Project Proposal - Airways Transport Network

Team Members

Likhil Kumar Rachuri (lkr46@drexel.edu)

Vuthej Krishna Reddy (vv334@drexel.edu)

Motivation

Huge data is being generated with the increasing demand for air-traffic, allowing the need to analyze the data more deeply to come up with interesting facts and trends. These trends will further help the competitors to better understand the needs of the users and to provide better facilities and connectivity.

Demand can be forecasted by observing trends and patterns from the network built using past data. This puts the airline companies in a better position for supply and demand.

Above points triggered the interest in us to select this topic for our term project.

Research Questions

From our observations, we would like to unearth interesting facts and trends which suggests a better understanding of the working of Airways Transportation Network. These include –

- Busiest Airport
- Least busy Airport
- Centrality - Influential Airports
- Eigenvector Centrality
- Group-based Community Detection
- Robust Communities

To also achieve a working model which will deliver the shortest path between the source and destination using necessary tools like NetworkX, UCINET, Gephi.

Introduction and Data Collection

U.S. Department of Transportation provides interesting data through Bureau of Transportation Statistics website www.transports.bts.gov. This website includes information of all airline carriers operating within United States of America.

Dataset we are interested in provides information about the source and destination of a flight within United States of America for the year 2019.

Airports are treated as Nodes and a flight traveling from one airport (source) to another (destination) establishes a connection/tie between the nodes. i.e. Source and Destination.

Current dataset consists of 1,786 nodes with 263,846 ties.

Source of our dataset : https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=292

UNIQUE_CARRIER_NAME	ORIGIN_AIRPORT_ID	ORIGIN	DEST_AIRPORT_ID	DEST	MONTH
Empire Airlines Inc.	10140	ABQ	11413	DRO	3
Empire Airlines Inc.	10140	ABQ	11711	FMN	3
Empire Airlines Inc.	10194	AFW	12278	ICT	3
Empire Airlines Inc.	10194	AFW	12896	LBB	3
Empire Airlines Inc.	10194	AFW	13158	MAF	3
Empire Airlines Inc.	10194	AFW	15370	TUL	3
Empire Airlines Inc.	10299	ANC	10299	ANC	3
Empire Airlines Inc.	10299	ANC	11555	ENA	3
Empire Airlines Inc.	10299	ANC	11630	FAI	3
Empire Airlines Inc.	10299	ANC	12184	HOM	3
Empire Airlines Inc.	10666	BLI	11762	FRD	3
Empire Airlines Inc.	10666	BLI	14747	SEA	3
Empire Airlines Inc.	11050	CLM	14747	SEA	3
Empire Airlines Inc.	11198	CVO	14057	PDX	3

Above table displays few instances/rows of data from our current dataset.

Unique_Carrier_Name : This is a unique Carrier Name allotted to every individual airline. When the same name has been used by multiple carriers, a numeric suffix is used for earlier users. For example, ‘Air Caribbean’, ‘Air Caribbean (1)’.

Origin_Airport_ID : Origin Airport, Airport ID. An identification number assigned by U.S. Department of Transportation to identify a unique airport.

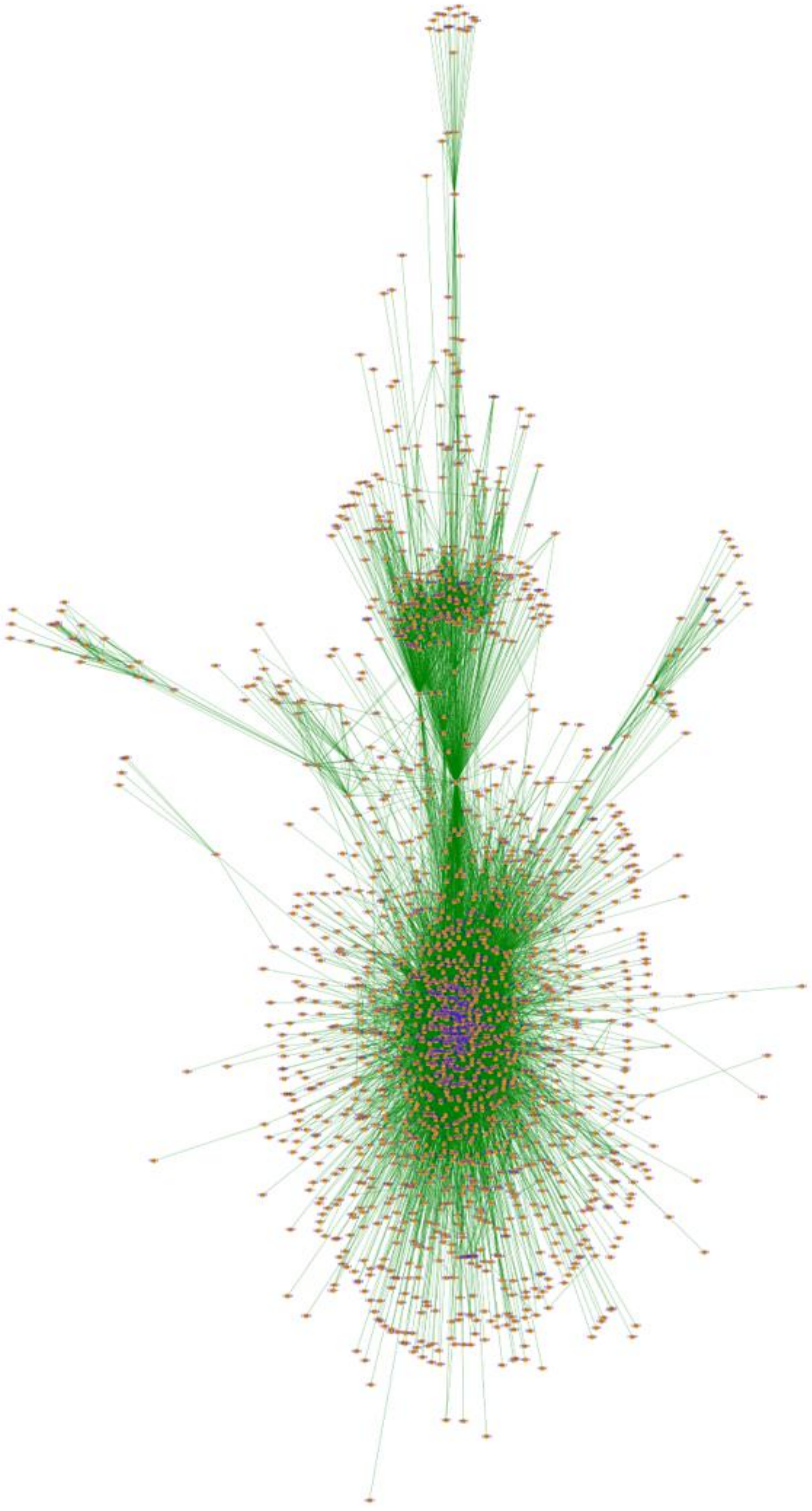
Origin : Origin Airport Code.

Dest_Airport_ID : Destination Airport, Airport ID. An identification number assigned by U.S Department of Transportation to identify a unique airport.

Dest : Destination Airport Code.

Month : This field provides the time period in months within a year.

Visualization of the network built from the current dataset.



Intended Methodology

To extract meaningful data from the interactions/ties displayed in the above network, certain techniques needs to be implemented on our network such as –

- Community Detection/Graph Clustering
- Clique Percolation Method (To identify overlapping communities)
- Focus-plus-Context based Visualization
- Fisheye View
- Fractal View
- Further methods with Descriptive and Statistical Models will be included in this project as we learn new methods in INFO-623 in the coming weeks.

It is important to apply these techniques since its hard to understand the characteristics and properties of underlying clusters of data when there are numerous connections making the network densely connected.

In order to implement these methodologies, we wish to utilize the tools NetworkX, UCINET and Gephi. For NetworkX, we will be using Python platform for programing.

References

Bureau of Transportation Statistics website - www.transports.bts.gov

Source of our dataset - https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=292

Knowledge about Focus-plus-Context based Visualization, Fisheye View, Fractal View methodologies - https://s3.us-east-1.amazonaws.com/blackboard.learn.xythos.prod/5a3199fc4282a/12756466?response-content-disposition=inline%3B%20filename%2A%3DUTF-8%27%27Analyzing%2520the%2520Terrorist%2520Social%2520Networks%2520with%2520Visualization%2520Tools%25281%2529.pdf&response-content-type=application%2Fpdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20200429T172852Z&X-Amz-SignedHeaders=host&X-Amz-Expires=21600&X-Amz-Credential=AKIAIL7WQYDOOHAZJGWQ%2F20200429%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=cded52f436063d2bd8a81285bd63e89a3c3935945a94c020b7b39f4a01ef700d