MH8351 Web Analytics Project

Air Routes Network Analysis

Chan Ka Wei Sherice (G1902324A)
Koh Chin Nam Andrew (G1901964B)
Lim Guowei (G1901873A)
Lee Sin Tat)G1800658J)

01

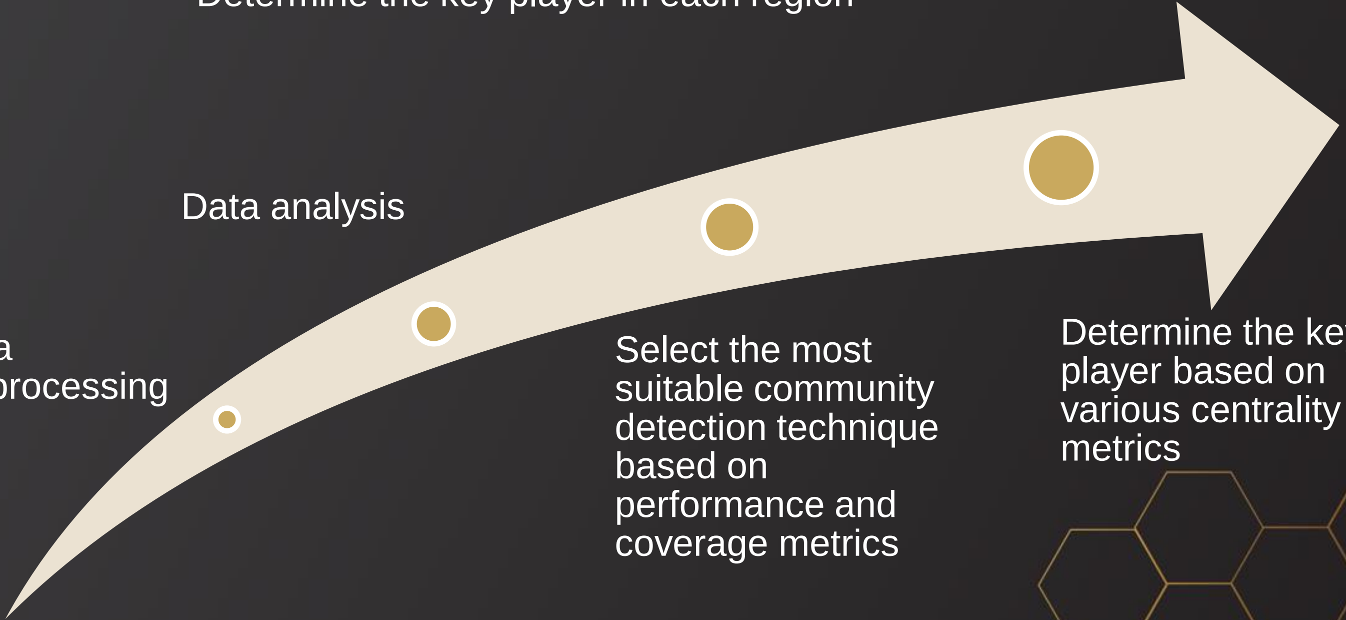
INTRODUCTION

Introduction

- According to International Air Transport Association (IATA), there were 3.8 billion air travelers in 2016. This number is expected to double to 7.2 billion passengers by 2035.
- Growth in air travel will generate new routes and require more infrastructure such as new airports and new planes.
- The air transportation system can be represented as a network.

Objectives and Approach

- Identify geographic regions based on flight routes graph data
- Determine the key player in each region



Data preprocessing

Data analysis

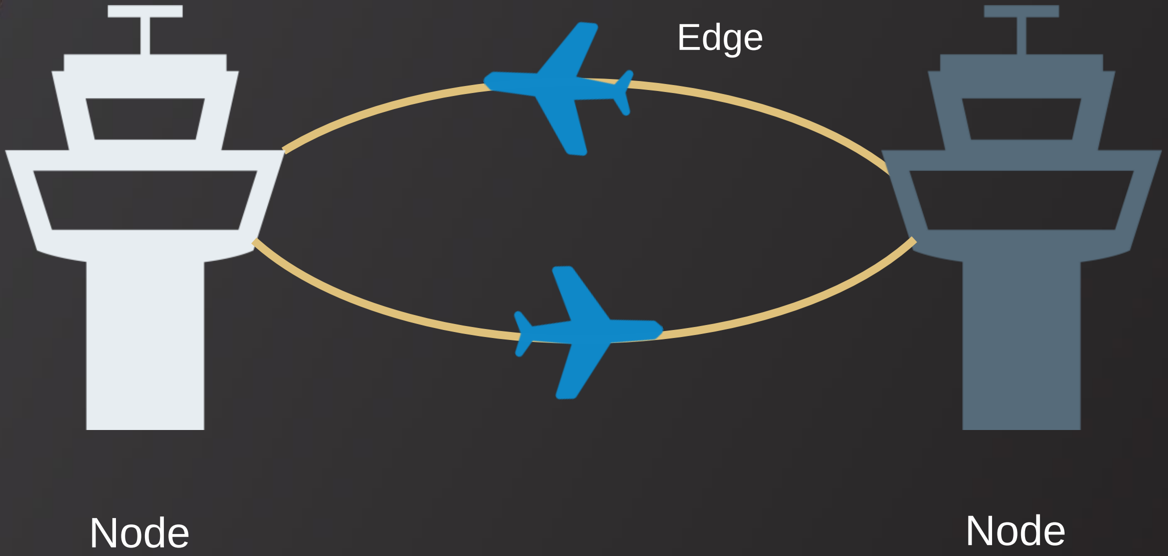
Select the most suitable community detection technique based on performance and coverage metrics

Determine the key player based on various centrality metrics

02

METHODS

Definition and Assumptions



- Undirected, non-weighted graph
- Network excludes airports that are not involved in routes data

Community Detection

Algorithm	Description
Clauset-Newman-Moore modularity maximization	The algorithm initially associates each node with a community. Then, it repeatedly combines the communities of which the union produces the highest increase to the modularity of the community structure.
Label propagation	Every node is initialised with a unique label. Iteratively, labels are reassigned such that each node takes the majority label of its neighbours.
Louvain best partition	The method first looks for "small" communities by optimising modularity locally, then aggregates nodes belonging to same community and build a new network. Repeat the process iteratively until maximum modularity is obtained.

Clustering Methods

Weakly Supervised Methods:

Non-Weighted vs. Weighted (Distance) Graphs
Spectral Clustering
Modification: Pre-define centres during K-Means

Key Player Detection

Degree centrality
Betweenness centrality
VoteRank centrality

Further Analysis: Airlines

Airlines Network Behaviour

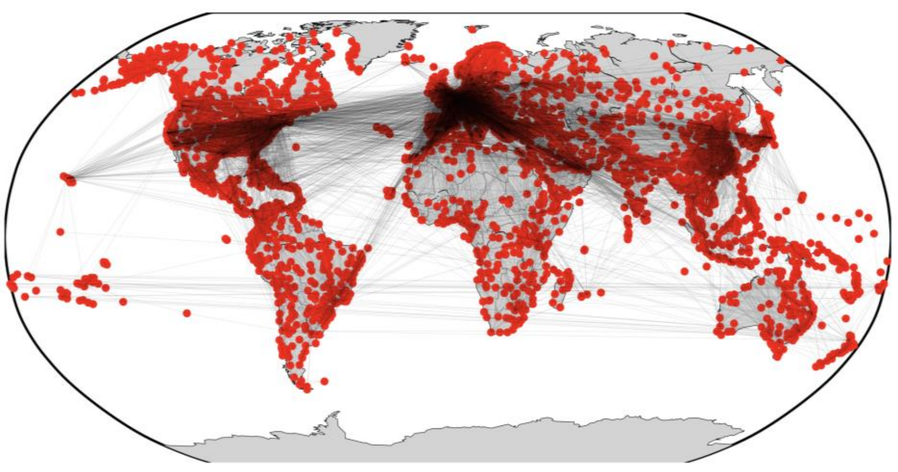
Competitor Analysis: Community Similarity

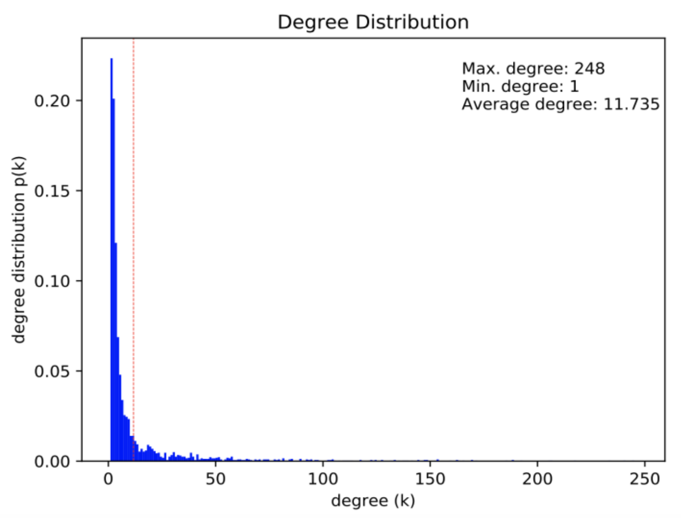
Strongly-Connected Components

03

RESULTS

Network Measures



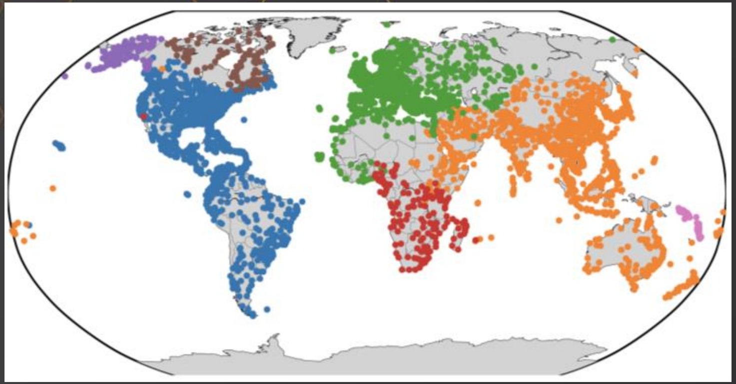


- Number of nodes: 3214
- Number of edges: 18858
- Density: 0.365%
- Average clustering coefficient: 0.492

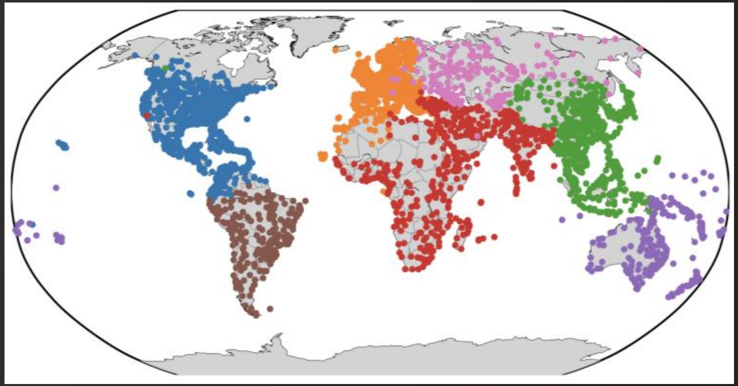
- Power law distribution

Community Detection

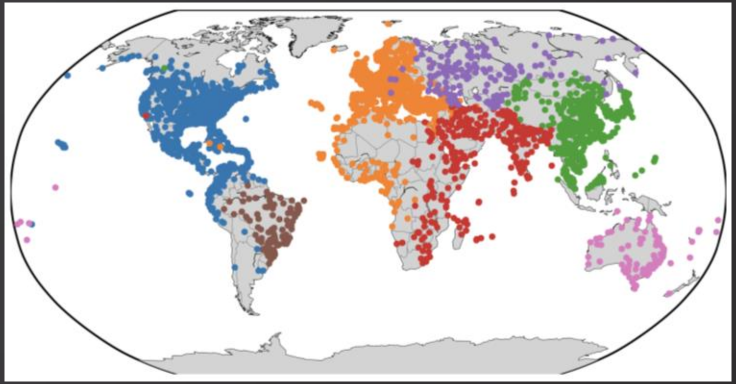
Clauset-Newman-Moore modularity maximization



Louvain best partition



Label propagation

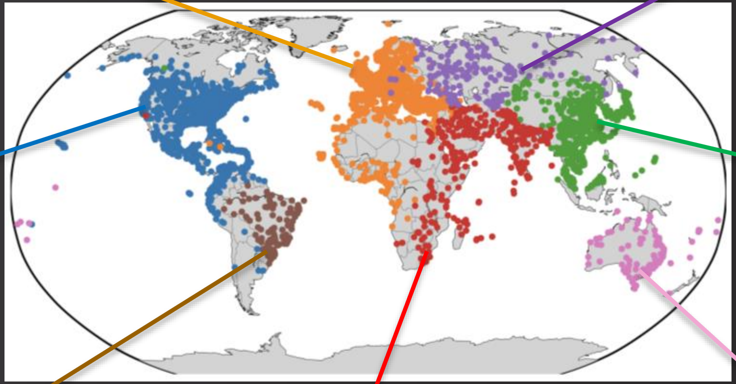


	Clauset-Newman-Moore modularity maximization	Label propogation	Louvain best partition
Performance	0.796	0.892	0.866
Coverage	0.904	0.852	0.867

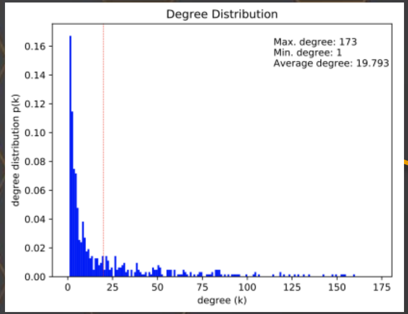
- Top 7 biggest clusters are plotted, else excluded
- Relatively similar performance and coverage
- Label propagation is down-selected for key player analysis based on visual closeness with known territories

Community Detection

Label propagation

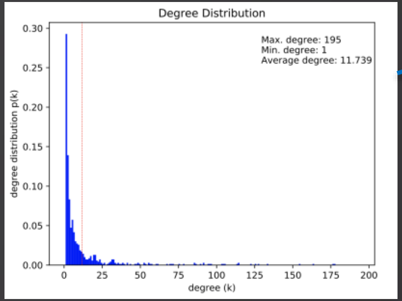


2

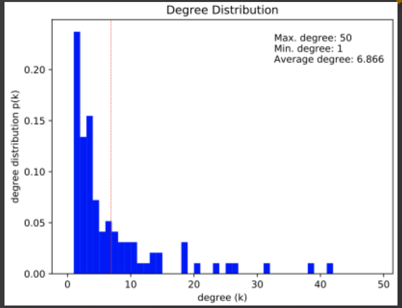


1

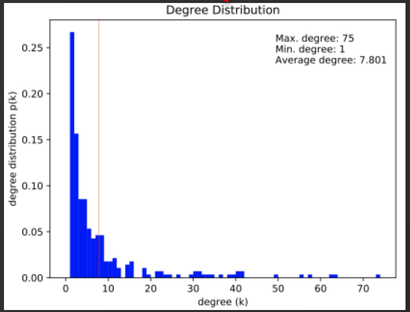
Largest



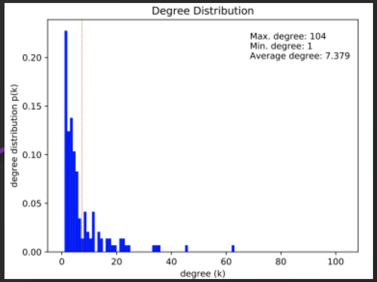
6



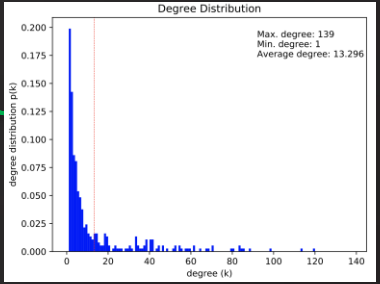
4



5

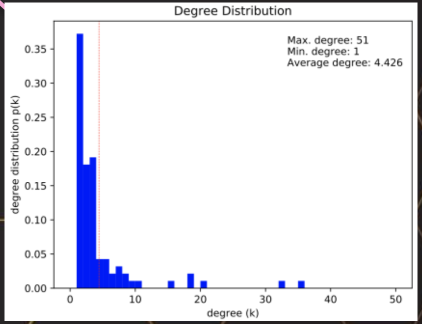


3



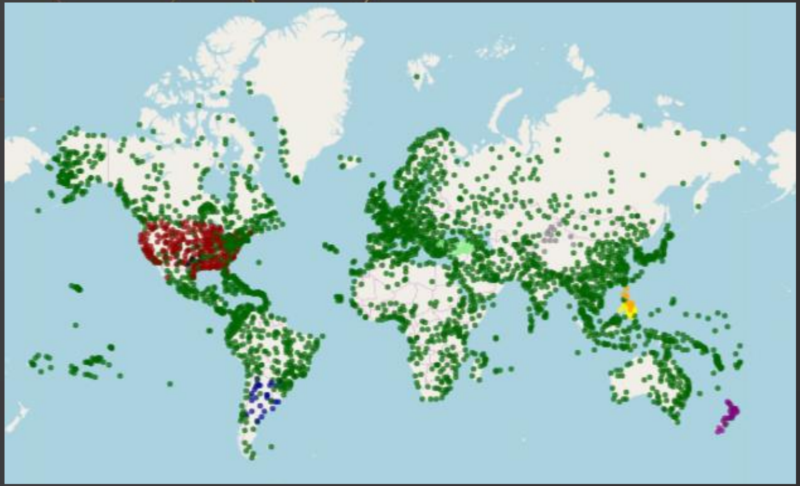
7

Smallest

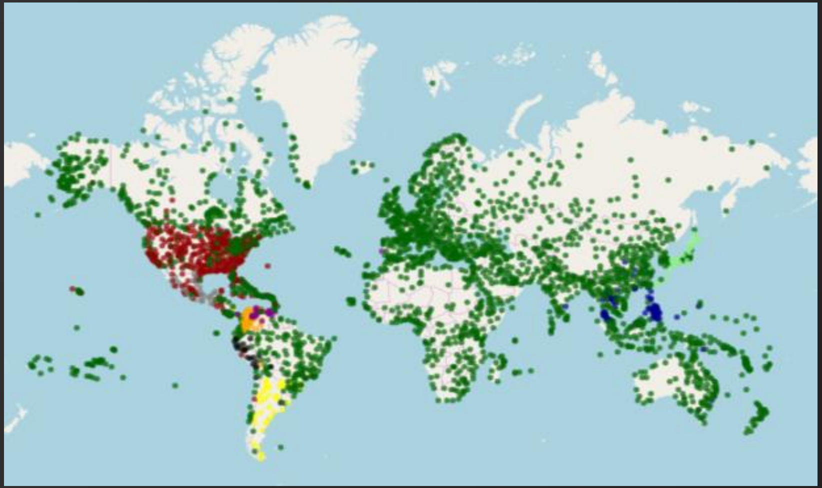


Clustering Methods: Spectral Clustering

Spectral Clustering (Non-weighted)



Spectral Clustering (Weighted - Distance)



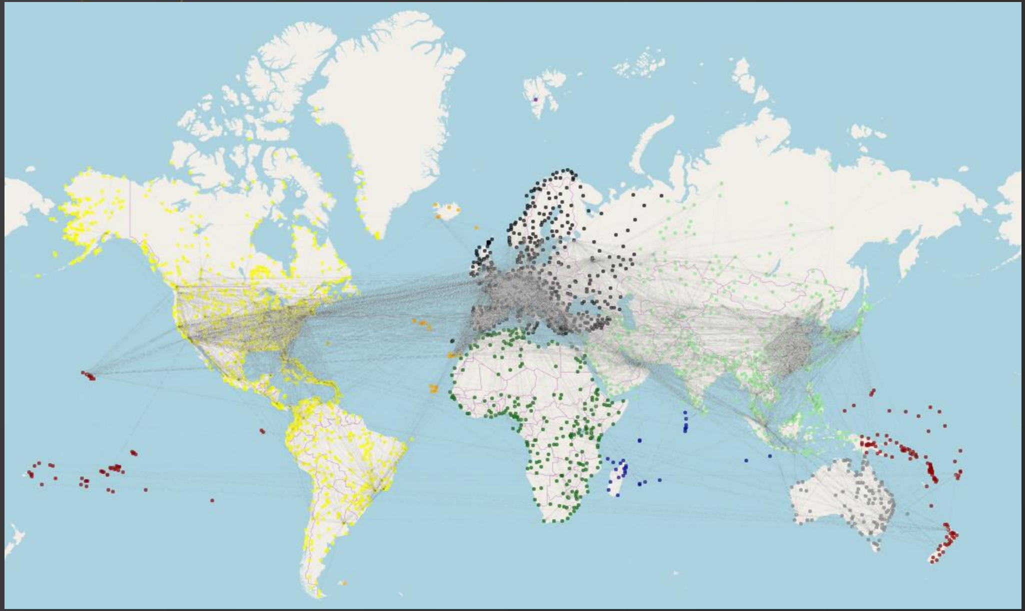
	Spectral Clustering (Non-weighted)	Spectral Clustering (Weighted)
Performance	0.363	0.225
Coverage	0.792	0.912

- Poor performance: Mainly one community detected
- Causes: Random initialization trap (Note: Spectral Clustering involve a final procedure: K-Means clustering on the Lapacian Eigenmap)
- Causes: Numerous inter-cluster (or regions) connections

Clustering Methods: Pre-define the Centres for Spectral Clustering

Original Dataset Mapping

Note: Above coloring is based assigned nine regions in the original dataset



Using ground-truth for better K-means clustering in the final step of Spectral Clustering*:

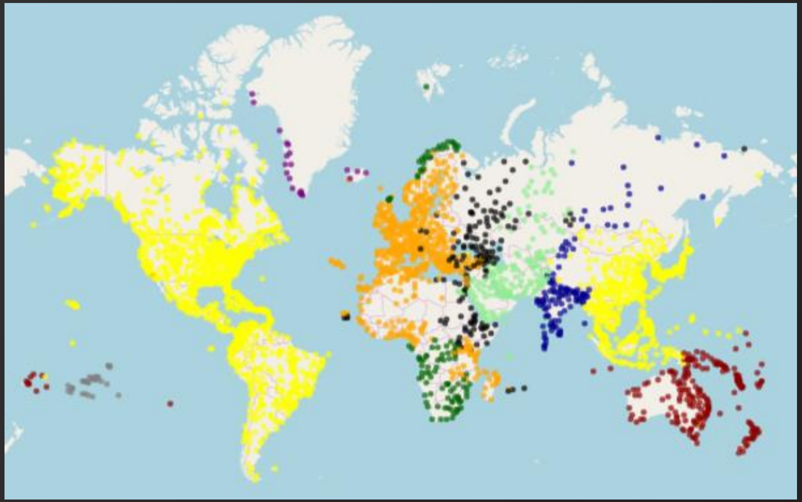
- Establish a hub/key airport for each region (stated in data set) using centrality measures
- Nine hubs for nine regions: Europe, Asia, America, Africa, Indian, Artic, Australia, Pacific, Atlantic.
- Modify the final step of Spectral Clustering: Perform K-means on the Lapacian Eigenmap using pre-defined hubs as cluster centers instead of letting it run randomly.
- Minimise erratic clustering

**Note: Method proposed was referenced from https://github.com/lchenbb/NTDS2019_Project*

Clustering Methods: Spectral Clustering (Modified the K-Mean Step)

Modified Final Step:
K-Means Clustering using Ground Truth (Non-weighted)

Modified Final Step:
K-Means Clustering using Ground Truth (Weighted - Distance)

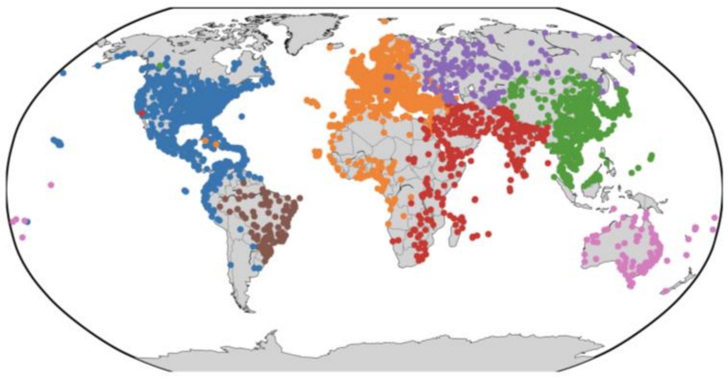


	Spectral Clustering (Non-weighted)	Spectral Clustering (Weighted)	Spectral Clustering (added ground truths for the K-mean step)	
			(Non-weighted)	(Weighted)
Performance	0.363	0.225	0.807	0.663
Coverage	0.792	0.912	0.767	0.847

- Improvement from random initialization.
- Poorer performance as compared to autonomous community detection methods (e.g. label propagation)

Key Players

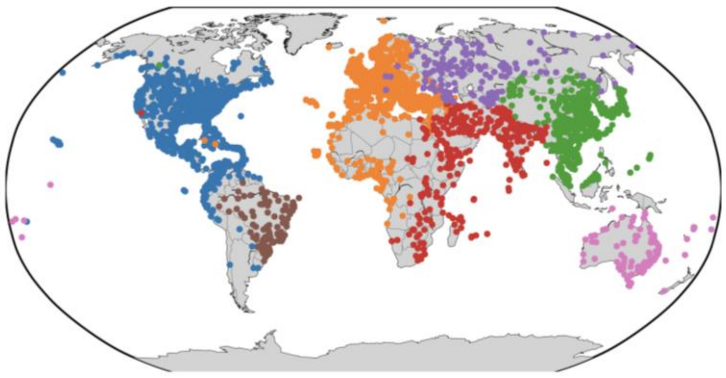
blue region					
	Airport ID	degree_centrality	Name	City	Country
0	3682	0.28	Hartsfield Jackson Atlanta International Airport	Atlanta	United States
1	3830	0.254	Chicago O'Hare International Airport	Chicago	United States
2	3670	0.253	Dallas Fort Worth International Airport	Dallas-Fort Worth	United States
3	3751	0.234	Denver International Airport	Denver	United States
4	3550	0.221	George Bush Intercontinental Houston Airport	Houston	United States
	Airport ID	betweenness_centrality	Name	City	Country
0	3682	0.104	Hartsfield Jackson Atlanta International Airport	Atlanta	United States
1	3751	0.102	Denver International Airport	Denver	United States
2	3670	0.098	Dallas Fort Worth International Airport	Dallas-Fort Worth	United States
3	3830	0.09	Chicago O'Hare International Airport	Chicago	United States
4	3550	0.074	George Bush Intercontinental Houston Airport	Houston	United States
	Airport ID	voterank_centrality	Name	City	Country
0	3682	1	Hartsfield Jackson Atlanta International Airport	Atlanta	United States
1	3830	2	Chicago O'Hare International Airport	Chicago	United States
2	3670	3	Dallas Fort Worth International Airport	Dallas-Fort Worth	United States
3	3751	4	Denver International Airport	Denver	United States
4	3550	5	George Bush Intercontinental Houston Airport	Houston	United States



- Hartsfield Jackson Atlanta International Airport has the highest centrality based on all 3 metrics

Key Players

	red region				
	Airport ID	degree_centrality	Name	City	Country
0	2188	0.268	Dubai International Airport	Dubai	United Arab Emirates
1	2072	0.261	King Abdulaziz International Airport	Jeddah	Saudi Arabia
2	3093	0.225	Indira Gandhi International Airport	Delhi	India
3	2997	0.221	Chhatrapati Shivaji International Airport	Mumbai	India
4	2191	0.204	Sharjah International Airport	Sharjah	United Arab Emirates
	Airport ID	betweenness_centrality	Name	City	Country
0	2072	0.19	King Abdulaziz International Airport	Jeddah	Saudi Arabia
1	2188	0.183	Dubai International Airport	Dubai	United Arab Emirates
2	2997	0.137	Chhatrapati Shivaji International Airport	Mumbai	India
3	3093	0.133	Indira Gandhi International Airport	Delhi	India
4	813	0.129	OR Tambo International Airport	Johannesburg	South Africa
	Airport ID	voterank_centrality	Name	City	Country
0	2188	1	Dubai International Airport	Dubai	United Arab Emirates
1	2072	2	King Abdulaziz International Airport	Jeddah	Saudi Arabia
2	3093	3	Indira Gandhi International Airport	Delhi	India
3	2997	4	Chhatrapati Shivaji International Airport	Mumbai	India
4	2191	5	Sharjah International Airport	Sharjah	United Arab Emirates



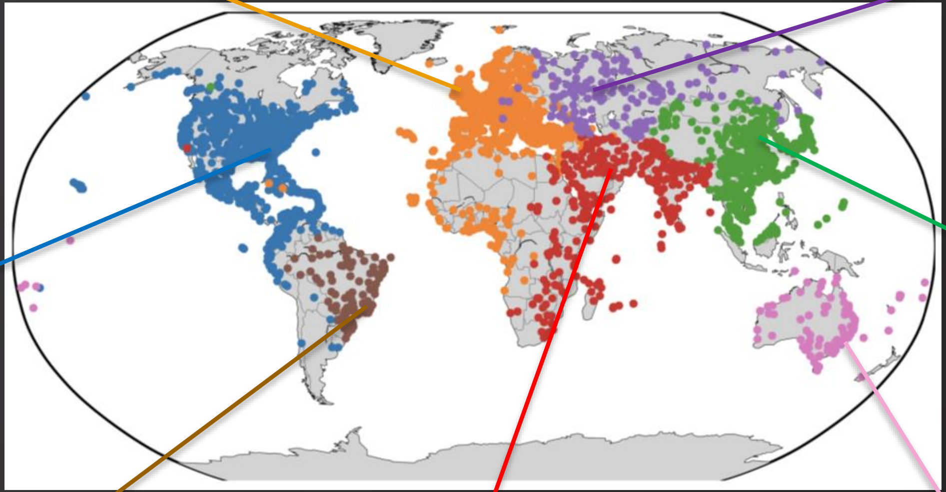
- Dubai International Airport exhibits 2nd highest centrality based on betweenness but 1st based on degree and voterank

- Hence, Dubai International Airport is selected as key player in red region based on voting scheme

Key Players

Amsterdam Airport Schiphol, Amsterdam, Netherlands

Domodedovo International Airport, Moscow, Russia



Hartsfield Jackson Atlanta International Airport, Atlanta, United States

Beijing Capital International Airport, Beijing, China

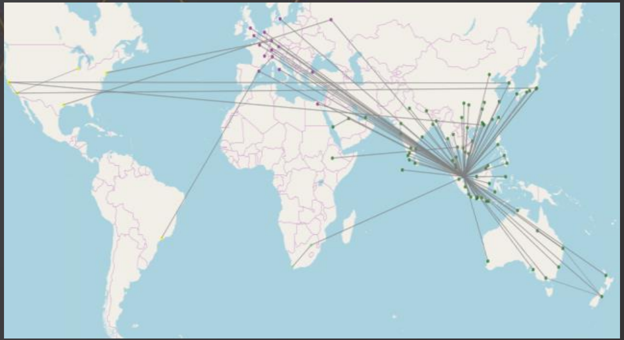
Viracopos International Airport, Campinas, Brazil

Dubai International Airport, Dubai, United Arab Emirates

Sydney Kingsford Smith International Airport, Sydney, Australia

Further Analysis: Airlines Network Behaviour

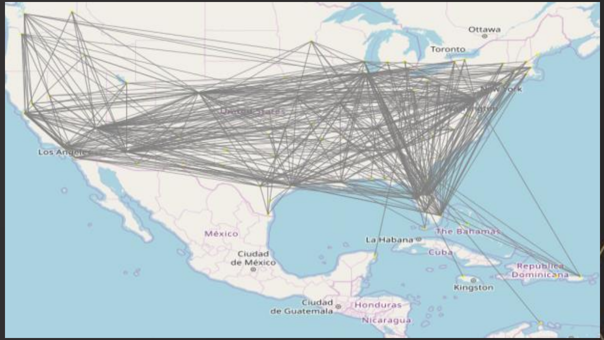
Singapore Airlines (International)



JetBlue Airways (US-Domestic)

United Emirates Airlines (International)

South West Airlines (US-Domestic)



- International airlines spread across various colored clusters as compared to single cluster for domestic.
- Most have single hub at home country and weak clustering (esp. international).
- Graph clustering tendency may differ: South West (US) exhibit stronger clustering/connectivity in West, Central US as compared to JetBlue Airways.
- Affect choice of airlines reward program by customers in different region.
- Evaluate competition and market structure

Further Analysis: Cluster Outreach of Competitor Airlines

Name	Unique_airports	No_of_airports	Communities	Communities_no	Competitor_Cluster_Advantage	Subject_Cluster_Advantage	Cluster_Jaccard
Singapore Airlines	[1107, 3885, 3341, 3316, 2006, 2994, 580, 1218...	102	[{4105, 16, 8209, 4130, 4160, 4161, 4162, 4166...	12	[]	[]	1.000000
Air China	[1107, 3370, 3364, 2006, 3406, 6404, 3395, 737...	186	[{4105, 16, 8209, 4130, 4160, 4161, 4162, 4166...	11	[{6372, 6373, 6374, 7558, 6795, 7470, 3380}]	[{9888, 9889, 3250, 3886}, {3240, 3241, 3243, ...	0.769231
Emirates	[253, 248, 2188, 1107, 3341, 2006, 3320, 3339,...	135	[{4105, 16, 8209, 4130, 4160, 4161, 4162, 4166...	11	[{6289, 4217, 6337, 6238}]	[{9888, 9889, 3250, 3886}, {3240, 3241, 3243, ...	0.769231
Korean Air	[2006, 3930, 580, 2340, 3682, 2560, 3797, 2650...	118	[{2006, 2007, 2009, 2011, 2012, 2014, 2015, 20...	11	[{6372, 6373, 6374, 7558, 6795, 7470, 3380}]	[{9888, 9889, 3250, 3886}, {3240, 3241, 3243, ...	0.769231
Asiana Airlines	[1107, 4059, 2006, 2279, 2908, 3930, 3682, 379...	105	[{4105, 16, 8209, 4130, 4160, 4161, 4162, 4166...	9	[]	[{2560, 2562, 2564, 2569, 2570, 2572, 2575, 25...	0.750000

Top 4 (Jaccard Coefficient) Airlines

Differences in Cluster Sets (Subject vs Competitor, vice versa) are identified

(Jaccard Coefficient) Descending Order

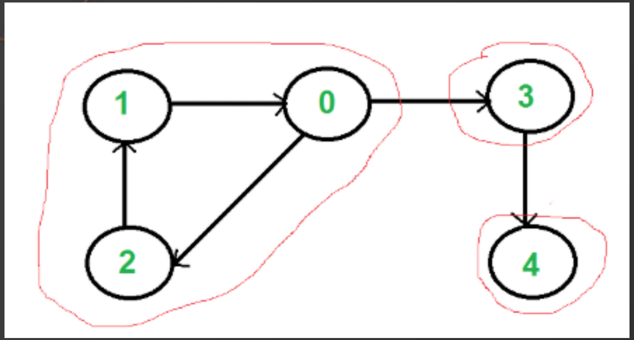
Note: Jaccard Coefficient = (Similar Community Set) / (Union of Both Subject and Competitor Community Set)

Search for closest competitors:

- Subject: Singapore Airlines (SQ) vs Other Airlines
- Use Jaccard Coefficient to detect cluster type similarity (i.e. if share same cluster grouping) . Higher coefficient could represent more shared competition.
- Air China, Emirates Airlines (see left - Top 4 coefficient listed) are well-known top competitors of SQ. Korean Air and Asiana could inconspicuous and worth monitoring.
- Identified clusters present in competitor but absent from Singapore Airlines, vice versa. May be opportunities to expand.

Further Analysis: Strongly Connected Components of Airlines Network

What is Strongly Connected Component (SCC)?



A directed graph is strongly connected if there is a <u>path between all pairs of vertices</u>. There are 3 strongly connected components above (Example: Node 3 can't travel to 0,1,2).

Configuration (Example): {3, 1, 1}

Possible insights:

- Search for possible disconnected communities within airlines network
- Implications: Loss of synergy / route opportunities / customers
- The strongly connected components of the Airline Network matter: {99 airports, 1 airport, 1 airport, 1 airport} may indicate a reasonable mix . Many large clusters (such as 30:40:32} may not be desirable as means any disconnection affect a large market (many airports or community).

Consider this:

Airline A's (SCC):

99 1 1 1

Airline B's (SCC):

30 40 32

Further Analysis: Strongly Connected Components of Airlines

Possible disconnection between all pairs of vertices / airports visited by Singapore Airlines

Concern: Route opportunities / Loss of customers

Search for Disconnected Airports in Singapore Airlines Network:

- 102 airports visited by SQ
- Performed "Strongly_Connected_Components" algorithm on "Directed" Graph of Singapore Airlines routes.
- Four strongly connected components were detected:
- 1 large component of 99 airports and 3 other small independent components {i.e. 99 airports, 1 airport, 1 airport, 1 airport}.
- One example: SQ flies to Lyon, France from Singapore but no returning route. Possible reason: Partnership with other airlines for return.

Singapore Airlines' Strongly Connected Component Network:

04

CONCLUSION AND RECOMMENDATION

Conclusion and recommendation

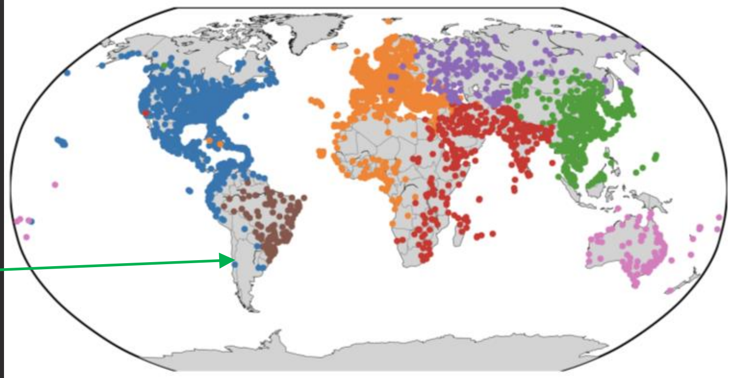
More data for graphing:

(1) Partnership between Airlines for flight routes (more accurate analysis of Airlines network).
(2) Temporal data for routes could enable graph filtering for more insight. Graph network might be dynamic over timeframe.
(3) Route profit/cost/time could be considered as weight

STAR ALLIANCE

Methods to explore:

Consider clique merging algorithms such as LCMA to include data points before centrality calculation



Appendix

Community detection output

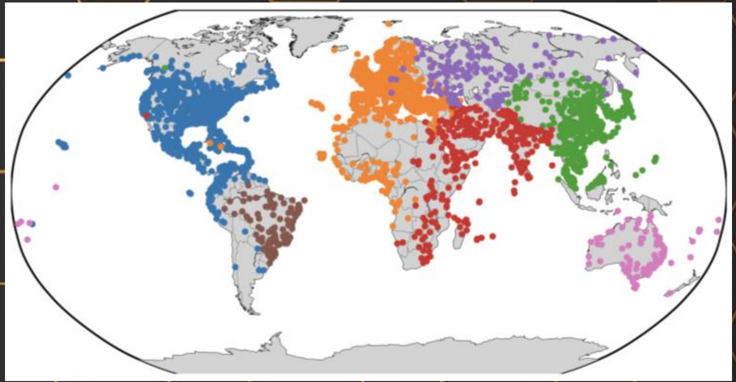
Clauset-Newman-Moore Modularity maximization							
sub_graph	**0**	**1**	**2**	**3**	**4**	**5**	**6**
Number of nodes	900	874	719	145	131	92	43
Number of edges	4686	4367	6876	292	192	133	69
Density (%)	1.158	1.145	2.664	2.797	2.255	3.177	7.641
Average clustering coefficient	0.5	0.539	0.465	0.487	0.44	0.297	0.526
Performance	0.796						
Coverage	0.904						

Label propogation							
sub_graph	0	1	2	3	4	5	6
Number of nodes	697	628	372	281	145	97	94
Number of edges	4091	6215	2473	1096	535	333	208
Density (%)	1.687	3.157	3.584	2.786	5.125	7.152	4.759
Average clustering coefficient	0.488	0.455	0.551	0.527	0.515	0.432	0.483
Performance	0.892						
Coverage	0.852						

Louvain best partition							
sub_graph	**0**	**1**	**2**	**3**	**4**	**5**	**6**
Number of nodes	733	511	490	463	238	211	159
Number of edges	4086	5323	2830	1745	494	575	676
Density (%)	1.523	4.085	2.362	1.632	1.752	2.595	5.382
Average clustering coefficient	0.509	0.44	0.531	0.549	0.495	0.432	0.505
Performance	0.866						
Coverage	0.867						

Appendix

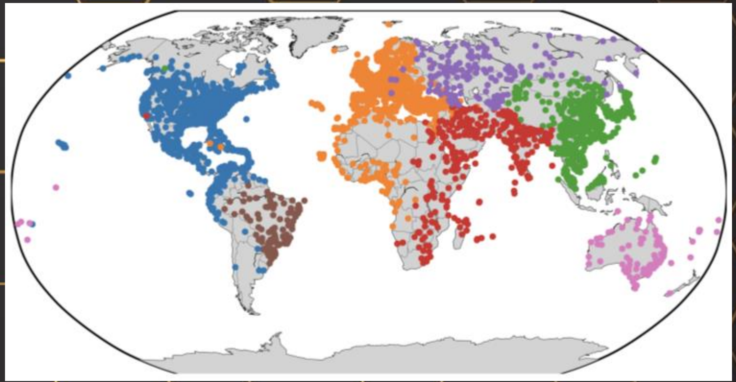
Key players



	blue region				
	Airport ID	degree_centrality	Name	City	Country
0	3682	0.28	Hartsfield Jackson Atlanta International Airport	Atlanta	United States
1	3830	0.254	Chicago O'Hare International Airport	Chicago	United States
2	3670	0.253	Dallas Fort Worth International Airport	Dallas-Fort Worth	United States
3	3751	0.234	Denver International Airport	Denver	United States
4	3550	0.221	George Bush Intercontinental Houston Airport	Houston	United States
	Airport ID	betweenness_centrality	Name	City	Country
0	3682	0.104	Hartsfield Jackson Atlanta International Airport	Atlanta	United States
1	3751	0.102	Denver International Airport	Denver	United States
2	3670	0.098	Dallas Fort Worth International Airport	Dallas-Fort Worth	United States
3	3830	0.09	Chicago O'Hare International Airport	Chicago	United States
4	3550	0.074	George Bush Intercontinental Houston Airport	Houston	United States
	Airport ID	voterank_centrality	Name	City	Country
0	3682	1	Hartsfield Jackson Atlanta International Airport	Atlanta	United States
1	3830	2	Chicago O'Hare International Airport	Chicago	United States
2	3670	3	Dallas Fort Worth International Airport	Dallas-Fort Worth	United States
3	3751	4	Denver International Airport	Denver	United States
4	3550	5	George Bush Intercontinental Houston Airport	Houston	United States

Appendix

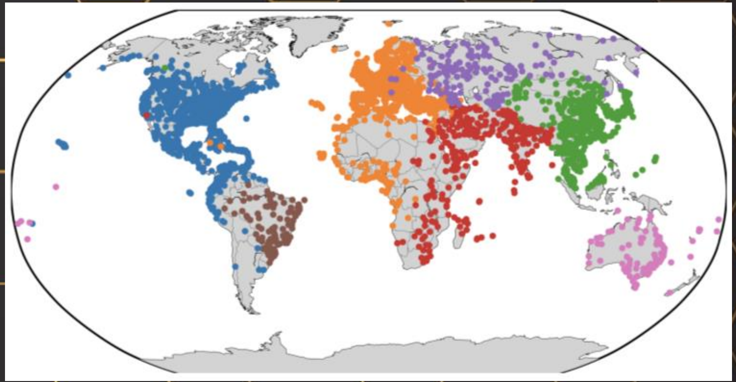
Key players



	Airport ID	degree_centrality	Name	City	Country
orange region					
	Airport ID	**degree_centrality**	**Name**	**City**	**Country**
0	580	0.276	Amsterdam Airport Schiphol	Amsterdam	Netherlands
1	1382	0.254	Charles de Gaulle International Airport	Paris	France
2	346	0.246	Munich Airport	Munich	Germany
3	548	0.244	London Stansted Airport	London	United Kingdom
4	1701	0.242	Atatürk International Airport	Istanbul	Turkey
	Airport ID	**betweenness_centrality**	**Name**	**City**	**Country**
0	1701	0.097	Atatürk International Airport	Istanbul	Turkey
1	1382	0.069	Charles de Gaulle International Airport	Paris	France
2	644	0.055	Oslo Lufthavn	Oslo	Norway
3	3941	0.055	Eleftherios Venizelos International Airport	Athens	Greece
4	580	0.054	Amsterdam Airport Schiphol	Amsterdam	Netherlands
	Airport ID	**voterank_centrality**	**Name**	**City**	**Country**
0	580	1	Amsterdam Airport Schiphol	Amsterdam	Netherlands
1	1382	2	Charles de Gaulle International Airport	Paris	France
2	548	3	London Stansted Airport	London	United Kingdom
3	1701	4	Atatürk International Airport	Istanbul	Turkey
4	346	5	Munich Airport	Munich	Germany

Appendix

Key players



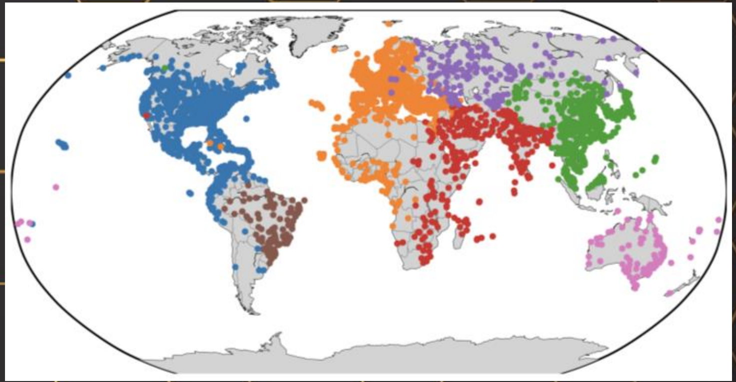
	Airport ID	degree_centrality	Name	City	Country
			green region		
0	3364	0.375	Beijing Capital International Airport	Beijing	China
1	3370	0.321	Guangzhou Baiyun International Airport	Guangzhou	China
2	3406	0.305	Shanghai Pudong International Airport	Shanghai	China
3	3395	0.264	Chengdu Shuangliu International Airport	Chengdu	China
4	3382	0.237	Kunming Changshui International Airport	Kunming	China

	Airport ID	betweenness_centrality	Name	City	Country
0	3364	0.153	Beijing Capital International Airport	Beijing	China
1	2359	0.12	Tokyo Haneda International Airport	Tokyo	Japan
2	3406	0.084	Shanghai Pudong International Airport	Shanghai	China
3	3157	0.076	Don Mueang International Airport	Bangkok	Thailand
4	3370	0.076	Guangzhou Baiyun International Airport	Guangzhou	China

	Airport ID	voterank_centrality	Name	City	Country
0	3364	1	Beijing Capital International Airport	Beijing	China
1	3370	2	Guangzhou Baiyun International Airport	Guangzhou	China
2	3406	3	Shanghai Pudong International Airport	Shanghai	China
3	3395	4	Chengdu Shuangliu International Airport	Chengdu	China
4	3382	5	Kunming Changshui International Airport	Kunming	China

Appendix

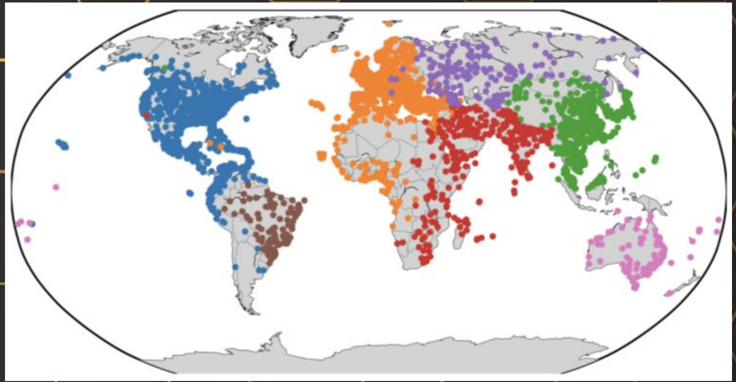
Key players



	red region				
	Airport ID	degree_centrality	Name	City	Country
0	2188	0.268	Dubai International Airport	Dubai	United Arab Emirates
1	2072	0.261	King Abdulaziz International Airport	Jeddah	Saudi Arabia
2	3093	0.225	Indira Gandhi International Airport	Delhi	India
3	2997	0.221	Chhatrapati Shivaji International Airport	Mumbai	India
4	2191	0.204	Sharjah International Airport	Sharjah	United Arab Emirates
	Airport ID	betweenness_centrality	Name	City	Country
0	2072	0.19	King Abdulaziz International Airport	Jeddah	Saudi Arabia
1	2188	0.183	Dubai International Airport	Dubai	United Arab Emirates
2	2997	0.137	Chhatrapati Shivaji International Airport	Mumbai	India
3	3093	0.133	Indira Gandhi International Airport	Delhi	India
4	813	0.129	OR Tambo International Airport	Johannesburg	South Africa
	Airport ID	voterank_centrality	Name	City	Country
0	2188	1	Dubai International Airport	Dubai	United Arab Emirates
1	2072	2	King Abdulaziz International Airport	Jeddah	Saudi Arabia
2	3093	3	Indira Gandhi International Airport	Delhi	India
3	2997	4	Chhatrapati Shivaji International Airport	Mumbai	India
4	2191	5	Sharjah International Airport	Sharjah	United Arab Emirates

Appendix

Key players



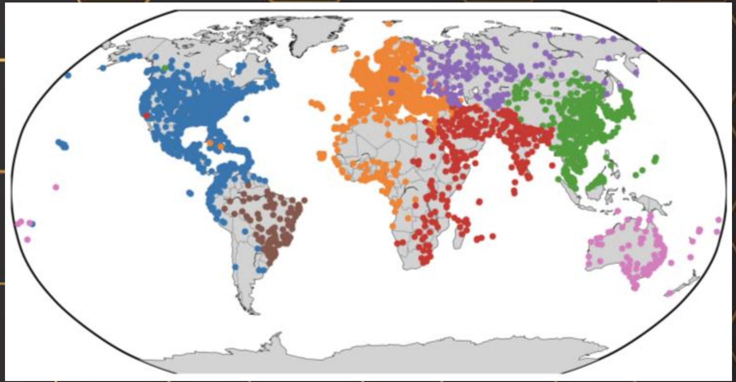
	Airport ID	degree_centrality	Name	City	Country
			purple region		
0	4029	0.722	Domodedovo International Airport	Moscow	Russia
1	2948	0.431	Pulkovo Airport	St. Petersburg	Russia
2	2975	0.312	Koltsovo Airport	Yekaterinburg	Russia
3	4078	0.243	Tolmachevo Airport	Novosibirsk	Russia
4	2988	0.236	Vnukovo International Airport	Moscow	Russia

	Airport ID	betweenness_centrality	Name	City	Country
0	4029	0.528	Domodedovo International Airport	Moscow	Russia
1	2948	0.183	Pulkovo Airport	St. Petersburg	Russia
2	2988	0.101	Vnukovo International Airport	Moscow	Russia
3	2975	0.095	Koltsovo Airport	Yekaterinburg	Russia
4	2923	0.054	Yakutsk Airport	Yakutsk	Russia

	Airport ID	voterank_centrality	Name	City	Country
0	4029	1	Domodedovo International Airport	Moscow	Russia
1	2948	2	Pulkovo Airport	St. Petersburg	Russia
2	2975	3	Koltsovo Airport	Yekaterinburg	Russia
3	2988	4	Vnukovo International Airport	Moscow	Russia
4	4078	5	Tolmachevo Airport	Novosibirsk	Russia

Appendix

Key players



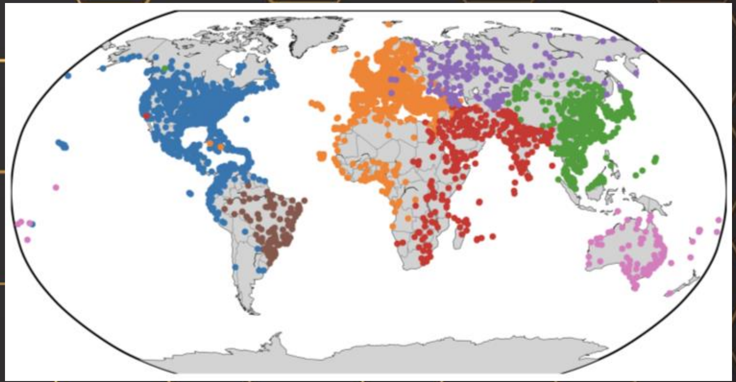
	Airport ID	degree_centrality	Name	City	Country
			brown region		
0	2578	0.521	Viracopos International Airport	Campinas	Brazil
1	2564	0.427	Guarulhos - Governador André Franco Montoro Int	Sao Paulo	Brazil
2	2531	0.396	Presidente Juscelino Kubistschek International Air	Brasilia	Brazil
3	2537	0.323	Tancredo Neves International Airport	Belo Horizonte	Brazil
4	2618	0.271	Congonhas Airport	Sao Paulo	Brazil

	Airport ID	betweenness_centrality	Name	City	Country
0	2578	0.291	Viracopos International Airport	Campinas	Brazil
1	2531	0.206	Presidente Juscelino Kubistschek International Air	Brasilia	Brazil
2	2564	0.141	Guarulhos - Governador André Franco Montoro Int	Sao Paulo	Brazil
3	2537	0.123	Tancredo Neves International Airport	Belo Horizonte	Brazil
4	2548	0.111	Marechal Rondon Airport	Cuiaba	Brazil

	Airport ID	voterank_centrality	Name	City	Country
0	2578	1	Viracopos International Airport	Campinas	Brazil
1	2564	2	Guarulhos - Governador André Franco Montoro Int	Sao Paulo	Brazil
2	2531	3	Presidente Juscelino Kubistschek International Air	Brasilia	Brazil
3	2537	4	Tancredo Neves International Airport	Belo Horizonte	Brazil
4	2618	5	Congonhas Airport	Sao Paulo	Brazil

Appendix

Key players



pink region					
	Airport ID	degree_centrality	Name	City	Country
0	3361	0.548	Sydney Kingsford Smith International Airport	Sydney	Australia
1	3320	0.376	Brisbane International Airport	Brisbane	Australia
2	3339	0.344	Melbourne International Airport	Melbourne	Australia
3	3341	0.215	Adelaide International Airport	Adelaide	Australia
4	3322	0.194	Cairns International Airport	Cairns	Australia
	Airport ID	betweenness_centrality	Name	City	Country
0	3361	0.462	Sydney Kingsford Smith International Airport	Sydney	Australia
1	3320	0.227	Brisbane International Airport	Brisbane	Australia
2	3341	0.163	Adelaide International Airport	Adelaide	Australia
3	3339	0.157	Melbourne International Airport	Melbourne	Australia
4	3351	0.121	Perth International Airport	Perth	Australia
	Airport ID	voterank_centrality	Name	City	Country
0	3361	1	Sydney Kingsford Smith International Airport	Sydney	Australia
1	3320	2	Brisbane International Airport	Brisbane	Australia
2	3339	3	Melbourne International Airport	Melbourne	Australia
3	3341	4	Adelaide International Airport	Adelaide	Australia
4	3322	5	Cairns International Airport	Cairns	Australia