



Università
Ca'Foscari
Venezia

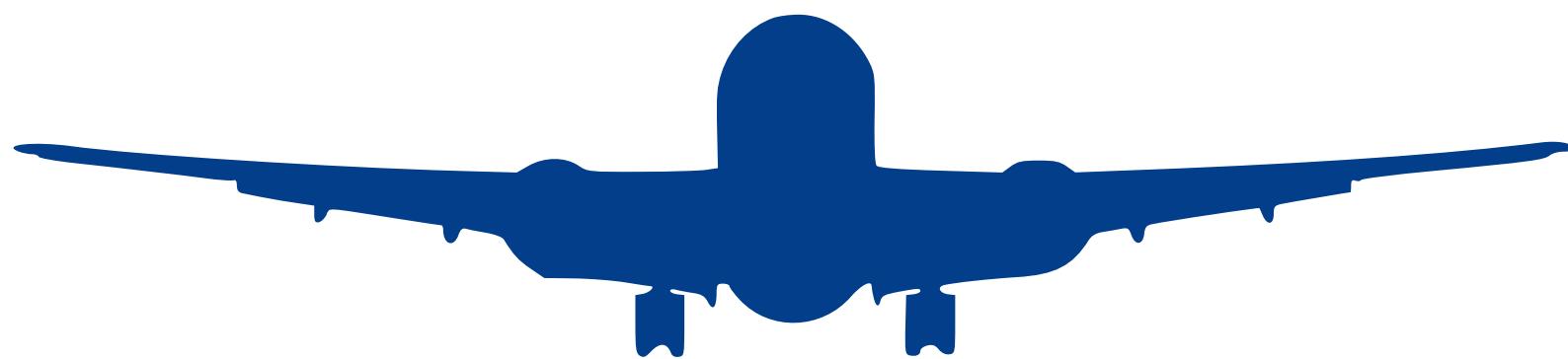
Social Network Analysis

Report

Prepared by
Hakim Haddaoui 892639

Index

Abstract	03
Introduzione	03
• Contesto del problema	
• Obiettivo del progetto	
• Motivazione e rilevanza della domanda di ricerca	
Metodologia	04
• Raccolta ed Analisi dei dati	
• Elaborazione dei dati	
• Tecniche applicate	
• Misure di centralità	
• Strumenti utilizzati	
Processo di Analisi	06
• Prima Fase	
• Seconda Fase	
• Terza Fase	
• Quarta Fase risultati	
Limiti e Bias	16
• limiti dei dati utilizzati	
Bibliografia	17
• Dataset e Fonti Dati	
• Software e Librerie	
• Metodologia e Concetti Teorici	
Conclusioni	18



Abstract

Questa ricerca analizza la rete globale del trasporto aereo per identificare gli aeroporti chiave utilizzando misure di centralità, come betweenness e degree. Il dataset di OpenFlights è stato elaborato per esaminare i principali hub globali e le loro connessioni, evidenziando il ruolo strategico di specifici aeroporti nel facilitare i flussi internazionali. L'analisi ha permesso di individuare due tipologie di aeroporti chiave: quelli che fungono da hub connettivi e quelli che svolgono un ruolo cruciale nel controllo del flusso tra diverse regioni, mostrando come la loro combinazione sia fondamentale per la struttura e l'efficienza della rete globale del trasporto aereo.

Introduzione

Contesto del problema



Il trasporto aereo rappresenta uno degli elementi chiave della mobilità globale, influenzando sia il commercio internazionale che il turismo, la logistica e la comunicazione fra i vari paesi.

Gli aeroporti fungono da nodi vitali nella rete di trasporto globale, e alcuni di essi, noti come **hub**, giocano un ruolo centrale nel facilitare flussi di traffico internazionali.

La posizione geografica di questi hub, insieme alle loro infrastrutture e alle connessioni con altre destinazioni, determina la loro importanza in un sistema globale di trasporto.

Obiettivo del progetto

L'obiettivo principale di questo progetto è identificare gli **aeroporti chiave**, denominati hub nella rete globale di trasporto aereo, analizzando come la loro posizione geografica e le rotte in entrata e uscita da essi influiscono sulla rete del traffico aereo internazionale.

Ho definito un aeroporto chiave come un aeroporto che possiede un elevato volume di traffico oppure funge da "gatekeeper" per altre destinazioni, agendo come ponte strategico tra diverse parti della rete. Senza questi aeroporti, ci sarebbero significative ripercussioni sul flusso globale del traffico aereo, compromettendo la connettività e la stabilità della rete.

Motivazione e rilevanza della domanda di ricerca

La domanda di ricerca "**Quali sono gli aeroporti chiave nella rete globale di trasporto aereo?**" nasce dalla mia passione per i viaggi. Negli anni, ho avuto la fortuna di visitare più di 20 paesi in 5 continenti ed ho sempre trovato affascinante l'idea di esplorare i flussi di traffico aereo e scoprire come alcuni aeroporti siano cruciali nel connettere persone e culture a livello globale.



Metodologia

Raccolta ed Analisi dei dati

Per questa analisi, i dati sono stati raccolti dal dataset **OpenFlights**, una fonte pubblica che fornisce gratuitamente informazioni dettagliate sulle reti aeree. Il dataset include dati sui collegamenti aerei, come i codici IATA degli aeroporti di partenza e arrivo, le coordinate geografiche e i nomi degli aeroporti.

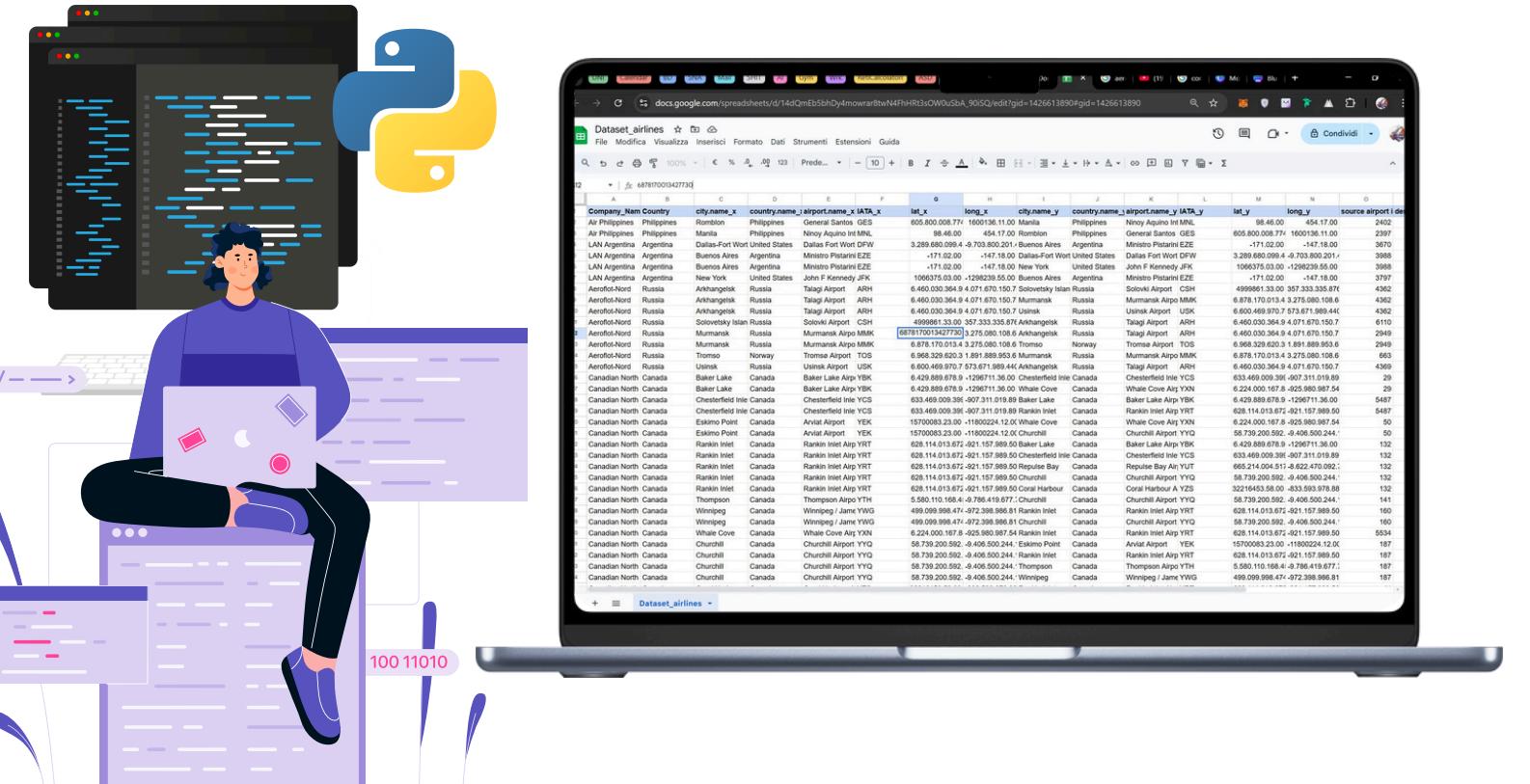
Ho utilizzato tre diversi dataset, integrati e pre-elaborati attraverso tecniche di **data enrichment**, **filtering** e **aggregation**, con l'obiettivo di ottenere un dataset completo e coerente per la mia analisi.

I dati sono stati acquisiti direttamente dalla piattaforma **Kaggle**, che offre una versione ben strutturata e facilmente accessibile del database OpenFlights.



Elaborazione dei dati

L'intera fase di elaborazione dei dati è stata condotta utilizzando il linguaggio di programmazione **Python**, sfruttando librerie come **Pandas** e **NumPy** per la manipolazione dei dati.



Metodologia

Tecniche applicate

Per analizzare i dati, ho applicato alcuni principi della teoria delle reti sociali, in modo da fornirmi un quadro per studiare le interazioni tra i vari aeroporti. In questo contesto, **ogni aeroporto è rappresentato come un nodo e ogni rotta aerea come un arco della rete.**

Un aspetto fondamentale della mia analisi è l'inclusione della **posizione geografica** di ciascun nodo, definita dai suoi attributi di **longitudine e latitudine**. Questo arricchimento ha permesso di analizzare non solo le connessioni tra gli aeroporti, ma anche di **visualizzare la distribuzione spaziale della rete aerea globale**, fornendo una visione più dettagliata delle relazioni geografiche tra i nodi.

Misure di centralità

Ho utilizzato due misure principali di centralità per valutare l'importanza relativa di ciascun nodo nella rete, in modo da identificare gli hub:

- **Degree Centrality**

Questa misura **rappresenta il numero totale di connessioni (entrate e uscite) di un nodo.**

Un valore elevato indica che l'aeroporto è molto connesso e potrebbe fungere da hub continentale o globale.

- **Betweenness Centrality**

Questa misura **riflette il ruolo di un nodo come intermediario nei percorsi più brevi tra altri nodi.**

Un valore elevato indica che il nodo associato all'aeroporto è cruciale per il transito aereo e potrebbe essere un punto chiave in caso di interruzioni di altri nodi.

Un'interruzione di uno di questi aeroporti potrebbe quindi causare significative ripercussioni sui flussi aerei globali.

Strumenti utilizzati

Per la visualizzazione dei dati e la produzione di plot ho utilizzato **Python** e una serie di librerie specifiche che hanno facilitato il processo di elaborazione, analisi e rappresentazione grafica. Le librerie principali utilizzate sono:

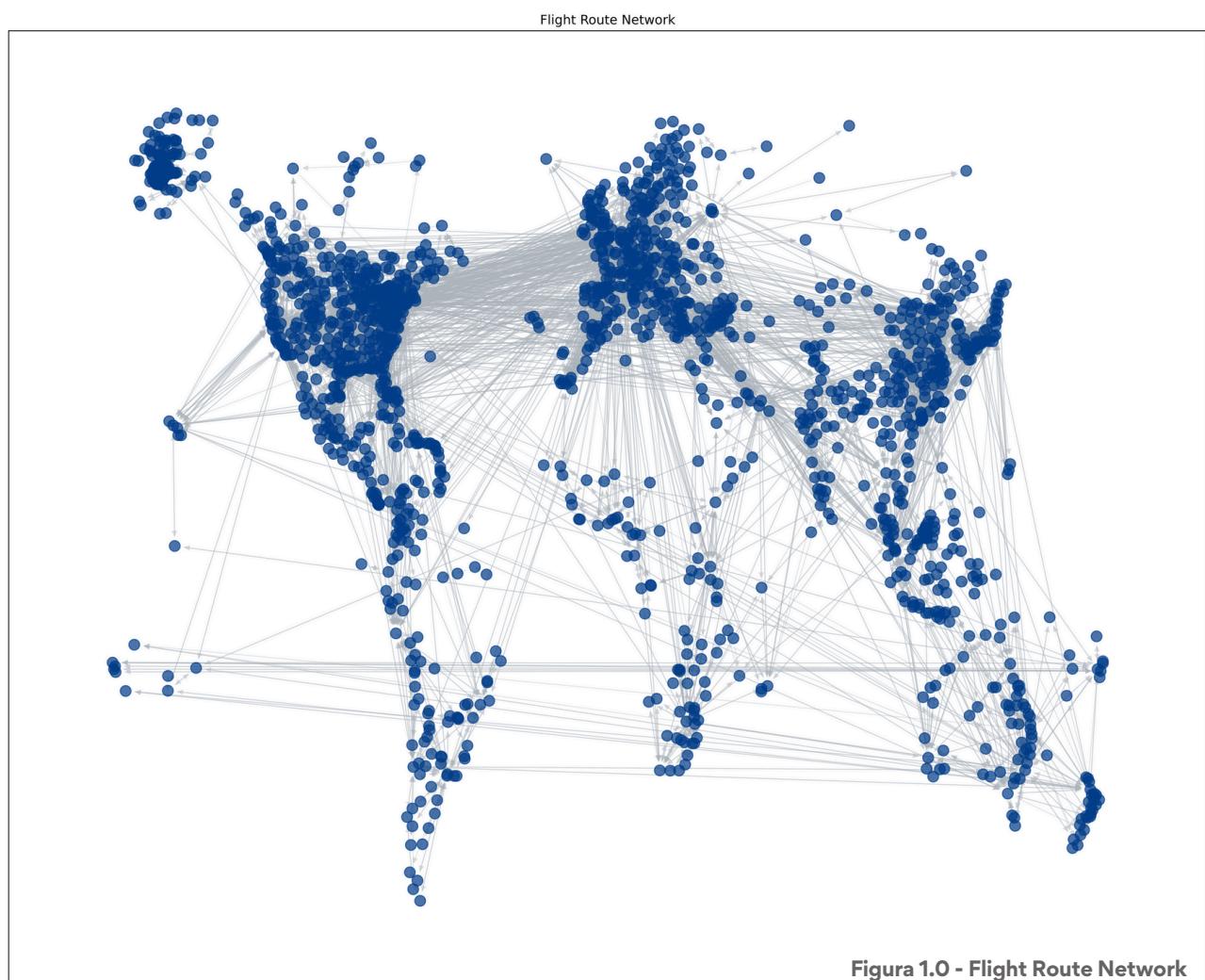
- **NumPy:** per operazioni numeriche efficienti e gestione di Data structure.
- **Pandas:** per la manipolazione e l'analisi dei dati.
- **NetworkX:** per la produzione dei vari plot.
- **Matplotlib:** per la produzione di grafici.
- **Cartopy:** per avere una mappa globale
- **Seaborn:** per la produzione di grafici

Processo di Analisi

Il processo di analisi è stato condotto seguendo una serie di fasi, con l'obiettivo di arrivare passo per passo ad **identificare i principali hub della rete aerea globale**

Prima Fase

La prima fase ha visto la creazione di un **grafo orientato** per rappresentare la rete aerea globale. Ogni aeroporto è stato trattato come un nodo, mentre ogni rotta aerea è stata rappresentata come un arco tra due nodi. A ogni nodo è stato assegnato un attributo che include le coordinate geografiche di longitudine e latitudine il che compone una sorta di Layout al nostro grafo.



Nella figura 1.0 possiamo notare che gli attributi delle coordinate associate ad ogni nodo della rete, sebbene non collegate a una mappa, hanno posto le basi per una rappresentazione spaziale della rete.

Guardando la struttura del grafo, non vi ricorda qualcosa?

Se aggiungiamo una mappa geografica, come mostrato nella figura 1.1, otterremo un grafo ancora più preciso e tangibile della rete aerea globale. In questo modo, le connessioni tra gli aeroporti, potranno essere visualizzate come veri e propri legami tra punti sparsi sulla superficie della Terra, rivelando con maggiore chiarezza la distribuzione geografica dei flussi aerei e la posizione degli aeroporti principali.

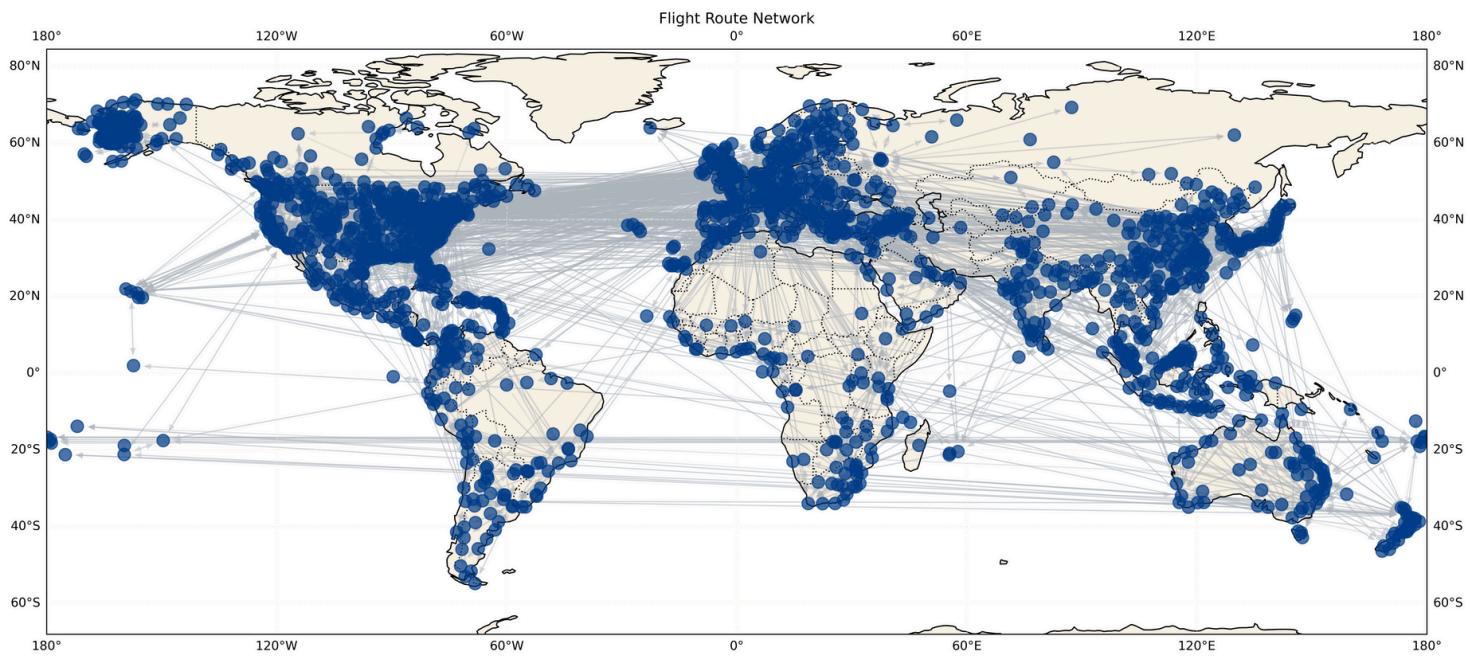


Figura 1.1 - Flight Route Network with world map

Seconda Fase

Dopo aver esaminato la rete globale nel suo complesso, ho scelto di concentrarmi su ciascun continente separatamente per analizzare le rotte da ogni aeroporto e le relative destinazioni verso altri continenti. Questo approccio mi ha permesso di osservare come il traffico aereo è distribuito tra i vari continenti. In particolare, ho studiato il numero e il volume delle rotte in uscita per ciascun continente come primo passo fondamentale per l'identificazione degli aeroporti chiave.

Questa analisi preliminare mi ha permesso di individuare quei continenti che, per il volume nelle rotte, potrebbero ospitare hub strategici.

Continente	Numero di rotte in uscita dal continente	Num. di rotte in entrata nel continente
Africa	85	121
Americhe	695	763
Asia	416	506
Europa	356	499
Oceania	105	129

Africa

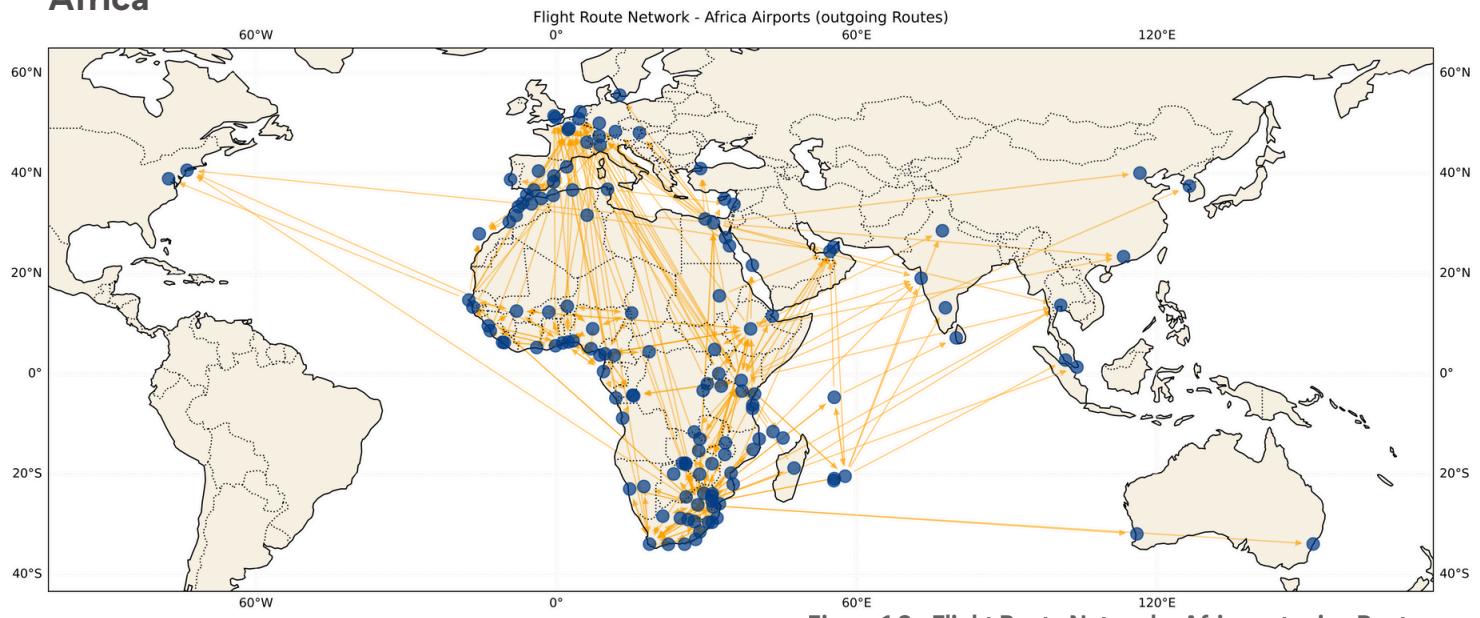


Figura 1.2 - Flight Route Network - Africa outgoing Routes

America

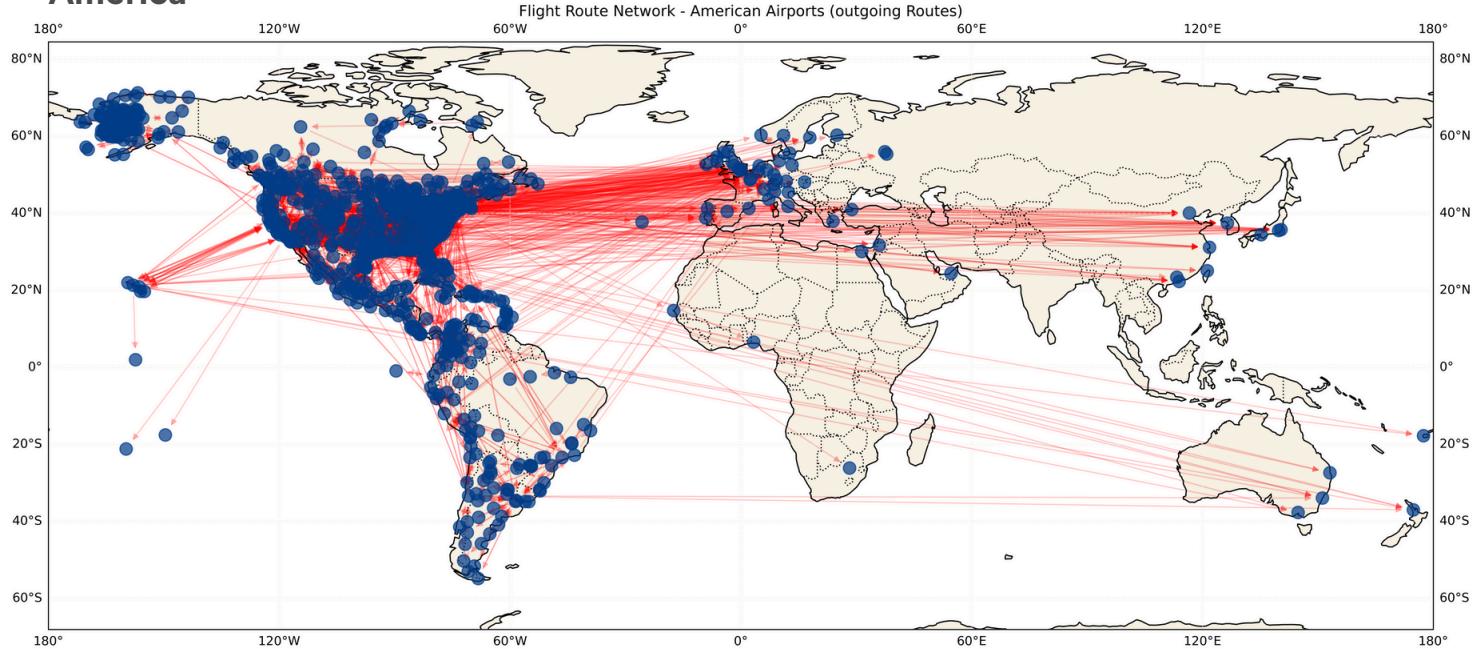


Figura 1.3 - Flight Route Network - America outgoing Routes

Asia

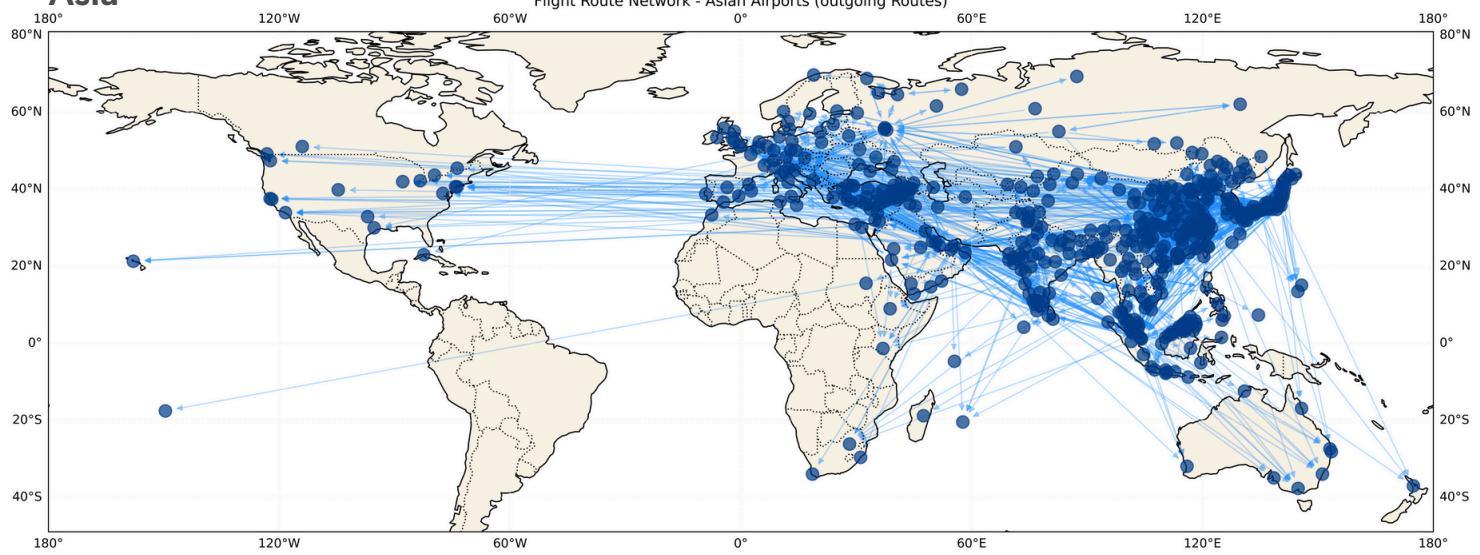


Figura 1.4 - Flight Route Network - Asia outgoing Routes

Europa

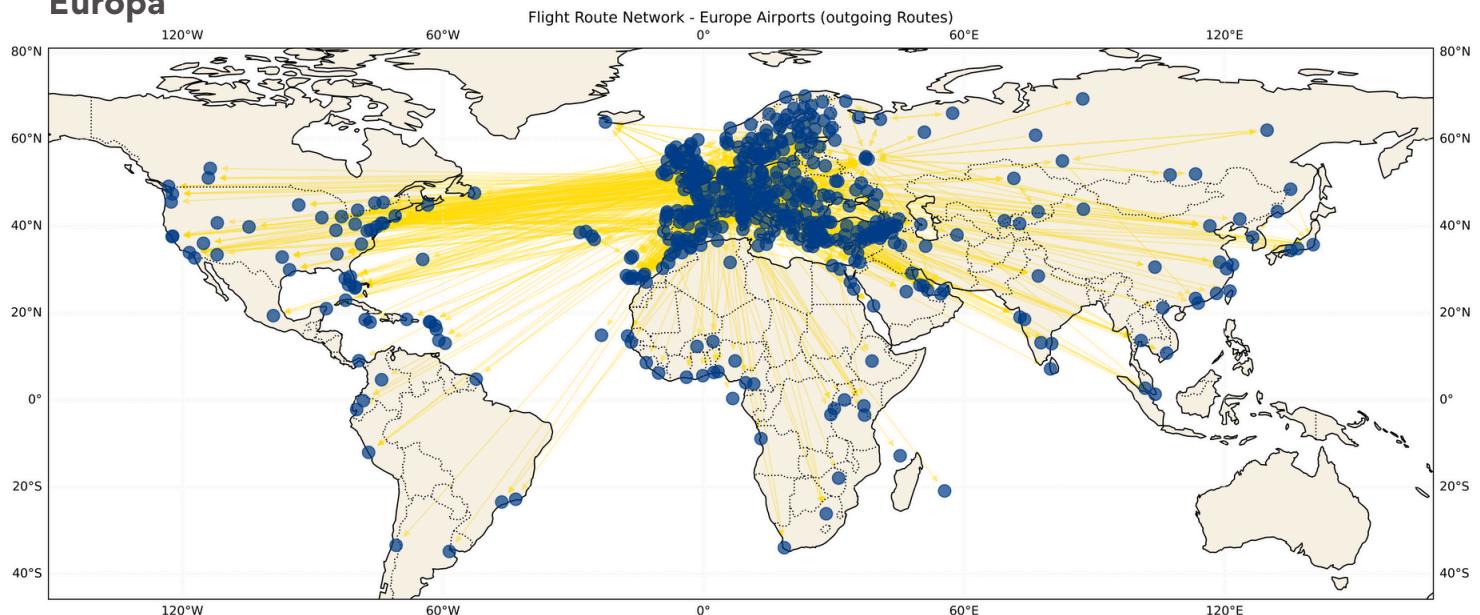


Figura 1.5 - Flight Route Network - Europa outgoing Routes

Oceania

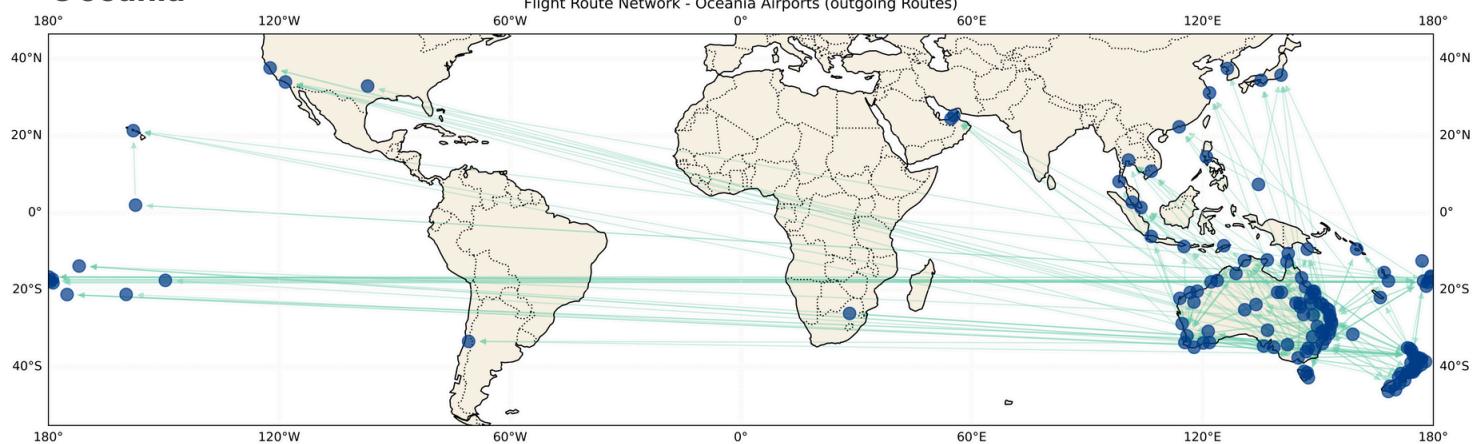


Figura 1.6 - Flight Route Network - Oceania outgoing Routes

Dati i risultati della seconda fase, è emerso che **i principali hub potrebbero trovarsi in America, Europa e Asia, dato il volume delle rotte osservato in questi continenti.**

Per verificare questa ipotesi, nella terza fase ho applicato diverse **misure di centralità** ai nodi del grafo, come descritto in precedenza, al fine di identificare gli aeroporti chiave nella rete e confrontarli con le osservazioni preliminari.

Ogni misura di centralità fornisce una prospettiva diversa sull'importanza di un aeroporto nella rete globale.

Questo approccio mi ha consentito di esaminare in modo più dettagliato il ruolo di ciascun aeroporto all'interno della rete e **verificare se i risultati quantitativi confermano le tendenze osservate nella fase preliminare.**

Terza Fase

Top 10 Airports by Degree Centrality:

1. Airport: **ATL** (Hartsfield Jackson Atlanta International Airport), **Centrality: 0.2151**
2. Airport: **ORD** (Chicago O'Hare International Airport), **Centrality: 0.1815**
3. Airport: **LAX** (Los Angeles International Airport), **Centrality: 0.1394**
4. Airport: **PEK** (Beijing Capital International Airport), **Centrality: 0.1388**
5. Airport: **VIE** (Vienna International Airport), **Centrality: 0.1322**
6. Airport: **DEN** (Denver International Airport), **Centrality: 0.1316**
7. Airport: **LHR** (London Heathrow Airport), **Centrality: 0.1286**
8. Airport: **CDG** (Charles de Gaulle International Airport), **Centrality: 0.1262**
9. Airport: **FRA** (Frankfurt am Main International Airport), **Centrality: 0.1244**
10. Airport: **AMS** (Amsterdam Airport Schiphol), **Centrality: 0.1214**

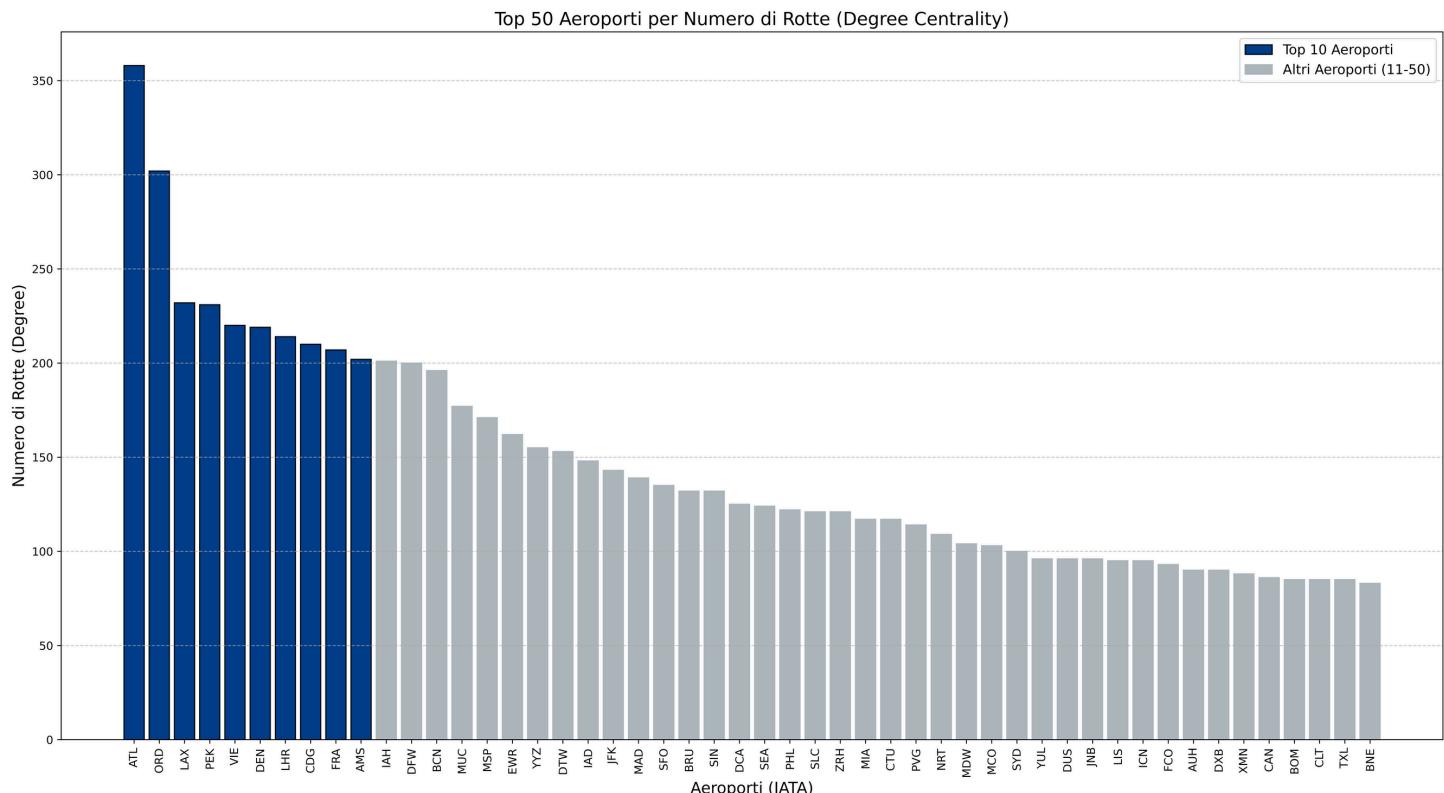


Figura 1.7 - Top 50 Aeroporti con Maggior Numero di Rotte (Entrata/Uscita)

Nella **figura 1.7**, viene mostrata chiaramente la distribuzione del carico di rotte sugli aeroporti.

Si osserva che gli aeroporti **ATL (Hartsfield-Jackson Atlanta International Airport)** e **ORD (Chicago O'Hare International Airport)** presentano un volume significativamente maggiore di rotte sia in entrata che in uscita rispetto agli altri aeroporti.

Degree Centrality

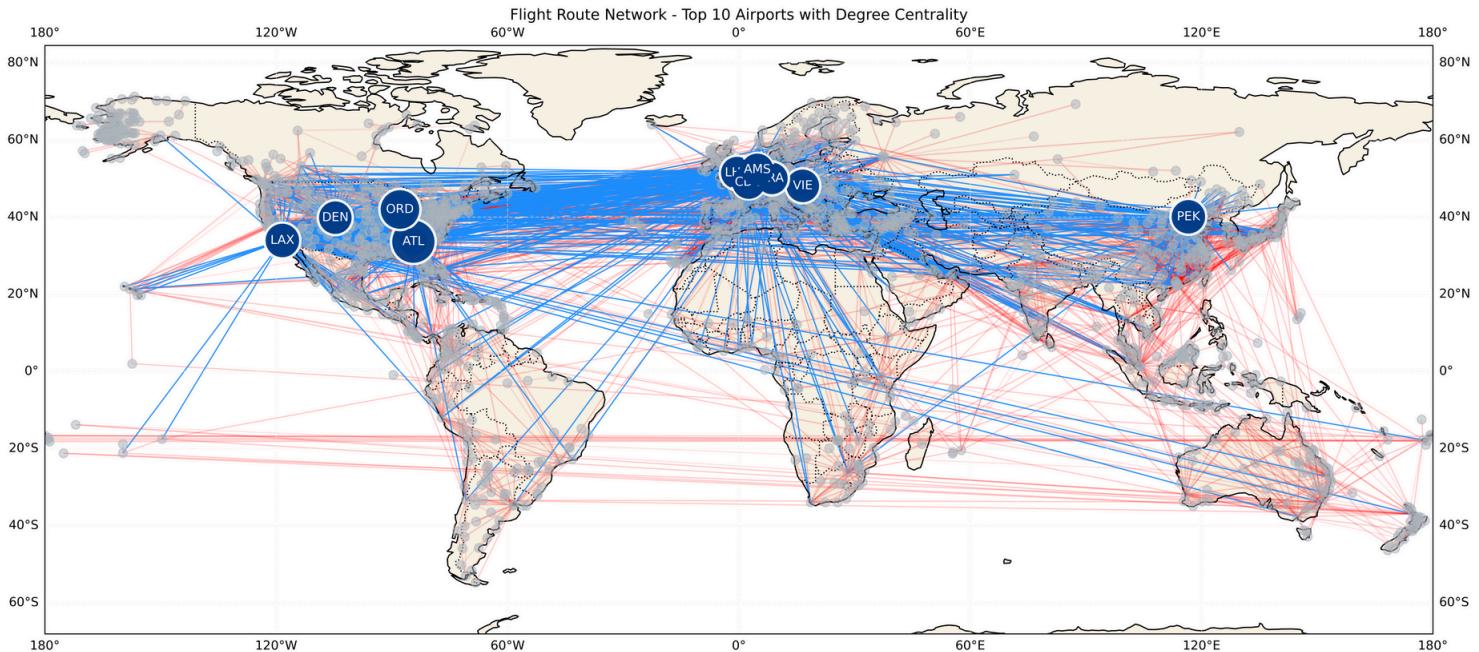


Figura 1.8 - Top 10 Aeroporti con Maggior Numero di Rotte (Entrata/Uscita) in confronto con il resto della rete aerea

Degree centrality only routes from top 10 airports

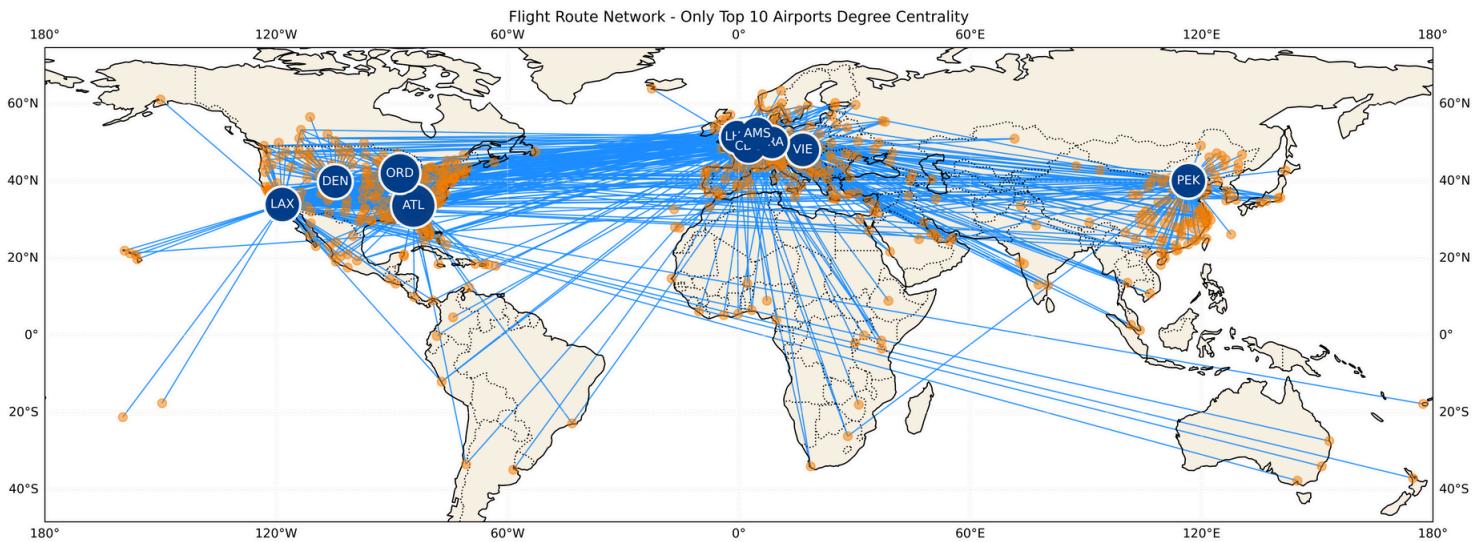


Figura 1.9 - Rotte dai 10 aeroporti con il maggior numero di rotte in entrata e in uscita, insieme alle possibili destinazioni.

I due plot (Figura 1.8 e Figura 1.9) illustrano i risultati ottenuti applicando la funzione di **Degree Centrality** alla rete aerea globale.

La **Figura 1.8** rappresenta l'intera rete, includendo tutti gli aeroporti e le relative rotte. In particolare, vengono evidenziati i principali aeroporti connessi da rotte in entrata e in uscita colorate in blu, mentre le rotte degli altri aeroporti sono colorate di rosso. I 10 aeroporti principali individuati attraverso il calcolo della Degree Centrality riflettono le assunzioni fatte nella fase 2 dell'analisi.

La **Figura 1.9**, invece, mostra esclusivamente i nodi con il grado più elevato, ossia gli aeroporti principali, insieme a tutte le rotte che li collegano agli altri aeroporti. Questo grafico mette in evidenza il carico di connessioni che questi aeroporti principali hanno con gli altri nodi della rete, enfatizzando il loro ruolo cruciale nel connettere la rete aerea globale.

Questo aspetto è particolarmente rilevante per la mia analisi, **in quanto consente di individuare gli aeroporti con il maggior volume di traffico** e di evidenziare la loro capacità di agire come hub a livello nazionale e internazionale.



Top 10 Airports by Betweenness Centrality:

1. Airport: **LAX** (Los Angeles International Airport), **Centrality: 0.1346**
2. Airport: **SIN** (Singapore Changi Airport), **Centrality: 0.0829**
3. Airport: **PEK** (Beijing Capital International Airport), **Centrality: 0.0777**
4. Airport: **LHR** (London Heathrow Airport), **Centrality: 0.0775**
5. Airport: **ANC** (Ted Stevens Anchorage International Airport), **Centrality: 0.0759**
6. Airport: **ATL** (Hartsfield Jackson Atlanta International Airport), **Centrality: 0.0648**
7. Airport: **CDG** (Charles de Gaulle International Airport), **Centrality: 0.0611**
8. Airport: **AMS** (Amsterdam Airport Schiphol), **Centrality: 0.0579**
9. Airport: **ORD** (Chicago O'Hare International Airport), **Centrality: 0.0555**
10. Airport: **FRA** (Frankfurt am Main International Airport), **Centrality: 0.0518**

Betweenness Centrality

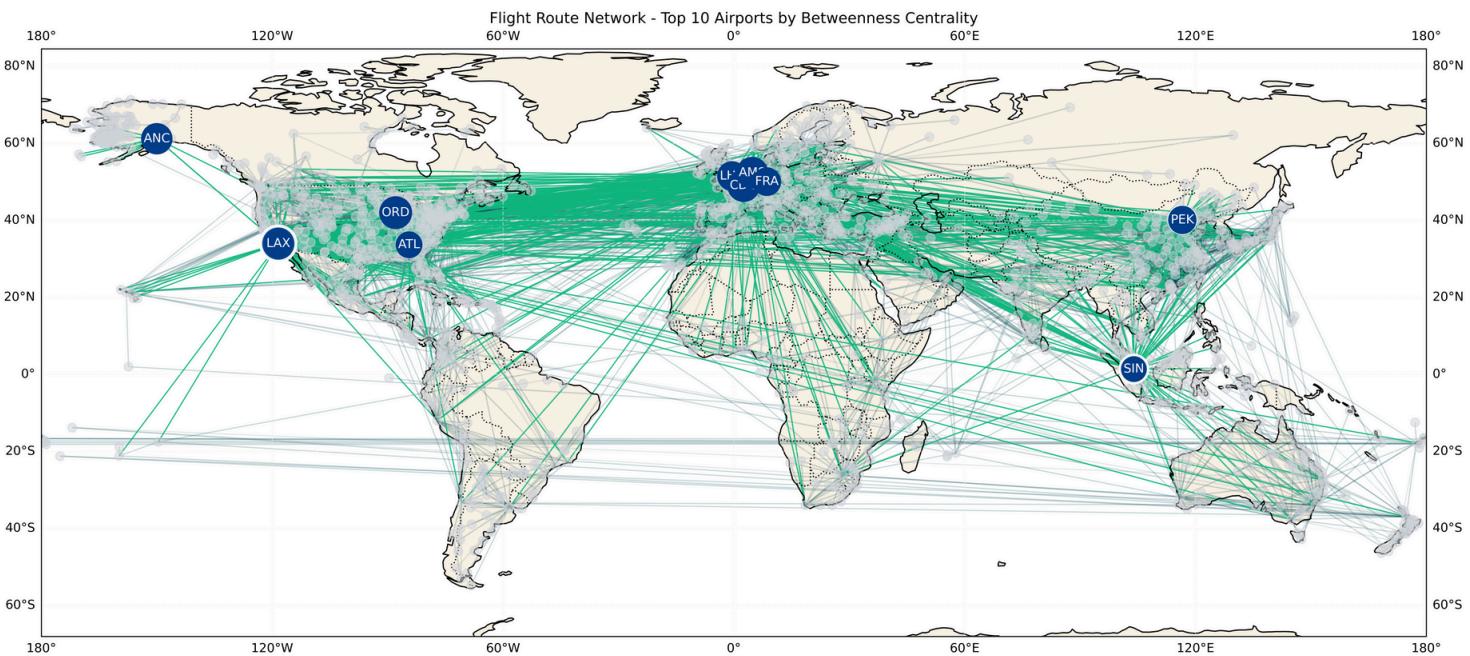


Figura 2.0 - Top 10 Aeroporti con Maggiore valore di Betweenness in confronto con il resto della rete aerea

Betweenness Centrality only routes from top 10 airports

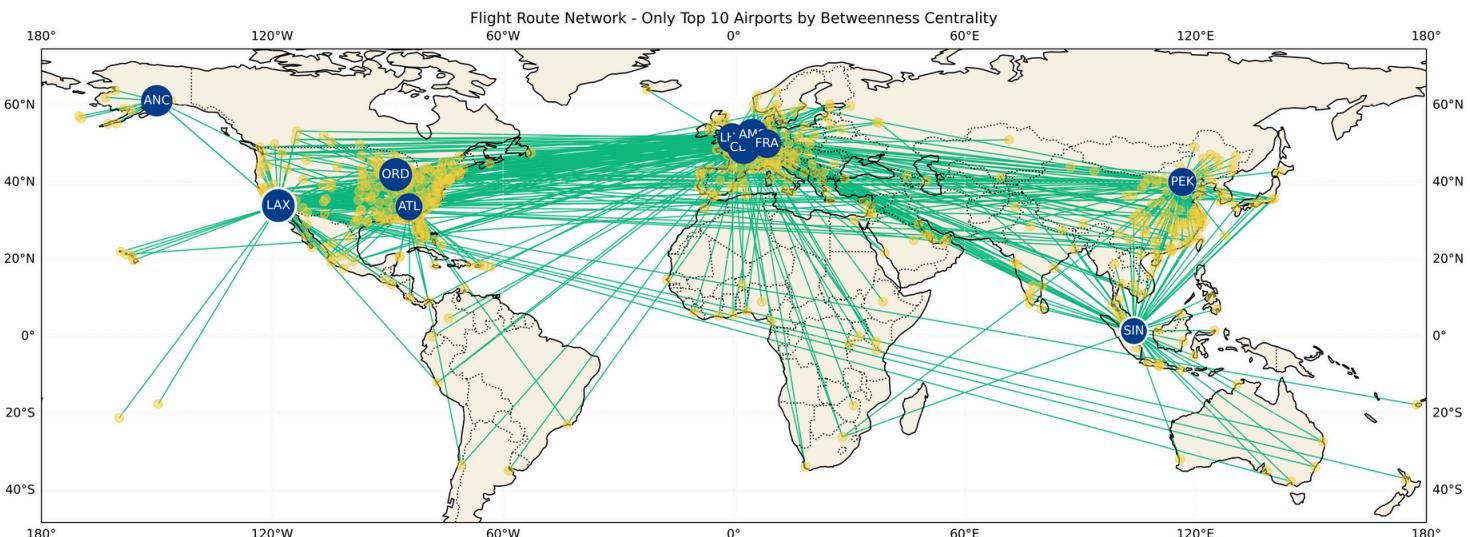


Figura 2.1 - Rotte dai 10 aeroporti con il maggiore valore di Betweenness, insieme alle possibili destinazioni.

I due plot (**Figura 2.0 e Figura 2.1**) mostrano i risultati ottenuti applicando la **Betweenness Centrality** alla rete aerea globale.

Questa metrica identifica gli aeroporti che fungono da ponti strategici nella rete, svolgendo un ruolo cruciale nel connettere diverse regioni geografiche e nel garantire un flusso continuo del traffico aereo.

La **Figura 2.0** rappresenta l'intera rete, includendo tutti gli aeroporti e le relative rotte. In questo grafico, i principali hub con elevata Betweenness Centrality sono messi in evidenza, permettendo di individuare gli aeroporti che occupano posizioni strategiche per il transito tra diverse aree geografiche.

La **Figura 2.1**, invece, mostra esclusivamente i nodi con i punteggi di Betweenness Centrality più alti. Questo grafico si concentra sugli aeroporti cardine che agiscono come punti di transito essenziali, enfatizzando il loro ruolo nella connettività globale della rete.

Rispetto ai risultati ottenuti con la Degree Centrality, si osservano alcune corrispondenze, poiché diversi aeroporti altamente connessi appaiono anche tra quelli con una Betweenness Centrality elevata.

Tuttavia, emergono alcune differenze: aeroporti con molte connessioni dirette potrebbero avere un punteggio di Betweenness Centrality basso se non svolgono un ruolo centrale nel transito tra nodi distanti. Questo perché la Betweenness Centrality misura il grado in cui un aeroporto funge da intermediario nelle rotte, enfatizzando la sua importanza nel collegare aeroporti distanti o regioni diverse.

Di conseguenza, questa misura è particolarmente utile per identificare quegli aeroporti che, pur non avendo un volume di traffico elevato, sono cruciali per la robustezza e la resilienza della rete aerea globale.



Quarta Fase risultati

Con l'applicazione delle misure di centralità, ho elaborato una lista dei 10 aeroporti più importanti a livello globale per ogni misura. Questi aeroporti sono stati selezionati in base alla loro centralità nella rete aerea.

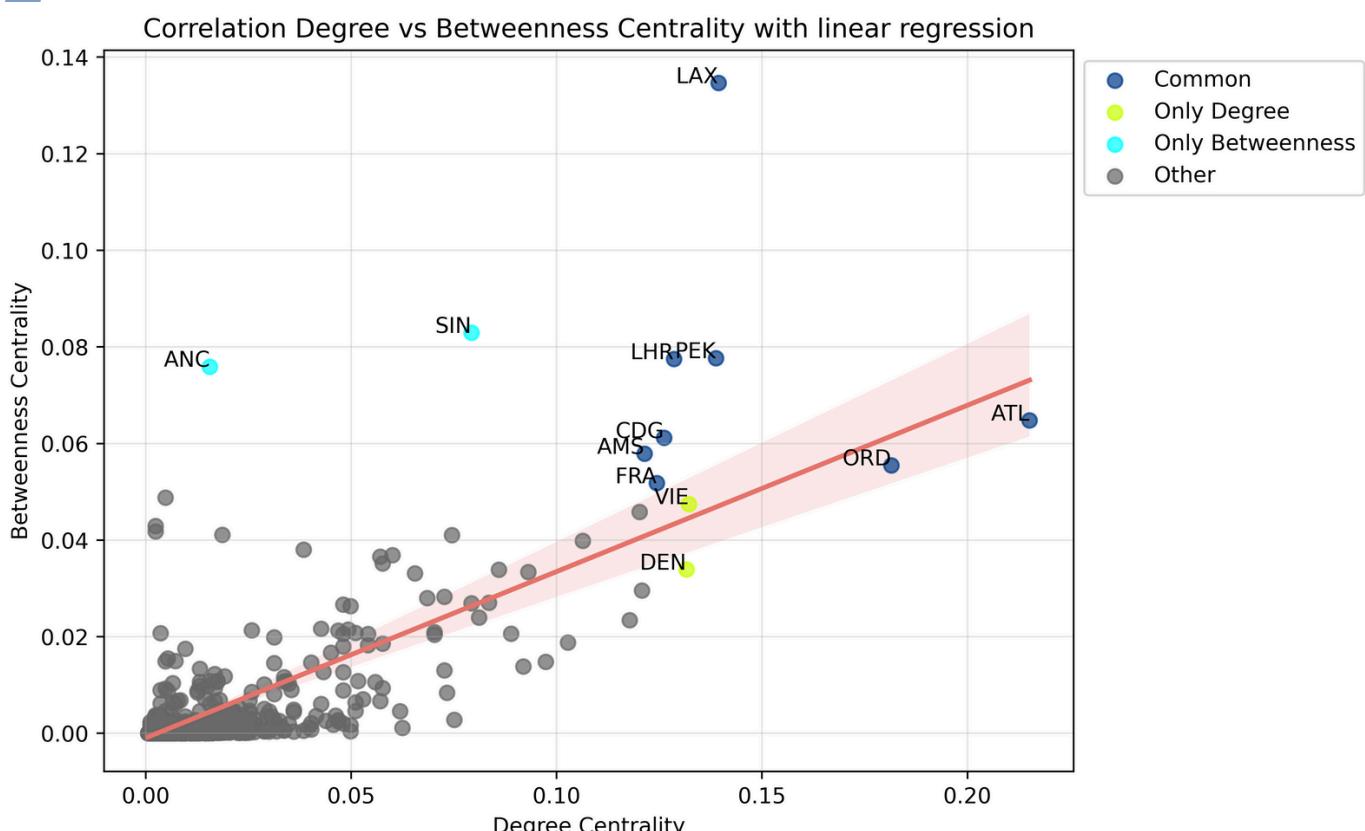
Per rispondere alla mia domanda di ricerca, ovvero **identificare gli aeroporti chiave nella rete globale di trasporto aereo**, ho deciso di combinare i risultati ottenuti dalle due misure di centralità.

La **Degree Centrality** evidenzia gli aeroporti che gestiscono il maggior volume di traffico aereo, considerando il numero totale di rotte in entrata e in uscita. Questi aeroporti rappresentano i nodi più connessi della rete e rivestono un ruolo fondamentale nella gestione della mobilità a livello regionale e globale.

D'altra parte, la **Betweenness Centrality** individua quegli aeroporti che fungono da ponti strategici, consentendo il collegamento tra diverse regioni geografiche e svolgendo un ruolo cruciale nel transito aereo globale. Questi aeroporti garantiscono la continuità e la robustezza della rete, agendo come snodi intermedi essenziali per il traffico internazionale.

Vediamo inoltre che, calcolando la correlazione tra le due misure, si ottiene un valore pari a **0.7646**, indicando una relazione positiva moderatamente forte tra le due misure. Questo risultato suggerisce che molti aeroporti altamente connessi, secondo la Degree Centrality, tendono anche a svolgere un ruolo strategico come snodi intermedi, evidenziato dalla Betweenness Centrality.

Tuttavia, è importante notare che la correlazione non è perfetta. Ciò implica che ci sono aeroporti con molte connessioni dirette (alto grado di Degree Centrality) che non necessariamente fungono da punti di transito critici nella rete (bassa Betweenness Centrality), e viceversa. Questa discrepanza sottolinea il valore complementare delle due misure per identificare gli aeroporti chiave.



Per comprendere meglio la relazione tra le due misure, è stato realizzato uno scatter plot (**Figura 2.2**) che rappresenta la Degree Centrality sull'asse x e la Betweenness Centrality sull'asse y. La linea rossa mostra la regressione lineare, evidenziando la relazione positiva tra le due misure, mentre l'area ombreggiata rappresenta l'intervallo di confidenza.

Questo grafico mette in evidenza come, all'aumentare della Degree Centrality, la Betweenness Centrality tenda generalmente a crescere. Tuttavia, i punti distribuiti al di sopra e al di sotto della linea di regressione mostrano che non tutti gli aeroporti con un alto numero di connessioni dirette svolgono un ruolo altrettanto rilevante come snodi intermedi, e viceversa.

Combinando i risultati di entrambe le centralità, possiamo costruire un quadro più completo degli aeroporti chiave nella rete globale di trasporto aereo. Da un lato, individuiamo gli aeroporti che gestiscono il maggior volume di traffico e, dall'altro, quelli che svolgono un ruolo cruciale nel garantire la connettività globale.

- **Aeroporti con elevato volume di traffico**

Questi aeroporti, caratterizzati da un'alta Degree Centrality, svolgono un ruolo fondamentale nella gestione del traffico aereo, garantendo un numero significativo di connessioni dirette con altre destinazioni. Essi rappresentano i principali punti di accesso alla rete aerea globale e sono cruciali per la mobilità regionale e internazionale.

Aeroporti chiave:

- **ATL (Hartsfield-Jackson Atlanta International Airport)**
- **ORD (O'Hare International Airport)**
- **LAX (Los Angeles International Airport)**

Questi aeroporti assicurano l'efficienza operativa della rete, offrendo un'ampia gamma di rotte dirette e riducendo la necessità di trasferimenti intermedi.

- **Aeroporti "gatekeeper" strategici**

Questi aeroporti, evidenziati da un'alta Betweenness Centrality, fungono da nodi intermedi essenziali, consentendo la connessione tra diverse regioni geografiche. Sono fondamentali per la robustezza della rete globale e per mantenere il flusso del traffico aereo internazionale.

Aeroporti chiave:

- **LHR (London Heathrow Airport)**
- **PEK (Beijing Capital International Airport)**
- **SIN (Singapore Changi Airport)**
- **LAX (Los Angeles International Airport)**

Questi aeroporti agiscono come ponti strategici, riducendo il rischio di frammentazione della rete e migliorandone la resilienza complessiva.

In conclusione, possiamo affermare che i nodi chiave identificati dalla combinazione delle due misure di centralità rappresentano gli aeroporti chiave nella rete globale di trasporto aereo. Aeroporti come **ATL**, **ORD** si distinguono per il loro elevato volume di traffico, mentre aeroporti come **LHR**, **PEK**, **LAX** e **SIN** emergono come snodi strategici cruciali per la connettività globale. La complementarità delle due misure offre una visione completa del ruolo di ciascun aeroporto nella rete, permettendoci di identificare i principali hub a livello mondiale.

Limiti e Bias

Limiti dei Dati Utilizzati

Il dataset utilizzato per l'analisi proviene da OpenFlights e fornisce informazioni sulle rotte aeree e sugli aeroporti. Tuttavia, vi sono delle limitazioni e bias su questi dati:

- **Incompletezza:** Il dataset potrebbe non includere tutte le rotte aeree globali o potrebbe presentare delle lacune.
- **Aggiornamento dei Dati:** I dati purtroppo risalgono al 2011 e coprono solo il mese di Gennaio, limitando la rappresentatività delle informazioni. Ad esempio, le nuove rotte o i cambiamenti nelle frequenze di volo o le fluttuazioni stagionali del traffico.
- **Nuove tecnologie:** Non è presente l'impatto di innovazioni tecnologiche o di nuove strategie aziendali che hanno trasformato il settore dell'aviazione, come l'introduzione di aeromobili a basso consumo o l'aumento delle rotte low-cost.

In futuro, sarebbe interessante approfondire l'analisi considerando cluster continentali per identificare gli aeroporti chiave a livello di continente. Questo permetterebbe di verificare se i risultati ottenuti a livello globale rispecchiano le dinamiche osservate su scala regionale, fornendo una visione più dettagliata delle reti di trasporto aereo e delle loro caratteristiche.

Bibliografia

Dataset e Fonti Dati

OpenFlights (2011). OpenFlights Airports Database [Dataset]. Disponibile su Kaggle.

Software e Librerie

- Python Software Foundation. (2024). Python Programming Language. <https://python.org>
- NumPy Community. (2023). NumPy: The fundamental package for scientific computing with Python. <https://numpy.org>
- The Pandas Development Team. (2024). pandas: Powerful data analysis tools for Python. <https://pandas.pydata.org>
- NetworkX Developers. (2023). NetworkX: Network Analysis in Python. <https://networkx.org>
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90-95.
- Met Office. (2023). Cartopy: A cartographic python library with matplotlib support. <https://scitools.org.uk/cartopy>
- Waskom, M. (2023). Seaborn: Statistical Data Visualization. <https://seaborn.pydata.org>

Metodologia e Concetti Teorici

- Zollo, F. (2024). Social Network Analysis [Lezioni universitarie]. Università Ca' Foscari Venezia, Italia.
- Zollo, F. (2024). Dispense del corso di Social Network Analysis [Materiale didattico]. Moodle Ca' Foscari, <https://moodle.unive.it>

Conclusioni

Dopo aver seguito ogni fase del processo, siamo giunti a una conclusione ***identificando una lista dei principali aeroporti chiave nella rete aerea globale***. Questi hub strategici ***si distribuiscono tra il centro dell'Europa, l'America centrale e l'Asia, coprendo la maggior parte del traffico aereo mondiale***, come da nostra prima ipotesi.

Sono loro che, come nodi di congiunzione, garantiscono solidità e struttura a questa rete intricata, rendendo possibile il flusso continuo di persone, culture ed esperienze.

Questa configurazione riflette non solo l'importanza geografica di tali hub, ma evidenzia anche la situazione economica e sociale dei Paesi che li ospitano. Gli aeroporti situati in nazioni economicamente forti sono spesso dotati di infrastrutture moderne e capacità operative avanzate, consolidando il loro ruolo chiave nella rete di trasporto aereo globale.

