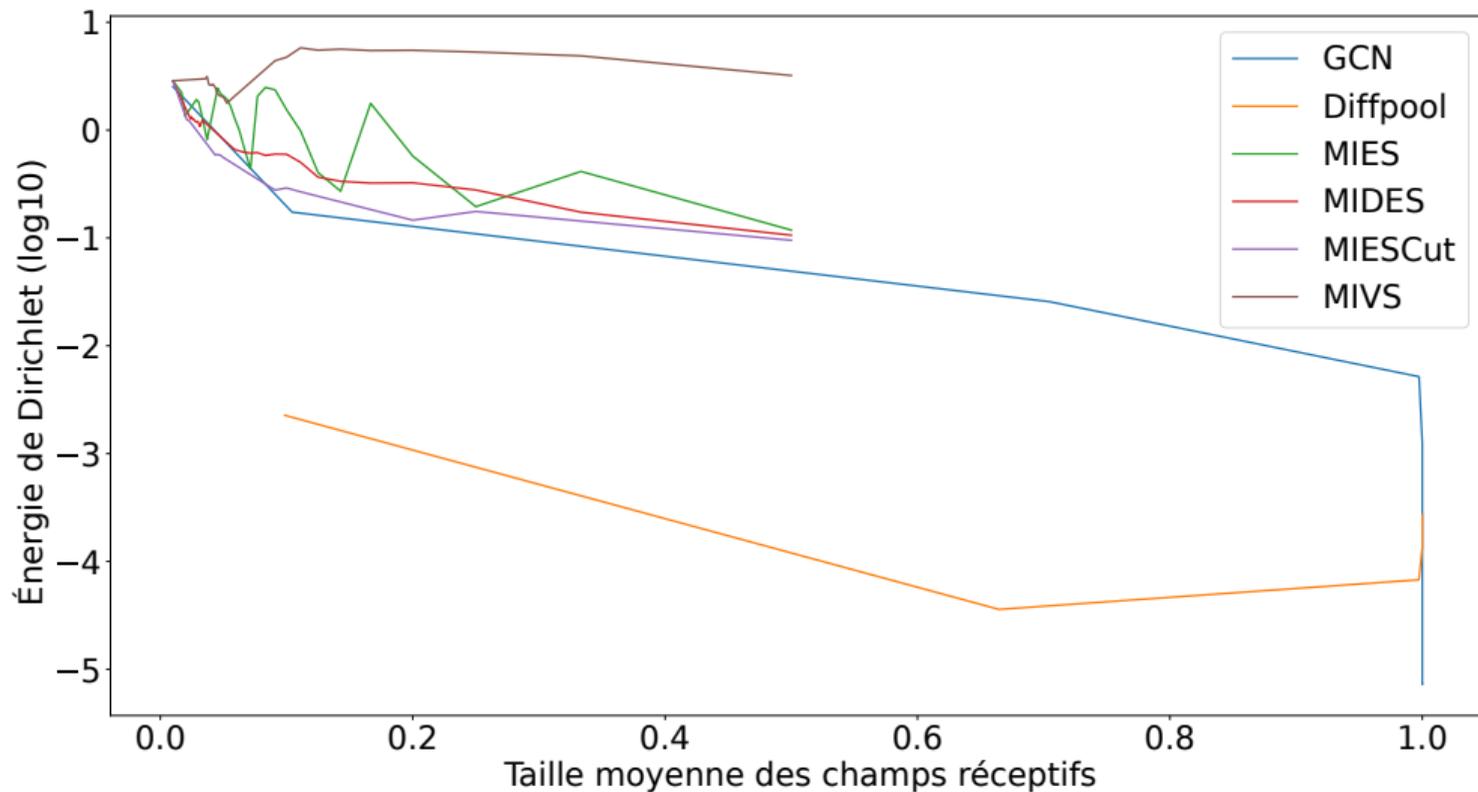


Classification - Méthodes au plus un

Méthodes	D&D	NCI109	REDDIT-BINARY	REDDIT-5K	REDDIT-12K
GCN+MIVS ⁺	74.79 ± 2.68	71.75 ± 1.62	88.15 ± 3.99	53.05 ± 1.44	46.94 ± 2.95
GCN+MIES ⁺	74.77 ± 2.51	59.32 ± 3.69	64.63 ± 11.50	32.81 ± 2.40	18.64 ± 2.21
GCN+MIESC ⁺	75.28 ± 2.42	61.29 ± 3.36	59.75 ± 12.01	41.05 ± 2.38	24.50 ± 3.81
GCN+MIDES ⁺	78.56 ± 3.48	53.95 ± 4.83	51.53 ± 6.17	33.90 ± 3.01	20.46 ± 1.29
Notre meilleure méthode	77.74 ± 2.85 (GCN+MIESC ⁺ Pool)	73.43 ± 2.27 (GCN+MIESP ⁺ Pool)	88.73 ± 4.43 (GCN+MIVSP ⁺ Pool)	54.40 ± 1.27 (GCN+MIESP ⁺ Pool)	47.51 ± 3.05 (GCN+MIESC ⁺ Pool)

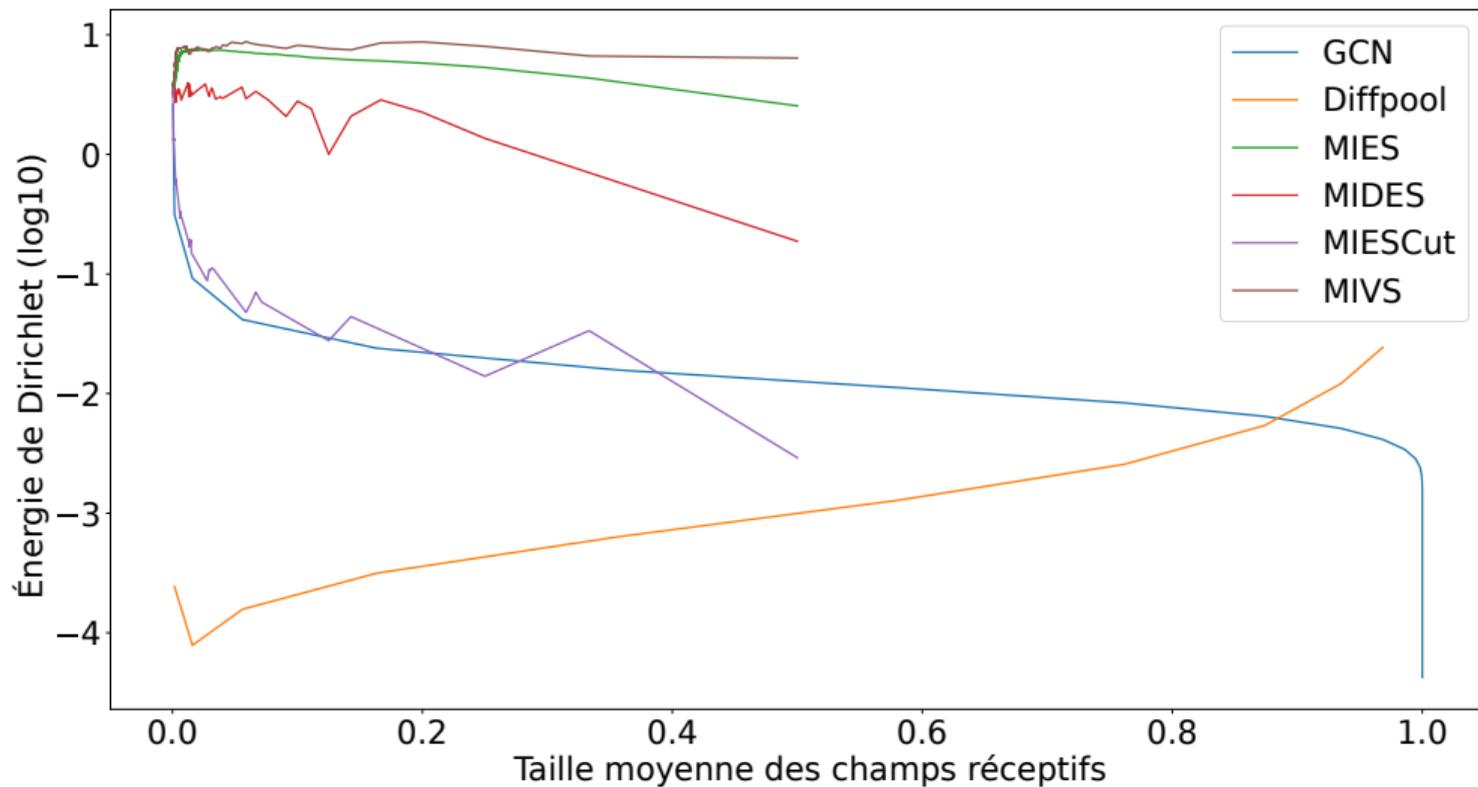


GCN et méthodes purement pooling



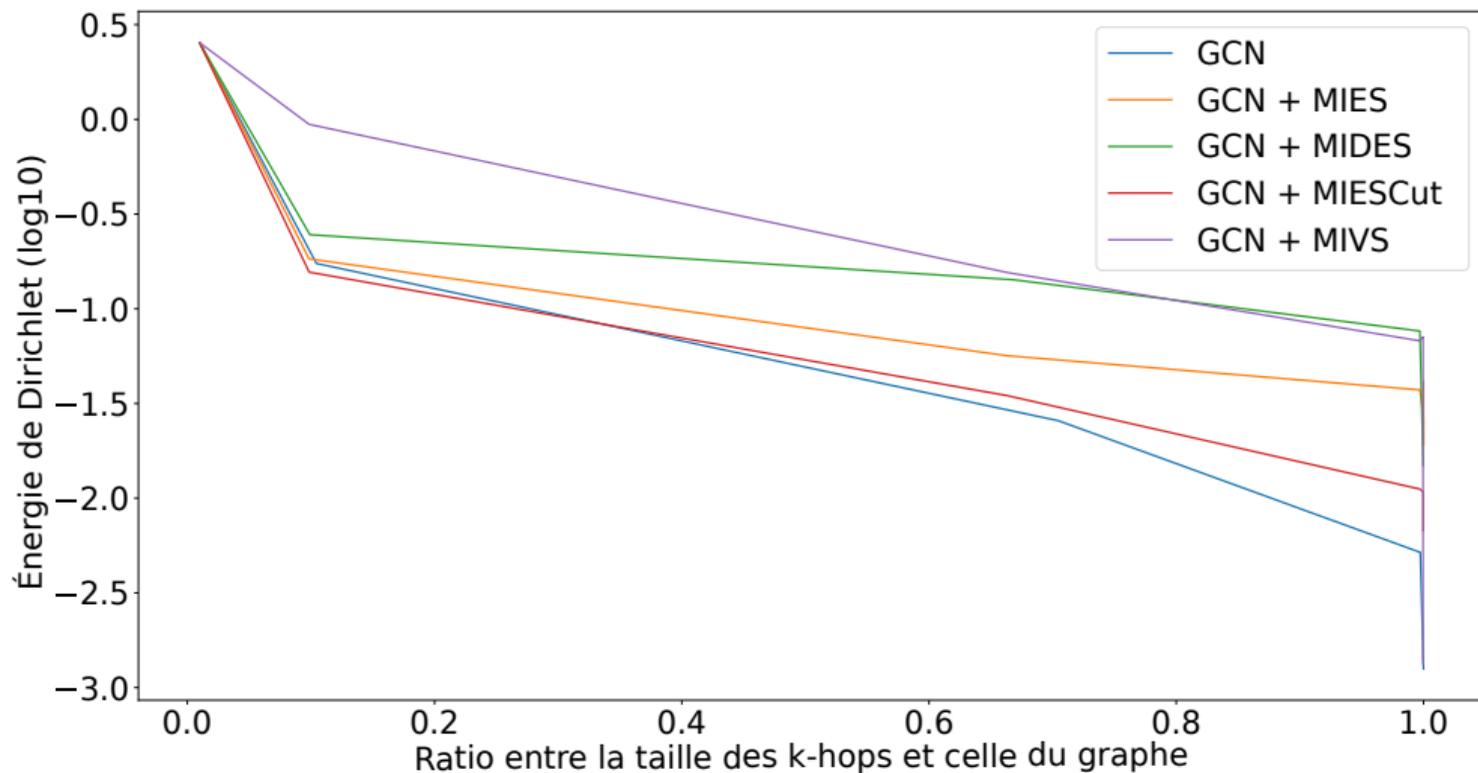


GCN et méthodes purement pooling



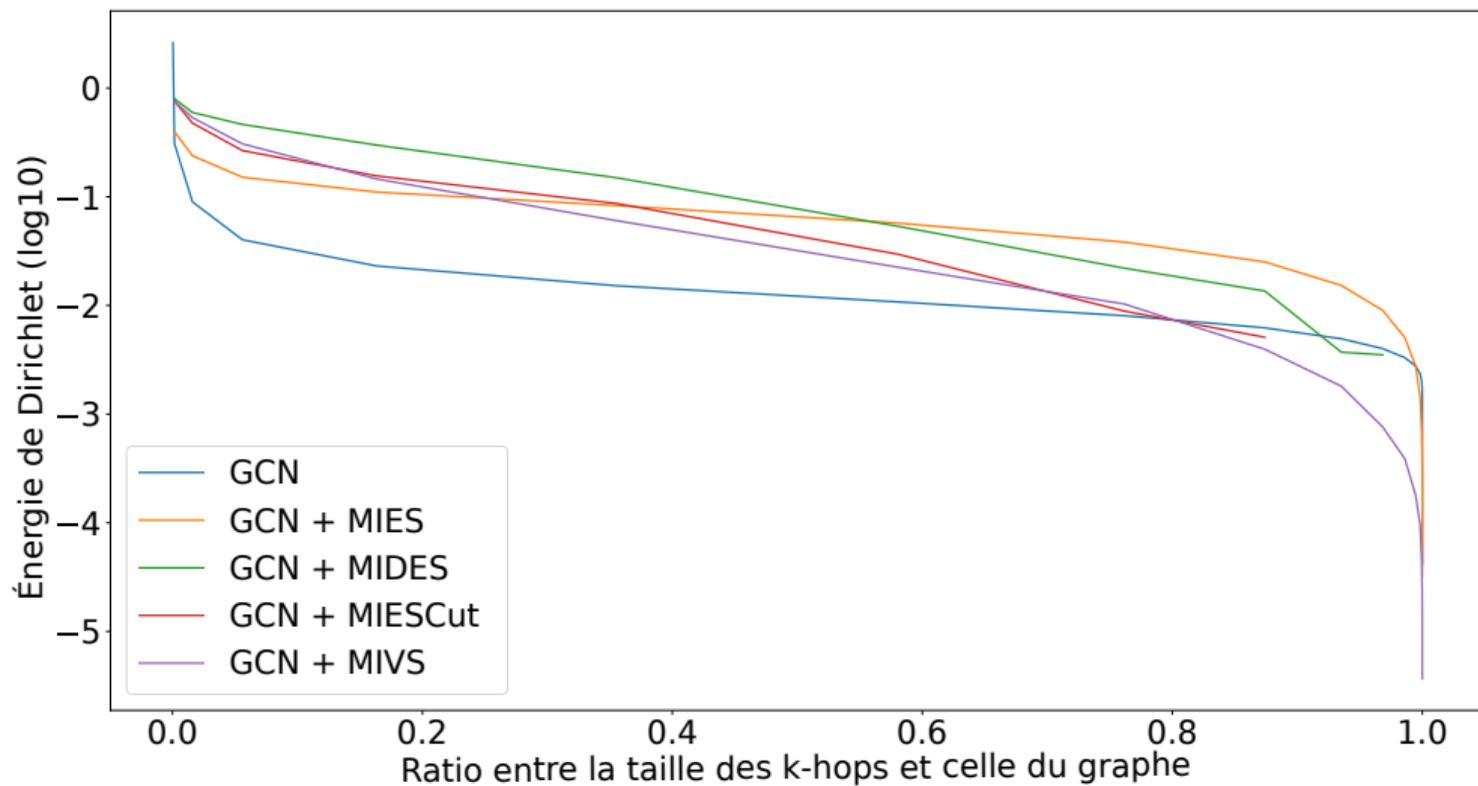


GCN et alternance GCN/pooling =1



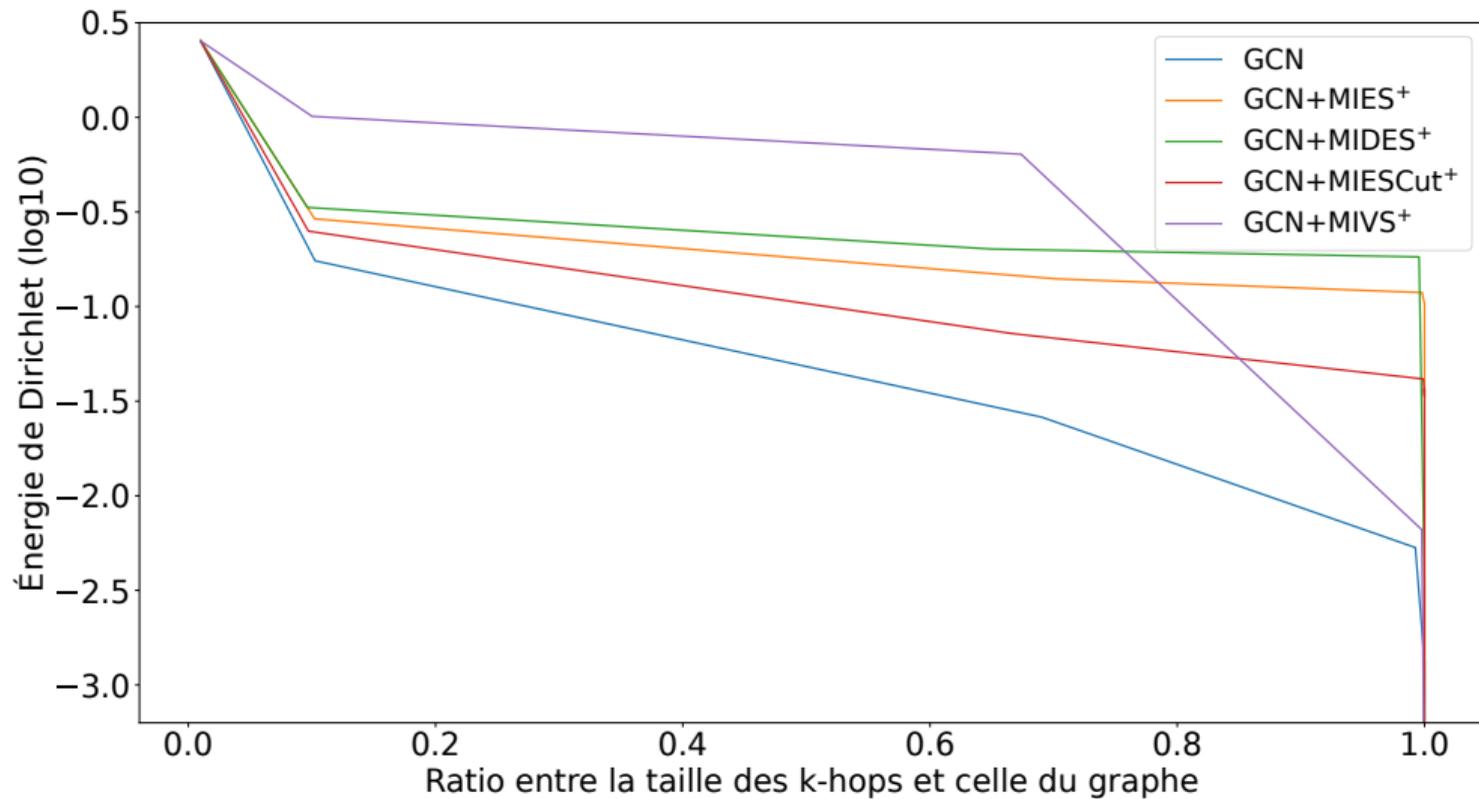


GCN et alternance GCN/pooling =1



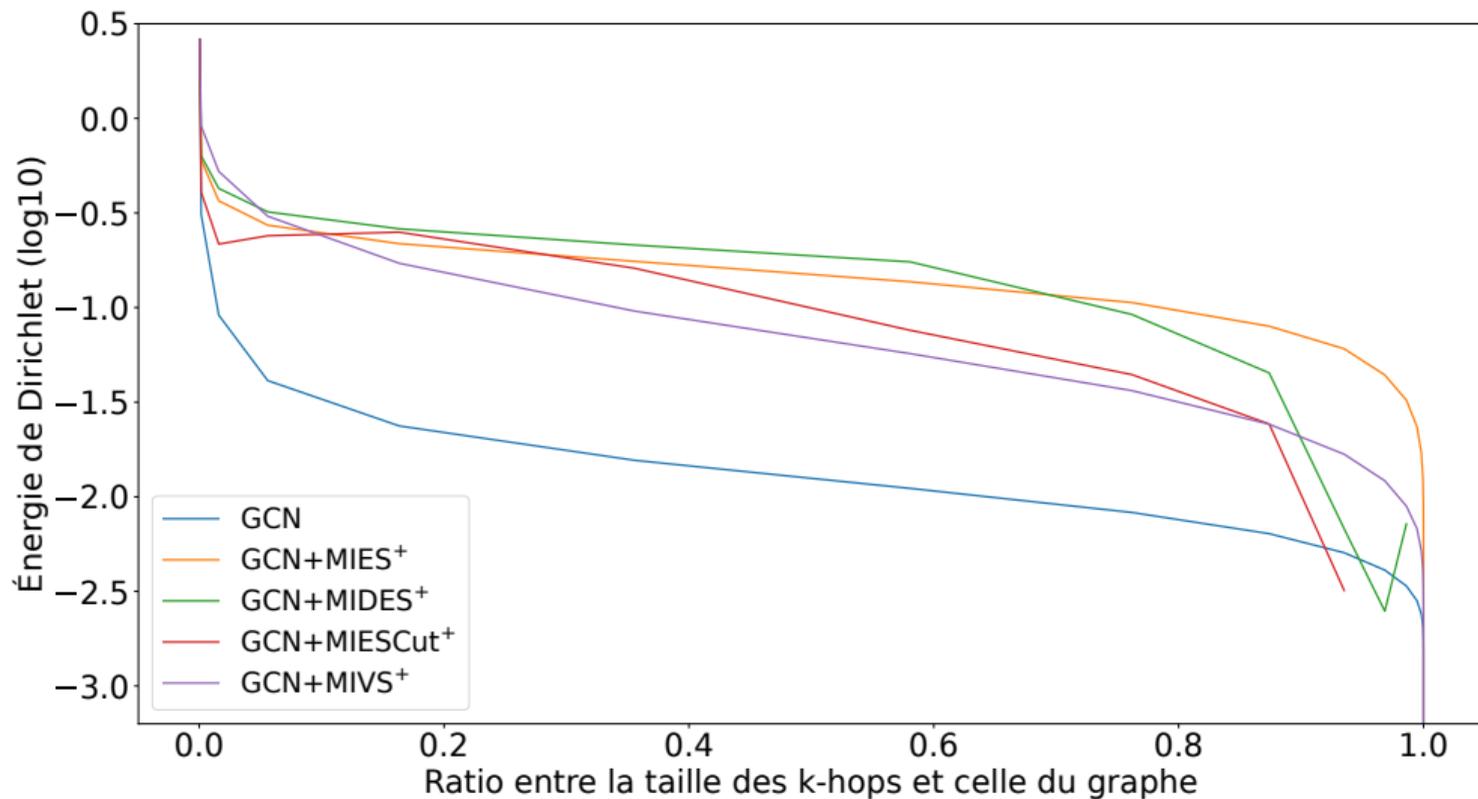


GCN et alternance GCN/pooling au -1



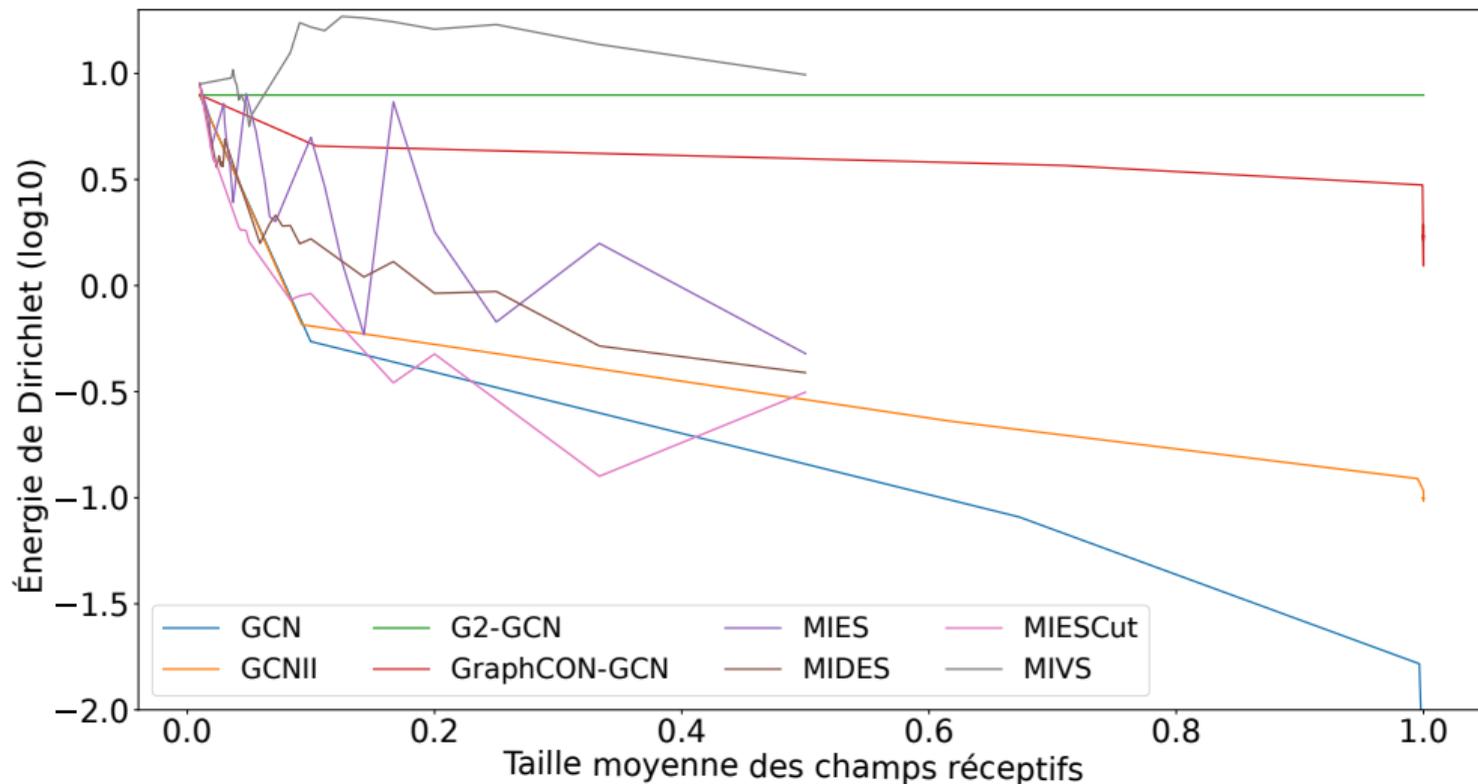


GCN et alternance GCN/pooling au -1



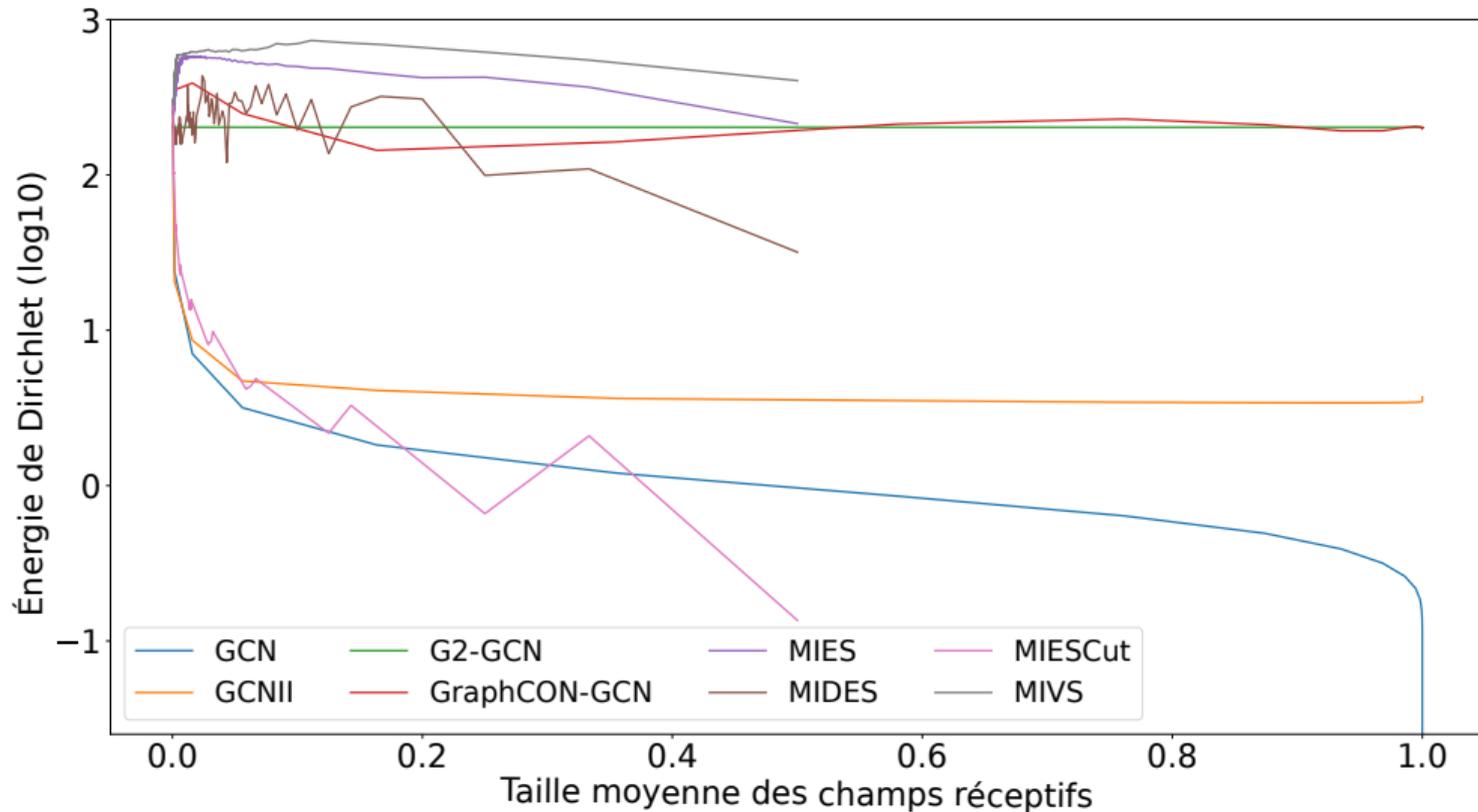


État de l'art comparaison Over-smoothing





État de l'art comparaison Over-smoothing





Références I



BIANCHI, Filippo Maria, Daniele GRATTAROLA et Cesare ALIPPI (2020). « Spectral Clustering with Graph Neural Networks for Graph Pooling ». In : *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*. T. 119. Proceedings of Machine Learning Research. PMLR, p. 874-883.



CANGEA, Catalina et al. (2018). « Towards Sparse Hierarchical Graph Classifiers ». In : *CoRR* abs/1811.01287.



DHILLON, Inderjit S., Yuqiang GUAN et Brian KULIS (2007). « Weighted Graph Cuts without Eigenvectors A Multilevel Approach ». In : *IEEE Trans. Pattern Anal. Mach. Intell.* 29.11, p. 1944-1957.



DIEHL, Frederik (2019). « Edge Contraction Pooling for Graph Neural Networks ». In : *CoRR* abs/1905.10990. URL : <http://arxiv.org/abs/1905.10990>.



DIEHL, Frederik et al. (2019). « Towards graph pooling by edge contraction ». In : *ICML 2019 workshop on learning and reasoning with graph-structured data*.



Références II



Hongyang et Shuiwang JI (2019). « Graph U-Nets ». In : *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Sous la dir. de Kamalika CHAUDHURI et Ruslan SALAKHUTDINOV. T. 97. Proceedings of Machine Learning Research. PMLR, p. 2083-2092.



GILMER, Justin et al. (2017). « Neural Message Passing for Quantum Chemistry ». In : *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*. Sous la dir. de Doina PRECUP et Yee Whye TEH. T. 70. Proceedings of Machine Learning Research. PMLR, p. 1263-1272.



GRATTAROLA, Daniele et al. (2024). « Understanding Pooling in Graph Neural Networks ». In : *IEEE Transactions on Neural Networks and Learning Systems* 35.2, p. 2708-2718.



HUANG, Jingjia et al. (2019). « AttPool : Towards Hierarchical Feature Representation in Graph Convolutional Networks via Attention Mechanism ». In : *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*. IEEE, p. 6479-6488.



Références III



KNIAZEV, Boris, Graham W. TAYLOR et Mohamed R. AMER (2019). « Understanding Attention and Generalization in Graph Neural Networks ». In : *Advances in Neural Information Processing Systems 32 : Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*. Sous la dir. d'Hanna M. WALLACH et al., p. 4204-4214.



LEE, Junhyun, Inyeop LEE et Jaewoo KANG (2019). « Self-Attention Graph Pooling ». In : *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Sous la dir. de Kamalika CHAUDHURI et Ruslan SALAKHUTDINOV. T. 97. Proceedings of Machine Learning Research. PMLR, p. 3734-3743.



MA, Yao et al. (2019). « Graph Convolutional Networks with EigenPooling ». In : *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*. Sous la dir. d'Ankur TEREDESAI et al. ACM, p. 723-731.

Références IV

- 
- 
- MA Zheng et al. (2020). « Path Integral Based Convolution and Pooling for Graph Neural Networks ». In : *Advances in Neural Information Processing Systems 33 : Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*. Sous la dir. d'Hugo LAROCHELLE et al.
- 
- MEER, Peter (1989). « Stochastic image pyramids ». In : *Comput. Vis. Graph. Image Process.* 45.3, p. 269-294.
- 
- RANJAN, Ekagra, Soumya SANYAL et Partha P. TALUKDAR (2020). « ASAP : Adaptive Structure Aware Pooling for Learning Hierarchical Graph Representations ». In : *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, p. 5470-5477.
- 
- YING, Zhitao et al. (2018). « Hierarchical Graph Representation Learning with Differentiable Pooling ». In : *Advances in Neural Information Processing Systems 31 : Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*. Sous la dir. de Samy BENGIO et al., p. 4805-4815.



-  ZHANG, Zhen, Jiajun BU, Martin ESTER, Jianfeng ZHANG, Zhao LI et al. (2023). « Hierarchical Multi-View Graph Pooling With Structure Learning ». In : *IEEE Trans. Knowl. Data Eng.* 35.1, p. 545-559.
-  ZHANG, Zhen, Jiajun BU, Martin ESTER, Jianfeng ZHANG, Chengwei YAO et al. (2019). « Hierarchical Graph Pooling with Structure Learning ». In : *CoRR* abs/1911.05954.