

DÉMO LIVE · MEETUP IAAM

Drug Repurposing

par embeddings de graphe

Trouver de nouveaux usages aux médicaments, avec l'IA

Saber GRAF — PhD Neurosciences · PharmD · Fondateur, SG AI Solutions





QUI PARLE ?

Saber GRAF

PharmD · PhD Neurosciences · 12+ ans en électrophysiologie et data science

- Corporate IT & AI Strategy Manager, Neuroservice Alliance (CRO neurosciences)
- Fondateur de SG AI Solutions, conseil IA santé & biotech
- Pipelines de données santé, RAG, agents, plateforme cloud interne (Django · Next.js · Azure)

À l'intersection de trois mondes qui se parlent rarement : pharmacologie, neurosciences, IA.

Lire à la fois la prédiction ET la biologie qui la justifie, c'est tout le sujet de ce talk.

Le drug repurposing, c'est quoi ?

Trouver un nouvel usage médical pour un médicament déjà approuvé — en réduisant durée, coût et risque.

Dimension	Départ de zéro	Repositionné
Durée	~12 ans	~6 ans
Coût	~2,6 Md\$	~300 M\$
Taux d'échec	90 %	Bien plus faible

Exemples connus : Viagra (angine → dysfonction érectile) · Thalidomide (oncologie) · Aspirine (prévention cardiovasculaire)

Le mur combinatoire

3 000

**Médicaments
approuvés**

10 000

Maladies

30 M

Paires candidates

On ne peut pas tester **30 millions de paires** au laboratoire. Le machine learning peut-il les classer pour concentrer les expériences coûteuses sur les meilleurs candidats ?

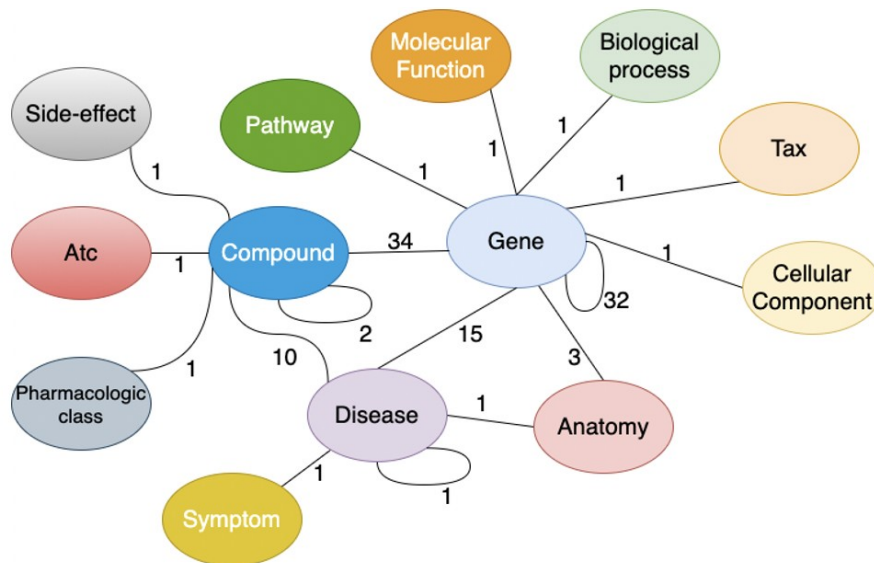
Reformuler en machine learning

Graphe de connaissances

Un graphe avec des nœuds typés et des arêtes typées.

Nœuds : Médicament · Gène · Maladie · Voie · Anatomie

Arêtes : treats · interacts_with · regulates · expressed_in



Drug repurposing = prédiction de lien

Prédire les arêtes manquantes de type (**Médicament, traite, Maladie**) — trouver les connexions les plus plausibles.

📄 La recette : DRKG → TransE → Scorer → Classifier → Ouvrir la boîte noire

Les données : DRKG

Drug Repurposing Knowledge Graph · Amazon, 2020

97K

Nœuds

107

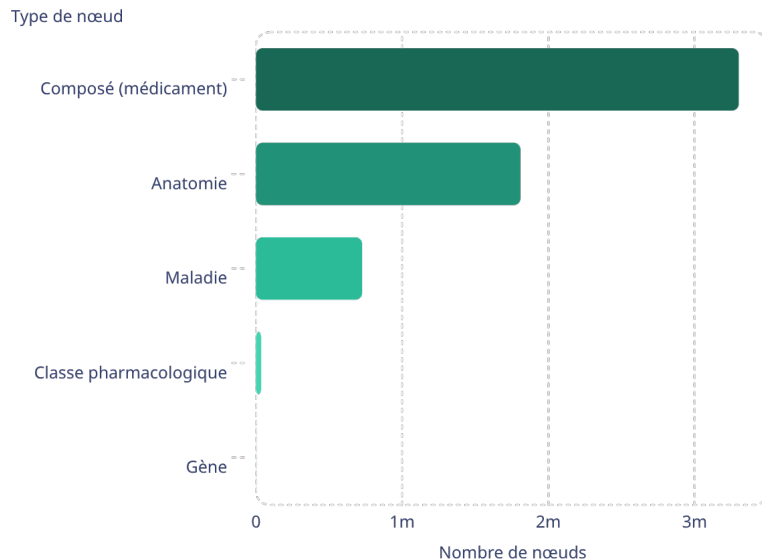
Types de relations

5,87M

Arêtes (triplets)

6

Bases sources



Distribution des nœuds

Les composés et structures anatomiques dominent le graphe, tandis que les gènes — pourtant centraux en biologie — sont sous-représentés.

Autres graphes biomédicaux

Hetionet (2017) · 47K nœuds, l'ancêtre du repurposing par graphe (projet Rephetio).

PrimeKG (Harvard, 2023) · 129K nœuds, centré médecine de précision.

OpenBioLink (2020) · ~5M arêtes, pensé pour benchmarker le link prediction.

TransE en 60 secondes

Idée centrale

$$e_h + e_r \approx e_t$$

Chaque relation est un **vecteur de translation** dans l'espace d'embedding.

Scoring : $\text{score}(h, r, t) = -\|e_h + e_r - e_t\|$

Plus le score est élevé, plus le triplet est plausible.

Pourquoi TransE ?

Conçu pour les graphes : une relation = un triplet (h, r, t), pas du texte.

Scalable + pré-entraîné : tourne sur 5,87M arêtes, zéro entraînement ici.

Un score classable : range les candidats, exactement le link prediction.

Limite assumée : relations 1-à-N gérées par ComplEx / RotatE.

Intuition : de Word2vec au biomédical

Généraliste

$\text{embedding}(\text{Paris}) + \text{embedding}(\text{capitale_de}) \approx \text{embedding}(\text{France})$

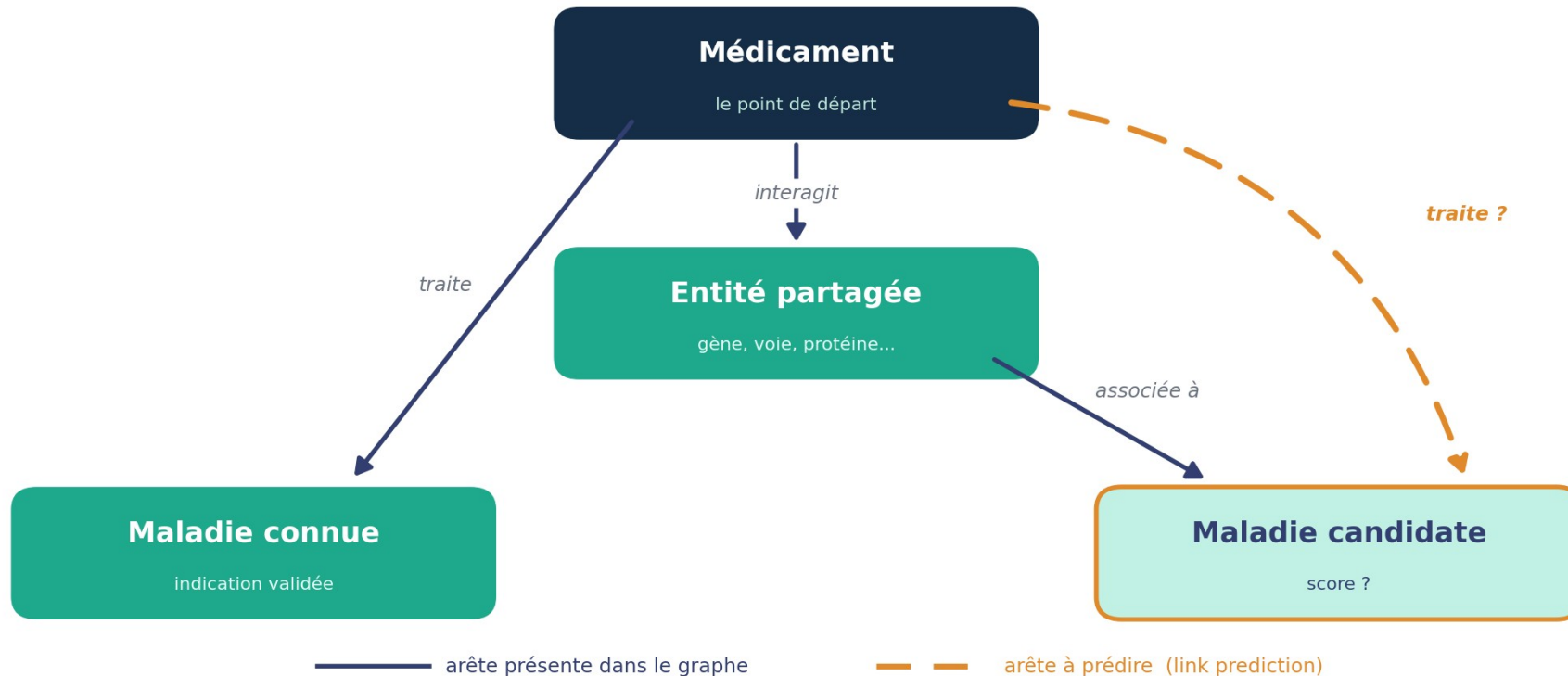
Biomédical

$\text{embedding}(\text{metformine}) + \text{embedding}(\text{traite}) \approx \text{embedding}(\text{diabète_type_2})$

Ce qu'on a dans le graphe, ce qu'on cherche

CE QU'ON A (déjà dans le graphe)

CE QU'ON CHERCHE (à prédire)



La base contient des arêtes connues (traite, cible, associé à). On demande au modèle de scorer une arête absente : (Médicament, traite ?, Maladie). C'est exactement le croisement médicament / maladie que la démo va prédire.

La question : le liraglutide

Agoniste du récepteur GLP-1 — approuvé en 2010 pour le diabète de type 2 (Victoza / Saxenda).

Étant donné tout le graphe de connaissances, **quelles maladies le liraglutide pourrait-il traiter ?**

1

Récupérer

Tous les nœuds Disease::* — 5 103 maladies

2

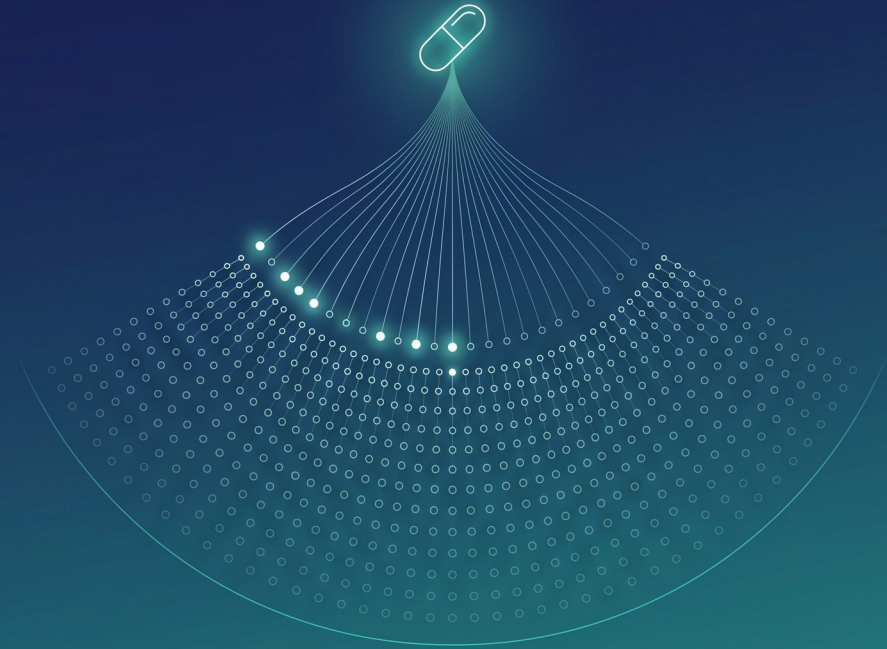
Scorer

(liraglutide, traite, d) pour chaque d

3

Classer

Du meilleur score au pire



Top 10 des indications prédites pour le liraglutide

Rang	Maladie	Score / Statut
1	Diabète de type 2	-11,86 · origine
2	Syndrome des ovaires polykystiques	-12,79
3	Rectocolite hémorragique	-12,83 · en essais
4	Dermatite atopique	-13,08
5	Lithiase rénale	-13,16
6	Paludisme	-13,18
7	Glaucome	-13,22
8	Cancer du sein	-13,27
9	Cancer de la peau	-13,30
10	Hypertension	-13,32

Alzheimer au rang 41 / 5 103 — top 0,8 %. Le KG est figé avant 2020 ; l'essai liraglutide (ELAD, phase 2b) a été mené dans les années 2010 et publié en 2025. Le modèle a signalé une hypothèse que le domaine teste activement.

Le modèle retrouve-t-il le connu ?

Sur 5 103 maladies candidates, voici où se situent quatre indications connues :

#1

Diabète type 2

Indication d'origine, 2010

#10

Hypertension

Essais cardiovasculaires en cours

#22

Obésité

Saxenda, approuvé 2014

#41

Alzheimer

Phase 2b (ELAD) : primaire manqué, secondaires positifs

Le KG est figé avant 2020 ; l'essai liraglutide a été mené dans les années 2010, publié en 2025. → Le modèle a signalé une hypothèse que le domaine teste activement.

Ouvrir la boîte noire

On a vu la prédiction. Cherchons maintenant l'explication dans le graphe.
Comment le modèle a-t-il trouvé la connexion Alzheimer ?



Comment on procède

On a un score élevé mais opaque. On va chercher l'explication directement dans le graphe.

1

Partir de la prédiction

Le couple (médicament, traite ?, maladie) : un score opaque.

2

Chercher les voisins communs

Quels nœuds relient médicament ET maladie ? Un voisin = un pont.

3

Lire les 2 arêtes du pont

Les 2 relations du pont donnent l'explication, vérifiable dans la base.

Le piège du chemin direct

0 chemin direct en 2 sauts dans 6M d'arêtes.
Pourtant le modèle classe la maladie au rang 41 / 5 103.

La géométrie des embeddings porte le signal, pas la recherche de chemin brute.

LE PONT IDENTIFIÉ

Un seul gène partagé relie deux maladies sans rapport

CÔTÉ MÉDICAMENT (DIABÈTE)

Liraglutide

le médicament

imite

GLP-1

l'hormone imitée

découpe

CÔTÉ ALZHEIMER

Alzheimer

la maladie

*s'accumule
→ cause*

β-amyloïde

la protéine du cerveau

découpe

une seule enzyme,
deux cibles

NÉPRILYSINE

l'enzyme commune ✂ (les ciseaux)

Les mêmes ciseaux (la néprilysine) découpent l'hormone visée par le médicament ET la protéine d'Alzheimer. Le modèle a trouvé ce pont sans rien connaître à la biologie.

La biologie en clair

Une enzyme, c'est une paire de ciseaux moléculaires. Chacune ne coupe que quelques cibles bien précises.

Fait 1 • GLP-1

Le médicament (liraglutide) imite une hormone, le GLP-1.
La néprilysine, ces ciseaux, le découpent.

Fait 2 • β -amyloïde

Les mêmes ciseaux découpent aussi le β -amyloïde, la protéine qui s'accumule et cause Alzheimer.

Une seule enzyme, deux maladies sans rapport. Le modèle a déduit ce lien de la seule géométrie du graphe.

ÉTAT DE L'ART 2026

Où en est le domaine en 2026 ?

Le drug discovery par IA est jeune (~5-7 ans), zéro médicament approuvé à ce jour.

173

Médicaments IA

En essais cliniques actifs

15-20

Phase 3 pivot

Attendus en 2026

2027

Premières approbations FDA

Vague attendue 2027-2030



2020

BenevolentAI prédit le baricitinib pour le Covid-19 via un graphe de connaissances (48 h)



2022

La FDA approuve le baricitinib — premier médicament signalé par IA sur le marché



2024

Rentosertib (Insilico Medicine) : premier médicament par IA générative, Phase 2 positive pour la FPI



2025

La FDA publie sa première guidance IA ; qualifie AIM-NASH, premier outil IA pour critères d'essais

Quatre points à retenir

1 • Embeddings génériques

Drug repurposing, recommandation, détection de fraude — mêmes maths.

2 • Pré-entraîné = économie

Zéro entraînement ici. Juste des vecteurs en dim 400 téléchargés.

3 • Interprétabilité atteignable

Parcourir le graphe après la prédiction. Trouver les ponts : ici, la néprilysine.

4 • Transfert inter-domaines réel

Un antidiabétique relié à Alzheimer via une enzyme partagée. Aucun a priori biologique.

Poursuivre l'exploration



Facile

- Refaire le pipeline avec un autre médicament (aspirine, sildénafil)
- Comparer TransE vs ComplEx, DistMult, RotatE (DGL-KE)
- Tracer la distribution des prédictions



Intermédiaire

- Entraîner un R-GCN sur un sous-graphe DRKG (~20 min sur GPU T4)
- Évaluer avec MRR et Hits@10 sur un test mis de côté
- Ajouter gènes ou voies comme features annexes



Recherche

- Combiner embeddings de graphe et LLM pour explications en langage naturel
- Construire son propre KG depuis PubMed et une ontologie
- Boucle d'active learning avec retour du labo de biologie

Références

DRKG · TransE (Bordes et al., NeurIPS 2013) · DGL-KE · Nature Medicine 2025 (essai liraglutide Alzheimer)

Conclusion : ce que ça change

▣ Hypothèses testables

Des candidats priorisés, pas une liste aléatoire de 30 millions.

⚡ Accélération R&D

De 12 ans à ~6 ans, de 2,6 Md\$ à ~300 M\$.

▣ Boîte noire ouverte

Lire la prédiction *et* la biologie qui la justifie.

Le repurposing par IA ne repose plus sur la chance, c'est du signal dans le bruit, amplifié par le ML.

DÉMO BONUS

Essayez-le vous-même

Le même modèle, transformé en app : choisissez un médicament, voyez les maladies candidates classées, et le chemin dans le graphe qui explique chaque prédiction.

drug-repurposing.sg-ai-solutions.com

Données publiques (DRKG). Démonstration, pas un avis médical.



ON GARDE LE CONTACT

Merci

Des questions ?

Saber GRAF

PharmD · PhD Neurosciences · Fondateur, SG AI Solutions

saber.graf@sg-ai-solutions.com

sg-ai-solutions.com



Me retrouver sur LinkedIn