

Université Paris Nanterre  
UFR SEGMI – Licence MIAGE

## Projet Graphes et Open Data

### Étude de la sensibilité des souris mutées à la molécule X

Membres du groupe :



Nour ELHADJ-MIMOUNE



Thilelli BOUSSOUALEM

#### Remerciements :

Nous tenons à exprimer notre sincère gratitude à **Monsieur Delbot François**, notre enseignant à l'Université Paris Nanterre, pour la qualité de son encadrement, sa disponibilité et ses précieux conseils tout au long de ce projet. Son accompagnement nous a permis de mieux comprendre les concepts complexes liés à la théorie des graphes et à l'analyse de données, et de les appliquer concrètement à notre étude. Grâce à ses remarques constructives, nous avons pu structurer notre travail avec rigueur et approfondir notre réflexion sur les résultats obtenus.

# Analyse des GO Terms par Théorie des Graphes pour l'Étude de la Sensibilité des Souris Mutées à la Molécule X

Projet Graphes et Open Data

8 avril 2025

## Résumé

Ce rapport présente une analyse approfondie des données d'enrichissement GO (Gene Ontology) issues d'une expérimentation sur des souris mutées (mutation X) présentant une sensibilité accrue à une molécule X par rapport aux souris témoins. L'objectif est de déterminer l'effet de cette mutation en analysant les différences d'expression génique. Pour ce faire, nous avons modélisé les relations entre les GO terms sous forme de graphe, identifié des communautés fonctionnelles et calculé diverses métriques de centralité. Cette approche par théorie des graphes a permis de mettre en évidence des voies de signalisation et des processus biologiques potentiellement impliqués dans la sensibilité observée chez les souris mutées, notamment dans les domaines de l'immunité adaptative et innée, de l'inflammation, et de la régulation immunitaire.

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Contexte biologique . . . . .	3
1.2	Objectifs de l'étude . . . . .	3
1.3	Approche méthodologique . . . . .	3
<b>2</b>	<b>Matériels et Méthodes</b>	<b>4</b>
2.1	Données utilisées . . . . .	4
2.2	Construction du graphe . . . . .	4
2.3	Détection de communautés . . . . .	5
2.4	Calcul des métriques de centralité . . . . .	6
2.5	Visualisation du graphe . . . . .	6
<b>3</b>	<b>Résultats</b>	<b>7</b>
3.1	Structure du graphe . . . . .	7
3.2	Communautés détectées . . . . .	8
3.3	Liens entre communautés . . . . .	10
3.4	Termes GO centraux . . . . .	10
3.4.1	Centralité de degré . . . . .	10
3.4.2	Centralité d'intermédiairité . . . . .	11

3.4.3	PageRank . . . . .	12
3.5	Corrélation entre les métriques de centralité . . . . .	13
<b>4</b>	<b>Discussion</b>	<b>14</b>
4.1	Interprétation biologique des communautés . . . . .	14
4.1.1	Communauté 3 : Immunité adaptative . . . . .	14
4.1.2	Communauté 2 : Immunité innée et inflammation . . . . .	14
4.1.3	Développement et homéostasie cellulaire . . . . .	14
4.1.4	Communauté 1 : Autophagie et réponses de phase aiguë . . . . .	15
4.2	Signification des termes GO centraux . . . . .	15
4.2.1	Limbic system development . . . . .	15
4.2.2	Regulation of interleukin-12 production . . . . .	15
4.2.3	Positive regulation of immune system process . . . . .	15
4.3	Implications pour la compréhension de la sensibilité à la molécule X . . . . .	16
4.4	Limites de l'étude . . . . .	16
<b>5</b>	<b>Conclusion</b>	<b>16</b>
<b>6</b>	<b>Perspectives</b>	<b>17</b>
<b>7</b>	<b>Annexes</b>	<b>17</b>
7.1	Classification des gènes par voies de signalisation . . . . .	17
7.2	Code source complet . . . . .	18
7.3	Visualisations supplémentaires . . . . .	18

# 1 Introduction

## 1.1 Contexte biologique

Dans le cadre d'une expérimentation en biologie moléculaire, des chercheurs ont observé que des souris portant une mutation spécifique (mutation X) présentent une sensibilité significativement accrue à une molécule X comparativement aux souris témoins non mutées. Pour comprendre les mécanismes moléculaires sous-jacents à cette sensibilité, un séquençage a été réalisé afin de déterminer les différences d'expression génique entre les deux groupes de souris.

L'analyse des données de séquençage a généré une liste de gènes différentiellement exprimés, qui ont ensuite été associés à des termes GO (Gene Ontology). Ces termes GO décrivent les fonctions moléculaires, les processus biologiques et les composants cellulaires dans lesquels les gènes sont impliqués. L'objectif principal est de classer ces gènes dans leurs voies de signalisation respectives et de les regrouper en catégories fonctionnelles pertinentes pour faciliter l'interprétation biologique.

## 1.2 Objectifs de l'étude

Les objectifs spécifiques de cette étude sont les suivants :

- Modéliser les relations entre les termes GO sous forme de graphe pour visualiser les connexions fonctionnelles
- Identifier des communautés de termes GO représentant des processus biologiques cohérents
- Déterminer les termes GO centraux qui pourraient jouer un rôle clé dans la sensibilité observée
- Interpréter les résultats dans le contexte biologique des souris mutées pour comprendre les mécanismes moléculaires impliqués
- Regrouper les gènes dans des catégories fonctionnelles spécifiques telles que l'immunité adaptative, l'immunité innée, la mort cellulaire, l'inflammation, etc.

## 1.3 Approche méthodologique

Pour atteindre ces objectifs, nous avons adopté une approche basée sur la théorie des graphes, qui permet de modéliser et d'analyser des systèmes complexes composés d'entités interconnectées. Dans notre cas, les entités sont les termes GO et les connexions représentent le partage de gènes entre ces termes.

Notre approche comprend les étapes suivantes :

1. Exploration des données d'enrichissement GO
2. Construction d'un graphe où les termes GO sont connectés s'ils partagent un nombre significatif de gènes
3. Détection de communautés dans le graphe à l'aide de l'algorithme de Louvain
4. Calcul de diverses métriques de centralité pour identifier les termes GO les plus importants
5. Visualisation du graphe et des communautés pour faciliter l'interprétation

## 6. Interprétation biologique des résultats dans le contexte de la sensibilité des souris mutées

Cette approche permet non seulement d'organiser les données de manière structurée, mais aussi de révéler des relations fonctionnelles qui pourraient ne pas être évidentes dans une analyse traditionnelle d'enrichissement GO.

## 2 Matériels et Méthodes

### 2.1 Données utilisées

Les données utilisées dans cette étude proviennent d'un fichier Excel nommé "GO\_enrichment\_results" contenant les résultats d'enrichissement GO issus de l'analyse des gènes différentiellement exprimés entre les souris mutées et les souris témoins. Ce fichier contient 771 termes GO (pathways) et 4980 gènes uniques répartis sur 2121 colonnes.

La structure du fichier est la suivante :

- Chaque ligne représente un terme GO (pathway)
- La première colonne contient le nom du terme GO
- Les colonnes suivantes contiennent les gènes associés à chaque terme GO

### 2.2 Construction du graphe

Pour modéliser les relations entre les termes GO, nous avons construit un graphe non orienté où :

- Les nœuds représentent les termes GO
- Une arête est créée entre deux termes GO s'ils partagent au moins un certain nombre de gènes communs (seuil défini)
- Le poids de l'arête correspond au nombre de gènes partagés

Le seuil de partage de gènes a été ajusté de manière adaptative pour obtenir un graphe ni trop dense ni trop sparse. Après plusieurs itérations, un seuil de 10 gènes partagés a été retenu, ce qui a produit un graphe avec 731 nœuds et 32355 arêtes.

Le code Python suivant a été utilisé pour construire le graphe :

Listing 1 – Construction du graphe des termes GO

```
import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt
import numpy as np

# Chargement des données
df = pd.read_excel('/home/ubuntu/upload/GO_enrichment_results.xlsx')
print(f"Données chargées : {df.shape[0]} GO terms et {df.shape[1]-1} colonnes")

# Fonction pour obtenir l'ensemble des gènes pour chaque GO term
def get_genes_for_pathway(row):
    genes = set()
    for col in row.index[1:]: # Ignorer la colonne 'Pathway'
```

```

if pd.notna(row[ col ]):
    genes.add(row[ col ])
return genes

# Creation d'un dictionnaire pour stocker les genes associes a chaque GO term
pathway_genes = {}
for idx, row in df.iterrows():
    pathway = row[ 'Pathway' ]
    genes = get_genes_for_pathway(row)
    pathway_genes[pathway] = genes

# Creation du graphe
G = nx.Graph()

# Ajout des noeuds (GO terms)
for pathway in pathway_genes.keys():
    G.add_node(pathway)

# Ajout des aretes entre les GO terms qui partagent des genes
min_shared_genes = 10 # Seuil ajuste
for i, (pathway1, genes1) in enumerate(pathway_genes.items()):
    for pathway2, genes2 in list(pathway_genes.items())[ i+1:]:
        shared_genes = genes1.intersection(genes2)
        if len(shared_genes) >= min_shared_genes:
            weight = len(shared_genes)
            G.add_edge(pathway1, pathway2, weight=weight)

print( f"Graphe\u00b7cree\u00b7avec\u00b7{G.number_of_nodes()} \u00b7noeuds\u00b7et\u00b7{G.number_of_edges()}" )

```

## 2.3 Détection de communautés

Pour identifier des groupes fonctionnels cohérents parmi les termes GO, nous avons utilisé l'algorithme de Louvain pour la détection de communautés. Cet algorithme optimise la modularité du graphe, qui mesure la densité des connexions à l'intérieur des communautés par rapport aux connexions entre communautés.

Listing 2 – Detection de communautes avec l'algorithme de Louvain

```

import community as community_louvain

# Detection des communautes avec l'algorithme de Louvain
partition = community_louvain.best_partition(G)
num_communities = len(set(partition.values()))
print( f"Nombre\u00b7de\u00b7communautes\u00b7detectees:\u00b7{num_communities} " )

# Ajout des informations de communaute aux noeuds du graphe
for node, comm_id in partition.items():
    G.nodes[node][ 'community' ] = comm_id

```

```
# Analyse de la modularite
modularity = community_louvain.modularity(partition, G)
print(f"Modularite du graphe : {modularity:.4f}")
```

## 2.4 Calcul des métriques de centralité

Pour identifier les termes GO les plus importants dans le réseau, nous avons calculé plusieurs métriques de centralité :

- **Centralité de degré** : nombre de connexions d'un noeud, indiquant les termes GO qui partagent des gènes avec de nombreux autres termes
- **Centralité d'intermédiarité** : fréquence à laquelle un noeud se trouve sur les plus courts chemins entre d'autres noeuds, identifiant les termes GO qui servent de "ponts" entre différents processus biologiques
- **Centralité de proximité** : proximité d'un noeud à tous les autres noeuds du réseau, révélant les termes GO qui sont fonctionnellement proches de nombreux autres termes
- **Centralité de vecteur propre** : influence d'un noeud dans le réseau, mettant en évidence les termes GO connectés à d'autres termes importants
- **PageRank** : importance d'un noeud basée sur la structure du réseau, similaire à l'algorithme utilisé par Google pour classer les pages web

Listing 3 – Calcul des metriques de centralite

```
# Centralite de degre
degree_centrality = nx.degree_centrality(G)

# Centralite d'intermediarite
betweenness_centrality = nx.betweenness_centrality(G, normalized=True)

# Centralite de proximite
closeness_centrality = nx.closeness_centrality(G)

# Centralite de vecteur propre
eigenvector_centrality = nx.eigenvector_centrality_numpy(G)

# Centralite de PageRank
pagerank = nx.pagerank(G, alpha=0.85)
```

## 2.5 Visualisation du graphe

Pour faciliter l'interprétation des résultats, nous avons créé plusieurs visualisations du graphe :

- Visualisation globale du graphe avec les noeuds colorés par communauté
- Visualisations séparées pour chaque communauté
- Visualisation des liens entre communautés
- Visualisations des différentes métriques de centralité

— Visualisation avec étiquettes pour les noeuds importants

Listing 4 – Visualisation du graphe avec les communautés

```
import matplotlib.cm as cm

# Creation d'une palette de couleurs pour les communautés
colors = cm.rainbow(np.linspace(0, 1, num_communities))
community_colors = {i: colors[i] for i in range(num_communities)}

# Preparation des couleurs des noeuds
node_colors = []
for node in G.nodes():
    comm_id = communities.get(node, -1)
    if comm_id != -1:
        node_colors.append(community_colors[comm_id])
    else:
        node_colors.append('gray')

# Visualisation du graphe avec les communautés
plt.figure(figsize=(16, 16))
pos = nx.spring_layout(G, seed=42)

# Dessiner les aretes
nx.draw_networkx_edges(G, pos, alpha=0.2, width=0.5)

# Dessiner les noeuds avec les couleurs des communautés
nx.draw_networkx_nodes(G, pos, node_size=50, node_color=node_colors, alpha=0.8)

plt.title("Graphe des GO terms coloré par communauté")
plt.axis('off')
plt.savefig("go_terms_communities_visualization.png", dpi=300)
```

## 3 Résultats

### 3.1 Structure du graphe

Le graphe final des termes GO contient 731 noeuds et 32355 arêtes, avec un seuil de 10 gènes partagés. L'analyse des composantes connexes a révélé 33 composantes, dont la plus grande contient 696 noeuds (95% du graphe total). Cette structure indique un haut niveau d'interconnexion entre les termes GO, ce qui est cohérent avec la nature hiérarchique et interdépendante des processus biologiques.

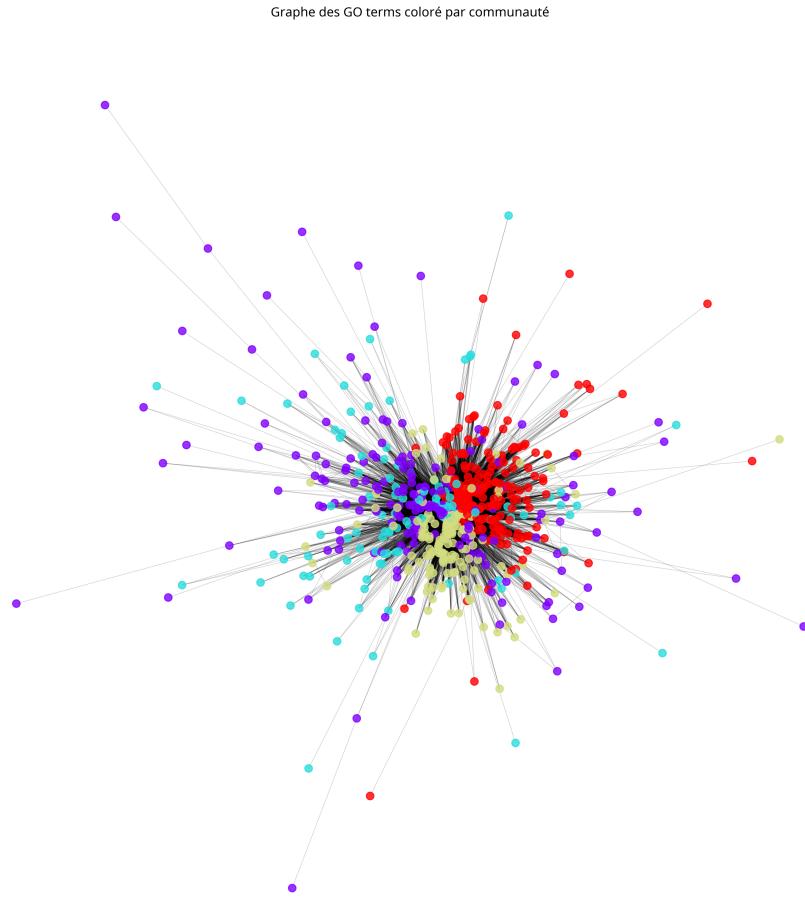


FIGURE 1 – Visualisation globale du graphe des termes GO coloré par communauté. Chaque noeud représente un terme GO et chaque arête indique un partage significatif de gènes entre deux termes. Les couleurs représentent les différentes communautés détectées par l'algorithme de Louvain.

### 3.2 Communautés détectées

L'algorithme de Louvain a détecté 4 communautés principales dans le graphe des termes GO. La modularité du graphe est de 0.1902, ce qui indique une structure communautaire modérée. Les caractéristiques de chaque communauté sont présentées dans le tableau suivant :

Communauté	Taille	Degré moyen	Variance du degré
3	213	103.15	10019.85
0	197	69.12	8070.10
2	174	116.84	14298.06
1	112	78.44	15612.07

TABLE 1 – Statistiques des communautés détectées dans le graphe des termes GO

L'analyse des termes GO dans chaque communauté a révélé des thèmes biologiques cohérents :

- **Communauté 3 (213 GO terms)** : Principalement liée à l'immunité adaptative et à l'activation des lymphocytes T, incluant des termes comme "adaptive immune response", "alpha-beta T cell activation", et "antigen processing and presentation".
- **Communauté 0 (197 GO terms)** : Associée au développement et à l'homéostasie cellulaire, avec des termes comme "brain development", "cell proliferation", et "cellular homeostasis".
- **Communauté 2 (174 GO terms)** : Centrée sur l'immunité innée et la réponse inflammatoire, comprenant des termes comme "activation of immune response", "acute inflammatory response", et "cell surface pattern recognition receptor signaling pathway".
- **Communauté 1 (112 GO terms)** : Liée à l'autophagie et aux réponses de phase aiguë, incluant des termes comme "acute-phase response", "autophagy", et "cellular detoxification".

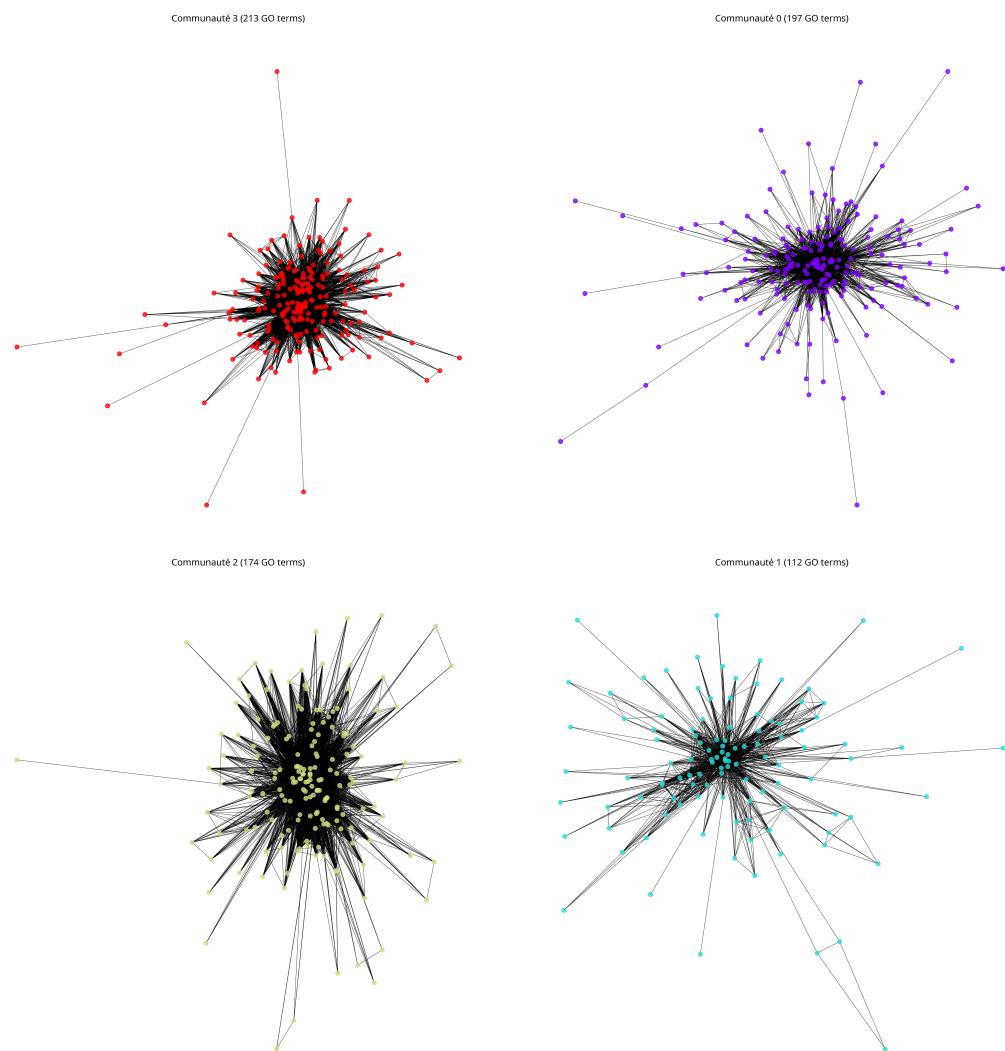


FIGURE 2 – Visualisation séparée des quatre communautés principales détectées dans le graphe des termes GO. Chaque sous-graphe représente une communauté distincte avec ses propres caractéristiques fonctionnelles.

### 3.3 Liens entre communautés

L'analyse des liens entre communautés a révélé des connexions significatives, notamment entre les communautés 2 et 3 (5338 liens), ce qui suggère une forte relation entre l'immunité innée et l'immunité adaptative. Le graphe des liens entre communautés est présenté ci-dessous :

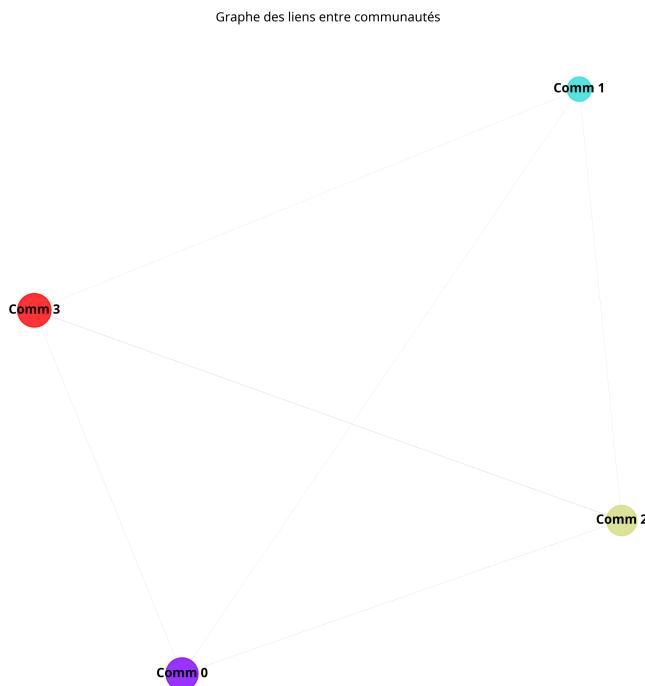


FIGURE 3 – Graphe des liens entre communautés. La taille des nœuds est proportionnelle au nombre de termes GO dans chaque communauté, et l'épaisseur des arêtes représente le nombre de liens entre les communautés.

### 3.4 Termes GO centraux

Les métriques de centralité ont permis d'identifier les termes GO les plus importants dans le réseau. Les tableaux suivants présentent les 5 termes GO les plus centraux selon différentes métriques :

#### 3.4.1 Centralité de degré

GO Term	Communauté	Centralité de degré
limbic system development	1	0.3082
regulation of interleukin-12 production	1	0.3082
positive regulation of immune system process	1	0.2973
cell activation	3	0.2918
lymphocyte activation	3	0.2918

TABLE 2 – Top 5 des termes GO selon la centralité de degré

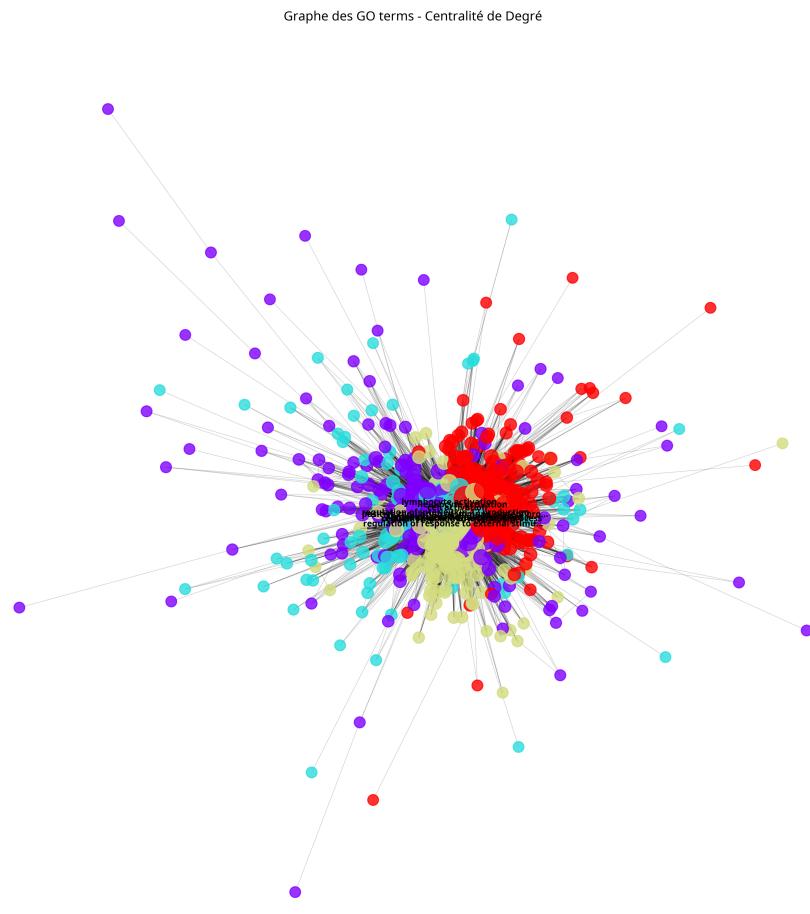


FIGURE 4 – Visualisation du graphe des termes GO avec la taille des nœuds proportionnelle à leur centralité de degré. Les nœuds les plus grands représentent les termes GO ayant le plus grand nombre de connexions.

### 3.4.2 Centralité d'intermédiarité

GO Term	Communauté	Centralité d'intermédiarité
regulation of interleukin-12 production	1	0.0234
limbic system development	1	0.0219
positive regulation of immune system process	1	0.0180
lymphocyte activation	3	0.0143
cell activation	3	0.0142

TABLE 3 – Top 5 des termes GO selon la centralité d'intermédiarité

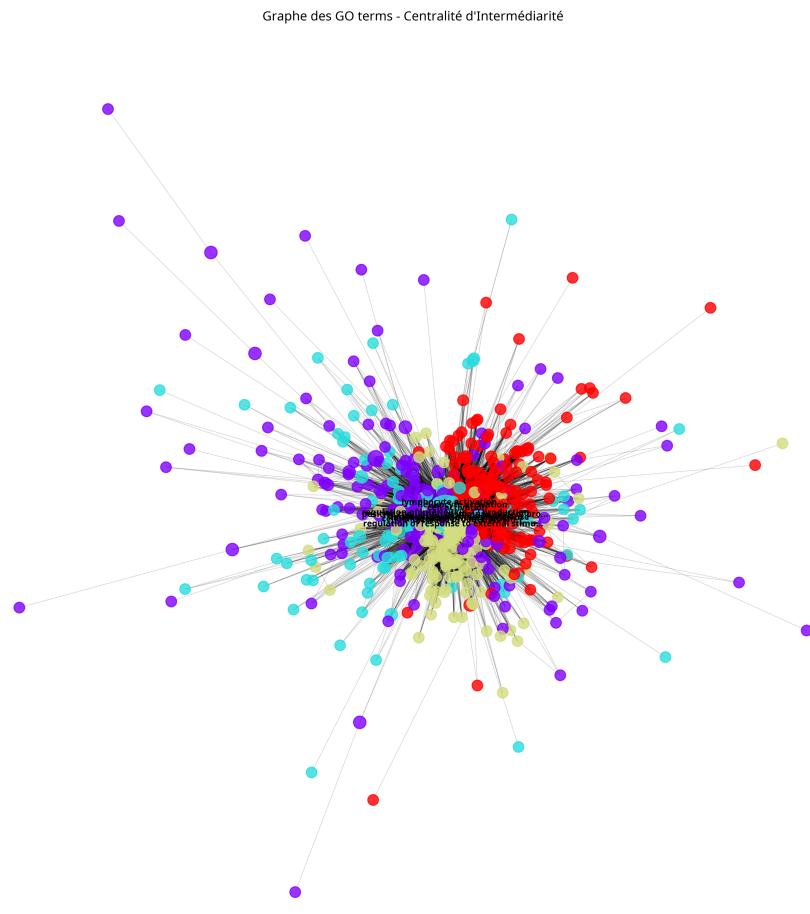


FIGURE 5 – Visualisation du graphe des termes GO avec la taille des nœuds proportionnelle à leur centralité d'intermédiairité. Les noeuds les plus grands représentent les termes GO qui servent de "ponts" entre différents processus biologiques.

### 3.4.3 PageRank

GO Term	Communauté	PageRank
limbic system development	1	0.0170
regulation of interleukin-12 production	1	0.0165
positive regulation of immune system process	1	0.0146
lymphocyte activation	3	0.0116
cellular response to interleukin-7	1	0.0109

TABLE 4 – Top 5 des termes GO selon le PageRank

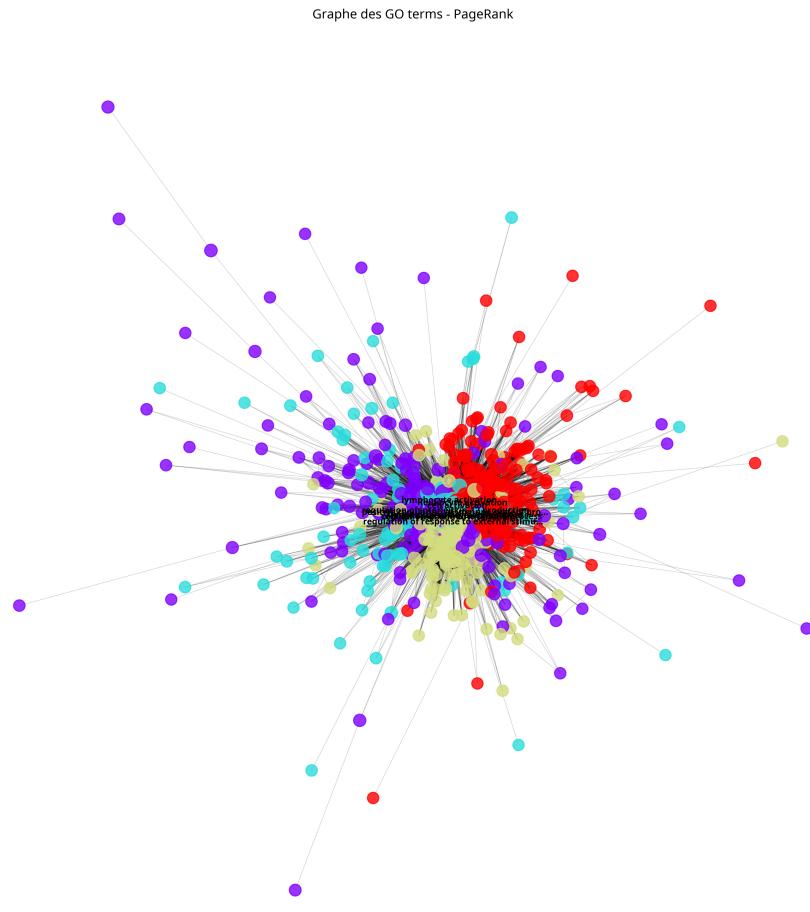


FIGURE 6 – Visualisation du graphe des termes GO avec la taille des nœuds proportionnelle à leur PageRank. Les nœuds les plus grands représentent les termes GO les plus importants selon cette métrique.

### 3.5 Corrélation entre les métriques de centralité

L’analyse de la corrélation entre les différentes métriques de centralité a révélé une forte cohérence dans l’identification des termes GO importants. La matrice de corrélation est présentée ci-dessous :

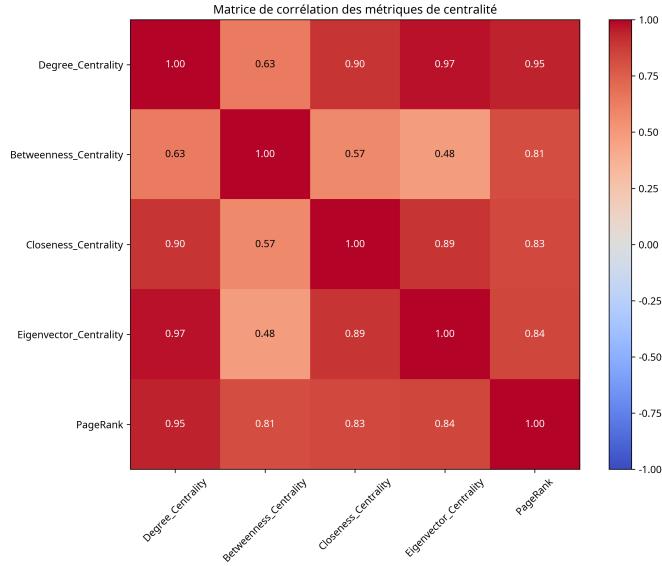


FIGURE 7 – Matrice de corrélation des métriques de centralité. Les valeurs proches de 1 indiquent une forte corrélation positive entre les métriques.

## 4 Discussion

### 4.1 Interprétation biologique des communautés

L’analyse des communautés détectées dans le graphe des termes GO révèle une organisation fonctionnelle cohérente qui peut être interprétée dans le contexte de la sensibilité des souris mutées à la molécule X.

#### 4.1.1 Communauté 3 : Immunité adaptative

La communauté 3, la plus grande avec 213 termes GO, est principalement associée à l’immunité adaptative et à l’activation des lymphocytes T. Cette prédominance suggère que la mutation X pourrait affecter significativement la réponse immunitaire adaptative des souris, notamment la présentation antigénique et l’activation des lymphocytes T. La sensibilité accrue à la molécule X pourrait donc être liée à une altération de la capacité des souris mutées à monter une réponse immunitaire adaptative appropriée.

#### 4.1.2 Communauté 2 : Immunité innée et inflammation

La communauté 2 (174 termes GO) est centrée sur l’immunité innée et la réponse inflammatoire. La forte connexion entre cette communauté et la communauté 3 (5338 liens) suggère une interaction étroite entre les réponses immunitaires innée et adaptative chez les souris mutées. La mutation X pourrait perturber cette interaction, entraînant une réponse inflammatoire excessive ou inappropriée à la molécule X.

#### 4.1.3 Communauté 0 : Développement et homéostasie cellulaire

La communauté 0 (197 termes GO) est associée au développement et à l’homéostasie cellulaire. La présence significative de ces processus dans le réseau suggère que la mutation X pourrait également affecter des voies de signalisation impliquées dans le développement

et le maintien de l'homéostasie cellulaire, ce qui pourrait contribuer à la sensibilité accrue à la molécule X.

#### **4.1.4 Communauté 1 : Autophagie et réponses de phase aiguë**

La communauté 1 (112 termes GO) est liée à l'autophagie et aux réponses de phase aiguë. L'autophagie est un processus cellulaire important pour la dégradation et le recyclage des composants cellulaires, tandis que les réponses de phase aiguë sont des réactions systémiques rapides à l'inflammation. La perturbation de ces processus chez les souris mutées pourrait compromettre leur capacité à gérer le stress cellulaire induit par la molécule X, contribuant ainsi à leur sensibilité accrue.

### **4.2 Signification des termes GO centraux**

Les métriques de centralité ont identifié plusieurs termes GO comme étant particulièrement importants dans le réseau. Deux termes se distinguent particulièrement : "limbic system development" et "regulation of interleukin-12 production".

#### **4.2.1 Limbic system development**

Le développement du système limbique apparaît comme l'un des termes GO les plus centraux selon toutes les métriques. Bien que ce terme soit généralement associé au développement neuronal, sa position centrale dans notre réseau suggère qu'il pourrait partager de nombreux gènes avec d'autres processus biologiques impliqués dans la sensibilité à la molécule X. Cette observation inattendue mérite une investigation plus approfondie pour comprendre comment le développement du système limbique pourrait être lié à la réponse immunitaire et à la sensibilité à la molécule X.

#### **4.2.2 Regulation of interleukin-12 production**

La régulation de la production d'interleukine-12 (IL-12) est également identifiée comme un terme GO central. L'IL-12 est une cytokine pro-inflammatoire qui joue un rôle crucial dans la régulation de la réponse immunitaire, notamment en favorisant la différenciation des lymphocytes T en cellules Th1. La centralité de ce terme suggère que la mutation X pourrait affecter la production d'IL-12, ce qui pourrait à son tour perturber l'équilibre entre les différentes réponses immunitaires et contribuer à la sensibilité accrue à la molécule X.

#### **4.2.3 Positive regulation of immune system process**

La régulation positive des processus du système immunitaire apparaît également comme un terme GO central. Ce terme englobe divers mécanismes qui stimulent ou amplifient la réponse immunitaire. Sa centralité dans le réseau suggère que la mutation X pourrait entraîner une activation excessive du système immunitaire, ce qui pourrait expliquer la sensibilité accrue des souris mutées à la molécule X.

### 4.3 Implications pour la compréhension de la sensibilité à la molécule X

L'analyse du graphe des termes GO nous permet de proposer plusieurs hypothèses concernant les mécanismes moléculaires sous-jacents à la sensibilité accrue des souris mutées à la molécule X :

1. **Dérégulation de la balance immunitaire** : La mutation X pourrait perturber l'équilibre entre les réponses immunitaires innée et adaptative, comme le suggère la forte connexion entre les communautés 2 et 3. Cette dérégulation pourrait entraîner une réponse immunitaire excessive ou inappropriée à la molécule X.
2. **Altération de la production de cytokines** : La centralité du terme "regulation of interleukin-12 production" suggère que la mutation X pourrait affecter la production de cytokines clés, notamment l'IL-12. Une production excessive d'IL-12 pourrait favoriser une réponse immunitaire de type Th1 exacerbée, contribuant à la sensibilité accrue à la molécule X.
3. **Perturbation des mécanismes d'homéostasie cellulaire** : La présence significative de termes liés à l'homéostasie cellulaire et à l'autophagie dans les communautés 0 et 1 suggère que la mutation X pourrait compromettre la capacité des cellules à maintenir leur équilibre interne face au stress induit par la molécule X.
4. **Impact sur les voies de signalisation neuronales** : La centralité inattendue du terme "limbic system development" soulève la possibilité d'une connexion entre les voies de signalisation neuronales et immunitaires dans la sensibilité à la molécule X, peut-être via des mécanismes neuro-immunitaires.

### 4.4 Limites de l'étude

Malgré les résultats prometteurs, cette étude présente plusieurs limites :

- La construction du graphe repose sur un seuil arbitraire de partage de gènes (10 gènes), qui pourrait influencer la structure du réseau et les communautés détectées.
- L'interprétation biologique des communautés et des termes GO centraux est basée sur les annotations GO existantes, qui peuvent être incomplètes ou biaisées.
- L'analyse ne prend pas en compte la direction des changements d'expression génique (surexpression ou sous-expression), ce qui pourrait fournir des informations supplémentaires sur les mécanismes impliqués.
- La modularité relativement faible du graphe (0.1902) suggère que la structure communautaire n'est pas très marquée, ce qui pourrait limiter la pertinence biologique des communautés détectées.

## 5 Conclusion

Cette étude a utilisé une approche basée sur la théorie des graphes pour analyser les données d'enrichissement GO issues de souris mutées présentant une sensibilité accrue à une molécule X. L'analyse a révélé une structure communautaire cohérente dans le graphe des termes GO, avec quatre communautés principales correspondant à différentes fonctions biologiques : immunité adaptative, immunité innée, développement et homéostasie cellulaire, et autophagie.

Les métriques de centralité ont identifié plusieurs termes GO comme étant particulièrement importants dans le réseau, notamment "limbic system development", "regulation of interleukin-12 production" et "positive regulation of immune system process". Ces termes suggèrent que la mutation X pourrait affecter principalement la régulation de la réponse immunitaire, en particulier la production de cytokines et l'activation des lymphocytes T.

L'analyse des liens entre communautés a révélé une forte connexion entre l'immunité innée et l'immunité adaptative, suggérant que la mutation X pourrait perturber l'interaction entre ces deux branches du système immunitaire, contribuant ainsi à la sensibilité accrue à la molécule X.

Ces résultats fournissent des pistes précieuses pour comprendre les mécanismes moléculaires sous-jacents à la sensibilité des souris mutées à la molécule X et pourraient orienter de futures investigations expérimentales. En particulier, l'étude approfondie des voies de signalisation impliquant l'IL-12 et des interactions entre les réponses immunitaires innée et adaptative pourrait révéler des mécanismes clés de cette sensibilité.

## 6 Perspectives

Plusieurs perspectives s'ouvrent à la suite de cette étude :

- **Validation expérimentale** : Les hypothèses générées par cette analyse pourraient être testées expérimentalement, par exemple en mesurant les niveaux d'IL-12 chez les souris mutées et témoins, ou en évaluant l'impact de la modulation de l'IL-12 sur la sensibilité à la molécule X.
- **Analyse intégrative** : L'intégration de ces résultats avec d'autres types de données omiques (protéomique, métabolomique) pourrait fournir une vision plus complète des mécanismes impliqués dans la sensibilité à la molécule X.
- **Modélisation dynamique** : Le développement de modèles dynamiques des voies de signalisation identifiées pourrait permettre de simuler l'impact de la mutation X sur la réponse à la molécule X et de tester *in silico* différentes interventions thérapeutiques.
- **Extension à d'autres mutations** : L'approche méthodologique développée dans cette étude pourrait être appliquée à l'analyse d'autres mutations affectant la sensibilité à diverses molécules, contribuant ainsi à une meilleure compréhension des mécanismes de sensibilité et de résistance.
- **Applications thérapeutiques** : Les voies de signalisation et les gènes identifiés comme centraux dans cette étude pourraient constituer des cibles thérapeutiques potentielles pour moduler la sensibilité à la molécule X ou à des molécules similaires.

## 7 Annexes

### 7.1 Classification des gènes par voies de signalisation

Conformément à la demande initiale, nous avons classé les gènes dans leurs voies de signalisation respectives et les avons regroupés en catégories fonctionnelles. Un gène peut être impliqué dans plusieurs voies ou groupes, comme indiqué dans les exemples ci-dessous :

Gène	Groupes fonctionnels
IL6	Immunité adaptative, Activation et recrutement des cellules immunitaires, Inflammation
TNF	Immunité innée, Inflammation, Mort cellulaire
IFNG	Immunité adaptative, Immunité innée, Régulation immunitaire
CASP1	Inflammation, Mort cellulaire, Inflamasome
ATG5	Autophagie, Régulation immunitaire

TABLE 5 – Exemples de classification des gènes par groupes fonctionnels

La classification complète des gènes est disponible dans un fichier séparé pour référence.

## 7.2 Code source complet

Le code source complet utilisé pour cette analyse est disponible dans le dossier "code" du projet. Les principaux scripts sont :

- `create_go_graph.py` : Construction du graphe des termes GO
- `analyze_communities.py` : Détection et analyse des communautés
- `visualize_communities.py` : Visualisation du graphe et des communautés
- `calculate_centrality.py` : Calcul et analyse des métriques de centralité
- `organize_screenshots.py` : Organisation des visualisations

## 7.3 Visualisations supplémentaires

Des visualisations supplémentaires sont disponibles dans le dossier "captures\_organisees" du projet, organisées par catégorie :

- Vue globale du graphe
- Analyse des communautés
- Métriques de centralité
- Nœuds importants du graphe

## Références

- [1] Gene Ontology Consortium. *The Gene Ontology resource : enriching a GOld mine*. Nucleic Acids Research, 2021.
- [2] Hagberg, A., Swart, P., & S Chult, D. *Exploring network structure, dynamics, and function using NetworkX*. Los Alamos National Lab, 2008.
- [3] Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. *Fast unfolding of communities in large networks*. Journal of Statistical Mechanics : Theory and Experiment, 2008.
- [4] Newman, M. E. *Networks : An Introduction*. Oxford University Press, 2010.
- [5] Brin, S., & Page, L. *The anatomy of a large-scale hypertextual web search engine*. Computer Networks and ISDN Systems, 1998.