

Projet Graph Neural Network

Prédiction de liens d'aliments

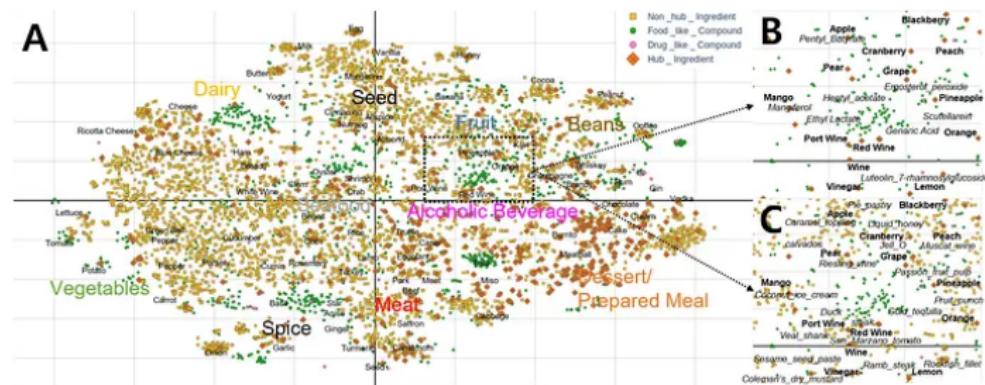
Erwann Lesech — Aymeric Le Riboter

EPITA — Majeure SCIA-G

5 janvier 2026

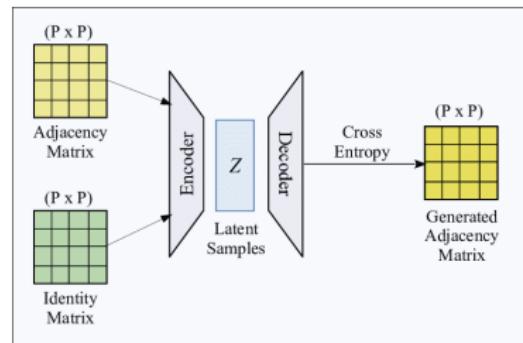
Introduction

- Les Graph Neural Networks modélisent des entités inter-connectées
- Application au domaine culinaire via le **FlavorGraph**
- Objectif : adapter et améliorer un modèle existant de prédiction de liens



Approche Initiale

- Graphe hétérogène de 1561 ingrédients
- 111 355 arêtes pondérées = affinités gustatives
- Tâche : **prédiction de liens**
- Modèle utilisé : **Graph Auto-Encoder (GAE)**
- Features très simples (ID des nœuds)
- Métrique principale : **AUC**



Modèles déjà entraînés

- **Encodeurs testés**

- GCN (GCNConv $\times 2$)
- GraphSage (SAGEConv $\times 2$)

- **Décodeur commun**

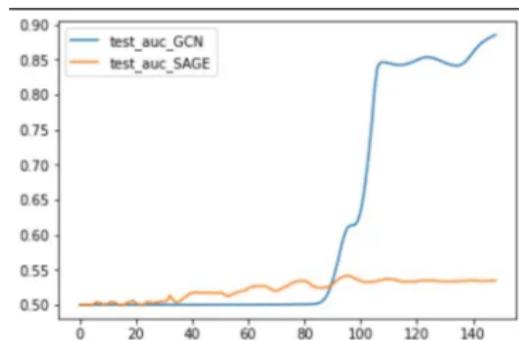
- Produit scalaire des embeddings
- Score : $s_{ij} = z_i^\top z_j$
- Apprentissage avec BCEWithLogitsLoss

- **Configuration**

- Adam, lr = 0.01
- 128 dimensions cachées \rightarrow 64 pour embeddings

- **Constats**

- Performances correctes mais limitées (recettes)
- Faible expressivité des features



Améliorations Proposées

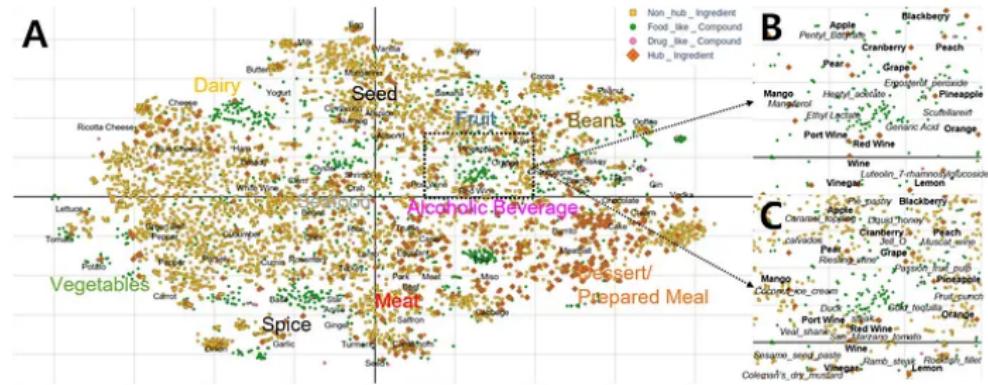
Trois axes majeurs :

- ① Enrichissement des features
- ② Diversification des architectures de l'encoder
- ③ Ajout de métriques d'évaluation orientées recommandation

Enrichissement des Features

Nouvelles représentations des nœuds :

- **Structurelles** : degré des nœuds, normalisation z-score
- **Chimiques** : nombre de composés associés à l'ingrédient
- **Catégorielles** : one-hot sur 7 catégories d'aliments



Nouveaux Modèles Testés

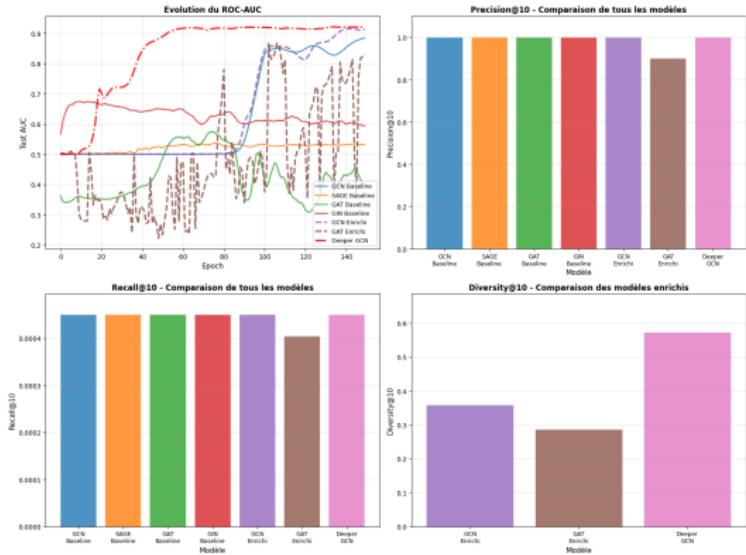
Architectures implémentées :

- GAT — mécanisme d'attention (4 têtes)
- GIN — test Weisfeiler-Lehman
- DeeperGCN — connexions résiduelles
- EnrichedNet / EnrichedGAT — ajout des nouvelles features

Métriques Avancées

- Precision@K
- Recall@K
- Diversity@10 — variété des catégories recommandées

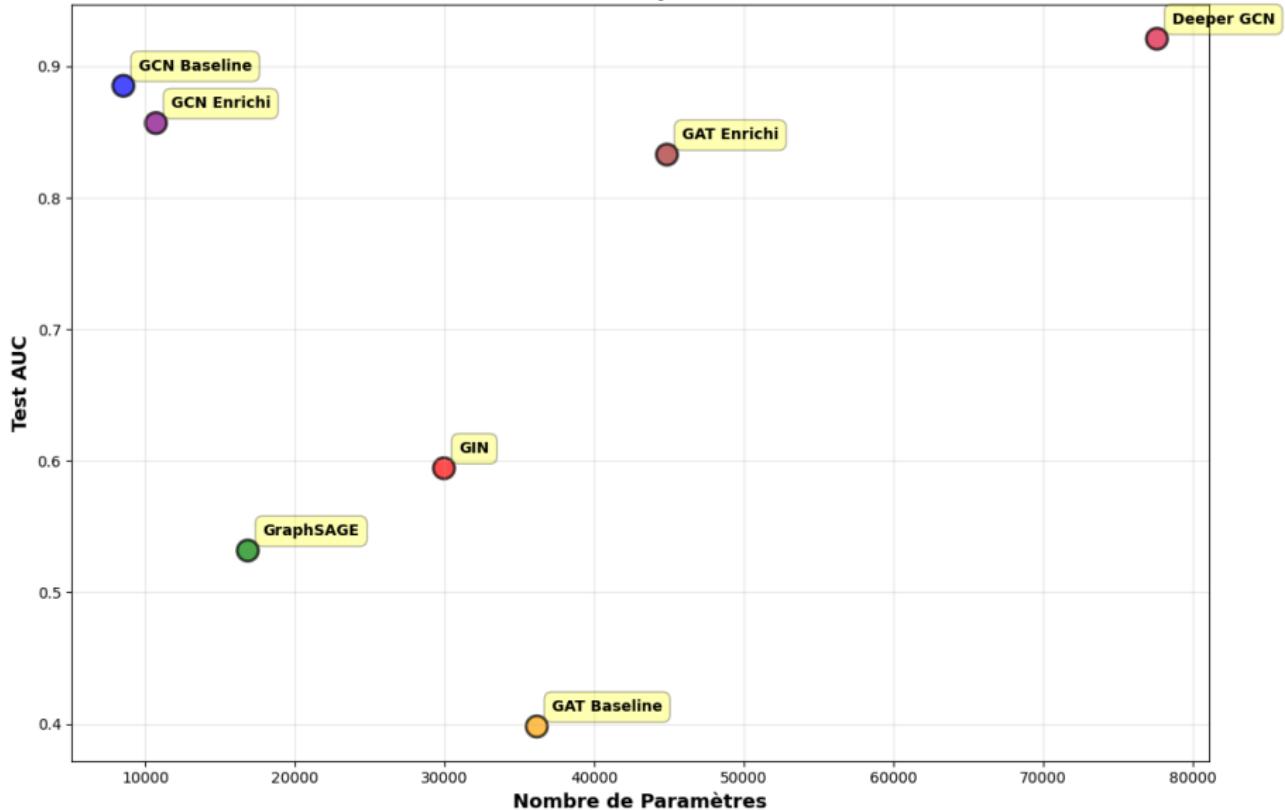
Comparatif des performances



Modèle	AUC	P@10	R@10	D@10
GCN Baseline	0.8852	1.0000	0.0004	-
SAGE Baseline	0.5318	1.0000	0.0004	-
GAT Baseline	0.5000	1.0000	0.0004	-
GIN Baseline	0.5000	1.0000	0.0004	-
GCN Enrichi	0.9132	1.0000	0.0004	0.3571
GAT Enrichi	0.8329	0.9000	0.0004	0.2857
Deeper GCN	0.9209	1.0000	0.0004	0.5714

Lien entre performance et complexité

Performance vs Complexité du Modèle



Exemple Qualitatif

Génération depuis **cheese** :

- **GAT Baseline** : cheese → chai_tea_mix → arm_roast → beef_gravy_mix → bird_seed (combinaisons incohérentes)
- **GCN Enrichi** : cheese → egg → baking_powder → bay_leaf → brown_sugar → all_purpose_flour (séquence cohérente pour une préparation)
- **Deeper GCN** : cheese → apple_cider → apple → american_cheese → baking_powder → almond_butter (combinaison créative et équilibrée)

Conclusion

- Importance capitale des features dans les GNNs
- Les métriques de ranking complètent utilement l'AUC
- Perspectives :
 - Embeddings pré-entraînés d'ingrédients
 - Features culturelles, géographiques, saisonnières
 - Tentatives de VGAE ou encore SEAL



Merci pour votre attention