

# Projet Graph Neural Network

## Prédiction de liens d'aliments

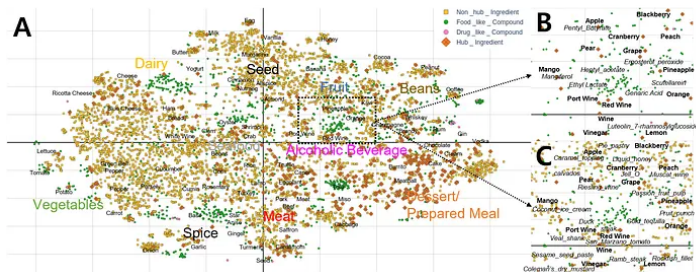
Erwann Lesech — Aymeric Le Riboter

EPITA — Majeure SCIA-G

5 janvier 2026

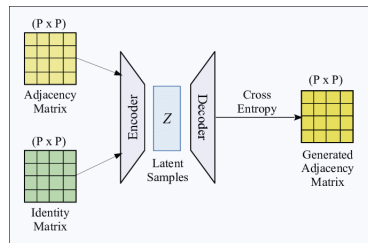
# Introduction

- 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**



- **Encodeurs testés**

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

- **Décodeur commun**

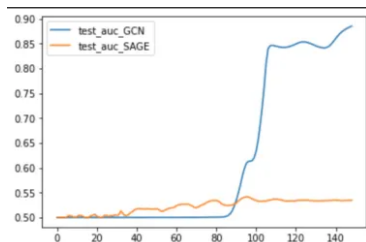
- Produit scalaire des embeddings
- Score :  $s_{ij} = z_i^\top z_j$
- Apprentissage avec BCEWithLogitsLoss

- **Configuration**

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

- **Constats**

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



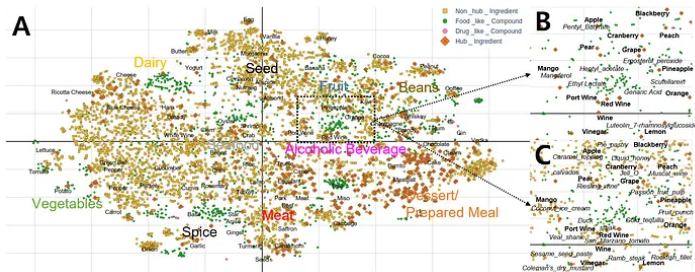
Trois axes majeurs :

- 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



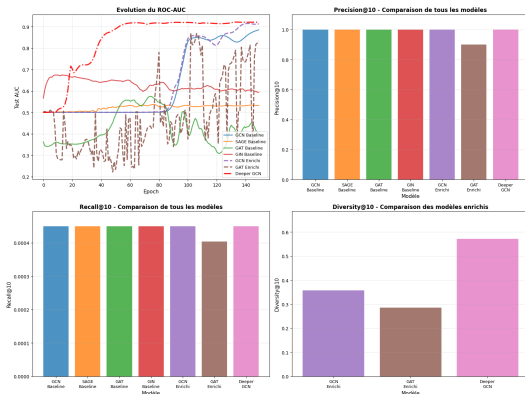
Architectures implémentées :

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

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

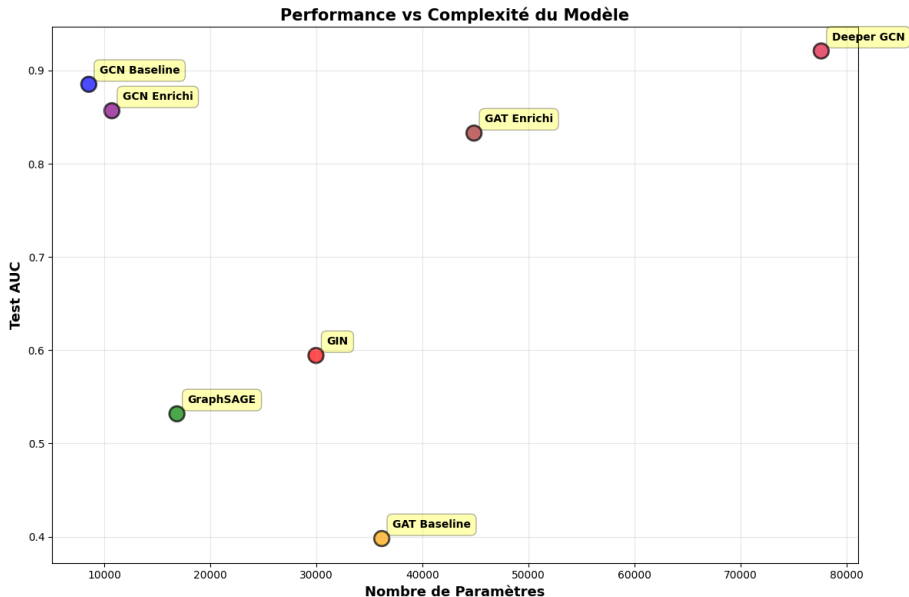


# Comparatif des performances



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	<b>0.9209</b>	<b>1.0000</b>	0.0004	<b>0.5714</b>

# Lien entre performance et complexité



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