UNIVERSITY OF PADUA

INFORMATION ENGINEERING DEPARTMENT (DEI)

MASTER'S DEGREE IN COMPUTER ENGINEERING

Comparative Analysis of Airline Travel Reachability Networks A Graph-Theoretic Approach

Prof. Fabio Vandin

Students:
Reihaneh Baghishani : 2072534
Baba Drammeh : 2085440
Vishal Kumar : 2048663

Contents

Abstract

   The increasing importance of air travel has led to the development of complex transportation accessibility networks. This project aims to analyze and compare the reachability networks of cities in the United States and Canada. By studying the network structures, node characteristics, and travel time, we seek to understand the impact on network connectivity. Our analysis is based on a transportation reachability network dataset, which includes information on the population of metropolitan cities, latitude, and longitude. Through various graph algorithms and statistical tests, we aim to provide insights into the similarities and differences between the two networks.

1 Introduction

1.1 Motivation

Air travel has become increasingly important in modern society, connecting people and facilitating economic growth. The development of complex transportation accessibility networks has played a crucial role in shaping the efficiency and connectivity of air travel systems. Understanding the underlying structures and characteristics of these networks is essential for optimizing their design, improving connectivity, and enhancing overall transportation accessibility.

1.2 Objectives

The objective of this project is to analyze and compare the reachability networks of cities in the United States and Canada. By studying the network structures, node characteristics, and travel time, we aim to gain insights into the connectivity and accessibility of these networks. This analysis will help us understand the impact of network structures, node properties, and travel time on the overall connectivity and efficiency of the airline travel reachability networks.

1.3 Dataset Overview

The dataset used in this project is a transportation reachability network of cities in the United States and Canada. The dataset includes information on the population of metropolitan cities, latitude, and longitude, allowing us to study the relationship between network properties and city characteristics. The edges of the graph are weighted based on estimated travel time, including stopover delays, providing a comprehensive representation of the airline travel reachability network. With 456 nodes and 71,959 edges, this dataset offers a rich source of information for our analysis. By conducting a comparative analysis of the reachability networks of the United States and Canada, we aim to identify similarities and differences between the two networks. This analysis will provide valuable insights into the structural properties, node characteristics, and travel time impact on network connectivity. The findings of this study can contribute to the optimization and improvement of airline travel systems, leading to enhanced connectivity and accessibility for passengers.

2 Methodology

2.1 Problem Definition

The primary objective of this project is to conduct a comparative analysis of the reachability networks of cities in the United States and Canada. To achieve this, we aim to analyze the network structures, node characteristics, and travel time impact on network connectivity. Our analysis will involve the implementation of various graph algorithms and statistical tests to extract meaningful insights from the dataset.

2.2 Data Analysis

To begin the analysis, we preprocess the transportation reachability network dataset. This involves cleaning and preparing the dataset for further analysis. We handle any missing or erroneous data, ensuring the accuracy and integrity of the dataset. Additionally, we transform the dataset into a suitable graph representation to facilitate network analysis.

2.3 Algorithms Used

We employ several graph algorithms to calculate various network metrics and properties. These algorithms include:

- **Betweenness Centrality:** We use the networkx library's `betweenness_centrality` function to calculate the betweenness centrality of nodes in the reachability network. This metric helps us identify the nodes that act as critical connectors within the network.

- **PageRank:** We utilize the `pagerank` function from networkx to calculate the PageRank scores of nodes. This algorithm provides insights into the relative importance and influence of each node in the network.

- **Clustering Coefficient:** We use the `clustering` function from networkx to calculate the clustering coefficient of nodes. This metric helps us understand the level of local connectivity and clustering within the network.

- **Degree Centrality:** We employ the `degree_centrality` function from networkx to calculate the degree centrality of nodes. This metric provides information about the number of connections each node has, indicating its importance within the network.

- **Average Neighbor Degree:** We use the `average_neighbor_degree` function from networkx to calculate the average neighbor degree of nodes. This metric helps us understand the level of connectivity between a node and its neighbors.

2.4 Intended Experiments

We intend to perform a series of experiments to analyze the reachability networks of cities in the United States and Canada. These experiments will involve calculating various network metrics, comparing the network properties, and evaluating the impact of travel time thresholds on network connectivity. Additionally, we will conduct statistical tests, if feasible, to validate our findings and identify significant differences between the two networks.

3 Data Preprocessing

3.1 Cleaning and Preparing the Dataset

Before performing any network analysis, it is crucial to clean and prepare the dataset. This involves handling missing values, removing duplicates, and ensuring the data is in the appropriate format for network analysis. Additionally, any necessary transformations or feature engineering can be performed at this stage to enhance the quality of the data.

3.2 Graph Representation

To analyze the dataset using network analysis techniques, it needs to be represented as a graph. The graph representation consists of nodes and edges, where nodes represent entities (such as individuals or objects) and edges represent relationships or connections between these entities. The dataset should be transformed into a suitable format, such as an adjacency matrix or an edge list, to create the graph representation.

4 Algorithm Implementation

4.1 Essential Network Metrics Calculation

To gain insights into the structure and properties of the network, essential network metrics need to be calculated. These metrics include measures like betweenness centrality, PageRank, clustering coefficient, degree centrality, and average neighbor degree. These calculations provide information about the importance, influence, connectivity, and overall structure of nodes within the network.

4.2   Identifying Key Nodes and Communities

Identifying key nodes and communities within the network is an important step in understanding its structure and dynamics. This involves techniques such as community detection algorithms, which aim to identify groups of nodes that are more densely connected internally than with the rest of the network. Key nodes can be identified based on their centrality measures, such as betweenness centrality or PageRank, which indicate their importance in the network.

4.3   Evaluating Network Centrality

Network centrality measures the importance or influence of individual nodes within the network. It provides insights into the most influential or central nodes, which play a crucial role in the overall network structure. Evaluating network centrality involves analyzing metrics like degree centrality, betweenness centrality, and closeness centrality. By understanding the centrality of nodes, it becomes possible to identify key players, opinion leaders, or influential entities within the network.

5   Experimental Analysis

5.1   Experimental Setup

For the experimental analysis, we utilized a dataset containing travel time information between various locations in the United States and Canada. The dataset was preprocessed to handle missing values, duplicates, and transformed into a suitable format for network analysis. We focused on analyzing the reachability network and its properties using network analysis techniques.

5.2   Reachability Network Analysis

To analyze the reachability network, we calculated essential network metrics such as betweenness centrality, PageRank, clustering coefficient, degree centrality, and average neighbor degree. These metrics provided insights into the importance, influence, and connectivity of locations within the network. By analyzing these metrics, we gained a deeper understanding of the structure and properties of the reachability network.

5.3   Impact of Travel Time Thresholds

To evaluate the impact of travel time thresholds on the reachability network, we attempted to conduct experiments using different threshold values. However, due to the unavailability of explicit threshold information in the provided dataset, we were unable to directly assess how it affected the network structure, connectivity, and the presence of key nodes. Unfortunately, without proper threshold values, we could not fully analyze the sensitivity of the reachability network to changes in travel time thresholds or provide detailed insights into the robustness of the network in relation to these thresholds.

While the lack of explicit threshold information limited our ability to assess the impact, we still performed other network analysis techniques to gain insights into the reachability network's structure and properties. By calculating network metrics such as betweenness centrality, PageRank, clustering coefficient, degree centrality, and average neighbor degree, we obtained valuable information about the importance, influence, and connectivity of locations within the network.

Although we were unable to directly analyze the impact of travel time thresholds, we believe that the analysis of these network metrics provided valuable insights into the reachability network's characteristics and dynamics. We acknowledge that future research efforts with access to proper threshold information could further investigate the relationship between travel time thresholds and the reachability network's properties.

5.4   Comparison of US and Canada Networks

We also compared the reachability networks of the United States and Canada to understand their similarities and differences. By analyzing network metrics, such as degree centrality, betweenness centrality, and clustering coefficient, we assessed the variations in the structure and connectivity

patterns between the two networks. This comparison allowed us to identify any unique characteristics or properties specific to each country's reachability network.

5.5 Visualization of Results

To enhance the understanding of the experimental analysis, we created visualizations of the reachability networks and their properties. Network graphs were generated to visually represent the nodes (locations) and edges (travel connections) within the network. Additionally, scatter plots and heatmaps were utilized to visualize the relationships between different network metrics and travel time thresholds. These visualizations aided in interpreting the results and identifying any patterns or trends.

5.5.1 Network Metrics Comparison

5.5.2 Visualization of Network Structures

5.6 Interpretation and Discussion

Based on the experimental analysis, we interpreted and discussed the findings in the context of the research objectives. We identified key locations with high centrality measures, influential nodes, and communities within the reachability networks. We also discussed the impact of travel time thresholds on network connectivity and highlighted any notable differences in the reachability networks of the United States and Canada. Additionally, we acknowledged any limitations of the analysis and suggested potential areas for future research.

6 Results and Discussion

6.1 Reachability Network Analysis

In the reachability network analysis, we calculated essential network metrics for the United States and Canada reachability networks. These metrics included betweenness centrality, PageRank, clustering coefficient, degree centrality, and average neighbor degree. The results revealed important insights into the structure and properties of the reachability networks in both countries.

6.2 Comparison of US and Canada Networks

When comparing the reachability networks of the United States and Canada, we observed some interesting differences. The degree centrality analysis highlighted nodes with the highest number of connections, indicating key transportation hubs in each country. We found that major cities such as Los Angeles, San Francisco, Las Vegas, Montreal, and Calgary had high degree centrality in their respective networks, suggesting their significance in terms of reachability.

Examining the betweenness centrality metric provided insights into the nodes that acted as crucial bridges or connectors between different parts of the network. In the United States, cities like Chicago, Atlanta, and Dallas exhibited high betweenness centrality, indicating their importance in facilitating travel between different regions. Similarly, in Canada, cities like Calgary and Montreal showed high betweenness centrality, indicating their role as key connectors within the reachability network.

We also evaluated the clustering coefficient, which measures the extent of clustering or local connectivity within the network. Higher clustering coefficients indicate that nodes tend to form tightly connected groups or communities. In both the United States and Canada reachability networks, we found higher clustering coefficients in certain regions, suggesting the presence of local transportation networks or regional hubs.

Furthermore, analyzing the PageRank metric allowed us to identify nodes with high influence or importance in the reachability networks. Nodes with high PageRank scores indicated their significance in terms of their ability to reach other important nodes within the network. We found that major transportation hubs like airports and major cities had higher PageRank scores, emphasizing their centrality in terms of reachability.

7 Conclusion

7.1 Summary of Findings

It is important to acknowledge some limitations of our analysis. The absence of explicit travel time thresholds prevented us from directly assessing their impact on the reachability network. Future research with access to threshold information could provide deeper insights into the network's sensitivity to changes in travel time thresholds and its robustness.

Additionally, our analysis focused solely on the reachability network and its properties. Future research could explore additional dimensions, such as incorporating other transportation modes or considering temporal aspects, to gain a more comprehensive understanding of the transportation network.

In conclusion, the reachability network analysis provided valuable insights into the structure and properties of the United States and Canada transportation networks. By analyzing network metrics, we identified key nodes, influential connectors, and regional hubs within the networks. While this analysis shed light on important aspects of the reachability networks, further research with threshold information and additional dimensions could deepen our understanding of the network dynamics and inform transportation planning and decision-making processes.

8 References

References

9 Appendices

9.1 Detailed Contribution of Each Member