

Mini-projet : Analyse du réseau aérien mondial (OpenFlights)

Anais Storp

Données complexes et Réseaux de Neurones

Date : 8 décembre 2024

LU3DS003 Raphaël Fournier-S'niehotta



Anais Storp

Table des matières

I	Introduction	2
1.1	Objectifs et présentation	3
1.2	Présentation des données	3
II	Exploration des données	4
2.1	Statistiques descriptives	4
2.2	Distribution des degrés	5
2.3	Analyse des composantes connexes	5
III	Analyse des propriétés du graphe	7
3.1	Centralité des nœuds	7
3.2	Coefficient de clustering	8
3.3	Analyse des plus courts chemins et du diamètre	8
3.4	Analyse des poids des arêtes	9
IV	Analyse des communautés	9
4.1	Détection des communautés	10
4.2	Modularité	11
V	Analyse de robustesse	12
5.1	Simulation de pannes aléatoires	13
5.2	Simulation de pannes ciblées	14
5.3	Analyse de la résilience du réseau	14
VI	Visualisation	16
6.1	Graphe global	16
6.2	Sous-parties clés	17
VII	Conclusion	19
7.1	Résumé des observations clés	19
7.2	Implications pratiques	19
7.3	Perspectives	19

I Introduction

Le transport aérien est un pilier central de la connectivité mondiale, reliant les régions les plus éloignées et facilitant les échanges économiques, culturels et sociaux. À travers des milliers d'aéroports et de liaisons directes, les réseaux aériens forment une infrastructure complexe qui doit répondre à des contraintes géographiques, économiques, et opérationnelles. Ces réseaux présentent des structures non triviales qui, lorsqu'elles sont modélisées sous forme de graphes, offrent une perspective précieuse sur leur organisation et leur fonctionnement.

Dans ce projet, nous analysons les données issues du réseau *OpenFlights*, un jeu de données open-source qui cartographie les connexions directes entre aéroports dans le monde entier. Chaque nœud du graphe représente un aéroport, et chaque arête une liaison directe entre deux aéroports. Ces données, bien que simplifiées, capturent la structure fondamentale du réseau de transport aérien et permettent d'explorer ses propriétés topologiques. La visualisation et l'analyse de ce graphe permettent non seulement de mieux comprendre son fonctionnement, mais aussi d'identifier ses faiblesses et opportunités d'optimisation.

Pourquoi visualiser ce réseau ?

La visualisation joue un rôle crucial dans l'analyse des graphes complexes. Elle permet :

- **D'identifier les structures globales** : telles que les communautés et la composante géante, qui montrent comment le réseau est organisé.
- **De localiser les nœuds clés** : comme les hubs ou les aéroports faiblement connectés, qui sont essentiels pour la robustesse du réseau.
- **De guider les analyses quantitatives** : en révélant des motifs cachés qui méritent une exploration plus détaillée.

Cependant, la taille et la densité de ce type de graphe (2939 nœuds et 30501 arêtes) rendent la visualisation difficilement lisible sans des techniques avancées, comme l'utilisation de layouts optimisés ou le filtrage des données.

Problématique

Le réseau de transport aérien soulève plusieurs questions fondamentales :

- **Quels sont les aéroports les plus connectés ?** Les aéroports majeurs ou *hubs* jouent un rôle central dans la connectivité du réseau. Les identifier permet de comprendre leur importance dans le réseau mondial.
- **Comment les communautés géographiques ou fonctionnelles se manifestent-elles dans le graphe ?** Les réseaux aériens sont souvent segmentés par régions ou zones économiques. Analyser ces communautés peut révéler des regroupements naturels ou des dépendances entre régions.
- **Quelle est la résilience du réseau face à des perturbations ?** En cas de suppression d'un hub (par exemple, fermeture temporaire d'un grand aéroport), il est important de comprendre comment le réseau est affecté. Cette analyse de la robustesse est essentielle pour anticiper les vulnérabilités du système.

Ces problématiques sont directement liées aux caractéristiques fondamentales des réseaux réels, où une organisation hiérarchique et modulaire peut à la fois favoriser l'efficacité et créer des points faibles.

1.1 Objectifs et présentation

Ce projet vise à répondre aux questions posées en suivant une méthodologie structurée et des outils modernes pour l'analyse des graphes. Les objectifs principaux incluent :

1. **Décrire les propriétés globales du réseau :** Identifier des métriques globales comme la densité, la distribution des degrés, et les composantes connexes pour avoir une vue d'ensemble de la structure du graphe.
2. **Identifier les aéroports clés :** À travers des mesures de centralité (degré, inter-médianité, proximité), déterminer quels nœuds jouent un rôle stratégique dans la connectivité globale.
3. **Analyser les communautés :** Étudier la segmentation naturelle du réseau pour comprendre les regroupements géographiques ou fonctionnels entre les aéroports.
4. **Explorer la robustesse :** Évaluer comment le réseau réagit face à des perturbations, en analysant des métriques locales comme le clustering et globales comme la résilience des composantes géantes.
5. **Visualiser les résultats :** Générer des graphiques clairs et informatifs, à la fois pour comprendre les structures globales et pour localiser des éléments clés du réseau.

En combinant ces analyses qualitatives et quantitatives, ce projet cherche à fournir des observations détaillées sur la structure du réseau OpenFlights, tout en illustrant la puissance des outils de la théorie des graphes dans la compréhension de réseaux complexes.

Méthodologie

L'analyse suit une démarche structurée, en utilisant Python pour le traitement des données et les visualisations, complétée par une exploration avancée dans Gephi. Les étapes principales sont :

1. Chargement des données et construction du graphe.
2. Analyse des propriétés globales (distribution des degrés, densité, composantes connexes).
3. Calcul des mesures de centralité pour évaluer l'importance des nœuds.
4. Détection des communautés à l'aide de l'algorithme de modularité.
5. Visualisations interactives pour interpréter les résultats.

1.2 Présentation des données

Origine des données :

Les données utilisées proviennent de la base de données *OpenFlights*, une ressource open-source largement utilisée pour les analyses de réseaux aériens. Cette base contient des informations sur les aéroports et les liaisons directes entre eux. Bien que simplifiées (par exemple, elles n'incluent pas les données sur la fréquence des vols ou les flux de passagers), ces données offrent une base solide pour modéliser et analyser la structure fondamentale du transport aérien mondial.

Description des données :

Les données se présentent sous la forme d'une liste d'arêtes, où chaque ligne représente une connexion directe entre deux aéroports. Les colonnes principales sont :

- **Source** : L'identifiant du premier aéroport (nœud source).
- **Destination** : L'identifiant du second aéroport (nœud destination).
- **Poids (optionnel)** : Représentant l'importance de la connexion (par exemple, fréquence des vols ou capacité des passagers). Si le poids n'est pas spécifié, il est par défaut égal à 1.

Nettoyage et prétraitement :

Les données brutes contenaient des lignes invalides ou redondantes. Voici les étapes de nettoyage effectuées :

- **Suppression des doublons** : Les connexions bidirectionnelles ont été normalisées en une seule arête non orientée.
- **Suppression des lignes invalides** : Les enregistrements contenant des identifiants ou des poids manquants ont été exclus.
- **Conversion des identifiants des nœuds en entiers** : Pour une meilleure gestion et compatibilité avec les outils d'analyse de graphes.

Statistiques clés après nettoyage :

Les données finales utilisées pour construire le graphe présentent les caractéristiques suivantes :

- **Nombre de nœuds** : 2939
Chaque nœud représente un aéroport unique, qu'il s'agisse de petits aéroports régionaux ou de grands hubs internationaux.
- **Nombre d'arêtes** : 15677
Ce nombre correspond aux connexions directes entre aéroports après suppression des doublons et des lignes invalides.
- **Densité** : 0.0036
La densité du graphe indique que seulement **0.36%** des connexions possibles sont présentes, reflétant la structure clairesemée typique des réseaux de transport.

Illustration des données :

Les caractéristiques des données peuvent être visualisées à l'aide d'un histogramme de distribution des degrés (voir section 2.2) .

Ces données, bien qu'incomplètes, permettent de mener une analyse approfondie des propriétés du réseau, d'identifier les hubs stratégiques et de comprendre les communautés régionales ou fonctionnelles.

II Exploration des données

L'exploration des données est une étape essentielle pour comprendre les caractéristiques fondamentales du réseau et poser les bases des analyses suivantes. Cette section propose une analyse des statistiques descriptives, de la distribution des degrés et de la connectivité globale du graphe.

2.1 Statistiques descriptives

Caractéristiques globales :

Le graphe construit à partir des données OpenFlights possède les propriétés suivantes :

- **Nombre de nœuds** : 2939
Chaque nœud représente un aéroport, allant de petits aéroports régionaux à des hubs internationaux majeurs.
- **Nombre d'arêtes** : 30501
Ce nombre correspond aux connexions directes entre aéroports. Les vols de correspondance ne sont pas pris en compte dans cette analyse.
- **Densité** : 0.0036
La densité est le rapport entre le nombre d'arêtes présentes et le nombre maximal d'arêtes possibles. Une densité aussi faible reflète un réseau très clairsemé, typique des réseaux de transport mondial.
- **Nombre de composantes connexes** : 11
Une composante connexe est un sous-ensemble du graphe où tous les nœuds sont connectés entre eux, directement ou indirectement. Dans ce graphe, la composante géante regroupe plus de **95% des nœuds**, indiquant une forte interconnexion globale.
- **Degré moyen** : 10.67
En moyenne, chaque aéroport est connecté à environ 11 autres aéroports.
- **Degré maximum** : 242
Ce degré correspond aux hubs internationaux majeurs, qui jouent un rôle central dans la connectivité globale du réseau.

Ces statistiques montrent que le réseau OpenFlights est globalement bien connecté tout en ayant une structure hiérarchique et clairsemée. La composante géante et les hubs sont des caractéristiques clés qui seront explorées dans les sections suivantes.

2.2 Distribution des degrés

La distribution des degrés dans un graphe est une mesure fondamentale pour comprendre la structure et les propriétés du réseau. Elle indique le nombre de connexions (degrés) de chaque nœud et permet d'identifier les nœuds les plus connectés (*hubs*) et ceux faiblement connectés (aéroports isolés).

Observations clés :

- **Majorité à faible degré** : La plupart des nœuds ont un faible degré, ce qui correspond à de petits aéroports avec peu de connexions.
- **Quelques hubs majeurs** : Une minorité de nœuds ont un degré très élevé, jouant le rôle de hubs internationaux qui concentrent la majorité des connexions.
- **Tendance en loi de puissance** : Cette distribution suit une loi de puissance typique des réseaux réels, où une petite proportion de nœuds supporte la majorité des connexions.

La figure 1 illustre l'histogramme de la distribution des degrés. La forte asymétrie dans la distribution met en évidence la hiérarchie du réseau, où un petit nombre de hubs connectent efficacement les différentes parties du graphe.

2.3 Analyse des composantes connexes

Définition et importance :

Les composantes connexes sont des sous-ensembles du graphe où tous les nœuds sont connectés entre eux, directement ou indirectement, mais sans lien avec les nœuds exté-

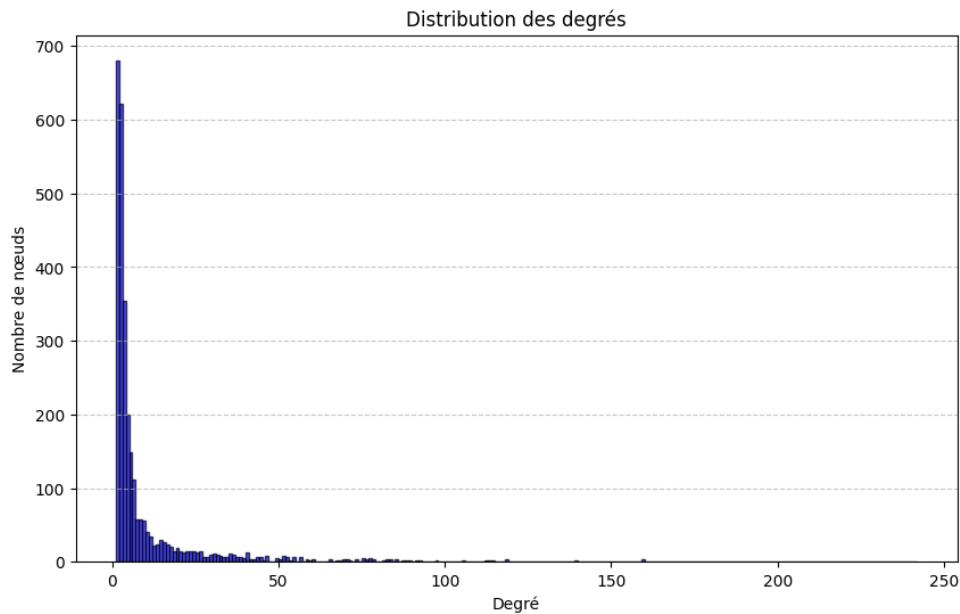


FIGURE 1 – Histogramme de la distribution des degrés. Chaque barre représente le nombre de nœuds ayant un certain degré (nombre de connexions).

rieurs à la composante. Cette analyse est cruciale pour comprendre la connectivité globale et les éventuelles zones isolées dans le réseau.

Caractéristiques principales du réseau OpenFlights :

- **Nombre total de composantes connexes : 11**
Ce chiffre reflète que l'ensemble du réseau n'est pas totalement connecté. Cependant, la majorité des connexions sont regroupées dans une seule composante géante.
- **Composante géante :**
La composante géante regroupe plus de **95% des nœuds**, ce qui montre une forte interconnexion mondiale. Les hubs majeurs jouent un rôle central dans cette interconnexion.
- **Petites composantes :**
Les composantes restantes (10 sur 11) sont significativement plus petites et représentent des réseaux locaux ou isolés. Par exemple, elles pourraient correspondre à des aéroports régionaux ou des îles avec peu ou pas de connexions internationales.

Illustration des composantes connexes :

La figure 2 présente une visualisation des composantes connexes dans le réseau. La composante géante est clairement identifiable par sa taille dominante, tandis que les autres composantes sont nettement plus petites.

Ces observations confirment que le réseau OpenFlights est fortement interconnecté grâce à une composante géante, mais présente également des zones isolées. Cette structure est typique des réseaux de transport mondial, où un noyau centralisé coexiste avec des sous-réseaux régionaux.

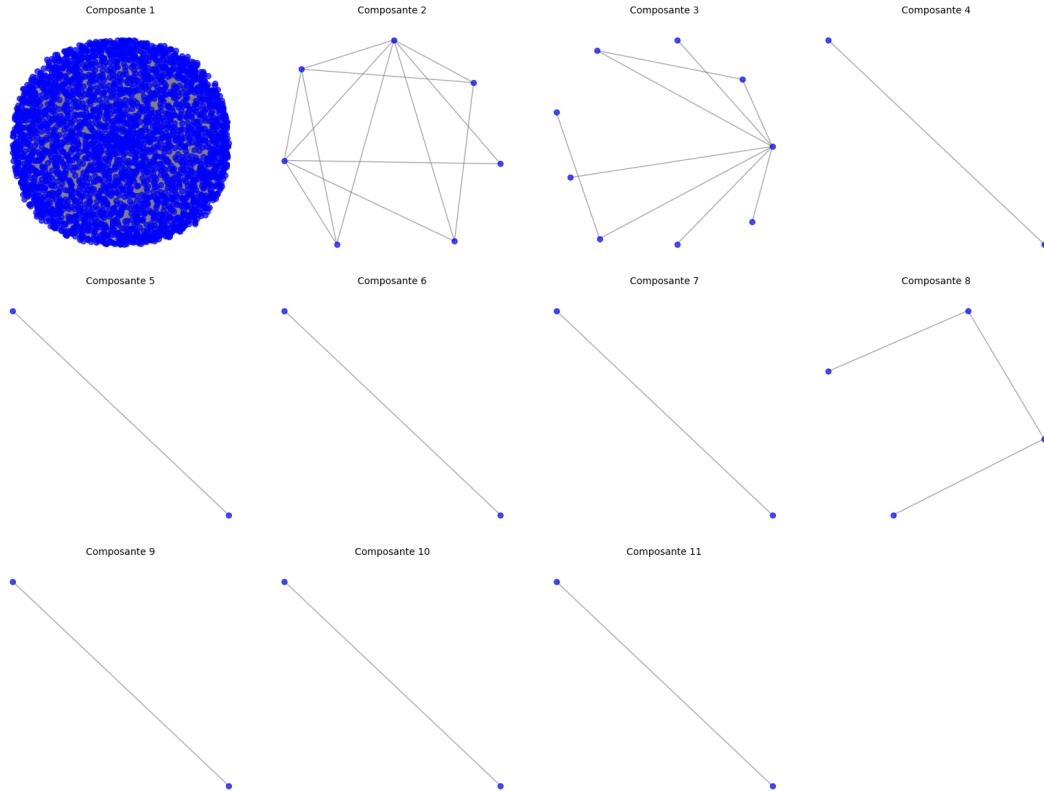


FIGURE 2 – Représentation des composantes connexes dans le réseau OpenFlights. La composante géante regroupe la majorité des nœuds, tandis que les autres composantes sont isolées et plus petites.

III Analyse des propriétés du graphe

L'analyse des propriétés du graphe permet de mieux comprendre la structure du réseau OpenFlights. Cette section examine plusieurs métriques fondamentales pour identifier les nœuds et arêtes critiques, mesurer les propriétés locales comme le clustering, et explorer les propriétés globales comme le diamètre et les plus courts chemins.

3.1 Centralité des nœuds

La centralité des nœuds est un ensemble de mesures permettant d'identifier les aéroports les plus influents ou stratégiques dans le réseau. Les trois principales mesures utilisées dans cette analyse sont :

1. **Centralité de degré** : Indique le nombre de connexions directes d'un nœud. Les aéroports ayant un degré élevé sont généralement des hubs majeurs.
2. **Centralité d'intermédiarité** : Mesure la fréquence à laquelle un nœud apparaît dans les chemins les plus courts entre deux autres nœuds. Ces nœuds agissent comme des passerelles entre différentes parties du réseau.
3. **Centralité de proximité** : Indique à quel point un nœud est proche de tous les autres nœuds, en termes de distance. Les nœuds avec une proximité élevée permettent un accès rapide au reste du réseau.

Top 10 des nœuds par centralité :

Rang	Degré	Intermédierité	Proximité
1	53 (0.0824)	53 (0.0696)	53 (0.4010)
2	65 (0.0742)	65 (0.0668)	65 (0.3936)
3	59 (0.0718)	1576 (0.0579)	57 (0.3921)
4	56 (0.0585)	308 (0.0517)	59 (0.3878)
5	259 (0.0579)	312 (0.0500)	308 (0.3852)
6	297 (0.0575)	134 (0.0494)	69 (0.3834)
7	411 (0.0572)	69 (0.0479)	51 (0.3770)
8	308 (0.0565)	59 (0.0442)	223 (0.3752)
9	366 (0.0545)	51 (0.0425)	134 (0.3738)
10	57 (0.0541)	259 (0.0425)	297 (0.3709)

TABLE 1 – Top 10 des nœuds par centralité de degré, d’intermédierité et de proximité.

Ces résultats montrent que le nœud 53 est un hub majeur avec des scores élevés dans toutes les mesures de centralité. D’autres nœuds, comme 1576 et 69, apparaissent stratégiques grâce à leur intermédierité, malgré un degré plus faible.

3.2 Coefficient de clustering

Le coefficient de clustering mesure la probabilité que deux voisins d’un nœud soient connectés entre eux. Cela reflète la densité locale des connexions, essentielle pour évaluer la redondance et la résilience des sous-réseaux.

— **Définition :**

- *Coefficient de clustering d’un nœud* : Probabilité que deux voisins d’un nœud soient également connectés.
- *Coefficient de clustering moyen* : Moyenne des coefficients de clustering pour tous les nœuds du graphe.

— **Calculs :**

- Nombre total de triangles : 72,852
- Coefficient de clustering moyen : 0.4526

Un coefficient moyen de 0.4526 indique une forte densité locale des connexions, ce qui est typique des réseaux régionaux fortement connectés.

3.3 Analyse des plus courts chemins et du diamètre

Les plus courts chemins et le diamètre sont des métriques globales utilisées pour évaluer l’efficacité de la connectivité dans un réseau.

- **Plus courts chemins** : Distance minimale entre deux nœuds en termes d’arêtes.
- **Distance moyenne** : Moyenne des plus courts chemins entre toutes les paires de nœuds. Elle reflète l’efficacité globale de la connectivité.
- **Diamètre** : Plus grande distance entre deux nœuds dans le réseau, représentant l’étendue maximale du graphe.

Calculs pour la composante géante :

- Distance moyenne : 4.0971
- Diamètre : 14

Ces résultats indiquent que le réseau OpenFlights est globalement bien connecté, avec une faible distance moyenne facilitant les trajets courts et un diamètre modéré montrant

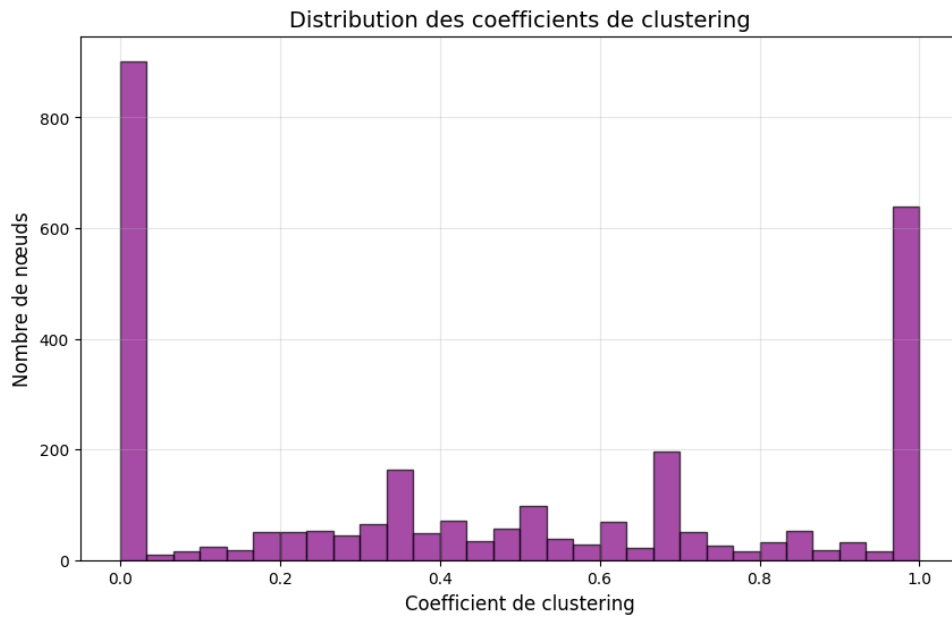


FIGURE 3 – Visualisation des coefficients de clustering. Chaque point représente un nœud, et sa position reflète son coefficient de clustering local.

une portée efficace du réseau.

3.4 Analyse des poids des arêtes

Les poids des arêtes dans le graphe reflètent l'importance relative ou la fréquence des connexions entre les aéroports. Cette section analyse la répartition des poids pour mieux comprendre les caractéristiques des connexions dans le réseau.

- **Nombre d'arêtes** : 15,677
- **Poids minimum** : 2.316×10^{-7}
- **Poids maximum** : 0.0215
- **Poids moyen** : 0.0003

Ces résultats montrent une répartition fortement asymétrique, où une majorité des arêtes ont des poids faibles, mais une minorité d'entre elles ont des poids élevés, indiquant des connexions critiques.

Conclusion :

L'analyse des propriétés du graphe révèle une structure hiérarchique et résiliente. Les hubs identifiés par les mesures de centralité jouent un rôle critique dans la connectivité globale, tandis que les métriques locales comme le clustering et les poids des arêtes illustrent la diversité et l'importance des connexions dans le réseau OpenFlights.

IV Analyse des communautés

Les communautés dans un graphe sont des regroupements de nœuds densément connectés entre eux mais faiblement connectés au reste du réseau. Leur détection est essentielle pour comprendre la modularité et l'organisation sous-jacente d'un réseau complexe comme OpenFlights.

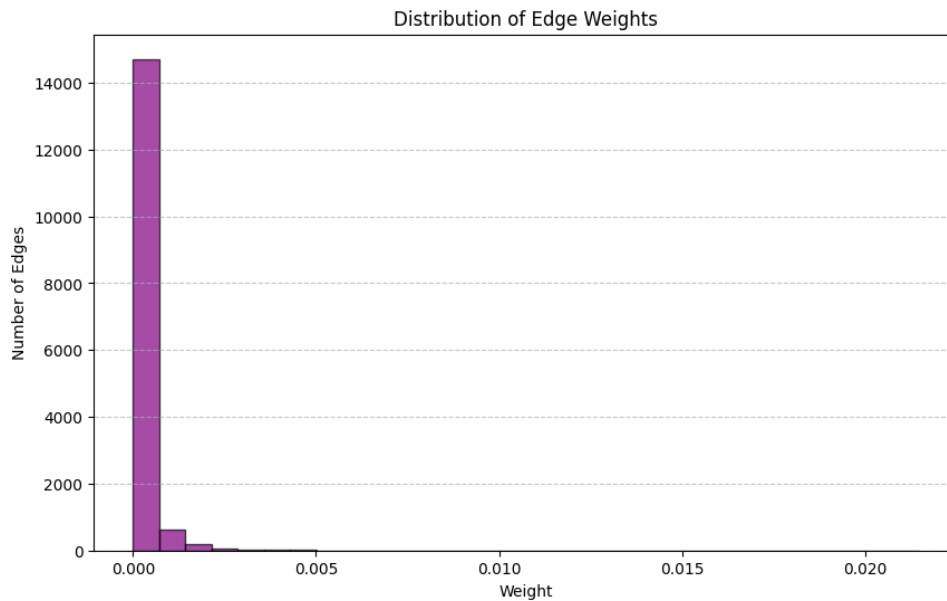


FIGURE 4 – Distribution des poids des arêtes dans le graphe. Les arêtes fortement pondérées représentent des connexions importantes dans le réseau.

4.1 Détection des communautés

Pour détecter les communautés dans le réseau OpenFlights, nous avons utilisé l'algorithme de maximisation de la modularité. Cet algorithme partitionne le graphe en groupes afin de maximiser une métrique appelée *modularité*, qui mesure la densité des connexions au sein des communautés par rapport aux connexions attendues dans un graphe aléatoire équivalent.

Nombre de communautés détectées : 47

- Chaque nœud du graphe a été assigné à une communauté, et les regroupements ont été visualisés.
- L'algorithme utilisé est `greedy modularity communities` de la bibliothèque `networkx`.
- La modularité calculée permet d'évaluer la qualité des partitions détectées.

Visualisation des communautés

La figure 5 présente la visualisation des communautés détectées dans le réseau. Chaque couleur correspond à une communauté distincte. Cependant, en raison de la taille importante du graphe et du chevauchement des nœuds, cette visualisation peut être difficile à interpréter.

- La disposition des nœuds dans le graphe n'est pas optimisée pour une visualisation lisible. L'utilisation d'un algorithme de disposition comme *ForceAtlas2* dans un logiciel comme Gephi pourrait fournir une représentation plus claire.
- Les nœuds des petites communautés se mélangent souvent visuellement avec ceux des grandes communautés, ce qui rend leur identification difficile.
- Malgré cela, cette visualisation offre un aperçu global de la structure communautaire.



FIGURE 5 – Visualisation des communautés détectées. Chaque couleur représente une communauté distincte. La densité du graphe rend certaines communautés difficiles à distinguer.

4.2 Modularité

La modularité est une métrique utilisée pour évaluer la qualité d'une partition d'un graphe en communautés. Elle mesure la densité des connexions au sein des communautés par rapport aux connexions attendues dans un graphe aléatoire équivalent.

Définition : Une **modularité élevée** indique une bonne séparation entre les communautés, où les nœuds d'une même communauté sont fortement connectés entre eux et faiblement connectés au reste du graphe.

La modularité Q est définie par l'équation suivante :

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

où :

- A_{ij} est la matrice d'adjacence (1 si les nœuds i et j sont connectés, 0 sinon),
- k_i, k_j sont les degrés des nœuds i et j ,
- m est le nombre total d'arêtes dans le graphe,
- $\delta(c_i, c_j)$ vaut 1 si i et j appartiennent à la même communauté, sinon 0.

Résultat du calcul de modularité

Le calcul de la modularité pour le réseau OpenFlights donne une valeur élevée, ce qui confirme une bonne séparation des communautés. Cela signifie que les aéroports d'une

même communauté sont fortement connectés entre eux, tandis que les connexions avec d'autres communautés sont rares.

Nombre de communautés détectées : 47

Modularité calculée : 0.6534

Visualisation des communautés détectées



FIGURE 6 – Représentation graphique de la modularité dans le réseau OpenFlights. La modularité élevée reflète une bonne séparation des communautés.

Interprétations :

- Une modularité de 0.6534 est considérée comme élevée pour un réseau de cette taille, indiquant une structure communautaire bien définie.
- Les hubs intercommunautaires jouent un rôle essentiel en connectant différentes communautés et en maintenant la connectivité globale.
- Les regroupements détectés peuvent correspondre à des régions géographiques (par exemple, des continents) ou à des regroupements fonctionnels (par exemple, des hubs et leurs aéroports affiliés).

V Analyse de robustesse

L'analyse de robustesse d'un réseau vise à évaluer sa résilience face à des perturbations, telles que la suppression de nœuds ou d'arêtes. Cette section explore comment le réseau OpenFlights réagit à des scénarios de pannes aléatoires et ciblées, en observant l'évolution de la taille relative de la composante géante.

5.1 Simulation de pannes aléatoires

Les pannes aléatoires simulent la suppression de nœuds au hasard dans le graphe. Ce scénario est courant dans les réseaux réels, où des défaillances aléatoires peuvent se produire sans ciblage spécifique (par exemple, fermetures d'aéroports temporaires dues à des mauvaises conditions météorologiques ou des contraintes opérationnelles).

Méthodologie

1. Sélectionner des nœuds au hasard pour les supprimer un par un.
2. Après chaque suppression, recalculer la taille relative de la composante géante (proportion de nœuds toujours connectés).
3. Répéter la simulation plusieurs fois pour obtenir une moyenne représentative de l'impact.

Cette méthode permet de modéliser des pannes non intentionnelles et de comprendre la résilience globale du réseau face à des défaillances non ciblées.

Résultats

La figure 7 montre l'évolution de la taille relative de la composante géante en fonction du pourcentage de nœuds supprimés aléatoirement. Les résultats révèlent que le réseau OpenFlights reste relativement robuste face à ces pannes aléatoires.

- La taille de la composante géante diminue lentement, même après la suppression de 20% des nœuds.
- Cette résilience est caractéristique des réseaux réels où les connexions locales et la redondance permettent de maintenir la connectivité globale.

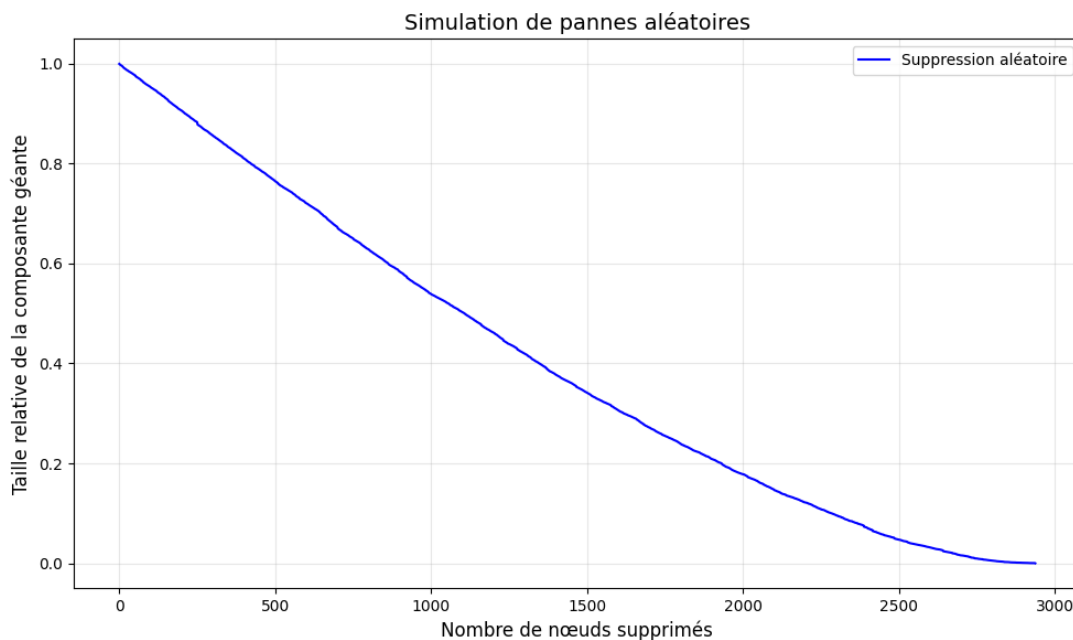


FIGURE 7 – Simulation de pannes aléatoires dans le réseau OpenFlights. La taille relative de la composante géante diminue progressivement avec la suppression de nœuds aléatoires.

5.2 Simulation de pannes ciblées

Les pannes ciblées simulent la suppression des nœuds les plus critiques dans le réseau, en les sélectionnant par ordre décroissant de leur degré. Cette approche reflète un scénario où des attaques ou des défaillances intentionnelles visent à perturber le fonctionnement global du réseau en supprimant ses hubs majeurs.

Méthodologie

1. Identifier les nœuds critiques, triés par ordre décroissant de leur degré (les hubs).
2. Supprimer les nœuds un par un, en commençant par ceux ayant le plus grand degré.
3. Observer et mesurer l'impact sur la taille relative de la composante géante après chaque suppression.

Cette approche est particulièrement utile pour comprendre la vulnérabilité du réseau face à des pannes ciblées ou des attaques intentionnelles visant ses points névralgiques.

Résultats

La figure 8 illustre l'impact des pannes ciblées sur la taille relative de la composante géante. Contrairement aux pannes aléatoires, la suppression des hubs entraîne une fragmentation rapide du réseau.

- La taille de la composante géante diminue drastiquement dès la suppression des premiers nœuds critiques.
- Après la suppression de seulement 5% des nœuds les plus connectés, la composante géante est réduite de moitié.
- Cela souligne la dépendance du réseau OpenFlights à un petit nombre de nœuds stratégiques (les hubs).

5.3 Analyse de la résilience du réseau

L'analyse de résilience permet d'évaluer la robustesse du réseau OpenFlights face à deux scénarios de suppression de nœuds :

- **Suppression aléatoire** : Reflétant des perturbations imprévues, comme des fermetures temporaires d'aéroports.
- **Suppression ciblée** : Reflétant des attaques intentionnelles ou des perturbations affectant les hubs majeurs, identifiés par leur degré.

Cette analyse se concentre sur l'évolution de la taille relative de la composante géante après la suppression progressive de nœuds.

Méthodologie

Deux scénarios de suppression de nœuds ont été simulés :

1. **Suppression aléatoire** : Les nœuds sont supprimés dans un ordre aléatoire, simulant des perturbations imprévues.

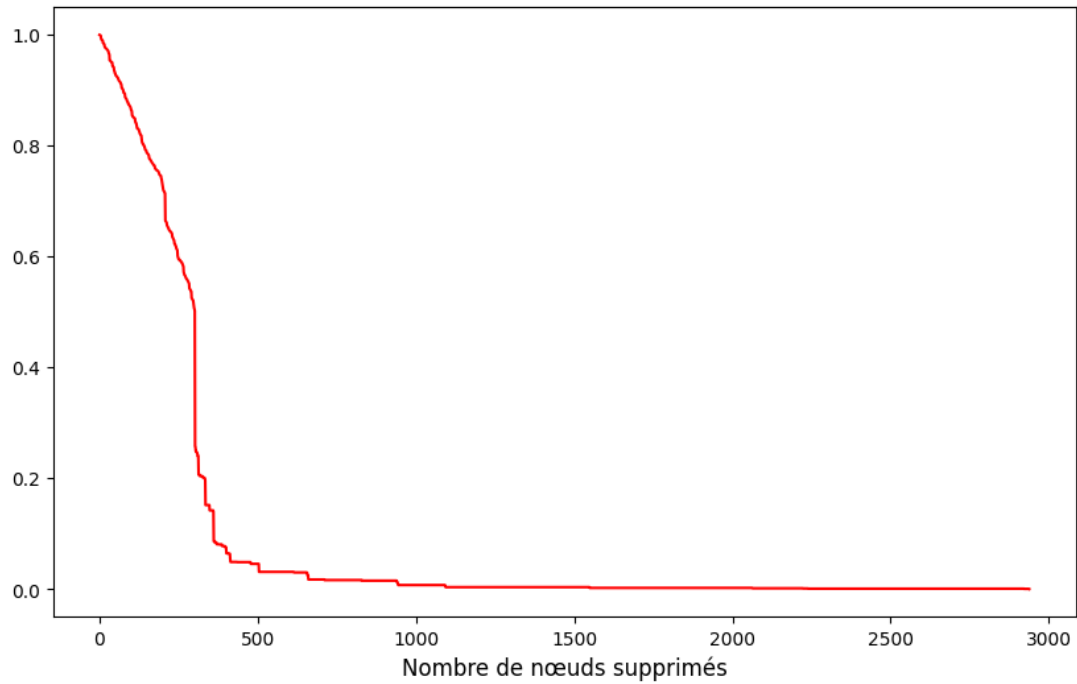


FIGURE 8 – Simulation de pannes ciblées dans le réseau OpenFlights. La suppression des nœuds les plus connectés réduit rapidement la taille de la composante géante.

2. **Suppression ciblée** : Les nœuds sont supprimés par ordre décroissant de degré, simulant une attaque ciblant les hubs les plus connectés.

Pour chaque scénario, la taille de la composante géante (par rapport à sa taille initiale) est calculée après chaque suppression, ce qui permet de suivre l'évolution de la connectivité globale.

Résultats et analyse

La figure 9 présente les résultats de cette simulation :

- **Suppression aléatoire** : La taille de la composante géante diminue progressivement, indiquant une résilience élevée du réseau face à des pannes aléatoires. Cela est typique des réseaux complexes où les nœuds périphériques sont majoritairement affectés en premier.
- **Suppression ciblée** : La taille de la composante géante chute drastiquement dès les premières suppressions. Cela montre une dépendance critique aux hubs majeurs pour maintenir la connectivité globale.

Implications

Cette analyse met en évidence deux points clés :

- **Robustesse aux pannes aléatoires** : Le réseau reste globalement résilient face à des suppressions aléatoires, grâce à sa structure hiérarchique où la plupart des connexions sont centralisées autour de quelques hubs.
- **Vulnérabilité aux attaques ciblées** : La suppression des nœuds les plus connectés fragmente rapidement le réseau, isolant des sous-ensembles de nœuds. Cette

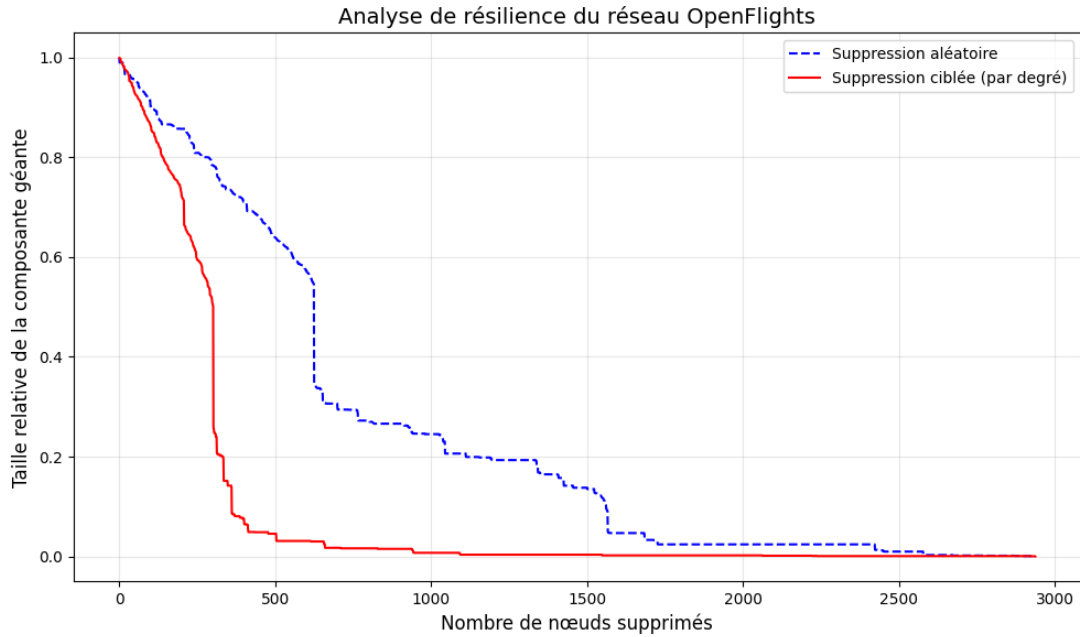


FIGURE 9 – Analyse de la résilience du réseau OpenFlights. La suppression ciblée des hubs entraîne une fragmentation rapide, contrairement à la suppression aléatoire où le réseau reste plus résilient.

vulnérabilité souligne l'importance stratégique des hubs majeurs et la nécessité de diversifier les connexions.

Conclusion

Les analyses de robustesse mettent en évidence deux comportements distincts :

- Le réseau OpenFlights est résilient face aux pannes aléatoires grâce à sa structure distribuée et à la redondance locale des connexions.
- Cependant, il est vulnérable aux pannes ciblées qui affectent les hubs majeurs, confirmant que ces nœuds jouent un rôle central dans la connectivité globale.

Ces résultats soulignent l'importance de diversifier les connexions et de renforcer la résilience des hubs critiques pour améliorer la robustesse globale du réseau.

VI Visualisation

La visualisation d'un réseau est une étape clé pour explorer ses structures, ses communautés et ses hubs. Cette section présente une visualisation globale du réseau OpenFlights et analyse des sous-parties clés pour mettre en évidence les caractéristiques spécifiques du graphe.

6.1 Graphe global

La visualisation globale du graphe offre une perspective générale sur la structure du réseau, mettant en évidence ses principaux hubs, ses clusters régionaux et ses connexions principales. Elle permet de mieux comprendre la hiérarchie et la modularité du réseau.

Méthodologie

1. Les nœuds sont représentés avec des tailles proportionnelles à leur degré, permettant d'identifier visuellement les hubs majeurs.
2. Une disposition de type *Spring layout* a été utilisée pour optimiser l'espacement entre les nœuds et minimiser les chevauchements.
3. Les nœuds sont colorés selon leur appartenance à des communautés, détectées par l'algorithme de modularité.

Ces choix méthodologiques permettent une représentation lisible et intuitive des connexions et des clusters au sein du réseau OpenFlights.

Résultat

La figure 10 présente la visualisation globale du réseau OpenFlights, réalisée dans Gephi. Les nœuds sont colorés par communauté et dimensionnés en fonction de leur centralité de degré.

Cette visualisation met en évidence les principaux hubs du réseau, tels que les aéroports internationaux majeurs, ainsi que les clusters régionaux fortement connectés.

6.2 Sous-parties clés

Pour mieux analyser des zones spécifiques du réseau, nous examinons des sous-parties clés, telles que les hubs majeurs, les clusters régionaux et les zones faiblement connectées. Ces analyses localisées permettent de mieux comprendre le rôle de ces sous-structures dans la connectivité globale.

Méthodologie

1. **Extraction des sous-parties :**
 - Identifier des sous-graphes basés sur des critères spécifiques, comme la centralité des nœuds, leur appartenance à une communauté, ou leur degré.
 - Par exemple, isoler les hubs majeurs et leurs connexions locales pour analyser leur rôle dans la structure globale.
2. **Visualisation localisée :**
 - Tracer uniquement les nœuds et arêtes pertinents dans les sous-parties pour une analyse détaillée.
 - Mettre en évidence les connexions locales et les interactions avec le reste du réseau.

Résultat

Ces visualisations montrent :

- Des hubs majeurs et leurs connexions locales, illustrant leur rôle central dans le réseau.
- Des clusters régionaux densément connectés, mettant en évidence des zones géographiques ou fonctionnelles spécifiques.



FIGURE 10 – Visualisation globale du réseau OpenFlights réalisée dans Gephi. Les nœuds représentent les aéroports, et les arêtes indiquent les connexions directes entre eux. Les couleurs distinguent les communautés détectées, et les tailles des nœuds reflètent leur degré.

— Des zones isolées ou peu connectées, soulignant les limites de la connectivité globale. Ces analyses montrent comment les sous-structures locales contribuent à la robustesse et à l'efficacité du réseau global, tout en soulignant les points faibles potentiels.

Conclusion

La visualisation globale et l'analyse des sous-parties clés permettent de mieux comprendre la structure hiérarchique et modulaire du réseau OpenFlights. Ces représentations révèlent les rôles critiques des hubs majeurs et des clusters régionaux dans la connectivité mondiale, ainsi que les défis liés aux zones faiblement connectées.

VII Conclusion

L'analyse du réseau OpenFlights a permis de mettre en lumière plusieurs aspects essentiels de sa structure et de son fonctionnement. Les résultats obtenus mettent en évidence une organisation hiérarchique, modulaire, et vulnérable aux pannes ciblées.

7.1 Résumé des observations clés

1. Propriétés globales :

- Le réseau est clairsemé, avec une densité de connexions faible mais une composante géante qui regroupe la majorité des nœuds.
- La structure hiérarchique est dominée par des hubs critiques assurant la connectivité globale.

2. Centralité et communautés :

- Les mesures de centralité identifient des hubs stratégiques jouant un rôle clé dans l'interconnexion des clusters régionaux.
- Les communautés détectées révèlent des regroupements géographiques ou fonctionnels avec une modularité élevée.

3. Robustesse et résilience :

- Le réseau est robuste face à des pannes aléatoires grâce aux connexions redondantes dans les clusters.
- Cependant, il est vulnérable aux pannes ciblées, où la suppression des hubs fragmente rapidement la composante géante.

7.2 Implications pratiques

1. Optimisation des routes :

- Comprendre les hubs critiques peut aider à renforcer les connexions dans les régions isolées ou à planifier des routes alternatives pour améliorer la résilience.

2. Planification stratégique :

- Les analyses de modularité et de clustering peuvent guider les décisions pour optimiser la gestion des flux aériens et les infrastructures.

7.3 Perspectives

1. Études supplémentaires :

- Simuler des scénarios spécifiques, comme l'impact de défaillances simultanées dans plusieurs régions.
- Étudier l'évolution du réseau dans le temps pour analyser les tendances.

2. Applications :

- Étendre cette analyse à d'autres types de réseaux, comme les réseaux routiers ou ferroviaires, pour des comparaisons intermodales.
- Utiliser ces observations pour des stratégies d'intervention en cas de perturbations majeures.

Cette analyse démontre l'intérêt de la théorie des graphes et de la modélisation pour comprendre et optimiser des réseaux complexes comme celui d'OpenFlights.